UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXVIII CICLO DEL DOTTORATO DI RICERCA IN APPLIED DATA SCIENCE AND ARTIFICIAL INTELLIGENCE

Machine Learning and Cryptocurrency Markets: Methods and Evidence

Settore scientifico-disciplinare: INF/01

DOTTORANDO
Luca Pennella

COORDINATORE
PROF. Francesco Pauli

SUPERVISORE DI TESI
PROF. Nicola Torelli

CO-SUPERVISORE DI TESI
PROF. Letterio Galletta
Dr. Francesco Biasiol

ANNO ACCADEMICO 2024/2025

UNIVERSITÀ DEGLI STUDI DI TRIESTE

APPLIED DATA SCIENCE & ARTIFICIAL INTELLIGENCE

Università degli Studi di Trieste

Ph.D. in Applied Data Science & Artificial Intelligence

XXXVIII cycle

Machine Learning and Cryptocurrency Markets: Methods and Evidence

Candidate
Luca Pennella

Supervisors
Prof. Nicola Torelli
Co-Supervisors
Prof. Letterio Galletta
Dr. Francesco Biasiol

A Mamma, Papà, Martina e Valentina.

Summary

Cryptocurrency markets and blockchain-based financial infrastructures generate data at unprecedented scale and granularity, while also introducing new sources of risk, new market microstructures, and fast-evolving regulatory debates. At the same time, modern machine learning (ML) offers strong predictive performance but is often criticized for limited transparency, creating a recurring tension between accuracy and interpretability whenever model outputs may influence economic, legal, or policy decisions. This thesis develops and applies reproducible, interpretable, and domain-tailored methodologies at the intersection of explainable ML and digital-asset economics, organized around three complementary objectives: (i) designing explainable ML pipelines for complex socio-economic phenomena, (ii) characterizing investor heterogeneity and regulatory attitudes in crypto markets using international survey data, and (iii) measuring micro-level token circulation and systematizing decentralized derivatives protocols in Decentralized Finance (DeFi).

On the methodological side, the thesis proposes risk-sensitive and interpretable classification workflows that preserve predictive quality while supporting credible interpretation at both local and global levels. It introduces *X-SPIDE*, an explainable pipeline for detecting smart Ponzi contracts on Ethereum that combines heterogeneous feature families and shows that compact feature sets can retain strong discriminative performance while enabling structured post-hoc explanations useful for forensic analysis; to facilitate replication and follow-up research, the accompanying work releases the extraction pipeline and curated datasets used in the study. The same interpretability-first approach is transferred to computational political science through an explainable framework for voting-intention prediction based on repeated survey data, where ensemble models, systematic feature reduction, and SHAP-based explanations highlight that value-based attitudes retain substantial explanatory content beyond demographics and support interpretable, coalition-specific profiles. The thesis also addresses a key fragility of applied classification pipelines and imbalanced learning under class overlap, by showing via controlled simulations that oversampling effectiveness depends not only on imbalance ratios but also on data geometry, motivating overlap-aware preprocessing decisions rather than default oversampling. Complementing feature-attribution approaches, it introduces Decision Predicate Graphs (DPG), a model-specific global interpretability tool for tree ensembles that represents decision predicates as a graph and enables structural diagnostics such as centrality, communities, and class-specific constraints even when rule-based summaries become too large to inspect.

On the empirical side, the thesis uses international survey evidence to connect crypto-market narratives to individual behaviors and beliefs. Focusing on memecoins, it documents that memecoin holders form a distinct segment within the broader crypto popula-

tion, characterized by systematic demographic and psychological regularities that translate into identifiable trading practices, including a stronger reliance on socially mediated speculation and higher derivative usage. Building on a broader survey infrastructure, it also studies support for regulatory domains including oversight, KYC requirements, and taxation, showing that preferences are heterogeneous and systematically related to perceived market illegitimacy and to personal exposure to crypto wealth, which is associated with lower support for regulation.

On the DeFi side, the thesis develops protocol-tailored measurement tools and unifying conceptual frameworks. It introduces a micro-velocity methodology for Lido's liquid staking tokens, including share-denominated reconstruction needed for rebasing assets, and provides evidence of persistently high velocity coupled with strong concentration of turnover in a small set of large addresses; the corresponding pipeline and datasets are released to support reproducibility and reuse. The thesis further systematizes decentralized derivatives protocols through a unified representation of actors, flows, and design principles, and operationalizes the framework via a tuple-based formalism and a reproducible simulation environment. Overall, the thesis advances a coherent framework, interpretable ML by design, empirically grounded measurement, and reproducible artifacts, while being explicit about scope conditions: survey-based findings are time- and sample-dependent, simulation results invite validation under broader parameterizations, and DeFi markets evolve rapidly, motivating future work on robustness, cross-chain extensions, and continuous monitoring under changing market regimes.

Acknowledgements

I would like to thank my supervisors, Prof. Nicola Torelli and Prof. Letterio Galletta, for their valuable guidance, for always being available for discussion, and for giving me the freedom to explore and experiment throughout this journey.

I also thank Rachael and SWG, and in particular Francesco Biasiol, for supporting this PhD programme.

I thank the DEAMS, especially Francesco, Gino, Gioia, Leonardo, Roberta, Susanna, and Vincenzo, for always making me feel welcome, for their helpful advice, and for the opportunity to contribute to teaching activities.

I thank IMT School for Advanced Studies Lucca, especially Fabio, Lillo, and Pietro, for the work we have carried out together over these years, and for enabling me to deepen my knowledge and discover a research field that has become central to my interests.

I thank the Blockchain Center at the University of Zurich, and especially Claudio, for offering me a new experience, not only from a scientific perspective but also on a personal level. And thanks to Benji, sort of an Italian friend in Zurich, I owe you a pack of cigarettes.

Thanks to all my friends in Trieste for making me feel a little more at home during these years spent here.

Thanks to Leo, Luca and Ludovica for being by my side since we were kids.

Thanks to Carlo, Dario, Ennio, Ermanno, Giovanni, and Silvia for being a safe harbour and such precious presences; thanks to you, life is definitely more beautiful.

Thanks to Zia Angela and Christian, an extension of my family, without them, this milestone would not have been possible.

Thanks to Mamma and Papi: there are no suitable words. I could not have wished for more.

Thanks to Martina: these have been complex years, full of challenges, but together we can overcome anything. Life without you cannot exist.

Thanks to Valu, for loving me in a way words cannot describe; thank you for being close to me through these years filled with so much love, but also distance, for being home regardless of geography. I love you.

Contents

Chapter 1

Introduction

The last decade has witnessed the rapid emergence of digitally native financial infrastructures. Blockchain-based systems, cryptoassets, and Decentralized Finance (DeFi) protocols have expanded from niche experiments to a global ecosystem where value is issued, traded, and transformed through software [172, 319]. Over the same period, general-purpose AI technologies, ranging from mature machine learning (ML) techniques to more recent advances in large language models (LLMs) and agentic AI, have begun to reshape economic, social, and informational processes at scale. These developments have been enabled by the growing availability of large, complex datasets that must be collected, organized, and analyzed in order to make sense of contemporary socio-economic phenomena.

Within this broader landscape, blockchain-based finance occupies a somewhat distinct position. While AI and ML are already deeply embedded in many decision-making pipelines, the full impact of blockchain infrastructures and DeFi on financial intermediation is still unfolding. On the one hand, smart-contract platforms and tokenized assets have already created new markets, organizational forms, and governance mechanisms. On the other, the long-term configuration of this ecosystem, and its interaction with traditional finance, regulation, and everyday economic behavior, remains uncertain. In this sense, blockchain technologies combine realized change with a still-open range of possible futures.

This work aims to develop an empirical and methodological framework at the intersection of explainable machine learning and cryptocurrency markets to study complex socio-economic phenomena, with a specific focus on mechanisms and behaviors directly shaped by blockchain technologies, including DeFi

This thesis places at the intersection of these new developments: it views digital asset markets not only as a new financial domain, but also as a laboratory for testing and extending machine learning methods, with particular emphasis on interpretability, survey-based measurement, and protocol-level analytics [139, 204]. The overarching aim is to use ML and quantitative methods to broadly understand complex economic and social phenomena, with a specific focus on those aspects of the economy and society that are directly shaped by blockchain technologies, such as DeFi.

Digital asset markets differ from traditional financial ones along several dimensions that are directly relevant for empirical analysis and modeling. First, many economic interactions are fully recorded on public ledgers, giving rise to rich transactional and contractual data that can be processed and modeled at scale. Second, the programmability of

smart contracts enables new financial primitives, such as perpetual futures, automated market makers, and liquid staking tokens, whose behavior is constrained by code and governance rather than by centralized intermediaries [311, 319]. Third, participation is global, pseudonymous, and highly heterogeneous, combining retail investors, professional funds, and protocol treasuries within the same infrastructure. These features jointly create opportunities and challenges for ML-based analysis. On the one hand, they facilitate large-scale, fine-grained measurement; on the other, they raise questions about model robustness, biases, and interpretability in high-stakes settings such as fraud detection, risk management, and policy design [76].

A further complication is that digital asset markets are not only shaped by on-chain mechanics, but also by off-chain beliefs, narratives, and regulatory expectations. The same address-level transaction history can correspond to very different investor profiles and motivations. Understanding the behavior of crypto users and the social meaning attached to novel instruments therefore requires integrating behavioral and attitudinal data – for instance from surveys – with on-chain and market-based evidence. This motivates a multi-layered research agenda, in which methodological advances in explainable ML, empirical survey analytics, and DeFi measurement are developed jointly rather than in isolation [139, 204].

Across all parts, the thesis adopts interpretability and reproducibility as methodological constraints for empirical analysis in high-dimensional, rapidly evolving socio-economic systems. Concretely, it combines (i) explainable ML pipelines for risk-sensitive prediction and model understanding, (ii) survey-based measurement to recover interpretable investor and attitude profiles that cannot be inferred from on-chain traces alone, and (iii) protocol-level measurement and conceptual systematization to link behavioral outcomes to programmable market design. Together, these components form a unified framework to analyze blockchain-based markets as socio-technical systems, where observed outcomes emerge from the interaction between code-level mechanisms, heterogeneous participants, and evolving regulatory expectations.

The concrete trajectory of this thesis reflects these intertwined motivations. It originated in the study of smart-contract fraud on Ethereum, where the rarity of Ponzi schemes relative to the universe of contracts, and the intentional similarity between malicious and legitimate behavior, raise acute challenges of class imbalance, class overlap, and model interpretability. Building explainable detection pipelines in this context naturally led to questions about generalizability: to what extent can similar methods be transferred to other domains, such as the prediction of voting intentions, while preserving interpretability and robustness? In parallel, the need to model rare but substantively important outcomes motivated a close examination of oversampling strategies and their interaction with data geometry. Across these applications, a central concern has been to privilege models and workflows that remain as interpretable as possible, or that can be made transparently explainable, especially when predictions bear directly on economic or political decisions.

1.1 Motivations and Challenges

From a methodological perspective, a fundamental motivation of this work is the dichotomy between predictive performance and interpretability. In many application do-

mains relevant to digital asset markets, including financial fraud detection, risk modeling, and policy evaluation, opaque "black-box" models are difficult to justify to regulators, affected users, and domain experts. Explainable AI (XAI) seeks to address this issue by providing local or global explanations of model predictions, for example through feature importance measures, surrogate models, or structure-aware representations of decision logic [145]. However, in practice, XAI tools are often integrated into complex pipelines without a systematic assessment of their stability, fidelity, or interaction with data characteristics such as class imbalance and class overlap. In high-stakes settings, several authors have argued for the preference of interpretable or rigorously explained models over purely black-box predictors [261].

Fraud detection on blockchains illustrates these challenges clearly. Ponzi schemes and related scams are typically rare events in the space of all smart contracts, and malicious behavior can intentionally resemble legitimate activity [50, 85]. The resulting data are highly imbalanced and may exhibit substantial overlap between fraudulent and non-fraudulent classes. In such settings, reporting only aggregate performance metrics is insufficient: one must also understand why a classifier flags specific contracts as fraudulent, how robust those signals are, and how they interact with the underlying data geometry. Recent work on interpretable Ponzi detection models underscores this need, but typically remains confined to specific architectures or explanation tools [120, 237].

The need for transparent modeling extends well beyond fraud detection on-chain. Political behavior, public attitudes toward regulation, and survey-based measurement of investor types face analogous issues. Here, the primary units of analysis are individuals rather than contracts, but the methodological problems, high-dimensional covariates, complex interactions, and the risk of overfitting, are similar. When models are used to infer ideological structures, identify key determinants of voting intentions, or map support for regulatory regimes, interpretability is critical for substantive conclusions and for democratic accountability [139, 204]. This motivates a broader use of explainable ML pipelines that incorporate hyperparameter optimization, model comparison, and post-hoc explanation in a transparent and reproducible fashion, for instance via Shapley-value-based diagnostics and interpretable ensembles [206].

On the domain side, cryptocurrency markets raise at least three interconnected challenges. First, investor populations are heterogeneous and rapidly evolving. Phenomena such as memecoin trading are often dismissed as purely speculative, yet they may reveal deeper patterns of risk-taking, social belonging, and financial experimentation. Recent studies document how crypto investors differ from traditional retail traders in terms of demographics, financial literacy, and behavioral biases, and they highlight the existence of distinct investor archetypes [308]. Systematically profiling these users requires survey instruments that capture socio-demographic traits, behavioral patterns, and psychological dispositions, and analytical methods that can uncover interpretable investor profiles.

Secondly, the regulatory environment for cryptoassets is in the process of being established. The primary objective of policymakers is to achieve a balance between consumer and investor protection, financial integrity, innovation, and competitiveness. However, there is a lack of empirical evidence regarding how different segments of the crypto population perceive regulation, and how those perceptions relate to actual portfolio exposure and risk. Understanding these attitudes is crucial for anticipating acceptance or

resistance to policy proposals.

Third, at the protocol level, DeFi introduces new forms of financial intermediation that call for new measurement tools. Liquid staking tokens, for example, simultaneously encode claims on validator rewards and highly composable collateral in DeFi protocols [135]. The circulation patterns of these tokens have implications for systemic risk, liquidity, and governance. However, traditional monetary aggregates and velocity measures are insufficient for capturing the micro-level dynamics of token reuse and staking flows.

Similarly, DeFi derivatives, ranging from perpetual futures to options-like instruments and synthetic assets, implement payoff structures through smart contracts, liquidity pools, and oracle mechanisms that deviate in important ways from their traditional finance counterparts. A unified conceptual map of these designs is needed to compare protocols, understand risk transfer, and inform both practitioners and regulators.

Across these domains, the main challenge is to build a coherent analytical toolkit that integrates methodological development and empirical application. The thesis responds to this challenge by combining (i) advances in explainable ML for high-stakes classification and model interpretation, (ii) survey-based analytics of digital asset users and their regulatory preferences, and (iii) empirical measurement and theoretical systematization of DeFi protocols.

1.2 Research Objectives

This thesis is organized around three primary objectives, each corresponding to one part of the dissertation.

1.2.1 (i): Explainable machine learning for complex socio-economic phenomena

The first objective is to design and evaluate machine-learning pipelines that explicitly balance predictive performance and interpretability in settings where model outputs can influence economic or political decisions. This objective is articulated through the following research questions:

- How can we build an explainable detection pipeline for smart contract Ponzi schemes on Ethereum that remains competitive with purely performance approaches while yielding meaningful explanations of fraudulent behavior? This question is addressed in Chapter 2.

- To what extent can such pipelines be transferred to other domains, such as the prediction of voting intentions, while preserving interpretability and robustness? Chapter 3 investigates this transfer.

- How do class imbalance and class overlap interact with oversampling techniques such as SMOTE and ROSE, and under which conditions does oversampling improve or harm generalization? This issue is examined in Chapter 4.

- Can we construct global, structure-preserving representations of tree ensembles, such as Decision Predicate Graphs, that provide faithful and human-navigable summaries of decision logic? Chapter 5 introduces and evaluates this approach.

Chapters in the first part of the thesis address these questions by introducing X-SPIDE, extending explainable pipelines to the study of Italian voting intentions, analyzing the role of class overlap in oversampling for imbalanced data, and proposing Decision Predicate Graphs as a tool for global interpretation of tree-based ensembles.

1.2.2 (ii): Characterize investor types and regulatory attitudes in cryptocurrency markets

The second objective is to connect digital-asset markets with survey-based evidence on investor profiles and perceptions of regulation. The main research questions are:

- Do memecoin investors constitute a distinct subgroup within the broader crypto population in terms of socio-demographic traits, behavioral patterns, and psychological dispositions? Evidence is provided in Chapter 6.
- How are attitudes toward cryptomarket regulation, including oversight, KYC requirements, and taxation, distributed across different investor archetypes and levels of crypto wealth exposure? Chapter 7 provides the empirical analysis.

The second part of the thesis addresses these questions using global survey data on crypto users. It applies modern survey analytics to identify memecoin-specific investor types and to relate perceptions of cryptomarket regulation to observed behavioral and attitudinal profiles.

1.2.3 (iii): Measure micro-level circulation and systematize derivatives in DeFi

The third objective is to develop empirical and theoretical tools tailored to DeFi protocols. This leads to the following questions:

- How can we define and operationalize a notion of micro-velocity for liquid staking tokens that captures the intensity and concentration of token reuse across addresses and protocols? This challenge is affrontata in Chapter 8
- What does the on-chain circulation of Lido's stETH and wstETH reveal about user behavior, protocol composability, and the concentration of activity across different types of addresses? These empirical findings are reported in Chapter 8.
- How can we construct a unified conceptual framework that positions perpetual futures, options-like instruments, and synthetic asset protocols within a common design space, highlighting both analogies and departures from traditional derivatives? This framework is developed in Chapter 9.

The third part of the thesis develops a micro-velocity framework for liquid staking tokens, accompanied by open-source code and datasets, and proposes a systematization of DeFi derivatives that organizes existing protocols into a coherent taxonomy.

1.3 Publications

We list below the journal and conference papers that form the basis of this thesis published by the time of writing:

- Pennella, L., Pinelli, F., & Galletta, L. (2025). *X-SPIDE: An eXplainable Machine Learning Pipeline for Detecting Smart Ponzi Contracts in Ethereum.* IEEE Access [244].
- Pennella, L., & Fabbrucci Barbagli, A. G. (2025). *Explainable Machine Learning for Predicting Voting Intentions: A Study of Italian Politics.* International Journal of Data Science and Analytics [242].
- Pennella, L., Di Credico, G., & Torelli, N. (2025). *Exploring the Role of Class Overlap in Oversampling Methods for Imbalanced Data.* Proceedings of the Statistics and Data Science Conference [Waiting for Publication].
- Arrighi, L., Pennella, L., Marques Tavares, G., & Barbon Junior, S. (2024). *Decision predicate graphs: Enhancing interpretability in tree ensembles.* In World Conference on Explainable Artificial Intelligence (pp. 311-332). Cham: Springer Nature Switzerland [36].
- Balietti, S., Celebi C., Pennella, L., & Tercero-Lucas, D. (2026). *Meme Money, Real People: Decoding the Crypto Memecoin Crowd* Available at SSRN 6021706 [46].
- Kremer, S., Pennella, L., Yurdabak, M. K., & Balietti, S. (2025). *Public Perceptions of Cryptomarket Regulation: Investor Profiles and Attitudes.* Available at SSRN 5557139 [192].
- Kraner, B., Pennella, L., Vallarano, N., & Tessone, C. J. (2025). *Money in Motion: Micro-Velocity and Usage of Ethereum's Liquid Staking Tokens.* In 7th Conference on Advances in Financial Technologies (AFT 2025) (pp. 9-1). Schloss Dagstuhl–Leibniz-Zentrum für Informatik [188].
- Pennella, L., Saggese, P., Pinelli, F., & Galletta, L. (2025). *A Unified Framework and Comparative Study of Decentralized Finance Derivatives Protocols.* ArXiv e-prints. Art. no. arXiv:2512.19113. doi:10.48550/arXiv.2512.19113 [245].

1.4 Contributions

This thesis develops an empirical and methodological framework at the intersection of explainable machine learning and cryptocurrency markets. Across three complementary research streams, it contributes: (i) explainable machine learning pipelines for risk-sensitive classification and model understanding, (ii) empirical evidence on investor heterogeneity and regulatory preferences using international survey data, and (iii) measurement and conceptual tools for protocol-level analysis in DeFi, spanning token circulation and derivatives design.

1.4.1 Objective (i): Explainable ML for complex socio-economic phenomena

The thesis proposes and evaluates explainable pipelines that remain competitive while enabling auditability and interpretation. It introduces an interpretable framework for Ponzi smart-contract detection on Ethereum, demonstrates transferability of interpretability-first modeling principles to voting-intention prediction, and clarifies when oversampling improves generalization by showing that performance depends on data geometry and class overlap rather than imbalance alone. It also contributes a global interpretability method for tree ensembles, Decision Predicate Graphs, to support structure-aware inspection when rule lists become impractical.

1.4.2 Objective (ii): Investor heterogeneity and regulatory attitudes in cryptocurrency markets

Using international survey data, the thesis documents systematic heterogeneity in crypto participation and identifies memecoin holders as a distinct subgroup with recognizable demographic, behavioral, and psychological patterns. It further links regulatory preferences to individual profiles, showing that support for different regulatory domains varies meaningfully across the population and is associated with perceived market illegitimacy and exposure to crypto wealth.

1.4.3 Objective (iii): DeFi measurement and derivatives systematization

At the protocol level, the thesis develops a micro-velocity methodology tailored to liquid staking tokens, including share-denominated reconstruction for rebasing assets, and provides evidence on circulation intensity, concentration of turnover, and a progressive shift toward wstETH consistent with composability. It also proposes a unified framework for decentralized derivatives protocols that formalizes actors, flows, and design principles, operationalized through a tuple-based representation and a reproducible simulation environment for comparative analysis.

1.4.4 Scope and limitations

The empirical findings inherit domain-specific constraints. Survey-based results reflect the sampled time window and self-reported measures; evidence on overlap and oversampling is derived from controlled simulations and calls for broader external validation; and protocol-level analyses depend on public on-chain data and on rapidly evolving DeFi market structure. These limitations motivate future work on robustness across regimes, scaling to additional datasets and protocols, and extending the proposed tools to new market and regulatory contexts.

Part I

Explainable machine learning for complex socio-economic phenomena

Part I focuses on explainable machine learning as a methodological framework for studying complex socio-economic phenomena, where predictive accuracy alone is insufficient and model outputs must be interpretable, auditable, and scientifically meaningful. The aim is to design and apply explainability-oriented pipelines that bridge high-performing models with transparent, domain-relevant evidence.

This part brings together four complementary contributions. The first presents X-SPIDE, a framework for detecting Ponzi smart contracts on Ethereum that integrates heterogeneous on-chain and contract-level features with structured post-hoc explanations to support forensic analysis. The second study develops an interpretability-first approach to predicting voting intentions from repeated survey data, combining feature reduction and SHAP-based explanations to reveal value-based dimensions that remain informative beyond demographics and to derive interpretable coalition-specific profiles. The third investigation analyzes imbalanced learning under class overlap through controlled simulations, showing that the effectiveness of oversampling depends on data geometry and motivating preprocessing strategies that explicitly account for overlap. The fourth contribution introduces Decision Predicate Graphs (DPG), a global interpretability method for tree ensembles that summarizes decision logic as a graph and enables structural diagnostics when rule-based inspection becomes impractical.

Chapter 2

X-SPIDE: An eXplainable Machine Learning Pipeline for Detecting Smart Ponzi Contracts in Ethereum

2.1 Introduction

Blockchain is revolutionizing how individuals and companies exchange digital assets without the control of a central authority. This technology has been successfully exploited for deploying new economic applications, e.g., cryptocurrencies [225] and Decentralized Finance [229]. However, shortly after this technology gained widespread adoption and its economic significance grew, it began to attract the interest of malicious users seeking to exploit the pseudonymity of these platforms and the absence of regulation [221]: on the one hand, they utilize cryptocurrencies to facilitate untraceable currency transfers, evading scrutiny by authorities; on the other hand, they deliver scams to deceive honest users willing to make revenue through cryptocurrencies. Nowadays, many types of scams can be found on blockchain platforms, such as exploits, hacks, and phishing [51]. It is estimated that fraudulent activities in the Bitcoin space amassed a minimum of 11 million USD between 2011 and 2014 [299]. Among the various scams, Ponzi schemes have approached the blockchain world, first on Bitcoin [299] and more recently on Ethereum [50]. *Ponzi schemes* are fraudulent investment operations where older investors obtain returns from new investors' money rather than legitimate business activities. Although the actual conditions to gain money depend on the specific rules of the scheme, a common feature is that participants who want to redeem their investments have to make new participants join the scheme. Participants who join later are the most likely to lose their money. Thus, the development of automatic techniques to counter these scams is required to protect average users and to allow them to participate safely in the blockchain economy. Our focus here is on Ethereum, one of the most famous blockchain systems, and smart contracts to deliver Ponzi schemes called *smart Ponzi contracts*. Although several papers have addressed Ponzi scheme detection on Ethereum [49, 86, 87, 203, 306, 324], there is still no conclusive solution. On the one hand, there is a lack of publicly available datasets that can be used to train and test effective classifiers. On the other hand, the problem has been tackled so far by considering only the optimization of the classifier's performance as a target, paying no attention to

explaining and interpreting classification results. Interpretability and explainability are crucial when classifier decisions have economic consequences. An automatic and explainable technique for classifying smart contracts is needed to fill such gaps, which can be safely used as the backbone for developing new fraud detection tools.

Contributions In this chapter, we contribute towards filling this gap by introducing X-SPIDE (*XAI Smart Ponzi Identification and Detection*), an explainable machine-learning pipeline to detect smart Ponzi contracts. Besides the classification of contracts and Ponzi scheme detection, X-SPIDE allows for comparing the results of different classifiers trained on different datasets. Moreover, it provides a procedure to refine the set of features while maintaining comparable classification performance to the original feature set. Finally, it relies on *eXplainable Artificial Intelligence* (XAI) techniques and tools to explain the model and carefully inspect misclassified contracts. We train and test our pipeline on a new and reusable dataset featuring 7446 unique real-world smart contracts. Within this dataset, 6566 (88.18%) are identified as non-Ponzi contracts, and 880 (11.82%) as Ponzi contracts. The dataset contains information about the transaction history of the contracts as well as their bytecode. We apply a thorough cleaning process on this dataset to detect and remove duplicates and non-compliant contracts (some bytecodes were not retrievable as the contracts appeared destroyed or were not recoverable on the blockchain). Then, we split the clean dataset into eight distinct datasets characterized by different bytecode features but the same transaction features. More precisely, four datasets incorporate features extracted from the *creation bytecode*, i.e., the code used to create and deploy the contract on the blockchain (creation code) plus the code of the contract itself (runtime code) and transaction features. The other four datasets instead include solely features obtained from the *deployed bytecode* (runtime code only) and the transaction features. We conduct several experiments on these datasets to train and test our machine learning pipeline. As a side effect, our experiments also allow us to answer the following six research questions: (1) Which is the pair (model, dataset) that provides the highest Recall?; (2) How does the usage of creation code and deployed code impact classification performance?; (3) What is the impact of account features on the classification performance?; (4) Does a subset of features exist that ensures performance comparable to the entire dataset? (5) What are the most important features?; (6) What are the characteristics of the misclassified contracts?

In conclusion, the main contributions of our work can be summarized in three main aspects: first, we define a new machine learning pipeline that incorporates eXplainable Artificial Intelligence (XAI) outputs to elucidate model behavior. Second, we introduce a new dataset that is publicly available for use. Last, through experiments, we analyze the different performances between created and deployed bytecode, investigate the significance of features, and analyze the key characteristics of misclassified contracts.

Structure of the contribution In Section 2.2, we provide the reader with the relevant background to understand the technical development of the experiment. In Section 2.3, we compare our work with the relevant literature. Section 2.4 formally defines our pipeline. In Section 2.5, we introduce our datasets, our features, and how we collect them. Section 2.6 reports on our experimental evaluation and the results we achieved. Finally, in Section 2.7, we draw some conclusions and discuss possible future work.

This is an extended version of the conference paper [130]. We extend the previous work in several directions: (1) we precisely define our pipeline X-SPIDE, characterizing what the input and the output of each stage are; (2) we introduce here a large dataset comprising more smart contracts and updated values of the features; we also devise a systematic process that cleans the dataset by addressing duplicates and non-compliant contracts; (3) our machine learning pipeline work considering both transaction and bytecode features; (4) we study how different types of bytecode impact classification performance; (5) we experiment with a different set of feature combinations (see Table 2.2), selecting the best one; (6) we study the interpretability of our model using different XAI techniques such as Shapley values and partial dependency plots, and we provide an interpretation of the results considering the smart contract's bytecode.

2.2 Background

Here, we provide the required information about blockchain systems, the Ethereum platform, and Ponzi schemes to clarify the context and the terminology for readers unfamiliar with this application domain.

2.2.1 A glimpse of Ethereum

A blockchain platform is a dynamic network of peer-to-peer nodes that maintain a replicated data structure, called the *blockchain*, that globally records the occurrence of certain events. The network nodes own a local copy of the blockchain and update it upon reception of special messages, called *transactions*. To ensure the blockchain's consistency, these systems rely on a consensus protocol that imposes a total order on the updates performed by the nodes.

Ethereum is one of the most famous blockchain platforms. In the recent version, it uses a Proof-of-Stake consensus algorithm where special nodes, called validators, propose blocks containing transaction bundles. The consensus is designed to be cryptographically and economically secure, requiring potential attackers to hold a significant amount of ether (ETH in short), the Ethereum cryptocurrency. A reward system incentivizes honest participation, while penalties deter malicious behavior among stakers.

A distinguishing feature of Ethereum is that it provides a decentralized *virtual machine* that can execute programs, called *contracts* and written in the EVM bytecode language. The EVM bytecode is a low-level language where a program/contract is a sequence of instructions, each characterized by its own opcode and its operands on the stack. From a technical point of view, Ethereum implements a state transition system. The current state of this transition system is stored on the blockchain and includes all the accounts and their balances. There are two types of accounts: the *external accounts* that users control and *contract accounts* that are controlled by a contract's bytecode. Intuitively, contracts can be seen as objects in object-oriented languages: they have fields representing the state of the contract and a set of functions that users or other contracts can invoke. We discuss an example of a smart contract in the next sub-section. An account is uniquely identified by its *address*. Every account is equipped with a particular field, called the *balance*, which represents the amount of ETH owned by the account. Users can send transactions, called *external transactions*, to the Ethereum net-

work. Through transactions, a user can *(i)* create new contracts; *(ii)* invoke a function of a contract; *(iii)* transfer Ether to contracts or other users. All external transactions are recorded on the blockchain. When a contract receives an external transaction, it reacts by executing one of its functions that, in turn, fires other transactions. These transactions are called *internal transactions* and are not recorded on the blockchain but still impact the balance of users and the state of other contracts.

2.2.2 Smart Ponzi contracts

Ponzi schemes are classic frauds concealed as "high-yield" investment programs. The initiator of the scheme generates returns for existing investors through revenue paid by new investors rather than from legitimate business activities or profits of financial trading. More in general, the U.S. Securities and Exchange Commission (SEC)[1] defines Ponzi schemes as:

> A Ponzi scheme is an investment fraud that involves the payment of purported returns to existing investors from funds contributed by new investors. Ponzi scheme organizers often solicit new investors by promising to invest funds in opportunities that generate high returns with little or no risk. With little or no legitimate earnings, Ponzi schemes require a constant flow of money from new investors to continue. Ponzi schemes inevitably collapse, most often when it becomes difficult to recruit new investors or when a large number of investors ask for their funds to be returned.

Although the actual conditions to gain money depend on the specific rules of the scheme, a common feature is that a user who wants to redeem her investment has to make new users join the scheme. In this way, the schemes create a pyramid of investors, where the initiator is at the top, and the investors at level $l + 1$ compensate for the investment of those at level l. Once a scheme collapses because no more investors join it, those at the top levels of the pyramid gain money, while those at the bottom lose it.

The spread of cryptocurrency and smart contracts has created new opportunities to deploy this kind of fraud. Indeed, it is possible to find samples of smart contracts implementing Ponzi schemes, called *smart Ponzi contracts*, deployed on the main blockchain platforms like Ethereum and Bitcoin. This application focuses on the Ethereum platform. According to the literature [49] smart Ponzi contracts have several attractive features that make them useful for delivering scams:

1. The initiator of a smart Ponzi could stay anonymous, since deploying the contract on the blockchain and withdrawing money from it only requires an Ethereum account that does not reveal her real identity.

2. Once deployed on the blockchain, smart contracts are "unmodifiable" and "unstoppable". Thus, no central authority could terminate the execution of the scheme, seize the money, and refund the victims.

3. Since the code of smart contracts is public and immutable, and its execution is automatically enforced by the blockchain platform, investors may believe that no

[1] See https://www.sec.gov/spotlight/enf-actions-ponzi.shtml

one can take advantage of their money and that they could eventually gain the declared interests.

The most significant feature of a smart Ponzi is the policy used to redistribute new investments among participants, i.e., how the money flows. This requires a smart Ponzi to maintain a data structure storing participants' information and to implement a strategy for redistributing dividends.

Identifying redistribution behavior is crucial to classify a contract as a smart Ponzi. Also, it is challenging because many other kinds of contracts, e.g., gambling games, may have similar behavior, which may induce many false positives. Bartoletti et al. [50] proposed the following four requirements to classify a smart contract as a Ponzi scheme:

R1 the contract redistributes money to the investors according to a given logic;

R2 the contract receives money only from the investors;

R3 each investor makes a profit if a certain number of investors subsequently join the contract, investing a certain amount of money;

R4 the later an investor joins the contract, the higher the risk for the investor to incur a loss.

A smart contract is classified as a smart Ponzi when it satisfies *all* four requirements. Note that requirement R1 rules out contracts that provide users with some assets but do not implement a logic to distribute them to participants, e.g., tokens; requirement R2 ensures that a participant invests a certain amount in joining the contract; requirement R3 demands a constant flow of new investments for investors to make a profit; requirement R4 characterizes the fraudulent nature of smart Ponzi contracts because it reflects the fact that making a profit for investors is likely impossible after a certain point in time: too many victims must join the scheme for the contract to have enough money to reward all the participants. Thus, the scheme collapses when this happens. Note that the requirements above impose no condition on whether the money was received or not by the initiator of the scheme. We will study the need for such a condition in our experimental evaluation of Section 2.6.

Typically, smart Ponzi contracts are categorized into four types according to their redistribution strategies: *Tree-scheme*, *Chain-scheme*, *Waterfall-scheme*, and *Handover-scheme*. Since we do not consider these categories here, we refer the interested reader to the relevant literature [50].

We now clarify a smart Ponzi contract through a small yet real example. Consider the code snippet in Figure 2.1. The code is written in Solidity, a high-level, object-oriented programming language primarily used for creating smart contracts. It features static typing and supports inheritance, reusable libraries, and complex user-defined types, ultimately compiling into EVM bytecode. The code is extracted from a contract named *Multiplier* that is deployed on Ethereum with the address

`0x30D1B797365F936300055A704A902124467B8b14`

and implements a chain scheme. The code snippet resembles a class declaration and consists of two parts: the declaration of fields and data, and the declaration of two functions.

```
contract Multiplier {
  //Address of the "promoter" of the contract: she receives fee for each transaction
  address constant private PROMO = 0x5D5fe29339592eEb51c43E54F0a81cA7642B6d2b;
  //Percent received by the "promoter"
  uint constant public PROMO_PERCENT = 7;
  //How many percent for your deposit to be multiplied
  uint constant public MULTIPLIER = 121;

  //The deposit structure holds all the info about the made deposits
  struct Deposit {
      address depositor; //The depositor address
      uint128 deposit;   //The deposit amount
      uint128 expect;    //How much we should pay out (initially it is 121% of deposit)
  }

  Deposit[] private queue;  //The queue of investors
  uint public currentHead = 0; //The index of the first depositor in the queue.

  //This function receives all the deposits stores them  in the queue and make immediate payouts
  function () public payable {
      if(msg.value > 0){
          //Add the investor into the queue. Mark that he expects to receive 121% of deposit back
          queue.push(Deposit(msg.sender, uint128(msg.value), uint128(msg.value*MULTIPLIER/100)));

          //Send some promo to enable this contract to leave long-long time
          uint promo = msg.value*PROMO_PERCENT/100;
          PROMO.send(promo);

          //Pay to first investors in line
          pay();
      }
  }

  //Used to distribute all the money on contract to the first investors starting from the head of queue
  function pay() private {
      // The current balance of the contract
      uint128 money = uint128(address(this).balance);

      //Cycle on the queue
      for(uint i=0; i<queue.length; i++){
          uint idx = currentHead + i;  //the index of the currently first investor

          Deposit storage dep = queue[idx]; //the info of the first investor

          if(money >= dep.expect){  //If we have enough money on the contract to fully pay to investor
              dep.depositor.send(dep.expect); //Send money to him
              money -= dep.expect;            //update money left

              //the investor is fully paid and she is removed from the queue
              delete queue[idx];
          }else{
              //No enough money, so partially pay to investor
              dep.depositor.send(money); //Send to her the money left
              dep.expect -= money;       //Update her expected amount
              break;
          }
      }

      currentHead += i; //Update the index of the current first investor
  }
}
```

Figure 2.1: Example of Smart Ponzi Contract: the code of the *Multiplier* contract

Lines 2 to 18 are constant and variable definitions used to record the state of the contract: PROMO is the address of the initiator of the contract; PROMO_PERCENT is the percentage of the investment that each participant pays to the initiator as a fee; MULTIPLIER is a constant denoting how much participant's investment must be multiplied; the structure `Deposit` records the address, the amount of deposit, and the expected reward of each investor. The array `queue` stores all the investors in order of arrival, and the integer `currentHead` is the index of the head of the queue. The contract receives Ether from investors through a transaction, and when this happens, it executes the special function with no name, called *fallback function* at line 20.

The received Ether is automatically transferred to the contract balance that is accessible to the programmer via the read-only variable `balance`. Information about the received transaction is accessible via the special object `msg`. In this case, the fallback function, if the received amount is positive (i.e., `msg.value > 0`), records the address (`msg.sender`), the investment and the expected reward of the investor (computed taking into account the factor MULTIPLIER) in the queue (line 23). Then, it calculates and pays the fee to the initiator (lines 28-27). Finally, it pays previous investors through the private function `pay`. This function (lines 35-61) scans the queue and pays the expected revenue to the investors as long as the balance of the contract is positive. Once an investor has received her reward, she is removed from the queue. From this example, it is easy to see that the scheme's initiator receives some money for each new investor and that investor A will be repaid only when she is at the front of the queue. This event occurs only when enough new investors join the scheme to repay the old investors preceding A in the queue.

2.3 Related work

Since the inception of Bitcoin in 2009, cryptocurrencies and blockchain systems have attracted the attention of cybercriminals, who exploit them to carry out potentially untraceable scams. Since the full transaction history is publicly available and provides accurate records of user behavior, several papers [51, 71, 85, 86, 87, 118, 166, 171, 174, 175, 199, 203, 241, 290, 306, 307, 317, 320, 324] have proposed machine learning techniques to detect possible frauds and scams. Below, we describe the contribution of these papers, and then we highlight the distinctive features of our work compared to them.

Bartoletti et al. [49] study the problem of identifying Ponzi schemes in Bitcoin using data mining techniques. In particular, they released a public dataset of Bitcoin addresses and an open-source tool to build such a dataset; then, they apply different classification algorithms (Random Forest, Bayes Network, and Ripper) and systematically evaluate them to identify the best discriminating features for detecting Ponzi schemes in Bitcoin. In a subsequent paper, Bartoletti et al. [50] consider smart Ponzi contracts in Ethereum. First, they define four behavioral criteria that characterize a contract as a smart Ponzi and produce a dataset with several samples satisfying such criteria. Then, they perform several analyses by hand on a subset of their dataset, e.g., about the security of their code and the fairness of the distribution policy. Chen et al. [86] provide a reusable dataset with real-world samples and evaluate different classifiers to determine which presents the best classification performance. They use two different classes of features: the account features taken from the transaction history and code features extracted from the

contract's bytecode. Chen et al. [87] propose SADPonzi, a detection approach based on symbolic execution. This approach analyzes the bytecode of contracts to extract semantic information about the execution and identifies investor-related transfer behavior and the distribution strategies adopted by the scheme. SADPonzi performs the classification only by looking at the code, not the transaction history. Wang et al. [306] propose a classifier based on Long-short Term Memory Network for detecting smart Ponzi. Fan et al. [118] propose PonziTect, a smart Ponzi contract detection method based on ordered boosting that classifies contracts considering only their bytecode. They adopt data augmentation to solve the problem of imbalanced samples among the two classes by increasing the proportion of smart Ponzi contracts at the boundary. Lou et al. [203] use convolutional neural networks to build a smart Ponzi detector. They focus mainly on bytecode features, and their pipeline is quite standard: First, they transform smart contracts into single-channel images and then adopt the spatial pyramid pooling method to ensure that the generated images have the same size. Ibba et al. [166] builds a machine learning model that uses account features and the bytecode of the contracts. They also consider Solidity source code and apply text classification techniques to extract further features to use in the classification. They tested their approach with decision trees, support vector machines, and naive Bayes. Jin et al. [174] propose HFAug, a generic Heterogeneous Feature Augmentation module that can be adapted to various existing Ponzi detection methods. The module captures heterogeneous information associated with account behavior. Zheng et al. [324] propose MulCas, a method for detecting smart Ponzi schemes that works solely on the bytecode. Peng et al. [241] use eight classification algorithms such as Logistic Regression, Decision Trees, Support Vector Machine, Random Forests, Extremely Randomized Trees, Gradient Boosting Machines, XGBoost, and LightGBM to build a model that detects Smart Ponzi schemes. They focus solely on the functionality based on the opcodes of smart contracts on Ethereum. Zhang et al. [320] introduce a method for identifying smart Ponzi using an enhanced version of the LightGBM algorithm. Their approach combines bytecode features with user transaction data and opcode frequencies. Chen et al. [85] focused on extracting features from user accounts and operation codes from smart contracts to build a classification model for detecting latent Ponzi schemes. Jin et al. [175] propose a dual-channel early warning framework dubbed Ponzi-Warning. Their proposal performs feature extraction and fusion on code and transaction features. They present a temporal evolution augmentation strategy for generating transaction graph sequences, increasing the data scale, and introducing temporal information. Yu et al. [317] model Ponzi scheme identification and detection as a node classification task and propose a detection model based on graph convolutional networks. Jacinta et al. [171] propose an approach that detects Ponzi schemes on Ethereum using random forest, neural network, and K-nearest neighbor. Cai et al. [71] devised a methodology that employs a graph to represent transaction-related semantics within a smart contract and uses a graph convolutional network to identify potential Ponzi-like transaction patterns. Wang et al. [307] present a detection approach that integrates opcode context analysis with the Adaptive Boosting (AdaBoost) algorithm. The methodology employs the n-gram algorithm to capture extensive opcode features related to contracts and integrates them with features derived from contract accounts. Below, we compare our work with the above-cited papers. This paper's main novelties and distinctive features are the introduction of an explainable machine learn-

ing pipeline and the study of the impact of different types of bytecodes on classification performances. Moreover, while other papers in the literature focus solely on optimizing classification performance, we prioritize finding the right balance between performance and interoperability. We believe that this focus becomes crucial when decisions can impact economic systems like Ethereum. More precisely, we consider both account and code features to detect smart Ponzi schemes. Additionally, we enhance the existing dataset by introducing new transaction features to capture better requirements R1-R4 of Section 2.2 and incorporating additional contracts. Furthermore, we integrate various Explainable Machine Learning techniques to foster a transparent understanding of the decision-making of our model. Going beyond traditional analysis, we scrutinize the classifier's response to misclassified contracts. Through these techniques, we aim to pinpoint a subset of features demonstrating comparable performance and contribute to an enhanced understanding and interpretation of the classification model. Additionally, we encourage further investigation and exploration within this domain by publicly making our dataset available for further research. Finally, Feng et al. [119] propose IDPonzi, an interpretable model for detecting Ponzi schemes on the blockchain. Compared to our approach, their dataset is smaller than ours, containing 200 Ponzi schemes, whereas our dataset includes 880. Moreover, their model utilizes only bytecode features and achieves lower performance in terms of Recall. From an explainability perspective, in contrast to theirs, our pipeline includes not only SHAP values but also partial dependence plots. Table 2.1 provides an overview of the characteristics of various related works and highlights the key differences, both among themselves and in comparison to this contribution. A checkmark (✓) means the presence of a specific feature, while a crossmark (✗) indicates its absence.

2.4 Methodology and pipeline design

This section introduces X-SPIDE, an interpretable machine-learning pipeline for identifying smart Ponzi contracts. Our pipeline goes beyond just classifying contracts and detecting Ponzi schemes, it also allows for *(i)* an easy comparison of the results produced by various classifiers trained on different datasets; *(ii)* feature refinement to reduce the feature set used by the model while ensuring that the classification performance remains comparable to the original set; *(iii)* explainability through XAI techniques and tools to clarify the model's decisions and thoroughly examine any misclassified contracts. More precisely, our pipeline consists of three stages: S_1 model and dataset selection; S_2 feature selection; S_3 model explanation. The first stage S_1 receives as input a set of classifiers C, a set of datasets D, and an evaluation metric m (e.g., Recall, Precision, AUC, etc.). It trains and tests all combinations of classifiers and datasets using a grid-search and cross-validation procedure and returns in output the classifier $c_i \in C$ (with its optimized hyperparameters) and the dataset $d_j \in D$ that outperforms the other combinations according to the metric m.

Formally, let CL be the set of classifiers, DS the set of datasets, M the set of evalua-

Table 2.1: A summary comparing related work with the present study.

Authors	Year	Code Features	Account Features	Creation vs. Deployed Bytecode	Data availability	XAI
Chen et al. [85]	2018	✓	✓	✗	✓	✗
Chen et al. [86]	2019	✓	✓	✗	✓	✗
Peng et al. [241]	2020	✓	✗	✗	✗	✗
Fan et al. [118]	2020	✓	✗	✗	✗	✗
Lou et al. [203]	2020	✓	✗	✗	✗	✗
Chen et al. [87]	2021	✓	✗	✗	✓	✗
Wang et al. [306]	2021	✓	✓	✗	✗	✗
Ibba et al. [166]	2021	✓	✓	✗	✗	✗
Zhang et al. [320]	2021	✓	✓	✗	✗	✗
Yu et al. [317]	2021	✗	✓	✗	✗	✗
Jin et al. [175]	2022	✓	✓	✗	✗	✗
Zheng et al. [324]	2022	✓	✓	✗	✓	✗
Jacinta et al. [171]	2023	✓	✗	✗	✗	✗
Cai et al. [71]	2023	✓	✓	✗	✗	✗
Wang et al. [307]	2023	✓	✓	✗	✗	✗
Feng et al. [119]	2024	✓	✗	✗	✗	✓
This contribution	2025	✓	✓	✓	✓	✓

tion metrics, we define the function $S_1 \colon \wp(CL) \times \wp(DS) \times M \mapsto CL \times DS$ as follows:

$$
\begin{aligned}
&S_1(C, D, m) = (c_i, d_j) \\
&\text{where} \quad c_i \in C \text{ and } d_j \in D \\
&\text{and} \quad \forall c_k \in C \setminus \{c_i\}. \forall d_k \in D \setminus \{d_j\}.\ m(c_i, d_j) \geq m(c_k, d_k)
\end{aligned}
$$

This stage produces the best pair of classifiers and a dataset, and allows for other analyses: we can investigate how different classifiers perform on various datasets, gaining insights from the features within each dataset. Moreover, considering different metrics, the stage helps understand which dataset is more suitable for optimizing a specific metric. For instance, a certain dataset with specific features might yield a higher recall than others. This dataset could be a better choice if the analyst aims to minimize false negatives by ensuring that as few threats as possible go undetected.

The second stage of the pipeline S_2 takes in input the classifier c_i and the dataset d_j produced by S_1, and it finds a subset of the features of d_j that increases or, at least, preserves the performance achieved by the classifier c_i when trained on d_j. Intuitively, S_2 adopts the Recursive Feature Elimination (RFE) algorithm that trains and tests c_i in several runs to detect the less contributing features to be removed in the next run. This process continues until the performances remain within a given threshold ϵ or the number of features to be considered reaches 10. Formally, let F_d denote the set of features of a dataset d, we define the function $S_2 \colon CL \times DS \times M \mapsto DS$ as follows:

$$
\begin{aligned}
&S_2(c, d, m) = d' \\
&\text{where} \quad |m(c, d') - m(c, d)| \leq \epsilon \text{ and } F_{d'} \subseteq F_d \text{ and } \epsilon \geq 0
\end{aligned}
$$

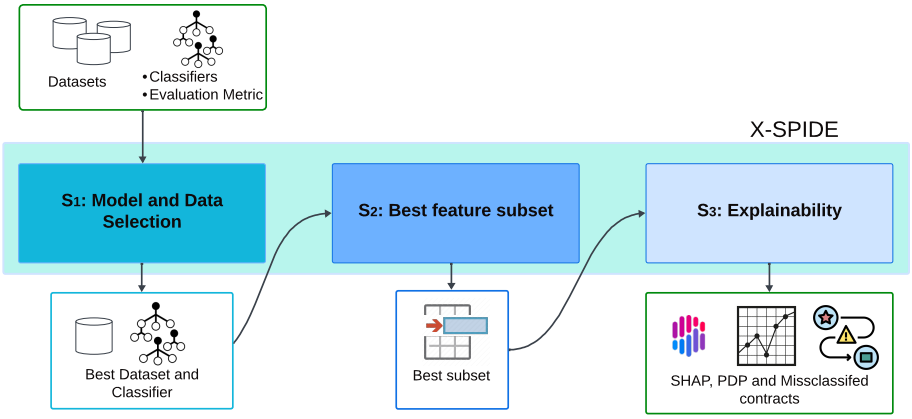


Figure 2.2: Structure and dataflow of the X-SPIDE pipeline

The stage S_2 of the pipeline allows us to identify a smaller set of features that ensures a good classification performance with respect to the optimized metric, to work with a more manageable dataset, and to improve the time for training and testing. Moreover, this stage also suggests which features mainly contribute to an accurate classification.

The stage S_3 provides results that can be used to describe the model behavior. In particular, S_3 computes the Shapley values Sh, partial dependencies PD for each smart contract for all considered features, and the misclassified contracts MC. More formally, we can define S_3 as the following:

$$S_3(c, d) = (Sh, PD, MC)$$

where c is a classifier and d is a dataset. We describe below how S_3 computes the elements of the resulting triple. The computation of Shapley values Sh assesses the contribution of each feature to the model's predictions, evaluating the model against all possible feature combinations. In practice, the Shapley value of a feature i for a specific instance x (a smart contract in our case), in symbols $Sh(x, i)$, is calculated as follows:

$$Sh(x, i) = \sum_{S \subseteq \{1,\dots,p\} \setminus \{i\}} \frac{|S|!(p - |S| - 1)!}{p!} [c(x_S \cup \{i\}) - c(x_S)]$$

where

- S represents a feature subset excluding feature i;
- $|S|$ denotes the cardinality of the set S;
- x_S is the sample x restricted to features in subset S;
- $c(x_S \cup \{i\})$ is the classifier prediction when feature i is included;
- $c(x_S)$ is the classifier prediction without feature i;
- p is the total number of features.

Intuitively, the formula above computes the contribution of feature i by comparing predictions with and without i across all possible feature subsets S, weighed by the subset size and total feature count. Finally, these contributions are summed to determine the Shapley value for feature i. Therefore, the set Sh includes all the $Sh(x, i)$ for all features

i and all the instances x in the dataset d. Shapley values are considered as XAI global and local models.

The partial dependencies PD determine the relationship between a feature i and the predicted class, i.e., Ponzi or non-Ponzi. The partial dependence of a feature i is computed as follows:

$$PD(i) = \frac{1}{|d|} \sum_{m=1}^{|d|} \left(\frac{1}{|S|} \sum_{w \in S} c(x[i \mapsto w]) \right)$$

where

- $|d|$ is the total number of samples in the dataset;
- S is the set of values taken by the feature i in the dataset d;
- $|S|$ is the size of S;
- $x[i \mapsto w]$ is the sample x where the value of the feature i is replaced with w (the values of the other features are unchanged).

Intuitively, the formula above is computed by varying the value of the feature of interest and calculating the average classifier prediction while keeping the value of other features constant. These values are useful for interpreting classifier responses to feature changes and can reveal non-linear or complex relationships between the features and the model output. The partial dependencies are XAI global models since they consider all instances and give a statement about the global relationship of a feature with the predicted outcome.

The set of Sh allows us to plot the model's global behavior as beeswarm plots of the feature importance. The set of DP helps investigate the relationship between the target response and a selected set of input features through Partial dependence plots (PDPs). Finally, stage S_3 helps study misclassified contracts MC and gain further insights regarding the behavior of the model and the specificity of these contracts. In particular, it applies SHAP on the subsets False Positive and False Negative independently, then it retrieves their code, enabling manual inspection by analysts, and finally, it supports the study of the feature importance for specific contracts.

Figure 2.2 summarizes the structure of our pipeline and describes what each stage takes as input and returns as output. In Section 2.6, we implement our pipeline using Python and the Scikit-learn library, and we adopt XAI library Shap[2] and partial dependence plots (PDPs) [126].

2.5 Dataset and feature descriptions

The contracts used to train and test our pipeline for the classification of Ponzi and non-Ponzi contracts are obtained by merging the lists of addresses and the corresponding labels of refernces [50, 86, 87, 174, 324]. We obtained a new list with 7962 unique contracts. However, some bytecodes were not retrievable as the contracts appeared destroyed. After compiling a comprehensive list of accessible smart contracts, we download the values

[2] https://shap.readthedocs.io/en/latest/index.html

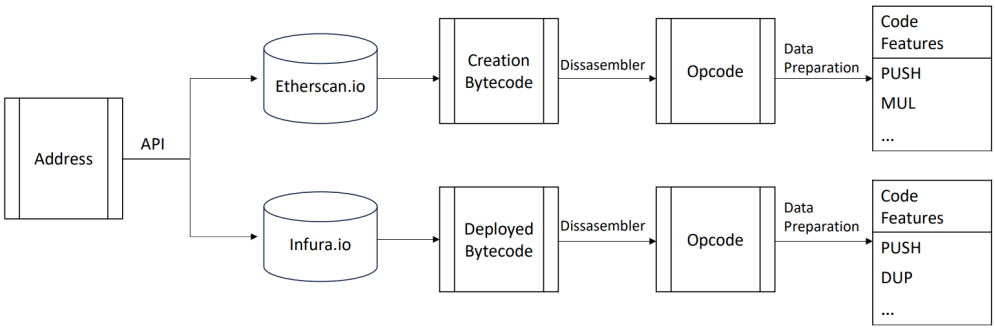


Figure 2.3: Scheme of the framework implemented for opcode extraction.

of the relevant features. Upon completing this initial analysis and cleanup, we obtain a dataset with 7446 contracts, comprising 6566 non-Ponzi and 880 Ponzi contracts.

Each contract is enriched with 5 types of features: (1) creation bytecode with absolute frequency; (2) creation bytecode with weighted frequency; (3) deployed bytecode with absolute frequency; (4) deployed bytecode with weighted frequency; (5) account and transaction history features. The first four types are obtained by analyzing the bytecode associated with each smart contract. The last one takes into account the information associated with each smart contract about its activities on the blockchain. We used eight possible combinations of these families of features, and each combination led to a dataset. Applying X-SPIDE to different combinations of features allows us to understand the contribution of each feature's family to the classification process and compare the relative performance. In the following, we describe the process and the type of features we extracted from the bytecode and transaction activities.

2.5.1 Code features

Bytecode is an abstract instruction set designed for efficient execution by a software interpreter or a virtual machine. Unlike human-readable source code, bytecode is expressed in a numerical format: the bytecode comprises a series of bytes, each of which refers to a specific operation whose semantics is specified by the Ethereum yellow paper [312]. For example, the operation with opcode $0x02$ represents the multiplication at the EVM level (denoted with the name MUL).

The reason for adopting bytecode is its widespread use in various aspects of smart contracts analysis, such as contract vulnerability and category classification [50, 250, 312]. In particular, we first define 4 types of features, taking into account the bytecode associated with each smart contract.

The first two feature sets consider the creation bytecode: the first set is obtained by counting the number of occurrences of each specific opcode in the code (call it absolute frequency); whereas the second considers the percentage of the occurrences (call it weighted frequency). The other two sets of features are similar, but consider the deployed bytecode.

We compute the values of these features as follows. We gather the creation bytecode associated with each smart contract from the online service etherscan.io,[3] and the

[3] https://etherscan.io/

deployed bytecode from the online service infura.io.[4] Additionally, we disassemble the bytecode into an equivalent series of opcodes with *evmdis*,[5], a tool to disassemble all instructions in bytecode. Furthermore, we unify opcodes with the same logical function into a single one. For example, the various variants of the PUSH instruction (from PUSH1 to PUSH32) are considered a single generic instruction PUSH. Thus, we associated each contract with two lists of opcodes, one for the creation bytecode and one for the deployed bytecode. For both lists, we compute the absolute and weighted frequency of each opcode, obtaining 4 lists to be associated with the corresponding smart contract.

A summary of the steps we implemented for extracting code features is in Figure 2.3. At the end of the extraction, we obtained 76 different features corresponding to the frequency of different opcodes within the various contracts analyzed.

2.5.2 Account Features

The intrinsic characteristics of a Ponzi scheme determine its behavior. There are at least three characteristics of Ponzi contracts: (1) these contracts usually send Ether to accounts once investing to the contract; (2) some accounts receive more payments than investments. For example, the creator who charges fees frequently from the contract; and (3) the contract balance may be low, as a Ponzi scheme always tries to maintain an image of fast and high returns.

Below, we report the list of features for each contract:

1. *Balance*: the difference between the amount of ETH in input and the ETH in output;
2. *Lifetime*: the difference between the time of the first and the last transaction made or received;
3. *Tx_in*: the number of input transactions;
4. *Tx_out*: the number of output transactions;
5. *Investment_in*: the number of input transactions that deposit an amount of ETH in the contract;
6. *Payment_out*: the number of output transactions paying an amount of ETH;
7. *#addresses_paying_contract*: the number of distinct addresses that paid the contract;
8. *#addresses_paid_by_contract*: the number of distinct addresses paid by the contract;
9. *Mean_v1*: the average of the differences of the number of input/output transactions from/to the same address;
10. *Mean_v2*: the average of the differences of the amount of ETH received and paid by the contract involving the same address;
11. *Sdev_v1*: the standard deviation of the differences in the number of input and output transactions involving the same address;
12. *Sdev_v2*: the standard deviation of the differences between the amount of ETH in and out involving the same address;
13. *Paid_rate*: the ratio between Tx_in and Tx_out;

[4] https://www.infura.io/
[5] https://pyevmasm.readthedocs.io/en/latest/index.html

14. *Paid_one*: the ratio between the number of investors paid many times and the number of total investors;
15. *Investment_in/Tx_in*: the ratio between Investment_in and Tx_in;
16. *Payment_out/Tx_out*: the ratio between Payment_out and Tx_out;
17. *Percentage_some_tx_in*: the percentage of active days with at least one input transaction during the contract lifetime;
18. *Sdev_tx_in*: the standard deviation of the number of transactions per day;
19. **Percentage_some_tx_out**: the percentage of active days with at least one output transaction during the contract lifetime;
20. **Sdev_tx_out**: the standard deviation of the number of transactions in output per day;
21. **Initiator_get_eth_wo_investing**: this feature is 1 if the contract initiator has earned ETH without any investment, 0 otherwise;
22. **Initiator_get_eth_investing**: this feature is 1 if the initiator has earned ETH investing in the contract, 0 otherwise;
23. **Initiator_no_eth**: this feature is 1 if the Initiator has obtained no ETH investing in the contract, 0 otherwise.

The first 18 features are inherited from the literature, while the last 5 in **bold** are the ones we introduce. Below, we briefly comment on some of the features, explaining the rationale behind them. For example, Feature 14 estimates how often a contract interacts with accounts it already knows. A high value of this feature means more interactions (we expect it to happen for a smart Ponzi). Features 15 and 16 aim to capture the requirement R3: the contract redistributes money to the investors according to a given logic by measuring the percentage of transactions distributing ether among Ethereum addresses. Since we expect that non-Ponzi contracts present a lower percentage than Ponzi ones, these features are uniquely based on Ether exchanges. Features 17 and 19 monitor the number of input and output transactions per day over time. A small value of Feature 21 (respectively 19) indicates that the contract was active for a few days, considering input transactions (output, respectively, for Feature 19). On the contrary, a high value means the contract presented a more regular activity over its lifetime. We also consider the standard deviation of the number of daily transactions to capture the variability during the contract life. Feature 15 considers input and output transactions, respectively. This feature measures whether a contract sends or receives transactions only in a few days or regularly. In particular, we expect that Ponzi contracts will have a short lifetime in which they will receive several user investments. Indeed, since making a profit for investors is likely impossible after a certain time, the scheme collapses, and the contract will no longer receive new investments. The last three features, 21, 22, and 23, verify whether the initiator received money from the contract. Indeed, in smart Ponzi contracts, the initiator usually receives a certain amount of money, even without an initial investment. We add these features to study whether receiving a certain amount of money identifies this fraud. Therefore, we expect most contracts labeled as Ponzi to have Feature 21 and Feature 22 equal to 1. On the contrary, Feature 23 equals 1 for non-Ponzi contracts since it is quite unusual to find the initiator of a smart Ponzi that receives no Ether. Our experimental evaluation shows that the new features improve the classification perfor-

mance. These experiments and a thorough description of the features are reported in the conference version of this work [130].

2.5.3 Datasets

As previously stated in stage S_1 of the pipeline, we process 8 datasets, selecting the best one, and we assess the impact of various features on classification quality. Subsequently, a ninth dataset was derived to identify a concise and efficient set of features during the stage S_2.

In particular, we aim to evaluate how the creation and deployed bytecode influence the classification task. Therefore, we consider two groups of datasets, one containing features extracted from the creation bytecode and another containing features related to the deployed bytecode.

The datasets utilizing creation bytecode are as follows:

- *Transaction plus frequency opcode* dataset, which includes the absolute frequency of opcodes and account features (7446 smart contracts x 99 features).
- *Transaction plus weighted opcode* dataset, which includes the weighted frequency of opcodes and account features (7446 smart contracts x 99 features).
- *Only frequency opcode* dataset, containing solely the absolute frequency of creation opcodes (7021 smart contracts x 76 features).
- *Only weighted opcode* dataset, incorporating the weighted frequency of creation opcodes (7021 smart contracts x 76 features) exclusively.

During the cleaning process of the creation bytecode, we identified 415 duplicate samples with the associated labels. Additionally, we encountered 10 instances where distinct labels were linked to the same bytecode at varying addresses. Since we want to devise a fully automated process, and since there is no best policy to solve these conflicts, all these samples were discarded.

Whereas the following datasets use the deployed bytecode:

- *Transaction plus frequency opcode* dataset, which includes the absolute frequency of opcodes and account features (7446 smart contracts x 99 features).
- *Transaction plus weighted opcode* dataset, which includes the weighted frequency of opcodes and account features (7446 smart contracts x 99 features).
- *Only frequency opcode* dataset, comprising uniquely the absolute frequency of deployed opcodes (6621 smart contracts x 76 features).
- *Only weighted opcode* dataset, presenting only the weighted frequency of deployed opcodes (6621 smart contracts x 76 features).

During the cleaning process of the deployed bytecode, we discovered 819 duplicate bytecodes, including their associated labels. Moreover, there were 6 occurrences where distinct labels were mapped to the same bytecode at different addresses. As with the creation bytecode, we discarded all such samples to maintain a fully automated process.

Table 2.2: A summary of the features included in the various datasets.

Dataset	Account Features	Frequency Opcode	Weighted Opcode	non-Ponzi Sample	Ponzi Sample	Total
Creation Bytecode						
Transaction plus frequency opcode	✓	✓	✗	6566	880	7446
Transaction plus weighted opcode	✓	✗	✓	6566	880	7446
Only frequency opcode	✗	✓	✗	6258	763	7021
Only weighted opcode	✗	✗	✓	6258	763	7021
Deployed Bytecode						
Transaction plus frequency opcode	✓	✓	✗	6566	880	7446
Transaction plus weighted opcode	✓	✗	✓	6566	880	7446
Only frequency opcode	✗	✓	✗	5952	669	6621
Only weighted opcode	✗	✗	✓	5952	669	6621
Best features set	✓	✗	✓	6566	880	7446

The rationale behind these decisions is to account for and manage the presence of duplicated bytecodes and their relationship with different addresses and their respective transaction histories.

Table 2.2 summarizes the presence or absence of the different types of features across the datasets. A checkmark (✓) denotes the inclusion of a type of feature, while a cross-mark (✗) indicates its absence. Additionally, a ninth dataset is generated during stage S_2 by minimizing the number of features through the application of RFE. This dataset comprises a subset of account and code features, specifically 47 variables based on the deployed bytecode. Further details about this procedure are in Section 2.6.

2.6 Experimental evalution

In this section, we explore the results obtained by applying X-SPIDE to the 8 datasets that were introduced earlier. These results, together with the inspection of the intermediate stages, answer the following research questions:

RQ1 What configuration (comprising model and dataset) maximizes Recall performance?

RQ2 How do the different types of code features of Section 2.5, derived from creation and deployed bytecode, impact the classification of smart contracts?

RQ3 Does the inclusion of account features improve classification performance compared to using code features alone?

RQ4 Does there exist a subset of sub-features that demonstrates comparable performance to the entire dataset?

RQ5 What are the most important features in detecting Smart Ponzi contracts, and how do they influence the classification outcome?

RQ6 What common peculiarities are found in misclassified contracts, i.e., false positives and false negatives, and how do they influence the incorrect classification?

2.6.1 Stage 1: best model and dataset selection

Here, we describe the stage S_1 of X-SPIDE and answer **RQ1**, **RQ2** and **RQ3**. As said, S_1 takes as input a set of classifiers C, a set of datasets D, a metric m, and assesses the performance of the classifiers in C trained on the datasets in D to identify the pair (classifier, dataset) that outperforms the other combinations according to m. In our experiments, the set of classifiers C includes Decision Tree [250], Random Forest [63], and the Light Gradient Boosting Machine Classifier (LGBMC) [179]. We selected these algorithms because they offer strong classification performance, especially with tabular and unbalanced datasets [218]. Additionally, their behavior is easier to explain compared to more complex methods like neural networks. In summary, they strike a good balance between classification performance and explainability. Whereas for the datasets D, we use the ones introduced in Section 2.5.

To perform the model selection, stage S_1 of X-SPIDE uses a grid search procedure with cross-validation to fine-tune the hyperparameters and thus optimize each classifier's performance. The grid search splits each dataset into an 85% training set and a 15% test set, with stratification based on the target variable. In our experiments, we employ two optimization metrics (i.e., the parameter m of S_1): Area Under the Curve (AUC) and Recall. The cross-validation process of the pipeline splits the training data into ten folds. The model (i.e., classifier and corresponding hyperparameters) undergoes training and testing ten times, using each fold as a validation set. We compute the average score for the metric of interest over these ten tests. Finally, stage S_1 outputs the best pair (classifier, dataset), such that the classifier $c_i \in C$ (with its optimized hyperparameters) and the dataset $d_j \in D$ outperform the other combinations according to the metric m. Specifically, the chosen classifier presents the highest average AUC when optimizing for AUC or the highest average Recall when optimizing for it. For the sake of generality, the pipeline, to perform its optimality selection, also computes the standard metrics *Accuracy*, *AUC*, *F1*, *Precision*, and *Recall* on the test set, considering the optimal hyperparameter values for each dataset and classifier. The results of this step are reported in Table 2.3. The table shows the previously mentioned metrics for all possible combinations of datasets and classifiers where the results optimize AUC and Recall. Typically, when dealing with fraud detection, security issues, and similar cases, Recall is the metric one wants to maximize since capturing as many fraudulent instances as possible is important. Maximizing Recall ensures we effectively identify all fraudulent cases or security breaches within our system. Table 2.3 shows that the LGBMC classifier, optimized for Recall on *Transaction plus weighted opcode* from deployed bytecode, performs the best. The optimal pair is highlighted in **purple**. This pair represents the output of the stage S_1. In particular, the hyperparameters obtained with the grid search for LGBMC are the following: 140 estimators, maximum depth of 15, learning rate of 0.1, subsampling of 0.5 for columns, regularization alpha of 0.2 (L1 regularization term), and regularization lambda of 1 (L2 regularization term).

In summary, our answer to **RQ1** is:

Answers to RQ1:

- When considering the metrics as a whole, the LGBMC classifier achieves the best performance on the *Transaction plus weighted opcode dataset* from deployed

bytecode with recall optimization (Table 2.3).

We further investigate the results of Table 2.3 to extract useful insights on the classification performance, related to the characteristics of the datasets, and to answer questions **RQ2** and **RQ3**. First, the results of the classification task show generally very good performance in all the considered metrics. In more detail, we observe that results obtained using the datasets generated by the deployed bytecode exhibit generally higher Recall than the ones obtained by the creation bytecode. There are two exceptions (2 cases over 12 experiments) when we apply Decision Tree and LGBMC on *Only frequency opcode* from creation bytecode. This consideration leads us to conclude that datasets derived from deployed bytecodes tend to provide higher Recall performance regardless of which classifier is adopted. This is also confirmed by Table 2.4, which summarises the top classifiers' performance across the four datasets. The table is computed by averaging the results obtained across the deployed or creation bytecode datasets for each metric. From the results, it emerges that *deployed bytecode* datasets exhibit better average performance in Recall and Accuracy. This is reasonable, considering the deployed bytecode contains only the code executed at runtime when some contract methods are invoked. Instead, the creation code also includes the code necessary for the creation and deployment of the contract on the blockchain, thus, the malicious behavior may be less evident.

Answers to RQ2:

- Generally, the deployed bytecode provides better classification performances than the creation bytecode alone (Table 2.3) when we consider the Recall as a metric to be optimized.
- All the datasets provide very good classification performances for the various classifiers according to the considered metrics (Table 2.3).

We consider the different performance achieved by the dataset containing or not transaction features to address **RQ3**. We observe two different behaviors depending on whether the dataset includes the creation or deployed bytecode. For the first case, the performance of the classifiers seems not to be impacted by transaction features, regardless of the optimized metric. For instance, considering the recall optimization results (top rows of Table 2.3), the recall achieved using the dataset *Transaction plus frequency opcode* is lower than the one obtained with *Only frequency opcode* for two cases out of three. An opposite behavior emerges considering the other group of datasets that contains code features from deployed bytecode, namely *Transactions plus weighted opcode* and *Only weighted opcode*. In these cases, the former shows higher performance than the latter. The datasets with transaction features always show better results in the case of deployed bytecode, regardless of the type of classifier and optimized metric.

Answers to RQ3:

- Although there are subtle variations in performance across datasets, the dataset incorporating account features exhibits slightly superior performance compared to those utilizing only code features (Table 2.3).

Table 2.3: Results from the grid search comparing three classifiers across datasets, including *Transaction plus frequency opcode*, *Transaction plus weighted opcode*, and *only frequency opcode*, considering both creation and deployed bytecode.

Creation Bytecode	Metric / Classifier	AUC	Accuracy	F1	Precision	Recall
Recall Optimization						
Transaction plus frequency opcode	Decision Tree	0.832	0.930	0.705	0.705	0.705
	LGBMC	0.978	0.962	0.841	0.841	0.841
	Random Forest	0.970	0.961	0.809	0.949	0.705
Transaction plus weighted opcode	Decision Tree	0.849	0.942	0.747	0.768	0.727
	LGBM	0.975	0.961	0.835	0.828	0.841
	Random Forest	0.980	0.962	0.812	0.959	0.705
Only frequency opcode	Decision Tree	0.859	0.933	0.713	0.667	0.765
	LGBMC	0.973	0.957	0.812	0.782	0.843
	Random Forest	0.971	0.953	0.747	0.892	0.643
Only weighted opcode	Decision Tree	0.818	0.929	0.657	0.636	0.680
	LGBMC	0.950	0.958	0.767	0.863	0.690
	Random Forest	0.955	0.964	0.804	0.881	0.740
Auc Optimization						
Transaction plus frequency opcode	Decision Tree	0.832	0.930	0.705	0.705	0.705
	LGBMC	0.973	0.962	0.824	0.925	0.742
	Random Forest	0.974	0.962	0.817	0.959	0.712
Transaction plus weighted opcode	Decision Tree	0.849	0.942	0.747	0.768	0.727
	LGBMC	0.976	0.965	0.840	0.919	0.773
	Random Forest	0.979	0.965	0.837	0.935	0.758
Only frequency opcode	Decision Tree	0.859	0.933	0.713	0.667	0.765
	LGBMC	0.975	0.955	0.805	0.770	0.843
	Random Forest	0.971	0.953	0.747	0.892	0.643
Only weighted opcode	Decision Tree	0.827	0.923	0.667	0.633	0.704
	LGBMC	0.969	0.954	0.776	0.817	0.739
	Random Forest	0.976	0.959	0.792	0.891	0.713
Deployed Bytecode	Metric / Classifier	AUC	Accuracy	F1	Precision	Recall
Recall Optimization						
Transaction plus frequency opcode	Decision Tree	0.841	0.928	0.706	0.686	0.727
	LGBMC	0.973	0.961	0.835	0.828	0.841
	Random Forest	0.964	0.962	0.812	0.959	0.705
Transaction plus weighted opcode	Decision Tree	0.873	0.944	0.769	0.757	0.780
	LGBMC	**0.973**	**0.962**	**0.840**	**0.825**	**0.856**
	Random Forest	0.974	0.964	0.829	0.951	0.735
Only frequency opcode	Decision Tree	0.825	0.933	0.673	0.657	0.690
	LGBMC	0.954	0.953	0.771	0.752	0.790
	Random Forest	0.950	0.959	0.757	0.928	0.640
Only weighted opcode	Decision Tree	0.818	0.929	0.657	0.636	0.680
	LGBMC	0.950	0.958	0.767	0.863	0.695
	Random Forest	0.955	0.964	0.804	0.881	0.740
Auc Optimization						
Transaction plus frequency opcode	Decision Tree	0.841	0.928	0.706	0.686	0.727
	LGBMC	0.975	0.970	0.866	0.902	0.833
	Random Forest	0.964	0.962	0.812	0.959	0.705
Transaction plus weighted opcode	Decision Tree	0.873	0.944	0.769	0.757	0.780
	LGBMC	0.974	0.968	0.859	0.887	0.833
	Random Forest	0.974	0.964	0.829	0.951	0.735
Only frequency opcode	Decision Tree	0.825	0.933	0.673	0.657	0.690
	LGBMC	0.961	0.951	0.782	0.886	0.700
	Random Forest	0.954	0.967	0.811	0.947	0.710
Only weighted opcode	Decision Tree	0.818	0.929	0.657	0.636	0.680
	LGBMC	0.950	0.958	0.767	0.863	0.690
	Random Forest	0.955	0.964	0.804	0.881	0.740

Table 2.4: Comparison of Performance Metrics between Created Bytecode and Deployed Bytecode.

Pipeline	Mean AUC	SD AUC	Mean Accuracy	SD Accuracy	Mean F1	SD F1	Mean Precision	SD Precision	Mean Recall	SD Recall
Deployed Bytecode	0.966	0.013	0.958	0.004	0.807	0.033	0.807	0.035	0.810	0.048
Creation Bytecode	0.971	0.007	0.957	0.005	0.816	0.022	0.852	0.051	0.806	0.036

- When optimizing Recall, LGBMC consistently demonstrates superior performance across all metrics (Table 2.3).
- When optimizing AUC, LGBMC outperforms other classifiers, maintaining high values across all metrics (Table 2.3).

2.6.2 Stage 2: feature selection and most relevant features

The stage S_2 of X-SPIDE receives as input the best pair (model, dataset) (i.e., LGBMC, *Transaction plus weighted opcode dataset* from deployed bytecode with Recall optimization) and determines if there exists a subset of features that improves the quality of the model. This stage returns a reduced subset of features and allows us to answer **RQ4**. To address such a problem, stage S_2 considers the number of features as a further hyperparameter to be optimized. In practice, it performs a grid search procedure with cross-validation to optimize the Recall and the number of features.

To tune this last hyperparameter, we start taking all the features of the considered dataset, perform a grid search and a 5-fold cross-validation with the other hyperparameters of the LGBMC classifier, and optimize for Recall, obtaining the best-performing hyperparameter configuration.

As for the previous experiments, the dataset is divided into an 85% training set (6566 samples) and a 15% test set (880 samples). Then, we adopt the *Recursive Feature Elimination* (RFE) algorithm to remove, at each run, the less important feature. This process continues until we have only 10 features left to evaluate. As a result, we obtain the highest mean Recall with a subset of *Transaction plus weighted opcode* from the deployed bytecode dataset, which consists of 47 features, including eight related to transactions and the remainder to bytecode. The LGBMC classifier trained with the RFE algorithm outperforms previous solutions with only 47 features. The best-performing feature set with LGBMC achieved the following metrics: AUC: 0.977, Accuracy: 0.969, F1: 0.867, Precision: 0.870, Recall: 0.864. This is significant considering the smaller size of the dataset compared to other datasets that included both transactions and bytecode (99 features) or only frequency opcode (76 features), as shown in Table 2.3. Stage S_2 outputs the resulting dataset, which we call *Best feature set.* The hyperparameters for the selected classifier are the following: 140 estimators, a maximum depth of 15, a learning rate of 0.1, a subsample of columns set to 0.8, alpha (L1 regularization term) of 0.2, and lambda (L2 regularization term) set to 1.

Additionally, we examine the confusion matrices of the model using all the features against *Best features set* (see Figure 2.4). The confusion matrix summarizes the number of correctly classified samples, namely true positives (TP); the number of false positives (FP); the number of false negatives (FN); and the number of true negatives (TN). From

Predicted class

	N	P
N	971	14
P	22	110

Full Dataset

Predicted class

	N	P
N	968	17
P	18	114

Best features set

Figure 2.4: Confusion matrices: confusion matrices of the best classifier on *Transaction plus weighted opcode dataset* (Full Dataset) and *Best features set*, where we indicate with N and P the non-Ponzi and Ponzi classes, respectively.

the matrices in the figure, we observe that the classifier trained with *Best features set* presents a higher number of TP and a lower number of FN. More precisely, the figure reveals that 4 false negatives are now correctly classified. In summary, the answer to **RQ4** is:

Answers to RQ4:

- The classification outcome seems to depend more on information from the contract code than the transaction behavior, even if this last one is relevant for the classification (Table 2.3).
- The best set of features consists of 47 features, including account and code features. The top twelve include ten code features and two transaction features (Figure 2.5).
- The LGBMC classifier trained with the RFE algorithm outperforms previous solutions while utilizing only 47 features.
- The classifier trained with the best features increases the number of correctly classified contracts (Figure 2.4).

2.6.3 Stage 3: model explanation with XAI

The stage S_3 of X-SPIDE takes in input the best classifier from S_1 and the dataset obtained with S_2, then it computes the Shapley values Sh, the partial dependences PD, and the misclassified contracts MC. The first two results of S_3 help understand the importance of the features for the classification model. The last result, namely, the misclassified contracts MC, undergoes further investigation to identify the features that misled the classifier.

The Shapley values can be used to visually investigate the impact of the most important features on classification. Partial dependences can be represented in plots called Partial Dependence Plots (PDPs) to analyze the relationship between the target response and specific features of interest while marginalizing the values of all other input features. Below, we describe the analyses on the results of S_3 to explain our model results and their classification performance.

Feature importance

The sets Sh and DP allow us to measure how individual features influence the behavior of Ponzi and non-Ponzi contracts and also help answer **RQ5**.

Figure 2.5 shows the beeswarm plot of the feature importance based on Shapley values for the test set. The plot provides a concise summary of how the top features in the dataset influence the model output. Each instance is represented by a single dot, positioned on the x-axis based on its corresponding Shapley value. The dots 'pile up' along each feature row, indicating the density. The color of the dots represents the original value of the feature, with blue indicating lower values and red showing higher values.

We can determine each feature's positive or negative effect on the prediction outcome by analyzing the plot. The figure shows that a high value of *CALLDATALOAD*,[6] *SUB*,[7] *MULMOD*,[8] *SGT*,[9] *paid_one* positively impacts the classification, indicating a higher likelihood of being labeled as a smart Ponzi contract. Conversely, a low value of *GASLIMIT*,[10] *POP*,[11] *CALLVALUE*,[12] and *tx_in* assists the classifier in identifying the instance as a smart Ponzi contract.

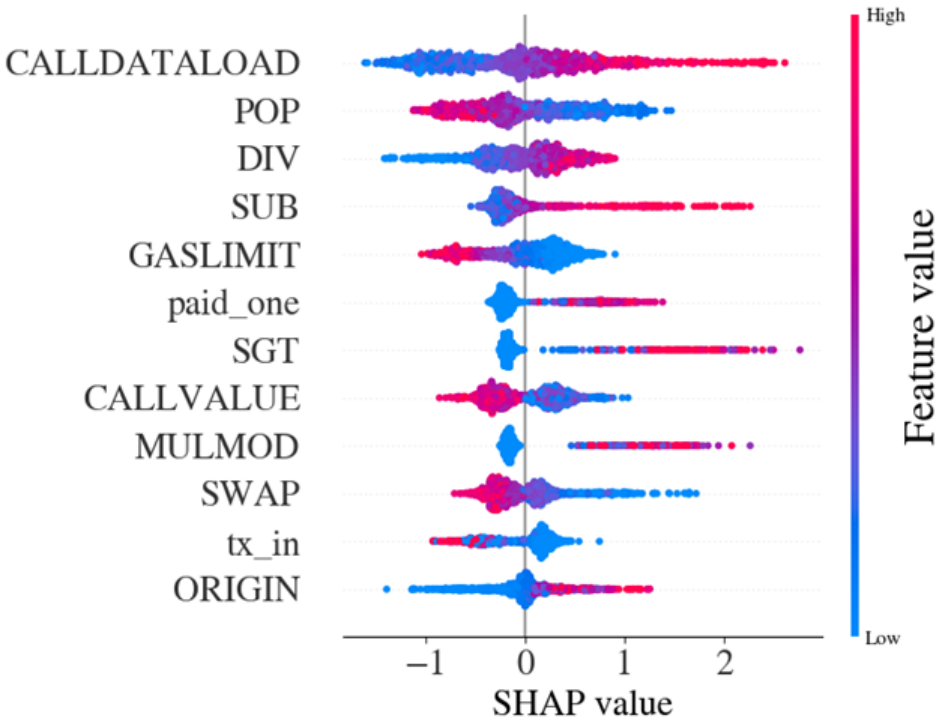


Figure 2.5: Shapley Values of Best feature set: the feature importance of the top 12 features in terms of Shapley values.

[6]The instruction loads 32 bytes of the transaction data on the virtual machine stack.

[7]The instruction subtracts the top two elements of the stack and pushes the result back on the stack.

[8]The instruction performs a modulo multiplication of the top 2 elements with the 3rd element and then pushes the result back on the stack.

[9]The instruction signed greater-than comparison.

[10]The instruction gets the block gas limit.

[11]The instruction removes and returns a word from the stack.

[12]The instruction gets deposited value by the instruction/transaction responsible for this execution.

Furthermore, it is interesting that two of the top twelve features concern the transaction history.

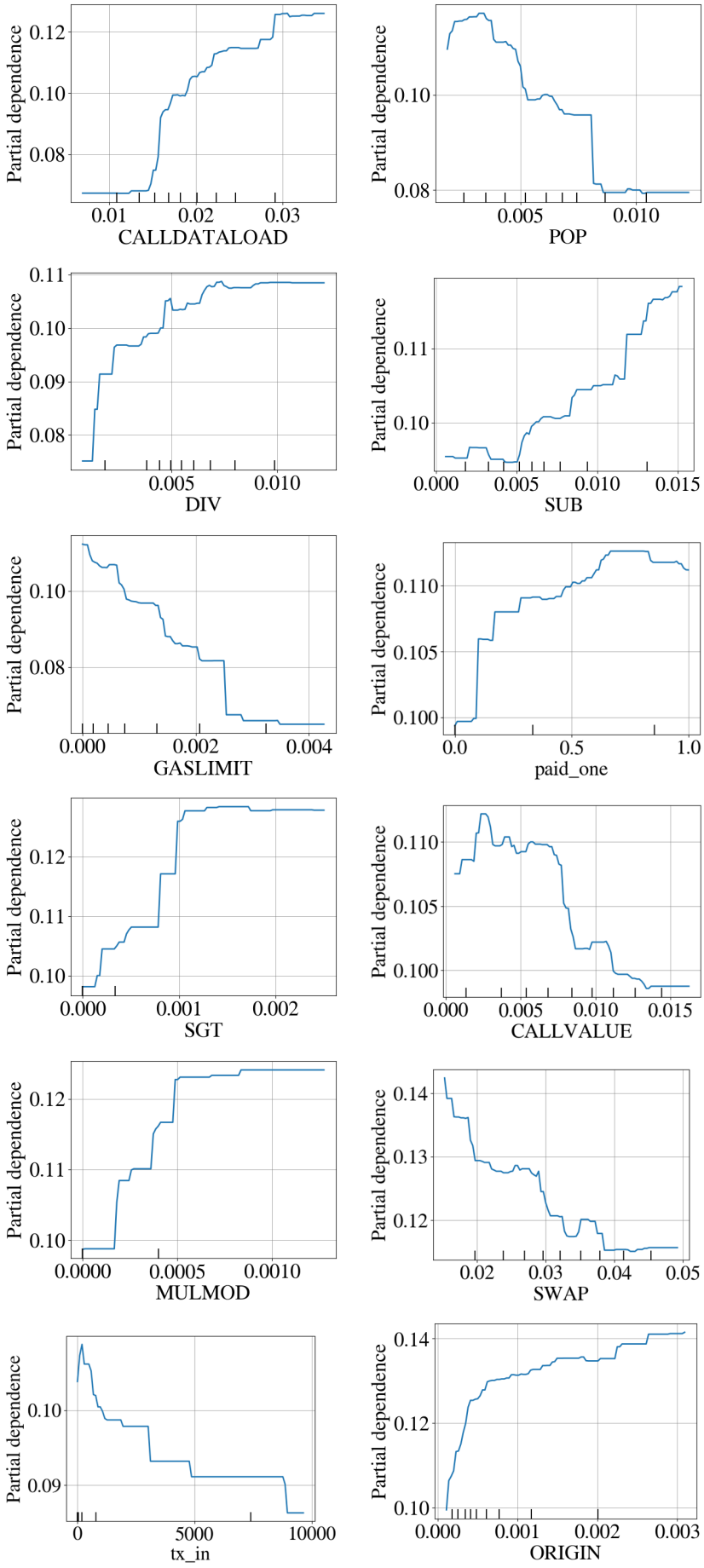


Figure 2.6: Partial Dependence plot of the top 12 features.

Partial dependence plots (PDPs) provide a useful visual and analytical tool to explore the relationship between the target response and a selected set of input features. In our case, we focus on the top twelve features identified through the Shapley values. In Figure 2.6, we present the PDP graphs for each feature within the top 12 features related to the target variable, whether it is classified as Ponzi or non-Ponzi. For instance, the PDP plot for the *CALLDATALOAD* feature in Figure 2.6 illustrates the impact of the proportion of *CALLDATALOAD* opcodes on the likelihood of a smart contract being classified as Ponzi. A higher proportion of *CALLDATALOAD* opcodes is associated with more positive cases in the target variable. This could be explained by considering that a conventional smart Ponzi contract retains a data structure in its storage, which stores users' addresses and investments. These data are included in the transaction input and must be stored in the contract's maintained data structure. They must initially be pushed onto the stack and subsequently loaded. Similar patterns can be observed for *paid_one*. This feature may positively affect classifying an instance as a smart Ponzi because, typically, such contracts do not provide users with many services besides investing some money and, in case, repaying them. Also, the features *DIV*, *MULMOD*, and *SGT* positively impact the classifier, but giving an intuitive explanation of the reason is tricky.

Conversely, features like *GASLIMIT* negatively correlate with the target response. This could be explained by the fact that this instruction appears the same in the code of all smart contracts, both Ponzi and non-Ponzi.

It is important to note that these observations focus on individual features, considering one feature at a time. However, they provide valuable insights into the marginal effects of each feature on the target response.

Answers to RQ5:

- By Figures 2.5 and 2.6 it is clear that opcode features such as *CALLDATALOAD*, *SUB*, and *MULMOD* with high Shapley values may identify some low-level behavior of smart Ponzi contract, e.g., maintaining a data structure to store users' investments or redistributing the reward to investors.
- By Figures 2.5 and 2.6 we see that opcode features such as *GASLIMIT*, *POP* and *CALLVALUE* with low Shapley values may indicate that those instructions are not sufficient to discriminate between the two classes because they appear in the same way in the code of all smart contracts.
- Figures 2.5 and 2.6 show that transaction features such as *paid_one* have high Shapley values, whereas *tx_in* with low Shapley Values may identify some behavioral aspects of smart Ponzi contracts, e.g., these contracts may receive few input transactions with a certain amount of money and tend to pay just one investor.

Analysis of misclassified contracts

The stage S_3 of X-SPIDE, also leverages XAI techniques to automatically analyze misclassified contracts and to identify which contract characteristics contribute to misclassification. Initially, S_3 applies SHAP to the contracts in the subsets *False Positive* and *False Negative*. Subsequently, given the small number of misclassified contracts, S_3 se-

lects them and returns their code, enabling their manual inspection. Finally, S_3 also supports the analysis of specific contracts. Later in this section, we specifically apply it to two pairs of contracts, one for False Positives and one for False Negatives, highlighting their peculiar characteristics through SHAP investigation. Finally, we answer to research question **RQ6**.

As said, we apply SHAP to the subset of False Positive contracts. The results are in Figure 2.7, where we report the most important features. Comparing this plot with the one in Figure 2.5, we note that the feature *tx_in* is replaced by the *SSTORE* one, while the other most important features remain the same. Considering this small set of False Positive contracts, we observe that the *MULMOD* and *SGT* features become more important for Ponzi classification. However, the values of the features push the classifier to label these contracts as Ponzi, making it difficult to distinguish them from the real Ponzi.

To better understand the nature of these contracts, S_3 supports manual inspection. More precisely, it retrieves the addresses of the misclassified contracts[13] and permits an analyst to inspect their verified source code, published on etherscan.io, when available.[14] From this manual inspection, we notice that most False Positive contracts are related to gambling games, lotteries, and tokens. The classifier may have misclassified them because they share characteristics with Ponzi schemes, such as receiving payments from users, storing user addresses in internal data structures, paying fees to the contract creator, and implementing a mechanism to pay out a winner or redeem tokens. However, other similar contracts (tokens, games, and lotteries) are correctly classified in our dataset.

Moreover, stage S_3 allows an analyst to select specific contracts and generates SHAP's bar plots. As an example, we analyze here two false positive contracts.[15] In this scenario, we expect to observe features commonly associated with Ponzi contracts. In the top plot of Figure 2.9, the features *CALLDATALOAD* and *POP* exhibit values similar to those found in Ponzi contracts, and both contribute heavily to the classifier's decision to assign a target value of 1. Conversely, the features *paid_one* and *SGT* have low values, typically associated with non-Ponzi contracts. A similar pattern is observed in the second false positive contract in Figure 2.9. Notably, features such as *SGT*, *MULMOD*, and *CALLDATALOAD* (highlighted in red) strongly influence the classifier to result in a Ponzi classification. On the other hand, the features *POP*, *Investment_in/tx_in*, and *paid_one* push the classifier towards a non-Ponzi classification.

For the False Positives, our analysis indicates that these contracts exhibit several features typically associated with smart Ponzi contracts, making it difficult for the classifier to correctly label them.

The same process is applied to False Negative contracts. The results of the SHAP analysis are in Figure 2.8: also, for this set, the most important features match those of Figure 2.5 except for *tx_in*, which is replaced by the feature associated with the instruction *STOP*. For the set of False Negative contracts, we notice that the feature ORIGIN is the most relevant, while the feature *DIV* lost its importance. Regarding the overall

[13] Available in our online repository.

[14] We cannot provide a conclusive analysis for contracts whose source code is not public.

[15] Addresses `0x927a4e90c3728f04cc373cd4c445daafa9e54df7`, `0xd51a87caaa567677abac451b00e0a0a18a992b49`.

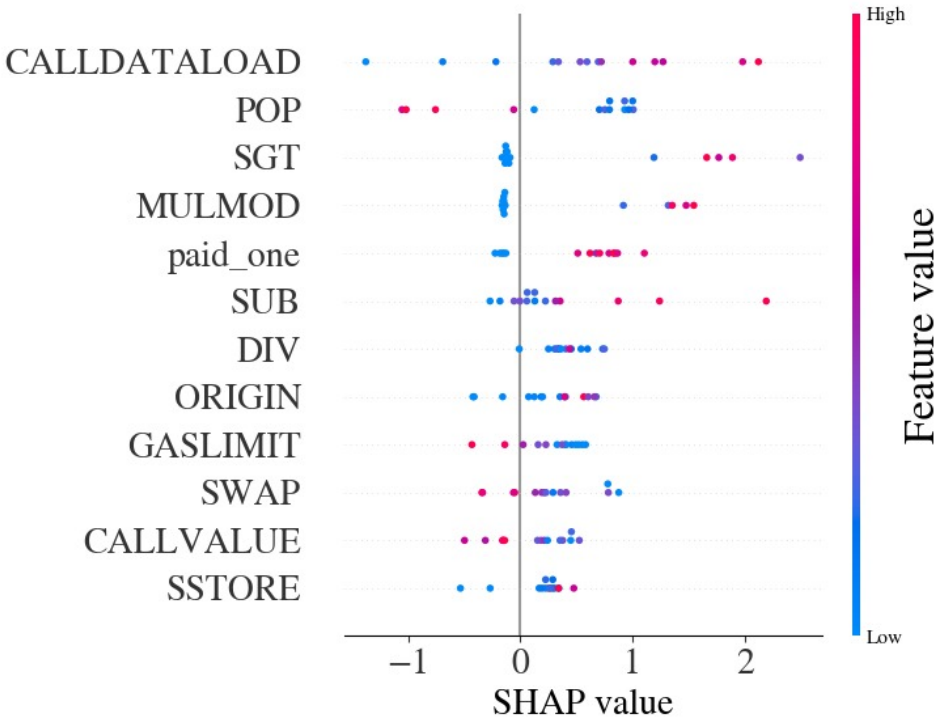


Figure 2.7: Shapley Values of False Positives: The feature importance of the top 12 features in terms of Shapley values.

characteristics, these contracts differ from the typical Ponzi by a low value of *CALLDATALOAD*, *SGT*, and *MULMOD*, pushing the classification towards Not-Ponzi.

Then, stage S_3 retrieves the code of the False Negative to enable the manual inspection of the contract code. From a manual inspection, we found that their code is complex with intricate program logic. However, for some of the contracts where we are able to understand their behavior: reading the code makes clear that they either hide a pyramid scheme within a seemingly legitimate contract type (such as a gambling game or token) or use a complex or overly simplistic logic for distributing rewards. In some cases, the author deliberately conceals the contract's true nature to avoid detection by users.

Finally, we select two false negative contracts misclassified by the model.[16] As shown in Figure 2.10, the SHAP values for these contracts are relatively low, indicating that the classifier tends to label them as non-Ponzi contracts due to the absence or limited presence of significant Ponzi-related features. For instance, features like *SUB* and *CALLDATALOAD* have values similar to non-Ponzi contracts, which might explain why the classifier failed to detect their Ponzi-like characteristics.

Answers to RQ6:

- The analysis reveals that false positives have some behavioral aspects similar to Ponzi schemes. The analysis of the features with XAI techniques (Figure 2.7 and 2.9) confirms that feature values resemble typical characteristics found in Ponzi

[16] Addresses `0xbdb8b73aea0c43118ce8834c91d50ae8bbd5ed32`, `0x1e997b4a256e11071167a1d08e98a9f5dde0bf72`.

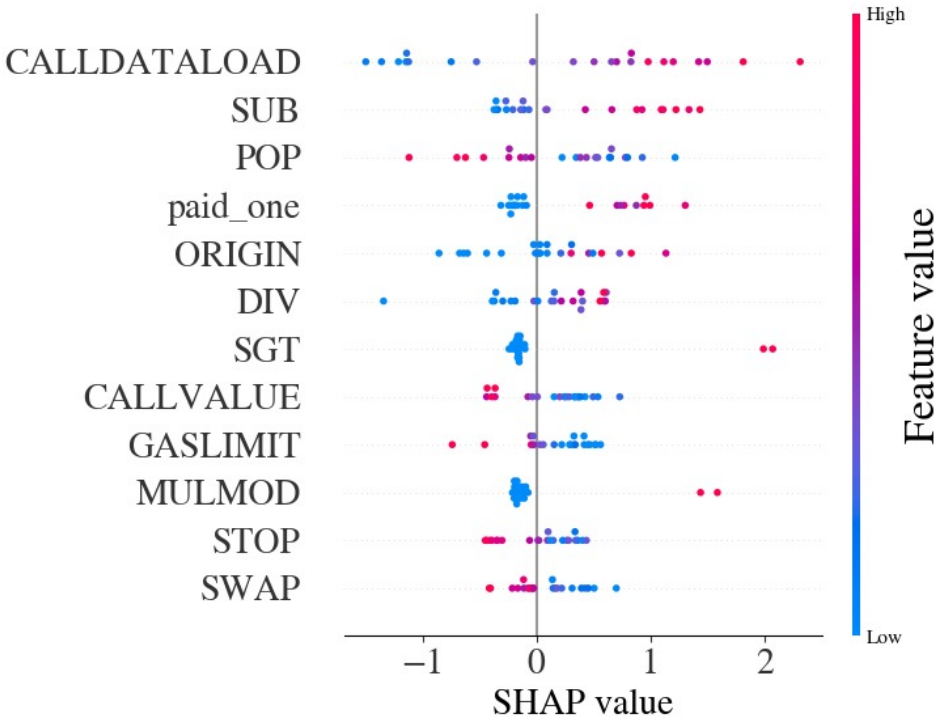


Figure 2.8: Shapley Values of False Negatives: The feature importance of the top 12 features in terms of Shapley values.

contracts, so this explains why the classifier labeled them as Ponzi.

- The manual inspection reveals that false negatives are contracts that hide a pyramid scheme inside a non-fraudulent contract, or implement a singular distribution logic. The analysis of the features with XAI techniques (Figure 2.10) confirms that such contracts often lack features typically associated with Ponzi schemes, so this explains why the classifier labeled them as non-Ponzi.

2.7 Conclusion and future work

This contribution presented X-SPIDE, an automatic explainable machine learning pipeline for detecting smart Ponzi contracts on Ethereum. X-SPIDE consists of three stages. The first stage S_1 assesses the performance of multiple classifiers trained using different datasets to identify the best pair (model, dataset), and allowed us to study how the different bytecode types impact the classification of Ponzi and non-Ponzi contracts. The results of our experiments show that deployed bytecode achieves better performance. Given the best pair (model, dataset), the second stage S_2 identifies a small yet effective set of features that ensures a good classification quality and enables us to explain the classifier's decisions. The last stage S_3 performs two different analyses relying on XAI techniques. The first investigates how the identified features impact classification and how they are related to possible fraudulent behavior. The second consists of an inspection of misclassified contracts. These analyses not only enhance the interpretability of our findings but also reinforce the robustness of our pipeline. Moreover, to train and

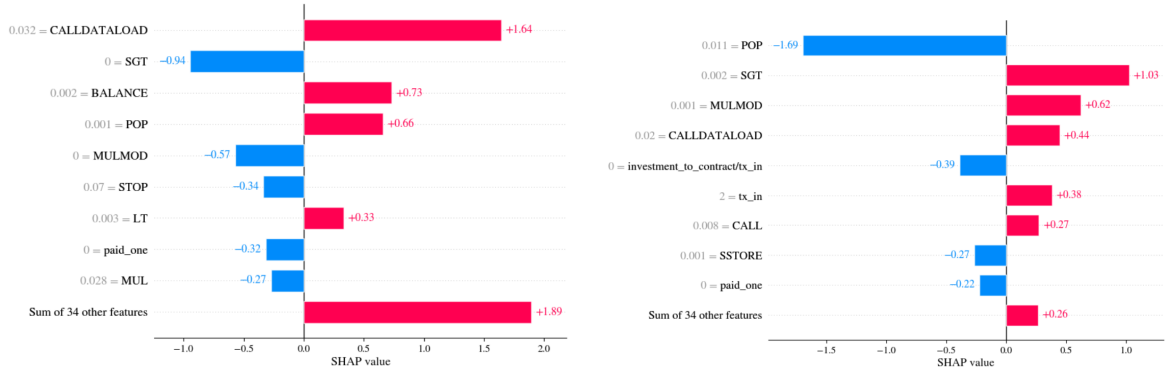


Figure 2.9: Bar Plot of False Positive contracts: Plots showing the impact of features in contract classification.

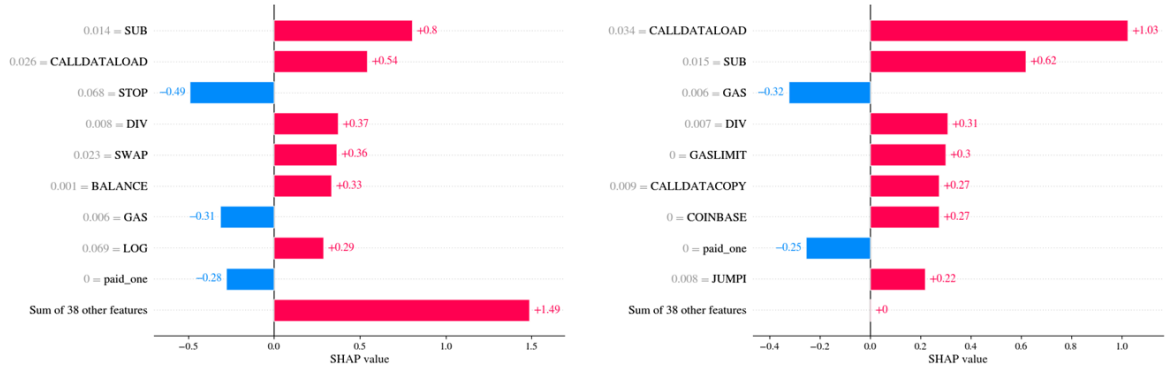


Figure 2.10: Bar Plot of False Negative contracts: Plots showing the impact of features in contract classification.

test our pipeline, we released new reusable datasets with 7446 unique real-world smart contracts that can be used for future research. The dataset contains transaction history and bytecode information and captures several behavioural aspects of contracts during their lifetime.

Future work aims to extend our pipeline in different directions. First, we intend to optimize the best feature extraction procedure. Then, we plan to improve our approach to extract opcode features by the bytecode. In particular, we aim to identify high-level behavioural patterns from sequences of instructions and possibly apply deep learning techniques to minimize the feature engineering effort. As a further direction, we will study how to make our pipeline robust against smart Ponzi contracts that use adversarial evasion techniques to conceal their fraudulent nature, e.g., hiding inside apparently harmless contracts. We plan to apply techniques similar to malware evasion countermeasures [19]. Moreover, we will study how the performance of our pipeline degrades over time as the Ponzi scheme evolves, and we will devise a retraining strategy that allows us to update our model, taking into account new training samples. Additionally, we plan to refine the requirements of being a smart Ponzi proposed by Bartoletti et al. [50]. Another line of research will focus on identifying active Smart Ponzi contracts nearing collapse and refining our features to capture the financial posture of contracts. We will explore the generalizability of this approach to other contract types and other forms of

scams on Ethereum, such as phishing, as one of the most promising. Finally, we plan to integrate our model inside an existing Ethereum Wallet to warn the user when she is about to invest in a suspicious contract.

Data availability

The datasets and the notebooks used for the experiments presented in this study are available online at https://github.com/LucaPennella/x-spide-smart-ponzi-detection.

Chapter 3

Explainable Machine Learning for Predicting Voting Intentions: A Study of Italian Politics

3.1 Introduction

The comprehension of voting intentions is considered a pivotal objective within the domains of political science and computational social science. While demographic variables such as age, gender, and education have traditionally played a key role in explaining electoral preferences, recent scholarship has increasingly highlighted the importance of value-based variables, such as attitudes toward globalization, religious influence, and drug legalization in shaping voter behaviour [111, 169].

This study investigates the predictive power of both demographic and value-oriented features within the Italian political context, leveraging data from the SWG and Rachael Monitoring surveys conducted between 2017 and 2019 [284]. These surveys administered by SWG, one of the leading survey institutes in Italy [285], with the support of Rachael [251], comprise a total of 4,500 observations, with 1,500 respondents surveyed each year. The dataset is stratified by region, gender, age, and urbanization level, and captures a wide range of socio-demographic and attitudinal variables relevant to electoral behavior.

We adopt a machine learning framework to model voting intentions across the three major political coalitions that emerged in the 2018 Italian general elections: the center-left *(Sx/CSx)*, the center-right *(Dx/CDx)*, and the Movimento 5 Stelle (*M5S)*. Our methodology employs tree-based ensemble classifiers: Random Forest [157], Light Gradient Boosting Machine Classifier (LightGBM) [180], and XGBoost [84], which are optimized via Bayesian hyperparameter tuning and evaluated using cross-validation. We also used Decision Tree as baseline model.

To improve model interpretability, we apply eXplainable Artificial Intelligence (XAI) techniques, including Recursive Feature Elimination (RFE) and SHapley Additive exPlanations (SHAP) [207]. These tools enable us to identify the most relevant features and provide transparent insights into the structure of political preferences in the Italian electorate.

This chapter makes four contributions: (i) it proposes an explainable predictive frame-

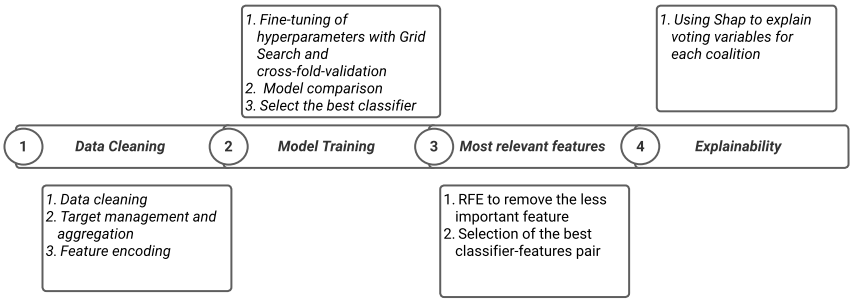


Figure 3.1: Flow diagram of the analysis process

work that jointly models demographic and value-based drivers of voting intentions; (ii) it provides a systematic comparison of tree-based ensemble methods with a transparent baseline, optimized via Bayesian hyperparameter tuning; (iii) it delivers interpretable evidence on coalition-specific voter archetypes through SHAP-based explanations and parsimonious RFE-selected models; and (iv) to foster transparency, reproducibility, and reuse, it publicly releases the full codebase and an anonymized version of the dataset under licenses at https://github.com/LucaPennella/xai-voting-intentions-italy.

The overall analytical workflow is summarized in Figure 3.1, which outlines the sequential phases of data processing, model training, feature selection, and interpretability.

To guide our investigation, we address the following research questions:

RQ1 Which classification model achieves the highest F1 in predicting voting intentions?

RQ2 Is it possible to identify a reduced subset of features that retains predictive performance comparable to the full dataset?

RQ3 Which features are most relevant for distinguishing among voters aligned with different political coalitions *(Dx/CDx, M5S, Sx/CSx)*?

Our results emphasize the added value of incorporating value-based predictors into voting intention models. By revealing distinct patterns of support across political coalitions, the findings contribute to the broader literature on computational electoral behavior and illustrate how XAI methodologies can enhance both the accuracy and interpretability of predictive models.

The remainder of the contribution is organized as follows. Section 2 reviews the related literature and critically examines the main factors influencing voting intentions. Section 3 details the data sources, the variables employed, and the data preparation procedures. Section 4 describes the methodological framework, while Section 5 reports the empirical results. Finally, Sections 6 and 7 present the discussion, draw the main conclusions, and suggest avenues for future research.

3.2 Related Work

In this section, we present an exhaustive introduction to some of the works from different sectors that have applied Explainable Artificial Intelligence.

In recent years, artificial intelligence (AI) has become fundamental in many fields, enhancing data-driven and automated decision-making processes. Its growing use raises important questions regarding transparency, along with issues of explainability, interoperability, accountability, and ethics [215]. These concerns are relevant to researchers and companies, especially in industries where AI-driven decisions involve significant risks that require different solutions and strategies [140]. Increasing attention has been given to making these models readable and reducing the most pressing problem that makes AI and ML methods a black box [58], resulting in an inscrutable process that users are usually requested to rely on and trust [255]. In this context, XAI is directly related to the pressing challenge of providing tools that help technical and non-technical stakeholders understand algorithmic decisions or to better use it as a tool to support their findings. In this sense, some studies have shown the adaptability of ML as a tool for robustness check, integrating traditional models with the parametric approaches [177]. Moreover, as shown in the literature, AI and machine learning algorithms have been used in different fields [164, 165, 177, 240, 243], showing their high reliability and adaptability in terms of applications, and challenges [25, 58, 305].

To enhance transparency, fairness, and trustworthiness [33, 143, 288], XAI techniques have emerged as a solution to both technical and non-technical stakeholders to better understand algorithmic outcomes and procedures [36, 114, 173]. In credit scoring, Chen et al. [88] examined fairness and explainability, while Bussmann et al. [68] proposed a model balancing accuracy and transparency. Other studies explore organisational ethical guidelines [41], and in medicine, XAI has supported patient-flow forecasting [283].

The application of machine learning algorithms is also growing in political and social sciences and, in particular, in the analysis of voting behaviour and electoral prediction. A variety of studies have been conducted in this area: Tumasjan et al. [292] demonstrated the potential of Twitter to predict elections by analysing political sentiment on social media. Burscher et al. [67] applied supervised machine learning to code policy issues, highlighting its adaptability across diverse contexts, Jordan et al. [183] applied Fuzzy forest to the voting data from 2016 Cooperative Congressional Election showing the importance of using ML alghortihm as tools to find the best predictors and find element to support (or not) present thoeries in social sciences.

This chapter is inserted into the topic of voting intentions, where value-based variables have gained prominence in understanding voter behaviour. Inglehart and Norris [168] analysed the rise of populism [53, 72, 327] through cultural values, illustrating how economic and cultural grievances shape voter preferences [57, 124]. Schwartz's theory of basic human values [271] has been central in political psychology in explaining electoral choices[1].

Additionally, socio-demographic variables such as age, gender, and Education have been crucial in predicting voting behaviour [29] investigated how internet use influences

[1]The theory identifies ten basic values, such as security, conformity, achievement, and universalism, that are recognised across cultures and arranged in a circular structure reflecting their motivational compatibilities and conflicts.

political knowledge across various socio-demographic groups, underscoring its impact on electoral participation. Franklin [125] conducted a comprehensive analysis of voter turnout, emphasising demographic factors in electoral dynamics.

Moreover, an increasing interest has been devoted to the analysis of voters' intentions and the different aspects of social factors: Milesi [219] has described the impact of moral values, emotions, and the impact of media and social factors. Waldvogen et. al [303] pointed out the role that emotions can play, affecting the way voters filter and process information, shaped by debates and information such as newspapers that are central in political elections. Particular attention has been given to e-campaigning and political engagement having a central role in electoral performance [80] and the importance of political engagement [133]. Coalition dynamics have been shown to profoundly influence voter behaviour. Laver and Schofield [195] analysed coalition politics in Europe, examining their impact on electoral outcomes. Mershon and Shvetsova [217] studied party system changes, demonstrating how legislative coalitions shape voter preferences. Other influential factors are, e.g., social relationships and structures or their values.

In this various context, where XAI methods have been successfully applied in domains such as finance [68, 88] and healthcare [283], their use in political science remains limited and fragmented. Existing electoral prediction studies for Italy and other European contexts often rely on either traditional statistical approaches or opaque machine learning models, typically prioritising predictive performance over interpretability [67, 292]. Moreover, research on Italian voting behaviour has tended to examine socio-demographic and value-based predictors: socio-demographics dominate turnout and participation studies [29, 125], while value frameworks such as Schwartz's theory [271] are explored primarily in political psychology rather than integrated into predictive models. To our knowledge, no prior work has systematically combined socio-demographic and value-based variables within an interpretable ML framework to model Italian voting intentions. Given these issues, our work do not only want to focus hypothesis testing, but wants to open the floor to a broader discussion on the use of ML algorithm to perform variable selection to develop a descriptive portrait of the voting intension by focusing on a time span going from 2017 and 2019 and by integrating it to the broader discussion of the political behavior theories in Italy. This tool will provide an example of how scholars from different disciplines can use it to support past results or find new insights.

Finally, this study addresses that gap by combining these dimensions in a LightGBM pipeline, enabling both accurate prediction and an explainable mapping of the attitudinal and demographic drivers.

3.3 Data

This section provides a detailed description of the data used in this study, outlining the sources, characteristics, and preparation methods. The subsections below describe the survey data collected for this study, and the data preparation process. Each subsection aims to clarify the nature and scope of the data, as well as the methods used to ensure its relevance and reliability for our analysis.

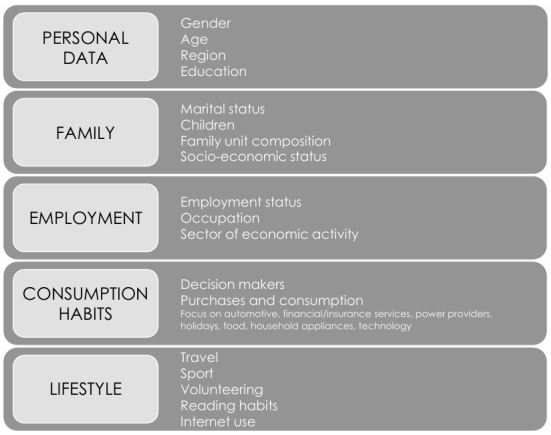


Figure 3.2: An outline of the profiling data.

3.3.1 Survey Data

The data analyzed in this study come from the annual SWG and Rachael Monitoring survey. The collected data encompass a wide range of areas, including socio-demographic, behavioural, purchasing, media consumption, values, and psycho-attitudinal information, as illustrated in Figure 3.2.

The survey targets adults residing in Italy and is stratified by region and degree of urbanization, with quotas for age and gender. Data were collected annually from 2017 to 2019 through online interviews administered via the CAWI method. Participants were drawn from a proprietary online panel under predefined selection criteria. No survey weights were applied; all analyses treat records as individual level observations. For model development, we adopted a temporally separated evaluation design: the training set comprises the 2017 and 2018 waves (n = 2,164 after cleaning and preprocessing), and the test set comprises the 2019 wave (n = 1,155 after cleaning and preprocessing). The survey investigates 14 value areas and trends, updated according to the historical period:

1. *Values:* Individual freedom and rights; Military interventionism; Merit vs. equality; Foundational values; Centrality of Catholic values; Openness to other religions; Fascism

2. *Individual:* Individual inadequacy; Included and excluded; Disorientation; Nostalgia and lost future; Future

3. *Community and Common Good:* Common good; Community; Civic sense and associationism; Solidarity and altruism; Social capital; Volunteering

4. *Immigration:* Attitude towards immigration; Perception of the phenomenon

5. *Environment and Sustainability:* Sensitivity to environmental issues; Individual actions towards sustainability

6. *Safety:* Safety regarding crime; Public order

7. *Gender Issues:* Topics concerning gender gap and gender-based violence

8. *Youth:* Trust in youth; Generational comparison

9. *Education, Science, and Technology:* School, training, and culture; Scientific approach; Opinions on scientific topics; Technological innovation
10. *Family:* Concerns and anxieties for children; Fear of having children
11. *Work:* Employment situation; Social class and future; Public and private employees; Risk propensity; Entrepreneurial model; Self-employment vs. employment; Territoriality; Job mobility
12. *Economy:* Economic security; Consumption and economic prospects
13. *Italian State:* Patriotism and national unity; Federalist push; Italian pride
14. *Globalism:* Europeanism; Globalization
15. *Southern Italy:* North vs. South and regionalism
16. *Politics:* Representativeness and trust in the party system; Trade unions; Politicization and citizen influence; Radicalization
17. *Populism:* Definitions and actions
18. *Taxes:* Opinions and actions
19. *Spirituality:* Religious beliefs, meditation
20. *Epidemics and Vaccines:* Concerns about new epidemics, vaccines, and the healthcare system

3.3.2 Data Description and Preparation

The initial dataset comprises 4,504 users and 160 variables. Below, we present the tables for gender, age, education level, and geographic distribution of the sample and the Italian population based on the 2018 census [170].

Gender	Sample (%)	Population (%)
Male	50	49.7
Female	50	51.3

Table 3.1: Gender distribution of the sample and the Italian population

The gender distribution shows a close alignment between the sample and the population, with the sample consisting of an equal percentage of males and females, closely mirroring the population percentages.

In terms of age distribution, the sample data is broadly representative of the population, with some variances across different age groups. The sample slightly underrepresents the youngest age group (15-24) and the oldest age group (65-90), while the middle age groups show closer alignment with the population percentages.

The education level distribution indicates that the sample has a higher proportion of individuals with middle and high education levels compared to the general population.

Age Group	Sample (%)	Population (%)
15-24	7.8	11.4
25-34	15.1	12.7
35-44	18.9	15.7
45-54	20.0	18.9
55-64	16.5	16.0
65-90	21.2	22.6

Table 3.2: Age distribution of the sample and the Italian population

Education Level	Sample (%)	Population (%)
Low (no title, primary, lower secondary)	18.0	50.9
Middle (upper secondary, post-secondary non-tertiary)	47.4	35.1
High (tertiary)	35.0	14.1

Table 3.3: Education level distribution of the sample and the Italian population

Region	Sample (%)	Population (%)
North-West	27.0	26.8
North-East	20.0	19.4
Center	19.3	19.8
South	22.5	23.1
Islands	11.1	10.9

Table 3.4: Geographic distribution of the sample and the Italian population

The geographic distribution of the sample closely matches that of the Italian population, with minimal differences across the regions. The proportions for each area in the sample are nearly identical to those of the population.

Overall, the sample data aligns well with the demographic characteristics of the Italian population for gender and geographic distribution, with some variations observed in age and education level distributions.

The data preparation classification pipeline involves selecting only demographic and value-related variables.

Since the data originate from surveys, encoding is necessary for all predictive variables of interest.

To reflect the structure of the 2018 Italian general election, we aggregate parties into three pre-electoral coalitions: *Sx/CSx* (centre-left), *Dx/CDx* (centre-right), and *M5S*. This aggregation enables a more precise analysis of political tendencies and voter behaviour in our setting. By focusing on these three coalitions, we aim to capture the political dynamics and ideological cleavages salient in Italy during the 2018 contest, and to assess how demographic and value-related factors shape preferences and alignments.

After recoding, we obtain three labelled groups: *Sx/CSx* with n=1,348 respondents, *Dx/CDx* with n=1,094, and *M5S* with n=869. Respondents indicating parties outside these coalitions (n=74) and those not reporting a vote intention (n=150) are excluded from the supervised classification analyses.

Because one of our objectives is to characterise voter archetypes, we do not include

respondents who reported being undecided in the predictive models. Given the substantive importance of indecision, we provide an exploratory analysis of undecided voters in Appendix C.

After feature construction and listwise deletion, the analytical dataset used for supervised modelling contains 3,319 observations and 160 variables.

The encoding process involves converting categorical data into numerical formats, which is essential for applying statistical and machine learning algorithms. This step ensures the data are in a suitable format for subsequent analysis, enabling the extraction of meaningful patterns and insights. The demographic variables include age, gender, education level, region, and urbanization level, while the value-related variables cover attitudes towards individual freedom, immigration, environmental sustainability, and economic security, among others. After applying encoding, the number of columns in our dataset is 188.

By carefully selecting and encoding the dataset, we aim to provide a robust foundation for our analysis, allowing us to explore the relationships between demographic factors, personal values, and political affiliations. This approach enhances the validity of our findings and ensures that our analysis accurately represents the socio-political landscape of Italy during the study period.

3.4 Methods

3.4.1 Research Questions

The research question for this analysis focused on various aspects. The first two questions examined the model, specifically considering F1 and the impact of features on the model's performance. The third question aimed to identify the most crucial variables for each category of voter.

RQ1 Which classification model achieves the highest F1 in predicting voting intentions?

RQ2 Is it possible to identify a reduced subset of features that retains predictive performance comparable to the full dataset?

RQ3 Which features are most relevant for distinguishing among voters aligned with different political coalitions *(Dx/CDx, M5S, Sx/CSx)*?

3.4.2 RQ1 – Model Selection and F1 Optimisation

We compared four tree-based classifiers, a single decision tree taken as baseline, a random-forest ensemble, LightGBM and XGBoost, because these algorithms routinely achieve strong accuracy on tabular data while retaining a level of interpretability that deep neural networks rarely provide [218]. For each learner we carried out an independent Bayesian hyper-parameter search with *Optuna* [21]; the optimiser was run for 200 trials per model and the macro-F_1 score was used as the sole optimisation target. Macro-F_1 was preferred over overall accuracy because the three voter blocs (*Dx/CDx*, *Sx/CSx*, *M5S*) are moderately imbalanced. Accuracy can be inflated when a classifier over-predicts the majority bloc, whereas macro-F_1, the unweighted harmonic mean of per-class precision

and recall, gives equal importance to every class and is therefore recommended for imbalanced multi-class problems where all categories are substantively relevant [138, 266].

Each trial was evaluated by ten-fold stratified cross-validation on the 2017–2018 portion of the survey (shuffle enabled, `random_state=42`); the mean macro-F_1 across the ten folds was returned to the sampler. At the end of the search we retained the hyper-parameter vector that maximised macro-F_1 together with the complete vector of fold-wise metrics (accuracy, precision, recall and macro-F_1). These artefacts served as input for the subsequent significance analysis. Following Dietterich [109] and Demšar [107], we compared models on the cross-validation folds with both a paired t-test (parametric) and the Wilcoxon signed-rank test (non-parametric). To verify that our conclusions generalise to unseen data we retrained each tuned classifier on all 2017–2018 records and evaluated it on the 2019 hold-out sample; on this independent test set we estimated the sampling distribution of the macro-F_1 difference via a paired bootstrap with $10{,}000$ resamples, following Efron and Tibshirani [289] and Bouckaert and Frank [59]. A model was declared superior only if (i) its ten-fold mean macro-F_1 was significantly higher ($p < 0.05$) in at least one of the two paired tests *and* (ii) the 95% bootstrap confidence interval for the test-set difference excluded zero; otherwise the models were deemed statistically equivalent and the final choice relied on secondary criteria such as computational efficiency and explanatory power.

3.4.3 RQ2 – Can a Reduced Feature-set Match the Full Model?

Step 1 – Recursive Feature Elimination (RFE). Starting from the complete set of $d = 188$ predictors, we applied *recursive feature elimination* with a LightGBM core estimator that used the fixed hyper-parameters selected in RQ1. At each iteration RFE ranks variables by split gain, removes the single weakest predictor (`step=1`) and refits the model, thereby generating a nested sequence $d, d-1, \ldots, 11$ feature subsets.

Step 2 – Fast local hyper-parameter tuning. To accommodate the change in dimensionality we performed a lightweight Bayesian search (Optuna, 10 trials) only when the current subset size was within ± 2 variables of the best macro-F_1 observed so far. Each trial was evaluated by ten-fold stratified cross-validation (*StratifiedKFold*, *random state*=42) and optimised against the same objective used in RQ1, namely macro-F_1.

Step 3 – Hold-out test. For every subset we retrained the best configuration (fixed or light-tuned) on the 2017–2018 data and measured its performance on the 2019 hold-out set. The subset that maximised test macro-F_1 was taken as the answer to RQ2.

3.4.4 RQ3 – Which features Drive Each Voting Party?

The third research question is explanatory: we aim to identify those survey variables that most strongly increase or decrease the probability that a respondent will declare an intention to vote for *Dx/CDx*, *M5S* or *Sx/CSx*.

Modelling set-up The analysis builds on the 93-feature LightGBM model derived in RQ2. The classifier is re-trained on the complete 2017–2018 sample and evaluated and

explained on the independent 2019 hold-out set.

Explainability framework. The entire procedure is implemented with the SHAPlibrary[2] by Lundberg Et al. [207]. We use the beeswarm plot of the feature importance based on SHAP values. The plot provides a concise summary of how the top features in the dataset influence the model output. Dot colours, reflecting original feature values (blue for lower, red for higher), reveal each feature's impact on prediction outcomes. Each instance is represented by a single dot, positioned on the x-axis based on its corresponding SHAP value. The dots 'pile up' along each feature row, indicating the density. The color of the dots represents the original value of the feature, with blue indicating lower values and red showing higher values. We can determine each feature's positive or negative effect on the prediction outcome by analyzing the plot. The plots are generated for *Dx/CDx*, *M5S* and *Sx/CSx*.

3.5 Results

3.5.1 RQ1 - Model comparison

This section reports predictive performance results following the evaluation protocol outlined above. We first present cross-validated metrics on the 2017–2018 data, using macro-F_1 as the primary criterion given the moderate class imbalance across the three blocs. Pairwise comparisons across folds are assessed with both a paired t-test and a Wilcoxon signed-rank test. We then evaluate each tuned classifier on the 2019 hold-out sample and quantify between-model differences via paired bootstrap confidence intervals. We report, for each learner, aggregate metrics (accuracy, precision, recall, F_1).

Cross-validation performance (2017–2018) After Bayesian tuning (200 trials per model) every candidate classifier was re-evaluated with the same ten stratified folds used during optimisation. Table 3.5 reports the fold–wise mean and standard deviation of the *macro*-F_1.

Table 3.5: Ten–fold cross-validation macro-F_1 (training years 2017–2018).

Model	Mean F_1	SD
Decision Tree	0.548	0.027
Random Forest	0.617	0.027
XGBoost	0.659	0.028
LightGBM	0.649	0.042

Pairwise significance was assessed on the ten fold scores with both the paired t-test and the Wilcoxon signed-rank test.

XGBoost and LightGBM outperform both Random Forest and the single Decision Tree. The mean difference in macro-F_1 between XGBoost and Random Forest is -0.042

[2] https://shap.readthedocs.io/en/latest/index.html

(RF lower), with $p = 8\times10^{-4}$ for the paired t-test and $p = 0.002$ for the Wilcoxon signed-rank test. LightGBM vs Random Forest yields an analogous gap of -0.032 ($p = 0.008$ and $p = 0.014$ respectively). Both ensembles are also superior to the Decision Tree by 0.07–0.11 F_1 points ($p < 0.01$). In contrast, the difference between LightGBM and XGBoost is only -0.01 and remains non-significant ($p = 0.40$ / $p = 0.70$).

Hold-out performance (survey year 2019) Each tuned model was retrained on the complete 2017–2018 data and applied to the independent 2019 sample. Table 3.6 summarises the main quality metrics; 95 % bootstrap confidence intervals (CI) were computed using $10{,}000$ paired resamples.

Table 3.6: Hold-out performance on the 2019 survey ($1{,}155$ cases). All figures are means over $10{,}000$ paired bootstrap resamples; brackets give the corresponding 95 % confidence interval.

Model	Accuracy	Precision	Recall	Macro-F_1
Decision Tree	0.663 [0.635, 0.690]	0.623 [0.592, 0.653]	0.624 [0.597, 0.650]	0.618 [0.590, 0.646]
Random Forest	0.693 [0.667, 0.719]	0.661 [0.628, 0.693]	0.645 [0.621, 0.670]	0.634 [0.606, 0.663]
XGBoost	0.689 [0.661, 0.715]	0.661 [0.630, 0.691]	0.647 [0.621, 0.673]	0.646 [0.617, 0.674]
LightGBM	0.685 [0.658, 0.711]	0.656 [0.625, 0.685]	0.643 [0.617, 0.669]	0.641 [0.613, 0.669]

On the independent 2019 test set the three ensembles converge: the macro-F_1 scores are 0.646 for XGBoost, 0.641 for LightGBM and 0.634 for Random Forest, with partially overlapping 95% bootstrap confidence intervals $\big([0.617,\ 0.674], [0.613,\ 0.669], [0.606,\ 0.663]\big)$.

Significance tests on both the cross-validation folds and the hold-out predictions show no measurable difference between XGBoost and LightGBM, whereas Random Forest and the Decision Tree are clearly inferior. On unseen data the three ensembles are statistically indistinguishable, with XGBoost showing a marginal but non-significant edge.

Because LightGBM matches XGBoost in predictive performance while training and scoring markedly faster, and because it provides native SHAP explanations [208], we adopt LightGBM as the reference model for all subsequent research questions.

3.5.2 RQ2 – Evaluating Whether a Compact Predictor Set Matches Full-Model Performance

Table 3.7 and the confusion matrix in Fig. 3.3 report the hold-out (2019) results obtained by the LightGBM classifier trained on the optimal 93-feature subset.

Error structure. The *Sx/CSx* is recovered most reliably: 86 % of its voters are correctly identified (421/487) with a precision of 0.73. The *Dx/CDx* voters display balanced precision and recall (0.75, 0.72); their misclassifications split almost evenly between the remaining blocs (60 labelled *M5S*, 45 labelled *Sx/CSx*). The *M5S* remains the hard class: its recall is only 0.42, with errors shared between the two mainstream coalitions (60 labelled *Dx/CDx*, 109 labelled *Sx/CSx*). This mirrors the socio-demographic overlap and its smaller weight in the training years.

Table 3.7: Hold-out metrics for the 93-feature LightGBM

Class	Support	Precision	Recall	F_1
Dx/CDx	379	0.75	0.72	0.73
M5S	289	0.56	0.42	0.48
Sx/CSx	487	0.73	0.86	0.79
Macro avg	–	0.681	0.668	0.669
Accuracy		0.706		

Macro vs. micro evaluation. Macro-F_1, which assigns equal weight to every class, drops to 0.669 precisely because of the limited *M5S* recall, whereas the weighted F_1 and overall accuracy exceed 0.69 and 0.70, respectively. The result confirms the appropriateness of macro-F_1 as the optimisation target: it prevents the minority party from being ignored while the majority classes dominate the global score.

Take-away for RQ2. With only 93 predictors, the model matches and slightly surpasses the 188-feature baseline (macro-F_1 $+1.4$ pp, accuracy $+0.3$ pp) while halving dimensionality and inference time. Further improvements should target the *M5S* via engineered features that capture anti-establishment sentiment or cost-sensitive training.

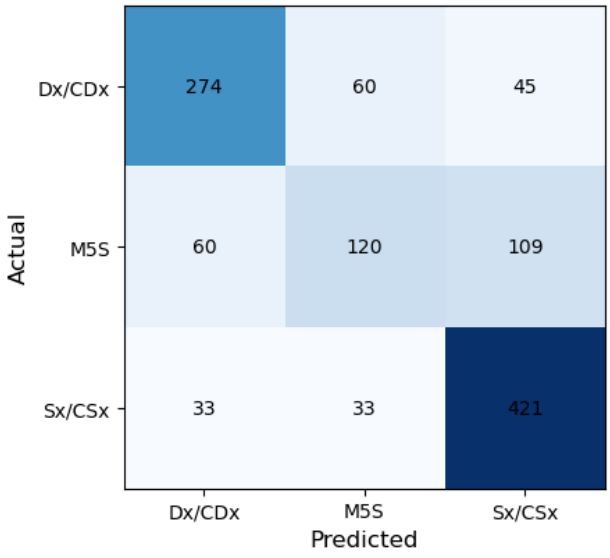


Figure 3.3: Confusion matrix for the 93-feature LightGBM on the 2019 hold-out sample.

3.5.3 RQ3 - Most important features for each type of voter

To answer to **RQ3** we adopt the XAI library Shap[3] by Lundberg et al. [207]. This library allows the visual investigation of the impact of the most important features on the clas-

[3] https://shap.readthedocs.io/en/latest/index.html

sification. The Shapley values can be used to visually investigate the impact of the most important features on classification.

Dx/CDx (centre-right and right-wing voters). Figure 3.4 shows the beeswarm plot of the feature importance based on SHAP values for the test set for *Rx/CDx* voters.

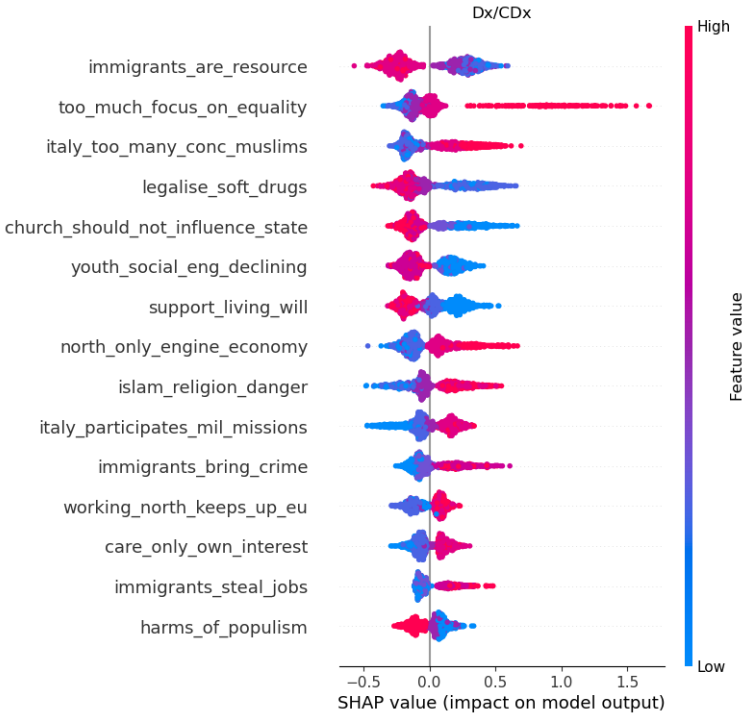


Figure 3.4: Comparison of the top 15 features resulted from the application SHAP values for Right/Center-Right voters.

The conservative electorate is primarily characterised by attitudes towards immigration, social equality and religion. A negative perception of immigration consistently raises the log-odds of a *Dx/CDx* vote: respondents who disagree with *"immigrants are a resource"* and agree with statements such as *"immigrants bring crime"*, *"immigrants steal jobs"* or *"Muslims receive too many concessions"* are pushed rightwards by sizeable positive SHAP values. The same constellation is documented in recent electoral studies that link anti-immigrant sentiment to far-right success across Western Europe [102, 259]. Likewise, judging that society places *too much focus on equality* and regarding Islam as a general danger both increase the probability of a conservative choice. Conversely, liberal stances support for the legalisation of soft drugs or opposition to Church influence on the state-tend to pull voters away from *Dx/CDx*. Regional and socio-economic items such as *"the North is the only engine of Italy's economy"* or *"working in the North keeps pace with the EU"* appear with smaller, more centred clouds, indicating heterogeneity within the bloc.

M5S (Movimento 5 Stelle voters). In Figure 3.5 the features that drive *M5S* predictions depict an electorate that is populist, sceptical of traditional parties and environmentally interventionist. Low concern for the *harms of populism* (blue dots on the right

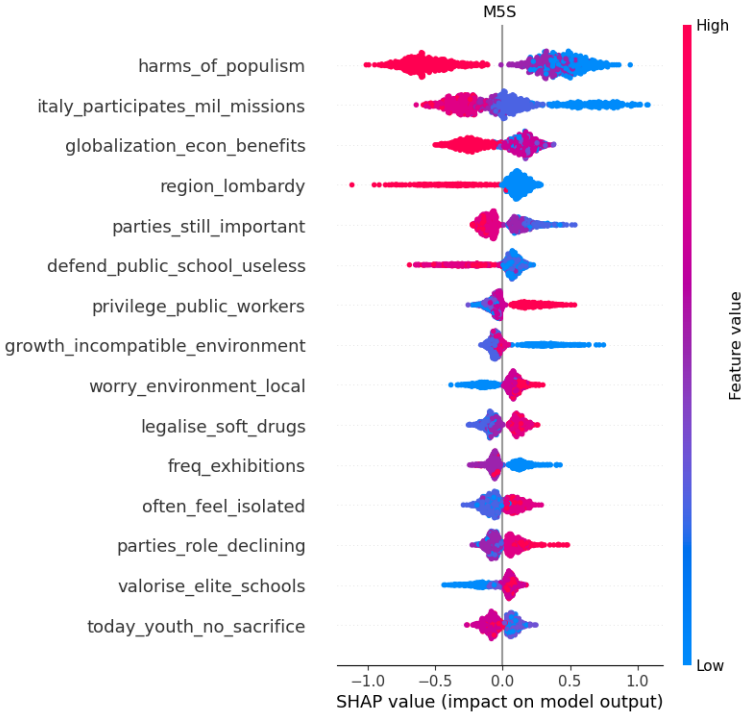


Figure 3.5: Comparison of the top 15 features resulted from the application SHAP values for M5S voters.

side of the axis) markedly increases the log-odds, as does agreement with *"Italy participates too much in foreign military missions"* and *"economic growth is incompatible with environmental protection"*, both resonating with *M5S* rhetoric [53, 300]. Distrust in established institutions emerges through the negative effect of believing that *parties are still important* and the positive contribution of judging a *defence of public school as useless* [54]. Residence in *Lombardy* lowers the probability of an *M5S* vote, in line with the region's stronger alignment with mainstream coalitions. Feelings of social isolation, frequent exhibition attendance and scepticism toward elite schools provide secondary but noticeable influences.

Sx/CSx (centre-left and left-wing voters). The Figure 3.6 Progressive voters are distinguished by pro-EU, pro-immigrant and anti-populist orientations. Agreement with *"the harms of populism"* and the notion that *"Italy has been modernised by the EU"* yields the largest positive SHAP values, strongly tilting predictions toward *Sx/CSx* [136]. Additional positive contributions arise from supporting EU membership and viewing immigrants as a resource, whereas considering *resistance values outdated* produces a negative shift. A pronounced *Italian rather than European identity* decreases the log-odds, whereas endorsing the *common-good perspective for Italy* increases them. Sociodemographic attributes such as higher *age* and employment in the *public sector* also exert moderate positive effects, reflecting the traditional base of the centre-left. Finally, perceiving personal insecurity or prioritising merit over equality nudges voters away from the progressive bloc, mirroring the ideological divide observed for *Dx/CDx*.

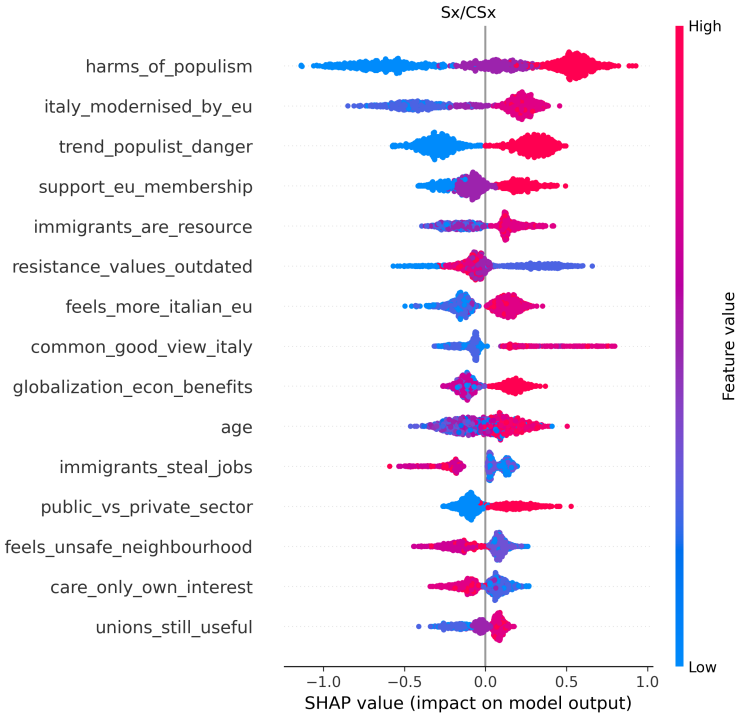


Figure 3.6: Comparison of the top 15 features resulted from the application SHAP values for Left/Center-Left voters.

Cross-parties comparison and takeaway. A juxtaposition of the three beeswarm plots confirms that the model captures three orthogonal attitudinal axes. First, the immigration–multiculturalism dimension sharply separates the conservative bloc (negative immigration valence) from the progressive bloc (positive valence), while generating ambiguous signals for the anti-establishment voters. Second, attitudes toward European integration exert opposite influences on the left and on *M5S*, whereas they are largely neutral for *Dx/CDx*; this pattern suggests that pro-EU items have become a hallmark of the centre-left after 2017, whereas Euro-scepticism is increasingly monopolised by the *M5S* electorate. Third, populism-related statements (*harms of populism*, *populist danger*) display mirror-image effects: rejecting the existence of a populist threat is the single strongest driver of an *M5S* classification, whereas acknowledging that threat is the most salient cue for a progressive vote; the centre-right remains almost orthogonal to this axis.

Taken together, the SHAP evidence indicates that the reduced-feature model encodes substantive political cleavages rather than spurious demographic correlates. Immigration, Europe and populism dominate the ideological space, while socio-economic variables play a secondary role once attitudes are accounted for.

Interpreting the global importance profile. To balance interpretability and completeness, we complement the class-specific beeswarms with a coalition-wise reading of the top-30 features ranked by global importance (absoulte mean SHAP), reported in Appendix B. Across blocs, the cumulative importance profile displays a clear elbow around ranks 10–15, after which contributions taper off; thus, items in positions 11–30 remain

informative but provide incremental lift relative to the leading set. For *Dx/CDx*, the dominant signals cluster around immigration, social-equality attitudes, and religiosity, while regional and socio-economic markers populate much of the long tail with smaller, more centered SHAP clouds. For *M5S*, anti-establishment orientations and foreign-policy and environmental trade-offs concentrate the top mass, whereas institutional distrust and some geographic markers mostly appear between ranks 11–30 with compressed magnitudes. For *Sx/CSx*, pro-EU, pro-immigration and anti-populist stances occupy the top positions, while identity- and security-related items sit largely in the tail and refine, rather than redefine, the decision boundary. This coalition-wise pattern justifies showing top-10 beeswarms in the main text while documenting 11–30 globally in the appendix, thereby bounding cognitive load without suppressing relevant information.

3.6 Discussion

In this section, we will discuss the results obtained by using the related literature related to the different factions: 1)*Dx/CDx*: Lega, Fratelli D'Italia (*FDI*), Forza Italia (*FI*), 2) Sx/CSx: Partito Democratico (*PD*), and 3) Movimento 5 Stelle (*M5S*).

Lega bases its electoral strength on the migrant crisis, positioning itself alongside other right-wing populist movements in Europe, emphasising national sovereignty. *M5S* has centred its consensus on the crisis of economic and political representation. Despite the two shared governments in 2018, two distinct paths emerged: *M5S* focused on anti-elitarian governance, Lega prioritised the migration crisis [72]. *FDI* has also built its identity around the migration crisis, which serves as a key predictor for both Lega and *FDI* voting behaviour [249]. *MS5* supporters have shifted from a left-leaning stance (opposed to spending cuts) to positions to the right (opposed to spending increases), and show interest in full employment, opposing EU-wide spending if financed through additional taxes [124].

Leadership dynamics have shaped party identities. A key factor for *FDI* is the figure of Giorgia Meloni, who has emerged as a fundamental element in shifting voting support from Lega to *FDI*, reflecting a leader-centred approach, reinforced by ideological alignment and negative perception of other rivals [249]. In the broader context, socioeconomic inequalities also play a significant role, along with political and cultural factors: the right and centre-right have strong electoral bases in regions with middle-class impoverishment, with declining wages, and a reduced gap between the rich and middle class [57]. Although it retains some working-class roots, with high employment but low wages, the centre-right has strengthened its ties to economic privilege and metropolitan regions in Central Italy.

Lega gains support in areas where middle-class income declines and approaches the poor, where wealth disparity is smaller, average wealth is lower, and work is more precarious [56]. It remains rooted in some Northern regions characterised by high incomes and inequality, but also attracts disappointed voters from the centre-left and other centre-right parties [57].

FI and *FdI* maintain their traditional strength in the South, characterised by low income and education levels. Meanwhile, the League has grown by attracting protest votes, especially in regions with more part-time work. The centre-left, historically stronger in wealthier regions with high employment and low inequality in disposable income,

shifted toward areas with greater wealth and inequality after 2008, losing support among precarious and part-time workers and graduates—the groups that previously supported them [57]. *MS5* has gained more support in impoverished areas with low incomes and precarious employment, especially in the South [56]. Generally, there has been an effort to politicise EU-related issues to attract voters interested in this topic [79].

Based on the literature, the SHAP-based analysis has deepened these insights, identifying three key attitudes related to 1) migrations with a negative attitude of the right/centre-right voters, with a left/centre-left voters holding a positive view and *M5S* exhibits a mixed profile, 2) Europe: Euroscepticism characterises the *M5S* voters with right/center right holding a neutral position, 3) Populism: *M5S* aknowledge the harms of populism, while the right/center Right electorate remains neutral.

Overall, all these parties present themselves as anti-system parties, with their identity and existence rooted in contrasting themselves from other parties through strong anti-party propaganda to highlight differences, leading or forming poles of opposition, or by introducing new issues into political discourse [328].

A major strength of the approach is its balance between predictive power and interpretability. Recursive feature elimination reduces survey length and computation time while retaining substantive cleavages, and SHAP offers class-specific explanations that domain experts can readily translate into political insight. The pipeline is fully automated yet remains transparent, thus satisfying the growing methodological demand for explainable machine learning in the social sciences.

About the limitation, first, the analysis relies on self-reported survey answers, which remain exposed to social desirability bias despite weighting procedures. Second, the model captures a static 2017–2019 snapshot; external shocks may alter the salience of individual predictors. Third, although the macro-balanced metric mitigates class imbalance, performance on *M5S* voters (recall 0.42) indicates that more latent or behavioural features may be needed to model protest constituencies. Finally, while SHAP enhances transparency, interpretation still hinges on expert judgment, stressing the importance of interdisciplinary collaboration between data scientists and political analysts.

The reduced-feature LightGBM encodes genuine ideological dimensions rather than spurious demographics, demonstrating that value-based variables are indispensable for fine-grained electoral prediction. By combining compact survey design with explainable machine learning, the study contributes a replicable blueprint for computational political science and related fields.

3.7 Conclusion and future work

This study demonstrates that integrating value-based variables with traditional socio-demographics markedly improves the modelling of Italian voting intentions. Using a modular pipeline that combines recursive feature elimination (RFE), Bayesian hyper-parameter optimisation, and post-hoc explainability, we identify a LightGBM classifier trained on 93 predictors that achieves a macro-$F_1 = 0.669$ and an accuracy of 0.706 on an unseen 2019 hold-out set, outperforming the full 188-feature baseline. Dimensionality is roughly halved without loss of interpretability, and inference becomes more efficient.

SHAP analyses highlight three largely orthogonal attitudinal axes. Immigration, European integration, and populism structure bloc membership: centre-right voters are

associated with anti-immigration and conservative religious stances; progressive voters with pro-EU and pro-immigrant orientations; and the *M5S* electorate with anti-establishment, eurosceptic, and environmental positions. These results underscore the explanatory leverage of value-laden survey items.

Our conclusions are specific to the 2017–2019 survey waves, the Italian political context, and the operationalisation of variables used here. While the pipeline is modular and could, in principle, be adapted to other settings. To support reproducibility and facilitate such assessments, we publicly release the full codebase and an anonymised version of the dataset at https://github.com/LucaPennella/xai-voting-intentions-italy.

Future work includes: (i) enriching the feature space with time-sensitive sources (e.g., social-media sentiment) to capture rapid opinion shifts; (ii) evaluating external validity by testing the framework on additional Italian waves and other European electorates, and exploring adaptation to non-electoral settings with appropriate outcome measures; and (iii) examining cost-sensitive retraining and data augmentation to improve recall for ideologically heterogeneous blocs such as *M5S*.

Chapter 4

Exploring the Role of Class Overlap in Oversampling Methods for Imbalanced Data

4.1 Introduction

When dealing with imbalanced classification problems, data-level solutions (i.e., modify the distribution of the training set to restore balance by adding or removing instances from the training dataset), such as undersampling and oversampling, are among those that proved successful in building a supervised learning algorithm. The popularity of these methods is due to the simplicity of their application since they are implemented as a step of the data preparation phase in popular software for managing the statistical and machine learning pipeline.

The main element that leads to consider when re-balancing the sample is usually the imbalance ratio (the ratio between the number of units in the prevalent class and the number of units in the rare class), and it is expected that the effectiveness of these solutions is related to cases with high imbalance. Recently, it has been argued that class overlap could matter when dealing with class imbalance [302]. Datasets frequently display varying degrees of class overlap and imbalance, hence it becomes imperative to investigate effective strategies for enhancing classification accuracy in these situations. In this work, we explore the relationship between class overlap and class imbalance: the aim is to find guidance to tune oversampling strategies when the degree of class overlap varies.

To this end, the contribution examines the performance of oversampling techniques, particularly the Synthetic Minority Over-sampling Technique (SMOTE) [82] and Random Over-Sampling Examples (ROSE) [216] under different imbalance ratios with increasing levels of class overlap.

The structure of the chapter is as follows: the section 4.2 introduces the related work; the section 4.3 illustrates the simulation scheme and provides some evidence; and the section 4.4 includes concluding remarks.

4.2 Related Work

Class imbalance and class overlap are two factors that can influence the performance of supervised classification models. While traditional approaches have primarily focused on imbalance ratios, recent studies indicate that class overlap may also play a role in classification performance.

Santos et al. [267] propose a framework that unifies the study of class overlap and imbalance, emphasizing that overlap introduces uncertainty in decision boundaries. Also Ho and Basu [159], show that problem difficulty is influenced by many other factors factors beyond class imbalance.

To address class overlap, Vuttipittayamongkol and Elyan [301] propose a neighborhood based undersampling method that selectively removes majority class instances in overlapping regions, improving classifier performance. In a related study, Vuttipittayamongkol et al. [302] demonstrate that class overlap has a more severe impact on classifier performance than class imbalance alone.

Napierala and Stefanowski [228] analyze different types of minority class instances, emphasizing the importance of local data characteristics in classification performance. The authors suggest that preprocessing techniques should be considered not relying exclusively on imbalance measures.

However, in the previously cited studies, behavior of oversampling methods under varying degrees of class overlap have not been explicitly dealt with. Our work is aimed at analyzing the relationship between class overlap and the effectiveness of oversampling techniques. More specifically, we considered both SMOTE and ROSE to generate synthetic data. Some limited simulations, with different overlap and imbalance levels, provide evidence on possible adaptation of oversampling strategies to improve classification performance.

4.3 Some Evidence From Simulation Studies

The datasets employed in this experiment were constructed to investigate the impact of varying degrees of class overlap and imbalance on the performance of classification models.

We employ a modified version of the data presented in [216], wherein the rare class is conceptualized as a semi-circle devoid of the prevalent class, which is typically distributed in an elliptical manner and possesses elliptical contours. This approach results in the formation of non-linear decision borders.

We consider two levels of class overlap: high ($o = 0.9$) and moderate ($o = 0.6$). The overlap is defined as the convex hull defined over the majority class.

The convex hull of a set of points S in a Euclidean space is the smallest convex set that contains S. The Convex Hull Overlap Region metric evaluates the overlap between classes by considering the convex hull of the points belonging to the majority class in a multi-dimensional feature space. It is defined as the ratio of the number of points from the minority class inside the convex hull of the majority class to the total number of points in the minority class. This metric is formulated as:

$$\text{Overlap Convex Region} = \frac{n_{\text{inside}}}{n_{\text{minority}}}$$

where:

- n_{inside}: number of minority class points inside the convex hull of the majority class.
- n_{minority}: total number of minority class points.

The procedure to calculate this measure involves the following steps:

1. Construct the convex hull around the points of the majority class.
2. Determine how many points of the minority class fall inside this convex hull.
3. Compute the ratio of the minority class points inside the convex hull to the total number of minority class points.

The goal is to introduce a measure of overlap which emphasizes the intrusion of the minority class points within the space spanned by points in the majority class. It is particularly useful in high-dimensional spaces, providing insights into the geometric relationships between the two classes.

We generated two synthetic populations with 100,000 units characterized by overlap level o with a tolerance level of +5%. The mean vector for both classes is set to $\mu_0 = (0, 0)$. The covariance matrix for the minority class is defined as

$$\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix},$$

and, according to the overlapping level o, the covariance matrices for the majority class are

$$\Sigma_0^{0.9} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Sigma_0^{0.6} = \begin{pmatrix} 0.2 & 0 \\ 0 & 1 \end{pmatrix},$$

where the subscripts 0.9 and 0.6 refer to the high and moderate overlap levels. For each population, we extracted four different test sets of 10,000 samples with different imbalance ratios (IR = 1, 9, 19, and 99) with no common units, and with the remaining 60,000 units, we sampled four training sets with the same four imbalance ratios.

Figures 4.1 and 4.2 present the distributions of the training dataset for a single simulation, showcasing varying levels of overlap and IR values. The first figure illustrates a scenario with high overlap, while the second figure shows moderate overlap.

This structured approach to dataset generation enables a comprehensive evaluation of the impact of overlap and imbalance on classification performance; at the end, we have four training sets and four test sets (one balance and the other imbalance) for each degree of overlapping.

We applied both SMOTE and ROSE to the training data to generate synthetic samples for the minority class in order to balance class distribution by oversampling minority instances. As for the use of ROSE, we let h_{mino}, the shrinking factor of the smoothing parameter of minority class, to take on values 0.25, 0.5, and 1. We expect that the two methods can perform differently because SMOTE generates new data points within the

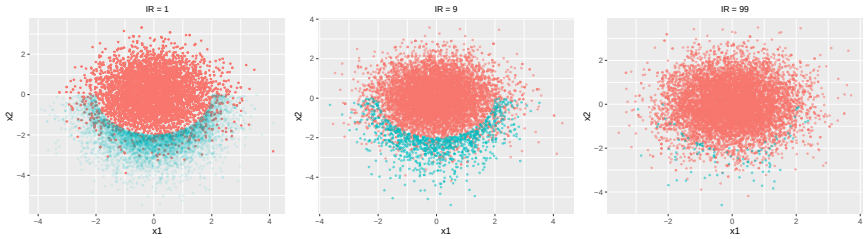


Figure 4.1: Distribution of the training dataset for a single simulation with high overlap and varying IR values

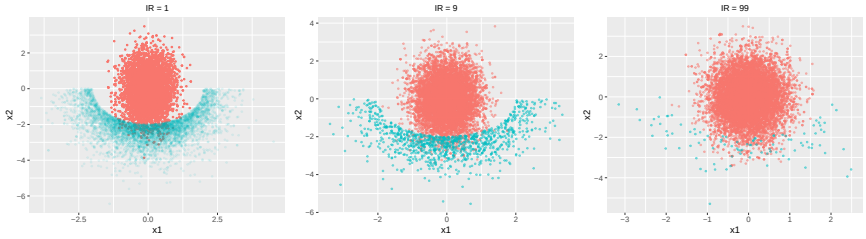


Figure 4.2: Distribution of the training dataset for a single simulation with moderate overlap and varying IR values

convex hull including the points in the minority class, while ROSE generates data according to a smoothed bootstrap approach.

We adopted Decision Tree as our base classification model. We carried out 50 simulations for each experiment, enhancing confidence in the obtained results. The classifier's performances trained on imbalanced data were compared with those obtained using balanced data. The threshold classification value for models trained with imbalanced data is obtained maximizing the Youden's index, while is set to 0.5 for the model trained on balanced data.

The models were assessed using standard classification metrics, i.e., recall and balanced accuracy. For brevity, we do not report results (i) for the imbalance ratio of 19, (ii) for the training test balanced using ROSE with $h_{\text{mino}} = 0.5$, and (iii) for the balanced test sets. However, performances are in line with those presented.

Table 1 shows that, in the case of non-linear decision borders, oversampling results in enhanced performance in terms of recall and balance accuracy across all levels of imbalance ratio and overlapping. In particular, the ROSE method with $h_{\text{mino}} = 1$ exhibits the best recall performances in each scenario under examination. About balance accuracy, SMOTE exhibits the best performance when there is moderate overlap and for a high overlap with an imbalance ratio of 9, whereas ROSE with $h_{\text{mino}} = 0.25$ demonstrates the best balance accuracy when there is high overlap. Regarding standard deviation, results obtained from the imbalanced training sets have, on average, higher variability than those derived from the balanced datasets.

4.4 Discussion

Results from the limited simulation studies with varying levels of overlap and imbalance suggest that oversampling could be more beneficial when severe overlap exists. More-

IR	Dataset	High Overlap ($o = 0.9$)				Moderate Overlap ($o = 0.6$)			
		b-accuracy		Recall		b-accuracy		Recall	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	Balanced	0.946	0.01	0.984	0.02	0.979	0.00	0.998	0.00
9	Imbalanced	0.730	0.05	0.843	0.02	0.889	0.01	0.971	0.02
	SMOTE	**0.941**	0.01	0.975	0.01	**0.914**	0.01	0.994	0.01
	ROSE h_{mino} =0.25	**0.941**	0.01	0.982	0.01	0.906	0.01	0.998	0.03
	ROSE h_{mino} =1	0.927	0.01	**0.989**	0.01	0.872	0.01	**1.000**	0.00
99	Imbalanced	0.727	0.10	0.728	0.10	0.807	0.07	0.754	0.15
	SMOTE	0.928	0.03	0.935	0.06	**0.906**	0.02	0.961	0.04
	ROSE h_{mino} =0.25	**0.936**	0.01	0.971	0.04	0.898	0.01	0.993	0.02
	ROSE h_{mino} =1	0.916	0.01	**0.974**	0.03	0.863	0.02	**0.999**	0.04

Table 4.1: Results of 50 bootstrap simulations of dataset comparisons based on overlap and imbalance ratio, with performance metrics including balanced accuracy (b-accuracy) and recall. The balanced dataset is used as the benchmark. The best performances are highlighted in bold.

over, if a synthetic index measuring class overlap was available, it could help decide if and to what extent oversampling has to be used and which oversampling method could lead to better results.

This line of research seems fruitful, and more extended simulation and application to some real data sets would be necessary to confirm what has been observed. We used a simple overlap measure for illustration, similar to the one already proposed in the literature [302]. Still, more research is needed to suggest a more general index that deals with a larger set number of features, possibly including some categorical measures. Furthermore, the case of data that exhibits a clustered structure and that usually calls for more complex oversampling methods [121], and where measurement of the degree of overlap is not straightforward, should be considered.

Chapter 5

Decision Predicate Graphs: Enhancing Interpretability in Tree Ensembles

5.1 Introduction

Artificial intelligence, although still under strong development, is now a consolidated and widely used tool. This is thanks to the continuous growth of computing power, which allows the use of increasingly complex and computationally expensive machine learning (ML) methods. The challenges presented by modern-world problems are growing in complexity as well as the proposed solutions.

Dealing with large quantities of data and frequently encountering unbalanced datasets are still significant obstacles in addressing many real-world issues; however, tree-based ensemble algorithms offer several advantages in overcoming these challenges, including robustness to noise and outliers, scalability to large datasets, automatic handling of missing values, and the ability to capture complex relationships and interactions within the data [213, 218].

Without delving into the intricacies of various algorithms, exhaustively described by Hastie et al. [152], the process of learning tree-based ensemble models involves training multiple decision trees to combine their predictions, optimising performance, and enhancing generalisation capabilities. While these models indeed offer concrete solutions to a diverse array of problems, developers and users are confronted with new challenges. The training outcome yields an exceptionally complex model, commonly referred to as an opaque-box model (also known as the black-box problem), whose internal workings are not transparent or easily interpretable [90]. In response to this context, the Decision Predicate Graph (DPG) is proposed in this work.

Drawing inspiration from the expanding theme of eXplainable Artificial Intelligence (XAI), we designed a graph structure to tackle transparency and explainability challenges inherent in tree-based ensemble models. This facilitates a better understanding of the intricate choices underlying these ML models. DPG is created with inspiration from the concept of aggregating random forests (RF) [64, 158], as introduced by Gossen and Steffen [137]. The approach proposed in [137] suggests visualising the decisions within the RF by combining the branches of the tree base learners into a single and compact

decision diagram. The concept behind DPG is to convert a generic tree-based ensemble model for classification into a graph, a defined and studied structure with known properties. In this graph, nodes represent predicates, i.e., the feature-value associations present in each node of every tree, while edges denote the frequency with which these predicates are satisfied during the model training phase by the samples of the dataset. The DPG structure enables comprehending the choices made by the model, enhancing transparency and understandability. Moreover, it allows the exploitation of graph properties to develop metrics and algorithms facilitating the analysis of the ensemble model. This, in turn, aids in understanding the decisions it makes, easing the task of visualising the graph which can be vast and complicated for larger models with numerous tree base learners.

DPG serves as a model-specific tool offering a comprehensive interpretation of tree-based ensemble models. It provides descriptive metrics that enhance the understanding of the decisions inherent in the model, offering valuable insights. This tool proves particularly useful for models that are *a priori* considered satisfactory in terms of performance.

Our work contributes in the following ways:

- we introduce DPG, a novel interpretability structure that transforms an opaque-box tree-based ensemble model into an enriched graph;
- we present the algorithm used to create DPG, accompanied by pseudo-code to enhance understanding and facilitate replication, complete with its asymptotic complexity;
- we provide the interpretation of three metrics from graph theory, enriching the model comprehension and gaining insights;
- we demonstrate the use of the proposed method through two case studies: the application of DPG to two RF models, respectively, on the Iris dataset [122] and a challenging dataset.

It is important to highlight that these results are achieved in a generic fashion, utilising a standard classifier on well-established datasets. Significantly, we intentionally avoided incorporating scenario-specific heuristics. Therefore, we posit that our aggregation approach has the potential for widespread application across a diverse range of related scenarios.

5.2 Literature Review

Complex yet effective models have become increasingly prevalent, especially in fields where the outcomes bear significant importance and require a heightened level of sensitivity. In these scenarios, there is a growing demand for models to exhibit transparency, accountability, and a comprehensive understanding. Consequently, the discussion on XAI has experienced significant growth, leading to an extensive body of literature [18, 144, 149].

In XAI, as suggested by Dwivedi et al. [113], classifications are often based on scope, distinguishing between global interpretability, which reveals overall data trends and provides insights into the entire model, and local interpretability, which elucidates the

reasoning behind specific predictions for individual instances. Another classification criterion involves how interpretability is achieved: intrinsic interpretability relies on straightforward model structures (e.g., concise decision trees or linear models), while post hoc interpretability involves methods applied after model training [18]. Interpretation methods are further categorised as model-specific or model-agnostic. Model-agnostic tools are versatile, capable of being employed with any ML model, and are applied post-training (e.g., SHAP [205], LIME [258], PDP [126], and Anchors [257]). In contrast, model-specific interpretation tools are tailored to specific model classes, such as our proposal centred around tree-based ensemble models.

In this context, XAI tools, especially those providing global interpretations, become valuable instruments for understanding tree-based ensemble models. These models are widely used in addressing diverse problem domains, as highlighted in several surveys [35, 91, 147]. As a result, the number of studies delving into model-specific techniques tailored for ensembles of trees has also increased.

The first significant study is proposed by Mashayekhi and Gras [214]. The authors introduced *RF+HC*, an approach that employs a hill climbing algorithm in RF to search for a decision set. This rule set reduces the number of decisions dramatically, which significantly improves the comprehensibility of the underlying model built by RF. Similarly, Hara and Hayashi [150] exploit Bayesian model selection to extract the decision set. These approaches share similarities with our method; however, our proposal extends beyond the extraction of decisions from the RF. Visualising the decisions of tree-based ensemble models and simultaneously complementing them with metrics developed by graph theory makes DPG more adaptable and holistic. This approach allows for obtaining insights into the model beyond just decision information.

Zhao et al. [322] proposed *iForest*, a visual analysis system specifically designed to interpret RF models and their predictions. They built a feature view to illustrate the relationships between input features and outcome predictions and proposed a design that summarises multiple decision paths based on feature occurrences and ranges, allowing users to explore and understand the partitioning logic of these paths. The iForest, like numerous visualisation systems, faces significant challenges related to scalability and interpretability when dealing with large ensembles. To overcome this challenge, our approach does not prioritise visualisation but instead strives for a comprehensive understanding of the ensemble's logic.

Hatwell et al. [153] contributed with *Collection of High Importance Random Path Snippets* (CHIRPS), a method that incorporates the explanation of RF classification for each data instance and extracts a decision path from each tree in the forest, resulting in a set of decisions that elucidate the classification process. However, this method is limited to rules extraction and lacks insight into the model's structure. Additionally, it does not incorporate metrics to explain the logic of the tree-based ensemble model.

Another technique focused on visualising decisions underlying RF models is introduced by Neto and Paulovich [231]. *Explainable Matrix* (ExMatrix) employs a matrix-like visual metaphor, where rows represent decisions, columns denote features, and cells encapsulate decision predicates, thereby facilitating the scrutiny of models and the audit of classification outcomes. The visualisation capability of ExMatrix for global visualisation has limitations in terms of scalability because the number of decisions increases significantly with the number of trees in large ensembles. Moreover, ExMatrix layouts

can rapidly become challenging to explore, while the complexity of the model increases. As mentioned earlier, our approach enables us to retrieve information without relying solely on visualisation.

Dedja et al. [106] introduced an approach denoted as *Building Explanations through a LocalLy AccuraTe Rule EXtractor* (BELLATREX), which is designed to explain the forest predictions for a given test instance by a set of logical rules based on the features of the dataset. However, a potential limitation lies in the computational complexity of the approach. While explaining a single prediction is quick, applying the method to a complete dataset becomes computationally expensive. Furthermore, BELLATREX focuses on instance-level explanations rather than providing a global perspective. In addition, BELLATREX uses clustering techniques to simplify the representation and decision logic, whereas our approach uses graphs to avoid simplifications that can lead to loss of information.

Various studies [108, 146, 297, 325] proposed several tree similarity metrics through the process of clustering based on tree representations. However, these approaches, while beneficial for interpretability, require simplifying the model through techniques such as selection, pruning, and frequency analysis, which can result in information loss.

A number of works [137, 222, 224, 278] established a connection between tree-based ensemble models and graph theory. The foundational concept of these techniques has been explored by [123, 167, 230, 236, 286, 304, 326]. These works established the theory that introduces the transformation of decision trees into graphs, aiming for more efficient and non-redundant tree structures. Nakahara et al. [224] and Silva et al. [278] works focus on performance optimisation, with Gossen and Steffen [137] and Murtovi et al. [222] being the sole contributors that employed these techniques for interpretability purposes.

In particular, Gossen and Steffen [137] introduced the *Algebraic Decision Diagram* (ADD), aiming to transform tree-based ensemble models into bipartite graphs. The ADD serves as an alternative construction to RF, providing an additional predictive model that functions as a surrogate. ADD proves valuable for specific aspects of interpretability, such as outcome explanation, logic of *majority vote*, and visualising the path. Acting as a surrogate model, their primary contributions lie in the ability to provide a simple, optimised model. In comparison, our objective is to extend their ideas by fully leveraging graph theory. This goes beyond the transposition of tree-based ensemble models into a graph; rather, it involves using graph theory and its associated metrics to gain insights into the functionality of the models.

ForestGUMP, an online tool developed by Murtovi et al. [222], is designed for generating ADDs. This tool provides valuable information such as graph visualisation, hypothetical sample path display, and logic of majority vote. However, it has some limitations in terms of the number of usable trees (only 20) and in visualising problems that involve multiple features and classification choices, making graph navigation complex. In our work, while we enable visualisation, our focus is on utilising graph-related metrics, and we do not incorporate the simplification of the analysed models.

5.3 Decision Predicate Graphs

DPGs capture the details of a tree-based ensemble model and learned dataset specifics, emphasising predicate paths while preserving crucial decision points. Essentially, DPG converts complex ensemble models into a graph structure, where nodes represent predicates made by the model and edges denote the occurrence of these predicates during model training.

In this segment, we present the formalisation of DPG, elucidate the algorithm employed in its development through pseudo-code and its asymptotic complexity, and provide a detailed exposition of various metrics and properties essential for comprehending the tree-based ensemble model. Moreover, we meticulously outline the advantages of metrics and articulate why they can serve as a valuable complementary aid to graph visualisation, particularly in overcoming its inherent limitations.

As previously mentioned in 5.1, DPG is tailored for tree-based ensemble models designed specifically for classification tasks.

5.3.1 Definition

Let $\mathcal{M}_n$ be the tree-based ensemble model consisting of n tree base learners $T(x; \Theta_b)$, where x is a generic sample, and Θ_b characterizes the bth learner in terms of split variables, cutpoints at each node, and terminal-node values. More specifically, Θ_b includes:

- all the splitting conditions associated with each jth internal node n_{bj} based on a specific feature f_{bj} and a threshold (for numerical features) or a set of possible values (for categorical features) t_{bj};
- the values assigned to leaf nodes c_b.

Let $\mathcal{D}$ be the training set on which $\mathcal{M}_n$ is trained. Every base learner is trained on a dataset $\mathcal{D}_b$, where $\mathcal{D}_b$ is a subset of $\mathcal{D}$. We define $\mathcal{O}$ as the set of logical operations: $\mathcal{O} = \{\leq, >, =, \neq\}$.

The predicate set $\mathcal{P}(\mathcal{M}_n)$, for an ensemble method $\mathcal{M}_n$ is the set obtained by the union of the set of all the triples $p = (f_{bj}, o, t_{bj})$, where $o \in \mathcal{O}$, and f_{bj}, $t_{bj} \in \Theta_b$, and the set of all leaf nodes c_b, for all n tree base learners of $\mathcal{M}_n$. The triples p are called decisions, while the elements of $\mathcal{P}(\mathcal{M}_n)$ are called predicates.

A Decision Predicate Graph (DPG($\mathcal{M}_n$)) for a model $\mathcal{M}_n$ is a directed weighted graph $(\mathcal{P}, E)$ where:

- $\mathcal{P}$ is the set of nodes, which corresponds to the predicate set $\mathcal{P}(\mathcal{M}_n)$;
- E is the set of edges, where each directed edge connects two predicates if and only if there exists an element in the training set $\mathcal{D}$ that satisfies both conditions specified by the predicates in an immediately consecutive manner. The weight of the edge corresponds to the number of training set elements satisfying these consecutive conditions.

For conciseness, from this point onward, we will use the acronym DPG to refer to the graph, indicating that it is constructed based on a model.

5.3.2 From Ensemble to a DPG

We introduce an algorithm, outlined in Algorithm 1, for constructing the DPG based on a tree-based ensemble model, by traversing all tree base learners with the training samples.

Algorithm 1: Construct DPG from Ensemble Tree Model

Input: Ensemble tree model $\mathcal{M}_n$, Training set $\mathcal{D}$
Output: DPG($\mathcal{M}_n$)

Initialise empty set DPG($\mathcal{M}_n$);
foreach T *(learner)* ***in*** $\mathcal{M}_n$ **do**
    Initialise empty predicate set $\mathcal{P}$ and edge set E;
    foreach x *(sample)* ***in*** $\mathcal{D}$ **do**
        Initialise empty predicate set $\mathcal{P}_x$ and edge set E_x;
        // To obtain the predicates path for x on the tree T
        ($\mathcal{P}_x$, E_x) ← TRAVERSING(T, x);
        Add ($\mathcal{P}_x$, E_x) to $\mathcal{P}$ and E;
DPG($\mathcal{M}_n$) ← AGGREGATING($\mathcal{P}$, E);
return *DPG*($\mathcal{M}_n$);

The algorithm iterates over each base learner in the ensemble tree model $\mathcal{M}_n$ and each training sample x in the training set $\mathcal{D}$.

To clarify, the TRAVERSING function follows the predicate path of a particular input sample x through the decision tree T, starting from the root node and navigating to the appropriate leaf node based on the feature values of x. Meanwhile, the AGGREGATING function processes the predicates and edges obtained from TRAVERSING the decision trees for all samples into a single graph representation, DPG($\mathcal{M}_n$), by taking the union of $\mathcal{P}$ and computing the frequency of elements of E.

The algorithm presents a systematic methodology for constructing DPG. The overall asymptotic complexity can be formally expressed as follows:

$$O(b \times s \times (k + k^2)) = O(b \times s \times k^2)$$

where:

- b is the number of learners in the ensemble,
- s is the number of samples in the training set, $|\mathcal{D}|$, and
- k represents the size of the ($\mathcal{P}_x$, E_x) processed by the TRAVERSING and AGGREGATING functions.

This analysis takes into account the linear time complexity $O(k)$ for the TRAVERSING function and the quadratic time complexity $O(k^2)$ for the AGGREGATING function. Our *Python* 3.10 implementation is accessible here[1].

[1] https://github.com/LeonardoArrighi/DPG

5.3.3 DPG interpretability

In this section, we enumerate and elucidate some of the advantages that DPG can offer. The graph-based nature of DPG provides significant enhancements in the direction of a complete mapping of the ensemble structure. Weighted directed graphs, such as DPGs, are studied structures with well-established properties that enable the identification or construction of useful metrics and algorithms. It is crucial to emphasise that all the observations presented in this section are valuable for comprehending and analysing the obtained model.

Visualisation.

DPG provides an immediate advantage by allowing the visualisation of the entire tree-based ensemble models through a single comprehensive graph. Similar to the idea proposed by Gossen and Steffen [137], consolidating all individual basic learners within the model into a unified graph provides a holistic representation of the decision-making process. This visualisation not only elucidates the decisions made by the learners but also reveals the intricate relationships between them. Consequently, it facilitates a comprehensive understanding of the utilised features and, more importantly, the associations between features and their values that enable the model to accurately classify a sample into a specific class.

Another noteworthy aspect of DPG lies in the representation of weights for pair-wise nodes. This feature enables a discerning analysis of the most significant path through predicates, shedding light on decisions consistently employed by numerous learners or across multiple samples. This insight not only highlights the prevalence of certain decisions but also opens avenues for targeted enhancement strategies, focusing on those influential aspects within the model.

Moreover, by traversing all the possible paths between predicates in reverse, starting from one of the classes, we can discern the essential characteristics that a sample must possess to be classified into a specific class. This capability facilitates the *a priori* elimination of certain elements from the dataset when considering the particular class.

Nevertheless, we acknowledge that while visualisation is a valuable tool, its effectiveness diminishes with an increasing number of tree base learners. A multitude of tree base learners implies an increase in decisions and, consequently, an abundance of predicates. As a result, the size of the graph, in terms of nodes, grows proportionally with the model's scale. This expansion can render the graph illegible or impractical to visualise due to its intricate complexity. To address this challenge, we provide additional tools that complement the visualisation, aiding in the extraction of model properties and facilitating a more comprehensible understanding.

One approach is based on the desire and feasibility of determining the specific characteristics a sample must exhibit to be assigned to a particular class. Taking inspiration from the *outcome explanation problem* introduced by Gossen and Steffen [137], to enhance the immediacy and effectiveness of this analysis, we provide an aggregation of predicates, referred to as *constraint*, which represent intervals associated with the features of each class. The constraints are defined as follows: for a given class identified in the DPG, we list all nodes connected by a path originating from the node itself and culminating in the class. For each feature within the node predicates, we delineate the

most extensive possible interval using the values associated with the features. This interval is defined by two endpoints. The minor endpoint is the smallest value within the set of values less than the feature, while the major endpoint is the largest value within the set of values greater than the feature. If either of these two sets is empty, the interval is deemed infinite. Each class has its constraints for every feature contributing to the classification of the samples.

It is important to note that constraints are not a substitute for visualisation; instead, they offer insights resembling those promoted by it. It is anticipated that additional graph measures could complement insights from constraints, providing similar interpretations as visualisation.

Centrality.

The centrality of a node is defined as a number or rank corresponding to the node position within the network. By observing centrality, we can make considerations that allow us to better understand the process hidden in the ensemble method. The notion of centrality encompasses a wide range of metrics. In this section, we explore those metrics that offer the most insightful information.

According to Brandes [61], we define the *betweenness centrality* (BC) of a node as the fraction of all the shortest paths between every pair of nodes of the graph passing through the considered node. Let DPG = $(\mathcal{P}, E)$ be the graph and $s, t, v \in \mathcal{P}$ three vertexes of DPG, we can denote with $\sigma(s, t)$ the number of shortest paths between s and t and with $\sigma(s, t|v)$ the number of shortest paths between s and t passing through v. Then, the BC of the node $v \in \mathcal{P}$ is defined as:

$$BC(v) = \sum_{s,t \in \mathcal{P}} \frac{\sigma(s, t|v)}{\sigma(s, t)}.$$

All details and observations about BC can be read in [61]. BC serves as a relevant metric for gaining a deeper insight into the significance of decisions within the ensemble model. We can observe that a node with a higher BC value has a more significant influence on the flow of information within the graph; nodes with high BC can be considered potential bottleneck nodes because they play a crucial role in facilitating interactions between different parts of the DPG. For this reason, we can assert that these nodes are meaningful to understanding the tree-based ensemble models: in all tree base learners, the decision contained in the node is essential to classify the elements of the dataset. We highlight that the significance extends beyond the characteristic itself; it encompasses the value associated with it.

According to Mones et al. [220], we define the *local reaching centrality* (LRC) of a node v of the DPG as the proportion of other nodes reachable from node v via outgoing edges. LRC can be generalised to weighted graphs by measuring the average weight of a given directed path starting from node v (more details are available on [220]). The LRC serves as a metric for assessing the importance of DPG's nodes. It gauges the extent to which decisions contained in these nodes are employed by diverse tree base learners for classifying samples in the training set. This, in turn, reflects the importance of these decisions in the classification of new samples. The LRC offers a comprehensive perspective on the concept of feature importance (FI) by extending its definition to encompass the

values associated with features across various decisions. Additionally, the prominence of paths between highlighted predicates indicates their frequent utilisation, providing insights into how new samples can be classified with fewer decisions.

Community.

While there is no single definition of a community, we can observe structures similar to communities in the DPG. According to Radicchi et al. [252], we define a *community* as a subset of nodes of the DPG characterised by dense interconnections between its elements and sparse connections with the other nodes of the DPG that do not belong to the community. Based on the properties of DPG, we employed *asynchronous label propagation* algorithm, proposed by Raghavan et al. [253], to detect communities within the graph. The core concept of the algorithm involves each graph node determining its community membership based on the majority of its neighbours. Without delving into details, which can be appreciated in [253], the algorithm comprises a series of steps: each node initially possesses a unique label. As these labels diffuse through the graph, closely connected groups of nodes converge on a common label. These consensus groups then expand outward until further expansion becomes impractical. After this label propagation process, nodes sharing the same labels are identified as belonging to the same community. This process is iterated until each node in the network aligns its label with the community that includes the maximum number of its neighbouring nodes. The algorithm is defined asynchronous, as each node receives updates without waiting for updates on the remaining nodes. Identifying communities in the DPG provides insights into the ensemble model: visualising them allows us to discover groups of nodes that similarly contribute to the classification of samples.

By employing the asynchronous label propagation algorithm, we observe that each formed community is associated with a class. Within these communities, the features utilised by the ensemble model to classify samples belonging to the community's class are emphasised. Once again, the association between features and values plays a key role, highlighting the specific decisions made by the learners. To quote Raghavan et al. [253]:

> Communities in social networks can provide insights about common characteristics or beliefs among people that make them different from other communities.

Similarly, we observe that communities within the DPG offer a valuable understanding of the characteristics for samples to be assigned to a particular community class. This intuition extends to identifying predominant features and those that play a marginal role in the classification process. Moreover, it is noteworthy that communities also provide insights into the entire dataset and the complexity of the problem. A community comprising a substantial number of nodes, each associated with different predicates often involving distinct features, indicates that the model makes diverse decisions to assign samples to the community class. This implies that the model encounters challenges in classifying samples for this particular class, and data from different classes are not easily distinguishable within the dataset.

Finally, communities, functioning as sub-graphs, can be used to visualise the decisions made in the ensemble model, enabling the identification of a specific class. This

replaces the complex illustration of the DPG, especially when we are focused on visualising a single class and dealing with many tree base learners. We summarised the utility of discussed properties and metrics in 5.1.

Table 5.1: Summary of Constraints, Betweenness Centrality (BC), Local Reaching Centrality (LRC), and Community, featuring provided definitions and their utility in offering insights into tree-based ensemble models.

Property	**Definition**	**Utility**
Constraints	The intervals of values for each feature obtained from all predicates connected by a path that culminates in a given class.	Calculate the classification boundary values of each feature associated with each class.
BC	Quantifies the fraction of all the shortest paths between every pair of nodes of the graph passing through the considered node.	Identify potential bottleneck nodes that correspond to crucial decisions.
LRC	Quantifies the proportion of other nodes reachable from the local node through its outgoing edges.	Assess the importance of nodes similarly to feature importance, but enrich the information by encompassing the values associated with features across all decisions.
Community	A subset of nodes of the DPG which is characterised by dense interconnections between its elements and sparse connections with the other nodes of the DPG that do not belong to the community.	Understanding the characteristics of nodes to be assigned to a particular community class, identifying predominant predicates, and those that play a marginal role in the classification process.

5.4 Empirical Results and Discussion

In this section, we demonstrate the effectiveness of DPG to the well-known Iris dataset [122] and a synthetic multiclass dataset[2]. Each experiment in this section was conducted using the RF classifier, with variations limited to the number of tree base learners. The implementation is available here[3]. Finally, we discuss potential enhancements to DPG

[2] https://github.com/LeonardoArrighi/DPG/datasets
[3] https://github.com/LeonardoArrighi/DPG

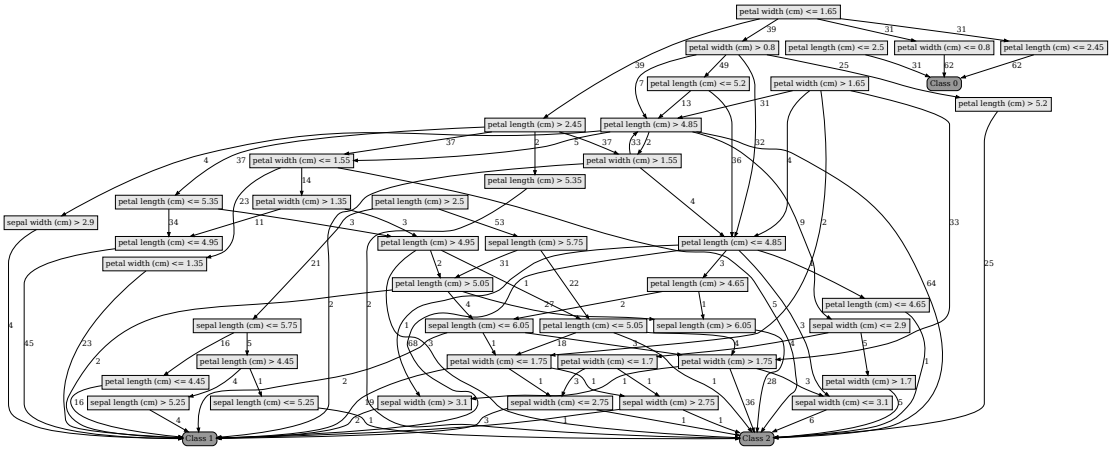


Figure 5.1: DPG of the RF composed of 5 tree base learners trained on Iris dataset.

and explore further development opportunities.

5.4.1 DPG: Iris insights

The first case study concerns the classification of the Iris dataset [122]. The simplicity, manageability, versatility, and relevance of this dataset make it an interesting and relevant resource for discussions and demonstrations of interpretability in ML. The dataset comprises measurements of sepals and petals for iris flowers encompassing three distinct species with a total of four features and three classes. The RF was selected as the tree-based ensemble model due to its well-established reputation and high-performance capabilities.

To conduct the classification, we partitioned the dataset into training and test sets, following a 80-20% proportion, respectively. A seed value of 42 was established for randomness control, and the number of base tree learners was set to 5. The RF performances, evaluated on the test set, are summarised in the confusion matrix in 5.2. The model demonstrates 100% accuracy.

Table 5.2: Confusion matrix depicting the performance evaluation of the RF model with 5 base tree learners on the test set.

	Prediction		
Ground truth	Class 0	Class 1	Class 2
Class 0	19	0	0
Class 1	0	13	0
Class 2	0	0	13

After training the model, we applied the algorithm outlined in 5.3.2 to obtain the DPG, which can be visualised in 5.1. Then, we can analyse the obtained graph using the metrics and algorithms proposed in 5.3.3, verifying their utility and effectiveness in gaining insights into the model. It is important to note that the DPG leads to the

calculation of both global metrics, referring to the overall graph, and metrics at the level of individual nodes.

To illustrate the effectiveness and one of the advantages of employing DPG, we highlight the constraints for the different classes in the 5.3. The class-specific constraints delineate the necessary characteristics a sample must exhibit to be assigned to that particular class by the tree-based ensemble model. This insight contributes to a better understanding of how the model utilises features for effective classification.

Table 5.3: Constraints for each class based on the DPG for an RF model within 5 tree base learners.

Class	Constraints
0	`petal width (cm)` ≤ 1.65 `petal length (cm)` ≤ 2.50
1	$5.25 <$ `sepal length (cm)` ≤ 6.05 $0.80 <$ `petal width (cm)` ≤ 1.75 $2.45 <$ `petal length (cm)` ≤ 5.35 $2.75 <$ `sepal width (cm)` ≤ 2.90
2	$5.75 <$ `sepal length (cm)` ≤ 6.05 $0.80 <$ `petal width (cm)` ≤ 1.75 $2.45 <$ `petal length (cm)` ≤ 5.35 $2.75 <$ `sepal width (cm)` ≤ 3.10

The first metric under discussion is the BC of the nodes, as depicted in 5.4, where we identify potential bottleneck nodes. These nodes encapsulate significant information, particularly representing decisions made by numerous tree base learners. We quickly discern that the decision associated with `petal length (cm)` and the value 4.85 is pivotal, as it is frequently relied upon by multiple basic learners and is essential for successful classification.

Furthermore, the LRC metric provides additional information. Examining 5.5a, we can emphasise the most crucial predicates influencing the decision-making process of the tree-based ensemble model. This includes not only identifying the most frequently used features but also recognising the associated values that lead to significant and divisive splits in the dataset across various basic learners. As observed in 5.5, a comparison between the LRC of the nodes and the FI, calculated on the same model on which DPG is based, suggests that the metric may rank the predicates similarly. FI is calculated using the *Mean Decrease Impurity* (MDI) algorithm introduced by Breiman [64]. This compar-

Table 5.4: Top eight predicates by evaluating their BC obtained from the DPG based on an RF model consisting of 5 tree base learners.

Predicate	BC
petal length (cm) > 4.85	0.053
petal length (cm) <= 4.85	0.036
petal width (cm) > 1.55	0.034
sepal length (cm) <= 6.05	0.032
petal length (cm) > 4.95	0.028
petal length (cm) > 4.65	0.022
petal width (cm) <= 1.75	0.022
petal width (cm) <= 1.55	0.021

ison also provides additional information about the values used in the decisions and the frequency of paths extending the concept of FI.

Table 5.5: Comparison of the top eight predicates by evaluating their LRC obtained from the DPG based on an RF model consisting of 5 tree base learners (5.5a), alongside the FI of the same model (5.5b) calculated exploiting MDI algorithm.

(a) LRC evaluation

Predicate	LRC
petal width (cm) <= 1.65	1.531
petal length (cm) > 2.45	0.919
petal width (cm) > 0.80	0.874
petal length (cm) > 2.50	0.699
petal width (cm) > 1.65	0.618
petal length (cm) <= 5.20	0.565
petal width (cm) > 1.55	0.540
sepal length (cm) > 5.75	0.332

(b) FI evaluation

Feature	FI
petal length (cm)	0.550
petal width (cm)	0.373
sepal length (cm)	0.054
sepal width (cm)	0.023

By employing the global metric community, we identified the presence of three distinct communities. 5.6 illustrates the association between each community and a distinct class obtained by applying the asynchronous label propagation algorithm to the DPG. We can affirm that each node within a community contains decisions that significantly contribute to the accurate classification of samples belonging to a specific class. For instance, when applying the predicates of Community 3 (comprising two features and two predicates) to the test set and traversing from the root node, it achieved 100% accuracy for Class 0, the class delineated in the mentioned community.

Finally, upon comparing the 5.6 and the 5.2, we can state that the communities facilitate the comprehension of how the model addresses the classification problem. Examining the number of decisions and features utilised in each community reveals that differentiating between Class 1 (Community 1) and Class 2 (Community 2) poses a greater challenge for the model. This indicates the difficulty in effectively separating samples within the dataset into their respective classes. This difficulty becomes apparent when

Table 5.6: Communities obtained from the DPG based on an RF model composed of 5 tree base learners. The table shows the number of predicates belonging to each community, the number of features in the community nodes, and the class involved in each community.

Community	# Predicates	# Features	Class
Community 1	23	4	1
Community 2	18	4	2
Community 3	4	2	0

visualising the dataset across the features, as in 5.2. Conversely, the community encompassing Class 0 (Community 3) consists of fewer predicates, signifying that it is more distinguishable from other classes, as confirmed in the 5.2.

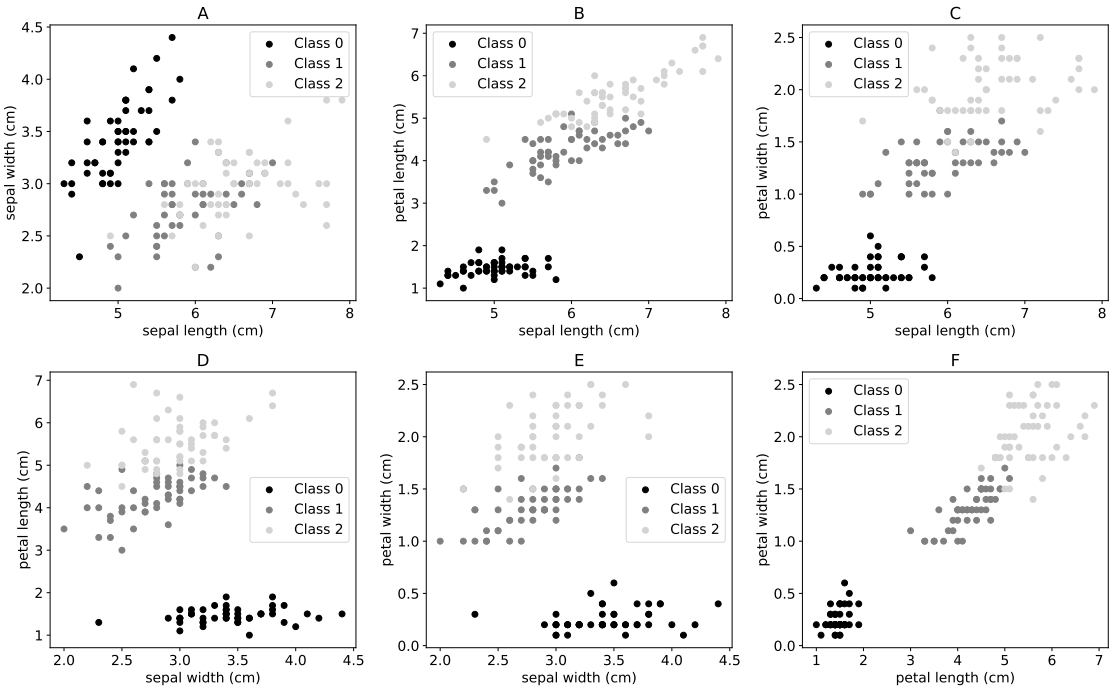


Figure 5.2: Two-dimensional depiction of the Iris dataset, employing feature pairs in each graph for visual representation.

5.4.2 Comparing to the Graph-based Solutions

As outlined in the 5.2, Gossen and Steffen [137] and Murtovi et al. [222] conducted studies on the interpretability of tree-based ensemble models exploiting graph structures. We compared DPG with their proposed method by examining their outcomes and potential insights. Using the same tree-based ensemble model we studied in 5.4.1 as input, we generated the ADD displayed in 5.3. The first noticeable distinction from DPG lies in the ADD structure, as it forms a bipartite graph. We can observe that the ADD is

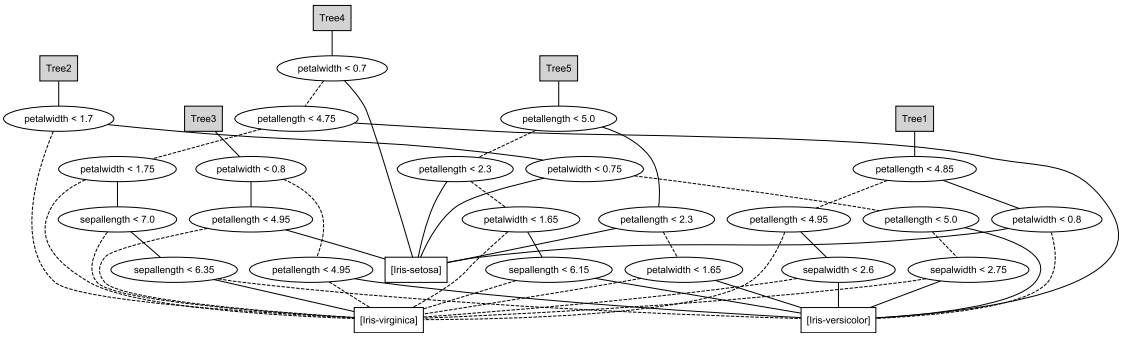


Figure 5.3: ADD of an RF model with 5 tree base learners induced for Iris dataset.

generated from the trained model, albeit without fully leveraging the training dataset. Consequently, the evaluation of connections between nodes and the assessment of the significance of decisions made by different tree base learners are not fully exploited. This implies that each branch carries equal weight and impact in the diagram, and a classification error significantly influences the structure by introducing an incorrect path. Another crucial difference is that ADD does not provide graph metrics, leaving the user the interpretation of the diagram and potentially missing out on relevant information. Moreover, an additional limitation, as indicated in the studies by [137, 222], involves the challenge that emerges when generating ADDs and dealing with large ensembles. Visualisation becomes intricate, even with a modest count of 20 tree base learners. In contrast, DPG allows the computation of both global and local metrics, even with a higher number of tree base learners.

To further examine these aspects, we use two RF models, one with 20 tree learners and the other with 100, to analyse a complex multiclass problem with a dataset comprising 4 classes, 1000 samples, and 16 features. This introduces a four-class problem that was randomly generated. The dataset was created using the `make_classification`[4] function from `scikit-learn`. The following setup has been maintained for training both models. We divided the dataset into training and test sets, with an 80-20% ratio. We fixed a seed value of 42 for randomness control to ensure reproducibility. The RF performances, assessed on the test set, are summarised in the confusion matrices presented in 5.7. The model with 100 tree base learners shows better performance in every parameter, reaching an overall accuracy of 66%, outperforming the model with 20 tree base learners, which barely reaches 58%.

We emphasise that even in this context, DPG is a useful tool. Both DPG and ADD present intricate visualisations with 20 tree base learners. However, DPG overcomes this obstacle by providing metrics that can still offer valid insights into the model. The first insight is displayed in 5.8, where we provide constraints for the four classes of the dataset. Constraints, even in this complex scenario, allow the visualisation of intervals where sample features should be situated for precise classification into their respective classes.

The BC metric helps identifying potential bottleneck nodes. Upon observing 5.9a, we can see that there is not a large difference between the BC values associated with the

[4] https://scikit-learn.org/stable/modules/generated/sklearn.datasets.make_classification.html

Table 5.7: Confusion matrices of the RF models with 20 tree base learners (RF 20) and with 100 tree base learners (RF 100) tested on the synthetic dataset.

	Prediction (RF 20)				Prediction (RF 100)			
Ground truth	Class 0	Class 1	Class 2	Class 3	Class 0	Class 1	Class 2	Class 3
Class 0	38	4	12	6	38	4	14	4
Class 1	5	31	3	5	4	32	2	6
Class 2	11	2	29	2	5	2	33	4
Class 3	10	13	10	19	9	9	5	29

Table 5.8: Constraints for each class based on the DPG for an RF model within 100 tree base learners.

Class 0	Class 1	Class 2	Class 3
$-5.87 < \text{F1} \leq 5.74$	$-5.79 < \text{F1} \leq 5.72$	$-5.76 < \text{F1} \leq 5.72$	$-5.76 < \text{F1} \leq 5.72$
$-2.64 < \text{F2} \leq 2.63$	$-2.61 < \text{F2} \leq 2.61$	$-2.61 < \text{F2} \leq 2.61$	$-2.61 < \text{F2} \leq 2.63$
$-5.24 < \text{F3} \leq 3.75$	$-5.38 < \text{F3} \leq 3.75$	$-5.24 < \text{F3} \leq 3.75$	$-5.24 < \text{F3} \leq 3.75$
$-4.80 < \text{F4} \leq 4.37$	$-5.15 < \text{F4} \leq 4.37$	$-4.80 < \text{F4} \leq 4.37$	$-4.80 < \text{F4} \leq 4.37$
$-2.61 < \text{F5} \leq 2.44$	$-2.61 < \text{F5} \leq 2.71$	$-2.61 < \text{F5} \leq 2.44$	$-2.61 < \text{F5} \leq 2.44$
$-2.29 < \text{F6} \leq 2.58$	$-2.29 < \text{F6} \leq 2.58$	$-2.29 < \text{F6} \leq 2.58$	$-2.29 < \text{F6} \leq 2.58$
$-2.82 < \text{F7} \leq 2.47$	$-2.82 < \text{F7} \leq 2.47$	$-2.82 < \text{F7} \leq 2.47$	$-2.82 < \text{F7} \leq 2.86$
$-4.62 < \text{F8} \leq 4.74$	$-4.62 < \text{F8} \leq 4.74$	$-4.62 < \text{F8} \leq 4.74$	$-4.62 < \text{F8} \leq 4.74$
$-2.40 < \text{F9} \leq 2.59$	$-2.40 < \text{F9} \leq 2.59$	$-2.40 < \text{F9} \leq 2.71$	$-2.40 < \text{F9} \leq 2.59$
$-4.71 < \text{F10} \leq 4.32$	$-4.71 < \text{F10} \leq 4.32$	$-4.71 < \text{F10} \leq 5.43$	$-4.71 < \text{F10} \leq 4.32$
$-2.87 < \text{F11} \leq 2.77$	$-2.87 < \text{F11} \leq 2.77$	$-2.87 < \text{F11} \leq 2.86$	$-2.87 < \text{F11} \leq 2.77$
$-2.42 < \text{F12} \leq 2.37$	$-2.42 < \text{F12} \leq 2.37$	$-2.42 < \text{F12} \leq 2.37$	$-2.42 < \text{F12} \leq 2.37$
$-4.28 < \text{F13} \leq 5.01$	$-4.28 < \text{F13} \leq 5.01$	$-4.28 < \text{F13} \leq 5.01$	$-4.28 < \text{F13} \leq 5.01$
$-7.31 < \text{F14} \leq 8.33$	$-7.31 < \text{F14} \leq 8.33$	$-7.31 < \text{F14} \leq 8.33$	$-7.31 < \text{F14} \leq 8.33$
$-4.46 < \text{F15} \leq 4.19$	$-4.46 < \text{F15} \leq 4.19$	$-4.46 < \text{F15} \leq 4.19$	$-4.97 < \text{F15} \leq 4.5$
$-4.70 < \text{F16} \leq 4.21$	$-4.70 < \text{F16} \leq 4.21$	$-4.70 < \text{F16} \leq 4.52$	$-4.70 < \text{F16} \leq 4.21$

predicates. Therefore, we can assume that there are no bottleneck nodes.

Examining the information provided in 5.9b, the LRC underscores which predicates significantly impact the decision-making process of the ensemble model.

Another insight can be obtained by employing the global metric community. In this scenario, we identified the presence of four distinct communities, displayed in 5.10. We note that each community contains a distinct class. Furthermore, upon observing the table, we can conclude that each community exhibits a high number of involved features and predicates, confirming the complexity of the classification problem.

5.4.3 Potential Improvements

Several avenues await exploration in the future. The primary aim is to reduce the computational cost of DPG, as many real-world problems involve large datasets that do not scale well with the current implementation of DPG. Expanding the application scope of

Table 5.9: Top eight predicates by evaluating their BC (5.9a), and top eight predicates by evaluating their LRC (5.9b), both obtained from the DPG based on an RF model consisting of 100 tree base learners.

(a) BC evaluation

Predicate	BC
F15 > 1.17	0.018
F15 <= 1.61	0.015
F12 > 0.20	0.015
F12 > 0.41	0.014
F4 <= 0.33	0.014
F8 > 0.36	0.014
F1 <= -1.10	0.013
F11 <= 1.16	0.013

(b) LRC evaluation

Predicate	LRC
F7 <= 1.62	15.812
F1 <= 3.10	14.475
F14 > -1.78	13.313
F4 > -2.97	13.158
F5 > -1.92	13.065
F4 > -1.36	12.989
F1 <= 2.49	12.986
F13 > 1.98	12.920

Table 5.10: Communities obtained from an RF model composed of 100 tree base learners. The table shows the number of predicates belonging to each community, the number of features in the community nodes, and the class involved in each community.

Community	# Predicates	# Features	Class
Community 1	7767	16	2
Community 2	2149	16	0
Community 3	2351	16	3
Community 3	2100	16	1

DPG is another key goal, including its utility in explaining models relevant to regression-type problems. Given DPG's applicability to any model and dataset, we aim to introduce new tests and use cases to delve deeper into the method. This includes proposing applications to novel datasets and exploring their compatibility with other tree-based ensemble models. Furthermore, while this contribution introduces certain metrics and algorithms derived from graph theory, the field offers extensive possibilities for future exploration. In the future, we plan to introduce new tools associated with DPG to enhance the interpretation of tree-based ensemble models.

5.5 Conclusion

In this work, we introduced Decision Predicate Graphs (DPG) as a novel model-specific tool for tree-based ensemble interpretability. DPG is obtained from a trained model and data, ensuring the maintenance of its performance. The concept behind DPG is to convert an opaque-box tree-based ensemble model into an enriched graph. DPG enables graph-based evaluations and the identification of model decisions towards facilitating comparisons between features and their associated values while offering insights into the entire model. In particular, we introduced Betweenness Centrality, Local Reach-

ing Centrality, Community and Constraints as useful metrics and properties towards improving and extending the XAI interpretability approaches. While DPG is still considered an evolving work, its potential is substantial, given the robust underlying theory and the versatility of the tool. Furthermore, the effervescent research on graphs, knowledge graphs, and complex networks might strengthen the possibilities grounded in DPG. As the next step of our current research, we expect to apply DPG to provide local interpretability and to enhance the scalability of the current implementation.

Part II

Characterize investor types and regulatory attitudes in cryptocurrency markets

Part II moves from methodological development to empirical evidence on cryptocurrency markets, focusing on market participants and their beliefs. The aim is to characterize investor heterogeneity and to examine how individual perceptions and exposure to cryptoassets relate to regulatory preferences, using survey data as a consistent measurement layer.

This part brings together two complementary studies. The first profiles memecoin holders as a distinct subgroup within the broader crypto population, documenting systematic demographic and psychological patterns and linking them to recognizable trading behaviors and greater engagement with speculative practices. The second investigation analyzes support for different regulatory domains, including KYC requirements and taxation, and shows that attitudes are heterogeneous and systematically associated with perceived market illegitimacy and with personal exposure to crypto wealth, which correlates with lower support for regulation.

Chapter 6

Meme Money, Real People: Decoding the Crypto Memecoin Crowd

6.1 Introduction

On February 16, 2025, Argentine President Javier Milei endorsed the LIBRA memecoin, a cryptocurrency that experienced a meteoric rise and an equally dramatic collapse [186]. Initially promoted as a tool to support Argentine projects, LIBRA's price skyrocketed to over $4 per token, only to plummet below $0.50 within hours. Investors accused the token's creators of orchestrating a "rug pull", a scheme in which early insiders sell off their holdings after driving up the price, leaving retail investors with substantial losses [233]. The fallout intensified scrutiny of the memecoin market[1], a segment of the cryptocurrency ecosystem that has become increasingly characterized by speculative investment or fraud [178].

Unlike cryptocurrencies such as Bitcoin or Ether, which were originally designed to create alternative financial systems resistant to state control [69, 226], memecoins have emerged as a distinct phenomenon within the Web3 ecosystem [202]. Rather than serving a clearly defined financial or technological purpose, they are largely shaped by cultural narratives and social media dynamics. Their appeal is driven primarily by speculation and online virality, and many exhibit limited practical utility [191, 274]. Memecoins market value is often dictated by celebrity endorsements[2], online communities, and social media hype. The intersection between politics and memecoin speculation has become increasingly visible, illustrated not only by the LIBRA initiative but also by the launch of politically branded tokens such as $TRUMP[3] and $MELANIA coins.[4] This connection between political identity and speculative digital assets is more than symbolic. [128] examine the effects of Donald Trump's official $TRUMP memecoin, introduced during the 2024 U.S. presidential campaign, and document significant volatility spillovers and contagion across major cryptocurrencies.

As memecoins gain traction within the wider cryptocurrency ecosystem, understand-

[1] Appendix A provides a comprehensive overview of the historical development of memecoins.

[2] In January 2022, Tesla began accepting Dogecoin (DOGE) as payment for select merchandise items, such as the "Cyberwhistle" and the "Giga Texas Belt Buckle" (see https://www.foxbusiness.com/markets/elon-musk-tesla-dogecoin-merchandise?).

[3] See https://www.cbsnews.com/news/trump-launches-own-meme-coin-cryptocurrency/

[4] See https://www.bbc.com/news/articles/c98y47vrv2jo.

ing of the retail investors driving this activity remains limited. Key aspects such as their demographic traits, investment strategies, risk tolerance, trust in government, perceptions of financial markets, and familiarity with both traditional and crypto-related financial education remain underexplored. Recent research highlights that memecoins have begun to constitute a distinct market niche, marked by extreme volatility, unique community dynamics, and particular investor behavior [189]. The objective of our manuscript is to investigate whether memecoin investors constitute a unique subset within the broader crypto investment ecosystem, differing in their motivations, behavior, and approach to risk.

Our results show that memecoin investors consistently differ from the broader crypto population across several dimensions. Individuals with specific knowledge about memecoins are significantly more likely to invest in them, suggesting that familiarity with this asset class plays a central role in adoption. Being male, active trading behavior in the crypto sphere, use of leverage, and ownership of NFTs also emerge as strong predictors of memecoin investment, indicating a pattern of engagement with high-risk and speculative financial instruments. Second, we find that portfolio diversification within the crypto space further increases the likelihood of holding memecoins, as each additional non-memecoin token correlates positively with memecoin ownership. Our interpretation is that memecoin investors may not be casual participants but rather deeply embedded in the crypto ecosystem. Additionally, lower scores on the need-to-belong scale are consistently associated with memecoin investment, pointing to a psychological profile marked by individualism and reduced social conformity. These findings remain robust across all model specifications and highlight a distinct behavioral and financial profile among memecoin investors.

This manuscript contributes to the recent and burgeoning literature on retail investor behavior, speculative trading dynamics, and cryptocurrency market participation. First, we extend the academic work related to this emerging asset class in the crypto ecosystem. Some studies highlight that memecoins exhibit unique characteristics such as community-driven valuation and a strong dependence on online trends [182, 194, 200]. In fact, [287] stress that social media plays a key role in influencing memecoin valuation. [182] use network modeling to analyze Twitter interactions among memecoin communities, identifying key nodes that amplify market sentiment. The speculative nature of memecoins has also been widely discussed. [268] explore fraudulent activities associated with memecoins, including pump-and-dump schemes and honeypots and [316] apply quantile connectedness models to examine the impact of memecoins on broader financial markets, suggesting that they serve as conduits for speculative shocks. In parallel, a growing number of studies have approached memecoin investment from a behavioral finance perspective: for instance, [246] explore the connection between meme asset speculation and gambling behaviors, finding that investors in memecoins often display high levels of overconfidence and risk-seeking tendencies, and [31] examine the psychological and social motivations behind memecoin adoption, identifying both extrinsic drivers—such as the pursuit of social belonging—and intrinsic traits like novelty-seeking.

Second, this study adds to the literature on retail participation in cryptocurrency markets. So far, research has rarely differentiated between investor profiles within the crypto space, often treating cryptocurrency users as a relatively uniform group[5]. Sur-

[5]There are some exceptions. [45] study the profile of NFT investors, finding that they are a particular

vey data from Canada, Austria, Japan and the U.S. show that cryptocurrency investors are predominantly male, younger, better educated, financially literate, and more willing to accept financial risk [38, 127, 156, 281]. In Brazil and Spain, recent findings point out that Brazilian crypto investors tend to be young men with high self-perceived investment expertise and elevated risk tolerance [96], while in Spain, adoption remains lower among women due to limited prior investment experience, lower familiarity with cryptocurrencies, and concerns related to security and perceived financial risks [26].

Our results that memecoin investors have low need for social belonging is related to the current research on the behavioral and psychological features of crypto investors. Some papers report that Bitcoin investors exhibit distinct personality traits and higher levels of psychological distress [181], others that crypto users tend to have greater levels of stress, loneliness, and involvement in gambling and online games [235]. [198] find that individuals who own cryptocurrencies are more likely to hold conspiratorial beliefs, the presence of dark personality traits and use alternative or non-mainstream social media platforms more frequently.

Our results offer practical policy implications for regulators and market designers seeking to mitigate the systemic, financial and consumer risks posed by speculative crypto-assets [176], particularly memecoins. First, existing disclosure regimes are insufficient for addressing the behaviorally distinct and informationally disadvantaged segment of memecoin investors. To address this gap, policymakers should mandate more explicit volatility and liquidity risk labels for memecoins listed on centralized exchanges, along with enhanced transparency around token distribution, developer holdings, and project governance structures. Moreover, given the prominent role of social media in shaping market sentiment, regulatory authorities should extend enforcement tools to cover promotional activity by influencers. These are examples of measures that would reduce asymmetries in information access and help restore a degree of market integrity where traditional fundamentals are largely absent.

Second, policy responses should account for the fact that memecoins are driven by digital culture and gamified investment behavior, which conventional financial education frameworks do not adequately address. Investor outreach initiatives should be redesigned to reflect the media habits of this demographic, using platform-specific, visually engaging formats disseminated through social networks that often amplify speculative behavior. At the same time, the rapid growth of memecoin markets has raised broader regulatory and financial stability concerns. [232] provide a systematic review of meme stocks and their parallels with memecoins, emphasizing the risk of manipulation and extreme volatility. Similarly, [27] show that meme assets are prone to price explosiveness and vulnerable to coordinated trading activity. These dynamics challenge the capacity of existing regulatory frameworks to protect retail investors and ensure orderly market functioning. More recently, [247] highlight the political implications of memecoins, focusing on the rise of 'PolitiFi' tokens and their potential influence on election campaigns. Together, these developments underscore the need for stronger international coordination to address regulatory arbitrage, where malicious actors exploit inconsistent rules across jurisdictions. Establishing shared standards for token vetting, exchange listing practices, and enforcement mechanisms would enhance global market integrity and reduce retail exposure to high-risk, manipulative assets.

group within the broad crypto universe.

The structure of the chapter is as follows. Section 2 details the data sources, describes the construction of the dataset, and explains the empirical methodology employed to identify the effects under study. Section 3 presents the main findings, accompanied by robustness checks to assess their validity. Section 4 offers concluding remarks.

6.2 Data and methodology

6.2.1 Data Collection

The data utilized in this study are drawn from the second edition of the State of Crypto Survey.[6] That edition of the State of Crypto Survey was conducted between December 2024 and February 2025, using a broad array of social media platforms to reach a global audience. Special emphasis was placed on X (formerly Twitter), widely recognized as a key hub for crypto-related discourse, and Reddit, which hosts a range of active and technically engaged cryptocurrency communities. To promote inclusivity and ensure representation across global crypto communities, the survey was translated into 30 languages.

Data were collected through the survey from 2,219 respondents, 1,414 of whom completed it. Respondents self-reported to be from more than 150 countries, with the largest shares coming from Nigeria (9.1%), Turkey (8.7%), Ethiopia (5.0%), Germany (4.2%), Spain (3.7%), and the United States (3.7%). The most common languages are English (54.3%), Turkish (10.57%), Spanish (6.7%), Amharic (5.2%), Chinese (4.6%), German (3.1%), and Italian (3.1%).

The survey gathered information about several variables of crypto and non-crypto investors, including demographic and socioeconomic variables (gender, education, age, race, employment situation), financial literacy, risk profiles and investing profiles. Additionally, the survey explored issues such as inequality, trust in the government, and the perceived purpose of crypto.

The survey was implemented in nodeGame [42], was preregistered on AsPredicted.org, received IRB clearance (EK Mannheim 40/2023), and was checked for GDPR compliance by the University of Mannheim's Data Protection Officer. For additional information about how the data were treated, see Appendix C.

6.2.2 Main data

The aim of this study is to examine whether memecoin investors constitute a distinct subgroup within the broader population of cryptocurrency holders. To this end, we restrict our sample to individuals who own or have owned cryptocurrencies. Memecoin

[6] The State of Crypto Survey is a cross-disciplinary academic initiative launched in 2021, dedicated to examining the societal implications of cryptocurrency as a cultural, social, technological, and economic phenomenon. The research team operates independently, with no ties to any crypto projects, thereby upholding the principles of transparency and academic neutrality. Emphasizing inclusivity, the survey is available in multiple languages, and all results are shared publicly—both as open-access research and in formats accessible to broader audiences. Every project within the initiative adheres to a strict non-deception policy, undergoes rigorous ethics committee review, and complies fully with GDPR regulations to guarantee secure data handling and participant confidentiality. For more information, visit: https://stateofcrypto.net.

ownership serves as our primary variable of interest, indicating whether a respondent holds (or has held) at least one memecoin. To characterize the profile of memecoin investors, we analyze data across three main categories of explanatory variables: (i) socioeconomic characteristics and risk preferences, (ii) investment behavior and financial experience, and (iii) attitudes toward cryptocurrencies and broader societal issues.

Table D16 presents a summary of the explanatory variables incorporated into the empirical analysis to capture the potential characteristics associated with memecoin investors.

Socio-economic and risk profile variables. These are variables that capture sociodemographic and socioeconomic indicators of cryptocurrency investors in our survey including the educational level, their age range, gender, their level of crypto literacy[7], financial literacy, their working status, if they work in the crypto sphere, and race.[8] This category also encompasses risk profile variables, including risk seeking.

Investing profile. These are variables that capture the behavior of cryptocurrency owners as investors, such as when they were first interested in crypto, percentage of their portfolio invested in crypto, percentage of their overall crypto portfolio in Bitcoin, number of cryptocurrencies currently owned, whether they are a bond or stock investor, if they have ever used options or other crypto derivatives, if they have ever used leverage to fund cryptocurrency investments, and if they have ever used crypto yield farming.

Attitudes. These variables capture the attitudes of crypto users towards taxes (if they think that all cryptocurrency gains should be taxed), scams (if they think that cryptocurrencies facilitate money laundering and scams more than cash or other means of payment), their government (how much of the time they think they can trust in their government), staking (if they are worried that governance protocols based on staking coins will allow the richest individuals and groups to buy the votes (stakes) they need to implement the governance they want) and mass-control (if they are worried about cryptocurrencies being used as an instrument of mass control). Finally, we also include in this category variables that capture the general opinion of crypto users about the purpose of cryptocurrencies.

Table 6.2 provides summary statistics for the variables included in the benchmark specification.[9] In our main specification, 47% of crypto users own or have owned at least one memecoin. From our respondents, the average age is 34.55 years old, 84% are male, and around 12% identify themselves as crypto traders.

6.2.3 Methodology

To test the hypothesis that memecoin investors exhibit distinct characteristics compared to the broader population of cryptocurrency investors, we employ a logistic regression model.

$$Pr(Y_i = 1|X_i) = \Lambda(\beta_0 + \beta_1 E_i + \beta_2 I_i + \beta_3 A_i + \epsilon_i) \tag{6.1}$$

[7]We construct three different crypto indexes: i) "Know crypto general" that assess the knowledge about crypto in five questions (maximum achievable score = 5), ii) "Know crypto Meme" that assess the general knowledge about NFTs in one question, and iii) "Know crypto" that combines both previous variables (maximum score achievable of 6).

[8]Race was divided in twelve categories (see Appendix B for additional details).

[9]Table 6.3, presents the main correlations among the variables employed.

Table 6.1: Potential features of Memecoin investors

Socio-economic and risk profile		
Variable	Code	Definition
Place of residence	*Country*	Country in which the respondent currently resides.
Race	*Race*	Race to which one classifies herself to.
Gender	*Gender*	Gender one identifies herself to (being male = 1, rest (women, non-binary, other) = 0.
Age	*Age*	Age range at the time of answering the survey.
Education	*Edu*	Maximum educational level achieved.
Work status	*Is working*	Working status of the respondent.
Working in crypto	*Works in crypto*	If the respondent works in the crypto sector or not
Memecoins literacy	*Know Meme*	Index that measures the level of Memecoins literacy.
NFT knowledge	*Know crypto NFT*	Index that measures the level of NFT literacy.
Cryptocurrency literacy	*Know crypto no Meme*	Index that measures the level of crypto literacy.
Financial literacy	*Know fin.*	Index that measures the level of financial literacy.
Trader	*isTrader*	Self-identified as a person who works in crypto and as a trader (in crypto).
Risk seeking	*Risk seeking*	Whether a participant chose to open more than 50 boxes in the Bomb risk elicitation task [100].
Investing profile		
Variable	Code	Definition
Interest in cryptocurrencies	*When interested*	When the respondent got interested in crypto for the first time.
Wealth invested in cryptocurrencies	*Wealth inv.*	Percentage of the overall investment portfolio in crypto.
Bitcoin share	*invested In BTC Z*	Percentage of overall crypto portfolio in Bitcoin.
Investment in derivatives	*Derivatives*	Whether the person invests in derivatives or not.
Farming	*Farming*	Whether the person engages in farming activities or not.
Leverage	*Leverage*	Opinion on whether the respondent has ever used leverage (borrowed money) to fund your cryptocurrency investments.
Own NFT	*hasNFT*	NFT Owner
Number of different cryptocurrencies held (No Memecoins)	*tot Cryptos nomeme*	Number of different cryptocurrencies held by an investor.
Attitudes towards crypto and society		
Variable	Code	Definition
Need to belong	*Need to belong*	Self-reported strength of the respondent's "need to belong" (desire for social acceptance and inclusion).
FOMO	*FOMO*	Self-reported extent to which the respondent experiences FoMO (fear of missing out)
Identity checks	*KYC*	Opinion on customer identification checks (e.g., Know-Your-Customer) when using crypto services
Scams	*Scams*	Opinion on whether cryptocurrencies facilitate money laundering and scams more than cash or other means of payment.
Taxes	*Taxes*	Opinion on whether all cryptocurrency gains should be taxed.
Trust in government	*Trust in gov.*	Opinion on whether how much of the time the respondent thinks she can trust the government in the country where she lives to do what is right.
Trust in regulation	*Trust in reg.*	Opinion on Better regulation of the crypto sphere would increase my trust in using crypto assets
Mass control	*Mass control*	Opinion on whether individuals are worried about cryptocurrencies been used as an instrument of mass control.

Note: additional details are provided in Appendix B.

where, $\Lambda(\cdot)$ denotes the standard logistic cumulative distribution function. $Y_{i,t}$ is a binary variable that takes the value 1 if individual i owns or has owned at least one memecoin, and 0 otherwise. E_i is a vector of socio-economic characteristics at the individual level, I_i represents variables related to the individual's investment profile, and $A_{i,t}$ captures individual attitudes and behavioral traits. Standard errors are clustered by

Table 6.2: Summary statistics

	Mean	SD	Min	Max	Median	N
Age	34.44	11.75	16.00	80.00	31.50	1013
Gender	0.84	0.37	0.00	1.00	1.00	1001
Edu. Level	2.88	1.27	0.00	5.00	3.00	1011
Crypto Trader	0.12	0.33	0.00	1.00	0.00	1013
Know Meme	0.33	0.47	0.00	1.00	0.00	1013
Know Crypto (ex Meme)	2.03	1.19	0.00	4.00	2.00	1013
Know Finance	1.55	1.00	0.00	3.00	2.00	1011
When Interested	124.76	44.74	1.00	192.00	133.00	1013
Risk Seeking	0.28	0.45	0.00	1.00	0.00	891
Wealth Crypto	36.89	37.30	0.00	100.00	24.00	1013
#Cryptos no meme	2.85	3.22	0.00	25.00	2.00	1013
Farming	0.26	0.44	0.00	1.00	0.00	1012
Derivatives	0.22	0.41	0.00	1.00	0.00	1007
Leverage	0.20	0.40	0.00	1.00	0.00	1013
Scams	1.84	1.46	0.00	4.00	2.00	1007
Taxes	1.50	1.47	0.00	4.00	1.00	1013
Trust in Gov.	0.92	0.82	0.00	3.00	1.00	1004
FOMO	1.36	1.10	0.00	4.00	1.00	1009
Need to Belong	1.22	1.17	0.00	4.00	1.00	1013
Own NFT	0.26	0.44	0.00	1.00	0.00	1013
Own Meme	0.47	0.50	0.00	1.00	0.00	1013
Regulations	2.32	1.41	0.00	4.00	3.00	1009
KYC	2.24	1.36	0.00	4.00	2.00	1004

continent."

The logistic regression model is estimated using maximum likelihood estimation. In the logit specification, the error term is assumed to follow a standard logistic distribution, with mean 0 and variance $\pi^2/3$. In all cases, average marginal effects are reported to facilitate interpretation of the results.

Table 6.3: Correlation table (N=1,013)

	belong	kyc	regu	gov	fomo	taxes	scams	leverage	NFT	coins	interest	wealth	trader	risk	edu	k fi	k NM	kM	age
Gender	-.02	-.09	-.07	.04	-.01	-.04	.00	.12*	.05	.10	-.17***	.16***	-.04	-.08	-.04	.10	.09	.05	.05
age	-.21***	.01	.02	-.05	-.07	.11*	-.04	-.02	-.04	.12*	-.26***	.02	-.18***	-.08	.25***	.16***	.07	-.01	
k M	.01	-.06	-.05	-.06	.11	-.11	-.10	.21***	.30***	.29***	-.02	.27***	.15***	.01	-.03	.14**	.40***		
k NM	-.08	-.05	.00	.00	-.01	-.03	-.11	.07	.21***	.29***	-.13**	.19***	.10	-.05	.14***	.30***			
k fi	-.17***	-.03	.01	.05	-.07	.12*	.10	-.02	-.02	.04	-.16***	.00	-.09	-.07	.19***				
edu	-.14**	.08	.04	.05	-.02	.16***	.08	-.10	-.03	.03	-.18***	-.06	-.11	.02					
risk	.10	.03	.03	.00	.09	-.03	.04	-.05	.01	-.03	.09	.00	.04						
trader	.19***	.07	.01	.02	.10	-.13**	-.05	.21***	.21***	.14**	.03	.14**							
wealth	-.02	-.23***	-.17***	-.14***	.07	-.27***	-.21***	.27***	.30***	.40***	-.10								
interest	.19***	.08	.13**	-.04	.08	-.04	.05	-.06	-.08	-.14***									
coins	-.09	-.11	-.06	-.08	.04	-.14***	-.21***	.23***	.28***										
NFT	.05	-.08	-.05	-.01	.14**	-.05	-.09	.24***											
leverage	.06	-.11	-.05	-.07	.12*	-.11	-.07												
scams	-.02	.16***	.17***	.18***	.06	.27***													
taxes	.01	.26***	.27***	.30***	-.02														
fomo	.37***	.02	.10	.03															
gov	.06	.17***	.14**																
regu	.17***	.36***																	
kyc	.14***																		

Notes: regu is trust in regulations, gov is trust in government, coins is the total number of non-memecoins held, risk is 1 if participant is risk seeking, wealth, is the proportion of wealth invested in crypto, NFT is if participant has held at least one NFT in the last year, k fi is financial knowledge, k NM is knowledge of crypto, excluded memecoins, k M is knowledge of memecoins. The symbols ***, **, and * indicate statistical significance at the 0.1%, 1%, 5% level, respectively.

6.3 Main results

6.3.1 Baseline regressions

Our main results (see Table 4) suggest that memecoin investors are better characterized by market engagement and broader participation in speculative crypto sub-sectors than by standard socio-demographics.

Gender is positively associated with memecoin ownership in specifications (1) and (2): being male increases the predicted probability by about 8–9 percentage points. The estimate attenuates to roughly 5 percentage points and is no longer statistically significant in (3). This pattern is consistent with the broader gender gap documented in crypto participation [26, 160], and with mechanisms typically invoked in the literature, higher risk tolerance, overconfidence, and speculative trading preferences [47, 161], as well as cultural alignment with irony-driven online communities [65, 318]. At the same time, the lack of robustness across specifications suggests caution in interpreting gender as an independent, stable predictor in this setting.

Age, in contrast, shows no measurable association. In fact, the reader may think that the absence of any age effect is unexpected. Popular narratives often associate memecoin activity with younger investors, citing their higher digital exposure, greater familiarity with social platforms, and higher risk tolerance [75, 209, 227, 275]. However, once the model accounts for behavioral factors—such as trading frequency, leverage use, and familiarity with specific asset types—age no longer explains variation in investment behavior. Our result suggests that age may operate as a proxy for experience, portfolio composition, or timing of market entry. When those dimensions are included directly in the model, the independent contribution of age disappears. In this context, it is not age itself that matters, but the traits and behaviors more frequently, but not exclusively, observed among younger individuals.

Behavioral and portfolio variables provide the clearest separation. Memecoin-specific knowledge is strongly and stably related to ownership: respondents answering the memecoin question correctly are about 13–15 percentage points more likely to hold memecoins across specifications. Active crypto trading is also consistently associated with ownership (approximately 13–19 percentage points), indicating that memecoin participation is tightly linked to hands-on market involvement. Portfolio breadth matters: each additional non-memecoin crypto asset held increases the probability by roughly 4 percentage points, and NFT ownership adds about 16–20 percentage points. This is in line with evidence that broader crypto exposure correlates with engagement in adjacent speculative markets such as NFTs [45].

Intensity-related measures reinforce this profile. A larger wealth allocation to crypto and individuals who developed an interest in cryptocurrencies more recently are more likely to invest in memecoins.are both positively associated with memecoin ownership, albeit with small marginal effects per unit. Leverage use is robustly positive (about 11–13 percentage points), consistent with leverage being a channel for higher volatility exposure and short-term speculation [20, 282]. By contrast, the task-based risk preference measure does not display a robust association once controls are included.

Institutional and psychological correlates are present but weaker. Greater tax salience is negatively associated with memecoin ownership (around 2 percentage points), consistent with frameworks linking compliance sensitivity to institutional constraints [184,

Table 6.4: Baseline regressions ATE

	(1)	(2)	(3)	(4)
Gender	0.088 (0.042)*	0.084 (0.042)*	0.052 (0.038)	
Age	0.007 (0.005)			
Edu	−0.013 (0.012)	−0.008 (0.011)	−0.009 (0.011)	
Know Meme	0.147 (0.036)***	0.143 (0.034)***	0.127 (0.031)***	0.137 (0.031)***
Know crypto no Meme	−0.003 (0.014)			
Know fin.	−0.022 (0.015)	−0.020 (0.015)		
Crypto trader	0.185 (0.059)**	0.167 (0.058)**	0.132 (0.051)**	0.138 (0.049)**
Risk Seeking	−0.005 (0.031)	−0.004 (0.031)		
When interested	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)**	0.001 (0.000)***
Wealth inv.	0.001 (0.000)*	0.001 (0.000)*	0.001 (0.000)***	0.001 (0.000)***
Leverage	0.117 (0.041)**	0.117 (0.040)**	0.114 (0.038)**	0.131 (0.038)***
Has NFT	0.196 (0.039)***	0.192 (0.038)***	0.171 (0.036)***	0.165 (0.036)***
Num non-memecoins held	0.039 (0.005)***	0.040 (0.005)***	0.042 (0.005)***	0.040 (0.005)***
Need to belong	−0.037 (0.014)*	−0.038 (0.013)**	−0.034 (0.012)**	−0.033 (0.012)**
FOMO	0.002 (0.014)			
KYC	−0.002 (0.012)			
Trust regulations	0.016 (0.011)	0.016 (0.010)	0.014 (0.010)	
Scams	0.003 (0.010)	0.001 (0.010)	0.000 (0.009)	
Taxes	−0.024 (0.011)*	−0.023 (0.011)*	−0.021 (0.010)*	−0.019 (0.009)*
Trust in Gov.	−0.017 (0.019)	−0.022 (0.019)	−0.013 (0.018)	
Num. Obs.	856	867	984	1013

Notes: The table reports the results of the estimation of the logit regression. The dependent variable is the ownership of memecoins. Average marginal effects presented. Standard errors clustered by continent are reported in parentheses. The symbols ***, **, *, and + indicate statistical significance at the 0.1%, 1%, 5% and 10% level, respectively. The models include also a constant and continent fixed effect, not reported for brevity.

273]. Trust in regulation is weakly positive in some specifications, suggesting that participation in speculative segments does not necessarily coincide with outright rejection of oversight [270, 298]. Finally, need to belong is consistently negative (roughly 3–4 percentage points per unit), aligning with the interpretation that memecoin participation may be facilitated by a lower reliance on conventional social affiliation motives within a culture often characterized by humor, satire, and informal group dynamics [196, 201].

6.3.2 Robustness checks

We ran an extensive battery of robustness checks (documented in Appendix D) to assess whether the baseline associations are driven by data-quality concerns, duplicated participation, or specific sample compositions [43]. Across all checks, the results are highly stable in sign and magnitude for the main correlates of memecoin ownership. In particular, memecoin-specific knowledge remains a strong and precisely estimated predictor, and the same holds for market engagement and intensity proxies: active crypto trading, leverage use, and broader involvement in speculative crypto segments (NFT ownership and holding more non-memecoin assets). The psychological correlate is likewise robust: "need to belong" remains consistently negative, and "tax salience" retains a small but stable negative association.

The stability of these findings is visible under each restriction. Excluding respondents whose IP geolocation does not match the self-reported country of residence yields near-identical estimates for the core variables. Removing observations that share the same IP, email, or crypto address (separately) does not materially affect coefficient magnitudes, suggesting that potential multiple participation is not driving the results. Sim-

ilarly, applying strict response-time filters (and additional response-level timing exclusions) leaves the main effects essentially unchanged; if anything, the association between memecoin knowledge and ownership becomes slightly stronger under these stricter screens.

We also tested robustness to alternative definitions of problematic responses. Excluding likely duplicates identified through similarity in demographic answers (using increasingly conservative thresholds) produces virtually identical estimates. Dropping implausible answer patterns (e.g., unusually inconsistent education/age combinations or extreme portfolio reports) does not alter the substantive conclusions. Results also remain stable when excluding respondents whose memecoin ownership is internally inconsistent across survey items, and when restricting the sample to respondents who consented to broader data sharing. Finally, multilevel specifications closely match the baseline models, indicating limited residual heterogeneity once individual covariates are accounted for.

Overall, the robustness checks support the interpretation that memecoin ownership is primarily associated with higher engagement in crypto markets and participation in adjacent speculative sub-sectors, while demographic correlates are comparatively weaker and more sensitive to specification and sample restrictions.

6.3.3 Further analysis

Age and the myth of the Memecoin generation?

The baseline specifications treat age as a linear predictor and yield no robust association with memecoin ownership. Given the common narrative linking memecoins to younger cohorts [75, 209, 227, 275], we further explore whether age effects are (i) non-linear or (ii) concentrated in specific generational cohorts.

Table 5 (columns 1–4) adds a quadratic term. The linear age component is weakly positive in some specifications, while the squared term is negative, but the overall evidence is not sufficiently strong to support a clear, stable inverted-U relationship. In other words, once the model conditions on engagement and portfolio characteristics, age does not emerge as a robust independent determinant of memecoin ownership.

To assess whether any age-related heterogeneity reflects discrete cohort differences rather than smooth variation, we recode age into three mutually exclusive groups, Gen Z (reference), Millennials, and Older respondents, motivated by the idea that cohorts differ in market entry conditions, institutional attitudes, and digital-cultural exposure [75, 209, 275]. Columns 5–8 show that Millennials are consistently more likely to hold memecoins than Gen Z, with average marginal effects of roughly 3–6 percentage points (statistically significant in most specifications). By contrast, the Older group does not differ significantly from Gen Z. These results are more consistent with a mild concentration among Millennials than with the notion of a uniquely "younger" memecoin generation; liquidity constraints and limited market experience among the youngest cohort may still play a role [77].

Table 6.5: Variation regressions ATE: Non-linear Age

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender	0.089*	0.087*	0.053*		0.089*	0.087*	0.053*	
	(0.043)	(0.044)	(0.025)		(0.039)	(0.040)	(0.024)	
Age	0.021+	0.023+	0.018	0.011				
		(0.012)	(0.012)	(0.012)	(0.009)			
Age squared	−0.001	−0.001	−0.001	−0.001				
					(0.049)	(0.048)	(0.038)	(0.041)
Millennials					0.058+	0.060+	0.038*	0.034*
					(0.034)	(0.035)	(0.019)	(0.017)
Other Gen.					0.028	0.029	0.001	0.003
		(0.001)	(0.001)	(0.001)	(0.001)			
Edu	−0.015	−0.013	−0.012*		−0.013	−0.012	−0.010+	
		(0.009)	(0.009)	(0.006)		(0.009)	(0.009)	(0.006)
Know Meme	0.147***	0.144***	0.126***	0.136***	0.146***	0.143***	0.127***	0.137***
	(0.027)	(0.030)	(0.023)	(0.022)	(0.028)	(0.031)	(0.024)	(0.022)
Know crypto no Meme	−0.003				−0.002			
		(0.017)			(0.017)			
Know fin.		−0.022	−0.022			−0.021	−0.021	
		(0.029)	(0.027)			(0.029)	(0.027)	
Crypto trader	0.186*	0.174**	0.135***	0.141***	0.187*	0.175**	0.134***	0.140***
	(0.078)	(0.065)	(0.030)	(0.025)	(0.077)	(0.063)	(0.029)	(0.024)
Risk Seeking	−0.004	−0.001			−0.005	−0.002		
	(0.016)	(0.015)			(0.015)	(0.014)		
When interested	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Wealth inv.	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	0.116***	0.116***	0.113***	0.131***	0.115***	0.115***	0.113***	0.131***
	(0.024)	(0.031)	(0.033)	(0.032)	(0.024)	(0.032)	(0.033)	(0.031)
Has NFT	0.195***	0.190***	0.171***	0.165***	0.192***	0.187***	0.168***	0.162***
	(0.023)	(0.021)	(0.031)	(0.029)	(0.025)	(0.023)	(0.034)	(0.031)
Num non-memecoins held	0.039***	0.039***	0.041***	0.040***	0.040***	0.040***	0.042***	0.040***
	(0.007)	(0.007)	(0.006)	(0.005)	(0.007)	(0.007)	(0.006)	(0.005)
Need to belong	−0.037***	−0.038***	−0.034***	−0.032***	−0.037***	−0.037***	−0.034**	−0.032**
	(0.007)	(0.009)	(0.010)	(0.010)	(0.008)	(0.010)	(0.010)	(0.010)
FOMO	0.003				0.003			
	(0.010)				(0.010)			
KYC	−0.002				−0.002			
	(0.014)				(0.014)			
Trust regulations	0.016	0.015+	0.013+		0.016	0.016+	0.014+	
	(0.011)	(0.009)	(0.008)		(0.011)	(0.009)	(0.008)	
Scams	0.004	0.003	0.001		0.003	0.003	0.001	
	(0.016)	(0.015)	(0.012)		(0.017)	(0.016)	(0.012)	
Taxes	−0.024*	−0.025*	−0.022*	−0.020*	−0.025*	−0.026*	−0.023*	−0.021*
	(0.010)	(0.011)	(0.009)	(0.009)	(0.010)	(0.011)	(0.010)	(0.009)
Trust in Gov.	−0.016	−0.019	−0.012		−0.017	−0.020	−0.013	
	(0.019)	(0.020)	(0.013)		(0.019)	(0.020)	(0.014)	
Num.Obs.	856	867	984	1013	856	867	984	1013

Notes: The table reports the results of the estimation of the logit regression. The dependent variable is the ownership of memecoins. Average marginal effects presented. Standard errors clustered by continent are reported in parentheses. The symbols ***, **, *, and + indicate statistical significance at the 0.1%, 1%, 5% and 10% level, respectively. The models include also a constant and a continent fixed effect, not reported for brevity.

Risk engagement beyond leverage?

To test whether the baseline leverage effect captures a broader propensity to engage with high-risk crypto-finance activities, we replace leverage with alternative binary indicators: prior experience with derivatives and exposure to yield farming (Table 6). Both proxies reflect engagement with more complex or speculative activities, but they involve different incentives and time horizons.

The derivatives indicator is positive across specifications and becomes statistically significant in the more parsimonious models, with average marginal effects of about 7–8 percentage points (columns 3–4). However, the estimate is smaller and imprecise in the richer specifications (columns 1–2), suggesting that derivatives experience overlaps with other engagement measures (e.g., active crypto trading and portfolio breadth) rather than constituting an independent driver throughout.

By contrast, the yield-farming indicator is close to zero and not statistically significant in any specification (columns 5–8). This pattern is consistent with farming being more closely related to capital lock-up, incentive optimization, and in some cases relatively passive strategies [39], which may not align with the fast-moving, narrative- and hype-driven dynamics typical of memecoin participation.

6.4 Conclusion

In this study we shed light on the behavioral and demographic correlates of memecoin ownership. Our results indicate that memecoin holders constitute a distinct subgroup within the broader cryptocurrency ecosystem, primarily differentiated by market engagement rather than by standard demographics. Across specifications and robustness checks, memecoin ownership is consistently associated with more intensive participation in crypto markets: active trading, leverage use, broader portfolios, and NFT ownership. Memecoin-specific knowledge is also a strong and stable correlate, suggesting that participation in this niche reflects targeted exposure and familiarity rather than general financial sophistication. Overall, the evidence supports the view that memecoin investment is closely tied to high-intensity crypto participation and involvement in adjacent speculative sub-sectors.

The demographic evidence is more nuanced. The association with gender is positive but not fully robust across specifications, indicating that any gender gap is smaller and more model-dependent than often assumed. Age does not emerge as a robust independent predictor once behavioral and portfolio controls are included, and a quadratic specification provides limited support for a systematic non-linear relationship. However, cohort-based estimates suggest a mild concentration among Millennials relative to Gen Z, consistent with the possibility that memecoin participation is shaped by cohort-specific market entry conditions and constraints, rather than by age per se.

Institutional and psychological correlates provide additional insight, although with smaller magnitudes. Memecoin ownership is negatively associated with tax salience and with the need-to-belong measure, while trust in regulation shows, at most, weakly positive associations in some specifications. Taken together, these results are consistent with a profile characterized by a higher tolerance for speculative environments and a comparatively lower reliance on conventional social affiliation motives. Importantly,

Table 6.6: Variation regressions ATE: Derivatives and Farming

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender	0.090*	0.088*	0.055*		0.095*	0.093*	0.061*	
	(0.044)	(0.044)	(0.025)		(0.044)	(0.045)	(0.026)	
Age	0.008				0.007			
	(0.006)				(0.007)			
Edu	−0.015	−0.010	−0.010		−0.016	−0.011	−0.012	
	(0.010)	(0.010)	(0.008)		(0.010)	(0.009)	(0.007)	
Know Meme	0.159***	0.157***	0.136***	0.149***	0.160***	0.154***	0.135***	0.148***
	(0.031)	(0.032)	(0.024)	(0.022)	(0.030)	(0.035)	(0.028)	(0.026)
Know crypto no Meme	−0.003				−0.006			
	(0.019)				(0.021)			
Know fin.	−0.024	−0.023			−0.023	−0.022		
	(0.027)	(0.026)			(0.025)	(0.023)		
Crypto trader	0.170*	0.159**	0.118***	0.122***	0.202**	0.186***	0.145***	0.151***
	(0.074)	(0.060)	(0.025)	(0.026)	(0.062)	(0.049)	(0.015)	(0.013)
Risk Seeking	−0.006	−0.006			−0.010	−0.009		
	(0.013)	(0.013)			(0.014)	(0.014)		
When interested	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Wealth inv.	0.001***	0.001***	0.002***	0.002***	0.001***	0.001***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Farming					0.000	−0.002	0.019	0.015
					(0.068)	(0.065)	(0.067)	(0.057)
Derivatives	0.045	0.035	0.070*	0.075*				
	(0.041)	(0.041)	(0.031)	(0.030)				
Has NFT	0.200***	0.198***	0.176***	0.172***	0.204***	0.201***	0.177***	0.175***
	(0.024)	(0.022)	(0.030)	(0.027)	(0.027)	(0.029)	(0.038)	(0.034)
Num non-memecoins held	0.041***	0.042***	0.042***	0.041***	0.042***	0.043***	0.044***	0.043***
	(0.008)	(0.007)	(0.005)	(0.004)	(0.007)	(0.006)	(0.005)	(0.004)
Need to belong	−0.039***	−0.040***	−0.035**	−0.033***	−0.036***	−0.037***	−0.032***	−0.031***
	(0.007)	(0.010)	(0.011)	(0.010)	(0.006)	(0.007)	(0.009)	(0.008)
FOMO	0.008				0.006			
	(0.010)				(0.010)			
KYC	−0.006				−0.007			
	(0.015)				(0.014)			
Trust regulations	0.017	0.016	0.015+		0.017	0.015	0.014	
	(0.013)	(0.010)	(0.009)		(0.012)	(0.010)	(0.009)	
Scams	0.005	0.003	0.001		0.004	0.002	0.001	
	(0.016)	(0.016)	(0.012)		(0.017)	(0.017)	(0.013)	
Taxes	−0.024**	−0.024*	−0.022**	−0.020*	−0.023*	−0.023*	−0.021*	−0.020*
	(0.009)	(0.010)	(0.008)	(0.008)	(0.009)	(0.010)	(0.008)	(0.008)
Trust in Gov.	−0.017	−0.022	−0.013		−0.020	−0.025	−0.017+	
	(0.016)	(0.014)	(0.011)		(0.017)	(0.016)	(0.010)	
Num.Obs.	852	864	981	1009	856	868	985	1014

Notes: The table reports the results of the estimation of the logit regression. The dependent variable is the ownership of memecoins. Average marginal effects presented. Standard errors clustered by continent are reported in parentheses. The symbols ***, **, *, and + indicate statistical significance at the 0.1%, 1%, 5% and 10% level, respectively. The models include also a constant and a continent fixed effect, not reported for brevity.

these patterns remain stable under stringent data-quality restrictions and alternative sample definitions, increasing confidence that they are not artifacts of response noise or duplication.

From a policy perspective, the findings suggest that interventions targeting memecoin-related risks should focus less on demographic segmentation and more on behavioral exposure. Users who actively trade, employ leverage, and participate in NFT markets appear systematically more likely to engage with memecoins and may therefore face disproportionate downside risk. Standard financial education and generic disclosures may have limited effectiveness in this context; instead, platform-level measures—such as clearer risk labeling on extreme volatility and liquidity, and friction or warnings around leveraged access—may be more directly aligned with the relevant behaviors. Given the cross-border nature of memecoin markets and their susceptibility to manipulation and regulatory arbitrage, greater international coordination on listing standards and transparency requirements would further reduce retail exposure to speculative shocks.

Chapter 7

Public Perceptions of Cryptomarket Regulation: Investor Profiles and Attitudes

7.1 Introduction

Over the past decade, cryptocurrencies have evolved from a technological innovation into a burgeoning, yet still nascent financial ecosystem. As cryptoassets and associated services continue to grow both in scale and adoption, addressing their miscellaneous risks has become a pressing issue for policymakers and regulators [234]. This urgency has intensified following a series of infamous events (e. g., FTX bankruptcy [97, 129] and Bybit hack [190]), the use of funds in illicit activities [78, 81, 112], and growing concerns over investor protection, financial stability, and market integrity [34, 37, 40, 176].

In response to the recognized regulatory gaps, jurisdictions worldwide are increasingly engaging in the enforcement of legal frameworks aimed at bringing greater transparency, accountability, and oversight over the cryptomarket. Notable initiatives include the European Union's Markets in Crypto-Assets (MiCA) regulation [117], the U.S. Securities and Exchange Commission's enforcement actions [264], and the Financial Action Task Force's (FATF) guidelines on anti-money laundering. Collectively, these measures reflect a growing consensus on the need to govern crypto operations.

Despite this regulatory momentum, there is yet a limited understanding of how cryptoasset investors relate to such interventions. While industry stakeholders and advocacy groups are often invited to private consultations on the scope and design of crypto regulation, empirical evidence on broader public sentiment is remarkably scarce. Existing user studies have largely concentrated on adoption drivers, risk perceptions, security practices, and investment behavior of cryptoasset investors [e. g., 16, 44, 60, 211]. In contrast, the general public's attitudes toward cryptomarket regulation are underexplored, limiting our understanding of the social legitimacy and acceptance of the emerging legal frameworks.

This chapter addresses this gap by empirically examining individuals' opinions on regulatory actions in the cryptomarket. Using unique survey data from over 1,400 participants from across the globe (68% of those self-reported to be cryptoasset users), we examine which subpopulations of this community are more likely to support or op-

pose regulatory interventions, and how these two polar groups differ in terms of their *socio-demographic characteristics*, *investment profiles*, and broader *belief systems*. Understanding these nuances is not only academically relevant, but also essential for designing effective and legitimate legal frameworks that resonate with the values and expectations of the communities they aim to protect.

Our contribution is twofold. Drawing on a large global survey sample, we provide novel empirical evidence on individuals' attitudes toward regulatory oversight in cryptomarkets. Second, we advance behavioral finance research and inform policy design by documenting systematic heterogeneity in the public acceptance of crypto regulation.

The chapter is structured as follows: Section 7.2 briefly reviews related literature; Section 7.3 outlines the methodology and survey data; Section 7.4 presents the empirical results; Section 7.5 concludes.

7.2 Related work

The state-of-the-art literature on crypto regulation has evolved from early legal scholarship discussing regulatory challenges [e. g., 131, 162, 197, 291] to emerging studies that critically evaluate the introduced compliance practices and their effectiveness [239]. Of particular relevance to our problem domain are event studies analyzing cryptocurrency market reactions to regulatory announcements. Specifically, [37] find that the introduction of crypto-specific legal frameworks tends to elicit positive market responses, while [93] show that anticipated regulations are often perceived as *bad news*, resulting in negative price returns. While informative, these studies provide inconsistent results and infer public sentiment indirectly from price dynamics, thereby overlooking more nuanced differences in investors' characteristics and beliefs. We extend this line of research using an empirical approach grounded in an online survey and a large sample of participants. Complementarily, technical work on explainable on-chain enforcement informs the feasibility, rather than the sentiment, of regulation [243].

Other related works include behavioral studies of cryptoasset users, their motivations, mental models, security behaviors, and risk perceptions [e. g., 15, 16, 193, 211]. To the best of our knowledge, no prior research has explored participants' views on crypto regulation or specific legal frameworks.

7.3 Methodology

7.3.1 Data Collection

The data utilized in this study are drawn from the second edition of the State of Crypto Survey[1]. The 2024/25 edition was fielded between December 2024 and February 2025 on several social media platforms, with targeted recruitment on *X* (formerly Twitter) and *Reddit*, two major channels for crypto discourse.

[1]The State of Crypto Survey is an independent, cross-disciplinary initiative launched in 2021 to study the cultural, social, technological, and economic ramifications of crypto-assets. The research team has no ties to any crypto project, adopts a strict non-deception policy, undergoes a regular ethics review, and releases anonymised data and documentation under an open-access license (https://stateofcrypto.net).

The survey yielded 2,219 entries, of which 1,414 were fully completed. Participation was voluntary and not incentivized. To maximise inclusiveness, the survey was translated into 30 languages and attracted respondents from more than 150 countries. The average completion time was about 10 minutes. Further details on the language and country distributions are reported in 11.

The instrument collected detailed information on socio-demographic characteristics (e. g., gender, age, education, race/ethnicity, employment status), financial and crypto literacy, risk-and-portfolio profiles as well as on psychological, trust, and broader attitudes toward cryptocurrencies.

The study was implemented in `nodeGame` [42], preregistered on AsPredicted[2], approved by the Institutional Review Board of the University of Mannheim (EK Mannheim 40/2023), and conducted in full compliance with the GDPR requirements according to the University's of Mannheim Data Protection Officer.

7.3.2 Survey Data

For the empirical analysis, we organise the survey items into three conceptual domains: (i) socio-economic background and risk propensity, (ii) investing behaviour, and (iii) attitudes toward crypto, regulation, and society. These domains are organized to reflect a progression from relatively stable individual characteristics to reported behavior and, finally, to evaluative beliefs, allowing for a structured assessment of how background traits and actions shape regulatory attitudes. 11 presents a summary of the explanatory variables included into the analysis.

Socio-economic and risk-profile variables This domain includes standard socio-demographic indicators (i. e., highest educational attainment, country of residence, self-identified race/ethnicity, age bracket, gender identity, and employment status) as well as an indicator for whether the respondent works in the crypto sector. Knowledge-based variables are captured by indices of general cryptocurrency literacy, memecoin-specific literacy, a composite crypto literacy score, and a standard financial literacy index.

Investing profile To characterise respondents' behaviour, we consider the share of total wealth allocated to cryptoassets, the date at which respondents first became interested in crypto and the number of cryptocurrencies held. We also include indicators for ownership of NFTs and memecoins.

Personal beliefs Policy and societal orientations are measured through items on the perceived role of crypto in facilitating scams and money laundering, concerns that crypto could become an instrument of mass control, trust in the national government, confidence that tighter regulation would increase personal trust in crypto assets, support for mandatory customer-identification checks (KYC), and views on whether all crypto gains should be taxed. All Likert-scale items are coded so that higher values correspond to stronger agreement with the stated proposition.

[2]Preregistration ID: https://aspredicted.org/wmp6-ynk6.pdf.

7.3.3 Approach

Our analysis is based on a series of logistic regression models assessing how various explanatory variables relate to individuals' perceptions of cryptomarket regulation. We operationalize *regulation* through three items: *"Better regulation of the crypto sphere would increase my trust in using crypto assets." (Trust in reg.)*, *"I think all cryptocurrency gains should be taxed." (Taxes)*, and *"How do you feel about customer identification checks (e. g., Know-Your-Customer) when using crypto services?" (KYC checks).* Together, these items capture analytically distinct areas of regulation: trust in regulation reflects perceptions of institutional legitimacy; taxation captures the integration of cryptoassets into existing fiscal regimes; and KYC checks operationalize compliance-oriented regulation aimed at consumer protection and anti–money laundering.

For the analysis, we dichotomize each ordinal variable of interest into a binary outcome by excluding neutral responses and coding the remaining responses as either favorable (1) or unfavorable (0). The resulting binary outcomes are modeled using multivariate logistic regressions, specified as:

$$p(Y_i = 1 \mid X_i) = \frac{\exp(\beta_0 + X_i'\beta)}{1 + \exp(\beta_0 + X_i'\beta)},$$

where Y_i is a binary outcome variable for respondent i, and X_i is a vector of individual-level covariates. Standard errors are clustered by continent.

7.4 Results

Table 7.1 reports results of the logistic regression models estimated by maximum likelihood. The final specification retains only those control variables that proved most relevant in tests of alternative models using the available survey items. For completeness, 11 contains results of the model specifications analyzing the effects of each category of the final explanatory variables. In addition, we ran a series of robustness checks to validate our main results (see 11). Descriptive statistics of the variables used in the final specification are reported in 11. *Note:* taxation of crypto gains (*Taxes*) was not preregistered but was later included as an additional dependent variable given its conceptual relevance to our research problem.

Consistent patterns emerge across the models. With respect to investment behavior, respondents holding a substantial share of their wealth in cryptoassets consistently exhibit skepticism toward regulation. This points to concerns about regulatory overreach, threats to privacy, and a perceived tension between regulation and the foundational principles of decentralization and financial autonomy underlying cryptocurrencies. As for the attitudinal variables, individuals who associate cryptoassets with money laundering and scams are more supportive of regulation. This aligns with the view that regulatory oversight can mitigate perceived risks of illicit activity, thereby enhancing social legitimacy and public trust.

There are also model-specific significant effects:

Trust in reg Older respondents, recent entrants, those with a strong need to belong, and those concerned about the potential use of cryptocurrencies as a tool for mass control tend to view regulation as trust-enhancing. While some of these effects are intuitive,

Table 7.1: Results of the logistic regression models

	Trust in reg.		KYC checks		Taxes	
Socio-demographic variables						
Male	−0.042	(0.044)	−0.026	(0.017)	−0.079	(0.042)
Age	0.015*	(0.006)	0.010	(0.008)	0.020***	(0.004)
Education	0.020	(0.017)	0.054***	(0.010)	0.034*	(0.016)
When interested	0.001*	(0.000)	0.001*	(0.000)	0.000	(0.000)
Works in crypto	−0.066*	(0.028)	−0.009	(0.068)	−0.043*	(0.020)
Knows crypto	0.019	(0.013)	0.000	(0.023)	−0.003	(0.012)
Knows finance	0.012	(0.015)	0.007	(0.015)	0.024	(0.016)
Investment profile						
Wealth invested	−0.002**	(0.001)	−0.002***	(0.001)	−0.002***	(0.000)
Number of cryptos held	0.000	(0.005)	0.000	(0.007)	0.003	(0.006)
Has NFT	−0.026**	(0.010)	−0.062	(0.066)	0.071	(0.039)
Has memecoins	0.010	(0.022)	−0.054*	(0.028)	−0.053	(0.032)
Attitudinal variables						
Scams	0.036*	(0.016)	0.034***	(0.009)	0.056***	(0.010)
Mass control	0.016*	(0.008)	−0.011	(0.011)	0.024*	(0.010)
Trust in government	0.077	(0.049)	0.088**	(0.033)	0.109***	(0.020)
Need to belong	0.065**	(0.024)	0.030*	(0.014)	0.026	(0.021)
FoMO	0.007	(0.030)	−0.016	(0.017)	−0.012	(0.020)
Num. obs.	839		753		854	
AIC	989.8		911.0		972.0	
BIC	1093.9		1012.7		1076.5	
Log likelihood	-472.908		-433.494		-464.017	
RMSE	0.44		0.44		0.43	

Note: Average treatment effects are reported. Continent-adjusted standard errors in parentheses. Significance levels: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

we draw on procedural justice theory [293] and prior research [105] to interpret the influence of the need to belong on perceptions of procedural fairness. In addition, regulation may be perceived not only as an instrument of control but as a safeguard against power concentration and a lack of accountability in the decentralized ecosystem.

Interestingly, crypto workers and NFT holders do not associate regulation with greater trust in cryptoassets. For these groups, trust appears to stem from personal experience, technical expertise, or ideological commitment, rendering external oversight burdensome or even threatening.

Customer identity verification KYC procedures are viewed more favorably by highly educated respondents, recent market entrants, and those with greater trust in government and a stronger need to belong. In contrast, memecoin holders have a negative attitude toward KYC checks, which may reflect hesitance to reveal their identities in largely speculative and short-term crypto operations.

Tax treatment of cryptoassets Older and more educated respondents, as well as those who express higher levels of institutional trust or who associate crypto-assets with fraud and mass control, tend to hold more favorable views toward the taxation of crypto-related income. For these groups, taxation is regarded not only as a legitimate

instrument of state governance but also as a necessary means of mitigating perceived risks in crypto markets. By contrast, individuals employed within the crypto sector tend to oppose taxation, which they may perceive as conflicting with the principle of financial autonomy. Their resistance may also reflect economic self-interest, as taxation directly affects income derived from crypto-related activities.

Summary Overall, our findings highlight substantial heterogeneity in public perceptions of crypto regulation. Two patterns, however, remain consistent: individuals who associate cryptocurrencies with fraud and money laundering are more supportive of regulation, whereas those who allocate a large share of their wealth to crypto-assets tend to oppose regulatory initiatives. These effects persist across the additional robustness checks (see 11).

7.5 Conclusion

Our results reveal a clear divide in how people perceive regulation of crypto markets. Those who view cryptocurrencies as risky or vulnerable to fraud are more comfortable with stronger oversight. In contrast, individuals who have invested more heavily in crypto tend to worry that legal rules could limit their financial freedom or slow innovation. These differences show that there is no single "public view" on crypto regulation. People's preferences depend largely on whether they see cryptoassets as a threat or an opportunity. Taken together, the findings suggest that policymakers will need to strike a careful balance. Regulations that build trust and protect consumers are likely to gain support, but they will be more effective if designed in a way that still allows space for innovation and responsible participation. In this sense, understanding people's expectations is essential for shaping regulatory frameworks that feel both fair and forward-looking.

Our study has certain limitations. As is common in empirical research, potential self-selection and response biases may restrict the generalizability of our results. In addition, the analysis focuses on three dimensions of regulation, leaving other relevant areas (e. g., mining) outside the scope. Future research could also attempt to cover region-specific samples or employ mixed-methods designs to better capture the heterogeneity of public sentiment.

Part III

Measure micro-level circulation and systematize derivatives in DeFi

Part III focuses on decentralized finance (DeFi) from a protocol-level perspective. The objective is twofold: first, to develop an analytics pipeline that captures how key DeFi tokens circulate on-chain; second, to provide a unified framework for describing and studying decentralized derivatives protocols in a reproducible and comparable way.

This part brings together two complementary contributions. The first introduces a micro-velocity methodology for Lido's liquid staking tokens (stETH and wstETH), including the share-denominated reconstruction required by rebasing assets, and provides evidence on token circulation patterns, turnover concentration, and a progressive shift toward wstETH consistent with composability and cross-protocol integration; the associated extraction pipeline and curated datasets are released to support reproducibility and reuse. The second contribution systematizes decentralized derivatives protocols (perpetuals, options-like instruments, and synthetics) through a unified representation of actors, flows, and design principles, and operationalizes this framework via a tuple-based formalism and a reproducible simulation environment to study protocol dynamics under controlled market conditions.

Chapter 8

Money in Motion: Micro-Velocity and Usage of Ethereum's Liquid Staking Tokens

8.1 Introduction

The transition to Proof-of-Stake (PoS) in Ethereum [70] has resulted in both expected and unexpected effects on the network. One notable outcome is the rise of third-party liquid staking providers [135, 269]. Staking in PoS protocols involves locking a specific amount of tokens (the stake) to participate in the consensus process. In Ethereum, stakers lock at least 32 ETH, enabling them to issue attestations in consensus epochs and participate in block proposal lotteries [70, 315]. This stake acts as collateral to ensure honest behaviour; any dishonest activity may result in a penalty known as slashing. Additionally, staking requires participants to accept an opportunity cost. During the lock-up period, the staked tokens become illiquid, meaning they cannot be used in transactions or for other purposes. Liquid staking services address this limitation [92].

Lido, the dominant provider of liquid staking services, allows users to convert ETH into *stETH*, a liquid token. Holders of *stETH* continue to earn staking rewards, minus a fee paid to the service provider. This fee compensates providers like Lido, the biggest LST protocol, for managing staking responsibilities, including tasks that, if mishandled, could result in penalties or fund losses [98]. In addition to *stETH*, Lido also provides *wstETH*, a wrapped version of the token designed to maintain a stable ratio relative to the underlying staked ETH. *wstETH* is particularly suited for integration with decentralized finance (DeFi) protocols, as it is ERC-20 compliant and reflects staking rewards through an increasing exchange rate rather than token balance growth.

Liquid staking's rapid adoption has raised concerns about possible unintended impacts on the PoS ecosystem. Critics argue that widespread use of liquid staking could lead to outcomes such as block cartelization, ambiguity in protocol governance, and increased coupling of capital risk with protocol risk [116]. Currently, PoS systems offer limited ways to mitigate the growth of liquid staking, with existing approaches focusing primarily on encouraging voluntary self-regulation by service providers [116].

Liquid staking introduces a novel concept in the realm of digital assets, one that lacks a direct real-world comparison. It effectively allows money to multiply within users'

digital wallets, a phenomenon largely absent from traditional financial systems[1]. This unique characteristic raises important questions about the nature and classification of liquid staking tokens (LSTs) such as *stETH* and its wrapped counterpart, *wstETH*.

LSTs can be viewed as debt instruments: the staked tokens resemble the principal of a bond, the staking yield corresponds to interest payments, and the eventual return of staked tokens parallels principal repayment. This perspective aligns with their role as inflation-resistant stores of value. However, LSTs also exhibit unique characteristics, such as their use in DeFi applications as collateral or for yield farming, positioning them closer to programmable money. Additionally, LSTs can be seen as derivative instruments, with their value tied to the underlying staked tokens. This interpretation underscores the complexity of LSTs and the challenges of fitting them into existing financial categories. The future trajectory of Ethereum's PoS ecosystem may depend on how these assets are perceived and used. If LSTs are primarily viewed as inflation-resistant stores of value, they may function like digital bonds. However, their liquidity and utility in transactions and DeFi applications could allow them to evolve into a new form of programmable money, which is secured against inflation from monetary expansion.

Liquid staking has also introduced new risks to the Ethereum ecosystem. For example, the adoption of liquid staking services has raised concerns about centralisation. As of August 2022, Lido controlled a 30.1 % market share of total staked ETH, highlighting the significant influence these services can have on the network. Yet today, as of May 2025, the centralisation is only slightly attenuated, as Lido still has a 27 % market share of total staked ETH.

Moreover, the introduction of LSTs has increased the overall leverage in the crypto market. By enabling staked assets to be used as collateral, LSTs create new opportunities for yield farming but also introduce risk sources. The *stETH* crisis in 2022 demonstrated how the perceived 1:1 peg between *stETH* and ETH can break under market stress, leading to liquidity issues for overleveraged participants [269]. As the Ethereum ecosystem continues to evolve post-merge, the role and impact of liquid staking will likely remain a critical area of focus. The balance between providing liquidity and maintaining network security, as well as the potential for new financial products and services built on LSTs, will shape the future of Ethereum's PoS system and the broader DeFi landscape.

In this work, we contribute to this debate by offering the first comprehensive analysis of LST usage and circulation dynamics through the lens of money micro velocity [74, 94, 309], a granular measure of monetary activity at the individual account level. This approach enables us to detect patterns of financial behaviour that aggregate velocity metrics overlook, such as the concentration of activity among high net worth actors and the emergence of high-frequency intermediaries in an ostensibly decentralized ecosystem.

Our contributions are fourfold:

- Methodological innovation: We adapt the micro velocity framework to rebasing tokens by reconstructing transfer histories in share denominated units, overcoming

[1] A useful comparison can be made with so-called *"helicopter money"*, a monetary policy tool in which central banks distribute funds directly to households. However, in most advanced economies, such transfers are not the sole mechanism of money creation, nor are they directly tied to individual wealth. Another relevant analogy is a system of central bank digital currency (CBDC), where money exists solely in interest-bearing accounts managed by the central bank, unlike physical cash, which bears no yield.

limitations in event log availability due to the late introduction of the TransferShares event in the Lido protocol.

- Empirical insights: Using on-chain data from December 2020 to November 2024, we compute and analyze both global and disaggregated micro velocities of *stETH* and *wstETH*, revealing a strong concentration of transactional activity in a small subset of large accounts.
- Comparative analysis: We contrast the behavior of rebasing *stETH* and non-rebasing *wstETH* LSTs, highlighting how ERC-20 compliance and integration with DeFi influence token usage and reutilization patterns.
- Open science and reproducibility: To support transparency and enable future research, we release all tools and datasets used in this study at `https://github.com/LucaPennella/money-in-motion-lsts`. This includes two open-source tools: one for extracting and indexing on-chain event logs, and another for querying historical contract state variables. We also provide two curated datasets: one containing `Transfer` events for *wstETH*, and another with `TransferShares` records for *stETH*. In addition, we publish the raw event logs and contract state variables used in our analysis.

The remainder of the chapter is structured as follows. In the *Related Work* section, we discuss prior research most relevant to our study. The *Methods* section introduces the foundations of the micro velocity framework, provides an overview of the Lido platform and its on-chain architecture, the *Data* section details the data collection process, and describes the preprocessing steps used to adapt micro velocity analysis to both *stETH* and *wstETH*. In the *Results* section, we present the global micro velocity trends and their decomposition by user categories, followed by an analysis of balance dynamics and a characterization of the Lido ecosystem. Finally, the *Discussion* section interprets the empirical findings and outlines directions for future research.

8.2 Related Work

The concept of *micro velocity*, the turnover rate of money measured at the level of individual addresses rather than in aggregate, has recently enriched the economics of digital currencies [74, 94, 95, 309]. Whereas traditional velocity treats all holders symmetrically, the micro perspective exposes behavioural diversity that is especially pronounced on blockchains. Empirical work shows large cross-agent dispersion: wealthier accounts transact far more frequently and high-velocity *intermediary entities* channel a disproportionate share of flows, hinting at latent centralisation in seemingly decentralised networks [74, 94].

Complementary research documents macro-structural asymmetries. Makarov and Schoar [212] reveal the dominance of a handful of players in Bitcoin, while network studies trace persistent hierarchical patterns and accumulative advantage across multiple chains [73, 104]. Wealth inequality metrics such as the Gini coefficient broadly echo these findings [265]. Governance-oriented work further shows how power concentrates through on-chain accountability mechanisms [223], and similar network tests have begun to quantify decentralisation in lending protocols such as Aave [32].

A third strand examines *liquid staking protocols* (LSPs). Gogol et al. provide the first systematic taxonomy and risk map of liquid-staking designs [134, 135]; Scharnowski [269] analyzes basis spreads in price discovery; and Tzinas & Zindros [294] model the principal–agent tension between delegators and node operators.

Despite extensive work on native token circulation and the design of liquidity staking protocols, the intersection of staking, liquidity provision, and transactional dynamics at the agent level remains underexplored. Researchers have not yet systematically applied micro-velocity frameworks beyond native tokens, while LSP research rarely incorporates address-level behavioral analysis. In this chapter, we bridge these gaps by extending the micro-velocity methodology to Lido's rebasing (*stETH*) and wrapped (*wstETH*) tokens, uncovering distinctive patterns of liquidity concentration and transformation unique to liquid staking tokens (LSTs). In addition to velocity metrics, we incorporate address-level balance analysis, providing a deep dive into the temporal behavior of key actors. To support transparency and reproducibility, we also release curated open datasets and modular software tools that enable the community to replicate and extend our results. Together, these contributions offer a comprehensive empirical foundation for the study of liquid staking economies at scale.

8.3 Methodology

To analyze the circulation dynamics of LSTs, we adopt a micro-velocity framework that measures transactional activity at the level of individual accounts. This method allows us to move beyond aggregate metrics and identify heterogeneous behaviour among different user groups, capturing both active and passive usage patterns.

We first introduce the theoretical basis of micro velocity and how it can be adapted to account-based blockchains. Then, we outline its implementation for rebasing tokens like *stETH*, including the handling of share-based accounting. Finally, we describe how this framework is extended to the wrapped, non-rebasing token *wstETH*, enabling a comparative analysis of LST behavior within the Lido ecosystem.

8.3.1 Micro Velocity

Consider an ERC-20 token account i endowed with a certain amount of tokens. Following [94] we define the probability distribution of tokens holding times of agent i at time t as $P_i^t(\tau)$:

$$P_i^t(\tau) = \frac{w_i^t(\tau)}{M_i(t)}$$

where:

- $w_i^t(\tau)$ is the total amount of tokens held by i at time t with holding times τ and,
- $M_i(t) = \sum_\tau w_i^t(\tau)$ is the total amount of tokens held by i at time t.

The Micro Velocity of agent i is then defined as:

$$V_i(t) = \sum_\tau \frac{1}{\tau} P_i^t(\tau) = \sum_\tau \frac{1}{\tau} \frac{w_i^t(\tau)}{M_i(t)} \tag{8.1}$$

$V_i(t)$ dimension is $block^{-1}$ (i.e., measured per block). From V_i we can derive the total velocity MV of the token as:

$$M(t)V(t) = \sum_i M_i V_i \tag{8.2}$$

In order to compute $P_i^t(\tau)$ for an account-based blockchain (e.g., Ethereum), we adopt a Last-In-First-Out (LIFO) policy for token spending, where users first spend the most recently received funds; if those are insufficient, older balances with different ages are considered. This approach is economically meaningful, as it naturally distinguishes "liquid" money akin to Mo or M1 in traditional macroeconomics, which sits at the top of the wallet, from "illiquid" money, typically held as longer-term savings or investment. This LIFO assumption aligns with prior empirical work [94], which also tested First-in-First-Out (FIFO) and random mixing policies and found negligible differences in velocity estimates. Intuitively, this reflects typical user behavior: everyday spending tends to use the most recently received or most accessible funds. For the numerical results, we relied on the Python package `MicroVelocityAnalyzer`[2].

8.3.2 Lido Platform Data Overview

Lido is a liquid staking platform that enables users to stake their ETH in a liquid and fractionalized manner. Through Lido, users can acquire *stETH* tokens by depositing ETH, which Lido stakes on their behalf. In exchange, *stETH* or its wrapped version, *wstETH* tokens are transferred to the user's account, representing the staked ETH. Unlike standard staked ETH, which remains locked as collateral, these tokens retain liquidity, allowing users to transfer them freely. This feature enables users to invest amounts below the 32 ETH required to initiate a validator, without the operational responsibilities typically associated with validator duties.

As the staked ETH accumulates rewards, Lido distributes these rewards to *stETH* holders through a process known as rebasing [135]. During rebasing, the *stETH* balance in each user's account is recalculated daily to reflect the earned rewards, effectively increasing the user's *stETH* balance over time[3]. In contrast, *wstETH* does not rebase. Instead, it reflects staking rewards through a continuously increasing exchange rate relative to ETH. As a result, *wstETH* are fully interoperable with DeFi platforms where predictable token balances and ERC-20 compatibility are essential and can be bridged across multiple blockchains [135].

The data collected for this study includes the transfer records related to both *stETH* and *wstETH* tokens, capturing user transactions and the evolution of staking positions over time. Additionally, we extract the internal state of the *stETH* smart contract, which is necessary to complement transfer information, particularly for accurately interpreting reward distributions, as will be discussed in the following sections.

[2] https://github.com/fdecollibus/MicroVelocityAnalyzer/tree/parallelised_plus_bilance

[3] Of consideration for the interested reader: the core Lido repository (current implementation): https://github.com/lidofinance/lido-dao/tree/master and a primer on Lido: https://lido.fi/static/Lido:Ethereum-Liquid-Staking.pdf.

stETH Shares

The *stETH* contract employs an internal share-based accounting mechanism to track each user's stake in the Lido-controlled ETH pool. When a user acquires *stETH* tokens, they effectively purchase shares in the Lido staking pool, representing the ETH they have staked. The conversion rate between *stETH* and shares is dynamic and depends on the total ETH staked with Lido.

Two functions facilitate the conversion between *stETH* tokens and shares:

- `getSharesByPooledEth`: This function calculates the number of shares corresponding to a specified amount of *stETH*, based on the formula:

$$\text{shares} = \frac{\text{stETH amount} \times \text{total shares}}{\text{total pooled ETH}} \tag{8.3}$$

- `getPooledEthByShares`: This function determines the amount of *stETH* corresponding to a specific number of shares, following the inverse formula:

$$\text{stETH amount} = \frac{\text{shares} \times \text{total pooled ETH}}{\text{total shares}} \tag{8.4}$$

These functions ensure that the protocol can dynamically adjust share-to-token ratios, accurately reflecting changes in the total pooled ETH and thus ensuring that each user's stake aligns with the current total pool distribution [4].

Rebasing

As rewards accrue from staked ETH, the *stETH* balance for each user changes to reflect these earned rewards, without affecting the user's underlying shares. This process, known as *rebasing*, recalculates each user's *stETH* balance based on their share of the total pooled ETH. The calculation follows the formula from 8.4[5]:

$$\text{balanceOf(account)} = \frac{\text{shares[account]} \times \text{totalPooledEther}}{\text{totalShares}}$$

where:

- **shares**: A mapping of each user's share count, updated with every Ether deposit, representing their fractional ownership in the staking pool.
- **totalShares**: The sum of all shares across accounts, used to proportionally distribute the pooled ETH.
- **totalPooledEther**: The total ETH held by the protocol, defined as the sum of `bufferedBalance`, `beaconBalance`, and `transientBalance`, where:
  - **Buffered balance**: ETH stored in the contract, yet to be deposited.

[4] Source: https://github.com/lidofinance/lido-dao/blob/master/contracts/0.4.24/StETH.sol

[5] Source of the formula and example: https://docs.lido.fi/contracts/lido#rebase

- **Transient balance**: ETH submitted to the Deposit contract, pending visibility in the beacon chain state, defined as the difference between DEPOSITED_VALIDATORS_POSITION and BEACON_VALIDATORS_POSITION measured in ETH.
- **Beacon balance**: ETH held in validator accounts, reported by oracles and serving as the primary contributor to *stETH* supply.

Rebasing events are triggered by the `handleOracleReport` function, which emits the `TokenRebased` event to reflect updated values for `totalPooledEther` and `totalShares`. It is important to note that rebasing only affects the value of these two variables, which in turn affects the balance calculation of each account in the *stETH* smart contract, but the rebasing doesn't iterate over all the accounts to update their balance, which would prove unfeasible and uneconomical.

Table 8.1: Summary of recorded events, highlighting that `TransferShares` events began later than `Transfer` events.

Token	Event	Number of Records	First Record Block	Last Record Block
stETH	Transfer	2,792,968	11,480,187	21,145,533
stETH	TransferShares	2,519,615	14,860,275	21,145,533
wstETH	Transfer	1,420,359	11,888,810	21,145,533

Minting Events and Initial Token Distribution

For the scope of the present work, we are not interested in the intricate web of smart contracts that manage the deposit of ETH and minting of *stETH*. What we care about is the trace (if any) that minting (and consequently new *stETH* users' adoption) leaves on the *stETH* token transfer records.

This process is observable through a standard on-chain event pattern: every token minting is recorded as a transfer from Ethereum zero address [6] to the minting user account, i.e. the user who's depositing ETH to buy *stETH*. This is consistent throughout all *stETH* implementations in time.

ETH Redemption and Shares Burning

For the scope of our research, which is computing the micro velocity of *stETH* shares, the burning of shares is not important, as burned shares are virtually motionless, meaning they are not to be transferred anymore, and as such their weight in the velocity computation becomes negligible as time passes. As of the Lido 2.0 implementation, deployed on May 12th, 2023, following the Shappella upgrade, the burning address[7] has been introduced. Because *stETH* token burning became relevant with the implementation of stake withdrawal (as *stETH* tokens then became redeemable as well), it is reasonable to assume that we can ignore burning addresses previously set. Since the burning address

[6]Ethereum zero address: 0x0000000000000000000000000000000000000000

[7]Burning address: 0xD15a672319Cf0352560eE76d9e89eAB0889046D3

only received tokens and never sent any (as expected for such an address), we do not consider it in the velocity calculation.

Lido Events

Events in Solidity serve as a bridge between smart contracts and external applications. When state changes occur on the blockchain, events provide a way to log and track these changes. The event data gets stored in transaction logs alongside the blocks, making them accessible to applications monitoring the blockchain.

In this study, we focused our attention on two particular events, which are crucial for tracking the flow of *stETH* and share allocations, as well as for understanding changes in pooled ETH over time, which directly impacts user balances and protocol dynamics:

- `Transfer`: This standard ERC-20 event logs token transfers, enabling tracking of *stETH* movements between accounts.
- `TransferShares`: This event logs share transfers specifically, complementing the ERC-20 standard by tracking the underlying share movements not captured by token-only events.

As of the current Lido implementation (version 2.0) these two events are emitted on the Ethereum chain together when a transaction in *stETH* tokens is emitted. This reflects the atomic relation between *stETH* shares and tokens.

Lido Variables

Constant state variables in Solidity serve as immutable values that are determined at compile time and stored directly in the contract's bytecode rather than in storage. This design choice makes them highly gas-efficient since reading these values doesn't require accessing blockchain storage. They're commonly used for fixed values like role identifiers, configuration parameters, or mathematical constants that won't change throughout the contract's lifetime.

In this study, we were interested in recovering the conversion rate between *stETH* shares and tokens, which, as described in 8.3.2, required the collection of the following constant state variables of Lido smart contracts:

- `lido.Lido.depositedValidators`
- `lido.Lido.beaconValidators`
- `lido.Lido.beaconBalance`
- `lido.Lido.bufferedEther`
- `lido.StETH.totalShares`

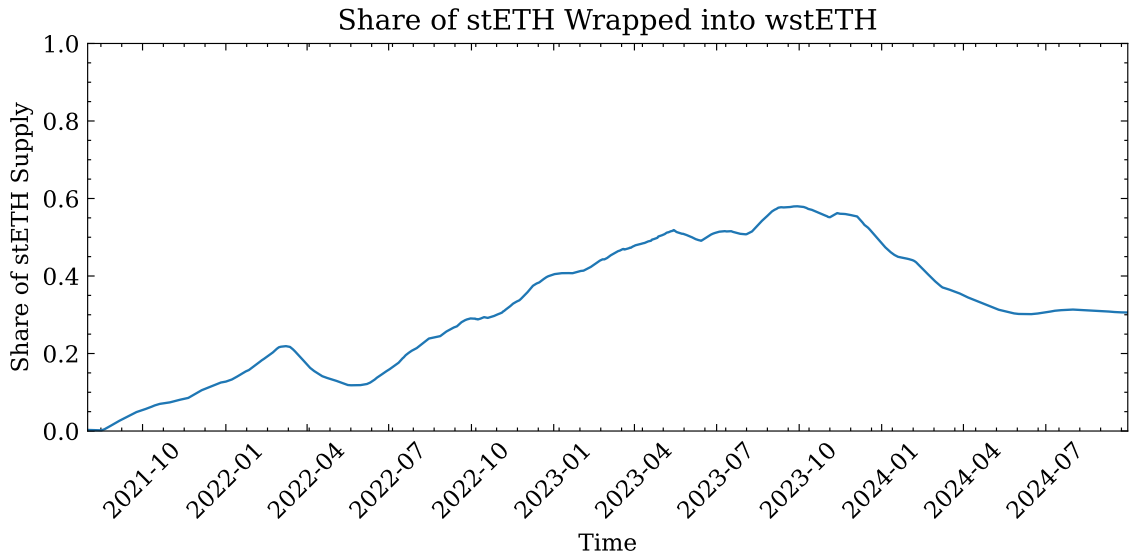


Figure 8.1: Evolution of the share of stETH wrapped into wstETH over time. The line represents the 30-day moving average of the proportion of total stETH supply held by the wstETH contract. This metric captures the relative demand for tokenized, non-rebasing stETH across the analyzed period.

wstETH Mechanics and Tracking

wstETH is the wrapped, non-rebasing version of *stETH*, designed to improve compatibility with decentralized finance (DeFi) protocols. Unlike *stETH*, which periodically updates account balances to reflect accrued staking rewards (rebasing), *wstETH* maintains fixed balances per account. Instead, it encodes staking rewards via an increasing exchange rate relative to *ETH*, updated daily in sync with *stETH*'s rebase events.

This structure makes *wstETH* fully compliant with the ERC-20 standard and particularly suitable for smart contract integration, where deterministic token behavior is essential. As a result, tracking *wstETH* on-chain is straightforward: the token emits standard `Transfer` events, and no share-based accounting is required.

In our analysis, we leverage the complete transfer history of *wstETH* to compute micro velocity without needing to reconstruct internal state variables or account for rebasing. While this limits granularity compared to *stETH*, it reflects realistic usage conditions and supports a comparative evaluation across LST designs. 8.1 illustrates the proportion of total stETH supply wrapped into wstETH over time.

We now turn to the data sources and processing steps required to apply this framework to the Lido ecosystem.

8.4 Data

To implement the micro-velocity framework described in the previous section, we constructed a dataset that captures both the transactional and structural characteristics of Lido's liquid staking tokens. This required overcoming several challenges stemming from the rebasing nature of *stETH*, the evolving protocol design, and the need for consistent longitudinal tracking of token behavior at the address level.

In what follows, we describe the data sources, tools, and procedures used to collect, reconstruct, and process the relevant on-chain information. This includes the development of custom software to extract event logs and smart contract state variables, as well as the reconstruction of share-denominated transfer histories necessary for accurate velocity computation.

8.4.1 Software Tools

Many different tools exist to collect and analyze Ethereum token data, see [323]. Given the rebase nature of *stETH*, the token is not fully ERC-20 compliant, which implies specific tools have to be developed. After a careful analysis of Lido smart contracts, we developed two software tools, the `ethereum-event-tracker` and the `ethereum-variable-tracker`, which were integrated in the data pipeline to produce the data we use in the present work.

The `ethereum-event-tracker` repository[8] was used to retrieve event records from Ethereum logs used in this study. 8.1 provides a summary of the recorded events, including record counts and the range of blocks in which these events were recorded. The `ethereum-variable-tracker` was used to retrieve the state of the constant state variables listed in the previous section for each block of Lido deployment. It is immediate to observe in 8.1, that the number of `TransferShares` events is fewer than the number of `Transfer` events, which does not sound consistent with Lido implementation. This follows the fact that the `TransferShares` event was introduced to the protocol as part of LIP-11[9], an update designed to support the protocol's adjustments in anticipation of the Ethereum Merge[10]. As described in 8.3.2, *wstETH* emits standard ERC-20 `Transfer` events, enabling straightforward tracking

8.4.2 Data Processing

In order to compute the micro velocity of *stETH* tokens, we had to account for the token's rebasing nature. An external observer who could only record transfers in *stETH* tokens would not be able to infer the real balances of an account, due to the daily change in account balances denominated in *stETH* tokens. Current methods to compute micro-velocity on Ethereum tokens rely on the collection of token transfers, and especially on the consistency of account balances denominated in the reference token. This makes it fundamental for our research to have access to the full record of *stETH* transfers expressed in shares.

Because of the late introduction of the `TransferShares` event, the collection of share-denominated transfers is incomplete for the early period of *stETH* existence. The same is not true for the `Transfer` events: as ERC-20 compatible events, these events were instantiated from the first deployment of the *stETH* contract. To fill the gap, we recovered the state of the *stETH* smart contract internal constants necessary to compute the tokens to share conversion rate and reconstruct the value in shares of each transfer originally recorded in *stETH* tokens, using the formula described in 8.3.2. This was done

[8] https://gitlab.uzh.ch/bdlt/ethereum-event-tracker

[9] https://github.com/lidofinance/lido-improvement-proposals/blob/develop/LIPS/lip-11.md

[10] https://ethereum.org/en/roadmap/merge/

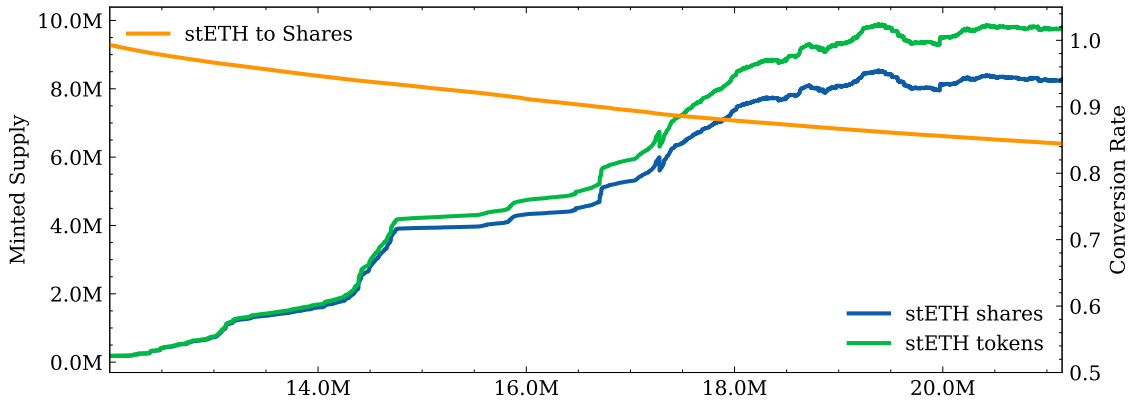


Figure 8.2: The left axis measures the the total amount of *stETH* tokens (green line) and shares (blue line) in millions. The right axis measures the conversion rate from *stETH* token to shares (orange line). On the x axis time is measured in Ethereum blocks.

for each block of Lido deployment. By doing so, we were able to reconstruct a clean database of all *stETH* transfers denominated in shares.

In 8.2 we can observe the total supply of *stETH* tokens and shares, as recovered from the constant state variable `lido.StETH.totalShares` from the *stETH* smart contract. As the supply grows, the conversion rate diminishes, meaning that the value of a single share is increasing, as the total Ethereum pooled in Lido validators increases and, as such, the rewards they bear. While this result is not novel per se, it stands to testify to the consistency of the data collection methodology. We plot the supply alongside the *stETH* to Shares conversion rate, which is used to convert token transfers into shares transfers and complete the dataset.

We also compute micro velocity separately for *wstETH*, the non-rebasing and ERC-20 compliant counterpart. Unlike *stETH*, *wstETH* maintains fixed balances and encodes staking rewards via an increasing exchange rate relative to ETH, thereby eliminating the need for share-based accounting. As such, the standard set of `Transfer` events is sufficient for computing micro velocity. While *wstETH* does not offer the granularity of share-level tracking, its simplicity and compatibility with DeFi protocols make it an important asset for comparative analysis. The inclusion of *wstETH* velocity allows us to capture a broader spectrum of user behaviour across Lido's liquid staking products.

8.5 Results

Following the methodology described in 8.3.1, we calculate the individual *stETH* and *wstETH* velocities of all token accounts during the period under consideration, from block 11,480,187 (December 18, 2020) to block 21,145,533 (November 18, 2024). Using 8.2, we derive the global velocity $V(t)$ for both assets, computed as the aggregate of individual micro velocities. The results are depicted in 8.3, revealing three key features: consistently high values, a two-phase dynamic, and several distinct spikes.

By comparing the velocity we observe to the velocities reported on Proof-of-Work (PoW) and ERC-20 tokens in the literature [74, 94], we find that the global *stETH* and *wstETH* velocity is strikingly large. We find that both tokens exhibit significantly higher

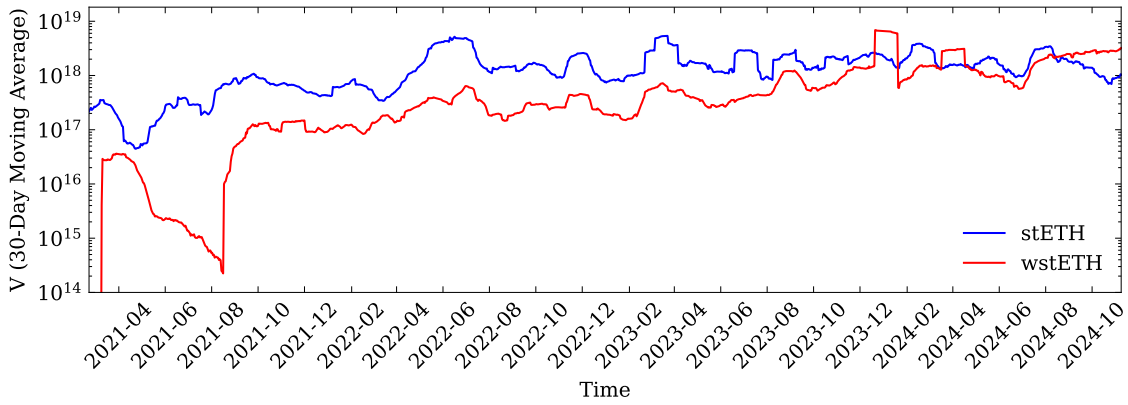


Figure 8.3: The global velocity $V(t)$ of *stETH* and *wstETH* tokens (see 8.2, the dimension is $block^{-1}$, averaged weekly.)

turnover. From a usage standpoint, we interpret this as evidence of Lido's success in achieving its mission: facilitating economic activity by unlocking the financial potential of staked tokens while maintaining Ethereum's Proof-of-Stake consensus and system security. Additionally, as interest-bearing tokens, *stETH* and *wstETH* provide an attractive alternative to ETH or other tokens as a store of value. They not only hedge against the effects of monetary expansion, thus appreciating in nominal terms, but also enable economic transactions.

Table 8.2: Summary of address categories by received *stETH*. The *Thresholds* column defines the lower and upper bounds used to assign accounts to categories.

Category	Thresholds (*stETH*)	Total Received (*stETH*)	Count
Whale	$x \geq 10000$	141703559.627	1011
Orca	$3000 \leq x < 10000$	6981034.142	1278
Dolphin	$1000 \leq x < 3000$	4074250.128	2378
Fish	$100 \leq x < 1000$	4547189.463	14441
Shrimp	$10 \leq x < 100$	1636947.564	49982
Krill	$1 \leq x < 10$	416861.963	113749
Plankton	$x < 1$	68469.071	307188

From 8.3, we observe a two-phase trajectory: an initial growth phase followed by a mature phase. The growth phase occurred between the end of 2020 and the first half of 2022, during which velocity increased, reflecting greater adoption and the maturation of Lido's economic role. Notably, Lido's merge-ready protocol upgrade took place in May 2022[11], coinciding with the Ethereum Merge, and introduced a more sophisticated version of the protocol. While this observation aligns with the natural growth in attention surrounding Lido and Ethereum's transition from PoW to PoS, it also validates our methodology, demonstrating its ability to capture significant on-chain economic events.

[11] https://research.lido.fi/t/announcement-merge-ready-protocol-service-pack/2184

A significant spike in *stETH* velocity is observed in March 2023. This aligns with user anticipation of the Shappella upgrade (April 2023), which introduced staking reward withdrawals, and the release of Lido 2.0 (June 2023). These events likely stimulated economic activity by unlocking previously illiquid capital.

Despite being a newer asset with a smaller user base, *wstETH* shows a rapidly increasing velocity, converging with *stETH* from early 2023 onward. This is particularly notable given the lower number of total addresses involved (8.3). We interpret this as a sign of higher per-token turnover, likely reflecting *wstETH*'s enhanced suitability for DeFi applications due to its fixed-balance, ERC-20-compatible design. The asset's structure encourages use in smart contracts and collateralized protocols, leading to intensive re-utilization. Notably, the global velocity of *wstETH* increased above that of its base token *stETH*. If this trend continues, differences in velocity may indicate a preferred role for staked Ether: *wstETH* functioning as an inflation-resistant form of smart money, and *stETH* serving more as a passive savings instrument.

To further investigate the source of the large velocity, we utilized the micro foundation of velocity and decomposed the global metric by contributor categories. Following the user taxonomy introduced by Lido[12], we classified all *stETH* accounts into seven categories based on the amount of *stETH* received during the study period (December 2020 to November 2024). 8.2 and 8.3 summarize these categories and their statistics for *stETH* and *wstETH*.

Table 8.3: Summary of address categories by received *wstETH*. The *Thresholds* column defines the lower and upper bounds used to assign accounts to categories.

Category	Thresholds (*wstETH*)	Total Received (*wstETH*)	Count
Whale	$x \geq 10000$	149285851.223	780
Orca	$3000 \leq x < 10000$	4618320.295	850
Dolphin	$1000 \leq x < 3000$	2457660.110	1401
Fish	$100 \leq x < 1000$	2084328.725	6224
Shrimp	$10 \leq x < 100$	413353.129	11285
Krill	$1 \leq x < 10$	54465.652	14063
Plankton	$x < 1$	8030.266	56518

For *stETH*, whale accounts, that received at least 10,000 tokens, represent just 0.25 % of all addresses (1,011 out of more than 480,000) yet collectively received approximately 141.7 million *stETH*, a dominant share of the total distribution. This confirms a well-documented trend in blockchain economies: a small number of large accounts command a disproportionate amount of monetary flow [73, 104, 296]. Mid-tier categories (e.g., Orcas and Dolphins) also hold substantial quantities, while the vast majority of accounts (Plankton, Krill, Shrimp) received comparatively minor volumes.

A similar concentration is observed for *wstETH*. Although whales are even fewer in number (780 addresses), they received nearly 149.3 million tokens, again constituting the majority of the total supply. Notably, the distribution for *wstETH* is even more top-heavy: small holders (Plankton, Krill, Shrimp) are fewer in number than in the *stETH*

[12] https://blog.lido.fi/analysis-of-steth-user-behaviour-patterns/

dataset and received a smaller aggregate amount (~476k tokens vs. over 2.1 million for *stETH*). This suggests a more targeted adoption of *wstETH* among sophisticated or institutional users, possibly driven by its composability and compatibility with DeFi infrastructure.

Taken together, the distributions reported in 8.2 and 8.3 indicate that the usage of liquid staking tokens exhibits a marked degree of asymmetry. In both *stETH* and *wstETH*, a small number of high volume addresses, classified as whales, receive the majority of tokens, while the vast majority of accounts hold relatively modest amounts. This pattern suggests the coexistence of two distinct user profiles: on one hand, large entities that actively transact substantial volumes of LSTs, likely for integration within DeFi protocols; on the other, a broader base of smaller holders whose participation is limited to token acquisition, likely to passively benefit from staking rewards.

When examining the decomposition of micro velocity by account category for both *stETH* and *wstETH* (8.4a and 8.4b), a consistent concentration pattern is observed. In each case, whale accounts representing less than 1 % of the total account for the overwhelming majority of total velocity. For *stETH*, whales contribute approximately 99 % of the aggregate velocity over the entire observation period. A comparable concentration is found for *wstETH*, where high-volume accounts similarly dominate the transactional dynamics.

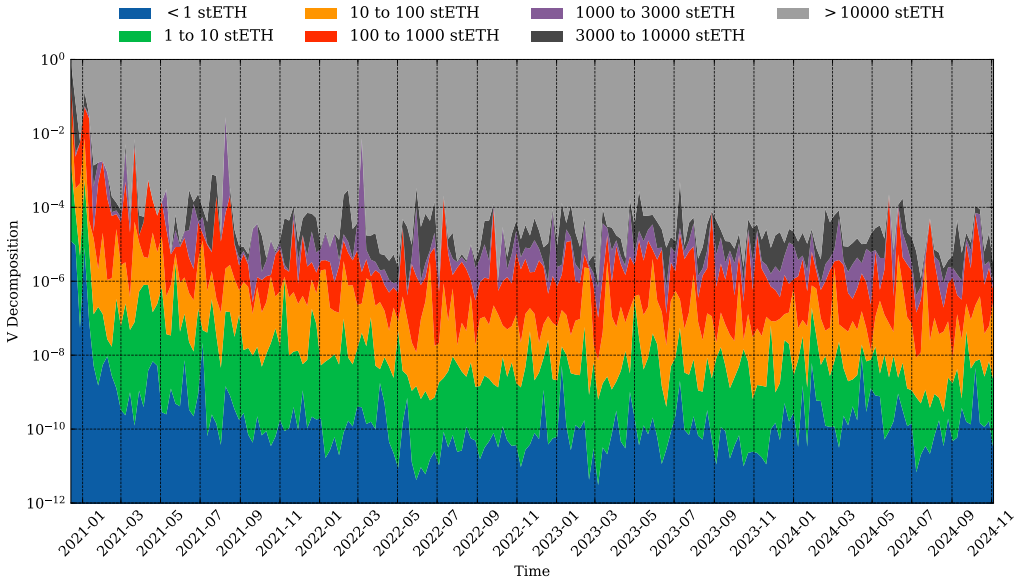
Outside the whale category, the remaining velocity contributions are substantially lower and display a more uniform distribution across mid and small volume groups. This stratification is more clearly visible in the log-scale representation of velocity shares, which reveals the limited role of lower-tier accounts in driving token circulation. These findings suggest that, while liquid staking tokens are widely distributed in terms of address count, the effective circulation is predominantly sustained by a narrow subset of large participants.

The observed concentration of velocity among whale accounts is likely influenced by the integration of liquid staking tokens into decentralized finance applications. Large accounts may correspond to actors such as DeFi protocols, liquidity providers, or institutional agents that engage in high-frequency and high-volume transactions, thereby contributing disproportionately to total velocity. This mechanism appears consistent across both *stETH* and *wstETH*, although the fixed-balance structure and DeFi-oriented design of *wstETH* may further incentivize its use in automated financial operations. This analysis allows us to identify two distinct user typologies for Lido (and, by extension, for LSTs):

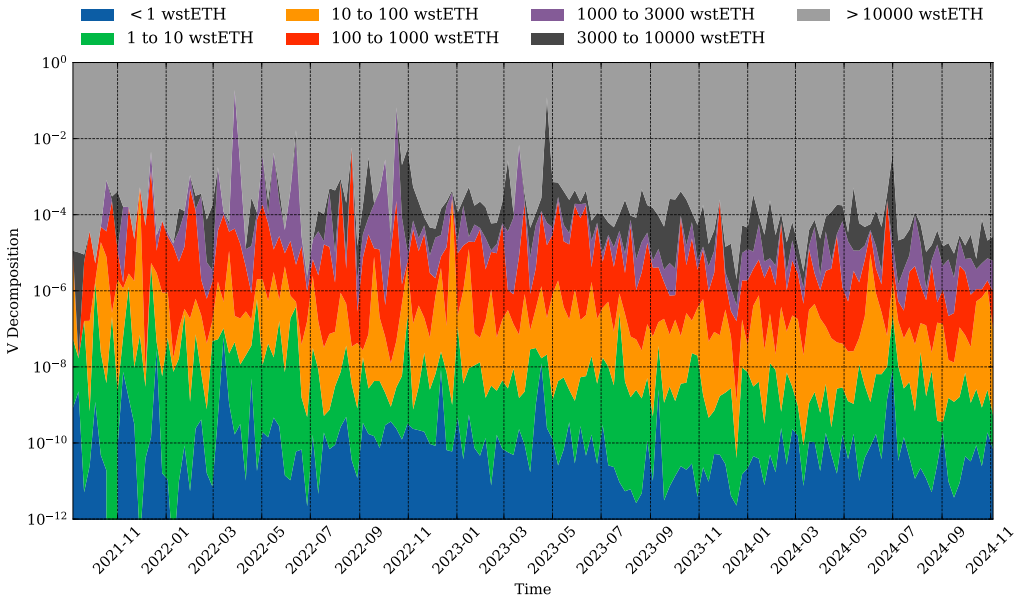
1. A limited set of high-capacity accounts that receive and transact large volumes of liquid staking tokens, likely reflecting their integration into decentralized finance infrastructures.

2. A broad base of smaller accounts that primarily hold LSTs in a passive manner, likely to access staking rewards without running a validator, and that exhibit low transactional activity.

Beyond aggregate measures of micro-velocity, we further explore the behaviour of the largest token holders through account-level balance dynamics. 8.5 plots the 30-day moving averages of *wstETH* balances for the five largest non-exchange holders, prominent DeFi actors such as Aave, Balancer, Spark and SkyMoney.



(a) Micro velocity shares decomposition for *stETH*, by categories defined in 8.2. Expressed in log scale.



(b) Micro velocity shares decomposition for *wstETH*, by categories defined in 8.2. Expressed in log scale.

Figure 8.4: Comparison of micro velocity decomposition across address categories for *stETH* and *wstETH*. Both distributions are highly concentrated in Whale addresses.

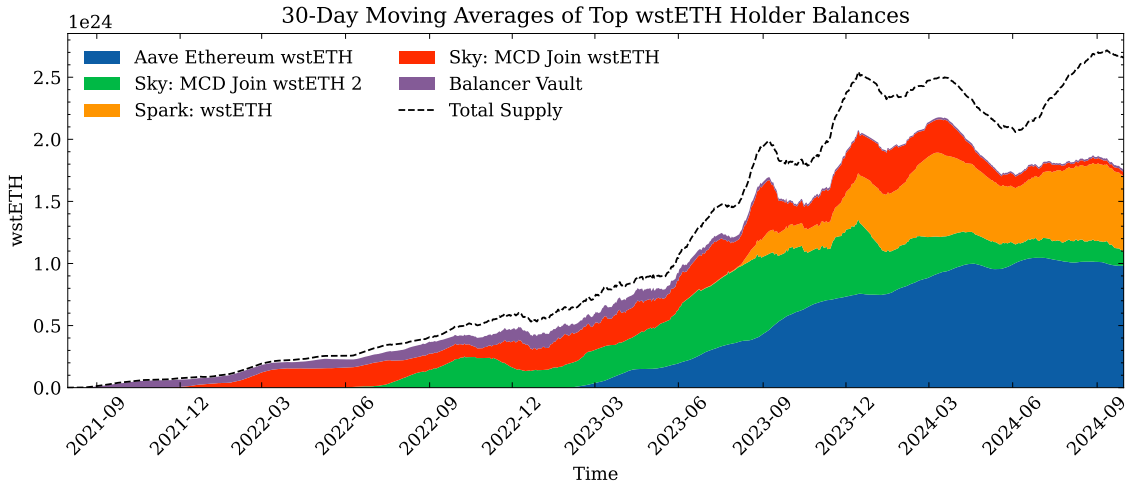


Figure 8.5: 30-day moving averages of *wstETH* balances for the top five holder addresses. These include major DeFi protocols such as AAVE, Spark, Balancer, and SkyMoney.

Table 8.4: Identified Ethereum Addresses and Associated Protocol Entities.

Address	**Protocol Entities**
`0x0b925ed163218f6662a35e0f0371ac234f9e9371`	Aave Ethereum wstETH
`0x12b54025c112aa61face2cdb7118740875a566e9`	Spark: wstETH
`0x248ccbf4864221fc0e840f29bb042ad5bfc89b5c`	SkyMoney: MCD Join wstETH 2
`0x10cd5fbe1b404b7e19ef964b63939907bdaf42e2`	SkyMoney: MCD Join wstETH
`0xba12222222228d8ba445958a75a0704d566bf2c8`	Balancer Vault

8.4 lists these contracts. The Balancer V2 Vault is the upgradeable contract that escrow-holds every Balancer pool's assets and executes swaps, joins, exits and external asset-management hooks; whenever a pool contains *wstETH*, the Vault's inventory rises accordingly, explaining its position as the single largest holder outside centralised exchanges. The Aave V3 *wstETH* aToken and the homologous Spark Lend receipt token serve as interest-bearing ERC-20 wrappers. SkyMoney accepts *wstETH* as collateral through two join adapters[13], whose dual-adapter architecture enables differentiated risk parameters. Collectively, these five contracts account for the overwhelming majority of *wstETH* held outside centralised exchanges, illustrating the token's deep integration into liquidity-provision, yield-generation and collateral-management workflows across the contemporary DeFi stack.

The time series reveals stable yet distinct trajectories, with entities like AAVE and Spark consistently maintaining large holdings, suggesting long-term integration of *wstETH* into lending and collateral strategies.

It is noteworthy that an examination of the top five accounts already reveals a significant concentration of wstETH holdings, consistent with the broader concentration patterns observed in the velocity decomposition.

[13] https://developers.sky.money/protocol/core/join/

8.6 Discussion & Conclusion

In this work, we provided a detailed analysis of the *stETH* and *wstETH* token transfer protocols, offering novel insights into the behaviour of Lido users by computing and decomposing the micro velocity of both liquid staking tokens (LSTs). By leveraging on-chain data, we captured the transactional activity of accounts across the entire lifespan of these assets, revealing both structural patterns and temporal dynamics.

Our findings show that both *stETH* and *wstETH* exhibit remarkably high global velocity, especially compared to previously studied Proof-of-Work assets and standard ERC-20 tokens. This supports the hypothesis that LSTs effectively unlock the liquidity of staked assets without undermining the security guarantees of Ethereum's Proof-of-Stake consensus. As a result, these tokens circulate actively within the ecosystem while simultaneously accruing staking rewards. The increasing global velocity of *wstETH* may serve as a proxy for its depth of integration and functional usability within DeFi applications, especially when compared to the velocity of its base token.

We also observe a growing proportion of *stETH* that is wrapped into *wstETH*, underscoring a demand for non-rebasing tokens, likely driven by their ERC-20 compliance and deterministic behaviour in smart contract environments.

A key outcome of our decomposition is the identification of a highly asymmetric distribution of transactional activity. A small subset of accounts, classified as whales, is responsible for the vast majority of observed velocity in both *stETH* and *wstETH*, despite constituting less than 1 % of all users. These high-wealth actors likely represent institutional agents, DeFi protocols, and liquidity providers who integrate LSTs into automated or high-frequency financial strategies. Conversely, a large number of smaller accounts appear to use LSTs in a passive manner, primarily as a means of earning staking rewards without running a validator.

Further insights emerge when focusing on top holders' balance trajectories (8.5). We observe persistent large-scale holdings by DeFi-native actors where a strong concentration is already evident among the top five addresses. This suggests a deliberate integration of the wrapped token into lending, collateralization, and composable DeFi protocols.

Taken together, these findings illustrate the emergence of a dual user base: on one hand, high-frequency DeFi participants actively leveraging LSTs within financial protocols; on the other, a broad population of low-activity holders passively engaging with staking infrastructure. This duality underscores the financial maturity and evolving specialization of LSTs in Ethereum's ecosystem.

To place our findings in proper context, the same micro-velocity lens must now be applied to other LSTs on Ethereum (e.g., *cbETH*, *rETH*) and to liquid-staking assets on external PoS chains. A cross-token panel will calibrate what "high" or "low" velocity really means and let us classify assets by behavioural archetype: passive store-of-value, actively rehypothecated collateral, bridge-hopping fuel, and so forth. Because micro-velocity reacts at the transaction level, it can surface shifts in user behaviour long before coarse indicators such as raw volume or TVL respond, making it a candidate early-warning metric for researchers, protocol teams, and on-chain risk monitors.

For future work, it would be worthwhile to expand the analysis to *yield farming* behaviours, whereby LSTs are recursively used as collateral to borrow and reinvest in additional staking positions. Such strategies, implemented through yield optimizers and

money market protocols, may amplify both token velocity and systemic leverage. Identifying these patterns on-chain could help quantify endogenous risk and clarify their impact on the monetary dynamics of liquid staking ecosystems.

These results ultimately serve to reinforce the pivotal role of Lido within the context of Ethereum's staking and DeFi landscape. By providing liquid staking infrastructure that balances accessibility, composability, and economic utility, Lido has not only become the dominant LST provider but also a structural pillar of post-Merge Ethereum. It is imperative to comprehend the dynamics that are facilitated by metrics such as micro velocity in order to grasp the evolving monetary and systemic properties of Proof-of-Stake ecosystems.

Data Availability

The datasets and tools used in this study are publicly available at https://github.com/LucaPennella/money-in-motion-lsts

Chapter 9

A Unified Framework and Comparative Study of Decentralized Finance Derivatives Protocols

9.1 Introduction

Derivatives protocols are an established category of Decentralized Finance (DeFi) applications that enable users to create, trade, or manage derivatives products. These financial instruments are implemented through smart contracts deployed on a Distributed Ledger Technology (DLT) like Ethereum. Their value is derived from an underlying asset or index, such as cryptocurrencies, fiat currencies, commodities, or other financial indicators.

Despite a strong initial interest within the DeFi community, the growth of derivatives protocols experienced a contraction after mid 2022. However, as Figure 9.1 shows, the total U.S. dollar value of digital assets staked in the smart contracts of major derivatives protocols - known as the *Total Value Locked* (TVL) - has started increasing again rapidly since 2024, driven mainly by new blockchain platforms, such as Solana and Arbitrum. This renewed growth is primarily due to the recent surge of new market entrants that are challenging existing incumbent decentralized applications implemented on the Ethereum blockchain. This scenario suggests an evolving, competitive landscape within the DeFi derivatives sector and a renewed interest from developers and DeFi users.

Derivatives protocols are commonly mentioned as one of DeFi's core decentralized applications (dApps), along with Protocols for Lending Funds (PLFs), Decentralized Exchanges (DEXs) with Automated Market Makers (AMMs), Yield Aggregators, and Liquid Staking protocols [99, 135, 151, 272, 310, 313]. Like other DeFi dApps, derivatives protocols rely on blockchain technology and web services to offer decentralized financial instruments. In the best case scenario, they might shape the financial and web ecosystems through algorithmic automation, competitive financial engineering, and new forms of openness to the financial sector [39]. In contrast to other well-researched DeFi applications, however, they remain relatively understudied by the scientific literature. Indeed, existing studies on derivatives in the crypto ecosystem mostly focus on tokens [142, 210] or on products traded on centralized exchanges [155]. To the best of our knowledge, no study systematically analyzes decentralized derivatives protocols and the services they offer.

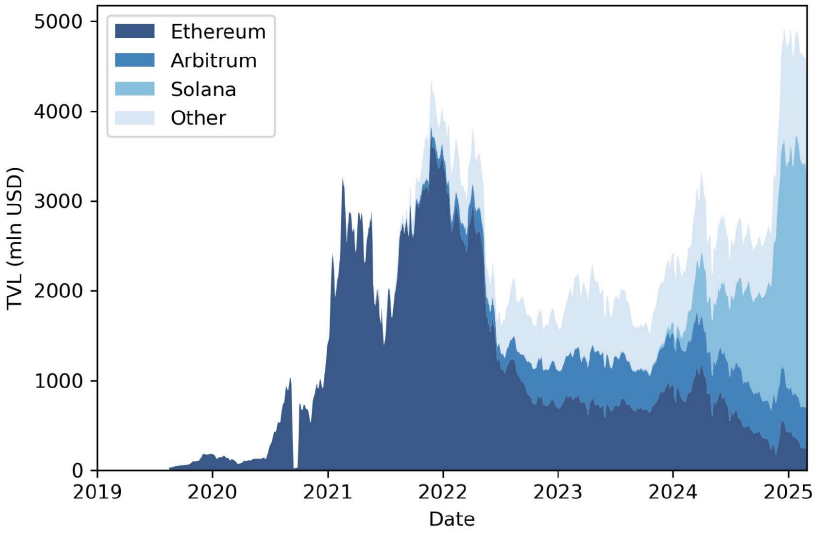


Figure 9.1: Total Value Locked (TVL) evolution over time for the derivative protocols reported in Table 9.1, distinguished by chain. Data extracted from DeFiLlama [13].

This chapter addresses this research gap by systematically reviewing existing DeFi derivative protocols, examining their design features, operational mechanisms, and the evolving market dynamics that influence their role in the DeFi ecosystem. We follow a structured process that consists of selecting a sample of major derivatives protocols (whose list is reported in Table 9.1), thoroughly analyzing their documentation and sources, and collecting their fundamental characteristics.[1] Our systematization allows us to highlight the similarities and differences among these protocols, including the derivative products they offer, their design principles, and the underlying cryptoassets and collateral they manage. As a result of this analysis, we devise a conceptual framework that generalizes and captures the main features of this class of protocols, describes the main entities involved in derivatives issuing and trading, the protocols' structures and dynamics, and further introduces fundamental metrics driving their behavior. More specifically, our framework includes: (i) a formal representation, expressed through tuples, of the derivative cryptoassets managed by these protocols; (ii) a taxonomy of the main actors and entities involved in their functioning; (iii) a characterization of protocol behavior for creating, trading and managing cryptoassets, as well as the identification of their core financial functionalities; (iv) finally, a simulation environment building on these formal specifications acting as a tool that generalizes protocol behavior. Through our framework, we thus aim to provide a more comprehensive understanding of the working principles of derivatives protocols and their implications for decentralized finance's evolution.

The remainder of the chapter is structured as follows: Section 9.2 provides relevant background information on derivatives in traditional finance and their evolution within Decentralized Finance. In Section 9.3, we review the existing literature on derivatives and existing frameworks related to the formalization and implementation of derivative

[1] The entire list of sources used is reported in Table D25 in the Appendix.

Table 9.1: List of protocols analyzed (top-5 by TVL on DeFiLlama [13] for the categories Derivatives - issuing perpetuals -, Synthetics, Options). Data includes the protocol category, its main chain, TVL in US$ on 31st Dec. 2024 and the TVL percentage variation within 2024.

Protocol	Category	Main Chain	TVL (m USD)	TVL Change (2024)
Jupiter	Perpetuals	Solana	1720.88	+1864.36%
Drift Trade	Perpetuals	Solana	802.55	+595.48%
GMX	Perpetuals	Arbitrum	696.06	-29.07%
dYdX	Perpetuals	dYdX	431.56	+23.69%
Hyperliquid	Perpetuals	Hyperliquid	398.65	+151.84%
Synthetix	Synthetics	Ethereum	381.33	-50.23%
Derive	Options	Ethereum	82.44	+2202.17%
Alchemix	Synthetics	Ethereum	67.79	-7.15%
Youves	Synthetics	Tezos	47.60	+10.02%
GammaSwap	Options	Arbitrum	27.36	+34860.89%
Deri V4	Options	Linea	24.67	+96.20%
Metronome	Synthetics	Ethereum	15.88	+1.97%
Hegic	Options	Arbitrum	13.40	-10.65%
SOFA.org	Options	Arbitrum	6.94	+1230.15%
Taiga	Synthetics	Acala	2.76	-85.64%

contracts in DeFi. Section 9.4 presents the methodology we adopted to select the derivative protocols analyzed and the data sources we relied on. Section 9.5 introduces a mathematical formalization of derivative contracts, while Section 9.6 presents an overview of the key entities involved and the operational dynamics of decentralized derivatives protocols, offering a generalized framework that synthesizes their main participants and examines exchange model dynamics. In Section 9.7 we complement the theoretical framework with Monte Carlo simulations to analyze the interactions between price variations of the underlying assets and protocol design under varying conditions. Finally, Section 9.8 concludes and discusses future lines of investigation. The appendix provides additional tables, the entire list of sources utilized, and further details on derivatives in traditional finance.

9.2 Background

In this section, we provide background information to position derivatives protocols into the landscape of DeFi and relate them to their corresponding equivalents in traditional finance.

9.2.1 Derivatives contracts in traditional finance (TradFi)

A derivative contract in TradFi is a financial instrument whose value derives from an underlying variable(s), ranging from the price of a traded asset to the performance of an index, an underlying commodity, or a currency rate, among others. It represents an agreement between two parties to exchange value based on changes in the valuation of the underlying over time.

The global derivatives market is massive and complex [256], thus a comprehensive discussion of all existing derivatives contracts goes beyond the goals of this study. In the

following, we introduce key elements and the primary instruments traded in traditional finance.

Forward contracts are customized contracts between two parties, typically traded over-the-counter (OTC), to buy or sell an asset at a specified future date for a price previously agreed upon. OTC refers to a kind of market where transactions occur bilaterally between participants, without the involvement of a centralized exchange. *Futures contracts* are similar to forwards, but they are standardized and traded on exchanges. Periodic funding payments from long to short positions ensure that their price tracks and ultimately converges to the spot price of the underlying asset. *Options* grant the buyer the right, but not the obligation, to buy (call) or sell (put) an asset at a specified price within a defined period. This advantage comes at the cost of an up-front fee. *Swaps* are agreements between two parties to exchange cash flows or other financial instruments. The most prominent ones are interest rate swaps, credit default swaps, currency and equity swaps.

Derivatives are typically traded on centralized exchanges or over-the-counter (OTC). While the two mechanisms exhibit differences, in both cases several intermediaries are involved in the contract creation, from brokers, who facilitate user access to exchanges support negotiations, to financial institutions, that create and manage the trading of the derivatives, and clearinghouses, that act as a central counterparty to guarantee transaction settlement and risk mitigation. More details on these products are given in Appendix **??**.

9.2.2 Derivatives contracts in crypto Centralized Finance (CeFi)

Derivatives based on blockchain technology can either refer to conventional derivatives whose underlying is a cryptocurrency or to products that leverage blockchain technology to create a derivative. The former are financial instruments typically created and traded through centralized cryptocurrency exchanges (CEXs) such as Binance or BitMex.[2] These trading platforms enable customers to trade cryptocurrencies or derivatives built on top of them in a centralized ecosystem, with trades executed on a platform rather than on a blockchain and the exchange acting as a custodian and counterparty. CEXs and other cryptoasset service providers are part of Centralized Finance (CeFi), an ecosystem that mirrors traditional financial systems despite being based on distributed ledger technologies [262].

Whilst derivatives offered by CEXs share similarities with TradFi products, the CeFi derivatives market has distinct characteristics. One primary difference is that the market of derivatives in CeFi is largely dominated by perpetuals, a type of derivative contract similar to futures, but without an expiration date. Interestingly, perpetuals were first proposed by [276] in 1993 with no relation to cryptocurrency markets. However, they were not implemented in practice for a long time, until they started gaining significant traction and wider adoption in cryptocurrency markets, when they were implemented and pioneered by BitMEX [55]. Perpetuals, like futures, rely on a funding-fee-based approach, a mechanism where traders periodically pay each other a fee to keep the contract's price aligned. This system was later adopted also by other cryptocurrency-specific derivative protocols, such as everlasting options (a novel category of derivatives [5] which will be

[2] https://www.binance.com/en and https://www.bitmex.com/

discussed in greater detail in the subsequent sections).

9.2.3 Derivatives contracts in Decentralized Finance (DeFi)

The second type of blockchain-based derivatives comprises instruments that are natively built on blockchain infrastructure. These are financial products created by Derivatives protocols, i.e., DeFi applications that leverage DLTs to replicate existing financial instruments in a decentralized system. Unlike derivatives offered in centralized ecosystems, these financial instruments are issued and traded automatically and deterministically on-chain through smart contracts. Users interact directly with the contracts, rather than with other users, to open and close their positions – a mechanisms known as the peer-to-pool model [39]. Trading and clearing mechanisms are meant to be implemented on-chain too, seeking to eliminate the reliance on centralized intermediaries.

According to leading DeFi data aggregators, derivative protocols are commonly grouped into three categories: perpetuals (often listed under 'derivatives'), synthetics, and options (see Table 9.1).

Perpetuals are similar to those implemented in CEXs such as BitMex; Synthetic assets are a type of contract created to track the performance of various assets without requiring their ownership; and Options assets mimic their equivalent in traditional finance.

DeFi derivatives typically reference crypto-native assets as the underlying asset, but may also track traditional financial indices or synthetic representations of real-world assets. Notably, DeFi derivatives may thus rely on oracles, i.e., blockchain-based services that feed external (off-chain) data to a blockchain or smart contract, to correctly price derivatives [115].

Note that issuing and trading derivatives on cryptocurrencies takes place mostly on centralized cryptocurrency exchanges at the moment, with daily trading volume in the order of magnitude of billions [17]. However, although we acknowledge that the CeFi market is larger than that of DeFi today, the former does not differ substantially from traditional derivatives, while the latter represents a rapidly growing segment of the crypto market and a significant innovation in financial infrastructure design. For this reason, we will focus on this category in the remainder of the chapter.

9.3 Related Work

Since most derivatives trading volume still occurs on centralized exchanges, prior studies have primarily examined perpetual derivatives and other instruments issued by centralized exchanges [254, 280], focusing e.g. on price discovery dynamics [23], market liquidity and quality [260], contract design and market microstructure [103], and deriving no-arbitrage pricing for perpetual futures under frictionless assumptions [17, 155].

In contrast, DeFi derivative protocols remain comparatively understudied. To date, the DeFi applications that have received more attention by the academic community are protocols for loanable funds [48, 141, 188] and DEXs with AMMs [52, 314], that respectively enable users to lend (borrow) and trade crypto assets. More recently, however, the DeFi derivatives ecosystem has gained attention too. A stream of literature focuses on DeFi-based perpetuals: [83] systematize differences between CeFi and DeFi perpetuals; [30] formally model two DeFi perpetual designs; and [110] introduce an AMM-based

mechanism emphasizing liquidity dynamics and price robustness. Compared to perpetuals, the DeFi options and synthetics market is smaller (as shown in Table 9.1) and less studied. [279] examine decentralized options designs, while [28] compare CeFi and DeFi options, arguing that higher costs in DeFi reflect thinner liquidity and limited retail demand.

Recent research has drawn parallels between classical and data-driven models: [187] finds jump-diffusion and stochastic-volatility models outperform Black–Scholes for BTC and ETH options; [66] shows tree-based ML with high-frequency volatility estimates exposes larger inefficiencies relative to traditional assets. At the protocol layer, [321] proposes UP-BLOC, a universal options framework with unlocked collateral and signature-based logic. Also risk in crypto derivatives has been analyzed from hedging and margin perspectives: [24] minimizes variance and liquidation risk jointly; [89] advocate higher margins for highly leveraged positions using extreme-value evidence.

In summary, existing studies on DeFi derivatives protocols focus on individual derivative protocol categories or investigate a limited set of features, rather than comparing their full design space across protocols. By contrast, we adopt a broader perspective, analyzing these protocols - perpetuals, options (expiring and everlasting), and synthetics - altogether, and comparing both their financial elements and protocol dynamics within a unified formal approach. To the best of our knowledge, no prior work (i) builds such a unified cross-protocol conceptualization; (ii) maps actors, components, and operational flows across protocol designs; and (iii) complements it with a simulation framework that quantifies how protocol parameters and market conditions affect profitability and liquidation risk of a position; thus differentiating our work from instrument- or protocol-specific studies.

9.4 Methodology and Data

We describe the methodology used to select the derivative protocols for our systematization and the data sources analyzed to build our framework. We follow a structured process that consists of three main phases: protocol selection, data source analysis, and derivation of the formal framework. In these phases, we combine a data-driven approach for protocol selection with qualitative and technical analysis to extract relevant information. Below, we detail each of them.

Protocol selection To identify the most prominent derivatives protocols, we rely on DeFiLlama [13], a prominent DeFi data aggregator that tracks metrics like TVL (Total Value Locked), protocol performance, and categorizes DeFi protocols according to the financial services they offer. In addition to the *Derivatives* category, which comprises primarily protocols issuing perpetuals, DeFiLlama includes other related categories, such as Synthetics, Options, Risk Curators, and Insurance.[3] The latter two mostly focus on coverage against on-chain risks, and we consider them to be beyond the goal of our study. We therefore focus on the three subcategories *Perpetuals*, *Synthetics*, and *Options*. For each category, we selected the top five protocols based on Total Value Locked (TVL)

[3] A full list of DeFiLlama categories can be found at https://defillama.com/categories

serving as a proxy for user adoption and capital engagement.[4] The rationale behind our selection is that, given the highly dynamic nature of DeFi, covering all derivatives protocols would be challenging; therefore, we focus on the top ones, which are the most widely adopted and assume that their mechanisms and dynamics are representative of the largest part of the ecosystem.

Once we identified a list of candidate protocols, we performed a filtering step as follows. We manually verified the accessibility of each protocol's public interface, including the website, the trading application (dApp), and the documentation. Protocols lacking accessible interfaces or documentation were excluded. For instance, Outcome Finance [12] was excluded due to inaccessibility, while Opyn [11] was omitted because its latest dApp was only available on testnet and not yet accessible to all users.

Table 9.1 reports the results of our selection process. More specifically, it summarizes the information on the five most relevant derivative protocols ranked by TVL for the three protocol categories analyzed. Moreover, the table shows that since 2024, the most relevant protocols have been deployed on chains alternative to Ethereum, with steep growth rate increases, especially on Solana and Arbitrum.

Protocol Analysis For each of the protocols of Table 9.1, we conducted an in-depth manual analysis of their design and operational mechanisms, relying on the following sources to retrieve all the information our framework relies on:

- official documentation such as whitepapers, developer documentation, and user guides;
- public interfaces such as dApps and websites;
- smart contract source code, when available via Etherscan, GitHub, or protocol explorers;
- online communities of the protocols, e.g., official Discord channels, and technical articles [4, 9, 10] for implementation-specific insights.

Table D25 reports the entire list of sources we used for each analyzed protocol.

Framework Derivation Following an inductive approach, we started from the collected data and performed a generalization and abstraction process. We cleaned, filtered, compared, and harmonized the data and information collected in the previous phase. This refined data and information served to extract the essential elements of each class of protocols, reorganize and summarize their common features, and lay the foundation for the formal framework presented in the following sections.

More specifically, this framework includes:

- A formal representation of the main derivative cryptoassets traded on or across these protocols;

[4] While we acknowledge that TVL is an imperfect metric currently lacking a standardized methodology [210, 263], it is widely accepted as a measure of performance and scale of DeFi protocols and DeFi more broadly.

Table 9.2: Summary Table with terms and variables introduced in our framework.

Asset	**Definition**
Perpetual	$\langle U, C, L, S, P_{\text{entry}}, F_m \rangle$
	Futures contracts without an expiration date, allowing traders to hold positions indefinitely.
Expiring Option	$\langle U, C, L, S, P_{\text{strike}}, T_{\text{expiry}} \rangle$
	Grants the right, but not the obligation, to buy (call) or sell (put) an asset at a predefined strike price before or at expiration.
Everlasting Option	$\langle U, C, L, S, P_{\text{strike}}, F_m \rangle$
	Option contract without expiration, with continuous funding.
Synthetic	$\langle U, C, P_U, P_C \rangle$
	Tokenized derivatives tracking the price of an underlying asset.
Asset Components	**Definition**
Underlying asset (U)	The asset on which the derivative contract is based.
Collateral (C)	The asset posted to support a position.
Leverage (L)	The ratio of borrowed funds to the trader's capital, thereby enabling a greater exposure to the underlying asset.
Strategy (S)	Trading direction: $+1$ for long, -1 for short.
Entry Price (P_{entry})	Price at which the position is opened.
Strike Price (P_{strike})	Price at which the option can be exercised.
Expiration Date (T_{expiry})	Last valid day to exercise an option.
The funding mechanism (F_m)	Transfers cash flows between long and short positions to ensure that the derivative price converges toward a protocol-specific reference value.
The Underlying Price (P_{U})	Reference price of the underlying U.
The Collateral Price (P_{C})	Reference price of the Collateral C

- A taxonomy of the main actors and entities (e.g., traders, liquidity providers, oracles, clearinghouses) involved in the trading and crypto-asset management;
- A description of the key dynamics and trading flows enabled by the protocols;
- A mathematical characterization of core financial functions provided by these protocols (reported in the Appendix).

The following sections detail each of the items above. Our framework facilitates a structured comparison of heterogeneous derivatives protocols and provides foundational tools for future theoretical and applied research in decentralized financial engineering.

9.5 Cryptoassets of Decentralized Derivative Protocols

To begin, we identify and formalize the cryptoassets managed by derivative protocols.

We cover three categories: *Perpetuals*, *Options*, and *Synthetics*. For each of them, we introduce a tuple-based formal notation that provides a rigorous representation of their core elements and operating logic and serves as the foundation for the developments in the next sections. It can also be used as a reference for prototypical implementations of these instruments.

Table 9.2 summarizes the main terms and notions used in this section.

Appendix 11 provides supporting notation and definitions, together with the explicit formulae for these and other concepts used throughout the contribution (e.g., profit and losses or PnL, margin requirements, and collateralization ratios).

9.5.1 Perpetuals

Perpetuals are futures contracts without an expiration date, allowing traders to hold positions indefinitely. They enable traders to take positions on the price movements of an underlying asset without the need to renew contracts as they would with traditional futures.

In our framework, a perpetual futures contract is a tuple:

$$\langle U, C, L, S, P_{\text{entry}}, F_m \rangle$$

where the underlying asset (U) is the financial instrument on which the perpetual futures contract is based, and that serves as the reference for the pricing and valuation of the contract.

The Collateral (C) refers to the assets the buyers of the contract must deposit to secure their positions and to manage the associated risks.

Using collateral allows traders to increase their exposure beyond their initial capital, enabling leveraged positions. The Leverage (L) is thus defined as the ratio of borrowed funds to the trader's capital, thereby enabling a greater exposure to the underlying asset. Protocols allow leverage from 1.1x up to over 100x; traders usually borrow assets from a liquidity pool to create a leveraged position. This borrowing incurs a *borrow rate fee*, which is paid to the pool. Protocols also specify a *maintenance margin* – i.e., the minimum equity required to keep a leveraged position open. If equity approaches the threshold, traders may need to add collateral or reduce exposure.

The strategy (S) indicates the trader's position:

- Long Position (when $S = +1$): The trader expects the underlying asset's price to increase.
- Short Position (when $S = -1$): The trader expects the underlying asset's price to decrease.

The entry price (P_{entry}) represents the price at which a trader initiates a perpetual futures position. The funding mechanism (F_m) transfers periodic cash flows between opposite sides to align the derivative with a reference price (e.g., perpetuals vs. spot price, everlasting option vs. benchmark). A positive rate implies payments from longs to shorts, and vice versa for negative rates.

Table 9.3 reports information on the underlying assets and collateral available on the perpetual protocols investigated.

Table 9.3: Overview of selected perpetual protocols (snapshot: *[April 2025]*). Supported Assets are grouped by categories; each cell encodes (underlying, collateral) support from left to right: ●○ = underlying only, ○● = collateral only, ●● = both, ∞ = not supported.

Protocol	Supported Assets								#U	#C
	L1	L2	DeFi	Meme	Gaming	Forex	RWA	Stable		
Jupiter Exchange	●●	∞	∞	∞	∞	∞	∞	○●	3	5
Drift Trade	●●	●○	●○	●○	∞	∞	∞	○●	49	2
GMX (Arbitrum)	●●	●●	●●	●●	●●	∞	∞	○●	49	49
dYdX	●○	●○	●○	●○	●○	●○	●○	●●	671	1
Hyperliquid	●●	●●	●●	●●	●●	∞	∞	○●	45	45

In particular, the term " L1" refers to native blockchain cryptocurrencies such as SOL (the native cryptocurrency of the Solana blockchain), BTC (the Bitcoin cryptocurrency), and ETH (the native cryptocurrency of Ethereum). The term "L2" refers to tokens associated with Layer 2 protocols [132], i.e. blockchain protocols built upon existing blockchain platforms, e.g., Ethereum, (*base layer*) to address scaling issues or high transaction cost in the base layer. Whereas we denote with the term "Meme" those crypto tokens like Dogecoin and Pepe that are created as ERC-20 tokens and that are inspired by Internet jokes, memes, or cultural references. The label "DeFi" includes governance and utility tokens of decentralized finance protocols, such as lending platforms, decentralized exchanges, liquid-staking protocols, or yield aggregators, including AAVE [1], LDO [8], and UNI [295].

The category "Gaming" collects tokens used as in-game currencies or platform tokens within blockchain-based games and metaverse environments. The category "Forex" comprises instruments whose payoff is linked to traditional foreign-exchange rates (e.g., exposure to TRY/USD or EUR/USD). The label "RWA" (Real-World Assets) denotes tokens that represent on-chain claims on off-chain assets. Finally, the category "Stable" includes cryptocurrencies designed to maintain a relatively stable value with respect to an external reference (typically a fiat currency such as the U.S. dollar).

Table 9.3 highlights the heterogeneous design choices adopted by decentralized perpetual protocols in terms of supported underlyings and eligible collateral. Jupiter Exchange lists a small set of Layer 1 assets (SOL, ETH, and wBTC) as both underlying assets and primary collateral, while additionally accepting the stablecoins USDC and USDT as collateral only. Drift Trade and GMX list a comparable range of underlying assets across most categories in Table 9.3, but differ in their collateral design: Drift restricts collateral to SOL and USDC, whereas GMX allows any listed underlying asset to be posted as margin. Hyperliquid follows a collateral policy that is analogous to GMX. dYdX as a unique behavour with the largest number and cateogories of underlying, support also forex like TRY-USD but with a very low volume negotied, similarly RWA, support PAXG-USD but without a high volume of negotioan. On the other hand dYdX support only USDC as collateral. dYdX supports by far the broadest universe of underlying assets, spanning all categories, including less standard forex pairs such as TRY–USD (Turkish lira versus U.S. dollar) and real-world-asset tokens such as PAXG–USD (a gold-backed token paired with the U.S. dollar). These instruments typically concentrate limited trading activity compared with major crypto pairs, but illustrate the long-tail nature of the platform's listing policy. Despite this extensive underlying asset coverage, dYdX accepts only USDC

as collateral.

The differences between the types of assets listed as underlyings and those accepted as collateral are informative about each protocol's risk-management policy. In most cases, protocols restrict collateral to assets that are highly liquid and widely recognized within the ecosystem.

Protocols also differ in maximum leverage offered and policies on maintenance margin, as well as in how they compute borrowing rates and other fees. Tables 9.4 and 9.5 respectively summarize protocol-specific settings for these specifications. Interestingly, we observe that maximum leverage varies from 3X to above 100X, implying large differences in risk exposure for users. Fee management reports similarities, with most protocols imposing both one-off fees upon opening/closing positions and ongoing hourly fees to implement funding mechanisms.

Table 9.4: Maintenance-margin rules and maximum leverage offered by leading perpetual protocols

Platform	Maintenance-margin policy	Maximum leverage
Jupiter Exchange	Maintenance margin is computed from the net exposure after fees: $\text{Mm} = \text{price} \pm \frac{\lvert \text{C} - \text{close fee} - \text{borrow fee} - \text{size/L} \rvert \times \text{price}}{\text{size}}$, with "+" for shorts and "−" for longs.	100 × on Solana; 150 × on Ethereum and Wrapped Bitcoin.
Drift	Asset-specific schedule: maintenance margin ranges from 3 % (most liquid) to 16.67 % (least liquid).	101 × on SOL, ETH, BTC; 3–10 × on other tokens.
GMX	Not disclosed in the public documentation.	100 × on major tokens; 50 × on others.
dYdX	Asset-dependent range between 0.05 % and 10 %.	50 × on BTC and ETH; 20 × on SOL; 5–10 × on other tokens.
Hyperliquid	Maintenance margin equals one-half of the initial margin at max leverage. With max leverage from 3 × to 40 ×, this yields 16.7 % to 1.25 %.	3–40 ×, depending on the asset.

Table 9.5: Fee structure of leading perpetual protocols

Platform	Ongoing hourly fees	One-off fees (open/close)
Jupiter Exchange	Borrow fee only: hourly borrow fee = utilization ratio × hourly borrow rate × position size. No funding fee.	Flat 0.06 % of notional each time a position is opened or closed; extra *price-impact fee* when the trade moves the pool price.
Drift	Funding fee: every hour the trader pays/receives one-twenty-fourth of the percentage gap between the 1-h EWMA mid-price and the oracle reference price.	Market-specific base fee 0.03–0.10 %; additional "insurance / borrow" surcharge 0.75–5.0 %.
GMX	Single hourly charge that combines funding and borrow components (formula in protocol docs).	Base fee 0.04–0.06 % of notional, discounted by volume tiers; *price-impact fee* for large orders.
dYdX	Funding premium: $(\max(0, \text{Bid} - \text{Index}) - \max(0, \text{Index} - \text{Ask}))/\text{Index}$.	Base fee 0.025–0.05 % with volume tiers; ongoing trader-rewards programme can offset part of the cost.
Hyperliquid	Hourly funding payment: funding rate = average premium index + clamp(interest rate − premium, $-0.0005, 0.0005$).	Maker–taker tier schedule: 0–0.045 % of notional.

9.5.2 Options

Options contracts allow holders to buy or sell an underlying asset at a specified price, known as the strike price, granting the possibility, but not the obligation, to exercise

a their right within a specific period. Typically, this price is decided at purchase time and remains constant throughout the contract's lifespan. These instruments can be categorized into two types: expiring options, which are similar to those in traditional finance, and everlasting options, which introduce a perpetual mechanism without a predetermined expiration date. Below, we formally characterize these two types of options, highlighting their structural similarities and differences.

Expiring Options

In our framework, an expiring option contract is a tuple:

$$\langle U, C, L, S, P_{\text{strike}}, T_{\text{expiry}} \rangle$$

where U is the underlying asset, C is the collateral, and L is the leverage, with the same meanings as in perpetual futures.

The price P_{strike} is the predetermined price at which the contract holder has the right to buy or sell the underlying asset U. In expiring options, the strike price is an essential parameter used to determine the option's intrinsic value at the expiration time.

Options in traditional finance can be very complex financial instruments. In the nascent crypto ecosystem, we have mainly two possible strategies (S) for these contracts, called *call* and *put*:

- A *call option* grants the holder the right to buy the underlying asset U at the strike price P_{strike} on or before the expiration date.
- A *put option*, on the other hand, grants the holder the right to sell U at the price P_{strike} upon expiration.

The expiration date T_{expiry} is the limit upon which the holder needs to buy or sell the underlying asset U according to the strategy S adopted.

Everlasting Options

Everlasting options are a novel class of derivatives that allow traders to maintain an option position indefinitely without the need for contract expiration or manual rollovers. Their structure is similar to the one of expiring options but with a key distinction: as they lack an expiration date, they derive their value from an embedded funding mechanism, which balances positions between long and short participants.

Formally, in our framework an everlasting option contract is a tuple:

$$\langle U, C, L, S, P_{\text{strike}}, F_m \rangle$$

where the first four elements of the tuple retain the same meaning as previously described for expiring options, while the last element – the strike price P_{strike} – acts as a reference for determining the option's intrinsic value and remains constant throughout the option's lifetime.[5]

[5]We note that a thorough analysis of the pricing of options lies beyond the scope of this study. For a detailed treatment of pricing methodologies for perpetual options, we refer the reader to the existing literature [248].

Table 9.6: Overview of top Option protocols - underlying and collateral used

Protocol	Underlying	Collateral	Option Type
Derive	ETH, BTC	USDC	Expiring
Hegic	ETH, BTC	USDC.e	Expiring
SOFA.org	ETH, BTC	USDT, crvUSD, stETH, RCH	Expiring
Deri V4	SOL, ETH, BTC, BNB, TON, SUI	USDC, ETH, USDT, WBTC, DAI, ARB, LINK, DERI, LUSD, wstETH	Everlasting
GammaSwap	weETH, WETH, WBTC, PENDLE, PEPE, ARB, GS	WETH, USDC, USDC.e, USDo++, WBTC	Everlasting

Table 9.6 reports the underlying assets and the collaterals supported by the options protocols we consider in this study. Compared to perpetuals, options are limited to a smaller number of cryptoassets, primarily stablecoins such as USDC and USDT, or major native cryptoassets like ETH and BTC. Additionally, protocols offering expiring options focus mainly on these two underlyings, ETH and BTC. In contrast, perpetual contracts include a wider variety of underlying assets and different types of collateral.

9.5.3 Synthetics

Synthetic assets are digital financial instruments created through tokenization, namely the process of representing real-world or financial assets as blockchain-based tokens. These assets can replicate the value and performance of financial instruments such as stocks, commodities, currencies, or other cryptocurrencies without requiring direct ownership of the underlying assets. Additionally, they can provide exposure to a cryptoasset through another cryptoasset.

In our framework, a synthetic asset is formally a tuple:

$$\langle U, C, P_{\mathrm{U}}, P_{\mathrm{C}} \rangle$$

where U is the underlying asset to track, C is the collateral deposited to facilitate the minting of the synthetic token, P_{U} is the price of U, which is continuously tracked, typically through an oracle, to ensure an accurate valuation. The last element of the tuple P_{C} represents the price of the collateral.

Table 9.7 reports the underlying assets and the collaterals supported by the synthetic protocols analyzed. Synthetix V3 supports assets such as sUSD, sBTC, sETH, and sLINK. Legacy assets, including sAMZN, sTSLA, and sAAPL (synthetic representations of Amazon, Tesla, and Apple stock), are no longer actively used and, for this reason, are not on our table. A comprehensive list is available on Etherscan.[6]

9.5.4 Structural Comparison of DeFi Derivative Cryptoassets

After introducing a formal notation for each individual derivative asset category, we compare them along four dimensions, by considering (i) whether they have a fixed expi-

[6] https://etherscan.io/tokens/label/synthetix?subcatid=3-0&size=7&start=0&col=3&order=desc

Table 9.7: Overview of top Synthetic protocols - underlying and collateral used

Protocol	Underlying	Collateral
Synthetix	sUSD, sETH, sBTC, sLINK	Layer 1, Layer 2, DeFi
Alchemix	alUSD, alETH	ETH, WETH, DAI, USDC, USDT, FRAX
Youves	uUSD, uBTC, uXTZ, uXAU, uDEFI	TEZ, tzBTC, SIRS
Metronome Synth	msUSD, msETH, msBTC, msOP	ETH, WBTC, DAI, USDC, FRAX, sfrxETH, vaETH, vaUSDC, vaFRAX, vacbETH, varETH, vastETH, OP, vaOP
Taiga	tDOT	DOT, LDOT

ration, (ii) how they behave with respect to the underlying price, (iii) whether transfers occur during the life of a position, and (iv) how traded prices are aligned to a target.

The presence or absence of T_{expiry} separates contracts that end at a specific date from those that can be held over time: expiring options include T_{expiry}, while perpetuals and everlasting options do not. When P_{strike} is present, as in options, the rule depends on how the underlying price compares to that reference; when it is absent—as in perpetuals and synthetics—the rule depends on the evolution of the underlying price itself or on the pair (P_U, P_C). Transfers governed by F_m occur during the life of the position; this holds for perpetuals and everlasting options, whereas expiring options and synthetics proceed without such transfers. The role of L and S is shared by perpetuals and options, which capture directional exposure with adjustable size, while synthetics omit L and S and are instead created and later redeemed against collateral using (P_U, P_C). Entry and exit follow the same logic: positions with F_m may experience intermediate transfers over time, whereas expiring options concentrate the final rule at T_{expiry}, and synthetics focus on mint and redeem operations using the reference prices defined above.

This alignment carries forward to the next section, where the same fields drive the sequence of actions and data paths for opening, managing, and closing positions.

9.6 Economic Agents and Components of Decentralized Derivatives Protocols

So far, we have provided a static description of the main assets of the various derivative contracts.

After formalizing the main DeFi derivative contracts, we introduce a conceptual model that synthesizes the core entities that participate in the protocols' activity and captures the operational dynamics and the exchange models that govern trade execution. Our conceptualization highlights that perpetual and option protocols involve similar entities and exhibit similar dynamics; thus, we first describe them together. Then, we describe the specific peculiarities of synthetic protocols.

9.6.1 Entities involved in Perpetual and Option Protocols

Upon the creation and trading of a derivative, three key economic agents (liquidity providers, traders, and liquidators or keepers)[7] interact with the derivative protocol, and

[7]Governance users, i.e. holders of governance tokens that grant voting rights and decision-making power, also play a crucial role in the protocol activity. However, a description of their role goes beyond our goal. For further details, see e.g. [148, 185].

Table 9.8: Summary of the entities and mechanisms involved in derivatives protocols.

Entity	Definition
Trader	A user who opens, manages, and closes positions, often using leverage and collateral.
Liquidity Provider (LP)	A user supplying capital to a protocol in exchange for rewards and LP tokens.
Liquidity Pool	A smart contract holding pooled assets to support leverage and trading activity.
Oracle	An entity that provides price data to smart contracts.
Clearinghouse	A component of a derivatives protocol that enforces margin requirements and triggers liquidations.
Matching engines	Matching engines specify how orders are paired with a counterparty and then finalized in protocol state. The most common implementations include order book systems and Pool-counterparty.

several protocol components (liquidity pools, oracles, matching engine, clearinghouse) are involved. Table 9.2 summarizes the entities involved in derivatives protocols.

Liquidity providers (LPs) are users that supply assets to a *liquidity pool*, i.e. one or more smart contracts utilized to store reserves of crypto-assets. The liquidity can be used by the protocol for several purposes, such as to enable borrowing for leverage-based financial activity. In return for providing liquidity, an LP receives a Liquidity Pool token, representing the share of the pool's total assets, and allowing an LP to claim a proportional share of trading fees and other incentives generated by the pool. The specific reward strategies for liquidity provision can vary across protocols.

Traders are instead users who hold a perpetual future or option position. Typically, traders can open leveraged positions by providing collateral and borrowing the remaining liquidity from the protocol through its liquidity pool.

To create and trade a position, a way to price assets is necessary. *Oracles* are computer systems external to blockchain platforms that provide off-chain data to smart contracts running there. They are an essential component for derivative protocols that rely on external data sources such as real-world asset prices and market indices.[8] Notably, oracles may differ in their trust models [22]: centralized ones depend on a single entity, ensuring efficiency but posing risks such as single points of failure and data manipulation; decentralized oracles, like Chainlink [3] and Band Protocol [2], aggregate instead data obtained independently by multiple nodes, enhancing security and trustworthiness.

Matching engines specify how orders are paired with a counterparty and then finalized in protocol state. We identify two primary models: in the *order book* model, trades are matched on a continuous limit order book (on- or off-chain), with price discovery endogenous to the book. A continuous ledger of bids and asks enables price discovery via direct market interaction. In contrast, the *pool-counterparty* (peer-to-pool) matching model eliminates the order book and enables trading against liquidity pools. Execution

[8]Note that oracles can also derive data from other smart contracts or DEXs. In this case, they are known as on-chain oracles. Off-chain oracles are instead those that pull data from external sources, such as traditional financial markets or web APIs.

prices are typically derived from one or more external oracles and adjusted by price impact proportional to the trade size and fee functions of pool utilization; liquidity conditions therefore affect price impact and fees.

Finally, the clearing function is executed by smart contracts or by modules at the protocol level that together act as an on-chain clearinghouse. These components compute profit and loss using current market prices, determine the required initial and maintenance margins, and check whether an account can be liquidated based on the account's balances and positions. The system repeatedly compares the account's equity (collateral plus unrealized PnL) with the maintenance requirement as prices update; when equity nears this level, traders must add collateral or cut exposure, otherwise, they incur into liquidations. These do not require centralized approval: external agents can call the contract's liquidation function once the on-chain condition is met.[9]

Table 9.9: Matching engine, counterparty structure, and oracle design across selected derivatives protocols

Protocol	**Matching engine**	**Liquidity Pool**	**Oracle**
Jupiter Exchange	Pool–counterparty	Single index pool (SOL, ETH, WBTC, USDC, USDT); pool PnL and a share of fees accrue to LP	Edge (primary); Chainlink and Pyth for verification/backup
Drift	Hybrid	No single counterparty pool.	Pyth
GMX	Pool–counterparty	Per-market GM pools; optional GLV vaults allocate across markets; long/short tokens back positions	Chainlink
dYdX v4	Order book	No counterparty pool	Chain-native validator-aggregated oracle
Hyperliquid	Order book	No counterparty pool	Weighted median of major venues (Binance, OKX, Bybit, Kraken, KuCoin, Gate.io, MEXC, Hyperliquid)
Derive	Order book	No counterparty pool	Block Scholes oracle
Hegic	Pool–counterparty	Single-asset pools (ETH, WBTC)	Chainlink
SOFA.org	Market-makers	No counterparty pool	Chainlink
Deri V4	Pool–counterparty	Unified cross-chain pool	Oraclum and Pyth
GammaSwap	Pool–counterparty	Borrows LP tokens from external Pools	Oracleless

[9]Maintenance requirements apply to margined instruments such as perpetual futures and everlasting options; fully paid, non-margined traditional options are generally not subject to maintenance tests.

9.6.2 Dynamics of Perpetual and Option Protocols

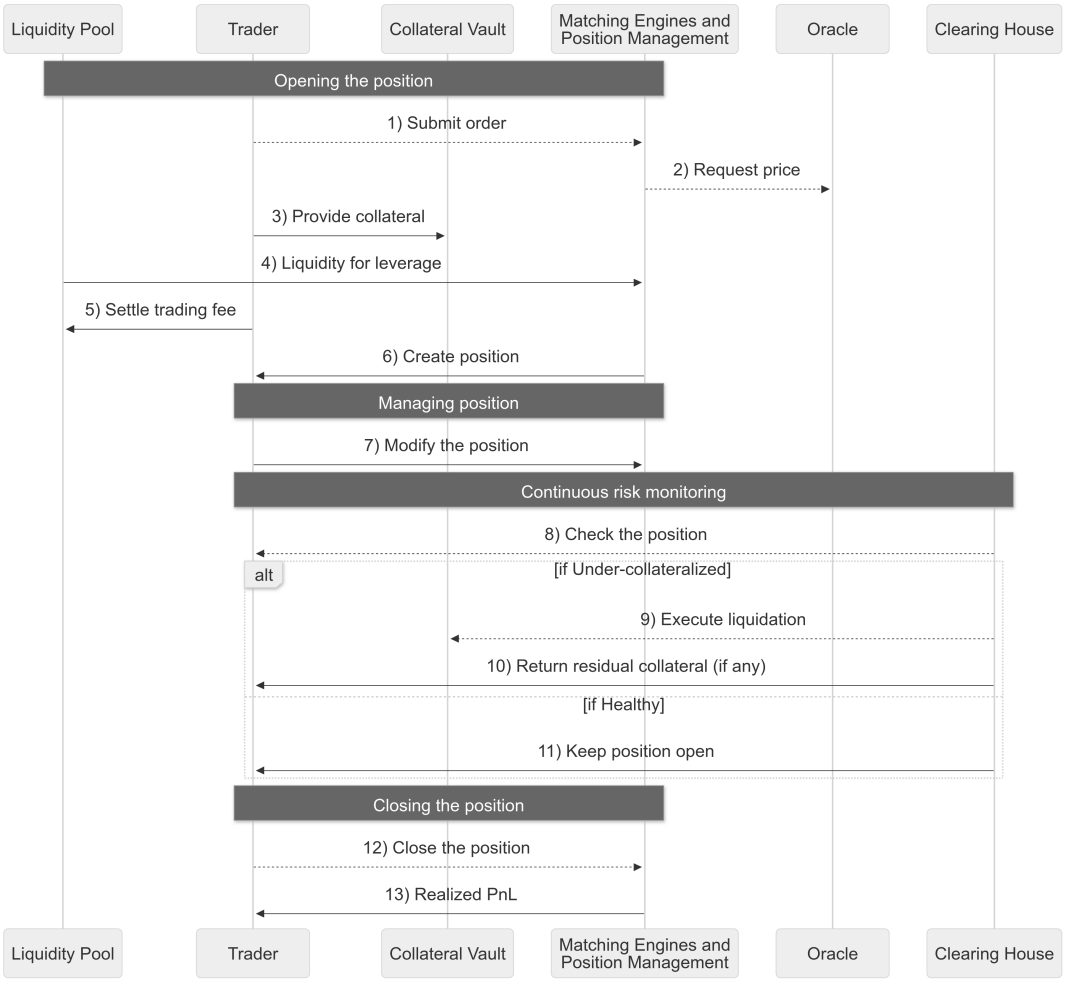


Figure 9.2: Sequence diagram of the trading workflow in perpetual and option protocols. Grey boxes represent the main entities, while dark grey boxes denote the main actions, further decomposed into sub-actions. The 'alt' box illustrates alternative execution paths. Dashed arrows indicate informational interactions without monetary transfers, solid arrows represent value transfers, and bidirectional arrows denote reciprocal interactions.

We now provide an overview of the operational dynamics of perpetual and option derivatives protocols.

A typical derivatives protocol exhibits two main operational flows. The first is the *liquidity provision flow* that captures the process through which liquidity providers supply and withdraw capital. The second is the *trading flow*, shown in Figure 9.2, which encompasses the sequence of actions a trader follows to open, manage, and close positions. The following subsections describe the exchange models and these operational flows in detail. Formal definitions and further details are provided in Appendix 11.

Liquidity provision flow

The liquidity flow specifies how a protocol governs the supply, utilization, and withdrawal of assets, and clarifies the structure of liquidity provisioning as well as the role of liquidity providers (LPs) in market operations.

LPs contribute assets to a liquidity pool and receive LP tokens that represent a claim on the pool's net asset value. Two accounting designs are common. For reward-bearing tokens, the LP token's unit value appreciates as fees and rewards accumulate; instead, in the case of rebase tokens, the protocol increases the token supply to reflect increased rewards. For example, in Jupiter [7], appreciation is driven by protocol-generated fees that accumulate to the pool, allowing the LP token's value to increase over time.

Finally, Liquidity withdrawal proceeds via a burn-and-redeem mechanism: LP tokens are burnt and the protocol returns the corresponding share of the pool. Depending on the implementation, withdrawals may be subject to fees or limits to preserve pool stability and reduce liquidity shocks.

Trading flow

This second mechanism describes how a position is opened, managed, and closed within a protocol. We decompose this flow into three phases: (i) *opening the position*, (ii) *managing the position*, and (iii) *closing the position*. Below, we detail these phases and analyze how trading operations interact with the protocol's infrastructure.

Opening the position A trader initiates a position by submitting an order through the protocol's matching engine (Action 1, Fig. 9.2). The protocol then queries the oracle for the current reference price of the underlying asset (Action 2). Next, the trader posts collateral (Action 3). If leverage is applied, additional liquidity is sourced from the liquidity pool, incurring the corresponding fees (Actions 4 and 5). The order is executed and the position is opened once the protocol's validation and risk checks succeed (Action 6).

At entry, the position's economic exposure depends on posted collateral, leverage, and contract specifications. The notional value NV serves as the reference base for several computations (trading fees, borrowing costs, and funding transfers) and updates with size adjustments over the life of the position. As shown in Fig. 9.2, leverage is provided by a liquidity pool (Action 4), and the trader pays a borrowing fee (e.g., the Gauntlet model used in Jupiter [6]). Trading fees (Action 5) are typically assessed on NV and charged at entry.

Managing the position Once opened, the trader may add or withdraw collateral, increase or reduce exposure, or partially close the position (Action 7). The clearinghouse monitors margin on a continuous basis (Action 8). If account equity falls to or below the maintenance requirement, liquidation becomes eligible to be invoked (Action 9); any residual collateral, net of costs, is returned when available (Action 10). Note that some designs implement partial liquidations; we abstract from these details here.

Positions are marked to market as prices update. Unrealized PnL is the profit or loss that would result if the position were closed at the current mark price; it is computed from the signed exposure and the difference between the current mark and the entry mark, and it is reported net of accrued fees and transfers (such as execution costs,

funding payments, and borrowing charges). Before full close, equity evolves with mark-to-market changes (unrealized PnL), periodic transfers (e.g., funding), borrowing costs, one-off execution fees, and cash movements due to collateral deposits/withdrawals.

Closing the position A position can be closed manually by the trader or automatically via stop-loss (SL) or take-profit (TP) triggers that execute when predefined price levels are reached (Actions 12–13). Realized PnL is determined at the time of closure and immediately booked to the account's equity.

9.6.3 Synthetic Protocols

Synthetic protocols generate exposure by minting a synthetic asset against posted collateral, thereby creating a position that must remain over-collateralized. While they share key components with perpetual and option protocols, such as the oracle and the clearinghouse, their interactions are structured around collateralized debt positions rather than margin accounts. Figure 9.3 illustrates the life cycle: minting (submit, price, post collateral, credit the synthetic), continuous risk monitoring (collateralization checks and potential liquidation), and redemption (repay/burn and unlock collateral). Further details on notation and formulas are provided in Appendix 11.

Minting An investor submits a mint request to the position-management module (Action 1, Fig. 9.3). The protocol then queries the oracle for reference prices of both the collateral and the target synthetic asset (Action 2). The investor posts collateral to the collateral vault (Action 3). Once risk constraints are validated, the protocol mints the synthetic asset and credits it to the investor (Action 4). Minting defines both the number of synthetic units issued and the corresponding debt, valued at the synthetic's reference price; together, these determine the position's economic exposure.

Continuous risk monitoring After issuance, the clearinghouse continuously tracks each position's collateralization ratio (CR), the value of posted collateral relative to outstanding debt, based on oracle price updates (Action 5). Protocols enforce a minimum collateral ratio to preserve solvency. When a position approaches this threshold, the investor can restore its safety by adding collateral, reducing exposure through partial repayment or burn, or both. If the CR falls below the minimum, the liquidation is triggered: the protocol (or external liquidators, where applicable) burns synthetic units to retire debt and sells collateral to cover shortfalls (Action 6), returning any remaining collateral after costs (Action 7). Positions that remain solvent continue unchanged (Action 8).

Redemption To close a healthy position, the investor repays the debt by burning the synthetic asset (Action 9). Once the debt is extinguished, the protocol unlocks the collateral (Action 10) and returns it to the investor (Action 11). At redemption, any remaining profit or loss, after redemption fees, stability charges, and execution costs, is realized and reflected in the investor's final proceeds.

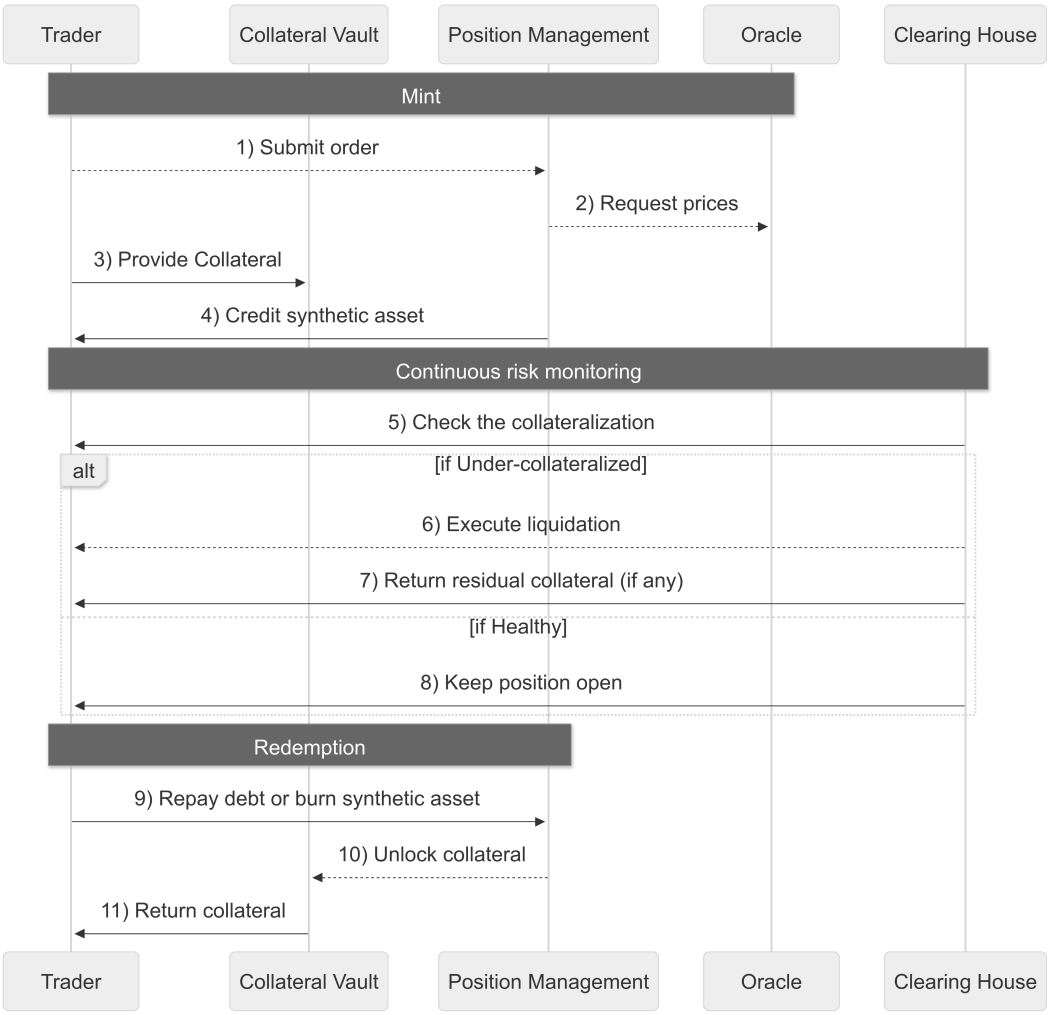


Figure 9.3: Sequence diagram of the trading workflow in synthetic protocols. Grey boxes represent the main entities, while dark grey boxes denote the main actions, further decomposed into subsequent sub-actions. The 'alt' box illustrates alternative execution paths. Dashed arrows indicate informational interactions without monetary transfers, whereas solid arrows represent value transfers.

9.7 Simulation Framework

We complement our systematization of DeFi derivative protocols with numerical simulations that approximate how positions evolve under specific market conditions. Specifically, we develop a Monte Carlo framework in Python that models the life cycle of derivative contracts under stochastic price dynamics, protocol-specific fees, and liquidation mechanisms. All code is publicly available on our GitHub,[10] providing a reproducible benchmark for testing and a foundation for future extensions.

The framework builds on the tuple-based representation introduced in Section 9.5. This unified structure ensures that both analytical and simulation-based components are

[10] https://github.com/LucaPennella/sok-defi-derivatives

governed by the same formal specifications, preserving consistency between theoretical modeling and computational experimentation.

Scope of Application In principle, the framework can simulate perpetuals, everlasting options, and synthetic assets. In this work, however, we focus on perpetual futures, as they represent the most widely used class of derivatives in practice. Their interplay of leverage, margining, and funding mechanisms makes them a comprehensive test for operationalizing our tuple-based formalism. The associated repository also includes prototype simulations for the other contract types.

Stochastic Dynamics In our simulations, the underlying asset price P_t is modeled as a function of time t using a geometric Brownian motion, a standard stochastic process in derivatives pricing [163, 277]:

$$dP_t = \mu P_t dt + \sigma P_t dW_t,$$

where W_t is a standard Brownian motion capturing the random component, μ represents the expected rate of return, and σ denotes volatility.

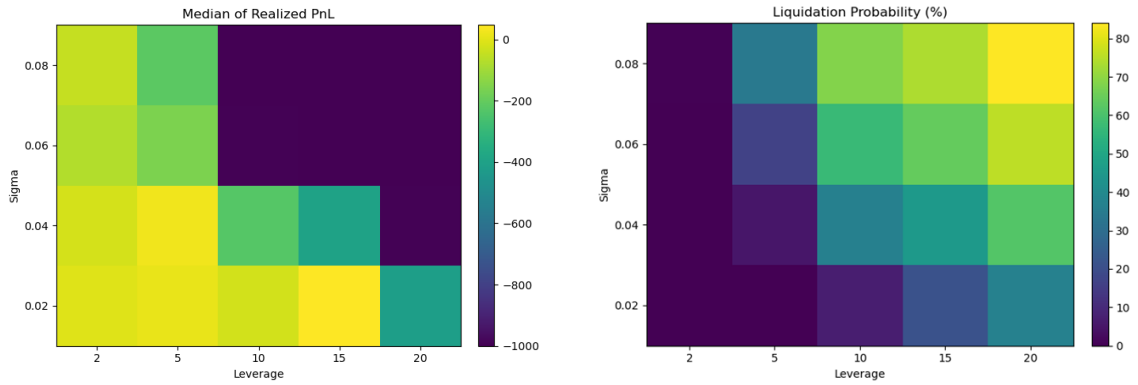
Each simulation run generates a single realization of P_t over a fixed horizon T. Alongside price trajectories, we record the evolution of key quantities such as trader equity, collateral ratios, liquidation events, and terminal profit and loss (PnL). Multiple runs are aggregated to produce summary statistics and visualize outcome distributions, enabling risk assessment and performance evaluation under stochastic conditions.

Perpetual Simulation The simulation pipeline comprises three components:

- Single-path simulations tracing the joint evolution of prices, unrealized PnL, equity, and margin ratios, with explicit inclusion of trading, borrowing, and funding fees.
- Batch simulations over different scenarios in order to estimate distributional outcomes (e.g., realized PnL) and risk metrics (e.g., liquidation probability) over fixed horizons.
- Sensitivity analyses that quantify how leverage, volatility, and transaction costs affect risk and profitability. Heatmaps display the bivariate dependence of realized PnL and liquidation probability on (σ, L), while tornado diagrams rank the univariate impact of each parameter around a baseline configuration..

To derive general insights, we conduct 200 simulation runs with 7-day horizons across parameter grids spanning volatility values $\sigma \in \{0.02, 0.04, 0.06, 0.08\}$ and leverage levels $L \in \{2, 5, 10, 15, 20\}$. Figure 9.4 summarizes the results. Liquidation probability increases monotonically with both parameters, exhibiting a clear transition that separates stable from fragile regimes. Median realized PnL declines with higher volatility and leverage. Together, these surfaces reveal the trade-off between amplified returns and elevated liquidation risk.

To identify the main drivers of short-term risk, we complement the heatmaps with a univariate sensitivity (tornado) analysis around the baseline configuration. For each parameter, we apply symmetric $\pm 20\%$ multiplicative shocks (holding others fixed) and



(a) Median realized PnL over (σ, L).

(b) Liquidation probability over (σ, L).

Figure 9.4: Sensitivity analysis from Monte Carlo simulations of perpetuals. Higher volatility and leverage reduce median PnL and increase liquidation risk.

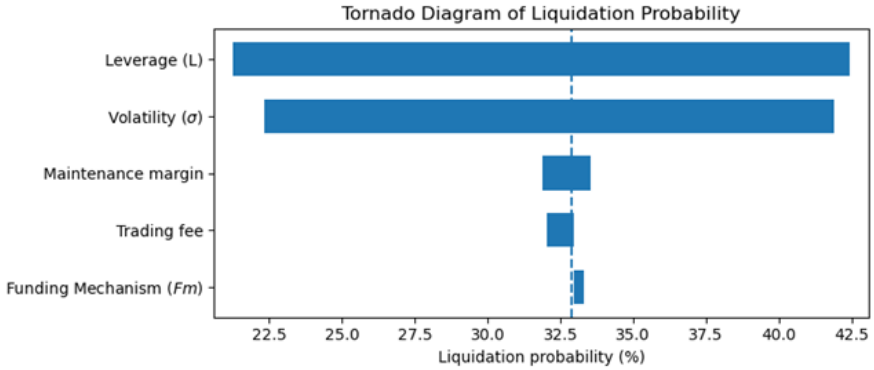


Figure 9.5: Tornado diagram for 7-day liquidation probability. Bars show the change in the metric under symmetric parameter shocks around the baseline (vertical dashed line). Leverage and volatility dominate; fee- and margin-related parameters are second order over this horizon.

measure the corresponding change in 7-day liquidation probability (baseline $\approx 33\%$). The ranking is clear: leverage is the dominant factor, with $\pm 20\%$ shocks shifting liquidation probability by roughly -8 to $+7$ percentage points (p.p.), while volatility (σ) induces slightly larger variations of about -11 to $+9$ p.p. By contrast, fee parameters, the maintenance margin ratio, and funding mechanics exhibit small, local effects, typically below 2 p.p., with funding contributing less than 1 p.p. These directional effects align with economic intuition: leverage tightens the margin buffer, while higher volatility broadens return distributions, pushing positions toward liquidation boundaries.

We further adapt the simulation settings to replicate parameters used by specific protocols such as Jupiter and dYdX, estimating the behavior of representative perpetual futures positions. For illustration, we simulate a long perpetual position on Solana configured with values inspired by Jupiter Exchange. The simulation spans a 7-day horizon across grids of leverage $L \in \{2, 5, 10, 15, 20, 50, 100\}$ and volatility $\sigma \in \{0.02, 0.04, 0.06, 0.08\}$, each repeated 500 times. Key parameters include:

- initial collateral: 1000.0 units,

Table 9.10: Liquidation probability (%) for a long Solana perpetual varying volatility σ and leverage L.

σ	$L = 2$	$L = 5$	$L = 10$	$L = 15$	$L = 20$	$L = 50$	$L = 100$
0.02	0.0	0.0	5.6	20.2	35.0	72.8	85.4
0.04	0.0	3.4	32.4	50.2	61.0	86.0	92.0
0.06	0.0	16.0	53.2	65.2	75.2	88.2	94.8
0.08	0.0	27.4	63.0	75.2	80.8	91.0	94.8

- trading fee rate: 0.0006 (on open and close),
- borrow fee rate: 0.000027 per time step,
- maintenance margin rate: 0.002556,
- slippage on entry: 20 basis points,
- funding fees: excluded in this configuration.

The goal is to compute liquidation probabilities under varying levels of leverage and volatility. Table 9.10 reports the results. At low volatility ($\sigma = 0.02$), liquidation risk remains limited even at moderate leverage, whereas at high volatility ($\sigma = 0.08$) it exceeds 75% for $L \geq 15$ and approaches certainty at extreme leverage. This nonlinear pattern highlights how risk grows disproportionately with volatility and position size.

Encoding perpetual mechanics as tuples ensures that both pathwise equity dynamics and aggregate statistics follow the same underlying specification. In this sense, the tuple serves as an executable formalism, transforming abstract definitions into observable outcomes. These simulations validate the tuple framework as a flexible and reproducible research tool for analyzing DeFi derivatives.

9.8 Conclusions

This contribution conducted the first systematic study and analysis of derivatives protocols in DeFi. While derivatives instruments represent a central component of traditional and centralized financial markets, their decentralized counterparts are comparatively understudied within the academic literature, especially their operational dynamics and protocol structures. This chapter addresses this gap by identifying, analyzing, and formalizing the leading DeFi protocols, categorized as perpetual futures, synthetic assets, and options, into a coherent and general conceptual framework.

Drawing from protocol documentation, user interfaces, deployed smart contracts, and community-driven sources, we provide a formal representation of Decentralized Finance derivatives instruments. The resulting framework captures the main design principles, protocol actors, trading flows, and core architecture that underpin decentralized derivatives protocols. By abstracting the structural components and financial functions across various protocols, our framework provides the tools for systematically comparing them and gaining a better understanding of their role in the broader decentralized financial ecosystem.

We acknowledge that, given the 'early-stage' phase of the DeFi derivatives ecosystem and the limited or fragmented documentation of some protocols, our framework is not without limitations. First, our protocol selection process used TVL to select the protocols to analyze and excluded those lacking accessible code or documentation. Although pragmatic, our approach might have framed the framework only on "top-tier" protocols that are already mature and well-documented, leaving out interesting aspects of newer, smaller, or less documented protocols, which might contain innovative mechanisms. Another limitation of our framework is that it is constrained by the availability of public data, the lack of protocol standardization. Moreover, our analysis is framed within a specific time window on December 31, 2024, and the "top-tier" protocols considered may change due to the rapid pace at which the ecosystem evolves. Nonetheless, we believe our framework offers a solid and pragmatic foundation for analyzing decentralized derivatives protocols, and serves as a starting point for further academic and practical inquiry in the field. As the DeFi derivatives continue to evolve, rigorous and accessible conceptual tools such as those proposed here will become increasingly valuable for researchers, developers, and regulators alike.

Future work could expand our analysis, framework and simulation to additional protocols, develop protocol-specific taxonomies for second-layer features, and study user behavior and market efficiency in real trading environments.

Chapter 10

Conclusions

This thesis set out to develop an empirical and methodological framework at the intersection of explainable machine learning and cryptocurrency markets, with the aim of providing interpretable and reproducible tools to study complex socio-economic phenomena shaped by blockchain technologies. As outlined in Chapter 1, the guiding premise has been that predictive performance alone is insufficient in these domains: models must also be auditable, transparent, and connected to domain-relevant evidence. Across the three parts of the dissertation, this perspective has informed both methodological developments and empirical applications, linking heterogeneous data sources, protocol-level measurements, and social attitudes within a coherent analytical program.

The first set of research questions concerned the role of explainability in high dimensional financial and social data. Chapters 3 and 4 demonstrated that interpretability-oriented pipelines can effectively bridge complex models and substantive insights. The X-SPIDE framework presented in Chapter 3 showed that post-hoc explanations can support forensic reasoning in smart-contract analytics, while Chapter 4 illustrated how SHAP-driven representations reveal value dimensions in survey data that remain informative beyond demographics. Chapters 5 and 6 further strengthened this perspective: the investigation on class overlap in Chapter 5 clarified how data geometry conditions the success of oversampling strategies, and Chapter 6 introduced Decision Predicate Graphs as a structural tool for inspecting tree ensembles when rule-based reasoning becomes impractical. Together, these chapters respond to the question of how explainability can be embedded throughout the modelling pipeline rather than added as an ex-post layer.

The second set of questions addressed the social foundations of cryptocurrency markets. Chapters 7 and 8 provided empirical evidence that investor populations are far from homogeneous. Chapter 7 identified memecoin holders as a recognizable subgroup with distinct psychological and behavioral profiles, while Chapter 8 showed that regulatory attitudes vary systematically with perceived market illegitimacy and personal exposure to crypto wealth. These findings clarify how beliefs and identities mediate the relationship between technological innovation and policy preferences, directly addressing the research questions formulated in Chapter 1 regarding the interaction between individual perceptions and institutional design.

The third set of questions focused on decentralized finance from a protocol-level perspective. Chapters 9 and 10 delivered two complementary contributions. Chapter 9 developed an analytics pipeline for measuring the micro-velocity of Lido's liquid stak-

ing tokens, documenting patterns of token circulation, turnover concentration, and the growing role of wstETH within composable DeFi infrastructures. Chapter 10 proposed a unified formalism for decentralized derivatives protocols and operationalized it through a tuple-based representation and simulation environment. These chapters respond to the question of how DeFi can be studied as a socio-technical system governed by explicit mechanisms rather than as a collection of isolated applications.

Several limitations must be acknowledged. First, as discussed in Chapters 4 and 8, survey-based evidence is subject to measurement error and sampling biases, while on-chain data used in Chapters 3 and 9 capture actions but not intentions. Second, the interpretability techniques employed in Chapters 3, 4, and 6 remain approximations of model behavior and may be sensitive to specification choices and feature engineering decisions. Third, the protocol analyses in Chapters 9 and 10 focus on specific ecosystems and time windows; the rapid evolution of DeFi and the impact of exogenous shocks limit the external validity of some conclusions.

In conclusion, and in line with the vision articulated in Chapter 1, this dissertation has argued that understanding cryptocurrency and DeFi requires an interdisciplinary toolkit in which explainability, social measurement, and protocol analytics are developed jointly rather than in isolation. By connecting the contributions of Chapters 3 to 10 within a unified framework, the thesis provides a more transparent and scientifically grounded account of digital financial systems and lays foundations for both academic inquiry and informed policy debate.

Chapter 11

Appendices

Appendix Chapter 3

A Explainability method

Shapley values assess each feature's impact on model predictions by examining all possible feature combinations. The Shapley value for feature i in instance x, denoted as Shap(x, i), is calculated using the following formula:

$$Shap(x,i) = \sum_{S \subseteq \{1,\dots,p\} \setminus \{i\}} \frac{|S|!(p-|S|-1)!}{p!} [c(x_S \cup \{i\}) - c(x_S)]$$

Here, S represents a feature subset excluding feature i, $|S|$ denotes the size of set S, x_S is the sample x with features restricted to subset S, $c(x_S \cup \{i\})$ is the classifier prediction with feature i included, $c(x_S)$ is the classifier prediction without feature i, and p is the total number of features.

Shapley values offer insights into how each feature contributes to predictions and are considered both global and local models in Explainable Artificial Intelligence.

B Extended Features Analysis

This appendix reports, for each coalition (*Dx/CDx*, *M5S*, *Sx/CSx*), the top-30 features by global importance computed as mean absolute SHAP values on the 2019 hold-out sample. These rankings provide a compact overview of where explanatory mass concentrates and how quickly it decays beyond the leading predictors. Consistent with 5.2, the cumulative importance curve exhibits an elbow around 10–15 features, indicating that the long tail (ranks 11–30) contributes marginal yet non-negligible refinements to class separation. Coalition-wise summaries in 5.2 synthesise these patterns; the detailed top 30 lists below document the full global importance profile for transparency.

C Undecided voters

This appendix focuses on the analysis of those respondents who did not provide a clear voting intention, indicating "undecided" as their answer. The number of respondents who reported this preference is 695. All background variables (socio–demographic and

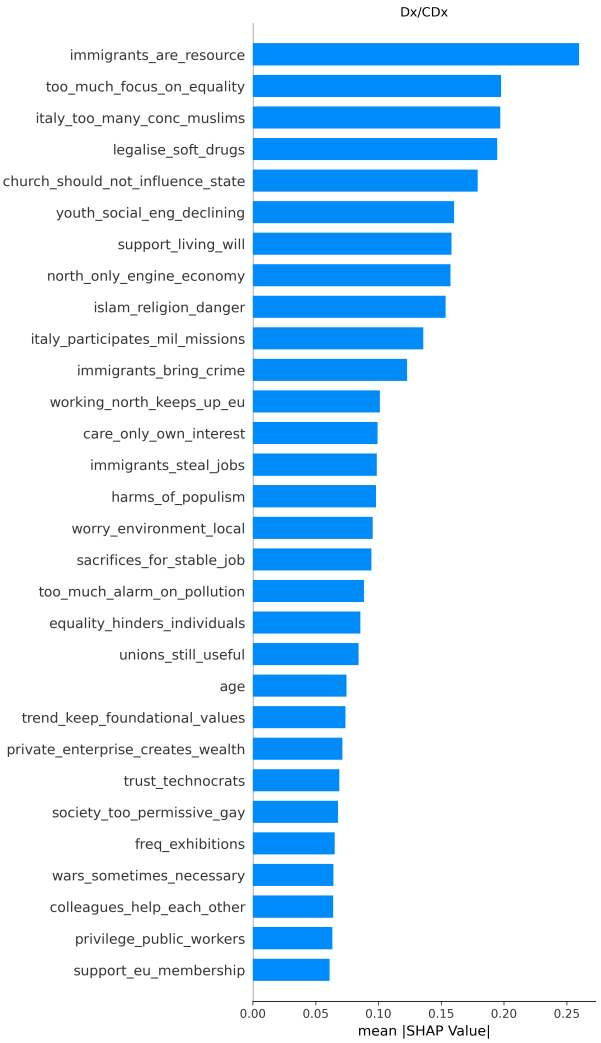


Figure 11.1: Top 30 features by absolute mean SHAP. Global importance ranking for *Dx/CDx* on the 2019 hold-out. The distribution shows an elbow around ranks 10–15; items beyond rank 15 contribute incremental lift

attitudinal) are retained; identification fields and the raw vote label are excluded. Since most variables are categorical, they are transformed via one–hot encoding, while numeric items are standardised.

We apply a k–Nearest Neighbours classifier, with $k = 21$ chosen after a grid search. This choice reflects a compromise between excessive sensitivity to noise (as with $k = 1$) and dilution of profile–specific signals (as with large k).

For each undecided respondent, the algorithm considers the fifteen most similar profiles among those with known vote intention and assigns the party preference.

The great majority of undecided respondents are mapped to the three largest parties: *Sx/CSx* (431 cases, 62%), *Dx/CDx* (186 cases, 26.8%), *M5S* (78 cases, 11.2%). These results indicate that although respondents claim indecision, their profiles align with the electorates of the major parties.

Despite the results there are some fundamental limitations: 1) nearest–neighbor allocation reflects relative profile similarity, not actual behavioural choice; the method

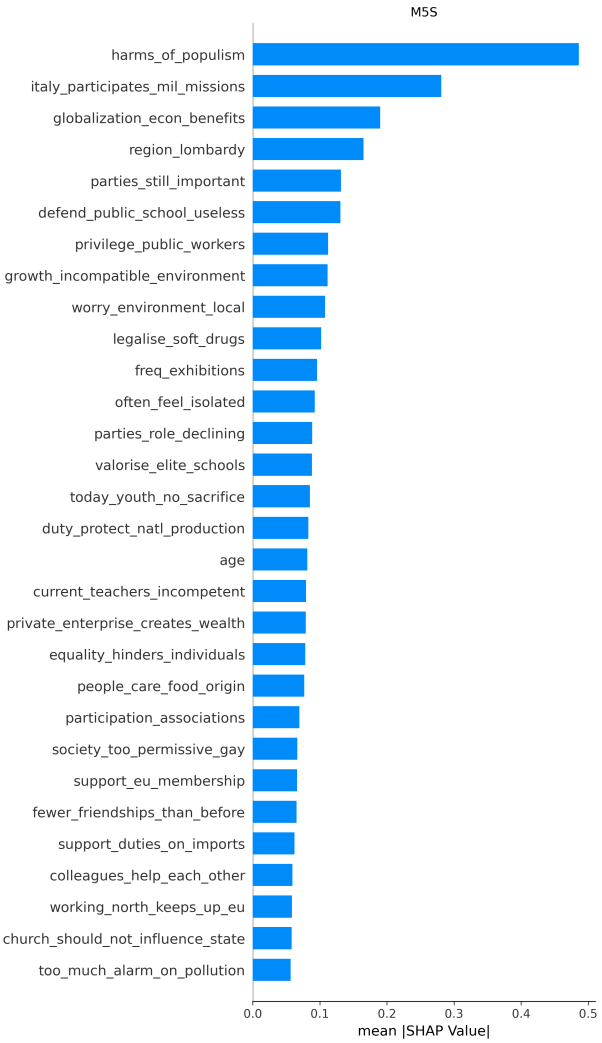


Figure 11.2: Top 30 features by absolute mean SHAP. Global importance ranking for *M5S* on the 2019 hold-out. Top mass concentrates on anti-establishment and policy-trade-off items; the long tail refines class boundaries

captures proximity in attitudinal space rather than forecasting a vote, 2) the imbalanced data, larger parties are over–represented among neighbors and thus more likely to attract undecided cases, 3), the choice of k and of Euclidean distance is conventional; alternative specifications (cosine distance, different k) produce variations in marginal parties but do not alter the dominance of the three main blocs.

D Unsupervised analysis of within-group heterogeneity among *M5S voters*

This appendix documents an unsupervised exploration of heterogeneity among respondents who declare an intention to vote for *M5S*). Categorical survey variables were represented through one–hot encoding, and clustering was performed with K-Means in the high–dimensional feature space. Several values of k were compared using internal criteria (silhouette with cosine distance, Calinski–Harabasz, Davies–Bouldin) together with

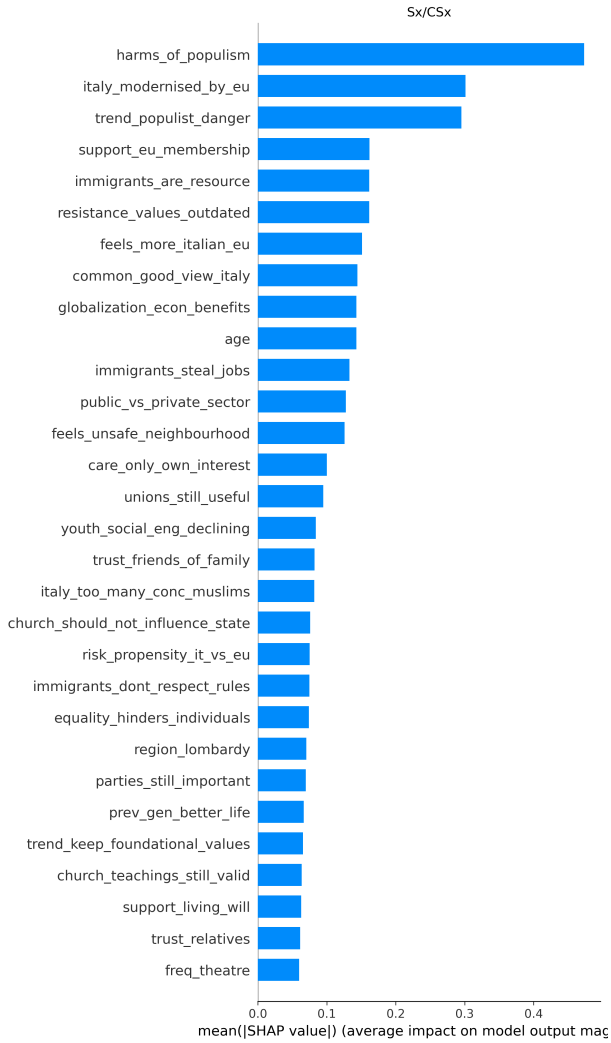


Figure 11.3: Top 30 features by absolute mean SHAP. Global importance ranking for *Sx/CSx* on the 2019 hold-out. Pro-EU and pro-immigration items dominate; identity/security variables mostly appear in the tail.

substantive readability; we retain $k = 3$ as a parsimonious and interpretable partition. Figure 11.4 reports a two-dimensional PCA projection of the feature space with points coloured by cluster.

The projection shows substantial overlap across clusters, which is consistent with very low internal separation (sample cosine silhouette ≈ 0.003) that is typical of sparse, high–dimensional survey data. Despite the overlap, three coherent tendencies emerge. A first subgroup is civic–cultural and technocratic, with stronger orientation to the local common good, higher cultural participation, and greater trust in technocrats. A second subgroup is euro–progressive and sceptical of populism, with more favourable views of EU membership and less alarmist attitudes on Islam and immigration, alongside higher support for civil–rights items. A third subgroup is more eurosceptic and securitarian, with greater concern about immigration and Islam, lower support for the EU, lower cultural participation, and weaker support for selected civil–rights items. These labels are descriptive summaries of the dominant markers within each cluster.

Estimated vote	Count	Percentage (%)
Sx/CSx	431	62.0 %
Dx/CDx	186	26.8 %
M5S	78	11.2 %

Table 11.1: Distribution of the votes after KNN

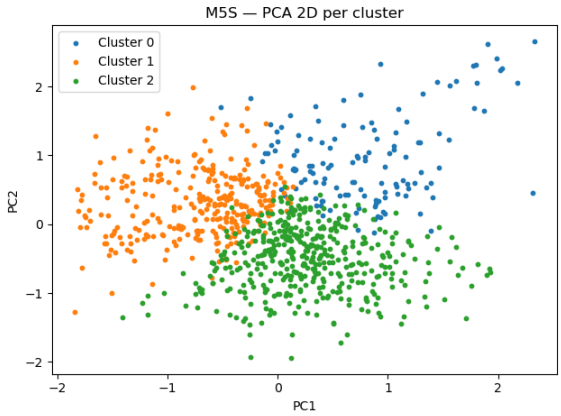


Figure 11.4: Two–dimensional PCA projection of the M5S sub–sample, points coloured by K–Means clusters ($k = 3$).

Two caveats apply. First, the low silhouette indicates that clusters should be interpreted as latent tendencies rather than sharply separated classes. Second, results depend on representation choices (one–hot encoding and PCA projection for visualisation) and on the selected number of clusters; nearby solutions such as $k = 2$ or $k = 4$ emphasise adjacent aggregations of the same underlying attitudinal axes.

Appendix Chapter 6

A A brief history of Memecoins

The history of memecoins reflects the broader evolution of cryptocurrency culture, shaped by the intersection of online humor, speculative investing, and grassroots community participation. The first recognised memecoin, Dogecoin[1], was launched in December 2014 by Billy Markus and Jackson Palmer. Originally created as a parody of Bitcoin, Dogecoin has since gained significant popularity and remains the most prominent memecoin in terms of both cultural relevance and market capitalization.[2] The rise of Ethereum in 2015 introduced a new phase for memecoins. With the ERC-20 token standard, the process of creating new cryptocurrencies became significantly more accessible [101]. During the 2017 Initial Coin Offering (ICO) boom, numerous projects leveraged memetic branding, though few gained long-term traction. One notable example was the Useless

[1] https://foundation.dogecoin.com/

[2] Despite its satirical origins, Dogecoin swiftly amassed a devoted following, particularly on platforms such as Reddit and X, with Elon Musk's 2020 tweet being a notable example of support for its adoption.

Ethereum Token (UET)[3], a satirical project explicitly marketed as having no value or purpose, yet it successfully raised 310 ETH. Between 2016 and 2018, the concept of collectible digital assets intersected with meme culture, laying the foundation for future NFT-based memecoins. The Rare Pepe phenomenon, which originated from the trading of tokenized Pepe the Frog memes on the Counterparty protocol, was one of the earliest examples of meme-related assets with provable scarcity.[4]

Following Dogecoin's rise, new dog-themed memecoins began appearing, often launched on the Ethereum blockchain and incorporating additional technical features. Shiba Inu (SHIB)[5] was released in August 2020 and described itself as the "Dogecoin Killer." Floki Inu (FLOKI)[6], inspired by Elon Musk's tweet about naming his pet dog "Floki," exemplified the growing influence of celebrity endorsements in memecoin valuations. Dogelon Mars (ELON)[7] combined themes from Dogecoin and Musk's association with space exploration and gained visibility after a portion of its supply was sent to Ethereum cofounder Vitalik Buterin, who later donated those tokens to charity. During the 2021 bull run, memecoins became more prominent in online financial discussions. The GameStop (GME) short squeeze, organized by users of the WallStreetBets subreddit[8], brought attention to the collective behavior of retail investors in speculative markets. In the years that followed, particularly after the 2022 bear market, activity around memecoins began to include projects on the Solana blockchain[9]. Solana's network structure, characterized by low transaction costs and high processing speed, enabled efficient deployment of new tokens.

B Additional variables' description

Tables 11.2 and **??** provides additional information about the main variables used in the empirical analysis. The third column presents the categories in which each variable was divided.

C Data Collection and Curation

We collected all the survey data through the platform State of Crypto (https://stateofcrypto.net) with a customized version of the NodeGame framework [42]. We opened the survey to the general public on December 7th 2024: anyone aged 18 (15 with the permission of their legal guardians) or more could join the study; participation was fully voluntary and responses were not incentivized. We closed the survey on February 20th. On average, the survey lasted about 10 minutes.

C.1 Survey Structure

The State of Crypto Survey spanned through the following sections:

[3] See https://uetoken.com/
[4] See https://rarepepes.com/
[5] https://shibatoken.com/
[6] https://floki.com/
[7] https://dogelonmars.com/
[8] https://www.reddit.com/r/wallstreetbets/
[9] See https://www.coingecko.com/en/categories/solana-meme-coins

Table 11.2: Additional variables description and categories

Socio-economic and risk profile		
Variable	Code	Categories
Place of residence	*Country*	Multiple countries.
Race	*Race*	Black or of African descent, East Asian, Hispanic or Latino, Middle Eastern, Multiracial, Native American, Pacific Islander, or Indigenous Australian, South Asian, Southeast Asian, White / Caucasian, Not listed.
Gender	*Gender*	Being male = 1, rest = 0.
Age	*Age*	14 or less; 15-17; 18-21; 22-25; 26-29; 30-33; 34-37; 38-41; 42-45; 46-49; 50-53; 54-57; 58-61; 62-65; 66-69; 70-73; 74-77; 78 or more.
Education	*Edu*	Less than high school degree; High school degree; Vocational education and training; Bachelor degree (e.g., BA, BS); Master degree (e.g., MA, MS); Doctorate (e.g., PhD, EdD) or professional degree (e.g., Med).
Work status	*Is working*	Self-employed; Part-time employed; Full-time employed; Retired; Unemployed: Looking for work; Unemployed: Unable to work due to sickness or ill health; Not working and not looking for work.
Working in crypto	*Works in crypto*	Main occupation; Side job; No.
Cryptocurrency literacy	*Know crypto general*	Index that measures the level of crypto literacy (maximum achievable score: 5).
Memecoins literacy	*Know Meme*	Index that measures the level of Memecoin literacy.
NFT knowledge	*Know crypto NFT*	Index that measures the level of NFT literacy.
Financial literacy	*Know finance*	Index that measures the level of financial literacy.
Crypto trader	Crypto trader	Yes = 1, no = 0.
Risk seeking	*Risk seeking*	Whether a participant chose to open more than 50 boxes in the Bomb risk elicitation task [100].

1. *Entry*: age-check, informed consent requested and GDPR informative, bot captcha.
2. *Demographics*: info about participant and relationship to crypto.
3. *Quiz*: knowledge about crypto and traditional financial.
4. *Behavioral*: risk preferences.
5. *Investing*: portfolio, investing behavior and confidence.
6. *Society*: inequality, trust in governance, carbon emissions.
7. *Privacy*: regulations, mass control, taxes.

The full list of questions is available in the anonymized OSF repository: https://osf.io/96zvy?view_only=82ab1b433b6e4759a3a510e21f3897ef.

C.2 Ethical and legal clearance

Prior to launching our survey, we preregistered our research at AsPredicted.org, with a copy available on OSF (https://osf.io/57jdb?view_only=82ab1b433b6e4759a3a510e21f3897ef). Most importantly, we obtained ethical review clearance from the Ethics Committee of the University of Mannheim (EK Mannheim 40/2023) and worked thoroughly with the Data Protection Team of the University of Mannheim to verify the compliance of our methods and infrastructure with the GDPR regulations.

C.3 Recruitment

We recruited participants by promoting the survey with paid ads on social media, mostly on Twitter (now X) via the account @stateof_crypto. In addition, we announced the survey on several other avenues with the following activities:

- direct mails targeted to influential academic and personalities in the crypto space,
- direct mails to influencers in the Financial Independence Retire Early (FIRE) community,
- direct mails to past participants of crypto conferences and events,
- messages in several Telegram and Discord channels about crypto and investing,
- collaborations with local influencers for African communities,
- messages in researchers' and translators' own social networks.

C.4 Data quality

3,041 respondents started the survey, 2,219 passed the initial screening (confirmed their age and solved a captcha) and gave their informed consent for participation, and 1,414 completed it (being the preregistration target 2,000). We followed anBd improved upon the data quality procedures in [45], as we outlined below.

A common technique to improve data quality in surveys are attention checks—also called "instructional manipulation checks" [238]. These checks include trap questions with nonsensical items, e.g., 30th February, or instructed response items, e.g., "Please select the third item to this question." However, they are known to have side effects [43], therefore we decided against their use for two main reasons. First, they degrade user experience and may generate psychological reactance [62]; given that our survey was not incentivized, their use might have significantly increased the dropout rate, reducing power and increasing Type II error [14]. Second, attention checks may interfere with research hypotheses and do not rule out confounding variables [154]. Therefore, we opted for the post-collection quality checks described below.

To prevent bot access we relied on H-Captcha, a captcha-based professional service that proved reliable in the previous edition of the State of Crypto Survey [45]; consequently, we did not use "honeypots" in favor of a leaner and more accessible setup.

Finally, we monitored response times throughout the survey. As preregistered, we excluded respondents whose completion times for either the overall survey or any section were faster than two standard deviations below the mean; to prevent undue influence of extreme delays, cases taking longer than one hour were removed from the standard-deviation computation (0.05% of cases). In addition, we took the conservative approach to removed all participants that speeded out through the survey with an average time per question below three seconds.

D Robustness checks

Herein we report the results of our robustness checks, including different model specifications.

D.1 Geolocation

We geolocated the IP address of all respondents and compared it with their self-reported country of residence. Three participants claimed to be from Antarctica; this is highly unlikely and furthermore their IP geo-location did not match their answers. Because it signals low-effort/spam behavior, we removed these three observations from all analyses.

As a robustness check, we re-ran our baseline regression, removing 319 observations for which the country of residence and IP do not match (see Table D1), and find that all our main results substantially hold.

Table D1: Robustness checks for Geolocation: removed obs with mismatched IP.

	(1)	(2)	(3)	(4)
Gender	0.097 (0.059)	0.089 (0.061)	0.063 (0.035)+	
Age	0.011 (0.006)+			
Edu	−0.006 (0.009)	0.000 (0.009)	0.000 (0.007)	
Know Meme	0.136 (0.035)***	0.137 (0.042)**	0.122 (0.034)***	0.137 (0.031)***
Know crypto no Meme	0.003 (0.020)			
Know fin.	−0.044 (0.024)+	−0.037 (0.022)+		
Crypto trader	0.191 (0.086)*	0.174 (0.073)*	0.136 (0.041)***	0.126 (0.048)**
Risk Seeking	0.020 (0.016)	0.018 (0.017)		
When interested	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Wealth inv.	0.001 (0.000)***	0.001 (0.000)***	0.002 (0.000)***	0.002 (0.000)***
Leverage	0.150 (0.018)***	0.149 (0.030)***	0.136 (0.027)***	0.152 (0.028)***
Has NFT	0.162 (0.032)***	0.160 (0.032)***	0.144 (0.038)***	0.140 (0.037)***
Num non-memecoins held	0.041 (0.009)***	0.042 (0.009)***	0.045 (0.009)***	0.044 (0.008)***
Need to belong	−0.039 (0.004)***	−0.038 (0.007)***	−0.035 (0.008)***	−0.035 (0.007)***
FOMO	0.008 (0.010)			
KYC	−0.010 (0.011)			
Trust regulations	0.014 (0.011)	0.014 (0.011)	0.011 (0.011)	
Scams	0.003 (0.015)	0.001 (0.016)	−0.005 (0.012)	
Taxes	−0.022 (0.013)+	−0.021 (0.014)	−0.020 (0.012)+	−0.017 (0.010)+
Trust in Gov.	−0.008 (0.013)	−0.019 (0.010)+	−0.008 (0.008)	
Num.Obs.	720	726	831	855
AIC	731.7	735.6	843.3	865.6
BIC	850.7	836.5	937.7	936.8
Log.Lik.	−339.833	−345.810	−401.637	−417.787
RMSE	0.39	0.40	0.40	0.40

D.2 Potentially duplicated responses

Because the survey is public, anyone with the access link could join it. We set a cookie to remember if a survey has already been started, but this alone does not ensure that the same person does not participate multiple times. Being the survey unincentivized, the

risk is small, nonetheless herein we try to quantify multiple participation, by looking for specific indicators.

In particular, we used a two-pronged strategy. First, we looked for respondents who shared the same IP, email, or crypto address. Sharing the same IP address is not alarming; this is common, for example, among students in the same dorm or users of the same Internet Service Provider (ISP) at different physical locations when IPs are dynamically assigned. Sharing the same crypto or email address is more concerning, but it does not, per se, mean that the same person took the survey twice. In fact, especially in developing countries, the same physical computer and the same email address could be shared among members of the same family, relatives, and friends.

Overall, we identified a total of 6.2% of respondents sharing either the same IP address (7.9%), the same email address (2.9%), or the same crypto address (0.4%).[10] Sharing the same IP address is not alarming; this is common, for example, for students in the same dorm or even for users of the same Internet Service Provider (ISP) at different physical locations, when IPs are dynamically assigned. Sharing the same crypto or email address is more concerning, but still does not mean per se that the same person took the survey twice. In fact, specially in developing countries, the same physical computer and the same email address could have been shared across members of the same family, relatives, and friends. Tables D2 and D3 show our baseline regressions for Model 1 and Model 4, removing potential duplicates and our results substantially hold.

Second, to further investigate the issue, we created an index of similarity ranging from zero (fully dissimilar) to one (identical), representing the share of identical answers to the questions in the demographics section of the survey. We computed this index for all pairwise comparisons of respondents not suspected to have taken the survey multiple times (94% of the sample). The average similarity of any two user is about 0.15; considering only those respondents who completed the survey, this number increases to about 0.19. Note that this number indicates a rather low level of similarity between any two users, nonetheless we used it as one of the baselines for the robustness checks. Here all the sets for potentially duplicated responses:

- *Chance*: similarity score above chance (0.19; 77 obs);
- *50*: similarity score above 50% (51 obs);
- *75*: similarity score above 75% (9 obs);

Tables D4 and D5 show our baseline regressions for Model 1 and Model 4, removing potential duplicates and again our results substantially hold.

D.3 Unlikely sets of answers

Some responses or combinations of responses are more unlikely than others. We define the following unlikely sets that we later use in the robustness analyses.

- *Ret50*: retired below age 50 (73 obs);
- *Doc22*: holding a doctorate below age 22 (7 obs);

[10] The sets are overlapping.

Table D2: Robustness checks for IP, email, crypto for the logistic model (1)

	ip	email	crypto
Gender	0.083 (0.049)+	0.101 (0.039)**	0.089 (0.042)*
Age	0.008 (0.006)	0.007 (0.007)	0.007 (0.007)
Edu	−0.012 (0.007)	−0.011 (0.007)	−0.013 (0.009)
Know Meme	0.147 (0.031)***	0.147 (0.029)***	0.147 (0.029)***
Know crypto no Meme	−0.002 (0.018)	−0.008 (0.018)	−0.003 (0.017)
Know fin.	−0.023 (0.031)	−0.022 (0.030)	−0.022 (0.029)
isTrader	0.200 (0.078)*	0.145 (0.075)+	0.176 (0.077)*
Risk Seeking	−0.008 (0.019)	−0.011 (0.018)	−0.005 (0.015)
When interested	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Wealth inv.	0.001 (0.000)**	0.001 (0.000)***	0.001 (0.000)***
Leverage	0.106 (0.028)***	0.120 (0.020)***	0.119 (0.023)***
Has NFT	0.183 (0.033)***	0.196 (0.024)***	0.193 (0.023)***
Num non-memecoins held	0.041 (0.009)***	0.040 (0.007)***	0.039 (0.007)***
Need to belong	−0.039 (0.007)***	−0.038 (0.009)***	−0.036 (0.007)***
FOMO	0.004 (0.010)	0.004 (0.011)	0.002 (0.010)
KYC	0.000 (0.015)	−0.002 (0.015)	−0.004 (0.013)
Trust regulations	0.016 (0.013)	0.015 (0.011)	0.016 (0.010)
Scams	0.006 (0.018)	0.006 (0.017)	0.003 (0.016)
Taxes	−0.025 (0.009)**	−0.025 (0.010)*	−0.023 (0.010)*
Trust in Gov.	−0.016 (0.023)	−0.020 (0.020)	−0.019 (0.019)
Num. Obs.	804	828	849
AIC	827.6	851.0	869.0
BIC	949.5	973.7	992.4
Log.Lik.	−387.782	−399.494	−408.511
RMSE	0.40	0.40	0.40

Table D3: Robustness checks for IP, email, crypto for the logistic model (4)

	ip	email	crypto
Know Meme	0.135 (0.025)***	0.128 (0.026)***	0.135 (0.023)***
Crypto trader	0.154 (0.027)***	0.114 (0.028)***	0.130 (0.026)***
When interested	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Wealth inv.	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Leverage	0.122 (0.031)***	0.136 (0.030)***	0.135 (0.031)***
Has NFT	0.149 (0.040)***	0.164 (0.032)***	0.161 (0.031)***
Num non-memecoins held	0.042 (0.005)***	0.041 (0.005)***	0.041 (0.005)***
Need to belong	−0.034 (0.010)***	−0.034 (0.011)**	−0.032 (0.010)**
Taxes	−0.019 (0.009)*	−0.022 (0.009)*	−0.019 (0.009)*
Num. Obs.	949	978	1005
AIC	976.4	1003.2	1027.8
BIC	1049.2	1076.5	1101.5
Log.Lik.	−473.188	−486.604	−498.905
RMSE	0.40	0.40	0.40

- *AllCryptos*: all possible coins reported to be owned (7 obs).

- *BF1std*: bought the first crypto asset more than one standard deviations before claimed to have gotten interested in crypto (67 obs).

Table D4: Robustness checks for duplicated responses for the logistic model (1)

	Chance	50	75
Gender	0.102 (0.045)*	0.102 (0.044)*	0.093 (0.046)*
Age	0.008 (0.006)	0.008 (0.006)	0.007 (0.007)
Edu	−0.014 (0.008)+	−0.014 (0.008)+	−0.014 (0.008)+
Know Meme	0.149 (0.031)***	0.149 (0.029)***	0.151 (0.028)***
Know crypto no Meme	−0.009 (0.019)	−0.007 (0.018)	−0.006 (0.018)
Know fin.	−0.024 (0.030)	−0.024 (0.029)	−0.024 (0.028)
isTrader	0.153 (0.072)*	0.162 (0.073)*	0.180 (0.082)*
Risk Seeking	−0.011 (0.018)	−0.007 (0.019)	−0.002 (0.015)
When interested	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Wealth inv.	0.001 (0.000)**	0.001 (0.000)**	0.001 (0.000)**
Leverage	0.117 (0.022)***	0.114 (0.023)***	0.103 (0.029)***
Has NFT	0.183 (0.030)***	0.186 (0.030)***	0.188 (0.031)***
Num non-memecoins held	0.042 (0.009)***	0.041 (0.009)***	0.041 (0.009)***
Need to belong	−0.035 (0.007)***	−0.034 (0.008)***	−0.032 (0.007)***
FOMO	0.002 (0.012)	0.001 (0.012)	0.002 (0.011)
KYC	0.000 (0.016)	0.001 (0.016)	−0.001 (0.014)
Trust regulations	0.015 (0.012)	0.014 (0.013)	0.014 (0.012)
Scams	0.010 (0.018)	0.009 (0.019)	0.007 (0.018)
Taxes	−0.024 (0.009)**	−0.024 (0.009)**	−0.023 (0.010)*
Trust in Gov.	−0.017 (0.023)	−0.018 (0.023)	−0.013 (0.022)
Num. Obs.	799	808	827
AIC	826.9	831.8	846.9
BIC	948.6	953.8	969.6
Log.Lik.	−387.438	−389.880	−397.470
RMSE	0.40	0.40	0.40

Table D5: Robustness checks for duplicated responses for the logistic model (4)

	Chance	50	75
Know Meme	0.128 (0.027)***	0.129 (0.024)***	0.137 (0.024)***
Crypto trader	0.121 (0.023)***	0.125 (0.024)***	0.134 (0.027)***
When interested	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Wealth inv.	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Leverage	0.133 (0.028)***	0.128 (0.030)***	0.122 (0.032)***
Has NFT	0.148 (0.038)***	0.156 (0.039)***	0.156 (0.038)***
Num non-memecoins held	0.043 (0.005)***	0.042 (0.005)***	0.042 (0.005)***
Need to belong	−0.032 (0.010)**	−0.032 (0.011)**	−0.030 (0.009)**
Taxes	−0.019 (0.009)*	−0.019 (0.008)*	−0.017 (0.009)+
Num. Obs.	941	952	979
AIC	975.6	981.8	1002.9
BIC	1048.3	1054.7	1076.2
Log.Lik.	−472.815	−475.920	−486.473
RMSE	0.41	0.40	0.40

Tables D6 and D7 confirms once more that our main results are robust even when potentially duplicated responses are removed.

Table D6: Robustness checks for unlikely responses for the logistic model (1)

	Ret50	Doc22	All Cryptos	BF1std
Gender	0.088 (0.044)*	0.088 (0.042)*	0.088 (0.042)*	0.080 (0.038)*
Age	0.007 (0.007)	0.007 (0.007)	0.007 (0.007)	0.008 (0.007)
Edu	−0.013 (0.009)	−0.013 (0.009)	−0.013 (0.009)	−0.013 (0.009)
Know Meme	0.145 (0.029)***	0.147 (0.028)***	0.148 (0.028)***	0.143 (0.035)***
Know crypto no Meme	−0.002 (0.017)	−0.003 (0.016)	−0.003 (0.017)	0.000 (0.016)
Know fin.	−0.021 (0.028)	−0.022 (0.029)	−0.022 (0.029)	−0.020 (0.029)
isTrader	0.185 (0.079)*	0.185 (0.079)*	0.186 (0.079)*	0.179 (0.081)*
Risk Seeking	−0.004 (0.015)	−0.005 (0.015)	−0.005 (0.015)	−0.009 (0.016)
When interested	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Wealth inv.	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Leverage	0.117 (0.024)***	0.117 (0.023)***	0.117 (0.024)***	0.131 (0.023)***
Has NFT	0.196 (0.023)***	0.196 (0.024)***	0.197 (0.024)***	0.189 (0.029)***
Num non-memecoins held	0.039 (0.007)***	0.039 (0.007)***	0.039 (0.007)***	0.039 (0.008)***
Need to belong	−0.037 (0.007)***	−0.037 (0.007)***	−0.037 (0.007)***	−0.036 (0.008)***
FOMO	0.002 (0.010)	0.002 (0.010)	0.002 (0.010)	0.006 (0.011)
KYC	−0.002 (0.013)	−0.002 (0.014)	−0.002 (0.014)	−0.003 (0.015)
Trust regulations	0.014 (0.011)	0.016 (0.011)	0.016 (0.011)	0.014 (0.010)
Scams	0.002 (0.016)	0.003 (0.016)	0.003 (0.016)	0.004 (0.019)
Taxes	−0.024 (0.010)*	−0.024 (0.010)*	−0.024 (0.010)*	−0.022 (0.009)*
Trust in Gov.	−0.018 (0.019)	−0.017 (0.019)	−0.017 (0.019)	−0.013 (0.017)
Num. Obs.	849	856	852	827
AIC	867.4	871.9	871.9	849.8
BIC	990.8	995.5	995.3	972.5
Log.Lik.	−407.719	−409.966	−409.937	−398.915
RMSE	0.40	0.40	0.40	0.40

Table D7: Robustness checks for unlikely responses for the logistic model (4)

	Ret50	Doc22	All Cryptos	BF1std
Know Meme	0.138 (0.025)***	0.137 (0.022)***	0.138 (0.022)***	0.142 (0.030)***
Crypto trader	0.138 (0.028)***	0.138 (0.028)***	0.138 (0.028)***	0.121 (0.025)***
When interested	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Wealth inv.	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Leverage	0.129 (0.031)***	0.131 (0.032)***	0.131 (0.032)***	0.144 (0.033)***
Has NFT	0.165 (0.030)***	0.165 (0.031)***	0.166 (0.031)***	0.162 (0.034)***
Num non-memecoins held	0.040 (0.005)***	0.040 (0.005)***	0.040 (0.005)***	0.041 (0.006)***
Need to belong	−0.033 (0.010)***	−0.033 (0.010)**	−0.033 (0.010)**	−0.031 (0.011)**
Taxes	−0.020 (0.009)*	−0.019 (0.009)*	−0.019 (0.009)*	−0.017 (0.008)*
Num. Obs.	1004	1013	1009	978
AIC	1025.1	1031.1	1031.0	999.8
BIC	1098.8	1104.9	1104.8	1073.0
Log.Lik.	−497.554	−500.527	−500.501	−484.877
RMSE	0.40	0.40	0.40	0.40

D.4 Inconsistent memecoin ownership

We cross-validated self-reported memecoin ownership using two items: (i) the checklist of cryptoassets currently held and (ii) a yes/no question on whether respondents held any memecoins over the last 12 months. We flagged as inconsistent memecoin owners

those respondents who selected at least one memecoin in the holdings checklist but answered "No" to the 12-month ownership question. To assess robustness, we excluded these cases and re-estimated the baseline logistic models and our main results still hold (see Table D8).

In addition, we implemented an even more strigent filter to eliminate the influence of potentially unserious participants: we also removed all respondents with an average time-per-answer < 5 seconds (this post hoc response-level filter complements the pre-registered, participant-level rule). The results presented in Table D9 are consistent with those presented in the main text.

Table D8: Robustness to excluding inconsistent memecoin ownership

	(1)	(2)	(3)	(4)
Gender	0.082 (0.045)+	0.078 (0.047)+	0.048 (0.019)*	
Age	0.005 (0.007)			
Edu	−0.016 (0.008)*	−0.012 (0.008)	−0.012 (0.008)	
Know Meme	0.166 (0.030)***	0.160 (0.031)***	0.140 (0.025)***	0.150 (0.023)***
Know crypto no Meme	−0.002 (0.018)			
Know fin.	−0.023 (0.030)	−0.021 (0.029)		
Crypto trader	0.164 (0.086)+	0.149 (0.071)*	0.114 (0.033)***	0.123 (0.031)***
Risk Seeking	0.003 (0.018)	0.004 (0.019)		
When interested	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Wealth inv.	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***	0.002 (0.000)***
Leverage	0.114 (0.029)***	0.117 (0.039)**	0.113 (0.040)**	0.131 (0.038)***
Has NFT	0.196 (0.032)***	0.192 (0.029)***	0.173 (0.036)***	0.167 (0.034)***
Num non-memecoins held	0.037 (0.007)***	0.038 (0.007)***	0.039 (0.006)***	0.038 (0.005)***
Need to belong	−0.035 (0.009)***	−0.038 (0.013)**	−0.034 (0.014)*	−0.033 (0.013)*
FOMO	−0.003 (0.014)			
KYC	−0.009 (0.016)			
Trust regulations	0.018 (0.007)*	0.016 (0.007)*	0.015 (0.006)*	
Scams	0.001 (0.015)	−0.001 (0.013)	−0.003 (0.011)	
Taxes	−0.020 (0.010)+	−0.020 (0.011)+	−0.018 (0.009)+	−0.017 (0.009)+
Trust in Gov.	−0.017 (0.018)	−0.022 (0.014)	−0.014 (0.013)	
Num. Obs.	819	830	938	967
AIC	824.1	825.9	948.1	975.8
BIC	946.5	929.7	1045.0	1048.9
Log.Lik.	−386.062	−390.929	−454.054	−472.896
RMSE	0.39	0.39	0.40	0.40

D.5 Data sharing

Participants could decide to share their responses with *all* the State of Crypto researchers or only with researchers at Uni Mannheim, the data processor according to GDPR (participants below 18 years old were automatically excluded from broader sharing). Here we remove those participants who decided not to share more broadly their data, and replicate the main results in Table D10.

Table D9: Robustness check: removed inconsistent memecoin ownership and avg. time-to-answer less than 5 sec.

	(1)	(2)	(3)	(4)
Gender	0.073 (0.043)+	0.069 (0.046)	0.040 (0.018)*	
Age	0.006 (0.006)			
Edu	−0.018 (0.008)*	−0.014 (0.008)+	−0.013 (0.008)+	
Know Meme	0.168 (0.032)***	0.161 (0.032)***	0.143 (0.025)***	0.149 (0.023)***
Know crypto no Meme	−0.003 (0.019)			
Know fin.	−0.021 (0.032)	−0.019 (0.031)		
Crypto trader	0.163 (0.093)+	0.147 (0.078)+	0.108 (0.040)**	0.120 (0.039)**
Risk Seeking	−0.011 (0.016)	−0.010 (0.017)		
When interested	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Wealth inv.	0.001 (0.000)*	0.001 (0.000)*	0.001 (0.000)***	0.001 (0.000)***
Leverage	0.113 (0.034)***	0.116 (0.042)**	0.109 (0.042)**	0.127 (0.040)**
Has NFT	0.202 (0.036)***	0.196 (0.036)***	0.171 (0.042)***	0.166 (0.040)***
Num non-memecoins held	0.037 (0.008)***	0.038 (0.007)***	0.039 (0.006)***	0.038 (0.005)***
Need to belong	−0.036 (0.008)***	−0.039 (0.012)***	−0.036 (0.013)**	−0.035 (0.013)**
FOMO	−0.002 (0.013)			
KYC	−0.006 (0.017)			
Trust regulations	0.017 (0.007)*	0.016 (0.007)*	0.015 (0.007)*	
Scams	0.003 (0.013)	0.001 (0.012)	−0.001 (0.010)	
Taxes	−0.022 (0.010)*	−0.022 (0.011)*	−0.020 (0.009)*	−0.020 (0.009)*
Trust in Gov.	−0.018 (0.019)	−0.023 (0.016)	−0.014 (0.014)	
Num. Obs.	801	811	916	943
AIC	808.6	810.4	930.9	954.4
BIC	930.4	913.8	1027.3	1027.1
Log.Lik.	−378.290	−383.218	−445.440	−462.176
RMSE	0.39	0.39	0.40	0.40

D.6 Random-effect modeling

We also tried a different modeling approach, namely multilevel regressions. First, we nested participants into continents, continent being a random effect (fixed effect removed). Results are nearly identical (see Table D11); using region instead of continent (13 groups instead of 6) as random effect, we obtain qualitatively very similar results (see Table D12). The very low variance of both random effects indicates that after conditioning on individual characteristics, there is no meaningful continent-level or region-level heterogeneity in memecoin adoption.

Appendix Chapter 7

A Additional Model Specifications

This appendix reports logistic regression results for models including a single category of the control variables.

Table D10: Robustness check: Only those who agreed to share with *all* State of Crypto researchers

	(1)	(2)	(3)	(4)
Gender	0.102 (0.044)*	0.096 (0.043)*	0.062 (0.039)	
Age	0.012 (0.005)*			
Edu	−0.007 (0.012)	−0.002 (0.012)	−0.003 (0.012)	
Know Meme	0.141 (0.037)***	0.137 (0.035)***	0.120 (0.033)***	0.130 (0.032)***
Know crypto no Meme	−0.002 (0.014)			
Know fin.	−0.029 (0.016)+	−0.025 (0.016)		
Crypto trader	0.193 (0.062)**	0.171 (0.061)**	0.142 (0.054)**	0.141 (0.052)**
Risk Seeking	0.005 (0.032)	0.004 (0.032)		
When interested	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Wealth inv.	0.001 (0.000)*	0.001 (0.000)*	0.001 (0.000)***	0.001 (0.000)***
Leverage	0.138 (0.044)**	0.140 (0.043)**	0.133 (0.041)**	0.151 (0.041)***
Has NFT	0.174 (0.041)***	0.174 (0.040)***	0.152 (0.038)***	0.146 (0.037)***
Num non-memecoins held	0.038 (0.006)***	0.040 (0.006)***	0.042 (0.005)***	0.040 (0.005)***
Need to belong	−0.034 (0.015)*	−0.036 (0.014)*	−0.032 (0.013)*	−0.031 (0.013)*
FOMO	0.004 (0.015)			
KYC	−0.009 (0.012)			
Trust regulations	0.016 (0.011)	0.015 (0.011)	0.014 (0.010)	
Scams	0.002 (0.011)	−0.001 (0.010)	0.001 (0.010)	
Taxes	−0.024 (0.011)*	−0.024 (0.011)*	−0.023 (0.011)*	−0.022 (0.010)*
Trust in Gov.	−0.026 (0.020)	−0.034 (0.020)+	−0.023 (0.018)	
Num. Obs.	771	781	889	915
AIC	773.3	779.2	898.8	926.2
BIC	894.2	881.8	994.6	998.5
Log.Lik.	−360.657	−367.623	−429.414	−448.094
RMSE	0.39	0.39	0.40	0.40

Table D11: Random-effect model, continent as random effect.

	(1)	(2)	(3)	(4)
Gender	0.077 (0.041)+	0.076 (0.041)+	0.044 (0.038)	
Age	0.007 (0.005)			
Edu	−0.017 (0.012)	−0.012 (0.011)	−0.012 (0.011)	
Know Meme	0.152 (0.036)***	0.145 (0.034)***	0.130 (0.032)***	0.139 (0.031)***
Know crypto no Meme	−0.005 (0.014)			
Know fin.	−0.028 (0.015)+	−0.027 (0.015)+		
Crypto trader	0.215 (0.058)***	0.194 (0.057)***	0.153 (0.052)**	0.157 (0.050)**
Risk Seeking	0.001 (0.031)	0.001 (0.031)		
When interested	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Wealth inv.	0.001 (0.000)*	0.001 (0.000)*	0.001 (0.000)***	0.001 (0.000)***
Leverage	0.129 (0.041)**	0.128 (0.040)**	0.123 (0.038)**	0.139 (0.038)***
Has NFT	0.204 (0.039)***	0.197 (0.038)***	0.177 (0.037)***	0.170 (0.036)***
Num non-memecoins held	0.038 (0.005)***	0.039 (0.005)***	0.041 (0.005)***	0.040 (0.005)***
Need to belong	−0.030 (0.014)*	−0.033 (0.013)*	−0.029 (0.012)*	−0.028 (0.012)*
FOMO	0.000 (0.014)			
KYC	0.002 (0.011)			
Trust regulations	0.018 (0.011)+	0.020 (0.010)+	0.017 (0.010)+	
Scams	0.003 (0.010)	0.002 (0.010)	0.000 (0.009)	
Taxes	−0.026 (0.011)*	−0.025 (0.011)*	−0.022 (0.010)*	−0.020 (0.009)*
Trust in Gov.	−0.023 (0.019)	−0.028 (0.019)	−0.016 (0.018)	
Num. Obs.	856	867	984	1013
R2 Marg.	0.490	0.488	0.458	0.446
R2 Cond.		0.488	0.463	0.451
AIC	869.3	872.0	1003.3	1031.7
BIC	973.9	957.8	1081.5	1085.9
RMSE	0.40	0.40	0.40	0.40

Table D12: Random-effect model, region as random effect.

	(1)	(2)	(3)	(4)
Gender	0.077 (0.041)+	0.076 (0.041)+	0.044 (0.038)	
Age	0.007 (0.005)			
Edu	−0.017 (0.012)	−0.012 (0.011)	−0.013 (0.011)	
Know Meme	0.152 (0.036)***	0.145 (0.034)***	0.129 (0.032)***	0.139 (0.031)***
Know crypto no Meme	−0.005 (0.014)			
Know fin.	−0.028 (0.015)+	−0.027 (0.015)+		
Crypto trader	0.215 (0.058)***	0.195 (0.056)***	0.156 (0.052)**	0.159 (0.050)**
Risk Seeking	0.001 (0.031)	0.001 (0.031)		
When interested	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Wealth inv.	0.001 (0.000)*	0.001 (0.000)*	0.001 (0.000)***	0.001 (0.000)***
Leverage	0.129 (0.041)**	0.128 (0.040)**	0.125 (0.038)**	0.140 (0.038)***
Has NFT	0.204 (0.039)***	0.197 (0.038)***	0.179 (0.037)***	0.172 (0.036)***
Num non-memecoins held	0.038 (0.005)***	0.039 (0.005)***	0.041 (0.005)***	0.040 (0.005)***
Need to belong	−0.030 (0.014)*	−0.032 (0.013)*	−0.029 (0.013)*	−0.027 (0.012)*
FOMO	0.000 (0.014)			
KYC	0.002 (0.011)			
Trust regulations	0.018 (0.011)+	0.020 (0.010)+	0.018 (0.010)+	
Scams	0.003 (0.010)	0.002 (0.010)	0.000 (0.009)	
Taxes	−0.026 (0.011)*	−0.025 (0.011)*	−0.023 (0.010)*	−0.020 (0.009)*
Trust in Gov.	−0.023 (0.019)	−0.028 (0.018)	−0.016 (0.018)	
Num. Obs.	856	867	984	1013
R2 Marg.	0.490	0.488	0.460	0.446
R2 Cond.			0.462	0.451
AIC	869.3	872.0	1003.8	1032.2
BIC	973.9	957.8	1082.1	1086.4
RMSE	0.40	0.40	0.40	0.40

Table D13: Only socio-demographics control variables

	Trust in reg.		KYC checks		Taxes	
Male	−0.032	(0.039)	−0.058*	(0.026)	−0.083	(0.059)
Age	0.011*	(0.005)	0.009	(0.009)	0.015*	(0.006)
Education	0.030	(0.021)	0.059***	(0.012)	0.055***	(0.015)
When interested	0.001***	(0.000)	0.001**	(0.000)	0.000	(0.000)
Works in crypto	−0.067*	(0.027)	−0.027	(0.066)	−0.037	(0.042)
Knows crypto	−0.007	(0.011)	−0.031	(0.029)	−0.033***	(0.008)
Knows finance	0.032**	(0.012)	0.034**	(0.011)	0.048***	(0.013)
Num.Obs.	912		822		925	
AIC	1146.7		1073.8		1171.1	
BIC	1209.3		1135.1		1233.9	
Log.Lik.	−560.348		−523.915		−572.565	
RMSE	0.46		0.47		0.46	

Table D14: Only investment-related control variables

	Trust in reg.		KYC checks		Taxes	
Wealth invested	−0.002	(0.001)	−0.003**	(0.001)	−0.003***	(0.000)
Num. of cryptos held	−0.003	(0.005)	−0.002	(0.006)	0.000	(0.005)
Has NFT	−0.023	(0.017)	−0.030	(0.048)	0.059	(0.038)
Has memecoins	0.007	(0.026)	−0.046	(0.024)	−0.085*	(0.038)
Num.Obs.	1034		908		1056	
AIC	1315.3		1159.2		1323.1	
BIC	1364.7		1207.3		1372.7	
Log.Lik.	−647.652		−569.578		−651.560	
RMSE	0.47		0.47		0.46	

Table D15: Only attitudinal control variables

	Trust in reg.		KYC checks		Taxes	
Scams	0.043***	(0.011)	0.051***	(0.014)	0.074***	(0.009)
Mass control	0.025*	(0.010)	0.003	(0.016)	0.034	(0.018)
Trust in Gov.	0.071*	(0.035)	0.084	(0.055)	0.119***	(0.031)
Need to belong	0.059***	(0.016)	0.035*	(0.017)	0.022	(0.014)
FoMO	0.015	(0.021)	−0.019	(0.019)	−0.016	(0.012)
Num.Obs.	1021		901		1041	
AIC	1244.4		1158.0		1234.4	
BIC	1298.6		1210.9		1288.8	
Log.Lik.	−611.213		−568.008		−606.182	
RMSE	0.45		0.47		0.45	

B Survey Implementation and Variable Definitions

B.1 Survey implementation details

The 2024/25 edition of the State of Crypto Survey was fielded between December 2024 and February 2025. Recruitment took place on several social media platforms, with targeted advertisements on *X* and *Reddit*. The survey yielded 2,219 entries, of which 1,414 were fully completed. Participation was fully voluntary and not incentivized.

To maximise inclusiveness, the survey was translated into 30 languages. The most common languages of completion are English (54.3%), Turkish (10.57%), Spanish (6.7%), Amharic (5.2%), Chinese (4.6%), German (3.1%), and Italian (3.1%). Respondents reported locations in more than 150 countries, with the largest shares coming from Nigeria (9.1%), Turkey (8.7%), Ethiopia (5.0%), Germany (4.2%), Spain (3.7%), and the United States (3.7%).

On average, the survey lasted about 10 minutes.

B.2 Survey variables: wording and coding

Table D16: Explanatory variables: definitions and coding

Socio-demographic variables		
Variable	Code	Definition
Male	*Male*	Dummy variable equal to 1 if the respondent identifies as male, 0 otherwise.
Age	*Age*	Age group of the respondent at the time of the survey.
Education	*Education*	Highest level of education completed by the respondent.
Interest in cryptocurrencies	*When interested*	Year when the respondent first became interested in cryptocurrencies.
Working in crypto	*Works in crypto*	Dummy variable equal to 1 if the respondent works in the crypto sector, 0 otherwise.
Cryptocurrency literacy	*Knows crypto*	Index measuring the respondent's level of general cryptocurrency literacy.
Financial literacy	*Knows finance*	Index measuring the respondent's level of financial literacy.
Investment profile		
Variable	Code	Definition
Wealth invested in cryptocurrencies	*Wealth invested*	Share (in %) of the respondent's overall investment portfolio held in cryptoassets.
Own NFT	*Has NFT*	Dummy variable equal to 1 if the respondent owns at least one NFT, 0 otherwise.
Own memecoins	*Has memecoins*	Dummy variable equal to 1 if the respondent owns at least one memecoin, 0 otherwise.
Number of different cryptocurrencies held	*Number of cryptos held*	Number of different cryptocurrencies currently held by the respondent.
Attitudes towards crypto and society		
Variable	Code	Definition
Scams	*Scams*	Perception of whether cryptocurrencies facilitate money laundering and scams more than cash or other means of payment.
Mass control	*Mass control*	Degree to which the respondent is worried that cryptocurrencies could be used as an instrument of mass control.
Trust in government	*Trust in government*	Degree to which the respondent thinks the government in the country where they live can be trusted to do what is right.
Need to belong	*Need to belong*	Self-reported strength of the respondent's "need to belong" (desire for social acceptance and inclusion).
FoMO	*FoMO*	Self-reported extent to which the respondent experiences FoMO (fear of missing out).

C Descriptive Statistics

This appendix reports descriptive statistics of the control and dependent variables as well as the cross-correlation matrix obtained from the observations used in regression models.

Table D17: Descriptive statistics

	Mean	SD	Min	Max	Median	N
Male	0.84	0.37	0.00	1.00	1.00	966
Age	34.61	11.85	16.00	80.00	31.50	966
Education	2.87	1.28	0.00	5.00	3.00	966
When interested	125.20	44.70	1.00	192.00	133.00	966
Works in crypto	0.31	0.46	0.00	1.00	0.00	966
Knows crypto	2.40	1.44	0.00	5.00	2.00	966
Knows finance	1.57	1.00	0.00	3.00	2.00	966
Wealth invested	37.08	37.14	0.00	100.00	25.00	966
Number of cryptos held	3.24	3.62	0.00	28.00	2.00	966
Has NFT	0.26	0.44	0.00	1.00	0.00	966
Has memecoins	0.48	0.50	0.00	1.00	0.00	966
Scams	1.83	1.48	0.00	4.00	2.00	966
Mass control	1.53	1.31	0.00	4.00	1.00	966
Trust in Gov.	0.91	0.82	0.00	3.00	1.00	966
Need to belong	1.21	1.18	0.00	4.00	1.00	966
FoMO	1.36	1.10	0.00	4.00	1.00	966
Trust in reg. recoded	0.65	0.48	0.00	1.00	1.00	839
KYC checks recoded	0.60	0.49	0.00	1.00	1.00	753
Taxes recoded	0.38	0.49	0.00	1.00	0.00	854
Europe	0.33	0.47	0.00	1.00	0.00	966
Asia	0.28	0.45	0.00	1.00	0.00	966
Africa	0.24	0.43	0.00	1.00	0.00	966
North America	0.07	0.25	0.00	1.00	0.00	966
South America	0.07	0.26	0.00	1.00	0.00	966
Oceania	0.01	0.12	0.00	1.00	0.00	966

Table D18: Cross-correlation matrix based on the observation used in the regression models

	Oceania	South America	North America	Africa	Asia	Europe	Works in crypto	Wealth invested	Trust. Gov.	totCryptos	Scams	Mass control	Male	Knows finance	Knows crypto	When interested	Has NFT	FoMO	Education	Need to belong	Age	Has memecoins	KYC checks
Trust in reg.	-.01	-.16***	-.06	.10	.07	-.04	.00	-.18***	.14**	-.07	.17***	.10	-.08	.03	-.01	.13*	-.06	.09	.05	.19***	.01	-.04	.37***
KYC checks	-.03	-.08	-.07	.17***	.00	-.07	.02	-.23***	.17***	-.11	.15***	.00	-.09	-.02	-.07	.08	-.10	.01	.10	.14**	.01	-.13*	
Has memecoins	.01	-.10	.00	.08	.09	-.11	.17***	.35***	-.11	.47***	-.13*	-.01	.09	-.08	.24***	.03	.36***	.10	-.09	-.01	-.01		
Age	-.03	.08	.16***	-.26***	.08	.05	-.16***	.03	-.04	.12*	-.04	.01	.06	.16***	.06	-.26***	-.04	-.08	.26***	-.19***			
Need to belong	-.01	-.07	-.09	.39***	-.06	-.21***	.17***	-.01	.05	-.07	-.03	.08	-.03	-.16***	-.05	.18***	.06	.37***	-.13*				
Education	-.02	.02	.07	-.24***	.00	.18***	-.02	-.05	.07	.02	.09	-.03	-.04	.20***	.11	-.18***	-.02	-.03					
FoMO	-.04	-.09	-.02	.19***	-.04	-.07	.08	.08	.01	.05	.07	.05	-.01	-.08	.04	.08	.14**						
Has NFT	-.01	-.11	.04	.07	.04	-.06	.26***	.29***	-.02	.30***	-.09	-.03	.06	-.03	.28***	-.08							
When interested	.02	-.03	-.12*	.19***	.00	-.10	-.04	-.11	-.03	-.14**	.06	.01	-.17***	-.15***	-.11								
Knows crypto	.04	-.06	.02	-.09	-.01	.11	.21***	.26***	-.02	.32***	-.11	.02	.09	.28***									
Knows finance	.00	.04	.09	-.28***	-.05	.23***	-.06	-.01	.06	.02	.10	.00	.08										
Male	.03	.06	.00	-.19***	.12	.02	-.03	.16***	.03	.10	.00	-.09											
Mass control	-.06	-.06	.01	.02	.08	-.05	.00	-.05	.00	-.01	.18***												
Scams	-.04	-.01	-.05	-.04	.01	.07	-.09	-.21***	.18***	-.20***													
totCryptos	-.01	-.06	.06	-.06	.08	-.02	.18***	.39***	-.09														
Trust. Gov.	.00	-.16***	.00	-.06	-.08	.22***	.03	-.14**															
Wealth invested	.03	-.05	.02	-.12*	.12	.01	.11																
Works in crypto	-.03	-.12	-.06	.28***	-.04	-.12																	
Europe	-.08	-.20***	-.18***	-.39***	-.43***																		
Asia	-.08	-.18***	-.16***	-.35***																			
Africa	-.07	-.16***	-.15***																				
North America	-.03	-.07																					
South America	-.03																						

D Robustness Checks

D.1 Duplicate Responses

Participation to the survey was not gated, anyone with the access link could join it. A cookie would remember if a survey has been already started, but this alone does not ensure that the same person does not participate multiple times. Being the survey unincentivized, the risk is small, nonetheless herein we try to quantify multiple participation, by looking for specific robustness check.

To investigate the issue, we created an index of similarity ranging from zero (fully dissimilar) to one (identical), representing the share of identical answers to the questions in the demographics section of the survey. The average similarity of any two users is about 0.15; considering only those respondents who completed the survey, this number increases to about 0.19. Note that this indicates a rather low level of similarity between any two users; nonetheless, we used it as one of the baselines for the robustness checks Here are all the sets for potentially duplicated responses:

- *Chance*: similarity score above 0.19;
- *50*: similarity score above 50
- *75*: similarity score above 75

For brevity, we report the regression results for the first model (*Trust in reg.*) and find that the results are unchanged across all models.

and also applied a robustness check for respondents who shared the same IP, email, or crypto address: we removed all such records. Sharing the same IP address is not alarming; this is common, for example, among students in the same dorm or users of the same Internet Service Provider (ISP) at different physical locations when IPs are dynamically assigned. Sharing the same crypto or email address is more concerning, but it does not, per se, mean that the same person took the survey twice. In fact, especially in developing countries, the same physical computer and the same email address could be shared among members of the same family, relatives, and friends.

We report the regression only for the first model, but the results are consistent across all models.

D.2 Unlikely Sets of Answers

Some responses, or combinations of those, are less likely than others. We define the following unlikely sets, which we use in the robustness analyses.

- *Ret30*: retired before age 30;
- *Doc22*: holding a doctorate before age 22;
- *AllCryptos*: reporting ownership of all possible coins;
- *BF2std*: bought the first crypto asset more than two standard deviations earlier than the stated time of becoming interested in crypto;
- *BF1std*: bought the first crypto asset more than one standard deviation earlier than the stated time of becoming interested in crypto.

Table D19: Robustness checks for duplicated responses

	Chance		50		75	
Male	−0.012	(0.054)	−0.012	(0.051)	−0.025	(0.045)
Age	0.016**	(0.005)	0.015**	(0.005)	0.016**	(0.006)
Education	0.019	(0.017)	0.017	(0.018)	0.020	(0.018)
When interested	0.001*	(0.000)	0.001*	(0.000)	0.001*	(0.000)
Works in crypto	−0.072*	(0.029)	−0.074**	(0.027)	−0.065*	(0.030)
Knows crypto	0.014	(0.018)	0.015	(0.018)	0.017	(0.017)
Knows finance	0.006	(0.010)	0.009	(0.012)	0.009	(0.013)
Wealth invested	−0.002**	(0.001)	−0.002**	(0.001)	−0.002**	(0.001)
Num. of cryptos held	−0.002	(0.005)	−0.001	(0.005)	0.000	(0.005)
Has NFT	−0.022	(0.018)	−0.018	(0.016)	−0.018	(0.015)
Has memecoins	0.012	(0.028)	0.012	(0.028)	0.004	(0.029)
Scams	0.037*	(0.017)	0.036*	(0.017)	0.037*	(0.018)
Mass control	0.023	(0.012)	0.022*	(0.011)	0.019*	(0.009)
Trust in Gov.	0.070	(0.052)	0.069	(0.052)	0.071	(0.050)
Need to belong	0.070*	(0.027)	0.070**	(0.027)	0.067**	(0.025)
FoMO	0.004	(0.030)	0.004	(0.030)	0.010	(0.029)
Num.Obs.	768		773		793	
AIC	920.1		927.9		951.7	
BIC	1022.3		1030.2		1054.6	
Log.Lik.	−438.062		−441.944		−453.856	
RMSE	0.44		0.44		0.44	

Table D20: Robustness checks for IP, email, crypto

	IP		Email		Crypto	
Male	−0.017	(0.056)	−0.029	(0.044)	−0.049	(0.038)
Age	0.016**	(0.005)	0.016**	(0.005)	0.016**	(0.006)
Education	0.021	(0.017)	0.021	(0.015)	0.022	(0.016)
When interested	0.001	(0.001)	0.001*	(0.000)	0.001*	(0.000)
Works in crypto	−0.054	(0.029)	−0.071**	(0.027)	−0.056	(0.031)
Knows crypto	0.015	(0.019)	0.017	(0.018)	0.018	(0.014)
Knows finance	0.010	(0.011)	0.005	(0.011)	0.010	(0.014)
Wealth invested	−0.002**	(0.001)	−0.002**	(0.001)	−0.002*	(0.001)
Num. of cryptos held	−0.003	(0.005)	0.000	(0.005)	−0.001	(0.005)
Has NFT	−0.022	(0.018)	−0.025	(0.018)	−0.023**	(0.008)
Has memecoins	0.011	(0.031)	0.002	(0.024)	0.006	(0.023)
Scams	0.037*	(0.018)	0.035*	(0.017)	0.036*	(0.016)
Mass control	0.024	(0.013)	0.021*	(0.010)	0.018*	(0.009)
Trust in Gov.	0.067	(0.053)	0.070	(0.050)	0.075	(0.047)
Need to belong	0.069**	(0.026)	0.065**	(0.025)	0.063**	(0.024)
FoMO	0.002	(0.030)	0.009	(0.030)	0.009	(0.029)
Num.Obs.	772		790		815	
AIC	922.5		941.9		971.3	
BIC	1024.8		1044.7		1074.8	
Log.Lik.	−439.241		−448.969		−463.653	
RMSE	0.44		0.44		0.44	

Table D21: Robustness checks for unlikely responses

	Ret30		Doc22		AllCryptos		BF1std		BF2std	
Male	−0.048	(0.038)	−0.048	(0.038)	−0.048	(0.037)	−0.037	(0.039)	−0.040	(0.041)
Age	0.016**	(0.005)	0.016**	(0.006)	0.016**	(0.006)	0.016**	(0.006)	0.016*	(0.006)
Education	0.020	(0.016)	0.022	(0.016)	0.023	(0.016)	0.023	(0.017)	0.027	(0.017)
When interested	0.001*	(0.000)	0.001*	(0.000)	0.001*	(0.000)	0.001*	(0.000)	0.001*	(0.000)
Works in crypto	−0.057	(0.031)	−0.056	(0.031)	−0.060	(0.032)	−0.064	(0.036)	−0.070	(0.049)
Knows crypto	0.018	(0.014)	0.018	(0.014)	0.017	(0.014)	0.024	(0.014)	0.024	(0.016)
Knows finance	0.010	(0.014)	0.010	(0.014)	0.011	(0.014)	0.005	(0.015)	0.006	(0.012)
Wealth invested	−0.002*	(0.001)	−0.002*	(0.001)	−0.002*	(0.001)	−0.001*	(0.001)	−0.002*	(0.001)
Num. of cryptos held	−0.001	(0.005)	−0.001	(0.005)	0.001	(0.005)	−0.001	(0.005)	−0.001	(0.005)
Has NFT	−0.025**	(0.008)	−0.023**	(0.008)	−0.026***	(0.007)	−0.032**	(0.011)	−0.038	(0.020)
Has memecoins	0.006	(0.023)	0.006	(0.023)	0.003	(0.022)	−0.007	(0.023)	−0.001	(0.023)
Scams	0.036*	(0.016)	0.036*	(0.016)	0.036*	(0.016)	0.035*	(0.016)	0.034	(0.018)
Mass control	0.018	(0.009)	0.018*	(0.009)	0.017	(0.009)	0.018*	(0.008)	0.015*	(0.006)
Trust in Gov.	0.075	(0.047)	0.075	(0.047)	0.073	(0.046)	0.075	(0.048)	0.069	(0.042)
Need to belong	0.063**	(0.023)	0.064**	(0.024)	0.064**	(0.023)	0.066**	(0.025)	0.067**	(0.024)
FoMO	0.009	(0.029)	0.009	(0.029)	0.011	(0.029)	0.008	(0.031)	0.002	(0.029)
Num.Obs.	814		818		814		792		745	
AIC	968.8		971.7		967.8		944.0		891.5	
BIC	1072.2		1075.3		1071.3		1046.8		993.0	
Log.Lik.	−462.396		−463.854		−461.921		−450.002		−423.745	
RMSE	0.44		0.44		0.44		0.44		0.44	

D.3 Geolocation

We geolocated the IP addresses of all respondents and compared them with their self-reported country of residence. In sum, we define the following set that we test in the robustness analysis:

- *IP mismatch*: country of residence and IP do not match.

We reported the results for all three logistic regression models.

D.4 Response Times

We monitored response times throughout the survey. As preregistered, we excluded respondents whose completion times for either the overall survey or any section were faster than two standard deviations below the mean; to prevent undue influence of extreme delays, cases taking longer than one hour were removed from the standard-deviation computation.

For the main analyses, we implemented an item-level filter: all responses with time-per-answer < 3 seconds were excluded. We also computed a robustness check with removing the respondents with time-per-answer < 5 seconds.

D.5 Bot Detection

Unlike [44], we did not use bots *honeypots* because the captcha proved reliable in the previous survey and this allowed us a leaner and more accessible setup.

E Ethical and regulatory clearance

Prior to launching our survey, we preregistered our research at AsPredicted.org (https://osf.io/5unmq?view_only=82ab1b433b6e4759a3a510e21f3897ef). Most

Table D22: Robustness checks for Geolocation

	Trust in reg.		KYC checks		Taxes	
Male	−0.051	(0.038)	−0.014	(0.037)	−0.030	(0.037)
Age	0.017*	(0.007)	0.013	(0.008)	0.023***	(0.004)
Education	0.022	(0.018)	0.047***	(0.012)	0.037***	(0.011)
When interested	0.001**	(0.000)	0.001*	(0.000)	0.000	(0.000)
Works in crypto	−0.064	(0.035)	−0.015	(0.060)	−0.058**	(0.022)
Knows crypto	0.024	(0.016)	0.001	(0.019)	0.001	(0.008)
Knows finance	−0.002	(0.014)	−0.007	(0.014)	0.017	(0.019)
Wealth invested	−0.002*	(0.001)	−0.002**	(0.001)	−0.002***	(0.000)
Num. of cryptos held	0.000	(0.004)	0.001	(0.006)	0.005	(0.007)
Has NFT	−0.014	(0.015)	−0.078	(0.053)	0.087	(0.051)
Has memecoins	−0.001	(0.021)	−0.077*	(0.031)	−0.077*	(0.037)
Scams	0.039*	(0.017)	0.036***	(0.010)	0.046***	(0.009)
Mass control	0.008	(0.011)	−0.006	(0.016)	0.028**	(0.010)
Trust in Gov.	0.090*	(0.045)	0.097**	(0.034)	0.101***	(0.026)
Need to belong	0.058*	(0.027)	0.018	(0.010)	0.036	(0.020)
FoMO	0.012	(0.027)	−0.004	(0.022)	−0.025	(0.017)
Num.Obs.	699		619		714	
AIC	813.2		741.9		813.4	
BIC	913.3		839.3		914.0	
Log.Lik.	−384.581		−348.966		−384.724	
RMSE	0.43		0.44		0.43	

Table D23: Robustness checks for time-per-answer < 5 seconds

	Trust in reg.		KYC checks		Taxes	
Male	−0.048	(0.038)	−0.009	(0.021)	−0.073*	(0.034)
Age	0.016**	(0.006)	0.012	(0.008)	0.021***	(0.005)
Education	0.022	(0.016)	0.056***	(0.009)	0.033*	(0.016)
When interested	0.001*	(0.000)	0.001*	(0.000)	0.000	(0.000)
Works in crypto	−0.056	(0.031)	0.001	(0.063)	−0.041*	(0.018)
Knows crypto	0.018	(0.014)	−0.002	(0.024)	−0.001	(0.010)
Knows finance	0.010	(0.014)	−0.001	(0.012)	0.025	(0.018)
Wealth invested	−0.002*	(0.001)	−0.002***	(0.001)	−0.002***	(0.000)
Num. of cryptos held	−0.001	(0.005)	0.000	(0.007)	0.003	(0.007)
Has NFT	−0.023**	(0.008)	−0.063	(0.067)	0.067	(0.047)
Has memecoins	0.006	(0.023)	−0.049	(0.029)	−0.059	(0.031)
Scams	0.036*	(0.016)	0.031***	(0.009)	0.054***	(0.010)
Mass control	0.018*	(0.009)	−0.010	(0.012)	0.024*	(0.010)
Trust in Gov.	0.075	(0.047)	0.088**	(0.033)	0.108***	(0.017)
Need to belong	0.064**	(0.024)	0.028*	(0.012)	0.029	(0.020)
FoMO	0.009	(0.029)	−0.014	(0.017)	−0.014	(0.020)
Num.Obs.	818		731		832	
AIC	971.7		886.8		949.5	
BIC	1075.3		987.8		1053.4	
Log.Lik.	−463.854		−421.375		−452.745	
RMSE	0.44		0.44		0.43	

importantly, we obtained ethical review clearance from the Ethics Committee of the University of Mannheim (EK Mannheim 40/2023) and worked thoroughly with the Data Protection Team of the University of Mannheim to verify the compliance of our methods and infrastructure with the GDPR regulations.

E.1 Internal Consistency

Table D24: Descriptive Statistics and Correlations for Regulatory Items

Variable	Mean	SD	Min	Max	(1)	(2)
Regulations (1)	0.632	0.482	0	1	—	
KYC (2)	0.594	0.491	0	1	0.39^{***}	—
Taxes (3)	0.372	0.484	0	1	0.33^{***}	0.27^{***}

Note: *** $p < .001$.

To assess whether the three regulatory-attitudes items captured a common underlying construct, we estimated a one-factor maximum likelihood factor analysis (with varimax rotation). The factor loadings indicate that all three items load on a single latent dimension (Regulations = 0.68, KYC = 0.56, Taxes = 0.48). These values fall within the typical range for short attitudinal scales and suggest that each item contributes meaningfully to the shared construct. The corresponding uniqueness values (0.53, 0.68, and 0.77) show that although the items share common variance, each also retains item-specific variance, which is expected given that they measure distinct facets of regulatory attitudes. The single factor accounts for 34% of the total variance (SS = 1.02), a level consistent with other concise three-item attitudinal measures. Taken together, these results support treating the items as indicators of a unidimensional regulatory-attitudes scale.

A principal components analysis of the correlation matrix further reinforces this conclusion. The first principal component accounts for 55.3% of the total variance, indicating a substantial shared regulatory-attitude dimension underlying all three indicators. The remaining variance is distributed across the second (24.6%) and third (20.2%) components, reflecting item-specific content differences, namely trust in regulation, identity verification requirements, and taxation—rather than coherent secondary dimensions.

Finally, we assessed the internal consistency of the three indicators. Corrected item–total correlations are moderate to strong (0.46–0.55), and all items exhibit positive r.drop values (0.35–0.42), indicating that each contributes to the overall reliability of the scale. These values fall within the expected range for a three-item attitudinal measure. As anticipated, the taxation item shows slightly weaker psychometric performance than the items measuring trust in regulation and support for identity checks, consistent with its narrower substantive focus. Overall, the three items demonstrate acceptable internal consistency while capturing complementary aspects of respondents' regulatory preferences.

In summary, the factor analysis, principal components analysis, and internal consistency metrics all converge to show that the three items jointly form a coherent and empirically defensible unidimensional measure of regulatory attitudes.

Appendix Chapter 9

A Derivatives contracts in traditional finance

This section provides additional details on the derivative securities traded in traditional financial markets discussed in Section 9.2. We outline their main properties, how they are traded, and describe their payoff functions. This serves as a support for readers that are not familiar with such concepts. A deeper discussion can be found in [163].

Forward contracts are contracts between two parties to buy or sell an asset at a specified future date T for a price previously agreed upon, the *delivery price* K. It differs from the *spot price*, S_0, the price one would pay for immediate delivery (i.e., at time $t = 0$) of the asset. The party agreeing to buy opens a long position, whereas the one agreeing to sell opens a short position. The delivery price can differ from the spot price at maturity S_T, that is, the actual spot price on the day the contract expires – and this determines the profitability of the contract for the involved parties. The payoffs of the buyer and seller are respectively given by the formulas:

$$\Pi_b = S_T - K, \qquad \Pi_s = K - S_T$$

Figure 11.5 illustrates graphically how the payoff of a long position changes as a function of S_T. The payoff of a short position is symmetrical with respect to the x-axis.

Forwards are private, non-standardized, large-size contracts whose terms (e.g., quantity, quality, expiration) can be negotiated directly between the parties. These types of contracts are traded in over-the-counter (OTC) markets, typically by large institutional participants, interacting through a network of brokers or electronic communication networks. In contrast to exchange markets with visible order books and transparent price discovery, in OTC markets, prices are set by large financial institutions that act as market makers, i.e. they provide liquidity by quoting buy and sell prices at which they are ready to accept incoming orders. Typically, forwards have one specified delivery date and are settled at the end of the contract term.

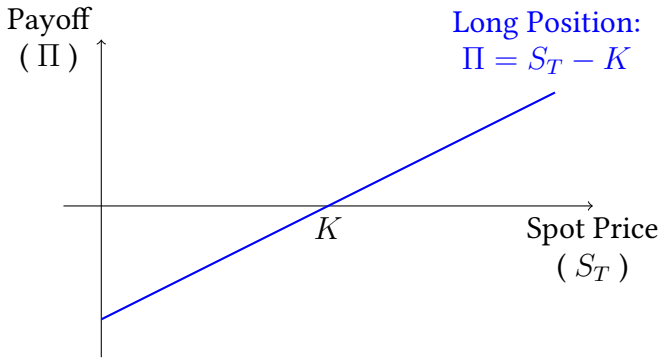


Figure 11.5: Payoff diagram with the spot price at expiration (S_T) on the x-axis and the payoff on the y-axis.

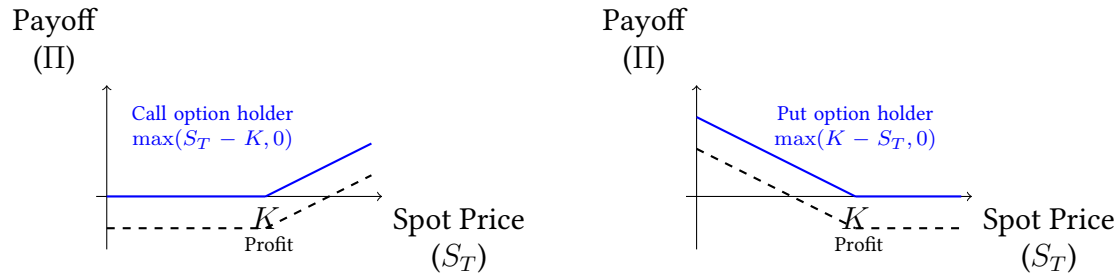


Figure 11.6: Payoff diagram with the spot price at expiration (S_T) on the x-axis and the payoff on the y-axis.

Futures are similar to forward contracts; however, these are standardized and more liquid contracts whose terms are uniform across markets. Therefore, futures are traded primarily in exchange-traded markets. Buyers and sellers submit their orders to the exchange. Orders are organized into an order-book that sorts them based on price and quantity bidded (or asked). Orders are then matched according to specific algorithms. When this occurs, a trade is executed, and orders are cleared from the order-book. The price discovery is the process by which an asset price is determined through supply and demand in the marketplace.

In terms of mathematical structure at expiration, both forward and future contracts have identical payoff functions.

The main difference resides in that futures contracts are settled on a daily basis, while forward contracts are settled only at expiration. Users entering into a future contract are required to deposit an amount known as the initial margin, i.e. a percentage of the contract's total value, to ensure they can fulfill their obligations. As futures are traded, their value might vary. At the end of each trading day, a trade reflecting such changes is settled, leading to funds moving from long to short positions (or vice versa). This process is called marking-to-market. Thus, gains or losses are realized incrementally each day as the contract price changes.

Options are instruments that grant the possibility, but not the obligation, to exercise a certain right on an underlying asset within a specific period. The most common ones are call and put options. They respectively allow the option buyer (holder) to purchase or sell the underlying asset at a strike (also called exercise) price K. The seller of the option is also called the writer. American and European options, respectively, refer to contracts that can be exercised at any time or only on the expiration date itself.

In terms of payoff functions, the payoffs of the holder and writer of a call option are respectively given by the formulas:

$$\Pi_h = \begin{cases} S_T - K & \text{if } S_T > K \\ 0 & \text{if } S_T \leq K \end{cases} \qquad \Pi_w = \begin{cases} K - S_T & \text{if } S_T > K \\ 0 & \text{if } S_T \leq K \end{cases}$$

The payoffs of the holder and writer of a put option are respectively given by the formulas:

$$\Pi_h = \begin{cases} 0 & \text{if } S_T \geq K \\ K - S_T & \text{if } S_T < K \end{cases} \qquad \Pi_w = \begin{cases} 0 & \text{if } S_T \geq K \\ S_T - K & \text{if } S_T < K \end{cases}$$

Notably, the holder of the option must pay, for this right, a premium to the writer of the option. Therefore, the final profit accounts for the premium paid (received) to buy (sell) the option. Figure 11.5 illustrates graphically how the payoff and profit of a call (left) and put (right) position changes for the holder as a function of S_T. The payoff of a short position is symmetrical with respect to the x-axis. The buyer's maximum loss is limited to the original purchase price of the option contract, while its potential gains are theoretically unlimited if the spot price rises far above the strike price.

B Data resources

Table D25 reports the data sources used for this study.

C Perpetuals Protocols - Fees

D Formal Definitions and Formulas

This appendix provides compact formal definitions for the metrics used in Section 9.7.

Notation and Conventions

- P_t: underlying mark (or spot) price at time t; P_{entry}: entry mark.
- Q: signed exposure in *units of underlying* for perpetuals ($Q > 0$ long, $Q < 0$ short).
- C: posted collateral at entry (assume net of upfront entry fees unless noted).
- L: leverage; NV: notional exposure at entry.
- $S \in \{+1, -1\}$: position sign ($+1$ long, -1 short).
- F_{open}, F_{close}: one-off execution fees at entry/exit (trading, gas, price impact).
- Φ_t: cumulative time-dependent fees up to t (funding, borrowing, stability, etc.).
- r_m: maintenance margin rate; $M_{m,t}$: maintenance margin requirement at t.

D.1 Perpetual Futures

Position size and notional value

$$Q \;=\; \frac{C\,L}{P_{\text{entry}}}, \qquad \text{NV} \;=\; C \times L. \tag{11.1}$$

Unrealized and realized PnL

$$\text{UPnL}_t = S\,Q\,(P_t - P_{\text{entry}}), \tag{11.2}$$

$$\text{RPnL}_t = \text{UPnL}_t - F_{\text{open}} - F_{\text{close}} - \Phi_t. \tag{11.3}$$

Funding and borrowing. With per-interval funding rate F_t over Δt:

$$\text{FundingPayment}_t \;=\; S \times \text{NV} \times F_t \times \Delta t. \tag{11.4}$$

Table D25: Data resources on DeFi protocols and documentation

Source	Description	Link
Alchemix	Official Documentation	https://docs.alchemix.fi/
Derive	Official Documentation	https://docs.derive.xyz/docs/about-derive
Deri	Official Documentation	https://docs.deri.io/
Deri	V4 white paper	https://github.com/deri-protocol/whitepaper/blob/master/deri_v4_whitepaper.pdf
Deri	V3 white paper	https://github.com/deri-protocol/whitepaper/blob/master/deri_v3_whitepaper.pdf
Deri	V2 white paper	https://github.com/deri-protocol/whitepaper/blob/master/deri_v2_whitepaper.pdf
Deri	V1 white paper	https://github.com/deri-protocol/whitepaper/blob/master/deri_whitepaper.pdf
Deri	white paper Exchange Everlasting Option	https://github.com/deri-protocol/whitepaper/blob/master/deri_everlasting_options_whitepaper.pdf
Deri	white paper Pricing Everlasting Option with Interest Rate	https://github.com/deri-protocol/whitepaper/blob/master/deri_everlasting_options_with_interst_rate.pdf
Deri	white paper Pricing Everlasting Option	https://github.com/deri-protocol/whitepaper/blob/master/Pricing_Continuously_Funded_Everlasting_Options.pdf
Drift Protocol v2	Official Documentation	https://docs.drift.trade/
Drift Protocol v0	White Paper	https://cdn.prod.website-files.com/611580035ad59b20437eb024/61293b57e3103934ddc5535f_v0%20Devnet%20Feature%20Paper%20-%20Revision%201.1.pdf
dYdX	Support Documentation	https://help.dydx.trade/en/
dYdX	Official Documentation	https://docs.dydx.exchange/
GammaSwap	Official Documentation	https://docs.gammaswap.com/
GMX	Official Documentation	https://gmx-docs.io/docs/intro/
Hegic	Official Documentation	https://www.hegic.co/app#/learn/
Hegic	White Paper	https://github.com/hegic/whitepaper/blob/master/Hegic%20Protocol%20Whitepaper.pdf
Hyperliquid	Official Documentation	https://hyperliquid.gitbook.io/hyperliquid-docs
Jupiter	Support Documentation	https://support.jup.ag/hc/en-us
Jupiter	Developer Documentation	https://dev.jup.ag/
Metronome	Official Documentation	https://docs.metronome.io/metronome-2.0/master
Opyn	Official Documentation	https://opyn.gitbook.io/opyn-hub
Paradigm	White Paper Power Perpetuals	https://www.paradigm.xyz/2021/08/power-perpetuals
Paradigm	Everything is a perp	https://www.paradigm.xyz/2024/03/everything-is-a-perp
Paradigm	Floor Perps	https://www.paradigm.xyz/2021/08/floor-perps
Paradigm	Everlasting Option	https://www.paradigm.xyz/2021/05/everlasting-options
Perp v2	Official Documentation	https://support.perp.com/
Solscan	2° Perp On-chain Transaction	https://solscan.io/tx/39aHLoPEEjpsN625BhuGpgYR2Li1nLFgdRZASnrEPvMCQU1swbWR551iTw81uW7h5udaQVwFDzJ2Mh19E8N8Bc8h
Solscan	1° Perp On-chain Transaction	https://solscan.io/tx/4JxAdCXhGsgZeujHuDc7aE8b4NPWuSEegME2T2UppufjWvW8Df7ypzBtmagQTG2qmXCrtYzBjsXETDBrfmJxsGqj
Synthetix	Official Documentation	https://docs.synthetix.io/
Taiga	Official Documentation	https://docs.taigaprotocol.io/
Youves	Official Documentation	https://docs.youves.com/
Youves	White Paper	https://docs.youves.com/assets/files/Quantitative_Paper_-_Synthetic_USD_on_Tezos-5148c0a85eda3c82147a59f9995009c5.pdf

Table D26: Glossary of technical terms used in perpetual-futures fee tables

Term	Meaning
Ask price	Lowest price at which a seller is currently willing to sell the asset.
Bid price	Highest price at which a buyer is currently willing to purchase the asset.
Index price	Volume-weighted average of spot prices across several exchanges, used as the "fair" reference price for the perpetual contract.
Mid-price	Simple average of best bid and best ask: $\frac{\text{Bid}+\text{Ask}}{2}$.
EWMA	Moving average that gives more weight to recent prices; decay factor is chosen by the protocol (e.g. 1-hour window).
Oracle price	Price delivered by an external data feed (oracle) rather than the protocol's own order book.
Funding fee / funding rate	Hourly (or 8-hour, etc.) payment exchanged between long and short traders to keep the perp price aligned with the index price. Positive funding means longs pay shorts; negative means the opposite.
Borrow fee	Interest charged for borrowing liquidity from the pool (leveraged positions). Computed continuously, settled hourly.
Utilization ratio	Locked liquidity ÷ Total liquidity in pool; indicates how "full" the lending pool is.
Clamp function	$\text{clamp}(x, a, b)$ limits a value x to the interval $[a, b]$; values below a return a, above b return b. Used in Hyperliquid's funding formula.
Price-impact fee	Extra spread charged when a trade is large enough to move the pool price beyond a set threshold.
Maker / taker	Fee schedule where "makers" add liquidity (post orders) and "takers" remove it (hit existing orders); makers usually pay lower fees.
Premium (funding premium)	Relative difference between the perp market price and the index price; determines who pays the funding fee.

Position equity and margining.

$$\text{Equity}_t \;=\; C_0 + \text{UPnL}_t - \Phi_t, \qquad M_{m,t} \;=\; r_m\,\text{NV}_t \tag{11.5}$$

Liquidation is triggered when $\text{Equity}_t \le M_{m,t}$.

D.2 Options (Expiring and Everlasting)

Position value and PnL (mark-to-market). Let n be the number of contracts and κ the contract multiplier (set $\kappa{=}1$ if not applicable). Entry premium per contract is π; mark value per contract at time t is V_t.

$$\text{UPnL}_t = s\,n\kappa\,(V_t - \pi), \tag{11.6}$$

$$\text{RPnL}_t = \text{UPnL}_t - F_{\text{open}} - F_{\text{close}} - \Phi_t. \tag{11.7}$$

At expiry T with strike K:

$$V_T = \begin{cases} \max(P_T - K, 0) & \text{call}, \\ \max(K - P_T, 0) & \text{put}. \end{cases} \tag{11.8}$$

For everlasting options, some designs apply a funding term to align the perpetual option's mark with a benchmark. When present, this accrues in the same way as perps and falls within the scope of the function Φ_t.

D.3 Synthetic Protocols

User posts c units of collateral with price $P_t^{(C)}$ (collateral value $V_t^{\text{coll}} = c\,P_t^{(C)}$) and mints Q^{synth} units of a synthetic with reference price $P_t^{(U)}$ (debt value $V_t^{\text{debt}} \;=\; Q^{\text{synth}} P_t^{(U)}$).

Minimum collateralization:

$$CR_t := \frac{V_t^{\text{coll}}}{V_t^{\text{debt}}} = \frac{c\, P_t^{(C)}}{Q^{\text{synth}} P_t^{(U)}} \;\geq\; CR_{\min}. \tag{11.9}$$

PnL for investors With entry/exit reference prices $P_{\text{entry}}^{(U)}$ and $P_{\text{exit}}^{(U)}$ on a net holding of Q^{synth}:

$$\text{RPnL} = Q^{\text{synth}} \left(P_{\text{exit}}^{(U)} - P_{\text{entry}}^{(U)}\right) - F_{\text{open}} - F_{\text{close}}. \tag{11.10}$$

Liquidation Liquidation triggers when $CR_t \leq CR_{\text{liq}}$; liquidators (or an automated module) burn synths, seize collateral up to a penalty, and target a post-liquidation CR above threshold.

Bibliography

[1] Aave liquidity protocol. URL https://app.aave.com/dashboard.

[2] Band protocol. URL https://bandprotocol.com.

[3] Chainlink: The foundation for onchain finance. URL https://chain.link/.

[4] Dune — crypto analytics powered by community. URL https://dune.com/home.

[5] Everlasting options. URL https://www.paradigm.xyz/2021/05/everlasting-options.

[6] Gauntlet - dual slope borrowing rate model implementation and recommendations. URL https://www.jupresear.ch/t/gauntlet-dual-slope-borrowing-rate-model-implementation-and-recommendations-12-19-24/29072.

[7] JLP economics | jupiter station. URL https://station.jup.ag/guides/jlp/JLP-Economics.

[8] Lido liquid staking. URL https://lido.fi.

[9] Medium. URL https://medium.com/.

[10] Messari. URL https://messari.io/.

[11] Opyn markets | uniswap, but for perps. URL https://www.opyn.co/.

[12] Outcome Finance DefiLlama. URL https://defillama.com/protocol/outcome-finance.

[13] Defillama. https://defillama.com, 2025. Accessed: January 2025.

[14] J. D. Abbey and M. G. Meloy. Attention by Design: Using Attention Checks to Detect Inattentive Respondents and Improve Data Quality. *Journal of Operations Management*, 53:63–70, 2017.

[15] S. Abramova and R. Böhme. Perceived Benefit and Risk as Multidimensional Determinants of Bitcoin Use: A Quantitative Exploratory Study. In *Proceedings of the Thirty Seventh International Conference on Information Systems (ICIS)*, Dublin, Ireland, 2016.

[16] S. Abramova, A. Voskobojnikov, K. Beznosov, and R. Böhme. Bits Under the Mattress: Understanding Different Risk Perceptions and Security Behaviors of Crypto-Asset Users. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. doi: 10.1145/3411764.3445679. URL https://doi.org/10.1145/3411764.3445679.

[17] D. Ackerer, J. Hugonnier, and U. Jermann. Perpetual futures pricing. Technical report, National Bureau of Economic Research, 2024.

[18] A. Adadi and M. Berrada. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6:52138–52160, 2018. ISSN 2169-3536. doi: 10.1109/ACCESS.2018.2870052.

[19] A. Afianian, S. Niksefat, B. Sadeghiyan, and D. Baptiste. Malware dynamic analysis evasion techniques: A survey. *ACM Comput. Surv.*, 52(6), nov 2019. ISSN 0360-0300. doi: 10.1145/3365001.

[20] F. Ahmmed, B. Y. Boadi, and M. Guillemette. Margin trading and cryptocurrency investment among us investors: Evidence from the national financial capability study. *Journal of Risk and Financial Management*, 18(7):373, 2025.

[21] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2623–2631, 2019.

[22] H. Al-Breiki, M. H. U. Rehman, K. Salah, and D. Svetinovic. Trustworthy Blockchain Oracles: Review, Comparison, and Open Research Challenges. *IEEE Access*, 8:85675–85685, 2020. ISSN 2169-3536. doi: 10.1109/ACCESS.2020.2992698. Conference Name: IEEE Access.

[23] C. Alexander, J. Choi, H. Park, and S. Sohn. Bitmex bitcoin derivatives: Price discovery, informational efficiency, and hedging effectiveness. *Journal of Futures Markets*, 40(1):23–43, 2020.

[24] C. Alexander, J. Deng, and B. Zou. Hedging with automatic liquidation and leverage selection on bitcoin futures. *European Journal of Operational Research*, 306(1): 478–493, 2023.

[25] S. Ali, T. Abuhmed, S. El-Sappagh, K. Muhammad, J. M. Alonso-Moral, R. Confalonieri, R. Guidotti, J. Del Ser, N. Díaz-Rodríguez, and F. Herrera. Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence. *Information fusion*, 99:101805, 2023.

[26] S. L. N. Alonso, J. Jorge-Vázquez, P. A. Rodríguez, and B. M. S. Hernández. Gender Gap in the Ownership and use of Cryptocurrencies: Empirical Evidence from Spain. *Journal of Open Innovation: Technology, Market, and Complexity*, 9(3-100103), 2023.

[27] A. Aloosh, S. Ouzan, and S. J. H. Shahzad. Bubbles across Meme Stocks and Cryptocurrencies. *Finance Research Letters*, 49:103155, 2022.

[28] A. Andolfatto, S. Naik, and L. Schönleber. Decentralized and centralized options trading: A risk premia perspective.

[29] E. Anduiza Perea, A. Gallego Dobón, and L. Jorba. Internet use and the political knowledge gap in spain. *Revista Internacional de Sociología*, 70(1):0129–151, 2012.

[30] G. Angeris, T. Chitra, A. Evans, and M. Lorig. A primer on perpetuals. *SIAM Journal on Financial Mathematics*, 14(1):SC17–SC30, 2023.

[31] E. Anton, M. Aptyka, T. D. Oesterreich, and F. Teuteberg. To the Moon with Dogecoin! Disentangling the Causalities behind Extrinsic and Intrinsic Motivations for Memecoin Investments. *Journal of Decision Systems*, pages 1–35, 2024.

[32] Z. Ao, L. W. Cong, G. Horvath, and L. Zhang. Is decentralized finance actually decentralized? a social network analysis of the aave protocol on the ethereum blockchain. *CoRR*, abs/2206.08401, 2022. doi: 10.48550/arXiv.2206.08401. URL https://doi.org/10.48550/arXiv.2206.08401.

[33] N. Aoki, T. Tatsumi, G. Naruse, and K. Maeda. Explainable ai for government: Does the type of explanation matter to the accuracy, fairness, and trustworthiness of an algorithmic decision as perceived by those who are affected? *Government Information Quarterly*, 41(4):101965, 2024.

[34] M. Aquilina, G. Cornelli, J. Frost, and L. Gambacorta. Cryptocurrencies and Decentralised Finance: Functions and Financial Stability Implications. Technical Report BIS Paper No. 156, Bank for International Settlements, April 2025. URL https://www.bis.org/publ/bppdf/bispap156.pdf. Accessed: 2025-07-15.

[35] M. Aria, C. Cuccurullo, and A. Gnasso. A comparison among interpretative proposals for random forests. *Machine Learning with Applications*, 6:100094, 2021. ISSN 2666-8270. doi: 10.1016/j.mlwa.2021.100094.

[36] L. Arrighi, L. Pennella, G. Marques Tavares, and S. Barbon Junior. Decision predicate graphs: Enhancing interpretability in tree ensembles. In *World Conference on Explainable Artificial Intelligence*, pages 311–332. Springer, 2024.

[37] R. Auer and S. Claessens. Regulating Cryptocurrencies: Assessing Market Reactions. *BIS Quarterly Review September 2018*, pages 51–65, 2018.

[38] R. Auer and D. Tercero-Lucas. Distrust or Speculation? The Socioeconomic Drivers of US Cryptocurrency Investments. *Journal of Financial Stability*, 62: 101066, 2022.

[39] R. Auer, B. Haslhofer, S. Kitzler, P. Saggese, and F. Victor. The technology of decentralized finance (defi). *Digital Finance*, 6(1):55–95, 2024.

[40] P. D. Azar, G. Baughman, F. Carapella, J. Gerszten, A. Lubis, J. P. Perez-Sangimino, D. E. Rappoport W, C. Scotti, N. Swem, A. Vardoulakis, et al. The Financial Stability Implications of Digital Assets. *Economic Policy Review*, 30(2):1–48, 2024.

[41] N. Balasubramaniam, M. Kauppinen, A. Rannisto, K. Hiekkanen, and S. Kujala. Transparency and explainability of ai systems: From ethical guidelines to requirements. *Information and Software Technology*, 159:107197, 2023. ISSN 0950-5849.

[42] S. Balietti. nodeGame: Real-time, Synchronous, Online Experiments in the Browser. *Behavior Research Methods*, 49:1696–1715, 2017.

[43] S. Balietti. From Online Experiments to Big Experimental Data. *Balietti, S.(2023). From Online Experiments to Big Experimental Data. In: T. Yasseri (Ed.), Handbook of Computational Social Science. Edward Elgar Publishing Ltd*, 2022.

[44] S. Balietti, C. Celebi, and D. Tercero-Lucas. From Crypto to NFTs: Identifying the New Wave of Digital Investors. *International Review of Financial Analysis*, 104: 104172, 2025. ISSN 1057-5219. doi: https://doi.org/10.1016/j.irfa.2025.104172.

[45] S. Balietti, C. Celebi, and D. Tercero-Lucas. From Crypto to NFTs: Identifying the New Wave of Digital Investors. *International Review of Financial Analysis*, 104: 104172, 2025.

[46] S. Balietti, C. Celebi, L. Pennella, and D. Tercero-Lucas. Meme money, real people: Decoding the crypto memecoin crowd. *Available at SSRN 6021706*, 2026. doi: http://dx.doi.org/10.2139/ssrn.6021706.

[47] B. M. Barber and T. Odean. Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics*, 116(1):261–292, 2001.

[48] M. Bartoletti, J. H.-y. Chiang, and A. L. Lafuente. SoK: Lending pools in decentralized finance. In *Financial Cryptography and Data Security. FC 2021 International Workshops*, pages 553–578. Springer. ISBN 978-3-662-63958-0. doi: 10.1007/978-3-662-63958-0_40.

[49] M. Bartoletti, B. Pes, and S. Serusi. Data mining for detecting bitcoin ponzi schemes. *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*, pages 75–84, 2018.

[50] M. Bartoletti, S. Carta, T. Cimoli, and R. Saia. Dissecting ponzi schemes on ethereum: Identification, analysis, and impact. *Future Gener. Comput. Syst.*, 102: 259–277, 2020. doi: 10.1016/j.future.2019.08.014.

[51] M. Bartoletti, S. Lande, A. Loddo, L. Pompianu, and S. Serusi. Cryptocurrency scams: Analysis and perspectives. *IEEE Access*, 9:148353–148373, 2021. doi: 10.1109/ACCESS.2021.3123894.

[52] M. Bartoletti, J. H.-y. Chiang, and A. Lluch-Lafuente. A theory of automated market makers in defi. *Logical Methods in Computer Science*, 18, 2022.

[53] A. Benasaglio Berlucchi. Populism without host ideologies: A new home for voters with exclusionary attitudes in italy's five star movement? *Party politics*, 28(5): 811–825, 2022.

[54] C. J. Bickerton and C. I. Accetti. 'techno-populism'as a new party family: the case of the five star movement and podemos. *Contemporary Italian Politics*, 10(2): 132–150, 2018.

[55] BitMex. Perpetual contracts guide - BitMEX, 2025. URL https://www.bitmex.com/app/perpetualContractsGuide.

[56] F. Bloise, D. Chironi, and M. Pianta. Inequality and voting in italy's regions. *Territory, Politics, Governance*, 9(3):365–390, 2021.

[57] F. Bloise, D. Chironi, D. della Porta, and M. Pianta. Inequality and elections in italy, 1994–2018. *Italian Economic Journal*, 10(1):1–23, 2024.

[58] F. Bodria, F. Giannotti, R. Guidotti, F. Naretto, D. Pedreschi, and S. Rinzivillo. Benchmarking and survey of explanation methods for black box models. *Data Mining and Knowledge Discovery*, 37(5):1719–1778, 2023.

[59] R. R. Bouckaert and E. Frank. Evaluating the replicability of significance tests for comparing learning algorithms. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 3–12, Heidelberg, 2004. Springer.

[60] E. Bouma-Sims, H. Hassan, A. Nisenoff, L. F. Cranor, and N. Christin. "It was honestly just gambling": Investigating the Experiences of Teenage Cryptocurrency Users on Reddit. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*, pages 333–352, Philadelphia, PA, Aug. 2024. USENIX Association. ISBN 978-1-939133-42-7.

[61] U. Brandes. On variants of shortest-path betweenness centrality and their generic computation. 30(2):136–145, 2008. ISSN 0378-8733. doi: 10.1016/j.socnet.2007.11.001.

[62] J. W. Brehm. *A Theory of Psychological Reactance*. Academic Press, 1966.

[63] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[64] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324.

[65] M. Brichta. Fanning money: the cultural economy and participatory politics of dogecoin. *International Journal of Communication*, 17:19, 2023.

[66] A. Brini and J. Lenz. Pricing cryptocurrency options with machine learning regression for handling market volatility. *Economic Modelling*, 136:106752, 2024.

[67] B. Burscher, R. Vliegenthart, and C. H. De Vreese. Using supervised machine learning to code policy issues: Can classifiers generalize across contexts? *The ANNALS of the American Academy of Political and Social Science*, 659(1):122–131, 2015.

[68] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock. Explainable machine learning in credit risk management. *Computational Economics*, 57(1):203–216, 2021.

[69] V. Buterin. A Next-Generation Smart Contract and Decentralized Application Platform. *Ethereum White Paper, https://ethereum.org/en/whitepaper/ (site accessed on 05.29.2021)*, 2013.

[70] V. Buterin, D. Hernandez, T. Kamphefner, K. Pham, Z. Qiao, D. Ryan, J. Sin, Y. Wang, and Y. X. Zhang. Combining GHOST and casper. *CoRR*, abs/2003.03052, 2020. doi: 10.48550/arXiv.2003.03052. URL https://arxiv.org/abs/2003.03052.

[71] J. Cai, B. Li, J. Zhang, and X. Sun. Ponzi scheme detection in smart contract via transaction semantic representation learning. *IEEE Transactions on Reliability*, 2023. doi: 10.1109/TR.2023.3319318.

[72] M. Caiani and E. Padoan. Populism and the (italian) crisis: The voters and the context. *Politics*, 41(3):334–350, 2021.

[73] C. Campajola, R. Cristodaro, F. M. D. Collibus, T. Yan, N. Vallarano, and C. J. Tessone. The evolution of centralisation on cryptocurrency platforms. *CoRR*, abs/2206.05081, 2022. doi: 10.48550/arXiv.2206.05081. URL https://doi.org/10.48550/arXiv.2206.05081.

[74] C. Campajola, M. D'Errico, and C. J. Tessone. Microvelocity: rethinking the velocity of money for digital currencies, 2023.

[75] J. Campino and S. Yang. Decoding the cryptocurrency user: An analysis of demographics and sentiments. *Heliyon*, 10(5), 2024.

[76] F. Carapella and N. Swem. Decentralized finance (defi): Transformative potential & associated risks. Finance and Economics Discussion Series 2022-057, Board of Governors of the Federal Reserve System, 2022. URL https://www.federalreserve.gov/econres/feds/files/2022057pap.pdf.

[77] F. Carozzi. Credit constraints and the composition of housing sales. farewell to first-time buyers? *Journal of the European Economic Association*, 18(3):1196–1237, 2020.

[78] C. Carpentier-Desjardins, M. Paquet-Clouston, S. Kitzler, and B. Haslhofer. Mapping the DeFi Crime Landscape: An Evidence-based Picture. *Journal of Cybersecurity*, 11(1):1–19, 2025.

[79] L. Carrieri. The limited politicization of european integration in italy: lacking issue clarity and weak voter responses. *Italian Political Science Review/Rivista Italiana di Scienza Politica*, 50(1):52–69, 2020.

[80] A. Ceron and G. d'Adda. E-campaigning on twitter: The effectiveness of distributive promises and negative campaign in the 2013 italian election. *New Media & Society*, 18(9):1935–1955, 2016.

[81] Chainalysis. The 2025 Crypto Crime Report. Annual report, New York, NY, USA, feb 2025. Available at: https://www.chainalysis.com/reports/2025-crypto-crime-report/.

[82] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16: 321–357, 2002.

[83] E. Chen, M. Ma, and Z. Nie. Perpetual future contracts in centralized and decentralized exchanges: Mechanism and traders' behavior. *Electronic Markets*, 34(1): 35, 2024.

[84] T. Chen and C. Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, 2016. Association for Computing Machinery. ISBN 978-1-4503-4232-2.

[85] W. Chen, Z. Zheng, J. Cui, E. Ngai, P. Zheng, and Y. Zhou. Detecting ponzi schemes on ethereum: Towards healthier blockchain technology. In *Proceedings of the 2018 World Wide Web Conference*, pages 1409–1418, 2018. doi: 10.1145/3178876.3186046.

[86] W. Chen, Z. Zheng, E. C.-H. Ngai, P. Zheng, and Y. Zhou. Exploiting blockchain data to detect smart ponzi schemes on ethereum. *IEEE Access*, 7:37575–37586, 2019. doi: 10.1109/ACCESS.2019.2905769.

[87] W. Chen, X. Li, Y. Sui, N. He, H. Wang, L. Wu, and X. Luo. Sadponzi: Detecting and characterizing ponzi schemes in ethereum smart contracts. *Proc. ACM Meas. Anal. Comput. Syst.*, 5(2), 2021. doi: 10.1145/3460093.

[88] Y. Chen, P. Giudici, K. Liu, and E. Raffinetti. Measuring fairness in credit ratings. *Expert Systems with Applications*, 258:125184, 2024.

[89] Z. Cheng, J. Deng, T. Wang, and M. Yu. Liquidation, leverage and optimal margin in bitcoin futures markets. *Applied Economics*, 53(47):5415–5428, 2021.

[90] R. Chimatapu, H. Hagras, A. Starkey, and G. Owusu. Explainable AI and fuzzy logic systems. In D. Fagan, C. Martín-Vide, M. O'Neill, and M. A. Vega-Rodríguez, editors, *Theory and Practice of Natural Computing*, pages 3–20. Springer International Publishing, 2018. ISBN 978-3-030-04070-3. doi: 10.1007/978-3-030-04070-3_1.

[91] H. Chipman, E. George, and R. McCulloch. Making sense of a forest of trees. *Proceedings of the 30th Symposium on the Interface*, 29, 1998.

[92] T. Chitra and A. Evans. Why stake when you can borrow? *CoRR*, abs/2006.11156, 2020. doi: 10.48550/arXiv.2006.11156. URL https://arxiv.org/abs/2006.11156.

[93] A. Chokor and E. Alfieri. Long and Short-term Impacts of Regulation in the Cryptocurrency Market. *The Quarterly Review of Economics and Finance*, 81:157–173, 2021.

[94] F. M. D. Collibus. *The ethereum ecosystem from a transaction network perspective*. PhD thesis, University of Zurich, Zürich, April 2024.

[95] F. M. D. Collibus, C. Campajola, and C. J. Tessone. The microvelocity of money in ethereum. *EPJ Data Sci.*, 14(1):11, 2025. doi: 10.1140/epjds/s13688-024-00518-6. URL https://doi.org/10.1140/epjds/s13688-024-00518-6.

[96] J. A. Colombo and L. Yarovaya. Are Crypto and Non-crypto Investors Alike? Evidence from a Comprehensive Survey in Brazil. *Technology in Society*, (102468), 2024.

[97] T. Conlon, S. Corbet, and Y. Hu. The Collapse of the FTX Exchange: The End of Cryptocurrency's Age of Innocence. *The British Accounting Review*, page 101277, 2023. ISSN 0890-8389. doi: https://doi.org/10.1016/j.bar.2023.101277.

[98] M. Cortes-Goicoechea, T. Mohandas-Daryanani, J. L. Muñoz-Tapia, and L. Bautista-Gomez. Autopsy of ethereum's post-merge reward system. In *IEEE International Conference on Blockchain and Cryptocurrency, ICBC 2023, Dubai, United Arab Emirates, May 1-5, 2023*, pages 1–9. IEEE, 2023. doi: 10.1109/ICBC56567.2023.10174942. URL https://doi.org/10.1109/ICBC56567.2023.10174942.

[99] S. Cousaert, J. Xu, and T. Matsui. Sok: Yield aggregators in defi. In *2022 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, pages 1–14. IEEE, 2022.

[100] P. Crosetto and A. Filippin. The "bomb" Risk Elicitation Task. *Journal of Risk and Uncertainty*, 47:31–65, 2013.

[101] P. Cuffe. The Role of the ERC-20 Token Standard in a Financial Revolution: the Case of Initial Coin Offerings. *IEC-IEEE-KATS Academic Challenge*, 2018.

[102] L. Davis and S. S. Deole. Immigration and the rise of far-right parties in europe. *ifo DICE Report*, 15(4):10–15, 2017.

[103] R. De Blasis and A. Webb. Arbitrage, contract design, and market structure in bitcoin futures markets. *Journal of Futures Markets*, 42(3):492–524, 2022.

[104] F. M. De Collibus, C. Campajola, G. Caldarelli, and C. J. Tessone. Patterns and centralisation in ethereum-based token transaction networks. *Frontiers in Physics*, 12, 2024. doi: https://doi.org/10.3389/fphy.2024.1305167.

[105] D. De Cremer and S. L. Blader. Why Do People Care about Procedural Fairness? The Importance of Belongingness in Responding and Attending to Procedures. *European Journal of Social Psychology*, 36(2):211–228, 2006.

[106] K. Dedja, F. K. Nakano, K. Pliakos, and C. Vens. BELLATREX: Building explanations through a LocaLly AccuraTe rule EXtractor. *IEEE Access*, 11:41348 – 41367, 2023. ISSN 2169-3536. doi: 10.1109/ACCESS.2023.3268866.

[107] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, 7(Jan):1–30, 2006.

[108] H. Deng. Interpreting tree ensembles with inTrees. *International Journal of Data Science and Analytics*, 7(4):277–287, 2019. ISSN 2364-4168. doi: 10.1007/s41060-018-0144-8.

[109] T. G. Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation*, 10(7):1895–1923, 1998.

[110] T. Do, T.-A. Pham, and T. Tran. Novel perpetual futures market model based on a family of asymptotic power curves. In *International Conference on Blockchain*, pages 69–83. Springer, 2024.

[111] G. D'Souza, H. Zhang, W. D'Souza, R. Meyer, and M. Gillison. Moderate predictive value of demographic and behavioral characteristics for a diagnosis of HPV16-positive and HPV16-negative head and neck cancer. 46(2):100–104, 2010.

[112] D. Dupuis, D. Smith, and K. Gleason. Old Frauds with a New Sauce: Digital Assets and Space Transition. *Journal of Financial Crime*, 30(1):205–220, 2023.

[113] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, and R. Ranjan. Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*, 55(9):194:1–194:33, 2023. ISSN 0360-0300. doi: 10.1145/3561048.

[114] S. Elhishi, A. M. Elashry, and S. El-Metwally. Unboxing machine learning models for concrete strength prediction using xai. *Scientific Reports*, 13(1):19892, 2023.

[115] S. Eskandari, M. Salehi, W. C. Gu, and J. Clark. Sok: Oracles from the ground truth to market manipulation. In *Proceedings of the 3rd ACM Conference on Advances in Financial Technologies*, pages 127–141, 2021.

[116] Ethereum Foundation. The risks of lsd. https://notes.ethereum.org/@djrtwo/risks-of-lsd, 2022. URL https://notes.ethereum.org/@djrtwo/risks-of-lsd. Accessed: 2025-08-06.

[117] European Parliament and Council of the European Union. Regulation (EU) 2023/1114 of 31 May 2023 on markets in crypto-assets (MiCA), June 2023. https://eur-lex.europa.eu/eli/reg/2023/1114/oj/eng.

[118] S. Fan, S. Fu, H. Xu, and C. Zhu. Expose your mask: Smart ponzi schemes detection on blockchain. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, 2020. doi: 10.1109/IJCNN48605.2020.9207143.

[119] X. Feng, Q. Shi, X. Li, H. Liu, and L. Wang. Idponzi: An interpretable detection model for identifying smart ponzi schemes. *Engineering Applications of Artificial Intelligence*, 136:108868, 2024. ISSN 0952-1976.

[120] X. Feng, Q. Shi, X. Li, H. Liu, and L. Wang. IDPonzi: An interpretable detection model for identifying smart ponzi schemes. *Engineering Applications of Artificial Intelligence*, 136:108868, 2024. doi: 10.1016/j.engappai.2024.108868.

[121] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera. *Learning from imbalanced data sets*, volume 10. Springer, 2018.

[122] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936. ISSN 2050-1439. doi: 10.1111/j.1469-1809.1936.tb02137.x.

[123] A. M. Florio, P. Martins, M. Schiffer, T. Serra, and T. Vidal. Optimal decision diagrams for classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 7577–7585, 2023. doi: 10.1609/aaai.v37i6.25920.

[124] F. Franchino and F. Negri. The fiscally moderate italian populist voter: Evidence from a survey experiment. *Party Politics*, 26(2):176–190, 2020.

[125] M. N. Franklin. *Voter Turnout and the Dynamics of Electoral Competition in Established Democracies since 1945*. Cambridge University Press, Cambridge, 2004.

[126] J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[127] H. Fujiki. Who Adopts Crypto Assets in Japan? Evidence From the 2019 Financial Literacy Survey. *Journal of the Japanese and International Economies*, 58:101107, 2020.

[128] L. Galati and S. Perdichizzi. From Zero to Hero: Memecoins' Spillover Effects in Cryptocurrency Markets. *Economics Letters*, page 112381, 2025.

[129] L. Galati, A. Webb, and R. I. Webb. Financial Contagion in Cryptocurrency Exchanges: Evidence from the FTT Collapse. *Finance Research Letters*, 67:105747, 2024. ISSN 1544-6123. doi: https://doi.org/10.1016/j.frl.2024.105747.

[130] L. Galletta and F. Pinelli. Explanable ponzi schemes detection on ethereum. In *The 39th ACM/SIGAPP Symposium On Applied Computing (SAC '24)*, page 1014–1023, 2024. doi: 10.1145/3605098.3636060.

[131] C. Gamble. The Legality and Regulatory Challenges of Decentralised Cryptocurrency: A Western Perspective. *Int'l Trade & Bus. L. Rev.*, 20:346, 2017.

[132] A. Gangwal, H. R. Gangavalli, and A. Thirupathi. A survey of layer-two blockchain protocols. *J. Netw. Comput. Appl.*, 209:103539, 2023. doi: 10.1016/J.JNCA.2022.103539.

[133] R. K. Gibson and I. McAllister. Online social ties and political engagement. *Journal of Information Technology & Politics*, 10(1):21–34, 2013.

[134] K. Gogol, B. Kraner, M. Schlosser, T. Yan, C. J. Tessone, and B. Stiller. Empirical and theoretical analysis of liquid staking protocols. *CoRR*, abs/2401.16353, 2024. doi: 10.48550/arXiv.2401.16353. URL https://doi.org/10.48550/arXiv.2401.16353.

[135] K. Gogol, Y. Velner, B. Kraner, and C. Tessone. Sok: Liquid staking tokens (lsts) and emerging trends in restaking. *arXiv preprint arXiv:2404.00644*, 2024.

[136] A. C. Goldberg, E. J. van Elsas, and C. H. De Vreese. The differential impact of eu attitudes on voting behaviour in the european parliamentary elections 2019. *Journal of Contemporary European Studies*, 32(4):1323–1342, 2024.

[137] F. Gossen and B. Steffen. Algebraic aggregation of random forests: towards explainability and rapid evaluation. *International Journal on Software Tools for Technology Transfer*, 25(3):1–19, 2021. ISSN 1433-2787. doi: 10.1007/s10009-021-00635-x.

[138] C. Goutte and E. Gaussier. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *European conference on information retrieval*, pages 345–359, Heidelberg, 2005. Springer.

[139] J. Grimmer, M. E. Roberts, and B. M. Stewart. Machine learning for social science: An agnostic approach. *Annual Review of Political Science*, 24:395–419, 2021. doi: 10.1146/annurev-polisci-053119-015921.

[140] H. Guan, L. Dong, and A. Zhao. Ethical risk factors and mechanisms in artificial intelligence decision making. *Behavioral Sciences*, 12(9):343, 2022.

[141] L. Gudgeon, S. Werner, D. Perez, and W. J. Knottenbelt. DeFi protocols for loanable funds: Interest rates, liquidity and market efficiency. In *Proceedings of the 2nd ACM Conference on Advances in Financial Technologies*, AFT '20, pages 92–112. Association for Computing Machinery.

[142] L. Gudgeon, S. Werner, D. Perez, and W. J. Knottenbelt. Defi protocols for loanable funds: Interest rates, liquidity and market efficiency. In *Proceedings of the 2nd ACM Conference on Advances in Financial Technologies*, pages 92–112, 2020.

[143] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM Comput. Surv.*, 51(5), Aug. 2018.

[144] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM Comput. Surv.*, 51(5), 2018. ISSN 0360-0300. doi: 10.1145/3236009.

[145] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM Computing Surveys*, 51 (5):93, 2019. doi: 10.1145/3236009.

[146] B. Gulowaty and M. Woźniak. Extracting interpretable decision tree ensemble from random forest. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2021.

[147] M. Haddouchi and A. Berrado. A survey of methods and tools used for interpreting random forest. In *2019 1st International Conference on Smart Systems and Data Science (ICSSD)*, pages 1–6, 2019. doi: 10.1109/ICSSD47982.2019.9002770.

[148] J. Han, J. Lee, and T. Li. A review of dao governance: Recent literature and emerging trends. *Journal of Corporate Finance*, page 102734, 2025.

[149] A. Hanif, X. Zhang, and S. Wood. A survey on explainable artificial intelligence techniques and challenges. In *2021 IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW)*, pages 81–89, 2021. doi: 10.1109/EDOCW52865.2021.00036. ISSN: 2325-6605.

[150] S. Hara and K. Hayashi. Making tree ensembles interpretable: A bayesian model selection approach. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, pages 77–85. PMLR, 2018. ISSN: 2640-3498.

[151] C. R. Harvey. *DeFi and the Future of Finance*. John Wiley & Sons, 2021.

[152] T. Hastie, R. Tibshirani, and J. Friedman. Additive models, trees, and related methods. In T. Hastie, R. Tibshirani, and J. Friedman, editors, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, pages 295–336. Springer, 2009. ISBN 978-0-387-84858-7. doi: 10.1007/978-0-387-84858-7_9.

[153] J. Hatwell, M. M. Gaber, and R. M. A. Azad. CHIRPS: Explaining random forest classification. *Artificial Intelligence Review*, 53(8):5747–5788, 2020. ISSN 1573-7462. doi: 10.1007/s10462-020-09833-6.

[154] D. J. Hauser, P. C. Ellsworth, and R. Gonzalez. Are Manipulation Checks Necessary? *Frontiers in Psychology*, 9, 2018. doi: 10.3389/fpsyg.2018.00998.

[155] S. He, A. Manela, O. Ross, and V. von Wachter. Fundamentals of perpetual futures. *arXiv preprint arXiv:2212.06888*, 2022.

[156] C. S. Henry, K. P. Huynh, and G. Nicholls. Bitcoin Awareness and Usage in Canada. *Journal of Digital Banking*, 2(4):311–337, 2018.

[157] T. K. Ho. Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282 vol.1, 1995.

[158] T. K. Ho. Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282 vol.1, 1995. doi: 10.1109/ICDAR.1995.598994.

[159] T. K. Ho and M. Basu. Complexity measures of supervised classification problems. 24(3):289–300. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[160] R. Hoechenberger, D. Hummel, and J. Seitz. Do women shy away from cryptocurrency investment? cross-country evidence from survey data. In *International Conference on Data Management, Analytics & Innovation*, pages 69–76. Springer, 2023.

[161] A. O. Hoffmann, T. Post, and J. M. Pennings. How investor perceptions drive actual trading and risk-taking behavior. *Journal of Behavioral Finance*, 16(1):94–103, 2015.

[162] S. D. Hughes. Cryptocurrency Regulations and Enforcement in the US. *W. St. UL Rev.*, 45:1, 2017.

[163] J. C. Hull and S. Basu. *Options, futures, and other derivatives*. Pearson Education India, 2016.

[164] T. Hulsen. Explainable artificial intelligence (xai): Concepts and challenges in healthcare. *AI*, 4(3):652–666, 2023. ISSN 2673-2688.

[165] M. C. Iban and S. S. Bilgilioglu. Snow avalanche susceptibility mapping using novel tree-based machine learning algorithms (xgboost, ngboost, and lightgbm) with explainable artificial intelligence (xai) approach. *Stochastic Environmental Research and Risk Assessment*, 37(6):2243–2270, 2023.

[166] G. Ibba, G. A. Pierro, and M. Di Francesco. Evaluating machine-learning techniques for detecting smart ponzi schemes. In *2021 IEEE/ACM 4th International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB)*, pages 34–40, 2021. doi: 10.1109/WETSEB52558.2021.00012.

[167] D. Ignatov and A. Ignatov. Decision stream: Cultivating deep decision trees. In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 905–912. IEEE, 2017. doi: 10.1109/ICTAI.2017.00140.

[168] R. F. Inglehart and P. Norris. Trump, brexit, and the rise of populism: Economic have-nots and cultural backlash. *HKS Working paper no. RWP16-026*, 2016.

[169] A. Islam. The usefulness of value-based variables in predicting voting intentions. *Journal of Political Studies*, 45(3):123–145, 2022.

[170] Istat. Istat.it - results of the permanent population census. URL https://www.istat.it/en/censuses/population-and-housing/results.

[171] O. I. Jacinta, A. E. Omolara, M. Alawida, O. I. Abiodun, and A. Alabdultif. Detection of ponzi scheme on ethereum using machine learning algorithms. *Scientific Reports*, 13(1):18403, 2023. doi: 10.1038/s41598-023-45275-0.

[172] E. Jiang, B. Qin, Q. Wang, Z. Wang, Q. Wu, J. Weng, X. Li, C. Wang, Y. Ding, and Y. Zhang. Decentralized finance (DeFi): A survey. *arXiv preprint arXiv:2308.05282*, 2023. URL https://arxiv.org/abs/2308.05282.

[173] P. Jiang, H. Suzuki, and T. Obi. Xai-based cross-ensemble feature ranking methodology for machine learning models. *International Journal of Information Technology*, 15(4):1759–1768, 2023.

[174] C. Jin, J. Jin, J. Zhou, J. Wu, and Q. Xuan. Heterogeneous feature augmentation for ponzi detection in ethereum. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 69(9):3919–3923, 2022. doi: 10.1109/TCSII.2022.3177898.

[175] J. Jin, J. Zhou, C. Jin, S. Yu, Z. Zheng, and Q. Xuan. Dual-channel early warning framework for ethereum ponzi schemes. In *Big Data and Social Computing: 7th China National Conference, BDSC 2022*, pages 260–274. Springer, 2022. doi: 10.1007/978-981-19-7532-5_17.

[176] H. Joebges, H. Herr, and C. Kellermann. Crypto Assets as a Threat to Financial Market Stability. *Eurasian Economic Review*, pages 1–30, 2025.

[177] S. Jordan, H. L. Paul, and A. Q. Philips. How to cautiously uncover the "black box" of machine learning models for legislative scholars. *Legislative Studies Quarterly*, 48(1):165–202, 2023.

[178] A. Kalacheva, P. Kuznetsov, I. Vodolazov, and Y. Yanovich. Detecting Rug Pulls in Decentralized Exchanges: The Rise of Meme Coins. *Available at SSRN 4981529*, 2024.

[179] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.

[180] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. LightGBM: a highly efficient gradient boosting decision tree. In *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, pages 3149–3157, Red Hook, 2017. Curran Associates Inc.

[181] H. J. Kim, J. S. Hong, H. C. Hwang, S. M. Kim, and D. H. Han. Comparison of Psychological Status and Investment Style between Bitcoin Investors and Share Investors. *Frontiers in Psychology*, 11:502295, 2020.

[182] J. H. Kim and H. W. Park. Identifying Networked Patterns in Memecoin Twitter Accounts Using Exponential Random Graph Modeling. *IT Professional*, 25(6):82–89, 2024.

[183] S.-y. S. Kim, R. M. Alvarez, and C. M. Ramirez. Who voted in 2016? using fuzzy forests to understand voter turnout. *Social Science Quarterly*, 101(2):978–988, 2020.

[184] E. Kirchler, E. Hoelzl, and I. Wahl. Enforced versus voluntary tax compliance: The "slippery slope" framework. *Journal of Economic psychology*, 29(2):210–225, 2008.

[185] S. Kitzler, S. Balietti, P. Saggese, B. Haslhofer, and M. Strohmaier. The governance of decentralized autonomous organizations: A study of contributors' influence, networks, and shifts in voting power. In *International Conference on Financial Cryptography and Data Security*, pages 313–330. Springer, 2024.

[186] O. Knight. Javier Milei Backtracks on $4.4B Memecoin After 'Insiders' Pocket $87M. *CoinDesk*, 2025. URL https://www.coindesk.com/business/2025/02/15/javier-milei-backtracks-on-usd4-4b-memecoin-after-insiders-pocket-usd87m. Accessed: 2025-06-13, available at: https://www.coindesk.com/business/2025/02/15/javier-milei-backtracks-on-usd4-4b-memecoin-after-insiders-pocket-usd87m.

[187] J. Kończal. Pricing options on the cryptocurrency futures contracts. *arXiv preprint arXiv:2506.14614*, 2025.

[188] B. Kraner, L. Pennella, N. Vallarano, and C. J. Tessone. Money in Motion: Micro-Velocity and Usage of Ethereum's Liquid Staking Tokens. In *7th Conference*

on Advances in Financial Technologies (AFT 2025), volume 354 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 9:1–9:18, Dagstuhl, Germany, 2025. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi: 10.4230/LIPIcs.AFT.2025.9.

[189] D. Krause. Beyond the Hype: A Meme Coin Reality Check for Retail Investors. *SSRN Electronic Journal*, 2024.

[190] D. Krause. The $1.4 Billion Bybit Hack: Cybersecurity Failures and the Risks of Cryptocurrency Deregulation, 2025.

[191] D. Krause. Risks of Investing in Meme Coins: A Case Study of the $TRUMP Coin. *SSRN Electronic Journal*, 2025.

[192] S. Kremer, L. Pennella, M. K. Yurdabak, and S. Balietti. Public perceptions of cryptomarket regulation: Investor profiles and attitudes. *Available at SSRN 5557139*, 2025. doi: http://dx.doi.org/10.2139/ssrn.5557139.

[193] K. Krombholz, A. Judmayer, M. Gusenbauer, and E. Weippl. The Other Side of the Coin: User Experiences with Bitcoin Security and Privacy. In *Financial Cryptography and Data Security (FC)*, pages 555–580, Berlin, Heidelberg, 2017. Springer. doi: 10.1007/978-3-662-54970-4_33. URL https://fc16.ifca.ai/preproceedings/33_Krombholz.pdf.

[194] E. Lansiaux, N. Tchagaspanian, and J. Forget. Community impact on a Cryptocurrency: Twitter Comparison example between Dogecoin and Litecoin. *Frontiers in Blockchain*, 5:829865, 2022.

[195] M. Laver and N. Schofield. *Multiparty government: The politics of coalition in Europe*. University of Michigan Press, Ann Arbor, 1998.

[196] M. R. Leary and K. M. Kelly. Belonging motivation. *Handbook of individual differences in social behavior*, 400409, 2009.

[197] J. Lee and F. L'heureux. A Regulatory Framework for Cryptocurrency. *European Business Law Review*, 31(3), 2020.

[198] S. Littrell, C. Klofstad, and J. E. Uscinski. The Political, Psychological, and Social correlates of Cryptocurrency Ownership. *PloS one*, 19(7):e0305178, 2024.

[199] X. F. Liu, X.-J. Jiang, S.-H. Liu, and C. K. Tse. Knowledge discovery in cryptocurrency transactions: A survey. *IEEE Access*, 9:37229–37254, 2021. doi: 10.1109/ACCESS.2021.3062652.

[200] H.-W. Long, H. Li, and W. Cai. CoinCLIP: A Multimodal Framework for Evaluating the Viability of Memecoins in the Web3 Ecosystem. *arXiv preprint arXiv:2412.07591*, 2024.

[201] H.-W. Long, N.-M. Wong, and W. Cai. Bridging Culture and Finance: A Multimodal Analysis of Memecoins in the Web3 Ecosystem. *arXiv preprint arXiv:2412.04913*, 2024.

[202] H.-W. Long, N.-M. Wong, and W. Cai. Bridging culture and finance: A multimodal analysis of memecoins in the web3 ecosystem. In *Companion Proceedings of the ACM on Web Conference 2025*, pages 1158–1161, 2025.

[203] Y. Lou, Y. Zhang, and S. Chen. Ponzi contracts detection based on improved convolutional neural network. In *2020 IEEE International Conference on Services Computing (SCC)*, pages 353–360, 2020. doi: 10.1109/SCC49832.2020.00053.

[204] I. Lundberg, J. E. Brand, and T. Jeon. Researcher reasoning meets computational capacity: Machine learning for social science. *Social Science Research*, 108:102807, 2022. doi: 10.1016/j.ssresearch.2022.102807.

[205] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 4768–4777. Curran Associates Inc., 2017. ISBN 978-1-5108-6096-4.

[206] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

[207] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 4768–4777, Red Hook, 2017. Curran Associates Inc.

[208] S. M. Lundberg, G. G. Erion, and S.-I. Lee. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*, 2018.

[209] J. Luo, S. Zhang, and C. Zhang. Drivers of investment intentions across diverse cryptocurrency categories. *Finance Research Letters*, 77:107024, 2025.

[210] Y. Luo, Y. Feng, J. Xu, and P. Tasca. Piercing the veil of tvl: Defi reappraised. *arXiv preprint arXiv:2404.11745*, 2024.

[211] A. Mai, K. Pfeffer, M. Gusenbauer, E. Weippl, and K. Krombholz. User Mental Models of Cryptocurrency Systems – A Grounded Theory Approach. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, pages 341–358. USENIX Association, Aug. 2020. ISBN 978-1-939133-16-8.

[212] I. Makarov and A. Schoar. Blockchain analysis of the bitcoin market. *SSRN Electronic Journal*, October 2021. doi: 10.2139/ssrn.3942181.

[213] A. Malekloo, E. Ozer, M. AlHamaydeh, and M. Girolami. Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. *Structural Health Monitoring*, 21(4):1906–1955, 2022. doi: 10.1177/14759217211036880.

[214] M. Mashayekhi and R. Gras. Rule extraction from random forest: the RF+HC methods. In D. Barbosa and E. Milios, editors, *Advances in Artificial Intelligence*, Lecture Notes in Computer Science, pages 223–237. Springer International Publishing, 2015. ISBN 978-3-319-18356-5. doi: 10.1007/978-3-319-18356-5_20.

[215] B. Memarian and T. Doleck. Fairness, accountability, transparency, and ethics (fate) in artificial intelligence (ai) and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, 5:100152, 2023.

[216] G. Menardi and N. Torelli. Training and assessing classification rules with imbalanced data. *Data mining and knowledge discovery*, 28:92–122, 2014.

[217] C. Mershon and O. Shvetsova. *Party system change in legislatures worldwide: Moving outside the electoral arena*. Cambridge University Press, New York, 2013.

[218] I. D. Mienye and Y. Sun. A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*, 10:99129–99149, 2022. doi: 10.1109/ACCESS.2022.3207287.

[219] P. Milesi. Moral foundations and voting intention in italy. *Europe's Journal of Psychology*, 13(4):667–687, Nov. 2017.

[220] E. Mones, L. Vicsek, and T. Vicsek. Hierarchy measure for complex networks. *PloS one*, 7(3):e33799, 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0033799.

[221] T. Moore. The promise and perils of digital currencies. *International Journal of Critical Infrastructure Protection*, 6(3):147–149, 2013.

[222] A. Murtovi, A. Bainczyk, G. Nolte, M. Schlüter, and B. Steffen. Forest GUMP: a tool for verification and explanation. *International Journal on Software Tools for Technology Transfer*, 25(3):287–299, 2023. doi: 10.1007/s10009-023-00702-5.

[223] K. Nabben and P. D. Filippi. Accountability protocols? on-chain dynamics in blockchain governance. *Internet Policy Rev.*, 13(4):1–22, 2024. ISSN 2197-6775. doi: 10.14763/2024.4.1807. URL https://doi.org/10.14763/2024.4.1807.

[224] H. Nakahara, A. Jinguji, S. Sato, and T. Sasao. A random forest using a multi-valued decision diagram on an FPGA. In *2017 IEEE 47th International Symposium on Multiple-Valued Logic (ISMVL)*, pages 266–271, 2017. doi: 10.1109/ISMVL.2017.40. ISSN: 2378-2226.

[225] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf, 2008.

[226] S. Nakamoto. Bitcoin: A Peer-to-peer Electronic Cash System. *White paper*, 2008.

[227] A. Nani. The doge worth 88 billion dollars: A case study of dogecoin. *Convergence*, 28(6):1719–1736, 2022.

[228] K. Napierala and J. Stefanowski. Types of minority class examples and their influence on learning classifiers from imbalanced data. 46(3):563–597.

[229] E. Napoletano. Decentralized finance is building a new financial system. https://www.nasdaq.com/articles/decentralized-finance-is-building-a-new-financial-system-2021-04-02, 2021. (last access 2022).

[230] S. Needham and D. L. Dowe. Message length as an effective ockham's razor in decision tree induction. In *International Workshop on Artificial Intelligence and Statistics*, pages 216–223. PMLR, 2001. ISSN: 2640-3498.

[231] M. P. Neto and F. V. Paulovich. Explainable matrix - visualization for global and local interpretability of random forest classification ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1427–1437, 2020. ISSN 1077-2626. doi: 10.1109/TVCG.2020.3030354.

[232] H. Nobanee and N. O. D. Ellili. What do we know about Meme Stocks? A Bibliometric and Systematic Review, Current Streams, Developments, and Directions for Future Research. *International Review of Economics & Finance*, 85:589–602, 2023.

[233] C. Nugent. Memecoin Scandal rocks Argentina's Javier Milei. *Financial Times*, 2025. URL https://www.ft.com/content/27bcc19e-d422-4fac-ac08-5b76c1095e52. Accessed: 2025-06-13, available at: https://www.ft.com/content/27bcc19e-d422-4fac-ac08-5b76c1095e52.

[234] D. G. Ocampo, N. Branzoli, and L. Cusmano. *Crypto, Tokens and DeFi: Navigating the Regulatory Landscape*. Bank for International Settlements, Financial Stability Institute, 2023.

[235] A. Oksanen, E. Mantere, I. Vuorinen, and I. Savolainen. Gambling and Online Trading: Emerging Risks of Real-time Stock and Cryptocurrency Trading Platforms. *Public Health*, 205:72–78, 2022.

[236] J. Oliver. Decision graphs - an extension of decision trees. Citeseer, 1992.

[237] I. J. Onu, A. E. Omolara, M. Alawida, O. I. Abiodun, and A. Alabdultif. Detection of ponzi scheme on Ethereum using machine learning algorithms. *Scientific Reports*, 13:18403, 2023. doi: 10.1038/s41598-023-45275-0.

[238] D. M. Oppenheimer, T. Meyvis, and N. Davidenko. Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power. *Journal of Experimental Social Psychology*, 45(4):867–872, 2009.

[239] M. Ordekian, I. Becker, T. Moore, and M. Vasek. Raising the Bar: Assessing Historical Cryptocurrency Exchange Practices in Light of the EU's MiCA and DORA Regulation. Technical Report 5327395, SSRN, June 2025.

[240] S. Patil, K. R. Patil, C. R. Patil, and S. S. Patil. Performance overview of an artificial intelligence in biomedics: a systematic approach. *International Journal of Information Technology*, 12(3):963–973, 2020.

[241] J. Peng and G. Xiao. Detection of smart ponzi schemes using opcode. In *Blockchain and Trustworthy Systems: Second International Conference, BlockSys 2020*, pages 192–204. Springer, 2020. doi: 10.1007/978-981-15-9213-3_15.

[242] L. Pennella and A. G. Fabbrucci Barbagli. Explainable machine learning for predicting voting intentions: a study of italian politics. *International Journal of Data Science and Analytics*, 21(1):54, 2026.

[243] L. Pennella, F. Pinelli, and L. Galletta. X-spide: An explainable machine learning pipeline for detecting smart ponzi contracts in ethereum. *IEEE Access*, 13:85037–85055, 2025.

[244] L. Pennella, F. Pinelli, and L. Galletta. X-SPIDE: An explainable machine learning pipeline for detecting smart ponzi contracts in Ethereum. *IEEE Access*, 13:85037–85055, 2025. doi: 10.1109/ACCESS.2025.3569565.

[245] L. Pennella, P. Saggese, F. Pinelli, and L. Galletta. A unified framework and comparative study of decentralized finance derivatives protocols, 2025. URL https://arxiv.org/abs/2512.19113.

[246] K. S. Philander. Meme Asset Wagering: Perceptions of Risk, Overconfidence, and Gambling Problems. *Addictive Behaviors*, 137:107532, 2023.

[247] J. Proelss, D. Schweizer, and S. Sévigny. PolitiFi: Just another Meme, or Instrumental for Winning Elections? *Finance Research Letters*, 72:106533, 2025.

[248] D. Protocol. Pricing continuously funded everlasting options. URL https://github.com/deri-protocol/whitepaper/blob/master/Pricing_Continuously_Funded_Everlasting_Options.pdf.

[249] L. Puleo, G. Carteny, and G. Piccolino. Giorgia on their minds: Vote switching to fratelli d'italia in the italian general election of 2022. *Party Politics*, 31(4):609–622, 2025.

[250] J. R. Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

[251] Rachael. Rachael. URL https://rachael.swg.it/.

[252] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. Defining and identifying communities in networks. In *Proceedings of the National Academy of Sciences*, volume 101, pages 2658–2663, 2004. doi: 10.1073/pnas.0400054101.

[253] U. N. Raghavan, R. Albert, and S. Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 76(3):036106, 2007. doi: 10.1103/PhysRevE.76.036106.

[254] R. Rahimian and J. Clark. A shortfall in investor expectations of leveraged tokens. *Full version of paper to appear at Advances in Financial Technologies (AFT)*, 2024.

[255] A. Rai. Explainable ai: From black box to glass box. *Journal of the academy of marketing science*, 48:137–141, 2020.

[256] S. Release. Otc derivatives statistics at end-december 2023. *Bank for International Settlements*, 2023.

[257] M. Ribeiro, S. Singh, and C. Guestrin. Anchors: High-precision model-agnostic explanations. In *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, pages 1527–1535, 2018. ISBN 978-1-57735-800-8.

[258] M. T. Ribeiro, S. Singh, and C. Guestrin. "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1135–1144. Association for Computing Machinery, 2016. ISBN 978-1-4503-4232-2. doi: 10.1145/2939672.2939778.

[259] S. Roupakias and M. Chletsos. Immigration and far-right voting: evidence from greece. *The Annals of Regional Science*, 65(3):591–617, 2020.

[260] Q. Ruan and A. Streltsov. Perpetual futures contracts and cryptocurrency market microstructure. *Available at SSRN 4218907*, 2022.

[261] C. Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1: 206–215, 2019. doi: 10.1038/s42256-019-0048-x.

[262] P. Saggese, E. Segalla, M. Sigmund, B. Raunig, F. Zangerl, and B. Haslhofer. Assessing the solvency of virtual asset service providers: are current standards sufficient? *Applied Economics*, pages 1–16, 2024.

[263] P. Saggese, M. FrĂświs, S. Kitzler, B. Haslhofer, and R. Auer. Towards verifiability of total value locked (tvl) in decentralized finance. *arXiv preprint arXiv:2505.14565*, 2025.

[264] A. Saggu, L. Ante, and K. Kopiec. Uncertain Regulations, Definite Impacts: The Impact of the U.S. Securities and Exchange Commission's Regulatory Interventions on Crypto Assets. *Finance Research Letters*, 72:106413, 2025. ISSN 1544–6123. doi: https://doi.org/10.1016/j.frl.2024.106413.

[265] A. R. Sai, J. Buckley, and A. L. Gear. Characterizing wealth inequality in cryptocurrencies. *Frontiers Blockchain*, 4:730122, 2021. doi: 10.3389/fbloc.2021.730122. URL https://doi.org/10.3389/fbloc.2021.730122.

[266] T. Saito and M. Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432, 2015.

[267] M. Santos, P. Abreu, N. Japkowicz, A. Fernández, and J. Santos. A unifying view of class overlap and imbalance: Key concepts, multi-view panorama, and open avenues for research. 89:228–253.

[268] J. Scharfman. Meme Coins, Honeypots, and Artificial Intelligence-Enabled Crypto Fraud. In *The Cryptocurrency and Digital Asset Fraud Casebook, Volume II: DeFi, NFTs, DAOs, Meme Coins, and Other Digital Asset Hacks*, pages 221–249. Springer, 2024.

[269] S. Scharnowski and H. Jahanshahloo. The economics of liquid staking derivatives: Basis determinants and price discovery. *Journal of Futures Markets*, 45(2):91–117, 2025. doi: 10.1002/fut.22556.

[270] L. C. Schaupp, M. Festa, K. G. Knotts, and E. A. Vitullo. Regulation as a pathway to individual adoption of cryptocurrency. *Digital Policy, Regulation and Governance*, 24(2):199–219, 2022.

[271] S. H. Schwartz. An overview of the schwartz theory of basic values. *Online readings in Psychology and Culture*, 2(1):11, 2012.

[272] F. Schär. Decentralized finance: on blockchain-and smart contract-based financial markets. 103(2):153–174.

[273] W. R. Scott. *Institutions and organizations: Ideas and interests*. Sage Publications, 2008.

[274] SEC. Staff Statement on Meme Coins. *Statement, Division of Corporation Finance, Securities and Exchange Commission*, 2025.

[275] M. Shea. Is every memecoin just a scam? experts on whether andrew tate and trump are fleecing their followers. *The Guardian*, 2025. ISSN 0261-3077. URL https://www.theguardian.com/technology/2025/may/30/is-memecoin-scam-crypto-trump.

[276] R. J. Shiller. Measuring asset values for cash settlement in derivative markets: hedonic repeated measures indices and perpetual futures. *The Journal of Finance*, 48(3):911–931, 1993.

[277] S. E. Shreve et al. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer, 2004.

[278] O. Silva, A. Silva, I. Moreira, J. Nacif, and R. Ferreira. RDSF: Everything at same place all at once - a random decision single forest. In *Anais do XIII Simpósio Brasileiro de Engenharia de Sistemas Computacionais*, 2023.

[279] S. F. Singh, P. Michalopoulos, and A. Veneris. Option contracts in the defi ecosystem: Motivation, solutions, & technical challenges. In *2024 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, pages 1–7. IEEE, 2024.

[280] K. Soska, J.-D. Dong, A. Khodaverdian, A. Zetlin-Jones, B. Routledge, and N. Christin. Towards understanding cryptocurrency derivatives: A case study of bitmex. In *Proceedings of the Web Conference 2021*, pages 45–57, 2021.

[281] H. Stix. Ownership and Purchase Intention of Crypto-assets – Survey Results. *Oesterreichische Nationalbank Working Papers*, (226):1–42, 2019.

[282] J.-O. Strych. The impact of margin trading and short selling by retail investors on market price efficiency: Empirical evidence from bitcoin exchanges. *Finance Research Letters*, 47:102689, 2022.

[283] T. Susnjak and P. Maddigan. Forecasting patient flows with pandemic induced concept drift using explainable machine learning. *EPJ Data Science*, 12(1):11, 2023.

[284] SWG. SWG.it/observatory, . URL https://www.swg.it/observatory.

[285] SWG. SWG.it, . URL https://www.swg.it/home-en.

[286] P. J. Tan and D. L. Dowe. MML inference of decision graphs with multi-way joins and dynamic attributes. In T. T. D. Gedeon and L. C. C. Fung, editors, *AI 2003: Advances in Artificial Intelligence*, Lecture Notes in Computer Science, pages 269–281. Springer, 2003. ISBN 978-3-540-24581-0. doi: 10.1007/978-3-540-24581-0_23.

[287] C. Tandon, S. Revankar, and S. S. Parihar. How can we predict the Impact of the Social Media Messages on the Value of Cryptocurrency? Insights from Big Data Analytics. *International Journal of Information Management Data Insights*, 1(2): 100035, 2021.

[288] P. P. Thulasiram. Explainable artificial intelligence (xai): Enhancing transparency and trust in machine learning models, 2025.

[289] R. J. Tibshirani and B. Efron. An introduction to the bootstrap. *Monographs on statistics and applied probability*, 57(1):1–436, 1993.

[290] A. Trozze, J. Kamps, E. A. Akartuna, F. J. Hetzel, B. Kleinberg, T. Davies, and S. D. Johnson. Cryptocurrencies and future financial crime. *Crime Science*, 11(1):1–35, 2022.

[291] K. V. Tu and M. W. Meredith. Rethinking Virtual Currency Regulation in the Bitcoin Age. *Wash. L. Rev.*, 90:271, 2015.

[292] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the international AAAI conference on web and social media*, volume 4, pages 178–185, 2010.

[293] T. R. Tyler. Procedural Justice, Legitimacy, and the Effective Rule of Law. *Crime and justice*, 30:283–357, 2003.

[294] A. Tzinas and D. Zindros. The principal-agent problem in liquid staking. In A. Essex, S. Matsuo, O. Kulyk, L. Gudgeon, A. Klages-Mundt, D. Perez, S. Werner, A. Bracciali, and G. Goodell, editors, *Financial Cryptography and Data Security. FC 2023 International Workshops - Voting, CoDecFin, DeFi, WTSC, Bol, Brač, Croatia, May 5, 2023, Revised Selected Papers*, volume 13953 of *Lecture Notes in Computer Science*, pages 456–469, Cham, 2023. Springer. ISBN 978-3-031-48806-1. doi: 10.1007/978-3-031-48806-1_29. URL https://doi.org/10.1007/978-3-031-48806-1_29.

[295] Uniswap. Uniswap interface, 2025. URL https://app.uniswap.org/.

[296] N. Vallarano, C. J. Tessone, and T. Squartini. Bitcoin transaction networks: An overview of recent results. *Frontiers in Physics*, 8:286, 2020. doi: 10.3389/fphy.2020.00286. URL https://doi.org/10.3389/fphy.2020.00286. Published: December 3, 2020. Section: Physics of Networks.

[297] A. Van Assche and H. Blockeel. Seeing the forest through the trees: Learning a comprehensible model from an ensemble. In J. N. Kok, J. Koronacki, R. L. d. Mantaras, S. Matwin, D. Mladenič, and A. Skowron, editors, *Machine Learning: ECML 2007*, pages 418–429. Springer, 2007. ISBN 978-3-540-74958-5. doi: 10.1007/978-3-540-74958-5_39.

[298] T. Van Der Linden and T. Shirazi. Markets in Crypto-Assets Regulation: Does it Provide Legal Certainty and Increase Adoption of Crypto-Assets? *Financial innovation*, 9(1):22, 2023.

[299] M. Vasek and T. Moore. There's no free lunch, even using bitcoin: Tracking the popularity and profits of virtual currency scams. In *Financial Cryptography and Data Security*, pages 44–61. Springer, 2015.

[300] D. Vittori. Party change in anti-establishment parties in government: the case of five stars movement and syriza. *Italian Political Science*, 13(2):78–91, 2018.

[301] P. Vuttipittayamongkol and E. Elyan. Neighbourhood-based undersampling approach for handling imbalanced and overlapped data. 509:47–70.

[302] P. Vuttipittayamongkol, E. Elyan, and A. Petrovski. On the class overlap problem in imbalanced data classification. *Knowledge-based systems*, 212:106631, 2021.

[303] T. Waldvogel, P. König, U. Wagschal, B. Becker, and S. Weishaupt. It's the emotion, stupid! emotional responses to televised debates and their impact on voting intention. *Open Political Science*, 5(1):13–28, 2022.

[304] C. S. Wallace. *Statistical and Inductive Inference by Minimum Message Length*. Information Science and Statistics. Springer-Verlag, 2005. ISBN 978-0-387-23795-4. doi: 10.1007/0-387-27656-4.

[305] G. Wang, Y. Guo, W. Zhang, S. Xie, and Q. Chen. What type of algorithm is perceived as fairer and more acceptable? a comparative analysis of rule-driven versus data-driven algorithmic decision-making in public affairs. *Government Information Quarterly*, 40(2):101803, 2023.

[306] L. Wang, H. Cheng, Z. Zheng, A. Yang, and X. Zhu. Ponzi scheme detection via oversampling-based long short-term memory for smart contracts. *Knowledge-Based Systems*, 228:107312, 2021. ISSN 0950-7051. doi: j.knosys.2021.107312.

[307] M. Wang and J. Huang. Detecting ethereum ponzi schemes through opcode context analysis and oversampling-based adaboost algorithm. *Computer Systems Science & Engineering*, 47(1), 2023.

[308] P. Wang and Z. Bai. Decoding the crypto investor profile: How financial literacy, investment experience and age shape cryptocurrency investment decisions. *Humanities and Social Sciences Communications*, 12:1785, 2025. doi: 10.1057/s41599-025-06068-0.

[309] Y. Wang, N. Ding, and L. Zhang. The circulation of money and holding time distribution. *Physica A: Statistical Mechanics and its Applications*, 324(3):665–677, 2003. ISSN 0378-4371. doi: https://doi.org/10.1016/S0378-4371(03)00074-8.

[310] S. Werner, D. Perez, L. Gudgeon, A. Klages-Mundt, D. Harz, and W. Knottenbelt. Sok: Decentralized finance (defi). In *Proceedings of the 4th ACM Conference on Advances in Financial Technologies*, pages 30–46, 2022.

[311] S. M. Werner, D. Perez, L. Gudgeon, A. Klages-Mundt, D. Harz, and W. J. Knottenbelt. SoK: Decentralized finance (DeFi). In *Proceedings of the 4th ACM Conference on Advances in Financial Technologies (AFT '22)*, pages 30–46, 2022. doi: 10.1145/3558535.3559780.

[312] G. Wood et al. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper*, 151(2014):1–32, 2014.

[313] J. Xu and Y. Feng. Reap the harvest on blockchain: A survey of yield farming protocols. *IEEE Transactions on Network and Service Management*, 20(1):858–869, 2022.

[314] J. Xu, K. Paruch, S. Cousaert, and Y. Feng. SoK: Decentralized exchanges (DEX) with automated market maker (AMM) protocols. 55(11):238:1–238:50.

[315] T. Yan, S. Li, B. Kraner, L. Zhang, and C. J. Tessone. A data engineering framework for ethereum beacon chain rewards: From data collection to decentralization metrics. *Scientific Data*, 12(1):519, 2025. ISSN 2052-4463. doi: 10.1038/s41597-025-04623-7. URL https://doi.org/10.1038/s41597-025-04623-7.

[316] I. Yousaf, L. Pham, and J. W. Goodell. The Connectedness between Meme Tokens, Meme Stocks, and other Asset Classes: Evidence from a Quantile Connectedness Approach. *Journal of International Financial Markets, Institutions and Money*, 82: 101694, 2023.

[317] S. Yu, J. Jin, Y. Xie, J. Shen, and Q. Xuan. Ponzi scheme detection in ethereum transaction network. In *Blockchain and Trustworthy Systems: Third International Conference, BlockSys 2021*, pages 175–186. Springer, 2021. doi: 10.1007/978-981-16-7993-3_14.

[318] D. Zelinsky. 'to the moon!': Elon musk, dogecoin, and the political economy of charismatic leadership. *Journal of Cultural Economy*, 17(3):297–313, 2024.

[319] D. A. Zetzsche, D. W. Arner, and R. P. Buckley. Decentralized finance. *Journal of Financial Regulation*, 6(2):172–203, 2020. doi: 10.1093/jfr/fjaa010.

[320] Y. Zhang, W. Yu, Z. Li, S. Raza, and H. Cao. Detecting ethereum ponzi schemes based on improved lightgbm algorithm. *IEEE Transactions on Computational Social Systems*, 9(2):624–637, 2021. doi: 10.1109/TCSS.2021.3088145.

[321] Z. Zhang, C. Xu, and C. Jiang. Practical blockchain-based options contract. *IEEE Transactions on Services Computing*, 2024.

[322] X. Zhao, Y. Wu, D. L. Lee, and W. Cui. iForest: Interpreting random forests via visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 25(1): 407–416, 2019. ISSN 1941-0506. doi: 10.1109/TVCG.2018.2864475.

[323] P. Zheng, Z. Zheng, J. Wu, and H. Dai. Xblock-eth: Extracting and exploring blockchain data from ethereum. *IEEE Open J. Comput. Soc.*, 1:95–106, 2020. doi: 10.1109/OJCS.2020.2990458. URL https://doi.org/10.1109/OJCS.2020.2990458.

[324] Z. Zheng, W. Chen, Z. Zhong, Z. Chen, and Y. Lu. Securing the ethereum from smart ponzi schemes: Identification using static features. *ACM Trans. Softw. Eng. Methodol.*, nov 2022. ISSN 1049-331X. doi: 10.1145/3571847.

[325] Y. Zhou and G. Hooker. Interpreting models via single tree approximation, 2016.

[326] B. Zhu and M. Shoaran. Tree in tree: from decision trees to decision graphs. *Advances in Neural Information Processing Systems*, 34:13707–13718, 2021.

[327] M. Zulianello. Varieties of populist parties and party systems in europe: From state-of-the-art to the application of a novel classification scheme to 66 parties in 33 countries. *Government and Opposition*, 55(2):327–347, 2020.

[328] M. Zulianello et al. New anti-system parties and political competition: the case of italy. *New opportunities and impasses: theorizing and experiencing politics*, page 344, 2014.