



**UNIVERSITÀ
DEGLI STUDI
DI TRIESTE**

UNIVERSITÀ DEGLI STUDI DI TRIESTE

**XXXIII CICLO DEL DOTTORATO DI RICERCA IN
INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE**

**COMPUTATIONAL MODEL TO ALERT EARLY THE
LEVEL OF ACHIEVEMENT OF THE GRADUATION
PROFILE OF UNIVERSITY ENGINEERING STUDENTS**

Settore scientifico-disciplinare:

ING-INF/06 BIOINGEGNERIA ELETTRONICA E INFORMATICA

**DOTTORANDO
HÉCTOR MIGUEL ZELADA VALDIVIESO**

**COORDINATORE
PROF. ALBERTO TESSAROLO**

**SUPERVISORE DI TESI
PROF. AGOSTINO ACCARDO**

ANNO ACCADEMICO 2020/2021

RESUMEN

Hoy en día el logro de las competencias del perfil de egreso es una evidencia muy importante que muestra la calidad de los programas universitarios. Sin embargo, las universidades deben esperar a que sus estudiantes terminen o lleguen al final de sus estudios para poder medir exactamente el logro de éstas, lo cual trae problemas que cuando algunos no alcancen los niveles esperados ya no haya el tiempo suficiente para poder tomar acciones correctivas. En ese sentido, en la literatura, existen varias investigaciones relacionadas con el perfil de egreso y otras que se centraron en la predicción del rendimiento académico en cursos presenciales o virtuales, en la predicción de la deserción estudiantil, en predecir la motivación académica del estudiante, en predecir la colocación de un estudiante en un empleo, en predecir el promedio final global del estudiante universitario o predecir el retraso en la graduación, más no hay estudios que hayan abordado directamente el tema de la predicción del nivel de logro del perfil del egresado universitario de ingeniería. Debido a esto, la investigación tuvo como objetivo general desarrollar un modelo computacional basado en Machine Learning que permita alertar tempranamente el nivel de logro del perfil de egreso de los estudiantes universitarios de Ingeniería, de tal manera que se tenga información oportuna para tomar decisiones correctivas anticipadamente y no esperar hasta el final de los estudios para poder obtener este resultado. Para esto, se siguió la metodología CRISP-DM, se recolectaron datos de 1982 egresados de diferentes programas de ingeniería de una universidad peruana y haciendo uso de Matlab y algoritmos de Machine Learning, al final se creó un modelo bastante preciso (accuracy: 96.7%) utilizando el algoritmo SVM y se obtuvo que las características relacionadas con las calificaciones obtenidas hasta el IV ciclo de estudios en los cursos de Matemáticas y Física, son los mejores predictores para pronosticar el nivel de logro del perfil de egreso de los estudiantes de ingeniería.

Palabras clave: Modelo computacional, Machine Learning, Predicción, Perfil de Egreso, Programas de Ingeniería

ABSTRACT

Today, the achievement of the graduate profile competencies is very important evidence that shows the quality of university programs. However, universities must wait for their students to finish or reach the end of their studies in order to accurately measure their achievement, which brings problems that when some do not reach the expected levels there is not enough time to take action corrective. In this sense, in the literature, there are several investigations related to the graduate profile and others that focused on the prediction of academic performance in face-to-face or virtual courses, in the prediction of student desertion, in predicting the student's academic motivation, in predicting the placement of a student in a job, in predicting the final global average of the university student or predicting the delay in graduation, but there are no studies that have directly addressed the subject of the prediction of the level of achievement of the profile of the university graduate engineering. Due to this, the general objective of the research was to develop a computational model based on Machine Learning that allows early warning of the level of achievement of the graduation profile of engineering university students, in such a way that timely information is available to make corrective decisions in advance. and not wait until the end of the studies to obtain this result. For this, the CRISP-DM methodology was followed, data was collected from 1982 graduates of different engineering programs of a Peruvian university and using Matlab and Machine Learning algorithms, in the end a fairly accurate model was created (accuracy: 96.7%) using the SVM algorithm and it was found that the characteristics related to the grades obtained up to the IV cycle of studies in Mathematics and Physics courses are the best predictors to predict the level of achievement of the graduation profile of engineering students.

Keywords: Computational model, Machine Learning, Prediction, Graduate Profile, Engineering Programs

INDICE

RESUMEN	2
ABSTRACT	3
I. INTRODUCCIÓN.....	8
II. REVISIÓN DE LA LITERATURA.....	16
III. MATERIALES Y MÉTODOS	42
IV. RESULTADOS	45
4.1. ANÁLISIS DEL PERFIL DE EGRESO DE LOS ESTUDIANTES DE INGENIERÍA.....	45
4.2. ANÁLISIS, PREPARACIÓN DE DATOS Y SELECCIÓN DE CARACTERÍSTICAS.....	51
4.3. DESARROLLO Y PRUEBAS DEL MODELO PREDICTIVO.....	70
4.4. CREACIÓN DE LA APLICACIÓN PARA PREDECIR NIVEL DE LOGRO DEL PERFIL DE EGRESO	78
V. DISCUSIÓN.....	80
VI. CONCLUSIONES	82
VII. REFERENCIAS.....	84

LISTA DE TABLAS

TABLA I. PERFIL DE EGRESO (STUDENTS OUTCOMES) PARA PROGRAMAS DE INGENIERÍA PROPUESTOS POR ICACIT Y ABET.....	46
TABLA II. NIVELES DE LOGRO DEL PERFIL DE EGRESO	47
TABLA III. ASIGNATURAS COMUNES DE LOS PROGRAMAS DE INGENIERÍA ..	48

TABLA IV. CLASIFICACIÓN DE ASIGNATURAS POR ÁREA.....	49
TABLA V. RELACIÓN DE ATRIBUTOS O CARACTERÍSTICAS DEL ESTUDIANTE	50
TABLA VI. CANTIDAD DE EGRESADOS POR ESCUELA Y AÑO DE EGRESO.....	53
TABLA VII. RESUMEN DESCRIPTIVO DE LAS VARIABLES NUMÉRICAS	60
TABLA VIII. VALORES OUTLIERS POR CADA ATRIBUTO NUMÉRICO.....	66
TABLA IX. CODIFICACIÓN DE VARIABLES CATEGÓRICAS	67
TABLA X. ORDEN DE RELEVANCIA DEL PREDICTOR PARA EL MODELO EN BASE A MRMR VS. COEFICIENTE DE CORRELACIÓN.....	68
TABLA XI. CARACTERÍSTICAS SELECCIONADAS	69
TABLA XII. EVALUACIÓN DE ALGORITMOS DE CLASIFICACIÓN DE MACHINE LEARNING.....	71
TABLA XIII. CANTIDAD DE DATOS DE PRUEBA POR PROGRAMA Y AÑO DE EGRESO.....	74
TABLA XIV PRECISIÓN (ACCURACY) DEL MODELO DURANTE LAS PRUEBAS	75

LISTA DE FIGURAS

Fig. 1 Metodología Cross Industry Standard Process for Data Mining (CRISP-DM) [69] .	44
Fig. 2 Algoritmos de clasificación de Machine Learning [72].....	44
Fig. 3 Esquema de alineación del perfil de egreso con students outcomes para ingeniería .	45
Fig. 4 Esquema de malla curricular de asignaturas comunes de los programas de ingeniería	49
Fig. 5 Cantidad de egresados por escuela profesional hasta el año 2021	52

Fig. 6 Cantidad de egresados por escuela y año de egreso.....	53
Fig. 7 Cantidad de egresados por escuela profesional y modalidad con la que ingresó.....	54
Fig. 8 Porcentaje de egresados según sexo.....	55
Fig. 9 Cantidad de egresados masculinos (M) y femeninos (F) por escuela	55
Fig. 10 Porcentaje de egresados según el tipo de colegio (público/privado) donde estudió	56
Fig. 11 Cantidad de egresados por escuela y tipo de colegio	56
Fig. 12 Cantidad de egresados por escuela y ubicación de colegio.....	57
Fig. 13 Porcentaje de egresados según el nivel de logro del perfil de egreso	57
Fig. 14 Porcentaje de egresados por escuela y su nivel de logro del perfil de egreso	58
Fig. 15 Mapa de calor del nivel de logro del perfil de egreso por escuela profesional	59
Fig. 16 Diagrama de distribución de frecuencias de las notas obtenidas en todas las asignaturas comunes	61
Fig. 17 Diagrama de distribución de las calificaciones obtenidas en las asignaturas comunes con respecto a la media.....	62
Fig. 18 Relación entre calificaciones de asignaturas del área de matemática	63
Fig. 19 Relación entre calificaciones de asignaturas del área de Física	64
Fig. 20 Relación entre calificaciones de asignaturas del área de humanidades	65
Fig. 21 Matriz de confusión para data de validación al entrenar el modelo	72
Fig. 22 Esquema del algoritmo SVM [75]	73
Fig. 23 Función de base radial (RBF) o gaussiana	73
Fig. 24 Código de entrenamiento del Modelo Computacional.....	74
Fig. 25 Matriz de confusión durante las pruebas del modelo con el programa de Ingeniería Civil	76

Fig. 26 Matriz de confusión durante las pruebas del modelo con el programa de Ingeniería Industrial.....	76
Fig. 27 Matriz de confusión durante las pruebas del modelo con el programa de Ingeniería Mecánica.....	77
Fig. 28 Matriz de confusión durante las pruebas del modelo con el programa de Ingeniería de Sistemas	77
Fig. 29 Aplicación para predecir el nivel de logro del perfil de egreso de un estudiante.....	78
Fig. 30 Aplicación para predecir el nivel de logro del perfil de egreso de varios estudiantes	79

I. INTRODUCCIÓN

El perfil de egreso de un programa universitario son un conjunto de competencias generales y específicas que deben lograr los estudiantes como resultado final del proceso de formación profesional [1]. Asimismo, en la literatura académica, podemos encontrar otras denominaciones relacionadas con estas competencias del egresado. En el caso de las acreditadoras ABET e ICACIT, utilizan el término Resultados del Estudiante (Students Outcomes) [2] [3], en cambio, la OECD (Organización para la Cooperación y el Desarrollo Económico), utiliza la denominación de resultados de aprendizaje en la educación superior [4].

El logro de las competencias establecidas en el perfil de egreso de un programa universitario tiene un rol muy importante en la educación superior y ha sido el tema de muchas discusiones en diferentes lugares [5]. Hoy en día, es una evidencia muy importante que muestra la calidad educativa de un programa universitario, debido a que muchos grupos de interés esperan que las universidades preparen adecuadamente a los estudiantes mediante el desarrollo de competencias relevantes. En ese sentido, actualmente los programas universitarios, no sólo deben centrarse en el proceso de enseñanza-aprendizaje, además deben preocuparse por los resultados. Es decir que deben verificar el resultado final del proceso a través de la evaluación del logro del perfil de egreso del programa [1] [2] [3].

La evaluación del logro del perfil del egresado universitario se centra en medir directa o indirectamente las competencias genéricas y específicas de los egresados. En algunos países, estos instrumentos de evaluación son diseñados por entidades externas a las universidades

[6][7]. En otros países, se evalúan mayormente los conocimientos aislados en las asignaturas y se asume que la aprobación satisfactoria del plan curricular garantiza el logro del perfil del egresado [8]. En cambio, en otras universidades en la fase terminal de la carrera incluyen actividades integradoras, tales como proyectos, trabajos de investigación aplicada, etc. que están orientados a evaluar estas competencias [9]. En ese mismo sentido, la OECD viene trabajando en desarrollar una plataforma global para la Evaluación de los Resultados del Aprendizaje en la Educación Superior (AHELO), que permitirá a los gobiernos, instituciones de educación superior y a los estudiantes evaluar y comparar los resultados de aprendizaje logrados con los estándares internacionales y de esta manera poder establecer planes de mejora [4].

Sin embargo, en varias universidades a pesar de que los perfiles de egreso de los planes curriculares se elaboran tomando como referencia las necesidades del mercado laboral y la opinión de los grupos de interés, vienen teniendo dificultades al final con el nivel de logro de las competencias del perfil de egreso. PayScale y Future Workplace, en Estados Unidos, en base a un estudio realizado, mostraron que existe una brecha entre las habilidades con las que los estudiantes salen de la universidad y las necesarias para el éxito en el mundo profesional. El 50% de los empleadores manifestaron que las habilidades esenciales en los recién graduados a menudo no las tienen [10].

De igual manera, un estudio realizado en Europa, refleja la distancia que existe entre las competencias requeridas por el mercado laboral y las que se adquieren en la universidad [11]. Asimismo, un estudio en Cataluña muestra que el 42% de las organizaciones empleadoras declaran haber tenido dificultades para contratar y el 60% de los encuestados manifiesta que

la principal dificultad que han afrontado en el proceso de contratación es la falta de solicitantes con las competencias necesarias [12].

Asimismo, en América Latina, el Banco Mundial, en base a un estudio global realizado por McKinsey, da a conocer que sólo el 42% de los empleadores manifiestan que los recién egresados están aptos para el mercado laboral [13]. Del mismo modo, en un estudio realizado en México por el CIDAC (Centro de Investigación para el Desarrollo, A.C.), menciona que los recién egresados de las universidades no cuentan con las competencias requeridas para insertarse en el mercado laboral. El 26% de las empresas encuestadas afirman que es difícil encontrar a trabajadores jóvenes que cumplan con el perfil del puesto requerido[14].

De igual manera, en el Perú, en donde existen 142 universidades a nivel nacional (51 públicas y 91 privadas), CEPLAN (Centro Nacional de Planeamiento) publica en el Diario el Comercio que el 60% de las empresas manifiesta que los recién egresados de universidades e institutos no están aptos para empezar a trabajar [15]. En esa misma línea, ManpowerGroup, indica que el 46% de las empresas tienen problemas para cubrir sus vacantes laborales, y entre las razones se indican que 20% es por falta de competencias técnicas, y el 8% es por falta de habilidades blandas [16]. Asimismo, en un artículo de MBA & Educación Ejecutiva, indica que el 60% de los profesionales jóvenes peruanos trabajan en algo para lo que no han estudiado, manifestando además que hay muchas empresas que buscan trabajadores, pero no los encuentran con las competencias y la formación que requieren [17]. De igual manera, en un artículo de Universia Perú, se indica que según un estudio publicado por el Banco Central de Reserva (BCR), el 40% de los profesionales

peruanos están subempleados y que esta cifra ha aumentado considerablemente en comparación con años anteriores, donde la tasa de subempleados llegaba al 29% [18].

Este problema también abarca a los egresados de los programas de ingeniería. La empresa Randstad, menciona que existen deficiencias de habilidades en ingeniería; cita al estudio de Global Engineering Report, quienes reportan que las empresas están alertando deficiencias considerables en los conocimientos y habilidades que deben tener los ingenieros. De igual manera, hace referencia al estudio del IET (Institución de Ingeniería y Tecnología), donde se menciona que el 62% de las empresas que contratan ingenieros dan a conocer que los recién egresados no tienen las competencias apropiadas, lo cual hace que se incremente la brecha entre la oferta de trabajos en ingeniería y las competencias de los graduados[19].

En Perú, en una universidad privada, en donde se han definido los siguientes niveles de logro globales de las competencias del perfil de egreso: Sobresaliente, Notable, Suficiente, Insuficiente y No logrado, luego de analizar los datos registrados de los egresados de la Facultad de Ingeniería hasta el año 2021, se encontró que el 32% no logró el perfil de egreso, obteniendo sólo el 11% un nivel sobresaliente y el 20% un nivel notable. Estando el mayor porcentaje (38%) con un nivel suficiente; es decir que sólo lograron los indicadores mínimos establecidos para cada competencia del perfil de egreso.

Debido a esto, en la literatura podemos encontrar varios estudios que han tratado de determinar los factores asociados con el rendimiento académico de un estudiante universitario, encontrando que existen factores socioeconómicos, familiares, escolares,

académicos, demográficos, institucionales y personales que pueden influir en el desempeño académico de un estudiante durante su formación profesional [20] [21].

En ese sentido, las universidades deben implementar mecanismos de seguimiento de sus estudiantes y poder evaluar el nivel de logro de las competencias del perfil de egreso oportunamente, de tal manera que puedan tomar acciones correctivas a tiempo. Esto debido a que el problema de que un estudiante universitario no logre las competencias del perfil de egreso trae como consecuencias la dificultad para evidenciar la calidad de los programas universitarios, que las empresas o instituciones no consigan el personal calificado que necesitan, disminución del índice de empleabilidad de los egresados y finalmente que las universidades no estén contribuyendo con uno de los objetivos del desarrollo sostenible que tiene que ver con la Educación de calidad.

En los últimos años han habido avances considerables con respecto a teorías, métodos, técnicas o herramientas computacionales que pueden aplicarse en la educación, tal como lo podemos ver en el uso de los kits de Lego Mindstorms en la enseñanza de robótica en las escuelas, el aprendizaje basado en aplicaciones de realidad virtual, sistemas de aprendizaje adaptativo, los sistemas de tutoría inteligente, el aprendizaje en línea, el procesamiento del lenguaje natural, dispositivos de lectura digital inteligentes, la tecnología de traducción automática, el Flipped Classroom , la gamificación, el Learning Analytics y la aplicación del Big Data, la minería de datos y el Machine Learning en la educación [22] [23]. Sin embargo, por diferentes motivos, aún varias universidades no hacen uso de estas tecnologías.

Del mismo modo, en la literatura, existen varias investigaciones que han tratado ampliamente temas relacionados con las competencias del egresado, resultados del aprendizaje [24][25][26][27][28] o los resultados del estudiante propuestos por la acreditadora ABET [29][30][31][32][33][34][35]. Sin embargo, estos estudios, se centraron en proponer acciones de mejora para el proceso de enseñanza-aprendizaje, mejorar la evaluación final de los resultados del aprendizaje o en determinar cuáles son los factores que influyen en el rendimiento académico de los estudiantes. Asimismo, existe abundante literatura con respecto al desarrollo de modelos de predicción en la educación [36][37][38], los cuales se centraron mayormente en la predicción del rendimiento académico en cursos, semestres o años de la educación presencial [39][40][41][42][43][44][45], predicción del rendimiento académico en cursos MOOC (Massive Open Online Course) [46][47][48][49][50][51][52][53][54], predicción de la deserción o abandono de los estudiantes [55][56], pronosticar el tiempo de retraso en la graduación [57][58] o predicción que si requiere tutoría o la motivación académica del estudiante [59][60] o la predicción de para conseguir empleo por parte de los estudiantes [61].

Sin embargo, el tema de la predicción del nivel de logro del perfil de egreso de un estudiante universitario de ingeniería no ha sido abordado directamente, encontrándose mayormente algunos estudios que buscan predecir el rendimiento académico final o el promedio final global del estudiante universitario [62][63][64][65][66][67][68]. Debido a esto se planteó el siguiente problema general de investigación: ¿De qué manera haciendo uso del Machine Learning se podría alertar tempranamente el nivel de logro del perfil de egreso de un estudiante universitario de ingeniería en una universidad peruana?

En base a todo esto, se planteó como objetivo general: Desarrollar un modelo computacional basado en Machine Learning que permita alertar tempranamente el nivel de logro del perfil de egreso que tendría un estudiante universitario de ingeniería en una universidad peruana. Para esto, se establecieron como objetivos específicos: Determinar las características o atributos del estudiante de ingeniería que permiten obtener la mayor precisión a la hora de predecir el nivel de logro de su perfil de egreso, evaluar los diferentes algoritmos de clasificación de Machine Learning que nos permitan crear el mejor modelo predictivo del nivel de logro del perfil de egreso de un estudiante universitario de ingeniería, crear un modelo utilizando las características y el algoritmo de Machine Learning seleccionado y someterlo a pruebas con datos nuevos de todos los egresados del año 2020 y 2021 de diferentes programas de ingeniería para poder evaluar su comportamiento y finalmente determinar con cuanto tiempo de anticipación es posible llegar a predecir el nivel de logro del perfil de egreso que obtendría un estudiante universitario de ingeniería al finalizar sus estudios.

En ese sentido, la investigación se justifica debido a que permite contar con un modelo basado en Machine Learning para predecir el nivel de logro del perfil de egreso que tendría un estudiante universitario de ingeniería y de esta manera se puedan proponer acciones de mejora oportunamente; es decir antes que los estudiantes terminen sus estudios. Lo cual apoyará a poder tener mejores resultados finales de las competencias del egresado que trae consigo que las universidades puedan acreditar la calidad de sus programas, que las instituciones de educación superior universitaria contribuyan con la educación de calidad que es uno de los objetivos del desarrollo sostenible, que los egresados estén mejor preparados y sean más empleables y que las empresas o instituciones puedan conseguir el profesional

calificado que necesitan, reduciendo de esta manera la brecha que existe actualmente entre las competencias del egresado y las necesidades del mercado. Asimismo, esta investigación permitió afianzar un mayor conocimiento con respecto a modelos computacionales basados en Machine Learning que permitan predecir el nivel de logro del perfil del egresado universitario de ingeniería, esto debido a que, en la literatura revisada no hay estudios que directamente hayan abordado este tema. Asimismo, contribuyó con la generación de conocimiento, sobre cuáles son los atributos o características del estudiante de ingeniería que influyen más significativamente en el logro de su perfil de egreso, cuáles son los algoritmos que muestran mejor precisión al momento de predecir tempranamente los resultados y que tan temprano es posible conseguir esto.

II. REVISIÓN DE LA LITERATURA

En la literatura, existen varias investigaciones que han tratado ampliamente temas relacionados con las competencias del egresado, resultados del aprendizaje [24][25][26][27][28] o los resultados del estudiante propuestos por la acreditadora ABET [29][30][31][32][33][34][35]. Sin embargo, estos estudios, se centraron en proponer alternativas para mejorar el proceso de enseñanza aprendizaje, mejorar la evaluación final de los resultados del aprendizaje o en determinar cuáles son los factores que influyen en el rendimiento académico de los estudiantes. Asimismo, existe abundante literatura con respecto al desarrollo de modelos de predicción en la educación [36][37][38], los cuales se centraron mayormente en la predicción del rendimiento académico en cursos, semestres o años de la educación presencial [39][40][41][42][43][44][45], predicción del rendimiento académico en cursos online [46][47][48][49][50][51][52][53][54], predicción de la deserción o abandono de los estudiantes [55][56], pronosticar el tiempo de retraso en la graduación [57][58] o predicción que si requiere tutoría o la motivación académica del estudiante [59][60] o la predicción de para conseguir empleo por parte de los estudiantes [61].

Sin embargo, el tema de la predicción del nivel de logro del perfil de egreso de un estudiante universitario de ingeniería no ha sido abordado directamente, encontrándose mayormente estudios que buscan pronosticar el rendimiento académico final o el promedio final global del estudiante universitario [62][63][64][65][66][67][68][69].

En la revisión de la literatura, se encontró que las investigaciones utilizaron algoritmos de Machine Learning para el diseño de sus modelos de predicción, siendo los más utilizados:

Neural Networks, Decision Tree, Support Vector Machines, Logistic Regression, Random Forest, Naive Bayes y K-Nearest Neighbors. Asimismo, se encontró que la mayoría de los autores concluyen que los algoritmos que mostraron mayor precisión fueron: Neural Networks y Random Forest, seguidos por Support Vector Machines y Decision Tree con precisiones que van desde 66% hasta 98%. Del mismo modo, en los diferentes modelos de predicción utilizaron varios tipos de características del estudiante como: Datos personales (sexo, edad, estado civil, lugar de residencia, ocupación del padre, ocupación de la madre, situación económica familiar, colegio de procedencia, religión, nacionalidad, etc.), modalidad de ingreso, asistencias a clases, calificaciones y características sobre su comportamiento en aplicaciones online.

Debido a esto, podemos concluir que existen algunos vacíos en el campo del conocimiento tales como:

- No se encontraron estudios que directamente hayan abordado el tema de alerta temprana o predicción del nivel de logro del perfil de egreso de ingeniería o Students Outcomes como los establece las agencias acreditadoras de ABET o ICACIT. Las investigaciones se centran mayormente en la predicción del rendimiento académico de cursos (presenciales / online), en la predicción de la deserción estudiantil o en la predicción de la calificación final del estudiante.
- Existen vacíos de conocimiento acerca de cuáles son los atributos o características de los estudiantes que podrían servir para poder predecir el nivel de logro del perfil de egreso de los estudiantes de un programa de ingeniería.
- Existen vacíos de conocimiento acerca de con cuánto tiempo de anticipación es posible predecir el nivel de logro del perfil del egresado de ingeniería; es decir

mínimamente datos de cuántos semestres de estudio se necesitan para poder obtener una mayor precisión en la predicción.

- Existen vacíos de conocimiento acerca de cuáles son los algoritmos de clasificación del Machine Learning que tienen una mejor precisión a la hora de predecir tempranamente el nivel de logro del perfil del egresado de ingeniería.

A continuación, se describen algunos trabajos de investigación relacionados con el tema de estudio:

H. Zelada-Valdivieso (2021), en su artículo titulado “Model based on Machine Learning to alert early on the academic performance that a Systems Engineering student would have upon graduation”, propone la creación de un modelo basado en Machine Learning para alertar tempranamente el nivel del logro del rendimiento académico de los estudiantes de Ingeniería de Sistemas en una Universidad Peruana. En su investigación establece que el rendimiento académico se mide por medio del promedio ponderado final que obtiene un estudiante al finalizar su plan de estudios. Asimismo, determina que las características que están más relacionadas con el nivel del rendimiento académico alcanzado son las notas que obtienen en las diferentes asignaturas de su plan de estudio y que sabiendo estas notas hasta el V ciclo es posible anticipar cuál sería el rendimiento académico final de un estudiante y de esta manera los programas puedan proponer acciones correctivas a tiempo; es decir antes que terminen sus estudios. Al final llega a determinar que el modelo entrenado Ensemble (Subspace discriminant) es el que obtiene la mayor precisión (accuracy): 81,4%. Este artículo se desarrolló durante los estudios del doctorado en ingeniería industrial y de la información de

la Universidad de Trieste – Italia y se encuentra publicado en la revista IEEE Xplore la cual se encuentra indizada con Scopus (<https://ieeexplore.ieee.org/document/9675117>) [69]

Hu et al. (2016), hicieron una revisión sistemática de la literatura académica existente entre el 2002 y 2016 sobre predicción de los resultados de aprendizaje de los estudiantes usando Learning Analytics (LA). Encontraron que el desempeño en un curso es el que aparece con mayor frecuencia como objetivo de la predicción, seguido de la retención o deserción escolar. Además, encontraron que las categorías de características más usadas son: características demográficas, registro y rendimiento del historial del estudiante, registro y rendimiento del estudiante en el curso actual, actividad y características del curso, características de comportamiento de aprendizaje, características auto informadas entre otras características poco claras. En total contabilizaron 14 algoritmos diferentes, siendo cinco los más utilizados en la mayoría de los artículos: Árboles de decisión, redes neuronales, clasificación basada en agrupación, algoritmos basados en reglas y Naive Bayes. Las precisiones informadas estuvieron dentro del rango de 60-98% para la mayoría de los casos [36] .

Del Río and Pineda (2016), hicieron una revisión de la literatura académica existente entre el 2011 y agosto del 2016 en el área de minería de datos aplicada al dominio educativo en un entorno tradicional en el aula en las Instituciones de Educación Superior (56 artículos). En su estudio, encontraron que las variables predictoras (características) más utilizadas son los datos académicos (89.3%), seguido de calificaciones y datos no académicos (7.2%). Además, encontraron que, entre los métodos de predicción, los más utilizados han sido: Clasificación (71.4%), agrupación (8.9%), reglas de asociación (7.1%), líneas de regresión (5.4%), otros (7.2%). Asimismo, con respecto a las herramientas de software, encontraron

que el 28.6% utilizó WEKA, pero el 60.7% de los artículos no especifica el software utilizado. Asimismo, se indica que los artículos revisados llegaron a predecir el rendimiento académico de los estudiantes, más no comentan nada acerca de la precisión que lograron[37].

Alper and Çataltepe (2012), realizaron una investigación para predecir los resultados de éxitos de los estudiantes (Aprobación o Falla) en diferentes cursos de la carrera de Ingeniería Informática de la UIT (Istanbul Technical University). Para esto utilizaron como características además de las calificaciones de los cursos convencionales, las medidas de rendimiento de los estudiantes que corresponden a los resultados ABET en los cursos. Estas características fueron seleccionadas utilizando el algoritmo de selección de características mRMR (minimum Redundancy Maximum Relevance). Los modelos predictivos que diseñaron fueron entrenados empleando métodos de Machine Learning: Naive Bayes, Multilayer Perceptron, SVM, Logistic Regression, Bayesian Logistic Regression con datos de tres cursos diferentes. Al final obtuvieron que el algoritmo de Regresión logística bayesiana obtuvo en promedio una mejor precisión: 78.47%. En esta investigación se utilizaron bibliotecas de Weka Java [39].

Verma, Singh, and Verma (2016) realizaron una investigación cuyo objetivo fue determinar cuáles son las características personales y preuniversitarias necesarias para poder predecir confiablemente el desempeño del estudiante en una universidad (Excelente, Muy Bueno, Bueno, Promedio y Malo). Para esto recolectaron datos de 10330 estudiantes, descritos por 20 características, que incluyen género, año de nacimiento, lugar de residencia y país, tipo de educación previa, perfil y lugar de la educación anterior, puntaje total de la educación anterior, año de admisión a la universidad, examen de admisión y puntaje obtenido,

especialidad, semestre actual, puntaje total de la universidad, etc. Para entrenar sus modelos, utilizaron los clasificadores WEKA, que incluye un algoritmo común de árbol de decisión C4.5 (J48), dos clasificadores bayesianos (Naive Bayes y Bayes Net), un algoritmo de vecino más cercano (IBk) y dos aprendices de reglas (OneR y JRip). Al final, obtuvieron que las tasas de predicción varían entre 52 y 67%, no precisando cual es el algoritmo más preciso, debido a que los clasificadores tienen un rendimiento diferente para las cinco clases. Además, concluyen que los atributos de datos relacionados con el puntaje de admisión en la universidad y el número de fracasos en los exámenes universitarios de primer año se encuentran entre los factores que más influyen en el proceso de clasificación. La herramienta de software que se utiliza para la implementación del proyecto es el software de fuente abierta WEKA, que ofrece una amplia gama de métodos de clasificación para la minería de datos [62].

Haskins (2017), realizó una investigación para poder apoyar a los profesores en determinar si un estudiante tendrá éxito o no en un módulo universitario. Para esto, las principales características de los estudiantes que utilizaron fueron: La calificación más alta y más baja del estudiante, porcentaje de evaluaciones completadas por el estudiante, porcentaje de tiempo que ha transcurrido desde el inicio del módulo, los años que lleva en la universidad y la frecuencia con la que el estudiante obtiene calificaciones por debajo del promedio del aula. Los modelos predictivos se entrenaron utilizando técnicas de aprendizaje supervisado de Machine Learning como: redes neuronales artificiales y árboles de decisión. Al final obtuvieron que individualmente, tanto la red neuronal como el árbol de decisión tienen una precisión superior al 77% y una sensibilidad superior al 83% en un conjunto de datos de prueba [40].

Kondo, Okubo, y Hatanaka (2017), realizaron una investigación donde proponen un enfoque para la detección de estudiantes académicamente en riesgo, para lo cual predicen el promedio de calificación (GPA) obtenido en base a los datos del sistema de gestión de aprendizaje (LMS). Para esto, seleccionaron 7 características del estudiante: Tasa de asistencia, número de ejecuciones del reproductor, número total de operaciones durante la noche, número de inicios de sesión, número de lecciones iniciadas, número de actividades completadas y la duración de la sesión iniciada. Los modelos predictivos se entrenaron utilizando los métodos de Machine Learning como: Support Vector Machine, Logistic regression y Random Forest. Como herramientas utilizaron Python 3.6.0 y el paquete Scikit-Learn. Al final obtuvieron que el modelo de bosque aleatorio mostró el comportamiento más estable y la mejor precisión, más no especifica el porcentaje logrado [46].

Ndlovu y Ngala (2018), realizaron una investigación para predecir el rendimiento académico de los estudiantes de primer año en la universidad de Johannesburgo antes del resultado del examen de fin de año. Para esto, utilizaron un conjunto de datos de 247 instancias que consta de 5 atributos: Resultados promedio de la escuela secundaria, Autoestudio, Profesor competente, Asistencia a clase y Resultados promedios del primer semestre. Los modelos predictivos se entrenaron utilizando algoritmos de Machine Learning como: Bayesian Networks (BN), Artificial Neural Network (ANN), Naive Bayes (NB), Decision Trees y Support Vector Machine (SVM). Además, se usó SPSS para determinar qué atributos eran significativos y algoritmos de aprendizaje automático alojados en WEKA. Al final obtuvieron que el mejor modelo se construyó utilizando máquinas de vectores de soporte (SVM) con una precisión del 72.87% [41].

Nicoletti y Marqués (2018), realizaron una investigación para pronosticar el rendimiento final (Aprobado o Fallido) de los estudiantes en el curso de algoritmos y programación de la carrera de informática de una universidad de Brasil. Para esto, utilizaron las siguientes características: Fecha de nacimiento, estado civil, sexo, nivel de instrucción, grado en vestibular, Calificación en la Evaluación 1, Calificación en la Evaluación 2, calificación en Lógicas y Matemáticas Discretas. Para entrenar su modelo emplearon el algoritmo de Machine Learning J48 de Weka. Al final obtuvieron para su modelo un rendimiento mayor del 80% [42].

Yang y Li (2018), realizaron una investigación para estimar el rendimiento de los estudiantes de secundaria en una escuela de China. Para esto, recopilaron datos de 60 estudiantes de secundaria utilizando atributos relacionados con el rendimiento académico de 6 asignaturas y atributos no relacionados con el rendimiento académico. Este experimento utilizó Red Neuronal de Propagación Regresiva (BP-NN) basada en la clasificación como método de entrenamiento, y trabajó con 14 capas ocultas. Al final obtuvieron que los resultados de la evaluación muestran que las herramientas propuestas pueden dar resultados correctos y más precisos, y también ofrecer una mejor comprensión del progreso del estudiante [63].

Wham (2017), realizó un estudio para pronosticar los resultados de un estudiante universitario en un curso. Para esto, contó con 5.5 millones de registros únicos de más de 250,000 estudiantes únicos. Utilizó características predictivas que van desde los logros académicos previos del estudiante (GPA de la escuela secundaria, rango de clase de la escuela secundaria, puntajes del SAT, GPA acumulativo, etc.), las características demográficas del

estudiante (género, edad, etc.) participación en el campus (número de créditos inscritos, estado del atleta, tipo de solicitud, etc.) y datos históricos del curso individual e instructor (calificación promedio, nivel del curso, etc.). Para entrenar su modelo utilizaron algoritmos de aprendizaje automático basado en árbol de gradiente aumentado y XGBoost y utilizaron el lenguaje de programación estadística R y Python. Al final encontró que la precisión del modelo no estaba altamente vinculada a ninguna disciplina en particular. Más bien, un patrón emergente fue que los cursos que tendían a tener las predicciones menos precisas parecían tomarse más comúnmente al principio de la carrera de un estudiante, mientras que los cursos que tendían a tener las predicciones más precisas tendían a tomarse más adelante en la carrera académica [43].

Kaur, Umesh, y Singh (2018), realizaron una investigación para predecir el rendimiento académico del estudiante y de esta manera poder trabajar en las debilidades de los estudiantes oportunamente mediante el uso de pedagogías y enfoques correctos. Los autores, manifiestan que son muchos los factores que intervienen en el rendimiento del estudiante: Social, económico, personal, cultural, geográfico, institucional, medio ambiente, etc. y que los algoritmos de minería de datos ayudan a encontrar esos factores que están influyendo con el rendimiento del estudiante. En este estudio utilizaron 1735 instancias con 37 atributos, entre ellos calificaciones de los estudiantes, género, educación del papá, educación de la mamá, trabajo de la mamá, trabajo del papá, hábitos, etc. Para diseñar sus modelos se basan en el enfoque de aprendizaje automático, utilizando Support Vector Machine, Naive Bayes, Regression Logistic, C4.5 y el LMT (árbol de decisión de regresión logística, que combina un modelo de árbol de inducción con un modelo de regresión logística en las hojas del árbol).

Al final obtuvieron que el algoritmo híbrido LMT muestra una mayor precisión que los demás con un 99.48% [64].

Shingari y Kumar (2018), realizaron un trabajo de investigación para predecir el rendimiento académico (calificación final) de los estudiantes utilizando técnicas de minería de datos. Para esto, utilizaron datos de 36 estudiantes de ingeniería del primer año, tomando sus calificaciones. Para diseñar sus modelos utilizaron R como lenguaje de programación y los siguientes algoritmos Decision Tree, Random Forest, Linear Model y Neural Network. Al final obtuvieron que el árbol de decisión es el algoritmo de clasificación más adecuado, pero no precisaron cuál fue el porcentaje de precisión [44].

Saheed et al. (2018), realizaron una investigación para predecir el rendimiento de los estudiantes basado en técnicas de minería de datos. Para esto, el conjunto de datos se obtuvo de una Universidad privada de la parte norte de Nigeria. Las características analizadas fueron: Edad, Estado civil, Religión, Sexo, Nacionalidad, Genotipo, Ocupación de los padres, Ocupación de las madres, Curso solicitado de admisión, Curso admitido, Nivel, Modo de entrada, Año de entrada. En este estudio se indica que se han propuesto varios algoritmos estadísticos y de aprendizaje automático para analizar, predecir y clasificar el desempeño de los alumnos. Sin embargo, estos algoritmos de clasificación aún plantean problemas en términos de la clasificación de rendimiento. En ese sentido plantean un método para predecir el rendimiento de los estudiantes mediante el uso del dicotomizador iterativo 3 (ID3), C4.5, el árbol de Clasificación y Regresión (CART) y el uso de Weka. Al final obtuvieron que el C4.5 obtuvo un mejor desempeño que los demás algoritmos con una precisión de 98.4% [65].

Hoffait y Schyns (2017), realizaron un estudio para la detección temprana de estudiantes de la Universidad de Lieja que tendrían potencial dificultad para completar el primer año académico. Para esto, utilizaron datos de 6845 estudiantes con las siguientes características: Género, nacionalidad, el campo de estudio, información sobre su educación anterior, si se han registrado a la edad prevista de 17 años o más, si su plan de estudios incluía latín o griego, nivel en matemáticas y si estaba becado o no. Otros factores socioeconómicos, como el nivel educativo y la ocupación de sus padres no se tomaron en cuenta, porque estos datos no se registran y son difíciles de obtener, pero incluso cuando se han recopilado para un subconjunto limitado de estudiantes, no mejora los resultados. Para el diseño de sus modelos utilizaron Logistic regression, Artificial neural networks, Decision Tree y Random Forest. Al final obtuvieron que no todos los factores socioeconómicos deben incluirse en los experimentos finales, ya que además de ser difíciles de recopilar, su contribución no resulta significativa y que los algoritmos estándar de minería de datos todos proporcionar una tasa de clasificación correcta en torno al 70%, sin embargo, han demostrado que es posible aumentar la precisión de las predicciones agregando clases "inciertas" [45].

Fahri y Isa (2018), realizaron una investigación para predecir los estudiantes que necesitaran tutoría basándose en la motivación, el aprendizaje y la inteligencia emocional. Para esto, trabajaron con 65 atributos, los cuales fueron agrupados en 3 categorías, motivación del estudiante, interés de aprendizaje y la inteligencia emocional. Para diseñar sus modelos utilizaron métodos de minería de datos como la red neuronal de perceptron multicapa y el árbol de decisiones. Al final obtuvieron que el método de red neuronal de perceptron multicapa tuvo mejor resultado al predecir desde el principio (primer semestre) con una precisión del 88% [59].

Lakkaraju et al. (2015), realizaron una investigación para identificar a los estudiantes de la escuela secundaria de Estados Unidos que estarían en riesgo de no terminar la escuela a tiempo. Para esto utilizaron 24 características de los estudiantes, tales como su promedio de calificaciones, sus inasistencias, sus tardanzas, su género, atributos demográficos, etc. Los modelos predictivos se entrenaron utilizando algoritmos de clasificación de Machine Learning tales como: bosques aleatorios (RF), Adaboost (AB), regresión logística (LR), máquinas de vectores de soporte (SVM) y árboles de decisión (DT). Ejecutaron 100 iteraciones para cada modelo de clasificación. Durante cada iteración, cada modelo lo entrenaron en $N - k$ puntos de datos elegidos al azar, donde N es el tamaño de todo el conjunto de datos y $k = 0.01 * N$. Al final obtuvieron que con el algoritmo Random Forest se logró mejores resultados con una precisión del 70% al octavo grado del alumno [57].

Al-Shabandar et al. (2017), realizaron una investigación para predecir los resultados de aprendizaje en cursos masivos abiertos en línea (MOOC) utilizando datos de la Universidad de Harvard. Para este estudio se seleccionaron 9 características que tienen que ver con los patrones de comportamiento de los usuarios expresados mediante la secuencia de clics del estudiante que representa el número de eventos de usuarios relacionados con las vistas de conferencias de video, la interacción del contenido del curso, el acceso a las tareas y las publicaciones en foros de discusión. Además, utilizaron la edad, género, antecedentes educativos, la fecha de registro en el curso y la última actividad del estudiante. Los modelos predictivos se entrenaron usando varios algoritmos de aprendizaje automático como: Support Vector Machine, Decision Tree, Naive Bayes, Random Forest, Logistic regression, Feedforward Neural Network, Self-Organised Map y Linear Discriminant Analysis. Al final

obtuvieron que Random Forest (RF) y Support Vector Machine (SVM) lograron un rendimiento ideal, con los valores de precisión de 98.81% y 98.51% respectivamente [47] .

Conijn and Zaanen (2017), realizaron una investigación para predecir el rendimiento de los estudiantes a partir de los datos del sistema de gestión del aprendizaje Moodle de la Universidad de Tecnología de Eindhoven. Para esto, utilizaron 13 características del estudiante (número de sesiones, tiempo total en línea, duración media de la sesión de inicio de sesión, número total de clics, número promedio de clics por sesión, regularidad de inicios de sesión -frecuencia, tiempo sin actividad, número de mensajes en el foro, número de mensajes enviados, número de pruebas participadas, promedio de calificaciones, número de asignaciones participadas y calificaciones de asignación). Los modelos predictivos fueron entrenados usando algoritmos de Machine Learning como: Neural Networks, Naive Bayes, k-Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machine and Logistic Regression. Al final obtuvieron que la red neuronal supera a los otros seis clasificadores utilizados en términos de precisión con un 66.1% [48] .

Cazarez y Martin (2018), realizaron un estudio para evaluar la precisión de tres tipos de Redes Neuronales al intentar predecir el desempeño del estudiante en la educación superior online. Lo que buscaron fue determinar si existe diferencia significativa entre la precisión de de un modelo estadístico y un modelo con redes neuronales a la hora de predecir la calificación final de un curso de educación superior online. Para este estudio utilizaron las siguientes características de los estudiantes: La cantidad de tiempo que permanece en el curso y el número de ingresos realizados en determinado tiempo, las actividades online desarrolladas y las calificaciones obtenidas. Para entrenar su modelo utilizaron estos 3 tipos

de Redes Neuronales: Base radial (RNBR), perceptron multicapa con retropropagación (RNPMR) y regresión generalizada (RNRG). Al final obtuvieron que mediante la comparación estadística de las precisiones de predicción de las redes neuronales con la de una regresión lineal estadística, que las tres redes neuronales obtuvieron una mayor precisión de predicción y por lo tanto proponen a las redes neuronales como técnicas para una pronta predicción e identificación de estudiantes en riesgo de fracaso [49].

Ma, Yang, y Zhou (2018), realizaron un estudio para predecir las tasas de aprobación de los estudiantes en la educación online. Para esto, utilizaron 27 características representativas y completas de los alumnos entre las cuales tenemos: Sexo, edad, título, antecedentes de estudios, día de estudio, actividad de comunicación, numero de cursos opcionales, numero de tópicos, numero de respuestas a los tópicos, etc. En sus modelos utilizaron el algoritmo del árbol de decisión (DT), el algoritmo de máquina de vectores de soporte (SVM) y el algoritmo de red neuronal profunda (DNN). Al final encontraron que, en función de la interpretabilidad, la facilidad y el tiempo de ejecución (velocidad) de los tres algoritmos, fue mejor usar el algoritmo DT para predecir las tasas de aprobación de los estudiantes [50].

Hussain et al. (2018), realizaron un estudio para predecir dificultades de los estudiantes en el curso de diseño digital online (Calificación final). Para esto, utilizaron los datos registrados por un sistema de aprendizaje con tecnología mejorada (TEL) denominado educación electrónica digital y suite de diseño (DEEDS). Entre las características seleccionadas que usaron tenemos: tiempo promedio de inactividad, Número total de actividades y número promedio de pulsaciones. Para entrenar sus modelos utilizaron algoritmos de aprendizaje automático como: máquinas de vectores de soporte (SVM), redes neuronales artificiales

(ANN), clasificadores de Naive Bayes, regresión logística, y árboles de decisión. Al final obtuvieron que las ANN y las SVM logran una mayor precisión que otros algoritmos [51].

Kim, Vizitei y Ganapathi (2018), realizaron un estudio para predecir el rendimiento estudiantil (graduación de los estudiantes) en cursos en línea abiertos masivos (MOOC). Para esto, utilizaron datos de los estudiantes de dos programas de Udacity Nanodegree, teniendo como características los eventos de los estudiantes. Un evento representa a un usuario que realiza una acción específica (por ejemplo, ver un video, leer una página de texto, intentar un cuestionario o recibir una calificación en un proyecto). Para diseñar el modelo utilizaron métodos de aprendizaje profundo (GritNet). Al final demostraron la superioridad de GritNet utilizando datos de estudiantes de los programas Nanodegree de Udacity y dos propiedades novedosas de GritNet: No necesita ninguna ingeniería de características; es decir puede aprender de la entrada en bruto y además puede operar en cualquier evento de datos del estudiante asociado con una marca de tiempo incluso cuando está muy desequilibrado [52].

Nguyen, Nguyen, y Nguyen (2018), realizaron una investigación para pronosticar los resultados de aprendizaje de los estudiantes en cursos de Blended Learning basado en Learning Analytics. Para esto, utilizaron datos de los alumnos que participan en el sistema Moodle LMS empleando las siguientes características: Número de visitas y publicaciones de los alumnos sobre los materiales, la frecuencia de uso, el nivel de atención en los materiales, en qué momentos, las discusiones del curso, visitas y publicaciones de alumnos en los foros del curso, la evaluación de las tareas, envío de trabajos temprano y a tiempo, etc. Para diseñar sus modelos emplearon técnicas de clasificación del aprendizaje automático. Al final obtuvieron que el 75% de los estudiantes tienen resultados cercanos a los resultados

predichos con una precisión de más del 50% y que los resultados de la prueba muestran que si el rendimiento académico del estudiante se predice a partir de la primera semana del curso (3ª semana), los resultados serán bastante erróneos [53].

Chen et al. (2018), realizaron un estudio para predecir tempranamente los resultados de aprendizaje en cursos cortos en línea a través de las conductas de aprendizaje. En este estudio se centran en diseñar y evaluar un método que permita la predicción de resultados de aprendizaje en cursos cortos online donde hay pocos alumnos matriculados, y menos cantidad de calificaciones disponibles. Para esto recoge las mediciones del comportamiento de los alumnos con el contenido del curso (cantidad de clics) y la forma en que interactúan en las redes de aprendizaje sociales (SLN) (foros) del curso. Para evaluar el método se tomaron 3 cursos que tienen una duración de 2 semanas y que tiene un resultado final aprueba o desaprueba definido por el instructor. En el diseño del modelo se utilizaron 6 clasificadores de predicción: Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Forward Neural Network (ANN), Random Forest (RF), K-Nearest Neighbor (KNN) y Gradient Boosting (XGB) debido a que muestran buen desempeño en la predicción de resultados del estudiante. Al final concluyen que el modelo ayuda a los instructores en la detección temprana de resultados de aprendizaje y que la elección del algoritmo varía según el curso y el conjunto de características. Además, agregan que, con la excepción de ANN, hay al menos un par de cursos y características para las que cada algoritmo se desempeña mejor, y manifiestan que el bajo rendimiento de las redes neuronales se puede deber a los tamaños de muestra pequeños de estos cursos. Asimismo, agregan que los atributos de SLN se convirtieron en el conjunto de comportamientos más útil para la predicción a lo largo del tiempo, mientras que los atributos de contenido proporcionaron una mejor calidad para la detección "más temprana" en los primeros días [54].

Pérez et al. (2018), realizaron un estudio para predecir la deserción estudiantil de primer año en la universidad Bernardo O'Higgins - UBO (Santiago de Chile). Para este estudio utilizaron datos académicos, sociales o familiares, los cuales inicialmente se identificaron en base a la literatura revisada. Al final llegaron a identificar diecisiete variables, asociadas con la deserción estudiantil. Finalmente, seleccionaron siete variables predictoras y obtuvieron una precisión de su modelo del 86,4 % [55].

Fernández et al. (2018), realizaron una investigación para predecir la deserción estudiantil. Para este estudio utilizaron como características medidas demográficas, académicas y biopsicosociales con las etiquetas de salida (desertor / no desertor). Para el diseño del modelo de predicción de la deserción, utilizaron métodos de clasificación basada en K vecinos más cercanos y máquinas de vectores de soporte. Al final obtuvieron un 85.64% de precisión, concluyendo que el sistema propuesto apunta a servir como una herramienta útil para planificar estrategias para evitar que los estudiantes salgan de la universidad sin terminar sus estudios [56].

Ojha (2017), realizó una investigación para predecir el retraso, no retraso, un año o 2 a más años de retraso en la graduación de los estudiantes de la Universidad de Nuevo México (UNM). Para esto, seleccionaron 19 características entre pre institucionales e institucionales (escuela secundaria, promedio de la escuela secundaria, número de horas de crédito tomadas por un estudiante hasta el segundo semestre en UNM, género, lugar de residencia original, graduado reciente de secundaria, promedios finales del primer y segundo semestre en UNM, estado de residencia actual, etc.) utilizando los siguientes métodos: Selección hacia adelante

con Regresión lineal, la Eliminación hacia atrás con Regresión lineal y la Regresión de lazo. Los modelos predictivos se entrenaron utilizando algoritmos de Machine Learning como: Máquinas de vectores de soporte (SVM), Procesos gaussianos (GP) y Máquinas de Boltzmann profundo (DBM). Al final obtuvieron una precisión en la predicción entre el 68 y 75%. Determinaron que los algoritmos DBM superan a las SVM en términos de precisión y recuperación para clases individuales [58].

Durđević (2017), realizó una investigación para predecir la motivación académica del estudiante en función de su comportamiento en los cursos del sistema de gestión del aprendizaje (LMS) en una universidad de Croacia. Para esto utilizó 4 características de los estudiantes en función a si vio o no vio la asignación, el foro de discusión, el cuestionario y los recursos. Los modelos predictivos se entrenaron utilizando métodos de Machine Learning como redes neuronales, árboles de decisión y máquinas de vectores de soporte. Para establecer si existe una diferencia significativa en el rendimiento de los modelos, se utilizó una prueba t de la diferencia en proporciones. Al final obtuvo que, aunque todos los modelos tuvieron éxito, el modelo de red neuronal demostró ser el más exitoso en detectar la motivación académica de los estudiantes según su comportamiento en el curso de LMS con una precisión del 76.92 % [60].

Sreenivasa, Swapna, y Praveen (2018), realizaron una investigación para predecir la posibilidad de colocación de cada estudiante en un empleo. Para esto, utilizaron como características SSC%, B. Tech% y puntuación de Co-cubo. Para diseñar sus modelos utilizaron algoritmos de aprendizaje automático J48, Naive Bayes, Random Forest y Random Tree en la herramienta Weka y Regresión lineal múltiple, regresión logística binomial,

Partición recursiva y árbol de regresión (rpart), árbol de inferencia condicional (ctree) y algoritmos de red neuronal (nnet) en R Studio. Al final obtuvieron que los algoritmos de Random Forest y Random Tree dieron una precisión del 100%, en cambio J48 dio una precisión del 88.89% y Naive Bayes obtuvo un 61.10%. Además, encontraron que el porcentaje de B. Tech es el que más ha contribuido en la predicción [61].

Xu, Moon, y Van Der Schaar (2017), realizaron una investigación para predecir el rendimiento del estudiante en programas de pregrado en la Universidad de California en Los Ángeles (UCLA). Esta investigación buscó construir un sistema que pueda realizar un seguimiento continuo del rendimiento académico de los estudiantes y predecir con precisión su rendimiento futuro, como cuando es probable que se gradúen y sus GPA (promedios de calificaciones) finales estimados. Para esto, seleccionaron las siguientes características de los estudiantes: Rasgos preuniversitarios de los estudiantes (GPA de la escuela secundaria y puntajes de matemáticas SAT), los cursos (incluyendo conferencias y laboratorios) que los estudiantes toman en cada trimestre académico, los créditos del curso y las calificaciones obtenidas. Además, hicieron un estudio para seleccionar que cursos están relacionados con el rendimiento final del estudiante. Los datos se recopilaron en el departamento de Ingeniería Mecánica y Aeroespacial de UCLA, conteniendo datos de 1169 estudiantes de pregrado anónimos inscritos en el programa en dos áreas diferentes (Ingeniería mecánica e Ingeniería aeroespacial). Los modelos predictivos se entrenaron usando 4 métodos de Machine Learning: Linear regression, Logistic regression, Random Forest y k-Nearest Neighbors (kNN). Al final concluyeron que el bosque aleatorio realiza el mejor rendimiento, no especifican que porcentaje de precisión obtuvo y que los estudiantes con puntajes más altos en el SAT también obtienen un GPA final más alto en el programa de pregrado. Sin embargo,

el GPA de la escuela secundaria casi no se correlaciona con el GPA final en la universidad, lo que sugiere que el GPA de la escuela secundaria tiene un poder predictivo más débil que los puntajes del SAT [66].

Shanthini, Vinodhini, y Chandrasekaran (2018), realizaron una investigación para pronosticar el rendimiento académico de los estudiantes universitarios de ciencias e ingenierías. Para esto, utilizaron como características del estudiante: Sexo, edad, y los calificativos de 10 asignaturas de primer y segundo semestre y recolectaron datos de 401 estudiantes de pregrado que estaban matriculados en el 2014 y 2015. En este trabajo presentaron un modelo que utiliza métodos conjuntos. Un método conjunto es un enfoque de aprendizaje que combina varios modelos para resolver un problema. Utilizaron técnicas de clasificación de árbol de decisión basadas en cuatro algoritmos de aprendizaje representativos: Adaboost, Bagging, Dagging y Grading. REPTree se utiliza como el método de árbol de decisión para la meta del aprendizaje. Al final obtuvieron que Adaboost es el mejor clasificador de decisión de metadatos para predecir el resultado del alumno y que se hace necesario un análisis adicional para comprender mejor y mejorar estos resultados [67].

Martins, Miguéis, y Tec (2018), realizaron una investigación para predecir el éxito académico global de los estudiantes de las licenciaturas del Instituto Politécnico de Bragança. Para esto, seleccionaron 11 características entre datos académicos de los 6 primeros semestres, datos de carácter sociodemográfico y de acceso a la enseñanza superior, los cuales a su vez fueron clasificadas en dos importantes subgrupos: variables con resultados curriculares semestrales acumulados y variables intemporales, que son variables cuyos valores no se alteran a lo largo del recorrido del alumno. Para entrenar su modelo utilizaron el algoritmo Random Forest y

recolectaron datos que involucran registros de grupos de alumnos bastante heterogéneos, procedentes de más de medio centenar de licenciaturas que cubren las más diversas áreas educativas impartidas en las cinco escuelas de la institución. Al final obtuvieron un $R^2 = 0.79$ y encontraron que el éxito del estudiante depende de la escuela de donde proviene, de factores del contexto curricular de desempeño académico del estudiante y del nivel educativo de la madre [68].

Alyahyan y Düşteğör (2020), realizaron una revisión sistemática de la literatura para poder proporcionar un conjunto de pautas para los educadores que deseen aplicar técnicas de minería de datos para predecir el éxito de los estudiantes. Investigaron acerca de los factores relacionados con la predicción del éxito académico de los estudiantes (rendimiento académico). Al final concluyen que los datos demográficos de los estudiantes, los atributos psicológicos, su actividad de aprendizaje electrónico, son los factores mayormente reportados [38].

Poblete y Villa (2011), realizaron una investigación donde proponen un Sistema de Evaluación Basado en Situaciones Complejas (SEBSCO), el cual consiste en evaluaciones complejas que tienen como propósito evaluar el nivel de desempeño en competencias genéricas y específicas para los estudiantes universitarios [24].

Hensel y Robinson (2014), realizaron una investigación que tuvo como objetivo demostrar la manera en que la evaluación que hace el empleador sobre los estudiantes puede ser utilizada para evaluar el resultado de aprendizaje de los estudiantes y proponer mejoras en los programas de ingeniería. Además, al final recomiendan que se puedan hacer estudios

futuros donde se estudie la correlación entre las métricas de rendimiento observadas y los factores que influyen en el rendimiento de los estudiantes, así como la correlación cruzada entre la evaluación que hace el empleador y la autoevaluación de los estudiantes [25].

Martín (2015), propone un modelo para medir el grado de logro de las competencias en la educación superior. El modelo establece resultados de aprendizaje para cada competencia, luego define descriptores para cada resultado de aprendizaje, además define indicadores de logro para cada descriptor y el alineamiento con los contenidos del curso, luego selecciona los procedimientos e instrumentos de evaluación y se realiza el diseño de los ítems (tareas) a evaluar. Posteriormente construye la matriz de evidencias: Ítems, Contenidos e indicadores de logro y luego hace la revisión del diseño de los ítems de las pruebas establecidas. Finalmente probó la aplicabilidad del modelo en la medición de la competencia en el curso de Álgebra Lineal.

Estrada (2016), propone un modelo para evaluar los resultados de aprendizaje de competencias en varias disciplinas. Esta investigación, se centró en evaluar resultados de aprendizaje a nivel de asignaturas, concluyendo que la metodología propuesta resulta importante en lo académico frente a los problemas actuales que enfrenta la formación profesional cuando se divide entre disciplinas que no permite que se aborde todo objeto de conocimiento. Además, observaron que existen diferencias significativas de los resultados de aprendizaje entre la evaluación tradicional con los resultados obtenidos con la metodología propuesta [27].

Cruz (2017), realizó una investigación donde desarrolló un sistema para la evaluación continua del logro del perfil del egresado. Este sistema permite a la Escuela de Sistemas obtener información sobre el cumplimiento de las competencias en los estudiantes durante un determinado periodo académico. Utilizó Open UP, Framework Kumbia PHP y MYSQL. Los resultados muestran que el sistema apoya la gestión en la evaluación continua de las competencias del perfil de egreso. El sistema, es una aplicación transaccional que permite primero registrar las competencias del perfil de egreso del programa de estudio, también se puede registrar los cursos que aportan al logro de cada competencia, y además permite tomar exámenes finales que incluyen temas relacionados con la competencia a evaluar.. Al final se obtienen los reportes del nivel de logro del perfil del egresado. En base a esto concluye finalmente que el sistema apoya la evaluación del logro del perfil del egresado [28].

Shuman, Besterfield-Sacre, y Mcgourty (2005) se plantearon 2 interrogantes con respecto a los resultados del estudiante que plantea la acreditadora ABET para los programas de Ingeniería: ¿Se pueden enseñar? ¿Pueden ser evaluados?, al final de su investigación concluyen en base a varios casos que citaron que, si se pueden enseñar las habilidades profesionales que propone ABET, pero para esto se necesita un formato moderno de educación en Ingeniería, que utilice el aprendizaje activo, cooperativo y a través del servicio. Asimismo, concluye que, si pueden ser evaluados y que cada resultado del estudiante de ABET, generalmente está siendo evaluado en diversos grados y con mucho trabajo para hacer esto, agregando que el tema de cómo mejorar la evaluación es un desafío aún para varios investigadores [29].

Yue (2007), realiza una investigación sobre la evaluación efectiva de los resultados del estudiante basado en los resultados de los cursos, teniendo en cuenta el modelo de la acreditadora ABET para computación. La propuesta plantea que se defina para cada resultado del estudiante de ABET unos indicadores y luego se relacionan éstos con los cursos, los cuales a su vez tienen sus propios objetivos que posteriormente serán medidos utilizando encuestas a los estudiantes y la evaluación de un proyecto final. De esta manera los resultados que se obtengan en los cursos servirán para determinar los resultados que se obtengan con respecto a los resultados del estudiante de ABET. Finalmente concluye que este tipo de evaluación permitió la acreditación de un programa de sistemas de información computacional y recomienda que se enfatice más en la evaluación directa que en las encuestas a los estudiantes [30].

Bassett y Duan (2012), proponen un procedimiento de evaluación para vincular los resultados del aprendizaje del curso con los resultados del estudiante propuestos por la acreditadora ABET. El procedimiento establece que cada curso tenga sus objetivos, resultado de aprendizaje, contenidos y actividades, los cuales a su vez deben estar bien relacionados con el resultado del estudiante propuesto por ABET con el cual esté asociado. Para demostrar este procedimiento, utilizaron un curso de la escuela de Ingeniería Mecánica Eléctrica evaluando los logros de los estudiantes de acuerdo con un formulario de evaluación propuesto [31].

Imam et al. (2016), en su estudio indican que los resultados de aprendizaje de un curso y los resultados del estudiante (Students Outcomes) deben evaluarse para la acreditación ABET. En ese sentido, una pregunta en una evaluación debe apuntar a un resultado de aprendizaje

de un curso, así como a los resultados del estudiante y debe representar un nivel requerido de la competencia. Debido a esto proponen un sistema experto (ES) que permita seleccionar la pregunta más adecuada para evaluar un resultado del estudiante según los requerimientos de acreditación de ABET. Para la implementación de ES, se consideran siete atributos de una pregunta, los resultados de aprendizaje y los niveles de la taxonomía de Bloom. Una base de datos contiene todos los datos sobre un curso, incluidos los temas del contenido del curso, los resultados del aprendizaje del curso y la matriz de relaciones resultados de aprendizaje de un curso y resultado del estudiante ABET. La base de conocimientos del ES presentado contiene un conjunto de preguntas, cada una con etiquetas de los atributos especificados que consolidan las opiniones de los expertos. Luego de esto, con la base de reglas implícita, el motor de inferencia encuentra la mejor pregunta posible que satisface los atributos requeridos. El motor de inferencia del sistema experto utiliza técnicas de búsqueda y coincidencia para encontrar las preguntas en la base de conocimientos que coincidan con el criterio del usuario, devolviendo finalmente la mejor elección de pregunta desde la base de conocimiento [34].

Tasadduq et al. (2017), realizaron una investigación que permita transformar los datos de cumplimiento de los resultados de aprendizaje del curso (CLO) y de esta manera obtener los datos del cumplimiento de los resultados de estudiante (SO) de ABET. En ese sentido, teniendo en cuenta la naturaleza difusa de las métricas de CLO y SO, propusieron un algoritmo de lógica difusa para extraer los datos de cumplimiento de SO de los datos de satisfacción de CLO para cualquier curso dado. Para esto, establecieron las funciones de pertenencia para las variables difusas, es decir, las relaciones CLO, SO y CLO-SO y se definieron un conjunto de 24 reglas formando la base de reglas del algoritmo de lógica difusa.

El algoritmo fue implementado y probado en MATLAB, concluyendo finalmente que la idea presentada ayudará a los instructores y administradores de los programas académicos que buscan la acreditación ABET y que este trabajo es una aplicación única de la teoría de conjuntos difusos no presentada anteriormente en la literatura y que se requiere de investigación adicional para encontrar las mejores funciones de membresía y el efecto de las variaciones de los parámetros léxicos en la salida [35].

III. MATERIALES Y MÉTODOS

La investigación, para el desarrollo de los objetivos, siguió principalmente las fases que establece la metodología Cross Industry Standard Process for Data Mining (CRISP-DM) [70] (Fig. 1):

Comprensión del problema: En esta fase, se analizaron los diferentes planes curriculares de las escuelas de la Facultad de Ingeniería (Ingeniería de Sistemas, Ingeniería Civil, Ingeniería Industrial e Ingeniería Mecánica). Se identificaron las competencias del perfil de egreso (Students Outcomes) con las que trabaja cada programa y se analizó la forma y escalas con que se mide actualmente el nivel de logro del perfil de egreso. Asimismo, se determinaron aquellas asignaturas de los planes de estudio que son comunes a todos los programas de ingeniería y los atributos de los estudiantes con los que cuenta la universidad que podrían servir como posibles predictores del nivel de logro del perfil de egreso.

Comprensión de los datos: En esta fase, se recolectaron datos de 1982 egresados hasta el año 2021 de las diferentes escuelas de la Facultad de Ingeniería de una Universidad Peruana. Datos de 710 egresados de la Escuela de Ingeniería Industrial, 678 egresados de la Escuela de Ingeniería de Sistemas, 424 egresados de la Escuela de Ingeniería Civil y 170 egresados de Ingeniería Mecánica. Luego de esto, se realizó un análisis exploratorio de los datos. Todos los datos fueron obtenidos manteniendo el anonimato del estudiante de una Base de Datos que está en Microsoft SQL SERVER.

Preparación de los datos: En esta fase, se evaluó si era necesario realizar una limpieza (Data Cleaning), reducción (Data Reduction) o transformación (Data Transformation) de algunos datos. Luego de esto, se seleccionaron las características o atributos del estudiante que nos permitan entrenar el modelo y obtener una mayor precisión. Para esto se combinaron técnicas de selección de características: métodos de filtro (Relación de las características con la variable de salida) y algoritmos de aprendizaje automático. Se utilizó el algoritmo de máxima relevancia, mínima redundancia (MRMR)[71] y los demás algoritmos que están disponibles desde la App Classification Learner de Matlab.

Modelado: En esta fase, se separaron los datos en 2 conjuntos, uno para entrenamiento y otro para las pruebas del modelo. De los 1982 egresados de ingeniería, se tomó como datos de prueba a todos los egresados del 2020 y 2021 que representan el 25% (494 egresados) y se entrenó el modelo con el 75% restante (1488 egresados). Luego de esto, haciendo uso de MatLab y los datos de entrenamiento, se evaluaron los diferentes algoritmos de clasificación de Machine Learning (Fig. 2) para determinar el algoritmo que nos da la mayor precisión. El método de validación utilizado fue de Cross-Validation con 5 folds. Posteriormente, utilizando el algoritmo seleccionado y los datos de entrenamiento, se construyó el modelo computacional que permite predecir el nivel de logro del perfil de egreso de nuevos estudiantes de Ingeniería.

Prueba del modelo: En esta fase, utilizando los datos de pruebas (494 egresados), se probó el modelo entrenado con todos los egresados de Ingeniería del año 2020 y 2021 (Civil: 180, Industrial: 199, Mecánica: 54 y Sistemas: 61). Asimismo, se construyeron las matrices de

confusión en base a los datos de validación y de pruebas para poder observar el comportamiento del modelo.

Despliegue: Como parte de esta fase, se creó un programa en MatLab que permite utilizar el modelo entrenado y predecir resultados futuros.

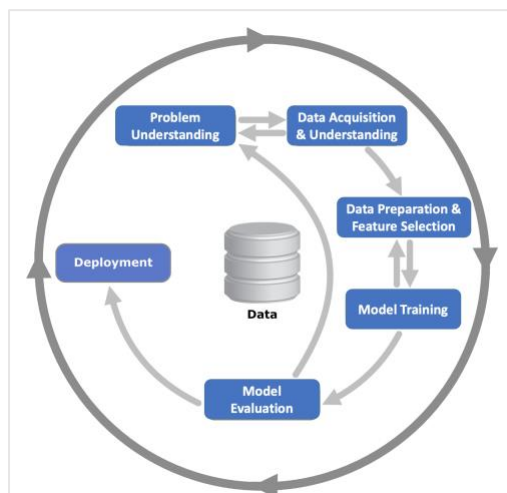


Fig. 1 Metodología Cross Industry Standard Process for Data Mining (CRISP-DM) [69]

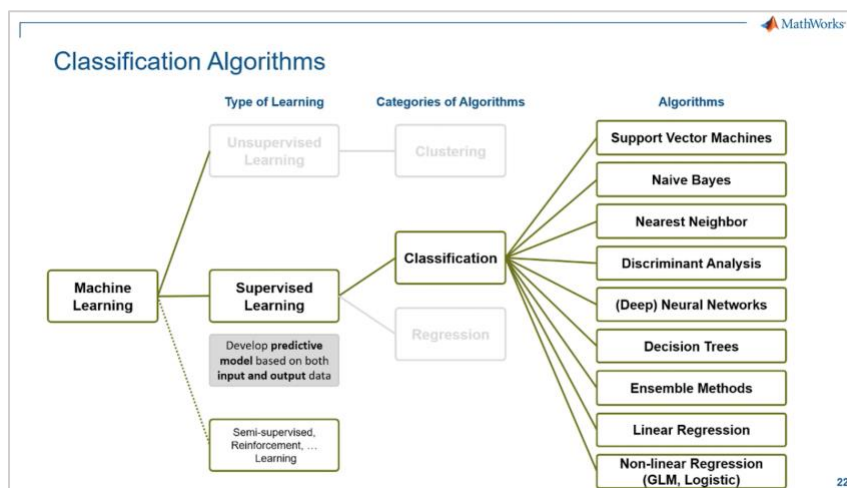


Fig. 2 Algoritmos de clasificación de Machine Learning [72]

IV. RESULTADOS

4.1. ANÁLISIS DEL PERFIL DE EGRESO DE LOS ESTUDIANTES DE INGENIERÍA

La Facultad de Ingeniería de la universidad peruana bajo estudio cuenta con las siguientes carreras profesionales: Ingeniería de Sistemas, Ingeniería Industrial, Ingeniería Civil e Ingeniería Mecánica, las cuales dentro de sus planes curriculares han definido competencias genéricas y específicas para sus perfiles de egreso. Estas competencias del perfil de egreso han sido alineadas con el perfil de egreso (students outcomes) que propone la agencia acreditadora ABET e ICACIT para los programas de ingeniería; las mismas que tienen que ser evaluadas para determinar el nivel de logro por parte de los estudiantes (Fig. 3).

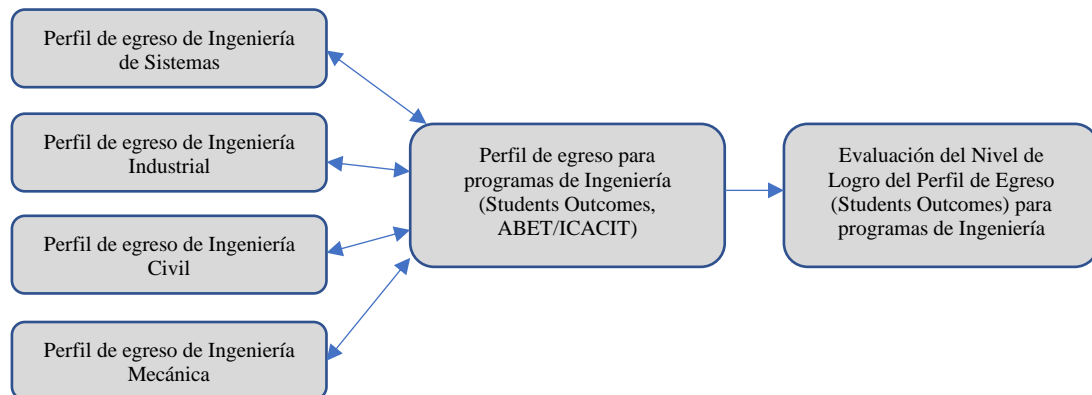


Fig. 3 Esquema de alineación del perfil de egreso con students outcomes para ingeniería

ICACIT es una agencia acreditadora especializada en programas de computación, ingeniería y tecnología en ingeniería que promueve la mejora continua de la calidad educativa de los programas, garantizando que éstos cumplan con los más elevados estándares internacionales que aseguren que los graduados se encuentren listos para

ejercer su profesión[73]. ICACIT, dentro de su modelo calidad, el cual está basado en el modelo de ABET, propone 12 students outcomes [74] los cuales se respaldan en el Washington Accord, que es considerado como un estándar internacional para el reconocimiento de los programas de ingeniería en diferentes países del mundo.

TABLA I. PERFIL DE EGRESO (STUDENTS OUTCOMES) PARA PROGRAMAS DE INGENIERÍA PROPUESTOS POR ICACIT Y ABET

Código	Perfil de egreso (Students Outcomes) para programas de ingeniería	Descripción
RE-I01	Conocimientos de Ingeniería	Aplica conocimientos de matemáticas, ciencias e ingeniería en la solución de problemas complejos de ingeniería
RE-I02	Análisis de Problemas	Identifica, formula, busca información y analiza problemas complejos de ingeniería para llegar a conclusiones fundamentadas usando principios básicos de matemáticas, ciencias naturales y ciencias de la ingeniería
RE-I03	Diseño o Desarrollo de Soluciones	Diseña soluciones para problemas complejos de ingeniería y diseña sistemas, componentes o procesos para satisfacer necesidades deseadas dentro de restricciones realistas en los aspectos de salud pública y seguridad, cultural, social, económico y ambiental
RE-I04	Investigación	Conduce indagaciones de problemas complejos de ingeniería usando conocimientos basados en la investigación y métodos de investigación incluyendo el diseño y la conducción de experimentos, el análisis y la interpretación de información, y la síntesis de información para producir conclusiones válidas
RE-I05	Uso de Herramientas Modernas	Crea, selecciona y utiliza técnicas, habilidades, recursos y herramientas modernas de la ingeniería y las tecnologías de la información, incluyendo la predicción y el modelamiento, con la comprensión de sus limitaciones
RE-I06	Ingeniería y Sociedad	Aplica el razonamiento informado mediante el conocimiento contextual para evaluar cuestiones sociales, de salud, de seguridad, legales y culturales y las consecuentes responsabilidades relevantes para la práctica profesional de la ingeniería
RE-I07	Medio Ambiente y Sostenibilidad	Comprende y evalúa el impacto de las soluciones a problemas complejos de ingeniería en un contexto global, económico, ambiental y social
RE-I08	Ética	Aplica principios éticos y se compromete con la ética profesional y las responsabilidades y normas de la práctica de la ingeniería
RE-I09	Trabajo Individual y en Equipo	Se desenvuelve eficazmente como individuo, como miembro o líder de equipos diversos
RE-I10	Comunicación	Se comunica eficazmente, mediante la comprensión y redacción de informes y documentación de diseño, la realización de exposiciones, y la transmisión y recepción de instrucciones claras
RE-I11	Gestión de Proyectos	Demuestra conocimiento y comprensión de los principios de gestión en ingeniería y la toma de decisiones económicas, y su respectiva aplicación

Fuente: ICACIT [73]

Los programas de ingeniería para evaluar los niveles de logro de su perfil de egreso han elegido asignaturas claves de los últimos ciclos, donde se desarrollan actividades integradoras que les permiten medir los indicadores de desempeño que han definido para cada competencia de su perfil de egreso (Students Outcomes). Llegando al final a establecer un nivel de logro global para cada estudiante en base al promedio de las calificaciones (escala vigesimal) obtenidas (Tabla II). Este proceso se realiza en cada semestre, el tema es que luego de obtener estos resultados es demasiado tarde como para poder tomar acciones correctivas a tiempo, esto debido a que los estudiantes están terminando sus estudios. Debido a esto, se definió como objetivo del modelo: Predecir el nivel de logro del perfil de egreso de los estudiantes universitarios de ingeniería antes que terminen sus estudios.

TABLA II. NIVELES DE LOGRO DEL PERFIL DE EGRESO

Promedio de Calificación	Nivel
De 0 a 8	No logrado
De 9 a 13	Insuficiente
De 14 a 15	Suficiente
De 16 a 17	Notable
De 18 a 20	Sobresaliente

En ese sentido, para poder lograr el perfil de egreso, los programas de ingeniería dentro de sus planes curriculares han definido un plan de estudios que tiene una duración de 5 años distribuidos en 10 semestres académicos. Estos planes de estudios fueron analizados llegando a identificar aquellas asignaturas comunes o equivalentes

en todos los planes de estudios de los programas de ingeniería (Tabla III), esto debido a que se tiene como propósito crear un modelo que sirva para cualquier programa de ingeniería. Asimismo, se hizo un esquema de malla curricular para poder ver la ubicación y relación entre todas las asignaturas comunes que tienen todos los programas de ingeniería (Fig. 4). Además se hizo una clasificación de las asignaturas por área del conocimiento Tabla IV.

TABLA III. ASIGNATURAS COMUNES DE LOS PROGRAMAS DE INGENIERÍA

Semestre	Asignatura del plan vigente	Asignaturas equivalentes en planes de estudios antiguos	Variable para el modelo
I	Matemática básica	Matemática básica	I_MAT BAS
	Física de los cuerpos rígidos	Física I	I_FIS I
	Estrategias para el aprendizaje autónomo	Metodología del trabajo intelectual	I_MTI
	Comprensión de textos y redacción básica	Lengua y comunicación I Lenguaje y comunicación I Comunicación I	I LENG I
II	Cálculo de una variable	Cálculo I Análisis matemático I Matemática para ingenieros I	II_CALC I
	Comprensión y redacción de textos académicos	Lengua y comunicación II Lenguaje y comunicación II Comunicación II	II LENG II
	Física de la masa y energía	Física II	II_FIS II
III	Cálculo de varias variables	Cálculo II Análisis matemático II Matemática para ingenieros II	III_CALC II
	Filosofía	Filosofía	III_FIL
	Electricidad y Magnetismo	Física III	III_FIS III
IV	Ecuaciones Diferenciales	Cálculo III Análisis matemático III Matemática para ingenieros III	IV_CALC III
	Antropología Filosófica	Antropología Filosófica	IV_ANT FIL

TABLA IV. CLASIFICACIÓN DE ASIGNATURAS POR ÁREA

Área	Asignaturas
Matemáticas	Matemática Básica
	Cálculo I
	Cálculo II
	Cálculo III
Física	Física I
	Física II
	Física III
Humanidades	Metodología del Trabajo Intelectual
	Lenguaje y comunicación I
	Lenguaje y comunicación II
	Filosofía
	Antropología

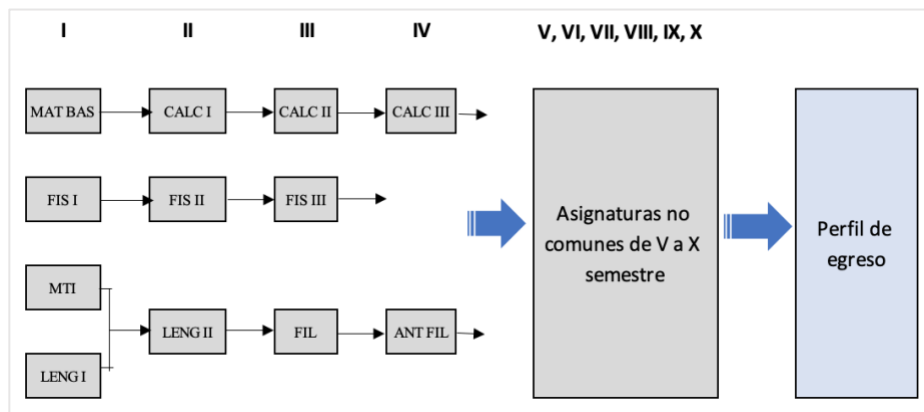


Fig. 4 Esquema de malla curricular de asignaturas comunes de los programas de ingeniería

Además, teniendo en cuenta la literatura revisada y los diferentes datos que se recogen o generan durante la formación profesional de los estudiantes, se identificaron las características o atributos de éstos que pudieran servir como posibles predictores del nivel de logro del perfil de egreso (Tabla V).

TABLA V. RELACIÓN DE ATRIBUTOS O CARACTERÍSTICAS DEL ESTUDIANTE

Característica	Descripción	Tipo
SEXO	Indica si el estudiante es de sexo masculino (1) o femenino (0)	Categórica
TIPO_COLEGIO	Muestra si el estudiante terminó la secundaria en un colegio privado (0) o público (1)	Categórica
UBICACION_COL	Muestra si el colegio donde terminó secundaria está ubicado en la misma localidad que la universidad (1) o no (0)	Categórica
TIPO_SEMESTRE_INGRESO	Indica si el estudiante ingresó a la universidad en el primer semestre (1) o en el segundo (0) semestre del año	Categórica
EDAD_INGRESO	Indica la edad que tenía el estudiante cuando ingresó a la universidad	Numérica
PLAN_ESTUDIO	Es el plan de estudio que tiene asignado el estudiante y que debe completar a lo largo de toda su formación universitaria	Categórica
ESCUELA	Es la escuela profesional o programa al que pertenece el estudiante: Ingeniería de Sistemas, Ingeniería Civil, Ingeniería Industrial e Ingeniería Mecánica	Categórica
MODALIDAD_INGRESO	Modalidad del proceso de admisión con el que ingresó el estudiante. Puede ser ordinaria 1 (Examen de admisión, test DHAC, evaluación preferente) o extraordinaria 0 (Escuela pre, primeros puestos, traslado externo, graduados y titulados, etc.)	Categórica
SEMESTRE_INGRESO	Muestra el semestre en que ingresó el estudiante a la universidad. Ejemplo 2015-I, 2015-II, 2016-I, etc.	Categórica
I_MAT_BAS	Promedio de las calificaciones finales que obtuvo cuando llevó la asignatura de Matemática Básica o una equivalente en su plan de estudios	Numérica
I_MAT_BAS_VECES	Número de veces que se matriculó en la asignatura de Matemática Básica.	Numérica
I_FIS_I	Promedio de las calificaciones finales que obtuvo cuando llevó la asignatura de Física de los cuerpos rígidos o una equivalente en su plan de estudios	Numérica
I_FIS_I_VECES	Número de veces que se matriculó en la asignatura de Física de los cuerpos rígidos.	Numérica
I_MTI	Promedio de las calificaciones finales que obtuvo cuando llevó la asignatura de Estrategia para el aprendizaje autónomo o una equivalente en su plan de estudios	Numérica
I_MTI_VECES	Número de veces que se matriculó en la asignatura de Estrategia para el aprendizaje autónomo.	Numérica
I LENG_I	Promedio de las calificaciones finales que obtuvo cuando llevó la asignatura de Comprensión de textos y redacción básica o una equivalente en su plan de estudios	Numérica
I LENG_I_VECES	Número de veces que se matriculó en la asignatura de Comprensión de textos y redacción básica.	Numérica
II_CALC_I	Promedio de las calificaciones finales que obtuvo cuando llevó la asignatura de Cálculo de una variable o una equivalente en su plan de estudios	Numérica
II_CALC_I_VECES	Número de veces que se matriculó en la asignatura de Cálculo de una variable	Numérica

TABLA V. RELACIÓN DE ATRIBUTOS O CARACTERÍSTICAS DEL ESTUDIANTE

Característica	Descripción	Tipo
II LENG II	Promedio de las calificaciones finales que obtuvo cuando llevó la asignatura de Comprensión y redacción de textos académicos o una equivalente en su plan de estudios	Numérica
II LENG II VECES	Número de veces que se matriculó en la asignatura de Comprensión y redacción de textos académicos	Numérica
II FIS II	Promedio de las calificaciones finales que obtuvo cuando llevó la asignatura de Física de la masa y energía o una equivalente en su plan de estudios	Numérica
II FIS II VECES	Número de veces que se matriculó en la asignatura de Física de la masa y energía	Numérica
III CALC II	Promedio de las calificaciones finales que obtuvo cuando llevó la asignatura de Cálculo de varias variables o una equivalente en su plan de estudios	Numérica
III CALC II VECES	Número de veces que se matriculó en la asignatura de Cálculo de varias variables	Numérica
III FIS III	Promedio de las calificaciones finales que obtuvo cuando llevó la asignatura de Electricidad y magnetismo o una equivalente en su plan de estudios	Numérica
III FIS III VECES	Número de veces que se matriculó en la asignatura de Electricidad y magnetismo	Numérica
III FIL	Promedio de las calificaciones finales que obtuvo cuando llevó la asignatura de Filosofía o una equivalente en su plan de estudios	Numérica
III FIL VECES	Número de veces que se matriculó en la asignatura de Filosofía	Numérica
IV CALC III	Promedio de las calificaciones finales que obtuvo cuando llevó la asignatura de Ecuaciones diferenciales o una equivalente en su plan de estudios	Numérica
IV CALC III VECES	Número de veces que se matriculó en la asignatura de Ecuaciones diferenciales	Numérica
IV ANT FIL	Promedio de las calificaciones finales que obtuvo cuando llevó la asignatura de Antropología filosófica o una equivalente en su plan de estudios	Numérica
IV ANT FIL VECES	Número de veces que se matriculó en la asignatura de Antropología filosófica	Numérica

4.2. ANÁLISIS, PREPARACIÓN DE DATOS Y SELECCIÓN DE CARACTERÍSTICAS

Análisis de datos

En esta fase, manteniendo el anonimato del estudiante, se recolectaron datos de 1982 egresados de las diferentes escuelas de la Facultad de Ingeniería de una Universidad

Peruana hasta el año 2021 y fueron procesados haciendo uso de Matlab. Como se puede ver en la Fig. 5, se obtuvieron datos de 710 egresados de la Escuela de Ingeniería Industrial, 678 egresados de la Escuela de Ingeniería de Sistemas, 424 egresados de la Escuela de Ingeniería Civil y 170 egresados de Ingeniería Mecánica. Estos datos fueron obtenidos de una Base de Datos en Microsoft SQL SERVER.

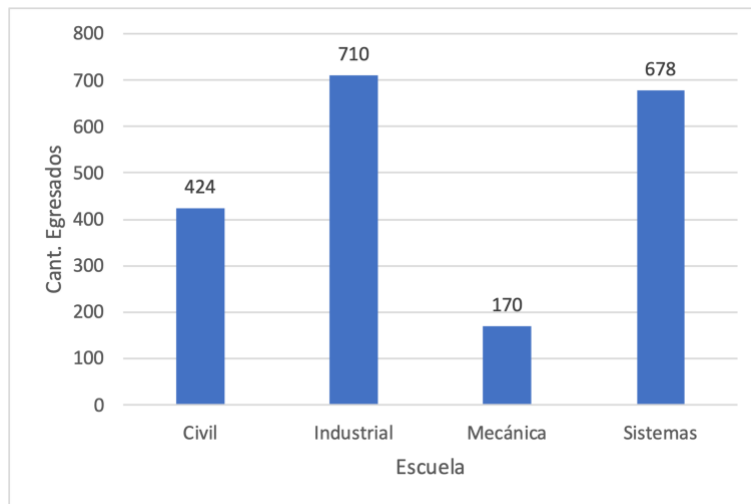


Fig. 5 Cantidad de egresados por escuela profesional hasta el año 2021

Esta cantidad de egresados por escuela en cada año ha ido aumentando como se puede ver en la Fig. 6 y Tabla VI, siendo la escuela de Ingeniería Industrial junto con la escuela de Ingeniería de Sistemas las que más egresados tienen hasta el año 2021, esto debido a que son las 2 escuelas más antiguas en la Facultad. Sin embargo, en el año pasado, la escuela de Ingeniería Civil es la que más egresados ha tenido.

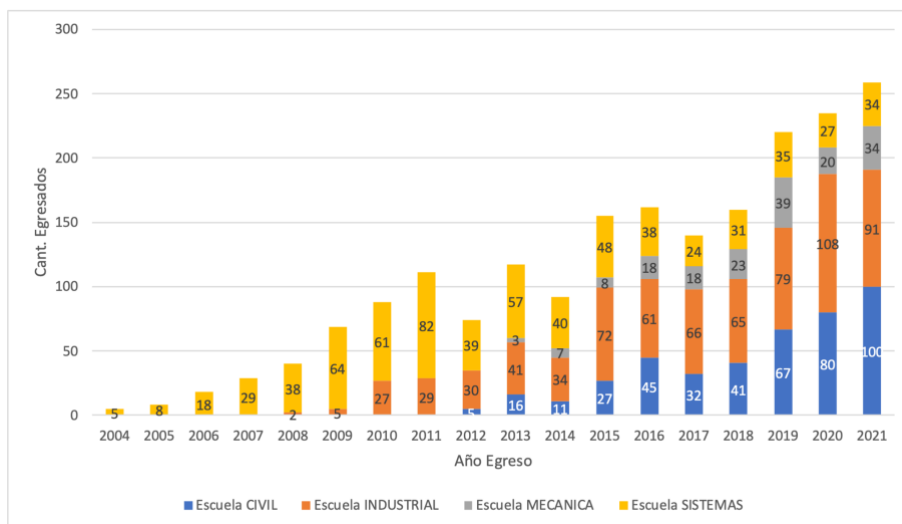


Fig. 6 Cantidad de egresados por escuela y año de egreso

TABLA VI. CANTIDAD DE EGRESADOS POR ESCUELA Y AÑO DE EGRESO

Año Egreso	Escuela				Total
	CIVIL	INDUSTRIAL	MECÁNICA	SISTEMAS	
2004	0	0	0	5	5
2005	0	0	0	8	8
2006	0	0	0	18	18
2007	0	0	0	29	29
2008	0	2	0	38	40
2009	0	5	0	64	69
2010	0	27	0	61	88
2011	0	29	0	82	111
2012	5	30	0	39	74
2013	16	41	3	57	117
2014	11	34	7	40	92
2015	27	72	8	48	155
2016	45	61	18	38	162
2017	32	66	18	24	140
2018	41	65	23	31	160
2019	67	79	39	35	220
2020	80	108	20	27	235
2021	100	91	34	34	259
Total	424	710	170	678	1982

Asimismo, también se recolectó como dato la modalidad de admisión por la que ingresó el estudiante en su momento. La universidad tiene varias modalidades de ingreso agrupadas en 2: Modalidad de Ingreso Ordinaria (examen de admisión, test DHAC y preferente) y la modalidad extraordinaria (Traslado externo, primeros puestos, graduados y titulados, beca 18 y la escuela pre universitaria). Como se puede ver en la Fig. 7, la mayor cantidad de estudiantes que ya egresaron ingresó en su momento por la modalidad de examen de admisión, seguido del test DHAC, escuela pre universitaria y evaluación preferente.

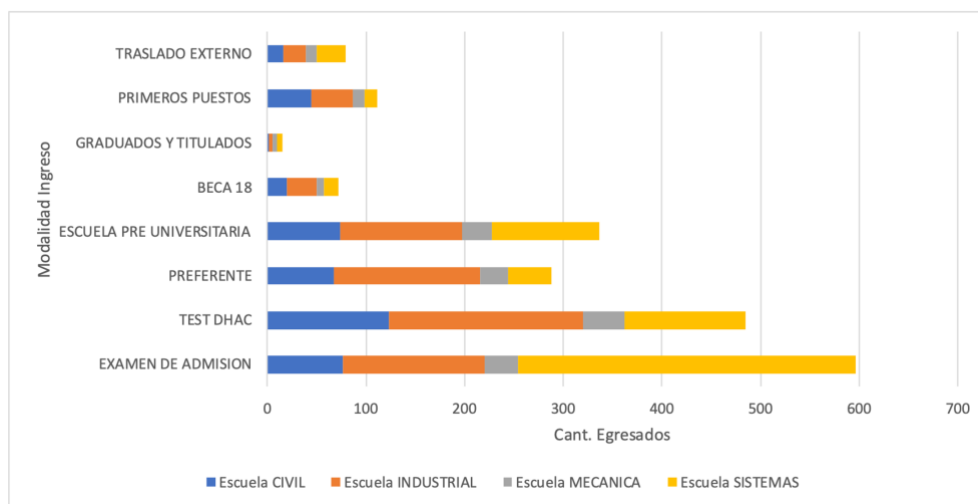


Fig. 7 Cantidad de egresados por escuela profesional y modalidad con la que ingresó

De los 1982 egresados de ingeniería que tiene la universidad hasta el año 2021, el 63% son hombres y el 37% son mujeres (Fig. 8). Sin embargo, la escuela de Ingeniería Industrial es la única escuela de ingeniería donde la cantidad de egresados mujeres es mayor que los hombres, y en la escuela de Ingeniería Mecánica son muy pocas las mujeres que han egresado con respecto a los hombres (Fig. 9).

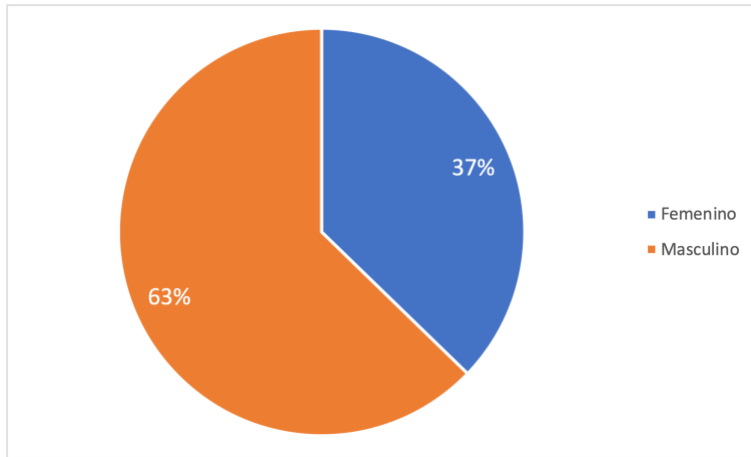


Fig. 8 Porcentaje de egresados según sexo

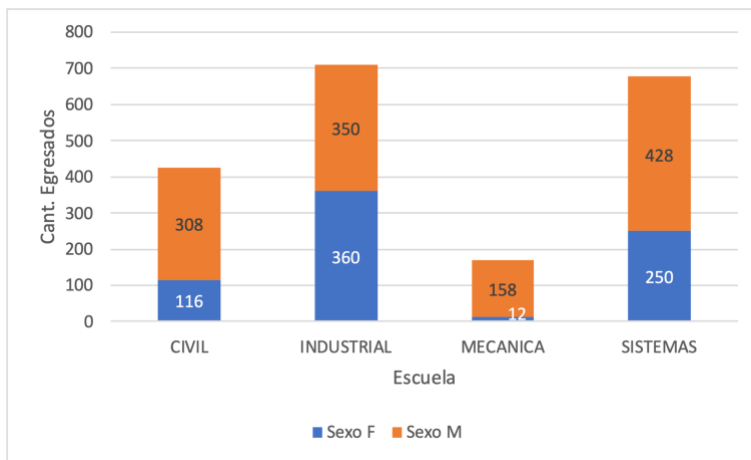


Fig. 9 Cantidad de egresados masculinos (M) y femeninos (F) por escuela

Con respecto al tipo de colegio donde estudió el egresado antes de ingresar a la universidad (Fig. 10), el 58% proviene de un colegio privado y el 42% de un colegio público, entendiéndose esto en el sentido que la universidad también es privada. Asimismo, al analizar por cada escuela observamos también lo mismo con excepción de la escuela de Ingeniería Civil, donde la mayoría de los egresados proviene de un colegio público (Fig. 11).

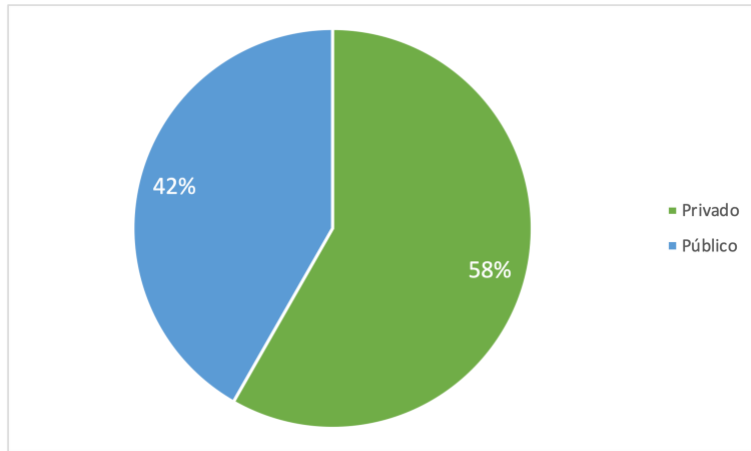


Fig. 10 Porcentaje de egresados según el tipo de colegio (público/privado) donde estudió

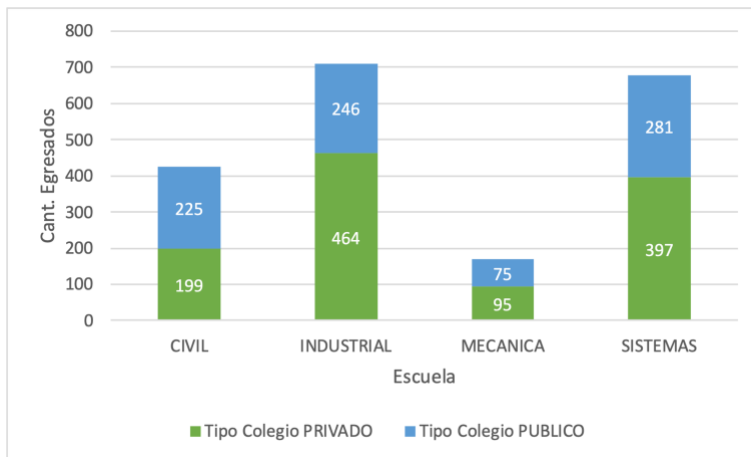


Fig. 11 Cantidad de egresados por escuela y tipo de colegio

Del mismo modo, se analizó la ubicación del colegio (Fig. 12). Encontrándose que la mayoría de los egresados provenían de la misma localidad donde está ubicada la universidad (Lambayeque – Perú)

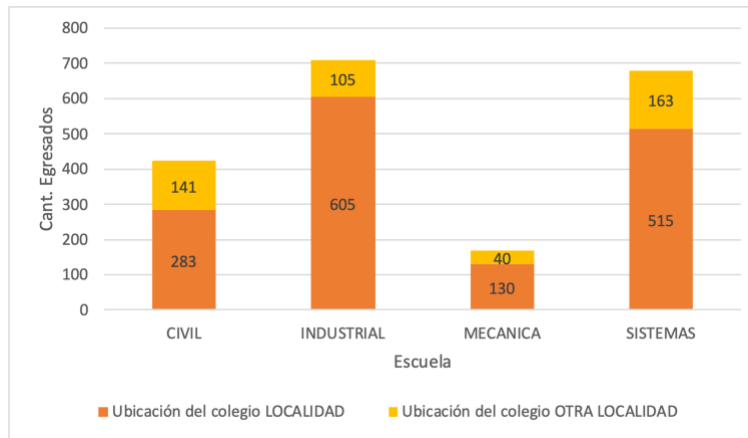


Fig. 12 Cantidad de egresados por escuela y ubicación de colegio

Al analizar la variable de salida (Fig. 13); es decir la variable objetivo o etiqueta que predecirá el modelo, nos encontramos que, de los 1982 egresados, el 32% no logró el perfil de egreso, sólo el 11% de los egresados logró un nivel sobresaliente y el 20% un nivel notable. El mayor porcentaje (38%) alcanzó un nivel suficiente; es decir que lograron en promedio los indicadores mínimos establecidos para cada competencia del perfil de egreso. Sin embargo, si lo analizamos por escuela tal como se puede ver en la Fig. 14, el porcentaje de egresados que no lograron las competencias del perfil de egreso (No logrado + insuficiente) es mayor en Ingeniería Industrial (34%) seguido de Ingeniería de Sistemas (33%).

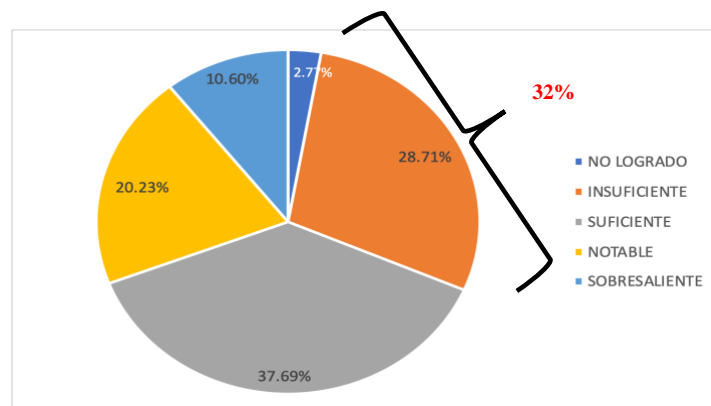


Fig. 13 Porcentaje de egresados según el nivel de logro del perfil de egreso

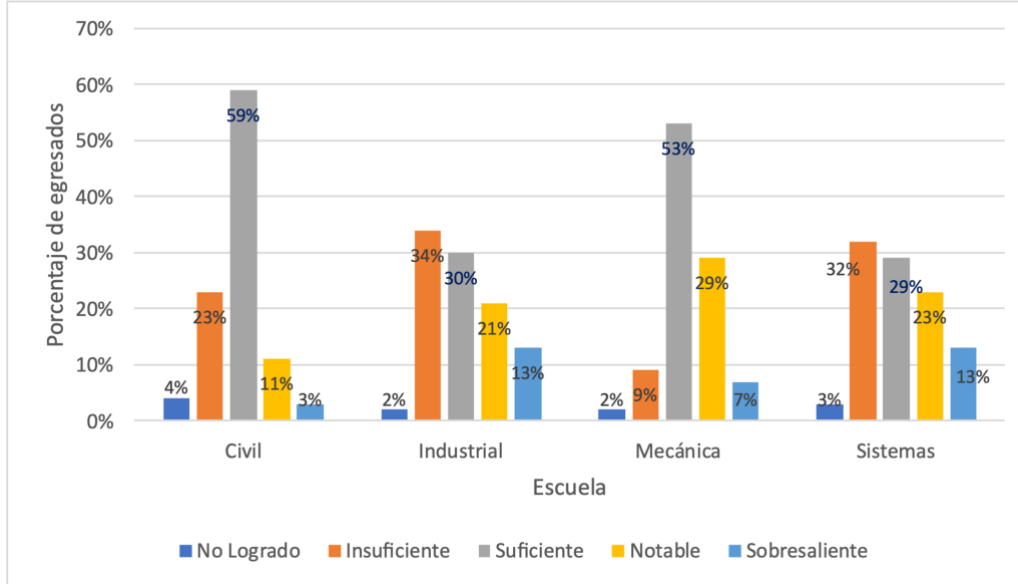


Fig. 14 Porcentaje de egresados por escuela y su nivel de logro del perfil de egreso

Asimismo, se elaboró un mapa de calor (Fig. 15) para poder ver donde está ubicada la mayor cantidad de egresados de cada escuela según el nivel de logro del perfil de egreso alcanzado. Encontrando que en la escuela de Industrial e Ingeniería de Sistemas, la mayor cantidad de sus egresados está en el nivel de logro insuficiente y suficiente. Sin embargo, en las otras escuelas la mayor cantidad de los egresados está en el nivel de logro suficiente.



Fig. 15 Mapa de calor del nivel de logro del perfil de egreso por escuela profesional

Para el análisis de las variables numéricas, haciendo uso de Matlab, se determinaron las medidas de tendencia central, valores faltantes, atípicos y la relación entre los datos.

En la Tabla I, se muestra un resumen descriptivo de las variables numéricas. Podemos ver que la edad a la que ingresa la mayoría de los estudiantes es 17 años. La nota más baja que se ha encontrado en alguna de las asignaturas comunes a todas las escuelas es de 5 y la nota más alta es de 20. Con respecto a las veces que un alumno llevó una misma asignatura tenemos que hay estudiantes que han llevado desde 1 hasta 7 veces. Cabe mencionar que cuando un estudiante lleva un curso varias veces la nota que se está tomando para el análisis es la nota promedio. Con respecto al área de Matemáticas: El calificación promedio en la asignatura de Matemática Básica es 13.93, en Cálculo I es de 14.69, en Cálculo II es de 14.36 y en Cálculo III es de 13.99. En el área de Física: La nota promedio en el curso de Física I es de 13.91, en el curso de Física II es de 14.47 y en el de Física III es de 14.53. Del mismo modo con relación a las asignaturas del área de humanidades, tenemos que la nota promedio

de Metodología del Trabajo Intelectual es de 14.95, de Lenguaje y Comunicación I es de 14.57, de Lenguaje y Comunicación II es de 14.25, de Filosofía es de 15.28 y en Antropología Filosófica es de 15.27. Asimismo, como se puede ver en la Fig. 16, que la gran mayoría de los egresados obtuvieron un calificativo promedio de 14, con excepción en Lengua y comunicación II donde la nota fue 15 y en Filosofía 16

TABLA VII. RESUMEN DESCRIPTIVO DE LAS VARIABLES NUMÉRICAS

FEATURE	SIZE	MIN	MAX	MEAN	MEDIAN	MODE	STD	CANT_MISSING
EDAD_INGRESO	1982	15	41	18.13	18	17	2.11	0
I_MAT_BAS	1982	5	20	13.93	14	14	2.84	0
I_MAT_BAS_VECES	1982	1	6	1.35	1	1	0.57	0
I_FIS_I	1982	5	20	13.91	14	14	2.77	0
I_FIS_I_VECES	1982	1	6	1.35	1	1	0.57	0
I_MTI	1982	10	20	14.95	15	14	1.94	0
I_MTI_VECES	1982	1	4	1.15	1	1	0.37	0
I LENG_I	1982	5	20	14.57	15	14	2.17	0
I LENG_I_VECES	1982	1	5	1.24	1	1	0.46	0
II_CALC_I	1982	5	20	14.69	15	14	2.36	0
II_CALC_I_VECES	1982	1	7	1.17	1	1	0.57	0
II LENG_II	1982	5	20	14.25	15	15	2.36	0
II LENG_II_VECES	1982	1	4	1.30	1	1	0.49	0
II_FIS_II	1982	5	20	14.47	14	14	2.35	0
II_FIS_II_VECES	1982	1	6	1.14	1	1	0.47	0
III_CALC_II	1982	5	20	14.36	14	14	2.51	0
III_CALC_II_VECES	1982	1	6	1.24	1	1	0.62	0
III_FIS_III	1982	5	20	14.53	15	14	2.40	0
III_FIS_III_VECES	1982	1	16	1.15	1	1	0.55	0
III_FIL	1982	5	20	15.28	16	16	2.22	0
III_FIL_VECES	1982	1	4	1.13	1	1	0.38	0
IV_CALC_III	1982	5	20	13.99	14	14	2.82	0
IV_CALC_III_VECES	1982	1	6	1.34	1	1	0.57	0

TABLA VII. RESUMEN DESCRIPTIVO DE LAS VARIABLES NUMÉRICAS

FEATURE	SIZE	MIN	MAX	MEAN	MEDIAN	MODE	STD	CANT_MISSING
IV_ANT_FIL	1982	7	20	15.27	15	14	2.00	0
IV_ANT_FIL_VECES	1982	1	5	1.10	1	1	0.37	0

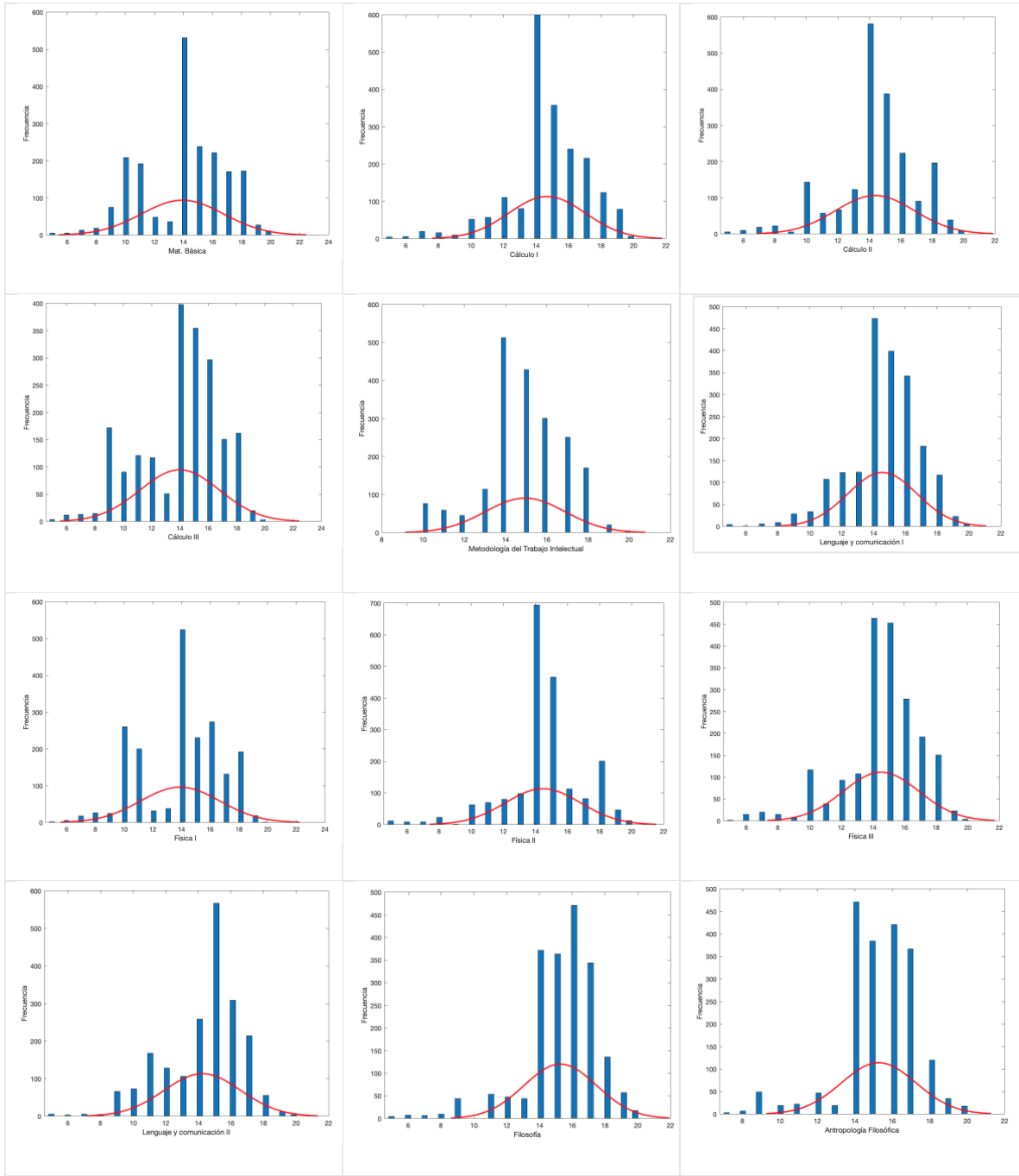


Fig. 16 Diagrama de distribución de frecuencias de las notas obtenidas en todas las asignaturas comunes

Como se puede ver en la Fig. 17, se realizaron diagramas boxplot para poder visualizar y comparar la distribución y la tendencia central de las calificaciones obtenidas en todas las asignaturas comunes. Si bien no se habían encontrado datos faltantes, aparentemente si existen outliers, sin embargo, cabe mencionar que la universidad trabaja en una escala vigesimal (De 0 a 20) y que estos datos son lo que realmente obtuvieron los estudiantes y se encuentran registrados en la base de datos institucional.

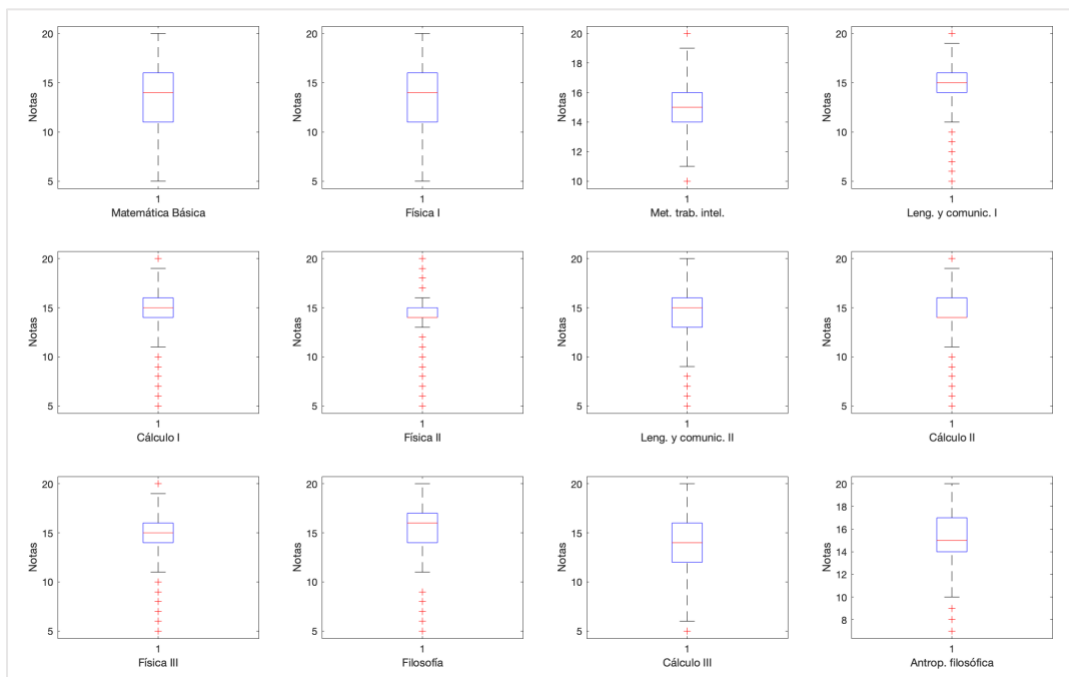


Fig. 17 Diagrama de distribución de las calificaciones obtenidas en las asignaturas comunes con respecto a la media

Luego, haciendo uso de Matlab se analizó la dispersión entre los datos por medio de diagramas de dispersión y calculando el coeficiente de correlación. Este análisis se hizo evaluando cada asignatura con su pre-requisito en base a la Fig. 4 y la clasificación por áreas de conocimiento a la que pertenecen éstas (Tabla IV).

En la Fig. 18, se puede ver que existe una relación positiva o directa entre las asignaturas de matemáticas, siendo más fuerte la relación entre la asignatura de Matemática Básica de primer ciclo con la asignatura de Cálculo I de segundo ciclo.

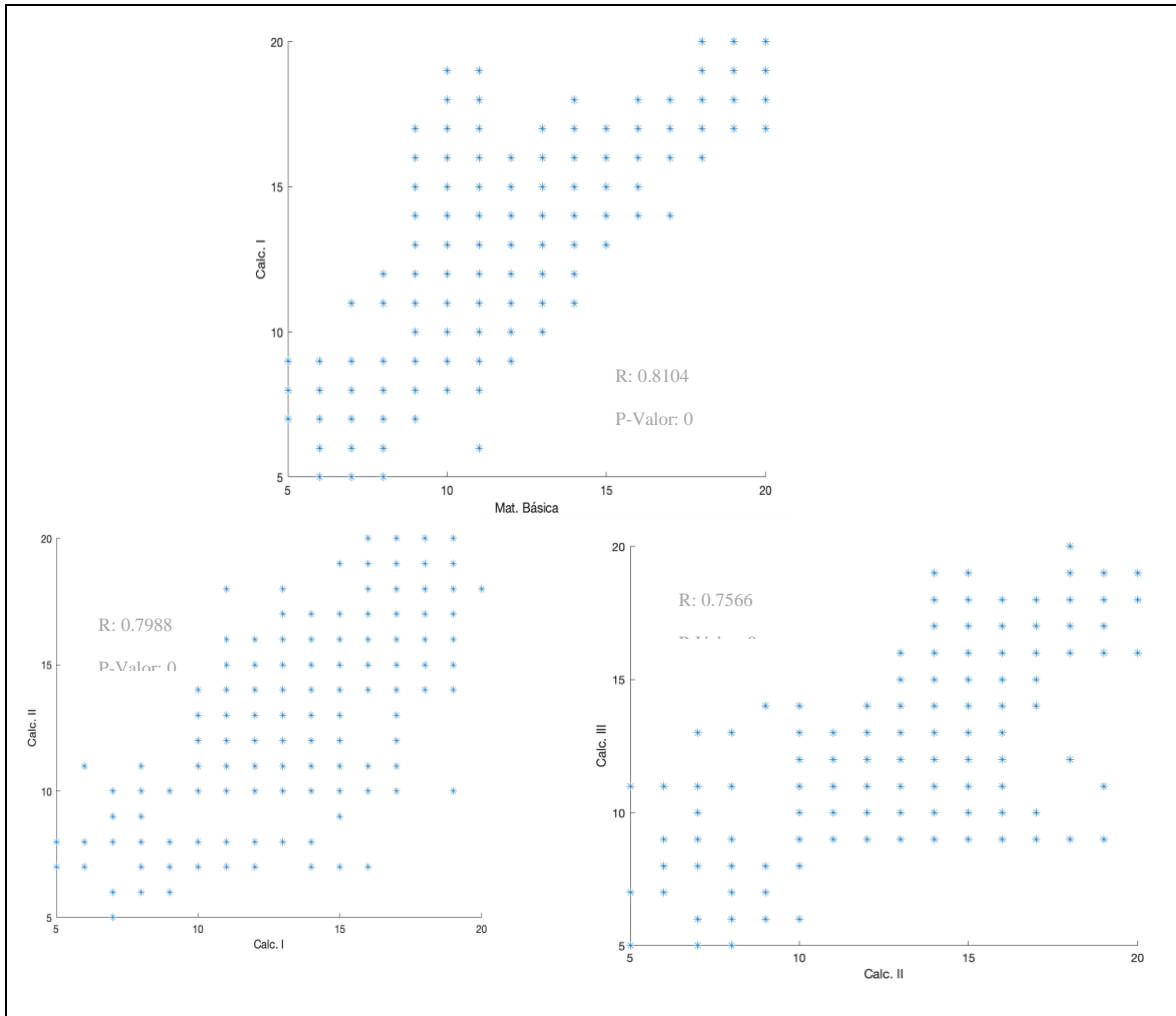


Fig. 18 Relación entre calificaciones de asignaturas del área de matemática

Con respecto a las asignaturas del área de física, se puede ver en la Fig. 19 que existe una relación positiva entre ellas, siendo más fuerte la relación entre la asignatura de Física II de segundo ciclo con la asignatura de Física III de tercer ciclo.

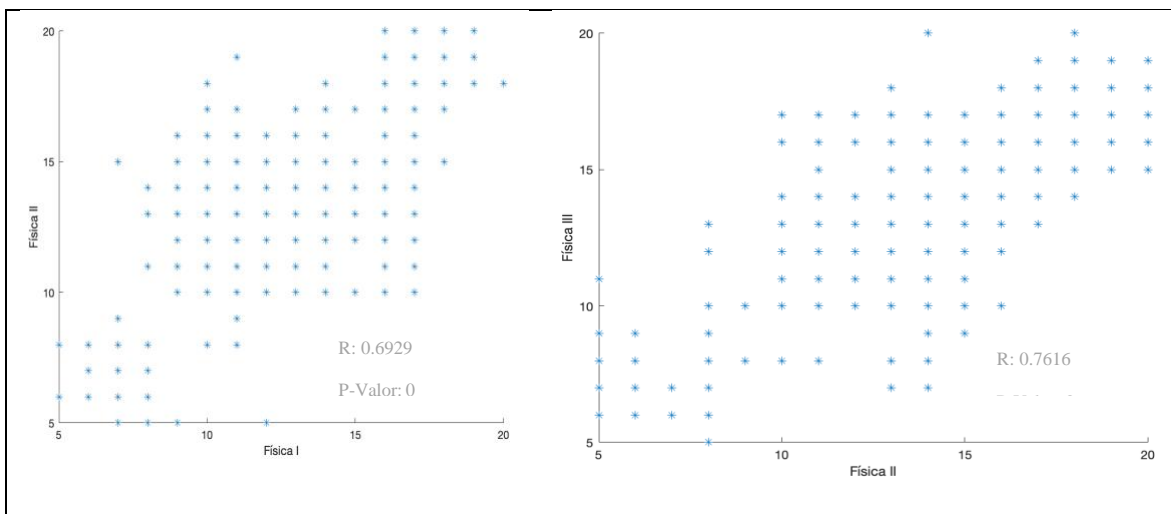


Fig. 19 Relación entre calificaciones de asignaturas del área de Física

Sin embargo, con respecto al área de humanidades, podemos ver en la Fig. 20 que existen relaciones más débiles entre las asignaturas de esta área con respecto a las observadas en las otras. Siendo más débil la relación entre Lenguaje y Comunicación II de segundo ciclo con Filosofía de tercer ciclo.

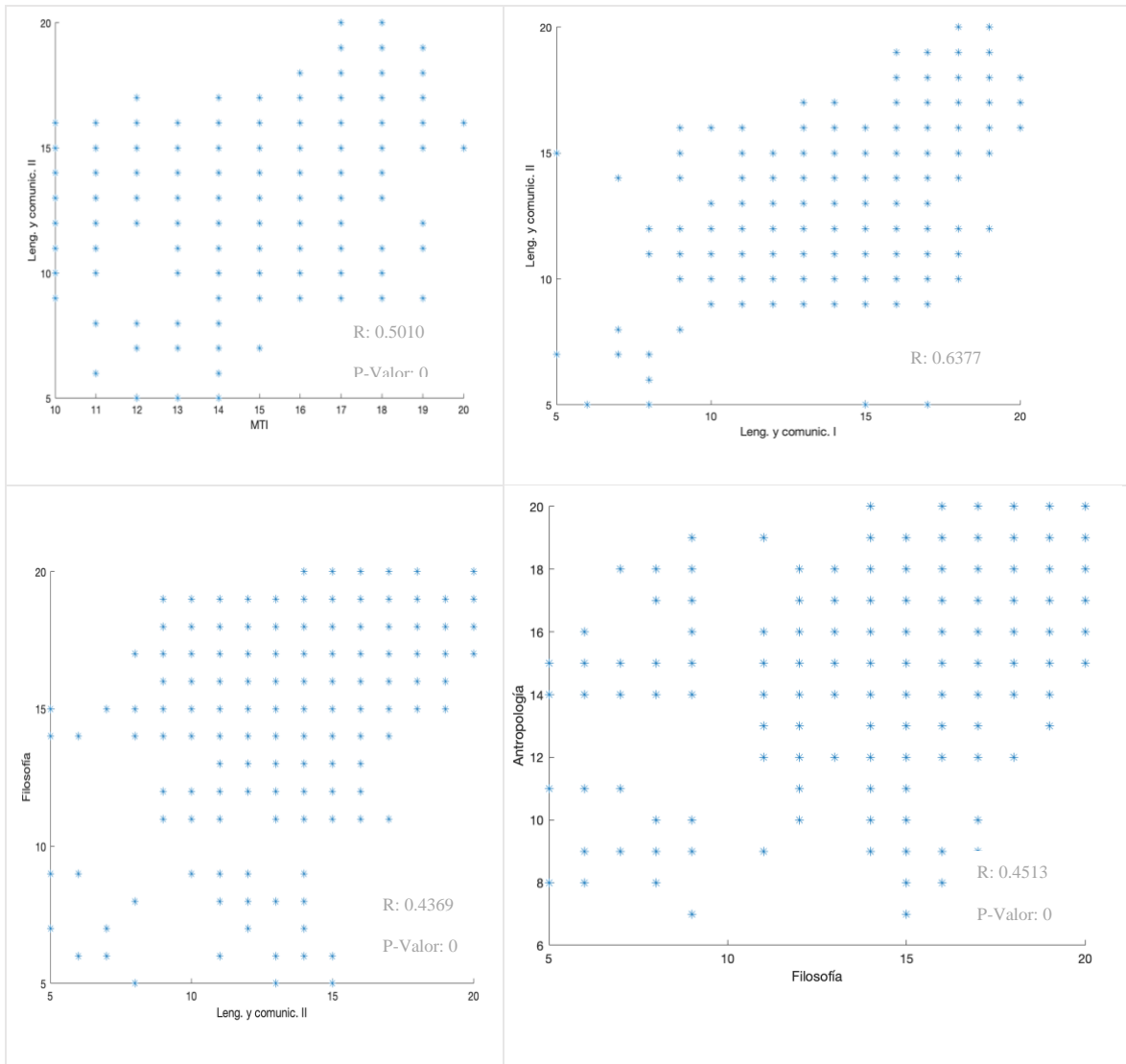


Fig. 20 Relación entre calificaciones de asignaturas del área de humanidades

Preparación de datos

En lo que respecta a la preparación de los datos, no se encontraron valores faltantes (NaN), esto debido a que los datos recolectados, son datos obligatorios para los sistemas de información de la universidad. Para el caso de los outliers (Tabla VIII) fueron tratados haciendo uso de filloutliers de MatLab. Asimismo, con respecto a las variables categóricas, fueron codificadas como se muestra en la tabla Tabla IX

TABLA VIII. VALORES OUTLIERS POR CADA ATRIBUTO NUMÉRICO

Característica	Media	STD	Cant. Outliers	Val Outliers
EDAD_INGRESO	18.13	2.11	36	25, 26, 27,28, 29, 3,31,38, 39, 41
I_MAT_BAS	13.93	2.84	6	5
I_MAT_BAS_VECES	1.35	0.57	13	4, 5, 6
I_FIS_I	13.91	2.77	2	5
I_FIS_I_VECES	1.35	0.57	15	4, 5, 6
I LENG_I	14.57	2.17	21	5, 6, 8
I LENG_I_VECES	1.24	0.46	22	3, 4, 5
I_MTI	14.95	1.94	0	
I_MTI_VECES	1.15	0.37	4	3, 4
II_CALC_I	14.69	2.36	31	5, 6, 7
II_CALC_I_VECES	1.17	0.57	62	3,4,5,6,7
II_FIS_II	14.47	2.35	30	5,6,7
II_FIS_II_VECES	1.14	0.47	70	3,4,6
II LENG_II	14.25	2.36	15	5,6,7
II LENG_II_VECES	1.30	0.49	23	3,4
III_CALC_II	14.36	2.51	16	5,6
III_CALC_II_VECES	1.24	0.62	29	4,5,6
III_FIL	15.28	2.22	30	5,6,7,8
III_FIL_VECES	1.13	0.38	31	3,4
III_FIS_III	14.53	2.40	37	5,6,7
III_FIS_III_VECES	1.15	0.55	46	3,4,6,16
IV_ANT_FIL	15.27	2.00	59	7,8,9
IV_ANT_FIL_VECES	1.10	0.37	27	3,4,5
IV_CALC_III	13.99	2.82	4	5
IV_CALC_III_VECES	1.34	0.57	14	4,5,6

TABLA IX. CODIFICACIÓN DE VARIABLES CATEGÓRICAS

Variable	Dato	Código
SEXO	Masculino	1
	Femenino	0
TIPO_COLEGIO	Público	1
	Privado	0
UBICACION_COL	Colegio donde terminó secundaria está ubicado en lambayeque	1
	Colegio donde terminó secundaria ubicado en otro dpto distinto a lambayeque	0
MODALIDAD_INGRESO	Examen de admisión	1
	Test dhac	
	Preferente	
	Escuela pre universitaria	
	Primeros puestos	0
	Traslado externo	
	Graduados y titulados	
	Beca 18	
	Deportista destacado	
	Personas con discapacidad	

Selección de características

Para realizar la selección de características, se combinaron métodos de filtro en base a la relación de las variables predictoras con la variable de salida y algoritmos de aprendizaje automático. Para esto, se obtuvo el coeficiente de correlación de cada variable predictora, luego se probó con el algoritmo de máxima relevancia, mínima redundancia (MRMR)[71]. Los resultados se pueden ver en la Tabla X. Ambos métodos coinciden que las siguientes variables predictoras no son muy relevantes: sexo, tipo de colegio donde estudió, tipo de semestre en el que ingresó, edad de ingreso y modalidad de ingreso. Además, las notas de los

cursos de las asignaturas de matemáticas y física tienen una relación más fuerte que las notas obtenidas en los cursos de humanidades.

TABLA X. ORDEN DE RELEVANCIA DEL PREDICTOR PARA EL MODELO EN BASE A MRMR VS. COEFICIENTE DE CORRELACIÓN

Orden de relevancia del predictor (MRMR)	Característica	Coefficiente de correlación
1	I_FIS_I	0.9128
2	III_FIS_III_VECES	-0.3818
3	I_MTI	0.5560
4	I_MAT_BAS_VECES	-0.6791
5	UBICACION_COL	0.0133
6	IV_ANT_FIL	0.4474
7	II_FIS_II	0.8172
8	IV_CALC_III_VECES	-0.6995
9	I_MAT_BAS	0.9136
10	II LENG_II_VECES	-0.5878
11	II_CALC_I_VECES	-0.4283
12	I LENG_I	0.6585
13	I_FIS_I_VECES	-0.7294
14	IV_CALC_III	0.8877
15	TIPO_SEMESTRE_INGRESO	0.1116
16	III_CALC_II	0.8741
17	III_FIL	0.5531
18	II_CALC_I	0.8846
19	III_FIS_III	0.8855
20	II LENG_II	0.7363
21	III_CALC_II_VECES	-0.5374
22	SEXO	-0.1048
23	I LENG_I_VECES	0.6585
24	II_FIS_II_VECES	-0.5015

25	IV_ANT_FIL_VECES	-0.2033
26	I_MTI_VECES	-0.3411
27	EDAD_INGRESO	-0.0037
28	III_FIL_VECES	-0.3296
29	MODALIDAD_INGRESO	-0.0679
30	TIPO_COLEGIO	-0.0375

Para llegar finalmente a seleccionar las características que se muestran en la Tabla XI, se hizo uso de la App de Clasification Learner de Matlab, donde se evaluaron aún más los atributos, llegando a determinar que las características relacionadas con las veces que llevó una asignatura ya no eran necesarias incluirlas pues no aportaba mayor precisión al modelo. Esto se debe en gran parte a que la nota que se ha recolectado para cada asignatura es la nota promedio que se obtiene de todas las veces que haya llevado la misma, por lo que en cierta forma la nota ya tiene incluido la ocurrencia de las veces. Asimismo, se hizo necesario incluir asignaturas hasta el IV ciclo para poder mejorar la precisión del modelo, por lo que podemos decir que mínimamente a partir de IV ciclo podemos predecir el nivel de logro de las competencias del perfil de egreso para un estudiante de un programa de ingeniería y además haya aprobado las asignaturas o alguna equivalente en su plan de estudios de Matemática Básica, Física I, Cálculo I, Física II, Cálculo II, Física III y Cálculo III

TABLA XI. CARACTERÍSTICAS SELECCIONADAS

Característica	Semestre
Seleccionada	
I_MAT_BAS	I
I_FIS_I	I
II_CALC_I	II
II_FIS_II	II

III_CALC_II	III
III_FIS_III	III
IV_CALC_III	IV

4.3. DESARROLLO Y PRUEBAS DEL MODELO PREDICTIVO

En esta etapa, se separaron los datos en 2 conjuntos, uno para el entrenamiento y otro para las pruebas del modelo. De los 1982 egresados de ingeniería, se tomó como datos de prueba a todos los egresados del 2020 y 2021 que representan el 25% (494 egresados) y se entrenó el modelo con el 75% restante (1488 egresados). Luego de esto, haciendo uso de MatLab y los datos de entrenamiento en base a las características seleccionadas, se evaluaron los diferentes algoritmos de clasificación de Machine Learning (Fig. 2) para determinar el algoritmo que nos da la mayor precisión. El método de validación utilizado fue de Cross-Validation con 5 folds.

Selección del algoritmo con mayor precisión

En la Tabla XII se muestran los resultados de precisión (Accuracy) obtenidos luego evaluar los diferentes algoritmos de clasificación de Machine Learning utilizando los datos de entrenamiento, las características seleccionadas y la aplicación de Classification Learner de Matlab como herramienta.

TABLA XII. EVALUACIÓN DE ALGORITMOS DE CLASIFICACIÓN DE MACHINE LEARNING

Algoritmo	Precisión (Accuracy)
Support Vector Machines (Medium Gaussian SVM)	96.7%
Nearest Neighbor Classifiers (KNN)	85.6%
Neural Network Classifiers	92.6%
Discriminant Analysis	95.7%
Decision Trees	95.1%
Naive Bayes Classifiers	95.8%
Ensemble Classifiers	95.4%

Creación y validación del modelo

Para crear el modelo se utilizaron datos de 1488 egresados de las diferentes escuelas de ingeniería (Civil, Industrial, Mecánica y Sistemas) que representan el 75% de todos los datos recolectados. De estos 1488 egresados, se utilizaron los datos en base a las características seleccionadas (Tabla XI) y además, se utilizó el algoritmo de Support Vector Machines (Medium Gaussian SVM) que es el que nos había dado la mayor precisión en la evaluación realizada (Tabla XII). Para observar el comportamiento del modelo con los datos de validación que ha usado durante su entrenamiento se elaboró la matriz de confusión como se puede ver en la Fig. 21

En la Fig. 21, la matriz de confusión muestra que nuestro modelo tuvo algunos pocos problemas para distinguir entre el nivel de logro insuficiente con el no logrado (4 casos),

entre el notable y suficiente (14 casos), para distinguir entre sobresaliente y notable (6 casos), entre suficiente y sobresaliente (6 casos), entre suficiente y notable (2 casos). Sin embargo, se confundió más cuando predijo 17 casos como nivel suficiente cuando en realidad eran insuficientes.

True Class	INSUFICIENTE	403				17
	NO LOGRADO	4	28			
	NOTABLE			295	6	2
	SOBRESALIENTE				162	6
	SUFICIENTE			14		551
		INSUFICIENTE	NO LOGRADO	NOTABLE	SOBRESALIENTE	SUFICIENTE
		Predicted Class				

Fig. 21 Matriz de confusión para data de validación al entrenar el modelo

Nuestro modelo trabaja con Support Vector Machines (Medium Gaussian SVM), SVM es un algoritmo de aprendizaje automático supervisado que busca encontrar un hiperplano que separe de la mejor forma posible dos clases diferentes de puntos de datos; es decir con el margen más amplio entre las dos clases. Este margen viene a ser el ancho máximo de la región que esta paralela al hiperplano que no tiene puntos de datos en su parte interior (Fig. 22).

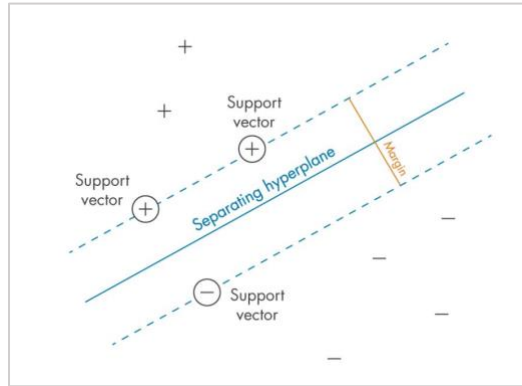


Fig. 22 Esquema del algoritmo SVM [75]

El modelo creado internamente construye varios clasificadores binarios y utiliza una función Kernel para transformar las características. El modelo trabaja con el tipo Medium Gaussian SVM y la función kernel viene definida por la fórmula de la Fig. 23, donde σ representa la anchura del kernel.:

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right)$$

Fig. 23 Función de base radial (RBF) o gaussiana

Fuente: Mathworks [75]

Los parámetros de configuración del modelo se muestran en la función creada en Matlab para entrenar el modelo (Fig. 24).

```

function [trainedClassifier, validationAccuracy] = trainClassifier(trainingData)

% Extract predictors and response
inputTable = trainingData;
predictorNames = {'I_MAT_BAS', 'I_FIS_I', 'II_CALC_I', 'II_FIS_II', 'III_CALC_II', 'III_FIS_III', 'IV_CALC_III'};
predictors = inputTable(:, predictorNames);
response = inputTable.NIVEL_LOGRO_PE; % Variable de salida

% Train a classifier
% This code specifies all the classifier options and trains the classifier.
template = templateSVM(...
    'KernelFunction', 'gaussian', ...
    'PolynomialOrder', [], ...
    'KernelScale', 2.6, ...
    'BoxConstraint', 1, ...
    'Standardize', true);
classificationSVM = fitcecoc(...
    predictors, ...
    response, ...
    'Learners', template, ...
    'Coding', 'onesone', ...
    'ClassNames', categorical({'INSUFICIENTE'; 'NO LOGRADO'; 'NOTABLE'; 'SOBRESALIENTE'; 'SUFICIENTE'}));

% Create the result struct with predict function
predictorExtractionFcn = @(t) t(:, predictorNames);
svmPredictFcn = @(x) predict(classificationSVM, x);
trainedClassifier.predictFcn = @(x) svmPredictFcn(predictorExtractionFcn(x));

% Add additional fields to the result struct
trainedClassifier.RequiredVariables = {'III_CALC_II', 'III_FIS_III', 'II_CALC_I', 'II_FIS_II', 'IV_CALC_III', 'I_FIS_I', 'I_MAT_BAS'};
trainedClassifier.ClassificationSVM = classificationSVM;

% Perform cross-validation
partitionedModel = crossval(trainedClassifier.ClassificationSVM, 'Kfold', 5);

% Compute validation accuracy
validationAccuracy = 1 - kfoldLoss(partitionedModel, 'LossFun', 'ClassifError');

```

Fig. 24 Código de entrenamiento del Modelo Computacional

Fuente: Matlab Classification Learner R2021a

Pruebas del modelo en los diferentes programas de ingeniería

En esta etapa, utilizando los datos de pruebas de 494 egresados, se probó el modelo entrenado con todos los egresados de cada escuela de la Facultad de Ingeniería en los años 2020 y 2021 (Tabla XIII).

TABLA XIII. CANTIDAD DE DATOS DE PRUEBA POR PROGRAMA Y AÑO DE EGRESO

Escuela	Año Egreso		Total
	2020	2021	
CIVIL	80	100	180
INDUSTRIAL	108	91	199
MECÁNICA	20	34	54
SISTEMAS	27	34	61
Total	235	259	494

Durante las pruebas del modelo, se obtuvieron los resultados que se muestran en la Tabla XIV y en las matrices de confusión de la Fig. 25, Fig. 26, Fig. 27 y Fig. 28. En todos los casos se obtuvo una precisión (Accuracy) alta que varía entre 92.2% y 96.7% comprobando de esta manera la exactitud del modelo creado para predecir el nivel del logro del perfil de egreso de un estudiante de ingeniería que haya cursado mínimamente el 4to ciclo y además tenga aprobadas todas sus asignaturas de matemáticas y física de su plan de estudio equivalentes a las mostradas en la Tabla IV.

TABLA XIV PRECISIÓN (ACCURACY) DEL MODELO DURANTE LAS PRUEBAS

Escuela	Precisión (Accuracy) con datos nuevos
CIVIL	92.2%
INDUSTRIAL	97%
MECANICA	96.3%
SISTEMAS	96.7%

En la Fig. 25, para el caso de los egresados 2020 y 2021 de la Escuela de Ingeniería Civil, la matriz de confusión muestra que nuestro modelo tiene algunos pocos problemas para distinguir entre el nivel de logro insuficiente con el notable (4 casos) y con no logrado (2 casos), para distinguir entre notable y suficiente (3 casos) o entre suficiente con insuficiente (5 casos).

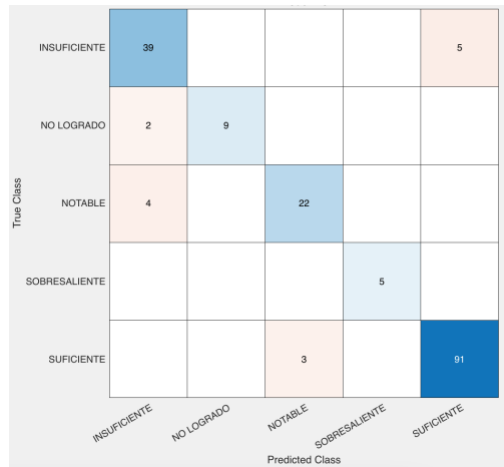


Fig. 25 Matriz de confusión durante las pruebas del modelo con el programa de Ingeniería Civil

En la Fig. 26, para el caso de los egresados 2020 y 2021 de la Escuela de Ingeniería Industrial, la matriz de confusión muestra que nuestro modelo tiene algunos pocos problemas para distinguir entre el nivel notable con el suficiente (3 casos) o entre suficiente con insuficiente (3 casos).

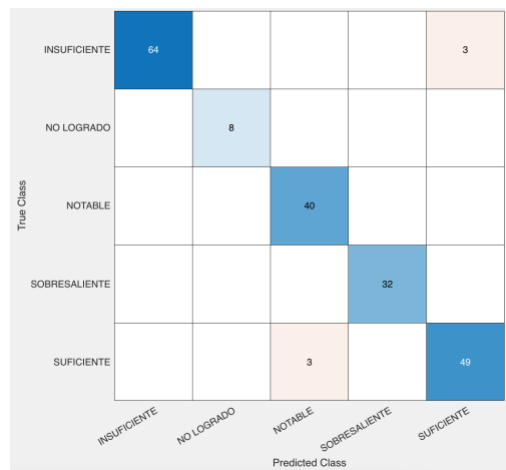


Fig. 26 Matriz de confusión durante las pruebas del modelo con el programa de Ingeniería Industrial

En la Fig. 27, para el caso de los egresados 2020 y 2021 de la Escuela de Ingeniería Mecánica, la matriz de confusión muestra que nuestro modelo tiene algunos pocos problemas para distinguir entre el nivel suficiente con insuficiente (2 casos).

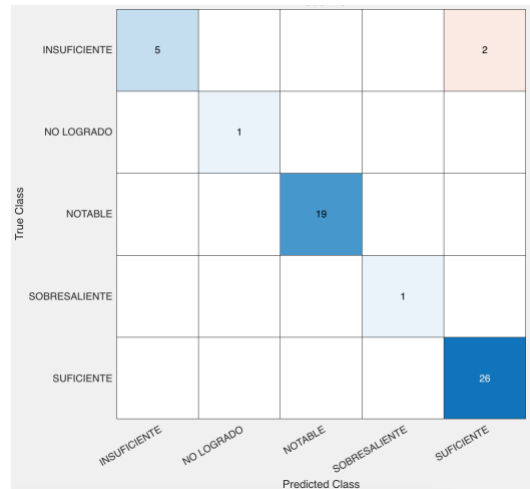


Fig. 27 Matriz de confusión durante las pruebas del modelo con el programa de Ingeniería Mecánica

En la Fig. 28, para el caso de los egresados 2020 y 2021 de la Escuela de Ingeniería de sistemas, la matriz de confusión muestra que nuestro modelo tiene algunos pocos problemas para distinguir entre el nivel suficiente con insuficiente (2 casos).

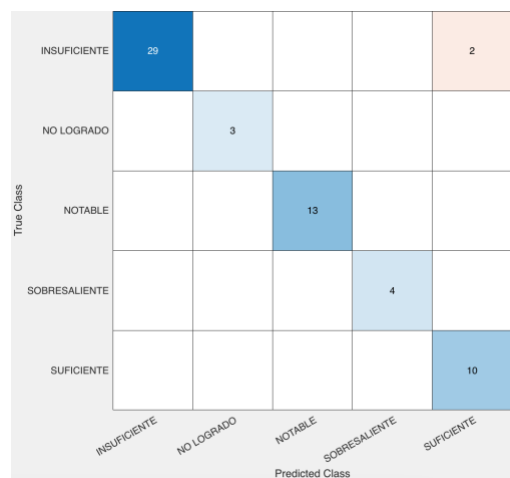


Fig. 28 Matriz de confusión durante las pruebas del modelo con el programa de Ingeniería de Sistemas

4.4. CREACIÓN DE LA APLICACIÓN PARA PREDECIR NIVEL DE LOGRO DEL PERFIL DE EGRESO

En la Fig. 29, se muestra la aplicación creada en Matlab, donde se puede ingresar el promedio de las notas de cada asignatura para un estudiante y al final predice el nivel de logro de perfil de egreso que tendría ese estudiante.

Calificación	Nota
I: Matematica Basica	7
I: Fisica I	8
II: Calculo I	5
II: Fisica II	8
III: Calculo II	8
III: Fisica III	7
IV: Calculo III	6

Nivel de logro del perfil de egreso: NO LOGRADO

Fig. 29 Aplicación para predecir el nivel de logro del perfil de egreso de un estudiante

Asimismo, en la Fig. 30, se muestra la aplicación creada en Matlab, donde permite importar un archivo de excel con las calificaciones de varios estudiantes y al final predice para cada uno de ellos el nivel de logro del perfil de egreso, alertando en rojo aquellos estudiantes que obtendrían un nivel no logrado o insuficiente.

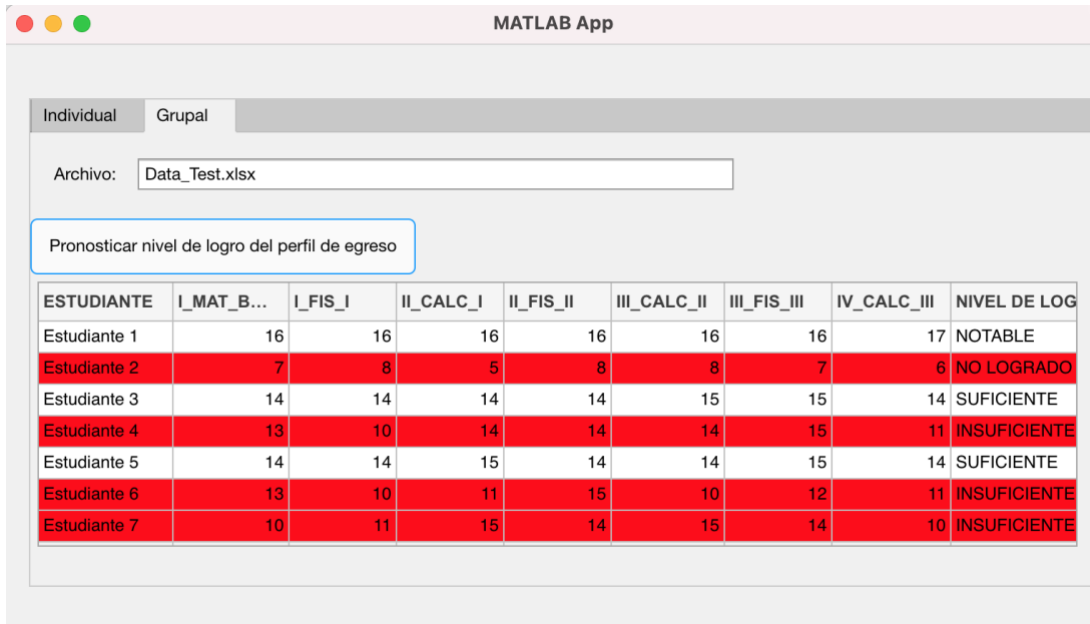


Fig. 30 Aplicación para predecir el nivel de logro del perfil de egreso de varios estudiantes

V. DISCUSIÓN

El perfil de egreso establecido en los planes curriculares de cada programa de estudio y que debe ser logrado por los estudiantes universitarios al finalizar su carrera, es un tema que ha sido ampliamente abordado en varias investigaciones [24][25][26][27][28], ya sea utilizando el termino de perfil de egreso o refiriendose a éste como resultados del estudiante (Students Outcomes) tal como lo establecen las agencias acreditadoras ABET e ICACIT [29][30][31][32][33][34][35]. Estos estudios, se centraron en proponer alternativas para mejorar el proceso de enseñanza aprendizaje, mejorar la evaluación final de los students outcomes o en determinar cuáles son los factores que influyen en el rendimiento académico de los estudiantes. En cambio, otros autores orientaron sus investigaciones al desarrollo de modelos de predicción en la educación [36][37][38], centrandose mayormente en la predicción del rendimiento académico en cursos, semestres o años de la educación presencial [39][40][41][42][43][44][45], predicción del rendimiento académico en cursos MOOC (Massive Open Online Course) [46][47][48][49][50][51][52][53][54], predicción de la deserción o abandono de los estudiantes [55][56], pronosticar el tiempo de retraso en la graduación [57][58], predicción que si se requiere de tutoría o de la motivación académica del estudiante [59][60] o la predicción para conseguir empleo por parte de los estudiantes [61]. Sin embargo, no se encontraron estudios que directamente hayan abordado el tema de alerta temprana o predicción del nivel de logro del perfil de egreso o Students Outcomes en los programas de ingeniería, niveles de logro que son medidos por los programas de estudios universitarios como parte de su proceso de mejora continua establecidos en sus modelos de calidad. Lo que si se encuentra mayormente son estudios que buscan pronosticar el

rendimiento académico final o el promedio final global del estudiante universitario [62][63][64][65][66][67][68][69]. Debido a esto, la presente investigación contribuye con un modelo basado en Machine Learning que permita predecir el nivel de logro del perfil de egreso de los estudiantes universitarios de las carreras de Ingeniería, de tal manera que se puedan tomar acciones correctivas de manera oportuna y no esperar hasta el final de los estudios para poder obtener este resultado, debido a que ya no habría mucho tiempo para poder proponer alguna mejora. En este trabajo, en primer lugar se determinó que las características del estudiante relacionadas con las calificaciones obtenidas en los cursos de Matemáticas (Matemática Básica, Cálculo I, Cálculo II y Cálculo III) y los cursos de Física (Física I, Física II y Física III) o algunos de sus equivalentes en otros planes de estudio, son los mejores predictores del nivel de logro del perfil de egreso para los estudiantes universitarios de los programas de ingeniería. Esto coincide con los resultados de [62] [66] y con un artículo publicado durante mis estudios del doctorado [69] relacionado con el tema, donde también se determina que las calificaciones de los estudiantes en las asignaturas son los principales predictores de un rendimiento académico futuro. En segundo lugar, se determinó que el algoritmo Support Vector Machine (SVM) es el algoritmo de Machine Learning que nos da la mayor precisión (accuracy): 96.7% a la hora de predecir el nivel de logro del perfil de egreso de los estudiantes universitarios de ingeniería tanto con los datos de entrenamiento como con datos nuevos (datos de prueba). Esto coincide con [69] y [47] que también obtuvieron mejor precisión al utilizar SVM, con 81.4% y 98.5% respectivamente, todo esto a diferencia de otras investigaciones que obtuvieron mejores resultados con árboles de decisión y redes neuronales.

VI. CONCLUSIONES

En esta investigación, haciendo uso de técnicas estadísticas y de machine learning, luego de recolectar datos de 1982 egresados de diferentes programas de Ingeniería de una universidad peruana, se llegó a determinar que las características del estudiante relacionadas con las calificaciones obtenidas en los cursos de Matemáticas (Matemática Básica, Cálculo I, Cálculo II y Cálculo III) y los cursos de Física (Física I, Física II y Física III) o algunos de sus equivalentes en otros planes de estudio, son los mejores predictores del nivel de logro del perfil de egreso para los estudiantes universitarios de los programas de ingeniería, a diferencia de los cursos de Humanidades (Metodología del Trabajo Intelectual, Lenguaje y Comunicación I, Lenguaje y Comunicación II, Filosofía y Antropología Filosófica) y otros atributos como la edad, sexo, tipo de colegio donde estudió, procedencia y modalidad de ingreso que no fueron necesarios incluir ya que no aportaban significativamente en la precisión del modelo.

Asimismo, se evaluaron los diferentes algoritmos de clasificación de machine learning: Support Vector Machines (SVM), Nearest Neighbor Classifiers (KNN), Neural Network Classifiers, Discriminant Analysis, Decision Trees, Naive Bayes Classifier y Ensemble Classifiers, llegando a la conclusión que el algoritmo SVM nos aportó la mayor precisión tanto en la validación durante el entrenamiento como en las pruebas a la hora de predecir el nivel de logro del perfil de egreso.

Además, haciendo uso de Matlab, teorías de machine learning y las calificaciones en ciertas asignaturas de los estudiantes de ingeniería, se llegó a crear un modelo computacional que

permita predecir anticipadamente el nivel de logro del perfil de egreso (No logrado, insuficiente, suficiente, notable o sobresaliente) de los estudiantes de los programas de ingeniería, de tal manera que se cuente con información que permita tomar acciones correctivas de manera oportuna y no esperar hasta el final de los estudios universitarios para poder recién obtener este resultado, debido a que ya no habría mucho tiempo para poder proponer alguna mejora. Este modelo obtuvo una precisión (accuracy) de 96.7% durante la etapa de entrenamiento y cuando fue sometido a datos nuevos de los estudiantes de Ingeniería Civil, Ingeniería Industrial, Ingeniería Mecánica e Ingeniería de Sistemas, que han egresado en los años 2020 y 2021 obtuvieron precisiones del 92% al 97% (Civil: 92.2%, Industrial: 97%, Mecánica: 96.3%, Sistemas: 96.7%).

Cuando se creó el modelo, se hizo necesario incluir las asignaturas de matemática y física que generalmente en los diferentes programas de ingeniería se desarrollan hasta cuarto ciclo, por lo que podemos concluir que mínimamente cuando un estudiante haya cursado el cuarto ciclo y ya tenga sus calificaciones completas de matemáticas y física será posible predecir anticipadamente su nivel de logro del perfil de egreso.

Finalmente todo esto, contribuyó a cubrir algunos vacíos que se encontraron producto de la revisión de la literatura , donde no se ha abordado directamente el tema de alerta temprana o predicción del nivel de logro del perfil de egreso de ingeniería o Students Outcomes como los establece las agencias acreditadoras de ABET o ICACIT. Las investigaciones se centran mayormente en la predicción del rendimiento académico de cursos (presenciales / online), en la predicción de la deserción estudiantil o en la predicción de la calificación final del estudiante.

VII. REFERENCIAS

- [1] SINEACE, “Modelo de Acreditación para Programas de Estudios de Educación Superior Universitaria,” 2016. Accessed: Sep. 27, 2018. [Online]. Available: www.sineace.gob.peContenido
- [2] ABET, “Engineering Accreditation Commission CRITERIA FOR ACCREDITING ENGINEERING PROGRAMS,” 2017. Accessed: Sep. 27, 2018. [Online]. Available: <http://www.abet.org/wp-content/uploads/2018/02/E001-18-19-EAC-Criteria-11-29-17.pdf>
- [3] ICACIT, “Criterios de Acreditación Programas de Ingeniería. Ciclo de Acreditación 2019,” 2018. Accessed: Sep. 28, 2018. [Online]. Available: www.icacit.org.pe
- [4] OECD, “AHELO Main Study - OECD,” 2015. <http://www.oecd.org/education/skills-beyond-school/ahelo-main-study.htm>
- [5] L. Praslova, “Adaptation of Kirkpatrick’s four level model of training criteria to assessment of learning outcomes and program evaluation in Higher Education,” *Educational Assessment, Evaluation and Accountability*, vol. 22, no. 3, pp. 215–225, 2010, doi: 10.1007/s11092-010-9098-7.
- [6] D. Nusche, “Assessment of learning outcomes in higher education,” *OECD Education Working Papers, No. 15*, no. 15, 2008, doi: <http://dx.doi.org/10.1787/244257272573>.
- [7] C. Coggi, “Theoretical models and evaluation tools of learning outcomes in Higher Education,” *Revista Italiana de Investigación Educativa*, vol. 16, p. 14, 2016, [Online]. Available: <http://ojs.pensamultimedia.it/index.php/sird/article/view/1773/1705>

- [8] E. ; Cano and M. Fernández, “Competencias de los egresados del Espacio Europeo de Educación Superior: relatos de vida de los nuevos estudiantes universitarios frente a los antiguos licenciados,” *REDU. Revista de Docencia Universitaria*, vol. 14, no. 2, p. 187, 2016, doi: 10.4995/redu.2016.5933.
- [9] A. Anaya, “Reflexiones sobre el logro del perfil y atributos requeridos del ingeniero químico egresado de las instituciones educativas,” *Educación química*, vol. 20, no. 1, pp. 70–74, 2009, doi: 10.1016/S0187-893X(18)30009-0.
- [10] PayScale y Future Workplace, “PayScale y Future Workplace Release 2016 Workforce-Skills Informe de preparación,” 2016.
<https://www.payscale.com/about/press-releases/payscale-and-future-workplace-release-2016-workforce-skills-preparedness-report>
- [11] N. Cabrera, M. López, and M. Portillo, “Las competencias de los graduados y su evaluación desde la perspectiva de los empleadores,” *Estudios Pedagógicos*, vol. 3, pp. 69–87, 2016, doi: 10.4067/S0718-07052016000400004.
- [12] C. Agencia para la Calidad del Sistema Universitario, “Empleabilidad y Competencias de los Recién Graduados: La Opinión de Empresas e Instituciones.Principales resultados del estudio de empleadores 2014.” 2015.
http://www.aqu.cat/doc/doc_25023446_1.pdf
- [13] Wendy Cunningham, “Muchas universidades no están conectadas con el mundo laboral,” 2015.
<http://www.bancomundial.org/es/news/feature/2015/02/21/universidades-latinoamericanas-desconectadas-realidad-laboral> (accessed Oct. 06, 2018).

- [14] CIDAC México, “Encuesta de Competencias Profesionales 2014.¿Qué buscan y no encuentran las empresas en los profesionistas jóvenes?,” 2014. Accessed: Sep. 28, 2018. [Online]. Available: www.cidac.org
- [15] CEPLAN, “Las universidades no evalúan calidad de sus egresados,” 2015. <https://elcomercio.pe/economia/ejecutivos/ceplan-universidades-evaluan-calidad-egresados-370938> (accessed Oct. 07, 2018).
- [16] ManpowerGroup, “Encuesta de Escasez de Talento. Resultados Perú 2016/2017,” 2016. Accessed: Oct. 02, 2018. [Online]. Available: <https://www.manpower.com.pe/mpintranet/publicaciones/4466-9358284738328.pdf>
- [17] MBA & Educación Ejecutiva - América Economía, “Perú y el desafío de acortar la brecha entre oferta y demanda laboral,” 2015. <https://mba.americaeconomia.com/articulos/reportajes/peru-y-el-desafio-de-acortar-la-brecha-entre-oferta-y-demanda-laboral> (accessed Sep. 28, 2018).
- [18] UNIVERSIA PERU, “Perú tiene un 40% de profesionales universitarios subempleados,” 2015. <http://noticias.universia.edu.pe/empleo/noticia/2015/01/14/1118197/peru-40-profesionales-universitarios-subempleados.html> (accessed Oct. 04, 2018).
- [19] E. Kramers, “Mejorando la brecha entre habilidades y empleo en ingeniería | Randstad,” 2017. <https://www.randstad.es/tendencias360/mejorando-la-brecha-entre-habilidades-y-empleo-en-ingenieria/> (accessed Nov. 13, 2018).
- [20] F. F. Esparza Paz, R. Sánchez-Chávez, S. Esparza-Zapata, E. Esparza-Zapata, and Á. Villacrés-Lara, “Factores de rendimiento académico en estudiantes universitarios, componentes de calidad de la educación superior. Estudio de caso Facultad de Administración de Empresas, Escuela Superior Politécnica de Chimborazo,”

Innovaciones Educativas, vol. 22, no. 33, pp. 46–61, Dec. 2020, doi: 10.22458/ie.v22i33.2893.

- [21] J. A. Gutiérrez-Monsalve, J. Garzón², and A. M. Segura-Cardona, “Factors associated to academic performance in university students,” *Formacion Universitaria*, vol. 14, no. 1, pp. 13–24, 2021, doi: 10.4067/S0718-50062021000100013.
- [22] B. Chair *et al.*, “Artificial intelligence and life in 2030, estudio de cien Años sobre Inteligencia Artificial: Informe del Panel de Estudio 2015-2016,” 2016, doi: <https://ai100.stanford.edu>.
- [23] M. Tord, “La Inteligencia Artificial en la Educación Superior,” 2018. <https://medium.com/8wires/la-inteligencia-artificial-en-la-educaci%C3%B3n-superior-ba067c5969f5> (accessed Oct. 15, 2018).
- [24] M. Poblete and a Villa, “SEBSCO, una experiencia alternativa para evaluar competencias,” *Aula Abierta*, vol. 39, p. 16, 2011, [Online]. Available: http://www.izenpe.com/s15-4812/es/contenidos/informacion/dsi_rr_hh/es_rr_hh/adjuntos/SEBSCO, una experiencia alternativa para evaluar competencias.pdf
- [25] E. Hensel and R. Robinson, “Direct Assessment of Student Learning Outcomes by Analysis of Performance Evaluation of Student Employees,” pp. 1–10, 2014, [Online]. Available: <http://www.laccei.org/LACCEI2014-Guayaquil/RefereedPapers/RP193.pdf>
- [26] M. Martín, “Un Modelo De Medida De Competencias En La Universidad: La Evaluación De Los Resultados De Aprendizaje,” Universidad Rey Juan Carlos, 2015. [Online]. Available: <https://ciencia.urjc.es/handle/10115/13562>

- [27] J. Estrada, “Propuesta Metodológica Para La Evaluación De Resultados De Aprendizaje Con Enfoque De Competencias Transdisciplinares,” 2016, Accessed: Oct. 10, 2018. [Online]. Available: www.grupocieg.org
- [28] I. D. Cruz, “Sistema de Información SICPE para la evaluación continua del logro del perfil del egresado de la carrera de Ingeniería de Sistemas,” Universidad Peruana Unión, 2017. Accessed: Oct. 09, 2018. [Online]. Available: http://repositorio.upeu.edu.pe/bitstream/handle/UPEU/975/Ian_tesis_Bachiller_2017.pdf?sequence=6
- [29] L. Shuman, M. Besterfield-Sacre, and J. MCGourty, “The ABET ‘Professional Skills’- Can They Be Taught? Can they Be Assessed?,” *Journal of Engineering*, no. January, 2005, doi: 10.1002/j.2168-9830.2005.tb00828.x.
- [30] K. Yue, “Effective course-based learning outcome assessment for ABET accreditation of computing programs,” *Journal of Computing Sciences in Colleges*, vol. 22, no. 4, pp. 252–259, 2007, doi: 10.1109/ICSENS.2011.6127228.
- [31] K. ; Bassett and S. Duan, “Integration of Course Learning Outcomes with Program Assessment for ABET Accreditation,” pp. 1–8, 2012, [Online]. Available: <https://proceedings.asmedigitalcollection.asme.org/proceeding.aspx?articleid=1750951>
- [32] E. Essa, A. Dittrich, S. Dascalu, and F. C. Harris, “ACAT: A web-based software tool to facilitate course assessment for ABET accreditation,” *ITNG2010 - 7th International Conference on Information Technology: New Generations*, pp. 88–93, 2010, doi: 10.1109/ITNG.2010.224.
- [33] M. H. Imam and I. A. Tasadduq, “Evaluating the Satisfaction of ABET Student Outcomes from Course Learning Outcomes through a Software Implementation,”

International Journal of Quality Assurance in Engineering and Technology Education, vol. 2, no. 3, pp. 21–33, 2012, doi: 10.4018/ijqaete.2012070102.

- [34] M. ; Imam, I. Tasadduq, A.-R. Ahmad, and F. Aldosari, “An Expert System for Assessment of Learning Outcomes for ABET Accreditation,” *International Journal of Educational and Pedagogical Sciences*, vol. 10, no. 1, p. 4, 2016, [Online]. Available: <https://waset.org/publications/10003532/an-expert-system-for-assessment-of-learning-outcomes-for-abet-accreditation->
- [35] I. Tasadduq, M. Imam, A. ; Ahmad, and F. Aldosari, “Obtaining ABET student outcome satisfaction from course learning outcome data using fuzzy logic,” *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 7, pp. 3069–3081, 2017, doi: 10.12973/eurasia.2017.00705a.
- [36] X. Hu, C. Cheong, W. Ding, and M. Woo, “A Systematic Review of Studies on Predicting Student Learning Outcomes Using Learning Analytics,” *ACM Digital Library*, vol. 72, pp. 414–422, 2016, doi: 10.1016/j.procs.2015.12.157.
- [37] C. A. del Río and J. A. Pineda, “Predicting academic performance in traditional environments at higher-education institutions using data mining: A review,” *Ecos de la Academia*, vol. 4, no. December, pp. 185–201, 2016, [Online]. Available: <https://www.researchgate.net/publication/311761954%0D>
- [38] E. Alyahyan and D. Düşteğör, “Predicting academic success in higher education: literature review and best practices,” *International Journal of Educational Technology in Higher Education*, vol. 17, no. 1. Springer, Dec. 01, 2020. doi: 10.1186/s41239-020-0177-7.
- [39] M. E. Alper and Z. Çataltepe, “Improving course success prediction using ABET course outcomes and grades,” *CSEDU 2012 - Proceedings of the 4th International*

Conference on Computer Supported Education, vol. 2, pp. 222–229, 2012, [Online].

Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0->

84864875945&partnerID=40&md5=fe58c02fd6b6f5fd1bc81e5cb2974c3d

- [40] B. Haskins, “Machine Learning as a Tool for Identifying Opportunities for Intervention in a University Module,” in *Pattern Recognition Association of South Africa and Robotics and Mechatronics*, 2017, vol. 36, no. 3, pp. 215–219. doi: 10.1109/RoboMech.2017.8261143.
- [41] B. Ndlovu and O. Ngala, “The Design of Predictive Model for the Academic Performance of Students at University Based on Machine Learning,” *J. of Electrical Engineering*, vol. 6, no. 4, pp. 229–237, 2018, doi: 10.17265/2328-2223/2018.04.006.
- [42] M. C. Nicoletti and M. G. M. Marques, “A Data Mining Approach for Forecasting Students’ Performance,” in *Iberian Conference on Information Systems and Technologies, CISTI*, 2018, pp. 1–7. doi: 10.23919/CISTI.2018.8399389.
- [43] D. Wham, “Forecasting student outcomes at university-wide scale using machine learning,” *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17*, pp. 576–577, 2017, doi: 10.1145/3027385.3029467.
- [44] I. Shingari and D. Kumar, “Predicting Student Performance Using Classification Data Mining Techniques,” vol. 6, no. 7, pp. 43–48, 2018, [Online]. Available: http://www.ijcseonline.org/pub_paper/7-IJCSE-04256.pdf
- [45] A. S. Hoffait and M. Schyns, “Early detection of university students with potential difficulties,” *Decision Support Systems*, vol. 101, pp. 1–11, 2017, doi: 10.1016/j.dss.2017.05.003.
- [46] N. Kondo, M. Okubo, and T. Hatanaka, “Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data,” *Proceedings - 2017 6th IIAI*

- International Congress on Advanced Applied Informatics, IIAI-AAI 2017*, pp. 198–201, 2017, doi: 10.1109/IIAI-AAI.2017.51.
- [47] R. Al-Shabandar, A. Hussain, A. Laws, R. Keight, J. Lunn, and N. Radi, “Machine learning approaches to predict learning outcomes in Massive open online courses,” *Proceedings of the International Joint Conference on Neural Networks*, vol. 2017-May, pp. 713–720, 2017, doi: 10.1109/IJCNN.2017.7965922.
- [48] R. Conijn and M. Zaanen, “Predicting student performance with Neural Networks,” no. May, 2017, [Online]. Available: <http://arno.uvt.nl/show.cgi?fid=143628>
- [49] R. L. U. Cazarez and C. L. Martin, “Neural networks for predicting student performance in online education,” *IEEE Latin America Transactions*, vol. 16, no. 7, pp. 2053–2060, 2018, doi: 10.1109/TLA.2018.8447376.
- [50] X. Ma, Y. Yang, and Z. Zhou, “Using Machine Learning Algorithm to Predict Student Pass Rates In Online Education,” *ACM Digital Library*, pp. 156–161, 2018, doi: 10.1145/3220162.3220188.
- [51] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, “Using machine learning to predict student difficulties from learning session data,” *Artificial Intelligence Review*, pp. 1–27, 2018, doi: 10.1007/s10462-018-9620-8.
- [52] B.-H. Kim, E. Vizitei, and V. Ganapathi, “GritNet: Student Performance Prediction with Deep Learning,” 2018, [Online]. Available: <http://arxiv.org/abs/1804.07405>
- [53] V. Nguyen, Q. Nguyen, and V. Nguyen, “A Model to Forecast Learning Outcomes for Students in Blended Learning Courses Based On Learning Analytics,” pp. 35–41, 2018, doi: 10.1145/3268808.3268827.
- [54] W. Chen, C. G. Brinton, D. Cao, A. Mason-singh, C. Lu, and M. Chiang, “Early Detection Prediction of Learning Outcomes in Online Short-Courses via Learning

- Behaviors,” *IEEE Transactions on Learning Technologies*, vol. 1382, no. c, pp. 1–14, 2018, doi: 10.1109/TLT.2018.2793193.
- [55] A. M. Pérez, C. R. Escobar, M. R. Toledo, L. B. Gutierrez, and G. Reyes, “Prediction model of first-year student desertion at Universidad Bernardo O’Higgins (UBO),” pp. 1–22, 2018, doi: 10.1590/S1678-4634201844172094.
- [56] J. Fernández, A. Rojas, G. Daza, D. Gómez, A. Alvarez, and Á. Orozco, “Student Desertion Prediction Using Kernel Relevance Analysis,” *Progress in Artificial Intelligence and Pattern Recognition. IWAIPR 2018. Lecture Notes in Computer Science*, vol. 11047, 2018, doi: 10.1007/978-3-030-01132-1.
- [57] H. Lakkaraju *et al.*, “A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes,” *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’15*, pp. 1909–1918, 2015, doi: 10.1145/2783258.2788620.
- [58] T. Ojha, “Prediction of Graduation Delay Based on Student Characteristics and Performance,” *Electrical and Computer Engineering ETDs Engineering*, 2017, [Online]. Available: https://digitalrepository.unm.edu/cgi/viewcontent.cgi?referer=https://www.google.es/&httpsredir=1&article=1363&context=ece_etds
- [59] M. U. Fahri and S. M. Isa, “Data Mining to Prediction Student Achievement based on Motivation, Learning and Emotional Intelligence in MAN 1 Ketapang,” no. June, pp. 53–60, 2018, doi: 10.5815/ijmecs.2018.06.07.
- [60] I. Đurđević, “Machine learning methods in predicting the student academic motivation,” *Croatian Operational Research Review*, vol. 8, no. 2, pp. 443–461, 2017, doi: 10.17535/crorr.2017.0028.

- [61] K. Sreenivasa, N. Swapna, and P. Praveen, "Educational data mining for student placement prediction using machine learning algorithms," *International Journal of Engineering & Technology*, vol. 7, no. 12, pp. 43–46, 2018, [Online]. Available: www.sciencepubco.com/index.php/IJET
- [62] K. Verma, A. Singh, and P. Verma, "A review on predicting student performance using data mining method," *MMW Fortschr Med*, vol. 3, no. 1, pp. 127–132, 2016, [Online]. Available: <http://queens.ezp1.qub.ac.uk/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=23270045>
<http://resolver.ebscohost.com/openurl?sid=OVID:medline&id=pmid:23270045&id=doi:&issn=1438-3276&isbn=&volume=154&issue=21&spage=5&d>
- [63] F. Yang and F. Li, "Study on student performance estimation, student progress analysis, and student potential prediction based on data mining," *Computers & Education*, pp. 97–108, 2018, doi: 10.1016/j.compedu.2018.04.006.
- [64] A. Kaur, N. Umesh, and B. Singh, "Machine Learning Approach to Predict Student Academic Performance," vol. 6, no. Iv, pp. 734–742, 2018, [Online]. Available: <https://www.ijraset.com/files/serve.php?FID=15594>
- [65] Y. ; Saheed, T. ; Oladele, A. ; Akanni, and W. Ibrahim, "Student Performance Prediction Based On Data Mining," vol. 37, no. 4, pp. 1087–1091, 2018, [Online]. Available: <http://nijotech.com/index.php/nijotech/article/view/1874/1359>
- [66] J. Xu, K. H. Moon, and M. van der Schaar, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs," *IEEE Journal on Selected Topics in Signal Processing*, vol. 11, no. 5, pp. 742–753, 2017, doi: 10.1109/JSTSP.2017.2692560.

- [67] A. Shanthini, G. Vinodhini, and R. M. Chandrasekaran, "Predicting students' academic performance in the University using meta decision tree classifiers," *Journal of Computer Science*, vol. 14, no. 5, pp. 654–662, 2018, doi: 10.3844/jcssp.2018.654.662.
- [68] M. P. G. Martins, V. L. Miguéis, and I. Tec, "Uma Metodologia de Data Mining para Prever o Desempenho de Estudantes de Licenciatura," 2018. doi: 10.23919 / CISTI.2018.8399175.
- [69] H. Zelada-Valdivieso, "Model based on Machine Learning to alert early on the academic performance that a Systems Engineering student would have upon graduation," *2021 IEEE 1st International Conference on Advanced Learning Technologies on Education & Research (ICALTER)*, 2021, doi: 10.1109/ICALTER54105.2021.9675117.
- [70] J. Gallardo, "Metodología para el desarrollo de proyectos en minería de datos CRISP-DM," pp. 1–12, 2017, [Online]. Available: http://www.oldemarrodriguez.com/yahoo_site_admin/assets/docs/Documento_CRISP-DM.2385037.pdf
- [71] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," 2005. [Online]. Available: <http://crd.lbl.gov/>
- [72] MathWorks, *Machine Learning with MATLAB*. 2019. Accessed: Jun. 27, 2022. [Online]. Available: <https://la.mathworks.com/content/dam/mathworks/ebook/gated/machine-learning-workflow-ebook.pdf>

- [73] “ICACIT.” <https://www.icacit.org.pe/web/es/component/content/article/16-eventos/190-reunion-de-coordinadores-de-acreditacion-octubre-03-icacit.html>
(accessed Jun. 21, 2022).
- [74] ICACIT, “Criterios de Acreditación para Programas de Pregrado,” 2020. [Online].
Available: www.icacit.org.pe,
- [75] “Support Vector Machine (SVM) - MATLAB & Simulink.”
<https://la.mathworks.com/discovery/support-vector-machine.html> (accessed Jun. 29,
2022).