UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO

FACULTAD DE INGENIERÍA

ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

Aplicación web basado en machine learning para predecir el rendimiento académico en matemática del tercer grado en un colegio militar

TESIS PARA OPTAR EL TÍTULO DE

INGENIERO DE SISTEMAS Y COMPUTACIÓN

AUTOR

Edwin Manuel Bazan Fernandez

ASESOR

Jose Segundo Castillo Zumaran

https://orcid.org/0000-0001-5613-5519

Chiclayo, 2025

Aplicación web basado en machine learning para predecir el rendimiento académico en matemática del tercer grado en un colegio militar

PRESENTADA POR

Edwin Manuel Bazan Fernandez

A la Facultad de Ingeniería de la
Universidad Católica Santo Toribio de Mogrovejo
para optar el título de

INGENIERO DE SISTEMAS Y COMPUTACIÓN

APROBADA POR

Hector Miguel Zelada Valdiviezo
PRESIDENTE

Miguel Angel Diaz Espino
SECRETARIO

Jose Segundo Castillo Zumaran
VOCAL

Dedicatoria

Este trabajo de investigación está dedicado, con profundo agradecimiento y admiración, a todas las personas que han contribuido a mi formación académica y personal.

A mis padres, por su amor incondicional, su apoyo constante y sus sabias enseñanzas, que han sido el pilar fundamental de mi vida. Gracias por creer en mí y por brindarme las herramientas necesarias para alcanzar mis metas.

Agradecimiento

A mis compañeros de estudio, por su colaboración, camaradería y apoyo mutuo. Los momentos compartidos, los retos superados juntos y las discusiones enriquecedoras que han hecho de esta experiencia algo inolvidable.

A los profesores, profesionales y asesor cuya guía, conocimiento y paciencia han sido invaluables en mi camino académico. Su dedicación y pasión por la enseñanza han sido una fuente constante de inspiración.

Resumen del informe de similitud

Aplicación web basado en machine learning para predecir el rendimiento académico en matemática del tercer grado en un colegio militar

INFORME DE ORIGINALIDAD

15%	14%	6%	6%
ÍNDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	tesis.usat.edu.pe	3%
2	www.coursehero.com	1%
3	revistas.up.ac.pa	1%
4	hdl.handle.net	<1%
5	ojs.urbe.edu	<1%
6	alicia.concytec.gob.pe	<1%
7	revistas.ulasalle.edu.pe	<1%
8	www.grafiati.com	<1%
9	Mestas Yucra, Edwin Edgar. "Modelo basado en Deep Learning para predecir el ingreso de estudiantes a la UNA Puno", Universidad Nacional del Altiplano de Puno (Peru) Publicación	<1%
10	repositorio.ulima.edu.pe	<1%

Índice

Resumen 6

Abstract 7

Introducción 8

Revisión de literatura 10

- **Antecedentes** 10
- **Bases Teóricas Científicas** 13

Materiales y métodos 16

- **Tipo de Investigación** 16
- **Técnicas de Investigación** 16
- **Metodologia de Desarrollo** 17

Resultados y discusión 20

Conclusiones 52

Recomendaciones 53

Referencias 54

Anexos 57

Resumen

En esta investigación se determinaron ciertas características claves que influyen en el rendimiento académico estudiantil, con la finalidad de predecir aquellos estudiantes que tengan alta probabilidad de desaprobar. Entre los principales problemas abordados se encuentran los factores que afectan el rendimiento académico, la deserción estudiantil y el alto porcentaje de desaprobados, la cual genera un incremento en los costos operativos, como la ampliación de horas laborales de los docentes y el uso intensivo de recursos e infraestructura. Para abordar esta problemática se propuso desarrollar una aplicación web basado en Machine Learning permitiendo predecir el rendimiento académico, teniendo en cuenta los siguientes objetivos específicos: determinar cuáles son las variables para predecir con mayor grado de precisión el rendimiento académico, determinar el algoritmo que nos proporcione la mayor precisión al predecir el rendimiento académico, crear un modelo de predicción en base a los datos recolectados y el algoritmo seleccionado, desarrollar la aplicación web como herramienta de interacción, validar la aplicación web construida en base a técnicas de pruebas de ingeniería de software. La metodología empleada es CRISP-DM, utilizada en proyectos de minería de datos y Machine Learning. Durante el desarrollo se evaluaron tres algoritmos de regresión, obteniendo los resultados: Algoritmo Regresión Lineal (ECM:14.15, RMSE:3.76, Índice de Concordancia:99.99988%), algoritmo Gaussian Naive Bayes ECM y el algoritmo regresor de Árbol de Decisión (ECM: 8.49, RMSE:91, Índice de Concordancia: 99.99992%), tras evaluar los resultados se seleccionó el algoritmo con mejor desempeño para implementar el modelo de predicción del rendimiento académico de los nuevos estudiantes.

Palabras Claves: Aprendizaje Automático, Rendimiento Académico, Modelo de predicción, algoritmos de entrenamiento, CRISP-DM.

Abstract

This research identified key characteristics that influence students' academic performance, with the aim of predicting those who are at high risk of failing. The main issues addressed include factors affecting academic performance, student dropout, and the high percentage of failing students, which leads to increased operational costs such as extended teaching hours and intensive use of resources and infrastructure.

To address this problem, the development of a web application based on Machine Learning was proposed to predict academic performance, considering the following specific objectives: to identify the variables that most accurately predict academic performance; to determine the algorithm that provides the highest accuracy in academic performance prediction; to create a prediction model based on the collected data and the selected algorithm; to develop the web application as an interactive tool; and to validate the web application based on software engineering testing techniques.

The methodology used was CRISP-DM, commonly applied in data mining and Machine Learning projects. During the development, three regression algorithms were evaluated, yielding the following results: Linear Regression (MSE: 14.15, RMSE: 3.76, Concordance Index: 99.99988%), Gaussian Naive Bayes, and Decision Tree Regressor (both with MSE: 8.49, RMSE: 2.91, Concordance Index: 99.99992%). Based on these results, the algorithm with the best performance was selected to implement the predictive model for forecasting the academic performance of new students.

Keywords: Machine Learning, Academic Performance, Prediction Model, training algorithms, CRISP-DM.

Introducción

Según la Organización para la Cooperación y el Desarrollo Económico (O.C.D.E), la cual es un organismo internacional conformado por 38 países, se propuso medir si los estudiantes están preparados para aplicar sus conocimientos y habilidades en lectura, matemática y ciencias mediante el programa internacional de evaluación de los estudiantes (P.I.S.A), este programa es aplicado cada tres años en setenta y nueve países, permitiendo así evaluar su rendimiento académico.

En la evaluación P.I.S.A realizado en el año 2022, los resultados se clasificaron en dos categorías, la categoría internacional y la categoría Latinoamericana. En la categoría internacional los cinco países con mayor puntaje fueron Singapur (575 puntos), Macao (552 puntos), Taipéi (547 puntos), Hong Kong (540 puntos) y Japón (536 puntos), en la categoría Latinoamericana los cinco países con mejor desempeño fueron Chile (412 puntos), Uruguay (409 puntos), México (395 puntos), Perú (391 puntos) y Costa Rica (385 puntos), dichos resultados reflejan el impacto sobre las interrupciones por la pandemia de Cobit19, el bajo nivel de ingresos y otras circunstancia propias de cada país. [1].

Según García [2], quien realizó una investigación en la Universidad César Vallejo, los principales problemas identificados fueron el alto porcentaje de estudiantes reprobados y el índice de deserción. Por su parte, Bermúdez [3] en su estudio realizado en la Universidad Nacional del Callao, destacó como problemas el bajo nivel académico de los estudiantes en los últimos años y la necesidad de identificar a los estudiantes reprobados.

Esta investigación se aplicó en un colegio militar del tercer grado de educación secundaria, con la finalidad de identificar los problemas de la institución, relacionados con la identificación de variables que afectan el rendimiento académico, con la finalidad de disminuir el porcentaje de desaprobados, prediciendo aquellos estudiantes que podrían desaprobar o aprobar el curso de matemáticas, reduciendo el pago del docente por las horas de estudio acumuladas y así disminuir el incremento del costo que genera el uso continuo de recursos e infraestructura.

Basado en el análisis anterior se formula la siguiente pregunta de investigación: ¿Cómo puede un modelo basado en Machine Learning predecir el rendimiento académico de los estudiantes del tercer grado de educación secundaria en el curso de matemática? Desde el ámbito científico se justifica porque contribuye al avance del conocimiento sobre machine learning, proporcionando una base sólida para futuras investigaciones, enfocados en la predicción del rendimiento académico en instituciones educativas como colegios, institutos y universidades pedagógicas. Desde el ámbito tecnológico se justifica por el desarrollo de la

aplicación web basada en Machine Learning, uso de herramientas de software libre como MySQL, Python, Bootstrap, Django. Se utilizo MySQL como gestor de base de datos para registrar los datos necesarios, se utilizó Python como lenguaje de programación para el desarrollo de la aplicación web, se utilizó el framework Bootstrap para realizar la interfaz gráfica de Usuario (G.U.I), que sirve como interacción entre la aplicación web y el usuario para su correcto funcionamiento, se utilizó el framework Django para el desarrollo e implementación Aplicación web. Los recursos de Hardware y software necesarios son los mismos que los colaboradores utilizan dentro de la institución para el cumplimiento de sus funciones, adicionalmente se realizó el entrenamiento de tres algoritmos de predicción, con la finalidad de seleccionar el algoritmo más optimo que permita generar el modelo de predicción mejorando así el rendimiento académico de los nuevos estudiantes. Desde el ámbito social se justifica porque está demostrado que gracias al uso de aplicaciones web, software libre, sistemas Información se puede mejorar significativamente el rendimiento académico de los estudiantes, permite mejorar la calidad de los servicios ofrecidos, optimizar los procesos de enseñanza y aprendizaje.

Un buen rendimiento académico en los estudiantes es clave para el progreso y la prosperidad de la sociedad, fomentar el desarrollo del capital humano cualificado, adicionalmente permite aplicar sus habilidades y conocimientos para impulsar el crecimiento económico, la innovación, el desarrollo social y cultural, contribuye directamente al desarrollo de una sociedad más preparada, innovadora y prospera. Desde el ámbito económico se justifica porque reduce el porcentaje de estudiantes desaprobados, reduce el costo del pago al docente por disminuir las horas de estudio, permite reducir al padre de familia o apoderado los costos asociados por desaprobar uno o más curso o por repetir el año académico, optimiza los recursos y maximiza el retorno de inversión de la institución.

Después de identificar los problemas y justificar el desarrollo de esta investigación, se plantea como objetivo principal desarrollar una aplicación web basada en machine learning para predecir el rendimiento académico que será logrado en base a los siguientes objetivos específicos, que es determinar cuáles son las variables para predecir con mayor grado de precisión el rendimiento académico, para determinar cuál es el algoritmo que nos proporciona la mayor precisión al predecir el rendimiento académico, facilitando la creación del modelo de predicción en base a los datos recolectados y el algoritmo seleccionado, para desarrollar la aplicación web como herramienta de interacción, y finalmente validar la aplicación web construida en base a técnicas de pruebas de la ingeniería de software, basándose en técnicas, normas, practicas.

Revisión de literatura

Antecedentes

Antecedentes Internacionales

Según Rico et at [4], este estudio abordó el alto porcentaje de desaprobados en los cursos de ingeniería y el bajo desempeño académico de los estudiantes en una universidad pública de México, su objetivo fue desarrollar una metodología para construir modelos predictivos del rendimiento académico a partir de las características de 228 estudiantes de ingeniería, para evaluar estos modelos se emplearon diferentes métricas de desempeño, los datos fueron recolectados al inicio del ciclo académico y se utilizaron tres técnicas de aprendizaje automático: Naïve Bayes, Vecinos Más Cercanos y Árbol de Decisión, los modelos se evaluaron según la precisión, la tasa de verdaderos positivos y la tasa de verdaderos negativos. Los resultados obtenidos fueron los siguientes: Arbol de Decisión su porcentaje de precisión fue del 59.2%, tasa de verdaderos positivos es del 67.9% y tasa de verdaderos negativos es del 46.8%, Vecinos Más Cercanos su porcentaje de precisión del 64.5%, tasa de verdaderos positivos es del 85.1% y tasa de verdaderos negativos es del 35.1%, Naïve Bayes su porcentaje de precisión del 62.7%, tasa de verdaderos positivos es del 70.1% y tasa de verdaderos negativos es del 52.1%.. El estudio concluyó que el algoritmo de Vecinos Más Cercanos es el más adecuado. Además, se determinó que la metodología propuesta puede aplicarse en otras instituciones educativas para predecir el rendimiento académico de sus estudiantes, este antecedente se relaciona con la investigación en curso, ya que ambos estudios utilizan tres algoritmos para predecir el rendimiento académico a partir de ciertas variables específicas extraídas de un conjunto de registros.

En la investigación de Sánchez et al. [5], se identificaron problemas como el alto índice de estudiantes desaprobados, el incremento de la deserción escolar y el bajo rendimiento académico en una o varias asignaturas, el objetivo principal fue diseñar y automatizar un modelo predictivo del rendimiento académico en el Instituto Politécnico Nacional de México (IPN), este modelo se aplicó a 86 estudiantes para predecir su desempeño, los 94 registros recopilados se dividieron aleatoriamente en dos grupos de 47 registros cada uno. La precisión del modelo en el primer grupo fue del 59.57%, mientras que en el segundo alcanzó el 68.09%. El estudio concluyó que el modelo predictivo es accesible tanto para docentes como para expertos en minería de datos, además, permitió identificar las principales actividades académicas que influyen en el rendimiento estudiantil, facilitando la implementación de estrategias de mejora dentro de la institución educativa. Este antecedente es relevante, ya que

utilizó un conjunto de datos históricos para entrenar el modelo y posteriormente aplicarlo a nuevos estudiantes.

Antecedentes Nacionales

La investigación de Gutiérrez et al. [6], realizada en la Facultad de Ciencias Empresariales y Económicas de la Universidad de Lima, abordó la dificultad de identificar a estudiantes en riesgo académico, entre los problemas detectados se encontraban el incumplimiento de los plazos de estudio y el bajo rendimiento en los cursos de sus respectivas carreras, el estudio tuvo como objetivo evaluar técnicas de aprendizaje automático para predecir el rendimiento académico, utilizando los algoritmos Árbol de Decisión, Bosques Aleatorios, Redes Neuronales y Máquinas de Soporte Vectorial. Los resultados de los algoritmos de predicción fueron, el algoritmo Arbol de decisión tubo una precisión del 73% y su curva de ROC fue del 72.9%, el algoritmo de Bosques Aleatorios tubo una precisión del 80.31% y su curva ROC fue del 70.2%., el algoritmo Máquinas de Soporte Vectorial tubo una precisión del 72.44% y su curva ROC fue del 0.02%, el algoritmo Redes Neuronales tubo una precisión del 78.74% y su curva ROC fue del del 71.1%., Se concluyó que el algoritmo de Bosques Aleatorios ofrece la mayor precisión en la predicción del rendimiento académico. Este estudio es relevante para la presente investigación, ya que empleó diversas variables para determinar el desempeño estudiantil, implementó cuatro algoritmos de predicción y estableció cuál presentaba el mayor porcentaje de precisión para predecir el rendimiento académico.

En el estudio de Gianira et al. [9], se identificaron como principales problemas la clasificación de los estudiantes según su rendimiento académico y la alta deserción debido al bajo desempeño, para abordar estos desafíos, se propuso un modelo de *machine learning* que optimiza el proceso de clasificación en distintos niveles de rendimiento. Los resultados mostraron que el sistema tradicional requería un tiempo promedio de 218.19 minutos para la clasificación, mientras que el modelo propuesto redujo este tiempo a 55.42 minutos, logrando una disminución de 162.77 minutos (74.60%). Esto permitió a la institución obtener información actualizada en menos tiempo. El análisis de los resultados concluyó que la metodología de machine learning utilizada mejoró la clasificación académica y contribuyó a reducir la deserción estudiantil al agilizar la disponibilidad de información clave. Este antecedente es relevante para la presente investigación, ya que demuestra cómo el uso de machine learning puede mejorar el desempeño académico al optimizar el proceso de clasificación estudiantil.

Antecedentes Locales

En la investigación realizada por García [5], realizada en la I.E. Sara Buyón, se identificó como problemática el bajo rendimiento académico de los estudiantes, influenciado por factores psicológicos. Para abordar esta situación, se propuso un sistema de predicción del rendimiento académico basado en diversos factores, utilizando minería de datos y la metodología CRISP-DM para analizar la información de los estudiantes se implementaron los algoritmos Random Forest y Decisión Tree como base del modelo de predicción, que posteriormente fue automatizado mediante una interfaz que permite ingresar datos y obtener predicciones en tiempo real. Los resultados mostraron que el algoritmo Decisión Tree alcanzó una precisión del 94.4%, mientras que Random Forest obtuvo un 95.8%, la investigación concluyó que Random Forest fue el algoritmo más preciso (95.8%) y se utilizó para la automatización del modelo. Además, se determinó el grado de importancia de cada variable en la predicción del rendimiento académico, aportando un antecedente clave en la selección y aplicación de modelos predictivos en el ámbito educativo.

En la investigación de Vargas [6], realizada en una universidad privada, se identificó como problemática la ausencia de un sistema informático para predecir el desempeño docente. Esto generaba una evaluación sujeta a la subjetividad de la comisión encargada, retrasos en el análisis debido al almacenamiento de grandes volúmenes de datos en hojas de cálculo y dificultades para obtener información en tiempo real con alta precisión. Como solución, se propuso la implementación de un sistema web basado en técnicas de machine learning para predecir el desempeño docente en educación superior. Los resultados obtenidos tiendo en cuenta la precisión de los algoritmos es de Regresión logística 58.3% de precisión, Máquinas de soporte vectorial (SVM) 59.8% de precisión, Árbol de decisión (Decision Tree Classifier): 87.3% de precisión, Bosques aleatorios (Random Forest Classifier): 90.5% de precisión, K-Vecinos más cercanos (K-Neighbors Classifier): 87.6% de precisión, Naive Bayes Gaussiano (Gaussian NB): 53.9% de precisión. Este antecedente fue seleccionado debido a su similitud con la presente investigación, ya que ambos estudios desarrollaron un sistema informático como herramienta tecnológica de predicción que permite predecir el desempeño docente en educación superior.

Bases Teóricas Científicas

Rendimiento Académico:

Mide el nivel de aprendizaje y logro de un estudiante en un periodo específico, evaluado mediante exámenes, tareas, proyectos y otras actividades académicas definidas por las instituciones educativas. En esta investigación, este concepto se aplicó en el desarrollo de un modelo de predicción basado en Machine Learning, que utiliza datos históricos para evaluar y anticipar el desempeño estudiantil, su objetivo es mejorar el rendimiento académico de los nuevos estudiantes antes de finalizar el año escolar [7]

Grados Académicos:

Son títulos, constancias o certificados otorgados por instituciones educativas como universidades, institutos o colegios, que reconocen formalmente el nivel de estudios alcanzado en un campo específico. Representan un indicador del conocimiento y habilidades adquiridos.

En esta investigación, el concepto se aplica a los tres grados académicos de la institución: tercer, cuarto y quinto grado, destacando el tercero por presentar el mayor porcentaje de desaprobados en matemáticas. [8]

Aplicación Web:

Es un programa que se ejecuta en un servidor remoto y se accede a través de un navegador web mediante una conexión a internet. Es independiente de la plataforma y el sistema operativo, y todas las actualizaciones se realizan en el servidor para garantizar que el usuario siempre tenga la última versión. No requiere instalación ni descarga, solo conexión a internet. En esta investigación, se adoptó este enfoque para desarrollar una aplicación web interactiva y accesible, diseñada para procesar datos históricos, entrenar y visualizar algoritmos de predicción. Además, permite generar y mostrar el modelo de predicción, facilitando la estimación del rendimiento académico de nuevos estudiantes con el algoritmo más optimo [5]

Variables - características:

Son las propiedades o atributos que describen los datos de entrada, son utilizados para entrenar los algoritmos de predicción utilizando los datos históricos y también para generar el modelo de predicción utilizando los datos actuales. En esta investigación se ha tenido en cuenta la investigación según el autor Ruiz, en su investigación considero 32 variables o características para generar un modelo de predicción, basado en dicha investigación se implementó la funcionalidad que la aplicación web permita seleccionar y validar un primer archivo en formato Excel que contiene las características o variables de los datos históricos con la finalidad de generar los algoritmos de entrenamiento, adicionalmente la aplicación web también permite

seleccionar y validar un segundo archivo en formato Excel que contiene las características o variables de los datos actuales con la finalidad de generar el modelo de predicción [9].

Sistemas de Predicción:

Es un conjunto de algoritmos basados en modelos matemáticos o estadísticos que predicen eventos futuros, tendencias o comportamientos, utilizando datos históricos o en tiempo real para analizar patrones y generar pronósticos que faciliten la toma de decisiones. En esta investigación se desarrolló una aplicación web que presenta tres algoritmos de entrenamiento y un modelo de predicción para estimar el rendimiento académico de los estudiantes de tercer grado [7].

Inteligencia Artificial:

Es un área de la informática dedicada a crear sistemas que realizan tareas que suelen requerir el conocimiento de un experto, con el fin de mejorar la toma de decisiones. En este proyecto se implementaron 3 algoritmos de predicción y un modelo de predicción, basado en inteligencia artificial, utilizando características que influyen en el rendimiento académico para estimar el desempeño de los estudiantes. [10]

Ramas de la Inteligencia Artificial:

Machine Learning o Aprendizaje Automático:

Es una rama o sub conjunto de la inteligencia artificial que se enfoca en desarrollar algoritmos y modelos de predicción, permitiendo a las maquinas aprender de los datos sin necesidad de ser programadas de programación especifica. En este proyecto se desarrolló una aplicación web basado en Machine learning que permita predecir el rendimiento académico de los estudiantes del tercer grado. [11]

Técnicas Machine Learning o Aprendizaje Automático:

Aprendizaje Supervisado:

Es una técnica que aprende patrones apartir de datos etiquetados (entradas y salidas), para realizar tareas autónomas y predecir nuevos datos, mejorando la toma de decisiones. En este proyecto se desarrollaron tres algoritmos y un modelo de predicción con veinte variables, tanto independientes como dependiente, para estimar el desempeño de los estudiantes de tercer grado. [12]

Tipos de Aprendizaje Supervisado

Algoritmos de Regresión

Son algoritmos que identifican patrones en datos históricos, generan una función matemática para predecir nuevos datos y minimizan el error al analizar relaciones entre variables, ya sean lineales o no lineales. Esta investigación desarrolla tres

algoritmos de regresión para entrenar un modelo con datos históricos y seleccionar el más preciso para predecir datos de nuevos estudiantes. [13]

Tipos o clasificación de los Algoritmos de Regresión

Estos algoritmos son ideales para modelar la relación entre variables independientes y dependientes, utilizando datos históricos para entrenamiento o datos actuales para generar modelos de predicción. Además, recomienda emplear tres algoritmos de regresión para predecir el rendimiento académico.

Regresión Lineal Múltiple:

Modela la relación entre las variables mediante una línea recta. Este algoritmo se utilizó para determinar cómo las variables independientes influyen en la variable dependiente, que corresponde a la nota del estudiante.

Regresión de Bosque Aleatorio o clasificador de Bosque aleatorio:

Utiliza múltiples árboles de decisión para promediar predicciones, mejorando la precisión y robustez. Se empleó este algoritmo para analizar las relaciones entre las variables independientes y la variable dependiente.

Árbol de decisión o regresión de Árbol de decisión:

Se utiliza para predecir valores continuos dividiendo los datos en subconjuntos según las características más relevantes. Este algoritmo se empleó para modelar la relación entre las variables independientes y la dependiente [14]

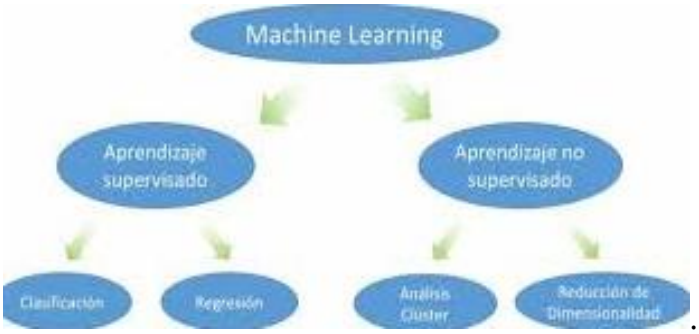


Fig. 1. Estructura de los algoritmos

En esta interfaz se muestra los algoritmos de predicción, clasificados por las técnicas de aprendizaje supervisado y no supervisado, que estado dentro del aprendizaje automático.

Modelo de Predicción

Es una herramienta que utiliza un determinado algoritmo de predicción, para aprender patrones y relaciones subyacentes, con el objetivo de hacer predicciones sobre datos nuevos o futuros, se incluyó este concepto porque la aplicación web

desarrollada selecciona en forma automática el algoritmo de entrenamiento más preciso con la finalidad de generar el modelo y predecir el rendimiento académico de los nuevos estudiantes, este modelo tiene los siguientes reportes : Estudiantes que podrían aprobar o desaprobar antes de finalizar el año académico, porcentaje de estudiantes que podrían aprobar o desaprobar el curso [15]

Materiales y métodos

Tipo de Investigación

Según el manual de Frascati [16], esta investigación es de tipo aplicada, ya que busca reducir la alta tasa de desaprobación mediante predicciones basadas en variables, sus categorías y la nota inicial del estudiante, con el objetivo de mejorar el rendimiento académico de los nuevos estudiantes.

Técnicas de Investigación

TABLA I

TÉCNICAS E INSTRUMENTOS PARA LA RECOLECCIÓN DE DATOS

Técnica	Instrumento	Población	Sustento
Entrevista	Grabadora de audio o video, apuntes registrados.	Áreas o departamentos que generan el registro de los datos necesarios.	Determinar las áreas o departamentos que genera los registros de datos. Determinar los requerimientos de Software y Hardware necesario aplicación web
Observación	Reportes del listado de Aprobados y desaprobados por grado académico, curso matemática del 2018-2022	Estudiantes del tercero, cuarto y quinto grado, curso matemática	Visualizar los reportes de los tres grados académicos con mayor porcentaje de desaprobados.
Registro Datos	Diversidad de fuente de datos históricos y actuales.	Estudiantes del curso de matemática del tercer grado secundaria.	Recopilar los datos históricos y los datos actuales datos en un archivo Excel que tiene las características de los estudiantes, las categorías de cada variable y una nota inicial del tercer grado en el curso de matemática de años académicos anteriores y del año actual.
Encuesta	Cuestionario de funcionalidad y usabilidad aplicación web	Jefe del departamento de evaluación y registro	Evaluar a través de un cuestionario la funcionalidad y Usabilidad de la aplicación web con el jefe de departamento de evaluación y registro.

En la tabla I se detalla las técnicas e instrumentos aplicados para la recolección de datos, indicando la población objetivo y el sustento que justifica cada procedimiento. Estas acciones garantizaron la obtención de información pertinente para el desarrollo del modelo predictivo.

TABLA II

MÉTODOS APLICADOS EN LA INVESTIGACIÓN Y SU SUSTENTO

Método	**Sustento por el cual será empleado en la investigación**
Analítico	Se aplico en la fase 1, entendimiento del negocio, consistió en el análisis de las características o variables, las categorías de cada variable, que influyen en el rendimiento académico en el curso de matemática del tercer grado, se seleccionó el algoritmo con mejor desempeño para implementar el modelo de predicción del rendimiento académico de los nuevos estudiantes.
Deductivo	Se aplico en la fase 3, preparación de los datos, que implico transformar los datos en un formato adecuado para su análisis, utilizando los algoritmos de regresión, para poder generar el modelo de predicción.
Implementación	Se aplico en la fase 5, evaluación, que permitió que la Aplicación web basada en Machine Learning, muestre en forma correcta los algoritmos de entrenamiento, el modelo de predicción, a su vez verificar que los mismos cumplan con los objetivos planteados al iniciar el proyecto, utilizando técnicas, herramientas, para evaluar los resultados.

En la tabla II, se detalla los métodos utilizados en el desarrollo de la investigación, especificando la justificación de su aplicación en las distintas fases, desde el análisis de variables hasta la implementación y evaluación del modelo predictivo.

Metodologia de Desarrollo

En este punto se listaron todas las actividades realizadas en cada fase de la metodología **CRISP-DM** propuesta en esta investigación.

Iteración #1- Primera Fase - Bussiness Understanding – Entendimiento del negocio:

Se desarrollaron ocho actividades, las cuales se detallan a continuación:

Se solicitó y analizó el organigrama del colegio

Se identificaron las necesidades del colegio.

Se determinaron las causas que motivaron el desarrollo del proyecto.

Se establecieron la misión y visión del negocio

Se definieron los objetivos del negocio.

Se seleccionó y aplicó el modelo pedagógico institucional.

Se identificaron los factores que influyeron en el rendimiento académico.

Se especificaron los requerimientos funcionales y no funcionales de la aplicación web.

Iteración #2 - Segunda Fase - Data Understanding – Entendimiento o comprensión de los Datos: Se desarrollaron cinco actividades, las cuales se detallan a continuación:

Se recolectaron los datos iniciales.

Se describieron los datos obtenidos.

Se cuantificaron los datos iniciales.

Se estandarizaron y formatearon las fuentes de datos.

Se analizaron los resultados y se determinó el porcentaje de aprobados y desaprobados.

Iteración #3 - Tercera Fase - Data Preparación - Preparación de los Datos: Se desarrollaron nueve actividades, las cuales se detalla a continuación:

Se cuantificó el total de datos históricos y datos actuales iniciales a utilizar.

Se evaluó la calidad de los datos históricos y de los nuevos datos iniciales.

Se descartaron los datos vacíos, incorrectos o inconsistentes de la data histórica y de la nueva data a utilizar.

Se realizó la depuración de los datos históricos y de los nuevos datos.

Se efectuó la integración de los datos históricos y actuales a utilizar.

Se determinaron y ubicaron los datos de entrenamiento y los nuevos datos.

Se seleccionaron los datos de entrenamiento y los nuevos datos.

Se listaron las variables a emplear con sus respectivas categorías.

Se formatearon los datos históricos y actuales para su correcta lectura en la aplicación web.

Iteración #4 - Cuarta Fase: Modeling – Modelado: Se desarrollaron Diesi siete actividades, las cuales se detalla a continuación:

Se descargaron e instalaron, mediante la URL, todas las herramientas necesarias para los modelos de entrenamiento y validación.

Se actualizó el gestor de paquetes Pip.

Se activaron las librerías de XAMPP, PhpMyAdmin y MySQL.

Se estableció la conexión entre la aplicación web y MySQL.

Se configuró la migración Django – MySQL.

Se creó y configuró un usuario tipo super usuario en Django – MySQL.

Se configuró la interacción entre PhpMyAdmin – MySQL y la integración MySQL – Django – Python.

Se integró la aplicación web Django dentro del entorno de Visual Studio Code.

Se configuraron las plantillas para el uso de las interfaces de la aplicación web.

Se gestionaron las tablas y registros de la base de datos de la aplicación web.

Se diseñaron y desarrollaron los módulos de acceso de usuarios.

Se diseñaron y desarrollaron los modelos de entrenamiento y validación.

Se implementó la plantilla e interfaz para generar los modelos de entrenamiento y validación.

Se elaboró un cuadro comparativo para identificar el algoritmo con mayor porcentaje de prediccion.

Se diseñó y desarrolló el modelo de predicción.

Iteración #5 - Quinta Fase - Evaluatión – Evaluación: Se desarrollaron tres actividades, las cuales se detalla a continuación:

Se evaluó el rendimiento del modelo de predicción seleccionado.

Se comparo los resultados entre los algoritmos de entrenamiento implementados.

Se valido los resultados del modelo seleccionado con expertos o usuarios finales.

Se analizo si el algoritmo seleccionado es el más optimo.

Se identifico los posibles limitaciones o errores del modelo

Se volvió a revisar todo el proceso de desarrollo.

Iteración #6 - Sexta Fase - Deployment – Despliegue – Implantación: Se desarrollaron seis actividades, las cuales se detalla a continuación

Se instalo y configuro todo el desarrollo de la aplicación en el servidor web final, utilizando las librerías, paquetes necesarios.

Se aplicaron las técnicas de pruebas durante el desarrollo de la aplicación web.

Se presentaron las recomendaciones finales basadas en la metodología.

Se elaboró un resumen de la metodología CRISP-DM.

Resultados y discusión

1. Resultados de acuerdo a los objetivos de la Investigación

a. Determinar cuáles son las variables para predecir con mayor grado de precisión el rendimiento académico en el curso de Matemática de 3er año de secundaria.

TABLA III

TESIS CONSULTADAS Y SU APLICACIÓN EN LA INVESTIGACIÓN

Tesis	Autor / Autores	Acciones Realizadas
Rendimiento académico de estudiantes en educación superior: predicciones de factores influyentes a partir de árboles de decisión	Diaz landa, Brenda Nicole; melean romero, Rosana Alejandra; marín-rodriguez, William Joel	La tesis fue analizada y considerada como referencia para la implementación de los algoritmos de entrenamiento; no obstante, no se empleó para definir la estructura de las variables ni el tipo de predicción
Sistema web basado en técnicas de machine learning para predecir el desempeño de los docentes en el nivel de educación superior universitaria	Chatila Vargas Toscanelli, Habana Sabrá	Se analizó esta tesis y se empleó como referencia para el desarrollo de la aplicación web; sin embargo, no se utilizó para definir las posibles variables a considerar.
Modelo de machine learning para la clasificación de estudiantes de acuerdo a su rendimiento académico en el centro de idiomas de la universidad nacional del santa	Espinoza airac, gianira Xiomara; león muñoz, eduar Fabián	Se analizo esta tesis y se utilizó como referencia para determinar las técnicas y métodos de la investigación, pero no se aplicó para definir las posibles variables a utilizar en la investigación.
Predicción de Rendimiento Académico de alumnos usando Machine Learning	Gutiérrez Villaverde, Herberth; Linares Barbero, Manuela; Agüero Correa, Ángel; Pérez Nuñe, Jhelly	El análisis de la tesis permitió determinar que se utilizaron treinta y dos variables para predecir el rendimiento académico de los ingresantes de la facultad de ciencias empresariales y económicas de la universidad de Lima

En la Tabla III se revisaron las tesis consideradas como antecedentes de la investigación, detallando a sus autores y las acciones realizadas en cada caso. Su propósito es mostrar cómo estos trabajos previos sirvieron de referencia para orientar la metodología y el diseño del modelo predictivo desarrollado en el presente estudio. En particular, se tomó como base la tesis de Gutiérrez et al. [11], que propone el uso de treinta y dos variables. A partir de esta propuesta, se seleccionaron las variables empleadas en la presente investigación y, adicionalmente, se implementó la Iteración 3, correspondiente a la Fase 03: Preparación de los datos, conforme a la metodología CRISP-DM

TABLA IV

VARIABLES CONSIDERADAS EN LA INVESTIGACIÓN RELACION EL RENDIMIENTO ACADÉMICO

N	Variable	Definición	Grado Importancia	Relación-rendimiento académico	Fuente de datos.
1	Código estudiante	Identificador que se asigna a cada estudiante dentro de un sistema educativo	Baja	No existe relación directa.	Registros académicos, Historial Académico
2	Edad	Cantidad años cumplidos por un estudiante	Baja	No existe relación directa	Registros académicos, Historial Académico
3	Sexo	Genero asignado al estudiante	Baja	No existe relación directa	Registros académicos, Historial Académico
4	DNI	Identificador que se asigna a cada persona por el sistema gubernamental	Baja	No existe relación directa	Registros académicos, Historial Académico
5	Año Académico	Periodo Escolar o universitario en que el estudiante está inscrito, analizar desempeño académico	Baja	Se selecciono los años académicos: 2018, 2019, 2020, 2021,2022(algoritmos entrenamiento) 2023 (modelo predicción)	Registros académicos, Historial Académico
6	Grado y Sección	Representa la etapa educativa y el grupo del estudiante	Baja	Se selecciono todas las secciones del tercer grado del año 2018 al 2023	Registros académicos, Historial Académico
7	Promedio	Medida cuantitativa que refleja el desempeño del estudiante durante periodo específico.	Alta	Resume nivel de éxito estudiantes, predicciones posibles resultados	Registros académicos, Historial Académico
8	Familiar que representa	Son los factores relacionados con el entorno familiar del estudiante.	Alta	Apoyo emocional, social, económico del estudiante núcleo familia, impacto significativo desempeño escolar, dinámicas de apoyo.	Registros académicos, Historial Académico
9	Grado Instrucción Padre	Es el nivel educativo alcanzado por el padre o tutor del estudiante.	Alta	Oportunidades de aprendizaje, apoyo al estudiante, recursos disponibles al estudiante.	Registros académicos, Historial Académico
10	Grado Instrucción Madre	Es el nivel educativo alcanzado por la madre o tutor del estudiante.	Alta	Oportunidades de aprendizaje, apoyo al estudiante, recursos disponibles al estudiante.	Registros académicos, Historial Académico
11	Desinterés del Rendimiento académico	Es la falta de motivación, compromiso o interés del estudiante en las actividades escolares	Alta	Métodos inadecuados Apatía, desconexión del estudiante en sus actividades escolares	Registros académicos, Historial Académico

12	Motivación Rendimiento académico	Nivel de interés, entusiasmo, compromiso del estudiante en sus estudios y metas educativas	Alta	Esfuerzo del estudiante en el aprendizaje, importancia educación en la vida del estudiante	Registros académicos, Historial Académico
13	Comunicación con el estudiante	Calidad, frecuencia y efectividad interacciones estudiante, docente, padres o tutores, compañeros.	Alta	Influye calidad del aprendizaje, la motivación, el bienestar emocional, mejora progreso académico.	Registros académicos, Historial Académico
14	Exceso de Alcohol.	Consumo frecuente o elevado de bebidas alcohólicas que puede interferir en el desempeño educativo estudiante.	Alta	Afecta el funcionamiento cognitivo, comportamiento, interfiere capacidad del estudiante, reduce desempeño optimo	Registros académicos, Historial Académico
15	Técnicas de Aprendizaje	Son las estrategias y métodos estudiante utiliza adquirir, procesar y retener información	Alta	Influyen forma estudiante procesa, retiene y aplica la información, mejora la comprensión, la mejoría, organización contenida.	Registros académicos, Historial Académico
16	Maltrato Físico	Presencia de cualquier forma agresión física al estudiante, golpes, empujones, quemaduras, lesiones causadas intencionalmente	Alta	Abuso físico que Afecta el bienestar emocional, físico, psicológico, falta de concentración, baja participación actividades académicas.	Registros académicos, Historial Académico
17	Maltrato Psicológico	Cualquier forma de abuso emocional, o verbal que pueda causar daño psicológico al estudiante	Alta	Interfiere en la capacidad de aprendizaje y desempeño escolar dentro o fuera del colegio.	Registros académicos, Historial Académico
18	Padres separados.	Es la situación familiar de los padres, podría ser: separados, divorciados, viven en hogares diferentes	Alta	Generan cambios emocionales, sociales, económicos, que afectan la estabilidad y el entorno educativo	Registros académicos, Historial Académico
19	Discusiones	Es la frecuencia, intensidad y naturaleza de los conflictos verbales en el entorno del estudiante, ya sea en el hogar, la escuela, interacciones sociales.	Alta	Afectan directamente el bienestar emocional, social y cognitivo, estas pueden interferir capacidad del estudiante para concentrarse, participar y superarse académicamente.	Registros académicos, Historial Académico
20	Proceso Aprendizaje Colegio	Conjunto de métodos, recursos, dinámicas pedagógicas, define el entorno educativo que influye en los estudiantes para generar conocimiento.	Alta	Es importante ya que abarca las metodologías, recursos, practicas pedagógicas, entorno escolar del estudiante para adquirir conocimiento y desarrollar habilidades	Registros académicos, Historial Académico

21	Asistencia a clases, Horas de estudio, Calificaciones anteriores, Participación en clase, Acceso a recursos educativos, Autoeficacia percibida, Nivel socio económico, Interés en el curso, Horas de descanso, Actividades fuera del colegio, Relación con el docente, Interés en el curso	Algunas áreas registraron esta variable en los estudiantes, no se utilizó esta variable predicción, no todos los estudiantes tienen registrado esta variable.

En la tabla IV resume las 20 variables consideradas en la investigación, detallando su importancia, relación con el rendimiento académico y fuente de datos. Se identificaron variables de baja relevancia sin relación directa, y se priorizaron aquellas de alta influencia asociadas a factores personales, familiares, sociales y pedagógicos. Un conjunto adicional de variables no se incluyó en la predicción debido a la falta de registros completos.

TABLA V

VARIABLES CONSIDERADAS EN LA INVESTIGACIÓN - CLASIFICACIÓN

N	Variable	Características	Medición	codificación	Observaciones
1	Código Estudiante	Identifica a un estudiante dentro de una institución o de un sistema académico.	Nominal: Solo identifica, no hay jerarquías ni magnitud	Texto continuo sin espacios, seguido de un guion, año en formato entero de cuatro dígitos, guion número entero de tres dígitos. Ejemplo: Est-2018-001	Es una etiqueta individual, se usa para identificar al estudiante dentro del colegio.
2	Edad	Identifica el tiempo de vida que tiene el estudiante	Numérica continua: se mide en años cumplidos	Entero positivo	Se validó que el rango de edades esté comprendido entre 14 y 16 años.
3	Sexo	Identifica el sexo o genero del estudiante.	Nominal: no hay relación de mayor a menor entre las categorías	Diferencia a los estudiantes, masculino o femenino	Se confirmó dos géneros o categorías.
4	DNI	Numero único que identifica legalmente al estudiante.	Nominal: No tiene jerarquía ni valor cuantitativo	Valor numérico de 8 dígitos que no se puede hacer operaciones matemáticas.	No se divulgaron datos personales del estudiante; se utilizaron únicamente para identificar el registro.
5	Año académico	Identifica el periodo escolar o ciclo escolar del estudiante	Ordinal: Esta representado por números, pero funcionan como etiquetas	Se registró solo el año con cuatro dígitos numéricos	Se estandarizó como dato numérico para facilitar el ordenamiento cronológico
6	Grado y sección	Identifica el nivel educativo y grupo especifico al que pertenece el estudiante dentro de la institución	Nominal: Clasifica a los estudiantes en grupos sin jerarquía numérica directa.	Tercero A, Tercero B, Tercero C, Tercero D, Tercero E, Tercero F	Se determinó que el número de secciones depende de la cantidad de estudiantes matriculados, variando entre

					cuatro y seis secciones.
7	Nota Promedio Curso	Identifica la calificación numérica obtenida del estudiante	De razón : Tiene cero absoluto y permite realizar operaciones matemáticas(sumas, promedio, desviación).	Se formateo como número entero positivo, sin decimales y redondeado al valor más próximo	Se validó que las notas se encuentren dentro del rango permitido de 0 a 20.
8	Familiar que representa	Persona que asume la representación legal o administrativa del estudiante dentro de la institución	Nominal: Solo clasifica en categorías, sin orden jerárquico	Analiza si el representante del estudiante influye en su rendimiento académico.	Cada categoría es representada por un número, pero no se admite operaciones matemáticas, solo conteos y frecuencias.
9	Grado Instrucción Padre	Nivel educativo formal alcanzado por el padre del estudiante	Ordinal: Las categorías se clasifican de acuerdo a un orden lógico.	Se registró el nivel educativo del padre correspondiente a cada estudiante	Se registraron y codificaron los niveles educativos alcanzados por el padre de cada estudiante.
10	Grado Instrucción Madre	Nivel educativo formal alcanzado por la Madre del estudiante	Ordinal: Las categorías se clasifican de acuerdo a un orden lógico.	Se registró el nivel educativo de la Madre correspondiente a cada estudiante	Se registraron y codificaron los niveles educativos alcanzados por la Madre de cada estudiante.
11	Padres separados.	Situación familiar en la que los padres del estudiante no conviven como pareja.	Nominal: Clasifica por categorías sin orden jerárquico.	No admite operaciones matemáticas, solo conteos por cada categoría	Permite analizar si existe relación entre la situación familiar y el rendimiento académico.
12	Desinterés del rendimiento académico	Muestra la motivación, el compromiso o interés que muestra el estudiante	Nominal: Puede clasificarse en varias categorías, niveles, sin orden jerárquico.	No admite operaciones matemáticas, solo conteos por cada categoría.	Permite relacionar la actitud del estudiante con sus calificaciones y hábitos de estudio
13	Motivación Rendimiento Académico	Muestra el interés, esfuerzo, perseverancia que muestra el estudiante	Nominal: Puede clasificarse en varias categorías, niveles, sin orden jerárquico.	No admite operaciones matemáticas, pero si operaciones jerárquicas.	Permite relacionar con las calificaciones, hábitos de estudio, asistencia.
14	Comunicación con el estudiante	Evalúa la interacción entre el estudiante, el docente, el personal administrativo, padres de familia	Nominal: Puede clasificarse en varias categorías, niveles, sin orden jerárquico.	Determino el grado, la calidad de interacción, diálogo entre el estudiante y sus padres, tutores o docentes**.**	Permite relaciónar la calidad de comunicación, con el desempeño académico, la conducta y la motivación.
15	Exceso de alcohol	Indica la frecuencia o presencia del consumo excesivo de alcohol del estudiante	Nominal: Puede clasificarse en varias categorías, niveles, sin orden jerárquico.	Consumo frecuente y elevado de bebidas alcohólicas del estudiante	Genera otras consecuencias como ausentismo, trastornos del sueño, baja participación.
16	Técnicas de	Depende del comportamiento y	Nominal: Puede clasificarse en varias	Variedad, frecuencia, la cantidad exacta o aproximada de	Es cognitiva, estratégica, está relacionada con

	Aprendizaje	estrategias del estudiante.	categorías, niveles, sin orden jerárquico.	técnicas empleadas, todo ello a través de categorías	hábitos de estudio y métodos de aprendizaje
17	Maltrato Físico	Acciones que causan daño físico al estudiante	Ordinal: Con jerarquía, se mide por frecuencia	Indica si el estudiante a sufrido agresiones físicas (golpes, empujones, castigos corporales).	Puede ser causado por el propio estudiante, el docente, el apoderado, el padre de familia o algún familiar cercano, sus propios amigos.
18	Maltrato Psicológico	Conjunto de acciones o actitudes verbales y no verbales	Ordinal: Con jerarquía, se mide por frecuencia	Afecta la autoestima, la estabilidad emocional y el bienestar psicológico del estudiante.	Evalúa la percepción del estudiante sobre el impacto emocional.
19	Discusiones	Intercambio verbal cargado de conflicto contra el estudiante.	Ordinal: Con jerarquía, se mide por frecuencia	Generan tensión emocional y afecta el bienestar del estudiante	Se genera en el colegio, el entorno familiar, dentro de la convivencia del estudiante.
20	Proceso de aprendizaje Colegio	Conjunto de actividades cognitivas, metodológicas y sociales	Nominal: Puede clasificarse en varias categorías, niveles, sin orden jerárquico.	Las categorías permiten alcanzar los objetivos académicos.	Está relacionada con métodos de enseñanza, participación del estudiante, desarrollo de actividades.
21	Asistencia a clases, Horas de estudio, Calificaciones anteriores, Participación en clase, Acceso a recursos educativos, Autoeficacia percibida, Nivel socio económico, Interés en el curso, Horas de estudio		Algunas áreas registrar esta variable en los estudiantes, no se utilizó esta variable predicción, no todos los estudiantes tienen registrado esta variable.		

En la tabla V se reúne las variables consideradas en la investigación, indicando sus características, tipo de medición, codificación y observaciones. Se incluyen variables de identificación y control (código del estudiante, edad, sexo, DNI, año académico y grado-sección), útiles para organizar y validar los registros, aunque sin relación directa con el rendimiento académico, también se incorporan variables académicas, personales, familiares y sociales como promedio de notas, instrucción de los padres, motivación, comunicación, técnicas de aprendizaje, maltrato y discusiones priorizadas por su influencia en el desempeño escolar. Finalmente, se registran variables complementarias (asistencia, horas de estudio, participación, acceso a recursos y nivel socioeconómico) que no se utilizaron en la predicción por carecer de información completa en todos los estudiantes.

.

TABLA VI

CLASIFICACIÓN Y DESCRIPCIÓN DE VARIABLES CONSIDERADAS EN LA INVESTIGACIÓN

N	Variable	Tipo Variable	Descripción	Interpretación
1	Código Estudiante	Cualitativa Nominal: identifica al estudiante sin orden ni magnitud.	Manejo eficiente de la información académica.	Identifica a un estudiante, no es un número para sumar o promediar.
2	Sexo	Cualitativa nominal Dicotómica: Tiene dos categorías, masculino y femenino	Existen diferencias cognitivas, psicológicas, sociales, de comprensión, entre niñas y niños.	Describe una característica biológica del estudiante.
3	Edad	Cuantitativa continua: Puede asumir cualquier valor dentro de un rango y se expresa en números.	Aplicar técnicas de aprendizaje de acuerdo a la edad del estudiante, existe una relación desarrollo cognitivo, emocional, el conocimiento, las experiencias acumuladas con la edad del estudiante	Determina la edad biológica del estudiante, no todos los estudiantes tienen la misma edad en el salón, estan entre 12 a 14 años
4	Grado y sección	Cualitativa ordinal: Existe un orden jerárquico, dentro del grado y sección	Identificar el grado y sección con mayor porcentaje de aprobados y desaprobados, con la finalidad de aplicar estrategias adecuadas, precisas y efectivas para disminuir el porcentaje desaprobados y aumentar porcentaje de aprobados, determinar las necesidades del grado y sección.	Determina el grado y sección del estudiante, de acuerdo a la organización del colegio.
5	DNI	Cualitativa Nominal: solo sirve para identificar, no indica cantidad	Vincular múltiples fuentes de datos, identificar, analizar el historial académico del estudiante	Numeración que identifica a una persona, Identificación Única sin orden, ni valor cuantitativo.
6	Año académico	Cuantitativa discreta: Se expresa en números enteros	Identificar las estrategias planteadas en los años académicos de mayor porcentaje de desaprobados o de aprobados, identificar las necesidades específicas y el diseño de intervenciones educativas adecuadas.	Permitió establecer un orden cronológico, pero sin realizar operaciones matemáticas significativas, mayor o menor entre años académicos.
7	Nota Promedio	Cuantitativa continua: Toma valores numéricos dentro del rango de notas de 0 a 20.	Variable que se desea predecir, determinar la evaluación del estudiante, el desempeño del estudiante, cuantificar el rendimiento académico del estudiante.	Determina la nota promedio del estudiante registrado por el docente.
8	Familiar que representa	Cualitativa nominal: Categórica: Indica la categoría del familiar responsable, sin orden jerárquico.	Frecuencia con que acude el familiar al colegio para verificar el rendimiento académico del estudiante, refleja el apoyo emocional, educativo, socio económico y cultural del estudiante.	Identifica el familiar que representa al estudiante dentro del colegio. identifica la relación del representante, sin tener orden.
9	Grado Instrucción Padre	Cualitativa Ordinal: Representa niveles educativos, con orden jerárquico	Proporciona un ambiente mas favorable para el aprendizaje del estudiante, por ayuda del padre, aspectos del entorno educativo y socio económico del estudiante.,	Determina el grado instrucción del padre actualmente. Nivel educativo en un orden Jerárquico: primaria, secundaria. bajo, medio, alto

			impacto significativo del estudiante, aunque la madre este ausente por temas laborales, personales.	
10	Grado Instrucción Madre	Cualitativa Ordinal: Representa niveles educativos, con orden jerárquico	Proporciona un ambiente más favorable para el aprendizaje del estudiante, por ayuda de la Madre, aspectos del entorno educativo y socio económico del estudiante., impacto significativo del estudiante, aunque la madre este ausente por temas laborales, personales.	Determina el grado instrucción de la Madre actualmente. Nivel educativo en un orden Jerárquico: primaria, secundaria. bajo, medio, alto
11	Desinterés en el rendimiento académico	Cualitativa Ordinal: Suele medirse en categorías o niveles, lo cual implica un orden.	Refleja la motivación, los hábitos de estudio, participación en clase, la actitud del estudiante hacia el aprendizaje, permite mejorar el rendimiento académico del estudiante conociendo el interés o desinterés del estudiante.	Identifica el interés que tiene el estudiante en el rendimiento académico.
12	Motivación Rendimiento académico.	Cualitativa Ordinal: Mide niveles de motivación en categorías o niveles.	Determina el comportamiento, las actitudes, el compromiso, las metas, hábitos de estudio del estudiante, evaluación mas completa y precisa del rendimiento académico	Identifica la motivación del rendimiento académico que tiene el estudiante en el rendimiento académico.
13	Comunicación con el Estudiante	Cualitativa Ordinal: Categorías distintas sin orden.	Refleja la calidad, frecuencia, entre el docente, familiar que representa, y el estudiante. Facilita un entorno de aprendizaje positivo, para mejorar el rendimiento académico del estudiante.	Identifica la comunicación que tienen, los docentes, personal administrativo, personal auxiliar, oficiales con el estudiante.
14	Exceso de Alcohol.	Cualitativa Ordinal: Nivel de frecuencia o cantidad, con orden jerárquico	Refleja los variados efectos sobre la salud física, mental, el comportamiento tanto de los familiares que representa, de los estudiantes.	Identifica el exceso de alcohol del familiar que representa, familiares que conviven con el estudiante y que reflejen directa e indirectamente su rendimiento académico.
15	Técnicas de Aprendizaje	Cualitativa Ordinal: Nivel de frecuencia o cantidad, sin orden jerárquico	Métodos de los estudiantes para aprender y generar conocimiento, impacto significativo en la capacidad para comprender y retener información, mejorar las técnicas de aprendizaje para mejorar el desempeño académico y las relaciones interpersonales de los estudiantes.	Demuestra todas las técnicas de aprendizaje utilizadas tanto de los familiares, familiar que representa del estudiante y que reflejan su rendimiento académico.
16	Maltrato físico.	Cualitativa Ordinal: Nivel de frecuencia o cantidad, con orden jerárquico	Mostrar los efectos negativos de la salud física, mental, emocional y la conducta del estudiante, promover la seguridad y bienestar de los estudiantes, para mejorar el desempeño académico y la calidad de vida.	Identifica el maltrato físico del familiar que representa, familiares, efectos negativos de la salud física, efectos que refleja en su rendimiento académico del estudiante
17	Maltrato Psicológico	Cualitativa Ordinal: Categórica, indica abuso psicológico de los padres al estudiante.	Mostrar los efectos negativos, profundos, ya sea a corto o largo plazo, deterioro de la salud mental, emocional y la conducta del estudiante, mejorar las relaciones interpersonales,	Identifica el maltrato Psicológico del familiar que representa, familiares, efectos negativos de la salud emocional y psicológica, que refleja en su rendimiento académico. del estudiante

			bienestar emocional y psicológico, mejora su desempeño académico y su calidad de vida.	
18	Padres separados.	Cualitativa Ordinal: Categórica, Indica estado separación padres, sin orden	Muestra los impactos generados por la separación o divorcio de los padres de familia, efectos significativos en la salud emocional, psicológica, el cambio de ambiente de estudio, influye en su desempeño académico del estudiante.	Determina los efectos negativos cuando los padres estan separados y lo que refleja en su rendimiento académico del estudiante.
19	Discusiones	Cualitativa ordinal: Categórica, que indica problemas o discusiones existentes entre los padres y el estudiante	Muestra los impactos directos o indirectos de los conflictos y las tensiones del hogar, como influye en la vida de los estudiantes, brindar apoyo psicológico al estudiante en situaciones familiares conflictivas, mejorar su desempeño académico y bienestar familiar.	Identifica los efectos negativos que se generan las discusiones dentro de su familia, representante familiar, conflictos que interfieren en el rendimiento académicos del estudiante.
20	Proceso Aprendizaje en el colegio	Categórica Ordinal: Categórica que indica calidad percibida del proceso de aprendizaje.	permite evaluar cómo los factores académicos, pedagógicos y contextuales influyen en el rendimiento y desarrollo del estudiante.	Determina el proceso aprendizaje dentro del colegio, en su familia, representante familiar.
21	Asistencia a clases, Horas de estudio, Calificaciones anteriores, Participación en clase, Acceso a recursos educativos, Autoeficacia percibida, Nivel socio económico, Interés en el curso, Horas de estudio		Algunas áreas registrar esta variable en los estudiantes, no se utilizó esta variable predicción, no todos los estudiantes tienen registrado esta variable.	

En la Tabla VI se presentan las variables incluidas en la investigación, clasificadas en cualitativas y cuantitativas, junto con su definición, medición e interpretación. Se incorporan variables de identificación y control (código del estudiante, sexo, edad, DNI, año académico y grado-sección), que permiten organizar los registros sin influir en el rendimiento académico. Asimismo, se consideran variables académicas, personales, familiares y sociales (promedio de notas, nivel educativo de los padres, motivación, comunicación, técnicas de aprendizaje, maltrato, separación y discusiones familiares), priorizadas por su incidencia en el desempeño escolar. Finalmente, se identifican variables complementarias (asistencia, participación, horas de estudio, acceso a recursos y nivel socioeconómico) que no fueron usadas en la predicción por falta de registros completos.

TABLA VII

RESUMEN DE VARIABLES SELECCIONADAS PARA EL ESTUDIO

N	Variable	Tipo Dato	Categorías	Importancia
1	Código estudiante	Cadena, string	1	Se seleccionó la variable porque facilita la identificación, el seguimiento y la gestión eficiente de la información académica, permitiendo análisis más precisos y completos. Esto favorece una toma de decisiones informada y la aplicación de estrategias educativas efectivas
2	Edad	Entero, integer	1	Se seleccionó la variable porque los reportes evidencian los porcentajes de aprobados y desaprobados según el rango de edad, lo que posibilita aplicar técnicas de aprendizaje acordes a cada etapa. La institución reconoce que la edad influye en el desarrollo cognitivo, emocional y en la acumulación de experiencias educativas
3	Sexo	Cadena, string	2	Se seleccionó la variable porque los reportes reflejan los porcentajes de aprobados y desaprobados según el sexo del estudiante, lo que permite diseñar estrategias para reducir la desaprobación en ambos grupos. La institución reconoce diferencias cognitivas, psicológicas, sociales y contextuales, lo que facilita intervenciones educativas más efectivas y adaptadas a sus necesidades.
4	DNI	Cadena, string	1	Se seleccionó la variable porque permite identificar el historial académico del estudiante, optimizando la gestión y el análisis de datos educativos. Asimismo, facilita la integración de múltiples fuentes de información, asegurando la calidad y precisión de los datos, y ofreciendo una visión integral y detallada del desempeño académico en distintos contextos
5	Año académico	Fecha, date	5	Se seleccionó la variable porque los reportes evidencian los porcentajes de aprobados y desaprobados según el año académico, lo que permite analizar las estrategias aplicadas en los periodos con mayores índices de aprobación o desaprobación. Asimismo, ofrece un marco para comprender la complejidad del currículo, el desarrollo cognitivo y emocional de los estudiantes y las expectativas de evaluación, facilitando una valoración más precisa y contextualizada del rendimiento, así como la identificación de necesidades y el diseño de intervenciones educativas adecuadas.
6	Grado-sección	Cadena, string	6	Se seleccionó la variable porque permite identificar el grado y la sección con mayor porcentaje de aprobados y desaprobados, lo que facilita la implementación de estrategias para reducir la desaprobación y aumentar la aprobación. Ofrece una visión detallada del contexto educativo, permite una evaluación precisa del rendimiento, la detección de necesidades específicas y la comprensión del impacto de las diferencias entre grados y secciones en el desempeño académico. Esto contribuye al diseño de intervenciones más efectivas para apoyar a todos los estudiantes
7	Nota Promedio del curso	Entero, integer	1	Se seleccionó la variable porque permite evaluar de forma continua el desempeño del estudiante mediante el registro de notas a lo largo del año académico, verificando su progreso en asignaturas específicas. Proporciona una medida directa y cuantificable del rendimiento, lo que favorece el ajuste de estrategias educativas personalizadas y efectivas, impulsando el desarrollo académico integral

8	Familiar que representa	Cadena, string	15	Se seleccionó la variable porque los reportes registran la frecuencia con la que el familiar responsable asiste a las reuniones convocadas por el colegio para informar sobre el rendimiento académico. La participación de este familiar —padre, madre, hermano mayor, tío, primo u otro representante legal— refleja el nivel de apoyo emocional, educativo, socioeconómico y cultural brindado al estudiante
9	Grado de instrucción del padre,	Cadena, string	13	Se seleccionó la variable porque refleja aspectos clave del entorno educativo y socioeconómico del estudiante, permitiendo una evaluación más completa y precisa de su contexto. Contribuye al desarrollo integral del alumno, dado que los padres con mayor nivel de instrucción suelen ofrecer un entorno más favorable para el aprendizaje, influyendo significativamente en el rendimiento académico, incluso en situaciones donde el padre está ausente por motivos laborales o personales
10	Grado de instrucción de la madre	Cadena, string	13	Se seleccionó la variable porque evidencia factores del entorno educativo y socioeconómico del estudiante, facilitando una evaluación más completa y precisa de su realidad. Las madres con mayor nivel de instrucción suelen proporcionar un ambiente propicio para el aprendizaje, lo que impacta de manera significativa en el rendimiento académico, incluso cuando la madre está ausente por motivos laborales o personales
11	Desinterés del Rendimiento académico	Cadena, String	15	Se seleccionó la variable porque refleja el nivel de interés o desinterés del estudiante hacia su desempeño académico, lo que influye en su motivación, hábitos de estudio, participación en clase y actitud hacia el aprendizaje. Su análisis permite diseñar estrategias que fomenten la motivación y mejoren el rendimiento académico. Asimismo, el interés familiar evidencia el grado de involucramiento del apoderado en el seguimiento del desempeño y las actividades escolares.
12	Motivación del Rendimiento académico	Cadena, String	15	Se seleccionó la variable porque refleja el comportamiento y las actitudes del estudiante hacia el aprendizaje, influyendo en su compromiso, metas, hábitos de estudio y conducta en el aula. Su análisis permite evaluar el rendimiento académico de manera integral, identificar necesidades específicas y diseñar intervenciones que fortalezcan la motivación para mejorar el desempeño
13	Comunicación con el estudiante	Cadena, String	15	Se seleccionó la variable porque representa la calidad y frecuencia de la comunicación entre estudiantes, docentes, padres y compañeros, influyendo significativamente en el aprendizaje. Una interacción efectiva fomenta apoyo, retroalimentación, motivación y desarrollo socioemocional, creando un entorno positivo que mejora el rendimiento académico, facilita la evaluación del desempeño e identifica necesidades para diseñar intervenciones que promuevan una comunicación abierta y constructiva.
14	Exceso de Alcohol	Cadena, String	15	Se seleccionó la variable porque impacta la salud física y mental, los hábitos de estudio, la calidad del sueño, las funciones cognitivas, el comportamiento y la disciplina, así como las interacciones sociales y académicas. También está asociada a riesgos académicos y legales. Su análisis permite fomentar estilos de vida saludables que contribuyan a mejorar el desempeño académico
15	Técnicas de aprendizaje	Cadena, String	15	Se seleccionó la variable porque refleja las estrategias y métodos que los estudiantes emplean para aprender, influyendo en la comprensión, retención de información y

				eficiencia en el estudio. También incide en el desarrollo metacognitivo, la motivación, la actitud y el aprendizaje a largo plazo. Su análisis promueve el uso de técnicas efectivas, mejorando el rendimiento académico y las relaciones interpersonales
16	Maltrato físico	Cadena, String	15	Se seleccionó la variable porque evidencia los efectos negativos en la salud física, mental y emocional del estudiante, así como en su autoestima, comportamiento, aprendizaje y relaciones interpersonales. Su análisis permite promover la seguridad y el bienestar estudiantil, mejorando el desempeño académico y la calidad de vida
17	Maltrato psicológico	Cadena, String	15	Se seleccionó la variable porque muestra los efectos profundos y duraderos en la salud mental, emocional y conductual del estudiante, afectando su autoestima, aprendizaje, comportamiento y relaciones interpersonales. Su análisis permite impulsar el bienestar psicológico y emocional, contribuyendo a mejorar el rendimiento académico y la calidad de vida
18	Padres separados,	Cadena, String	15	Se seleccionó la variable porque evidencia el impacto que la separación o divorcio parental puede generar en la vida del estudiante, afectando su salud emocional, dinámica familiar, recursos socioeconómicos, comportamiento y relaciones interpersonales. Su análisis permite brindar apoyo en contextos familiares difíciles, favoreciendo el bienestar integral y el rendimiento académico
19	Discusiones,	Cadena, String	15	Se seleccionó la variable porque muestra el impacto directo e indirecto de los conflictos familiares en la salud emocional, el ambiente de estudio, la conducta, las relaciones interpersonales y el bienestar general del estudiante. Además, permite identificar situaciones familiares adversas para ofrecer apoyo oportuno, mejorando su desempeño académico y bienestar integral.
20	Proceso de Aprendizaje en el colegio	Cadena, String	15	Se seleccionó la variable porque integra factores esenciales de la enseñanza, como el ambiente escolar, las metodologías, el currículo, los recursos, la evaluación, la atención individualizada y la participación estudiantil. Estos elementos influyen directamente en la calidad educativa y, por ende, en el rendimiento académico.
21	Asistencia a clases, Horas de estudio, Calificaciones anteriores, Participación en clase, Acceso a recursos educativos, Autoeficacia percibida, Nivel socio económico, Interés en el curso, Horas de estudio		Algunas áreas registrar esta variable en los estudiantes, no se utilizó esta variable predicción, no todos los estudiantes tienen registrado esta variable.	

En la tabla VII muestra las variables consideradas en el estudio, organizadas por tipo de dato, categorías e importancia. Incluye datos de identificación, académicos, familiares, sociales y personales que influyen en el rendimiento escolar, priorizando aquellas con mayor impacto en el análisis. También se registran variables complementarias que no fueron utilizadas en la predicción debido a la falta de información completa en todos los estudiantes.

b. Determinar cuál es el algoritmo que nos proporciona la mayor precisión al predecir el rendimiento académico en el curso de Matemática de 3er año de secundaria.

TABLA VIII

TESIS REVISADAS Y TÉCNICAS DE PREDICCIÓN UTILIZADA

N	Titulo Investigación	Autor/autores	Tipo Algoritmo	Algoritmos	Detalles
1	Rendimiento académico de estudiantes en educación superior: predicciones de factores influyentes a partir de árboles de decisión	Diaz landa, Brenda Nicole; melean romero, Rosana Alejandra; marín-rodriguez, William Joel	Algoritmos de clasificación predictiva	Árboles de Decision	Cuando la predicción es de categorías o clases aprobar/reprobar, alto/bajo rendimiento
2	Sistema web basado en técnicas de machine learning para predecir el desempeño de los docentes en el nivel de educación superior universitaria	Chatila Vargas Toscanelli, Habana Sabrá	Algoritmos Basados en Redes Neuronales	Redes Neuronales Artificiales (ANN)	Útiles cuando hay grandes volúmenes de datos y relaciones complejas.
3	Modelo de machine learning para la clasificación de estudiantes de acuerdo a su rendimiento académico en el centro de idiomas de la universidad nacional del santa	Espinoza airac, gianira Xiomara; león muñoz, eduar Fabián	Algoritmos Basados en Redes Neuronales (Deep Learning)	Redes Neuronales Artificiales	Predicción numérica, clasificación, análisis de patrones.
4	Predicción de Rendimiento Académico de alumnos usando Machine Learning	Gutiérrez Villaverde, Herberth; Linares Barbero, Manuela; Agüero Correa, Angel; Pérez Nuñe, Jhelly	Algoritmos de Regresión (Predicción de valores numéricos)	Regresión Lineal Regresión Ridge / Lasso / Elastic Net	Permite predecir valores numéricos

En la Tabla VIII se presentan las cuatro principales tesis analizadas, con sus respectivos autores, el tipo de algoritmo empleado, el algoritmo aplicado y la justificación de su selección. Asimismo, en la Iteración N.º 1 – Fase IV: Modelización de Datos de la metodología CRISP-DM, basada en el planteamiento de Gutiérrez et al. [11], se evaluaron diversos algoritmos de regresión para identificar aquellos con mayor precisión en la predicción del rendimiento académico, considerando la estructura de los datos históricos generados por los sistemas administrativos y transaccionales, las variables definidas.

TABLA IX

IDENTIFICACION ALGORITMOS REGRESION [18]

Algoritmo	Importancia	Justificación
Regresión Lineal Simple	Busca la Relación de una variable dependiente y una o más variables independientes	No se utilizó este algoritmo porque se determinó 19 variables independientes y una variable dependiente o de predicción
Regresión Lineal Múltiple	Busca las relaciones existentes entre múltiples variables independientes con la variable dependiente.	Si se utilizó este algoritmo porque se busca determinar cómo las variables independientes influyen en el rendimiento académico que es la nota del estudiante, ósea la variable dependiente.
Regresion Ridge o Ridge Regression.	Busca reducir la relación que existe entre las variables independientes con respecto a la variable dependiente	No se utilizó este algoritmo porque no se busca reducir las relaciones que existen entre las variables independientes y la variable dependiente.
Regresion Lasso o Lasso regression.	Busca establecer la relación en forma automática entre las variables independiente y dependiente.	No se utilizó este algoritmo porque la selección de variables tanto independientes como dependientes es de acuerdo a la estructura de los datos generados en el colegio.
Regresion Elastic Net	Busca reducir y seleccionar las variables independientes y dependientes.	No se utilizó este algoritmo porque no se busca ni reducir las relaciones de las variables ni que la selección sea de manera automática.
Support Vector Regression (SVR)	Busca dividir las variables independientes en clases o categorías para poder relacionarla con la variable dependiente.	No se utilizó este algoritmo porque no se busca clasificar a las variables independientes.
Árbol de decisión para regresión o Regresor de Árbol de decisión	Busca modelar la relación que existe entre las variables independientes y la variable dependiente.	Si se utilizó este algoritmo porque se desea saber o modelar la relación que existe entre las variables independientes y la variable dependiente.
Random Forest Regresor o Clasificador de bosque aleatorio	Busca establecer la relación existente entre una variable independiente y la variable dependiente, de acuerdo a la relación probada anteriormente	Si se utilizó este algoritmo porque se busca establecer las relaciones existentes entre cada una de las variables independientes con relación a la variable dependiente.

En la tabla IX sistematiza los algoritmos de regresión evaluados en la investigación, destacando su importancia y la justificación de uso o descarte. Se utilizaron la Regresión Lineal Múltiple, el Árbol de Decisión para Regresión y el Random Forest Regressor, debido a su capacidad para modelar y explicar la relación entre múltiples variables independientes y el rendimiento académico (variable dependiente). En contraste, algoritmos como la Regresión Lineal Simple, Ridge, Lasso, Elastic Net y Support Vector Regression fueron descartados, ya que no se ajustaban a los objetivos analíticos ni a la estructura de los datos del estudio, donde la selección de variables se realizó de manera intencional y no automática.

Resultado obtenido de los algoritmos de entrenamiento

Porcentajes de Desempeño – Algoritmo Entrenamiento.

TABLA X

RESULTADOS COMPARATIVOS DE LOS ALGORITMOS DE ENTRENAMIENTO

Algoritmo Predicción	Porcentaje Desempeño.
Regresión Lineal	78.83%
Clasificador Bosque Aleatorio	86.34%
Regresor de Árbol de Decisión	84.88%

En la tabla X muestra los resultados comparativos de los algoritmos aplicados para la predicción del rendimiento académico. El algoritmo Regresión Lineal obtuvo un rendimiento del 78.83%, el algoritmo Regresor de Árbol de Decisión obtuvo un rendimiento del 84.88%. El mejor rendimiento fue el algoritmo Clasificador Bosque Aleatorio, que logró un porcentaje del 86.34%, consolidándose como el algoritmo más eficiente para realizar el modelo de predicción de los nuevos estudiantes.

TABLA XI

METRICAS EVALUACION DE LOS ALGORITMOS DE PREDICCION

Algoritmos	ECM	Raíz del ECM	Índice de concordancia
Regresión Lineal	Tiene un error medio cuadrático 14.50 puntos de los datos históricos registrados en escala de 0 a 20, teniendo un error exagerado.	Tiene un error promedio de 3.76 puntos, es decir el entrenamiento realizado difiere de la nota real por 3.76 puntos, es decir difiere mucho.	El valor de 0.9999 significa que el 99.999% de las veces el algoritmo de entrenamiento mantiene el orden relativo entre las notas reales y las notas de entrenamiento.
Clasificador de Bosque Aleatorio-Naive bayes gaussiano	El algoritmo entrenado en promedio se equivoca al predecir por una magnitud de 8.49 puntos, se considera moderado o no tan graves	Tiene un error promedio de 2.91 puntos, es decir el entrenamiento realizado difiere de la nota real por 2.91 puntos, se considera aceptable toma decisiones	El valor de 0.9999 significa que el 99.999% de las veces el algoritmo de entrenamiento mantiene el orden relativo entre las notas reales y las notas de entrenamiento.
Regresor de Árbol de Decisión	El algoritmo entrenado en promedio se equivoca al predecir por una magnitud de 8.49 puntos, se considera moderado o no tan graves	Tiene un error promedio de 2.91 puntos, es decir el entrenamiento realizado difiere de la nota real por 2.91 puntos, se considera aceptable toma decisiones	El valor de 0.9999 significa que el 99.999% de las veces el algoritmo de entrenamiento mantiene el orden relativo entre las notas reales y las notas de entrenamiento.

En la tabla XI, se realizó el análisis comparativo de los algoritmos aplicados, se concluye que el algoritmo de Regresión Lineal presentó un desempeño deficiente debido a su elevado error (ECM = 14.50; Raíz ECM = 3.76), lo que limita su utilidad en la predicción del rendimiento académico. En cambio, el algoritmo Clasificador de Bosque Aleatorio con el algoritmo Naive

Bayes Gaussiano **y el** Regresor de Árbol de Decisión mostraron un error considerablemente menor (ECM = 8.49; Raíz ECM = 2.91), lo que los posiciona como modelos más adecuados y aceptables para este contexto. No obstante, todos los algoritmos alcanzaron un índice de concordancia de 0.9999, evidenciando una alta capacidad para mantener la coherencia en el orden de las notas reales y las notas de predicción.

c. **Crear un modelo de predicción en base a los datos recolectados y el algoritmo seleccionado.**

TABLA XII

SÍNTESIS DE ACCIONES EN EL PROCESO DE MODELADO PREDICTIVO

Proceso / acción	Acciones Realizadas
Definir claramente lo que se quiere predecir	Se definió la variable de predicción y el tipo de predicción a realizar.
Contar los años académicos históricos para generar los algoritmos de entrenamiento.	Se Identifico los años históricos que se ha tenido en cuenta para implementar los algoritmos de entrenamiento.
Generar los porcentajes de predicción de cada algoritmo de entrenamiento.	Se determino el porcentaje de predicción de cada algoritmo de entrenamiento y utilizar el algoritmo con mayor porcentaje para generar el modelo de predicción.
Listar los estudiantes del año académico actual para generar el modelo de predicción	Se identifico a todos los estudiantes registrados antes de generar el modelo de predicción.
Realizar análisis exploratorio (detectar datos faltantes, inconsistencias o valores atípico)	Se realizo el análisis exploratorio en las variables utilizadas en los algoritmos de entrenamiento y en las variables utilizadas en el modelo de predicción.
Identificar que las variables en los algoritmos de entrenamiento sean las mismas que las variables en el modelo de predicción	Se identifico que las variables utilizadas en los algoritmos de entrenamiento sean las mismas que las variables utilizadas en el modelo de predicción.
Codificar las variables categóricas seleccionadas	Se codifico todas las variables categóricas seleccionadas para la creación del modelo de predicción
Determinar que los nuevos estudiantes tengan registrado todas las variables seleccionadas	Se determino que los nuevos estudiantes tengan registrado todas las variables seleccionadas en el modelo de predicción.
Seleccionar el algoritmo con mayor porcentaje de predicción para generar el modelo de los nuevos estudiantes.	Se selecciono el algoritmo de entrenamiento con mayor porcentaje de predicción para generar el modelo de los nuevos estudiantes.

En la tabla XII, presenta las acciones realizadas en el proceso de construcción del modelo de predicción del rendimiento académico, que comprendió la definición de la variable objetivo, la selección, la preparación de datos históricos, el análisis exploratorio, la codificación de variables categóricas. Finalmente, se eligió el algoritmo con mayor precisión para generar el modelo aplicado a los estudiantes del año académico actual.

TABLA XIII

ESTRATEGIAS - DEPURAR DATOS HISTORICOS Y ACTUALES

Estrategia	Procedimiento	Importancia
Detección datos Faltantes o Nulos	Se detectaron registros vacíos o nulos y se determinaron que algunos deben ser eliminados o completados mediante imputación con valores similares	Se evito errores durante el entrenamiento de algoritmos de machine learning que no toleran valores nulos, mejoro la calidad del análisis estadístico, asegurando que los resultados sean válidos y confiables y se preservo la integridad del conjunto de datos.
Identificación datos duplicados	Se detectaron registros duplicados y se determinaron que algunos deben ser eliminarlos, algunos deben ser fusionados con la información redundante.	Se redujo el volumen del dataset innecesariamente sin aportar nueva información, se mejoró la calidad y fiabilidad del análisis, al trabajar con datos únicos y representativos, se optimizo el desempeño de los algoritmos.
Corrección de formatos	Se estandarizaron los tipos de datos y se convirtieron a un formato unificado.	Se estandarizaron los datos para que sean compatibles con las herramientas de análisis y modelado, se redujo los errores de procesamiento **y** evitar fallos en la lectura o interpretación de los datos, se facilitó la automatización de tareas y el uso eficiente de técnicas estadísticas y de machine learning.
Identificación de valores atípicos	Se determinaron que algunos datos deben ser corregidos, otros deben ser transformados, otros deben ser eliminados, a través de métodos estadísticos rango intercuartílico (IQR), z-score, boxplots.	Se optimizo el modelo al evitar que este se vea influenciado por datos extremos no representativos, se evitó interpretaciones erróneas en análisis estadísticos, se decidió si el dataset deben ser corregidos, eliminados o analizados por separado.
Validación de rangos esperados	Se verifico que los datos estén en un rango lógico, se corrigió los errores encontrados.	Se detecto los errores de entrada o captura de datos, como notas o edades del estudiante fuera del rango o imposibles, se evitó que datos inválidos afecten el análisis estadístico o el entrenamiento de machine learning
Categorización inconsistente	Se unifico las variables o categorías similares y se corrigió los errores ortográficos encontrados.	Se permitió unificar los valores categóricos**,** asegurando coherencia y mejor interpretación de los datos, Se **e**vito divisiones erróneas en el análisis estadístico, donde una misma categoría es tratada como varias.
Reglas de Negocio	Se aplico rangos o reglas específicas para realizar la validación de los datos, se cruzó los datos entre columnas para determinar consistencia.	Se valido la coherencia interna de los datos con respecto a las condiciones reales del entorno, permitió prevenir errores lógicos o inconsistencias que podrían pasar desapercibidas con validaciones técnicas básicas.
Visualización de Datos	Se mostro los datos a través de gráficos, histogramas, mapas de color, para detectar patrones o errores no evidentes.	Se facilito la exploración y el análisis de grandes volúmenes de información de forma intuitiva, se permitió detectar errores, valores atípicos y comportamientos inesperados en los datos.

Revisión cruzada con fuentes confiables	Se Comparo los datos existentes con otras áreas para verificar su validación	Se aseguro la integridad y calidad de los datos, especialmente en contextos donde se requiere alta precisión, se evitó la propagación de información incorrecta que podría afectar negativamente los resultados del análisis o la toma de decisiones.
Documentar decisiones de limpieza	Se registro las modificaciones realizadas para conocer la versión actualizada	Se facilito la revisión y validación del proceso por otros analistas o investigadores, Se aseguro la coherencia del análisis**,** especialmente en proyectos colaborativos o a largo plazo.

En la Tabla XIII se presentan las estrategias de limpieza y validación de datos, que garantizan la calidad e integridad del conjunto de datos, optimizan la fiabilidad de los resultados, el desempeño de los algoritmos de Machine Learning, facilitando la interpretación y toma de decisiones en la predicción del rendimiento académico.

TABLA XIV

PORCENTAJE ESTRATEGIA – DATOS HISTORICOS

Estrategia	Cantidad	Porcentaje
Detección datos Faltantes o Nulos	973	27.11%
Identificación datos duplicados	2	0.06%
Corrección de formatos	242	6.74%
Identificación de valores atípicos	60	1.67%
Validación de rangos esperados	130	3.62%
Categorización inconsistente	568	15.83%
Reglas de Negocio	135	3.76%
Visualización de Datos	693	19.31%
Revisión cruzada con fuentes confiables	627	17.47%
Documentar decisiones de limpieza	159	4.43%
Totales	**3589**	**100%**

La Tabla XJV presenta el total de registros iniciales de los datos históricos. De los 3,589 registros originales, la aplicación de estrategias de limpieza y validación permitió obtener 818 registros válidos, con todas las variables y categorías necesarias, garantizando la calidad, coherencia e integridad del conjunto de datos para la implementación de los algoritmos de entrenamiento y la selección del algoritmo optimo.

TABLA XV

PORCENTAJE ESTRATEGIA - DATOS ACTUALES

Estrategia	Cantidad	Porcentaje
Detección datos Faltantes o Nulos	336	13.05%
Identificación datos duplicados	481	18.69%
Corrección de formatos	140	5.44%
Identificación de valores atípicos	256	9.95%
Validación de rangos esperados	154	5.98%
Categorización inconsistente	413	16.05%
Reglas de Negocio	454	17.64%
Visualización de Datos	190	7.38%
Revisión cruzada con fuentes confiables	126	4.90%
Documentar decisiones de limpieza	24	0.93%
Totales	2574	**100%**

La Tabla XV presenta el total de registros iniciales de los datos actuales. De los 2,574 registros originales, la aplicación de estrategias de limpieza y validación permitió obtener 180 registros válidos, con todas las variables y categorías necesarias, garantizando la calidad, coherencia e integridad del conjunto de datos para la implementación del modelo de predicción.

TABLA XVI

DESCRIPCIÓN DATOS HISTÓRICOS – ACTUALES – RENDIMIENTO ACADÉMICO

Tipo de Datos	Definición	Registros	Años Académico
Datos Históricos	Registro de estudiantes del curso de matemática del tercer grado de diferentes años académicos.	Entre 800 a 900 registros	5 años académicos anteriores: 2018, 2019, 2020, 2021, 2022
Datos Actuales	Registro de estudiantes del curso de matemática del tercer grado del año académico actual.	Entre 150 a 180 registros	2023

En la tabla XVI presenta los datos utilizados para el análisis del rendimiento académico en matemáticas de tercer grado, destacando los registros históricos (800–900 registros de 2018 a 2022) y los actuales (150–180 registros de 2023), proporcionando una base para la implementación de algoritmos de predicción y el análisis comparativo entre años.

TABLA XVII

DESEMPEÑO ACADÉMICO DE ESTUDIANTES POR GRADO Y AÑO ACADÉMICO (2018-2022)

Grado Académico	Aprobados	Aprobados %	Desaprobados	Desaprobados %	Año Académico	Total
Tercer Grado	100	31.25 %	220	68.75 %		
Cuarto Grado	150	55.56 %	120	44.44 %	2018	870
Quinto Grado	200	71.43 %	80	28.57 %		
Tercer Grado	150	37.60 %	250	62.50 %		
Cuarto Grado	200	66.67 %	100	33.33 %	2019	900
Quinto Grado	150	75.00 %	50	25.00 %		
Tercer Grado	130	34.21 %	250	65.79 %		
Cuarto Grado	180	54.55 %	150	45.45 %	2020	940
Quinto Grado	130	56.52 %	100	43.48 %		
Tercer Grado	80	25.81 %	230	74.19 %		
Cuarto Grado	180	60.00 %	120	40.00 %	2021	860
Quinto Grado	150	60.00 %	100	40.00 %		
Tercer Grado	80	25.00 %	240	75.00 %		
Cuarto Grado	180	60.00 %	120	40.00 %	2022	900
Quinto Grado	200	71.43 %	80	28.57 %		

En la tabla XVII muestra el desempeño académico de los estudiantes de tercer, cuarto y quinto grado entre 2018 y 2022, indicando los porcentajes de aprobación y desaprobación. Se observa que los estudiantes de quinto grado presentan los mayores índices de aprobación, mientras que los de tercer grado registran los más bajos, lo que evidencia variaciones en el rendimiento y resalta la necesidad de estrategias educativas enfocadas en mejorar los resultados en los niveles más afectados.

TABLA XVIII

PROCESO CONVERSION DATOS

Proceso	Características
Limpieza Datos	Se Reemplazo o elimino valores nulos o ausentes, se rectificó valores incorrectos o inconsistentes, se eliminó registros duplicados para evitar sesgos, error o defecto en el problema del estudio.
Transformación de Datos:	Se normalizo la fuente de datos, estableciendo escalas en un rango definido, se realizó ajustes en las fuentes de datos estableciendo estándares para resolver el problema de estudio, se asignó un valor único a cada categoría de una determinada variable, se agrupo las fuentes de datos en variables o categorías respectivas.
Reducción de Dimensionalidad:	Se selecciono las variables o características más importantes, se generó nuevas características o variables que muestren información relevante utilizando las fuentes de datos.
Creación de nuevos atributos	Se combino las variables existentes con las nuevas variables consideradas importantes, se asignó un nuevo formato a las fuentes de datos existentes
Manejo de valores extremos:	Se estableció un rango aceptable a los valores de la fuente de datos seleccionada, se redujo los valores extremos estableciendo rangos predefinidos que solucionen el problema de estudio., reducir los valores extremos
Formateo y estructura:	Se convirtió el tipo de dato de las variables y categorías de las fuentes de datos, se estableció nuevo orden de las columnas, variables o categorías, se modificó el nombre de las variables, categorías de las fuentes de datos iniciales.
Manejo de datos desbalanceados:	Se aumento el número de registros con menor frecuencia, se redujo el número de registros con mayor frecuencia.

Consolidación de datos:	Se unió o combino múltiples fuentes de datos, se reorganizo los datos para facilitar el análisis para la solución del problema planteado.
Codificación de datos temporales:	Convertir el tipo de datos de las variables, categorías que se va a utilizar en un formato estándar, ver las variables que se van a utilizar o crear nuevas variables de acuerdo al problema a resolver.
Filtrado y segmentación:	Excluir datos irrelevantes o atípicos (alejan del problema) según el problema a resolver.
Categorías-Variables	Todas las categorías de las variables fueron codificadas con valores numéricos enteros, ya que el proceso de minería de datos requiere que estas se representen numéricamente. Originalmente, los valores eran de tipo cadena, fecha o carácter, por lo que se asignó un número entero ascendente a cada categoría, comenzando desde 1 hasta el total de categorías por variable

En la tabla XVIII se realizaron procedimientos de limpieza, transformación y reducción de dimensionalidad con el fin de asegurar la consistencia y calidad de los datos. Asimismo, se generaron nuevas categorías, se controlaron valores atípicos, se reestructuró la base de datos y se aplicaron técnicas de balanceo de variables, obteniéndose finalmente un conjunto de datos depurado y adecuado para el análisis académico.

Procedimiento creación del modelo predicción

Para cumplir con este objetivo, se aplicó la Fase IV – Modelización de Datos siguiendo la metodología CRISP-DM. La aplicación web permite seleccionar y validar un primer archivo Excel para entrenar los algoritmos y calcular sus porcentajes de desempeño, y posteriormente seleccionar un segundo archivo Excel para generar el modelo predictivo de nuevos estudiantes utilizando el algoritmo previamente seleccionado.

TABLA XIX

PROCEDIMIENTO ALGORITMO MAYOR PORCENTAJE PREDICCION-MODELO PREDICCION

Pasos	Detalle
Paso 01	Verificar si se muestra los algoritmos de entrenamiento utilizando los datos históricos, para determinar el algoritmo más optimo con la finalidad de mostrar el modelo de predicción, sino se muestra los algoritmos de entrenamiento no se podrá mostrar el modelo de predicción.
Paso 02	Sino se ha generado los algoritmos de entrenamiento, acceder a la opción de la aplicación web algoritmos de entrenamiento, para poder seleccionar el archivo Excel generado de los datos históricos y realizar las validaciones correspondientes.
Paso 03	Verificar que el archivo en formato Excel de datos históricos este actualizado y tenga como estructura todas las variables o características, las categorías de cada variable y la nota final del curso de matemática.
Paso 04	Seleccionar el archivo Excel para que la aplicación web, realice las validaciones correspondientes y registre el archivo Excel.
Paso 05	Verificar si existen mensajes de error generados por la aplicación web al registrar el archivo Excel de los datos históricos.
Paso 06	La aplicación web muestra el entrenamiento de los tres algoritmos de predicción, utilizando el archivo Excel cargado previamente y que contiene todos los datos históricos
Paso 07	Acceder a la opción de la aplicación web, modelo de predicción, para poder seleccionar el archivo Excel generado correspondiente a los datos actuales y realizar las validaciones correspondientes.
Paso 08	Seleccionar el archivo Excel para que la aplicación web, realice las validaciones correspondientes y registre el archivo Excel.
Paso 09	Verificar si existen mensajes de error generados por la aplicación web al registrar el archivo Excel de los datos actuales.
Paso 10	La aplicación web muestra el modelo de predicción de los nuevos estudiantes, utilizando el algoritmo clasificador de bosque aleatorio, adicionalmente muestra la opción o link para descargar

	el archivo Excel de los datos actuales cargados previamente, adicionalmente muestra la opción o link para descargar el archivo Excel de los datos actuales cargados previamente y la opción para eliminar el archivo Excel.

En la tabla XIX describe los pasos para la carga, validación y procesamiento de los archivos Excel en la aplicación web, desde la verificación de los datos históricos y el entrenamiento de los algoritmos hasta la generación del modelo predictivo de los nuevos estudiantes, asegurando la consistencia de la información y ofreciendo opciones de descarga o eliminación de los archivos.

d. Desarrollar la aplicación web como herramienta de interacción.

Para cumplir con este objetivo, se aplicó la Iteración Fase IV – Evaluación siguiendo la metodología CRISP-DM, abarcando dos procesos: la planificación del proyecto y la implementación del desarrollo de la aplicación web, lo que permite garantizar una ejecución eficiente, alineada con los objetivos y dentro de los plazos y recursos disponibles.

TABLA XX

PASOS PLANIFICACIÓN - DESARROLLO DE LA APLICACIÓN WEB - METODOLOGÍA CRISP-DM

Pasos	Proceso	Detalle
Paso 01	Definición del Proyecto	Se definieron el objetivo, alcance y justificación del proyecto en la Fase 01 de la metodología CRISP-DM, enfocada en la comprensión del negocio.
Paso 02	Identificación de los interesados	Los interesados en el proyecto incluyen estudiantes, padres de familia y usuarios de la aplicación web.
Paso 03	Requisitos y especificaciones	En la Fase 02 de CRISP-DM (comprensión de los datos) se definieron entrevistas, encuestas, requisitos funcionales y no funcionales, así como los comportamientos esperados de la aplicación web.
Paso 04	Desarrollo del plan del proyecto:	El cronograma, la asignación de roles, recursos, responsabilidades y el presupuesto para el desarrollo de la aplicación web se definieron en las Fases 01 y 02 de la metodología CRISP-DM, centradas en la comprensión del negocio y de los datos.
Paso 05	Diseño de la arquitectura y la interfaz	La arquitectura de la aplicación web, las tecnologías, los diseños de interfaz, frameworks, bases de datos, lenguajes de programación y prototipos funcionales se definieron en la Fase 04 – Modelado de la metodología CRISP-DM.
Paso 06	Desarrollo de la estructura del trabajo	La asignación y división de tareas, incluidas las subtareas desarrolladas, se definieron en la Fase 05 (evaluación) de la metodología CRISP-DM.

En la tabla XX muestra el resume de los pasos de planificación y desarrollo de la aplicación web según la metodología CRISP-DM, incluyendo la definición del proyecto, identificación de interesados, recopilación de requisitos, planificación de recursos - cronograma, diseño de arquitectura e interfaz, estructuración de tareas, asegurando un desarrollo organizado y alineado con los objetivos del proyecto.

TABLA XXI

FASES - PASOS DESARROLLO, IMPLEMENTACIÓN - MANTENIMIENTO APLICACIÓN WEB

Pasos	**Proceso**	**Detalle**
Paso 01	Preparación y configuración Inicial	La configuración del entorno y la selección de tecnologías para el desarrollo de la aplicación web se definieron en la Fase 04 (modelado) de la metodología CRISP-DM.
Paso 02	Desarrollo del Back-end	El desarrollo de la base de datos, la instalación y configuración de librerías externas para el desarrollo de la aplicación se definieron en la Fase 04 (modelado) de la metodología CRISP-DM.
Paso 03	Desarrollo del Front-end	El diseño de la interfaz de usuario y la implementación de la funcionalidad se definieron en la Fase 04 (modelado) de la metodología CRISP-DM.
Paso 04	Integración continua y control de versiones:	La gestión del código fuente, su integración con las librerías externas y la detección temprana de posibles problemas se definieron en las Fases 04 (modelado) y 06 (despliegue) de la metodología CRISP-DM.
Paso 05	Pruebas y verificación	Las pruebas individuales de cada módulo, la integración de módulos, las validaciones de requisitos funcionales, la mejora de la experiencia del usuario y las pruebas de rendimiento de la aplicación web se definieron en las Fases 05 (evaluación) y 06 (despliegue) de la metodología CRISP-DM.
Paso 06	Despliegue	La configuración de los entornos de producción, las pruebas finales, la documentación necesaria, la implementación de la aplicación web en producción y la supervisión del rendimiento en tiempo real se definieron en las Fases 05 (evaluación) y 06 (despliegue) de la metodología CRISP-DM.
Paso 07	Mantenimiento y actualizaciones	El soporte técnico para resolver consultas de clientes, corregir errores y mejorar el rendimiento en producción se ha definido en las fases 04 (Modelado), 05 (Evaluación) y 06 (Despliegue) de la metodología CRISP-DM.
Paso 08	Documentación y formación	La documentación del código, herramientas, librerías, conexiones, manual de usuario y formación para garantizar el funcionamiento de la aplicación web se ha definido en las fases 05 (Evaluación) y 06 (Despliegue) de la metodología CRISP-DM.
Paso 09	Revisión Pot-Implementacion	La revisión final para evaluar el funcionamiento, identificar mejoras y documentar lecciones aprendidas y mejores prácticas se ha definido en la fase 06 (Despliegue) de la metodología CRISP-DM.

En la tabla XXI resume los pasos del desarrollo, implementación y mantenimiento de la aplicación web bajo la metodología CRISP-DM, incluyendo configuración inicial, desarrollo de back-end-front-end, integración continua, pruebas, despliegue, mantenimiento, documentación, revisión post-implementación, asegurando la eficiencia, calidad y sostenibilidad del proyecto.

e. Validar la aplicación web construida en base a técnicas de pruebas de la ingeniería de software: técnicas, normas, practicas.

Para alcanzar este objetivo, se aplicó la Iteración #6: Fase Deployment – Despliegue siguiendo la metodología CRISP-DM. Adicionalmente, se realizaron encuestas a los usuarios, pruebas de funcionalidad, usabilidad, rendimiento, compatibilidad, con el fin de validar el desarrollo y garantizar el correcto funcionamiento de la aplicación web.

Encuesta para medir la funcionalidad y usabilidad aplicación web: Permite mostrar información importante sobre la experiencia del usuario en el funcionamiento de la aplicación web.

TABLA XXII

VALORACIÓN DE LAS OPCIONES ENCUESTA APLICACIÓN WEB

Valoración	% Calidad	Clasificación	Criterio de Valoración	Conclusión	Explicación
1	25%	Mala	La aplicación web debe ser desarrollado nuevamente.	Rechazado	En desacuerdo
2	50%	Regular	Se deben hacer grandes modificaciones en la aplicación web.	Rechazado	Ni de acuerdo, ni en desacuerdo.
3	75%	Buena	Se necesitan hacer pequeñas modificaciones en la aplicación web.	Aceptado	De acuerdo
4	100%	Excelente	No se requieren hacer modificaciones en la aplicación web.	Aceptado	Muy de acuerdo

En la tabla XXII, resume la valoración de la calidad de la aplicación web, indicando que las clasificaciones “Mala” y “Regular” requieren rechazo o modificaciones importantes, mientras que “Buena” y “Excelente” son aceptadas, con pequeñas o ninguna modificación, asegurando la funcionalidad y calidad de la aplicación.

TABLA XXIII

RESULTADOS DE LA EVALUACIÓN - USABILIDAD Y FUNCIONALIDAD APLICACIÓN WEB

N	Pregunta	Resultados
1	La aplicación web cumple con todos los requerimientos.	4/4
2	La aplicación web siempre hizo lo que yo estaba esperando.	4/4
3	Tiene consistencia interna (uniformidad de las pantallas, menús, reportes, mensajes, opciones, etc.)	4/4
4	Los resultados producidos por la aplicación web son correctos.	4/4
5	Se conecta y opera fácilmente con otro sistema	1/4
6	Los términos usados están estandarizados (por ejemplo. Adicionar, insertar, agregar)	4/4
7	Los símbolos (iconos) usados están estandarizados.	4/4
8	Previene el acceso no autorizado (sea accidental o deliberado) en la aplicación web y datos	4/4
9	Existen controles de acceso adecuado a la aplicación web, subsistemas, funciones, etc.	4/4
10	El uso de la aplicación web es fácil de aprender	4/4

11	Los comandos de la aplicación web se aprenden rápidamente	4/4
12	La información de la aplicación web se presenta de manera clara y entendible	4/4
13	La terminología usada se relaciona bien con el trabajo que hago.	4/4
14	Los mensajes que aparecen en la pantalla son claros.	4/4
15	La organización de los menús o listas de información parece bastante lógica.	4/4
16	Las interfaces o pantallas representan los objetos (por ejemplo, íconos) con análogos en mi ambiente de trabajo	4/4
17	La presentación de la aplicación web es muy atractiva.	4/4
18	Los mensajes de entradas de datos son entendibles	4/4
19	Los mensajes de prevención de error son adecuados	4/4
20	El tipo, tamaño y formato de texto es correcto.	4/4
21	Es fácil moverse de una parte de una tarea a otra.	4/4
22	Los errores de ingreso de datos los puedo corregir fácilmente.	4/4
23	Hay tareas en la aplicación web que requieren de demasiados pasos.	4/4
24	Tengo que buscar ayuda la mayoría de las veces cuando uso la aplicación web.	4/4
25	La información de ayuda dada por la aplicación web es muy útil.	4/4

En la tabla XXIVIII se muestra la evaluación de la aplicación web y evidencia un alto nivel de satisfacción de los usuarios, obteniendo puntajes máximos (4/4) en la mayoría de los criterios relacionados con funcionalidad, usabilidad, consistencia, seguridad y presentación visual. Únicamente el aspecto de conectividad con otros sistemas presentó una valoración baja (1/4), lo que constituye el principal punto de mejora. En conjunto, los resultados reflejan que la aplicación es eficiente, confiable, fácil de usar**,** destacando su claridad, estandarización y atractivo visual.

Pruebas funcionales: Aseguran que la aplicación funcione según lo esperado y cumpla con los requisitos especificados, permitiendo identificar y resolver problemas, antes que sea lanzado a producción.

Aplicaciones Utilizadas: Programas o herramientas para gestionar las pruebas funcionales:

Selenium: Herramienta para la automatización de pruebas funcionales en aplicaciones web.

Cypress: Herramienta para la facilidad de uso y la velocidad.

TestCafe: Herramienta funciona navegador.

Requerimiento funcional o caso de prueba: Registrar Datos Históricos

Descripción: Registrar el archivo actualizado de datos históricos.

Precondición: La página de registrar datos históricos debe estar accesible.

Pasos: Verificar que el archivo de datos históricos sea el actualizado, verificar que la

estructura del archivo Excel tenga las variables, las categorías de cada variable y la nota del estudiante de los datos históricos, hacer clic en el botón cargar datos históricos, verificar que se haya registrado el archivo Excel y se haya generado el entrenamiento de los algoritmos de predicción.

Resultados Esperados: El usuario es redirigido a la página donde se muestra el entrenamiento de los algoritmos de predicción después de haber registrado los datos históricos.

Resultados Obtenidos: El usuario fue redirigido correctamente a la página donde se mostró el entrenamiento de los algoritmos de predicción.

Interfaces Aplicación Web

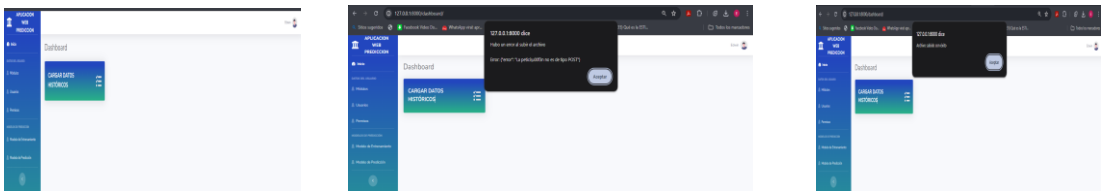


Fig. 2. Caso de Prueba – Registrar Datos Históricos

En las tres interfaces se muestran el proceso que se sigue al registrar el archivo Excel con los datos históricos, primero se muestra la opción de registrar datos históricos, después los errores generados al registrar un archivo con diferente formato o estructura, en la tercera interfaz se muestra mensaje de confirmación al registrar el archivo Excel correctamente.

Requerimiento funcional o caso de prueba: Algoritmos de Entrenamiento.

Descripción: Mostrar los tres algoritmos de entrenamiento desarrollados.

Precondición: La página de algoritmos de entrenamiento debe estar accesible.

Pasos: Click en la opción algoritmo entrenamiento Regresión Lineal, click en la opción algoritmo entrenamiento clasificador de bosque aleatorio, click en la opción algoritmo de entrenamiento regresor de árbol de decisión.

Resultados Esperados: El usuario es redirigido a la opción de cualquiera de los tres algoritmos de entrenamiento que se ha seleccionado.

Resultados Obtenidos: El usuario fue redirigido correctamente a la opción que se ha seleccionado de cualquiera de los tres algoritmos de entrenamiento seleccionado.

Interfaces Aplicación Web.

Fig. 3. Caso de Prueba – Algoritmos de Entrenamiento

En esta interfaz se muestra los tres algoritmos de entrenamiento generados después de registrar el archivo Excel con la estructura de datos, teniendo en cuenta las características o variables, las categorías de cada variable y la nota registrada del estudiante.

Requerimiento funcional o caso de prueba: Registrar Datos actuales.

Descripción: Registrar el archivo actualizado de datos actuales.

Precondición: La página de registrar datos actuales debe estar accesible.

Pasos: Verificar que el archivo de datos actuales sea el actualizado, verificar que la estructura del archivo Excel tenga las variables, las categorías de cada variable y la nota del estudiante de los datos actuales, hacer clic en el botón cargar datos actuales, verificar que se haya registrado el archivo Excel y se haya generado el modelo de predicción utilizando el algoritmo más preciso.

Resultados Esperados: El usuario es redirigido a la página donde se muestra el modelo de predicción después de haber registrado los datos actuales.

Resultados Obtenidos: El usuario fue redirigido correctamente a la página donde se mostró el modelo de predicción de los nuevos estudiantes

Interfaces Aplicación Web

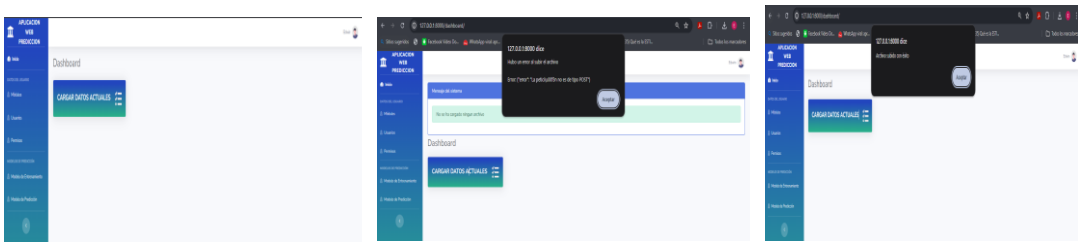


Fig. 4. Caso de Prueba – Registrar Datos Prueba.

En las tres interfaces se muestran el proceso que se sigue al registrar el archivo Excel con los datos actuales, primero se muestra la opción de registrar datos actuales, después los errores generados al registrar un archivo con diferente formato o estructura, en la tercera interfaz se muestra mensaje de confirmación al registrar el archivo Excel correctamente.

Requerimiento funcional o caso de prueba: Modelo de Predicción

Descripción: Mostrar el modelo de predicción utilizando el algoritmo más preciso.

Precondición: La página de modelo de predicción debe estar accesible.

Pasos: Click en la opción Modelo de predicción.

Resultados Esperados: El usuario es redirigido a la opción modelo de predicción utilizando el algoritmo más preciso de los nuevos estudiantes.

Resultados Obtenidos: El usuario fue redirigido correctamente y se visualiza el modelo de predicción.

Pruebas de usabilidad: Aseguran que la aplicación sea fácil de usar, intuitiva, satisfaga las necesidades del usuario final.

Aplicaciones Utilizadas: Ayudan a evaluar la experiencia del usuario.

UserTesting: Permite recopilar comentarios de usuarios reales, mientras interactúas con tu aplicación.

Lookback: Herramienta que permite hacer pruebas de usabilidad em vivo o grabadas.

TestCafe: Herramienta funciona navegador.

HotJar : Herramienta que analiza comportamiento usuarios, grabación de sesiones.

Tarea 1: Registrar Datos Históricos.

Objetivo: Evaluar la facilidad con que el usuario puede registrar o cargar los datos históricos.

Pasos: Navegar a la página o hacer click en la opción de registrar datos históricos, verificar que el archivo de datos históricos sea el actualizado, verificar que la estructura del archivo Excel tenga las variables, las categorías de cada variable y la nota del estudiante de los datos históricos, cargar el archivo de datos históricos correctamente, verificar los mensajes generados.

Métricas: Tiempo para cargar el archivo de datos históricos, numero de errores cometidos al registrar los datos históricos, comentarios verbales de los usuarios al cargar los datos históricos, nnivel de satisfacción al registrar o cargar los datos históricos.

Tarea 2: Algoritmos de Entrenamiento.

Objetivo: Evaluar la rapidez con que el usuario puede generar los tres algoritmos de entrenamiento.

Pasos: Navegar a la página o hacer click en la opción de modelos de entrenamiento, verificar que los algoritmos de entrenamiento sean realizados en forma correcta, verificar que las opciones de cada algoritmo de entrenamiento se muestren en forma

correcta.

Métricas: Tiempo para cargar los algoritmos de entrenamiento realizados, numero de errores cometidos al generar los algoritmos de entrenamiento, comentarios verbales de los usuarios sobre los algoritmos de entrenamiento, nivel de satisfacción al generar los algoritmos de entrenamiento.

Tarea 3: Registrar Datos Actuales.

Objetivo: Evaluar la facilidad con que el usuario puede registrar los datos actuales.

Pasos: Navegar a la página o hacer click en la opción de registrar datos actuales, verificar que el archivo de datos actuales sea el actualizado, verificar que la estructura del archivo Excel tenga las variables, las categorías de cada variable y la nota del estudiante de los datos actuales, cargar el archivo de datos actuales correctamente y verificar los mensajes generados.

Métricas: Tiempo para cargar el modelo de predicción realizados, numero de errores cometidos al generar el modelo de predicción, comentarios verbales de los usuarios sobre el modelo de predicción, nivel de satisfacción al generar el modelo de predicción.

Tarea 4: Modelo de Predicción

Objetivo: Evaluar la facilidad con que el usuario puede generar el modelo de predicción

Pasos: Navegar a la página o hacer click en la opción de modelo de predicción, verificar que el modelo de predicción sea realizado en forma correcta y verificar que las opciones del modelo de predicción se muestren en forma correcta.

Métricas: Tiempo de carga del modelo de predicción, numero de errores generados, comentarios verbales de los usuarios y nivel de satisfacción al generar el modelo.

Pruebas de rendimiento: Determinan funcionamiento o carga de la aplicación web en diferentes niveles de usuarios virtuales.

Aplicaciones Utilizadas: Evalúan la capacidad de respuesta, estabilidad, escalabilidad aplicación bajo diferentes condiciones de carga.

ApacheJmeter: Permite medir el rendimiento de aplicaciones web y otros servicios.

Gatling: Herramienta que permite generar grandes cantidades de tráfico simultaneo.

LoadRunner: Herramienta que mide el comportamiento aplicación bajo carga.

Optimizar consultas a la base de datos, ampliar CPU y memoria del servidor e implementar caché para reducir la carga.

la aplicación web soporta 100 usuarios concurrentes con un tiempo de respuesta

aceptable, no obstante, es necesario optimizar el rendimiento para mejorar estabilidad y capacidad de respuesta en picos de carga.

Prueba Carga: Aplicación web funciona con 100 usuarios conectados al mismo tiempo.

Prueba Estrés: Aplicación web falla con 101 usuarios conectados al mismo tiempo.

Prueba capacidad: Aplicación web falla con 101 usuarios conectados tratando de subir el archivo Excel de datos históricos o de datos actuales al mismo tiempo

Prueba de resistencia: La aplicación web en un periodo de 3 horas constantes de funcionamiento, falla con 101 usuarios conectados, tratando de subir el archivo Excel con datos históricos y el archivo Excel de los estudiantes nuevos al mismo tiempo.

Pruebas de compatibilidad: Garantiza el correcto funcionamiento de la aplicación web y una experiencia de usuario consistente en todos los entornos.

Aplicaciones Utilizadas: .

BrowserStack: Herramienta que permite probar aplicación web en diferentes navegadores y dispositivos reales.

Sauce Labs : Herramienta que permite acceder a varios navegadores y dispositivos.

CrossBrowserTesting: Herramienta que permite realizar pruebas en diferentes navegadores y dispositivos.

TABLA XXV.

NAVEGADORES WEB - SISTEMAS OPERATIVOS

Navegadores Web Probados	**Resultados**	**Sistemas Operativos probados**	**Resultados**
Google Chrome:	Aplicación web compatible, sin errores	PC (Windows)	Funciona correctamente, sin errores de diseño ni funcionalidad.
Firefox	Aplicación web compatible, sin errores.	Mac (macOS)	Funciona correctamente, sin errores de diseño ni funcionalidad
Safari:	Aplicación web no compatible, presenta errores	iPhone (iOS):	Presenta errores en diseño y funcionalidad
Edge	Aplicación web no compatible, presenta errores	Android Phone, (Android)	Presenta errores en diseño y funcionalidad
Opera	Aplicación web no compatible, presenta errores	iPad (iOS):	Presenta errores en diseño y funcionalidad
		Android Tablet (Android):	Presenta errores en diseño y funcionalidad

En la Tabla XXV se evidencia que la aplicación web es plenamente compatible en Google Chrome y Firefox sobre Windows y macOS, sin presentar errores de diseño ni funcionalidad. No obstante, se identificaron incompatibilidades y fallas en otros navegadores

(Safari, Edge y Opera**)** y en sistemas móviles **(**iOS y Android**)**. Se recomienda optimizar el diseño responsive en dichos entornos para garantizar una experiencia de uso óptima.

2. Discusión

El producto acreditable es una aplicación web basada en machine learning, diseñada para predecir el rendimiento académico de estudiantes de tercer grado en matemáticas y desarrollada para cumplir los objetivos establecidos. La herramienta procesa las características o variables ofreciendo una solución innovadora y poco utilizada en el ámbito educativo.

Según los autores "Gutiérrez *et al*. [11] y el autor Páez [19] coinciden con esta investigación en desarrollar un modelo para predecir el rendimiento académico de los estudiantes mediante algoritmos de machine learning. No obstante, esta investigación se diferencia en desarrollar una aplicación web interactiva que permite mostrar el entrenamiento de tres algoritmos de predicción con la finalidad de generar un modelo de predicción utilizando el algoritmo más preciso de los nuevos estudiantes del curso de matemática del tercer grado secundaria.

En las investigaciones realizadas por el autor Ocaña [20], y el autor Ruiz [9], coincide en que las variables que influyen en el rendimiento académico permiten generar modelos para predecir el desempeño estudiantil, identificar a los estudiantes en riesgo de desaprobar y a quienes pueden mejorar su rendimiento. Sin embargo, el principal problema señalado por los autores es el alto costo de implementar un servidor web remoto para desarrollar la aplicación web. A diferencia de ellos, esta investigación utilizó un servidor web local para implementar la aplicación y mostrar las predicciones."

3. Producto Acreditable

a. Interfaces.

Fig. 5. Framework Bootstrap

Se utilizó el framework Bootstrap 5.0 para diseñar las interfaces de la aplicación web, integrándolo con Python. Las tecnologías empleadas incluyen HTML5, CSS, JavaScript y JQuery para funcionalidades específicas. El proceso de carga, instalación, configuración e integración de estas tecnologías se detalla en la fase 04 del modelado según la metodología CRISP-DM.

b. Arquitectura

Fig. 6. Requerimientos funcionales y no funcionales

Se diseñó la arquitectura, el desarrollo y la codificación, definiendo los requerimientos funcionales, no funcionales y los módulos de desarrollo para garantizar el correcto funcionamiento, según lo detallado en la fase 04 del modelado utilizando la metodología CRISP-DM.

c. Infraestructura Tecnológica.

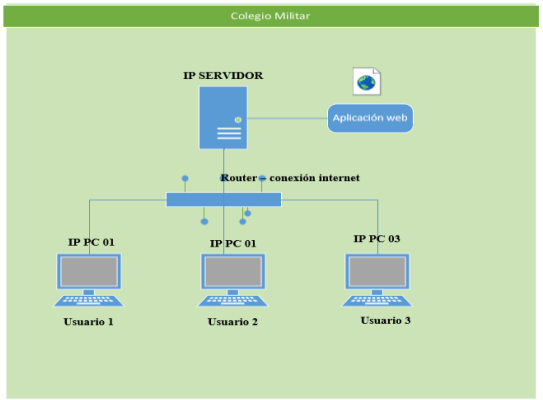


Fig. 7. Infraestructura Tecnológica

Hardware

PC de escritorio: Permite el acceso de los usuarios registrados a la aplicación web.

Router: Dispositivo que proporciona conexión a internet por cable o WiFi.

Internet: Servicio indispensable para que los usuarios utilicen la aplicación web

Servidor back-end: Equipo donde se aloja y ejecuta la funcionalidad de la aplicación web.

Software

Arquitectura de la aplicación web: Se diseñó definiendo la estructura y relación de sus componentes para garantizar su correcto funcionamiento.

Conectividad: Los terminales de los usuarios (PCs) se conectan a través de internet desde el colegio y acceden a la aplicación web mediante la dirección IP del servidor.

Servidor back-end: Aquí se instaló la aplicación web, no requieren licencias de uso y se utilizó las siguientes aplicaciones:

Bootstrap: Para el diseño de la interfaz de usuario.

Python: Como lenguaje de programación principal.

MySQL: Para la gestión de bases de datos.

Librerías externas: Para funcionalidades adicionales de la aplicación web.

4. **Impactos Esperados**

a. **Impacto ético:** Se garantizo el uso responsable de los datos personales ya sea para fines educativos y de mejora académica. se estableció la transparencia y equidad de los datos evitando sesgos o discriminación entre estudiantes, atreves de las capacitaciones en el uso de la aplicación web se explicó los resultados obtenidos, se promovió el uso correcto de herramientas tecnológicas para disminuir la cantidad de estudiantes desaprobados.

b.

c. **Impacto de seguridad:**

d. **Impacto Institucional:** Fortalece la gestión institucional mediante la toma de decisiones basada en datos, optimiza los recursos humanos y materiales enfocados en los estudiantes de bajo rendimiento académico, mejora la imagen institucional, calidad de servicio al utilizar nuevas tecnologías para disminuir la cantidad de estudiantes desaprobados.

Conclusiones

En el análisis de las variables, familiar que representa, grado instrucción del padre, grado instrucción de la madre, desinterés del rendimiento académico, motivación rendimiento académico, comunicación con el estudiante, exceso de alcohol, técnicas de aprendizaje, maltrato físico, maltrato psicológico, padres separados, discusiones, proceso de aprendizaje en el colegio, se identificó que estas son las más precisas para predecir el rendimiento académico

Entre los algoritmos implementados, regresión lineal, boque aleatorio y árbol de decisión, se demostró que el algoritmo con mayor precisión es el bosque aleatorio, que permite predecir el rendimiento académico.

Se demostró que el algoritmo clasificador de bosque aleatorio es el más preciso, y se utilizó para generar el modelo de predicción que mejora el rendimiento académico de los nuevos estudiantes.

Se desarrolló la aplicación web como herramienta tecnológica de interacción, basada en aprendizaje automático, que incluye algoritmos de entrenamiento y un modelo de predicción para evaluar el rendimiento académico de los nuevos estudiantes. Además, se alineó con los requerimientos funcionales y no funcionales, siguiendo la metodología CRISP-DM.

La aplicación web desarrollada ha sido validada mediante pruebas exhaustivas, que incluyen pruebas de caja negra, caja blanca, análisis de riesgos y contingencias, un plan de pruebas, instrumentos para medir funcionalidad y usabilidad, y cumplimiento de las normas ISO/IEC 25010.

Recomendaciones

Para optimizar la predicción del rendimiento académico, se propone un seguimiento detallado de las variables del modelo, incorporando nuevas variables relevantes y ajustando las más influyentes, mediante un análisis de sensibilidad, precisión y margen de error para evaluar cómo los cambios afectan los resultados.

Se recomienda añadir un módulo de gestión de algoritmos que permita al usuario administrar los algoritmos de predicción según sus necesidades.

Para garantizar la integridad de los datos, se propone realizar copias de seguridad periódicas para proteger la información generada por la aplicación web y mejorar el rendimiento académico.

Se sugiere ofrecer capacitaciones constantes a los usuarios de la aplicación web mediante talleres, materiales de apoyo y manuales claros y accesibles, para que comprendan las capacidades y funcionalidades de la herramienta.

Referencias

[1] M. Cormann, «El estado del aprendizaje y la equidad en la educación,» *OECD,* vol. Volumen I, nº 1, p. 3, 2023.

[2] J. D. Garcia Dionisio, «Machine Learning para predecir el rendimiento académico de los estudiantes universitarios,» , 2021.

[3] F. H. LAYZA BERMUDEZ, «MODELO MATEMÁTICO DE PREDICCIÓN DEL RENDIMIENTO ACADÉMICO DEL CURSO MATEMÁTICA II EN LA FACULTAD DE INGENIERÍA QUÍMICA, UNAC 2019,» Callao, Callao, 2020.

[4] A. Rico Páez y N. D. Gaytán Ramírez, «Modelos predictivos del rendimiento académico,» *Revista de Investigacion educativa de la Radio Ch,* vol. Volumen 13, nº ISSN: 2448-8550, p. 18 paginas, 2022.

[5] L. J. Garcia Peredo, «SISTEMA PREDICTIVO DE RENDIMIENTO ACADÉMICO EN BASE A FACTORES INFLUYENTES EN ESTUDIANTES DEL 1º SECUNDARIA EN UN COLEGIO DE LAMBAYEQUE,» Universidad Catolica Santo Toribio de Mogrovejo, Lima-Peru, 2021.

[6] H. S. Chatila Vargas Toscanelli , «Sistema web basado en técnicas de machine learning para predecir el desempeño de los docentes en el nivel de educación superior universitaria,» Usat, Chiclayo, 2023.

[7] Y. G. Ortiz., «Estudiantes universitarios con bajo rendimiento académico,,» *Edu Mesentro,* p. 7, 14 Septiembre 2020.

[8] J. Tourón, «LA PREDICCION DEL RENDIMIENTO ACADEMICO DE LOS ALUMNOS UNIVERSITARIOS: ALGUNAS IMPLICACIONES PEDAGOGICAS,» *https://www.researchgate.net/publication/265268986,* vol. 3, nº 3, p. 11, 2021.

[9] E. L. Ruiz1, «Habilidades no cognitivas y rendimiento académico. Un estudio empírico en alumnos de secundaria,» *Revista de Educación, Innovación y Formación,* vol. 7, nº 27, p. 21, 2022.

[10] B. N. Diaz Landa, R. A. Meleán Romero y W. . J. Marín-Rodriguez, «RENDIMIENTO ACADÉMICO DE ESTUDIANTES EN EDUCACIÓN SUPERIOR: PREDICCIONES DE FACTORES INFLUYENTES A PARTIR DE ÁRBOLES DE DECISIÓN,» Revista de Estudios Interdisciplinarios en Ciencias Sociales, Chiclayo-Peru, 2022.

[11] H. Gutiérrez Villaverde, M. Linares Barbero, A. Agüero Correa y J. Pérez Nuñe, «Predicción de Rendimiento Académico de alumnos usando Machine Learning,» UNIVERSIDAD DE LIMA FACULTAD DE CIENCIAS EMPRESARIALES Y ECONÓMICAS CARRERA DE NEGOCIOS INTERNACIONALES , Lima-´Peru, 2022.

[12] J. Calvo, M. A. Guzman y D. Ramos, Machine Laearning una pieza clave en la transformacion de los modelos de Negocio, : Manager Solutions, 2020.

[13] B. N. Diaz Landa, R. A. Meleán Romero y W. . J. Marín-Rodriguez, «RENDIMIENTO ACADÉMICO DE ESTUDIANTES EN EDUCACIÓN SUPERIOR: PREDICCIONES DE FACTORES INFLUYENTES A PARTIR DE ÁRBOLES DE DECISIÓN,» Revista de Estudios Interdisciplinarios en Ciencias Sociales, Chiclayo-Peru, 2022.

[14] D. H. Ramirez, «El Machine Learning a travez de los tiempos y los aportes a la humanidad.,» Pereira, 2021.

[15] R. Edel Navarro, «El rendimiento academico: Concepto, investigacion y desarrollo.,» *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación,* vol. Volumen 1, nº 2, p. 16, 2023.

[16] Frascati, «GUÍA PARA LA RECOPILACIÓN Y PRESENTACIÓN DE INFORMACIÓN SOBRE LA INVESTIGACIÓN,» Fundación Española para la Ciencia y la Tecnología, FECYT, 2020.

[17] G. X. ESPINOZA AIRAC y E. F. LEON MUÑOZ, «MODELO DE MACHINE LEARNING PARA LA CLASIFICACIÓN DE ESTUDIANTES DE ACUERDO A SU RENDIMIENTO CADÉMICO EN EL CENTRO DE IDIOMAS DE LA UNIVERSIDAD NACIONAL DEL SANTA,» Chimbote - PERÚ, 2022.

[18] A. Rico Páez y N. D. Gaytán Ramírez, «Modelos predictivos del rendimiento académico,» *Revista de Investigacion educativa de la Radio Ch,* vol. 13, nº ISSN: 2448-8550, p. 18, 2022.

[19] A. R. Páez, «Diseño de un modelo para automatizar la predicción del rendimiento académico en estudiantes del IPN,» Revista Iberoamericana para la investigacion y desarrollo educativo., Mexico City-Mexico, 2023.

[20] J. Ocaña-Zúñiga, «Variables cognitivas de los estudiantes universitarios: su relación con dedicación al estudio y rendimiento academico,» *PSICUMEX,* vol. volumen 10, nº Numero1, p. 14, 2022.

[21] C. E. Castillo Diestra y J. L. Gutierrez Gutierrez , «Instrumento para medir la funcionalidad y Usabilidad de un Software Comercial,» *Revista Cientifica San Pedro, Conocimiento para el Desarrollo,* vol. 2, nº 2, p. 10, 2022.

[22] J. A. C. Huancas, «Aplicación web y móvil basado en el Algoritmo de Optimización de Colonia de Hormigas para apoyar el proceso de ventas en la panadería Mishka,» Universidad Catolica Santo Toribio de Mogrovejo, Chiclayo-Peru, 2024.

[23] A. Rico Paes y D. Sanchez Guzman, «Diseño de un modelo para automatizar la predicción del rendimiento académico en estudiantes del IPN,» *Revista Iberoamericana para la investigacion y el Desarrollo Educativo ISSN,* vol. ISSN, nº 7467, p. 21, 2020.

[24] L. J. Garcia Peredo, «Sistema predictivo del rendimiento academico en base a factores influyentes en estudiantes del primero de secundaria en un colegio de lambayeque,» Usat, chiclayo, 2021.

Anexos

Anexo 01: constancia de aprobación del producto acreditable en el colegio militar.

GOBIERNO REGIONAL LAMBAYEQUE

I.E.P.M "Colegio Militar Elías Aguirre"
CM: 0452581

"AÑO DEL BICENTENARIO DE LA CONSOLIDACIÓN DE NUESTRA INDEPENDENCIA Y DE LA CONMEMORACIÓN DE LAS HEROICAS BATALLAS DE JUNÍN Y AYACUCHO"

CARTA N° 001- 2024

ASUNTO: CONSTANCIA DE APROBACIÓN DEL PRODUCTO ACREDITABLE DE TRABAJO DE INVESTIGACIÓN.:

"APLICACIÓN WEB BASADA EN MACHINE LEARNING PARA PREDECIR EL RENDIMIENTO ACADÉMICO EN EL ÁREA DE MATEMÁTICA DE LOS ESTUDIANTES DEL TERCER GRADO DE EDUCACIÓN SECUNDARIA EN UN COLEGIO MILITAR EN LAMBAYEQUE".

De nuestra consideración:

Quien suscribe y en representación de la Institución Educativa Pública Militar "Colegio Militar Elías Aguirre" en mi calidad de subdirectora del Departamento de Evaluación, es grato dirigirme a usted con la finalidad de saludarlo cordialmente, asimismo, mediante la presente se suscribe que, habiendo revisado el producto acreditable de la tesis que lleva como título: **"APLICACIÓN WEB BASADA EN MACHINE LEARNING PARA PREDECIR EL RENDIMIENTO ACADÉMICO EN EL ÁREA DE MATEMÁTICA DE LOS ESTUDIANTES DEL TERCER GRADO DE EDUCACIÓN SECUNDARIA EN UN COLEGIO MILITAR EN LAMBAYEQUE"**, presentado por el estudiante EDWIN MANUEL BAZÁN FERNÁNDEZ, identificado con código universitario N°011120112I de la Escuela de Ingeniería de Sistemas y Computación de la universidad "Santo Toribio de Mogrovejo", se certifica la culminación exitosa del correcto funcionamiento y el cumplimiento de los requisitos establecidos, aportando a satisfacer las necesidades de la Institución, habiéndole dado las facilidades, permisos y apoyo pertinentes en acceder a la información necesaria para el desarrollo de la tesis; asimismo mencionar que se tomaron las medidas sanitarias de seguridad.

Reiteró mi estima y consideración

Chiclayo, mayo de 2024

Colegio Militar "Elías Aguirre" Departamento de Evaluación

Dra. Betty Ydali Vargas Tocto
JEFA DEL DPTO. DE EVALUACIÓN

Anexo 02: Manual de Usuario

1. **Modulo Registro Usuario**

Interfaz Registrar Usuario

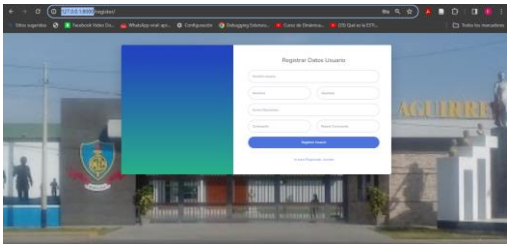


Fig. 8. Interfaz Registrar Usuario

En esta interfaz, el usuario debe ingresar los datos requeridos para registrar una nueva cuenta y obtener acceso a la aplicación web

2. **Modulo Acceso – Login de Usuarios**

Interfaz Acceso Aplicación Web

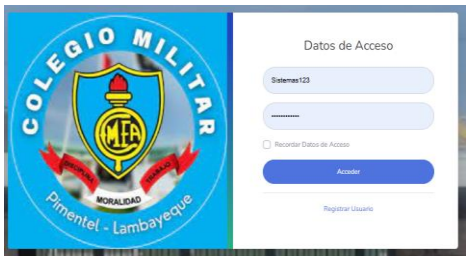


Fig. 9. Interfaz acceso Aplicación web

En esta interfaz, el usuario debe ingresar sus datos de acceso (usuario y contraseña), las cuales fueron registradas en la base de datos al momento de crear su cuenta en la aplicación web, la opción registrar usuario permite crear un nuevo usuario que tendrá acceso a la aplicación web.

3. **Modulo Algoritmos de entrenamiento**

a **Cargar Datos Históricos**

- **Interfaz Cargar Datos Históricos**

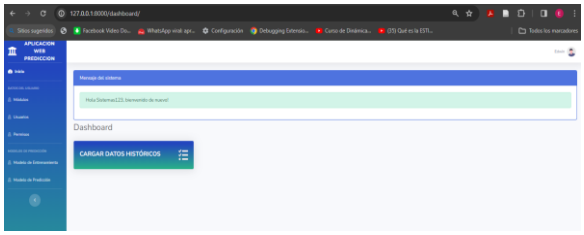


Fig. 10 Interfaz Cargar datos históricos

En esta interfaz, el usuario puede cargar los datos históricos de los estudiantes mediante un archivo en formato Excel. En la prueba inicial, se cargaron datos correspondientes al periodo 2018–2022. Al momento de la carga, la aplicación web realiza automáticamente las

siguientes validaciones y procesos: Verifica el formato del archivo cargado, comprueba que el archivo Excel cumpla con los formatos requeridas, genera los modelos de entrenamiento, Selecciona el algoritmo más óptimo para generar el modelo de predicción de los nuevos estudiantes.

b Validaciones Archivo Datos Históricos

- **Interfaz Validaciones Archivo datos históricos.**

Fig. 11. Interfaz Validaciones datos históricos

En esta interfaz se muestran todos los mensajes generados durante la lectura del archivo Excel con los datos históricos requeridos. La aplicación web valida que el archivo cumpla con las siguientes condiciones: Formato Excel, total de columnas (variables) configuradas previamente en la aplicación web, tener un tamaño diferente al archivo registrado, incluir todas las categorías correspondientes a cada variable previamente asignada.

c Listado algoritmos de Entrenamiento.

- **Interfaz Listado algoritmos de Entrenamiento**

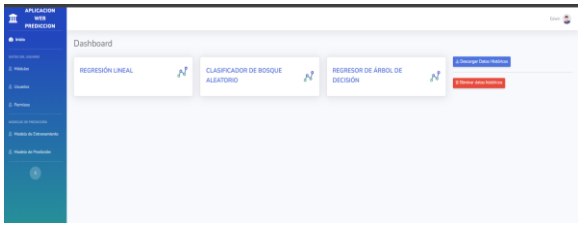


Fig. 12- Interfaz Listado de algoritmos de entrenamiento

Esta interfaz se muestra una vez que la aplicación valida el archivo Excel cargado con los datos históricos de los estudiantes. La información es procesada y se presentan los tres algoritmos de entrenamiento disponibles: Regresión Lineal (Linear Regression), Clasificador de Bosque Aleatorio (GaussianNB) y Regresor de Árbol de Decisión (DecisionTreeRegressor). Cada algoritmo utiliza el 80% de los datos para el entrenamiento y el 20% para la validación. Además, la interfaz permite descargar o eliminar el archivo Excel cargado.

d Estructura de cada algoritmo de Entrenamiento

- **Interfaz Grafico estadístico de Barras**

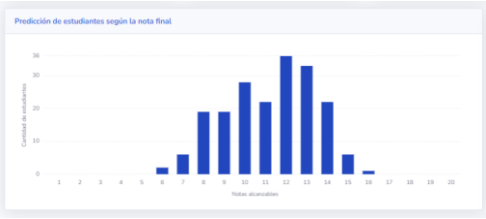


Fig. 13. Cantidad de estudiantes según nota- algoritmo entrenamiento

En esta interfaz se muestra un gráfico de barras que representa la distribución del total de estudiantes según la nota de predicción, dentro de un rango de 0 a 20. El gráfico se genera de manera dinámica a partir de los datos históricos o de los datos actuales cargados mediante el archivo Excel. Asimismo, se actualiza automáticamente según el algoritmo de predicción seleccionado o en función del modelo de predicción aplicado a los nuevos estudiantes.

- **Interfaz grafico estadístico de pastel**

Fig. 14. Gráfico de Pastel- Porcentaje de Aprobados y Desaprobados

En esta interfaz se presenta un gráfico de pastel que muestra el porcentaje de aprobados y desaprobados. El gráfico se genera dinámicamente a partir de los datos históricos registrados al momento de cargar el archivo Excel y se actualiza cada algoritmo de entrenamiento implementado.

- **Interfaz grado Importancia Variables algoritmos de entrenamiento**

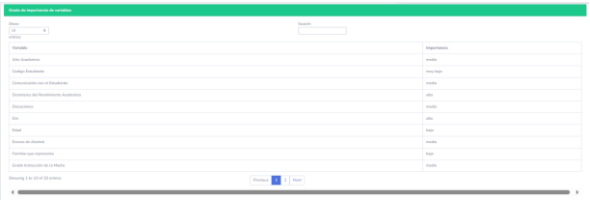


Fig. 15. Grafico grado Importancia variables utilizadas

En esta interfaz se presenta el grado de importancia de las variables utilizadas en los algoritmos de entrenamiento, las cuales consta de: muy bajo, bajo, medio, alto, dichos grados depende de como las variables influyen en el entrenamiento de los algoritmos de entrenamiento.

- **Interfaz Listado de Aprobados y Desaprobados**

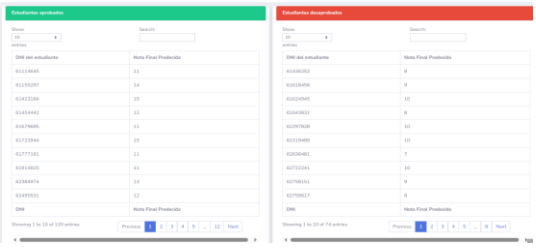


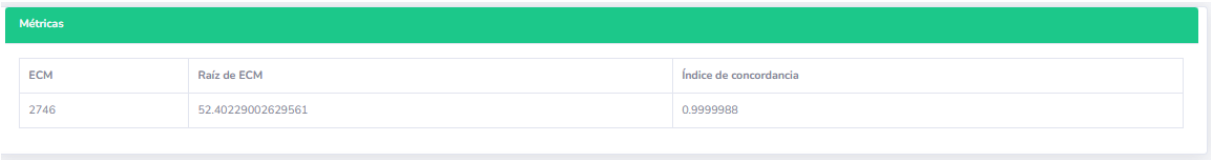
Fig. 16 . Interfaz Aprobados y desaprobados – algoritmo entrenamiento

- **Interfaz Cuadro Comparativo – Algoritmos de entrenamiento**

Fig. 17. Interfaz porcentajes de predicción – algoritmos de entrenamiento

En esta interfaz se presentan los porcentajes de exactitud obtenidos por cada algoritmo de predicción implementado. Su finalidad es identificar el algoritmo con el mayor porcentaje de acierto, el cual será seleccionado automáticamente por la aplicación web para generar el modelo de predicción aplicado a los nuevos estudiantes.

- **Cantidades evaluación del algoritmo de entrenamiento**



Métricas

ECM	Raíz de ECM	Índice de concordancia
2746	52.40229002629561	0.9999988

Fig. 18. Evaluación - Algoritmo de entrenamiento

En esta interfaz se presentan los resultados obtenidos al evaluar cada algoritmo de entrenamiento. Con base en estas métricas, la aplicación web selecciona automáticamente el algoritmo más óptimo para generar el modelo de predicción.

e **Opciones archivo datos históricos – algoritmos de entrenamiento**

- **Opción eliminar archivo datos históricos**

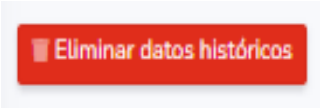


Fig. 19. Opción Eliminar Datos Históricos

En esta interfaz permite eliminar el archivo de Excel que corresponde a los datos históricos que se han cargado previamente en la opción de cargar datos históricos.

- **Opción mensaje confirmación – eliminar archivo datos históricos**

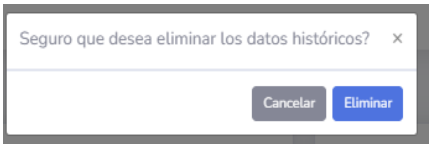


Fig. 20. Opción mensaje confirmación

En esta interfaz muestra un mensaje de confirmación al intentar eliminar el archivo Excel de datos históricos cargado. Si el usuario selecciona Eliminar, se borran todos los algoritmos de entrenamiento generados con dichos datos. La aplicación habilita la opción Cargar Datos Históricos, permitiendo subir un nuevo archivo para regenerar los algoritmos de entrenamiento. Si el usuario selecciona Cancelar, la aplicación no realiza ninguna acción y conserva los datos históricos existentes.

- **Opción Descargar archivo Datos Históricos**

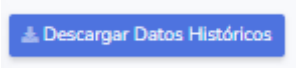


Fig. 21. Opción descargar datos históricos

En esta interfaz ofrece la opción de descargar el archivo de datos históricos previamente cargado. Esta funcionalidad permite al usuario verificar si el archivo excel corresponde a la versión correcta. Si el archivo es el correcto, puede mantenerse sin cambios, si el archivo no corresponde, el usuario puede utilizar la opción eliminar datos históricos y, posteriormente, subir la versión correcta para volver a generar los algoritmos de entrenamiento.

4. Modulo Modelo Predicción

a Cargar Datos Actuales.

- **Interfaz Cargar Datos Actuales**

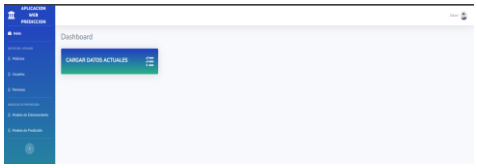


Fig. 22. Interfaz Cargar Datos actuales

En esta interfaz, el usuario puede cargar los datos históricos de los estudiantes mediante un archivo en formato Excel. En la prueba inicial, se cargaron los datos correspondientes al periodo 2018–2022. Al momento de la carga, el sistema realiza automáticamente las siguientes validaciones y procesos: Verifica el formato del archivo cargado, comprueba que el archivo Excel cumpla con los formatos requeridas, genera los modelos de entrenamiento, Selecciona el algoritmo más óptimo para generar el modelo de predicción de los nuevos estudiantes.

b Validaciones archivo datos actuales

- **Interfaz Validaciones Archivo Datos Actuales**

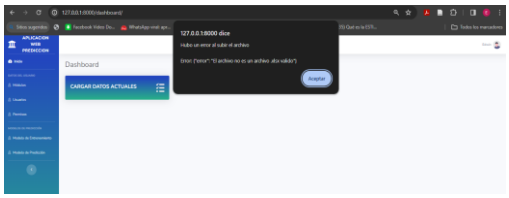


Fig. 23. Interfaz validaciones datos actuales.

Esta interfaz permite al usuario cargar un archivo Excel con los datos de los nuevos estudiantes, al realizar la carga, la aplicación web ejecuta automáticamente las siguientes validaciones: el formato del archivo, que permite verificar que el archivo cargado sea compatible, la estructura del excel, permite comprobar las características requeridas, consistencia de datos, permite identificar y descarta posibles errores o inconsistencias. Una vez superadas las validaciones, la aplicación web selecciona el algoritmo más óptimo entre los previamente entrenados y lo utiliza para generar el modelo de predicción del rendimiento académico de los nuevos estudiantes.

c Opciones Archivo Datos Actuales – modelo de predicción

- **Opción Eliminar archivo Datos actuales**

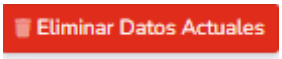


Fig. 24. Opción Eliminar Datos Actuales

En esta interfaz permite eliminar el archivo de Excel que corresponde a los datos actuales que se han cargado previamente en la opción de cargar datos actuales.

- **Opción Mensaje confirmación – eliminar datos actuales**

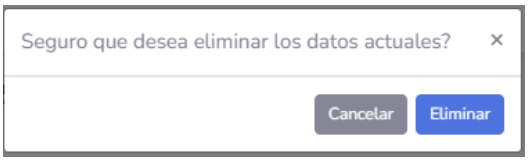


Fig. 25. Opción mensaje confirmación

La interfaz muestra un mensaje de confirmación al intentar eliminar el archivo Excel de datos actuales cargado. Si el usuario selecciona Eliminar, se borra el modelo de predicción generados con dichos datos. La aplicación habilita la opción Cargar Datos actuales permitiendo subir un nuevo archivo para regenerar el modelo de predicción. Si el usuario selecciona Cancelar, la aplicación no realiza ninguna acción y conserva los datos actuales existentes.

- **Opción Descargar datos actuales**

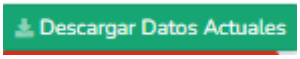


Fig. 26. Opción descargar datos actuales

En esta interfaz ofrece la opción de descargar el archivo de datos actuales previamente cargado. Esta funcionalidad permite al usuario verificar si el archivo excel corresponde a la versión correcta. Si el archivo es el correcto, puede mantenerse sin cambios, si el archivo no corresponde, el usuario puede utilizar la opción eliminar datos actuales y, posteriormente, subir la versión correcta para volver a generar el modelo de predicción

d Estructura del modelo predicción.

- **Interfaz Grafico estadístico de Barras**

Fig. 27 - Cantidad de estudiantes según nota predicción - modelo predicción

En esta interfaz se muestra un gráfico de barras que representa la distribución del total de estudiantes según la nota de predicción, dentro de un rango de 0 a 20. El gráfico se genera de manera dinámica a partir de los datos actuales cargados mediante el archivo Excel. Asimismo, se actualiza automáticamente según el modelo de predicción, que a su vez se genera seleccionando el algoritmo de entrenamiento con mayor porcentaje de predicción y este se aplica a los nuevos estudiantes.

- **Interfaz grafico estadístico de pastel**

Fig. 28. Gráfico de Pastel- Porcentaje de Aprobados y Desaprobados

En esta interfaz se presenta un gráfico de pastel que muestra el porcentaje de aprobados y desaprobados. El gráfico se genera dinámicamente a partir de los datos actuales registrados al momento de cargar el archivo Excel y se actualiza de acuerdo con el modelo de predicción de los nuevos estudiantes

- **Interfaz grado Importancia Variables modelo de predicción**

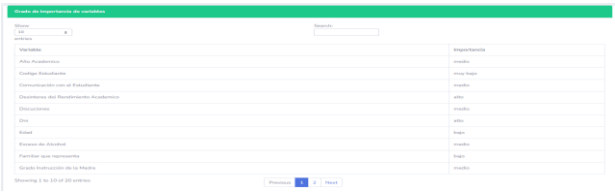


Fig. 29. Gráfico grado Importancia variables utilizadas

En esta interfaz se presenta el grado de importancia de las variables utilizadas en el modelo de predicción, las cuales consta de: muy bajo, bajo, medio, alto, dichos grados depende de cómo las variables influyen en el modelo de predicción de los nuevos estudiantes.

- **Interfaz Listado de Aprobados y Desaprobados**

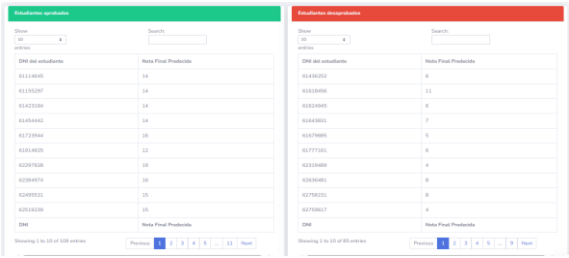


Fig. 30 . Interfaz Aprobados y desaprobados – modelo predicción

En esta interfaz se muestran dos tablas separadas con el listado de estudiantes aprobados y desaprobados. Cada tabla incluye una opción de búsqueda por DNI, la nota mostrada es según el modelo de predicción generado.

- **Resumen rendimiento académico 2024**

Resumen de predicción del año académico actual

Año actual	Estudiantes matriculados		Estudiantes desaprobados		Estudiantes aprobados			
					Promedio < 14		Promedio >= 14	
	Total	Porcentaje	Total	Porcentaje	Total	Porcentaje	Total	Porcentaje
2024	194	100	85	43.81443298969072	17	15.596330275229358	92	84.40366972477064

Fig. 31 . Interfaz de predicción año académico actual

Esta interfaz presenta un resumen del rendimiento académico de los nuevos estudiantes, generado a partir del modelo de predicción con mayor porcentaje de acierto del año 2024, del total de 194 estudiantes matriculados: predicción de desaprobados 85 (43.81%), predicción de aprobados 109 (56.19%), distribuidos en 92 estudiantes con nota mayor o igual a 14 (84.40%), 17 estudiantes con nota menor a 14 (15.60%**).**

- **Cantidades evaluación del modelo predicción**

Métricas

ECM	Raíz de ECM	Índice de concordancia
1418	37.656340767525464	0.9999993

Fig. 32. Evaluación – Modelo Predicción

En esta interfaz se presentan los resultados obtenidos al evaluar el modelo de predicción, utilizando el algoritmo más optimo.

5. **Modulo Listado Módulos aplicación web.**

- **Interfaz Listado Módulos Aplicación web**

Fig. 33. Interfaz Listado Módulos

Esta interfaz muestra las opciones disponibles de cada módulo implementado en la aplicación web, accesibles únicamente mediante el perfil de usuario administrador**.**

6. **Modulo Listado Permisos por módulos Aplicación web.**

- **Interfaz agrupación permisos por modulo**

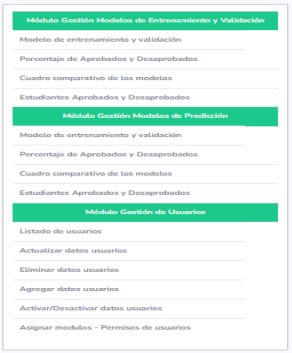


Fig. 34. Permisos – Modulo Aplicación web

En esta interfaz muestra los permisos disponibles en cada módulo implementado en la aplicación web, accesibles únicamente mediante el perfil de usuario administrador**.**

7. **Modulo Gestión de Usuarios**

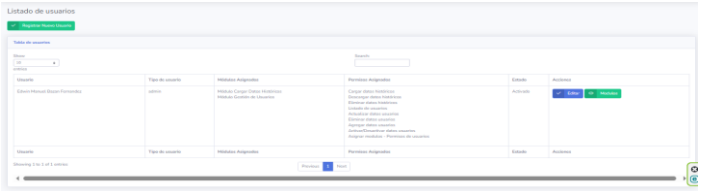


Fig. 35. Modulo Gestión de Usuarios

En está interfaz permite administrar a los usuarios que acceden a la aplicación web. La información se presenta en una tabla con los siguientes campos: Usuarios, muestra los apellidos y nombres de los usuarios registrados, Tipo de Usuario, indica el perfil o rol asignado a cada usuario, Módulos, detalla los módulos habilitados para cada usuario, Permisos, especifica los permisos otorgados dentro de cada módulo asignado, Estado, permite identificar si el usuario se encuentra activo/desbloqueado o inactivo/bloqueado. Esta interfaz está disponible únicamente para el perfil de administrador.

a Opción Registrar Usuario

- **Interfaz agregar nuevo usuario**

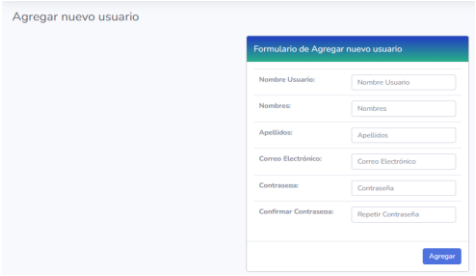


Fig. 36. Interfaz agregar nuevo usuario

Esta interfaz permite registrar una nueva cuenta de usuario en la aplicación web. Los campos a completar son: Nombre de usuario, identificador con el que el usuario accederá a la aplicación, Nombres, Nombres del usuario, Apellidos**:** Apellidos del usuario, Correo electrónico: Dirección de correo electrónico del usuario, Tipo de usuario**:** Perfil o rol asignado al usuario dentro de la aplicación, Activación**:** Estado de la cuenta (activa o inactiva), Una vez completados todos los campos, haga clic en Agregar para registrar la nueva cuenta de usuario.

b Opción Modificar Usuario

- **Interfaz actualizar usuario registrado**

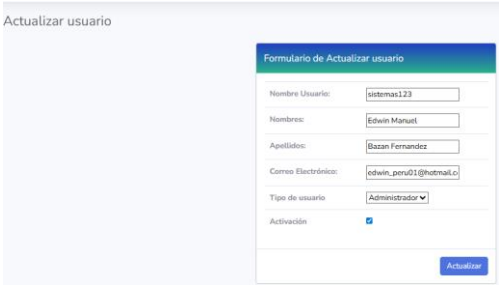


Fig. 37. Interfaz modificar datos usuarios

Esta interfaz permite actualizar la cuenta de usuario registrada en la aplicación web. Los campos a modificar son: Nombre de usuario, identificador con el que el usuario accederá a la aplicación, Nombres, Nombres del usuario, Apellidos**:** Apellidos del usuario, Correo electrónico: Dirección de correo electrónico del usuario, Tipo de usuario**:** Perfil o rol asignado

al usuario dentro de la aplicación, Activación**:** Estado de la cuenta (activa o inactiva), Una vez actualizado todos los campos, haga clic en actualizar para modificar los campos del usuario registrado.

c Opción Asignar Módulos y permisos – usuario

- **Interfaz asignación módulos y permisos - usuario**

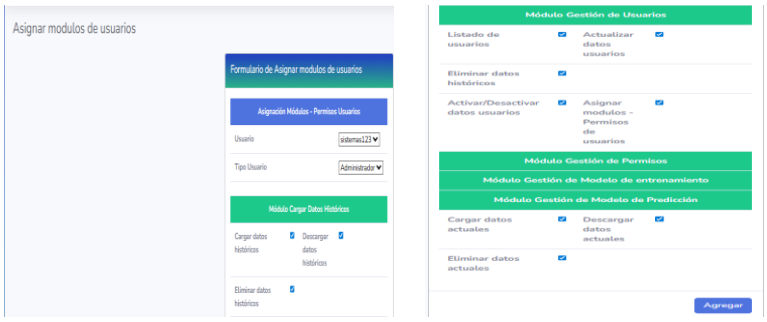


Fig. 38. Interfaz módulos – permisos usuario seleccionado

Esta interfaz permite asignar módulos y permisos al usuario seleccionado, Los datos del usuario y el tipo de usuario se muestran únicamente de forma informativa y no pueden ser modificados, **l**os permisos se agrupan por módulo. Al seleccionar un permiso de cualquier módulo, este se asigna automáticamente. Los permisos disponibles por módulo son: Módulo Algoritmos de entrenamiento, **c**argar datos históricos, descargar datos históricos, eliminar datos históricos, módulo gestión de usuarios, registrar usuario, modificar usuario, listado de usuarios, asignar módulos y permisos, activar/desactivar usuarios, módulo modelo de predicción, actualizar datos actuales, eliminar datos actuales, descargar datos actuales, finalmente, el usuario debe hacer clic en Asignar para guardar la configuración de módulos y permisos seleccionados.