



III Memoria SOCID
"Actas del III Congreso de Investigación Científica"
Sociedad Científica de Docentes
Universidad Nacional "Siglo XX"



**MODELO DE PREDICCIÓN DEL RENDIMIENTO
ACADÉMICO EMPLEANDO EL APRENDIZAJE
AUTOMÁTICO**

Jorge Villcáez Castillo Ph.D.
jorgevill2015@gmail.com

Ingeniería Informática
Universidad Nacional "Siglo XX"
Llallagua - Bolivia

RESUMEN -

La presente investigación enfoca el desarrollo de un modelo de predicción del rendimiento académico en la educación superior para el Instituto Técnico "PÍO XII" de la ciudad de Llallagua empleando el aprendizaje automático, para identificar factores que afectan el bajo rendimiento de los estudiantes.

La problemática identificada es el porcentaje considerable de estudiantes que presentan bajo rendimiento y estudiantes que abandonan sus estudios por diversos factores, ante ello se planteó el objetivo general de implementar un modelo de predicción del rendimiento académico para identificar los factores de bajo rendimiento.

Se utilizó el paradigma de investigación naturalismo científico o positivista, el enfoque de investigación fue el mixto, es decir cualitativo, cuantitativo y empírico multidisciplinario. Los tipos de investigación utilizados fueron la aplicada, la propositiva transformadora y documental.

Los resultados obtenidos permitieron identificar 23 factores que inciden en el rendimiento académico.

Palabras clave: Aprendizaje Automático, Ciencia de Datos, Modelo de Regresión, Logística, Predicción, Rendimiento Académico.

ABSTRACT

This research focuses on the development of an academic performance prediction model in higher education for the "PÍO XII" Technical Institute in the city of Llallagua, using machine learning to identify factors affecting student underachievement.



III Memoria SOCID “Actas del III Congreso de Investigación Científica” Sociedad Científica de Docentes Universidad Nacional “Siglo XX”

The identified problem is the significant percentage of students who present low achievement and students who drop out of their studies due to various factors. Therefore, the general objective was to implement an academic performance prediction model to identify factors of underachievement.

The scientific naturalism or positivist research paradigm was used, and the research approach was mixed, that is, qualitative, quantitative, and multidisciplinary empirical. The types of research used were applied, transformative propositional, and documentary.

The results obtained allowed us to identify 23 factors that influence academic performance.

Keywords: Academic Performance, Data Science, Logistic Regression Model, Machine Learning, Prediction.

1. INTRODUCCIÓN

Actualmente los sistemas de predicción, basados en el uso de algoritmos de Machine Learning o aprendizaje automático, permiten desarrollar predicciones y estimaciones con los datos con los que las instituciones cuentan a partir de las conductas de las personas, generando información necesaria para la toma de decisiones. Estos sistemas de predicción están conformados por un grupo de técnicas que construyen y estudian nuevas previsiones a través de una rama de la inteligencia artificial denominada aprendizaje automático.

1.1. Problemática de la investigación

La situación problemática identificada con respecto al rendimiento educativo de los estudiantes del Instituto Técnico “PÍO XII” de la ciudad de Llalagua, principalmente de resume en los siguientes aspectos:

- Se tiene una población estudiantil con diversidad de antecedentes académicos, demográficos, culturales, socioeconómicos y otros, lo que hace que sus condiciones no sean iguales y puede ocasionar diferencias en cuanto al rendimiento académico. Se tiene que un 12.7% de estudiantes provienen de colegios de área dispersa en comparación con un 87.3% del área urbana. Se tiene que un 20.7% de estudiantes provienen del interior y un 79.3% provienen del municipio de Llalagua. Un 53.3% de estudiantes trabaja en el día. Un 36.7% de estudiantes del instituto, estudia en diferentes Carreras de la Universidad Nacional “Siglo XX” en el día.
- Existe dificultad en la identificación anticipada de estudiantes que tienen bajo rendimiento, ya que no se tienen políticas y estrategias institucionales, que permitan identificar de manera temprana el bajo rendimiento de los estudiantes.
- El Instituto Técnico “PÍO XII” tiene recursos limitados económicos y humanos para brindar apoyo personalizado a todos los estudiantes que tienen bajo rendimiento. Los ingresos del Instituto se obtienen solamente del costo de matrícula de los estudiantes que es de Bs 100.-, ya que el Instituto es de carácter fiscal; aproximadamente se tiene alrededor de 600 estudiantes por gestión, lo que permite al Instituto tener un ingreso aproximado de Bs 60.000.- por gestión, los cuales son destinados a cubrir las diversas necesidades del instituto para su funcionamiento. El instituto cuenta con 25 docentes, 3 directivos y 2 administrativos.
- El Instituto tiene gran número de estudiantes a los que no se puede atender de manera personalizada y ver qué estudiantes de bajo rendimiento pueden ser sujetos de un trato diferente para mejorar su rendimiento. Como promedio se tiene 600 estudiantes por gestión académica para 25 docentes, lo que muestra que la relación estudiante/docente sea de 24 estudiantes por docente, lo que dificulta una atención personalizada.



III Memoria SOCID

“Actas del III Congreso de Investigación Científica”

Sociedad Científica de Docentes
Universidad Nacional “Siglo XX”

- Carencia de estrategias efectivas de apoyo y tutoría para los estudiantes que presentan bajo rendimiento. El instituto no cuenta con políticas, reglamentos, normativas y estrategias que permitan brindar apoyo por parte de los docentes hacia los estudiantes identificados con bajo rendimiento.
- Porcentaje considerable de estudiantes que reprueban asignaturas en las distintas carreras del Instituto. Se tiene un 36.7% de estudiantes que reprobaron alguna asignatura.
- Porcentaje considerable de estudiantes que reprueban en alguna asignatura en un bimestre determinado. Se tiene un 37.3% de estudiantes que reprobaron alguna asignatura en un determinado bimestre.
- Estudiantes que están nivelando asignaturas, como consecuencia de su bajo rendimiento. Se tiene un 10% de estudiantes que se encuentran nivelando algunas asignaturas.
- Estudiantes que están arrastrando asignaturas, como consecuencia de su bajo rendimiento. Se tiene un 10% de estudiantes que se encuentran arrastrando algunas asignaturas.

1.2. Formulación del problema científico

¿Cómo mejorar de manera efectiva el rendimiento académico de los estudiantes del Instituto Técnico “PÍO XII” de la ciudad de Llalagua con la implementación del Modelo de Predicción del Rendimiento Académico?

1.3. Objetivos de la investigación

1.3.1. Objetivo general

Implementar un modelo de predicción del rendimiento académico en la educación superior para el Instituto Técnico “PÍO XII” de la ciudad de Llalagua empleando el aprendizaje automático, para identificar factores de bajo rendimiento académico.

1.3.2. Objetivos específicos

- Obtener los datos relevantes de los estudiantes del Instituto Técnico “PÍO XII” mediante un cuestionario aplicado a los estudiantes de Segundo y Tercer curso de las Carreras con las que cuenta el Instituto y la exploración de calificaciones.
- Realizar el Preprocesamiento y Análisis exploratorio de datos obtenidos para la selección de variables, para garantizar la calidad de los datos mediante herramientas de ciencia de datos.
- Aplicar modelos de Aprendizaje Automático y herramientas de Ciencia de Datos, con el fin de identificar factores y patrones que influyen en el rendimiento académico.
- Obtener el modelo de predicción empleando algoritmos de aprendizaje automático supervisados para la predicción del rendimiento académico de los estudiantes.
- Evaluar los resultados del modelo predictivo obtenido para predecir el rendimiento académico.

1.4. Justificación

La investigación propuesta aborda una problemática vigente en el ámbito educativo en el Instituto Técnico “PÍO XII” de la ciudad de Llalagua, y tiene el potencial de generar un impacto en la mejora de la educación superior, la igualdad de condiciones de acceso a la educación y la toma de decisiones anticipadas y efectivas en las instituciones educativas de educación superior.

1.5. Hipótesis

El modelo de predicción del rendimiento académico en la educación superior para el Instituto Técnico "PÍO XII" la ciudad de Llallagua empleando el aprendizaje automático, permite identificar los factores influyentes en el rendimiento académico de los estudiantes del Instituto Técnico "PÍO XII" de la ciudad de Llallagua.

Variables

- Variable Independiente: Modelo de predicción del rendimiento académico en la educación superior para el Instituto Técnico "PÍO XII" la ciudad de Llallagua empleando el aprendizaje automático.
- Variable Dependiente: Factores influyentes en el rendimiento académico de los estudiantes del Instituto Técnico "PÍO XII" de la ciudad de Llallagua.

1.5. Revisión de Literatura

1.5.1. Predicción en machine learning

Los sistemas de predicción, basados en el uso de algoritmos de Machine Learning o aprendizaje autónomo permiten desarrollar predicciones y estimaciones con los datos con los que las empresas cuentan a partir de las conductas de sus usuarios, generando valor para las empresas.

Los sistemas de predicción están conformados por un grupo de técnicas que construyen y estudian nuevas previsiones a través de una rama de la inteligencia artificial denominada aprendizaje automático (Agencia Vasca de Desarrollo Empresarial, 2018).

1.5.2. Rendimiento académico

La mayoría de investigadores concuerdan en que el rendimiento académico es el resultado del aprendizaje producido por la interacción didáctica y pedagógica del docente y estudiante. El rendimiento académico es una medida de las capacidades respondientes o indicativas que manifiestan, en forma estimativa, lo que una persona ha aprendido como consecuencia de un proceso de instrucción o formación. También se considera que el rendimiento académico es el producto que da el alumnado en los centros de enseñanza y que habitualmente se expresa a través de las calificaciones escolares (Tacilla Cardenas, Vásquez Villanueva, Verde Avalos, & Colque Díaz, 2020).

El rendimiento académico depende en gran parte de la forma en la que se obtienen los resultados de los aprendizajes, estos resultados se logran en un periodo académico determinado, en los cuales se evalúa de manera cualitativa y cuantitativa para saber si se alcanzó los objetivos propuestos (Estrada García, 2018, pág. 224).

1.5.3. Educación superior

La educación superior es aquella que contempla la última fase del proceso de aprendizaje académico, es decir, aquella que viene luego de la etapa secundaria. Es impartida en las universidades, institutos superiores o academias de formación técnica. La enseñanza que ofrece la educación superior es a nivel profesional (Pedraja Rejas, Rodríguez Ponce, & Labraña, 2022).

La educación superior le brinda al estudiante la oportunidad de capacitarse académicamente para luego poder entrar al campo laboral. Es por esto que en ella se estudian un conjunto de asignaturas consideradas como obligatorias y otras consideradas optativas, con el fin de ejercer una profesión que esté asociado con ellas. Es importante que la persona antes de cursar estudios superiores tenga muy claro su vocación profesional y cómo se encuentra el mercado laboral en relación con la carrera que desea cursar.



III Memoria SOCID “Actas del III Congreso de Investigación Científica” Sociedad Científica de Docentes Universidad Nacional “Siglo XX”

La educación superior no sólo se dedica a formar profesionales, sino que también tiene una función orientada hacia la investigación y ésta a su vez debe estar vinculada con la sociedad, ya que muchos de los conocimientos obtenidos de las investigaciones deben beneficiar a la sociedad. Ya que lo ideal es que todo egresado de una universidad debe estar consciente que es un ser social, por lo tanto, debe servirle no sólo a él mismo, sino a la sociedad a la cual se encuentra integrado (Tintero, 2023).

1.5.4. Machine learning

El Machine Learning es un término amplio que agrupa varias estrategias analíticas cuyo propósito es el desarrollo de algoritmos para extraer información de los datos ya sea para explicación, clasificación o predicción. Pese a que se suele considerar como sinónimo de inteligencia artificial, es importante precisar que la inteligencia artificial es una clasificación aún más amplia que incluye tanto técnicas para el análisis de datos estructurados como el aprendizaje automático y datos no estructurados como procesamiento de lenguaje natural. (Pedrero, Grandón, Ureta, & Cortéz, 2021, pág. 249).

Los algoritmos de Machine Learning pueden ser clasificados en: supervisados, no supervisados y parcialmente supervisados. La principal diferencia entre estos tres tipos de algoritmos es la presencia o ausencia de una variable de resultado o dependiente, a la cual se le denomina marca o etiqueta (Pedrero, Grandón, Ureta, & Cortéz, 2021, pág. 249).

1.5.5. Algoritmos de predicción aplicados al análisis del rendimiento académico en educación superior.

Para la predicción de resultados en el análisis del rendimiento académico, se utilizan diversos algoritmos de Machine Learning para predecir y comprender mejor los patrones detrás de los resultados educativos de los estudiantes.

Entre algunos algoritmos predictivos de Machine Learning más comúnmente aplicados para la temática planteada, se pueden destacar los siguientes con sus respectivas características:

Regresión Logística: Utilizado para predecir resultados binarios o multinomiales, como el éxito o fracaso académico, basándose en variables predictoras (Martínez Pérez, 2020).

Árboles de Decisión y Bosques Aleatorios: Los árboles de decisión son modelos que dividen los datos en nodos basados en características para predecir una variable objetivo. Los bosques aleatorios son conjuntos de múltiples árboles de decisión que promedian las predicciones individuales para mejorar la precisión (Díaz Martínez, 2021).

Redes Neuronales Artificiales: Estos modelos se basan en una estructura de nodos interconectados para aprender patrones complejos en los datos. Son útiles para predecir el rendimiento académico considerando múltiples variables (Gil Vera & Quintero López, 2021).

Algoritmo Naive Bayes: Permite crear modelos predictivos de categorización binaria o múltiple (Castro Solis, 2024).

Algoritmo K-Nearest Neighbors (K-NN): Clasifica los datos basándose en la similitud con los vecinos más cercanos en un espacio de características. Puede agrupar a los estudiantes con características académicas similares (Arango Palacio & Bernal Vélez, 2019).

Support Vector Machines (SVM): SVM es un algoritmo de aprendizaje supervisado que se puede usar tanto para problemas de clasificación como de regresión. Puede ser aplicado para clasificar a los estudiantes en diferentes grupos de rendimiento basándose en datos multidimensionales (Garzón Barrero, Sánchez Pineda, & Londoño Pinilla, 2023).

2. MATERIALES Y MÉTODOS

2.1. Paradigma de la investigación

Por las características del tema propuesto como tesis doctoral en el área de la Inteligencia Artificial y la Educación, se considera utilizar la epistemología del NATURALISMO CIENTÍFICO, que busca explicar que el conocimiento y la cognición se

establecen en términos de procesos naturales y científicos, donde se rechaza de que el conocimiento requiere una base epistemológica sobrenatural o trascendental y que la ciencia es la mejor manera de comprender el mundo y obtener el conocimiento confiable (Jaramillo Echeverri, 2003).

2.2. Enfoque de la investigación

Por las características de la investigación propuesta, el enfoque de investigación aplicada será el enfoque de tipo mixto, es decir tendrá un enfoque cualitativo y cuantitativo (Barrón Tirado, 2015).

Además de que la investigación, se basa en un enfoque empírico y multidisciplinario que combina conceptos y métodos de la educación, la inteligencia artificial y la ciencia de datos.

2.3. Tipo de investigación

El tipo de investigación, principal, que se utilizará en el presente trabajo de investigación es la investigación aplicada y el propositivo transformador, porque la el modelo de predicción desarrollado incidirá en transformar la situación actual de los estudiantes con bajo rendimiento en el Instituto, ya que la propuesta a implementarse permitirá cambiar los aspectos negativos con una solución positiva (Cabrera G., 2018).

De la misma manera se pretende utilizar la Investigación documental, la Investigación exploratoria, la Investigación explicativa y otros que permitirán el logro de los objetivos planteados (Hernández Sampieri, Fernández Collado, & Baptista Lucion, 2016).

2.4. Universo, población y muestra

Universo

El Universo está constituido por el total de estudiantes matriculados en el Instituto Técnico “PIO XII” de la ciudad de Llalagua. El instituto cuenta con tres carreras: Secretariado Ejecutivo, Contaduría General y Sistemas Informáticos. El número de estudiantes es de 622, lo que constituirá el universo sujeto de estudio.

Población

La población está constituida por los estudiantes Segundo y Tercer año de las carreras de Secretariado Ejecutivo, Contaduría General y Sistemas Informáticos del Instituto Técnico “PIO XII” de la ciudad de Llalagua. Por lo que la población es de 286 estudiantes. En la Tabla 5, se puede apreciar el detalle de número de estudiantes que conforman la población bajo estudio.

Muestra

Para la obtención de la muestra se utiliza la expresión matemática del muestreo aleatorio simple y ajustado, que se expresa mediante la ecuación mostrada en la Figura 1.

$$n = \frac{Z^2 pqN}{(NE^2) + (Z^2 pq)} \quad (1)$$

Figura 1. Expresión del Muestreo Aleatorio
Fuente (Torres Bardales, 2018).



III Memoria SOCID “Actas del III Congreso de Investigación Científica” Sociedad Científica de Docentes Universidad Nacional “Siglo XX”

Tabla 1: *Tabla de valores para obtener la muestra*

Z	N	p	q	E	n
Nivel de Confianza (95%)	Población de estudio	Probabilidad de éxito	Probabilidad de error	Error de estimación	Tamaño de la muestra
1.96	286	0.50	0.50	0.05 (5%)	150

Fuente: Elaboración propia

2.5. Métodos, técnicas e instrumentos

● Métodos empíricos

ENCUESTA: Para realizar un diagnóstico de relevamiento de información de los factores que intervienen en el rendimiento académico de los estudiantes.

ENTREVISTA: Para recopilar información de las autoridades y docentes del Instituto con respecto a los resultados de rendimiento académico de los estudiantes en las diferentes asignaturas de las Carreras del Instituto.

EXPERIMENTACIÓN: Se utilizará para el desarrollo del modelo de aprendizaje automático de predicción.

● Métodos teóricos

MÉTODO DEDUCTIVO – INDUCTIVO: Permitirá obtener información sobre la problemática de los factores que intervienen en el rendimiento académico de los estudiantes (Avendaño Osinaga, 2008)

MÉTODO ANÁLISIS – SÍNTESIS: Permitirá descomponer y procesar los datos recolectados de los estudiantes (Avendaño Osinaga, 2008).

2.6. Recopilación de Datos

Para la recolección de datos que permitan brindar información de carácter académico, personal y sociodemográfico de los estudiantes, se elaboró el cuestionario para la aplicación con los estudiantes.

Para la elaboración del cuestionario aplicado, las preguntas se plantearon con la finalidad de recopilar información de los siguientes aspectos, que inicialmente se consideraron como factores influyentes el rendimiento académico de un estudiante:

- Datos Personales
- Datos Académicos
- Datos Socioeconómicos

2.6.1. Validación del cuestionario para la recolección de datos de estudiantes

Para medir la confiabilidad del cuestionario aplicado, se utilizó el estadígrafo alfa de Cronbach, que es una medida estadística utilizada para evaluar la confiabilidad interna de un conjunto de preguntas o ítems en un cuestionario. En otras palabras, ayuda a medir cuán consistentes y relaciones entre sí son las preguntas en una escala de medición (Nina Cuchillo & Nina Cuchillo, 2021).



III Memoria SOCID “Actas del III Congreso de Investigación Científica” Sociedad Científica de Docentes Universidad Nacional “Siglo XX”

$$\alpha = \frac{K}{K-1} \left[\frac{\sum_{i=1}^K s_i^2}{Var(q_i)} \right]$$

Figura 2 Expresión del estadígrafo Alfa de Cronbach
Fuente: (Nina Cuchillo & Nina Cuchillo, 2021).

2.6.2. Recolección de datos del cuestionario aplicado

Para la recolección de datos, se aplicó el cuestionario validado a los estudiantes de Segundo y Tercer Grado de las Carreras de Secretariado Ejecutivo, Contaduría General y Sistemas informáticos.

Para ello se recurrió a la herramienta de Google Formularios para que el llenado de las encuestas por parte de los estudiantes se lo realice por Internet, el enlace del formulario es el siguiente:
https://docs.google.com/forms/d/e/1FAIpQLSc7wAaDu2aDPiAJ4b6iKWXBzRTATZ1kk3zS4FQmnlKrYPzxc/viewform?usp=sf_link

The screenshot shows a Google Form titled "Cuestionario para Estudiantes del Instituto Técnico 'PIO XII'". The form is in Spanish and includes a description of the study, instructions for respondents, and a section for personal data. The personal data section has two fields: "1. Apellidos y nombres:" and "2. Edad:". The form is displayed in a web browser window.

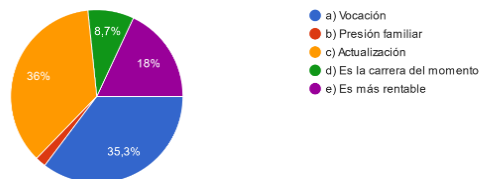
Figura 2. Recolección de datos de los estudiantes
Fuente: Elaboración Propia.

2.6.3. Recolección de datos del rendimiento académico

Se realizó la recolección de datos del rendimiento académico de los estudiantes, consiguiendo los resultados de las calificaciones obtenidas por los estudiantes en las diferentes asignaturas de las Carreras correspondientes de las planillas calificaciones de los documentos del Instituto, como ser planillas de calificaciones, centralizadores de calificaciones y otros. Dichos datos fueron proporcionados por Dirección Académica del Instituto.

[illegible]

Se realizó el siguiente análisis e interpretación de los resultados del cuestionario aplicado a los estudiantes del Instituto Técnico “PÍO XII” de los indicadores más relevantes.



24



III Memoria SOCID “Actas del III Congreso de Investigación Científica” Sociedad Científica de Docentes Universidad Nacional “Siglo XX”

3. RESULTADOS (PROPUESTA)

Para el presente proyecto se consideró implementar el MODELO DE REGRESIÓN LOGÍSTICA, ya que permite modelar la relación entre una variable dependiente binaria y una o más variables independientes y se utiliza cuando el resultado es binario (por ejemplo, éxito/fallo, sí/no, 1/0). Además de que el modelo permite clasificar.

El modelo matemático de regresión logística se puede describir mediante la siguiente ecuación:

$$P(y|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

Donde:

- $P(y=1|X)$ es la probabilidad de que el evento de interés ocurra (por ejemplo, que la respuesta sea 1) dado el conjunto de variables independientes X .
- β_0 es el intercepto (término independiente).
- $\beta_1, \beta_2, \dots, \beta_n$ son los coeficientes asociados a las variables independientes $x_1, x_2, \dots, x_n$.
- $x_1, x_2, \dots, x_n$ son las variables independientes (variables predictors).

3.1. Modelo propuesto de predicción del rendimiento académico

En base a las consideraciones teóricas de los párrafos anteriores, se pudo establecer el modelo de predicción del rendimiento académico propuesto en la investigación, modelo que puede clasificar a estudiantes que tiene Rendimiento Normal o Rendimiento Bajo.

El modelo inicialmente permite identificar las variables que se constituyen como factores influyentes en el rendimiento académico para la selección de dichas variables. En base a esas variables identificadas, se construye el modelo de predicción del rendimiento académico.

Para ello el modelo de regresión logística realizó:

- La clasificación de las variables importantes que se considerarán en el modelo final a partir de las variables iniciales de la respectiva tabla, es decir el modelo permitió seleccionar las variables que tienen mayor incidencia en el rendimiento académico, para ello el modelo consideró para su selección los coeficientes proporcionados por el modelo de regresión logística. El modelo implementado para seleccionar las características es donde se muestra expresión matemática de regresión logística, con su respectivo estudio.

$$\begin{aligned} \text{Logit}(P) = & \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 \\ & + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} + \beta_{15} X_{15} + \beta_{16} X_{16} + \beta_{17} X_{17} \\ & + \beta_{18} X_{18} + \beta_{19} X_{19} + \beta_{20} X_{20} + \beta_{21} X_{21} + \beta_{22} X_{22} + \beta_{23} X_{23} + \beta_{24} X_{24} + \beta_{25} X_{25} \\ & + \beta_{26} X_{26} + \beta_{27} X_{27} + \beta_{28} X_{28} + \beta_{29} X_{29} + \beta_{30} X_{30} + \beta_{31} X_{31} + \beta_{32} X_{32} + \beta_{33} X_{33} + \beta_{34} X_{34} \\ & + \beta_{35} X_{35} + \beta_{36} X_{36} + \beta_{37} X_{37} + \beta_{38} X_{38} + \beta_{39} X_{39} + \beta_{40} X_{40} + \beta_{41} X_{41} \end{aligned} \quad (4)$$

Figura 5. Modelo inicial para selección de características importantes

Fuente: Elaboración propia

- Una vez identificadas las variables predictoras, se implementó el modelo de predicción del rendimiento académico mediante el modelo de regresión logística mediante la siguiente expresión matemática:



III Memoria SOCID “Actas del III Congreso de Investigación Científica” Sociedad Científica de Docentes Universidad Nacional “Siglo XX”

$$\text{Logit}(P) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} + \beta_{15} X_{15} + \beta_{16} X_{16} + \beta_{17} X_{17} + \beta_{18} X_{18} + \beta_{19} X_{19} + \beta_{20} X_{20} + \beta_{21} X_{21} + \beta_{22} X_{22} + \beta_{23} X_{23} \quad (5)$$

Figura 6. Modelo propuesto de predicción del Rendimiento Académico

Fuente: Elaboración propia

- Implementación del modelo de regresión logística

Producto de la implementación del Modelo de Regresión Logística para la predicción del rendimiento académico que se realizó en Python y en los resultados obtenidos de la implementación del modelo

- Resultados obtenidos de la implementación

```

Terminal 1/A X
Accuracy: 0.8666666666666667
Confusion Matrix:
[[16  2]
 [ 2 10]]
Classification Report:

```

	precision	recall	f1-score	support
0	0.89	0.89	0.89	18
1	0.83	0.83	0.83	12
accuracy			0.87	30
macro avg	0.86	0.86	0.86	30
weighted avg	0.87	0.87	0.87	30

Figura 7. Resultados de la Implementación del Modelo de Predicción

Fuente: Elaboración propia

- Coeficientes del modelo de regresión logística obtenida

Luego del entrenamiento realizado con el modelo de Regresión Logística, se obtuvieron los siguientes coeficientes del modelo matemático

Terminal de IPython

```

Terminal 1/A X
Intercepto: 0.224741
Coeficientes:
Coeficiente B1= -2.538767
Coeficiente B2= -0.384553
Coeficiente B3= 0.745004
Coeficiente B4= -0.824087
Coeficiente B5= 0.521478
Coeficiente B6= 0.695532
Coeficiente B7= 0.388551
Coeficiente B8= 0.855316
Coeficiente B9= -0.501497
Coeficiente B10= -0.563063
Coeficiente B11= -0.527943
Coeficiente B12= -0.443124
Coeficiente B13= -0.075197
Coeficiente B14= 0.443570
Coeficiente B15= 0.225071
Coeficiente B16= 0.552286
Coeficiente B17= 0.106103
Coeficiente B18= 0.342694
Coeficiente B19= -0.207298
Coeficiente B20= -0.011372
Coeficiente B21= -0.163828
Coeficiente B22= -0.202774
Coeficiente B23= -0.241850
  
```

Figura 8. Coeficientes del modelo de Regresión Logística
Fuente: Elaboración propia

Reemplazando los valores obtenidos en el modelo propuesto, con los coeficientes, el modelo matemático es:

$$\begin{aligned}
 \text{Logit}(P) = & 0.224741 - 2.538767 X_1 - 0.384553 X_2 + 0.745004 X_3 - 0.824087 X_4 \\
 & + 0.521478 X_5 + 0.695532 X_6 + 0.388551 X_7 + 0.855316 X_8 - 0.501497 X_9 \\
 & - 0.563063 X_{10} - 0.527943 X_{11} - 0.443124 X_{12} - 0.075197 X_{13} \\
 & + 0.443570 X_{14} + 0.225071 X_{15} + 0.552286 X_{16} + 0.106103 X_{17} \\
 & + 0.342694 X_{18} - 0.207298 X_{19} - 0.011372 X_{20} - 0.163828 X_{21} \\
 & - 0.202774 X_{22} - 0.241850 X_{23}
 \end{aligned}
 \tag{6}$$

Figura 9. Modelo de Predicción del Rendimiento Académico
Fuente: Elaboración propia

- Uso del modelo para la predicción del rendimiento académico



III Memoria SOCID “Actas del III Congreso de Investigación Científica” Sociedad Científica de Docentes Universidad Nacional “Siglo XX”

Una vez obtenido el modelo, éste se usa para predecir el Rendimiento Académico de un estudiante, ingresando por teclado los 23 factores identificados.

```
Terminal de IPython
Terminal I/A X
CALIFICACIÓN IB 2024 ==> 15
CUMPLES CON TUS TAREAS DE MANERA PUNTUAL: Siempre:3, A veces:2, Nunca:1 ==> 2
ESTUDIAS EN OTRO LUGAR: Si:1, No:0 ==> 1
ACCESO A COMPUTADORA: Si:1, No:0 ==> 0
COLEGIO DE PROCEDENCIA: Área urbana:0, Área dispersa:1 ==> 1
FALTAS A CLASES: No me faltó nunca:1, Me faltó algunas veces:2, Siempre me faltó:3 ==> 2
CASA DONDE VIVES: Propia:1, De mis padres:2, Alquiler:3, Anticrético:4, Otro:5 ==> 4
ESTADO CIVIL: Soltero:1, Casado:2, Concubinado:3, Divorciado:4 ==> 2
MOTIVO DE INGRESO AL INSTITUTO: Vocación:1, Presión familiar:2, Actualización:3, Es la carrera del momento:4, Es más rentable:5 ==> 1
ZONA DONDE VIVES: Llalagua:1, Siglo XX:2, Salvador:3, Villarroel:4, Cancañiri:5, 22 de Diciembre:6, Sakamarka:7,
9
PROCEDENCIA: Municipio de Llalagua:0, Del interior del país:1 ==> 1
LA INFRAESTRUCTURA DEL INSTITUTO CONSIDERAS QUE ES: Muy adecuado:1, Adecuado:2, Poco adecuado:3 ==> 2
GESTIÓN DE INGRESO ==> 2018
LA BIBLIOTECA MEJORARÍA TU FORMACIÓN: Mucho:1, Poco:2, Nada:3 ==> 2
ASISTES PUNTUALMENTE A CLASES: Siempre:1, A veces:2, Nunca:3 ==> 2
¿CUAL ES LA FINALIDAD DEL EXAMEN: Aprobar el examen:1, Vencer el curso:2, Medir tu grado de conocimientos y habilidades
TRABAJAS EN EL DÍA: Si:1, No:0 ==> 0
EDAD ==> 45
EL EQUIPAMIENTO DE LABORATORIOS CONSIDERAS QUE ES: Muy adecuado:1, Adecuado:2, Poco adecuado:3 ==> 2
GÉNERO: Masculino:1, Femenino:0 ==> 1
MOMENTO DE EVALUACIÓN: Al final del bimestre:1, Al inicio del bimestre:2, Durante el bimestre:3 ==> 1
CARRERA: Secretariado Ejecutivo:1, Contaduría General:3, Sistemas Informáticos:2 ==> 3
EL INTERNET EN EL INSTITUTO MEJORARÍA TU FORMACIÓN: Mucho:1, Poco:2, Nada:3 ==> 2

RESULTADOS DE LA PREDICCIÓN
Predicción de Rendimiento Académico= 1
RENDIMIENTO ACADÉMICO BAJO
```

Figura 10. Predicción del Modelo para el caso de Rendimiento Académico Bajo
Fuente: Elaboración propia

```
Terminal I/A X
In [181]: runfile('C:/ProgramasPython/PredicirV112.py', wdir='C:/ProgramasPython')
CALIFICACIÓN IB 2024 ==> 90
CUMPLES CON TUS TAREAS DE MANERA PUNTUAL: Siempre:3, A veces:2, Nunca:1 ==> 3
ESTUDIAS EN OTRO LUGAR: Si:1, No:0 ==> 0
ACCESO A COMPUTADORA: Si:1, No:0 ==> 1
COLEGIO DE PROCEDENCIA: Área urbana:0, Área dispersa:1 ==> 0
FALTAS A CLASES: No me faltó nunca:1, Me faltó algunas veces:2, Siempre me faltó:3 ==> 1
CASA DONDE VIVES: Propia:1, De mis padres:2, Alquiler:3, Anticrético:4, Otro:5 ==> 1
ESTADO CIVIL: Soltero:1, Casado:2, Concubinado:3, Divorciado:4 ==> 1
MOTIVO DE INGRESO AL INSTITUTO: Vocación:1, Presión familiar:2, Actualización:3, Es la carrera del momento:4, Es más rentable:5 ==> 1
ZONA DONDE VIVES: Llalagua:1, Siglo XX:2, Salvador:3, Villarroel:4, Cancañiri:5, 22 de Diciembre:6, Sakamarka:7, Catavi:8, Uncia:9,
1
PROCEDENCIA: Municipio de Llalagua:0, Del interior del país:1 ==> 0
LA INFRAESTRUCTURA DEL INSTITUTO CONSIDERAS QUE ES: Muy adecuado:1, Adecuado:2, Poco adecuado:3 ==> 1
GESTIÓN DE INGRESO ==> 2022
LA BIBLIOTECA MEJORARÍA TU FORMACIÓN: Mucho:1, Poco:2, Nada:3 ==> 1
ASISTES PUNTUALMENTE A CLASES: Siempre:1, A veces:2, Nunca:3 ==> 1
¿CUAL ES LA FINALIDAD DEL EXAMEN: Aprobar el examen:1, Vencer el curso:2, Medir tu grado de conocimientos y habilidades:3 ==> 3
TRABAJAS EN EL DÍA: Si:1, No:0 ==> 0
EDAD ==> 25
EL EQUIPAMIENTO DE LABORATORIOS CONSIDERAS QUE ES: Muy adecuado:1, Adecuado:2, Poco adecuado:3 ==> 1
GÉNERO: Masculino:1, Femenino:0 ==> 1
MOMENTO DE EVALUACIÓN: Al final del bimestre:1, Al inicio del bimestre:2, Durante el bimestre:3 ==> 3
CARRERA: Secretariado Ejecutivo:1, Contaduría General:3, Sistemas Informáticos:2 ==> 2
EL INTERNET EN EL INSTITUTO MEJORARÍA TU FORMACIÓN: Mucho:1, Poco:2, Nada:3 ==> 1

RESULTADOS DE LA PREDICCIÓN
Predicción de Rendimiento Académico= 0
RENDIMIENTO ACADÉMICO NORMAL
```

Figura 11. Predicción del Modelo para el caso de Rendimiento Académico Normal
Fuente: Elaboración propia

4. CONCLUSIONES

- Se pudo obtener los datos relevantes de los estudiantes del Instituto Técnico “PÍO XII” aplicando el cuestionario, donde se pudo recolectar información de diferentes aspectos de los estudiantes. Lo anteriormente mencionado se logró mediante el diseño y validación del cuestionario aplicado.
- Se logró realizar el preprocesamiento y análisis exploratorio de los datos obtenidos para la selección de variables, para garantizar la calidad de los datos mediante herramientas de ciencia de datos, donde se obtuvo los datos seleccionados, codificados, estandarizados y balanceados para la posterior obtención del modelo de predicción.
- Se logró aplicar el modelo de Regresión Logística y herramientas de Ciencia de Datos para identificar los factores y patrones que influyen en el rendimiento académico. Los factores que influyen en el rendimiento académico se obtuvieron en base a la importancia de incidencia de las variables independientes en la variable de Rendimiento Académico.



III Memoria SOCID “Actas del III Congreso de Investigación Científica” Sociedad Científica de Docentes Universidad Nacional “Siglo XX”

- Se identificaron 23 factores que influyen en el rendimiento académico de un estudiante, las cuales están ordenadas de acuerdo a su importancia de incidencia de mayor a menor.

Tabla 2. Variables identificadas como factores que influyen en el rendimiento académico

No.	Variable
1	Calificaicion1B2024
2	CumpleTarea
3	Estudias
4	AccesoComputadora
5	ProcedenciaColegio
6	FaltasClases
7	CasaDondeVives
8	EstadoCivil
9	MotivoIngreso
10	ZonaDondeVives
11	Procedencia
12	Infraestructura
13	GestionIngreso
14	MejoraBiblioteca
15	PuntualidadClases
16	FinalidadExamen
17	TrabajasDia
18	Edad
19	EquipamientoLaboratorio
20	Genero
21	MomentoEvaluacion
22	Carrera
23	MejoraInternet

Fuente: Elaboración propia



III Memoria SOCID “Actas del III Congreso de Investigación Científica” Sociedad Científica de Docentes Universidad Nacional “Siglo XX”

Los patrones de rendimiento académico bajo y rendimiento académico normal se obtuvieron en base al análisis de las tasas de conversión de cada variable, lo que permitió la Identificación de Patrones de Rendimiento Académico.

Tabla 3. Patrones de Rendimiento Académico

No.	Variable	Patrón de Rendimiento Bajo	Patrón de Rendimiento Normal
1	CumpleTarea	Estudiantes que nunca o a veces cumplen con sus tareas	Estudiantes que siempre cumplen con sus tareas
2	Estudias	Que si estudian	Que no estudian
3	AccesoComputadora	No tiene acceso a computadora en casa	Tiene acceso a computadora en casa
4	ProcedenciaColegio	Procedencia del colegio sea del área dispersa	Procedencia del colegio sea del área urbana
5	FaltasClases	Los que siempre se faltan o se faltan algunas veces	Los que no se faltan
6	CasaDondeVives	Los que viven en alquiler y otros	Los que tienen casa propia, de sus padres y anticrético
7	EstadoCivil	Que sean casados, concubinados o divorciados	Que sean solteros
8	MotivoIngreso	Por presión familiar y los que consideran que es más rentable	Los que ingresaron por vocación, actualización y consideran que es la carrera del momento
9	ZonaDondeVives	Estudiantes que viven en zonas alejadas del Instituto	Estudiantes que viven en zonas cercanas al Instituto
10	Procedencia	Del interior del país	Del municipio de Llallagua
11	Infraestructura	Los que consideran que la infraestructura es muy adecuada	Los que consideran que la infraestructura no es adecuada
12	MejoraBiblioteca	Los estudiantes que consideran que la biblioteca no mejora su formación	Los estudiantes que consideran que la biblioteca mejora en poco y mucho en su formación
13	PuntualidadClases	Estudiantes que a veces y nunca son puntuales	Estudiantes que siempre son puntuales
14	FinalidadExamen	Estudiantes que la finalidad es vencer el curso y solo aprobar el examen	Estudiantes que consideran que el examen permite medir el grado de conocimientos y habilidades
15	TrabajosDia	Estudiantes que trabajan en el día	Estudiantes que no trabajan en el día
16	EquipamientoLaboratorio	Los que consideran que el equipamiento no es adecuado	Los que consideran que el equipamiento es adecuado
17	Genero	Estudiantes de género Masculino	Estudiantes de género Femenino
18	MomentoEvaluacion	Al final del bimestre	Al inicio y durante el bimestre
19	Carrera	En Contaduría General	En Sistemas Informáticos y Secretariado Ejecutivo
20	MejoraInternet	Los estudiantes que consideran que el acceso a Internet en el Instituto no mejora su formación	Los estudiantes que consideran que el acceso a Internet en el Instituto si mejora su formación

No.	Variable	Patrón de Rendimiento Bajo	Patrón de Rendimiento Normal
21	Edad	Mayores a 25 años	Menores a 25 años
22	CalificacionIB2024	Calificación menor a 63 puntos	Calificación mayor a 63 puntos
23	GestionIngreso	Estudiantes que ingresaron antes de la gestión 2020	Estudiantes que ingresaron desde la gestión 2020 hacia adelante

Fuente: Elaboración propia

Hallazgos Importantes: Se puede señalar que los factores más influyentes en el bajo rendimiento académico son los siguientes:

- El promedio de reprobación último de calificación obtenida
- Si el estudiante estudia en otra institución
- Estudiantes que no tienen acceso a un equipo de computación en su domicilio
- La procedencia del colegio del área dispersa
- El lugar de procedencia del estudiante del interior
- Si la casa donde vive el estudiante no es propia
- El estado civil del estudiante si es casado o concubinado
- La asistencia y puntualidad a clases es irregular



III Memoria SOCID “Actas del III Congreso de Investigación Científica” Sociedad Científica de Docentes Universidad Nacional “Siglo XX”

- La zona donde vive es alejada del Instituto
- Si el estudiante trabaja en el día
- La edad del estudiante, estudiantes mayores a 25 años tienen bajo rendimiento
- La carrera que cursa el estudiante, si es estudiante de Contaduría General, tiene más tendencia al bajo rendimiento.
- La inexistencia de una biblioteca en el Instituto
- La inexistencia de Internet en el Instituto
- La infraestructura y equipamiento de los laboratorios del instituto, los cuales se consideran poco adecuados.

REFERENCIAS

- Arango Palacio, M. I., & Bernal Vélez, M. J. (2019, agosto 15). Predicción de enfermedades del corazón usando el algoritmo K-Nearest Neighbors. *Revista Universidad EAFIT*, 3(1), 10–15. <https://www.researchgate.net/publication/364476395>
- Avendaño Osinaga, R. (2008). *Metodología de la investigación* (2.ª ed.). Educación y Cultura.
- Barrón Tirado, M. C. (2015). Concepciones epistemológicas y práctica docente: Una revisión. *Revista de Docencia Universitaria*, 13(1), 35–56. <https://doi.org/10.4995/redu.2015.6436>
- Cabrera G., G. C. (2018). *Elementos básicos del estudio y la investigación*. Latinas Editores.
- Castro Solis, Y. E. (2024, enero 15). Comparación de algoritmos de machine learning. *Revista Electrónica sobre Tecnología, Educación y Sociedad*, 11(21), 120–124. <https://orcid.org/0009-0001-1806-5641>
- Chucos Baquerizo, N., & Vega Ventocilla, E. J. (2022, febrero 15). Evaluación de algoritmos de machine learning. *Revista Multidisciplinar Ciencia Latina*, 6(1). <https://ciencialatina.org/index.php/cienciala/article/view/1843/2638>
- Dennett, D. (1995). *La conciencia explicada: Una teoría interdisciplinar* (1.ª ed.). Ediciones Paidós. <https://datelobueno.com/wp-content/uploads/2014/05/La-conciencia-explicada.pdf>
- Díaz Martínez, M. A. (2021, marzo 30). Árboles de decisión como metodología para determinar el... *Revista Lasallista de Investigación*, 18(2), 94–100. <http://www.scielo.org.co/pdf/rlsi/v18n2/1794-4449-rlsi-18-02-94.pdf>
- Garzón Barrero, J., Sánchez Pineda, N., & Londoño Pinilla, D. F. (2023, diciembre 17). Evaluación comparativa de los algoritmos. *Ciencia e Ingeniería Neogranadina*, 33(2), 131–148. <http://www.scielo.org.co/pdf/cein/v33n2/1909-7735-cein-33-02-131.pdf>
- Gil Vera, V., & Quintero López, C. (2021, junio 10). Predicción del rendimiento académico estudiantil con redes. *Información Tecnológica*, 32(6), 221–224. <https://www.scielo.cl/pdf/infotec/v32n6/0718-0764-infotec-32-06-221.pdf>
- Guerrero Pino, G. (2015). El sistema epistemológico de Quine. *Universitas Philosophica*, 64(32), 26–39. <http://www.scielo.org.co/pdf/unph/v32n64/v32n64a03.pdf>
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2016). *Metodología de la investigación* (4.ª ed.). McGraw-Hill.
- Jaramillo Echeverri, L. G. (2003). ¿Qué es epistemología? Mi mirar epistemológico y el progreso de la ciencia. *Cinta de Moebio*, (18), 174–178. <file:///C:/Users/Jorge%20Villc%C3%A1ez/Downloads/publicadorcdm,+Journal+manager,+26135-85751-1-CE.pdf>



III Memoria SOCID “Actas del III Congreso de Investigación Científica” Sociedad Científica de Docentes Universidad Nacional “Siglo XX”

Kuhn, T. S. (2004). La estructura de las revoluciones científicas (2.^a ed.). Fondo de Cultura Económica. https://www.bfa.fcnym.unlp.edu.ar/catalogo/doc_num.php?explnum_id=2721

Martínez Pérez, J. R. (2020, marzo 18). Regresión logística y predicción del bajo rendimiento académico de estudiantes en la carrera Medicina. Revista Electrónica Dr. Zoilo Marinello Vidaurreta, 45(4), 25–30. https://revzoilomarinellosldcu/index.php/zmv/article/view/2230/pdf_691

Mora Pineda, J. (2022, diciembre 16). Modelos predictivos en salud basados en aprendizaje de máquina. Revista Médica Clínica Las Condes, 33(1), 583–590. <https://www.sciencedirect.com/science/article/pii/S0716864022001213>

Nina Cuchillo, J., & Nina Cuchillo, E. E. (2021, mayo 23). Análisis de confiabilidad: Cálculo de coeficiente alfa de Cronbach usando el software SPSS. https://d1wqtxts1xze7.cloudfront.net/67404272/NINA_CUCHILLO_CONFIABILIDAD_CRONBACH_SPSS-libre.pdf

Pedraja Rejas, L., Rodríguez Ponce, E., & Labraña, J. (2022, julio 25). ¿Qué sabemos de la cultura académica? Revisión del concepto en la literatura en educación superior. Revista de la Universidad de Tarapacá, 48(2), 1–10. <https://doi.org/10.1590/S1678-4634202248240831>

Quine, W. O. (1959). Palabra y objeto (1.^a ed.). Editorial Labro S. A. https://kupdf.net/download/120542497-quine-palabra-y-objeto_58a482db6454a7ef0ab1e8e3_pdf

Ramírez, J., Gómez, J. H., & Ortega Herrera, A. D. (2023, enero 14). Análisis de las técnicas de machine learning para la predicción de deserción de estudiantes universitarios. Investigación y Acción, 3(2), 17–25. <https://investigaciones.uniatlantico.edu.co/revistas/index.php/Invefor/article/view/3833>

Tacilla Cárdenas, I., Vásquez Villanueva, S., Verde Ávalos, E. E., & Colque Díaz, E. (2020, junio 7). Rendimiento académico: Universo muy complejo para el quehacer pedagógico. Muro de la Investigación, 3(1), 53–60. <https://doi.org/10.17162/rmi.v5i2.1325>

Torres Bardales, C. (2018). Orientaciones básicas de metodología de la investigación (3.^a ed.). El Fulgor.

SOBRE EL AUTOR

Docente investigador con artículos publicados en las Revistas:

Revista SECON-INF del Instituto Técnico “PÍO XII” con el artículo: Virtualización de Sistemas Operativos.

Revista CIENCIA Y TECNOLOGÍA INFORMÁTICA de la Carrera Ingeniería Informática de la Universidad Nacional “Siglo XX” con el artículo: Aplicación de las Matemáticas en el Modelado de un Brazo Robótico y Simulación con Matlab y Simulink.

Revista CIENCIA Y TECNOLOGÍA “SIGLO XX” de la Dirección General de Investigación de la Universidad Nacional “Siglo XX” con el artículo: Casco Minero Inteligente.

Revista SINERGIA del Centro de Conocimiento Integral y de Innovación con el artículo: Modelos de Aprendizaje Automático en Inteligencia Artificial.

Revista SINERGIA del Centro de Conocimiento Integral y de Innovación con el artículo: Algoritmos Predictivos de Machine Learning y su Impacto en el Rendimiento Académico de la Educación Superior.



III Memoria SOCID
“Actas del III Congreso de Investigación Científica”
Sociedad Científica de Docentes
Universidad Nacional “Siglo XX”



Fig 12. Fotografía de presentación de la ponencia