



**UNIVERSIDAD CÉSAR VALLEJO**

**FACULTAD DE INGENIERÍA Y ARQUITECTURA  
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Machine learning para precedir el rendimiento académico en la IE  
N<sup>o</sup>116 Abraham Valdelomar San Juan de Lurigancho 2022.

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:**

Ingeniero de Sistemas

**AUTORES:**

Acosta Portocarrero, Antony Christian ([orcid.org/0000-0001-6191-1666](https://orcid.org/0000-0001-6191-1666))

Ruiz Vargas, Paul ([orcid.org/0000-0001-9003-2695](https://orcid.org/0000-0001-9003-2695))

**ASESOR:**

Mg. Pacheco Pumaleque, Alex Abelardo ([orcid.org/0000-0001-9721-0730](https://orcid.org/0000-0001-9721-0730))

**LÍNEA DE INVESTIGACIÓN:**

Sistemas Información y Comunicaciones

**LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:**

Desarrollo económico, empleo y emprendimiento

LIMA – PERÚ

2022

## Dedicatoria

Esta investigación está dedicada a nuestros hijos e hijas, quienes han sido y seguirán siendo nuestro motor que nos permita seguir adelante por el camino de la vida y construir nuestro futuro con la honestidad y humildad.

## Agradecimiento

A través de estas líneas expresamos nuestro profundo agradecimiento al Mg. Pacheco Pumaleque, Alex Abelardo por su contribución como asesor y mentor durante el desarrollo de esta tesis, fundamental para el éxito de este proyecto.

También nuestro reconocido agradecimiento a vuestros padres por haber hecho de nosotros persona perseverante en nuestros objetivos.

Y asimismo agradecer a los docentes de la Universidad, colegas que compartieron críticas y posiciones sobre el desarrollo de la investigación.

## Índice de Contenidos

Carátula .....	i
Dedicatoria .....	ii
Agradecimiento.....	iii
Índice de Contenidos .....	iv
Índice de Tablas.....	v
Indicé de Figuras .....	vi
Resumen.....	vii
Abstract.....	viii
I. INTRODUCCIÓN .....	1
II. MARCO TEÓRICO .....	5
III. METODOLOGÍA .....	17
3.1 Tipo y diseño de investigación .....	18
3.1.1. Tipo de Investigación: .....	18
3.1.2. Diseño de Investigación:.....	18
3.2 Variable y operacionalización .....	19
3.2.1. Variable Independiente Matching Learning.....	19
3.2.2. Definición Conceptual de la Variable Dependiente: Rendimiento Académico. ....	19
3.3 Población, muestra y muestreo probabilístico .....	20
3.4. Técnicas e instrumentos de recolección de datos .....	21
3.5. Procedimiento.....	22
3.6. Método de análisis de datos .....	23
3.7. Aspectos éticos.....	23
IV. RESULTADOS .....	24
V. DISCUSIÓN .....	43
VI. CONCLUSIONES .....	48
VII. RECOMENDACIONES .....	50
REFERENCIAS .....	52
ANEXOS.....	55

## Índice de Tablas

Tabla 1 Matriz de confusión .....	22
Tabla 2 Población de estudio.....	29
Tabla 3 Ficha técnica del instrumento .....	30
Tabla 4 Expertos que validaron herramientas de recolección de datos.....	31
Tabla 5 Tabla de calificaciones.....	34
Tabla 6 Matriz de confusión – Árbol de decisión .....	34
Tabla 7 Matriz de observación- Árbol de decisión .....	35
Tabla 8 Matriz de confusión – Máquina de Vectores SVM .....	35
Tabla 9 Matriz de observación- Máquina de Vectores SVM .....	36
Tabla 10 Matriz de Confusión- Red Bayesiana.....	36
Tabla 11 Matriz de observación- Red Bayesiana .....	37
Tabla 12 Tabla cruzada – cálculo de precisión con algoritmo árbol de decisión .....	38
Tabla 13 Tabla cruzada – cálculo de precisión con algoritmo SVM.....	39
Tabla 14 Tabla cruzada – cálculo de precisión con algoritmo Red – Bayesiana .....	40
Tabla 15 Cuadro comparativo de resultados según el indicador precisión .....	41
Tabla 16 Tabla cruzada – cálculo de sensibilidad con árbol de decisión.....	41
Tabla 17 Tabla cruzada – cálculo de sensibilidad con algoritmo Máquina de Vectores de Soporte - SVM .....	42
Tabla 18 Tabla cruzada – cálculo de sensibilidad con algoritmo Red – Bayesiana .....	44
Tabla 19 Cuadro comparativo de resultados según el indicador precisión .....	45
Tabla 20 Tabla cruzada – cálculo de especificidad con árbol de decisión .....	45
Tabla 21 Tabla cruzada – cálculo de especificidad con algoritmo SVM .....	46
Tabla 22 Tabla cruzada – cálculo de especificidad con algoritmo Red – Bayesiana.....	47
Tabla 23 Cuadro comparativo de resultados según el indicador precisión .....	48
Tabla 24 Resumen de cuadro comparativo de Algoritmos .....	49
Tabla 25 Medida de Kappa De cohen-Árbol de decisión .....	49
Tabla 26 Medida de Kappa De cohen-SVM.....	50
Tabla 27 Medida de Kappa De Cohen-Red Bayesiana .....	50

## Indicé de Figuras

Figura 1 Ejemplo de clasificación con SVM .....	17
Figura 2 Kernel Lineal, Kernel polinómico y Gaussian kernel respectivamente.....	18
Figura 3 Estructura de árbol de decisión .....	18
Figura 6 fases resumidas de la metodología KDD.....	22
Figura 7 Fórmula para calcular la sensibilidad.....	23
Figura 8 Fórmula para calcular la Especificidad .....	23
Figura 9 Fórmula para calcular la precisión .....	24
Figura 10 Diseño pre-experimental con un solo grupo .....	25
Figura 11 Nivel de Kappa de Cohen .....	36

## Resumen

En esta investigación se elaboró un modelo predictivo de Machine Learning, para predecir el rendimiento académico de los alumnos del 4to. Y 5to. Año de secundaria de la IE N°116 “Abraham Valdelomar – San Juan de Lurigancho, para lograr el objetivo se utilizó la metodología KDD, así mismo herramientas como SPSS statistic Ver. 26 y SPSS Modeler Ver.18.

El objetivo de la investigación es determinar en qué porcentaje Machine Learning permite predecir el rendimiento académico con precisión, sensibilidad y especificidad, con el fin de poder identificar a los alumnos con probabilidad de éxito o fracaso.

Para la investigación se utilizó una población de 101 alumnos, y se utilizó la totalidad de la población como muestra. El estudio es de tipo aplicada, con un diseño de investigación experimental de tipo pre-experimental de un solo grupo, lo que se obtendrá luego de aplicar el modelo predictivo de Machine Learning se podrá observar con los resultados y validar la medición.

Se obtuvo como resultado para los indicadores de precisión, sensibilidad y especificidad para el algoritmo Árbol de Decisión los siguientes valores porcentuales: (Precisión-84.73, Sensibilidad-61.54, Especificidad-90.47), Máquina de vectores SVM (Precisión-100.00, Sensibilidad-100.00, Especificidad-100.00) y Red-Bayesiana (Precisión-99.60, Sensibilidad-99.00, Especificidad-99.75), con base a estos resultados se concluye que el algoritmo que pudo predecir el rendimiento académico de los estudiantes de la IE N°116 “Abraham Valdelomar” con mejores resultados fue Máquina de vectores (SVM) con un valor de 100%.

**Palabras clave:** Machine Learning, rendimiento académico, métricas de precisión.

## Abstract

In this research, a predictive model of Machine Learning was developed to predict the academic performance of 4th grade students. And 5th. High school year of IE No. 116 "Abraham Valdelomar - San Juan de Lurigancho, to achieve the objective the KDD methodology was used, as well as tools such as SPSS statistic Ver. 26 and SPSS Modeler Ver.18.

The objective of the research is to determine in what percentage Machine Learning allows predicting academic performance with precision, sensitivity and specificity, in order to be able to identify students with a probability of success or failure.

For the investigation, a population of 101 students was used, and the entire population was used as a sample. The study is of an applied type, with a pre-experimental experimental research design of a single group, what will be obtained after applying the Machine Learning predictive model can be observed with the results and validate the measurement.

The following percentage values were obtained for the indicators of precision, sensitivity and specificity for the Decision Tree algorithm: (Precision-84.73, Sensitivity-61.54, Specificity-90.47), SVM vector machine (Precision-100.00, Sensitivity-100.00 , Specificity-100.00) and Red-Bayesiana (Presicion-99.60, Sensitivity-99.00, Specificity-99.75), based on these results it is concluded that the algorithm that could predict the academic performance of the students of the IE N°116 "Abraham Valdelomar" with the best results was Vector Machine (SVM) with a value of 100%.

**Keywords:** Machine Learning, academic performance, precision metrics.



## I. INTRODUCCIÓN

Una de las tareas más importantes a nivel mundial es la observación simultánea de los datos con el procesamiento estadístico. En tal sentido como enfoque principal de este estudio es entender como el aprendizaje automático se transforma en la base del procesamiento de datos a gran escala, en los sectores siguientes: Tecnologías de la información, salud, educación y turismo. Estos sectores mencionados han encontrado en las últimas décadas solución a esta compleja tarea del aprendizaje y predicción, disciplinas procedentes de la ingeniería artificial. A esta herramienta informática se le llama Machine Learning. (C. Batanero,2019).

Su aplicación en el sector Educación sería de mucha utilidad, ya que los sistemas son capaces de la evaluación de cada estudiante y a la vez realizando un plan de trabajo con necesidades, también permitirá al maestro enfocarse en reemplazar los defectos de cada estudiante. Por lo expresado líneas arriba nos permite afirmar que unas de las tareas del Machine Learning es dotar a la educación de un recurso faltante en la misma, que para muchos es la calidad.

En el Perú, el desempeño estudiantil es uno de los temas más discutidos y de mayor preocupación para nuestras autoridades y padres de familia. No solo en nuestro país, sino también en muchos otros países latinoamericanos y otros continentes (Lamas, 2015).

Lo que significaría el éxito o el fracaso del estudiante, que es considerado importante la institución que lo acoge y sobre todo del futuro del país. Por lo tanto, es necesario identificar objetivamente a todos los estudiantes que se desempeñan bien o mal en su ciclo de aprendizaje para sugerir mejoras en su educación.

El COVID-19 ha provocado muchos cambios en el mundo. Algunos de estos han tenido un impacto directo en la educación, pasando del aprendizaje presencial al aprendizaje virtual, y han impactado el rendimiento escolar de los adolescentes y, por lo tanto, se están extendiendo por todo el país y en todo el mundo, superando las expectativas en el hogar y en las escuelas. (Ramírez Abad, 2022).

En este sentido, los algoritmos de aprendizaje automático nos permiten predecir los resultados del aprendizaje con una precisión similar, lo que nos permite identificar a los estudiantes que corren el riesgo de tener un bajo rendimiento.

Asimismo, es razonable decir lo siguiente: La problemática de la IE N°116 "ABRAHAM VALDLEOMAR" es la deserción de los alumnos por problemas de pandillaje, drogadicción y problemas económicos a que conlleva a los estudiantes a no culminar sus estudios secundarios, ante esta problemática se plantea brindar diversas opciones educativas y así poder lograr un buen desarrollo estudiantil y pueda tener buenos profesionales para que tengan una mejor calidad de vida. Se justifica teóricamente ya que permite medir las habilidades y talentos (resultados académicos) de los alumnos de la IE No. 116 Abraham Valdelomar para que continúan con su desarrollo mental y estado social. Además, la Justificación práctica es brindar conocimientos a los docentes de la IE N° 116 Abraham Valdelomar para que tengan una visión no solo de la deserción educativa de la IE N°116; sino también de la realidad (educación social) presente en la educación institucional y la sociedad peruana, lo cual contribuirá a la formación de líderes talentosos y competitivos que la sociedad necesita.

Además, se propone el siguiente objetivo general, determinar qué el aprendizaje automático predice el rendimiento de aprendizaje del IE N°116 Abraham Valdelomar. como objetivos específicos; Primero, determinar en qué medida el aprendizaje automático predice con precisión los resultados de aprendizaje de la IE N° 116 Abraham Valdelomar-San Juan de Lurigancho; como Segundo: determinar en qué medida el aprendizaje automático predice con sensibilidad los resultados de aprendizaje de la IE N° 116 Abraham Valdelomar-San Juan de Lurigancho y finalmente determinar en qué medida el aprendizaje automático predice con especificidad los resultados de aprendizaje de la IE N° 116 Abraham Valdelomar-San Juan de Lurigancho.

Estos objetivos nos permiten establecer la siguiente hipótesis general: El aprendizaje automático predice el rendimiento académico de la IE N°116 Abraham Valdelomar-San Juan de Lurigancho. Como hipótesis específicas; en primer lugar: El aprendizaje automático predice con precisión el rendimiento académico de la IE N°116 Abraham

Valdelomar-San Juan de Lurigancho, en segundo lugar: El aprendizaje automático predice con sensibilidad el rendimiento académico de la IE N°116 Abraham Valdelomar-San Juan de Lurigancho, y finalmente, el aprendizaje automático predice con especificidad el rendimiento académico de la IE N°116 Abraham Valdelomar- San Juan de Lurigancho.

## **II. MARCO TEÓRICO**

Para el estudio y el desarrollo de esta tesis se consultaron fuentes internacionales y nacionales, las mismas que se detallan a continuación.

Como dice Candía (2019), en su disertación preparatoria en Perú “Predicting UNSAAC Students' Academic Performance from Their Admissions Data Using Machine Learning Algorithms”, dijo que ha desarrollado modelos predictivos utilizando el método CRISP-DM y herramientas WEKA, tomando en cuenta datos demográficos y factores educativos. Datos de análisis de estudiantes del 2014-I al 2018-I por la Universidad Nacional San Antonio Abad del Cusco, totalizando 12,698 estudiantes matriculados manifiesta que la investigación es cuantitativa, correlacional y no experimental. De manera similar, el algoritmo utilizado es el siguiente: la precisión del árbol de decisión J48 es del 67,3 %, el algoritmo del bosque aleatorio es del 69,4 %, el algoritmo del vecino más cercano es del 63,8 %, la función logística es del 68 % y finalmente la cognición multicapa. Es 68%. Estudia conceptos bien definidos y estructurados para dar cuenta de factores relacionados con los resultados del aprendizaje (Candia, 2019).

Según Menacho Chiok (2017 p. 2), “Aplicación de las técnicas de minería de datos para predecir el rendimiento académico fueron “desarrolladas en Perú”, expresó su preocupación por los resultados de aprendizaje de los estudiantes y concluyó que el problema más importante para la educación superior es resolver el problema, y las técnicas de minería de datos (TMD) aplicadas a los datos producidos en las instituciones educativas probadas se convertirán en una herramienta eficaz para predecir el efecto de la enseñanza de los estudiantes. En general, los resultados mostraron que la red Naive Bayes (red bayesiana que muestra gráficamente la dependencia e independencia de tres variables aleatorias, especialmente la independencia condicional) tuvo la tasa de clasificación correcta más alta (71,0). (Menacho Chiok, 2017).

Según Chahuán (2019) publicó un artículo sobre la predicción del rendimiento de los estudiantes mediante el aprendizaje automático en la Segunda Conferencia Internacional sobre Avances en Ciencia y Tecnología (ICAST) (Instituto de Tecnología de la Información KJ Somaiya, Mumbai, India). Según el objetivo principal es crear un motor de aprendizaje que pueda predecir el GPA (Promedio de

calificaciones) de un estudiante (2015-2019) basado en datos de cursos anteriores de informática. Las técnicas utilizadas son técnicas de regresión como ANN, árboles de decisión, SVM (Support Vector Machines), bosques aleatorios y regresión lineal. Se concluyó que el método de regresión lineal múltiple es óptimo con un error cuadrático de 0.00, un error cuadrático medio de 0.2, un error absoluto de 0.19 y un R.- al cuadrado de 0.90. La importancia de este estudio es la posibilidad de comparar modelos con mayor precisión, en este caso métodos de adquisición de regresión múltiple. (Chauhan, 2019).

El estudio es tan importante que concluye que el GPA (Grade Point Average) es un término utilizado para cuantificar las calificaciones obtenidas por los estudiantes. Esta técnica se puede utilizar para mejorar el rendimiento académico a través de la técnica de regresión descrita anteriormente.

Según Burman (2019) en (Conferencia Internacional de Amity sobre Inteligencia Artificial 2019 (AICAI, Dubái, Emiratos Árabes Unidos) utilizando un vector de soporte para predecir el rendimiento de aprendizaje de los estudiantes; según él, el objetivo es ayudar a los estudiantes a mejorar su aprendizaje a través de la inteligencia artificial. Creó una predicción. modelo basado en una función de vector de soporte (SVM) que clasifica a los estudiantes en tres categorías: alto, medio y bajo. El método de desarrollo propuesto por el autor incluye seis etapas: entrada de datos con un clasificador de máquina de vector de soporte (SVM), una etapa de entrenamiento con núcleos lineales y radiales, modelo de prueba de paso y estudio comparativo. Se recolectó información a través de cuestionarios sobre criterios psicológicos, motivaciones, estrategias psicosociales, estrategias de enseñanza per, según estilos de enseñanza y nivel socioeconómico y un total de 1000 conjuntos de datos.

En resumen, se recolectaron datos para análisis, el 70% de los cuales fue para capacitación y el 30% para validación o prueba. Para dar fidelidad al modelo a la sensibilidad, especificidad y precisión de la validación del modelo, se obtuvieron mejores resultados con una función de base de radio del 90 % en comparación con un kernel lineal del 64 %. A partir de este trabajo, el algoritmo SVM para Support Vector Machines y su modelo predictivo son instructivos. (Burman & Som, 2019).

Para obtener datos, realizamos una encuesta a los estudiantes matriculados desde 2022 con 36 preguntas, teniendo en cuenta otros factores confirmados en la encuesta, como el auto concepto, la motivación, los estilos de aprendizaje, los factores socioculturales, la educación de los padres, la inteligencia emocional. A partir de este estudio, se concluye que se tendrán en cuenta el procedimiento de compilación de datos y los factores de rendimiento.

Para respaldar completamente el estudio, utilizamos referenciales teóricos que explican la enseñanza automática, algoritmos supervisados como Máquina de Vectores de Soporte - SVM, Árboles de Decisión y Red Bayesiana; utilizando Metodología KDD, los resultados de evaluar la precisión de los modelos predictivos y, en última instancia, los resultados de la ciencia.

Según Arthur Samuel (1959) definió por primera vez el término aprendizaje automático como la rama de la ciencia que permite a las computadoras aprender sin programación explícita” (Bhavsar et al., 2017). Él autor Tom Mitchell define el Machine Learning en uno de sus libros “MACHINE LEARNING” como el estudio de algoritmos computacionales que mejoren automáticamente su rendimiento en función de la experiencia. (Álvaro, 2018). Se puede argumentar que el aprendizaje automático es la parte de la inteligencia artificial que permite que las computadoras aprendan de algoritmos que mejoran con la experiencia.

Como dice Segura-Morales & Loza-Aguirre (2017) en su investigación realizada en Las Vegas, NV, EE. UU. “(Predecir el rendimiento académico basado en factores socioeconómicos usando árboles de decisión). Su objetivo es determinar cómo los factores socioeconómicos afectan la educación tomando datos socioeconómicos y científicos, seguidos de la aplicación de técnicas de algoritmos de clasificación y métodos de estudio automático. Esta investigación nos explica que el factor económico afecta a los alumnos en su rendimiento académico debido a que no cuentan con los recursos pertinentes debidos como materiales de escritorio y tecnológicos por eso se ve afectado su rendimiento académico (Segura-Morales & Loza-Aguirre, 2017).



Como dice Díaz (2021) en Arbitraje Científico (Venezuela) Telos. Su propósito es evaluar el progreso de los estudiantes en la Maestría en Educación. En este sentido, utilizando el algoritmo WEKA J 8, se construye el modelo utilizando tecnología de árboles de decisión, minería de datos y herramientas de inteligencia artificial, teniendo en cuenta factores como educación, familia, nivel socioeconómico, hábitos y costumbres. Usando el coeficiente kappa de Cohen, la precisión del modelo con 237 estudiantes de una universidad pública en Perú fue del 66%. En conjunto, los resultados muestran un método para crear un sistema de evaluación de estudiantes basado en un tipo de resultado de aprendizaje. Esta clasificación ayuda a identificar a los estudiantes que probablemente tengan problemas de rendimiento en una etapa temprana. Por lo tanto, las medidas simultáneas y mitigadoras pueden implementarse de inmediato. (Díaz, 2021).

Según Vega (2019) Los algoritmos de aprendizaje automático son utilizados en su tesis en Perú, "Modelo predictivo del producto docente de los estudiantes del plan de estudios básico de la Universidad Ricardo Palma". Su objetivo es utilizar técnicas de aprendizaje automático para predecir el número de estudiantes exitosos y no exitosos. Todo el estudio está respaldado por referencias teóricas, incluido el aprendizaje automático, y se describe utilizando algoritmos supervisados como Support Vector Artifice (SVM), un algoritmo de seguimiento de entrenamiento para muchos problemas de clasificación y regresión, árboles de decisión, K-Neighbors y redes neuronales; El método KDD es un proceso asistido por un profesional de la industria, también conocido como agente inteligente, para calcular las métricas de demanda y, en última instancia, los resultados de aprendizaje del modelo predictivo.(Vega, 2019).

Arthur Samuel (1952) fue el primero en describir el aprendizaje automático como "el campo donde las computadoras pueden aprender sin ser programadas directamente". El algoritmo mejora automáticamente su rendimiento empíricamente. Se refiere a un programa de computadora que aprende tareas a partir de la experiencia y utiliza métricas de rendimiento.(Alvaro, 2018a).

El aprendizaje automático se utiliza para medir el rendimiento académico de los alumnos de la IE N°116” ABRAHAM VALDELOMAR”, al usar el Maching Learning como herramienta para poder estimar cuales son los problemas que afectan a los estudiantes del 4to y 5to se secundaria y poder brindar una solución a la institución, y capacitar a los docentes con nuevas técnicas de estudios a través del uso del machine Learning.

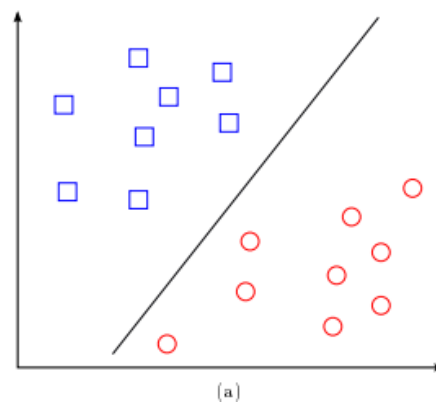
Según Hinostraza citado en el artículo de Cárdenas “El aprendizaje automático a través de los siglos y su aporte al desarrollo humano” 2018, página 8), menciona que existen tres tipos de aprendizaje automático, el primero tipo: a través de algoritmos guiados; que usa etiquetas para categorizar datos para identificar patrones que luego se pueden usar en otros grupos de entrada. El autor menciona que el segundo tipo se llama: algoritmos de aprendizaje no supervisado, donde no se utilizan etiquetas y la información se obtiene buscando patrones en los datos de entrada. Se puede concluir que el factor humano no interfiere en este proceso. Finalmente, el último tipo: llamado Empoderando el aprendizaje esto es cuando los datos no están etiquetados, pero después de unos pasos y un tiempo determinado, el sistema recibe una respuesta con actualizaciones; después de esta declaración. Lo que el sistema aprende del análisis de datos a través de prueba y error (Cárdenas, 2018).

Otro punto a considerar en este estudio es lo que Datademia (Data Academy) afirma que el aprendizaje profundo (deep learning) es una forma de aprendizaje automático donde una máquina intenta simular el cerebro humano utilizando una red neuronal artificial con más de tres capas que permiten predecir con alta precisión.(Datademia, 2022).

En este trabajo de investigación se utilizó un algoritmo de tipeo supervisado. Esto se debe a que las etiquetas se usaron para definir el desempeño académico en categorías (excelente, excelente, bueno, regular, reprobado) según la investigación de otras partes interesadas. También pudimos observar que las SVM (Support Vector Machines) lograron resultados cercanos al 100%, por lo que utilizamos principalmente este algoritmo para comparar con algoritmos de árboles de decisión, redes bayesianas.

De acuerdo a Carmona (2016). Vector Machine: SVM es un algoritmo originalmente destinado a problemas de clasificación. Sin embargo, más recientemente se ha utilizado para resolver problemas de regresión, agrupamiento y clasificación múltiple. También dice: "Para las tareas de clasificación, SVM es un tipo de clasificador lineal porque produce separadores lineales, también conocidos como hiperplanos". (Carmona, 2016).

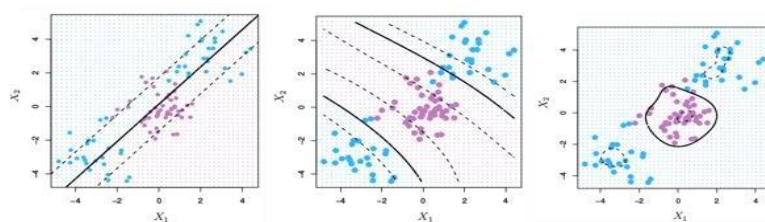
Figura 1 Ejemplo de clasificación con SVM



Fuente: Carmona (2016)

Siguiendo con el autor Carmona (2016) También se menciona que se pueden obtener resultados óptimos cuando las clases son linealmente separables, de lo contrario es necesario utilizar una función generadora multidimensional con el fin de encontrar un hiperplano. El autor afirma: Entre los dos vectores discurre en un nuevo espacio dimensional que es diferente del espacio original. Los núcleos más utilizados son lineales, polinómicos y gaussianos (RBF).

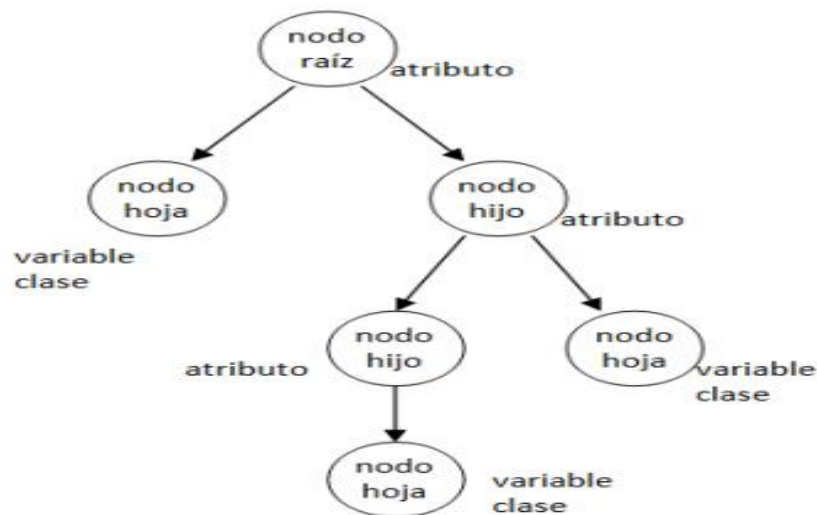
Figura 2 Kernel Lineal, Kernel polinómico y Gaussian kernel respectivamente



Fuente: Carmona (2016)

Otro modelo utilizado para nuestra investigación es el algoritmo de árbol de decisión según Charris (2018), "Su enfoque principal es el aprendizaje inductivo basado en estructuras y observaciones lógicas". Se puede representar como un diagrama de bloques porque su estructura es similar. Se utilizan para representar y clasificar una serie de condiciones que se dan secuencialmente para resolver un problema. El árbol comienza desde el nodo raíz correspondiente al atributo raíz; se ramifica en nodos secundarios que representan reglas de decisión y finalmente termina con nodos hoja que representan una decisión o resultado. (Charris.L, 2018).

Figura 3 Estructura de árbol de decisión

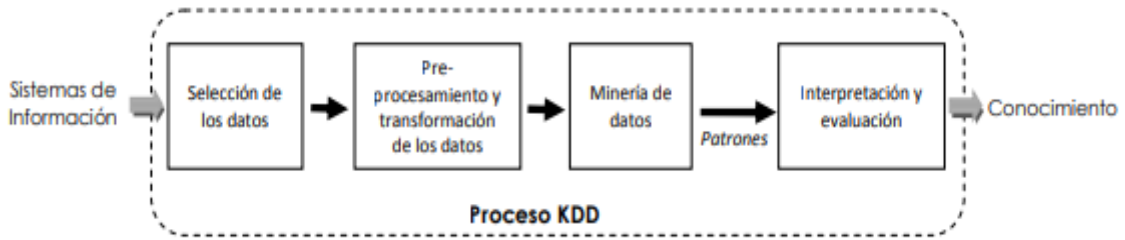


Fuente: Charris (2018)

Barrientos (2009) También mencionó que los algoritmos más utilizados en las decisiones relacionadas con el valor simple, de precisión y bajo informática en su rendimiento son los siguientes: ID3, "El algoritmo de la diferencia. Entre los datos para el análisis es el proceso de las piezas y la victoria, Maximización de la información recibida utilizada como el mejor expediente de selección de atributos para dividir los datos en capas homogéneas ", además de tener el algoritmo J48, su característica es crear un método repetitivo, también puede usar los números y eliminar las propiedades de la creación de la creación de la creación de Nuevas reglas, esto no está relacionado con otra función actual. (Barrientos, 2009).



Figura 5 fases resumidas de la metodología KDD



Fuente: Moine (2013)

**Primera fase,** Selección de los datos sobre los que se trabajará para encontrar un nuevo conocimiento.

**Segunda fase,** Preprocesamiento de datos en el que se procesan datos inválidos y faltantes para seleccionar datos de buena calidad.

**Tercera fase,** Transforme los datos y reduzca el tamaño para reducir el número de variables, y las variables no afectan.

**Cuarta fase,** Minería de datos, se genera en el tercer paso donde se obtienen los modelos de interés según la tarea de minería de datos que realicemos (descriptiva o predictiva).

**Quinta fase,** Interpretación y evaluación nuevo conocimiento descriptivo o predictivo utilizando precisión, sensibilidad y especificidad (Moine, 2013).

Tabla 1: Matriz de confusión

		Predicción	
		Positivos	Negativos
Verdaderos	Positive	(VP) Verdaderos Positivos	(FN) Falsos Negativos
	Negative	(FP) Falsos Positivos	(VN) Verdaderos Negativos

Fuente: Arias (2020)

Al realizar una predicción en esta matriz de confusión se pueden observar cuatro resultados diferentes:

**Resultado uno**, verdaderos positivos (VP) son los casos que son predichos de manera acertada por el modelo, es decir que son positivos y efectivamente se clasificaron como positivos.

**Resultado dos**, falsos positivos (FP) son los casos que el modelo clasificó como positivos, pero realmente son negativos

**Resultado tres**, falsos negativos (FN) es el número de muestras que el modelo le otorgó la etiqueta de negativo y realmente son positivos.

**Resultado cuatro**, verdaderos negativos (VN) son los casos que el modelo predice como negativo y efectivamente lo son.

**La Sensibilidad** ("Recall" o "Sensitivity"), "Representa el número real de predicciones positivas del número total de predicciones reales que calificaron como positivas. En otras palabras, esta métrica muestra qué tan bien el modelo predice casos positivos"...(Mishra, 2018).

Figura 6 Fórmula para calcular la sensibilidad

$$SENSIBILIDAD = \frac{VP}{VP + FN} \times 100$$

Fuente: Mishra (2019)

**La especificidad**, "Esto corresponde a la proporción del total de casos negativos que el modelo considera positivos. En otras palabras, el error del modelo predice casos negativos".(Mishra, 2018).

Figura 7 Fórmula para calcular la Especificidad

$$ESPECIFICIDAD = \frac{VN}{VN + FP} \times 100$$

Fuente: Mishra (2019)

**La Precisión (Recall)**, "Esta es una métrica que proporciona información sobre el rendimiento del modelo en términos de la cantidad de muestras clasificadas como falsos negativos, es decir, la cantidad de predicciones fallidas".(Mishra, 2018).

Figura 8 Fórmula para calcular la precisión

$$PRECISIÓN = \frac{VP + TN}{VP + VN + FP + FN} \times 100$$

Fuente: Mishra (2019)

Otro aspecto muy importante de esta investigación son los factores que influyen en los resultados del aprendizaje y sus predictores. Para muchos escritores, el objetivo principal de una institución educativa es producir buenos estudiantes, y muchos factores incluyen su cercanía y aprecio por sus estudiantes. El comportamiento basado principalmente en los esfuerzos de calificación del maestro puede llevar al éxito o al fracaso del estudiante.(Burgueño, 2019).

Bajo dicha premisa autores como Hernández (2013, p.20), identificar los resultados del aprendizaje como resultado de factores relevantes y relacionados con el alumno, como el estilo de vida y los comportamientos de salud adoptados, para que los jóvenes con estilos de vida y comportamientos más saludables puedan demostrar un mejor rendimiento académico. (Hernández, 2015). Por otro lado, Navarro (2014), Esto nos dice que el rendimiento académico está relacionado con la evaluación periódica de un estudiante y que se deben considerar los factores externos que rodean a ese estudiante, como las aulas, los departamentos, los centros de aprendizaje, etc., las pandillas, las drogas y el acoso escolar. (Navarro, 2003).

De lo que han escrito varios actores en sus documentos anteriores, se puede inferir que los resultados del aprendizaje están impulsados por muchos factores, lo que permite múltiples direcciones de investigación para determinar cómo afectan los resultados del aprendizaje.



### **III. METODOLOGÍA**

### 3.1 Tipo y diseño de investigación

#### 3.1.1. Tipo de Investigación:

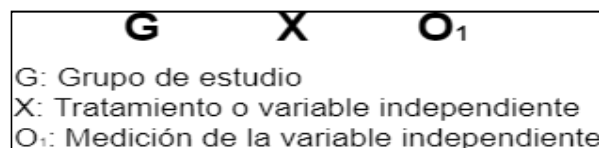
Esta investigación es de tipo aplicada, para el autor (Murillo, 2009, p.12), “Se caracteriza por el hecho de que está destinado a la realización o familiarizarse, mientras que otros conocimientos son adquiridos después de la implementación y sistematización de la práctica investigativa”. Según Schubert (2017, p. 5), “Deriva de prácticas sociales que producen resultados que pueden ser utilizados, pero no necesariamente utilizados en la producción, muchas veces por razones de costo”.

La investigación es aplicable porque se emplea un aprendizaje automático para resolver la predicción de los resultados del aprendizaje, lo que permite identificar a los estudiantes con un alto grado de precisión.

#### 3.1.2. Diseño de Investigación:

Así mismo, el diseño del estudio es de experimental de tipo pre- experimental, en el cual se acopia información de un grupo de estudiantes, representados por sus datos, a través de un cuestionario, luego se aplica aprendizaje automático y luego se obtiene una observación para realizar un sondeo. Según Hernández (2014, p. 141) en su libro Métodos de investigación, “Esto sugiere que una prueba inicial Implica estimulación o tratamiento grupal, y luego medir una o más variables para ver cuántas personas hay en ese grupo. ” Bernal (Bernal, p. 153, 2010) “Descripción en su obra “Métodos de investigación” de un proyecto preexperimental con un grupo. Este es un plan no controlado en el que solo se toma otra medida.

Figura 9 Diseño pre-experimental con un solo grupo



**G:** Datos de los estudiantes de ingeniería de sistemas

**X:** Machine Learning.

**O:** Métricas de precisión.

## **3.2 Variable y operacionalización**

### **3.2.1. Variable Independiente Maching Learning**

Una variable de aprendizaje automático es un tipo de variable cuantitativa. Según (McKinsey & Company), se basa en algoritmos que aprenden de los datos sin usar programación basada en reglas.

#### **Definición Conceptual de Maching Learning.**

El aprendizaje automático es un área de inteligencia artificial que utiliza algoritmos para permitir que las computadoras reconozcan y predigan patrones en grandes cantidades de datos.

#### **Definición Operacional de Machine Learning**

El aprendizaje automático se trata básicamente de identificar patrones o tendencias "ocultos" en los datos al automatizarlos con varios algoritmos. Por tanto, es muy importante no solo elegir el algoritmo más adecuado (y su posterior parametrización para cada tarea), sino también disponer de una gran cantidad de datos de la calidad adecuada.

### **3.2.2. Definición Conceptual de la Variable Dependiente: Rendimiento Académico.**

En este estudio se utilizan indicadores para medir la variable dependiente: a) Precisión; b) Sensibilidad y c) Especificidad esto se mide como un porcentaje utilizando la guía de observación según el Anexo N°02

#### **Indicadores**

Esta investigación muestra tres indicadores para una variable independiente, como primer indicador tenemos: Precisión, el segundo indicador es: Sensibilidad y el tercer indicador es: Especificidad.

#### **Escala de Medición**

Para esta variable se toma en cuenta la escala de razón la cual se caracteriza por medir datos cuantitativos y no existen valores negativos.

### 3.3 Población, muestra y muestreo probabilístico

#### Población

Es un conjunto de elementos o individuos que corresponden al inmueble objeto de estudio.(Fuentes, 2004, p. 18). Para Ojeda, las poblaciones son elementos disponibles o unidades de análisis que pertenecen a un campo de estudio específico.(Ojeda, 2020). En la IE N°116 “Abraham Valdelomar” se estará tomó la población de los alumnos estudiantes de secundaria con total de 101 alumnos.

A continuación, se describe la población de estudio de la presente investigación:

Tabla 2 Población de estudio

Población	Cantidad		Indicador
	Pre-test	Post-test	
Alumnos	101	101	Predecir con Precisión el Rendimiento Académico de la IE N°116.
Alumnos	101	101	Predecir la Sensibilidad del Rendimiento Académico de la IE N°116.
Alumnos	101	101	Predecir con Especificidad el Rendimiento Académico de la IE N°116.

Fuente: Elaboración propia

#### Muestra

Según (Hernández-Sampieri & Mendoza, 2018) definen que la muestra es un subconjunto de la población o universo de su interés, sobre qué datos relevantes se deben recopilar y deben ser representativos de ese conjunto. Para nuestra investigación la población objetivo estuvo conformada por una población de 101 registros de alumnos, los mismos que fueron obtenidos a través de un cuestionario.

### **Muestreo No probabilístico por conveniencia**

Un método de muestreo no probabilístico en el que las muestras se seleccionan de una población simplemente porque son convenientes para el investigador. (Muestreo No Probabilístico: Definición, Tipos y Ejemplos, 2020).

#### **3.4. Técnicas e instrumentos de recolección de datos**

Según García, “Las preguntas son la forma tradicional de recolectar y recolectar datos en las ciencias sociales. Su versatilidad les permite ser utilizadas tanto como herramientas de investigación como de evaluación de personas, procesos y cursos. (García, 2003, p.2), el autor García Córdova define un cuestionario como un sistema de preguntas coherentemente estructuradas de importancia lógica y psicológica, expresión en lenguaje sencillo y claro. Esto permite recolectar datos de fuentes primarias (García, 2002, p.7). En este estudio se realizó con un cuestionario (35 preguntas) el cual fue aplicado a alumnos del 4to. Y 5to. de secundaria, el mismo que fue registrado en la herramienta Google Formulario, las preguntas se formularon en relación con artículos relacionados con el tema de investigación.

Tabla 3 Ficha técnica del instrumento

<b>Nombre del Instrumento</b>	<b>Cuestionario</b>
Autor	Acosta Portocarrero, Antony Christian y Ruiz Vargas Paul
Año	2022
Descripción	Cuestionario de preguntas
Tipo de Instrumento	
Objetivo	Machine Learning para predecir el rendimiento académico de la IE N°116 Abraham Valdelomar san Juan de Lurigancho 2022
Indicadores	a) Predecir con Precisión el Rendimiento Académico de la IE N°116 b) Predecir la Sensibilidad del Rendimiento Académico de la IE N°116 c) Predecir con Especificidad el Rendimiento Académico de la IE N°116
Número de Alumnos	101
Aplicación	Directa

Fuente: Elaboración propia

## Validez

Mediante la revisión por pares, las herramientas de recopilación de datos se validaron con 3 expertos que señalaron (Hernández-Sampieri & Mendoza, 2018) que la validación de la herramienta era esencial, ya que con ello asegura que está usando correctamente lo necesario para medir la variable de estudio

En la siguiente tabla se muestra la lista de expertos que participaron en la validación.

Tabla 4 Expertos que validaron herramientas de recolección de datos

DNI	Grado Académico Apellidos y Nombres	Institución donde labora	Calificación
44147992	Magister Alan Leoncio, Fierro Barriales	Universidad Cesar Vallejo	Aplicable
80593277	Doctor Pedro Saturnino Peña Huapaya	Catedrático	Aplicable
08820518	Magister Orlando Gustavo Diaz López	Catedrático	Aplicable

Fuente: Elaboración propia

### 3.5. Procedimiento

El propósito de este estudio es usar el aprendizaje automático para predecir los resultados de rendimiento de los estudiantes IE No. 116 "ABRAAM VALDELOMAR", que es la variable dependiente "predicción de resultados de aprendizaje" que se basó en estudios similares. A nivel nacional e internacional para descubrir una solución por el momento.

Por lo tanto, las variables dependientes e independientes se exploran en detalle, reuniendo información de tesis y artículos científicos con el propósito de conseguir la base teórica y preliminar a partir de ellas y así obtener las medidas e indicadores del número estable. Los datos fueron recolectados de los estudiantes para este proyecto de investigación de la IE N°116 "ABRAHAM VALDELOMAR" mediante un cuestionario el cual fue generado por la aplicación de formularios de Google para

visualizar las preguntas calculadas en este estudio, posteriormente se desarrollaron los métodos KDD (Knowledge Discovery in Data bases), estos son esencialmente procesos automatizados que combinan detección y análisis. Este proceso consiste en extraer patrones de los datos en forma de reglas o funciones que puedas analizar.

### **3.6. Método de análisis de datos**

La información para el estudio se recopiló mediante un cuestionario diseñado en la herramienta en línea Formularios virtuales de Google, luego se revisaron y eliminaron los datos inconsistentes y luego se transformaron las variables de rendimiento con la herramienta SPSS Statistic View 26. De cuantitativo a cualitativo, SPSS Modeler ver. finalmente se utilizó. 18 para visualizar la información puntual de la variable dependiente (Presión, Sensibilidad y Especificidad). Se utilizaron estadísticas predictivas para obtener esta información.

### **3.7. Aspectos éticos**

En la realización de esta investigación se tuvo cuidado de garantizar una conducta ética, tal como lo establece la Resolución del Consejo Universitario de César Vallejo 03 0-2021-UCV (Resolución No. Integridad en la Ciencia). La investigación realizada y aprobada por investigadores universitarios es consistente con los estándares globales de integridad, responsabilidad y rigor científico. De esta forma, se reconoce la responsabilidad y obligación de recolectar y utilizar datos científicos, justos y honestos. Y, como último recurso, se utilizó el programa Turnitin para comprobar la originalidad de la investigación y el cumplimiento de las normas anti plagio.

#### **IV. RESULTADOS**



Los resultados del estudio basados en medidas de precisión, sensibilidad y especificidad se detallan aquí. Compare cuatro algoritmos de aprendizaje para determinar el mejor algoritmo utilizando el índice Cohen Kappa para la prueba de hipótesis. Esto indica tanto la equivalencia como el nivel de importancia. De igual manera, el MINEDU recomienda utilizar las siguientes escalas para medir los resultados de aprendizaje. Del mismo modo, la Universidad de Granada clasifica sus calificaciones en Perú y crea una tabla de conversión de calificaciones específica del país de la siguiente manera:

**Tabla 5 Tabla de calificaciones**

Excelente	19-20	5
Distinguido	17-18.9	4
Bueno	14-16.9	3
Regular	11-13.9	2
Reprobado	0-10.9	1

Fuente: Universidad de granada

A continuación, se muestra la solución de la matriz de confusión obtenida con SPSS Modeler Ver18. para los siguientes algoritmos.

### Algoritmo: Árbol de decisión

**Tabla 6 Modelo de confusión - árbol de decisión**

		Predicción					
		1	2	3	4	5	
1	Rendimiento	Reprobado	0	7	0	0	0
2	Regular	0	22	5	0	0	
3	Bueno	0	4	33	7	0	
4	Distinguido	0	0	4	17	0	
5	Excelente	0	0	0	0	2	

Fuente: Elaboración propia

**Tabla 7 : Modelo de observación - Árbol de decisiones**

clase	Medidas			
	TP	TN	FP	FN
1	0	94	0	7
2	22	63	11	5
3	33	48	9	11
4	17	73	7	4
5	2	75	0	0

Fuente: Elaboración propia

De un total de 101 registros de datos de entrenamiento para el algoritmo Árbol de decisión se encontró 0 estudiantes con calificaciones reprobatorias (TP) y 7 estudiantes con calificaciones incorrectas (FN), 22 estudiantes con calificaciones regulares (TP), 5 estudiantes incorrectas (FN), 33 estudiantes con calificaciones buenas (TP), 11 estudiantes tuvieron calificaciones incorrectas (FN), 17 estudiantes fueron obtuvieron calificaciones distinguidas (TP) y 4 estudiantes obtuvieron calificaciones incorrectas (FN) y finalmente 2 estudiantes obtuvieron calificaciones de excelente (T) y 0 estudiantes resultaron con calificaciones incorrectas (FN).

### Redes Máquina de Vectores – SVM

Tabla 8 :Modelo de confusión – Máquina de Vectores SVM

		Predicción				
		1	2	3	4	5
1	Rendimiento Reprobado	7	0	0	0	0
2	Regular	0	27	0	0	0
3	Bueno	0	0	44	0	0
4	Distinguido	0	0	0	21	0
5	Excelente	0	0	0	0	2

Fuente: Elaboración propia

Tabla 9: Modelo de Observación - SVM. Vector Machine

Clase	Medidas			
	TP	TN	FP	FN
1	7	94	0	0
2	27	74	0	0
3	44	57	0	0
4	21	80	0	0
5	2	99	0	0

Fuente: Elaboración propia

De un total de 101 conjuntos de datos sobre datos de entrenamiento para el algoritmo SVM, se pudo identificar a 7 estudiantes con calificaciones reprobatorias (TP) y 27 estudiantes con calificaciones regulares (TP), 44 estudiantes con calificaciones buenas (TP), 21 estudiantes fueron identificados con calificaciones de distinguido (TP) y finalmente se pudo identificar 2 estudiantes con calificaciones de excelente (TP).

### Red Bayesiana

Tabla 10: Modelo de Confusión- Red Bayesiana

Rendimiento		Predicción				
		1	2	3	4	5
1	Reprobado	7	0	0	0	0
2	Regular	0	26	1	0	0
3	Bueno	0	0	44	0	0
4	Distinguido	0	0	0	21	0
5	Excelente	0	0	0	0	2

Fuente: Elaboración propia

Tabla 11: Modelo de observación - red bayesiana

clase	Medidas			
	TP	TN	FP	FN
1	7	94	0	0
2	26	74	0	1
3	44	56	1	0
4	21	80	0	0
5	2	99	0	0

Fuente: Elaboración propia

Del total de 101 registros de datos de entrenamiento para el algoritmo Árbol de decisión se encontró 7 estudiantes con calificaciones reprobatorias (TP) y 0 estudiantes con calificaciones incorrectas (FN), 26 estudiantes con calificaciones regulares (TP), 1 estudiantes incorrectas (FN), 44 estudiantes con calificaciones buenas (TP), 0 estudiantes tuvieron calificaciones incorrectas (FN), 21 estudiantes fueron obtuvieron calificaciones distinguidas (TP) y 0 estudiantes obtuvieron calificaciones incorrectas (FN) y finalmente 2 estudiantes obtuvieron calificaciones de excelente (T) y 0 estudiantes resultaron con calificaciones incorrectas (FN).

Entonces, se presentan los datos obtenidos para las hipótesis formuladas en este estudio. Para obtener estos resultados se utilizó la herramienta estadística SPSS para mostrar el grado de concordancia entre dos observaciones y se utilizó el exponente Kappa de Cohen. Según Manterola (2018, p. 261), la índice kappa de Cohen es una forma de correlación y, como todos los demás índices, puede variar de -1 a 1, donde 1 representa el acuerdo perfecto entre dos observadores. de Salas (2019, pág. 2) lo define como “Es el valor que tiende a fortalecer la concordancia entre diferentes evaluadores”, como se muestra en la Figura 11.

Figura 10 : Clase Cappa de Cohen

Valores	Interpretación
< 0,01	No acuerdo
0,01 - 0,20	Ninguna a escaso
0,21 - 0,40	Regular o razonable
0,41 - 0,60	Moderado
0,61 - 0,80	Substancial
0,81 - 1,00	Casi perfecto

Fuente: Manterola (2018)

HE1: El aprendizaje automático predice con precisión el rendimiento de los alumnos de la IE N° 116 “Abraham Valdelomar”.

### Árbol de decisión

Tabla 12: Tabla cruzada – Cálculo preciso con algoritmo de árbol de decisión

		Predicción				Total	
		1	2	3	4		
Rendimiento	1	Recuento	7	0	0	0	7
		% dentro	de100,0%	0,0%	0,0%	0,0%	100,0%
	Rendimiento						
	2	Recuento	22	5	0	0	27
		% dentro	de81,5%	18,5%	0,0%	0,0%	100,0%
	Rendimiento						
	3	Recuento	4	33	7	0	44
		% dentro	de9,1%	75,0%	15,9%	0,0%	100,0%
	Rendimiento						
	4	Recuento	0	4	17	0	21
		% dentro	de0,0%	19,0%	81,0%	0,0%	100,0%
	Rendimiento						
	5	Recuento	0	0	0	2	2
		% dentro	de0,0%	0,0%	0,0%	100,0%	100,0%
	Rendimiento						

Total	Recuento	33	42	24	2	101
	% dentro de Rendimiento	32,7%	41,6%	23,8%	2,0%	100,0%

Fuente: SPSS Statistics Ver.26

$$\text{PRECISIÓN} = \frac{VP + VN}{VP + VN + FP + FN} * 100 = \frac{(64+380)}{(64+380+40+40)} = \frac{444}{524} * 100 = 84.73$$

Interpretación: El aprendizaje automático permite utilizar un algoritmo de árbol de decisión para predecir los resultados de aprendizaje de los alumnos IE N°116 "Abraham Valdelomar" con una precisión del 84,73%..

Máquina de vectores de soporte (SVM)

Tabla 13: Tabla cruzada – Cálculo preciso con algoritmo SVM

		Predicción					Total	
		1	2	3	4	5		
Rendimiento	1	Recuento	7	0	0	0	0	7
		% dentro de Rendimiento	100,0%	0,0%	0,0%	0,0%	0,0%	100,0%
	2	Recuento	0	27	0	0	0	27
		% dentro de Rendimiento	0,0%	100,0%	0,0%	0,0%	0,0%	100,0%
	3	Recuento	0	0	44	0	0	44
		% dentro de Rendimiento	0,0%	0,0%	100,0%	0,0%	0,0%	100,0%
	4	Recuento	0	0	0	21	0	21
		% dentro de Rendimiento	0,0%	0,0%	0,0%	100,0%	0,0%	100,0%
	4	Recuento	0	0	0	0	2	2
		% dentro de Rendimiento	0,0%	0,0%	0,0%	0,0%	100,0%	100,0%
Total	Recuento	7	27	44	21	2	101	
	% dentro de Rendimiento	6,9%	26,7%	43,6%	20,8%	2,0%	100,0%	

Fuente: SPSS Statistics Ver. 26.

$$\text{PRECISIÓN} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{VN} + \text{FP} + \text{FN}} * 100 = \frac{(101+404)}{(101+404+0+0)} = \frac{(505/505)}{} * 100 = 100$$

Interpretación: El aprendizaje automático permite utilizar un algoritmo de Vectores de soporte-SVM para predecir los resultados de aprendizaje de los alumnos IE N°116 "Abraham Valdelomar" con una precisión del 100.00%.

**Red – Bayesiana**

Tabla 14: Tabla cruzada – Cálculo preciso con algoritmo de redes Bayesianas

		Predicción					Total	
		1	2	3	4	5		
Rendimiento	1	Recuento	7	0	0	0	0	7
		% dentro	de100,0%	0,0%	0,0%	0,0%	0,0%	100,0%
		Rendimiento						
	2	Recuento	0	27	0	0	0	27
		% dentro	de0,0%	100,0	0,0%	0,0%	0,0%	100,0%
		Rendimiento		%				
	3	Recuento	0	0	44	0	0	44
		% dentro	de0,0%	0,0%	100,0	0,0%	0,0%	100,0%
		Rendimiento			%			
	4	Recuento	0	0	0	21	0	21
		% dentro	de0,0%	0,0%	0,0%	100,0	0,0%	100,0%
		Rendimiento				%		
	5	Recuento	0	0	0	0	2	2
		% dentro	de0,0%	0,0%	0,0%	0,0%	100,0	100,0%
		Rendimiento					%	
Total	Recuento	7	27	44	21	2	101	
	% dentro	de6,9%	26,7%	43,6%	20,8%	2,0%	100,0%	
	Rendimiento							

Fuente: SPSS Statistics Ver.26

$$\text{PRECISIÓN} = \frac{VP + VN}{VP + VN + FP + FN} * 100 = \frac{(100+403)}{(100+403+1+1)} = (503/505) * 100 = 99.60$$

Interpretación: El aprendizaje automático permite utilizar un algoritmo de Redes Bayesianas para predecir los resultados de aprendizaje de los alumnos IE N°116 "Abraham Valdelomar" con una precisión del 99.60%.

Tabla 15: Cuadro comparativo de resultados según el indicador precisión

ALGORITMO	RESULTADO (%)
Árbol de decisión	84.73
SVM	100.00
Red - Bayesiana	99.60

Fuente: SPSS Statistics Ver.26

Interpretación: Se observa en la tabla 18 que el método que tiene mejor solución para el indicador de precisión es la "Máquina de Vectores (SVM)" con un 100%, a continuación, seguido del "algoritmo de la Red – Bayesiana" con un 99.60% y por último el algoritmo de precisión "Árbol de Decisión" con un resultado igual a 84.73%.

HE2: El aprendizaje automático predice con sensibilidad el rendimiento académico de los alumnos de la IE N°116- "Abraham Valdelomar".

**Árbol de decisión**

Tabla 16 Tabla cruzada – cálculo de sensibilidad con árbol de decisión

		Predicción				Total	
		1	2	3	4		
Rendimiento	1	Recuento	7	0	0	0	7
		% dentro de Rendimiento	de100,0%	0,0%	0,0%	0,0%	100,0%



2	Recuento	22	5	0	0	27
	% dentro Rendimiento	de81,5%	18,5%	0,0%	0,0%	100,0%
3	Recuento	4	33	7	0	44
	% dentro Rendimiento	de9,1%	75,0%	15,9%	0,0%	100,0%
4	Recuento	0	4	17	0	21
	% dentro Rendimiento	de0,0%	19,0%	81,0%	0,0%	100,0%
5	Recuento	0	0	0	2	2
	% dentro Rendimiento	de0,0%	0,0%	0,0%	100,0%	100,0%
Total	Recuento	33	42	24	2	101
	% dentro Rendimiento	de32,7%	41,6%	23,8%	2,0%	100,0%

Fuente: SPSS Statistics Ver.26

$$\text{SENSIBILIDAD} = \frac{\text{VP}}{\text{VP} + \text{FN}} * 100 = \frac{64}{64+40} = \frac{64}{104} * 100 = 61.54$$

Interpretación: El aprendizaje automático permite utilizar un algoritmo de Árbol de Decisión para predecir los resultados de aprendizaje de los alumnos IE N°116 "Abraham Valdelomar" con una sensibilidad del 61.54%.

### Máquina de vectores SVN

Tabla 17 Tabla cruzada – cálculo de sensibilidad con algoritmo Máquina de Vectores de Soporte – SVM

		Predicción					Total	
		1	2	3	4	5		
Rendimiento	1	Recuento	7	0	0	0	0	7
		% dentro	de100,0%	0,0%	0,0%	0,0%	0,0%	100,0%
		Rendimiento						
	2	Recuento	0	27	0	0	0	27
		% dentro	de0,0%	100,0%	0,0%	0,0%	0,0%	100,0%
		Rendimiento						
	3	Recuento	0	0	44	0	0	44
		% dentro	de0,0%	0,0%	100,0	0,0%	0,0%	100,0%
		Rendimiento			%			
	4	Recuento	0	0	0	21	0	21
		% dentro	de0,0%	0,0%	0,0%	100,0	0,0%	100,0%
		Rendimiento				%		
4	Recuento	0	0	0	0	2	2	
	% dentro	de0,0%	0,0%	0,0%	0,0%	100,0	100,0%	
	Rendimiento					%		
Total	Recuento	7	27	44	21	2	101	
	% dentro	de6,9%	26,7%	43,6%	20,8%	2,0%	100,0%	
	Rendimiento							

Fuente: SPSS Statistics Ver.26

$$\text{SENSIBILIDAD} = \frac{\text{VP}}{\text{VP} + \text{FN}} * 100 = (101) / (0) = (101/0) * 100 = 100.00$$

Interpretación: El aprendizaje automático permite utilizar un algoritmo de Máquina de vectores-SVN para predecir los resultados de aprendizaje de los alumnos IE N°116 "Abraham Valdelomar" con una sensibilidad del 100.00%.

## Red – Bayesiana.

Tabla 18 Tabla cruzada – cálculo de sensibilidad con algoritmo Red – Bayesiana

		Predicción					Total	
		1	2	3	4	5		
Rendimiento	1	Recuento	7	0	0	0	0	7
		% dentro	de100,0%	0,0%	0,0%	0,0%	0,0%	100,0%
		Rendimiento						
	2	Recuento	0	27	0	0	0	27
		% dentro	de0,0%	100,0	0,0%	0,0%	0,0%	100,0%
		Rendimiento		%				
	3	Recuento	0	0	44	0	0	44
		% dentro	de0,0%	0,0%	100,0	0,0%	0,0%	100,0%
		Rendimiento			%			
	4	Recuento	0	0	0	21	0	21
		% dentro	de0,0%	0,0%	0,0%	100,0	0,0%	100,0%
		Rendimiento				%		
	5	Recuento	0	0	0	0	2	2
		% dentro	de0,0%	0,0%	0,0%	0,0%	100,0	100,0%
		Rendimiento					%	
Total	Recuento	7	27	44	21	2	101	
	% dentro	de6,9%	26,7%	43,6%	20,8%	2,0%	100,0%	
	Rendimiento							

Fuente: SPSS Statistics Ver.26

$$\text{SENSIBILIDAD} = \frac{\text{VP}}{\text{VP} + \text{FN}} * 100 = \frac{(100)}{(100+1)} = (100/101) * 100 = 99.00$$

Interpretación: El aprendizaje automático permite utilizar un algoritmo de Redes Bayesianas para predecir los resultados de aprendizaje de los alumnos IE N°116 "Abraham Valdelomar" con una sensibilidad del 99.00%.

Tabla 19: Cuadro comparativo de resultados según el indicador precisión

ALGORITMO	RESULTADO (%)
Árbol de decisión	61.54
SVM	100.00
Red - Bayesiana	99.00

Fuente: SPSS Statistics Ver.26

Interpretación: En la Tabla 19 se concluye que el método que tiene mejor solución para el indicador de precisión es la "Máquina de Vectores (SVM)" con un 100%, a continuación, seguido del "algoritmo de la Red – Bayesiana" con un 99.00% y por último el algoritmo de precisión "Árbol de Decisión" con un resultado igual a 61.54%.

HE3: El aprendizaje automático predice con especificidad el rendimiento académico de los alumnos de la IE N°116 "Abraham Valdelomar".

### Árbol de decisión

Tabla 20 Tabla cruzada – cálculo de especificidad con árbol de decisión

		Predicción				Total	
		1	2	3	4		
Rendimiento	1	Recuento	7	0	0	0	7
		% dentro	de100,0%	0,0%	0,0%	0,0%	100,0%
		Rendimiento					
	2	Recuento	22	5	0	0	27
		% dentro	de81,5%	18,5%	0,0%	0,0%	100,0%
		Rendimiento					
	3	Recuento	4	33	7	0	44
		% dentro	de9,1%	75,0%	15,9%	0,0%	100,0%
		Rendimiento					
	4	Recuento	0	4	17	0	21
		Rendimiento					

	% dentro	de0,0%	19,0%	81,0%	0,0%	100,0%
	Rendimiento					
5	Recuento	0	0	0	2	2
	% dentro	de0,0%	0,0%	0,0%	100,0%	100,0%
	Rendimiento					
Total	Recuento	33	42	24	2	101
	% dentro	de32,7%	41,6%	23,8%	2,0%	100,0%
	Rendimiento					

Fuente: SPSS Statistics Ver.26

$$\text{ESPECIFICIDAD} = \frac{\text{VN}}{\text{VN} + \text{FP}} * 100 = \frac{380}{380+40} = \frac{380}{420} * 100 = 90.47$$

Interpretación: El aprendizaje automático permite utilizar un algoritmo de Árbol de decisión para predecir los resultados de aprendizaje de los alumnos IE N°116 "Abraham Valdelomar" con una especificidad del 90.47%.

### Máquina de vectores de soporte (SVM)

Tabla 21 Tabla cruzada – cálculo de especificidad con algoritmo SVM

		Predicción					Total	
		1	2	3	4	5		
Rendimiento	1	Recuento	7	0	0	0	0	7
		% dentro	de100,0%	0,0%	0,0%	0,0%	0,0%	100,0%
		Rendimiento						
	2	Recuento	0	27	0	0	0	27
		% dentro	de0,0%	100,0	0,0%	0,0%	0,0%	100,0%
		Rendimiento						
	3	Recuento	0	0	44	0	0	44
		% dentro	de0,0%	0,0%	100,0	0,0%	0,0%	100,0%
		Rendimiento						
	4	Recuento	0	0	0	21	0	21
		% dentro	de0,0%	0,0%	0,0%	100,0	0,0%	100,0%
		Rendimiento						

5	Recuento	0	0	0	0	2	2
	% dentro Rendimiento	de0,0%	0,0%	0,0%	0,0%	100,0%	100,0%
Total	Recuento	7	27	44	21	2	101
	% dentro Rendimiento	de6,9%	26,7%	43,6%	20,8%	2,0%	100,0%

Fuente: SPSS Statistics Ver.26

$$\text{ESPECIFICIDAD} = \frac{VN}{VN + FP} * 100 = (404) / (404+0) = (404/404) * 100 = 100.00$$

Interpretación: El aprendizaje automático permite utilizar un algoritmo de Máquina de Vectores-SVM para predecir los resultados de aprendizaje de los alumnos IE N°116 "Abraham Valdelomar" con una especificidad del 100.00%.

### Red – Bayesiana

Tabla 22 Tabla cruzada – cálculo de especificidad con algoritmo Red – Bayesiana

		Predicción					Total	
		1	2	3	4	5		
Rendimiento	1	Recuento	7	0	0	0	0	7
		% dentro Rendimiento	de100,0%	0,0%	0,0%	0,0%	0,0%	100,0%
	2	Recuento	0	27	0	0	0	27
		% dentro Rendimiento	de0,0%	100,0%	0,0%	0,0%	0,0%	100,0%
	3	Recuento	0	0	44	0	0	44
		% dentro Rendimiento	de0,0%	0,0%	100,0%	0,0%	0,0%	100,0%
	4	Recuento	0	0	0	21	0	21
		% dentro Rendimiento	de0,0%	0,0%	0,0%	100,0%	0,0%	100,0%

5	Recuento	0	0	0	0	2	2
	% dentro Rendimiento	de0,0%	0,0%	0,0%	0,0%	100,0%	100,0%
Total	Recuento	7	27	44	21	2	101
	% dentro Rendimiento	de6,9%	26,7%	43,6%	20,8%	2,0%	100,0%

Fuente: SPSS Statistics Ver.26

$$\text{ESPECIFICIDAD} = \frac{\text{VN}}{\text{VN} + \text{FP}} * 100 = \frac{403}{403+1} = \frac{403}{404} * 100 = 99.7$$

Interpretación: El aprendizaje automático permite utilizar un algoritmo de Redes Bayesianas para predecir los resultados de aprendizaje de los alumnos IE N°116 "Abraham Valdelomar" con una especificidad del 99.70%.

**Tabla 23** Cuadro comparativo de resultados según el indicador precisión

ALGORITMO	RESULTADO (%)
Árbol de decisión	90.47 %
SVM	100.00 %
Red - Bayesiana	99.70 %

Fuente: SPSS Statistics Ver.26

Interpretación: La Tabla 23 se concluye que el método que tiene mejor solución para el indicador de precisión es la "Máquina de Vectores (SVM)" con un 100%, a continuación, seguido del "algoritmo de la Red – Bayesiana "con un 99.70% y por último el algoritmo de precisión "Árbol de Decisión" con un resultado igual a 90.47%.

### 5.1. Hipótesis General:

Ho: El Machine Learning no permite predecir el rendimiento académico de los estudiantes de la IE N°116" Abraham Valdelomar"-San Juan de Lurigancho.

H1: El Machine Learning permite predecir el rendimiento académico de los estudiantes de la IE N°116" Abraham Valdelomar"-San Juan de Lurigancho.

**Tabla 24** Resumen de cuadro comparativo de Algoritmos

ALGORITMO	INDICADORES		
	PRECISIÓN	SENSIBILIDAD	ESPECIFICIDAD
Árbol de decisión	84.73 %	61.54 %	90.47 %
SVM	100.00 %	100.00 %	100.00 %
Red- Bayesiana	99.60 %	99.00 %	99.75 %

Fuente: SPSS Statistics Ver.26

Después de procesar los algoritmos de aprendizaje automático en Spss Modeler Ver. 18 según la tabla 24, se puede concluir que los resultados obtenidos del Árbol de Decisión, la Máquina Vectores (SVM), Red Bayesiana, con relación a la "precisión", "la sensibilidad" y "la especificidad", da como resultado que el algoritmo de Máquina de Vectores (SVM), es el mejor para predecir el rendimiento académico, porque otorga puntajes del 100% en todos los indicadores.

De manera similar, se utilizó la índice kappa de Cohen para determinar la afinidad y el nivel de significación entre dos observaciones. Los resultados obtenidos de cada algoritmo se muestran a continuación.

**Tabla 25** Medida de Kappa De cohen-Árbol de decisión

Medidas simétricas

	Valor	Error estándar asintóticoa	T aproximadab	Significación aproximada
Medida de acuerdo	deKappa ,608	,063	9,518	,000
N de casos válidos	101			

Fuente: Elaboración Propia



Interpretación: El valor kappa es 0,608, por lo que existe una concordancia significativa entre el valor observado originalmente y el valor pronosticado del “algoritmo del árbol de decisión”, nos muestra un valor de significación inferior al 5 %. Por tanto, podemos confirmar que los valores obtenidos de la medida Kappa de Cohen y los valores obtenidos del modelo de árbol de decisión nos dan confianza para aceptar la hipótesis alternativa y rechazar la hipótesis nula.

**Tabla 26** Medida de Kappa De cohen-SVM

Medidas simétricas

	Valor	Error estándar asintóticoa	T aproximadab	Significación aproximada
Medida deKappa	1,000	,000	16,202	,000
acuerdo				
N de casos válidos	101			

Fuente: Elaboración Propia

Dado que el valor kappa representa un valor de 1, también existe una coincidencia casi perfecta entre el valor observado original y el valor predicho del algoritmo SVM con un valor de significación inferior al 5 %, por lo que se pueden aceptar hipótesis alternativas y nulas. La hipótesis fue rechazada utilizando el algoritmo SVM.

**Tabla 27** Medida de Kappa De Cohen-Red Bayesiana

Medidas simétricas

	Valor	Error estándar asintóticoa	T aproximadab	Significación aproximada
Medida deKappa	,986	,014	15,969	,000
acuerdo				
N de casos válidos	101			

Fuente: Elaboración Propia

Dado que el valor kappa muestra un valor de 0,986, también existe una coincidencia casi perfecta entre el valor observado inicialmente y el valor predicho por el algoritmo de red bayesiano, con un valor de significación inferior al 5 %, por lo tanto, al usar el algoritmo bayesiano rojo nos da confianza para aceptar la hipótesis alternativa y rechazar la hipótesis nula.

## V. DISCUSIÓN

A continuación, se muestra una discusión realizada utilizando tres métricas sensibilidad, especificidad y precisión, influenciadas por la variable dependiente, denominada rendimiento académico y la variable independiente denominada el aprendizaje automático.

### **Con relación al indicador N° 1. Precisión**

Para este indicador se tomaron modelos de aprendizaje automático (árbol de decisión, máquina vectorial, redes bayesianas), luego de compararlos entre sí, se encontró que el modelo de algoritmo de máquina vectorial (SVM) prometía un 100% de precisión, por lo que encontramos que pronosticar el rendimiento escolar de los estudiantes es lo más apropiado.

El mismo indicador fue sometido a análisis inferencial según la prueba estadística del índice Kappa Cohen, y como resultado, el valor del valor kappa es igual a 1, por lo que existe una concordancia casi perfecta entre los valores observados. al principio y los resultados. Los datos pronosticados del algoritmo vectorial SVM también nos muestran un valor de significancia menor al 5%, podemos concluir que al utilizar el algoritmo SVM Vector Machine se acepta la hipótesis alternativa y se rechaza la hipótesis nula.

La solución de nuestro algoritmo se compara (Burman y Som, 2019) utilizando el mismo algoritmo (SVM) obtenido con una precisión de 90 grados, mientras que el mismo modelo logró una precisión de 100 grados. Otros autores como (Candia, 2019) lograron una precisión de 69% utilizando random forest en su estudio de referencia, Hamoud (2018) encontró que el algoritmo j 8 logra una precisión de 62%, su modelo de Segura (2018) 67. 1. % utilizando árboles de gradiente, Vega (2019) por su parte alcanza un 89% con el algoritmo XGBoosting y finalmente Alsalma (2019) un 90% utilizando redes neuronales..

### **Con relación al indicador N° 2. Sensibilidad**

Para el índice de sensibilidad del modelo de aprendizaje automático, la medida obtenida en nuestro estudio fue del 100% con el algoritmo de máquina vectorial (SVM), que el autor (Burman y Som, 2019) consideró para medir la sensibilidad de su modelo (SVM) a obtener. el resultado. El 90%, en cambio (Vega, 2019) llegaba hasta el 95% y Hamoud (2018) hasta el 63%.

Sobre este indicador también se realizó un análisis de inferencia estadística del índice Kappa Cohen, el cual nos permite comprobar en un primer momento la compatibilidad de los valores observados con los valores predichos por el algoritmo vector máquina SVM. podemos concluir que se acepta la hipótesis alternativa utilizando el algoritmo SVM Vector Machine y se rechaza la hipótesis nula

### **Con relación al indicador N° 3. Especificidad**

Finalmente, el índice de especificidad para los estudios con Máquinas Vectoriales (SVM), el 100%, mientras que los Árboles de Decisión alcanzaron el 90.47%, Chi. el 82,00% y Red Bayesiana 99.75%, respectivamente. Como otros autores (Burman & Som, 2019) y (Vega, 2019) Usando las mismas métricas, logramos resultados del 91 % para el algoritmo SVM y del 83 % para el algoritmo XGBoosting.

Para este indicador también se realizó el analices inferencial estadística del índice Kappa cohen que nos permite verificar la concordancia con los valores Observados en su primer momento y los valores predichos por el Algoritmo Maquina de Vectores-SVM, entonces podemos concluir que se acepta la Hipótesis alterna al utilizar el algoritmo Maquina de Vectores SVM y se rechaza la Hipótesis nula

Con relación al Objetivo General

De acuerdo con la explicación anterior, podemos confirmar que el aprendizaje automático predice el rendimiento académico de los estudiantes de la IE No. 116 "Abraham Valdelomar" porque se encontró que es un indicador aceptable para la variable dependiente "Rendimiento Académico".

Indicador de Precisión. Se utilizaron cuatro modelos de algoritmos de aprendizaje automático; Árboles de Decisión, Máquinas Vectoriales, Redes Bayesianas. En segundo lugar, el modelo mostró resultados del 100% al algoritmo Vector Machine (SVM) después de completar el proceso. Además, el análisis inferencial según la prueba estadística del índice Kappa de Cohen mostró un valor kappa igual a 1. Por lo tanto, la coincidencia de valores es casi perfecta. Los primeros valores observados y valores predichos del algoritmo de máquina vectorial (SVM) también muestran valores significativos por debajo del 5%. Usando estos valores, podemos rechazar la hipótesis nula y aceptar la alternativa.

Indicador de Sensibilidad. Este indicador tuvo en cuenta el uso de algoritmos de indicadores de precisión. Después del procesamiento y la capacitación, SVM-Vector Machines obtuvo un porcentaje del 100 % y el análisis obtuvo el mismo resultado que la métrica de precisión. Usando estos valores, podemos rechazar la hipótesis nula y aceptar la alternativa.

Indicador de Especificidad. Finalmente, el indicador mencionado se procesó con los mismos algoritmos utilizados para los indicadores de precisión y sensibilidad, obteniendo los mismos resultados, de los cuales podemos concluir el rechazo de la hipótesis nula y la aceptación de la hipótesis alternativa.

La metodología de este estudio se basa en el campo del análisis educativo. Tiene como finalidad proponer un modelo con algoritmos de aprendizaje automático que utilicen el aprendizaje automático para generar la capacidad de predecir el progreso del estudiante IE N° 116 "AVRAHAM VALDELOMAR". Para ello, se contrastarán los datos de 101 alumnos de 2022 mediante un cuestionario de 36 preguntas.

Para el tratamiento de la información se utilizaron los siguientes medios técnicos: Usamos SP versión 26 para importar la base de datos, definir las variables y sus respectivos valores, y luego usamos SPSS Modeler versión 18, que nos permite crear modelos de aprendizaje automático.

En cuanto a los indicadores, se puede decir que son de gran utilidad para esta tesis, los cuales pueden medir con precisión y afectar los resultados obtenidos de la variable dependiente. También cabe señalar que este trabajo brinda conocimientos a estudiantes de instituciones públicas y no públicas sobre el entorno estudiantil.

Después de ejecutar los procesos en modelos de algoritmos de aprendizaje automático y encontrar valores aceptables para predecir los resultados del aprendizaje, podemos concluir que el uso del aprendizaje automático nos permite aceptar la hipótesis alternativa que describe "El aprendizaje automático predice el rendimiento académico de los estudiantes". estudiante IE #116 "Abraham Valdelomar" y rechazó la hipótesis nula.

En nuestra investigación utilizamos la metodología KDD (Knowledge Discovery from Databases), un proceso que nos permite identificar modelos válidos, nuevos,

potencialmente útiles y en su mayoría comprensibles que, por sus características, son mucho más inteligentes y requieren menos tiempo, llevar a cabo autores como Chahuán (2019), Canagareddy (2019), Hamoud (2018), Segura (2018) y Soto (2015) utilizaron KDD como referencia para el origen de datos. El autor (Burman and Som, 2019) utilizó su metodología, la cual consta de seis pasos: conjunto de datos de entrada, implementación del algoritmo, entrenamiento con kernel lineal, kernel radial, prueba y estudio comparativo, y finalmente autores como Alsalman (2019), (Candia 2019), (Vega, 2019) y Chiheb (2017) querían estandarizar en cada etapa, es decir profundizar en las tareas y funciones, por lo que se utilizó la metodología CRISP-DM, para poder predecir el rendimiento Académico.

A continuación, se desarrolló un modelo predictivo utilizando SPSS Modeler Ver. 18 debido a su interfaz amigable e interactiva que permitió desarrollar el modelo de manera más rápida y eficiente, mientras que otros autores como Canagareddy (2019), Alsalman (2019), Candia (2019), Vega (2019), Hamoud (2018), Chiheb (2017) y Soto (2015) utilizaron la herramienta Weka, y Segura (2018) utilizó RapidMiner, en esta investigación determino usar el SPP Moleder para poder diseñar el Machine learning y medir el rendimiento académico.

## **VI. CONCLUSIONES**



En la presente investigación realizada se ha llegado a las siguientes conclusiones:

Primera, después de utilizar varios algoritmos como árboles de decisión, máquinas vectoriales (SVM) y red bayesiana, se puede concluir que el modelo de aprendizaje automático proporciona la mayor precisión del rendimiento académico de los estudiantes, lo que demuestra que los vectores se muestran en porcentaje del 100%

Segunda, de manera similar, utilizando varios modelos algorítmicos como árboles de decisión, máquinas vectoriales (SVM) y red bayesiana, se encontró que el modelo de aprendizaje automático proporciona una mejor sensibilidad, es decir, predice casos positivos verdaderos IE N<sup>o</sup>116 "Abraham Valdelomar" el éxito en el aprendizaje del estudiante es. Vector Machine-SVM 100%.

Tercera, además, con la ayuda de varios modelos de algoritmos, como árboles de decisión, máquinas vectoriales (SVM) y red bayesiana, se podría argumentar que el modelo de aprendizaje automático ofrece una mayor especificidad, es decir, se pueden predecir casos positivos verdaderos. El éxito del estudio de los alumnos de IE N<sup>o</sup>116 "Abraham Valdelomar" es Vector Machine-SVM 100%.

Cuarta, por lo tanto, podemos concluir que la Máquinas de Vectores (SVM) y la Red Bayesiana, permiten predecir el rendimiento académico de los estudiantes de la IE No 116 "Abraham Valdelomar" con alta precisión, sensibilidad y especificidad. Además, la similitud entre las observaciones se puede probar utilizando la índice kappa de Cohen, que arrojó 1, lo que indica una casi perfecta concordancia

## **VII. RECOMENDACIONES**

Las recomendaciones para futuras investigaciones son las siguientes:

Dado que el modelo actualmente se entrena con datos escasos obtenidos a través de un cuestionario, recomendamos aplicar otros algoritmos de aprendizaje automático a los datos históricos de los estudiantes para ampliar el panorama predictivo.

Para comprender mejor la situación de los estudiantes durante su estancia en el colegio, recomendamos realizar investigaciones en instituciones privadas y nacionales que sean relevantes para su desempeño académico.

Recomendamos recopilar datos de diferentes colegios y estudiantes de diferentes grados académicos para crear un modelo aplicable a todas las Entidades privadas y públicas.

Recomendamos construir modelos predictivos que utilicen más factores relacionados con el rendimiento académico, como los siguientes: Factores psicológicos, salud, gestión del tiempo, entre otros.

Se recomienda realizar investigaciones sobre el uso de algoritmos combinados para mejorar los algoritmos construidos sobre un solo algoritmo.

Fomentar la creación de software de apoyo institucional para la evaluación de nuevos estudiantes

## REFERENCIAS

- Alvaro. (2018a, April 30). *Machine Learning, Data Science Y Analítica Avanzada*. ¿Qué Es El Aprendizaje Automático o Machine Learning?
- Alvaro. (2018b, April 30). *¿Qué es el aprendizaje automático o machine learning?* <https://machinelearningparatodos.com/que-es-el-aprendizaje-automatiko-o-machine-learning/>
- Barrientos. (2009). *Árboles de decisión como herramienta en el diagnóstico médico*. <https://www.medigraphic.com/pdfs/veracruzana/muv-2009/muv092c.pdf>
- Batanero. (2019). *Aprendizaje por refuerzo y técnicas profundas aplicadas a un sistema de recomendación de venta al por menor*. [file:///C:/Users/soporte/Downloads/anguiano\\_batanero\\_elyo\\_tfm.pdf](file:///C:/Users/soporte/Downloads/anguiano_batanero_elyo_tfm.pdf)
- Bhavsar, P., Safro, I., Bouaynaya, N., Polikar, R., & Dera, D. (2017). Machine Learning in Transportation Data Analytics. *Data Analytics for Intelligent Transportation Systems*, 283–307. <https://doi.org/10.1016/B978-0-12-809715-1.00012-2>
- Burgueño. (2019). *La relación profesor-alumno en la metodología*. <file:///C:/Users/soporte/Downloads/La%20relaci%C3%B3n%20profesor-alumno%20en%20la%20metodolog%C3%ADa.pdf>
- Burman, I., & Som, S. (2019). Predicting Students Academic Performance Using Support Vector Machine. *2019 Amity International Conference on Artificial Intelligence (AICAI)*, 756–759. <https://doi.org/10.1109/AICAI.2019.8701260>
- Candia. (2019). *Predicción Del Rendimiento Académico De Los Estudiantes De La Unsaac A Partir De Sus Datos De Ingreso Utilizando Algoritmos De Aprendizaje Automático* [Universidad Nacional de San Antonio Abad del Cusco]. <http://hdl.handle.net/20.500.12918/4120>
- Cárdenas, J. M. (2018). *El Machine Learning A Través De Los Tiempos, Y Los Aportes A La Humanidad Denniye Hinestroza Ramírez*. <https://repository.unilibre.edu.co/bitstream/handle/10901/17289/EL%20MACHINE%20LEARNING.pdf?sequence=1&isAllowed=y>
- Carmo, J. (2016, November). *Tutorial de Máquina de Vectores de Soporte SVM*. [https://www.researchgate.net/publication/263817587\\_Tutorial\\_sobre\\_Maquinas\\_de\\_Vectores\\_Soporte\\_SVM](https://www.researchgate.net/publication/263817587_Tutorial_sobre_Maquinas_de_Vectores_Soporte_SVM)
- Charris, L. (2018). *Análisis comparativo de algoritmos de árboles de decisión en el procesamiento de datos biológicos*. [file:///C:/Users/soporte/Downloads/emelamed,+Diagramando+3%20\(1\).pdf](file:///C:/Users/soporte/Downloads/emelamed,+Diagramando+3%20(1).pdf)
- Datademia. (2022). *¿Qué es Deep Learning y qué es una red neuronal - Datademia*. <https://datademia.es/blog/que-es-deep-learning-y-que-es-una-red-neuronal>
- Díaz-Landa, B., Meleán-Romero, R., & Marín-Rodríguez, W. (2021). Rendimiento académico de estudiantes en Educación Superior: predicciones de factores influyentes a partir de árboles de decisión. *Telos Revista de Estudios Interdisciplinarios En Ciencias Sociales*, 23(3), 616–639. <https://doi.org/10.36390/telos233.08>
- Fayyad. (1996). *El proceso de descubrimiento de conocimiento en bases de datos (KDD)*. <https://ediciones.ucc.edu.co/index.php/ucc/catalog/download/36/40/230?inline=1>

- Fuentes Gallego, C. (2004). Cálculo del tamaño de la muestra Formación continuada. In *Matronas Profesión* (Vol. 5, Issue 18). <https://www.federacion-matronas.org/wp-content/uploads/2018/01/vol5n18pag5-13.pdf>
- García. (2002). *Resumen Del Libro: El Cuestionario*.
- García. (2003). *El Cuestionario Como Instrumento De Investigación/Evaluación*. [http://www.univsantana.com/sociologia/El\\_Cuestionario.pdf](http://www.univsantana.com/sociologia/El_Cuestionario.pdf)
- García. (2008). *Sistemas de Tutorización Inteligente Basados en Redes Bayesianas*. <file:///C:/Users/soporte/Downloads/SistemasDeTutorizacionInteligenteBasadosEnRedesBay-2585646.pdf>
- García José R. (2019). *Extracción de Conocimiento para la Predicción y Análisis de los Resultados de la Prueba de Calidad de la Educación Superior en Colombia*. <https://www.scielo.cl/pdf/formuniv/v12n4/0718-5006-formuniv-12-04-00055.pdf>
- González Ignacio. (2004). *Realización de un análisis discriminante explicativo del rendimiento académico en la universidad*. <https://revistas.um.es/rie/article/view/98751>
- Guiselle. (2007). *Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública*. <https://www.redalyc.org/pdf/440/44031103.pdf>
- Hernández. (2015). *Rendimiento académico y correspondencias con indicadores de salud física y psicológica*. <https://core.ac.uk/download/pdf/75988294.pdf>
- Lamas, H. A. (2015). School Performance. *Propósitos y Representaciones*, 3(1), 351–386. <https://doi.org/10.20511/pyr2015.v3n1.74>
- Landa. (2019, February 19). *¿Qué es KDD y Minería de Datos?* Página Web. <https://fcojlanda.me/es/ciencia-de-los-datos/kdd-y-mineria-de-datos-espanol/>
- López. (2007). *Las redes bayesianas como herramientas de modelado en psicología*. <https://www.redalyc.org/pdf/167/16723218.pdf>
- Menacho Chiok, C. H. (2017). Predicción del rendimiento académico aplicando técnicas de minería de datos. *Anales Científicos*, 78(1), 26. <https://doi.org/10.21704/ac.v78i1.811>
- Mishra. (2018). *Metrics to Evaluate your Machine Learning Algorithm*. <https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>
- Mohamed. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5/6), 318. <https://doi.org/10.1504/IJTEL.2012.051815>
- Moine. (2013). *Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo*. UNIVERSIDAD NACIONAL DE LA PLATA FACULTAD DE INFORMÁTICA .
- Muestreo no probabilístico: definición, tipos y ejemplos*. (2020). <https://www.questionpro.com/blog/es/muestreo-no-probabilistico/>
- Murillo. (2009). *Investigación Aplicada*. 12–12. [https://www.google.com/search?q=La+investigacion+aplicada%2C+segun+murillo&rlz=1C1GCEU\\_esPE971PE971&oq=La+investigacion+aplicada%2C+segun+murillo&aqs=chrome..69i57j33i10i160l2j33i22i29i30.22629j0j7&sourceid=chrome&ie=UTF-8](https://www.google.com/search?q=La+investigacion+aplicada%2C+segun+murillo&rlz=1C1GCEU_esPE971PE971&oq=La+investigacion+aplicada%2C+segun+murillo&aqs=chrome..69i57j33i10i160l2j33i22i29i30.22629j0j7&sourceid=chrome&ie=UTF-8)

- Navarro. (2003, December). *El Rendimiento Académico: Concepto, Investigación y Desarrollo*.  
 file:///C:/Users/soporte/Downloads/EL%20RENDIMIENTO%20ACAD%C3%89MICO.pdf
- Ojeda, P. C. (2020). *Sesión 4 Universo, población y muestra*.  
<https://www.aacademica.org/cporfirio/18.pdf>
- Ramírez Abad. (2022, August 4). *Universidad Estatal Península De Santa Elena Facultad De Ciencias Sociales Y De La Salud Carrera De Gestión Social Y Desarrollo*.  
<https://repositorio.upse.edu.ec/handle/46000/7050>
- Rendón-Macías, M. E., Villasís-Keever, M. A., & Miranda-Novales, M. G. (2016). Estadística descriptiva. *Revista Alergia Mexico*, 63(4), 397–407.
- Rojas Luis. (2013, January 3). *Validez Predictiva De Los Componentes Del Promedio De Admisión A La Universidad De Costa Rica Utilizando El Género Y El Tipo De Colegio Como Variables Control*.  
<https://revistas.ucr.ac.cr/index.php/aie/article/view/11707/18183>
- Segura-Morales, M., & Loza-Aguirre, E. (2017). Using Decision Trees for Predicting Academic Performance Based on Socio-Economic Factors. *2017 International Conference on Computational Science and Computational Intelligence (CSCI)*, 1132–1136. <https://doi.org/10.1109/CSCI.2017.197>
- Silva. (2018). *Estrategia par ategia para la evaluación r aluación rápida de la estabilidad de un ápida de la estabilidad de un sistema de potencia empleando medición fasorial y aprendizaje supervisado*.  
[https://ciencia.lasalle.edu.co/cgi/viewcontent.cgi?article=1167&context=ing\\_electrica](https://ciencia.lasalle.edu.co/cgi/viewcontent.cgi?article=1167&context=ing_electrica)
- Tourón, J. (1985). *La predicción del rendimiento académico: Procedimientos, resultados e implicaciones*. <https://hdl.handle.net/10171/18774>
- Vega. (2019). *Modelo de pronóstico de rendimiento académico de alumnos en los cursos del programa de estudios básicos de la Universidad Ricardo Palma usando algoritmos de Machine Learning* [Tesis, Universidad Ricardo Palma usando]. <https://hdl.handle.net/20.500.14138/2914>
- zapata. (2014, March). *Método basado en clasificadores k-NN parametrizados con algoritmos genéticos y la estimación de la reactancia para localización de fallas en sistemas de distribución*.  
<https://www.redalyc.org/pdf/430/43030033020.pdf>

## **ANEXOS**

## Anexo 1: Matriz de Consistencia

**TITULO: MACHINE LEARNING PARA PREDECIR EL RENDIMIENTO ACADEMICO DE LA IE No116 ABRAHAM VALDELOMAR  
SAN JUAN DE LURIGANCHO 2022**

**AUTORES: ACOSTA PORTOCARRERO, ANTONY CHRISTIAN  
RUIZ VARGAS, PAUL.**

Problemas	Objetivos	Hipótesis	Variables	Dimensión	Indicadores	Instrumento	Metodología
PG: ¿En qué medida Machine Learning permite predecir el rendimiento académico de la IE No116 Abraham Valdelomar San Juan de Lurigancho 2022	<b>OG:</b> Determinar en qué medida Machine Learning permite predecir el rendimiento académico de la IE No116 Abraham Valdelomar San Juan de Lurigancho 2022	<b>HG:</b> El Machine Learning predice el rendimiento académico de la IE N°116 Abraham Valdelomar.	<b>Variable Independiente:</b> Maching Learning				
Problemas específicos	Objetivos específicos	Hipótesis específicas					
¿En qué medida el Machine Learning predice con precisión el rendimiento académico de la IE N°116 Abraham Valdelomar San Juan de Lurigancho?	Determinar En qué medida el Machine Learning predice con precisión el rendimiento académico de la IE N°116 Abraham Valdelomar San Juan de Lurigancho?	El Machine Learning predice con precisión el rendimiento académico de la IE N°116 Abraham Valdelomar San Juan de Lurigancho	<b>Variable dependiente:</b> <b>Rendimiento Académico</b>	Métricas de precisión	Accuracy= (TP+TN/TP+TN+FP+FN) *100	Formulario	Tipo de Investigación: Aplicada Diseño de Investigación: Experimental de tipo pre-experimental



¿En qué medida el Machine Learning predice la sensibilidad del rendimiento académico de la IE N°116 Abraham Valdelomar San Juan de Lurigancho?	Determinar En qué medida el Machine Learning predice la sensibilidad del rendimiento académico de la IE N°116 Abraham Valdelomar San Juan de Lurigancho?	El Machine Learning predecir la sensibilidad del rendimiento académico de la IE N°116 Abraham Valdelomar de San Juan de Lurigancho			Sensibilidad = $(TP/TP+FN) \cdot 100$	Formulario	
¿En qué medida el Machine Learning permite con especificidad el rendimiento académico de la IE N°116 Abraham Valdelomar San Juan de Lurigancho?	Determinar En qué medida el Machine Learning permite con especificidad el rendimiento académico de la IE N°116 Abraham Valdelomar San Juan de Lurigancho?	El Machine Learning predice con especificidad el rendimiento académico de la IE N°116 Abraham Valdelomar de San Juan de Lurigancho			Especificidad= $(TN/TN+FP) \cdot 100$	Formulario	

## Metodología

TIPO Y DISEÑO	POBLACIÓN Y MUESTRA	TÉCNICAS E INSTRUMENTOS	ESTADÍSTICA POR UTILIZAR
<p><b>Tipo:</b> Aplicada</p> <p><b>Enfoque:</b> Cuantitativo</p> <p><b>Diseño:</b> Experimental – Pre-Experimental</p>	<p><b>Población:</b> 101 observaciones</p> <p><b>Tamaño de muestra:</b> 101 observaciones</p> <p><b>Muestreo:</b> NO-Probabilístico conveniencia</p>	<p><b>Técnicas:</b> Observación</p> <p><b>Instrumentos:</b> Formulario</p>	<p><b>Descriptiva:</b> (Rendón-Macías et al., 2016) mencionan que la estadística descriptiva es la rama de la estadística que hace recomendaciones sobre cómo resumir los datos de las encuestas de forma clara y sencilla en forma de gráficos, tablas, figuras o gráficos.</p> <p>Para el análisis descriptivo se calculará la media de los datos recolectados por cada indicador en las etapas del pre test y post test, para poder visualizar el efecto de la variable independiente sobre la variable dependiente.</p> <p><b>Inferencial:</b> Se procesaron los datos recolectados con el Coeficiente kappa de Cohen para comprobar la concordancia observada en un conjunto de datos, para compararlo con la estadística predictiva y contrastar la hipótesis general y específica.</p>

**Anexo: N° 2**  
**Operacionalización de Variables:**

**TITULO : MACHINE LEARNING PARA PREDECIR EL RENDIMIENTO ACADEMICO DE LA IE N°116 ABRAHAM VALDELOMAR-SAN JUAN DE LURIGANCHO 2022**

**AUTORES: ACOSTA PORTOCARRERO, ANTONY CHRISTIAN  
RUIZ VARGAS, PAUL.**

Variable	Definición	Definición Operacional	Dimensiones	Indicadores	Escala de Medición
Independiente: Maching Learning	“El aprendizaje automático se basa en algoritmos que pueden aprender de los datos sin depender de la programación basada en reglas” (McKinsey & Co)				Razón
Dependiente: rendimiento Académico	Partiendo del concepto de rendimiento académico muchos autores entre los cuales tenemos a Jiménez (como se citó en Navarro, 2014, párr.11) lo define como el “nivel de conocimiento demostrado por el alumno en un área o materia comparado con la norma de edad y nivel académico”,	Para medir la mejora del rendimiento académico se dan usos de las métricas de precisión, la cual será obtenida a través de la una herramienta de Machine Learning.	Métricas de Precisión	Accuracy= $(TP+TN/TP+TN+FP+FN) *100$ Sensibilidad = $(TP/TP+FN) *100$ Especificidad= $(TN/TN+FP) *100$	Porcentaje

### ANEXO 3: INSTRUMENTO DE OBSERVACION PARA EL ALGORITMO SVM

Tipo de Prueba	Post Test
Investigador	Acosta Portocarrero Antony christian & Ruiz Vargas Paul
Fecha de inicio	

Algoritmo	SVM
-----------	-----

Matriz de confusión:

		predicción	
		Positive	Negative
observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Negative	(TN) True Negative

Métricas a Evaluar:

Ítem	Indicador	Media	Formula	Precisión
1	Precisión (Exactitud)	Razón	$Accuracy = (TP+TN/TP+TN+FP+FN) * 100$	
2	Especificidad	Razón	$Especificidad = (TN/TN+FP) * 100$	
3	Sensibilidad	Razón	$Sensibilidad = (TP/TP+F N) * 100$	

Otras Métricas:

ítem	Indicador	Medida	Formula	Precisión
1	F 1 SCORE	Razón	$2*(Recall*precisión)/(Recall+Precision)$	

Tipo de Prueba	Pre Test
Investigador	Acosta Portocarrero Antony christian & Ruiz Vargas Paul
Fecha de inicio	

Algoritmo	SVM
-----------	-----

Matriz de confusión:

		predicción	
		Positive	Negative
observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Negative	(TN) True Negative

Métricas a Evaluar:

Ítem	Indicador	Medida	Formula	Precisión
1	Precisión (Exactitud)	Razón	$Accuracy = (TP + TN / TP + TN + FP + FN) * 100$	
2	Especificidad	Razón	$Especificidad = (TN / TN + FP) * 100$	
3	Sensibilidad	Razón	$Sensibilidad = (TP / TP + FN) * 100$	

Otras Métricas:

ítem	Indicador	Medida	Formula	Precisión
1	F 1 SCORE	Razón	$2 * (Recall * precisión) / (Recall + Precisión)$	

## INSTRUMENTO DE OBSERVACION PARA EL ALGORITMO ARBOL DE DECISION

Tipo de Prueba	Post Test
Investigador	Acosta Portocarrero Antony christian & Ruiz Vargas Paul
Fecha de inicio	

Algoritmo	Árbol de decisión
-----------	-------------------

Matriz de confusión:

		predicción	
		Positive	Negative
observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Negative	(TN) True Negative

Métricas a Evaluar:

Ítem	Indicador	Media	Formula	Precisión
1	Precisión (Exactitud)	Razón	$Accuracy = (TP+TN/TP+TN+FP+FN) * 100$	84.73 %
2	Especificidad	Razón	$Especificidad = (TN/TN+FP) * 100$	61.54 %
3	Sensibilidad	Razón	$Sensibilidad = (TP/TP+FN) * 100$	90.47 %

Tipo de Prueba	PreTest
Investigador	Acosta Portocarrero Antony christian & Ruiz Vargas Paul
Fecha de inicio	

Algoritmo	Árbol de decisión
-----------	-------------------

Matriz de confusión:

		predicción	
		Positive	Negative
observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Negative	(TN) True Negative

Métricas a Evaluar:

Ítem	Indicador	Media	Formula	Precisión
1	Precisión (Exactitud)	Razón	$Accuracy = (TP+TN/TP+TN+FP+FN) * 100$	
2	Especificidad	Razón	$Especificidad = (TN/TN+FP) * 100$	
3	Sensibilidad	Razón	$Sensibilidad = (TP/TP+FN) * 100$	

## INSTRUMENTO DE OBSERVACION PARA EL ALGORITMO SVM

Tipo de Prueba	Post Test
Investigador	Acosta Portocarrero Antony christian & Ruiz Vargas Paul
Fecha de inicio	

Algoritmo	SVM
-----------	-----

Matriz de confusión:

		predicción	
		Positive	Negative
observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Negative	(TN) True Negative

Métricas a Evaluar:

Ítem	Indicador	Media	Formula	Precisión
1	Precisión (Exactitud)	Razón	$Accuracy = (TP + TN / TP + TN + FP + FN) * 100$	100.00 %
2	Especificidad	Razón	$Especificidad = (TN / TN + FP) * 100$	100.00 %
3	Sensibilidad	Razón	$Sensibilidad = (TP / TP + FN) * 100$	100.00 %



Tipo de Prueba	Pre Test
Investigador	Acosta Portocarrero Antony christian & Ruiz Vargas Paul
Fecha de inicio	

Algoritmo	SVM
-----------	-----

Matriz de confusión:

		predicción	
		Positive	Negative
observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Negative	(TN) True Negative

Métricas a Evaluar:

Ítem	Indicador	Media	Formula	Precisión
1	Precisión (Exactitud)	Razón	$Accuracy = (TP + TN / TP + TN + FP + FN) * 100$	
2	Especificidad	Razón	$Especificidad = (TN / TN + FP) * 100$	
3	Sensibilidad	Razón	$Sensibilidad = (TP / TP + FN) * 100$	

## INSTRUMENTO DE OBSERVACION PARA EL ALGORITMO REDES BAYESIANAS

Tipo de Prueba	Post Test
Investigador	Acosta Portocarrero Antony christian & Ruiz Vargas Paul
Fecha de inicio	

Algoritmo	Redes Bayesianas
-----------	------------------

Matriz de confusión:

		predicción	
		Positive	Negative
observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Negative	(TN) True Negative

Métricas a Evaluar:

Ítem	Indicador	Media	Formula	Precisión
1	Precisión (Exactitud)	Razón	$Accuracy = (TP + TN / TP + TN + FP + FN) * 100$	99.60 %
2	Especificidad	Razón	$Especificidad = (TN / TN + FP) * 100$	99.00 %
3	Sensibilidad	Razón	$Sensibilidad = (TP / TP + FN) * 100$	99.75 %

Tipo de Prueba	Pre Test
Investigador	Acosta Portocarrero Antony Christian & Ruiz Vargas Paul
Fecha de inicio	

Algoritmo	Redes Bayesianas
-----------	------------------

Matriz de confusión:

		predicción	
		Positive	Negative
observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Negative	(TN) True Negative

Métricas a Evaluar:

Ítem	Indicador	Media	Formula	Precisión
1	Precisión (Exactitud)	Razón	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100$	
2	Especificidad	Razón	$Especificidad = \frac{TN}{TN+FP} * 100$	
3	Sensibilidad	Razón	$Sensibilidad = \frac{TP}{TP+FN} * 100$	

#### Anexo 4: Certificado de Validez Expertos

#### Certificado de validez de contenido del instrumento

#### Validación del Experto N°1

**Apellidos y Nombre del experto:** Mg. Fierro Barriales Alan Leoncio

**Nombre del instrumento:** Ficha de Registro

**Autores:** Acosta Portocarrero, Antony Christian y Ruíz Vargas Paul

**Título de investigación:** Machine Learning para predecir el rendimiento académico de los estudiantes de la IE N°116 Abraham Valdelomar, San Juan de Lurigancho 2022.

INDICADORES	CRITERIOS	Deficiente 0-20%	Regular 21-50%	Buen o51- 70%	Muy Buen o 71-80%	Excelent e81- 100%
Claridad	Promedio de Validación					95
Objetividad	Esta expresado en conducta observable.					95
Actualidad	Es adecuado al avance de la ciencia.					95
Organización	Existe una organización lógica					95
Suficiencia	Comprende los aspectos de cantidad y calidad.					95
Intencionalidad	Adecuado para valorar aspectos del sistema metodológico y científico					95
Consistencia	Está basado en aspectos teóricos, científicos acordes a la tecnología educativa.					95
Coherencia	Está basado en aspectos teóricos, científicos acordes a la tecnología educativa.					95
Metodología	Responde al propósito del trabajo bajo los objetivos a lograr					95
Pertinencia	El instrumento es adecuado al tipo de investigación.					95
<b>Promedio de Validación</b>						

Promedio de Valoración: 95%

Observaciones:

## Validación del Experto N°1

N°	INDICADORES	Claridad <sup>1</sup>		Pertinencia <sup>2</sup>		Relevancia <sup>3</sup>		Sugerencias
		Si	No	Si	No	Si	No	
1	EXACTITUD	X		X		X		
2	SENSIBILIDAD	X		X		X		
3	ESPECIFICIDAD	X		X		X		

Observaciones (precisar si hay suficiencia):

---

Opinión de aplicabilidad:   Aplicable    Aplicable después de corregir    No aplicable

Apellidos y nombres del juez evaluador: FIERRO BARRIALES, ALAN LEONCIO

DNI: 44147992

Lima, 24 de agosto 2022

Especialista: Metodólogo    Temático

Grado: Maestro    Doctor

<sup>1</sup> Claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

<sup>2</sup> Pertinencia: Si el ítem pertenece a la dimensión.

<sup>3</sup> Relevancia: El ítem es apropiado para representar al componente o dimensión específica del constructo

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión



Fierro Barriales, Alan Leoncio  
DNI 44147992  
Universidad Cesar Vallejo

## Certificado de validez de contenido del instrumento

### Validación del Experto N°2

**Apellidos y Nombre del experto:** Dr. Orlando Gustavo Díaz López

**Nombre del instrumento:** Ficha de Registro

**Autores:** Acosta Portocarrero, Antony Christian y Ruíz Vargas Paul

**Título de investigación:** Machine Learning para predecir el rendimiento académico de los estudiantes de la IE N°116 Abraham Valdelomar, San Juan de Lurigancho 2022.

INDICADORES	CRITERIOS	Deficiente 0-20%	Regular 21-50%	Bueno 51- 70%	Muy Bueno 71-80%	Excelente 81-100%
Claridad	Promedio de Validación					93
Objetividad	Esta expresado en conducta observable.					93
Actualidad	Es adecuado al avance de la ciencia.					93
Organización	Existe una organización lógica					93
Suficiencia	Comprende los aspectos de cantidad y calidad.					93
Intencionalidad	Adecuado para valorar aspectos del sistema metodológico y científico					93
Consistencia	Está basado en aspectos teóricos, científicos acordes a la tecnología educativa.					93
Coherencia	Está basado en aspectos teóricos, científicos acordes a la tecnología educativa.					93
Metodología	Responde al propósito del trabajo bajo los objetivos a lograr					93
Pertinencia	El instrumento es adecuado al tipo de investigación.					93
<b>Promedio de Validación</b>						

Promedio de Valoración: 93%

Observaciones:

## Validación del Experto N°2

Variable: RENDIMIENTO ACADÉMICO

N°	INDICADORES	Claridad <sup>1</sup>		Pertinencia <sup>2</sup>		Relevancia <sup>3</sup>		Sugerencias
		Si	No	Si	No	Si	No	
1	EXACTITUD	X		X		X		
2	SENSIBILIDAD	X		X		X		
3	ESPECIFICIDAD	X		X		X		

Observaciones (precisar si hay suficiencia):

Opinión de aplicabilidad:    Aplicable [ X ]            Aplicable después de corregir [ ]            No aplicable [ ]

Apellidos y nombres del juez evaluador: Dr. Orlando Gustavo Díaz López    DNI: 08820518

Especialista: Metodólogo [X]    Temático [ ]

Grado: Maestro [ ]    Doctor [X]    Ingeniero [ ]

<sup>1</sup> claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

<sup>2</sup> pertinencia: Si el ítem pertenece a la dimensión.

<sup>3</sup> relevancia: El ítem es apropiado para representar al componente o dimensión específica del constructo

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión

Lima, 29 de agosto de 2022



Dr. Orlando Gustavo Díaz López  
DNI N° 08820518

## Certificado de validez de contenido del instrumento

### Validación del Experto N°3

**Apellidos y Nombre del experto:** DR. Pedro Peña Huapaya

**Nombre del instrumento:** Ficha de Registro

**Autores:** Acosta Portocarrero, Antony Christian y Ruíz Vargas Paul

**Título de investigación:** Machine Learning para predecir el rendimiento académico de los estudiantes de la IE N°116 Abraham Valdelomar, San Juan de Lurigancho 2022.

INDICADORES	CRITERIOS	Deficiente 0-20%	Regular 21-50%	Bueno 51- 70%	Muy Bueno 71-80%	Excelente 81-100%
Claridad	Promedio de Validación					99
Objetividad	Esta expresado en conducta observable.					99
Actualidad	Es adecuado al avance de la ciencia.					97
Organización	Existe una organización lógica					95
Suficiencia	Comprende los aspectos de cantidad y calidad.					98
Intencionalidad	Adecuado para valorar aspectos del sistema metodológico y científico					99
Consistencia	Está basado en aspectos teóricos, científicos acordes a la tecnología educativa.					98
Coherencia	Está basado en aspectos teóricos, científicos acordes a la tecnología educativa.					95
Metodología	Responde al propósito del trabajo bajo los objetivos a lograr					97
Pertinencia	El instrumento es adecuado al tipo de investigación.					97
<b>Promedio de Validación</b>						

Promedio de Valoración: 97.4%

Observaciones:



### Validación del Experto N°3

#### Variable: RENDIMIENTO ACADÉMICO

Nº	INDICADORES	Claridad <sup>1</sup>		Pertinencia <sup>2</sup>		Relevancia <sup>3</sup>		Sugerencias
		Si	No	Si	No	Si	No	
1	EXACTITUD	X		X		X		
2	SENSIBILIDAD	X		X		X		
3	ESPECIFICIDAD	X		X		X		

Observaciones (precisar si hay suficiencia):

---

Opinión de aplicabilidad:    **Aplicable [  ]**            **Aplicable después de corregir [  ]**            **No aplicable [  ]**

Apellidos y nombres del juez evaluador:    **DR. PEDRO PEÑA HUAPAYA**            **DNI: 80593277**

Especialista: **Metodólogo [  ]**    **Temático [  ]**

**Grado: Maestro [  ]    Doctor [  ]    Ingeniero [  ]**


<sup>1</sup> **Claridad:** Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

<sup>2</sup> **Pertinencia:** Si el ítem pertenece a la dimensión.

<sup>3</sup> **Relevancia:** El ítem es apropiado para representar al componente o dimensión específica del constructo

**Nota:** Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión

Lima, 29 agosto de 2022



DR. PEDRO PEÑA HUAPAYA  
DNI 80593277

## Anexo 5: Constancia de Grados

### Constancia de Grados y títulos de validadores (SUNEDU)

#### VALIDADOR 1:

18/9/22, 11:00



**PERÚ**

Ministerio de Educación

Superintendencia Nacional de  
Educación Superior Universitaria

Dirección de Documentación e  
Información Universitaria y  
Registro de Grados y Títulos

### REGISTRO NACIONAL DE GRADOS ACADÉMICOS Y TÍTULOS PROFESIONALES

Graduado	Grado o Título	Institución
FIERRO BARRIALES,ALAN LEONCIO DNI 44147992	<b>INGENIERO DE SISTEMAS</b> Fecha de diploma: 08/07/2013 Modalidad de estudios: -	UNIVERSIDAD PRIVADA CÉSAR VALLEJO <b>PERU</b>
FIERRO BARRIALES,ALAN LEONCIO DNI 44147992	<b>BACHILLER EN INGENIERIA DE SISTEMAS</b> Fecha de diploma: 17/05/2013 Modalidad de estudios: -  Fecha matrícula: Sin información (***) Fecha egreso: Sin información (***)	UNIVERSIDAD PRIVADA CÉSAR VALLEJO <b>PERU</b>
FIERRO BARRIALES,ALAN LEONCIO DNI 44147992	<b>MAESTRO EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN</b> Fecha de diploma: 10/12/18 Modalidad de estudios: PRESENCIAL  Fecha matrícula: 20/01/2017 Fecha egreso: 19/08/2018	UNIVERSIDAD PRIVADA CÉSAR VALLEJO <b>PERU</b>

## Constancia de Grados y títulos de validadores (SUNEDU)

### VALIDADOR 2:

29/8/22, 12:18



**PERÚ**

Ministerio de Educación

Superintendencia Nacional de  
Educación Superior Universitaria

Dirección de Documentación e  
Información Universitaria y  
Registro de Grados y Títulos

### REGISTRO NACIONAL DE GRADOS ACADÉMICOS Y TÍTULOS PROFESIONALES

Graduado	Grado o Título	Institución
DIAZ LOPEZ, ORLANDOGUSTAVO DNI 08820518	<b>LIC. SOCIOLOGIA</b> Fecha de diploma: 24/05/1990 Modalidad de estudios: -	UNIVERSIDAD NACIONAL FEDERICO VILLARREAL <b>PERU</b>
DIAZ LOPEZ, ORLANDOGUSTAVO DNI 08820518	<b>LICENCIADO EN SOCIOLOGIA</b> Fecha de diploma: 24/05/1990 Modalidad de estudios: -	UNIVERSIDAD NACIONAL FEDERICO VILLARREAL <b>PERU</b>
DIAZ LOPEZ, ORLANDOGUSTAVO DNI 08820518	<b>BACHILLER EN SOCIOLOGIA</b> Fecha de diploma: 12/09/1986 Modalidad de estudios: -  Fecha matrícula: Sin información (***) Fecha egreso: Sin información (***)	UNIVERSIDAD NACIONAL FEDERICO VILLARREAL <b>PERU</b>
DIAZ LOPEZ, ORLANDOGUSTAVO DNI 08820518	<b>BACHILLER EN EDUCACION</b> Fecha de diploma: 05/12/2007 Modalidad de estudios: -  Fecha matrícula: Sin información (***) Fecha egreso: Sin información (***)	UNIVERSIDAD PRIVADA SANPEDRO <b>PERU</b>
DIAZ LOPEZ, ORLANDOGUSTAVO DNI 08820518	<b>LICENCIADO EN EDUCACION SECUNDARIA EN LA ESPECIALIDAD DE HISTORIA, GEOGRAFIA Y CIENCIAS SOCIALES</b> Fecha de diploma: 18/11/2009 Modalidad de estudios: -	UNIVERSIDAD SAN PEDRO <b>PERU</b>
DIAZ LOPEZ, ORLANDOGUSTAVO DNI 08820518	<b>MAGISTER EN DOCENCIA UNIVERSITARIA E INVESTIGACION PEDAGOGICA</b> Fecha de diploma: 22/12/2011 Modalidad de estudios: -  Fecha matrícula: Sin información (***) Fecha egreso: Sin información (***)	UNIVERSIDAD SAN PEDRO <b>PERU</b>

<p>DIAZ LOPEZ, ORLANDOGUSTAVO DNI 08820518</p>	<p><b>DOCTOR EN GESTIÓN Y CIENCIAS DE LA EDUCACIÓN</b> Fecha de diploma: 04/04/16 Modalidad de estudios: PRESENCIAL</p> <p>Fecha matrícula: 17/03/2012 Fecha egreso: 31/07/2015</p>	<p>UNIVERSIDAD SAN PEDRO <i>PERU</i></p>
--	---	--

# Constancia de Grados y títulos de validadores (SUNEDU)

## Validador 3

29/8/22, 12:38



**PERÚ**

Ministerio de Educación

Superintendencia Nacional de  
Educación Superior Universitaria

Dirección de Documentación e  
Información Universitaria y  
Registro de Grados y Títulos

### REGISTRO NACIONAL DE GRADOS ACADÉMICOS Y TÍTULOS PROFESIONALES

Graduado	Grado o Título	Institución
PEÑA HUAPAYA, PEDROSATURNINO DNI 80593277	<b>BACHILLER EN ECONOMIA</b> Fecha de diploma: 10/01/1986 Modalidad de estudios: -  Fecha matrícula: Sin información (***) Fecha egreso: Sin información (***)	UNIVERSIDAD DE SAN MARTÍN DEPORRES <b>PERU</b>
PEÑA HUAPAYA, PEDROSATURNINO DNI 80593277	<b>ECONOMISTA</b> Fecha de diploma: 05/06/1986 Modalidad de estudios: -	UNIVERSIDAD DE SAN MARTÍN DEPORRES <b>PERU</b>
PEÑA HUAPAYA, PEDROSATURNINO DNI 80593277	<b>MAESTRO EN MARKETING Y NEGOCIOS INTERNACIONALES</b> Fecha de diploma: 16/03/2011 Modalidad de estudios: -  Fecha matrícula: Sin información (***) Fecha egreso: Sin información (***)	UNIVERSIDAD NACIONAL FEDERICOVILLARREAL <b>PERU</b>
PEÑA HUAPAYA, PEDROSATURNINO DNI 80593277	<b>DOCTOR EN PSICOLOGÍA</b> Fecha de diploma: 31/03/17 Modalidad de estudios: <b>PRESENCIAL</b>  Fecha matrícula: 09/03/2015 Fecha egreso: 09/07/2015	UNIVERSIDAD ALAS PERUANAS S.A. <b>PERU</b>

## Anexo 6: Instrumento Para Recolección De Datos

Item	Factor	Componente	Valor
1	Personal	¿Género?	Masculino / Femenino
2		¿Edad?	Dato Numérico
3		¿Grado de estudios?	
4		¿Distrito en el que vives?	
5		¿Trabajas?	SI / NO
6		¿Consumes drogas?	SI / NO
7		¿Consumes licor?	SI / NO
8		¿Llegas puntual a las clases?	Siempre/Casi siempre/Nunca
9		¿Pertenece a alguna pandilla?	SI / NO
10	Auto concepto	¿Qué tan responsable se considera?	Responsable/ Muy Responsable Irresponsable/Muy responsable/
11		¿Considera posible terminar el año escolar?	SI / NO / Tal vez
12		¿Cómo te consideras en los estudios?	Excelente / Bueno / Regular / Malo
13		¿Cómo te consideras para estos cursos: ¿Matemáticas, Física, Religión y Lenguaje?	Muy bueno / Bueno / Regular / Muy malo / Malo
14	Motivación	¿Te gusta estudiar?	SI / NO
15		¿Estudian en un buen ambiente físico?	SI / NO
16		¿Te gusta estudiar en grupo?	SI / NO / Tal vez
17		¿Cuántas horas al día son dedicadas al estudio?	Respuesta corta (Cadena)
18		¿Si te dieran a elegir entre estudiar o no estudiar? ¿Estudiarías?	SI / NO
19		¿Cuál de las siguientes redes sociales usas para buscar información para tus tareas?	Google / Facebook / Twitter / Instagram / Wasap / Tik Tok
20	Socio cultural	¿Cuántos integrantes hay en su grupo familiar?	Dato Numérico
21		¿Cuántos trabajan en tu familia?	Dato Numérico
22		¿Cuántos de su familia estudian?	Dato Numérico
23		¿Cuántos de sus familiares reciben pensión del Estado?	Dato Numérico
24		¿Dependes de tus padres?	SI / NO
25		¿Dependes de alguna persona?	Padre / Madre / Tío / Tía / Otra
26	Nivel de Educación de los padres	¿Nivel de educación de su padre?	Inicial/ Primaria / Secundaria / Técnico / Superior
27		¿Qué nivel de educación posee su madre?	Inicial/ Primaria / Secundaria / Técnico / Superior
28	Emociones	¿De qué manera el stress influye en su vida escolar?	Muy negativamente / Negativamente / Muy Positivamente / Muypositivamente
29		¿Tiene empatía hacia sus compañeros?	Sin empatía / Muy poca empatía / Muy empático / Empático / Indiferente
30		¿Cuándo obtienes malas notas? ¿Cómo Te Sientes?	Frustrado / Molesto / Triste / Todas / Ninguna
31		¿Qué cantidad de amigos te gustaría tener?	Muchos / Pocos / Regular / Ninguno
32	Económico	¿Cantidad de horas que trabaja a la semana?	Dato numérico
33		¿Ingreso familiar aproximado mensualmente?	Dato numérico
34		¿Sus estudios iniciales en que establecimiento lo realizó?	Público / Privado / En casa
35		¿Establecimiento en el que estudia es?	Público / Privado
36	Variable objetivo	¿Promedio ponderado durante el año?	Dato numérico

## Anexo 7: Base de Datos

Figura 1: Datos importados de Google Forms a Excel

	AD	AE	AF	AG	AH	AI	AJ	AK	AL
1	Tiene empatía hacia sus	Cuando obtienes malas r	Que cantidad de amigos	Cantidad de horas que tr	Ingreso familiar aproxima	Sus estudios iniciales en	Establecimiento en el qu	Promedio ponderado durante el año	
7	Empático	Frustrado	Regular	Ninguna	4000soles	Público	Público		16
8	Muy empático	Frustrado	Regular	ninguna	4500 soles	Privado	Público		17
9	Empático	Todas	Regular	No trabajo	5000 soles	Privado	Público		16
10	Muy empático	Triste	Muchos	NOP TRABAJO	3000 SOLES	Público	Público		15
11	Empático	Frustrado	Pocos	9horas	2500soles	Público	Público		15
12	Empático	Triste	Regular	No trabajo	4000 soles	Privado	Público		18
13	Empático	Triste	Muchos	60	3500	Público	Público		20
14	Muy empático	Todas	Muchos	12 horas	2200 soles	Público	Público		13
15	Empático	Todas	Regular	0	5000	Público	Público		17
16	Muy empático	Triste	Regular	NO TRABAJO	3500 SOLES	Público	Público		17
17	Muy empático	Frustrado	Regular	60	3500	Público	Público		20
18	Empático	Molesto	Pocos	NO TRABAJO	5000 SOLES	Público	Público		16
19	Empático	Triste	Muchos	8	3000	Público	Público		15
20	Empático	Frustrado	Regular	8	3000	Privado	Público		18
21	Empático	Triste	Regular	NO TRABAJO	4500 SOLES	Privado	Público		16
22	Empático	Todas	Regular	5	1300	Público	Público		14
23	Empático	Triste	Regular	NO TRABAJO	6000 SOLES	Privado	Público		17
24	Muy empático	Triste	Regular	No trabajo	4000	Público	Público		18

Fuente: Elaboración Propia

Figura 2: Definición de las Variables

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Marcatemporal	Fecha	40	0	Marca temporal	Ninguna	Ninguna	11	Derecha	Escala	Entrada
2	Género	Númérico	8	0	Genero	{1, Masculin...	Ninguna	9	Derecha	Nominal	Entrada
3	Edad	Númérico	8	0	Edad	Ninguna	Ninguna	6	Derecha	Escala	Entrada
4	Gradodeestudios	Númérico	8	0	Grado de estud...	{1, 5to_año}...	Ninguna	9	Derecha	Nominal	Entrada
5	Distritoenelquevives	Númérico	8	0	Distrito en el qu...	{1, San juan...	Ninguna	14	Derecha	Nominal	Entrada
6	Trabajas	Númérico	8	0	Trabajas	{1, Si}...	Ninguna	5	Derecha	Nominal	Entrada
7	Consumesdrogas	Númérico	8	0	Consumes drog...	{1, Si}...	Ninguna	6	Derecha	Nominal	Entrada
8	Consumeslicor	Númérico	8	0	Consumes licor	{1, Si}...	Ninguna	3	Derecha	Nominal	Entrada
9	Llegaspuntualalascases	Númérico	8	0	Llegas puntual ...	{1, Siempre}...	Ninguna	12	Derecha	Nominal	Entrada
10	Perteneceaalgunapandilla	Númérico	8	0	Pertenece a al...	{1, Si}...	Ninguna	3	Derecha	Nominal	Entrada
11	Quétanresponsableseconsidera	Númérico	8	0	Qué tan respon...	{1, Muy res...	Ninguna	17	Derecha	Nominal	Entrada
12	Considera posibleterminarrelañoescolar	Númérico	8	0	Considera posi...	{1, Si}...	Ninguna	7	Derecha	Nominal	Entrada
13	Cómoteconsideras enlosetudios	Númérico	8	0	Cómo te consid...	{1, Excelent...	Ninguna	9	Derecha	Nominal	Entrada
14	Cómoteconsideras paraestoscursosMatemáticasFisi...	Númérico	8	0	Cómo te consid...	{1, Muy Bue...	Ninguna	9	Derecha	Nominal	Entrada
15	Tegustaestudiar	Númérico	8	0	Te gusta estudiar	{1, Si}...	Ninguna	7	Derecha	Nominal	Entrada
16	Estudianenunbuenambiente físico	Númérico	8	0	Estudian en un ...	{1, Si}...	Ninguna	7	Derecha	Nominal	Entrada
17	Tegustaestudiarengroupo	Númérico	8	0	Te gusta estudi...	{1, Si}...	Ninguna	7	Derecha	Nominal	Entrada
18	Cuántashorasaldíasondedicadaalestudio	Númérico	8	0	Cuántas horas ...	Ninguna	Ninguna	12	Derecha	Nominal	Entrada
19	Sitedieranaelegirentrestudiaronoestudiar ¿Estudiar...	Númérico	8	0	Si te dieran a el...	{1, Si}...	Ninguna	14	Derecha	Nominal	Entrada
20	Cuálde lassiguientesredessocialesusas para buscarin...	Númérico	8	0	Cuál de las sig...	Ninguna	Ninguna	13	Derecha	Nominal	Entrada
21	Cuántos integrantes hay en su grupo familiar	Númérico	8	0	Cuántos integra...	Ninguna	Ninguna	12	Derecha	Nominal	Entrada
22	Cuántostrabajan en su familia	Númérico	8	0	Cuántos trabaja...	Ninguna	Ninguna	12	Derecha	Nominal	Entrada
23	Cuántos de su familia estudian	Númérico	8	0	Cuántos de su f...	Ninguna	Ninguna	12	Derecha	Nominal	Entrada
24	Cuántos de sus familiares reciben pensión del Estado	Númérico	8	0	Cuántos de sus...	Ninguna	Ninguna	12	Derecha	Nominal	Entrada

Fuente: Elaboración Propia.

**Figura 3:** Salida de datos con nodo resultado, tipo tabla - SPSS

	Género	Edad	Grado de estudios	Distrito en el que vive	Trabaja	Consumes drogas	Consumes licor	Llega puntual a las clases	Pertenece a algún pandilla	¿Quién es responsable?
1	2.000	16.0...	1.000	1.000	2.000	2.000	2.000	3.000	2.000	
2	1.000	16.0...	1.000	1.000	2.000	2.000	2.000	1.000	2.000	
3	1.000	16.0...	1.000	1.000	2.000	2.000	2.000	1.000	1.000	
4	2.000	17.0...	1.000	1.000	2.000	2.000	2.000	1.000	2.000	
5	2.000	15.0...	1.000	2.000	2.000	2.000	2.000	3.000	2.000	
6	2.000	16.0...	1.000	1.000	2.000	2.000	2.000	1.000	2.000	
7	2.000	15.0...	2.000	1.000	1.000	2.000	2.000	3.000	2.000	
8	2.000	15.0...	2.000	1.000	2.000	2.000	2.000	1.000	2.000	
9	1.000	16.0...	1.000	1.000	2.000	2.000	2.000	1.000	2.000	
10	1.000	16.0...	1.000	1.000	2.000	2.000	2.000	3.000	2.000	
11	2.000	14.0...	2.000	1.000	2.000	2.000	2.000	3.000	2.000	
12	2.000	16.0...	1.000	1.000	2.000	2.000	2.000	1.000	2.000	
13	1.000	15.0...	2.000	1.000	1.000	2.000	2.000	3.000	2.000	
14	1.000	16.0...	1.000	1.000	1.000	2.000	2.000	1.000	2.000	
15	2.000	15.0...	2.000	1.000	1.000	2.000	2.000	3.000	2.000	
16	1.000	17.0...	1.000	1.000	1.000	1.000	2.000	1.000	1.000	
17	1.000	15.0...	2.000	1.000	2.000	2.000	2.000	3.000	2.000	
18	1.000	16.0...	1.000	1.000	2.000	2.000	2.000	3.000	2.000	
19	1.000	15.0...	2.000	1.000	2.000	2.000	2.000	1.000	2.000	
20	1.000	15.0...	2.000	1.000	2.000	2.000	2.000	3.000	2.000	

Fuente: Elaboración propia

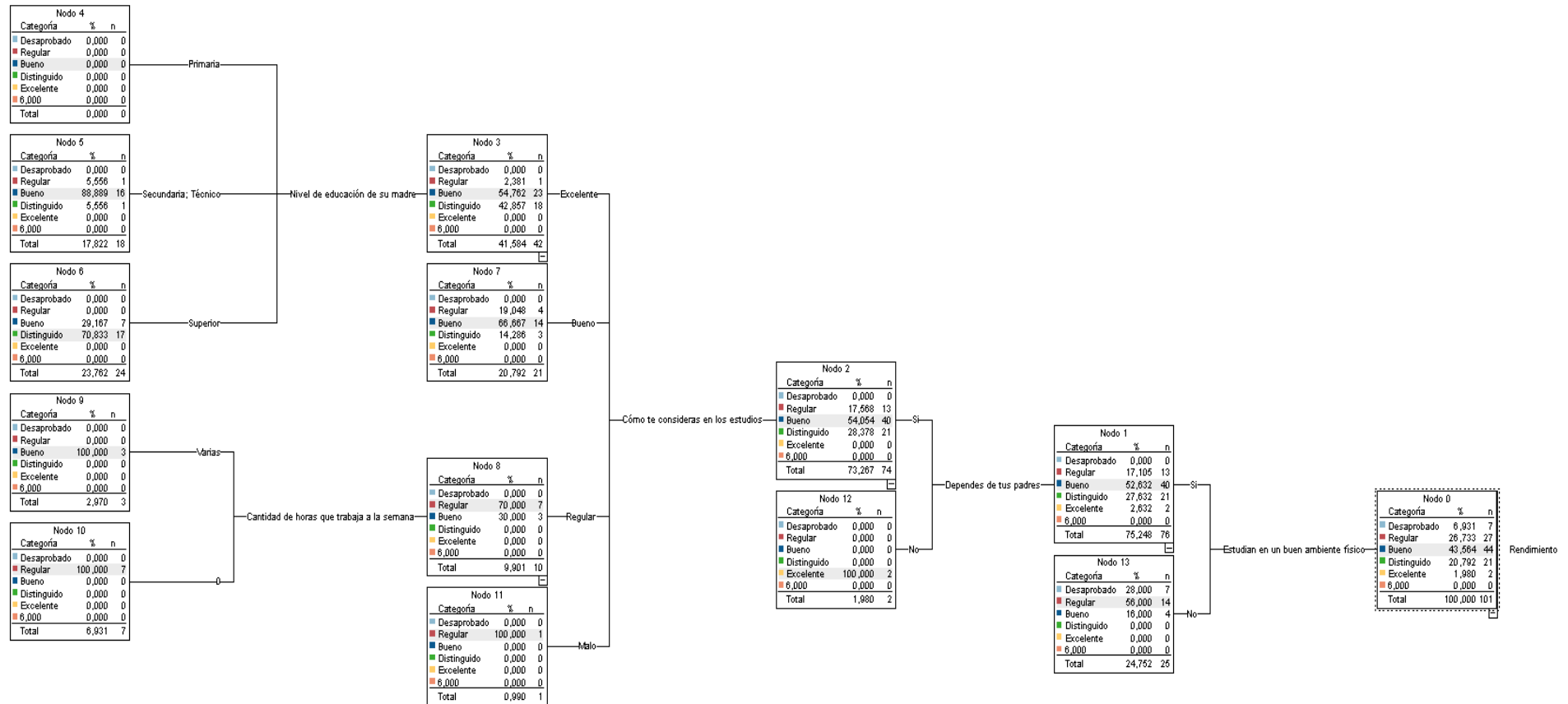
**Figura 4:** Aplicación del nodo tipo - SPSS Modeler Ver.18

Campo	Medida	Valores	No se enc...	Comprobar	Rol
Género	Nominal	"1","2"		Ninguno	Entrada
Edad	Continuo	[14.0,18.0]		Ninguno	Entrada
Grado de est...	Nominal	"1","2"		Ninguno	Entrada
Distrito en elq...	Nominal	"1","2"		Ninguno	Entrada
Trabaja	Nominal	"1","2"		Ninguno	Entrada
Consumes d...	Nominal	"1","2"		Ninguno	Entrada
Consumes li...	Nominal	"1","2"		Ninguno	Entrada
Llega puntu...	Nominal	"1","2","3"		Ninguno	Entrada
Pertenece a	Nominal	"1","2"		Ninguno	Entrada

Fuente: Elaboración Propia



**Figura 5:** Árbol de decisión – SPSS – Modeler Ver. 18



Fuente: Elaboración Propia

## Anexo 8: Autorización para realizar la investigación



UNIVERSIDAD CÉSAR VALLEJO

### Datos Generales

Nombre de la Organización	RUC
IE N°116 “ABRAHAM VALDLEOMAR “	20511525650
Nombre del titular o representante legal	DNI
Rommel Lizandro Crispín	10678927

### Consentimiento:

De conformidad con lo establecido en el artículo 7º, literal “F” del Código de Ética en Investigación de la Universidad César Vallejo (\*), autorizo [ X ], no autorizo [ ] publicar la Identidad de la Organización, en la cual se lleva a cabo la investigación:

Nombre del trabajo de investigación	
Machine Learning para predecir el rendimiento académico de la IE N°116 Abraham Valdelomar San Juan de Lurigancho 2022	
Nombre del Programa Académico	
Escuela Profesional de Ingeniería de Sistemas	
Autor	DNI
Acosta Portocarrero, Antony Christian Ruiz Vargas, Paul	44125358 25599478

En caso de autorizarse, soy consciente que la investigación será alojada en el Repositorio Institucional de la UCV, la misma que será de acceso abierto para los usuarios y podrá ser referenciada en futuras investigaciones, dejando en claro que los derechos de propiedad intelectual corresponden exclusivamente al autor (a) del estudio.

Lima, 12 agosto del 2022

Mg. Rommel Lizandro Crispín  
DIRECTOR

(\*). Código de Ética en Investigación de la Universidad César Vallejo-Artículo 7º, literal “F” Para difundir o publicar los resultados de un trabajo de investigación es necesario mantener bajo anonimato el nombre de la institución donde se llevó a cabo el estudio, salvo el caso en que haya un acuerdo formal con el gerente o director de la organización, para que se difunda la identidad de la institución. Por ello, tanto en los proyectos de investigación como en los informes o tesis, no se deberá incluir la denominación de la organización, pero sí será necesario describir sus características

## **Anexo 9: Desarrollo del sistema de Machine Learning para predecir el rendimiento Académico de los estudiantes de la IE N°116.**

El presente documento describe paso a paso los procesos que se tuvieron en cuenta para el desarrollo del modelo Machine Learning para predecir el rendimiento académico de los estudiantes de la IE N°116- "Abrahán Valdelomar"-San Juan de Lurigancho, haciendo uso de método KDD (Knowledge Discovery in Data bases), la cual está dividida en 5 etapas. A continuación, se dará más detalle de lo que se realizó en cada etapa.

Javier Landa, describe "como un proceso utilizado para llevar a cabo la extracción automatizada de conocimiento partiendo de grandes volúmenes de datos, el cual es de naturaleza iterativa, por lo tanto, es aplicable tantas veces como sea necesario hasta obtener la información necesaria.(Javier Landa, 2019).

Nuestra investigación tiene como finalidad predecir el rendimiento académico de los estudiantes de la IE N°116- "Abrahán Valdelomar", con el objetivo de lograr identificar aquellos estudiantes con probabilidades de bajo rendimiento académico, buscando que el colegio en base a los resultados tome las mejores decisiones para el éxito del alumno.

### **Etapas 1: Etapa de selección de datos**

Para obtener los datos del alumno, se diseñó un cuestionario de 36 preguntas las cuales estuvieron diseñadas con determinantes: personales (9 preguntas), concepto (4 preguntas), motivación (6 preguntas), culturales (6 preguntas), educación de los padres (2 preguntas), emocional (4 preguntas), económicos (3 preguntas), escuela origen (1 preguntas) y variable objetiva (1 pregunta). Estas preguntas se registraron en un formulario de Google y enviadas a estudiantes de la IE N°116-" Abraham Valdelomar" obteniendo un total de 101 registros.

**Figura 1:** Cuestionario utilizando Google Forms

**Factores que influyen en el rendimiento escolar**

---

*Buenas buenas días/buenas tardes Estimados alumnos, somos bachilleres en Ingeniería de Sistemas que estamos realizando una investigación para nuestra Tesis de titulación acerca del "Rendimiento Académico de los estudiantes del 4to. y 5to. grado de secundaria" de la institución a las cuales ustedes pertenecen. De tal manera que los resultados que se obtengan de esta investigación servirán a sus autoridades y profesores de la institución para que tomen medidas correctivas para el éxito de los alumnos. Para lo cual solicitamos su colaboración para responder honestamente las preguntas de este formulario. Agradecemos su valiosa colaboración para el éxito de nuestra investigación.*

---

Fuente: Elaboración propia

## **Etapa 2: Etapa de preprocesamiento / limpieza de datos**

Después de terminada la recolección de los datos a través de formulario de Google, se obtuvo 101 registros.

A la tabla que se generó en Excel, por medio del formulario se procedió a una limpieza de aquellos datos que fueron mal ingresados. Nos referimos a la información que digito el estudiante. Por ejemplo: su distrito, la cantidad de horas que trabaja, la cantidad de horas que dedica al estudio, el promedio ponderado, entre otros.

En segundo lugar, de los datos exportados de formulario google a Excel, equivalentes a un total de 101 registros estudiantiles. Se hizo un análisis de los datos, corroborando que no existiera datos nulos, también se homogenizaron las respuestas con relación a las siguientes preguntas:

- ¿Cuántas horas al día son dedicadas al estudio?
- ¿Cuántos de sus familiares reciben pensión del Estado?
- ¿Cuántos integrantes hay en su grupo familiar?
- ¿Cuántos de sus familiares reciben pensión del Estado?
- ¿Cantidad de horas que trabaja a la semana?
- ¿Ingreso familiar aproximado mensualmente?

A estas columnas se les había ingresado la palabra horas junto con la cantidad numérica. Sólo se dejó la cantidad numérica.

**Figura 2: Datos importados de Google Forms a Excel**

	AD	AE	AF	AG	AH	AI	AJ	AK	AL
1	Tiene empatía hacia sus	Cuando obtienes malas	Que cantidad de amigos	Cantidad de horas que tr	Ingreso familiar aproxima	Sus estudios iniciales en	Establecimiento en el qu	Promedio ponderado durante el año	
7	Empático	Frustrado	Regular	Ninguna	4000soles	Público	Público	16	
8	Muy empático	Frustrado	Regular	ninguna	4500 soles	Privado	Público	17	
9	Empático	Todas	Regular	No trabajo	5000 soles	Privado	Público	16	
10	Muy empático	Triste	Muchos	NO TRABAJO	3000 SOLES	Público	Público	15	
11	Empático	Frustrado	Pocos	9horas	2500soles	Público	Público	15	
12	Empático	Triste	Regular	No trabajo	4000 soles	Privado	Público	18	
13	Empático	Triste	Muchos	60	3500	Público	Público	20	
14	Muy empático	Todas	Muchos	12 horas	2200 soles	Público	Público	13	
15	Empático	Todas	Regular	0	5000	Público	Público	17	
16	Muy empático	Triste	Regular	NO TRABAJO	3500 SOLES	Público	Público	17	
17	Muy empático	Frustrado	Regular	60	3500	Público	Público	20	
18	Empático	Molesto	Pocos	NO TRABAJO	5000 SOLES	Público	Público	16	
19	Empático	Triste	Muchos	8	3000	Público	Público	15	
20	Empático	Frustrado	Regular	8	3000	Privado	Público	18	
21	Empático	Triste	Regular	NO TRABAJO	4500 SOLES	Privado	Público	16	
22	Empático	Todas	Regular	5	1300	Público	Público	14	
23	Empático	Triste	Regular	NO TRABAJO	6000 SOLES	Privado	Público	17	
24	Muy empático	Triste	Regular	No trabajo	4000	Público	Público	18	

Fuente: Elaboración propia

Seguidamente, se realizó la definición de las variables, las cuales están asociadas a las preguntas del cuestionario, se definió que los valores sean de tipo numérico con un ancho de 8 dígitos, en la columna “Decimales” se utilizó para la variable Ingreso familiar aproximado mensualmente 1 decimal y para la variable Promedio ponderado durante el año 2 decimales. Luego en la columna “Valores”, se añadió las alternativas por cada pregunta y su equivalencia en números, también en la columna “Medida” se seleccionó entre ordinal, escalar y nominal según corresponde y finalmente en la columna “Rol” se seleccionó el tipo de dato en caso se eligió para todas las variables de tipo Entrada.

Tabla que muestra los valores numéricos que se asignaron cada variable del formulario de Google.

Genero	
Femenino	2
Masculino	1
Actitud_responsable	
Muy Responsable	1
Responsable	2
Irresponsable	3
Muy irresponsable	4
Aptitud_culminar_ciclo_exitoso	
Muy en desacuerdo	1
En desacuerdo	2
No sabe	3
De acuerdo	4
Muy de acuerdo	5
Persona_inteligente	
Muy en desacuerdo	1
En desacuerdo	2
No sabe	3
De acuerdo	4
Muy de acuerdo	5
Gusto_carrera_profesional_elegida	
Si	1
No	2
Estudiar_noEstudiar	
Si	1
No	2
Dependiente_Independiente_padre	
Independiente	1
Dependiente	2
Dependiente_Independiente_madre	
Independiente	1
Dependiente	2
Nivel_educación_madre	
Primaria	1
Secundaria	2
Técnico	3
Superior	4
Nivel_educación_padre	
Primaria	1

Secundaria	2
Técnico	3
Superior	4
Reacción_problema_entorno_social	
Reacciona con violencia física o verbal	1
Intenta conversar sobre la situación	2
Con indiferencia	3
Influencia_stress_vida_academica	
Muy negativamente	1
Negativamente	2
No influye	3
Positivamente	4
Muy positivamente	5
Nivel_empatia_compañeros	
Sin empatía	1
Muy poco empático	2
Indiferente	3
Poco empático	4
Muy empático	5
Tipo_establecimiento_estudio_primario	
Público	1
Privado	2
Tipo_establecimiento_estudio_secundario	
Público	1
Privado	2

Luego de realizar la limpieza de datos y generar la tabla con los valores que serán asignados a las variables se procedió a la importación de los datos de la tabla Excel al SPSS Statistic Ver.26, aquí se agregó la variable de “Rendimiento”. A la que se asignó valores numéricos para cada opción 1= “Reprobado”, 2= “Regular”, 3= “bueno”, 4= “Distinguido” y 5= “Excelente”

**Figura 3: Definición de variables – SPSS Statistics**

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Marcatemporal	Fecha	40	0	Marca temporal	Ninguna	Ninguna	11	Derecha	Escala	Entrada
2	Género	Numérico	8	0	Genero	{1, Masculin...	Ninguna	9	Derecha	Nominal	Entrada
3	Edad	Numérico	8	0	Edad	Ninguna	Ninguna	6	Derecha	Escala	Entrada
4	Grado de estudios	Numérico	8	0	Grado de estud...	{1, 5to_año}...	Ninguna	9	Derecha	Nominal	Entrada
5	Distrito en el que vive	Numérico	8	0	Distrito en el qu...	{1, San juan...	Ninguna	14	Derecha	Nominal	Entrada
6	Trabajas	Numérico	8	0	Trabajas	{1, Si}...	Ninguna	5	Derecha	Nominal	Entrada
7	Consumes drogas	Numérico	8	0	Consumes drog...	{1, Si}...	Ninguna	6	Derecha	Nominal	Entrada
8	Consumes licor	Numérico	8	0	Consumes licor	{1, Si}...	Ninguna	3	Derecha	Nominal	Entrada
9	Llegas puntual a las clases	Numérico	8	0	Llegas puntual ...	{1, Siempre}...	Ninguna	12	Derecha	Nominal	Entrada
10	Perteneces a alguna pandilla	Numérico	8	0	Perteneces a al...	{1, Si}...	Ninguna	3	Derecha	Nominal	Entrada
11	¿Qué tan responsable se considera	Numérico	8	0	¿Qué tan respon...	{1, Muy res...	Ninguna	17	Derecha	Nominal	Entrada
12	¿Considera posible terminar la educación	Numérico	8	0	Considera posi...	{1, Si}...	Ninguna	7	Derecha	Nominal	Entrada
13	¿Cómo se considera en los estudios	Numérico	8	0	Cómo te consid...	{1, Excelent...	Ninguna	9	Derecha	Nominal	Entrada
14	¿Cómo se considera para estos cursos Matemáticas Físicas	Numérico	8	0	Cómo te consid...	{1, Muy Bue...	Ninguna	9	Derecha	Nominal	Entrada
15	¿Te gusta estudiar	Numérico	8	0	Te gusta estudiar	{1, Si}...	Ninguna	7	Derecha	Nominal	Entrada
16	¿Estudias en un buen ambiente físico	Numérico	8	0	Estudian en un ...	{1, Si}...	Ninguna	7	Derecha	Nominal	Entrada
17	¿Te gusta estudiar en grupo	Numérico	8	0	Te gusta estudi...	{1, Si}...	Ninguna	7	Derecha	Nominal	Entrada
18	¿Cuántas horas al día se dedica al estudio	Numérico	8	0	Cuántas horas ...	Ninguna	Ninguna	12	Derecha	Nominal	Entrada
19	¿Si te dieran a elegir entre estudiar o no estudiar	Numérico	8	0	Si te dieran a el...	{1, Si}...	Ninguna	14	Derecha	Nominal	Entrada
20	¿Cuál de las siguientes redes sociales usas para buscar información	Numérico	8	0	Cuál de las sig...	Ninguna	Ninguna	13	Derecha	Nominal	Entrada
21	¿Cuántos integrantes hay en su grupo familiar	Numérico	8	0	Cuántos integra...	Ninguna	Ninguna	12	Derecha	Nominal	Entrada
22	¿Cuánto trabaja en su familia	Numérico	8	0	Cuántos trabaja...	Ninguna	Ninguna	12	Derecha	Nominal	Entrada
23	¿Cuántos de su familia estudian	Numérico	8	0	Cuántos de su f...	Ninguna	Ninguna	12	Derecha	Nominal	Entrada
24	¿Cuántos de sus familiares reciben pensión del Estado	Numérico	8	0	Cuántos de sus...	Ninguna	Ninguna	12	Derecha	Nominal	Entrada

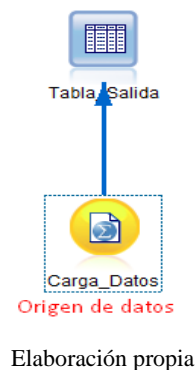
Fuente: Elaboración propia

### Etapa 3: Etapa de transformación y reducción

Para esta etapa, se utilizó la herramienta SPSS Modeler Ver.18, para crear el modelo predictivo Machine Learning, y también se creyó conveniente deshabilitar algunas variables no influyentes como por ejemplo “el distrito en el que vive” y “establecimiento en el que estudias”.

Se realizó la carga de la base de datos, la cual fue creada a partir de la herramienta SPSS Statistic Ver.26, para esta carga se utilizó un “nodo origen” de tipo “Archivo de Statistic”, para validar la importación de datos se utilizó un “nodo resultado” de tipo Tabla, el cual permite visualizar los datos ya convertidos en datos numéricos. Como se muestra en la figura 4 y figura 5 nos muestra la salida de los datos en tipo tablas.

**Figura 4:** Carga de datos con nodo origen - SPSS Modeler.



**Figura 5:** Salida de datos con nodo resultado, tipo tabla - SPSS

La imagen muestra una ventana de SPSS con el título 'Tabla\_Salida (35 campos, 101 registros) #1'. La interfaz incluye una barra de menú con 'Archivo', 'Editar' y 'Generar', y una barra de herramientas. La tabla principal tiene las siguientes columnas: Género, Edad, Grado de estudios, Distrito en el que vive, Trabaja, Consumes drogas, Consumes licor, Llegas puntual a las clases, Pertenece a algún pandilla y Qué tan responsable es. Se muestran los primeros 20 registros de la tabla.

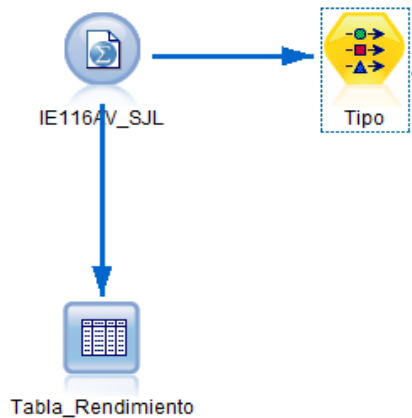
	Género	Edad	Grado de estudios	Distrito en el que vive	Trabaja	Consumes drogas	Consumes licor	Llegas puntual a las clases	Pertenece a algún pandilla	Qué tan responsable es
1	2.000	16.0...	1.000	1.000	2.000	2.000	2.000	3.000	2.000	
2	1.000	16.0...	1.000	1.000	2.000	2.000	2.000	1.000	2.000	
3	1.000	16.0...	1.000	1.000	2.000	2.000	2.000	1.000	1.000	
4	2.000	17.0...	1.000	1.000	2.000	2.000	2.000	1.000	2.000	
5	2.000	15.0...	1.000	2.000	2.000	2.000	2.000	3.000	2.000	
6	2.000	16.0...	1.000	1.000	2.000	2.000	2.000	1.000	2.000	
7	2.000	15.0...	2.000	1.000	1.000	2.000	2.000	3.000	2.000	
8	2.000	15.0...	2.000	1.000	2.000	2.000	2.000	1.000	2.000	
9	1.000	16.0...	1.000	1.000	2.000	2.000	2.000	1.000	2.000	
10	1.000	16.0...	1.000	1.000	2.000	2.000	2.000	3.000	2.000	
11	2.000	14.0...	2.000	1.000	2.000	2.000	2.000	3.000	2.000	
12	2.000	16.0...	1.000	1.000	2.000	2.000	2.000	1.000	2.000	
13	1.000	15.0...	2.000	1.000	1.000	2.000	2.000	3.000	2.000	
14	1.000	16.0...	1.000	1.000	1.000	2.000	2.000	1.000	2.000	
15	2.000	15.0...	2.000	1.000	1.000	2.000	2.000	3.000	2.000	
16	1.000	17.0...	1.000	1.000	1.000	1.000	2.000	1.000	1.000	
17	1.000	15.0...	2.000	1.000	2.000	2.000	2.000	3.000	2.000	
18	1.000	16.0...	1.000	1.000	2.000	2.000	2.000	3.000	2.000	
19	1.000	15.0...	2.000	1.000	2.000	2.000	2.000	1.000	2.000	
20	1.000	15.0...	2.000	1.000	2.000	2.000	2.000	3.000	2.000	

Fuente: Elaboración propia



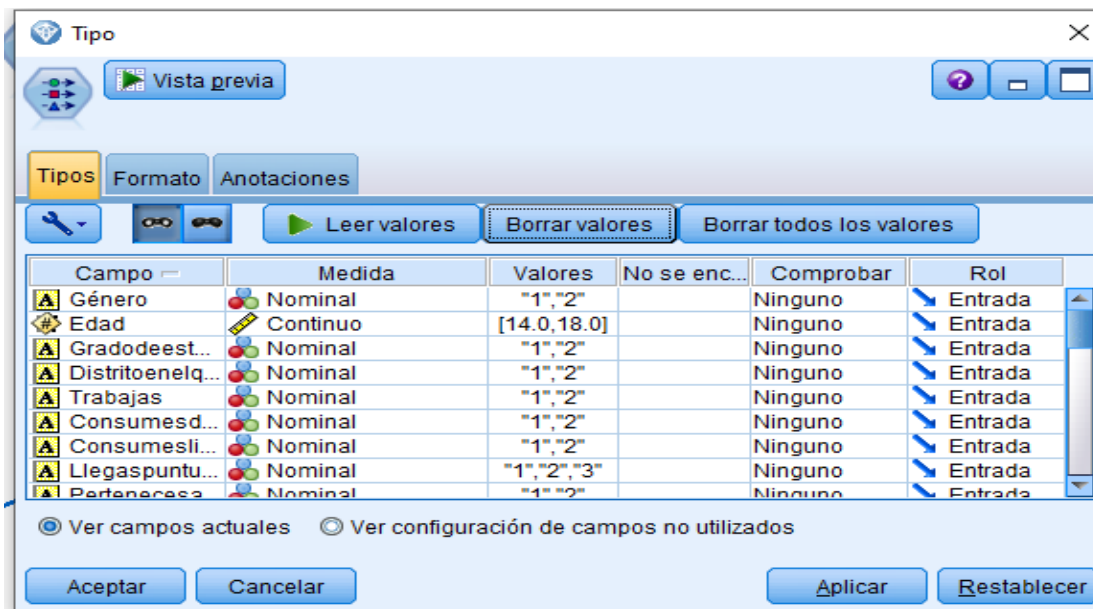
Con el objetivo de definir nuestras variables de entrada y variable de destino, se utilizó un “nodo tipo” y seguidamente se cargaron los valores numéricos para cada variable como lo muestra la figura 6 y en la figura 7 es el resultado del nodo tipo.

**Figura 6:** Aplicación del nodo tipo - SPSS



Fuente: Elaboración Fuente propia

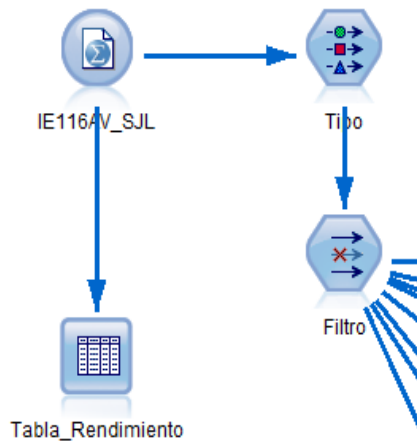
**Figura 7:** Aplicación del nodo tipo - SPSS Modeler Ver.18



Elaboración: Fuente Propia

Para deshabilitar las variables que no influyentes, se usó el “nodo filtro “según como se muestra en la Figura 8.

**Figura 8:** Aplicación del nodo filtro - SPSS Modeler Ver. 18

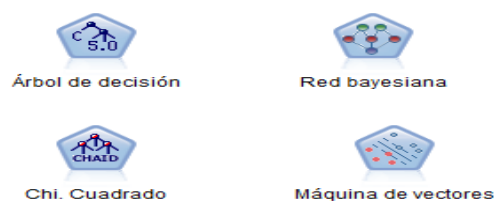


Fuente: Elaboración propia

#### **Etapa 4: Minería de datos**

En esta etapa, se eligió distintos algoritmos de aprendizaje automático con la finalidad de comparar resultados y obtener el mejor modelo predictivo. En esta investigación se usaron los modelos como: Árbol de Decisión, Máquina de Vectores de Soporte-SVM y Red Bayesianano, estos modelos son proporcionados por la herramienta SPSS Modeler Ver.18, según como se muestra en la Figura 9.

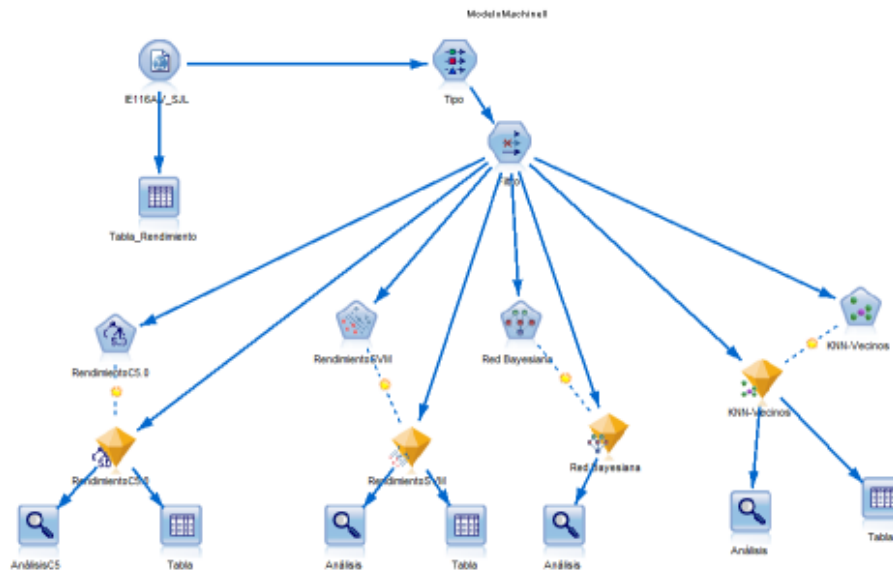
**Figura 9:** Algoritmos de aprendizaje automático - SPSS Modeler Ver.18



Fuente: Elaboración propia

A continuación, se muestra el modelo generado para cada algoritmo de aprendizaje como se detalla en la figura 10.

**Figura 10:** Proyecto de rendimiento académico - SPSS – Modeler Ver. 18

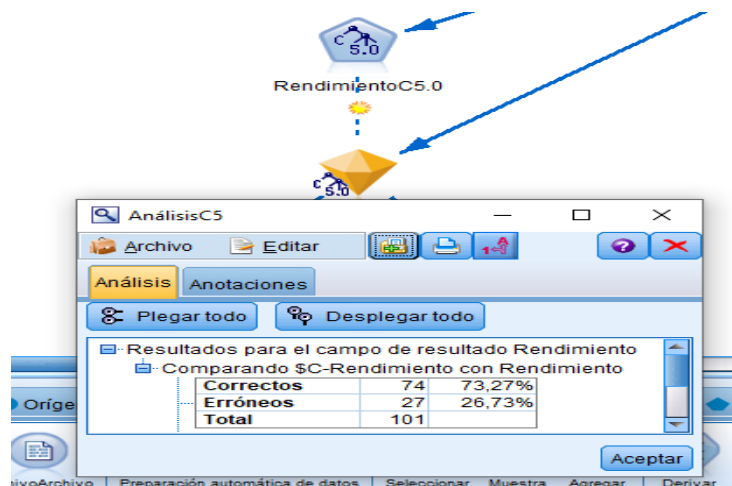


Fuente: Elaboración propia

### **Etapas 5: Interpretación**

**Árbol de decisión – C5**, es una herramienta que nos permite obtener información de las variables con mayor relevancia en relación a la variable rendimiento. Sin embargo el modelo elegido alcanzó una precisión de 73.27% con una tasa de error de 26.73% como se detalla en la figura 11.

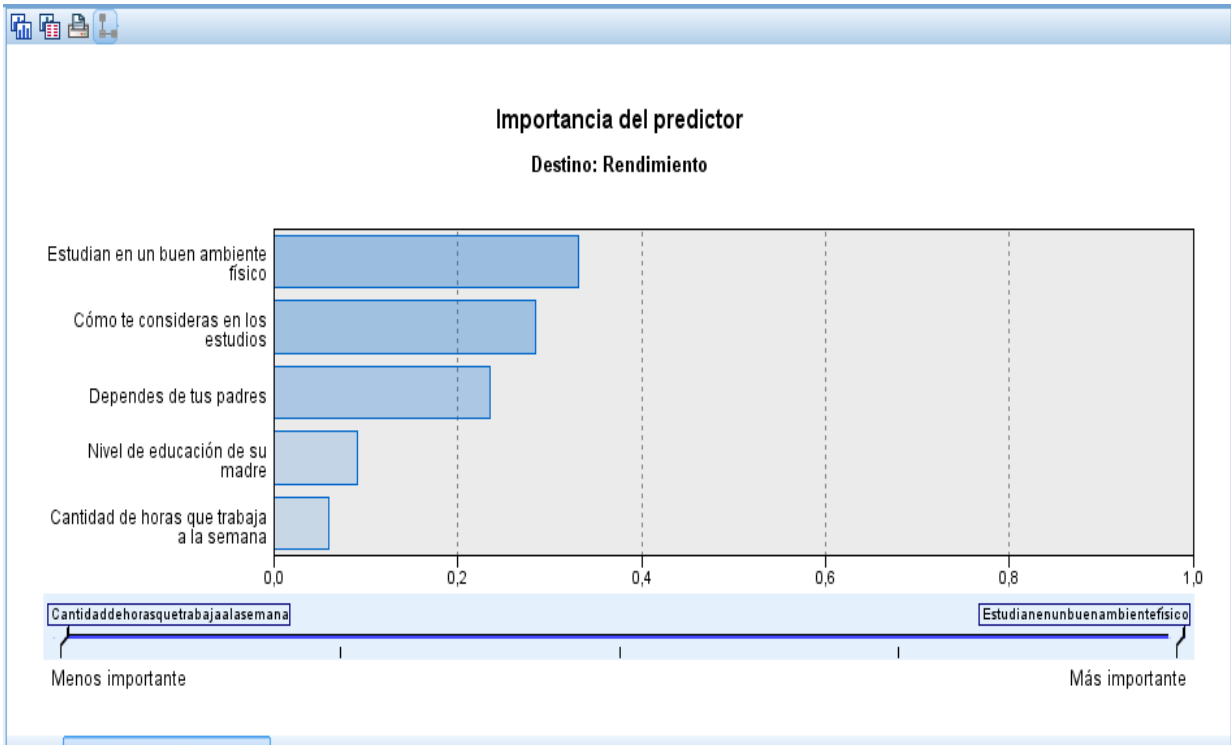
**Figura 11:** Precisión de modelo utilizando árbol de decisión – SPSS – Modeler Ver. 18



Modeler Fuente: Elaboración propia

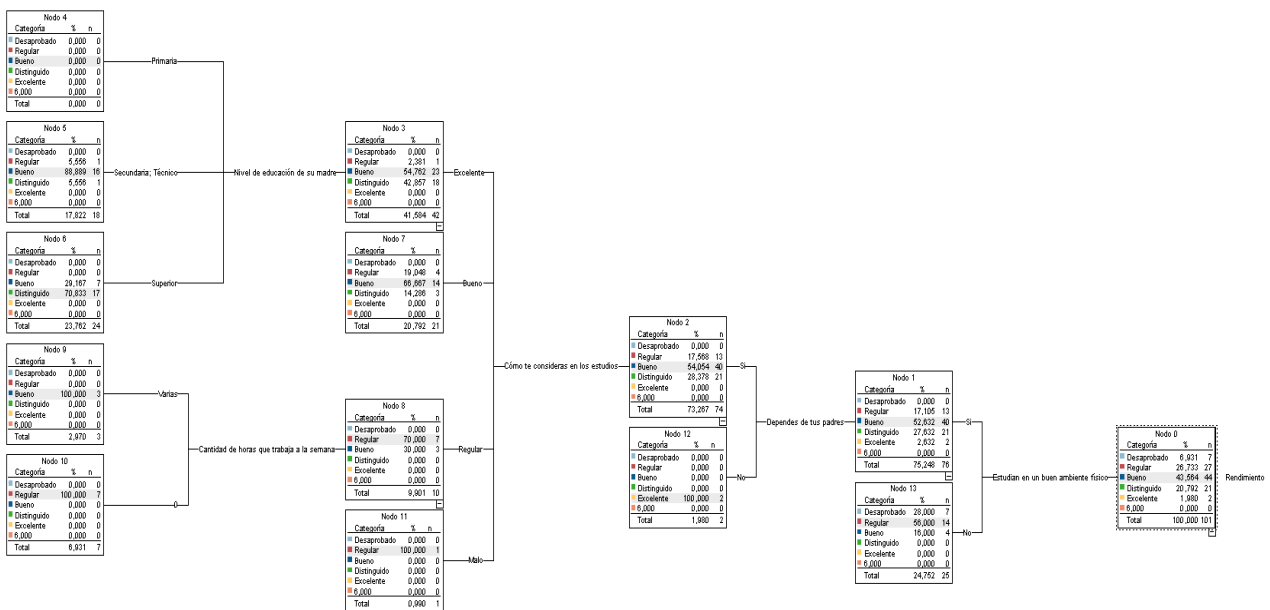
En la Figura 12 se muestra los resultados gráficos que muestra el Algoritmo Árbol de Decisión.

**Figura 12:** Variables con relevancia utilizando árbol de decisión SPSS – Modeler Ver. 18



Modeler Fuente: Elaboración propia

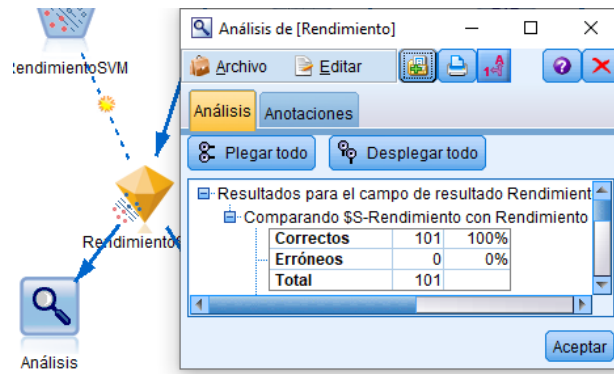
**Figura 13:** Árbol de decisión – SPSS – Modeler Ver. 18



Elaboración propia

**Máquina de Vectores (SVM):** Mediante esta herramienta Podemos obtener valores con una precisión del 100% una tasa de error de 0% según como se muestra en la Figura 14.

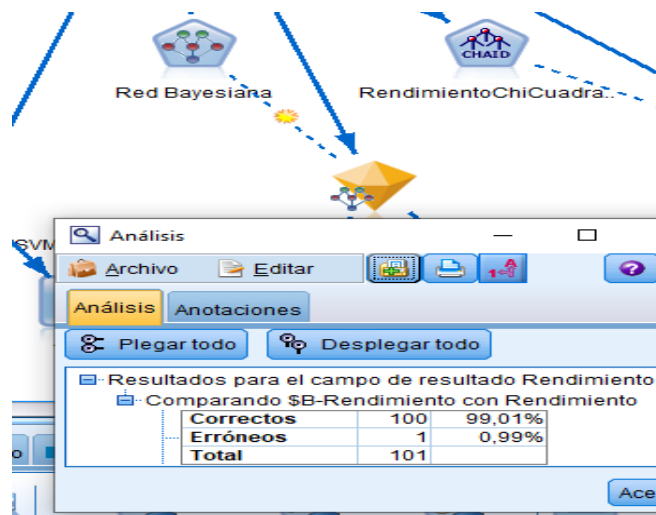
**Figura 14:** Precisión de modelo utilizando SVM - SPSS Modeler Ver. 18



Fuente: Elaboración propia

**Red Bayesiana:** Mediante esta herramienta Podemos obtener valores con una precisión del 99.01% una tasa de error de 0.99%, como se detalla en la figura 15.

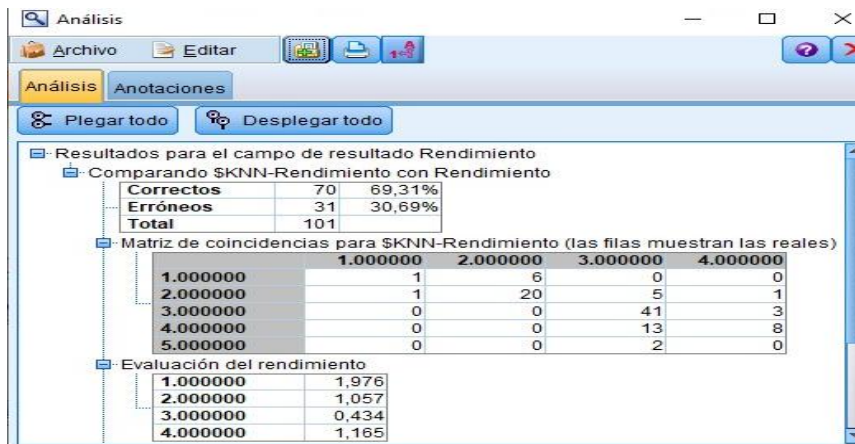
**Figura 15:** Precisión de modelo utilizando Red Bayesiana - SPSS Modeler ver. 18



Fuente: Elaboración Propia

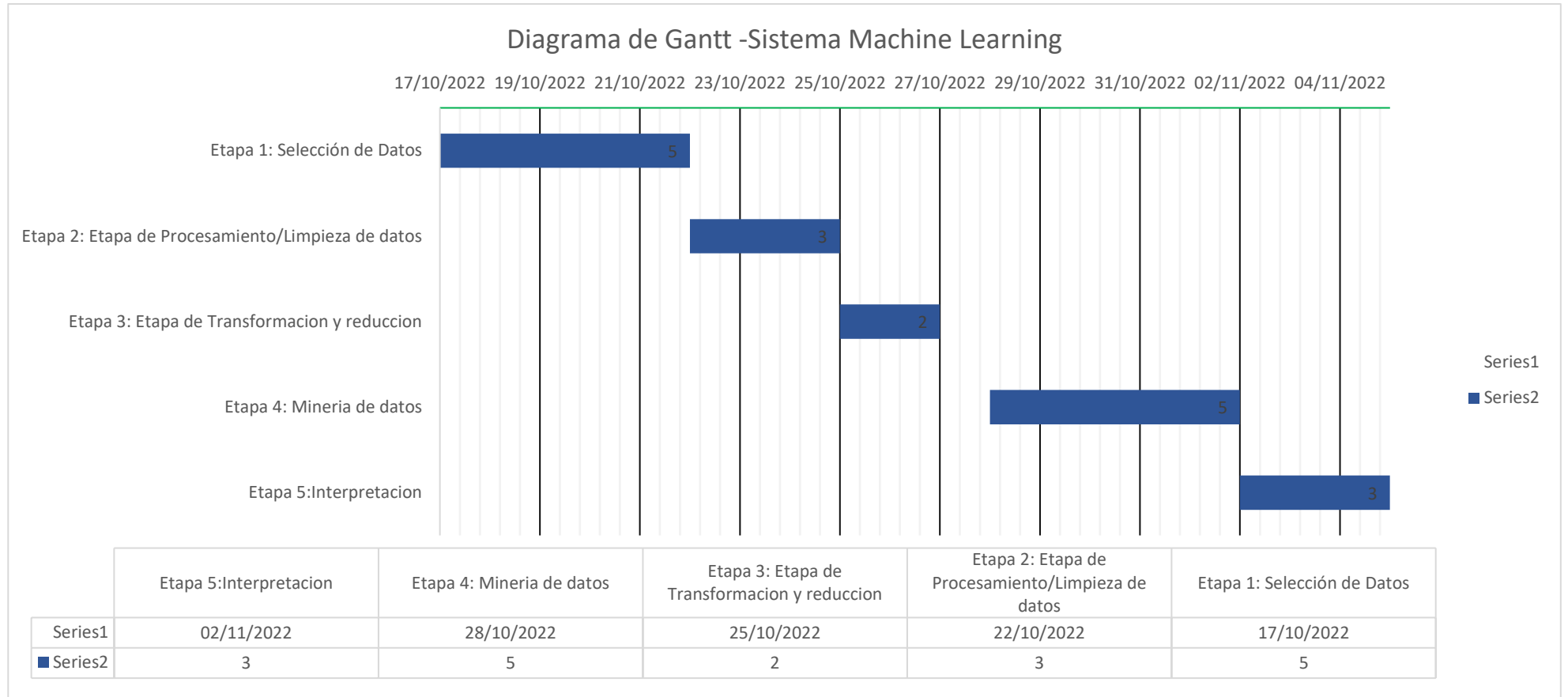
**KNN-Vecino:** Mediante esta herramienta Podemos obtener valores con una precisión del 69.31% una tasa de error de 30.69%, como se detalla en la figura 16.

**Figura 16:** Precisión de modelo utilizando KNN-Vecino - SPSS Modeler ver. 18



Fuente. Elaboración Propia

**Figura 17:** Diagrama de Gantt: Sistema Machine Learning Para medir el Rendimiento académico.



**Prototipo del sistema**

**Figura 18:** Diagrama De Caso De Uso

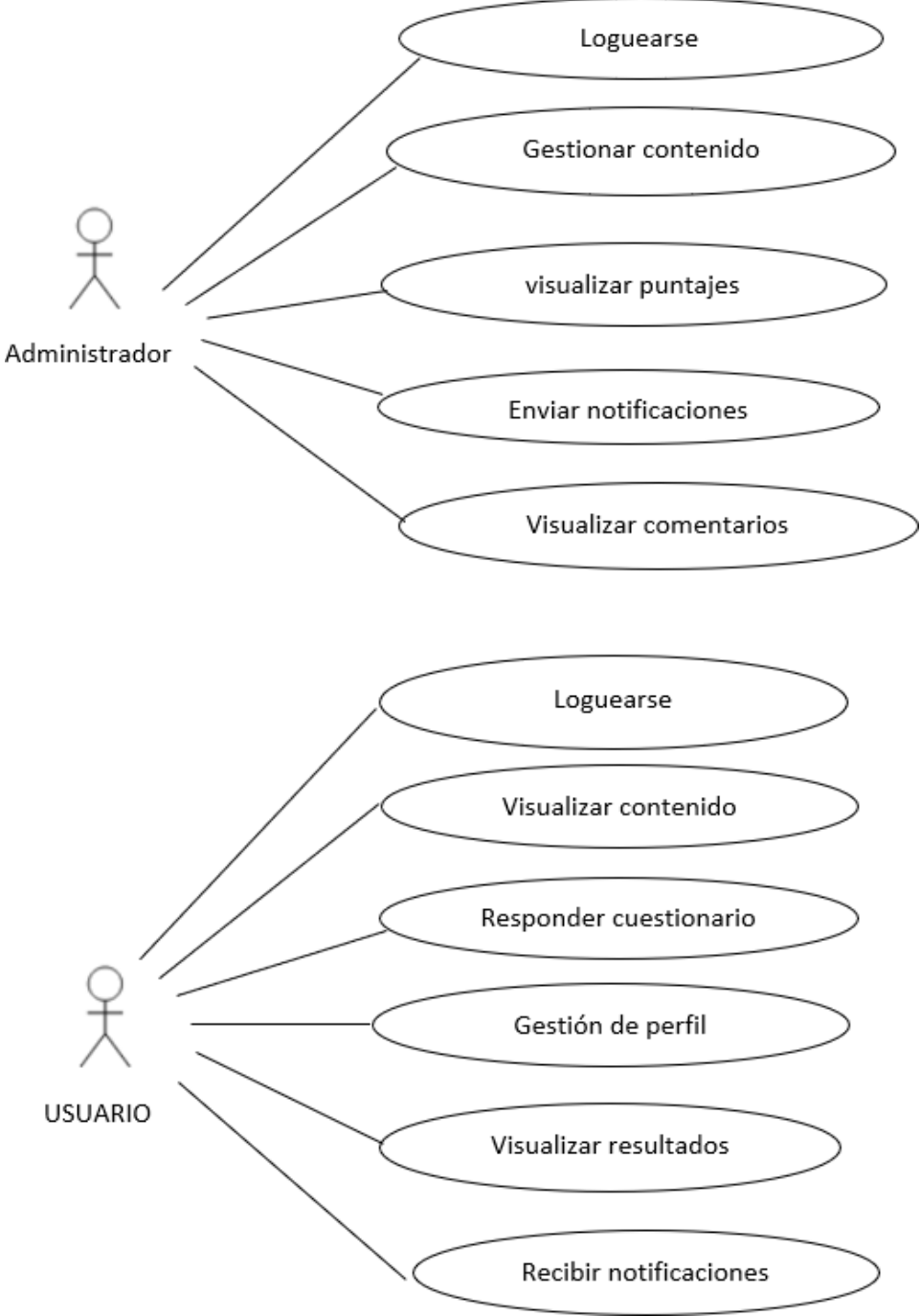


Figura 18. Diagrama de casos de uso



Figura 19: Diagrama De Clase

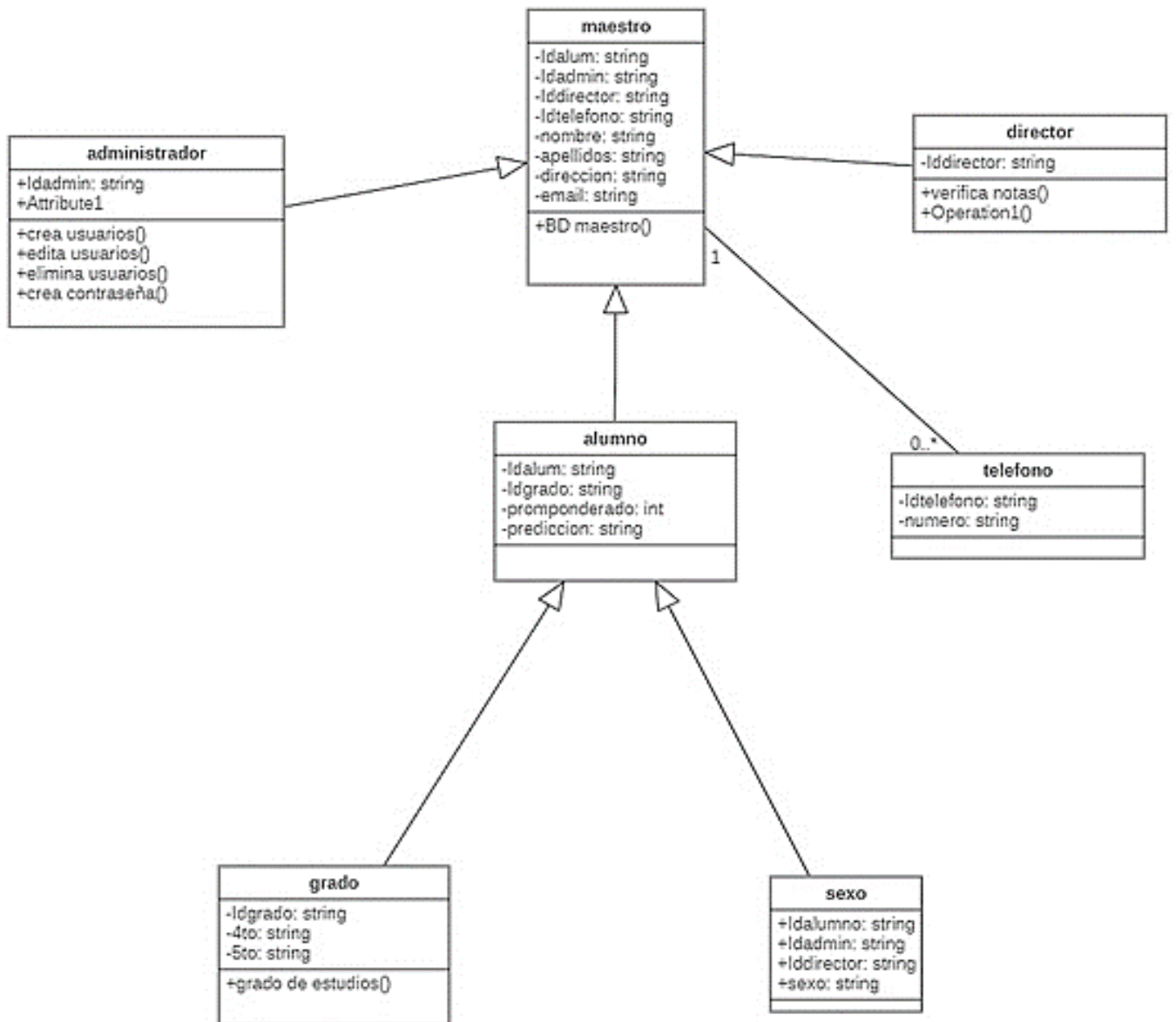


Figura 19. Diagrama de clase

Para el Desarrollo de la Aplicación del Modelo de SVM, Machine Learning, se utilizó el Framework Laravel.

Laravel es un framework de código abierto para desarrollar aplicaciones y servicios web con PHP 5, PHP 7 y PHP 8, su filosofía es desarrollar código PHP de forma elegante y simple, evitando el "código espagueti" (código parece asemejarse a un plato de espaguetis, es decir, un montón de hilos intrincados y anudados). Fue creado en 2011 y tiene una gran influencia de framework como Ruby on Rails, Sinatra y ASP.NET MVC.

**El Framework Laragon:** Crea entornos de desarrollo para PHP y otros stack de tecnologías, de una manera cómoda, rápida y con prestaciones altamente profesionales., como se muestra en la Figura 19.

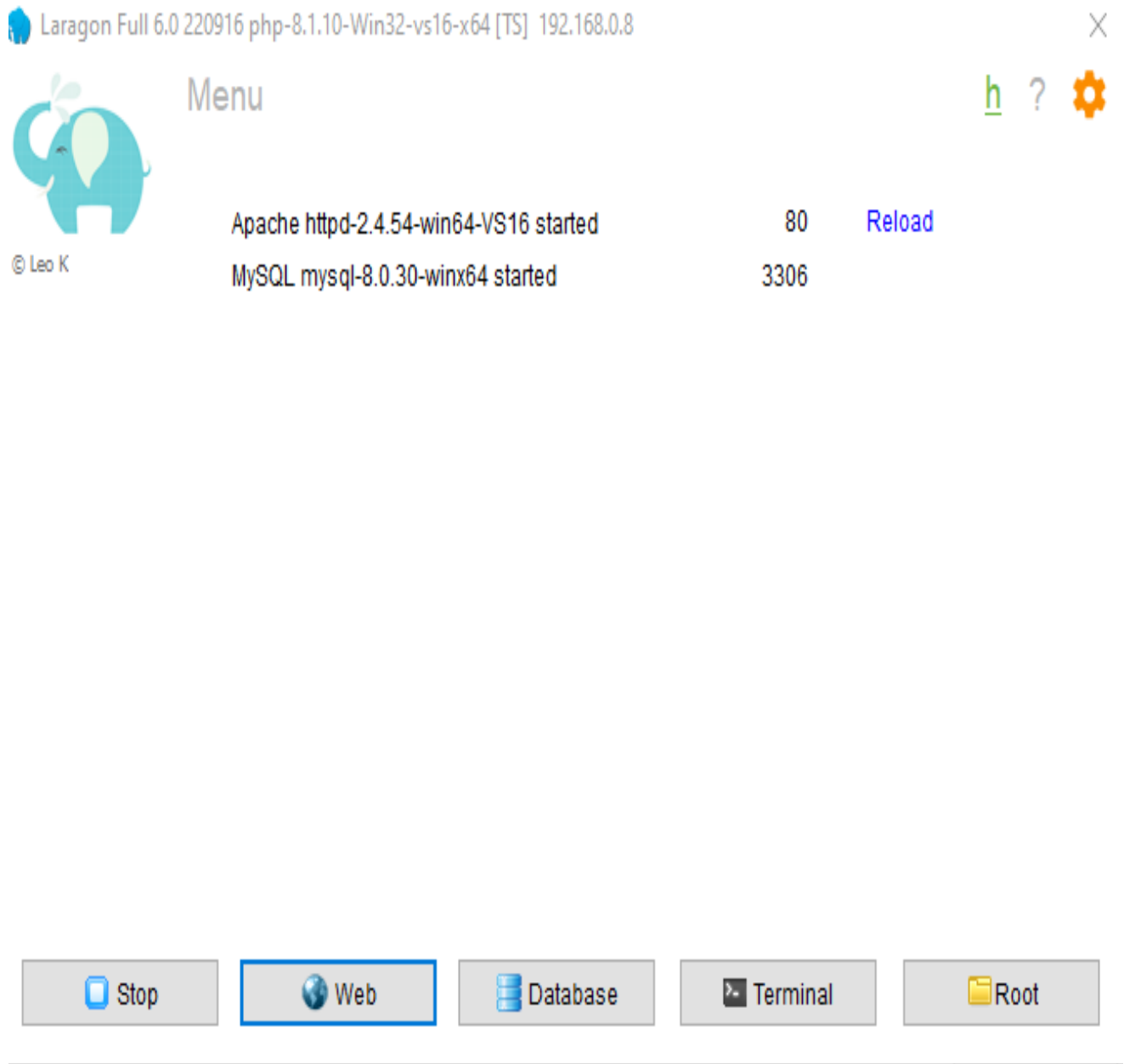
**Figura 19:** Aplicacion al Framework Laragon



```
app > Http > Controllers > AlumnoController.php > AlumnoController > store
1 <?php
2
3 namespace App\Http\Controllers;
4 use Illuminate\Http\Request;
5 use App\Models\Alumno;
6
7 class AlumnoController extends Controller
8 {
9     protected $alumnos;
10
11     public function __construct(Alumno $alumnos)
12     {
13         $this->alumnos = $alumnos;
14     }
15
16     public function index()
17     {
18         $alumnos = $this->alumnos->obtenerAlumnos();
19         return view('alumnos.lista', ['alumnos' => $alumnos]);
20     }
21 }
```

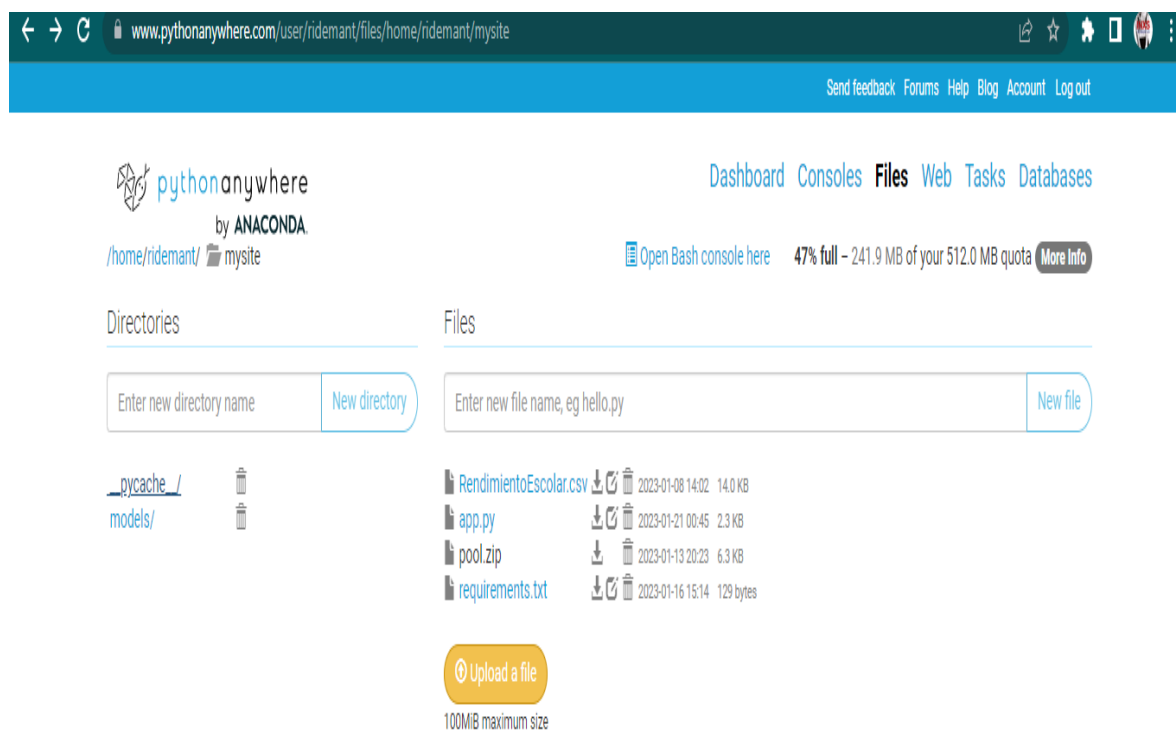
**Laragon** es una opción relativamente Nueva para la creación de lo que llamamos el entorno de desarrollo, es decir, todo un conjunto de programas necesarios para desarrollar aplicaciones. Sirve para trabajar con PHP, pero también con otros lenguajes del lado del servidor, como Node, Python o Ruby, como se muestra en la figura 20.

**Figura 20:** Aplicacion Laragon



Para subir el Código de Python con el cual se entrena al modelo SVM de Machine Learning se eligió un servidor de Python, como se muestra en la figura 21.

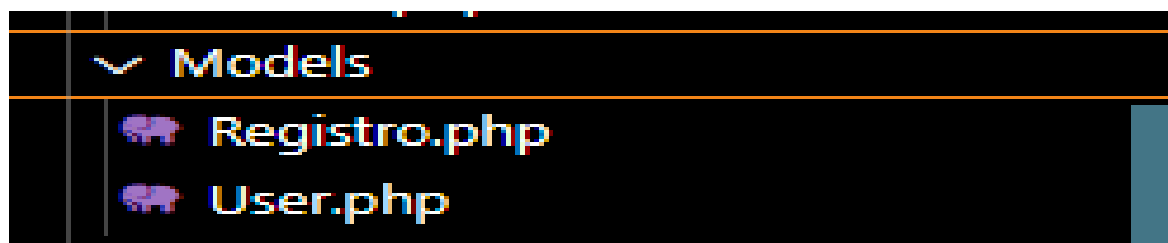
**Figura 21:** Conexión al Servidor Python.



Luego para realizar el desarrollo del código fuente usamos el ide Visual studio Código donde se usó la tecnología MVC para crear el Modelo, las Vistas y Controles para el uso del Sistema de Predicción del Rendimiento Académico.

**Modelo:** Contiene toda la lógica de la aplicación que no está contenida en una vista o un controlador. El modelo debe contener toda la lógica de negocios de la aplicación, la lógica de validación y la lógica de acceso a la base de datos, como se muestra en la Figura 22.

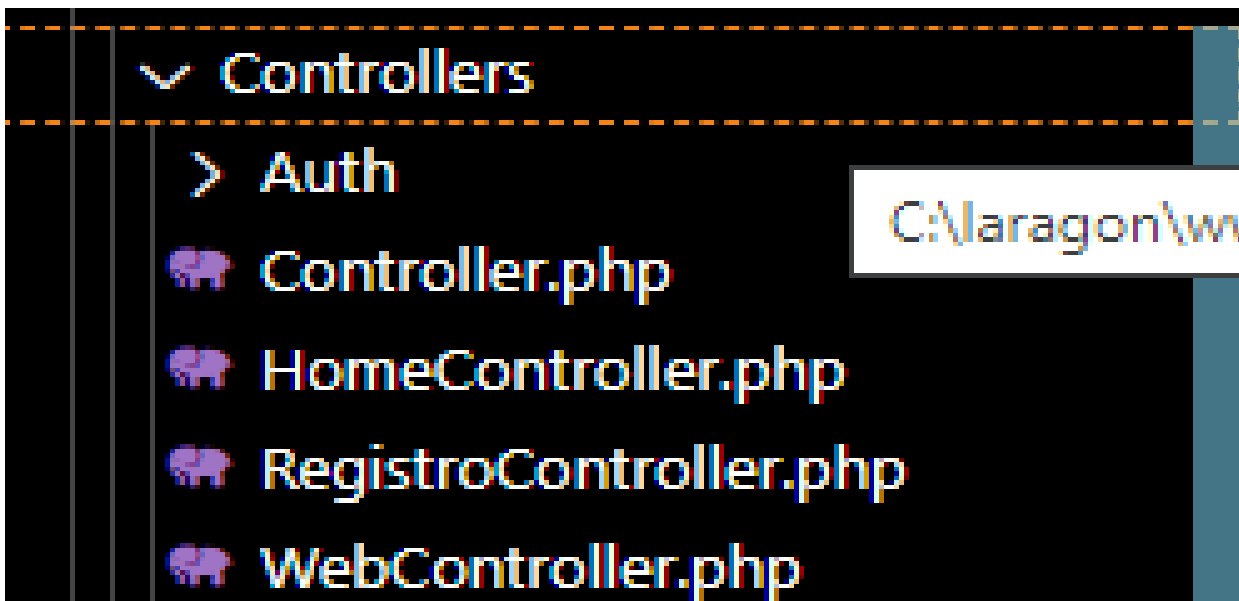
**Figura 22:** Configuración Y codificación del modelo

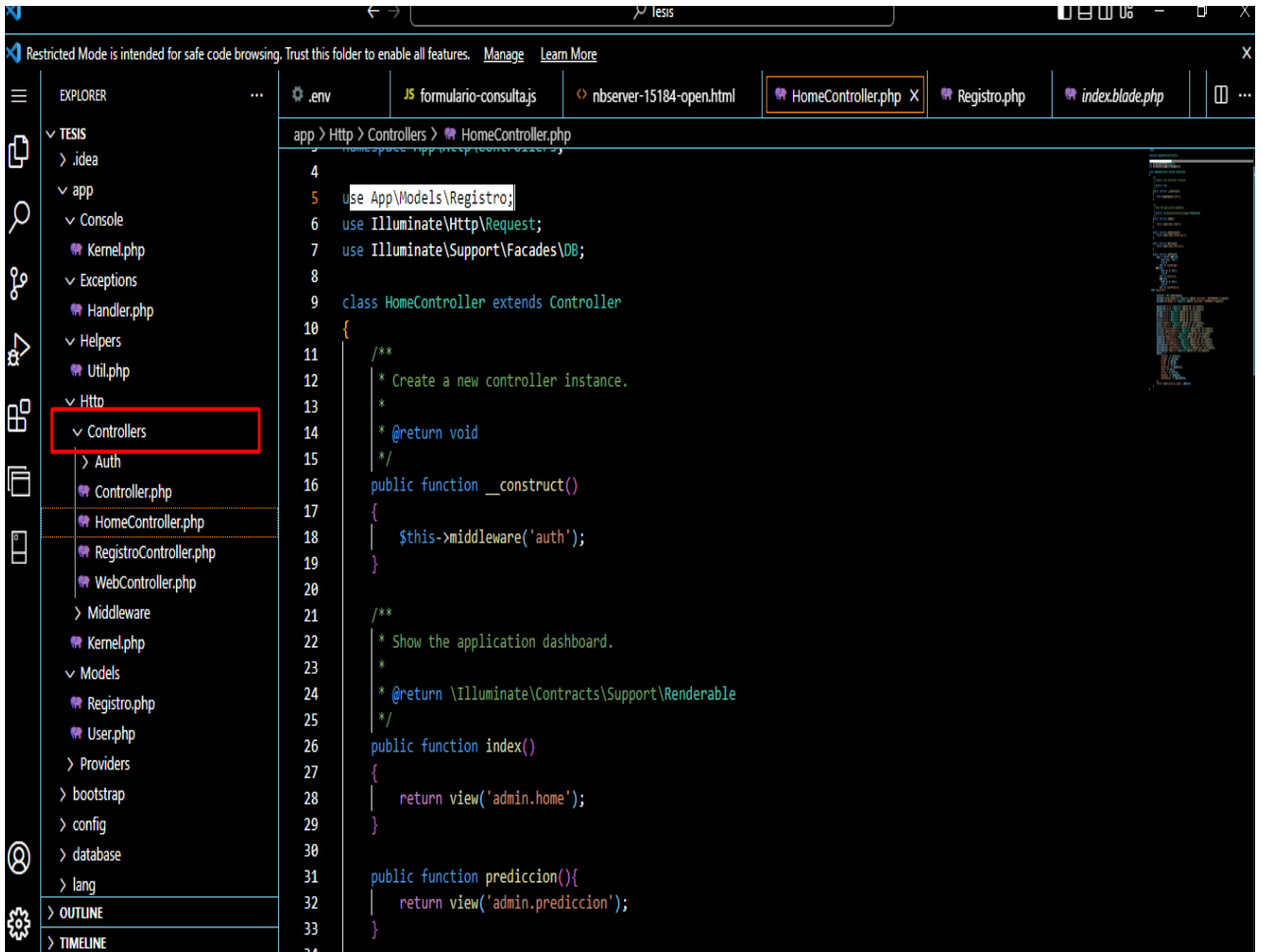


```
Restricted Mode is intended for safe code browsing. Trust this folder to enable all features. Manage Learn More
JS formulario-consulta.js nserver-15184-open.html HomeController.php Registro.php User.php X
EXPLORER
TESSIS
> .idea
> app
  Console
  Kernel.php
  Exceptions
  Handler.php
  Helpers
  Util.php
  Http
  Controllers
    Auth
      Controller.php
      HomeController.php
      RegistroController.php
      WebController.php
    Middleware
      Kernel.php
  Models
    Registro.php
    User.php
  Providers
  bootstrap
  config
  database
  lang
  > OUTLINE
  > TIMELINE
app > Models > User.php
6 use Illuminate\Database\Eloquent\Factories\HasFactory;
7 use Illuminate\Foundation\Auth\User as Authenticatable;
8 use Illuminate\Notifications\Notifiable;
9 use Laravel\Sanctum\HasApiTokens;
10
11 class User extends Authenticatable
12 {
13     use HasApiTokens, HasFactory, Notifiable;
14
15     /**
16      * The attributes that are mass assignable.
17      *
18      * @var array<int, string>
19      */
20     protected $fillable = [
21         'name',
22         'email',
23         'password',
24     ];
25
26     /**
27      * The attributes that should be hidden for serialization.
28      *
29      * @var array<int, string>
30      */
31     protected $hidden = [
32         'password',
33         'remember_token',
34     ];
35
36     /**
```

**Controlador:** Es quien actúa como intermediario entre el Modelo y la Vista, gestionando el flujo de información entre ellos y las transformaciones para adaptar los datos a las necesidades de cada uno, como se muestra en la figura 23.

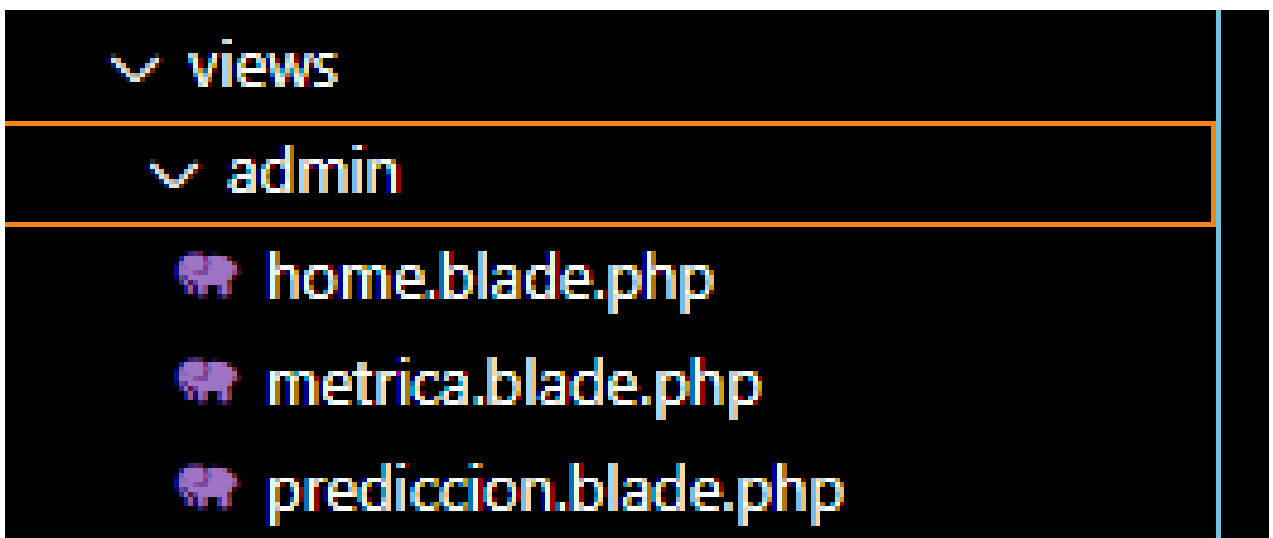
**Figura 23:** Configuración Y codificación del controlador





**Vista:** o interfaz de usuario, que compone la información que se envía al cliente y los mecanismos interacción con éste, como se detalla en la figura 24.

**Figura 24:** Configuración Y codificación de las vistas.



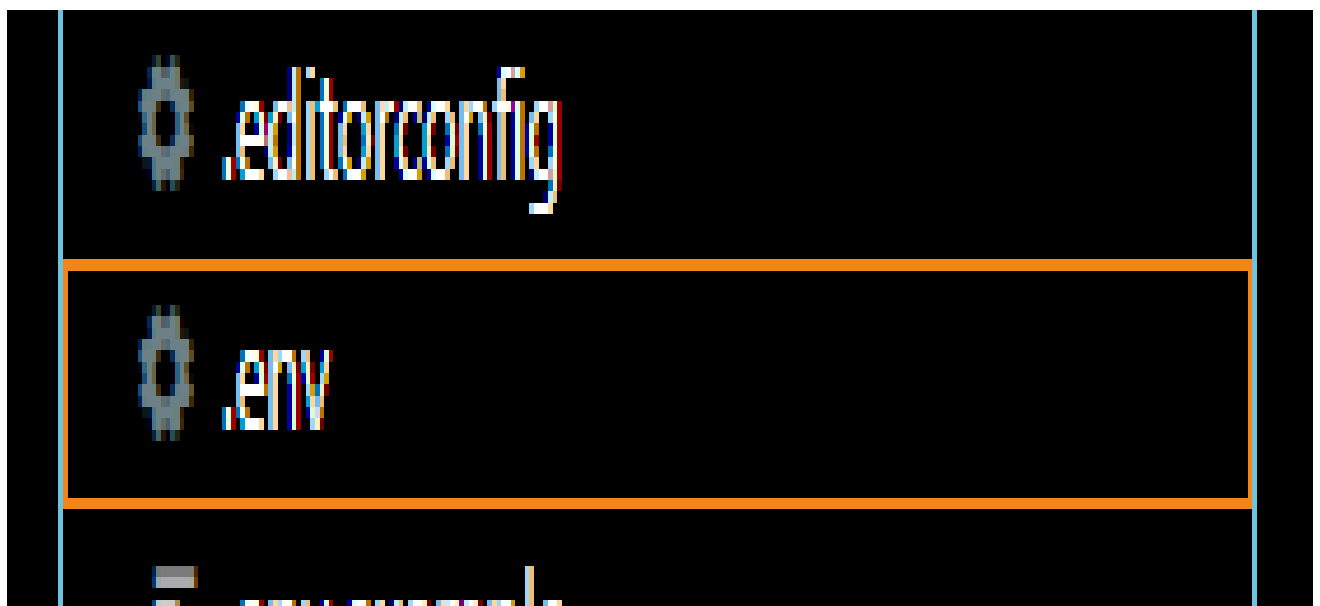
```
Restricted Mode is intended for safe code browsing. Trust this folder to enable all features. Manage Learn More
EXPLORER
TESIS
public
assets
admin
formulario-consultajs
reporteajs
css
img
js
vendor
.htaccess
faviconico
index.php
reporte.xlsx
robots.txt
resources
css
js
views
admin
home.blade.php
metrica.blade.php
prediccion.blade.php
auth
grafico
index.blade.php
OUTLINE

resources > views > admin > prediccion.blade.php
1 @extends('layouts.app')
2 @section('content')
3     <div class="pagetitle">
4         <h1>Formulario de predicci3n</h1>
5     <nav>
6         <ol class="breadcrumb">
7             <li class="breadcrumb-item"><a href="{{route('home')}}">Inicio</a></li>
8             <li class="breadcrumb-item active">Predicci3n</li>
9         </ol>
10    </nav>
11 </div><!-- End Page Title -->
12 <section class="section">
13     <div class="row">
14         <div class="col-lg-12">
15             <div class="card">
16                 <div class="card-body">
17                     <h5 class="card-title">Formulario</h5>
18
19                     <!-- General Form Elements -->
20                     <form id="formularioConsulta">
21
22                         <div class="row">
23                             <div class="row">
24                                 <div class="row mb-3 col-lg-6" >
25                                     <label class="col-sm-4 col-form-label fw-bold">DNI del estudiante</label>
26                                     <div class="col-sm-6">
27                                         <input class="form-control" type="text" step="0.01" name="dni" id="c">
28                                         <small class="text-danger" id="mensajedni"></small>
29                                     </div>
30                                 </div>

```

Conexi3n a la BD, como se detalla en la figura 25.

Figura 25: Configuraci3n Y codificaci3n de la BD.



Restricted Mode is intended for safe code browsing. Trust this folder to enable all features. Manage Learn More

EXPLORER

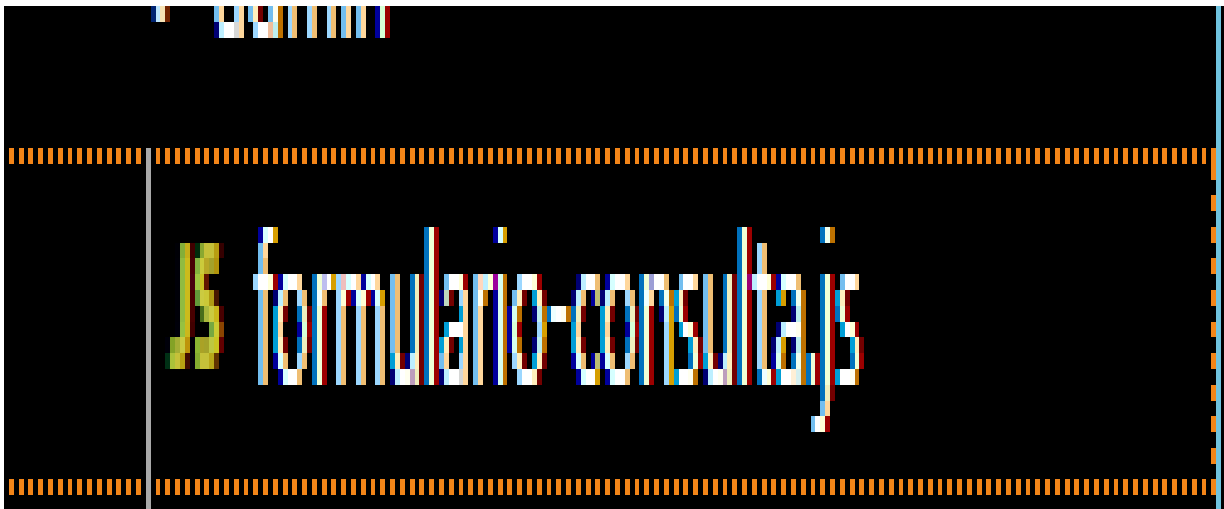
- TESIS
  - prediccion.blade.php
  - auth
    - grafico
      - index.blade.php
    - layouts
    - registro
      - home.blade.php
      - welcome.blade.php
  - routes
    - api.php
    - channels.php
    - console.php
    - web.php
  - storage
  - tests
  - vendor
  - .editorconfig
  - .env
  - .env.example
  - .gitattributes
  - .gitignore
  - artisan
  - composer.json
  - composer.lock
- OUTLINE
- TIMELINE

JS formulario-consulta.js

```
1 APP_NAME=Laravel
2 APP_ENV=local
3 APP_KEY=base64:LoqwG436ZSaMfCbcNkyytf1uzyJke9tMY2gEbknF6RI=
4 APP_DEBUG=true
5 APP_URL=http://localhost
6
7 LOG_CHANNEL=stack
8 LOG_DEPRECATIONS_CHANNEL=null
9 LOG_LEVEL=debug
10
11 DB_CONNECTION=mysql
12 DB_HOST=127.0.0.1
13 DB_PORT=3306
14 DB_DATABASE=rendimiento
15 DB_USERNAME=root
16 DB_PASSWORD=
17
18 BROADCAST_DRIVER=log
19 CACHE_DRIVER=file
20 FILESYSTEM_DISK=local
21 QUEUE_CONNECTION=sync
22 SESSION_DRIVER=file
23 SESSION_LIFETIME=120
24
25 MEMCACHED_HOST=127.0.0.1
26
27 REDIS_HOST=127.0.0.1
28 REDIS_PASSWORD=null
29 REDIS_PORT=6379
30
31 MAIL_MAILER=smt
```

**Directorio:** Donde se consume el API, de Python del proceso de predicción, como se detalla en la figura 26.

**Figura 26:** Configuración Y codificación del directorio.





```

Restricted Mode is intended for safe code browsing. Trust this folder to enable all features. Manage Learn More
EXPLORER
TESIS
  WebController.php
  Middleware
  Kernel.php
  Models
  Registro.php
  User.php
  Providers
  bootstrap
  config
  database
  lang
  public
    assets
      admin
        JS formulario-consultajs
        JS reporteajs
        css
        img
        js
        vendor
        .htaccess
        favicon.ico
        index.php
        reporte.xlsx
        robots.txt
    OUTLINE
    TIMELINE
public > assets > admin > JS formulario-consultajs > click() callback > success > url
1
2
3 $(".btnPredecir").click(function (){
4     var estado =true;
5     for(let i = 1;i <32 ; i++){
6         let selectorMensaje = '#mensaje'+i;
7         $(selectorMensaje).html('');
8         let elemento = 'p'+i;
9         let respuesta = validar(elemento);
10        if (respuesta){
11            estado = false;
12        }
13    }
14
15    if ($('#p32').val()==' '){
16        $('#mensaje32').html('Por favor, complete este campo');
17        estado = false;
18    }
19    if ($('#p32').val().length>20 || $('#p32').val().length<0){
20        $('#mensaje32').html('Ingrese un número de 0 a 20');
21        estado = false;
22    }
23
24    var dni = $('#dni').val();
25    if ($.isNumeric(dni) && dni.length === 8){
26        $('#mensajedni').html('')
27    }else {
28        $('#mensajedni').html('Ingrese un valor numérico de 8 dígitos')
29        estado = false;
30    }
31

```

**Gráficos:** Muestra la estadística de los Alumnos con problemas de Alcohol, drogadicción o deserción, también nos muestra el rendimiento académico de los estudiantes, como se muestra en la figura 27.

**Figura 27:** Configuración Y codificación de los gráficos.

```

Restricted Mode is intended for safe code browsing. Trust this folder to enable all features. Manage Learn More
EXPLORER
TESIS
  resources.txt
  resources
    css
    js
  views
    admin
      home.blade.php
      metrica.blade.php
      prediccion.blade.php
    auth
    grafico
      index.blade.php
    layouts
      app.blade.php
    registro
      index.blade.php
      home.blade.php
      welcome.blade.php
    routes
      api.php
      channels.php
      console.php
      web.php
    storage
    tests
    OUTLINE
    TIMELINE
resources > views > grafico > index.blade.php
1 @extends('layouts.app')
2 @section('content')
3     <section class="section">
4         <div class="row">
5             <div class="col-lg-12">
6                 <div class="card">
7                     <div class="card-body">
8                         <h5 class="card-title">Problemas sociales</h5>
9
10                        <!-- Column Chart -->
11                        <div id="columnChart"></div>
12
13                        <script>
14                            document.addEventListener("DOMContentLoaded", () => {
15                                new ApexCharts(document.querySelector("#columnChart"), {
16                                    series: [{
17                                        name: 'Si',
18                                        data: [{{ $licor['si'] }}, {{ $droga['si'] }}, {{ $pandilla['si'] }}]
19                                    }, {
20                                        name: 'No',
21                                        data: [{{ $licor['no'] }}, {{ $droga['no'] }}, {{ $pandilla['no'] }} ]
22                                    }
23                                ],
24                                chart: {
25                                    type: 'bar',
26                                    height: 350
27                                },
28                                plotOptions: {
29                                    bar: {
30                                        horizontal: false,
31                                        columnWidth: '55%',
32                                        endingShape: 'rounded'
33                                    }
34                                }
35                            });
36                        </script>
37                    </div>
38                </div>
39            </div>
40        </div>
41    </section>
42 @endsection

```

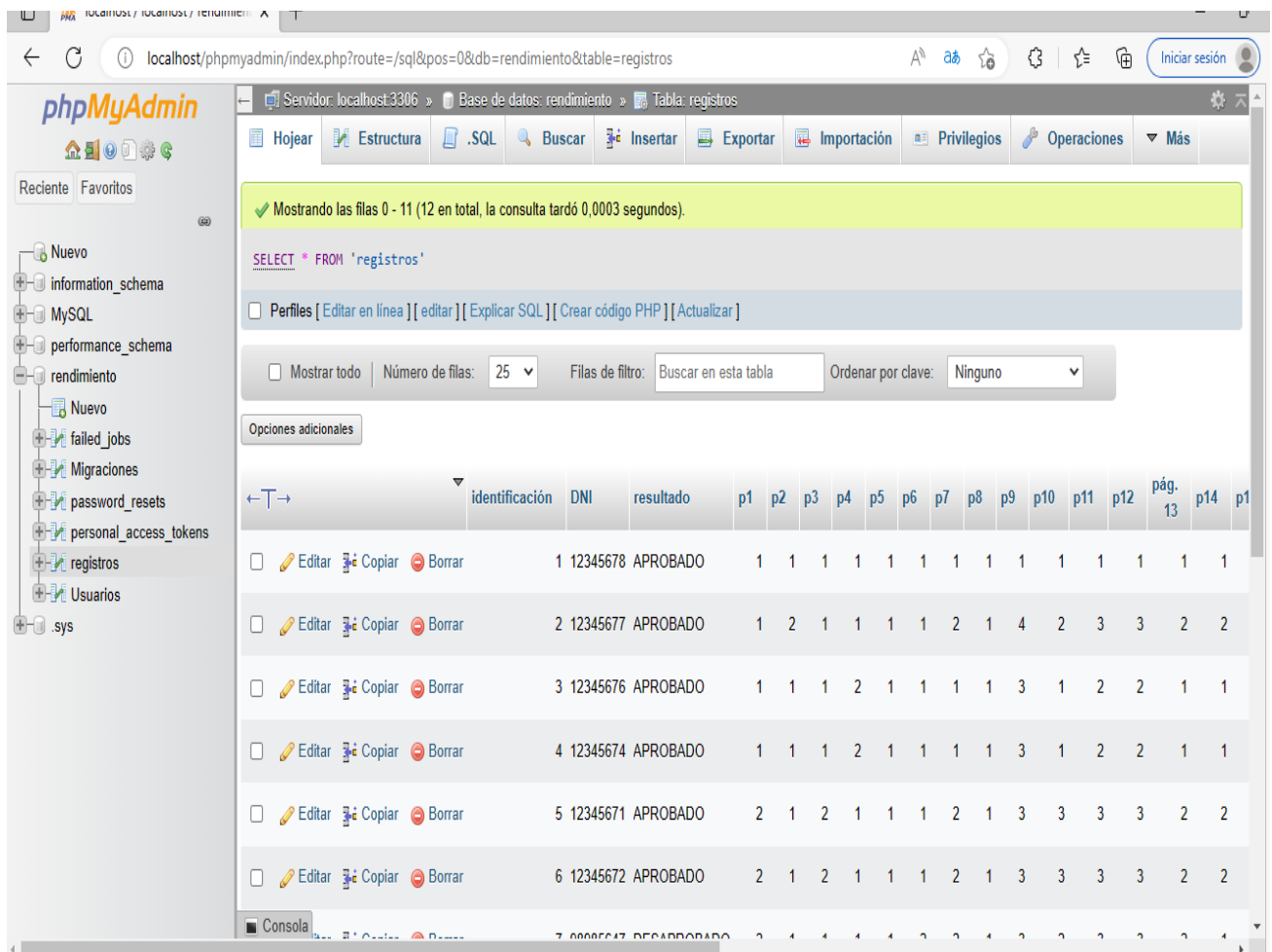
## Base de Datos:

Para la elaboración de la base de datos se usó MySQL, como se detalla en la figura 28.

**MySQL:** Es un sistema de gestión de base de datos de código abierto basado en el lenguaje de consultas estructurado y funciona con todas las plataformas.

- Se asocia con las aplicaciones web y publicaciones en línea.
- Permite acceder y almacenar a través de múltiples motores de almacenamiento.

**Figura 28:** Conexión de la base de datos con el servidor MySQL



The screenshot shows the phpMyAdmin interface for a MySQL database named 'rendimiento'. The table 'registros' is selected, and the SQL query 'SELECT \* FROM `registros`' is displayed. The table has 17 columns: 'identificación', 'DNI', 'resultado', and 16 performance metrics (p1 to p16). The data is as follows:

	identificación	DNI	resultado	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	pág. 13	p14	p15	p16
<input type="checkbox"/>	1	12345678	APROBADO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<input type="checkbox"/>	2	12345677	APROBADO	1	2	1	1	1	1	2	1	4	2	3	3	2	2	2	2
<input type="checkbox"/>	3	12345676	APROBADO	1	1	1	2	1	1	1	1	3	1	2	2	1	1	1	1
<input type="checkbox"/>	4	12345674	APROBADO	1	1	1	2	1	1	1	1	3	1	2	2	1	1	1	1
<input type="checkbox"/>	5	12345671	APROBADO	2	1	2	1	1	1	2	1	3	3	3	3	2	2	2	2
<input type="checkbox"/>	6	12345672	APROBADO	2	1	2	1	1	1	2	1	3	3	3	3	2	2	2	2

Finalmente, se usó WinSCP es una aplicación de Software Libre, Su función principal es facilitar la transferencia segura de archivos entre dos sistemas informáticos, el local y uno remoto que ofrezca servicios SSH. Es aquí donde se aloja el sistema para que se pueda ver en web en cualquier parte del mundo con la siguiente url: <https://tesis.enkargasac.com/>.



# WinSCP

Free SFTP, SCP, S3 and FTP client for Windows

[Home](#)

[News](#)

[Introduction](#)

[Download](#)

[Install](#)

## Introducción

WinSCP es una aplicación de Software Libre. WinSCP es un cliente SFTP gráfico para Windows que emplea SSH. También se puede seguir usando la versión anterior del protocolo. Su función principal es facilitar la transferencia segura de archivos entre dos sistemas informáticos, el local y uno remoto que ofrezca servicios SSH.

Esta página es una pequeña introducción en castellano, ya que la mayor parte de la documentación de WinSCP se encuentra únicamente en inglés, así como la mayor parte del contenido referido en los enlaces.

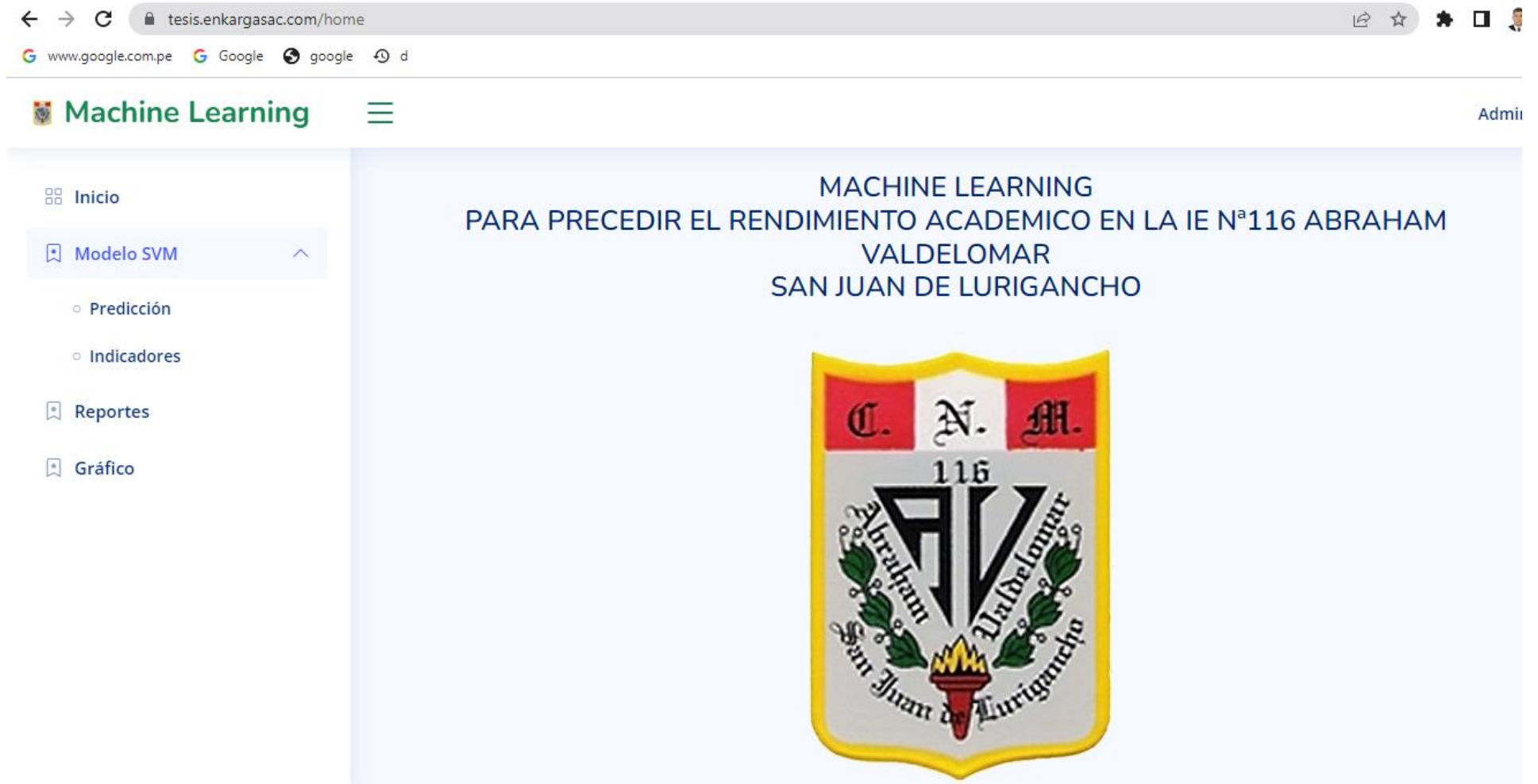
**Figura 29:** Registro de Usuario (Administrador)



The image shows a user registration form for 'ABRAHAM VALDELOMAR'. At the top, there is a logo featuring a shield with a scale of justice and a sword, with the text 'ABRAHAM VALDELOMAR' below it. The main heading of the form is 'Ingresa con tu cuenta'. Below this, there are two input fields: 'Email' and 'Contraseña'. The 'Email' field has a small grey box with an '@' symbol on the left. At the bottom of the form is a blue button labeled 'Ingresar'.

**Fuente:** Elaboración Propia

Figura 30: Pantalla de Inicio



Fuente: Elaboración Propia

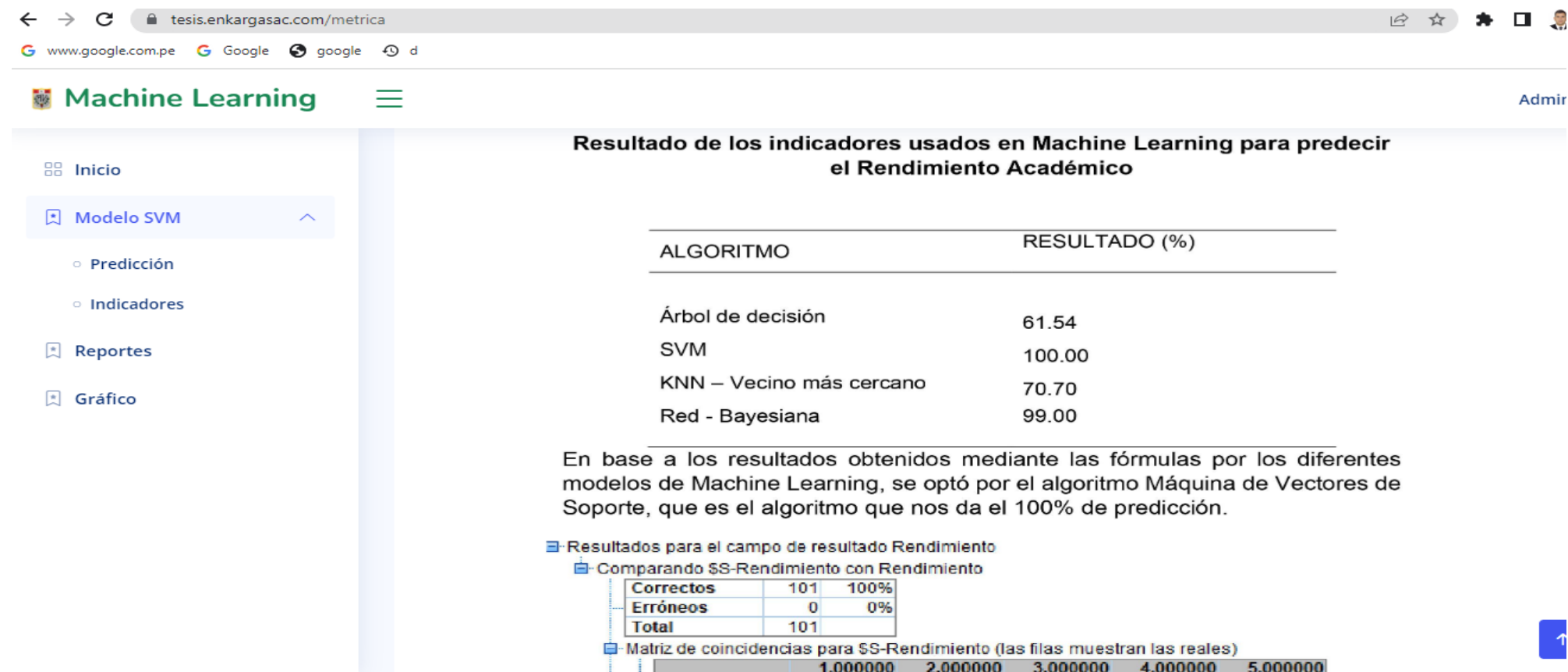
Figura 31: Modelo SVM Predicción

The screenshot shows a web browser window with the URL [tesis.enkargasac.com/prediccion](https://tesis.enkargasac.com/prediccion). The page features a navigation menu on the left with options: Inicio, Modelo SVM (selected), Predicción, Indicadores, Reportes, and Gráfico. The main content area is titled 'Formulario' and contains the following fields:

- DNI del estudiante:
- Género:
- Edad:
- Grado de estudios:
- Trabajas:
- Consumes drogas:
- Consumes licor:
- Llegas puntual a las clases:
- Perteneces a alguna pandilla:
- Qué tan responsable se considera:
- Considera posible terminar el año escolar:

Fuente: Elaboración Propia

Figura 32: Modelo SVM-Indicador



Fuente: Elaboración Propia

# Machine Learning

- Inicio
- Modelo SVM
  - Predicción
  - Indicadores
- Reportes
- Gráfico

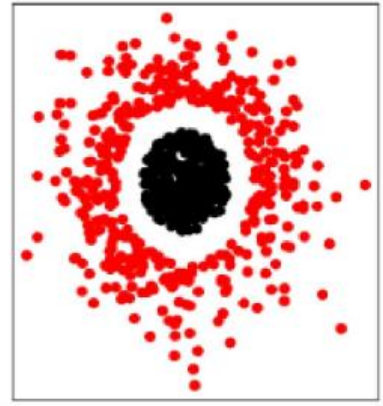
Total	101				
Matriz de coincidencias para SS-Rendimiento (las filas muestran las reales)					
	1.000000	2.000000	3.000000	4.000000	5.000000
1.000000	7	0	0	0	0
2.000000	0	27	0	0	0
3.000000	0	0	44	0	0
4.000000	0	0	0	21	0
5.000000	0	0	0	0	2

Fórmulas:

$$SENSIBILIDAD = \frac{VP}{VP + FN} \times 100$$

$$ESPECIFICIDAD = \frac{VN}{VN + FP} \times 100$$

$$PRECISIÓN = \frac{VP + TN}{VP + VN + FP + FN} \times 100$$



Elaboración: Fuente Propia



**Figura 33:** Reporte de Rendimiento académico

tesis.enkargasac.com/reportes

Machine Learning

Admir

Reportes

10 entries per page

Search...

Imprimir Exportar Excel

Dni	Predicción	Alcohol	Pandilla	Droga	Sueldo Padres	Sexo	Edad	Trabaja	Promedio Ponderado	Fecha
12345678	APROBADO	Si	Si	Si	Hasta 2000	MASCULINO	14 a 15	Si	18	18-01-2023 14:32:29
12345677	APROBADO	Si	Si	Si	Hasta 2000	MASCULINO	16 a 17	Si	11	18-01-2023 16:14:09
12345676	APROBADO	Si	Si	Si	Hasta 2000	MASCULINO	14 a 15	No	14	18-01-2023 16:51:06
12345674	APROBADO	Si	Si	Si	Hasta 2000	MASCULINO	14 a 15	No	14	18-01-2023 16:51:32

Fuente: Elaboración Propia

**Figura 34:** Grafica de los problemas Sociales de los estudiantes



Fuente: Elaboración Propia

**Figura 35:** Reporte exportado a Excel

DNI	Predicción	Género	Edad	G. Estudios	Trabajas?	Drogas	Licor	Puntualidad	Pandilla	Responsabilidad	Terminar estudios
12345678	APROBADO	MASCULINO	14 a 15	5to.	Si	Si	Si	Siempre	Si	Muy responsable	Si
12345677	APROBADO	MASCULINO	16 a 17	5to.	Si	Si	Si	Nunca	Si	Muy irresponsable	No
12345676	APROBADO	MASCULINO	14 a 15	5to.	No	Si	Si	Siempre	Si	Irresponsable	Si
12345674	APROBADO	MASCULINO	14 a 15	5to.	No	Si	Si	Siempre	Si	Irresponsable	Si
12345671	APROBADO	FEMENINO	14 a 15	4to.	Si	Si	Si	Nunca	Si	Irresponsable	No
12345672	APROBADO	FEMENINO	14 a 15	4to.	Si	Si	Si	Nunca	Si	Irresponsable	No
98985647	DESAPROBADO	FEMENINO	14 a 15	5to.	Si	Si	No	Nunca	Si	Irresponsable	No
25599421	APROBADO	FEMENINO	14 a 15	4to.	No	No	No	Siempre	Si	Muy responsable	Si
25599422	APROBADO	FEMENINO	14 a 15	4to.	No	No	Si	Siempre	Si	Muy responsable	Si
25599426	APROBADO	FEMENINO	14 a 15	4to.	Si	Si	Si	Nunca	Si	Irresponsable	No
25599478	APROBADO	MASCULINO	16 a 17	5to.	No	No	No	Siempre	No	Responsable	Si
12341234	APROBADO	MASCULINO	14 a 15	5to.	Si	Si	Si	Siempre	Si	Muy responsable	Si

Fuente: Elaboración Propia

## Anexo:10 Artículo Científico

ANALISIS DEL RENDIMIENTO ACADEMICO MEDIANTE EL USO DE MACHINE LEARNING

Variable Independiente: Machine Learning

variable dependiente: Rendimiento Académico

ACOSTA PORTOCARRERO, ANTONY CHRISTIAN

Correo: [aacostapo01@ucvvirtual.edu.pe](mailto:aacostapo01@ucvvirtual.edu.pe)

RUIZ VARGAS, PAUL

Correo: [rvargaspa01@ucvvirtual.edu.pe](mailto:rvargaspa01@ucvvirtual.edu.pe)

### RESUMEN:

El rendimiento académico de un estudiante de la IE N°116 generalmente se mide a través de sus calificaciones, las cuales derivan en una situación académica normal o deficiente, que a su vez depende de diversos factores. El machine Learning es la forma más efectiva de lidiar con este problema. Por lo cual, esta investigación fue encontrar los principales predictores de la situación académica de un estudiante de la IE N°116 luego de que transcurrieron 9 meses de estudio. El proceso de desarrollo de un sistema de aprendizaje automático para predecir los resultados del aprendizaje se divide en 4 etapas. Etapa de selección de Datos: se recolectan datos por Google formulario. Etapa de Procesamiento-Limpieza de datos: Se procesa los datos y se depura los datos inservibles. Etapa de transformación y reducción: se Usó el Machine Learning con los algoritmos. Etapa minería de datos: se usa los algoritmos de aprendizaje automático para comparar los resultados y ver el mejor modelo predictivo. En ese sentido se aplicó el método KDD (Knowledge Discovery in Database), usando 4 cuatro algoritmos de clasificación: árbol de decisión, máquinas de vectores SVM, KNN-vecinos y Red Bayesiana. El modelo con la técnica Maquina de vectores- SVM resultó el más adecuado para la mayoría de las métricas empleadas en el estudio, principalmente, para identificar estudiantes en peligro de reprobación. Contribución- El trabajo conjunto de todos los participantes en la IE N°116 permite mejorar el progreso de los Estudiantes mediante el uso de Machine Learning con el uso de algoritmos.

**Palabras clave:** análisis de datos; aprendizaje automático; Evaluación de competencias; autoevaluación, minería de datos; rendimiento académico.

#### ABSTRACT:

**Purpose-** The academic performance of a student at IE No. 116 is generally measured through her grades, which result in a normal or deficient academic situation, which in turn depends on various factors. Machine Learning is the most effective way to deal with this problem. Therefore, this research was to find the main predictors of the academic situation of a student of IE No. 116 after 9 months of study. **Method-** The process of developing a machine learning system to predict learning outcomes is divided into 4 stages. **Data selection stage:** data is collected by Google form. **Data Processing-Cleaning Stage:** The data is processed and useless data is purged. **Transformation and reduction stage:** Machine Learning was used with the algorithms. **Data mining stage:** machine learning algorithms are used to compare the results and see the best predictive model. **Results.** In this sense, a data mining technique was applied to select predictor variables and 4 four classification algorithms were applied: decision tree, SVM vector machines, KNN-neighbors and Bayesian Network. Then, those models with the best values of sensitivity, specificity and Accuracy were chosen, where the machine learning model that achieves the highest precision in student performance is a vector machine. **Contribution-** The joint work of all the participants in IE N°116 allows to improve the progress of the Students through the use of Machine Learning with the use of algorithms.

**Keywords:** data analysis; machine learning; Competency assessment; self-assessment, data mining; academic performance

## Introducción

En el Perú, el desempeño estudiantil es uno de los temas más discutidos y de mayor preocupación para nuestras autoridades y padres de familia. No solo en nuestro país, sino también en muchos otros países latinoamericanos y otros continentes (Lamas, 2015). El aprendizaje automático o aprendizaje automático aplicado (ML) en la educación se puede definir como la recopilación, el análisis y la difusión de datos sobre los alumnos para comprender y optimizar los aspectos relevantes del proceso de enseñanza y aprendizaje. (Mohamed, 2012). El concepto de rendimiento académico es un concepto complejo y tiene muchas interpretaciones diferentes, por lo que se considera un concepto multidimensional, relativo y contextual. (González Ignacio, 2004).

Según Menacho Chiok (2017 p. 2), en su artículo desarrollado en Perú titulado Aplicación de técnicas de minería de datos para predecir el rendimiento académico; planteó preocupaciones sobre los resultados de aprendizaje de los estudiantes y concluye que el asunto más importante que deben abordar las instituciones de educación superior, y las técnicas de minería de datos (TMD) aplicadas a los datos generados en instituciones educativas probadas Conviértase en una herramienta eficaz para predecir el efecto de enseñanza de los estudiantes. Se concluye que los resultados mostraron que la red Naive Bayes (una red bayesiana que muestra gráficamente la dependencia y la independencia entre tres variables aleatorias, especialmente la independencia condicional), que tuvo la tasa de clasificación correcta más alta (71,0%) (Menacho Chiok, 2017)

Hay diferentes definiciones de rendimiento académico. A algunos les gusta (Guiselle, 2007), que el logro académico es la suma de muchos factores diferentes y complejos que afectan a un estudiante (Tourón, 1985), argumenta que el logro académico es el resultado del aprendizaje iniciado por intervenciones pedagógicas iniciadas por el maestro y generadas por el estudiante. (Rojas Luis, 2013). conceptualizarlos como una serie de factores que rodean el resultado final de los esfuerzos del alumno.; y(. García

José R), lo consideran un factor determinante del éxito y fracaso de los estudiantes, por lo que debe ser considerado como uno de los aspectos importantes a la hora de analizar el resultado del proceso educativo.

Este estudio describe la importancia de utilizar el aprendizaje automático para medir el rendimiento de los estudiantes de 4to y 5to grado en instituciones públicas utilizando algoritmos de evaluación comparativa: árbol de decisión, SVM, KNN-Vecino y el algoritmo bayesiano. red para mejorar el rendimiento de los estudiantes.

Finalmente, se recolectan datos para análisis, 70% para entrenamiento y 30% para validación o prueba. Evaluación de la sensibilidad, la especificidad y la precisión de la validación del modelo, con un 90 % de radio base que funciona mejor que un 64 % de kernels lineales. A partir de este estudio, la implementación de SVM y los índices actuarán como sistema de referencia para la evaluación del modelo (Burman & Som, 2019)

**MÈTODO:** En la literatura revisada se observó que existen pocos trabajos en Perú que utilizan técnicas de aprendizaje automático para el diseño de modelos predictivos del rendimiento académico, a pesar del beneficio potencial que pueden tener en el desempeño académico de estudiantes, específicamente, la predicción del rendimiento académico ofrece la oportunidad de elaborar planes de prevención de reprobación estudiantil mediante la realización de estrategias de intervención en lugar de estrategias de recuperación académica. Es decir, estos modelos permiten a los profesores e instituciones educativas realizar intervenciones desde el principio del curso y no al final cuando es demasiado tarde para realizar alguna acción para evitar la reprobación del estudiante.

En este artículo se propone una metodología basada en recopilar información de estudiantes al inicio, o incluso antes, de un curso de una asignatura, como se muestra en la Figura 1. Posteriormente, se construyen modelos predictivos que permiten predecir el rendimiento académico que obtendrán futuros estudiantes al finalizar el curso. Finalmente, se evalúan los modelos y se comparan con base en métricas representativas.

**Figura 1:** Metodología Propuesta



Fuente: Elaboración Propia

Este trabajo se fundamenta en el campo del análisis educativo, el objetivo es presentar un modelo de un algoritmo de aprendizaje automático para determinar la capacidad de predecir los resultados de aprendizaje de los estudiantes de la IE N°116” Abraham Valdelomar”, en esta investigación participaron 101 alumnos mediante un cuestionario de 36 preguntas.

#### Etapa 1: Etapa de selección de datos

Para obtener datos sobre los estudiantes se elaboró un cuestionario con 36 preguntas, las preguntas se escribieron en una hoja de Google y se enviaron a estudiantes 4to y 5to grado de la IE N°116 recibió 101 postulaciones. Este tipo de características fueron seleccionados debido a que son más simples de recolectar y han sido utilizados en otros estudios de predicción del rendimiento académico (Shahiri et al., 2015). Según la figura 2

**Figura 2:** Cuestionario utilizando Google Forms

**Factores que influyen en el rendimiento escolar**

Buenas buenas días/buenas tardes Estimados alumnos, somos bachilleres en Ingeniería de Sistemas que estamos realizando una investigación para nuestra Tesis de titulación acerca del "Rendimiento Académico de los estudiantes del 4to. y 5to. grado de secundaria" de la institución a las cuales ustedes pertenecen. De tal manera que los resultados que se obtengan de esta investigación servirán a sus autoridades y profesores de la institución para que tomen medidas correctivas para el éxito de los alumnos. Para lo cual solicitamos su colaboración para responder honestamente las preguntas de este formulario. Agradecemos su valiosa colaboración para el éxito de nuestra investigación.

Fuente: Elaboración propia

#### Etapa 2: Etapa de preprocesamiento / limpieza de datos

Se analizo 101 Preguntas después de completar la recopilación de datos a través del formulario de Google. En el caso de una tabla creada en Excel



mediante un formulario, eliminamos los datos que se ingresaron incorrectamente. Nos referimos a la información ingresada por los estudiantes. Por ejemplo, su condado, horas trabajadas, horas dedicadas a estudiar, promedio ponderado, etc. En segundo lugar, de los datos exportados de Google From a Excel, corresponde a 87 registros de estudiantes. Se realizó el análisis de datos, se confirmaron los datos nulos faltantes y también se normalizaron las respuestas. Figura 3.

**Figura 3:** Datos importados de Google Forms a Excel

	AD	AE	AF	AG	AH	AI	AJ	AK	AL
1	Tiene empatía hacia sus	Cuando obtienes malas r	Que cantidad de amigos	Cantidad de horas que tr	Ingreso familiar aproxim	Sus estudios iniciales en	Establecimiento en el qu	Promedio ponderado durante el año	
7	Empático	Frustrado	Regular	Ninguna	4000soles	Público	Público	16	
8	Muy empático	Frustrado	Regular	ninguna	4500 soles	Privado	Público	17	
9	Empático	Todas	Regular	No trabajo	5000 soles	Privado	Público	16	
10	Muy empático	Triste	Muchos	NO TRABAJAJO	3000 SOLES	Público	Público	15	
11	Empático	Frustrado	Pocos	8horas	2500soles	Público	Público	15	
12	Empático	Triste	Regular	No trabajo	4000 soles	Privado	Público	18	
13	Empático	Triste	Muchos	60	3500	Público	Público	20	
14	Muy empático	Todas	Muchos	12 horas	2200 soles	Público	Público	13	
15	Empático	Todas	Regular	0	5000	Público	Público	17	
16	Muy empático	Triste	Regular	NO TRABAJAJO	3500 SOLES	Público	Público	17	
17	Muy empático	Frustrado	Regular	60	3500	Público	Público	20	
18	Empático	Molesto	Pocos	NO TRABAJAJO	5000 SOLES	Público	Público	16	
19	Empático	Triste	Muchos	8	3000	Público	Público	15	
20	Empático	Frustrado	Regular	8	3000	Privado	Público	18	
21	Empático	Triste	Regular	NO TRABAJAJO	4500 SOLES	Privado	Público	16	
22	Empático	Todas	Regular	5	1300	Público	Público	14	
23	Empático	Triste	Regular	NO TRABAJAJO	6000 SOLES	Privado	Público	17	
24	Muy empático	Triste	Regular	No trabajo	4000	Público	Público	18	

Fuente: Elaboración propia.

### Etapa 3: Etapa de transformación y reducción

En este punto, también es razonable utilizar SPSS Modeler versión 18 para crear un modelo predictivo de aprendizaje automático que excluya algunas variables no afectadas. Al cargar una base de datos creada con SPSS Statistics, esta carga utiliza un "Nodo de origen" de tipo "Archivo de estadísticas" para examinar un registro de datos de tipo tabla "Nodo de resultado" que le permite visualizar los datos numéricos generados por los datos. Según la Figura 4.

Figura 4: Carga de datos con nodo origen - SPSS Modeler



Fuente: Elaboración propia

#### Etapa 4: Minería de datos

En este paso, se seleccionan varios algoritmos de aprendizaje automático para comparar los resultados y obtener el mejor modelo de predicción. En este estudio se utilizaron árboles de decisión, máquinas vectoriales y K-NN. Estos algoritmos son proporcionados por SPSS Modeler. Según la figura 5.

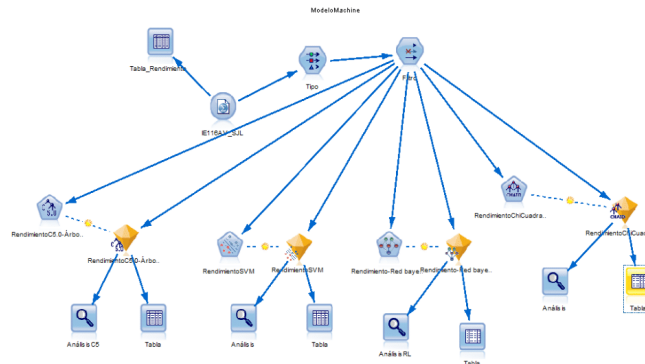
**Figura 5:** Algoritmos de aprendizaje automático - SPSS Modeler Ver.18



Fuente: Elaboración propia

A continuación, se muestra los modelos creados por cada algoritmo de aprendizaje según la Figura 6.

**Figura 6:** Proyecto de rendimiento académico – SPSS Modeler



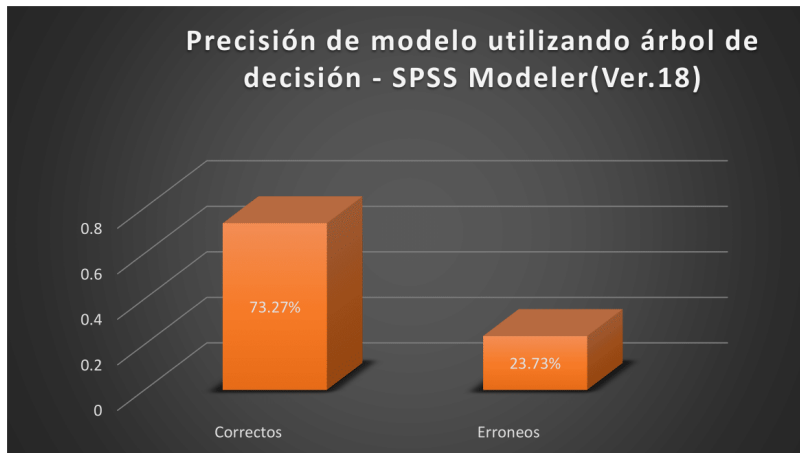
Fuente: Elaboración propia

**RESULTADOS:** La construcción de modelos predictivos requiere evaluarlos, es decir, comprobar que arroje resultados suficientemente satisfactorios, por lo que se requiere el planteamiento de métricas de evaluación. En este trabajo se utiliza la exactitud, tasa de verdaderos positivos y la tasa de verdaderos negativos (Durairaj y Vijitha, 2014). La exactitud fue definida en la sección anterior como la cantidad de predicciones correctas con respecto al total de las predicciones. La tasa de verdaderos positivos es una métrica que se calcula dividiendo el número de registros predichos como positivos entre el total de registros positivos, en este trabajo, representaría las predicciones correctas de estudiantes aprobados con respecto al total de aprobados. De manera similar, la tasa de verdaderos negativos es el número de registros predichos como reprobados con respecto al total de registros de estudiantes que reprobaron.

#### **Árbol de decisión – C5,**

Esta herramienta proporciona información sobre qué variables son más importantes para las variables de rendimiento. Esta herramienta te permite visualizar que el modelo logra un 73,27% de precisión y un 26,73% de error, según la figura 7.

Figura 7: Precisión de modelo utilizando árbol de decisión - SPSS Modeler

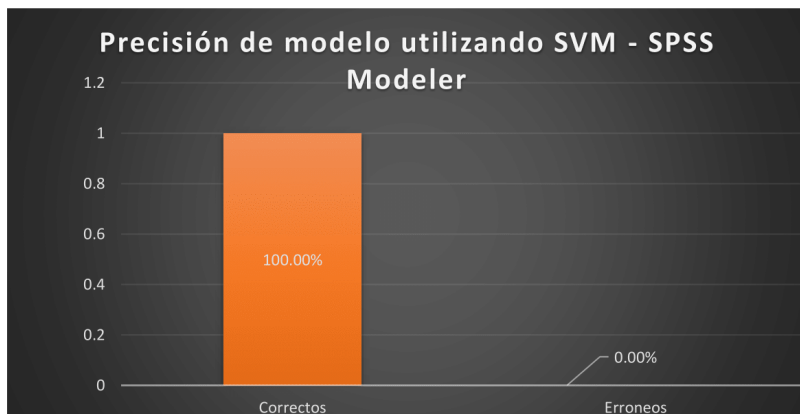


Fuente: Elaboración propia

#### Máquina de Vectores (SVM):

Esta herramienta proporciona información sobre qué variables son más importantes en relación con la variable de rendimiento. Esta herramienta nos permite visualizar que el modelo ha logrado un 100 % de precisión con un 0 % de tasa de error, según como se detalla en la Figura 8

Figura 8: Precisión de modelo utilizando SVM - SPSS Modeler Ver. 18

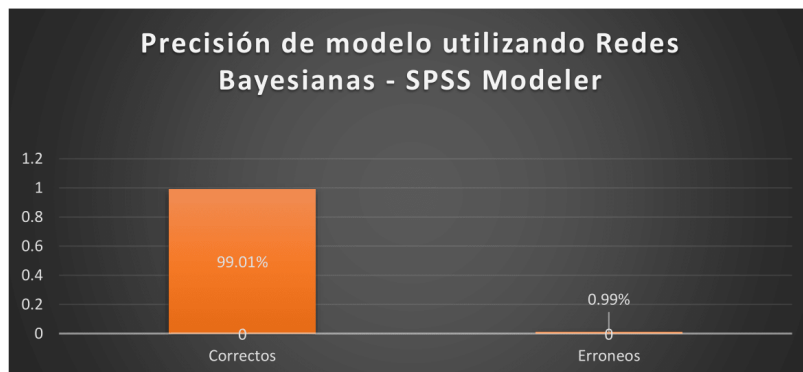


Fuente: Elaboración propia

### Red Bayesiana:

Esta herramienta proporciona información sobre qué variables son más importantes en relación con la variable de rendimiento. Esta herramienta nos permite visualizar que el modelo logra un 99,01% de precisión con una tasa de error del 0,99%, según como se detalla en la Figura 9.

**Figura 9:** Precisión de modelo utilizando Red Bayesiana - SPSS Modeler Ver18.

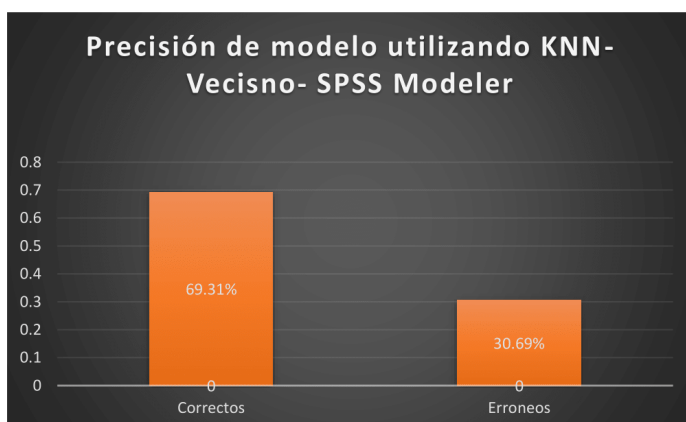


Fuente: Elaboración propia

### KNN-vecinos:

Esta herramienta proporciona información sobre qué variables son más importantes en relación con la variable de rendimiento. Esta herramienta nos permite visualizar que el modelo logra un 69,31% de precisión con una tasa de error del 30,69%, según como se detalla en la Figura 10.

**Figura 10:** Precisión de modelo utilizando KNN-vecinos - SPSS Modeler Ver18.



Fuente: Elaboración propia.

#### **DISCUSIÓN**

En la Figura 7, el aprendizaje automático predice el desempeño de los alumnos de la IE N° 116 - "Abraham Valdelomar"; con una precisión del 73.27% utilizando el algoritmo del árbol de decisión.

En la Figura 8, el aprendizaje automático nos permite predecir el rendimiento académico de los estudiantes de la IE N° 116 -"Abraham Valdelomar"; con un 100 % de precisión utilizando el método Máquina de Vectores (SVM).

En la Figura 9; Explicación: Machine Learning predice el rendimiento académico de los alumnos de la IE N°116"Abraham Valdelomar" - San Juan de Lurigancho con 99.01% de precisión usando el método Redes-Bayesianas.

En la figura 10; El aprendizaje automático predice el desempeño de los alumnos de 4to y 5to de la IE No. 116 - "Abraham Valdelomar" - San Juan de Lurigancho; con una precisión del 69.31% utilizando el método KNN-Vecinos.

En la literatura existen trabajos que utilizan técnicas de aprendizaje similares a los utilizados en la presente investigación. Juárez et al. (2014) realizaron un estudio en el que participaron 104 estudiantes y el valor más alto de exactitud fue de 80%. Salal et al. (2019) emplearon 649 registros de estudiantes y el valor

mayor de exactitud obtenido fue 76.7%. Para ambos trabajos se requirieron alrededor de 30 o más atributos correspondientes a datos personales, de domicilio, entre otros, además, solo se calculó la exactitud como métrica de evaluación.

Castrillón et al. (2020) utilizaron únicamente la técnica de aprendizaje automático árbol de decisión y 22 atributos. Se emplearon 460 registros de estudiantes para el modelo y se obtuvo una exactitud del 91%, sin embargo, una de las razones de este valor puede deberse a que la exactitud fue calculada con métodos que introducen poca aleatoriedad a la evaluación, entre los cuales fueron la validación cruzada con dos particiones y otras evaluaciones que consistieron en dividir los datos en una partición para crear el modelo y otra partición para realizar las predicciones. Estos métodos de evaluación con poca aleatoriedad tienden a favorecer al modelo que se ajusta más al conjunto de datos de entrenamiento (Hernández et al., 2004).

## **CONCLUSIONES**

Se concluyó que el rendimiento académico de la IE N°116 se puede mejorar utilizando el aprendizaje automático, los algoritmos utilizados fueron de la más alta precisión, lo que se refleja en que lograron los mejores resultados en cuanto a métricas de evaluación con precisión del 100%. Además, permite que todos los docentes de la institución aprendan nuevas tecnologías para enseñar a los estudiantes.

## REFERENCIAS

- Burgos, C., Campanario, M. y otros cuatro autores, Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout, <https://doi.org/10.1016/j.compeleceng.2017.03.005>, Computers and Electrical Engineering, 66, 541-556 (2018).
- Castrillón, O., Sarache, W., y Ruiz, S., Predicción del rendimiento académico por medio de técnicas de inteligencia artificial, <https://doi.org/10.4067/S0718-50062020000100093>, Revista Formación Universitaria, 13, 93-102 (2020) .
- Contreras, L., y Rodríguez, J., Big data: An exploration toward the improve of the academic performance in higher education, [https://doi.org/10.1007/978-3-319-93803-5\\_59](https://doi.org/10.1007/978-3-319-93803-5_59). Lecture Notes in Computer Science, 10943, 627-637 (2018) .
- De La Hoz, E., De La Hoz, E., y Fontalvo, T. Methodology of Machine Learning for the classification and Prediction of users in Virtual Education Environments, <https://doi.org/10.4067/S0718-07642019000100247>, Información Tecnológica, 30, 247-254 (2019).
- Dyckhoff, A., Zielke, D. y otros tres autores, Design and implementation of a learning analytics toolkit for teachers, Educational Technology and Society, 15, 58-76 (2012) [ Links ]
- Escudero, T., Indicadores del rendimiento académico una experiencia en la Universidad de Zaragoza - Ministerio de Educación y Cultura, 1º edición, 251-262. Centro de Publicaciones, España (1999) [ Links ]
- Fernandes, E., Holanda, M., y otros tres autores, Educational data mining: Predictive analysis of academic performance of public-school students in the capital of Brazil, <https://doi.org/10.1016/j.jbusres.2018.02.012>, Journal of Business Research, 94, February 2018, 335-343. (2019) [ Links ]
- Garbanzo, R., y María, G., Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública, Revista Educación, 31,1, 43-63 (2007) [ Links ]
- García, J., Sánchez, P., Orozco, M., y Obredor, S., Extracción de Conocimiento para la Predicción y Análisis de los Resultados de la Prueba de Calidad de la Educación Superior en Colombia, <https://doi.org/10.4067/S0718-50062019000400055>, Revista Formación Universitaria, 12, 4, 55-62 (2019) [ Links ]
- García, K., Learning Analytics as an analysis factor of university academic performance, CEUR Workshop Proceedings, 2231, 42-50 (2019) [ Links ]
- Jahangiri, A., y Rakha, H., Applying Machine Learning Techniques to Transportation Mode Recognition Using Mobile Phone Sensor Data, <https://doi.org/10.1109/TITS.2015.2405759>, IEEE Transactions on Intelligent Transportation Systems, 16 ,5, 2406-2417 (2015) [ Links ]
- Khan, I. A., y Choi, J. T. An Application of Educational Data Mining (EDM) Technique for Scholarship Prediction, <https://doi.org/10.14257/ijseia.2014.8.12.03>, International Journal of Software Engineering and Its Applications, 8 ,12, 31-42 (2014). [ Links ]



- Lonn, S., Aguilar, S. y Teasley, S., Investigating student motivation in the context of a learning analytics intervention during a summer bridge program, <https://doi.org/10.1016/J.CHB.2014.07.013>, *Computers in Human Behavior*, 47, 90-97 (2015). [ Links ]
- MEN-Sistema Nacional de Información de la Educación Superior., MEN-Sistema Nacional de Información de la Educación Superior, (2017) [ Links ]
- Nieto, Y., García, V., Montenegro, C., y Crespo, R., Supporting academic decision making at higher educational institutions using machine learning-based algorithms, <https://doi.org/10.1007/s00500-018-3064-6>, *Soft Computing*, 23, 4145-4153 (2018) [ Links ]
- Alvaro. (2018, April 30). MACHINE LEARNING, DATA SCIENCE Y ANALÍTICA AVANZADA. ¿Qué Es El Aprendizaje Automático o Machine Learning?
- Menacho Chiok, C. H. (2017). Predicción del rendimiento académico aplicando técnicas de minería de datos. *Anales Científicos*, 78(1), 26. <https://doi.org/10.21704/ac.v78i1.811>
- Murillo. (2009). *Investigación Aplicada*. 12-12. [https://www.google.com/search?q=La+investigacion+aplicada%2C+segun+murillo&rlz=1C1GCEU\\_esPE971PE971&oq=La+investigacion+aplicada%2C+segun+murillo&aqs=chrome..69i57j33i10i160l2j33i22i29i30.22629j0j7&sourceid=chrome&ie=UTF-8](https://www.google.com/search?q=La+investigacion+aplicada%2C+segun+murillo&rlz=1C1GCEU_esPE971PE971&oq=La+investigacion+aplicada%2C+segun+murillo&aqs=chrome..69i57j33i10i160l2j33i22i29i30.22629j0j7&sourceid=chrome&ie=UTF-8)
- Nithya, P., Umamaheswari, B., y Umadevi, A. A Survey on Educational Data Mining in Field of Education, *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* 16,1, 145-153 (2019) [ Links ]
- Noel, M., Ayán, R., Ángel, M., y Díaz, R., Indicadores de rendimiento de estudiantes universitarios, <http://10-4438/1988-592X-RE-2011-355-033>, *Revista de Educación*, 355, 467- 492 (2011) [ Links ]
- Oblinger, G., Campbell, J., y otros dos autores, *Academic Analytics: A New Tool for a New Era*, *Research in Higher Education*, 1(2), 727-742 (2007) [ Links ]
- Osmanbegović, E., y Suljić, M., Data mining approach for predicting student performance, *Journal of Economics and Business*, 10, 1, 20-30 (2012). [ Links ]
- Page, M., Gaviria, J., y Gómez, C., *Hacia un modelo causal del rendimiento académico*, Ministerio de educación, 1ª edición, 25-230. Centro de publicaciones y secretaria General, España, (1990) [ Links ]
- Palmer, S., y Stuart., Modelling engineering student academic performance using academic analytics, *The International journal of engineering education*, 29,1,132-138 (2013) [ Links ]
- Porto, A., y Gresia, L. Di., Performance of University students and their determinants. *Revista de economía y estadística*, 42,1, 93-113 (2005) [ Links ]
- Radhwan, A., Abbas, A, y Ali, S., Popular Decision Tree Algorithms of Data Mining Techniques, *International Journal of Computer Science and Mobile Computing*, 6,6, 133-142 (2017) [ Links ]
- Ramesh, V., Parkavi, P., y Ramar, K., Predicting Student Performance: A Statistical and Data Mining Approach, <https://www.ijcaonline.org/archives/volume63/number8/10489-5242>, *International Journal of Computer Applications*, 63,8,35-39 (2013). [ Links ]
- Rojas, L., Validez predictiva de los componentes del promedio de admisión a la universidad de costa rica utilizando el género y el tipo de colegio como variables control, <https://revistas.ucr.ac.cr/index.php/aie/article/view/11707/18183>, *Revista Electrónica Actualidades Investigativas En Educación*, 13(1), 17-25, (2013) [ Links ]
- Salcedo, A., Desertion in Colombian Universities, [http://www.alfaguia.org/alfaguia/files/1319043663\\_03.pdf](http://www.alfaguia.org/alfaguia/files/1319043663_03.pdf), *Revista Academia y Virtualidad*, 3(1), 50-60 (2010) [ Links ]

- Sánchez, P. y García, J., A new methodology for neural network training ensures error reduction in time series forecasting, <https://doi.org/10.3844/jcssp.2017.211.217>, Journal of Computer Science, 13, 211-217 (2017) [ Links ]
- Santosh, K., AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data, <https://doi.org/10.1007/s10916-020-01562-1>, Journal of Medical Systems, 44(5), 1-5 (2020) [ Links ]
- Tourón, J., La predicción del rendimiento académico: Procedimientos y resultados, <http://dadun.unav.edu/handle/10171/18774>, Revista Española de Pedagogía, 1(25), 168-182 (1985) [ Links ]
- Zaffar, M., Hashmani, M. A., Savita, K. S., y otros tres autores, A Study of Feature Selection Algorithms for Predicting Students Academic Performance, <https://10.14569/IJACSA.2018.090569>



**UNIVERSIDAD CÉSAR VALLEJO**

**FACULTAD DE INGENIERÍA Y ARQUITECTURA  
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

### **Declaratoria de Autenticidad del Asesor**

Yo, PACHECO PUMALEQUE ALEX ABELARDO, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA ESTE, asesor de Tesis titulada: "MACHINE LEARNING PARA PRECEDIR EL RENDIMIENTO ACADEMICO EN LA IE Nª116 ABRAHAM VALDELOMAR SAN JUAN DE LURIGANCHO 2022.", cuyos autores son ACOSTA PORTOCARRERO ANTONY CHRISTIAN, RUIZ VARGAS PAUL, constato que la investigación tiene un índice de similitud de 17.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

LIMA, 15 de Diciembre del 2022

<b>Apellidos y Nombres del Asesor:</b>	<b>Firma</b>
PACHECO PUMALEQUE ALEX ABELARDO <b>DNI:</b> 41651279 <b>ORCID:</b> 0000-0001-9721-0730	Firmado electrónicamente por: AAPACHECOP el 18- 12-2022 18:02:33

Código documento Trilce: TRI - 0490203