



**UNIVERSIDAD TECNOLÓGICA  
INDOAMÉRICA**

**FACULTAD DE INGENIERÍA Y TECNOLOGÍAS DE LA  
INFORMACIÓN Y LA COMUNICACIÓN**

**CARRERA DE INGENIERÍA INDUSTRIAL**

**TEMA:**

---

**BIG DATA COMO HERRAMIENTA PARA EL ANÁLISIS DEL DESEMPEÑO  
ACADÉMICO EN LA EDUCACIÓN**

---

Trabajo de Integración Curricular previo a la obtención del título de Ingeniero Industrial

**Autor(a)**

Chucuri Real Boris Gabriel

**Tutor(a)**

PhD. Ayala Chauvin Manuel Ignacio

AMBATO – ECUADOR

2022

**AUTORIZACIÓN POR PARTE DEL AUTOR PARA LA CONSULTA,  
REPRODUCCIÓN PARCIAL O TOTAL, Y PUBLICACIÓN ELECTRÓNICA  
DEL TRABAJO DE INTEGRACIÓN CURRICULAR**

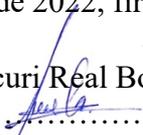
Yo Chucuri Real Boris Gabriel, declaro ser autor del Trabajo de Integración Curricular con el nombre “Big Data como herramienta para el análisis del desempeño académico en la educación”, como requisito para optar al grado de Ingeniero industrial y autorizo al Sistema de Bibliotecas de la Universidad Tecnológica Indoamérica, para que con fines netamente académicos divulgue esta obra a través del Repositorio Digital Institucional (RDI-UTI).

Los usuarios del RDI-UTI podrán consultar el contenido de este trabajo en las redes de información del país y del exterior, con las cuales la Universidad tenga convenios. La Universidad Tecnológica Indoamérica no se hace responsable por el plagio o copia del contenido parcial o total de este trabajo.

Del mismo modo, acepto que los Derechos de Autor, Morales y Patrimoniales, sobre esta obra, serán compartidos entre mi persona y la Universidad Tecnológica Indoamérica, y que no tramitaré la publicación de esta obra en ningún otro medio, sin autorización expresa de la misma. En caso de que exista el potencial de generación de beneficios económicos o patentes, producto de este trabajo, acepto que se deberán firmar convenios específicos adicionales, donde se acuerden los términos de adjudicación de dichos beneficios.

Para constancia de esta autorización, en la ciudad de Ambato, a los 14 días del mes de septiembre de 2022, firmo conforme:

Autor: Chucuri Real Boris Gabriel

Firma: .....  .....

Número de Cédula: 1805304282-2.

Dirección: Tungurahua, Píllaro, La Matriz.

Correo Electrónico: borisgabriel4562@gmail.com

Teléfono: 0983295011

## **APROBACIÓN DEL TUTOR**

En mi calidad de Tutor del Trabajo de Integración Curricular “BIG DATA COMO HERRAMIENTA PARA EL ANÁLISIS DEL DESEMPEÑO ACADÉMICO EN LA EDUCACIÓN” presentado por Chucuri Real Boris Gabriel, para optar por el Título de Ingeniero Industrial,

### **CERTIFICO**

Que dicho Trabajo de Integración Curricular ha sido revisado en todas sus partes y considero que reúne los requisitos y méritos suficientes para ser sometido a la presentación pública y evaluación por parte los Lectores que se designe.

Ambato 14, de septiembre del 2022

.....  
PhD. Manuel Ignacio Chauvin Ayala

## DECLARACIÓN DE AUTENTICIDAD

Quien suscribe, declaro que los contenidos y los resultados obtenidos en el presente Trabajo de Integración Curricular, como requerimiento previo para la obtención del Título de Ingeniero Industrial, son absolutamente originales, auténticos y personales y de exclusiva responsabilidad legal y académica del autor

Ambato 14, septiembre de 2022



.....

Boris Gabriel Chucuri Real  
180530428-2

## **APROBACIÓN DE LECTORES**

El Trabajo de Integración Curricular ha sido revisado, aprobado y autorizada su impresión y empastado, sobre el Tema: BIG DATA COMO HERRAMIENTA PARA EL ANÁLISIS DEL DESEMPEÑO ACADÉMICO EN LA EDUCACIÓN, previo a la obtención del Título de Ingeniero Industrial, reúne los requisitos de fondo y forma para que el estudiante pueda presentarse a la sustentación del Trabajo de Integración Curricular.

Ambato, 14 de septiembre de 2022

.....

MSc. Patricio Eduardo Sánchez Díaz  
LECTOR

.....

PhD. Pedro Fernando Escudero Villa  
LECTOR

## **DEDICATORIA**

Este trabajo de titulación va dedicado a mis amados padres, quienes han sido mis mentores y me han brindado su apoyo moral, además de que con su esfuerzo y dedicación me han podido ofrecer todos los recursos necesarios para llegar a este punto de mi preparación académica.

También va dedicado a una persona muy especial en mi vida que a formado parte de este proceso y me a dado la mano en momentos oportunos y necesarios, Mónica Barreros.

Finalmente quiero dedicar este trabajo de titulación a mi querida hermana quien ha sido mi motivación emocional día a día y me ha brindado su apoyo incondicional en cada etapa de mi vida.

## **AGRADECIMIENTO**

Quiero agradecer a Dios quien ha derramado la fortuna y bendición de llegar a esta etapa de mi vida.

A la Universidad Indoamérica y a todos sus docentes quienes han dedicado su tiempo en compartir sus conocimientos y experiencia para nutrir mi formación académica y personal, principalmente al PhD. Ignacio Ayala Chauvin y el Dr. Pedro Escudero Villa quienes me han acompañado en la elaboración de este proyecto y han sabido guiarme con fe y paciencia, nutriéndome con su vasto conocimiento lo que ha permitido la culminación de este proyecto.

Finalmente, quiero agradecer a todas las personas que me han brindado su apoyo para llegar a esta etapa de mi vida.

## Tabla de Contenido

Resumen Ejecutivo .....	14
Abstract .....	15
CAPITULO I .....	16
INTRODUCCIÓN .....	16
Problematización.....	18
Antecedentes .....	19
Justificación .....	20
Objetivo general:.....	21
Objetivos Específicos:.....	21
CAPITULO II .....	22
METODOLOGÍA .....	22
Área de estudio.....	22
Enfoque .....	22
Justificación de la metodología .....	22
Investigación Aplicada.....	22
Diseño del trabajo .....	23
Procedimiento para obtención y análisis de datos .....	24
Población y muestra .....	25
Hipótesis .....	25
Cronograma.....	25
CAPITULO III.....	28
DESARROLLO DE LA INVESTIGACIÓN .....	28
Descripción de la Unidad Educativa .....	28
Identificación del área de estudio y Organización de Datos.....	29
Desarrollo del Modelo.....	32
Análisis del comportamiento de datos.....	33
Correlación entre atributos .....	45
Clasificadores Utilizados.....	47

Máquinas de Soporte Vectorial .....	47
Evaluación de la Predicción .....	49
Arboles de Decisión Bosque Aleatorio .....	49
Evaluación de la Predicción .....	51
Regresión Logística.....	52
Evaluación de la Predicción .....	53
CAPITULO IV.....	54
Resultados y Discusión .....	54
Contraste con otras investigaciones .....	57
Verificación de la hipótesis.....	58
Componente Ambiental .....	60
CAPITULO V .....	61
CONCLUSIONES Y RECOMENDACIONES .....	61
Conclusiones: .....	61
Recomendaciones: .....	62
Bibliografía: .....	63
ANEXOS: .....	65

## Índice de Tablas

<b>Tabla 1.</b> Población y muestra .....	25
<b>Tabla 2.</b> Codificación de datos personales por estudiante .....	29
<b>Tabla 3.</b> Codificación de asignaturas de la Escuela “Isabel la Católica” .....	30
<b>Tabla 4.</b> Modelo de matriz de datos por paralelo en la Escuela “Isabel la Católica” .....	31
<b>Tabla 5.</b> Matriz de datos de la institución educativa definidas por variables .....	32
<b>Tabla 6.</b> Cálculo de Chi cuadrada .....	59

## Índice de Figuras

<b>Figura 1.</b> Bajo desempeño académico en los estudiantes de la Escuela de Educación Básica Isabel La Católica .....	18
<b>Figura 2.</b> Flujograma del Diseño del trabajo .....	23
<b>Figura 3.</b> Proceso de obtención y análisis de datos para el desarrollo del modelo predictivo ....	24
<b>Figura 4.</b> Cronograma para la elaboración del proyecto .....	26
<b>Figura 5.</b> Diagrama del cronograma para la elaboración del proyecto.....	27
<b>Figura 6.</b> Ubicación de la Escuela de Educación Básica “Isabel la Católica” .....	28
<b>Figura 7.</b> Histograma del número de estudiantes de la institución educativa dividida por sexo.	34
<b>Figura 8.</b> Histograma del número de estudiantes de la institución educativa dividido por la dirección domiciliaria dentro de la ciudad de Píllaro. ....	34
<b>Figura 9.</b> Histograma del número de estudiantes de la institución educativa dividido por el tipo de vivienda. ....	35
<b>Figura 10.</b> Histograma del número de estudiantes de la institución educativa dividido por condiciones económicas.....	36
<b>Figura 11.</b> Histograma de calificaciones de los estudiantes de la institución educativa en la asignatura de Ciencias Naturales.....	36
<b>Figura 12.</b> Histograma de calificaciones de los estudiantes de institución educativa en la asignatura de Educación Artística. ....	37
<b>Figura 13.</b> Histograma de calificaciones de los estudiantes de institución educativa en la asignatura de Educación Física. ....	38
<b>Figura 14.</b> Histograma de calificaciones de los estudiantes de la institución educativa en la asignatura de Estudios Sociales.....	38
<b>Figura 15.</b> Histograma de calificaciones de los estudiantes de la institución educativa en la asignatura de Inglés.....	39
<b>Figura 16.</b> Histograma de calificaciones de los estudiantes de la institución educativa en la asignatura de Lenguaje.....	39
<b>Figura 17.</b> Histograma de calificaciones de los estudiantes de la institución educativa en la asignatura de Matemática.....	40
<b>Figura 18.</b> Gráfica del comportamiento de datos por puntos en la asignatura de Ciencias Naturales vs años.....	41
<b>Figura 19.</b> Visualización del comportamiento de datos por puntos en la asignatura de Educación Artística vs años. ....	41
<b>Figura 20.</b> Visualización del comportamiento de datos por puntos en la asignatura de Educación Físicas vs años.....	42
<b>Figura 21.</b> Visualización del comportamiento de datos por puntos en la asignatura de Estudios Sociales vs años. ....	43

<b>Figura 22.</b> Visualización del comportamiento de datos por puntos en la asignatura Ingles vs años. .....	43
<b>Figura 23.</b> Visualización del comportamiento de datos por puntos en la asignatura Lenguaje vs años. ....	44
<b>Figura 24.</b> Visualización del comportamiento de datos por puntos en la asignatura Matemática vs años.....	45
<b>Figura 25.</b> Matriz de correlación de datos de los estudiantes de institución educativa. ....	46
<b>Figura 26.</b> Diagrama de correlación de datos de los estudiantes de institución educativa. ....	47
<b>Figura 27.</b> Mapeo de características y frontera de decisión entre 2 clases de datos con SVM..	48
<b>Figura 28.</b> Matriz de confusión con Maquinas de soporte vectorial del rendimiento académico de los estudiantes de la institución educativa. ....	49
<b>Figura 29.</b> Bosque Aleatorio vs Árbol de decisión.....	50
<b>Figura 30.</b> Matriz de confusión con Arboles Aleatorios del rendimiento académico de los estudiantes de la institución educativa. ....	51
<b>Figura 31.</b> Matriz de confusión con Regresión Logística del rendimiento académico de los estudiantes de la institución educativa. ....	53
<b>Figura 32.</b> Diagrama de confusión con Maquinas de soporte vectorial en el rendimiento académico de los estudiantes de la institución educativa. ....	55
<b>Figura 33.</b> Matriz de confusión después del entrenamiento con Maquinas de soporte Vectorial den el rendimiento académico de los estudiantes de la institución educativa. ....	56
<b>Figura 34.</b> Porcentaje de calificaciones del rendimiento académico de los estudiantes de la institución educativa.....	57
<b>Figura 35.</b> Distribución de la Chi cuadrado hasta su punto crítico.....	60

## Índice de anexos

<i>Anexo 1: Concentrados físicos de las calificaciones anuales de la Escuela de Educación Básica Isabel la Católica los periodos de 2013 al 2015</i> .....	65
<b>Anexo 2:</b> Concentrados físicos de las calificaciones anuales de la Escuela de Educación Básica Isabel la Católica-Búsqueda de información por estudiante.....	66
<b>Anexo 3:</b> Búsqueda de concentrados físicos de los ciclos académicos faltantes de la Escuela de Educación Básica Isabel la Católica dentro del departamento de secretaria y vicerrectorado. ...	67
<b>Anexo 4:</b> Vista frontal de la fachada de la Escuela de Educación Básica Isabel la Católica, ubicada en el cantón Pillaro.....	68
<b>Anexo 5:</b> Concentrados físicos e informes de las calificaciones anuales de la Escuela de Educación Básica Isabel la Católica en años anteriores. ....	69
<b>Anexo 6:</b> Búsqueda de concentrados físicos faltantes en la base de datos, de las calificaciones anuales de la Escuela de Educación Básica Isabel la Católica en años anteriores.....	70
<b>Anexo 7:</b> Modelo de ficha socioeconómica por alumno de la Escuela de Educación Básica Isabel la Católica la que se extrajo información por estudiante. ....	71
<b>Anexo 8:</b> Formato digital de la ficha socioeconómica utilizada en la Escuela de Educación Básica Isabel la Católica hoja número 1. ....	72
<b>Anexo 9:</b> Formato digital de la ficha socioeconómica utilizada en la Escuela de Educación Básica Isabel la Católica hoja número 2. ....	73
<b>Anexo 10:</b> Líneas de Código 1 .....	74
<b>Anexo 11:</b> Líneas de Código 2 .....	75
<b>Anexo 12:</b> Líneas de Código 3 .....	76
<b>Anexo 13:</b> Líneas de Código 3 .....	77
<b>Anexo 14:</b> Líneas de Código 4 .....	78
<b>Anexo 15:</b> Líneas de Código 5 .....	79
<b>Anexo 16:</b> Líneas de Código 6 .....	80
<b>Anexo 17:</b> Líneas de Código 7 .....	81
<b>Anexo 18:</b> Líneas de Código 8 .....	82
<b>Anexo 19:</b> Líneas de Código 9 .....	83

**UNIVERSIDAD TECNOLÓGICA INDOAMÉRICA**  
**FACULTAD DE INGENIERÍA Y TECNOLOGÍAS DE LA INFORMACIÓN**  
**Y COMUNICACIÓN**  
**CARRERA DE INGENIERÍA INDUSTRIAL**

**TEMA:**

BIG DATA COMO HERRAMIENTA PARA EL ANÁLISIS DEL DESEMPEÑO ACADÉMICO EN LA EDUCACIÓN

**Autor:** Chucuri Real Boris Gabriel.

**Tutor:** PhD. Ayala Chauvin Manuel Ignacio.

### **Resumen Ejecutivo**

El presente proyecto de titulación, se enfoca en calcular el nivel de aprendizaje de la escuela de educación Básica Isabel La Católica con tres modelos clasificadores utilizados en Python que permitieron determinar los puntos febles en la educación de la escuela. Para lograr este análisis se utilizó 3 modelos en donde intervinieron: árboles aleatorios, regresión logística y máquinas de soporte vectorial. Estos modelos, fueron probados y entrenados con datos reales de varios ciclos académicos contando con más de dos mil datos desde el ciclo académico 2013 hasta el ciclo académico 2021. Además, dentro de cada modelo se añadieron características relacionadas a la socioeconomía del estudiante relacionadas al género, vivienda y economía. Finalmente, el modelo que se destaca de esta aplicación es el modelo con máquinas de soporte vectorial, el cual obtuvo un mejor porcentaje de precisión, con el 96,06% evidenciando resultados prometedores con un porcentaje del 56% de calificaciones muy buenas que representan el desempeño académico de la escuela de educación básica y es replicable en otros contextos educativos.

**Descriptores:** big data, desempeño académico, educación, máquinas de soporte vectorial

**UNIVERSIDAD TECNOLÓGICA INDOAMÉRICA**  
**FACULTAD DE INGENIERÍA Y TECNOLOGÍAS DE LA INFORMACIÓN**  
**Y COMUNICACIÓN**  
**CARRERA DE INGENIERÍA INDUSTRIAL**

**TEMA:**

“BIG DATA AS A TOOL FOR THE ANALYSIS OF ACADEMIC PERFORMANCE IN EDUCATION”

**Autor:** Chucuri Real Boris Gabriel.

**Tutor:** PhD. Ayala Chauvin Manuel Ignacio.

**Abstract**

The current research focuses on analyzing the academic performance at an educational level of elementary students. The purpose of this study is to identify performance patterns in students. This study aims to take action and predict the academic performance of the coming new students through the training program based on a classifier algorithm, which will provide tools and guidelines for teachers on the development of teaching. With the premise of poor academic performance at basic education levels; this research focuses on creating a predictive model of academic performance to determine weaknesses within education. The analysis presents three models with random-tree- graphs, logistic regression, and support vector machines. It is worth saying that the models of this study were tested and trained with real data from several academic cycles. In addition, socio-economic status and grading records of students had to be gathered to get reliable data. To conclude, the deployment of the proposal implies handling support vector machines, which leads to high percentage levels of accuracy. The model shows satisfactory results that present the academic

performance of basic education schools; furthermore, the model is replicable in other educational contexts.

**Keywords:** academic performance, big data, education, support vector

## CAPITULO I

### INTRODUCCIÓN

Durante años, se conoce al Big data como una herramienta de gran utilidad para el manejo de datos (Hernández-Leal, Duque-Méndez y Moreno-Cadavid 2017); se menciona que es lo último en tendencia con la aparición de nuevas tecnologías (Sakti 2018), además de que el manejo y buen uso de esta herramienta, tiene un sin fin de beneficios en muchos campos (Enriquez ,Carolina - Segura, Angela - Tovar 2013), pero ¿Qué es el Big Data?, Gartner define al Big Data como: un gran volumen o variedad de información que demanda formas costeables e innovadoras de procesamiento de información y permiten ideas extendidas, toma de decisiones y automatización del proceso (Salgado 2014).

Actualmente, la utilización de datos se ha convertido en un principio que permite llegar cada vez a más personas a través de muchos medios (Salgado 2014). El uso de las nuevas tecnologías en compañía con el Big data, han promovido múltiples ventajas (Fernández-Morales y Vallejo-Casarín 2014) y pueden potenciar la gestión educativa y el aprovechamiento de tecnologías digitales (Niemann y Pisla 2021) ; se menciona que las nuevas modalidades de estudio, en compañía con estas tecnologías tienen mayores ventajas para el acercamiento al estudiante (Escobar Borja y Mercado Pérez 2019); además que esto ha permitido un cambio notable en varios paradigmas de la educación tradicional (Sakti 2018), sobre todo en países que buscan la adaptabilidad

de los nuevos requerimientos de la sociedad moderna con la educación (Fernández-Morales y Vallejo-Casarín 2014).

Sin embargo, una de las acciones más comunes con el uso de nuevas tecnologías, es el desecho de una gran cantidad de datos (Sakti 2018); debido a que los métodos informáticos tradicionales, no permiten procesarlos y ser aprovechados dentro del proceso enseñanza-aprendizaje (Bonami 2020).

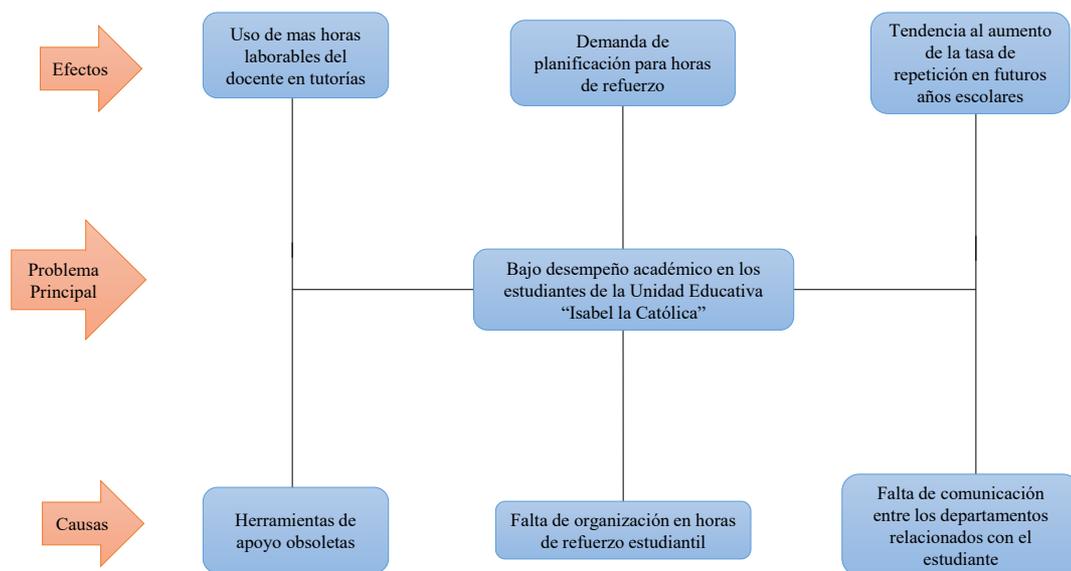
Por otro lado, en varios países los métodos del proceso de enseñanza-aprendizaje se han visto en constante cambio debido a factores externos que han puesto en desafío la adaptabilidad de la educación (Puyol Moreno 2014); además, se menciona que el método tradicionalista utilizado en muchas instituciones académicas carecía del uso de herramientas tecnológicas llevadas por un buen método de enseñanza (Enriquez ,Carolina - Segura, Angela - Tovar 2013).También, el sistema tradicional de educación, principalmente en países latinoamericanos, asegura que el proceso de desarrollo aprendizaje se ha basado en factores relacionados con el desempeño y experiencia del docente, mas no en un método y herramientas de apoyo(Salgado 2014). Es así como en Ecuador las nuevas modalidades de educación implementadas de manera forzada en los último años como la virtual, han crecido lentamente a diferencia de otros países (Escobar Borja y Mercado Pérez 2019); que ya poseían la experiencia de un acercamiento e implementación de la educación virtual, con el uso de herramientas de apoyo de que permiten aprovechar mejor los recursos informáticos, y brindar una mejor experiencia al estudiante (Fernández-Morales y Vallejo-Casarín 2014) .

Ahora bien, la importancia de analizar datos yace con la necesidad de tomar buenas decisiones y proporcionar mejores resultados ante cualquier proceso sin importar el campo de estudio (Bonami 2020); para ello el análisis predictivo es una de las técnicas más utilizadas, permitiendo basarse en datos obtenidos con anterioridad y predecir un resultado futuro con la ayuda de datos anteriores (Centeno Martín-Romero 2020); es así como el comportamiento del desempeño académico estudiantil se puede predecir para poder tomar mejores decisiones, fundamentándose en la relación entre la información obtenida en el pasado, para obtener un resultado del comportamiento

futuro (Sakti 2018), teniendo como objetivo plantear una buena estrategia y mejorar u optimizar la situación actual dentro del proceso enseñanza-aprendizaje.

## Problematización

El bajo desempeño académico es uno de los principales problemas con el que constantemente muchas instituciones académicas se enfrentan a lo largo de los años (Uribe-Enciso y Carrillo-García 2014), generalmente en países latinoamericanos esta problemática se ha podido identificar por causas tanto internas como externas, relacionadas con el estado salud, situación socioeconómica, estados individuales a cada estudiante, disfuncionalidad familiar y/o estados relacionado a múltiples insuficiencias en el sistema educativo (Uribe-Enciso y Carrillo-García 2014), al igual que las herramientas de apoyo, que influyen en el desempeño académico del estudiante (Enriquez, Carolina - Segura, Angela - Tovar 2013).



**Figura 1.** Bajo desempeño académico en los estudiantes de la Escuela de Educación Básica Isabel La Católica

Elaborado por: Chucuri, Boris (2022).

Por ello, en la figura 1 bajo un árbol de problemas se detalla la problemática dentro de la Unidad Educativa “Isabel la Católica “ actualmente ubicada en Píllaro Provincia de Tungurahua; en el que se menciona las causas evidentes de la problemática dentro de la unidad de educación básica, al igual que los efectos que provoca el bajo desempeño académico de la unidad educativa, por ello se describe que tres de las principales

causas de la problemática se centran en: la utilización de herramientas de apoyo obsoletas, la falta de comunicación entre departamentos relacionados directamente con el estudiante y la falta de organización en horas de refuerzo académico.

## **Antecedentes**

El análisis del Big data es el recurso más importante para entender, mejorar y potenciar la educación (Sakti 2018), por ello el análisis Big data y sus herramientas son tecnologías que brindan soporte y gestión al desarrollo de nuevas metodologías en el proceso de enseñanza – aprendizaje (Niemann y Pisla 2021). Además de que estas tecnologías se basan en comprender todo el proceso para poder evaluar y optimizar una mejora continua al entorno educativo buscando una mejor acción futura (Fernández-Morales y Vallejo-Casarín 2014), como una mejor planificación o un plan de estudios actualizado con el fin de alcanzar la adaptabilidad educativa y proporcionar un apoyo a todo el personal (Niemann y Pisla 2021), además de poder identificar logros al igual que patrones en los estudiantes, facilitando la toma de acciones en la gestión educativa del alumnado (Niemann y Pisla 2021).

Es así que con ayuda del Big data análisis, es posible manejar y construir un nuevo sistema de educación más actualizado y poder medir las aptitudes de cada estudiante (Hernández-Leal, Duque-Méndez y Moreno-Cadavid 2017); además de que con programas de apoyo se facilita la relación de los estudiantes y una carrera que funcione con un empleo proyectado en años posteriores, relacionado con sus habilidades académicas (Sakti 2018). Puesto que el sistema creado por medio de la analítica en Big data se puede modelar a la realidad de diferentes modalidades en educación (Centeno Martín-Romero 2020), así como una comunicación más amigable con el alumnado para mantener nuevas posibilidades de educación, ante la busca de mejores metodologías educativas en donde la tecnología participa como principal herramienta (Bonami 2020) , esto ayuda a los educadores y otras partes interesadas a que el aprendizaje se mantenga en un mejor control, ante el proceso de educación para la mejora continua y el conocimiento (Bonami 2020).

## Justificación

La **importancia** del presente proyecto radica en fortalecer la facilidad del cambio en diversos modelos educativos y proporcionar un mejor enfoque del comportamiento de los estudiantes dentro de la Unidad Educativa “Isabel la Católica” para desarrollo de aprendizaje, teniendo una herramienta de análisis que permita predecir, así como observar los datos obtenidos de un grupo de estudiantes con la finalidad de facilitar la toma de decisiones relacionados al desempeño académico para una mejora continua.

Además, el **impacto** en la investigación afecta de manera positiva en el uso de herramientas tecnológicas que permiten seguir actualizando el modelo y organización de datos como parte de la gestión de la Unidad educativa, siendo los estudiantes al igual que la Unidad educativa los principales **beneficiarios**. También, se analizó de forma más objetiva el desempeño académico de un grupo de estudiantes en distintos ciclos académicos, hoy por hoy muchos países como china consideran que la mejor adaptabilidad en la sociedad está acompañada de las tecnologías dentro de la educación, que pueden apoyar al desarrollo académico como lo es con la utilización de plataformas de apoyo y Big data que permiten obtener proyección para una mejora de sus resultados y objetivos (Fernández-Morales y Vallejo-Casarín 2014).

Es así como la investigación se considera **factible**, puesto que la Unidad educativa cuenta con una base de datos en desarrollo en la que se almacenan la mayor parte de la información de cada estudiante, a la que se pudo tener acceso y poder realizar la extracción de la información y la realización de entrevistas no estructuradas con el cuerpo administrativo de la Unidad Educativa, además de utilizar herramientas de Open Source o también llamadas de acceso abierto. La **utilidad** de esta investigación está estrechamente relacionada con el estudio y comportamiento del desempeño académico por medio de un previo análisis con el fin de facilitar las acciones que dentro del desarrollo del proceso de gestión educativa y mejorarlo con el uso de herramientas tecnológicas.

### **Objetivo general:**

- Analizar el desempeño académico de la Escuela de Educación Básica “Isabel La Católica” mediante herramientas de Big Data.

### **Objetivos Específicos:**

- Analizar la información que dispone la unidad educativa mediante un registro de Excel.
- Procesar la información mediante técnicas de filtrado de datos.
- Realizar el modelo predictivo utilizando las herramientas de Big Data y técnicas de aprendizaje supervisado, con el fin de generar predicciones de los estudiantes de la institución educativa.
- Evaluar las predicciones de los estudiantes de la institución “Isabel La Católica” para el periodo 2016-2021 como resultado del modelo predictivo.

## **CAPITULO II**

### **METODOLOGÍA**

#### **Área de estudio**

**Línea de Investigación:** Software y gestión de las tecnologías de la información

**Campo:** Ingeniería Industrial

**Área:** Bases de datos

**Aspecto:** Desempeño académico

**Objeto de estudio:** Análisis del desempeño académico en la educación

#### **Enfoque**

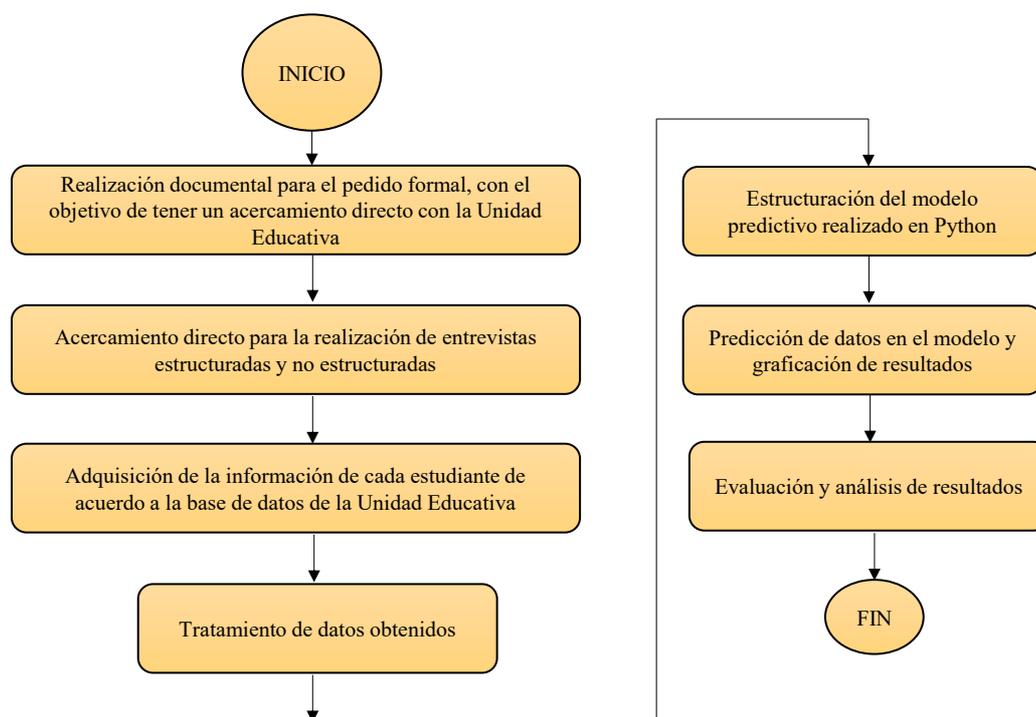
El enfoque del presente proyecto fue cuantitativo debido a que se trabajó con los datos percibidos de la Escuela de Educación Básica “Isabel la Católica” del cantón Píllaro, obteniéndolos por medio de un acercamiento directo a la Escuela ya mencionada con una solicitud formal y varias entrevistas no estructuradas para obtener la información necesaria, por ello el proyecto se basó en el tratamiento de datos para determinar la situación actual y poder obtener la proyección respectiva del desempeño académico estudiantil.

#### **Justificación de la metodología**

#### **Investigación Aplicada**

El presente proyecto se considera con una metodología de campo y bibliográfico documental, ya que se basó en la aplicación de un modelo predictivo con fuentes relacionadas al tema, además de que fue necesario el acercamiento directo a la Escuela de Educación Básica en repetidas ocasiones para la obtención detallada de la información del desempeño académico y la comunicación directa con el cuerpo administrativo.

## Diseño del trabajo



**Figura 2.** Flujograma del Diseño del trabajo

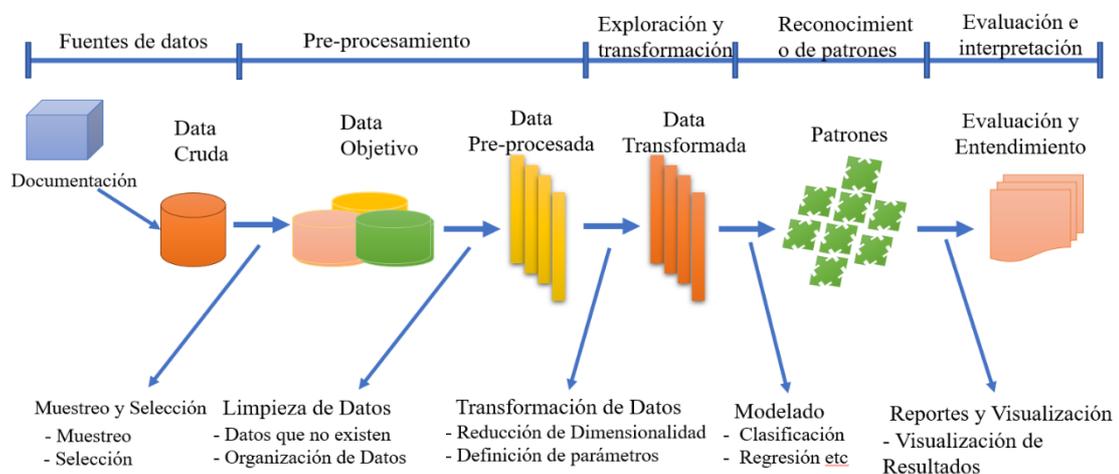
Elaborado por: Chucuri, Boris (2022).

Dentro del desarrollo de la investigación se planteó la gestión necesaria para proceder con el pedido formal y acceso a toda la información documentada de forma física y digital que pueda ser de utilidad ( notas y promedios académicos por ciclos , fichas socioeconómicas y otros ) además se solicitó el acceso libre a la base de datos a utilizar , en este caso se presenta la información almacenada un documentos de Excel , posteriormente se obtuvo un tratamiento de datos que facilite la utilización de los mismos , determinando promedios por ciclos académico y asignatura , y finalmente se

estructuró el modelo predictivo mediante el software Python que permitió visualizar las gráficas de los resultados respectivos mediante Júpiter.

### Procedimiento para obtención y análisis de datos

En este proyecto se pudo obtener los datos necesarios para la investigación, dentro de la Escuela de Educación Básica “Isabel La Católica” ubicada en el cantón Píllaro, es así que mediante la revisión de la base de datos se pudo determinar la organización en la cual se encontraban distribuidos los diferentes cursos por aula, nivel y asignatura, así como ciclo académico, además de que se pudo tener acceso a los datos de cada estudiante previo a una solicitud formal a la escuela para el posterior tratamiento de datos, y tener entrevistas previas no estructuradas con el personal administrativo que brindó información de importancia, por otro lado se pudo tener acceso a las fichas de cada estudiante donde se detalla información personal y familiar. Es así como con la integración de todos estos datos se pudo determinar una organización propicia luego de una purificación de datos para la implementación del modelo predictivo con la herramienta Python.



**Figura 3.** Proceso de obtención y análisis de datos para el desarrollo del modelo predictivo

Elaborado por: Chucuri, Boris (2022).

## **Población y muestra**

En el presente trabajo se logró identificar la población con un total de 2101 estudiantes, pero para facilidad del entrenamiento del algoritmo se establece la muestra de 2000 estudiantes detallado en la Tabla 1. Dicha información fue recolectada por acercamiento y comunicación con el cuerpo administrativo de la institución, todos estos datos pertenecientes al periodo 2014 – 2021.

**Tabla 1.** Población y muestra

<b>Usuarios</b>	<b>Número</b>	<b>Porcentaje</b>
Muestra de estudio	2071	100 %
Entrenamiento del algoritmo	1998	96,57%

Elaborado por: Chucuri, Boris (2022).

## **Hipótesis**

El análisis Big data determinará el nivel de desempeño de la Escuela de Educación Básica “Isabel la Católica”.

## **Cronograma**

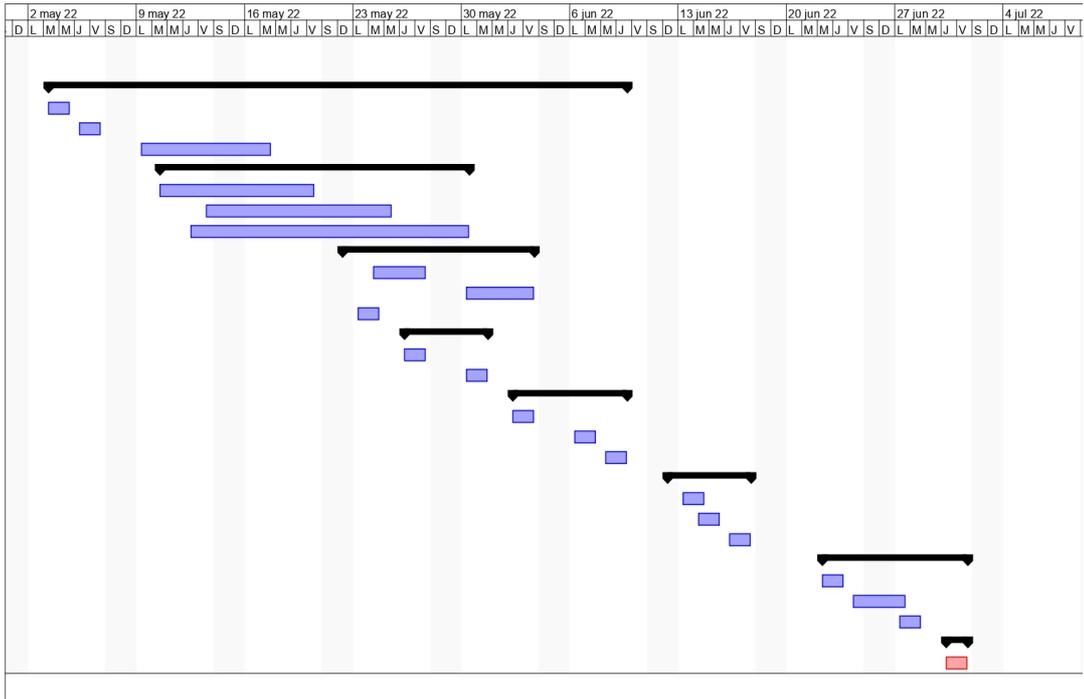
Para la realización del presente proyecto se plantea el cronograma de actividades detallado en la figura 4, con el objeto de cumplir con cada una de las asignaciones necesarias que llevan a la realización del proyecto y en la consecución de los objetivos planteados. Para lo cual se destaca un total de 29 actividades.

		Nombre	Duracion	Inicio	Terminado
1		<b>Acercamiento a la Unidad Educativa</b>	<b>28 days?</b>	<b>03/05/22 8:00</b>	<b>09/06/22 17:00</b>
2		Realización documental para pedido formal	2 days	03/05/22 8:00	04/05/22 17:00
3		Realización de entrevistas al personal Administrativo	2 days	05/05/22 8:00	06/05/22 17:00
4		Adquisición de información necesaria	7 days?	09/05/22 8:00	17/05/22 17:00
5		<b>Preprocesamiento</b>	<b>14 days</b>	<b>10/05/22 13:00</b>	<b>30/05/22 13:00</b>
6		Muestreo y selección de datos	8 days	10/05/22 13:00	20/05/22 13:00
7		Limpeza y organización de datos	8 days	13/05/22 13:00	25/05/22 13:00
8		Definición de parámetros	12 days	12/05/22 13:00	30/05/22 13:00
9		<b>Transformación de Data</b>	<b>10 days</b>	<b>22/05/22 8:00</b>	<b>03/06/22 17:00</b>
10		Exportación de datos al software utilizado	4 days	24/05/22 8:00	27/05/22 17:00
11		Realización de diagramas e histogramas	5 days	28/05/22 8:00	03/06/22 17:00
12		Análisis del comportamiento de datos	2 days	22/05/22 8:00	24/05/22 17:00
13		<b>Reconocimiento de patrones</b>	<b>4 days</b>	<b>26/05/22 8:00</b>	<b>31/05/22 17:00</b>
14		Visualización de patrones por la data	2 days	26/05/22 8:00	27/05/22 17:00
15		Definición de parametros a evaluar	2 days	28/05/22 8:00	31/05/22 17:00
16		<b>Realización del primer modelo</b>	<b>6 days</b>	<b>02/06/22 8:00</b>	<b>09/06/22 17:00</b>
17		Definición de parámetros	2 days	02/06/22 8:00	03/06/22 17:00
18		Entrenamiento del modelo	2 days	04/06/22 8:00	07/06/22 17:00
19		Evaluación de resultados	2 days	08/06/22 8:00	09/06/22 17:00
20		<b>Realización del segundo modelo</b>	<b>5 days</b>	<b>12/06/22 8:00</b>	<b>17/06/22 17:00</b>
21		Definición de parámetros	2 days	12/06/22 8:00	14/06/22 17:00
22		Entrenamiento del modelo	2 days	14/06/22 8:00	15/06/22 17:00
23		Evaluación de resultados	2 days	16/06/22 8:00	17/06/22 17:00
24		<b>Realización del tercer modelo</b>	<b>8 days</b>	<b>22/06/22 8:00</b>	<b>01/07/22 17:00</b>
25		Definición de parámetros	2 days	22/06/22 8:00	23/06/22 17:00
26		Entrenamiento del modelo	2 days	24/06/22 8:00	27/06/22 17:00
27		Evaluación de resultados	2 days	26/06/22 8:00	28/06/22 17:00
28		<b>Comparación de modelos</b>	<b>2 days</b>	<b>30/06/22 8:00</b>	<b>01/07/22 17:00</b>
29		Evaluación de resultados	2 days	30/06/22 8:00	01/07/22 17:00

**Figura 4.** Cronograma para la elaboración del proyecto

Elaborado por: Chucuri, Boris (2022).

En la figura 5 se observa el proceso de elaboración del proyecto en el cual se detallan las actividades previas a desarrollarse en correspondencia al cronograma de actividades expuesta en la figura 4, en consecuencia, facilitara el control y seguimiento del avance del proyecto.



**Figura 5.** Diagrama del cronograma para la elaboración del proyecto

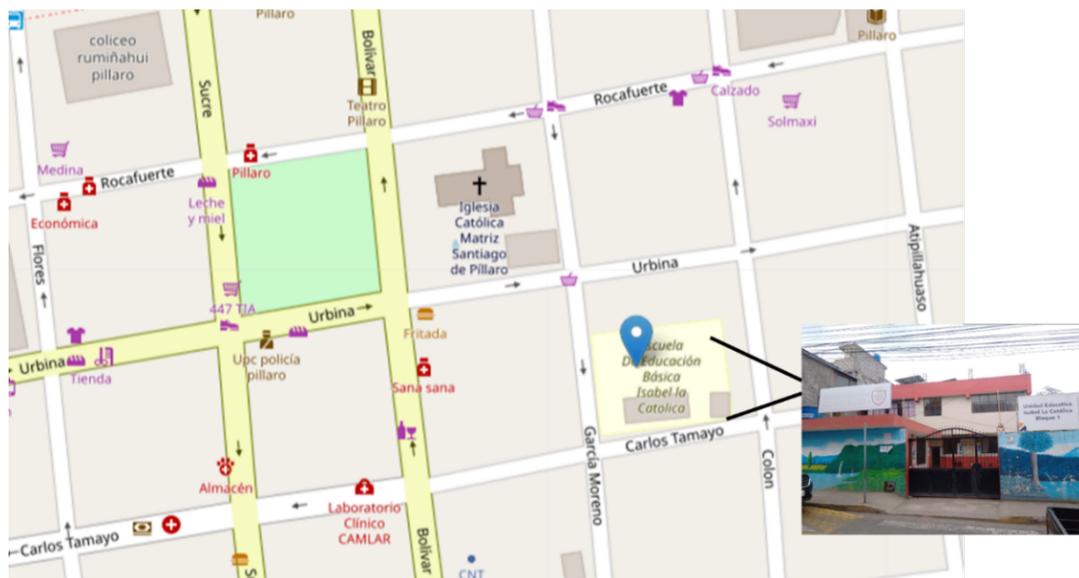
Elaborado por: Chucuri, Boris (2022).

## CAPITULO III

### DESARROLLO DE LA INVESTIGACIÓN

El presente proyecto se propone el análisis del desempeño académico de la Escuela de Educación Básica “Isabel La Católica” mediante herramientas de Big Data conjuntamente aplicando modelos tres modelos diferentes de machine learning para la predicción del rendimiento académico de los estudiantes.

#### Descripción de la Unidad Educativa



**Figura 6.** Ubicación de la Escuela de Educación Básica “Isabel la Católica”

**Fuente:** (GoogleMaps 2022)

La Escuela de Educación Básica Isabel la Católica se encuentra ubicada en la provincia de Tungurahua cantón “Píllaro”, calles García Moreno y Carlos Tamayo como se puede observar en la figura, la misma que se dedica a la prestación de servicios de formación académica en nivel primario.

Dentro de las jornadas matutina y vespertina que se encuentran en funcionamiento en la Escuela de Educación Básica “Isabel la Católica” se extrajeron datos de la jornada matutina para el análisis en donde se hallan cursos de segundo a séptimo grado divididos en diversos paralelos compuestos entre 20 a 31 estudiantes, además que las materias de estudio que conforman la malla de la escuela son: Ciencias Naturales, Matemáticas, Educación Física, Educación artística, Ciencias Sociales, Ingles, Lengua y literatura

### **Identificación del área de estudio y Organización de Datos**

El presente estudio está enfocado en el análisis y predicción del desempeño académico de los estudiantes de la Escuela de Educación Básica “Isabel la Católica”, mediante el uso de herramientas tecnológicas de Big Data y utilización de clasificadores que permitirán determinar el nivel de desempeño académico.

Dentro de la Escuela de Educación Básica “Isabel la Católica”, se tomaron datos de 3 cortes de estudiantes, en donde los datos de los periodos iniciales se encuentran en las promociones; 2013-2014, 2014-2015, 2015-2016 y finalizan en las promociones; 2018-2019, 2019-2020 y 2020-201 respectivamente. Es así como se organizaron las calificaciones de los estudiantes por asignatura, al igual que datos personales extraídos de las fichas socioeconómicas de cada estudiante de forma cuantitativa como se muestra en las siguientes tablas.

**Tabla 2.** Codificación de datos personales por estudiante

<b>DIRECCIÓN</b>	<b>CÓDIGO</b>
Norte	0
Sur	1
Este	2
Oeste	3
<b>ECONOMÍA</b>	
Regular	0
Buena	1
Muy Buena	2
Excelente	3
<b>VIVIENDA</b>	
Propia	0
Arrendada	1

Prestada	2
Anticresis	3
Con préstamo	4
<b>SEXO</b>	
Hombre	1
Mujer	0

Elaborado por: Chucuri, Boris (2022).

En la tabla 1, se presenta la codificación utilizada en el algoritmo, teniendo numeraciones del 1 al 3, en el caso de la dirección, si el estudiante vive al: norte de la ciudad tiene el valor de 0, Sur de la ciudad tiene el valor de 1, Este de la ciudad tiene el valor de 2 y Oeste de la ciudad tiene el valor de 3, de igual forma la economía tiene numeraciones en donde 0 representa una economía regular en el estudiante, 1 representa una economía buenas, muy buena se le asigna el valor de 2 y excelente un valor de 3. También se numeran aspectos de la vivienda, en donde si la vivienda es: propia se la asigna un valor de 0, si es arrendada se le asigna un valor de 1, si es anticresis s le asigna un valor de 3, si es con préstamo un valor de 4, por último, se asignan valores relacionados al sexo del estudiante, si es hombre se le asigna el valor de 1 y si es mujer se l asigna el valor de 0.

**Tabla 3.** Codificación de asignaturas de la Escuela “Isabel la Católica”

Notas	Asignatura
P1	Ciencias Naturales
P2	Educación Artística
P3	Educación Física
P4	Ciencias Sociales
P5	Ingles
P6	Lenguaje
P7	Matemática

Elaborado por: Chucuri, Boris (2022).

Analizando los datos recolectados en la Escuela de Educación Básica Isabel La Católica, se procedió a organizarlos dentro una matriz en la cual se colocaron únicamente las calificaciones y los códigos asignados anteriormente, en aspectos relacionados al estudiante como: sexo, dirección, economía, vivienda y todas las asignaturas percibidas en el plantel.

**Tabla 4.** Modelo de matriz de datos por paralelo en la Escuela “Isabel la Católica”

	SEXO	DIR	ECO	VIV	P1	P2	P3	P4	P5	P6	P7
E1	0	0	2	0	9,79	9,66	9,92	9,87	9,62	9,87	9,86
E2	1	2	2	0	9,59	9,16	9,76	9,58	9,32	9,64	9,57
E3	0	3	1	1	9,54	9,53	9,41	9,48	9,24	9,39	9,44
E4	0	1	2	0	8,72	8,88	8,67	8,74	8,17	8,46	8,4
E5	1	2	1	1	8,79	8,58	8,98	8,74	8,77	8,81	8,81
E6	1	1	1	0	8,94	8,96	9,09	9,07	8,77	8,83	8,59
E7	1	3	1	1	7,58	7,9	8,26	7,97	7,96	7,1	7,29
E8	0	2	1	1	8,22	8,45	8,33	8,57	8,16	7,67	7,75
E9	0	3	2	0	8,89	9,02	8,91	8,87	8,96	8,67	8,56
E10	0	1	1	0	8,99	9,05	9,04	8,94	8,74	8,81	8,84
E11	1	3	2	2	9,52	9,62	9,41	9,39	9,35	9,31	9,41
E12	1	3	2	0	9,91	9,79	9,66	9,82	9,69	9,9	9,9
E13	0	0	1	0	9,65	9,75	9,79	9,72	9,51	9,55	9,52
E14	1	3	1	1	8,51	8,66	8,96	8,57	8,56	8,26	8,15
E15	1	0	1	0	9,83	9,97	9,7	9,87	9,69	9,94	9,87
E16	1	3	2	0	8,94	9	8,88	8,97	8,69	8,91	9,02
E17	0	3	1	1	8,59	8,38	8,48	8,24	8,76	8,05	8,22
E18	0	3	2	0	8,28	8,25	8,56	8,33	8,32	7,98	7,89
E19	0	2	2	0	9,98	9,78	9,68	9,86	9,77	9,93	9,98
E20	0	3	1	1	9,75	9,76	9,68	9,72	9,68	9,66	9,52
E21	1	2	1	1	7,55	7,82	8,23	7,96	8	8,03	7,74
E22	1	1	1	0	9,78	9,66	9,81	9,89	9,64	9,78	9,8
E23	1	0	2	0	9,55	9,51	9,4	9,46	9,3	9,37	9,39
E24	0	0	1	0	9,38	9,03	8,87	9,49	9	9,25	9
E25	0	0	1	0	9,56	9,56	9,53	9,43	9,65	9,43	9,61
E26	0	1	1	0	9,78	9,81	9,76	9,78	9,44	9,86	9,74
E27	0	1	2	0	9,66	8,87	9,89	9	9,75	7,55	7,74
E28	0	0	3	1	9,49	9,53	9,68	9,46	9,66	9,86	8,56
E29	1	0	2	1	8,59	8,25	8,38	9,75	9,55	7,82	9,81
E30	1	2	2	0	7,55	8	8,38	7	8,1	7,41	7,2
E31	1	1	2	0	8,38	9,78	8,25	8,48	9,75	9,37	9

Elaborado por: Chucuri, Boris (2022).

Así es como se pudo relacionar todos los grados de estudio de la Escuela de Educación Básica Isabel la Católica por paralelo, de tal modo que fue necesario agruparlos nuevamente en una sola matriz de datos en el que se pueda especificar el periodo o año de inicio del año lectivo, así como el año de finalización asignadas por las

variables: AI, AF respectivamente, al igual que variables relacionadas con el grado de estudio y promedio de calificación final, definidas por GE y PROMEDIO.

**Tabla 5.** Matriz de datos de la institución educativa definidas por variables

ALUMNO	SEXO	DIR	ECO	VIVI	P1	P2	P3	P4	P5	P6	P7	GE	AI	AF
E1	1	2	1	1	7,93	7,87	8,08	8,07	9,33	8,60	8,34	2B	2013	2014
E2	1	1	1	1	9,54	9,64	9,62	9,68	9,44	8,67	8,45	2B	2013	2014
E3	1	2	1	1	7,81	8,62	8,46	8,90	8,28	8,45	9,45	2B	2013	2014
E4	1	1	2	1	9,24	9,47	9,44	9,38	9,18	8,54	8,45	2B	2013	2014
E5	0	2	1	0	7,97	8,39	8,66	8,92	9,10	8,34	8,34	2B	2013	2014
E6	1	1	2	1	8,51	8,44	8,37	9,08	9,49	9,56	8,45	2B	2013	2014
E7	0	2	1	0	9,26	8,96	9,23	9,07	9,23	8,45	8,56	2B	2013	2014
E8	1	2	2	1	8,23	8,07	8,48	8,89	9,49	8,96	9,67	2B	2013	2014
E9	0	2	1	0	8,84	8,60	8,96	8,39	8,95	8,56	8,97	2B	2013	2014
E10	1	1	2	1	8,97	9,14	9,29	9,30	8,98	8,45	8,12	2B	2013	2014
E11	0	1	1	0	8,90	9,18	9,07	8,89	9,07	9,45	9,45	2B	2013	2014
E12	1	1	2	1	9,32	8,90	9,41	9,44	9,13	8,56	8,34	2B	2013	2014
E13	0	0	1	0	7,96	8,03	8,45	8,52	9,52	9,67	8,56	2B	2013	2014
E14	1	0	1	4	9,57	9,68	9,58	9,21	9,04	8,34	9,67	2B	2013	2014
E15	0	0	0	4	9,68	9,69	9,56	9,54	9,48	9,67	8,56	2B	2013	2014
E16	1	1	1	0	8,69	8,41	8,81	9,53	8,94	8,45	9,56	2B	2013	2014
E17	0	0	0	4	8,50	8,27	8,73	9,26	8,31	9,23	8,34	2B	2013	2014
E18	0	1	1	1	9,16	9,40	9,04	9,52	8,28	8,97	8,56	2B	2013	2014
E19	1	0	2	2	7,24	7,43	8,09	8,83	8,99	8,45	9,67	2B	2013	2014
E20	0	1	1	1	8,69	8,88	8,95	8,96	9,23	8,40	9,67	2B	2013	2014
E21	1	2	2	2	9,53	9,47	9,52	9,44	9,82	8,10	8,43	2B	2013	2014
E22	0	1	1	1	9,01	8,79	8,92	8,41	9,63	8,73	8,70	2B	2013	2014
E23	0	0	0	2	9,13	8,63	9,11	9,24	9,18	9,28	8,45	2B	2013	2014

Elaborado por: Chucuri, Boris (2022).

Mediante la elaboración de la tabla 4 se pudo clasificar de una forma más idónea los datos obtenidos de la Escuela de Educación Básica Isabel la Católica, con el objetivo de plantear la pertenencia de las calificaciones de cada estudiante, por su respectivo ciclo académico, con la finalidad de analizar el comportamiento del desempeño académico en el transcurso del tiempo.

### **Desarrollo del Modelo**

### **Python 3.10.5**

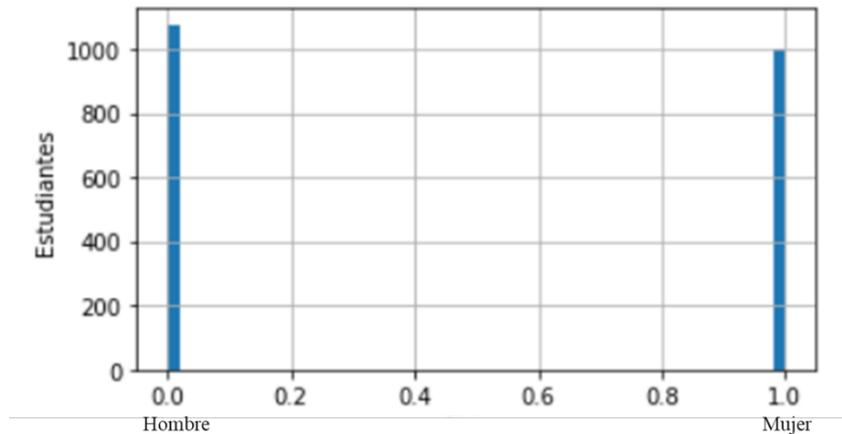
En el presente estudio, se utilizó el software Python versión 3.10.5 como herramienta de Big Data para poder determinar la predicción de los datos obtenidos, herramienta de código abierto que es utilizada para la solución de problemas de manera sencilla y eficiente por su versatilidad y lenguaje áreas (Challenger Pérez, Díaz Ricardo y Becerra García 2014), además de poseer un gran procesamiento de buenas señales en áreas como la matemática, estadística, así como disponibilidad en librerías de visualización en ramas similares (Challenger Pérez, Díaz Ricardo y Becerra García 2014). Por otra parte Python cuenta con una estrecha relación con scikit-learn o módulo de aprendizaje automático que se complementa de manera sencilla con su interfaz lo que lo hace ideal en algoritmos de análisis de datos y estadística (Barupal y Fiehn 2019).

### **Jupyter**

Además, un recurso de mucha importancia, en el que se enfocó la visualización del código fue Jupyter, que es una aplicación que facilita la interacción de trabajo vinculados a proyectos relacionados al análisis, estadística e innovación tecnológica, por ello se considera un cuaderno digital de mucha importancia para el mejor entendimiento cuando se trabaja con sentencias lógicas, también tiene una interfaz amigable con el usuario y contiene librerías de trabajo muy completas que lo hacen ideal para el desarrollo de proyectos.

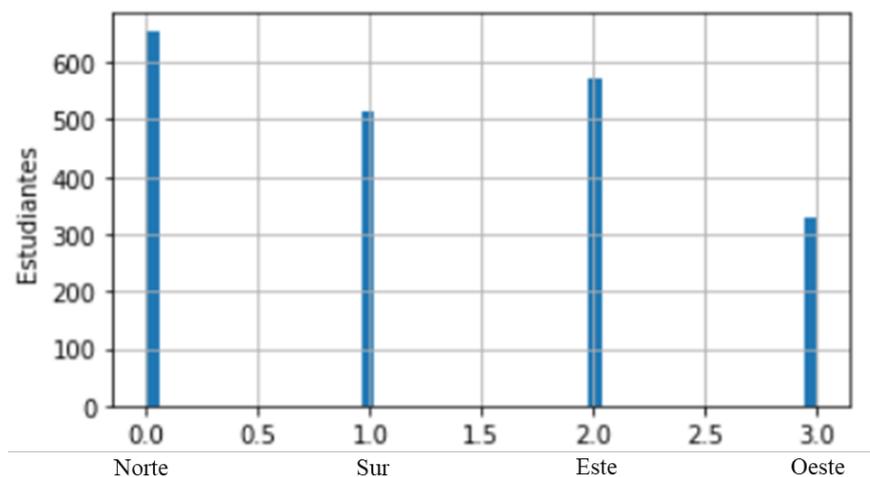
### **Análisis del comportamiento de datos**

Dentro de la elaboración del modelo se realizó un proceso de preparación de datos dentro del software Python en el que se analizó el comportamiento de los datos de cada estudiante por medio de la realización de algunos diagramas e histogramas que se presentan a continuación.



**Figura 7.** Histograma del número de estudiantes de la institución educativa dividida por sexo. Elaborado por: Chucuri, Boris (2022).

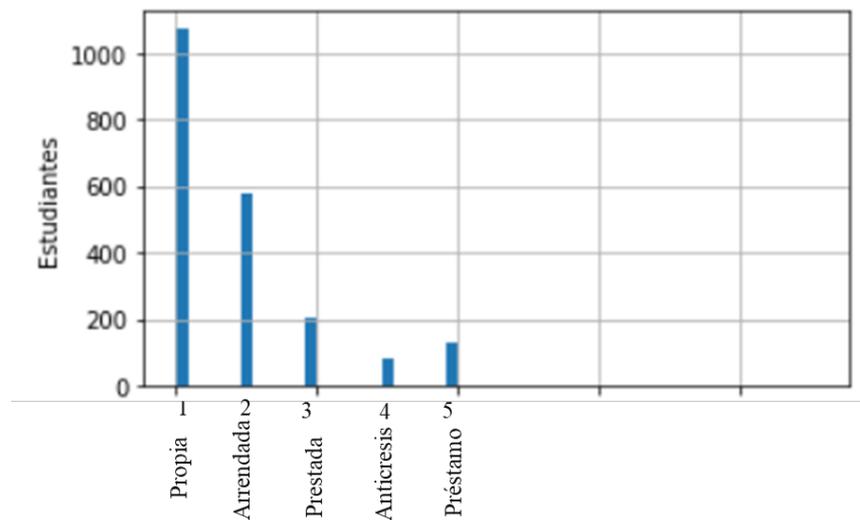
En la figura 7 se presenta la clasificación del número de estudiantes presentes en la escuela de educación Básica Isabel La Católica, divididos por sexo en el que se observa que el mayor número de estudiantes están conformados por hombres con un total de 1000, existiendo una diferencia mínima de 100 estudiantes con el valor de las mujeres que centra un valor de 900. Esta diferencia mínima puede ser a casusa de que a lo largo de los años los centros educativos han optado por una educación mixta ya que en años anteriores se repercutía a una educación solo masculina.



**Figura 8.** Histograma del número de estudiantes de la institución educativa dividido por la dirección domiciliaria dentro de la ciudad de Píllaro.

Elaborado por: Chucuri, Boris (2022).

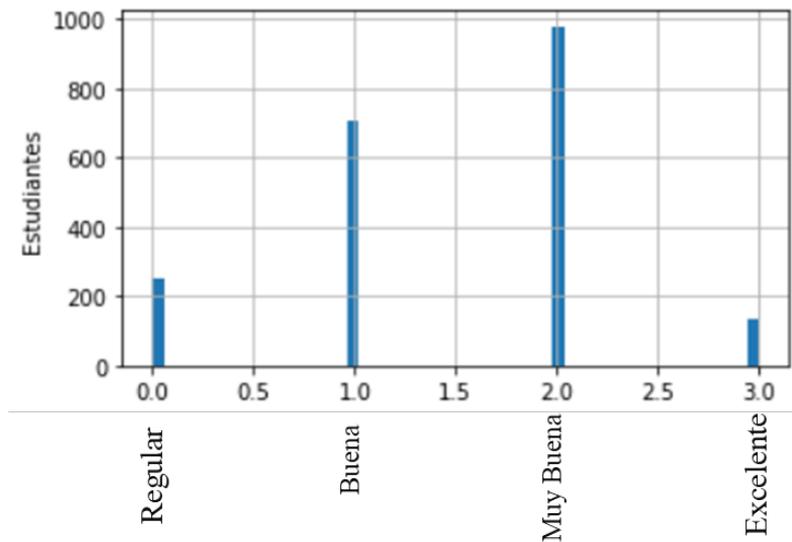
En la Figura 8 se observa la distribución de todos los estudiantes de la Escuela de Educación Básica Isabel la Católica, en 4 sectores del cantón Píllaro, en donde se puede mencionar que el mayor número de estudiantes que asisten a la unidad educativa pertenecen al Norte, mientras que el menor número de estudiantes pertenecen al Sur.



**Figura 9.** Histograma del número de estudiantes de la institución educativa dividido por el tipo de vivienda.

Elaborado por: Chucuri, Boris (2022).

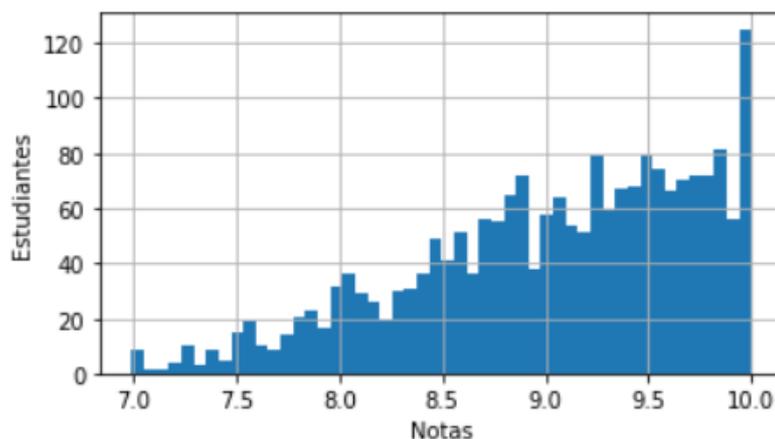
En la figura 9 se puede observar la clasificación del número de estudiantes de la Escuela de Educación Básica Isabel la Católica, dividido en 5 grupos, de acuerdo al tipo de vivienda de cada estudiante, en consecuencia, se menciona que el mayor número de estudiantes reside en una vivienda propia, además que los grupos más pequeños mencionan residir en viviendas prestadas, o arrendadas.



**Figura 10.** Histograma del número de estudiantes de la institución educativa dividido por condiciones económicas

Elaborado por: Chucuri, Boris (2022).

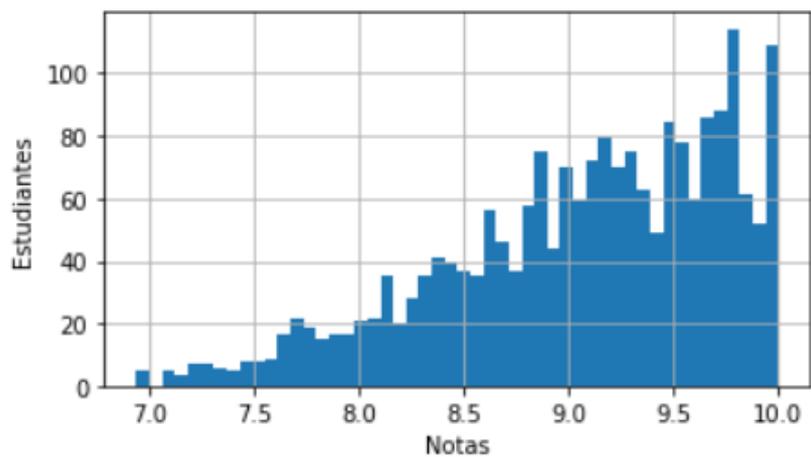
En la figura 10 se observa la clasificación de los estudiantes de la Escuela de Educación Básica Isabel la Católica, dividida en 4 grupos relacionados a su nivel económico, en donde se puede mencionar que el número más grande de estudiantes tiene un nivel económico muy bueno y bueno, mientras los grupos más pequeños de estudiantes tienen un nivel económico regular y excelente.



**Figura 11.** Histograma de calificaciones de los estudiantes de la institución educativa en la asignatura de Ciencias Naturales.

Elaborado por: Chucuri, Boris (2022).

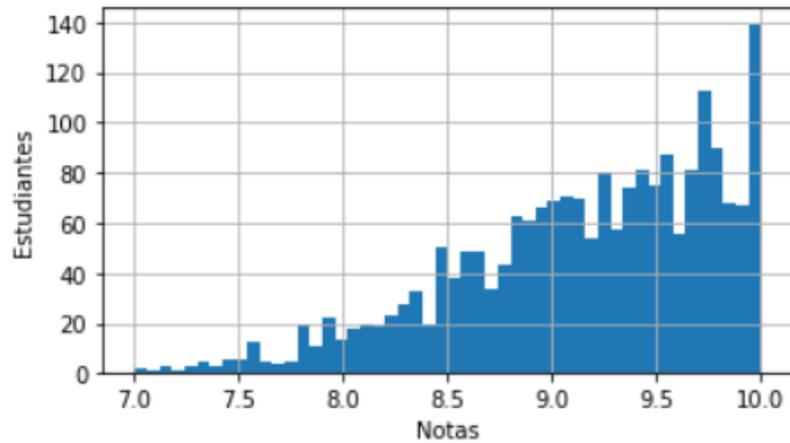
En la figura 11, se puede visualizar el comportamiento de los datos, en la asignatura de Ciencias Naturales en calificaciones de 7 a 10 puntos en diferentes grupos, por ello, se puede analizar que existe un pequeño número de estudiantes de la Escuela Isabel la Católica se encuentran con calificaciones sobresalientes mayores a 9 puntos, mientras que un pequeño grupo de estudiantes tienen la mayor parte de calificaciones no menores a 7 puntos.



**Figura 12.** Histograma de calificaciones de los estudiantes de institución educativa en la asignatura de Educación Artística.

Elaborado por: Chucuri, Boris (2022).

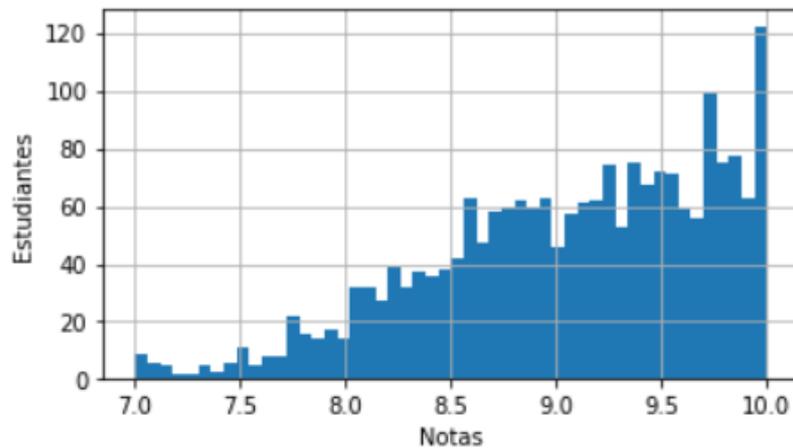
En la figura 12, se puede visualizar el comportamiento de los datos, en la asignatura de Educación Artística en calificaciones de 7 a 10 puntos en diferentes grupos, por tanto, se puede analizar que existe un pequeño grupo de estudiantes que tienen calificaciones menores a 10 puntos y mayores a 9,5 puntos, por otra parte, se observa que un pequeño grupo de estudiantes tienen calificaciones deficientes que son menores a 7 puntos.



**Figura 13.** Histograma de calificaciones de los estudiantes de institución educativa en la asignatura de Educación Física.

Elaborado por: Chucuri, Boris (2022).

En la figura 13, se puede visualizar el comportamiento de los datos, en la asignatura de Educación Física en calificaciones de 7 a 10 puntos en diferentes grupos, por ello, se interpreta que un grupo de estudiantes tiene calificaciones mayores a 8,5 puntos y un grupo de estudiantes tiene calificaciones sobresalientes de 10 puntos.

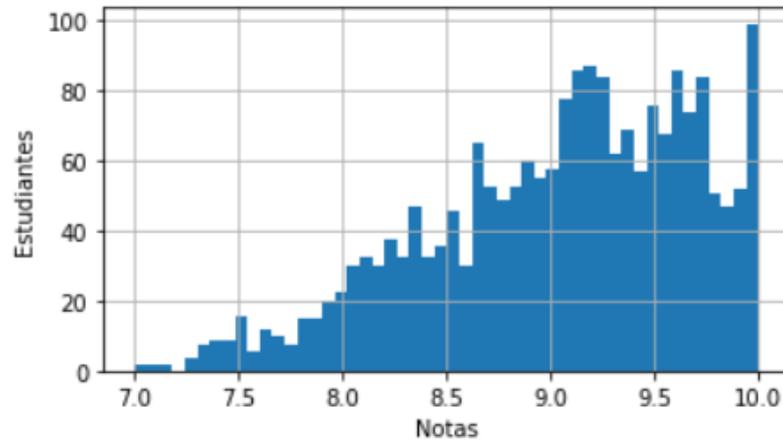


**Figura 14.** Histograma de calificaciones de los estudiantes de la institución educativa en la asignatura de Estudios Sociales.

Elaborado por: Chucuri, Boris (2022).

En la figura 14, se puede visualizar el comportamiento de los datos, en la asignatura de Estudios Sociales en calificaciones de 7 a 10 puntos en diferentes grupos, por tanto, se menciona que el mayor número de estudiantes posee calificaciones buenas entre 8,5

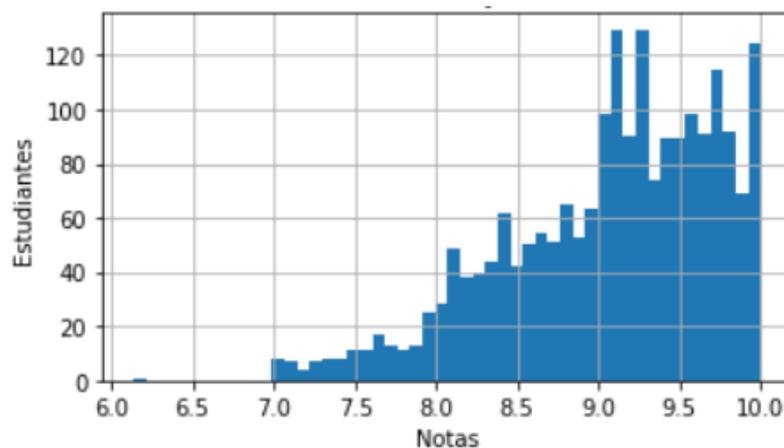
y 9,5 puntos, mientras que existe un pequeño grupo de estudiantes con calificaciones de 7 y 10 puntos.



**Figura 15.** Histograma de calificaciones de los estudiantes de la institución educativa en la asignatura de Inglés.

Elaborado por: Chucuri, Boris (2022).

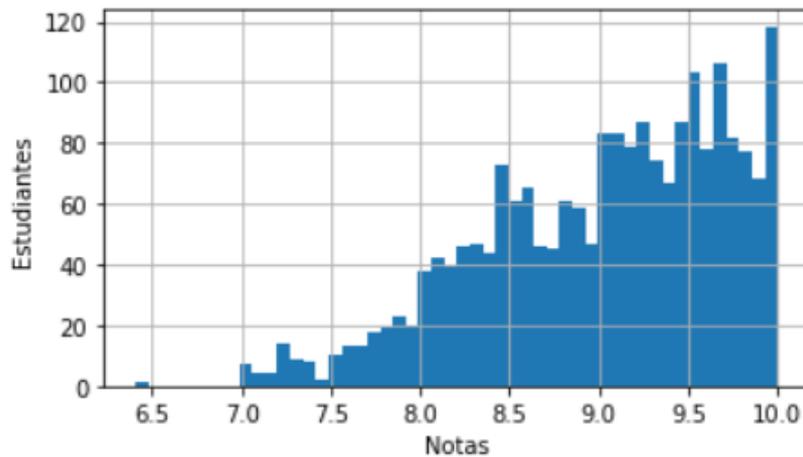
En la figura 15, se puede visualizar el comportamiento de los datos, en la asignatura de Inglés en calificaciones de 7 a 10 puntos en diferentes grupos, por tanto, se encontró la existencia de un mayor sesgo entre las calificaciones de todos los estudiantes además de que existe un pequeño grupo de estudiantes con calificaciones mayores a 7 puntos y menores a 7,5 puntos.



**Figura 16.** Histograma de calificaciones de los estudiantes de la institución educativa en la asignatura de Lenguaje.

Elaborado por: Chucuri, Boris (2022).

En la figura 16, se puede visualizar el comportamiento de los datos, en la asignatura de Lenguaje en calificaciones de 7 a 10 puntos en diferentes grupos, por tanto, se encontró estudiantes con calificaciones deficientes menores a puntos, sin embargo, el mayor número de estudiantes tiene calificaciones muy buenas, mayores a los 8,5 puntos.

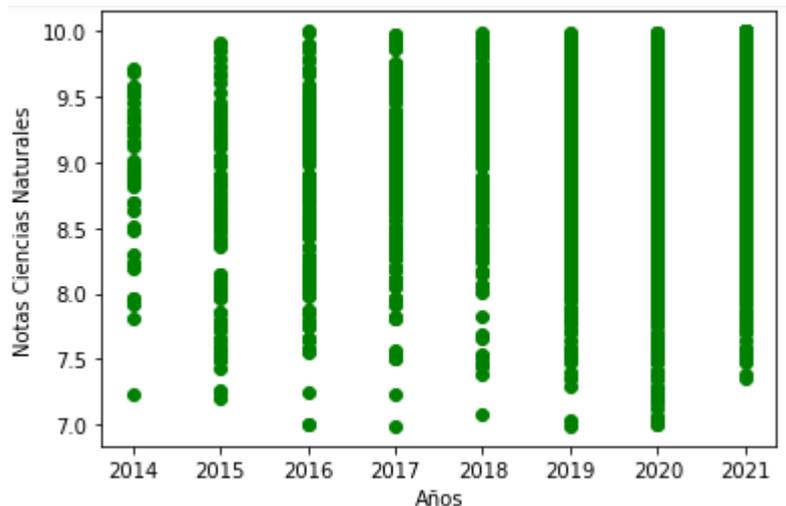


**Figura 17.** Histograma de calificaciones de los estudiantes de la institución educativa en la asignatura de Matemática.

Elaborado por: Chucuri, Boris (2022).

En la figura 17, se puede visualizar el comportamiento de los datos, en la asignatura de Matemática en calificaciones de 7 a 10 puntos en diferentes grupos, por tanto, se pudo identificar que existen estudiantes con calificaciones deficientes mejores a los 7 puntos, además de que el mayor grupo de estudiantes tiene calificaciones mayores a los 8,5 puntos.

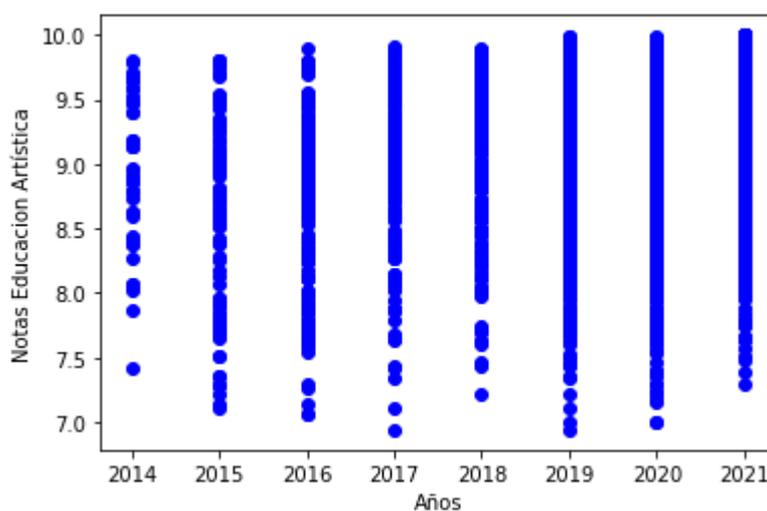
Asimismo, se estudió el comportamiento de los datos con la realización de diversos diagramas por asignatura, tomando como base las calificaciones de los estudiantes de 7 a 10 puntos, y los periodos académicos transcurridos a partir del año 2014 hasta el año 2021.



**Figura 18.** Gráfica del comportamiento de datos por puntos en la asignatura de Ciencias Naturales vs años.

Elaborado por: Chucuri, Boris (2022).

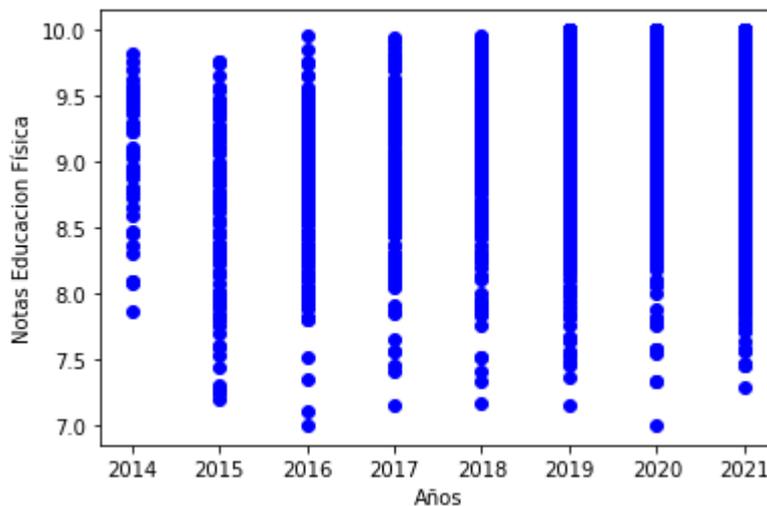
En la figura 18, se puede visualizar el comportamiento de las calificaciones de los estudiantes en la asignatura de Ciencias Naturales en 8 ciclos académicos, por ello se identificó que durante los periodos académicos de los años 2016,2017,2019 y 2020 existían estudiantes con notas regulares, correspondientes a 7 puntos mientras que en el año 2021 se denota una ligera mejora en el rendimiento de los estudiantes con notas desde 7,5 puntos , por otro lado se pudo observar que el mejor ciclo académico en la asignatura fue en el 2014 con la mayoría de estudiantes con notas superiores los 7,5 puntos.



**Figura 19.** Visualización del comportamiento de datos por puntos en la asignatura de Educación Artística vs años.

Elaborado por: Chucuri, Boris (2022).

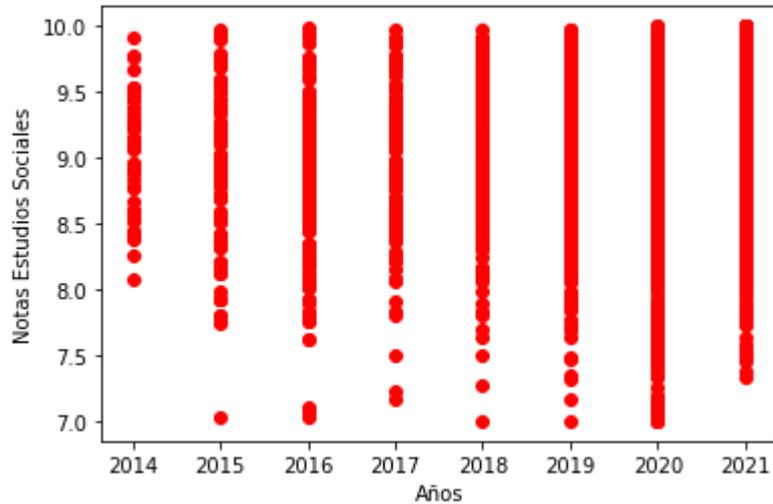
En la figura 19, se puede visualizar el comportamiento de las calificaciones de los estudiantes en la asignatura de Educación Artística en 8 ciclos académicos, en consecuencia, se identificó que existen grupos de estudiantes con calificaciones menores a 7 puntos en los años 2017 y 2019, además de que el mejor año en rendimiento académico de los estudiantes fue en el año 2014.



**Figura 20.** Visualización del comportamiento de datos por puntos en la asignatura de Educación Física vs años.

Elaborado por: Chucuri, Boris (2022).

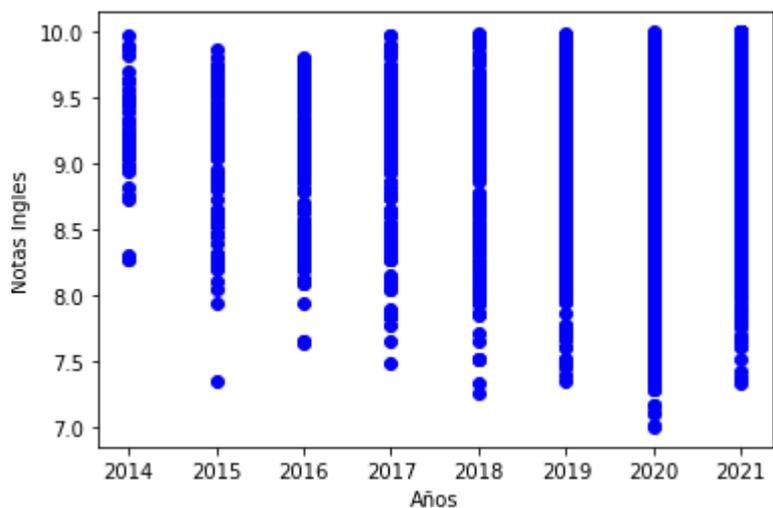
En la figura 20, se puede visualizar el comportamiento de las calificaciones de los estudiantes en la asignatura de Educación Física en 8 ciclos académicos, en donde se pudo identificar que existen grupos de estudiantes con calificaciones menores a 7 puntos en los años 2016 y 2020, además, se identificó que en el año 2014 existen un mejor rendimiento con calificaciones superiores a 8 puntos en la mayor parte de estudiantes.



**Figura 21.** Visualización del comportamiento de datos por puntos en la asignatura de Estudios Sociales vs años.

Elaborado por: Chucuri, Boris (2022).

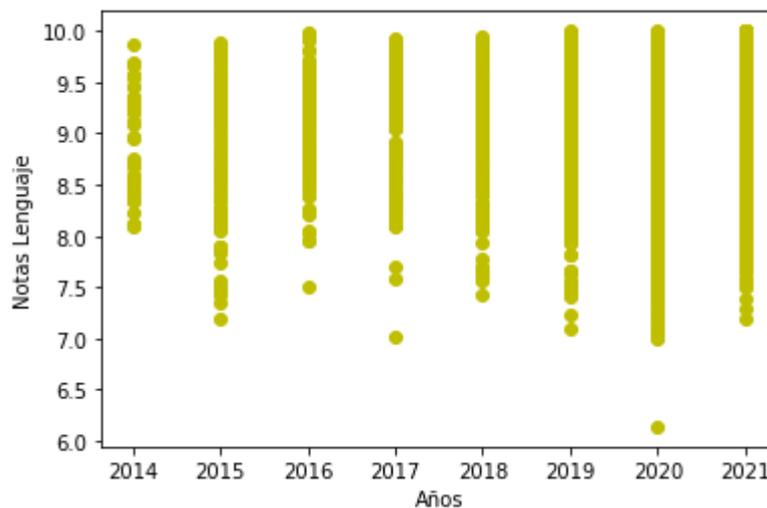
En la figura 21, se puede visualizar el comportamiento de las calificaciones de los estudiantes en la asignatura de Estudios Sociales en 8 ciclos académicos, en donde se pudo identificar que con excepción del año 2014 y 2021 las calificaciones de los estudiantes con el rendimiento más bajo se encontraban a partir de los 7 puntos, sin embargo durante el año 2017 se pudo identificar que la mayor parte de los estudiantes obtuvieron calificaciones mayores a 8 puntos, con un comportamiento similar pero no mejor al del año 2014.



**Figura 22.** Visualización del comportamiento de datos por puntos en la asignatura Inglés vs años.

Elaborado por: Chucuri, Boris (2022).

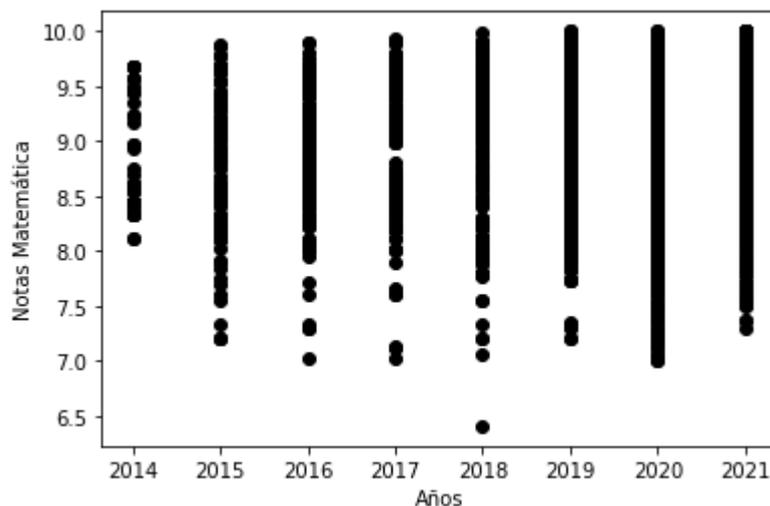
En la figura 22, se puede visualizar el comportamiento de las calificaciones de los estudiantes en la asignatura de Inglés en 8 ciclos académicos, por ello se pudo identificar que la mayoría de los estudiantes durante el año 2014, tiene calificaciones mayores a 8,5 puntos siendo el año con el mejor rendimiento, por otra parte durante el año 2020 se identificó que los estudiantes tienen calificaciones regulares desde los 7 puntos, con un comportamiento similar a los años 2018,2019 y 2021.



**Figura 23.** Visualización del comportamiento de datos por puntos en la asignatura Lenguaje vs años.

Elaborado por: Chucuri, Boris (2022).

En la figura 23, se puede visualizar el comportamiento de las calificaciones de los estudiantes en la asignatura de Lenguaje en 8 ciclos académicos, por ello se puede identificar que existen grupos de estudiantes con calificaciones deficientes entre los 6 y 6,5 puntos en el año 2020, además el comportamiento entre notas de 7 puntos y superiores es similar en casi todos los ciclos académicos, con excepción del año 2014 siendo el año con calificaciones superiores a 8 puntos.



**Figura 24.** Visualización del comportamiento de datos por puntos en la asignatura Matemática vs años.

Elaborado por: Chucuri, Boris (2022).

En la figura 24, se puede visualizar el comportamiento de las calificaciones de los estudiantes en la asignatura de Matemática en 8 ciclos académicos, así pues, se encontró estudiantes con calificaciones deficientes menores a 6,5 puntos en el año 2018, además de encontrar estudiantes con notas regulares en los años 2016,2017,2019,2020 y 2021, por otro lado, el año 2014 los estudiantes obtuvieron notas superiores a 8 puntos.

También, se realizó la correlación de atributos de todos los datos de los estudiantes, mediante el método de Pearson, utilizando parámetros como la covarianza entre estudiantes de forma que se pueda medir de manera cuantitativa los resultados mediante una matriz que se presenta a continuación.

### **Correlación entre atributos**

	SEXO	DIR	ECO	VIVI	P1	P2	P3
SEXO	1.000000	0.083239	0.029012	0.024995	-0.013121	-0.031699	-0.024842
DIR	0.083239	1.000000	0.078612	-0.018557	-0.029138	-0.019249	-0.007350
ECO	0.029012	0.078612	1.000000	-0.096282	0.030446	0.046410	0.035516
VIVI	0.024995	-0.018557	-0.096282	1.000000	-0.049431	-0.044446	-0.039452
P1	-0.013121	-0.029138	0.030446	-0.049431	1.000000	0.903928	0.288269
P2	-0.031699	-0.019249	0.046410	-0.044446	0.903928	1.000000	0.280768
P3	-0.024842	-0.007350	0.035516	-0.039452	0.288269	0.280768	1.000000
P4	-0.015484	-0.049761	0.043827	-0.034422	0.903708	0.870152	0.282770
P5	-0.031791	-0.019659	-0.018675	-0.035730	0.287344	0.262169	0.079934
P6	-0.019615	-0.031655	0.047837	-0.042681	0.774263	0.718902	0.215661
P7	0.004602	-0.026370	0.047699	-0.029575	0.800618	0.761486	0.225657

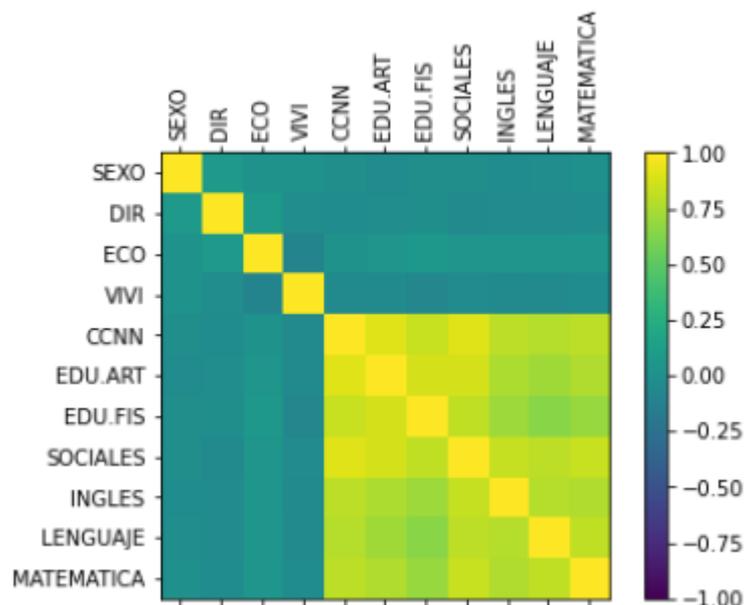
  

	P4	P5	P6	P7
SEXO	-0.015484	-0.031791	-0.019615	0.004602
DIR	-0.049761	-0.019659	-0.031655	-0.026370
ECO	0.043827	-0.018675	0.047837	0.047699
VIVI	-0.034422	-0.035730	-0.042681	-0.029575
P1	0.903708	0.287344	0.774263	0.800618
P2	0.870152	0.262169	0.718902	0.761486
P3	0.282770	0.079934	0.215661	0.225657
P4	1.000000	0.297012	0.801027	0.831056
P5	0.297012	1.000000	0.285892	0.277422
P6	0.801027	0.285892	1.000000	0.810643
P7	0.831056	0.277422	0.810643	1.000000

**Figura 25.** Matriz de correlación de datos de los estudiantes de institución educativa.

Elaborado por: Chucuri, Boris (2022).

Se pudo identificar mediante la matriz de correlación realizada en la figura 25, que la relación entre las variables como el sexo, dirección, economía y vivienda es muy baja o casi nula, entre las calificaciones de las diversas asignaturas de los estudiantes de la Escuela Isabel la Católica, por ello se observó que la relación de datos existe entre asignaturas como Ciencias Naturales y Estudios Sociales, con una correlación fuerte positiva entre ambas variables. Por ello, se presenta el diagrama de correlación entre todas las variables del estudio en las que se puede observar la relación por colores entre los atributos ya mencionados a continuación.



**Figura 26.** Diagrama de correlación de datos de los estudiantes de institución educativa.

Elaborado por: Chucuri, Boris (2022).

Por otro lado, para la preparación de datos se pudo definir los parámetros de clasificación de su comportamiento por medio de decisiones entre calificaciones de: (7 a 7.9), (8 a 8.9), (9 a 9.9) y 10, definiéndolas como calificaciones regulares, buenas, muy buenas y excelentes, respectivamente y en caso de ser menores a 7 puntos se categorizaron como deficientes.

## Clasificadores Utilizados

### Máquinas de Soporte Vectorial

Las máquinas de soporte vectorial SVM (Support Vector Machine), son modelos de aprendizaje automático y clasificación utilizados para la división entre 2 o más clases de datos por medio de fronteras de decisión (Betancourt 2005); con la ayuda de muy pocos datos de aprendizaje, estos datos son mapeados con un método kernel a un área dimensional, en la que visualmente se separan por una frontera que permite distinguir su clasificación (Betancourt 2005). Así pues, la conceptualización matemática en un modelo sencillo se describe a continuación.

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 = 0 \tag{1}$$

Generalizada en "P" dimensiones

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \dots \beta_p x_p = 0 \tag{2}$$

En donde

$$X = x_1, x_2 \dots x_p \tag{3}$$

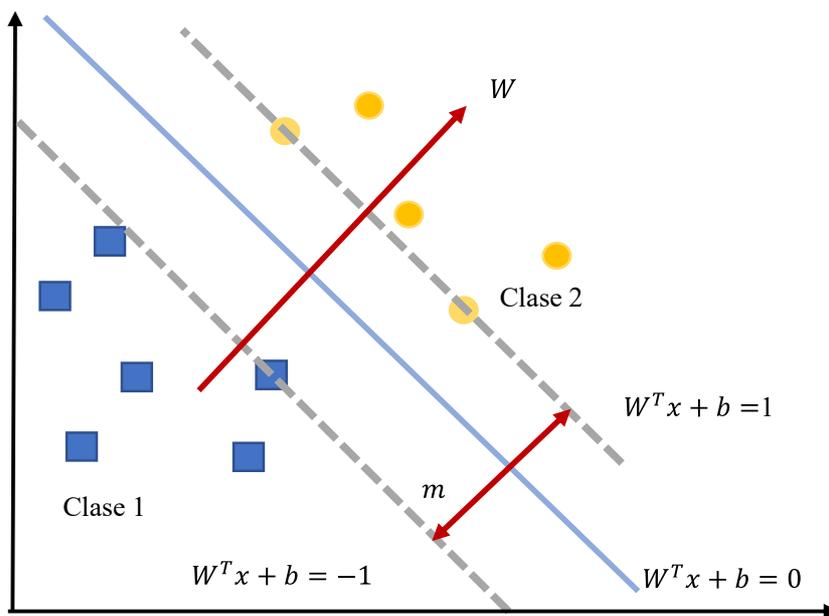
Al no satisfacer la ecuacion:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \dots \beta_p x_p < 0 \tag{4}$$

En donde

$\beta$  = número de dimensiones  
dependedientes

$X$  = punttos definidos por los vectores



**Figura 27.** Mapeo de características y frontera de decisión entre 2 clases de datos con SVM

**Fuente:** (Betancourt 2005)

## Evaluación de la Predicción

```
Matriz Confusion
[[147  0  4  0]
 [  0  2  3  0]
 [  3  2 232  0]
 [  2  0  0 191]

Presicion
0.966183574879227
Error
0.033816425120772986
Recuperacion
0.966183574879227
```

**Figura 28.** Matriz de confusión con Maquinas de soporte vectorial del rendimiento académico de los estudiantes de la institución educativa.

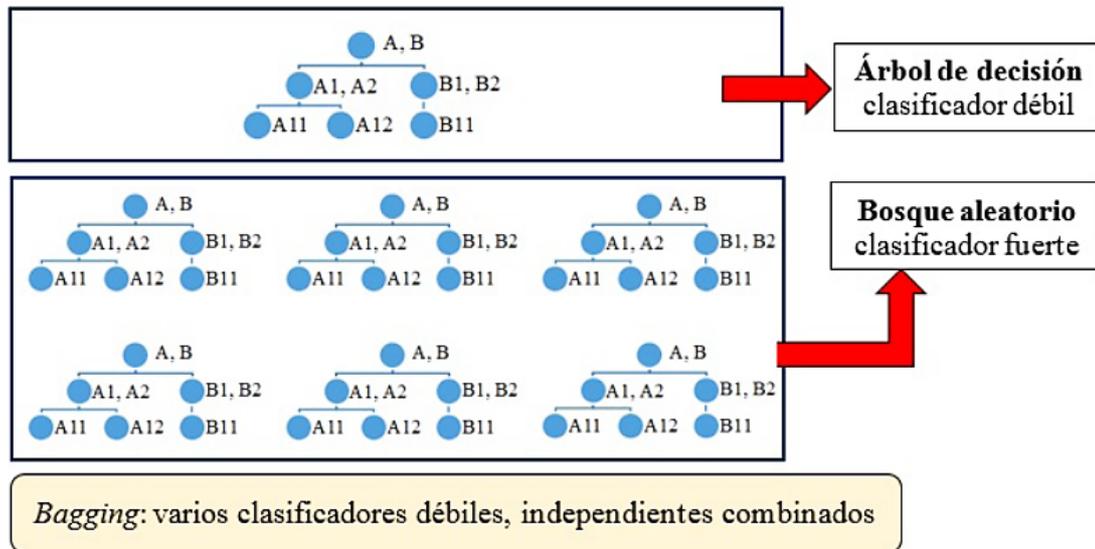
Elaborado por: Chucuri, Boris (2022).

En base a los antecedentes mencionados anteriormente, relacionados al funcionamiento de las máquinas de soporte vectorial, se realizó la predicción de los datos de los estudiantes en función de las métricas con las que funciona la máquina de soporte vectorial, así pues, se consiguió entrenar al algoritmo con el 99,8% de datos totales para poder obtener la matriz de confusión presentada como el resultado de la predicción con la utilización del 0,2% de datos totales. También, se pudo conseguir los resultados de la precisión, error y recuperación en el proceso de predicción con el cual funciona el algoritmo. Por ello, se puede establecer el resultado de 147 datos con predicción de un rendimiento académico “Bueno” con 4 datos considerados como “Muy bueno” que en realidad pertenecen a un rendimiento académico “Bueno” (falsos negativos). Además, en base al resultado de la matriz de confusión se obtiene una precisión del 96,6% y un error de 3,3% en la predicción en base al entrenamiento previo del algoritmo.

## Arboles de Decisión Bosque Aleatorio

Los árboles de decisión o Random Forest, es una técnica de clasificación dentro de la minería de datos; esta técnica está compuesta por un conjunto de árboles de decisión

que permiten manejar un conjunto de datos más grande y varias variables a la vez (Medina-Merino y Ñique-Chacón 2017). Es así como cada árbol de decisión compone un subconjunto de datos para determinar el análisis con el uso de todas las variables para la construcción del modelo (Medina-Merino y Ñique-Chacón 2017).



**Figura 29.** Bosque Aleatorio vs Árbol de decisión

**Fuente:** (Medina-Merino y Ñique-Chacón 2017)

## Evaluación de la Predicción

```
Matriz Confusion
[[140  0  3  0]
 [  0  0  7  0]
 [ 11  0 229  0]
 [  7  0  0 17]]
Presicion
0.9323671497584541
Error
0.03140096618357491
Recuperacion
0.9323671497584541
```

**Figura 30.** Matriz de confusión con Árboles Aleatorios del rendimiento académico de los estudiantes de la institución educativa.

Elaborado por: (Chucuri 2022).

Gracias al entrenamiento previo del algoritmo, se pudo obtener el resultado de la predicción del desempeño académico con el 0,2% de datos restantes de los estudiantes, así pues, la matriz de confusión obtenida, figura 30 , explica que 143 estudiantes tienen un rendimiento académico “Bueno” con 3 estudiantes con rendimiento académico “Muy Bueno” que en realidad debieron categorizarse como buenos (falsos positivos), así pues se visualiza el porcentaje de error en el algoritmo con 4,3% en margen de error y con una precisión en la predicción del 93% por ello identifica 17 estudiantes con un rendimiento académico “Regular”.

$$b = \tau - t \quad (5)$$

$$v = \frac{1}{R - 1} \quad (6)$$

$$\sum_{r=1}^R (T - t)^2 = R \quad (7)$$

*En donde:*

$b = \text{nodos aleatorios}$

$t = \text{nodo desicional}$

$\beta = \text{número de dimensiones}$

$t = \text{nodo desicional}$

$T = \text{Nodos totales}$

$R = \text{numero de variables dependientes}$

## **Regresión Logística**

La regresión logística, es un modelo que está formado por una o varias funciones lineales, en el que generalmente se utiliza para pronosticar posibles ocurrencias en intervalos entre 1 y 0 en función de las variables (Alba y Calle 2016); en consecuencia a ello, es necesario la construcción de una función matemática, que permita aplicarse al modelo con la variable real y la variable de interés para encontrar evento probable (Alba y Calle 2016).

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ en donde } i = 1, 2, \dots, n$$

*Entonces*

$$\text{Si } y = 1, \quad \varepsilon_i = 1 - \beta_0 - \beta_1 x_i \quad (9)$$

*Entonces*

$$\text{Si } y = 0, \quad \varepsilon_i = 1 - \beta_0 - \beta_1 x_i \quad (10)$$

*En donde*

$\beta = \text{número de dimensiones}$

$E = \text{variables dependientes}$

$n = \text{Probabilidad de evento de interés}$

$t = \text{nodo desicional}$

$T = \text{Nodos totales}$

## Evaluación de la Predicción

```
Matriz Confusion
[[139  0  2  2]
 [  0  1  6  0]
 [  5  0 235  0]
 [  2  0  0 22]]
Presicion
0.9589371980676329
Error
0.04347826086956519
Recuperacion
0.9589371980676329
```

**Figura 31.** Matriz de confusión con Regresión Logística del rendimiento académico de los estudiantes de la institución educativa.

Elaborado por: (Chucuri 2022).

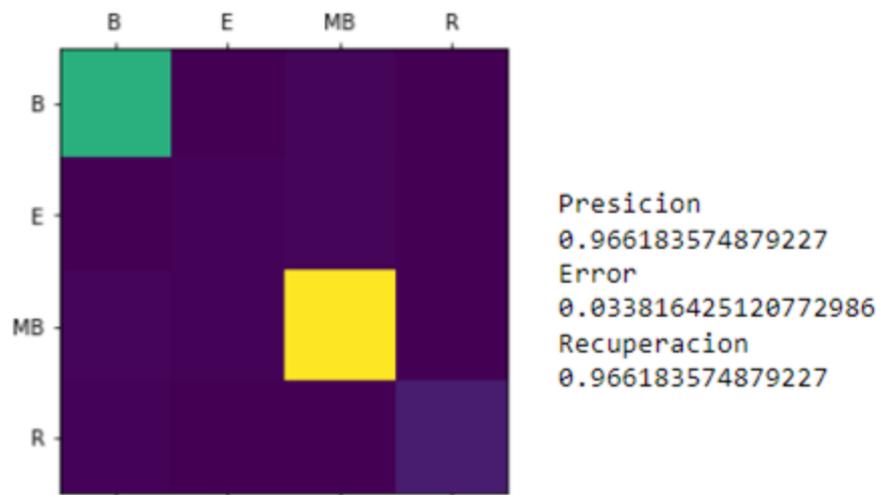
También, se pudo desarrollar una nueva predicción con Regresión Logística, después del entrenamiento previo del algoritmo y como resultado, se evidencia la matriz de confusión figura 31, así pues se obtuvo 139 estudiantes que predicen estar dentro de un grupo “Bueno” de desempeño académico, con 4 datos como falsos positivos que están catalogados como “Regulares” y “Muy Buenos”, sin embargo deberían clasificarse como “Buenos”, por ello se menciona que el porcentaje de precisión de la predicción del modelo tiene un porcentaje del 95% y porcentaje de error con 4,3%.

## CAPITULO IV

### Resultados y Discusión

En el presente proyecto se realizó el análisis de desempeño académico de la Escuela de Educación Básica Isabel la Católica mediante Python, así pues, se utilizó el método de Pearson para encontrar la correlación de datos en los diversos ciclos académicos, también se pudo desarrollar diversos histogramas en los que se pudo visualizar el comportamiento en las características de los estudiantes y diferentes asignaturas. Además, previo a la preparación de datos se utilizaron tres clasificadores distintos; Maquinas de soporte vectorial, Arboles aleatorios y Regresión logística, para predecir el desempeño académico en cada uno de ellos, así pues se consideraron métricas de precisión, exactitud y sensibilidad establecidos en el entrenamiento de la librería de aprendizaje skitlearn, también se consideró el cambio de kernel de la función dentro de cada clasificador con base radial considerando todas el caso del proyecto como multi-varial, también se utilizó un Big data del 99,8% de más de 2000 calificaciones y características de estudiantes en 8 ciclos académicos.

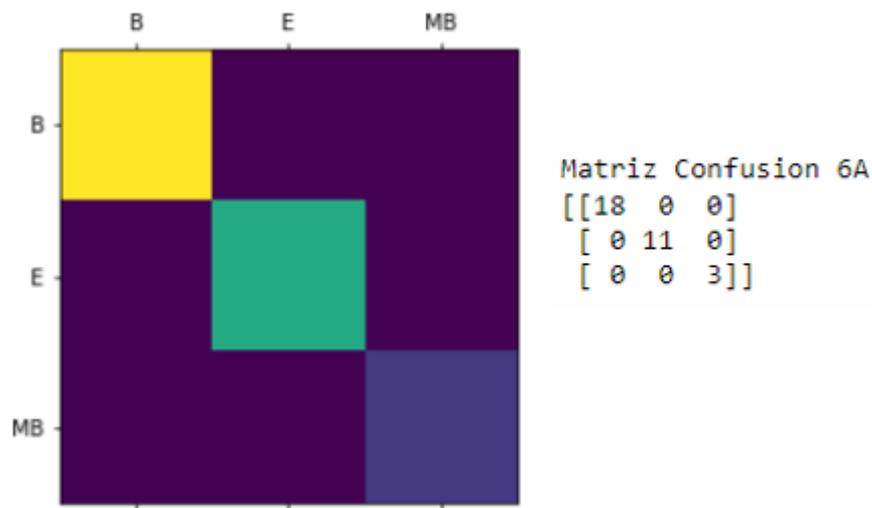
Por ello se considera la utilización del modelo predictivo con Maquinas de soporte vectorial, tomando en cuenta un error de 0.3% y un porcentaje de precisión en la predicción del 96,6% en el entrenamiento del modelo, que se puede visualizar de forma gráfica a continuación.



**Figura 32.** Diagrama de confusión con Maquinas de soporte vectorial en el rendimiento académico de los estudiantes de la institución educativa.

Elaborado por: (Chucuri 2022).

En consecuencia, a ello, la figura 32 se enuncia el diagrama de confusión que representa la predicción de datos del rendimiento académico de la escuela Isabel la Católica en función del entrenamiento del modelo predictivo. Posteriormente se realizó la evaluación de predicciones con máquinas de soporte vectorial, así pues, se vio la necesidad de utilizar nuevos datos de un nuevo curso de la escuela Isabel la Católica, con la finalidad de evitar predicciones erróneas y el sobre entrenaamiento del modelo predictivo, como resultado a continuación se presenta la predicción de una matriz de confusión del sexto grado con el respectivo diagrama de confusión, en el que establece que 18 estudiantes tendrán un desempeño académico “Bueno”, 11 estudiantes con el desempeño académico “Excelente” y 3 estudiantes con el desempeño académico “Muy bueno”, por ello se analiza que no existirán estudiantes con desempeño académico “Regular” o “Deficiente” es decir, no existirán estudiantes con calificaciones menores a 7 puntos en este paralelo.

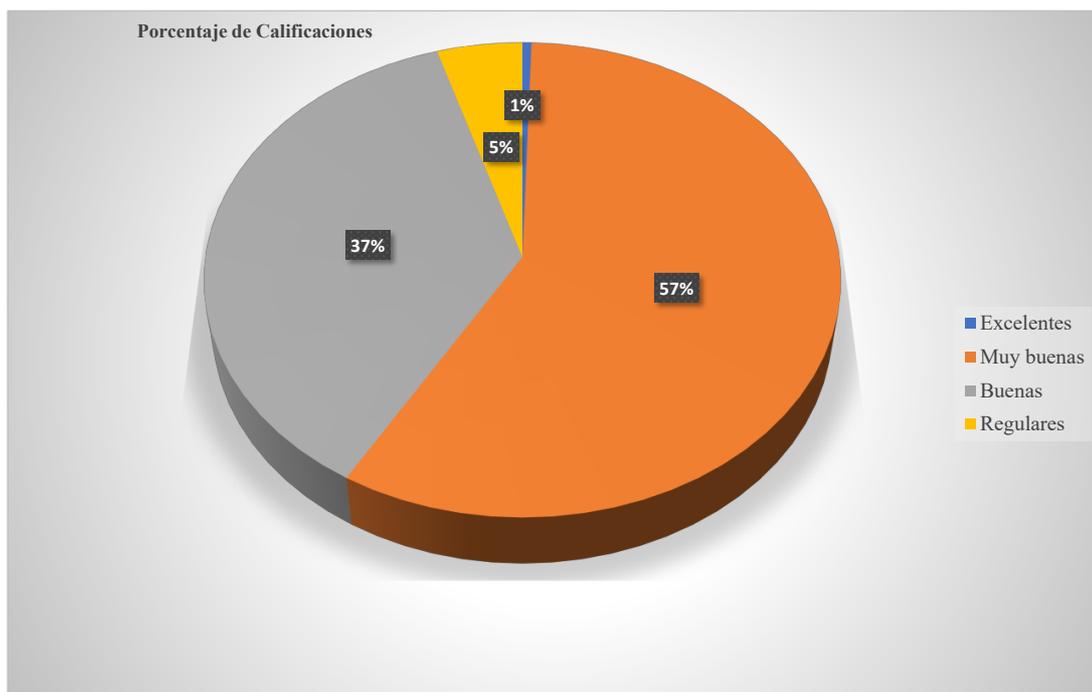


**Figura 33.** Matriz de confusión después del entrenamiento con Maquinas de soporte Vectorial den el rendimiento académico de los estudiantes de la institución educativa.

Elaborado por: (Chucuri 2022).

Así podemos analizar que el desempeño académico de la Escuela “Isabel la Católica” en la predicción sexto “A” tiene calificaciones aceptables y no se encuentran estudiantes con calificaciones menores a 7 puntos dentro de la predicción, en el modelo utilizado en el ciclo académico 2021 - 2022.

En consecuencia, se presenta en la siguiente gráfica, la predicción por ciclo escolar del desempeño académico de la escuela “Isabel la Católica” en donde se utilizaron los datos de entrenamiento para el modelo predictivo.



**Figura 34.** Porcentaje de calificaciones del rendimiento académico de los estudiantes de la institución educativa.

Elaborado por: (Chucuri 2022).

### **Contraste con otras investigaciones**

El sistema de académicos del Ecuador está basado en, modelos procesos de aprendizajes no adecuados para los estudiantes, por lo que la educación impartida no es la adecuada y se categoriza como mediocre. Por ello diversas problemáticas a largo plazo como; la deserción escolar en el nivel secundario y superior, al igual que la pobreza en la sociedad se pueden atribuir a la falta de relación entre docente y estudiante, así como la falta de decisiones apropiadas ante el bajo desempeño académico en los estudiantes (Núñez-Naranjo, Ayala-Chauvin y Riba-Sanmartí 2021).

Por otro lado, muchos países actualmente están utilizando múltiples modalidades y procesos aprendizaje en dónde la minería de datos y herramientas Big data permiten dar un mejor enfoque al análisis del rendimiento académico, puesto que se da un seguimiento en base al comportamiento de los estudiantes permitiendo tomar mejores decisiones para mejorar el desempeño académico (Bonami 2020).

En base a ello, el modelo predictivo del presente proyecto permite obtener un mejor reporte del desempeño académico de la escuela de Educación Básica "Isabel la Católica" con la finalidad de tener resultados claros, de las necesidades académicas de los estudiantes y tomar decisiones a largo plazo de forma más rápida y eficiente.

En (Fernández-Morales y Vallejo-Casarín 2014) se menciona que, frente a los constantes cambios tecnológicos, la educación debe evolucionar para poder ofrecer una mejor experiencia al estudiante con el uso de herramientas de Inteligencia artificial (IA) y poder obtener los resultados que la sociedad busca en la actualidad.

Por ello , el primer paso a un proceso de evolución y adaptabilidad tecnológica dentro de la escuela Isabel la Católica, es una de las principales ventajas , por otro lado es necesario avanzar a una implementación del modelo mediante la utilización de una plataforma que nutra aún más el modelo predictivo , y pueda brindar una mejor experiencia en su uso , por lo cual se considera una desventaja en el proyecto, sin embargo el análisis realizado brindó una mejora significativa a la base de datos, puesto que se utilizó información en años anteriores que no constaban en la matriz , además la visualización cualitativa y visual del comportamiento del desempeño académico así como la predicción podrán brindar mejor enfoque a los docentes de la institución así como a su administración , facilitando y mejorando la toma de decisiones para el rendimiento académico de la escuela.

## **Verificación de la hipótesis**

Mediante el método de chi cuadrado planteado a continuación se comprueba y evalúa la hipótesis del trabajo de investigación, así pues, se resume el grado de libertad como 13 unidades.

$$V = (r - 1)(c - 1) \quad (11)$$

$$V = (14 - 1)(2 - 1) \quad (12)$$

$$V = (13)(1) \quad (13)$$

$$V = 13 \text{ o grado de libertad} \quad (14)$$

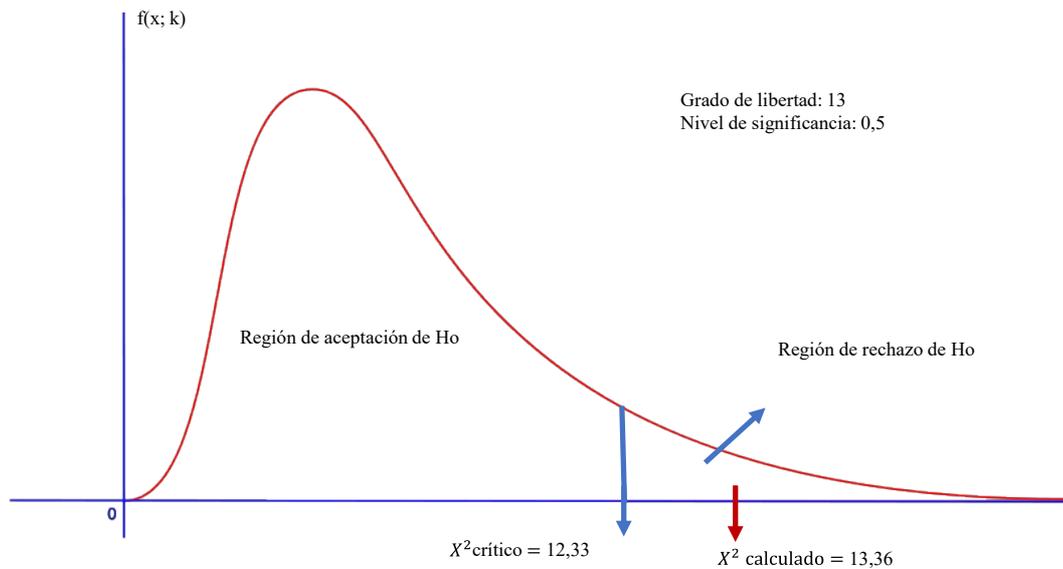
$$X^2 = \sum \frac{(F_o - F_e)^2}{F_e} \quad (15)$$

Teniendo en cuenta los cálculos necesarios para la frecuencia obtenida y la frecuencia esperada, se procede a enunciar los resultados del método en el desempeño académico de las diferentes asignaturas obtenidas, así como el desempeño a niveles deficientes de las mismas asignaturas, con el fin de evaluar la relación de del desempeño académico actual y el desempeño académico deficiente.

**Tabla 6.** Cálculo de Chi cuadrada

<b>Desempeño Académico</b>	<b>Frecuencia Obtenida (Fo)</b>	<b>Frecuencia Esperada (Fe)</b>	<b>X<sup>2</sup></b>
Estudios Sociales	9,51	7,03	0,87105
	4,00	7,03	1,30908
Ciencias Naturales	9,48	7,03	0,85007
	4,00	7,03	1,30908
Educación Artística	9,26	7,03	0,70400
	4,00	7,03	1,30908
Matemática	8,92	7,03	0,50531
	4,00	7,03	1,30908
Inglés	8,84	7,03	0,46334
	4,00	7,03	1,30908
Educación Física	8,73	7,03	0,40860
	4,00	7,03	1,30908
Lenguaje	8,71	7,03	0,39901
	4,00	7,03	1,30908
<b>Total</b>	91,45	98,48	13,36

Elaborado por: Chucuri, Boris (2022).



**Figura 35.** Distribución de la Chi cuadrado hasta su punto crítico.

Elaborado por: (Chucuri 2022).

Por consiguiente, en la figura 35 se visualiza los resultados del método aplicado y comprueba que; el valor de chi cuadrado está fuera dentro de los límites de aceptación de la hipótesis, por ello se menciona que las evidencias del estudio no son lo suficientemente contundentes para la aprobación de la hipótesis planteada.

### Componente Ambiental

El análisis Big data de la Escuela de educación básica “Isabel la Católica”, permitirá reducir la utilización de recursos físicos, así como energéticos, para evitar la degradación ambiental de forma local y contribuir al cuidado del medio ambiente. Esto permitirá reducir el número de reuniones grupales entre docentes para la toma de decisiones del desempeño académico del alumnado, así como la reducción del uso de infraestructura y consumo energético dentro de la unidad educativa, igual que la utilización de documentación física que justifique el plan de horas de refuerzo a los estudiantes con bajo desempeño académico , puesto que el modelo predictivo justificara el bajo desempeño académico al os grupos de estudiantes que puedan necesitar tutoría académica, y toda esta información se manejaría de forma digital, en conjunto con la base de datos de la unidad educativa.

## CAPITULO V

### CONCLUSIONES Y RECOMENDACIONES

#### **Conclusiones:**

En el presente proyecto técnico se realizó el análisis del desempeño académico de la Escuela de Educación Básica “Isabel La Católica”, para ello se efectuó la extracción y análisis de la información que dispone la unidad educativa, mediante un registro de Excel, luego se procedió a procesar la información, mediante técnicas de filtrado de datos y poder tener una mejor organización de la información obtenida, también se procedió a plantear un modelo predictivo del desempeño académico, mediante herramientas Big data, también se hizo la comparación de tres clasificadores, con la finalidad de determinar cual tendría mejor adaptabilidad al caso de estudio, finalmente se evaluó las predicciones con métricas de sensibilidad, precisión y exactitud dentro del modelo aplicado, con datos de entrenamiento y post entrenamiento.

En (Núñez-Naranjo, Ayala-Chauvin y Riba-Sanmartí 2021) predicen el desempeño académico de nivel superior en un conjunto de datos relacionados con la deserción académica por medio de máquinas de soporte vectorial con dos variables, sin embargo en el presente estudio se utilizó restricciones con múltiples variables para poder predecir el desempeño académico en un nivel de educación inicial, teniendo ventajas dentro el campo académico para poder predecir el desempeño académico de los estudiantes y poder proponer estrategias de seguimiento y mejora continua desde edad temprana, siendo un proceso idóneo para el aprendizaje. Por otro lado, es importante considerar ampliar el modelo del software para establecer diversas características que faciliten la interpretación de resultados, así como su integración dentro de una plataforma de apoyo académico dirigido al docente. En resumen, la utilización de herramientas Big data permitirá al docente plantear mejores estrategias de enseñanza-

aprendizaje, tras el análisis y evaluación de datos obtenidos por medio del modelo, y así poder tener mejores resultados en el rendimiento académico de la unidad educativa.

**Recomendaciones:**

- ❖ Ampliar el modelo predictivo con la integración de su funcionamiento dentro de la plataforma de apoyo dirigido al docente.
- ❖ Integrar características que faciliten el análisis e interpretación de resultados dentro del software.
- ❖ Actualizar los datos relacionados a la información socioeconómica de los estudiantes con nuevas encuestas dentro de la unidad educativa para mejorar la interpretación de resultados.
- ❖ Integrar datos relacionados con la formación y experiencia de los docentes de la unidad educativa, para encontrar nuevos patrones relacionados con el desempeño académico y mejorar la predicción del modelo.

## Bibliografía:

- ALBA, D. y CALLE, J., 2016. Aplicación De Técnicas De Machine Learning Basado En Información Sísmica Para Profundizar La Probabilidad De Terremotos Mediante El Uso De Regresión Logística Y Redes Neuronales. [en línea], pp. 260. Disponible en: <http://repositorio.ug.edu.ec/handle/redug/11741>.
- BARUPAL, D.K. y FIEHN, O., 2019. Generating the blood exposome database using a comprehensive text mining and database fusion approach. *Environmental Health Perspectives*, vol. 127, no. 9, pp. 2825–2830. ISSN 15529924. DOI 10.1289/EHP4713.
- BETANCOURT, G.A., 2005. Las maquinas de soporte vectorial. *Scientia et Technica*, vol. 27, no. 27, pp. 67–72.
- BONAMI, B.P., 2020. Educación , Big Data e Inteligencia Artificial : Metodologías mixtas en plataformas digitales Education , Big Data and Artificial Intelligence : Mixed methods in digital platforms. *Revista Comunicar* [en línea], no. 65, pp. 43–52. Disponible en: <http://web.a.ebscohost.com.ezproxy.uniminuto.edu/ehost/pdfviewer/pdfviewer?vid=1&sid=248de2a7-f633-431e-a68c-a468b69fdb0f%40sdc-v-sessmgr01>.
- CENTENO MARTÍN-ROMERO, A., 2020. Big Data. Técnicas de machine learning para la creación de modelos predictivos para empresas. [en línea], pp. 1–57. Disponible en: <https://repositorio.comillas.edu/xmlui/handle/11531/45878>.
- CHALLENGER PÉREZ, I., DÍAZ RICARDO, Y. y BECERRA GARCÍA, R., 2014. El lenguaje de programación Python/The programming language Python. *Revista Ciencias Holguín* [en línea], vol. 20, pp. 1–13. ISSN 1027-2127. Disponible en: <http://www.linuxjournal.com/article/2959>.
- ENRIQUEZ ,CAROLINA - SEGURA, ANGELA - TOVAR, J., 2013. Factores de Riesgo Asociados a bajo rendimiento academicoen escolares de Bogota. *Invesitigaciones Andina* [en línea], vol. 15, pp. 654–666. Disponible en: <https://n9.cl/tgnxf>.
- ESCOBAR BORJA, M. y MERCADO PÉREZ, M., 2019. *Big data: un análisis documental de su uso y aplicación en el contexto de la era digital*. 2019. S.l.: s.n.
- FERNÁNDEZ-MORALES, K. y VALLEJO-CASARÍN, A., 2014. Online Education: A Perspective Based on the Experiences of Each Country. *Revista de Educación y Desarrollo* [en línea], no. 29, pp. 1–11. Disponible en: [http://www.cucs.udg.mx/revistas/edu\\_desarrollo/anteriores/29/029\\_Fernandez.pdf](http://www.cucs.udg.mx/revistas/edu_desarrollo/anteriores/29/029_Fernandez.pdf).
- HERNÁNDEZ-LEAL, E.J., DUQUE-MÉNDEZ, N.D. y MORENO-CADAVID, J., 2017. Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. *TecnoLógicas*, vol. 20, no. 39, pp. 15–38. ISSN 0123-7799. DOI 10.22430/22565337.685.
- MEDINA-MERINO, R.F. y ÑIQUE-CHACÓN, C.I., 2017. Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python. *Interfases*, vol. 0, no. 010, pp. 165. DOI 10.26439/interfases2017.n10.1775.
- NIEMANN, J. y PISLA, A., 2021. Big Data en la educación. *Mechanisms and Machine Science*, vol. 90, pp. 91–105. ISSN 22110992. DOI 10.1007/978-3-030-56449-0\_8.

- NÚÑEZ-NARANJO, A.F., AYALA-CHAUVIN, M. y RIBA-SANMARTÍ, G., 2021. Prediction of University Dropout Using Machine Learning. , vol. 2, pp. 396–406. DOI 10.1007/978-3-030-68285-9\_38.
- PUYOL MORENO, J., 2014. Una aproximación a Big Data. *Revista de Derecho de la UNED (RDUNED)*, no. 14, pp. 471. ISSN 1886-9912. DOI 10.5944/rduned.14.2014.13303.
- SAKTI, B.P., 2018. Big Data en la Educacion: Beneficios e Impacto de la Analitica de Datos. , vol. 12, no. 4, pp. 336–347.
- SALGADO, M., 2014. Oracle apuesta por Big Data con tecnología y proyectos. *Oracle apuesta por Big Data con tecnología y proyectos* [en línea], Disponible en: <https://www.computerworld.es/big-data/oracle-apuesta-por-big-data-con-tecnologia-y-proyectos>.
- URIBE-ENCISO, O.L. y CARRILLO-GARCÍA, S., 2014. Relationship between reading and writing, Academic performance and university dropout. *Entramado*, vol. 10, no. 2, pp. 272–285. ISSN 1900-3803.

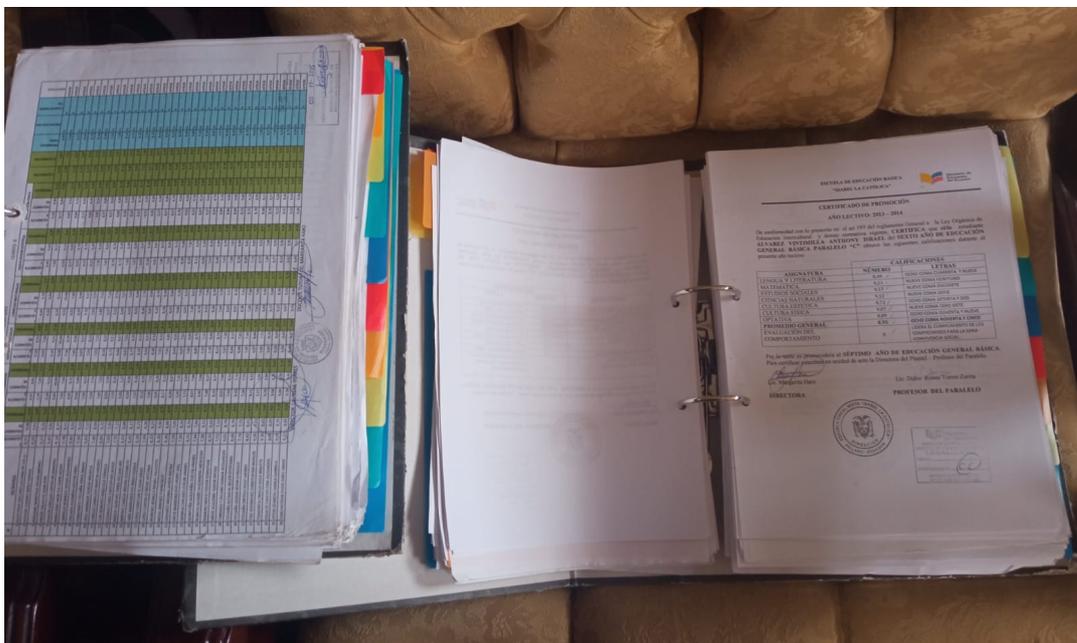
## ANEXOS:

*Anexo 1: Concentrados físicos de las calificaciones anuales de la Escuela de Educación Básica Isabel la Católica los periodos de 2013 al 2015*



Elaborado por: Chucuri, Boris (2022).

**Anexo 2: Concentrados físicos de las calificaciones anuales de la Escuela de Educación Básica Isabel la Católica-Búsqueda de información por estudiante**



Elaborado por: Chucuri, Boris (2022).

*Anexo 3: Búsqueda de concentrados físicos de los ciclos académicos faltantes de la Escuela de Educación Básica Isabel la Católica dentro del departamento de secretaria y vicerrectorado.*



Elaborado por: Chucuri, Boris (2022).

*Anexo 4: Vista frontal de la fachada de la Escuela de Educación Básica Isabel la Católica, ubicada en el cantón Pillaro*



Elaborado por: Chucuri, Boris (2022).

*Anexo 5: Concentrados físicos e informes de las calificaciones anuales de la Escuela de Educación Básica Isabel la Católica en años anteriores.*



Elaborado por: Chucuri, Boris (2022).

*Anexo 6: Búsqueda de concentrados físicos faltantes en la base de datos, de las calificaciones anuales de la Escuela de Educación Básica Isabel la Católica en años anteriores*



Elaborado por: Chucuri, Boris (2022).

Anexo 7: Modelo de ficha socioeconómica por alumno de la Escuela de Educación Básica Isabel la Católica la que se extrajo información por estudiante.

Ministerio de Educación

**UNIDAD EDUCATIVA "ISABEL LA CATÓLICA"**  
DEPARTAMENTO DE CONSEJERÍA ESTUDIANTIL Año Lectivo: 2021-2022

**L- DATOS DE IDENTIFICACIÓN / INFORMACIÓN**

Apellidos y nombres del/ la estudiante:		Cevallos Chicaiza Marilyn Sabella			
Grado / Año:	Tercero	Paralelo:	C	Jornada:	Matutina
Cédula de Ciudadanía:	1851062230	Lugar y fecha de nacimiento: (dd/mm/aa)		Ambato 25/08/2014	
Nombre del/la representante legal:		Monica Chicaiza Lopez			
Dirección:		Calle Principal: Sucre		Calle Secundaria: Urbina	
Referencia: Frente al garaje de la policía nacional					
Provincia:	Tungurahua	Cantón:	Pillaro	Parroquia:	Pillaro
Telefonos:		092864496		Etnia:	mestiza
En caso de emergencia llamar a: Monica Chicaiza					

**2. DATOS FAMILIARES:**

Madre: Monica Chicaiza	Padre: José Cevallos	*Representante: Monica Chicaiza
Cédula de identidad: 1804190732	Cédula de identidad: 1804306015	Cédula de identidad: 1804190732
Profesión: Médico	Profesión: Peluquero	Profesión: Médico
Lugar: Ambato	Lugar: Pillaro	Lugar: Ambato
Teléfono: 0992864496	Teléfono: 0995946738	Teléfono: 0992864496
Instrucción:	Instrucción:	Instrucción:
Primaria	Primaria	Primaria
Secundaria	Secundaria	x Secundaria
Otros	x Otros	Otros

\* Este casillero se completa solamente si el estudiante se encuentra bajo el cuidado de otra persona que no sean sus progenitores.

**3.- REFERENCIAS FAMILIARES DEL/LA ESTUDIANTE:**

**3.1 Personas con quien vive el estudiante**

Madre	Padre	Hermanos	Representante/Tutor	Otros	Lugar entre hermanos:
x	x	x			2

Nº	Apellidos y Nombres/Hermanos	Edad	Institución	Grado/Curso/Jornada:
1	Cevallos Chicaiza Bryan Alexander	16	UNIDAD EDUCATIVA JORGE ÁLVAREZ	TERCERO DE BACHILLERATO

Detallar el nombre, edad, institución de los hermanos en orden de mayor a menor / Si es hijo único escribir esa observación

**3.2 Descripción de la estructura familiar:** Describir datos relevantes de su familia, algún detalle que sea importante conocer

Observaciones: .....

**3.3 Antecedentes familiares:**

Familiares con algún tipo de discapacidad/Vulnerabilidad: SI..... NO..x.....

Determinar quién: Nombre: ..... Parentesco: .....

Tipo de discapacidad/ Vulnerabilidad: .....

**4.- DATOS DE SALUD:**

El estudiante tiene algún tipo de discapacidad: SI  No

Física	Intelectual	Visual	Auditiva	Multidiscapacidad	Psicológica
--------	-------------	--------	----------	-------------------	-------------

Nº Carné: .....

El estudiante tiene alguna condición médica específica: SI  No  Determinar cuál: .....

Especificar medicamentos que utiliza: .....

Observación: .....

**5. Condiciones de vivienda**

ión: Av. Amazonas N34-451 y Av. Atahualpa.  
postal: 170507 / Quito-Ecuador  
no: 593-2-396-1300 / www.educacion.gob.ec

 **Gobierno del Encuentro** | Juntos lo logramos

Elaborado por: Chucuri, Boris (2022).

**Anexo 8: Formato digital de la ficha socioeconómica utilizada en la Escuela de Educación Básica Isabel la Católica hoja número 1.**



República del Ecuador

Ministerio de Educación

UNIDAD EDUCATIVA "ISABEL LA CATÓLICA"  
DEPARTAMENTO DE CONSEJERÍA ESTUDIANTIL Año Lectivo: 2021-2022

**1.- DATOS DE IDENTIFICACIÓN/ INFORMACIÓN**

Apellidos y nombres del/ la estudiante:			
Grado /Año:		Paralelo:	Jornada:
Cédula de Ciudadanía:		Lugar y fecha de nacimiento: (dd/mm/aa)	
Nombre del/la representante legal:			
Dirección:	Calle Principal:	Calle Secundaria:	
Referencia:			
Provincia:	Cantón	Parroquia	Barrio/sector
Teléfonos:		Etnia:	
En caso de emergencia llamar a:			

**2. DATOS FAMILIARES:**

Madre:	Padre:	*Representante:
Cédula de identidad:	Cédula de identidad:	Cédula de identidad:
Profesión:	Profesión:	Profesión:
Lugar:	Lugar:	Lugar:
Teléfono:	Teléfono:	Teléfono:
Instrucción:	Instrucción:	Instrucción:
Primaria	Primaria	Primaria
Secundaria	Secundaria	Secundaria
Otros	Otros	Otros

*\* Esta casilla se completa solamente si el estudiante se encuentra bajo el cuidado de otra persona que no sean sus progenitores.*

**3.- REFERENCIAS FAMILIARES DEL/LA ESTUDIANTE:**

**3.1 Personas con quien vive el estudiante**

Madre	Padre	Hermanos	Representante/Tutor	Otros	Lugar entre hermanos:
Nº	Apellidos y Nombres/Hermanos		Edad	Institución	Grado/Curso/Jornada:

*Detallar el nombre, edad, institución de los hermanos en orden de mayor a menor / Si es hijo único escribir esa observación*

**3.2 Descripción de la estructura familiar:** Describir datos relevantes de su familia, algún detalle que sea importante conocer

Observaciones:.....

**3.3 Antecedentes familiares:**

Familiares con algún tipo de discapacidad/Vulnerabilidad: SI..... NO.....

Determinar quién: Nombre: .....

Tipo de discapacidad/ Vulnerabilidad: .....Parentesco: .....

**4.- DATOS DE SALUD:**

El estudiante tiene algún tipo de discapacidad: Sí  No

Física	Intelectual	Visual	Auditiva	Multidiscapacidad	Psicológica
Nº Carné					

El estudiante tiene alguna condición médica específica: Sí  No  Determinar cuál: .....

Especificar medicamentos que utiliza:.....

**Anexo 9: Formato digital de la ficha socioeconómica utilizada en la Escuela de Educación Básica Isabel la Católica hoja número 2.**



República del Ecuador

Ministerio de Educación

Observación: .....

**5. Condiciones de vivienda**

Propia	Arrendada	Prestada	Anticresis	Con préstamo		
Luz eléctrica	Agua Potable	SSH	Cable	Teléfono	Celular	Computador
Internet	Pozo séptico	Tanque Reservorio	Acequia	Alcantarillado		

Breve descripción de la vivienda: (casa, departamento, cuarto, etc.) .....

**6.- DATOS ACADÉMICOS/ RENDIMIENTO ESCOLAR**

- ✓ Desde que año escolar el/la estudiante ingresó a la institución:....
- ✓ Institución educativa de la que procede: .....

**6.1 Datos Académicos:**

Asignaturas de preferencia del estudiante:

Matemática	Lengua	EESS	CCNN	Inglés	Otras
------------	--------	------	------	--------	-------

Asignaturas en las que ha tenido dificultad:

Matemática	Lengua	EESS	CCNN	Inglés	Otras
------------	--------	------	------	--------	-------

**6.2. HISTORIA ESCOLAR**

A repetido algún año, cual	Debido a				
Dificultades Académicas:	Cuales				
Apoyo recibido:	Pedagógico	Logopédico	Psicológico	Fisioterapéutico	Otro

*Describir si en algún momento el estudiante acudió a apoyo psicológico, lugar, edad, razón*

Observación: .....

**7.- PSICODANAMNÉSIS PERSONAL**

**7.1. Embarazo y parto**

EMBARAZO					
Edad de la madre	Accidentes	Medicamentos	Emociones fuertes	Riesgo	
PARTO					
A término	Prematuro	Tardío	Cesárea	Normal	Parto inducido
Llanto	Otros	<i>Describir detalles del parto</i>			

**7.2.- Datos del/ la niño/a:**

Al nacer		Edad que empezó a								
Peso	Talla	Sostuvo la cabeza	Sentarse	Gatear	Caminar	Correr	Hablar	Lactancia (Hasta que edad tomó seno)	Biberón (Hasta que edad tomó biberón)	Control esfínteres (Edad en la que le quitó el pañal)

**7.3 Enfermedades (desde la infancia hasta la actualidad)**

Enfermedades	Accidentes	Alergias	Cirugías	Perdidas de conocimiento	Otros
--------------	------------	----------	----------	--------------------------	-------

**8.- Autorización**

Autorizo a los funcionarios del Departamento de Consejería Estudiantil "DECE" realicen entrevistas, intervenciones, derivaciones (instituciones externas correspondientes según sea el caso detectado MSP – Fiscalía – JCPD), de acuerdo a lo que requiera mi representado/a con el fin de precautelar el bienestar físico y emocional garantizando su permanencia en el sistema educativo.

SI .....

NO .....

Observación: .....

**Profesional DECE responsable**

CI: .....

**Representante**

CI: .....

### *Anexo 10: Líneas de Código 1*

Importación de librerías y datos mediante un archivo con extensión csv, para la generación de matrices previas, obtención de la descripción rápida de datos, verificación de categorías existentes, descripción estadística rápida y verificación de correlación existente.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
data=pd.read_csv("Rendimiento Academico1.csv")
print(data.info())
ata["GE"].value_counts()
estadistica=data.describe()
ata_train=data.iloc[:,1:12]#
correlacion = data_train.corr(method='pearson')
print(correlacion)
data_rd=data.drop(['ALUMNO', 'GE','AI','AF'],axis=1)
corr_matrix = data_rd.corr()
corr_matrix["Promedio"].sort_values(ascending=False)
```

Elaborado por: Chucuri, Boris (2022).

### *Anexo 11: Líneas de Código 2*

Declaración e impresión del sesgo en los datos obtenidos para el análisis del comportamiento de resultados, impresión y visualización de datos en ciclos académicos en las materias de Educación Artística y Educación Física, Ciencias Naturales y Estudios Sociales.

```
sesgo=data_train.skew()
print(sesgo)
x_a=data.iloc[:,14]
y_ccnn=data.iloc[:,5]
plt.plot(x_a,y_ccnn,'go')
plt.ylabel("y CCNN")
plt.xlabel('x Años')
plt.show()
#P2    EDU.ARTIS.
y_art=data.iloc[:,6]
plt.plot(x_a,y_art,'bo')
plt.ylabel("y EDU.ARTIS")
plt.xlabel('x Años')
plt.show()
#P3    EDU.FISICA
#P4    SOCIALES
y_s=data.iloc[:,8]
plt.plot(x_a,y_s,'ro')
plt.ylabel("y Sociales")
plt.xlabel('x Años')
plt.show()
```

Elaborado por: Chucuri, Boris (2022).

### *Anexo 12: Líneas de Código 3*

Impresión y visualización de datos en ciclos académicos en las materias de Inglés, Lenguaje y Matemática. Para visualizar su comportamiento y revisión de patrones.

```
#P5    INGRES
      y_s=data.iloc[:,9]

plt.plot(x_a,y_s,'ro')
plt.ylabel("y Ingles")
plt.xlabel('x Años')
plt.show()

#P6    LENGUAJE
      y_l=data.iloc[:,10]
plt.plot(x_a,y_l,'yo')
plt.ylabel("y Lenguaje")
plt.xlabel('x Años')
plt.show()

#P7    MATEMATICA
      y_m=data.iloc[:,11]
plt.plot(x_a,y_m,'ko')
plt.ylabel("y MATEMATICA")
plt.xlabel('x Años')
plt.show()
```

Elaborado por: Chucuri, Boris (2022).

### *Anexo 13: Líneas de Código 3*

Impresión de Diagramas en las diversas asignaturas, revisión del sesgo por asignatura e impresión del diagrama de correlación con los datos importados.

```
axarr=data_train.hist(bins=50, figsize=(20,15))
for ax in axarr.flatten():
    ax.set_xlabel("X")
    ax.set_ylabel("y")
plt.show()
data[INGLES].hist()
plt.title('INGLES')
plt.xlabel('x Notas')
plt.ylabel('y Ingles' )
plt.show()
"""diagrama de matriz de correlación"""
atributos=list(correlacion.columns)
print(atributos)
fig = plt.figure()
ax = fig.add_subplot(111)
cax = ax.matshow(correlacion, vmin=-1, vmax=1)
fig.colorbar(cax)
ticks = np.arange(0,11,1)
#
ax.set_xticks(ticks)
ax.set_yticks(ticks)
ax.set_xticklabels(atributos,rotation=90)
ax.set_yticklabels(atributos)
plt.show()
```

Elaborado por: Chucuri, Boris (2022).

#### *Anexo 14: Líneas de Código 4*

preparación de datos con los parámetros de clasificación requeridos e importación de librerías y transformación de datos necesarios para el entramiento del algoritmo.

```
x=data.iloc[:,1:13]
y=[]
promedio=np.array(data.iloc[:,15].values)
for i in promedio:
    if i >=7 and i<8:
        y.append("R")
    elif i>=8 and i<9:
        y.append("B")
    elif i>=9 and i<10:
        y.append("MB")
    elif i==10.00:
        y.append("E")
    else:
        y.append("D")
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import StandardScaler
atributos_cat=['GE']
atributos_num=['SEXO','DIR','ECO','VIVI','CCNN','EDU.ART','EDU.FIS','SOCIALES','ING
LES','LENGUAJE','MATEMATICA']
trans_Atributos = ColumnTransformer([
    ("num",StandardScaler(),atributos_num),
    ("cat", OneHotEncoder(), atributos_cat),
])x_porcess=trans_Atributos.fit_transform(x)
#print(x_porcess)
le = LabelEncoder()
y = le.fit_transform(y)
print(le.classes_)
```

Elaborado por: Chucuri, Boris (2022).

**Anexo 15:** Líneas de Código 5

Funcionamiento del algoritmo con las funciones de Máquinas de Soporte Vectorial y especificación del cambio de kernel en una función no lineal.

```
from sklearn.model_selection import train_test_split
x_train,x_val,y_train,y_val =train_test_split(x_porcess,y,test_size=0.2)
from sklearn.svm import SVC
C=10 #C=0.1 c=0.01 c=0.001 c=10 (1-100)

svc= SVC(kernel='rbf', gamma ='auto',C=C)
svc.fit(x_train, y_train)
y_pred= svc.predict(x_val)
```

Elaborado por: Chucuri, Boris (2022).

### *Anexo 16: Líneas de Código 6*

Funcionamiento del algoritmo con las funciones de Regresión Logística y especificación de las librerías a utilizar con el porcentaje de error y precisión del modelo.

```
from sklearn.linear_model import LogisticRegression
softmax_reg = LogisticRegression(multi_class="multinomial",solver="saga")
softmax_reg.fit(x_train, y_train)
y_pred_rl=softmax_reg.predict(x_val)

cm_rl=confusion_matrix(y_val,y_pred_rl)
print("Matriz Confusion")
print(cm_rl)
print("Presicion")
p_rl=precision_score(y_val, y_pred_rl,average='micro')
print(p_rl)
e_rl=1-p
print("Error")
print(e_rl)
print("Recuperacion")
r_rl=recall_score(y_val, y_pred_rl,average='micro')

print(r_rl)
```

Elaborado por: Chucuri, Boris (2022).

### *Anexo 17: Líneas de Código 7*

Funcionamiento del algoritmo con las funciones de Random Forest o Árboles Aleatorios y especificación de las librerías a utilizar con el porcentaje de error y precisión del modelo.

```
from sklearn.ensemble import RandomForestClassifier
rnd_clf = RandomForestClassifier(n_estimators=50, max_leaf_nodes=8, random_state=42)
rnd_clf.fit(x_train, y_train)
y_pred_rf = rnd_clf.predict(x_val)
cm_rf=confusion_matrix(y_val,y_pred_rf)
print("Matriz Confusion")
print(cm_rf)
print("Presicion")
p_rf=precision_score(y_val, y_pred_rf,average='micro')
print(p_rf)
e_rf=1-p
print("Error")
print(e_rf)
print("Recuperacion")
r_rf=recall_score(y_val, y_pred_rf,average='micro')
print(r_rf)
```

Elaborado por: Chucuri, Boris (2022).

### **Anexo 18: Líneas de Código 8**

Evaluación de la predicciones y especificación del porcentaje de precisión y error por medio de una matriz de confusión.

```
from sklearn.metrics import confusion_matrix,precision_score,recall_score
cm=confusion_matrix(y_val,y_pred)
print("Matriz Confusion")
print(cm)
np.random.seed(2)
fig,ax = plt.subplots()
#print(ax)
ax.matshow(cm,interpolation ='nearest')
ax.set_xticklabels(['B', 'E', 'MB','R'])
ax.set_yticklabels(['B', 'E', 'MB','R'])
#fig.colorbar(im)
plt.show()
print("Presicion")
p=precision_score(y_val, y_pred,average='micro')#0.93
print(p)
e_sv=1-p#0.06
print("Error")
print(e_sv)
print("Recuperacion")
r=recall_score(y_val, y_pred,average='micro')
print(r)
```

Elaborado por: Chucuri, Boris (2022).

### *Anexo 19: Líneas de Código 9*

Evaluación de las predicciones con datos nuevos de un nuevo curso para el análisis, impresión de una matriz de confusión de la nueva predicción y su respectivo diagrama de confusión.

```
data_7a=pd.read_csv("test7a.csv")
x_test_7a=data_7a.iloc[:,1:13]
y_test_7a=[]
promedio=np.array(data_7a.iloc[:,15].values)
for i in promedio:
    if i >=7 and i<8:
        y_test_7a.append("R")
    elif i>=8 and i<9:
        y_test_7a.append("B")
    elif i>=9 and i<10:
        y_test_7a.append("MB")
    elif i==10.00:
        y_test_7a.append("E")
    else:
        y_test_7a.append("D")
x_test=trans_Atributos.transform(x_test_7a)
y_test_7a= le.transform(y_test_7a)
y_pred_7a= svc.predict(x_test)
print("Matriz Confusion 7A")
print(cm_7a)
np.random.seed(2)
fig,ax = plt.subplots()
ax.matshow(cm_7a,interpolation ='nearest')
ax.set_xticklabels(['B', 'E', 'MB'])
ax.set_yticklabels(['B', 'E', 'MB'])
plt.show()
```

Elaborado por: Chucuri, Boris (2022).