



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA



MAESTRÍA EN GESTIÓN DE SISTEMAS DE INFORMACIÓN E INTELIGENCIA DE NEGOCIOS

ANÁLISIS DEL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES DE LA UNIDAD EDUCATIVA "ARISTÓTELES SCHOOL" APLICANDO HERRAMIENTAS DE BUSINESS INTELLIGENCE Y MINERÍA DE DATOS PARA MEJORAR EL PROCESO DE ENSEÑANZA APRENDIZAJE

AUTOR: LUIS ANTONIO MANTILLA GALARZA

DIRECTOR: MSC RAMIRO NANAC DELGADO RODRIGUEZ





AGENDA

- ❖ Antecedentes
- ❖ Justificación
- ❖ Objetivos
- ❖ Marco Teórico
- ❖ Desarrollo del problema
- ❖ Conclusiones
- ❖ Recomendaciones



ANTECEDENTES

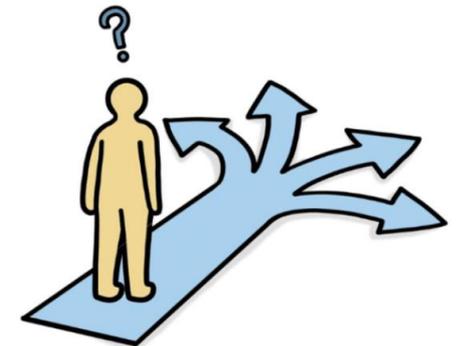
- ❖ Mejorar la calidad de la educación
- ❖ Se lleva la información Académica en hojas excel y no permite un mejor analisis
- ❖ La toma de decisiones deben ser acertadas y basada en el analisis de los datos mas relevantes



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

JUSTIFICACIÓN

- ❖ El fracaso escolar se puede definir como un fenómeno dado en los alumnos/as que presentan bajo rendimiento escolar, discapacidades en las materias instrumentales como: lectura, escritura y cálculo. Es decir, no alcanza el nivel de rendimiento medio esperado para su edad y nivel pedagógico.
- ❖ El presente proyecto pretende proveer a las autoridades de una herramienta de análisis de rendimiento al respecto de sus estudiantes para así tomar decisiones dentro de su competencia con el propósito de mejorar la imagen que tiene la institución con el público o comunidad.



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

OBJETIVOS

- OBJETIVO GENERAL

- ❖ Analizar el rendimiento académico de los estudiantes de la unidad educativa "Aristóteles School" mediante minería de datos, business intelligence y la implementación de un dashboard para proponer mejoras al modelo educativo y garantizar

- OBJETIVOS ESPECÍFICOS

- ❖ Realizar una revisión para determinar posibles soluciones y recomendaciones existentes para analizar el rendimiento académico de los alumnos de una institución educativa.
- ❖ Analizar las herramientas de minería de datos y Business Intelligence existentes en el mercado con el objeto de determinar las más adecuadas para el análisis del rendimiento académico de los estudiantes de la Unidad Educativa "Aristóteles School"
- ❖ Implementar un modelo orientado hacia la minería de datos con el propósito de predecir si un estudiante es propenso al supletorio
- ❖ Evaluar el modelo implementado mediante protocolos de validación con técnicos, usuarios finales y autoridades



MARCO TEORICO

Analítica del aprendizaje: El análisis del aprendizaje aprovecha la gran cantidad de información que generan los estudiantes durante el proceso educativo con la finalidad de mejorar la calidad del aprendizaje y el rendimiento académico (García, 2015)

Minería de datos (Data mining): La revolución digital ha hecho posible capturar, procesar, almacenar, distribuir y difundir fácilmente la información digital. Con los tremendos avances en computación y tecnologías relacionadas y la expansión de su uso en varios aspectos de la vida, se recopilan y almacenan grandes cantidades de información en bases de datos. (Riquelme et al., 2006).

Descubrir conocimiento de este enorme volumen de datos es un reto en sí mismo. La minería de datos (MD) es un intento de buscarle sentido a la explosión de información que actualmente puede ser almacenada.



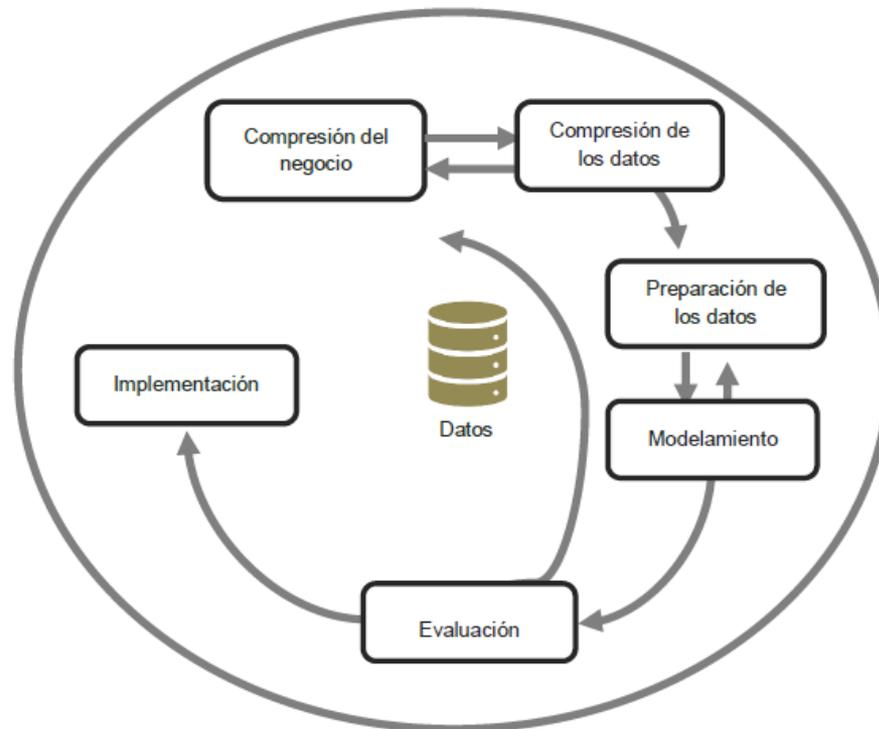
MARCO TEORICO

Business Intelligence: El Business Intelligence, BI o la inteligencia de negocios combina análisis de negocios, minería, visualización, herramientas e infraestructura de datos, además de prácticas recomendadas para ayudar a las empresas o entidades a tomar decisiones basadas en los datos. En la práctica, implementar la inteligencia de negocios moderna implica contar con una vista integral de todos los datos de la organización. Además, consiste en usar estos datos para impulsar el cambio, eliminar las ineficiencias y adaptarse rápidamente a los cambios del mercado o la demanda.



DESARROLLO DEL PROBLEMA

Para el Desarrollo del problema se utiliza la metodología CRISP-DM



Comprensión
del negocio

Comprensión
de los datos

DESARROLLO DEL PROBLEMA

Se identificó las características puntuales del rendimiento de los estudiantes pues la mejor manera de abordar esta temática fue la aproximación con base en las calificaciones. Cada uno de los aspectos se manifestaron con base en tablas descriptivas.

La recolección de datos empleados para esta investigación corresponde a la población estudiantil de la institución objeto de estudio. El periodo de recolección abarca a los años lectivos 2009 a 2020.



DESARROLLO DEL PROBLEMA

Preparación de los datos

Con los datos obtenidos en las fases anteriores, se hizo una selección, integración y limpieza para eliminar data redundante, incompleta con el propósito de crear y limpiar cualquier valor atípico que pueda entorpecer el desarrollo de la investigación.

Es este punto en especial, la idea en general de la preparación de los datos es identificar la calidad de la información encontrada, debido a que al momento de trabajar con datos en crudo estos poseen algunas complicaciones como por ejemplo errores de tipeo, subrayado y desde luego información no característica acorde a la temática de análisis.



DESARROLLO DEL PROBLEMA

Selección de la muestra

Se establece que las edades comprendidas están entre 15 a 19 años donde se produce principalmente en el transcurso del primer año de la enseñanza media superior, este corresponde el primer años de bachillerato o cuarto curso, unas de las razones implicaría el cambio de la Educación general Básica al bachillerato, lo cual significa un cambio importante en el entorno del estudiante, motivo por el cual es ese grupo en cual se centra esta investigación



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DESARROLLO DEL PROBLEMA



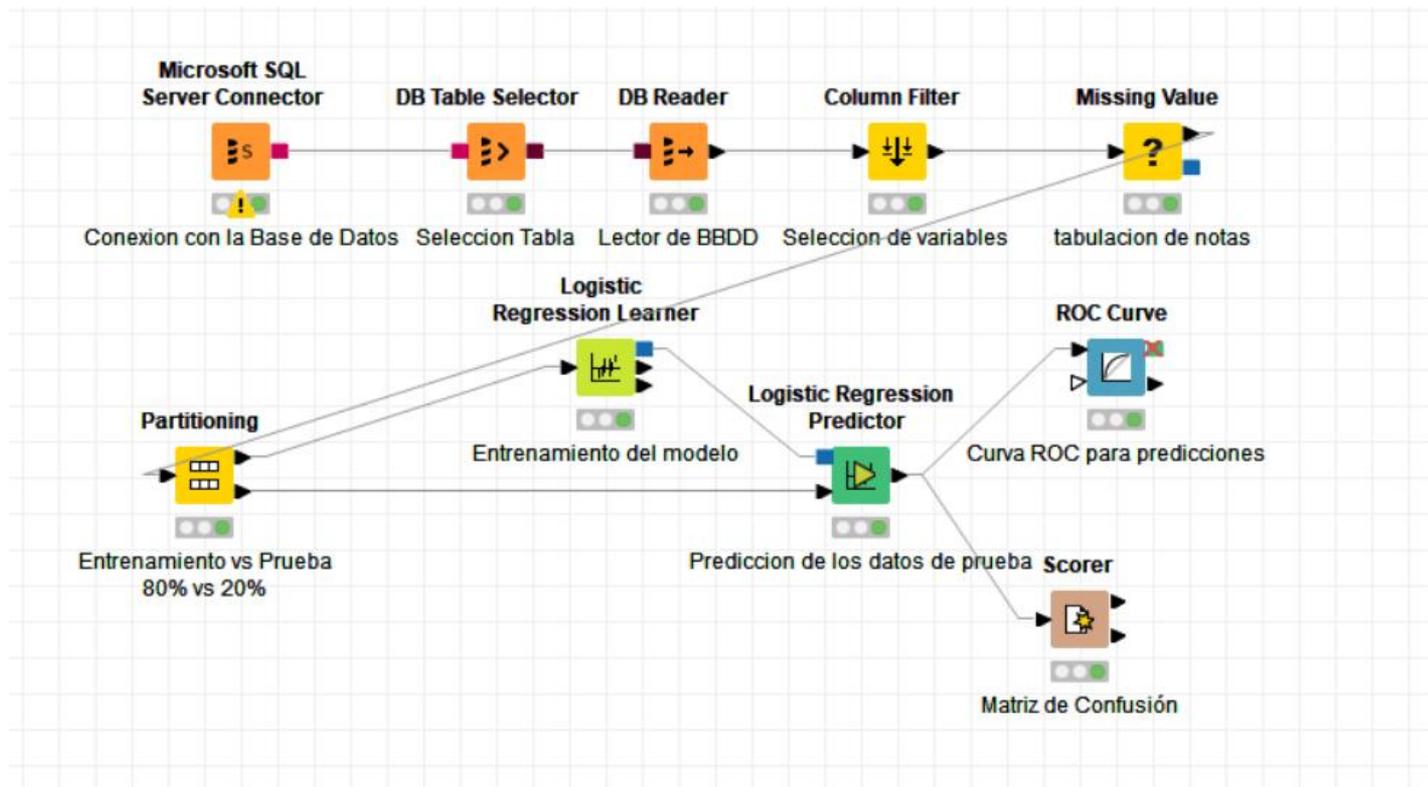
Realizada la preparación de los datos que incluye la limpieza de datos , se utilizo los siguientes modelos de predicción:

- ❖ Regresión logística
- ❖ Random Forest
- ❖ Arbol de decision



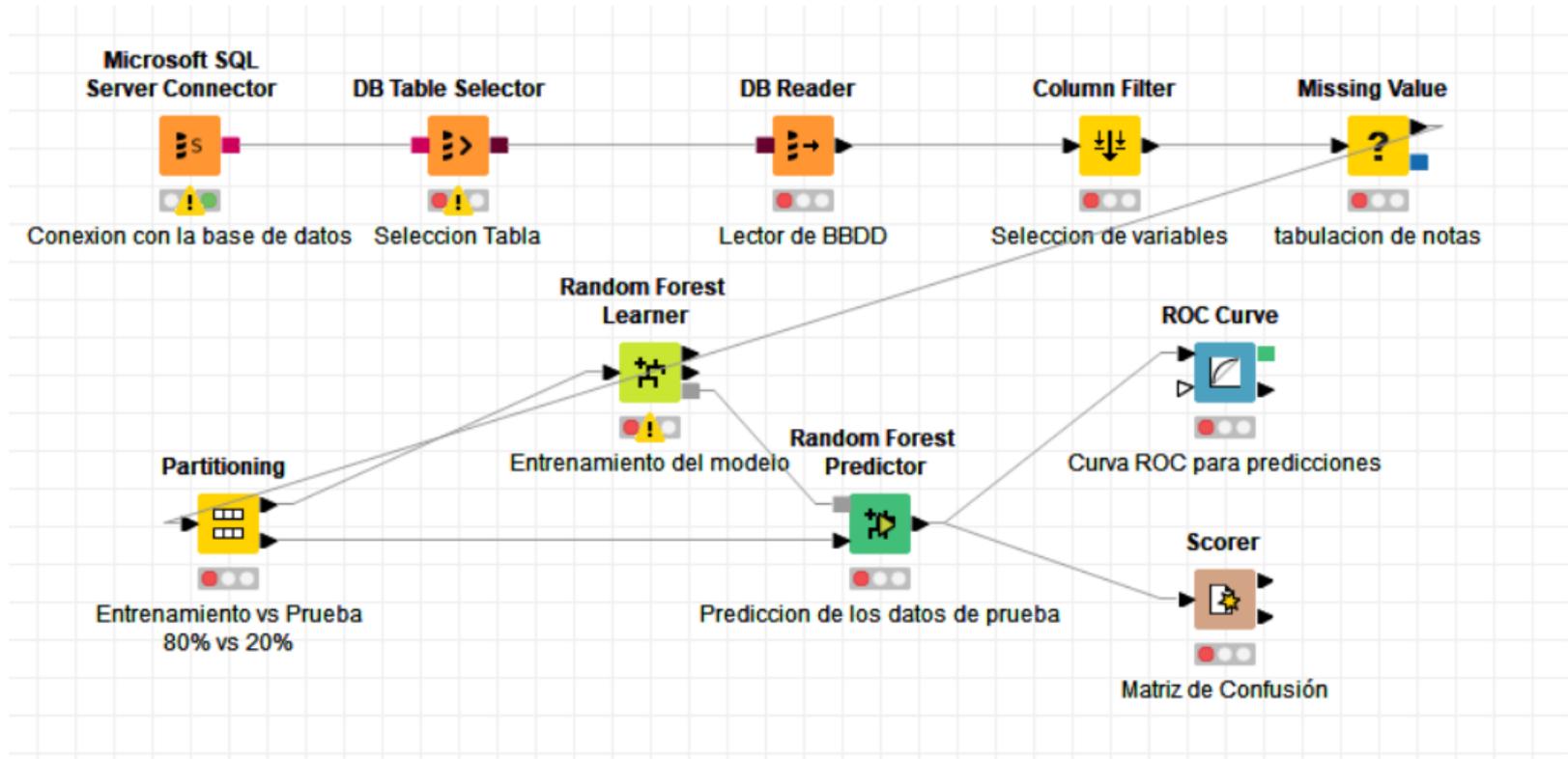
DESARROLLO DEL PROBLEMA

Regresión Logística



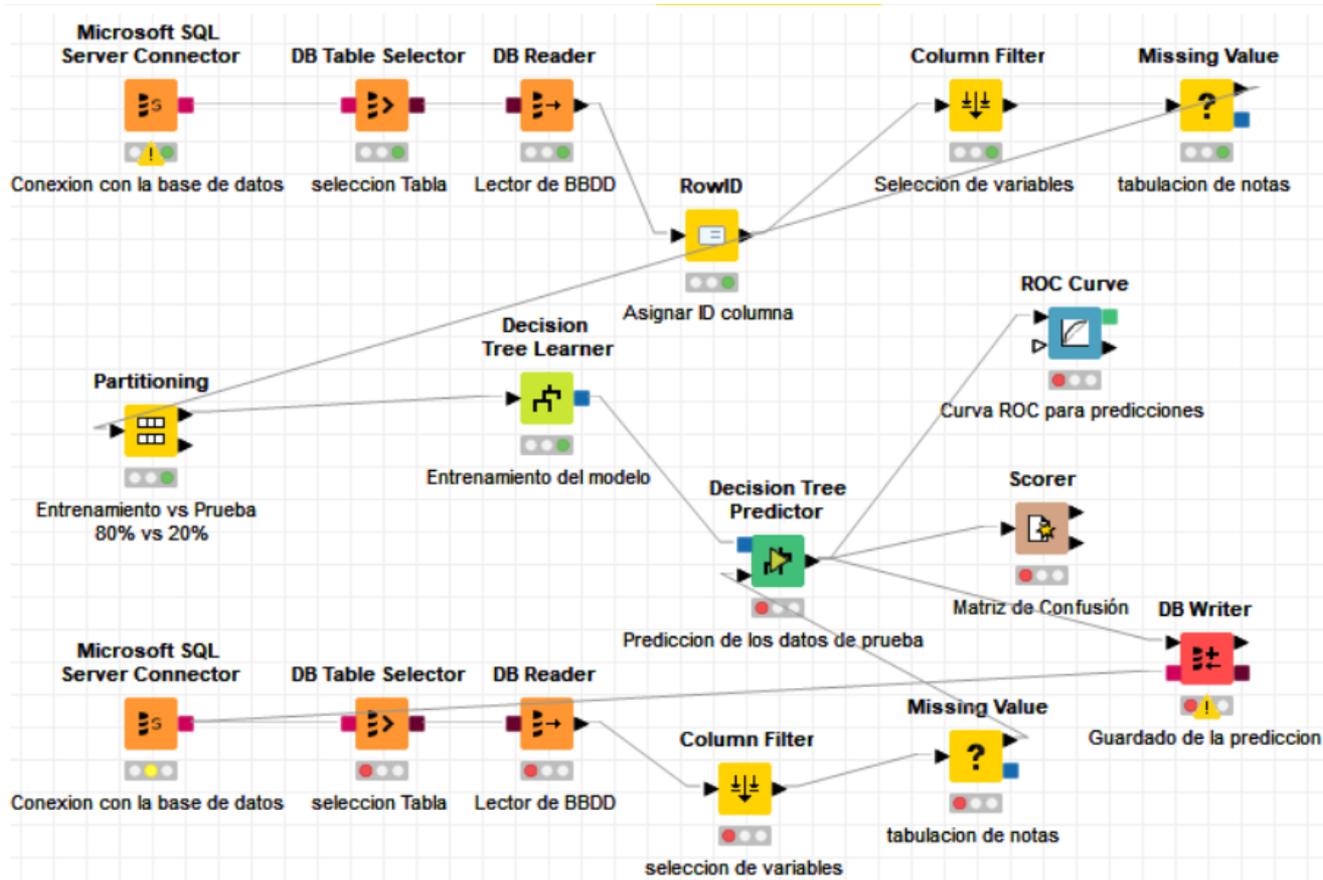
DESARROLLO DEL PROBLEMA

Random Forest



DESARROLLO DEL PROBLEMA

Arbol de decisión



DESARROLLO DEL PROBLEMA

Evaluación

Es uno de los pasos más importantes de la metodología, ya que un análisis descriptivo permite la derivación del desempeño de las tareas y actividades que involucran y se basan en la minería de procesos.

Por otro lado, al hablar de la parte predictiva se intentó utilizar técnicas de aprendizaje automático para aprender el comportamiento de los datos y hacer predicciones con base en si un estudiante está orientándose hacia el supletorio.



DESARROLLO DEL PROBLEMA



Evaluación

La evaluación fue llevada a cabo a partir del Análisis ROC (Receiver Operating Characteristic o característica operativa del receptor), esta es una alternativa útil para la evaluación con precisión de las predicciones de modelo ya que traza la sensibilidad frente a la especificidad de una prueba de clasificación

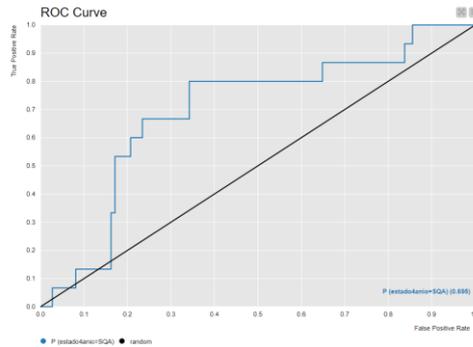


ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

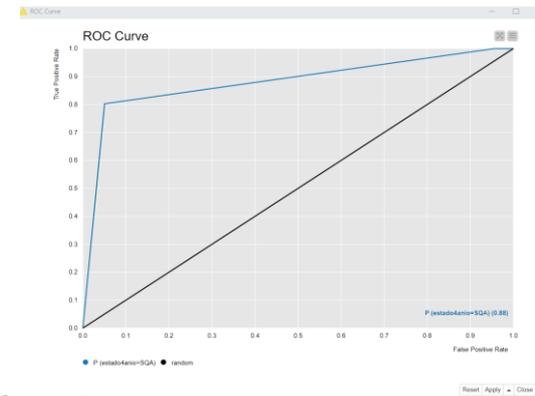
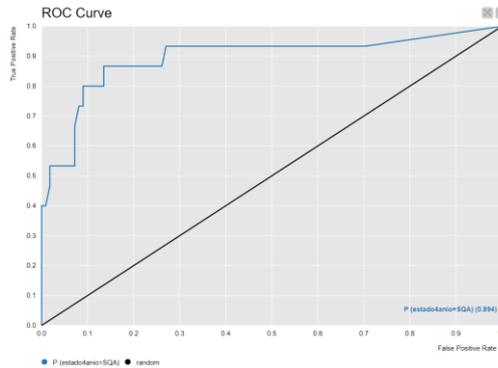
DESARROLLO DEL PROBLEMA

Curva ROC Regresión logística

Curva ROC Arbol decision



Curva ROC Random Forest



Coeficiente de Kappa

Modelo	Coeficiente Kappa	Grado de concordancia
Regresión Logística	0.029	Insignificante
Random Forest	0.54	Moderado
Árbol de decisión	0.74	Sustancial



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DESARROLLO DEL PROBLEMA



Depliegue

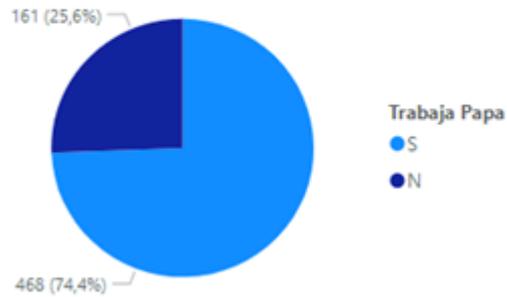
Finalmente, esta última etapa permitió hacer referencia a la implementación del modelo predictivo desarrollado. Para facilitar su uso y muestra, se propone la creación de un dashboard con los datos encontrados.



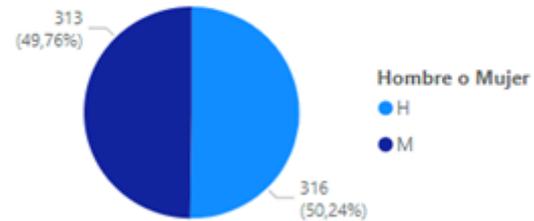
ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DESARROLLO DEL PROBLEMA

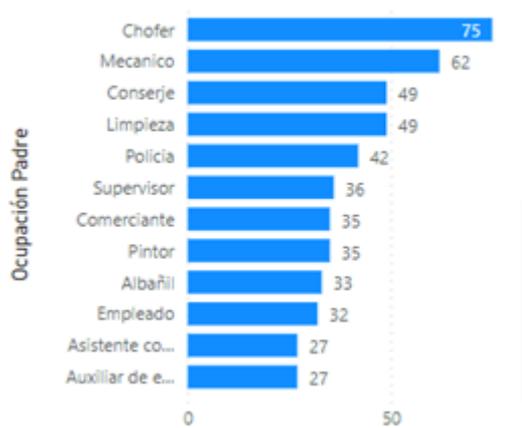
Porcentaje de Papás de estudiantes que trabajan



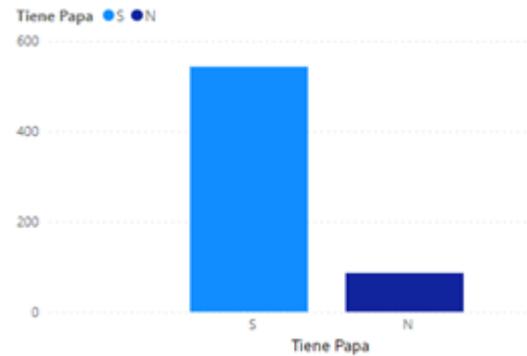
Porcentaje de estudiantes Hombres o Mujeres



Ocupacion de los padres de los estudiantes



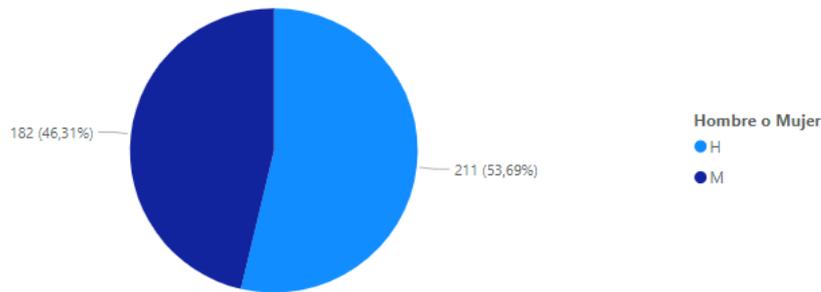
Alumnos que tienen Papá



DESARROLLO DEL PROBLEMA

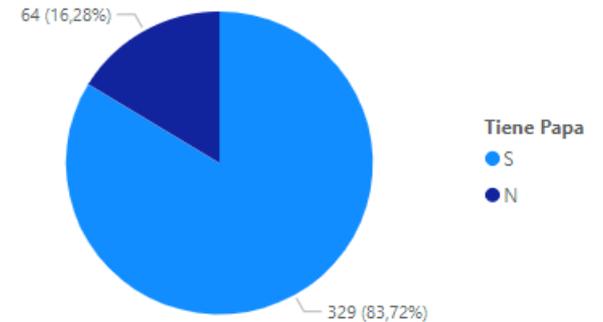
Proyeccion que se quede de año por sexo

Count of P (estado4anio=SQA) by Hombre o Mujer



Proyeccion que tenga Papa los estudiantes de 4to año que pasaron de año

Proyeccion que tengan Papa los estudiantes de 4to año que pasaron de año



CONCLUSIONES

- ❖ La revisión teórica abordada en la investigación permite inferir que el Business Inteligente es una herramienta poderosa para el proceso de toma de decisiones estratégicas para la Unidad Educativa “Aristóteles School” por medio del uso eficiente de los datos, además junto con la técnica de la minería de datos permite construir modelos eficientes para predecir un resultado esperado en los estudiantes, a partir de la información recogida se evidencio que la mayoría de estudiantes de 4to año de Colegio son propensos a perder el año comparando con otros grados, por tal motivo se realizó el modelo y predicción de datos sobre este grupo en específico.
- ❖ Mediante la información de los estudiantes de 4to año de Colegio se realizo el proceso de minería de datos, las características que influirán en la reprobación del año escolar son los padres en unión libre (47.58%) y la carencia de una figura paterna (83.72%).



CONCLUSIONES

- ❖ Para evaluar los modelos predictivos sobre el resultado de Aprobado o Desaprobado del año escolar de los estudiantes de la Unidad Educativa “Aristóteles School”, se empleó la matriz de confusión por medio de la cual se obtuvo que el modelo de regresión logística, tuvo una exactitud de 85%, árbol de decisión del 92% y Random Forest de 93%, respecto a la concordancia de la matriz de confusión se ha determinado un coeficiente Kappa de 0.029 para el modelo de regresión logística lo cual indica que tiene una concordancia insignificante, para el modelo Random Forest la concordancia fue de 0.54 siendo de grado moderado, y el modelo de árbol de decisiones obtuvo un valor Kappa de 0.74 siendo de grado sustancial.
- ❖ Basados en la métrica de la exactitud y el coeficiente Kappa, se concluye que el modelo predictivo que presenta mejores características tanto a nivel de proporción de observaciones que el clasificador predice correctamente y el grado de concordancia de los datos es el modelo del árbol de decisiones ya que obtuvo 92% y 0.74 respectivamente.



RECOMENDACIONES

- ❖ Si bien el modelo analizado presenta una tendencia de aprobados y reprobados, precisamente esto ayuda a la institución educativa a la toma de decisiones respecto a la calidad de la educación. Pero hay que tomar en cuenta que estos datos origen están en hojas Excel, por lo que se recomienda migrar toda esta información en un sistema gestor que utilice una base de datos.
- ❖ Se debe emplear información relevante correspondiente a las variables, es decir considerar las características que se consideran que influyen en este caso el rendimiento académico para de esta manera obtener mejores métricas de evaluación.



RECOMENDACIONES

- ❖ Se recomienda aplicar las técnicas de minería de datos complementada con información socioeconómica con la finalidad de mejorar el modelo de predicción en cuanto a los factores que influyen en la reprobación del año escolar.
- ❖ Los resultados de cada técnica de minería de datos dependen de los datos disponibles por lo que se recomienda obtener una gran cantidad de datos confiable y suficiente para un correcto análisis.

