



**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.**

Mantilla Galarza, Luis Antonio

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios

Trabajo de titulación, previo a la obtención del título de Magister en Gestión de Sistemas de  
Información e Inteligencia de Negocios

Ing. Delgado Rodríguez, Ramiro Nanac, PhD

16 de agosto de 2023



### Document Information

Analyzed document	TESIS_MANTILLA_ANTONIO_VFINAL.docx (D171142637)
Submitted	2023-06-22 17:39:00
Submitted by	Ramiro Delgado
Submitter email	pg.docenterdr@uniandes.edu.ec
Similarity	2%
Analysis address	pg.docenterdr.unia@analysis.urkund.com

### Sources included in the report

<b>SA</b>	<b>Tesis_vs24.0.pdf</b> Document Tesis_vs24.0.pdf (D29754380)	 2
<b>SA</b>	<b>Tesis_MSGIIN_DianaPoma_2017.docx</b> Document Tesis_MSGIIN_DianaPoma_2017.docx (D30682071)	 4
<b>SA</b>	<b>M2.878_20212_PEC4.2 - Redacci3n de la memoria (Entrega Final)_17762751.txt</b> Document M2.878_20212_PEC4.2 - Redacci3n de la memoria (Entrega Final)_17762751.txt (D139363672)	 2
<b>SA</b>	<b>40.pdf</b> Document 40.pdf (D108767047)	 1
<b>SA</b>	<b>TSP-JURGEN JOSTEIN ALVAREZ CHANCASANAMPA.pdf</b> Document TSP-JURGEN JOSTEIN ALVAREZ CHANCASANAMPA.pdf (D144084624)	 1
<b>SA</b>	<b>B0.487_20221_PEC 2: Realizaci3n/Implementaci3n del proyecto_18992068.txt</b> Document B0.487_20221_PEC 2: Realizaci3n/Implementaci3n del proyecto_18992068.txt (D155061985)	 2
<b>W</b>	URL: <a href="http://www.dspace.uce.edu.ec/bitstream/25000/18214/1/T-UCE-0010-FIL-331.pdf">http://www.dspace.uce.edu.ec/bitstream/25000/18214/1/T-UCE-0010-FIL-331.pdf</a> Fetched: 2022-03-21 19:37:27	 1
<b>SA</b>	<b>TESIS - INVESTIGACION Dina Leon.pdf</b> Document TESIS - INVESTIGACION Dina Leon.pdf (D142648347)	 2
<b>W</b>	URL: <a href="https://ivypanda.com/essays/education-technologies/">https://ivypanda.com/essays/education-technologies/</a> Fetched: 2022-06-16 05:34:14	 1

### Entire Document



Firmado electr3nicamente por:  
RAMIRO NANAC  
DELGADO RODRIGUEZ

Ing. Delgado Rodríguez, Ramiro Nanac, PhD

**DIRECTOR**



**Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología**

**Centro de Posgrados**

**Certificación**

Certifico que el trabajo de titulación, "**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.**" fue realizado por el señor **Mantilla Galarza Luis Antonio**; el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizado en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

**Sangolquí, 16 de agosto de 2013**



Delgado Rodríguez, Ramiro Nanac

**Director**

**C.C.: 170701917-8**



**Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología**

**Centro de Posgrados**

**Responsabilidad de Autoría**

Yo **Mantilla Galarza, Luis Antonio**, con cédula de ciudadanía n°1718852831, declaro que el contenido, ideas y criterios del trabajo de titulación: **"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."** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

**Sangolquí, 16 de agosto de 2023**

.....  
Mantilla Galarza Luis Antonio

C.C.: 171885283-1



**Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología**

**Centro de Posgrados**

**Autorización de Publicación**

Yo **Mantilla Galarza, Luis Antonio**, con cédula de ciudadanía n°171885283-1, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **“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.”** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

**Sangolquí, 16 de Agosto de 2023**

.....  
Mantilla Galarza Luis Antonio

C.C.: 171885283-1

## **Dedicatoria**

El resultado de este proyecto va dedicado con amor a mi hija Ruth Mantilla que me brindo gran motivo de alegría y ahora sé que me vera desde el cielo.

A mi esposa Alexandra, mi Padres Luis y Marina además de mi hermano que me brindo el apoyo necesario para culminar este trabajo.

## Agradecimiento

Agradezco a Dios por cada día cuidarme y darme la salud, la vida y por lo bondadoso que ha sido conmigo.

A mi ángel, mi hija Ruth Mantilla, gracias por brindarme el significado de la vida.

A mi esposa Alexandra Caiza, por todo su amor y apoyo incondicional.

A mis Padres Luis Mantilla y Marina Galarza por su amor y darme el significado de superarse como persona.

A la Universidad de las Fuerzas Armadas – ESPE, en especial a los coordinadores: Tatiana Gualotuña y Diego Marcillo

## Índice General

Índice General.....	8
Índice de tablas.....	10
Índice de figuras.....	12
Resumen.....	13
Abstract.....	14
Capítulo 1.....	15
Introducción.....	15
Antecedentes .....	15
Justificación e Importancia y Alcance.....	20
Planteamiento del Problema.....	22
<b>Objetivo General</b> .....	22
<b>Objetivos Específicos</b> .....	22
Capítulo 2.....	23
Marco Teórico.....	23
Introducción.....	23
Señalamiento de variables.....	23
<i>Fundamentación de la Variable Independiente</i> .....	24
<i>Fundamentación de la Variable dependiente</i> .....	33
Capítulo 3.....	34
Marco Metodológico.....	34
Metodología empleada.....	34
<i>Comprensión de los datos</i> .....	35

<i>Preparación de los datos</i> .....	45
<i>Modelado</i> .....	63
<i>Evaluación</i> .....	70
<i>Despliegue</i> .....	82
Capítulo 4.....	89
Análisis y resultados .....	89
Análisis de los datos obtenidos en la depuración de la base de datos .....	89
<i>Análisis socioeconómico de la familia de los estudiantes</i> .....	89
<i>Análisis rendimiento académico</i> .....	94
Conclusiones.....	104
Recomendaciones .....	105
Bibliografía .....	106

## Índice de tablas

<b>Tabla 1.-</b> Número de Estudiantes que pierden el año por año lectivo .....	18
<b>Tabla 2.-</b> Número de Estudiantes en Examen Remedial/Supletorio por año lectivo .....	19
<b>Tabla 3.-</b> Número de Estudiantes matriculados por año lectivo.....	36
<b>Tabla 4.-</b> Descripción de datos e información del estudiante: Lengua y Literatura y Matemática .....	38
<b>Tabla 5.-</b> Descripción de datos e información del estudiante; Ciencias Naturales y Estudios Sociales.....	39
<b>Tabla 6.-</b> Descripción de datos e información del estudiante; Dibujo y Música.....	40
<b>Tabla 7.-</b> Descripción de datos e información del estudiante; Educación Estética y Educación Física.....	41
<b>Tabla 8.-</b> Descripción de datos e información del estudiante; Lengua Extranjera y Computación .....	42
<b>Tabla 9.-</b> Descripción de datos e información del estudiante; segundo año de Bachillerato General Unificado nómina de alumnos .....	43
<b>Tabla 10.-</b> Descripción de datos e información del estudiante con relación a la información de la madre .....	44
<b>Tabla 11.-</b> Descripción de datos e información del estudiante con relación a la información del padre .....	45
<b>Tabla 12.-</b> Limpieza de datos.....	46
<b>Tabla 13.-</b> Limpieza de datos; notas.....	47
<b>Tabla 14.-</b> Limpieza de datos; notas.....	48
<b>Tabla 15.-</b> Limpieza de datos; notas remedial.....	49
<b>Tabla 16.-</b> Limpieza de datos; notas inglés .....	50
<b>Tabla 17.-</b> Limpieza de datos; notas inglés supletorio .....	51
<b>Tabla 18.-</b> Limpieza de datos; notas supletorio .....	52
<b>Tabla 19.-</b> Limpieza de datos; notas supletorio .....	53
<b>Tabla 20.-</b> Limpieza de datos; notas global .....	54
<b>Tabla 21.-</b> Limpieza de datos; notas global .....	55
<b>Tabla 22.-</b> Limpieza de datos; notas global (supletorio).....	56
<b>Tabla 23.-</b> Limpieza de datos; notas global (supletorio).....	57

<b>Tabla 24.-</b> Limpieza de datos; notas global (supletorio).....	58
<b>Tabla 25.-</b> Limpieza de datos; notas global (supletorio).....	59
<b>Tabla 26.-</b> Limpieza de datos; notas global (supletorio).....	60
<b>Tabla 27.-</b> Limpieza de datos; notas global (supletorio).....	61
<b>Tabla 28.-</b> Limpieza de datos; notas global (supletorio).....	62
<b>Tabla 29.-</b> Ventajas y desventajas de Random Forest .....	68
<b>Tabla 30.-</b> Ventajas y desventajas del árbol de decisión .....	69
<b>Tabla 31.-</b> Valoración del coeficiente Kappa .....	81
<b>Tabla 32.-</b> Resultados del coeficiente Kappa .....	82
<b>Tabla 33.-</b> Probabilidad de aprobación o reprobación de año escolar en estudiantes de cuarto año .....	87
<b>Tabla 34:</b> <i>Porcentaje de estudiantes aprobados y reprobados durante el periodo de estudio..</i>	94
<b>Tabla 35:</b> <i>Promedio por quimestre .....</i>	97
<b>Tabla 36:</b> <i>Promedio obtenido para el quimestre 1 .....</i>	99
<b>Tabla 37:</b> <i>Promedio obtenido para el segundo quimestre .....</i>	101

## Índice de figuras

<b>Figura 1.-</b> <i>Modelo Jerárquico de procesos CRISP-DM</i> .....	30
<b>Figura 2.-</b> <i>El proceso de la Metodología CRISP-DM</i> .....	35
<b>Figura 4.-</b> <i>Construcción del ETL</i> .....	65
<b>Figura 6.-</b> <i>Random Forest</i> .....	68
<b>Figura 7.-</b> <i>Árbol de Decisión</i> .....	70
<b>Figura 8.-</b> <i>Matriz de confusión</i> .....	71
<b>Figura 9.-</b> <i>Curva ROC de la Regresión Logística</i> .....	73
<b>Figura 10.-</b> <i>Matriz de confusión de la Regresión Logística</i> .....	73
<b>Figura 11.-</b> <i>Curva ROC de Random Forest</i> .....	75
<b>Figura 12.-</b> <i>Matriz de confusión de Random Forest</i> .....	76
<b>Figura 13.-</b> <i>Árbol de problema aplicado a los datos</i> .....	78
<b>Figura 14.-</b> <i>Curva ROC de Árbol de Decisión</i> .....	79
<b>Figura 15.-</b> <i>Matriz de confusión de Árbol de Decisión</i> .....	79
<b>Figura 16.-</b> <i>Ejecución del modelo y guardado del resultado de predicción en la Base de Datos</i> .....	83
<b>Figura 19.-</b> <i>Dashboard de BI (Business Intelligence)</i> .....	86
<b>Figura 21.-</b> <i>Proyección de aprobación o reprobación de año escolar en estudiantes de cuarto año</i> .....	88
<b>Figura 22.-</b> <i>Género de los estudiantes de la institución educativa</i> .....	89
<b>Figura 23.-</b> <i>Presencia del padre en el hogar</i> .....	90
<b>Figura 24.-</b> <i>Presencia de la madre en el hogar</i> .....	91
<b>Figura 25.-</b> <i>Estado civil de los padres</i> .....	92
<b>Figura 26.-</b> <i>Trabajo del padre</i> .....	93
<b>Figura 27.-</b> <i>Trabajo de la madre</i> .....	93
<b>Figura 28.-</b> <i>Porcentaje de aprobados y reprobados</i> .....	96
<b>Figura 29.-</b> <i>Promedio general para cada quimestre</i> .....	98
<b>Figura 31.-</b> <i>Promedio obtenido para el segundo quimestre</i> .....	102

## Resumen

En el ámbito educativo es primordial la toma de decisiones, en décadas pasadas se realizaban cuando finalizaba el año escolar, actualmente con la tecnología actual se puede predecir o tener una herramienta que nos ayude a mejorar la calidad de la educación en cualquier época del año lectivo. La presente tesis realiza recoge la información académica y social de los alumnos del Colegio Particular Aristóteles Bilingüe con esta información se realiza un proceso de limpieza de datos, para luego evaluar 3 modelos de predicción que son: regresión logística, Random Forest y Árbol de decisión. Construidos estos modelos se los evalúa mediante la matriz de confusión, la curva de ROC (Característica Operativa del Receptor) y el coeficiente de Kappa. Siendo el árbol de decisión el escogido por tener 92% de exactitud y 0,74 de coeficiente de Kappa que lo vuelve el más sustancial de acuerdo con el grado de concordancia de los valores precedentemente nombrados. Elegido el modelo de árbol de decisión se procede a ejecutar y guardar la predicción de que los alumnos 4to año aprueben o reprobren el año escolar. Además, se construye un Tablero de Inteligencia de Negocios con información relevante a la institución y también un apartado con información relevante a la proyección que se realizó en el modelo propuesto.

*Palabras clave:* minería de datos educativos, modelos de predicción, predicción de rendimiento académico, deserción escolar, tablero institucional escolar de inteligencia de negocios.

## Abstract

In the educational field it is essential to make decisions, in past decades they were made at the end of the school year, nowadays with current technology we can predict or have a tool to help us improve the quality of education at any time of the school year. This thesis collects the academic and social information of the students of the Colegio Particular Aristóteles Bilingüe with this information a process of data cleaning is carried out, to then evaluate 3 prediction models which are: logistic regression, Random Forest and Decision Tree. Once these models are built, they are evaluated by means of the confusion matrix, the ROC (Receiver Operating Characteristic) curve and the Kappa coefficient. Being the decision tree the one chosen for having 92% accuracy and 0.74 Kappa coefficient which makes it the most substantial according to the degree of agreement of the values previously mentioned. Once the decision tree model has been chosen, the prediction that 4th grade students will pass or fail the school year is executed and stored. In addition, a Business Intelligence Board is built with information relevant to the institution and a section with information relevant to the projection made in the proposed model.

*Keywords:* educational Data Mining, prediction models, academic performance prediction, school dropout, business intelligence school institutional board.

## **Capítulo 1**

### **Introducción**

#### **Antecedentes**

Los directivos de la institución educativa “Aristóteles Bilingüe” siempre se han propuesto mejorar la calidad de la educación en general, se analizan los resultados de los estudiantes al finalizar el año lectivo para la toma de decisiones a nivel administrativo. Si bien los profesores al entrar a la institución reciben una inducción y seguimiento esto no es suficiente ya que existen problemas con los alumnos y esto deteriora la imagen de la institución. Lo que ha motivado a las autoridades a buscar una alternativa para identificar de manera particular el nivel de rendimiento académico de los estudiantes y realizar un seguimiento continuo durante el año escolar; esto se lograría implementando una analítica de datos que permita identificar de manera temprana los problemas de bajo rendimiento para tomar acciones de mejora académica como tutorías o sesiones de refuerzo. Es importante anotar que la institución en la actualidad no dispone de ninguna herramienta de análisis de datos y las estadísticas se obtienen utilizando Excel y de manera reactiva; es decir al fin del año lectivo, lo que se traduce en propuestas que se aplican con nuevos estudiantes

El bajo rendimiento de los estudiantes, es el resultado de varios factores de riesgo, que surgen de la combinación y acumulación de muchas barreras y desventajas que afectan a los estudiantes a lo largo de sus vidas. Muchos niños, adolescentes y jóvenes, alrededor del mundo, están inmersos en este fenómeno tipo cascada que los sumerge en un abismo de desmotivación, debido a que los promedios de calificaciones obtenidos en pruebas escritas son inferiores a los

necesarios para aprobar el curso según lo determina el análisis estadístico comparativo entre varios países europeos y latinoamericanos con los que se ha declarado el bajo rendimiento académico en la región (Muñoz, 2018). La Organización para la Cooperación y el Desarrollo Económicos (OCDE) presentó un estudio trienal sobre los alumnos de 15 años de 72 países participantes denominado Programa para la Evaluación Internacional de Alumnos (PISA), que inicialmente se dio con países de economía media y actualmente Ecuador ha ingresado a este proceso en el que se analizarán los resultados académicos comparables con países latinoamericanos (Muñoz, 2018).

PISA ayuda a identificar las características de los sistemas educativos de mayor rendimiento, lo que puede permitir a gobiernos y educadores reconocer políticas educativas efectivas que se adapten a sus contextos locales, regionales, nacionales o mundiales.

PISA define: Estudiantes de “bajo rendimiento” como aquellos que puntúan por debajo del nivel 2 en las pruebas de Matemáticas, Lectura y Ciencias de PISA. El nivel 2 es considerado como el nivel básico de conocimiento que se requiere para participar plenamente en una sociedad moderna. Los estudiantes que puntúan en el nivel 1 pueden responder preguntas con instrucciones claras y relaciones sencillas que requieran utilizar una sola fuente de información, pero no pueden enfrentarse a la resolución de problemas que requieran razonamiento complejo (PISA IN FOCUS, 2015).

El tema de rendimiento escolar es un problema que preocupa a nivel global, sus estudios datan del siglo pasado, lo que motivó hacia inicios de siglo XXI a la realización de nuevos estudios de casos sobre escuelas outliers (valores atípicos) positivas y negativas hasta los más actuales. Esto fue llevado por “Investigación Iberoamericana sobre eficacia escolar” efectuada por la OEI (Organización de estados Iberoamericanos 2007) y Estudiantes de bajo rendimiento; al respecto de identificar por qué se quedan atrás y cómo se les puede ayudar. En lo que se plantean causas,

consecuencias y alternativas de solución al bajo rendimiento de los estudiantes, haciendo hincapié en las estadísticas, que son el referente de la existencia comprobada de esta problemática.

A nivel de país, el estado ecuatoriano invierte en educación el 4% del presupuesto anual del estado, que se destina entre otras cosas a sueldos de docentes, gastos de infraestructura, servicios básicos, servicios complementarios de colación escolar, uniformes y textos.

El 16 de julio de 1986 se firmó el acuerdo ministerial No 068 con el que legalmente se crea la Escuela particular "Aristóteles Bilingüe". El 7 de octubre de 1986 se inaugura el primer año lectivo 1986-1987, con la concurrencia de 30 alumnos de los diferentes barrios del sector; en especial de Guajalo, Guamaní, Chillogallo, La Ecuatoriana, Turubamba, El Beaterio y otros. Esto llevó a que progresivamente se creen los diferentes niveles hasta la consecución de los bachilleratos.

A partir del décimo año de vida, y hasta la fecha la institución "ARISTÓTELES BILINGÜE", cumpliendo con todos los requisitos pasa a ser miembro de FEDERAL. (FEDERACIÓN DE ESTABLECIMIENTOS PARTICULARES LAICOS).

Entre las principales actividades institucionales que se han venido desarrollando desde su creación están:

- La institución ha venido participando en mingas con la comunidad para el mejoramiento de los espacios verdes
- Desde inicios la escuela ha venido realizando cada año las demostraciones académicas con la participación de niños.
- La institución organiza eventos culturales como concursos de poemas, libro leído, ortografía, dibujo y pintura.

- Cada año los diferentes grados y cursos vienen realizando visitas de observaciones a diferentes lugares del centro histórico como museos, planetario, ferias de Tecnología y lugares representativos.
- En algunos años la institución ha sido invitada a participar a nivel inter escolar en concursos de libro leído obteniendo los primeros lugares.
- La institución ha participado en los festivales artísticos, convocados por la supervisión Provincial de Educación.
- Los estudiantes del colegio han venido exponiendo “Casa Abierta”, año tras año.
- Internamente se han organizado los juegos deportivos
- En el año lectivo 2003-2004 se concreta el Himno Institucional
- Para conmemorar el veinte y cinco aniversarios se organizó la Casa Abierta y la revista escolar, con la participación de todos quienes hacemos la gran familia ARISTÓTELES BILINGÜE.

En el contexto académico, se cuenta con información de la institución “Aristóteles School”, esta se presenta en la Tabla 1 la cual evidencia el total de estudiantes que pierden el año y en la Tabla 2 se presenta el número de estudiantes que deben rendir el examen remedial o supletorio, estas son un indicador del bajo rendimiento de los estudiantes.

**Tabla 1.-**

*Número de Estudiantes que pierden el año por año lectivo*

<b>Año Lectivo</b>	<b>Número de Estudiantes que pierden el año</b>
2011-2012	16

2012-2013	7
2014-2015	10

*Nota.* Colegio Aristóteles Bilingüe

**Tabla 2.-**

*Número de Estudiantes en Examen Remedial/Supletorio por año lectivo*

<b>Año Lectivo</b>	<b>Examen /Supletorio</b>	<b>Remedial</b>
2015-2016	11	
2016-2017	12	
2017-2018	25	
2018-2019	18	

*Nota.* Colegio Aristóteles Bilingüe

La minería de datos, también conocida como descubrimiento de conocimiento en Bases de datos (sus siglas en inglés son “KDD – Knowledge Discovery in Databases”), es el campo que nos permite descubrir información nueva y potencialmente útil de grandes cantidades de datos (Jimenez & Alvarez, 2010)

Esta tipología de método se destaca en la optimización de campañas de marketing. Identificar el riesgo de los estudiantes, detectar el fraude en el sector público y privado, analizar el crimen y el comportamiento peligroso, contrarrestar el terrorismo, clasificar y predecir pacientes en medicina, optimizar el reclutamiento de talentos y otras áreas donde hay grandes volúmenes. Es valioso extraer información de los datos. (Escobar et al., 2017).

La aplicación automatizada de algoritmos de Minería de Datos permite detectar patrones en los datos eficientemente y a partir de éstos, derivar información implícita en ellos, de tal forma que

se pueda comprobar qué tan útiles son las predicciones que se derivan de los datos para una organización en general. Dichos algoritmos de Minería de Datos se encuentran en continua evolución y se desarrollan como resultado de la colaboración entre campos de investigación tales como bases de datos, reconocimiento de patrones, inteligencia artificial, sistemas expertos, estadística, visualización, recuperación de información, y computación de altas prestaciones (García & Acevedo, 2010).

El análisis de grandes cantidades de datos para resolver problemas educativos ha llevado al desarrollo de la Minería de Datos Educativos (EDM). Es una disciplina que se dedica al desarrollo de métodos de exploración de datos de entornos relacionados con la educación en un esfuerzo por comprender mejor a los estudiantes, profesores y otras personas involucradas en su entorno educativo para mejorar los procesos y el aprendizaje. (Escobar et al., 2017).

Con la incorporación de las tecnologías de la información y la comunicación (TIC) en las instituciones educativas surge la acumulación de grandes cantidades de datos, que no son utilizados en la toma de decisiones, estos datos por su naturaleza requieren de un procesamiento sustancial que duplica los esfuerzos de trabajo, pero este proceso posibilita la generalización y la información procesos de enriquecimiento que permiten la adaptación de la información existente. La mayoría de los métodos de extracción se pueden aplicar directamente, mientras que otros deben adaptarse a problemas educativos específicos. Por otro lado, existe trabajo académico previo que orienta la aplicación de procedimientos específicos de minería de datos a problemas explícitos de aprendizaje (Romero & Ventura, 2010).

### **Justificación e Importancia y Alcance**

María del Carmen Caballero Ruíz, afirma: 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 (Ruíz, 2010).

¿En qué momento se inicia el fracaso como tal?... El debate se centra ahora en la secundaria obligatoria, pero hay que tener en cuenta que la escuela genera también un porcentaje de fracaso inducido que se inicia en la primaria. Existe la posibilidad que este aspecto pueda ser puntualizado desde el fracaso; también se debe revisar esta etapa en la cual no hay suficientes medidas compensadoras, ni tampoco una retención preferente por la diversidad. Es un fracaso temprano, camuflado muchas veces por la promoción automática que por una mayor interacción con el profesorado caracterizado por no asentar las destrezas básicas (lectura, escritura, procedimiento matemático básico) tras una serie de años de escolarización (Muñoz, 2018).

Es por tal motivo que 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.

La investigación requirió la recolección y análisis de la información con la finalidad de definir que las fuentes de datos a ser utilizadas sean confiables, así como también los datos con valores atípicos para que sean excluidos a fin que no alteren los resultados. A continuación, se desarrolló el dashboard de Business Intelligence con minería de datos, con el propósito de evaluar los resultados conjuntamente con los datos históricos. El Dashboard tomó en cuenta el 80% de la data de la base de datos obtenidos para el aprendizaje, el 20% restante se empleó para pruebas.

## **Planteamiento del Problema**

No cabe duda, que el bajo rendimiento académico constituye un problema para todo el sistema educativo y las instituciones inmersas en el mismo; pues, cuando el aprendizaje no alcanza niveles de calidad significa que hay debilidades en el proceso de aprendizaje y en el tiempo dedicado al trabajo personal o autónomo (Muñoz, 2018).

La unidad educativa “Aristóteles School”, en el transcurso del periodo 2009-2020 se evidenció una disminución en la media del rendimiento académico institucional lo que ha motivado a las autoridades a buscar una alternativa para identificar de manera particular el nivel de rendimiento académico de cada estudiante y desde luego realizar un seguimiento continuo durante el año escolar. Este aspecto se puede lograr implementando una herramienta de análisis de datos que permita identificar de manera temprana los problemas de bajo rendimiento y poder tomar acciones de mejora académica como tutorías o sesiones de refuerzo. Es importante anotar que la institución en la actualidad no dispone de ninguna herramienta de análisis de datos y las estadísticas se obtienen utilizando Excel y de manera reactiva; es decir, al final del quimestre o al fin de año, lo que se traduce en propuestas que se aplican con nuevos estudiantes.

## **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 un aprendizaje significativo.

## **Objetivos Específicos**

**A:** Realizar una revisión de literatura para determinar posibles soluciones y recomendaciones existentes para analizar el rendimiento académico de los alumnos de una institución educativa

**B:** 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”

**C:** Implementar un modelo orientado hacia la minería de datos con el propósito de predecir si un estudiante es propenso al supletorio

**D:** Evaluar el modelo implementado mediante protocolos de validación con técnicos, usuarios finales y autoridades

## **Capítulo 2**

### **Marco Teórico**

#### **Introducción**

#### **Señalamiento de variables**

Una vez elaborados los objetivos de investigación, se procede a definir la variable dependiente e independiente.

- Variable Independiente: Implementación de políticas y acciones de mejora académica basadas en un dashboard de business intelligence.
- Variable Dependiente: Rendimiento académico.

## ***Fundamentación de la Variable Independiente***

### **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)

Este tipo de análisis permite usar los datos asociados con el aprendizaje de los estudiantes y generar informes que sean útiles para los docentes (actividades y progreso de los estudiantes), para los estudiantes (retroalimentación) y para los administradores (incremento de aulas de clase, tasa de graduación, etc.). (Brown, 2011)

Learning analytics es la medición, recopilación, análisis y presentación de datos sobre los alumnos y sus contextos, a efectos de entender y optimizar el aprendizaje y los entornos en los que ocurren los sucesos de aprendizaje (Siemens, et al., 2011)

Por otra parte, Johnson y otros indican que el learning analytics busca aprovechar los avances en la minería de datos, la interpretación y modelación, para mejorar la comprensión de la enseñanza y el aprendizaje para adaptar eficazmente la educación a los estudiantes. (Johnson et al., 2011)

Según Siemens et al., (2011), estas definiciones nos permiten corroborar dos afirmaciones planteadas por

- El learning analytics utiliza datos pre-existentes.
- El learning analytics puede ser usado para manejar grandes conjuntos de datos, que no son posibles de manejar de forma manual.

## **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.

Es conocida la frase: “los datos en bruto raramente son beneficiosos directamente”. Su verdadero valor se basa en:

- (a) la habilidad para extraer información útil la toma de decisiones o la exploración; y,
- (b) la comprensión del fenómeno gobernante en la fuente de datos.

En muchos dominios, el análisis de datos fue tradicionalmente un proceso manual. Uno o más analistas familiarizados con los datos, con la ayuda de técnicas estadísticas, proporcionaban resúmenes y generaban informes. En efecto, el analista hacía de procesador de preguntas sofisticado. Sin embargo, tal enfoque cambió como consecuencia del crecimiento del volumen de datos. Cada vez es más común encontrarse con bases de los datos con un número de ejemplos del orden de  $10^9$  o superior y  $10^3$  dimensiones. Cuando la escala de manipulación de datos, exploración e inferencia va más allá de la capacidad humana, se necesita la ayuda de las tecnologías informáticas para automatizar el proceso (Riquelme et al., 2006).

## **Técnicas de Data Mining para Problemas Descriptivos**

El objetivo de la resolución de problemas descriptivos es encontrar los detalles de los datos. Por ejemplo, es común agrupar clientes con características similares para enviar notificaciones más personalizadas. Otro detalle puede ser encontrar asociaciones de productos que se venden juntos (Romero J. , 2019).. Las siguientes técnicas están disponibles para esto:

- **Análisis de Segmentación**

La intención de esta técnica es encontrar grupos similares, homogéneos en los datos. Para su solución se construyen modelos de aprendizaje no supervisados o clustering.

- **Análisis de Asociación**

El objetivo es obtener relaciones entre datos, relaciones desconocidas que tengan sentido. El ejemplo clásico más famoso es el análisis de los carritos de compras en los supermercados. Según los informes, Walmart utilizó el modelo de transacciones para descubrir que el viernes un grupo de clientes no solo trajo cerveza, sino que también compró pañales. Así que puso el pañal junto a la cerveza, aumentando así las ventas de ambos productos. (businessintelligence.info, 2009)

## **Técnicas Minería de Datos para problemas Predictivos**

Se usa para obtener modelos que se usarán para aplicarlo en datos futuros., esencialmente para predecir comportamientos. En inteligencia artificial se llaman modelos de aprendizaje supervisado. Las variables utilizadas pueden ser categóricas y numéricas (Romero J. , 2019).

## **A. Clasificación**

Se refiere a modelos en los que las variables a predecir tienen valores fijos y son contables. Se utilizan variables categóricas. Por ejemplo, se puede clasificar un evento en el tipo de visitantes como: empresarios, emprendedores y espectadores. Otro ejemplo es predecir si un cliente comprará un determinado producto.

## **B. Predicción de Valores**

Como dice su nombre, se intenta predecir valores, por lo que más ocupa son variables numéricas. Con frecuencia lo que se obtiene es la probabilidad de un evento: la probabilidad de que un cliente siga con su tarjeta de crédito, o de que adquiera una tarjeta adicional.

- **Algoritmos de Minería de datos**

### **Árboles de decisión**

#### **Redes neuronales artificiales**

Es una secuencia de decisiones dispuestas jerárquicamente, como las ramas de un árbol. Estos algoritmos aceptan datos numéricos y categóricos. Este algoritmo se aplica a menudo para tareas de clasificación, agrupación y predicción. Si predicen categorías, a menudo se denominan árboles de clasificación. Si son numéricos y están destinados a ser predichos, se denominan árboles de regresión.

- **Análisis en Componentes Principales**

Es una técnica multivariante, que se utiliza con el propósito de reducir la dimensionalidad del conjunto de datos. Es decir, si tenemos un gran número de variables  $X$  y  $N$ ; Obtenga suficientes

variables C y M mínimas; Cuáles serán los componentes principales que representan la información en los datos. Este algoritmo trabaja con variables numéricas.

- **Análisis Factorial**

Al igual que el análisis en componentes principales, el objetivo del análisis factorial también tiene como objetivo disminuir la dimensión de los datos. Es una especie de combinación que condensan la información completa. El algoritmo está diseñado para trabajar con variables cuantitativas.

- **Análisis de Correspondencia**

El análisis de correspondencias es una herramienta adecuada para realizar una tarea. Se utilizan dos vértices, análisis de correspondencia simple, que evalúa dos variables; Se basa en la tabla de contingencia. y el análisis de correspondencia múltiple, que considera más de dos variables considerando la tabla de Burt (Fuente , 2011).

- **Escalamiento Multidimensional**

Es usado para representar de manera gráfica a través de un mapa perceptual las similitudes que tienes objetos de una nube de datos, considerando el posicionamiento entre ellos. se parece mucho al análisis clúster. La diferencia radica en que en este modelo no se conocen las variables para determinar la semejanza, mientras que en el clúster sí (Fuente , 2011)..

- **El algoritmo K-means**

La agrupación en clústeres de K-means es uno de los algoritmos de aprendizaje automático no supervisado más simples y populares. En general, los algoritmos no supervisados hacen inferencias a partir de un conjunto de datos usando solo vectores de entrada sin mencionar resultados conocidos o etiquetados.

- **Algoritmo del vecino k más cercano**

K-Nearest Neighbors se denota por ser uno de los algoritmos de clasificación más básicos pero esenciales en el aprendizaje automático. Pertenece al campo del aprendizaje supervisado y encuentra aplicaciones profundas en el reconocimiento de patrones, minería de datos y detección de intrusos. En los algoritmos de minería de datos, es uno de los más famosos (Chaud, 2016).

- **Naive Bayes**

Naive Bayes es un algoritmo probabilístico de aprendizaje automático basado en el Teorema de Bayes, utilizado en una amplia variedad de tareas de clasificación. En este post, obtendrá una comprensión clara y completa del algoritmo Bayes Ingenuo y todos los conceptos necesarios para que no haya lugar a dudas ni a lagunas en la comprensión.

- **Algoritmo CART**

El algoritmo de CART está estructurado como una secuencia de preguntas, cuyas respuestas determinan cuál será la siguiente pregunta, si es que la hay. El resultado de estas preguntas es una estructura en forma de árbol donde los extremos son nodos terminales en cuyo punto no hay más preguntas.

Los principales elementos de CART (y cualquier algoritmo de árbol de decisión) son:

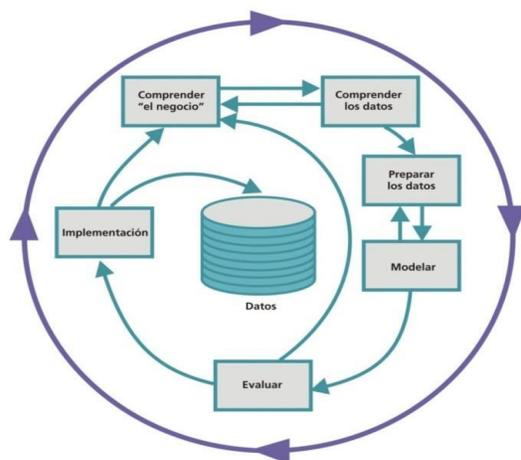
- Reglas para dividir los datos en un nodo basadas en el valor de una variable;
- Reglas de parada para decidir cuándo una rama es terminal y ya no se puede dividir;
- y
- Finalmente, una predicción para la variable objetivo en cada nodo terminal.

## Metodología cross-industry standart process for data mining (crisp-dm)

El CRISP-DM fue diseñado por DaimlerChrysler a fines de 1996. Es importante resaltar que este visionario ya estaba por delante de la mayoría de las organizaciones industriales y comerciales en la aplicación de minería de datos en sus operaciones comerciales. SPSS (entonces ISL) había brindado servicios basados en minería de datos desde 1990 y lanzó el primer banco comercial de minería de datos Clementine en 1999 (Giraldo & Jiménez, 2013), el modelo Jerárquico de procesos CRISP-DM se puede observar en la Figura 1.

**Figura 1.-**

*Modelo Jerárquico de procesos CRISP-DM*



*Nota.* La figura mostrada responde al nombre de Metodología CRISP-DM. Elaborado por: (Cárdenas et al., 2018).

El énfasis se encuentra en procesos de alto nivel que sean lo suficientemente generales para cubrir todas las posibles aplicaciones de explotación de la información. Esta metodología define un ciclo de vida del proyecto de explotación de la información que define las principales fases de dicho proyecto (Giraldo & Jiménez, 2013).

## Etapas de la metodología CRISP-DM

La metodología CRISP-DM consta de seis etapas:

**1. *Comprensión del problema o negocio:*** Esta es la etapa más importante, ya que, si no se tiene una correcta comprensión del problema, o negocio, de nada servirán las etapas siguientes. Las actividades principales de esta etapa son:

- *Identificación del problema:* Consiste en entender y delimitar la problemática, así como identificar los requisitos, supuestos, restricciones y beneficios del proyecto.
- *Determinación de objetivos:* Puntualiza las metas a lograr al proponer una solución basada en un modelo de minería de datos.
- *Evaluación de la situación actual:* Especifica el estado actual antes de implementar la solución de minería de datos propuesta, a fin de tener un punto de comparación que permita medir el grado de éxito del proyecto.

**2. *Comprensión de datos:*** Las actividades principales de esta etapa son:

- *Recolección de datos:* Consiste en obtener los datos a utilizar en el proyecto identificando las fuentes y las técnicas empleadas.
- *Descripción de datos:* Identifica el tipo, formato, volumetría y significado de cada dato.
- *Exploración de datos:* Radica en aplicar pruebas estadísticas básicas que permitan conocer las propiedades de los datos a fin de entenderlos lo mejor posible.

**3. *Preparación de datos:*** Generalmente esta es la etapa que consume más tiempo en el proyecto, y es donde se seleccionan los datos que se transforman de acuerdo con los resultados de la etapa anterior a fin de utilizarlos en la etapa de modelado (Cortina , 2015).

Las actividades principales de esta etapa son:

- *Limpieza de datos*: Aplicación de diferentes técnicas, por ejemplo, normalización de datos, discretización de campos numéricos, tratamiento de valores ausentes, tratamiento de duplicados e imputación de datos.
  - *Creación de indicadores*: Genera indicadores que potencien la capacidad predictiva de los datos a partir de los datos existentes y ayuden a detectar comportamientos interesantes para modelar.
  - *Transformación de datos*: Cambia el formato o estructura de ciertos datos sin modificar su significado, a fin de aplicarles alguna técnica particular en la etapa de modelado.
- 4. Modelado:** En esta etapa se obtiene propiamente el modelo de minería de datos. Sus actividades principales son:
- *Selección de técnica de modelado*: Elige la técnica apropiada de acuerdo con el problema a resolver, los datos disponibles, las herramientas de minería de datos disponibles, así como el dominio de la técnica elegida.
  - *Selección de datos de prueba*: En algunos tipos de modelos se requiere dividir la muestra en datos de entrenamiento y de validación.
  - *Obtención del modelo*: Genera el mejor modelo mediante un proceso iterativo de modificación de parámetros del mismo.
- 5. Evaluación del Modelo:** durante esta fase, la calidad del modelo se determina con base en el análisis de algunas medidas estadísticas del modelo, la comparación de resultados con resultados anteriores o el análisis de resultados con el apoyo de expertos en el área del problema. En base a los resultados de esta fase, se toma la decisión de continuar con

la fase final de la metodología, volver a una de las fases anteriores o incluso empezar de cero con un nuevo proyecto.

- 6. *Implementación del modelo:*** este paso explota el conocimiento adquirido a través del modelo a través de funciones concretas. También es importante documentar los resultados de manera clara para el usuario final y asegurarse de que todos los pasos de la metodología estén debidamente documentados para examinarlos y aprender del proyecto. De igual manera, monitorear stocks para detectar áreas de oportunidad o nuevos problemas (Espinosa, 2020)

### ***Fundamentación de la Variable dependiente***

#### **Rendimiento Académico**

El rendimiento escolar es un problema que preocupa profundamente a estudiantes, padres de familia, docentes y autoridades; Y no solo en nuestro país, sino también en muchos otros países de América Latina y otros continentes. La complejidad del rendimiento académico comienza con su concepto. A veces se habla de mérito académico, rendimiento académico o éxito académico, pero en general, la diferencia de concepto se explica solo por cuestiones semánticas, ya que se utilizan como sinónimos. Por convención se ha estipulado que se debe utilizar el rendimiento académico en la población universitaria y el rendimiento académico en la población de educación básica regular y alternativa (Lamas, 2015).

De hecho, la evaluación del rendimiento escolar tiene un doble interés: por un lado, indica en qué medida los alumnos logran los aprendizajes por los que se esfuerzan al máximo; Por otro lado, permite conocer la eficacia de la escolarización, ya que no es fácil que la escuela logre objetivos complejos e intangibles como la adquisición de valores, la formación del carácter, la construcción de hábitos de estudio y trabajo, el amor a la cultura, entre otros. Si no logra, al menos, objetivos menos complejos y más concretos, como los fines del aprendizaje, que en todo caso son siempre

absolutamente necesarios tanto por la tradición como por los actores de la educación y de la sociedad en su conjunto (Lamas, 2015).

## Capítulo 3

### Marco Metodológico

Es imprescindible elegir una metodología que permita alcanzar los objetivos planteados. La aplicación del procedimiento apropiado ayuda a realizar la investigación de forma estructurada y ordenada. Adicionalmente permite definir el uso de métodos, técnicas, instrumentos, estrategias y procedimientos que posibilitan la consecución del proyecto planteado. La acertada selección de los métodos que se adapten a las necesidades del proyecto de investigación es un factor fundamental pues facilita el desarrollo de las actividades y desde luego dar respuesta a los objetivos planteados.

#### Metodología empleada

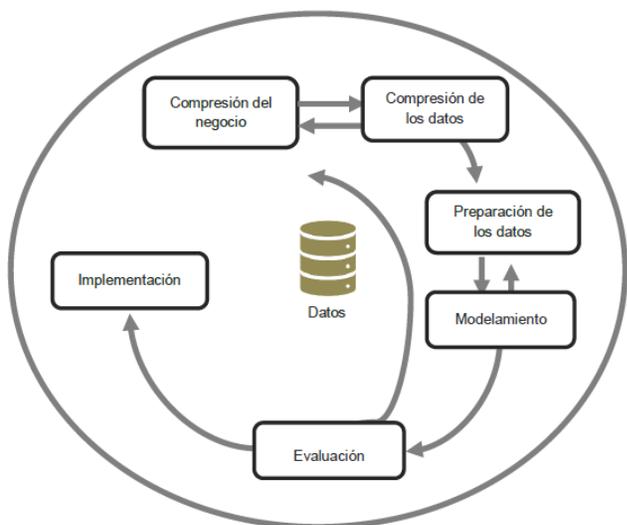
Cross Industry Standard Process for Data Mining (CRISP-DM) es un modelo de proceso, de minería de datos que describe una manera en la que los expertos en esta materia abordan el problema, además de ser una de las metodologías que ha tenido más apoyo de las empresas privadas y organismos públicos (Galán, 2015).

Este método se edifica en 6 fases, estas se pueden observar de forma gráfica en la Figura 2, las cuales permiten la organización y desarrollo de un proyecto de Data Mining, la sucesión de fases no tiene que ser de manera ordenada, por lo que algunas son bidireccionales. Es decir que de

una fase en concreto se puede volver a una fase anterior para poder revisarla (Gondar Nores, 2004).

**Figura 2.-**

*El proceso de la Metodología CRISP-DM*



*Nota.* Proceso de la metodología CRISP-DM Elaborado por: (Mancilla et al., 2020)

### ***Comprensión de los datos***

Esta etapa, como su nombre lo indica, trata de comprender y contextualizar la situación sobre la cual se va a desarrollar el trabajo. Es decir, involucra de manera completa a la institución objeto de estudio. Para lograrlo, primero se definieron las características puntuales de los individuos que fueron evaluados, los cuales fueron 5567 alumnos, los cuales corresponden al número total de alumnos desde el año 2009 al 2020, esto se evidencia en la Tabla 3. Como segundo punto 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 Escuela y Colegio Aristóteles Bilingüe es una institución ubicada en el Sur de Quito (Ecuador) en el sector de Guajalo, destacada por la enseñanza del inglés en los primeros niveles de educación básica. En su larga trayectoria posee alrededor de 37 años de vida institucional y es una de las primeras instituciones creadas en el sur de Quito.

**Tabla 3.-**

*Número de Estudiantes matriculados por año lectivo*

<b>Año Lectivo</b>	<b>Número de Estudiantes</b>
2009-2010	728
2010-2011	714
2011-2012	724
2012-2013	774
2013-2014	578
2014-2015	491
2015-2016	408
2016-2017	347
2017-2018	274
2018-2019	260
2019-2020	269
<b>Total</b>	<b>5567</b>

*Nota.* Colegio Aristóteles Bilingüe

## **Recolección de datos**

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. Esta información fue recabada con el consentimiento del rector de la institución puesto que al ser información de índole privada es necesario que esta sea obtenida por medio de todos los reglamentos y firmas de consentimiento en caso de ser necesario. Es necesario acotar que la información en crudo fue obtenida por medio de una hoja de cálculo en Excel.

A continuación, en la Tabla 4, Tabla5, *Nota*. Colegio Aristóteles Bilingüe

Tabla 6, Tabla 7, Tabla 8 se presentan la base de datos obtenida en la que se presenta la información a ser analizada.

Tabla 4.-

Descripción de datos e información del estudiante: Lengua y Literatura y Matemática

OCTAVO AÑO DE EDUCACIÓN GENERAL BÁSICA SUPERIOR																																				
ORD	APELLIDOS Y NOMBRES	LENGUA Y LITERATURA													MATEMÁTICA																					
		I Bloque	II Bloque	III Bloque	PROMEDIO BLOQUES	0,80	Examen Quimestral	PROME. FINAL	PROM. FINAL SIN FORMULA	I Bloque	II Bloque	III Bloque	PROMEDIO BLOQUES	0,80	Examen Quimestral	PROM. FINAL	PROM. FINAL SF.	TA FINAL	PLETORIO	I Bloque	II Bloque	III Bloque	PROMEDIO BLOQUES	0,80	Examen Quimestral	PROME. FINAL	PROM. FINAL SIN FORMULA	I Bloque	II Bloque	III Bloque	PROMEDIO BLOQUES	0,80	Examen Quimestral	PROM. FINAL	PROM. FINAL SF.	TA FINAL
1	Nómina de Estudiantes	7,96	7,27	6,32	7,38	5,91	1,54	7,45	7,45	8,33	8,69	8,53	8,52	6,81	1,60	8,41	8,41	7,93		8,76	8,48	8,50	8,58	6,86	2,00	8,86	8,86	8,94	8,66	8,60	8,73	6,99	2,00	8,99	8,99	8,93
2		6,30	5,70	5,40	5,80	4,64	1,40	6,04	6,04	4,67	5,52	6,60	5,60	4,48	1,60	6,08	6,08	6,06	ETOPRIO	7,00	4,80	7,90	6,57	5,25	1,80	7,05	7,05	7,52	6,06	6,76	6,78	5,42	1,74	7,16	7,16	7,11
3		9,19	8,10	8,50	8,60	6,88	1,80	8,68	8,68	7,87	8,51	8,02	8,13	6,51	1,80	8,31	8,31	8,50		8,34	9,68	8,44	8,82	7,06	2,00	9,06	9,06	8,46	8,58	8,06	8,37	6,69	1,50	8,19	8,19	8,63
4		7,00	7,00	4,90	6,30	5,04	1,12	6,16	6,16	3,97	7,26	6,00	5,74	4,59	1,60	6,19	6,19	6,18	ETOPRIO	6,40	6,40	7,60	6,80	5,44	1,56	7,00	7,00	7,26	8,14	6,72	7,37	5,90	1,56	7,46	7,46	7,23
5		8,14	7,50	7,40	7,68	6,14	1,80	7,94	7,94	7,28	8,16	8,40	7,95	6,36	2,00	8,36	8,36	8,15		8,60	8,76	8,00	8,45	6,76	2,00	8,76	8,76	8,14	8,36	7,26	7,92	6,34	1,56	7,90	7,90	8,33
6		6,84	6,93	3,66	5,81	4,65	0,20	4,85	4,85	4,34	1,00	1,00	2,11	1,69	1,00	2,69	2,69	3,77	MEDIAL	7,84	7,46	6,92	7,41	5,93	1,20	7,13	7,13	7,82		7,82	6,34		6,34		7,13	
7		7,71	7,66	7,64	7,67	6,14	1,78	7,92	7,92	8,42	7,63	8,17	8,07	6,46	2,00	8,46	8,46	8,19		8,58	9,20	9,74	9,17	7,34	2,00	9,34	9,34	9,30	9,38	9,47	9,38	7,51	2,00	9,51	9,51	9,43
8		8,10	7,71	7,66	7,82	6,26	1,52	7,78	7,78	8,47	8,05	8,07	8,20	6,56	2,00	8,56	8,56	8,17		8,82	8,96	8,96	8,91	7,13	1,86	8,99	8,99	9,52	9,04	8,84	9,13	7,31	1,44	8,75	8,75	8,87

Nota. Colegio Aristóteles Bilingüe

Tabla 5.-

Descripción de datos e información del estudiante; Ciencias Naturales y Estudios Sociales

OCTAVO AÑO DE EDUCACIÓN GENERAL BÁSICA		CIENCIAS SOCIALES															CIENCIAS NATURALES																		
ORD	APELLIDOS Y NOMBRES	I Bloque	II Bloque	III Bloque	PROMEDIO BLOQUES	0,80	Examen Quimestral	PROME. FINAL	PROM. FINAL SIN FORMULA	I Bloque	II Bloque	III Bloque	PROMEDIO BLOQUES	0,80	Examen Quimestral	PROM. FINAL	PROM. FINAL SF.	TA FINPLETORIO	I Bloque	II Bloque	III Bloque	PROMEDIO BLOQUES	0,80	Examen Quimestral	PROME. FINAL	Promedio Digitado	I Bloque	II Bloque	III Bloque	PROMEDIO BLOQUES	0,80	Examen Quimestral	PROM. FINAL	PROM. FINAL SF.	TA FINPLETORIO
		1	Nómina de Estudiantes	7,32	7,87	8,98	8,06	6,45	1,76	<b>8,21</b>	8,21	8,80	8,85	9,00	8,88	7,11	1,87	<b>8,98</b>	8,98	<b>8,60</b>	8,44	8,32	7,02	7,93	6,34	1,50	<b>7,84</b>	7,84	8,98	8,22	8,82	8,67	6,94	1,76	<b>8,70</b>
2	5,00	5,00		3,40	4,47	3,57	0,80	<b>4,37</b>	4,37	8,40	8,10	8,77	8,42	6,74	1,65	<b>8,39</b>	8,39	<b>6,38</b>	7,31	5,25	7,78	6,78	5,42	1,60	<b>7,02</b>	7,02	7,58	4,80	6,56	6,31	5,05	1,68	<b>6,73</b>	6,73	<b>6,88</b>
3	7,25	7,51		9,35	8,37	6,70	1,72	<b>8,42</b>	8,42	8,40	8,90	8,76	8,69	6,95	1,85	<b>8,80</b>	8,80	<b>8,61</b>	7,26	8,52	8,24	8,01	6,41	2,00	<b>8,41</b>	8,41	7,84	8,38	7,60	7,94	6,35	1,66	<b>8,01</b>	8,01	<b>8,21</b>
4	6,84	6,31		8,84	8,00	6,40	1,52	<b>7,92</b>	7,92	6,17	7,75	8,01	7,31	5,85	1,32	<b>7,17</b>	7,17	<b>7,55</b>	6,38	6,38	7,92	6,89	5,51	1,50	<b>7,01</b>	7,01	7,20	7,20	6,44	6,95	5,56	0,44	<b>6,00</b>	6,00	<b>6,51</b>
5	8,42	8,04		9,38	8,28	6,62	1,50	<b>8,12</b>	8,12	8,35	9,10	9,20	8,88	7,11	2,00	<b>9,11</b>	9,11	<b>8,62</b>	6,72	8,52	7,52	7,59	6,07	1,76	<b>7,83</b>	7,83	8,68	7,44	7,60	7,91	6,33	1,15	<b>7,48</b>	7,48	<b>7,66</b>
6	7,50	7,20		8,52	7,41	5,93	1,00	<b>6,93</b>	6,93	7,42			7,42	5,94		<b>5,94</b>		<b>6,93</b>	8,22	<b>8,44</b>	6,62	7,76	6,21	1,10	<b>7,31</b>	7,31									<b>7,31</b>
7	8,72	8,32		9,52	8,85	7,08	1,80	<b>8,88</b>	8,88	8,60	9,65	9,34	9,20	7,36	1,95	<b>9,31</b>	9,31	<b>9,10</b>	8,46	8,56	8,46	8,49	6,79	1,76	<b>8,55</b>	8,55	8,78	8,58	9,04	8,80	7,04	2,00	<b>9,04</b>	9,04	<b>8,80</b>
8	8,74	8,00		9,28	8,34	6,67	1,64	<b>8,31</b>	8,31	8,80	9,25	9,61	9,22	7,38	1,85	<b>9,23</b>	9,23	<b>8,77</b>	8,40	8,66	7,84	8,30	6,64	2,00	<b>8,64</b>	8,64	8,46	8,48	8,76	8,57	6,85	1,80	<b>8,65</b>	8,65	<b>8,65</b>

Nota. Colegio Aristóteles Bilingüe



Tabla 7.-

Descripción de datos e información del estudiante; Educación Estética y Educación Física

OCTAVO AÑO DE EDUCACION GENERAL BASICA SUPERIOR		EDUCACIÓN ESTÉTICA														EDUCACIÓN FÍSICA																				
No	NOMINA	PRIMER QUIMESTRE							SEGUNDO QUIMESTRE							PRIMER QUIMESTRE							SEGUNDO QUIMESTRE													
		I Bloque	II Bloque	III Bloque	PROMEDIO BLOQUES	80%	Examen Quimestral	PROM. FINAL I QUIMESTRE	PROM. FINAL I QUIM SIN FORMULA	I Bloque	II Bloque	III Bloque	PROMEDIO BLOQUES	80%	Examen Quimestral	PROM. FINAL II QUIMESTRE	PROM. FINAL II QUIM SIN FORMULA	I Bloque	II Bloque	III Bloque	PROMEDIO BLOQUES	80%	Examen Quimestral	PROM. FINAL I QUIMESTRE	PROM. FINAL I QUIM SIN FORMULA	I Bloque	II Bloque	III Bloque	PROMEDIO BLOQUES	80%	Examen Quimestral	PROM. FINAL II QUIMESTRE	PROM. FINAL II QUIM SIN FORMULA			
1		6,55	6,90	7,54	7,00	5,60	1,10	6,70	6,70	7,86	7,00							9,60	9,60	8,00	9,07	7,25	1,60	8,85	8,85	8,00	8,80							8,40	6,72	6,72
2		RETIRADO																																		
3		6,15	7,00	6,75	6,63	5,31	0,70	6,01	6,01	7,62	7,70							9,60	8,40	8,80	8,93	7,15	2,00	9,15	9,15	8,20	8,80							8,50	6,80	6,80
4		8,65	9,50	9,21	9,12	7,30	1,80	9,10	9,10	9,26	9,16							9,00	9,40	8,80	9,07	7,25	1,40	8,65	8,65	8,20	8,40							8,30	6,64	6,64
5		6,10																																		
6		7,95	8,36	8,29	8,20	6,56	1,40	7,96	7,96	8,21	7,92							9,00	9,20	8,60	8,93	7,15	1,40	8,55	8,55	7,60	7,60							7,60	6,08	6,08
7		8,35	8,60	8,04	8,33	6,66	1,00	7,66	7,66	8,76	8,08							8,80	9,20	9,40	9,13	7,31	1,40	8,71	8,71	8,00	8,20							8,10	6,48	6,48
8	Nómina de	8,15	8,38	8,35	8,29	6,63	1,20	7,83	7,83	8,67	8,47							7,20	9,20	9,00	8,47	6,77	1,40	8,17	8,17	7,80	8,40							8,10	6,48	6,48
9	Estudiantes	6,70	6,55	7,20	6,82	5,45	1,40	6,85	6,85	7,91	7,65							9,40	9,40	9,20	9,33	7,47	2,00	9,47	9,47	8,00	9,00							8,50	6,80	6,80
10		7,60	8,25	7,53	7,79	6,23	1,30	7,53	7,53	7,54	7,52							9,00	9,20	7,40	8,53	6,83	1,40	8,23	8,23	8,40	8,20							8,30	6,64	6,64
11		RETIRADO																																		
12		7,68	7,98	8,00	7,88	6,31	1,20	7,51	7,51	7,60	7,51							9,20	9,60	8,80	9,20	7,36	1,60	8,96	8,96	8,40	9,40							8,90	7,12	7,12
13		7,45	7,35	7,46	7,42	5,94	1,40	7,34	7,34	7,68	7,39							9,40	8,80	8,60	8,93	7,15	1,40	8,55	8,55	9,60	9,40							9,50	7,60	7,60
14		8,66	8,64	8,19	8,50	6,80	1,63	8,42	8,42	8,89	8,26							8,60	9,20	8,40	8,73	6,99	1,40	8,39	8,39	8,00	8,80							8,40	6,72	6,72
15		8,20	8,08	7,86	8,05	6,44	1,40	7,84	7,84	7,78	7,43							9,60	9,60	8,80	9,33	7,47	2,00	9,47	9,47	9,60	9,80							9,70	7,76	7,76
16		6,65	6,75	6,63	6,68	5,34	1,00	6,34	6,34	7,48	6,80							9,80	9,40	8,80	9,33	7,47	1,60	9,07	9,07	8,40	8,80							8,60	6,88	6,88
17		6,40	6,75	6,63	6,59	5,27	0,80	6,07	6,07	7,66	6,95							9,00	9,20	8,40	8,87	7,09	1,40	8,49	8,49	7,60	7,00							7,30	5,84	5,84
18		9,40	8,65	8,15	8,73	6,99	1,50	8,49	8,49	8,87	8,36							9,40	9,40	8,60	9,13	7,31	1,80	9,11	9,11	8,40	8,60							8,50	6,80	6,80
19		6,90	7,55	7,53	7,33	5,86	1,50	7,36	7,36	7,55	5,28							9,60	9,60	8,60	9,27	7,41	1,80	9,21	9,21	8,00	8,60							8,30	6,64	6,64
20		5,20	6,08	6,55	5,94	4,75	1,00	5,75	5,75	7,24								9,00	9,40	8,20	8,87	7,09	1,40	8,49	8,49	8,20							8,20	6,56	6,56	

Nota. Colegio Aristóteles Bilingüe

Tabla 8.-

Descripción de datos e información del estudiante; Lengua Extranjera y Computación

OCTAVO AÑO DE EDUCACION GENERAL BASICA SUPERIOR		LENGUA EXTRANJERA												OPTATIVA (COMPUTACIÓN)												PROMEDIO GENERAL		PROMEDIO ANUAL						
No	NOMINA	PRIMER QUIMESTRE				80%	20%	PROM. FINAL I QUIMESTRE	PROM. FINAL I QUIMESTRE FORMULA	SEGUNDO QUIMESTRE					20%	PROM. FINAL I QUIMESTRE	PROM. FINAL I QUIMESTRE FORMULA	PRIMER QUIMESTRE				80%	20%	PROM. FINAL I QUIMESTRE	PROM. FINAL I QUIMESTRE FORMULA	SEGUNDO QUIMESTRE			20%	PROM. FINAL I QUIMESTRE	PROM. FINAL I QUIMESTRE FORMULA	10	20	
		I Bloque	II Bloque	III Bloque	PROMEDIO BLOQUES	Examen Quimestral	I Bloque			II Bloque	III Bloque	PROMEDIO BLOQUES	80%	Examen Quimestral	I Bloque			II Bloque	III Bloque	PROMEDIO BLOQUES	80%	Examen Quimestral	I Bloque			II Bloque	III Bloque		PROMEDIO BLOQUES					80%
1		7,51	8,23	7,40	7,71	6,17	1,40	7,57	7,57	7,90	8,43				8,12	6,49	6,49	5,61	7,71	7,08	6,80	5,44	0,90	6,34	6,34	6,86	6,33	6,60	5,28	5,28		7,55		
2		RETIRADO																								RETIRADO								
3		6,21	7,16	6,45	6,61	5,29	0,73	6,02	6,02	5,63	8,20			6,92	5,53	5,53	6,17	7,49	7,37	7,01	5,61	1,10	6,71	6,71	6,52	5,51	6,02	4,81	4,81		7,10			
4		9,78	9,93	10,00	9,90	7,92	2,00	9,92	9,92	9,86	10,00			9,93	7,94	7,94	9,45	9,65	9,60	9,57	7,65	2,00	9,65	9,65	8,80	8,90	8,85	7,08	7,08		9,43			
5		6,03	RETIRADO		6,03	4,82		4,82									5,64			5,64	4,51		4,51			RETIRADO								
6		8,16	8,87	8,73	8,59	6,87	1,22	8,09	8,09	7,04	8,53			7,79	6,23	6,23	8,11	8,70	7,62	8,14	6,51	1,00	7,51	7,51	8,40	6,15	7,28	5,82	5,82		8,28			
7		7,80	8,49	8,34	8,21	6,57	1,56	8,13	8,13	7,61	8,90			8,26	6,60	6,60	7,73	8,29	7,71	7,91	6,33	1,40	7,73	7,73	8,26	7,77	8,02	6,41	6,41		8,22			
8		7,08	7,90	7,60	7,53	6,02	1,10	7,12	7,12	7,18	8,57			7,88	6,30	6,30	7,09	8,10	7,92	7,70	6,16	1,00	7,16	7,16	7,36	6,58	6,97	5,58	5,58		7,78			
9		8,59	8,48	7,76	8,28	6,62	1,40	8,02	8,02	8,19	8,49			8,34	6,67	6,67	7,00	8,77	7,21	7,66	6,13	0,90	7,03	7,03	7,52	6,77	7,15	5,72	5,72		7,87			
10		7,88	8,59	8,44	8,30	6,64	1,05	7,69	7,69	6,91	7,69			7,30	5,84	5,84	6,49	8,75	7,33	7,52	6,02	1,46	7,48	7,48	6,66	5,87	6,27	5,01	5,01		7,88			
11		RETIRADO																								RETIRADO								
12		7,16	8,17	7,28	7,54	6,03	1,22	7,25	7,25	7,37	8,15			7,76	6,21	6,21	7,18	8,15	7,48	7,60	6,08	0,92	7,00	7,00	7,18	6,22	6,70	5,36	5,36		7,64			
13		8,70	9,22	8,42	8,78	7,02	1,56	8,58	8,58	8,36	8,56			8,46	6,77	6,77	7,75	8,29	7,79	7,94	6,35	1,46	7,81	7,81	7,36	7,92	7,64	6,11	6,11		8,13			
14		8,64	7,96	9,42	8,67	6,94	1,69	8,63	8,63	8,20	9,25			8,73	6,98	6,98	8,62	8,97	8,35	8,65	6,92	1,70	8,62	8,62	8,36	6,50	7,43	5,94	5,94		8,75			
15		7,11	8,01	6,28	7,13	5,71	1,31	7,02	7,02	6,61	6,89			6,75	5,40	5,40	6,04	7,78	7,18	7,00	5,60	1,00	6,60	6,60	7,46	5,98	6,72	5,38	5,38		7,75			
16		6,82	8,02	6,03	6,96	5,57	0,89	6,46	6,46	6,25	7,41			6,83	5,46	5,46	6,15	7,61	7,53	7,10	5,68	0,76	6,44	6,44	8,14	5,69	6,92	5,53	5,53		7,31			
17		6,70	8,44	7,06	7,40	5,92	1,22	7,14	7,14	6,59	7,04			6,82	5,45	5,45	7,03	7,79	7,96	7,59	6,07	1,06	7,13	7,13	6,32	7,01	6,67	5,33	5,33		7,38			
18		7,16	9,00	8,10	8,09	6,47	1,11	7,58	7,58	7,16	9,11			8,14	6,51	6,51	7,65	8,32	8,00	7,99	6,39	0,86	7,25	7,25	6,52	6,65	6,59	5,27	5,27		8,11			
19		9,32	7,25	7,69	8,09	6,47	1,83	8,30	8,30	7,31	8,49			7,90	6,32	6,32	7,85	8,24	8,93	8,34	6,67	2,00	8,67	8,67	9,47	6,15	7,81	6,25	6,25		8,35			
20		7,08	8,59	4,64	6,77	5,42	0,89	6,31	6,31	4,96				4,96	3,97	3,97	6,25	7,48	7,33	7,02	5,62	0,60	6,22	6,22	RETIRADO						6,58			

Nota. Colegio Aristóteles Bilingüe

Tabla 9.-

Descripción de datos e información del estudiante; segundo año de Bachillerato General Unificado nómina de alumnos

SEGUNDO AÑO E.G.B								
Nº	NÓMINA	C.I. Estudiante	Fecha de Nacimiento	Edad	Dirección	Nombre del Representante	C.I. Representante	Teléfonos
1	ALVAREZ LISINTUÑA AMBAR HEDELINE	1750981134	18/3/2011		LUCHA DE LOS POBRES CALL	ALVAREZ EDWIN	502369226	985837474
2	ANDRADE FUENTES MELANY SCARLETH	1751684125	6/9/2011		TURUBAMABA BAJO CALLE	FUENTES MIRYAN	2100484126	995986271
3	BEDON ESTRELLA ARIEL SEBASTIAN	1751317197	15/6/2011		QUITUMBE CALLE RUMICHAC	ESTRELLA GLORIA	1725032609	979259371
4	BEDOYA CHIRIBOGA KATHERINNE SOFIA							
5	CALERO SALAS RODRIGO ALEXANDER	1750928762	4/2/2011		BARRIO VALLE DEL SUR CALL	SALAS JENNY	1716330871	987837003
6	CAÑAR CAMPOVERDE RENATA ELIZABETH	1751697309	3/6/2011		LUCHA DE LOS POBRES CALL	CAMPOVERDE VERONICA	173588190	987342265
7	CHECA LANDETA EMILY ABIGAIL							
8	CHIU BALANTE ARIANA VALENTINA	1752874659	20/10/2011		CALLE BORBON OE2-104 OE2	BALANTE JULI	915200125	992417983
9	CHORLANGO LEON ALEJANDRA ABIGAIL	#####	21/10/2011		LUCHA DE LOS POBRES	LEÓN VANESSA		0959551705/09612266
10	EGAS SANGUANO DANNA CAMILA	1752328250	26/10/2011		TURUBAMBA TENIENTE HUHO	SANGUANO LOURDES	1710508993	995744257
11	ESPIN RONQUILLO ALEJANDRO DAVID							
12	GUERRA POZO CRISTOFER MATEO	1751082726	1/2/2011		LA VICTORIA BAJA EL PEDRE	MARGARITA POZO	401544978	984007843
13	HARO TOAPANTA SANTIAGO JOSUE	1751949551	23/11/2011		AV.21 DE AGOSTO S26-05 Y	TOAPANTA CRISTINA	1726686817	987198517
14	HERAS BONIFAZ JAHNINA ELIZABETH	1750868224	8/2/2011		LUCHA DE LOS POBRES MAN	BONIFAZ ALEXANDRA		0995625665/2734859
15	LANDETA CHALACAN TIFANNY ANAHI							
16	LEMA CÚÑEZ JOSUE FRANCISCO	1750940270	1/2/2011		LA CALCETA S19-33 Y CHAC	LEMA PEDRO	201440328	991460996
17	NAVARRETE CARRION GABRIEL MARTIN	1752094837	18/12/2011		LUCHA DE LOS POBRES CALL	LAPO JOHANNA	1708339633	995266307
18	OVIDO VELASQUEZ MEGAN SCARLETH	1750792242	10/1/2011		CALLE 21 DE AGOSTO	OVIDO BAYRON	1103953046	999859267
19	PAGUAY TOAPANTA DAMARIS MIKAELA	1751932920	18/8/2011		NUEVA AURORA GENERAL A	TOAPANTA LISBET	1725246969	23654302
20	PITISACA SALTOS STEVEN GRABIEL	1751858521	27/9/2011		ARGELIA,PORTAL DE GUAJA	SALTOS SONNIA	1719507152	997247829

Nota. Colegio Aristóteles Bilingüe

Tabla 10.-

Descripción de datos e información del estudiante con relación a la información de la madre

Nombre de la Madre	C.I.Madre	Ocupación	Lugar de trabajo	Telefono	Estado Civil	Nombre del Padre	C.I.Padre	Ocupación	Lugar de
LISINTUÑA CRUZ MAYRA JEANETH	0503054587	Empleada privado	Av Amazonas y Fran	0988721084	Casada	ALVAREZ BARRIONUEVO EDWIN	0502369226	AGENTE DE SEGURIDAD	
FUENTES SANCHEZ MIRYAN PIEDA	2100484126	AMA DE CASA	NO	022190411	Soltera	ANDRADE ZAMBRANO MAURIC	1713743019	Policia Nacional	
SALAS GONZALEZ JENNY ALEXANE	1716330871	QUEHACER DOMESTIC	QUITUMBE, CDLA.	023072432	Casada	CALERO CAIZA JAIME RODRIG	1712429164	EMPLEADO PRIVADO	QUITUMI
CAMPOVERDE JIMÉNEZ VERÓNIC	1713588190	Ama de casa	ninguna	0	Unión Libre	CAÑAR ESPINOZA FABIAN JOSÉ	1714874086	Policia Nacional	
BALANTE HIDALGO YULI ROXANA	0915200125	EMPLEADO PÚBLICO	TERMINAL TERRESTRE	2907005	Casada	CHIU ESTRADA SANTIAGO MARTÍN		NINGUNA	
SANGUANO MONTAGUANO TERE	1710508993	Cajera	TURUBAMBA - ARCA	022690543	Divorciada	EGAS ESPINDOLA GUILLERMO EDUARDO		Chofer de taxi	
POZO HERNANDEZ MARGARITA C	0401544978	empaquetamiento	shirs y portugal	2243507	Casada	GUERRA POZO MARCO GUILLER	0401370598	chofer	santa bc
TOAPANTA CONTENTO CRISTINA	1726686817	EMPLEADO PRIVADO	AV. 10 DE AGOSTO	0987198517	Soltera	HARO MORALES JOSE LUIS	1717743254	EMPLEADO PRIVADO	
MARIA FAVIOLA CUÑEZ CUJI	0201409414	QUE HACERES DOMÉSTI	Calçeta s19-33 y Ch	0991460996	Casada	PEDRO RODRIGO LEMA PUCHA	0201440328	CONSERJE	GONZAL

Nota. Colegio Aristóteles Bilingüe

**Tabla 11.-**

*Descripción de datos e información del estudiante con relación a la información del padre*

<b>C.I.Padre</b>	<b>Ocupación</b>	<b>Lugar del trabajo</b>	<b>Teléfono</b>	<b>Estado Civil</b>
0502369226	AGENTE DE SEGURIDAD			Divorciado
1713743019	Policia Nacional		2685233	Unión Libre
1712429164	EMPLEADO PRIVADO	QUITUMBE, CDLA. VALLES D	023072432	Casado
1714874086	Policia Nacional		2685233	Unión Libre
TÍN	NINGUNA			Casado
DUARDO	Chofer de taxi			Soltero
0401370598	chofer	santa barbara		Casado
1717743254	EMPLEADO PRIVADO			Soltero
0201440328	CONSERJE	GONZALEZ SUAREZ Y CORUÑA		Casado
1722503826	MONITOR DE CALL CENTER	ÑAQUITO, AV. 6 DE DICIEM	2923251	Soltero
1720170503	CHOFER			Soltero
1719912845	GESTOR ARTISTICO	CUILA ARCELA, PORTAL DE CUI	0990844787	Casado

*Nota.* Colegio Aristóteles Bilingüe

### ***Preparación de los datos***

Con los datos obtenidos en la Fase 1, 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 dato 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 tpeo, subrayado y desde luego información no característica acorde a la temática de análisis. La limpieza de datos se expone con base en las siguientes tablas descriptivas, denotando las siguientes características:

Tabla 12.-

Limpieza de datos

ORD	APELLIDOS Y NOMBRES	Lengua y Literatura Nota 1	Lengua y Literatura Nota 2	Matemática y Literatura	SUPLETORIO Lengua y Literatura	Matemática Nota 1	Matemática Nota 2	Matemática Nota 3	SUPLETORIO Matemática	Estudios Sociales Nota 1	Estudios Sociales Nota 2	Estudios Sociales Nota 3	SUPLETORIO Estudios Sociales	Ciencias Naturales Nota 1	Ciencias Naturales Nota 2	Ciencias Naturales Nota 3
1	Nómina de Estudiantes	5,33	7,01	6,17	SUPLETORIO	6,63	7,62	7,13		6,70	7,30	7,00		5,06	8,93	
2		8,20	8,87	8,54		8,84	9,10	8,97		9,46	9,59	9,53		8,48	8,52	
3		6,33	6,75	6,54	SUPLETORIO	5,66	5,48	5,57	SUPLETORIO	7,18	6,83	7,01		5,80	8,20	
4		5,40	7,00	6,20	SUPLETORIO	6,96	5,16	6,06	SUPLETORIO	6,87	7,13	7,00		5,10	8,90	
5		7,30	8,82	8,06		8,00	8,52	8,26		8,77	9,46	9,12		8,64	7,57	
6		5,28	8,71	7,00		5,23	8,76	7,00		8,41	8,66	8,54		5,90	8,10	
7		6,19	7,88	7,04		6,81	7,26	7,04		8,44	7,71	8,08		8,23	7,91	
8		5,46	8,53	7,00		5,98	8,01	7,00		8,16	8,60	8,38		7,17	7,14	
10		7,61	8,42	8,02		7,40	8,22	7,81		8,63	9,27	8,95		7,84	7,76	
11		7,14	7,58	7,36		9,12	8,09	8,61		9,26	8,68	8,97		8,06	7,74	
12		7,20	7,58	7,39		7,33	6,67	7,00		8,53	7,45	7,99		7,33	6,77	
13		6,52	7,47	7,00		8,83	7,80	8,32		7,89	7,52	7,71		6,72	8,23	
14		5,38	7,60	6,49	SUPLETORIO	5,90	4,61	5,26	SUPLETORIO	6,72	7,28	7,00		4,42	9,58	
15		5,33	7,37	6,35	SUPLETORIO	4,96	6,01	5,49	SUPLETORIO	7,05	7,25	7,15		4,90	9,10	
16		7,33	7,77	7,55		6,27	7,89	7,08		8,31	8,53	8,42		8,17	7,59	

Nota. Colegio Aristóteles Bilingüe

Tabla 13.-

Limpieza de datos; notas

Ciencia Naturales Nota 3	FORIO Ciencias N	Educación Cult y Art Nota 1	Educación Cult y Art Nota 2	Educación Cult y Art Nota 3	In Cult y Art SUPLE	Educación Física Nota 1	Educación Física Nota 2	Educación Física Nota 3	Educación Física SUPLETORIO	Inglés Nota 1	Inglés Nota 2	Inglés Nota 3	Inglés SUPLETORIO
7,00		5,49	8,51	7,00		6,49	7,50	7,00		6,57	7,51	7,04	
8,50		8,26	8,57	8,42		7,86	7,96	8,01		8,67	9,02	8,85	
7,00		6,83	7,29	7,06		5,94	8,06	7,00		7,32	8,03	7,68	
7,00		6,13	7,86	7,00		6,81	7,19	7,00		5,89	6,20	6,05	SUPLETORIO
8,11		7,49	8,93	8,21		7,71	8,09	7,90		8,14	9,10	8,62	
7,00		6,70	8,01	7,36		8,61	7,98	8,30		6,77	7,87	7,87	
8,07		6,35	7,70	7,03		7,41	7,25	7,33		7,72	7,74	7,73	
7,16		6,13	7,98	7,06		7,50	7,63	7,57		6,54	8,38	7,46	
7,80		7,53	8,52	8,03		7,70	8,35	8,03		8,06	8,78	8,42	
7,90		7,86	6,31	7,09		8,28	9,45	8,87		8,37	8,68	8,53	
7,05		7,68	6,79	7,24		8,16	8,47	8,32		7,99	7,67	7,83	
7,48		7,19	7,30	7,25		7,08	7,52	7,30		8,65	8,80	8,73	
7,00		5,77	8,22	7,00		6,79	7,21	7,00		5,90	7,79	6,85	SUPLETORIO
7,00		3,99	10,00	7,00		6,55	7,45	7,00		5,92	7,23	6,58	SUPLETORIO
7,88		7,29	7,24	7,27		7,82	7,66	7,74		7,33	8,29	7,81	

Nota. Colegio Aristóteles Bilingüe

Tabla 14.-

Limpieza de datos; notas

ORD	APELLIDOS Y NOMBRES	matemática nota 1	matemática nota 2	temática nota	SUPLETORIO MATEM	Física nota 1	Física nota 2	Física nota 3	SUPLETORIO FISI	Química Nota1	Química Nota 2	Química Nota	SUPLETORIO QUIM	Biología Nota1	Biología Nota2	Biología Nota
1	Nómina de Estudiantes	4,54				4,69				3,98		1,99	REMEDIAL	3,58		
2		7,00	7,26	7,13		7,55	7,33	7,44		6,54	4,95	5,75	SUPLETORIO	7,31		8,07
3		5,70	6,85	6,28	SUPLETORIO	3,67	4,93	4,30	REMEDIAL	4,61	6,11	5,36	SUPLETORIO	5,53		8,51
4		5,56	6,24	5,90	SUPLETORIO	5,37	4,70	5,04	SUPLETORIO	5,39	3,48	4,44	REMEDIAL	5,94		8,06
5		6,60	7,40	7,00		6,50	7,50	7,00		7,00	3,61	5,31	SUPLETORIO	6,16		7,84
6		9,88	9,47	9,68		9,61	9,35	9,48		9,99	9,51	9,75		9,51		9,56
7		7,06	7,55	7,31		6,20	7,79	7,00		6,59	5,72	6,16	SUPLETORIO	5,62		8,38
8		4,81	6,38	5,60	SUPLETORIO	4,35	5,66	5,01	SUPLETORIO	5,69	3,91	4,80	REMEDIAL	4,91		9,08
9		7,38	7,57	7,48		7,00	7,13	7,07		7,72	6,45	7,09		7,23		7,64
10		6,08	6,14	6,11	SUPLETORIO	5,34	5,77	5,56	SUPLETORIO	7,12	4,83	5,98	SUPLETORIO	6,83		7,16
11		7,08	7,33	7,21		6,53	7,38	7,02		6,35	5,35	5,85	SUPLETORIO	6,16		7,91
12		7,38	6,80	7,09		6,73	7,27	7,00		7,90	7,65	7,78		7,15		8,35
13		3,52		3,52	REMEDIAL	2,44	0,00	2,44	REMEDIAL	3,73		1,87	REMEDIAL	4,27		
14		3,39		3,39	REMEDIAL	3,02		3,02	REMEDIAL	4,30		2,15	REMEDIAL	4,97		

Nota. Colegio Aristóteles Bilingüe

Tabla 15.-

Limpieza de datos; notas remedial

Biología Nota	SUPLETORIO Biología	Historia Nota1	Historia Nota2	Historia Nota3	SUPLETORIO HISTORIA	Ed ciudadania Nota1	Ed ciudadania Nota2	ciudadania Nota	SUPLETORIO ED. CIUDADANIA	Filosofía Nota1	Filosofía Nota2	Filosofía Nota	SUPLETORIO FILOSOFÍA	lenguaje y literatura	lenguaje y literatura	lenguaje y literatura	SUPLETORIO LENGUA Y LITERATURA
1,79	REMEDIAL	6,24				7,41				6,24				4,09	0,00	4,09	REMEDIAL
7,69		8,17	9,44	7,20		8,33	9,37	8,85		8,93	9,51	9,22		6,80	7,88	7,34	
7,02		7,84	9,08	7,59		6,90	8,77	7,84		8,21	9,05	8,63		6,58	7,41	7,00	
7,00		7,45	8,70	6,09		8,38	8,40	8,39		7,46	9,03	8,25		6,12	7,88	7,00	
7,00		7,71	9,18	6,39		8,03	8,69	7,88		8,42	8,95	8,69		7,20	7,23	7,22	
9,54		9,62	9,97	9,74		8,55	9,89	9,22		9,88	9,93	9,91		8,27	9,31	8,79	
7,00		7,15	8,47	7,10		6,60	8,53	7,57		7,08	8,72	7,90		6,53	7,50	7,02	
7,00		7,23	8,73	6,32		7,46	8,54	8,00		7,23	8,74	7,99		6,14	7,87	7,01	
7,44		7,90	9,77	8,11		7,51	9,52	8,40		8,52	9,67	9,10		7,36	8,16	7,76	
7,00		8,06	9,46	7,15		7,05	8,87	7,96		8,27	9,18	8,73		6,59	7,40	7,00	
7,04		7,92	9,35	7,35		6,98	9,48	8,23		7,23	9,46	8,35		7,01	7,81	7,41	
7,75		8,51	8,93	8,29			9,36	8,87		8,54	9,50	9,02		7,79	8,01	7,90	
2,14	REMEDIAL	5,74			SUPLETORIO	6,47		6,47	SUPLETORIO	7,06		7,06		4,79	0,00	2,40	REMEDIAL
2,49	REMEDIAL	6,86			SUPLETORIO	6,35		7,55		7,61		7,61		6,30	0,00	3,15	REMEDIAL

Nota. Colegio Aristóteles Bilingüe

Tabla 16.-

Limpieza de datos; notas inglés

Ingles Nota1	Ingles Nota2	Ingles Nota3	SUPLETORIO INGLÉS	ccion Cult Art	ccion Cult Art	ccion Cult Art	SUPLETORIO ECA	Educación Física Nota 1	Educación Física Nota 2	Educación Física Nota 3	ORIO EDUCACION	Emprendimiento y gestión nota1	Emprendimiento y gestión nota2	Emprendimiento y gestión nota3	EMPRENDIMIENTO Y GESTION
4,35				6,13	6,98			4,11		4,11	REMEDIAL	3,51			
6,41	7,94	8,73		7,23	7,76	7,50		5,95	7,44	6,70	SUPLETORIO	7,07	6,96	7,02	
6,03	8,07	8,56		6,95	7,86	7,41		5,87	7,29	6,58	SUPLETORIO	6,84	7,19	7,02	
5,85	6,63	7,83	SUPLETORIO	7,22	8,46	7,84		6,74	7,02	6,97	SUPLETORIO	5,88	6,02	5,95	SUPLETORIO
6,25	5,88	7,42	SUPLETORIO	8,10	8,31	8,20		7,08	8,95	8,02		7,00	7,00	7,00	
8,86	9,16	9,55		8,51	8,11	8,31		9,23	9,27	9,25		9,01	9,21	9,11	
7,48	7,56	8,14		6,77	7,22	7,00		7,04	7,50	7,27		5,19	6,01	5,60	SUPLETORIO
4,60	5,26	7,00	REMEDIAL	7,40	7,99	7,70		5,66	7,18	6,42	SUPLETORIO	4,44	4,60	4,52	REMEDIAL
5,87	8,12	8,90		8,26	8,41	8,12		8,41	7,97	8,19		6,71	7,28	7,00	
9,06	9,04	9,11		6,66	8,36	7,51		7,01	6,13	6,57	SUPLETORIO	4,07	6,63	5,35	
6,73	7,36	8,41		6,86	8,43	7,42		8,18	8,93	8,56		5,62	7,00	6,31	SUPLETORIO
6,38	7,65	8,58		7,27		8,10		7,38	8,34	7,86		7,08	7,73	7,41	
2,38				5,25	0,00	5,25	SUPLETORIO					3,31			
5,32				6,21		6,21	SUPLETORIO					5,01			

Nota. Colegio Aristóteles Bilingüe

Tabla 17.-

Limpieza de datos; notas inglés supletorio

ORD	APELLIDOS Y NOMBRES	Matemática Nota 1	Matemática Nota 2	Matemática Nota 3	Supletorio Matemática	Física Nota 1	Física Nota 2	Física Nota 3	Supletorio Física	Química Nota 1	Química Nota 2	Química Nota 3	Supletorio Química	Biología Nota 1	Biología Nota 2	Biología Nota 3
1	Nómina de Estudiantes	6,22	6,12	6,17	SUPLETORIO	7,01	7,77	7,39		7,80	6,97	7,39		8,14	6,60	
2		5,66	7,65	6,66	SUPLETORIO	6,71	7,75	7,23		7,38	6,61	7,00		7,15	6,84	
3		7,46	7,96	7,71		6,16	7,91	7,08		8,02	6,27	7,15		7,27	7,44	
4		6,15	7,02	6,59	SUPLETORIO	5,84	8,19	7,02		5,46	8,53	7,00		8,54	7,18	
5		5,19	5,78	5,49	SUPLETORIO	5,17	8,82	7,00		6,75	7,37	7,06		7,60	7,37	
6		7,09	7,27	7,18		7,03	7,10	7,07		8,60	8,14	8,37		9,10	8,52	
7		7,58	7,34	7,46		8,13	7,60	7,87		7,11	6,91	7,01		7,76	7,36	
8		7,83	7,84	7,84		7,59	7,90	7,75		7,62	7,39	7,51		7,22	7,11	
9		7,06	7,38	7,22		7,74	7,06	7,40		5,76	8,25	7,01		7,16	7,41	
10		8,10	8,10	8,10		8,35	8,20	8,28		7,73	7,59	7,66		7,35	7,98	
11		7,37	6,16	6,77	SUPLETORIO	6,17	7,92	7,05		8,41	6,16	7,29		7,00	7,40	

Nota. Colegio Aristóteles Bilingüe

Tabla 18.-

Limpieza de datos; notas supletorio

Biología Nota 3	Supletorio Biología	Historia y ciencias sociales nota 1	Historia y ciencias sociales nota 2	Historia y ciencias sociales nota 3	Historia y Ciencias	Lengua y Literatura Nota 1	Lengua y Literatura Nota 2	Lengua y Literatura Nota 3	Supletorio Lengua y Literatura	Inglés Nota 1	Inglés Nota 2	Inglés Nota 3	Supletorio Inglés	Educación Física Nota 1	Educación Física Nota 2	Educación Física Nota 3	Supletorio Educación Física
7,37		7,41	9,00	8,21		7,61	7,29	7,45		7,63	7,04	7,34		7,32	8,85	8,09	
7,00		6,43	8,92	7,68		7,25	7,36	7,31		7,16	7,94	7,55		7,06	8,99	8,03	
7,36		7,43	8,06	7,75		7,14	7,81	7,48		7,22	6,78	7,00		7,62	8,00	7,81	
7,86		6,81	8,28	7,55		6,97	7,66	7,32		6,21	6,85	6,53	SUPLETORIO	7,05	8,20	7,63	
7,49		8,12	7,92	8,02		7,70	7,51	7,61		6,71	7,50	7,11		7,27	8,93	8,10	
8,81		7,47	8,72	8,10		7,33	7,42	7,38		8,68	8,02	8,35		7,27	8,24	7,76	
7,56		7,92	8,42	8,17		7,86	7,23	7,55		7,82	7,95	7,89		7,56	9,47	8,52	
7,17		8,14	9,00	8,57		8,45	7,75	8,10		8,09	8,14	8,12		7,90	8,59	8,25	
7,29		7,87	7,93	7,90		7,34	7,23	7,29		6,76	8,13	7,45		7,04	8,73	7,89	
7,67		8,19	9,28	8,74		7,82	7,66	7,74		7,61	8,11	7,86		7,79	9,01	8,40	
7,20		5,23	8,77	7,00		6,11	7,94	7,03		6,17	6,65	6,41	SUPLETORIO	7,04	8,34	7,69	

Nota. Colegio Aristóteles Bilingüe



Tabla 20.-

Limpieza de datos; notas global

ORD.	APELLIDOS Y NOMBRES	LENGUA Y LITERATURA A NOTA 1	LENGUA Y LITERATURA A NOTA 2	LENGUA Y LITERATURA A NOTA 3	MATEMATICA NOTA 1	MATEMATICA NOTA 2	MATEMATICA NOTA 3	ESTUDIOS SOCIALES NOTA 1	ESTUDIOS SOCIALES NOTA 2	ESTUDIOS SOCIALES NOTA 3	CIENCIAS NATURALES NOTA 1	CIENCIAS NATURALES NOTA 2	CIENCIAS NATURALES NOTA 3	EDUCACION CULTURAL Y ARTISTICA NOTA 1	EDUCACION CULTURAL Y ARTISTICA NOTA 2
1	Nómina de Estudiantes	9,15	9,29	9,22	9,40	8,77	9,09	9,58	9,35	9,47	9,28	9,19	9,24	9,16	10,00
2		9,71	9,91	9,81	9,81	9,86	9,84	9,93	9,90	9,92	9,94	9,97	9,96	9,79	10,00
3		9,10	9,30	9,20	9,15	8,75	8,95	9,65	9,49	9,57	9,55	9,45	9,50	9,59	10,00
4		7,92	7,94	7,93	6,95	7,21	7,08	7,97	8,24	8,11	7,65	8,01	7,83	9,37	10,00
5		9,87	9,98	9,93	9,88	9,93	9,91	9,82	9,97	9,90	9,94	9,97	9,96	9,95	10,00
6		8,37	8,72	8,55	8,41	7,66	8,04	9,08	8,72	8,90	7,99	8,36	8,18	9,37	10,00
7		7,72	7,31	7,52	7,64	7,30	7,47	8,98	7,76	8,37	9,11	8,13	8,62	9,59	10,00
8		8,04	7,61	7,83	7,69	6,99	7,34	8,77	7,46	8,12	8,05	8,12	8,09	9,16	10,00
9		7,53	6,91	7,22	6,90	7,10	7,00	8,10	7,76	7,93	7,36	7,56	7,46	9,32	10,00
10		8,12	8,29	8,21	7,15	7,23	7,19	9,00	8,72	8,86	8,87	8,55	8,71	9,37	10,00
11		9,55	9,90	9,73	9,81	9,31	9,56	9,95	9,98	9,97	9,92	9,95	9,94	9,92	10,00
12		7,38	7,98	7,68	6,88	7,19	7,04	8,66	7,93	8,30	8,18	8,92	8,55	9,27	10,00
13		8,33	8,24	8,29	7,46	7,21	7,34	9,20	8,35	8,78	8,60	8,74	8,67	9,37	10,00
14		8,27	8,71	8,49	8,59	7,94	8,27	8,90	8,81	8,86	8,62	8,52	8,57	9,32	10,00
15		7,45	7,80	7,63	6,79	7,34	7,07	8,53	7,48	8,01	8,52	7,88	8,20	9,37	10,00
16		8,91	9,23	9,07	8,94	8,14	8,54	9,53	8,82	9,18	9,29	9,35	9,32	9,73	10,00
17		8,82	9,04	8,93	8,75	8,25	8,50	9,41	8,85	9,13	9,56	9,21	9,39	9,79	10,00
18		9,69	9,91	9,80	9,73	9,56	9,65	9,92	9,65	9,79	9,90	9,95	9,93	9,89	10,00
19		8,08	8,66	8,37	8,15	7,51	7,83	9,24	8,38	8,81	8,72	8,50	8,61	9,79	10,00
20		8,88	8,58	8,73	8,49	7,65	8,07	9,69	8,73	9,21	9,32	9,21	9,27	9,48	10,00
21		8,81	8,51	8,66	8,24	7,57	7,91	9,64	8,88	9,26	8,91	8,84	8,88	9,53	10,00

Nota. Colegio Aristóteles Bilingüe

Tabla 21.-

Limpieza de datos; notas global

MATEMATICA NOTA 3	ESTUDIOS SOCIALES NOTA 1	ESTUDIOS SOCIALES NOTA 2	ESTUDIOS SOCIALES NOTA 3	CIENCIAS NATURALES NOTA 1	CIENCIAS NATURALES NOTA 2	CIENCIAS NATURALES NOTA 3	EDUCACION CULTURAL Y ARTISTICA NOTA 1	EDUCACION CULTURAL Y ARTISTICA NOTA 2	EDUCACION CULTURAL Y ARTISTICA NOTA 3	EDUCACION FISICA NOTA 1	EDUCACION FISICA NOTA 2	EDUCACION FISICA NOTA 3	INGLÉS NOTA 1	INGLÉS NOTA 2	INGLÉS NOTA 3
9,09	9,58	9,35	9,47	9,28	9,19	9,24	9,16	10,00	9,58	9,09	8,79	8,94	8,79	9,44	9,12
9,84	9,93	9,90	9,92	9,94	9,97	9,96	9,79	10,00	9,90	9,20	9,63	9,42	9,89	9,78	9,84
8,95	9,65	9,49	9,57	9,55	9,45	9,50	9,59	10,00	9,80	8,64	9,00	8,82	9,52	9,45	9,49
7,08	7,97	8,24	8,11	7,65	8,01	7,83	9,37	10,00	9,69	9,27	9,32	9,30	7,67	8,07	7,87
9,91	9,82	9,97	9,90	9,94	9,97	9,96	9,95	10,00	9,98	9,52	9,44	9,48	9,70	9,79	9,75
8,04	9,08	8,72	8,90	7,99	8,36	8,18	9,37	10,00	9,69	9,11	9,07	9,09	8,09	8,01	8,05
7,47	8,98	7,76	8,37	9,11	8,13	8,62	9,59	10,00	9,80	9,20	9,16	9,18	7,65	7,52	7,59
7,34	8,77	7,46	8,12	8,05	8,12	8,09	9,16	10,00	9,58	8,89	8,99	8,94	7,77	7,16	7,47
7,00	8,10	7,76	7,93	7,36	7,56	7,46	9,32	10,00	9,66	8,73	8,89	8,81	6,28	7,81	7,05
7,19	9,00	8,72	8,86	8,87	8,55	8,71	9,37	10,00	9,69	9,37	9,15	9,26	7,66	8,16	7,91
9,56	9,95	9,98	9,97	9,92	9,95	9,94	9,92	10,00	9,96	9,00	9,11	9,06	9,77	9,78	9,78
7,04	8,66	7,93	8,30	8,18	8,92	8,55	9,27	10,00	9,64	8,89	9,52	9,21	5,86	8,14	7,00
7,34	9,20	8,35	8,78	8,60	8,74	8,67	9,37	10,00	9,69	9,79	9,57	9,68	7,03	7,61	7,32
8,27	8,90	8,81	8,86	8,62	8,52	8,57	9,32	10,00	9,66	9,32	9,25	9,29	8,25	8,67	8,46
7,07	8,53	7,48	8,01	8,52	7,88	8,20	9,37	10,00	9,69	8,27	8,73	8,50	6,92	7,17	7,05
8,54	9,53	8,82	9,18	9,29	9,35	9,32	9,73	10,00	9,87	9,48	10,00	9,74	9,43	9,36	9,40
8,50	9,41	8,85	9,13	9,56	9,21	9,39	9,79	10,00	9,90	8,64	8,95	8,80	8,16	8,46	8,31
9,65	9,92	9,65	9,79	9,90	9,95	9,93	9,89	10,00	9,95	9,47	9,48	9,48	9,69	9,83	9,76
7,83	9,24	8,38	8,81	8,72	8,50	8,61	9,79	10,00	9,90	9,27	9,32	9,30	8,60	8,45	8,53
8,07	9,69	8,73	9,21	9,32	9,21	9,27	9,48	10,00	9,74	8,89	9,00	8,95	9,05	8,52	8,79
7,91	9,64	8,88	9,26	8,91	8,84	8,88	9,53	10,00	9,77	9,57	9,25	9,41	8,01	8,28	8,15

Nota. Colegio Aristóteles Bilingüe

Tabla 22.-

Limpieza de datos; notas global (supletorio)

ORD	APELLIDOS Y NOMBRES	matematica nota 1	matematica nota 2	tematica nota 3	SUPLETORIO MATEMA	Fisica nota 1	Fisica nota 2	Fisica nota 3	SUPLETORIO FISIC	Química Nota1	Química Nota 2	Química Nota	SUPLETORIO QUIM	Biología Nota1	Biología Nota2	B
1	Nómina de Estudiantes	7,06	7,14	7,10		7,05	7,49	7,27		7,53	9,10	8,32		7,48	8,49	
2		5,27	3,42	4,35	REMEDIAL	5,45	3,69	4,57	REMEDIAL	5,73	3,69	4,71	REMEDIAL	3,58	4,85	
3		8,46	7,17	7,82		8,64	7,47	8,06		8,21	8,79	8,50		8,94	8,91	
4		6,42	4,57	5,50	SUPLETORIO	6,31	5,55	5,93	SUPLETORIO	6,88	6,17	6,53	SUPLETORIO	6,61	4,97	
5		8,22	6,14	7,18		7,47	7,01	7,24		6,24	6,71	6,48	SUPLETORIO	6,52	8,30	
6					RETIRADA					RETIRADA	4,34			RETIRADA	5,79	
7		8,26	7,05	7,66		8,38	7,58	7,98		6,83	6,63	6,73	SUPLETORIO	8,05	7,56	
8		8,56	7,26	7,91		8,35	7,56	7,96		8,15	7,41	7,78		8,79	7,87	
9		7,56	5,18	6,37	SUPLETORIO	7,48	5,42	6,45	SUPLETORIO	7,03	7,19	7,11		7,35	7,68	
10		7,13		7,13	RETIRADA	6,88	0,00			RETIRADA	6,90		6,90	RETIRADA	7,00	
11		5,83	5,51	5,83	SUPLETORIO	6,64	4,85	6,64	SUPLETORIO	5,75	6,54	6,15	SUPLETORIO	7,02	6,27	
12		7,32	6,06	6,84	SUPLETORIO	7,52	5,59	6,56	SUPLETORIO	7,41	6,61	7,01		7,46	6,76	
13		7,94	6,30	7,12		7,49	7,30	7,40		7,02	7,55	7,29		7,75	7,11	
14		6,97	5,98	6,48	SUPLETORIO	6,38	4,31	6,38	SUPLETORIO	4,52	6,08	5,30	SUPLETORIO	4,84	5,17	
15		9,16	7,92	8,54		8,26	8,25	8,26		7,93	8,56	8,25		7,23	8,98	
16		5,76	3,61	4,69	REMEDIAL	5,97	6,06	6,02	SUPLETORIO	6,71	4,58	5,65	SUPLETORIO	6,49	4,66	
17		7,81	7,55	7,68		8,46	7,98	8,22		9,28	8,39	8,84		9,10	9,09	
18		7,33	2,47	4,88	REMEDIAL	6,11	3,51	4,81	REMEDIAL	4,19	3,55	3,87	REMEDIAL	5,74	4,05	
19		9,48	8,85	9,17		9,11	8,35	8,73		9,28	9,06	9,17		9,14	9,19	
20		7,61	7,67	7,64		6,28	7,48	6,88	SUPLETORIO	6,05	8,01	7,03		7,04	7,75	
21		8,14	7,25	7,70		7,28	7,64	7,46		8,35	8,71	8,53		8,11	7,25	
22		5,75	5,83	5,79	SUPLETORIO	7,01	6,53	6,77	SUPLETORIO	6,24	7,77	7,01		7,09	6,09	
23		7,14	7,79	7,47		7,56	7,56	7,56		6,86	7,58	7,22		6,37	7,65	

Nota. Colegio Aristóteles Bilingüe





Tabla 25.-

Limpieza de datos; notas global (supletorio)

ORD	APELLIDOS Y NOMBRES	matemática nota 1	matemática nota 2	temática nota 3	LETORIO MATEM	Física Química nota 1	Física Química nota 2	Física Química nota 3	LETORIO FISICA QU	Economi a Nota 1	Economi a Nota 2	Economi a Nota 3	PLETORIO Econo	Biología Nota1	Biología Nota2	
1	Nómina de Estudiantes	5,89	4,99	5,44	SUPLETORIO	6,46	6,50	6,48	SUPLETORIO	7,47	7,61	7,54		7,13	7,80	
2		6,72	5,94	6,33	SUPLETORIO	7,07	7,44	7,26		7,53	6,87	7,20		8,40	6,69	
3		8,40	6,77	7,59			8,33	7,63	7,98		9,13	8,34	8,74		9,16	7,48
4		9,28	8,73	9,01			9,20	8,92	9,06		9,57	9,68	9,63		9,37	9,40
5		9,56	8,79	9,18			9,05	8,80	8,93		9,45	9,83	9,64		9,24	8,75
6		5,57	5,83	5,70	SUPLETORIO	6,20	5,73	5,97	SUPLETORIO	6,34	8,28	7,31		6,41	7,69	
7		5,90	3,69	4,80	REMEDIAL		5,98	3,75	4,87	REMEDIAL	6,82	7,28	7,05		7,87	7,56
8		7,01	6,98	7,00			7,39	7,27	7,33		7,64	8,85	8,24		7,15	7,30
10		8,60	8,36	8,48			8,14	7,96	8,05		8,60	9,48	9,04		8,25	9,43
11		8,91	8,04	8,48			8,59	8,19	8,39		8,93	9,40	9,17		8,84	9,04
12		7,00	6,20	6,60	SUPLETORIO		7,13	7,00	7,07		7,53	8,66	8,10		8,47	8,32
13		8,67	7,64	8,16			8,53	8,89	8,71		8,02	9,52	8,77		9,16	9,02
14		7,56	7,40	7,48			6,18	6,34	6,26	SUPLETORIO	8,67	8,36	8,52		7,77	7,39
15		8,60	8,36	8,48			8,14	7,96	8,05		8,60	9,48	9,04		8,25	9,43
16		8,91	8,04	8,48			8,59	8,19	8,39		8,93	9,40	9,17		8,84	9,04
17		7,00	6,20	6,60	SUPLETORIO		7,13	7,00	7,07		7,53	8,66	8,10		8,47	8,32
18		8,67	7,64	8,16			8,53	8,89	8,71		8,02	9,52	8,77		9,16	9,02
19		7,56	7,40	7,48			6,18	6,34	6,26	SUPLETORIO	8,67	8,36	8,52		7,77	7,39
20		8,67	7,64	8,16			8,53	8,89	8,71		8,02	9,52	8,77		9,16	9,02

Nota. Colegio Aristóteles Bilingüe

Tabla 26.-

Limpieza de datos; notas global (supletorio)

Biología Nota	SUPLETORIO Biología	Historia y Ciencias Sociales Nota1	Historia y Ciencias Sociales Nota2	Historia y Ciencias Sociales Nota3	HISTORIA y CIENCIAS	Ed ciudadania Nota1	Ed ciudadania Nota2	Ciudadanía Nota	SUPLETORIO ED. CIUDAD	Filosofía Nota1	Filosofía Nota2	Filosofía Nota	SUPLETORIO FILOSOFÍA	Lingüística y literatura	Lingüística y literatura	Lingüística y literatura	SUPLETORIO LENGUA Y LINGÜÍSTICA
7,47		8,03	8,27	8,15		8,16	9,10	8,63		7,20	9,10	8,15		8,53	8,14	8,34	
7,55		8,88	8,37	8,63		8,05	8,53	8,29		8,95	8,74	8,85		6,25	8,19	7,22	
8,32		8,32	7,74	8,03		8,34	8,00	8,17		8,93	8,95	8,94		8,83	8,58	8,71	
9,39		9,05	8,69	8,87		9,30	9,49	9,40		9,51	9,25	9,38		8,86	8,55	8,71	
9,00		9,19	9,14	9,17		9,56	9,61	9,59		9,06	9,28	9,17		8,60	8,87	8,74	
7,05		7,41	7,78	7,60		6,79	7,41	7,10		7,03	7,28	7,16		7,96	6,42	7,19	
7,72		5,50	6,34	5,92	SUPLETORIO	5,37	7,38	6,38	SUPLETORIO	6,62	7,84	7,23		7,98	7,46	7,72	
7,23		8,79	8,15	8,47		7,13	8,39	7,76		8,45	8,53	8,49		8,62	8,13	8,38	
8,84		9,01	9,06	9,04		9,01	9,33	9,17		9,26	9,02	9,14		7,89	8,69	8,29	
8,94		9,16	8,95	9,06		9,39	9,30	9,35		8,98	9,49	9,24		8,39	9,10	8,75	
8,40		8,72	8,10	8,41		8,95	9,15	9,05		8,26	9,10	8,68		8,38	8,21	8,30	
9,09		8,94	9,08	9,01		9,37	9,32	9,35		9,49	9,26	9,38		8,61	8,37	8,49	
7,58		8,07	7,79	7,93		7,53	7,76	7,65		7,15	7,05	7,10		8,31	7,93	8,12	
8,84		9,01	9,06	9,04		9,01	9,33	9,17		9,26	9,02	9,14		7,89	8,69	8,29	
8,94		9,16	8,95	9,06		9,39	9,30	9,35		8,98	9,49	9,24		8,39	9,10	8,75	
8,40		8,72	8,10	8,41		8,95	9,15	9,05		8,26	9,10	8,68		8,38	8,21	8,30	
9,09		8,94	9,08	9,01		9,37	9,32	9,35		9,49	9,26	9,38		8,61	8,37	8,49	
7,58		8,07	7,79	7,93		7,53	7,76	7,65		7,15	7,05	7,10		8,31	7,93	8,12	
9,09		8,94	9,08	9,01		9,37	9,32	9,35		9,49	9,26	9,38		8,61	8,37	8,49	

Nota. Colegio Aristóteles Bilingüe

Tabla 27.-

Limpieza de datos; notas global (supletorio)

Extranjera	Extranjera	Extranjera	RIO LENGUA EX	cion Cult Art	cion Cult Art	cion Cult Art	SUPLETORIO ECA	Educación Física Nota 1	Educación Física Nota 2	Educación Física Nota 3	ORIO EDUCACION	MPUTACION NOT	MPUTACION NOT	MPUTACION NOT	TORIO COMPUT	Empleamiento y gestión nota1	Empleamiento y gestión nota2
7,10	8,11	7,61		8,68	9,07	8,88		8,61	7,31	7,96		4,55	6,12	5,34	SUPLETORIO	4,69	5,6
7,14	6,85	7,00		8,05	7,74	7,90		8,55	7,59	8,07		6,92	6,50	6,71	SUPLETORIO	5,51	6,1
8,06	7,51	7,79		9,12	8,01	8,57		9,23	8,72	8,98		8,35	5,84	7,10		8,30	7,4
9,02	8,27	8,65		8,75	8,91	8,83		9,13	9,21	9,17		8,42	8,29	8,36		9,16	8,2
9,43	7,95	8,69		8,69	9,15	8,92		9,16	8,72	8,94		7,80	8,70	8,25		9,14	8,8
4,12	4,41	4,27	REMEDIAL	8,03	8,95	8,49		8,77	8,88	8,83		5,00	3,40	4,20	REMEDIAL	4,47	5,5
5,14	4,80	4,97	REMEDIAL	7,79	7,92	7,86		7,91	8,20	8,06		7,00	3,95	5,48	SUPLETORIO	5,68	5,1
6,75	5,98	6,75	SUPLETORIO	8,21	8,72	8,47		7,80	7,00	7,40		5,70	6,25	5,98	SUPLETORIO	5,73	6,5
9,06	8,54	8,80		8,95	8,90	8,93		9,52	9,48	9,50		7,78	7,42	7,60		7,69	7,8
9,23	9,08	9,16		9,01	8,96	8,99		7,84	7,11	7,48		8,21	8,48	8,35		8,15	8,4
6,86	7,90	7,38		8,69	8,42	8,56		7,93	8,17	8,05		5,70	7,10	6,40	SUPLETORIO	7,00	7,7
8,21	8,23	8,22		8,96	8,65	8,81		7,07	9,09	8,08		8,00	8,42	8,21		7,84	8,2
6,41	6,63	6,52	SUPLETORIO	8,16	8,00	8,08		6,29	8,83	7,56		7,03	5,34	6,19	SUPLETORIO	5,04	6,4
9,06	8,54	8,80		8,95	8,90	8,93		9,52	9,48	9,50		7,78	7,42	7,60		7,69	7,8
9,23	9,08	9,16		9,01	8,96	8,99		7,84	7,11	7,48		8,21	8,48	8,35		8,15	8,4
6,86	7,90	7,38		8,69	8,42	8,56		7,93	8,17	8,05		5,70	7,10	6,40	SUPLETORIO	7,00	7,7
8,21	8,23	8,22		8,96	8,65	8,81		7,07	9,09	8,08		8,00	8,42	8,21		7,84	8,2
6,41	6,63	6,52	SUPLETORIO	8,16	8,00	8,08		6,29	8,83	7,56		7,03	5,34	6,19	SUPLETORIO	5,04	6,4
8,21	8,23	8,22		8,96	8,65	8,81		7,07	9,09	8,08		8,00	8,42	8,21		7,84	8,2

Nota. Colegio Aristóteles Bilingüe

Tabla 28.-

Limpieza de datos; notas global (supletorio)

SUPLETORIO ECA	Educación Física Nota 1	Educación Física Nota 2	Educación Física Nota 3	ORIO EDUCACION	MPUTACION NOT	MPUTACION NOT	MPUTACION NOT	ETORIO COMPUT	Emprendimiento y gestión nota1	Emprendimiento y gestión nota2	Emprendimiento y gestión nota3	EMPRESARIADO	PSICOLOGIA NOTAS	PSICOLOGIA NOTAS	PSICOLOGIA NOTAS	ETORIO PSICOL
	8,61	7,31	7,96		4,55	6,12	5,34	SUPLETORIO	4,69	5,64	5,17	SUPLETORIO	8,56	8,56	8,83	
	8,55	7,59	8,07		6,92	6,50	6,71	SUPLETORIO	5,51	6,15	5,83	SUPLETORIO	8,61	8,61	8,92	
	9,23	8,72	8,98		8,35	5,84	7,10		8,30	7,46	7,88		8,60	8,60	8,46	
	9,13	9,21	9,17		8,42	8,29	8,36		9,16	8,28	8,72		9,42	9,42	9,29	
	9,16	8,72	8,94		7,80	8,70	8,25		9,14	8,65	8,90		8,91	8,91	8,57	
	8,77	8,88	8,83		5,00	3,40	4,20	REMEDIAL	4,47	5,54	5,01	SUPLETORIO	8,35	8,35	8,46	
	7,91	8,20	8,06		7,00	3,95	5,48	SUPLETORIO	5,68	5,15	5,68	SUPLETORIO	7,72	7,72	7,38	
	7,80	7,00	7,40		5,70	6,25	5,98	SUPLETORIO	5,73	6,57	6,15	SUPLETORIO	7,99	7,99	8,54	
	9,52	9,48	9,50		7,78	7,42	7,60		7,69	7,82	7,76		8,91	8,91	9,05	
	7,84	7,11	7,48		8,21	8,48	8,35		8,15	8,46	8,31		9,00	9,00	9,19	
	7,93	8,17	8,05		5,70	7,10	6,40	SUPLETORIO	7,00	7,72	7,36		8,53	8,53	8,47	
	7,07	9,09	8,08		8,00	8,42	8,21		7,84	8,25	8,05		9,04	9,04	9,02	
	6,29	8,83	7,56		7,03	5,34	6,19	SUPLETORIO	5,04	6,40	5,72	SUPLETORIO	8,06	8,06	7,94	
	9,52	9,48	9,50		7,78	7,42	7,60		7,69	7,82	7,76		8,91	8,91	9,05	
	7,84	7,11	7,48		8,21	8,48	8,35		8,15	8,46	8,31		9,00	9,00	9,19	
	7,93	8,17	8,05		5,70	7,10	6,40	SUPLETORIO	7,00	7,72	7,36		8,53	8,53	8,47	
	7,07	9,09	8,08		8,00	8,42	8,21		7,84	8,25	8,05		9,04	9,04	9,02	
	6,29	8,83	7,56		7,03	5,34	6,19	SUPLETORIO	5,04	6,40	5,72	SUPLETORIO	8,06	8,06	7,94	
	7,07	9,09	8,08		8,00	8,42	8,21		7,84	8,25	8,05		9,04	9,04	9,02	

Nota. Colegio Aristóteles Bilingüe

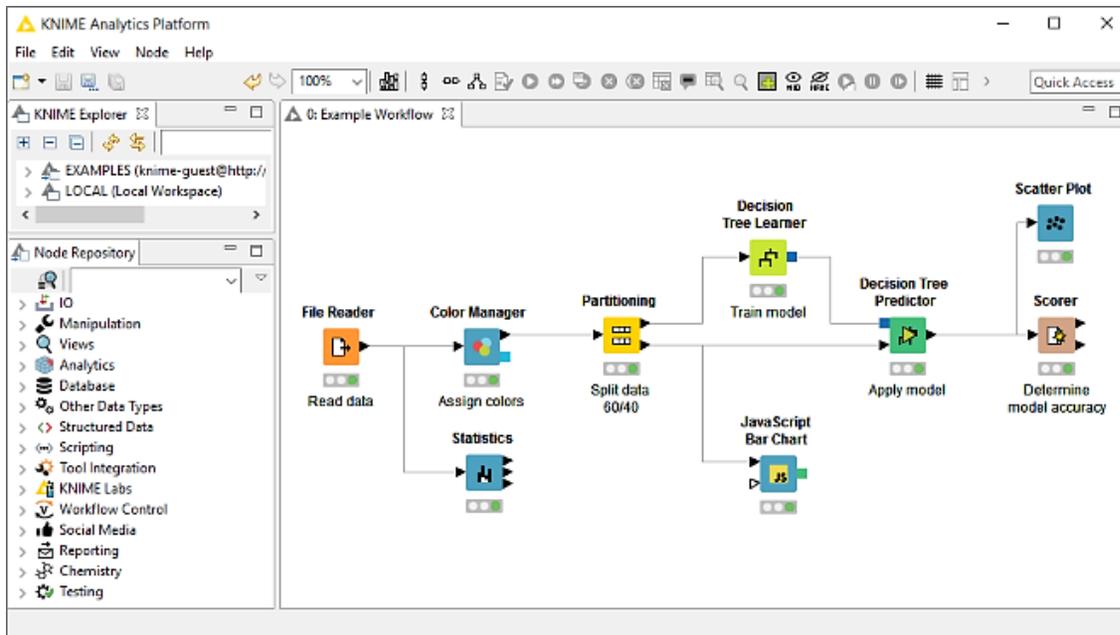
### **Selección de muestra**

De acuerdo con lo mencionado por la Comisión Económica para América Latina y el Caribe en la mayoría de los países de América Latina se produce la pérdida escolar y deserción en escalas muy altas, el reporte de la CEPAL establece que las edades comprendidas están entre 15 a 19 años (CEPAL, 2002) 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 en ese grupo donde se centra esta investigación.

### ***Modelado***

Antes de comenzar con el modelado, se determinó el formato de la base de datos a utilizar en cada modelo y una característica con base en el año de extracción de datos. Esta base de datos fue un modelo limpio de datos atípicos y previamente caracterizado. Una vez establecida la limpieza de los datos fue necesario emplear el programa Knime; herramienta ETL y de minería, análisis de datos mediante el cual se procedió a construir el ETL dado origen a la base de datos para el modelo.

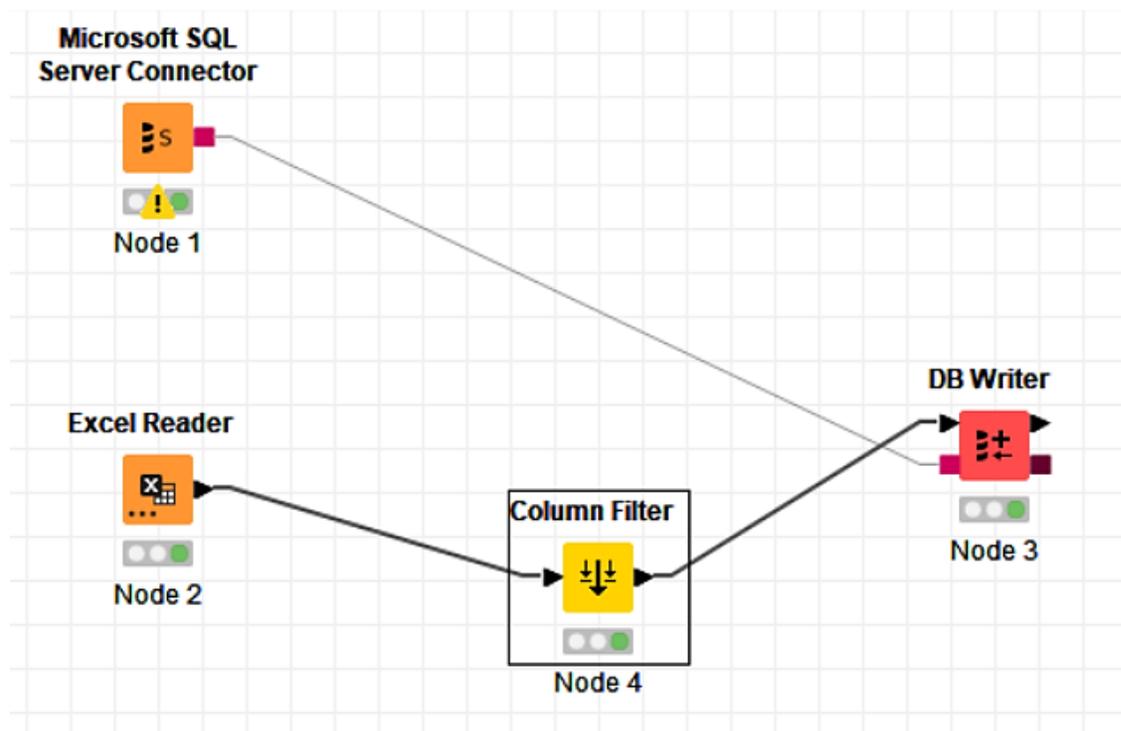
Figura 3.-

*Interfaz Knime*

Nota. Interfaz Heramienta Knime

Figura 4.-

Construcción del ETL



Nota. ETL

### Modelos de predicción

Para la construcción del modelo de base de datos se hizo una comparación de 3 modelos: regresión logística, Random Forest y Árbol de Decisión, los cuales se describen a continuación.

#### a) Regresión Logística

Es un modelo estadístico empleado para la predicción de la presencia o ausencia de un evento basado en la relación de una característica, obteniendo como resultado según varios criterios predictores un resultado determinado (IBM, 2022).

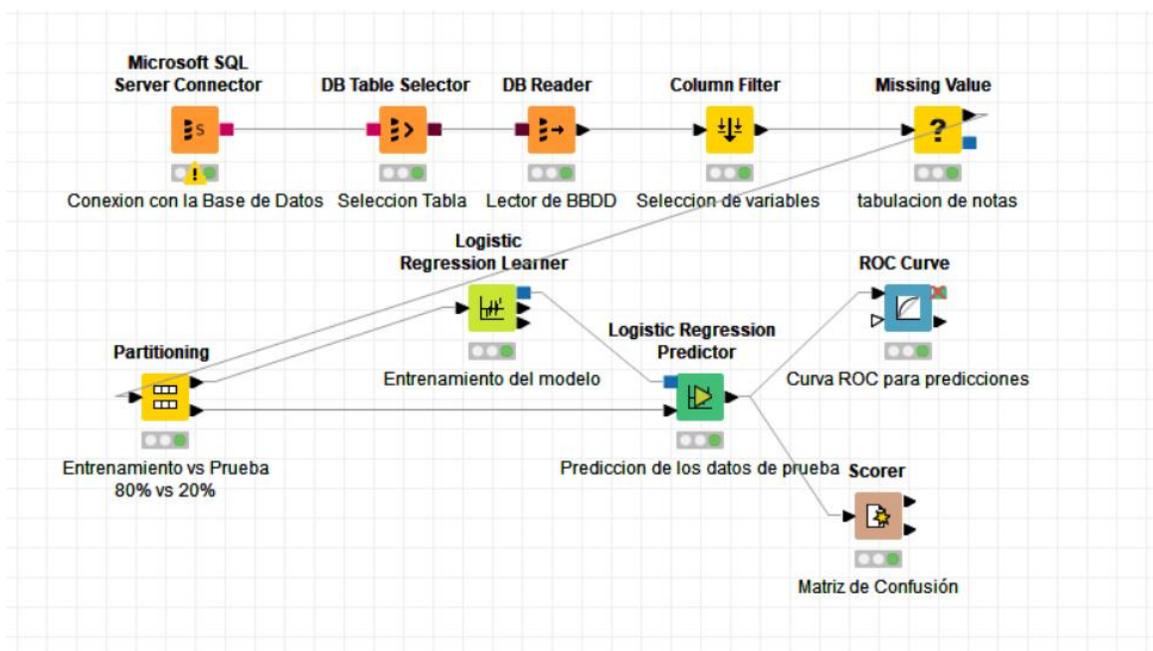
En el modelo de regresión logística pueden identificarse tres tipos:

- **Regresión logística binaria:** es aquella en la cual solo hay dos respuestas categóricas, el ejemplo más claro es la respuesta referente a la aprobación o reprobación de un estudiante.
- **Regresión logística multinomial:** es aquella en la cual las respuestas obtenidas representan a tres o más variables sin orden en específico.
- **Regresión logística ordinal:** contempla respuestas de tres o más variables, pero considera un orden específico de medición.

La construcción de un modelo de regresión logística se realiza en base a determinadas suposiciones, entre ellas:

- La regresión logística binaria requiere una respuesta binaria.
- El resultado esperado debe representarse por 1, si es lo deseado y 0 si no lo es.
- Incluir solo variables significativas.
- Las variables independientes deben mantener su independencia entre sí.
- El modelo se aplica solo a tamaños de muestras masivos.

Figura 5.\_

*Modelo de regresión logística*

Nota. Regresión Logística

## b) Random Forest

Es un modelo general de aprendizaje automático que puede realizar tareas de regresión y clasificación. También utiliza técnicas de reducción de tamaño, maneja valores perdidos, valores atípicos y otras operaciones básicas de exploración de datos. Es un método de aprendizaje por conjuntos que combina varios modelos débiles en un solo modelo fuerte (Gonzalez, 2018).

También conocido como bosque aleatorio, el modelo implementa múltiples algoritmos de árboles de decisión en lugar de uno solo, para la clasificación de nuevos objetos por atributo, cada árbol de decisión proporciona una clasificación, y la decisión final que recibe la mayor cantidad de "votos" es la conjetura del algoritmo.

Este modelo presenta algunas ventajas y desventajas, estas se resumen en la siguiente tabla.

**Tabla 29.-**

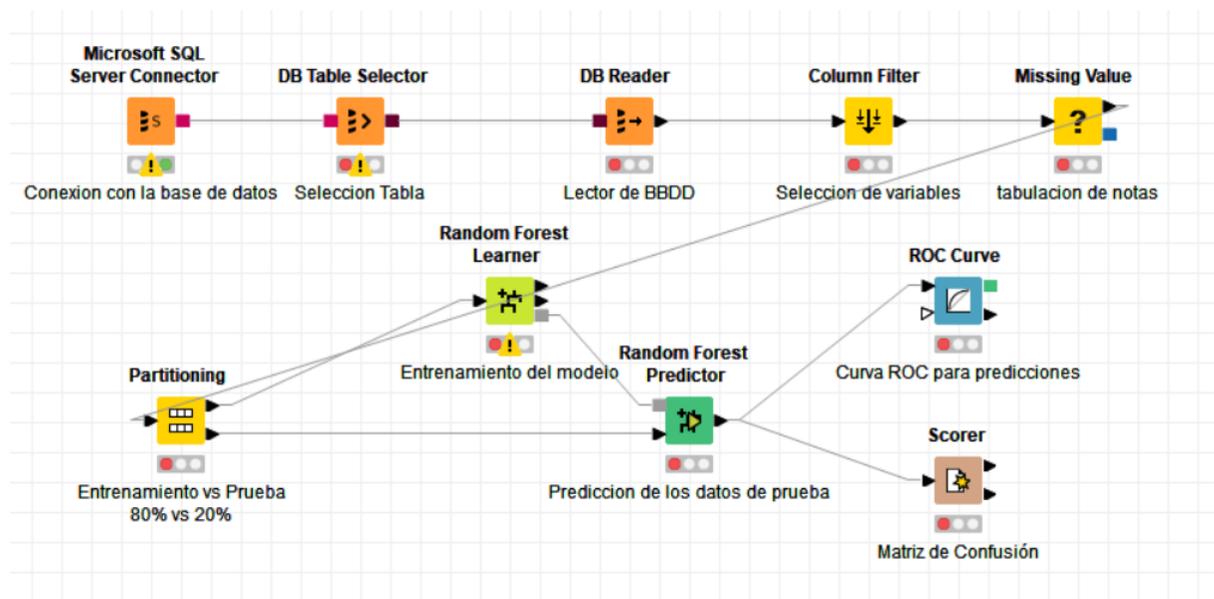
*Ventajas y desventajas de Random Forest*

Ventajas	Desventajas
Capacidad de resolver problemas de clasificación y regresión	Poco control sobre lo que hace el modelo
Procesamiento de grandes cantidades de datos	Posibilidad solo de probar diferentes parámetros y datos aleatorios
Efectividad en estimación de datos faltantes manteniendo la precisión.	
Probabilidad en Machine learning	

*Nota.* Adaptado de (Gonzalez, 2018)

**Figura 6.-**

*Random Forest*



Nota. Modelo Random Forest

### c) **Árbol de Decisión**

Los árboles de decisión son gráficos o modelos predictivos que evalúan ciertos parámetros para determinar resultados en función de las métricas utilizadas (Oliveira, 2021).

El propósito de crear un modelo de árbol de decisiones es contribuir a la evaluación de las posibles consecuencias de un conjunto particular de acciones y establecer el resultado.

Para Ferrero (2020) un árbol de decisión es un modelo predictivo que divide el espacio de estimadores agrupando observaciones con respuestas similares o valores de variables dependientes.

Este modelo presenta algunas ventajas y desventajas, estas se resumen en la siguiente tabla.

**Tabla 30.-**

*Ventajas y desventajas del árbol de decisión*

<b>Ventajas</b>	<b>Desventajas</b>
Construcción fácil	Sobre ajuste, es decir, el valor no se estima con el mismo índice de acierto.
Predicciones en base a predictores más importantes	Creación de outliers, que no predicen bien para nuevos casos.
Ante la falta de datos predice por medio del promedio de las ramas	Ineficientes en modelos de regresión
Útil para variables dependientes e independientes	Modelos complejos

Categorizar variables numéricas

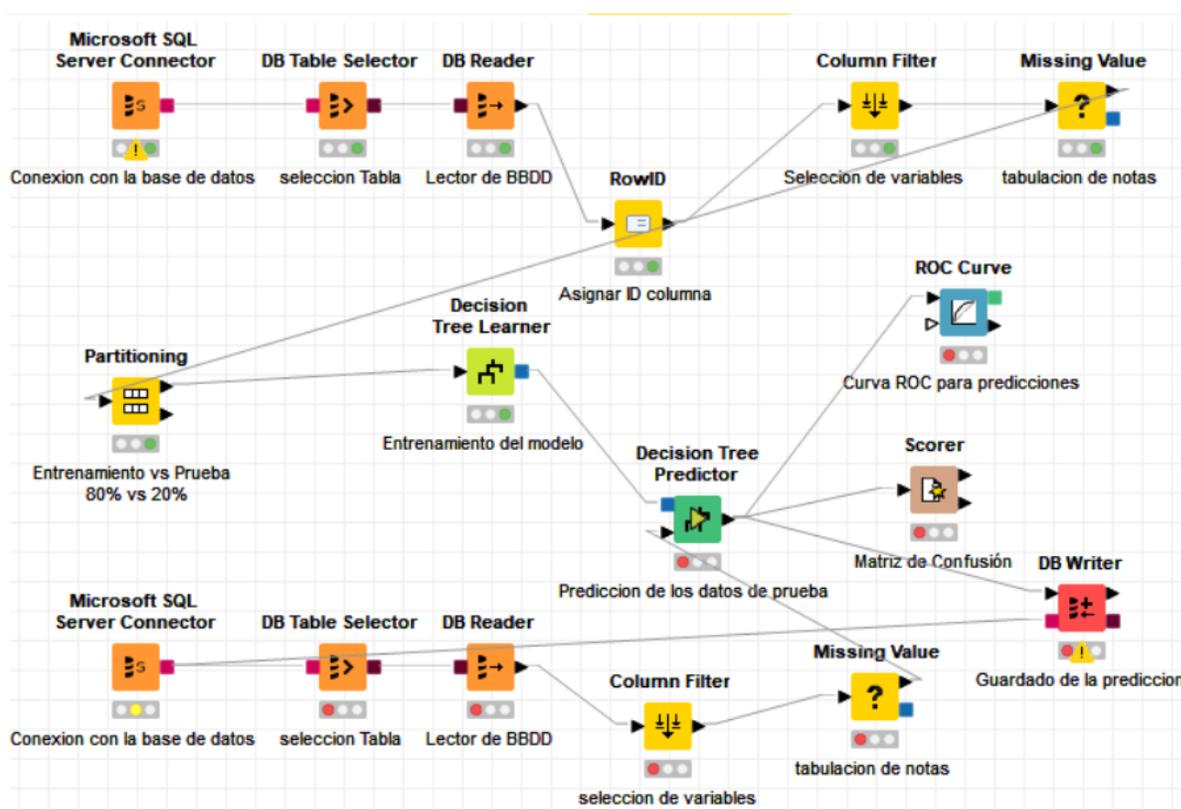
Arboles sesgados ante variables de diferentes dimensiones

Perdida de datos en categorizaciones numéricas continuas

Nota. Adaptado de (Ferrero, 2020)

Figura 7.-

Árbol de Decisión



Nota. Modelo Árbol de Decisión

## 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. Este tipo de análisis apunta a una caracterización totalmente completa, puesto que se presentan los datos relevantes como indicadores clave de desempeño del alumnado. 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.

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 (IBM, 2021).

En ese mismo contexto se aplica también la matriz de confusión o tabla de contingencia es una herramienta empleada en la inspección y evaluación del desempeño de un modelo de predicción (Sotaquirá, 2022). En cada fila se representa el número de predicciones de cada clase y en las columnas las instancias de la clase real.

La descripción de cada elemento de la matriz es la siguiente:

### Figura 8.-

*Matriz de confusión*

		Valores reales	
		Positivo	Negativo
Valores Predicción	Positivo	Verdaderos positivos (VP)	Falsos Positivos (FP)
	Negativo	Falsos negativos (FN)	Verdaderos negativos (VN)

*Nota.* Descripción Matriz de confusión

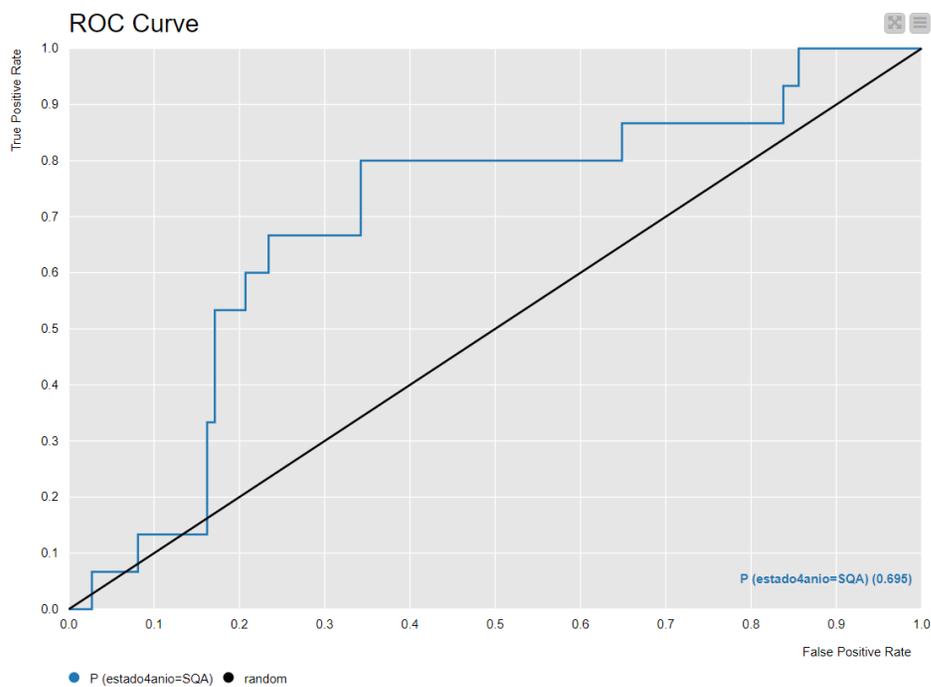
- **Verdadero positivo (VP):** número de ejemplos positivos que el modelo predice como positivos.
- **Falso positivo (FP):** número de ejemplos negativos que el modelo predice como positivos.
- **Falso negativo (FN):** número de ejemplos positivos que el modelo predice como negativos. FN
- **Verdadero negativo (VN):** número de ejemplos negativos que el modelo predice como negativos.

Los parámetros de evaluación usado son: Exactitud o accuracy, Recall o sensibilidad, Precisión, F1 score:

- **Exactitud o accuracy:** la fracción de predicciones que el modelo realizó correctamente.
- **Recall o sensibilidad:** indica la proporción de ejemplos positivos que están identificados correctamente por el modelo entre todos los positivos reales.
- **Precisión:** esta métrica está determinada por la fracción de elementos clasificados correctamente como positivo entre todos los que el modelo ha clasificado como positivos.
- **F1 score:** combina las métricas Precisión y Recall para dar un único resultado. Esta métrica es la más apropiada cuando tenemos conjuntos de datos no balanceados.

A continuación, se presentan los resultados de los tres modelos:

#### a) Regresión Logística

**Figura 9.-***Curva ROC de la Regresión Logística*

*Nota. Curva ROC*

**Figura 10.-***Matriz de confusión de la Regresión Logística*

estado4ani...	PA	SQA
PA	106	5
SQA	14	1

Correct classified: 107      Wrong classified: 19  
 Accuracy: 84,921 %      Error: 15,079 %  
 Cohen's kappa ( $\kappa$ ) 0,029

*Nota. Matriz de confusión*

- **Exactitud o accuracy**

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN}$$

$$\text{Exactitud} = \frac{106 + 1}{106 + 1 + 5 + 14}$$

$$\text{Exactitud} = 0,85$$

El modelo acierta el 85% de las veces

- **Recall o sensibilidad**

$$\text{Recall} = \frac{VP}{VP + FN}$$

$$\text{Recall} = \frac{106}{106 + 14}$$

$$\text{Recall} = 0.88$$

El modelo es capaz de identificar un 88% de los estudiantes que aprobarán

- **Precisión**

$$\text{Precisión} = \frac{VP}{VP + FP}$$

$$\text{Precisión} = \frac{106}{106 + 5}$$

$$\text{Precisión} = 0.95$$

El modelo acierta con una precisión del 95%

- **F1 score:**

$$\text{Exactitud} = \frac{2 * \text{Precisión} * \text{Recall}}{\text{Precisión} + \text{Recall}}$$

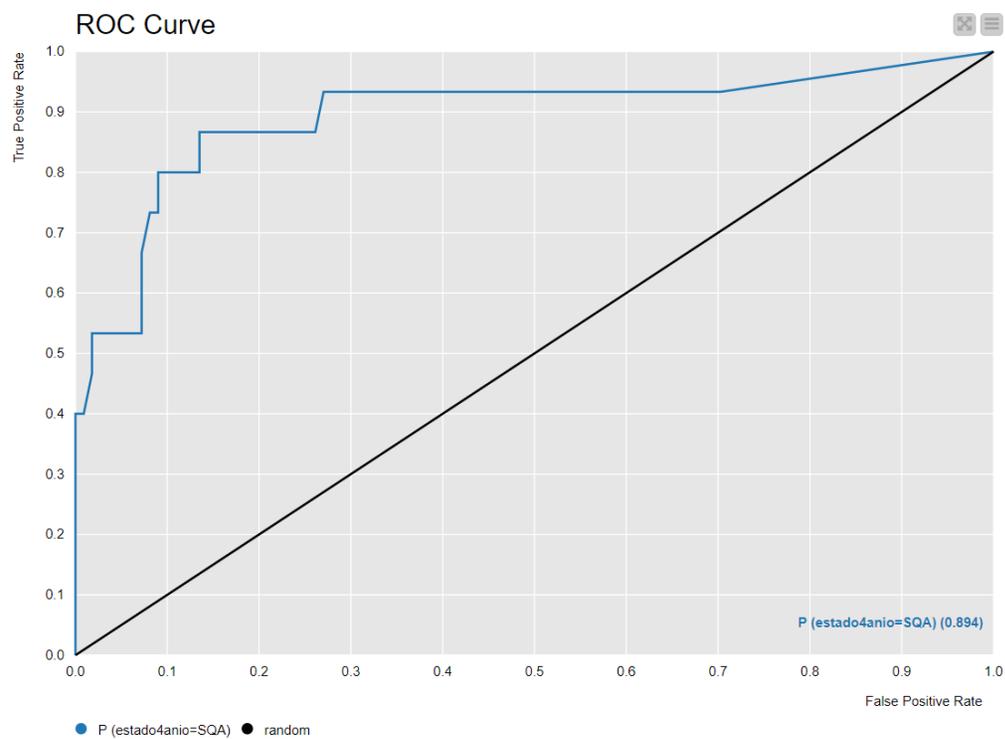
$$\text{Exactitud} = \frac{2 * 0.95 * 0.88}{0.95 + 0.88}$$

$$\text{Exactitud} = 0.92$$

## b) Random Forest

Figura 11.-

Curva ROC de Random Forest



Nota. Curva ROC

**Figura 12.-***Matriz de confusión de Random Forest*

estado4anio \ Prediction (e...	PA	SQA
PA	111	0
SQA	9	6

Correct classified: 117      Wrong classified: 9  
 Accuracy: 92,857 %      Error: 7,143 %  
 Cohen's kappa (κ) 0,54

*Nota.* Matriz de confusión

- **Exactitud o accuracy**

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN}$$

$$\text{Exactitud} = \frac{111 + 6}{111 + 6 + 0 + 9}$$

$$\text{Exactitud} = 0,93$$

El modelo acierta el 93% de las veces

- **Recall o sensibilidad**

$$\text{Recall} = \frac{VP}{VP + FN}$$

$$\text{Recall} = \frac{111}{111 + 9}$$

$$\text{Recall} = 0.93$$

El modelo es capaz de identificar un 93% de los estudiantes que aprobarán

- **Precisión**

$$\text{Precisión} = \frac{VP}{VP + FP}$$

$$\text{Precisión} = \frac{111}{111 + 0}$$

$$\text{Precisión} = 1$$

El modelo acierta en con una precisión del 100%

- **F1 score:**

$$\text{Exactitud} = \frac{2 * \text{Precisión} * \text{Recall}}{\text{Precisión} + \text{Recall}}$$

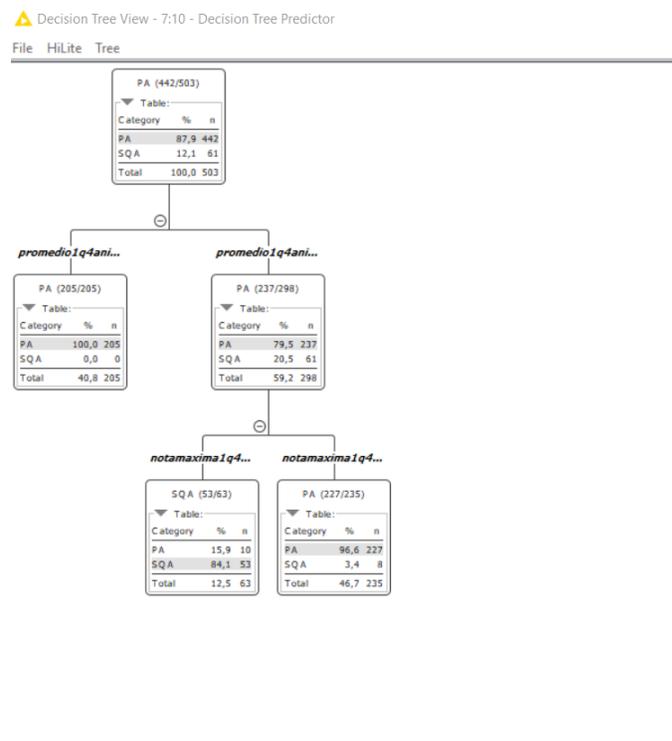
$$\text{Exactitud} = \frac{2 * 1 * 0.93}{1 + 0.93}$$

$$\text{Exactitud} = 0.96$$

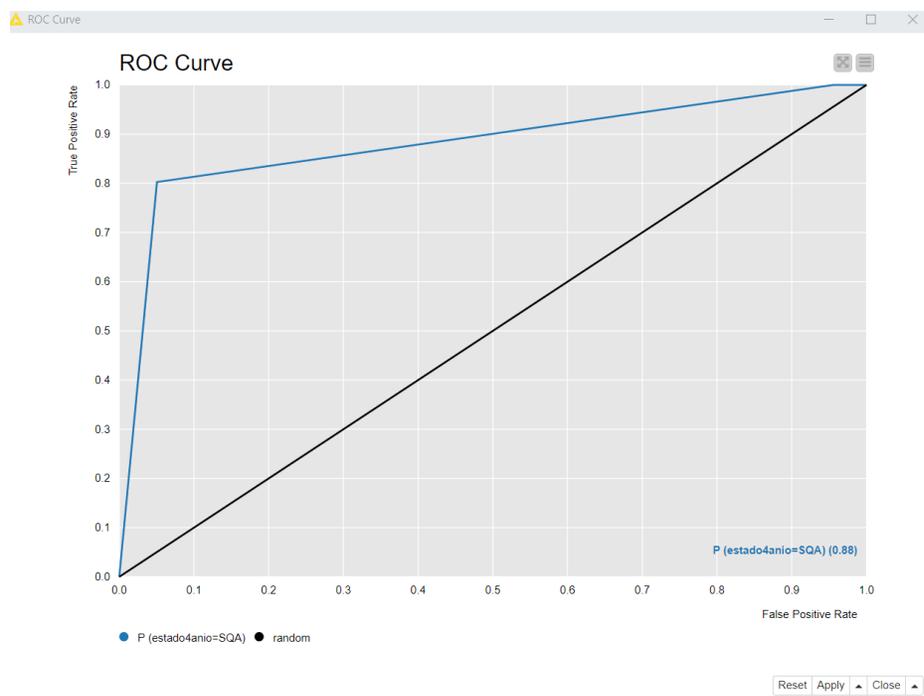
c) **Árbol de Decisión**

Figura 13.-

Árbol de problema aplicado a los datos



Nota. Diagrama árbol de decisión

**Figura 14.-***Curva ROC de Árbol de Decisión**Nota. Curva ROC***Figura 15.-***Matriz de confusión de Árbol de Decisión*

The figure is a window titled "Confusion Matrix - 7:..." displaying a confusion matrix. The matrix has two rows and two columns. The columns are labeled "PA" and "SQA". The rows are labeled "PA" and "SQA". The values in the matrix are: PA/PA: 301, PA/SQA: 16, SQA/PA: 15, SQA/SQA: 61. Below the matrix, there are summary statistics: "Correct classified: 362", "Wrong classified: 31", "Accuracy: 92,112 %", "Error: 7,888 %", and "Cohen's kappa ( $\kappa$ ) 0,748".

estado4ani...	PA	SQA
PA	301	16
SQA	15	61

Correct classified: 362      Wrong classified: 31  
 Accuracy: 92,112 %      Error: 7,888 %  
 Cohen's kappa ( $\kappa$ ) 0,748

*Nota. Matriz de confusión*

- **Exactitud o accuracy**

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN}$$

$$\text{Exactitud} = \frac{301 + 61}{301 + 61 + 16 + 15}$$

$$\text{Exactitud} = 0,92$$

El modelo acierta el 92% de las veces

- **Recall o sensibilidad**

$$\text{Recall} = \frac{VP}{VP + FN}$$

$$\text{Recall} = \frac{301}{301 + 15}$$

$$\text{Recall} = 0.95$$

El modelo es capaz de identificar un 95% de los estudiantes que aprobarán

- **Precisión**

$$\text{Precisión} = \frac{VP}{VP + FP}$$

$$\text{Precisión} = \frac{301}{301 + 16}$$

$$\text{Precisión} = 0.95$$

El modelo acierta en con una precisión del 95%

- **F1 score:**

$$\text{Exactitud} = \frac{2 * \text{Precisión} * \text{Recall}}{\text{Precisión} + \text{Recall}}$$

$$\text{Exactitud} = \frac{2 * 0.95 * 0.95}{0.95 + 0.95}$$

$$\text{Exactitud} = 0.95$$

### **Coeficiente Kappa**

Es un coeficiente que posibilita la determinación de la concordancia entre datos, aplicada a la matriz de confusión, contribuye a la determinación de si la clasificación observada es concordante con la clasificación realizada por el clasificador, es decir la predicha, la escala de valoración del coeficiente Kappa se presenta a continuación en la Tabla 31.

**Tabla 31.-**

*Valoración del coeficiente Kappa*

<b>Rango del Coeficiente Kappa</b>	<b>Grado de concordancia</b>
< 0,00	Sin acuerdo
>0,00 - 0,20	Insignificante
0,21 - 0,40	Discreto
>0,41 - 0,60	Moderado
0,61 - 0,80	Sustancial
0,81 - 1,00	Casi perfecto

*Nota.* (Menacho, 2017)

Los resultados del coeficiente Kappa para cada modelo se presentan en la Tabla 32.

**Tabla 32.-**

*Resultados del coeficiente Kappa*

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

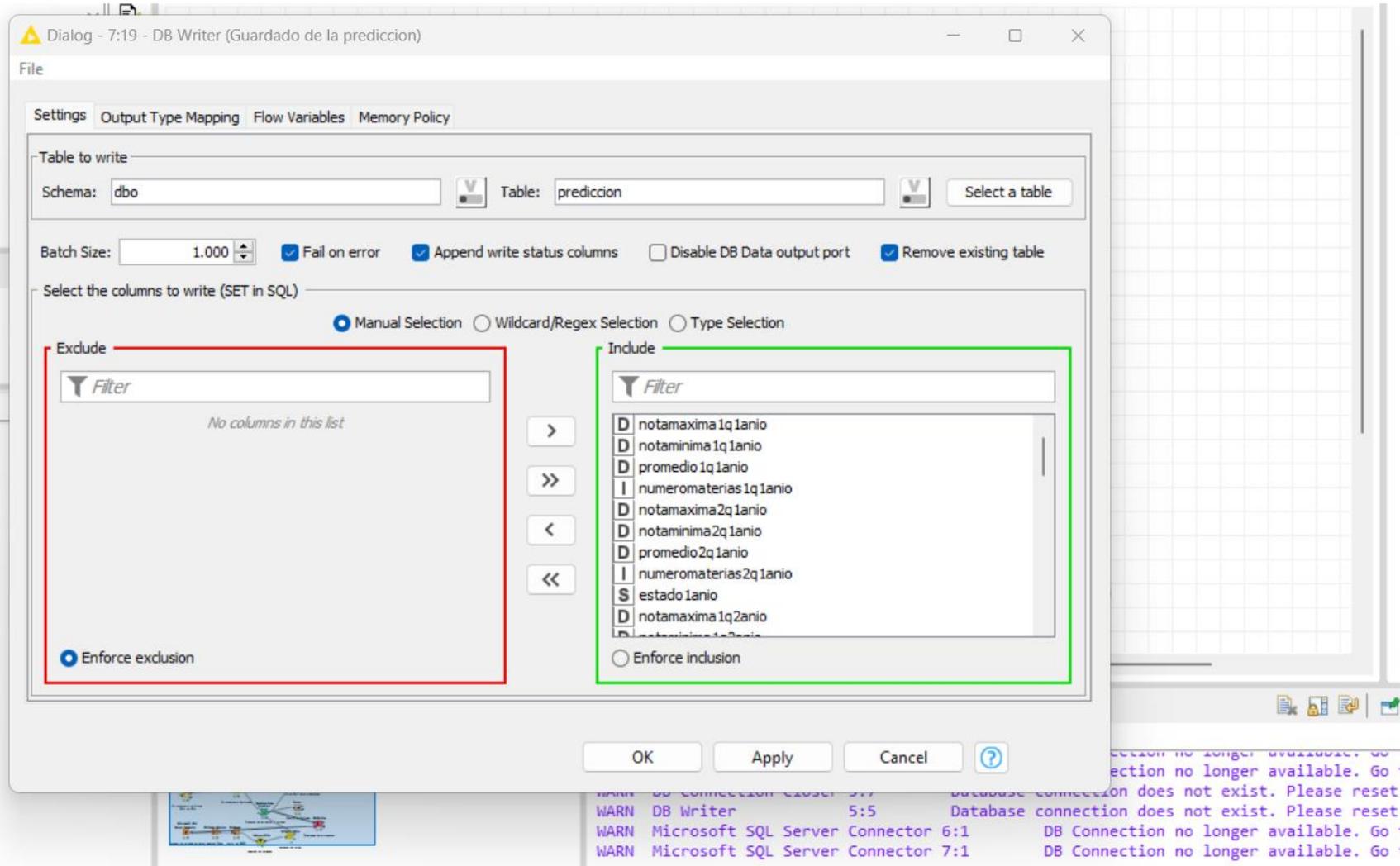
*Nota.* (Menacho, 2017)

### ***Despliegue***

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.

Figura 16.-

Ejecución del modelo y guardado del resultado de predicción en la Base de Datos

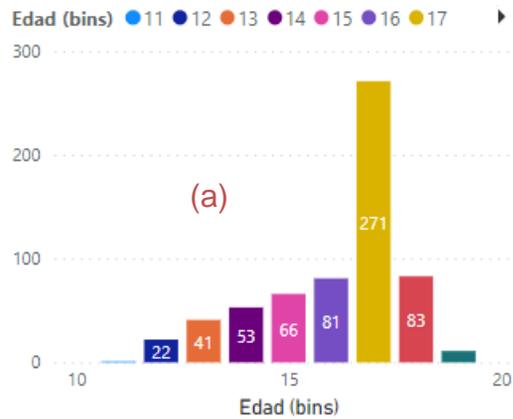


Nota. Ejecución del modelo

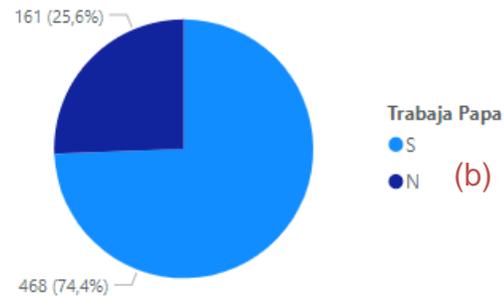
Figura 17.-

Dashboard de BI (Business Intelligence)

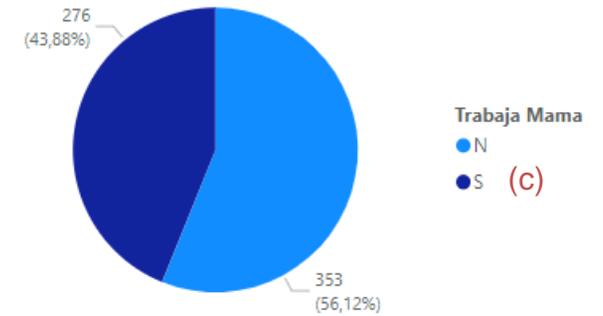
Grupos por edad Estudiantes



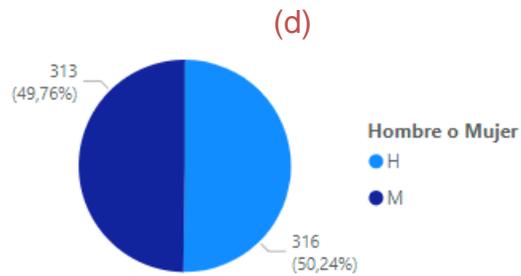
Porcentaje de Papás de estudiantes que trabajan



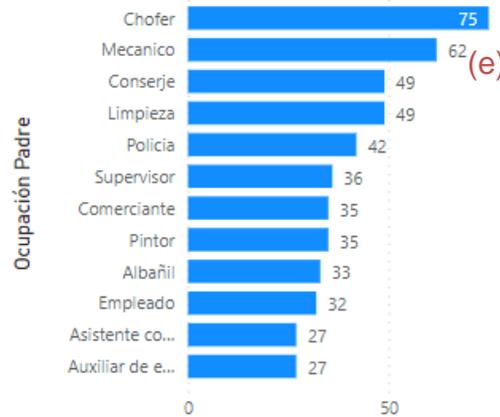
Porcentaje de Mamás de estudiantes que trabajan



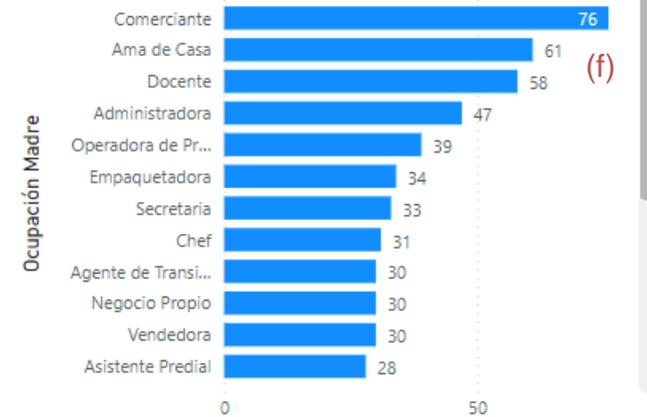
Porcentaje de estudiantes Hombres o Mujeres



Ocupacion de los padres de los estudiantes



Ocupacion de las madres de los estudiantes

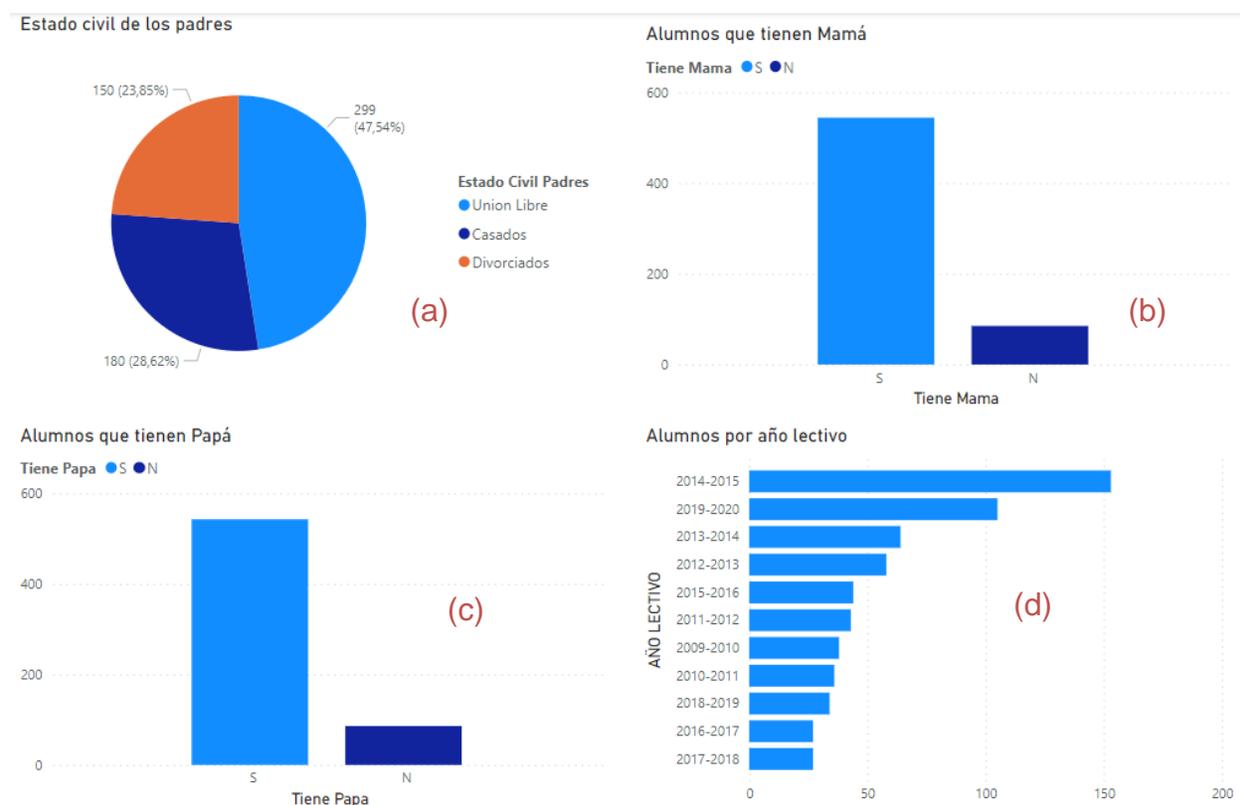


Nota. Dashboard de Business Intelligence

La Figura (a) presenta la gráfica del filtro edad y el número de estudiantes, en la que el mayor número de estudiantes están en la edad de 17 años. Así mismo en la Figura (b) se analiza si los padres de familia trabajan obteniendo que el 74.4% si trabaja y el 25.6% no lo hace, y de forma similar en la Figura (c) se analiza a las madres evidenciando que el 56.12% no trabaja y el 43,88% si trabaja. En la Figura (d) se presenta que el 50.21% de son hombres y el 49.76% son estudiantes mujeres y finalmente en la Figura (e) y Figura (f) se presenta de forma dinámica las ocupaciones que poseen los padres y las madres de los estudiantes respectivamente.

### Figura 18.-

#### Dashboard de BI (Business Intelligence)



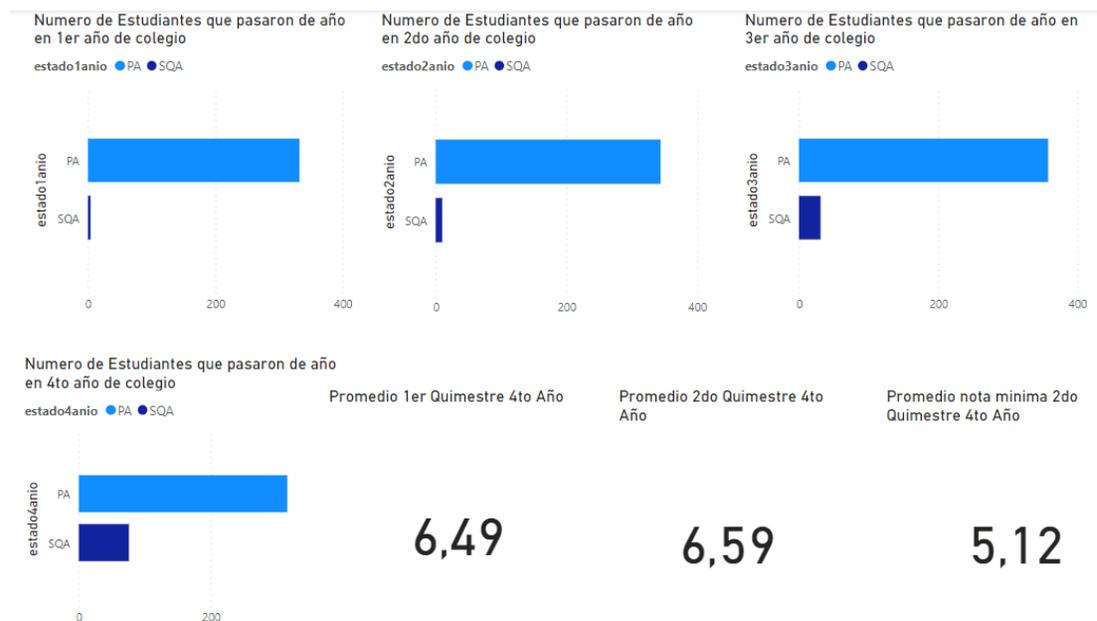
Nota. Dashboard de Business Intelligence

En la Figura 18 (a) se analizó el estado civil de los padres teniendo que 47.54% están en unión libre, 23.85% son divorciados y 28.62% son casados, en la Figura 18(b) y Figura 18(c) se detalla

si los estudiantes tienen madre y padre respectivamente en los dos gráficos se observa una predominancia del sí, y en la Figura 18(d), se presenta las métricas de los estudiantes a lo largo del periodo de 2002 a 2020.

**Figura 19.-**

*Dashboard de BI (Business Intelligence)*



*Nota.* Dashboard de Business Intelligence

En la Figura se presenta los criterios de aprobación y reprobación de los estudiantes en los años 1, 2, 3, 4, donde PA significa Paso de año y SQA significa Se quedó de año, en todos los casos se evidencia que existe una mayoría que aprueba el ciclo escolar.

En la Tabla 33 se presenta los resultados de la probabilidad P de que se apruebe o no el año escolar cada estudiante analizado y en la Figura se presenta el resumen de si corresponden a hombres y mujeres, obteniendo que el 53.69% son hombres con probabilidad de reprobación y el 46.31% son mujeres que pueden reprobare.

**Tabla 33.-**

*Probabilidad de aprobación o reprobación de año escolar en estudiantes de cuarto año*

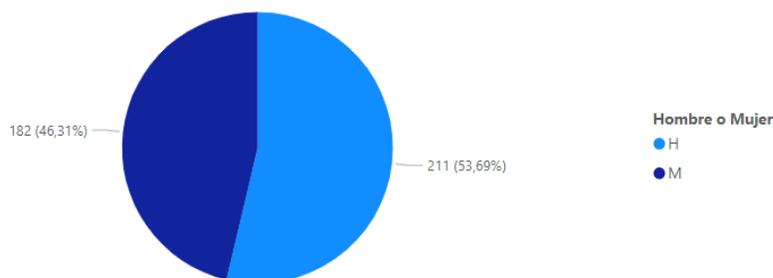
NOMBRES Y APELLIDOS	estado4anio	P (estado4anio=SQA)	P (estado4anio=PA)	Prediction (estado4anio)
ABAD MARTINEZ JONATHAN SANTIAGO	PA	0,03	0,97	PA
ABATA GUAMANZARA ALEXIS MARCELO	PA	0,03	0,97	PA
AGUILAR CAMPAÑA JORDAN ALEXIS	PA	0,03	0,97	PA
AGUIRRE ESTEVEZ CARLOS DAVID	PA	0,03	0,97	PA
ALAVA MENENDEZ ADDYN EMANUEL	PA	0,03	0,97	PA
ALAVA URBINA FIDEL ANDERSON	PA	0,03	0,97	PA
ALAVA URBINA TERESA YAMILET	PA	0,03	0,97	PA
ALBUJA BONILLA JOHN FRANKLIN	SQA	0,84	0,16	SQA
ALCOCER SANTAMARIA DENISSE JOSETH	PA	0,03	0,97	PA
ALDAZ NARANJO FRANCIS BERNARDA	PA	0,03	0,97	PA
ALMACHI CHANCUSIG MARILYN NICOLE	PA	0,03	0,97	PA
ALMEIDA LOPEZ BRANDON AARON	SQA	0,84	0,16	SQA
ALTAMIRANO VILLALOBOS ANIBAL ISSAC	PA	0,03	0,97	PA
ALVARADO IDROVO RICARDO GABRIEL	PA	0,03	0,97	PA
ALVARADO VALENCIA JOSE ALEJANDRO	PA	0,03	0,97	PA
ANALUISA BORJA STEFANY DANIELA	PA	0,03	0,97	PA

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

*Nota.* Predicción de aprobación o reprobación

**Figura 20.-**

*Proyección de aprobación o reprobación de año escolar en estudiantes de cuarto año por género*



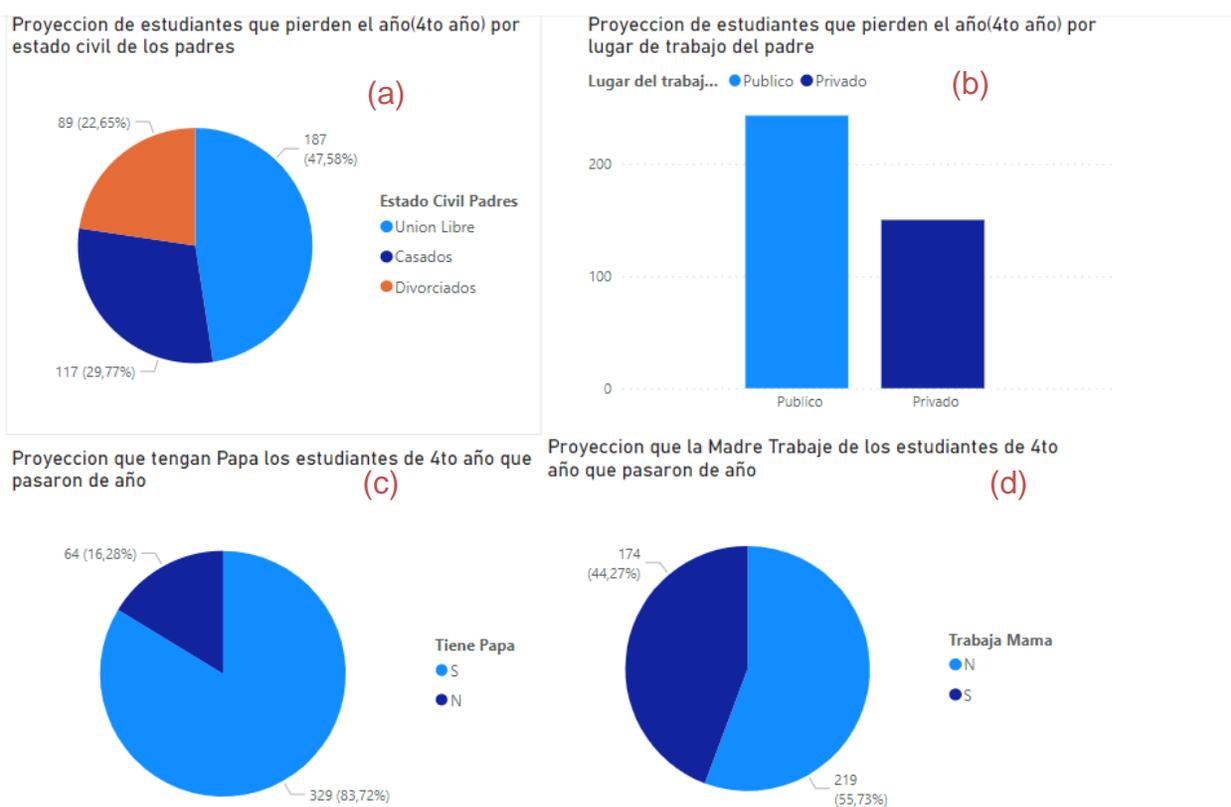
*Nota.* Proyección de aprobación o reprobación según el género

En la Figura (a), se presenta la probabilidad de pérdida de año donde los estudiantes propensos a reprobación provienen con el 47.58% de padres en unión libre, 22.65% de padres divorciados y

29.77% de padres casados. En la Figura (b) se evidencia que el mayor porcentaje de los estudiantes que no aprueban el año escolar corresponden a aquellos cuyos padres laboran en el sector público siendo un valor inferior a aquellos que laboran en el sector privado, por otro lado en la Figura (c) y Figura (d) se observa que los estudiantes tienen a no pasar de año cuando no cuentan con una de las figuras paternas, siendo así que el 83.72% pierden cuando no tienen padre y el 55.73% cuando no tienen madre.

### Figura 21.-

*Proyección de aprobación o reprobación de año escolar en estudiantes de cuarto año*



*Nota.* Proyección de aprobación o reprobación de año escolar

## Capítulo 4

### Análisis y resultados

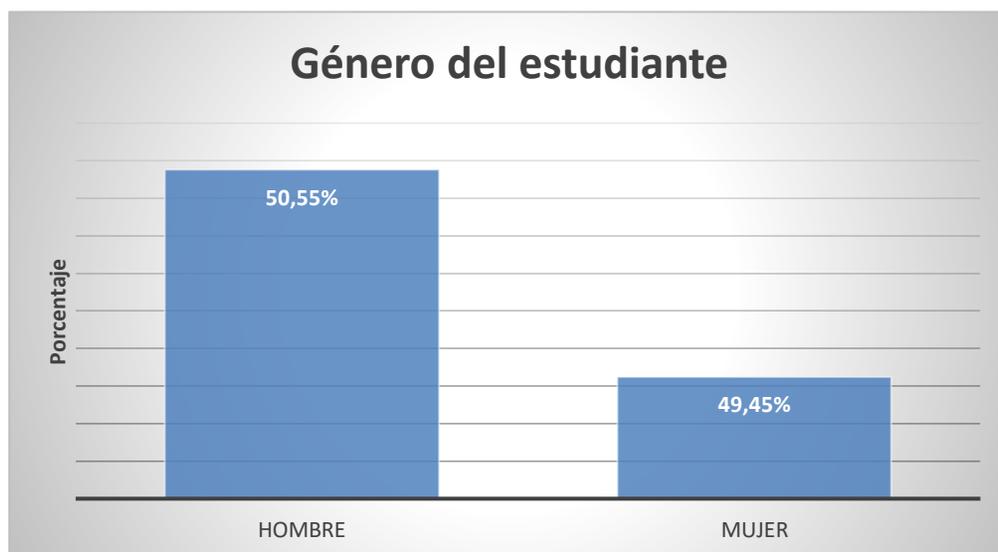
#### Análisis de los datos obtenidos en la depuración de la base de datos

En este apartado se lleva a cabo el análisis, posterior a la tabulación de los datos obtenidos por medio de la metodología CRISP-DM de minería de datos. En primera instancia, se analizaron las variables relacionadas con la ocupación de los padres de familia de los estudiantes tomados en consideración para el desarrollo del estudio. En las siguientes figuras se detallan los resultados obtenidos. Es importante mencionar que la muestra seleccionada contempla un intervalo aproximado de 11 años, es decir, abarca desde el año lectivo 2009 – 2010 hasta el año lectivo 2019 – 2020.

#### *Análisis socioeconómico de la familia de los estudiantes*

**Figura 22.-**

*Género de los estudiantes de la institución educativa*

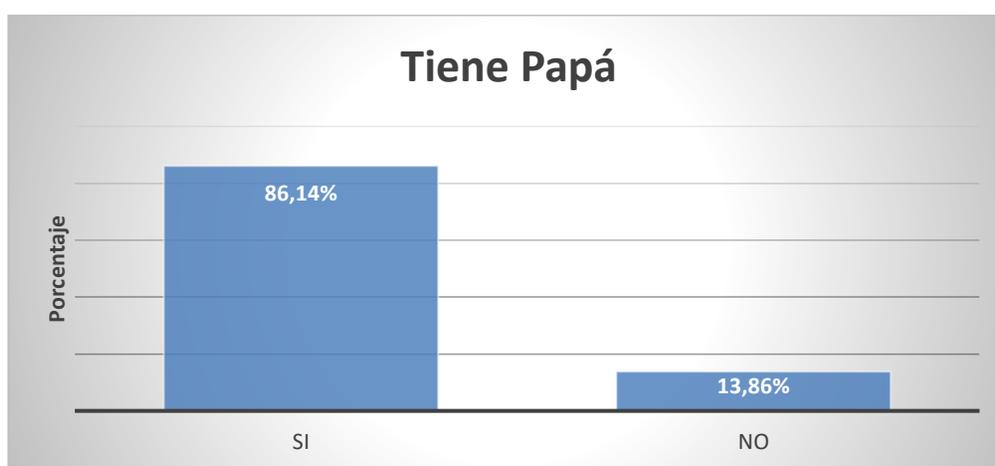


*Nota.* Imagen elaborada a partir de los datos

En lo que respecta al género de los estudiantes que formaron parte del presente estudio no se presentan diferencias significativas al momento del análisis de los datos. Se puede apreciar que el 50,55% de estudiantes son hombres, mientras que el 49,45% son mujeres. Esta situación permite apreciar que la muestra de estudiantes seleccionada es homogénea, es decir, una muestra pareja.

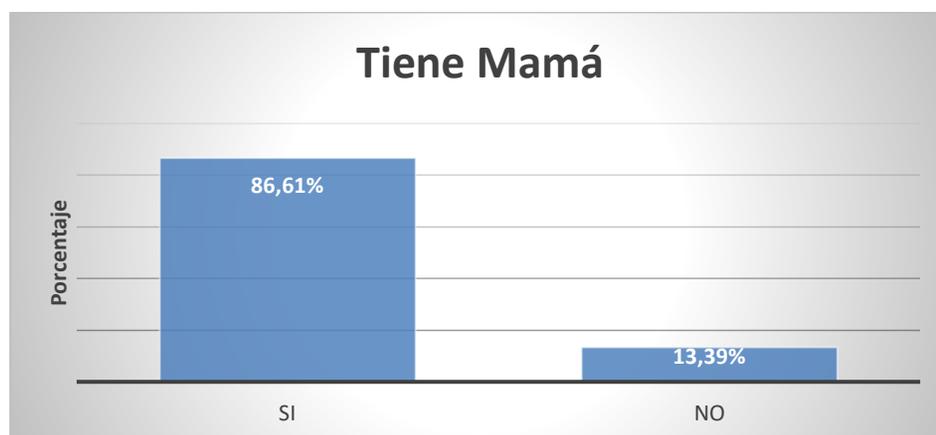
**Figura 23.-**

*Presencia del padre en el hogar*



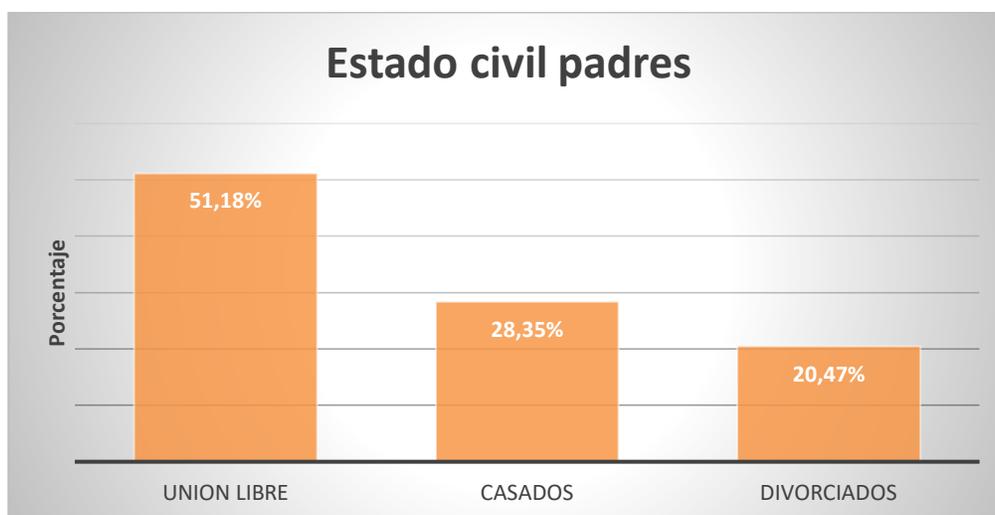
*Nota.* Imagen elaborada a partir de los datos

Con relación a la presencia del padre en el hogar, el 86,14% tienen padre, mientras que el 13,86% no tienen padre. Se aprecia que la mayoría de estudiantes analizados para el periodo de tiempo seleccionado cuentan con su padre en casa, lo que conlleva a establecer que estos cuentan con el apoyo del mismo para el desarrollo de sus actividades académicas.

**Figura 24.-***Presencia de la madre en el hogar*

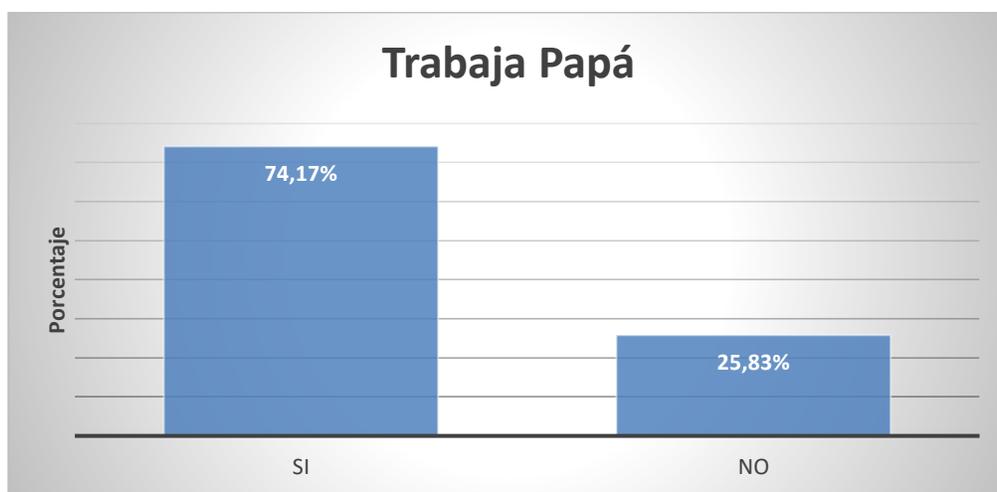
*Nota.* Imagen elaborada a partir de los datos

Como un complemento de la figura anterior, cerca del 86,61% de los estudiantes tomados en consideración para el estudio cuentan con su madre en casa, mientras que el 13,39% no cuentan con su madre en casa. Los porcentajes obtenidos para esta pregunta son similares a las cifras obtenidas para la presencia del padre en casa, no obstante, no se puede afirmar si los estudiantes que no tienen madre no tienen padre tampoco.

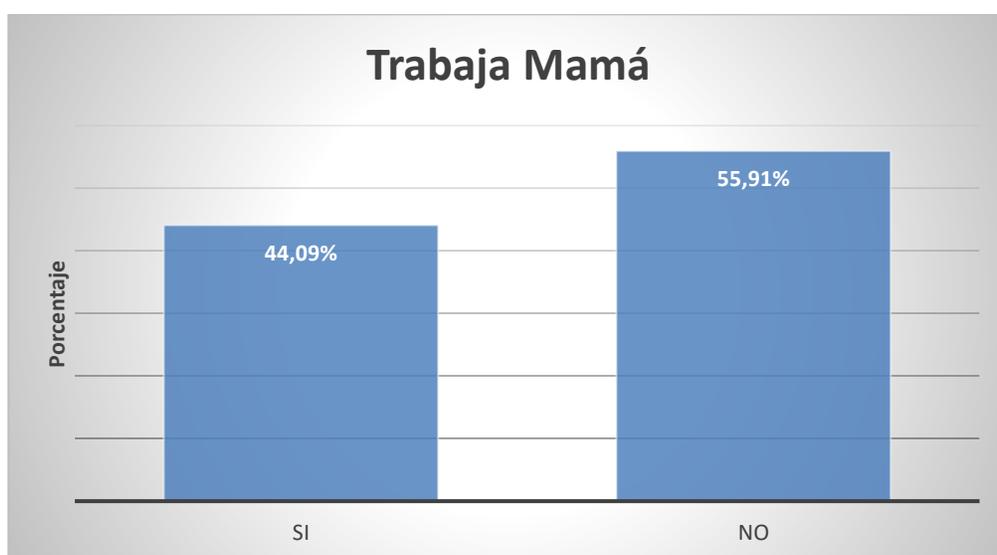
**Figura 25.-***Estado civil de los padres*

*Nota.* Imagen elaborada a partir de los datos

En lo que respecta al estado civil de los padres, se puede apreciar que los resultados obtenidos no presentan una diferencia significativa en lo referente a unión libre, matrimonios o divorcios. Se aprecia que la gran mayoría de padres de familia viven en unión libre (51,18%). El porcentaje de personas casados se sitúa en 28,35%. Finalmente, el porcentaje de divorcios se ubica en un 20,47%.

**Figura 26.-***Trabajo del padre*

*Nota.* Imagen elaborada a partir de los datos

**Figura 27.-***Trabajo de la madre*

*Nota.* Imagen elaborada a partir de los datos

Con respecto a la empleabilidad de los padres de familia, para el caso de los padres, el 74,17% poseen un trabajo, mientras que el 25,83% no poseen un trabajo. Estas cifras se contrastan directamente con los porcentajes obtenidos para las madres de familia, donde cerca del 55,91% no tienen trabajo, mientras que el 44,09% disponen de una plaza de trabajo. Se puede apreciar que los padres de familia tienen un mayor nivel de empleabilidad con respecto a las madres de familia. Esta situación se explica con el hecho relacionado de la tradición de que el hombre debe procurar por la situación económica de su hogar, mientras que la mujer debe quedarse en casa al cuidado de los niños y los quehaceres domésticos.

### ***Análisis rendimiento académico***

Posterior al análisis socioeconómico realizado, se procedió al análisis de los resultados adquiridos con base en la depuración de las notas para cada estudiante durante el periodo seleccionado. En primera instancia se analizó el porcentaje de estudiantes aprobados y dejados en suspenso durante el periodo de estudio seleccionado, obteniendo los resultados mostrados en la siguiente tabla y figura.

**Tabla 34:**

*Porcentaje de estudiantes aprobados y reprobados durante el periodo de estudio*

<b>Porcentaje de aprobados y reprobados</b>			
<b>Año lectivo</b>	<b>Estado</b>	<b>Frecuencia</b>	<b>Porcentaje</b>
2009-2010	Aprobado	155	88,1%
	Reprobado	21	11,9%
	<b>Total</b>	<b>176</b>	
2010-2011	Aprobado	199	90,9%
	Reprobado	20	9,1%
	<b>Total</b>	<b>219</b>	

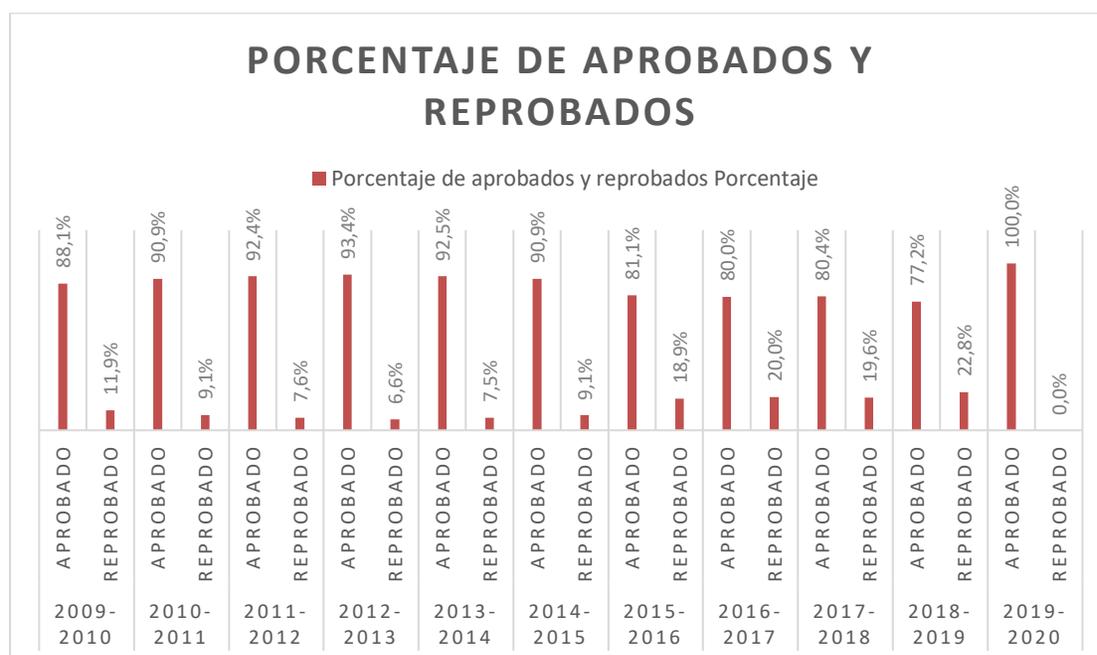
<b>Porcentaje de aprobados y reprobados</b>			
<b>Año lectivo</b>	<b>Estado</b>	<b>Frecuencia</b>	<b>Porcentaje</b>
2011-2012	Aprobado	244	92,4%
	Reprobado	20	7,6%
	<b>Total</b>	<b>264</b>	
2012-2013	Aprobado	282	93,4%
	Reprobado	20	6,6%
	<b>Total</b>	<b>302</b>	
2013-2014	Aprobado	270	92,5%
	Reprobado	22	7,5%
	<b>Total</b>	<b>292</b>	
2014-2015	Aprobado	240	90,9%
	Reprobado	24	9,1%
	<b>Total</b>	<b>264</b>	
2015-2016	Aprobado	120	81,1%
	Reprobado	28	18,9%
	<b>Total</b>	<b>148</b>	
2016-2017	Aprobado	96	80,0%
	Reprobado	24	20,0%
	<b>Total</b>	<b>120</b>	
2017-2018	Aprobado	82	80,4%
	Reprobado	20	19,6%
	<b>Total</b>	<b>102</b>	
2018-2019	Aprobado	88	77,2%
	Reprobado	26	22,8%

Porcentaje de aprobados y reprobados			
Año lectivo	Estado	Frecuencia	Porcentaje
	<b>Total</b>	<b>114</b>	
2019-2020	Aprobado	105	100,0%
	Reprobado	0	0,0%
	<b>Total</b>	<b>105</b>	

Nota. Tabla elaborada a partir de los datos

**Figura 28.-**

*Porcentaje de aprobados y reprobados*



Nota. Imagen elaborada a partir de los datos

Los resultados obtenidos en la figura anterior permiten apreciar diferentes aspectos y dimensiones con base en las notas de los estudiantes. A continuación, se presentan los puntos más importantes del análisis realizado a los datos anteriores:

- Durante todo el periodo de estudio (2009 – 2020) se puede apreciar que el porcentaje de aprobación de niveles tiene una media de 90%, con respecto al aproximado de 10% de estudiantes que se quedaron en suspenso durante el periodo de estudio.
- El año 2020 presenta un porcentaje de aprobación del 100%. Esta situación se debe a la pandemia de COVID – 19 que surgió durante el año mencionado, lo que generó que los estudiantes en su totalidad aprueben el año lectivo, debido a que el Ministerio de Educación dispuso que ningún estudiante puede perder el año por la emergencia sanitaria mundial del COVID, que fue una limitante del adecuado acceso a la educación.
- El año lectivo con menor porcentaje de aprobación es el periodo 2018 – 2019. En este periodo en particular, el porcentaje de aprobación se ubica en 77,2%, presentando un nivel de estudiantes en supletorio o reprobados mayor al 20%.
- Sin contar el año 2020, el año con mayor porcentaje de aprobación dentro del periodo de estudio analizado es el año lectivo 2012 – 2013. En este año se presentó un porcentaje de aprobación de 93,4%; presentando también un porcentaje de reprobación de 6,6%.

**Tabla 35:***Promedio por quimestre*

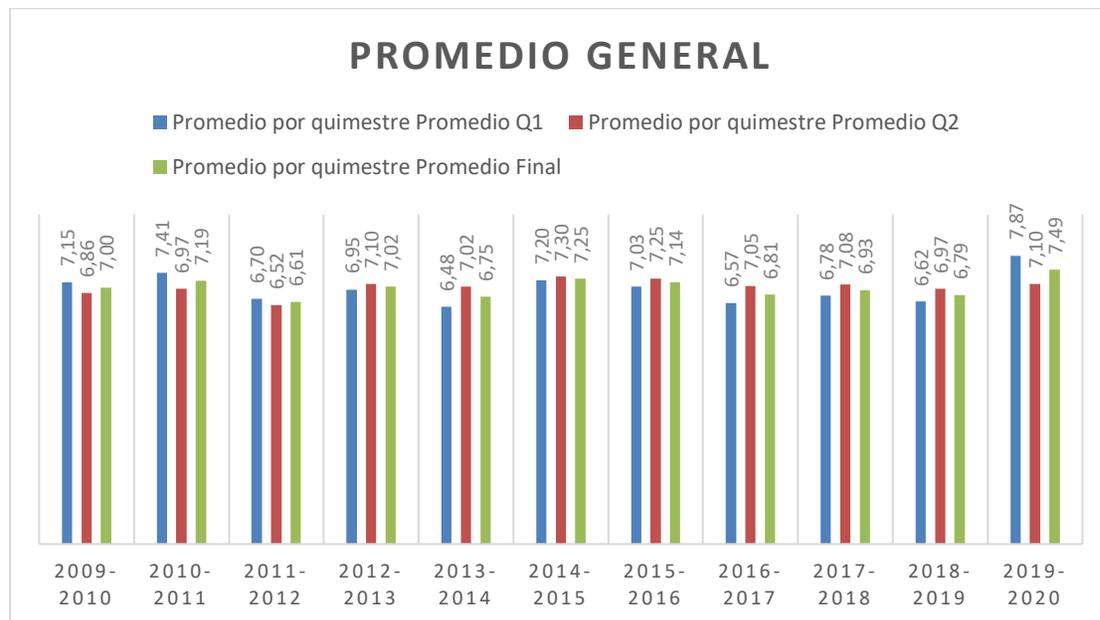
<b>Promedio por quimestre</b>			
<b>Año lectivo</b>	<b>Promedio Q1</b>	<b>Promedio Q2</b>	<b>Promedio Final</b>
2009-2010	7,15	6,86	7,00
2010-2011	7,41	6,97	7,19
2011-2012	6,70	6,52	6,61
2012-2013	6,95	7,10	7,02
2013-2014	6,48	7,02	6,75
2014-2015	7,20	7,30	7,25

2015-2016	7,03	7,25	7,14
2016-2017	6,57	7,05	6,81
2017-2018	6,78	7,08	6,93
2018-2019	6,62	6,97	6,79
2019-2020	7,87	7,10	7,49

Nota. Tabla elaborada a partir de los datos

**Figura 29.-**

*Promedio general para cada quimestre*



Nota. Imagen elaborada a partir de los datos

En lo que respecta al promedio general obtenido para cada año lectivo analizado, se puede analizar lo siguiente de la figura anterior.

- El promedio final de cada uno de los años lectivos analizados oscila en un promedio de 6,5 a 7,5

- Como en el análisis de aprobados y quedados a suspenso, el año 2020 presenta la particularidad de ser el año de mayor promedio de los analizados durante el estudio.
- El año lectivo con menor promedio de los analizados fue el 2011 – 2012, con un promedio general de 6,61.
- Se aprecia que las notas del segundo parcial, en su mayoría, son las que definen la clasificación promedio final, puesto que los estudiantes presentan un mejor desempeño en este periodo, a comparación del primer parcial, en donde presentan un menor nivel de rendimiento académico.

Con respecto al análisis del rendimiento académico para cada quimestre, en las siguientes tablas y figuras se presentan los resultados depurados por quimestre y por rango de aprobación o suspenso.

**Tabla 36:**

*Promedio obtenido para el quimestre 1*

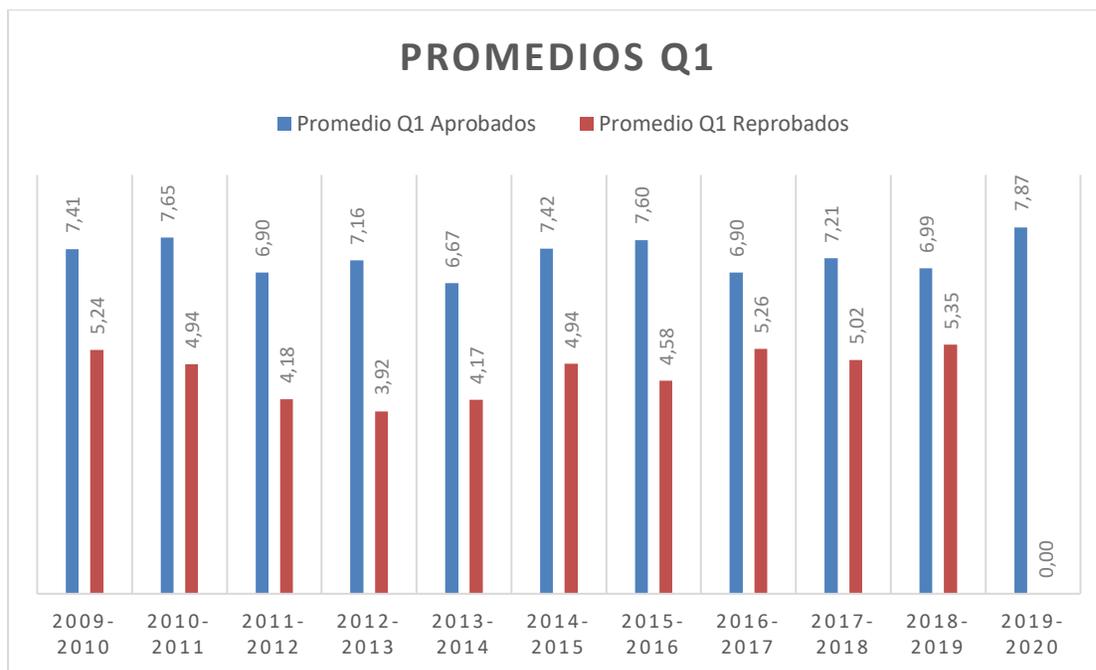
<b>Promedio Q1</b>		
<b>Año lectivo</b>	<b>Aprobados</b>	<b>Reprobados</b>
2009-2010	7,41	5,24
2010-2011	7,65	4,94
2011-2012	6,90	4,18
2012-2013	7,16	3,92
2013-2014	6,67	4,17
2014-2015	7,42	4,94
2015-2016	7,60	4,58
2016-2017	6,90	5,26
2017-2018	7,21	5,02

2018-2019	6,99	5,35
2019-2020	7,87	0,00

*Nota.* Tabla elaborada a partir de los datos

**Figura 30.-**

*Promedio obtenido para el quimestre 1*



*Nota.* Imagen elaborada a partir de los datos

De la tabla y figura anteriores se puede analizar los siguiente:

- El promedio de notas de los estudiantes aprobados para el primer parcial oscila en el intervalo de 7 a 7,6 (exceptuando al año 2020 del análisis). Esta información contrasta directamente con el promedio que presentaron los estudiantes que se quedaron en suspenso o perdieron el año, miso que se encuentra en un intervalo comprendido entre 4 a 5,4.

- El año de peor rendimiento para los estudiantes aprobados fue el periodo lectivo 2013 – 2014, con un promedio de 6,67; mientras que el año lectivo de menor rendimiento para los estudiantes de perdida fue el periodo 2012 – 2013, con un promedio de 3,92.
- Con respecto al mejor nivel de rendimiento, el periodo lectivo 2010 – 2011 fue el de mejor promedio para los estudiantes aprobados, con un puntaje de 7,65; mientras que para los estudiantes en perdida el mejor promedio se presentó durante el periodo lectivo 2018 – 2019, con un puntaje de 5,35.

Se puede apreciar entonces que el puntaje promedio de las notas obtenidas para el primer parcial es homogéneo. Es importante destacar el hecho de que los estudiantes reprobados presentaron un promedio menor a 6 durante el primer quimestre, situación que no pudo ser complementada y solventada en el segundo quimestre.

**Tabla 37:**

*Promedio obtenido para el segundo quimestre*

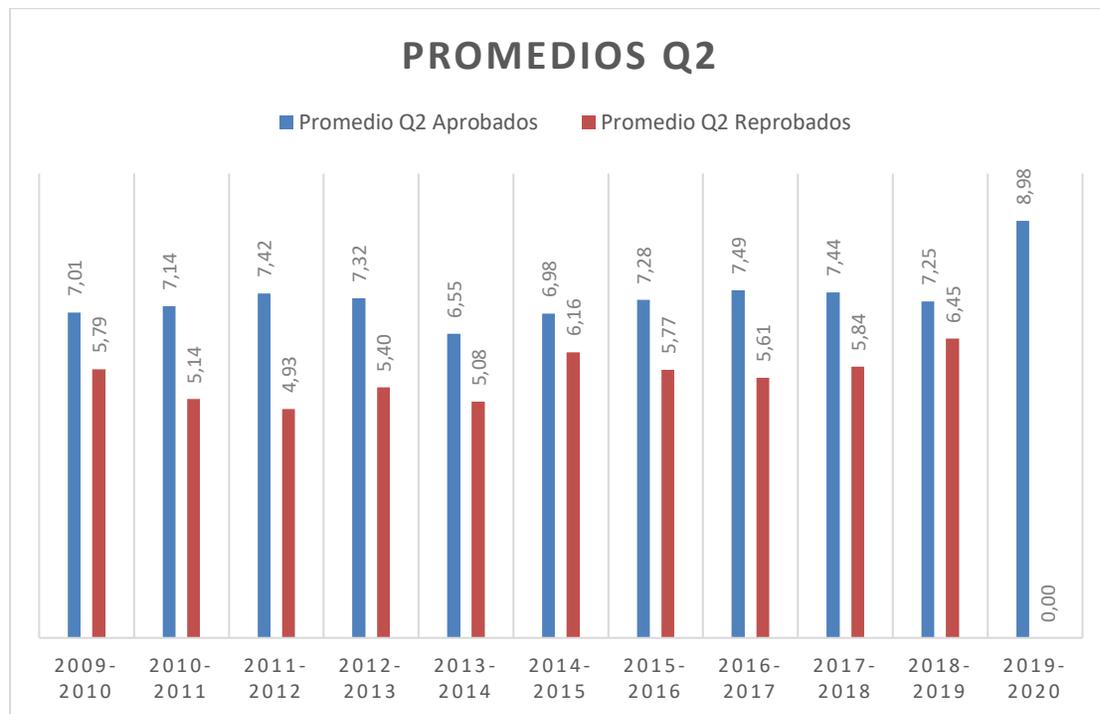
<b>Promedio Q2</b>		
<b>Año lectivo</b>	<b>Aprobados</b>	<b>Reprobados</b>
2009-2010	7,01	5,79
2010-2011	7,14	5,14
2011-2012	7,42	4,93
2012-2013	7,32	5,40
2013-2014	6,55	5,08
2014-2015	6,98	6,16
2015-2016	7,28	5,77
2016-2017	7,49	5,61
2017-2018	7,44	5,84

Promedio Q2		
Año lectivo	Aprobados	Reprobados
2018-2019	7,25	6,45
2019-2020	8,98	0,00

Nota. Tabla elaborada a partir de los datos

**Figura 31.-**

*Promedio obtenido para el segundo quimestre*



Nota. Tabla elaborada a partir de los datos

Con relación a los promedios obtenidos por los estudiantes para el segundo quimestre en cada uno de los años lectivos analizados, se pueden mencionar los siguientes resultados:

- El promedio general de los estudiantes aprobados se mantiene en un intervalo de 6,5 a 7,5, manteniendo la tendencia del primer parcial.
- Del mismo modo, las notas obtenidas por los estudiantes quedados en suspenso para el segundo quimestre se presentan en un intervalo de 4,9 a 6,4. Se puede apreciar que el intervalo del promedio aumenta con respecto al quimestre anterior.
- Nuevamente se excluye al año 2020 del análisis, puesto que durante el segundo quimestre del año lectivo 2019 – 2020 dio inicio la pandemia de COVID – 19, por ende, las actividades y notas se vieron alteradas con respecto a los años lectivos anteriores.
- Como se mencionó en el apartado anterior, los estudiantes que no presentaron una mejoría en el primer quimestre no emplean el segundo quimestre, lo que genera que estos se queden a suspenso o pierdan el año.

## Conclusiones

- La revisión teórica abordada en la investigación permite inferir que el Business Inteligente es un conjunto de herramientas y metodologías para el análisis y 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%).
- 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.
- 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.

## Bibliografía

- Bharara, S. (2017). Application of learning analytics using clustering data Mining for Students' disposition analysis. *Education and Information Technologies*.
- Brown, M. (2011). *Learning Analytics: the coming third wave*. Retrieved from <https://net.educause.edu/ir/library/pdf/ELIB1101.pdf>
- businessintelligence.info. (2009, 10 29). *Business Intelligence*. Retrieved from Business Intelligence facil: <https://www.businessintelligence.info/dss/ejemplo-data-mining-panales-y-cerveza.html>
- Carbonell, M. C. (2006). *Cómo evitar el fracaso escolar en secundaria*. Madrid: Narcea, S.A. Ediciones.
- Cárdenas , C., Gonzáles , S., Nahuel , R., Herrera , P., & Ferrada , L. (2018). *Diseño de un modelo predictivo de pesquisa cardiovascular utilizando Árboles de Decisión: propensión de pacientes a presentar diabetes tipo 2, hipertensión arterial o dislipidemia: Estudio piloto, comuna de Quellón, Chiloé*. *Rev Chil Cardiol*. Retrieved from [https://www.scielo.cl/scielo.php?script=sci\\_arttext&pid=S0718-85602018000200126](https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-85602018000200126)
- Castrillon, O., Sarache, W., & Ruiz, S. (2020). Predicción de las principales variables que conllevan al abandono estudiantil por medio de técnicas de minería de datos. *Formacion Universitaria*.
- Castrillon, O., Sarache, W., & Ruiz, S. (2020). Predicción del rendimiento académico por medio de técnicas de Inteligencia artificial . *Formacion Universitaria*.
- CEPAL. (2002). *Panorama Social de America Latina 2001-2002*. Retrieved from Panorama Social de America Latina 2001-2002: [https://repositorio.cepal.org/bitstream/handle/11362/1213/1/S026433\\_es.pdf](https://repositorio.cepal.org/bitstream/handle/11362/1213/1/S026433_es.pdf)
- Chasipanta, M. (2017). Subsistema de refuerzo académico y su incidencia en el rendimiento académico de los estudiantes del Bachillerato General Unificado del Colegio Nacional

Conocoto, de la ciudad de Quito, año lectivo 2016-2017. *Proyecto de Investigación presentado como requisito previo a la obtención del Título de Licenciada en Ciencias de la Educación mención Comercio y Administración.*

Chaud, C. (2016). Comprensión lectora y rendimiento escolar en el área de comunicación en alumnos de primer año de secundaria en una institución educativa estatal y no estatal del Distrito de Surco. *TESIS PARA OPTAR EL GRADO DE MAESTRA EN PSICOLOGÍA*  
*TESIS PARA OPTAR EL GRADO DE MAESTRA EN PSICOLOGÍA.*

Cortina , V. (2015). *Aplicación de la metodología CRISP.DM a un proyecto de minería de datos en el entorno universitario.* Universidad Carlos III de Madrid. Retrieved from [https://e-archivo.uc3m.es/bitstream/handle/10016/22198/PFC\\_Victor\\_Galan\\_Cortina.pdf](https://e-archivo.uc3m.es/bitstream/handle/10016/22198/PFC_Victor_Galan_Cortina.pdf)

Data Warehouse. (2015, Junio 15). Retrieved from <https://www.areatecnologia.com/informatica/data-warehouse.html>

Educación, E. M. (2013). *Resultados de las pruebas Ser Estudiante.* Retrieved from <https://educación.gob.ec/noticias>

Escobar, H., Alcivar, M., & Marquez de la Plata, C. (2017). IMPLEMENTACIÓN DE MINERÍA DE DATOS EN LA GESTIÓN ACADÉMICA DE LAS INSTITUCIONES DE EDUCACIÓN SUPERIOR. *Didáctica y Educación.*

Espinosa, J. (2020). Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. *Ingeniería, investigación y tecnología.*

Ferrero, R. (2020). *QUÉ SON LOS ÁRBOLES DE DECISIÓN Y PARA QUÉ SIRVEN.* Retrieved from [maximaformacion.es: https://www.maximaformacion.es/blog-dat/que-son-los-arboles-de-decision-y-para-que-sirven/](https://www.maximaformacion.es/blog-dat/que-son-los-arboles-de-decision-y-para-que-sirven/)

Fuente , S. (2011). *Análisis correspondencias simples y múltiples.* UNAM. Retrieved from <https://www.fuenterrebollo.com/Economicas/ECONOMETRIA/REDUCIR-DIMENSION/CORRESPONDENCIAS/correspondencias.pdf>

Galán, V. (2015). Aplicacion de la Metodología CRISP-DM a un proyecto de Minería de datos en el entorno Universitario. *Universidad Carlos III.*

- García, D. (2015). CONSTRUCCIÓN DE UN MODELO PARA DETERMINAR EL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES BASADO EN LEARNING ANALYTICS (ANÁLISIS DEL APRENDIZAJE), MEDIANTE EL USO DE TÉCNICAS MULTIVARIANTES. *Tesis Doctoral*.
- García, J., & Acevedo, A. (2010). ANÁLISIS PARA PREDICCIÓN DE VENTAS UTILIZANDO MINERÍA DE DATOS EN ALMACENES DE VENTAS DE GRANDES SUPERFICIES. *TESIS DE GRADO*.
- Giraldo, J., & Jiménez, J. (2013). Caracterización del Proceso de Obtención de Conocimiento y Algunas Metodologías para Crear Proyectos de Minería de Datos. *Revista Latinoamericana de Ingeniería de Software*.
- Gondar Nores, J. (2004). Metodologías para la Realización de Proyectos de Data Mining.
- Gonzalez, L. (2018, Marzo 23). *Aprendizaje Supervisado: Random Forest Classification*. Retrieved from aprendeia.com: <https://aprendeia.com/aprendizaje-supervisado-random-forest-classification/>
- Heredero, C., Lozano, I., & Alcala, G. (1998). EL PROCESO DE IMPLANTACIÓN DEL DATA WAREHOUSE EN LA ORGANIZACIÓN: ANÁLISIS DE UN CASO. *Investigaciones Europeas de Dirección y Economía de la Empresa*.
- IBM. (2021, Julio 12). *Análisis ROC*. Retrieved from ibm.com: <https://www.ibm.com/docs/es/spss-statistics/beta?topic=features-roc-analysis>
- IBM. (2022, Septiembre 13). *Regresión Logística*. Retrieved from ibm.com: <https://www.ibm.com/docs/es/spss-statistics/saas?topic=regression-logistic>
- Ijaz, S. (2020). Educational Data Mining: A Review and Analysis of Student's Academic Performance. *International Conference on Intelligent Technologies and Applications*. Springer.
- Jimenez, A., & Alvarez, H. (2010). Minería de Datos en la Educación. *Universidad Carlos III de Madrid*.

Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). *The 2011 Horizon Report*. Retrieved from <http://net.educause.edu/ir/library/pdf/hr2011.pdf>

Lamas, H. A. (2015). Sobre el rendimiento escolar.

Lati, D. (2020, Abril 20). <http://t21.com.mx/opinion/columna-invitada/2020/04/27/datos-nuevo-petroleo>. Retrieved from <http://t21.com.mx/opinion/columna-invitada/2020/04/27/datos-nuevo-petroleo>: <http://t21.com.mx/opinion/columna-invitada/2020/04/27/datos-nuevo-petroleo>

Liberati, A. A. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *British Medical Journal*, 339.

Mancilla , G., Leal , P., Ortiz , A., & Vidal , C. (2020). *Factores asociados al éxito de los estudiantes en modalidad de aprendizaje en línea: un análisis en minería de datos*. Retrieved from [https://www.scielo.cl/scielo.php?script=sci\\_arttext&pid=S0718-50062020000600023&lng=pt&nrm=iso&tlng=pt#f3](https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-50062020000600023&lng=pt&nrm=iso&tlng=pt#f3)

Menacho, C. (2017). Predicción del rendimiento académico aplicando técnicas de minería de datos. *Anales Científicos*, 78(1), 226-33. doi:<http://dx.doi.org/10.21704/ac.v78i1.811>

Muñoz, G. F. (2018). *Análisis del rendimiento académico en los/as estudiantes de octavo año*. Quito.

Oliveira, J. (2021, Marzo 14). *¿Qué es un árbol de decisiones y cómo se hace?* Retrieved from crehana.com: <https://www.crehana.com/blog/estilo-vida/hacer-arbol-de-decisiones/>

Pallavi Amitava, B. (2016). A systematic review of factors linked to poor academic performance of disadvantaged students in science and maths in schools. *Cogent Education*.

PISA IN FOCUS. (2015). Retrieved from <https://www.oecd.org/pisa/pisa-2015-results-in-focus-ESP.pdf>

Riquelme, J., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias . *Inteligencia Artificial Vol. 10 No 29*.

- Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State-of-the-Art. *Transaction on Systems, Man, and Cybernetics*,.
- Romero, J. (2019). *Técnicas y algoritmos de Minería de Datos*. Retrieved from <https://jorgeromero.net/tecnicas-y-algoritmos-de-mineria-de-datos/>
- Ruíz, M. d. (2010). Fracaso escolar, Una realidad en nuestras aulas. *Revista de la Educación en Extremadura*, 154.
- Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Buckingham, S., & Ferguson, R. (2011). *Open Learning Analytics : an integrated & modularized platform Proposal to design , implement and evaluate an open platform to integrate heterogeneous learning analytics techniques*. Retrieved from <http://solaresearch.org/OpenLearningAnalytics.pdf>
- Solis, F., & Aguiar, R. (2017). Análisis del papel del involucramiento de la familia en la escuela secundaria y su repercusión en el rendimiento académico. *SciELO*.
- Sotaquirá, M. (2022, Septiembre 9). *La Matriz de Confusión*. Retrieved from [codificandobits.com: https://www.codificandobits.com/blog/matriz-de-confusion/](https://www.codificandobits.com/blog/matriz-de-confusion/)
- Torrecilla, F. J. (2007). *Investigación iberoamericana sobre eficacia escolar*.