

**Modelo de Gestión TI Apoyado en la Minería de Datos como Estrategia de Predicción de la
Deserción de los Estudiantes en el Nivel de Secundaria en la Institución Educativa de
Coyongal**

Francelina Romero Reyes

Asesor

Erika María Sandoval Valero

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI
Magister en Gestión de Tecnología y la Información

2023

Erika Maria Sandoval Valero

Nombre Director de Trabajo de Grado

Diana Marcela Vasquez Bravo

Jurado

Fernando José Díaz Martínez

Jurado

Dedicatoria

Es un esfuerzo y compromiso que he dispuesto con responsabilidad lograr este título de Magíster

y por esta razón lo dedico a:

Dios Todopoderoso y mi Redentor por su orientación y dotación de sabiduría a través de su

Espíritu Santo fortaleciéndome cada día en este proceso.

Mis tres hijas Keyssi, Catalina y Kenny como también a mi hijo Ángel por su amor y apoyo

incondicional.

Mi esposo Hector Manuel por su importante acompañamiento y dedicación.

Mis hermanas, hermanos y todos aquellos familiares y amigos que de uno u otra forma han

aportado un granito de arena en este proyecto.

Agradecimientos

Existe en la vida, no una sino, varias oportunidades que se aprovechan gracias a la intervención de entidades y personas que han facilitado el proceso de formación de este título de Magister, por

esta razón y muchas otras agradezco a:

Dios Padre Celestial por su grandiosa bendición y misericordia guardándome siempre en la salud y prosperidad en este proyecto.

La Universidad Nacional Abierta y a Distancia (UNAD), por su orientación y apoyo en este proceso de formación.

Mi asesora Erika María Sandoval Valero, por su dedicación y acompañamiento en el desarrollo de esta importante tesis de grado.

La Institución Educativa de Coyongal, su comunidad en general (Rector, Coordinador, estudiantes y padres de familia), por su valiosa colaboración y apoyo en el desarrollo de las actividades correspondientes a la elaboración de esta tesis de grado.

A todos aquellos que de una u otra forma me han colaborado incondicionalmente en este proceso.

Resumen

Este trabajo de grado propone el modelo basado en árboles de decisión con Algoritmo CART para predecir la deserción escolar como proyecto de investigación en la Maestría Gestión de Tecnología y de la Información. Este modelo de gestión se basa en la implementación del análisis que involucra actores representativos en el proceso escolar (estudiantes, docentes, coordinadores, psicorientadora, rector y secretaria), con el objetivo de encontrar factores influyentes en la deserción escolar de los estudiantes en el nivel de secundaria utilizando como variable relacionada con la matrícula cada año, rendimiento académico por grupo, las estrategias de permanencia, prácticas pedagógicas, la motivación académica y el sistema de evaluación. Se utilizará un sistema híbrido basado en herramientas de análisis de infraestructura de negocios, empresarial y de solución TI, apoyadas en técnicas para minería de datos desde una visión metodológica que determine el nivel de deserción en los últimos 6 años (2015 a 2020).

Palabras Claves: Minería de datos, estrategias, deserción, predecir, permanencia, algoritmo, arboles de decisión.

Abstract

This article proposes the model based on decision trees with CART algorithm to predict school dropout as a research project in the master's degree in information and technology Management. This management model is based on the implementation of the analysis that involves representative actors in the school process (students, teachers, coordinators, counselor, rector and secretary), with the aim of finding influential factors in the school dropout of students at the level secondary school using as a variable related to permanence strategies, pedagogical practices, academic motivation and the evaluation system. A hybrid system will be used based on business infrastructure analysis tools, business, and IT solution, supported by techniques for data mining from a methodological vision that determines the level of attrition in the last 3 years (2018, 2019 and 2020).

Keywords: Data mining, strategies, attrition, predict, permanence, algorithm, decision trees.

Tabla de Contenido

Introducción	13
Justificación	14
Caracterización problema de la investigación	16
Título.....	16
Problema	16
Formulación	17
Descripción de la problemática.....	17
Contextualización de la problemática.....	19
Antecedentes	25
Marco teórico	32
Marco conceptual.....	36
Deserción estudiantil.....	36
La deserción en el nivel de secundaria.	37
Gestión TI.	38
Gestión de proyecto TI.....	42
Minería de datos.....	44
La minería de datos como estrategia de predicción de la deserción escolar.....	45
El Algoritmo CART para predecir el nivel de deserción escolar	47
Diseño Metodológico.....	48
Caracterización de la población y muestra	52
Selección de las técnicas a utilizar en el desarrollo del trabajo	53

El método de clasificación basada en árboles de decisión en la predicción de la deserción estudiantil.....	54
Técnicas de minería de datos utilizadas en estudios sobre deserción estudiantil.	58
Procesamiento y flujo de datos para la predicción.....	61
Desarrollo de los objetivos propuestos en esta investigación.....	63
Fase 1. Realización del análisis diagnóstico.	63
Fase 2. Aplicación de la técnica algoritmos CART y árboles de decisiones.	80
Fase 3. Resultados predictivos del modelo de gestión TI.	96
Propuesta del modelo de Gestión TI apoyado en la Minería de datos para predecir la deserción en el nivel de secundaria.....	105
Evaluación de la predicción del nivel de deserción en los años 2021 -2025.....	115
Conclusiones.....	120
Recomendaciones.....	122
Referencias.....	123

Lista de Tablas

Tabla 1 <i>Fuentes bibliograficas que fundamentan la investigación</i>	26
Tabla 2 <i>Población estudiantil</i>	53
Tabla 3 <i>Análisis de matricula a partir de la minería de datos de los años 2014 a 2020</i>	59
Tabla 4 <i>VN: Valoración Nacional</i>	67
Tabla 5 <i>Cuadro de matriculas año 2014</i>	69
Tabla 6 <i>Cuadro de matriculas año 2015</i>	71
Tabla 7 <i>Cuadro de matriculas año 2016</i>	72
Tabla 8 <i>Cuadro de matriculas año 2017</i>	73
Tabla 9 <i>Cuadro de matriculas año 2018</i>	74
Tabla 10 <i>Cuadro de matriculas año 2019</i>	75
Tabla 11 <i>Cuadro de matriculas año 2020</i>	88
Tabla 12 <i>Los indicadores con relación a la variable de la eficiencia interna</i>	90
Tabla 13 <i>Resultados de la Institución Educativa de las Pruebas Icfes Saber 11</i>	91
Tabla 14 <i>Transformación de variables con la minería de datos</i>	93
Tabla 15 <i>Hallar el promedio porcentual en Rstudio</i>	97
Tabla 16 <i>Comprobación de los atributos aprobados, desertados, retirados y reprobados en la variable de rendimiento académico</i>	98

Lista de Figuras

Figura 1 <i>Estructura organizacional administrativa</i>	20
Figura 2 <i>Mapa de procesos estructura organizacional</i>	21
Figura 3 <i>Perspectivas claves que determinan la eficiencia educativa</i>	24
Figura 4 <i>Modelo de toma de decisiones</i>	39
Figura 5 <i>Diseño metodológico de la investigación</i>	49
Figura 6. <i>Análisis de matrícula a partir de la minería de datos por años</i>	55
Figura 7 <i>Árbol de decisiones de la variable matrícula</i>	56
Figura 8 <i>Árbol de decisiones de la variable rendimiento académico</i>	57
Figura 9 <i>Modelo de técnica de la minería de datos</i>	58
Figura 10 <i>El flujo de datos consistió en la metodología basado en 5 etapas</i>	61
Figura 11 <i>Mineria de datos a partir de la ficha de matrícula</i>	62
Figura 12. <i>Matriz de relación de asignatura aprobadas y no aprobadas</i>	63
Figura 13 <i>Esquema de la planeación de la selección de los datos</i>	64
Figura 14 <i>Análisis predictivo de la deserción escolar mediante patrones y atributos</i>	65
Figura 15 <i>Resumen histórico de la matrícula en comparación con años anteriores</i>	76
Figura 16 <i>Clasificación de la variable matrícula por años</i>	77
Figura 17 <i>Análisis comparativo de los resultados de las Pruebas Saber</i>	79
Figura 18 <i>Lista de notas por período como estrategia de minería de datos</i>	81
Figura 19 <i>Planillas de notas en físico en los años de 2014 a 2020</i>	82
Figura 20 <i>Planillas de asistencia en físico y actualizada en Punto Edu</i>	83
Figura 21 <i>Modelo de gestión PMBOK</i>	84
Figura 22 <i>Flujo de procesos de gestión PMBOK</i>	85

	11
Figura 23 <i>Archivo de Excel como base de datos de eficiencia interna por año</i>	87
Figura 24 <i>Comparación de los puntajes de cada competencia evaluada pruebas Icfes</i>	92
Figura 25 <i>Creación de la librería al instalar Rstudio</i>	94
Figura 26 <i>Ejecución de la librería en Rstudio</i>	94
Figura 27 <i>Modelo de evaluación mediante las propiedades plot, print y main en Rstudio</i>	95
Figura 28 <i>Modelo de predicción hallando el promedio de estudiantes matriculados</i>	100
Figura 29 <i>Modelo de predicción hallando el promedio de estudiantes retirados</i>	101
Figura 30 <i>Proceso histórico del estudio predictivo con árbol de clasificación y regresión</i>	102
Figura 31 <i>Árbol de decisión en el atributo de retirados</i>	103
Figura 32 <i>Gráfica del modelo de predicción del atributo aprobado</i>	103
Figura 33 <i>Atributo de estimación del promedio de reprobados por año</i>	104
Figura 34 <i>Representación del árbol de decisiones detectando riesgos de deserción</i>	106
Figura 35 <i>Comparación de los atributos en Rstudio del grado sexto</i>	108
Figura 36 <i>Gráfica que predice el porcentaje de deserción en el grado sexto</i>	109
Figura 37 <i>Comparación de los atributos de matrícula, aprobados y reprobados de séptimo</i>	109
Figura 38 <i>Gráfica que predice el porcentaje de deserción del grado séptimo</i>	111
Figura 39 <i>Comparación de los atributos de matrícula, aprobación y reprobación de octavo</i>	112
Figura 40 <i>Gráfica que predice el porcentaje de deserción de octavo</i>	113
Figura 41 <i>Comparación de los atributos de matrícula, aprobados y reprobados de noveno</i>	114
Figura 42 <i>Gráfica que predice el porcentaje de deserción de noveno</i>	114
Figura 43 <i>Ejecución de la serie temporal y observación para la predicción en Rstudio</i>	116
Figura 44 <i>Los paquetes instalados permiten la clasificación de categorías predictivas</i>	117
Figura 45 <i>Organización de la información históricamente</i>	117

Figura 46 *Hallazgos del modelo predictivo con arbol de decisiones en histograma.*

Introducción

El uso de la Minería de Datos o Data Mining, como soporte a las decisiones en las actividades de negocio, requiere mucho más que la aplicación de sofisticadas técnicas como redes neuronales o árboles de decisión sobre las tablas de datos. En este sentido, la minería de datos muestra comportamientos de predicción a través de la clasificación de datos tomados de fuentes primarias, como uno de los pasos del proceso de descubrimiento de conocimiento en base de datos (KDD) y por otro lado como un proceso que consta de diferentes fases, en las cuales se utilizan como apoyo, técnicas relacionadas con la estadística, el reconocimiento de patrones y algoritmos de aprendizaje, entre otras (Castañeda y Rodríguez, 2003).

Con el fin de hallar resultados válidos se plantea un modelo de gestión TI mediante la construcción con base en la ejecución de un análisis diagnóstico que haga visible los requerimientos para el desarrollo de un proyecto de data mining para que provea los mecanismos para la predicción de deserción estudiantil. Se realiza un estudio estadísticos de documentos extraídos de plataformas confiables y efectivos permitiendo la minería de datos como estrategia para predecir la deserción.

Justificación

Hacer Data Mining sobre datos que se actualizan a menudo es un desafío. Por consiguiente, hay varios problemas que necesitan ser investigados extensamente, antes de que se pueda llevar a cabo lo que se conoce como Data Mining en tiempo real. El uso de indicadores en pro de medir la bondad, aplicabilidad, la relevancia y la novedad, de los resultados de la Minería de Datos, pueden resultar en ocasiones muy subjetivos, pero los negocios necesitan contar con algún medio que les permita medir el interés y el impacto del conocimiento que se puede obtener al aplicar Minería de Datos. De igual manera, la intervención y experiencia del tomador de decisiones es relevante para establecer algunas medidas y poder calcular los indicadores antes mencionados (Mercano y Talavera, 2007).

Es importante destacar que la disciplina denominada Minería de Datos estudia métodos y algoritmos que permiten la extracción automática de información sintetizada que permite caracterizar las relaciones escondidas en la gran cantidad de datos; también se pretende que la información obtenida posea capacidad predictiva, facilitando el análisis de los datos de forma eficiente. Bajo la denominación de "minería de datos" se han agrupado recientemente diversas técnicas estadísticas y del aprendizaje automático (Inteligencia Artificial) enfocadas, principalmente, a la visualización, análisis, y modelización de información de bases de datos masivas (Beltrán, s.n.f.).

Objetivos

Objetivo General

Desarrollar un modelo de gestión TIC apoyado en la minería de datos para predecir la deserción de los estudiantes en el nivel de secundaria en la Institución Educativa de Coyongal

Objetivos Específicos

Realizar un análisis diagnóstico para identificar el nivel de deserción utilizando los resultados de matrículas del I SIMAT, el informe comparativo de las pruebas ICSE y los resultados de las pruebas saber entre los años 2015 a 2020 como minería de datos.

Aplicar las técnicas de predicción algoritmos CART y árboles de decisiones como minería de datos para predecir la deserción clasificando las variables: el rendimiento académico en PUNTO EDU, los resultados de las Pruebas ICFES y Saber 11, y la autoevaluación institucional en los últimos 6 años (2014 a 2020) en los grupos del nivel de secundaria.

Proponer Modelo de Gestión TI para predecir el nivel de porcentaje de deserción en los años 2021 – 2025 a partir de los resultados arrojados en los análisis estadísticos de los años 2014 a 2020.

Caracterización Problema de la Investigación

Título

Modelo de gestión tic apoyado en la minería de datos como estrategia de predicción de la deserción de los estudiantes en el nivel de secundaria en la institución educativa de coyongal

Problema

En este apartado se plantea el problema de forma descriptiva a partir de la formulación, la cual conlleva a la investigación el análisis de minería de datos relacionados con la deserción escolar en la Institución Educativa de Coyongal y de esta manera proponer un modelo de gestión para predecir la deserción escolar.

La investigación parte de la problemática denominada “deserción de los estudiantes en el nivel de secundaria”. Este fenómeno afecta a la población estudiantil de la Institución Educativa de Coyongal, notándose en el alto porcentaje de estudiantes que abandonan su proceso académico. (Ángel y Facundo, 2009), dicen que la deserción implica un desperdicio de recursos y esfuerzos, tanto públicos como privados, que deja muchos sinsabores y frustraciones en los afectados. De igual forma, (Rodríguez et al, 2016), afirma que los costos sociales de la deserción estudiantil son extremadamente altos, lo cual repercute en un incremento de las tasas de criminalidad y decremento de la tasa de crecimiento económico y muchos otros índices como aumento de la inequidad o la brecha entre ricos y pobres.

Según Hernández, en el año 2017 el municipio de Magangué tuvo una matriculas atendida oficial de 27150 alumnos, en el 2018 el reporte fue de 27340, es decir aumentó la matricula atendida, lo que es factor para la distribución de los recursos del Sistema General de Participaciones para la educación del componente de calidad –matricula oficial. Pero los altos indicadores de deserción, que fue para el 2018 de 820 alumnos, y de reprobación, que alcanzó

2400 alumno, hacen que esos cupos tengan que ser reemplazos para la próxima vigencia de matrícula, en este caso 2019, porque la mayoría de las veces estos jóvenes o dejan el sistema educativo o pasan a los modelos flexibles de educación. Para el funcionario, lo altos indicadores de deserción y reprobación podría estar pasando por la mala aplicación de los docentes del decreto 1290 de 2009, que reglamentó precisamente la evaluación del aprendizaje y promoción de los estudiantes de los niveles de educación básica y media. Es decir, la entidad territorial de Magangué llegó a un porcentaje de 8.7% de reprobados, alta con la media nacional que alcanzó en el 2017 4.4% (Secretaria de Educacion de Magangué, 2023).

Formulación

¿Cómo desarrollar un modelo de gestión TIC apoyado en la minería de datos para predecir la deserción de los estudiantes en el nivel de secundaria en la Institución Educativa de Coyongal?

Descripción de la Problemática

La calidad educativa se ha convertido en la iniciativa de las instituciones para contribuir el seguimiento de un sistema de evaluación a través de criterios referenciales estandarizados a nivel nacional e internacional. De este modo las políticas educativas orientan los niveles de aprendizaje de los estudiantes, la gestión educativa y las prácticas pedagógicas en el desarrollo de competencias y habilidades para la integración social (MEN, 2010)

El sistema colombiano ha considerado asistir la problemática que en particular se presenta en las instituciones educativas, en especial en las zonas rurales, debido a los diferentes factores que inciden en la deserción escolar, el abandono de los estudios de los estudiantes dimensiona la determinación del diseño y creación de estrategias que motiven a los estudiantes que presentan debilidades en su proceso formativo para que se apoyen en la importancia de la

formación educativa en su proyecto de vida. La ministra de Educación María Victoria Angulo (2020), afirma que la cifra de deserción es del 2,7%, muy similar a los años anteriores. Cabe resaltar que esta cifra causó alarma por cuanto se estaba iniciando la problemática de la pandemia.

En este sentido, es importante definir desde el ángulo individual se apunten factores principalmente familiares, es por ello, que se encuentran dudas en el momento de reflexionar si la deserción se decide en forma voluntaria o involuntaria, consciente o inconsciente o también puede ser por cambios de estrategias escolares, sociales o naturales. Para encontrar la claridad a estas dudas es necesario formular interrogantes ¿cómo se puede analizar el fenómeno de la deserción escolar?

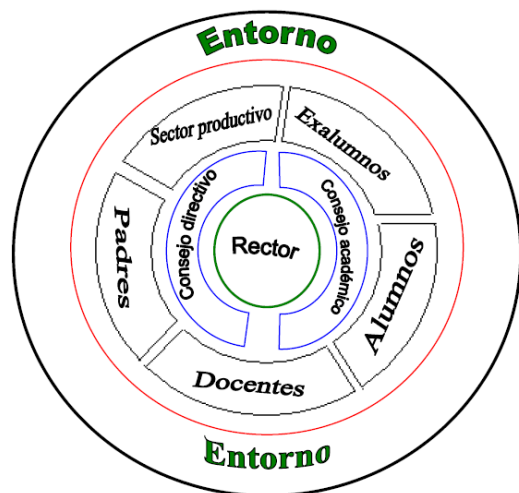
Reconociendo que la deserción escolar es la contundente y sistemática evidencia que ratifica que el fracaso escolar afecta de manera principal y mayoritariamente a quienes pertenecen a los sectores más pobres en las distintas sociedades: mientras más pobres, vulnerables y excluidos son los estudiantes, mayores son sus probabilidades de no aprender lo necesario, de no alcanzar buenos desempeños, de reprobado grado, de dejar de asistir a clases, o finalmente desertar definitivamente del sistema escolar (Roman, 2013).

Por lo tanto, la deserción escolar no es un problema ajeno en la Institución Educativa de Coyongal, puesto al alto porcentaje reflejado cada año, situación que preocupa a las entidades receptoras como son la Secretaría de Educación Municipal, Departamental y el Ministerio de Educación Nacional. Es decir, el alto nivel de deserción estudiantil en el nivel de secundaria en las instituciones educativas oficiales es uno de los problemas principales que enfrentan. Se estima, que la tasa de deserción anual está en el orden del 41,9%, en el año 2015, el 40,5% en el año 2016, en 49,6%, en el año 2017 y el 47,2% en el año 2018 (ISCE) según el informe emitido

por el Ministerio de Educación Nacional (MEN). En un periodo de tiempo normal, solamente logran culminar la media, un promedio del 41% de los que se matriculan en la Institución (SIMAT). Para ello, es necesario acudir a teorías específicas relacionadas con la problemática.

Contextualización de la Problemática

La Institución Educativa de Coyongal, para el logro de la Misión y Visión que se ha propuesto tiene una estructura organizacional que se puede apreciar en el siguiente organigrama.

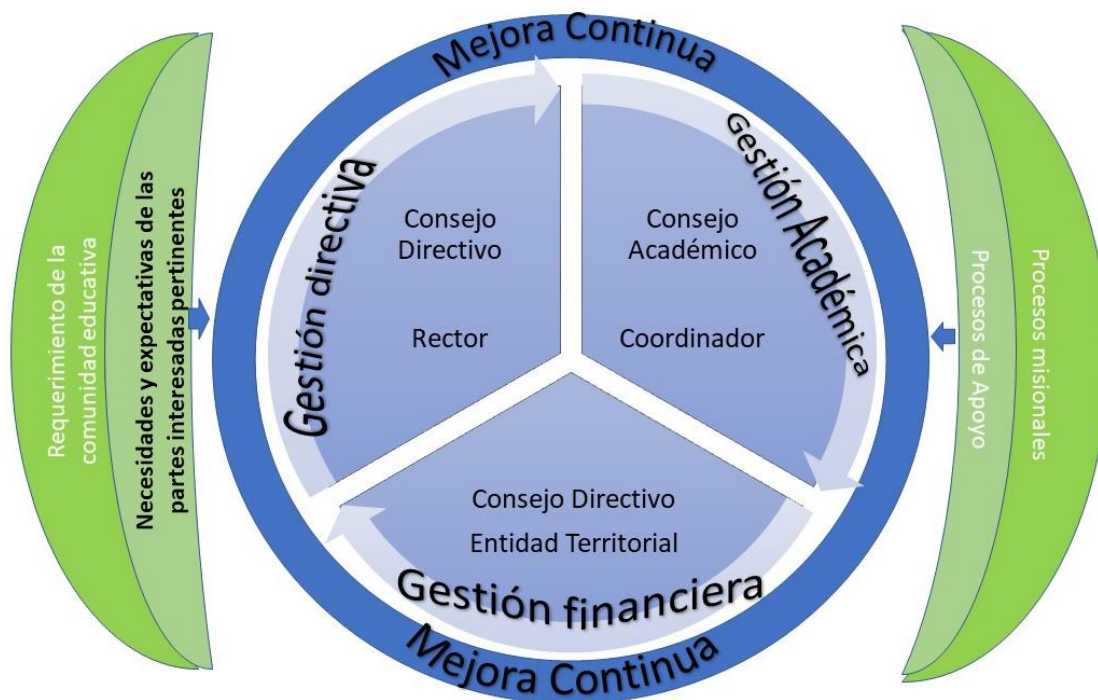
Figura 1*Estructura organizacional administrativa*

Fuente. Proyecto Educativo Institucional (PEI), 2019.

El mapa de proceso incluye las cuatro gestiones que articulan la estructura organizacional administrativa para satisfacer las necesidades de la población estudiantil: gestión directiva, gestión académica, gestión financiera y gestión comunitaria incluida en los procesos de apoyo. Esta área da soporte al trabajo institucional, tiene a su cargo todos los procesos de apoyo a la gestión académica, la administración de la planta física, los recursos y los servicios, el manejo del talento humano y el apoyo financiero y contable.

Figura 2

Mapa de procesos estructura organizacional.



Fuente. Propia del investigador

Contextualizando la Institución Educativa de Coyongal, por ser una entidad de carácter oficial funciona con base a las políticas del Ministerio de Educación Nacional emanado por la Ley 115 de 1994, presentando un organigrama con funcionalidad directa a la población estudiantil. De este modo, ofrece la administración cuyas funciones las cumple el rector, la gestión académica con las funciones de los coordinadores y los procesos misionales fundamentados en el diseño de proceso curricular en la prestación del servicio educativo.

De hecho, en la búsqueda de la mejora continua, el sistema educativo de Colombia ha desarrollado el Programas Todos Aprender con el fin de determinar la calidad y responsabilidad de la implementación de los procesos formativos de los estudiantes, con el fin de detectar los

factores que inciden en la deserción de los estudiantes y así diseñar estrategias que garanticen la permanencia y culminación de su proceso de formación en las instituciones.

En este contexto, se caracteriza el problema que generaliza la preocupación en la Institución Educativa de Coyongal, teniendo en cuenta que la deserción escolar en el nivel de secundaria consiste en el abandono del proceso académico por parte de los estudiantes, provocado por diferentes factores internos o externos, que se generan tanto al interior del sistema educativo como en contextos de tipo social, familiar, individual y del entorno. Es decir, el número de alumnos que, inscritos en un grado cualquiera de un año escolar determinado, no continúa estudios el año siguiente, sea por abandonar la escuela antes de haber finalizado el curso, o por abandonarla después de finalizado, habiendo sido los alumnos aprobado o no aprobado (Oviedo y Diaz, 1982 en Lugo, 2013, p. 292).

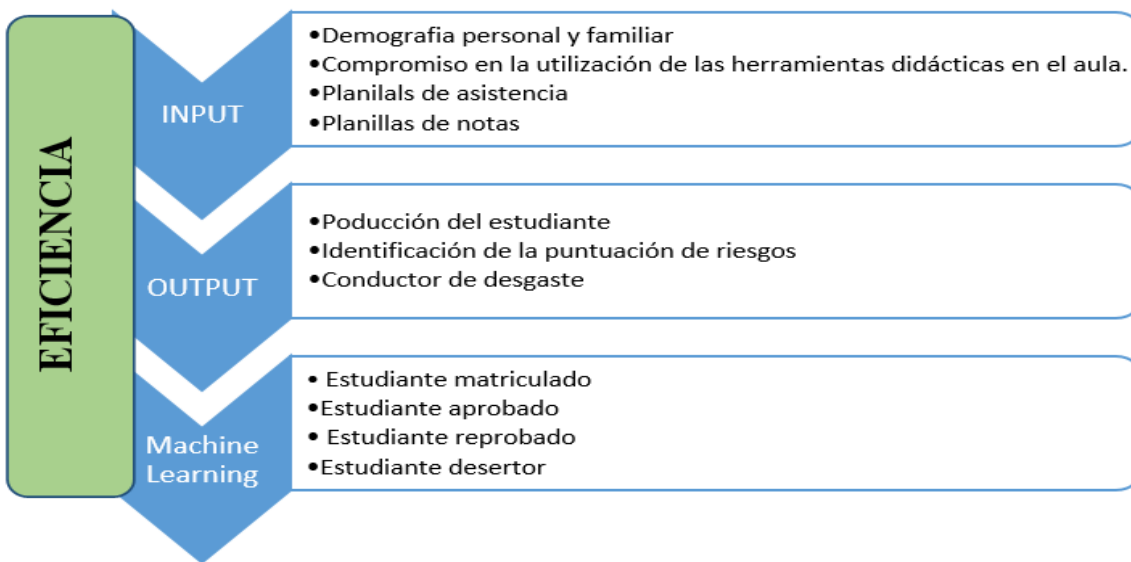
Por lo tanto, la deserción está relacionada con los procesos educativos institucionales que conlleva a determinar múltiples perspectivas en la conceptualización de la desaprobación, en cuanto al consenso a que la institución Educativa de Coyongal articula e implementa los procesos de evaluación y la innovación. Se identifica un sinnúmero de bases de datos que se utilizarán como estrategias de minería de datos para predecir la deserción en el nivel de secundaria. También permitió analizar y reflexionar acerca de los tres factores claves del sistema educativo mediante la autoevaluación institucional como se detalla a continuación:

Teniendo en cuenta, que la deserción es un problema de política educativa y está relacionada con procesos de calidad. Se determinaron múltiples perspectivas en la conceptualización de la calidad, en cuanto al consenso a que ella se articula con la evaluación y la innovación, y también en la necesidad de reflexionar acerca de los tres factores claves del sistema educativo: el primero, la eficiencia que determina el INPUT denominado como el aporte

del estudiante aspectos como son: demografía familiar y personal; el grado de compromiso en la utilización de las herramientas didácticas en el aula, la asistencia y las notas. El segundo, se establece el OUTPUT denominado como la producción del estudiante identificando la puntuación de riesgos, el conductor de desgaste. Y el tercero Machine Learning denominado como el modelo de predicción o identificación del estudiante modelo desertor.

Figura 3

Perspectivas claves que determinan la eficiencia educativa



Fuente. Propia del investigador

El fenómeno de la deserción incrementa cada año en las instituciones oficiales Colombia. En el año 2018 como se observa en el análisis documental realizado para identificar el problema se disminuyó un gran porcentaje de la deserción escolar por cuanto en el componente eficiencia muestra la tasa de aprobación en las cuatro versiones del ISCE (2018). En este contexto, el proyecto encamina el estudio y análisis de la deserción escolar en los niveles de secundaria y media mediante la prueba inicial de comportamiento de datos y la aplicación de algoritmos en la plataforma del Simat con el objetivo de predecir el porcentaje de deserción que posee cada estudiante matriculado en los tres primeros periodos escolares determinando un sistema de inteligencia artificial que proporciona Azure Machine Learning Stue generando un árbol de decisión. De esta manera, se determina un modelo de Gestión TIC mediante herramientas estadísticas que permitan analizar y luego comprender las posibles causas de la deserción escolar.

Antecedentes

Para este trabajo de grado para optar e título de magister en Gestión de las TI, se realizó una revisión bibliográfica, con el objeto de avocar conceptos y teorías acerca del uso de minería de datos, predicción de la deserción en los niveles de secundaria y modelo de gestión TI, en los procesos educativos, mediante la consulta de material bibliográfico en diferentes tipos de documentos, con la finalidad de proporcionar sustento científico a las temáticas que se abordan en el desarrollo de este proyecto. Para ello, se clasifican referentes pertinentes con el problema: se consulta bibliografía de fuentes primarias de autores que han abordado la problemática desde los contextos internacionales y nacionales.

Tabla 1*Fuentes bibliográficas que fundamentan la investigación*

Fuentes bibliográficas				
Nombre	Tipo	Año	Aportes	Tipo de investigación
Internacionales				
Ordoñez Briceño Karla	Tesis de grado	2009	Contribuir con la solución del problema de la deserción estudiantil en la Universidad Nacional Daniel Alcides Carrión se plantea realizar un estudio comparativo de técnicas de minería de datos para predecir la deserción estudiantil universitaria en la región de Pasco.	El diseño empleado fue descriptivo cuasiexperimental, de tipo cualitativo y cuantitativo, cuyo análisis mostró una diferencia significativa entre la deserción estudiantil y promedio de notas, a un nivel de confianza del 95% y nivel de significancia del 5%
Espinola Hoffmann Viola y Claro	Artículo	2009	Su investigación destaca la magnitud del problema de la deserción juvenil y propone alternativas de política que	Su metodología trata de políticas basadas en la descentralización de decisiones y

Stuardo			podrían conducir a un	responsabilidades a las
Juan			aumento gradual del	administraciones
Pablo.			porcentaje de estudiantes que	locales y en la acción
			completan su escolaridad y	coordinada de todos los
			así, enfrenten mejores	niveles de la
			opciones en su futuro la	administración
			investigación ha demostrado	educativa, incluyendo a
			que cuando los padres se	las familias y a la
			involucran en las actividades	comunidad.
			escolares de sus hijos, mejoran	
			la asistencia, la conducta, el	
			desempeño académico y la	
			motivación de los estudiantes	
			por asistir al liceo.	
Valero	Articulo	2003	Las técnicas de minería de	Para esta investigación
Orea			datos permiten obtener	se utilizaron
Sergio,			conocimiento oculto en	específicamente dos
Salvador			grandes cantidades de datos	algoritmos: el
Vargas			con información valiosa que,	algoritmo de árboles de
Alejandro			al explotarse, ofrece ventajas	clasificación C4.5 y el
y García			competitivas a las	algoritmo de los k
Alonso			organizaciones con la técnica	vecinos más cercanos
Marcela			de Minería de Datos con la	en la Universidad

Herramienta Weka un Repositorio abierto de la Universidad de Cantabria en México.	Tecnológica de Izúcar de Matamoros, tomando como base de análisis los datos del estudio socioeconómico del EXANI-II, elaborado por el CENEVAL, mismo que se aplica desde el 2003 en nuestra institución.
--	--

Nacionales

Aristizábal Fúquene, Jorge Alexánder	Tesis de grado	2016	Su tesis trata con el diseño de un modelo metodológico para minería de datos en un aula de educación presencial y el análisis de las reflexiones de los actores (profesores, administrativos docentes, estudiantes e investigador) en entorno al acto pedagógico luego de su implementación	Se realizó un estudio comparativo de corte cualitativo y cuantitativo contando con la participación en el estudio de 10 estudiantes de 32 matriculados del grado 10 como muestra probabilística mediante
---	-------------------	------	--	---

encuesta tipo
cuestionario y registro
documental de
evaluación de
aprendizaje en grupos
focales. Se encontró
que la utilización de
esta técnica de minería
de datos complementa
el logro del objetivo
general de la
investigación puesto
que el modelo
metodológico para
minería de datos
incluye la analítica de
visualizaciones como
medio para hacer más
accesible la minería de
datos a la comunidad
en general.

Escobar Rubio Jaquelinne	Tesis de grado	2015	Esta tesis se enfoca en indagar los factores que permitan descubrir el porqué de la deserción escolar, es decir, del abandono de algunos estudiantes en los planteles educativos, para el caso de la investigación en la Institución Educativa Normal Superior María Inmaculada.	Se aplicó un estudio de tipo descriptivo, apoyado en el método cuantitativo con padres o acudientes de estudiantes desertores 166 y 66 estudiantes. se encontraron factores que conllevan a la deserción como: la situación económica, modelo familiar nuclear y mononuclear, entre otros.
Varón Martínez Fernelly	Tesis de grado	2017	identificar una tendencia de retiro escolar con diversidad de variables en un periodo de tiempo específico dentro del marco temporal seleccionado	Realizó un estudio exploratorio, utilizando el análisis cuantitativo y cualitativo de la comunidad educativa relacionada con este fenómeno educativo y mostrando un especial interés por el ser.

Camargo García Aníbal José	Trabajo de investigación	2020	Este proyecto de investigación es crear un modelo para la predicción de la deserción de estudiantes de pregrado en la Universidad de la Costa - CUC, a partir del análisis de diferentes factores socioeconómicos y académicos.	El estudio requirió de la ejecución de una serie de fases: caracterización, experimentación, desarrollo y evaluación.
-------------------------------------	-----------------------------	------	---	---

Fuente. Propia del investigador

Marco Teórico

En esta investigación se determina la deserción escolar como aquel estudiante que matriculó su año escolar en la institución con su respectivo proceso en el SIMAT y por causas desconocidas antes de finalizar el ciclo escolar deja de presentar sus deberes escolares y de asistir a las clases. Para esto, es necesario citar a los siguientes autores:

Desde el punto de vista pedagógico la deserción enfatiza la predicción a través de aspectos tales como problemas familiares, desmotivación, desconcentración, afectación psicológica y otros. Por lo tanto, (Pérez y Merino, 2008) afirman que la familia es la que ejerce un peso clave a la hora de que un niño en edad escolar lleve a cabo la citada deserción. Y es que el hecho de que este deje de asistir a clase está en relación, en muchas ocasiones, con las características de su entorno. Así, cuando el mismo tiene escasos recursos económicos, no tiene una residencia fija y además los padres no tienen formación académica, se dan los factores para que el pequeño no acuda al colegio.

De acuerdo a lo expuesto anteriormente la deserción se produce en factores como la baja identidad familiar, es decir los padres no muestran interés en la formación de su hijo y los hermanos no han continuado sus estudios. De este modo, se suman otros problemas para el aprendizaje, como son responsabilidades en la casa con la familia, enfermedades y la repitencia que incide en los primeros años de la secundaria.

Asimismo, Ethingon (1991), incorpora la teoría de las “conductas de logro”, avocando atributos como la perseverancia, la elección y el desempeño, principalmente esbozando que el nivel de aspiraciones y las expectativas de éxito del estudiante son elementos fundamentales para explicar la deserción. De hecho, el autor afirma que la motivación escolar depende del entorno familiar, que está estrechamente relacionado con el apoyo que le brindar con la integración

familiar y referencia su rendimiento académico. De este modo el estudiante puede evaluar su éxito o su fracaso. También muestra el modelo de integración del estudiante y explica que la mayor presencia está en el ambiente académico y social que además contribuye a un compromiso institucional más fuerte lo afecta la decisión del alumno de permanecer o desertar (Spady, 1970; Tinto, 1975; Cabrera, et al. 1993).

De esta manera, Moreno (2013) avoca al Plan Nacional de Desarrollo en el contexto de la Política Educativa “Educación de Calidad”, como el Plan de Desarrollo Municipal y el Plan de Ordenamiento Territorial, tienen como propósito fundamental disminuir las brechas en la educación en todos sus niveles, anotando que el término brechas educativas hace referencia a situaciones de inequidad en las condiciones y posibilidades de acceso al aprendizaje dentro de los procesos educativos. Se debe propender por el desarrollo de estrategias que brinden oportunidades de acceso, calidad y permanencia de la población escolar en el sistema educativo; es decir, que las políticas educativas, como lo establece la misma Constitución Nacional, la Ley General de Educación y demás normas concordantes, supuestamente están orientadas a asegurar las condiciones de permanencia y calidad en el sistema educativo de la población estudiantil hasta finalizar sus estudios.

La misma Constitución Política de 1991 en su Artículo 67 establece: “El Estado, la sociedad y la familia son responsables de la educación, que será obligatoria entre los cinco y los quince años de edad y que comprenderá como mínimo, un año de preescolar y nueve de educación básica”, quiere decir esto que si la educación es fundamental en una sociedad, es primordial que los niños que ingresen al sistema escolar permanezcan en él, y por ende el Gobierno a través de las políticas públicas debe evitar la deserción escolar.

Teniendo en cuenta lo anterior, el MEN ha medido la deserción a través de la tasa de deserción intra-anual, es decir, el porcentaje de estudiantes que dejan de estudiar durante el transcurso del año académico, en comparación con los inicialmente matriculados. De esta manera, los resultados de la Encuesta Nacional de Deserción (ENDE), evidencian la complejidad y heterogeneidad del fenómeno. No es simplemente el problema económico sino variables de diversas dimensiones: familia, Institución Educativa, contexto social e individuo.

Por consiguiente, la deserción ha mostrado características que permiten predecir y conocer el porcentaje de estudiantes con probabilidad de desertar. (Verdejo, 2021) afirma que el sistema de predicción se realiza mediante métodos de oferta y filtrado del contenido digital de plataformas destinadas a la transmisión ofrece el trabajo de algoritmos. De este modo, funciona con base a una fórmula prediseñada por el investigador utilizando patrones con registros almacenados en un determinado tiempo o periodo. Algo si es cierto, que la inteligencia artificial no ayuda a predecir el futuro, pero, si ayuda a analizar mediante plataformas para encontrar la problemática, luego reflexionar y tomar decisiones para el futuro. De igual manera, (Tinto, 1975). Plantea que el abandono de los estudios puede tener carácter tanto transitorio y/o permanente, de acuerdo con el tiempo que dure el alejamiento del estudiante de las instituciones de educación. La deserción es transitoria, cuando el alumno decide revertir la situación de abandono, en la medida de que éste no tenga impedimento para retomar sus estudios, y, se torna permanente cuando existen causas concretas que le impiden volver a cursar o retomar sus estudios (Avila, 2021, p. 18).

Por lo anterior, teóricamente se encontró que para esbozar la probabilidad de desertores en el nivel de secundaria se pudo hacer a través de Árboles de decisión lo cual permitió obtener de forma visual patrones de decisión bajo los cuales operan una determinada población, a partir

de datos históricos almacenados. (Ávila, 2021), afirma que su principal ventaja es la facilidad de interpretación. Esta técnica ha tenido aplicabilidad en diversas disciplinas del conocimiento y es muy útil cuando se tiene una propiedad conocida para un conjunto de elementos, pero, no se conoce esa misma propiedad en un elemento concreto. La clasificación basada en árboles de decisión es tal vez una de las técnicas que más se ha popularizado en Data Mining o minería de datos, los árboles de decisiones se usan en investigación de operaciones para describir modelos jerárquicos de decisión y su consecuencia, en Data Mining su uso tiene que ver con modelos analíticos encaminados a realizar predicciones. (Ramírez & Grandón, 2018) “Esta técnica no paramétrica clasifica una población en un modelo de segmentos de tipo ramas que construyen un árbol invertido, y luego este modelo se utiliza para predecir una variable objetivo”. p3.

Marco Conceptual

En virtud a la realidad reflejada en la Institución Educativa de Coyongal, se optó a la persuasión de árboles de decisión como técnicas de clasificación de datos permitiendo organizar un cuadro comparativo que precise la predicción de la deserción teniendo en cuenta variables académicas y cada uno de los factores que ayudaron a encontrar el nivel de porcentaje de deserción escolar. De esta manera se conceptualizan los procesos de análisis mediante las técnicas de minería de datos diseñando un modelo de predicción del estado académico de los estudiantes en los tres primeros periodos escolar por año.

Deserción Estudiantil.

En esta investigación se determina la deserción escolar como aquel estudiante que matriculó su año escolar en la institución con su respectivo proceso en el SIMAT y por causas desconocidas antes de finalizar el ciclo escolar deja de presentar sus deberes escolares y de asistir a las clases. Para esto, es necesario citar a los siguientes autores:

Bachman, Green, & Wirtanen (1971), definen que la deserción escolar se ocasiona por aquellos estudiantes que interrumpen su asistencia a la escuela por varias semanas por diferentes razones, exceptuando aquellos por enfermedad.

Morrow (1985), define a la deserción cuando un estudiante el cual estuvo inscrito en la escuela, deja la misma por un largo periodo de tiempo y no se inscribió en otro colegio. Donde, no se toman en cuenta, a los estudiantes que estuvieron enfermos o fallecieron

Fitzpatrick & Yoels (1992), se refieren a la deserción, cuando un estudiante deja la escuela sin graduarse, independientemente si regresan o reciben algún certificado equivalente.

Lavado & gallegos (2005), elaboran su propia definición partiendo de las definiciones anteriores, donde llegaron a establecer que la deserción escolar se da siempre y cuando los individuos que habiendo asistido a la escuela el año anterior, en el año actual o corriente no lo están haciendo, exceptuando solo a aquellos que han dejado de asistir por diversos motivos.

La Deserción en el Nivel de Secundaria.

Es importante anotar el resultado de las investigaciones de los autores, Jara, Velarde, Gordillo, Guerra, León, Arroyo y Figueroa (2008), quienes manifiestan que las causas más específicas del bajo rendimiento académico son las dificultades que en cuanto a capacidad de autoaprendizaje y aprendizaje presentan los estudiantes. Se entiende que este proceso depende de la motivación del estudiante desde el uso de diferentes recursos necesarios en el desarrollo de su aprendizaje.

Es necesario describir los factores que motivan la deserción escolar. Es por ello, que se cita al autor Elis y Moreno (2002) quien resume los factores de la deserción escolar en líneas: la primera, debido a problemas de rendimiento académico teniendo en cuenta el alto porcentaje en sus debilidades y lo relacionado con la edad, en el nivel de secundaria entre los grados 6° a 9° se da un gran porcentaje de repitencia lo que conlleva al estudiante presentar edad mayor con referente a los demás compañeros. La segunda, es la falta de interés y compromiso ante sus responsabilidades escolares y el acompañamiento negativo por parte de sus padres. Y la tercera, son factores relacionado con los problemas familiares, testimonio de los estudiantes mismos tales como la obligación a los quehaceres del hogar, el embarazo y la maternidad a temprana edad.

Es importante anotar el resultado de las investigaciones de los autores, Jara, Velarde, Gordillo, Guerra, León, Arroyo y Figueroa (2008), quienes manifiestan que las causas más específicas del bajo rendimiento académico son las dificultades que en cuanto a capacidad de

autoaprendizaje y aprendizaje presentan los estudiantes. Se entiende que este proceso depende de la motivación del estudiante desde el uso de diferentes recursos necesarios en el desarrollo de su aprendizaje.

Gestión TI.

La utilización de herramientas y prácticas para gestionar aplicaciones e infraestructuras innovadoras adopta procesos ágiles en las instituciones, para ello, es necesario avocar las plataformas híbridas y la transformación digital como base de las soluciones de gestión.

Por lo tanto, la Gestión Tecnológica de la Información son recursos tecnológicos relacionados al tratamiento, almacenamiento y protección de las informaciones; involucra también la colecta, selección, comparación y distribución de datos. Un ejemplo son los softwares que se utilizan en la gestión Top-down que consiste en el análisis de la información que permite determinar la mejor opción de toma de decisiones. Es decir, aplica un enfoque de las decisiones relacionadas con la inversión y son evaluadas desde una mirada de información global, hasta ir abordando los valores y las variables más detalladas y específicas como lo muestra el siguiente esquema:

Figura 4

Modelo de toma de decisiones.



Nota. Se muestra la optimización en forma de pirámides del tp-down y down-top para la toma de decisiones de la gestión TI.

En este sentido, la gestión TI administra todas aquellas actividades de soluciones de una entidad o institución con el objetivo de optimizar el proceso de análisis facilitando la eficacia y aumentando la productividad de la calidad del servicio a través de la implementación de plataformas como herramientas que garantizan la infraestructura de TI y en este contexto funcionan los procesos correctamente. También, actúa como mecanismo de control y asegura la protección de la información optimizando la toma de decisiones. En conclusión, los procesos aplicados mediante Gestión TI garantiza el control total de los resultados proporcionado y alineados con los objetivos, en este caso con la predicción de la deserción escolar.

Sin embargo, este proceso funciona mediante el trabajo colaborativo enfocado desde la administración para la promoción de la seguridad de los datos con el objetivo de minimizar riesgos que influyen en la planeación de gestión preventiva anticipando soluciones para eventuales problemas a enfrentar.

De este modo, la gestión de TI es responsabilidad de los ejecutivos y de la junta directiva y hacen parte de ella el liderazgo, las estructuras organizativas y los procesos para garantizar que las mencionadas tecnologías de la institución sustenten y extiendan las estrategias y objetivos institucionales. (Brandis et al, 2014). La gestión TI es el eje principal de los procesos relacionados con los proyectos organizados con el objeto de mejorar la calidad de los procedimientos internos de la institución. Rahimi et al. (2016), explican que las TI ofrecen una amplia gama de oportunidades a las organizaciones para automatizar, informar y transformar un negocio. Aguilar Alonso et al. (2017) sostienen que las organizaciones dependen cada vez más de las TI para la toma de decisiones con el propósito de sostener el crecimiento del negocio.

Por otro lado, Chen y Wub (2011) resaltan la necesidad de contar con un conjunto de habilidades básicas y prácticas de gestión que necesitan los gerentes de TI, relacionadas con la alineación, la sincronización y la convergencia de tecnología y negocios, la capacidad de gestión de las mismas y las competencias de gestión, entre otras posibilidades.

La gestión de TI, según Huang et al. (2011), se centra en tareas rutinarias que se realizan a diario, como control eficiente, asignación y gestión de diversas operaciones de servicios de TI, es decir, se centra en el suministro eficaz de servicios y productos de gestión eficiente de las operaciones de tales tecnologías. Por esta razón, se asegura una excelente gestión institucional teniendo en cuenta que gestión de TI es el proceso de supervisión de todos los asuntos relacionados con las operaciones y recursos de tecnología de la información dentro de una

organización. De hecho, la gestión de TI asegura que todos los recursos tecnológicos y los empleados asociados sean utilizados correctamente y de tal manera que proporcione valor para la organización. Es decir, la gestión de TI efectiva permite a una organización optimizar los recursos y la dotación de personal, mejorar los procesos de negocio y de comunicación, y aplicar mejores prácticas. (Blog RSM Perú, 2019).

Desde el punto de vista educativo, el manejo de TI debe apuntarle a la alineación de los objetivos de las mismas con los de la organización, al aprovechamiento de oportunidades y la generación de mayor rentabilidad, al uso equilibrado, equitativo y justo de los recursos destinados a TI y a la minimización del riesgo (Marulanda Echeverry, López Trujillo y Cuesta Iglesias, 2009). En este sentido, Salazar Saavedra y Vela Londoño (2012) explican que la gestión de servicios de TI es una disciplina basada en procesos, enfocada a alinear los servicios de las mismas proporcionados con las necesidades de las empresas, con énfasis en los beneficios que puede percibir el cliente final. Para Arias Londoño y Sánchez Vélez (2013), la gestión de TI se cimienta en la realización de procesos técnicos y en la calidad de servicios orientados de manera intencional hacia el cliente, la eficiencia operativa y la capacidad de respuesta a las necesidades emergentes, mediante la implementación de políticas de cambio que sean rápidas y seguras.

Sin embargo, la comunicación es un componente importante en el sector educativo y aún más si se integra con la gestión TI. Es por ello, que la gestión TIC suple los aspectos tecnológicos, humanos y organizativos de una empresa. Las TIC, al posibilitar nuevos diseños organizativos, dan lugar a nuevas formas y procedimientos de gestión, a nuevas estrategias, nuevas relaciones y formas de integración de las personas. Su impacto en la organización y su importancia radica en la apropiación que de ellas hagan los individuos y en el éxito de

integración de las mismas a través de sus aplicaciones a los procesos y procedimientos. (Moreno et al., 2014).

De hecho, la gestión de las TIC en las instituciones educativas debe entenderse como un proceso intencionado de comprensión e intervención en las relaciones entre ciencia, tecnología y sociedad, que integra la transferencia y la apropiación tecnológica y la innovación, en un mismo plano, bajo el desarrollo de las capacidades de difusión y absorción en la gestión del conocimiento. Es tarea de las instituciones educativas no solo potenciar las actitudes, aptitudes, habilidades, destrezas y competencias de los miembros de su organización, tanto individualmente como colectivamente, sino también la de crear redes de información, comunicación e intercambio de los suyos, como de éstos con sus pares y con el entorno, de manera de crear las capacidades. (Moreno et al., 2014). Como lo señala IBM, “el entorno de TI se compone por una multitud de hardware, red y software que incluyen más no se limitan a computadoras, servidores, enrutadores, aplicaciones, microservicios y tecnologías móviles”. (Red Techtarget, 2021).

Gestión de Proyecto TI

La sociedad actual se enfoca en tendencias referentes a procesos permanentes de mejora tanto de productos como de servicios. Por esta razón, es importante adoptar un modelo de gestión que sirva como referente a la aplicación de conocimientos, procesos, habilidades, herramientas y técnicas puede tener un impacto positivo en el éxito de un proyecto. (López, 2017). Este modelo de gestión organiza todos los recursos (oportunidades educativas) y a las personas en un sentido y propósito determinado. López (2017), le denomina modelo de gestión de calidad por ser la manera de concebir, planificar, organizar, dirigir, administrar, evaluar y rendir cuentas para lograr mejores resultados en los aprendizajes. Dicha gestión se realiza a

través de tres estrategias generales: organizativa, de formación continua y acompañamiento. De acuerdo con el modelo, la gestión son acciones institucionales y pedagógicas que entrelazadas entre sí logran los aprendizajes.

Es inevitable incitar la gestión de proyecto TI sin antes valorar la siguiente frase: “Somos aquello que hacemos repetidamente. La excelencia, pues, no es un acto sino un hábito”.

(Aristóteles). De esta manera, la determinación de un modelo de gestión TI, contempla la gestión de calidad aplicada por Malcolm Baldrige a la empresa Servicios Ortuño y Castellón R.L. (ORTYCAST), de Estelí. El modelo Malcolm Baldrige según Camisón (2006), hace referencia a que las organizaciones demuestren una gestión de la calidad efectiva, mediante la formación y la implicación de los empleados, en la mejora de la empresa. (Rodríguez y Sandoval, 2017).

El modelo Baldrige integra un conjunto de variables y criterios de calidad en la calidad del diseño y en la prevención. Es decir, “Visión a largo plazo del futuro”. Esta gestión está basada en datos y hechos que involucran la responsabilidad social y la orienta mediante análisis a resultados que pueden ser mejorados. El enfoque de la gestión por proceso es lo que fundamenta la gestión de calidad de Baldrige, determinando la información mediante clasificación y análisis de datos e información que apoya los procesos claves de la organización y el rendimiento de la organización. (Valle, 2017).

En conclusión, la gestión de proyectos según PMI (2013), es la aplicación de conocimientos, habilidades, herramientas y técnicas a las actividades del proyecto con el fin de satisfacer sus necesidades. Según Carvalho y Rabechini (2011), complementan esta definición, mencionando la norma ISO 10006, que dice que el manejo del proyecto incluye la planificación,

organización, supervisión y control de todos los aspectos del proyecto en un proceso continuo para lograr sus objetivos.

Minería de Datos

Partiendo de la identificación de la minería de datos como estrategia es necesario citar a Rivero (2012), en su tesis de grado clasificación de clientes mediante técnicas de minería de datos para la empresa Comertex S.A., anota como objetivo “diseñar un modelo conceptual utilizando técnicas de minería de datos para la construcción de perfiles de los clientes que permita a la organización tomar decisiones estratégicas sobre los segmentos actuales de la empresa”.

Por consiguiente, es importante afirmar que este autor busca diseñar un modelo de gestión TIC basado en minería de datos como técnica que permitió la generación de nuevas estrategias de comercialización desde la perspectiva “No dirigida” partiendo de datos para la obtención de patrones, que mediante la aplicación de varias herramientas de infraestructura de negocio ayudó a resolver necesidades de la empresa, es importante anotar la estructura de la técnica iniciando con la consolidación, depuración y adecuación de los datos para el posterior tratamiento estadístico y de minería de datos.

Otro aporte referente en esta propuesta es el del autor Aristizabal (2017) en su tesis de grado doctoral, presenta el diseño de un modelo metodológico para minería de datos educativos en el aula de educación media presencial cuya implementación y resultado contribuyan a la reflexión sobre el acto pedagógico. El autor explica: “partiendo de los objetivos se logró reflexionar desde los constructos los cuales dieron origen a las variables y que delimitaron la investigación desde las prácticas de enseñanza, el enganchamiento estudiantil y la evaluación, los cuales se utilizaron como ejes para la discusión de los resultados”.

Se puede afirmar, que la anterior tesis parte desde los referentes de comportamiento de Hall y Simeral (2015) quienes establecen categorías de contribuciones de los diversos actores los modelos de gestión TIC que consiste en un sistema cíclico que incluye cuatro fases como: los datos, la preparación, la analítica de visualizaciones y el conocimiento.

De acuerdo al referente Martínez (2012), presenta su tesis de grado para Magister, aplicación de técnicas de minería de datos para mejorar el proceso de control de gestión en ENTEL, cuyo objetivo es “reducir el tiempo de cálculo de los indicadores de servicios privados, para lo cual se aplicó el modelamiento multidimensional, técnicas de minería de datos y automatización de procesos, y de este modo poder entregar información más oportunamente...”.

Teniendo en cuenta que este autor aplica los procesos metodológicos conocidos como Knowledge Discovery in Databases (KDD) implementando estrategias como CRISP-DM usada para el desarrollo de la técnica de minería de datos. Las actividades las desarrolló en cuatro fases: el levantamiento de las métricas, la selección y el procesamiento de las fuentes de datos, y luego aplicó la técnica Clustering para obtener grupos de elementos don datos de clientes y servicios generando un modelo de clasificación para asignar acuerdos mediante la similitud y así estimar los ingresos percibidos con la ayuda de datos.

La Minería de Datos como Estrategia de Predicción de la Deserción Escolar

En cuanto a la deserción se encontraron referentes que involucran la técnica de la minería de datos como apoyo para la predicción de la deserción como es el autor Ordoñez (2013), quien plantea un modelo de minería de datos aplicando la metodología CRISM – DM contando con la ayuda del análisis de la información que los estudiantes proporcionan a las bases de datos del sistema académico Syllabus y el entorno de aprendizaje para lograr obtener patrones de

comportamiento para conocer cuáles son las posibles causas que motivan a los estudiantes desertar de sus estudios de primera etapa universitaria.

Es muy relevante este referente, ya que, utilizan una metodología que permite crear el modelo de gestión TIC para la generación de proyectos de minería de datos pretendiendo obtener un modelo de análisis de datos con la ayuda de la implementación de algoritmo de inteligencia artificial ya incorporado en la herramienta de procesamiento de datos weka para conocer las causas de la deserción en la primera etapa universitaria.

La autora Escobar (2015) en su monografía utiliza herramientas para identificar los factores asociados a la deserción escolar en la Institución Educativa Superior María Inmaculada en el nivel de secundaria periodo escolar 2014.

Esta monografía se refiere a la construcción de una primera evidencia de la emergencia de conductas negativas acompañadas por una trayectoria educacional marcada por la indisciplina, la inasistencia frecuente a clase, el desinterés y bajo rendimiento académico, es por eso que se requiere indagar la deserción en esta institución y así ayudar a solucionar en gran parte esta situación por la que atraviesa muchos jóvenes.

Según Carvajal (2012) propone estrategias que reduzcan el impacto de las causas académicas en la deserción. También divide la deserción en tres tipos: la deserción precoz, temporal y tardía señalando al estudiante de tres maneras: el desertor del grado, el desertor de la institución y el desertor del sistema educativo. Basándose en estas categorizaciones, se facilita el análisis y la selección de datos para predecir la deserción en un grado en específico como es en el caso de noveno del nivel de secundaria.

Otro aporte referencial es el de Varón (2017) este reporte “analiza el problema de la deserción estudiantil en la Institución Educativa Siete de Agosto, de Cali, por medio del estudio

de los factores internos y externos y el análisis del desarrollo de los procesos educativos como causantes del retiro escolar”. En este artículo se sugieren también estrategias encaminadas a disminuir el riesgo de desertar mediante la descripción realizando un estudio exploratorio y utilizando el análisis cuantitativo y cualitativo de la comunidad educativa relacionada con este fenómeno educativo, con un especial interés por el ser. Con esta estrategia se busca identificar una tendencia de retiro escolar con diversidad de variables en un periodo de tiempo específico dentro del marco temporal seleccionado.

El Algoritmo CART para Predecir el Nivel de Deserción Escolar

El algoritmo CART de Leo Breiman (Alcover, R., Benlloch, J., Blesa, P., Calduch, M. A., Celma, M., Ferri, C., & Robles, 2007), realiza particiones binarias, con el objetivo que la media de cada rama sea diferente y, por tanto, discrimine con suficiente precisión, un número adecuado de particiones, para asignar a cada hoja un valor cercano a la media de los elementos que caen en ella. Este algoritmo genera arboles de fácil interpretación con resultados óptimos (Ara et al., 2015), lo que se considera una ventaja, al crear modelos predictivos (Lin, 2015).

El modelado CART se basa en metodología implementada en árboles de clasificación y regresión. Expertos en estadísticas de la Universidad de Stanford y la Universidad de California en Berkeley (1984), incorporan extensiones de modela CART para mejorar los resultados de análisis y predicción. Esto, le ha conllevado al diseño de un modelo predictivo desde el enfoque de minería de datos basados en “bagging” y la potenciación. (Minitan Insights).

Diseño Metodológico

El proyecto que se presenta sintetizó conceptos y teorías a través de una revisión bibliográfica y bases de datos que proporcionó el sustento científico abordando la problemática en el respectivo desarrollo de la investigación. De este modo, el diseño metodológico consistió en la interpretación de hipótesis mediante la investigación cuantitativa y por supuesto el logro de los objetivos a través de la investigación cualitativa. En este proceso se dieron respuestas a las preguntas de investigación comprobando la hipótesis planteada y concibiendo el desarrollo de los objetivos.

En este sentido, se consideró aplicar una metodología de tipo mixta porque se enfoca en información cualitativa en cuanto a la descripción de los aspectos que inciden en la problemática de deserción escolar y es cuantitativa ya que, se espera que arroje datos variables como resultado de la investigación relacionada con el porcentaje de deserción escolar por cada año lectivo.

(Lerma González, 2016).

Figura 5*Diseño metodológico de la investigación*

Nota. La figura representa detalladamente lo planteado por Calderón (2011) en su presentación de la Investigación educativa.

Metodológicamente se abordó un enfoque proyectivo, toda vez que se persiguió la utilización o aplicación de tecnología que permitiera realizar el procesamiento y análisis de la información de grupos y grados en el nivel de secundaria almacenada en bases de datos, partiendo de la aplicación de técnicas de análisis para identificar el porcentaje de la deserción escolar por periodos anuales. Por otra parte, también permitió planear y desarrollar un proceso metódico de exploración y búsqueda que incluye actividades en las cuales se describe, se compara, se analiza, se explica y predice (Hurtado, 2000).

Por consiguiente, Hurtado (2000), dice que la investigación proyectiva se apoya en la creación de una propuesta ya sea un modelo, plan, etc. como recurso para la solución que responda a una necesidad de tipo práctico, presentada en un espacio geográfico determinado, grupo social o institución. Este tipo de investigación se da en un área específica del conocimiento, en el cual se realiza un diagnóstico de las necesidades, se analizan las causas que las producen y se buscan tendencias futuras. Además, Hurtado (2000) sostiene que “La

investigación proyectiva se ocupa de cómo deberían ser las cosas, para alcanzar unos fines y funcionar adecuadamente. La investigación proyectiva involucra creación, diseño, elaboración de planes, o de proyectos”.

Por otro lado, (Gómez, 2006) señala que, bajo la perspectiva cuantitativa, la recolección de datos es equivalente a medir. De acuerdo al autor la medición esboza la asignación de números a objetos y eventos de acuerdo a ciertas reglas. Muchas veces el concepto se hace observable a través de referentes empíricos asociados a él. Por ello, desde el enfoque cuantitativo se pretende obtener la explicación de una realidad social en una comunidad educativa desde la perspectiva interna y objetiva. la intención es buscar la exactitud de mediciones o indicadores sociales con el fin de generalizar sus resultados a poblaciones o situaciones amplias. Trabajan fundamentalmente con el número, el dato cuantificable (Galeano, 2004).

En este sentido, la recolección de información desde la minería de datos permite cuantificar numéricamente e indica con facilidad lo que interesa medir. (Namakforoosh, 2005), explica que un instrumento de medición considera tres características principales:

La validez. Se refiere al grado en que la prueba está midiendo lo que en realidad se desea medir.

Confiabilidad: se refiere a la exactitud y a la precisión de los procedimientos de medición.

Factibilidad: se refiere a los factores que determinan la posibilidad de realización, que son tales como: factores económicos, conveniencia y el grado en que los instrumentos de medición sean interpretables.

Ahora bien, la orientación de la calidad educativa se promueve a través de las acciones de mejoramiento, mediante el análisis e interpretación válida y confiable, En efecto, una adecuada evaluación, que tome en consideración los avances de las ciencias de la cognición, de la pedagogía y de la administración, aporta elementos para una acertada toma de decisiones en los

distintos ámbitos educativos, tales como: los procesos de enseñanza-aprendizaje, la formulación de políticas, programas y proyectos, la asignación de recursos y el perfeccionamiento de los procesos curriculares, pedagógicos y de gestión (Fernández, 2005).

En este orden de ideas, (Gaviria & Barrientos, 2001; Barrientos-Marín, 2008; Correa, 2004; Chica-Gómez, Galvis-Gutiérrez & Ramírez-Hassan, 2010; Gómez, 2014; Hernández-Angulo, 2015), se basan en información procesada mediante un análisis estadístico, donde fundamentalmente se consideran variables y relaciones primarias, sin tener en cuenta las verdaderas interrelaciones, que por lo general están ocultas y que se pueden descubrir utilizando un tratamiento más complejo de los datos, que es posible con la minería de datos. Por lo tanto, la aplicación de las técnicas y herramientas de la minería de datos a partir de la deserción se conoce como data mining. Un estudio muy relevante en los últimos años, pues se pueden utilizar técnicas para explicar y/o predecir cualquier fenómeno dentro del campo educativo (Timarán-Pereira et al., 2013a; 2013b).

Asimismo, para seleccionar el patrón en la realización de la predicción es necesario la clasificación con árboles de decisión considera clases disjuntas, de forma que el árbol conducirá a una y sólo una hoja, asignando una única clase a la predicción (Hernández & Lorente, 2009). De hecho, con esta técnica se permite obtener un modelo que predice los nuevos casos de estudiantes, cuáles son los factores socioeconómicos, académicos e institucionales asociados al buen (por encima de la media) o mal (por debajo de la media) desempeño académico (Timarán, Caicedo e Hidalgo, 2019).

De este modo, se aborda el diseño de investigación descriptivo porque puntualiza y centra la población sujeta a la investigación describiendo aspectos que incurren en la deserción. Por ello, se proponen las siguientes fases para alcanzar los objetivos propuestos:

Primera fase: Realización de un análisis diagnóstico para identificar el nivel de deserción a través de los resultados del SIMAT, pruebas ICSE y pruebas saber entre los años 2015 a 2020.

Segunda fase: Aplicación de la técnica algoritmos CART y árboles de decisiones utilizando la minería de datos como estrategia para predecir la deserción con las variables, el rendimiento académico en PUNTO EDU, las Pruebas Icfes Saber 11 y la autoevaluación institucional en los últimos 6 años (2014 a 2020) en los grupos del nivel de secundaria.

Tercera fase: Evaluación de los resultados predictivos arrojados en los análisis estadísticos para presentar propuesta de Modelo de Gestión Ti para predecir el nivel de deserción en los años 2021 -2025.

Caracterización de la Población y Muestra

La Institución Educativa de Coyongal se encuentra en zona rural al sur del municipio de Magangué, a orillas del Río Magdalena, factores que inciden en el proceso académico de los estudiantes debido a la distancia y dificultad de transporte, no cuenta con herramientas didácticas y/o oportunidad de asistencia a otros programas que inviten al estudiante valorar la permanencia en su proceso académico y formativo. Uno de los problemas es el factor social, económico y familiar. Por lo que, la distancia de sus viviendas para ir a la escuela es lejana, el servicio de transporte escolar no es garantizado, el desorden alimenticio de los estudiantes por cumplir con un horario establecido para tomar el transporte escolar y las constantes lluvias debido al cambio de clima.

En este sentido, la Institución Educativa de Coyongal cuenta con una población estudiantil matriculada de 1.078 estudiantes de géneros hombres y mujeres con edades que oscilan entre 5 a 17 años de edad. La matrícula se distribuye en cuatro niveles educativos:

Tabla 2*Población estudiantil*

		NIVELES EDUCATIVOS			
		Preescolar	Primaria	Secundaria	Media
Cantidad de	De 5 años	210			
estudiantes	Edades entre		430		
por niveles	6 a 10 años				
	Edades entre			346	
	11 a 14 años				
	Edades entre				92
	15 a 17 años				
	Porcentajes	19,48%	39,88%	32,09%	8,53

Nota. Se muestran los porcentajes totales de la población seleccionada para la intervención en la investigación.

La muestra corresponde al 32,09% de la población estudiantil matriculada en el nivel de secundaria. Cantidad de estudiantes objeto de investigación para realizar el análisis estadístico en la predicción de la deserción escolar.

Selección de las Técnicas a Utilizar en el Desarrollo del Trabajo

La gran ventaja de representar la predicción de la deserción mediante la técnica de minería de datos representó el conocimiento extraído de la técnica de árboles de decisión mediante un diagrama de flujo relacionando atributos que validaron cada una de las variables de concepto académico. También se fundamentó la técnica de algoritmos CART para formar arboles de decisiones clasificando un conjunto de datos estadísticos con metodología cuantitativa

logrando así la predicción de la deserción en un periodo que permita diseñar estrategia de permanencia.

El Método de Clasificación Basada en Árboles de Decisión en la Predicción de la Deserción Estudiantil

Esta técnica consistió en el estudio comparativo de algoritmos con el fin de detectar al estudiante con probabilidad de deserción. Se diseñó un diagrama de flujo usando variables concretas que ayudaron clasificar atributos porcentuales realizando comparación entre cada una de las variables. Este método de clasificación se utilizó para estimar los parámetros del modelo en dos conjuntos de datos: estudiantes matriculados y estudiantes desertados. La clasificación se supervisa mediante el paradigma estadístico de regresión logística y de inteligencia artificial. Se comprobó el comportamiento de datos registrados utilizando un muestreo aleatorio aplicando como resultados dos variables binarias: el 0 un valor real y el 1 un valor estimado. De este modo, para encontrar los aciertos se estima pertinente el análisis mediante Rstudio de la siguiente forma:

Análisis comparativo de matriculado y desertado: el árbol de decisión se diseñó mediante un grupo de nodos que albergaron segmentos de valor con una serie de reglas asignando entradas y salidas. Después de organizar los documentos descargados de la plataforma el SIMAT, se realizó la tabla de en la que se seleccionaron las variables generando factores en los que se le asignaron valores y así la función del programa de Excel ayudó a reevaluar mediante el sistema de Algoritmos CART. A continuación, se explica el proceso:

En la siguiente tabla se organizaron las variables que permitieron estructurar de manera estadística y porcentual la predicción:

Figura 6

Análisis de matrícula a partir de la minería de datos de los años 2014 a 2020.

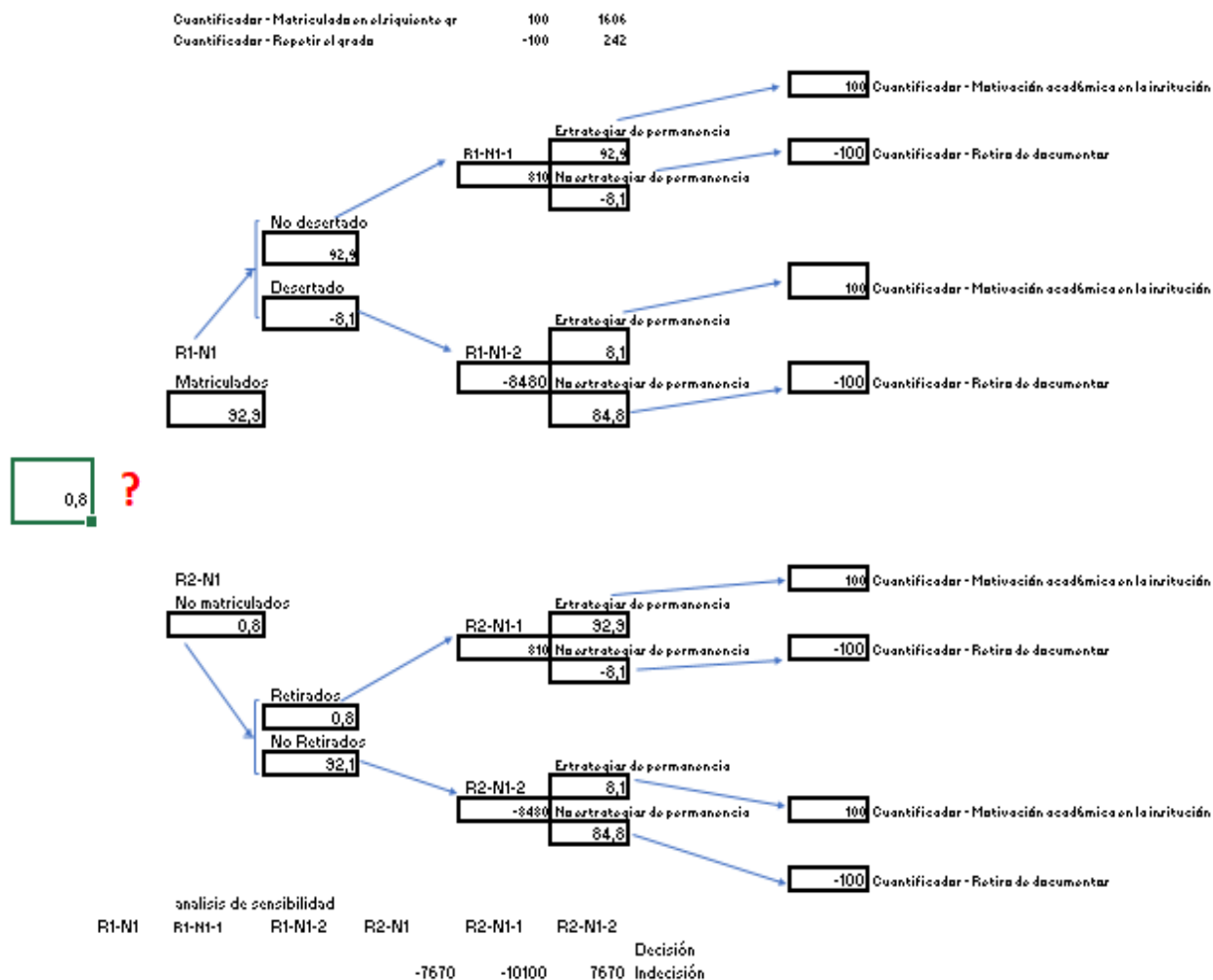
ANÁLISIS DE MATRÍCULA																												
NIVEL DE SECUNDARIA																												
	2014				2015				2016				2017				2018				2019				2020			
	6	7	8	9	6	7	8	9	6	7	8	9	6	7	8	9	6	7	8	9	6	7	8	9	6	7	8	9
MATRICULADOS	87	56	70	59	81	47	45	60	84	70	41	38	94	83	41	36	102	64	61	29	93	95	59	51	75	92	84	51
APROBADOS	47	45	60	47	70	41	38	53	83	41	36	30	64	61	29	31	95	59	51	24	92	84	51	51	64	58	72	42
REPROBADOS	12	5	4	3	5	2	3	2	0	11	1	3	7	9	3	0	2	1	1	0	0	3	2	0	1	3	2	2
DESERTADOS	26	6	6	7	5	4	4	3	1	16	4	5	21	13	8	4	3	3	6	3	1	8	6	0	8	11	9	7
RETIRADOS	2	0	0	2	1	0	0	2	0	2	0	0	2	0	1	0	2	1	3	2	0	0	0	0	2	3	1	2
	40	11	10	12	11	6	7	7	1	29	5	8	30	22	12	4	7	5	10	5	1	11	8	0	11	17	12	11

Nota. Se muestra la relación de las variables por año creando el compormiento histórico en la clasificación de los datos.

A partir de esta cantidad de datos en representación de las variables: matriculados, reprobados, desertados y retirados por cada año (2014, 2015, 2016, 2017, 2018, 2019 y 2020). En estos valores se determinaron respuestas concretas y precisas colocando en el factor el valor de 0.8 teniendo en cuenta el cuantificador matriculado en el siguiente grado y matriculados repitiendo el grado.

Figura 7

Árbol de decisiones de la variable matricula



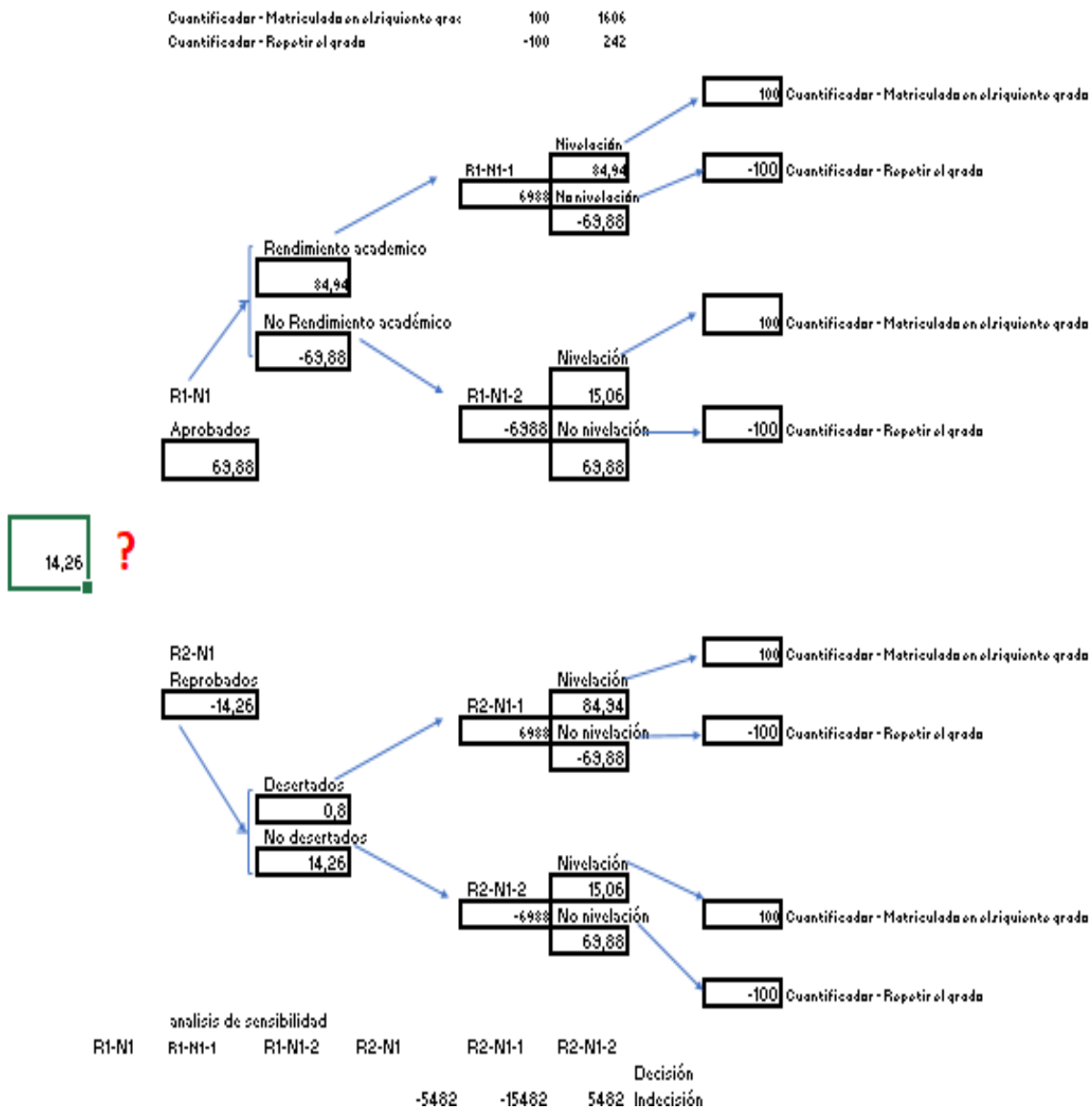
Fuente. Propia del investigador.

Análisis comparativo de rendimiento académico. Este atributo expresó condiciones desde las instancias clasificadas indicando caminos desde la raíz disponibles para validar el muestreo con porcentajes exactos obtenidos del valor 0.8. Las variables utilizadas de la minería de datos relacionada con la minería de datos fueron aprobados, reprobados, nivelados, no

nivelados y desertados. Para ello, se utilizaron los documentos arrojados de la plataforma Punto Edu por cada periodo de los cuatro periodos del año.

Figura 8

Árbol de decisiones de la variable rendimiento académico



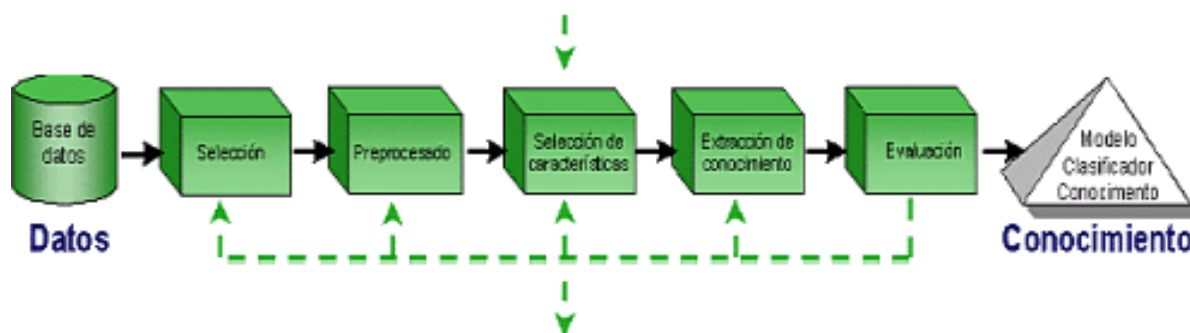
Fuente. Propia del investigador

Técnicas de Minería de Datos Utilizadas en Estudios sobre Deserción Estudiantil

Esta permitió la minería de datos provenientes de la inteligencia computacional (inteligencia artificial, machine learning, clasificación de patrones, etc.) y de la estadística (análisis multivariado, inferencia, teoría estadística del aprendizaje, regresión, etc.). estas técnicas utilizan patrones basadas en algoritmos, más o menos sofisticados que se aplican sobre un conjunto de datos para obtener unos resultados (Gonzalez y Llaugel, 2015). La técnica se determinó bajo los criterios de cálculos a partir de un modelo proporcionado para buscar patrones de comportamiento de las variables procesadas en el conjunto de datos. Para ello, se utilizaron los siguientes parámetros:

Figura 9

Modelo de técnica de la minería de datos



Fuente. Gonzales y Hernández (2006). Técnicas de minería de datos.

En primer lugar, se hizo agrupamientos de registros de datos utilizando el usuario de dirección de grupos de la Plataforma Punto Edu (con previa solicitud a cada docente), centrando la minería de datos en el rendimiento académico de cada asignatura y cada área por cada periodo del año. Es de aclarar que la Institución Educativa de Coyongal cuenta con cuatro periodos divididos en un periodo de 10 semanas escolares en el año. Al finalizar cada periodo se realiza consejo académico, en él se evalúan porcentualmente la aprobación de cada asignatura y la

superación de cada área. Su sistema de evaluación cuenta con una valoración estipulada de la siguiente manera:

Tabla 3

VN: Valoración Nacional

DS	(DESEMPEÑO SUPERIOR):	DE 90 A 100	(SUPERACIÓN CON MÉRITOS DE LOS DESEMPEÑOS EN RELACIÓN CON LAS ÁREAS OBLIGATORIAS Y FUNDAMENTALES).
DA	(Desempeño Alto):	De 80 a 89	(Superación de los desempeños en relación con las áreas obligatorias y fundamentales).
DB	. (Desempeño Básico):	De 60 a 79	(Superación de los desempeños necesarios en relación con las áreas obligatorias y fundamentales).
DBJ	(Desempeño Bajo):	De 20 a 59	. (Dificultad para superar los desempeños en relación con las áreas obligatorias y fundamentales.).

Nota. La tabla muestra el Sistema Institucional de Evaluación de la INSTEC.

De acuerdo a esta valoración se entrenaron algoritmos predictivos seleccionando la técnica supervisada precisando estadísticamente similitudes entre varios atributos: asistencia del estudiante e inasistencia permanente y continua. También se tiene en cuenta el valor promedio en el grupo y el valor promedio porcentual en cada asignatura. Esta cantidad de datos permitieron analizar el nivel de aprobación y reprobación de cada estudiante, evaluar la asignatura con más alto índice de reprobación en el periodo académico. Para ello, se abordó el modelo predictivo a partir de tres enfoques:

El enfoque de selección de datos: se realizó selección de datos del estudiante de acuerdo a la ficha de matrícula utilizada en la institución según la información arrojada en el SIMAT. Se hizo registro mediante la minería de datos organizando la ficha de matrícula de cada estudiante matriculado para los grados de 6°, 7°, 8° y 9° de la Institución Educativa de Coyongal.

El enfoque de estado académico: se clasificaron valores lineales y no lineales basadas en características explicativas discriminatorias y separadas teniendo en cuenta la valoración del sistema de evaluación institucional, estrategias motivacionales en la práctica pedagógica, la gestión institucional y estrategias de permanencias de los años 2014 a 2020.

El enfoque de categorías predictivas: se definieron reglas asociativas predictivas basada en árboles de decisiones a través de Algoritmos CART utilizando el lenguaje de programación Rstudio. El modelo predictivo se diseñó con la estrategia de la minería de datos.

Procesamiento y Flujo de Datos para la Predicción

Figura 10

El flujo de datos consistió en la metodología se basó en cinco etapas



Nota. Se muestra el flujo de la minería de datos procesados en cinco etapas.

Este método representó el modelo basado en arboles de decisiones con algoritmos CART, proporcionando una matriz que contiene los datos de los estudiantes tomados de la ficha de matrícula adjunta en el SIMAT. Se tomaron datos como la identificación, nombres y apellidos, genero, fecha de nacimiento, lugar de residencia y grado matriculado.

Figura 11

Minería de datos a partir de la ficha de matrícula que contiene los datos del estudiante.

INSTITUCIÓN EDUCATIVA DE COYONGAL

Creada por ordenanza N° 020 de 2.002 y el Decreto 143 de 2.003, emanados de la Asamblea departamental y la Gobernación de Bolívar
Aprobado según Resolución N° 35.18 de Octubre 8 de 2.019, emanada de la Secretaría de Educación Municipal de Magangaj
Registro DANIE 2134.30002337 - NIT: 800181494-2
Coyongal, Magangaj, Bolívar

FOLIO:		NÚMERO DE MATRICULA:		CÓDIGO:		SEDE:		LRM002 - 23/11/2021	
APELLIDOS Y NOMBRES DEL ESTUDIANTE		FECHA DE NACIMIENTO			LUGAR DE NACIMIENTO		DOCUMENTO DE IDENTIDAD		LUGAR DE EXPEDICIÓN
		DÍA	MES	AÑO	MUNICIPIO	DEPARTAMENTO	R.C.	T.I.	
DIRECCIÓN DE RESIDENCIA		BARRIO			N° DE CELULAR		CORREO ELECTRÓNICO		
POSEE CONDICIÓN DE DESPLAZADO		SI	NO		POSEE ALGUNA DISCAPACIDAD		SI	NO	CUÁL:
DATOS PERSONALES DE LOS PADRES DE FAMILIA Y/O ACUDIENTE									
APELLIDOS Y NOMBRES DE LA MADRE		DOCUMENTO DE IDENTIDAD			OCUPACIÓN		N° DE CELULAR		CORREO ELECTRÓNICO
APELLIDOS Y NOMBRES DEL PADRE		DOCUMENTO DE IDENTIDAD			OCUPACIÓN		N° DE CELULAR		CORREO ELECTRÓNICO
APELLIDOS Y NOMBRES DEL ACUDIENTE		DOCUMENTO DE IDENTIDAD			OCUPACIÓN		N° DE CELULAR		PARENTESCO
INSTITUCIÓN EDUCATIVA DONDE CURSÓ LOS GRADOS ANTERIORES		GRADO CURSADO		AÑO		INSTITUCIÓN EDUCATIVA DONDE CURSÓ LOS GRADOS ANTERIORES		GRADO CURSADO	
COMPROMISO: Los firmantes aceptamos el Proyecto Educativo Institucional - PEI y cumplir con lo establecido en el Manual de Convivencia de la Institución Educativa de Coyongal									
MATRICULA Y/O RENOVACIÓN				MATRICULA Y/O RENOVACIÓN				FIRMAS	
DÍA	MES	AÑO	GRADO	DÍA	MES	AÑO	GRADO	ACUDIENTE	ESTUDIANTE
CANCELACIÓN DE MATRICULA				CANCELACIÓN DE MATRICULA				FIRMAS	
DÍA	MES	AÑO	MOTIVO	ACUDIENTE		ESTUDIANTE		SECRETARIA	RECTOR
OBSERVACIONES:									

Fuente. SIMAT, Institución Educativa de Coyongal.

Por otro lado, se diseñó una segunda matriz que contiene datos académicos relacionados con el periodo académico, asignaturas aprobadas y no aprobadas, promedio de asistencia e inasistencia, promedio de notas según el periodo y estrategias motivacionales y de permanencia.

Figura 12.

Matriz de relación de asignaturas aprobadas y no aprobadas, promedio de asistencia.

ETC	ESTADO	JERARQUIA	INSTITUCIO	DANE	CALENDARI	SECTOR	SEDE	CODIGO_DA	CONSECUTIV	ZONA_SEDE	JORNADA	GRADO_COT	GRUPO	MODELO	MOTIVO	FECHAINI	FECHAFIN	NUJ
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	TOLU	2.134E+11	2.134E+13	RURAL	MAÑANA	5	501	EDUCACIÓN TRADICION	27/12/2021	20:52	SAMAKO211
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	ROMA	2.134E+11	2.134E+13	RURAL	MAÑANA	5	501	EDUCACIÓN TRADICION	23/12/2021	20:54	BALAKO1355
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	PUNTA DE C	2.134E+11	2.134E+13	RURAL	MAÑANA	2	201	EDUCACIÓN TRADICION	25/12/2021	14:41	SADAKO213
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	SANTA COIT	2.134E+11	2.134E+13	RURAL	MAÑANA	5	501	EDUCACIÓN TRADICION	31/01/2022	15:58	YALAKO5496
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	SANTA COIT	2.134E+11	2.134E+13	RURAL	MAÑANA	5	501	EDUCACIÓN TRADICION	25/01/2022	10:57	YASAKO3811
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	PLAYA DE LF	2.134E+11	2.134E+13	RURAL	MAÑANA	1	101	EDUCACIÓN TRADICION	25/12/2021	14:32	FIOAKO1255
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	PLAYA DE LF	2.134E+11	2.134E+13	RURAL	MAÑANA	3	301	EDUCACIÓN TRADICION	25/12/2021	14:35	YABAOK1251
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	TOLU	2.134E+11	2.134E+13	RURAL	MAÑANA	2	201	EDUCACIÓN TRADICION	27/12/2021	20:47	ISAKO12568
022	MAGANGUE	RETRADO	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	9	902	EDUCACIÓN NO ATENDI	9/09/2022	20:12	LULAKU2052
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	4	402	EDUCACIÓN TRADICION	30/12/2021	13:56	ANDAKU13
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	3	301	EDUCACIÓN TRADICION	30/12/2021	13:53	MOSAKU958
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	TOLU	2.134E+11	2.134E+13	RURAL	MAÑANA	3	301	EDUCACIÓN TRADICION	27/12/2021	20:48	YARAKU564
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	PUNTA DE C	2.134E+11	2.134E+13	RURAL	MAÑANA	3	301	EDUCACIÓN TRADICION	21/02/2022	8:39	YONAKU149
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	SANTA PABL	2.134E+11	2.134E+13	RURAL	MAÑANA	3	301	EDUCACIÓN TRADICION	23/12/2021	21:06	LOAKU2758
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	9	901	EDUCACIÓN TRADICION	31/12/2021	11:29	YIRAKU2210
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	SANTA PABL	2.134E+11	2.134E+13	RURAL	MAÑANA	1	101	EDUCACIÓN TRADICION	23/12/2021	21:03	YONAKU111
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	ROMA	2.134E+11	2.134E+13	RURAL	MAÑANA	4	401	EDUCACIÓN TRADICION	23/12/2021	20:53	KARAGL205
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	PUNTA DE C	2.134E+11	2.134E+13	RURAL	MAÑANA	2	201	EDUCACIÓN TRADICION	25/12/2021	14:41	ESNAI20671
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	PLAYA DE LF	2.134E+11	2.134E+13	RURAL	MAÑANA	1	101	EDUCACIÓN TRADICION	28/01/2022	16:17	TIAAI48147
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	5	501	EDUCACIÓN TRADICION	21/02/2022	8:34	EDÉAUM1885
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	FIN DE SEM	26	2601	PROGRAMA PARA JÓVEN	6/04/2022	8:27	ROSAUM2028
022	MAGANGUE	RETRADO	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	7	701	EDUCACIÓN NO ATENDI	8/04/2022	11:27	ORIAUM1206
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	0	TRO1	EDUCACIÓN TRADICION	14/09/2022	16:29	SARAL87171
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	8	803	EDUCACIÓN TRADICION	9/03/2022	20:17	BRALB1354
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	3	301	EDUCACIÓN TRADICION	30/12/2021	13:53	MARALB177
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	4	401	EDUCACIÓN TRADICION	30/12/2021	13:56	SARALB1536
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	10	1001	EDUCACIÓN TRADICION	31/12/2021	11:39	SEBALB2016
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	5	501	EDUCACIÓN TRADICION	30/12/2021	13:58	DARALB2121
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	3	301	EDUCACIÓN TRADICION	17/03/2022	11:20	SAMALB9901
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	PUNTA DE C	2.134E+11	2.134E+13	RURAL	MAÑANA	2	201	EDUCACIÓN TRADICION	25/12/2021	14:41	MARALB585
022	MAGANGUE	MATRICULA	MAGANGUE	I.E. DE COYC	2.134E+11	A	OFICIAL	I.E. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	4	401	EDUCACIÓN TRADICION	10/03/2022	9:30	YUAAIB3774
003	MARANGU	RETRADO	MAGANGUE	I.F. DE COYC	2.134E+11	A	OFICIAL	I.F. COYONG	2.134E+11	2.134E+13	RURAL	MAÑANA	10	1002	EDUCACIÓN NO ATENDI	27/01/2022	17:23	MARALB667

Fuente. Archivos de Instec (2022)

Desarrollo de los Objetivos Propuestos en esta Investigación

Se planteó el desarrollo de un modelo de gestión TI apoyado en la minería de datos para predecir la deserción de los estudiantes en el nivel de secundaria en la Institución Educativa de Coyongal teniendo en cuenta que la investigación esbozó un enfoque metodológico permitiendo la preparación, el análisis y el conocimiento de los resultados y la contribución a la reflexión de la sistematización realizada a través de la organización y planteamiento del estudio mediante el desarrollo de las siguientes fases:

Fase 1. Realización del Análisis Diagnóstico.

En esta primera fase se desarrolló el primer objetivo de la investigación que consistió en la realización del análisis diagnóstico para identificar el nivel de deserción a través de los resultados del SIMAT, pruebas ICSE y pruebas saber entre los años 2015 a 2020. En el proceso de la planeación se revisaron documentos de matrícula de los estudiantes por cada año en cada

uno de los grados en el nivel de secundaria en la base de datos del Simat. Se hizo un análisis comparativo de las pruebas ICSE y pruebas saber por cada año desde el año 2014 a 2020.

Como inicio de la fase se sistematizaron las siguientes actividades:

Selección de documentos de matrículas en el Simat diseñando categorías de análisis comparativos por cada año escolar en cada uno de los grados del nivel de secundaria.

Aplicación del análisis comparativo de los resultados de las pruebas saber entre los años 2015 a 2020 de los grados 5° y 9°.

La recolección documental se basó en la investigación histórica para determinar variables cuantitativas y cualitativas mostrando un proceso práctico y real. La recolección de documentos para la minería de datos siguió secuencias ordenadas según lo indica la investigación histórica según (Best, 1982), con el propósito de establecer sucesos, ocurrencias o eventos en un ámbito que interesó en la investigación teniendo en cuenta la metodología y el modo como se enfocan los problemas de deserción y se buscan las causas que conllevaron el alto porcentaje en la población estudiantil en el nivel de secundaria en la Institución Educativa de Coyongal. Por ello, se realizó la planeación en el siguiente esquema:

Figura 13

Esquema de la planeación de la selección de los datos.



Nota. Proceso de selección de documentos para la minería de datos.

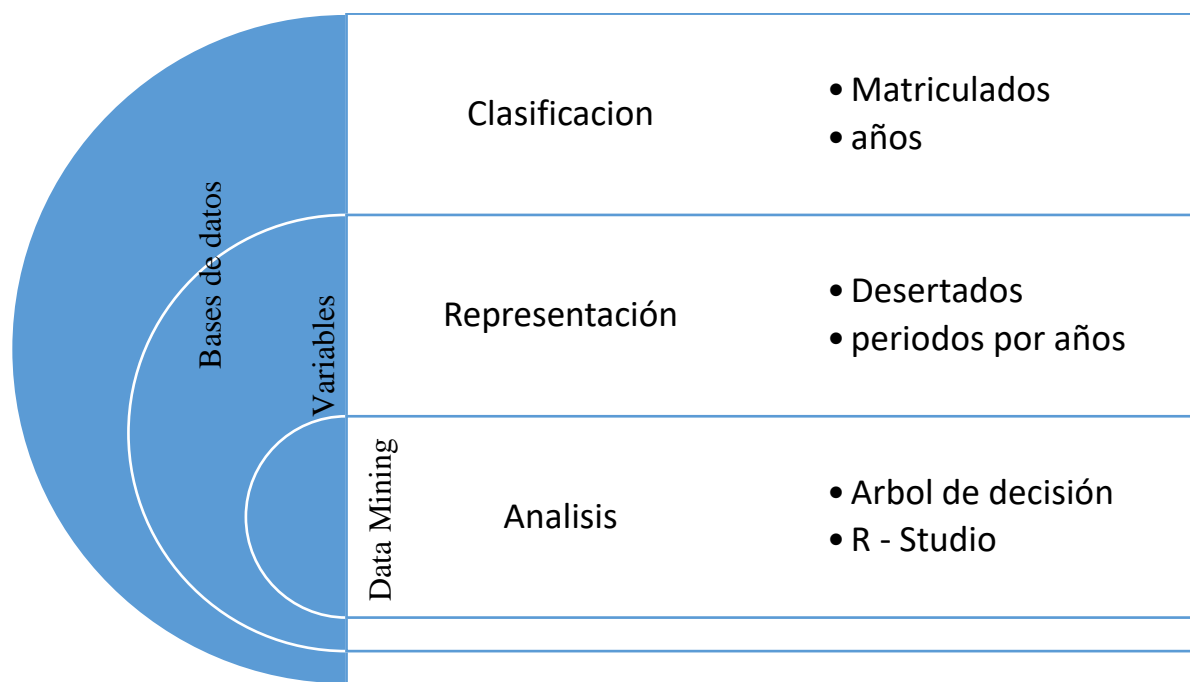
De igual manera, se plantea el modelo de gestión TI apoyado en la minería de datos para predecir la deserción mediante la selección y organización de documentos y variables determinadas en el análisis que esbozaron la predicción de la deserción como fuente primaria. La

minería de datos evoca la exploración de datos de grandes almacenamientos para descubrir el fenómeno de la deserción en la Institución Educativa de Coyongal. La técnica se empleó estadísticamente de manera automatizada con patrones, atributos y variables que arrojan una sola fuente como el nivel de deserción escolar.

Para validar la minería de datos al ámbito de gestión TI se analiza previamente el uso de patrones que sirvieron de búsqueda para realizar el análisis predictivo de la deserción escolar.

Figura 14

Análisis predictivo de la deserción escolar mediante patrones, atributos y variables.



Nota. Se muestra la clasificación, representación y análisis de las variables mediante patrones y atributos.

Por tanto, la minería de datos como estrategia de la predicción de la deserción, se aplicó en la plataforma de SIMAT como fuente primaria para hallar resultados verídicos determinando un modelo de gestión TI y su ejecución, generando un modelo predictivo en Algoritmo CART de clasificación y regresión encontrando patrones de comportamientos fijando variables

independientes factibles como la matrícula y el rendimiento académico, en espera de este modelo la identificación de factores que inciden en la deserción escolar en el nivel de secundaria de la Institución Educativa de Coyongal. El presente trabajo de investigación se basa en la recolección de información desde un método mixto, como menciona Sampieri (2010) representan un conjunto de procesos sistemáticos, empíricos y críticos de investigación e implican la recolección y el análisis de datos cualitativos y cuantitativos, así como su integración y discusión conjunta, para realizar inferencias producto de toda la información recabada y lograr un mayor entendimiento del fenómeno bajo estudio.

Actividad 1 de la Primera Fase. Se realizó la selección de documentos de matrículas en el Simat diseñando categorías de análisis comparativos por cada año escolar en cada uno de los grados del nivel de secundaria. La clasificación consistió en la recopilación de la información para la minería de datos de los grados 6°, 7°, 8° y 9° con los años 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021 desarrollando el objetivo de identificar la cantidad de estudiantes matriculados por años.

Tabla 4*Cuadro de matrículas año 2014*

Cuenta de GRADO_COD	Etiquetas											Total, general	
	Etiquetas de fila	0	1	2	3	4	5	6	7	8	9		10
BOCAS DE													
GUAMAL			9	5									14
I.E. COYONGAL -													
SEDE PRINCIPAL	29	27	24	27	45	49	91	77	83	48	48	31	579
ISLA PERICO		2	2	4									8
PLAYA DE LAS													
FLORES	25	21	12	19	12	15							104
PUNTA DE													
CARTAGENA	29	17	18	20	25	20							129
ROMA	16	4	4	4									28
SANTA COITA	19	8	11	20	12	11							81
SANTA MONICA	14	13	10	12	16	20							85
SANTA PABLA	13	11	8	11	26	25							94
TOLU	17	23	30	19	19	28							136
Total general	162	135	124	136	155	168	91	77	83	48	48	31	1258

Nota. La tabla representa el número de estudiantes por grado en cada una de las sedes de la Institución arrojando así el estado de matrículas del año 2014.

La recopilación de la minería de datos teniendo en cuenta la variable de la matrícula correspondió en la organización de los grados de 6°, 7°, 8° y 9° en el año 2014 observándose en la tabla valores cuantificables que permitieron determinar un promedio porcentual que ayudó organizar de maneja secuencial patrones de análisis para determinar la deserción escolar. Se realizó la comparación en el árbol de decisiones globalizando la cantidad porcentual de los estudiantes matriculados en el año 2014.

Tabla 5*Cuadro de matrículas año 2015*

Cuenta de GRADO_COD	Etiquetas de columna													Total, general
	0	1	2	3	4	5	6	7	8	9	0	1		
BOCAS DE GUAMAL	5	7	8											20
I.E. COYONGAL - SEDE PRINCIPAL	40	33	23	29	23	1	0	0	5	0	0	9		583
ISLA PERICO				2										2
PLAYA DE LAS FLORES	20	13	18	13	14	0								88
PUNTA DE CARTAGENA	22	22	16	19	16	4								109
ROMA	6	16	3	6										31
SANTA COITA	10	15	14	6	14	3								62
SANTA MONICA	22	10	17	9	18	9								85
SANTA PABLA	17	21	19	12	17									86
TOLU	24	23	20	25	18	6								126
Total general	166	0	8	1	0	3	0	0	5	0	0	9		1192

Nota. La tabla representa el número de estudiantes por grado en cada una de las sedes de la Institución arrojando así el estado de matrículas del año 2015.

La estadística del cuadro de matrícula del año 2015 muestra diferencias de cantidad de estudiantes matriculados por cuanto se decidió conglomerar secuencialmente la técnica de minería de datos y establecer comparaciones con el total de matrícula con el año anterior. Se notó también el aumento de matrícula en el grado sexto en la tabla de este año.

Tabla 6*Cuadro de matrículas año 2016*

Cuenta de GRADO_COD	Etiquetas de columna												Total general	
	0	1	2	3	4	5	6	7	8	9	0	1		
BOCAS DE GUAMAL	1	3	5	2	1									12
I.E. COYONGAL - SEDE PRINCIPAL							8	4	4	6	4	2		
PLAYA DE LAS FLORES	20	28	20	45	30	31	1	7	5	0	4	9		480
PUNTA DE CARTAGENA	9	15	14	11	14	20								83
ROMA	12	19	18	24	16	11								100
SANTA COITA	1	5	1	8	4									19
SANTA MONICA	4	16	7	13	9	7								56
SANTA PABLA	8	16	16	12	8	12								72
TOLU	8	16	15	9	13	9								70
	27	22	28	20	24	15								136
Total, general	90	0	4	4	9	5	1	7	5	0	4	9		1028

Nota. La tabla representa el número de estudiantes por grado en cada una de las sedes de la Institución arrojando así el estado de matrículas del año 2016.

Se interpretó en la tabla 3, la cual representó la cantidad de estudiantes matriculados en el año 2016 que bajó el promedio de matrícula en el grado sexto por cuanto se realiza el análisis organizando variables y realizando la comparación con los años anteriores se determinó la minería de datos en la variable de matrícula para el propósito planteado en esta investigación que se trató de determinar el nivel deserción.

Tabla 7*Cuadro de matrículas año 2017*

Cuenta de GRADO_COD	Etiquetas de												Total, general
	columna	0	1	2	3	4	5	6	7	8	9	0	
Etiquetas de fila	0	1	2	3	4	5	6	7	8	9	0	1	
BOCAS DE GUAMAL	4	4	3	5									16
I.E. COYONGAL - SEDE PRINCIPAL	23	23	24	24	47	33	4	0	1	8	3	4	504
PLAYA DE LAS FLORES	8	13	18	14	11	7							71
PUNTA DE CARTAGENA	17	11	13	18	20	16							95
ROMA	5	3	6	1	8	8							31
SANTA COITA	5	12	11	7	14	7							56
SANTA MONICA	8	12	18	14	14	6							72
SANTA PABLA	10	8	13	15	11	9							66
TOLU	21	24	26	32	24	24							151
		11	13	13	14	11	8	7	4	3	5	4	
Total general	101	0	2	0	9	0	4	0	1	8	3	4	1062

Nota. La tabla representa el número de estudiantes por grado en cada una de las sedes de la Institución arrojando así el estado de matrículas del año 2017.

Para el uso del análisis fue pertinente organizar la información en un conjunto categorizado de acuerdo a la reducción de números de matrículas interpretando características y segmentos relacionados con la deserción. Para el año 2017 se reflejó diferencias en comparación con los años anteriores en cuanto a la matrícula de los grados 6°, 7°, 8° y 9°.

Tabla 8*Cuadro de matrículas año 2018*

Cuenta de GRADO_COD	Etiquetas de columna											Total		
	0	1	2	3	4	5	6	7	8	9	0	1	1	1
Etiquetas de fila														
BOCAS DE GUAMAL	3	2	4	4										13
I.E. COYONGAL - SEDE PRINCIPAL	38	26	24	25	22	52	4	3	1	6	5	2		518
PLAYA DE LAS FLORES	9	7	15	15	10	12								68
PUNTA DE CARTAGENA	11	20	8	11	22	17								89
ROMA	3	5	5	4	1									18
SANTA COITA	17	8	9	7	6	10								57
SANTA MONICA	18	8	9	14	18	9								76
SANTA PABLA	11	11	7	13	10	8								60
TOLU	22	20	20	23	28	19								132
		10	10	11	11	12	9	8	4	3	3	4		
Total general	132	7	1	6	7	7	4	3	1	6	5	2		1031

Nota. La tabla representa el número de estudiantes por grado en cada una de las sedes de la Institución arrojando así el estado de matrículas del año 2018.

La comprensión de los datos organizado en la minería de datos permitió el análisis del comportamiento de la matrícula de los grados 6°, 7°, 8° y 9°. En el año 2018 estadísticamente se notó una diferencia de cantidad de estudiantes en comparación con los años anteriores.

Tabla 9*Cuadro de matrículas año 2019*

Cuenta de GRADO_COD	Etiquetas											Total	
	de columna												
Etiquetas de fila	0	1	2	3	4	5	6	7	8	9	10	11	general
BOCAS DE GUAMAL	2		2	2	3								9
I.E. COYONGAL - SEDE													
PRINCIPAL	44	38	30	29	18	19	102	64	61	29	31	29	494
PLAYA DE LAS FLORES	23	11	11	12	14	9							80
PUNTA DE CARTAGENA	21	10	17	8	10	18							84
ROMA	8	5	6	5									24
SANTA COITA	9	18	8	6	11	4							56
SANTA MONICA	15	12	7	6	16	12							68
SANTA PABLA	8	11	12	5	8	10							54
TOLU	24	20	27	16	17	20							124
Total general	154	125	120	89	97	92	102	64	61	29	31	29	993

Nota. La tabla representa el número de estudiantes por grado en cada una de las sedes de la Institución arrojando así el estado de matrículas del año 2019.

La comprensión de los datos permitió modelar la información relacionada con la matrícula lo que ayudó la visualización estadística identificando la variación de cantidad de estudiantes matriculados en comparación con los años anteriores. La tabla 6 muestra una diferencia significativa entre los demás años.

Tabla 10*Cuadro de matrículas año 2020*

Cuenta de GRADO_COD	Etiquetas de columna											Total general	
	0	1	2	3	4	5	6	7	8	9	0		1
I.E. COYONGAL - SEDE PRINCIPAL	21	48	35	38	3	0	3	5	9	1	4	5	532
PLAYA DE LAS FLORES	18	22	13	17	9	1							90
PUNTA DE CARTAGENA	21	18	6	15	8	0							78
ROMA	6	6	6	6	7	3							34
SANTA COITA	10	11	18	7	7	9							62
SANTA MONICA	5	16	8	6	5	5							55
SANTA PABLA	10	9	12	13	4	0							58
TOLU	15	21	25	23	9	5							118
Total general	106	151	121	128	89	99	99	59	52	22	25	25	1027

Nota. La tabla representa el número de estudiantes por grado en cada una de las sedes de la Institución arrojando así el estado de matrículas del año 2020.

La cantidad de información utilizada permitió la clasificación documental para realizar un análisis globalizado, tomando el histórico de matrículas de años anteriores a los caracterizados para el estudio, estableciendo diferencias entre los años inmediatamente anterior.

El análisis documental se determinó como fuente primaria porque está relacionada en términos de tiempo y espacio con el evento de la deserción teniendo en cuenta el comportamiento y ocurrencia en el periodo académico anual para luego, interpretar y predecir en un futuro. De este modo, se realizó el procedimiento de los algoritmos CART permitieron la descripción de los datos en la investigación usando la estrategia de la minería de datos, se relacionaron los datos con características relevantes a la deserción y recopilados en la plataforma del SIMAT en cuanto a la matrícula. En este caso, se organizaron en tablas para la graduación porcentual considerando a los estudiantes desertores para ejecutar el árbol de decisiones usando el modelo secuencial de parámetros aleatorios.

Figura 15

Resumen histórico de la matrícula en comparación con años anteriores

SEDES	2011	2012	2013	2014	2015	2016	2017	2018	2019
MATRICULA	1258	1192	1121	1097	1028	1062	1031	993	1027
DIF. CON EL AÑO ANTERIOR (%)		-5%	-6%	-2%	-6%	3%	-3%	-4%	3%
DIF. CON EL AÑO 2011 (%)		-5%	-11%	-13%	-18%	-16%	-18%	-21%	-18%

Nota. Se muestra en valor porcentual el histórico de matrículas por años.

Figura 16

Clasificación de la variable matrícula por años



Nota. Se muestra la cantidad de estudiantes matriculados por años en la INSTEC

Esta actividad esbozó el planteamiento del modelo de gestión TI desde la perspectiva metodológica mediante la identificación de las buenas prácticas que soportan la permanencia en la Institución Educativa de Coyongal, aclarando que las buenas prácticas tratan de todas las estrategias aplicadas institucionalmente para que los estudiantes superen sus debilidades tanto académicas como motivacionales durante el periodo académico. No obstante, se encontró en el análisis un factor determinante tratándose del alto nivel de deserción escolar en periodos anuales. Luego, basados en el resultado que alertó de manera preocupante en la investigación se enfocó en el comportamiento de la matrícula por cada año utilizada como estrategia de minería de datos usando métodos basados en la organización de la biblioteca para la gestión del Servicio de TI. En esta biblioteca (IT Infrastructure Library) permitió construir un proceso de análisis de alta calidad confrontando los indicadores de desempeño desde la perspectiva de la matrícula en la institución. Como resultado se pudo observar variación comportamental en la matrícula, lo que conllevó a utilizar otras variables como minería de datos.

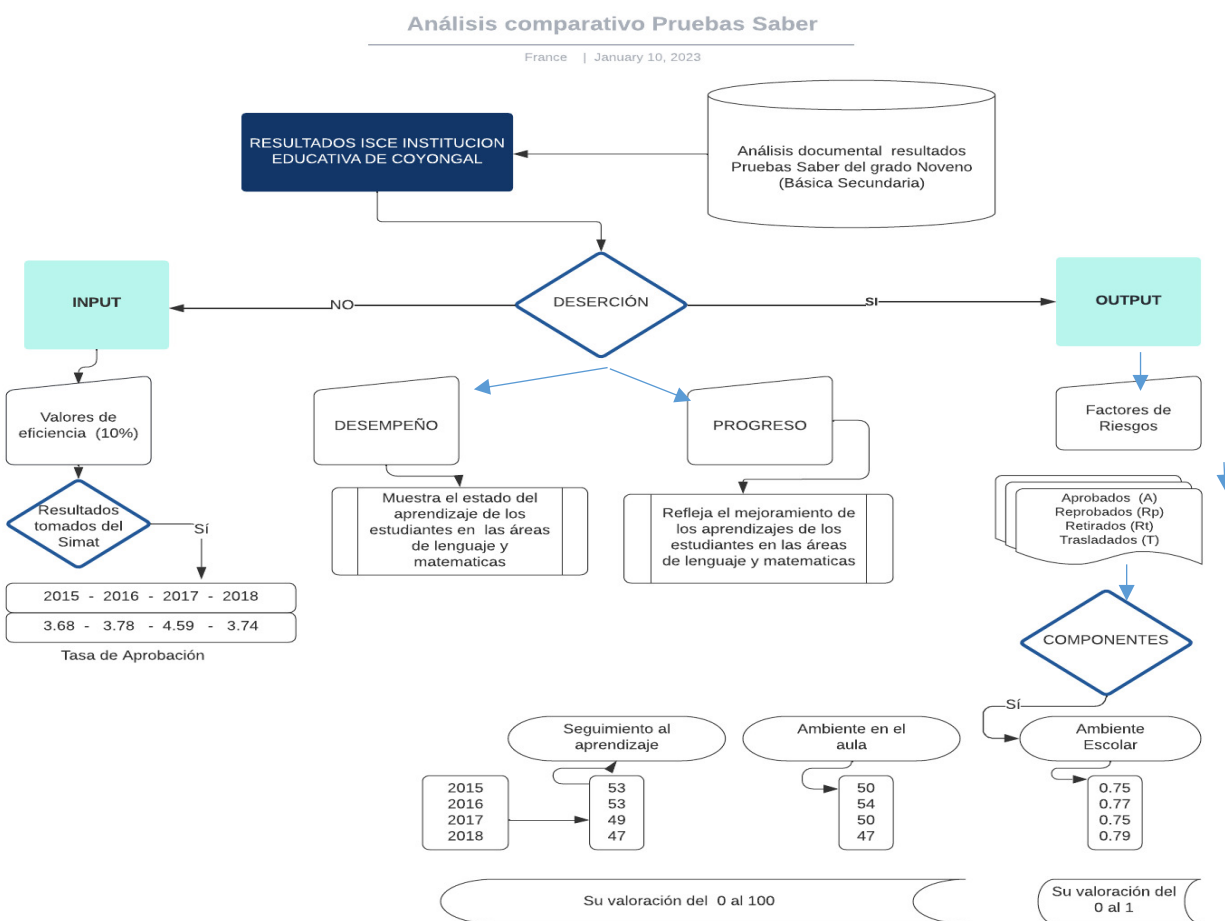
Actividad 2 de la primera fase. Aplicación del análisis comparativo de los resultados de las pruebas saber entre los años 2015 a 2020 de los grados 5° y 9°.

La investigación histórica ayudó a clasificar para luego comparar los resultados del Índice Sintético de Calidad de cada uno de los años en el que se implementó la Prueba Saber en los grados 3°, 5° y 9°. En primer lugar, se acudió a realizar el análisis documental mediante el reporte de la excelencia (ISCE), de los años 2015, 2016, 2017 y 2018 seleccionando la variable cuantitativa de cada uno de los cuatro componentes que componen esta evaluación. En segundo lugar, se ha tenido en cuenta que el Índice Sintético de la Calidad Educativa (ISCE) es un indicador numérico que se enmarca entre la escala de 1 a 10 el cual es entregado por cada nivel educativo (primaria, secundaria y media) a todos los Establecimientos Educativos (EE) del país desde el año 2015. Este índice es calculado por el Instituto Colombiano para la Evaluación de la Educación (ICFES) a partir de los resultados de las pruebas Saber y la eficiencia interna de los Establecimientos Educativos, y es divulgado por el Ministerio de Educación Nacional, (MEN, 2022).

Desde el punto de vista histórico se interpretó en un grado de porcentaje estadístico el valor de cada componente teniendo en cuenta la variable que conlleva al alto nivel de deserción en la Institución Educativa de Coyongal. A partir de los valores encontrados por cada año y comparados en cada uno de los grados 3°, 5° y 9° se encontró la necesidad de mejorar la calidad educativa institucional en los componentes de desempeño y ambiente escolar. Como se muestra en la siguiente figura:

Figura 17

Análisis comparativo de los resultados de las Pruebas Saber. Aplicando Algoritmos CART.



Fuente. Propia del investigador.

El modelo de gestión TI mediante algoritmos CART arrojó resultados literales en los que se pueden determinar las variables incidentes en el alto nivel de deserción como son los factores de riesgos. Se comparó el valor con cada uno de los años teniendo en cuenta cada componente con la tasa de aprobación y se encontró que el análisis documental ha tenido comportamientos diferentes por cada año, sin embargo, la estrategia de minería de datos mostró que la cantidad de información analizada debe ser complementada con más historia documental para el análisis.

Fase 2. Aplicación de la Técnica Algoritmos CART y Árboles de Decisiones.

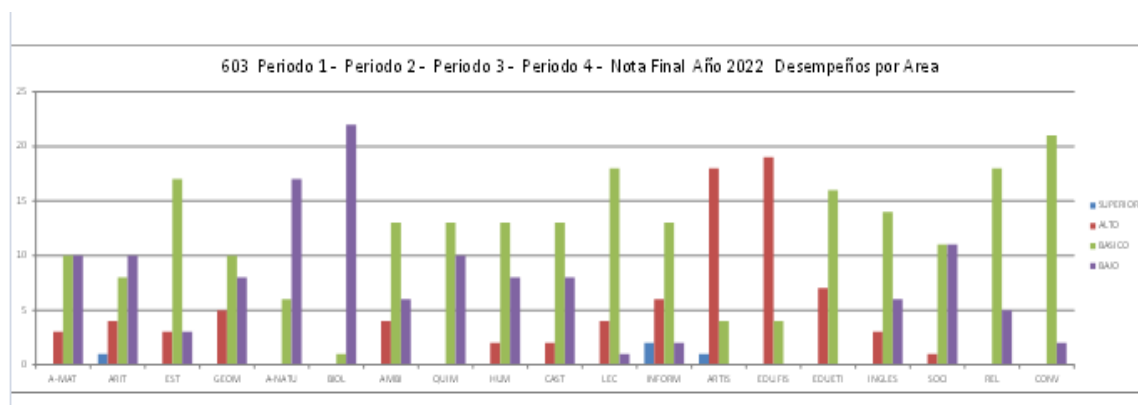
En esta segunda fase se desarrolla el segundo objetivo determinando la técnica de algoritmo CART para el análisis estadístico en R-Studio, aplicando la técnica de árboles de decisiones y utilizando la minería de datos como estrategia para predecir la deserción con las variables: rendimiento académico en PUNTO EDU, Pruebas Icfes Saber 11 y la autoevaluación institucional en los últimos 6 años (2014 a 2020) en los grupos del nivel de secundaria. La minería de datos consistió en la recopilación de la información adquirida en el SIMAT, clasificando la cantidad de estudiantes matriculados y luego realizando la comparación con los estudiantes aprobados, reprobados, desertados y retirados para predecir la deserción en los grados del nivel de secundaria.

La estrategia permitió la estructuración del modelo de gestión TI compilando una estructura jerárquicamente en cinco procesos: iniciación, planeación, ejecución, monitoreo, control y cierre caracterizando variables relacionadas con el rendimiento académico. En esta primera actividad de la fase 2, se realizó el proceso de iniciación; se accedió a la plataforma Punto Edu descargando los informes académicos por periodos de cada estudiante correspondientes a los grados de 6°, 7°, 8° y 9°. Luego se diseñó la biblioteca de los promedios de notas por estudiantes teniendo en cuenta la valoración de cada área de aprendizaje.

Figura 18

Lista de notas por periodo como estrategia de minería de datos

INSTITUCION EDUCATIVA DE COYONGAL																																							
MAGANGUE Bolívar																																							
DESEMPEÑOS POR AREAS DETALLADOS POR GRUPO																																							
Periodo 1 - Periodo 2 - Periodo 3 - Periodo 4 - Nota Final AÑO: 2022																																							
GRUPO: 603 - I.E. COYONGAL SEDE PRINCIPAL - 23 ESTUDIANTES																																							
VALORACION	A-MAT		ARIT		GEOM				A-NATU		BIOL		AMBI		QUIM		HUM		CAST		LEC		INFORM		ARTIS		EDU.FIS		EDU.ETI		INGLES		SOCI		REL		CONV		
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	
SUPERIOR	0	0	1	4,3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	8,7	1	4,3	0	0	0	0	0	0	0	0	0	0	0	0		
ALTO	3	13	4	17	3	13	5	22	0	0	0	0	4	17	0	0	2	8,7	2	8,7	4	17	6	26	18	78	19	83	7	30	3	13	1	4,3	0	0	0		
BASICO	10	44	8	35	17	74	10	44	6	26	1	4,3	13	57	13	57	13	57	13	57	18	78	13	57	4	17	4	17	16	70	14	61	11	48	18	78	21	91	
BAJO	10	44	10	44	3	13	8	35	17	74	22	96	6	26	10	44	8	35	8	35	1	4,3	2	8,7	0	0	0	0	0	0	0	6	26	11	48	5	22	2	8,7



Nota. La figura muestra gráficamente el puntaje de la valoración del estudiante de los cuatro periodos del año lectivo.

De igual manera, en el proceso de planeación se gestionó el alcance de notas de cada área por periodo realizando la comparación de la valoración promedio por cada periodo, siendo que la Institución Educativa de Coyongal estructura dicha valoración en 60 puntos por cada periodo del año superando cada estudiante cuatro (4) periodo en el año. Es decir, el estudiante debe alcanzar 240 punto en cada área superando evaluaciones de tipo cualitativa y cuantitativa según lo muestra la figura. Para determinar el nivel de deserción se utilizó el proceso de Algoritmos CART en el programa R-Studio.

Figura 19

Planillas de notas en físico en los años 2014, 2015, 2016, 2017, 2018, 2019 y 2020

		MAGANGUE Bolívar																																										
		DESEMPEÑO POR ASIGNATURA EN CADA GRADO																																										
		PRIMERO			SEGUNDO			TERCERO			CUARTO			QUINTO			SEXTO			SEPTIMO			OCTAVO			NOVENO			TRANSICION															
		DS	DA	DB	DS	DA	DB	DS	DA	DB	DS	DA	DB	DS	DA	DB	DS	DA	DB	DS	DA	DB	DS	DA	DB	DS	DA	DB	DS	DA	DB	DS	DA	DB	DS	DA	DB							
7	EDUCACIÓN ARTÍSTICA	4	29	59	3	9	37	52	9	18	28	62	4	24	44	57	5	7	43	63	32	51	23	17	30	34	2	14	22	7														
8	CASTELLANO	6	20	48	21	9	8	65	25	9	12	59	32	19	19	68	24	7	25	59	22	7	19	41	16	1	11	26	2	2	8	38	15	7	22	17								
9	EDUCACIÓN FÍSICA	2	63	26	3	11	49	42	5	38	37	34	3	33	19	74	4	21	65	27	77	5	1	2	12	23	3	8	53	5	2	42	2											
10	EDUCACIÓN ÉTICA Y VALORES	9	41	42	3	14	30	58	5	6	25	77	4	18	32	76	4	13	53	47	18	64	1	7	33	2	6	54	1	7	27	6	4											
11	INFORMÁTICA	4	25	63	3	3	24	64	17	2	10	92	8	17	21	84	8	5	21	87	6	58	17	21	11	19	10	6	48	11	2	22	21	1										
12	INGLÉS	3	14	73	5	4	20	62	21	4	15	67	26	9	14	96	11	3	20	69	21	12	21	30	20	3	8	24	5	1	21	37	7	2	12	26	6							
13	MATEMÁTICAS	11	14	53	17	12	11	61	23	6	20	61	25	19	22	74	15	9	33	58	13																							
14	NATURALES	6	20	57	12	4	19	67	17	7	24	60	21	18	12	86	14	8	23	68	14																							
15	EDUCACION RELIGIOSA	5	27	60	3	29	17	55	6	2	26	74	10	18	29	78	5	18	49	46	7	66	10	2	30	8	16	48	1	1	13	25	5											
16	SOCIALES	6	18	67	4	6	16	72	13	4	13	75	20	24	18	74	14	5	26	65	17	17	36	30	2	36	2	11	54	1	1	26	18	1										
17	CONVIVENCIA	37	43	14	2	43	13	51	17	48	39	8	41	24	65	58	48	7	22	22	31	8	2	32	6	21	3	33	8	21	18	4	3	36	31	20	1							
18	DIMENSION COGNITIVA				1																																	9	38	34	7			
19	DIMENSION COMUNICATIVA				1																																	12	34	33	9			
20	DIMENSION ESPIRITUAL				1																																	14	36	33	5			
21	DIMENSION ESTETICA				1																																	11	45	30	2			
22	DIMENSION SOCIOAFECTIVA				1																																	10	32	41	5			
23	DIMENSION CORPORAL				1																																	12	47	28	1			
24	DIMENSION ETICA				1																																	10	42	31	5			
25	BIOLOGÍA																					4	18	44	17	1	24	15	43	29	26	19												
26	ED. AMBIENTAL																					18	62	3	9	25	6	10	17	38	1	2	16	25	3									
27	FÍSICA																					1	16	20	1	1	3	36	26	2	1	42												
28	QUÍMICA																					14	42	27	2	17	21	1	7	36	22	4	26	16										
29	ALGEBRA																													3	7	41	11	9	32	5								
30	GEOMETRÍA																					4	11	23	45	1	5	28	6	6	52	8	3	34	9									
31	ESTADÍSTICA																					6	24	40	13	4	18	2	16	2	7	44	13	3	33	10								
32	ARITMÉTICA																					18	19	14	32	5	4	26	5	3	1													
33	LECTURA CRÍTICA																					22	26	23	11	2																		

Nota. Resumen de la planilla de notas ´por estudiantes en cada año lectivo.

En el proceso de ejecución se administraron herramientas necesarias alineando la misión, visión y objetivos institucionales desde la medida de factores cualitativos influyentes en la deserción, codificando cada uno de estos factores: estudiantes matriculados, aprobados, reprobados, retirados y desertado en el periodo lectivo anual. Para el análisis la estrategia de minería de datos permitió acudir a los archivos de lista de asistencia de los estudiantes. Estos documentos se encuentran firmados por cada docente responsable de la asignatura que practica en la institución.

Para complementar la minera de datos del rendimiento académico se calculó teniendo en cuenta la frecuencia del periodo anual el porcentaje de asistencia de cada estudiante en los grados de sexto, séptimo, octavo y noveno lo que permitió predecir el posible abandono escolar antes de culminar el periodo anual.

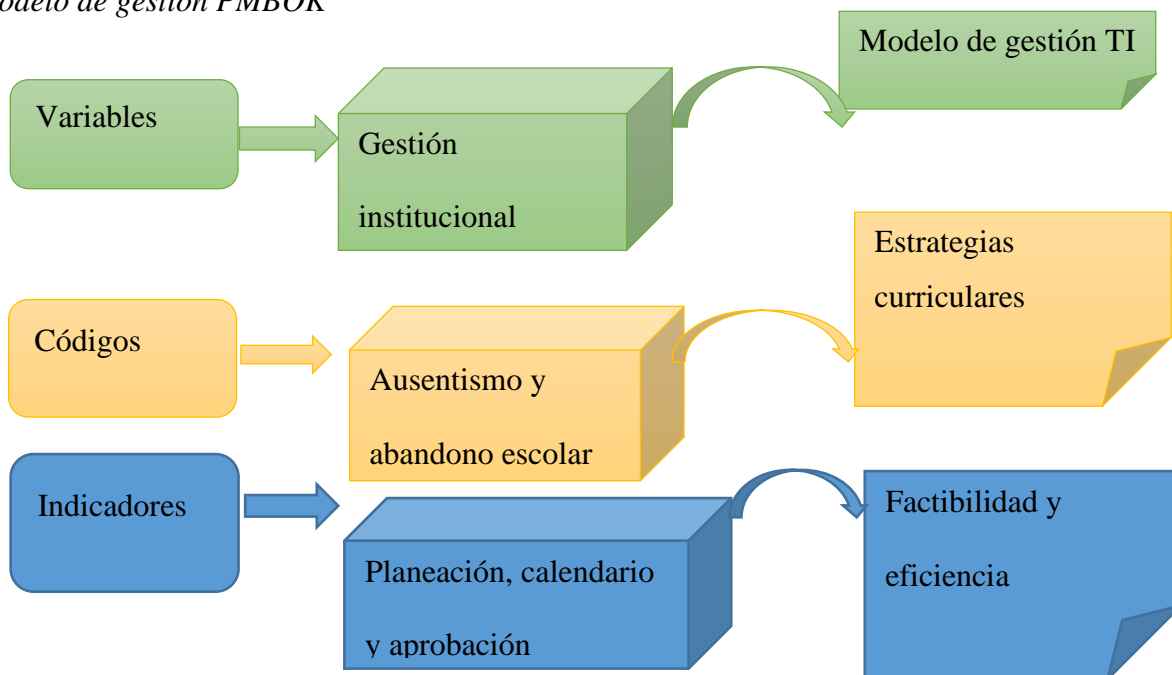
Figura 20

Planillas de asistencia en físico y actualizada en Punto Edu

	Apellidos y Nombre	Tipo	Horas	Hora
1	ALVAREZ JAQUE MARIANA LUCIA	Asistió		
2	ALVAREZ JAQUE MIRIAN EDITH	Asistió		
3	ALVAREZ RUZ JHOVY JOSE	Asistió		
4	ALVAREZ RUZ OSCAR JOSE	Asistió		
5	ARAQUE OBREGON ESTEBAN	Asistió		
6	BELÉIDO AREVALO JOSE ISAAC	Asistió		
7	CABARCAS RAMOS DIOMEDES MANUEL	Asistió		
8	DONADO REYES DIANY LUZ	Asistió		
9	DONADO VELLILA IVAN DARIO	Asistió		
10	GARAY VARGAS ANGEL MAURICIO	Asistió		
11	MEJIA BELLO JOAN SEBASTIAN	Asistió		
12	MENDOZA PAVARES SANTIAGO	Asistió		

Nota. Se muestra el histórico de la planilla de asistencia en Punto Edu.

En el proceso de monitoreo, se confrontaron los procesos de aprendizaje identificando eventos pasados para predecir tendencias futuras de deserción escolar. El modelo de gestión TI alineó procesos estructurados de planeación curricular, recursos físicos, humanos y financieros y de esta manera conllevó a la definición de roles y responsabilidades administrativas institucional.

Figura 21*Modelo de gestión PMBOK*

Nota. Proceso de monitoreo del modelo de gestión TI

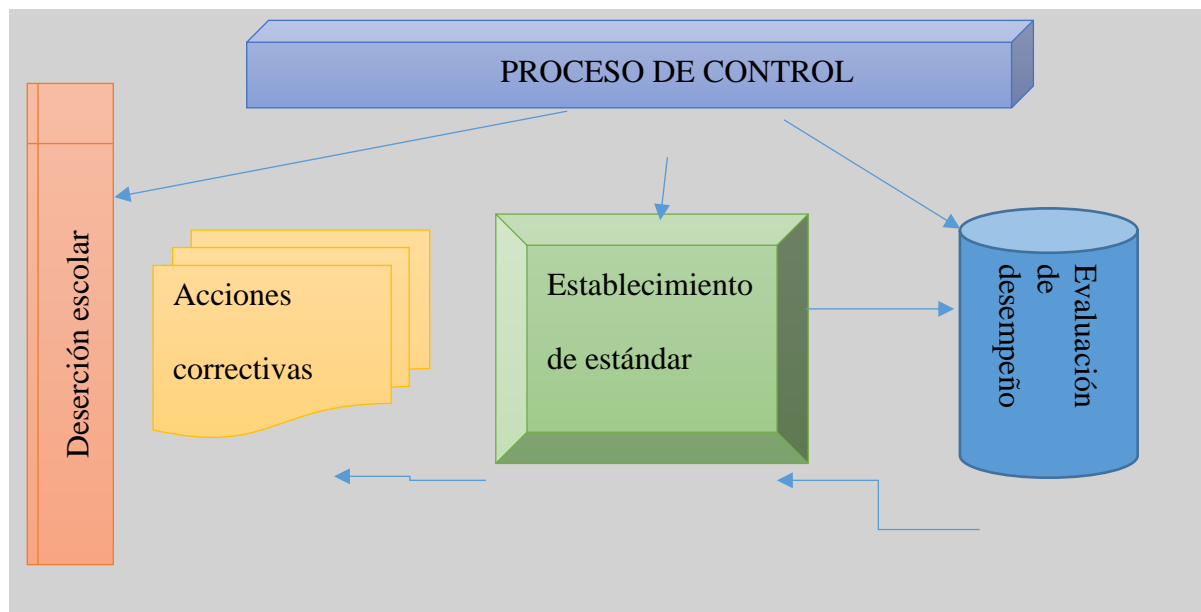
El proceso de **control** se esbozó la variable del rendimiento académico desde la perspectiva estratégica de la minería de datos siguiendo el proceso de planificación y recepción de documentos teniendo en cuenta los criterios de evaluación y la medición de los resultados conforme a la ejecución de los factores incidentes en la deserción escolar. Por ello, se estructuró la predicción desde la determinación secuencial de cada uno de los aspectos relacionados con el rendimiento académico.

En la comprensión de los datos se identificó el problema de la calidad de los procesos educativos en un conjunto de datos para formular hipótesis específicas convenientes del rendimiento académico. La librería de datos en esta variable se compone de cuatro archivos por cada año correspondiente a cada periodo escolar. Cada archivo contiene las sábanas de notas de cada estudiante por grado desde sexto a noveno determinando el área con promedio más bajo y

posibles de alcanzar el próximo periodo. Por esta razón, se gestiona el flujo de procesos PMBOK.

Figura 22

Flujo de procesos de gestión PMBOK



Nota. Comprensión de datos en el proceso de control de la deserción escolar.

El rendimiento académico se consideró como una variable determinante para la predicción de la deserción por cuanto estableció componentes descriptivos que permitió identificar al posible estudiante frágil o vulnerable a la deserción en el tercer periodo del año escolar. Los procesos de seguimiento educativo generaron el monitoreo de desempeño y ambientes educativos influyentes en el fracaso escolar. En este sentido, el estado actual de los procesos que predicen la deserción escolar, se aplicó la implementación de componentes concibiendo un diseño de gestión de servicio TI, creando un listado de reconocimiento en los procesos estandarizados institucionalmente.

La actividad 2 de la fase 2 consistió en la ejecución analítica del formato de eficiencia interna en el programa de Excel. El sistema educativo de la INSTEC establece cada año la promoción de la población estudiantil matriculada de un grado a otro manteniendo la fluidez y sostener el proceso de escolar el periodo lectivo anual. A esta capacidad demostrada de retención de la población en permanencia escolar se le denomina eficiencia interna. De hecho, la dimensión de análisis se enfocó principalmente en la medición del flujo de alumnos matriculados, con el fin de posibilitar un diagnóstico del mismo en términos de eficiencia, como metas del análisis predictivo en la deserción escolar (IDECE, s.f.).

De este modo, se realizó el análisis basado en tres componentes denominados indicadores incluyentes en el sistema analítico con algoritmos CART, mediante e entorno integrado en R-Studio: la tasa de repitencia, la tasa de promoción efectiva y la tasa de abandono interanual. Esto facilitó la predicción de la promoción y la deserción. Para esbozar directamente al análisis estadístico se crea un archivo con el documento de eficiencia interna de cada año inmediatamente anterior, 2014 a 2019.

Figura 23

Archivo en Excel como base de datos de eficiencia interna por cada año.

	B	D	F	G	H	I	J	K	L	M	N	O	P	Q	R
1															
2	1: PROMOCIÓN AÑO ANTERIOR SECTOR OFICIAL (INSTITUCIONES OFICIALES)														
3	MUN_CODIC	CODIGO_DANE_SE DE	TIPO_JORNAL	CARÁCT	GRAD	NUM_HOMBRES APROBADO	NUM_MUJERES APROBADO	NUM_HOMBRES EROBADO	NUM_MUJERES EROBADO	NUM_HOMBRES DESERTARCI	NUM_MUJERES DESERTARCI	NUM_HOMBRES TRASLADAI	NUM_MUJERES TRASLADAI	TOT	NOMBRE SEDE
4	430	213430002337	2	1	0	16	25	0	0	3	3	0	0	47	COYONGAL
5	430	213430002337	2	1	1	18	28	2	0	2	0	0	0	50	COYONGAL
6	430	213430002337	2	1	2	21	8	1	0	0	0	0	0	30	COYONGAL
7	430	213430002337	2	1	3	9	13	3	3	1	0	0	0	29	COYONGAL
8	430	213430002337	2	1	4	6	9	1	0	4	2	0	0	22	COYONGAL
9	430	213430002337	2	1	5	10	9	0	0	0	0	0	0	19	COYONGAL
10	430	213430002337	2	1	6	38	38	17	2	4	0	0	0	99	COYONGAL
11	430	213430002337	2	1	7	24	22	7	4	5	2	0	0	64	COYONGAL
12	430	213430002337	2	1	8	23	20	12	4	0	2	0	0	61	COYONGAL
13	430	213430002337	2	1	9	6	17	5	5	0	0	0	0	33	COYONGAL
14	430	213430002337	2	1	10	12	11	5	3	0	0	0	0	31	COYONGAL
15	430	213430002337	2	1	11	16	14	0	0	0	0	0	0	30	COYONGAL
16	430	213430002337	2	1	23									0	COYONGAL
17	430	213430002337	2	1	24									0	COYONGAL
18	430	213430002337	2	1	25									0	COYONGAL
19	430	213430002337	2	1	26									0	COYONGAL
20	430	213430001900	2	1	0	1	1	0	0	0	0	0	0	2	BOCAS DE GUAMAL
21	430	213430001900	2	1	1	0	0	0	0	0	0	0	0	0	BOCAS DE GUAMAL
22	430	213430001900	2	1	2	1	1	0	0	0	0	0	0	2	BOCAS DE GUAMAL
23	430	213430001900	2	1	3	1	1	0	0	0	0	0	0	2	BOCAS DE GUAMAL
24	430	213430001900	2	1	4	1	2	0	0	0	0	0	0	3	BOCAS DE GUAMAL
25	430	213430000539	2	1	0	8	11	0	0	4	1	0	0	24	PUNTA DE CARTAGENA
26	430	213430000539	2	1	1	4	5	1	2	0	0	0	0	12	PUNTA DE CARTAGENA
27	430	213430000539	2	1	2	8	9	0	0	0	0	0	0	17	PUNTA DE CARTAGENA
28	430	213430000539	2	1	3	6	3	0	1	0	0	0	0	10	PUNTA DE CARTAGENA
29	430	213430000539	2	1	4	4	7	0	0	0	0	0	0	11	PUNTA DE CARTAGENA
30	430	213430000539	2	1	5	10	6	2	1	0	0	0	0	19	PUNTA DE CARTAGENA
31	430	213430002400	2	1	0	11	10	0	0	1	2	0	0	24	PLAYA DE LAS FLORES
32	430	213430002400	2	1	1	3	6	1	1	0	0	0	0	11	PLAYA DE LAS FLORES

Nota. Comparación de las variables matriculas con estudiantes promovidos

De manera secuencial se hizo referencia a la proporción de estudiantes matriculados que cumplieron con los requisitos académicos exigidos y lograron aprobar el grado matriculado y luego pueden ser promovidos al siguiente grado. Se hizo relación de variables de eficiencia interna midiendo en las formas: cualitativa y cuantitativa, dos sucesos para respaldar la toma de decisiones para predecir la deserción.

Tabla 11

Los indicadores con relación a la variable de la eficiencia interna

Indicadores de eficiencia interna					
Tasa de aprobación	ta_n	Matriculados M_n	n = grado $n_1 =$ Sexto $n_2 =$ Séptimo $n_3 =$ Octavo $n_4 =$ Noveno	$A_n =$ Número de estudiantes aprobados en el grado matriculado	$M_n =$ cantidad de estudiantes matriculados en el grado.
Tasa de repitencia	$TRep_n$	Tasa de repitencia en el grado matriculado	n = grado $n_1 =$ Sexto $n_2 =$ Séptimo $n_3 =$ Octavo $n_4 =$ Noveno	$Rep_n =$ Número de estudiantes repitentes en el grado matriculado	$M_n =$ cantidad de estudiantes matriculados en el grado.
Tasa de reprobación	TR_n	Tasa de Reprobados	n = grado $n_1 =$ Sexto $n_2 =$ Séptimo $n_3 =$ Octavo $n_4 =$ Noveno	$R_n =$ Número de estudiantes que reprobaron	$M_n =$ cantidad de estudiantes matriculados en el grado
Tasa de abandono y/o deserción	ta_n	Tasa de desertores Interanual TDI_n	n = grado $n_1 =$ Sexto $n_2 =$ Séptimo $n_3 =$ Octavo $n_4 =$ Noveno	$DI_n =$ Número de estudiantes que abandonan el sistema escolar antes de terminar el año lectivo (Desertores).	$M_n =$ cantidad de estudiantes matriculados en el grado.
Frecuencia	Anual				

Nota. La relaciona la formula adaptada para hallar el promedio de los indicadores de la eficiencia interna de la INSTEC.

Los indicadores de la eficiencia interna clasificaron porcentualmente a partir de la frecuencia anual la cantidad de estudiantes aprobados en el grado matriculado según documentos del SIMAT, estudiantes repitentes y estudiantes reprobados. Con estos indicadores se calcula la cantidad de estudiantes posibles de abandono escolar en la frecuencia interanual determinando el porcentaje de estudiantes desertores. Para ello, se dividió el número de estudiantes desertores de cada uno de los grados sexto, séptimo, octavo y noveno entre la cantidad de estudiantes matriculados en cada grado del mismo año y multiplicado por 100. Esta medida porcentual hallada, toma valores de intervalos de 0 y 1 permitiendo diseñar el modelo de gestión TI a partir de la minería de datos de la eficiencia interna. Se identificaron las necesidades metodológicas cualitativas interpretando la proporción de estudiantes posibles de desertar en el grado calculado.

Los datos generados de las pruebas Icfes Saber 11° de los años 2014 a 2020 generaron la minería de datos en el desarrollo de métodos de descubrimiento que permitieron comprender la importancia de las estrategias de permanencia utilizadas en la Institución Educativa de Coyongal. En la actividad 3 de la fase 2, se analizaron técnicas estadísticas para encontrar patrones de análisis automáticos para predecir fenómenos en un futuro. Este fenómeno, trató de la deserción organizando un conjunto de variables que intervinieron en el estudio. Se estructura el modelo de descubrimiento de patrones de desempeño académico con árbol de decisiones combinando los procesos de extracción de datos de estudiantes aprobados, reprobados y desertados para luego determinar los posibles estudiantes desertados en el periodo del año escolar.

Tabla 12*Patrones de análisis automáticos para predecir fenómenos en un futuro*

	2014	2015	2016	2017	2018	2019	2020
Nº DE ESTUDIANTES	47	30	44	40	30	26	54
PROMEDIO ARITMETICO	220	219	225	213	210	209	222
MAYOR PUNTAJE	278	273	309	278	289	260	290

Nota. Comparación de estudiantes matriculados y su promedio aritmético en las Pruebas Icfes

Para ello, se hizo la recopilación de la base de datos proporcionada por el Icfes sobre el rendimiento de las pruebas saber 11 en los años 2014 al 2020 con los datos descriptivos de las variables evaluadas: lectura crítica, matemáticas, sociales y ciudadanas, ciencias naturales e inglés, además de la intervención de los aspectos socioeconómicos y culturales. El registro de los datos referentes a las Pruebas Icfes permitió la unificación de un archivo para estructurar la librería tomada para el análisis algorítmica en el software Rstudio con indicadores simplificados que calcularon el promedio del nivel de desempeño de las variables evaluadas de cada uno de los estudiantes inscritos.

Tabla 13*Resultados de la Institución Educativa de las Pruebas Icfes Saber 11*

Nombre del campo	Tipo de campo	Descripción
Código- DANE	Texto	Establecimiento educativo
Inscritos	Numérica	Total de estudiantes del establecimiento educativo que participan en la evaluación.
Promedio	Numérica	Promedio de los puntajes de los estudiantes dentro del Establecimiento Educativo
Desviación	Numérica	Desviación estándar del puntaje de los estudiantes dentro del Establecimiento Educativo
Insuficiente	Numérica	Porcentaje de estudiantes en el nivel de desempeño insuficiente
Mínimo	Numérica	Porcentaje de estudiantes en el nivel de desempeño mínimo
Satisfactorio	Numérica	Porcentaje de estudiantes en el nivel de desempeño satisfactorio
Avanzado	Numérica	Porcentaje de estudiantes en el nivel de desempeño avanzado

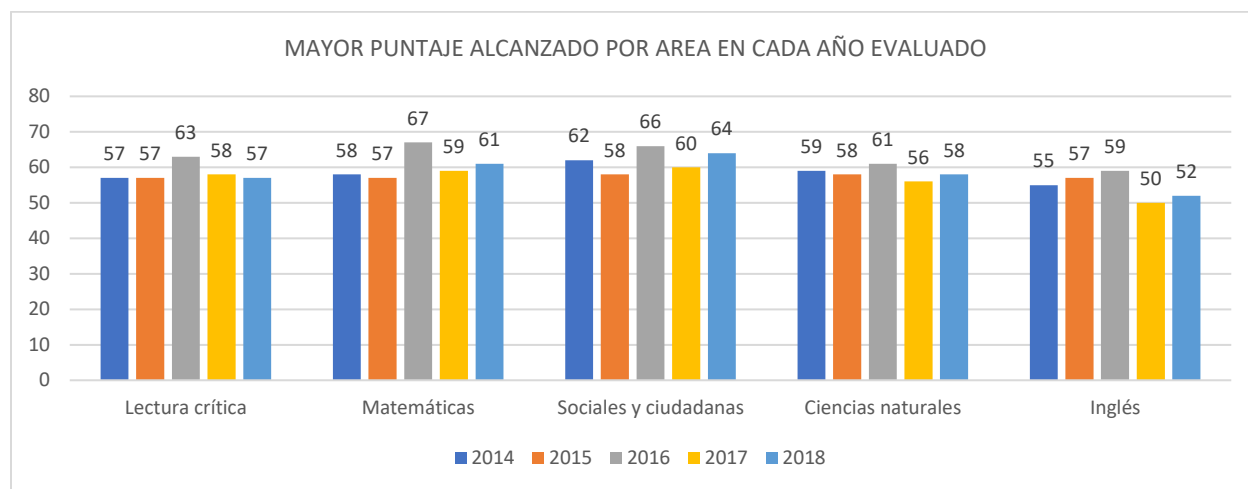
Fuente. Icfes, 2020

Encontrando de esta manera que el mayor puntaje se encuentra en un promedio variable de 63% de lectura crítica, 67% de matemáticas, 66% Sociales y ciudadanas, 61% ciencias naturales y 59% de inglés en el año 2016.

Figura 24

Comparación de los puntajes de cada competencia evaluada por año en las Pruebas Icfes Saber

11



Fuente. (MEN, 2019)

La variable de las pruebas Icfes saber 11, se detectaron atributos de datos atípicos utilizando el lenguaje de programación Rstudio con diagrama de Pareto para los atributos: lectura crítica, matemáticas, sociales y ciudadanas, ciencias naturales e inglés. Se encontraron cinco datos específicos en el año 2016 que mostraron un valor porcentual determinando el año de mayor aprobación en el desempeño académico.

Se concluye el análisis de la fase 2 utilizando variables que determinaron la técnica de algoritmos CART, se procede a la ejecución de los datos organizados y clasificados para diseñar el modelo de gestión TI mediante Rstudio. Este permitió realizar los análisis comparativos teniendo en cuenta las variables según las categorías: matriculado, aprobado, reprobado, repitente y desertado; tomando los datos de 1218 estudiantes matriculados almacenados en hojas de cálculos y luego fueron transformados a variables según los atributos clasificados de cada variable:

Tabla 14*Transformación de variables con la minería de datos*

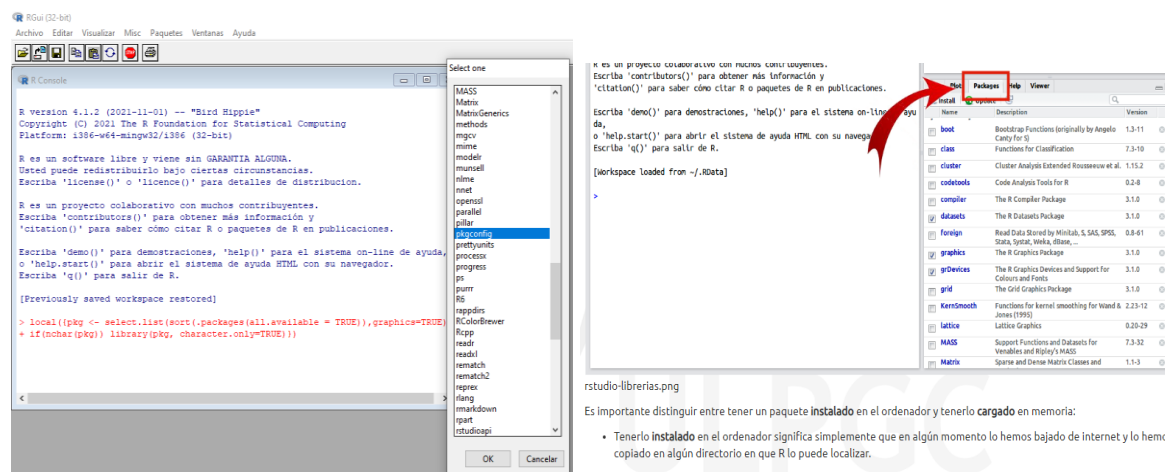
Variables	Minería de datos
Estudiantes matriculados	Plataforma del Simat Grados: sexto, séptimo, octavo y noveno
Rendimiento académico	Punto Edu: Lista de asistencia Lista de notas
Eficiencia interna	Documento en Excel por año
Pruebas Icfes Saber 11	Resultado de la evaluación por año

Nota. Clasificación y organización de la minería de datos según las variables.

De esta manera, para la generación del modelo predictivo seleccionó el árbol de decisión de manera conjunta con el algoritmo CART del lenguaje de programación Rstudio para contar con variables nominales cuantitativas y cualitativas. Se seleccionaron los datos del estudiante registrando en el archivo de la librería la cantidad de estudiantes matriculados por grado en los años de 2014 a 2020. Se aprovecharon las múltiples funciones de Rstudio importando los archivos para la realización de las transformaciones y el ajuste de las evaluaciones de los enfoques de minería de datos. Se utilizó package para colecciona las funciones de datos y códigos almacenando cada una de las carpetas conforme a las variables con cada atributo para la predicción de la deserción.

Figura 25

Creación de la librería al instalar R Studio



Nota. Al instante de instalar R Studio por defecto se incorporan numerosos paquetes que facilitan funcionalmente la ejecución del análisis.

Figura 26

Ejecución de la Librería en Rstudio

```
library()
```

Nota. Incorporación de los datos en la librería

El árbol de decisiones se diseñó el programa de Excel y luego se hizo la comparación con los resultados del árbol de decisiones de Rstudio. Se organizó la librería (plot) para estructurar las variables de la minería de datos correspondiente al modelo predictivo. Se cargan los paquetes de archivos estructurados en Excel de la eficiencia interna por que se detecta el numero de estudiantes en cada uno de los atributos: matriculados, aprobados, reprobados, retirados y desertados.

Figura 27

Modelo de evaluación mediante las propiedades `plot`, `print` y `main` en Rstudio

The screenshot shows the RStudio interface. The source editor contains the following R code:

```

1 install.packages("plot")
2 library(plot)
3 plot(final1$noveno5 ~ final1$Años , xlab= "Años", ylab = "Retirados
4     main = "Estudiantes retirados (grado noveno)", pch = 16)
5

```

The console shows the execution of several `plot` commands:

```

R 4.1.2 ~ /
> plot(final1$septimo ~ final1$Años , xlab= "Años", ylab = "Reprobados ", main = "Estudiantes reprobados(grado séptim
o)", pch = 16)
> plot(final1$octavo ~ final1$Años , xlab= "Años", ylab = "Reprobados ", main = "Estudiantes reprobados(grado octav
o)", pch = 16)
> plot(final1$noveno ~ final1$Años , xlab= "Años", ylab = "Reprobados ", main = "Estudiantes reprobados(grado noven
o)", pch = 16)
> plot(final1$sexto1 ~ final1$Años , xlab= "Años", ylab = "Desertados", main = "Estudiantes desertados(grado sexto)",
pch = 16)
> plot(final1$septimo1 ~ final1$Años , xlab= "Años", ylab = "Desertados", main = "Estudiantes desertados(grado séptim
o)", pch = 16)
> plot(final1$octavo1 ~ final1$Años , xlab= "Años", ylab = "Desertados", main = "Estudiantes desertados(grado octav
o)", pch = 16)
> plot(final1$noveno1 ~ final1$Años , xlab= "Años", ylab = "Desertados", main = "Estudiantes desertados(grado noven
o)", pch = 16)
> plot(final1$sexto2 ~ final1$Años , xlab= "Años", ylab = "Retirados ", main = "Estudiantes retirados (grado sexto)",
pch = 16)
> plot(final1$septimo3 ~ final1$Años , xlab= "Años", ylab = "Retirados ", main = "Estudiantes retirados (grado séptim
o)", pch = 16)
> plot(final1$octavo4 ~ final1$Años , xlab= "Años", ylab = "Retirados ", main = "Estudiantes retirados (grado octav
o)", pch = 16)
> plot(final1$noveno5 ~ final1$Años , xlab= "Años", ylab = "Retirados ", main = "Estudiantes retirados (grado noven
o)", pch = 16)

```

Nota. Medición y procesamiento de la variable rendimiento académico

Se utilizó la función de entrenamiento como modelo de evaluación para predecir las salidas de un conjunto de instancias con las propiedades `plot`, `print` y `main` para predecir `fit`. El modelo predictivo determinó los valores de las notas del rendimiento académico arrojados en la Plataforma Punto Edu para comparar los atributos de aprobados, reprobados, desertados y retirados. Para ello se midió el rendimiento [`repetir`] + [`preprocesamiento`]. Durante esta etapa se interpretó la variable, la cual presentó una probabilidad de estudiantes con tendencia a desertar en los grados sexto, séptimo, octavo y noveno con la clasificación en árboles de diversos paquetes en R (`rpart`, `RWeka`, `evtree`, `C50`, `party.rpart`).

Estos conceptos consisten en:

plot.train

Visualizar gráfico del modelo de predicción con diferentes ajustes de configuración `print.train`

Descripción en formato texto de los resultados

predict.train

Predecir nuevos ejemplos (por ejemplo, datos de test)

\$finalModel

Visualizar gráfico del modelo de predicción con diferentes ajustes de configuración

Después de organizada la minería de datos en la librería en R Studio se procede a evaluar los resultados predictivos arrojados en los análisis estadísticos para presentar propuesta de Modelo de Gestión Ti para predecir el nivel de deserción en los años 2021 -2025 categorizados en la siguiente fase.

Fase 3. Resultados Predictivos del Modelo de Gestión TI.

La tercera fase consistió en la ejecución del tercer objetivo el cual consistió en evaluar los resultados predictivos arrojados en los análisis estadísticos para presentar propuesta de Modelo de Gestión Ti para predecir el nivel de deserción en los años 2021 -2025. Para la elaboración de la tabla se sumó el total de estudiantes matriculados en los grados sexto, séptimo, octavo y noveno sacando el promedio anual de los años 2014, 2015, 2016, 2017, 2018, 2019 y 2020, para determinar la tasa promedio porcentual de los aprobados en los cuatro periodos del año escolar.

Tabla 15

Hallar el promedio porcentual en RStudio mediante la formula determinando la frecuencia.

Estudiantes matriculados por año					
Grado	Cantidad	Formula	Proceso	Total promedio	Promedio porcentual
Sexto	675	$M = \frac{\sum X_0}{n}$	$\frac{=91+130+81+84+94+102+93}{7}$	96.42	35,19%
Séptimo	516	$M = \frac{\sum X_0}{n}$	$\frac{=77+80+47+70+83+64+95}{7}$	73.71	26,90%
Octavo	395	$M = \frac{\sum X_0}{n}$	$\frac{=83+65+45+41+41+61+59}{7}$	56,42	20,59%
Noveno	332	$M = \frac{\sum X_0}{n}$	$\frac{=48+70+60+38+36+29+51}{7}$	47,42	17,30%
Total promedio de matrículas en el nivel de secundaria en el periodo de los siete años.				271,97	99,98%

Nota. La tabla muestra la frecuencia determinada de la población estimada de cada grado en total de los 7 años tomados para el análisis para hallar la moda.

A partir del promedio de matrícula se calculó el promedio de la variable de rendimiento académico comparando los atributos de aprobados, reprobados, retirados y desertados en RStudio según lo muestra la siguiente tabla:

Tabla 16

Comprobación de los atributos aprobados, desertados, retirados y reprobados en la variable de rendimiento académico.

The screenshot shows the RStudio interface. The top part displays a data table with 13 columns and 7 rows. The columns are labeled 'Séptimo', 'Octavo', 'Noveno', 'sexto1', 'septimo1', 'octavo1', 'novenio1', 'sexto2', 'septimo3', 'octavo4', and 'novenio5'. The rows contain numerical values. Below the table, the R console shows a series of R commands for plotting data from the table. The commands use the 'plot' function to create plots for each column, with labels for 'Años', 'Reprobados', 'Desertados', and 'Retirados'. The main titles for the plots are 'Estudiantes reprobados', 'Estudiantes desertados', 'Estudiantes desertados', 'Estudiantes desertados', 'Estudiantes retirados', 'Estudiantes retirados', 'Estudiantes retirados', and 'Estudiantes retirados'. The console also shows the R version 'R 4.1.2' and the file path '~/'.

to	Séptimo	Octavo	Noveno	sexto1	septimo1	octavo1	novenio1	sexto2	septimo3	octavo4	novenio5
12	5	4	3	26	6	7	2	0	0		
5	2	3	2	5	4	4	3	1	0	0	
0	11	1	3	1	16	4	5	0	2	0	
7	9	3	0	21	13	8	4	2	0	1	
2	1	1	0	3	3	6	3	2	1	3	
0	3	2	0	1	8	6	0	0	0	0	
1	3	2	2	8	11	9	7	2	3	1	

Nota. Creación de la base de datos por cada grado en los años determinados históricamente.

Se enumeró la minería de datos a partir del modelo de regresión lineal en función predictiva teniendo en cuenta la población de matrícula de cada grado en cada año estimado transfiriendo los intervalos de los indicadores de la variable rendimiento académico:

```

predict.lm(object, newdata, se.fit = FALSE, scale = NULL, df = Inf,
  interval = c("none", "confidence", "prediction"),
  level = 0.95, type = c("response", "terms"),
  terms = NULL, na.action = na.pass, pred.var = res.var/weights,
  weights = 1, ...)

```

Este cálculo consistió en el ajuste del modelo de regresión para interpretar el número de estudiantes desertados en función con el conjunto de datos que arrojaron los atributos de la lista

de notas y listas de asistencia con el fin de que se muestre el lenguaje predictivo de R Studio comparando las situaciones de aprobación y reprobación de los años 2014, 2015, 2016, 2017, 2018, 2019 y 2020. Para ello se utilizó el modelo:

```
mod <- lm(Employed ~ Unemployed + Year, data=longley)
```

```
library(broom)
```

```
tidy(mod, quick=TRUE)
```

La minería de datos consistió en la comparación de las siete situaciones:

En el año 2014 con 91 en sexto, 77 en séptimo, 83 en octavo y 48 estudiantes en noveno

En el año 2015 con 130 en sexto, 80 en séptimo, 65 en octavo y 70 estudiantes en noveno

En el año 2016 con 81 en sexto, 47 en séptimo, 45 en octavo y 60 estudiantes en noveno

En el año 2017 con 84 en sexto, 70 en séptimo, 41 en octavo y 38 estudiantes en noveno

En el año 2018 con 94 en sexto, 85 en séptimo. 41 en octavo y 36 estudiantes en noveno

En el año 2019 con 102 en sexto, 64 en séptimo, 61 en octavo y 29 estudiantes en noveno

En el año 2020 con 93 en sexto, 95 en séptimo, 59 en octavo y 51 estudiantes en noveno.

Para ajustar la predicción se calcularon en R Studio los siguientes valores:

```
inicio <- data.frame(Year=c(2014,2015,2016,2017,2018,2019,2020),
```

```
Unemployed=sexto(91,130,81,84,94,102,93),
```

```
Unemployed=séptimo(77,80,47,79,85,64,95),
```

```
Unemployed=octavo(83,65,45,41,41,61,59),
```

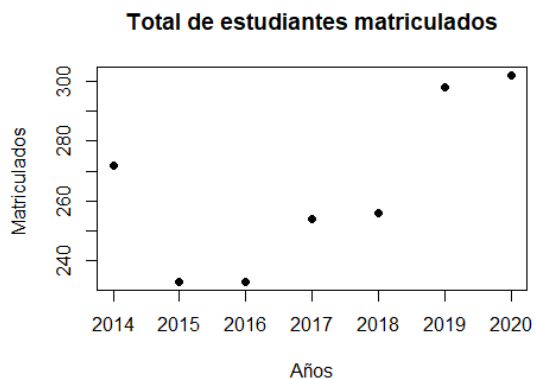
```
Unemployed=noveno(48,70,60,38,36,29,51),
```

```
predict(object=mod, newdata=nuevo)
```

Hallando entonces el nuevo modelo de predicción con el esquema en árbol de decisiones el promedio porcentual y la varianza de matrícula en cada año.

Figura 28

Modelo de predicción hallando el promedio de estudiantes matriculados.

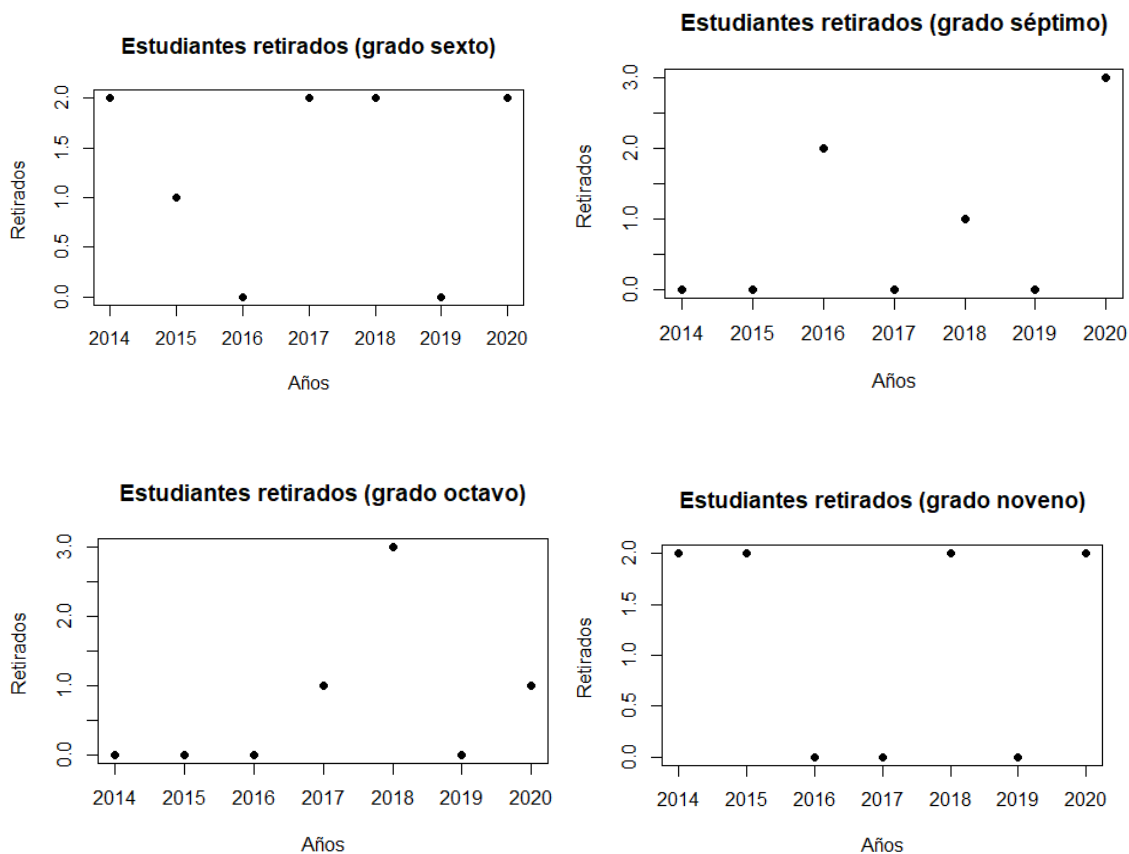


Nota. Evaluación de la variable de estudiantes matriculados

Por consiguiente, se determinó el valor porcentual de los estudiantes retirados con el fin de predecir algorítmicamente la cantidad de estudiantes posibles a desertar comparando el grado seleccionado en cada año estipulado en el estudio. Haciendo comparación de los grados con los años lectivos para el estudio y luego determinar el año que mostró un nivel más alto de estudiantes retirados.

Figura 29

Modelo de predicción hallando el promedio de estudiantes retirados.

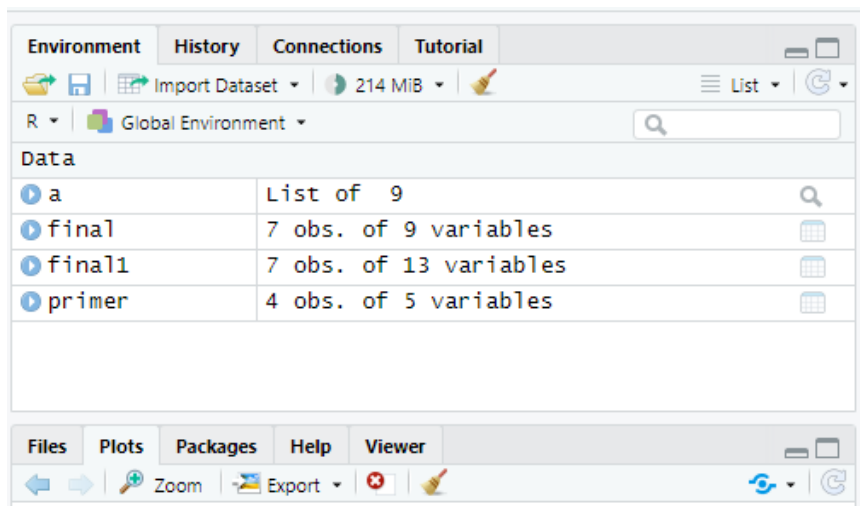


Nota. Evaluación del modelo de gestión hallando el promedio de la estudiantes retirados por grado.

Se hizo la comparación de estudiantes matriculados con los estudiantes retirados en cada grado por año esto permitió la calificación y la regresión de la minería de datos en R Studio.

Figura 30

Proceso histórico del estudio predictivo con árbol de clasificación y regresión

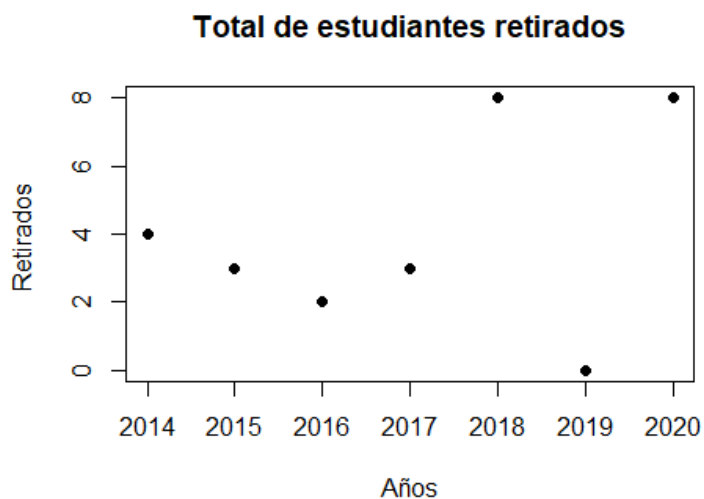


Nota. Se establece el árbol de clasificación y regresión interpretando el comportamiento histórico

Se utilizó la técnica CART para obtener una función que permitiera predecir a través de la variable de estudiantes retirados realizando el estudio predictivo con árbol de clasificación y regresión. El árbol de clasificación dio resultados a los objetivos con técnica discreta separando los datos cualitativos y cuantitativos para interpretar el comportamiento de los procesos de permanencia de los estudiantes durante los cuatro periodos del año y así predecir el porcentaje de estudiantes posibles a abandonar su proceso escolar antes de terminar el año escolar. Arrojando un total clasificado en el árbol de decisiones.

Figura 31

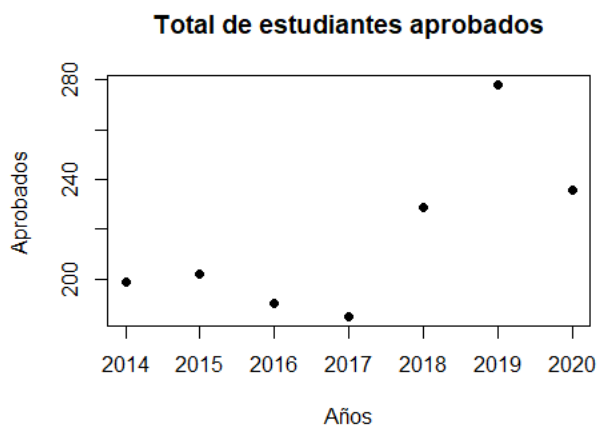
Árbol de decisión en el atributo de retirados



Nota. Hallazgo histórico del comportamiento del árbol de decisiones del atributo de retirados

Figura 32

En el atributo de aprobados en total el modelo de predicción arrojó el árbol de decisiones gráficamente.



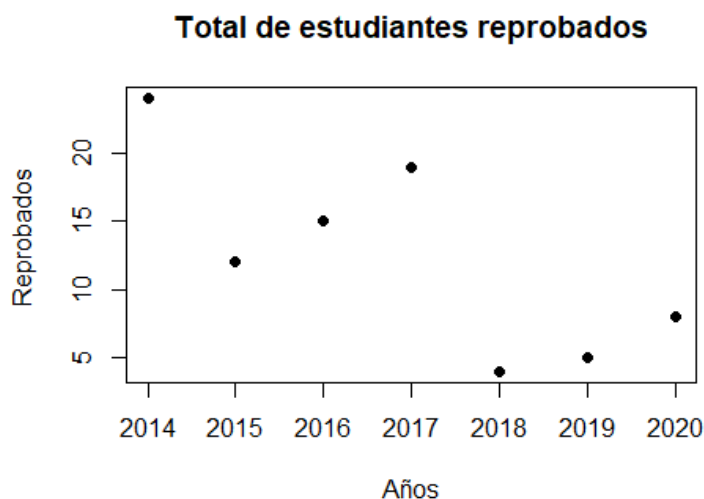
Nota. Hallazgo histórico del comportamiento del árbol de decisiones del atributo de aprobados.

En este atributo se condiciona al estudiante que ha sido matriculado y ha alcanzado con el puntaje promedio en los cuatro periodos de cada año escolar. Notándose en la gráfica como resultado del árbol de decisiones en R un alto porcentaje de estudiantes aprobado en el año 2019. Esta condición estima que la cantidad de estudiante aprobado es promovido al siguiente grado escolar. De este modo se enunció a partir de este modelo el estado del estudiante en el sistema educativo la derivación de las mediciones que sirvieron como parámetros o factores explicativos de deserción escolar.

Otro atributo interesante en este estudio es la determinación del total de estudiantes reprobados que consistió en el conjunto de estimaciones cualitativas relacionadas con la falta de alcance en el promedio de notas y/o la inasistencia en los periodos escolares durante el año lectivo.

Figura 33

Atributo que determinó la estimación de promedio de estudiantes reprobados por año.



Nota. Hallazgo histórico del comportamiento del árbol de decisiones del atributo de reprobados

La categorización de la reprobación en los años determinados para el estudio se encontró un alto índice en el año 2014 superando este porcentaje en el año 2018. De esta forma se infirieron comportamientos cualitativos que permitieron describir las causas de la reprobación:

Comp 1 = no aprobar más de tres áreas

Comp 2 = problemas familiares

Comp 3 = problemas de salud

Comp 4 = desmotivación académica

El reconocimiento de la importancia del atributo de reprobación para la Institución Educativa de Coyongal esbozó cada uno de los factores influyentes en las causas de la reprobación por cuanto afecta el proceso de mejora continua en el servicio educativo. En este sentido, la clasificación de datos sintetizó la información en transferida en el análisis para determinar el modelo de gestión TI para predecir la deserción.

Propuesta del Modelo de Gestión TI apoyado en la Minería de Datos para Predecir la Deserción en el Nivel de Secundaria.

En esta actividad de la fase 3 se desarrolló la propuesta del modelo de Gestión TI apoyado en la minería de datos para predecir la deserción en el nivel de secundaria finalizando el análisis de la cantidad de estudiantes desertados, se organizaron los datos por grado en los años anotados teniendo en cuenta las tres categorías anteriormente totalizadas.

El modelo de gestión TI constituye el análisis de riesgo de estudiantes que deserte en el periodo del año lectivo mediante la construcción de gráficos estadísticos. Las alternativas de cálculo dar solución al problema arrojando resultados de probabilidades de deserción a través de nodos representados por cuadros, nodos de decisión representados por círculos, ramificaciones de alternativas combinando otros procesos representados con flechas y alternativas rechazadas

que muestran resultados que no se deben tener en cuenta pero que son importante en el proceso de predicción.

El modelo de gestión de predicción consiste en el diseño de un esquema que estructura cada evento de ocurrencia, cuantificando los valores estipulados para cada variable: matricula, aprobado, reprobado y desertado. El modelo de probabilidad se define por un margen, puesto que es más exacto en la estadística del evento. En primer lugar, se dibujó el árbol en el programa de Excel para determinar el margen de probabilidad de las variables: matriculados y desertados.

Se determinan dos ramas:

Matriculas

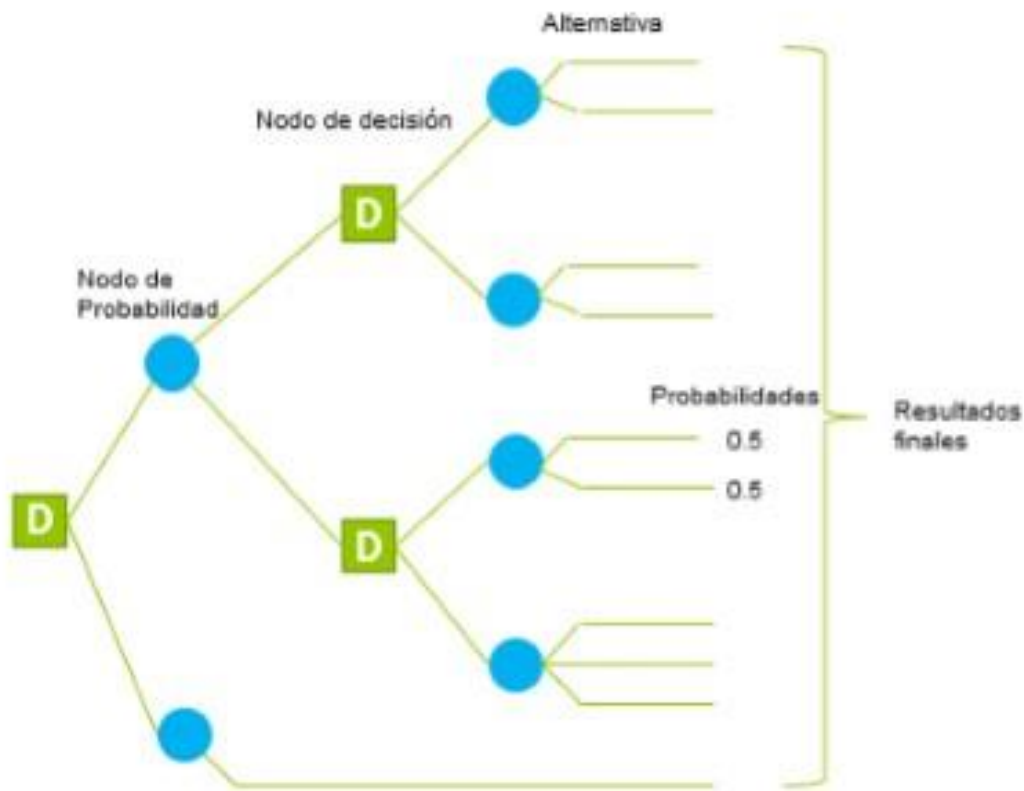
- R1-N1 determina el estudiante matriculado
- R2-N1 determina el estudiante no matriculado.

Rendimiento Académico

- R1-N1 determina el estudiante aprobado
- R2-N1 determina el estudiante reprobado.

Figura 34

Representación del árbol de decisiones detectando los riesgos de deserción escolar

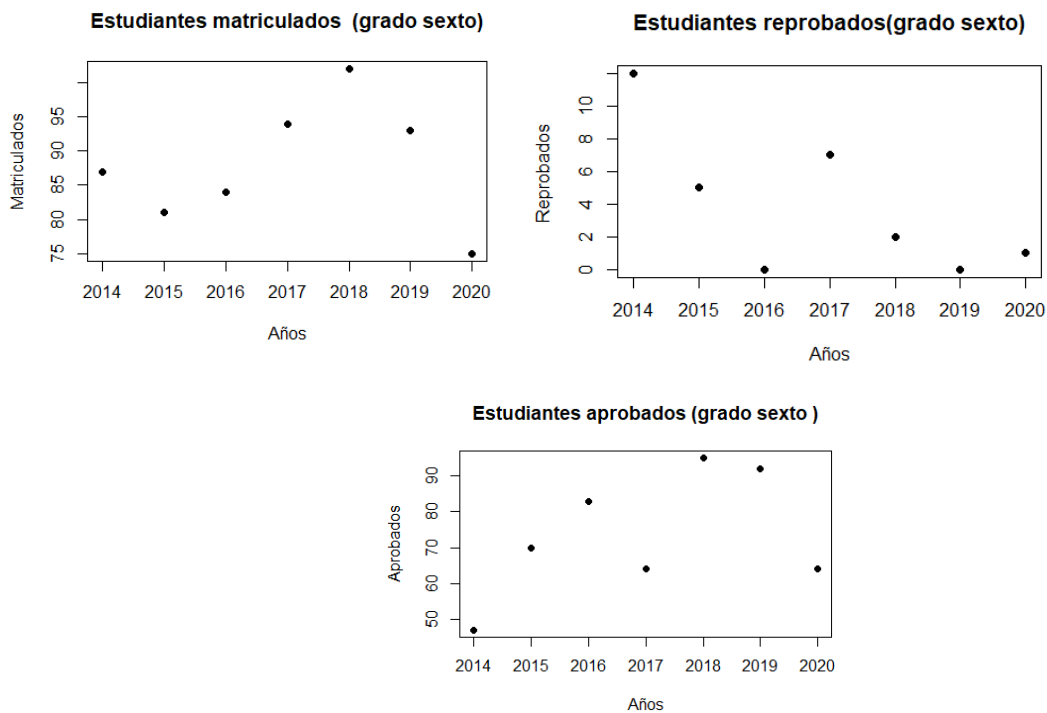


Fuente. Gutierrez (2019). Proyecto de gestión de riesgo. Ministerio de Lectura

Para determinar el cálculo a partir del nodo de decisión se plantearon probabilidades reales encontradas en el análisis documental que arrojó la fase anterior de la investigación. En este punto, se evaluaron los resultados predictivos arrojados en los análisis estadísticos que se presentaron en la propuesta de Modelo de Gestión TI para predecir el nivel de deserción en los años 2021 -2025 tomando la minería de datos como estrategia de selección, organización y clasificación de información en periodos pasados y presentes en Rstudio.

Figura 35

Comparación de los atributos en Rstudio por grado arrojando los siguientes resultados en el grado sexto.

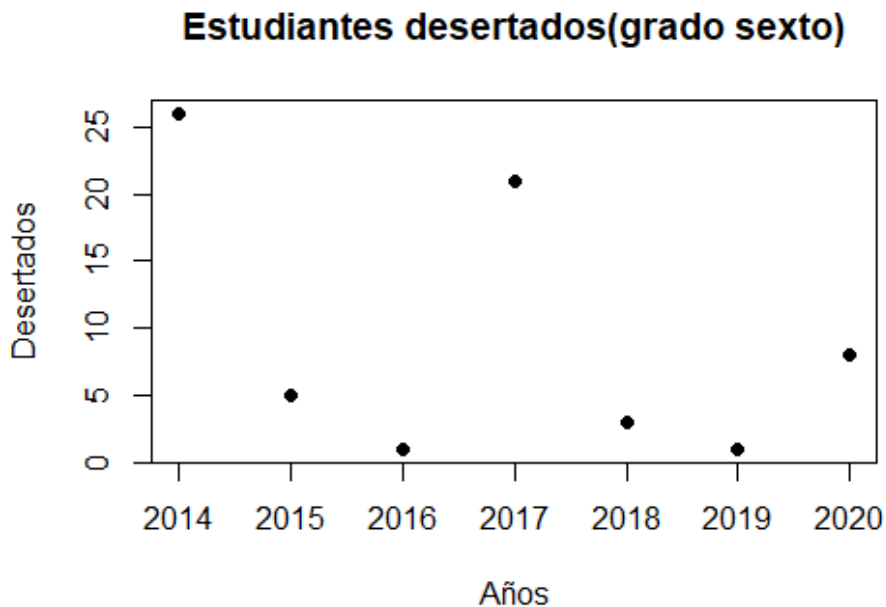


Nota. Comparación de los resultados de cada atributo en el grado sexto con Rstudio.

Interpretando en este modelo de clasificación y de regresión que en el atributo de reprobados discrimina de manera descriptiva las posibilidades de los estudiantes con comportamiento de ausencia u otras causas influyentes en la reprobación escolar. En la variable de rendimiento académico estima al estudiante reprobado aquel que no alcanza los 60 puntos promedios de cada periodo en el año escolar.

Figura 36

Gráfica que predice el porcentaje de estudiantes del grado sexto como desertores.

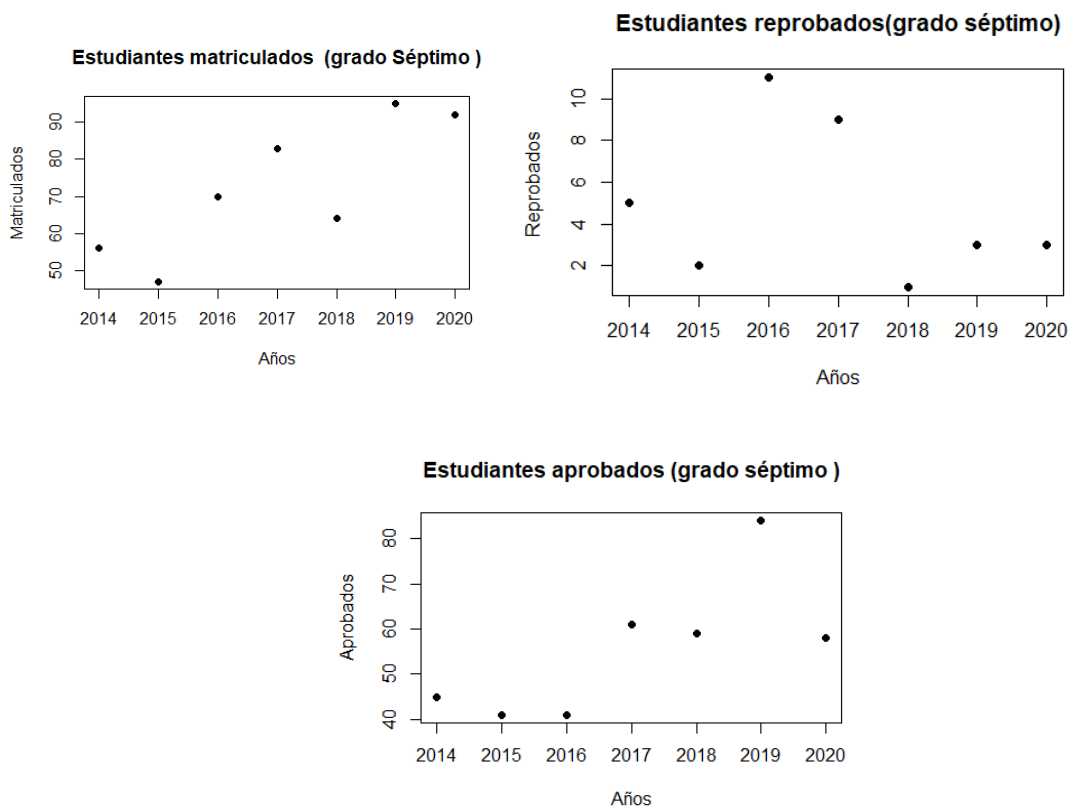


Nota. Histograma del atributo desertado en el grado sexto.

El árbol de decisiones halló que el mayor índice de reprobación se dio en los años 2014 y 2017 notándose mejora en el bajo nivel de deserción en el año 2019, luego aumentó en un porcentaje promedio en el año 2020.

Figura 37

Comparación de los atributos de matrícula, aprobación, reprobación y deserción en el grado séptimo arrojó en el árbol de decisiones de R Studio

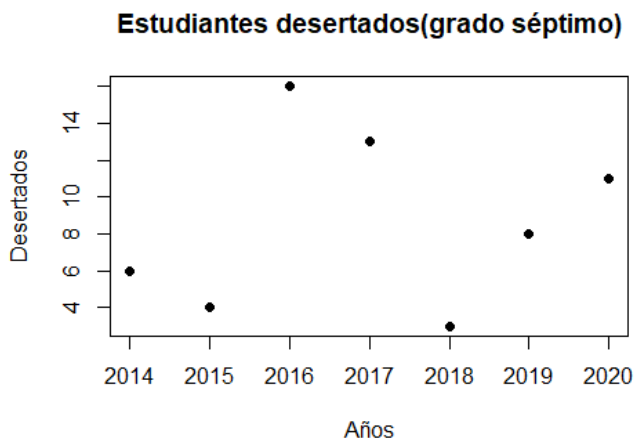


Nota. Histograma de los atributos aprobados, retirados y matriculados del grado séptimo.

En el grado séptimo se encontró un alto nivel de reprobación en los años 2016 y 2017 marcando una diferencia bastante notable en el año 2018. Sin embargo, se hipotetizaron causas que conllevaron a subir el porcentaje de reprobación en los años 2019 y 2020. De este modo, se calculó la predicción de deserción en el grado séptimo en el periodo determinado.

Figura 38

Gráfica que predice el porcentaje de estudiantes del grado séptimo como desertores.



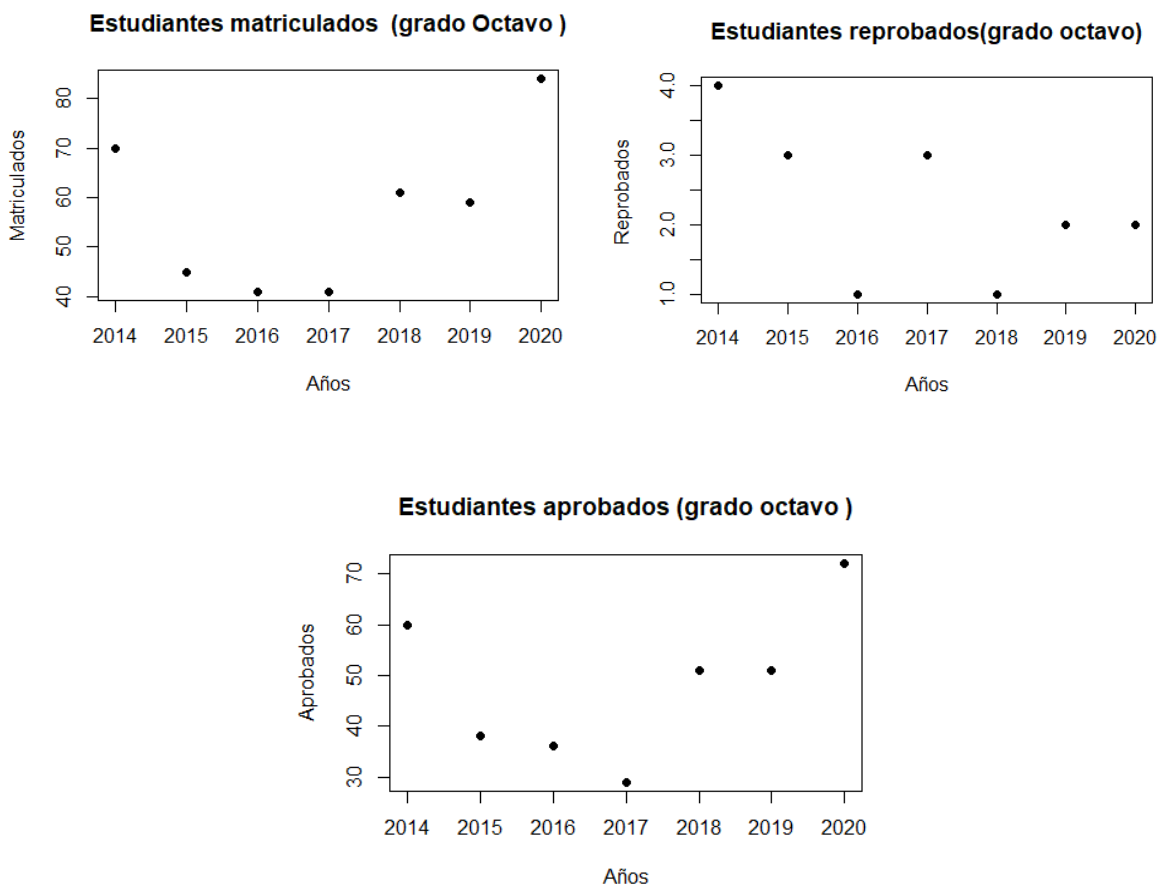
Nota. Histograma del atributo desertado en el grado séptimo.

El modelo predictivo a partir de la clasificación y regresión algorítmica arrojó en el grado séptimo un alto porcentaje de deserción en el año 2016, bajando este problema en el 2017. Sin embargo, en el 2018 disminuyó el nivel de deserción generado expectativas de permanencia en la Institución Educativa de Coyongal. Pero, se notó el aumento nuevamente de la deserción de este grado en los años 2019 y 2020. El enfoque de categorías predictiva constituyó estrategias de minería de datos en la variable de rendimiento académico por cuanto las bajas notas en un importante porcentaje de estudiantes que no superaron en los últimos años el promedio de cada periodo escolar correspondiente a los 60 puntos por cada periodo en cada asignatura del saber.

De igual manera, en el grado octavo se evaluó el criterio de ordenación de datos por atributos que correspondió a los estudiantes matriculados, aprobados, reprobados y desertados asignando una ponderación para hallar el promedio porcentual de acuerdo al análisis predictivo de clasificación y regresión en R estudio.

Figura 39

Comparación de los atributos de matrícula, aprobación, reprobación y deserción en el grado octavo arrojó en el árbol de decisiones de R Studio

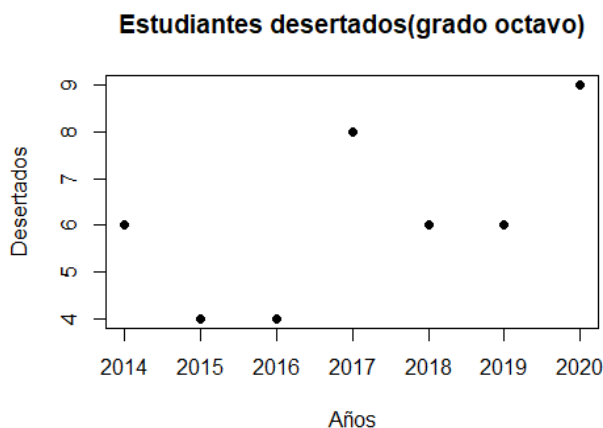


Nota. Histograma de los atributos aprobados, retirados y matriculados del grado octavo.

Los datos correspondientes a la variable de rendimiento académico arrojaron un alto porcentaje de reprobación en el año 2014, luego se muestra que fue disminuyendo con una importante diferencia hasta el año 2016. No obstante, en el año 2017 se evidencia un aumento significativo y preocupante, lo que permitió subrayar varios atributos como la asistencia, las notas y luego comparar con los estudiantes matriculados. La predicción en R Studio arrojó en esta comparación el nivel de deserción:

Figura 40

Gráfica que predice el porcentaje de estudiantes del grado octavo como desertores.



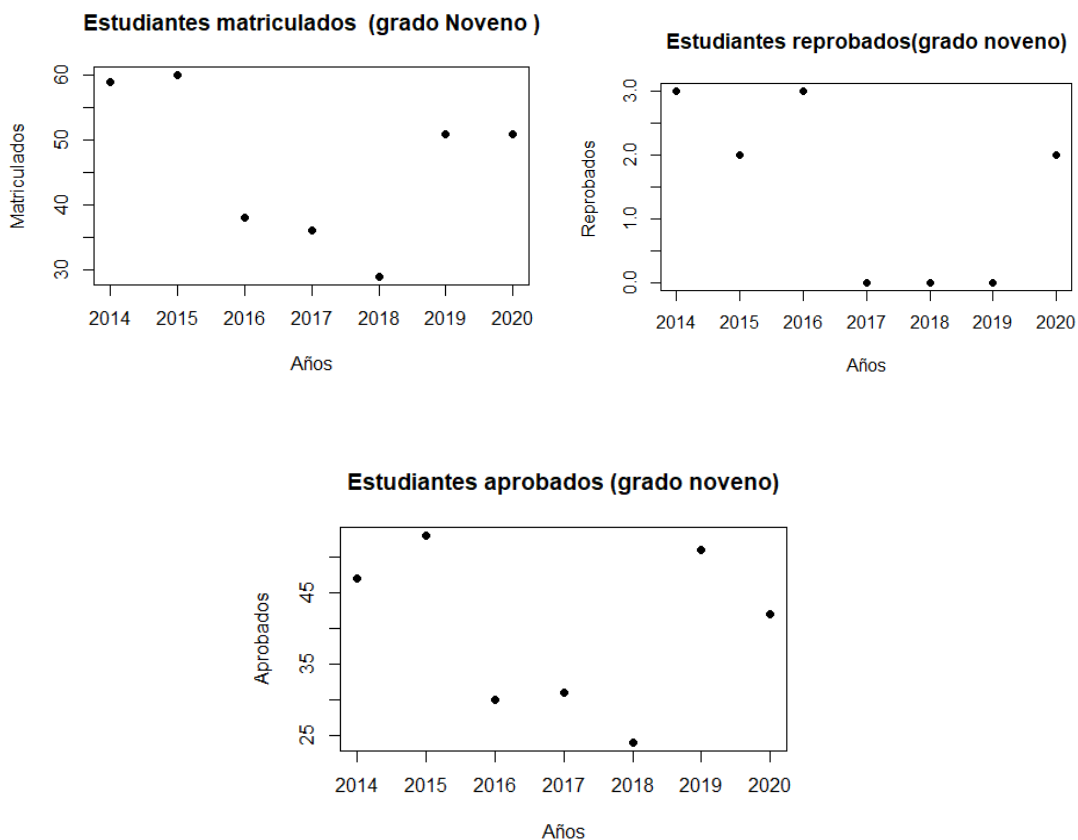
Nota. Histograma del atributo desertado en el grado octavo.

En el análisis de los datos recopilados de acuerdo a las cuatro variables: matrícula, rendimiento académico, eficiencia interna y pruebas Icfes Saber 11, transfirieron de manera alarmante el alto nivel de deserción en el año 2020 puesto que ya se habían ejecutados numerables de estrategias de permanencia motivadoras para el desempeño de los estudiantes. De este modo, se hace necesario e importante el modelo predictivo con árbol de decisiones mediante el enfoque de algoritmos CART en el lenguaje de programación R Studio, puesto que permitió arrojar resultados reales e histórico para el proceso de la formación educativa en la Institución de Coyongal.

Por consiguiente, se abordó el análisis predictivo en el grado noveno contando con los atributos correspondientes del rendimiento académico se enumeró la cantidad de estudiantes reprobados y luego se hizo la comparación con la cantidad de estudiantes matriculados, aprobados y retirados. En este caso la minería de datos mostró una importante variable que permitió la predicción de la deserción.

Figura 41

Comparación de los atributos de matrícula, aprobación, reprobación y deserción en el grado noveno arrojó en el árbol de decisiones de R Studio

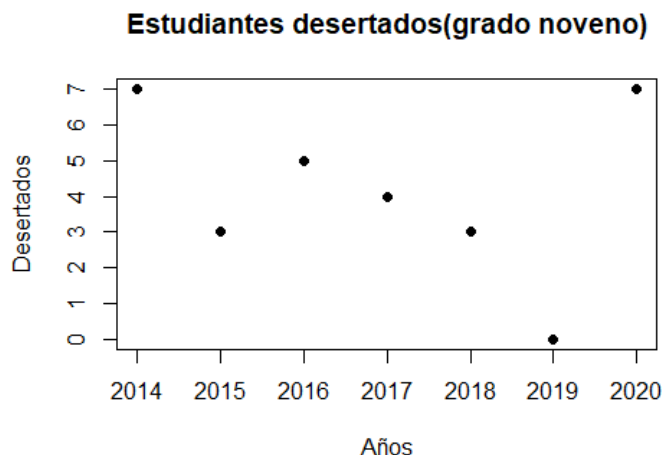


Nota. Histograma de los atributos aprobados, retirados y matriculados del grado noveno.

Luego de generar el árbol de decisiones en el Lenguaje de Programación RStudio se analiza de manera descriptiva desde el enfoque cuantitativo el índice de deserción del grado noveno.

Figura 42

Gráfica que predice el porcentaje de estudiantes del grado noveno como desertores.



Nota. Histograma del atributo desertado en el grado noveno.

Se pudo evidenciar en el análisis un alto índice de deserción en el año 2014, bajando notoriamente en el año 2015. Sin embargo, se elevó nuevamente en el año 2016, pero en el año 2019 bajó en un porcentaje bastante importante. Lo preocupante se notó en este modelo predictivo el aumento de la deserción en el año 2020 lo que motivó proyectar el análisis del año 2021 a 2025 para predecir la deserción.

Evaluación de la Predicción del Nivel de Deserción en los Años 2021 -2025.

Para finalizar el análisis predictivo se realizó la evaluación seleccionando los resultados más precisos de varios datos predictivos para cada ubicación de un cubo espacio y tiempo. Los datos históricos de los años anteriores que conformaron la minería de datos como la historia de matrícula, los informes académico por periodos, lista de asistencia, las pruebas Saber y la eficiencia basaron la historia de los análisis predictivos se estimaron valores futuros diseñando el modelo de gestión TI, es decir, el comportamiento sistemático de la minería de datos y cada una de las categorías clasificadas para la predicción abordaron tres enfoques: de selección de datos,

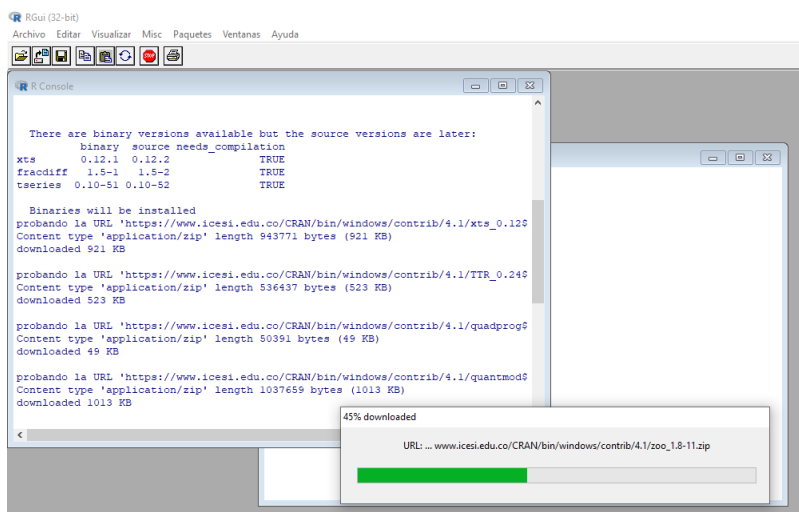
estado académico y categorías predictivas. Se realizó la regresión lineal estimando la tendencia histórica y el comportamiento de cada una de las categorías predictivas.

El enfoque de selección de datos permitió estudiar posibilidades de clasificación de los valores lineales encontrando resultados débiles. De este modo, se definieron reglas predictivas asociadas con árboles de decisiones a través de Algoritmos CART. En este sentido, se hizo necesario crear un histograma de cada grado por año, esto facilitó estimar el nivel de deserción por cada periodo de cada año (un año cuenta con dos periodos evaluativos en la categoría de rendimiento académico).

Se procedió a instalar paquetes `install.package(s) from local files`. Se consideró instalar dos paquetes `Fpp2` y `tidyverse`. En el paquete `fpp2` se obtuvieron dos funciones que ayudaron el pronóstico y la predicción que fueron `goog` y `qcement`. De este modo, se usó la función de `library()` teniendo en cuenta dos categorías: serie temporal que corresponde a la predicción y serie observada que concierne a la información histórica.

Figura 43

Ejecución de la serie temporal para la predicción y la serie observada para el histograma.



Nota. Se determina la url de las dos series para el modelo predictivo.

Figura 44

Los paquetes instalados mostraron el lenguaje algorítmico que permitió clasificar cada una de las categorías predictivas de manera cuantitativa.

```

RGui (32-bit)
Archivo Editar Visualizar Misc Paquetes Ventanas Ayuda

R Console
probando la URL 'https://www.icesi.edu.co/CRAN/bin/windows/contrib/4.1/fpp2_2.46
Content type 'application/zip' length 358544 bytes (350 KB)
downloaded 350 KB

package 'xts' successfully unpacked and MD5 sums checked
package 'TTR' successfully unpacked and MD5 sums checked
package 'quadprog' successfully unpacked and MD5 sums checked
package 'quantmod' successfully unpacked and MD5 sums checked
package 'fracdiff' successfully unpacked and MD5 sums checked
package 'generics' successfully unpacked and MD5 sums checked
package 'lmtest' successfully unpacked and MD5 sums checked
package 'timeDate' successfully unpacked and MD5 sums checked
package 'tseries' successfully unpacked and MD5 sums checked
package 'urca' successfully unpacked and MD5 sums checked
package 'zoo' successfully unpacked and MD5 sums checked
Error in read.dcf(file.path(pkgname, "DESCRIPTION"), c("Package", "Type")) :
  No se puede abrir la conexión
Además: Warning messages:
1: In readRDS(dest) : lzma decoding result 10
2: In download.file(url, destfile, method, mode = "wb", ...) :
  downloaded length 315392 != reported length 2402156
3: In unzip(zipname, exdir = dest) : error 1 al extraer del archivo zip
4: In read.dcf(file.path(pkgname, "DESCRIPTION"), c("Package", "Type")) :
  cannot open compressed file 'RcppArmadillo/DESCRIPTION', probable reason 'No S
>

```

Nota. Después de clasificar, seleccionar y procesar se transformaron los resultados de las cuatro categorías determinadas como minería de datos: matriculados, rendimiento académico, eficiencia interna y pruebas saber.

Figura 45

Organización de la información históricamente.

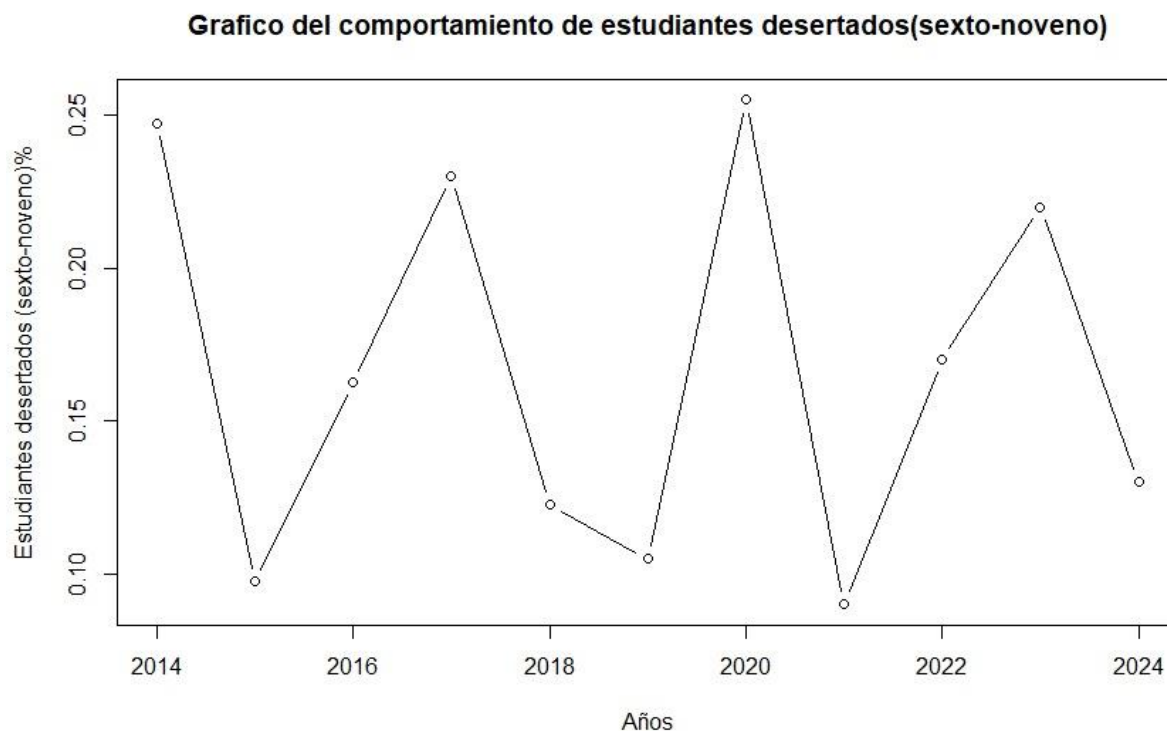
	periodo	sexto	septimo	octavo	noveno
1	2014	0.32	0.13	0.20	0.34
2	2015	0.12	0.04	0.03	0.20
3	2016	0.03	0.35	0.03	0.24
4	2017	0.28	0.19	0.23	0.22
5	2018	0.09	0.02	0.20	0.18
6	2019	0.03	0.15	0.21	0.03
7	2020	0.17	0.18	0.33	0.34

Nota. Confirmación hisorica de la predicción de acuerdo a la serie temporal.

La serie temporal correspondió en la confirmación de la predicción de la deserción para los años 2021, 2022, 2023 y 2024. Es importante anotar en esta investigación que en el procesamiento de datos se realizaron varias tareas que caracterizaron la fiabilidad y la factibilidad de los resultados arrojados en los próximos cuatro años. De acuerdo a la anterior tabla se prepararon los datos clasificados en orden teniendo en cuenta la serie observada de datos históricos y el comportamiento de la deserción por años. Arrojando los siguientes resultados:

Figura 46

Hallazgos del modelo predictivo con árbol de decisiones generando un histograma



Nota. Promedio porcentual de acuerdo al comportamiento histórico en el análisis de la minería de datos estimando el nivel de riesgo de la deserción escolar en el modelo predictivo.

Se hizo la imagen que muestra el promedio de cada grado por año. No se muestra una función regresiva de los datos, según el comportamiento de la deserción estimando el promedio

histórico Rstudio pronosticó el porcentaje de deserción para los años próximos. De este modo en la gráfica se observa que en el año 2014 el grado sexto hubo un 25% de deserción, en el año 2015 en un 10% de desertores y así sucesivamente en todos los grados subió al 16% de deserción en el nivel de secundaria, es decir, sexto, séptimo, octavo y noveno. El comportamiento de la serie histórica sube y baja y de acuerdo a las estimaciones algorítmicamente en el año 2023 subirá la deserción en un 19% siendo que en el 2024 vuelve y baja el nivel de deserción.

De este modo, la herramienta de Rstudio, ajustó la curva de forma paramétrica a cada uno de los valores de espacio y tiempo conformando el cubo de entrada y así se pudo dar la predicción de forma temporal. En la evaluación de los resultados del análisis predictivo se realizó comparación con otras herramientas predictivas y esto afirmó que Rstudio utilizó adecuadamente la tendencia predecible presentando ciclos estacionales importante en la que fusionó varios cubos en el histograma proporcionando valores porcentuales para cada periodo de tiempo.

Conclusiones

Las estimaciones de acuerdo al carácter público de la Institución Educativa de Coyongal, el análisis de las variables se agrupó de acuerdo a los grados de secundaria en los años 2014, 2015, 2016, 2017, 2018, 2019 y 2020 teniendo en cuenta factores descritos en el mapa de procesos de estructura organizacional correspondiente a las gestiones administrativas, directiva y académica. Estos factores incluyeron características de metodología cualitativa que permitieron enfocar la predicción de manera cuantitativa como son: genero, socioeconómico, individuales o familiares e institucionales.

El modelo predictivo con árboles de decisiones y algoritmos CART en R Studio distinguió el riesgo de deserción entre estudiantes hombres y estudiantes mujer notándose mayor riesgo en el género femenino por causas individuales o familiares (embarazo a temprana edad). En comparación con el factor socioeconómico disminuyó el riesgo de deserción por cuanto los estudiantes cuentan con estrategias de permanencia institucional (transporte escolar, restaurante escolar PAE y donación de recursos didácticos para su formación).

Por otro lado, en el factor institucional permitió identificar mayor riesgo de deserción en la variable de rendimiento académico especialmente en los atributos de la inasistencia y reprobación de áreas en los primero dos periodos del año, factor que influye en la deserción escolar.

Basados en la identificación de los anteriores factores se concluye que el modelo predictivo alertó a la Institución Educativa de Coyongal a la aplicación de un análisis predictivo en los primeros tres periodos del año escolar para identificar a los estudiantes de cada grado con riesgo a desertar y luego plantear nuevas estrategias de permanencias institucionales.

Se determinó el modelo de gestión para predecir la deserción a través de árboles de decisiones en Rstudio como proceso alternativo para identificar los riesgos del estudiante desertor. Se concluye que la Institución de Coyongal ha presentado atención al observar el impacto del modelo de gestión TI para calcular estadísticamente la probabilidad eventual de la deserción por cada año y así establecer estrategias institucionales que les permita evitar cuya problemática.

Recomendaciones

Se considera aplicar el modelo de predicción mediante la minería de datos no sólo en las instituciones de educación primaria, secundaria y media, si no también en instituciones de educación superior con la finalidad de detectar al estudiante posible desertor y disminuir los factores incidentes en la problemática con estrategias de permanencia.

Referencias Bibliograficas

Avila. (2021, p. 18). Modelo De Predicción De Deserción Estudiantil, Apoyado En Tecnologías De Data Mining.

<https://repository.unad.edu.co/bitstream/handle/10596/42544/Mavila.pdf?sequence=1&isAllowed=y>

Aristizabal (2017). Diseño y aportes de un modelo para minería de datos educativos en aulas de educación media de carácter presencial. <http://hdl.handle.net/11634/3945>

Best. (1982). *¿Cómo investigar en educación?* Madrid: Morata.

Consejería de Educación (s. f). Las TIC al servicio de un proyecto educativo. Junta de Andalucía

Gonzalez y Llaugel. (2015). Un Modelo Predictivo de Deserción Escolar para la República Dominicana. *PNUD-IDEICE*.

Escobar (2015). Factores asociados a la deserción escolar en la institución educativa normal superior maría inmaculada en el nivel de secundaria –periodo 2014.

<http://bibliotecadigital.univalle.edu.co/bitstream/10893/11299/1/0523728.pdf>

Gaviria (2018). Factores asociados a la deserción escolar en la Institución Educativa Normal Superior María Inmaculada en el nivel de secundaria -período 2014.

<http://hdl.handle.net/10893/11299>

Gonzales (2006). Herramienta de Minería de datos. Revista Cubana de Ciencias Informáticas ISSN: 1994-1536. RCCI Vol. 3, No. 3-4 JULIO-DICIEMBRE, 2009 p. 73-80.

<http://www.redalyc.org/pdf/3783/378343637009.pdf>

Martínez (2012). Aplicación de técnicas de minería de datos para mejorar el proceso de control de gestión de ENTEL. <http://repositorio.uchile.cl/handle/2250/112065>

Ordoñez (2013). Aplicación de técnicas de minería de datos para predecir la deserción de los estudiantes de primer ciclo de la modalidad abierta y a distancia de la UTPL.

<http://dspace.utpl.edu.ec/bitstream/123456789/7897/1/Ordonez%20Brice%C3%B1o%20Karla-%20Informatica.pdf>

Rivero (2012). Clasificación de clientes mediante técnicas de minería de datos para la empresa COMERTEX S.A. <http://tangara.uis.edu.co/biblioweb/tesis/2012/146524.pdf>

Silvera (2016). La evaluación y su incidencia en la deserción escolar: ¿Falla de un sistema, de las instituciones educativas, del docente o del estudiante? DOI:

<http://dx.doi.org/10.17081/eduhum.18.31.1381>

Varón (2017). El fenómeno de la deserción escolar en un contexto local: estudio de la política municipal. dixi 26. Mayo 2017. Pág. 85. doi:

<http://dx.doi.org/10.16925/di.v19i26.1953>

MEN. (2022). *Indice Sintetico de la Calidad Educativa ISCE*.

<https://www.mineduacion.gov.co/portal/micrositios-preescolar-basica-y-media/Evaluacion/Evaluacion-de-estudiantes/397385:Indice-Sintetico-de-la-Calidad-Educativa-ISCE>

Oviedo y Diaz. (1982 en Lugo, 2013, p. 292). La deserción estudiantil: ¿Realmente es un problema social? *Revista de Postgrado FACE-UC. Vol. 7 N° 12*.

Pérez y Merino. (2008). La desercion escolar. Obtenido de <https://definicion.de/desercion-escolar/>

Roman. (2013). Factores asociados al abandono y la deserción escolar en América Latina: Unca mirada en conjunto. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación* .

Timarán, Caicedo e Hidalgo. (2019). Árboles de decisión para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas Saber 11°. *Revista de Investigación, Desarrollo e Innovación*. doi:10.19053/20278306.v9.n2.2019.9184

Apendices.

Apendice A

Institución Educativa de Coyongal.



Resignificación de los procesos de Gestión académica y administrativa





Resignificación de estrategias de permanencia académica

Apendice B

Organización de espacio y tiempo por periodos académicos

INSTITUCIÓN EDUCATIVA DE COYONGAL CALENDARIO DE ACTIVIDADES PARA FINALIZACIÓN TERCER PERIODO ACADÉMICO Del 16 de Octubre de 2022 hasta el 2 de diciembre de 2022.			
ACTIVIDADES	FECHA DE INICIO	FECHA FINAL	RESPONSABLE
Registro en Punto Edu, indicadores de desempeño Cuarto periodo periodo	martes, 18 de octubre de 2022	miércoles, 09 de noviembre de 2022	Docentes
Actualización novedades en Punto Edu	martes, 18 de octubre de 2022	miércoles, 09 de noviembre de 2022	Docentes y Directivos
Registro de notas en Punto Edu, cuarto periodo académico. (Primaria)	martes, 18 de octubre de 2022	martes, 22 de noviembre de 2022	Docentes
Reunión evaluación y promoción (Preescolar y Primaria)	jueves, 24 de noviembre de 2022	jueves, 24 de noviembre de 2022	Docentes y Directivos
Finalización cuarto periodo académico (Primaria)	lunes, 28 de noviembre de 2022	lunes, 28 de noviembre de 2022	Docentes y Directivos
Registro de notas en Punto Edu, cuarto periodo académico. (Secundaria)	martes, 18 de octubre de 2022	jueves, 17 de noviembre de 2022	Docentes
Finalización Cuarto periodo académico (Secundaria)	viernes, 18 de noviembre de 2022	viernes, 18 de noviembre de 2022	Docentes y Directivos
Reunión evaluación y promoción (Secundaria)	viernes, 18 de noviembre de 2022	viernes, 18 de noviembre de 2022	Docentes y Directivos
Nivelaciones Especiales	lunes, 21 de noviembre de 2022	miércoles, 23 de noviembre de 2022	Docentes y Estudiantes
Entrega y revisión de Resultados Nivelacion especial.	jueves, 24 de noviembre de 2022	jueves, 24 de noviembre de 2022	Docentes y Directivos
Entrega de Informes Académicos Cuarto periodo (Preescolar y Primaria)	lunes, 28 de noviembre de 2022	lunes, 28 de noviembre de 2022	Docentes y Directivos
Entrega de Informes Académicos Cuarto periodo (Secundaria)	martes, 29 de noviembre de 2022	martes, 29 de noviembre de 2022	Docentes y Directivos
Grados (Preescolar y Primaria)	miércoles, 30 de noviembre de 2022	miércoles, 30 de noviembre de 2022	Docentes y Directivos
Grados (Secundaria)	jueves, 01 de diciembre de 2022	jueves, 01 de diciembre de 2022	Docentes y Directivos

Nota: Este calendario es de estricto cumplimiento, no se hará ningún tipo de modificación a las actividades programadas en sus respectivas fechas, por lo que se solicita a los docentes realizar todas las gestiones necesarias para el cumplimiento en los tiempos establecidos.

Apéndice C

Resignificación de planes curriculares e institucionalización de formatos



Planeación y reorganización de los planes de estudios institucionales