UNIVERSIDAD TECNICA DE COTOPAXI

DIRECCIÓN DE POSGRADO

MAESTRIA EN SISTEMAS DE INFORMACIÓN

MODALIDAD: PROPUESTA METODOLÓGICA Y TECNOLOGÍA AVANZADA

Titulo:

Modelo de análisis del rendimiento académico de la Unidad Educativa Personas Con Escolaridad Inconclusa. (P.C.E.I.) "Monseñor Leonidas Proaño" del cantón Latacunga, a través de minería de datos.

Trabajo de titulación previo a la obtención del título de magister en Sistemas de Información.

Autor:

Chancúsig Taipicaña Diego Marcelo

Tutora:

Albán Taipe Mayra Susana PhD.

LATACUNGA – ECUADOR

2020

APROBACIÓN DEL TUTOR

En mi calidad de Tutora del Trabajo de Titulación "Modelo de análisis del rendimiento académico de la Unidad Educativa Personas Con Escolaridad Inconclusa. (P.C.E.I.) Monseñor Leonidas Proaño del cantón Latacunga, a través de minería de datos." presentado por Chancúsig Taipicaña Diego Marcelo, para optar por el título magíster en Sistemas de Información.

CERTIFICO

Que dicho trabajo de investigación ha sido revisado en todas sus partes y se considera de que reúne los requisitos y méritos suficientes para ser sometido a la presentación para la valoración por parte del Tribunal de Lectores que se designe y su exposición y defensa pública.

Latacunga, mayo, 09, 2020

……

Albán Taipe Mayra Susana PhD.

CC.: 0502311988

APROBACIÓN TRIBUNAL

El trabajo de Titulación: Modelo de análisis del rendimiento académico de la Unidad Educativa Personas Con Escolaridad Inconclusa. (P.C.E.I.) Monseñor Leonidas Proaño del cantón Latacunga, a través de minería de datos, ha sido revisado, aprobado y autorizado su impresión y empastado, previo a la obtención del título de Magíster en Sistemas de Información; el presente trabajo reúne los requisitos de fondo y forma para que el estudiante pueda presentarse a la exposición y defensa.

Latacunga, junio, 09, 2020

.....

MSc. José Augusto Cadena Moreano
CC. 0501552798
Presidente del tribunal

.....

MSc. Alex Christian Llano Casa
CC. 0502589864
Lector 2

.....

MSc. Roberto Carlos Herrera Albarracín
CC. 0502310253
Lector 3

DEDICATORIA

El presente trabajo está dedicado a mi esposa Paulina Erazo y mi querida hija Paula Marcela, quienes con su ánimo, carisma y afectividad me han acompañado a culminar esta nueva experiencia académica que me ha permitido crear nuevas expectativas en mi profesión.

AGRADECIMIENTO

Expresar mi eterno agradecimiento a las autoridades, administrativos y docentes de pos grado de la Universidad Técnica de Cotopaxi por la oportunidad de formarme adacémicamente para la realización de nuevas expectativas. De una manera especial a la PhD. Mayra Susana Albán Taipe tutora del presente trabajo investigativo, demostrando su profesionalismo con una actitud positiva.

Muchas gracias.

RESPONSABILIDAD DE AUTORÍA

Quien suscribe, declara que asume la autoría de los contenidos y los resultados obtenidos en el presente trabajo de titulación.

Latacunga, mayo, 09, 2020

……
Diego Marcelo Chancúsig Taipicaña
0502309388

RENUNCIA DE DERECHOS

Quien suscribe, cede los derechos de autoría intelectual total y/o parcial del presente trabajo de titulación a la Universidad Técnica de Cotopaxi.

Latacunga, mayo, 09, 2020

……
Diego Marcelo Chancúsig Taipicaña
0502309388

AVAL DEL VEEDOR

Quien suscribe, declara que el presente Trabajo de Titulación: Modelo de análisis del rendimiento académico de la Unidad Educativa Personas Con Escolaridad Inconclusa. (P.C.E.I.) Monseñor Leonidas Proaño del cantón Latacunga, a través de minería de datos, contiene las correcciones a las observaciones realizadas por los lectores en sesión científica del tribunal.

Latacunga, junio, 23, 2020

.....

MSc. José Augusto Cadena Moreano
CC. 0501552798

UNIVERSIDAD TÉCNICA DE COTOPAXI
DIRECCIÓN DE POSGRADO

MAESTRÍA EN SISTEMAS DE INFORMACIÓN

Título: Modelo de análisis del rendimiento académico de la Unidad Educativa Personas Con Escolaridad Inconclusa. (P.C.E.I.) Monseñor Leonidas Proaño del cantón Latacunga, a través de minería de datos.

Autor: Chancúsig Taipicaña Diego Marcelo.

Tutora: PhD. Mayra Susana Albán Taipe.

RESUMEN

El objetivo principal de este trabajo es contribuir al proceso de predicción del rendimiento académico de los estudiantes de la Unidad Educativa de Personas Con Escolaridad Inconclusa (PCEI) Monseñor Leonidas Proaño de la ciudad de Latacunga mediante el estudio integral de técnicas y herramienta de análisis de minería de datos a partir de los factores de influencia como el social, económico y académico, determinando indicadores que detecten elementos que servirán a los docentes, autoridades y mentores educativos para mejorar el rendimiento académico del estudiante en el el proceso educativo. Una de las etapas de esta investigación fue el diseño de un modelo teórico de la retención estudiantil a través del software Statistical Package for Social Sciences (SPSS) a través de regresión lineal, mínimos cuadrados ordinarios que permitieron crear el modelo teórico del rendimiento académico Posteriormente este modelo siguió un proceso experimental con cuatro algoritmos de clasificación a través de técnicas de machine learning como J48, Random Forest, Naive Bayes y OneR, proceso que se utilizó para predecir la tasa de precisión del modelo propuesto. La implementación de estas técnicas permitió determinar que el algoritmo Naive Bayes presenta una tasa de precisión del 88.85% lo que indica que el modelo que se presenta es adecuado en términos de confiabilidad, los niveles de capa obtenidos a través del proceso experimental con un resultado del 0,86 indican que estos modelos son adecuados para predecir la retención estudiantil.

PALABRAS CLAVE: Modelo análisis de rendimiento académico secundaria, minería de datos.

UNIVERSIDAD TECNICA DE COTOPAXI
DIRECCION DE POSGRADO

MAESTRIA EN SISTEMAS DE INFORMACIÓN

Title: Performance academic analysis model of people with unfinished Schooling from Unidad Educativa Monseñor Leonidas Proaño (P.C.E.I.) from Latacunga city through data mining

Author: Chancúsig Taipicaña Diego Marcelo

Tutor: Mayra Susana Albán Taipe. PhD.

ABSTRACT

The main objective of this work is to contribute to the process of predicting the academic performance of the students of Unidad Educativa de Personas Con Escolaridad Inconclusa (PCEI) Monseñor Leonidas Proaño from Latacunga city by means of the integral study of techniques and tool of analysis of mining of data from the factors of influence such as the social, economic and academic, determining indicators that detect elements which will serve the teachers, authorities and educational mentors to improve the academic performance of the student in the process of education. One of the stages of this research was to design a theoretical model of student retention through the Statistical Package for Social Sciences (SPSS) software applying linear regression, ordinary least squares that allowed to create the theoretical model of academic performance. This model followed an experimental process with four classification algorithms through machine learning techniques such as J48, Random Forest, Naive Bayes and OneR, this process was used to predict the precision rate of the proposed model. The implementation of these techniques allowed us to determine that the Naive Bayes algorithm presents an accuracy rate of 88.85%, which indicates that the model presented is adequate in terms of reliability, the layer levels obtained through the experimental process with a result of 0.86 indicate that these models are adequate to predict students retention.

KEY WORDS: Secondary academic performance analysis model, data mining.

Yo, Fanny Rosario Villagrán Vergara con cédula de identidad número: 1708136724 Magister en Lingüística y Didáctica de la Enseñanza de Idiomas Extranjeros con número de registro de la SENESCYT:2016 -10 -03; **CERTIFICO** haber revisado y aprobado la traducción al idioma inglés el resumen del trabajo de investigación con el título: "Modelo de análisis del rendimiento académico de la Unidad Educativa Personas Con Escolaridad Inconclusa. (P.C.E.I.) Monseñor Leonidas Proaño del cantón Latacunga, a través de minería de datos" de Diego Marcelo Chancúsig Taipicaña, aspirante a magister en Sistemas de Información.

Latacunga, marzo, 10, 2020

…….....

Fanny Rosario Villagrán Vergara

1708136724

ÍNDICE DE CONTENIDOS

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

ÍNDICE DE FÓRMULAS

ABREVIACIONES

PCEI Personas con escolaridad inconclusa

EGB-S Educación General Básica Superior

BGU Bachillerato General Unificado

BT Bachillerato Técnico

SPSS Statistical Package for Social Sciences

KDD Knowledge Discovery in Databases

MCO Mínimos Cuadrados Ordinarios

LOEI Ley Orgánica Intercultural Bilingüe

DECE Departamento de Consejería Estudiantil

INTRODUCCIÓN

La revolución digital ha permitido realizar investigaciones de análisis gracias a la facilidad de capturar, procesar, almacenar, distribuir y transmitir información digitalizada generada por las nuevas sociedades para descubrir modelos interesantes y solucionar problemas cotidianos de la sociedad [1]. El progreso de las tecnologías de la información y comunicación y su evidente globalización ha permitido integrar diferentes tipos de datos y gestionar grandes bases de datos para interpretar la información y el conocimiento [2]. La manipulación de datos a gran escala ha superado la capacidad humana, siendo necesaria la ayuda de las tecnologías informáticas para automatizar el proceso a través de técnicas y herramientas ya que actualmente la mayoría de organizaciones o instituciones procuran gestionar las actividades relevantes de manera eficaz y eficiente apoyándose en las ciencias informáticas y no es ajeno en el sistema educativo donde se genera datos de la gestión educativa administrativa y académica desarrollada en plataformas virtuales para la comunidad educativa [3].

El rendimiento académico en las instituciones educativas ha permitido evidenciar la producción real del estudiante en actividades formales, por lo que es constante el desafío de mantener y/o mejorar la calidad académica, revisando periódicamente contenidos, estrategias y métodos de enseñanza para garantizar estándares de calidad, con la finalidad de establecer posibles soluciones y evitar la deserción estudiantil que es una de las causas evidentes que deriva del bajo rendimiento académico. Ante esta situación problémica, en los últimos treinta años ha surgido la preocupación de países de América Latina, en especial de México, por los estudiantes que no alcanzan niveles de excelencia o al menos estándares de conocimiento; realidad que requiere ser analizada para determinar las causas que influye en el rendimiento académico, el incremento de la deserción y abandono en el sistema educativo [4].

La Organización para la Cooperación y el Desarrollo Económicos (OCDE) en el año 2000 [5] reúne a treinta y dos países de distintos continentes, que participaron

en el Programa Internacional para la Evaluación de Estudiantes (PISA), con el objetivo de buscar mecanismos para detectar y describir los aspectos de enseñanza en el contexto personal, familiar y escolar que influyen y que se reflejan al aplicar las evaluaciones. A inicios de este nuevo siglo la Organización de Estados Iberoamericanos (OEI) en el año 2007 a través de la OCDE, publica informes finales donde establece causas, consecuencias y alternativas de solución del bajo rendimiento de los estudiantes [6]. En Ecuador el Ministerio de Educación, junto a la Investigación Iberoamericana sobre Eficacia Escolar (IIEE) y al CAB, (Convenio Andrés Bello, 2007) [6] ayudó a determinar los factores educativos asociados al desempeño educativo, y entre los más relevantes determina concluyendo que no solo lo que acontece en el aula tiene importancia para el desarrollo de los alumnos, sino incide también las características del país.

En el año 2014 el Ministerio de Educación del Ecuador y como aliado el Instituto Nacional de Evaluación Educativa (INEVAL) se adhiere al Programa Internacional para la Evaluación de Estudiantes para el Desarrollo (PISA-D), participando en octubre de 2017 por primera vez en las evaluaciones con 6100 estudiantes de 15 años pertenecientes a diversas instituciones educativas que cursaban desde el octavo año de Educación General Básica (E.G.B.) hasta el bachillerato, obteniendo un favorable desempeño en consideración al promedio de América Latina y el Caribe (ALC) sobre la media en lectura 408 puntos, en ciencias con 399 puntos y ligeramente bajo la media en matemática con 377 puntos [7]. Sin embargo, al cotejar los resultados de Ecuador con países participantes del resto del mundo la puntuación es inferior al promedio de la OCDE.

En la revista Panorama de la educación del año 2017 de propiedad de la fundación Santillana fueron publicados también los resultados de las evaluaciones del Programa Internacional para la Evaluación de Estudiantes (PISA) a nivel mundial siendo celebrado por países como Singapur con el mejor ranking en las tres asignaturas, siguiendo Japón, Estonia, Taiwán, Finlandia y China (Macao). En la búsqueda constante de un rendimiento académico óptimo y la calidad de educación en las instituciones educativas, el Estado Ecuatoriano en la Constitución de la

República del Ecuador publicado en el registro oficial 449 del 20 de octubre de 2008 considera en el artículo 347 numeral 1 "*Fortalecer la educación pública y la coeducación; asegurar el mejoramiento permanente de la calidad, la ampliación de la cobertura, la infraestructura física y el equipamiento necesario de las instituciones educativas públicas*".

De igual manera la Ley Orgánica de Educación Intercultural [LOEI] (2011). Quito: Registro Oficial, 417 en el artículo 22, literal dd [8] *"La Autoridad Educativa Nacional definirá estándares e indicadores de calidad educativa que serán utilizados para las evaluaciones por el Instituto Nacional de Evaluación Educativa. Los estándares serán al menos de dos tipos: curricular, referidos al rendimiento académico estudiantil y alineados con el currículo nacional obligatorio…"* artículo que al ser cumplidos los estudiantes mejorarían la vida y el desarrollo sostenible tanto individual como colectivo. El Instituto Nacional de Evaluación Educativa (INEVAL), la Organización para la Cooperación y el Desarrollo Económicos (OCDE) [5] junto al comité del Ministerio de Educación del Ecuador como revisor en el año 2018 publicó el informe nacional de la evaluación, obteniendo indicadores propios del país, con la finalidad de proporcionar información de calidad para constituir una base de datos para la investigación y el análisis orientados a mejorar políticas, destacando que el factor que mayor influencia existe en el rendimiento académico de los estudiantes de secundaria es el socioeconómico.

Por otra parte, Ecuador ha dedicado en la última década esfuerzos para llegar a la universalización de la educación con los lineamientos de los Objetivos de Desarrollo Sostenible (ODS) en especial el objetivo cuarto que pretende garantizar a las personas la inclusión y conclusión de los ciclos de enseñanza primaria, secundaria y lograr calidad y pertinencia del aprendizaje a lo logro de la vida. Tal es el caso que en el 2017 se registra en Educación General Básica (EGB) una tasa de matrícula del 96,2 %, y en el mismo año una tasa de matrícula del 71 % en el Bachillerato General Unificado (BGU). En el mismo año es preocupante el creciente porcentaje de deserción en EGB con un 2,1% y un 5,3% en BGU. De igual

forma, es preocupante el creciente porcentaje de no promovidos, con el 1,2% en EGB y el 3,4% de no promovidos para BGU. [9].

La Unidad Educativa Personas Con Escolaridad Inconclusa (PCEI) Monseñor Leonidas Proaño de la ciudad de Latacunga, institución fiscomisional por el convenio celebrado por el Subsistema de Educación Fiscomisional Semipresencial del Ecuador (SEFSE) y el Ministerio de Educación, menciona que en los últimos tres años es evidente la preocupación por el rendimiento académico como un factor crítico que está asociado a una alta tasa de deserción y abandono de los estudiantes. Esta compleja realidad educativa por su naturaleza al formar y educar a adolescentes, jóvenes, adultos, adultos mayores, grupos de personas vulnerables, personas privadas de la libertad (PPL´s) con la pre libertad y personas con necesidades educativas especiales no ha sido ajena a los retos que la educación exige para encontrar posibles soluciones frente al bajo rendimiento académico y la deserción estudiantil en todos sus niveles.

Por lo expuesto anteriormente, se propone diseñar un modelo de análisis del rendimiento académico a través de minería de datos: caso de estudio Unidad Educativa PCEI Monseñor Leonidas Proaño de la ciudad de Latacunga. Este modelo es planteado con la finalidad de conocer las causas del bajo rendimiento académico y proponer estrategias probabilísticas que permitan a los administradores de las instituciones educativas mejorar el proceso educativo en los estudiantes en el nivel de educación general básica superior y bachillerato para la toma de decisiones efectivas.

Para cumplir con el objetivo general es importante hacer referencia a la investigación bibliográfica, como primer objetivo específico para conocer el estado del arte de los modelos de análisis de datos del rendimiento académico a través de minería de datos**.** A continuación, se describe cada una de las tareas para el desempeño de los objetivos específicos de la investigación.

Tabla 1. Tareas de los objetivos específicos

Objetivos	Actividades (tareas)
Objetivo Específico 1: Estudiar sistemáticamente la literatura para conocer las concepciones teóricas y filosóficas de la minería de datos y su influencia en el análisis predictivo del rendimiento académico de los estudiantes de la Unidad Educativa de Personas Con Escolaridad Inconclusa Monseñor Leonidas Proaño.	Buscar e identificar las fuentes primarias de información.
	Analizar el contenido de las fuentes primarias.
	Definir el marco conceptual.
Objetivo Específico 2: Establecer a través de analítica de componentes la importancia de los indicadores académicos, sociales y económicos de los estudiantes de la Unidad Educativa de Personas Con Escolaridad Inconclusa Monseñor Leonidas Proaño de la ciudad de Latacunga.	Definición de la metodología.
	Definir las variables más relevantes para la construcción del modelo.
	Validación de variables a través de la analítica descriptiva de datos.
	Validación del modelo a través de encuesta.
Objetivo Específico 3: Aplicar técnicas de minería de datos para los indicadores en los estudiantes de la Unidad Educativa de Personas Con Escolaridad Inconclusa Monseñor Leonidas Proaño de la ciudad de Latacunga.	Aplicación de la metodología .
	Aplicación de técnicas de Minería de datos.
	Extracción de conocimiento.
	Evaluación del modelo de predicción.

Elaborado por: El investigador

Las investigaciones que se han dedicado a buscar soluciones en la construcción de modelos de análisis en el ámbito educativo, han experimentado un arduo trabajo por la cantidad de datos para identificar y encontrar información útil aprovechando el uso de herramientas de integración de datos y la aplicación de algoritmos mediante técnicas de minería de datos. Alvarado y Álvarez [10] manifiesta que la minería de datos en el proceso educativo en los últimos años se orienta al desarrollo de métodos con el uso de plataformas educativas para entender a los estudiantes y el entorno en el que aprenden. Nieto [11] manifiesta que la literatura ha permitido encontrar en la actualidad la existencia de modelos de análisis de datos en el campo educativo enfocado al estudio del rendimiento académico; pero la mayoría de investigaciones se han enfocado al nivel superior o universitario por la facilidad de

acceso a sólidas bases de datos que posee estas instituciones, y es precisamente por las limitadas investigaciones que, se pretende contribuir al estudio y el análisis de datos en el sistema educativo secundario, dando respuestas prácticas y tecnológicas a la problemática expuesta aplicando el conocimiento científico.

Este modelo propuesto tiene como caso de estudio a la Unidad Educativa PCEI "Monseñor Leonidas Proaño" de la ciudad de Latacunga, que requiere conocer las posibles causas del rendimiento académico, pretendiendo descubrir información útil y relevante que derriban no únicamente de los datos académicos, sino también del registro de asistencia a las tutorías presenciales, el comportamiento, la situación social y económica de los estudiantes usando técnicas de minería de datos. Pero es importante sugerir el complementar la aplicación del modelo con una adecuada orientación vocacional la cual podría ayudar a definir la afinidad, conocimientos, destrezas y habilidades del estudiante a futuro para profesionalizarse en carreras universitarias o profesión técnicas y artesanales. Por otra parte, el modelo propuesto puede servir como herramienta de apoyo el cual puede ser aplicado para su validación en las diez extensiones de las Unidad Educativa.

Para este trabajo es importante partir de la búsqueda de información en fuentes de consulta primarias para validar el conocimiento científico y cumplir los objetivos propuestos empezando por la investigación bibliográfica. Sampier, Collado y Lucio [12] manifiestan que la investigación científica cumple dos propósitos: una investigación básica que produce teorías, conocimiento y la investigación aplicada que resuelve problemas prácticos. Como primer paso en el proceso de investigación es la revisión a través de la investigación bibliográfica o documental como fuente primaria como aporte de datos. Ocampos [13] considera que las fuentes bibliográficas no se centran únicamente en fuentes documentales, sino también en otro tipo de fuentes y lo fundamental es considerar el acceso a estas. La investigación de campo requiere salir en busca de obtener la información amplia y útil. Por lo que es necesario puntualizar que este trabajo investigativo por la diversidad de la información recolectada en la data se utiliza la metodología mixta,

ya que se pretende mostrar y explicar el porqué de las cosas e identificar las causas reales.

Abreu [14] en su artículo plantea que este tipo de investigación permite explicar en detalle construyendo y ampliando las razones detrás de la teoría determinando la respuesta más acertada, comprobada y validando las predicciones de una teoría. Pereira [15] la investigación mixta permite obtener una perspectiva más desarrollada sobre los factores del análisis del rendimiento académico de los estudiantes, ya que se obtiene una mejor exploración y aprovechamiento de los datos analizados cuantitativa, ya que se recolecta, analiza y gestiona información científica con relación a las variables para dar respuesta a preguntas y prueba las hipótesis planteadas, estableciendo patrones exactos de comportamiento de la población ayudados por el uso de la estadística. Ocampos [13] prevé que se debe tener cuidado en las conclusiones obtenidas a partir de los datos. Por otro lado, la cualitativa que se basa en la recolección de datos sin medición numérica, como la observación, la descripción, las reflexiones culturales, como una cualidad del objeto de investigación.

Este trabajo de investigación emplea a la vez una investigación explicativa que se enfoca en el objeto de estudio a través de la relación entre variables para indagar las causas de los problemas y no únicamente a partir de una correlación estadística, construyendo así, teorías y agregando valor a las predicciones y a los principios científicos. Cazaú [11] señala que las investigaciones explicativas suelen tener más trascendencia a la hora de explicar el problema estableciendo la naturaleza de la relación entre una o más variables dependientes o independientes. Con el diagnóstico del problema se gestiona la búsqueda de información científica relacionada con las variables para demostrar la hipótesis propuesta, para ejecutar y posteriormente instaurar una relación teórica entre las variables de selección [12].

La investigación, en el estado actual del conocimiento con enfoque mixto es explicativo y correlacional. Los métodos cuantitativos de la investigación correspondieron a:

- **Teórico-analítico-sintético** que determinan los niveles con relación a indicadores sociales, económicos y académicos.
- **Teórico lógico-histórico** analiza la evolución histórica sobre el rendimiento académico en los estudiantes de educación general básica superior y bachillerato de la Unidad Educativa.
- **Teóricos por modelación** descomposición operacional sobre asignación de cualidades con relación a indicadores sociales, económicos y académicos.
- **Empírico por medición** comprueba los indicadores sociales, económicos y académicos.

Los paradigmas que sustentan esta metodología se originan de supuestos paradigmáticos distintos de la teoría del conocimiento. Este trabajo investigativo se enmarca en un paradigma naturalista con un enfoque cualitativo, en donde se quiere formar al ser humano. Este paradigma puede aparecer como el adecuado debido a la alta complejidad de los fenómenos educativos que no siempre serán mediciones numéricas.

Hipótesis

Si se desarrolla un modelo de análisis para evaluar el rendimiento académico de los estudiantes de la Unidad Educativa Personas Con Escolaridad Inconclusa Monseñor Leonidas Proaño de la ciudad de Latacunga, a través del uso de técnicas de minería de datos, entonces se podrá contribuir con información adecuada para la toma de decisiones y el diseño de estrategias educativas.

Definición de variables:

Variable dependiente (RA): rendimiento académico.

Variable independiente (A1): factores de rendimiento académico.

CAPÍTULO I. FUNDAMENTACIÓN TEÓRICA

1.1 Antecedentes.

Las primeras investigaciones del rendimiento académico comienzan aproximadamente a finales del siglo 19 e inicios del siglo 20 con estudios que se enfocaron únicamente aspectos cognitivos produciendo varias teorías. A mediados del siglo 20 investigadores revelan la importancia de los componentes afectivos y su influencia en el aprendizaje y el rendimiento académico; pero ya a fines del siglo veinte se conjuga los dos aspectos el cognitivo y el afectivos dando como resultados nuevos conocimientos, teorías y modelos, como el constructo aprendizaje autorregulado [16]. El rendimiento académico ha sido un tema de conocimiento e investigación por más de 70 años especialmente en Europa, evidenciando tendencias y teorías de donde se derivan hipótesis, y a partir de éstas hipótesis se realiza la investigación. En el ámbito científico pedagógico la teoría es considerada también como modelo que explica, predice y domina diversos fenómenos de la realidad, capaz de exponer predicciones correctas y al menos verificables [17].

Uno de los primeros trabajos de investigación sobre el estudio de la predicción del rendimiento académico aparece por Secadas [18] en México, en el año 1952, y es en aquella época cuando comienza a producir interés en el tema. En el año 1964 el trabajo de García [19] sobre el rendimiento académico como estudio de predicción escolar. En la década de los setenta los estudios de predicción escolar empiezan a tener base estadística que ayudan a determinar varias conductas. Posteriormente Nieto [11] en su libro “Hacia una teoría sobre el rendimiento académico en enseñanza primaria a partir de la investigación empírica: datos preliminares” compila y crea una tabla de distribución de frecuencias de investigaciones sobre el rendimiento académico desde el año 1970 al 2007, con una exposición descriptiva

de los resultados recolectados en la base de datos Redinet con un registro de 648 trabajos, donde se descarta 96 al no considerarlo como investigación, sino catalogados como estudios o reflexiones teóricas. En la tabla 2 se observa la clasificación y descripción en función de las cinco categorías establecidas en el sistema educativo.

Tabla 2: Distribución de frecuencia de investigaciones sobre rendimiento académico. Periodo 1970 – 2007

Períodos temporales	Primaria		Secundaria		Universidad		Pri/Secu.		Secu/Univ.		TOTAL	
	N	%	N	%	N	%	N	%	N	%	N	%
1970-1980	16	50,0	8	25,0	8	25,0	0	0,0	0	0,0	32	100.0
1981-1985	53	56,0	14	15,0	19	20,0	8	8,4	1	1,1	95	100.0
1986-1990	82	53,2	39	25,3	23	14,9	10	6,5	0	0,0	154	100.0
1991-1995	45	48,4	21	22,6	18	19,4	6	6,5	3	3,2	93	100.0
1996-2000	19	24,1	37	46,8	19	24,1	4	5,1	0	0,0	79	100.0
2001-2007	25	25,3	45	49,5	24	24,2	4	4,0	1	1,0	99	100.0
TOTAL	240	43,5	164	29,6	111	20,1	32	5,8	5	1,0	552	100.0

Fuente: Santiago Nieto [17].

Como podemos observar es en el periodo 2001 al 2007 donde se produce un despunte de trabajos del rendimiento académico a nivel secundario con 45 investigaciones que refleja un 49,5%. Aporta además el autor con conclusiones generales pero esenciales en la influencia y las causas del rendimiento académico en los cinco niveles estudiados, citándolos sin orden de importancia: el autoconcepto, el origen social, la capacidad intelectual, los factores emocionales, el factor socioeconómico y la adaptación escolar. Hay que considerar que, durante estos años suscitan cambios y reformas al currículo escolar por lo que podría haber inclinación de investigación a ciertas categorías variando así la cantidad de trabajos de investigación en cierto periodo. Nieto [17] concluye precisando que al revisar los trabajos de investigación del rendimiento académico no va más allá de crear teorías limitadas, de pronosticar tendencias que podrían ser no ciertas, pero poseen una presunta verdad; y propone trabajar en teorías puntuales que guíen la investigación hacia problemas específicos dentro de un marco efímero y en evolución.

Es necesario entender al rendimiento académico en su concepto y Jiménez [20] lo define como el sistema que mide logros y construye conocimiento en los estudiantes a través de didácticas, métodos cualitativos y cuantitativos que son evaluados. Guerrero, Cardona y Cuevas [21] el conocimiento es un fenómeno complejo de evaluar y expresarlo en calificaciones por ser resultado de múltiples factores extraescolares que se refiere a la situación socioeconómica del estudiante y factores intraescolares propios del sistema educativo. Erazo [22] señala que el rendimiento académico ya no está determinado ni vinculado únicamente a pasar de año por las calificaciones y promedios académicos sino porque existen factores complejos inmersos, cambiándolo a una situación fenomenológica.

1.2. Fundamentación epistemológica.

El rendimiento y desempeño académico o escolar son sinónimos que hacen referencia al mismo concepto, por lo contrario, el rendimiento académico no es sinónimo de aprovechamiento escolar, ya que el aprovechamiento escolar se define como el resultado del proceso enseñanza - aprendizaje, donde el nivel de éxito es responsabilidad del docente como del estudiante. Según el diccionario etimológico de Corominas [23] el término rendimiento académico aparece en textos escritos por el año 1580 como palabra derivada del verbo rendir. La enciclopedia de Pedagogía/Psicología El Tewab [24], publicada en el año de 1997 define al rendimiento académico como una relación entre lo obtenido y el esfuerzo empleado en obtenerlo. De igual manera, el autor Andrade [25] define al rendimiento académico como la medida de las capacidades que una persona ha aprendido como consecuencia de un proceso o formación, pero aun habiendo recibo un eficiente proceso de formación académica, se puede ver afectado por factores como malos hábitos de estudio, bajo aprovechamiento de actividades dentro del aula, el uso ineficiente de recursos, entre otros.

Ruíz [26] considera al bajo rendimiento académico como la discapacidad en materias instrumentales como la lectura, la escritura y el cálculo en estudiantes considerado como un fenómeno que posiblemente termine en fracaso escolar.

Existen varias investigaciones de carácter educativo que están orientadas a explicar sobre el rendimiento académico que van desde estudios exploratorios, descriptivos y explicativos que nos permiten entender y descubrir las variables asociadas al éxito o fracaso académico. Glasser [27] define al rendimiento académico con características bajas como el reflejo del fracaso de los hogares, y su contexto social, cultural y económico, exhortando a las personas involucrada en el proceso educativo y sociedad a responsabilizarse, propiciando un adecuado sistema escolar que pretenda ser exitoso y comprobable. Bricklin y Bricklin [28] realizan la investigación con estudiantes de una escuela elemental donde descubre que el factor cooperación y la apariencia física de los estudiantes son aspectos determinantes del rendimiento académico, así lo entienden los docentes, considerándolos como más inteligentes y de mayor rendimiento académico.

La importancia y necesidad del estudio del rendimiento académico, durante varios años se incluye y aplica tecnologías de la información que han permitido analizar, explicar y definir características y perfiles propios. Estos perfiles definen teorías y conclusiones que aplicados a un análisis inteligente de datos es posible aplicar minería de datos, eligiendo adecuadas herramientas y metodología para crear modelos de análisis informático que ayude probablemente a establecer estratégicas para mejorar el rendimiento académico en el sistema educativo secundario. Por otra parte, Fayyad [29] puntualiza que el análisis inteligente de datos es un proceso no trivial para identificar patrones válidos, novedosos, potencialmente útiles y comprensibles a partir de datos, para lo cual, Han et al. [30] determinan este proceso de análisis inteligente de datos en cinco fases interactivas e iterativas: integración y recopilación de datos, pre procesamiento de datos, minería de datos, evaluación e interpretación y difusión y uso de modelos, añadiendo que la minería de datos es una fase interdisciplinaria que permite predecir salidas e indagar las relaciones entre los datos provocando nuevos conocimientos.

Liñán y Pérez [31] determina que el estudio de minería de datos en la educación es similar a inteligencia de negocios y analítica de datos, de ahí su importancia al aplicar en el campo educativo, como lo afirma Alejandro Ballesteros Román, 2013

como un medio para descubrir modelos, tareas, métodos y algoritmos que explota datos proveniente del contexto educativo, con la finalidad de encontrar patrones de comportamiento y resultados para mejorar el rendimiento académico en los estudiantes [32]. Romero y Ventura [33] relacionan a la minería de datos en la educación con otras áreas como se observa en el gráfico 1, evidenciando la relación de minería de datos para la educación con otras áreas, determinando a la minería de datos educativa como una disciplina que desarrolla y aplica métodos explotando datos de repositorios que proceden del entorno educativo de diversas fuentes convencionales, sistemas informáticos, capacitación en línea o evaluaciones que continuamente proveen datos para ser analizados y determinar nuevas respuestas del comportamiento y cambios actitudinales de un grupo de estudiantes.

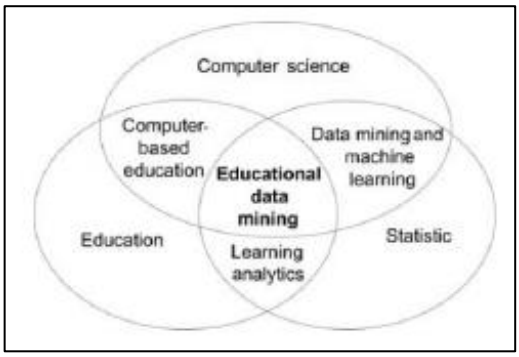


Gráfico 1: Principales áreas relacionadas con minería de datos para educación
Fuente: Cristóbal Romero [33]

Liñán y Pérez [31] señalan que la aplicación de la minería de datos en el campo educativo se debe a la necesidad de optimizar factores como:

- Mejorar el proceso de toma de decisiones basado en datos generados como se realiza en inteligencia de negocios y analítica de datos.
- Encontrar patrones en los datos y construir modelos de predicción de fácil adaptación.
- Almacenar y procesar la gran cantidad de datos académicos en tecnologías de computación vigentes.
- Reducir el abandono y pretender mejorar la calidad en la educación, a través de predicciones.

La Minería de datos educativa comparada con otras áreas presenta similitud en las metodologías, sin embargo, contiene cierta diferencia en el dominio, la data y el objetivo que pretende el proceso educativo. Papamitsiou y Economides [34] identifican las fortalezas, oportunidades, debilidades y amenazas (FODA) de la Minería de datos educativa como se observa en la tabla 3 y 4, en intersección con áreas como la educación, psicología, pedagogía y computación, generando información siendo analizada, toda acción de aprendizaje puede ser asilada, identificada y clasificada para crear patrones.

Tabla 3: Matriz FODA. Minería de Datos en Educación

Fortalezas	Debilidades
• Gran volumen de datos disponibles. • Uso de algoritmos poderosos y validados ya existentes. • Múltiples formas de visualizar la data. • Modelos más precisos de los usuarios para la mejora y personalización de los sistemas. • Obtener una visión de las estrategias de aprendizaje y sus resultados.	• Errores en la interpretación de los resultados debido a factores humanos. • Fuentes de datos heterogéneos. No existe todavía un estándar para los datos. • Los resultados en su mayoría son cuantitativos. Los métodos cualitativos no han brindado resultados significantes. • Sobrecarga de información. Sistemas complejos. • Incertidumbre debido a que solo los docentes o instructores con cierto nivel de habilidades pueden interpretar correctamente los resultados.
Oportunidades	**Amenazas**
• Estandarización de la data y mejora de compatibilidad entre las diferentes aplicaciones y herramientas. • Aprendizaje multimodal y afectiva. • Capacidad de auto-aprendizaje en sistemas inteligentes y autónomos. • Integración de los resultados obtenidos con otros sistemas de toma de decisiones. • Modelo de aceptación, describiendo usabilidad, expectativas, confiabilidad, entre otros.	• Aspectos éticos como privacidad de los datos. • Sobre-análisis. • Posibilidad de errores en la clasificación de patrones. • Confiabilidad: resultados contradictorios durante la implementación de modelos ya establecidos

Fuente: Papamitsiou y Economides [34]

Tabla 4: Aplicación de Minería de datos

Tarea	Objetivos/Descripción	Aplicaciones claves
Predicción	Inferir la variable objetivo desde la combinación de otras variables. Se aplican técnicas de clasificación, regresión y estimación de densidad.	Predicción de rendimiento académico y detectar comportamiento de estudiantes
Agrupamiento	Identificar grupos con características similares	Crear grupos de estudiantes con características similares en su patrón de interacción y aprendizaje.
Minería de relaciones	Estudio de relaciones entre variables y descubrimiento de reglas	Identificar relaciones en patrones de aprendizaje y diagnosticas dificultades en aprendizaje.
Destilación de datos	Representación de datos de manera más inteligible mediante resumen, visualización e interfaces interactivos.	Brindar herramientas de apoyo a los instructores para visualizar y analizar actividades relacionadas con sus estudiantes.
Descubrimiento de anomalías	Detección de individuos con diferencias significativas	Detección de estudiantes con dificultades o con aprendizaje irregular
Minería de texto.	Extracción de información de calidad desde los datos.	Análisis de contenido de foros, chats, páginas web y documentos.
Seguimiento de conocimiento	Estimar el dominio de habilidades de estudiantes, utilizando modelos cognitivos y los registros de la evaluación como evidencia.	Monitorear el nivel alcanzado por los estudiantes en el tiempo.
Factorización de matriz no negativa	Definición de matrices en base a resultados de evaluación que se descomponen en otras matrices que representan habilidades obtenidas	Evaluación de habilidades.

Fuente: Romero y Ventura [33]***, Shahiri*** [35]

Kotsiantis, Patriarchas y Xenos [36] afirman que a partir del año 2000 se puede encontrar trabajos investigativos donde se aplica minería de datos mediante técnicas estadísticas usando el análisis de regresión y correlación; posteriormente se conoce que la técnica más apropiada de minería de datos en educación es la predicción ya que crea modelos predictivos a través de clasificación, categorización y regresión. Romero [21] y Shahiri [35] ratifican que efectivamente la técnica más utilizada en minería de datos en educación es la predictiva mediante la clasificación y los algoritmos más aptos aplicarse es el árbol de decisiones, redes neuronales, bayesiano, K-nearest neighbor y Suppor Vector Machine, los cuales Young y Chávez [37] manifiestan que estos han permitido identificar factores importantes como los antecedentes académicos, el entorno social, la situación económica, la psicología, la demografía y la cultura, afectando al rendimiento académico de los estudiantes.

Es importante exponen algunos ejemplos representativos de trabajos del rendimiento académico utilizando técnicas estadísticas. Wang y Newlin [24] realizan el estudio a estudiantes de educación a distancia a través de cursos en línea utilizando regresión y correlación para predecir el rendimiento académico e identificar los factores que permiten pronosticar alumnos que permanecerán, continuarán y concluirán exitosamente sus estudios. Por otra parte, Martínez [38] recurre al análisis de funciones discriminantes para determinar e identificar los atributos para predicen la culminación exitosa de los estudiantes. Y finalmente Araque, Roldán y Salguero [39] aplicaron modelos de regresión lógica para determinar el riesgo de abandono y afirmar si el ocio es uno de los factores de abandono de los estudios.

Ahora algunos ejemplos de investigaciones del rendimiento académico utilizando minería de datos es aplicado por Veitch [40] quien realiza el análisis usando la técnica del árbol de decisión, para predecir las características de los estudiantes que abandonan el bachillerato en la educación secundaria. Superby, Vandamme, y Meskens [41] recurren a algoritmos como redes neuronales y árboles de decisión, clasificando a los estudiantes en categoría de bajo, medio y alto riesgo de reprobar para predecir si terminarán sus estudios secundarios exitosamente. Estos estudios e investigaciones López y Santín [42] se apoyan aplicando métodos de análisis matemático utilizando diferentes herramientas informáticas para descubrir patrones y tendencias que describen y comprenden mejor los datos y predecir eficazmente comportamientos futuros.. En los últimos años ha sido notorio el uso progresivo de herramientas de minería de datos que han contribuido para crear modelos en todo ámbito, por lo que es meritorio citar y definir conceptos de las herramientas más usadas en minería de datos en la educación:

Knime

Berzal et al. [43] Knime es un software libre programado en Java y desarrollado sobre la plataforma Eclipse. Este software es sencillo de manipular, y una de las mejores opciones por su diseño de tablas y gráficos interactivos.

Jhepwork

Maco [44] lo define a Jhepwork como un marco de trabajo de código abierto escrito en Java con la funcionalidad de analizar datos como una herramienta competitiva a herramientas comerciales. Su principal característica es multiplataforma.

Weka.

Desarrollada en la Universidad de Waikato, tiene la característica de ser gratuita, implementada en Java puede ser ejecutada en cualquier plataforma, a más de poseer técnicas de reprocesamiento, clustering, clasificación, regresión, visualización, selección de datos y modelado, su aprendizaje es muy rápido.

Esta herramienta tiene una amplitud de acceso a base de datos con SQL y conexión JDBC (Java DataBase Connecticity).

R-Commander.

Castillo [45] la define como una herramienta para el análisis estadístico que permite ingresar código propio para obtener resultados. Diseñado inicialmente por Robert Gentleman y Ross Ihaka quienes accedieron a aceptar modificaciones con código abierto para su desarrollo bajo ciertas condiciones de uso de licencias.

Rapidminer.

Maco [44] conocido anteriormente como YALE, es una herramienta de software libre, de código abierto de rápido aprendizaje, abarca minería de texto y análisis predictivo. En el año 2011 fue definida como la más utiliza ya que permite resumir el análisis y el uso de la información. Generalmente es utiliza en aplicaciones comerciales y de negocios, así como para investigación, educación, capacitación, creación rápida de prototipos y desarrollo de aplicaciones, y respalda los pasos del proceso de aprendizaje, la visualización de resultados, la validación y la optimización del modelo.

Para la predicción del rendimiento académico es fundamental elegir la metodología adecuada ya que permitirá realizar ordenadamente los procesos de minería y entender el proceso de descubrimiento de conocimiento. Dentro de las

metodologías más adecuadas y utilizadas por una gran cantidad trabajos de investigación se puede citar las siguientes.

Cross Industry Standard Process for Data Mining (CRISP-DM)

Hernández [44] se refiere a esta metodología donde se organiza por etapas designando tareas específicas y generales para cumplir el objetivo. Su funcionalidad es cíclica, con la facilidad de regresar desde alguna etapa a otra anterior como se observa en el gráfico 2.

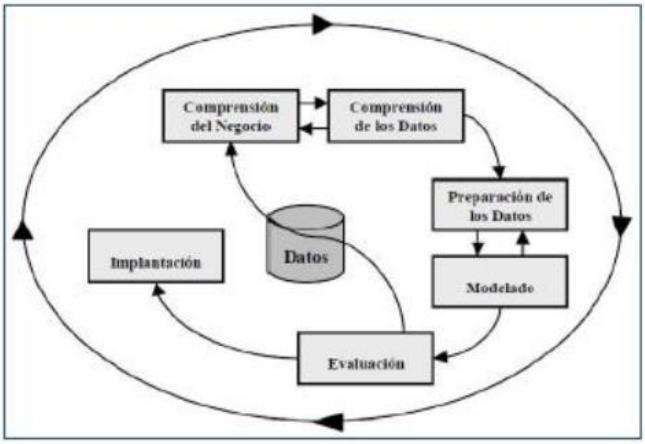


Gráfico 2: Fase del modelo CRISP-MD
Fuente: Hernández [44] "CRISP-MD 1.0: Step by step Data Mining guide

Cada una de estas etapas cumplen las siguientes funciones:

- Comprensión del negocio: determina los objetivos y requerimientos desde una perspectiva de negocio.
- Comprensión de los datos: elige y ordena los datos, para identificar las dificultades en la calidad de datos y obtener datos viables para el análisis.
- Preparación de los datos: selecciona los datos que van a una fase de limpieza, estructuración, integración y formateo.
- Modelamiento y evaluación: selecciona la técnica, edificando el modelo, para posteriormente ser sujeto a pruebas y evaluaciones.
- Despliegue del proyecto: muestra el modelo para intégralos en los procesos de toma de decisión.

Sample Explore Modify Model Access (SEMMA)

Moine, Haedo y Gordillo [46] es una metodología desarrollada por el Statistical Analysis Systems (SAS) Institute orientada a los aspectos técnicos minimizando el análisis y comprensión de la investigación. Esta metodología fue desarrollada para aplicarla sobre la herramienta de minería de datos SAS Entreprise Miner. Santos y Azevedo [47] muestran a la metodología SEMMA en un ciclo con cinco etapas para procesos de explotación de información como se muestra en el gráfico 3.

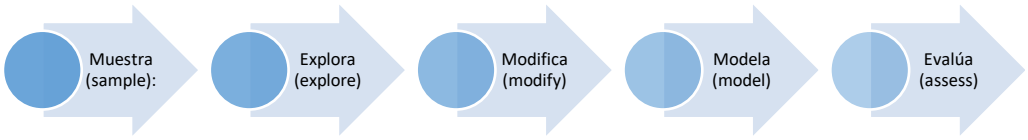


Gráfico 3: Fase del modelo SEMMA
Fuente: (SAS Institute, 2010)

Cada una de estas etapas cumplen las siguientes funciones:

- Muestra (sample): muestrear los datos para poder manipular
- Explora (explore): explorar los datos mediante la búsqueda de tendencias imprevistas y anomalías.
- Modifica (modify): modifica los datos mediante la creación, selección y transformación de variables.
- Modela (model): modela los datos con la ayuda de un software para predecir confiablemente el resultado.
- Evalúa (assess): evalúa los datos y estima su funcionalidad.

Knowledge Discovery in Databases (KDD)

Han et al. [30] afirma que para la extracción del conocimiento Knowledge Discovery in Databases (KDD) descubre conocimiento e información útil en los datos de repositorios de información que al explorados se puede determinar relaciones. Este proceso extrae información de calidad pudiendo ser usada para definir conclusiones basadas en relaciones o modelos. Las etapas consideradas en esta metodología se pueden ver en el gráfico 4.

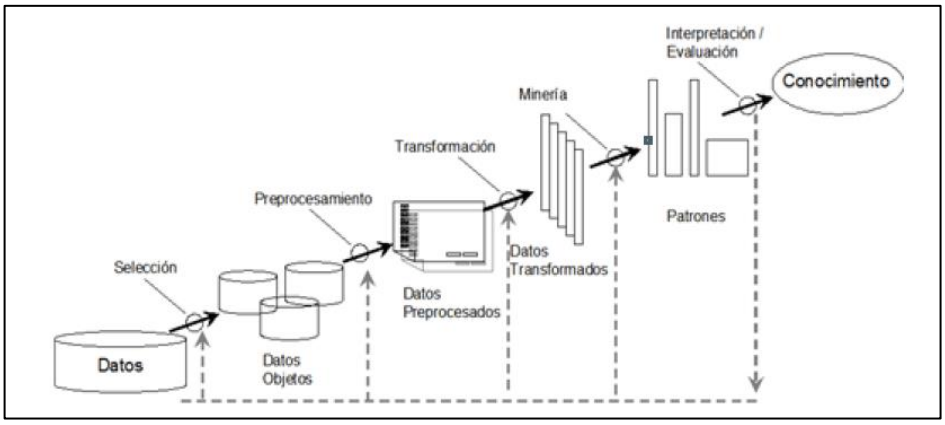


Gráfico 4: Knowledge Discovery in Databases (KDD)
Fuente: Han, J y Kamber [30]

Las funciones que cumplen cada una de estas etapas son las siguientes:

- Selección de datos: determina las fuentes de datos y el tipo de información.
- Pre procesamiento: preparación y limpieza de los datos extraídos, usando estrategias para manejar datos ausentes o inconsistentes.
- Transformación: es el tratamiento de los datos para generar variables apropiadas desde la existente sometidas a normalización.
- Minería de datos: aplica los métodos inteligentes para extraer patrones ocultos en los datos.
- Interpretación y evaluación: identifica los patrones con mayor relevancia evaluándolos.

Y es precisamente la metodología KDD que fue considerada como metodología específica para la construcción del modelo a través de minería de datos**.**

Después de las definiciones anteriores, la modelación aparece como el resultado de todo el proceso del análisis inteligente de datos. Mosterín y Estany [48] remontándose a la historia señalan que tras la segunda guerra mundial aparece ya el término modelo y empieza a ser usado con prioridad en las ciencias como la lógica, matemática e informática, posteriormente y no menos importante en las ciencias naturales como la física, química, biología entre otras. Jiménez [14]. en su libro Base de datos relacionales y modelado de datos, define al modelo de datos como un conjunto de herramientas conceptuales que explica los datos, sus relaciones, límites de integridad que afecta.

Shahiri [35] en la misma línea predictorias crea modelos a través de técnicas como la clasificación, regresión y categorización, destacando que la técnica más aplicada en educación efectivamente es la clasificación usando algoritmos como el árbol de decisiones, redes neuronales, bayesiano, entre otros. A continuación, se expone los principales trabajos investigativos de autores que han creado modelos de minería de datos en la educación desde diversos escenarios educativos. Por otra parte, Rojas [49] en su estudio de la afectividad educativa, como una consecuencia para el bajo rendimiento académico de los estudiantes dentro de la institución educativa, define como un factor relevante que influye directamente en el aprovechamiento académico de los estudiantes, visto desde la perspectiva de la motivación, el afecto y la inteligencia intrapersonal. En el Ecuador, como en Colombia y Argentina el autor refiere a la afectividad dentro de las aulas con el propósito de mejorar y establecer contextos cálidos en la comunidad educativa.

Esta investigación ha determinado que en los tres países analizados se constata de ambientes hostiles que causan deterioro en el desempeño estudiantil y la apremiante necesidad de establecer entornos de afecto positivos dentro del aula para un mejor desenvolvimiento académico conduciéndolos a un desarrollo integral del estudiante. Concluyendo que la educación es un proceso que va desde lo sentimental y emocional hasta lo racional. Mayancela [50] refiriéndose a los factores que originan un bajo rendimiento académico y la deserción del sistema escolar los agrupa en dos marcos interpretativos: *el extraescolar*, refiriéndose a la situación socioeconómica y al contexto familiar y *el intraescolar* que se refiere a la baja motivación para el estudio por la cosmovisión y las expectativas no satisfechas. Esta investigación basada en documentación de campo y bibliográfica, determina que existe un leve aumento de la deserción debido a la desintegración familiar por la migración, el ambiente escolar y el nivel de preparación profesional de los docentes.

En conclusión, la mayor parte de los estudios de minería de datos analizados en la educación secundaria se observa aplicaciones que detecta qué es lo que hay que hacer para que el estudiante mejore su rendimiento académico durante el proceso

aprendizaje, pero no está por demás citar realidades del análisis como la idea de riesgo al fracaso antes de iniciar el proceso educativo, apareciendo como posibles soluciones del rendimiento académico para tratar el problema de la deserción.

1.3. Fundamentación del estado del arte.

El historial de las publicaciones de minería de datos sobre el rendimiento académico en la educación secundaria, permite comprender los avances realizado a lo largo de estos cinco últimos años, por lo que se ha revisado el estado del arte. En los años noventa, países de Latinoamérica y principalmente México, comienza a promover investigaciones sobre bajo rendimiento académico en estudiantes, es así como Álvarez [51] en su investigación con un grupo de estudiantes de la promoción 2004 del Colegio de bachilleres del estado de Baja California, identifica y determina que 7 de cada 10 estudiantes enfrentan abandono, deserción y no aprobación del periodo escolar, realidad que ha predominado durante varios años. Cascón [52] en su estudio predictores del rendimiento académico se refiere al factor psicopedagógico de la inteligencia como algo predictivo, investigación que induce a revisar literatura sobre las inteligencias múltiples y aprendizajes escolares para entender de mejor manera el desarrollo de habilidades, capacidades y tipos de inteligencias.

En España, Muñoz [53] presenta un modelo para mejorar el rendimiento académico en alumnos de la Enseñanza Secundaria Obligatoria (ESO) mediante minería de datos, partiendo de la información personal y académica en particular de la didáctica de las matemáticas con el enfoque ontosemiótico, construye un sistema que detecta elementos e indicadores aportando información de trayectorias didácticas y mostrar si existe dispersión para análisis del departamento didáctico para mejorar o reorientar acciones en el proceso académico y contribuir su aprendizaje en otras asignaturas del bachillerato. Este autor pretende mejor el rendimiento académico a partir de la información académica, para detectar elementos que sirvan a la didáctica mejorar el proceso enseñanza – aprendizaje en las asignaturas de lenguaje y matemática e introducir el aprendizaje obtenido para trabajar sobre experiencias acumuladas y construir modelos que muestran el

comportamiento de los estudiantes, además de generar modelos descriptivos para explorar y comprender los datos iniciales e identificar patrones, relaciones y dependencias.

Huerta et al. [54] menciona que en México las evaluaciones nacionales de logro académico a nivel nacional de los estudiantes de bachillerato reportan que el 50 % tienen un nivel insuficiente en habilidades lectoras y de comunicación, así también un 60,7% en habilidades matemáticas Aparece una deserción alta por el bajo rendimiento académico, un 43% en estudiantes entre 15 y 19 años no estudia; por lo que es importante hacer un estudio de los factores que influyen el rendimiento académico. Aplicando técnicas de minería de datos para estudiar el impacto de las actividades diarias en el rendimiento académico de los estudiantes, para lo cual se aplicaron encuestas a 208 estudiantes y para el proceso de selección de atributos y clasificación de estudiantes según su rendimiento se aplicó WEKA 7,0. descubriendo patrones de aprobación y reprobación del nivel escolar. El algoritmo de clasificación utilizado RETree presentó resultados con un 84% de exactitud.

Knowles [55] menciona que la empresa Wisconsin Information System for Education Data Dashboard (WISEdash) crea una plataforma con un portal gráfico usando Dashboard, presentando datos educativos de los periodos escolares de las escuelas distritales de Wisconsin donde recoge información de estudiantes, docentes y cursos para ser almacenados en una data warehouse sirviendo como fuente de reportes y acceso a datos. Esta data educativa ha servido como repositorio para realizar análisis de los datos, aplicando técnicas y métodos de minería con la finalidad de hallar información oculta y útil. Piscoya [56] en su investigación de aplicación de técnicas de minería de datos para predecir la deserción en la educación básica regular en la región de Lambayeque Perú, presenta información de predicción de los estudiantes matriculados, obteniendo resultados a corto plazo para la toma de decisiones, aplicando técnicas predictivas con algoritmos de ETS y Redes Neuronales. Determinando que las redes neuronales auto regresiva presenta mejor confiabilidad en el nivel primario con un 91% y 96% en el secundario, con un promedio de tiempo mínimo en la obtención de datos.

Timarán, Hidalgo y Caicedo [57] realizan la investigación en Colombia para determinar el rendimiento académico a través de las pruebas Saber, aplicando en primera instancia cuadros estadísticos, sin evidenciar las verdaderas interrelaciones ocultas, posteriormente aplica minería de datos a la investigación mostrando resultados obtenidos basado en árboles de decisión para pronosticar modelos asociados al desempeño académico de los años lectivos 2015 y 2016, partiendo de datos académicos y socioeconómicos de la data del ICFES aplicando la metodología CRISP-MD, la herramienta WEKA y su algoritmo J48. Esta investigación concluye determinando que es mayor el porcentaje de estudiantes colombianos que tienen un desempeño académico bajo, comparado con el porcentaje de estudiantes que tienen un buen desempeño, definiendo que los atributos de mayor relevancia en los patrones descubiertos asociados al bajo desempeño académico, están en el estrato socioeconómico bajo. [58].

Ledesma [59] manifiesta que estos modelos, permiten al docente redirigir estrategias y adaptaciones curriculares, para evitar un posible fracaso y optando a tomar la decisión más oportuna. La exteriorización de los resultados y conclusiones de los estudios de cada una de las investigaciones que se han realizado a lo largo de los últimos diez años, han permitido visualizar los posibles patrones más próximos en el campo educativo secundario en el ámbito del rendimiento académico para poder adaptarlas, según su contextos y particularidades como posible solución para mejorar o perfeccionar el rendimiento académico con la toma de decisiones en cada una de las instituciones, gestionando así una verdadera educación de calidad.

Conclusiones Capítulo I.

- La revisión bibliográfica ha permitido observar que no existe un acercamiento consensuado sobre un método o algoritmo más adecuado a la hora de predecir las causas del rendimiento académico.
- Los conceptos y postulados de minería de datos en la educación como área multidisciplinaria permiten encontrar información útil, prediciendo comportamientos futuros como la finalidad de mejorar los procesos educativos relacionados con la enseñanza y aprendizaje.

- Al revisar la fundamentación del estado del arte de los modelos de análisis de datos a través de minería de datos, se observa la carencia de un registro historial de métodos o algoritmos más destacados para predecir el bajo rendimiento académico y su influencia del fracaso escolar a nivel secundario.
- La mayoría de investigaciones relacionados con este tema, están enfocados al nivel de educación universitaria y son muy escasos a nivel de educación general básica superior y bachillerato.

CAPÍTULO II. PROPUESTA.

2.1 Analítica descriptiva de la población estudiada.

La Unidad Educativa de modalidad semipresencial de educación extraordinaria, P.C.E.I. Monseñor Leonidas Proaño de la ciudad de Latacunga, es una institución fiscomisional alineada al acuerdo ministerial número 00040-A (ver anexo 1) referido al currículo integrado con los niveles de Educación General Básica Superior (EGB-S) y Bachillerato General Unificado (BGU), como también el Bachillerato Técnico (BT) con cuatro especialidades: Producción Agropecuaria, Industrias de la Confección, Instalaciones, Equipos y Máquinas Eléctricas y Administración y Contabilidad. La matriz de la Unidad Educativa está ubicada en la ciudad de Latacunga, cuenta con infraestructura propia y el permiso y autorización de funcionamiento del Ministerio de Educación del Ecuador, desde donde administra diez extensiones o denominados también Centros de Apoyo Tutorial (CAT´s) de la provincia de Cotopaxi, específicamente en: La Maná, Sigchos, Salcedo, Pujilí, Zumbahua, El corazón, Pucayacu, Moraspungo, Chugchilán y Maliguapamba.

La actividad académica está sujeta a la base legal de la estructura de los aportes académicos o calificaciones del art. 186 numeral 2 de la Ley Orgánica de Educación Intercultural (LOEI) referente a la evaluación formativa realizada durante el proceso de aprendizaje (ver anexo 2)**.** La misma que permite obtener resultados del cumplimiento de los objetivos de aprendizaje establecidos en el currículo y en los estándares de aprendizaje nacionales, a través de la escala de calificaciones cuantitativas según el art. 194 de la LOEI (ver anexo 3). La unidad educativa no ha podido obtener información digital sistematizada del proceso educativo, aún más desperdiciando la oportunidad de estudiar las causas de fenómenos o

acontecimientos que generan los datos que permiten descubrir posibles causas de varias problemáticas propias de la educación secundaria, obteniendo de manera tradicional información académica que se constata al final del primer quimestre con los informes, actas, documentos y reportes de los estudiantes desertores o no promovidos. Estos informes académicos y libros de actas de cada curso al ser revisados y analizados, en los tres últimos periodos académicos aparecen observaciones y resoluciones escritas por parte de los docentes de los aspectos más relevantes tales como:

- Registro de estudiantes matriculados que nunca asistieron a las tutorías presenciales
- Notificación de inspección de un registro alto de faltas injustificadas
- Alto porcentaje de estudiantes que tempranamente abandonan o desertan.
- Actas de retiro voluntario del estudiante por problemas económicos, de trabajo o violencia intrafamiliar.
- Actas de retiro voluntario de las estudiantes por maternidad.
- Informes de bajo rendimiento en evaluaciones parciales y quimestrales.
- Incumplimiento de tareas y trabajos en las tutorías presenciales.
- Actas de condicionamiento por faltas injustificadas y bajo rendimiento.

Toda esta información se estanca entre las hojas de los libros de actas privándose de descubrir información útil y valiosa, aún más, cuando la Unidad Educativa no está considerada para utilizar el portal virtual de servicios educativos Educar Carmenta del Ministerio de Educación del Ecuador. Por tal razón para tener una mejor descripción informativa de los datos de la población se ha desarrollado un sistema informático de ambiente web (ver anexo 4), que facilita la digitalización de datos, los mismo que ha sido desarrollado con el lenguaje de programación PHP v7 y la base de datos MariaDB. (ver anexo 5)**.** El sistema informático propuesto se desarrollado sigue el patrón de diseño Modelo-Vista-Controlador y para su diseño se recurre al framework Bootstrap lo cual le proporciona características responsivas, así mismo se ha empleado PHP por ser un lenguaje de programación

diseñado exclusivamente para el desarrollo de aplicaciones de ambiente web que en su versión 7.0 presenta mejoras en cuanto a factores de rendimiento y seguridad. Este sistema informático ha permitido ingresar registros de calificaciones de los estudiantes de los periodos lectivos 2016-2017, 2017-2018 y 2018-2019 (ver anexo 6), así como también los datos personales denominada ficha estudiantil (ver anexo 7) y datos socioeconómicos denominada ficha socioeconómica (ver anexo 8) con parámetros direccionados por la Psicóloga Anita Rocafuerte responsable del Departamento de Consejería Estudiantil (DECE). Reflejando una estadística de evidente preocupación de deserción que cada año se refleja en un alto porcentaje, situación que no ha disminuido por lo menos en los tres últimos periodos tal como se presenta en los Gráfico 5,6 y 7.

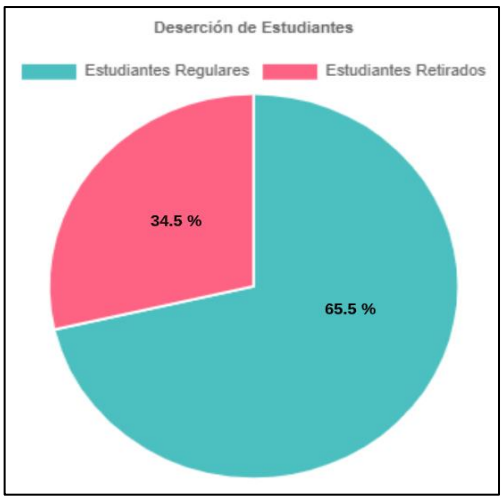


Gráfico 5: Periodo lectivo 2016-2017.
Fuente: Sistema Informático.

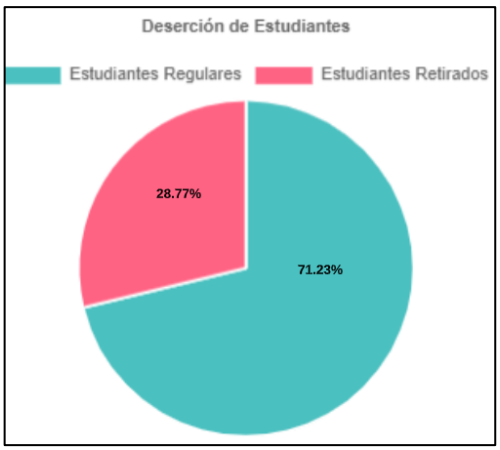


Gráfico 6: Periodo lectivo 2017-2018.
Fuente: Sistema Informático.

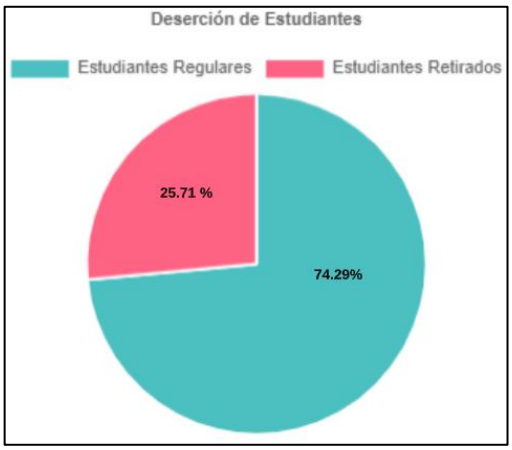


Gráfico 7: Periodo lectivo 2018-2019
Fuente: Sistema Informático

Inicialmente se obtiene datos de los estudiantes en los tres periodos lectivos, los mismo que están identificados en tres grupos de posibles factores que influyen el rendimiento académico. En la tabla 5 se presentan los factores y datos.

Tabla 5. Factores de influencia.

Factores	Datos
Personal	Datos de identificación Datos familiares Referencias familiares Datos de salud y vivienda Datos académicos/Historial
Académico	Rendimiento académico Calificaciones. Asistencia Comportamiento
Socioeconómico	Estatus económico Trabajo del estudiante Dependencia económica Situación laboral de los padres Nivel educativo delos padres Entorno económico

Elaborado por: El Investigador.

La tabla 6 presenta la estadística descriptiva de la población de un total de 350 estudiantes obtenida de la base de datos del sistema informático.

Tabla 6. Estadística descriptiva de la población estudiada.

Campo	Condición	Valor	Porcentaje
Calificaciones	≥ 7	254	73%
	< 7	96	27%
Ausencia a clases	> 8	101	29%
	< 8	249	71%
Computador	Si	25	7%
	No	325	93%
Discapacidad familiar	Si	8	2%
	No	342	98%
Etnia	Indígena	78	22%
	Mestiza	272	78%
Profesión madre	Profesión	7	2%
	Oficio	343	98%
Profesión padre	Profesión	50	14%
	Oficio	300	86%
Profesión tutor	Profesión	52	15%
	Oficio	298	85%
Formación padre	Primaria	271	77%
	Secundaria/superior	79	23%
Formación madre	Primaria	294	84%
	Secundaria/superior	56	16%
Formación tutor	Primaria	296	85%
	Secundaria/superior	54	15%
Ausencia padre	Si	86	25%
	No	264	75%
Ausencia madre	Si	21	6%
	No	329	94%
Número hermanos	≥ 2	182	52%
	< 2	168	48%
Lugar ocupa familia	Primogénito	115	33%
	No primogénito	235	67%
Estructura familiar	Disfuncional	104	30%
	Funcional	246	70%
Estado civil	Casado	68	19%
	Soltero/unión libre	282	81%
Discapacidad estudiante	Si	5	1%
	No	345	99%
Lugar de residencia	Latacunga	212	61%
	Fuera de Latacunga	138	39%
Número hijos	Si	46	13%
	No	304	87%
Vivienda	Si	190	54%
	No	160	46%
Servicios básicos (luz)	Si	349	100%
	No	1	0%
Repetición	Si	85	24%
	No	265	76%
Servicios básicos (ss.hh.)	Si	349	100%
	No	1	0%
Teléfono convencional	Si	64	18%
	No	286	82%

Tabla 7. Estadística descriptiva de la población estudiada. (Continuación)

Teléfono celular	Si	331	95%
	No	19	5%
Internet	Si	46	13%
	No	304	87%
Conducta	A,B	343	98%
	C,D	7	2%
Dificultades académicas	Si	158	45%
	No	192	55%
Servicios básicos (agua)	Si	349	100%
	No	1	0%
Actividad trabajo	Si	251	72%
	No	99	28%
Cabeza de familia	Si	142	41%
	No	208	59%

Fuente: Sistema Informático.
Elaborado por: El Investigador.

Población y Muestra.

Se presenta una población total de estudiantes de la Unidad Educativa PCEI Monseñor Leonidas Proaño de Latacunga matriculados en los niveles EGB-S, el BGU y el BT de los periodos lectivos de los años 2016 – 2017, 2017 – 2018, y 2018 – 2019 con el total de 1051 estudiantes clasificados y representados de la siguiente manera:

Tabla 8. Número de estudiantes por periodos lectivos.

Periodo Lectivo	N° Estudiantes
2016 – 2017	360
2017 – 2018	351
2018 – 2019	340
TOTAL	1051

Fuente: Secretaria e Inspección de la Unidad Educativa PCEI Monseñor Leonidas Proaño.
Elaborado por: El investigador

Para obtener la muestra se aplica la fórmula 1 cuando se conoce el total de la población y deseamos saber cuántos del total de la población tenemos que estudiar Pita [60].

$$n = \frac{\mathrm{N} * {Z_a}^2 * \mathrm{p} * \mathrm{q}}{d^2 * (N-1) + {Z_a}^2 * p * q} \quad (1)$$

Fórmula 1. Muestra de la población
Fuente: Pita

Donde:

n = tamaño de la muestra; N = total de la población (1051); Za = valor obtenido mediante niveles de confianza (1.96); d = precisión (4,3); p = proporción esperada (0,5) que maximiza el tamaño de la muestra y q = 1 – p (0,5).

Aplicando los valores en la fórmula se obtiene el siguiente resultado:

$$n = \frac{1051 * (1.96)^2 * 0.5 * 0.5}{(0.43)^2 * (1051 - 1) + (1.96)^2 * 0.5 * 0.5}$$

$$\boldsymbol{n = 347,8}$$

2.2 Analítica descriptiva de los datos

La información histórica estructurada obtenida del sistema informático de la unidad educativa relacionada con los datos de los estudiantes en factores como el rendimiento académico, la situación familiar y socioeconómico, se aplicaron técnicas de extracción, transformación y carga de datos a hojas de cálculo de Excel para crear el dataset con un formato apropiado para la aplicación de técnicas de minería de datos. Creada la dataset, se procedió a cargar la información al software Statistical Package for Social Sciences (SPSS) para tabular, análisis y probar la consistencia de los datos. Una vez ingresada la dataset a SPSS se realizó la estadística descriptiva que puede ser de dos maneras, Timarán et al. [57] una estadística de frecuencia o una estadística descriptiva, son formas de describir las propiedades de las distribuciones como la tendencia central, la posición, la dispersión y la forma. En nuestro caso de estudio se genera una estadística descriptiva de la dataset para analizar la dispersión a través del mínimo, máximo, media, desviación típica y varianza. En la tabla 8 generada por la herramienta SPSS, se visualiza la estadística descriptiva de la dataset para establecer valores cercanos.

Tabla 9. Estadística Descriptivo.

N°		Mínimo	Máximo	Media	Desviación Típica	Varianza
V1	350	0	1	0,72	0,449	0,202
V2	350	0	1	0,71	0,452	0,204
V28	350	0	1	0,07	0,254	0,065
V17	350	0	1	0,02	0,141	0,020
V5	350	0	1	0,22	0,414	0,172
V6	350	0	1	0,02	0,131	0,017
V7	350	0	1	0,14	0,346	0,12
V8	350	0	1	0,15	0,357	0,128
V9	350	0	1	0,22	0,418	0,175
V10	350	0	1	0,16	0,368	0,136
V11	350	0	1	0,15	0,36	0,13
V12	350	0	1	0,24	0,427	0,183
V13	350	0	1	0,94	0,239	0,057
V14	350	0	1	0,52	0,5	0,25
V15	350	0	1	0,33	0,47	0,221
V16	350	0	1	0,3	0,458	0,209
V4	350	0	1	0,19	0,391	0,153
V18	350	0	1	0,01	0,107	0,011
V19	350	0	1	0,6	0,49	0,24
V20	350	0	1	0,13	0,336	0,113
V21	350	0	1	0,54	0,499	0,249
V22	350	0	1	1	0,054	0,003
V30	350	0	1	0,24	0,429	0,184
V24	350	0	1	1	0,054	0,003
V25	350	0	1	0,18	0,384	0,147
V26	350	0	1	0,94	0,233	0,054
V27	350	0	1	0,13	0,336	0,113
V3	350	0	1	0,98	0,141	0,02
V29	350	0	1	0,45	0,498	0,248
V23	350	0	1	1	0,054	0,003
V31	350	0	1	0,71	0,454	0,206
V32	350	0	1	0,41	0,492	0,242
N° válido.	350					

Fuente: SPSS.

Confiabilidad de los datos.

Para valorar la solidez interna de la data se realizó la prueba de fiabilidad a través del coeficiente Alfa de Cronbach del software SPSS, con un resultado del 0,86 de un total de 350, valor aceptable dentro del rango según Oviedo y Campos [61].

Tabla 10. Resumen del procesamiento de los casos

		N°	%
Casos	Válidos	350	100
	Excluidos[a]	0	0
	Total	350	100,0

Fuente: SPSS

Tabla 11.Estadístico de fiabilidad

Alfa de Cronbach	N° de elementos
0,86	32

Fuente: SPSS

2.3 Diseño del modelo teórico de rendimiento académico.

Según Granados [62] las técnicas o métodos de regresión están consideradas por el tipo y la funcionalidad de las variables. La regresión lineal admite una relación entre dos variables determinadas en una forma lineal, idóneo para entender fenómenos relativamente complejos. Por otra parte Camacho, López y Arias [63] mencionan que las técnicas de regresión lineal pretenden conseguir la descripción de la relación entre variables y predecir los valores de una variable en función de otra, con la finalidad de ser representadas a través de diagramas de dispersión o nube de puntos. Esta técnica nos permitió constituir un modelo lineal en donde se forma la relación entre la variable dependiente y las variables independientes, entendiéndose que la relación de estas dos variables sigue una misma línea recta. En el caso particular de la construcción de este modelo se utilizó la regresión lineal para crear un modelo del rendimiento académico.

La construcción del modelo estadístico como lo muestra la tabla 11, permitió determinar a través de los datos históricos, académicos y socioeconómicos, los cuales son factores de mayor inciden e influencia positiva o negativa en el rendimiento académico de los estudiantes de la unidad educativa. Las causas por las cuales la tasa de retención es baja, para lo cual se utilizó técnicas de predicción como la regresión lineal que permitió determinar la influencia de las variables independientes sobre la variable dependiente, y lo más notable se determinó el nivel de significancia de las variables a través de un indicador estadístico que es la probabilidad, con un nivel de confianza de la técnica de 95% con un valor de p-value de 0,05.

Tabla 12. Modelo inicial de rendimiento académico estudiantil.

Dependent Variable: V28 Method: Least Squares Date: 04/18/20 Sample: 1 350 Included observations: 350				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
V3	0.000000	3.25E-16	0.000000	10.000
V30	4.83E-15	1.23E-15	3.933.418	0.0001
V31	-1.03E-16	3.13E-16	-0.327956	0.7432
V4	-4.78E-15	1.44E-15	-3.328.593	0.0010
V5	6.85E-16	3.11E-16	2.201.971	0.0284
V6	-4.38E-16	9.99E-16	-0.438400	0.6614
V7	9.02E-16	3.67E-16	2.457.947	0.0145
V8	-1.69E-16	3.67E-16	-0.459937	0.6459
V1	-4.57E-16	2.95E-16	-1.549.653	0.1222
V10	-5.53E-16	3.62E-16	-1.526.207	0.1279
V11	1.53E-15	4.73E-16	3.223.687	0.0014
V13	-2.89E-16	3.01E-16	-0.960819	0.3374
V14	-2.80E-17	3.05E-16	-0.091930	0.9268
V15	6.82E-16	3.33E-16	2.046.726	0.0415
V16	4.40E-15	8.14E-16	5.402.730	0.0000
V17	-2.34E-16	4.39E-16	-0.531815	0.5952
V18	1.28E-15	2.66E-16	4.818.117	0.0000
V19	6.75E-16	4.32E-16	1.563.424	0.1189
V2	2.74E-17	2.59E-16	0.105888	0.9157
V20	-6.76E-16	2.81E-16	-2.404.039	0.0168
V22	2.07E-14	1.28E-15	1.614.611	0.0000
V23	1.24E-16	3.24E-16	0.381039	0.7034
V24	-6.17E-17	3.63E-16	-0.169837	0.8652
V25	-2.61E-16	3.26E-16	-0.800826	0.4238
V26	-1.79E-15	5.71E-16	-3.140.067	0.0018
V27	1.65E-15	6.61E-16	2.495.895	0.0131
V29	1.45E-15	2.83E-16	5.110.886	0.0000
C	1.000.000	1.97E-15	5.09E+14	0.0000
Mean dependent var	1.000.000	S.D. dependent var		0.000000
S.E. of regression	2.31E-15	Sum squared resid		1.72E-27
Durbin-Watson stat	2.017.165			

Fuente: SPSS

Se determinó un modelo ajustado del rendimiento académico como lo muestra la tabla 12 en base a las variables que estadísticas dieron una influencia significativa con un valor a menos 0,05. Este proceso indica que estas variables son muy significativas para el diseño y construcción del modelo de rendimiento académico.

Estas variables influyen de manera positiva o negativa, es decir tienen un alta influyen.

Tabla 13: Modelo ajustado del rendimiento académico estudiantil

Dependent Variable: V28				
Method: Least Squares				
Date: 05/18/20				
Sample: 1 350				
Included observations: 350				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
V30	4.83E-15	1.23E-15	3.933.418	0.0001
V4	-4.78E-15	1.44E-15	-3.328.593	0.0010
V5	6.85E-16	3.11E-16	2.201.971	0.0284
V7	9.02E-16	3.67E-16	2.457.947	0.0145
V11	1.53E-15	4.73E-16	3.223.687	0.0014
V15	6.82E-16	3.33E-16	2.046.726	0.0415
V16	4.40E-15	8.14E-16	5.402.730	0.0000
V18	1.28E-15	2.66E-16	4.818.117	0.0000
V20	-6.76E-16	2.81E-16	-2.404.039	0.0168
V22	2.07E-14	1.28E-15	1.614.611	0.0000
V26	-1.79E-15	5.71E-16	-3.140.067	0.0018
V27	1.65E-15	6.61E-16	2.495.895	0.0131
V29	1.45E-15	2.83E-16	5.110.886	0.0000
C	1.000.000	1.97E-15	5.09E+14	0.0000
Mean dependent var	1.000.000	S.D. dependent var		0.000000
S.E. of regression	2.14E-15	Sum squared resid		1.54E-27
Durbin-Watson stat	1.996.507			

Fuente: SPSS

Residuos

López [64] los residuos son utilizados para comprobar si el modelo de regresión lineal es el adecuado, por lo que es útil hacer gráficos de los residuos para justificar la regresión. Los residuos representan un nivel variable partiendo de un rango de 0.0E con una tasa de incremento de + o - 0,05 a pesar que las desviaciones de la serie con respecto al nivel van continuas de variaciones.

El gráfico 9 presenta los residuos totales de las variables del modelo estadístico, donde se comprueba que existen residuos estandarizado e iguales a 10, 55 y 200 superior a la desviación estándar, existiendo valores atípicos, ya que son residuos tipificados mayores a tres en valor absoluto.

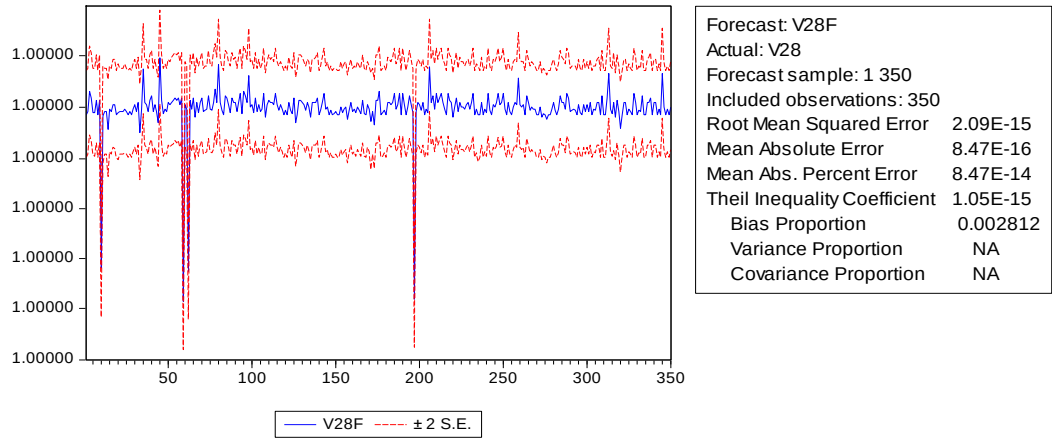


Gráfico 8. Residuos del modelo del rendimiento académico estudiantil.
Fuente: SPSS.

Mínimos Cuadrados Ordinarios (MCO)

Chirivella [65] considera al MCO como un método de estimación que realiza el ajuste del modelo de regresión lineal simple. Este método permite establecer una función lineal entre dos variables donde generalmente ***x*** representa a la variable dependiente y, por otro lado, ***y*** representa a la variable independiente. Hurtado [66] manifiesta que el MCO sirve para ajustar rectas a un conjunto de datos al aplicar la siguiente ecuación $y = mx + b$. Con base a lo que señalado por los autores se considera importante el análisis con MCO aplicado en las 13 variables del modelo que se muestra en el gráfico 10.

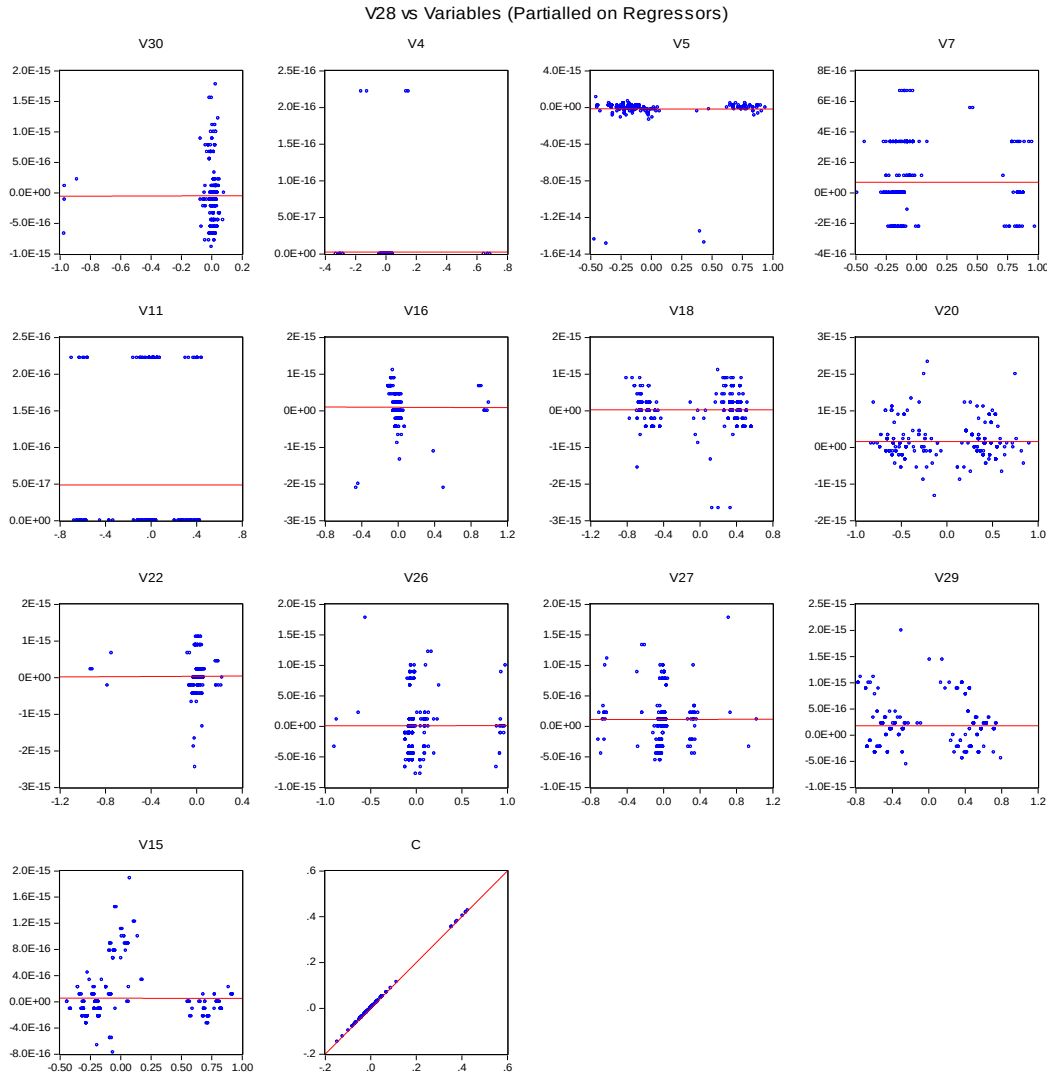


Gráfico 9. Mínimos ordinarios variables identificadas.
Fuente: SPSS.

Estimación del Modelo.

Para la estimación del modelo del rendimiento académico estudiantil, se aplica la técnica de regresión lineal para predecir los valores y la relación entre variables cuantitativas. La variable dependiente representada por ***x*** corresponde a la variable (V28) que hace referencia a dificultad académica y las variables independientes o predictoras corresponden a las variables identificadas en el modelo ajustado. La tabla 13 presenta la ecuación del modelo del rendimiento académico estudiantil propuesto y sus coeficientes.

Tabla 14. Estimación del modelo

Estimation Command:
LS V28 V30 V4 V5 V7 V11 V16 V18 V20 V22 V26 V27 V29 V15 C
Estimation Equation:
V28 = C(1)*V30 + C(2)*V4 + C(3)*V5 + C(4)*V7 + C(5)*V11 + C(6)*V16 + C(7)*V18 + C(8)*V20 + C(9)*V22 + C(10)*V26 + C(11)*V27 + C(12)*V29 + C(13)*V15 + C(14)
Substituted Coefficients:
V28 = 2.10094577692e-15*V30 - 4.14735520467e-15*V4 + 6.13496619995e-16*V5 + 7.3735651354e-16*V7 + 6.73041056068e-16*V11 + 4.07839070749e-15*V16 + 1.80602900616e-17*V18 - 2.85712529647e-16*V20 + 2.09704015828e-14*V22 - 7.96733880547e-16*V26 + 9.92390689167e-16*V27 + 9.75426220175e-16*V29 + 1.20458541376e-16*V15 + 1

Fuente: SPSS.

Variables del rendimiento académico estudiantil identificadas.

La tabla 14 contiene las variables que han sido identificadas a través del proceso experimental que sirvieron para la construcción del modelo teórico del rendimiento académico. Se presenta una breve descripción de las 13 variables.

Tabla 15: Descripción de las Variables del rendimiento académico estudiantil identificadas.

Id. Variable	Variables	Descripción
V30	Actividad de Trabajo	El estudiante mantiene un trabajo temporal o permanente para el sustento de su familia durante la semana.
V4	Estado Civil	Identifica si un estudiante es soltero y casado (unión libre).
V5	Etnia	Señala la etnia con la cual el estudiante se identifica .
V7	Profesión del Padre	Se refiere a la actividad laboral profesional que el padre de familia realiza.
V11	Ausencia padre	No tiene la figura paterna dentro de la estructura de la familia.
V15	Estructura Familiar	La identidad que la familia asume si es formalizada u otras formas de convivencia.
V16	Calificaciones	Se refiere a las calificaciones ~~notas~~ obtenidas en el proceso de evaluación formativa.
V18	Lugar de Residencia	Lugar donde vive el estudiante.
V20	Vivienda	Referida a si el lugar donde vive es propio o arrendado.
V22	Formación del Tutor/o representa te legal.	Se refiere nivel de preparación y formación académica que el tutor o representante legal tiene.
V26	Internet	Disponibilidad frecuente de acceso a internet.
V27	Computador	Disponibilidad frecuente de poseer o tener acceso a un computador.
V29	Repetición	Los años de repetición o no promovido en el proceso de educación en los niveles educativos.

Elaborado por: El Investigador

Hipótesis de Estudio.

En la tabla 15 se presentan las 13 variables con su respectivo código que han sido identificadas a través del proceso experimental con las cuales se plantean 13 hipótesis que sirvieron para la construcción del modelo teórico del rendimiento académico.

Tabla 16. Hipótesis del modelo de rendimiento académico estudiantil

Cod. Variable	Hipótesis
V30	H1: ¿El incremento en la actividad laboral del estudiante influye negativamente en la rendimiento académico?
V4	H2: ¿El estado civil influye positivamente en el rendimiento académico estudiantil?
V5	H3: ¿La condición étnica de la población influye negativamente en el rendimiento académico estudiantil?
V7	H4: ¿La profesión del Padre influye negativamente en el rendimiento académico estudiantil?
V11	H5: ¿La ausencia del Padre en el hogar influye negativamente en el rendimiento académico?
V15	H6: ¿La estructura familiar influye positivamente en el rendimiento académico del estudiante?
V16	H7: ¿Las calificaciones influye negativamente en el rendimiento académico?
V18	H8: ¿La distancia del lugar de residencia influye negativamente en el rendimiento académico?
V20	H9: ¿El vivir en un predio propio influye positivamente en el rendimiento académico estudiantil?
V22	H10: ¿La formación del académica del tutor/o representante legal influye positivamente en el rendimiento académico estudiantil?
V26	H11: ¿El limitado acceso a Internet influye negativamente en el rendimiento académico?
V27	H12: ¿La disponibilidad de un computador influye negativamente en el rendimiento académico estudiantil?
V29	H13: ¿La repetición de años escolares influye negativamente en el rendimiento académico?

Elaborado por: El Investigador

2.4 Modelo teórico del rendimiento académico estudiantil.

El modelo para determinar los factores que influyen en el rendimiento académico está conformado por trece los cuales se han graficado de la siguiente manera.

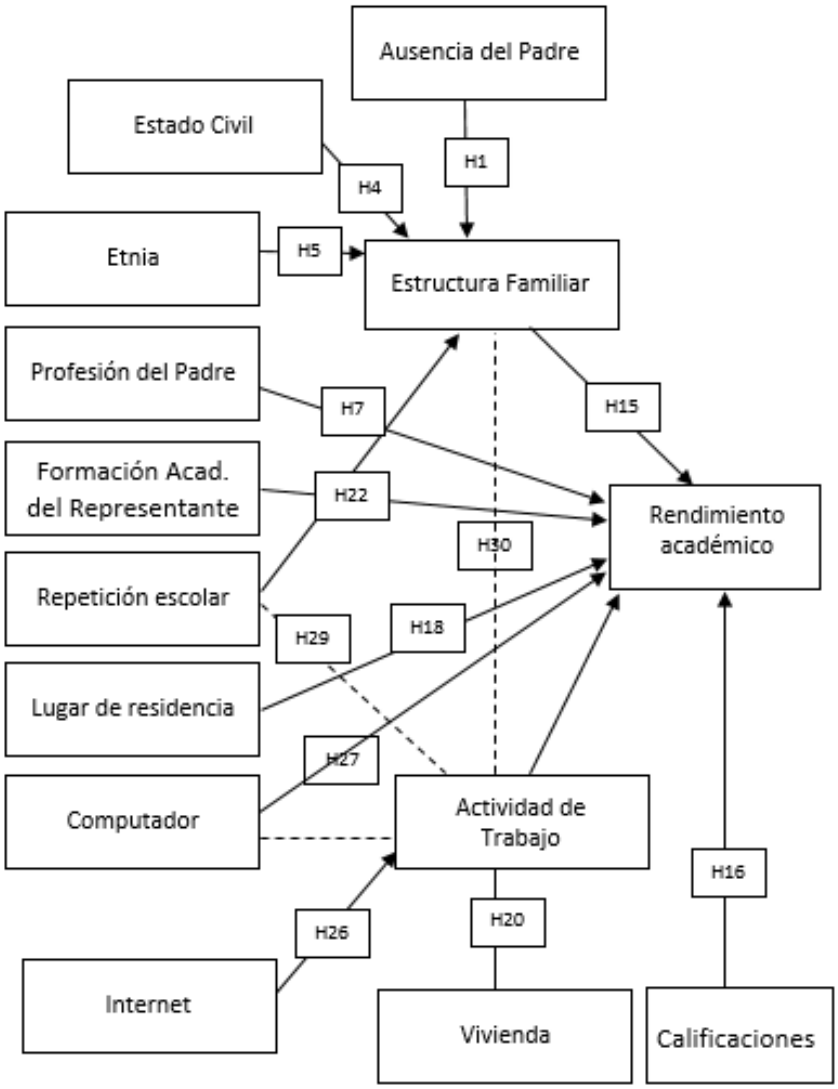


Gráfico 10. Modelo teórico del rendimiento académico estudiantil
Elaborado por: El Investigador.

El modelo conceptual propuesto se determinó que las trece variables obtenidas del proceso experimental estructura familiar y actividad de trabajo del estudiante tienen un efecto causal directo hacia el rendimiento académico estudiantil, estas dos variables aparecen como positivas. La variable computador tienen un efecto causal directo hacia el rendimiento académico.

Las variables lugar de residencia, formación del tutor y/o representante legal y profesión del padre también tiene también un efecto causal directo hacia el rendimiento académico. Mientras que la ausencia del padre de familia, el estado civil, etnia y repetición no presentan una relación causal directa con el rendimiento académico, más bien es un proxy a la estructura familiar. De igual manera repetición, computador, internet y vivienda, no presentan una relación causal directa con el rendimiento académico, más bien es un proxy a la actividad de trabajo del estudiante. Mientras tanto, la variable repetición de año presenta una relación causal hacia la variable estructura familiar y actividad de trabajo.

Conclusiones del Capítulo II.

- La validación del modelo se realizó aplicando la analítica en la herramienta SPSS, visualizando la relación de las variables.
- Se construyó el modelo estadístico a partir de los datos históricos, académicos y socioeconómicos, los cuales son los factores de mayor inciden e influencia positiva o negativa para lo cual se utilizó técnicas de predicción como la regresión lineal que permitió determinar la influencia de las variables independientes sobre la variable dependiente.
- Se determinó el nivel de significancia de las variables a través de un indicador estadístico que es la probabilidad, con un nivel de confianza de la técnica de 95% con un valor de p-value de 0,05.
- Se logró construir el modelo teórico del rendimiento académico estudiantil con técnicas estadísticas a partir de 32 variables de las cuales se ajustó el modelo en 13 variables que estadísticas dieron una influencia significativa con un valor a menos 0,05.

CAPÍTULO III. APLICACIÓN Y VALIDACIÓN DE LA PROPUESTA.

El objetivo de la investigación es proponer y desarrollar un modelo de análisis de datos del rendimiento académico y determinar la tasa de predicción del rendimiento académico estudiantil. El modelo desarrollado será probado en el caso de estudio Unidad Educativa PCEI Monseñor Leonidas Proaño de la ciudad de Latacunga. El cual se compone de cuatro fases que inician con la recolección de los datos familiares, socioeconómicos y académicos de los estudiantes, integración de la información desarrollada a través de un sistema de ambiente web para centralizar la información en una base de datos relación. La dataset parte de la base de datos del sistema informático y por último el análisis de las variables para lo cual se utiliza la herramienta Weka como software de minería de datos con lo cual se identifican los principales patrones que repercuten en el rendimiento académico a través de la aplicación de algoritmos decisión J48, random forest, naive bayes y OneR.

ETAPAS DE LA MINERIA DE DATOS.

3.1 Integración de los Datos

Balagueró [67] define al dataset como el conjunto de datos tabulados y comprendidos en una matriz de datos estadísticos. Para la integración de los datos se procedió a extraer los registros de la base de datos del sistema informático de los periodos lectivos 2016-2017, 2017-2018 y 2018-2019, que recoge factores como el social, académico y económico, de los estudiantes, la misma que fue agregada de manera satisfactoria, confiable y útil para la construcción del dataset, además se estructuró una matriz de Excel formada por un total de 1051 registros, adaptada y en algunos casos separando información incompleta o inadecuada.

Posteriormente se eligió una muestra, a través de casos de muestra aleatoria en SPSS, con un tamaño de muestra exacta de 350, cantidad que fue considerada del cálculo mediante de la población total. Este proceso de selección aleatoria generó identidad a cada registro con ceros y unos, la cual nos permitió selecciona los registros con el valor de uno y eliminando los registros con valor de cero, como se visualiza en el gráfico12.

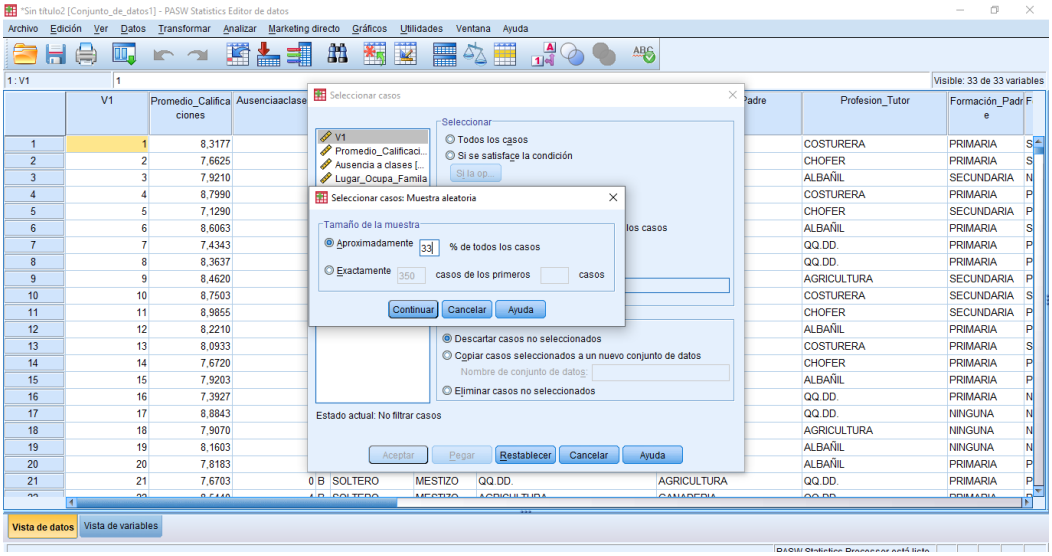


Gráfico 11. Muestra aleatoria
Fuente: SPSS

Para facilitar la ejecución de los algoritmos de minería, se transformó todos los registros de la dataset a número binario, designando entre 0 y 1.

3.2 Limpieza de los Datos

Timarán [57] considera muy importante esta etapa como un proceso de calidad en los datos, a través de aplicaciones e instrucciones para eliminar datos nulos, duplicados, remoción de ruidos (noisy data) y estrategias de manejo de datos desconocidos (missing empty). Los datos desconocidos (empy) no tienen un valor real, y los desaparecidos (missing) son valores no capturados. Los datos ruidos (noisy) en cambio son valores muy significativos que están fuera del rango. Gracias a estas instrucciones de la etapa de limpieza se reemplaza por valores más próximos con la utilización de métricas estadísticas como el mínimo y máximo, la media, la moda.

3.3 Pre procesamiento.

García [68] considera la selección de atributos de Weka para identificar atributos de un conjunto de datos libre de errores con variables lo más independiente considerando el peso asociado a aquellos valores característicos para definir a que clase pertenecen los datos. En Weka, la selección de atributos tiene varias opciones, para lo cual debemos seleccionar un método de búsqueda y de evaluación. Para la selección de las variables se aplicó el algoritmo GainRationAttributeEval, con el método Ranker. Al referirse a GainRationAttributeEval Ching [69] lo define como un algoritmo que evalúa la ganancia de información respecto a la dataset. Por su parte, Pascual et al. [70] señal que el método Ranker facilita la ordenación de atributos según su importancia. El resultado de este algoritmo y el método se presenta en la tabla 16.

Tabla 17. Selección de atributos

=== Attribute Selection on all input data ===	
Search Method:	
Attribute ranking.	
Attribute Evaluator (supervised, Class (nominal): 1 V5):	
Gain Ratio feature evaluator	
Ranked attributes:	
V4	0.045754
V30	.034907
V22	.034907
V16	0.020057
V7	.012481
V15	.007094
V20	.005399
V27	.005076
V18	.004344
V11	.001307
V26	.000655
V29	.000433
Selected attributes: 6,10,9,5,2,4,8,12,7,3,11,13 : 12	

Fuente: SPSS

3.4 Extracción del Conocimiento.

Timarán [57] esta fase busca y revelar patrones imprevistos significativos utilizando la clasificación y regresión. El modelo predictivo intenta considerar valores futuros de variables importantes, conocidas como variables dependientes, utilizando variables independientes o predictivas. Por otro lado, el modelo descriptivo reconoce patrones que explican y explora las propiedades de los datos; entre las tareas descriptivas encontramos los patrones secuenciales, clustering, reglas de asociación y las correlaciones, para lo cual se utilizó tres algoritmos: J48, random forest, naive bayes y OneR. Para aportar al conocimiento de las técnicas de minería de datos, se presenta en la tabla 17 la ficha técnica de cada uno de los cuatro algoritmos que se utilizaron en el proceso de predicción del modelo del rendimiento académico estudiantil.

Tabla 18. Ficha técnica de los algoritmos aplicados.

Método de Minería de datos	Algoritmo	Descripción	Autor(es)
Arboles	J48	J48 genera un árbol de decisión estadístico. Se escoge el atributo con mayor logro, normalizando como parámetro de decisión. Si las muestras de la lista pertenece a la misma clase se crea un nodo de hoja para el árbol.	Alemán et al. [78]
Bayes	Naive Bayes	Es usado para procesos de clasificación, ya que es muy efectivo y eficiente en la minería de datos. Trabaja con hipótesis y sus atributos son independientes entre sí, siempre que la variable se conozca.	Bedoya, López y Marulanda [79]
Árbol	Ramdom Forest	Este algoritmo utiliza métodos de clasificación supervisada, ya que es robusto y fácil de interpretar. Su función es realizar particiones sucesivas en el espacio de variables.	Castillo, García y Sarria [80]
Reglas	OneR	Este algoritmo selecciona el atributo que mejor revela la clase de salida. La característica propia del método de clasificación es la rapidez y buenos resultados.	Bedoya, López y Marulanda [79]

Elaborado por: El Investigador.

Los algoritmos elegidos para el proceso de predicción del modelo del rendimiento académico estudiantil se dividieron en dos partes, uno grupo de datos para la muestra de entrenamiento que corresponde al 80% (280 casos) y para la muestra de comprobación que corresponde al 20% (70 casos). Se aplicó la validación cruzada para evaluar los resultados del análisis estadístico. Para facilitar el conocimiento de los resultados se utilizó una matriz de evaluación que contiene una tasa de aciertos indicando el nivel de precisión de la clasificación de cada modelo, el estadístico de Kappa que muestra el coeficiente de concordancia de las variables, la f-measure que midió la confiabilidad del modelo y por último el área ROC, que indicó la exactitud global de la prueba. La aceptabilidad de los criterios evaluativos, determinaron que el valor máximo es de 1 y el mínimo valor de 0,50. Menacho [71] propone la escala de valoración de Kapa en la tabla 18.

Tabla 19. Valoración del coeficiente Kappa

Kappa	Grado de Concordancia
< 0,00	Sin acuerdo
> 0,00 – 0,20	Insignificante
0,21 – 0,40	Discreto
> 0,41 – 0,60	Moderado
0,61 – 0,80	Sustancial
0,81 – 1,00	Casi Perfecto

Fuente: Menacho [71]

Software Weka

Este software con licencia publica GNU de código abierto posee una interfaz gráfica de usuario. Tiene una colección de herramientas de visualización y algoritmos para el análisis de datos. En el gráfico 15 se presenta la herramienta Weka con el proceso de datos, clasificación, regresión, clustering, reglas de asociación y visualización. El proceso inicial en Weka se realizó definiendo un conjunto de 350 instancias y 31 atributos.

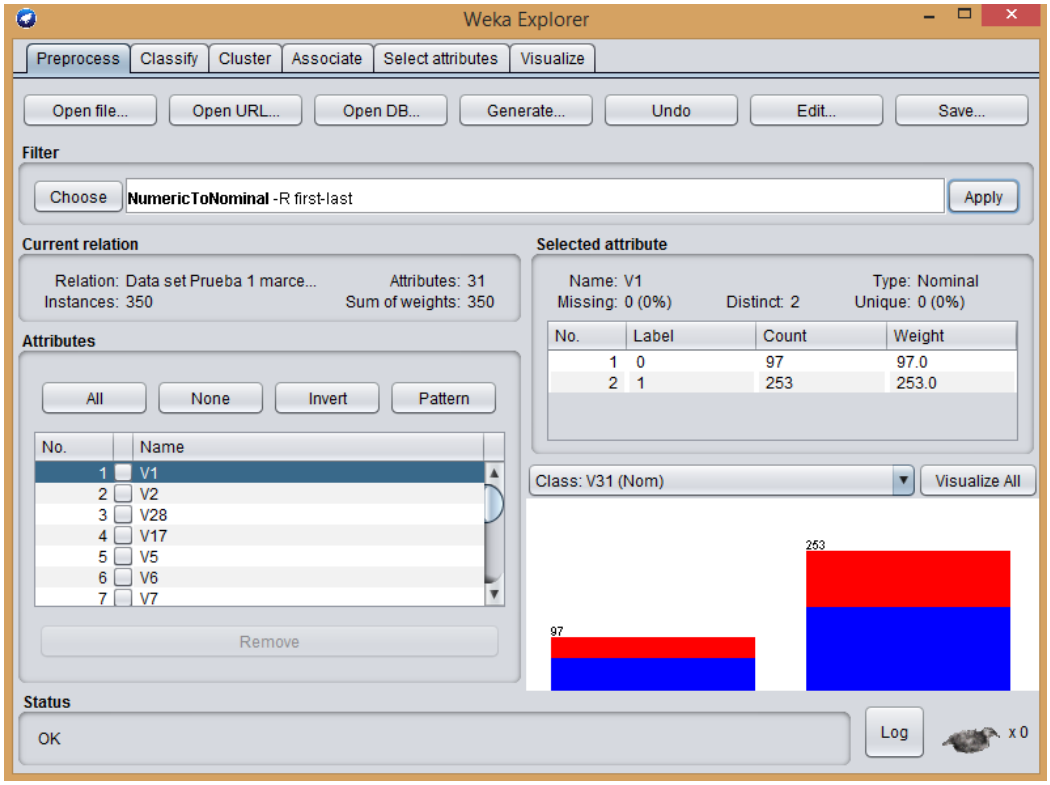


Gráfico 12. Dataset
Fuente: WEKA

Resultados

A continuación, se muestran la ejecución y resultados alcanzados por cada uno de los algoritmos aplicados para el modelo del rendimiento académico de la unidad.

ALGORITMO J48.

Sharini et al. [72] J48 es un algoritmo que utiliza la técnicas más comunes y accesibles a la hora de trabajar con predicción, por sus características sencillas y factibilidad de interpretar los resultados. Esta técnica tiene la facilidad de convertir un conjunto de datos en reglas de condición con el comando si-entonces. Pudiendo trabajar con variables numéricas y categóricas, algunos árboles más utilizados y construidos en base a algoritmos son el ID3 y C45. Hernández y Ferri [73] mencionan que una de las característica de esta técnica de minería es que trabaja con números finitos, es decir con una cantidad limitada de clases. Para el desarrollo del modelo del rendimiento académico se utilizó árboles de decisiones por ser una técnica de fácil y de sencilla interpretación. En el algoritmo J48 se visualiza los resultados del proceso de predicción en la tabla 19.

Tabla 20. Resultados del proceso de predicción mediante árboles de decisión Árbol J48

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: Data set Prueba 1 marcelo-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last
Instances: 350
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

```
J48 pruned tree
V30 = 0: 0 (209.0/6.0)
V30 = 1
|  V18 = 0
|  |  V15 = 0: 0 (70.0/6.0)
|  |  V15 = 1
|  |  |  V10 = 0: 0 (44.0/13.0)
|  |  |  V10 = 1: 1 (6.0/1.0)
|  V18 = 1
|  |  V24 = 0
|  |  |  V7 = 0
|  |  |  |  V11 = 0: 1 (2.0)
|  |  |  |  V11 = 1
|  |  |  |  |  V1 = 0: 0 (6.0/1.0)
|  |  |  |  |  V1 = 1
|  |  |  |  |  |  V8 = 0: 1 (4.0/1.0)
|  |  |  |  |  |  V8 = 1: 0 (3.0/1.0)
|  |  |  V7 = 1: 1 (2.0)
|  |  V24 = 1: 1 (4.0)

Number of Leaves  :  10
Size of the tree :  19
Time taken to build model: 0 seconds
```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	300	85.7143 %
Incorrectly Classified Instances	50	14.2857 %
Kappa statistic	0.1715	
Mean absolute error	0.1767	
Root mean squared error	0.3453	
Relative absolute error	81.3031 %	
Root relative squared error	105.1749 %	
Total Number of Instances	350	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,951	0,814	0,893	0,951	0,921	0,182	0,654	0,900	0
	0,186	0,049	0,348	0,186	0,242	0,182	0,654	0,254	1
Weigh.	0,857	0,720	0,826	0,857	0,838	0,182	0,654	0,820	

=== Confusion Matrix ===

a	b	<-- classified as
292	15	a = 0
35	8	b = 1

Fuente: WEKA.

En la tabla 19 el algoritmo inicia mostrando el nombre: J48, el nombre de la relación: Prueba 1, y el número de instancias en la relación: 350. Este modelo clasificador es un árbol de decisión en forma de texto que fue producido en los datos completos de entrenamiento. Como se puede ver, la primera división es en el atributo V30 en segundo nivel la división está en V18, en la estructura del árbol esta V15 y V24, seguido de las instancias V11 debajo de la estructura de del árbol hay 10 atributos, así como el número de nodos es 18. El resultado del algoritmo muestra que se ha clasificado 300 instancias correctas (86%) y 50 instancias incorrectas (14%) de un total de 350, lo que indica que el rendimiento académico se encuentra clasificado correctamente, además muestra el coeficiente del estadístico de Kappa con un grado de concordancia de 0,17 (insignificante).

La clase 0, presenta TP Rate (Tasa de verdaderos positivos) de 0,95 aciertos, una FP Rate (Tasa de falsos positivos) de 0,81, lo que se designa como un nivel de error elevado, una Precisión de 0,89 lo que indica un nivel de instancias correcta, clasificadas ligeramente bajo un Recall (cobertura) de 0,9. La clase 1, presenta TP Rate (Tasa de verdaderos positivos) de 0,18 aciertos, una FP Rate (Tasa de falsos positivos) de 0,04 lo que se designa como un nivel de error bajo, una Precisión de 0,34 lo que indica un nivel de instancias correcta, clasificadas ligeramente bajo un Recall (cobertura) de 0,18.

Según la matriz de confusión el algoritmo consideró lo siguiente: los números de instancias son los valores que encuentran de manera diagonal en la matriz, prediciendo como correctas y de un total de 350 instancias, J48 clasificó 300 instancias correctas. En el gráfico 14 se observa el árbol de decisión J48, en la que podemos ver los nodos más importantes y sus ramificaciones.

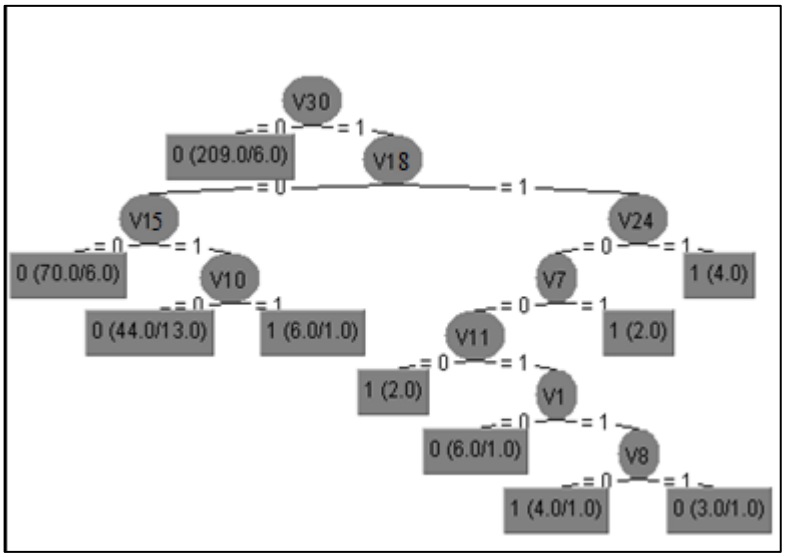


Gráfico 13. Tree visualizer.
Fuente: WEKA.

Análisis del costo/beneficio del modelo.

En el gráfico 15 se observa la compensación costo/beneficio del modelo de rendimiento académico. Esta metodología evalúa de forma exhaustiva determinando como lo indica el clasificador accurracy el modelo tiene una precisión de predicción del 87,71% lo cual aprueba la aplicación de esta técnica.

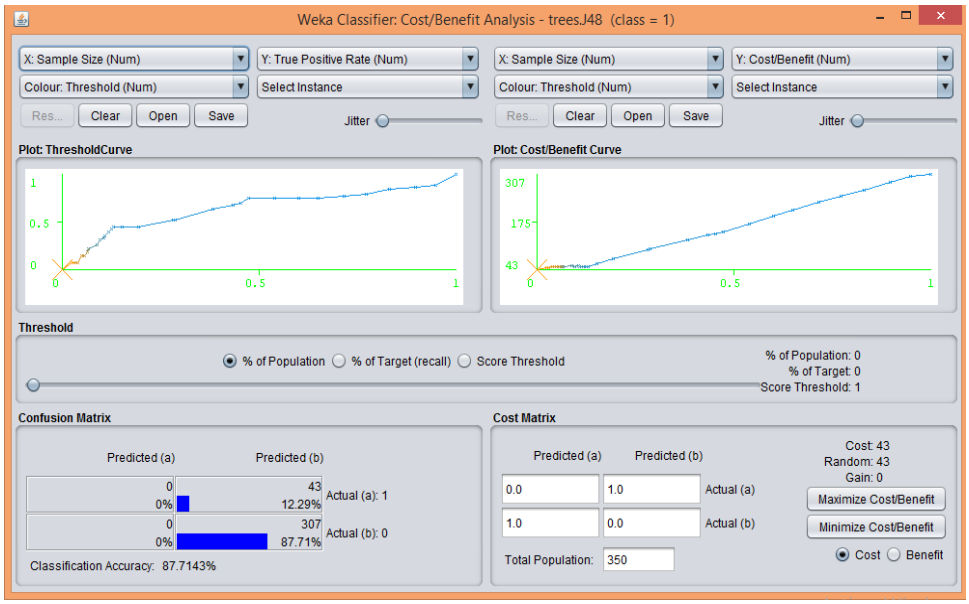


Gráfico 14. Cost/Benefit del algoritmo J48.
Fuente: WEKA.

ALGORITMO RANDOM FOREST

Breiman [74] la define como una técnica compuesta de predictores de árboles, para lo cual cada árbol depende de los valores que un vector. Su estructura presenta eficacia por la potencia que cada árbol presenta y en su conjunto, teniendo una estrecha correlación. Escobar [75] resalta la característica de esta técnica al correr de forma eficiente con bases de datos, aplicando el método experimental encontrando interacciones entre variables, permitiendo una estimación de las variables más importantes en la clasificación. Por lo que se consideró importante esta técnica en el caso de estudio por la interacción que realiza entre variables, y admitiendo una estimación de las variables más importantes. El algoritmo Random Forest (bosque al azar) muestra la información generada en la tabla 20.

Tabla 21. Resultados del proceso de predicción mediante el algoritmo Random Forest

=== Run information ===									
Scheme: weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -									
Relation: Data set Prueba 1 marcel0-weka.filters.unsupervised.attribute.NumericToNominal-									
Instances: 350									
Attributes: 31									
Test mode: 10-fold cross-validation									
=== Classifier model (full training set) ===									
RandomForest									
Bagging with 100 iterations and base learner									
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities									
Time taken to build model: 0.06 seconds									
=== Stratified cross-validation ===									
=== Summary ===									
Correctly Classified Instances	303	86.5714 %							
Incorrectly Classified Instances	47	13.4286 %							
Kappa statistic	0.0701								
Mean absolute error	0.1797								
Root mean squared error	0.3049								
Relative absolute error	82.6712 %								
Root relative squared error	92.8546 %								
Total Number of Instances	350								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,977	0,930	0,882	0,977	0,927	0,093	0,803	0,958	0
	0,070	0,023	0,300	0,070	0,113	0,093	0,803	0,337	1
Weigh.	0,866	0,819	0,811	0,866	0,827	0,093	0,803	0,882	
=== Confusion Matrix ===									
a b <-- classified as									
300 7 | a = 0									
40 3 | b = 1									

Fuente: Weka.

En la tabla 20 el algoritmo Random Forest muestra la relación: Prueba 1, el número de instancias: 350 y los atributos que contiene los datos y su identificador: 31. El resultado de este algoritmo muestra que se ha clasificado 303 instancias correctas (87%) y 47 instancias incorrectas (13%) de un total de 350, lo que indica un nivel de aciertos muy bueno, además muestra el coeficiente del estadístico de Kappa con un grado de concordancia bajo de 0,07 (insignificante). La clase 0, presenta TP Rate (Tasa de verdaderos positivos) de 0,97 aciertos, una FP Rate (Tasa de falsos positivos) de 0,93, lo que se designa como un nivel de error elevado, una Precisión de 0,88 lo que indica un nivel de instancias correcta, clasificadas ligeramente bajo un Recall (cobertura) de 0,97. La clase 1, presenta TP Rate (Tasa de verdaderos positivos) de 0,07 aciertos, una FP Rate (Tasa de falsos positivos) de 0,02 lo que se designa como un nivel de error bajo, una Precision de 0,3 lo que indica un nivel de instancias correcta, clasificadas ligeramente bajo un Recall (cobertura) de 0,07. Según la matriz de confusión el algoritmo consideró lo siguiente: los números de instancias son los valores que encuentran de manera diagonal en la matriz, prediciendo como correctas.

De un total de 350 instancias, Random Fores clasificó 303 instancias correctas y como incorrectas 47. En el gráfico 16, margincurve representa el margen de predicción como la diferencia entre la probabilidad pronosticada para la clase real y la probabilidad más alta.

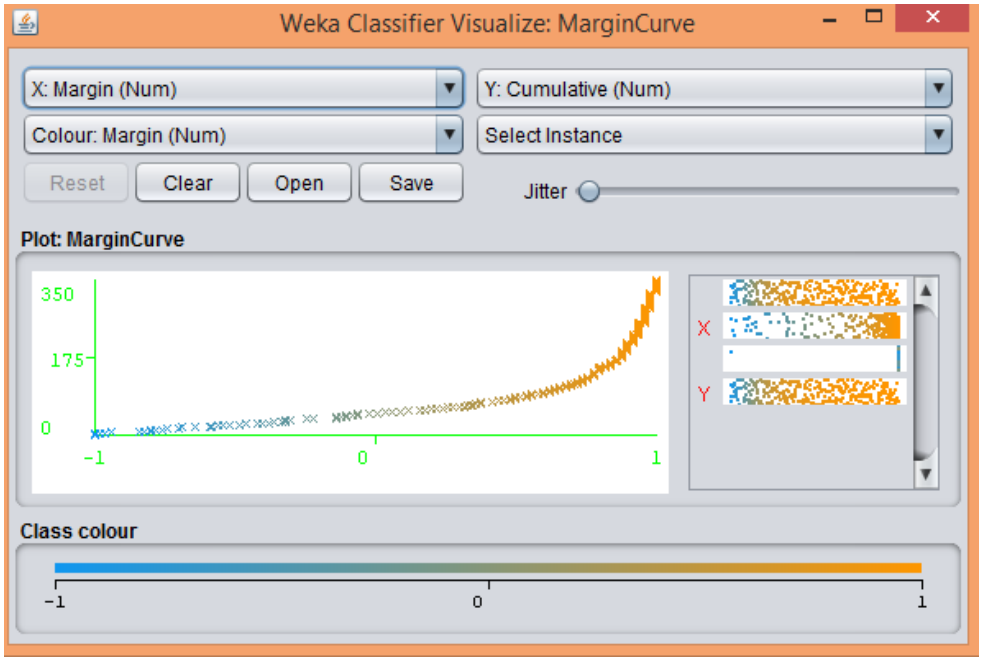


Gráfico 15. MarginCurve.
Fuente: Weka

Análisis del costo/beneficio del modelo

En el gráfico 17 se observa la compensación costo/beneficio del modelo con la metodología que evalúa de forma exhaustiva determinando el clasificador accurracy con una precisión de predicción del 87,71% lo cual aprueba la aplicación de esta técnica.

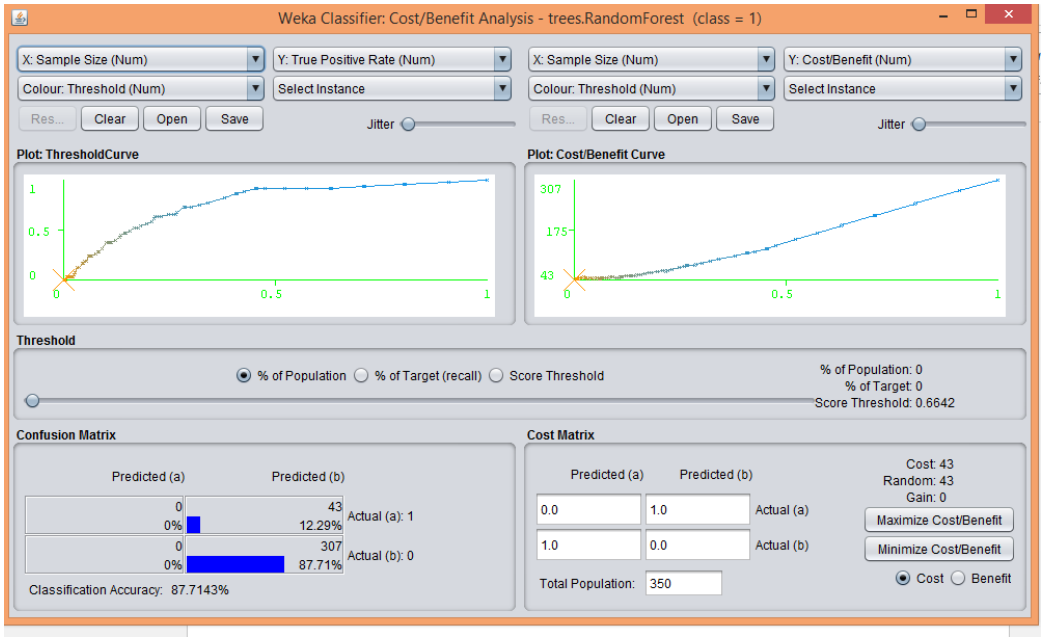


Gráfico 16. Cost/Benefit Random Forest,
Fuente: Weka

Es una gran combinación de árboles de decisión partir de los datos de entrada por lo que se altera el conjunto inicial de partida eligiendo diferentes variables, pero es mejor utilizar un conjunto de árboles a un solo árbol; esto provoca mayor estabilidad en el modelo permitiendo tener mayor acierto y garantizando un menor drawdown (mide el riesgo del sistema).

ALGORITMO NAIVE BAYES

Lemaire [76] en esta técnica se considera a las variables explicativas como independientes. Gironés et al. [77] define como una técnica de probabilidad condicionada de cada clase buscando combinar para maximizar las probabilidades. Hernández y Ferri [73] se refieren a este algoritmo con la particularidad del uso de distribuciones de probabilidad para cuantificar datos. Añade también [73] que a través de esta técnica se ha contribuido en la resolución de problemas relacionados a la inteligencia artificial. Mediante la estimación de las probabilidades de pertenencia, para lo cual se utiliza teoremas de Bayes. Según Beltrán [78] este algoritmo tiene la ventaja especial de ser aplicada en minería de datos para aprender en relaciones de dependencia y causalidad, combina conocimiento con datos, evita exagerado ajuste de los datos y acepta bases de datos incompleta. Esta técnica cuantifica las probabilidades acercándonos más a las posibles soluciones del modelo de rendimiento académico estudiantil para una buena toma de decisiones. Este algoritmo, se ha ejecutado con los siguientes resultados:

Tabla 22. Resultados del proceso de predicción mediante el algoritmo Naive Bayes

=== Run information ===
Scheme: weka.classifiers.bayes.NaiveBayes Relation: Data set Prueba 1 marcelo-weka.filters.unsupervised.attribute.NumericToNominal Instances: 350 Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
Naive Bayes Classifier Class Attribute 0 1 (0.88) (0.13)
==============================
V5 0 239.0 36.0 1 70.0 9.0 V7 0 266.0 37.0 1 43.0 8.0 V15 0 211.0 31.0 1 98.0 14.0 V16 0 295.0 43.0 1 14.0 2.0 V18 0 118.0 22.0 1 191.0 23.0 V20 0 146.0 16.0 1 163.0 29.0 V30 0 4.0 2.0 1 305.0 43.0 V26 0 268.0 39.0 1 41.0 6.0 V27 0 277.0 41.0 1 32.0 4.0 V29 0 167.0 27.0 1 142.0 18.0 Time taken to build model: 0.01 seconds
=== Stratified cross-validation === === Summary ===
Correctly Classified Instances 311 88.8571 % Incorrectly Classified Instances 39 11.1429 % Kappa statistic 0.2918 Mean absolute error 0.1634 Root mean squared error 0.2964 Relative absolute error 75.169 % Root relative squared error 90.2836 % Total Number of Instances 350

Tabla 23. Resultados del proceso de predicción mediante el algoritmo Naive Bayes. (Continuación)

=== Detailed Accuracy By Class ===								
TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,980	0,767	0,901	0,980	0,939	0,335	0,799	0,954	0
0,233	0,020	0,625	0,233	0,339	0,335	0,799	0,434	1
Weigh.0,889	0,676	0,867	0,889	0,865	0,335	0,799	0,890	
=== Confusion Matrix ===								
a b <-- classified as								
301 6 | a = 0								
33 10 | b = 1								

Fuente: Weka

En la tabla 21 empieza el algoritmo Naive Bayes muestra la relación: Prueba 1, el número de instancias en la relación: 350, además se muestra el modelo clasificador. Se visualiza un nodo raíz que pertenece a la variable dependiente denominada Clase y el conjunto de atributos independientes que contiene: Media, Desviación estándar, Suma de valor y precisión. El resultado del algoritmo muestra que se ha clasificado 311 instancias correctas (89%) y 99 instancias incorrectas (11%) de un total de 350, además muestra el coeficiente del estadístico de Kappa con un grado de concordancia de 0,29 (discreto). La clase 0, presenta TP Rate (Tasa de verdaderos positivos) de 0,98 aciertos, una FP Rate (Tasa de falsos positivos) de 0,76, lo que se designa como un nivel de error elevado, una Precisión de 0,90 lo que indica un nivel de instancias correcta, clasificadas ligeramente bajo un Recall (cobertura) de 0,98. La clase 1, presenta TP Rate (Tasa de verdaderos positivos) de 0,23 aciertos, una FP Rate (Tasa de falsos positivos) de 0,02 lo que se designa como un nivel de error bajo, una Precisión de 0,62 lo que indica un nivel de instancias correcta, clasificadas ligeramente bajo un Recall (cobertura) de 0,23.

Según la matriz de confusión el algoritmo consideró lo siguiente: los números de instancias son los valores que encuentran de manera diagonal en la matriz, prediciendo como correctas. De un total de 350 instancias, naive bayes clasificó 311 instancias correctas y 39 incorrectas. En el gráfico 18, margincurve representa el margen de predicción como la diferencia entre la probabilidad pronosticada para la clase real y la probabilidad más alta.

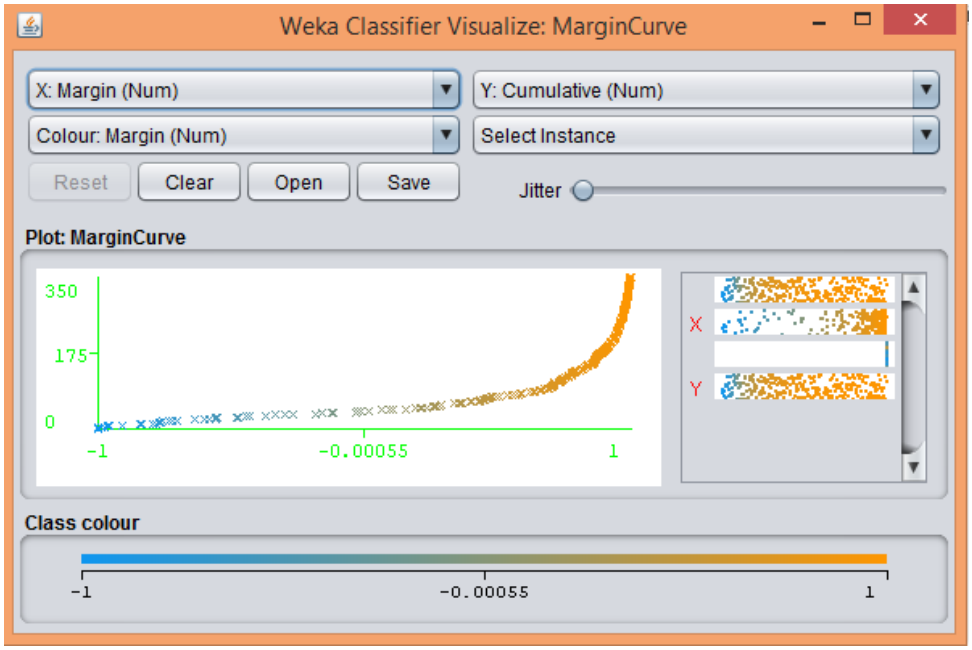


Gráfico 17. MarginCurve Naive Bayes
Fuente: Weka

En el gráfico 19 se observa la compensación costo/beneficio del modelo en el algoritmo naive bayes. Esta metodología determinó al clasificador accurracy con una precisión de predicción del 87,71% lo cual aprueba la aplicación de esta técnica.

.

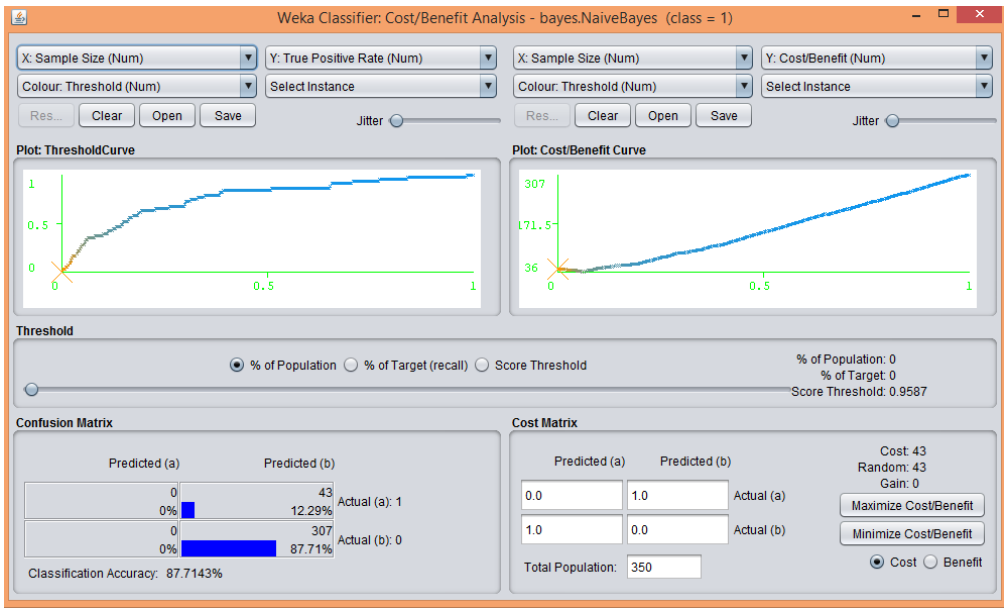


Gráfico 18. Cost/Benefit Naive Bayes.
Fuente: Weka

ALGORITMO ONER

Scalab [79] lo define como un clasificador muy sencillo y rápido. Los resultados que genera este algoritmo son muy buenos a comparación con otros algoritmos. Su característica principal es que selecciona el atributo que explique mejor la clase de salida. Este algoritmo pertenece a una regla mayoritaria sobre un solo atributo. Se consideró a OneR por la sencillez y la rápida respuesta de los resultados, haciéndola fácil la lectura de los resultados. El algoritmo OneR, perteneciente al método de reglas muestra los resultados en la tabla 22.

Tabla 24. Resultados del proceso de predicción mediante el algoritmo OneR

=== Run information ===								
Scheme: weka.classifiers.rules.OneR -B 6								
Relation: Data set Prueba 1 marcelo-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last								
Instances: 350								
Attributes: 31								
Test mode: 10-fold cross-validation								
=== Classifier model (full training set) ===								
V1:								
0 -> 0								
1 -> 0								
(307/350 instances correct)								
Time taken to build model: 0 seconds								
=== Stratified cross-validation ===								
=== Summary ===								
Correctly Classified Instances	307	87.7143 %						
Incorrectly Classified Instances	43	12.2857 %						
Kappa statistic	0							
Mean absolute error	0.1229							
Root mean squared error	0.3505							
Relative absolute error	56.5226 %							
Root relative squared error	106.7512 %							
Total Number of Instances	350							
=== Detailed Accuracy By Class ===								
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1,000	1,000	0,877	1,000	0,935	0,000	0,500	0,877	0
	0,000	0,000	0,000	0,000	0,000	0,000	0,500	0,123	1
Weigh.	0,877	0,877	0,769	0,877	0,820	0,000	0,500	0,784	
=== Confusion Matrix ===									
a b <-- classified as									
307 0 | a = 0									
43 0 | b = 1									

Fuente: Weka

En la tabla 22 el algoritmo OneR muestra la relación: Prueba 1, el número de instancias que es de 350, los atributos que contiene los datos y su identificador. Este algoritmo pose la característica de escoger el atributo que cree conveniente para desplegar resultados sobre la Clase. El resultado del algoritmo muestra que se ha clasificado 311 instancias correctas (89%) y 39 instancias incorrectas (11%) de un total de 350, lo que indica que el modelo del rendimiento académico se encuentra clasificado correctamente, además muestra el coeficiente del estadístico de Kappa con un grado de concordancia de 0,29 (insignificante). La clase 0, presenta TP Rate (Tasa de verdaderos positivos) de 1 acierto, una FP Rate (Tasa de falsos positivos) de 1, lo que se designa como un nivel de error elevado, una precisión de 0,87 lo que indica un nivel de instancias correcta, clasificadas ligeramente bajo un recall (cobertura) de 1. La clase 1, presenta TP Rate (Tasa de verdaderos positivos) de 0,00 aciertos, una FP Rate (Tasa de falsos positivos) de 0,00 lo que se designa como un nivel de error bajo, una precisión de 0,00 lo que indica un nivel de instancias correcta, clasificadas ligeramente bajo un Recall (cobertura) de 0,00.

Según la matriz de confusión el algoritmo consideró que los números de instancias son los valores que encuentran de manera diagonal en la matriz, prediciendo como correctas. De un total de 350 instancias, clasificó 307 instancias correctas y 43 incorrectas. En el gráfico 20 se observa margincurve de OneR.

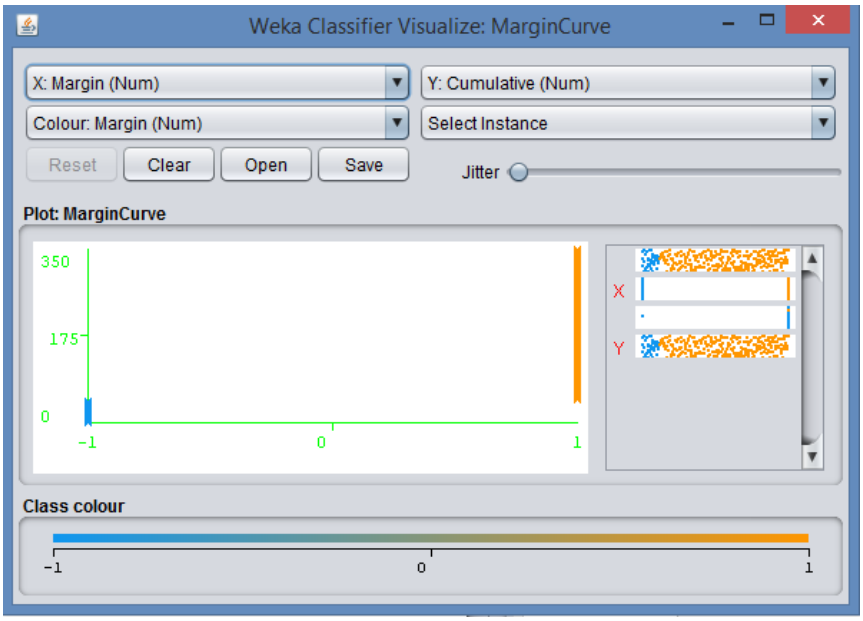


Gráfico 19. MarginCurve OneR.
Fuente: Weka

En el gráfico 21 se observa varias compensaciones de costo/beneficio del modelo. Esta metodología permitió evalúa los costes y beneficios con el objetivo de determinar es viable desde el punto de vista del bienestar social. Se puede observar una curva de elevación con un 87,71 del clasificador accurracy, lo cual aprobó la fiabilidad.

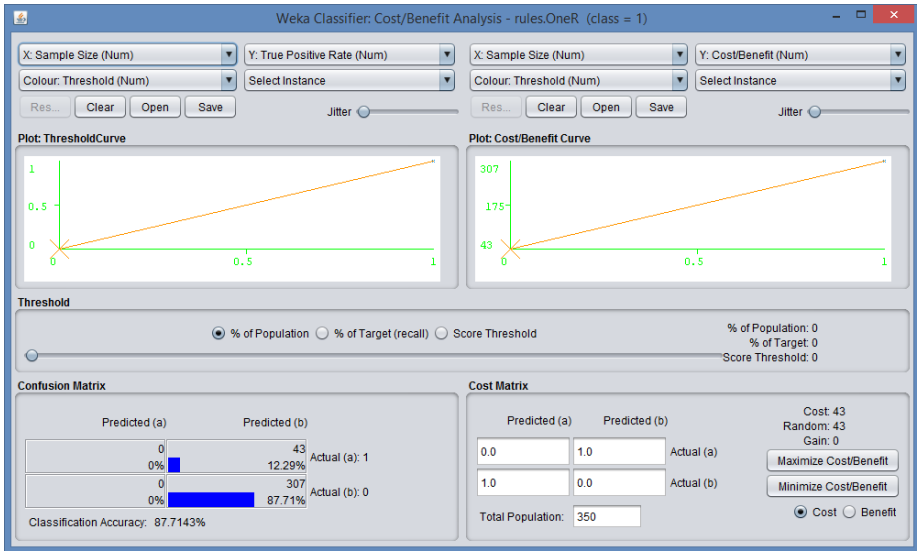


Gráfico 20. Cost/Benefit OneR.
Fuente: Weka

El gráfico 22, margincurve representa el margen de predicción como la diferencia entre la probabilidad pronosticada para la clase real y la probabilidad más alta.

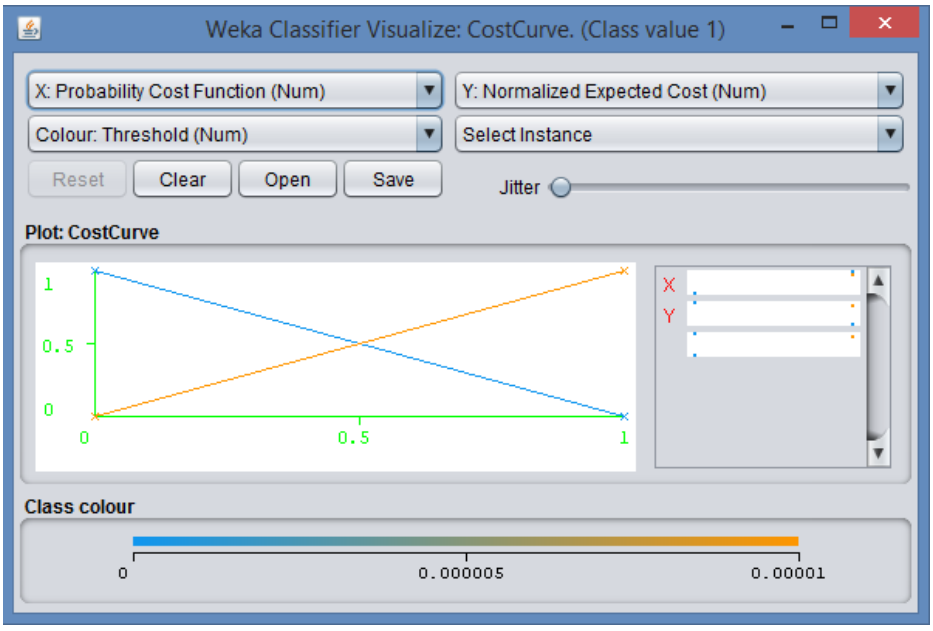


Gráfico 21. CostCurve OneR
Fuente: Weka

Evaluación de los modelos de predicción

En el gráfico 23 se puede visualizar la curva de ROC generada en puntos que ilustran las compensaciones de predicción, al estar en valor umbral al ser mayor a 0,5 quiere decir que la instancia se prediga como positiva.

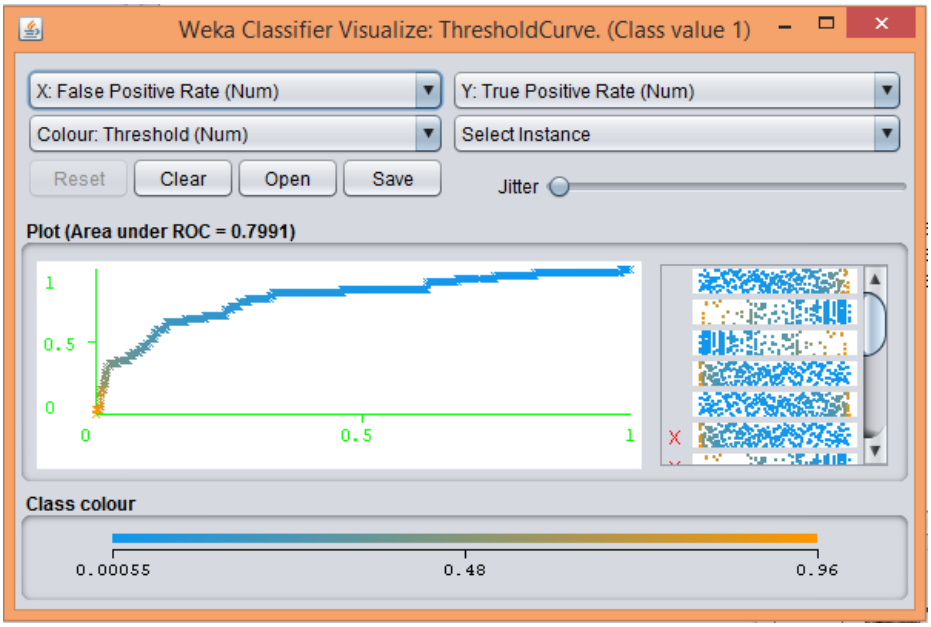


Gráfico 22. ThreshouldCurve Naive Bayes.
Fuente: Weka

Como resultado final una vez aplicadas cada una de las técnicas de minería de datos se emplearon métricas de evaluación para determinar que técnica o algoritmo tiene mejor resultado como se muestra en los cuatro modelos empleados y evaluados a través de las métricas obtenidas a partir de los resultados de la matriz de confusión, estadística de Kappa y el área bajo la curva ROC

Tabla 25. Métricas de evaluación

Métricas de Evaluación	J48	Naive Bayes	Random Forest	OneR
Instancias clasificadas correctas	300	311	303	307
Instancias clasificadas incorrectas	50	39	47	43
Precisión de la predicción (%)	85,72%	88,85%	86,57%	87,71%
Estadística de Kappa	0,17	0,29	0,07	0
Precisión	0,82	0.86	0,81	0,76
Cobertura	0,85	0,88	0,86	0,87
F-Measure	0,83	0,86	0,82	0,82
Área Bajo la curva ROC	0,65	0,79	0,8	0,5

Realizado por: El investigador

De acuerdo con la tabla 23 se puede visualizar que los cuatro algoritmos de clasificación presentan relativamente buenos resultados, los mismos que son similares entre sí. El resultado más alto se obtiene por la clasificación del Naive Bayes el cual presenta un porcentaje de precisión del 88,85%, lo que significa que, de las 350 instancias, 311 fueron clasificadas como correctas lo que indica que se encuentran clasificados correctamente.

El coeficiente kappa obtenido por el naive bayes posee el valor más alto de 0,29, lo que significa que el nivel de concordancia de las variables es discreto, según la escala de valores de kappa. La precisión y la cobertura están relacionadas, pero si la precisión aumenta la cobertura disminuye, y se nota que el algoritmo naive bayes aumenta la cobertura y para el J48 disminuye. El valor del estadístico F-menssure se encuentra cerca de 1, lo que indica que existe una alta confiabilidad del modelod. Finalmente, el resultado del Área ROC en todos los clasificadores poseen niveles de exactitud aceptable, ya que superan el nivel mínimo de 0,50 y tienen valores próximos a 1.

A continuación, se presenta el gráfico 24 con la tasa de precisión de los cuatro algoritmos aplicados al modelo de análisis del rendimiento académico.

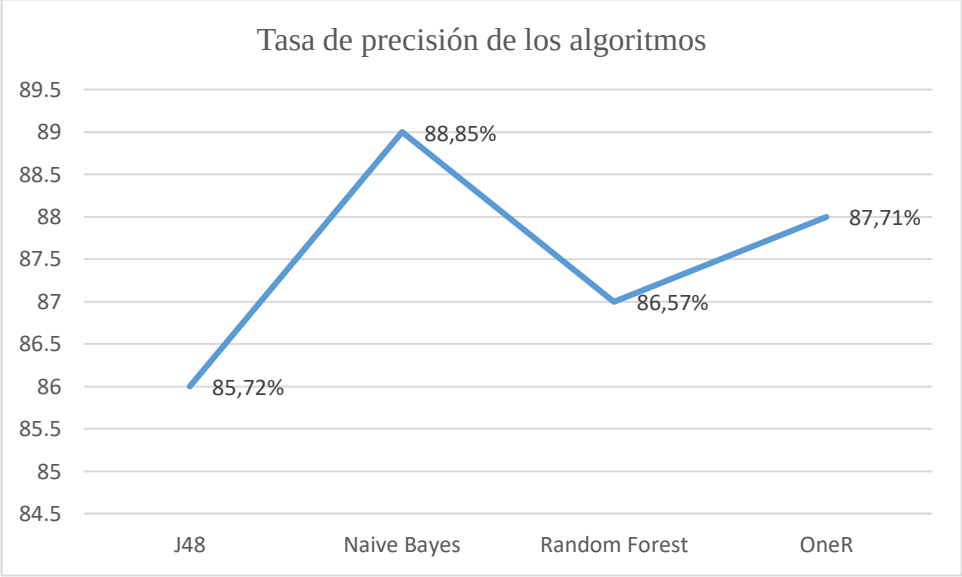


Gráfico 23. Tasa de precisión de los algoritmos.
Realizado por: El investigador

El nivel de precisión que se obtuvo por los algoritmos aplicados al conjunto de datos con un tamaño de 350 registros se muestra en el gráfico 24, el mayor nivel para algoritmo clasificador es el Naive Bayes, el nivel de precisión para los algoritmos se encuentra en intervalos desde 84,5% a 89,5%.

La revisión bibliográfica permitió encontrar estudios previos sobre la temática, los cuales permitieron tener una idea más clara de conceptos y factores que son parte del proceso académico de los estudiantes. Los resultados alcanzados del modelo permitieron afirmar que los factores que influyen en el rendimiento académico de los estudiantes son estadísticamente significativos con un nivel de confianza del 95% y con un p-value del 0,05%, finiquitando que todos los trece factores influyen en el rendimiento académico de los estudiantes de la unidad educativa considerada como caso de estudio. Por lo que se deja abierta una serie de factores que deben ser involucrados para el proceso experimental y contribuir a crear modelos en beneficio de la educación. Por otra parte, se ha conseguido demostrar la importancia de aplicar técnicas de minería de datos para realizar análisis estadístico y de predicción en base a la base de datos históricos, queda por aplicar este estudio en redes neuronales como en técnicas de inteligencia artificial para obtener modelos con un alto valor de precisión.

INDICADORES DEL MODELO DEL RENDIMIENTO ACADÉMICO

V30: Actividad trabajo

Canabal [80] afirma que el trabajar durante la formación académica está relacionado positivamente ya que el estudiante administra mejor su tiempo y quizá tiene mejores calificaciones. Por otra parte Sullana [81] determina que los grupos minoritarios son más propensos a realizar una actividad de trabajo ya que está asociado a un nivel bajo sociocultural y a grupos de desigualdad educativa, que se refleja a posterior en el rendimiento académico.

V4: estado civil

Malstrom [82] el estado civil soltero o casado es un factor predictivo significativo de mucha importancia. López y Tulcán [83] en un estudio de colegio PCEI del Carchi define que el estado civil no es un factor para la deserción y el bajo rendimiento académico, ya que existe un mínimo porcentaje que desertan sean casados o en unión libre. Por otra parte, Ferreyra [84] manifiesta que el estar casado no existe mayor influencia, pero al asociarlo con factores como el residir en la casa de los padres con la influencia que este genera y asistir a la educación secundaria son factores que disminuyen el rendimiento académico.

V5: Etnia

Romero [85] menciona que la población que se reconoce como perteneciente a una etnia se encuentra en desventaja con respecto a la no etnia ya que enfrenta situaciones socioeconómicas adversas, asociadas a bajos niveles de ingresos, educación escasa y pobreza, con una menor probabilidad de estar estudiando. Estas desventajas afectan negativamente al rendimiento académico de los estudiantes étnicos evidenciándose en las calificaciones.

V7: Profesión del padre

Kohl, Lengua y McMahon [86] este factor repercute en el rendimiento académico del estudiante, ya que los padres tienen dificultades de aprendizaje y no pueden ser apoyo a la hora de revisar o explicar obtenido malas experiencias educativas. Campos [87] señala que la profesión y un empleo permite garantizar un nivel de suficiencia económica para satisfacer los requerimientos educativos de sus hijos, sin embargo, al no contar con esa condición, deben priorizar sus esfuerzos en necesidades inmediatas, generando frustración ante las necesidades primarias, personales y familiares, antes que las necesidades más elevadas como la educación de los hijos.

V11: Ausencia del padre

Núñez [88] considera que los padres no solo influyen con sus palabras, sino también con sus actitudes ya que el estudiante va asumiendo un estilo de vida. Por otra parte

Rodríguez [89] señala que, en las familias con desajustes emocionales, desintegradas, o con ausencia del padre que tiene el rol de protector, visto como modelo a imitar y al estar ausente el estudiante pierde seguridad y presencia de un modelo paternal, siendo probable pobre rendimiento académico.

V:15 Estructura familiar

Valle, González y Frías [90] consideran que la estructura familiar pese a haber sido analizado durante varios años sigue incidiendo indirectamente en la formación académica de los estudiantes. Ruiz [91]manifiesta que el nivel cultural-educativo familiar es limitado haciendo que el interés de los padres por la educación sea mínimo. Dulanto [92] afirma que ningún sistema educacional y por consiguiente ningún estudiante puede alcanzar un buen nivel de rendimiento si no existe una estructura familiar sólida. López [93] la estructura de la familia tiene la funcionalidad de ayudar en la educación de los hijos transmitiendo valores éticos y morales y enseña a vivir y relacionarse con el resto de la sociedad.

V:16 Calificaciones

Page [94] indica que las calificaciones finales por ser una medida exacta y accesible posee un valor relativo como medida del rendimiento. Escudero [95] manifiesta que las calificaciones son una medida de los resultados de la enseñanza, pero no exactamente de calidad, sino por criterios, la valoración y calificación del aprendizaje del rendimiento académico por parte del docente.

V18: Lugar de residencia

Camacho y Ramos [96] señalan que el lugar de residencia con las características del barrio, el estrato socioeconómico, cultura y político y los posibles factores de violencia y la delincuencia, impide que los estudiantes ingresen a los sistemas educativos o tengan un adecuado rendimiento académico, ya que pueden ser acosados por grupos que promueven la deserción motivando a la creación de pandillas. Erazo [22] manifiesta que el estudiante debe lidiar con los problemas de barrio que son causas también de un rendimiento académico bajo.

V20: Vivienda

Los estudiantes con rendimiento académico alto tienen un mayor nivel socioeconómico contando con recursos y bienes. Gil [97] explica que si la vivienda está construida con materiales de buena calidad y dispone una habitación para el solo y sus padres trabajan, son estudiantes que no tendrán necesidad de trabajar, por lo contrario, si es bajo recursos y su vivienda no cuentan con todos los servicios viven hacinados es considerable que posiblemente los estudiantes trabajen y baje su rendimiento académico.

V22: Formación del representante legal

Jara at all [98] la formación académica del representante legal y/o padre de familia influye en el rendimiento académico, ya que un estudiante hijo de un obrero tiene el 4.8 de probabilidad de perder por lo menos en una asignatura que un estudiante hijo o representado de un juez o con un nivel académico alto.

V26: Internet

Campos [87] afirma que los estudiantes con estrato de mayores ingresos consideran a la televisión cable y el internet en fuente de información sumamente importantes para desarrollar sus actividades escolares, lo contrario de los estudiantes de estrato de menores ingresos considerando a la televisión abierta como único medio de información para su preparación académica.

V27: Uso de Computador

Jara at al [98] dice que la tenencia de recursos como el computador integran la complejidad del rendimiento y desafortunadamente el bajo rendimiento se asocia con estudiantes que tienen padres con bajas capacidades económicas para adquirirlas.

V29: Repetición académica

La repetición académica es uno de los factores con mayor incidencia de los estudiantes en los cursos inferiores. Según la OCDE [5] en Ecuador el 18,5% de los estudiantes señalan haber repetido al menos una vez, y a veces sin llegar a repetir

un curso formalmente por problemáticas como enfermedad, trabajo o cuidado de un familiar, por lo que existe mayor probabilidad de seguir desertando o teniendo un bajo rendimiento académico.

Conclusiones del Capítulo III

- Se aplicó las técnicas de minería de clasificación usando los algoritmos como: árbol de decisiones j48, random forest, naive bayes y oneR ejecutando la data. Los resultados demuestran que el Naive Bayes es el modelo que mayor precisión da en la prueba de entrenamiento aprendizaje aplicada a la data.

- El coeficiente kappa obtenido por el naive bayes fue el valor más alto de 0,29, lo que significa que el nivel de concordancia de las variables fue discreto, según la escala de valores de kappa.

- La f-menssure fue confiable, ya que se encuentra cerca del uno, y mientras el valor se acerque a 1, mayor es la confiabilidad.

- El resultado de Área ROC los clasificadores poseen nivel de exactitud aceptable, ya que superan el nivel mínimo de 0,50 y tienen valores próximos a 1.

- El nivel de precisión que se obtuvo por los algoritmos aplicados al conjunto de datos con un tamaño de 350 registros, es el naive bayes, en un nivel de precisión del 88,85%. Siendo Naive Bayes el que tiene mejor nivel de precisión.

CONCLUSIONES GENERALES.

Al finalizar el trabajo de investigación se presenta las principales conclusiones.

- La revisión de la literatura permitió la sustentación bibliográfica con un total de 140 documentos revisados entre libros, revistas indexadas, artículos científicos, acuerdos ministeriales, instructivos y tesis de grado, con los cuales se obtuvo el marco teórico. Por otra parte, se eligió la metodología y las herramientas informáticas más adecuadas y utilizadas en la minería de datos, las cuales posibilitó el análisis los datos y se estableció los principales factores del rendimiento académico.

- Se construyó una dataset a partir de la base de datos del sistema informático construido para la unidad educativa, con una población de 1050 estudiantes de todos los niveles de educación vigentes, los cuales permitieron obtener 13 variables convertidas en hipótesis probadas a través de procesos experimentales para el diseño del modelo conceptual, con los cuales se pudo evidenciar la influencia de los factores más influyentes en el rendimiento académico.

- Para la comprobación de la tasa de precisión del modelo propuesto se establece un proceso experimental con cuatro algoritmos de clasificación a través de técnicas de machine learning como j48, random forest, naive bayes y oneR. La implementación de estas técnicas permitió determinar la tasa de precisión de la predicción del modelo propuesto. A través de estas técnicas se puede determinar que existe un 89% de tasa de precisión a través de la técnica naive bayes, lo que indica que el modelo que se presenta es adecuado en términos de confiabilidad, los niveles de capa obtenidos a través del proceso experimental con un resultado del 0,86 indican que estos modelos son adecuados para este tipo de procedimientos.

RECOMENDACIONES

- A partir de la investigación propuesta se recomienda el análisis de nuevos factores del rendimiento académico, como el emocional ya que este aspecto no fue consideraron en esta investigación ya que se partió de la data histórica. Ya que se pudo evidenciar en la literatura que la afectividad es un factor muy influyente en el proceso de aprendizaje y por ende este factor recae en el rendimiento académico.
- Implementar al modelo de rendimiento académico nuevas técnicas de predicción que posiblemente podrían incrementar la tasa de precisión de estos modelos.
- La aplicación de algoritmo hibrido que pueda combinar estas técnicas y mejorar los modelos de precisión.
- Con la valoración del modelo se podría determinar estrategias para mitigar e incrementar la tasa de retención en la institución, evitando que las unidades educativas de modalidad semipresenciales desaparezcan como una opción de formación académica para las personas con escolaridad inconclusa.

.

REFERENCIAS BIBLIOGRÁFICAS.

[1] A. Luis, Cibersociedad: los retos sociales ante un nuevo mundo digital, Salamanca, 1997.

[2] J. Riquelme, «Minería de Datos: Conceptos y Tendencias,» *Revista Iberoamericana de Inteligencia Artificial,* vol. 10, nº 29, pp. 11-18, 2006.

[3] D. S.-G. a. R. G. S. Alejandro Ballesteros Román, «Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo,» *Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada, Unidad Legaria del Instituto Politécnico Nacional. Calzada Legaria,,* pp. 1-7, 2013.

[4] L. A. Alvares Aldaco, «"Comportamiento de la Deserción y Reprobación en el Colegio de Bachilleres del Estado de Baja California: Caso Plantel Ensenada",» de *X Congreso Nacional de Investigación Educativa. México*, México, 2009.

[5] OCDE, «El Programa PISA de la OCDE,» OCDE, Paris.

[6] T. F. J. Murillo, Investigación iberoamericana sobre eficacia escolar., Quito: Convenio Andrés Bello, 2007.

[7] J. A. Gross, «Educación en Ecuador. Resultados de PISA para el Desarrollo,» Ineval, Quito, 2018.

[8] M. d. Educación, Marco Legal Educativo, Quito: Editogran S.A., 2012.

[9] I. N. d. Educativa, «La educación en Ecuador: logros alcanzados y nuevos desafíos. Resultados educativos 2017-2018,» © Instituto Nacional de Evaluación Educativa,, Quito, 2018.

[10] H. Á. G. Álvaro Jiménez Galindo, «Minería de Datos en la Educación,» *Universidad Carlos III,* pp. 1-8, 2010.

[11] S. N. Marín, Hacia una teoría sobre el rendimiento académico en enseñanza primaria a partir de la investigación empírica: datos preliminares, España: BIBLID, 2008.

[12] R. H. C. C. F. &. L. P. B. Sampieri, Metodología de la nvestigación, vol. 6, México: Mcgraw-hill., 2014.

[13] M. C. Ocampo, «Métodos de Investigación académica. (versión 1.1),» Sede de Occidente, UCR, Costa Rica, 2017.

[14] J. L. Abreu, «Hipótesis, Método & Diseño de Investigación,» *Daena: International Journal of Good Conscience.,* vol. 7, nº 2, pp. 187-197, 2012.

[15] P. P. Zulay, «Los diseños de método mixto en la investigación en educación: Una experiencia concreta,» *Revista Electrónica Educare,* vol. XV, nº 1, pp. 42-58, 2011.

[16] F. Herrera, « ¿ Cómo interactúan el autoconcepto y el rendimiento académico, en un contexto educativo pluricultural?,» *Revista Iberoamericana de Educación,* vol. 34, nº 2, pp. 1-9, 2004.

[17] N. M. Santiago, HACIA UNA TEORÍA SOBRE EL RENDIMIENTO ACADÉMICO EN ENSEÑANZA PRIMARIA A PARTIR DE LA INVESTIGACIÓN EMPÍRICA: DATOS PRELIMINARES, Salamanca: Ediciones Universidad de Salamanca, 2008.

[18] S. Marcos, « Factores de personalidad y rendimiento escolar,» *Revista Española de pedagogía,* vol. X, nº 37, pp. 77-78, 1952.

[19] G. Y. J, «El pronóstico para los estudios del bachillerato elemental a nivel de ingreso,» *Psicología General y Aplicada,* vol. XIX, nº 73, pp. 523-526, 1964.

[20] J. M., «Competencia social: intervención preventiva en la escuela.Infancia y Sociedad. 24, pp. 21-,» nº 24, pp. 21-48, 2000.

[21] C. L. E. Guerrero, Á. M. S. Cardona y J. R. T. Cuevas, «Factores de riesgo asociados a bajo rendimiento académico en escolares de Bogotá,» *Investigaciones Andina,* vol. 15, nº 26, p. 108, 2013.

[22] O. A. Erazo, «EL RENDIMIENTO ACADÉMICO, UN FENÓMENO DE MÚLTIPLES RELACIONES Y COMPLEJIDADES,» *Revista Vanguardia Psicológica Clínica Teórica y Práctica,* vol. 2, nº 2, pp. 144-173, 2011.

[23] C. Joan, Diccionario crítico etimológico castellano e hispánico, Madrid: Gredos, 2000.

[24] Tawab, Enciclopedia de pedagogía/psicología., Barcelona: Ediciones Trébol, 1997.

[25] G. Andrade, «Predicción del rendimiento académico lingüístico y lógico-matemático por medio de las variables modificables de las inteligencias múltiples y del hogar,» *Revista Digital de Investigación y Nuevas Tecnologías - Contexto Educativo,* vol. 17, pp. 1-12, 2001.

[26] M. d. C. C. Ruíz, «Fracaso Escolar, Una realidad en nuestras aulas,» *Educación en Extremadura,* p. 154, 2010.

[27] G. W, «EL RENDIMIENTO ACADÉMICO:,» *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación ,* vol. 1, nº 2, pp. 1-15, 2003.

[28] B. B. y. B. M., Causas psicológicas del bajo rendimiento escolar., México: Librería Carlos Cesarman, S.A., 1988.

[29] F. U. M., «Data Mining and Knowledge Discovery: Making Sense out of Data.,» *IEEE Expert, Intelligent Systems & Their Applications,* pp. 20-25, 1996.

[30] J. P. y. K. M. Han Jiawei, Minería de datos: conceptos y técnicas ., Illinois: Elsevier, 2011.

[31] L. C. &. P. Á. A. J. Liñán, «Educational Data Mining and Learning Analytics: differences, similarities, and time evolution,» *International Journal of Educational Technology in Higher Education,* vol. 12, nº 3, pp. 98-112, 2015.

[32] B. R. A., «Minería de Datos Educativa Aplicada a la Investigación de Patrones de Aprendizaje en Estudiante en Ciencias,» Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada. Instituto Politécnico Nacional, México, (2012)..

[33] C. &. V. S. Romero, «Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,» vol. 3, nº 1, pp. 12-27, 2013.

[34] Z. K. &. E. A. A. Papamitsiou, «Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence,» *Educational Technology & Society,* vol. 17, nº 4, pp. 49-64, 2014.

[35] A. M. &. H. W. Shahiri, «A review on predicting student's performance using data mining techniques,» *Procedia Computer Science,* nº 72, pp. 412-422, 2015.

[36] P. K. &. X. M. Kotsiantis S., «{Un conjunto combinatorio incremental de clasificadores como técnica para predecir el rendimiento de los estudiantes en educación a distancia,» *Digital Library,* vol. 23, nº 6, pp. 529-535, 2010.

[37] P. A. &. C. E. L. Aloise-Young, « Not all school dropouts are the same: Ethnic differences in the relation between reason for leaving school and adolescent substance use,» *Psychol Sch,* vol. 39, nº 5, pp. 539-547, 2002.

[38] D. Martínez, «Predicting student outcomes using discriminant functions analysis.,» Lake Arrowhead, California, 2001.

[39] F. R. C. &. S. A. Araque, «Factors influencing University Drop Out Rates,» *Computers & Education,* vol. 53, pp. 563-574, 2009.

[40] V. William, «Identifying characteristics of high school dropouts: data mining with a decisión tree model,» *Annual meeting of the American educational Research Association,* vol. 1, pp. 1-11, 2004.

[41] J. F. V. J. P. &. M. N. Superby, «Determination of factors influencing the achivement of first-year university students using data mining methods,» *Educational data mining workshop,* pp. 1-8, 2006.

[42] M. d. Datos, César Pérez López, Daniel Santín González, Madrid: Thomson Ediciones Paraninfo, S.A., 2008.

[43] B. I. C. J. y. M. N. Berzal F., «Marcos de minería de datos basados en componentes.,» *Comunicaciones de la ACM,* vol. 45 , nº 12, pp. 97-100., 2002.

[44] Maco, «5 de los mejores software de minería de datos de Código Libre y Abierto,» 2016. [En línea]. Available: http;//blog.jmacoe.com/gestion_ti/base_de_datos/5-mejores-software-mineria-datos-codigo-libre-abierto. [Último acceso: 23 Enero 2020].

[45] A. J. S. Castillo, Métodos Estadísticos con R y R Commander, España: Creative Commons, 2010.

[46] A. S. H. y. S. G. Juan Miguel Moine, «Estudio comparativo de metodologías para minería de datos,» de *XII Workshop de Investigadores en Ciencias de la Computación*, Argentina, 2011.

[47] M. &. A. C. Santos, Data Mining – Descoberta de Conhecimento em Bases de Dados., FCA Publisher, 2005.

[48] M. Jesús, Conceptos y teorías en la ciencia, Madrid: Alianza Editorial, 1984.

[49] R. B. J. Campoverde, «Análisis comparativo sobre la afectividad como motivadora del proceso enseñanza-aprendizaje. Casos: Argentina, Colombia y Ecuador,» *Sophia,* vol. 12, nº 2, pp. 217-231, 2016.

[50] G. B. M. Acero, *agrupa en dos marcos interpretativos con variables de índole extraescolar, refiriéndose a la situación socioeconómica y al contexto familiar y la intraescolar que se refiere a la baja motivación para el estudio por la cosmovisión y las expectativas no sat,* Cuenca, 2016.

[51] Á. Aldaco, «Comportamiento de la Deserción y Reprobación en el Colegio de Bachillerato de Estados de Baja California,» de *Caso Plantel Ensenada. X Congreso Nacional de Investigación Educativa.* , Veracruz, 2009.

[52] C. I, «Análisis de las calificaciones escolares como criterio de rendimiento académico.,» 2000.

[53] D. A. M. Ledesma, «Modelos para la Mejora del Rendimiento Académico de Alumnos de la E.S.O. mediante Técnicas de Minería de Datos,» Murcia, 2015.

[54] R. J. C. N. M. L. B. F. a. R. V. Huerta Luis, «Minería de datos: Impacto de Actividades Cotidianas en el Rendimiento Estudiantil,» *International Journal of Innovation and Applied Studies ,* vol. 14, nº 4, pp. 927-935, 2016.

[55] K. Jared, «De agujas y pajares: construir un sistema de alerta temprana de abandomo escolar en todo el estado de Wisconsin,» *Journal of Educational Data Mining,* vol. 7, nº 3, pp. 18-67, 2015.

[56] P. O. L. Emir, «APLICACIÓN DE TÉCNICAS DE MINERIA DE DATOS PARA PREDECIR LA DESERCIÓN ESTUDIANTIL EN LA EDUCACIÓN BÁSICA EN LA REGIÓN DE LAMBAYEQUE,» Pimentel, 2016.

[57] H. I. C. S. H. A. y. A. J. Timarán S, «El proceso de descubrimiento de conocimiento en bases de datos,» *Universidad Cooperativa de Colombia,* pp. 63-86, 2016.

[58] R. C.-Z. J. &. H.-T. A. Timarán-Pereira, «Árboles de decisiones para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas saber 11°. Rev.investig.desarro.innov., 9 (2), xxx,» *ev.investig.desarro.innov,* vol. 9, nº 2, 2019.

[59] A. M. Ledesma, «Modelos para la Mejora del Rendimiento Académico de Alumnos de la E.S.O. mediante Técnicas de Minería de Datos,» Murcia, 2015.

[60] F. Pita, «Determinación del tamaño muestral,» *Cad Aten Primaria,* vol. 3, nº 138, pp. 1-6, 1996.

[61] O. H. y. C. A, «Aproximación al uso de Coeficiente Alfa de Cronbach,» *Revista Colombiana de Psiquiatría,* vol. XXXIV, nº 4, pp. 572-580, 2005.

[62] M. G. Roberto, «Modelos de regresión lineal múltiple,» Granada España, 2016.

[63] L. A. A. M. Carlos Camacho, «Regresión linela simple,» de *Apuntes no publicados de la asignatura Análisis de datos II de la licenciatura de Psicología*, Universidad de Sevilla, 2006.

[64] L. P. César, Minería de datos: técnicas y herramientas, Paraninfo, 2007.

[65] C. G. V., «Hipótesis en el modelo de regresión lineal por Nínimos Cuadrados Ordinarios,» 2015.

[66] H. C. E, «El método de mínimos cuadrados,» México, Facultad de Ciencias UNAM, 2016, pp. 1,2-5.

[67] B. T, «¿Qué son los datasets y los dataframes en el Big Data?,» 2018.

[68] G. Morate, Manual de Weka, 2008.

[69] C.-H. C. a. Y.-S. Chen., «Extracting rules of initial returns using attribute selection and entropy-based rough sets in electronic firm,» *FSKD,* pp. 146-150, 2007.

[70] B. A. M. A. y. M. A. Pascual M, *Un nuevo clasificador de préstamos bancarios a través de minería de datos.*

[71] Menacho, «Predicción del rendimiento académico aplicando técnicas de minería de datos,» *Revista de Facultad de Economía y Planificación,* vol. 78, pp. 26-33, 2017.

[72] S. A., «A review on predicting student´s performance using mining techniques,» *Procedia Computer Science,* pp. 412-422, 2012.

[73] H. y. Ferri, Introducción a la Minería de Datos, 2004.

[74] B. L, «Machine Learning,» *Bagging Predictors,* vol. 24, nº 2, pp. 123-140, 1996.

[75] B. W. y. P. Escobar H, «Análisis inteligente de datos aplicado al proceso de nivelación en la Universidad Estatal de Quevedo,» *Publicando,* vol. 3, nº 7, pp. 33-44, 2016.

[76] L. V, «A survey on Supervised,» *Classification on Data Streams,* nº 205, pp. 88-125, 2015.

[77] C. R. M. A. y. C. T. Gironés Roid, Minería de datos, Modelos y Algoritmos, Barcelona: 1 edición, 2017.

[78] B. M. Beatriz, «Introducción al proceso de descubrimiento de conocimiento en bases de datos (KDD),» Puebla, 2015.

[79] Scalab, «Grupo de investigación,» *Software de aprendizaje automático.*

[80] M. Canabal, « College student degree of participation in the labour force: determinants and relationship to school performance.,» *College Student Journal,* vol. 32, pp. 597-605, 1998.

[81] S. V., «Correspondencia entre estrategias de aprendizaje y rendimiento académico,» *Boletín de investigación Educacional,* vol. 15, pp. 70-88, 2000.

[82] M. E., « Predicting academic success in engineering graduate,» nº 47, 1994.

[83] A. N. &. T. C. J. V. López Cerón, «López Cerón, A. N., & Tulcán Cuasapud, J. V. (2018). Factores que inciden en la tasa de deserción y repitencia de la carrera de nutrición y salud comunitaria de la Universidad Técnica del Norte en el periodo 2009-2017,» Tulcán, 2018.

[84] F. M, «Determinantes del desempeño universitario: efectos heterogéneos en un modelo censurado,» (Doctoral dissertation, Universidad Nacional de La Plata)., Argentina, 2007.

[85] R. J, «Educacion, calidad de vida y otras desventajas económicas de los indígenas de Colombia,» Banco dela república, 2010.

[86] L. L. y. M. R. Kohl G., «Parent involvement in school conceptualizing multiple dimensions and their relations with family and demographic risk factors.,» *Journal of school psychology,* vol. 6, nº 38, pp. 501-523, 2000.

[87] W. Campos, «Factores socioeconómicos y rendimiento académico en estudiantes universitarios: una aproximación teórica. Moquegua : Magister SAC.,» Moquegua, 2006.

[88] N. J., *Motivación y aprendizaje Escolar. .,* Mexico, 1996.

[89] W. Rodríguez, *Teoría de la educación.,* Lima: Escuela Nueva., 1997.

[90] G. y. F. Valle C, «Estructura familiar y rendimiento escolar en niños de educación primaria de nivel socieconómico bajo,» *Anuario de investigación,* vol. 7, pp. 237-250, 2006.

[91] R. d. Miguel, «Factores familiares vinculados al bajo rendimiento,» *Revista complutense de educación,* vol. 12, nº 1, pp. 81-113, 2001.

[92] D. J., «Sugerencias para lograr una buena formación en valores familiares desde la adolecencia.,» Mexico, 2000.

[93] L. L., «La protección constitucional de la familia.,» *Revista de derechos sociales,* pp. 124-131, 2012.

[94] P. M., «Hacia un modelo causal del rendimiento académico,» *Madrid: Centro de publicaciones - Secretaria General Técnica, Ministerio de Educación y Ciencia,* 1990.

[95] E. T., Indicadores del rendimiento académico: una experiencia en Zaragoza., Zaragoza: Ministerio de Educación y Cultura. Centro de publicaciones, 1999, pp. 251-262.

[96] C. y. Ramos, «Componentes de fortaleciemiento ciudadano, jóvenes en riesgo y resocialización,» 2003.

[97] G. J, «Medición del nivel socioeconómico familiar en el alumnado de Educación Primaria.,» *Revista de Educación,* nº 362, pp. 1-17, 2013.

[98] D. V. H. G. G. G. G. L. I. A. C. &. F. M. Jara, «Factores influyentes en el rendimiento académico de estudiantes del primer año de medicina. In Anales de la Facultad de Medicina (Vol. 69, No. 3, pp.,» *In Anales de la Facultad de Medicina ,* vol. 69, nº 3, pp. 193-197, 2008.

[99] E. C. N. V. B. J. A. G. A. &. G. V. Aguilar, «Engagement vs performance: using electronic portfolios to predict first,» de *En LAK '14 Proceedings of the*, New York, USA, 2014.

[100] S. Buckingham, «Learning analytics.,» *Policy Brief,* pp. 2,3, 2012.

[101] R. R. K. G. José C. Riquelme, «Minería de Datos: Conceptos y Tendencias,» *Iteligencia Artificial,* vol. 10, nº 29, p. 13, 2006.

[102] M. J. R. Q. J. H. Orallo, Extracción Automática de Conocimiento en Bases de Datos e Ingeniería del Software, España, 2003.

[103] N. L. Q. Gil y C. A. Valencia, «Aplicación del proceso de KDD en el contexto de bibliomining: El caso Elogim,» *Interamericana de Bibliotecología,* vol. 35, nº 1, 2012.

[104] D. X. G. C. D. T. Héctor Oscar Nigro, «KDD (Knowledge Discovery in Databases): Un proceso centrado en el usuario,» de *VI Workshop de Investigadores en Ciencias de la Computación*, Tandil - Argentina, 2004.

[105] J. BIGUS, Data Mining With Neural Networks., Houston (TX, USA: McGraw-Hill., 1996.

[106] M. Q. M. G. F. &. P. M. Moreno, «Aplicación de técnicas de minería de datos en la construcción y validación de modelos predictivos y asociativos a partir de especificaciones de requisitos de software,» de *DIS 2001, Apoyo a la Decisión en Ingeniería del Software - Decision Support in Software Engineering, Proceedings of the II ADIS 2001 Workshop on Decision Support in Software Engineering,*, Salamanca, 2010.

[107] Y. R. S. y. A. D. Amador, «Herramientas de Minería de Datos,» *RCCI,* vol. 3, nº 3-4, pp. 73-80, 2009.

[108] M. R. A. &. T. A. Mohri, «Foundations of machine,» 2012.

[109] R. S. Pressman, Ingeniería del software, un enfoque práctico, México D.F.: McGrawHill, 2010.

[110] Y. Sarduy Domínguez, «El análisis de información y las investigaciones cuantitativa y cualitativa,» *Revista cubana de salud pública,* vol. 33, 2007.

[111] M. Miguel, «La investigación cualitativa (síntesis conceptual),» *Revista de investigación en psicología,* vol. 9, nº 1, pp. 123-146, 2006.

[112] C. E. Ramos, «Métodos y técnicas de investigación,» 2008. [En línea]. Available: http://www.gestiopolis.com/metodos-y-tecnicas-de-investigacion. [Último acceso: 23 05 2019].

[113] M. N. Q. a. N. V. Kalyankar, «“Drop Out Feature of Student Data for Academic Performance Using Decision Tree Techniques”,» *Global Journal of Computer Science and Technology,* vol. vol. 10, pp. pp. 2-5, 2010..

[114] D. S.-G. a. R. G. Alejandro Ballesteros Román, «Minería de datos educativa: Una herramienta para,» *Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada, Unidad Legaria,* vol. 7, nº 4, 2014.

[115] C. M. Vera, *PREDICCIÓN DEL FRACASO Y EL ABANDONO,* Cordoba, 2015.

[116] P. D. A. Sánchez, «La tendencia del abandono escolar en Ecuador: período 1994-2014,» *Valor Agregado,* vol. 3, nº 2, p. 34, 2015.

[117] P. S., «Modelling engineering student academic performance using academic,» *International journal of engineering education,* vol. 29, nº 1, pp. 132-138, 2013.

[118] S. M, «Academic analytics landscape at the University of Phoenix,» *In Proceedings of the 1st International Conference on Learning and Knowledge,* nº 122-126, 2011.

[119] H. A. G. Alvaro Jiménez Galindo, «Minería de datos en la Ecucación».

[120] R. C. &. V. Sebastián, «Data mining in education.,» *Diario Revisiones interdisciplinarias de Wiley,* vol. 3, nº 1, pp. 12-27, (2013).

[121] E. Tawab, Enciclopedia de Pedagogía/Psicología, Barcelona: Ediciones Trébol, 1997.

[122] R. 2. Camana, « Una Experiencia Personal: Pico y Pala en la Exploración y Visualización de Datos Electorales.,» *Tecnológica ESPOL –RTE,,* vol. 27, pp. 1-13.

[123] R. R. K. G. José C. Riquelme, «Inteligencia Artificial,» *Revista Iberoamericana de Inteligencia Artificial.,* nº No.29 , pp. 11-18., 2006.

[124] G. A. J. Ana Isabel Oviedo Carrascal, «Estudio sobre estilos de aprendizaje mediante minería de datos c o apoyo a la gestión académica en istituciones educativas,» *RISTI Revista Ibérica de Sistemas y Tecnologías de Información,* nº 29, pp. 1-13, 2018.

[125] P. Cazau, Introducción a la Investigación en Ciencias Sociales, Buenos Aires: Red de Psicología online, 2006.

[126] J. K. M. Han, Data Mining: Concepts and techniques, USA: Morgan Kaufmann Publisher, 2001.

[127] D. S.-G. a. R. G. S. Alejandro Ballesteros Román, «Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo,» Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada, Unidad Legaria del Instituto Politécnico Nacional., 2013, 2013.

[128] R. &. H. J. Wirth, «Towards a standard process model for data mining. In Proceedings,» de *4th international conference on the practical applications of knowledge discovery and data mining*, 2000.

[129] C. M. Y. Jiménez., Bases de datos relacionales y modelado de datos., IC Editorial, 2015., 2015.

[130] m. Y. J. Capel, Base de datos relacionales y modelado de datos, Málaga: IC Editorial, 2014.

[131] A. Parker, «Un estudio de variables que predicen el abandono de la educación a distancia,» *Revista Internacional de Tecnología,* vol. 1, nº 2, pp. 1-10, 1999.

[132] A. Y. &. N. M. H. Wang, «Predictors of web-based performance: the role of self-efficacy and reasons for taking an on-line class.,» *Computers in Human Behavior.,* vol. 18, pp. 151-163, 2002.

[133] M. L. Antonio, «Modelos para la mejora del rendimiento académico de alumnos de la ESO mediante técnicas de minería de datos. Proyecto de investigación:.,» Murcia, 2016.

[134] M. Daniel, «Predicting student outcomes using discriminant functions analysis.,» Lake Arrowhead, California, 2001.

[135] D. S. G. César Pérez López, Minería de Datos, Madrid: Thomson Ediciones Paraninfo, S.A., 2008.

[136] B. J. C. &. Z. M. E. G. Rojas, «Análisis comparativo sobre la afectividad como motivadora del proceso enseñanza-aprendizaje "Casos: Argentina, Colombia y Ecuador,» . *Sophia, ,* vol. 12, nº 2, pp. 217-231, 2016.

[137] B. S. y. J. M. Ruiz Omar, «Aplicación de minería de datos para detección de patrones en investigaciones biotecnológicas,» 11 03 2009. [En línea]. Available: http://www.dspace.espol.edu.ec/xmlui/handle/123456789/4719.

[138] A. Y. C. L. M. A. C. M. y. C. G. Diaz H., «Algoritmos de aprendizaje automático para clasificación de Splice Sites en secuencias genómicas,» *revista Cubana de Ciencias Informáticas,* vol. 9, pp. 155-170, 2015.

[139] L. M. M. C. Bedoya O., «Minería de datos en egresados de la Universidad de Caldas,» *Revista virtual. Universidad Católica del Norte,* vol. 6, pp. 110-124, 2016.

[140] S. F. C. G. García F., «Modificaciones del algoritmo Random Forest para su empleo en clasificación de imágenes de teledetección,» *Revista Tecnológica de la información geográfica,* vol. 1, pp. 359-368, 2016.

ANEXOS

Anexo 1: Currículo Integrado.

HORAS PEDAGÓGICAS POR CURSO											
		GRADO	OCTAVO			NOVENO			DECIMO		
		MODALIDAD	SEMIPRESENCIAL		PRESENCIAL O A DISTANCIA	SEMIPRESENCIAL		PRESENCIAL O A DISTANCIA	SEMIPRESENCIAL		PRESENCIAL O A DISTANCIA
	Áreas	ASIGNATURA	PRESENCIAL	AUTÓNOMO		PRESENCIAL	AUTÓNOMA		PRESENCIAL	AUTÓNOMA	
BÁSICA SUPERIOR	Lengua y Literatura	Lengua y Literatura	120	80	200	120	80	200	120	80	200
	Matemática	Matemática	120	80	200	120	80	200	120	80	200
	Ciencias Sociales	Estudios Sociales	80	80	160	80	80	160	80	80	160
	Ciencias Naturales	Ciencias Naturales	80	80	160	80	80	160	80	80	160
	Educación Cultural y Artística	Educación Cultural y Artística	40	40	80	40	40	80	40	40	80
	Educación Física	Educación Física	40	-	40	40	-	40	40	-	40
	Lengua Extranjera	Inglés	80	40	120	80	40	120	80	40	120
	Horas pedagógicas		560	400	960	560	400	960	560	400	960
	Horas adicionales a discreción para la flexibilización curricular (tutorías)		240	-	240	240	-	240	240	-	240
	Horas pedagógicas totales		800	400	1200	800	400	1200	800	400	1200

	ÁREAS	CURSO	HORAS PEDAGÓGICAS POR CURSO								
			PRIMERO			SEGUNDO			TERCERO		
		MODALIDAD	SEMIPRESENCIAL		PRESENCIAL O A DISTANCIA	SEMIPRESENCIAL		PRESENCIAL O A DISTANCIA	SEMIPRESENCIAL		PRESENCIAL O A DISTANCIA
		ASIGNATURAS	PRESENCIAL	AUTÓNOMO		PRESENCIAL	AUTÓNOMO		PRESENCIAL	AUTÓNOMO	
TRONCO COMÚN	Matemática	Matemática	80	40	120	80	40	120	80	80	160
	Ciencias Naturales	Física	40	40	80	40	40	80	80	80	160
		Química	40	40	80	40	40	80	80	40	120
		Biología	40	40	80	40	40	80	40	40	80
	Ciencias Sociales	Historia	40	40	80	40	40	80	80	40	120
		Educación para la Ciudadanía	40	40	80	40	40	80	-	-	-
		Filosofía	40	40	80	40	40	80	-	-	-
	Lengua y Literatura	Lengua y Literatura	80	40	120	80	40	120	80	80	160
	Lengua Extranjera	Inglés	80	40	120	80	40	120	80	40	120
	Educación Cultural y Artística	Educación Cultural y Artística	40	40	80	40	40	80	-	-	-
	Educación Física	Educación Física	40	-	40	40	-	40	40	-	40
	Módulo inter-áreas	Emprendimiento y Gestión	40	-	40	40	-	40	40	-	40
BACHILLERATO EN CIENCIAS	**Horas pedagógicas del tronco común**		**600**	**400**	**1000**	**600**	**400**	**1000**	**600**	**400**	**1000**
	Horas adicionales a discreción para tutorías		200	-	200	200	-	200	200	-	200
	Horas pedagógicas totales del Bachillerato en Ciencias		**800**	**400**	**1200**	**800**	**400**	**1200**	**800**	**400**	**1200**
BACHILLERATO TÉCNICO	Horas adicionales para Bachillerato Técnico		400	200	600	400	200	600	400	200	600
	Horas pedagógicas totales del Bachillerato Técnico		**1200**	**600**	**1800**	**1200**	**600**	**1800**	**1200**	**600**	**1800**

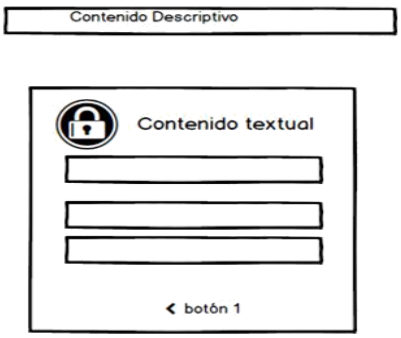
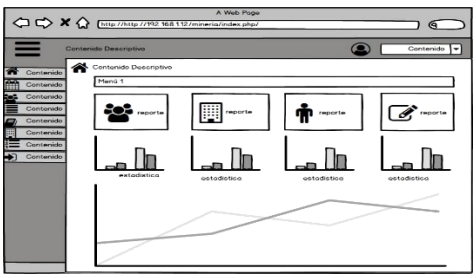
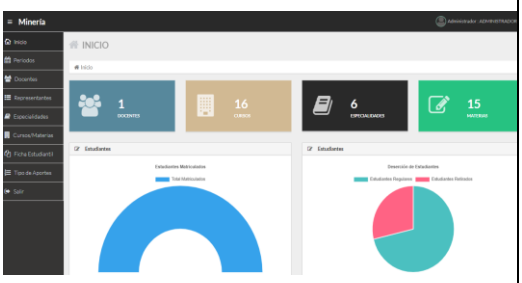
Anexo 2: Proceso de calificaciones

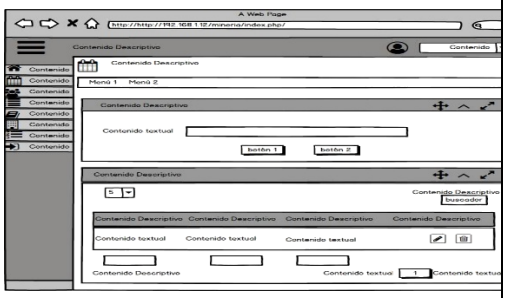
		Nota Final = (Quimestre 1 + Quimestre 2) / 2							
		Quimestre 1				Quimestre 2			
		Parcial 1	Parcial 2	Parcial 3	Examen Quimestral	Parcial 1	Parcial 2	Parcial 3	Examen Quimestral
100%	20%	Tareas	Tareas	Tareas		Tareas	Tareas	Tareas	
	20%	Actividades individuales	Actividades individuales	Actividades individuales		Actividades individuales	Actividades individuales	Actividades individuales	
	20%	Actividades grupales	Actividades grupales	Actividades grupales		Actividades grupales	Actividades grupales	Actividades grupales	
	20%	Lecciones	Lecciones	Lecciones		Lecciones	Lecciones	Lecciones	
	20%	Nota Sumativa	Nota Sumativa	Nota Sumativa		Nota Sumativa	Nota Sumativa	Nota Sumativa	
		80%			20%	80%			20%
		100%				100%			

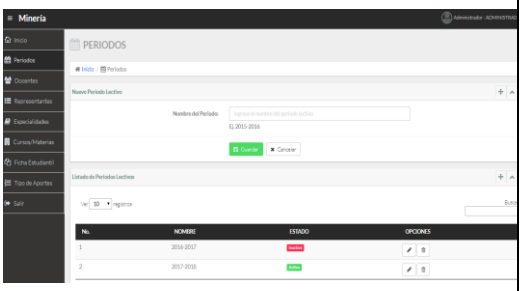
Anexo 3: Escala Cualitativa, cuantitativa del Ministerio de Educación del Ecuador

Escala cualitativa	Escala cuantitativa
Supera los aprendizajes requeridos.	10
Domina los aprendizajes requeridos.	9
Alcanza los aprendizajes requeridos.	7 - 8
Está próximo a alcanzar los aprendizajes requeridos.	5 - 6
No alcanza los aprendizajes requeridos.	≤ 4

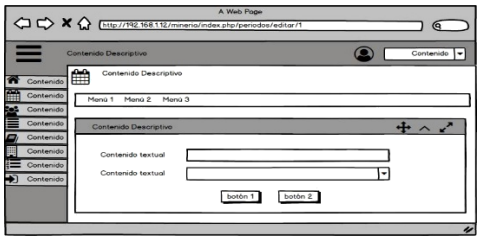
Anexo 4: Diseño del sistema informático

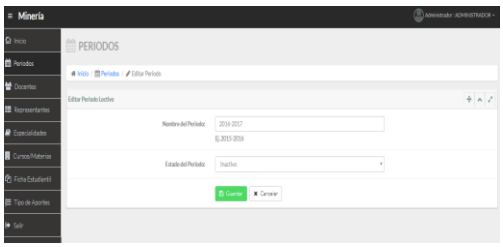
DISEÑO	RESULTADO
	 Formulario de Inicio de Sesión
	 Tablero de Indicadores Clave “DashBoard”



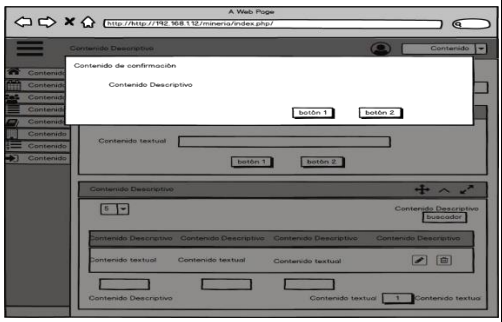


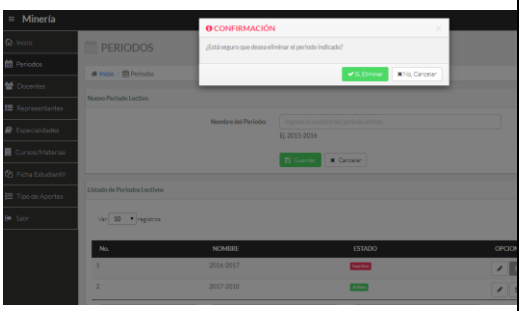
Interfaz de Gestión de Entidades





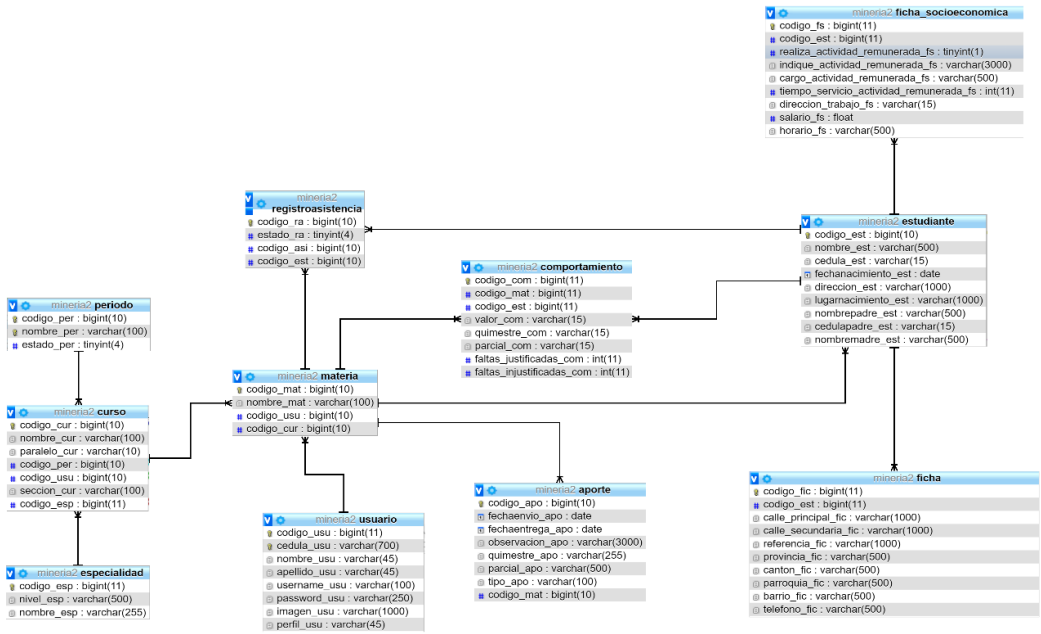
Interfaz de Edición de Entidades





Interfaz de Eliminación de Entidades

Anexo 5: La base de datos tiene el siguiente Diagrama Entidad Relación:



Anexo 6: Ficha estudiantil

Ficha Estudiantil Nº 0002

UNIDAD EDUCATIVA PCEI MONSEÑOR LEÓNIDAS PROAÑO LATACUNGA
DEPARTAMENTO DE CONSEJERIA ESTUDIANTIL
FICHA ESTUDIANTIL

Año Lectivo: 2018-2019

Los campos que "NO" se pueden editar se encuentran vinculados a otras áreas del sistema.

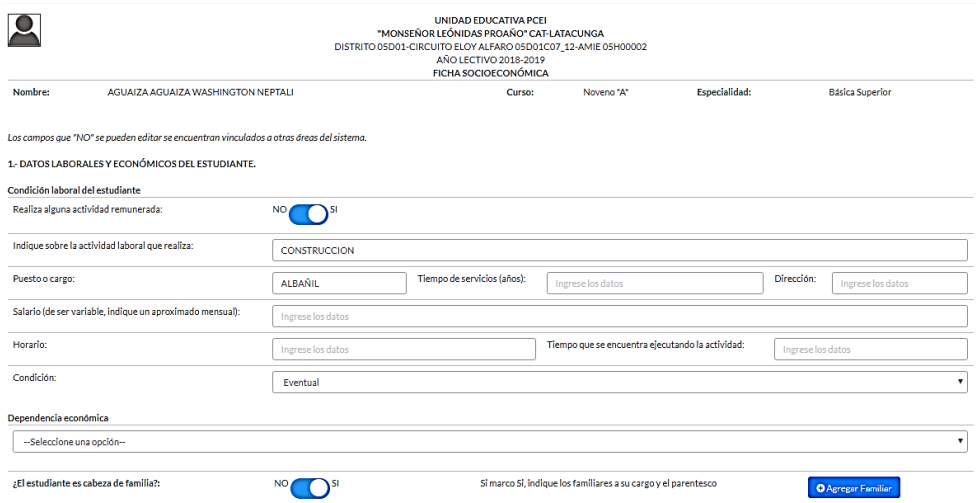
1.- DATOS DE IDENTIFICACIÓN/INFORMACIÓN.

Apellidos y nombres del/la estudiante:	AGUAIZA AGUAIZA WASHINGTON NEPTALI						
Grado/Año:	Noveno	Paralelo:	A	Jornada:	Educación General Básica - Básica Superior		
Cédula de Ciudadanía:	0504728502	Lugar Nacimiento:				Fecha Nacimiento:	29/03/2003
Nombre del/la representante legal:							
Dirección:	Calle Principal:	JUNTO AL CUERPO DE BOMBEROS		Calle Secundaria:	Ingrese calle secundaria		
Referencia:	Ingrese una referencia						
Provincia:	COTOPAXI	Cantón:	LATACUNGA	Parroquia:	TANICUCHI	Barrio/Sector	LASSO
Teléfonos:	0962653569			Etnia:	MESTIZO		
En caso de emergencia llamar a:	0968511202						

2.- DATOS FAMILIARES.

Madre:	COCHA MARIA	Padre:	AGUAIZA COCHA ANGEL MARIA	*Representante:	

Anexo 7: Ficha socioeconómica del estudiante



UNIDAD EDUCATIVA PCEI
"MONSEÑOR LEÓNIDAS PROAÑO" CAT-LATACUNGA
DISTRITO 05D01-CIRCUITO ELOY ALFARO 05D01C07_12-AMIE 05H00002
AÑO LECTIVO 2018-2019
FICHA SOCIOECONÓMICA

Nombre: AGUAIZA AGUAIZA WASHINGTON NEPTALI **Curso:** Noveno "A" **Especialidad:** Básica Superior

Los campos que "NO" se pueden editar se encuentran vinculados a otras áreas del sistema.

1.- DATOS LABORALES Y ECONÓMICOS DEL ESTUDIANTE.

Condición laboral del estudiante

Realiza alguna actividad remunerada: NO SI

Indique sobre la actividad laboral que realiza: CONSTRUCCION

Puesto o cargo: ALBAÑIL Tiempo de servicios (años): Ingrese los datos Dirección: Ingrese los datos

Salario (de ser variable, indique un aproximado mensual): Ingrese los datos

Horario: Ingrese los datos Tiempo que se encuentra ejecutando la actividad: Ingrese los datos

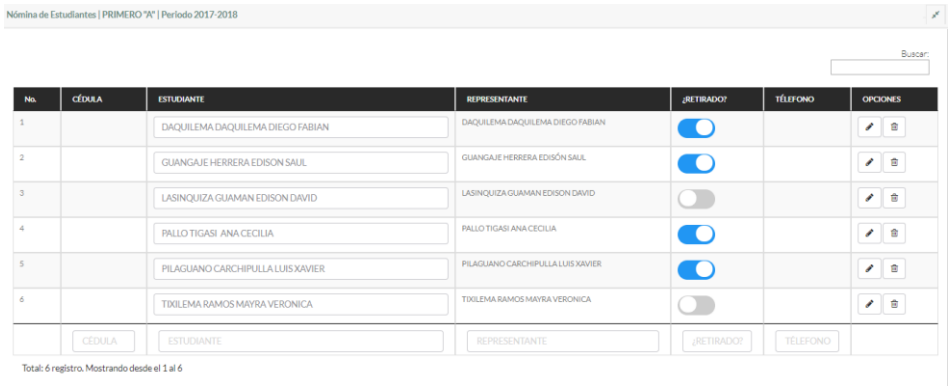
Condición: Eventual

Dependencia económica

--Seleccione una opción--

¿El estudiante es cabeza de familia?: NO SI Si marco SI, indique los familiares a su cargo y el parentesco Agregar Familiar

Anexo 8: Sistema informático de apoyo del modelo de análisis propuesto



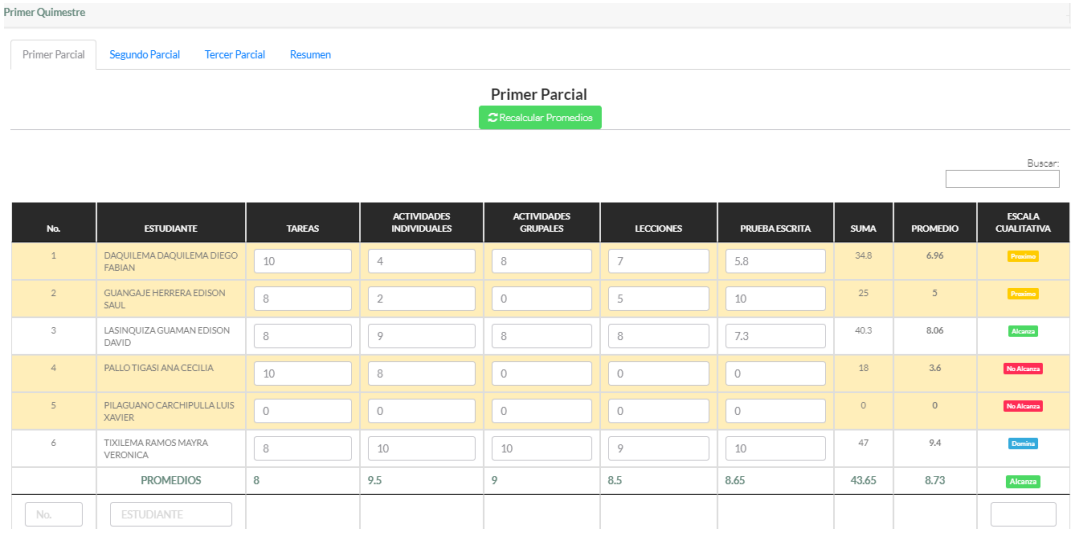
Nómina de Estudiantes | PRIMERO "A" | Período 2017-2018

Buscar:

No.	CÉDULA	ESTUDIANTE	REPRESENTANTE	¿RETIRADO?	TÉLEFONO	OPCIONES
1		DAQUILEMA DAQUILEMA DIEGO FABIAN	DAQUILEMA DAQUILEMA DIEGO FABIAN			
2		GUANGAJE HERRERA EDISON SAUL	GUANGAJE HERRERA EDISÓN SAUL			
3		LASINQUIZA GUAMAN EDISON DAVID	LASINQUIZA GUAMAN EDISON DAVID			
4		PALLO TIGASI ANA CECILIA	PALLO TIGASI ANA CECILIA			
5		PILAGUANO CARCHIPULLA LUIS XAVIER	PILAGUANO CARCHIPULLA LUIS XAVIER			
6		TIXILEMA RAMOS MAYRA VERONICA	TIXILEMA RAMOS MAYRA VERONICA			
	CÉDULA	ESTUDIANTE	REPRESENTANTE	¿RETIRADO?	TÉLEFONO	

Total: 6 registro. Mostrando desde el 1 al 6

Anexo 9: Sistema informático de apoyo del modelo de análisis propuesto.



Primer Quimestre

Primer Parcial | Segundo Parcial | Tercer Parcial | Resumen

Primer Parcial

Recalcular Promedios

Buscar:

No.	ESTUDIANTE	TAREAS	ACTIVIDADES INDIVIDUALES	ACTIVIDADES GRUPALES	LECCIONES	PRUEBA ESCRITA	SUMA	PROMEDIO	ESCALA CUALITATIVA
1	DAQUILEMA DAQUILEMA DIEGO FABIAN	10	4	8	7	5.8	34.8	6.96	Proximo
2	GUANGAJE HERRERA EDISON SAUL	8	2	0	5	10	25	5	Proximo
3	LASINQUIZA GUAMAN EDISON DAVID	8	9	8	8	7.3	40.3	8.06	Alcanza
4	PALLO TIGASI ANA CECILIA	10	8	0	0	0	18	3.6	No Alcanza
5	PILAGUANO CARCHIPULLA LUIS XAVIER	0	0	0	0	0	0	0	No Alcanza
6	TIXILEMA RAMOS MAYRA VERONICA	8	10	10	9	10	47	9.4	Domina
	PROMEDIOS	8	9.5	9	8.5	8.65	43.65	8.73	Alcanza
No.	ESTUDIANTE								

Anexo 101: Promedio de las puntuaciones asignadas mediante el criterio de expertos.

Requerimiento		**Puntuación**					**Validación (Si/No)**	**Verificación (Si/no)**
Descripción	**Criterio**	**1**	**2**	**3**	**4**	**5**		
El Modelo de análisis de datos del Rendimiento académico recoge información que permite dar respuesta al problema de investigación	Adecuado					x	Si	Si
	Pertinencia				x		Si	Si
El Modelo de análisis de datos del Rendimiento académico responde a los objetivos del estudio	Adecuado				x		Si	Si
	Pertinencia					x	Si	Si
La estructura del Modelo de análisis de datos del Rendimiento académico es adecuada	Adecuado				x		Si	Si
	Pertinencia					x	Si	Si
Los ítems del Modelo de análisis de datos del Rendimiento académico responden a los objetivos del estudio	Adecuado				x		Si	Si
	Pertinencia				x		Si	Si
La secuencia presentada ayuda en el desarrollo del Modelo de análisis de datos del Rendimiento académico	Adecuado				x		Si	Si
	Pertinencia				x		Si	Si
Los ítems del Modelo de análisis de datos del Rendimiento académico son claros y entendibles	Adecuado					x	Si	Si
	Pertinencia					x	Si	Si
El número de ítems del Modelo de análisis de datos del Rendimiento académico es adecuado para su aplicación	Adecuado				x		Si	Si
	Pertinencia				x		Si	Si

Anexo 11: Gastos Directos del Software

Gastos	Detalle	Cantidad	V. Unitario	Total
Software	PHP versión 7	1	Licencia Gratuita	$0.00
	Weka	1	Licencia Gratuita	$0.00
	Navegador de Internet Chrome/Firefox	1	Licencia Gratuita	$0.00
	MySQL	1	Licencia Gratuita	$0.00
	StarUML (Diagramas)	1	Licencia de prueba	$0.00
	Paquete de Office 2016 (Documentación)	1	$40.00	$40.00
	Internet	12 meses	$18.00	$216
Sistema	Desarrollo	470 puntos de función	$20.00	$9,400
Total				$9,656

Anexo 12: Gastos Directos Servidor.

Descripción:	Mínimo	Costos
Memoria	512 RAM	$250,00
Procesador	1GHz	
Unidad	Unidad de DVD ROM	
Espacio en disco disponible	8 GBYTES	
Pantalla y periféricos	Súper VGA (800x600) o superior	
Valor Total		**$250.oo**

Anexo 13: Gastos Indirectos.

Gastos Indirectos			
Detalle	**Cantidad**	**Valor Unitario**	**Total**
Pasajes	20	$1.00	$20.00
Alimentación	40	$2.50	$100.00
Comunicación	10	$3	$30.00
Copias	200	$0.03	$6.00
Esferos	4	$0.50	$2.00
Total			**$158.00**

Anexo 14: Gastos Totales.

Descripción	Valor
Total Gastos Directos	$9,906.00
Total Gastos Indirectos	$158.00
Gastos Directos + Gastos Indirectos	$10,064
Imprevistos (10%)	$1,006.40
Total	**$11,070.40**

Anexo 15: Autorización para la investigación.

UNIDAD EDUCATIVA P.C.E.I.
"Monseñor Leonidas Proaño"
Dirección: Av. Iberoamericana y México (Barrio San Felipe)
Teléfono: 032252225 e-mail: uedcotopaxi@gmail.com / distritolatacunga05h00002@gmail.com
Latacunga - Ecuador

"Nunca es tarde para aprender"

Latacunga, 23 Julio de 2019

MSc.
Juan Francisco Ulloa Aguilera
RECTORA DE LA UNIDAD EDUCATIVA P.C.E.I. "MONSEÑOR LEONIDAS PROAÑO"
Ciudad. -

Asunto: Solicitud de permiso para facilitar información documentada de los estudiantes de la Unidad Educativa.

Yo, CHANCÚSIG TAIPICAÑA DIEGO MARCELO con C.C. Nº 0502309388 en calidad de docente de la institución, solicito de la manera más comedida a Usted me facilite obtener información documentada de los estudiantes de la institución ya que me encuentro realizando mi tesis con el tema: **"Modelo de análisis de del rendimiento académico de la Unidad educativa P.C.E.I. "Monseñor Leonidas Proaño" del cantón Latacunga a través de minería de datos,** el mismo que se aplicará los resultados obtenidos de dicho modelo en la institución.

Por la gentil atención que se digne dar a la presente, le expreso mi agradecimiento.

De Usted, muy atentamente,

Ing. Diego Marcelo Chancúsig T.
DOCENTE DE LA U.E. PCEI "MLP"
0502309388

Anexo 16: Respuesta de la Autorización para la investigación.

UNIDAD EDUCATIVA P.C.E.I.
"Monseñor Leonidas Proaño"
Dirección: Av. Iberoamericana y México (Barrio San Felipe)
Teléfono: 032252225 e-mail: uedcotopaxi@gmail.com / distritolatacunga05h00002@gmail.com
Latacunga - Ecuador

Latacunga, 26 Julio de 2019

Ing.
Diego Marcelo Chancúsig T.
DOCENTE DE LA INSTITUCIÓN
En su despacho. -

A petición escrita del docente Ing. CHANCÚSIG TAIPICAÑA DIEGO MARCELO con C.C. Nº 0502309388 autorizo y dispongo el acceso a la documentación requerida de los estudiantes de la institución para el desarrollo del trabajo de tesis.

Atentamente,

Dr. Juan Olloa Aguilera
RECTOR - UEDC