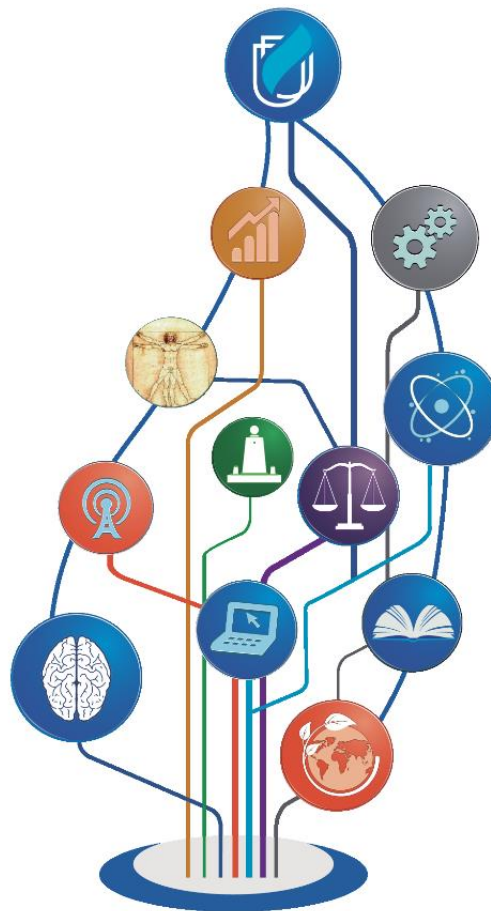


VII CONGRESO CIENTÍFICO INTERNACIONAL

“SOCIEDAD DEL CONOCIMIENTO:
RETOS Y PERSPECTIVAS. ACCIONES PARA
UN MUNDO SOSTENIBLE”



MEMORIAS

ISBN: 978-9942-960-75-7

Octubre 2022

TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN

NO.	TÍTULO PONENCIA	NOMBRE
001	ÁRBOLES DE DECISIÓN PARA CARACTERIZAR LOS PERFILES ACADÉMICOS DE LOS ESTUDIANTES DE TERCERO DE BACHILLERATO EN ECUADOR	Ing. Shirley Gabriela Alarcón Loza Karen Lissette Estacio Corozo, Mgtr.
002	CIENCIA ABIERTA: CONOCIMIENTO Y GRADO DE ACEPTACIÓN POR LA COMUNIDAD CIENTÍFICA	Gisella Aguilar Mera. Mgtr Jenny Garzón Balcázar. Mgtr
003	ECOLOGÍAS DE APRENDIZAJE DIGITAL Y SU INFLUENCIA EN LA PROTECCIÓN DEL PATRIMONIO CULTURAL INMATERIAL – CASO PODCASTS	Lourdes Parra Trelles, Mgtr. Iskra Sánchez Solórzano, Mgtr. Juan José Chacón Cornejo, Mgtr. Ricardo Coello Yagual, Mgtr.
004	EL USO DE OPENCL, COMO MECANISMO PARA LOGRAR PROGRAMACION PARALELA	Luis Espinoza Mendoza, Mgtr.
005	CLASES VIRTUALES, UN ANÁLISIS POST PANDEMIA	Erika Ascencio Jordán, Mgtr. Ing. George Ascencio Jordán

LÍNEA TEMÁTICA:



Tecnologías de la Información y la Comunicación

ISBN: 978-9942-960-75-7

Octubre 2022

001. ÁRBOLES DE DECISIÓN PARA CARACTERIZAR LOS PERFILES ACADÉMICOS DE LOS ESTUDIANTES DE TERCERO DE BACHILLERATO EN ECUADOR

DECISION TREES TO CHARACTERIZE THE ACADEMIC PROFILES OF STUDENTS IN THE THIRD OF HIGH SCHOOL IN ECUADOR

Autores:

Ing. Shirley Gabriela Alarcón Loza

Instituto Superior Tecnológico ARGOS

galarcon@tecnologicoargos.edu.ec

Karen Lissette Estacio Corozo, Mgtr.

Instituto Superior Tecnológico ARGOS

k_estacio@tecnologicoargos.edu.ec

RESUMEN

Los posibles desertores en la educación superior pueden anticiparse, mediante el análisis computacional de los datos, para disminuir el impacto en las entidades educativas. Este artículo tiene como objetivo caracterizar los perfiles académicos de los estudiantes de 3° Bachillerato General Unificado a través de un modelo predictivo que permita reconocer si un estudiante podría ubicarse en bajo rendimiento académico, desde la perspectiva de las características sociodemográficas y el tipo de financiamiento de las instituciones educativas secundarias. La base de datos incluyó el registro de aspectos sociodemográficos y promedios de varias asignaturas cursadas en el periodo lectivo 2018-2019, de una muestra de 2417 estudiantes. En la investigación, se consideró el Promedio Global para evaluar la condición de los estudiantes respecto del bajo rendimiento académico que incide en los perfiles académicos. El presente trabajo consistió en un diseño correlacional de corte transversal en el que se empleó la metodología KDD y, a partir del software WEKA, se ejecutaron algoritmos de clasificación para determinar atributos predictores, siendo el índice socioeconómico y el tipo de financiamiento de la institución educativa, los más significativos. Los resultados mostraron que, para la variable NL_GLOBAL, los dos mejores algoritmos de clasificación fueron REP Tree y J48, cuyo valor de precisión fue 67,3149% y 67,108%, respectivamente; mientras que para el área bajo la curva (ROC-Area) fue 0,616 y 0,551, respectivamente. Se sugiere desarrollar investigaciones relacionadas con minería de datos que promuevan el mejoramiento de la calidad educativa en las instituciones de educación secundaria y superior a nivel nacional.

Palabras clave: Educación secundaria, Algoritmos de clasificación, Árboles de decisión, Perfil académico.

ABSTRACT

Possible dropouts in higher education can be anticipated, through computational data analysis, to reduce the impact on educational entities. The objective of this article is to characterize the academic profiles of the students of the 3rd Unified General Baccalaureate through a predictive model that allows recognizing if a student could be located in low academic performance, from the perspective of the sociodemographic characteristics and the type of financing. secondary educational institutions. The database included the record of sociodemographic aspects and averages of various subjects taken in the 2018-2019 academic period, from a sample of 2,417 students. In the investigation, the Global Average was excluded to evaluate the condition of the students regarding the low academic performance that affects the academic profiles. The present work consisted of a cross-sectional correlational design in which the KDD methodology was used and, from the WEKA software, classification algorithms were executed to determine predictive attributes, being the socioeconomic index and the type of educational financing of the institution, the most significant. The results showed that, for the NL_GLOBAL variable, the two best classification algorithms were REP Tree and J48, whose precision value was 67.3149% and 67.108%, respectively; while for the area under the curve (ROC-Area) it was 0.616 and 0.551, respectively. It is suggested to develop research related to data mining that promotes the improvement of educational quality in secondary and higher education institutions at the national level.

Keywords: Secondary education, Classification algorithms, Decision trees, Academic profile.

INTRODUCCIÓN

En los últimos años, la tecnología ha desencadenado la generación de una gran cantidad de datos que se almacenan en servidores a nivel mundial sin recibir una evaluación, para luego eliminarse sin haber aplicado acciones. La tendencia actual radica en el uso de datos y transformarlos en información valiosa mediante el análisis computacional.

En ese contexto, nace la minería de datos como una subdisciplina de las ciencias de la computación que ha logrado mucho reconocimiento en los últimos años, dado que, puede ser usada para diferentes propósitos en las demás disciplinas de las ciencias (Valero y otros, 2022).

Dentro del ámbito de la educación, se encuentra el bajo rendimiento académico que es un estado en el que convergen diversos factores que colocan al estudiante en una situación de propensión de deserción escolar, las problemáticas derivadas del bajo rendimiento académico han sido temas comunes de preocupación y estudio en el contexto educativo (Ayala y otros, 2021). Si bien el rendimiento académico caracteriza individualmente al alumno, es posible obtener a partir de él, ciertas características generales de grupos de estudiantes; es allí donde radica la importancia de contar con métodos confiables de evaluación del rendimiento, partiendo de una correcta organización de los datos, que pueda brindar una visión clara de los inconvenientes en el desempeño de los estudiantes dentro de las instituciones educativas (La Red y otros, 2015).

Las instituciones de educación superior se ven en la necesidad de contar con mecanismos que ayuden a disminuir la deserción de los estudiantes en los primeros niveles de estudio (Amaya y otros, 2014). A partir de dicho fundamento, el objetivo del presente artículo es caracterizar los perfiles académicos de los estudiantes de 3° Bachillerato General Unificado a través de un modelo predictivo que permita reconocer si un estudiante podría ubicarse en bajo rendimiento académico, desde la perspectiva de las características sociodemográficas y el tipo de financiamiento de las instituciones educativas.

Para tal efecto, se organizaron los datos académicos y sociodemográficos de los estudiantes de la muestra seleccionada, se aplicó la metodología Knowledge Discovery in Databases (KDD), se utilizó el software Waikato Environment for Knowledge Analysis (WEKA) para la generación de árboles de decisión mediante la ejecución de algoritmos de clasificación y se interpretaron los resultados.

DESARROLLO

El bajo rendimiento académico es un estado que se puede presentar en cualquier etapa de la vida estudiantil, colocando a los estudiantes en un nivel de aprendizaje insuficiente ya que no han asimilado todos los conocimientos ni las destrezas impartidas en el periodo académico.

Aunado a lo anterior, la minería de datos ayuda a predecir posibles escenarios desfavorables en la educación, de tal manera que permite aprovechar los datos generados en cada periodo. La Tabla 1, recopila los trabajos más importantes acerca de situaciones académicas que han sido evaluadas por varios algoritmos y metodologías de minería de datos.

Tabla 1. Revisión bibliográfica sobre metodología y algoritmos de minería de datos

Ítem	Estudio	Metodología	Software	Algoritmos	Trabajos
1	Identificación de factores de riesgo en alumnos	KDD	WEKA	K-means A priori	(Flores y otros, 2017)
2	Predicción y Análisis de los Resultados de la Prueba de Calidad de la Educación Superior	KDD	MATLAB	Redes neuronales	(García y otros, 2019)
3	Factores asociados al éxito de los estudiantes	KDD CRISP-DM	SPSS Statistic	C5.0 CHAID	(Mancilla y otros, 2020)
4	Modelo predictivo de deserción estudiantil	KDD	WEKA	J48 IDE3	(Amaya y otros, 2014)
5	Detección de Patrones de Bajo Rendimiento Académico	KDD	TariyKDD	C4.5 EquipAsso	(Timarán, 2009)
6	Modelos predictivos de riesgo académico en carreras de computación	KDD	WEKA	J48 LMT Logistic MultilayerPerceptron Random Forest	(Ayala y otros, 2021)
7	Perfiles de Rendimiento Académico	SEMMA	IBM Data Warehouse Edition	C4.5	(La Red y otros, 2015)

8	Predicción de la deserción escolar	KDD	WEKA	C4.5 K vecinos más cercanos	(Valero y otros, 2022)
9	Identificación de patrones de deserción estudiantil	KDD	Data Mining Tool Orage	Árboles de decisión	(Calvache Fernández y otros, 2018)
10	Identificación de Patrones de Rendimiento Académico	CRISP-DM	WEKA	J48	(Solís y otros, 2019)
11	Identificación de estudiantes universitarios desertores	KDD		C4.5 REP Tree Bosques aleatorios SVM Bayes ingenuo	(Urbina y otros, 2021)
12	Predicción del rendimiento académico	No específica	WEKA	J48 Regresión Logística Redes neuronales Naive de Bayes	(Menacho, 2017)
13	Factores asociados al desempeño académico	CRISP-DM	WEKA	J48	(Timarán y otros, 2019)
14	Deserción escolar universitaria: patrones para prevenirla	KDD	WEKA	C4.5	(Cruz y otros, 2020)

Base de datos

El objeto de estudio se basó en 1777 instituciones educativas del Sistema Nacional de Educación, que ofertaron servicios de Educación General Básica (EGB) y Bachillerato General Unificado (BGU) en el régimen Costa y Sierra. Las características de los estudiantes fueron: estar matriculados en el 4°, 7°, 10° de EGB y 3° BGU; estudiar en instituciones con tipo de financiamiento público (fiscal y municipal), privado (particular) o

mixto (fiscomisional) ubicadas en zonas urbanas y rurales; pertenecer a cualquier nivel socioeconómico; cursar las asignaturas de Matemáticas, Lenguaje, Ciencias Naturales y Estudios Sociales.

En la Tabla 2 se muestra la población que corresponde a los 48383 estudiantes del periodo lectivo 2018-2019, que guardan las características mencionadas. Los estudiantes matriculados en el último año del BGU están próximos a graduarse lo que implica su posterior inserción a las universidades. A partir de este aspecto se seleccionó la muestra empleando la técnica no probabilística por conveniencia, considerando a los estudiantes de 3° Bachillerato General Unificado.

Tabla 2. Distribución de la población y muestra empleada

Población	Muestra	
	Estudiantes	Institución
48383 estudiantes	2417 estudiantes del 3° BGU	Área rural Financiamiento: público, privado y mixto

Para este estudio se consideró la base de datos de 2417 estudiantes con las características que se reflejan en la Tabla 3. Los aspectos sociodemográficos resultan del registro inicial que los padres de familia realizan para matricular a sus hijos; las asignaturas corresponden a los cuatro campos del conocimiento elegidos por el Instituto Nacional de Evaluación Educativa y, los niveles de logro están basados en los estándares de aprendizaje aprobados por el Ministerio de Educación.

Tabla 3. Características de los estudiantes 3° BGU, periodo lectivo 2018- 2019

Características	Descripción	Características	Descripción
Aspectos sociodemográficos	Código	Promedio General de las asignaturas	Matemática
	Edad		Lengua y
	Sexo		Literatura Ciencias
	Quintil		Naturales Estudios
	Etnia		Sociales

Institución	Financiamiento	Promedio Global según los niveles de logro	Insuficiente Elemental Satisfactorio Excelente
-------------	----------------	--	---

Como herramienta tecnológica se empleó el software WEKA (Frank y otros, 2017), versión 3.8.6, que es considerado como un software de minería de datos, de código abierto que aloja una serie de algoritmos de aprendizaje automático permitiendo el procesamiento de una gran cantidad de datos para su posterior análisis e interpretación.

METODOLOGÍA

El presente artículo es un diseño correlacional de corte transversal que emplea un Proceso de descubrimiento de conocimiento en bases de datos o KDD (Calvache-Fernández y otros, 2018) para el proceso de minería de datos, dado que, posee la particularidad de crear conjuntos de datos mediante el pre procesamiento y la limpieza de los mismos en su fase de selección y preparación. Es utilizado en varios escenarios porque comienza con un análisis y comprensión del problema organizacional antes de comenzar el proceso de minería (Moine y otros, 2011). Esta metodología consta de cinco fases que se describe a continuación:

Fase de Selección

Los datos académicos se generan en cada periodo lectivo en las instituciones educativas de Ecuador, siendo almacenados en el software correspondiente y recopilados en la Base de Información del Instituto Nacional de Evaluación Educativa, por lo que es imprescindible que reciban algún tipo de tratamiento para descubrir patrones de comportamiento que provean información valiosa del estado de la educación a nivel nacional y sus repercusiones en la educación superior. Al identificar este escenario, se planteó la presente investigación con el objetivo de caracterizar los perfiles académicos de los estudiantes de 3° BGU a través de un modelo predictivo que permita reconocer si un estudiante podría ubicarse en bajo rendimiento académico, desde la perspectiva de las características sociodemográficas y el tipo de financiamiento de las instituciones educativas, para que las instituciones de educación superior tomen decisiones oportunas al momento de recibir estudiantes en el primer nivel. Para esto, se utilizaron algoritmos de clasificación de minería de datos y a su vez, se validaron los más idóneos a partir de los datos recopilados de la muestra.

Fase de Pre procesamiento

En esta fase se utilizó una base de datos que contiene las características de los estudiantes y de las instituciones educativas, constituyendo 48383 registros de 32 campos diferentes. Estos campos fueron estandarizados y almacenados en Microsoft Excel, de tal manera que quedaron 2417 registros de 11 campos diferentes, mostrada en la Tabla 4.

Fase de Transformación

Para determinar el perfil académico de los estudiantes, se ha distinguido una arista denominada rendimiento académico y en éste se ha seleccionado el Promedio Global para evaluar la condición de los estudiantes respecto del bajo rendimiento académico. Con ese fundamento, es indispensable tener una variable dependiente, de tipo nominal. A partir de la misma se ejecutan varias técnicas de clasificación de minería de datos en WEKA. La variable dependiente que afecta directamente al rendimiento académico y, por ende, incide en el bajo rendimiento, es la variable Promedio Global (NL_GLOBAL).

Para realizar el modelo predictivo, se determinó que el rendimiento académico tiene variables independientes y una dependiente. En la Tabla 4, se detallan las variables que fueron ingresadas al software WEKA para conocer los perfiles académicos asociados con el rendimiento. De acuerdo con la política del Instituto Nacional de Evaluación Educativa, el estudiante debe mantener un nivel de logro Elemental (700-799 puntos), Satisfactorio (800-949 puntos) o Excelente (950-1000 puntos) en el promedio general, incluyendo las asignaturas de los cuatro campos del conocimiento. Se encuentra en bajo rendimiento académico si su nivel de logro es Insuficiente (nota inferior o igual a 699 puntos).

Cabe resaltar que en esta fase se discretizaron los valores numéricos de la tabla original, transformándolos a valores nominales como es el caso de EDAD, SEXO, QUINTIL, ETNIA, FINANC y NL_GLOBAL para favorecer el proceso de minería.

Tabla 4. Descripción de las variables consideradas para el perfil académico

Variables	Descripción	Tipo
COD	Identificación única del estudiante	Nominal
EDAD	Edad al momento del estudio (2018)	Numérica
	Sexo del estudiante	

SEXO		Nominal (Hombre / Mujer)
	Índice socioeconómico	
QUINTIL		Nominal (Q1, Q2, Q3, Q4, Q5)
	Autoidentificación étnica	
ETNIA		Nominal (Afroecuatoriano, Montubio, Indígena, Mestizo/Blanco)
	Tipo de financiamiento de la institución educativa	
FINANC		Nominal (Público, Privado, Mixto)
	Promedio de la asignatura de Matemáticas	
MAT_PROM		Numérica
	Promedio de la asignatura de Lenguaje	
LENG_PROM		Numérica
	Promedio de la asignatura de Ciencias Naturales	
CIENC_PROM		Numérica
	Promedio de la asignatura de Estudios Sociales	
ESSOC_PROM		Numérica
	Nivel de logro alcanzado por el promedio global	
NL_GLOBAL		Nominal Ordinal (Insuficiente, Elemental, Satisfactorio, Excelente)
(variable clase)		

Fase de Minería de datos

A partir del conjunto de datos de prueba, se evalúa el rendimiento del modelo considerando varios experimentos en relación a la variable dependiente NL_GLOBAL. Los algoritmos de clasificación empleados para caracterizar los perfiles académicos relacionado con el rendimiento, fueron los árboles de decisión J48 (Ramos-Jiménez y otros, 2017), J48Consolidate, Random Forest (Ramaswami y otros, 2019), Random Tree y REP Tree (Calvache Fernández y otros, 2018). Además, se aplicó el modo de evaluación y

validación denominado *Cross-Validation Folds* basado en N iteraciones, que proporciona una precisión promedio por variable clase ejecutada.

En la Tabla 5, se han utilizado dos indicadores para la evaluación de la calidad del algoritmo de clasificación, que son ROC-Area (Área bajo la curva) y Precisión (Menacho, 2017). El área bajo la curva es la razón de verdaderos positivos frente a la razón de los falsos positivos, el comportamiento perfecto de este indicador se da cuando se acerca a la unidad. Con respecto a la Precisión, se tiene el porcentaje de los valores positivos que clasificaron la prueba.

Tabla 5. Evaluación de la calidad de los algoritmos de clasificación

Algoritmo	NL_GLOBAL	
	ROC-Area	Precisión
J48	0,551	67,108%
J48Consolidated	0,608	38,5188%
Random Forest	0,643	66,0323%
Random Tree	0,641	65,784%
REP Tree	0,616	67,3149%

Fase de Interpretación

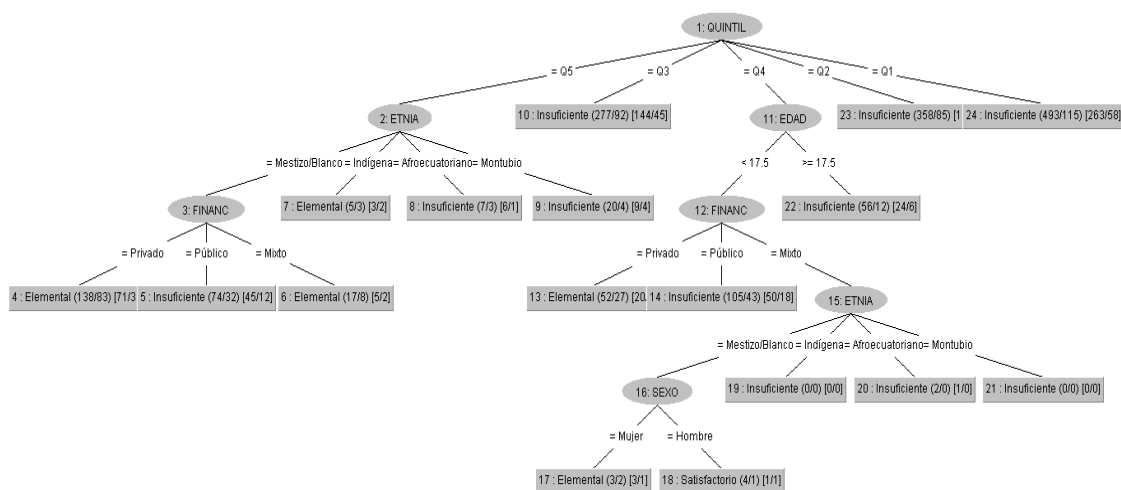
En el presente artículo, el modelo predictivo se generó por la aplicación de los algoritmos de clasificación en el software WEKA. Se realizaron varios ensayos, de tal manera que se escogieron dos algoritmos significativos. De acuerdo con la Tabla 5 se aplicó el algoritmo REP Tree para caracterizar los perfiles académicos de los estudiantes. En la Figura 1 se destacan las diecisiete reglas de conocimiento, indicando que, si hay estudiantes con bajo rendimiento académico, correspondientes a la categoría Insuficiente.

La interpretación de las reglas de conocimiento generadas, que declaran que existe bajo rendimiento académico, se explican a continuación: A) Si el valor QUINTIL es igual a 3, entonces si hay bajo rendimiento académico y se localizaron 277 estudiantes; B) Si el valor QUINTIL es igual a 2, entonces si hay bajo rendimiento académico y se localizaron 358 estudiantes; C) Si el valor QUINTIL es igual a 1, entonces no hay bajo rendimiento académico y se localizaron 493 estudiantes; D) Si el valor QUINTIL es igual a 5 y el valor de ETNIA es igual a Indígena, entonces si hay bajo rendimiento académico y se localizaron 5 estudiantes;

E) Si el valor QUINTIL es igual a 5 y el valor de ETNIA es igual a Afroecuatoriano, entonces si hay bajo rendimiento académico y se localizaron 7 estudiantes; F) Si el valor QUINTIL es igual a 5 y el valor de ETNIA es igual a Montubio, entonces si hay bajo rendimiento académico y se localizaron 20 estudiantes; G) Si el valor QUINTIL es igual a 4 y el valor de EDAD es mayor o igual a 17.5, entonces si hay bajo rendimiento académico y se localizaron 56 estudiantes; H) Si el valor QUINTIL es igual a 5 y el valor de ETNIA es igual a Mestizo/Blanco y el valor de FINANC es igual a Privado, entonces no hay bajo rendimiento académico y se localizaron 138 estudiantes; I) Si el valor QUINTIL es igual a 5 y el valor de ETNIA es igual a Mestizo/Blanco y el valor de FINANC es igual a Público, entonces si hay bajo rendimiento académico y se localizaron 74 estudiantes; J) Si el valor QUINTIL es igual a 5 y el valor de ETNIA es igual a Mestizo/Blanco y el valor de FINANC es igual a Mixto, entonces no hay bajo rendimiento académico y se localizaron 17 estudiantes; K) Si el valor QUINTIL es igual a 4 y el valor de EDAD es menor o igual a 17.5 y el valor de FINANC es igual a Privado, entonces no hay bajo rendimiento académico y se localizaron 52 estudiantes; L) Si el valor QUINTIL es igual a 4 y el valor de EDAD es menor o igual a 17.5 y el valor de FINANC es igual a Público, entonces si hay bajo rendimiento académico y se localizaron 105 estudiantes; M) Si el valor QUINTIL es igual a 4 y el valor de EDAD es menor o igual a 17.5 y el valor de FINANC es igual a Mixto y el valor de ETNIA es igual a Mestizo/Blanco y el valor SEXO es igual a Mujer, entonces no hay bajo rendimiento académico y se localizaron 3 estudiantes; N) Si el valor QUINTIL es igual a 4 y el valor de EDAD es menor o igual a 17.5 y el valor de FINANC es igual a Mixto y el valor de ETNIA es igual a Mestizo/Blanco y el valor SEXO es igual a Hombre, entonces no hay bajo rendimiento académico y se localizaron 4 estudiantes; O) Si el valor QUINTIL es igual a 4 y el valor de EDAD es menor o igual a 17.5 y el valor de FINANC es igual a Mixto y el valor de ETNIA es igual a Indígena, entonces no hay bajo rendimiento académico y se localizaron 0 estudiantes; P) Si el valor QUINTIL es igual a 4 y el valor de EDAD es menor o igual a 17.5 y el valor de FINANC es igual a Mixto y el valor de ETNIA es igual a Afroecuatoriano, entonces no hay bajo rendimiento académico y se localizaron 2 estudiantes; Q) Si el valor QUINTIL es igual a 4 y el valor de EDAD es menor o igual a 17.5 y el valor de FINANC es igual a Mixto y el valor de ETNIA es igual a Montubio, entonces no hay bajo rendimiento académico y se localizaron 0 estudiantes.

Figura 1

Reglas de conocimiento del algoritmo REP Tree para NL_GLOBAL



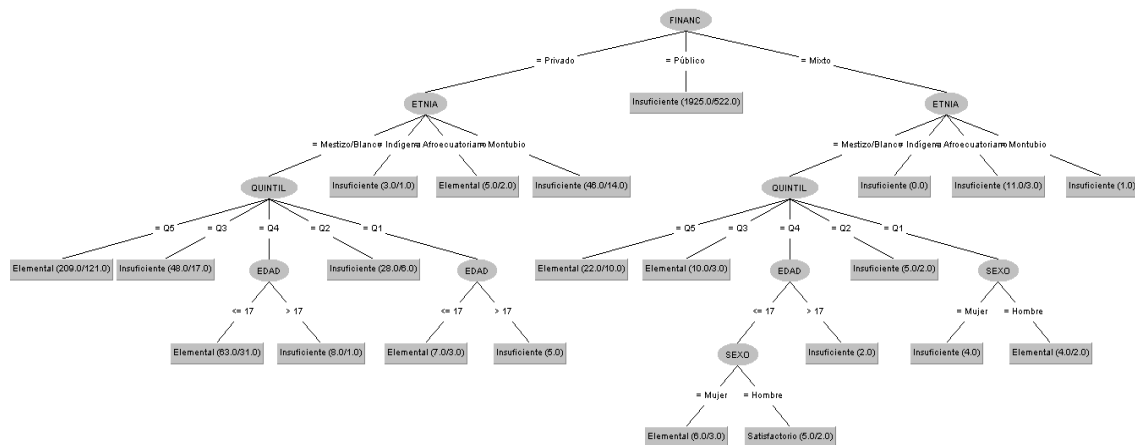
Del mismo modo, se aplicó el algoritmo J48 para caracterizar los perfiles académicos de los estudiantes. En la Figura 2 se destacan las veintidós reglas de conocimiento, indicando que, si hay estudiantes con bajo rendimiento académico, correspondientes a la categoría Insuficiente.

La interpretación de las reglas de conocimiento generadas que declaran que existe bajo rendimiento académico, se explican a continuación: A) Si el valor FINANC es igual a Público, entonces si hay bajo rendimiento académico y se localizaron 1925 estudiantes; B) Si el valor FINANC es igual a Privado, y el valor ETNIA es igual a Indígena, entonces si hay bajo rendimiento académico y se localizaron 3 estudiantes; C) Si el valor FINANC es igual a Privado, y el valor ETNIA es igual a Afroecuatoriano, entonces no hay bajo rendimiento académico y se localizaron 5 estudiantes; D) Si el valor FINANC es igual a Privado, y el valor ETNIA es igual a Montubio, entonces si hay bajo rendimiento académico y se localizaron 46 estudiantes; E) Si el valor FINANC es igual a Mixto, y el valor ETNIA es igual a Indígena, entonces no hay bajo rendimiento académico y se localizaron 0 estudiantes; F) Si el valor FINANC es igual a Mixto, y el valor ETNIA es igual a Afroecuatoriano, entonces si hay bajo rendimiento académico y se localizaron 11 estudiantes; G) Si el valor FINANC es igual a Mixto, y el valor ETNIA es igual a Montubio, entonces si hay bajo rendimiento académico y se localizaron 1 estudiantes; H) Si el valor FINANC es igual a Privado, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 5 entonces no hay bajo rendimiento académico y se localizaron 209 estudiantes; I) Si el valor FINANC es igual a Privado, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 3, entonces si hay bajo rendimiento académico y se localizaron

48 estudiantes; J) Si el valor FINANC es igual a Privado, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 2, entonces si hay bajo rendimiento académico y se localizaron 28 estudiantes; K) Si el valor FINANC es igual a Mixto, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 5, entonces no hay bajo rendimiento académico y se localizaron 22 estudiantes; L) Si el valor FINANC es igual a Mixto, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 3, entonces no hay bajo rendimiento académico y se localizaron 10 estudiantes; M) Si el valor FINANC es igual a Mixto, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 2, entonces si hay bajo rendimiento académico y se localizaron 5 estudiantes; N) Si el valor FINANC es igual a Privado, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 4, y el valor EDAD es menor o igual a 17, entonces no hay bajo rendimiento académico y se localizaron 63 estudiantes; O) Si el valor FINANC es igual a Privado, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 4, y el valor EDAD es mayor a 17, entonces si hay bajo rendimiento académico y se localizaron 8 estudiantes; P) Si el valor FINANC es igual a Privado, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 1, y el valor EDAD es menor o igual a 17, entonces no hay bajo rendimiento académico y se localizaron 7 estudiantes; Q) Si el valor FINANC es igual a Privado, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 1, y el valor EDAD es mayor a 17, entonces si hay bajo rendimiento académico y se localizaron 5 estudiantes; R) Si el valor FINANC es igual a Mixto, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 4, y el valor EDAD es mayor a 17, entonces si hay bajo rendimiento académico y se localizaron 2 estudiantes; S) Si el valor FINANC es igual a Mixto, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 1, y el valor SEXO es igual a Mujer, entonces si hay bajo rendimiento académico y se localizaron 4 estudiantes; T) Si el valor FINANC es igual a Mixto, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 1, y el valor SEXO es igual a Hombre, entonces no hay bajo rendimiento académico y se localizaron 4 estudiantes; U) Si el valor FINANC es igual a Mixto, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 4, y el valor EDAD es menor o igual a 17, y el valor SEXO es igual a Mujer, entonces no hay bajo rendimiento académico y se localizaron 6 estudiantes; V) Si el valor FINANC es igual a Mixto, y el valor ETNIA es igual a Mestizo/Blanco, y el valor QUINTIL es igual a 4, y el valor EDAD es menor o igual a 17, y el valor SEXO es igual a Hombre, entonces no hay bajo rendimiento académico y se localizaron 5 estudiantes.

Figura 2

Reglas de conocimiento del algoritmo J48 para NL_GLOBAL



RESULTADOS

Para la variable Promedio Global (NL_GLOBAL) se aplicaron varios algoritmos de los cuales el algoritmo REP Tree obtuvo un alto porcentaje en su precisión y margen de calidad (Maya y otros, 2018), según la evaluación mostrada en la Tabla 5. Al generarlo, desglosó diecisiete reglas de conocimiento que consideran la existencia de alto o bajo rendimiento, evaluando el nivel de logro alcanzado por los estudiantes en las cuatro asignaturas. Los resultados exponen que el índice socioeconómico es un factor predominante para determinar el perfil académico, ya que los estudiantes ubicados en el quintil 1, 2 y 3 poseen un nivel de logro Insuficiente. Por otra parte, los estudiantes asignados al quintil 5, se ven mayormente influenciados por el tipo de financiamiento de la institución educativa, en este caso se localizaron 74 estudiantes en un nivel de logro Insuficiente provenientes de instituciones públicas; mientras que 7 estudiantes afroecuatorianos y 20 montubios se encuentran en la misma situación. Finalmente, los estudiantes ubicados en el quintil 4 se ubican en un nivel de logro Insuficiente cuando son mayores o iguales a 17.5 años, y los menores de esta edad provienen de instituciones públicas, o vienen de instituciones mixtas y son afroecuatorianos. En otra posición, el hecho que los estudiantes tengan un índice socioeconómico alto, no garantiza que no se ubiquen en un nivel de logro Insuficiente.

Del mismo modo, se aplicó el algoritmo J48 obteniendo un buen porcentaje en su precisión y margen de calidad, según la evaluación mostrada en la Tabla 5. La calidad de este algoritmo también se ve reflejada en el conjunto de datos evaluado en la investigación realizada por Urbina y otros (2021), Mhetre y Nagar (2017), donde sus resultados fueron

los mejores. Al generarlo, desglosó veintidós reglas de conocimiento que consideran la existencia de alto o bajo rendimiento, evaluando el nivel de logro alcanzado por los estudiantes en las cuatro asignaturas. Los resultados exponen que el tipo de financiamiento de la institución educativa es un factor predominante para determinar el perfil académico, ya que los estudiantes matriculados en instituciones públicas poseen un nivel de logro Insuficiente. Por otra parte, los estudiantes matriculados en instituciones privadas que tienen un nivel Insuficiente son aquellos mestizos o blancos ubicados en el quintil 2 y 3, así como los estudiantes del quintil 1 y 4 que son mayores a 17 años. Finalmente, los estudiantes matriculados en instituciones mixtas con nivel Insuficiente son estudiantes afroecuatorianos y montubios, de la misma manera que estudiantes mestizos o blancos ubicados en el quintil 2, así como los ubicados en el quintil 4 mayores de 17 años, y los ubicados en el quintil 1 correspondientes al sexo femenino. En otra posición, el hecho que los estudiantes estén matriculados en instituciones privadas, no garantiza que no se ubiquen en un nivel de logro Insuficiente.

CONCLUSIONES

La minería de datos se establece como una herramienta que puede atender los aspectos del funcionamiento de las instituciones que ofertan servicios a nivel de Educación General Básica y Bachillerato General Unificado, así como instituciones de educación superior, en donde se generen grandes volúmenes de datos. Las actividades en la formación secundaria son de alto impacto social, porque tratan con estudiantes y su futuro desarrollo profesional, que aporta a la sociedad. Con dicho fundamento, se incita a tomar mejores decisiones, empleando modelos computacionales que permitan predecir escenarios para contribuir al mejoramiento del servicio educativo en las instituciones públicas, privadas y mixtas del Ecuador.

Para esta investigación, se evidenció que el índice socioeconómico y el tipo de financiamiento de la institución tienen una mayor influencia en el perfil académico de los estudiantes, por lo que el Estado debería tomar medidas para el mejoramiento de la calidad educativa de las instituciones públicas, ya que presentan una gran cantidad de estudiantes con nivel Insuficiente, ubicándolos en un promedio global inferior o igual a 699 puntos. Respecto del índice socioeconómico, el Estado debe implementar y mejorar las políticas públicas ya que los estudiantes con menos ingresos económicos representan una gran parte de la muestra que no alcanza los niveles de aprendizaje esperados. En relación, emplear un modelo predictivo que se enfoca en el perfil académico permite observar a los estudiantes que tienen bajo rendimiento, así como la institución de procedencia, para

alertarles oportunamente y evitar que sean posibles desertores en los primeros niveles de la educación superior, lo que sugiere otros problemas.

Vinculado a lo anterior, sería interesante identificar y evaluar los datos que registran las instituciones de educación superior, para generar nuevas investigaciones donde se considere a los estudiantes que reciben becas de ayuda y se observen otros factores que influyen en el rendimiento académico, entre otros aspectos. Respecto del enfoque tecnológico, se sugiere realizar modelos predictivos con otros algoritmos de clasificación.

REFERENCIAS BIBLIOGRÁFICAS

- Amaya, Y., Barrientos, E., & Heredia, D. (2014). *Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos*. <https://cutt.ly/4VQ6Fna>
- Ayala, E., López, R., & Menendez, V. (2021, Abril). Modelos predictivos de riesgo académico en carreras de computación con minería de datos educativos. *RED. Revista de Educación a Distancia*, 66(21), 1-30. <https://doi.org/10.6018/red.463561>
- Calvache Fernández, L., Alvarez Vallejo, V., Triviño Arbeláez, J., Quiceno Restrepo, C., & Pulgarin Giraldo, R. (2018). Aplicación de técnicas de minería de datos para la identificación de patrones de deserción estudiantil como apoyo a las estrategias de SARA. *Octava Conferencia Latinoamericana sobre el Abandono en la Educación Superior*. <https://revistas.utp.ac.pa/index.php/clabes/article/view/2021/2969>
- Calvache-Fernández, L., Álvarez-Vallejo, V., & Triviño-Arbeláez, J. (2018). Proceso KDD como apoyo a las estrategias del proyecto SARA. *Revista Educación en Ingeniería*, 13(26), 82-89. <https://doi.org/http://dx.doi.org/10.26507/rei.v13n26.916>
- Cruz, R., Urbina, A., & Camino, J. (2020). Deserción escolar universitaria: Patrones para prevenirla aplicando minería de datos educativa. *RELIEVE Revista Electrónica de Investigación y Evaluación Educativa*, 26(1). <https://doi.org/10.7203/relieve.26.1.16061>
- Flores, A., Alejo, R., & Reyes, A. (2017). Minería de datos aplicada para la identificación de factores de riesgo en alumnos. *Research in Computing Science*, 139, 177–189. <https://doi.org/10.13053/rcs-139-1-14>
- Frank, E., Witten, I., Hall, M., & Pal, C. (2017). *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier.

- García, J., Sánchez, P., Orozco, M., & Obredor, S. (2019, Febrero 1). Extracción de Conocimiento para la Predicción y Análisis de los Resultados de la Prueba de Calidad de la Educación Superior en Colombia. *Formación universitaria*, 12(4), 55-62. [10.4067/S0718-50062019000400055](https://doi.org/10.4067/S0718-50062019000400055)
- La Red, D., Karanik, N., Giovannini, M., & Pinto, N. (2015). Perfiles de Rendimiento Académico: Un Modelo Basado en Minería de Datos. *Campus Virtuales*, 5(1), 12-30. <http://www.uajournals.com/ojs/index.php/campusvirtuales/article/view/66/65>
- Mancilla, G., Leal, P., Sánchez, A., & Vidal, C. (2020, Diciembre). Factores asociados al éxito de los estudiantes en modalidad de aprendizaje en línea: un análisis en minería de datos. *Formación universitaria*, 13(6), 23-35. <https://doi.org/10.4067/S0718-50062020000600023>
- Maya, P., Aguilar, J., Zamora, R., & Barron, M. (2018). Diseño de un Modelo predictivo aplicando minería de datos para identificar causas de deserción estudiantil universitaria. *Technology & Society*, 11-39. <http://ijsts.org/index.php/STS3/article/view/34/48>
- Menacho, C. (2017, Enero 11). Predicción del rendimiento académico aplicando técnicas de minería de datos. *Anales Científicos*, 78(1), 26-33. <https://doi.org/10.21704/ac.v78i1.811>
- Mhetre, V., & Nagar, M. (2017). Classification based data mining algorithms to predict slow, average and fast learners in educational system using WEKA. *International Conference on Computing Methodologies and Communication (ICCMC)*, (pp. 475-479). <https://doi.org/10.1109/ICCMC.2017.8282735>
- Moine, J., Gordillo, S., & Haedo, A. (2011). Análisis comparativo de metodologías para la gestión de proyectos de minería de datos. In U. N. (UNLP) (Ed.), *XVII Congreso Argentino de Ciencias de la Computación*, (pp. 931-938). <http://sedici.unlp.edu.ar/handle/10915/18749>
- Ramaswami, G., Susnjak, T., Mathrani, A., Lim, J., & Garcia, P. (2019). Using educational data mining techniques to increase the prediction accuracy of student academic performance. *Information and Learning Science*, 120(7-8). <https://doi.org/10.1108/ILS-03-2019-0017>

- Ramos-Jiménez, G., Campo-Avila, J., Morales-Bueno, R., & Baena-García, M. (2017). Minería de datos educativos para la predicción personalizada del rendimiento académico. *Convencion Internacional 2017 Universidad Central "Marta Abreu" de Las Villas*. <https://cutt.ly/aVWqdGk>
- Solís, D., Alegría, D., Gutiérrez, E., Zapata, V., Vidal, F., & Timarán, R. (2019). Identificación de Patrones de Rendimiento Académico en las Pruebas Saber Pro entre 2012-2014, en las Competencias Lectura Crítica y Comunicación Escrita con Técnicas Predictivas de Minería de Datos. *Cuaderno Activa*, 11, 51-64. <https://cutt.ly/2VWuf9f>
- Timarán, R. (2009). *Detección de Patrones de Bajo Rendimiento Académico y Deserción Estudiantil con Técnicas de Minería de Datos*. <https://cutt.ly/iVWtVbE>
- Timarán, R., Caicedo, J., & Hidalgo, A. (2019). Árboles de decisión para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas Saber 11°. *Revista de Investigación Desarrollo e Innovación*, 9(2), 363-378. <https://doi.org/10.19053/20278306.v9.n2.2019.9184>
- Urbina, A., Téllez, A., & Cruz, R. (2021). Patrones que identifican a estudiantes universitarios desertores aplicando minería de datos educativa. *Revista Electrónica de Investigación Educativa*, 23, 1-15. <https://doi.org/10.24320/redie.2021.23.e29.3918>
- Valero, S., Salvador, A., & García, M. (2022). *Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos*. http://fcaenlinea.unam.mx/anexos/1566/1566_u6_act1b.pdf