

Estudio Sistemático de Literatura sobre la Minería de Datos Educativos para determinar Factores que afectan el Rendimiento Escolar

Systematic Mapping Study of Literature on Educational Data Mining to Determine Factors That Affect School Performance

Viviana Párraga & Juan Zaldumbide

Escuela de Formación de Tecnólogos

Escuela Politécnica Nacional

Quito, Ecuador

{viviana.parragav, juan.zaldumbide} @epn.edu.ec

Resumen — Este documento expone los resultados de realizar un Systematic Mapping Study (SMS) sobre la minería de datos educativos para determinar factores que afectan el rendimiento escolar en el sistema de educación superior. Como resultado de este análisis se obtuvieron 20 estudios primarios, donde se observa, como, la minería de datos es aplicada para predecir el rendimiento escolar y así disminuir índices de deserción del alumnado. Mediante este trabajo se brinda una perspectiva más integral desde el proceso de captura de información y consolidación automática hasta proponer un modelo de predicción basándose en el análisis y uso de técnicas de minería de datos, contribuyendo con nuevos enfoques de análisis a los estudios actuales, tomando en cuenta las características de la gestión educativa.

En conclusión, los estudios consideran que, al utilizar minería de datos educativos aplicada a grandes volúmenes de datos existente en las instituciones de educación superior, permitiría mejorar la toma de decisiones para garantizar una educación de alta calidad, destacándose los problemas de rendimiento desde el principio y proponiendo acciones correctivas. Además, estos estudios muestran técnicas de minería de datos utilizadas, tales como: regresión logística, árboles de decisión, bosques aleatorios, Naive Bayes para explorar datos de entornos educativos.

Palabras Clave – minería de datos educativos, estudio sistemático de literatura, bases de datos digitales, factores rendimiento escolar.

Abstract — This document presents the results of conducting a Systematic Mapping Study (SMS) on the mining of educational data to determine factors that affect school performance in the higher education system. As a result of this analysis, 20 primary studies were obtained, where it is observed how data mining is applied to predict school performance and thus decrease the dropout rates of students. Through this work a more integral perspective is offered from the process of information capture and automatic consolidation until proposing a prediction model based on the analysis and use of data mining techniques, contributing

with new analytical approaches to the current studies, taking consider the characteristics of educational management.

In conclusion, the studies consider that, by using educational data mining applied to large volumes of data existing in higher education institutions, it would improve decision making to guarantee a high-quality education, highlighting the performance problems from the beginning and proposing corrective actions. In addition, these studies show data mining techniques used, such as logistic regression, decision trees, random forests, Naive Bayes to explore data from educational environments.

Keywords - educational data mining, Systematic Mapping Study, digital databases, school performance factors.

I. INTRODUCCIÓN

Uno de los mayores desafíos de la Educación Superior a nivel mundial es la integración, permanencia y egreso de los estudiantes de educación superior. “Según la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco), el abandono de ese nivel de formación llega a 40 % [1] debido a diversos factores, entre ellos el bajo rendimiento escolar.

Es así que estudiar esta problemática se constituye en interés primordial de profesores, autoridades educativas y estudiantes, puesto sería significativo poder determinar los múltiples factores que pueden influir en el abandono, con el propósito de gestionar y reducir este comportamiento.

Para contribuir con la solución al problema de fracaso escolar varios autores han planteado utilizar la minería de datos para encontrar patrones ocultos en grandes volúmenes de datos, dándole una aplicación educativa.

Por ejemplo, Guo, Zhang, Xu, Shi, & Yang en el año 2016, realizaron un estudio en el que predecían el rendimiento del

estudiante en su educación vocacional, en función de su motivación. La técnica de minería de datos se utilizó en el proceso de examen, donde, mediante un algoritmo de clasificación de Naive Bayes, basado en parámetros del nivel de motivación como entrada, obtuvieron variables más precisas. [2]

En otro artículo se define a la capacidad de minería de datos como una poderosa herramienta para el análisis académico, ya que se puede utilizar en el proceso de descubrir tendencias y patrones ocultos que ayudan a las instituciones a pronosticar el rendimiento de los estudiantes. Con base en los resultados extraídos mediante redes logísticas y neuronales se revela que factores como el ingreso económico familiar, la raíz étnica y las habilidades lingüísticas tienen una asociación significativa con el rendimiento académico. [3]

Mientras tanto el artículo “Analyzing undergraduate students’ performance using educational data mining”, utiliza métodos de minería de datos para estudiar el rendimiento de los estudiantes de pregrado, primero prediciendo el rendimiento académico de los estudiantes al final de un programa de estudio de cuatro años y luego estudiando las progresiones típicas y combinándolas con los resultados de predicción. Los resultados indican que al centrarse en un pequeño número de cursos que son indicadores de un rendimiento particularmente bueno o deficiente, es posible brindar una advertencia y apoyo oportunos a los estudiantes de bajo rendimiento, y consejos y oportunidades para los estudiantes de alto rendimiento. [4]

Otros autores, proponen el uso de técnicas de minería analizando datos históricos para diseñar un plan de acción de tutoría, donde se logró reducir la tasa de deserción en un 14% con respecto a los años académicos anteriores en los que no se aplicó ningún mecanismo de prevención de deserción escolar. [5]

Por lo antes expuesto, realizar un SMS sobre la minería de datos educativo para determinar factores que afectan el rendimiento escolar, permitirá conocer los diferentes puntos de vista expuestos por algunos autores a través de sus artículos, basados en investigaciones realizados por ellos sobre la problemática y las posibles soluciones a la misma, en este caso utilizando la minería de datos educativa.

II. METODOLOGÍA

El mapeo sistemático de literatura es un estudio de alcance que analiza un amplio conjunto de estudios primarios (artículos, publicaciones) para identificar qué y cuantas evidencias hay disponibles sobre un determinado tema.

Por lo general este tipo de estudio se desarrolla cuando: el tema que se quiere estudiar es muy amplio, cuando se requiere estudiar la cantidad de evidencias, cuando se necesita realizar un análisis superficial de estudios y cuando el objetivo es identificar la literatura más relevante de un tema de investigación para guiar la misma por el camino más adecuado.

La información generada se puede obtener respondiendo a preguntas de investigación sobre un tema específico, para después continuar con mapeos sistemáticos completos y rigurosos, siendo un punto de partida para la realización de futuras revisiones sistemáticas sobre las respuestas encontradas.

Un SMS está compuesto por pasos que permiten al investigador obtener estudios primarios pertinentes de alta calidad, luego de realizar el respectivo proceso de análisis.

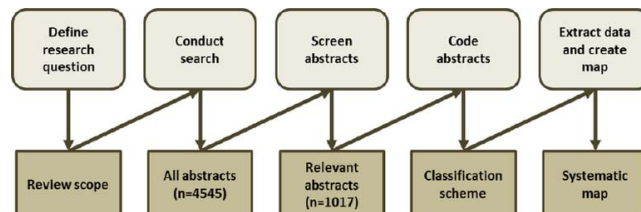


Figure 1. Proceso SMS [6]

A. Definición de Preguntas de Investigación

Se busca con este trabajo analizar la información sobre la minería de datos educativa para la identificación de patrones que influyen en el aprovechamiento académico y así conocer las causas generales de reprobación y repitencia de materias.

Las preguntas por resolver mediante este SMS son las siguientes:

RQ1: ¿Cómo se pueden determinar los factores que influyen en la deserción estudiantil en la educación superior?

RQ2: ¿Cómo se puede predecir el fracaso escolar y el abandono?

RQ3: ¿Qué decisiones se deben tomar y que acciones son pertinentes para obtener un mejor desempeño académico de los estudiantes?

RQ4: ¿Cuáles son los mejores modelos resultantes para dar solución al problema planteado?

RQ5: ¿Cuál es la metodología que mejor estructure el proceso de minería de datos utilizando los mejores modelos para analizar el comportamiento de los alumnos?

RQ6: ¿Cómo aplicar la inteligencia de Negocios para la toma de decisiones en el área administrativa, docente y psicopedagógica, y así evitar el rezago estudiantil y apoyar al alumnado?

B. Selección y clasificación de estudios

De acuerdo con las preguntas de investigación planteadas se procede a realizar:

1) *Revisión inicial*: Mediante una revisión inicial de literatura, se analizó a nivel del título, resumen, palabras clave y a groso modo el contenido de un conjunto de estudios que respondan a las preguntas de investigación planteadas, revisando si existen estudios relacionados.

2) *Integración del Grupo de Control*: El grupo de control conforman los estudios que cumplen con las características de la investigación analizado por los investigadores considerando el título del estudio, resumen y palabras claves. Con esto se facilita la ubicación de palabras claves que definan las cadenas de búsqueda que garanticen la cobertura de la literatura.

Los estudios del grupo de control analizados son los siguientes:

TABLE I. GRUPOS DE CONTROL

#	Grupos de control		
	<i>Estudio</i>	<i>Repositorio</i>	<i>Términos</i>
1	Predicting students achievement based on motivation in vocational school using data mining approach. [7]	IEEE XPORE	Data mining; educational institutions; human factors; pattern classification; academic achievement indicator; classification algorithm;
2	Modeling Academic Achievement of UUM Graduate Using Descriptive and Predictive Data Mining. [8]	Springer, Cham	Modeling Academic Achievement ; academic analysis; data Mining ; undergraduate students.
3	Application of learning analytics using clustering data Mining for Students' disposition analysis. [9]	Springer US	Learning analytics; Educational data mining; Disposition analytics; Academic analytics; Learning management systems
4	Data-driven intervention-level prediction modeling for academic performance. [10]	ACM New York	Predicting academic performance; prediction model; machine learning, technology in education; data mining
5	A machine learning framework to identify students at risk of adverse academic outcomes. [11]	ACM New York	evaluation metrics; applications; education; risk prediction
6	A survey on educational process mining. [12]	Scopus	Computer aided instruction; Data mining; Application area; Educational Applications; Educational data minings; (EDM); Educational environment; Educational process; Industry applications; Sequential-pattern mining; Visual representations
7	Predictive Modeling of Students Performance Through the Enhanced Decision Tree. [13]	Scopus	Classification (of information); Classifiers; Computer aided instruction; Data mining; Decision trees; StudentsTrees (mathematics)
8	Generating descriptive model for student dropout: a review of clustering approach. [14]	Scopus	Clustering; Dropout; Educational data mining; Retention Student performance
9	An Appraisal Model Based on a Synthetic Feature Selection Approach for Students' Academic Achievement. [15]	Scopus	Academic achievement; Data mining; Support vector machine; Synthetic feature selection approach (SFSA)

#	Grupos de control		
	<i>Estudio</i>	<i>Repositorio</i>	<i>Términos</i>
10	Analyzing undergraduate students' performance using educational data mining. [16]	Scopus	Data mining; Decision trees; Clustering; Performance prediction; Performance progression; Quality of educational processes
11	Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. [17]	Scopus	E-learning; Student dropout prediction; Educational data mining; Logistic regression model; Temporal data; Student dropout prevention; Tutoring action plan
12	A review on predicting student's performance using data mining techniques. [18]	Scopus	Student performance; educational data mining; performance prediction

3) *Construcción de la cadena de búsqueda:* Para la construcción de la cadena de búsqueda se analiza los estudios del grupo de control, encontrando palabras comunes entre estudios y las palabras propias que están direccionadas al objetivo de la presente investigación.

La cadena de búsqueda se forma con la combinación de las palabras que más se repiten en cada contexto para unir en la cadena se utilizó el conector OR y para concatenar con el contexto se usa el conector AND. Para la presente investigación se definieron 6 cadenas de búsquedas:

TABLE II. CADENAS DE BÚSQUEDA

CD#	<i>Cadenas de Búsqueda</i>
CD1	data AND mining OR decision AND trees AND academic AND achievement AND indicator
CD2	educational AND data AND mining AND student AND dropout AND prediction AND educational AND institutions
CD3	data AND mining AND academic AND analysis AND undergraduate AND students OR retention AND student AND performance
CD4	predicting AND academic AND performance OR quality AND of AND educational AND processes AND educational AND data AND mining
CD5	education OR educational AND applications AND data AND mining AND student AND dropout AND prediction
CD6	pattern AND classification OR evaluation AND metrics AND learning AND management AND systems AND prediction AND model

Cada cadena de búsqueda se analizó en el repositorio SCOPUS, donde se obtuvo los siguientes resultados:

TABLE III. ESTUDIOS DE CONTROL POR CADENA DE BÚSQUEDA

CD#	Estudios	Estudios de Grupo de Control
CD1	310	GC1, GC2, GC4, GC5, GC6, GC7, GC9, GC10
CD2	96	GC3, GC8, GC11, GC12
CD3	771	GC1, GC2, GC3, GC4, GC5, GC6, GC8, GC9, GC10.
CD4	844	GC3, GC4, GC8, GC10, GC11, GC12
CD5	295	GC1, GC2, GC4, GC5, GC6, GC9, GC10, GC11
CD6	4872	GC1, GC3, GC4, GC5.

De acuerdo con la tabla anterior se escoge como cadena de búsqueda definitiva la CD3, puesto que en ella se agrupan un mayor número de grupos de control, estableciendo la siguiente cadena de búsqueda:

((Data mining) AND (academic analysis) AND ((undergraduate students) OR (retention student performance)))

4) *Validación Cruzada de Estudios*: La validación cruzada permite garantizar que los estudios cumplan con los criterios de inclusión y exclusión.

Con base en el proyecto planteado se puede considerar como criterios de inclusión a los:

- Artículos sobre la minería de datos orientados a la predicción del fracaso escolar.
- Artículos sobre predicción de deserción de estudiantes.
- Artículos sobre aplicación de minería de datos en análisis académico.
- Artículos sobre clasificación de patrones basado en indicadores de logro académico.
- Artículos sobre modelos de predicción de rendimiento académico.

Además, se puede considerar como criterios de exclusión a los:

- Artículos sobre análisis de factores humanos en la repetición académica.
- Artículos sobre análisis predictivo mediante algoritmos matemáticos.
- Artículos sobre aprendizaje de máquina.
- Artículos de minería de datos que no estén orientados a la predicción de fracaso escolar.
- Artículos que hablen sobre big data.
- Artículos sobre el análisis del rendimiento en el estudio de minería de datos.
- Artículos acerca de técnica de enseñanza para mejorar aprendizaje.
- Artículos sobre técnica de aprendizaje y enseñanza.

La cadena de búsqueda fue aplicada en el repositorio digital SCOPUS, realizando configuraciones como: idioma, fuentes de

publicaciones fiables, año de publicación (2016-2018), estudios duplicados, área de interés, tipo de publicación (revistas y congresos), entre otros verificando coincidencias y discrepancias respecto a los criterios de inclusión y exclusión para determinar los estudios candidatos depurados en diferentes repositorios digitales.

De acuerdo con la figura 2. se puede observar que de 440 estudios CANDIDATOS(C), luego de la depuración se consiguió encontrar 83 registros CANDIDATOS DEPURADOS (CD).

Figure 2. Estudios Candidatos Depurados

A continuación, se procede a seleccionar los estudios a partir de una revisión de títulos y resúmenes de los estudios candidatos depurados, verificando coincidencias y discrepancias respecto a los criterios de inclusión y exclusión.

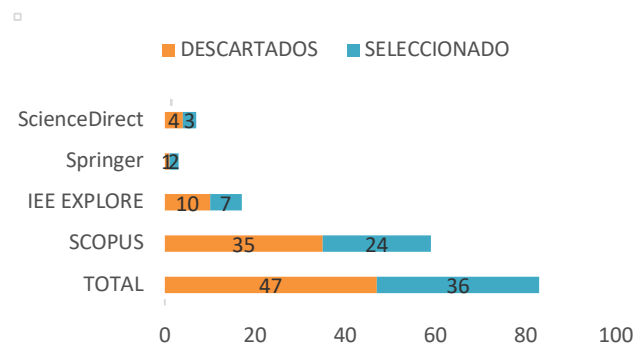


Figure 3. Estudios Seleccionados

De acuerdo con la figura 3, se puede observar que de 83 estudios CANDIDATOS DEPURADOS(CD), luego de la selección se consiguió encontrar 36 registros SELECCIONADOS (S).

Para finalizar el SMS se realiza una depuración de los estudios seleccionados y así encontrar los estudios PRIMARIOS, esto se lo obtiene descargando los artículos y analizándolos, donde se obtuvieron los resultados indicados en la tabla 4:

TABLE IV. ANÁLISIS DE ESTUDIOS SELECCIONADOS

Fuente	Descartados	Primarios	Seleccionados
SCOPUS	11	13	24

Fuente	Descartados	Primarios	Seleccionados
IEEE XPLORE	4	3	7
SPRINGER	1	1	7
SCIENCE DIRECT	0	3	3
TOTAL	16	20	36

De acuerdo con los datos obtenidos se puede encontrar 20 ESTUDIOS PRIMARIOS de los 440 estudios con los que se comenzó el análisis. Resumiendo todo el análisis la figura 4:

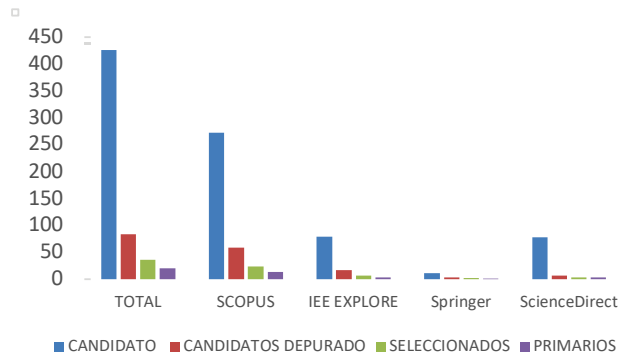


Figure 4. Resumen Estadístico de Estudios

Que permite visualizar el análisis en cada una de las bases digitales tal y como se muestra en la figura 5.

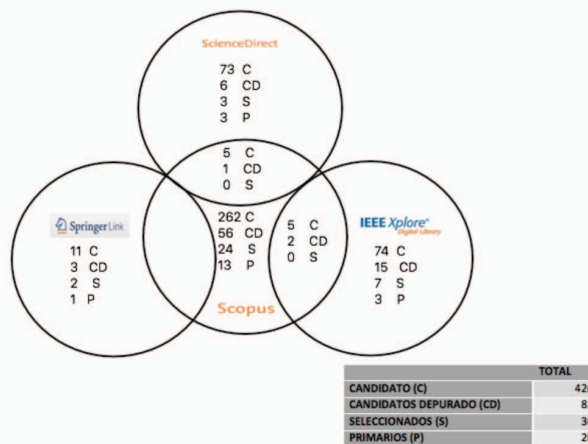


Figure 5. Resumen Final del Análisis de los Estudios

5) Extracción y Mapeo de Datos

La información considerada de los estudios primarios es la más relevante, y consecuente con los objetivos de investigación.

Para obtener dicha la información se plantea un proceso estructurado de extracción de información, que incluye: extracción de las características de los estudios primarios y extracción de la información específica objeto de la SMS.

Los estudios primarios recopilados mediante este análisis son:

- Generating descriptive model for student dropout: a review of clustering approach [19]
- Analyzing undergraduate students' performance using educational data mining[4]
- Improved student dropout prediction in Thai University using ensemble of mixed-type data clusterings [20]
- Depth and breadth of educational data mining: Researchers' point of view [21]
- Prediction of students performance using Educational Data Mining [22]
- Using data mining techniques to predict students at risk of poor performance [23]
- Using data mining techniques with open source software to evaluate the various factors affecting academic performance: A case study of students in the faculty of information technology [24]
- Predicting Students Performance in Educational Data Mining [2]
- Modeling academic achievement of UUM graduate using descriptive and predictive data mining [3]
- Predicting the risk of attrition for undergraduate students with time based modelling [25]
- Mining students' record to predict their performance in undergraduate degree[26]
- Academic Performance Predictors [27]
- Using cluster ensemble to improve classification of student dropout in Thai university [28]
- Employing Markov Networks on Curriculum Graphs to Predict Student Performance [29]
- Analyzing students records to identify patterns of students' performance [30]
- Knowledge discovery from academic data using Association Rule Mining [31]
- Educational Data Mining for Analysis of Studentsâ€™ Solutions [32]
- Predicting Critical Courses Affecting Students Performance: A Case Study [33]
- Predicting Students' Performance in University Courses: A Case Study and Tool in KSU Mathematics Department[34]
- Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector [35]

A continuación, se procede a revisar la información de los estudios primarios. Considerando la información de estos estudios como la más relevante y consecuente con los objetivos de investigación. Para obtener dicha información se plantea un proceso estructurado de extracción de información que incluye:

- Extracción de las características de los estudios primarios

- Extracción de la información específica objeto de la SMS

De forma general los estudios primarios responden a la disciplina de Computers & Education y Minería de Datos, y se considera un tema actual de estudio, debido a que los estudios seleccionados corresponden a los últimos años.

III. RESULTADOS

Luego de realizar la lectura de los estudios primarios, se indica que, un análisis de minería de datos se puede utilizar para destacar los problemas de rendimiento escolar desde el principio y proponer acciones correctivas[23]. Los autores Altujjar, Altamimi, Al-Turaiki y Al-Razgan describen que la predicción del rendimiento académico de los estudiantes es una de las aplicaciones más importantes de la minería de datos educativos.[33]

Los resultados de una minería de datos educativo, indican que al centrarse en un pequeño número de cursos que son indicadores de un rendimiento particularmente bueno o malo, es posible brindar una advertencia y apoyo oportunos a los estudiantes de bajo rendimiento, y consejos y oportunidades para los estudiantes de alto rendimiento.[4] [32]

También se hace hincapié en que los registros académicos históricos de los estudiantes pueden analizarse para explorar los factores detrás de fenómenos, para predecir el rendimiento de los estudiantes mediante un modelo de minería de datos. [34] [26]

Chai y Gibson probaron tres técnicas supervisadas de aprendizaje automático para desarrollar los modelos predictivos: regresión logística, árboles de decisión y bosques aleatorios en su trabajo de análisis de minería de datos educativos.[25] Por otro lado (Kaur, Singh, & Josan, 2015), utilizaron registros académicos de los estudiantes en algoritmos de clasificación como Perception multicapa, Naïve Bayes, SMO, J48 y REPTree utilizando WEKA como herramienta de código abierto. Como resultado, las estadísticas se generan en base a todos los algoritmos de clasificación y la comparación de los cinco clasificadores para predecir la precisión y para encontrar el algoritmo de clasificación de mejor rendimiento entre todos. [35] [23]

Mientras que Devasia, Vinushree y Hegde propusieron una aplicación basada en la web que hace uso de la técnica de minería Naive Bayesian para la extracción de información útil. [22] Y otros realizaron el análisis utilizando técnicas de aprendizaje automático y minería de datos para explorar datos de entornos educativos.[2]

Los datos demográficos de los estudiantes y el desempeño académico se utilizan para estudiar el patrón académico. Se recomienda utilizar una metodología como CRISP-DM, que consisten en la comprensión de datos y la preparación de estos. La herramienta de modelado y minería de datos utilizada identifica la correlación más significativa de las variables asociadas con el éxito académico basada en los últimos años de los datos demográficos y de rendimiento de los estudiantes, por ejemplo. [30]

Con esta idea, se han desarrollado varios métodos de minería de datos para la detección temprana de estudiantes en riesgo de deserción, de ahí la aplicación inmediata de medidas de asistencia. El estudio empírico sobre la recopilación de datos de tipo mixto relacionado con los detalles demográficos de los estudiantes, el rendimiento académico y el registro de inscripciones, sugiere que el enfoque propuesto suele ser más eficaz que varias técnicas de transformación de referencia, en diferentes clasificadores. [20] [28]

Es así, como la capacidad de predecir el rendimiento académico es una gran ventaja tanto para los estudiantes como para los administradores de la institución. Para los estudiantes, pueden ajustar la carga de trabajo, la dirección de la carrera, etc., si conocen su capacidad. Para los administradores e instructores, las advertencias tempranas facilitarían la intervención, permitiendo así un ambiente académico más exitoso. Además, los recursos institucionales se pueden utilizar de manera óptima, obteniendo así la eficiencia de la operación.[27]

IV. CONCLUSIONES

La minería de datos en el análisis del ámbito educativo es un tema muy extenso y cubierto por algunos artículos como lo revela este estudio, se puede observar como la minería de datos es aplicada principalmente para predecir el rendimiento escolar de los estudiantes y así disminuir índices de deserción de éstos. Al descubrir el conocimiento oculto de un gran volumen de datos educativos y aplicarlo adecuadamente para la toma de decisiones se puede garantizar una educación de alta calidad en cualquier institución académica.

El uso de minería de datos, puede contribuir de una manera estratégica en la planificación del área administrativa, docente y psicopedagógica, evitando de esta forma el bajo rendimiento académico y con esto el rezago estudiantil, apoyando en todo momento al alumnado con acciones pertinentes.

Existen diferentes técnicas de minería de datos, tales como: regresión logística, árboles de decisión, bosques aleatorios, Naive Bayesian para explorar datos de entornos educativos, desde diferentes puntos de vista.

En consecuencia, el presente estudio contribuirá con generación de conocimiento sobre la problemática planteada en la identificación de factores que influyen en el aprovechamiento académico de los estudiantes que de soporte a la toma de decisiones, evitando el bajo rendimiento académico y con esto el rezago estudiantil, mediante una perspectiva más integral desde el proceso de captura de información, consolidación automática y así proponer un modelo de pronóstico basándose en el análisis y uso de técnicas de minería de datos contribuyendo con nuevos enfoques de análisis a los estudios actuales, considerando las características de la gestión educativa.

Un Estudio Sistemático de Literatura permite simplificar el estudio de un tema determinados, debido a que presenta de manera resumida los artículos o trabajos ya desarrollados y considerados ya importante, reduciendo tiempos de investigación.

Como trabajo futuro se plantea, realizar un artículo sobre índices de retención alcanzados aplicando minería de datos educativos, aplicación de estas técnicas en gestión educativa

como contratación docente, estrategias de gestión de autoridades de instituciones de educación superior, procesos de acreditación entre otras.

REFERENCIAS

- [1] E. Telégrafo (2016, nov 10), La deserción universitaria bordea el 40%, [Online]. Available: <http://www.eltelegrafo.com.ec/noticias/sociedad/4/la-desercion-universitaria-bordea-el-40>.
- [2] B. Guo, R. Zhang, G. Xu, C. Shi, y L. Yang, "Predicting Students Performance in Educational Data Mining" in *Proceedings International Symposium on Educational Technology*, ISET, 2015, pp. 125-128.
- [3] F. Siraj, "Modeling Academic Achievement of UUM Graduate Using Descriptive and Predictive Data Mining", *Springer*, pp. 609-620, January 2016.
- [4] R. Asif, A. Merceron, S. A. Ali, y N. G. Haider, "Analyzing undergraduate students' performance using educational data mining", *Comput. Educ.*, vol. 113, pp. 177-194, October 2017.
- [5] C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano, y M. A. Martínez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout", *Comput. Electr. Eng.*, March 2017.
- [6] K. Petersen, R. Feldt, S. Mujtaba, y M. Mattsson, "Systematic Mapping Studies in Software Engineering" in *EASE'08 Proceedings*, Italy, of the 12th international conference on Evaluation and Assessment in Software Engineering, 2008, pp. 68-77.
- [7] J. N. Purwaningsih y Y. Suwarno, "Predicting students achievement based on motivation in vocational school using data mining approach", in *2016 4th International Conference on Information and Communication Technology (IColCT)*, 2016, pp. 1-5.
- [8] F. Siraj, "Modeling Academic Achievement of UUM Graduate Using Descriptive and Predictive Data Mining", in *Advanced Computer and Communication Engineering Technology*, Springer, Cham, 2016, pp. 609-620.
- [9] S. Bharara, S. Sabitha, y A. Bansal, "Application of learning analytics using clustering data Mining for Students' disposition analysis", *Educ Inf Technol*, pp. 1-28, oct. 2017.
- [10] M. Mgala y A. Mbogho, "Data-driven Intervention-level Prediction Modeling for Academic Performance", in *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development*, New York, NY, USA, 2015, pp. 2:1-2:8.
- [11] H. Lakkaraju et al., "A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes", in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2015, pp. 1909-1918.
- [12] A. Bogarín, R. Cerezo, y C. Romero, "A survey on educational process mining", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 1, 2018.
- [13] S. Sivakumar y R. Selvaraj, "Predictive modeling of students performance through the enhanced decision tree", *Lecture Notes in Electrical Engineering*, vol. 443, pp. 21-36, 2018.
- [14] N. Iam-On y T. Boongoen, "Generating descriptive model for student dropout: a review of clustering approach", *Human-centric Computing and Information Sciences*, vol. 7, no. 1, 2017.
- [15] C.-H. Cheng y W.-X. Liu, "An appraisal model based on a synthetic feature selection approach for students' academic achievement", *Symmetry*, vol. 9, no. 11, 2017.
- [16] R. Asif, A. Merceron, S. A. Ali, y N. G. Haider, "Analyzing undergraduate students' performance using educational data mining", *Computers & Education*, vol. 113, pp. 177-194, oct. 2017.
- [17] C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano, y M. A. Martínez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout", *Computers & Electrical Engineering*, mar. 2017.
- [18] A. M. Shahiri, W. Husain, y N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques", *Procedia Computer Science*, vol. 72, pp. 414-422, ene. 2015.
- [19] N. Iam-On y T. Boongoen, "Generating descriptive model for student dropout: a review of clustering approach", *Hum.-Centric Comput. Inf. Sci.*, vol. 7, no. 1, December 2017.
- [20] N. Iam-On y T. Boongoen, "Improved student dropout prediction in Thai University using ensemble of mixed-type data clusterings", *Int. J. Mach. Learn. Cybern.*, vol. 8, no. 2, pp. 497-510, April 2017.
- [21] A. A. Mehta y N. J. Buch, "Depth and breadth of educational data mining: Researchers' point of view" in *Proceedings of the 10th International Conference on Intelligent Systems and Control*, India, ISCO 2016, 2016.
- [22] T. Devasia, T. P. Vinushree, y V. Hegde, "Prediction of students performance using Educational Data Mining", in *Proceedings of 2016 International Conference on Data Mining and Advanced Computing*, India, SAPIENCE 2016, 2016, pp. 91-95.
- [23] Z. Alharbi, J. Cornford, L. Dolder, y L. I. De, "Using data mining techniques to predict students at risk of poor performance", in *Proceedings of 2016 SAI Computing Conference*, London, SAI 2016, 2016, pp. 523-531.
- [24] F. Hanandeh, M. Y. Al-Shannag, y M. M. Alkhaffaf, "Using data mining techniques with open source software to evaluate the various factors affecting academic performance: A case study of students in the faculty of information technology", *Int. J. Open Source Softw. Process.*, vol. 7, no. 2, pp. 72-92, January 2016.
- [25] K. E. K. Chai y D. Gibson, "Predicting the risk of attrition for undergraduate students with time based modelling", in *Proceedings of the 12th International Conference on Cognition and Exploratory Learning in the Digital Age*, Dublin, CELDA 2015, 2015, pp. 109-116.
- [26] L. Ramanathan, A. Geetha, y M. Khalid, "Mining students' record to predict their performance in undergraduate degree", *Int. J. Appl. Eng. Res.*, vol. 10, no. 1, pp. 1113-1142, 2015.
- [27] C. Lei y K. F. Li, "Academic Performance Predictors", in *Proceedings - IEEE 29th International Conference on Advanced Information Networking and Applications Workshops*, South Korea, WAINA 2015, 2015, pp. 577-581.
- [28] N. Iam-On y T. Boongoen, "Using cluster ensemble to improve classification of student dropout in Thai university", in *2014 Joint 7th International Conference on Soft Computing and Intelligent Systems, SCIS 2014 and 15th International Symposium on Advanced Intelligent Systems*, Japan, ISIS 2014, 2014, pp. 452-457.
- [29] A. Slim, G. L. Heileman, J. Kozlick, y C. T. Abdallah, "Employing markov networks on curriculum graphs to predict student performance", in *Proceedings - 2014 13th International Conference on Machine Learning and Applications*, USA, ICMLA 2014, 2014, pp. 415-418.
- [30] A. C. K. Hoe et al., "Analyzing students records to identify patterns of students' performance", in *International Conference on Research and Innovation in Information Systems*, Malaysia, ICRIS, 2013, pp. 544-547.
- [31] S. Ahmed, R. Paul, y A. S. M. L. Hoque, "Knowledge discovery from academic data using association rule mining", in *2014 17th International Conference on Computer and Information Technology*, Bangladesh, ICCIT 2014, 2003, pp. 314-319.
- [32] K. Vaculík, L. Nezvalova, y L. Popelínský, "Educational data mining for analysis of students' solutions", *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.*, vol. 8722, pp. 150-161, November 2014.
- [33] Y. Altujjar, W. Altamimi, I. Al-Turaiki, y M. Al-Razgan, "Predicting Critical Courses Affecting Students Performance: A Case Study", *Procedia Computer Science*, vol. 82, pp. 65-71, January 2016.
- [34] G. Badr, A. Algobail, H. Almutairi, y M. Almutery, "Predicting Students' Performance in University Courses: A Case Study and Tool in KSU Mathematics Department", *Procedia Computer Science*, vol. 82, pp. 80-89, January 2016.
- [35] P. Kaur, M. Singh, y G. S. Josan, "Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector", *Procedia Computer Science*, vol. 57, pp. 500-508, 2015.