

UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN



**REVISIÓN SISTEMÁTICA SOBRE LA PREDICCIÓN DEL RENDIMIENTO
ACADÉMICO EN ESTUDIANTES: TÉCNICAS Y ALGORITMOS**

**TRABAJO DE INVESTIGACIÓN PARA OPTAR EL GRADO ACADÉMICO DE
BACHILLER EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**

AUTOR

LUIS JESUS GARCIA PEREDO

ASESOR

JURY YESENIA AQUINO TRUJILLO

<https://orcid.org/0000-0003-1662-6406>

Chiclayo, 2020

Índice

1. Introducción	3
2. Metodología	4
2.1. Formulación de las preguntas de las preguntas de investigación.....	4
2.2. Definición de la estrategia de búsqueda.....	4
2.2.1. Términos de búsqueda.....	4
2.2.2. Recursos de literatura	5
2.2.3. Proceso de búsqueda.....	5
2.3. Selección de trabajos.....	5
2.3.1. Criterios de exclusión/inclusión	5
2.3.2. Criterios de calidad.....	6
3. Resultados y discusión	6
4. Conclusiones	8
5. Agradecimientos	8
6. Referencias	8

Revisión sistemática sobre la predicción del rendimiento académico en estudiantes: técnicas y algoritmos

A systematic review about the prediction of academic performance in students: techniques and algorithms

Luis Jesús García Peredo

Universidad Católica Santo Toribio de Mogrovejo, Av. San Josemaría Escrivá de Balaguer 855
73772317@usat.edu.pe

El objetivo de la siguiente investigación es realizar un análisis en base a la revisión de literatura sobre la predicción del rendimiento académico del estudiante mediante uso de técnicas y/o algoritmos. Se estableció una revisión sistemática con publicaciones de los últimos cinco años, las cuales están alojadas en 3 bases de datos digitales seleccionadas: ProQuest, ScienceDirect y IOPscience, en todas se aplicaron filtros de búsqueda en base a criterios de inclusión/exclusión y calidad, finalmente se obtuvieron 13 artículos con los que se responden a las preguntas planteadas. Durante el desarrollo de esta investigación se obtuvo un listado de los continentes y países que realizan más investigaciones sobre la temática abarcada, gráficos que reflejan la utilización de técnicas/algoritmos para predecir el rendimiento académico e información gráfica y tabular sobre las características/factores que se toman en cuenta para lograr la predicción del rendimiento académico. Al obtener los resultados se pudo descubrir que el continente más investigador con respecto a esta temática es Asia, siendo los países que más aportan India e Indonesia, además se halló que las técnicas/algoritmos más usados para este tipo de predicción son Naive Bayes, Decision Tree, Support Vector Machine y Neural Networks, por último, se definieron 5 factores que se utilizan mayormente para obtener predicciones sobre el rendimiento académico. Este artículo sirve para llenar un poco el vacío de información que hay en el Perú con respecto al tema y también como guía para investigaciones futuras que abarquen temáticas similares.

Palabras claves: predicción, rendimiento académico, estudiante, algoritmo, factores, revisión sistemática de literatura

The objective of the following research is to carry out an analysis based on the literature review on the prediction of student academic performance through the use of techniques and / or algorithms. A systematic review was established with publications from the last five years, which are housed in 3 selected digital databases: ProQuest, ScienceDirect and IOPscience, in all search filters were applied based on inclusion / exclusion criteria and quality, finally 13 articles were obtained with which the questions posed were answered. During the development of this research, a list was obtained of the continents and countries that carry out more research on the subject covered, graphics that reflect the use of techniques / algorithms to predict academic performance and graphic and tabular information on the characteristics / factors that are taken into account to achieve the prediction of academic performance. When obtaining the results, it was discovered that the most investigative continent with respect to this issue is Asia, with India and Indonesia being the countries that contribute the most, it was also found that the most used techniques / algorithms for this type of prediction are Naive Bayes, Decision Tree, Support Vector Machine and Neural Networks, finally, defined 5 factors that are used mostly to obtain predictions about academic performance. This article serves to fill a bit of the information gap that exists in Peru regarding the subject and also as a guide for future research covering similar topics.

Keywords: prediction, academic performance, student, algorithm, factors, systematic literature review

1. Introducción

La definición más global que se tiene del rendimiento académico, es la que lo establece como una variable esencial que

hace posible la evolución del Sistema Educativo, pero la definición que más se relaciona con esta investigación es la que establece Requena [1] explicando el rendimiento académico como una consecuencia de la relación entre la capacidad del estudiante y su desempeño. Aunque esta relación parezca algo

fácil de comprender, existe una serie de factores que influyen de manera directa, Oliver [2] es muy tajante al afirmar que mayormente se piensa que el rendimiento académico del estudiante varía de acuerdo a su organización con respecto a sus responsabilidades o a sus hábitos de estudio, pero no se evalúa más allá, no se toman en cuenta otros factores que también son muy influyentes en la obtención de sus resultados. Ahí es donde se da pie a que se genere un problema, las instituciones saben que cierta parte de sus estudiantes presentan un rendimiento académico irregular y los tratan de “ayudar” mediante reforzamiento académico, pero no da resultado, esto nos confirma que el punto clave sigue siendo que hay otros factores que no se están tomando en cuenta para saber cómo apoyar al estudiante. Estos factores que a simple vista no se pueden identificar ni analizar, pero con el uso de ciertas técnicas sí es posible, el descubrimiento de estos factores se da en los datos del estudiante que muchas veces la misma institución recopila y almacena. Datos que muchas veces se almacenan rutinariamente y no se aprovechan de manera correcta debido al desconocimiento, complejidad o volumen, por tanto, analizar esa data es de gran interés para hallar nuevas propuestas de mejora, en este caso predecir el rendimiento de los estudiantes es una de ellas. Dentro de este contexto entra a tallar la Educational Data Mining (EDM), la cual desarrolla y adapta métodos estadísticos, de aprendizaje automático y de minería de datos para estudiar datos pertenecientes al área educativa, formando parte de este grupo las técnicas/algoritmos. Además, su aplicación puede ayudar en el análisis de los procesos de aprendizaje de los estudiantes, considerando también su interacción con factores internos (ser humano) y factores externos (medio ambiente) [3].

Los resultados de obtener una predicción del rendimiento académico ayudan a identificar la influencia de factores en el desenvolvimiento del estudiante, en base a eso luego se pueden tomar decisiones o acciones de ayuda más concretas. De acuerdo a lo anteriormente mencionado, se decidió realizar una revisión de literatura enfocada a la predicción del rendimiento académico de estudiantes mediante el uso de técnicas y/o algoritmos. Durante la revisión nos centramos en los algoritmos que son más utilizados para realizar predicciones, las características o factores relacionados al estudiante que se toman en cuenta para lograr la predicción y los continentes/países de donde proceden las investigaciones. Incluso mediante este artículo también se busca sumar conocimiento respecto a la temática abarcada, aportar con la muestra de soluciones modernas que permitan aumentar el interés y el deseo de implementar este tipo de soluciones aquí en Perú, que a largo plazo no solo mejorarían nuestras capacidades, sino que también reforzarían la educación, la cual es uno de los pilares más importantes en la formación de una persona.

2. Metodología

Para el desarrollo de esta investigación se decidió realizar una revisión sistemática de literatura (SLC) de artículos científicos enfocados en el tema de esta investigación. La estructura de pasos a seguir en la SLC son la que utilizaron los autores [4] y [5]: formulación de las preguntas de investigación, definición de la estrategia de búsqueda, selección de trabajos aplicando criterios de exclusión/inclusión y criterios de calidad, finalmente se explican los resultados. Los pasos de la estructura se plantearon de la siguiente forma: sección 2.1 (formulación de las preguntas de investigación), sección 2.2 (definición de estrategia de

búsqueda), sección 2.3 (selección de trabajos) y sección 3 (resultados).

2.1. Formulación de las preguntas de las preguntas de investigación

En este punto se formularon las preguntas/incógnitas de la investigación (PI - x) que servirán como guía y orientación para realizar la búsqueda y selección de investigaciones, además estas preguntas serán respondidas durante el desarrollo de este artículo.

Para la formulación de estas se ha tomado como guía el modelo redactado en [6]. Ubicándonos en el marco de tiempo 2015 – 2020, se plantean las siguientes preguntas:

PI – 1: ¿En qué continentes y países se realizan más investigaciones sobre la predicción del rendimiento académico del estudiante?

PI – 2: ¿Qué tipos de algoritmos se utilizan para predecir el rendimiento académico de un estudiante?

PI – 3: ¿Cuáles son los factores o características que se toman en cuenta para predecir el rendimiento académico de un estudiante?

2.2. Definición de la estrategia de búsqueda

En base al modelo de los autores [7], la estructura de la estrategia de búsqueda seleccionada para esta investigación consta de 3 partes: términos de búsqueda, recursos de literatura y proceso de búsqueda. Cada parte fue desarrollada y presentada de manera concisa como se muestra a continuación:

2.2.1. Términos de búsqueda

Antes de iniciar con el proceso de búsqueda se definieron las palabras claves que se encargan de cubrir el objetivo principal de esta investigación, siendo: academic performance, prediction y algorithm.

Pese a que se establecen las palabras claves, los resultados de la búsqueda no solo contienen la terminología precisa, sino que también contienen sinónimos y términos relacionados. Debido a eso se decidió hacer uso de la opción de “búsqueda avanzada” que brindan las 3 bases de datos digitales seleccionadas, pero solo una contiene la opción de campos adicionales (SF), se utilizaron 3 campos: el primero con términos “academic performance”, el segundo con “prediction” y el tercero con el término “algorithm”.

Además, esta base de datos también permitió el uso de conectores en la búsqueda (AND, OR, NOT). El establecimiento de estos términos de búsqueda se encuentra detallados en la Tabla 1.

Tabla 1: Términos en campos de búsqueda de las bases de datos

DataBases	Campos de búsqueda (SF)			Conector de enlace
	Campo 1	Campo 2	Campo 3	
ProQuest (DB1)	Academic performance	prediction	algorithm	AND

ScienceDirect (DB2)	Academic performance prediction	No disponible	No disponible	No disponible
IOPscience (DB3)	Academic performance prediction	No disponible	No disponible	No disponible

2.2.2. Recursos de literatura

Para la búsqueda de artículos enfocados al tema tratado se decidió utilizar bases de datos digitales con acceso universitario, estas son: ProQuest (DB1), ScienceDirect (DB2) y IOPscience (DB3); además de brindar disponibilidad de acceso, estas bases de datos fueron escogidas debido al buen número de resultados obtenidos en base a la terminología de búsqueda.

2.2.3. Proceso de búsqueda

Luego de definir los campos de búsqueda, se procede a ejecutar la búsqueda en cada una de las bases de datos. Teniendo como tema principal “predicción del rendimiento académico”, en la primera base de datos se decidió que el campo 1 (SF1) y el campo 2 (SF2) deberían ser ubicados en el título de las investigaciones a encontrar, mientras que el campo 3 (SF3) no debería estar necesariamente en el título sino en cualquier parte.

Por otro lado, las 2 bases de datos restantes no cuentan con la opción de ubicación de búsqueda del campo, así que filtran de manera automática las coincidencias en los títulos.

Esta búsqueda se realizó el 29 de noviembre del 2020 y el número de resultados obtenidos se muestran en la Tabla 2 (la cantidad de resultados pueden mostrar variabilidad de acuerdo a las actualizaciones de las bases de datos digitales), cabe recalcar que a estos resultados posteriormente se le aplican criterios de exclusión/inclusión y calidad para finalmente obtener los artículos que se utilizarán en la investigación.

Tabla 2: Resultados obtenidos en la primera búsqueda

DataBase	Ubicación búsqueda campo	Resultados de búsqueda
ProQuest (DB1)	SF1: Título de investigación	17
	SF2: Título de investigación	
	SF3: Cualquier parte	
ScienceDirect (DB2)	-ubicación única-	99547
	SF1: Título de investigación	
IOPscience (DB3)	-ubicación única-	500
	SF1: Título de investigación	

El total de artículos hallados en la primera búsqueda fue de 100.064, todos estos artículos tomaron la característica de “relación con el tema” y pueden ser utilizarlos en la investigación.

2.3. Selección de trabajos

Para poder delimitar y refinar el número de coincidencias en la búsqueda, fue indispensable el establecimiento de criterios de exclusión/inclusión y criterios de calidad que se presentan en las siguientes subsecciones.

2.3.1. Criterios de exclusión/inclusión

Las reglas de inclusión que se tomaron en consideración para los primeros resultados obtenidos fueron: Tener en cuenta solo las investigaciones de tipo “Artículo Científico”. Para trabajar con investigaciones de países desarrollados se aceptaron los idiomas de inglés y español, además para no tener problemas con la antigüedad de las investigaciones, se estableció un rango de cinco años (2015-2020). Luego para tener un acceso total a la información encontrada, se seleccionaron solo las investigaciones de tipo “acceso abierto” (texto completo). Por último, se consideraron los resultados que estén relacionados con el área de Computer Science, esta opción está disponible solo en la DB2.

Por otro lado, las reglas de exclusión están basadas en el incumplimiento de los criterios de inclusión, descartando así los resultados que las presenten, asimismo para evitar la redundancia se decidió no tomar en cuenta los artículos duplicados.

Ambos tipos criterios fueron aplicados en las 3 bases de datos digitales, según su funcionalidad, se puede visualizar mejor en la Tabla 3.

Tabla 3: Establecimiento de criterios exclusión/inclusión

	Criterios de inclusión	Criterios de exclusión
Ofrecido por Digital DataBase	<ul style="list-style-type: none"> Año de publicación: 2015 – 2020 Tipo documento: Artículo Idioma: Inglés y Español Acceso: Abierto 	<ul style="list-style-type: none"> Año de publicación: fuera de rango especificado. Tipo documento: Diferente a tipo artículo. Idioma: diferente a Inglés y Español. Acceso: Cerrado Artículos duplicados.
Lectura	Área: Computer Science	Área: Diferente a Computer Science

Es importante dejar en claro que la aplicación de los criterios de exclusión/inclusión varía de acuerdo a la base de datos: La DB2 (ScienceDirect) requiere la selección de checkbox`s para

especificación de fechas, marcando así cada año de manera manual, mientras que las otras bases de datos permitían seleccionar un solo checkbox de “Últimos 5 años”. De la misma forma la DB1 no presenta como tal la opción de acceso abierto, pero la establece como checkbox “texto completo” la cual fue seleccionada. Otro detalle es el que presenta la DB2, la cual no cuenta con opción de acceso abierto, debido a eso el número de resultados obtenidos luego de aplicar los demás criterios de exclusión/inclusión pasarán al punto de filtrado manual para seleccionar solo los de acceso abierto.

Los criterios de exclusión/inclusión fueron aplicados de manera consecutiva (uno sobre otro) disminuyendo así el número de resultados en cada aplicación. Luego del filtrado la DB1 (ProQuest) obtuvo 8 artículos, la DB2 (ScienceDirect) obtuvo 3748 artículos y la DB3 (IOPscience) obtuvo 100 artículos, quedándonos con un total de 3856 artículos como se aprecia en la Tabla 4. A esa cantidad total de artículos obtenidos se le pasó por un filtrado manual, que consistió en leer el abstract y las conclusiones para asegurar el desarrollo del tema abarcado (aplicación de técnicas y/o algoritmos para la predicción del rendimiento académico), si la investigación guardaba relación con la terminología pero no se trabajaron los puntos que permiten dar respuesta a las preguntas planteadas en este artículo, quedaba descartada. También se descartaron las investigaciones de acceso cerrado que no se pudieron filtrar en la DB2. Finalmente, como resultado del filtrado manual se obtuvieron 13 artículos, los cuales se miden mediante criterios de calidad en el siguiente punto.

Tabla 4: Resultados obtenidos aplicando criterios de exclusión/inclusión

Criterio	DB1	DB2	DB3
Primera búsqueda	17	99547	500
Año de publicación: 2015 - 2020	12	29125	323
Tipo documento: Artículo	9	21654	299
Idioma: Inglés y Español	9	21654	300
Acceso: Abierto	8	No cuenta	100
Lectura: Área de Computer Science	No cuenta	3748	No cuenta

2.3.2. Criterios de calidad

Se evaluó la calidad de los artículos escogidos mediante el instrumento que utilizaron Dybå y Dingsøyr [7] en su investigación, el cual es aplicable y amoldable a otros estudios.

Este proceso inició con la valoración de los artículos en base a 6 preguntas planteadas (criterios de calidad), cada pregunta se califica con 3 opciones de respuesta “Sí”, “En Parte” y “No”. Para darle más robustez a la evaluación, también se tomó como base la investigación de Wen [8] que puntúa las opciones de respuesta: “Sí” = 1, “En parte” = 0,5 y “No” = 0. El valor final para la calidad de los artículos se calcula sumando las puntuaciones de las respuestas a las preguntas anteriormente definidas.

Tabla 5: Criterio de calidad - Preguntas

PREGUNTA	ID
¿La investigación tiene base científica?	PC1
¿Se recopilaron datos para poder cumplir con el objetivo de la investigación?	PC2
¿En su metodología de desarrollo se utilizaron técnicas / algoritmos?	PC3
¿Se logró predecir el rendimiento académico del estudiante?	PC4
¿Ha sido desarrollada por más de un autor?	PC5
¿Se definieron de manera clara los resultados?	PC6

Tabla 6: Puntuación para artículos en base a preguntas

	PUNTUACIÓN			Total	%
	Sí (1pt)	En parte (0.5pt)	No (0pt)		
PC1					
PC2					
PC3					
PC4					
PC5					
PC6					

Sólo permanecieron las investigaciones con un nivel de calidad aceptable, es decir, con un valor final de calidad superior a 3 (50% de la puntuación perfecta). Por tanto, tras aplicar los criterios de calidad, resultaron un total de 13 artículos definitivos que permitirán concluir con el desarrollo restante de esta investigación.

3. Resultados y discusión

Las preguntas planteadas se han respondido en base a los 13 artículos seleccionados para esta investigación.

PI – 1: ¿En qué continentes y países se realizan más investigaciones sobre la predicción del rendimiento académico del estudiante?

Esta pregunta se pudo responder identificando el país de procedencia de cada artículo, luego se agruparon según el continente al que pertenezca cada uno y finalmente se contabilizó la cantidad de artículos por agrupaciones. Esta agrupación se puede ver en la Tabla 5.

Como resultado se obtuvo que los continentes que más investigan sobre el tema son Asia (9 investigaciones) y Oceanía (2 investigaciones), los otros continentes que también realizan investigaciones acerca del tema, pero en menor cantidad son Europa (1) y América (1). De la misma forma los países que más

contribuyen a su continente con investigaciones son India e Indonesia ubicados en Asia.

Tabla 7: Continentes y países que más investigan sobre el tema

Continente	País	N.º Artículos por país	Total Continente
Asia	Pakistán	1	9
	Indonesia	2	
	Filipinas	1	
	Iraq	1	
	India	3	
	China	1	
Oceanía	Australia	1	2
	Nueva Zelanda	1	
Europa	Turquía	1	1
América	México	1	1
Total:		13	

PI- 2: ¿Qué técnicas/algoritmos son los que más se utilizan para predecir el rendimiento académico de un estudiante?

Para dar respuesta a esta pregunta se analizó cada artículo con el fin de poder identificar las técnicas/algoritmos que se utilizaron para predecir el rendimiento académico. Luego de identificar y contabilizar el número de veces que se utilizaron las técnicas/algoritmos, se obtuvo que la cifra de utilización más alta es 4 (Naive Bayes, Neuronal Network, Decision Tree y Support Vector Machine), después la cifra 3 (Logistic Regression), casi al final la cifra 2 (K-Nearest N., Fuzzy Algorithm, Random Forest) y por último la cifra 1 (Genetic Algorithm, Sequential Minimal Optimization), todas estas cifras se pueden visualizar en la Tabla 6.

Tabla 8: Técnicas/Algoritmos más utilizados para la predicción

Técnica/Algoritmo	Artículo	Nº Utiliza.
Neural networks	[9] [12] [16] [10]	4
Decision Tree	[12] [11] [16] [20]	4
Naive bayes	[21] [16] [15] [14]	4
Bayes Ingenuo	[19]	1
Logistic regression	[21] [17] [16]	3
Support Vector Machine	[12] [17] [16] [10]	4
K-Nearest Neighbor	[21] [16]	2
Sequential Minimal O.	[16]	1
Genetic algorithm	[13]	1
Fuzzy algorithm	[11] [18]	2
Random Forest	[21] [20]	2

Para poder representar mejor estas cifras se generó un gráfico de pastel que permite identificar de manera más rápida las técnicas/algoritmos más usados para la predicción del

rendimiento académico, además, cuenta con tonalidad de colores para diferenciar los parámetros. Este gráfico se puede observar en la Figura 1.

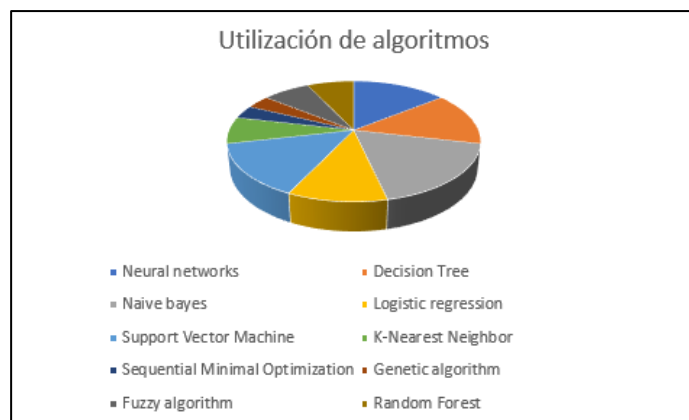


Figura 1: Técnicas/Algoritmos más usados para predicción – Gráfico Pastel

PI – 3: ¿Cuáles son los factores o características que se toman en cuenta para predecir el rendimiento académico de un estudiante?

Para responder esta pregunta se analizaron los artículos con respecto a la data utilizada para alimentar sus modelos, la información obtenida se organizó en un gráfico de barras horizontales con el fin de poder compararla entre sí. Como se puede observar en la Figura 2, hay factores que son utilizados casi siempre, siendo estos: el género, promedio final y apoyo familiar.

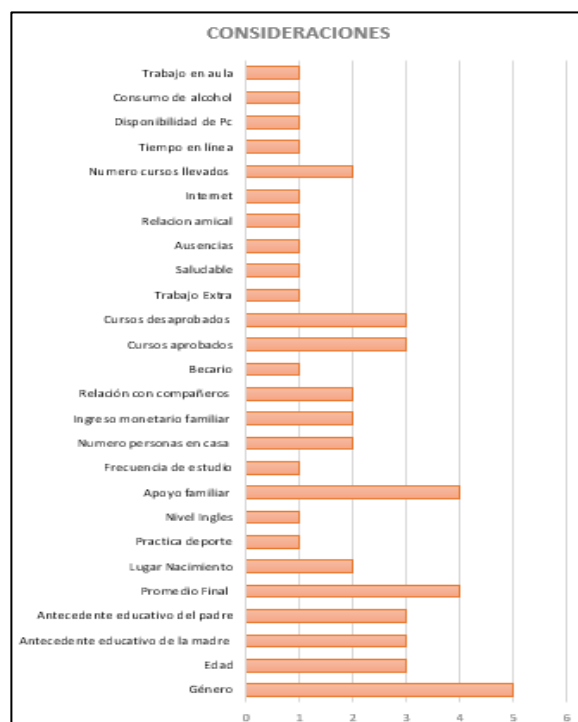


Figura 2: Consideración de factores para predicción – Barras Horizontales

4. Conclusiones

De acuerdo a los resultados que se obtuvieron, los cuales permitieron dar respuesta a las preguntas planteadas en la primera parte de esta investigación, se concluye que:

Los continentes que realizan más investigaciones acerca de la predicción del rendimiento académico son Asia (9) y Oceanía (2) teniendo un total de 11 investigaciones entre los dos, además los países que más aportan a estas cifras son India e Indonesia. De modo que se pueden tomar como buenos referentes para obtener conocimiento del tema o desarrollar soluciones.

Las técnicas/algoritmos que más se utilizan para predecir el rendimiento académico de los estudiantes son: algoritmo clasificador “Naive Bayes” (utilizado 4 veces), Neural Networks (utilizado 4 veces), Decision Tree (utilizado 4 veces) y Support Vector Machine (utilizado 4 veces).

Son muchos los factores/características que se toman en cuenta para poder predecir el rendimiento académico de un estudiante, ya sean factores internos como externos son tomados en cuenta, pero los que casi siempre están considerados son el género, promedio final, apoyo familiar, número de cursos matriculados y antecedentes educativos de los padres de familia.

5. Agradecimientos

Primero agradecer a Dios por acompañarme siempre, por brindarme salud y dotarme con las capacidades necesarias para afrontar cada desafío que pone la vida. También agradecer a mi familia por su amor y apoyo incondicional, siendo ellos mi motivación para lograr las metas planteadas. Finalmente, a la Ing. Jury por guiarme durante esta investigación, estar siempre al pendiente y brindarme palabras de aliento.

6. Referencias

[1] C. Jaspe. (2010, Julio 7). El Rendimiento escolar y estrategias de enseñanza [Online]. Available: <https://bit.ly/2UAANiB>. [Accessed: Nov 28, 2020].

[2] J. Oliver, “Factores asociados al rendimiento académico”, *Educación*, vol. 31, núm. 1, pp. 43-63, 2007.

[3] Laura Calvet Liñán and Ángel Alejandro Juan Pérez, “Educational Data Mining and Learning Analytics: differences, similarities, and time evolution”, *Universities and Knowledge Society Journal*, vol. 12, no. 3, pp. 98-112, 2015.

[4] T. M. Connolly, E. A. Boyle, E. MacArthur, T. Hainey y J. M. Boyle, “A systematic literature review of empirical evidence on computer games and serious games”, *Computers & Education*, vol. 59, no. 2, pp. 661-686, 2012.

[5] A. M. Navarro, M. P. Sancho y J. A. Medina Garrido, “BPMS para la gestión: una revisión sistemática de la literatura”, *Revista Española de Documentación Científica*, vol. 41, no. 3, p. e213, 2018.

[6] J. D. Velásquez, “Una Guía Corta para Escribir Revisiones Sistemáticas de Literatura Parte 3”, *Dyna*, vol. 82, no. 189, pp. 9-12, 2015.

[7] J. Wen, S. Li, Z. Lin, Y. Hu y C. Huang, “Systematic literature review of machine learning based software development effort estimation models”, *Information and Software Technology*, vol. 54, no. 1, pp. 41-59, 2012.

[8] Dybå, T.; Dingsøyr, T, “Empirical studies of agile software development: A systematic review”, *Information and Software Technology*, vol. 50, no. 10, pp. 833-859, 2008.

[9] Ali Centinkaya and Omer Kaan Baykan, “Prediction of middle school students’ programming talent using artificial neural networks”, *Engineering Science and Technology, an International Journal*, vol. 23, no. 6, pp. 1301-1307, December 2020.

[10] Hajra Waheed, Naif Radi and Julie Hardman, “Predicting academic performance of students from VLE big data using deep learning models”, *Computers in Human Behavior*, vol. 104, pp. 106-140, March 2020.

[11] Hashmia Hamsa, Simi Indiradevi and Jubilant Kizhakkethottam, “Student academic performance prediction model using decision tree and fuzzy genetic algorithm”, *Procedia Technology*, vol. 25, pp. 326-332, March 2016.

[12] Xing Xua, Jianzhong Wanga and Hao Peng, “Prediction of academic performance associated with internet usage behaviors using machine learning algorithms”, *Computers in Human Behavior*, vol. 98, pp. 166-173, April 2019.

[13] Al Farissi, Halina Mohamed Dahlan and Samsuryadi, “Genetic Algorithm Based Feature Selection With Ensemble Methods For Student Academic Performance Prediction”, *Journal of Physics: Conference Series*, vol. 1500, no. 3, October 2019.

[14] A Saifudin, Ekawati, Yulianti and T Desyani, “Forward Selection Technique to Choose the Best Features in Prediction of Student Academic Performance Based on Naïve Bayes”, *Journal of Physics: Conference Series*, vol. 1477, no. 1, January 2020.

[15] Jovel T. Rosado, Angelica P. Payne and Corzon B. Rebong, “eMineProve: Educational Data Mining for Predicting Performance Improvement Using Classification Method” in *IOP Conference Series: Materials Science and Engineering*, Harbin, China, 2019, pp. 1-9.

[16] Ali Salah Hashim, Wid Akeel Awadh and Alaa Khalaf Hamound, “Student Performance Prediction Model based on Supervised Machine Learning

Algorithms” in 2nd International Scientific Conference of Al-Ayen University, Thi-Qar, Iraq, 2019, pp. 1-19.

[17] Ralph Olusola Aluko, Olalekan Oshodi and Abiodun Olatunji, “Towards reliable prediction of academic performance of architecture students using data mining techniques”, *Journal of Engineering, Design and Technology*, vol. 18, no. 6, pp. 385-397, Marzo 2018.

[18] Anil Kumar Tiwari, G. Ramakrishna and Lokesh Kumar Sharma, “Academic performance prediction algorithm based on fuzzy data mining”, *IAES International Journal of Artificial Intelligence*, vol. 8, no. 1, pp. 26-32, Marzo 2019.

[19] Andrés Risco y Daniel Sánchez, “Análisis de datos de estudiantes de ingeniería para la predicción del rendimiento académico mediante técnica de clasificación bayesiana”, *Revista Dilemas Contemporáneos: Educación, Política y Valores*, vol. 13, no. 3, pp. 1-24, Agosto 2018.

[20] Sonali Kashyap, Shikhar Bajaj and Jabanjalin Hida, “Prediction of Student Academics Performance Based on their Previous Performance and Family Support”, *Research J. Pharm. and Tech*, vol. 10, no. 7, pp. 2075-2080, July 2017.

[21] Gomathy Ramaswami, Teo Susnjak and Anuradha Mathrani, “Using educational data mining techniques to increase the prediction accuracy of student academic performance”, *Information and Learning Sciences*, vol. 120, no. 2, pp. 451-467, July 2019.