Journal of Computer Science

Review

A Systematic Review Regarding the Prediction of Academic Performance

[1]Percy De-La-Cruz, [1]Rommel Rojas-Coaquira, [1]Hugo Vega-Huerta,
[1]José Pérez-Quintanilla and [2]Manuel Lagos-Barzola

[1]*Department of Computer Science, Universidad Nacional Mayor de San Marcos, Lima, Peru*
[2]*Department of Mathematics and Physics, San Cristóbal de Huamanga University, Ayacucho, Peru*

Article history
Received: 23-08-2022
Revised: 21-11-2022
Accepted: 07-12-2022

Corresponding Author:
Hugo Vega-Huerta
Department of Computer Science, Universidad Nacional Mayor de San Marcos, Lima, Peru
Email: hvegah@unmsm.edu.pe

Abstract: The prediction of students' academic performance is an area of great concern for universities and educational institutions since academic performance is one of the most important aspects of the learning process. To analyze this behavior, this study makes a critical analysis of the topic of interest and aims to review, analyze and summarize the latest research advances related to the prediction of academic performance. The systematic literature review method is applied to answer three questions: (1) What factors are determinants in predicting students' academic performance? (2) what methods are used to predict students' academic performance? and (3) what are the objectives and interests in predicting students' academic performance? After conducting the study of 50 outstanding articles, as results, we found that academic factor is the guideline for predicting academic performance; supervised machine learning is the most used technique, highlighting support vector machine, random forests, and neural networks; the most outstanding objectives for the application of prediction were: Student performance with 53%, risk of failure with 14%, search for student knowledge with 12%, avoid dropout with 12%, and decision making with 10%.

Keywords: Academic Performance Prediction Academic Performance of Students, Machine Learning, Predictive Models, Systematic Review of Literature

Introduction

The academic performance of students is a very important activity for schools and universities because it indicates how well students are educated. Therefore, academic performance has attracted the interest of the research community aiming to address issues, such as underachievement, student dropout, the risk of failure, and others. The prediction of the academic performance of a student should be done early as possible, so schools and universities can take appropriate actions.

Academic performance calculation is based on many factors, such as socioeconomic, personal, and environmental factors. The knowledge of these factors and student performance can help to prevent its impact. From the computational point of view, the literature offers a wide range of studies that attempts to predict student performance. Usually, such methods are based on Machine Learning (ML) techniques. Despite the advances, there is no consensus on the effectiveness of such algorithms to predict the academic performance of students. So, there is a strong need to search and know what are the factors that significantly determine those performances. The introduction of some predictive model to predict academic performance might help both institutions and students to take action to improve their performance, or at least, to avoid its dropping. This study reviews the research studies performed in this field between 2016 to 2022, aiming.

To provide a deep understanding of techniques and models to predict academic performance, which represents the performance of a student during a school period.

To compare the achievements of existing models and techniques taking into account different aspects, such as accuracy, effectiveness, and complexity.

To understand the factors that influence the estimation of academic performance.

To establish the research limitations and challenges that face current approaches to predict academic performance.

The remainder of this study is organized as follows. The first section describes the systematic review methodology that we adopted in this research, as well as the questions and objectives of this research. The second section analyzes in deep of the selected papers to answer the stated questions. Finally, the third section summarizes the findings and future guidelines of research.

Materials and Methods

This study is the result of a qualitative investigation of 54 research studies from January 2016 to July 2022, of which 51 studies have been selected for the systematic review and the remaining three have been taken as a reference. The studies are related to the prediction of academic performance, in which we recovered the determinant factor of performance, the technique of prediction, and its objectives. Selected studies come from different sources, such as book chapters, journals, and conferences. To perform the search and retrieval of studies, we used the indexer tool Scopus, which provides access to papers from a wide range of scientific institutions. Table 1 summarizes the main characteristics used for the literature review. It shows four groups of columns: Objectives, techniques, algorithms or methods, and the dominant factors for predicting academic performance. As follows, a brief description of chosen characteristic is provided.

Objectives

It answers the question: What for? Here, each reference has an interest for which the academic performance is used to.

Techniques

Here the type of technique used by different algorithms, methods, and tools is considered.

Algorithms and Methods

A set of algorithms and methods applied in each case are. For each study, a cell is marked with 'X' to indicate the selection. In the column of methods, for a study that has compared a set of algorithms, the best one is marked with the letter 'S' and it is highlighted with a shadow yellow.

Dominant Factors

Here, a factor or factors of high influence in academic performance is/are marked.

Research Planning

This study conducted a systematic review in which relevant studies to predict academic performance were identified, selected, and critically evaluated using various criteria, as presented in the results section. To organize our contributions, we formulated three key research questions:

Q1: What factors are determinants to predict academic performance?
Q2: What methods and algorithms are applicable to predict academic performance?
Q3: What are the goals and interests to predict academic performance?

The main objective of this review was to provide a comprehensive panorama of the academic performance prediction field and answer the above-proposed research questions, we adopted the PICO model (Petersen *et al*., 2015). Table 2 outlines four elements identified with the PICO strategy:

- Population: Refers to the research studies that address the academic performance of students
- Intervention: Refers to the models, techniques, and algorithms used to predict the academic performance of students
- Comparison: Refers to the variability of academic performance prediction among the reviewed models
- Outcome: Refers to the accurate results reported in the studies

The research studies reviewed belong to online scientific publications. These repositories include the ACM, IEEE, Science Direct, Springer, ISPJ, IGI, MDPI, and World Scientific digital libraries.

The searches were conducted using the Scopus tool to retrieve studies published from 2016 through 2022. For the search for process, the following keywords were considered: "Prediction performance academic student", "estimation performance academic student university", and "prediction performance academic machine learning".

Inclusion Criteria

Using the retrieval tool, we guarantee that returned papers are in the range of published years and restricted to the established keywords. Table 3 summarizes the first inclusion criteria for the retrieval of papers. Articles that did not meet these criteria were discarded.

To ensure the clarity of our retrieval strategy, we used the PRISMA systematic review (Moher *et al*., 2009), whose main steps are.

Identification

Identifies potential studies to be investigated by automatic or manual searches.

Selection

At this stage duplicate or irrelevant studies will be excluded. In addition, the inclusion criteria listed in Table 3 are strictly applied.

Table 1: Summary of the main features of the literature review

	Objectives		Techniques		Methods and algorithms			Factors		
References	Article type	Student desertion	Student Perform-ance	Decision Making	Risk of failure	Knowledge of students	Supervised machine learning	Unsupe-rvised Machine L.	Recomm-endation systems	Data mining techniques
Rimadana *et al.* (2019)	C		X				X			
Shanthini *et al.* (2018)	J		X				X			
Francis and Babu (2019)	J		X						X	
Gamao and Gerardo (2019)	J	X					X			
Xing *et al.* (2019)	J		X				X			
Kostopoulos *et al.* (2019)	J		X						X	
Deepali and Kamatchi (2019)	J		X				X			
Stephen *et al.* (2018)	J		X				X			
Rodrigues *et al.* (2019)	J		X				X			
Yang and Li (2018)	J					X			X	S
Waheed *et al.* (2020)	J			X			X			
Tsiakmaki *et al.* (2020)	J				X				X	
Karthikeyan *et al.* (2020)	J		X						X	
Jayaprakash *et al.* (2020)	C				X		X			
Rincón-Flores *et al.* (2020)	C					X	X			
Czibula *et al.* (2019)	J				X				X	
Tai Chui *et al.* (2020)	J				X		X			
Crivei *et al.* (2020)	C					X		X		
Popescu (2018)	J		X				X			
Thi Oanh (2017)	J					X		X		
Pandey and Taruna (2018)	C				X		X			
Hasan *et al* (2018)	C		X						X	
Kaunang and Rotikan (2018)	C		X						X	
Zhang *et al.* (2018)	C			X			X			
Saifudin and Desyani (2020)	J		X				X			
Echegaray-Calderon and Barrios-Aranibar (2015)	C		X				X			
Kamal and Ahuja (2019)	J		X						X	
Lau *et al.* (2019)	J		X				X			
Chohan *et al.* (2017)	J					X		X		
Ramaswami *et al.* (2019)	J		X						X	
Yao *et al.* (2019a)	J		X				X			
Imran *et al.* (2019)	J		X				X			
Roy and Garg (2017)	C	X							X	
Nandyala *et al.* (2019)	J			X					X	
Zhenpeng (2016)	C		X				X			
	C		X				X			
Buniyamin *et al.* (2016)	C		X						X	
Merchán (2016)	C	X							X	
Devasia *et al.* (2016)	C	X							X	
Iam-On (2017)	J	X							X	
Song *et al.* (2020)	J		X				X			
Jenitha *et al.* (2021)	J					X			X	
Hai-tao *et al.* (2021)	C				X		X			
Zulfikri *et al* (2021)	C		X				X			
Masangu *et al.* (2021)	J		X						X	
Erdem and Kaya (2021)	J		X				X			
Nibras and Walaa (2021)	C				X				X	
Mansouri *et al.* (2021)	J		X				X			
Cagliero *et al.* (2021)	J			X			X			
Sohaib and Fang (2021)	J			X			X			
Vega *et al.* (2022)	J	X							X	

References	BP-NN	ANN		SVM	NB	DT	SPR-AR	RF	SSR	RL	LMN-NR	CF	MLP
Rimadana *et al.* (2019)		X		S	X	X							
Shanthini *et al.* (2018)				S									
Francis and Babu (2019)	X	X	X	S									
Gamao and Gerardo (2019)			S	X									
Xing *et al.* (2019)		X											
Kostopoulos *et al.* (2019)		X					S						
Deepali and Kamatchi (2019)	X	S	S										
Stephen *et al.* (2018)											S		
Rodrigues *et al.* (2019)		X	X	X				S					
Yang and Li (2018)													
Waheed *et al.* (2020)	S	X						X					
Tsiakmaki *et al.* (2020)	S												
Karthikeyan *et al.* (2020)	X	X	S							X			
Jayaprakash *et al.* (2020)			X			S							
Rincón-Flores *et al.* (2020)						S							
Czibula *et al.* (2019)					S								
Tai Chui *et al.* (2020)		S											
Crivei *et al.* (2020)													
Popescu (2018)						X			S				
Thi Oanh (2017)		S							S				
Pandey and Taruna (2018)			S	S									
Hasan *et al* (2018)			X			S							
Kaunang and Rotikan (2018)				S		X							
Zhang *et al.* (2018)		X	X	X						S			
Saifudin and Desyani (2020)			S										
Echegaray-Calderon and Barrios-Aranibar (2015)	S									X			
Kamal and Ahuja (2019)	X	S	S	S		X							
Lau *et al.* (2019)	S												
Chohan *et al.* (2017)											S		
Ramaswami *et al.* (2019)			X			S		X					
Yao *et al.* (2019a)		X		X		X				X	X	S	
Imran *et al.* (2019)	X									X			
Roy and Garg (2017)			X							X			
Nandyala *et al.* (2019)			X			S		X					
Zhenpeng (2016)						X				X	S		
													S
Buniyamin *et al.* (2016)	X	X	X	X									
Merchán (2016)													
Devasia *et al.* (2016)			S										
Iam-On (2017)	S		X	X									
Song *et al.* (2020)				X		X				X			
Jenitha *et al.* (2021)	S	X	X										
Hai-tao *et al.* (2021)	X					X							
Zulfikri *et al* (2021)											S		
Masangu *et al.* (2021)		S		X		X				X	X		
Erdem and Kaya (2021)								S					
Nibras and Walaa (2021)		S	X							X			
Mansouri *et al.* (2021)													
Cagliero *et al.* (2021)		X	X	X		X							
Sohaib and Fang (2021)	X		X	S									
Vega *et al.* (2022)				X									

References	MLR	MTLTR	GFS	FN	K-means	Clustering	MMFA	J48	KNN	RAR	WTC	GA	PCA
Rimadana *et al.* (2019)													
Shanthini *et al.* (2018)													
Francis and Babu (2019)			S										
Gamao and Gerardo (2019)				S									
Xing *et al.* (2019)													
Kostopoulos *et al.* (2019)													
Deepali and Kamatchi (2019)													
Stephen *et al.* (2018)													
Rodrigues *et al.* (2019)													
Yang and Li (2018)													
Waheed *et al.* (2020)													
Tsiakmaki *et al.* (2020)													
Karthikeyan *et al.* (2020)					S								
Jayaprakash *et al.* (2020)													
Rincón-Flores *et al.* (2020)		X											
Czibula *et al.* (2019)													
Tai Chui *et al.* (2020)													
Crivei *et al.* (2020)							S			S			
Popescu (2018)						X							
Thi Oanh (2017)													
Pandey and Taruna (2018)		S											
Hasan *et al* (2018)													
Kaunang and Rotikan (2018)													
Zhang *et al.* (2018)													
Saifudin and Desyani (2020)													
Echegaray-Calderon and Barrios-Aranibar (2015)									S				
Kamal and Ahuja (2019)													
Lau *et al.* (2019)													
Chohan *et al.* (2017)													
Ramaswami *et al.* (2019)						X							
Yao *et al.* (2019a)													
Imran *et al.* (2019)					S								
Roy and Garg (2017)					S								
Nandyala *et al.* (2019)													
Zhenpeng (2016)		S			X								
Buniyamin *et al.* (2016)	S					X							
Merchán (2016)					S								
Devasia *et al.* (2016)													
Iam-On (2017)						X		S					
Song *et al.* (2020)											S		
Jenitha *et al.* (2021)													
Hai-tao *et al.* (2021)												S	
Zulfikri *et al* (2021)													
Masangu *et al.* (2021)													
Erdem and Kaya (2021)													
Nibras and Walaa (2021)													
Mansouri *et al.* (2021)													S
Cagliero *et al.* (2021)						X							
Sohaib and Fang (2021)													
Vega *et al.* (2022)													

References	CNN	GCN	LFCM	Others	Time	Acad-emic	Demogr-aphics	Beha-vior	Behavior internet	Social	Viewing behavior	Psychol-ogical	Economics
Rimadana *et al.* (2019)				X									
Shanthini *et al.* (2018)	X		X										
Francis and Babu (2019)			X	X	X								
Gamao and Gerardo (2019)			X										
Xing *et al.* (2019)						X							
Kostopoulos *et al.* (2019)			X				X						
Deepali and Kamatchi (2019)				X									
Stephen *et al.* (2018)			X					X					
Rodrigues *et al.* (2019)		X	X			X		X					
Yang and Li (2018)			X										
Waheed *et al.* (2020)			X	X									
Tsiakmaki *et al.* (2020)	X					X							
Karthikeyan *et al.* (2020)			X										
Jayaprakash *et al.* (2020)							X		X	X			
Rincón-Flores *et al.* (2020)			X										
Czibula *et al.* (2019)			X										
Tai Chui *et al.* (2020)			X										
Crivei *et al.* (2020)			X										
Popescu (2018)			X	X		X			X				
Thi Oanh (2017)			X										
Pandey and Taruna (2018)			X	X									
Hasan *et al* (2018)		X	X										
Kaunang and Rotikan (2018)		X	X	X			X						
Zhang *et al.* (2018)			X		X								
Saifudin and Desyani (2020)		X	X				X			X			
Echegaray-Calderon and Barrios-Aranibar (2015)			X	X	X		X		X	X			
Kamal and Ahuja (2019)			X	X	X		X						
Lau *et al.* (2019)			X				X			X			
Chohan *et al.* (2017)				X	X		X		X				
Ramaswami *et al.* (2019)			X										
Yao *et al.* (2019a)			X		X								
Imran *et al.* (2019)			X	X			X			X			
Roy and Garg (2017)		X	X		X		X			X			
Nandyala *et al.* (2019)			X	X			X	X					
Zhenpeng (2016)			X		X								
			X				X			X			
Buniyamin *et al.* (2016)			X	X	X								
Merchán (2016)	X		X	X			X						
Devasia *et al.* (2016)		X	X		X		X			X			
Iam-On (2017)			X	X									
Song *et al.* (2020)			X	X									
Jenitha *et al.* (2021)			X	X	X				X				
Hai-tao *et al.* (2021)			X										
Zulfikri *et al* (2021)			X										
Masangu *et al.* (2021)			X	X	X								
Erdem and Kaya (2021)			X	X			X			X			
Nibras and Walaa (2021)			X										
Mansouri *et al.* (2021)			X										
Cagliero *et al.* (2021)			X	X				X					
Sohaib and Fang (2021)	X		X										
Vega *et al.* (2022)			X										

Table 2: PICO model adopted in the literature review

Population	Intervention	Comparation	Output
Studies predicting student academic performance	Methods, techniques, and algorithms reviewed	Comparison of methods and algorithms	Quality and accuracy of approaches

Table 3: Inclusion criteria adopted in the literature review

Criteria	Description of criteria
Study approach	Studies predicting students' academic performance
Type of document	Articles, Conference journal, Books, book chapter
Year of publication	Studies published from 2016 to 2022
Language of publication	Only articles in English have been considered

Eligibility

This stage filters out those studies that do not follow the schemes determined by the PICO model in Table 2.

Inclusion

This is the result of applying the exclusions to launch the analysis of studies involving the prediction of academic performance.

DOI: 10.3844/jcssp.2022.1219.1231

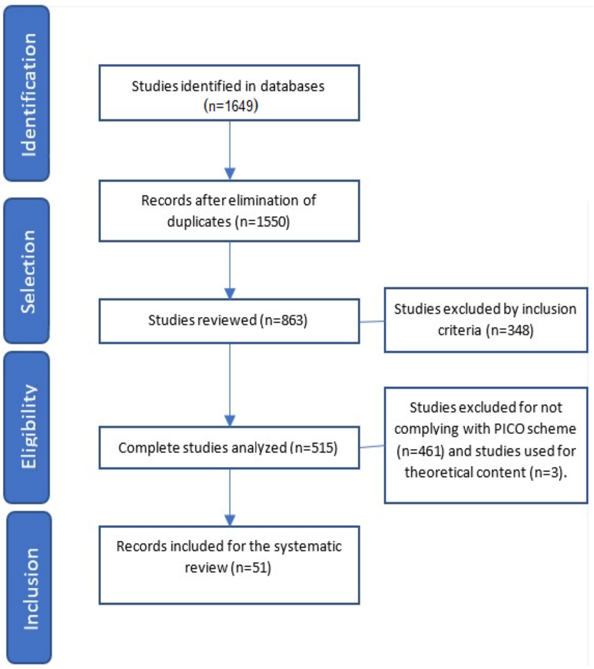

Fig. 1: PRISMA flowchart of the review methodology

Figure 1 illustrates the PRISMA steps. We can see that after the initial round of database searches it retrieved 1649 articles. After eliminating duplicates and applying the inclusion criteria, the number of studies was reduced to 515. Next, we excluded 461 articles that did not meet the PICO rules. Finally, we selected 54 papers, of which 51 are used for this review and three articles are used for theoretical support.

Data Extraction

Applying PRISMA (Moher *et al*., 2009), the selected group of studies was thoroughly analyzed to extract data to help answer our research questions. Data extracted are:

- The general information of the study (e.g., year of publication, type of paper)
- The algorithms used to predict the academic performance
- The Factors that determine the academic performance
- The data processing techniques used

To answer questions Q1, Q2, and Q3 we applied thematic analysis to the extracted data. The data were grouped and categorized according to the themes reported in the results section.

Results

This section presents general information about the articles analyzed, the techniques used to predict academic performance, the models and algorithms developed for prediction, and the factors that determine student academic performance.

Type of Paper and Year of Publication

A total of 51 studies were analyzed to help answer the questions posed in our research. Figure 2 shows that studies correspond to journals (33 studies, 65%) and conferences (18 studies, 35%).

Figure 3 shows the distribution in which studies were published.

Methods and Algorithms that Predict Academic Performance

Most methods or algorithms for predicting academic performance are ML-based classification algorithms because they provide higher accuracy against other proposals (Rimadana *et al*., 2019; Gamao and Gerardo, 2019; Xu *et al*., 2019; Vora and Kamatchi, 2019; Tsiakmaki *et al*., 2020; Jayaprakash *et al*., 2020). These classification algorithms are mainly: Support Vector Machine (SVM), Naive Bayes (NB), Decision Trees (DT), Random Forests (RF), and Artificial Neural Networks (ANN).

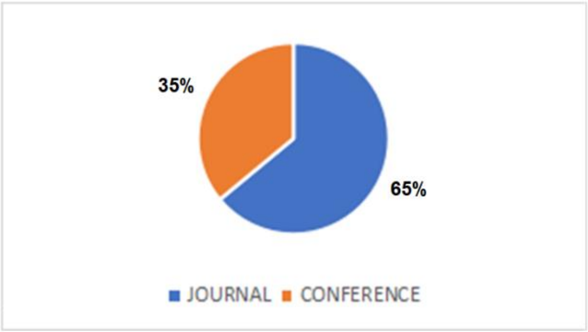

Fig. 2: Statistics on the type of document reviewed in the literature

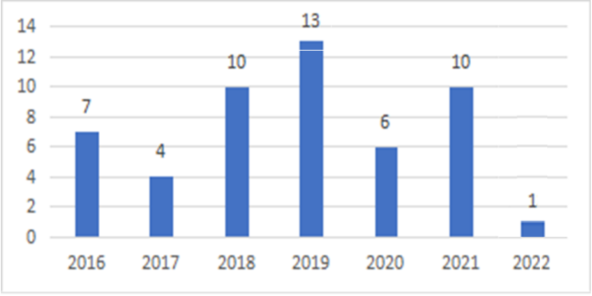

Fig. 3: Number of articles according to the year of publication

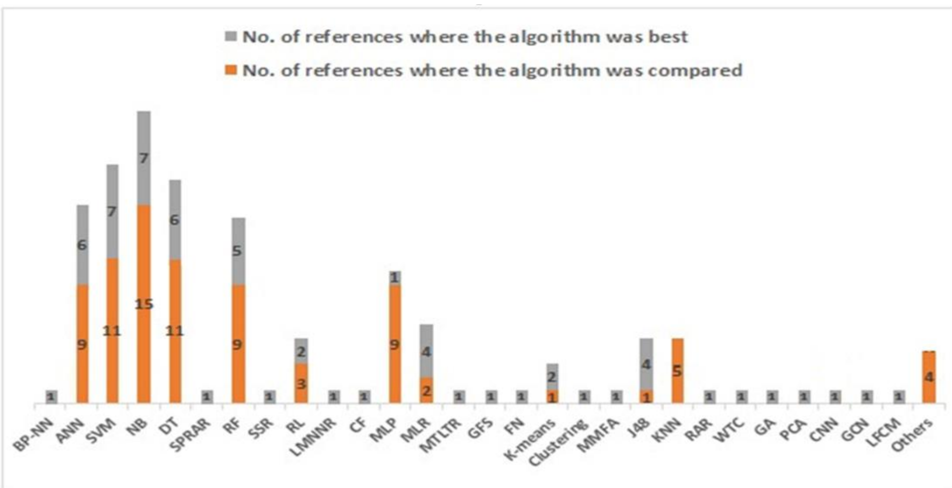

Fig. 4: Distribution of the algorithms and methods applied for prediction according to the literature review

It was found that 11 studies were based on the use of hybrid models to improve the accuracy of academic performance. These hybrid methods involve the integration of a set of algorithms (usually a pair) to achieve higher prediction accuracy compared to existing algorithms (Francis and Babu, 2019; Vora and Kamatchi, 2019; Karthikeyan *et al*., 2020; Crivei *et al*., 2020). According to (Fathian *et al*., 2016), hybrid models (as the 11 studies referred to) can be built using mainly bagging, boosting, and stacking approaches, these can increase performance because they can minimize classification errors and bias (Webb and Zheng, 2004) and to increase accuracy by 30% compared to any single model (Finlay, 2014).

We also found that 16 studies proposed an improved academic prediction performance. When compared with others it turned out to have better accuracy than this one (Gamao and Gerardo, 2019; Yang and Li, 2018; Czibula *et al*., 2019; Popescu and Leon, 2018; Mansouri *et al*., 2021).

Figure 4 shows the methods and algorithms most used to predict academic performance, where the lower part indicates that it has been put in comparison with others and the upper part indicates that it was the one that gave the best accuracy for each research case.

In Table 4, we categorize the 51 articles according to the main category used technique, which is: Supervised Machine Learning, Unsupervised Machine Learning, Recommender Systems, and Data Mining. These categories were identified according to the author's approach to predicting academic performance. Overall, 28 studies mark a trend of the best applied supervised machine learning technique for prediction, 19 studies applied the data mining technique, two studies opted for recommender systems, and one study on unsupervised machine learning.

Supervised Machine Learning

Supervised Learning (SL) are algorithms capable of reasoning from externally supplied instances to produce general hypotheses, which then make predictions about future instances (Rastrollo-Guerrero *et al*., 2020). Within this technique, we found algorithms and methods that were massively used in the literature. Among them, we highlight the SVM, RF the random forests, and NNs (Rimadana *et al*., 2019; Shanthini *et al*., 2018; Vora and Kamatchi, 2019; Jayaprakash *et al*., 2020; Pandey and Taruna, 2018; Imran *et al*., 2019). In the study (Popescu and Leon, 2018), the RF achieved a close approximation to the performance of undergraduate students that occurs in the classroom and improves as more data is fed into it.

Unsupervised Machine Learning

Unsupervised Learning (UL) is also known as classroom. The main difference between SL is that in UL the labels that the training set does not contain the labels that we like to predict. The study (Crivei *et al*., 2020), investigated the use of unsupervised machine learning methods, particularly principal component analysis and relational association rule mining to analyze academic performance. The authors proposed a new binary classification model called S PRAR (prediction of academic performance using relational association rules) to predict the outcome of a student in each academic discipline using Relational Association Rules (RAR).

Recommender Systems

In this technique, information is acquired explicitly by collecting user scores or implicitly by monitoring user behavior, such as visits to learning materials, downloaded documents, etc. Support vector machine, multiple linear regression, and collaborative filtering algorithms were identified for this technique (Tran *et al*., 2017; Chohan *et al*., 2017).

DOI: 10.3844/jcssp.2022.1219.1231

Table 4: Techniques that applied methods and algorithms for the prediction of academic performance.

Technique	Number of studies	Studies
Supervised machine learning	28 (55%)	Rimadana *et al*. (2019); Shanthini *et al*. (2018); Gamao and Gerardo (2019); Xu *et al*. (2019); Vora and Kamatchi (2019); Yang *et al*. (2018); Rodrigues *et al*. (2019); Waheed *et al*. (2020); Jayaprakash *et al*. (2020); Rincón-Flores *et al*. (2020); Chui *et al*. (2020); Popescu and Leon (2018); Pandey and Taruna (2018); Zhang *et al*. (2018); Saifudin and Desyani (2020); Echegaray-Calderon and Barrios-Aranibar (2015); Lau *et al*. (2019); Yao *et al*. (2019b); Imran *et al*. (2019); Li *et al*., 2016; Uddin and Lee, 2016; Song *et al*. (2021); Hai-Tao *et al*. (2021); Zulfikri *et al*. (2021); Erdem and Kaya (2021); Mansouri *et al*. (2021); Cagliero *et al*. (2021); Latif *et al*. (2021)
Unsupervised machine learning	1 (2%)	Crivei *et al*. (2020)
Recommender systems	2 (4%)	Tran *et al*. (2017); Chohan *et al*. (2017)
Data mining technique	20 (39%)	Francis and Babu (2019); Kostopoulos *et al*. (2019); Yang and Li (2018); Tsiakmaki *et al*. (2020); Karthikeyan *et al*. (2020); Czibula *et al*. (2019); Hasan *et al*. (2018); Kaunang and Rotikan (2018); Kamal and Ahuja (2019); Ramaswami *et al*. (2019); Roy and Garg (2017); Anvesh *et al*. (2019); Buniyamin *et al*. (2016); Rubiano and Garcia (2016); Devasia *et al*. (2016); Iam-On and Boongoen (2017); Jenitha *et al*. (2021); Masangu *et al*. (2021); Salih and Khalaf (2021); Vega *et al*. (2022)

Data Mining Technique

20 studies use this technique for academic performance prediction. We found that NNs, SVM, NB classifier, DT, Clustering, and J48 work successfully (Francis and Babu, 2019; Kostopoulos *et al*., 2019; Karthikeyan *et al*., 2020; Kaunang and Rotikan, 2018; Masangu *et al*., 2021; Salih and Khalaf, 2021). The study of (Yang and Li, 2018) uses a NN to estimate the performance/attributes of students according to students' prior knowledge as well as the performance/attributes of other students who have similar characteristics. In the study (Tsiakmaki *et al*., 2020), they implemented the deep NN with three hidden layers of neurons to predict the students' risk of failure.

Factors that Determine Academic Performance

We found that many factors influence the academic performance of students, which indicates the accuracy and quality of predictions (Rodrigues *et al*., 2019). There is a diversity of characteristics that influence academic performance and the selection of them is often difficult (Song *et al*., 2020). After a review of 51 studies, shown in Table 1, we found that the academic factor is the most relevant in 45 studies, followed by the demographic factors in 19 studies, 15 studies intended the social factors and 12 regarding the behavioral or conduct factors. To a lesser extent, we found the economic, time, psychological, internet behavior, and multimedia behavior factors with 9, 7, 5, 4, and 4 studies, respectively. Figure 5 outlines the distribution of the factors that influence academic performance and taking into consideration these factors would provide a greater impact and making a decision by schools. As follows, we provide a brief description of factors that influence academic performance.

Time

This factor uses time management skills as the preponderant factor, such as the time spent on a subject, time to study, time to take breaks, and travel time, among others. The study by Rimadana *et al*. (2019) proposes a system for predicting academic performance using time management characteristics. The process to solve the problem starts with obtaining the data through TSQ questionnaires.

Academic

This factor is the most important within this review and involves student grades, such as semester grade averages, past grades, and weighted grade averages, among others. The work of (Rubiano and Garcia, 2016) uses academic data of students from the Systems Engineering undergraduate program to predict academic performance by applying the J48 model.

Demographic

This factor is related to the general information of a student, in which data includes the attributes of age, gender, district of residence, and place of birth, among others. The study by Rodrigues *et al*. (2019) proposed a logistic regression model to predict academic performance from indicators of self-regulated behavior in LMS.

Social

Within the social factor are the relationship between parents, friends, family, communication and work aspects, and the surrounding social environment. The work in (Kostopoulos *et al*., 2019) built a model trained with social attributes extracted from the university.

Psychological

Among the psychological factors are motivation, perception, attention, and others (Echegaray-Calderon and Barrios-Aranibar, 2015; Chohan *et al*., 2017).

DOI: 10.3844/jcssp.2022.1219.1231

Economic

This factor has a high impact on academic performance. Economic problems of families, the possibility of parents paying for extra classes, living in pleasant conditions, scholarships, and family wealth, influence the performance of the student. The study (Erdem and Kaya, 2021) used additional statistical data from the PISA 2018 results as secondary data, together with economic data, parents' education, possessions, and financial support to provide a more reliable prediction.

Internet Behavior

Within this factor are all things everything related to the use of the internet, the time spent on it, searches performed, and sites searched, among others. The work in (Xu *et al*., 2019) studied the association between Internet usage behaviors and academic performance, and to predict academic performance they used an ML approach. As input data, they used online duration of navigation on the internet, internet traffic, the volume of data, and frequency of connections.

Multimedia Behavior

This factor involves the behavior toward the learning management systems, such as Moodle or other virtual classrooms. The interaction of a student with the resources furnished by the system (documents, videos, and other didactical material), provides good input data to train an ML model. The work in (Rodrigues *et al*., 2019) developed a model to predict academic performance Using behavioral data from an LMS platform used in a public university during a distance education course. Data was collected over seven years.

Table 5, shows the Factors' importance and how they affect performance.

Table 5: How Factors affect performance

Factor	Importance	How factors affect performance
Time	The investment of time in teaching and learning the process is correlated with the academic performance	Determining the number of hours/days spent by the of students or teachers in academic activities that contribute to the prediction of university academic performance
Academician	Factor that qualitatively involves academic performance based grades obtained from the student	It is directly related to academic performance; grades determine the academic situation of the student and their decision-making
Demographic	Factor that links student indicators such as age, gender, etc., with their Local environment	Determining these factors helps select and classify data to analyze academic performance
Social	This factor is linked to social and family contexts They determine the influence external to the student	Affect the student's performance according to the influence perceived by the family and society, which influences the learning process
Psychological	It can positively or negatively affect the academic performance of the student such as attitudes self-esteem, motivation among others	The psychological factor affects the student because causes a positive or negative stimulus that will determine the academic performance
Economic	This factor considers the financial support that the student receives (family or external) to improve their academic performance	This factor improves opportunities for access to resources, which improves the quality of the student learning
Internet behavior	This factor is associated with the use of the internet and the investigation of its proper use	Proper use helps the student to improve performance, but uncontrolled use can be distracting
Multimedia behavior	It involves the management of teaching and use of the resources provided by virtual educational platforms by the student and the teacher	Platforms such as Moodle or virtual classrooms and other digital resources help in the teaching process that facilitates better interaction with students

Table 6: Methods using the time factor

	Methods and algorithms						
References	SVM	NB	DT	RF	RL	J48	Factor TIME
Rimadana *et al*. (2019)	S	X	X				X
Rodrigues *et al*. (2019)	X	X	X		S		X
Hasan *et al*. (2018)		X		S			X
Kaunang and Rotikan (2018)			S	X			X
Saifudin and Desyani (2020)		S					X
Roy and Garg (2017)		X				S	X
Devasia *et al*. (2016)		S					X
Total (7)	1	2	1	1	1	1	

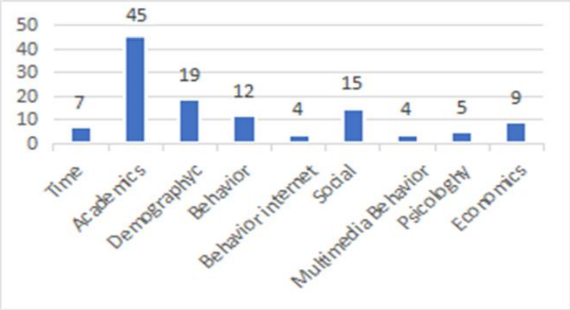

Fig. 5: Distribution of factors determining academic performance

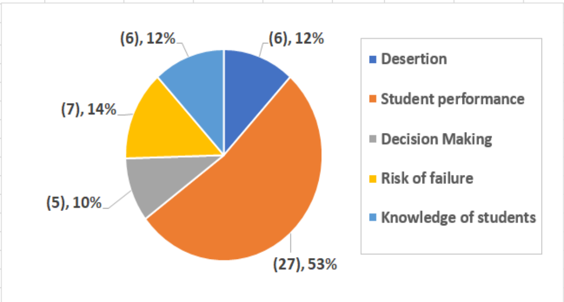

Fig. 6: Number and percentage of objectives for predicting student performance

Table 6 shows the data extracted from Table 1, concerning the Time factor, where it can be seen that seven references mention the existence of at least one algorithm that uses the Time factor for the prediction of academic performance, among which SVM, NB, DT, RF, RL, J48 stand out, of which two authors agree that the best algorithm, in this sense, is the NB (Naive Bayesian Classifier). Similarly, by analyzing Table 1, it can be determined which is the best algorithm for each factor.

Objectives of Interest

Based on our review, different objectives have guided the authors for academic performance prediction. These objectives can be compiled into 6 groups: Attrition, student performance, decision-making, risk of failure, recommended activities and resources, and student knowledge. Figure 6 shows the number and percentage of each of these objectives in the surveyed papers. Overall, student performance occupies 53% of studies, followed by risk of failure with 14%, the search for student knowledge was addressed in 12% of papers, the objective of avoiding dropout with 12%, and finally, the decision-making was addressed in 10% of articles.

Student Dropout (Desertion)

Six studies focused on dropout, in which they proposed predicting academic performance to solve this problem. In higher education, the increased attrition of students not only affect themselves but also the reputation of the institutions in which they study (Devasia *et al*., 2016). In addition, it has been established that in the first year of study, students have the highest risk of dropping out or not completing their degree on time (Iam-On *et al*., 2017).

Student Performance

One of the most essential objectives for institutions is the prediction of academic performance because they want to know the performance that students will have (low, medium, and high performance) to supervise them aiming to ensure their education completion (Saifudin and Desyani, 2020; Yao *et al*., 2019b; Song *et al*., 2020). Furthermore, it was identified that attending classes via online education platforms and their interaction with them, can be used to predict academic performance (Xu *et al*., 2019; Rodrigues *et al*., 2019). This objective is important and it is important to take into account in designing a predictive model (Song *et al*., 2020).

Decision Making

The reviewed studies argue that knowing the academic performance of students helps to feedback and pedagogical support to students, where interested students to improve their performance could receive reinforcement (Waheed *et al*., 2020). However, many times the decisions are unclear, and the interpretability and explainability of models are often limited. So, it is recommendable to enforce an explainable system of academic performance to reinforce both the students and the curriculum of institutions (Cagliero *et al*., 2021; Latif *et al*., 2021).

Risk of Failing or Failure

Education is considered essential for social progress and ensuring that students pass their courses, so it is of high concern to schools (Chui *et al*., 2020). Early identification of the risk of failure can help to identify weak students and to take corrective actions to avoid or mitigate risks (Pandey and Taruna, 2018). According to the study of (Jayaprakash *et al*., 2020), a model that can significantly reduce academic failure helps students to get prior consultation and more opportunities to decide on a successful path at a very early stage.

Knowledge of Students

A few studies give relevance to this objective. However, the performance prediction models work according to this characteristic of students. The work in (Chohan *et al*., 2017) concluded that students who often assist to libraries increase their knowledge and are more prone to get a successful academic performance. Moreover, under the insight of this objective, students should know their own capabilities to improve their performance. In another study by Tran *et al*. (2017), it was found that students who chose elective courses without knowing the required background, cause students study a little motivation. Then, a model that takes into account the necessary background for the selection of courses as an aid for students is desirable. The study by Rincón-Flores *et al*. (2020), found that a general picture of students' performance is enough to improve, and knowing their predictive scores at the beginning of the course will help them to improve in class.

Discussion

In this research, 51 articles aimed to predict the academic performance of higher-level students have been reviewed and the questions proposed in the methodology were answered. The following are some of the most relevant findings of the research.

In this review, research in the academic prediction of students at the higher level proposes different ways to analyze it. We have considered three points of view: According to the factors, the algorithms or methods, and the techniques.

Among the factors, the academic factor was identified as the most relevant for prediction accompanied by demographic, social, and behavioral data. These factors give better accuracy. However, the use of an excessive combination of factors might cause difficulties at the time of processing.

In the case of the algorithms or methods, it was determined that the classifiers were relevant in the experimental processes since they were objects of comparison and many times turned out to be the best ones. These algorithms were: NN, SVM, NB, DT, and RF. In addition, there are proposals that achieved a higher percentage of accuracy; however, they depend on the factors used and their objectives. Hybrid model proposals have also been considered, where they reached a better accuracy, although the process is more complex than the classical algorithms.

Finally, on the side of the techniques, it was found that supervised learning is the best for prediction, followed by data mining. Note that the most common objectives for prediction were student performance, risk of failure, dropout, and decision-making.

Conclusion and Future Works

In this study, a systematic literature review of 1649 articles related to the prediction of academic performance of university-level students was carried out, removing the duplicates and irrelevant ones, the number was reduced to 862 articles and excluding those that did not meet the inclusion criteria (less than 6 years old and that the methods used for the prediction are artificial intelligence) 515 were obtained, of which reviewing their abstracts and applying the PICO criterion, finally, 54 articles were selected, of which 3 served as theoretical support and 51 were the relevant articles that were completely analyzed as shown in Fig. 1. Articles were analyzed based on the proposed planning, consequently, the conclusions of the work were related to the three research questions that consider the factors, algorithms, methods or techniques and objectives.

In the factors, academics were identified as the most relevant for prediction, accompanied by demographic, social, and behavioral data. Although the excessive combination of several factors can cause problems at the time of processing.

In the case of algorithms or methods, it was determined that the classifiers were relevant in the experimental processes since they were the object of comparison and often proved to be the best. These algorithms were: Neural Networks (ANN), Support Vector Machine (SVM), Naive Bayesian Classifier (NB), Decision Trees (DT), and Random Forests (RF).

On the techniques side, we conclude that supervised machine learning is one of the best for prediction, followed by data mining methods.

The most outstanding objectives for the application of prediction were student performance at 53%, risk of failure at 14%, search for student knowledge at 12%, avoiding dropout at 12%, and decision-making at 10% as shown in Fig. 6.

Of the 51 studies analyzed, intelligent models were presented to predict the academic performance of students from different approaches and the accuracy of the prediction is relative, therefore, as a general conclusion, we can state that the prediction of academic performance depends on the selection of factors, algorithms, and techniques.

In a similar study, (Arifin *et al*., 2021) evaluated methods for predicting academic performance and including similar classifiers, but conclude that more research is required to recommend which is the most appropriate, so, in agreement with that study, as future work we hope to expand our research, considering a larger number of more recent references.

Funding Information

The authors have not received any financial support or funding to report.

Author’s Contributions

Percy De-La-Cruz: Generated the topic of the research, designed the research plan, and organized the study.

Rommel Rojas-Coaquira: Performed the literature review, data extraction and analysis, and drafting of the manuscript.

Hugo Vega-Huerta: Principal critical reviewer of the manuscript and gave the final approval of the version to be presented.

José Pérez-Quintanilla: Critical reviewer of the manuscript provided major improvements to the content.

Manuel Lagos-Barzola: Contributed to data analysis and validation of results.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Anvesh, K., Srilatha, M., Raghunadha Reddy, T., Gopi Chand, M., & Jyothi, G. D. (2019). Improving Student Academic Performance Using an Attribute Selection Algorithm. In *First International Conference on Artificial Intelligence and Cognitive Computing* (pp. 539-548). Springer, Singapore. https://doi.org/10.1007/978-981-13-1580-0_52

Arifin, M., Wibowo, A., & Warsito, B. (2021, September). Comparative Analysis on Educational Data Mining Algorithm to Predict Academic Performance. In *2021 International Seminar on Application for Technology of Information and Communication (iSemantic)* (pp. 173-178). IEEE. https://doi.org/10.1109/iSemantic52711.2021.9573185

Buniyamin, N., bin Mat, U., & Arshad, P. M. (2015, November). Educational data mining for prediction and classification of engineering students' achievement. In *2015 IEEE 7th International Conference on Engineering Education (ICEED)* (pp. 49-53). IEEE. https://doi.org/10.1109/ICEED.2015.7451491

Cagliero, L., Canale, L., Farinetti, L., Baralis, E., & Venuto, E. (2021). Predicting student academic performance by means of associative classification. *Applied Sciences*, *11*(4), 1420. https://doi.org/10.3390/app11041420

Chohan, T. M., Bhatti, R., & Naeem, S. B. (2017, September). Prediction of Academic Performance of the University Students Through Their Use of Library Electronic Resources and Their Self-efficacy. In *European Conference on Information Literacy* (pp. 557-567). Springer, Cham. https://doi.org/10.1007/978-3-319-74334-9_57

Chui, K. T., Fung, D. C. L., Lytras, M. D., & Lam, T. M. (2020). Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. *Computers in Human Behavior*, *107*, 105584. https://doi.org/10.1016/j.chb.2018.06.032

Crivei, L. M., Czibula, G., Ciubotariu, G., & Dindelegan, M. (2020, May). Unsupervised learning based mining of academic data sets for students' performance analysis. In *2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI)* (pp. 000011-000016). IEEE. https://doi.org/10.1109 / SACI49304.2020.9118835

Czibula, G., Mihai, A., & Crivei, L. M. (2019). S PRAR: A novel relational association rule mining classification model applied for academic performance prediction. *Procedia Computer Science*, *159*, 20-29. https://doi.org/10.1016/j.procs.2019.09.156

Devasia, T., Vinushree, T. P., & Hegde, V. (2016, March). Prediction of students performance using Educational Data Mining. In *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)* (pp. 91-95). IEEE. https://doi.org/10.1109/SAPIENCE.2016.7684167

Echegaray-Calderon, O. A., & Barrios-Aranibar, D. (2015, October). Optimal selection of factors using Genetic Algorithms and Neural Networks for the prediction of students' academic performance. In *2015 Latin America Congress on Computational Intelligence (LA-CCI)* (pp. 1-6). IEEE. https://doi.org/10.1109/LA-CCI.2015.7435976

Erdem, C., & Kaya, M. (2021). Socioeconomic status and wellbeing as predictors of students' academic achievement: Evidence from a developing country. *Journal of Psychologists and Counsellors in Schools*, 1-19. https://doi.org/10.1017/jgc.2021.10

Fathian, M., Hoseinpoor, Y., & Minaei-Bidgoli, B. (2016). Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods. *Kybernetes*. https://www.emerald.com/insight/content/doi/10.1108/K-07-2015-0172/full/html

Finlay, S. (2014). *Predictive analytics, data mining and big data: Myths, misconceptions and methods*. Springer.

Francis, B. K., & Babu, S. S. (2019). Predicting academic performance of students using a hybrid data mining approach. *Journal of medical systems*, *43*(6), 1-15. https://doi.org/10.1007/s10916-019-1295-4

Gamao, A. O., & Gerardo, B. D. (2019). Prediction-based model for student dropouts using modified mutated firefly algorithm. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(6), 3461-3469. https://doi.org/10.30534/ijatcse/2019/122862019

Hai-Tao, P., Ming-Qu, F., Hong-Bin, Z., Bi-Zhen, Y., Jin-Jiao, L., Chun-Fang, L., ... & Rui, S. (2021). Predicting academic performance of students in Chinese-foreign cooperation in running schools with graph convolutional network. *Neural Computing and Applications*, *33*(2), 637-645. https://doi.org/10.1007/s00521-020-05045-9

Hasan, R., Palaniappan, S., Raziff, A. R. A., Mahmood, S., & Sarker, K. U. (2018, August). Student academic performance prediction by using decision tree algorithm. In *2018 4th international conference on computer and information sciences (ICCOINS)* (pp. 1-5). IEEE. https://doi.org/10.1007/978-981-10-6602-3_16

Iam-On, N., & Boongoen, T. (2017). Improved student dropout prediction in Thai University using ensemble of mixed-type data clusterings. *International Journal of Machine Learning and Cybernetics*, *8*(2), 497-510. https://doi.org/10.1007/s13042-015-0341-x

Imran, M., Latif, S., Mehmood, D., & Shah, M. S. (2019). Student Academic Performance Prediction using Supervised Learning Techniques. *International Journal of Emerging Technologies in Learning*, *14*(14). https://doi.org/10.3991/ijet.v14i14.10310

Jayaprakash, S., Krishnan, S., & Jaiganesh, V. (2020, March). Predicting students academic performance using an improved random forest classifier. In *2020 international conference on emerging smart computing and informatics (ESCI)* (pp. 238-243). IEEE. https://doi.org/10.1109/ESCI48226.2020.9167547

Jenitha, T., Santhi, S., & Jeba, J. M. P. (2021). Prediction of Students' Performance based on Academic, Behaviour, Extra and Co-Curricular Activities. *Webology*, *18*(SI01), 262-279. https://doi.org/10.14704/WEB/V18SI01/WEB18058

Kamal, P., & Ahuja, S. (2019). An ensemble-based model for prediction of academic performance of students in undergrad professional course. *Journal of Engineering, Design and Technology*, *17*(4), 769-781. https://doi.org/10.1108/JEDT-11-2018-0204

Karthikeyan, V. G., Thangaraj, P., & Karthik, S. (2020). Towards developing hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation. *Soft Computing*, *24*(24), 18477-18487. https://doi.org/10.1007/s00500-020-05075-4

Kaunang, F. J., & Rotikan, R. (2018, October). Students' academic performance prediction using data mining. In *2018 Third International Conference on Informatics and Computing (ICIC)* (pp. 1-5). IEEE. https://doi.org/10.1109/IAC.2018.8780547

Kostopoulos, G., Kotsiantis, S., Fazakis, N., Koutsonikos, G., & Pierrakeas, C. (2019). A semi-supervised regression algorithm for grade prediction of students in distance learning courses. *International Journal on Artificial Intelligence Tools*, *28*(04), 1940001. https://doi.org/10.1142/S0218213019400013

Latif, S., XianWen, F., & Wang, L. L. (2021). Intelligent decision support system approach for predicting the performance of students based on three-level machine learning technique. *Journal of Intelligent Systems*, *30*(1), 739-749. https://doi.org/10.1515/jisys-2020-0065

Lau, E. T., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*, *1*(9), 1-10. https://doi.org/10.1007/s42452-019-0884-7

Li, Z., Shang, C., & Shen, Q. (2016, July). Fuzzy-clustering embedded regression for predicting student academic performance. In *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 344-351). IEEE. https://doi.org/10.1109/FUZZ-IEEE.2016.7737707

Mansouri, T., ZareRavasan, A., & Ashrafi, A. (2021). A Learning Fuzzy Cognitive Map (LFCM) approach to predict student performance. *Journal of Information Technology Education: Research*, *20*, 221-243. https://doi.org/10.28945/4760

Masangu, L., Jadhav, A., & Ajoodha, R. (2021). Predicting Student Academic Performance Using Data Mining Techniques. *Advances in Science, Technology and Engineering Systems Journal*, *6*(1), 153-163. https://doi.org/10.25046/aj060117

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2009). *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statemen*. https://doi.org/10.1136/bmj.b2700

Merchán, V. (2016). Mix it in rice, top it on karahis and enjoy it with BBQ. https://www.transfermarkt.com/gonzalo-merchan/profil/spieler/282416

Nibrasm S., & Walaa, Z. K. (2021). Prediction of student's performance through educational data mining techniques. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(3), 1708–1715. https://doi.org/10.11591/ijeecs.v22.i3.pp1708-1715

Nandyala, A. R., Darpe, A. K., & Singh, S. P. (2020). Damage localization in cross-ply laminated composite plates under varying temperature conditions using Lamb waves. *Meas. Sci. Technol*. 31, 064003. https://iopscience.iop.org/article/10.1088/1361-6501/ab6eca/meta

Pandey, M., & Taruna, S. (2018). An ensemble-based decision support system for the students' academic performance prediction. In *ICT Based Innovations* (pp. 163-169). Springer, Singapore. https://doi.org/10.1007/978-981-10-6602-3_16

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology, 64*, 1-18. https://doi.org/10.1016/j.infsof.2015.03.007

Popescu, E., & Leon, F. (2018). Predicting academic performance based on learner traces in a social learning environment. *IEEE Access*, *6*, 72774-72785. https://doi.org/10.1109/ACCESS.2018.2882297

Popescu, D. (2018). The Importance of Bearing Witness. *East European Politics and Societies*, 32(2), 315–319. https://doi.org/10.1177/0888325417745129

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 Sciences*, *120*(7/8), 451-467. https://doi.org/10.1108/ILS-03-2019-0017

Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., & Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: A review. *Applied sciences*, *10*(3), 1042. https://doi.org/10.3390/app10031042

Rimadana, M. R., Kusumawardani, S. S., Santosa, P. I., & Erwianda, M. S. F. (2019, December). Predicting student academic performance using machine learning and time management skill data. In *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 511-515). IEEE. https://doi.org/10.1109/ISRITI48646.2019.9034585

Rincón-Flores, E. G., López-Camacho, E., Mena, J., & López, O. O. (2020, April). Predicting academic performance with Artificial Intelligence (AI), a new tool for teachers and students. In *2020 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1049-1054). IEEE. https://doi.org/10.1109/EDUCON45650.2020.9125141

Rodrigues, R. L., Ramos, J. L. C., Silva, J. C. S., Dourado, R. A., & Gomes, A. S. (2019). Forecasting students' performance through self-regulated learning behavioral analysis. *International Journal of Distance Education Technologies (IJDET)*, *17*(3), 52-74. https://doi.org/10.4018/IJDET.2019070104

Roy, S., & Garg, A. (2017, October). Predicting academic performance of student using classification techniques. In *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)* (pp. 568-572). IEEE. https://doi.org/10.1109/UPCON.2017.8251112

Rubiano, S. M. M., & Garcia, J. A. D. (2016). Analysis of data mining techniques for constructing a predictive model for academic performance. *IEEE Latin America Transactions*, *14*(6), 2783-2788. https://doi.org/10.1109/TLA.2016.7555255

Saifudin, A., & Desyani, T. (2020, March). Forward selection technique to choose the best features in prediction of student academic performance based on Naïve Bayes. In *Journal of Physics: Conference Series* (Vol. 1477, No. 3, p. 032007). IOP Publishing. https://doi.org/10.1088 / 1742-6596 / 1477/3/032007

Salih, N. Z., & Khalaf, W. (2021). Prediction of studentâ€™ s performance through educational data mining techniques. *Indonesian Journal of Electrical Engineering and Computer Science*, *22*(3), 1708-1715. https://doi.org/10.11591/ijeecs.v22.i3.pp1708-1715

Shanthini, A., Vinodhini, G., & Chandrasekaran, R. M. (2018). Predicting Students' Academic Performance in the University Using Meta Decision Tree Classifiers. *J. Comput. Sci.*, *14*(5), 654-662. https://doi.org/10.3844/jcssp.2018.654.662

Song, X., Li, J., Sun, S., Yin, H., Dawson, P., & Doss, R. R. M. (2020). SEPN: a sequential engagement based academic performance prediction model. *IEEE Intelligent Systems*, *36*(1), 46-53. https://doi.org/10.1109/MIS.2020.3006961

Sohaib, A., Latifa, F. B., Wen, X., Iwendi, C., Li-Li, F. W., Mohsin, S. M., Han, Z., & Band, S. S. (2022). AI-empowered, blockchain and SDN integrated security architecture for IoT network of cyber physical systems. 181, 274-283. https://doi.org/10.1016/j.comcom.2021.09.029

Tran, T. O., Dang, H. T., Dinh, V. T., Truong, T. M. N., Vuong, T. P. T., & Phan, X. H. (2017). Performance prediction for students: A multi-strategy approach. BULGARIAN ACADEMY OF SCIENCES. https://doi.org/10.1515/cait-2017-0024

Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2020). Transfer learning from deep neural networks for predicting student performance. *Applied Sciences*, *10*(6), 2145. https://doi.org/10.3390/app10062145

Thi Oanh, N. (2017). Hanoi University of Science and Technology School of Information and Communication Technology. *Hanoi University of Science and Technology*. https://www.researchgate.net/profile/Thi-Oanh-Nguyen

Uddin, M. F., & Lee, J. (2016, August). Predicting good fit students by correlating relevant personality traits with academic/career data. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 968-975). IEEE. https://doi.org/10.1109/ASONAM.2016.7752357

Vega, H., Sanez, E., De La Cruz, P., Moquillaza, S., & Pretell, J. (2022). Intelligent System to Predict University Students Dropout. *International Journal of Online & Biomedical Engineering*, *18*(7). https://doi.org/https://doi.org/10.3991/ijoe.v18i07.30195

Vora, D. R., & Rajamani, K. (2019). A hybrid classification model for prediction of academic performance of students: A big data application. *Evolutionary Intelligence*, 1-14. https://doi.org/10.1007/s12065-019-00303-9

Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior*, *104*, 106189. https://doi.org/10.1016/j.chb.2019.106189

Webb, G. I., & Zheng, Z. (2004). Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques. *IEEE Transactions on Knowledge and Data Engineering*, *16*(8), 980-991. https://ieeexplore.ieee.org/abstract/document/1318582

Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior, 98,* 166-173. https://doi.org/10.1016/j.chb.2019.04.015

Xing, L., Zhang, H., Qi, R., Tsao, R., & Mine, Y. (2019). Recent Advances in the Understanding of the Health Benefits and Molecular Mechanisms Associated with Green Tea Polyphenols. *J. Agric. Food Chem*. 67(4), 1029-1043. https://doi.org/10.1021/acs.jafc.8b06146

Yang, F., & Li, F. W. (2018). Study on student performance estimation, student progress analysis and student potential prediction based on data mining. *Computers & Education*, *123*, 97-108. https://doi.org/10.1016/j.compedu.2018.04.006

Yang, S. J., Lu, O. H., Huang, A. Y., Huang, J. C., Ogata, H., & Lin, A. J. (2018). Predicting students' academic performance using multiple linear regression and principal component analysis. *Journal of Information Processing*, *26*, 170-176. https://doi.org/10.2197/ipsjjip.26.170

Yao, H., Lian, D., Cao, Y., Wu, Y., & Zhou, T. (2019a). Predicting academic performance for college students: A campus behavior perspective. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *10*(3), 1-21. https://doi.org/10.1145/3299087

Yao, H., Wei, Y., Huang, J., & Li, Z. (2019b). Hierarchically Structured Meta-learning. https://www.researchgate.net/publication/333102365_Hierarchically_Structured_Meta-learning

Zhang, X., Xue, R., Liu, B., Lu, W., & Zhang, Y. (2018, July). Grade prediction of student academic performance with multiple classification models. In *2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (pp. 1086-1090). IEEE. https://doi.org/10.1109/FSKD.2018.8687286

Zulfikri, M. F., Shaharudin, S. M., Rajak, N. A. A., & Ibrahim, M. S. B. (2021). Predictive Analytics on Academic Performance in Higher Education Institution during COVID-19 using Regression Model. *International Journal of Biology and Biomedical Engineering*. https://doi.org/10.46300/91011.2021.15.21

Zhenpeng, H. (2016). Nestlé Milkpak Goodness. https://www.researchgate.net/profile/Zhenpeng-Hu