

A Thematic Review on Predicting Student Performance Using Data Mining

Mohammad Izzuan Termedi, Aini Marina Ma'rof, Habibah binti
Ab. Jalil, Iskandar Ishak

Faculty of Educational Studies Universiti Putra Malaysia, Jalan Universiti 1, 43400 Serdang,
Selangor, Malaysia

Corresponding Author Email: izzues66@gmail.com

To Link this Article: <http://dx.doi.org/10.6007/IJAROSS/v13-i11/19563> DOI:10.6007/IJAROSS/v13-i11/19563

Published Date: 23 November, 2023

Abstract

Student performance, a key criterion for academic placement and success, can be effectively predicted and potentially improved using advanced data mining techniques. However, more studies delve into patterns and methods for forecasting student performance in this context. This study presents a comprehensive thematic review of the literature published between 2016 and 2022 to highlight prevailing trends and underline the significant gaps in this rapidly growing research domain. The review emphasizes that the highest number of related publications were seen in 2021, with a clear focus on higher educational institutions. Two major factors were recurrent in these studies: academic records and demographic information of the students, widely recognized as significant determinants of academic success. This observation reaffirms previous research asserting the substantial role of academic records in predicting student performance. Our analysis expands the scope of prior literature reviews by incorporating studies from Mendeley and Science Direct databases. Interestingly, we discovered a glaring research gap in predicting performance in Malaysian secondary schools using locally relevant datasets. Therefore, further exploration in this niche could benefit this geographical context. Furthermore, our review identified the need to investigate other variables that may impact student performance. There is vast potential for unearthing novel insights that could substantially enhance our understanding and predictive accuracy. Consequently, this study emphasizes the importance of continuing research in this field, integrating various datasets, geographical contexts, and potential predictors to better anticipate and address student performance across various educational scenarios.

Keywords: Student Performance, Data Mining Techniques, Thematic Review, Academic Records, Predictive Accuracy.

Introduction

Data mining has evolved as a critical method for analyzing massive amounts of information and generating valuable insights. The education industry recently began using data mining

approaches to forecast students' academic achievement (Shahiri et al., 2015). With data mining in education, educators may now make data-driven choices that can enhance student learning outcomes (Mokhtar et al., 2019). Predicting students' academic success by data mining is a burgeoning field of study that has piqued the interest of educators, academics, and policymakers alike. This topical review study aims to investigate trends in forecasting student performance using data mining. The review will examine the literature on data mining methods to predict student performance, the data sources employed in this research, and the performance criteria used to assess the prediction models.

Despite the extensive use of data mining methods in various disciplines, there is a need for more review articles in Malaysia that explicitly concentrate on data mining for predicting students' academic success. This research gap is problematic since forecasting student success is critical for educational institutions to become more data-driven in their decision-making processes. Furthermore, although data mining has been used in numerous studies to predict student performance, there is still a need for a comprehensive examination of the approach used.

Thus, this paper aims to fill this gap in the literature by conducting a systematic review of studies published from 2016 to 2022 that discuss the trends in predicting students' performance using data mining. The review will address the following research question:

What are the current trends in predicting students' performance using data mining discussed in publications from 2016 to 2022?

This review aims to shed light on the state-of-the-art techniques used in predicting students' academic performance, identify potential gaps in the methodology, and suggest future research directions. This review will benefit educators, policymakers, and researchers seeking to utilize data-driven approaches to improve students' learning outcomes.

Materials and Methods

This study's literature review methodology used Zairul's (2020) thematic analysis process; hence ATLAS.ti 9 was used to conduct the theme review. Thematic analysis is used to construct patterns and themes over thorough reading on the subject (Clarke & Braun, 2013). The next stage in utilizing data mining to predict student performance is to look for patterns and establish classifications to clarify prevailing tendencies. Several factors were used in making the literary selection. Articles that meet the following criteria will be considered: 1) published between 2016 and 2022; 2) include at least one topic-related term; and 3) center on utilizing data mining to predict student performance.

Table 1.

Search strings from Mendeley and Science Direct

Mendeley	“predict\$ student performance” AND “data mining” year:[2016 TO 2022]	71 results
Science Direct	“predict\$ student performance” AND “data mining”	110 results

Mendeley and Science Direct were used to conduct the literature search. 71 (Mendeley) and 110 (Science Direct) items were found in the initial search (Table 1). However, 128 publications were disregarded because they did not discuss students' performance using data mining or had premature results and anecdotes. It was also discovered that several of the articles were overlapped, had broken links, or were either incomplete or unavailable in their

entirety. As a result, 53 articles were left in the final manuscript for evaluation (see Figure 1). The articles were examined based on the publication year and the discussion style.

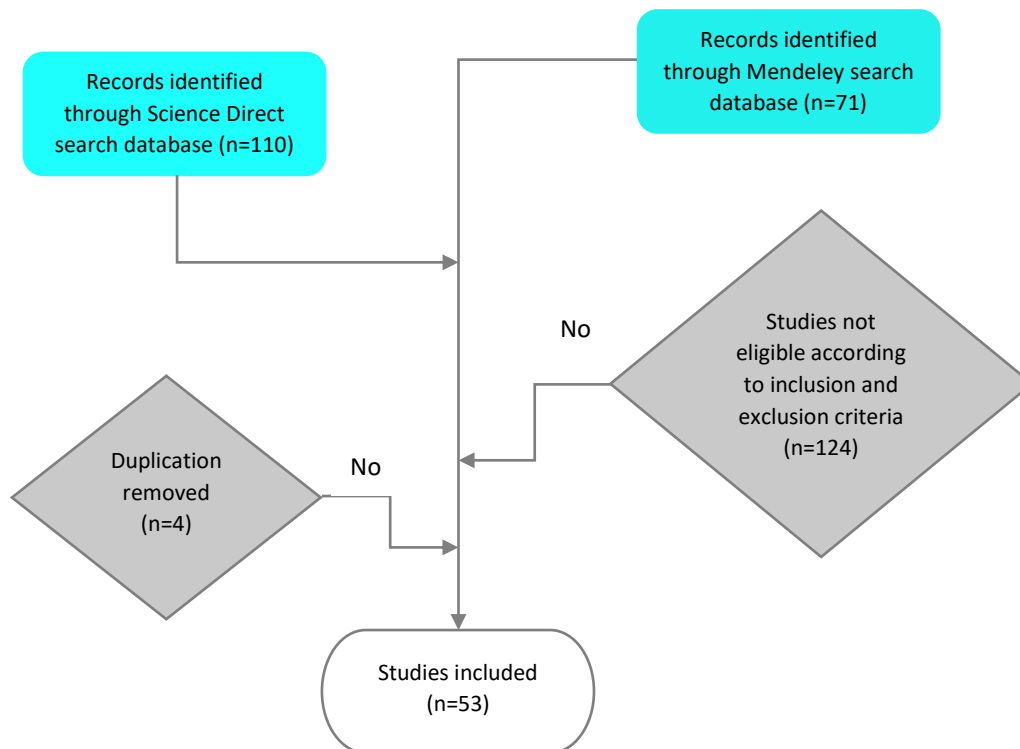


Figure 1. Inclusion and exclusion criteria in the thematic review

Results and Discussions

The study's findings are presented in two forms: quantitative and qualitative. The quantitative data provides a descriptive account of the results, emphasizing numerical values. In contrast, the qualitative data focuses on identifying and describing themes that emerge from the analysis of each article. The qualitative data provides an in-depth understanding of the data, highlighting essential insights that may not be captured through quantitative analysis alone. Additionally, this section will provide a summary of the key findings, drawing attention to noteworthy highlights that can be used to guide future research. Overall, this comprehensive presentation of the data in both forms will provide readers with a robust and nuanced understanding of the study's results.

Quantitative Data

Figure 2 shows the total number of publications published within the given period. The ascending line graph indicates that the number of research in this field is progressing in the correct direction. The most considerable number of articles were published in 2021, a total of fourteen. Ten articles were published in both 2018 and 2019, hence the overall number of publications remained the same. In the following year, there was a modest decline in published works, with just seven appearing in print. In 2016 and 2017, four and six papers were published, respectively. As of August 2022, there have been just two articles published. Table 2 displays the split of publications by authors and years.



Figure 2. Number of the paper published from 2016 to 2022

Table 2.

Publication breakdown according to the year

Author(s)	2016	2017	2018	2019	2020	2021	2022
Abdulazeez & Abdulwahab			/				
Akar & Altun		/					
Akçapınar et al.				/			
Alturki et al.						/	
Aluko et al.			/				
Amrieh et al.	/						
Asif et al.		/					
Badr et al.	/						
Basu et al.				/			
Begum & Padmannavar							/
Bernacki et al.					/		
Božić						/	
Burgos et al.		/					
Chango et al.						/	
Chaudhari et al.		/					
Costa et al.		/					
Dabhade et al.						/	
Delen et al.				/			
N. Roslan et al.						/	
Ezz & Elshenawy				/			
F. Gontzis et al.				/			
Fernandes et al.			/				
Galvan	/						
Gao et al.							/
Sergis et al.					/		
Sandoval et al.						/	
Guabassi et al.						/	

Helal et al.	/		
Hung et al.		/	
Injadat et al.		/	
Injadat et al.		/	
Kabathova & Drlik			/
Keržič et al.	/		
Kishan Das Menon & Janardhan		/	
Kostopoulos et al.	/		
Kostopoulos et al.		/	
Maheswari et al.			/
Mai et al., 2022		/	
Malini & Kalpana			/
Merchán Rubiano et al.	/		
Miguéis et al.	/		
Nabil et al.			/
Garg	/		
Oztekin	/		
Palacios et al.			/
Pallathadka et al.			/
Qazdar et al.		/	
Ramaswami et al., 201		/	
Riestra-González et al.			/
Sandoval et al.	/		
Tegegne & Alemu	/		
Waheed et al.		/	
Yusuf & John		/	

The articles focus on the countries depicted in Figures 3 and 4 since those countries are the ones for whom the datasets being examined were produced. The United States of America is in second place behind India in terms of the number of datasets cited in published studies, with four compared to India's total of eight. Since Greece, Nigeria, and Spain each have a total of three articles, these countries are tied for third position. Several nations' constitutions, including Brazil, Canada, Chile, Egypt, Jordan, Mexico, Pakistan, Portugal, Saudi Arabia, and Turkey, each have two articles. Other nations' constitutions contain less than two articles. Only one research that uses datasets from each of the following countries—Australia, Bosnia, China, Columbia, Ethiopia, Malaysia, Morocco, New Zealand, Slovakia, Slovenia, Taiwan, and Vietnam—has been published.

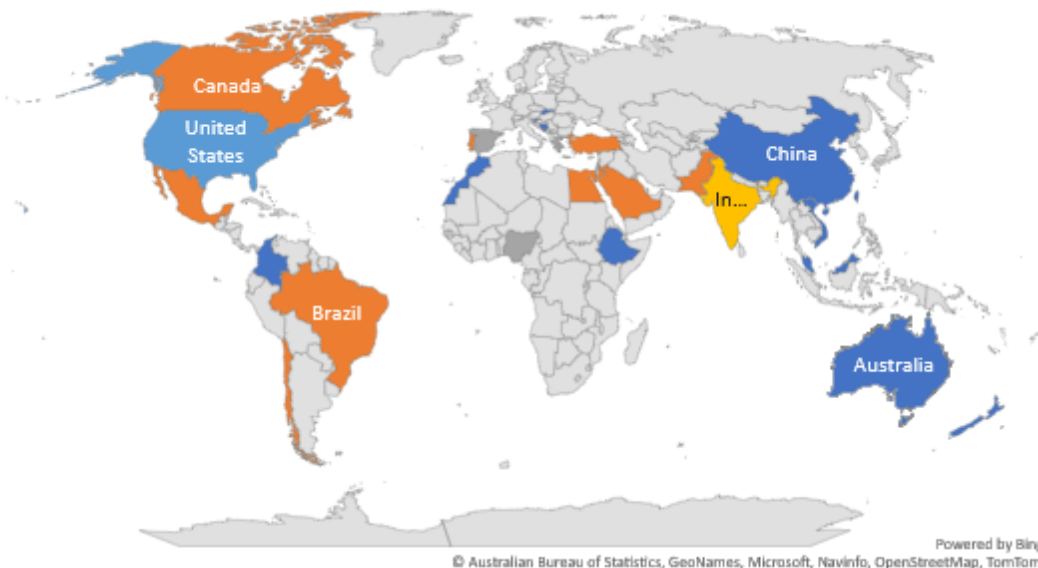


Figure 3. World map of datasets in the article

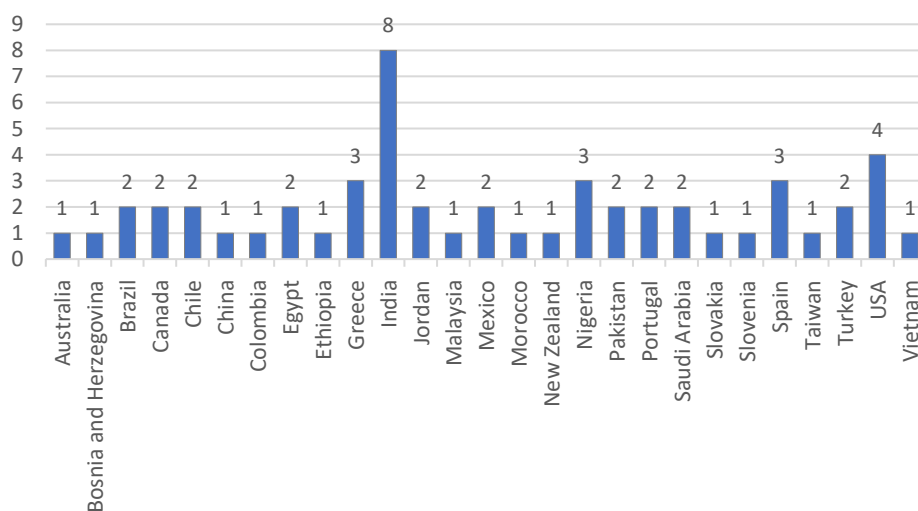


Figure 4. The country of datasets in the article

There were three main aspects to be discussed: the context of the studies of previous literature review, the factors that affect students’ performance and the data mining algorithms used to predict students’ performance. Looking at Figure 5, it seems that the majority of the research – a total of forty-six – is centered on educational establishments of higher education. The following five articles focused on secondary schools, while the final article focused on elementary or primary education.

The results indicate that most studies involving student performance prediction utilized datasets from higher education institutions. In their comprehensive literature analysis, Roslan and Chen (2022) revealed that of the fifty-eight publications analyzed, forty-nine included studies in higher education institutions, eight in secondary schools, and one in elementary schools. According to our findings, forty-six studies covered datasets from higher education institutions, five from secondary schools, and one from elementary schools. In addition, the systematic literature review conducted by Roslan and Chen (2022) revealed that no study involving secondary schools in Malaysia utilized datasets from Malaysia. The same applies to

our findings since the Malaysian research included higher education. Predicting the performance of pupils in Malaysian secondary schools is therefore a necessary objective of this research.

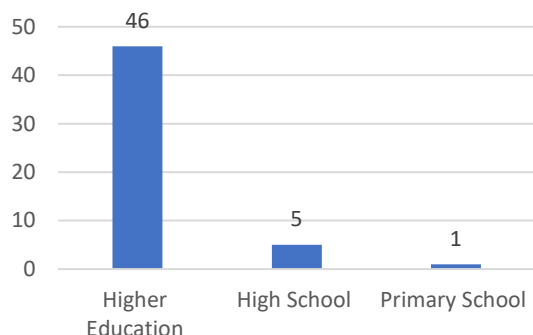


Figure 5. Study context of the article

Qualitative Data

This study extends the systematic literature reviews reported by Shahiri et al. (2015) and Abu Saa et al. (2019). Abu Saa et al. (2019) studied and assessed publications published between 2009 and 2018, while Shahiri et al. (2015) examined research published between 2002 and 2015. This research also complemented Roslan and Chen's thorough literature evaluation (2022). Roslan and Chen investigated articles published between 2015 and 2021 in Lens and Scopus, while our analysis analyzed articles published between 2016 and 2022 in Mendeley and Science Direct. This study's findings were consistent with the majority of the literature evaluated.

Factors Affecting Students' Performance

Figure 6 is a compilation of the factors that influence the overall performance of the students. The student's academic record is the factor that is mentioned the most often as a means of forecasting the student's performance. It accounts for 49% of all characteristics that were compiled. The portion that is published the second most often is the student demographics. The part that is reported the most is the student's academic record. These two facets are responsible for 81% of the total number of elements that have been gathered. A limited number of researchers focused their attention on a variety of ancillary factors, such as student settings (which were only investigated in 3% of the studies), course features (which were analyzed in 5% of the studies), and other elements (11% of the studies). The sections of the paper are broken down in Table 3 according to who wrote them and the variables that affected the students' achievement.

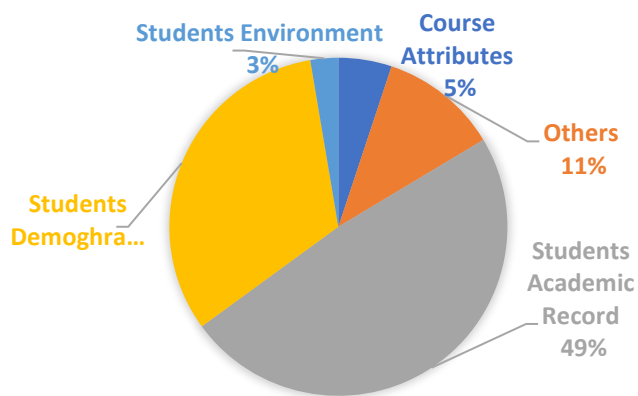


Figure 6. Factors affecting students' performance discussed in the article

Table 3.

Paper breakdown according to authors and factors affecting students' performance

Author(s)	Course Attributes	Students Academic Record	Students Demographic	Students Environment	Others
Abdulazeez & Abdulwahab		5	4		2
Akar & Altun		1	1	1	3
Akçapınar et al.		3			
Alturki et al.	1	6		1	
Aluko et al.		4			
Amrieh et al.		6	4		6
Asif et al.		2			
Badr et al.		1			
Basu et al.		12	9		
Begum & Padmannavar					
Bernacki et al.		1			
Božić					
Burgos et al.		1			
Chango et al.		2			
Chaudhari et al.		2		1	12
Costa et al.	1	8	5	1	
Dabhade et al.		8	6		4
Delen et al.		4	12		
N. Roslan et al.	1	4	7		
Ezz & Elshenawy		3			
F. Gontzis et al.	8	1	1		
Fernandes et al.		4	6	6	
Galvan	11	2			

Gao et al.	1		
Sergis et al.	1	2	2
Sandoval et al.	1		
Guabassi et al.	4	3	4
Helal et al.	3	3	
Hung et al.	5		1
Injadat et al.	1		
Injadat et al.	9		
Kabathova & Drlik	5		
Keržič et al.	2		
Kishan Das Menon & Janardhan	1		
Kostopoulos et al. 1	1	7	
Kostopoulos et al.	3	1	2
Maheswari et al.		5	8
Mai et al., 2022	4		
Malini & Kalpana	8	21	
Merchán Rubiano et al.		16	
Miguéis et al.	13	6	1
Nabil et al.	1		
Garg	6	6	1
Oztekin	18	9	
Palacios et al.	7	8	5
Pallathadka et al.	1		
Qazdar et al.	14	1	
Ramaswami et al., 201	2		
Riestra-González et al.	1		
Sandoval et al.	12	3	
Tegege & Alemu	5		
Waheed et al.			2
Yusuf & John	1		

The results indicate that student academic records and demographics are the most commonly reported elements in predicting student success. This finding is consistent with previous research that has shown academic records to predict student success strongly. Figure 6 illustrates that a student's academic record (GPA, past results, test scores, grades, marks, and attendance) is the most relevant element in determining their academic achievement. This is consistent with the findings of Shahiri et al. (2015) and Roslan and Chen (2022), who reported that nearly one-third of previous studies employed academic records such as CGPA. In a

similar line, Abu Saa et al. (2019) discovered, based on a survey of 36 research articles published between 2009 and 2018, that students' prior grades and internal ratings are the most prevalent predictors of student achievement. Asif et al. (2014) discovered that it is feasible to predict student performance only based on academic results, regardless of other determinants.

Students' demographics (gender, age, place of birth, sex, nationality, ethnic background, country of origin, marital status, parent's income, neighborhood, number of children, parents' occupation, parents' education level, and school location) are the second most influential factor affecting student performance in the results. Similarly, the results of Shahiri et al. (2015), Shingari & Kumar (2018), Hussain et al. (2018), Abu Saa et al. (2019), and Khanna et al. (2017) indicate that demographic factors are commonly included in the process of forecasting student performance.

However, the results also indicate that researchers have yet to look at miscellaneous aspects such as student environments, course attributes, and other elements as frequently. This may be due to the difficulty in measuring and quantifying these aspects or a need for more focus on these factors in previous research.

Overall, the findings reported in this study highlight the importance of considering both student academic records and demographics in predicting student success. However, it also suggests the need for more research into the impact of other factors, such as student environments and course attributes, on students.

Data Mining Algorithms

The results presented in this study reveal the most commonly used data mining algorithms in the literature, as depicted in Figure 7. The algorithms were categorized into five types: trees, functions, Bayes, rules, and lazy algorithms.

The findings suggest that the most commonly used data mining algorithms were trees and functions, each accounting for 36% of the studies, which comprise 72% of all the studies employed. Decision trees are a powerful tool for classification and prediction tasks, making them popular for data mining applications. Decision trees are prevalent in data mining for several reasons. First, decision trees are easy to interpret and visualize, which makes them helpful in communicating insights to non-technical stakeholders. Second, decision trees can handle numerical and categorical data, making them versatile. Third, decision trees can be used for classification and regression tasks, making them applicable to various problems (Hastie et al., 2009; Han et al., 2011; Witten et al., 2016). Conversely, functions are popular in data mining because they allow analysts to model complex relationships between variables and make predictions about new data. Functions are mathematical representations that describe how one variable (the dependent variable) is related to one or more other variables (the independent variables) (Hastie et al., 2009; Han et al., 2011; Witten et al., 2016).

The second most commonly used data mining algorithm was the Bayes algorithm, employed in only 14% of the studies. One of the main advantages of Bayesian methods is that they can handle complex models with many parameters. Unlike traditional statistical methods that often assume that the data follow a specific distribution, Bayesian methods allow analysts to specify a prior distribution for each parameter and then update these priors based on the observed data. This flexibility makes Bayesian methods useful for a wide range of applications in data mining, such as clustering, classification, and regression (Gelman et al., 2014; Koller et al., 2009; Murphy et al., 2012).

The lazy algorithm was the most commonly used in 10% of the studies. Lazy learning methods, also known as instance-based methods, are not as popular in data mining as eager learning methods because of their computational inefficiency and lack of generalization. Lazy learning methods do not build a model based on the entire training data but instead store the training data and make predictions based on the most similar instances to the new data. This approach can result in accurate predictions for small datasets but can be computationally expensive for large datasets because it requires comparing the new data to every instance in the training data (Aha et al., 1991; Hastie et al., 2009).

Finally, the least commonly used algorithm was rules, which were only employed in 4% of the studies. Rules-based methods involve extracting rules from the data that describe relationships between variables. These rules can be used to make predictions about new data or to identify patterns in the data. However, rules-based methods can be prone to overfitting, meaning they may perform well on the training data but poorly on new data (Han et al., 2011; Witten et al., 2016; Tan et al., 2019).

Table 3 shows the breakdown of papers according to authors and data mining algorithms. This table provides a more detailed analysis of the data mining algorithms employed in each study, along with the authors of the papers.

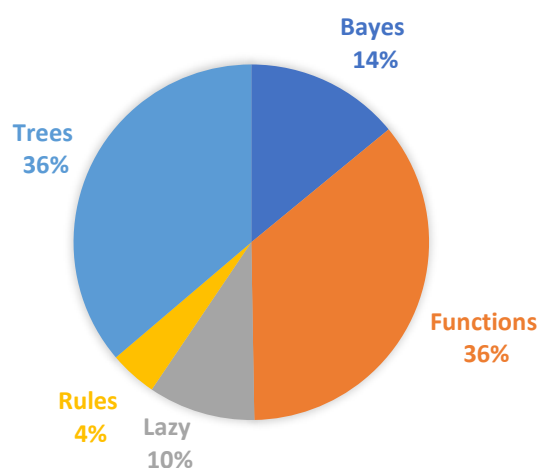


Figure 7. Data mining algorithms discussed in the article

Table 3.

Paper Breakdown According to the authors and data mining algorithms discussed in the article

Author(s)	Bayes	Functions	Lazy	Rules	Trees
Abdulazeez & Abdulwahab		1	1		1
Akar & Altun					2
Akçapınar et al.	1	2	1	1	2
Alturki et al.	1				5
Aluko et al.		2			
Amrieh et al.	1	1			1
Asif et al.	1	1	1	1	2

Badr et al.				1	
Basu et al.	1	2	1		2
Begum & Padmannavar					
Bernacki et al.	1	1			2
Božić	1		1		2
Burgos et al.		1			
Chango et al.				3	3
Chaudhari et al.	1				2
Costa et al.					
Dabhade et al.		2			
Delen et al.	1				
N. Roslan et al.					
Ezz & Elshenawy		2	1		
F. Gontzis et al.	2	2	1	1	2
Fernandes et al.					1
Galvan	1				
Gao et al.		1			
Sergis et al.		2			2
Sandoval et al.		1			
Guabassi et al.		2			2
Helal et al.	1	1		1	1
Hung et al.		1			2
Injadat et al.	1	3	1		1
Injadat et al.	1	3	1		1
Kabathova & Drlik	1	3			2
Keržič et al.	1	2	1		1
Kishan Das Menon & Janardhan	1	2	1		1
Kostopoulos et al.	1	1	2		2
Kostopoulos et al.		2	1		3
Maheswari et al.		2			2
Mai et al., 2022					
Malini & Kalpana		2			
Merchán Rubiano et al.	2				2
Miguéis et al.	1	1			4
Nabil et al.		3	1		1
Garg					3
Oztekin		3			1
Palacios et al.	1	2	1		2
Pallathadka et al.	1	1			2
Qazdar et al.		1			

Ramaswami et al., 201	1	1	1	1
Riestra-González et al.	1	3		1
Sandoval et al.		2		1
Tegegne & Alemu				1
Waheed et al.		3		
Yusuf & John		1	1	1

Conclusion and Future Studies

The study included quantitative and qualitative data to shed light on studies on forecasting student performance using data mining methods. According to the quantitative statistics, the number of published publications in this sector has increased over time, with the bulk of research focused on higher education institutions. According to the qualitative data, the student's academic record is the most generally mentioned factor influencing student achievement.

The outcomes of the study have implications for future research. One suggestion is to concentrate more on forecasting secondary school student performance, given that there is a need for more research in this area utilizing Malaysian datasets. In addition, while only one investigation has been undertaken in this area, future studies might use data mining methods to forecast student performance in elementary schools.

Another suggestion is to investigate other aspects that may impact student performance, such as student settings, course features, and other components. Further study might examine the links between these variables and student performance and how they can be incorporated into data mining systems to increase forecast accuracy.

Future research might also evaluate various data mining techniques to see which are the most successful in predicting student achievement. This might include running several algorithms on the same dataset and evaluating performance indicators like accuracy, precision, recall, and F1-score.

The research offers valuable insights by reviewing the literature on student performance prediction between 2016 and 2022. It identifies the growing interest in this field, particularly in 2021, with a primary focus on higher education. By reiterating the importance of academic records and demographic data as key factors in academic success, it strengthens existing knowledge. The study's methodology, which includes data from Mendeley and Science Direct databases, broadens the scope of the review.

Additionally, the research uncovers a significant research gap in predicting student performance in Malaysian secondary schools, emphasizing the need for context-specific investigations. It also encourages exploring new variables that can influence student performance, potentially leading to more accurate predictions. Overall, the research underscores the importance of ongoing research in this area, promoting the integration of diverse data sources, consideration of various contexts, and the exploration of a wider range of predictors to enhance our understanding of and support for student success in different educational settings.

Overall, the report presents a complete summary of studies on data mining strategies for forecasting student success. The results emphasize the need to consider many aspects that may impact student performance and investigate new areas of study, such as forecasting student success in secondary and elementary schools.

References

- Abdulazeez, Y., & Abdulwahab, L. (2019). Application of classification models to predict students' academic performance using classifiers ensemble and synthetic minority over sampling techniques. *Bayero Journal of Pure and Applied Sciences*, 11(2), 142. <https://doi.org/10.4314/bajopas.v11i2.17>
- Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, 6(1), 37-66.
- Akar, S. G. M., & Altun, A. (2020). Individual Differences in Learning Computer Programming: A Social Cognitive Approach. *Contemporary Educational Technology*, 8(3), 195–213. <https://doi.org/10.30935/cedtech/6196>
- Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop an early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, 16(1). <https://doi.org/10.1186/s41239-019-0172-z>
- Alturki, S., Alturki, N., & Stuckenschmidt, H. (2021). Using Educational Data Mining To Predict Students' Academic Performance For Applying Early Interventions. *Journal of Information Technology Education: Innovations in Practice*, 20, 121–137. <https://doi.org/10.28945/4835>
- Aluko, R. O., Daniel, E. I., Shamsideen Oshodi, O., Aigbavboa, C. O., & Abisuga, A. O. (2018). Towards reliable prediction of academic performance of architecture students using data mining techniques. *Journal of Engineering, Design and Technology*, 16(3), 385–397. <https://doi.org/10.1108/JEDT-08-2017-0081>
- Amrieh, A., Hamtini, T., & Aljarah, I. (2016). Mining Educational Data to Predict Student's academic Performance using Ensemble Methods. 9(8), 119–136.
- Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers and Education*, 113, 177–194. <https://doi.org/10.1016/j.compedu.2017.05.007>
- Badr, G., Algobail, A., Almutairi, H., & Almutery, M. (2016). Predicting Students' Performance in University Courses: A Case Study and Tool in KSU Mathematics Department. *Procedia Computer Science*, 82, 80–89. <https://doi.org/10.1016/j.procs.2016.04.012>
- Basu, K., Basu, T., Buckmire, R., & Lal, N. (2019). Predictive models of student college commitment decisions using machine learning. *Data*, 4(2), 1–18. <https://doi.org/10.3390/data4020065>
- Begum, S., & Padmannavar, S. S. (2022). Genetically Optimized Ensemble Classifiers for Multiclass Student Performance Prediction. *International Journal of Intelligent Engineering and Systems*, 15(2), 316–328. <https://doi.org/10.22266/ijies2022.0430.29>
- Bernacki, M. L., Chavez, M. M., & Uesbeck, P. M. (2020). Predicting achievement and providing support before STEM majors begin to fail. *Computers and Education*, 158(September 2019), 103999. <https://doi.org/10.1016/j.compedu.2020.103999>
- Božić, R. (2022). Predicting Study Programme Selection with Data Mining Classification Technique. *Ijeec - International Journal of Electrical Engineering and Computing*, 5(2), 94–101. <https://doi.org/10.7251/ijeec2102094b>
- Burgos, C., Campanario, M. L., Peña, D. de la, Lara, J. A., Lizcano, D., & Martínez, M. A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers and Electrical Engineering*, 66, 541–556. <https://doi.org/10.1016/j.compeleceng.2017.03.005>
- Chango, W., Cerezo, R., & Romero, C. (2021). Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses. *Computers and*

- Electrical Engineering, 89(October 2020).
<https://doi.org/10.1016/j.compeleceng.2020.106908>
- Chaudhari, K. P., Sharma, R. A., Jha, S. S., & Bari, R. J. (2017). Student Performance Prediction System using Data Mining Approach. *Ijarccce*, 6(3), 833–839.
<https://doi.org/10.17148/ijarccce.2017.63195>
- Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, 73, 247–256. <https://doi.org/10.1016/j.chb.2017.01.047>
- Dabhade, P., Agarwal, R., Alameen, K. P., Fathima, A. T., Sridharan, R., & Gopakumar, G. (2021). Educational data mining for predicting students' academic performance using machine learning algorithms. *Materials Today: Proceedings*, 47(xxxx), 5260–5267.
<https://doi.org/10.1016/j.matpr.2021.05.646>
- Delen, D., Topuz, K., & Eryarsoy, E. (2020). Development of a Bayesian Belief Network-based DSS for predicting and understanding freshmen student attrition. *European Journal of Operational Research*, 281(3), 575–587. <https://doi.org/10.1016/j.ejor.2019.03.037>
- Ezz, M., & Elshenawy, A. (2020). Adaptive recommendation system using machine learning algorithms for predicting student's best academic program. *Education and Information Technologies*, 25(4), 2733–2746. <https://doi.org/10.1007/s10639-019-10049-7>
- F. Gontzis, A., Kotsiantis, S., T. Panagiotakopoulos, C., & Verykios, V. S. (2022). A predictive analytics framework as a countermeasure for attrition of students. *Interactive Learning Environments*, 30(3), 568–582. <https://doi.org/10.1080/10494820.2019.1674884>
- Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Erven, G. Van. (2019). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*, 94(August 2017), 335–343.
<https://doi.org/10.1016/j.jbusres.2018.02.012>
- Galvan, P. (2016). Educational Evaluation and Prediction of School Performance through Data Mining and Genetic Algorithms. *International Journal of Advanced Engineering Research and Science*, 3(10), 215–220. <https://doi.org/10.22161/ijaers/3.10.34>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis* (3rd ed.). Chapman and Hall/CRC.
- Gao, L., Zhao, Z., Li, C., Zhao, J., & Zeng, Q. (2022). Deep cognitive diagnosis model for predicting students' performance. *Future Generation Computer Systems*, 126, 252–262.
<https://doi.org/10.1016/j.future.2021.08.019>
- Garg, R. (2018). Predicting Student Performance Of Different Regions Of Punjab Using Classification Techniques. 9(1), 236–240.
- Guabassi, I. El, Bousalem, Z., Marah, R., & Qazdar, A. (2021). Comparative Analysis of Supervised Machine Learning Algorithms to Build a Predictive Model for Evaluating Students' Performance. *International Journal of Online and Biomedical Engineering*, 17(2), 90–105. <https://doi.org/10.3991/ijoe.v17i02.20025>
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques*. Elsevier.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.
- Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D. J., & Long, Q. (2018). Predicting academic performance by considering student heterogeneity. *Knowledge-Based Systems*, 161, 134–146. <https://doi.org/10.1016/j.knosys.2018.07.042>

- Hung, H. C., Liu, I. F., Liang, C. T., & Su, Y. S. (2020). Applying educational data mining to explore students' learning patterns in the flipped learning approach for coding education. *Symmetry*, 12(2), 1–14. <https://doi.org/10.3390/sym12020213>
- Injadat, M. N., Moubayed, A., Nassif, A. B., & Shami, A. (2020). Multi-split optimized bagging ensemble model selection for multi-class educational data mining. *Applied Intelligence*, 50(12), 4506–4528. <https://doi.org/10.1007/s10489-020-01776-3>
- Kabathova, J., & Drlik, M. (2021). Towards predicting student's dropout in university courses using different machine learning techniques. *Applied Sciences (Switzerland)*, 11(7). <https://doi.org/10.3390/app11073130>
- Keržič, D., Aristovnik, A., Tomažević, N., & Umek, L. (2019). Assessing the impact of students' activities in e-courses on learning outcomes: a data mining approach. *Interactive Technology and Smart Education*, 16(2), 117–129. <https://doi.org/10.1108/ITSE-09-2018-0069>
- Kishan Das Menon, H., & Janardhan, V. (2020). Machine learning approaches in education. *Materials Today: Proceedings*, 43(xxxx), 3470–3480. <https://doi.org/10.1016/j.matpr.2020.09.566>
- Koller, D., & Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.
- 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(4), 1–19. <https://doi.org/10.1142/S0218213019400013>
- Kostopoulos, G., Kotsiantis, S., Pierrakeas, C., Koutsonikos, G., & Gravvanis, G. A. (2018). Forecasting students' success in an open university. *International Journal of Learning Technology*, 13(1), 26–43. <https://doi.org/10.1504/IJLT.2018.091630>
- Maheswari, K., Priya, A., Balamurugan, A., & Ramkumar, S. (2021). Analyzing student performance factors using KNN algorithm. *Materials Today: Proceedings*, xxx. <https://doi.org/10.1016/j.matpr.2020.12.1024>
- Mai, T. T., Bezbradica, M., & Crane, M. (2022). Learning behaviours data in programming education: Community analysis and outcome prediction with cleaned data. *Future Generation Computer Systems*, 127(September 2021), 42–55. <https://doi.org/10.1016/j.future.2021.08.026>
- Malini, J., & Kalpana, Y. (2021). Investigation of factors affecting student performance evaluation using education materials data mining technique. *Materials Today: Proceedings*, 47(xxxx), 6105–6110. <https://doi.org/10.1016/j.matpr.2021.05.026>
- Merchán Rubiano, S., Beltrán Gómez, A., & Duarte García, J. (2017). Engineering Students' Academic Performance Prediction using ICFES Test Scores and Demo-graphic Data. *Ingeniería Solidaria*, 13(21), 53–61. <https://doi.org/10.16925/in.v13i21.1729>
- Miguéis, V. L., Freitas, A., Garcia, P. J. V., & Silva, A. P. (2020). Predicting Students' Performance: Comparing Traditional Predictive Models with Survival Analysis. *Springer, Cham*, 7(6), 93. <https://doi.org/10.1007/s40593-020-00219-y>
- Mohammed, J., & Majeed, A. K. (2018). Predictive Data Mining for the Performance of Iraqi Students using Decision Tree. *Journal of Theoretical and Applied Information Technology*, 96(12), 3959–3966.
- Morais, A. S., Amandi, A., & Schiaffino, S. (2018). Using neural networks for the prediction of performance of students. *Journal of Computer Science and Technology*, 18(2), 265–273. <https://doi.org/10.2424/ijcsit.2018.2.2.05>

- Moreno García, C. F., Cortes Martinez, R. H., Cortes Martinez, J. J., Rodríguez, M. del P. C., & Leyva López, J. C. (2017). Predicting academic performance of Computer Engineering Students through Artificial Neural Networks. *Research in Computing Science*, 1332(c), 13–23.
- Nananukul, N., & Tongdhamachart, T. (2019). Predicting student academic performance in introductory programming courses. *Computers and Education*, 140, 103606. <https://doi.org/10.1016/j.compedu.2019.103606>
- Naseeb, N., & Atif, Y. (2018). Student's Performance Prediction: A New Approach in the Perspective of Students. 6(2), 13–20.
- Noel, N., & Doy, A. A. (2017). Predictive data mining and modeling of student's performance in E-learning. *Education and Information Technologies*, 22(2), 599–619. <https://doi.org/10.1007/s10639-1555-0>
- Ola, O., Lasisi, R., & Eweoya, I. (2019). Predicting Academic Performance: A Comparative Analysis of Machine Learning Algorithms. *International Journal of Interactive Mobile Technologies*, 13(8), 124–132. <https://doi.org/10.3991/ijim.v13i08.10466>
- Patil, A. G., & Chopade, R. (2021). A Predictive Data Mining Model for Academic Performance Improvement of Students. 2020 2nd International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2020, 692–696. <https://doi.org/10.1109/ICASERT48347.2020.9354117>
- Paul, R. A. J., Nagarajan, K., & Eswaran, R. (2020). Analysis of University Academic Performance Prediction Using Data Mining Techniques. **Materials Today: Proceedings*, xxxx. <https://doi.org/10.1016/j.matpr.2020.12.475>
- Pinon, L., Dorrnzoro, E., & López, A. S. (2021). Feature selection for predicting academic performance in higher education through machine learning techniques. *Computer Applications in Engineering Education*, 29(6), 1392–1408. <https://doi.org/10.1002/cae.22295>
- Pratap Reddy, K., & Deo, B. S. (2017). A survey of data mining methods and models for the prediction of educational performance. *Int J Adv Res Comput Sci*, 8(9), 1411–1416.
- Praveen, S. S., & Kumar, P. V. (2018). An overview on data mining approach for education performance prediction. *Materials Today: Proceedings*, 5(1), 2678–2684. <https://doi.org/10.1016/j.matpr.2017.11.487>
- Ramírez-Corona, N., Martínez-Trinidad, J. F., & Carrasco-Ochoa, J. A. (2022). Analysis of Students' Dropout in Virtual Learning Environments through Data Mining and Machine Learning: A Survey. *Computers*, 11(2), 102. <https://doi.org/10.3390/computers11020010>
- Rojas, G., Munoz, R., Zvare, R., & Mandler, B. (2021). Predictive models for academic performance in online exams. *Computers in Human Behavior*, 118(July 2020), 106677. <https://doi.org/10.1016/j.chb.2020.106677>
- Rokhman, F. A., & Ayu, M. A. (2018). Students' Performance Prediction Using Classification Algorithms. *Jurnal Ilmiah SISFOTENIKA (SISFOTENIKA)*, 8(1), 25–34. <https://doi.org/10.24912/sisfotenika.v8i1.1903>
- Sakar, C. O., Polat, S. O., Kursun, O., & Sert, D. E. (2019). Real-time prediction of student scores: A flipped classroom experiment with international postgraduate students. *Computers and Education*, 136, 118–132. <https://doi.org/10.1016/j.compedu.2019.03.001>
- Saldaña-Duenez, E. L., López, A. J. A., Tapia, L. A., & López, G. J. (2017). Learning Analytics in E-learning environments using multi-label classification. *Journal of Computer Science and Technology*, 17(1), 69–79. <https://doi.org/10.2424/ijcsit.2017.1.1.01>

- Sathik, M. M., & Chandrakala, S. (2018). Prediction of student performance in programming courses based on data mining classification models. *Materials Today: Proceedings*, 5(1), 2533–2538. <https://doi.org/10.1016/j.matpr.2017.11.398>
- Soni, A. K., Prajapati, P., & Patel, R. (2018). Predicting student performance for e-learning courses using the multilayer perceptron neural networks. *Journal of Ambient Intelligence and Humanized Computing*, 9(6), 1897–1904. <https://doi.org/10.1007/s12652-018-0753-1>
- Sreedhar, C., & Vasista, A. K. (2021). Predicting Student Academic Performance using Supervised Learning Algorithms. *Materials Today: Proceedings*, xxxx. <https://doi.org/10.1016/j.matpr.2021.02.622>
- Sreeja, S., & Pramodh, S. (2019). Enhancing Student Performance Prediction using Feature Selection. *Journal of Critical Reviews*, 6(5), 293–298.
- Tamada, T., Ono, R., & Takahashi, S. (2020). A Clustering-Based Approach to Predicting Student Dropout and Efficiently Selecting Early Warning Indicators in Online Courses. *Computers in Human Behavior*, 104(June 2019), 106171. <https://doi.org/10.1016/j.chb.2019.106171>
- Tanveer, R., Ashfaq, A., Abid, M., Kavitha, R., & Lee, S. (2022). A feature selection technique based on evolutionary search for predicting academic performance. *Journal of King Saud University - Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2021.11.019>
- Tsai, T. H., & Tseng, T. H. (2018). A review of user's behavior and performance in e-learning and learning analytics for the perspective of a comprehensive study. *Computers and Education*, 125, 389–411. <https://doi.org/10.1016/j.compedu.2018.07.013>
- Tüzün, H., Yılmaz-Soylu, M., Karakuş, T., İlker, Y., & Kızılkaya, G. (2009). The effects of computer games on primary school students' achievement and motivation in geography learning. *Computers & Education*, 52(1), 68–77. <https://doi.org/10.1016/j.compedu.2008.06.008>
- Verma, B., & Shah, P. (2020). Predicting Academic Performance of Engineering Students Using the Decision Tree Technique. *Materials Today: Proceedings*, 23(August 2019), 68–72. <https://doi.org/10.1016/j.matpr.2019.07.185>
- Vishnu, V. S., & Sri, R. (2018). Predicting Student Performance Using Machine Learning. *Ijarccce*, 7(1), 57–60. <https://doi.org/10.17148/ijarccce.2018.72106>
- Wang, Y. H., Tseng, C. C., Chou, C. H., Chan, T. W., & Chen, G. D. (2019). The role of personal and social factors in designing an affective student modeling system. *Computers & Education*, 131(March), 15–29. <https://doi.org/10.1016/j.compedu.2018.12.003>
- Yang, D. (2021). A robust data preprocessing technique for estimating the performance of predictive modeling in big data analytics. *Computers in Human Behavior*, 114(June 2020), 106572. <https://doi.org/10.1016/j.chb.2020.106572>
- Yılmaz, R. M. (2017). Predicting student academic achievement by the use of machine learning techniques. *Neural Computing and Applications*, 28(7), 1773–1781. <https://doi.org/10.1007/s00521-017-3055-8>
- Zain, J. M., Saad, M. R., & Abdullah, A. (2017). Predicting student academic performance using classification algorithms. *Journal of Theoretical and Applied Information Technology*, 95(10), 2351–2362.