Data and Metadata. 2025; 4:730
doi: 10.56294/dm2025730

REVIEW

Predictive analytics in education: machine learning approaches and performance metrics for student success - a systematic literature review

Analítica Predictiva en la Educación: Enfoques de Aprendizaje Automático y Métricas de Desempeño para el Éxito Estudiantil - Una Revisión Sistemática de la Literatura

Shoukath T K[1], Midhun Chakkaravarthy[1]

[1]Faculty of AI Computing and Multimedia, Lincoln University College. Malaysia.

Cite as: Shoukath TK, Chakkaravarthy M. Predictive analytics in education: machine learning approaches and performance metrics for student success - a systematic literature review. Data and Metadata. 2025; 4:730. https://doi.org/10.56294/dm2025730

Submitted: 25-04-2024 **Revised:** 22-08-2024 **Accepted:** 15-03-2025 **Published:** 16-03-2025

Editor: Dr. Adrián Alejandro Vitón Castillo

Corresponding Author: Shoukath T K

ABSTRACT

Higher education institutions rely on student performance to improve grades and enhance academic outcomes. Universities face challenges in evaluating student achievement, providing high-quality instruction, and analyzing performance in a dynamic and competitive context. However, due to limited research on prediction techniques and the critical factors influencing performance, making accurate forecasts is challenging. The utilization of educational data and machine learning has the potential to improve the learning environment. Ensemble models in educational data mining enhance accuracy and robustness by combining predictions from multiple models. Approaches such as bagging and boosting effectively mitigate the risk of overfitting. Machine learning techniques, including Support Vector Machines, Random Forests, K-Nearest Neighbors, Artificial neural networks, Decision Trees, and convolutional neural networks, have been employed in performance prediction. In this study, we examined 85 papers that focused on student performance prediction using machine learning, data mining, and deep learning techniques. The thorough analysis underscores the importance of various factors in forecasting academic performance, offering valuable insights for improving educational strategies and interventions in higher education contexts.

Keywords: Student Performance; Educational Data; Machine Learning; Deep Learning; Ensemble Models.

RESUMEN

Las instituciones de educación superior dependen del rendimiento estudiantil para mejorar las calificaciones y optimizar los resultados académicos. Las universidades enfrentan desafíos en la evaluación del desempeño estudiantil, la provisión de una enseñanza de alta calidad y el análisis del rendimiento en un contexto dinámico y competitivo. Sin embargo, debido a la investigación limitada sobre técnicas de predicción y los factores críticos que influyen en el rendimiento, realizar pronósticos precisos es un desafío. La utilización de datos educativos y el aprendizaje automático tienen el potencial de mejorar el entorno de aprendizaje. Los modelos de conjunto en la minería de datos educativos mejoran la precisión y la robustez al combinar predicciones de múltiples modelos. Enfoques como bagging y boosting mitigan eficazmente el riesgo de sobreajuste. Se han empleado técnicas de aprendizaje automático, incluidas Máquinas de Vectores de Soporte, Bosques Aleatorios, K-Nearest Neighbors, Redes Neuronales Artificiales, Árboles de Decisión y Redes Neuronales Convolucionales, en la predicción del rendimiento estudiantil. En este estudio, examinamos 85 artículos centrados en la predicción del rendimiento estudiantil utilizando técnicas de aprendizaje automático, minería de datos y aprendizaje profundo. El análisis detallado destaca la importancia de diversos factores en la predicción del rendimiento académico, ofreciendo valiosos conocimientos para mejorar estrategias e intervenciones educativas en contextos de educación superior.

Palabras clave: Rendimiento Estudiantil; Datos Educativos; Aprendizaje Automático; Aprendizaje Profundo; Modelos de Conjunto.

INTRODUCTION

The most essential metric for assessing students' capabilities is their academic performance.[1] This serves as the principal criterion by which educational institutions evaluate and select students. In an increasingly competitive educational landscape, many institutions face challenges in attracting prospective students.[2] Accurate prediction of academic performance is vital for fostering student development and enhancing educational standards.[3] However, student performance is influenced by a multitude of complex factors, including socioeconomic status and prior academic achievements.[4,5] Notably, despite the availability of statistical techniques, the analysis and prediction of student performance have been relatively underexplored in academic research.[6,7]

Educational Data Mining (EDM) is an emerging field described by the educational data mining community as being concerned with developing methods to explore and analyze the large-scale data generated from educational environments. These methods aim to better understand students and the contexts in which they learn.[8] Broadly, EDM involves applying data mining techniques to educational data to derive insights from various educational systems.[9,10,11] Predicting student success within these systems is challenging due to the increasing volume of data.[12] Current methods for predicting student performance in higher education settings are often deemed insufficient, highlighting the need for more effective strategies.[13] Additionally, further research is necessary to fully understand the factors influencing student performance.[14] Therefore, there is a clear need to identify the most critical elements that significantly impact student performance.[15]

Rao, A.S., et al.[16] have demonstrated a prediction model for student placement, leveraging a data mining perspective within the framework of Outcome-Based Education (OBE). This research developed a predictive model that considers students' prior academic and extracurricular achievements to forecast their placement into different categories, such as super dream companies, dream companies, and mass recruiter companies. The study also provided real-time experimental results and findings, along with performance indicators used to validate the model, aiming to assist educational institutions in achieving the milestones set by OBE.

Livieris, I.E., et al.[17] proposed a method for predicting secondary school students' performance using semi-supervised machine learning. Their study evaluates the effectiveness of two wrapper techniques for semi-supervised learning algorithms in forecasting students' final exam results. Additionally, the research includes an analysis of student interactions within a Learning Management System (LMS). Khakata, E., et al.[18] developed a stochastic modeling approach for predicting student performance in an internet-mediated environment. Their model, which uses Stochastic Differential Equations (SDE), emphasizes the role of student effort, investment costs, and strategy efficacy in achieving good performance. The study analyzes data from various higher education respondents to generate a performance trajectory for students. Abazeed, A., et al.[19] proposed a model for classifying and predicting student performance at the university level. Their approach uses a classification algorithm to extract hidden patterns from student records and develop a prediction model. This model helps in selecting appropriate learning paths, identifying students who require additional support, and pinpointing factors that could negatively impact performance, thus aiding in preventing potential failures.

Anuradha, C., et al.[20] conducted a comparative analysis of classification algorithms for predicting students' performance in end-of-semester exams at the university level. Their research highlights the varying accuracy of different classification approaches, noting that predictions for distinction-class students were less accurate compared to first-class students. They emphasize the need for further research using diverse data mining techniques to improve accuracy. Wang, et al.[21] developed a machine learning-based method for providing personalized feedback in an online learning environment. Their model incorporates fine-tuning and pre-training phases to enhance the efficiency of online learning through personalized feedback. The study involved a quasi-experimental setup with 62 participants 29 in the control group and 33 in the experimental group) to evaluate the model's validity and the impact of personalized feedback on cognitive load and learning outcomes.

The purpose of this paper is to conduct a literature review aimed at predicting student performance and identifying the most significant and well-researched factors that influence student success in higher education. The paper develops a comprehensive list of variables and characteristics believed to impact learning outcomes and student performance. To carry out this investigation, the existing body of literature has been thoroughly examined.

The key conceptual insights discussed in the paper include:

- Importance of Academic Performance and how it helps to make decisions regarding admissions, placements, and interventions also, to provide better support and guidance to students.
- Prior to student demographics, academic achievement, psychological traits, e-learning activity, and student environments are the major factors influencing student performance.

- The Machine learning techniques of k-nearest neighbors (KNN), support vector machines (SVM), random forest (RF) and decision trees (DT) which are mostly used to identify associations and patterns of data for predicting student performance.
- Course grades, exam/test scores, grade range/pass-fail, program graduation/retention/dropout rates, and GPA/CGPA are the metrices used to describe student performance, which provide insights into students' academic progress and outcomes.
- Hybrid and ensemble approaches, addressing class imbalance, and considering more complex performance metrics are the areas suggested for future research.

The structure of the article is as follows: The next section explains the data mining cycle, followed by a presentation of the discussions and results, and concluding with the final remarks.

Data Mining Cycle

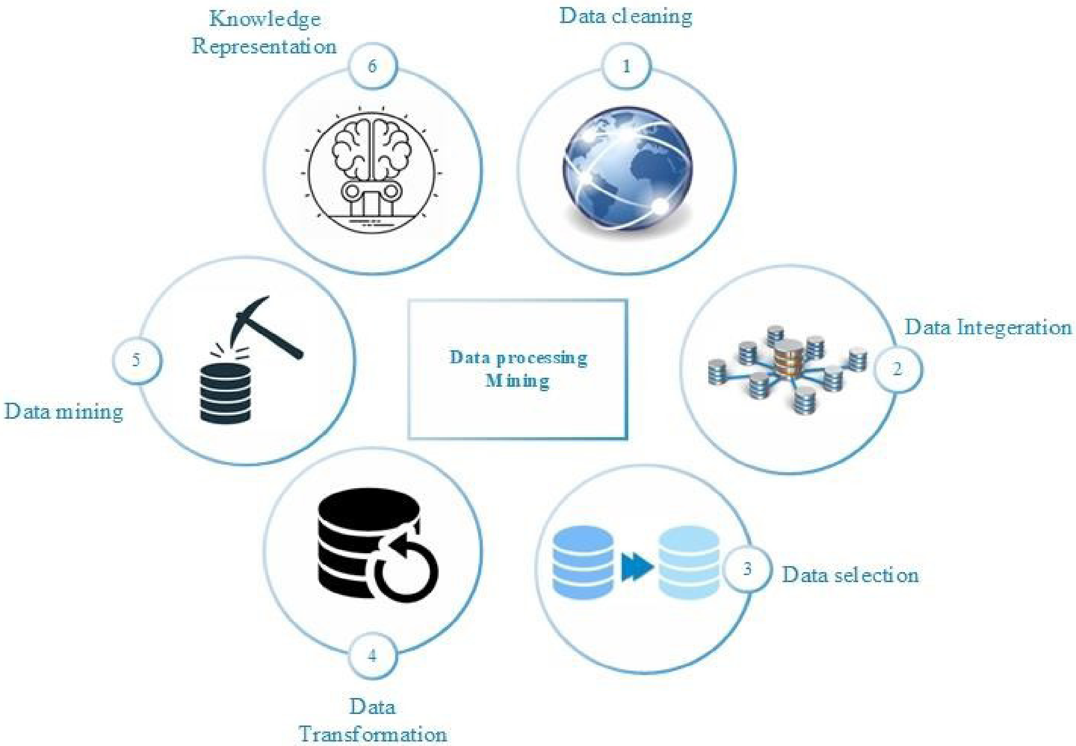

Figure 1. Data mining cycle

Figure 1 illustrates the data mining cycle, detailing how data is cleaned, integrated, selected, and transformed, and how data mining is subsequently performed.

METHOD

This study employs a systematic literature review (SLR) approach, as guided by Kitchenham et al.[22]. The approach consists of three phases: planning, conducting, and reporting. The planning and conducting phases are discussed in the subsections below, while the reporting phase is addressed in the results & discussions.

Planning phase

This phase includes determining the research questions, identifying search keywords, selecting sources, establishing inclusion and exclusion criteria, and formulating the data extraction strategy.

Research Questions

The research questions help maintain the focus of the review. They were designed using the Intervention, Population, Context (PIOC), and Outcomes criteria provided by Kitchenham et al.[22]. The PIOC criteria are shown in table 1.

The research questions addressed by this SLR are:

Q1. What are the features used for prediction – focus EDM studies?

Q2. What is the Machine-learning techniques used for student performance prediction?

Q3. Which metrics type were used for describing students' performance?

Table 1. PIOC	
Population	Educational institution (Student Performance)
Intervention	Machine learning Techniques /Methods for prediction
Outcomes	Features used for prediction, ML techniques/methods used for prediction, Metrics used for prediction
Context	Academic institutions

Search Term Used

The search terms were derived from the research questions outlined in subsections. The search string used in this study is: ("Data Mining in Education" OR "Data Mining" OR "Machine Learning" OR "Deep Learning") AND ("Student Performance Prediction" OR "Academic Performance Prediction") AND (Features OR Factors OR Metrics OR Techniques OR Methods OR Approaches).

Sources of Data

The objectives of this review were to conduct a comprehensive systematic literature review. To find primary data and relevant papers, five research databases were utilized: IEEE Xplore, SpringerLink, ScienceDirect, ACM Digital Library, and Scopus. Detailed investigations of these repositories were performed using various queries related to student performance prediction using machine learning techniques, covering the period from 2011 to 2024. Many research papers were retrieved through the pre-determined queries and were subsequently filtered to retain only the most relevant publications for this review.

Criteria for Exclusion and Inclusion

This study involves retrieving all papers from the search results that meet the criteria specified in table 2 below.

Table 2. Exclusion and Inclusion Criteria

Inclusion	Exclusion
Studies Related to Student Performance Prediction	Studies Not Related to Student Performance
Research papers published and accepted by journals or conferences that use blind peer review	Papers that are not experimentally conducted or do not propose validation methods.
Papers published from 2011 to 2024	Patents, business posters, editorials, short papers, conducted reviews, Wikipedia articles, technical reports, survey studies.
Papers written in English	Papers written in non-English languages.

Strategy of Data Extraction

Based on the research questions, a data extraction strategy is developed in this phase. Table 3 presents the high-level taxonomy.

Table 3. Data Layout

Artefact	Description
Features	Data features used for performance prediction include demographic data, previous academic achievements, and similar factors.
ML Techniques/Methods	Machine learning techniques/methods used for prediction encompass supervised and unsupervised learning, data mining, feature extraction, and statistical techniques.
Prediction Metrics	Details and descriptions of what is being predicted include pass/fail outcomes in annual exams, grades or grade ranges in courses, and dropout or graduation rates in programs.

Conducting the Review

This phase begins by searching the data sources using the search terms mentioned above. The initial results are listed in table 4.

Table 4. Data Sources with initial results		
Identifiers	**Databases**	**Results**
1.	Digital Library of IEEE Xplore	34
2.	Springer Link	29
3.	Science Direct	27
4.	Digital Library of ACM	19
5.	Scopus	38

Then, the abstract of each article was examined for inclusion, and only relevant articles were considered. After evaluation, 13 articles were removed due to duplication, and 19 articles were excluded for not meeting the inclusion criteria. Additionally, 23 articles were excluded for not being related to the scope of this study. Finally, 92 papers were selected for consideration. The number of articles selected each year is tabulated in table 5.

Table 5. No. of articles from 2011 to 2024		
Publication Year	**No of article**	**Article references**
2011	1	(30)
2012	2	(36,60)
2013	3	(40,72,49)
2014	1	(35)
2015	4	(55,20,23,26)
2016	1	(67)
2017	11	(33,34,48,51,52,56,58,73,19 ,27,28)
2018	3	(41,64,6)
2019	9	(62,76,1,7,9,15,16,17,25)
2020	34	(2,3,5,11,13,14,18,24,31,38,39, 42,43,47,49,50,53) (57,59,61,63,68,69,74,75,77,79, 80,81,82,83,85,86,87)
2021	9	(4,8,12,33,45,66,67,71,85)
2022	7	(38,46,47,55,10,21,28)
2023	4	(89,90,91,92)
2024	3	(93,94,95)

Figure 2 demonstrates the number of publications cited in databases per year and the publication ratio for the specified years.

Figure 2. Published articles during 2011-2024

https://doi.org/10.56294/dm2025730

RESULTS AND DISCUSSION

Factors influenced in predicting academic performance.

Soares, D.L., et al.[23] conducted a study to examine the relationship between intelligence and academic success as students' progress through their education. In this three-year longitudinal study, 284 Portuguese middle school students took reasoning tests with three subtests (numerical, verbal, and abstract) at the end of the seventh grade to assess intelligence. Academic grades were collected at the same time (prior academic achievement, AA7) and again at the end of the ninth grade (final academic achievement, AA9). The primary findings indicated that AA9 was a better predictor of intelligence when the mediating effect of AA7 was considered.

Cutumisu, M., et al.[24] studied the relationship between academic achievement and the spontaneous use of design thinking strategies. The study examined how students' learning during a 10–15-minute online activity correlated with their prior academic performance and their decisions to apply design-thinking tactics, such as seeking feedback and improving their work. Sixth-grade students created three digital posters, and after each poster, they had the option to request either affirmative or critical feedback before deciding whether to revise their poster. The findings indicated a positive correlation between seeking critical feedback and prior academic performance.

Cutumisu, M.[25] explored the relationship between prior academic performance and university achievements among law students. The study examined the National Matriculation Examination performance of 426 law students in relation to their subsequent academic success in higher education, using quantitative methods. The results demonstrated that prior performance was associated with both study success and academic advancement. Additionally, the findings indicated that law students who earned higher grades in mathematics courses completed their university studies more quickly and successfully.

Hettler, P.L.[26] proposed that demographics of students and the impact of team-based learning were explored. The findings show that there is no statistically significant difference based on gender.

Pincus, K.V., et al.[27] explored the impact of technological and financial factors on the landscape of higher education. The study considered student demographics, student debt levels, and public funding as financial factors, and task automation, skills, and competency training as technological factors. The study concluded that these factors have not yet significantly altered either the curriculum (what is taught) or pedagogy (how education is delivered).

Using a flipped classroom (FC) model, Yilmaz, R.[28] investigated the impact of e-learning readiness on student motivation and satisfaction. The study aimed to assess how students' readiness for e-learning influenced their satisfaction and motivation within the FC instructional context. The study involved 236 undergraduate students enrolled in a Computing I course using the FC model. Data were collected using three self-report tools: the E-learning Readiness Scale, the Satisfaction Scale, and the Motivated Strategies for Learning Questionnaire.

Oztekin, A.[29] developed a technique to evaluate the usability of e-learning systems using machine learning methods. The study proposed a new machine learning-based evaluation method for assessing system usage in e-learning. It combined decision trees, support vector machines, and neural networks with multiple linear regression to create predictive models. The proposed model revealed relationships underlying the overall usability of an e-learning system and its predictor elements. The rank-order relevance of the predictors was assessed using sensitivity analysis. A statistical measure called the severity index was developed using both usability scores and sensitivity values.

Klanja-Milievi, A., et al.[30] proposed a hybrid recommendation technique for personalizing and identifying learning styles. The study outlined the recommendation module of Protus, a programming system that dynamically adjusts to students' interests and skill levels. By analyzing learners' server logs and learning style assessments, the system identifies patterns in learning behaviors and processes clusters according to different learning methods. It then uses a prior method to mine frequently occurring sequences, evaluating learning styles and student interests to offer personalized educational material recommendations.

Tomasevic, N., et al.[31] developed a supervised data mining algorithm for predicting student performance and identifying those at high risk of dropping out. The model utilized state-of-the-art supervised machine learning algorithms for classification and regression tasks, achieving the highest overall precision with artificial neural networks. The study revealed that prior performance data and demographic engagement data did not significantly affect prediction precision.

Mubarak, A.A., et al.[32] designed a deep learning model, CONV-LSTM, to predict student dropout rates in massive open online courses (MOOCs). The model combines convolutional neural networks and long short-term memory to automatically extract features from MOOC logs and forecast whether students will complete or drop out of courses. The study addressed the issue of class imbalance, which can result in biased predictions and high false-negative rates. To improve prediction performance, the loss function employed a cost- sensitive approach that considered the different costs of misclassifications for false positives and false negatives.

Asif, R., et al.[33] examined undergraduate student performance using data mining techniques. The study focused on two areas: predicting academic performance after a four-year program and analyzing progression patterns. The findings identified two crucial student groups: high and low achievers. The study suggested providing possible warnings and support for low performers, as well as advice and opportunities for high achievers.

Costa et al.[34] proposed an evaluation of the effectiveness of techniques in mining for the early prediction of student failures. However, this paper also compares the effectiveness of educational data mining strategies in identifying students who are struggling in introductory courses. The study analyzes: (i) the effectiveness of data mining techniques in identifying students likely to fail at an early enough stage to allow for intervention, and the actions that can be taken to decrease the failure rate; (ii) the impact of data preprocessing and algorithm fine-tuning on improving prediction accuracy. Different machine learning techniques were applied using two distinct and independent data sources. The study demonstrated that the techniques predicted which students were most likely to fail. Moreover, it showed that some techniques, particularly those involving data preprocessing and algorithm fine-tuning, were more effective. The study revealed that the support vector machine technique significantly outperforms others in a statistically significant way.

Zwilling et al.[35] proposed a knowledge management system for higher education institutions as a solution for data mining student performance. A study comparing two different data mining methods was conducted to address the research topics using small student datasets. The findings revealed great promise, motivating higher education institutions to incorporate data mining software as a crucial component of their knowledge management systems.

Kabakchieva[36] developed a data mining classification algorithm for predicting student performance. The study demonstrated the enormous potential of data mining applications in university management, particularly in making university recruitment efforts more effective and attracting the most promising students. The aim was to create a data mining model to forecast student achievements based on pre-university academic performance, personal traits, and collegiate performance.

Feng et al.[37] proposed a study and prediction of student performance using academic educational data mining. This study analyzed and forecasted academic student performance using classification and clustering techniques, along with convolutional neural networks. The study first proposed a novel technique optimized by determining the clustering number, followed by the application of the K-means algorithm. The data was labeled for testing and training using convolutional neural networks. A model was then created to forecast future performance, which was evaluated using two metrics in cross-validation methods.

Abubakaria et al.[38] developed a neural network-based model for predicting student academic performance. Using 480 instances from a student dataset, where each student had 16 attributes, the Adam optimizer was used but resulted in a performance accuracy of less than 60 %. However, with the stochastic gradient descent optimizer, accuracy improved to over 75 %, with a stable final accuracy of 76,8 %. This suggests that the proposed neural network model may be reliable for predicting student academic performance.

The first research question of this SLR work was: (Q1) What are the features used for prediction in focused EDM studies?

Although a wide range of features have been examined in the literature regarding their influence on the prediction of academic success, this SLR revealed that prior academic achievement, student demographics, psychological traits, e-learning activity, and student environments are the most frequently reported predictive features. Figure 3 shows the factors predicted to influence academic student performance.

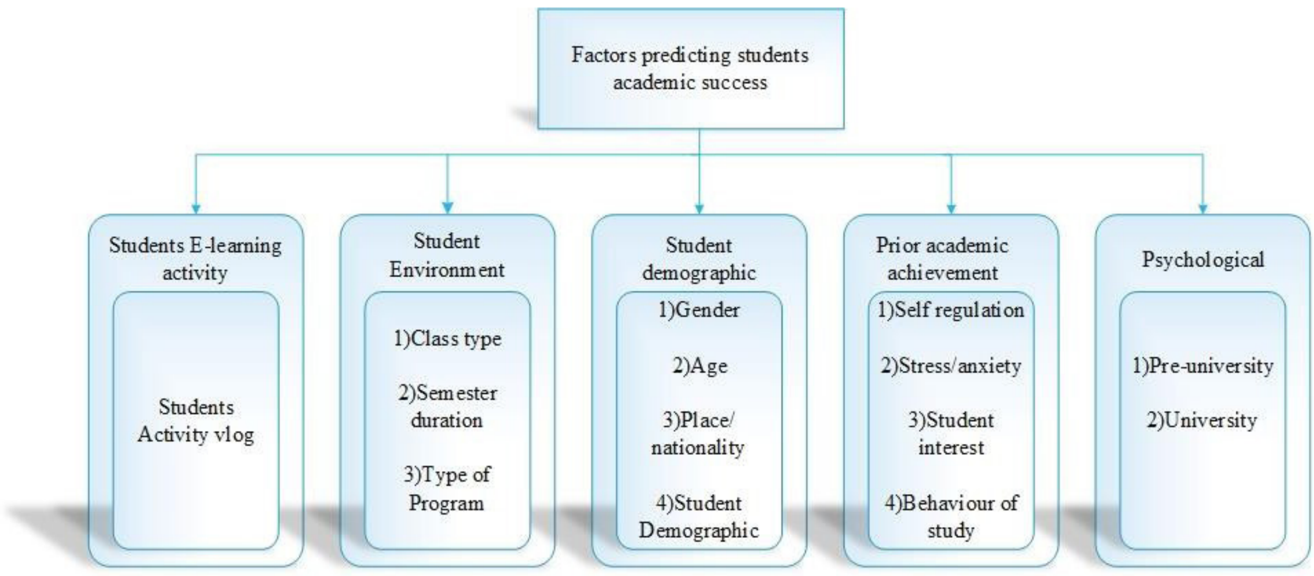

Figure 3. Factors predicting student academic performance

https://doi.org/10.56294/dm2025730

Machine Learning Techniques for Predicting Student Performance

This section reviews and summarizes the machine learning (ML) models most widely used in predicting student performance. It also details some of the latest studies, highlighting their findings and the evaluation metrics used for performance prediction.

Review of ML Models & Findings

Table 6 provides a review of previous machine learning models and findings.

Table 6. Review of ML Models in Previous Studies

Author/year	Technique	Findings
Mussa S. Abubakari, et al.[39]	Neural Network algorithm	Utilized deep learning with TensorFlow for predicting students' academic performance in educational data mining.
Injadat, et al.[40]	Quantitative, Graphical, and Statistical Techniques	Developed an ensemble model for educational data mining to systematically classify outliers.
Hussain, et al.[41]	Waikato Environment for Knowledge Analysis (WEKA)	Analyzed academic performance using WEKA for educational data mining. Explored course suitability for different student clusters.
Injadat, et al.[42]	Machine Learning Algorithms for Model Optimization	Selected Multi-split Optimized Bagging Ensemble Models for multi-class educational data mining.
Karthikeyan, et al.[43]	Hybrid Educational Data Mining Model (HEDM)	Created a hybrid model (HEDM) to facilitate and accurately assess student performance.
Aouifi, et al.[44]	K-Nearest Neighbors (KNN) and Multilayer Perceptron	Predicted student success using video clips and analyzed viewing behavior through educational data mining.
Ramaswami, et al.[45]	CatBoost Algorithm	Developed a generic model for educational data mining to predict student outcomes.
Feng, et al.[46]	K-means Algorithm and Convolutional Neural Network (CNN)	Analyzed and forecasted students' academic performance using educational data mining.

Support Vector Machine

Table 7 provides a review of previous models that utilize Support Vector Machines (SVM), detailing the techniques and findings.

Table 7. A Review of SVM Models in Previous Studies

Author/year	Techniques	Findings
Ghorbani et al.[47]	SVM-SMOTE	Compared different resampling methods using machine learning techniques to predict students' performance.
Kadambande, et al.[48]	Support Vector Machine (SVM)	Developed a performance prediction system for students.
Zulfiker, et al.[49]	SVM, Logistic Regression, Decision Tree, KNN, MLP, AdaBoost, Extra Tree Classifier	Applied various machine learning approaches to predict student performance at Bangladeshi private universities.
Wang, et al.[50]	HRNN+SVM	Used short-term sequential campus behaviors for student performance prediction.
Pushpa, et al.[51]	Support Vector Machine	Utilized machine learning to predict class results.
Arnedo, et al.[52]	SVM+ Black-box techniques	Enhanced the expressiveness of black-box models to forecast student performance.
Chui, et al.[53]	RTV-SVM	Predicted at-risk university students in a virtual learning environment using machine learning algorithms
Sarwat, et al.[54]	CGAN+SVM	Applied deep SVM and Conditional Generative Adversarial Networks for academic performance prediction.

Naive Bayes (NB) and K-nearest neighbor (KNN)

This section reviews and details the use of K-Nearest Neighbor (KNN) and Naive Bayes (NB) models in predicting student performance.

Shaziya et al. proposed a strategy using a Naive Bayes Classifier to predict students' semester examination results. This method forecasts student success based on expected final semester grades. It benefits all stakeholders in the education system—teachers, students, and educational institutions. The predictions can be used to support students in various ways.[55]

Hamoud et al. developed a method for predicting student success using Bayes algorithms. Their model recommends the optimal algorithm based on performance data. The study employed responses from student questionnaires and two Bayes algorithms: Naive Bayes and Bayes Network.[56]

Isa and Asril proposed a K-Nearest Neighbor (K-NN) approach to predict students' final grades. Their model achieved 93,2 % accuracy in predicting on-time study status, 91,5 % accuracy for overall year predictions, and 75,63 % for semester predictions, based on data from 1989 computer science students at BINUS University from 2016 to 2019.[57]

Brown suggested a K-Nearest Neighbor model for predicting math test scores. The K-NN method compares the Euclidean distance between a test record and training records. It examines the K most similar records and predicts the category most frequent among them.[58]

Mardolkar and Kumaran developed a K-Nearest Neighbor model to predict and prevent student dropout. Their study evaluated various settings and variants of the KNN method using data from ninth-grade students in an English-medium school. The model helps identify underperforming students and provides early intervention.[59]

Kabakchieva proposed a KNN method for predicting student performance to support university management. The model uses personal, pre-university, and university performance factors to predict student success. The research was conducted at one of Bulgaria's most prestigious universities.[60]

Singh and Pal introduced an ensemble technique to improve student performance prediction. They used Decision Tree (DT), Extra Tree (ET), Naive Bayes (NB), and K-Nearest Neighbors (KNN) algorithms, combining them with boosting and bagging methods to create an ensemble model. The study compared results from both approaches to select the best model.[61]

Fujita and Son proposed a neural-fuzzy modeling approach to forecast student performance. Their Multi-Input Multi-Output Student Academic Performance Prediction (MIMO SAPP) model addresses the prediction of future success after college enrollment. They introduced MANFIS-S (Multi Adaptive Neuro-Fuzzy Inference System with Representative Sets) to overcome limitations in existing methods, utilizing multiple parameters sets and a unique learning methodology.[62]

Random Forest (RF) and Decision Tree (DT)

This section reviews ML algorithms, specifically Decision Tree (DT) and Random Forest (RF), based on previous models. The techniques, findings, and authors are detailed in table 8 below.

Table 8. Review of Decision Tree (DT) and Random Forest (RF) Techniques in Prior Research

Author/year	Techniques	Findings
Hasan, et al.[63]	RF	Applied data mining techniques and video analytics for predicting student performance in higher education.
Ahmed et al.[64]	WERKA+RF	Explained the use of the Random Forest algorithm in the educational field.
Dass, et al.[65]	RF	Utilized a Random Forest model to predict student dropout in self-paced MOOCs.
Cam, et al.[66]	MLP+RF	Identified and predicted factors contributing to the learning performance of first- year university students using Decision Trees and Random Forest algorithms
Hamsa, et al.[67]	DT + Fuzzy Genetic Algorithm	Used genetic fuzzy algorithms and Decision Trees to predict academic performance.
Park et al.[68]	DT	Employed Decision Tree analysis to forecast student assessments of instruction.
Hoque, et al.[69]	DT	Analyzed results and predicted outcomes using the Decision Tree algorithm for university students.
Cam, et al.[70]	DT+RF	Identified and predicted factors contributing to the learning performance of first-year university students using Decision Trees and Random Forest algorithms.
Ning Fang et al.[71]	DT	Predicted student performance in high-impact, high-enrollment core engineering courses using Decision Trees.

Vera, et al.[72]	Genetic Programming Algorithm + DT	Used genetic programming and various data mining techniques to predict academic failure in students with high-dimensional and unbalanced data.
Crockett, et al. [73]	FuzzyRF + DT	Employed Fuzzy Decision Trees to predict learning styles in an intelligent tutoring system.

The second research question of this SLR study is: (Q2) What are the machine learning techniques used to predict student performance?

Support Vector Machines, K-Nearest Neighbor, Decision Trees, and Random Forest are among the most widely used machine learning techniques for predicting student performance. Deep learning techniques, such as Convolutional Neural Networks and Artificial Neural Networks, have also been extensively explored. The most used metrics to evaluate performance include overall accuracy, ROC AUC, recall, specificity, and precision. This SLR also revealed that hybrid and ensemble approaches can significantly improve prediction accuracy. However, the number of studies employing hybrid and ensemble methods is limited, indicating a clear need for more research in this area to enhance the prediction of student performance.

Primary Evaluation Metrics for ML Models

This section presents the primary evaluation metrics used for predicting student performance. The results from the review are detailed in table 9.

Table 9. Primary Evaluation Metrics

Author/ Publication Year	Objective	Methods	Evaluation Metric(s)
Zaffar et al.[74]	End-of-Term	SVM+ FCBF	Precision, Recall, Accuracy, F1 Score
Jiang et al.[75]	In-Term	Preference CD	MAE and RMSE
Gitinabard et al.[76]	Earlier prediction	RF, Logistic Regression, SVM,	F1 Score
Aydogdu[77]	End-of-Term	ANN	Accuracy
He et al.[78]	At-risk identification	Neural Networks	Accuracy
Mengash[79]	End-of-Term	ANN	F1Score 10-fold cross-validation, Recall, Accuracy and Precision
Yang et al.[80]	End-of-Term	RF	Precision Hold-out, ROC AUC, Accuracy, Recall and Specificity
Figueroa-Cañas et al.[81]	Performance and Annual Exam Dropout	Conditional Tree	Recall
Waheed et al.[82]	Early Prediction	DNN	Accuracy Hold-out Yousafzai et al., Recall and Precision
Deo et al.[83]	End-of-Term	ELM	RRMSE, root MSE, MAE and MAPE
Turabieh et al.[84]	End-of-Term	RNN Layered+ HHO	Accuracy
Wang et al.[85]	End-of-Term	Attention-based Hybrid RNN + SVM	Accuracy
Tsiakmaki et al.[86]	At risk Students	Transfer Learning and DNN	Accuracy
Yan et al.[87]	Performance of Student Predicted in Academic Competition	Model Ensembled (SVM/RF/ AdaBoost)	ROC AUC, F1 Score, Recall and Precision

Metrics Used for Describing the Performance

This section presents the findings on the types of metrics used to describe students' academic performance. Only quantitative metrics were considered in this study. The most widely used metric for describing performance is course grade, accounting for 27 % of the reviewed studies. Exam/test scores were used in 22 % of the studies, while grade range/PASS-FAIL (20 %), student at risk/retention (13 %), and GPA/CGPA (11 %) were also commonly used metrics for performance prediction. Figure 4 illustrates the percentage at which various metrics have been used to describe performance. Some papers used more than one metric, and these papers were counted under each relevant metric.

The third research question of this SLR study is: (Q3) What types of metrics are used to describe student performance?

Establishing a precise definition of academic performance is crucial for predicting students' success in higher education. Widely considered metrics in the literature include quantifiable measures such as pass/fail probability, course grades, retention, and successful graduation from a program. However, there is a need to explore more complex performance metrics, such as knowledge gain and the speed of course or assignment completion.

Additionally, there is a need to incorporate Explainable AI (XAI) into student performance prediction studies to gain a deeper understanding of the factors influencing academic outcomes. The integration of XAI techniques, such as SHAP and LIME, in the machine learning workflow enables educators to interpret and trust model predictions. This transparency is crucial for the adoption of AI-driven decision-making in education, allowing for more informed and effective interventions.

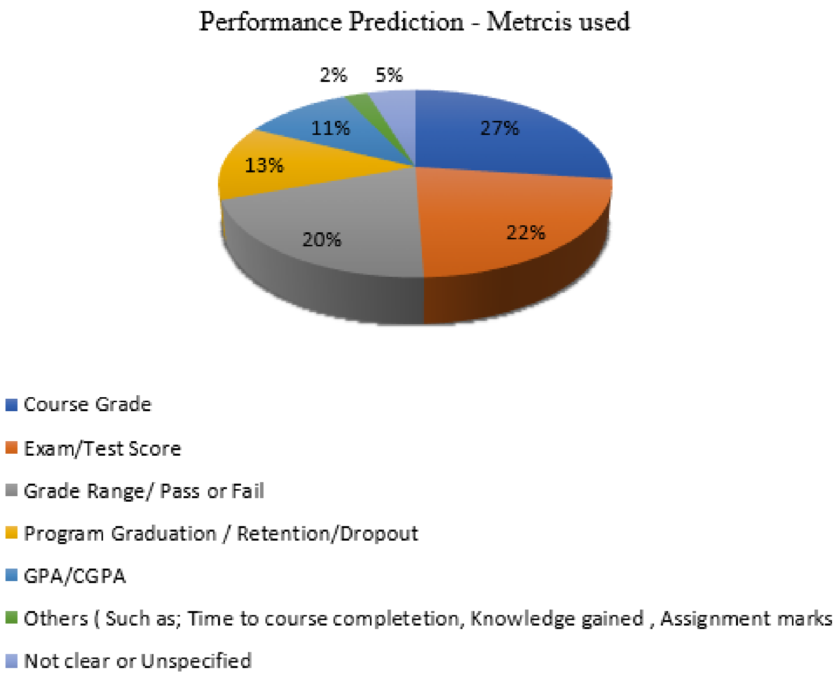

Figure 4. Metrics used for describing student performance

CONCLUSIONS

Predicting student performance has proven to be highly beneficial for both students and educators, enabling them to tailor and enhance their learning and teaching strategies. By accurately forecasting student outcomes, educators can identify at-risk students early, allowing for timely interventions that can significantly improve academic success. Students, on the other hand, can gain insights into their academic strengths and weaknesses, helping them to focus their efforts more effectively.

This article has reviewed and analyzed early research on various techniques used to predict student performance. Most studies have utilized internal assessments and cumulative grade point average (CGPA) data as key indicators in their prediction models. These metrics provide a quantifiable measure of student performance and are commonly employed in educational data mining (EDM).

Classification techniques have emerged as a dominant approach within the field of educational data mining. Among these, decision trees and neural networks are particularly prevalent. Decision trees offer a clear and interpretable model, making them a popular choice for educators looking to understand the factors contributing to student success or failure. Neural networks, on the other hand, provide a more complex and nuanced analysis, capable of uncovering patterns in data that may not be immediately apparent through traditional methods.

This analysis has also examined the broader context of data mining and the various machine learning techniques used to develop student performance prediction systems. Techniques such as Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Random Forest (RF), and deep learning models like Convolutional Neural Networks (CNN) and Artificial Neural Networks (ANN) have all been explored for their potential to enhance prediction accuracy.

https://doi.org/10.56294/dm2025730

Moreover, this study highlights the growing interest in hybrid and ensemble approaches, which combine multiple algorithms to improve prediction outcomes. Although still limited in number, studies employing these advanced techniques have shown promise in delivering more accurate and reliable predictions. However, there is a clear need for further research in this area to fully realize the potential of these methods.

In addition to the techniques used, this review has also shed light on the types of metrics commonly employed to describe student performance. Traditional metrics such as course grades, pass/fail status, and retention rates are widely used, but there is a growing recognition of the need to incorporate more complex metrics, such as knowledge gain and the speed of course or assignment completion. These advanced metrics could provide a more holistic view of student performance, offering deeper insights into the factors that contribute to academic success.

Additionally, there is a need to incorporate Explainable AI (XAI) into student performance prediction studies to gain a deeper understanding of the factors influencing academic outcomes. The integration of XAI techniques, such as SHAP and LIME, in the machine learning workflow enables educators to interpret and trust model predictions. This transparency is crucial for the adoption of AI-driven decision-making in education, allowing for more informed and effective interventions.

In conclusion, while significant progress has been made in the field of student performance prediction, there remains ample opportunity for further exploration and innovation. As educational institutions continue to adopt data-driven approaches, the integration of more sophisticated machine learning techniques and the development of more comprehensive performance metrics will be crucial in advancing the effectiveness of these predictive models.

BIBLIOGRAPHIC REFERENCES

1. Ajibade SSM, Ahmad NB, Shamsuddin SM. A heuristic feature selection algorithm to evaluate academic performance of students. In: 2019 IEEE 10th Control and System Graduate Research Colloquium (ICSGRC); 2019 Aug; IEEE. p. 110-4.

2. Joshi A, Vinay M, Bhaskar P. Impact of coronavirus pandemic on the Indian education sector: perspectives of teachers on online teaching and assessments. Interact Technol Smart Educ. 2020.

3. Huber SG, Helm C. COVID-19 and schooling: evaluation, assessment, and accountability in times of crises—reacting quickly to explore key issues for policy, practice and research with the school barometer. Educ Assess Eval Account. 2020;32(2):237-70.

4. Houlden S, Veletsianos G. The problem with flexible learning: Neoliberalism, freedom, and learner subjectivities. Learn Media Technol. 2021;46(2):144-55.

5. Horta H, Santos JM. Organizational factors, and academic research agendas: an analysis of academics in the social sciences. Stud High Educ. 2020;45(12):2382-97.

6. Kapur R. Factors influencing the students' academic performance in secondary schools in India. Univ Delhi. 2018;575-87.

7. Leal-Rodriguez AL, Albort-Morant G. Promoting innovative experiential learning practices to improve academic performance: Empirical evidence from a Spanish Business School. J Innov Knowl. 2019;4(2):97-103.

8. Singh M, Adebayo SO, Saini M, Singh J. Indian government E-learning initiatives in response to COVID-19 crisis: A case study on online learning in Indian higher education system. Educ Inf Technol. 2021;26(6):7569-607.

9. Hooshyar D, Pedaste M, Yang Y. Mining educational data to predict students' performance through procrastination behavior. Entropy. 2019;22(1):12.

10. Shin J, Chen F, Lu C, Bulut O. Analyzing students' performance in computerized formative assessments to optimize teachers' test administration decisions using deep learning frameworks. J Comput Educ. 2022;9(1):71-91.

11. Waheed H, Hassan SU, Aljohani NR, Hardman J, Alelyani S, Nawaz R. Predicting academic performance of students from VLE big data using deep learning models. Comput Human Behav. 2020;104:106189.

12. Nabil A, Seyam M, Abou-Elfetouh A. Prediction of students' academic performance based on courses' grades using deep neural networks. IEEE Access. 2021;9:140731-46.

13. Dien TT, Luu SH, Thanh-Hai N, Thai-Nghe N. Deep learning with data transformation and factor analysis for student performance prediction. Int J Adv Comput Sci Appl. 2020;11(8).

14. Song X, Li J, Sun S, Yin H, Dawson P, Doss RRM. SEPN: a sequential engagement based academic performance prediction model. IEEE Intell Syst. 2020;36(1):46-53.

15. Sultana J, Rani MU, Farquad MAH. Student's performance prediction using deep learning and data mining methods. Int J Recent Technol Eng. 2019;8(1S4):1018-21.

16. Rao AS, Aruna Kumar SV, Jogi P, Chinthan Bhat K, Kuladeep Kumar B, Gouda P. Student placement prediction model: A data mining perspective for outcome-based education system. Int J Recent Technol Eng (IJRTE). 2019;8:2497-507.

17. Livieris IE, Drakopoulou K, Tampakas VT, Mikropoulos TA, Pintelas P. Predicting secondary school students' performance utilizing a semi-supervised learning approach. J Educ Comput Res. 2019;57(2):448-70.

18. Khakata E, Omwenga V, Msanjila S. A stochastic modeling approach to student performance prediction on an internet-mediated environment. J Syst Integr. 2020;11(1).

19. Abazeed A, Khder M. A classification and prediction model for student's performance in university level. J Comput Sci. 2017;13(7):228-33.

20. Anuradha C, Velmurugan T. A comparative analysis on the evaluation of classification algorithms in the prediction of students' performance. Indian J Sci Technol. 2015;8(15):1-12.

21. Wang X, Zhang L, He T. Learning performance prediction-based personalized feedback in online learning via machine learning. Sustainability. 2022;14(13):7654.

22. Kitchenham B, Brereton OP, Budgen D, Turner M, Bailey J, Linkman S. Systematic literature reviews in software engineering–a systematic literature review. Inf Softw Technol. 2009;51(1):7-15.

23. Soares DL, Lemos GC, Primi R, Almeida LS. The relationship between intelligence and academic achievement throughout middle school: The role of students' prior academic performance. Learn Individ Differ. 2015;41:73-8.

24. Cutumisu M, Schwartz DL, Lou NM. The relation between academic achievement and the spontaneous use of design-thinking strategies. Comput Educ. 2020;149:103806.

25. Cutumisu M. Exploring the relationship between law students' prior performance and academic achievement at university. Educ Sci. 2019;9(3):236.

26. Hettler PL. Student demographics and the impact of team-based learning. Int Adv Econ Res. 2015;21(4):413-22.

27. Pincus KV, Stout DE, Sorensen JE, Stocks KD, Lawson RA. Forces for change in higher education and implications for the accounting academy. J Account Educ. 2017;40:1-18.

28. Yilmaz R. Exploring the role of e-learning readiness on student satisfaction and motivation in flipped classroom. Comput Human Behav. 2017;70:251-6.

29. Oztekin A, Delen D, Turkyilmaz A, Zaim S. A machine learning-based usability evaluation method for eLearning systems. Decis Support Syst. 2013;56:63-73.

30. Klašnja-Milićević A, Vesin B, Ivanović M, Budimac Z. E-Learning personalization based on hybrid recommendation strategy and learning style identification. Comput Educ. 2011;56(3):885-99.

31. Tomasevic N, Gvozdenovic N, Vranes S. An overview and comparison of supervised data mining techniques for student exam performance prediction. Comput Educ. 2020;143:103676.

https://doi.org/10.56294/dm2025730

32. Mubarak AA, Cao H, Hezam IM. Deep analytic model for student dropout prediction in massive open online courses. Comput Electr Eng. 2021;93:107271.

33. Asif R, Merceron A, Ali SA, Haider NG. Analyzing undergraduate students' performance using educational data mining. Comput Educ. 2017;113:177-94.

34. Costa EB, Fonseca B, Santana MA, de Araújo FF, Rego J. Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Comput Human Behav. 2017;73:247-56.

35. Natek S, Zwilling M. Student data mining solution–knowledge management system related to higher education institutions. Expert Syst Appl. 2014;41(14):6400-7.

36. Kabakchieva D. Student performance prediction by using data mining classification algorithms. Int J Comput Sci Manag Res. 2012;1(4):686-90.

37. Feng G, Fan M, Chen Y. Analysis and prediction of students' academic performance based on educational data mining. IEEE Access. 2022;10:19558-71.

38. Abubakaria MS, Arifin F, Hungilo GG. Predicting students' academic performance in educational data mining based on deep learning using TensorFlow. Int J Educ Manag Eng (IJEME). 2020;10(6):27-33.

39. Injadat M, Moubayed A, Nassif AB, Shami A. Systematic ensemble model selection approach for educational data mining. Knowl-Based Syst. 2020;200:105992.

40. Pal AK, Pal S. Analysis and mining of educational data for predicting the performance of students. Int J Electron Commun Comput Eng. 2013;4(5):1560-5.

41. Hussain S, Dahan NA, Ba-Alwib FM, Ribata N. Educational data mining and analysis of students' academic performance using WEKA. Indones J Electr Eng Comput Sci. 2018;9(2):447-59.

42. Injadat M, Moubayed A, Nassif AB, Shami A. Multi-split optimized bagging ensemble model selection for multi-class educational data mining. Appl Intell. 2020;50(12):4506-28.

43. El Aouifi H, El Hajji M, Es-Saady Y, Douzi H. Predicting learner's performance through video sequences viewing behavior analysis using educational data-mining. Educ Inf Technol. 2021;26(5):5799-814.

44. Ramaswami G, Susnjak T, Mathrani A. On developing generic models for predicting student outcomes in educational data mining. Big Data Cogn Comput. 2022;6(1):6.

45. Feng G, Fan M, Chen Y. Analysis and prediction of students' academic performance based on educational data mining. IEEE Access. 2022;10:19558-71.

46. Ghorbani R, Ghousi R. Comparing different resampling methods in predicting students' performance using machine learning techniques. IEEE Access. 2020;8:67899-911.

47. Kadambande A, Thakur S, Mohol A, Ingole AM. Predicting students' performance system. Int Res J Eng Technol. 2017;4(5):2814-6.

48. Zulfiker MS, Kabir N, Biswas AA, Chakraborty P, Rahman MM. Predicting students' performance of the private universities of Bangladesh using machine learning approaches. Int J Adv Comput Sci Appl. 2020;11(3).

49. Wang X, Yu X, Guo L, Liu F, Xu L. Student performance prediction with short-term sequential campus behaviors. Inf. 2020;11(4):201.

50. Pushpa SK, Manjunath TN, Mrunal TV, Singh A, Suhas C. Class result prediction using machine learning. In: 2017 International Conference on Smart Technologies for Smart Nation (SmartTechCon); 2017 Aug; IEEE. p. 1208-12.

51. Villagrá-Arnedo CJ, Gallego-Durán FJ, Llorens-Largo F, Compañ-Rosique P, Satorre-Cuerda R, Molina-Carmona R. Improving the expressiveness of black-box models for predicting student performance. Comput Hum Behav. 2017;72:621-31.

52. Chui KT, Fung DCL, Lytras MD, Lam TM. Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. Comput Hum Behav. 2020;107:105584.

53. Sarwat S, Ullah N, Sadiq S, Saleem R, Umer M, Eshmawi AA, Mohamed A, Ashraf I. Predicting Students' Academic Performance with Conditional Generative Adversarial Network and Deep SVM. Sensors. 2022;22(13):4834.

54. Shaziya H, Zaheer R, Kavitha G. Prediction of students performance in semester exams using a Naïve Bayes classifier. Int J Innov Res Sci Eng Technol. 2015;4(10):9823-9.

55. Hamoud A, Humadi A, Awadh WA, Hashim AS. Students' success prediction based on Bayes algorithms. Int J Comput Appl. 2017;178(7):6-12.

56. Asril T, Isa SM. Prediction of student's study period using K-Nearest Neighbor algorithm. Int J. 2020;8(6).

57. Brown JM. Predicting math test scores using k-nearest neighbor. In: 2017 IEEE Integrated STEM Education Conference (ISEC); March 2017; 104-106.

58. Mardolkar M, Kumaran N. Forecasting and Avoiding Student Dropout Using the K-Nearest Neighbor Approach. SN Comput Sci. 2020;1(2):1-8.

59. Kabakchieva D. Student performance prediction by using data mining classification algorithms. Int J Comput Sci Manag Res. 2012;1(4):686-90.

60. Singh R, Pal S. Machine learning algorithms and ensemble technique to improve prediction of student's performance. IJATCSE. 2020;9(3).

61. Son LH, Fujita H. Neural-fuzzy with representative sets for prediction of student performance. Appl Intell. 2019;49(1):172-87.

62. Hasan R, Palaniappan S, Mahmood S, Abbas A, Sarker KU, Sattar MU. Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. Appl Sci. 2020;10(11):3894.

63. Ahmed NS, Sadiq MH. Clarify of the random forest algorithm in an educational field. In: 2018 International Conference on Advanced Science and Engineering (ICOASE); October 2018;179-84.

64. Dass S, Gary K, Cunningham J. Predicting student dropout in self-paced MOOC course using random forest model. Inf. 2021;12(11):476.

65. Huynh-Cam T, Chen LS, Le H. Using decision trees and random Forest algorithms to predict and determine factors contributing to first-year university students' learning performance. Algorithms. 2021;14(11):318.

66. Hamsa H, Indiradevi S, Kizhakkethottam JJ. Student academic performance prediction model using decision tree and fuzzy genetic algorithm. Procedia Technol. 2016;25:326-32.

67. Park E, Dooris J. Predicting student evaluations of teaching using decision tree analysis. Assess Eval High Educ. 2020;45(5):776-93.

68. Hoque MI, Azad AK, Tuhin MAH, Salehin ZU. University students result analysis and prediction system by decision tree algorithm. Adv Sci Technol Eng Syst J. 2020;5:115-22.

69. Huynh-Cam TT, Chen LS, Le H. Using decision trees and random Forest algorithms to predict and determine factors contributing to first-year university students' learning performance. Algorithms. 2021;14(11):318.

https://doi.org/10.56294/dm2025730

70. Fang N, Lu J. Work in progress-a decision tree approach to predicting student performance in a high-enrollment, high-impact, and core engineering course. In: 2009 IEEE Frontiers in Education Conference; October 2009;1-3.

71. Márquez-Vera C, Cano A, Romero C, Ventura S. Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. Appl Intell. 2013;38(3):315-30.

72. Crockett K, Latham A, Whitton N. On predicting learning styles in conversational intelligent tutoring systems using fuzzy decision trees. Int J Hum-Comput Stud. 2017;97:98-115.

73. Zaffar M, Hashmani M, Savita K, Sajjad S, Rehman M. Role of FCBF Feature Selection in Educational Data Mining. Mehran Univ Res J Eng Technol. 2020;39:772-8.

74. Jiang P, Wang X. Preference Cognitive Diagnosis for Student Performance Prediction. IEEE Access. 2020;8:219775-87.

75. Gitinabard N, Xu Y, Heckman S, Barnes T, Lynch C. How Widely Can Prediction Models Be Generalized? Performance Prediction in Blended Courses. IEEE Trans Learn Technol. 2019;12:184-97.

76. Aydogdu S. Predicting student final performance using artificial neural networks in online learning environments. Educ Inf Technol. 2020;25:1913-27.

77. He Y, Chen R, Li X, Hao C, Liu S, Zhang G, Jiang B. Online At-Risk Student Identification using RNN-GRU Joint Neural Networks. [citation incomplete].

78. Mengash H. Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems. IEEE Access. 2020;8:55462-70.

79. Yang J, Devore S, Hewagallage D, Miller P, Ryan Q, Stewart J. Using machine learning to identify the most at-risk students in physics classes. Phys Rev Phys Educ Res. 2020;16:020130.

80. Figueroa-Cañas J, Sancho-Vinuesa T. Early Prediction of Dropout and Final Exam Performance in an Online Statistics Course. IEEE Rev Iberoam Tecnol Aprendiz. 2020;15:86-94.

81. Waheed H, Ul Hassan S, Aljohani N, Hardman J, Alelyani S, Nawaz R. Predicting academic performance of students from VLE big data using deep learning models. Comput Hum Behav. 2020;104:106189.

82. Deo R, Yaseen Z, Al-Ansari N, Nguyen-Huy T, Langlands T, Galligan L. Modern Artificial Intelligence Model Development for Undergraduate Student Performance Prediction: An Investigation on Engineering Mathematics Courses. IEEE Access. 2020;8:136697-136724.

83. Turabieh H, Azwari S, Rokaya M, Alosaimi WA, Alhakami AAW, Alnfiai W. Enhanced Harris Hawks optimization as a feature selection for the prediction of student performance. Comput. 2021;103:1417-38.

84. Wang X, Yu X, Guo L, Liu F, Xu L. Student Performance Prediction with Short-Term Sequential Campus Behaviors. Inf. 2020;11:201.

85. Tsiakmaki M, Kostopoulos G, Kotsiantis S, Ragos O. Transfer Learning from Deep Neural Networks for Predicting Student Performance. Appl Sci. 2020;10:2145.

86. Yan L, Liu Y. An Ensemble Prediction Model for Potential Student Recommendation Using Machine Learning. Symmetry. 2020;12:728.

87. Figueroa-Cañas J, Sancho-Vinuesa T. Early Prediction of Dropout and Final Exam Performance in an Online Statistics Course. IEEE Rev Iberoam Tecnol Aprendiz. 2020;15:86-94.

88. TK S, Midhunchakkravarthy. Academic Performance Prediction of At-Risk Students using Machine Learning Techniques. In: 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE); Greater Noida, India; 2023:1222-1227. doi: 10.1109/ICACITE57410.2023.10183199.

89. Orji F, Fatahi S, Vassileva J. Data-Driven Approach for Student Engagement Modelling Based on Learning Behaviour. In: 2023. p. 46. doi: 10.1007/978-3-031-35998-9_46.

90. Zhang H, Wu H, Li Z, Gong W, Yan Y. Machine Learning based Analysis of the Effect of Team Competition on College Students' Academic Performance. 2023. doi: 10.21203/rs.3.rs-3519859/v1.

91. Thi H, Hai L, Trung D. Using machine learning to identify influential factors and predict student academic performance in blended learning. J Sci Nat Sci. 2023;68:63-76. doi: 10.18173/2354-1059.2023-0006.

92. Sunil Kumar Das. AI-Powered Predictive Analytics in Financial Forecasting: Implications for Corporate Planning and Risk Management. Int J Intell Syst Appl Eng. 2024;12(21s):3512-3516. Available from: https://ijisae.org/index.php/IJISAE/article/view/6061.

93. Tk S. A Study on Predictive Modelling of Student Academic Performance using Machine Learning Method. J Inf Syst Eng Manag. 2024;10:78-88. doi: 10.52783/jisem.v10i1s.103.

94. Adewale M, Azeta A, Adebayo-Alli A, Sambo-Magaji A. Empirical Investigation of Multilayered Framework for Predicting Academic Performance in Open and Distance Learning. Electronics. 2024;13:2808. doi: 10.3390/electronics13142808.

FINANCING
No financing.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHORS CONTRIBUTION
Data curation: Shoukath T K, Midhun Chakkaravarthy.
Methodology: Shoukath T K, Midhun Chakkaravarthy.
Software: Shoukath T K, Midhun Chakkaravarthy.
Drafting - original draft: Shoukath T K, Midhun Chakkaravarthy.
Writing - proofreading and editing: Shoukath T K, Midhun Chakkaravarthy.