International Journal of Computer Applications Technology and Research
Volume 14–Issue 10, 59 – 67, 2025, ISSN:-2319–8656
DOI:10.7753/IJCATR1410.1012

A Systematic Review of the Machine Learning-Based Educational Data Mining Models for Student Performance Prediction

Millicent Murithi
Murang'a University of Technology
Department of Computer Science
Kenya

David Marangu
Murang'a University of Technology
Department of Information Technology
Kenya

Abstract: Predicting students' performance is one of the applications of Educational Data Mining (EDM). EDM uses big data from education setting to predict students' performance .This provides an avenue to track the academic progress of all students by educators. Machine-learning-based models, such as artificial neural networks, decision trees, and support vector machines, and their hybrids are the most frequently used models to predict students' performance. These models have ability to efficiently extract hidden and useful information from large educational datasets. Following PRISMA framework, the review conducts an exhaustive literature search, emphasizing on the machine learning models used in prediction of the student performance. Key findings highlight the efficacy of machine learning based model used in predicting students' performance, and their major drawbacks. It was clear that these models use irrelevant features (noisy data), limited number of students' features, some of which are not within the current context of a particular learning environment to predict. This systematic review provides insights into existing machine learning models used in student performance prediction, contributing to the field's advancement and providing guidance for researchers and practitioners.

***Keywords**:* Machine learning, student performance prediction, Predictive Modelling, educational data mining

1. Introduction

The discipline of educational data mining (EDM) is concerned with developing models that are able to explore on the big data coming from the educational settings, and use these models to understand students as well as the settings which they learn in, in a better way (Ayienda et al., 2021). Data mining incorporates multi-displinary techniques which aim at extracting valuable and intellectual insights from raw data (Bendangnuksung, P. P., 2018). Educational data mining is a combination of three main multidisplinary which includes computer science, education and statistics (Romero & Ventura, 2013). Predicting students' performance is one of the main applications of educational data mining models (Feng et al., 2022). Prediction of students' performance provides an avenue for intervention measures of low performing students at an early stage before it is too late (Ahmad et al., 2021). It can also be used by the educators to track the academic progress of all students, with an intention of improving the teaching and learning processes of the individual students as well as an aid towards improving an existing curriculum (Ahmad et al., 2021). Additionally E-learning, where students are enrolled in online courses is a rapidly growing. Some of the E-learning platforms include the Intelligent Tutoring Systems (ITS), Learning Management Systems (LMS), and Massive Open Online Courses (MOOC). These E learning platforms' information gathered over time can be utilized by EDM in developing student performance predictive models. The acquired information, over time, is processed and analyzed using different machine learning models. Machine learning is a sub-field of AI, where ML systems learn from the data, analyze patterns and predict the outcomes. ML models can automatically and quickly analyze bigger and more complex data with accurate results and avoid unexpected risks. Over time, education data mining models have evolved from the traditional manual-based methods to the current machine learning-based methods (Chango et al., 2021). Hybrid machine learning models have recently gained popularity as the most reliable educational data mining models used for predicting students' performance (Chatzimparmpas et al., 2020).

The rest of the paper is organized as follows: The research methodology employed in this review study is presented in the next section. The third section presents the results; discussion of the review process is then discussed. Lastly, the conclusion and future research direction are presented. This is drawn from the systematic review process.

2. Research Methods

A systematic literature review is performed with a research method that is unbiased and ensures completeness to evaluate all available research related to the respective field. The systematic literature review in this study involved a systematic exploration and theoretical analysis of existing scholarly works, focusing on machine learning models used for student performance prediction. The theoretical analysis is pivotal in identifying current approaches and features used in the student performance prediction. This work adopted the PRISMA approach of the systematic review to identify and analyse the literature. This process involved framing the questions for the review. In this study, research questions guided the review. Secondly, the identification and screening of the relevant literature is done. Lastly, inclusion and exclusion criteria for the publications which included identification and screening of the terms and resources to be used are done. A study of the quality assessment procedure is identified before finally discussing the research findings.

2.1 Research questions

This systematic review study is guided by the following research questions:
RQ1: Which machine learning models are used in the student performance prediction?
RQ2: What are the commonly used attributes in the student performance prediction?
RQ3: What are the prediction tasks performed by the machine learning models?
RQ4: What is the effectiveness of machine learning models used in the student performance prediction?

2.2 Data Sources

In order to carry out an extensive systematic literature review based on the objectives of this review, we exploited five research databases to find the primary data and to search for the relevant papers. These databases include; IEEE, ACM, Research Gate, Elsevier science direct, and Springer databases which are the biggest collection of citations and abstracts. These sources were chosen because they had published papers on the topic.

2.3 Search string and Query space

The following search terms were used to find out data from the databases according to our research questions:

i. EDM OR Performance OR eLearning OR Machine Learning
ii. Educational Data Mining OR Student Performance Prediction OR Evaluations of Students OR Performance Analysis of Students OR Learning Curve Prediction

A combination of specific keywords is used to search for the articles from the online libraries.

2.4 Criteria for selection

The inclusion and exclusion criteria are guided by the research questions. The identification of the literature meta-search on articles dedicated to the "student performance prediction" and "machine learning" as the main keywords. The articles published between 2018 and 2025 are identified with an exemption of classical references. This choice allowed for a full-date review of the literature to be undertaken. A total of 258 papers were obtained through an exhaustive search before the quality assessment was conducted. Quality assessment and selection criteria further ensured that only the articles that have a greater relevance to the research question were chosen.
The selection procedure of the paper is comprised of identification, screening, eligibility checking, and inclusion criteria meeting of the research papers.

2.5 Assessment on Quality

Quality assessment parameters take the following into account to achieve good quality results that answer the research questions:

i. To extract the relevant research materials, only reliable electronic scientific libraries mentioned above are utilized.
ii. Only the peer-reviewed journal paper or published conference proceedings are considered to ascertain the good quality results
iii. The process of selecting the relevant articles is balanced and uses the following criteria for assessment;
    a) if the paper was a primary
    b) if the study objective which is a student performance prediction, and
    c) If the paper discussion had clarity.

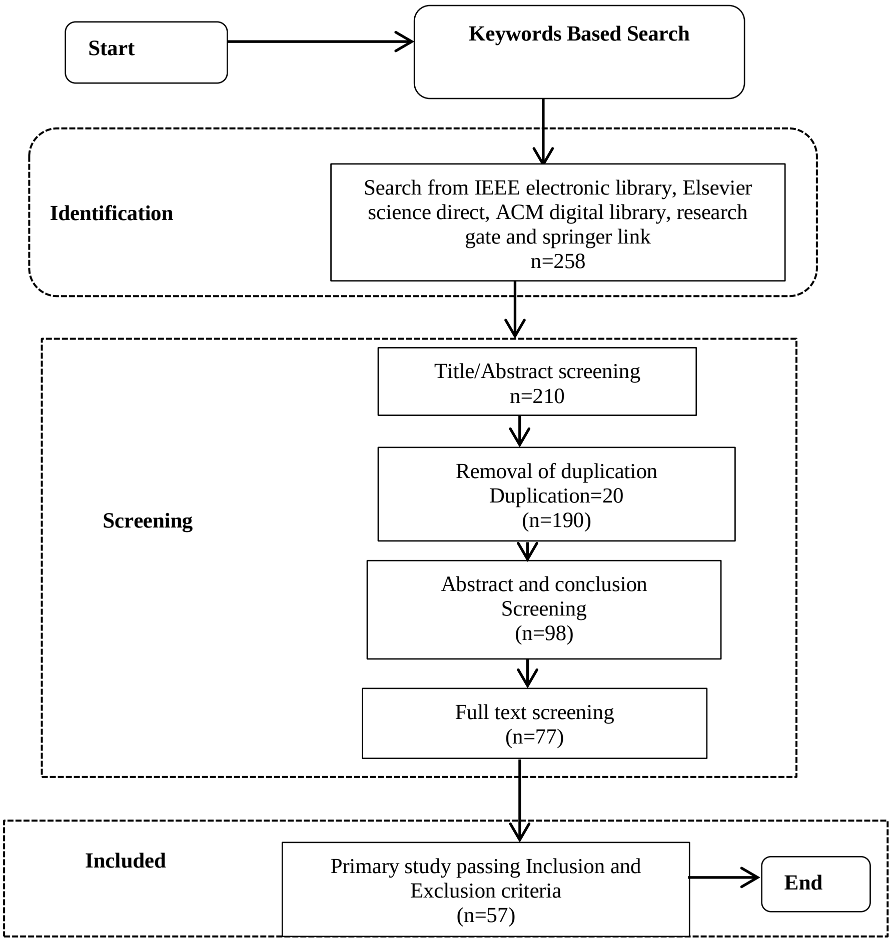

Fig 1: PRISMA Flow Chart for the Articles Selection

3. Results

In this section, we report the results of our systematic literature review study, where we address our RQs and elaborate on the findings derived from analysing the extracted data.

3.1 The machine learning models used in the student performance prediction

Due to the use of internet in education, there is availability of e-learning or web based education context which includes large amounts of information about teaching-learning interactions (Romero & Ventura, 2013);(Deuja et al., 2018). Students' performance prediction process is mostly done through machine learning and EDM models (Chen et al., 2020). Data, mainly available from online learning environments, can be used to improve students' learning (Wang, 2024).This section discusses the machine learning models used in the student performance prediction.

RQ1: Which machine learning models are used in the student performance prediction?

Figure 1 illustrates the distribution of the machine learning models used in the student performance prediction. The popularity of the hybrid models and neural networks has significantly increased in recent years; this indicates that predicting students' performance through hybrid machine learning and use of neural networks methods is attracting the attention of various researchers as shown in figure 2.

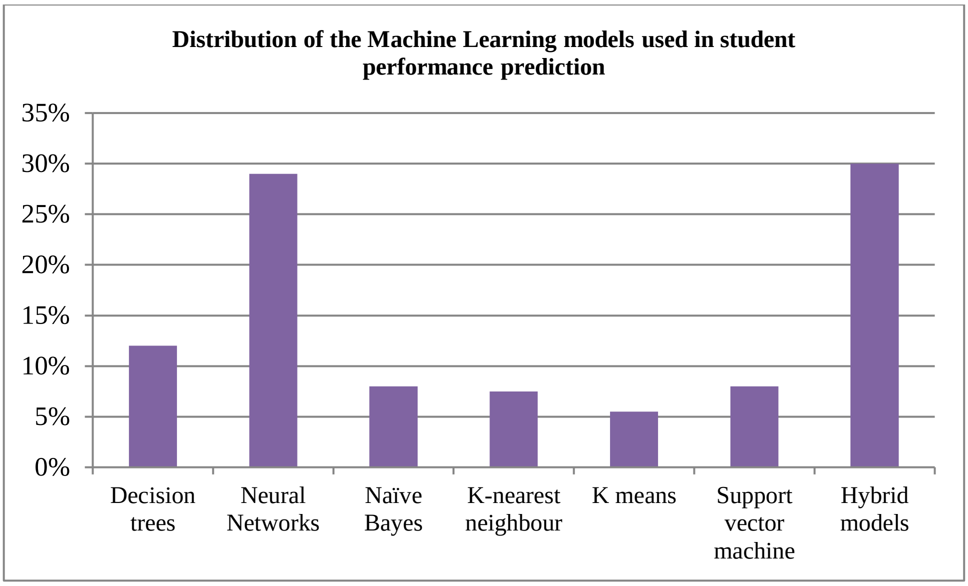

Figure2: Distribution of machine learning models used in the student performance prediction

3.2 Commonly used attributes in the student performance prediction

This section answers the research question 2 through an exhaustive research search and analysis.

RQ2: What are the commonly used attributes in the student performance prediction?

According to (Rizwan et al., 2025), the commonly used factors to predict students' performance are classified into five categories i.e. demographic, academic, family/personal, behavioural, psychological and institutional attributes. Chen et al., 2020, argue that academic and family attributes are the key deciding factors of student academic performance. In another study, the authors found that it is possible to predict the graduation performance in the fourth year at university using only pre-university marks and marks in fourth and second year's course (Chango et al., 2021). Table 1 summarizes the findings by category, and attributes used.

Table 1: summary of the attributes used in the student performance prediction

Attribute Category	Attributes Used	Reference
Academic	Grade ID; Section ID; Topic; Semester; Program; CGPA; Stage ID; visit to educational resources,	(Rizwan et al., 2025);(Chen et al., 2020); (Sokkhey P & Okazaki T, 2020)
Internal Assessment	Daily study time; University entrance examinations, Quizzes coursework, Assignments, Lab test, Examinations; Plagiarism, internal grade	Kumar & Aggarwal(Buschetto Macarini et al., 2019), (Alshabandar et al., 2020), (Xiaoming et al., 2022)
Demographic	Gender; Age; Nationality; Place of birth; Marital status; Guardian; Address, school location	(Xiaoming et al., 2022)
Behavioural and class attendance	visit to educational resources, viewing teachers announcement, commitment to discussion groups, levels of answering queries , Raised hands; Visited resources; student interaction with the learning management system and punctuality in classroom.	(Buschetto Macarini et al., 2019); (Braca et al., 2022); (Xiaoming et al., 2022)

Psychological	Learning strategies, Personality; Motivation, Contextual influences, Socio economic status, Approaches to learning	(Batool et al., 2023);(Alam & Mohanty, 2023)
Family/ Personal	Parent status,Parent satisfaction; Family size, Parent education, Income.	(Chen et al., 2020)

The most frequently used attributes are course work and CGPA which fall under the academic and internal assessment category. A total of 72% of the articles reviewed have utilized the internal assessment and academic group to predict the performance of the students. This is because internal assessment and academic factors have a significant academic potential. The learning behaviors, social factors, psychological factors and demographics have had influence in the prediction. Recently there has been innovation of researchers who started to collect multimodal data (such as eye gaze, facial expression and body gestures) for modeling (Deuja et al., 2018). The extent to which the multimodal data affects the learners' performance needs to be put in place when developing state of the art models during the future research endeavors. The use of static data to predict learning performance has resulted in some concerns, including ignoring student's actual efforts in the learning process and student educational records (such as quiz performance) as part of the final grade (Buschetto Macarini et al., 2019). However, few researchers in educational data mining have combined online behaviors with textual data for improving prediction performance (Ayienda et al., 2021). Therefore, combining online behaviors with textual data for prediction of performance is essential for improving prediction outcomes.

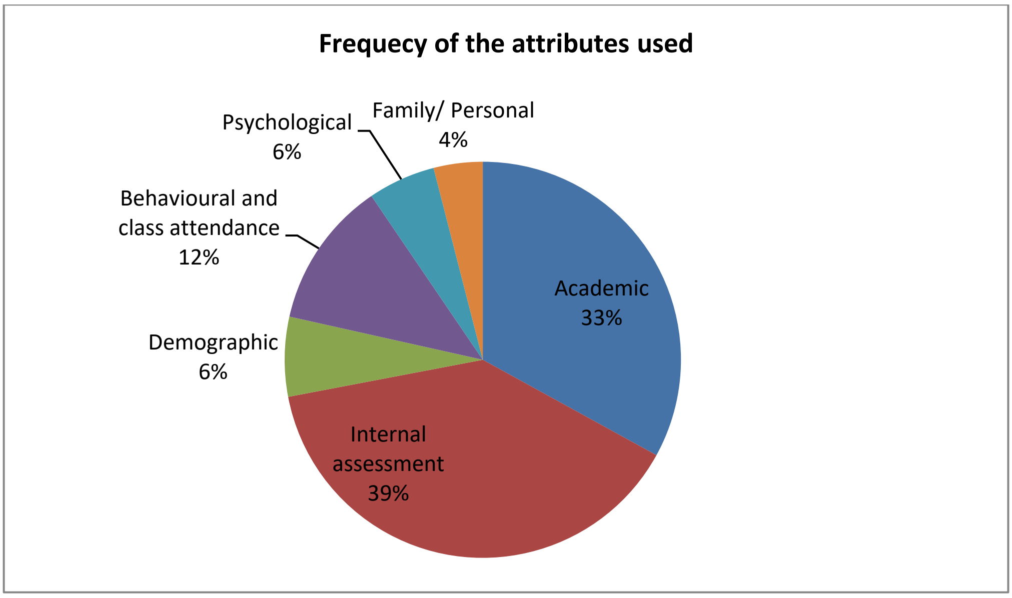

Figure 3: Frequency of the attributes used

3.3 The prediction tasks performed by the machine learning models

The literature reviews that machine learning models for student performance prediction are used to perform a range of predictive tasks with the aim to understand, forecast, or improve students' learning outcomes. Some of the popular tasks included; final grade prediction,drop out and retention risk, student learning behaviour. A few researchers were also done with aim of predicting the student graduation, collaboration of team in group work and knowledge mastery.

RQ3: What are the prediction tasks performed by the machine learning models?

Figure 3 shows the frequency of the prediction tasks performed by the machine learning models. This review provides valuable insights on the predictive tasks with final grade prediction and drop out and retention being the most popular task. This allows for personalized analysis and teaching improvement in the field. Additional tasks like student learning behaviour, team work collaboration and knowledge mastery allows educators to allocate teaching resources more effectively. It also allows for the creation of more targeted learning materials, enhancing the efficiency of the overall teaching-learning process.

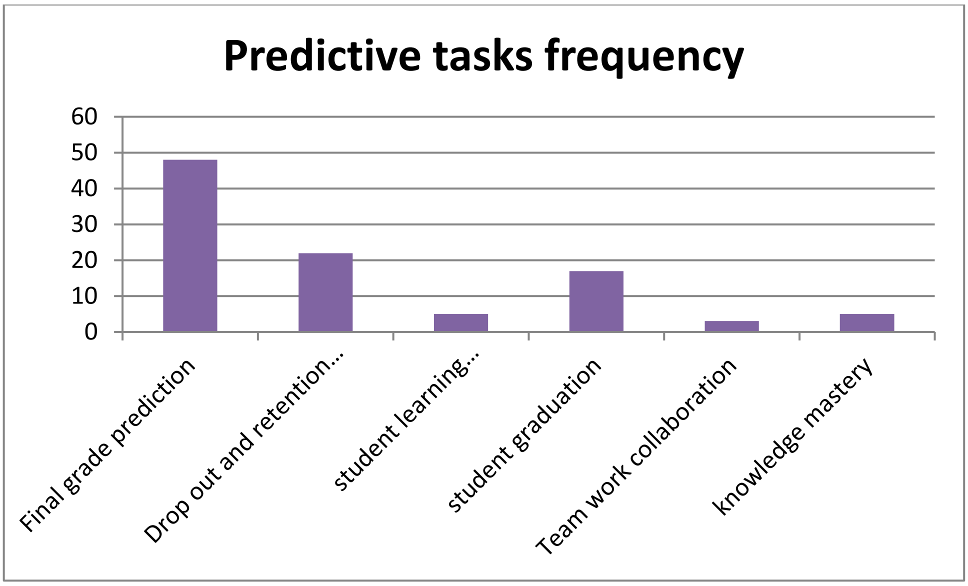

Figure 4: Predictive tasks performed by machine learning models

3.4 Machine learning models' effectiveness in the student performance prediction

This section evaluates how the effectiveness of the models has been evaluated. Table 2 highlights the machine learning models' evaluation criteria and the percentages achieved. The research question 4 was answered in this section.

RQ4: What is the effectiveness of machine learning models used in the student performance prediction?

Table 2: Evaluation metrics used for student performance prediction models

Reference	ML Model description	Evaluation
(Lu et al., 2025)	Dual-path attention mechanism, used personal information and behavioral data to predict	prediction accuracy of 72.43
(Biswas et al., 2025)	Recurrent neural network model to Predict and Monitor Student Performance	MSE of 12.5, an MAE of 2.6, and an R-squared of 0.65.
(Shen & Zhang, 2024)	Fusion algorithm that combines K-means clustering and LightGBM regression	Accuracy of 97.8%
(Ray et al., 2025)	Bagging Classifier ensemble models	Accuracy of 0.95, ROC AUC of 0.97, and F1 Score of 0.95,
(Alshanqiti & Namoun, 2020)	Hybrid machine learning approach (random forest, support vector machine, artificial neural network and extreme gradient boosting)	accuracy and root mean square error (RSME) as the evaluation metrics of their work. They used ten fold cross-validations
(Wafi et al., 2019)	k-nearest neighbor	82.6% accuracy
Selvam (Selvam, 2021)	Hybrid classification algorithms (radial basis function network (RBFN), multilayer perceptron, C4.5 and random forest algorithms)	accuracy of 76.4583%.
(Hui et al., 2025)	Artificial neural network and random Forest	accuracy of 91.08%.
(Sokkhey P & Okazaki T, 2020)	Hybrid algorithm(principal component analysis, decision trees, naive bayes, random forest and support vector machines)	Accuracy of 89.9%
(Wang, 2024)	GIWRF-SVM utilizing Gini Impurity based Weighted Random Forest (GIWRF) with Support Vector Machine (SVM) performance prediction rate.	Mean Absolute Error (MAE) of 9.24, Root Mean Square Error (RMSE) of 4.28 and Root

		Absolute Error (RAE) of 4.76
(Ouyang, 2024)	Improved Support Vector Machine (ISVM)	accuracy of 89.57%
(Tripathi & Reddy, 2025)	Random Forest classification	accuracy of 99 %
(Ram Pavan Kumar M et al., 2025)	Random Forest, Gradient Boosting, and Neural Networks, with deep learning models	accuracy of 89.3%.
(Ali et al., 2023)	Several machine learning classifiers were applied such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Support Vector Machine (SVM), and Naive Bayes (NB)	CNN-RNN outperforms other methods with an accuracy equal to 89.1%.
(Nabil et al., 2021)	Deep neural network (DNN), decision tree, random forest, gradient boosting, logistic regression, support vector classifier, and K-nearest neighbor.	accuracy of 89%,
(Niyogisubizo et al., 2022)	Hybrid of Random Forest (RF), Extreme Gradient Boosting (XGBoost), Gradient Boosting (GB), and Feed-forward Neural Networks (FNN) to predict student's dropout in university classes.	F1-Score of 0.92.
(Asselman et al., 2023)	Machine learning XGBoost algorithm	AUC of 71.08%

4. Discussion

i. Models' diversity in the student performance prediction

It is clear from the literature that thirty percent of the models used for the student performance prediction are the hybrid models. This is because they combine the strengths of different approaches. Neural networks which include ANN, RNN, and CNN have also gained popularity due to their ability to handle complex, non-linear patterns and sequential data from learning management systems. The supervised machine learning models have continually been used in the prediction due to availability of the input features from the EDM. Additionally they utilize clear performance metrics like accuracy, F1 score, precision and AUC.

ii. Trends in the attributes used in the student performance prediction

The typical attributes used in the prediction of students' performance usually include assessments, academics, learning behaviours, social factors, demographics, Family and psychological issues. It is clear that all factors which contribute to static and dynamic data are needed for the prediction of students' performance. Static data include demographics, self-report data and historical educational records, on the other hand, online behaviours, textual data and other multimodal data can be regarded as dynamic data (Bilal et al., 2022). Recently there has been innovation of researchers who collects multimodal data (such as eye gaze, facial expression and body gestures) for modelling (Kim et al., 2018). It is clear that the extent to which the multimodal data affects the learners' performance needs to be put in place when developing state of the art student performance prediction models.

iii. Diversity in the prediction tasks performed by the machine learning models

It is clear that a range of predictive tasks are performed by the ML models. Final grade prediction and drop out and retention risk have gained popularity as the main tasks performed. Others include student learning behaviour, student graduation, team work collaboration and knowledge mastery. It is evidenced that educational sector can utilize Data Mining (DM) and Machine Learning (ML) techniques to analyse and predict the academic performance of students in a reliable, sophisticated and timely manner in order to ensure academic success for every student.

iv. Performance evaluation metrics of machine learning models used in the student performance prediction

The study shows that accuracy is the most popular metric in evaluating prediction results. However recall, precision, f-measure, area under the curve, tenfold cross-validation procedure, space and time complexity has been used. Additionally error analysis metrics like Mean Square Error (RMSE) and Absolute Error have also been used in the evaluation. Visualization metrics like the confusion matrix have been used to visualize the performance of the classification algorithms.

5. Key Gaps

i. The use of static data to predict learning performance has resulted in some concerns; including ignoring student's actual efforts in the learning process and student educational records therefore use of real time data for prediction is very essential.

ii. There is need to combine online behaviours with textual data for improving prediction performance. The models should capture the students' features that are relevant and within the context of a particular students' learning environment.

6. Conclusion and Future Works

This work clearly demonstrates that the prediction of students' performance is critical in any education setup. Several machine learning models have been proposed, geared towards the prediction of students' performance. These models use educational datasets as the input and apply specific performance metrics during the evaluation process. The common metrics used in evaluation process include: F1 score, precision, accuracy, AUC, RMSE and recall. The datasets used as input into these models are educational datasets extracted from learning environments and the different predictor features within the datasets are categorized in different groups of the factors that affect students' performances i.e. student factors, family factors, school factors etc. These educational datasets contain both the static and dynamic data. However, we note that ensuring inclusivity of all the relevant features that can determine students' performance is critical to enhancing models' reliability and performance accuracy during the prediction process. Using predictor features within the context of a particular learning environment is not put into consideration when predicting students' performances with most of these models.

References

[1] Ahmad, M. S., Asad, A. H., & Mohammed, A. (2021). A Machine Learning Based Approach for Student Performance Evaluation in Educational Data Mining. *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, 187–192. https://doi.org/10.1109/MIUCC52538.2021.9447602

[2] Alam, A., & Mohanty, A. (2023). Predicting Students' Performance Employing Educational Data Mining Techniques, Machine Learning, and Learning Analytics. In R. S. Tomar, S. Verma, B. K. Chaurasia, V. Singh, J. H. Abawajy, S. Akashe, P.-A. Hsiung, & R. Prasad (Eds.), *Communication, Networks and Computing* (Vol. 1893, pp. 166–177). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-43140-1_15

[3] Ali, D. A., Aborizka, M., & Dahroug, A. (2023). Prediction of Student Performance by Using Machine Learning Techniques. *2023 4th International Conference on Artificial Intelligence, Robotics and Control (AIRC)*, 101–105. https://doi.org/10.1109/AIRC57904.2023.10303160

[4] Alshabandar, R., Hussain, A., Keight, R., & Khan, W. (2020). Students Performance Prediction in Online Courses Using Machine Learning Algorithms. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–7. https://doi.org/10.1109/IJCNN48605.2020.9207196

[5] Alshanqiti, A., & Namoun, A. (2020). Predicting Student Performance and Its Influential Factors Using Hybrid Regression and Multi-Label Classification. *IEEE Access*, *8*, 203827–203844. https://doi.org/10.1109/ACCESS.2020.3036572

[6] Asselman, A., Khaldi, M., & Aammou, S. (2023). Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. *Interactive Learning Environments*, *31*(6), 3360–3379. https://doi.org/10.1080/10494820.2021.1928235

[7] Ayienda, R., Rimiru, R., & Cheruiyot, W. (2021). Predicting Students Academic Performance using a Hybrid of Machine Learning Algorithms. *2021 IEEE AFRICON*, 1–6. https://doi.org/10.1109/AFRICON51333.2021.9571012

[8] Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H.-Y., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, *28*(1), 905–971. https://doi.org/10.1007/s10639-022-11152-y

[9] Bendangnuksung, P. P. (2018). Students' performance prediction using deep neural network. *International Journal of Applied Engineering Research*, *13*(2), 1171–1176.

[10] Bilal, M., Omar, M., Anwar, W., Bokhari, R. H., & Choi, G. S. (2022). The role of demographic and academic features in a student performance prediction. *Scientific Reports*, *12*(1), 12508. https://doi.org/10.1038/s41598-022-15880-6

[11] Biswas, U., Sinha, S., & De, P. (2025). Recurrent Neural Network Model to Predict and Monitor Student Performance. *2025 8th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)*, 1–6. https://doi.org/10.1109/IEMENTech65115.2025.10959550

[12] Braca, P., Millefiori, L. M., Aubry, A., Marano, S., De Maio, A., & Willett, P. (2022). Statistical Hypothesis Testing Based on Machine Learning: Large Deviations Analysis. *IEEE Open Journal of Signal Processing*, *3*, 464–495. https://doi.org/10.1109/OJSP.2022.3232284

[13] Buschetto Macarini, L. A., Cechinel, C., Batista Machado, M. F., Faria Culmant Ramos, V., & Munoz, R. (2019). Predicting Students Success in Blended Learning—Evaluating Different Interactions Inside Learning Management Systems. *Applied Sciences*, *9*(24), 5523. https://doi.org/10.3390/app9245523

[14] Chango, W., Cerezo, R., Sanchez-Santillan, M., Azevedo, R., & Romero, C. (2021). Improving prediction of students' performance in intelligent tutoring systems using attribute selection and ensembles of different multimodal data sources. *Journal of Computing in Higher Education*, *33*(3), 614–634. https://doi.org/10.1007/s12528-021-09298-8

[15] Chatzimparmpas, A., Martins, R. M., Jusufi, I., & Kerren, A. (2020). A survey of surveys on the use of visualization for interpreting machine learning models. *Information Visualization*, *19*(3), 207–233. https://doi.org/10.1177/1473871620904671

[16] Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *1*, 100002. https://doi.org/10.1016/j.caeai.2020.100002

[17] Deuja, R., Karna, R., & Kusatha, R. (2018). Data-Driven Predictive Analysis of Student Performance In College Using Neural Networks. *2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)*, 77–81. https://doi.org/10.1109/CCCS.2018.8586809

[18] Feng, G., Fan, M., & Chen, Y. (2022). Analysis and Prediction of Students' Academic Performance Based on Educational Data Mining. *IEEE Access*, *10*, 19558–19571. https://doi.org/10.1109/ACCESS.2022.3151652

[19] Hui, L., Shan, P., Hui, L., & Duan, S. (2025). *Artificial Intelligence for Student Performance Prediction in Blended Learning: A Systematic Literature Review*. https://doi.org/10.2139/ssrn.5311952

[20] Kim, B.-H., Vizitei, E., & Ganapathi, V. (2018). *Domain Adaptation for Real-Time Student Performance Prediction* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.1809.06686

[21] Lu, P., Gui, W., Lin, M., Tan, D., Sun, Y., Qu, S., & Du, Q. (2025). Research on Performance Prediction Model Based on Innovative Methods. *2025 13th International Conference on Information and Education Technology (ICIET)*, 165–170. https://doi.org/10.1109/ICIET66371.2025.11046326

[22] Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks. *IEEE Access*, *9*, 140731–140746. https://doi.org/10.1109/ACCESS.2021.3119596

[23] Niyogisubizo, J., Liao, L., Nziyumva, E., Murwanashyaka, E., & Nshimyumukiza, P. C. (2022). Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization. *Computers and Education: Artificial Intelligence*, *3*, 100066. https://doi.org/10.1016/j.caeai.2022.100066

[24] Ouyang, M. (2024). Evaluation and Prediction of Student's Learning Status in E-Commerce Smart Classroom based on Improved Support Vector Machine Algorithm. *2024 International Conference on Data Science and Network Security (ICDSNS)*, 1–4. https://doi.org/10.1109/ICDSNS62112.2024.10691105

[25] Ram Pavan Kumar M, V. T., Giridhar, N., Pavani, S., Srilekha, R., Harika, S., & Meghana, P. (2025). A Machine Learning Framework for Evaluating College Performance and Student Outcomes. *2025 8th International Conference on Trends in Electronics and Informatics (ICOEI)*, 1296–1300. https://doi.org/10.1109/ICOEI65986.2025.11013685

[26] Ray, S., Nawaz, A., & Ahmad, A. (2025). Enhancing Student Performance Prediction Through Ensembles of Machine Learning Models and Explainable Artificial Intelligence. *2025 13th International Conference on Information and Education Technology (ICIET)*, 160–164. https://doi.org/10.1109/ICIET66371.2025.11046309

[27] Rizwan, S., Nee, C. K., & Garfan, S. (2025). Identifying the Factors Affecting Student Academic Performance and Engagement Prediction in MOOC Using Deep Learning: A Systematic Literature Review. *IEEE Access*, *13*, 18952–18982. https://doi.org/10.1109/ACCESS.2025.3533915

[28] Romero, C., & Ventura, S. (2013). Data mining in education. *WIREs Data Mining and Knowledge Discovery*, *3*(1), 12–27. https://doi.org/10.1002/widm.1075

[29] Shen, T., & Zhang, Y. (2024). A Performance Prediction Model for Newly Enrolled University Students Based on The Fusion of K-Means and LightGBM Algorithms. *2024 3rd International Conference on Cloud Computing, Big Data Application and Software Engineering (CBASE)*, 40–44. https://doi.org/10.1109/CBASE64041.2024.10824518

[30] Sokkhey P & Okazaki T. (2020). Hybrid machine learning algorithms for predicting academic performance. *Int. J. Adv. Comput. Sci. Appl*, *11*(1), 32–41.

[31] Tripathi, A., & Reddy, K. (2025). Development of an Intelligent System for Predictive Analysis of Student Academic Performance Using Machine Learning Techniques. *2025 International Conference on Inventive Computation Technologies (ICICT)*, 1110–1117. https://doi.org/10.1109/ICICT64420.2025.11005033

[32] Wafi, M., Faruq, U., & Supianto, A. A. (2019). Automatic Feature Selection for Modified K- Nearest Neighbor to Predict Student's Academic Performance. *2019 International Conference on Sustainable Information Engineering and Technology (SIET)*, 44–48. https://doi.org/10.1109/SIET48054.2019.8986074

[33] Wang, Y. (2024). Educational Big Data Mining and Analysis Based on Support Vector Machine Performance Prediction Rate. *2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS)*, 1–5. https://doi.org/10.1109/ICIICS63763.2024.10859638

[34] Xiaoming, D., Ying, C., Xiaofang, Z., & Yu, G. (2022). Study on Feature Engineering and Ensemble Learning for Student Academic Performance Prediction. *International Journal of Advanced Computer Science and Applications*, *13*(5). https://doi.org/10.14569/IJACSA.2022.0130558