

# Towards Accurate Student Performance Prediction: An Assessment of Machine Learning Models and Metrics

Saurabh Charaya<sup>1</sup>, Sonia<sup>2\*</sup>

<sup>1</sup>School of Engineering & Technology, Om Sterling Global University, Hisar, Haryana (India)

<sup>2</sup>Ph.D. Scholar, School of Engineering & Technology, Om Sterling Global University, NH-52, Hisar-Chandigarh National Highway, Hisar-125001

\*Corresponding Author

DOI: <https://doi.org/10.51583/IJLTEMAS.2026.150100053>

Received: 20 January 2026; Accepted: 26 January 2026; Published: 06 February 2026

## ABSTRACT

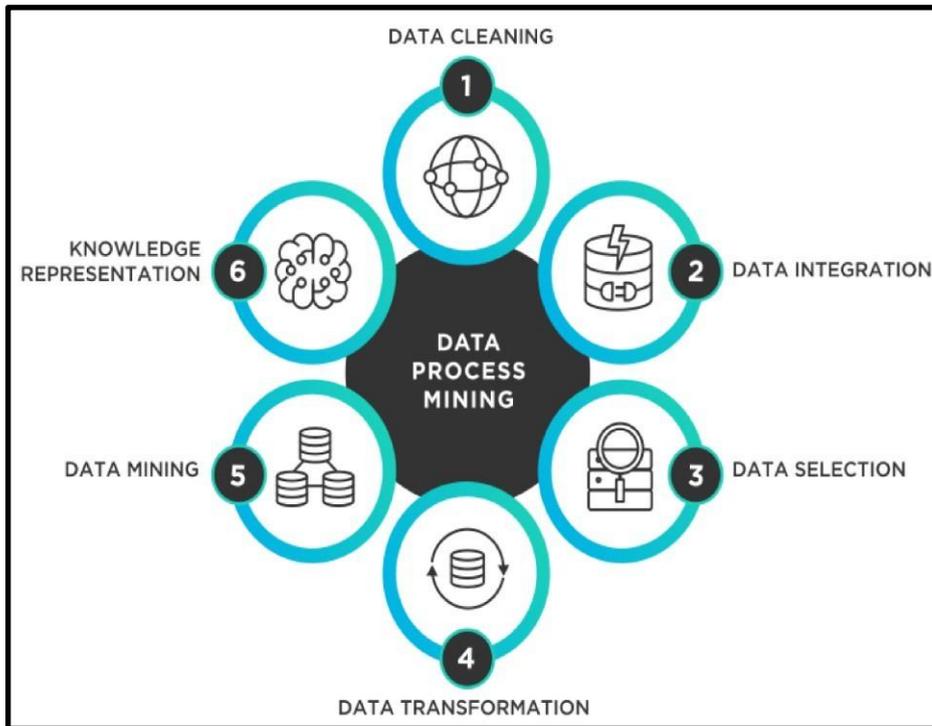
Predicting students' academic outcomes has become a central focus within Educational Data Mining (EDM) to support early academic interventions, minimize dropout risks, and promote personalized learning strategies. The availability of diverse educational data—ranging from demographic details and previous grades to behavioral engagement and digital learning activity—has encouraged the adoption of machine learning (ML) approaches for this purpose. This analysis synthesizes major ML techniques applied to student performance prediction, including regression models, decision trees, random forests, support vector machines, k-nearest neighbors, naïve Bayes classifiers, artificial neural networks, gradient boosting methods such as XGBoost and LightGBM, and deep learning models like CNNs and RNNs. Common evaluation measures—such as accuracy, precision, recall, F1-score, ROC-AUC, MAE, MSE, RMSE, and  $R^2$ —are also examined. Despite their potential, challenges persist, including inconsistent data quality, complex feature selection, privacy and ethical concerns, limited interpretability, poor generalization across institutions, and the need for frequent model updates. Future research should emphasize explainable AI (XAI), temporal modeling techniques, integration of behavioral and psychological indicators, and transfer learning to enhance scalability and adaptability. Overall, this study highlights the promise of ML-driven systems in improving educational outcomes when combined with ethical, interpretable, and context-aware design principles.

**Keywords:** Machine Learning, Student Performance Prediction, Data Mining, AI, Predictive Analytics

## INTRODUCTION:

Student performance prediction has become one of the critical research areas in educational data mining, as it facilitates early intervention approaches, identifies vulnerable students, and supports student-centered learning plans. As the problem of student dropout rates and low-level performance tends to gain increasing attention, proper academic performance prediction is one of the current priorities of education establishments on a global level (Kumar, 2024). The upsurge in the accessibility of educational data, both in quantity and in character, such as the student demographics, academic background, behavioral history, and engagement history, has created new opportunities to use machine learning (ML) methods to draw informed predictions. Machine learning models may be useful to extrapolate intricate patterns and unseen relations in data that might not be derived using customary statistical methods (Gul et al, 2025). In this paper, we will attempt to discuss and critically examine the different ML models that are usually applied in predicting student performance, as well as the criteria used to assess their overall performance. In this way, it will help to be able to give new directions to future work and practice in other educational environments to achieve better academic results. (Haziq & Chwen, 2022).

**Background and Related Work:** Educational Data Mining (EDM) is an emerging, but interdisciplinary, sub-discipline of data mining and machine learning, with the focus on teaching and learning through the application of data mining methods to educational data. The prediction of student performance has been given a lot of research has been conducted over the last ten years, which helps institutions to anticipate any academic problems and to improve student retention (Alnasyan et al.,2024; Hussain & Khan, 2023). The early researches were based on the traditional statistical models, and in recent years, the nature has changed to machine learning models because of the enhanced efficient performance in capturing the high-dimensional and considerably complex information.



**Figure 1: A Systematic Assessment assessment of Students’ Performance Prediction Using Machine Learning Techniques (Source: <https://www.mdpi.com/2227-7102/11/9/552>)**

Performance prediction has found application in various sectors of learning, such as K-12 learning institutions, mock colleges and universities, and Massive Open Online Courses (MOOCs). The types of these domains are different in their structure and rates of engagement in them, the point being that predictive analytics are useful in each of them to address the adjustments to interventions and enhance educational efficacy (Pelima et al., 2024). Some of the more common prediction variables are demographic information (i.e. age, gender, parental education), academic results (i.e., grades, exam scores), behavioural studies (i.e. attendance, participation), and digital records produced in educational management systems (i.e. frequency of login, submission patterns of assignments). The most recognizable datasets in this study are the UCI Student Performance Dataset and the Open University Learning Analytics Dataset (OULAD). The former consists of attributes that contribute to the academic performance of Portuguese students in school, whereas the latter provides full-fledged data on online learners. Such data sets are used as a test and comparison point for predictive models.

**Important Factors (Predictors) in Student Performance Prediction:** Many studies have examined what types of input features (predictors) most influence prediction accuracy. These broadly fall into several categories:

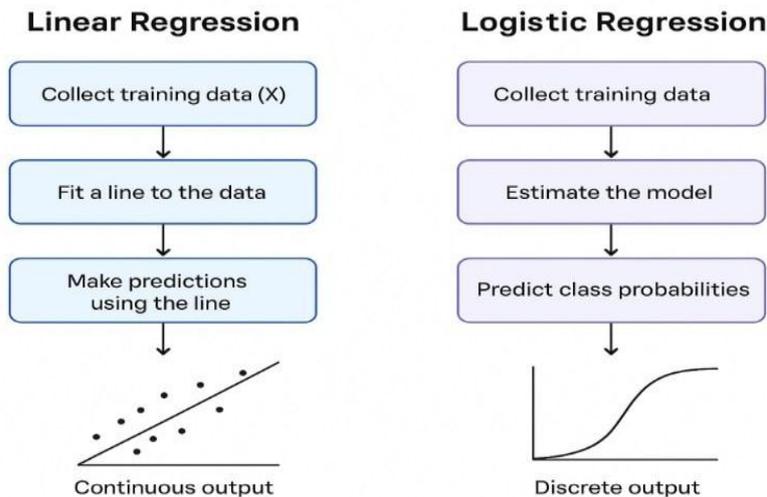
- **Demographic and Socio-economic Information:** This includes the factors like age, gender, parental education, household income, place of residence. These are relatively simple to collect and often show consistent correlation with student outcomes. (Haziq & Chwen, 2022)R

- **Academic History / Prior Performance:** Past grades, cumulative GPA, performance in earlier semesters, or performance in prerequisite courses are among the strongest predictors. For example, midterm grades are often used to predict final exam outcomes. (Yağcı , 2022)
- **Behavioural Engagement / Learning Activities:** Attendance, participation, assignment submission, frequency of login to Learning Management Systems (LMS), time spent on tasks, interaction logs (e.g. forum posts) etc. These provide dynamic and fine-grained information about how students engage with the learning materials. LMS data have been used in multiple studies to improve prediction accuracy.  
  
(Athanasios et al, 2025)
- **Course and Institutional Variables:** Course difficulty, teacher quality or pedagogical approach, class size, department, faculty, institutional resources, etc. These contextual variables can influence performance and sometimes interact with student attributes. (Yaosheng and Kimberly, 2025)
- **Psychological / Motivation / Cognitive Factor:** Less frequently used but gaining importance: students' motivation, self-regulation, prior domain knowledge, learning styles or skills. Some studies have proposed semantic or latent measures of domain knowledge to enhance early prediction.
- **Temporal / Dynamic Features:** How students' engagement or performance evolves over time (e.g., semester-wise performance, changes in behaviour over weeks). Temporal models can allow earlier warning of deteriorating performance and dynamic interventions (John et al, 2025)

#### **MACHINE LEARNING MODELS USED IN STUDENT PERFORMANCE PREDICTION:**

Machine learning models became efficient in the inference of student academic performance by learning the pattern using historical information. Every algorithm possesses its strengths and weaknesses regarding the characteristics of a dataset and the specific task of making predictions.

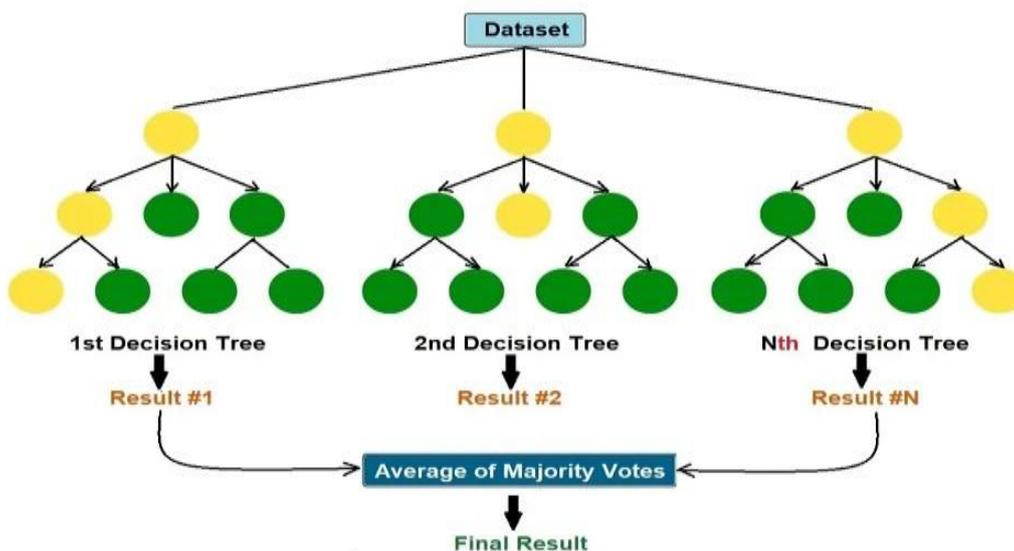
**Linear Regression / Logistic Regression:** Linear Regression is applied to the prediction of continuous outcomes such as grades, whereas Logistic Regression can be applied to the prediction of classification problems such as pass/fail outcomes. Such models are also interpretable and are computationally efficient. They, however, presume a linear association between variables, which restricts them from giving accurate readings in very complex situations (Zhu et al., 2024). Logistic regression has also been applied in the literature where students who were at-risk were identified in terms of past performance and attendance (Kim et al, 2022),



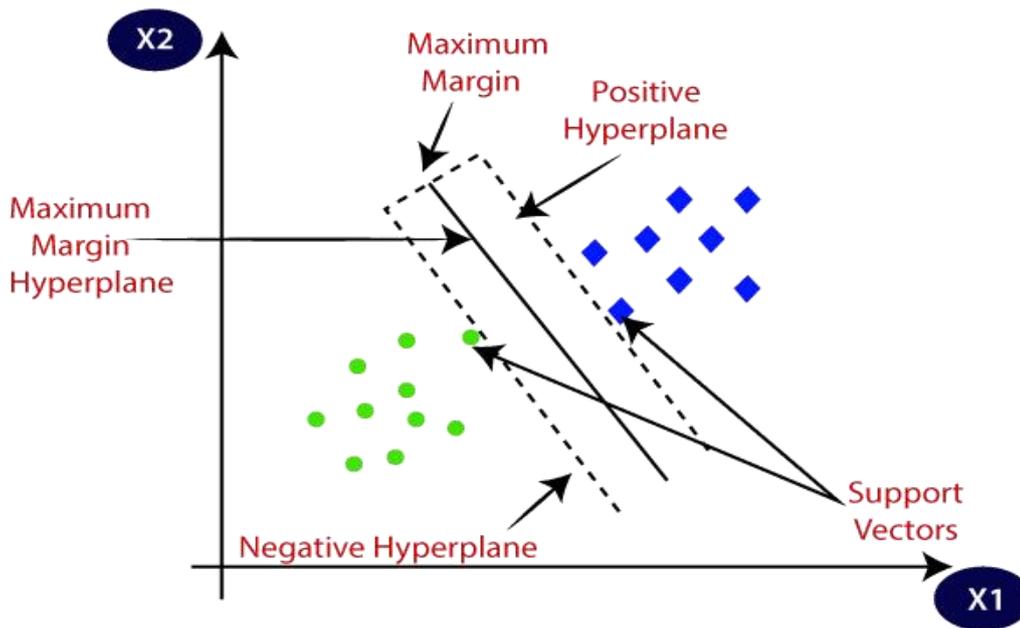
**Figure 2: Flow chart of Linear Regression / Logistic Regression**

**Decision Trees and Random Forests:** Decision trees are simple to interpret since the data is divided into features. Random Forests are more accurate because the outcomes of several trees are combined, thus giving fewer chances of overfitting (Wu et al., 2024). They process numerical data as well as categorical data and are noise resistant. Random Forests have also proved effective in educational studies, where they are used in classification problems, including predicting school dropout cases. (Natarajan, 2024; Kotsiantis, 2012)

**Support Vector Machines (SVM):** SVM determines the best hyperplane that separates classes. It is efficient to use with high-dimensional data and a binary classification task. But it has an advantage that it is computationally intensive and less interpretable. SVM has been applied in the prediction of success among the students attending MOOCs, where the feature spaces are large. (Boutahri & Tilioua, 2024). SVM is a supervised learning algorithm. It works well for binary classification and also supports multi-class classification using one-vs-one or one- vs -rest strategy. Margin maximization makes it less prone to over fitted.

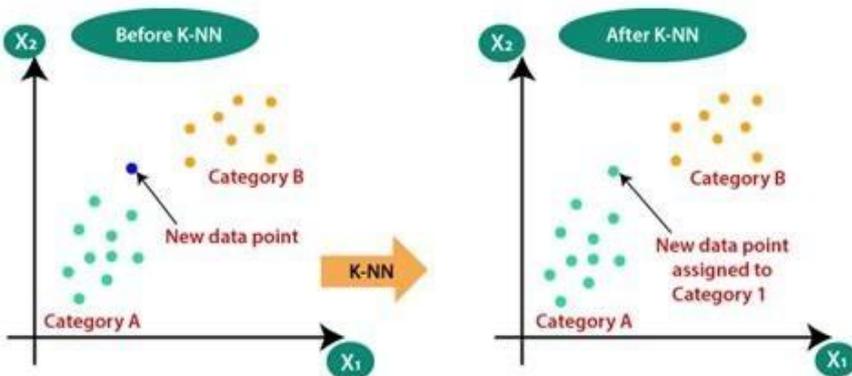


**Figure 3: Flow chart of Decision Trees and Random Forests**



**Figure 4: Flow chart of Support Vector Machines**

**K-Nearest Neighbors (KNN):** KNN is a non-parametric method that is used to classify a data point according to the highest percentage of the type of its nearest neighbors. It is easy, but efficient when it comes to small data sets. Nevertheless, it is too time-consuming with a huge set, and it is also sensitive to irrelevant features. KNN has been implemented to forecast the student scores given resemblance to the previous pupils. Working of kNN



**Figure 5: Flow chart of K-Nearest Neighbors**

**Naive Bayes:** According to this probabilistic model, the independence of features and the Bayes theorem are used. It is also efficient and performs well in terms of text or categorical data. Although it makes very strong assumptions, it has performed well in educational applications in early warning systems and sentiment analysis. (Peling et al 2017)

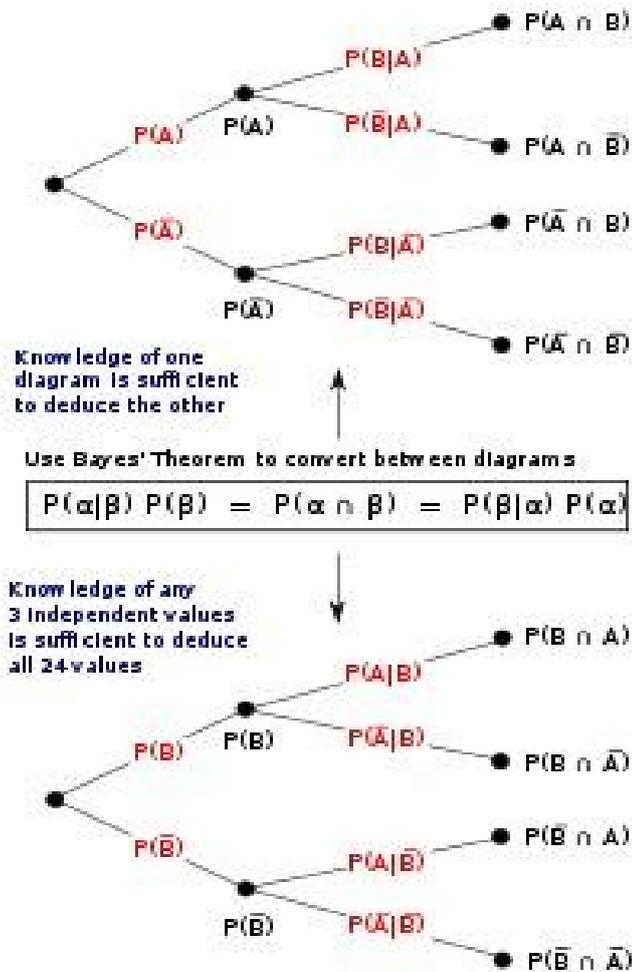
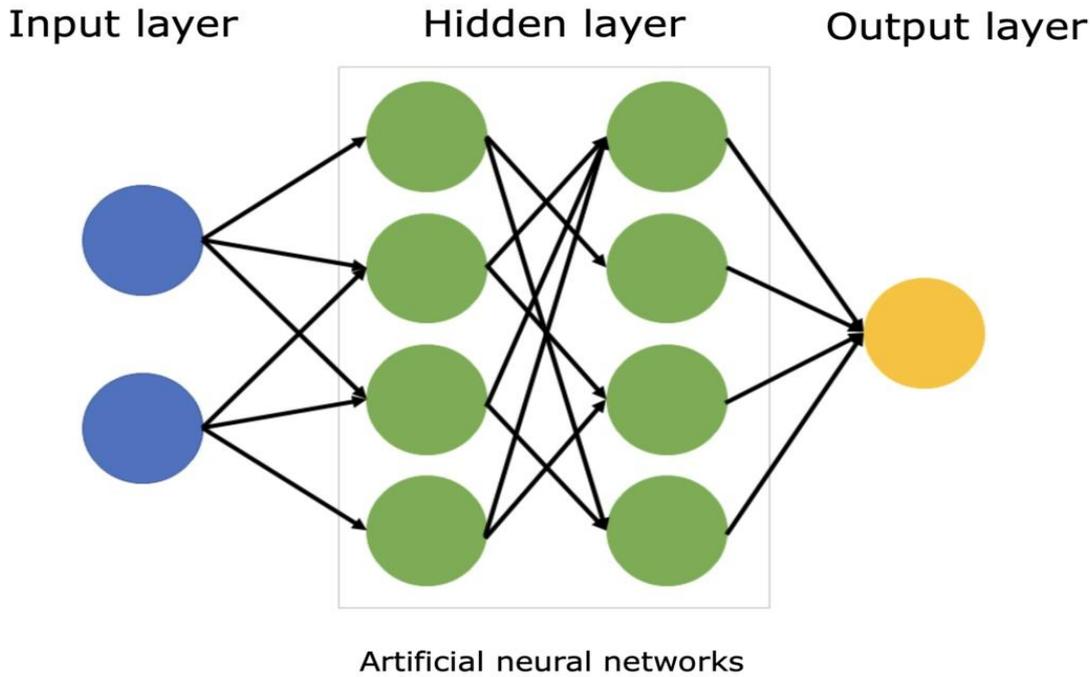


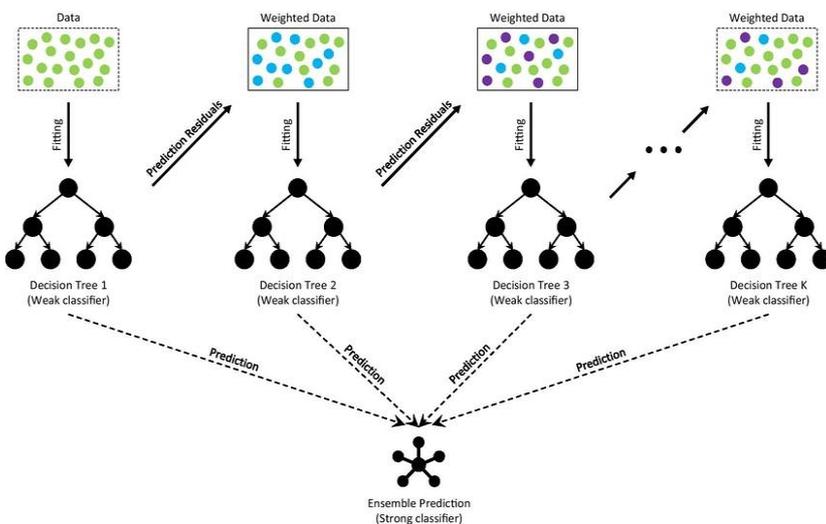
Figure 6: Flow chart of Naive Bayes

**Artificial Neural Networks (ANN):** ANNs imitate the way the brain functions by having neurons in it and can take in non-linear interactions. They are potent and need huge amounts of data and may be referred to as black boxes (Bressane *et al.*, 2024). ANNs have been used especially and quite effectively in predicting the grades of students and the risk of their dropping out, especially when the input features are many. (Ahmed, 2024)



**Figure 7: Flow chart of Artificial Neural Networks**

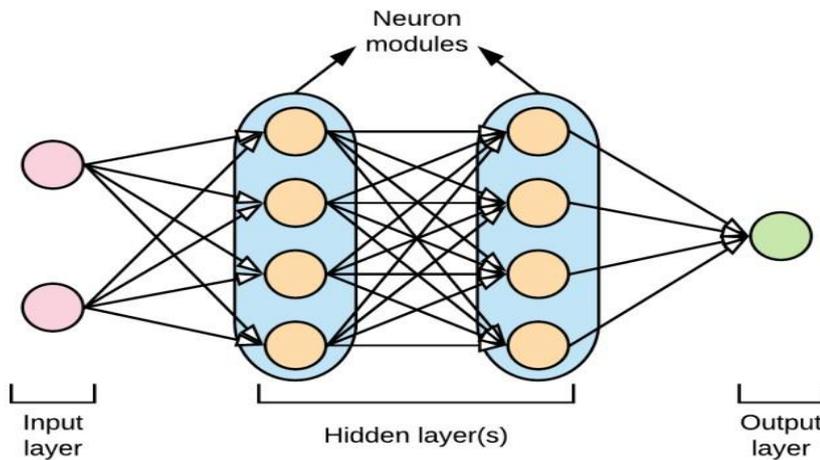
**Gradient Boosting XG Boost / Light GBM:** Such ensemble methods train sequentially in order to eliminate errors in earlier models. They are highly performative and deal with missing data and feature interactions adequately. The XGBoost model has better results on numerous educational datasets, and thus it is commonly used in competitions and as a research topic. Gradient Boosting is an ensemble machine learning method that builds a strong model by combining many weak learners (usually decision trees). Each new tree tries to correct the errors (residuals) made by the previous trees, using gradient descent optimization to minimize the overall loss.



**Figure 8: Flow chart of Gradient Boosting XGBoost / Light GBM**

**Deep Learning (CNN, RNN):** Recurrent Neural Networks (RNN) Deep learning models, particularly RNNs, are the best-suited to sequential data like time-stamped learning activities. The use of CNNs is limited unless one is handling educational material that relies on images (Oyucu *et al.*, 2024). RNNs have recently had success at modeling student progress in time-series learning settings, some of which is modeled by LMS clickstream data. (Archana & Jeevaraj, 2024; Banita et al, 2024)

**The Evaluation Metrics in Performance Prediction:** The effectiveness of the machine learning models in predicting the performance of the learners, one of the essential measures is to use appropriate measures that depend on the type of prediction, either regression or classification. In the case of regression tasks, when one tries to predict the outcomes that are continuous values rather than classes, the following metrics are the most common



**Figure 9: Flow chart of Deep Learning**

**Mean Absolute Error (Mae):** It Is the Average of the Size of the Errors as Opposed to Their Direction.

- **Mean Squared Error (MSE):** The larger errors are penalized much more, and hence it is outlier sensitive.
- **Root Mean Squared Error (RMSE):** The Square root of MSE, which scales the error back to the original unit.
- **R2 Score (coefficient of determination):** It shows how effectively the model can explain variance in the outcome, and the higher the value, the better the performance.

The most common criterion used in classification tasks, like the prediction of pass/fail or dropout risk, is:

- **Accuracy:** Percentage of predictions that are correct; however, it can be problematic when presented with unequal data.
- **Precision:** The ratio of true positives to the total predicted positives; of benefit when false positives are expensive.
- **Recall (Sensitivity):** True positive/actual positive; significant to find out at-risk students.
- **F1-score:** This is the harmonic mean of precision and recall, and is hence a compromise between the two.
- **ROC-AUC:** The standard of confusion in the two distinct classes over thresholds of the model.

The most important part is the choice of the metric to use, even in the case of education, where the data are frequently imbalanced (e.g., very few dropouts when compared to a larger number of successful students). Such measures as the F1-score and ROC-AUC are more trustworthy in that scenario. A reflective choice makes the model applicable in real-life education interventions.



**Figure 10: Analyzing and Predicting Students' Performance by Means of Machine Learning (Source: <https://www.mdpi.com/2076-3417/10/3/1042>) ISSUES AND CONSTRAINTS:**

Although machine learning has potential in forecasting student outcomes, a number of issues have remained. The quality of data is a big issue; the educational data that one might encounter lacks some data, contains inconsistent or noisy data, and data that may distort model predictions (Idowu *et al.*, 2024). Along with that, feature selection is tricky, and not all the variables (e.g., attendance, engagement) are equally useful, and irrelevant features may worsen the model accuracy. Ethical issues and privacy are also of great concern. There are controversial issues with sensitive personal information in educational data, such as concerns about what to consent to, whether there is security, and misuse of information. Unintentionally, models may introduce or enhance gender, socioeconomic status, or ethnic biases and then make unfair educational decisions (Sarker, 2021)

The second issue is that of model interpretability. Such sophisticated models as neural networks or ensemble methods are frequently black-boxes, and at least the teacher might not trust such results, and he/she would not be able to take any action based on it. In this respect, explainable AI is a work in progress (Ikegwu *et al.*, 2024). Finally, most of the models are not generalized. A model trained in one institution might not work in another because of other aspects such as different curriculum, grading, or demographics. The limitations mentioned above are steps to be taken concerning their clearance of the predictive models' accurate and morally right at an educational institution.

Although machine learning (ML) has considerable potential in forecasting student outcomes, several critical issues continue to impede its practical application in educational settings.

**Data Quality Issues:** The quality of data remains a major challenge. Educational datasets often contain missing values, inconsistent records, and noisy or erroneous data, all of which can distort model predictions (Idowu *et al.*, 2024). Data is typically collected from multiple heterogeneous sources such as learning management systems, attendance registers, and online assessments, and this heterogeneity increases the difficulty of preprocessing and integration. Without proper data cleaning and validation mechanisms, the reliability and reproducibility of predictions become questionable.

**Feature Selection Complexity:** Feature selection is another complex and critical aspect. Not all variables (e.g., attendance, engagement, demographic details) are equally useful, and the inclusion of irrelevant or redundant features can degrade model accuracy and increase computational burden. Manual feature engineering requires

domain expertise and is time-consuming, while automated methods may fail to capture contextual nuances. This imbalance contributes to inconsistent model performance across different datasets.

**Ethical and Privacy Concerns:** Ethical issues and data privacy are also significant. Educational datasets frequently contain sensitive personal information, raising concerns about informed consent, secure data storage, and potential misuse. There is a risk that models might unintentionally incorporate or amplify biases based on gender, socioeconomic status, or ethnicity, thereby producing unfair or discriminatory predictions. Such ethical lapses could negatively impact students' educational opportunities and trust in technology.

**Lack of Model Interpretability:** The interpretability of ML models is another key challenge. Sophisticated algorithms such as neural networks or ensemble methods are often considered "black boxes." Teachers and administrators may hesitate to trust these predictions if they do not understand how the model arrives at its decisions. This limits the models' practical utility for guiding interventions. Explainable AI (XAI) methods are emerging to address this, but they are still evolving and not yet widely integrated into educational systems (Ikegwu et al., 2024).

**Limited Generalizability:** Most ML models are not generalizable. A model trained in one institution may not perform well in another due to differences in curricula, grading systems, teaching styles, or demographic factors. This lack of external validity restricts large-scale adoption and scalability, as models often require complete retraining when moved to new contexts.

**Need for Continuous Updating:** Educational environments are dynamic, and changes in curricula, assessment patterns, or student behavior can cause model performance to deteriorate over time (concept drift). However, most institutions lack mechanisms for continuous model evaluation and updating, which threatens the long-term accuracy and reliability of predictions.

#### **Future Directions:**

Integration of Explainable AI (XAI), which would assist educators to understand the rationale behind their predictions and hence drive trust and accountability, is one of the potential fields. Study time and time-series modeling as well as LSTM (Long Short-Term Memory) networks, are popular in temporal modeling, particularly in online learning tasks where student dynamics change over time (Hussain *et al.*, 2024). Learning trajectories can be captured with such models and offer timely interventions.

The next important trend is integrating behavioral and psychological characteristics, including motivation, stress levels, and learning style, with the help of a survey or wearable data. These have the potential to complement models with more records than academia to provide a more subsumed picture of student success.

The transfer and few-shot learning may be utilized to make the models generalizable across various institutions, especially when the amount of data is small. The methods allow pre-trained models to adjust with small sets of data and a little amount of work (Manzoor *et al.*, 2024). Finally, the next phase of systems should be forecasted on the personalized prediction and adaptive learning environment, and they should target suggestions and materials according to the needs of particular students.

## **CONCLUSION**

This paper shows the increased importance of machine learning in the precise prediction of student performance, which provides worthwhile resources in early alerts, resource planning, and education planning. Many different models, including standard logistic regression, all the way to high-value deep learning approaches, have been promising with their special advantages dependent upon the available data scenario. Almost as significant is the selection of the evaluation measures, especially when dealing with imbalanced data, which is common in education.

There are several challenges to tackle, including the quality of data, ethical issues, and the interpretation of the model that should facilitate the responsible and successful implementation. In the future, a combination of explainable AI, temporal modeling, and behavioral data can play a great role in further complementing the model's relevance and accuracy. In the end, a combination of machine learning processes and learning objectives can contribute to the creation of more personalized, inclusive, and outcome-focused learning experiences in cases of students at all levels of education

## REFERENCES

1. Kumar, S., Kumar, H., Kumar, G., Singh, SP., Bijalwan, A., Diwakar, M., 2024. A methodical exploration of imaging modalities from dataset to detection through machine learning paradigms in prominent lung disease diagnosis: a review. *BMC Medical Imaging*, 24(1), 30. <https://link.springer.com/content/pdf/10.1186/s12880-024-01192-w.pdf>
2. Haziq, R., Chwen, JC., 2022. Educational Data Mining for Student Performance Prediction: A Systematic Literature Review (2015-2021). *International Journal of Emerging Technologies in Learning*, 17(05), 147-179. DOI: 10.3991/ijet.v17i05.27685
3. Alnasyan, B., Basher, M., Alassafi, M., 2024. The power of Deep Learning techniques for predicting student performance in Virtual Learning Environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, 100231. <https://www.sciencedirect.com/science/article/pii/S2666920X24000328>
4. Pelima, LR., Sukmana, Y., Rosmansyah, Y., 2024. Predicting university student graduation using academic performance and machine learning: a systematic literature review. *IEEE Access*, 12, 2345123465. <https://ieeexplore.ieee.org/iel7/6287639/6514899/10418878.pdf>
5. Yağcı, M., 2022. Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(11).
6. Athanasios, A., John, A., Markos, K., Alkiviadis, T., 2025. Predicting Student Performance and Enhancing Learning Outcomes: A Data-Driven Approach Using Educational Data Mining Techniques. *Computers*, 14(3), 83. <https://doi.org/10.3390/computers14030083>
7. Yaosheng, L., Kimberly, FC., 2025. Performance prediction using educational data mining techniques: a comparative study. *Discover Education*, 4(12), 112.
8. Zhu, Y., Ma, Y., Fan, F., Huang, J., Wu, K., Wang, G., 2024. Towards accurate infrared small target detection via edge-aware gated transformer. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. <https://ieeexplore.ieee.org/iel7/4609443/4609444/10496183.pdf>
9. John-Bosco, D., Icon, Mekala, MS., Ulric, SA., John, I., Eyad, E., 2025. Information & Communications Technology in Education Predicting student next-term performance in degree programs using AI-based approach: a case study from Ghana, *Cogent Education*, 12(1). <https://doi.org/10.1080/2331186X.2025.2481000>
10. Wu, S., Cao, Y., Cui, J., Li, R., Qian, H., Jiang, B., Zhang, W., 2024. A comprehensive exploration of personalized learning in smart education: From student modeling to personalized recommendations. *arXiv preprint arXiv:2402.01666*. <https://arxiv.org/pdf/2402.01666>
11. Natarajan, SK., Shanmurthy, P., Arockiam, D., Balusamy, B., Selvarajan, S., 2024. Optimized machine learning model for air quality index prediction in major cities in India. *Scientific reports*, 14(1), 6795. <https://www.nature.com/articles/s41598-024-54807-1.pdf>
12. Boutahri, Y., Tilioua, A., 2024. Machine learning-based predictive model for thermal comfort and energy optimization in smart buildings. *Results in Engineering*, 22, 102148. <https://www.sciencedirect.com/science/article/pii/S259012302400402X>
13. Bressane, A., Zwirn, D., Essiptchouk, A., Saraiva, ACV., de Campos Carvalho, FL., Formiga, JKS., Negri, RG., 2024. Understanding the role of study strategies and learning disabilities on student academic performance to enhance educational approaches: A proposal using artificial intelligence. *Computers and Education: Artificial Intelligence*, 6, 100196. <https://www.sciencedirect.com/science/article/pii/S2666920X23000759>

15. Ahmed, QO., 2024. Machine learning for intrusion detection in cloud environments: A comparative study. *Journal of Artificial Intelligence General Science (JAIGS)*. [https://www.researchgate.net/profile/Qazi-Omair-Ahmed/publication/387076339\\_Machine\\_Learning\\_for\\_Intrusion\\_Detection\\_in\\_Cloud\\_Environments\\_A\\_Comparative\\_Study/links/676173c2e9b25e24af5c961a/Machine-Learning-for-Intrusion-Detectionin-Cloud-Environments-A-Comparative-Study.pdf](https://www.researchgate.net/profile/Qazi-Omair-Ahmed/publication/387076339_Machine_Learning_for_Intrusion_Detection_in_Cloud_Environments_A_Comparative_Study/links/676173c2e9b25e24af5c961a/Machine-Learning-for-Intrusion-Detectionin-Cloud-Environments-A-Comparative-Study.pdf)
16. Oyucu, S., Doğan, F., Aksöz, A., Biçer, E., 2024. Comparative analysis of commonly used machine learning approaches for Li-ion battery performance prediction and management in electric vehicles. *Applied Sciences*, 14(6), 2306. <https://www.mdpi.com/2076-3417/14/6/2306/pdf>
17. Archana, R., Jeevaraj, PE., 2024. Deep learning models for digital image processing: a review. *Artificial Intelligence Review*, 57(1), 11. <https://link.springer.com/content/pdf/10.1007/s10462-023-10631-z.pdf>
18. Idowu, JA., 2024. Debiasing education algorithms. *International Journal of Artificial Intelligence in Education*, 1-31. <https://link.springer.com/content/pdf/10.1007/s40593-023-00389-4.pdf>
19. Ikegwu, AC., Nweke, HF., Anikwe, CV., 2024. Recent trends in computational intelligence for educational big data analysis. *Iran Journal of Computer Science*, 7(1), 103-129. [https://www.researchgate.net/profile/Henry-Nweke-3/publication/373738289\\_Recent\\_trends\\_in\\_computational\\_intelligence\\_for\\_educational\\_big\\_data\\_analysis/links/67c74dc28311ce680c7cb159/Recent-trends-in-computational-intelligence-for-educationalbig-data-analysis.pdf](https://www.researchgate.net/profile/Henry-Nweke-3/publication/373738289_Recent_trends_in_computational_intelligence_for_educational_big_data_analysis/links/67c74dc28311ce680c7cb159/Recent-trends-in-computational-intelligence-for-educationalbig-data-analysis.pdf)
20. Hussain, T., Yu, L., Asim, M., Ahmed, A., Wani, MA., 2024. Enhancing e-learning adaptability with automated learning style identification and sentiment analysis: a hybrid deep learning approach for smart education. *Information*, 15(5), 277. <https://www.sciencedirect.com/science/article/pii/S1532046424000406>
21. Manzoor, A., Qureshi, MA., Kidney, E., Longo, L., 2024. A review on machine learning methods for customer churn prediction and recommendations for business practitioners. *IEEE access*. <https://ieeexplore.ieee.org/iel7/6287639/6514899/10531735.pdf>
22. Gul, MN., Waseem, A., Muhammad, ZB., Abeer, A., Muhammad, A., 2025. Data driven decisions in education using a comprehensive machine learning framework for student performance prediction, 28, 153. <https://doi.org/10.1007/s10791-025-09585-3>
23. Baniata, LH., Kang, S., Alsharaiah, MA., Baniata, MH., 2024. Advanced deep learning model for predicting the academic performances of students in educational institutions. *Appl Sciences*, 14(5), 1963. <https://doi.org/10.3390/app14051963>
24. Hussain, S., Khan, MQ., 2023. Student-performulator: predicting students' academic performance at secondary and intermediate level using machine learning. *Ann Data Science*, 10(3), 637–55. <https://link.springer.com/article/10.1007/s40745-021-00341-0>
25. Sarker, IH., 2021. Machine Learning: Algorithms, real-world applications and research directions. *SN Comput. Science*, 2, 160. <https://link.springer.com/article/10.1007/s42979-021-00592-x>
26. Kotsiantis, SB., 2011. Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. *Artificial Intelligence Review*, 37(4), 331–344. <https://doi.org/10.1007/s10462-011-9234-x>
27. Kim, SJ., Bae, SJ., Jang, MW., 2022. Linear Regression Machine Learning Algorithms for Estimating Reference Evapotranspiration Using Limited Climate Data. *Sustainability*, 14, 11674. <https://doi.org/10.3390/su141811674>
28. Peling, IBA., Arnawan, IN., Arthawan, IPA., Janardana, IGN., 2017. Implementation of data mining to predict period of students study using naive bayes algorithm. *International Journal of Engineering and Emerging Technology*, 2(1), 53. [10.24843/IJEET.2017.v02.i01.p11](https://doi.org/10.24843/IJEET.2017.v02.i01.p11)