www.ijecs.in
International Journal of Engineering and Computer Science
Volume 14 Issue 02 February 2025, Page No. 26865-26870
ISSN: 2319-7242 DOI: 10.18535/ijecs/v14i02.4996

Students performance analysis using machine learning

Preethi Vajja [1], Dr. Srivaramangai Ramanujam [2*]

[1] University Department of Information Technology, University of Mumbai, Mumbai, India.

[2] University Department of Information Technology, University of Mumbai, Mumbai, India.

Abstract

Modern machine learning techniques create fresh footprints in the education landscape by utilizing predictive analytics for the student's performance assessment. Unlike standard assessments of academic performance which are rigid and subject to bias against the individual student, machine learning defines problems in predicting student performance and applies the different scales over these very narrow margins. Unstandardized assessment methods, the difference in the ways of learning by students, and feedback not given in real-time to guide assessments are among the factors that affect student performance. The data availability, feature selection, one of the most important challenges, and interpretation of the model are also identified as some of the most critical challenges. Also, these grading systems have the limitations that urge a shift toward automated, data-driven methods of clear improvements in predictive accuracy. Efficiency of prediction brought about by advanced machine learning techniques provides a basis on which one can forecast the academic performances of students. They use neural networks, decision trees and other ensemble methods through more efficient analysis. It also improves usage with precision by predicting academic results applicable to different student demography, besides being user-friendly. Establishing the prediction model of student performance will need a serious analysis of the prevailing challenges and also practical machine learning solutions. The study results will greatly contribute to improving the educational strategies, enhance the learning experiences of students, and optimize academic decision-making.

Keywords: Student performance prediction, machine learning in education, predictive analytics, academic success factors, data-driven learning, educational data mining, artificial intelligence in education, automated grading systems, learning behavior analysis, and model interpretability in education.

1. Introduction

For ages, student performance has stood as a prime axis of education systems for improving teaching styles and learning outcomes. It is based on static evaluations and, hence, cannot give an extensive view of student progress by traditional methods of assessments. Artificial intelligence and machine learning have worked wonders in the ability of predictive analytics to offer deep insights into students' performances, trends, and possible interventions involving personalized learning experiences. Various factors among others influence student performance: Academic records include attendance, level of engagement with education, and socio-economic background. More traditional evaluation techniques are adaptive in a way they would tend to generalize the assessment instead of specifying winter needs associated with his or her learning. Machine learning seems to propose giving a predictive pattern in student performance analysis considering data perspectives and recognizing associated patterns that could be impacted with

highlighting areas of improvement and thus suggesting tailor-made learning strategies. Some important challenges include issues such as data quality, interpretability of the model, and privacy even in a promising machine learning method applied to education. While developing a robust predictive model, special attention should also be given to security of data and ethical use of data so as to preserve trust in such analytical systems. Additionally, another directed dimension is the machine learning algorithms and feature sets that ensure high accuracy and reliability for performance predictions. This paper discusses current issues and challenges in analyzing student performance, as well as the relevance of machine learning techniques in mitigating these problems. The study, therefore, focuses on proposing an efficient model that would accurately predict a student's outcome for making good decisions in education and improving educational quality. This study therefore helps understand how technology can improve learning by optimizing different ways for students' success in technology-driven methods. This paves way for a better-prepared learning environment to support an increasingly diverse range of academic needs.

2. Literature Review

This study, on the face of it, engages machine learning and it bringing to bear on analysis of student performance. With the advancing age, many mechanisms have been researched, widely, to make predictions more accurate about the evolution of decisions. A profusion of machine learning mechanisms had been brought to bear to understand the variables affecting academic performance and thereby generating intervention strategies accordingly. Phauk and Okazaki [1] proposed hybrid models of machine learning to predict students' performance. It presents that adding principal component analysis (PCA) with other machine-learning algorithms is efficient for increasing accuracy in predictive. Cruz-Jesus et al. [2] pointed that artificial neural networks (ANNs), decision trees (DTs) and support vector machines (SVM) greatly help in modeling performance assessment towards a more empirical approach in the evaluation of academic performance. Bithari et al. [3] have used ensemble voting method providing a far better predictor than the single model. They concluded from the resultant proof that the ensemble techniques are capable of overcoming the prediction error and improving the reliability of prediction models. Oppong [4] thus concludes that neural networks stand tall among methods in student performance prediction. The indicatives of their study showed that deep learning models had rendered a better capture of what have been thought to be complex relationships in student data than traditional statistical methods. The combination of decision trees, naive bayes classifiers, and logistic regression methods thus laid the foundation for even-handed comparison and proves that hybrid models perform better. Classification performance was determined for Naive Bayes, Random Forest, and SVM by Ogwoka et al. [5]. Both Naive Bayes and SVM demonstrated the highest classification performance. Okereke et al. [6] scrutinise the prediction framework based on decision tree for student performance along with the techniques of preprocessing the data. Feature selection has been made the point in emphasis that improves the accuracy of the predictive model while adding that student's demographic and academic characteristics need to be such prime considerations in performance analysis. Urkude and Gupta . [7] advocate, therefore, for the construction of a Student Intervention System on the Support Vector Machine (SVM), Decision Tree, and Naïve Bayes algorithms, thus demonstrating that the SVM-based model classifies students at-risk of underperforming better than any investigators looked at. Performance Prediction Model for Students is under development through Decision Tree, Naïve Bayes, and Logistic regression Techniques, stated Hashim et al. [8]. It created a predictive model with a conclusion that Logistic Regression was good at predicting final grades while Decision Tree models efficiently gave high explanatory insight into feature importance, thus believing that practical applicability with much better prediction probability has to come from the combination.

Iqbal et al. [9] recognized machine learning in predicting students' grades. According to them, collaborative filtering could have effective personalization in proposing recommendations for every student, depending on the historical performance data the model was trained on. Echoing the aforementioned authors, the model can also be very much useful for working within the adaptive learning platform beyond just academic assistance on students' individual needs. Ojajuni et al. [10] employed Extreme Gradient Boosting (XGBoost), which is a machine learning algorithm, for higher accuracy prediction of students' academic performance over their traditional methods. The results further established that ensemble learning techniques would be competent in handling complex datasets from the education perspective. Deep Neural Networks were the most successful models out of many models that were seen by Vijayalakshmi et al. [11] in the prediction of student performance through machine learning. This study has its beginnings in the academic patterning by AI-based applications. The best support vector machines (SVM) model at 96% accuracy stated Ahmed in the report [12] was sound for predicting students' performance using machine learning. The paper elaborates the importance of choosing an appropriate classification model in educational analytics. Albreiki et al. [13] systematically reviewed the literature on student performance prediction and hence endorsed that machine learning approaches, particularly with regard to Decision Trees, are the current flavour of education, and further went on to comment on the trend among students towards AI-based tools fostering personalized learning. Rahman et al. [14], then, pushed forward with reviews of potential uses of artificial intelligence in which it is majorly machine learning predictive application within the performance of students. Their study had addressed different approaches regarding supervised learning methods, which included decision trees and linear models and were backed to academic context under which maximization of learning output would be attained. This works as the evidence by Al Mayahi et al. [15], who proposed a machine learning prototype for prediction of student performance for learning that indeed had showed SVC and Elastic Net performing well in the prediction of success of students or student success prediction. This shows that also these modelling approaches could enhance the prediction accuracy. To validate the above predictions, Orji et al. [16] incorporated psychological factors that affect modeling students' academic performance by machine learning techniques into their study. The attempt aims at creating a ground that abides cognitive-behavioral factors with AI techniques towards better understanding in the performance analysis of students. Dake et al. [17] experimented with Random Forest and Naive Bayes as preferred classification algorithms to ascertain the academic performance of students. This paper clearly illustrates how machine learning predictive models can assist the instructor in predicting the risk posed by students from the early beginning of the semester. Kalpana et al. [18] reviewed a number of machines learning algorithms, including Linear Regression, Support Vector Machines, etc., for predicting student performance more accurately. They analyzed that selection of features and hyper-parameter tuning has a great impact on prediction accuracy. Onker et al. [19] suggested supervised learning algorithms like Decision Trees and Random Forests to generate academic recommendations based on the attributes of student. They were convinced that the use of different models would optimize generalization and minimize bias to produce reliable predictions. Issah et al. [20] presented a systematic review of 84 works state the application of machine learning techniques in predicting student performance. They classified the techniques as Decision Trees, Random Forests, and Naive Bayes and presented comparative studies of their effectiveness. They concluded that ensembles generally outperform classification from individual methodologies.

3. Observations

It highlights some interesting observations based on existing literature regarding the analyses of the student performances using machine learning,

such as the deviations very intentionally made by the field with respect to diverse implementations for various machine learning algorithms and ensemble and hybrid approaches. The review of literature contains countless research papers which were recognized within the last decade with a focus on student performance analysis. Different methods of machine learning dominate the studies, and amongst those, Decision Trees, Logistic Regression, Naive Bayes, Support Vector Machines (SVMs), and Artificial Neural Networks (ANN) have been mostly studied concepts. Studies look forward to increasing predictive accuracy, thus identifying key elements. The approaches connected to such endeavors are including a combination between ensemble models and feature selection techniques like Principal Component Analysis (PCA) and the hybridization sometimes of different algorithms to be more powerful in prediction. By the way, nowadays uses of those techniques like Voting and Random Forest in terms of robustness and excellent performance have increased. Even if some progress has been made, problems still exist, such as cases of imbalanced data sets, absence of standardized metrics for evaluating various designs, and uncertainty plus interpretability of model predictions for educationalists and those stakeholders. Hyperparameter tuning, therefore, stands as a developing area in question, together with steps addressing overfitting as well as integrating explainable AI means, which make models more accountable and trustworthy for academic institutions. Another pretty amazing observation here is the fact that a number of works have been directed toward the automated predictive pipelines, rendering any human intervention in the analysis functions rather inefficient for the educational sector. Multi-source data fusion techniques were also considered for more reliable prediction performance based on academic records, attendance, socioeconomic status, and psychometric assessments. This mainline trend more optimally channels future studies toward interpretable machine learning paradigms, as well as embedding the domain knowledge of the lecturers and mending the data collection practices to yield performance information that would be accurate and thus actionable for better impact on educational policy and student strategies toward success.

4. Conclusion

The contested research tackles some challenges that are tied to student performance analyses in the era of intervention through advanced machine learning. These are imbalanced datasets, personalized learning strategies, feature selection issues, and predictive models' interpretability, which could limit the student's performance under sorry assumption. Poor assessment leads to failure in early identifying at-risk students and hampers remedial measures for improving their academic performance. And for that, a machinery analysis system, along with an action plan that employs machine learning, could harness gaps as early warning systems, real-time tracking of performance, adaptive learning insights, and intelligent feedback mechanisms. It should also have user-friendly analytical dashboards with actionable recommendations for educators and personalized learning pathways for students. Such systems, then, would require keeping evolving in a sustainable manner to keep their relevance by leveraging AI-powered predictive analytics, automated feature engineering, adaptive learning, in the cloud, and so on. Future work aims to improve the transparency of models should also explore raw multimodal data pools, like behavior and emotion analysis, to enhance student profiling and implement adequate biases mitigation techniques on educational data. AI-sustained innovation and further enhancement toward performance analyses will remain the heart of a future data-driven personalized learning environment that is equitable.

5. References

1. Sokkhey Phauk, Takeo Okazaki, 2020. Hybrid Machine Learning Algorithms for Predicting Academic Performance. Publication- International Journal of Advanced Computer Science and Applications (IJACSA) 11(1): 32-41.

2. Use Artificial Intelligence Methods to Assess Academic Achievement in Public High Schools of a European Union Nation. 2020. The Publication- Heliyon 6(6): e04081 Frederico Cruz-Jesus, Mauro Castelli, Tiago Oliveira, Ricardo Mendes, Catarina Nunes, Mafalda Sa-Velho, Ana Rosa-Louro.
3. Prediction of Academic Performance of Engineering Students Using Ensemble Method by Tek Bist Bithari, Sharan Thapa, Hari K.C. 2020. Publication: Technical Journal 2(1): 89-98, Nepal Engineers Association, Gandaki Province.
4. Stephen Opoku Oppong. Machine Learning Algorithms for Predicting Students' Performance: A Review. 2023. Publication: Asian Journal of Research in Computer Science 16(3): 128-148.
5. Thaddeus Matundura Ogwoka, Prof. Robert Obwocha Oboko, Prof. Christopher Kipchumba Chepken. 2024. Towards: Comparative Analysis of Machine Learning Classifier Models for Predicting Student Cognitive Load and Performance Outcomes in Moodle Learning Environment. Publication: East African Journal of Information Technology 7(1): 301-317.
6. Okereke GE, Mamah CH, Ukekwe EC, Nwagwu HC. 2020. Machine Learning Based Framework for Predicting Student's Academic. Publication -Physical Science Journal & Biophysics 4(2): 000145.
7. [7] Urkude Shubhangi, Kshitij Gupta. 2019. Student Intervention System Using Machine Learning Techniques. Publication- International Journal of Engineering and Advanced Technology (IJEAT) 8(6S3): 2061-2065.
8. Ali Salah Hashim; Wid Akeel Awadh; Alaa Khalaf Hamoud. 2020. "Student Performance Prediction Model Based on Supervised Machine Learning Algorithms". IOP Conference Series: Materials Science and Engineering 928(3).
9. Zafar Iqbal, Junaid Qadir, Adnan Noor Mian, Faisal Kamiran. 2017. "Machine Learning Based Student Grade Prediction: A Case Study". arXiv preprint arXiv:1708.08744.
10. Opeyemi Ojajuni, Foluso Ayeni, Olagunju Akodu, Femi Ekanoye, Samson Adewole, Timothy Ayo, Sanjay Misra, Victor Mbarika. 2021. "Predicting Student Academic Performance Using Machine Learning". Lecture Notes in Computer Science 12957:481-491. Springer.
11. V. Vijayalakshmi, K. Venkatachalapathy. 2019. "Comparison of Predicting Student's Performance Using Machine Learning Algorithms". International Journal of Intelligent Systems and Applications 11(12): 34-45.
12. Esmael Ahmed. 2024. "Student Performance Prediction Using Machine Learning Algorithms". Applied Computational Intelligence and Soft Computing 2024, Article ID 4067721.
13. Balqis Albreiki, Nazar Zaki, Hany Alashwal. 2021. "A Systematic Literature Review of Student Performance Prediction Using Machine Learning Techniques". Education Sciences 11(9): 552.
14. Noor Fadzillah Ab Rahman, Shir Li Wang, Theam Foo Ng, Amr S. Ghoneim. 2025. "Artificial Intelligence in Education: A Systematic Review of Machine Learning for Predicting Student Performance". Int Journal of Advanced Research in Applied Sciences and Engineering Technology 54(1): 198-221.
15. Khalfan Al Mayahi, Mahmood Al-Bahri. 2020. "Students' Academic Success Prediction Based on Machine Learning". Proceedings of the 2020 12th International Congress on Ultra-Modern Telecommunications and Control Systems and Workshops (ICUMT). IEEE.
16. Fidelia A. Orji, Julita Vassileva. 2022. "Machine Learning Approach for Predicting Students' Academic Performance and Study Strategies Based

on Their Motivation". arXiv preprint arXiv:2210.08186.

17. Delali Kwasi Dake, Daniel Danso Essel, Justice Edem Agbodaze. 2021. "Using Machine Learning to Predict Students' Academic Performance During COVID". Proceedings of 2021 International Conference on Computing, Computational Modelling and Applications (ICCMA). IEEE.
18. P. Kalpana, E. Arunmaran, S. Hanif, T. Deebak. 2020. "Student Performance Analysis Using Machine Learning". Publication -International Journal of Innovative Technology and Exploring Engineering (IJITEE) 9(6): 211-215.
19. Onker Vandana, Kumar Krishna Singh, Lamkuche Hemraj Shobharam, Kumar Sunil, Sharma Vijay Shankar, Chowdhary Chiranji Lal, Kumar Vijay. 2025. "Harnessing Machine Learning for Academic Insight: A Study of Educational Performance in Bhopal, India". Publication -Education and Information Technologies.
20. Iddrisu Issah, Obed Appiah, Peter Appiahene, Fuseini Inusah. 2023. "A Systematic Review of the Literature on Machine Learning Application for Determining the Attributes Influencing Academic Performance". Publication - Decision Analytics Journal 7: 100204.