www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 11, November 2024, pp : 1144-1149

e-ISSN : 2583-1062
Impact Factor : 7.001

A REVIEW ON MACHINE LEARNING APPROACHES FOR PREDICTING STUDENT DROPOUT RATES

Adnan Anwar Jamal Qureshi[1], Anas Tufail Ahmed Qureshi[2], Saad Nasir Khan[3], Zaibunnisa L. H. Malik[4]

[1,2,3]Computer Engineering M. H. Saboo Siddik Polytechnic Byculla, India.

[4]Guide, Computer Engineering (HOD) M. H. Saboo Siddik Polytechnic Byculla, India.

adnanqr1306@gmail.com

anasqureshi2424@gmail.com

saadkhan.n1696@gmail.com

zebamalik@yahoo.com

DOI: https://www.doi.org/10.58257/IJPREMS36902

ABSTRACT

This review examines the application of machine learning (ML) models in predicting student dropout and academic performance. Various algorithms, including deep learning and ensemble methods, have demonstrated high accuracy in forecasting student outcomes, offering educational institutions valuable tools for early intervention.

However, challenges such as model interpretability, data privacy concerns, and algorithmic bias persist. The review also highlights the need for more inclusive models that generalize well across diverse educational settings. Looking forward, the integration of ML with technologies like virtual reality (VR) could further enhance student engagement and retention, making education systems more adaptive and personalized.

1. INTRODUCTION

In recent years, the rapid growth of digital technologies has revolutionized various sectors, including education. One of the most significant advancements in this domain is the application of machine learning (ML) and artificial intelligence (AI) to predict student outcomes, particularly in addressing the issue of student dropout.

Educational institutions face a growing challenge with dropout rates, which can result in negative economic and social consequences. Predicting which students are at risk of dropping out can allow educators to intervene early, improving retention and ensuring students complete their programs successfully.

This has prompted the development of various machine learning-based predictive models to forecast student dropout, academic performance, and engagement, making education systems more data-driven and student-centered [10][3][15].

Traditional methods of student assessment and dropout prevention often rely on demographic or attendance data, which can fail to capture the full spectrum of variables influencing student success.

With the advent of machine learning techniques, including supervised learning algorithms, ensemble methods, and deep learning, there is now potential to use more complex data sources like student interactions, behavior patterns in Massive Open Online Courses (MOOCs), and psychological factors to better predict academic performance and dropout risks [27][14][5]. For instance, models based on deep learning, such as Deep Featured Spectral Scaling Classifiers, have shown promise in predicting academic success among students with mental health conditions like bipolar disorder [5].

Research in this area focuses not only on predictive accuracy but also on model interpretability and real-world applicability. As universities increasingly adopt AI-driven tools, models such as the Two-Layer Ensemble Machine Learning approach have been shown to significantly improve dropout prediction, especially in online learning environments [20][9].

Moreover, comparative studies reveal that supervised machine learning models, when integrated with real-time data, outperform traditional statistical methods in predicting student retention and academic outcomes [7][17][22].

Given the high stakes of student retention, the role of machine learning in education is critical not only in identifying students at risk but also in providing personalized interventions that can help keep them engaged. This review aims to explore the various machine learning approaches developed for predicting student dropout and academic success, comparing their effectiveness across different educational contexts and discussing their potential for future integration into higher education systems [2][16].

2. METHODS

a. Terminology:

To ensure clarity, several key terms are defined as they are central to this review:

1) Machine Learning (ML): Refers to computational models and algorithms that enable systems to learn from data and make decisions or predictions. In the context of education, ML models can predict student dropout rates, academic success, or performance based on historical data [16][20].

2) Student Dropout Prediction: This refers to the process of using data-driven techniques to identify students at risk of discontinuing their education. Various approaches, including decision trees, ensemble methods, and deep learning models, have been applied to this problem [2][28].

3) Ensemble Learning: A machine learning approach where multiple models are combined to produce a stronger predictive model. In student dropout prediction, ensemble methods such as stacking or bagging have been shown to outperform single models in terms of accuracy and precision [4][19].

4) Educational Data Mining (EDM): The use of data mining techniques to analyze educational data and identify patterns. This is commonly used to predict student performance, dropout likelihood, and other academic outcomes [14][7].

b. Search Strategy:

To gather relevant literature for this review, a comprehensive search was conducted across several academic databases, including IEEE Xplore, Springer, Elsevier, and MDPI. Key terms used for the search included "student dropout prediction," "machine learning in education," "academic performance prediction," and "deep learning for student retention." The search aimed to capture both recent and foundational studies published between 2010 and 2024 [18][25][3].

c. Selection Criteria:

The inclusion criteria for selecting research papers for this review were as follows:

1) Relevance to the Topic: Only studies focusing on machine learning applications for predicting student outcomes, such as dropout or academic performance, were included [17][9].

2) Recency: To ensure that the review reflects the most current developments, only papers published after 2010 were selected. This timeframe captures the advancements in AI and machine learning relevant to educational data mining [24][6].

3) Diverse Educational Contexts: Studies covering different educational levels (primary, secondary, and higher education) as well as different geographical regions were included to ensure a comprehensive understanding of the global challenges in student retention and academic success [29][10].

4) Type of Study: Both theoretical papers that discuss the methodologies behind machine learning models and empirical studies that apply these models to real-world educational datasets were included [5][21].

d. Data Analysis Approach:

The reviewed papers were analyzed based on the machine learning models they employed, their performance metrics (e.g., accuracy, precision, recall), and the type of educational data used (e.g., student demographics, attendance, behavior in online platforms). Additionally, each paper was compared in terms of its focus on interpretability versus predictive accuracy, a trade-off often encountered in machine learning [30][1].

3. RESULTS

a. Performance of Machine Learning Models:

Machine learning models have shown varied success in predicting student dropouts across different educational contexts. For example, ensemble methods, such as stacking and boosting, were reported to achieve high predictive accuracy. In a study by Niyogisubizo et al. [20], the ensemble approach yielded a precision and recall rate exceeding 90%, while single models, like logistic regression, demonstrated lower performance in predicting dropout risks [7][5]. In MOOCs, Wang and Wang [9] found that decision trees and random forests handled large, complex datasets effectively, improving dropout predictions with considerable accuracy [25].

b. Model Application in Real-World Settings**:**

In studies applying machine learning to real-world educational data, the predictive models performed well, though performance varied by region and demographic. For instance, Hegde and Prageeth [2] showed that models developed in Western contexts performed poorly when transferred to diverse settings, such as in developing countries, necessitating localized adaptations. Studies from Kim et al. [16] and Del Bonifro et al. [6] demonstrated that incorporating socio-economic and behavioral data improved the robustness of dropout predictions across different student populations [13].

c. Impact of Data Complexity:

Complex data sources, such as those from online learning platforms, were found to improve prediction accuracy when used with deep learning models. For example, deep learning models like the DFSSC [5] and the Deep FM-based predictive model [13] were effective in predicting dropout based on students' behavioral patterns and mental health data. Such models proved highly effective in MOOC platforms, where large datasets require sophisticated techniques to capture the nuances of student engagement and dropout likelihood [25][18].

4. DISCUSSION

a. Challenges with Model Interpretability:

Despite high performance, interpretability remains a challenge, particularly with complex models like deep learning and ensemble methods. Educators require insights into why a model predicts that a student will drop out, in order to intervene effectively. Models like neural networks, while powerful, are often difficult to interpret. Studies such as those by Delogu et al. [24] and Coussement et al. [27] emphasized the importance of combining accurate predictions with transparent decision-making processes, particularly in educational environments where understanding the underlying reasons is crucial [22][30].

b. Ethical and Privacy Concerns:

As machine learning models increasingly rely on sensitive student data, ethical concerns have emerged around data privacy. Studies by Mustafa et al. [10] and Kim et al. [16] have underscored the need for robust data protection policies, particularly in regions where data misuse could have far-reaching implications. Furthermore, algorithmic bias remains a significant challenge, particularly when models are trained on biased datasets that fail to represent the diversity of the student population. Researchers like Hegde and Prageeth [2] have called for more inclusive datasets that account for diverse educational backgrounds to mitigate this risk [26].

c. Potential for Improving Student Engagement:

Machine learning models have also demonstrated potential in enhancing student engagement by identifying at-risk students early and enabling timely interventions. Alruwais [13] and Villar et al. [7] found that real-time predictive feedback improved student retention rates, particularly in online classes and MOOCs, where dropout rates tend to be high. By providing personalized interventions, predictive models can keep students engaged, helping them complete their courses successfully [9][15].

d. Future Research and Application:

Moving forward, further integration of predictive models with emerging technologies, such as VR and AR, could provide more immersive and engaging learning environments. Mubarak et al. [25] suggested that these technologies could enhance student motivation by identifying disengagement in real time, allowing for instant interventions. Additionally, future research should focus on developing more inclusive models that can generalize across diverse student populations, expanding beyond higher education into secondary and vocational contexts [12][4][29].

5. CONCLUSION

Machine learning has proven to be a powerful tool in predicting The application of machine learning in predicting student dropout and academic performance has shown significant promise in improving retention and student success rates across various educational settings. From traditional in-person classes to online learning environments like MOOCs, machine learning models—particularly ensemble methods and deep learning algorithms—have demonstrated high predictive accuracy, allowing educational institutions to identify at-risk students early and intervene proactively [25][20][9]. However, the trade-off between model complexity and interpretability remains a major challenge, as educators need transparent and understandable models to make informed decisions. While simpler models like decision trees provide better interpretability, more complex models, such as deep learning, offer higher accuracy at the expense of clarity [13][24].

Ethical concerns surrounding data privacy and algorithmic bias also pose significant challenges to the widespread adoption of these technologies. With student data being a sensitive resource, stricter data protection frameworks and bias mitigation techniques are needed to ensure equitable outcomes across diverse student populations [10][16]. Nevertheless, the potential benefits of machine learning in education—particularly in enhancing student engagement, personalizing learning experiences, and preventing dropout—outweigh the risks if handled responsibly. Researchers like Mubarak et al. [25] have pointed toward future integration with technologies such as virtual reality (VR) and augmented reality (AR), which could further enhance student motivation and retention.

Moving forward, the development of more inclusive, interpretable, and adaptive machine learning models will be crucial to addressing the diverse needs of students globally. By focusing on ethical practices and transparency, the use of machine learning in education has the potential to revolutionize student retention and academic success across all levels of education [5][29][12].

6. REFERENCES

[1] Rahul Katarya, Aryan Garg, Jalaj Gaba, Varsha Verma (2021). A review on machine learning-based student's academic performance prediction systems. IEEE (ICAIS). DOI: 10.1109/ICAIS50930.2021.9395767

[2] Vinayak Hegde, Prageeth P P (2018). Higher Education Student Dropout Prediction and Analysis through Educational Data Mining. IEEE (ICISC). DOI: 10.1109/ICISC.2018.8398887

[3] Marcell Nagy, Roland Molontay (2018). Predicting Dropout in Higher Education based on Secondary School Performance. IEEE (INES). DOI: 10.1109/INES.2018.8523888

[4] P. Karpagalakshmi, S.R. Dhanashree, M. Abdul Wahidh, Dr. A. Rajesh (2024). Predicting College Dropout Rates Using Machine Learning: A Student Success Initiative. IEEE (ICCDS). DOI: 10.1109/ICCDS60734.2024.10560384

[5] S. Peerbasha, M. Mohamed Surputheen (2021). Prediction of Academic Performance in College Students with Bipolar Disorder Using Deep Featured Spectral Scaling Classifier (DFSSC). IJERT. DOI: 10.17577/IJERTV10IS100122

[6] Francesca Del Bonifro, Maurizio Gabbrielli, Giuseppe Lisanti, Stefano Pio Zingaro (2020). Student Dropout Prediction. Springer. DOI Link

[7] Alice Villar, Carolina Robledo Velini de Andrade (2024). Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study. Springer. DOI: https://doi.org/10.1007/s44163-023-00079-z

[8] Dr. D. Jayanthi, Mr. K. Naveen Akash, Mr. S. Sai Nithish, Ms. V. Nanthanavalli (2024). An Analysis of Dropout Students in the Education System of Gujarat. IEEE (ICCDS). DOI: 10.1109/ICCDS60734.2024.10560409

[9] Lutong Wang, Hong Wang (2019). Learning Behavior Analysis and Dropout Rate Prediction Based on MOOCs Data. IEEE (ITME). DOI: 10.1109/ITME.2019.00100

[10] Mohammad Nurul Mustafa, Linkon Chowdhury, Md. Sarwar Kamal (2012). Students Dropout Prediction for Intelligent System from Tertiary Level in Developing Country. IEEE (ICIEV). DOI: 10.1109/ICIEV.2012.6317441

[11] Kittinan Limsathitwong, Kanda Tiwatthanont, Tanasin Yatsungnoen (2018). Dropout Prediction System to Reduce Discontinue Study Rate of Information Technology Students. IEEE (ICBIR). DOI: 10.1109/ICBIR.2018.8391192

[12] Warit Tenpipat, Khajonpong Akkarajitsakul (2020). Student Dropout Prediction: A KMUTT Case Study. IEEE (IBDAP). DOI: 10.1109/IBDAP50342.2020.9245457

[13] Nuha Mohammed Alruwais (2023). Deep FM-Based Predictive Model for Student Dropout in Online Classes. IEEE (ACCESS). DOI: 10.1109/ACCESS.2023.3312150

[14] Anjana Pradeep, Smija Das, Jubilant J Kizhekkethottam (2015). Students Dropout Factor Prediction Using EDM Techniques. IEEE (ICSNS). DOI: 10.1109/ICSNS.2015.7292372

[15] Valentim Realinho, Jorge Machado, Luís Baptista, Mónica V. Martins (2022). Predicting Student Dropout and Academic Success. MDPI. DOI: https://doi.org/10.3390/data7110146

[16] Sangyun Kim, Euteum Choi, Yong-Kee Jun, Seongjin Lee (2023). Student Dropout Prediction for University with High Precision and Recall. MDPI. DOI: 10.3390/app13106275

[17] Dalia Abdulkareem Shafiq, Mohsen Marjani, Riyaz Ahamed Ariyaluran Habeeb, David Asirvatham (2022). Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review. IEEE (ACCESS). DOI: 10.1109/ACCESS.2022.3188767

[18] Havan Agrawal, Harshil Mavani (2015). Student Performance Prediction using Machine Learning. IJERT. DOI: 10.17577/IJERTV4IS030127

[19] Prashant Dixit, Dr. Harish Nagar, Prof. Sarvottam Dixit (2021). Decision Support System Model for Student Performance Detection using Machine Learning. IJERT. DOI: IJERTV10IS050017

[20] Jovial Niyogisubizo, Lyuchao Liao, Eric Nziyumva, Evariste Murwanashyaka, Pierre Claver Nshimyumukiza (2022). Predicting Student's Dropout in University Classes Using Two-Layer Ensemble Machine Learning Approach: A Novel Stacked Generalization. Elsevier. DOI: https://doi.org/10.1016/j.caeai.2022.100066

[21] Matti Vaarma, Hongxiu Li (2024). Predicting student dropouts with machine learning: An empirical study in Finnish higher education. Elsevier. DOI: 10.1016/j.techsoc.2024.102474

[22] Daniel Andrade-Girón, Juana Sandivar-Rosas, William Marín-Rodriguez, Edgar Susanibar-Ramirez, Eliseo Toro-Dextre, Jose Ausejo-Sanchez, Henry Villarreal-Torres, Julio Angeles-Morales (2023). Predicting Student Dropout Based on Machine Learning and Deep Learning: A Systematic Review. EAI. DOI: https://doi.org/10.4108/eetsis.3586

[23] Valentim Realinho, Jorge Machado, Luís Baptista, Mónica V. Martins (2022). Predicting Student Dropout and Academic Success. ERIC. DOI: https://doi.org/10.5281/zenodo.5777339

[24] Marco Delogu, Raffaele Lagravinese, Dimitri Paolini, Giuliano Resce (2024). Predicting Dropout from Higher Education: Evidence from Italy. Elsevier. DOI: https://doi.org/10.1016/j.econmod.2023.106583

[25] Ahmed A. Mubarak, Han Cao, Ibrahim M. Hezam (2021). Deep Analytic Model for Student Dropout Prediction in Massive Open Online Courses. Elsevier. DOI: https://doi.org/10.1016/j.compeleceng.2021.107271

[26] Steffen Wild, Lydia Schulze Heuling (2020). Student dropout and retention: An event history analysis among students in cooperative higher education. Elsevier. DOI: 10.1016/j.ijer.2020.101687

[27] Kristof Coussement, Minh Phan, Arno De Caigny, Dries F. Benoit, Annelies Raes (2020). Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model. Elsevier. DOI: 10.1016/j.dss.2020.113325

[28] Stefania Guzmán-Castillo, Franziska Körner, Julia I. Pantoja-García, Lainet Nieto-Ramos, Yulineth Gómez-Charris, Alex Castro-Sarmiento, Alfonso R. Romero-Conrado (2022). Implementation of a Predictive Information System for University Dropout Prevention. Elsevier. DOI: https://doi.org/10.1016/j.procs.2021.12.287

[29] Ari Melo Mariano, Arthur Bandeira de Magalhães Lelis Ferreira, Maíra Rocha Santos, Mara Lucia Castilho, Anna Carla Freire Luna Campêlo Bastos (2022). Decision Trees for Predicting Dropout in Engineering Course Students in Brazil. Elsevier. DOI: 10.1016/j.procs.2022.11.285

[30] Neema Mduma, Khamisi Kalegele, Dina Machuve (2019). A Survey of Machine Learning Approaches and Techniques for Student Dropout Prediction. Data Science Journal. DOI: 10.5334/dsj-2019-014