IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Literature Review On Early Intervention Strategies To Address Student Attrition With Predictive Analytics

Rupali Ambalal Jadhav
PhD Scholar, Atmiya University, Rajkot
Assistant Professor, Thakur Institute of Management Studies, Career Development & Research, Mumbai

Dr. Rupal Parekh
Assistant Professor
Atmiya University, Rajkot, Gujarat

Abstract: Student attrition, the premature withdrawal of students from educational programs, poses a significant challenge to higher education institutions globally. This research This study examines the application of predictive analytics, leveraging machine learning and data mining techniques, to identify at-risk students and implement timely interventions aimed at improving retention and academic success. We systematically review existing literature on predictive learning analytics (PLA) in higher learning, examining various methodological approaches, predictive models, and the factors influencing student attrition. The analysis highlights the advantages and disadvantages of various prediction models, emphasizing the need for responsible AI frameworks to address ethical concerns and ensure equitable outcomes. Furthermore, we discuss the integration of PLA with motivational interventions and the role of data warehousing in facilitating data-driven decision-making for enhanced student support. Finally, we identify key research gaps and propose future directions for refining PLA strategies to create more effective and inclusive learning environments.

***Index Terms* - Predictive analytics, student attrition, early intervention, student retention strategies, higher education, machine learning, data mining, responsible AI, educational outcomes, data-driven decision-making**

I. Introduction

High rates of student attrition represent a substantial financial and societal cost [1], impacting institutional reputation, funding, and the overall success of educational initiatives [2]. The problem is multifaceted, impacted by a variety of intricately interacting elements, such as academic achievement, socio-economic background, student engagement, and access to support services [3]. Traditional approaches to addressing attrition have often been reactive, focusing on interventions after students demonstrate signs of academic difficulty [1]. However, the increasing availability of student data, coupled with advancements in predictive analytics, presents a unique opportunity to adopt a proactive strategy.

Predictive learning analytics (PLA) analyzes student data and forecasts the probability of attrition using data mining and machine learning algorithms [4]. This makes it possible for educational institutions to spot pupils who are at risk early on their academic journey and provide timely treatments that are catered to their individual requirements. [3]. This proactive approach can significantly improve student retention rates, enhance academic performance, and ultimately optimize educational outcomes [4]. This research paper provides a comprehensive overview of the PLA approach to student attrition, focusing on its methodological underpinnings, practical applications, and ethical considerations.

II.Literature Review

Numerous studies demonstrate the potential of PLA to positively impact student retention and other aspects contributing to student success [4], [3]. Researchers have employed various machine learning algorithms, including decision trees [5], random forests [5], [6], support vector machines [5], logistic regression [5], and neural networks [6], [7], [8], to predict student outcomes. These models often incorporate a wide range of predictor variables, encompassing demographic information [9], academic performance [5], [2], course engagement [2], and online interaction data [6]. The choice of algorithm and predictor variables often depends on the specific context, data availability, and the desired level of prediction accuracy. Siti Dianah Abdul Bujang, A. Selamat, and O. Krejcar [5] created a predictive model utilizing logistic regression, support vector machines, decision trees, and random forests to predict student final grades. Their findings indicated that the J48 decision tree model accomplished the highest prediction accuracy (99.6%), demonstrating the potential for early detection of at-risk students [5]. In contrast, Mahdi-Reza Borna et al. [6] utilized random forests, XGBoost, and recurrent neural networks to predict student withdrawals and distinctions using clickstream information from the Open University Learning Analytics Dataset (OULAD). While their models effectively identified at-risk students, achieving an accuracy of approximately 77-78%, predicting high achievers proved more challenging [6]. This highlights the complexities inherent in predicting academic success and the need for more nuanced modelling approaches.

Furthermore, studies have explored the use of PLA to identify at-risk students in specific disciplines, such as engineering mathematics [2] and accounting [9]. These studies demonstrate the adaptability of PLA to different educational contexts and the potential for tailoring interventions to specific student populations. However, PLA's effectiveness depends on the completeness and quality of the data used to train the models [10]. Due of ethical and data privacy considerations, it is essential to develop responsible AI frameworks to guide the implementation and use of PLA in higher education [10].

Table 2: Research Gap Analysis

Study	Key Findings	Gap Analysis
McLean (NaN)[1]	Predictive analytics is being used to continuously improve higher education.	Need for comprehensive models integrating organizational goals.
Dart (NaN)[2]	Developing predictive models for engineering mathematics student achievement.	Implementation challenges due to lack of technical expertise.
Bacus & Cascaro (2024)[3]	Comprehensive analysis of the research on the effects of analytics for predictive learning in higher education.	Lack of longitudinal studies for long-term impact assessment.
Bujang et al. (2021)[4]	Supervised machine learning for student grade prediction.	Need for cross-institutional studies to generalize findings.
Borna et al. (2024)[5]	AI-assisted click data analysis was used to forecast student performance and learning evaluation.	Lack of comprehensive models integrating various data sources.
Berens et al. (2019)[6]	Used administrative student data and machine learning to predict student dropouts.	Need for cross-institutional studies to generalize findings.
Berens et al. (NaN)[7]	Used administrative student data and machine learning to predict student dropouts.	Practical implementation challenges.

Nakale & Amugongo (2023)[8]	Case study on predicting student attrition at the University of Namibia.	Limited research on personalization of interventions.
Tirado et al. (2024)[9]	suggested an AI architecture for learning analytics in higher education that is operationally responsible	Challenges in practical implementation due to data privacy concerns and resource constraints.
Ismaili & Besimi (2024)[10]	Developed a data warehousing framework for predictive analytics to identify at-risk students.	Limited research on personalization of interventions.
Sghir et al. (2022)[11]	Systematic review of advances in predictive learning analytics over the last decade.	Need for more longitudinal studies.
Tarmizi et al. (2019)[12]	analysis of student attrition through the use of data mining and big data analytics methods.	Practical implementation challenges and resource constraints.
Williams et al. (NaN)[13]	Predicting student attrition considering demographic, study-based, and psychological factors.	Implementation challenges due to lack of technical expertise.
Hernndez-de-Menndez et al. (2022)[14]	State of the art in learning analytics.	Need for more comprehensive models.
Ortiz et al. (2022)[15]	Academic achievement, admissions information, and membership in an ethnic or racial minority are all related to student retention.	Limited research on personalization of interventions.
Herodotou et al. (2020)[16]	Motivational interventions using PLAs can improve student retention.	Need for longitudinal studies to assess long-term impact.
Seidel & Kutieleh (2017)[17]	Used predictive analytics to target and improve first-year student attrition.	Practical implementation challenges and data privacy concerns.
Shafiq et al. (NaN)[18]	Systematic literature review on student retention using educational data mining and predictive analytics.	Lack of cross-institutional studies to generalize findings.

III. Methodological Approaches and Predictive Models

The development and application of PLA involves several key stages: Data pre-processing, data collection, model selection, feature engineering, model evaluation, and model training [4]. Data collection encompasses gathering relevant student information from multiple sources, such as learning management systems (LMS), student information systems (SIS), and other relevant databases [11]. Data pre-processing entails addressing problems such missing numbers and discrepancies while cleaning, converting, and getting the data ready for analysis [4]. Feature engineering focuses on selecting and transforming relevant variables to improve model accuracy and interpretability [12].

Model selection involves choosing a suitable machine learning algorithm in accordance with the data's properties and the research question [12]. Various algorithms, as mentioned previously, have been employed, each with its strengths and weaknesses. Decision trees, for example, offer high interpretability, while neural networks can capture complex relationships in the data but may be less transparent [12]. Model training involves using the pre-processed data to train the chosen algorithm, optimizing its parameters to maximize prediction accuracy [12]. Taking appropriate action, such as accuracy, precision, recall, F1-score, and AUC, model evaluation entails assessing the performance of the trained model [6], [9].

The choice of evaluation metrics is crucial and should align with the specific goals of the PLA application [12]. For example, if the primary goal is to identify at-risk students, recall (the capacity to accurately recognize all at-risk students) may be prioritized over precision (the capacity to accurately recognize only at-risk students) [6]. Furthermore, the robustness of the model should be assessed by testing its performance on different datasets and under various conditions [12]. Cross-validation techniques are commonly used to ensure that the model generalizes well to unseen data [2].

IV. Factors Influencing Student Attrition

Student attrition is a complex phenomenon influenced by a multitude of factors [13], [14]. These factors can be broadly categorized into academic, socio-economic, psychological, and institutional factors [14]. Academic factors include academic performance, course difficulty, and the student's perceived ability to succeed in their chosen field of study [2], [9]. Socio-economic factors encompass the student's financial situation, family background, and access to resources [14]. Psychological factors include student motivation, stress levels, and mental health [14]. Institutional factors involve the quality of teaching, the availability of support services, and the overall learning environment [14].

Several Research has looked into the relative importance of these factors in predicting student attrition [14], [15]. For instance, C. Williams et al. [14] found that Academic achievement during the first year of study was a strong indicator of attrition alongside psychological factors such as burnout and perceived stress. Other studies have highlighted the importance of socio-economic factors [14], particularly for students from disadvantaged backgrounds who may face greater challenges in accessing resources and support [16]. The relative importance of these factors may vary based on the particular institutional context and student population [13].

V. Interventions and Support Strategies

Identifying at-risk students through PLA is only the first step in mitigating attrition [17]. To help these students and improve their prospects of academic success, effective interventions are essential [17]. Academic advising, tutoring, mentoring, and financial help are just a few of the ways these interventions might be implemented [17]. Moreover, motivational interventions, such as personalized communication and encouragement from university staff, can significantly improve student retention outcomes [17].

C. Herodotou et al. [17] demonstrated the effectiveness of motivational interventions in improving student retention in a randomized controlled trial. Their study showed that students receiving personalized support from Student Support Teams (SSTs) were significantly more likely to complete their studies [17]. This highlights the importance of combining PLA with targeted support strategies to maximize the impact on student retention [17]. The design and implementation of interventions should be informed by a thorough understanding of the specific factors contributing to attrition in a given context [18]. Furthermore, the effectiveness of interventions should be continuously evaluated and refined based on their impact on student outcomes [18].

VI. Research Gaps and Future Directions

While PLA holds immense potential for enhancing student success, several research gaps remain [12], [19]. Further research is needed to refine predictive models, exploring more sophisticated algorithms and incorporating a wider range of predictor variables [12]. The development of more robust and reliable models is crucial to ensure that interventions are targeted effectively and efficiently [12]. Moreover, research is needed to investigate the long-term effect of PLA interventions on student outcomes, assessing not only retention rates but also academic achievement, career prospects, and overall well-being [19].

Further research should also focus on developing more personalized and adaptive interventions, tailoring support strategies to individual student needs and learning styles [19]. The integration of PLA with other educational technologies and pedagogical approaches could also enhance its effectiveness [19]. Finally, more research is needed to address the ethical implications of PLA, ensuring fairness, transparency, and accountability in its application [19]. This research should involve collaboration between data scientists, educators, and ethicists to develop responsible AI frameworks and guidelines for the use of PLA in higher education [19].

VII. Conclusion

Predictive learning analytics offers a powerful tool for enhancing student success and mitigating attrition in higher education. By leveraging data mining and machine learning techniques, institutions can proactively identify at-risk students and Put timely interventions into place to enhance academic performance and retention. However, the effective implementation of PLA requires careful consideration of methodological approaches, the selection of appropriate predictive models, and the ethical implications of using AI in educational decision-making. Future research should focus on refining predictive models, developing more personalized interventions, and addressing ethical concerns to ensure that PLA is used responsibly and equitably to create more effective and inclusive learning environments. The integration of PLA with robust data warehousing and responsible AI frameworks will be crucial for maximizing its impact on student success and achieving the full potential of data-driven educational strategies. This proactive approach can transform higher education, leading to improved student outcomes and a more equitable learning experience for all.

VIII. Acknowledgment

I am sincerely thanks Dr. Rupal Parekh for her invaluable guidance and support throughout this research. Also extend gratitude to Thakur Institute of Management Studies, Career Development & Research, Mumbai for the opportunity to pursue a Ph.D.

REFERENCES

[1] McLean, M.. NaN. "Continuous Improvement in Higher Education: A Change Model Using Predictive Analytics to Achieve Organizational Goals". None. https://doi.org/None

[2] Dart, S.. NaN. "Developing predictive models of student success in undergraduate engineering mathematics courses". None. https://doi.org/None

[3] Bacus, John A. and Cascaro, Rhodessa J.. 2024. "Impact of Predictive Learning Analytics in Higher Education: A Systematic Literature Review". International Conference on Educational and Information Technology. https://doi.org/10.1109/ICEIT61397.2024.10540840

[4] Bujang, Siti Dianah Abdul, Selamat, A., and Krejcar, O.. 2021. "A Predictive Analytics Model for Students Grade Prediction by Supervised Machine Learning". None. https://doi.org/10.1088/1757-899X/1051/1/012005

[5] Borna, Mahdi-Reza, Saadat, Hanan, Hojjati, Aref Tavassoli, and Akbari, Elham. 2024. "Analyzing click data with AI: implications for student performance prediction and learning assessment". Frontiers in Education. https://doi.org/10.3389/feduc.2024.1421479

[6] Berens, Johannes, Schneider, Kerstin, Grtz, Simon, Oster, Simon, and Burghoff, Julian. 2019. "Early Detection of Students at Risk - Predicting Student Dropouts Using Administrative Student Data from German Universities and Machine Learning Methods". None. https://doi.org/10.5281/ZENODO.3594771

[7] Berens, Johannes, Schneider, Kerstin, Grtz, Simon, Oster, Simon, and Burghoff, Julian. NaN. "Early Detection of Students at Risk Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods". Social Science Research Network. https://doi.org/10.2139/ssrn.3275433

[8] Nakale, Samuel Nghidengwa and Amugongo, L. M.. 2023. "Predicting Student Attrition: A Case Study of the University of Namibia Bachelor of Accounting (Chartered Accountancy) Programme". None. https://doi.org/10.1109/ETNCC59188.2023.10284930

[9] Tirado, Alba Morales, Mulholland, Paul, and Fernndez, Miriam. 2024. "Towards an Operational Responsible AI Framework for Learning Analytics in Higher Education". arXiv.org. https://doi.org/10.48550/arXiv.2410.05827

[10] Ismaili, Burim and Besimi, Adrian. 2024. "A Data Warehousing Framework for Predictive Analytics in Higher Education: A Focus on Student at-Risk Identification". SEEU Review. https://doi.org/10.2478/seeur-2024-0020

[11] N, Sghir, A, Adadi, and M, Lahmer. 2022. "Recent advances in Predictive Learning Analytics: A decade systematic review (2012-2022).".https://doi.org/10.1007/s10639-022-11536-0

[12] Tarmizi, Syaidatus Syahira Ahmad, Mutalib, Sofianita, Hamid, Nurzeatul Hamimah Abdul, and Rahman, S. A.. 2019. "A Review on Student Attrition in Higher Education Using Big Data Analytics and

Data Mining Techniques". International Journal of Modern Education and Computer Science. https://doi.org/10.5815/ijmecs.2019.08.01

[13] Williams, C., Freitas, S., and Dziurawiec, S.. NaN. "Predicting student attrition with "big data": Considering demographic, study-based

[14] and psychological factors using two large datasets". None. https://doi.org/None

[15] M, Hernndez-de-Menndez, R, Morales-Menendez, C, Escobar, and R, Ramrez Mendoza. 2022. "Learning analytics: state of the art"

[16] A, Ortiz, JV, Brown, M, Bains, and SL, Goffar. 2022. "Association Between Ethnic-Racial Minority Status, Admissions Data, and Academic Performance in Student Retention from a Physical Therapy Program in a Minority-Serving Institution.".

[17] Herodotou, C., Naydenova, G., Boroowa, Avinash, Gilmour, Alison, and Rienties, B.. 2020. "How Can Predictive Learning Analytics and Motivational Interventions Increase Student Retention and Enhance Administrative Support in Distance Education?". Journal of Learning Analytics. https://doi.org/10.18608/JLA.2020.72.4

[18] Seidel, Ewa and Kutieleh, Salah. 2017. "Using predictive analytics to target and improve first year student attrition". None.

[19] https://doi.org/10.1177/0004944117712310

[20] Shafiq, Dalia Abdulkareem, Marjani, Mohsen, Habeeb, Riyaz Ahamed Ariyaluran, and Asirvatham, D.. NaN. "Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review". IEEE Access. https://doi.org/10.1109/ACCESS.2022.3188767