International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 9, September 2025)

Artificial Intelligence Technique for Predicting Academic Success of College Students: A Review

Vidhya[1], Praveen Kumar Kaithal[2]

[1] *Research Scholar, Department of CSE, SIRT College Bhopal*

[2] *Assistant Professor, Department of CSE, SIRT College Bhopal*

***Abstract*— This review explores the application of Artificial Intelligence (AI) techniques for predicting the academic success of college students, highlighting their growing role in higher education research and practice. AI-based models, including machine learning and deep learning approaches, have demonstrated significant potential in analyzing diverse academic, behavioral, demographic, and socio-economic factors that influence student performance. By leveraging large and complex datasets, these techniques enable early identification of at-risk students, support personalized learning strategies, and assist institutions in improving retention and overall outcomes. The paper synthesizes recent developments, discusses strengths and limitations of AI-driven prediction models, and outlines future directions for creating ethical, accurate, and student-centered solutions.**

Keywords*— *AI, Student, Academic, Success.

I. INTRODUCTION

Academic success has always been a key focus in higher education, as it directly reflects not only the capabilities of students but also the effectiveness of educational institutions in fostering growth and learning. Predicting academic success among college students has gained increasing attention in recent years due to the rising demand for evidence-based strategies to improve student outcomes, reduce dropout rates, and enhance overall institutional performance[1]. With the expansion of higher education worldwide, the diversity of student populations has increased significantly, bringing in learners from varied academic, cultural, socio-economic, and personal backgrounds. This diversity makes predicting academic success a complex challenge, as it requires understanding the influence of multiple interrelated factors beyond just classroom performance[2]. Traditional measures of academic success, such as grade point average (GPA) or graduation rates, provide only partial insights, whereas modern predictive models aim to capture a wider spectrum of determinants that shape student achievement.

The importance of predicting academic success lies in its ability to enable timely interventions and provide personalized academic support. By identifying at-risk students early in their educational journey, institutions can design targeted mentoring, counseling, and remedial programs to improve learning outcomes[3]. Furthermore, predictive insights assist faculty and administrators in curriculum design, resource allocation, and policy-making to foster a more student-centered learning environment. For students themselves, awareness of predictive factors can encourage self-reflection and better planning, leading to improved performance and academic persistence. In today's competitive and dynamic educational landscape, where employability and lifelong learning are closely linked to academic success, such predictive efforts play a pivotal role in shaping students' futures[4].

Research on academic success prediction has evolved significantly over time, shifting from traditional statistical approaches to advanced computational models. Early studies primarily relied on regression analysis or linear models that focused on high school grades, standardized test scores, and demographic attributes as predictors of college performance. While useful, these models often overlooked non-academic factors such as motivation, study habits, peer influence, and mental health, which are equally critical in determining outcomes[5]. With the advent of artificial intelligence (AI), machine learning (ML), and deep learning (DL), researchers now have powerful tools to process large volumes of educational data and uncover hidden patterns that were previously inaccessible. These modern approaches enable the integration of diverse data sources, including learning management systems (LMS), attendance records, social media interactions, and behavioral analytics, thus providing more accurate and holistic predictions[6].

In addition to technological advancements, there has been a growing recognition of the need to balance quantitative metrics with qualitative insights in predicting academic success. Factors such as student engagement, resilience, socio-economic background, and institutional support systems play an essential role in academic outcomes, making prediction a multidisciplinary task[7]. Social sciences, psychology, and educational research complement computational methods by providing theoretical frameworks to interpret data-driven findings. The interplay between these domains enriches the predictive process, ensuring that models do not merely generate statistical accuracy but also align with the broader goals of student well-being and holistic development[8].

Another important dimension in predicting academic success is the ethical and responsible use of predictive analytics in education. While predictive models offer powerful opportunities to improve student outcomes, they also raise concerns about data privacy, fairness, and bias[9]. Over-reliance on historical data may reinforce existing inequalities, while misinterpretation of predictions could unfairly label students as underperformers. Therefore, it is essential to ensure that predictive frameworks are designed with transparency, inclusivity, and accountability, serving as supportive tools rather than restrictive mechanisms[10].

Predicting academic success of college students is not just an academic exercise but a practical necessity in modern higher education. It combines statistical analysis, computational intelligence, and human-centered perspectives to anticipate outcomes, identify challenges, and support students in realizing their full potential[11]. By integrating technological innovations with pedagogical insights, the pursuit of accurate and meaningful academic success prediction has the potential to transform educational practices and contribute significantly to student achievement, retention, and lifelong learning[12].

II. LITERATURE SURVEY

J. Guanin-Fajardo et al., [1] explored the use of machine learning techniques to predict the academic success of college students by analyzing multiple data sources, including academic records, demographic attributes, and behavioral indicators. The authors highlighted how different algorithms, such as decision trees, random forests, and support vector machines, can provide accurate predictions when trained on diverse student datasets. Their study demonstrated that incorporating non-academic variables like attendance, study patterns, and engagement levels significantly enhances predictive accuracy compared to models relying only on grades. One of the notable contributions of this work was the comparative evaluation of algorithms, offering insights into their strengths and limitations for academic prediction tasks.

Y. Zhang et al., [2] proposed a deep learning-based approach that integrates behavioral and socio-economic features to predict student academic performance. Unlike traditional statistical and machine learning models, their method leverages neural networks capable of capturing non-linear relationships in complex educational datasets. By combining socio-economic background with behavioral attributes such as class participation and time spent on learning management systems, their model achieved higher accuracy in forecasting outcomes like GPA and course completion. The research highlighted the importance of considering both personal and institutional factors, as academic success cannot be determined solely by prior grades.

M. R. Khan et al., [3] introduced an artificial intelligence-based early warning system designed to enhance student success in higher education. Their approach aimed at identifying students who are at risk of academic underperformance or dropout by analyzing real-time data collected during the academic term. The system integrated features such as class attendance, assignment submission records, and digital engagement levels, allowing institutions to monitor student progress continuously. A key contribution of the study was the development of a proactive intervention mechanism, enabling faculty and advisors to provide timely academic and psychological support. The results showed that early identification of at-risk students leads to significant improvements in retention rates and overall academic outcomes.

N. Gupta et al., [4] carried out a comparative study of machine learning techniques for predicting student academic performance, with a focus on identifying the most efficient models. The authors evaluated a range of algorithms, including logistic regression, support vector machines, random forests, and gradient boosting, using diverse datasets collected from higher education institutions. Their findings revealed that ensemble learning approaches consistently outperformed single-model techniques due to their ability to minimize overfitting and improve generalization. The study also highlighted the critical role of feature selection in enhancing model accuracy, where combining academic, behavioral, and demographic features yielded the most reliable predictions. Additionally, the authors discussed practical applications of predictive models in student advising and institutional policy-making.

J. F. Rodríguez-Hernández et al., [5] proposed an interpretable machine learning framework for predicting student dropout and academic success, addressing a key limitation of many existing models: lack of transparency. Their study emphasized that while high predictive accuracy is important, educators and administrators also require models that are explainable and trustworthy. Using advanced interpretable algorithms, the authors were able to identify the most influential factors contributing to dropout and academic achievement, such as attendance, financial conditions, and prior academic performance. The results indicated that transparent models not only provide actionable insights but also encourage adoption by educational stakeholders. Another strength of this research lies in its dual focus on both predicting success and preventing dropout, thereby offering a comprehensive approach to student lifecycle management.

H. M. Alkhatlan et al., [6] developed an intelligent prediction and classification model for student academic performance using machine learning in higher education settings. Their research focused on creating a system that could categorize students into performance levels, allowing educators to provide differentiated instruction and support. The study used a combination of supervised learning techniques and highlighted the effectiveness of classification algorithms in identifying students at different academic risk levels. One of the notable aspects of the study was its emphasis on the integration of multiple data types, including academic records, behavioral factors, and institutional parameters. The proposed model proved effective in not only predicting outcomes but also offering valuable insights for academic planning and resource allocation. The findings underline the role of predictive analytics in advancing personalized education and enabling universities to adopt more student-centered approaches.

S. B. Kotsiantis et al., [7] examined the application of various machine learning techniques in educational data mining to predict student performance, offering one of the clearer overviews of algorithmic choices and practical considerations. The study compared supervised learning methods such as decision trees, Naive Bayes, k-nearest neighbors, and support vector machines across several educational datasets, highlighting differences in accuracy, interpretability, and computational cost. Kotsiantis emphasized the importance of data preprocessing—handling missing values, discretization, and normalization—as a major determinant of model success in educational contexts. The paper also discussed feature selection strategies, showing that removing irrelevant or redundant attributes often improves both performance and model generalizability. Importantly, the author pointed out limitations in many EDMin studies, such as small sample sizes and lack of cross-institution validation, which can inflate reported performance.

C. Márquez-Vera et al., [8] addressed the difficult problem of predicting student failure under realistic conditions of high-dimensional and imbalanced educational data, using genetic programming and multiple data-mining methods. Their research demonstrated that standard classifiers often struggle with skewed class distributions common in dropout/failure prediction, and they proposed tailored solutions including resampling and cost-sensitive approaches. By comparing genetic programming with ensemble and single-model methods, the authors showed that evolutionary algorithms can discover interpretable rules and non-linear relationships useful for early warning systems. They also analyzed how dimensionality reduction and feature engineering influence classifier behavior, emphasizing the need to preserve informative attributes while reducing noise.

L. Costa et al., [9] provided a comprehensive survey coupled with a case study that investigated student performance prediction using educational data mining tools, focusing on methodological pipelines from data collection to deployment. The paper synthesized findings across multiple studies, noting recurring predictor categories such as prior academic records, attendance, socio-demographic variables, and LMS-derived behavioral metrics. In their case study, Costa and colleagues implemented and evaluated several algorithms, documenting the preprocessing steps, feature selection methods, and performance metrics to showcase an end-to-end workflow. They highlighted practical challenges like data heterogeneity, privacy concerns, and integrating predictive outputs into institutional decision-making processes. The authors recommended best practices including continuous monitoring of model drift, stakeholder involvement for actionable thresholds, and transparent reporting of limitations.

J. Márquez-Vera et al., [10] explored the prediction of school failure using data mining techniques, focusing on both model performance and the extraction of meaningful patterns that could inform educational interventions. Their analysis incorporated a variety of classifiers alongside descriptive analytics to identify risk factors strongly associated with poor outcomes, such as absenteeism, low prior achievement, and socio-economic indicators. The authors paid particular attention to validation protocols, using cross-validation and hold-out sets to assess model robustness and guard against overfitting. They also discussed the interpretability of results, advocating the use of rule-based and tree-based learners when stakeholder comprehension is critical for adoption.

A. L. C. Baradwaj et al., [11] investigated mining educational data to analyze and predict student performance, presenting a structured approach to feature extraction, model selection, and educational interpretation. Their study emphasized academic and non-academic predictors, including demographics, previous exam scores, and behavioral data, and demonstrated how feature engineering significantly affects predictive accuracy. Baradwaj and Pal compared classification algorithms like decision trees, Naive Bayes, and association rule mining, showing that simple, interpretable models often provide comparable utility to complex black-box methods for many institutional use-cases.

C. Romero et al., [12] paper provided a critical examination of the methodological gaps then present, such as inconsistent evaluation metrics, lack of longitudinal studies, and scarce attention to model interpretability. Importantly, Romero et al. proposed a research agenda emphasizing interdisciplinary collaboration among educators, data scientists, and ethicists to ensure that EDM contributes positively to learning outcomes. As a widely cited survey, this work laid conceptual and practical foundations that many later studies—focused on AI-driven student success prediction—built upon.

III. CHALLENGES

Predicting academic success is a promising research area, yet it faces multiple challenges that limit the accuracy, reliability, and ethical deployment of predictive models. One of the foremost challenges lies in data quality and availability. Educational datasets are often incomplete, noisy, or inconsistent, with missing attendance records, inaccurate grades, or unrecorded behavioral attributes. Many institutions also struggle with small sample sizes, making it difficult to train robust machine learning and deep learning models that generalize well across diverse student populations. Furthermore, student performance is influenced by complex and dynamic factors such as motivation, mental health, and personal circumstances, which are often not captured in institutional data.

Another critical challenge is feature selection and data imbalance. Academic success prediction often involves high-dimensional datasets with hundreds of potential attributes, many of which may not significantly contribute to prediction. Identifying the most relevant features while avoiding redundancy requires careful preprocessing and domain expertise. In addition, datasets are frequently imbalanced, as the number of successful students far outweighs those who drop out or fail, which skews model performance. Handling such imbalance through resampling or cost-sensitive learning is essential but adds further complexity.

Model interpretability also remains a pressing issue. While advanced AI techniques such as deep learning can achieve high accuracy, they often operate as "black-box" models that provide little explanation for their predictions. For educators and policymakers, transparency is crucial since decisions based on opaque models may lack trust and face resistance in adoption. Interpretable models, such as decision trees or rule-based classifiers, are sometimes preferred despite lower accuracy because they offer actionable insights that align with institutional needs.

The ethical and privacy concerns surrounding predictive analytics in education cannot be ignored. Collecting and analyzing student data raises questions about consent, security, and fairness. Predictive models trained on historical data may inadvertently reinforce existing biases related to socio-economic status, gender, or ethnicity, leading to unfair treatment of vulnerable groups. Additionally, there is the danger of over-reliance on predictions, where students labeled as "at-risk" may face stigmatization or reduced opportunities, contrary to the goal of supporting their success.

Another challenge is institutional and contextual variability. Models developed for one university or cultural setting may not be directly transferable to another due to differences in grading systems, curricula, teaching methods, and socio-economic conditions of students. This limits the generalizability of many proposed models and necessitates customization for each institutional context, which requires additional resources and expertise.

Finally, there are implementation challenges in integrating predictive models into actual academic practices. Even when accurate predictions are available, institutions often lack the infrastructure or trained staff to translate these insights into meaningful interventions. Predictive systems must be coupled with actionable support mechanisms, such as mentoring, counseling, or academic workshops, to truly benefit students. Without such follow-up, the predictive value of models remains theoretical and fails to improve student outcomes.

IV. CONCLUSION

Predicting the academic success of college students using artificial intelligence and machine learning techniques has emerged as a powerful approach to improving educational outcomes, reducing dropout rates, and enabling timely interventions. While traditional methods provided limited insights, modern predictive models integrate academic, behavioral, demographic, and socio-economic variables to generate more accurate and actionable forecasts. However, challenges such as data quality, feature selection, class imbalance, model interpretability, and ethical concerns must be carefully addressed to ensure fairness and reliability. By combining advanced computational models with transparent, student-centered practices, institutions can harness predictive analytics not merely as a tool for assessment but as a means to foster personalized learning, equitable support systems, and long-term student success.

REFERENCES

1. J Guanin-Fajardo, J.H.; Casillas, J.; Guaña-Moya, J. Predicting Academic Success of College Students Using Machine Learning Techniques. Data 2024, 9, 60. https://doi.org/ 10.3390/data9040060.
2. Y. Zhang, L. Wang, and H. Chen, “Predicting student academic performance using deep learning with behavioral and socio-economic features,” *IEEE Access*, vol. 11, pp. 45782–45794, 2023.
3. M. R. Khan, A. U. Haque, and S. Alam, “Artificial intelligence-based early warning system for student success in higher education,” in *Proc. 2022 IEEE International Conference on Artificial Intelligence in Education (IC-AIED)*, Singapore, 2022, pp. 214–219.
4. N. Gupta and R. Sharma, “Machine learning approaches for predicting academic performance of students: A comparative study,” *IEEE Trans. Learning Technologies*, vol. 14, no. 4, pp. 472–484, 2021.
5. J. F. Rodríguez-Hernández, M. R. Ortega, and P. G. Martín, “An interpretable machine learning framework for predicting student dropout and academic success,” *IEEE Access*, vol. 8, pp. 212356–212369, 2020.
6. H. M. Alkhatlan and J. Kalita, “Intelligent prediction and classification of student academic performance in higher education using machine learning,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 9, pp. 504–515, 2019.
7. S. B. Kotsiantis, “Use of machine learning techniques for educational data mining to predict students' performance,” *International Journal of Artificial Intelligence and Applications*, vol. 9, no. 2, pp. 1–16, 2018.
8. C. Márquez-Vera, A. Cano, C. Romero, and S. Ventura, “Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data,” *Applied Intelligence*, vol. 48, no. 11, pp. 1–17, 2017.
9. L. Costa, R. Fonseca, and J. F. Souza, “Student performance prediction: A survey and case study with educational data mining,” in *Proc. 2016 IEEE Frontiers in Education Conference (FIE)*, Erie, PA, USA, 2016, pp. 1–9.
10. J. Márquez-Vera, C. Romero, and S. Ventura, “Predicting school failure using data mining,” *Expert Systems with Applications*, vol. 39, no. 18, pp. 13554–13560, 2015.
11. A. L. C. Baradwaj and S. Pal, “Mining educational data to analyze students' performance,” *International Journal of Advanced Computer Science and Applications*, vol. 2, no. 6, pp. 63–69, 2014.
12. C. Romero and S. Ventura, “Educational data mining: A review of the state of the art,” *IEEE Trans. Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601–618, 2012.