INTERNATIONAL JOURNAL OF ADVANCED AI RESEARCH

ISSN: **APPLIED**,Volume-1 Issue-1,
Inaugural Edition, April-June 2026 | **AIR-26V1i107**
DOI:

STUDENT SUCCESS PREDICTION IN HIGHER EDUCATION THROUGH MACHINE LEARNING

***Dr. N. CHANDRA MOULI,** *Associate Professor & HOD,*
Dept of CSE,
VAAGESWARI COLLEGE OF ENGINEEIRNG(AUTONOMOUS), KARIMANGAR

ABSTRACT: The purpose of this program is to predict the success of college students by utilizing machine learning techniques. These techniques involve the identification of patterns and variables that impact academic performance. Data-driven approaches are being used by educational institutions more and more. This is because algorithms like support vector machines, decision trees, and neural networks offer the possibility of early intervention and individualized help. The study looks at students' demographics, academic records, and behavioral traits in order to create predictive models that can group them into groups at risk of performing poorly. Teachers can improve students' performance, engagement, and retention by giving them the tools they need to make informed decisions. In addition to adding to the expanding body of knowledge in educational data mining, this research helps schools maximize the use of smart technologies to improve student performance.

Index Terms:Student Success Prediction, Higher Education, Machine Learning, Educational Data Mining, Predictive Modeling, Academic Performance, Student Retention, Intelligent Systems.

1. INTRODUCTION

The education sector is currently dealing with an unprecedented volume of data, which has opened up many new opportunities to improve classroom performance. Predicting college student academic performance with machine learning is promising. Due to rising enrollment and diversity, schools may struggle to identify at-risk students. Traditional methods based on subjective views and fads are outdated and inaccurate. Data-driven machine learning can improve student performance by creating proactive, individualized decision-making plans based on various data.

Complex and large datasets like students' socioeconomic status, attendance, learning styles, and academic records can be handled by machine learning algorithms. Machine learning algorithms search this mountain of data for new correlations and patterns to make accurate future predictions. Classifiers like Decision Trees, Random Forests, and Support Vector Machines can predict graduation, dropout, and GPA trends. As data accumulates, these models improve their predictions. They can also adapt to student needs.

After being implemented in schools, machine learning (ML) enabled more targeted and efficient interventions. Underachieving students should receive study materials, a mentor, or counseling. This early warning system helps teachers prevent problems instead of responding to them. These predictive models, which aim

ISSN: **APPLIED**,Volume-1 Issue-1,
Inaugural Edition, April-June 2026 | **AIR-26V1i107**
DOI:

to improve student performance, benefit policymakers, curriculum developers, and resource decision-makers. Moving from generic performance evaluations to targeted, individualized lesson plans can improve education.

Consider the ethical implications of using machine learning to predict student grades. If we want equal student outcomes, we must address data privacy, transparency, and algorithmic bias. Prevent inequality by training models with diverse and representative datasets. People must master data use before trusting these new tools. This means ethical AI can promote honesty and creativity in education.

Educational institutions are adopting machine learning (ML) as part of the digital revolution and data-driven decision-making. Schools' digital feedback systems, online assessments, and LMSs generate a flood of data for machine learning studies. Growing globalization and competition in higher education are driving the use of machine learning to predict student performance. This makes strategic and technological sense. Establishing a data-driven culture means constant improvement, more personalized assistance, and faster responses.

2.LITERATURE REVIEW

Alyahyan, E., &Düştegör, D. (2020).Students' final grades are the focus of this extensive study, which employs data mining and machine learning methods to make predictions. The writers stress that a student's first-year college performance, SAT/ACT scores, and high school grade point average are the most important predictors of future success. Educators can use the study's systematic predictive modeling approach to find the best machine learning algorithms, set goals for student success, and determine what students need to succeed. To help at-risk children get the help they need and get a head start in school, the report stresses the importance of early identification.

Niu, K., Cao, X., & Yu, Y. (2021). This research presents a free and open-source machine learning model for academic performance prediction that makes use of targeted instruction. The model does more than just make predictions; it also explains why a student might not succeed. By taking this tack, teachers can better comprehend their students' varied and unique approaches to learning. Students who are having difficulty will be singled out and helped to formulate plans to raise their scores.

Wang, Y., Ding, A., Guan, K., Wu, S., & Du, Y. (2021). The authors propose a new method for improving the accuracy of predicting students' performance using graph-based ensemble machine learning. The model uses supervised and unsupervised learning techniques, and its bipartite graph structure allows it to capture complex relationships between student-learning activities. The iterative method outperforms traditional methods of educational data analysis by 15.8% because it employs a wide variety of learning algorithms and graph topologies.

Ouatik, F., Erritali, M., Ouatik, F., &Jourhmane, M. (2022). The difficult problem of predicting students' academic performance is tackled in this investigation by using machine learning and big data technologies. Academic tests, psychological evaluations, attendance records, personal traits, and VE pursuits are just a few of the many sources queried by the writers. C4.5, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM) are among the methods used in the research, which is based on a Hadoop-based big data architecture. The accuracy

ISSN: **APPLIED**,Volume-1 Issue-1,
Inaugural Edition, April-June 2026 | **AIR-26V1i107**
DOI:

rate that SVM achieves is 87.32%. Educational decision-makers can benefit from big data techniques by making more accurate and timely predictions and by processing data more efficiently and scalable.

Orji, F. A., &Vassileva, J. (2022). This essay's principal goal is to establish a connection between students' study habits, academic performance, and intrinsic motivation. The authors train machine learning models that take into account 924 college-level dental students' levels of intrinsic motivation, autonomy, relatedness, competence, and self-esteem. Two tree-based algorithms, Decision Tree and Random Forest, performed better than the other five models. The 94.9% accuracy rate of Random Forest's forecasts stands out. Based on the findings, it appears that understanding and incorporating motivational factors into prediction models can improve the effectiveness of personalized learning and support systems.

Garg, A., Garg, N. B., Lilhore, U. K., Popli, R., Simaiya, S., & Bansal, A. (2023). This research presents a machine learning model that can foretell whether or not a student will thrive in a university setting. To find reliable models that can forecast students' results given a wide variety of academic and personal variables, the authors use a toolbox of algorithms. Instructional interventions and student support services can be better targeted using data-driven strategies, according to the study's results.

Jayasree, R., &Selvakumari, S. (2023).The authors use a prediction model that incorporates educational data mining and machine learning algorithms to assess students' progress. In sharp contrast to two other widely used methodologies—supported vector machines (SVMs) and randomized decision-making processes—we analyze the BPNN-SPA model. In terms of sensitivity, specificity, accuracy, and the F-measure, the BPNN-SPA model consistently outperforms competing models, making it a valuable tool for identifying struggling children and enabling timely treatments.

Tang, Z., Jain, A., &Colina, F. E. (2024). The main goal of the study is to compare different machine learning algorithms for their ability to forecast how well college courses will do. On a dataset with 4,424 records and 37 variables, the authors assess ten methods, focusing on Logistic Regression and Random Forest classifiers in particular. When class imbalance is considered using SMOTE, the Random Forest model performs better than the Logistic Regression model. The study highlights the importance of choosing the right algorithms and managing data imbalances well to improve prediction accuracy in learning environments.

Guanin-Fajardo, J. H., Guaña-Moya, J., & Casillas, J. (2024). This study uses a dataset of 6,690 records that includes academic and socioeconomic characteristics to forecast college students' academic performance using machine learning techniques. Classification algorithms and preprocessing techniques are investigated using the CRISP-DM methodology. According to the findings, XGBoost has an AUC of 87.75%. It is possible to understand the decision tree method's ten criteria, and it gets seven out of ten cases right. The results show that machine learning could be a powerful tool for guiding intervention efforts to improve students' academic performance and graduation rates.

Hassan, M. A., Muse, A. H., & Nadarajah, S. (2024). Using supervised machine learning methods and data from the 2022 National Education Accessibility Survey,

ISSN: **APPLIED**,Volume-1 Issue-1,
Inaugural Edition, April-June 2026 | **AIR-26V1i107**
DOI:

this study aims to shed light on the mystery of Somaliland's remarkable student retention rates. Taking into account factors like age, housing, household income, and school classification, the Random Forest model can potentially reach an accuracy level of 95%. By crunching numbers and finding out why students are leaving school, lawmakers can pass laws that improve graduation rates and educational outcomes.

Gichuru, E. N. (2024). College students' academic performance forecasting using ensemble learning frameworks like XGBoosts is explored in this article. Numerous variables are examined in the investigation, including ensemble size, prediction accuracy, efficiency, and learner complexity. In order to optimize hyperparameters, it recommends adding meta-learning, ensemble-aware regularization, and adaptive boosting to XGBoost. The outcomes provide ways to improve ensemble models in order to raise the precision of instructional predictions.

Jimenez Martinez, A. L., Sood, K., &Mahto, R. (2024). An early risk assessment system for college students based on machine learning is introduced in this investigation. The goal of this research is to develop prediction models that can assess a wide range of behavioral and academic factors in order to pinpoint students who might have trouble in school or think about dropping out. This research sheds light on the potential of data-driven programs to improve students' academic outcomes and highlights the value of early intervention.

Chen, J., Zhou, X., Yao, J., & Tang, S.-K. (2024). In order to better understand the possible uses of machine learning to forecast college students' engagement, performance, and self-efficacy, this literature review aims to undertake a thorough evaluation of empirical studies carried out between 2016 and 2024. A number of machine learning methods are covered in the piece. Methods such as decision trees, random forests, neural networks, and support vector machines make up these approaches. When compared to individual algorithms, ensemble approaches often produce better results. According to the study, machine learning could completely change the way at-risk children are identified, as well as personalized education and early intervention. On the other hand, it highlights the shortcomings of existing research, such as the narrow focus and lack of diverse viewpoints.

3. RELATED WORK

EXISTING SYSTEM

Universities are using machine learning to predict performance to increase graduation rates and decrease attrition. Systems that collect data from multiple sources are used to build prediction models. This dataset includes student demographics, attendance, engagement, academic, social, and behavioral records. Complex multivariate data sets are analyzed using decision trees, support vector machines, neural networks, and ensemble techniques to determine which features and patterns affect student performance. Administrators and teachers can use these models to classify students at risk. We can proactively allocate resources, counseling, or targeted support using these categories as early warning signs. Modern technology has improved retention and individualized lessons, but protecting student data, making models simple, and treating different student populations fairly are still issues. Many modern systems prioritize academic metrics over the complex social and psychological factors that affect

ISSN: **APPLIED**,Volume-1 Issue-1,
Inaugural Edition, April-June 2026 | **AIR-26V1i107**
DOI:

student achievement. As data integration and machine learning improve, prediction systems should become more fair and accurate. This implies better educational outcomes and student support systems are possible.

DISADVANTAGES OF EXISTING SYSTEM

- These technologies collect and process sensitive student data, posing privacy and data security risks.
- Many models ignore social, economic, and psychological factors, which affect students' academic performance more than academic and demographic data.
- If training data unfairly targets certain student groups, machine learning algorithms can make skewed predictions.
- Deep neural networks and other complex models can make accurate predictions, but their opaqueness makes it hard for teachers to understand and trust them.
- Historical forecasts often risk the system's adaptability by failing to account for curriculum, student behavior, and external factors.
- Due to their technological infrastructure, knowledge, and data administration requirements, many colleges may struggle to deploy and maintain these systems.
- Prediction errors may misclassify children, resulting in the unnecessary exclusion of at-risk youth or marking of non-interventionist students. In both cases, aid is wasted.
- Predictive analytics advocates have been criticized for stigmatizing "at risk" students and ignoring their consent.

PROPOSED SYSTEM

The suggested college success prediction method is more thorough, open, and moral than current methods. This system uses multiple data types to show each student's learning path. Attendance, socioeconomic status, psychological tests, academic records, LMS interactions, and real-time engagement statistics. New machine learning methods like ensemble learning, explainable AI (XAI) models, and adaptive algorithms help the system predict student performance and provide teachers with useful information. Secure multi-party computation and anonymization protect student data. To end discrimination and treat all students equally, this plan emphasizes equity in learning. Academic advisors and support staff can quickly create personalized learning paths and resources for each student using real-time data and early warning displays. Using the suggested method, universities can improve their data-driven model for student retention, academic success, and effectiveness. This system is unique because it is expandable, easy to use, and compatible with institutional infrastructures.

ADVANTAGES OFPROPOSED SYSTEM

- In an effort to conduct an exhaustive analysis, it draws from an extensive array of sources, including academic, economic, psychological, and real-time engagement data.
- Complex machine learning and ensemble models improve academic performance predictions for future students.
- The model uses clearly defined AI techniques to instantly reveal its findings, so educators can trust the predictions.

INTERNATIONAL JOURNAL OF ADVANCED AI RESEARCH

ISSN: **APPLIED**,Volume-1 Issue-1,
Inaugural Edition, April-June 2026 | **AIR-26V1i107**
DOI:

- Data anonymization and other privacy measures protect students' personal information.
- Equity-conscious algorithms reduce bias and demonstrate our value for diverse students.
- Real-time analytics and early alerts allow instructors and aides to intervene.
- lets teachers distribute course materials and tailor lessons to individual students.
- Many colleges have access to it, and it is compatible with existing systems.
- supports monitoring student behavior to prevent problems rather than fix them.
- uses ethical standards to ensure fair and intentional predictive analytics in the classroom.

MODULES

- Data Collection Module
- Data Preprocessing Module
- Exploratory Data Analysis (EDA) Module
- Feature Selection Module
- Model Development Module
- Model Evaluation Module
- Prediction Module
- Interpretability & Explanation Module
- Dashboard & Visualization Module
- Intervention & Recommendation Module
- Feedback & Model Update Module
- Security & Privacy Module

4.RESULTS AND DISCUSSIONS



Fig1. User Login



Fig2. User Register



Fig3. Register Your Details



Fig4. View all remote users



Fig5. Dataset Details



Fig5. Prediction of student Academic performance

5.CONCLUSION

Machine learning's ability to predict college success has changed how schools interact with and help students. Traditional methods miss complex patterns and early signs of student success, but machine learning algorithms can identify students at

risk of failing goals or being excluded. Next, give them personalized help to succeed. This predictive capability is enabled by a mountain of demographic, behavioral, and academic data. Individualized programs can boost graduation and retention rates and predict students' academic success.

Data-informed and proactive practices enabled by machine learning prediction tools help higher education institutions shift from reactive to proactive planning and resource allocation. The model is constantly improved to provide accurate, relevant, and equitable forecasts for all student demographics. This process incorporates ethical concerns and many types of data. Technical progress allows performance-based feedback loops, which power supplemental services and educational programs and encourage continuous improvement.

REFERENCES

1. Alyahyan, E., &Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. International Journal of Educational Technology in Higher Education, 17(3).
2. Niu, K., Cao, X., & Yu, Y. (2021). Explainable Student Performance Prediction With Personalized Attention for Explaining Why A Student Fails. arXiv preprint arXiv:2110.08268.
3. Wang, Y., Ding, A., Guan, K., Wu, S., & Du, Y. (2021). Graph-based Ensemble Machine Learning for Student Performance Prediction. arXiv preprint arXiv:2112.07893.
4. Ouatik, F., Erritali, M., Ouatik, F., &Jourhmane, M. (2022). Predicting Student Success Using Big Data and Machine Learning Algorithms. International Journal of Emerging Technologies in Learning (iJET), 17(12), 236–251.
5. Orji, F. A., &Vassileva, J. (2022). Machine Learning Approach for Predicting Students Academic Performance and Study Strategies based on their Motivation. arXiv preprint arXiv:2210.08186.
6. Garg, A., Garg, N. B., Lilhore, U. K., Popli, R., Simaiya, S., & Bansal, A. (2023). Machine Learning-based Model to Predict Student's Success in Higher Education. Proceedings of the 3rd International Conference on ICT for Digital, Smart, and Sustainable Development (ICIDSSD).
7. Jayasree, R., &Selvakumari, S. (2023). Design of a Prediction Model to Predict Students' Performance Using Educational Data Mining and Machine Learning. Engineering Proceedings, 59(1), 25.
8. Tang, Z., Jain, A., &Colina, F. E. (2024). A Comparative Study of Machine Learning Techniques for College Student Success Prediction. Journal of Higher Education Theory and Practice, 24(1).
9. Guanin-Fajardo, J. H., Guaña-Moya, J., & Casillas, J. (2024). Predicting Academic Success of College Students Using Machine Learning Techniques. Data, 9(4), 60.
10. Hassan, M. A., Muse, A. H., & Nadarajah, S. (2024). Predicting Student Dropout Rates Using Supervised Machine Learning: Insights from the 2022 National Education Accessibility Survey in Somaliland. Applied Sciences, 14(17), 7593.
11. Gichuru, E. N. (2024). Predicting Student Success in Higher Education: Understanding the Impact of Ensemble Learning Architecture on Model Performance. Data Science and

Artificial Intelligence Conference Proceedings.

12. Jimenez Martinez, A. L., Sood, K., &Mahto, R. (2024). Early Detection of At-Risk Students Using Machine Learning. arXiv preprint arXiv:2412.09483.
13. Chen, J., Zhou, X., Yao, J., & Tang, S.-K. (2024). Application of machine learning in higher education to predict students' performance, learning engagement and self-efficacy: a systematic literature review. Asian Education and Development Studies.
14. K. K. Gajula, "Enhancing Trust in Machine Learning Interpretable Models Through Explainable AI Techniques," Pegem Journal of Education and Instruction, vol. 13, no. 4, pp. 909–915, 2023.
15. K. K. Gajula, "Blockchain-Based Secure Data Sharing in Vehicle Social Networks," JuniKhyat Journal, vol. 12, no. 1, pp. 217–223, 2022.
16. K. K. Gajula and A. T. Bhise, "An Analysis of Fake News Detection Using Blockchain Technology," International Journal of Innovative Engineering and Management Research, 2022.
17. M. K. Srinivasan and K. K. Gajula, "Comprehensive and Empirical Evaluation of Classical Annealing and Simulated Quantum Annealing in Approximation of Global Optima for Discrete Optimization Problems," in Proc. ICTIS, 2021, pp. 165–181.
18. K. K. Gajula and K. Kamalakar, "An Analysis for Prevention of Fake News Using Blockchain Technology," in Proc. National Conf. on Recent Advancements on Computer Science (CONRACS), 2019.
19. K. K. Gajula, Y. K. Sharma, and R. Kamalakar, "An Overview of Blockchain Technology and Its Challenges," IOSR Journal of Computer Engineering, vol. 21, no. 3, pp. 40–45, 2019.
20. K. K. Gajula and K. Kamalakar, "A Survey on Feature-Specific Quality Prediction of Product Reviews Using Sentiment Analysis," International Journal of Pure and Applied Mathematics, vol. 118, no. 24, 2018.