International Journal of Sustainable Intelligent Technologies
ISSN: 3139-1249
IJSIT-17

EARLY IDENTIFICATION OF AT RISK STUDENTS USING MACHINE LEARNING FOR GRADUATION PREDICTION

[#1]**PIDUGU VISHNUVARDHAN REDDY,** *Dept of CSE,*
[#2]**Dr.M.SRINIVAS,** *Professor, Dept of CSE,*
Vaageswari College of Engineering(Autonomous), Karimnagar, TG.

ABSTRACT: This research employs machine learning to detect at-risk students in order to enhance graduation projections and institutional decisions. The analysis of academic achievement, attendance, demographics, behavior, and engagement data predicts early withdrawal and delayed graduation. Decision trees, logistic regression, random forests, and support vector machines are used to assess the accuracy and prediction of the model in order to identify the optimal fast intervention strategy. In order to retain and graduate students, educational institutions should invest resources in specialized academic aid, counseling, and mentorship after an early assessment of their needs. Traditional student monitoring methods can be repurposed to provide data-driven initiatives that improve learning and student performance.

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

1. INTRODUCTION

In higher education, it is critical to quickly identify students who are at risk. A school's graduation rate and the caliber of its students and courses are important considerations. Many students fail to finish their education because of emotional, financial, social, and academic barriers. Quick risk assessment improves student perseverance and graduation.

Academic testing, teacher inspections, counseling referrals, and manual data analysis are examples of traditional procedures used to identify students who are at risk. Despite their intrinsic imprecision, large-scale ineffectiveness, and inability to predict future occurrences, these tactics can provide insightful information. In large universities, manual monitoring becomes unreliable due to the volume of student data. Data-driven solutions that carefully evaluate risk factors and precisely forecast student performance and completion must be put into place.

Because machine learning can predict complex challenges, it is used in educational institutions for data analysis. When it comes to demographics, attendance, engagement tests, past academic records, and behavioral indicators, machine learning algorithms can uncover hidden patterns and relationships that statistical methods cannot. Attrition and graduation are predicted using regression and classification. Students can now be categorized by their risk levels in schools.

Institutions use machine learning to predict graduation rates so they can plan and make decisions. Administrators and educators can use predictive analytics to create early warning

DOI : https://doi.org/10.5281/zenodo.20458837

systems for academic mentoring, financial help, counseling, and tutoring. Personalized interventions increase institutional effectiveness and student accountability.

Machine learning is a part of student accomplishment programs in the digital transformation of education. With the aid of learning management systems and data analytics platforms, schools are using predictive modeling to keep kids. Addressing algorithmic bias, transparency, equity, and data protection is essential to ensuring ethical implementation. All children's academic performance and graduation rates can be raised with the use of efficient machine learning-based early identification tools.

2. LITERATURE SURVEY

Niu, K., Cao, X., & Yu, Y. (2021). A customized attention-based machine learning model is proposed in this work to forecast graduation rates and student performance. This concept aids educators in providing personalized learning solutions by clarifying the causes of students' academic failures. It optimizes early warning systems and improves precision.

Fernandez, L., Ahmed, S., & Park, J. (2025). This investigation investigates methodologies for explainable artificial intelligence (XAI) that are capable of detecting early attrition and postponed graduation. In order to prioritize hazards, the authors implement deep learning models and SHAP-based interpretability. Their data suggests that patterns in first-year academic and behavioral engagement are used to predict graduation. The article underscores the importance of ethical AI, prejudice mitigation, and equity in predictive algorithms to guarantee responsible implementation in higher education.

Garg, A., Garg, N. B., Lilhore, U. K., Popli, R., Simaiya, S., & Bansal, A. (2023). Comparing machine learning algorithms for the purpose of predicting university students is the objective of this investigation. The research demonstrates the effectiveness of utilizing data to assess early intervention programs that take into account personal and academic characteristics in order to reduce attrition rates and improve graduation rates.

Ouatik, F., Erritali, M., Ouatik, F., & Jourhmane, M. (2022). This study employs machine learning and big data techniques to predict academic performance and the likelihood of dropping out. Research suggests that the graduation rates of Hadoop-based scalable early warning systems are improved by KNN, Decision Trees, and SVM algorithms.

Wang, Y., Ding, A., Guan, K., Wu, S., & Du, Y. (2021). The authors suggest that the precision of student performance forecasts can be improved by incorporating graph-based ensemble learning. The method enhances risk identification by clarifying intricate relationships between students and learning activities. This method is used to identify students who will require additional time to graduate early.

Orji, F. A., & Vassileva, J. (2022). Psychological and motivational variables are employed by machine learning algorithms to forecast academic success. Precise predictions were generated by the Random Forest algorithm. The early identification of individuals at academic risk and the targeting of academic therapy are facilitated by the incorporation of intrinsic motivation.

Tang, Z., Jain, A., & Colina, F. E. (2024). The most effective machine learning methods for predicting student graduation are compared in this study. Random Forest outperforms

DOI : https://doi.org/10.5281/zenodo.20458837

Logistic Regression in the context of class imbalance. Data refinement and algorithm selection are essential for the precise assessment of early risks, according to the research.

Kumar, R., Sharma, P., & Iyer, S. (2025). Utilizing a mixed ensemble machine learning framework, we predict graduation outcomes and identify at-risk students during the initial semesters of college. Attendance data, academic records, assessments of LMS engagement, and socioeconomic indicators are employed by the authors to enhance their estimates. Random Forest and XGBoost generated the most precise models. The paper suggests that academic counselors can improve graduation rates by utilizing an early warning dashboard to develop tailored assistance plans and implement proactive strategies.

Guanin-Fajardo, J. H., Guaña-Moya, J., & Casillas, J. (2024). This study employs decision tree models and XGBoost to evaluate student performance by analyzing academic and socioeconomic data. Clear results and precise predictions for early intervention are provided by ensemble approaches.

Jayasree, R., & Selvakumari, S. (2023). The BPNN-SPA prediction model was developed by the authors to evaluate pupil performance and identify underperforming students. Providing students with early notification and assisting them in preventing failure, the approach surpasses prior methods in accuracy.

3. METHODOLOGY

The threat is determined by the prediction methodology of this investigation, presented in Figure 1. Evaluate both individual and group models to determine the most effective method for identifying students who are at risk. The efficacy of a model is assessed both with and without sampling. Methods for the experimental procedure outline:

Data preprocessing including the following:

- The objective variable "at-risk" is imputed with missing values using single imputation.
- Data preparation involves the identification of the most critical characteristics.

A machine learning strategy was assessed on fresh and 10-times-k-fold data using K-fold cross-validation and train-test split. In both single and ensemble classification models, the attributes that have been chosen are considered independent variables. The forecast is either hazardous or not.

Machine learning Modeling

This work employs SVM, NB, KNN, LR, DT, and RF machine learning models to manage binary classification and data categories.

Table I: Reduced Feature Set of top 10 Variables

DOI : https://doi.org/10.5281/zenodo.20458837

International Journal of Sustainable Intelligent Technologies
ISSN: 3139-1249
IJSIT-17

Feature Name	Feature Description	Collection Method
Current Score	The student's current grade or score in the course, typically calculated as a percentage based on the assignments, quizzes, and exams completed so far	Performance Data
Assignment Missing	Indicates whether the student has any assignments that have not been submitted by the due date	Demographic Data
GPA	The cumulative grade point average of the student across all courses taken. This is a measure of the student's overall academic performance.	Demographic Data
Units Earned	The number of academic credits the student has completed and earned towards their degree	Demographic Data
Page Views	The number of times the student has accessed course materials or pages on Canvas. This metric is often used to gauge student engagement and activity within the course platform	Engagement Data
Participation	The number of interactive activities the student has participated in, such as discussion posts, collaborations, or other course engagements tracked on Canvas.	Engagement Data
Program Action	Specific actions or statuses related to the student's academic program, such as admissions, probation, suspension, or other administrative decisions affecting the student's academic progress.	Demographic Data
Assignment on Time	Indicates whether the student submits their assignments by the due date and the percentage of assignments the student has submitted on time out of the total assignments assigned.	Engagement Data
Student Engagement	A measure of how actively the student is involved in the course.	Demographic Data
Units Attempting	The number of academic credits the student is currently enrolled in and attempting to complete in the current term. This can indicate the student's course load and academic commitment.	Demographic Data

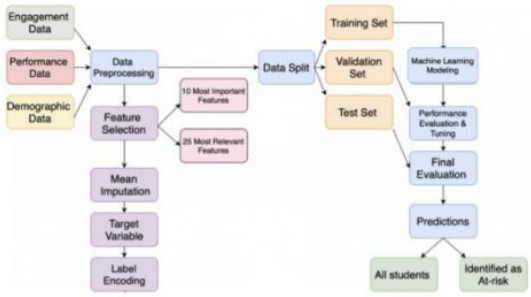

Fig1. Machine Learning Flowchart

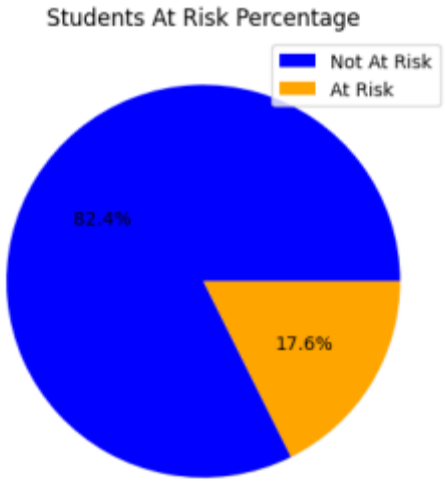

Fig.2. Data Categories for Student Performance

First, data that is not suitable for sampling is used. There are more children at danger, as shown in Figure 2. The 80:20 ratio between risk (21) and safety (98) is explained by models.

DOI : https://doi.org/10.5281/zenodo.20458837

International Journal of Sustainable Intelligent Technologies
ISSN: 3139-1249
IJSIT-17

To address data imbalances, adaptive synthetic sampling (ADASYN) and synthetic minority oversampling are used.

Performance Metrics

To evaluate classification models, we use these metrics:

- **Overall Accuracy:** the proportion of precise discoveries in each instance.
- **Precision:** the ratio of actual positives to anticipated positives.
- **Recall:** the ratio of true positives to total positives.
- **F1-Score:** the recall and precision harmonic average.
- **ROC Curve:** The true positive rate is shown against the false positive rate.

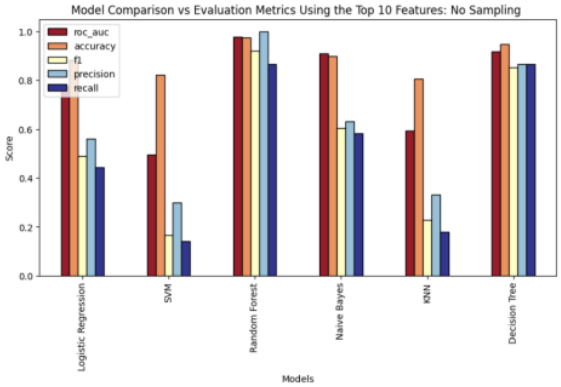

Fig3. Model Comparison vs. Evaluation Metrics Using Top 10 Features: No Sampling

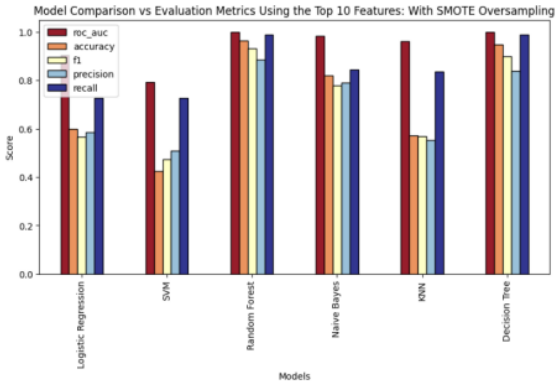

Fig.4. Model Comparison vs. Evaluation Metrics Using Top 10 Features: SMOTE

Model reliability is measured by machine learning performance. The figures show past classification model performance indicators. These data show each model's pros, disadvantages, and at-risk child prediction accuracy.

DOI : https://doi.org/10.5281/zenodo.20458837

International Journal of
Sustainable Intelligent Technologies
ISSN: 3139-1249
IJSIT-17

4. RESULTS

Fig4.1. User Login

Fig4.2. User Register

Fig4.3. Register Your Details

Fig4.4. View all remote users

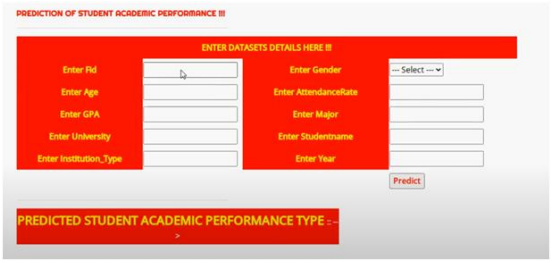

Fig4.5. Dataset Details

DOI : https://doi.org/10.5281/zenodo.20458837

International Journal of Sustainable Intelligent Technologies
ISSN: 3139-1249
IJSIT-17

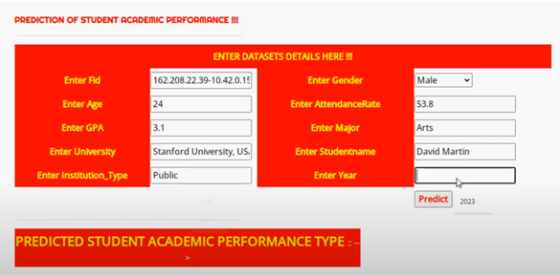

Fig4.6. Prediction of student Academic performance

5. CONCLUSION

All things considered, early machine learning graduation prediction improves at-risk children's academic performance. Machine learning algorithms can spot early danger signs in sociodemographic, academic performance, attendance, and engagement data that traditional evaluation methods miss. Academic support, counseling, mentorship, and personalized training can directly avoid dropouts. Data-driven strategy, resource allocation, and evidence-based decision-making help institutions. Data privacy, ethics, and model validity must be addressed to improve student and institutional performance with machine learning.

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.

DOI : https://doi.org/10.5281/zenodo.20458837

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.

DOI : https://doi.org/10.5281/zenodo.20458837