IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, January 2026
Impact Factor: 7.67

Predictive Analysis of Student Dropout

Atharwa Baban Pawar[1], Aryan Sachin Gangawane[2], Saksham Deepak Choundiye[3], Prof. N. S. Kharatmal[4]
Students, Computer Science and Engineering[1,2,3]
Lecture, Computer Science and Engineering[4]
Matsyodari Shikshan Sanstha College of Engineering and Polytechnic, Jalna, India
pawaratharwa75@gmail.com[1], Aryan.gangwane0025@gmail.com[2],
sakshamchoundiye01@gmail.com[3], nanditakhartmal27@gmail.com[4]

Abstract: *Dropout rates have remained a major problem for institutions of higher learning, and performance at the institutional level and academic and socio-economic development of students. Early identification of dyslexicsat-risk students allows schools to go beyond simply reacting to problems and take proactive steps to effectively address the issues of these individuals. By doing so interventions. The paper outlines an extensive framework for the predictive analysis of dropout in students. using data-driven and machine learning-based methods. In particular, the research becomes significant with respect to a multidimensional data set, involving the following variables: attendance trends, demographic characteristics, and socio-economic conditions. To overcome the limitations of traditional statistical methods to capture the complex and non-linear relationships of machine learning models are used. Comparison of ensemble learning techniques, such as Random Forest, Gradient Boosting (XGBoost) and LightGBM are used to determine the efficacy of models such as students having a high possibility of attrition. Additionally, methods of feature selection and data preprocessing. are used to improve the performance of models.*

Owing to the natural class imbalance problem in the student datasets, the number of dropout cases is much lower compared to regular cases. This class imbalance problem successful completion of models, the Synthetic Minority Over-sampling Technique (SMOTE) is employed to enhance sensitivity of the statistical models used, and reliability of the research implementation process. Findings from the experiment show, attendance regularity, and financial viability factors are among the most significant predictors of student persistence. The most accurate models are capable of achieving high levels of prediction accuracy and F1 scores, thus identification of high-risk students.

Keywords: *Dropout rates*