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 4, Issue 6, November 2024
Impact Factor: 7.57

Application of AI and Data Analysis for Classification of Student Success in Large-Scale Educational Dataset

Adarsh Reddy Bilipelli
Independent Researcher

Abstract: *A major obstacle in higher education is the early detection of students who are at risk of dropping out, which has major consequences for retention rates and students' general performance in the classroom. New possibilities for data-driven interventions are opening up as large-scale educational datasets like OULAD become more widely available. Traditional prediction methods, however, are frequently rendered ineffective by educational data due to its intrinsic complexity, heterogeneity, and imbalance. This research presents a scalable and powerful machine learning framework for early student success prediction that makes use of AI and sophisticated data analysis methods. After cleaning, imputation of missing values, normalization, one-hot encoding, and feature engineering using recursive feature elimination (RFE), the OULAD dataset was ready for preprocessing. After determining which variables were most predictive, split the dataset in half to use for training and testing. A total of three models, namely LGBM, ANN, and FCM-RF (Fuzzy C-Means with Random Forest), were trained and assessed using critical metrics. Outperforming ANN (accuracy: 78% and FCM-RF accuracy: 88.33%), experimental results show that LGBM obtained higher predictive performance with a precision of 94.40%, recall of 93.21%, and F1-score of 93.86%. The findings highlight the effectiveness of ensemble-based methods, particularly LGBM, in accurately identifying at-risk students, enabling timely and targeted interventions to enhance institutional outcomes.*

Keywords: Early prediction, machine learning, educational data mining, OULAD Dataset, LGBM Model, academic performance, student retention

DOI: 10.48175/IJARSCT-22564