

Literature Review

This review looks at two studies focused on predicting student dropout in higher education. The first study, Early Prediction of Students' Performance in Higher Education by Martins et al. (2021), used machine learning to predict dropout based on data from a Portuguese university. They used information like students' academic history and personal background. To handle the imbalance (since there were fewer dropouts compared to successful students), they used a method called SMOTE. They tested several models, and Gradient Boosting worked best overall, but it still struggled to predict dropouts accurately. The second study, Exploring Statistical Approaches for Predicting Student Dropout in Education by Goyal & Kumar (2023), reviewed different methods used to predict student dropout. They used data from various institutions and tested models like Decision Trees and Neural Networks. Again, SMOTE was used to fix the imbalance in the data. They found that Decision Trees, especially Random Forests, gave the best results, but challenges with data imbalance and different dropout factors led to mixed results.