

Abstract

Enrolling for the wrong programme by university students has, to an extent, contributed to the high rates of discontinuation on academic grounds, repeat year cases, change of programme after registration, interuniversity transfers, deferments to change programme, drop out cases, suspension over exam irregularities as well as to strikes. This study focused on finding a technological solution for reducing these cases by evaluating three tree-based predictive models and recommending the most predictive model to implement as a programme recommender. Data was collected in five selected public universities in Kenya using Google Forms. The respondents were 308 translating to 308 rows of data with 36 columns. Numpy, Pandas, Matplotlib, Sklearn, Seaborn, Scipy, Plotly python analytics libraries were deployed using Jupyter Notebook for Anaconda. The cleaned and processed dataset features had categorical variables thus one-hot-encoding technique was employed. Data was split for training and testing with the `random_state` set to 42. Gini index criteria was implemented.

The three models were evaluated on their performance from the optimally split data for training and test with a 80:20 ratio. Random Forest (RF) came out the most predictive at 99.3% followed by Gradient Boosting (XG Boost) at 90% then Decision Tree (DT) at 80.93%. The testing accuracy score for RF was 81.72%, XGBoost was at 75.72% and DT was at 76.34%. Confusion matrix criterion was implemented to evaluate the performance of the three models. The results of this study have demonstrated the high accuracy level of RF as the most predictive tree-based model for this real-world University crisis. The model is recommended for development as a system to be integrated into the KUCCPS portal. The integrated system is dubbed Programme Recommender which if launched would highly predict the best programme of study for application by university entrants.