

# **Predicting Undergraduates Dropouts Using Classification Techniques**

**Case Study on Degree Awarding Non-State  
Higher Education Institute**

## Abstract

Undergraduate dropout is one of the biggest concerns in higher education institutes. This has become a significant concern locally as well as globally. Student retention has gained more attention from university administrators, especially those in private-sector higher education institutes as the competition is quite high in the private sector. This research's main objective is to predict undergraduate dropouts in the Information Technology Degree Program of a non-state higher education institution in Sri Lanka.

This research has used secondary data, the examination results related to student's performances were collected from the examination department and students' demographic data were collected internally from the Information Technology Faculty. A review of the literature help to identify important variables that have been studied in the past, according to the findings undergraduate, age, gender, student performances, marks obtained for the English Language, whether the student has obtained a government Loan or not, and their pre-university education level like Advanced Level Stream were considered as the variables of this study. According to the analysis, it has been identified that there is no significant relationship between age and dropout or non-dropout classes and no significant association between Gender and the Dropout class, but it also identified through the analysis, that Advanced Level stream, Loan, and student performances (which is measured through GPA value) have a significant association with Dropout

Logistic Regression, Random Forest, Naïve Bayes, Artificial Neural Network (ANN), Decision Tree, and Support Vector Machine (SVM) classification techniques were used for the prediction. According to the results, SVM has the best F1 score which is 90%, ANN, Decision Tree, and Logistic regression got 88%, and Random Forest and Naïve Bayes have an 87% of F1 score. It has also been identified that dropouts are high in those who have done Advanced Level in Art Stream and under Other Category.

Therefore, before students get register from those categories if faculty can give them an aptitude test and select the relevant candidates, will be helpful to reduce the dropouts. Data mining techniques can improve the quality of education in non-state higher education institutes as this helps to identify the hidden patterns of educationally linked data. This research on undergraduate dropout in non-state higher education institutes is essential because it helps the institute administrators and policymakers to understand the reasons behind student attrition and devise strategies to address the issue. Private universities are businesses and losing students to dropouts can significantly impact their revenue and reputation. Understanding why students drop out is crucial to developing interventions that can improve retention rates. By conducting this research on predicting undergraduate dropout, private sector institutes can identify risk factors and develop proactive approaches to support at-risk students.