

# Model Development and Training Report

## INTRODUCTION

This report details the development of the machine learning models used in the classification of student dropout. It takes into consideration two main models, Random Forest classifier and Decision Trees Classifier.

These two models were employed in this project and the best performing model out of the two was selected.

## Model description

### 1. Decision Trees

- **Use case:** Can be used when feature interactions and non-linearity are important.
- **Architecture:** A tree structure where each internal node splits the data based on a feature, and each leaf represents a predicted class.
- **Advantages:** Captures non-linear relationships. o Easy to interpret and visualize.
- **Disadvantages:** Prone to overfitting on complex datasets unless properly tuned

### 2. Random Forest

- **Use case:** This is a more robust model than a single decision tree, it is very useful when higher accuracy is desired, and the dataset has many features.
- **Architecture:** It consists of an ensemble of decision trees. Each tree is built from a random sample of the data, and the final output is based on the majority vote (classification).

- **Advantages:** Reduces overfitting compared to a single decision tree. It Handles large numbers of features well.
- **Disadvantages:** It is Less interpretable than individual decision trees as it contains many trees.

## **Summary**

The Random forest classifier model had a better performance than the Decision Trees Classifier model and ultimately, the best performing model was used in model development.