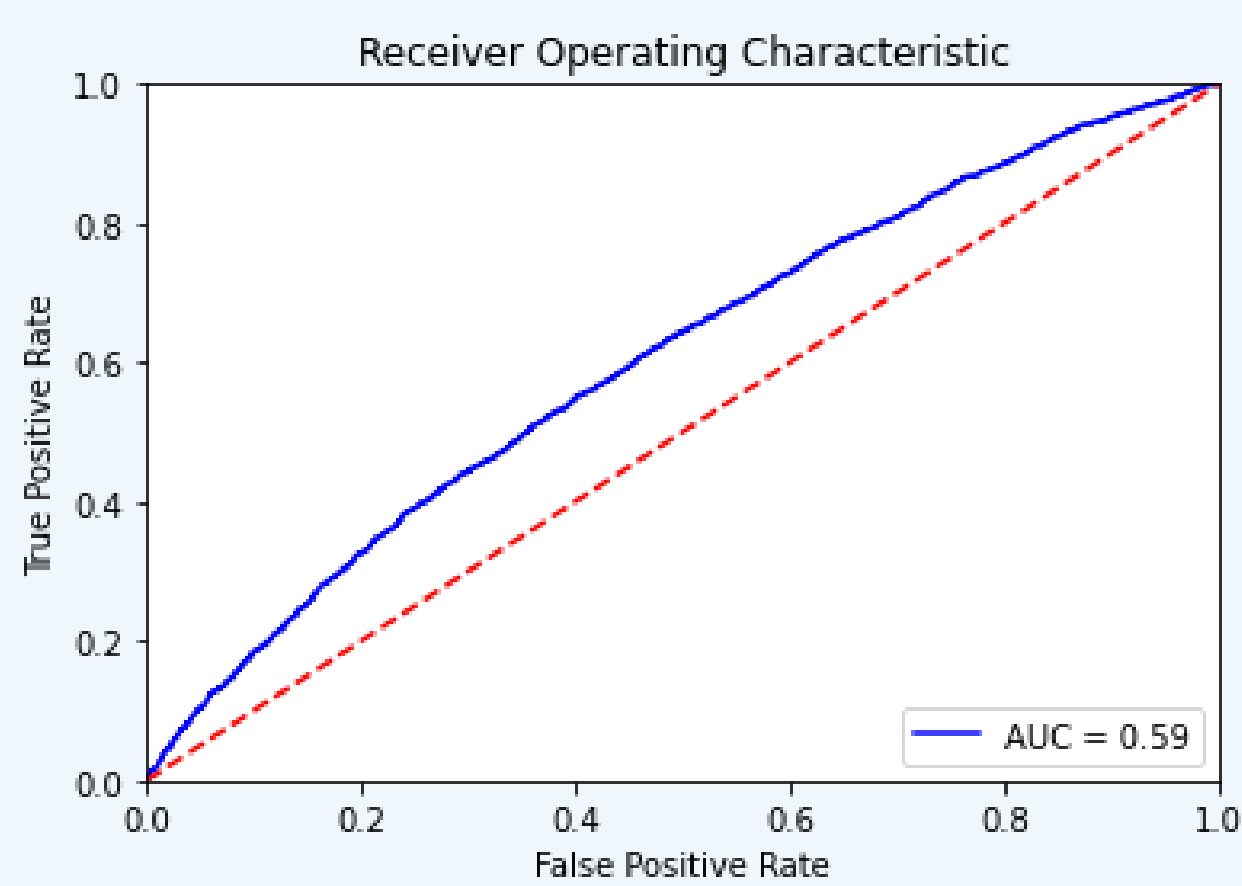


Univ.AI AI-1 Final Project: Churn Prediction

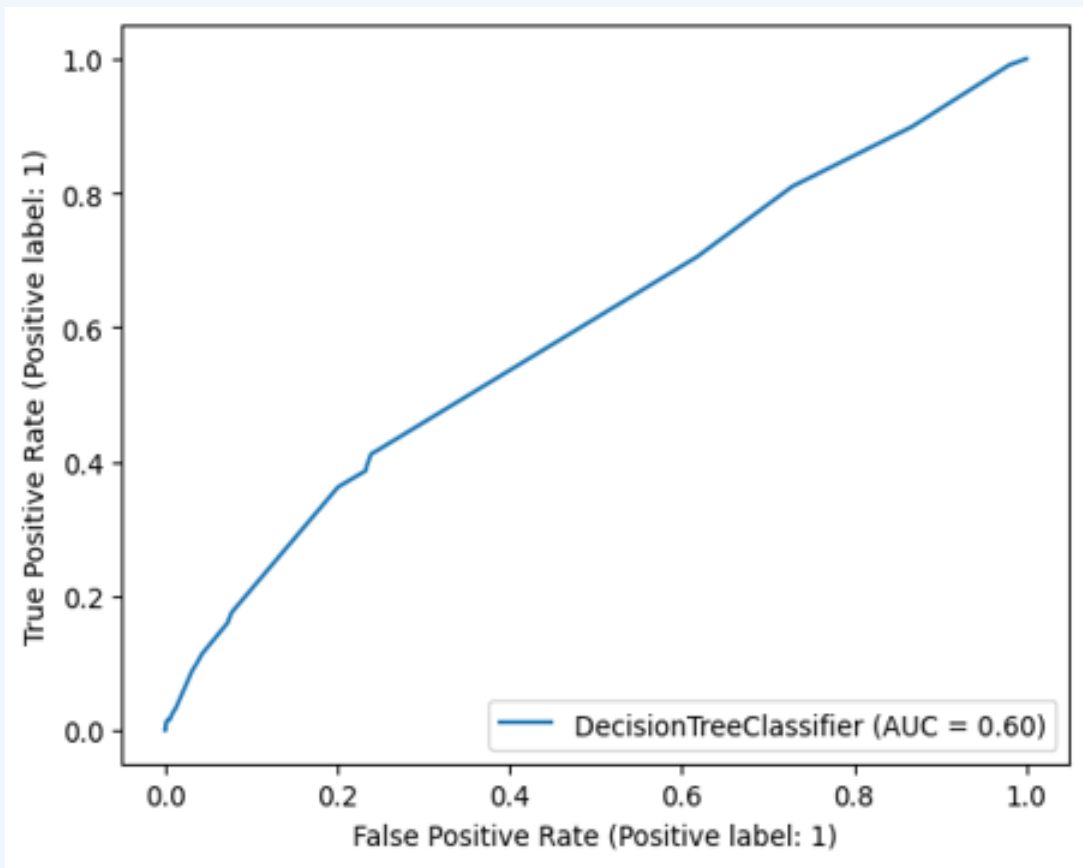
Background

We are given a real-life dataset of a telecom operator and tasked to predict the propensity of customers to switch provider (churn), and is a classical classification problem. This problem is important, as churning can cause losses to the operator, and predicting prospective churners can allow the operator to provide incentives against churning, thus reducing loss. Since it is a large dataset, with many variables and NaN data, dealing with it is a challenging yet rewarding problem. We will put four classification models and try to check their performance against a validation set drawn from the train data.

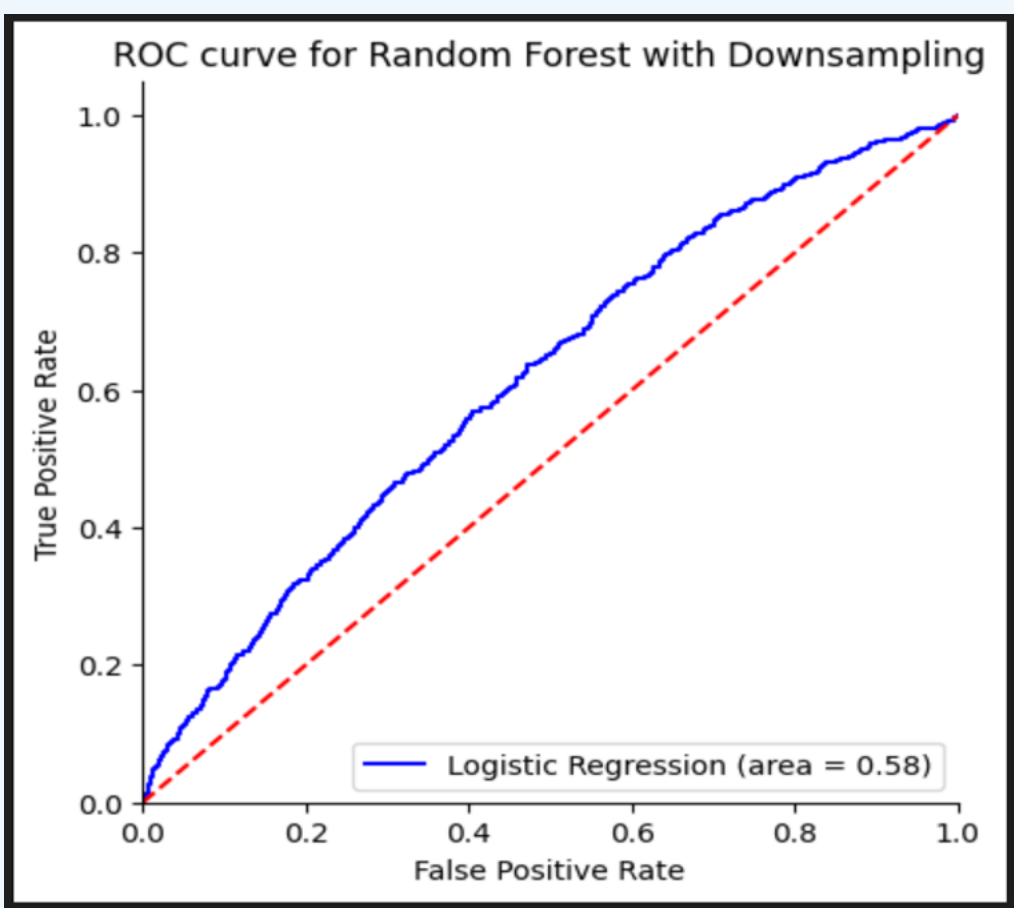
Logistic Regression with Upsampling



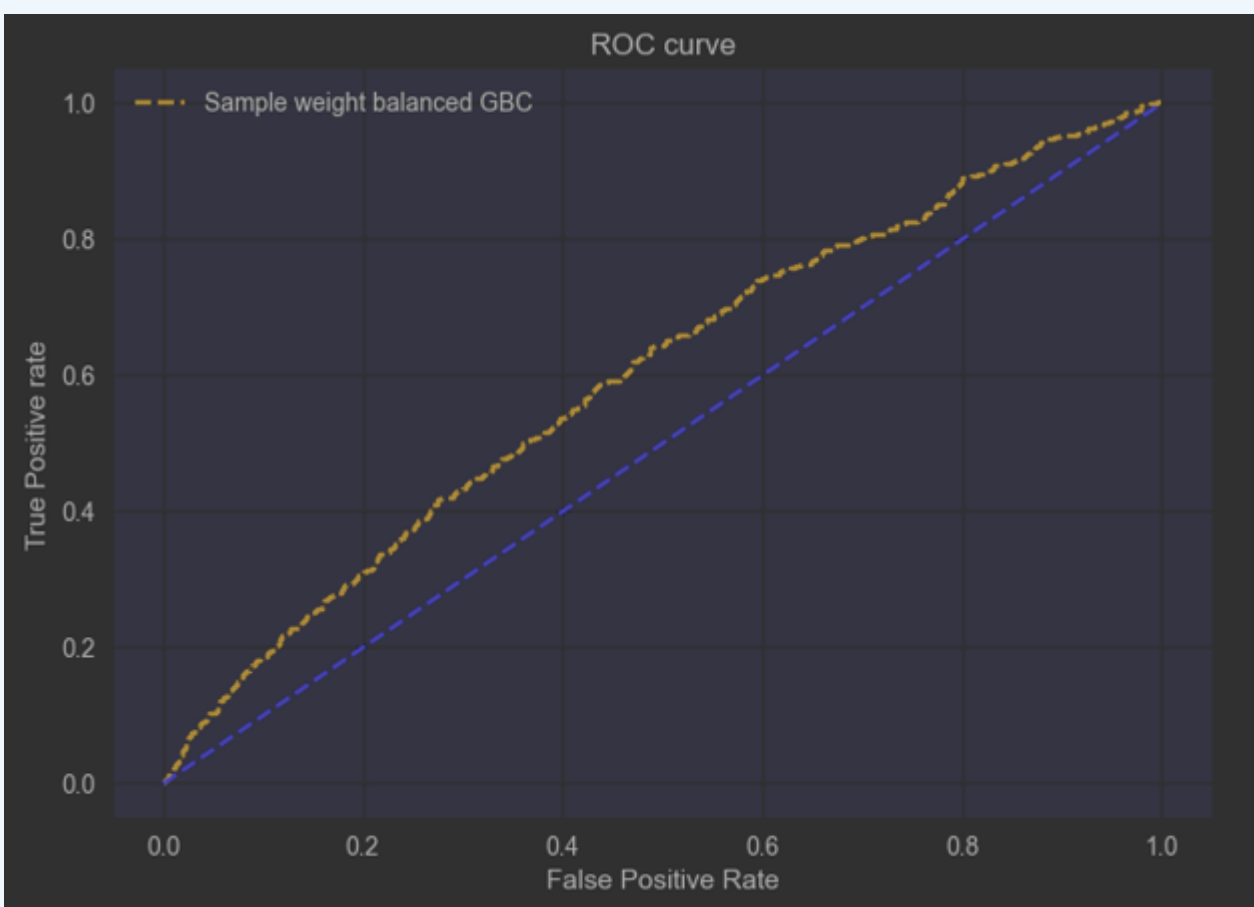
Decision Trees



Random Forests with Downsampling



Weight-balanced Gradient Boosting



Results

Table 1: Model with ROC AUC Score

Model Used	ROC-AUC Score
Logistic Regression	0.59
Decision Trees	0.60
Random Forest	0.58
Gradient Boosting	0.58
2009 SOTA	0.76

Takeaway

We see that the complication of the data makes it difficult to deal with, but we have made significant progress using what was learnt in AI - 1. More rigorous feature engineering, with finely tuned models on good hardware will lead to much better results.