

Predicting Customer Churn at Syriatel

...

Building an ML model that predicts from various features whether a customer would stick or churn.

Introduction

Syriatel wants to predict which customers are likely to drop them so as to put measures forehand to mitigate that.

There are records in the database detailing customer mobile services usage and whether they churned or not.

We have analyzed this data and developed a machine learning model that will predict whether a customer will drop Syriatel or not.

The Data

From Syriatel's database, necessary features for this analysis have been collated in this [file](#).

These features include number and length of calls made by a customer at various times of day, call charges, and plans that a customer has subscribed to.

The target feature is the churn column which tells us whether or not the customer churned.

Modelling

The data was first split into a training set and a test set.

Data was then analyzed and preprocessed for the Machine Learning task ahead. Highly correlated features were dropped and continuous variables standardized.

Two modelling techniques were then used to develop binary classifiers that predict whether a customer would churn or not.

- Decision Trees
- Logistic Regression

Decision Trees

We began with a simple Decision Tree model that implements default hyperparameters.

We fit the model to the training data and generated its predictions against “unseen” test data.

We evaluated it using the Area Under the Curve (AUC) metric and found the model overfitting, therefore not suitable.

To improve the model, we tuned different hyperparameters.

Decision Trees - Tuning

A range of values were tried for various Decision Tree model hyperparameters like maximum depth and maximum features.

The results of all tuning were analyzed through graphs and resultant AUC scores.

Optimal hyperparameter values were identified and applied to the final Decision Tree model.

There was a significant improvement to the earlier model, and consequently more confidence in predictive performance.

Logistic Regression

To investigate the performance of another modelling technique, we built a Logistic Regression model.

The initial basic model with default hyperparameters did not overfit, but had lower AUC scores compared to the earlier models.

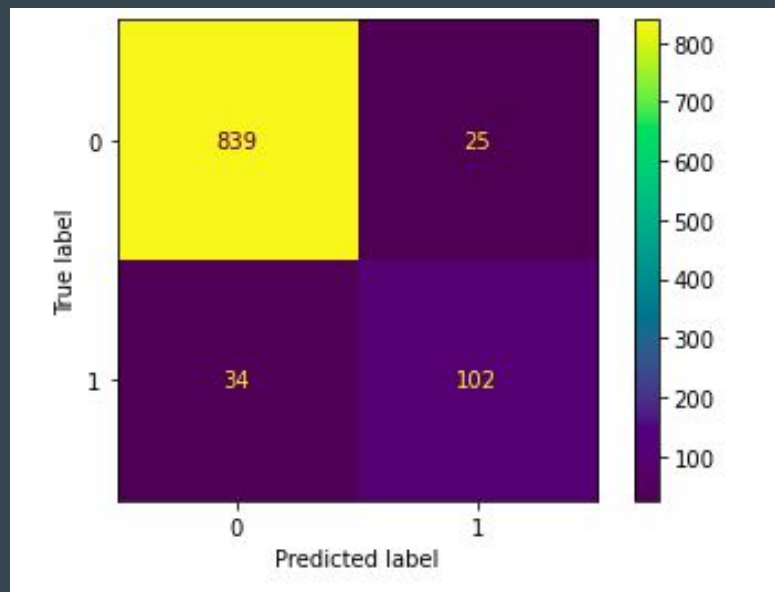
We tuned this model by applying various regularizations and reducing class imbalance in the target feature.

Though with an improved AUC score, this tuned model scored lower than the Decision Tree model.

Final Model

The selected final model was the tuned Decision Tree model.

This model has an accuracy score of 94% meaning that 94 out of 100 times, the model would give an accurate prediction of whether a customer would churn or not.



Evaluation & Recommendations

With a 94% accuracy, this model gives fairly confident predictions of customers likely to churn.

Syriatel will be able to use this model to predict from customer mobile services usage the likelihood of a customer dropping them albeit occasional false positives and false negatives.

Improvements to the model can be made as more data becomes available for training the model.

Next Steps

Integration of the model with Syriatel's system.

Deployment of the model.

Continuous improvements.

Thank You!



Collin Owino

owino.collin@gmail.com

[LinkedIn](#)