



# SYRIATEL TELCOM CHURN PREDICTION

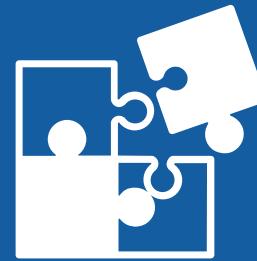
*Presented by: Mercy Jepkogei*



# Project Overview

This project aims to use machine learning algorithms to predict customer churn for Syriatel telecommunication ltd.





# Business understanding

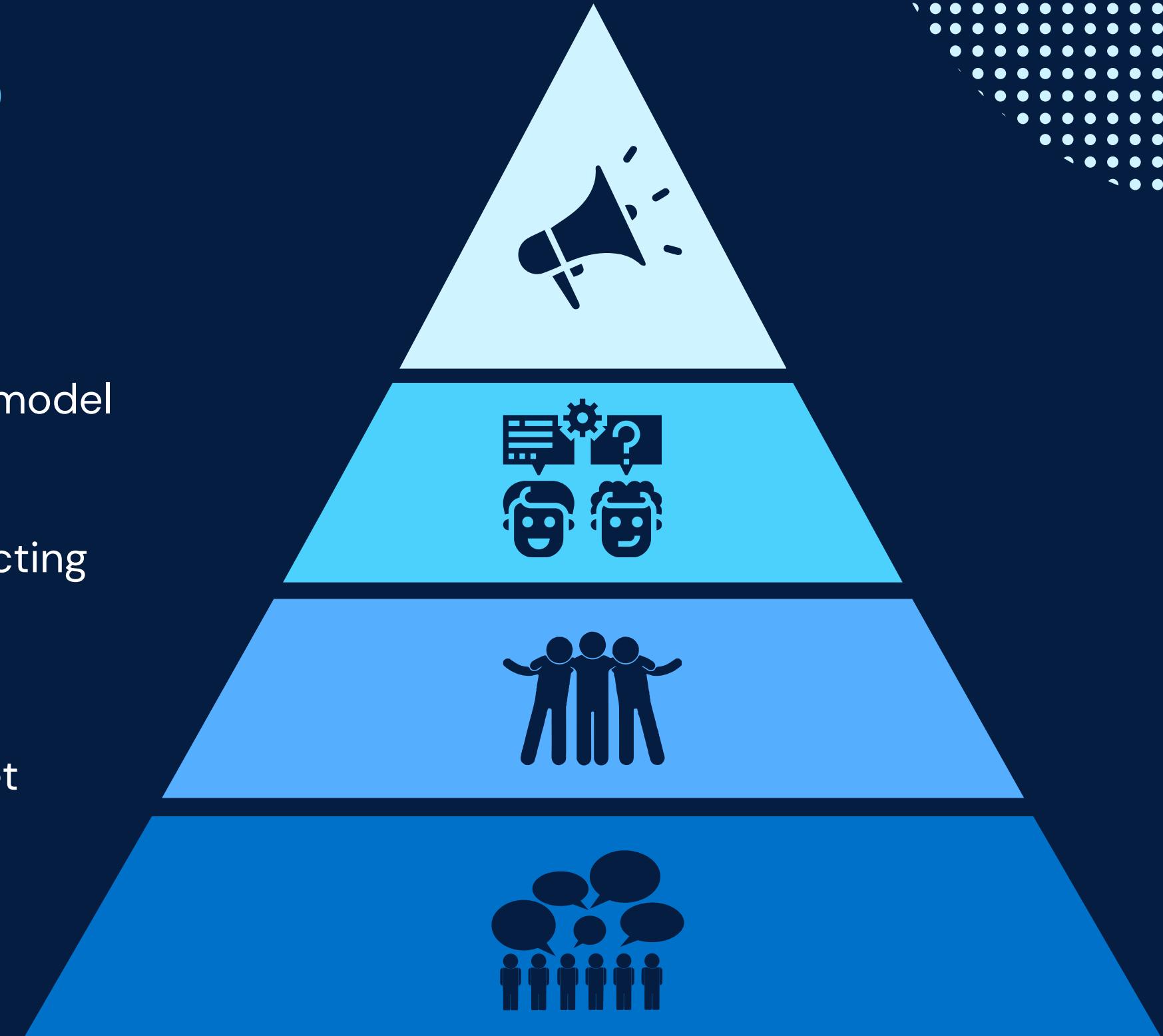


SyriaTel, a leading telecommunications company, urgently needs to address customer churn, as the company has been losing a significant amount of revenue due to customers who do not return. The project aims to develop a binary classification model to predict customer churn.



# PROJECT OBJECTIVES

- 01** To develop the best machine learning model for predicting customer churn.
- 02** To identify important features in predicting churn among customers.
- 03** To evaluate the effect of various data preprocessing techniques on the target outcome.



# DATA UNDERSTANDING

Features

21

Numerical  
features

11

Customer  
records  
3333

Dropped  
featuress

4

categorical  
features  
2

## Target Feature

Churn

Categorical variable -Yes or No

## Independent Variables

'number vmail messages', 'total day minutes',  
'total day calls','total eve minutes', 'total eve  
calls','total night minutes', 'total night calls', 'total  
intl minutes', 'total intl calls', 'customer service  
calls','total charge



# DATA PREPROCESSING

A photograph of a man with dark hair and glasses, wearing a dark long-sleeved shirt and jeans, sitting in a white chair and looking at a green tablet. He is positioned on the left side of the slide, partially obscured by a large blue triangle.

One Hot Encoding the categorical features; 'International plan' and 'voice mail plan'.

Standard scaller for Numerical Features from scikit learn Library.

Label Encoder for the target variable.

Train test Split for splitting data into training and testing sets

Your paragraph text

# MODELLING

Logistic  
Regression

Decision  
Tree  
Classifier

# LOGISTIC REGRESSION RESULTS

Train Data

ROC-AUC

**0.7694**

Accuracy

**76.94%**

Test Data

ROC-AUC

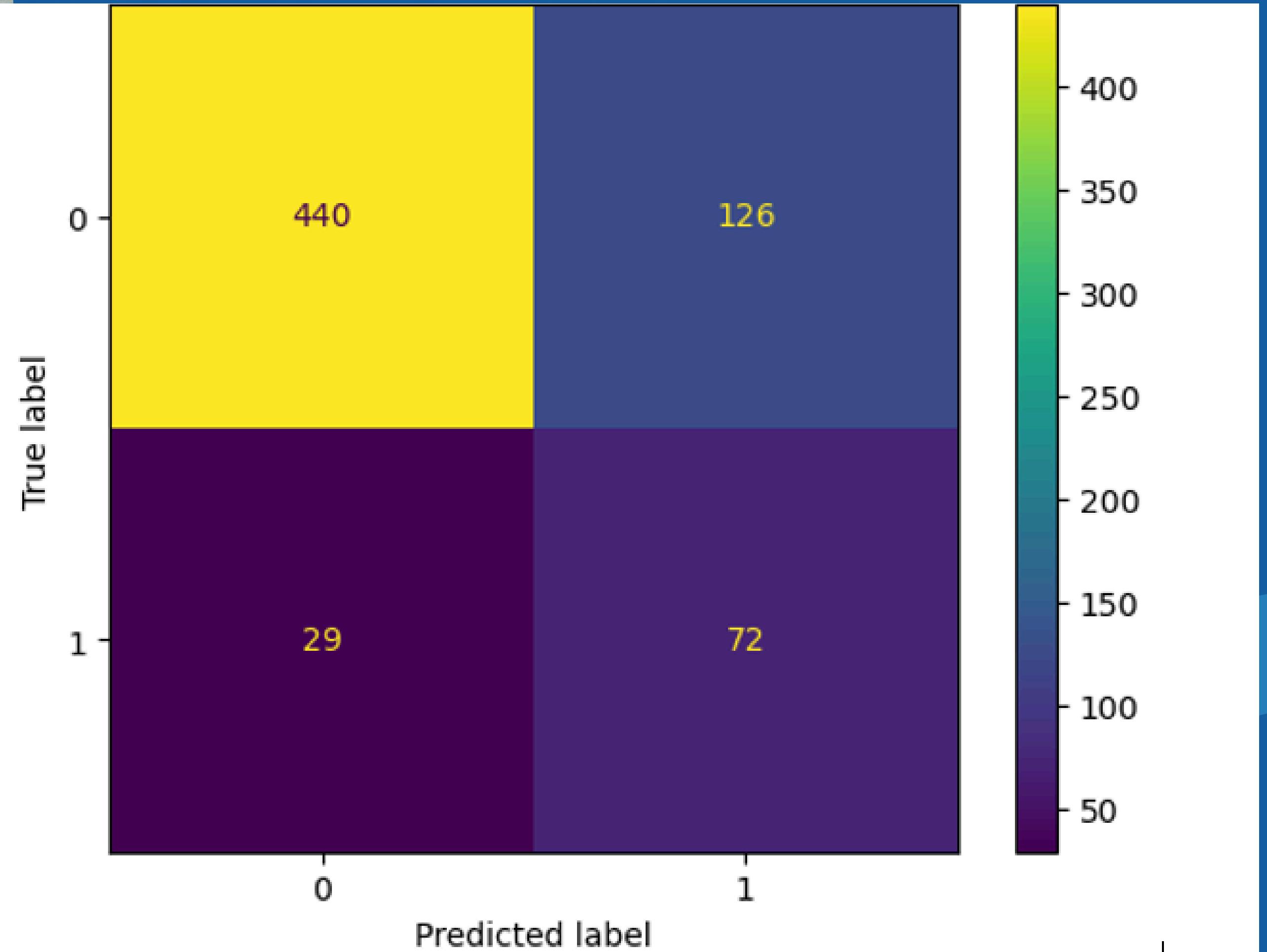
**0.7451**

Accuracy

**76.76%**



# L R CLASSIFICATION MATRIX



# DECISION TREE CLASSIFIER RESULTS

Train Data

ROC-AUC 1.00

Accuracy 1.0

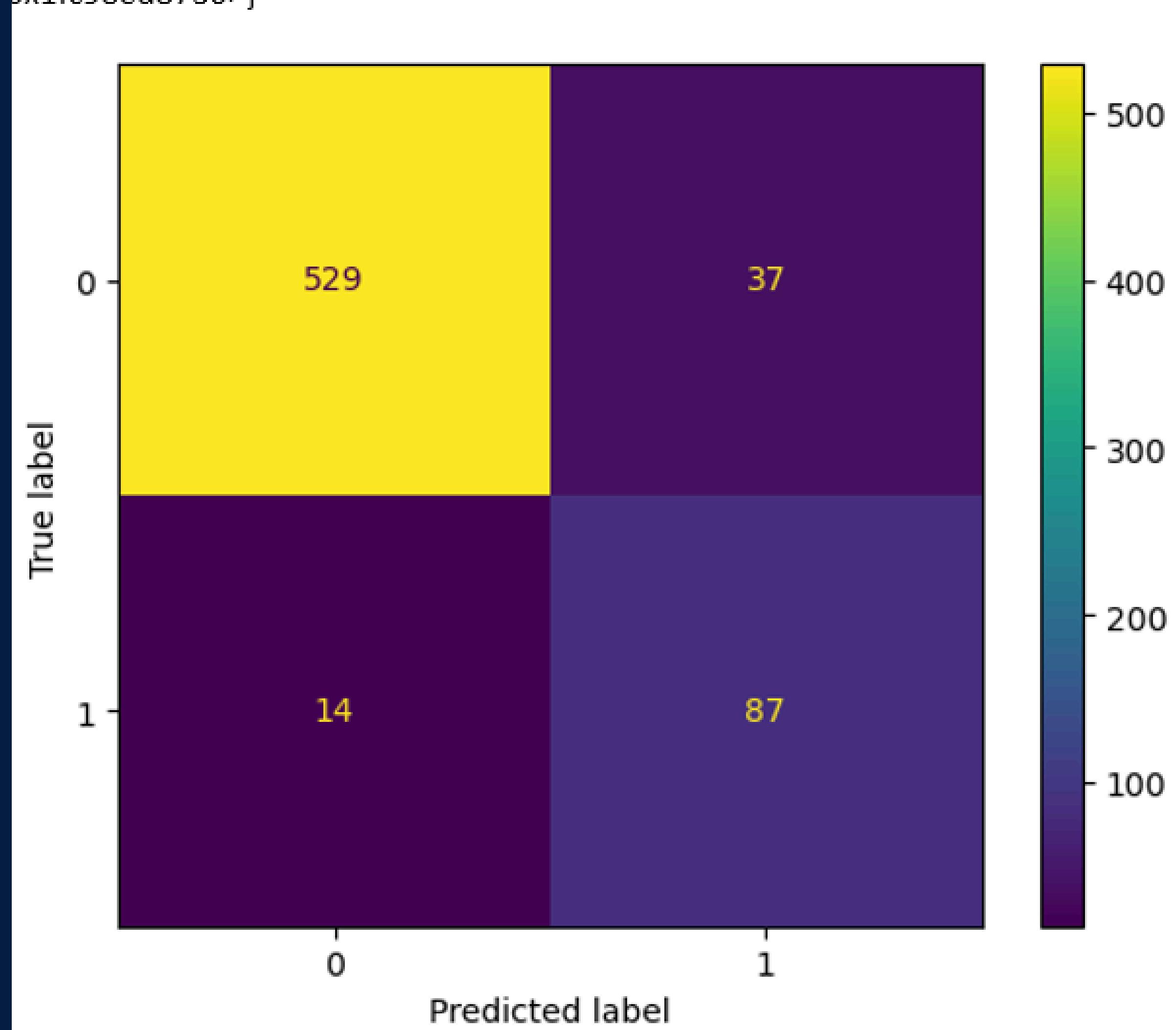
Test Data

ROC-AUC 0.898

Accuracy 92.35%



# Decision Tree Classifier matrix





# MODELLING INTERPRETATION

Logistic regression model exhibits level of discrepancy between the training and testing data.

High Number of 29 false negatives, 126 false positives.

This perfect score on Decision Classifier suggests that the model has learned the training data very well, but it could also be a sign of overfitting.



# DECISION TREE CLASSIFIER TUNING RESULTS

## Train Data

ROC-AUC **0.988**

Accuracy **98.83%**

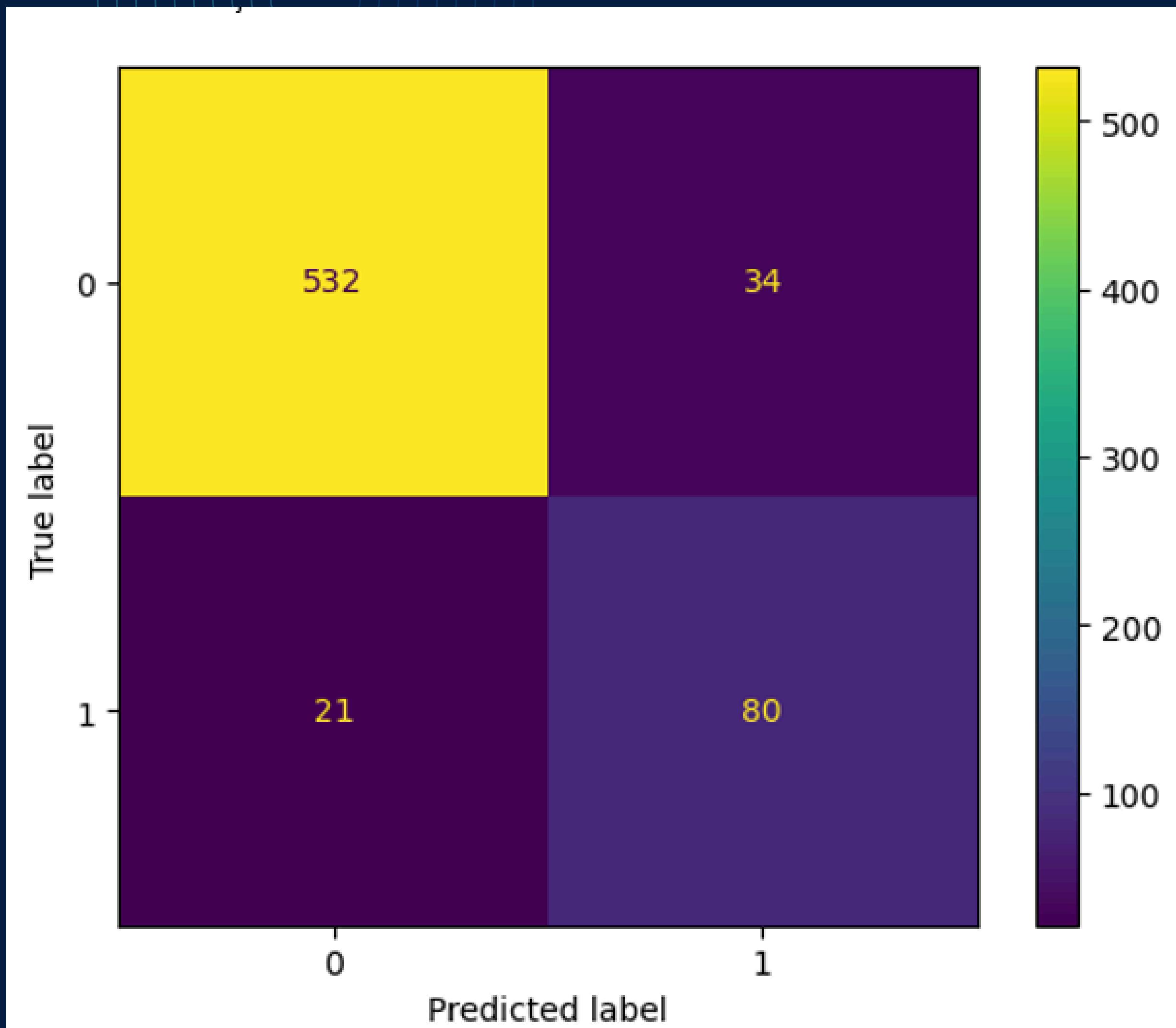
## Test Data

ROC-AUC **0.866**

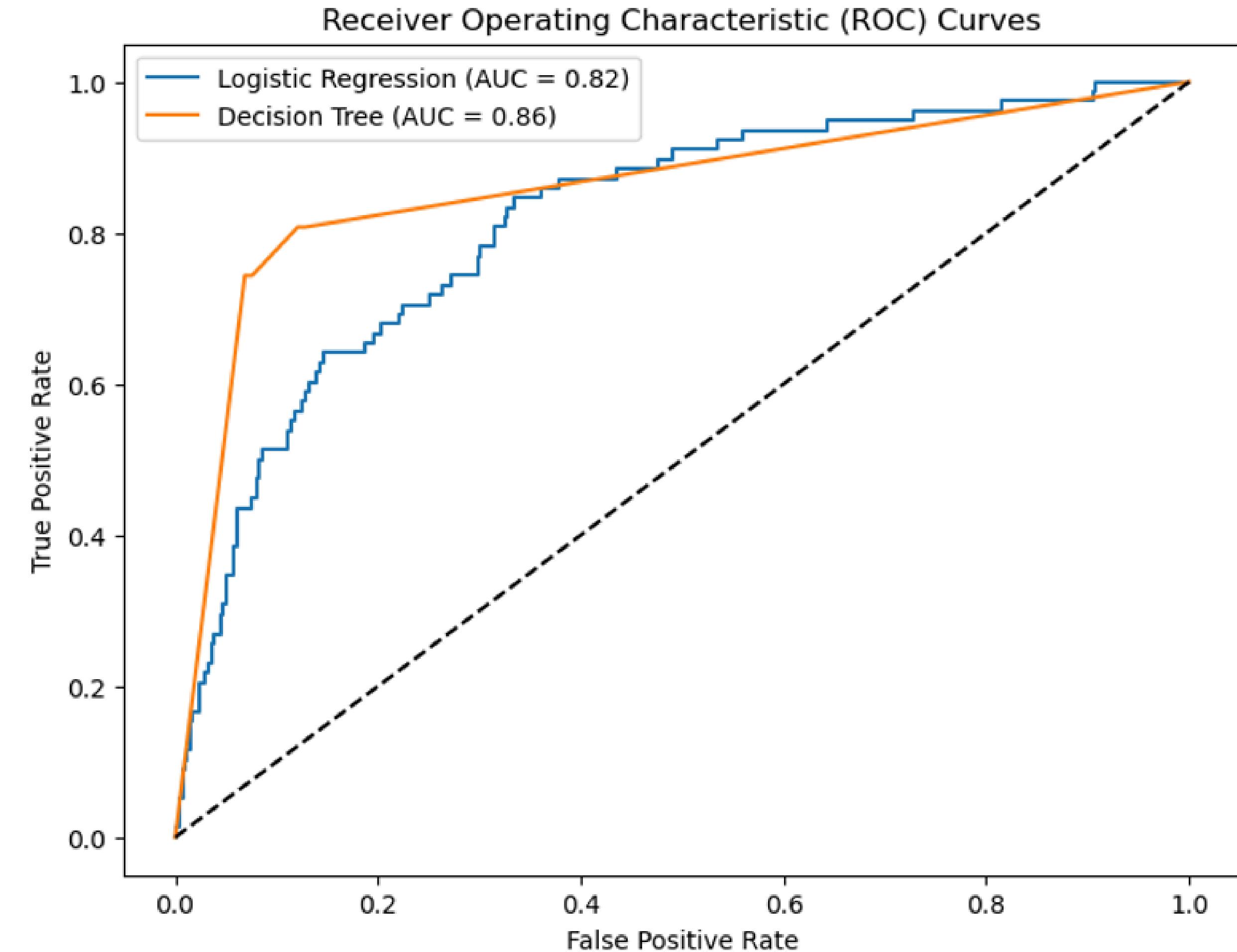
Accuracy **91.75%**



# Decision Tuning matrix



# MODEL EVALUATION





# FINDINGS



The decision tree classifier is a more better model for the Syriatel communication ltd. The decision tree boost of a slightly higher area under the curve, showcasing its superiority in distinguishing positive and negative classes.

The logistic regression on the other hand has an AUC of 0.82, falling slightly lower than the decision tree classifier. It however has a higher number of 126 false positives and 29 false negatives.

# CONCLUSION



1. Decision tree classifier after hyper parameter optimization is the best model strategy for Syriatel communication Ltd. It has the highest accuracy and AUC score, significantly outperforming logistic regression.



2. The analysis shows that 'total charges', in particular the aggregate sum of total day charge, total evening charge, total night charge and total international charge is the most important feature in Syriatel.



3. Effective data preprocessing, such as using SMOTE to address class imbalance and encoding categorical variables, is crucial for improving model performance. This highlights the significance of thorough data preparation in predictive analytics.



# Thank's For Watching



# RESOURCE PAGE

