SyriaTel-Customer-Churn-Prediction-Model

Endalkachew Dessalegne

Introduction

SyriaTel – Telecommunication Company

- Company interest to Identify customers as :
  - Customers who will churn
  - Customers who will stay
- Aim:
  - Reducing money lost due to customers who will churn

Key Business Questions

- which factors contribute most to customers churn?
- are there any predictable patterns in the dataset?
- which model can better predict customers who will churn?
- What corrective action can be taken to reduce number of customers who churn?

Data Overview

- Data sourced from [kaggle] (https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset)
- Over 3,330 records of customers
- Categorical Data of customer information:
- 20 Predictor features : Examples
  - Total call day minutes
  - Customer service calls
  - International plan
- Target feature - Churn :
  - whether customer churned or Stayed

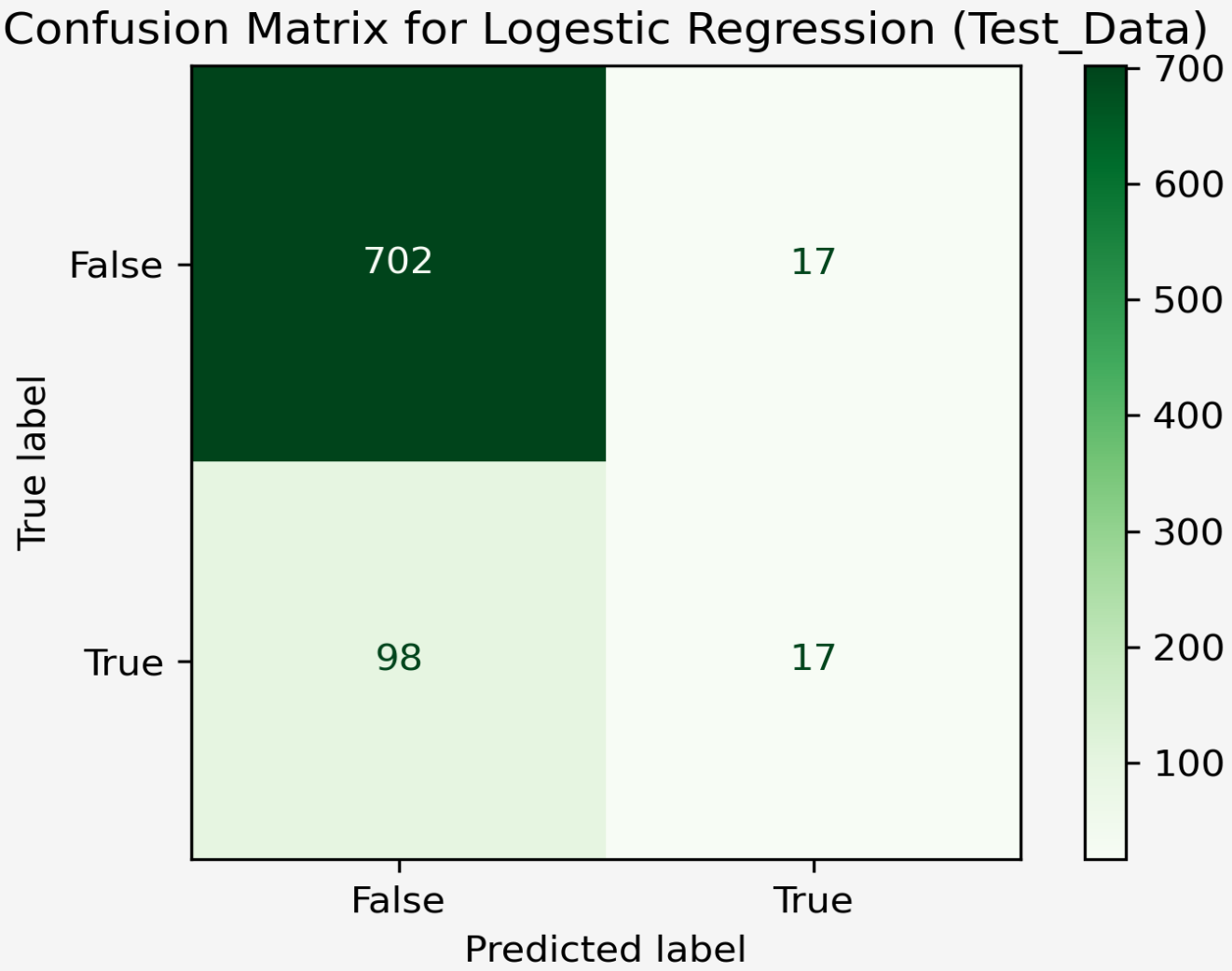
Methodology

- Data preprocessing
- Normalization of data
  - Encoding categorical features
  - Standard Scaler of numerical features
- Modeling
  - Logistic regression
  - Decision Tree
  - Random Forest
- Evaluation of Model
  - Precision score

Findings & Insights

- Predictive Models :
  - ➢ Predict whether a customer will churn or not
  - ➢ Evaluation of Performance of model:
    - Finding model mainly with high Precision
    - Focus on identifying customers who will churn
      - ❖ True – Customer will churn
      - ❖ False – Customer will stay
  - ➢ Precision - correctly predicted positive(TP) instances out of all instances predicted as positive (TP + FP)

Baseline Model: Logistic Regression

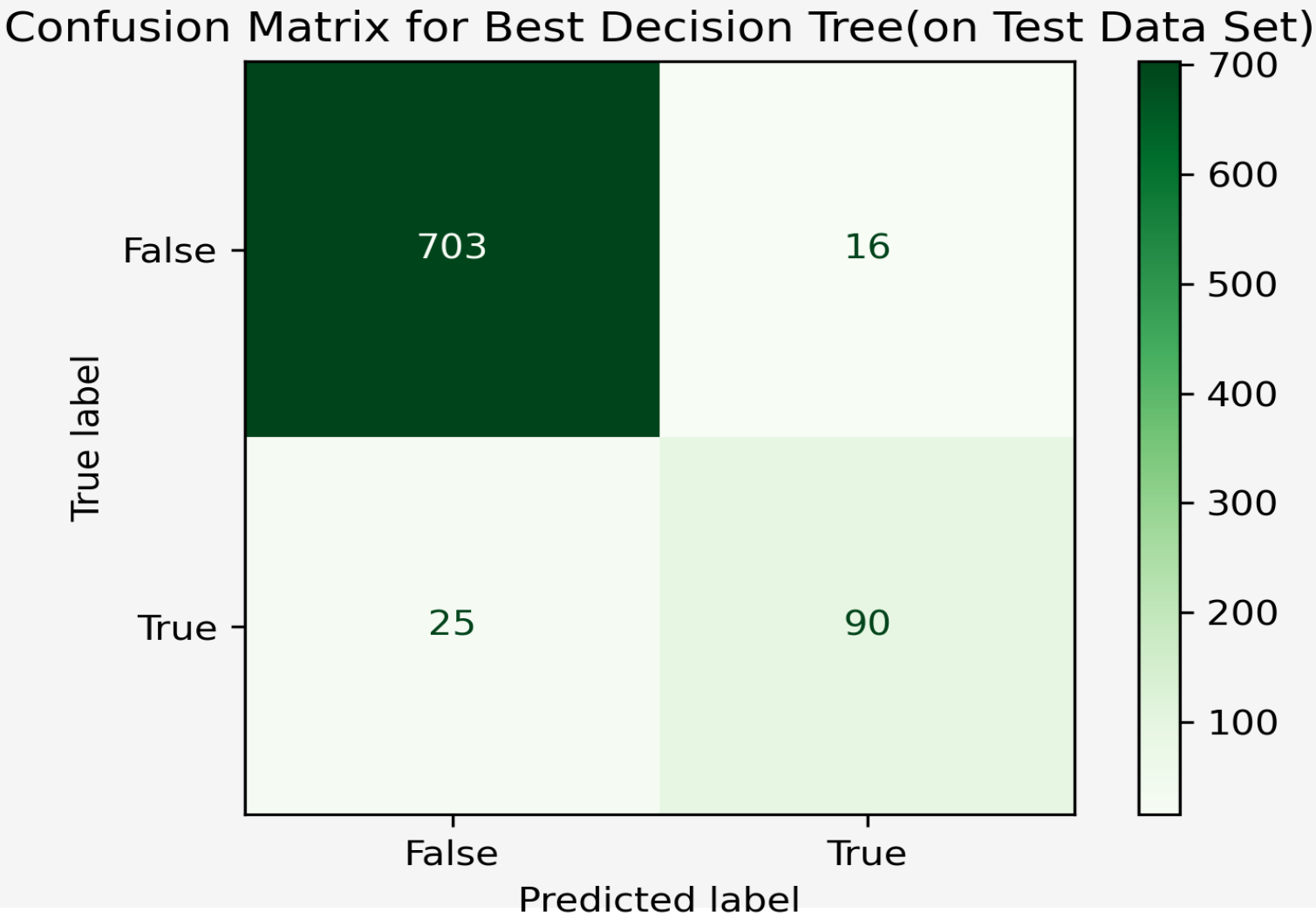


- **TP = 17**
- **FP = 17**

- **Precision** of model = 50 %

- Poor prediction

Optimized Decision Tree Model (After Hyperparameter Tuning)

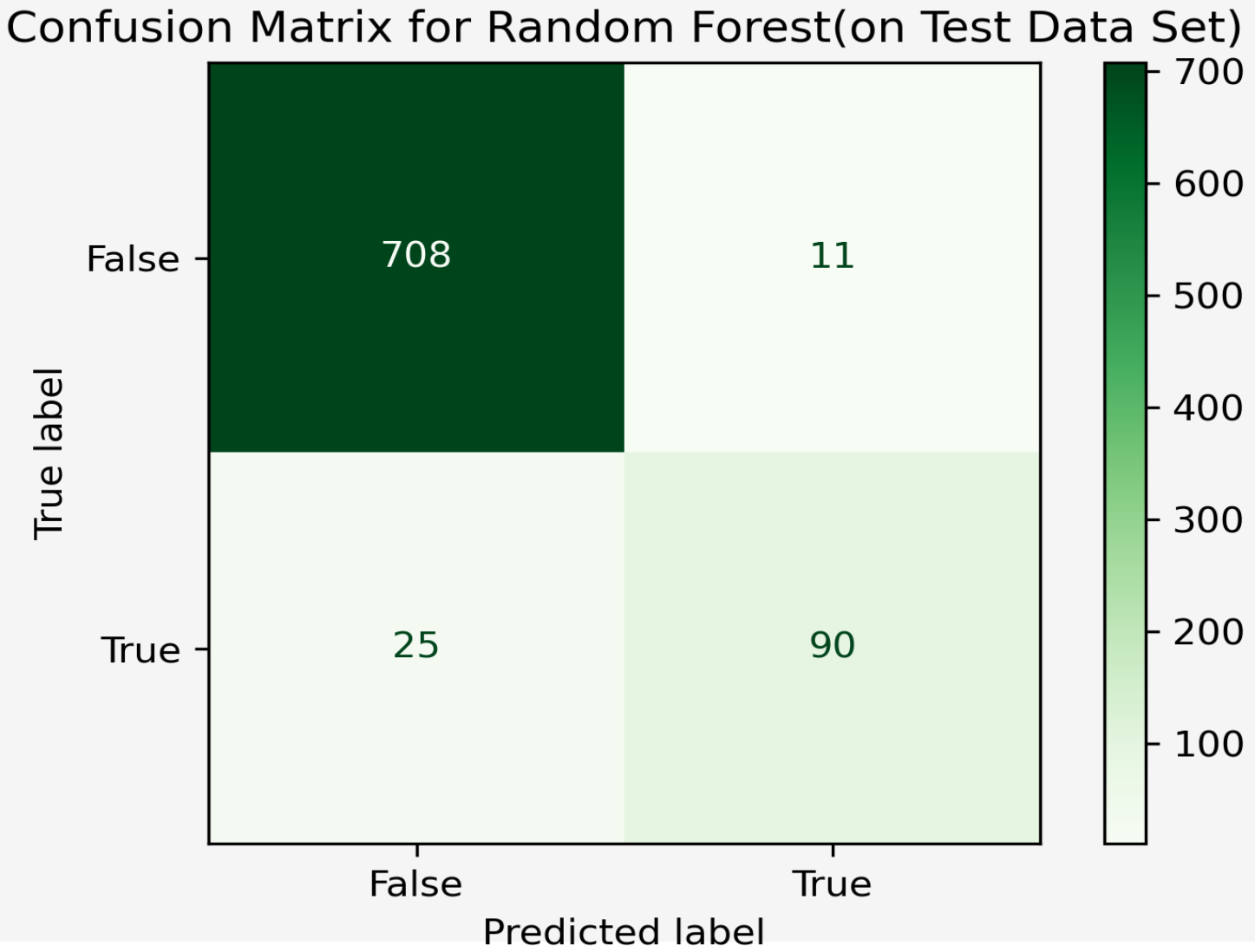


- **TP = 90**
- **FP = 16**

- **Precision** of model = 85 %

- Better prediction

Random Forest Model

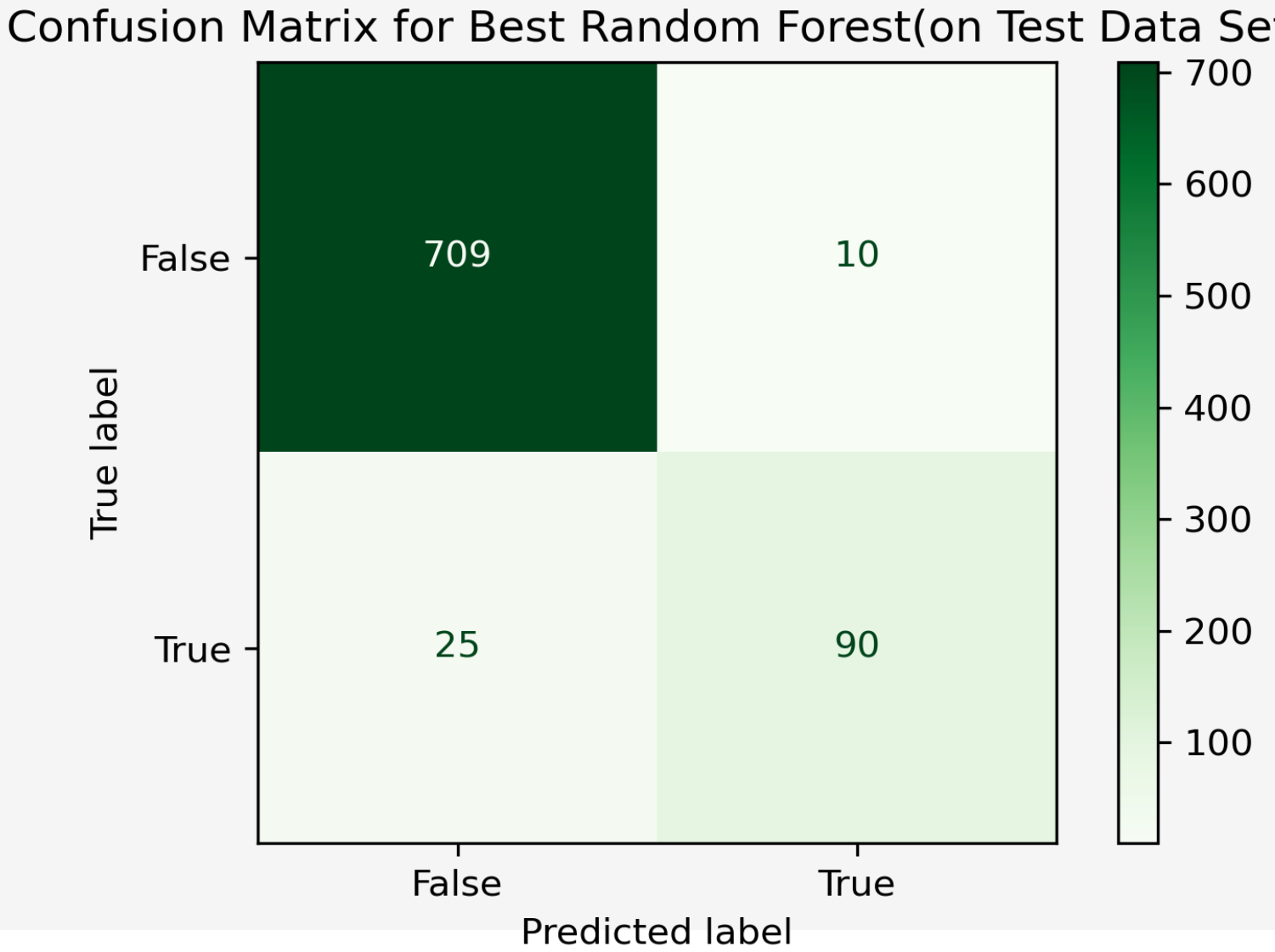


- **TP = 90**
- **FP = 11**

- **Precision** of model = 89 %

- Further improved prediction

Final Model : Random Forest Model (Optimized After Hyperparameter Tuning)

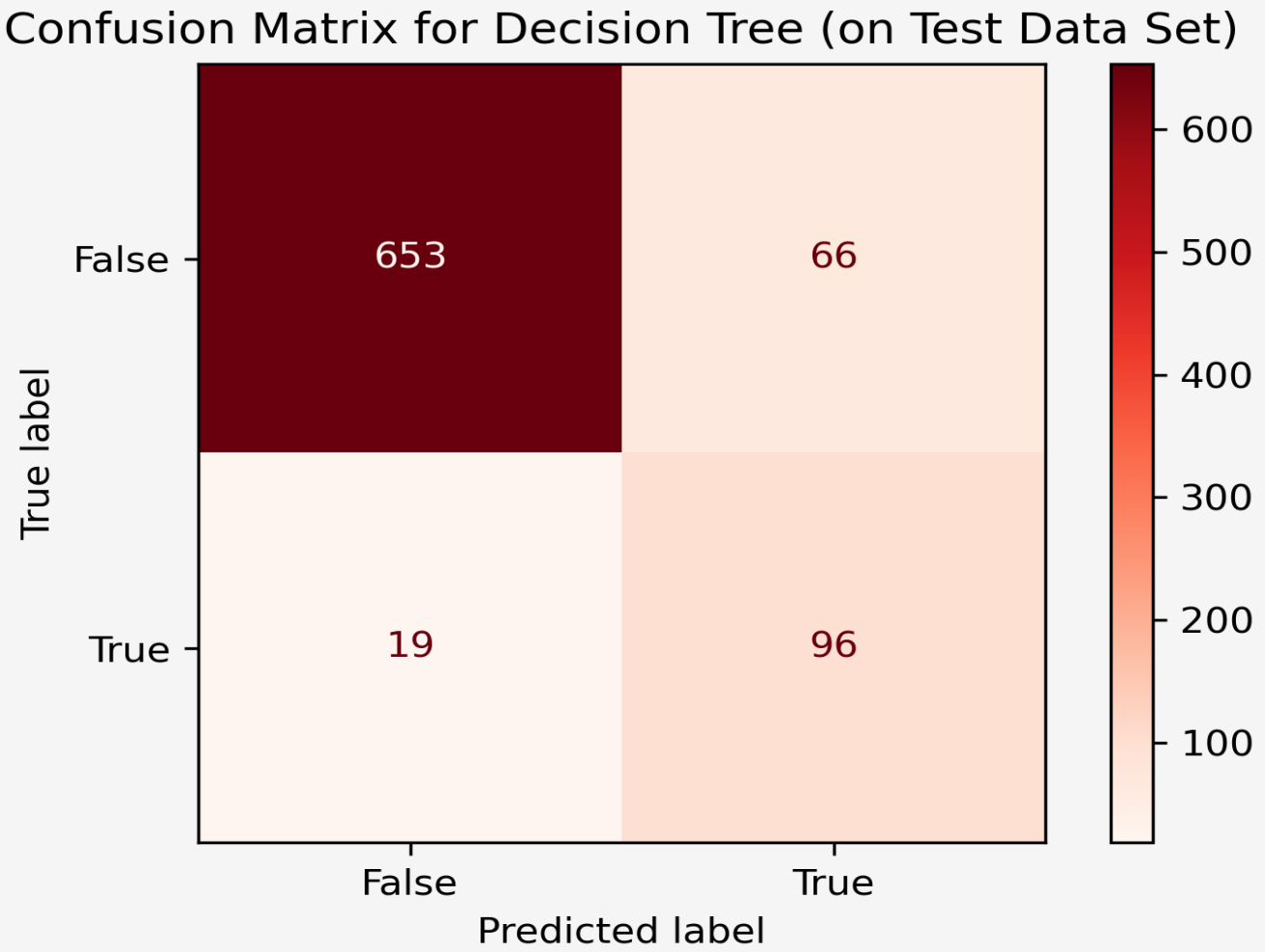


- **TP = 90**
- **FP = 10**

- **Precision** of model = 90 %

- Final precision score attained

Decision Tree Model



- **TP = 96**
- **FP = 66**

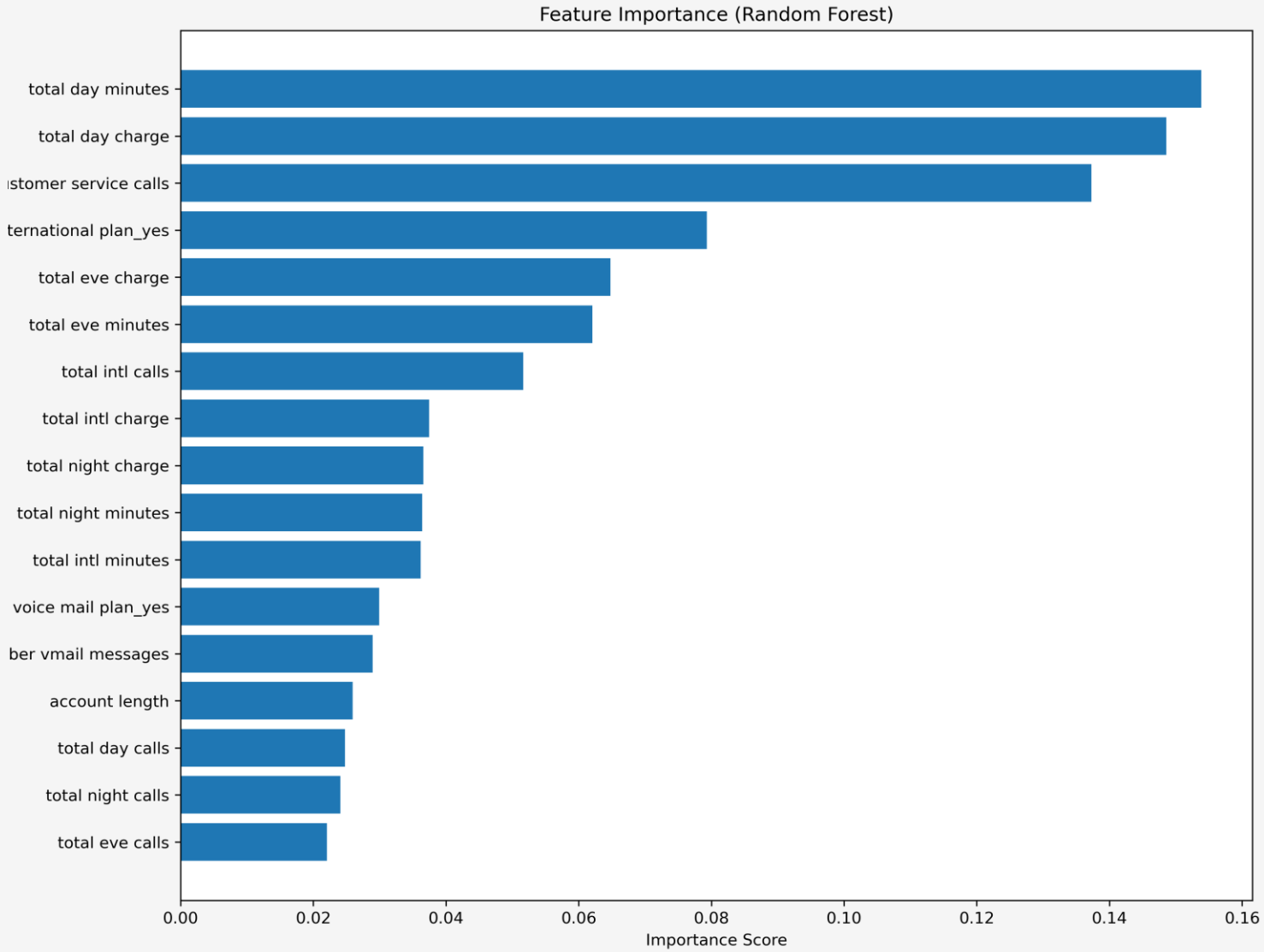
- **Precision** of model = 64 %

- Improved prediction

Summary of Evaluation Metrics of All Model

	Model	Accuracy	Precision	Recall	F1_Score
0	Logestic Regression(On Train Data)	0.853141	0.504274	0.160326	0.243299
1	Logestic Regression(On Test Data)	0.86211	0.5	0.147826	0.228188
2	Logestic Regression (SMOTE_Train_Data)	0.756922	0.751262	0.768184	0.759629
3	Logestic Regression (SMOTE_Test_Data)	0.697842	0.26936	0.695652	0.38835
4	DecisionTreeClassifier(On_Train_Data)	1	1	1	1
5	DecisionTreeClassifier(On_Test_Data)	0.898082	0.592593	0.834783	0.693141
6	Best_DecisionTree(On_Train_Data)	0.964786	0.951613	0.80163	0.870206
7	Best_DecisionTree(On_Test_Data)	0.950839	0.849057	0.782609	0.81448
8	RandoForest(On_Train_Data)	1	1	1	1
9	RandomForest(On_Test_Data)	0.956835	0.891089	0.782609	0.833333
10	Best_RandomForest(On_Train_Data)	0.978792	1	0.855978	0.922401
11	Best_RandomForest(On_Test_Data)	0.958034	0.9	0.782609	0.837209

- Which factors contribute most to customers churn?



- Top 5 features
  - Total day minutes
  - Total day charge
  - Customer service calls
  - International plan
  - Total eve charge

Conclusions

- Random Forest Classifier –is the best performing model with the highest precision
- Precision improved from 50% of the baseline model to 90% precision of RandomForest model
- Precision is the best evaluation metrics for model performance to maximize the correct prediction of customers who will churn
- Some features such as ‘Total day minutes’, ‘Total day charge’, ‘Customer service calls’, ‘International plan’ are more important features for prediction

Recommendations

- Random Forest model best predicts customers who will churn
- To reduce customers who will churn:
  - Offer competitive pricing and bundle plans to retain customers who use more daytime and evening minutes and pay higher charges
  - Frequent 'customer service calls' indicate customer dissatisfaction which may lead to churn . improv service quality and responsiveness
  - Customers with international plans, those who make overseas communication. Offer them packages and competitive rates.

Next steps

- Improve the predictive precision of the model by tuning different parameters.
- Find other models with better precision and accuracy.
- Consider each important feature for customer churn and address it accordingly

Thank You!