Telecom Customer Churn Prediction

FOR: TELECOM COMPANY

PREPARED BY:
EUNICE NDUATI

Overview

The aim of the project is to provide insight and come up with a model that predicts customer churn for Telecom company.

Business Problem

Telecom company is facing this challenge and would like to get more knowledge and understanding on the best way to maintain it's customer base as well as attract new customers on board.

Data Understanding

- Information is extracted from 'Telecom Customer Churn dataset'.
- The data was cleaned by checking and dealing with missing values, duplicates and checking the data types.

Data Understanding

- Some features were analyzed to help understand how they affect customer churn. The features analyzed were:

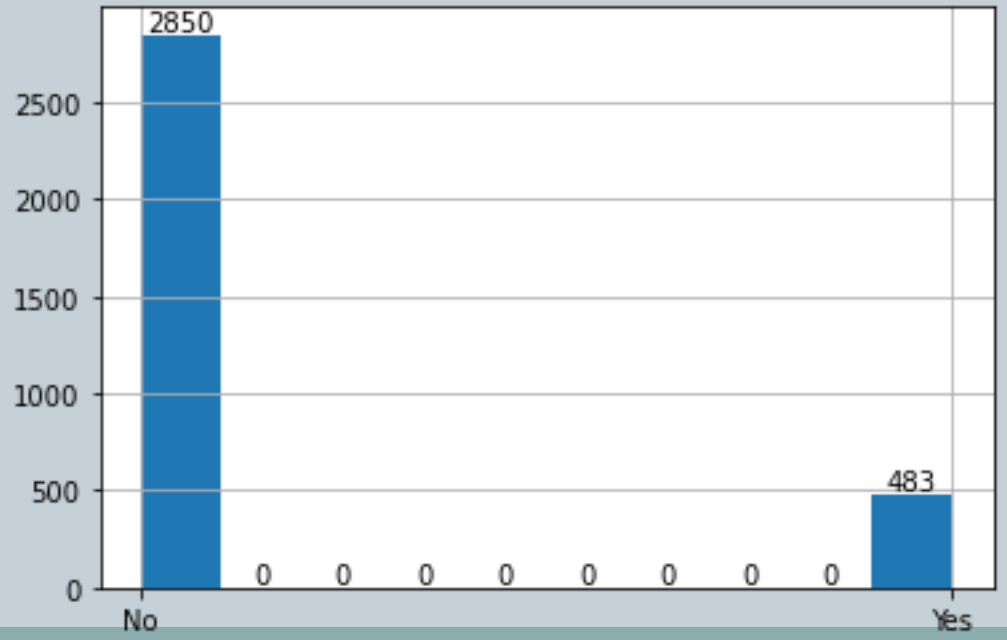
    1. Voice mail plan.
    2. International plan

Correlation Heatmap

- The correlation coefficient between each pair of variables in the data.
- -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 indicates no correlation.

Data Preprocessing

- After examining the class distribution in the target column “churn”. The analysis results show that there is a significant imbalance between the “Not Churn = 0” and “Churn = 1” classes. We use SMOTE to handle class imbalance



Modeling and Prediction

- Three models were used to train and predict data. These are:

1. Logistic regression Model
2. Decision Tree Classifier Model
3. Random Forest Classifier Model

Logistics Regression

- Precision Score: 0.68 Recall Score: 0.68 F1 Score: 0.68 Accuracy: 0.68.
- The Logistic Regression model achieves the lowest F1-score among the three models, with a value of 0.68. It shows relatively lower precision and recall scores, indicating that the model may struggle with accurately identifying positive cases. The accuracy of 0.68 suggests moderate performance.

Decision Tree Classifier

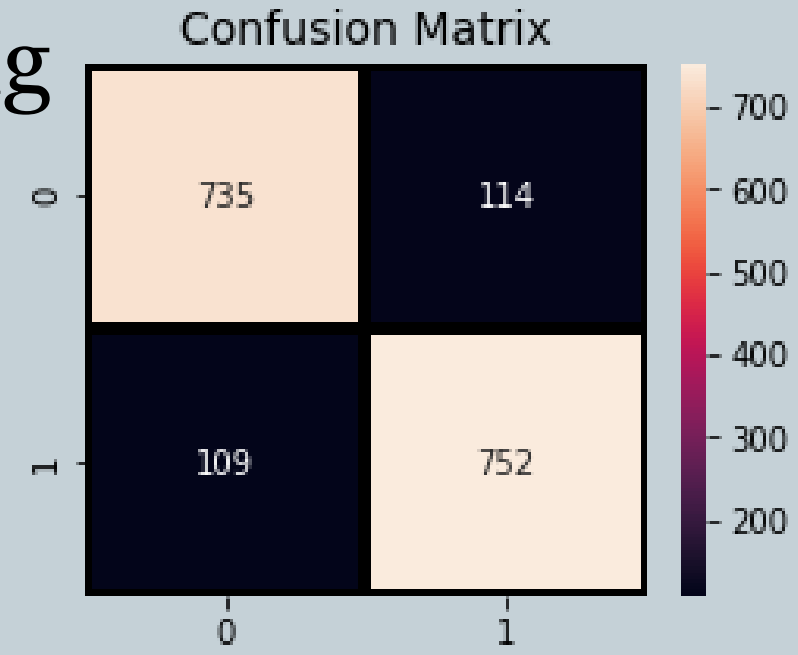
- Precision Score: 0.87 Recall Score: 0.83 F1 Score: 0.85 Accuracy: 0.85
- The Decision Tree Classifier model demonstrates a slightly lower F1-score of 0.85 compared to Random Forest Classifier.
- It achieves a relatively lower precision score, indicating that it has a lower probability of correctly identifying positive cases. However, the recall score suggests that it may miss some positive cases. The accuracy of 0.85 suggests good overall performance.

Random Forest Classifier

- Precision Score: 0.90 Recall Score: 0.95 F1 Score: 0.92 Accuracy: 0.92.
- The Random Forest Classifier Model achieves a decent F1-score of 0.92. It shows a reasonable balance between precision and recall, indicating that the model can make accurate predictions while capturing a good proportion of positive cases. The overall accuracy of 0.92 suggests that the model performs well in classifying churn.

Confusion Matrix

- Confusion matrix is a table that describes the performance of the model on test data, comparing the predicted results with the true values.



Confusion Matrix Results

- There are total 758+91= 849 actual non-churn values and the algorithm predicts 758 of them as non churn and 91 of them as churn. While there are 85+776= 861 actual churn values and the algorithm predicts 85 of them as non churn values and 776 of them as churn values.

Conclusion

- Upon reviewing the results, the top performing models in ranking from highest to lowest based on the provided metrics are:

1. Random Forest Classifier Model.
2. Decision Tree Classifier Model
3. Logistic Regression Model

Recommendation

- 1. Use Random Forest Classifier for future prediction.
- 2. Encourage more customers to subscribe to voicemail plan.
- 3. Consider coverage outside the country to retain temporal customers from abroad.


THANK YOU!