Customer Churn Prediction

PATRICK MAINA

Agenda

- ❖ Overview
- ❖ Project workflow
- ❖ Business Understanding
- ❖ Analysis
- ❖ Modeling
- ❖ Evaluation
- ❖ Conclusion
- ❖ Recommendations

Overview

- One of the major challenges facing telecommunication providers is customer churn – a scenario where users discontinue their service, mostly due to dissatisfaction from the provider.

- In order to mitigate this challenge, telecom companies are exploring churn prediction mechanisms, as well as the need to understand factors that contribute to customer churn.

Project Workflow

Recommendations

Conclusion

Evaluation

Business
Understanding

Data
Understanding

Modeling

01 02 03 04 05 06

Business Understanding

Business Problem

SyriaTel, a leading telecom provider, is experiencing a significant loss of customers who are choosing to leave its services for other competitors.

Project Objectives

- To determine the key characteristics and behavior patterns that contribute to customer churn.

- To build a robust predictive model that will identify customers with a high likelihood of discontinuing their service, with a recall of about 80%.

- To provide data-driven insights and recommendations that will proactively engage, and retain high-risk customers.

Data Understanding

- The dataset is a customer churn dataset from the telecom sector, obtained from Kaggle.

- It contains about 3333 records, and 21 columns, of which 4 columns are categorical and 17 columns are numerical.

- Some of the key features of the dataset include: **Call rates, State and Area Code, International and Voice Mail Plans, and Customer Service Calls.**
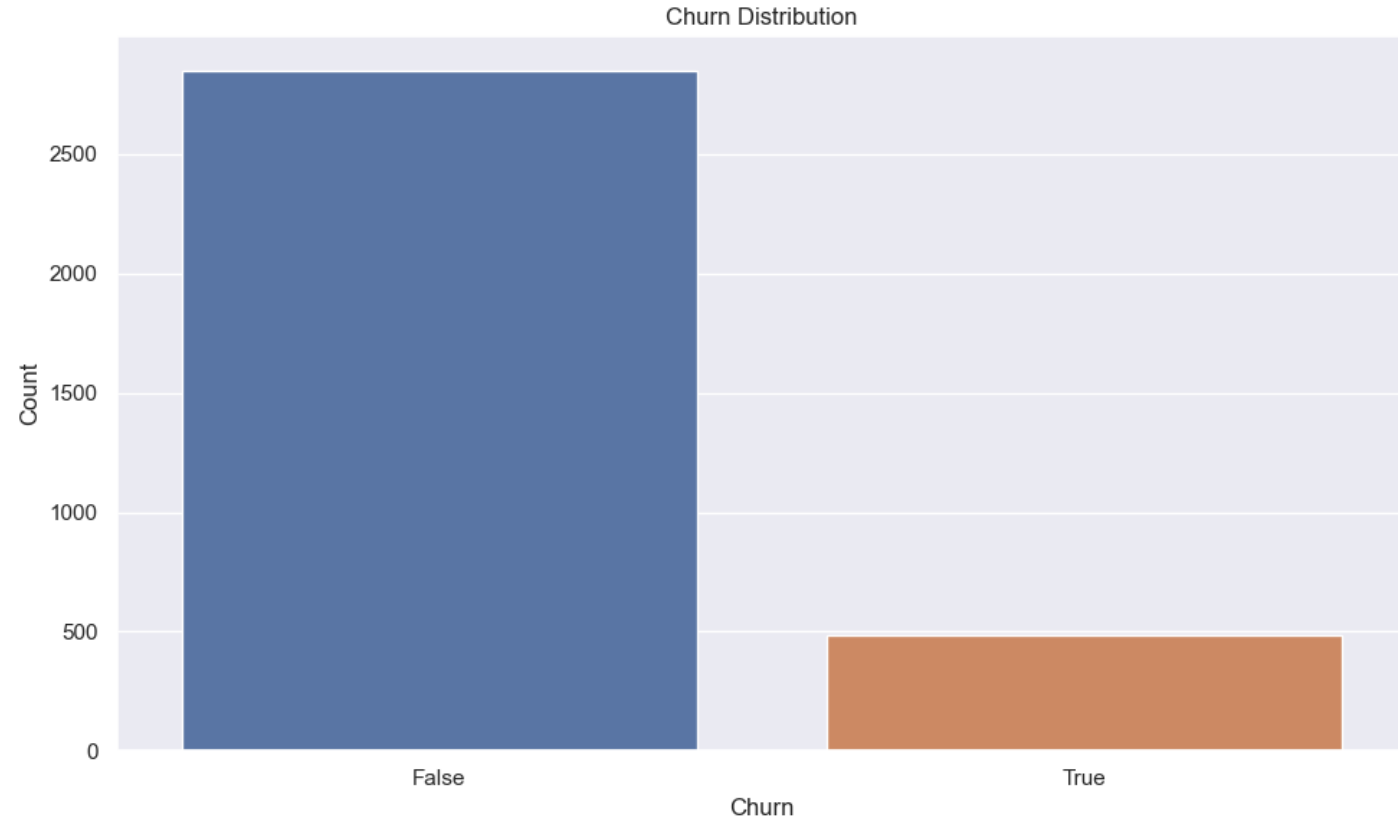
Analysis

Total number of customers in the dataset; *3333*

Number of customers who churned from the company; *483*
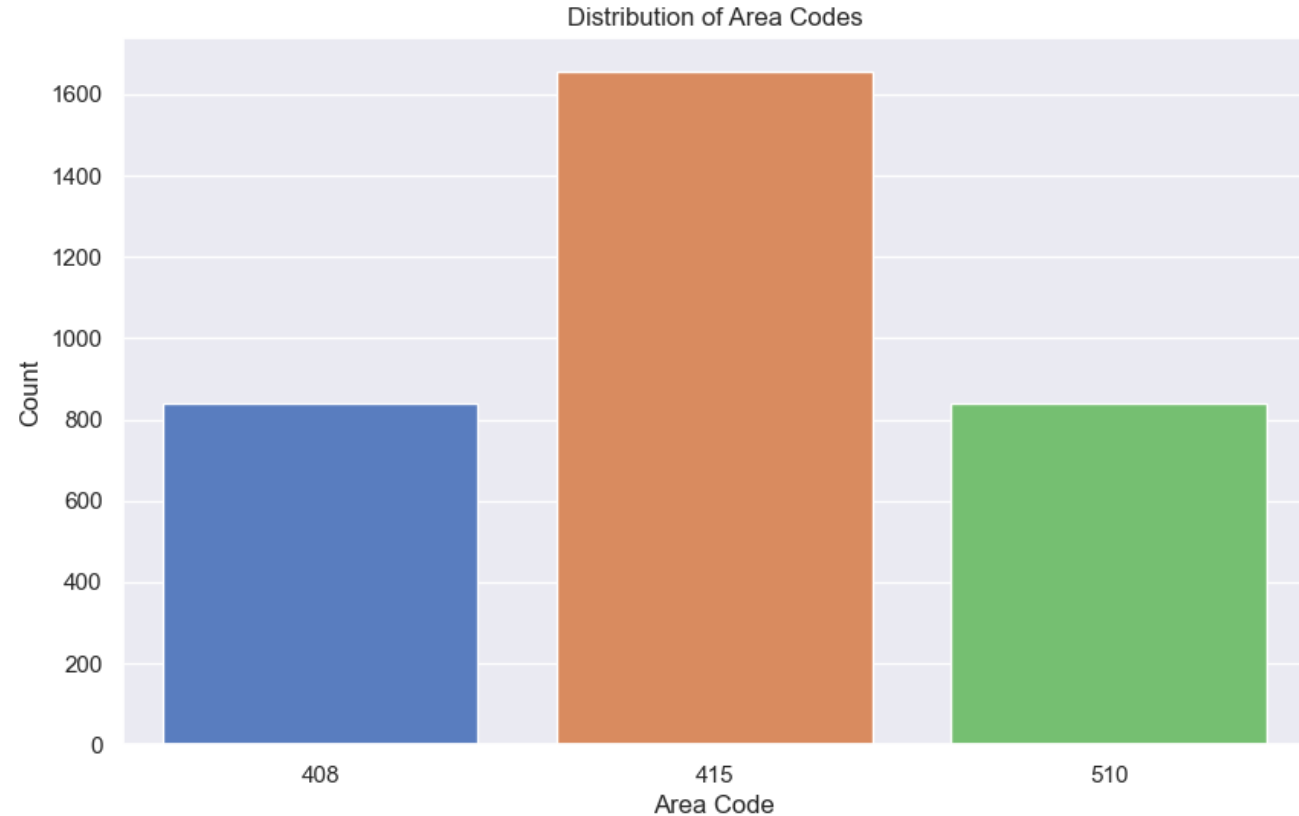
Percentage of churned customers; *14.5%*

Distribution of the Churn target

- Area code 415 has a higher number of customers with about *1655 customers*, which accounts for about 49.7%.

- Area codes 408 and 510 have a close number of customers:

Area 408: *838 customers*

Area 510: *840 customers*

Distribution of Area Codes

Distribution of customers by Area Code

- Customers who churn (churn=True) tend to have more customer service calls than those who don't.

- Additionally, majority of the customers who churn come from area codes 415 and 510
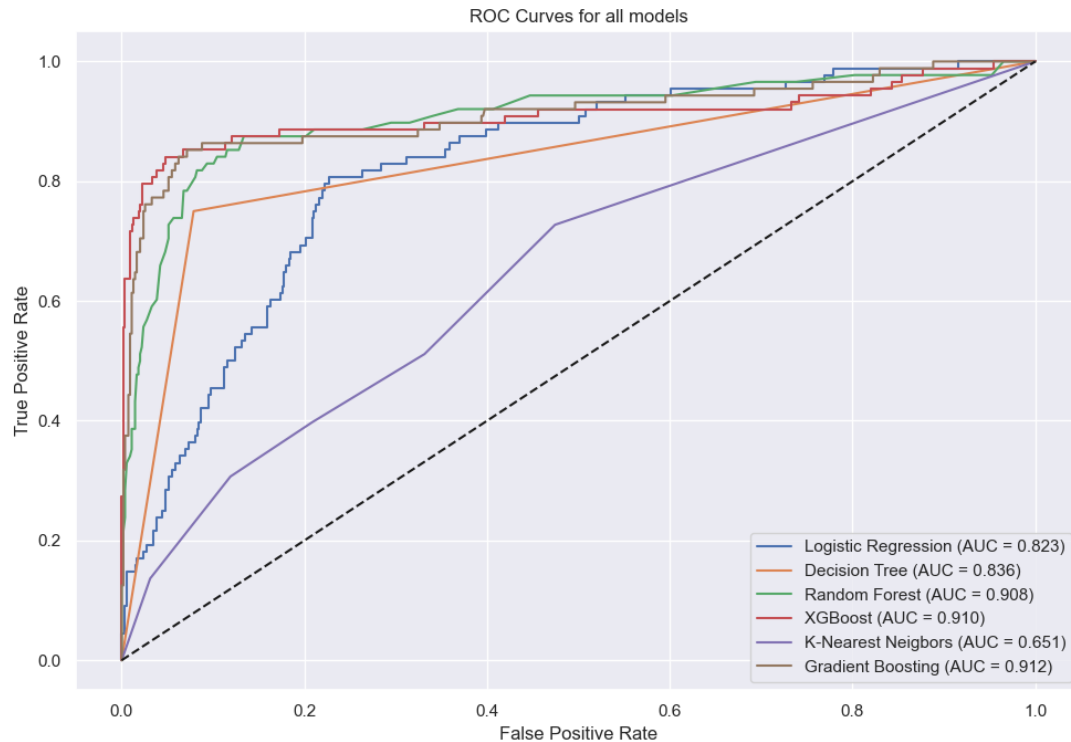

Customer Service Calls vs Churn

Relationship between customer service calls and the churn rate, by area code

Modeling

Algorithms used:

- Logistic Regression

- Decision Tree

- Random Forest

- XGBoost

- K-Nearest Neighbor

- Gradient Boosting

Evaluation

ROC Curves for all models

Top 3 models Based on Recall Score:

- Gradient Boosting - 0.807

- XGBoost – 0.795

- Decision Tree – 0.75

Top 3 models Based on ROC_AUC curve:

- Gradient Boosting – 0.912

- XGBoost – 0.910

- Random Forest – 0.908

Model Tuning Results

- The tuned Gradient Boosting model achieved a recall score of 0.81, which was similar to the untuned model score, and an AUC score of 0.921, which was an improvement from the AUC score of the untuned model.

- The tuned XGBoost model achieved a recall score of 0.82, which was a significant improvement from the recall score of the untuned model, and an AUC score of 0.911, which was a slight improvement from the AUC score of the untuned model.

Conclusion

- The recall score of the XGBoost model was 0.82, which surpassed our initial target of 0.8.

- From my analysis, the most important features that contributed significantly to the model performance are ***Customer Service calls, Total Day Charge and International Plan.***

- We achieved our objectives of identifying the key attributes that contribute to customer churn, and developing a customer churn prediction model with a recall score of 0.8 and above.

Recommendations

- Offer specialized discounts, loyalty rewards, or exclusive promotions in high-churn area codes such as 415, which will serve as an effective incentive to retain customers.

- Investing in comprehensive training programs for support staff, and implementing better issue/conflict resolution frameworks in order to enhance customer satisfaction, thus minimizing the rate of churn.

- Developing localized market efforts, personalized engagement strategies, and enhanced customer support in high-churn areas such as Texas and New Jersey, in order to strengthen customer loyalty and retention.

Thank you

Patrick Maina

Patrickmaina364@gmail.com

LinkedIn profile

GitHub Profile

Personal Website