SyriaTel Churn Prediction

Using Data to Understand and Reduce Churn

Overview

- **Objective**: To predict which customers are likely to churn
- **Goal**: Assist the business in proactively reducing churn by targeting at-risk customers
- **Approach**: Data analysis as well as machine learning to uncover key churn indicators

Business Understanding

Churn is a one of the biggest problems in the telecom industry.For Telecom companies it is important to attract new customers and at the same time prevent old ones from leaving their companies. There are different reasons why customers may choose to terminate their contracts, including pricing, poor customer service, change in customer preference, among others.

Due to the direct effect of churn on revenues, telecom companies apply machine learning models to predict churn on an individual customer basis and take countermeasures such as discounts, special offers, or other measures to keep their customers. Finding factors that contribute to customer churn is important so as to take the necessary actions in order to reduce this churn.

Data Understanding

The dataset contains data on the customers of a Telecom company. Each row represents a customer and the columns contain customer’s attributes.

There are over 3000 customer records analyzed and 21 different features including:

- Customer service calls
- Total charges
- Voicemail & international plans
- Churn (target)

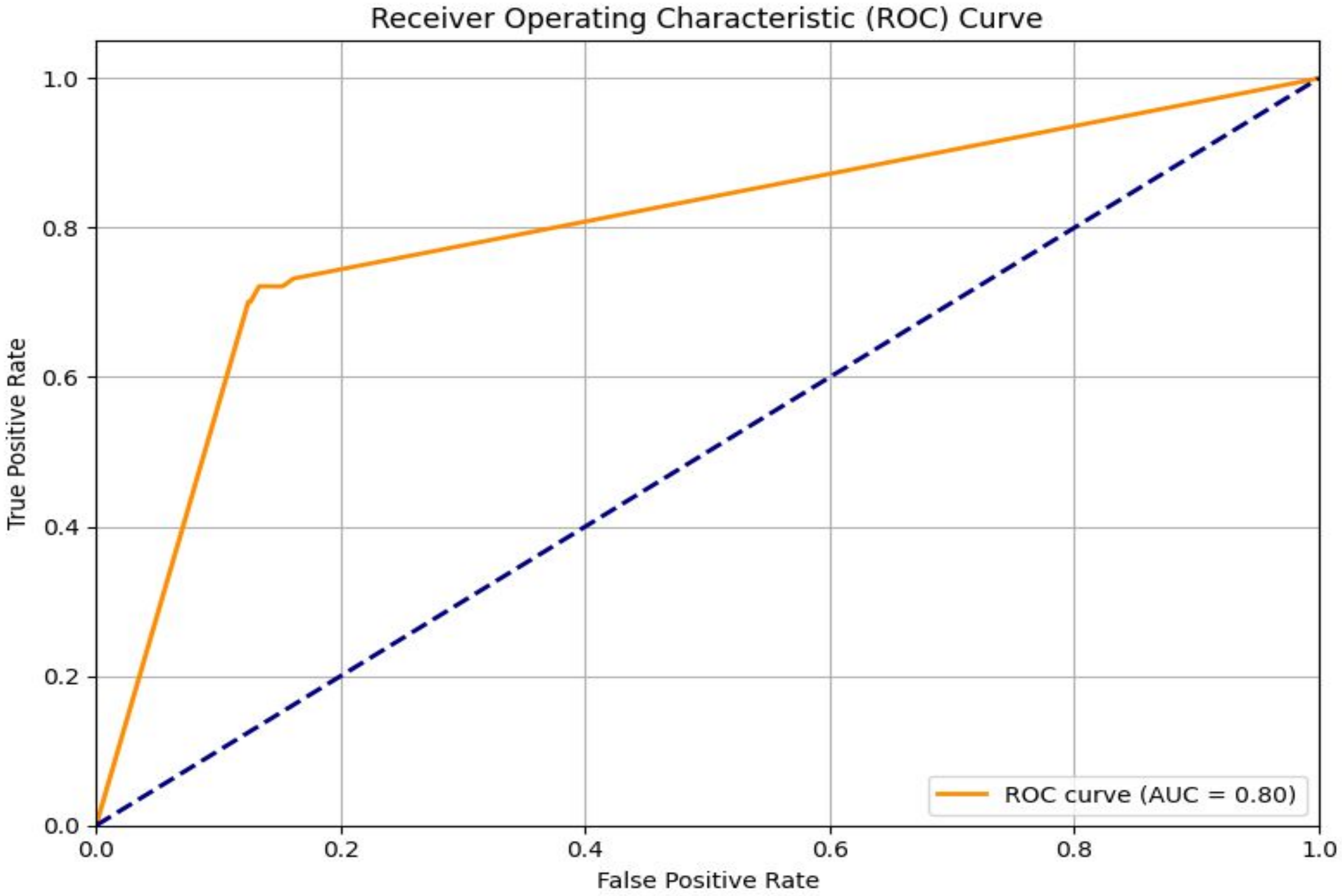
Our Modelling Approach

We decided to use multiple models, a logistic regression model, its hyperparameter-tuned version as well as a decision tree and its hyperparameter-tuned version.

The logistic regression model serves as a baseline for comparison.

We compared them using precision, recall, and AUC (Area Under Curve).

After comparison we found that the hyperparameter-tuned decision tree model performed the best especially the recall and the F1 score.


Receiver Operating Characteristic (ROC) Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (AUC = 0.80)

Receiver Operating Characteristic Curve and Area Under the Curve

The curve above represents the ROC AUC Curve for the Decision tree model.

It basically summarizes the performance of the model into a single number.

The closer the number is to 1.0 the better the model performance.

The model has an AUC of 0.80. This means that the model can distinguish between churners and non-churners 80% of the time.

This is a decent performance and suggests that the model is quite effective at identifying potential churners.

Decision Tree Evaluation

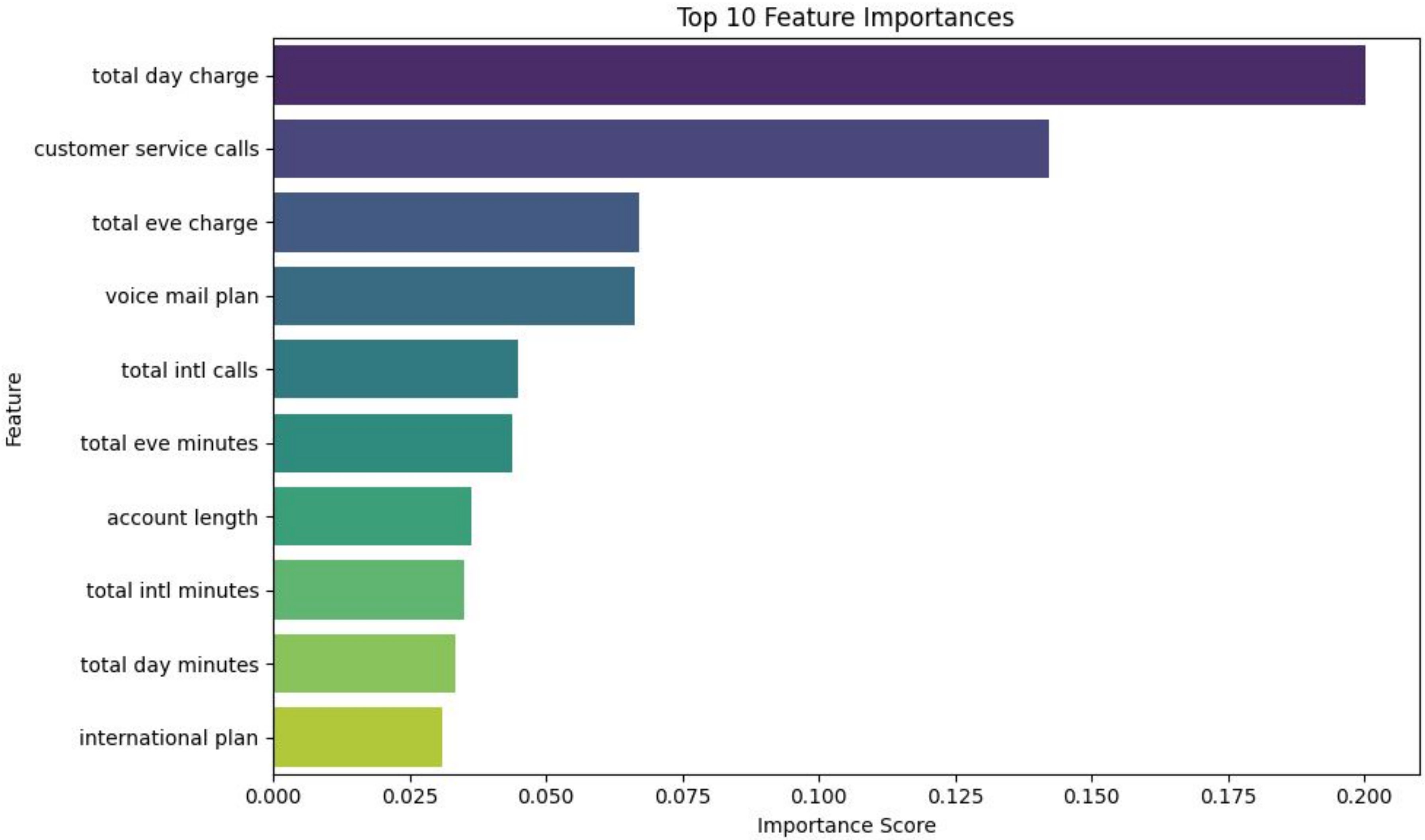
The Recall for churn is 72% - This represents how many churners we correctly flagged

Nearly 3 out of 4 actual churners correctly identified

ROC AUC: 0.80 shows a strong separation between churn and non-churn

Our model reliably identifies customers likely to churn

We can now put our energy into retaining the at risk customers before they leave



Top 10 Feature Importances
total day charge
customer service calls
total eve charge
voice mail plan
total intl calls
total eve minutes
account length
total intl minutes
total day minutes
international plan
Feature
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
Importance Score

Feature Importance

The top most influential features are:

- total day charge
- customer service calls
- total eve charge
- voicemail plan

Recommendations

- Offer loyalty discounts, tiered plans, or flat-rate unlimited day calling may reduce pressure from pricing.
- Consider campaigns that educate customers on optimizing their usage to lower perceived charges.
- Take note of users with multiple service calls in a short window as high risk.
- Track the efficiency of customer service resolution and improve on it.
- Create time-based promotions, for example, unlimited evening or weekend calling to reduce the perceived cost.
- Add alerts or reminders for nearing plan limits or high evening usage to reduce the feeling of surprise bills on customers.

Next Steps

For even better model performances:

Try more advanced models like XGBoost and Random Forest.

Use more recent data, or data that has more features to manipulate.

Build a real-time churn alert system.

Conclusion

The models built show encouraging results in churn prediction, especially when referring to decision trees with tuned hyperparameters. With proper deployment and thorough monitoring, the system could significantly improve customer retention efforts as they can zero in at-risk customers early.

By Kago Ruburu