




SyriaTel Customer Churn



Jessica Rodriguez
May 2022
Flatiron: Self-Paced



Overview

The goal of this project was to build a binary classifier to predict whether a customer would "soon" leave the telecommunications company, SyriaTel.

Using the provided dataset, I discovered that the current "churn" rate for SyriaTel is about 14%.

The following questions were addressed:

- * What features are the primary determinants of customer "churn"?
- * Are there any predictable patterns?
- * How can SyriaTel use these findings to implement cost-effective solutions?

Determining the Best Metric: Type I and Type II Errors

SyriaTel would save the most money by prioritizing the retention of current customers over the acquisition of new customers.

In statistics, a type I error is a false-positive result, meaning that a null hypothesis is rejected when it is actually true. A type II error is a false-negative result, meaning that a null hypothesis is not rejected when it is actually false.

For this project, incorrectly classifying a false-negative (type II error) would be worse than incorrectly classifying a false-positive. **A false negative would mean that the reality of a customer canceling would have been overlooked.**

My goal was to **build a classifier that minimizes false negatives.**

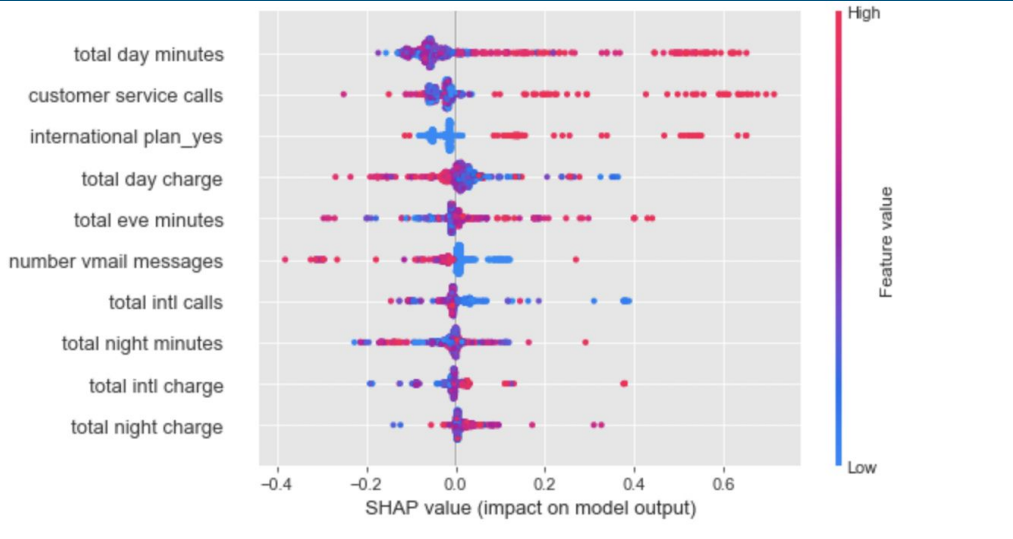
Building a Binary Classifier

After building and evaluating different machine learning models, I ultimately decided to use my decision tree classifier for further analyses.

The decision tree had an acceptable recall score of 74%. In binary classification, “recall” is the ratio of true positives over the number of false negatives.

Recall took priority over precision because my ultimate goal was to minimize false negatives.

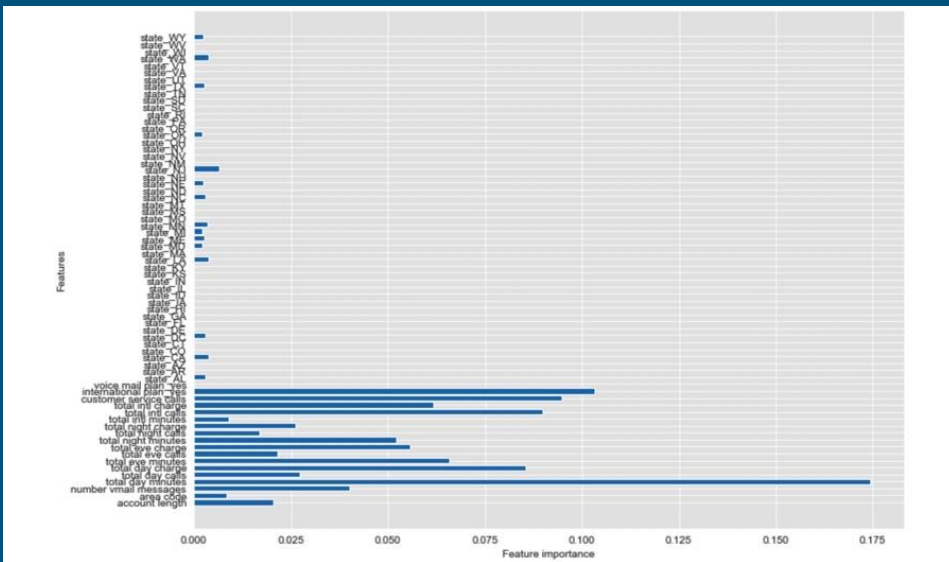
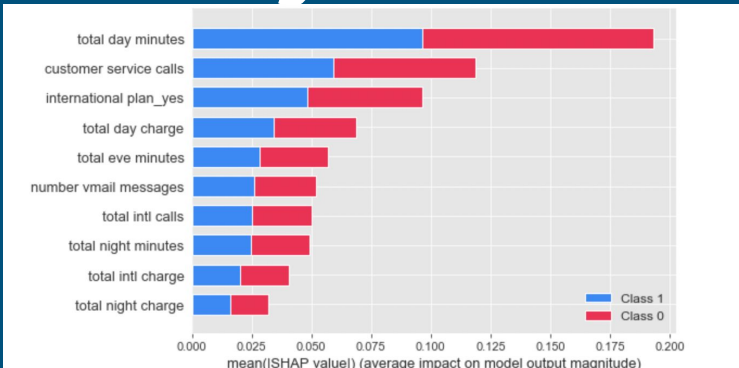
Visualizing Results of Model with SHAP



On the right side of the graph, the legend indicates that RED is HIGH feature value and BLUE is LOW feature value. On the y axis, the features are divided individually. The x axis ranks how significant the impact is.

If the feature has a tail going to the right, it means that those values are causing an impact on model output, pushing customer churn from zero (not churning) to one (churn).

Findings



Highest contributors of customer churn:

- Having an international plan
- Total number of day minutes
- High number of customer service calls
- Total number of evening minutes

Lowest contributors of customer churn:

- Number of Voicemail Messages
- Total international calls
- State
- Area code

It is evident that SyriaTel charges customers based on the number of minutes they use to make calls. The highest contributors of customer churn are factors that lead to a higher bill.

Findings and Recommendation

- It is evident that SyriaTel's current business model charges customers per minute of phone use. Along with high customer service calls, the highest contributors of churn are factors that lead to a higher bill.
- According to the data, the average charge per user is about \$60. My recommendation would be for the company to charge a flat monthly fee to the demographic of customers who are charged \$60 or less monthly and a higher tier plan for those who use their phones more.
- If SyriaTel can retain 247 more of the customers in this dataset, they would reach their goal of reducing their current churn rate from 14% to 7%. As their business model improves, it would be wise to compare their processes with their competitors', and continue to evaluate their churn rate with renewed perspective.

Thank you.