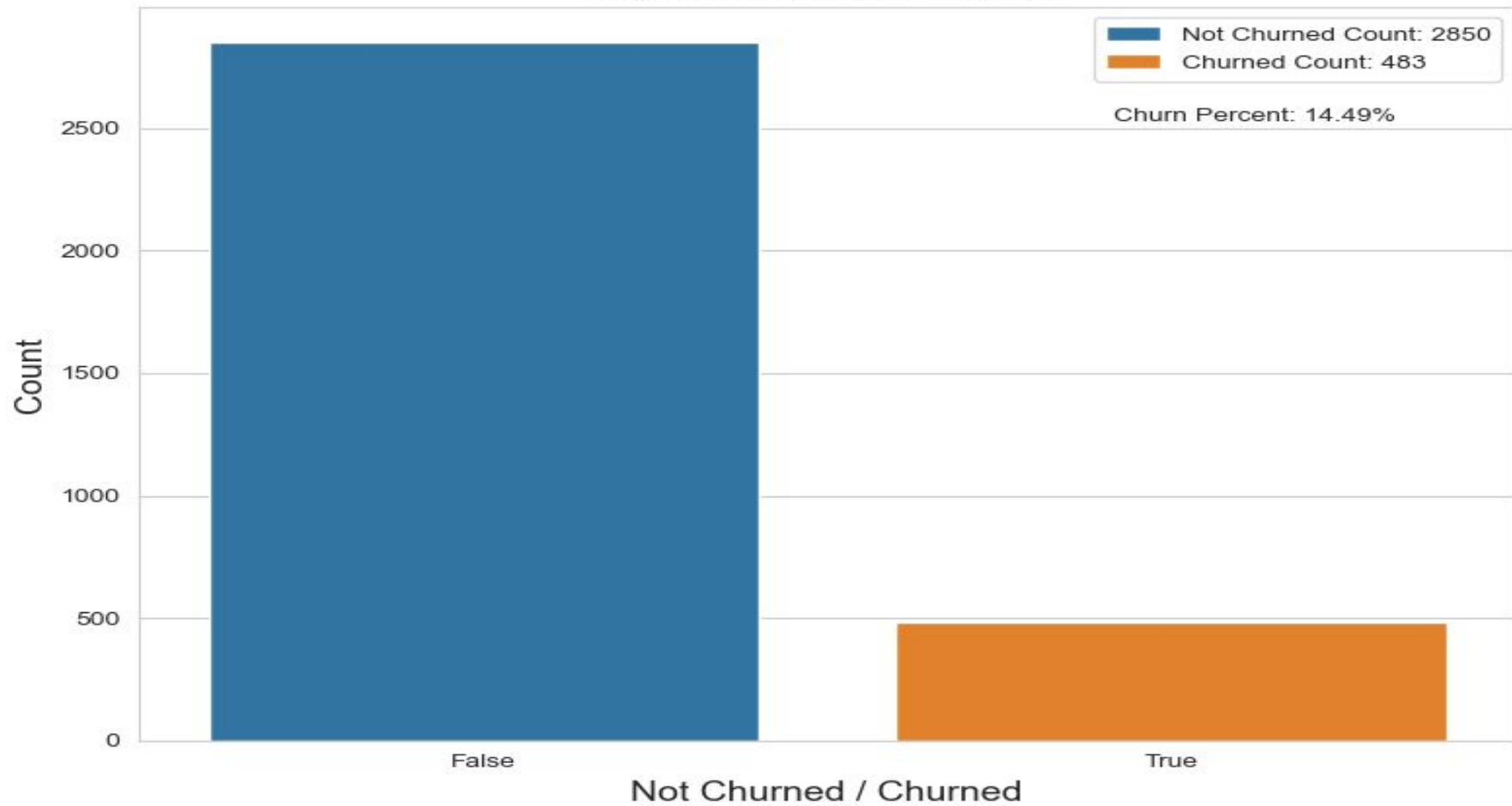# Customer Churn in Telecoms

By Jesse Moore

Overview

SyriaTel wishes to understand more about why customers cancel their service ("churn") to help them develop a more effective retention strategy.

Our goal is to create a predictive model to correctly analyze churn targeting for 80% recall.

'Recall' is a statistic that measures how accurately we predict True Positive cases, thus the higher the recall, the lower our number of false negatives.

A false negative means we have missed a customer that has churned, costing our client on average 5 to 7 times more than retaining an existing client.

Customer Churn Counts

| | Not Churned Count: 2850 |
| | Churned Count: 483 |

Churn Percent: 14.49%

Count

Not Churned / Churned

False   True

The first step is to build a baseline model using just one feature, 'customer service calls'. We'll evaluate the performance, and then scale up by adding more features and tuning hyperparameters using GridSearchCV.

This model performs poorly, with a Recall: 0.247, and predicting 73 False Negatives

Baseline Model
Confusion Matrix: Decision Tree (Validation Set)
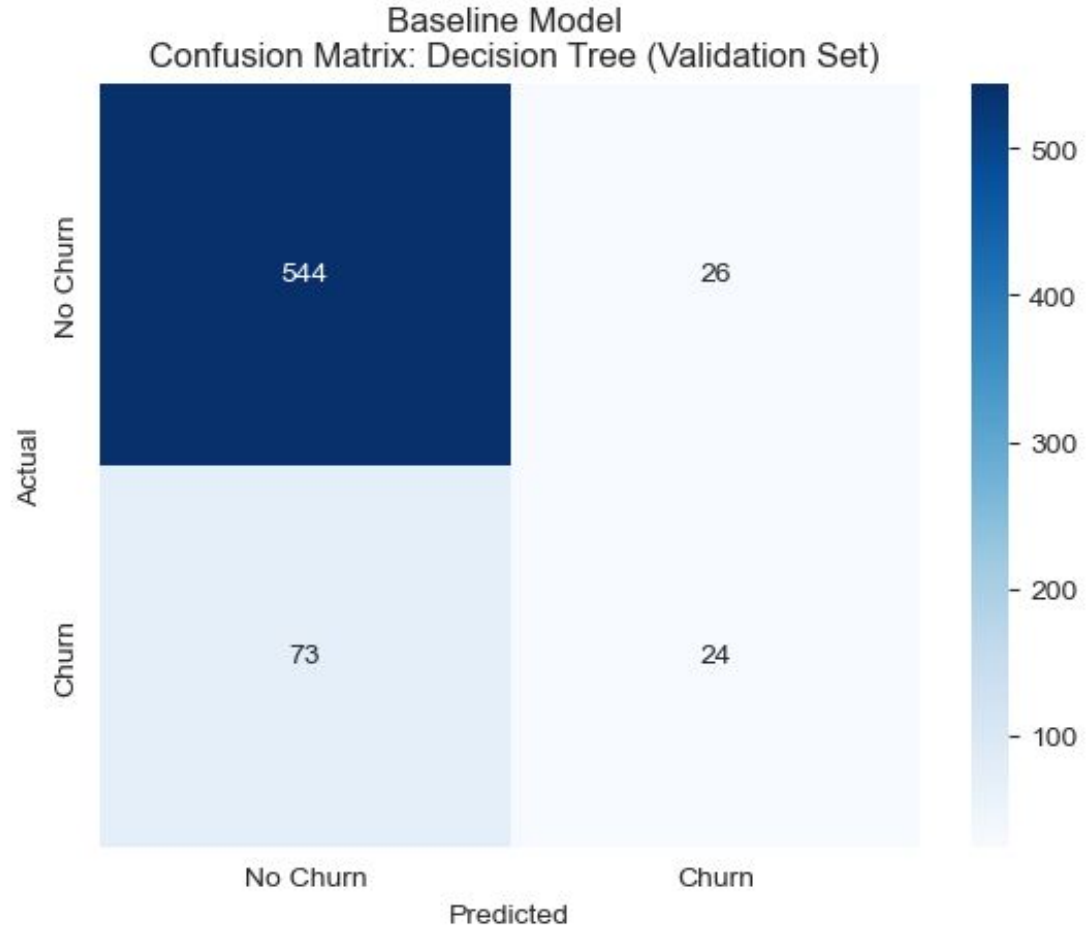
After several adjustments to our hyperparameters, reducing the number of min_sample_leaf and min_sample splits, and including all of the features, we achieve the Recall score of: 0.845 and 15 False Negative predictions.

This is an excellent model to test for our final model!

Ninth Model
Confusion Matrix: Decision Tree (Validation Set)
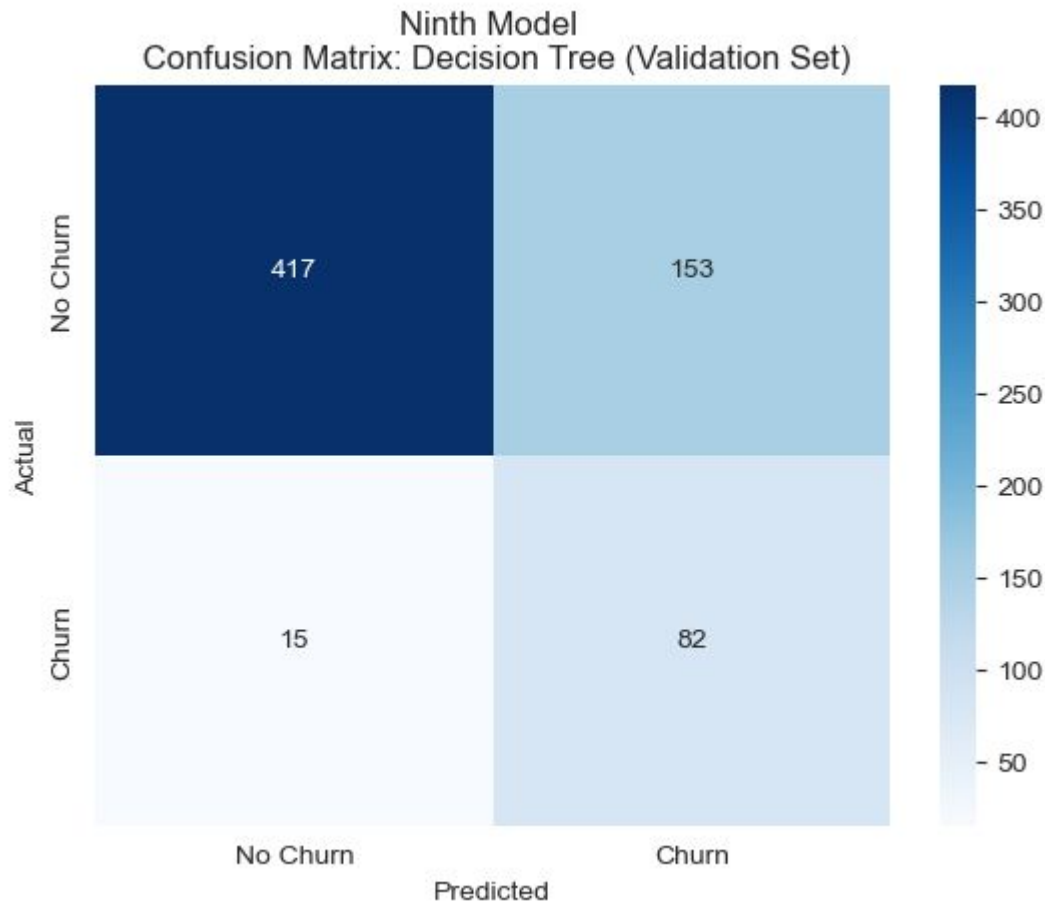
As we can see from our Confusion Matrix, we have minimized our false negative predictions to 16, from 73.

Test Set Final Evaluation:
  Accuracy: 0.744 | Precision: 0.340 | Recall: 0.833 | F1-Score: 0.483

With a recall of 0.833, we have achieved our client's goal of predicting with churn with an 80% recall success metric.

Confusion Matrix: Decision Tree (Test Set)
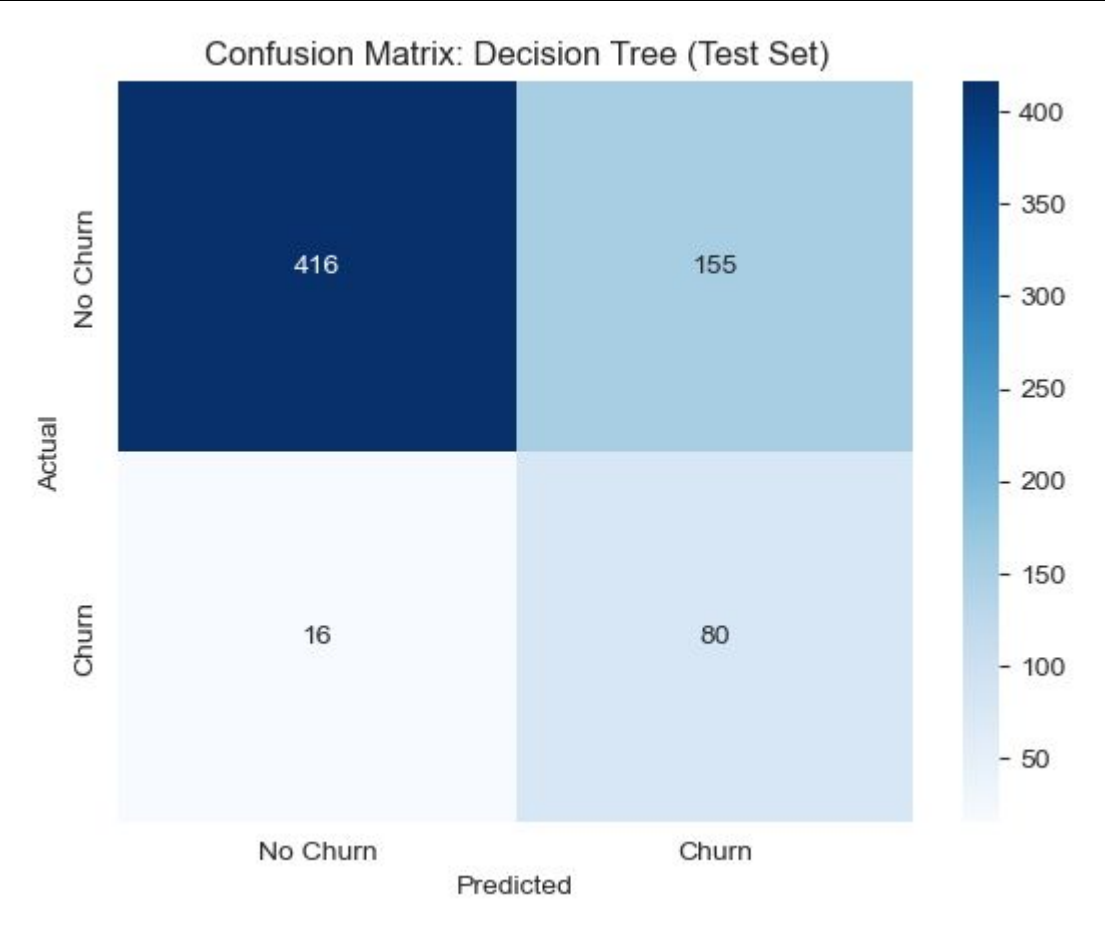
For more information:

[Project Repository](#)
[My Github](#)
[My LinkedIn](#)

Thank you for your time!