

SYRIATEL CUSTOMER CHURN PREDICTION

The background image is a surreal, apocalyptic cityscape. In the center, a large, red and white dome, resembling a giant balloon or a stylized building, dominates the skyline. Below it, a dense cluster of buildings, some appearing to be under construction or in ruins, sits on a hill. A construction crane is visible on the right side of the hill. The city is surrounded by a body of water in the foreground, and the sky is filled with large, billowing clouds. The overall tone is dramatic and somber.

Phase 3 project By. Rwenji Murengaitta



PROJECT OVERVIEW

Keeping customers happy is really important for businesses, especially in the competitive telecommunications industry. In any business it is cheaper and easier to maintain the customer base than to add new customers.

In line with customer retention we are developing a predictive model for SyriaTel that can predict which customers might leave. We're going to look at different kinds of information about the customers, like their location, how they use the service, if they have made any complaints, and their billing history. By doing this, we hope to find patterns and reasons why customers might leave.

BUSINESS PROBLEM

SyriaTel has experienced considerable financial setbacks as a result of a noticeable churn rate. The objective of this initiative is to develop an accurate predictive model that identifies customers at risk of churning within SyriaTel, a telecommunications company.

By proactively identifying potential instances of service discontinuation, the goal is to reduce customer attrition and maintain a larger customer base. Ultimately, this endeavor aims to help SyriaTel mitigate financial losses associated with churn, improve overall customer retention rates, and enhance business strategies to increase profitability.

Project Objectives

- To comprehend the elements influencing customer churn
- To foster customer loyalty
- To improve overall customer retention.

DATA UNDERSTANDING

The dataset contains data about customer churn for the SyriaTel Communications Company.

- The dataset had 3333 rows and 21 columns
- There are three main data types in the data; float, integer and object.

DATA CLEANING



No missing values.



No duplicates found

- Created new column "total_cost" to cover different times of the day/night
- Converted "area code" typed as numeric data to be categorical as it had only 3 unique values.

MODELLING

During EDA, discovered that there was perfect correlation between the following variables:

- * Total day minutes and total day charge
- * Total eve minutes and total eve charge
- * Total night minutes and total night charge
- * Total intl minutes and total intl charge

Chose to drop the following columns to deal with multicolleniariry: Total day minutes, Total eve minutes, Total night minutes and Total intl minutes.

Scaled the data to deal with outliers, used `train_test_split` method to divide our data into the training and testing data we need to perform our modeling. Used the SMOTE technique to deal with data imbalance.

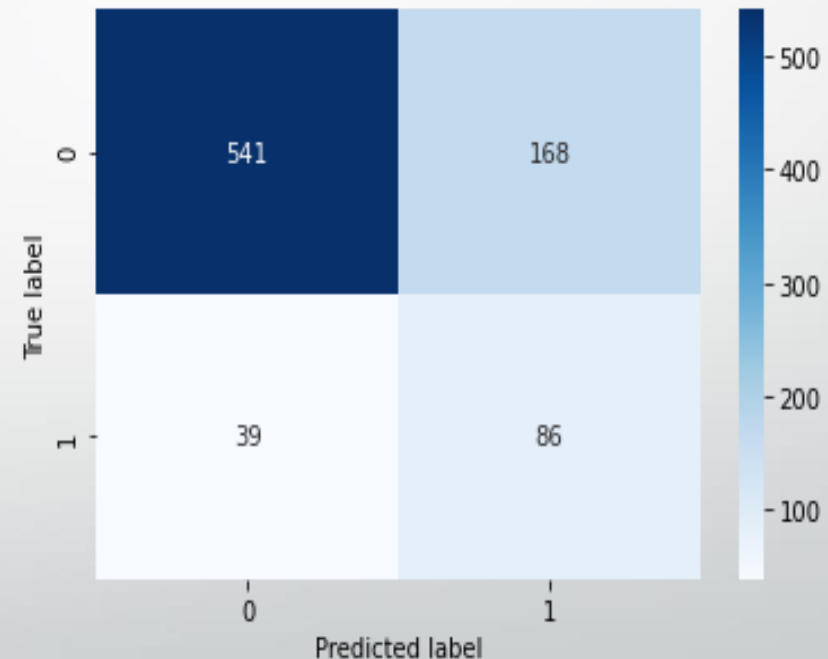
used 3 different models to check which performed better. The models were:

- * Logistic Regression.
- * Decision trees
- * Random Forest

BASELINE MODEL

Our baseline model was a Logistic Regression Model with churn as the predicted variable. The model had a recall score of 69% for no churn and accuracy of 75%

+-----+-----+-----+-----+				
Precision	Recall	F1-Score	Support	
+-----+-----+-----+-----+				
0.93	0.76	0.84	709	
0.34	0.69	0.45	125	
+-----+-----+-----+-----+				
-	-	-	834	
+-----+-----+-----+-----+				
0.64	0.73	0.65	-	
+-----+-----+-----+-----+				
0.84	0.75	0.78	-	
+-----+-----+-----+-----+				
Accuracy			0.75	
+-----+-----+-----+-----+				

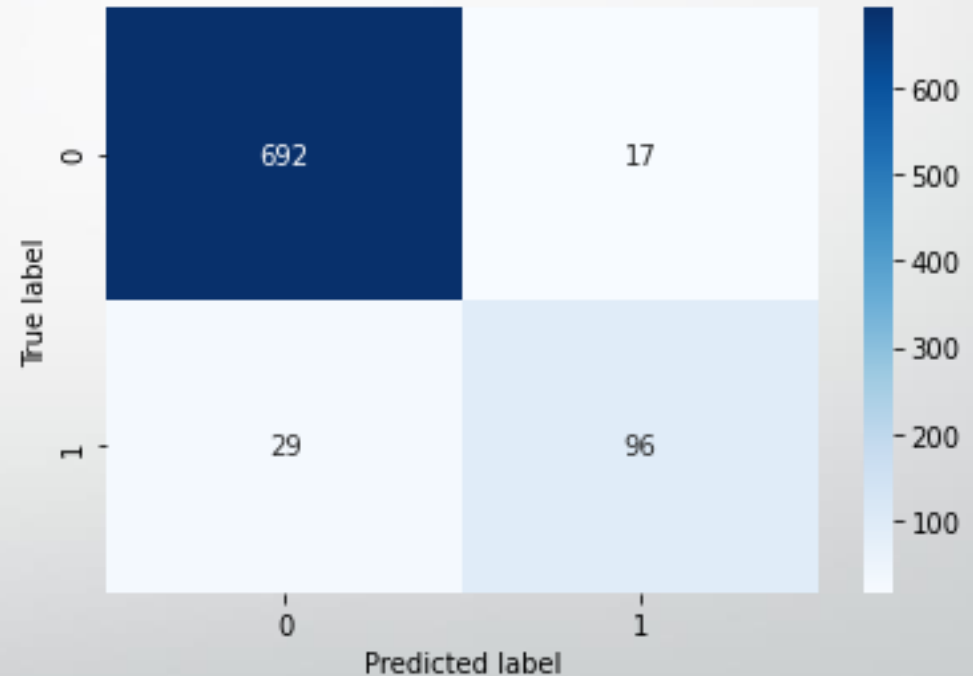


FINAL MODEL

The most robust model was the Random Forest one and after tuning the overall results were:

- Recall score of 0.77, which means model is able to capture 77% of the actual churned customers
- An accuracy of 94% meaning 94% of test data was correctly predicted.

+-----+-----+-----+-----+			
Precision	Recall	F1-Score	Support
+-----+-----+-----+-----+			
0.96	0.98	0.97	709
0.85	0.77	0.81	125
+-----+-----+-----+-----+			
-	-	-	834
+-----+-----+-----+-----+			
0.90	0.87	0.89	-
+-----+-----+-----+-----+			
0.94	0.94	0.94	-
+-----+-----+-----+-----+			
Accuracy			0.94
+-----+-----+-----+-----+			



CONCLUSION

The random forest emerged as the top-performing model for two key reasons. Firstly, it demonstrated a notable recall score of 0.77, indicating a 77% accuracy in identifying customers at risk of churn. This aligns perfectly with our goal of effectively recognizing and addressing churn.

The random forest model exhibited a commendable accuracy score of 0.94, signifying a 94% success rate in correctly classifying churn cases. This high accuracy underscores the model's effectiveness in predicting churn accurately, a critical factor for informed business decision-making.

The model's impressive ability to harmonize precision and recall is evident in its F1 score of 0.81. Combining these metrics into a single measure, the F1 score reflects a balanced evaluation of the model's performance. With a higher F1 score, the random forest model efficiently identifies true positives while minimizing both false positives and false negatives. In our context, the F1 score of 0.81 underscores the model's adeptness at striking a precise balance between precision and recall, ensuring accurate identification of churn cases.

RECOMMENDATIONS

SyriaTel can improve customer retention by:

- Using the Random Forest Classifier for predicting churn in real-time. Regularly monitoring the model ensures it stays effective, with updates adjusting to changing customer behaviors
- Analyzing feature importance helps identify areas for targeted service improvements and personalized retention efforts. What is observed is that call cost, duration of calls and charges related to voice mail services have a huge impact on the churn rate. Having better incentives for categories of customers who use these services are more likely to reduce churn.
- Work closely with retention teams ensures smooth integration and informed decision-making.
- By combining quantitative predictions with qualitative customer feedback, SyriaTel gains a thorough understanding for crafting nuanced retention strategies.
- Continuous model enhancements and investments in customer experience initiatives contribute to long-term success, optimizing resource allocation and nurturing lasting relationships in the telecommunications industry.