PREDICTING CUSTOMER CHURN FOR SYRIATEL

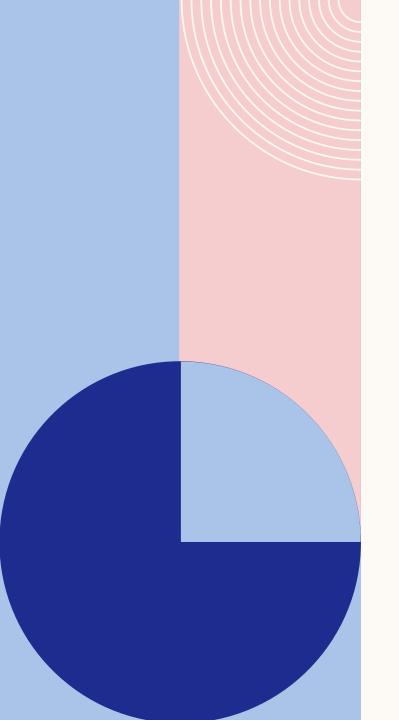
OVERVIEW

- This project aims to build a classifier to predict whether a customer will ("soon") stop doing business with SyriaTel, a telecommunications company.
- This is a binary classification problem with 2 possible outcomes:
- The customer will soon stop doing business with SyriaTel.
- The customer will not stop doing business with SyriaTel.

BUSINESS UNDERSTANDING

- The loss of customers in a company such as SyriaTel means loss of revenue and increased costs since acquiring new customers is more expensive than retaining existing ones.
- Predicting customer dropout in advance helps SyriaTel identify high-risk customers early therefore implementing cost-effective retention strategies leading to reduced revenue loss and improved customer satisfaction which leads to profitability to the telecommunication company.





KEY BUSINESS QUESTIONS

- What are some of the factors that contribute to customer dropout?
- How can SyriaTel leverage on these factors to come up with solutions?

DATA UNDERSTANDING

The dataset consists of customer records from SyriaTel, a telecommunications company with features that help predict whether a customer will stop using the service.

Some of these key features include:

- Account length
- Area code
- Phone number
- Total day calls
- Total intl calls

- Customer service calls
- Number vmail messages
- International plan
- Total night minutes

Understanding these patterns helps in building a predictive model that assists SyriaTel in retaining valuable customers.

CONTINUATION

Some of these key features include:

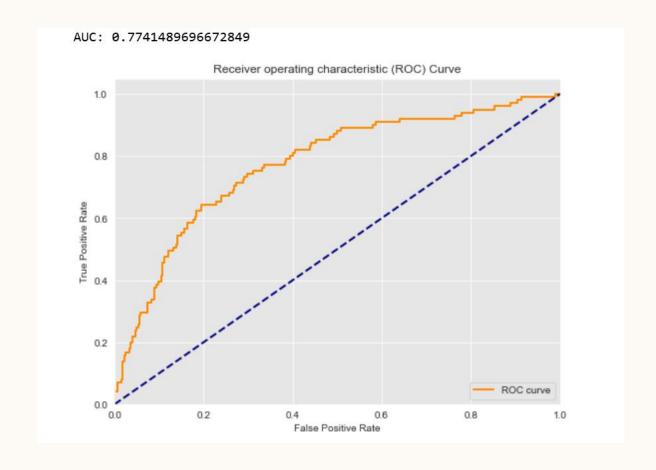
- Area code
- Phone number
- Total day calls
- Total intl calls
- Customer service calls
- Number vmail messages
- International plan
- Total night minutes

Understanding these patterns helps in building a predictive model that assists SyriaTel in retaining valuable customers.

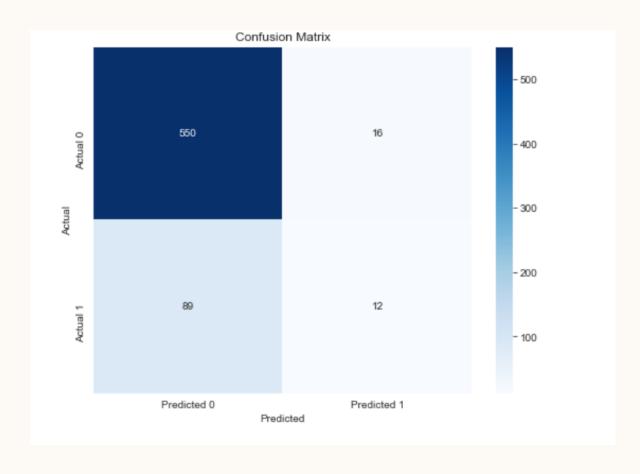
MODELING AND EVALUATION

- A higher AUC is shown indicating that SMOTE improved the model's performance on the minority class which in this case is churn.
- The Confusion Matrix displayed a lower f1score indicating a poor performance, meaning the model isn't effective at identifying churn.
- Our analysis developed a churn prediction model for SyriaTel, achieving a high accuracy of 93.55% using a tuned Decision Tree Classifier in this case GridSearchCV.

ROC CURVE



CONFUSION MATRIX



DECISION TREE

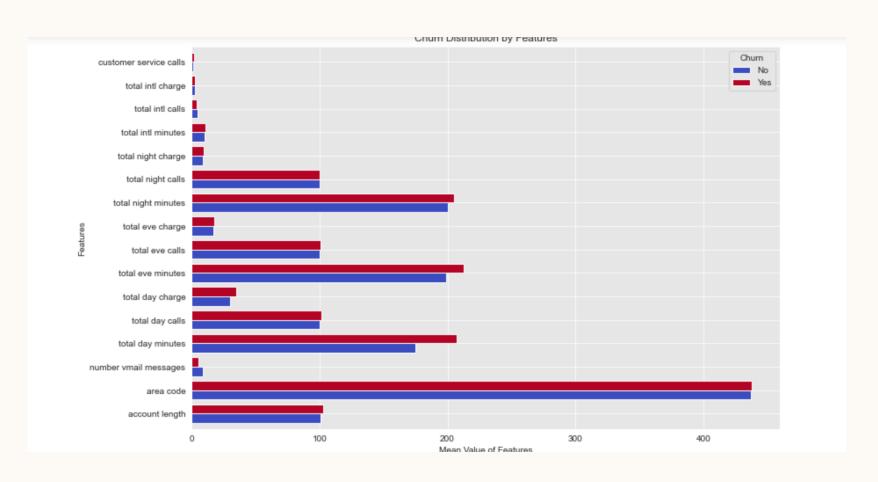
total day charge <= 0.749 gini = 0.246 samples = 2666 value = [2284, 382] class = No Churn

customer service calls <= 0.389 gini = 0.201 samples = 2492 value = [2209, 283] class = No Churn number vmail messages <= 0.127 gini = 0.49 samples = 174 value = [75, 99] class = Churn

international plan yes <= 0.5 gini = 0.144 samples = 2287 value = [2108, 179] class = No Churn total day minutes <= 0.491 gini = 0.5 samples = 205 value = [101, 104] class = Chum total eve charge <= 0.508 gini = 0.395 samples = 129 value = [35, 94] class = Chum international plan_yes <= 0.5 gini = 0.198 samples = 45 value = [40, 5] class = No Churn

gini = 0.091 samples = 2078 value = [1979, 99] class = No Chum gini = 0.473 samples = 209 value = [129, 80] class = No Churn gini = 0.294 samples = 106 value = [19, 87] class = Chum gini = 0.284 samples = 99 value = [82, 17] class = No Churn gini = 0.466 samples = 46 value = [29, 17] class = No Churn gini = 0.134 samples = 83 value = [6, 77] class = Churn gini = 0.0 samples = 38 value = [38, 0] class = No Churn gini = 0.408 samples = 7 value = [2, 5] class = Chum

CHURN DISTRIBUTION BY FEATURES



CONCLUSION

- Our analysis developed a churn prediction model for SyriaTel, achieving a high accuracy of 93.55% using a tuned Decision Tree Classifier.
- About 95% of customers predicted as non-churners was actually correct while 82% of those predicted as churners was equally correct(Precision).
- The model identifies 97% of non-churners and 73% of churners correctly meaning it still misses a percentage of churners (Recall).
- The model has a higher f1_score of 96% meaning it identifies non-churners with ease as compared to churners.

RECOMMENDATIONS

- Since recall for churners is 73%, the model correctly identifies many customers who are likely to leave. SyriaTel can target these high-risk customers by offering discounts or improved customer service.
- Ensure high-risk customers complaints are resolved in the shortest time possible and this reduces frustrations among them.
- If certain area codes have a significantly higher churn rate, SyriaTel should investigate whether these regions have poorer service quality, limited coverage, or strong competitors.
- The company can reach out to high-usage customers with special offers which prevents churn.

SUMMARY

Customer churn is a major challenge for SyriaTel impacting revenue and long-term growth. Using machine learning, we built a Decision Tree classifier that achieved 93.55% accuracy, effectively predicting which customers are likely to stop using services provided by SyriaTel. By taking note of these insights, SyriaTel can make steps to reduce churn, increase customer satisfaction and maximize revenue.

THANK YOU

Frankline Kipchumba
7frankipchumba@gmail.com