SYRIATEL CHURN PREDICTION ANALYSIS

GROUP PROJECT MEMBERS

- Joseph Kinuthia

- John Mark

- Peter Kariuki

- Collins Kanyiri

- Calvin Kipkirui

- Raphael Muthemba

INTRODUCTION

- In a rapidly evolving telecommunications environment, customer churn remains a significant challenge that directly impacts business profitability and sustainability.

- What is a customer churn? Also known as customer attrition, refers to the activity where a customer stops using a company's service for whatever reason.

- Churn analysis is the technique used to understand why customers leave, predict when a customer is likely to stop using and most importantly to determine which strategy to employ to retain customers.

- We will use this technique to provide SyriaTel, a telecommunication service provider, with the right information to implement targeted efforts for retaining their valuable customer base.

PROBLEM STATEMENT

- SyriaTel is experiencing a challenge in understanding why some of their customers are discontinuing from their service

- The client wants to retain as many customers as possible.

- The company needs a model that can infer the features that would make a customer to stop using their service.

- The model needs to be trained on data that accurately represents churn likelihood.

- The results can be used by different stakeholders such as the management, marketing team, customer support team to effectively improve customer retention.

OBJECTIVES

Main Objective

- To build a classifier developing accurate predictive models capable of identifying customers who are likely to churn soon. Thereby, gaining insightful information on how to avoid this and build customer satisfaction while boosting business sustainability.

Specific Objectives

1. To identify key features that significantly influence a customer to churn.
2. To develop a model that accurately estimates churn likelihood.
3. To evaluate the performance of the developed model.
4. To interpret and give relevant advice to the client.

RESEARCH QUESTIONS

1. Can a predictive model accurately forecast whether a customer is likely to churn based on the available attributes and usage metrics?

2. Which features contribute the most to the model's predictions?

3. How well does the developed model generalize to new, unseen data? Are there certain patterns that the model consistently struggles to capture?
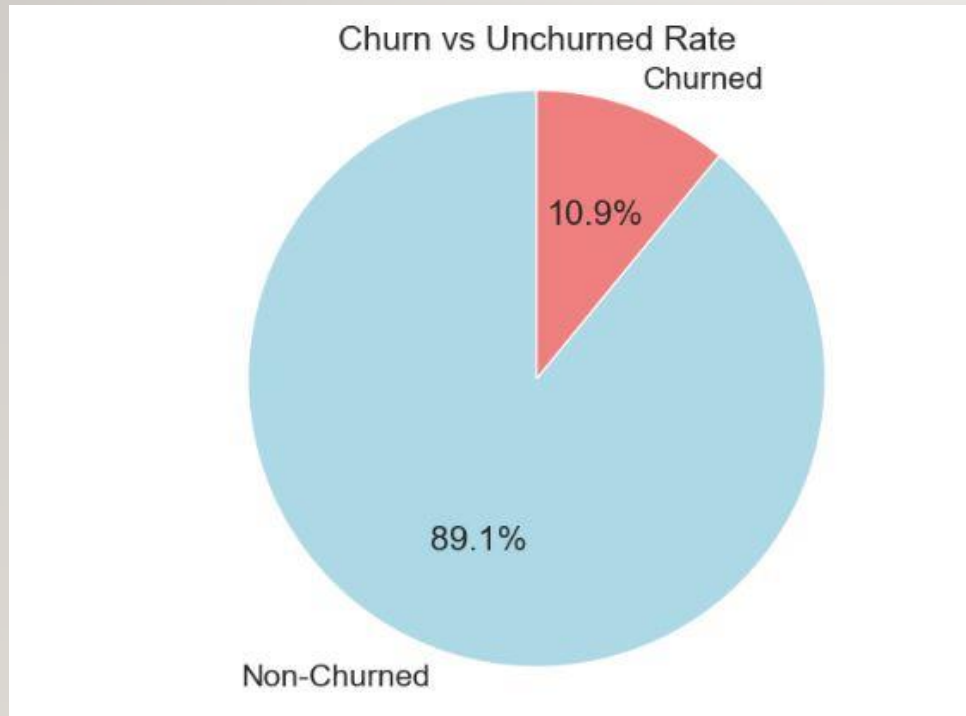
MODELS USED

1. Baseline Model

2. Random Forest Model

3. Decision Tree

4. Tuned Random Forest Model

5. Ensamble Methods

VISUALIZATIONS: CHURN COUNT
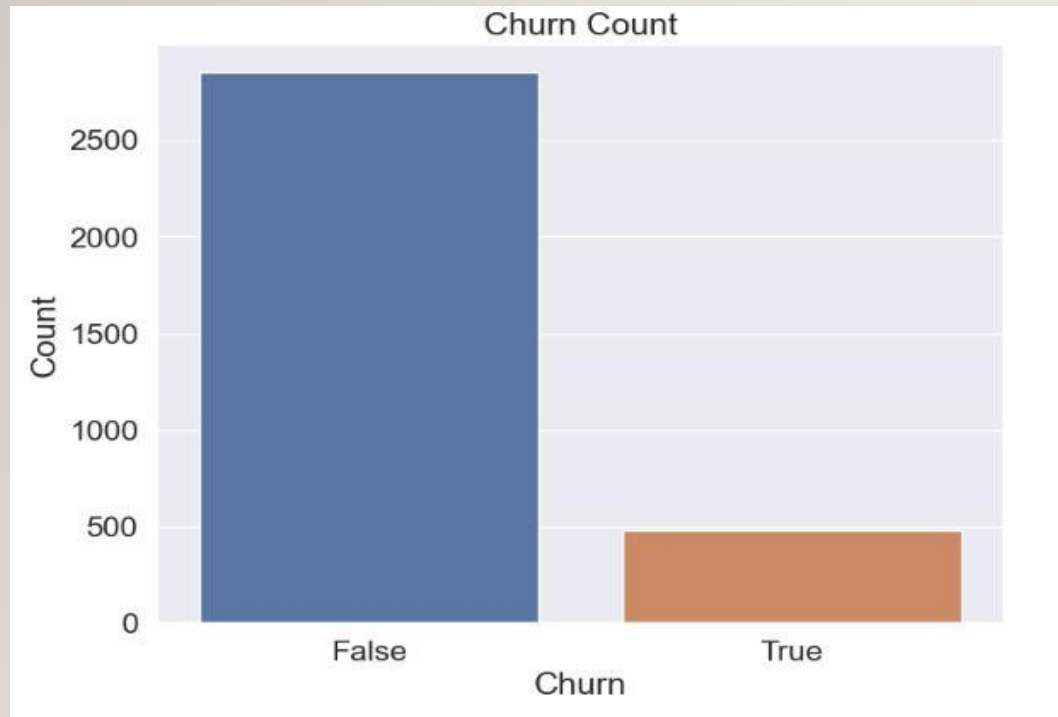
- We realize that the number of churns instances are just below 500.

VISUALIZATIONS: CHURN RATE

Churn vs Unchurned Rate

- We further realize that we have a churn rate of just about 11 %

VISUALIZATIONS: PAIR PLOT

Pair Plots of Numeric Features with Hue as 'Churn'

- It is evident that the majority of them exhibit a normal distribution, except for "Customer service calls."

- We also notice a relation between customer service calls and churning as after the fourth call the customer deregisters.

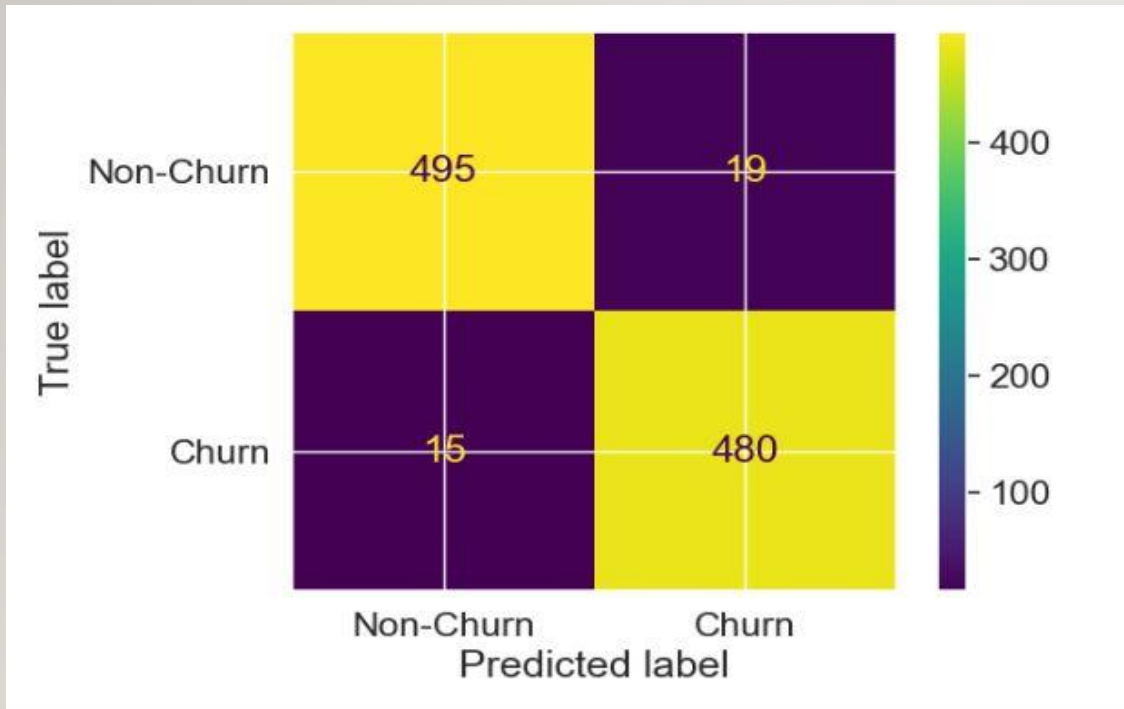# FINDINGS: RESEARCH QUESTION 1(LOGISTIC REGRESSION)
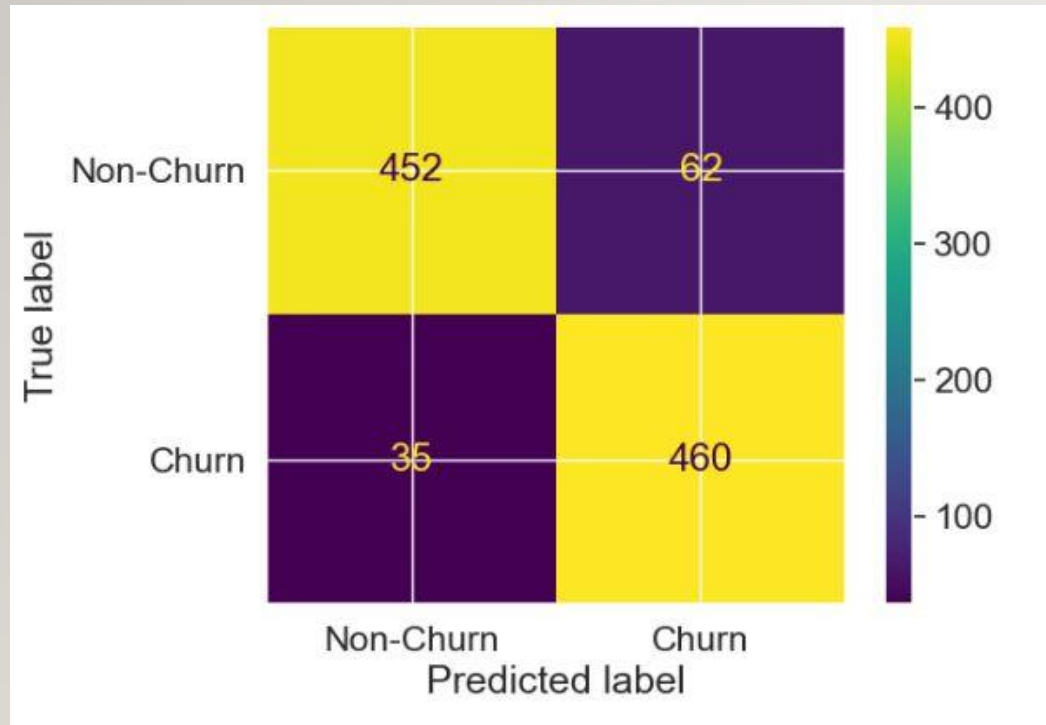
- We used this model as our baseline model so we will compare our results to this

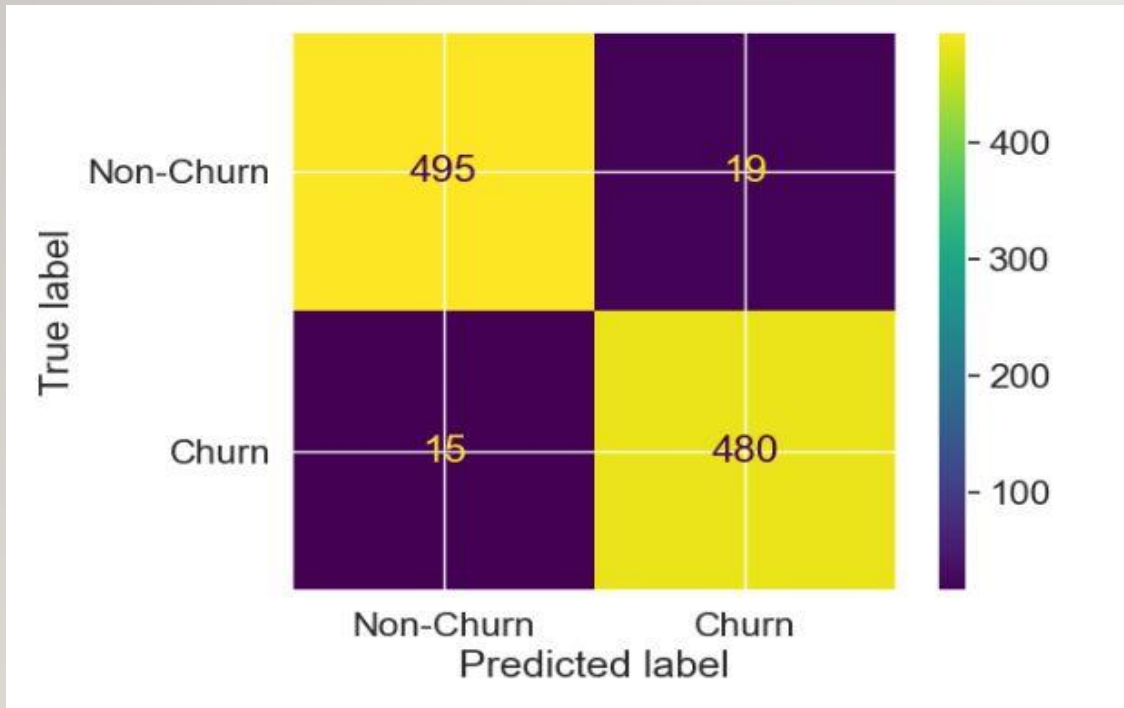- Accuracy: 0.90 (90%)

FINDINGS: RESEARCH QUESTION 1 (DECISION TREE)
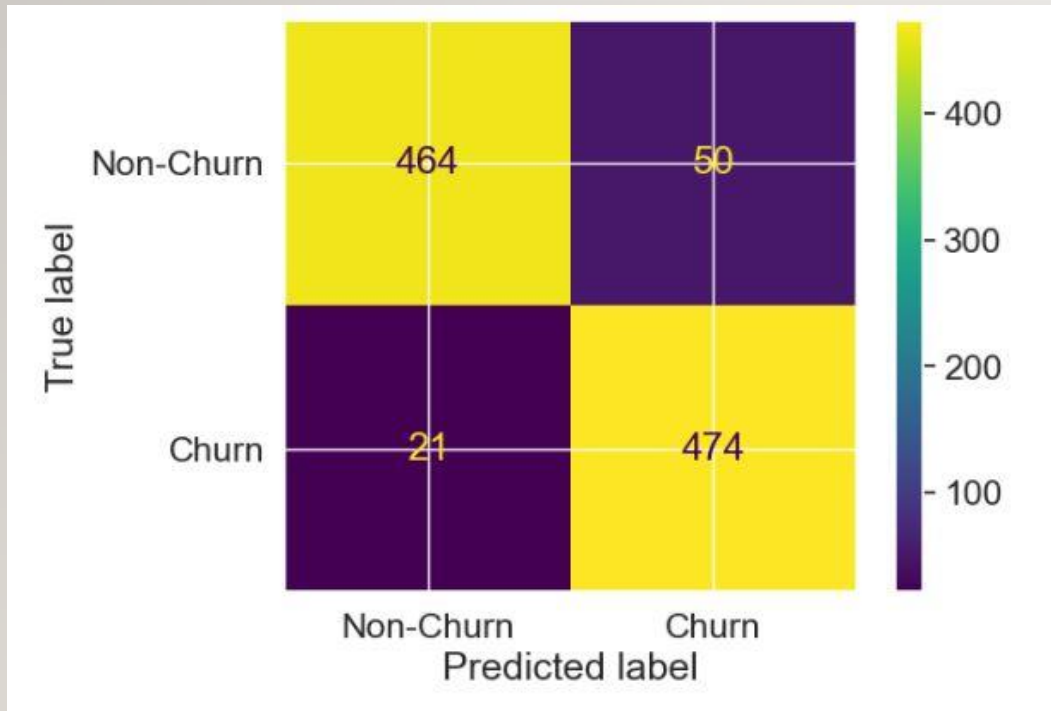
- Accuracy: 0.93 (93%)

- A test for overfitting found out that the model may be overfitting.

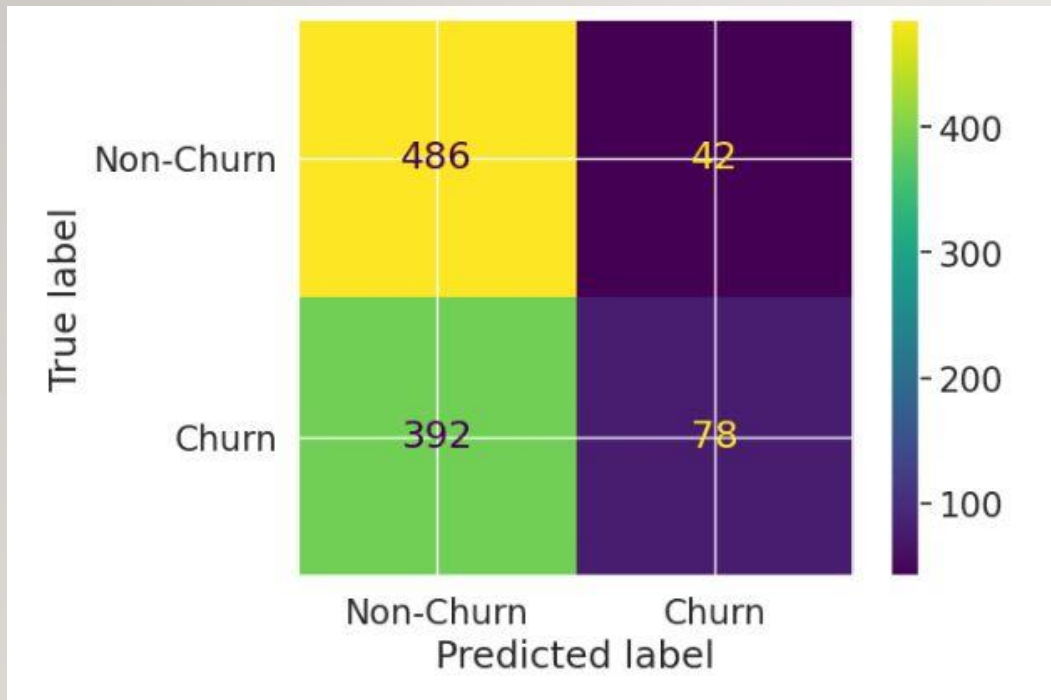# FINDINGS: RESEARCH QUESTION 1(RANDOM FOREST CLASSIFIER)

- Accuracy: 0.97 (97%)

- A test for overfitting found out that the model may be overfitting.

FINDINGS: RESEARCH QUESTION 1(TUNED RANDOM FOREST CLASSIFIER)

- Accuracy: 0.97 (97%)

- A test for overfitting found out that the model is likely not to be overfitting.

- Accuracy: 0.95 (95%)

- A test for overfitting found out that the model is likely not overfitting.

CONCLUSION

1.Model Performance: The Tuned Random Forest Classifier, which gives a success rate of 97%, and the Ensemble Method using XGBoost with a success rate of 95%. These models demonstrated the highest performance in predicting customer churn. Overall, they exhibited superior accuracy, precision, recall, F1 score, and ROC AUC score compared to other models. The models are also not overfitting and generalizes well to new data.

2.Imbalanced Data: The initial distribution of churned and non-churned customers in the dataset was imbalanced, with a significantly larger number of non-churned customers. This imbalance could have led to misleadingly high accuracy scores. Addressing this imbalance using techniques like SMOTE improved model performance.

3.Feature Importance: During model analysis, features such as "customer service calls" were identified as having a significant impact on churn prediction. Customers making multiple service calls were more likely to churn. Identifying such influential features can guide targeted efforts to reduce churn rates.

RECOMMENDATIONS

1. **Focus on Recall**: Given the nature of the problem, where identifying potential churners is crucial, we recommend placing more emphasis on recall for the positive churn class. This will help ensure that the company doesn't miss out on identifying customers who are likely to churn.
2. **Personalized Retention Strategies**: Leverage the insights gained from the predictive models to design targeted retention campaigns. Tailoring strategies based on individual customer behaviors and characteristics can improve customer satisfaction and loyalty.
3. **Enhanced Customer Support**: Use the churn prediction models to identify customers who might require additional support. Proactively addressing their concerns and issues can prevent them from churning.
4. **Resource Allocation**: Allocate resources more efficiently by focusing on customers who are more likely to churn. This will help optimize operational efficiency and minimize costs.
5. **Continuous Monitoring and Adaptation**: Customer behaviors and preferences evolve over time. Regularly update and retrain the churn prediction models to ensure they remain effective in capturing changing patterns.

THANK YOU

GROUP 10