

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 attributes that would make a customer to stop using their services.
- The model needs to be trained on data that accurately represents churn likelihood.
- The results can be used by different stakeholders such as the company management, marketing team and 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. This will help gain insightful information on how to retain customers 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.

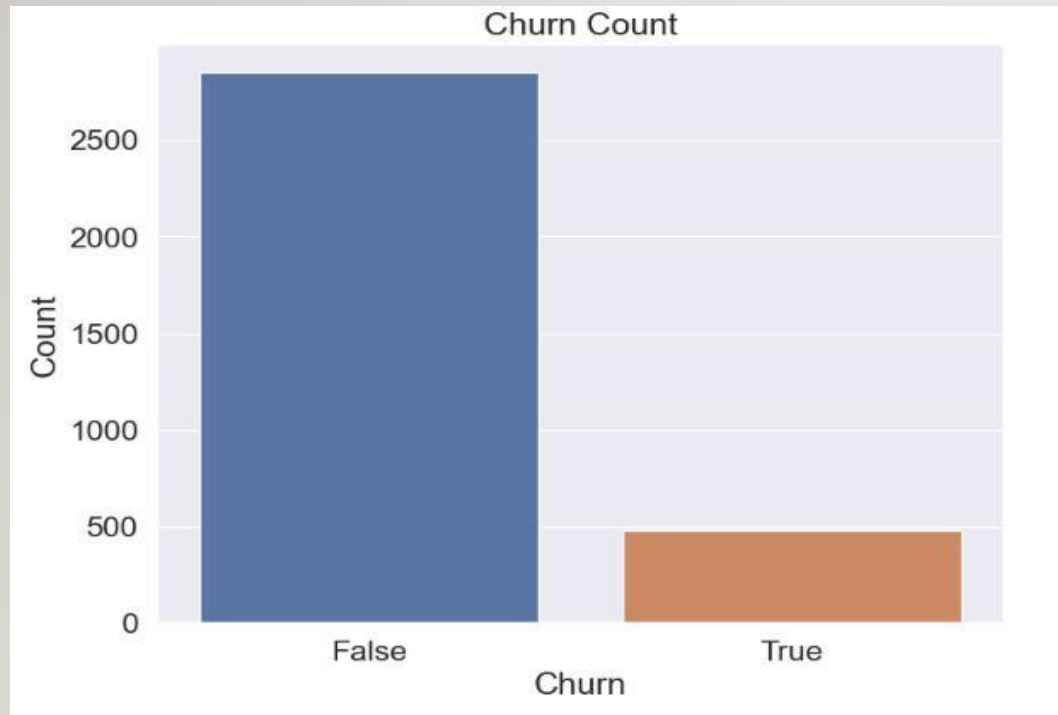
RESEARCH QUESTIONS

1. Can a predictive model accurately forecast whether a customer is likely to churn based on the available attributes and usage metrics?
2. How well does the developed model generalize to new, unseen data? Are there certain patterns that the model consistently struggles to capture?

MODELS USED

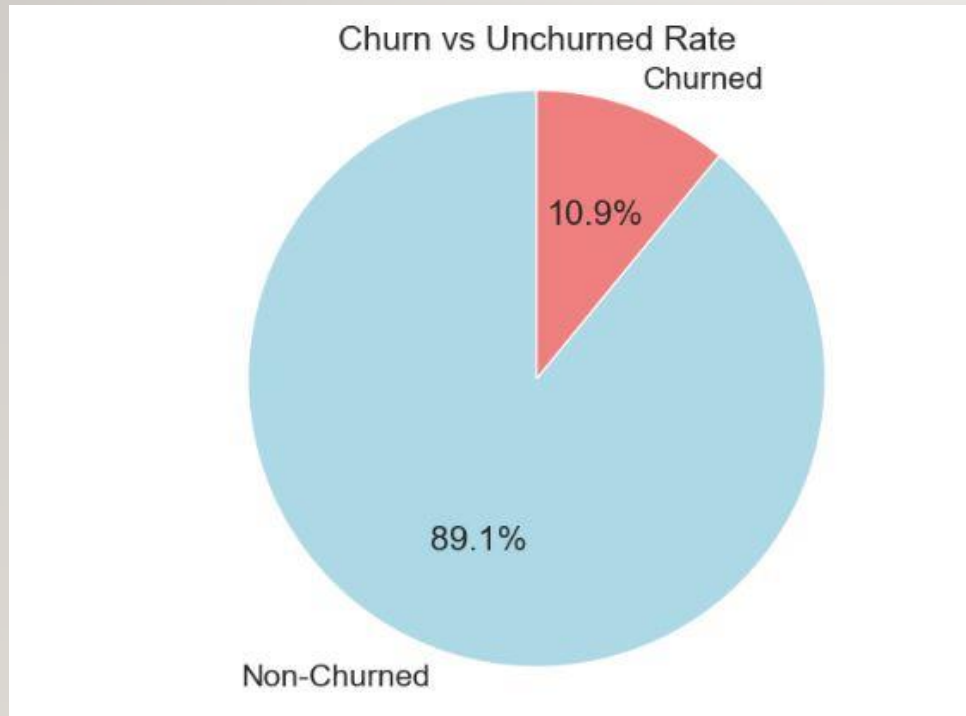
1. Baseline Model
2. Random Forest Model
3. Tuned Random Forest Model
4. Decision Tree
5. Hyperparameter Tuning Decision Tree Model
6. Ensemble Methods

VISUALIZATIONS: CHURN COUNT



- We realize that the number of churns instances are just below 500.

VISUALIZATIONS: CHURN RATE



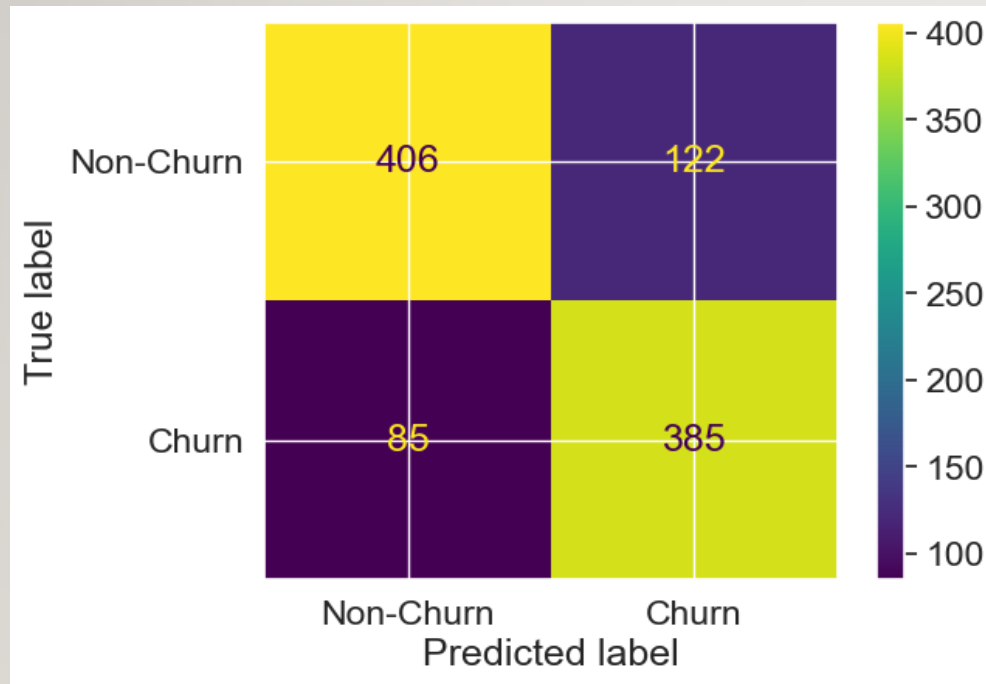
- We further realize that we have a churn rate of just about 11 % and non-churn rate of about 89%

VISUALIZATIONS: PAIR PLOT



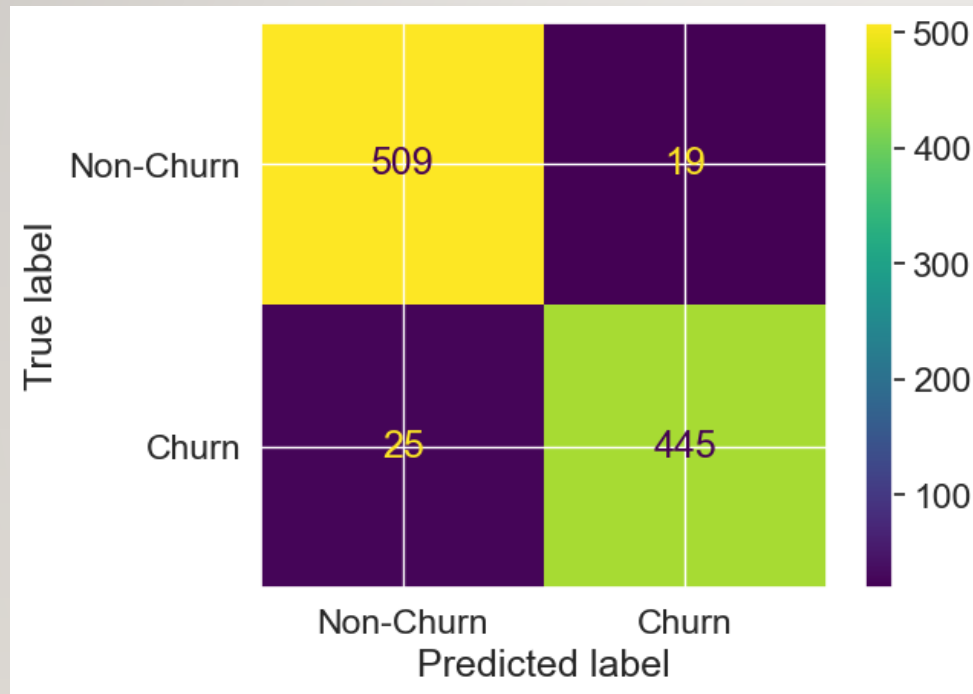
- It is evident that the majority of them exhibit a normal distribution, except for "Customer service calls."
- We also notice a relation between customer service calls and churning as after the fourth call the customer deregisters.

MODEL I: LOGISTIC REGRESSION



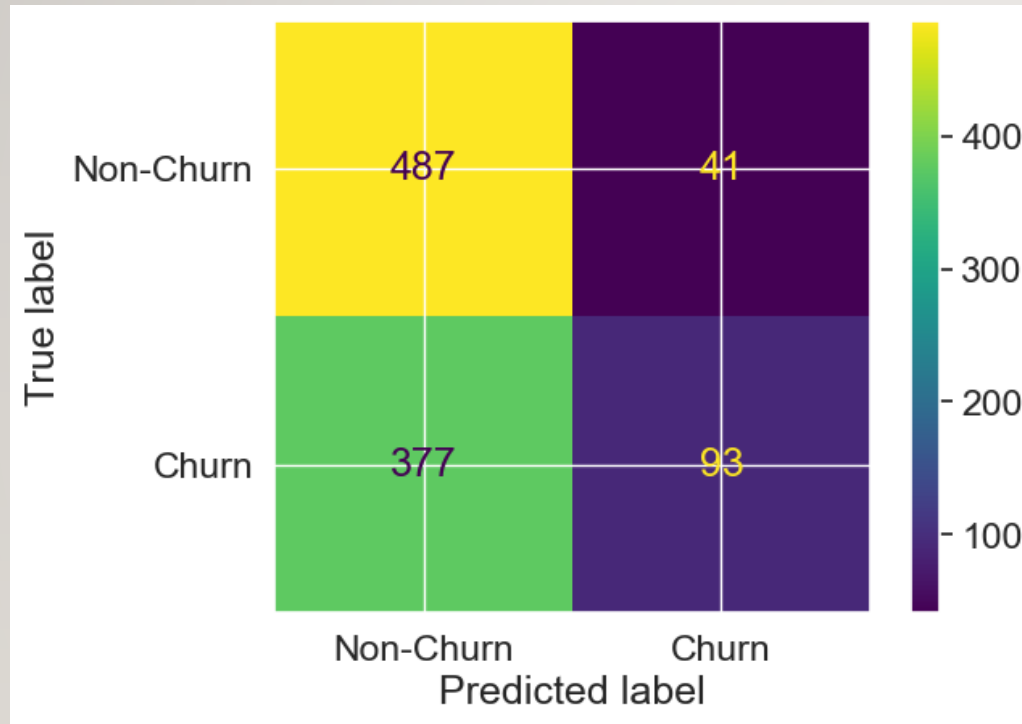
- We used this model as our baseline model so we will compare our results to this
- Accuracy: high accuracy of 0.79 (79%)
- Test for overfitting shows that the model is overfitting

MODEL 2: RANDOM FOREST CLASSIFIER



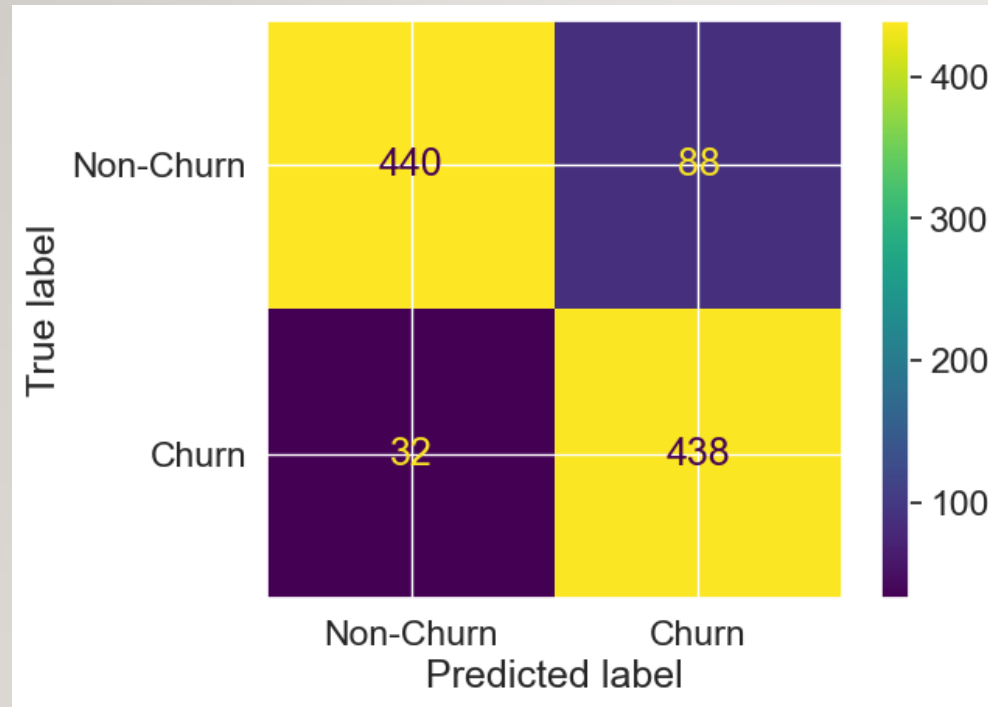
- Accuracy: high accuracy of 0.96 (96%)
- Test for overfitting shows that the model is overfitting

MODEL 3:TUNED RANDOM FOREST



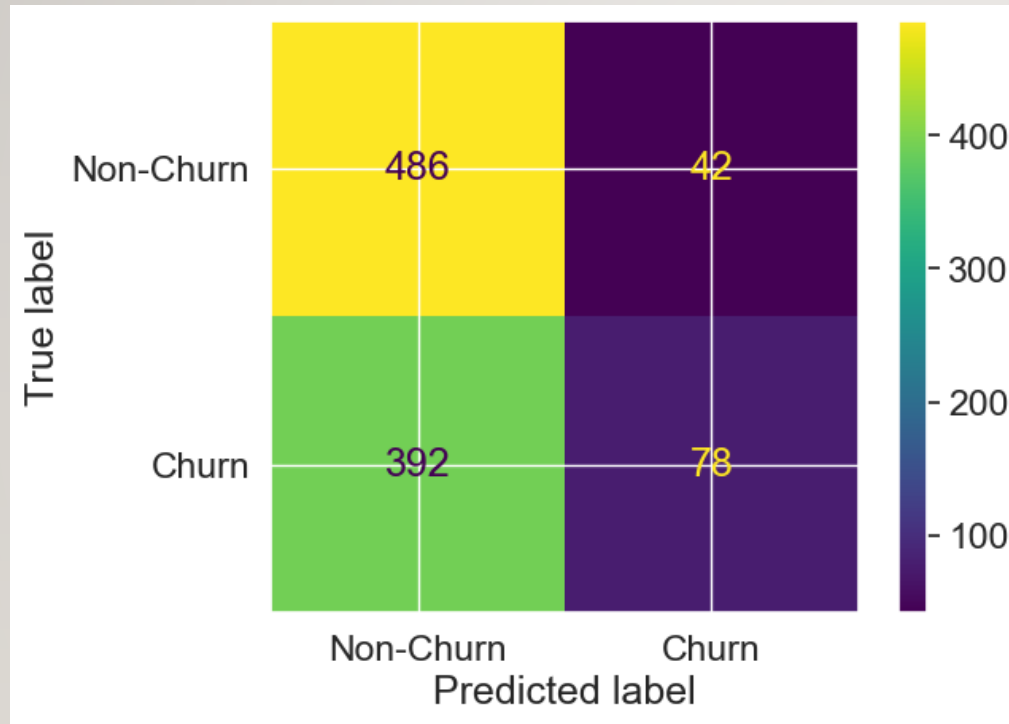
- Accuracy: medium accuracy of 0.58 (58%)
- Test for overfitting shows that the model is not overfitting

MODEL 4: DECISION TREE



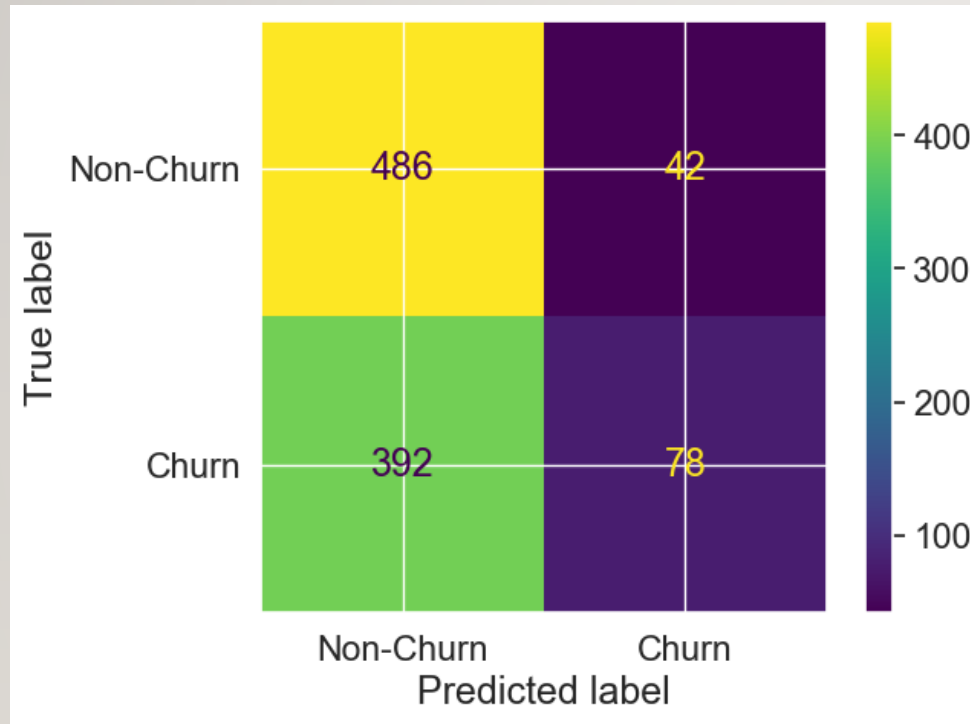
- Accuracy: high accuracy of 0.88 (88%)
- A test for overfitting found out that the model may be overfitting.

MODEL 5: HYPERPARAMETER TUNING



- Accuracy: 0.57 (57%)
- A test for overfitting found out that the model may not be overfitting.

MODEL 6: ENSEMBLE METHODS



- Accuracy: has medium accuracy of 0.57 (57%)
- A test for overfitting found out that the model is likely not overfitting.

CONCLUSION

- **Predictive Capability and Generalization:** The Random Forest Classifier emerged as the best-performing model in terms of predictive capability and generalization to new data. Evaluation of various metrics on both training and testing datasets shows its robustness in capturing complex relationships and patterns.
- **Trade-offs Between Precision and Recall:** The baseline Logistic Regression model had a relatively lower recall score compared to the Random Forest model, indicating a tendency to miss some churn cases.
- **Avoiding Overfitting:** The Tuned Random Forest model demonstrated the ability to mitigate overfitting concerns, highlighting the significance of hyperparameter tuning in achieving reliable model outcomes.
- **Challenges of Certain Models:** The Decision Tree and Ensemble Method using XGBoost exhibited struggles in generalization. The Decision Tree model suffered from overfitting, impacting its performance on the testing dataset., while the XGBoost model had medium accuracy of 0.57 (57%).

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