SyriaTel Customer Churn

Sylvia Sarange Manono

Overview

- SyriaTel is a telecom business facing customer churn challenges.
- **Customer churn:** when a customer ceases to be a customer.
- This project's objective is to predict customer churn and provide actionable insights to reduce customer attrition and associated costs.

Outline

- Business and Data Understanding
- Data Exploration
- Modelling
- Evaluation
- Recommendations
- Next Steps

Business and Data Understanding

Business Context:

- The telecom industry is highly competitive.
- Customer retention is crucial for maintaining revenue.
- Predicting churn helps SyriaTel proactively address potential losses.

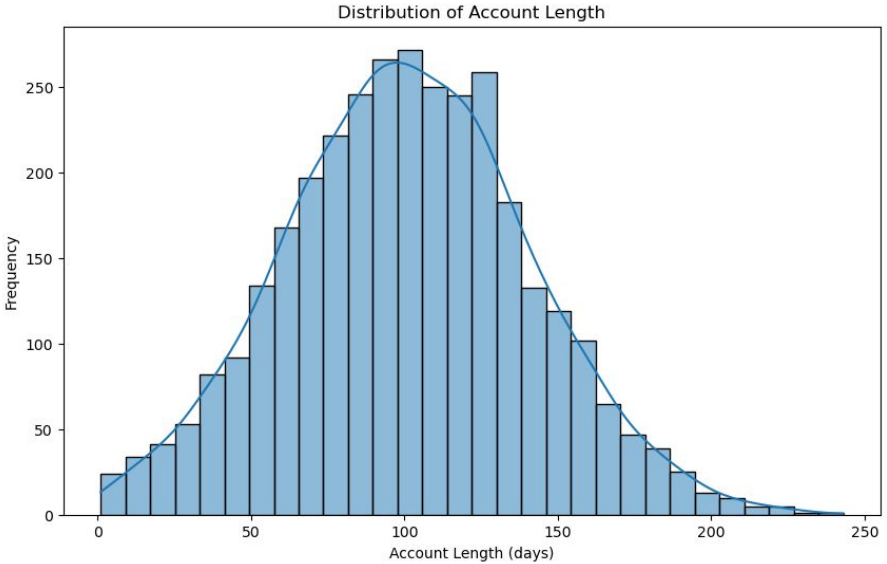
Data Overview:

- The dataset consists of **3,333 entries** and **21 columns**, with a **14.5% churn rate**, indicating class imbalance.
- Key features include account length, total day minutes, and customer service calls.
- There is need for data-driven insights on customer churn rate.

Data Exploration

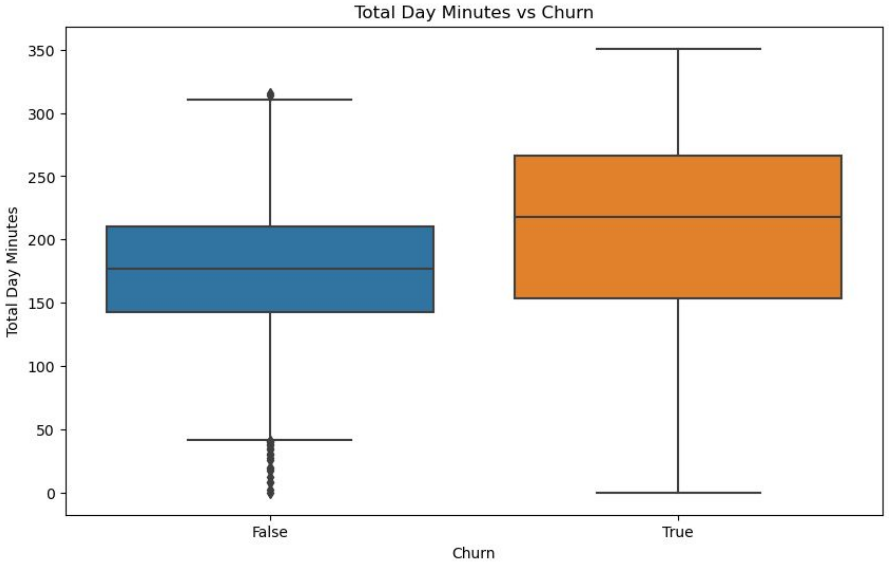
1. **Account Length**

- Customer tenure with SyriaTel varies widely.
- This variation suggests that customer loyalty or the likelihood of churn may be influenced by how long a customer has been with the company.



Data Exploration

2. **Total Day Minutes**
- Customers who churn tend to have a wider range of total day minutes.
- This suggests that higher or more variable usage during the day might be associated with a higher likelihood of churn.



Modelling

Why classification?

- Classification models help us predict whether a customer will stay with SyriaTel or leave (churn). Customers are sorted into two groups: those who are likely to stick around and those who might be at risk of leaving.
- Different models help to see which one can best predict customer behavior. This helps us understand which customers need attention and what actions SyriaTel can take to keep them.

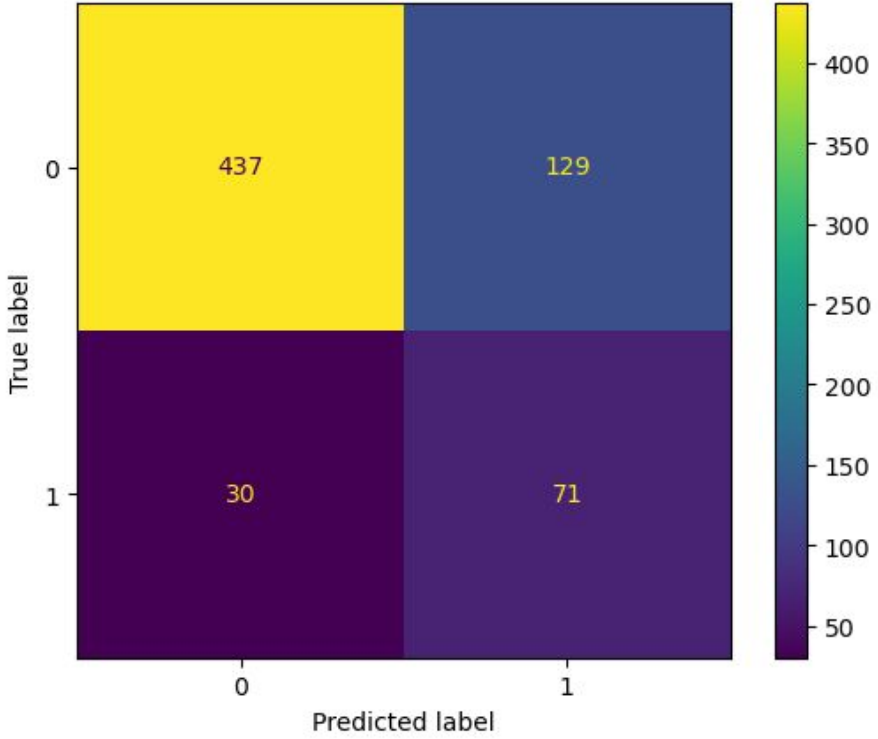
Modelling

- There are 3 models:

#	Model	Description
1	Baseline model	- Logistic Regression with an accuracy of **76.5%** on test data. The model performs better at predicting non-churners (class 0) than churners (class 1)
2	Improved models	- Decision Tree Classifier with accuracy rates of **93.4%** - Random Forest Classifier with accuracy rates of **87.4%,** respectively. Hyperparameter tuning further improved this model to an accuracy of **92.5%**

Model 1: Logistic Regression Model

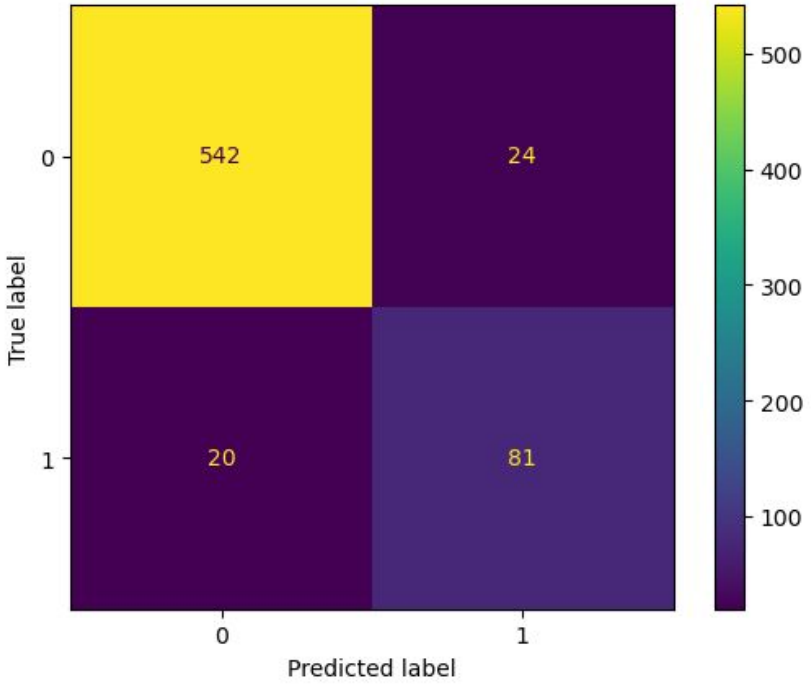
- Logistic regression had an accuracy of 76% on training data and 74.7% on testing data.
- The model performed consistently on training and test data, hence nosignificant signs of overfitting.
- However, cross-validation with 5 folds was done to improve accuracy.
- Training data accuracy remained at 76% and testing data dropped to 73.6%.



Confusion matrix after cross-validation

Model 2: Decision Tree Model

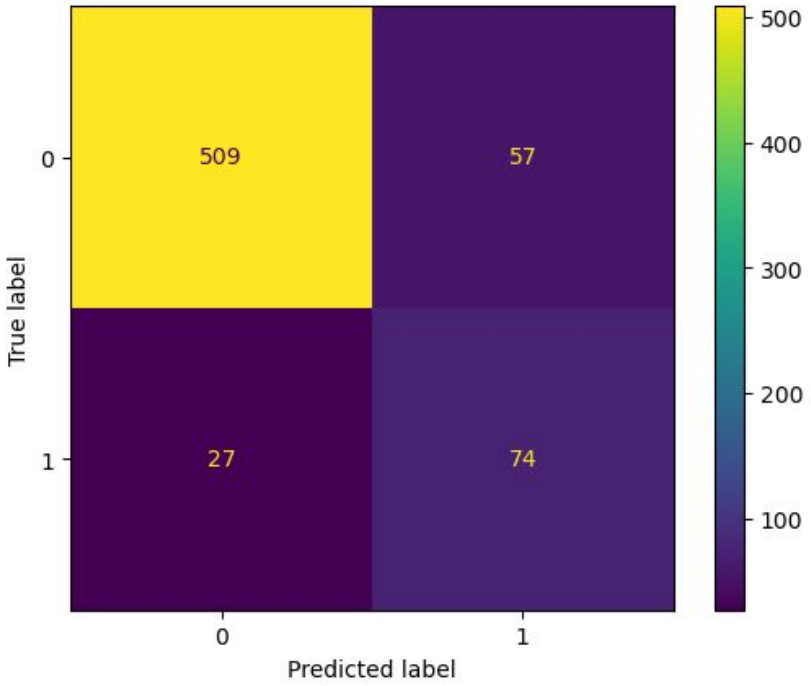
- The decision tree classifier model has an 89.2% accuracy level on training data and 88% on testing data in predicting customer churn rate.
- From the confusion matrix, it can be seen that the model reflects improved prediction capability, particularly in reducing misclassifications.



Decision tree confusion matrix

Model 3: Random Forest Model

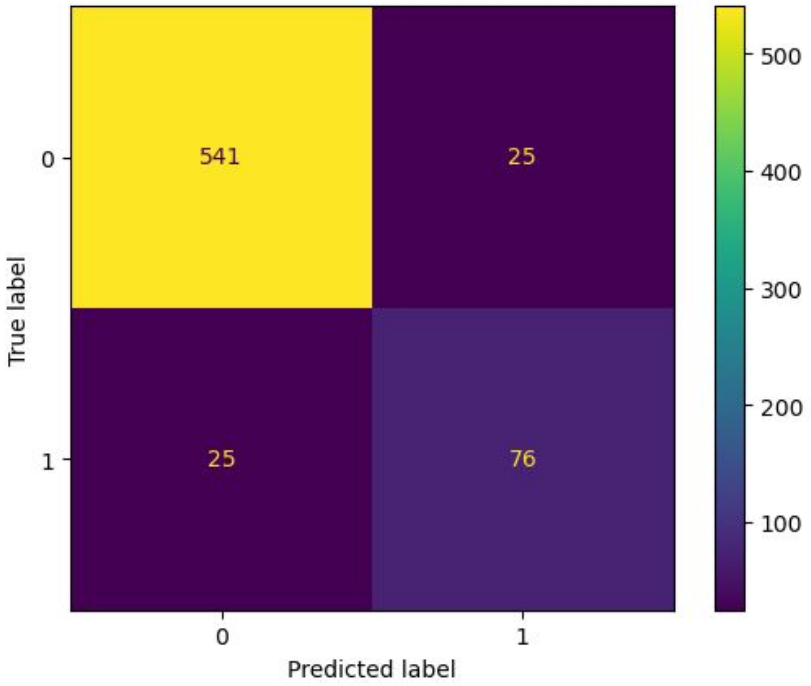
- The random forest classifier model has an 87.1% accuracy level on training data and 87.4% on testing data in predicting customer churn rate.



Random forest confusion matrix

Hyperparameter Tuning

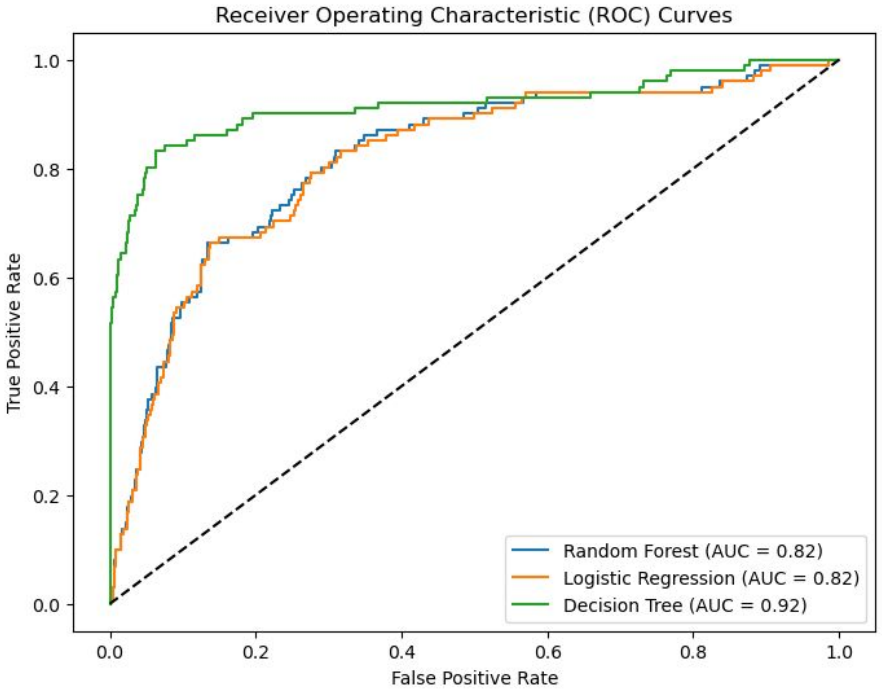
- After tuning the parameters for the random forest classifier using grid search our model improved on its perfomance as explained below:
- The accuracy of the random forest model is now at 92.5%.
- Trained data accuracy - 93.5%.
- Test data accuracy - 85.4%.



Random forest confusion matrix after tuning

Evaluation

- The decision tree model curve has the highest area under the curve (AUC) score (0.92), followed by the random forest and logistic regression curves (0.82).
- This shows that it has the best overall ability to predict customer churn.



Random forest confusion matrix after tuning

Recommendations

- With the decision tree mode, SyriaTel can quickly identify which customers are likely to leave and take specific actions to retain them, such as offering special deals or improving service quality.
- That way, the company can direct their interventions to focus on at-risk customers, which saves money and increases customer satisfaction.
- The model provides clear insights into customer behavior, allowing for more informed decisions that align with business goals.
- However, SyriaTel should regularly assess the model's performance, fine-tune parameters, and retrain the models to ensure optimal accuracy and effectiveness in predicting customer churn.

Next Steps

- **Deploy the model:** Integrate the Decision Tree model into SyriaTel's customer service system to start predicting churn in real-time.
- **Monitor and improve:** Regularly check how the model is performing and update it with new customer data to keep it accurate and relevant.
- **Customer engagement:** Use the insights from the model to develop new strategies for customer engagement, improving overall customer retention and satisfaction.

Thank You!

Email: sarangemanono@gmail.com
GitHub: @SarangeManono
LinkedIn: linkedin.com/in/sylviasmanono