



Machine Learning
Classification Model to
Predict Customer Churn

Recommendations to
SyriaTel on
Customers' features
that contribute to
Churn

Contents



- Industry Overview
- Business and Data Understanding
- Modelling
- Evaluation
- Recommendations
- Next Steps
- Thank you !

Industry Overview



- This project examines a dataset from SyriaTel, a telecommunications company with the aim of using classification prediction modelling, to predict the likelihood of a customer ceasing to do business with the company within a given period (churn)
- Acquiring new customers involves huge marketing costs, that include huge advertising budgets and commissions to sales agents. It therefore becomes imperative to retain those customers once they are acquired.
- It is the motivation of every telco company to understand the features or characteristics of a customer who is likely to 'churn'.
- With this understanding, the company can get ahead of the problem and develop initiatives that target these specific customers to discourage them from ceasing doing business with the company.

Business and Data Understanding

The SyriaTel data set consist of straightforward and well formatted data and is available on <https://www.kaggle.com/> It has customer usage (minutes/number of calls), customer choices of premium services columns, as well as other customer features like location, area code, phone number and account length <https://www.kaggle.com/datasets/becksdff/churn-in-telecoms-dataset> .

SyriaTel is intentional about reducing the high cost of churn. With this Model, their Data Analytics, Marketing and Revenue Assurance departments will be able to:-

- Identify the customer characteristics (features) that have the most impact on churn.
- Use the best model identified from the modelling iterative process to test future customers' data to predict the likelihood of a customer leaving the network.
- Develop proactive retention strategies specifically targeted to 'at risk' customers to discourage them from leaving the network.

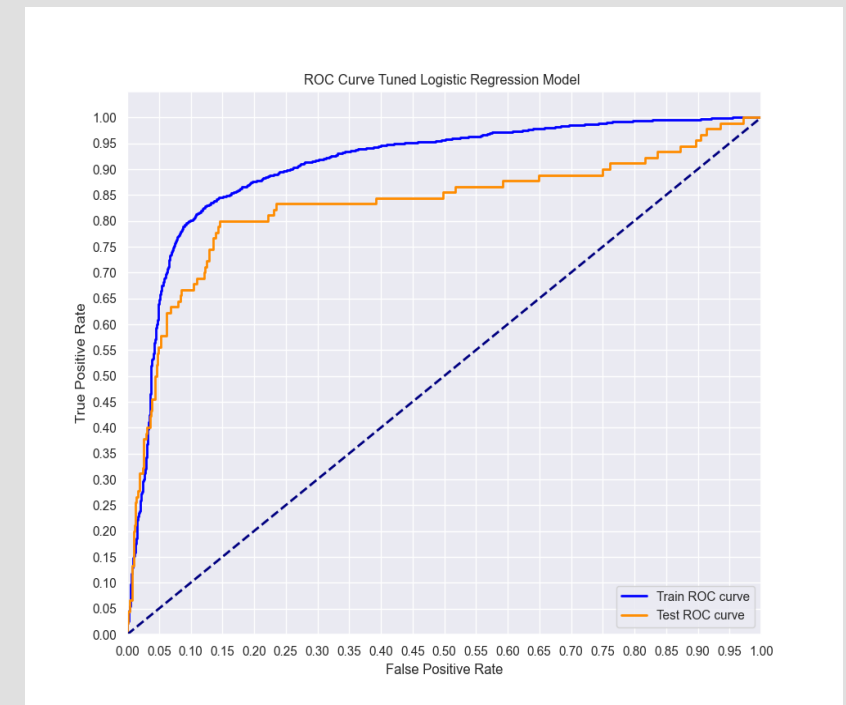
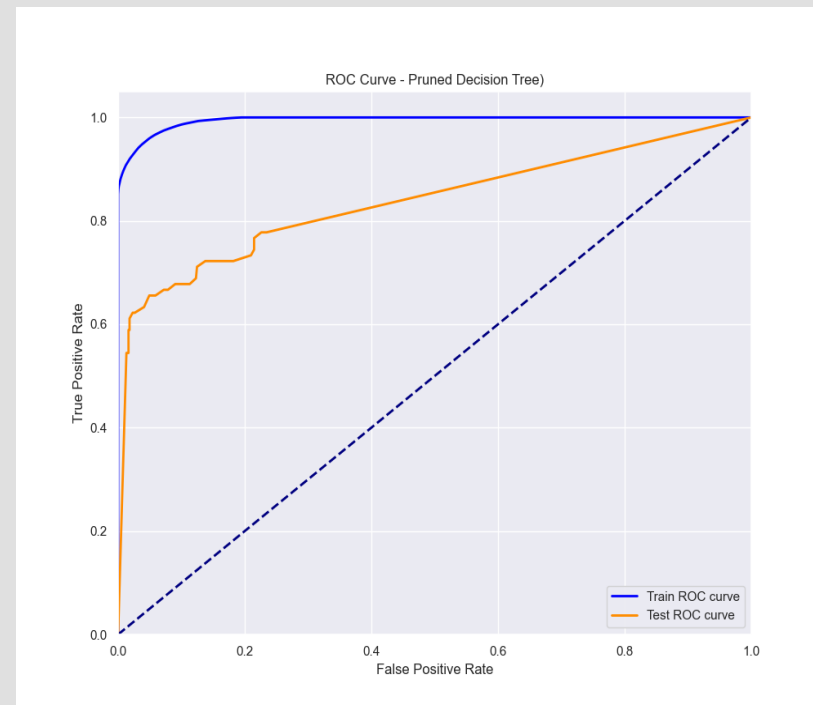
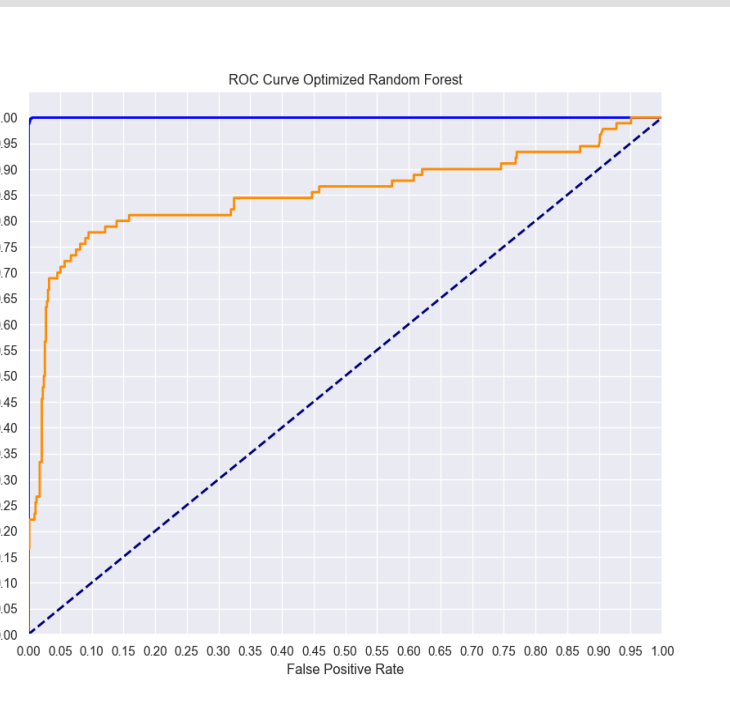
The Modeling Process

- Three machine learning algorithms, best suited for a classification problem like SyriaTel's were trained on the dataset provided. These models are Logistic Regression, Decision Tree and Random Forest.
- An iterative process was used on the split dataset with distinct training and test sets. Each of the model is trained on the training set and tested on the test set.
- Starting with a baseline model logistic model, the models are systematically optimized with machine learning techniques suitable for each model.
- In every step of the iterative modelling process, the following evaluation metrics are used to gauge the performance of each model:-
 - Accuracy – This metric measures how often the model gets the prediction right
 - Area Under the Curve(AUC) – This metric measures the ability of the model to distinguish between customers who churn and those who don't.
 - ROC (Receiver Operating Characteristic) Curve is a graphical representation used to evaluate the performance of a binary classification model.
- The model with the best metrics at the end of the iterative process will be recommended for deployment on SyriaTel's customers' data to predict the likelihood of churn.

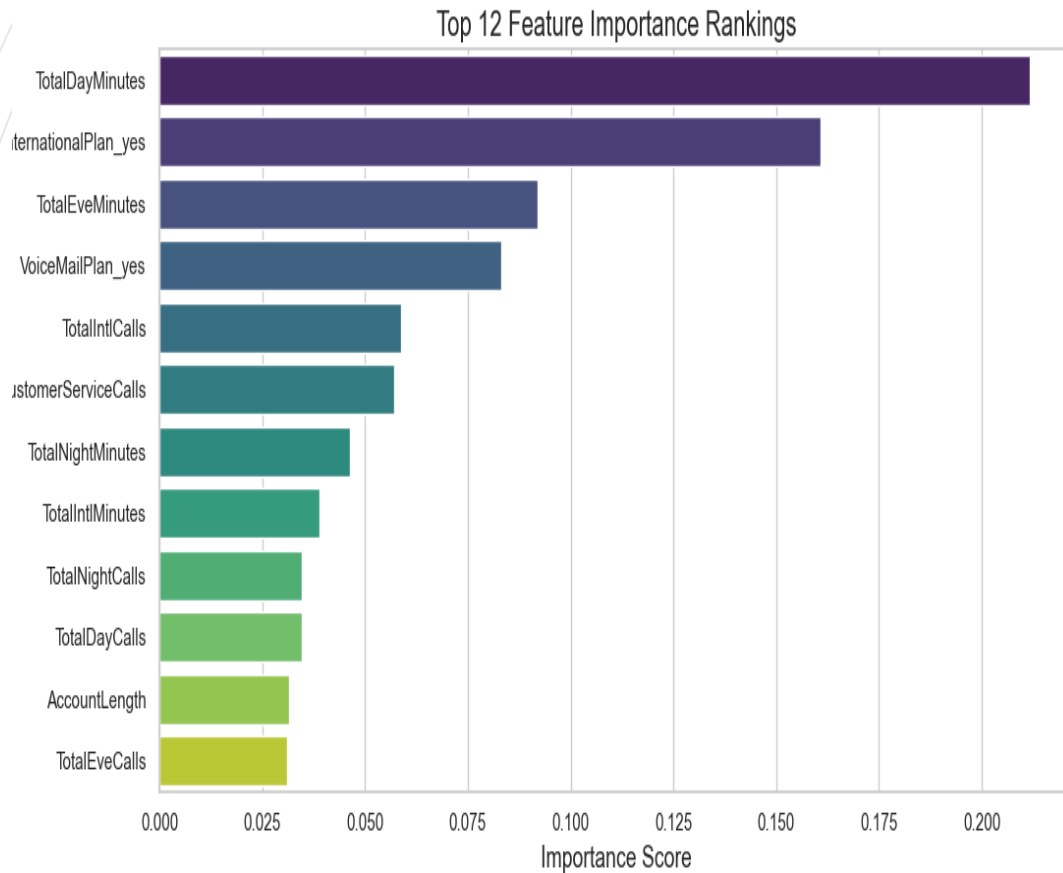
Model Evaluation

Metric	Best Logistic Model	Best Decision Tree Model	Best Random Forest Model
Train Accuracy	0.85	0.96	1.00
Test Accuracy	0.84	0.90	0.92
Train AUC	0.91	0.99	1.00
Test AUC	0.83	0.84	0.87

- A test accuracy score of 92% from the best Random Forest Model means that our model was able to predict on unseen data correctly 92% of the time.
- The test AUC of 0.87 means that there is an 87% chance that the model will correctly rank a randomly chosen churner higher than a randomly chosen non-churner.
- From the metrics the **Random Forest Model** is the best model and will be deployed to predict customer churn on SyriaTel's customer data.



Recommendations



- The performance of the 3 models is indicative that the features in the SyriaTel data set have good predictive power.
- Customers with high usage as indicated by Total Day Minutes and Total Evening Minutes are at a higher risk of churn. The Usage patterns during the day give the highest predictive power.
- Premium customers indicated by enrolment into the International Call Plan also has significant impact on churn. Voice Mail Plan enrolment also has a strong impact.
- Customer Service Calls also have an impact on churn, meaning a customer calling customer service more is more likely to churn.
- The frequency as indicated by the number of calls also has an impact, though to a lower extent than the actual time (minutes) spent on the call.

Next Steps



- **Model Improvement:** the SyriaTel will update the model with new data to maintain accuracy and adapt to any changes in customer behavior. These updates may result in a new choice of model selection.
- **Proactive Engagement:** Regularly check in with high-usage customers and premium customers (those enrolled into International and Voice Mail plans) to address any potential issues and enhance their experience.
- **Personalized Offers:** Provide targeted offers and discounts to high-risk customers based on their usage patterns and preferences.
- **Customer Service Improvement:** Analyze customer service interactions to identify common pain points and address them promptly.



Thank You !



Wambui Munene

wambui@icfoglobal.com