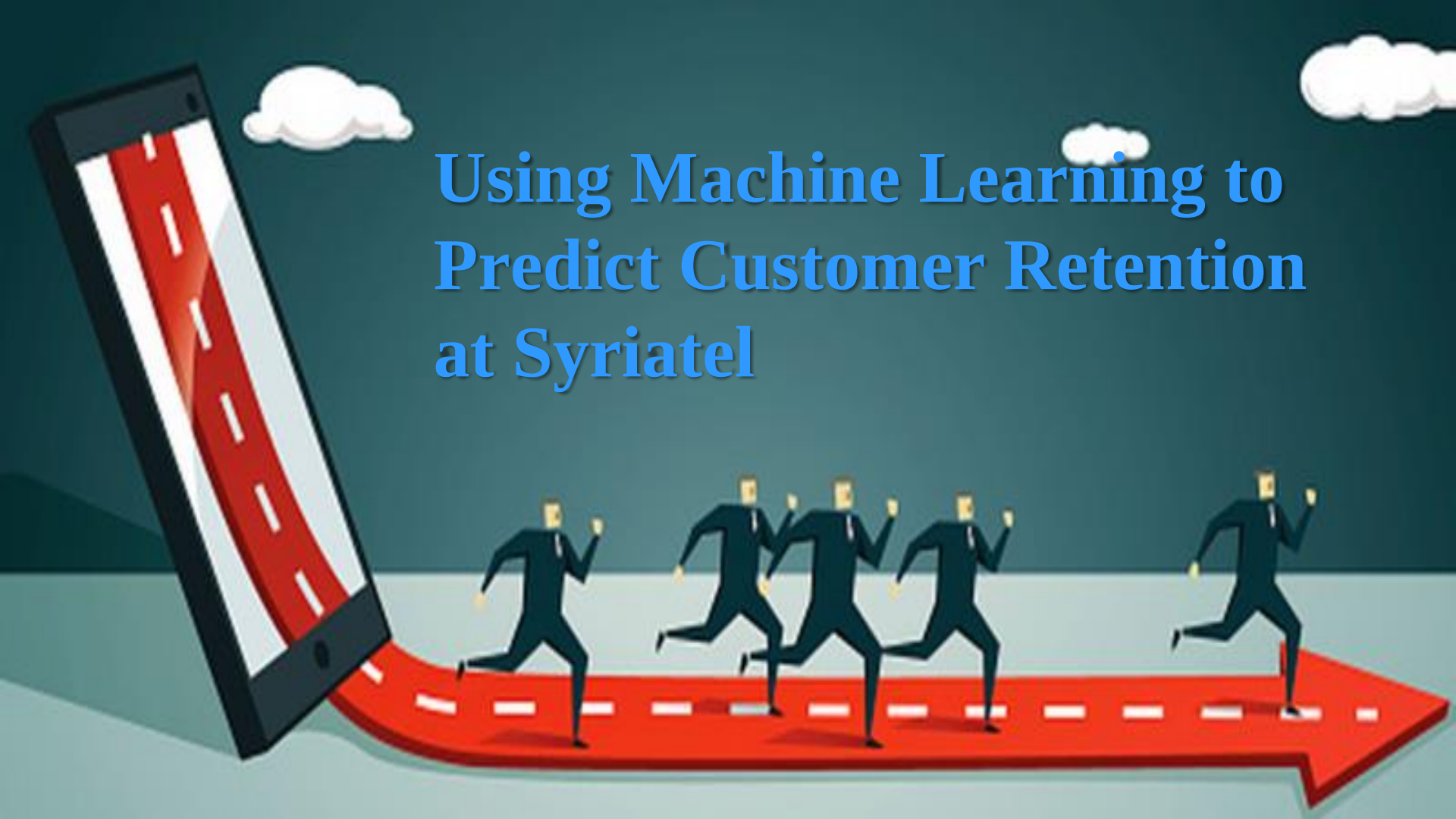
Using Machine Learning to Predict Customer Retention at Syriatel



Prepared by:

- Edward Opollo
- Cynthia Nasimiyu
- John Karanja
- Sheilah Machaha
- Julius Charles
- Sharon Kimutai
- Phelix Okumu

Overview

It costs from 5 to 25 times to acquire a new customer than retain an existing one while retaining existing clients by 5% results in profits augmenting by 25% to 95% - ‘*Ame Gello’*

Business Understanding

The telecommunication industry has become a very competitive one, particularly with emerging technological innovations which has given customers alternative communication channels.

We aim at creating a predictive business model which would enable Syriatel adapt strategies that would reduce churn, maintain and grow it’s customer base ,in a bid to sustaining overall growth and profitability.

Business Questions

- What are the most significant predictors of customer churn for Syriatel Mobile Telecom?

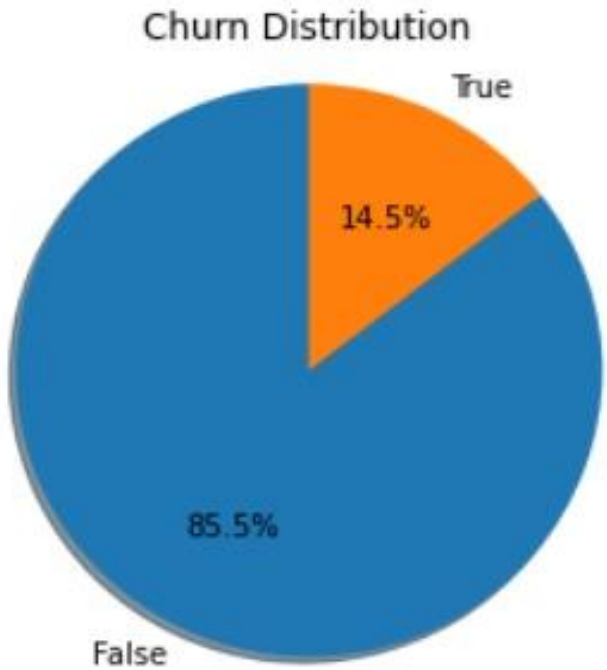
- Which Machine Learning Model is the most suitable in predicting Customer Churn?

- What strategies can Syriatel Mobile Telecom implement to retain customers and reduce churn rates?

Data Understanding

- The telecom dataset contains information about customer activity in relation to churn.
- It contains a total of 21 columns and 3333 rows. Each row having a unique identifier as the telephone number.
- It contains both numerical and categorical values.
- The data is basically a binary classification problem.
- There are no missing values and duplicates.

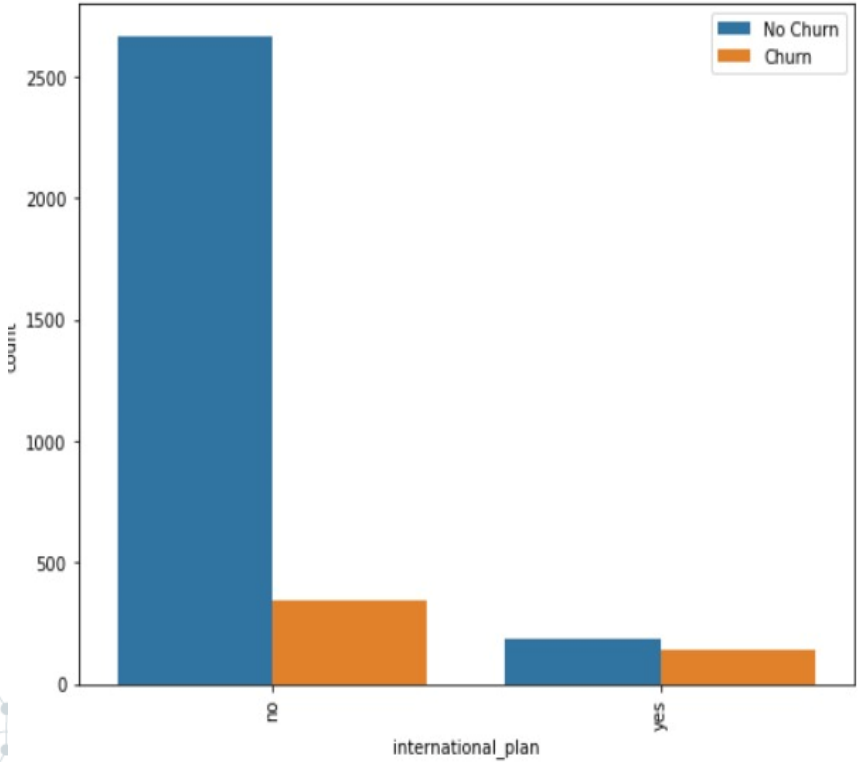
Exploratory Data Analysis

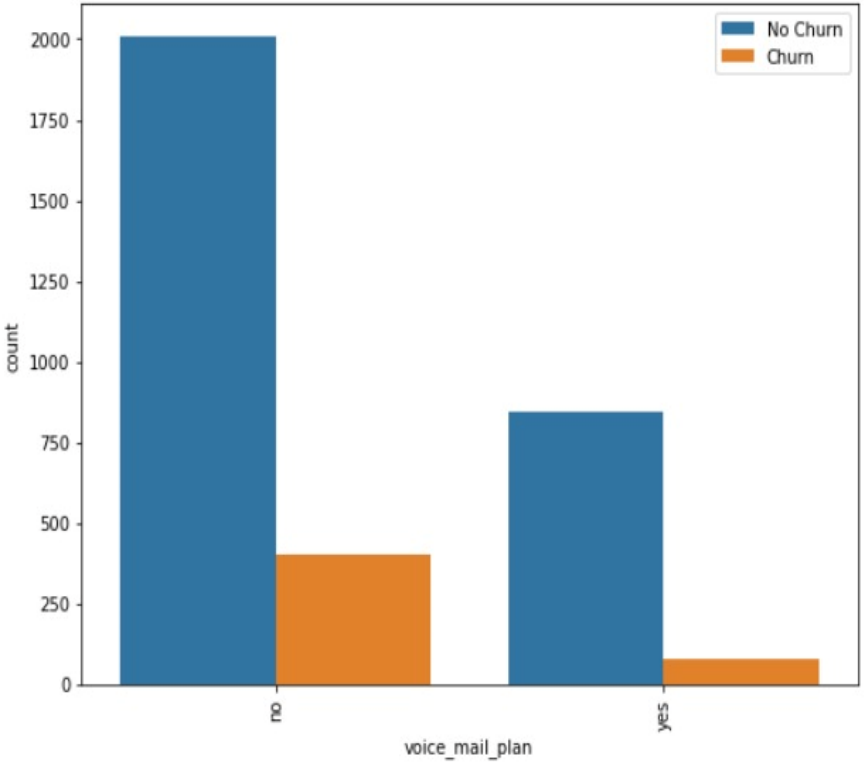


Of the 3,333 customers in the dataset, 483 have terminated their contract with the Telecom firm. That is 14.5% of customers lost.

- True = churned
- False = not churned

Bivariate Analysis



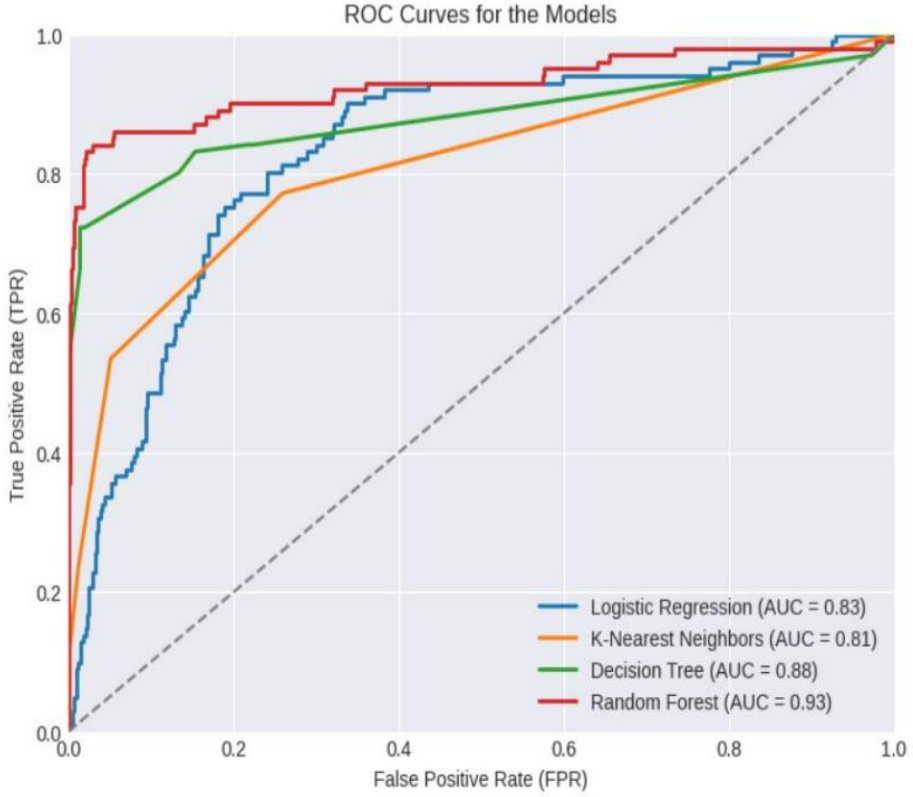


Modelling

Four different models were built to predict Churn and further improvements were made to achieve the best predictive results. They are:

- Logistic Regression Model
- K- Nearest Neighbors
- Decision Tree Classifier
- Random Forest Classifier

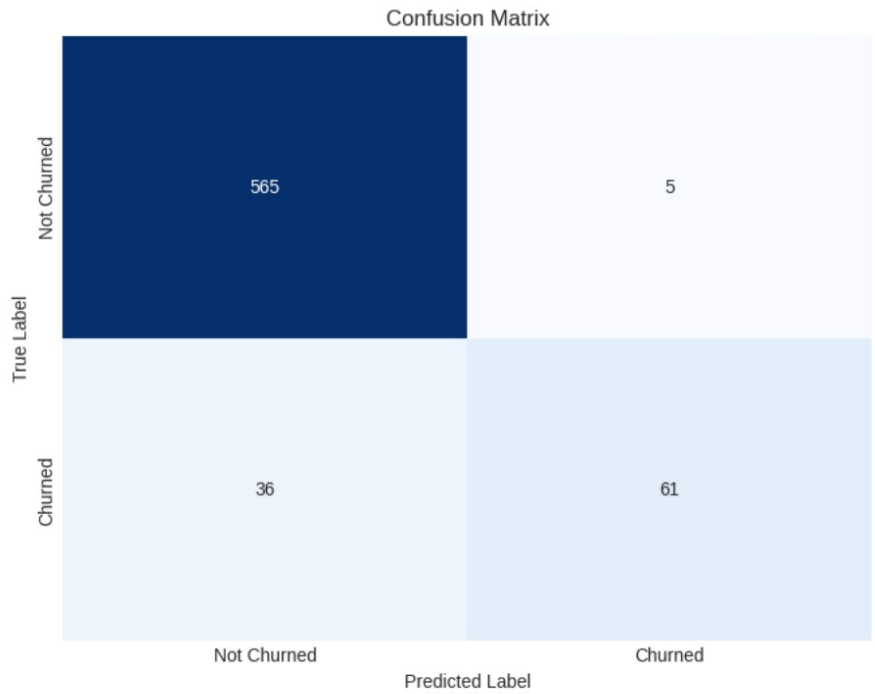
Model Performance



The Best Model is Random Forest Classifier

	Model	Accuracy (Test Set)	F1 Score (Test Set)	Recall (Test Set)	Precision (Test Set)
0	Logistic Regression	0.780000	0.510000	0.770000	0.390000
1	K-Nearest Neighbors Classifier	0.880000	0.410000	0.280000	0.800000
2	Decision Trees Classifier	0.950000	0.810000	0.730000	0.900000
3	Random Forest Classifier	0.940000	0.800000	0.690000	0.960000

Model Selected – Random Forest Classifier (reduced estimators)



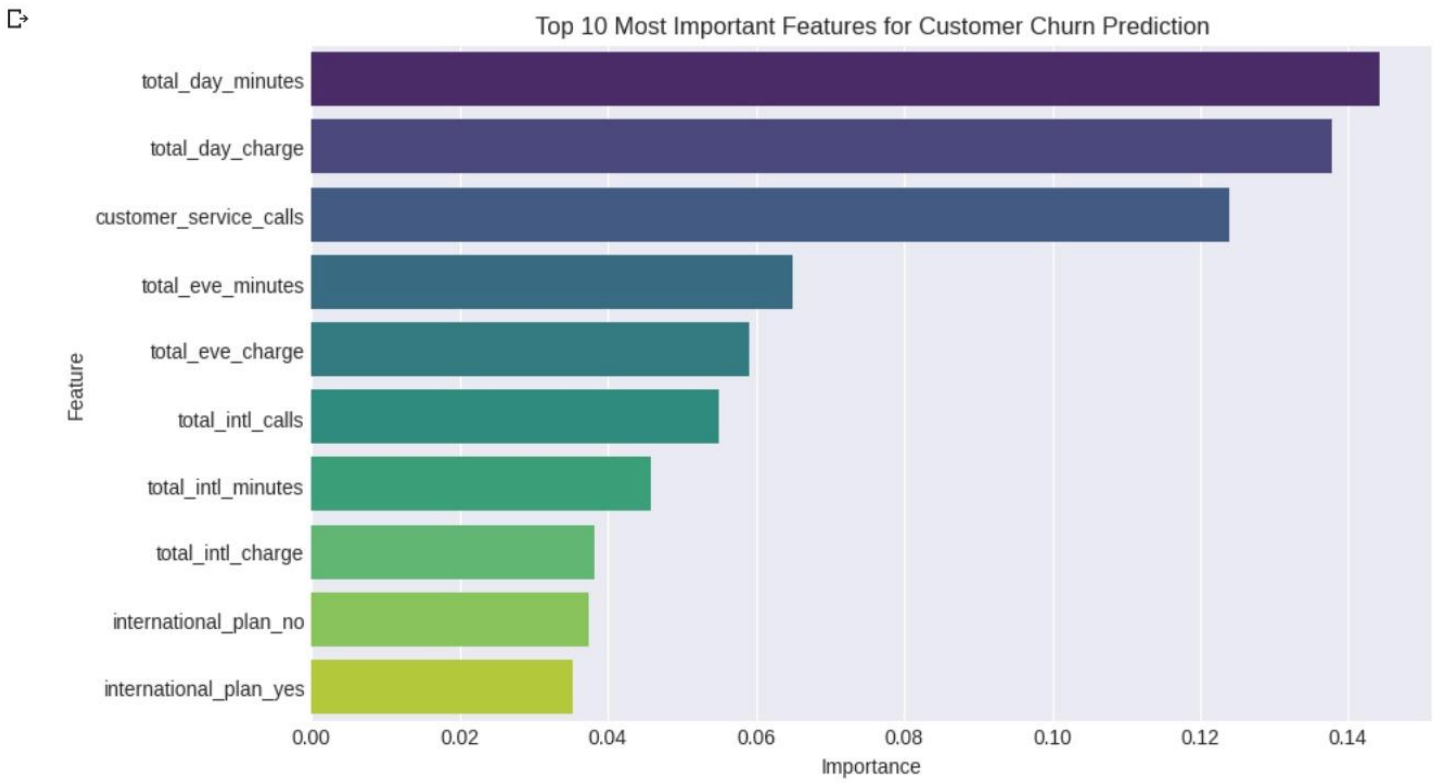
The model achieved :

- Accuracy : 95%
- Precision: 92%
- Recall: 62%
- F1 score: 75%
- Train score: 98%
- Test score: 94%

Model Interpretation

- The model's accuracy, reflecting its ability to correctly predict churn, is 95%
- Precision measures the accuracy of positive predictions, and a precision of 92% suggests that when the model predicts a positive outcome, it is correct 92% of the time
- Recall, (sensitivity), represents the model's ability to identify actual positive instances, this indicates that the model captures 62% of the total positive instances.
- The F1 score, which combines precision and recall, is 75%, indicating a balanced performance between precision and recall for the positive predictions made by the model.
- The model does not overfit as demonstrating its ability to generalize and make accurate predictions on new and unseen data.

Feature Importance



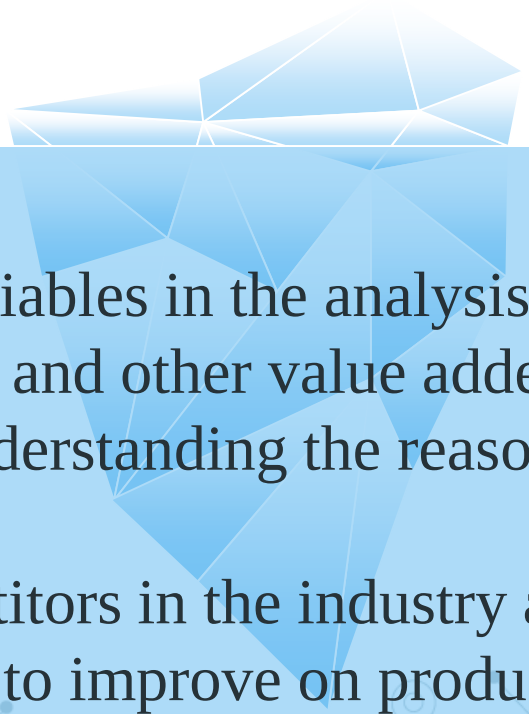
Recommendations

- Syriatel make use of the Random Forest Classifier as the primary model for predicting customer churn. This model has a higher ROC curve and strong overall performance.

- Customer Retention strategy that addresses key features in relation to call minutes and charges. These efforts could include personalized offers or discounts on day charges.

- We would recommend, that Syriatel comes up with strategies to reduce on Customer Service calls, as this is among the top features that would likely lead to Customer Churn. Example: come up IVR (Interactive Voice Response) and DIY (Do It Yourself) solutions.

Next Steps



- Exploration of more variables in the analysis incorporating features such as internet bundles and other value added services
- Further analysis into understanding the reasons for customers churning.
- Investigate what competitors in the industry are offering and come up with business strategies to improve on products and service offering.

Thanks!

Any questions?

References:

https://hbr.org/2014/10/the-value-of-keeping-the-right-customers

https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset

GitHub: https://github.com/charles4data/DSF-Phase-3-Project-Predicting-Customer-Churn-at-SyriaTel-with-Machine-Learning