

A REPORT ON THE DEVELOPING A CLASSIFIER THAT PREDICTS WHICH CUSTOMERS ARE LIKELY TO CHURN SO THAT SYRIATEL CAN TAKE APPROPRIATE ACTIONS AND REDUCE CUSTOMER ATTRITION.

1. BUSINESS UNDERSTANDING

1.1 BUSINESS OVERVIEW

SyriaTel is a telecommunication company based in Syria. The services of the company include voice and data services. Recently, the company has been concerned about the increased rate of customer churn that is resulting in high loss of revenue. Customer churn is when a customer discontinues their relationship with a company, which can result in lost revenue and reduced profits.

The company is looking to outsource a data scientist to help identify the contributing factors that are leading to customer's opting out on the services.

A data scientist who will leverage cutting-edge technology to develop a binary classifier that accurately predicts whether a customer is likely to churn or not. By analyzing a wide range of customer data including demographics, usage patterns, and service interactions, we will identify patterns associated with churn and develop strategies to mitigate these issues.

The success of the project will be determined by the accuracy of the classifier and the effectiveness of the strategies employed. Ultimately, efforts will translate into outcomes for SyriaTel and its customers.

1.2 BUSINESS OBJECTIVE

- **Main Objective**

The primary goal of this project is to identify the factors that contribute to customer churn and Develop a classifier that predicts which customers are likely to churn so that SyriaTel can take appropriate actions and reduce customer attrition.

- **specific objectives**

1. Conducting a comprehensive analysis of SyriaTel's customer data to identify patterns and trends that contribute to customer churn.
2. Determining which variables have the highest impact on customer churn in SyriaTel's customer base.
3. Building and testing a predictive model to accurately forecast the likelihood of customer churn.
4. Evaluating the performance of the predictive model and comparing it with other alternative models.
5. Identifying preventive measures that SyriaTel can take to reduce customer churn and retain more customers.
6. Developing a plan to implement the preventive measures based on the insights gained from the predictive model.

1.3 BUSINESS CRITERIA

Being able to develop a binary classifier that accurately predicts whether a customer is likely to churn or not and By analyzing a wide range of customer data including demographics, usage patterns, and service interactions, we will identify patterns associated with churn and develop strategies to mitigate these issues.

1.4 ASSESSING THE SITUATION

1.1.4 RESOURCE INVENTORY

1.4.1.1. DATASETS

SyriaTel Customer ChurnLinks ,find the data in kaggle as Churn in Telecom's dataset.

1.4.1.2 SOFTWARE USED

- Jupyter Notebook
- Git
- Github
- Numpy
- Pandas
- scikit learn

1.4.2 ASSUMPTIONS

- data provided is correct and upto date

1.4.3 CONSTRAINTS

In []:

2. DATA UNDERSTANDING

2.1 DATA UNDERSTANDING OVERVIEW

For this project, we are using the available dataset in Kaggle. The dataset is the SyriaTel Customer Churn link: <https://www.kaggle.com/datasets/beckssddf/churn-in-telecoms-dataset>
(<https://www.kaggle.com/datasets/beckssddf/churn-in-telecoms-dataset>)

2.2 DATA DESCRIPTION

We have one dataset available for this project.

We have no missing values

Data has both continuous and categorical features comprising of the following data types; objects, integers, float and booleans

we have a total of 3333 entries and 21 columns

3. DATA PREPARATION

This are the steps in preparing data

3.1 LOADING DATA

- load data from the excel sheets downloaded from kaggle as a csv file

3.2 CLEANING DATA

- Convert categorical data to
- remove irrelevant columns

3.3 ANALYSIS

We performed various analyses in our dataset such as Exploratory Data Analysis(Univariate Analysis, Bivariate Analysis) and also carried out data visualization

4. MODELING

4.1 PREPROCESS DATA

- Split and train data into(X_train, X_test, y_train, y_test)
- label encode categorical data
- normalize the numeric data

4.2 BUILDING AND EVALUATING MODELS

We decided to work with the following binary classifiers:

- KNN(K-Nearest Neighbours)
- Logistic Regression
- Decision Trees
- Random Forest

5. CONCLUSION

- Four models were explored to determine the best model for predicting customer churn.
- False Negatives would be costly, therefore we optimized for Recall.
- The best classifier was the Random Forest Classifier with a Recall of 0.75.
- 7.4% of the predictions are False Negatives which means that the model will 7.4% of the time predict that a customer will not churn yet the customer churns. The model can be optimized further to reduce the false negatives to below 2%.

6. RECOMMENDATIONS

- Determine the unique needs of the following customers and meet them;
 1. Heavy daytime callers. Come up with tariff incentives for them
 2. Customers with international plans. Come up with unique retention plans
 3. Customers who frequently call customer service. Assist customers proactively to reduce the need to call.

In []: