SYRIATEL CUSTOMER CHURN PREDICTION STUDY

Analysis on best prediction methods of customer churn

Overview

Background :

The telecommunications industry in the US has evolved drastically over the years,shaping the ations communications industry.

With the join of SyriaTel ,they improved services ranging from traditional landline telephone to broadband internet, cable television and mobile services

Today, the U.S. Telecommunications industry continues to be dynamic, with ongoing advancements in 5G technology, fiber-optic networks, and the convergence of services hence improving their communication sector.

OBJECTIVES :

1.)To investigate the relationship between the feature variables and the target variable.

2.)To check the relationship between numerical features.

3.)To investigate each feature and check for patterns

4.)To create a precise model that will be used to predict customer churn depending on a range of features.

DATA UNDERSTANDING

1. We first import the necessary libraries
2. Read the .csv file dataset - The dataset contains customers of SyriaTel and their information including whether they have churned or not. It contains 21 columns with 3333 entries
3. We then check for outliers using even boxplots
4. We then perform Exploratory Data Analysis on the dataset-(Univariate,Bivariate and Multivariate Analysis.)
5. After performing EDA , we found that there is no independence between these features and in turn it means we wiil not use Logistic Regression or any Naive Bayes model which assume independence of features.

MODELLING

We first do data preprocessing before the modelling ,this is done by creating functions that map the values

First create a function that brings all the classification metrics

We then start the modelling with 1.)decision trees

2.)K-Nearest Neighbors(KNN)

3.)Discriminant Analysis

4.)Random Forests

5.)XGBoosts

LIMITATIONS

1.)The classes were imbalanced. in that the `False` class was way more than the `True` class. This is why we had to use the SMOTE technique to mitigate the issue.

RECOMMENDATIONS :

From the results, we can see that all the models have a recall of more than 70% save for the KNN algorithm. However, this is a decrease from the results we achieved when we created the other models. Even with these low results we can see that the XGBoost algorithm gives more than 80% recall score which is a good sign and one that can be presented to the SyriaTel stakeholders. This is why this model will be chosen to predict customer churn.