MiTH

Predicting the customer churn in a bank

Problem Statement:

For any customer centric organization, customer churn/retention analysis is going to be extremely valuable. It is a market reality that the banks and other financial institutions are battling to attract each other's customers while retaining their own. Thus, Customer churn reduction is the central concern of most banking organizations as switching costs to the customer are low and acquisition cost to the organization is high. Churn reduces profitability as it means potential loss of future revenue and also losing the invested costs of acquisition. On the other hand, long term customers contribute to profitability through mere continued association as they are less costly to serve and through additional services and referrals. It is important to have data that underlies construction of purchase decision and also the underlying loyalty hooks. So a good deal of marketing budget is allocated to engage customer on an ongoing basis and prevent churn by designing new plans and offering freebees, discounts etc. Thus, it is important to predict the customers who are likely to churn and develop appropriate modalities for retention.
An international bank has a business problem with customer retention and observed high churn rates. The bank observed that alarming number of existing customers exit while it is getting increasingly difficult to acquire new customers. So they want to understand the hidden patterns in their customer behaviour by use of customer data which has early warning signs.This helps understand the drivers behind considering an alternative. The variables could be demographic, perception or behavioural.

You are expected to create an analytical and modelling framework to predict the customer churn in a bank based on the quantitative and qualitative features provided in the datasets.

Data:

1. Train.xlsx (to train the model)
2. Test.xlsx (to predict the Churn)

Objective:

You are expected to create an analytical and modeling framework to predict the customer churn (Churn: Yes/No) based on the quantitative and qualitative features provided in the data.

Main Tasks:

1. Exploratory Data Analysis using visualizations in R Notebook or Jupyter notebook format (Use only Train data for this task)
2. You are expected to build a framework that predicts whether a customer churns or not from the company.
3. You are expected to include the classification_report/confusion matrix on validation data in your code notebook.
4. Viva

Evaluation Metric: Recall for the level "Yes" on target attribute "Churn" with a minimum accuracy of 60%

Other Instructions:

1. Spend enough time on pre-processing and data understanding. Think of the problem from domain's perspective to build an efficient model.
2. Your final grader score carries much lower weightage than your overall approach which includes data exploration and model validation. Use your time wisely.