Bank Customer Churn

Prepared by Abdullah Al Imran (18201103)
CSE Fundamentals Review Bootcamp (BUET CSE Fest 2021)
Department of CSE
BUET, Dhaka, Bangladesh
+880 1716-146019
abdullah.al.imran@gmail.com
GitHub: alimranabdullah
LinkedIn: alimranabdullah

Introduction

Customer churn, also known as customer attrition, is a metric used to measure how many customers are leaving a business over a given period of time. In the context of business, churning means the loss of revenue. It costs more to acquire a new customer than it costs to retain an existing customer. So, it is important to know which customers are likely to churn so that the company can take proactive steps to retain them. In this notebook, we will explore a dataset from a bank and try to predict which customers are likely to churn based on their profile and activity with the bank.

Data Description

In this dataset we have 10000 rows and 14 columns. The dataset is available at this link: https://www.kaggle.com/shrutibhargava/churn-data.

Dataset Level

Each row represents a customer. Each column contains attributes about that customer. Number of rows indicates that there are 10000 customers in the bank.

Column Description

- RowNumber: Just the row number.
- CustomerId: A unique identifier for each customer. This is not a feature. It seems highly unlikely that a customer's ID would affect their churning. So, we will not include this column in our analysis.
- Surname: Customer's surname. This is also not a feature. It seems highly unlikely that a customer's surname would affect their churning. So, we will not include this column in our analysis.
- CreditScore: Customer's credit score. This is a feature. It seems likely that a customer's credit score affects their churning.
- Geography: Customer's country. This is a feature. It seems likely that a customer's country affects their churning.
- Gender: Customer's gender. This is a feature. It seems likely that a customer's gender affects their churning.
- Age: Customer's age. This is a feature. It seems likely that a customer's age affects their churning.
- Tenure: Number of years the customer has been with the bank.
- Balance: Customer's account balance.
- NumOfProducts: Number of products that a customer has purchased through the bank.
- HasCrCard: Whether a customer has a credit card or not.
- IsActiveMember: Whether a customer is an active member or not.
- EstimatedSalary: Customer's estimated salary.
- Exited: Whether a customer has churned or not. This is the target variable.

Preprocessing Steps

Since the dataset is already quite clean, we will only perform a few basic preprocessing steps. We will drop the columns that we think will not be useful for our analysis. We will also check for missing values and encode the categorical features. We will also check for outliers and see if we need to do something about them. We will also check the distribution of the target variable to see if we need to do something about it. We will also check the correlation between the features and the target variable to see if we can get some insights from the data. We will also check the correlation between the features to see if we can drop some of them. We will also check the distribution of the features to see if we need to do something about them. We will also check the distribution of the features with respect to the target variable to see if we can get some insights from the data