

# Bank Customer Churn

A report on predicting churn in bank customers

Momin Asadullah Khan

## **Introduction:**

Banks need to have satisfied customers for it to be successful and this report looks at bank customers and our ability to correctly predict if a bank customer will churn or not. From a business perspective, the bank is the client and they are concerned about potentially losing current customers. They want to be able to predict which customers are at risk of churning and be able to do outreach (in the form of promotions, etc) to convince them to retain their services with the bank.

## **Dataset:**

The dataset is of bank customers from the bank consisting of 14 feature columns and 1 output columns indicating if the customer has churned or not. The size of the dataset is 10000 rows x 15 columns and will be acquired from Kaggle in a csv format.

## **Methodology/Deliverables:**

The analysis will be done in Jupyter Notebooks with the first step being to load the data into a pandas dataframe and check out summary statistics, which show basic statistics, about the data. The next step will be data wrangling to make sure the data is in the correct format and to check for common problems like missing or null values, renaming columns. After making sure that the data is in the correct format, I will go ahead and perform exploratory data analysis to get a visual understanding of what the data looks like. I would be checking to see if there are any correlations in the dataset, exploring the distributions in the dataset and what the data is telling us visually. Once exploratory data analysis has been performed and we have looked at whether we can make inferences from the results, I will split the dataset into training and testing subsets. The model will be trained on the training dataset using different algorithms and cross validation while comparing which model performs best and why. I will analyze the classification report as well as look at the auc to see the validity of models. The model that performs best on the testing data will be chosen for the final analysis.

The deliverables for this project will include providing the Jupyter Notebook with all the code and analysis along with this report and a slide deck to present to the bank. In the presentation, I will provide answers to questions such as; what was the problem? What

steps were taken? Why they were taken? And, what is the end result? Is it helpful to us? What added value does it provide?

**References:**

<https://www.kaggle.com/barelydedicated/bank-customer-churn-modeling>