Title: Predicting Churn for Bank Customers

# **Bhargava Gaggainpali DSC680 - Applied Data Science** [bgaggainpali/bgaggainpali\_DSC680 (github.com)](https://github.com/bgaggainpali/bgaggainpali_DSC680)

Portfolio: [Bhargava Gaggainpali - Portfolio | Bhargava-Gaggainpali (bgaggainpali.github.io)](https://bgaggainpali.github.io/Bhargava-Gaggainpali/)

https://bgaggainpali.github.io/Bhargava-Gaggainpali/

# Which Domain?

In financial industry, banks are playing important role in challenging times like now, with COVID pandemic across the globe. People are losing jobs and financial institutions are facing more Customer churn in Bank Accounts. The increase in Customer Churn rate will result in significant financial loss to commercial banks. It is very critical for lending institutions like banks to have a prediction model to be able to predict customers churn to better serve the customers and reduce the churn.

# Which Data?

I have identified Churn\_Modelling.csv as source for my work, below is the Kaggle link. There are 10,000 observations in the dataset, each row in the dataset represents a Bank Customer Account. Given is the list of variables in the dataset.

Source File: <https://www.kaggle.com/adammaus/predicting-churn-for-bank-customers?select=Churn_Modelling.csv>

**Variable Description**

RowNumber Sequence Number

CustomerId Customer Account Number

Surname Customer Name

CreditScore Credit Score

Geography Location Country

Gender Male / Female

Age Customer Age

Tenure Period of time in Years as Customer to the Bank

Balance Balance amount in the Bank

NumOfProducts Number of Products availed by the Customer

HasCrCard Customer has Credit card

IsActiveMember Customer Active Member

EstimatedSalary Customer Salary

Exited Customer Churn value

# Research Questions? Benefits? Why analyze these data?

We could see many people close their bank accounts where they cannot maintain minimum balance in the account with the current pandemic Covid situation and it would impact financial budgeting for banks as it has limited assets which can handle the churn until certain extent only, so finding out the customers who are possibly to fall in the churn category will help banks to mitigate and balance the risk of the churn customers. In this paper I will try to predict to find, who are possibly to fall in as churn category and identify the customers to reduce this risk factor of bank to better serve the customers.

# What Method?

I see this problem as a classification issue, where we should try to understand and able to predict the customers, who might fall in to the category of ‘Bank Account Customer Churn’.

Planning to use supervised machine learning algorithm to work on the classification problem to be trained with algorithms like:

1. Logistic Regression

2. Decision Tree

3. Random Forest

Start with loading data into a data frame and then understand the data, then perform Exploratory Data Analysis (EDA) on the data set. EDA involves doing Univariate and Bivariate Analysis, identify missing values and outliers and fill the gaps with appropriate values. In the next step, building model with starting from logistic regression and observe the accuracy of the model. When the accuracy of the of the model is not high, then planning to use Decision Tree and Random Forest to achieve higher accuracy.

# Potential Issues?

Due to the low volume of the data that is available, I would assume model accuracy would be low. I am planning to use multiple algorithms, at least one would give expected accuracy in the range of 80-85. Also, I might encounter some data type errors while working with variables.

# Concluding Remarks

By taking a sample data set, with set of features, I would like to explore the features to find the pattern which majorly contribute the ‘Bank Account Customer Churn Prediction’ process and the order of importance. For this to achieve, will try to clean, validate, perform exploratory data analysis and also use the machine learning algorithm to build the predicting model. Based on the data and model efficiency, I would assume the accuracy of the model will be between 80%-85%.

**Reference:**

Soumya Sethuraman 2019 - Why Customers Leave & What Can Banks Do?

[Why Customers Leave & What Can Banks Do? | Tiger Analytics](https://www.tigeranalytics.com/blog/addressing-customer-churn-in-banking/)

Sina Esmaeilpour Charandabi - Kent State University - 2020 - Prediction of Customer Churn in Banking Industry

[(PDF) Prediction of Customer Churn in Banking Industry (researchgate.net)](https://www.researchgate.net/publication/342424673_Prediction_of_Customer_Churn_in_Banking_Industry)

Abbas Keramati, Hajar Ghaneei & Seyed Mohammad Mirmohammadi – 2016 -

Developing a prediction model for customer churn from electronic banking services using data mining

<https://jfin-swufe.springeropen.com/articles/10.1186/s40854-016-0029-6>

Diana Kaemingk - August 29, 2018 - Reducing customer churn for banks and financial institutions

[Reducing customer churn for banks and financial institutions // Qualtrics](https://www.qualtrics.com/blog/customer-churn-banking/)

Nelson Belém da Costa Rosa - November 2018 - Gauging and Foreseeing Customer Churn in the Banking Industry

[TGI0223.pdf (unl.pt)](https://run.unl.pt/bitstream/10362/71584/1/TGI0223.pdf)