Banking Customer Churn - Prediction

# **Vinay Nagaraj DSC680 - T302 Applied Data Science (2217-1)** <https://vinaynagaraj88.github.io/DataScience_Portfolio>

# Which Domain?

For my second project, I am planning to work in the Banking domain. Customers are the most important part of your business regardless of the industry. There would be no sales without customers and they are a critical factor when developing your marketing messaging and strategy. Customer Churn is the rate at which customers stop doing business with an entity. This is one of the most acknowledged problems in the banking sector and Banks are constantly looking at data/suggestions which could help them improve their customer service and retain their existing customers and also bring in new customers.

A customer’s banking relationship includes key journeys that range from onboarding and transacting to maintenance and problem resolution. Customers are central to a wave of new opportunities and challenges facing banking executives, with regulators increasingly expecting banks to deliver on more than just credit-risk management and associated capital requirements.

Banks often use customer churn analysis and customer churn rates as one of their key business metrics because the cost of retaining existing customers is far less than acquiring a new one. Customer churn prevention is one of the deciding factors when it comes to maximizing the revenues of any organization.

Through this project, I intend to focus on the behavior of bank customers who are more likely to leave the bank. I want to find out some striking behaviors of customers through Exploratory Data Analysis and later on use some of the predictive analytics techniques to determine the customers who are most likely to churn.

Below are several references I am planning to refer as part this project:

1. <https://rstudio-pubs-static.s3.amazonaws.com/565148_6e82a5c320f14869bf63e23bcf59ce9b.html>
2. <https://www.tigeranalytics.com/blog/addressing-customer-churn-in-banking/>
3. <https://www.researchgate.net/publication/342424673_Prediction_of_Customer_Churn_in_Banking_Industry>
4. <https://medium.com/@noah.fintech/creating-a-banking-customer-churn-model-1a2d0850f071>
5. <https://cmapspublic3.ihmc.us/rid=1MSXYH555-HXXB67-14YC/Tsai_Lu_Hybrid_Neural_Network_2009.pdf>
6. <https://www.semanticscholar.org/paper/A-Customer-Churn-Prediction-Model-in-Telecom-Using-Lu-Lin/66944ff8021d88cd646314fd3ba7a94d44d3de20>
7. <https://www.vttresearch.com/sites/default/files/julkaisut/muut/2006/customer_churn_case_study.pdf>
8. <https://arxiv.org/pdf/1912.11346.pdf>
9. <https://link.springer.com/chapter/10.1007/978-981-15-5243-4_12>
10. <https://ideas.repec.org/a/zna/indecs/v14y2016i2p116-124.html>

# Which Data?

I am planning to use the below dataset from Kaggle which contains details of a bank's customers and the target variable is a binary variable reflecting the fact whether the customer left the bank (closed his account) or he continues to be a customer. I am planning to use this data to train the model as part of this project.

Dataset Link - <https://www.kaggle.com/shrutimechlearn/churn-modelling>

The dataset contains 10,000 records with 13 attributes and one target variable. Demographically the information is about customers from Spain, France, and Germany. 55% of customers are Male, and 45% are Female, with an average age of 38.9 years.

Column Exited looks like the feature I will have to predict which will tell us whether the customer closed his account or not. So, feature “Exited” will mostly be my target variable.

# Research Questions? Benefits? Why analyze these data?

Below are some of the research questions I would like to find answers for as customer churn has become a big challenge for the banking sector and answering these research questions will help them understand the reason behind customer churn.

* Which are the most important features which leads to customer churn?
* Is there any correlation with age and customer churn?
* Is there any relation with gender and customer churn?
* Is there any relation with balance and customer churn?
* Is there any relation with Estimated Salary and customer churn?
* Which is the best prediction model to predict customer churn?
* What is the accuracy of the available classification models in predicting customer churn?

# What Method?

* I am planning on using Python language and work on Jupyter notebook for this project.
* I will initially clean the dataset for any missing values or outliers.
* Perform some EDA on the dataset to understand the trend of the data.
* I will perform Feature reduction/selection and then build or use some of the classification models to make predictions on employee attrition.

# Potential Issues?

My initial review of the data and the columns do provide some good insights and I don’t see any major issues or roadblocks at this point.

# Concluding Remarks

While banks are always on a lookout for new customers, it is very important to make sure to keep the existing customer base happy. Customer retention increases banks customers' lifetime value and boosts banks revenue. It also helps the bank to build amazing relationships with the customers. They trust the bank with their money based on the value they receive in exchange. Existing customers can drive repeat business, increase revenue, create brand ambassadors, defend against competition and gain valuable feedback.

The aim of this project is to analyze bank customer data and gain insight of the reasons of customer churn. It is important to focus on customers identified by the churn model. Bank and other financial institutions can greatly benefit from churn analysis. It will help them to retain their existing customer base.