Predict Customer Churn in Banking Industry

Author: Sathish Manthani

Bellevue University

1000 Galvin Rd S, Bellevue NE 68005

e-mail: smanthani@my365.bellevue.edu

# Abstract

Customers are the most valuable assets of any business. Retaining customers has become the basic need of the organizations regardless of the industry.

This document explores the various steps for predicting customer churn in banking industry using CRISP-DM methodology.

Predictive models provide insights into data after analyzing huge volumes of customer account balances and transactions data. Finance managers can probably coordinate with marketing and IT teams to identify the existing issues of such customers and target them for possible promotions can increase the chances of retention.

# Author Keywords

Data Science; Customer Churn; Predictive Analytics; CRISP-DM; Banking industry; Classification.

# ACM Classification Keywords

[CCS](https://dl.acm.org/ccs/ccs.cfm?id=0&lid=0)🡪[Information systems](https://dl.acm.org/ccs/ccs.cfm?id=10002951&lid=0.10002951) →  [Information systems applications](https://dl.acm.org/ccs/ccs.cfm?id=10003227&lid=0.10002951.10003227) →  [Data mining](https://dl.acm.org/ccs/ccs.cfm?id=10003351&lid=0.10002951.10003227.10003351) →  **Clustering**

[CCS](https://dl.acm.org/ccs/ccs.cfm?id=0&lid=0) →  [Computing methodologies](https://dl.acm.org/ccs/ccs.cfm?id=10010147&lid=0.10010147) →  [Machine learning](https://dl.acm.org/ccs/ccs.cfm?id=10010257&lid=0.10010147.10010257) →  [Machine learning approaches](https://dl.acm.org/ccs/ccs.cfm?id=10010293&lid=0.10010147.10010257.10010293) →  **Classification and regression trees**

# Introduction

In the current world, almost everyone has relationship with bank in some or the other way. Banks provide services through various channels, digital payment apps, internet banking, ATMs, brick-and-mortar locations, etc. The number of people who use banks are also increasing at a greater rate. Besides, customers have become increasingly demanding for quality of the service. This fuels a great competition amongst various banks, which in turn pressurizes banks to provide a greater quality of customer service. Also, customers have known to switch from one bank to another bank for various reasons like better app features, quick pay option, contactless payment support, customer service quality, low interest charges, nearest accessible brick-and-mortar location etc. Therefore, there is a greater need to develop a predictive model which can tell who’s going to leave next.

Emphasis on such predictive models in big banking companies is high because the cost of acquiring a new customer is at least 6 times higher than retaining an existing customer. Besides, an existing customer generates relatively more revenue than a new customer.

So, banks or credit card institutions must find ways to predict who is going to leave next. Banks should analyze such customer’s recent activity for any issues they faced and expedite resolving them, target them for promotions, and start communicating with them to ensure he/she stays satisfied with the services.

# Literature Review

## **Customer Churn**

According to Shamli Prakash (2018) (Reference [5]) Customer churn — or customer attrition is defined as the count of customers lost within a specific period of time, netting off the new acquisitions during that time. So, it is basically Newly acquired customers minus the lost customers. In other words, the metric tracks how successful or not you have been at keeping your customers engaged.

## **Stages of Customer Relationship**

Sahar F. Sabbeh (2018) describes four stages of customer relationships. 1. Customer identification 2. Customer attraction 3. Customer retention and 4. Customer development

### Customer Identification

This step involves identifying prospects who could be potential customers

### Customer Attraction

The identified prospects can be segmented into groups and sub-groups based on certain features. Different marketing techniques can be applied on each group to attract them.

### Customer Retention

This is the topic we are most interested in. Retaining existing customers is far more profitable than acquiring a new customer [21]. Customer retention includes all the required actions taken by the banking company to ensure retention of the customer. Predicting churn is an important part of customer retention.

### Customer Development

This step involves driving more spend from the existing customers. Targeting them with right offers and offering upgraded services to ensure they use our card for transactions than our competitors.

## **Customer Segmentation**

According to SAS article [11], customers need to be segmented into 4 groups, so that our marketing team can focus on high-valued, high-attrite risk customers more than others. The article highlights 4 segments of the customer.

Segments of the customer. 4 Segments are placed in 2-D graph. Risk is on Y-axis and Customer lifetime value is on X-axis.
Cultivate being on left bottom, Divest is on left top, Maintain is on right bottom and Aggressively Retain is on right top of the graph. 

Figure 1 shows the segments of customers

Maintain: These are customers with a high value and low attrition risk. This segment generally has best customers, and banks should continue to keep them happy.

Cultivate: These are customers with a low value and low attrition risk. Banks should promote their services to drive spend to make them high-value customers. These customers are not in attrite zone.

Divest: These are customers with a low value and high attrition risk. Banks revenue don’t get hit much by loosing these customers. However, efforts should be made to keep every customer and promote the services.

Aggressively Retain: These are customers with a high value and high attrition risk, and they are the ones banks should pay the most attention to. Banks revenue will be most hit if these customers continue to attrite. As soon as customers appear in the predictive model, banks should focus on them.

# CRISP-DM Methodology

## Business understanding

This phase involves understanding the features and characteristics of churners. [1]

## Data understanding

Zan [14] says “determine what data is

available to solve your business needs.”

So, this phase involves analyzing the historical customer data to see what made them churn.

In his research Zan said customers without any transaction for about two years have churned their cards eventually. Some of the variables he looked at include gender, incidents raised and transactional variables.

## Data preparation

Raw data should be transformed into useful data in this phase.

Building the variables for datasets takes place in this phase. [11]

This step involves fetching data from heterogeneous sources and performing quality checks on the data to ensure its properly loaded to the target database system.

According to SAS paper [11], some of the variables could be related to -

Transaction Recency: When was the last transaction took place.

External Deposit Recency: When was the last external deposit?

Recent Large Transaction: When was the recent large transaction took place?

Small Number of Transactions: Was there a recent small number of transactions?

Large Number of Transactions: Was there a recent large number of transactions?

A Seasonality Component: Determine if certain customers are more likely to churn during the summer, or during the Christmas season.

## Modeling

This step involves identifying a suitable machine learning model for predicting customer churn.

Since we don’t know the variables that ae contributing to the attrition of the customer, we’ll run the data through Unsupervised learning models to get insights into patterns of attrite customers.

## Evaluation and Deployment

These are final two stages which are focused on how the models helped in finding the customers that are being churned and assess the accuracy of the model to make it better.

In order to determine the accuracy and precision of the model, we’ll examine the output datasets for True positives, False positives, False negatives and True Negatives.

Let me define one of the above types.

False-negative (FN) refers to the number of positive data rows that were incorrectly

labeled as negative.

# Conclusion

In banking industry, CRISP-DM methodology can be used to identify the features and characteristics of the customers who are about to churn. Different variables of customers like demographic (gender, age, geographical location, profession, etc.,) and transactional (average daily balance, last transaction date, last large transaction, count of transactions, credit limit utilization, etc.,) variables need to be run through predictive models to get insight into the churn patterns. Marketing/IT/Call center teams will take appropriate actions to ensure the attrition-prone customers would stay satisfied with the banking services.

Also, we can use this predictive modeling framework to predict other aspects of customer behavior like when a customer would go on a vacation or buy a car so that targeting him/her with relevant offers would maximize the revenue.

# Acknowledgements

I thank the course instructor for posting relevant content for us to learn things quickly and in orderly fashion. His feedback on past submissions helped me improve. I also thank my online classmates who has been providing insightful comments. I’m learning more by reading their posts and referral links. I thank my family for supporting me throughout. Special thanks to the authors of the research papers which I mentioned in the references section below.

# References

1. Developing a prediction model for customer churn from electronic banking services using data mining <https://www.researchgate.net/publication/306388481_Developing_a_prediction_model_for_customer_churn_from_electronic_banking_services_using_data_mining>

1. Effective CRM using Predictive Analytics: A voluntary churn propensity model for credit card holders: <https://learning.oreilly.com/library/view/effective-crm-using/9781119011552/c06.xhtml>
2. Safari: Customer churn and retention:

<https://learning.oreilly.com/library/view/hands-on-data-science/9781789346343/82525596-ff2f-4eeb-86a5-cd72dd7c9509.xhtml>

1. Predicting Customer Churn For A Credit Card Company:

<https://www.galitshmueli.com/data-mining-project/predicting-customer-churn-credit-card-company>

1. AI 101: Understanding Customer Churn Management:

<https://towardsdatascience.com/ai-101-understanding-customer-churn-management-514416c17643>

1. Preventing Bank Customer Churn:

<https://www.kaggle.com/kmalit/bank-customer-churn-prediction>

1. Introduction to Churn Prediction in Python:

<https://www.datascience.com/blog/churn-prediction-python>

1. Predicting Customer Churn in Banking using Neural Networks:

<https://www.indecs.eu/2016/indecs2016-pp116-124.pdf>

1. Customer churn prediction in banking:

<https://www.slideshare.net/BU_Research_Methods/customer-churn-prediction-in-banking>

1. Predicting Potential Banking Customer Churn using Apache Spark ML and MLlib Packages: A Comparative Study: <https://thesai.org/Downloads/Volume9No11/Paper_96-Predicting_Potential_Banking_Customer.pdf>
2. Reducing Customer Attrition with Machine Learning for Financial Institutions: <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/1769-2018.pdf>
3. Predicting Financial Account Churn: <https://www.elderresearch.com/hubfs/Resources/Elder_Research_Case_Study_Predicting_Account_Churn_in_Retail_Banking.pdf>
4. [Junmei Zhong](https://dl.acm.org/author_page.cfm?id=99659372515), [William Li](https://dl.acm.org/author_page.cfm?id=99659376910) (2019)

“[Predicting Customer Churn in the Telecommunication Industry by Analyzing Phone Call Transcripts with Convolutional Neural Networks](https://dl.acm.org/citation.cfm?id=3319937)”

1. Zan M, Shan Z, Li L, Ai-jun L (2007) A predictive model of churn in telecommunications based on data mining, IEEE International Conference on Control and Automation ThAl-2 Guangzhou.
2. [Jagat Sastry Pudipeddi](https://dl.acm.org/author_page.cfm?id=87958796557), [Leman Akoglu](https://dl.acm.org/author_page.cfm?id=81414592499), [Hanghang Tong](https://dl.acm.org/author_page.cfm?id=81337494052) (2014) “ [User churn in focused question answering sites: characterizations and prediction](https://dl.acm.org/citation.cfm?id=2576965)” WWW '14 Companion: Proceedings of the 23rd International Conference on World Wide Web
3. [Marwa N. Abd-Allah](https://dl.acm.org/author_page.cfm?id=99658647486), [Samhaa R. El-Beltagy](https://dl.acm.org/author_page.cfm?id=81100218127), [Akram Salah](https://dl.acm.org/author_page.cfm?id=84460506657) (2017) “[DyadChurn: Customer Churn Prediction using Strong Social Ties](https://dl.acm.org/citation.cfm?id=3105832" \t "_self)” IDEAS 2017: Proceedings of the 21st International Database Engineering & Applications Symposium
4. [Shuting Xu](https://dl.acm.org/author_page.cfm?id=81408601242), [Shuhua Lai](https://dl.acm.org/author_page.cfm?id=81536716656), [Manying Qiu](https://dl.acm.org/author_page.cfm?id=81418592746) (2009) “[Privacy preserving churn prediction](https://dl.acm.org/citation.cfm?id=1529643)” SAC '09: Proceedings of the 2009 ACM symposium on Applied Computing
5. [Yiqing Huang](https://dl.acm.org/author_page.cfm?id=99658720032), [Fangzhou Zhu](https://dl.acm.org/author_page.cfm?id=99658720670), [Mingxuan Yuan](https://dl.acm.org/author_page.cfm?id=81100116352), [Ke Deng](https://dl.acm.org/author_page.cfm?id=99658717541), [Yanhua Li](https://dl.acm.org/author_page.cfm?id=81548016057), [Bing Ni](https://dl.acm.org/author_page.cfm?id=99658680485), [Wenyuan Dai](https://dl.acm.org/author_page.cfm?id=81333488189), [Qiang Yang](https://dl.acm.org/author_page.cfm?id=81548761456), [Jia Zeng](https://dl.acm.org/author_page.cfm?id=81558247556) (2015) “[Telco Churn Prediction with Big Data](https://dl.acm.org/citation.cfm?id=2742794)” SIGMOD '15: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data
6. [Marwa N. Abd-Allah](https://dl.acm.org/author_page.cfm?id=99658647486), [Akram Salah](https://dl.acm.org/author_page.cfm?id=84460506657), [Samhaa R. El-Beltagy](https://dl.acm.org/author_page.cfm?id=81100218127) (2014) “[Enhanced Customer Churn Prediction using Social Network Analysis](https://dl.acm.org/citation.cfm?id=2665997)” DUBMOD '14: Proceedings of the 3rd Workshop on Data-Driven User Behavioral Modeling and Mining from Social Media
7. [Mehmet Gök](https://dl.acm.org/author_page.cfm?id=99658764166), [Tansel Özyer](https://dl.acm.org/author_page.cfm?id=81100569218), [Jamal Jida](https://dl.acm.org/author_page.cfm?id=81548043584) (2015) “[A Case Study for the Churn Prediction in Turksat Internet Service Subscription](https://dl.acm.org/citation.cfm?id=2808821)” ASONAM '15: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015
8. Sahar F. Sabbeh (2018) “Machine-Learning Techniques for Customer Retention: A Comparative Study” (IJACSA) International Journal of Advanced Computer Science and Applications