Bank Customer Churn Association Rules (June 2019)

Jason Huggy, Student, Lewis University

Abstract—The purpose of this study was to use association rules to determine the reasons for customer churn at a European bank. This study looked at a variety of features, to include: salary, whether the customer has a credit card with the bank or not, how many products from the bank each customer uses, and basic information such as gender and age. Orange Canvas is utilized to prepare all data and to determine possible association rules amongst the data. The association rule method was able to find several strong correlations, such as customers with credit cards from the bank, male customers, customers aged 33-48, active members, customers who use at least two of the bank's products, and members from France all tend to stay with the bank. All of these relationships could prove useful to the bank in trying to improve its business model.

I. INTRODUCTION

This study focuses on the customers of an anonymous European bank, and more specifically, looks at the factors that impact customer churn. This study analyzes 10,000 customers fixated on 11 primary features, to include: credit score, country, gender, age, tenure, balance, number of products used, if the customer has a credit card with the bank, if the customer is an active member, estimated salary, and whether the customer was retained or lost. The outcome of this study is expected to help the bank manage its program to reduce customer churn.

Frequent items and association rule mining are utilized to compare these features, and to determine implied relations between features. Frequent item-sets are determined by counting the number of times each individual feature and combination of features occurs in the data. A support and confidence threshold are assigned to refine the data into association rules that demonstrate to have the biggest impact on customer churn.

II. METHODOLOGY

To conduct this study on bank customer churn, a dataset named *Churn Modeling* was imported from Kaggle.com [1]. This dataset includes 14 features, in which 11 are used for this study, to analyze the data on 10,000 customers. The features used include: credit score, country, gender, age, tenure, balance, number of products used, if the customer has a credit card with the bank, if the customer is an active member, estimated salary, and whether the customer was retained or lost. The 11 features used are described in detail below.

- *Credit score* is the customer's credit score ranging from 350 to 850 and is separated into categories. The categories are: *< 450; 450-550; 550-650; 650-750; ≥ 750.*
- *Geography* is the country that each customer lives in, to include France, Spain and Germany.
- *Gender* is male or female.
- *Age* is the age of the customer ranging from 18 to 92 and is separated into categories. The categories are: < 33; 33-48; 48-62; 62-77; ≥ 77.
- *Tenure* is the number of years each customer has been with the bank and is separated into categories. The categories are: *< 2; 2-4; 4-6; 6-8; ≥ 8*
- *Balance* is the bank balance of each customer and is separated into categories. The categories are: *< 50,179.62; 50,179.62-100,359.24; 100,359.24-150,538.85; 150,538.85-200,718.47; ≥ 200,718.47.*
- *NumOfProducts* in the number of bank products each customer uses and is separated into categories. The categories are: *< 2; 2; 3; > 3.*

- *HasCrCard* is a binary flag for whether the customer has a credit card with the bank or not. 0 means no, and 1 means yes.

- *IsActiveMember* is a binary flag for whether the customer is an active member with the bank or not. 0 means no, and 1 means yes.

- *EstimatedSalary* is the estimated salary for each customer and is separated into categories. The categories are: *< 40,007.76; 40,007.76-80,0003.94; 80,0003.94-120,000.12; 120,000.12-159,996.30; ≥159,996.30*

- *Exited* is a binary flag of whether the customer was retained or if the customer closed their account. 0 means no, and 1 means yes.

A. Creation of Association Rules

For this study, Orange Canvas, a data analysis tool, is used to pre-process the bank data and then create the association rules that meet a certain criterion. The only edit to the data needed was to categorize all the numerical features. Their breakdown is listed above. The Association Rules widget in Orange Canvas first calculates frequent occurrences among each feature of the data, and then applies a support and confidence threshold [2]. The support threshold assigns the minimum percentage of how often a rule must be applicable to the given data for an association to be considered. Support in this study is calculated as the number of times each rule occurs divided by the total number of customers. Confidence predicts how frequently a rule is true and is calculated as support for a rule divided by the support of the antecedent. The stronger the relationship, the higher the confidence. Samples must match or exceed the confidence threshold to be considered. For this study, a 40 percent support threshold is assigned and a 75 percent confidence threshold.

III. RESULTS & DISCUSSION

After running the Association Rules widget in Orange Canvas, the results are listed. Fig. 1. shows the output. The first two columns show the support and confidence for each rule, and the last two columns display the resulting rule. Overall, the parameters used found 6 association rules.

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.563	0.798	0.706	1.129	1.002	0.001	HasCrCard=1	→	Exited=0
0.456	0.835	0.546	1.459	1.049	0.021	Gender=Male	→	Exited=0
0.447	0.812	0.550	1.448	1.019	0.009	Age=33 - 48	→	Exited=0
0.442	0.857	0.515	1.546	1.077	0.031	IsActiveMember=1	→	Exited=0
0.424	0.924	0.459	1.735	1.161	0.059	NumOfProducts=2 - 2	→	Exited=0
0.420	0.838	0.501	1.588	1.053	0.021	Geography=France	→	Exited=0

Fig. 1. Orange Canvas association rule mining results from Bank Churn data

A. Customer Has Credit Card

The first line in the results shows the association between customers having a credit card and not closing their account. This rule had the highest support, meaning that 56 percent of the customers had a credit card and did not close their account. The confidence was 79 percent; meaning that within the cases of a customer having a credit card, there was a 79 percent chance that they would not close their account. This association means that there is a high likelihood that customers that have a credit card with the bank are less likely to close their account. With the confidence only at 79 percent, it expected that this will not be the case for every customer; however, this does show a strong relationship between a customer having a credit card and whether they choose to close their account. This would suggest that customers who open a credit card are more committed to the bank. That or else customers are too lazy to close their bank account knowing that they would have to change their credit card. This results in customers having to change their account information for all their automatic payments. Another reason for this occurrence could be because the bank offers better rates on their credit card for customers who choose to set-up bank accounts with them. More research should be conducted on this occurrence to further understand the reasoning behind it.

B. Male Customers

The next association rule listed shows a relation between males and the customers that do not close their accounts. 45 percent of the customers listed were male and did not close their account. The confidence that a male customer keeps open their account is 83 percent. This is a considerable amount and shows a strong correlation between male customers and whether a customer keeps open their account. This may suggest that male customers are more likely to commit to a bank for a long time, or that males are more likely to research what bank is best for them and then stick with it. The specific

reasons are not known, and more research should be put towards understanding this relation.

C. Age 33-48

44 percent of the customers who did not close their account comprised of people aged 33 to 48. The association rule results show that there is 81 percent confidence that customers aged 33 to 48 will not close their account. This is another significant detail. This may propose that customers aged 33 to 48 are more likely to stay with a bank due the period in life that they are in. At this age a person is typically settled into a job they will do for the rest of their life. They are also worried about a family, house, and bills. Younger people are still trying to figure out their life, so may bounce around banks. Older people may change to banks that will support them as they reach retirement. There are many possibilities, and this relationship should be researched on its own for added value.

D. Active Members

What determines an active member at this bank is not specified, but there is 85 percent confidence that active members will not close their account. An active customer may assume that a customer uses the bank's application or website to check their account often; or perhaps, consistently deposits or withdraws money from their account. This would infer that customers who do these types of activities more often are more likely to continue using the bank. More research may need to be done to determine what defines an active member, and what actions lead to less customer churn.

E. Customer Utilizes Two Products

Customers who use two of the bank's products and do not close their account consist of 42 percent of the total data, and this relates to a 92 percent confidence level to customers who will not close their account. More research should be done to determine which products lead to a lower churn rate, but this assumes that this population of customers utilizes two products such as a checking account and savings account. In the future, this association rule could be altered to show customers who use 2 products or more. It is assumed that any customer who uses at least 2 products from the bank is more likely to keep using the same bank.

F. French Members

Members from France who do not close their account consist of 42 percent of the customers in this study. There is an 83 percent confidence associated with customers from France will not close their account. The reason for this is unknown, but it is possible that the bank's options benefit people from France more often. Maybe French people are less likely to change banks. While the confidence is only 83 percent, this is still a high percentage and this instance should be researched further because of it.

G. Continued Research and Customer Focus

The results from this study show that there are several different studies that need to occur to get further value out of this study. The bank's management needs to look at why males, customers 33 to 48, active members, customers who utilize two products, and members from France tend to leave their accounts open. This population is more committed to the bank and it is advised that more be done to understand why.

As for reducing customer churn, it is recommended that the bank take two different options to improve the value of their business. One, the bank's management focuses on the populations of customers listed in this study. These members are less likely to close their accounts, so if they improve their marketing or support for the listed populations then the bank may see improved numbers amongst these groups. The other option is to focus on the population of customers not listed in this study's association rules. Put more focus on younger and older customers. Direct attention to female customers. Try to get inactive customers more active, and market the benefits of the bank's credit card more. Market more in Germany and Spain to improve the customer base in these Countries. Lastly, give customers a reason to want to use more than one of the bank's products. Whichever route the bank decides to go will most likely improve the bank's overall business.

IV. CONCLUSION

In closing, the purpose of this study was to determine what customers tend to close their accounts with an anonymous European bank. The study focused on 10,000 customers using 11 features, to include: credit score, country, gender, age, tenure, balance, number of products used, if the customer has a credit card with the bank, if the customer is an active member, estimated salary, and whether the

customer was retained or lost. The study found a that customers that use credit cards from the bank, male customers, customers aged 33-48, active members, customers who use at least two of the bank's products, and members from France all tend to stay with the bank. This assumes that the members that close their accounts tend to associate with the rest of the population not listed. It is advised that further research be conducted on each individual association found in this study. The bank must understand these relationships more in depth to get the most out of the business.

V. References

[1] Kaggle.com. (2019). *Churn Modelling*. [online] Available at: https://www.kaggle.com/shrutimechlearn/churn-modelling#Churn_Modelling.csv [Accessed 8 Jun. 2019].

[2] University of Ljubljana, "Association Rules in Orange," *Orange Blog*. [Online]. Available: https://blog.biolab.si/2016/04/25/association-rules-in-orange/. [Accessed: 08-Jun-2019].