

Prediction of the Reasons for Losing Credit Card Customers based on Multiple Models

Chengyu Huang^{1,a,*}

¹*McMaster University, Hamilton, ON, L8S 4L8, CA*

a. Chengyu2392373977@outlook.com

**corresponding author*

Abstract: The data in this article is taken from Kaggle, and the data is publicly available. The data mainly focuses on the nearly 20 different possible reasons that the bank is losing credit card customers, such as their monthly and yearly income or gender. However, some of the data in the original dataset is embodied in words. For this kind of data, the data needs to be converted into numbers and then put into the model for testing. It is also important to check the data whether it has the outliers. Putting the data set into the models to test the accuracy. By sorting the results of different models, the model with the best matching degree is finally obtained. Then, the highly correlated factors in the model are sorted to get the main reasons why banks lose credit card customers. In the end, providing some business advice for banks could be used to identify the same group of credit card customers, and providing some advice to banks could be used to save credit card customers.

Keywords: Credit card users, Machine learning, Statistics, K-NN model, Logistic model, Naive Bayes.

1. Introduction

A credit card is a physical card. When a credit card account is opened, the credit card company sets a credit limit. Within the credit limit, the cardholder can purchase goods, or services, or withdraw cash. The use of credit cards is very common in today's society, however, with the insecurity of credit cards and the emergence of more convenient and efficient mobile payment methods, gradually people are using credit cards less and have even decided to quit the credit card business. Therefore, analyzing the reasons for losing credit card customers is a potentially important source of value for financial institutions or banks.

So, I hope to use the data model to analyze the credit cardrelated problem. In the beginning, we should know the importance of data analysis. In the context of digital development, data analysis tends to be widely used by enterprises to improve their business since increasingly more companies have access to gain their customers' data and consumption data, whether from their sources or a third party [1]. Also, online companies can conduct data analysis to understand their customers' behavior and preferences [2], then they can make adjustments or apply some operation strategies to improve the number of customers, improve the conversion rate, which finally leads to the increase of the revenues. So, after we know the importance of using data to analyze types of questions, we need to understand what is the purpose of the article.

The main purpose of this research paper is to try to identify the main reasons why banks lose customers by constructing a model about bank credit card customers based on a set of actual data so that it can better screen the reasons why they may lose customers and can also be used to identify potential customer groups.

Credit cards are a long-term and profitable loan program for banks because credit card accounts take a revolving line of credit and therefore, banks have more aspects to monitor and manage them than other retail loans such as mortgages [3]. The high-interest balances generated when borrowers do not pay their credit card rooms on time are a major source of profit for banks [4]. Therefore, credit card customer churn is a huge loss for banks, however, it is particularly important to analyze the reasons for customer churn. There are objective reasons for customer churn, such as credit card fraud, in 2017, there were 1,579 data breaches and nearly 179 million records among which Credit card frauds were the most common form with 133,015 reports [5]. Meanwhile, along with this, in some countries, mobile payments have become an established payment mechanism similar to credit or debit cards. For example, the volume of mobile payment transactions in China is expected to exceed credit card transactions by 30% [6]. However, there are many objective reasons, such as customer gender, income, marital status, educational background, etc. This paper analyzes and gives recommendations based on these possible variables to help banks better deal with customer churn and predict potential customers.

2. Literature Review

Keeping an eye on credit card usage has always been a very important issue. Credit cards are being used more and more frequently in consumer spending, both in terms of profitability for banks and ease of use for cardholders, and they have taken over the dominant payment system. Expanding globalization and widespread use of the Internet have led to increased frequency of credit card use, especially in online shopping. A variety of online platforms have made it possible to constantly target potential customers through a potentially irresistible variety of deals and promotional holidays (e.g. Black Friday, Christmas, New Year). Similarly, credit card companies offer lucrative rewards (e.g. miles, points, cashback bonuses) for using credit cards instead of cash. According to the Federal Reserve Bank, 72% of U.S. consumers have at least one credit card as of 2014. However, a large consumer base doesn't mean they understand exactly what's behind the credit card. Many consumers, especially younger consumers, lack financial literacy and consumer experience, making it easy for merchants to accept misleading offers and abuse their credit. They show no concern for the expansion of credit. As of 2015, the average credit card debt of card-carrying households reached \$15,779 with an outstanding revolving balance of \$937.9 billion [7].

Yet with such a huge frequency of credit card use, attention to credit card customers is essential. In the past, however, banks have tended to focus their attention on credit card customers in terms of defaults. There are many ways to identify credit card defaults, such as the widely used classification method. Data mining is used to construct models, such as multicriteria linear programming, that classify the behavior of credit cardholders to determine the likelihood of default and late payment [8]. In addition, based on Knight's theoretical research, a model for studying credit card risk was creatively proposed, using a priori data on credit card customers' personal information and consumer behavior to distinguish uncertainty and risk. The data level is convincing enough, but the degree of relevance to consumers themselves is low, and the scope is too narrow by starting with behavior alone. However, at the same time, as Hyun's study [25] shows, standard credit risk models can be built to classify credit cardholders. Studies conducted through machine learning (ML) have shown the impact of credit risk models in identifying multiple banking risk problems.

While it is important for banks to control risk properly, it is also important for cardholders to protect their assets safely. The rapid and high volume of credit card transactions (CCTs) has led to a

significant increase in fraud cases. In 2017, a CyberSource report noted that fraudulent Latin American e-commerce credit card transactions (chargebacks) amounted to 1.4% of the total net value of the industry [9].

But now with the recessionary form of the economy, especially through the COVID-19 period, many financial institutions are on the verge of bankruptcy. This is because, during the pandemic, credit growth was negative during the downturn, while bank deposits (i.e., third-party funding) increased[10]. The negative impact of the recession, along with the questionable security of credit cards and the continuous loss of credit card customers has become a huge problem. Therefore, I will use a logistic regression model to examine the data and give the most intuitive reason for banks to lose credit card customers.

Therefore, this paper mainly uses a logistic regression model to examine the data and build a model to give banks the most intuitive reason for losing credit card customers, to help banks to better retain existing customers and explore potential customers.

3. Methodology

3.1. Analysis and Data Set

The credit card customer dataset is downloaded from the official website of Kaggle (<https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers>). This dataset consists of over 10,000 customers including their age, salary, marital status, credit card limit, credit card type, etc. with nearly 20 features.

First, the presentation of unified data is very important: for Attrition Flag, 1 represents Existing Customer, 2 represents Attrited Customer; for Gender, 1 represents Male, 2 represents Female; for Education level, 1 represents Graduate, 2 represents High School, 3 for others; for Marital Status, 1 for Married, 2 for Single, 3 for others; for Income Category, 1 for Less than \$40K, 2 for Between \$40K - \$60K, 3 represents others; for Card Category, 1 represents Blue, 2 represents Silver, and 3 represents others.

To better understand the variables in the dataset and check whether there are missing values or any abnormal values in the dataset, the descriptive analysis will be applied first. The descriptive statistics about the variables in the dataset are shown below:

Table 1: Descriptive of four elements.

	Transformed-Attrition-Flag	Customer-Age	Transformed-Gender	Dependent-count
N	10127	10127	10127	10127
Missing	0	0	0	0
Mean	0.839	46.3	1.53	2.35
Median	1	46	2	2
Standard deviation	0.367	8.02	0.499	1.30
Minimum	0	26	1	0
Maximum	1	73	2	5

Table 2: Descriptive of four elements.

	Transformed-Attrition-Flag	Customer-Age	Transformed-Gender	Dependent-count
N	10127	10127	10127	10127
Missing	0	0	0	0
Mean	2.18	1.69	2.12	1.08
Median	2	2	2	1
Standard deviation	0.876	0.715	0.899	0.319
Minimum	1	1	1	1
Maximum	3	3	3	3

Table 3: Descriptive of four elements.

	Transformed-Attrition-Flag	Customer-Age	Transformed-Gender	Dependent-count
N	10127	10127	10127	10127
Missing	0	0	0	0
Mean	35.9	3.81	2.34	2.46
Median	36	4	2	2
Standard deviation	7.99	1.55	1.01	1.11
Minimum	13	1	0	0
Maximum	56	6	6	6

Table 4: Descriptive of four elements.

	Transformed-Attrition-Flag	Customer-Age	Transformed-Gender	Dependent-count
N	10127	10127	10127	10127
Missing	0	0	0	0
Mean	8632	1163	7469	0.760
Median	4549	1276	3474	0.736
Standard deviation	9089	815	9091	0.219
Minimum	1438	0	3.00	0.00
Maximum	34516	2517	34516	3.40

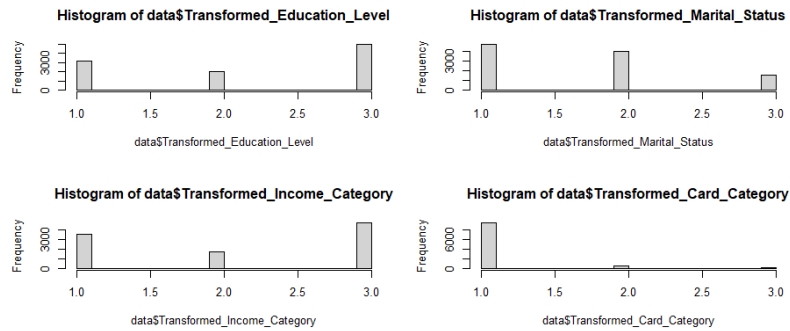
Table 5: Descriptive of four elements.

	Transformed-Attrition-Flag	Customer-Age	Transformed-Gender	Dependent-count
N	10127	10127	10127	10127
Missing	0	0	0	0
Mean	4404	64.9	0.712	0.275
Median	3899	67	0.702	0.176
Standard deviation	3397	23.5	0.238	0.276
Minimum	510	10	0.00	0.00
Maximum	18484	139	3.71	0.999

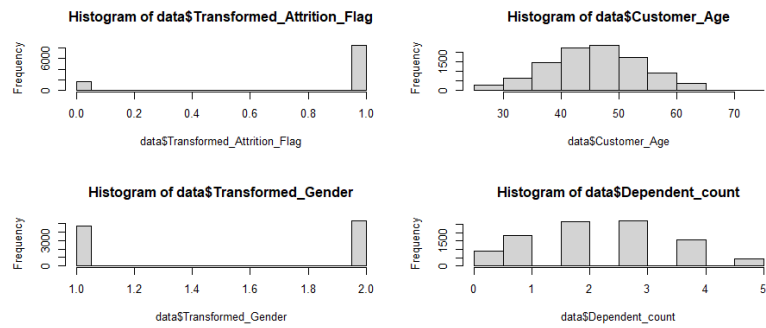
Table1-5 shows the descriptive statistics about all the variables in the wine quality dataset. According to the table above, it is obvious that there are no missing values in the dataset. In addition, according to common sense and the range (maximum-minimum) of the variables above, there are no abnormal values for these variables. Moreover, by observing the features of these variables in the dataset, it is not difficult to find out that all variables are continuous variables.

In detail, the average customer age is 8.3, while the minimum customer age is 26 and the maximum customer age is 73. The average dependent count is 2.35, while the minimum dependent count is 0 and the maximum dependent count is 5. As for the months on a book, the average value of months on the book is 35.9, while the minimum number of months on the book is 13 and the maximum customer age is 56, so the range of this chemical in wine is 43 ($56-13=43$). The average total relationship count is 3.81, while the minimum total relationship count is 1 and the maximum citric acid used is 6. As for the months inactive in 12 months, the average value of months inactive in 12 months is 2.34, while the range of t months inactive in 12 months is 6 ($6-0=6$). The average contacts count in 12 months is 2.46, while the minimum contacts count in 12 months is 0 and the maximum contacts count in 12 months is 6. As for the credit limit, the average value of the credit limit is 8632,

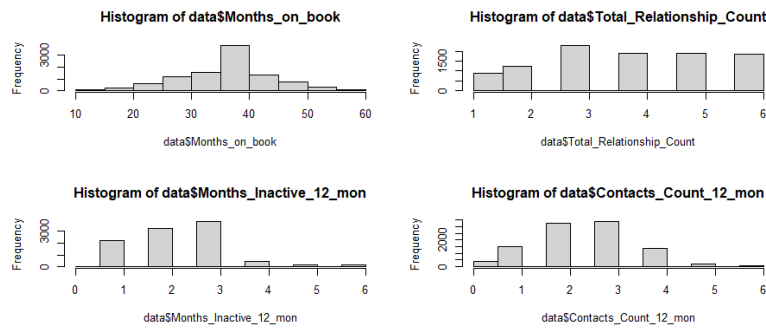
while the range of credit limit is 33078 ($34516-1438=33078$). The other elements could be read in the same way as the above data.



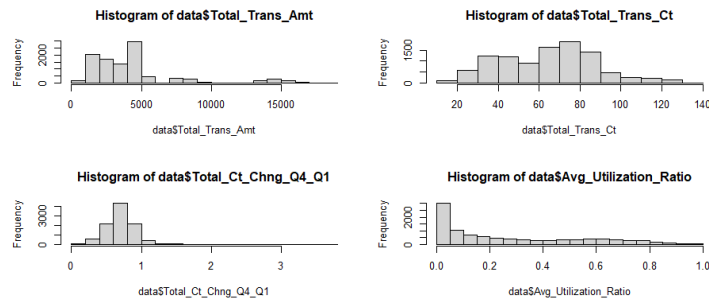
(a)



(b)



(c)



(d)

Figure 1: Histograms of four features in the data set.

From figure 1, all the variables in the dataset have been shown in the form of descriptive statistics tables. In the following part, the histograms of these variables will be plotted to understand the distributions of these variables. According to the figures above, the variables of customer age, dependent count, months on book, months inactive in 12 months, and total transaction count are approximately normal distribution while all other variables in the dataset seem to be right-skewed. Except for some transformed data which was words original, they have transformed to simple numbers 1,2,3, so the histogram for them is a little strange, but there is no influence to analyze the data.

3.2. Analytical Model

As stated above, the main task of this report is to understand the different kinds of elements in influencing lose and get the credit card customers, and the effects of these different kinds of influencing lose and get the credit card customers. Thus, the logistic regression model, model for KNN, and model for Naive Bayes will be applied in the following parts.

3.2.1. Logistic Regression Model

When there are no more than two outcomes in the regression model for predicting the dependent variable, the binomial logistic regression analysis will be applied instead of the simple regression analysis. As the goal of this report is to examine the impact of various kinds of reasons why the bank will lose credit card customers, the binomial logistic regression analysis will be applied.

3.2.2. K-NN Model

The main thinking of the k-NN algorithm is that if a sample belongs to a class if most of its k-nearest neighbors in the feature space belong to a certain class, which was considered has the same characteristics. When determining the classification decision, the only way to determines which the class belong to, is chosen by the class of the most recent sample or samples. The k-NN method is only relevant to a tiny number of near samples when making class decisions. The most important parameter is the K value, so the accuracy is starting from 1 to 8.

3.2.3. Model train for Naïve Bayes

Naive Bayes Classifier (Naive Bayes Classifier or NBC). In theory, the NBC model has highest accuracy compared to other classification methods. This is because the NBC model assumes each element are not dependent. However, this assumption is always wrong in reality. The idea of Naive Bayes is a posterior probability, which is to calculate the probability of occurrence of the event once again in the event of an event.

4. Analysis and results

4.1. Results and analysis of Binomial Logistic Regression Model

After figuring out the correlation coefficients between quality and some other variables in the dataset, the multivariate regression analysis will be conducted. And to begin with, all independent variables will be used to fit the regression model to predict the quality of the wine. The results of the multivariate regression model are shown below:

Model Fit Measures

Model	Deviance	AIC	BIC	R ² _{McF}
1	4757	4795	4933	0.467

Figure 2: Logistic Regression model 1 with all independent variables.

Model Coefficients - Transformed_Attrition_Flag

Predictor	Estimate	SE	Z	p
Intercept	-3.20142	0.45477	-7.040	< .001
Customer_Age	0.00399	0.00762	0.524	0.600
Transformed_Gender	-0.66479	0.09835	-6.759	< .001
Dependent_count	-0.13474	0.02956	-4.559	< .001
Transformed_Education_Level	-0.07085	0.04296	-1.649	0.099
Transformed_Marital_Status	-0.32587	0.05277	-6.176	< .001
Transformed_Income_Category	-0.02707	0.05753	-0.471	0.638
Transformed_Card_Category	-0.35685	0.13590	-2.626	0.009
Total_Relationship_Count	0.45005	0.02738	16.436	< .001
Months_on_book	0.00534	0.00761	0.702	0.483
Months_Inactive_12_mon	-0.50243	0.03766	-13.342	< .001
Contacts_Count_12_mon	-0.50605	0.03625	-13.961	< .001
Avg_Open_To_Buy	7.59e-6	6.35e-6	1.196	0.232
Total_Ct_Chng_Q4_Q1	2.81721	0.18719	15.050	< .001
Total_Amt_Chng_Q4_Q1	0.42207	0.18605	2.269	0.023
Total_Revolving_Bal	9.62e-4	6.97e-5	13.798	< .001
Total_Trans_Ct	0.11706	0.00368	31.822	< .001
Total_Trans_Amt	-4.75e-4	2.27e-5	-20.939	< .001
Avg_Utilization_Ratio	0.03933	0.24486	0.161	0.872

Note. Estimates represent the log odds of "Transformed_Attrition_Flag = 1" vs. "Transformed_Attrition_Flag = 0"

Figure 3: Logistic Regression model 1 with all independent variables.

Figures 2 and 3 has shown the results for the binomial logistic regression model with all independent variables. According to the figure, only the variables of gender, dependent count, marital status, relationship count, number of inactive months in 12 months, number of contacts in 12 months, total transaction amount, and change in total transaction amount is statistically significant in this logistic regression model since their p-values are smaller than 0.01. All other estimated coefficients are not statistically significant in this multivariate regression model.

As for the estimated coefficients which are statistically significant in this multivariate regression model, the indications of their estimated coefficients will be explained in detail. The estimated coefficient of Gender is -0.66, which indicates keeping other variables held constant, if the Gender increase by 1, then the bank will lose 0.66 credit card customers. The estimated coefficient of dependent count is -0.13, which indicates keeping other variables held constant, if the number of dependent count increases by 1, then the bank will lose 0.13 credit card consumers. The estimated coefficient of total marital status is -0.33, which indicates keeping other variables held constant, if the number of marital statuses increases by 1, then the bank will lose 0.33 credit card consumers. The estimated coefficient of total relationship count is 0.45, which indicates keeping other variables held constant, if the amount of total relationship counts increases by 1, then the bank will increase by 0.45 customers. The estimated coefficient of months inactive is -0.50, which indicates keeping other variables held constant, if the number of months inactive increases by 1, then the bank will lose 0.5 credit customers. The estimated coefficient of contacts count is -0.51, which indicates keeping other variables held constant, if the amount of contacts increases by 1, then the bank will lose 0.51 credit customers. It comes the same with other related elements.

In addition, the multiple R-squared value for this multivariate regression model is 0.467, which means 46.7% of the variation in the reason for losing credit card customers can be explained by these independent variables in the model.

It can be found from figure 3 that several variables did not pass the hypothesis test, that is, the parameters of several variables did not have statistical significance. So, the number of variables should be reduced next. Here, this report uses stepwise regression. Stepwise regression is based on the initial model, which gradually reduces a certain variable, then calculates the AIC value of each model, and selects the smallest AIC value until the variables can no longer be deleted. In other words, the final model has the lowest AIC value. The following output is a summary of the model.

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.5089   0.1161   0.2492   0.4538   2.2815

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.755273   0.315354  -5.566 2.61e-08 ***
Transformed_Gender -0.543865  0.089496  -6.077 1.22e-09 ***
Dependent_count -0.129960  0.033974  -3.825 0.000131 ***
Transformed_Marital_Status -0.337595  0.061330  -5.505 3.70e-08 ***
Total_Relationship_Count  0.543023  0.032291  16.817 < 2e-16 ***
Months_Inactive_12_mon  -0.435284  0.043246 -10.065 < 2e-16 ***
Contacts_Count_12_mon    -0.409383  0.041393  -9.890 < 2e-16 ***
Total_Trans_Ct           0.061715  0.002543  24.273 < 2e-16 ***
Total_Ct_Chng_Q4_Q1      2.792044  0.212107  13.163 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4: Logistic Regression model without insignificant elements.

Figure 4 above shows the logistic regression model without insignificant elements, when deleting the irrelevant elements, the reliabilities of the model will increase.

Next, check the performance in the test set, see the figure below (figure 5):

```
pred_glm
      n      y
0  298  343
1  106 3323
```

Figure 5: Performance of the New Logistic Regression model.

Because the result of logistic regression calculation is the probability of success, we judge the correct rate is the correct rate of judgment classification. Therefore, if the calculation result is greater than or equal to 0.5, the admission is successful. The final calculation accuracy rate is 88.97%

4.2. Results and analysis of the K-NN model

The KNN algorithm uses a majority voting mechanism. It collects data from the training dataset and then uses that data to make predictions on new records. The use of the KNN algorithm can compete with the most accurate models because it can make highly accurate predictions. Therefore, we can use the KNN algorithm for applications that require high accuracy but do not require a human-readable model.

Here comes with the results of the K-NN model:

	k <dbl>	accuracy <dbl>
1	1	0.0000000
2	2	0.8702703
3	3	0.8884521
4	4	0.8823096
5	5	0.8926290
6	6	0.8916462
7	7	0.8884521
8	8	0.8879607
9	NA	0.8936118
10	NA	0.8914005

1-10 of 21 rows

Figure 6: K-NN model results.

	k <dbl>	accuracy <dbl>
11	NA	0.8948403
12	NA	0.8950860
13	NA	0.8943489
14	NA	0.8909091
15	NA	0.8918919
16	NA	0.8909091
17	NA	0.8896806
18	NA	0.8877150
19	NA	0.8886978
20	NA	0.8886978

11-20 of 21 rows

Figure 7: K-NN model results.

	k <dbl>	accuracy <dbl>
21	NA	0.8889435

21-21 of 21 rows

Figure 8: K-NN model results.

From the figures 6-8, we can read the table that when K= 5 or 6, which means that we could take 5 or 6 elements, at that time, the accuracy reaches the highest value. However, if we take the suitable values to test the accuracy, we will find the result in the following table.

In the end, the accuracy obtained by this algorithm is equal to 87.49%, which can be regarded as a very good classification model.

knn	0	1
0	379	247
1	262	3182

Figure 9: K-NN model accuracy results.

4.3. Model training for Naive Bayes

In this model, I include all the variables in the algorithm. The following is a summary of the output of the model.

```
Confusion Matrix and Statistics

      Reference
Prediction 0    1
0      320   321
1      146  3283

Accuracy : 0.8853
95% CI : (0.8751, 0.8949)
No Information Rate : 0.8855
P-Value [Acc > NIR] : 0.5319

Kappa : 0.5137

McNemar's Test P-value : 8.161e-16

Sensitivity : 0.68670
Specificity : 0.91093
Pos Pred Value : 0.49922
Neg Pred Value : 0.95742
Prevalence : 0.11450
Detection Rate : 0.07862
Detection Prevalence : 0.15749
Balanced Accuracy : 0.79881
```

Figure 10: K-NN model results.

It can be found that the accuracy rate reaches 88.53% and the Kappa value is greater than 0.5, which indicates that this model performs well on the test set.

5. Discussion

5.1. Conclusion of Models

The accuracy of the four methods is

logistic regression 88.97%, K-NN 87.49%, Naïve Bayes 88.53%,

In this paper, by comparing the four methods, I finally found that the logistic regression model (specifically for binomial one) has the best prediction ability; it comes with Naive Bayes and KNN. Here, it cannot be concluded that the K-NN model is the best poor model because the data can be made more applicable to the model through variable transformation.

5.2. Shortcomings of Models

Moreover, the K-NN table has a few shortcomings, one is we could just use the K-NN table to identify how many elements we could use in our model, then we need to change the number of elements to reach the highest accuracy; secondly, K-NN is very sensitive to data size because it relies on computing distances. For features with higher scales, the calculated distance may be very high and may yield poor results. Therefore, it is recommended to scale the data before running KNN.

In addition, the general model has 2 shortcomings: 1. It is not necessarily correct to judge whether a bank will lose a customer with a limit of 0.5, because in the subsequent analysis, it was found that there are still many credit card customers, obviously not in line with common sense. 2. The processing of missing values is not very good, and direct deletion will reduce the overall variance of the data set. All the information provided by the sample was not fully used.

5.3. Suggestions for the Research

Based on the results obtained above, some suggestions can be made to help bank credit card managers to improve the probability to get new and save original credit card holders, and therefore increase bank revenue and expand bank services. To begin with, since the gender, dependent count, marital

status, months inactive in 12 months (No. of months inactive in the last 12 months), and contacts count in 12 months (No. of Contacts in the last 12 months) have negative effects on deciding which credit customer should be chosen, here comes with few ideas:

1. Try to find some men as potential credit card holders, and then try to give men some discounts for existing credit cardholders. There is a high possibility that they can be retained. A possible reason is that men who tend to hold more wealth may be more likely to be potential credit card users.

2. Try to find or retain some people with simple family members as credit card candidates. For such people, their family members are often simpler and therefore more willing to become credit card candidates.

3. For married credit card holders or candidates, try your best to win, they will often face some financial difficulties, so having a credit card can be a good way to tide over the difficulties, but getting married often means more sense of responsibility, So timely repayment is often not a big problem for them.

4. Pay attention to customers who have not been very active in the past 12 months, they are often unlikely to be loyal credit card customers because if they are not active in the banking business, they are also more likely to have a credit card business that requires timely repayment Not very interested.

5. Don't assume that the higher the number of contacts in the past twelve months, they will be active, and they may not be potential or loyal credit card customers.

6. For customers who hold more products, they tend to become loyal credit card customers, because the more products they hold in a bank, the more they have deposits in that bank, and they are also satisfied. Bank-related financial business so pays special attention and pay attention to this group of people.

7. Pay attention to those who have more total transactions. The more transactions they have in the bank, it means that their capital flows are all placed in this bank, which means they are very satisfied with this bank. Service mechanism, whether it is low transfer fee or high-interest rate, there are many possibilities for such people to become credit card customers.

8. For those with an increasing number of transactions in the first to fourth quarters, an upward trend in their number of transactions means that most of their turnover at the end of the year will be in this bank, and it also shows that they are very satisfied with the bank's service, and there is a high possibility of becoming a potential holder of the bank's credit card.

6. Conclusion

As introduced above, the main goal of this report is to understand different kinds of reasons why the bank is losing credit card customers and to examine the impacts of different reasons. To figure out these questions, descriptive statistics, data visualization (histograms), Pearson correlation coefficient, binomial regression analysis, K-NN model, and Naïve Bayes model have been applied in R studio. The results have shown that the gender, dependent count, marital status, total relationship count (Total no. of products held by the customer), months inactive in 12 months (No. of months inactive in the last 12 months), contacts count in 12 months (No. of Contacts in the last 12 months), total trans_Ct (Total Transaction Count (Last 12 months)) and Total_Ct_Chng_Q4_Q1 (Change in Transaction Count (Q4 over Q1)) tend to have a significant impact on the bank losing credit card customers. In addition, the estimated coefficients of gender, dependent count, marital status, months inactive in 12 months (No. of months inactive in the last 12 months), and contacts count in 12 months (No. of Contacts in the last 12 months) are negative while the estimated coefficients of total relationship count (Total no. of products held by the customer), total trans_Ct (Total Transaction Count (Last 12 months)) and Total_Ct_Chng_Q4_Q1 (Change in Transaction Count (Q4 over Q1)) are positive. These results obtained have relatively reliable since most of the assumptions of the multivariate regression model have been proved to meet.

References

- [1] Steve LaValle, E. L. (2010, December 21). *Big Data, analytics, and the path from insights to value*. MIT Sloan Management Review. Retrieved March 30, 2022, from <https://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/>
- [2] Zhong, Q., Liang, S., Cui, L., Chan, H. K., & Qiu, Y. (2019, June 3). *Using online reviews to explore consumer purchasing behavior in different cultural settings*, *Kubernetes - X-MOL*. X-MOL. Retrieved March 30, 2022, from <https://www.x-mol.com/paper/1368397301464850432?adv>
- [3] Butaru, F., Chen, Q., Clark, B., Das, S., Lo, A. W., & Siddique, A. (2016, August 8). *Risk and risk management in the credit card industry*. *Journal of Banking & Finance*. Retrieved March 30, 2022, from <https://www.sciencedirect.com/science/article/pii/S0378426616301340>
- [4] Agarwal, S., & Zhang, J. (2015, October 19). *A review of credit card literature: Perspectives from ...* Retrieved March 30, 2022, from <https://www.fca.org.uk/publication/market-studies/review-credit-card-literature.pdf>
- [5] Dornadula, V. N., & Geetha, S. (2020, February 27). *Credit card fraud detection using machine learning algorithms*. *Procedia Computer Science*. Retrieved March 30, 2022, from <https://www.sciencedirect.com/science/article/pii/S187705092030065X>
- [6] Liu, C., & Wang, M. (2018, October 19). *New ways to pay make life easier on the road*. *Chi-nadaily.com.cn*. Retrieved March 30, 2022, from <http://global.chinadaily.com.cn/a/201810/19/WS5bc93a59a310eff303283525.html>
- [7] Donou-Adonsou, F., & Basnet, H. C. (2019, March 19). *Credit card delinquency: How much is the internet to blame?* *The North American Journal of Economics and Finance*. Retrieved March 30, 2022, from <https://www.sciencedirect.com/science/article/abs/pii/S1062940818304868>
- [8] Sariannidis, N., Papadakis, S., Garefalakis, A., Lemonakis, C., & Kyriaki-Argyro, T. (2020, March 15). *Default avoidance on credit card portfolios using accounting, Demographical and exploratory factors: Decision making based on machine learning (ML) techniques - an-nals of operations research*. SpringerLink. Retrieved March 30, 2022, from <https://link.springer.com/article/10.1007/s10479-019-03188-0>
- [9] Suominen, K. (2017, January 1). *Accelerating digital trade in Latin America and the Carib-bean*. EconStor. Retrieved March 30, 2022, from <https://www.econstor.eu/handle/10419/173858>
- [10] Darjana, D., Wiryono, S. K., & Koesrindartoto, D. P. (2022, May 31). *The COVID-19 pan-demic impact on banking sector: Published in Asian Economics Letters*. *Asian Economics Letters*. Retrieved March 30, 2022, from <https://doi.org/10.46557/001c.29955>