Proceedings of the 2023 International Conference on Management Research and Economic Development
DOI: 10.54254/2754-1169/24/20230454

Root Cause Analysis of Credit Card Fraud

Yutong Zhang[1,a,*]

[1]*School of Mathematical Sciences, University of Sichuan Normal University, Chengdu, 610066, China*

a. 631952046@qq.com

**corresponding author*

Abstract: Fraud detection has always been a major problem in society, especially in the banking, insurance and medical sectors. As more and more people are using credit cards to pay, the fraud rates tend to get higher. This situation is a considerable risk for the credit card company. At the same time, credit cards and their related industries have become an indispensable part of the economic construction of various countries. Therefore through descriptive analysis, regression and factor analysis, this paper analyzes the root cause of fraudulent transactions and makes simple predictions. Discusses the characteristics of cardholders and influencing factors of fraudulent transactions, The subjective psychological factors have the biggest influence on fraudulent transactions and puts forward some suggestions to prevent fraudulent transactions.

Keywords: credit card fraud, logistic regression, characteristics of the cardholder

1. Introduction

Since the 1960s, the credit card industry in Europe, the United States, Japan and other capitalist countries has established a social personal credit system in the long-term development, and commercial banks have also established an anti-fraud system. Now Europe and some countries in order to deal with the credit card fraudulent transactions, has formulated the corresponding measures and methods. With the deepening of the development of the credit card industry, the industry risks are increasing, and the types and forms of risks are diversified and complicated. From the perspective of international market, credit card business is the focus of most of the big international banks and the main source of income. However, for the credit card issuing banks with low risk management level, if the risk is not well controlled, the credit card business will become the trigger for their profit decline or even loss.

Credit card fraud refers to the intentional use of forged, invalid credit card, using others' credit card defraud property, or a malicious overdraft related to the credit cards. Common credit card fraud including card fraud, fake application, fake credit card [1]. Nowadays, more and more people will use credit card to pay, convenient and quick, also can consume in advance, but it also has huge potential risks. Therefore, the risk of credit card fraud should be paid great attention to people.

2. Literature Review

In 2012, the global plastic card business reached $21.60 trillion. The losses related to the fraud were approximately $11.27 billion, or $5.22 ¢ per $100 (HSN Consultants Inc., 2013). However, this

DOI: 10.54254/2754-1169/24/20230454

number is growing rapidly, the total fraud losses have climbed to $56 billion by 2020 (Business Wire). The losses caused by credit card fraud are huge, and banks must reduce their own losses as a long-term goal [2].

In recent years, many scholars and experts at home and abroad are constantly exploring the field of anti-credit card fraud, and the comprehensive anti-credit card fraud research has not stopped. Based on spending statistics for holidays such as Christmas, New Year's and other religious holidays, cardholders spend more on those holidays [3,4]. Hongbo Shen, Hui Huang and Li Liao put forward relevant suggestions on the development of credit card consumption to accurately locate customer groups and understand customer needs. Different consumer groups have inconsistent demands for credit cards. For example, there are significant differences between women and men in their requirements for credit card cards [5]. Deng Ran believes that the regression results between the characteristics of credit card holders and the degree of credit card overdraft show that the younger people are, the higher the amount of credit card debt, and the older people are, the smaller the amount of credit card debt [6]. Across a stratified sample of Malaysians, age, previous credit card holdings, loan commitments, bad debt history and current account ownership influence the probability and level of credit card debt [7].

This shows that the fraud of credit card has a certain degree of recognition, and there are certain differences between fraud and non-fraud. Now most of the studies have conducted the influence of personal factors on credit card transactions. This paper analyzes the influence of external factors, such as the distance from home, the distance between two transactions, and the amount of transactions on credit card transactions.

Chapter 1, the Background. Introduce the research background of the credit card. Chapter 2, The Literature Review. This paper introduces the status of domestic and foreign research on credit card fraud, as well as the main work. Chapter 3, the basic analysis of credit card fraud characteristics. The data set was analyzed to obtain the basic characteristics of the cardholders. Chapter 4, cardholder characteristics analysis and prediction based on logistic regression. Chapter 5, the evaluation index of credit card fraud transactions. The factors affecting the fraudulent transactions in the credit card are obtained through the factor analysis. Chapter 6, Conclusion. Make identification and prevention recommendations for credit card fraudulent transactions.

3. Descriptive Statistical Analysis

3.1. The Crosstab of Fraud and Used_Pin_Number

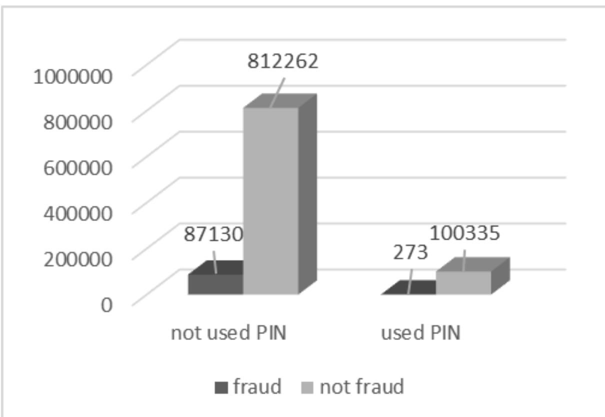

Figure 1: Fraud*PIN_number.

As Figure 1 shows, commercial banks should note when trying fraudulent transactions that cardholders will not use PIN code generally, but it does not rule out that transactions using PIN code are fraudulent. Similarly, it can be seen from Figure 1 that most cardholders do not set the PIN code. PIN code, full name is Personal Identification Number, is a personal identification number. It is similar to the password to prevent the leakage of credit card information from endangering the users' personal privacy. In China, there is usually no PIN code for credit cards, and there is no need to input during consumption. If you want to set it up, you can contact the card issuing bank. Now the credit card PIN code is more used in foreign countries.

3.2. The Crosstab of Fraud and Online Order

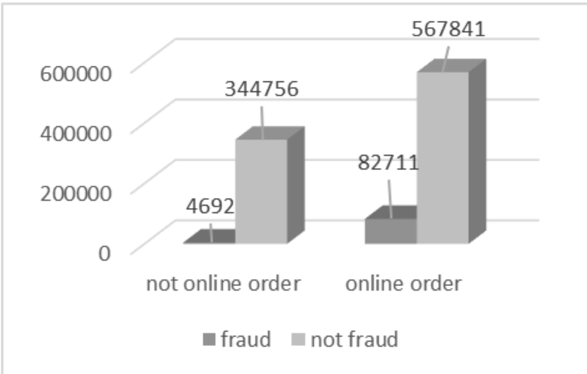

Figure 2: Fraud*Online_order.

According to Figure 2, the number of cardholders making fraudulent transactions online is about 17.63 times that of nononline transactions. Nowadays, many criminals conduct a credit card fraud by means of payment without card, this method uses credit cards only through channels such as phone, email and the Internet that do not need to show the real card. Criminals only need to provide the illegally obtained other people's credit card number, name, validity period and other information, so that they can make fraudulent transactions, causing huge personal and banking losses.

3.3. Compare Means

Table 1: Descriptive statistical analysis.

	Fraud	Average	Number of cases
Distance from home	0	22.83	912597
	1	66.26	87403
Distance from last transaction	0	4.30	912597
	1	12.71	87403
Ratio to median purchase price	0	1.42	912597
	1	6.01	87403

3.3.1. Distribution of distance from home

According to Table 1, it can be seen that the average distance from the home is 66.26, and the non-fraudulent transaction is 22.83. In fraudulent transactions, cardholders usually choose places relatively far from home. In criminal psychology, criminals are more likely to reveal their whereabouts if they commit near their hiding place than if they commit relatively far away This view also applies to fraudulent transactions.

DOI: 10.54254/2754-1169/24/20230454

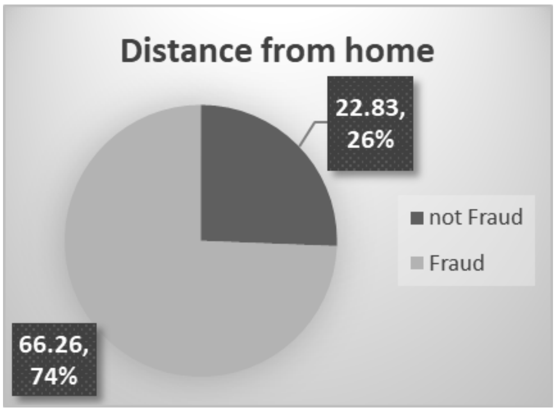

Figure 3: The distribution plot of distance_from_home.

3.3.2. Distribution of distance from last transaction

As Table 1 shows, the average distance between fraudulent transactions and the last transaction was 12.71, while the non-fraudulent transaction was 4.30, and fraudulent transactions tended to be longer from the previous transaction. In daily life, more and more places need to use credit cards to pay, and people's consumption frequency is also increasing. So when a card is not used for a long time, Some banks consider the security of the card, will take control measures against it.

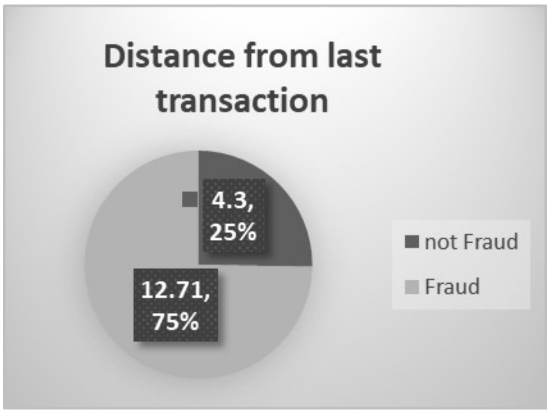

Figure 4: The distribution plot of distance_from_last_transaction.

3.3.3. Distribution of ratio to median purchase price

Finally, the average ratio of the fraudulent purchase price to the median purchase price is 6.01, much higher than the price of ordinary transactions. This suggests that fraudulent transactions may occur when large payments occur. At present, if a large transaction occurs in China, the financial institution is required to perform the following operations: The large transfer payment is reported by the financial institution through the relevant system and the payment transaction monitoring system. And report to the head office of the People's Bank of China on the second working day after the transaction occurs. Large cash receipts and payments are reported by the financial institution through its business processing system or in writing. The business shall be submitted to the local branch of the People's Bank of China within the second working day from the date of occurrence, and the local branch shall deliver the business to the head office of the People's Bank of China.

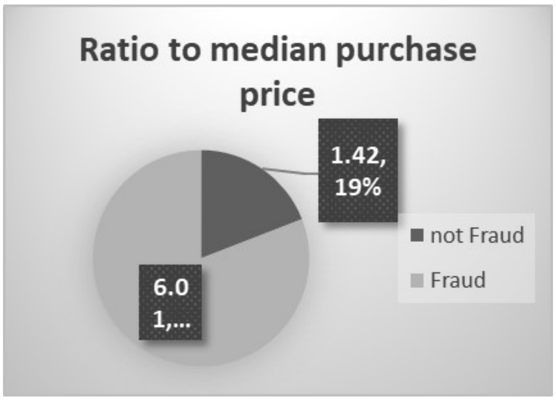

Figure 5: The distribution plot of Ratio to median purchase price.

4. Cardholder Characteristics Analysis Based on Logistic Regression

Whether the cardholder makes fraudulent transactions, and the behavioral characteristics of fraudulent transactions, influenced by a variety of factors, so this section try to establish a Logistic regression model, examining the influence of various factors on fraudulent transactions. By establishing the Logistic regression model, the quantitative analysis of the selected independent variables is conducted, and the influence rules on fraudulent transactions is obtained, and finally the characteristics of fraudulent transaction groups are described.

4.1. Independent Variables Were Determined Based on the Independent Sample T-test

Refer to Table 2, all variables were first tested for the independent-sample T-test and the statistically significant differences between fraudulent and non-fraudulent transactions were included: Distance from home, Distance from last transaction, Ratio to median purchase price, repeat_retailer, used_chip, used_pin_number, online_order.

Table 2: Independent-sample T-Test.

	F-value	significance	T-value
Distance from home	71930.84	0.000	-190.96
Distance from last transaction	23489.45	0.000	-92.31
Ratio to median purchase price	102322.20	0.000	-521.36
repeat_retailer	7.346	.007	1.357
used_chip	24292.070	.000	61.088
used_pin_number	52436.517	.000	100.801
online_order	662273.495	.000	-195.611

4.2. Test of the Model

4.2.1. Multicolliearity

Multicollinearity diagnosis was performed on the independent variables that passed the significance test. When the variance of regression coefficient is estimated by the least square method, the result of collinearity between independent variables is larger than that of no collinearity between independent variables. If the value of VIF is large, the degree of multicollinearity between variables is strong. When the value reaches 10, it indicates that there is serious collinearity between the independent variables, but the critical value will be different in different specific circumstances.

DOI: 10.54254/2754-1169/24/20230454

Refer to Table 3,We can see that the VIF index of all independent variables is less than 5, and it can be initially considered that the collinearity problem can be ignored. Therefore, the Logistic regression model can be used to explore the influence law on fraudulent transactions.

Table 3: Multicollinearity diagnosis.

	tolerance	VIF
Distance from home	.980	1.021
Distance from last transaction	1.000	1.000
Ratio to median purchase price	1.000	1.000
Repeat retailer	.980	1.021
Used chip	1.000	1.000
Used pin number	1.000	1.000
Online order	1.000	1.000

4.2.2. Omnibus Tests of Model Coefficients

In the comprehensive test of Model coefficients, the likelihood ratio test results of whether all parameters in the Logistic regression model are 0 are output in the model line. It can be seen from Table 4 that P = 0.000< 0.05, indicates that the OR value of at least one variable in the fitted model is statistically significant, that is, the model is generally significant.

Table 4: Omnibus tests of Model Coefficients.

Test	Chi-square	Significance
step	321585.167	7
model	321585.167	7

4.3. The Solution Results of This Model

With distance from home, distance from the last transaction, ratio and median purchase price, repeat retailers, use of chips, used pin number and online order as independent variables, and whether the transaction is fraudulent or not as dependent variables, the logistic regression model set in this paper is as follows:

$$P(Y|X=1)=\frac{exp(\beta 0+\beta 1X1+\cdots+\beta 5X5)}{1+exp(\beta 0+\beta 1X1+\cdots+\beta 5X5)} \quad (1)$$

$$P(Y|X=0)=1-\frac{exp(\beta 0+\beta 1X1+\cdots+\beta 5X5)}{1+exp(\beta 0+\beta 1X1+\cdots+\beta 5X5)}=\frac{1}{1+exp(\beta 0+\beta 1X1+\cdots+\beta 5X5)} \quad (2)$$

Among them, $P(Y|X=1)$ is recorded as the probability of fraudulent transaction, X_i is the independent variable (explanatory variable), β_i ------ Represents the contribution of the corresponding independent variable, β_0 ------ intercept term。

In this paper, the input method is selected, the significance level is set as 0.05, the cut-off value is set as 0.2, and the regression results are as follows

DOI: 10.54254/2754-1169/24/20230454

Table 5: Binary logistic regression model estimated results.

Independent variable name	coefficient (B)	standard error	Wald	significance	EXP (B)
distance_from_home	.015	.000	33427.188	.000	1.015
distance_from_last_transaction	.025	.000	11154.315	.000	1.026
ratio_to_median_purchase_price	.862	.003	92247.494	.000	2.368
repeat_retailer (1)	-.621	.016	1556.786	.000	.537
used_chip (1)	-1.049	.012	7362.279	.000	.350
used_pin_number (1)	-13.740	.159	7510.555	.000	.000
online_order (1)	6.651	.037	31938.137	.000	773.851
constant	-10.361	.044	56263.895	.000	.000

4.4. Analysis of Model Results

4.4.1. Forecast results

Table 6: Confusion matrix.

Confusion matrix		True value		Correct percentage
		Positive	Negative	
Predicted value	Positive	878813	33784	96.3%
	Negative	12781	74622	85.4%
Overall percentage				95.3%

According to Table 6 the Accuracy (ACC) of the model is 95.3%, Precision (PPV) is 96.3%, Sensitivity (TPR) or Recall is 98.6%, and Specificity (TNR) is 68.8%.

$$\text{F1 Score} = \frac{\text{2PrecisionRecall}}{Precision + Recall} \tag{3}$$

Therefore F1-Score is 0.97, indicating that this model is enough to get a sufficiently accurate prediction.

4.4.2. Feature analysis

The P value of the above independent variables is less than 0.05, that is, the influence of the independent variables is significant, so the following conclusions can be obtained:

First, the regression coefficient value of distance _from_ home is 0.015. The distance between the cardholder's transaction and his home will have a significant positive impact on whether it is a fraudulent transaction. Specifically, the odds ratio (OR value) is 1.015, which means that the increase of Y is 1.015 times higher.

Next, which means that the time interval between the cardholders' two transactions will have a significant positive impact on whether it is fraudulent. And an odds ratio (OR) of 1.026, which means that when the transaction interval increases by one unit, Y increases by 1.026 times.

Then, the regression coefficient value of ratio_to_median_purchase_price is 0.862, that is, the larger the price of the last one transaction than the previous transaction, the more likely the transaction is to be fraudulent. For each unit of ratio _ to _ median _ median _ purchase _ price, the increase of Y is 2.368 times.

Last, EXP (B) is the incidence rate, which is interpreted here as the ratio of the probability of fraudulent transactions to the probability of not occurring, then the probability of fraudulent transaction occurring in the same merchant is 0.537 times of the probability of non-same merchant. Transactions with cards are 0.350 times more likely to be fraudulent than transactions without cards. Online transactions were 773.851 times more likely to be fraudulent than offline orders.

5. Evaluation Indicators of Credit Card Fraud Transactions

5.1. Test of the Model

In this section, we adopt the method of factor analysis and utilize the idea of dimension reduction. According to the internal dependence of the correlation matrix of the original variables, the original variables are grouped by correlation, so that the correlation between variables in the same group is high, while the correlation between variables in different groups is low.

Table 7: KMO and Bartlett tests.

KMO Measure of Sampling Adequacy		0.500
Bartlett sphericity test	Approximate chi square	20718.599
	Df	21
	P	.000

According to Table 7, the KMO test result shows that the value of KMO is 0.500; meanwhile, the Bartlett sphericity test result shows that the significance P value is 0.000, showing significance at the level. Therefore, the null hypothesis is rejected, indicating that there is correlation between variables, and the data can be factor analysis.

5.2. The Number of the Extracted Factors

According to Figure 6, starting from the sixth principal component, the characteristic root value of the principal component begins to decline rapidly.

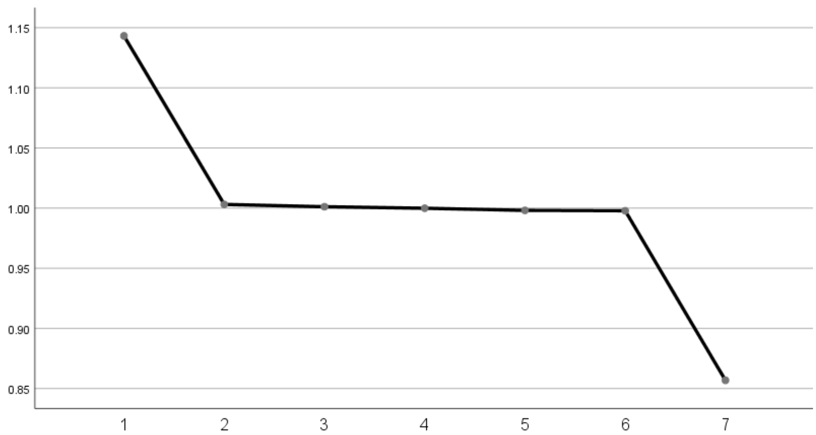

Figure 6: Lithotriptic diagram.

As shown in Table 8, under the condition that the contribution degree of factors' cumulative interpretation reaches 85%, we can choose to retain the six principal components.

DOI: 10.54254/2754-1169/24/20230454

Table 8: Total variance interpretation.

	Initial eigenvalue			Extracting sum of squares of loads		
component	total	variance percentage	cumulation %	total	variance percentage	cumulation %
1	1.143	16.331	16.331	1.143	16.331	16.331
2	1.003	14.329	30.660	1.003	14.329	30.660
3	1.001	14.302	44.962	1.001	14.302	44.962
4	1.000	14.284	59.247	1.000	14.284	59.247
5	.998	14.259	73.505	.998	14.259	73.505
6	.998	14.254	87.759	.998	14.254	87.759
7	.857	12.241	100.000			

5.3. Factor Score

Table 9: The component score coefficient matrix.

	component					
	1	2	3	4	5	6
Distance from home (A1)	.661	.007	-.006	.001	.012	.005
Distance from last transaction9 (A2)	-.003	.616	.151	.161	.483	-.582
Ratio to median purchase price (A3)	.000	.195	.729	.258	-.602	-.031
Repeat retailer (A4)	.661	.000	.010	.012	.001	.009
Used chip (A5)	-.009	.638	.003	-.115	.124	.751
Used pin number (A6)	-.009	-.415	.573	.179	.625	.276
Online order (A7)	-.008	-.002	-.339	.928	-.022	.149

With the continuous progress of the era of big data, factor analysis is widely used, and it can study a variety of objects, and can better apply statistical methods to the evaluation of financial-related indicators. Factor analysis can reduce the dimension of large sample data and simplify the design weight of the data. Even with a large sample size, the dimensionality of data can be reduced through factor analysis. It not only avoids a lot of complicated calculation to reduce the difficulty of calculation, but also solves the overlapping problem of factor information [8–12]. In this section, Table 8 shows the component score coefficient matrix obtained by SPSS25.0. When interpreting principal components, small coefficient (absolute values less than 0.6) is excluded. Therefore, we can obtain the scoring expression of F1, F2, F3, F4, F5, F6 and its common factors as:

$$F1 = 0.661A1 + 0.661A4,$$
$$F2 = 0.616A2 + 0.638A5,$$
$$F3 = 0.729A3,$$
$$F4 = 0.928A7,$$
$$F5 = -0.602A3 + 0.625A6,$$
$$F6 = 0.751A5$$

In this paper, F1 is interpreted as a psychological factor, and the transaction place chosen by the cardholder and whether the transaction is conducted for the same cardholder have equal effects to

DOI: 10.54254/2754-1169/24/20230454

the psychological factor, and the subjective psychological factors have the greatest influence on fraudulent transactions. Interpreting F2 as the trading status factor, whether the chip card trades affect this factor most. Interpreting F3 as the price fluctuation factor and F4 as the trading route factor. F5 is interpreted as a security awareness factor, and whether to use PIN code has the greatest impact on this factor, and the more abnormal the transaction price is, the weaker the cardholders' security awareness is. Interpreting F6 as a trading mode factor has a minimal impact on fraudulent transactions.

6. Conclusion

Today is the rapid development of the Internet era, credit card fraud has new characteristics and manifestations. Through the characteristic analysis of cardholders in this paper, credit card fraud is showing new features such as networking and novelty. These new and general features are intertwined, so on the basis of external supervision, doing a good job of internal response of commercial banks can effectively reduce the occurrence of credit card fraud.

6.1. Improve the Construction of the Anti-credit Card Fraud System

Improve the relevant laws and regulations of credit card fraud, and integrate the risk investigation of the industry into the regulatory system. Now in the era of Internet big data, credit card fraud and telecom fraud cases are still in high incidence, so the whole society should strengthen the research on the characteristics of the Internet. Cultivate talents in the field of Internet risk prevention, and continue to invest in research and development funds. At the same time, the public security departments should strengthen the cooperation with the industry regulatory authorities, the Internet service provider companies, and the big data analysis companies.

6.2. Strengthen Dynamic Monitoring and Risk Control

Because of Credit card’s special transaction form, cardholders are prone to impulse spending and overspend. At present, money laundering risk and fraud risk coexist. As a financial institution, it is necessary to constantly improve its own prevention system construction. First of all, conduct risk assessment before applying for credit cards, and do a good job in the first step of prevention. For large consumption, financial institutions should focus on the investigation. At the same time, the occurrence of multiple micropayments or abnormal transaction time is also worth the attention of financial institutions. Once abnormal credit card transactions are found, they should be immediately controll to reduce personal and social losses as far as possible.

6.3. Popularize Personal Card Safety Knowledge and Anti-fraud Knowledge

Preventing credit card fraud is not only the duty of the public security and relevant institutions, but also the obligation of the cardholder himself. As a cardholder, the main purpose is to ensure that your credit card is not lost and protect your personally identifiable information. If the credit card is lost accidentally, it should report the loss timely and freeze the account funds in time. At the same time, do not easily inform or fill in personal information, there are many places to fill in identity information, and people need to make sure that this is a legal channel before filling in. Furthermore, the public should understand the role of the PIN code and use the PIN as much as possible. Although it is cumbersome, cardholders can better protect their property safety and reduce personal losses. At the same time, if the cardholder wants to avoid the bank card stolen brush, it is best to call the bank to close the free secret payment function, or will reduce the limit, in case of stolen brush can reduce the loss. In addition, many third-party payment applications can open the secret

DOI: 10.54254/2754-1169/24/20230454

payment function, it is best to turn it off. It is important to improve the convenience, but how to balance the convenience and security of payment is worth thinking about by many payment institutions and cardholders.

In practice, however, normal transactions are far more than fraudulent ones. This paper only analyzes the characteristics of cardholders and establishes the prediction model, but does not solve the problem of sample imbalance, which can be used as a follow-up research direction to establish a more perfect credit card fraud detection model.

References

[1] Yihao Zhang, Danhong Sheng, Lifang Li, et al. Application and Research on Weighted Random Forest [J]. Computer Programming Skills and Maintenance, 2021 (4): 111-112

[2] HSN Consultants Inc. (2013). The Nilson report. Oxnard, CA: Author.

[3] Shen, A. (2011). INFOGRAPHIC: Americans are spending a whopping $704.18 on gifts this year. Retrieved December 6, 2015, from http://www. businessinsider.com/what-americans-spend-on-christmas-2011-12

[4] Visa Europe. (2011). Irish consumers to spend €257 million on Christmas nternet shopping. Retrieved December 6, 2015, from http://www.visa.ie/about-us/press-releases/irish-consumers-to-spend-257-million-onchristmas-internet-shopping

[5] Hongbo Shen, Hui Huang & Li Liao. (2013). Study on the overdraft behavior of cardholders in China's credit card market. Statistical study (10), 61-67. doi: 10.19343/j.cnki.11-1302/c.2013.10. 009.

[6] Ran Deng. (2018). Research on the impact of exogenous impact on cardholders' credit card consumption behavior (doctoral dissertation, Central University of Finance and Economics). https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CDFDLAST2019&filename=1019652399.nh

[7] Andrew K.G. Tan,Steven T. Yen & Yiing Jia Loke.(2019).Credit Card Holders, Convenience Users and Revolvers: A Tobit Model with Binary Selection and Ordinal Treatment. Journal of Applied Economics (2). doi:12043395.

[8] Oyekunle Janet Olufunmike; Ayoola Joshua Femi; Oyenuga Iyabode Favour; Masopa Adekunle Nurudeen; Adesiyan Adefowope Abdul Azeez Estimation of CPI of Some Sub-sahara African Countries: Panel Data Approach [J] American Journal of Theoretical and Applied Statistics,2019

[9] D. Gao, B. Zhang, and S. Li, "Spatial effect analysis of total factor productivity and forestry economic growth," Forests, vol. 12, no. 6, p. 702, 2021.

[10] B. Qla and B. Ms, "BALFA: a brain storm optimization-based adaptive latent factor analysis model," Information Sciences, vol. 578, pp. 913–929, 2021.

[11] W. Fukumoto, H. Mitani, Y. Kuno et al., "Incidence and factor analysis of laryngohyoid fractures in hanging individuals—computed tomography study," European Radiology, vol. 31, no. 10, pp. 7827–7833, 2021.

[12] Zhao Qiang. (2022). Quality Analysis and Key Factor Research in Japan's Economic Growth Based on Factor Analysis. Journal of Sensors. doi:10.1155/2022/9976494.