

ISSN Online: 2329-3292 ISSN Print: 2329-3284

Prediction of Default Probability of Credit-Card Bills

Yuhan Ma

University of Wisconsin-Madison, Madison, WI, USA Email: mayuhan190716@outlook.com

How to cite this paper: Ma, Y.H. (2020) Prediction of Default Probability of Credit-Card Bills. *Open Journal of Business and Management*, **8**, 231-244. https://doi.org/10.4236/ojbm.2020.81014

Received: October 31, 2019 Accepted: December 24, 2019 Published: December 27, 2019

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

http://creativecommons.org/licenses/by/4.0/





Abstract

The credit-card industry has existed for decades and is a product both of changing consumer habits and improved national incomes. Both the number of card issuers and issuing banks, and transaction volumes themselves, have increased significantly. Nonetheless, with the increase of credit-card transactions, overdue amounts and delinquency rates of credit-card loans have also become problems that cannot be ignored. The successful resolution of this issue is central to the successful future development of the industry. In this work, we have presented a credit-score model to reflect the attributes and credit ratings of clients. We merge 23 variables from the original dataset and 25 additional financial features that are mined from the original financial variables, then apply to the XGBoost model. The model itself provides the 13 most significant variables by listing them according to the calculated scores. It then predicts the probability of individuals' willingness to pay back a credit-card loan. At last, the default ratio will be converted to a credit-score system to understand the credit ratings of clients more intuitively. This model can make contributions to the resolution of default probability and is very helpful to the credit-card industry.

Keywords

Credit Card Industry, Default Probability, XGBoost, Data Mining

1. Introduction

Credit-card debts are among the most significant assets of many banks, and in the modern era, credit-card ownership is exceptionally widespread. People often prefer to obtain "advanced consumption" as well as loans from their financial institutions. The percentage of owning General-purpose credit cards such as Visa, MasterCard and Discover is 70.2 percent of all American households,

based on information from the U.S. Census Bureau [1]. Moreover, halfway through 2019, U.S. credit-card debt had already reached \$971 billion. This represented a 5% increase over 2018, reflecting a general upward trend [2]. Unlike large loans, such as mortgage loans and car loans, using a credit card to purchase goods is a more widespread and routine phenomenon in daily life.

On average, the revolving balance of one American household is approximately \$6829. Hence, finding a reliable means to assess clients' ability to repay credit card debt (and thus avoid costly defaults) is an essential task for banks. Who can be trusted with a card, for example, and what should a client's credit limit be? According to the Board of Governors of the Federal Reserve System, the charge-off rate in 2019 for all commercial banks in the United States is about 3.7% [2]. Obviously, these institutions should deploy every possible resource to predict and forestall default behaviour.

Currently, loan officers decide whether to approve applications based on credit history and personal experience. Nonetheless, these criteria are far from perfect. First, not every applicant has a credit history, so this approach will not work when these applicants ask for credit cards. Second, loan officers are often subjective: the standards applied by one are not necessarily those of another. In other words, their decisions can be heavily influenced by personal factors. Despite the acceptance that predicting credit-card default is highly significant for banks, at present most research papers focus on financial policies and banking regulation, to examine how these policies impact default probability. Research prioritizes theoretical derivation but pays insufficient attention to rigorous mathematical analysis. Some other research, meanwhile, relates loan default to "machine intelligence". Student loans and mortgages often receive more intensive scrutiny than credit-card debt. Therefore, studying credit-card charge-off rates through mathematical models seems both timely and relevant.

This paper will address the primary contributing factors to default probability by means of the XGBoost model. In addition, it predicts the probability of the default of a credit-card holder through the relevant information pertaining to that individual, then converts the probability into an intuitive credit score through "certain score mapping". The attributes and credit ratings of clients may be intuitively reflected in credit scores. Obviously, the higher the score, the better the credit; the lower the score, the lower the credit rating.

Moreover, according to the credit scores of users, credit agencies can implement relevant management measures, including reminders, collection procedures and other strategies. Clients with lower scores could be issued with payment reminders, while higher-scoring clients need not receive special attention. Institutions may take "further measures" with those with very low scores. Following such a strategy would allow banks and others to manage their credit exposure far more efficiently.

Research Problem

Credit cards are the growing trend in the market and are quickly replacing the

other mediums of transactions in the face of the market economy. With the ever-increasing usage of the medium, the constant threat of default is on the continuous rise. The credit card issuers are in a constant dilemma to recognize the key drivers that will represent the likelihood of default on the path of the customers. The decision to allocate the credit limit towards the individual customers is yet another major source of the dilemma faced by the issuers. Off lately, the market of credit card has expanded massively, bringing about the massive number of defaults. The credit card debt in the USA is more than ever putting pressure on the banks and other financial institutions to come up with a mechanism to understand the potential customers and reduce the risk of debt. The issuers need to come up with a mechanism that can evaluate the probability of default from the consumer end.

Research Objectives

The research aims at developing a mechanism to predict the credit card default beforehand and to identify the potential customer base that can be offered various credit instruments so as to invite minimum default.

2. Literature Review

2.1. Asymmetric Information Theory

According to information economics, if the information related to a transaction is asymmetrical, the progenitors of "adverse selection" and "moral hazard" will arise, which are important factors for the formation of default risk. Research by American economists Stiglitz and Weiss showed that adverse selection and improper incentives always exist in the credit market [3]. In terms of information, broadly speaking, banks and other lenders are in a position of relative weakness while borrowers are in a position of comparative strength. Banks often do not know the borrower's repayment motivation, repayment ability and "project risk" [4]. It is hardly a secret that, to obtain a short-term advantage, certain borrowers distort or conceal negative information in their dealings with lenders; more positive "information" may simply be invented. Financial institutions are aware of their information disadvantage. As a result, they tend to charge higher interest rates, which crowd out creditworthy borrowers, while those who continue actively looking for loans are more likely to represent potential non-performing loan risks. In other words, good clients can be squeezed out of the credit market by poor ones.

On the other hand, when financial institutions do provide borrowers with funds, those borrowers may have a strong motivation to hide certain facts of operation and profitability in order to obtain economic benefit. If the borrower uses the loan for high-risk investment projects after the loan contract is signed, once that project fails, he/she transfers the risk to the bank. The borrower may simply refuse to pay either principal or interest, thus effectively avoiding the debt obligation. As a result of this "moral hazard", the financial institution must absorb the loss of both interest and principal.

2.2. Previous-Study Review

In 2009, Yeh and Lien applied a standardised dataset and published a research paper, which compared the prediction accuracy of credit-card clients' default probability in terms of six data-mining techniques: these comprised K-nearest neighbour (KNN) classifiers, logistic regression (LR), discriminant analysis (DA), naive Bayesian (NB) classifiers, artificial neural networks (ANNs) and classification trees (CTs) [5]. In terms of error rates, the study detected no significant difference between the six. There was, conversely, a pronounced difference regarding area ratios. More importantly, artificial neural networks (ANNs) executed classification more accurately than the other five methods. A subsequent study (2016) was undertaken by Vendakesh and Jacob to assess the comparative performance of classifiers [6]. They found that Random Tree, Random Forest and IBK obtained good rates of accuracy.

Currently, papers addressing the application of machine learning to credit-card default tend to pay more attention to comparing machine-learning models to see which models exhibit better accuracy. Therefore, in this paper, instead of comparing different methods, we will instead merely apply one currently popular technique for mining data, to see how it can provide information about credit-card default probability.

As per **Figure 1** presented above, the credit card delinquency rates are showing an upward trend in both the small and large banks. However, the ratio seems to be more for smaller banks in comparison to the larger banks. It is mainly due to the credit mapping practices, and default mapping is more stringent in larger banks as compared to the smaller banks [7]. It has also been observed that the smaller banks often provide risky credits to attract and lure mass consumer base, often letting the security protocols aloof for short-term gains. This eventually leads to long-term debts.

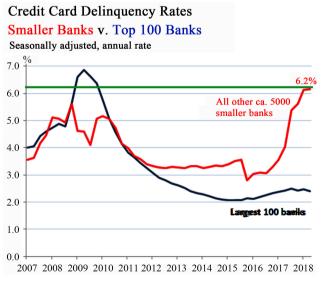


Figure 1. Credit card delinquency rate [7].

In a survey of banks and credit unions, it has been evidently stated that institutions have suffered extreme losses due to the credit card frauds. 2015 and 2016 have been taken as two base years to compare the scams arising due to credit cards. As the results shown in **Figure 2**, the survey indicates that the frauds due to credit card have been more in 2016 as compared to 2015. It is a clear indication that the frauds arising in the credit card segment have been on the growing trend as its popularity [8].

Figure 3 represents the credit card volume in the USA, indicating that the general-purpose transactions in the nation are more in comparison to the private-label transactions. The transactions show a positive trend. The general-purpose value has increased significantly; however, the private-label value has also demonstrated a positive growth but not as much as compared to the former.

3. Methodology

3.1. Research Variables

The dataset [10] comprises 24 variables, including clients' personal information (gender, age, marital status, education level) and financial information such as account balance, monthly bills and repayment status. Default payment, a binary variable, is meanwhile deployed as the response variable. Detailed descriptions and summary statistics of the explanatory variables are presented in **Table 1** and **Table 2**.

The client's ability and willingness to repay are reflected in his/her repayment details, which are themselves merely part of the wider financial data. For example, if the user has paid on time for the past six months and can cover the current month's bills, this probably means that the user has a strong willingness to repay and strong repayment ability. Hence, there is obviously a good chance that he/she will repay next month. Conversely, if the client has delayed repayment once or more in the previous six months, this may bode ill for future repayments.

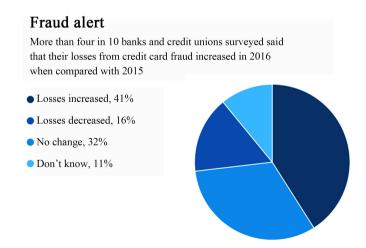


Figure 2. Credit card fraud index [8].

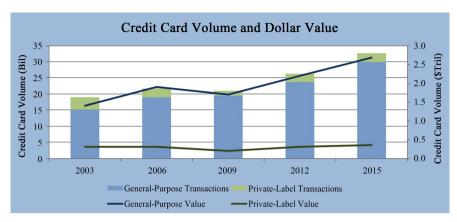


Figure 3. Credit card volume and dollar value [9].

Table 1. Description of explanatory variables in the dataset.

Number	Name	Description
1	Limit_Bal	The quantity of credit provided (NT dollars): this comprises both the individual's family credit and his/her consumer credit.
2	Sex	Gender (1 = male, 2 = female)
3	Education	Education level (1 = graduate school, 2 = university, 3 = high school, 0, 4, 5, 6 = others)
4	Marriage	Marital status (1 = married, 2 = single, 3 = divorce, 0 = others)
5	Age	Age (year)
6	Pay_0	Starting with repayment status as of September 2005 (0), six variables reflect past payment history. $(-2 = \text{no consumption}, -1 = \text{paid in full}, 0 = \text{the use of revolving credit}, 1 = \text{payment delay for one month}, 2 = \text{payment delay for two months},, 8 = \text{payment delay for eight months}, 9 = \text{payment delay for 9 months and above})$
7	Pay_2	The repayment status in August 2005
8	Pay_3	The repayment status in July 2005
9	Pay_4	The repayment status in June 2005
10	Pay_5	The repayment status in May 2005
11	Pay_6	The repayment status in April 2005
12	Bill_Amt1	Amount of bill statement (NT dollar) in September 2005
13	Bill_Amt2	Amount of bill statement (NT dollar) in August 2005
14	Bill_Amt3	Amount of bill statement (NT dollar) in July 2005
15	Bill_Amt4	Amount of bill statement (NT dollar) in June 2005
16	Bill_Amt5	Amount of bill statement (NT dollar) in May 2005
17	Bill_Amt6	Amount of bill statement (NT dollar) in April 2005
18	Pay_Amt1	Amount paid (NT dollar) in September 2005
10	Pay_Amt2	Amount paid (NT dollar) in August 2005
20	Pay_Amt3	Amount paid (NT dollar) in July 2005
21	Pay_Amt4	Amount paid (NT dollar) in June 2005
22	Pay_Amt5	Amount paid (NT dollar) in May 2005
23	Pay_Amt6	Amount paid (NT dollar) in April 2005

Table 2. Statistics summary of explanatory variables in the dataset.

Variable	Mean	Median	Min	Max
Limit_Bal	167,484	140,000	10,000	1,000,000
Sex	1.604	2	1	2
Education	1.853	2	0	6
Marriage	1.552	2	0	3
Age	35.49	34	21	79
Pay_0	-0.0167	0	-2	8
Pay_2	-0.1338	0	-2	8
Pay_3	-0.1662	0	-2	8
Pay_4	-0.2207	0	-2	8
Pay_5	-0.2662	0	-2	8
Pay_6	-0.2911	0	-2	8
Bill_Amt1	51,223	22,382	-165,580	964,511
Bill_Amt2	49,179	21,200	-69,777	983,931
Bill_Amt3	47,013	20,088	-157,264	1,664,089
Bill_Amt4	43,263	19,052	-170,000	891,586
Bill_Amt5	40,311	18,104	-81,334	927,171
Bill_Amt6	38,872	17,071	-339,603	961,664
Pay_Amt1	5664	2100	0	873,552
Pay_Amt2	5921	2009	0	1,684,259
Pay_Amt3	5226	1800	0	896,040
Pay_Amt4	4826	1500	0	621,000
Pay_Amt5	4799.4	1500	0	426,529
Pay_Amt6	5215.5	1500	0	528,666

These financial factors, rather than being separate features, are interrelated. Therefore, simply applying these variables to the model is not enough. This paper uses similar ideas to mine, in depth, the billing information and repayment information of sample users in the past six months to depict the default attributes that users may have.

The research variables can be divided into three types.

- 1) Proportional category: monthly payment amount/monthly bill amount, monthly bill amount/credit card amount/credit card amount. This category addresses both the usage rate of the amount in question and real repayment ability.
- 2) Frequency category: the frequency of having more than three months overdue, having more than four months overdue, and having more than five months overdue, no monthly consumption status, monthly full coverage of the number of bill repayments. The frequency of having monthly bills that exceed 8000 and more than 16,000, as well as the frequency of having monthly repayments that exceed 0. These data can be used to analyse clients' historical repayment behaviour.

3) Statistical category: maximum, minimum, average and uneven monthly bill amount; maximum, minimum, average and uneven monthly repayment amount.

In conclusion, a total of 42 connected variables were excavated, once feature engineering had been completed.

The research variables will aid in clarity of the purpose of the research. The individual variables will collectively help in understanding the credit viability of the customer. The personal information and the financial analysis of an individual will provide transparency to the issuers and institutions about the credit habits of the user. The personal data will provide demographic information about the customer, which will help in understanding and gaining a new customer base. Personal information is imperative in instances wherein the customer has no previous financial or credit record. As customer acquisition is critical to the organizations, the issuers scrutinize the personal information very minutely. Financial information puts the issuers at ease as the entire credit history of the customer can be viewed. The can check the credit ranking of the customer along with his default record and decide upon issuing the credit.

Issuing credit merely on the basis of personal information is more of a gamble for the issuers as the financial behavior of the customer is unknown, and hence, it poses a high risk. Therefore, these variables are scrutinized very stringently.

Research Design

A secondary research approach has been used by taking the already present information of the samples in the form of financial records and personal data. Exploratory research on the basis of the survey has been conducted. An inductive research format has been applied to gain maximum insight into the research problem. Definitive conclusions have been derived from the research with the help of quantitative data.

3.2. XGBoost Model

XGBoost is a machine-learning system centred on the "lifting tree" proposed by Chen (2016) [11], on the basis of a great deal of previous research work on the gradient lifting algorithm. It contains a set of iterated residual trees. Each tree is the residual of the N-1 tree before learning. Adding the output values of new samples predicted by each tree produces the predicted value of the final sample (2016). Nonetheless, unlike the commonly used gradient-boosting decision tree (GBDT), which only uses the first derivative information in optimisation, XGBoost expands the cost function in the second-order Taylor expansion and uses the first and second derivatives at the same time, which allows XGBoost to obtain credible results. Its main characteristics are as follows.

- 1) By supplementing the loss function with a regularisation term, XGBoost acquires a formidable anti-overfitting characteristic.
- 2) The second-order Taylor expansion is used to make the loss function more accurate.
 - 3) The efficacy of the model iteration is greatly enhanced by the qualities of

the parallel operation.

4) XGBoost supports column sampling, which reduces overfitting, reduces calculation and improves iteration efficiency.

Because of its advantageous algorithm, XGBoost has been deployed in the financial sector and elsewhere with increasing frequency in recent years. The present context allows its popularity and practicality to be more objectively assessed. Therefore, this paper takes XGBoost as its basic tool to complete the users' credit-card characteristic analysis.

4. Research Findings

4.1. Modelling

In this paper, a total of 30,000 samples of credit-card billing information and repayment information are modelled, together with some basic user information. The label is the repayment status of the user in October 2005. "1" indicates default and "0" indicates non-default. The distribution of repayment and non-repayment is shown in **Table 3**. As shown in this table, the default probability of the sample is 22.12%.

An outlier detection should be undertaken prior to merging the original basic data with the feature-engineering data and implementing the XGBoost model. The following three figures, **Figures 4-6** depict the distribution of several explanatory variables that possibly include outliers. Some variables do indeed reflect outliers that need to be discounted.

For LIMIT_BAL, more than 1,000,000 is an exception value. For BILL_AMT1 and BILL_AMT2, more than 1,000,000 is an abnormal value. For BILL_AMT3, more than 750,000 is an outlier. For PAY_AMT1, more than 75,000 is an abnormal value. For PAY_AMT2, more than 150,000 is an abnormal value. The remaining sample figure is 29,996, once the necessary exclusion of abnormal-value samples has taken place.

The training samples and test samples were randomly divided into train_data and test_data at a ratio of 8:2. 800 trees are selected in this case, due to the relatively small number of variables and samples. The maximum depth of the tree is set to 3, and the learning rate is 0.03. The final model on the test set has AUC score 0.779, where the Receiver Operating Characteristic (ROC) curve is represented in Figure 7.

As per the research conducted on the basis of the personal and the financial information of the customers. About 30,000 samples have been used a majority of which presented information about the financial information of the customers and a meagre amount of personal information about the customers [12]. The financial credentials of the customers played a pivotal role in estimating the default. The personal information of the customers did not play a much essential part; however, it helped in understanding the demographic data like the age, income, geographical location and employers' details of the defaulters [13]. As per the research methods applied, it can be seen that the willingness of the custom-

ers to pay the timely credits is on a positive end. An improvement in the debt payment has been established through the research. The personal data of the users did not help much in reaching this consensus.

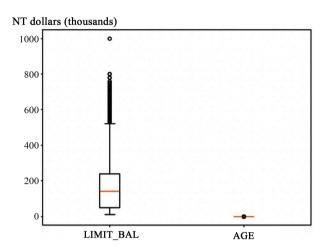


Figure 4. Outlier detection of the amount of given credit and age.

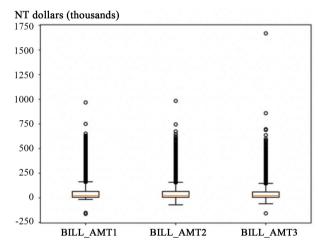


Figure 5. Outlier detection of bill statement in September, August, July 2005.

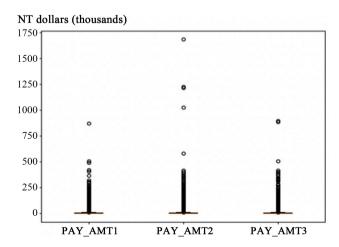


Figure 6. Outlier detection of the amount paid in September, August, July 2005.

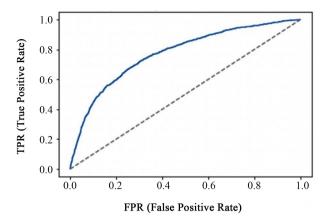


Figure 7. Receiver Operating Characteristic (ROC) Curve.

Table 3. Default probability from the collecting data.

Default Number	Not Default Number	Total Number	Default Ratio
6636	23,364	30,000	22.12%

The developed system of credit mapping intuitively estimates the default rate of the customers. The research by considering the five factors has clearly stated that the default of the credit is likely to reduce in the future. A high credit rating points towards low default and a low credit rating points towards a high default rate.

4.2. Results

The models above, and their associated results exhibit an improved distinguishing effect; the AUC is 0.779. Moreover, the model lists the features we applied in order of importance based on the index it calculated. **Table 4** provides the five most important variables. They are:

For the preceding six months, the number representing bill-repayment coverage reflects ability and willingness to repay: hence, this variable is highly significant. The ultimate goal of the model is to predict the default probability of users, but compared with the comparative model, the "probabilistic probability" paradigm cannot directly reflect the user's characteristics. Hence, in the context of a scorecard application, there is a particular mapping relationship between default probability and component numbers.

The research has been successful in utilizing the financial information of the customers and determining a credit rating mechanism. The rating will eventually help the issuers to check the credit score of the customers and evaluate their default status. The samples taken for the purpose of the research indicate a high credit ranking in the customers, which is a clear mark that the credit default is likely to decrease in the future. The sample size in however small and cannot be guaranteed, however, the similar mechanism can be used to check the feasibility of the credit and viability of the customer.

Table 4. Five most important variables affecting credit default probability.

Variables	Meaning
revolving_credit_cnt_recently_6m	The number for the full coverage of bills paid in the past six months
BILL_AMT2_vs_LIMIT_BAL_ration	The proportion of amount for bills to credit line in the preceding two months
BILL_AMT1	The amount relating to bills in the preceding month
PAY_AMT4_vs_BILL_AMT4	The proportion of the amount of repayment to the amount for bills in the past four months
PAY_AMT3_vs_BILL_AMT3	The proportion of the amount of repayment to the amount for bills in the past four months

There is a specific hypothesis:

Hypothesis: Assuming that the default normal ratio has an expected score and the default normal ratio doubles, the expected score is reduced by 50 points.

As follows:

$$Score = A - B * log(odds)$$
 (1)

$$Score-50 = A - B * log(2 * odds)$$
 (2)

Odds = y/1 - y. (1), (2) Solvable:

Score =
$$50 * log2((1-pro_y)/pro_y) + 450$$
 (3)

Therefore, the probabilistic default probability of the sample prediction is brought into (3) to obtain credit-card user ratings.

The relationship between default probability and sample score is in **Figure 8**, for the 6000 test samples cited above. The abscissa is the user credit score, and the ordinate is the default ratio. The trend clearly shows that high-level contract breach is proportionately lower while low-level breach is proportionately higher. When over 760 points, we can see that the default rate is 0. Generally, the relationship manifests an obvious trend.

5. Conclusions

The present paper has addressed the five principal factors that impact the probability of credit-card default. This probability has also been converted into a credit-score system that, via certain score mapping, can be understood intuitively. From the above data analysis, it can be concluded that XGBoost modelling is effective in predicting users' willingness to repay credit-card debt. As the hypothesis predicted, the default normal ratio has an expected score and the default normal ratio doubles: the expected score is reduced by 50 points, which provides a new paradigm for the study of credit-card default in the future.

The user's characteristics, or "qualifications", and his/her credit rating can be directly expressed via the credit score. The higher the score, the better the credit rating, and the lower the score, the lower the credit rating. More importantly, the credit card companies can arrange relevant credit-management measures

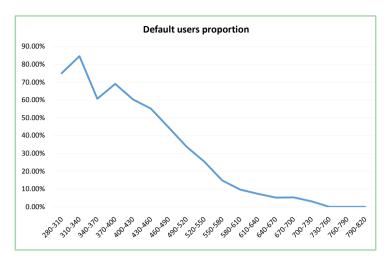


Figure 8. Relationship of default probability and credit score.

according to the user credit score, including strategies such as reminding and collecting. Clients with higher scores will require relatively little attention, while those with lower scores can receive tailored management proportionate to their higher risk. Companies may need to take additional measures for those clients with extremely low scores. Through this credit score system, the credit-management efficiency of banks and other credit institutions can be greatly improved.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] US Census Bureau (2010) Table 1188. Usage of General Purpose Credit Cards by Families: 1995 to 2007. https://www2.census.gov/library/publications/2010/compendia/statab/130ed/tables/11s1188.pdf
- [2] Wallet, N. (2019) American Household Credit Card Debt Study. https://www.nerdwallet.com/blog/credit-card-data/
- [3] Stiglitz, W. (2001) The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2001: Information for the Public. The Royal Swedish Academy of Sciences, Nobel Foundation. nobelprize.org.
- [4] Merikoski, M., Viitala, A. and Shafik, N. (2018) Predicting and Preventing Credit Card Default.
- [5] Yeh, I.-C. and Lien, C.-H. (2009) The Comparisons of Data Mining Techniques for the Predictive Accuracy of Probability of Default of Credit Card Clients. *Expert Systems with Applications*, 36, 2473-2480. https://doi.org/10.1016/j.eswa.2007.12.020
- [6] Venkattesh, A. and Jacob, G. (2016) Prediction of Credit-Card Defaulters: A Comparative Study on Performance of Classifiers. *International Journal of Computer Applications*, 145, 36-41. https://doi.org/10.5120/ijca2016910702

- [7] Ogundimu, E.O. (2019) Prediction of Default Probability by Using Statistical Models for Rare Events. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **182**, 1143-1162. https://doi.org/10.1111/rssa.12467
- [8] Sun, T. and Vasarhelyi, M. (2019) Predicting Credit Card Delinquency: An Application of the Decision Tree Technique. In: Dai, J., Vasarhelyi, M. and Medinets, A., Eds., Rutgers Studies in Accounting Analytics: Audit Analytics in the Financial Industry (Rutgers Studies in Accounting Analytics), Emerald Publishing Limited, 71-83. https://doi.org/10.1108/978-1-78743-085-320191006
- [9] https://www.philadelphiafed.org/consumer-finance-institute/statistics
- [10] Dataset from University of California, Irvine. http://archive.ics.uci.edu/ml/index.html
- [11] Chen, T. and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, 13-17 August 2016, 785-794. https://doi.org/10.1145/2939672.2939785
- [12] Leow, M. and Crook, J. (2016) A New Mixture Model for the Estimation of Credit Card Exposure at Default. *European Journal of Operational Research*, 249, 487-497. https://doi.org/10.1016/j.ejor.2015.10.001
- [13] Agarwal, S., Chomsisengphet, S., Liu, C., Song, C. and Souleles, N.S. (2018) Benefits of Relationship Banking: Evidence from Consumer Credit Markets. *Journal of Monetary Economics*, **96**, 16-32. https://doi.org/10.1016/j.jmoneco.2018.02.005