Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: 10.61784/tsshr3005

AN ANALYSIS OF EARLY WARNING FOR CREDIT CARD CUSTOMER CHURN

ChangFu Yang
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: 625087024@qq.com

Abstract: This paper aims to explore the development of credit cards in the Chinese market and the challenges posed by the rise of internet finance, while also analyzing customer churn issues and their management strategies. Since the introduction of credit cards to China in 1986, despite the initial lack of supporting facilities such as POS machines, credit cards have served as a monetary credit voucher, facilitating the small loan business of commercial banks. Over time, the credit card market has experienced rapid growth, becoming a vital channel for personal consumer loans. However, the emergence of internet finance has had a profound impact on the traditional banking business model, with customer churn becoming an increasingly prominent issue.
Keywords: Credit cards; Internet finance; Customer churn; Data mining; Risk management

1 INTRODUCTION

Since the introduction of credit cards in China in 1986, they have not only transformed traditional payment habits but also provided financial support for the small credit services of commercial banks. The widespread use of credit cards has enabled consumers to enjoy the convenience of "spend now, pay later," which has in turn spurred the rapid development of the market. However, with the rise of internet finance, traditional banking services are facing unprecedented challenges. Internet finance companies, taking advantage of the internet's convenience, have quickly attracted a large number of customers, breaking the closed loop of banking services, especially in areas that banks find difficult to reach, such as personal loans for college students.

The development of internet finance is not simply a matter of internet plus finance,but rather it is a new form of finance that relies on new technologies such as cloud computing and mobile networks, and combines with new types of services like online payments and social media. This transformation has not only reduced the cost of financial services but also improved efficiency. However, it has also led to an increase in the rate of customer attrition from traditional banks. According to data from the central bank, the growth rate of credit card issuance has been declining year by year since 2017, and the issue of customer attrition is becoming increasingly serious.

The application of machine learning to predict credit card customer churn has seen significant advancements. Studies like those by Xu et al.[1] and Nie et al. [2] have shown that hybrid models and logistic regression can achieve high accuracy in predicting customer churn. Lin et al.[3] and Caigny et al.[4] have further enriched the field with their innovative approaches combining different techniques. Ensemble methods, as explored by de Bock and van den Poel[5], have also proven effective. Collectively, these studies highlight the growing sophistication of predictive analytics in customer retention strategies.

This paper utilizes a bank customer dataset obtained from the Kaggle website, employing text mining techniques to extract the main characteristics of churned customers and conducting Exploratory Data Analysis (EDA) to gain a deeper understanding of the data. Subsequently, a Random Forest model is applied for feature extraction to identify key characteristics of churned customers, and effective customer churn control and risk prevention recommendations are proposed.

2 MODELING ANALYSIS

2.1 Model Introduction

Random Forest (RF) is an ensemble learning method with decision trees as its core building block. The algorithm improves the accuracy and robustness of classification by constructing multiple decision trees and integrating their prediction results. In a random forest, each decision tree classifies the input samples independently.

For the Random Forest algorithm, it is an ensemble classifier composed of K decision trees $h(X,\theta_k),k=1,2,\ldots,K$ as the basic classifiers. When a sample to be classified is input, the classification result output by the Random Forest is determined by the voting of the classification results of each decision tree. The sequence of random variables $\theta_k,k=1,2,\ldots,K$ is determined by the two main randomization ideas of the Random Forest: Bagging and feature subspace. Let N denote the number of training samples, and M represent the number of features. The construction algorithm is as follows:

1) The number of input features m is used to determine the decision outcome at a node in the decision tree (m < M).
2) With replacement sampling is performed N times from the N training samples to form a training set

(Bootstrap sampling).

3) For each node, m features are randomly selected, and the best way to classify based on these m features is calculated.
4) The individual trees are combined to form the Random Forest.

The process of training each decision tree is to train the entire random forest, and because each decision tree is independent of each other, its training can be carried out together, which will greatly improve the efficiency of the model. Each decision tree is trained in the same way and then combined to obtain K decision trees, and the required random forest model is formed. The samples to be predicted are obtained by ranking the weights of the problem solving, and the average of the output results of each decision tree is taken as the result of random forest prediction.

Random Forest Regression (RFR) is formed by the growth of decision trees related to a random vector θ, where the dependent variable is a continuous variable, and it is assumed that the training set is independently drawn from the distribution of a random variable. Let $h_i(x)$ be the regression model result of a single decision tree, then the predicted value of the Random Forest Regression is obtained by averaging the regression results of k decision trees $\{h(X,\theta_i), i=1,2,\ldots,k\}$, that is:

$$H(x)=\frac{1}{k}\sum_{i=1}^{k}h_i(x)$$

Where $H(x)$ represents the result of the combined regression model.

The Random Forest method uses the bootstrap resampling method to obtain different sample sets, and constructs decision tree regression models using these sample sets, thereby increasing the differences between models and enhancing the ability of extrapolation prediction.

In summary, the steps of the Random Forest algorithm are as follows:

Step 1: A set of samples $\{T_k, k=1,2,\ldots,K\}$ is calculated by the weights selected by a decision tree for the problem.

Step 2: Classification results:

1) Individual decision trees are generated by randomly drawing K samples with replacement from the training set;
2) During the generation of the decision tree, a subset of attributes is randomly selected with equal probability;
3) The above two steps are repeated to generate K decision trees;
4) Finally, the optimal result of the final vote is output.

2.2 Indicator Selection

In order to identify the key factors influencing users' credit card satisfaction and the indicators that can reflect the trend of users' credit card use intention, this study used text mining technology to screen out 18 key indicators from the hot words related to credit card churn. These metrics include: user age, gender, household size, education level, marital status, income level, credit card tier, credit card registration time (months), number of products held, number of inactive months in the last 12 months, number of contacts in the last 12 months, total revolving credit card balance, average open purchase credit limit, change in transaction value between Q4 and Q1, total transactions in the last 12 months, number of transactions, and average frequency of credit card usage.

2.3 Model Solving

The random forest model is a powerful tool that is able to assess the impact of different factors on the churn rate of credit card users. In this study, we quantified the effect of each independent variable on the dependent variable by adjusting the model parameters and training a random forest model by adjusting the model parameters and using the churn rate of bank credit card users as the dependent variable. Once the model was trained, we got an importance score for each variable, which reflects the relative impact of each factor on the churn rate.

The relative error of the model is 0.0425, indicating that the model has high prediction accuracy. The results of the analysis showed that the total number of transactions and the number of transactions were the two most important factors influencing user churn, and their importance scores were significantly higher than those of other variables. This is followed by changes in the total amount frozen and the number of transactions, which constitute the second most important group of factors influencing user churn. The importance score of change in transaction amount and average frequency of use is low, but still higher than other factors, making up the third echelon. The importance score for the number of products held is in the fourth tier. The mean values for age and open purchase credit are classified as the fifth band while the importance scores for credit card time on book, months of inactivity, and number of contacts are in the sixth bracket. Factors such as educational attainment, number of family members, gender, marital status, income rating, and other factors have relatively low importance scores, while credit card ratings have the lowest degree of influence among all factors. Through the analysis of the random forest model, we can clearly identify the main factors affecting

the churn of bank credit card users, and provide targeted strategy suggestions for banks. Random forest model impact factor importance ranking can be seen in Figure 1.

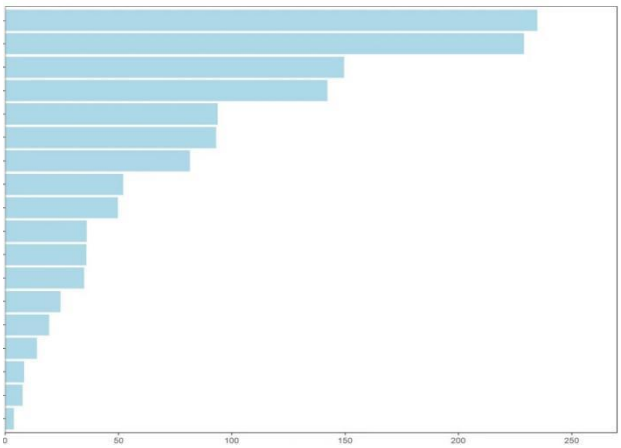

Figure 1 Random Forest Model Impact Factor Importance Ranking

According to the analysis results of the random forest model, we find that the main driving factors for the churn of bank credit card users include key economic indicators such as total transaction value and number of transactions. These findings point to the possibility that the credit card limits offered by banks for different levels of users may not be sufficient to meet their needs. Specifically, the growth of total transaction value is negatively correlated with the decrease in churn, suggesting that banks may be effective in reducing churn if they can provide credit lines that match users' spending power.In addition, an increase in the number of transactions is likewise associated with a lower churn rate. This means that frequent transaction activity may enhance users' dependence on and satisfaction with banking services. Therefore, banks can appropriately increase the frequency of credit card use on the premise of ensuring that users have the ability to repay, so as to stabilize and increase the user base.

Based on the importance of the influencing factors shown by the random forest model, banks can formulate targeted policies, such as adjusting credit limits and optimizing transaction experience, to reduce the churn rate of credit card users. In addition, by building a churn early warning model, banks can more effectively manage customers, identify high-risk customer groups, and take customer care measures to maximize the retention of these customers, thereby enhancing the ability of enterprises to resist customer churn risks. This data-driven strategy not only helps banks optimize customer relationship management, but also improves the personalization and competitiveness of banking services, ultimately leading to increased customer satisfaction and loyalty.

3 CONCLUSIONS AND RECOMMENDATIONS

The purpose of this paper is to use data mining technology to construct an early warning model for credit card customer churn for a bank. By quantitatively analyzing the churn problem of credit card customers and applying modern data mining methods such as random forest, an early warning model was successfully established, which can effectively monitor the risk of customer churn. This study combines statistical testing and data mining techniques to realize the integration of statistics and business practice, as well as the unification of qualitative and quantitative analysis methods. Through the test of the validation dataset, the model confirms its effectiveness in the early warning of credit card user churn, which provides strong support for solving the problem of user churn. The main results of this paper are summarized below:

1) Personalized marketing strategy: With the intensification of business competition, Internet financial enterprises need to implement personalized marketing for different credit card users. This study demonstrates the role of random forest model in quantifying user behavior characteristics and assisting marketing decision-making, and emphasizes the importance of data mining technology in banking business improvement and customer relationship management.
2) Feasibility of model establishment: Based on expert survey and statistical analysis, this study discusses the feasibility of establishing an early warning model for credit card customer churn and identifies the key factors affecting customer churn.
3) Empirical research: Through the case study of text data mining, a random forest prediction model is established, which provides an effective churn analysis framework for enterprises.

Credit card churn early warning analysis is an emerging field in credit card customer relationship management. The variability of customer needs and the expansion of choice increase the risk of customer churn. Therefore, understanding and maintaining valuable customers, improving the bank's competitiveness and reducing operational risks have become the key to the bank's business.

The user portrait system developed by the AI Lab provides the foundation for intelligent marketing in banks. Credit card churn warning models not only help target potential churned customers and improve customer retention, but can also be applied to all stages of the customer lifecycle, such as prospecting, nurturing high-value customers, increasing customer loyalty, and extending customer lifecycles. It is expected that more intelligent marketing model technologies and applications will be developed and applied in the future.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Xu, Y., Rao, C., Xiao, X., Hu, F. Novel Early-Warning Model for Customer Churn of Credit Card Based on GSAIBAS-CatBoost. CMES-Computer Modeling in Engineering & Sciences, 2023, 137(3).

[2] Nie, G., Rowe, W.G., Zhang, L., Tian, Y., Shi, Y. Credit card churn forecasting by logistic regression and decision tree. Expert Syst. Appl., 2011, 38: 15273-15285.

[3] Lin, C. S., Tzeng, G. H., Chin, Y. C. Combined rough set theory and flow network graph to predict customer churn in credit card accounts. Expert Systems with Applications, 2011, 38(1): 8-15.

[4] De Caigny, A., Coussement, K., De Bock, K. W. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. European Journal of Operational Research, 2018, 269(2): 760-772.

[5] De Bock, K. W., & Van den Poel, D. An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. Expert Systems with Applications, 2011, 38(10): 12293-12301.