ISSN 2616-5775 Vol. 5, Issue 10: 54-59, DOI: 10.25236/AJCIS.2022.051009

Research on Personal Credit Risk Assessment Based on Combination Weight and Shap Interpretable Machine Learning

Xuehan Peng*

School of Statistics, Capital University of Economics and Business, Beijing, 100000, China

Abstract: *With the continuous promotion of e-commerce platform installment payment and P2P credit platform, personal credit risk assessment is becoming more and more important for e-commerce and users Taking the data of Tianchi competition platform as a sample, this paper constructs an index system for personal credit risk assessment, and calculates the combined weight of each index through the critical method and entropy weight method, and weights it to obtain three factors: basic information, credit information and lending behavior information. Based on the quantified three indicators and the unbalanced sample algorithm, XGBoost and lightGBM are used to predict credit risk, it is found that the performance of these two methods is basically the same. This paper uses the SHAP-values interpretable machine learning method to explain the importance of these three factors. The empirical results show that the accuracy of XGBoost and lightGBM is higher than 80%, and the order of the importance of the three factors is: "basic information" is higher than "lending behavior information" and higher than "credit information". Finally, this paper puts forward relevant suggestions for the stable development of enterprise risk control and credit industry.*

Keywords: *credit risk assessment, portfolio weight, SHAP-values, machine learning*

1. Introduction

With the continuous development of e-commerce, personal credit business based on Internet platform has also risen rapidly, and it has become a necessary means to attract young people to consume. By developing personal credit business, users can cultivate the habit of borrowing and consumption, and bring greater consumption transaction volume and working capital to enterprises. However, high returns often coexist with risks, and e-commerce platforms must operate cautiously if they want to survive forever. On the one hand, due to the incomplete credit reporting system in China, the historical data accumulated by the Internet financial platform is short and the coverage is not wide enough; On the other hand, the Internet financial platform encourages ahead of time consumption, making some users fail to take into account the long-term accumulated amount of their arrears when paying instalments, which is a disaster for users and e-commerce platforms.

Accurate assessment of credit risk of credit applicants is very important for commercial banks, and scholars at home and abroad have been committed to this research. The consumer credit scoring model proposed by David Durand (1941) in risk element in consumer. lnstrument lending was the first to quantify the borrower's personal data such as income and deposit, so as to rate its credit, which is still of reference significance to the current personal credit rating research in China Li Shuguang (2003) divides personal credit scoring into three categories: personal behavior scoring, profit scoring and scoring considering economic environmental factors in personal credit evaluation research [1], summarizes the problems that need to be solved at home and abroad, and believes that the personal credit scoring model should include the consideration of macroeconomic factors Lu Wei (2004) analyzed the time span distribution based on data in the "comparative study of credit scoring methods based on China's real personal credit data" [2], and found that the life cycle of the model of personal credit scoring is relatively short in China. He believes that the practical model developed based on China's data should be updated at least once a year. This puts forward the requirement of strong updatability for the research of personal credit scoring model. On the whole, there has been a certain degree of research on personal credit risk assessment at home and abroad, and how to determine the index system for assessing personal credit risk, and then build a personal credit risk assessment model to provide an assessment direction for the credit risk of commercial banks, there is still a lot of content to be further studied. Abellan[3] applied different basic learners to credit scoring with bagging scheme

ISSN 2616-5775 Vol. 5, Issue 10: 54-59, DOI: 10.25236/AJCIS.2022.051009

in 2017. The results showed that the integrated model with decision tree as basic learner achieved the best credit scoring effect. Liaowen xiong[4] and others used a series of data preprocessing methods and feature selection method xgbfs based on embedded ideas to reduce the user credit data dimension and train the xgboost evaluation model, and finally realize the user credit risk evaluation. Although many researchers have explored e-commerce risk control from different perspectives, there are still several problems: First, the interpretability results of many methods are not unique in theory, for example, weight, gain and cover weights in XGBoost, which may explain different and contradictory results. The third point is about the quantification of risk control factors. At present, most methods do not combine the information content, volatility and correlation of indicators to effectively construct appropriate risk control factors.

This paper proposes a personal credit risk assessment model based on combination weight and shap interpretable machine learning. The innovation of this model lies in: firstly, the combined weight based on the combination of critical method and entropy method not only considers the amount of indicator information, but also measures the volatility and correlation of indicators. Secondly, shap interpretable machine learning method can explain the importance of indicators from the macro and micro levels. Taking the data of Tianchi competition platform as a sample, this paper constructs a personal credit risk assessment index system including primary indicators such as basic information, credit information and lending behavior information, and calculates the combined weight of each indicator through critical method and entropy weight method, and weights the indicators, so as to quantify the basic information, credit information and lending behavior information. Based on the smote unbalanced sample algorithm, this paper uses two machine learning methods, xgboost and lightgbm, to predict the default of samples. Based on the interpretable machine learning method of shap, the customer behavior and information are analyzed from the macro and micro levels respectively.

2. Theory and method

2.1 Construction of index system

Table 1: personal credit risk assessment index system

Primary index	Secondary indicators	Explanation of the meaning of secondary indicators	Positivity and Negativity
Essential information	X1	Years of employment (years)	Positivity
	X2	Housing ownership status of the borrower	Positivity
	X3	annual income	Positivity
Credit information	X4	Loan amount	Negativity
	X5	lending rate	Negativity
	X6	Installment amount	Positivity
	X7	Level of loan grade	Positivity
	X8	Debt to income ratio	Negativity
	X9	The number of default events in the borrower's credit file that are more than 30 days overdue in the past two years	Negativity
	X10	Number of open credit lines in the borrower's credit file	Negativity
	X11	Total credit turnover balance	Positivity
	X12	The amount of credit used by the borrower relative to all available revolving credits	Positivity
	X13	Total current credit limit in the borrower's credit file	Negativity
	X14	Indicate whether the loan is an individual application or a joint application with two co borrowers	Positivity
Lending behavior information	X15	Verification status	Positivity
	X16	The lower limit of the borrower's FICO at the time of loan issuance	Negativity
	X17	The upper limit range of the borrower's FICO at the time of loan issuance	Negativity
	X18	Number of derogatory public records	Negativity
	X19	Number of public records cleared	Negativity
	X20	Initial list status of loan	Negativity

The personal credit risk assessment index system established in this paper includes three first-class indicators, namely "basic information", "loan information" and "loan behavior information". Basic information refers to the annual income, employment years, housing ownership and other information

ISSN 2616-5775 Vol. 5, Issue 10: 54-59, DOI: 10.25236/AJCIS.2022.051009

of the credit applicant, which represents the repayment ability of the credit applicant. Credit information mainly includes the credit and liabilities of loans and credit cards, repayment details, asset disposal information, and guarantees for others. Lending behavior information refers to a series of behavioral characteristics of borrowers that may affect credit over a period of time. These three primary indicators contain several secondary indicators. These secondary indicators, their meanings and indicators are shown in the following table 1.

2.2 Personal credit risk assessment model based on combination weight and shap interpretable machine learning

2.2.1 Combination weight

This paper uses the combined weight based on entropy weight method and critical method. Both entropy method and critical method can scientifically and objectively calculate the weight of evaluation indicators, eliminate the indicators that make little contribution to the index system, and eliminate the subjective judgment bias However, the entropy method relies on the discreteness of the data itself, ignoring the possible correlation between the indicators. The change of indicators with small weight calculated by the entropy method may affect other indicators with large weight. The premise of the establishment of critical method is that the current evaluation indicators have been determined and improved, ignoring the possibility of the development and change of each evaluation indicator. To sum up, the combination of entropy method and critical method has perfect complementarity. By combining the two, we can not only fully consider the variability of each index data, but also take into account the correlation between data in the process of objective weighting.

2.2.2 SMOTE

Smote algorithm is an oversampling technology that deals with data imbalance and synthesizes minority classes. Its basic idea is to analyze minority samples and synthesize new samples according to minority samples and add them to the data set. To a certain extent, it can make the model avoid the problem of over fitting.

2.2.3 XGBoost and lightGBM

In this paper, the integrated models xgboost and lightgbm are used as prediction models. Compared with a single model, the integrated model has the advantages of enhancing the expression ability and reducing the error probability. Therefore, this paper chooses to use the integrated model.

2.2.4 Shap interpretable machine learning method

Shap values is an additive interpretation model built by lundberg[3] inspired by cooperative game theory in 2017. Its core is to calculate the shap values of each feature, so as to reflect the contribution of the feature to the prediction ability of the whole model. Compared with various feature importance measures based on the characteristics of the model itself, shap values have the additive consistency of the output results, which is consistent with the regression in the general sense: for each prediction sample, the model generates a prediction value, and the shap value is the value assigned to each feature in the sample[5,6,7].

In this paper, through the combination of entropy method and critical method, the combined weight of each index is calculated and weighted, and three first-class indexes are obtained. Based on the quantified three indicators and the processing of unbalanced sample algorithm, xgboost and lightgbm are used to predict default. For these two methods, the importance of these three indicators is explained by using the shap values interpretable machine learning method. Accordingly, this paper obtains the workflow diagram of the personal credit risk assessment system, as shown in Figure 1.

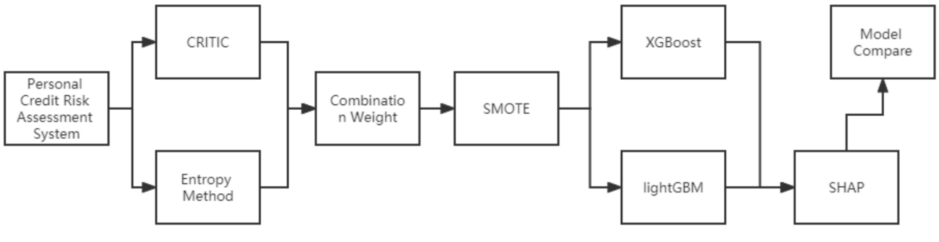

Figure 1: Workflow

ISSN 2616-5775 Vol. 5, Issue 10: 54-59, DOI: 10.25236/AJCIS.2022.051009

3. Empirical results

3.1 Data source and preprocessing

The data used in this paper comes from the competition data of Alibaba cloud Tianchi competition platform. The original sample is 800000 in total. After excluding the samples containing missing values, this paper randomly selects 2200 samples, including 200 default samples and 2000 non default samples. Normalize the extracted samples. Because the data extracted in this paper is unbalanced sample data, it needs to be processed by smote algorithm, that is, analyze and simulate a few categories of samples (default samples), and add manually simulated new samples to the data set, so that the categories in the original data are no longer seriously unbalanced, so as to make the subsequent analysis more accurate.

3.2 Comparison of prediction accuracy results of the model

This paper mainly uses five common classification performance indicators to evaluate the advantages and disadvantages of the model, namely accuracy, precision, recall, F1 value and AUC value. Accuracy, as the most basic evaluation index, is the percentage of all samples that are correctly classified. Precision is an indicator only for prediction results, and the percentage of samples classified as positive that have correct prediction classification. Recall is the percentage of true positive samples with correct prediction classification, also known as recall. F1 value is an index that integrates the output results of precision and recall, and the value range is from 0 to 1, where 1 represents the best output of the model, on the contrary, 0 represents the worst output of the model. The AUC value (area under curve) is the sum of the areas under the ROC curve. The larger the AUC value, the higher the accuracy of the model.

Based on the above five indicators, this paper compares the model accuracy of xgboost and lightgbm. The comparison results are shown in Table 2.

Table 2: Comparison of model accuracy

Model name	XGBoost	lightGBM
AUC	0.928(0.053)	0.937(0.046)
Accuracy	0.855(0.058)	0.868(0.054)
Precision	0.856(0.021)	0.87(0.014)
Recall	0.854(0.132)	0.866(0.128)
F1 Score	0.849(0.086)	0.862(0.079)

According to the analysis, the values of the evaluation indicators of xgboost and lightgbm models are very close, with an accuracy of about 86.1%, an accuracy of about 86.2%, a recall rate of about 86%, an F1 value of about 0.855, and an AUC value of about 0.932. The accuracy, precision, recall, F1 value and AUC value of the two models are very high, both above 85%. Therefore, these two machine learning methods can well evaluate and predict personal credit risk, which reflects the superiority of the integrated algorithm.

3.3 Explanatory results of predictive variables based on Shep values

At present, an important problem in the application of machine learning in practical business areas is that machine learning is difficult to make operators understand which indicators play a key role like linear regression. In other words, although the result can be trusted, the process may not be trusted. Aiming at this problem, this paper introduces the shap interpretation method, and strives to solve this problem in model interpretation.

According to the shake method, based on the machine learning model that has been trained, we calculate the contribution of the index shake value for each sample point once, so that the sum of all the index shake values of all the sample points is equal to the output of the sample on the model, that is, the probability of default. Taking the 40th and 45th loan applicants in the training set as examples, we use xgboost model and lightgbm model respectively to visualize the local interpretation of their contributions to the sample points under the shake method:

Arrows in different directions and lengths are used in the schematic diagram to indicate the direction and size of the effect on the probability of default. The arrow pointing to the right (red) indicates that the corresponding indicator has an increasing effect on the probability of default when taking the sample value, and the arrow pointing to the left (blue) indicates that the indicator has a

ISSN 2616-5775 Vol. 5, Issue 10: 54-59, DOI: 10.25236/AJCIS.2022.051009

decreasing effect on the probability of default when taking the sample value. The absolute value of the contribution of each part of the length corresponding to the indicator. Based on the prediction of the 40th loan applicant, for xgboost and lightgbm models, F1 and F2 indicators, i.e. basic information and credit information, have increased the default probability of output, while F3 indicator lending behavior information has reduced the default probability of output of the model. Finally, the default probability of output is 81% and 67% respectively. Based on the prediction of the 45th loan applicant, for the xgboost and lightgbm models, F1, F2 and F3 indicators, namely basic information, credit information and lending behavior information, have reduced the default probability of the model output, and finally we can get that the default probability of their output is 5% and 4% respectively. As shown in figure 2 and figure 3.

Figure 2: partial explanation of the shadow method for xgboost

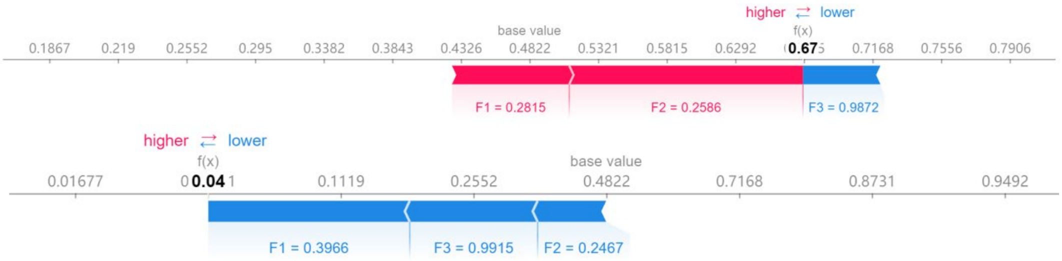

Figure 3: partial explanation of the shadow method for lightgbm

After studying the local interpretation (micro results) of the machine learning model by the shap method, this paper continues to study its macro results. Therefore, based on the interpretable machine learning method of shap values, this paper makes a global interpretation of xgboost and lightgbm models respectively, and the macro results are shown in Figure 4 and figure 5.

It can be seen from the analysis that the macro results of the shap method for these two models are very similar, F1 (basic information) and F3 (lending behavior information) are of high importance, and F2 (credit information) is of low importance. For the lightgbm model, for the two categories of default and non default, the importance of the three indicators is basically the same, that is, F1 and F3 are both more important, and F2 is less important.

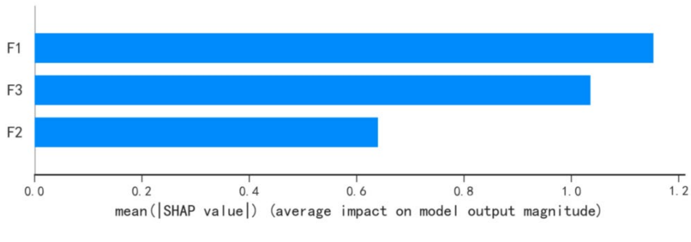

Figure 4: The macro explanation of SHAP for xgboost

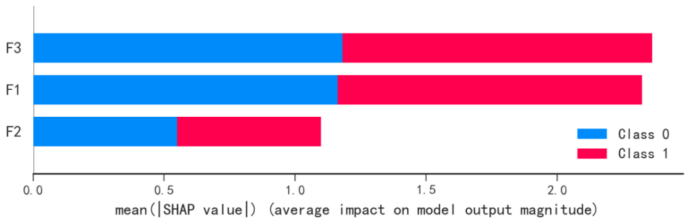

Figure 5: The macro explanation of Shep for lightgbm

ISSN 2616-5775 Vol. 5, Issue 10: 54-59, DOI: 10.25236/AJCIS.2022.051009

4. Conclusion and Prospect

This paper proposes a personal credit risk assessment model based on combination weight and shap interpretable machine learning. Experiments show that the model has high prediction accuracy and can output reasonable macro and micro results. Based on the results of this paper, we can put forward targeted suggestions for personal credit risk assessment. According to the macro results of the shap model, basic information and lending behavior information play a greater role in whether the credit applicant defaults. Therefore, e-commerce platforms need to pay attention to investigating the basic information of the credit applicant, such as annual income, employment years, real estate, and a series of behavioral characteristics of other borrowers that may affect credit[8-9].

References

[1] Shuguang Li. Research on personal credit evaluation. Beijing. Beijing Jiaotong University. June 2003.13-30

[2] Wei Lu. A comparative study of credit scoring methods based on real personal credit data in China [Dissertation]. Hefei. University of science and technology of China. 2004.45-48

[3] Abellán JMantas C J. Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring [J]. Expert Systems with Applications, 2014, 41(8): 3825-3830.

[4] Lundberg S M, Lee S I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems, ACM, 2017, 4765-4774.

[5] Wenxiong Liao, Bi Zeng, Tiankai Liang, Yayun Xu, Junfeng Zhao. Personal credit risk assessment method for high dimensional data[J]. Computer engineering and Application, 2020, 56(04):219-224.

[6] Arora N, Kaur P D. A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment [J]. Applied Soft Computing, 2020, 86: 105936.

[7] Marcílio W E, Eler D M. From explanations to feature selection: assessing shap values as feature selection mechanism. 2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI). Ieee, 2020: 340-347.

[8] Gramegna A, Giudici P. SHAP and LIME: an evaluation of discriminative power in credit risk [J]. Frontiers in Artificial Intelligence, 2021: 140.

[9] Zhang X, Yu L, Yin H, et al. Integrating data augmentation and hybrid feature selection for small sample credit risk assessment with high dimensionality [J]. Computers & Operations Research, 2022: 105937.