Artificial Intelligence and Human-Computer Interaction
Y. Ye and P. Siarry (Eds.)

doi:10.3233/FAIA240149

Explainability of Machine Learning in Credit Risk Assessment of SMEs

Boya HUANG [a], Feiyang ZHAO [a], Maohong TIAN [a,b], Dequan ZHANG [a,b]
Xintong ZHANG [b], Zuo WANG [a,b], Hualin LI [a,b,1] and Bingling CHEN [c]
[a] *Chongqing Institute of Engineering, Chongqing, China*
[b] *Shu Yi Xin Credit Management Co., LTD, Chongqing, China*
[c] *Department of Big Data Artificial Intelligence, Guangzhou Nanfang College, Gangzhou, China*

Abstract. Small and medium-sized enterprises (SMEs) have developed rapidly in China, bringing enormous opportunities and challenges. In this study, we aim to investigate methods that can accurately assess credit risks of SMEs, using machine learning algorithms, focus on explainability, customer default forecasting, and delinquency. This study focused on the enterprises' performance data and used the authorized invoice data of 425 SMEs in Chongqing. Machine learning algorithms, such as logistic regression, random forest, support vector machine, and soft voting ensemble learning methods, were used to establish a prediction classifier that was combined with the SHAP value to explain the feature contribution of a specific output. Therefore, Our study presented a strong correlation between the derived features and future delinquencies, which will enable in forecasting enterprises' business performance.

Keywords. Credit risk model, SMEs lending, explainability, data mining, machine learning.

1. Introduction

Credit is at the heart of not only banking but also the business as a whole. The history of several financial crises has made people realize that systematic risk is not independent; it can be transferred from other individual risks, such as credit, market, and liquidity risks. Thus, credit risk assessment is a critical issue in financial risk management.

Small and medium-sized enterprises (SMEs) have developed rapidly in the People's Republic of China (PRC) in recent years, accounting for over half of the gross domestic product (GDP) since 2019. However, it is almost impossible for most SMEs to apply for operational loans without sufficient assets, mortgages, and related guarantees. Credit loans have gradually become a vital channel for meeting an SMEs' financial demands. In addition, the available online information is ineffective in evaluating credit performance because most SMEs have insufficient historical records, and consequently posing difficulties in loan approval. Hence, the majority of financial institutions conduct due diligence for loan approving, resulting in human and time costs. In the traditional method, the final credit line is based on the risk appetite and historical experience of the

[1] Corresponding Author: Hualin Li, Chongqing Institute of Engineering, Shu Yi Xin Credit Management Co., LTD, China. E-mail: hualinli@hotmail.com.

investigator. Although the artificial method considerably avoids bad debt risks, financial institutions are limited in time and space. Furthermore, it affects high-quality SME development.

To solve the problem of information asymmetry, Yin et al. combined financial and judicial information to evaluate the potential credit risk [1]. With the rapid development and improvement of computing power and artificial intelligence technology, machine learning (ML) algorithms, such as support vector machine (SVM) [2], decision trees [3], and XGBoost [4], have become popular in modeling SMEs' credit risk, especially in detecting default probability [5]. Liu et al. applied KNN and SVM models to predict the default probability of online loan borrowers [6]. Chen et al. focused on resampling and cost-sensitive mechanisms to process imbalanced datasets by avoiding information asymmetry in the predicted model, and improving model performance [7]. Individual and ensemble ML methods are used in predict SME credit risk in supply chain finance [8]. However, owing to lack of transparency in traditional ML algorithms, measuring credit risk in an interpretable manner has become a challenge. Previous studies applied ML techniques and rule-based methods for better explanation of model results [9]. Besides, researches focused on explainable ML methods developed in the last three years to help stakeholders comprehend the main drivers of model-driven decisions [10-11]. The Local Interpretable Model-agnostic Explanation (LIME) identifies sparse explanations and fits a simple interpretable surrogate model that is locally consistent with the black box model [12-13]. Model-independent methods, such as Shapley's Additive Explanations (SHAP value) from cooperative game theory, attribute the marginal effects of individual variables to model predictions and reveal the dispersion, nonlinearity, and structural breaks between each feature and the target variable [14-15].

However, research on SMEs' credit risk is still under development. Owing to privacy considerations, existing literature primarily focused on conventional data to train models, and with limited explanation of the model results. In this work, we focus on enterprise performance data and use the authorized invoice data of 425 SMEs in Chongqing as the data source, and form the final invoice data set after a series of operations such as data masking, label selection and data cleaning. For a better feature performance, Decision Tree (DT) is used for segmentation, and the predictive ability of each feature on the target variable is measured by Weight of Evidence (WOE) and Information value (IV). Finally, the features are classified by machine learning algorithms such as logistic regression, random forest, support vector machine, and soft-voting integrated learning methods to achieve corporate credit risk prediction. For the performance comparison, we used a confusion matrix, KS value, and AUC value to choose the model with higher performance. For interpretability comparison, we applied the explainable ML method SHAP value to estimate the marginal contribution of each feature, either positively or negatively, to the target variable. Furthermore, to improve the transparency, auditability, and explainability of the forecast results, we determined the contribution of important features. Figure 1 illustrates the entire operational procedure of this research, from data preprocessing and feature selection to model training, ensembling, evaluation, and explanation.

2. Materials and methods

2.1. Date source

We randomly sampled and collected data from 425 enterprises in Chongqing, China. The authorized private invoice data contains 728,822 annually input valued added tax (VAT) data and 561,428 output VAT data from 2018 to 2020. Credit performance is defined as the label data that distinguishes an enterprise's creditworthiness. To protect the security of these SMEs' private information, all data are processed using Data Masking technology.

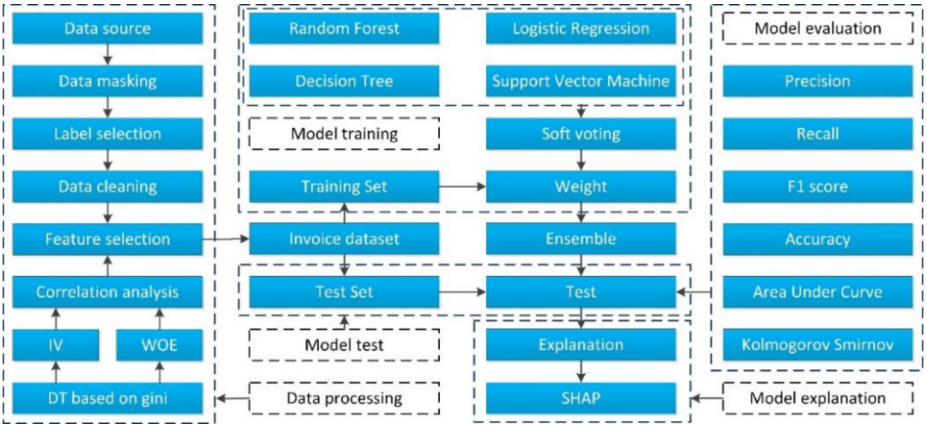

Figure 1. Operational framework.

Table 1 intuitively explains the business performance of SMEs based on annual sales and procurement data from 2016 to 2020. The performance data for 2020 are deficient due to the difficulties in data collection.

Table 1. Distribution of purchasing and selling for 425 companies from 2016-2020.

Data Source	Observations				
	2016	2017	2018	2019	2020
Sales Data	106K	3,054K	3,543K	3,438K	219K
Procurement Data	54K	1,995K	4,798K	5,575K	300K

The data were divided into training and test sets in the proportion of 7:3. Thus, invoice data of 327 enterprises were used for model training, while the remaining 98 enterprises were used for model evaluation.

2.2. Data preprocessing

2.2.1. Data cleansing

Several steps were taken to ensure data accuracy and integrity. Irrelevant content and superfluous information were removed, such as company names and addresses. Furthermore, all missing data were replaced with 0, while rows with more than a 50% default rate were deleted.

2.2.2. Feature engineering

The collected invoice data illustrate enterprises' performance during different periods. This data reflects invoice quantity, amount, and ratio, which helps in determination of reliable statistical and empirical relationship between certain features and the target variable.

From 2018 to 2020, 238 alternative features, including volatility-related features, development rate, concentration rate were acquired through the feature engineering process. Owing to the problem of data deficiency, 18 procurement-related indicators and 17 sales-related indicators from 2019 were obtained through the feature engineering process.

2.2.3. Feature selection

To include more characteristic values of invoice data and compensate for sample deficiency as well as calculation procedure, all discrete and continuous fields after feature engineering are binned through classification and regression tree methods generated by the Gini index.

Subsequent to the binning process, and based on the binning results, the Weight of Evidence (WOE) value and IV (Information Value) are calculated to measure the predictive power of each feature to the target variable [16] for further feature selection [17-18]. The mapping of variables from the original value to binned value can be completed by calculating WOE value of each group after binning process, with the mapping formula shown in Equation 1:

$$WOE_i = \ln\left(\frac{\frac{y_i^1}{y^1}}{\frac{y_i^0}{y^0}}\right) = \ln\left(\frac{y_i^1}{y_i^0}\right) - \ln\left(\frac{y^1}{y^0}\right) \tag{1}$$

where y_i^1 and y_i^0 represent the number of default and nondefault enterprises in the ith bin, while y^1 and y^0 represent the number of default and non-default companies in the total sample, respectively.

Accordingly, the IV value is obtained based on the WOE calculation, which is then used to measure the information value of a certain variable, that is, its forecasting ability. Assuming that the sample set is divided into N groups, the formula for calculating the IV value is given by Equation 2.

$$IV = \sum_{i=1}^{N} WOE_i * \left(\frac{y_i^1}{y^1} - \frac{y_i^0}{y^0}\right) \tag{2}$$

Besides, features with high correlation (Pearson correlation >0.7) and relatively low IV were deleted to avoid overfitting. Subsequent to the feature selection process, 27 variables remained for further modeling. Figure 2 shows the strongly correlated features and Table 2 illustrates theses 27 features.

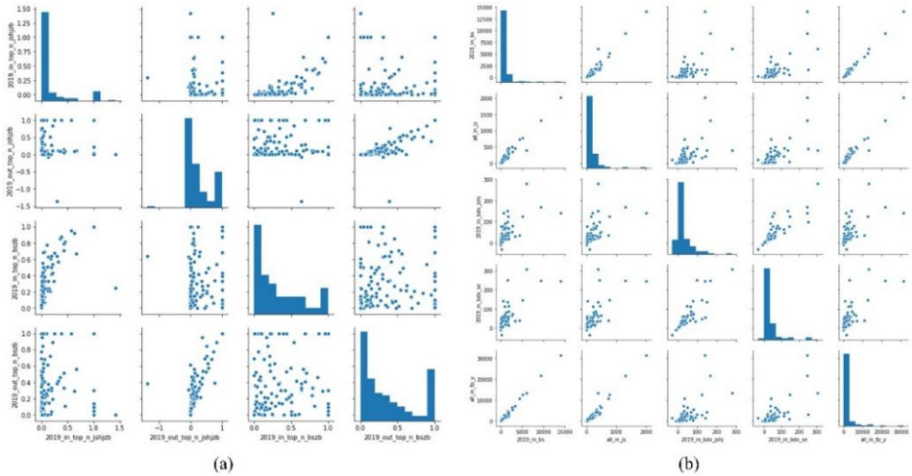

Figure 2. Correlation scatter plot of highly correlated features.

Table 2. Indicator description.

Order	Variable	Indicator
1	all_in_fp_y	Number of valid input invoices.
2	all_out_js	Number of output counter-parties.
3	2019_in_avg_sd	Average input tax rate in 2019.
4	all_out_fp_y	Number of valid output invoices.
5	2019_out_jshj	Amount of output invoices in 2019.
6	2019_in_jshj_zb_all	The amount of input invoices in 2019 compared with total sales amount from 2018 to 2020.
7	2019_out_bdx_se	Volatility of output invoices tax amount in 2019.
8	2019_out_se	Tax amount of output invoices in 2019.
9	all_out_fp_w	Number of invalid output invoices.
10	2019_in_bdx_se	Volatility of input invoices tax amount in 2019.
11	2019_out_czl_bs	Growth rate of output invoices in 2019.
12	2019_out_avg_sd	Average output tax rate in 2019.
13	2019_out_bdx_js	Volatility of counter-party numbers of output data in 2019.
14	2019_out_se_zb_all	The tax amount of sales data in 2019 compared with total input amount from 2018 to 2020.
15	2019_out_bdx_jshj	Volatility of output invoices amount in 2019.
16	2019_in_czl_bs	Growth rate of input invoices in 2019.
17	2019_out_top_n_jshjzb	The output invoices amount of top 5 counter-parties compared with total output amount in 2019.
18	2019_in_jshj	Amount of input invoices in 2019.
19	2019_out_jshj_zb_all	The amount of output invoices in 2019 compared with total sales amount from 2018 to 2020.
20	2019_in_se_zb_all	The tax amount of procurement data in 2019 compared with total input amount from 2018 to 2020.
21	2019_in_top_n_bszb	The input invoices number of top 5 counter-parties compared with total output number in 2019.
22	2019_out_js	Number of output counter-parties in 2019.
23	all_in_fp_w	Number of invalid input invoices.
24	all_in_js	Number of input counter-parties.
25	2019_out_bs	Number of output invoices in 2019.
26	2019_in_js	Number of input invoices in 2019.
27	2019_in_bdx_js	Volatility of counter-party numbers of input data in 2019.

2.3. Machine learning (ML) algorithm

Logistic regression, decision tree, and support vector machine are frequently used for classification. Logistic regression, as one of the most basic and widely used generalised linear regression models, can map the outcome of the classification problem $y \in (-\infty, \infty)$ to (0,1) using the Sigmoid function. SVM can map vectors to a high-dimensional space and build an optimal decision hyperplane in that space so that the samples between the two classes on either side of the closest plane distance is maximised, thus providing a good generalisation capability for the classification problem of whether or not to default. The DT algorithm is one of the most popular classification models in recent years, classifying invoice data by a series of visual rules to produce a final prediction.

However, with the development of machine learning algorithms, a single prediction model can no longer meet the accuracy requirements of prediction effectively. As a result, integrated models have emerged, such as Random Forest, which is an integrated model composed of multiple decision tree models. Due to the randomness, it has a good tolerance to outliers and noise, and thus can effectively avoid the overfitting problem in the DT algorithm, but its underlying model is still a single model, which is not obvious for the improvement of accuracy. We hope that the integrated learning underlay is not a single ML algorithm, but an ML method that uses a series of algorithms for learning, while using certain rules to integrate all the learning results to obtain better prediction results [19]. Thus even if a weak classifier obtains an incorrect prediction, other weak classifiers can correct these errors [20].

In this study, soft voting method is used for model integration, four common machine learning algorithms include logistic regression, decision tree, support vector machine, and random forest are used as the underlying model. The output weights of each individual classifier are combined with the prediction results to form the final integrated model prediction conclusions.

2.4. Performance evaluation

2.4.1. Traditional evaluation methods

The accuracy, precision, recall, and F1 score calculated from the confusion matrix [21], the Kolmogorov–Smirnov(KS) value obtained from the KS curve [22], and the area under curve (AUC) value obtained from the receiver operating characteristic (ROC) curve were used to evaluate the predictive performance of the models [23].

Table 3. Confusion matrix.

Confusion Matrix		**Predict**	
		0(-)	1(+)
Actual	0(-)	True Negative(TN)	False Positive(FP)
	1(+)	False Negative(FN)	True Positive(TP)

The main parameters of the confusion matrix are listed in Table 3. Six derivative indicators are calculated from the confusion matrix. The precision considers the forecast sample that evaluates the proportion of all samples predicted to be positive (class 1) as actually positive (class 1), thus focusing more on the correct prediction of the positive

sample results. However, a test can cheat and maximize this by only returning a positive result.

Recall considers the original sample and evaluates the proportion of samples that are predicted to be positive among all the samples that are actually positive (class 1). However, a test can cheat and maximize this by returning a positive result.

Owing to the drawback and trade-off between precision and recall rate, F_β-score, the weighted harmonic mean of the precision and recall, is employed to balance these two indicators. β represents the different emphasis on the recall and precision rates. As in the default prediction aspect, the recall rate and precision rate are as important as each other; therefore, β=1 is chosen in the F-score calculation.

The data were divided into ten intervals to calculate KS and ROC values. The KS curve was obtained from the cumulative difference between the default and non-default percentages in each interval; the larger the KS value, the stronger the risk discrimination ability of the model. The AUC value is the area under the ROC curve, which was obtained with the FPR as the horizontal axis and TPR as the vertical axis, the better the discrimination ability.

2.4.2. Explainable ML method

Traditional ML algorithms provide accurate prediction performance but usually lack explanatory power, which makes reluctance of financial institutions to use ML algorithms. Explainability tools, such as the SHAP value, help to increase the transparency of models, revealing that the ML algorithm is superior in terms of both classification performance and explainability.

The Shapley value, which assumes that each variable in the model is a player in a game, explains the prediction result through the marginal contribution of each feature. The Shapley value is the average marginal contribution of the features in all possible coalitions, which is an agnostic model independent tool that can interpret how each feature affects the final prediction in a technologically neutral manner.

Lundberg and Lee developed the Shapley Additive Explanations (SHAP) value, which aims to interpret the prediction by calculating the contribution of each feature to the target variable. Specifically, it calculates the Shapley value of each feature to measure its influence on the final output. SHAP specifies the interpretation of Equation 3.

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z_j' \tag{3}$$

The explanation model g(z′) of prediction f(x) is constructed by an additive feature attribution method, which decomposes the prediction into a linear function of the binary variables $z' \in \{0, 1\}^M$, where z'_j representing whether a feature exists in the decision-making process, while $z'_j = 1$ for observed. M is the number of input variables, and ϕ_j is the characteristic attributed Shapley value for feature j. For feature j, the Shapley value is the weighted sum of a single Shapley value for all possible combinations of features, including different orders.

$$\phi_j(val) = \sum_{S \subseteq \{x_1,\ldots,x_p\} \setminus \{x_j\}} \frac{|S|!(p-|S|-1)!}{P!} \left(val\left(S \cup \{x_j\}\right) - val(S)\right) \tag{4}$$

Equation 4 illustrates the calculation process for the Shapley value of feature j, where S is the subset of features used in the prediction model, x is the vector of features

in the explainable sample, p is the number of features, and val(S) represents the model prediction with feature combination S.

3. Results and discussions

Experiment materials consist of computers with Inter(R) UHD Graphics 620 and Python (Version 3.9.13). The involved libraries includes pandas, numpy, sklearn, calendar, seaborn, matplotlib, random and shap.

3.1. Model results

The performance indicators after the model fitting and prediction are listed in Table 4. F1-score, KS value, and AUC value are considered in the model evaluation. The logistic regression algorithm shows the lowest discrimination power, whereas RF has the best performance amongst the four single classifiers. However, the ensemble model integrated from the soft voting method improved the model performance significantly.

Table 4. Model results for classifiers.

	P	R	F1	KS	AUC
RF	0.9545	0.9130	0.9333	0.916/0.2577	0.9847
DT	0.9545	0.9130	0.9333	0.9445/0.2577	0.9730
LR	0.6129	0.8261	0.7037	0.7421/0.2165	0.9104
SVM	0.9474	0.7526	0.8571	0.859/0.2577	0.9630
Ensemble	1.0	0.9565	0.9778	0.973/0.2577	1.0

Historical credit data usually have many missing values owing to operational risk, and class-imbalanced problems caused by the few defaulted samples. Thereby making it difficult to effectively train the forecast model [24]. The soft voting ensemble method for heterogeneous classifiers considers different weights for each model, provides higher weights for models with better performance, and helps correct errors with each single classifier, and thus improving the discrimination ability of the ensemble learning model. The model results illustrate that higher the proportion of RF and DT, better the performance of ensemble models.

3.2. Model explanation

The trade-off between model complexity and model transparency, and the need to balance the high predictive accuracy brought by sophisticated ML models and inexplicable black boxes has motivated us to introduce explainable methods for further discussion [25]. Previous studies have rarely focused on explaining invoice fields. In this study, invoice related features are considered in the prediction model. SHAP value and the combination of their financial meanings, helps exploit the inherent relationship between default risk and invoice performance, and thereby providing an innovative aspect for credit risk evaluation and prediction.

For model explanation, we select the ensemble soft weighting model with the best performance for further discussion. The summary plot which focuses on the interpretation of the total samples' prediction and the decomposition plot for a single sample's forecast were both considered.

Figure 3 shows the overall impact of features, that is, the importance level of each feature to the target result, which is calculated by the absolute average effects of different

features in the model for each sample. According to the mean SHAP value results, the 2019_out_jshj, 2019_out_bs, 2019_out_se, all_out_fp_y, and all_out_fp_w features rank in the top five important positions that represent the total sales amount for 2019, the total number of sales invoices for 2019, the total tax amount of output invoices in 2019, the number of all valid sales invoices, and the number of all invalid sales invoices, respectively. The 2019 total sales amount ranks as the most critical situation for default prediction, with an impact significantly higher than those of the other features.

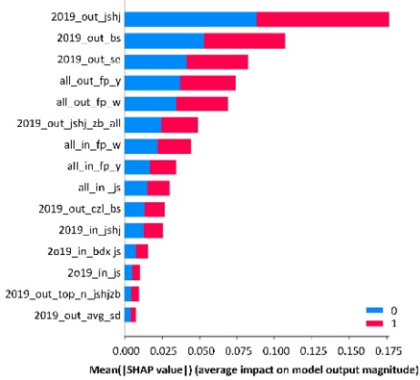

Figure 3. Overall impact of features to model output by SHAP value.

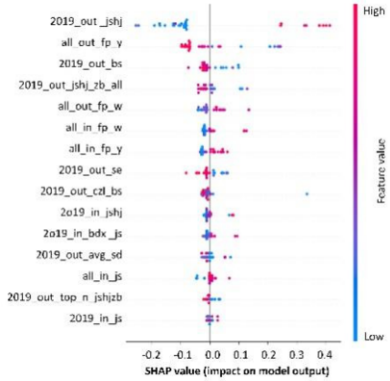

Figure 4. Decomposition plot for SHAP value.

The first two features in Figure 3 represent the sales amount and number in 2019, whereas the third feature represents the actual tax payment. From a financial perspective, the sales performance and tax payment in recent years reflect a company's income and operational conditions, as well as the cash flow and earning ability. Evaluates enterprises' credit capacity based on the company's liquid assets for debt repayment. The number of valid and invalid sales invoices reflects the frequency of transactions; for example, the same sales revenue received in one month or consequently received in several months may affect the cash flow condition significantly. In addition, the percentage of sales in a recent year against the preceding three years and the development rate of invoice numbers help reflect the trend of operational conditions. Furthermore, the amount of procurement invoice data reflects the concentration degree of the enterprise's upstream and downstream trading partners. If transaction counterparties are too concentrated, the enterprise has no upstream bargaining power, especially for the manufacturing enterprise. If the price of upstream materials rise significantly, low bargain power means it may be difficult to strive for a lower cost, thereby affecting the production and operation chains. When a large enterprise has problems, its downstream will be significantly affected and may even lead to a break in the capital chain, and consequently increasing the credit risk of the target enterprise.

Figure 4 shows the SHAP values of each feature for each sample through a scatter plot. The relationship between the size of the feature value and the predicted impact can be seen through color, as well as the distribution of its feature value. A large area indicates a large cluster of samples. The Colors indicate the size of the feature value, with red indicating high feature values, while blue indicating low feature values. For feature 2019_out_jshj, the higher the value of 2019_out_jshj, the more likely it is that the model will predict the company to default (with a positive SHAP value). Similarly, the lower the 2019_out_jshj value, the more likely it is to have a smaller model output value, meaning that the related company is less likely to default (negative SHAP value).

In Figure 4, a large number of 2019_out_jshj samples were clustered in the area with a negative SHAP value.

For a better understanding of the decision-making process, Figure 5 shows the decomposition of a single predicted output with two defaulted and non-defaulted companies. The data for all features were equally divided into ten intervals to determine the contribution of features and their values to the results. The redder the color, the higher the positive importance, and the bluer the color, the higher the negative importance. The figure clearly shows the advantages of the explainable model. This indicates the feature attribution to the target output, not only in a summarized view for the total sample but also differently and specifically for every single company in the test set.

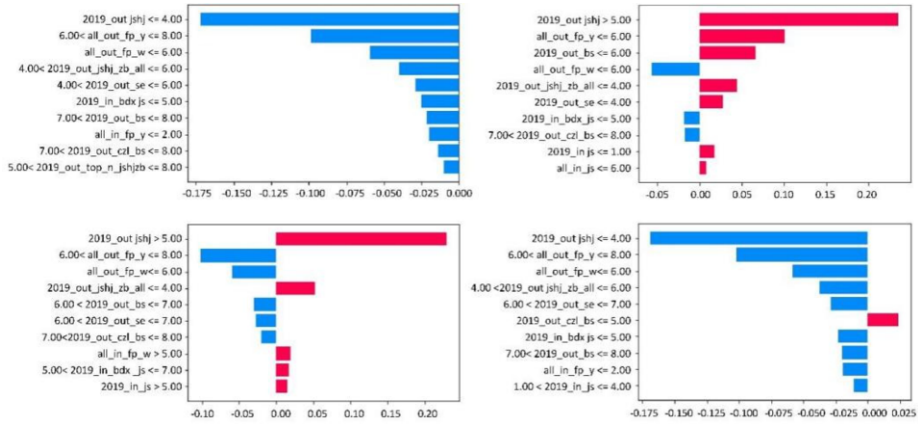

Figure 5. Contribution of each explanatory variable to the Shapley decomposition of four predicted output.

The top three important features for the two non-default companies are both located in the medium interval. The sales amount for the recent year is located in the first four intervals, the total number of valid sales invoices before the eighth interval, and invalid sales invoices number before the sixth interval, are common to both companies. Some social intermediaries who understand the loan access conditions and risk control process of commercial banks provide funds to many SMEs and defraud bank loans by falsely increasing the transaction volume of accounts and fake transactions. As a result, many fake invoices form part of the large number of invoices. Thus, the appropriate distribution of invoice amounts and numbers, which is neither too low nor too high, represents a relatively excellent operating condition and credit quality.

However, Figure 5 shows the most important features of the two defaulted companies differently. The 2019 total sales amount is larger than the fifth interval, and the percentage of the 2019 sales amount against the total sales amount from 2018 to 2021 for both enterprises. The total number of valid sales invoices and the number of 2019 sales invoices is smaller than the sixth interval, and the recent year's tax amount for company 2.

In other words, higher sales amounts in recent years with a lower percentage than the total sales amount from 2018 to 2020 may indicate the fake transactions mentioned before and poor recent business conditions, thus increasing the probability of default of these two companies. For Company 2, the small amount and number of sales invoices also indicate poor business performance, thereby increasing the company's credit risk. As for company 3, except for the common features, the relatively higher invalid number of procurement invoices and the higher volatility of counterparty numbers illustrate the

potential operational risk and unstable business performance; this reduces the company's credit willingness and capacity, and consequently increases credit risk.

4. Conclusions

This study chose invoice data for 425 enterprises in Chongqing. Four ML algorithms and one ensemble method were developed to establish a prediction classifier, and then combined with the SHAP value to explain the feature contribution of a specific output. Our study shows a strong correlation between the derived features and future delinquencies. More importantly, we provided new explainable aspects and related features, such as the invoice field, to forecast enterprises' business performance for further analysis.

However, only invoice data from limited SMEs were considered in this research. This may cause insufficient data problems and thus decrease the model generalization ability. Future research should not only focus on the invoice data but also increase the dimension of data sources. Furthermore, more samples should be collected to identify important features of target variables. For the integrated model part, ensemble learning modeling based on sequential approaches for precise prediction combined with interpretable ML methods should be considered. This helps humans understand the decision-making process of ML algorithms, reduce the black-box part, and increase model transparency. In addition, the effects of extreme events, such as green swan events in a climate disaster, should be considered while modeling credit risk, as they may provide different data patterns and business performances compared with historical ones. It will be interesting to see future solutions to these aforementioned challenges, and the increase in the ability of SME credit risk assessment.

Acknowledgments

Thanks to the Chongqing Institute of Engineering for providing the basic computing environment. Thanks to Shu Yi Xin Credit Management Co., LTD. for providing the relevant IoT data for use in this study.

This study was funded by the scientific research fund of Chongqing Institute of Engineering, China (2022gcky03, 2022xzcr05, 2022xzcr06, 2023xzcr04, 2023xzcr05).

References

[1] Yin C, Jiang C, Jain HK, Wang Z. Evaluating the credit risk of SMEs using legal judgments. Decision Support Systems. 2020;136:113364, doi: 10.1016/j.dss.2020.113364.

[2] Yao G, Hu XJ, Wang GX. A novel ensemble feature selection method by integrating multiple ranking information combined with an SVM ensemble model for enterprise credit risk prediction in the supply chain. Expert Systems with Application. 2022;200:117002, doi: 10.1016/j.eswa.2022.117002.

[3] Jabeur SB, Sadaaoui A, Sghaier A, Aloui R. Machine learning models and cost-sensitive decision trees for bond rating prediction. Journal of the Operational Research Society. 2020;71(8):1161-1179, doi: 10.1080/01605682.2019.1581405.

[4] Chang YC, Chang KH, Wu GJ. Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. Applied Soft Computing. 2018;73:914-20, doi: 10.1016/j.asoc.2018.09.029.

[5] Francesco C, Alessandro G, Giacomo M, Edward A. Rethinking SME default prediction: a systematic literature review and future perspectives. Scientometrics. 2021;126(3):2141-88, doi: 10.1007/s11192-020-03856-0.
[6] Liu Y, Yang ML, Wang YD, Li YS, Xiong TC. Applying machine learning algorithms to predict default probability in the online credit market: Evidence from China. International Review of Financial Analysis. 2022;79:101971, doi: 10.1016/j.irfa.2021.101971.
[7] Chen YR, Leu JS, Huang SA, Wang JT, Takada JI. Predicting Default Risk on Peer-to-Peer Lending Imbalanced Datasets. IEEE Access. 2021;9:1, doi: 10.1109/ACCESS.2021.3079701.
[8] Zhu Y, Xie C, Wang GJ, Yan XG. Comparison of individual, ensemble and integrated ensemble machine learning methods to predict China's SME credit risk in supply chain finance. Neural Computing & Applications. 2017;28(Supply 1):41-50, doi: 10.1007/s00521-016-2304-x.
[9] Tian MH, Li HL, Huang J, Liang J, Bu WB, Chen BL. Credit Risk Models using Rule-Based Methods and Machine-Learning Algorithms. Proceedings of the 2022 6th International Conference on Computer Science and Artificial Intelligence. 2022;203-209, doi: 10.1145/3577530.3577588.
[10] Kim DS, Shin S. The economic explainability of machine learning and standard econometric models-an application to the U.S. mortgage default risk. International Journal of Strategic Property Management. 2021;25(5):396-412, doi: 10.3846/ijspm.2021.15129.
[11] Bussmann N, Giudici P, Marinelli D, Papenbrock J. Explainable Machine Learning in Credit Risk Management. Computational Economics. 2021;57(1):203-216, doi: 10.1007/s10614-020-10042-0.
[12] Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier, ACM; 2016. doi: 10.1145/2939672.2939778.
[13] Bcker M, Szepannek G, Gosiewska A, Biecek P. Transparency, auditability, and explainability of machine learning models in credit scoring. Journal of the Operational Research Society. 2021;6:1-21, doi: 10.1080/01605682.2021.1922098.
[14] Ariza-Garzon MJ, Arroyo J, Caparrini A, Segovia-Vargas MJ. Explainability of a Machine Learning Granting Scoring Model in Peer-toPeer Lending. IEEE Access. 2020;8:64873-64890, doi: 10.1109/ACCESS.2020.2984412.
[15] Lundberg S, Lee SI. A unified approach to interpreting model predictions: Nips; 2017.
[16] Wu Y, Pan YW. Application Analysis of Credit Scoring of Financial Institutions Based on Machine Learning Model. Complexity. 2021;2021(1):9222617(1-12), doi: 10.1155/2021/9222617.
[17] Rosenblatt E. Calculating weight of evidence and information value. Credit Data and Scoring. 2020;2020:99-104, doi: 10.1016/b978-0-12-818815-6.00012-1.
[18] Shen L, Ross S. Information Value of Property Description: A Machine Learning Approach. Journal of Urban Economics. 2020;121:103299, doi: 10.1016/j.jue.2020.103299.
[19] Milford, TM. Review of Multiple regression and beyond: An introduction to multiple regression and structural equation modeling (2nd ed.). Journal of Educational Measurement. 2016;53(2):248-250, doi: 10.1111/jedm.12108.
[20] Li W, Ding S, Wang H, Chen Y, Yang SL. Heterogeneous ensemble learning with feature engineering for default prediction in peer-to-peer lending in China. World Wide Web. 2020;23(1):23-45, doi: 10.1007/s11280-019-00676-y.
[21] Mishra KN, Pandey SC. Fraud Prediction in Smart Societies Using Logistic Regression and k-fold Machine Learning Techniques. Wireless Personal Communications. 2021;119(2):1341-1367, doi: 10.1007/s11277-021-08283-9.
[22] Kovalev MS, Utkin LV. A robust algorithm for explaining unreliable machine learning survival models using the Kolmogorov-Smirnov bounds. Neural networks : the official journal of the International Neural Network Society. 2020;132(0):1-18, doi: 10.1016/j.neunet.2020.08.007.
[23] Carter JV, Pan J, Rai SN, Galandiuk S. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. Surgery. 2016;159(6):1638-1645, doi: 10.1016/j.surg.2015.12.029.
[24] Li H, Qiu H, Sun S, Chang J, Tu WT. Credit scoring by oneclass classification driven dynamical ensemble learning. Journal of the Operational Research Society. 2021;73(1):181-190, doi: 10.1080/01605682.2021.1944824.
[25] Moscato V, Picariello A, Sperlí G. A benchmark of machine learning approaches for credit score prediction. Expert Systems with Applications. 2021;165(0):113986, doi: 10.1016/j.eswa.2020.113986.