

Evaluating the credit risk of SMEs using legal judgments

Chang Yin^a, Cuiqing Jiang^{a,*}, Hemant K. Jain^b, Zhao Wang^a

^a School of Management, Hefei University of Technology, Hefei, Anhui, PR China

^b Gary W. Rollins College of Business, University of Tennessee–, Chattanooga, TN 37403, USA

ARTICLE INFO

Keywords:

Credit risk
Information asymmetry
Legal judgment
SME
Text mining

ABSTRACT

Loan application assessments of small and medium-sized enterprises (SMEs) are difficult because of information asymmetry. To mitigate the information asymmetry, this paper focuses on information found in legal judgments involving the company and its principles and combines this information with financial and firm-specific information to help evaluate the credit risk of SMEs. We propose a framework to identify legal judgments that are effective in predicting credit risk and extract relevant features that are contained within the effective legal judgments. Empirical evaluation shows that features extracted from effective legal judgments significantly improve the discrimination performance and granting performance of our model compared with the baseline model, which uses financial and firm-specific features only.

1. Introduction

Small and medium-sized enterprises (SMEs) significantly contribute to the economy by creating wealth and employment opportunities [16]. However, it is difficult for SMEs to get financial support from the credit market for further development, such as for expanding production capacity [25]. The main reason for this difficulty is information asymmetry between the SME and credit market, including an incomplete system of financial system records, information opacity, and so on [6,8]. Thus, the credit market prefers to lend money to large companies for higher profits and security rather than to SMEs [19,28].

To address the problem of information asymmetry, many studies have focused on non-financial information, such as the firm age, management style, number of employees, and characteristics of the board of directors, to evaluate credit risk [22,23]. Other studies have focused on the productive efficiency [29], business plan [1], and financial reports [34]. However, since SMEs lack mature management [9] and an information disclosure mechanism [5], there are significant challenges in obtaining non-financial information and verifying the authenticity of the information.

In this study, we use published legal judgments (which are referred to hereafter as judgments) involving the enterprise and/or its principles as non-financial information and combine it with the company's financial and firm-specific information (which are referred to hereafter as basic information) to evaluate the credit risk of manufacturing SMEs. Each judgment reflects a dispute between the enterprise and others that arose in the process of business operation. As we know, the legal

sanction in a judgment has a more negative effect on SMEs than on large companies. For example, if a court orders an SME to pay RMB 5,000,000 to another party, the payment may be a hardship that hurts the operation of the SME, or even causes it to close down. However, the same payment may be affordable for a large company. Additionally, judgments include rich information that reflects the risk of default of SMEs to some extent. Consider a case in which an enterprise as defendant was sued for a private lending dispute; the case may reflect the poor credit record of the enterprise because of a lack of willingness to repay and/or an insufficient ability to repay. This information could be useful in evaluating the enterprise's credit risk. Since the judgments are openly available on the Internet, using judgments reduces information collection costs for banks and guarantees the authenticity of the information. The challenges in using judgments are the difficulty in identifying and selecting effective judgments, and the difficulty in extracting important features from the varying structures of the judgment text, features that can improve the performance of the prediction model.

Our study aims to address these challenges by using a framework consisting of three stages. First, based on the legal lexicon and structuring rules, we extract structured information such as the judgment code, date of judgment, lawsuit status of the loan applicant, cause of action, judgment result, and amount awarded. Then, based on taxonomy, we build a method to classify the judgments into four categories based on the lawsuit status and judgment result, and identify the influence of each judgment category. Third, we use the chi-squared test and logistic regression method to identify features with high predictive

* Corresponding author.

E-mail address: jiangcuiq2017@163.com (C. Jiang).

power that can be extracted from effective judgments. We add these features to prediction model for predicting the loan default probability of SMEs. The empirical evaluation shows that the discrimination performance and granting performance of our model are significantly improved by the addition of the judgment information.

This study makes several important contributions to research and practice. To the best of our knowledge, it is the first study to identify the utility of judgment text in predicting the credit risk of SMEs. We find that not all judgments affect credit risk; the judgments with an effect are the ones where the lawsuit status of the loan applicant and judgment results are negative for an applicant (the negative lawsuit status includes defendant, appellee, and so on; the negative judgment results include paying money, freezing assets, and so on). We use text mining techniques to extract features from the judgment text that can significantly improve the discrimination performance of the prediction model. We find that two variables—the number of judgments arising from disputes about loan contracts, and the ratio of the award amount to the yearly income of the company being greater than 12.15%—significantly increase the prediction power of the model over that of the benchmark. In addition, we test the granting performance of the models and examine whether including judgments in evaluating loan applications can bring economic benefits to a bank due to a lower default rate. These results demonstrate that the use of judgments is an effective and low-cost approach for banks and other lending institutions to reduce the losses caused by loan defaults.

The remainder of the paper is organized as follows. The relevant literature is reviewed in the next section. In section 3, we present the proposed framework for mining valuable information from judgments. We describe the empirical evaluation and analysis in section 4. Finally, we conclude this study by summarizing our contributions and discussing future research directions in section 5.

2. Literature review

Decisions on an SME's loan application by banks are generally based on the creditworthiness [35]. Calabrese et al. [3] argued that if a bank can better predict whether or not SMEs will default on their loans, it would be more efficient for the credit system. Thus, accurate assessment of credit risk plays a crucial role in solving the problem of loan availability for SMEs. Credit risk evaluation has focused on two aspects: determining the features impacting credit risk, and predicting the possibility of default.

The features impacting credit risk can be classified into two types: financial features and non-financial features. Since financial features usually reflect the financial performance and repayment ability of an SME [37], many researchers have focused on the role of financial features in the credit risk evaluation of SMEs. Some financial ratios such as the operating profit ratio, current ratio, and total assets turnover are significant default predictors of SMEs [7]. Lin et al. [21] integrated the categories of financial variables into bigger classes and classified the numerical financial ratios into classes by coarse classification. They found that the coding weights generally gave better results when they added these processed classes into models. Gupta et al. [14] examined the effectiveness of operating cash flow information in modelling the credit risk and explaining the financial distress of UK SMEs.

Although financial ratios provide a meaningful and unbiased quantitative representation of the performance of SMEs, the previous research has highlighted the informational opacity of SMEs as a major problem. Many SMEs are not listed on financial markets, and their financial statements may be incomplete or unaudited [23]. Thus, several studies emphasized the significant role of non-financial information in determining the credit risk of SMEs due to the limited scale and scope of hard information [13]. Firm characteristics (size, age, geographical location of the enterprise, etc.) are usually considered to be associated with the bankruptcy risk of SMEs [38]. Psillaki et al. [29] found that managerial inefficiency as non-financial factor is an important ex-ante

indicator of a firm's financial risk, and that firms with more efficiency, more profit, and more liquid assets are less likely to fail. Moro et al. [24] found that a trusting relationship, which can provide more soft information to a loan manager, benefits both banks and SMEs. They examined the value of the manager's trust and found that the higher level of trust SMEs enjoyed from loan managers, the more credit those SMEs received. In addition, distribution and customer networks, supply chain information, the owners' CVs, and eventual awards received by the company are receiving increasing attention as non-financial information [1,33,39].

With the development of text mining technology, many studies began focusing on extracting non-financial information from text as a predictor (or predictors) to evaluate credit risk. Depending on the management quality and market position extracted from credit files, Grunert et al. [12] explored the role of non-financial factors in internal credit rating. They found that the combined use of financial and non-financial factors leads to a more accurate prediction of future default events than the single use of each of these factors. Sánchez et al. [30] utilized non-financial information extracted from audit reports (such as proportion of audited years, number of different auditors hired, average length of auditors' contracts, number of critical qualified audit reports, etc.) to forecast the financial distress of SMEs. The results indicated that the audit of distressed SMEs has several distinctive features: higher audit rotation, more qualified reports, and non-compliance with deadlines to approve and file the annual financial statements. Tsai et al. [34] applied sentiment analysis to study relations between soft information in financial reports and financial risk. They demonstrated the importance of financial sentiment words extracted from financial reports in credit risk prediction. However, many of the types of non-financial information mentioned above are difficult to reliably collect for SMEs. Therefore, these approaches are difficult to use in practice.

Our study strives to fill this gap by focusing on judgments published as text by courts to improve the prediction accuracy of the credit risk of SMEs. Judgments can be easily collected since in most countries they are public and available on the Internet. In China, they can be found at China Judgments Online, a website created by the Supreme People's Court of the People's Republic of China. This public nature guarantees the availability and the reliability of the data source. In addition, legal events, which can be tracked by judgments, can impact the daily operation of SMEs and significantly impact their credit risk.

The development of a prediction model to accurately estimate the probability of default has been a major stream of research. Researchers have used data-driven statistical approaches and artificial intelligence to develop prediction models. Statistical approaches are based on linear discriminant analysis [4] and logistic regression [11]. However, these approaches require various assumptions to be validated. The methods based on artificial intelligence, such as neural networks [2,15], support vector machines [18], and decision trees [26,31], usually outperform statistical models. However, a single-classifier model may not contain enough knowledge for enterprise credit evaluation [32]. To overcome this disadvantage, ensemble models have been widely applied in credit risk evaluation [27,39].

3. Proposed framework

We propose a framework to extract valuable information from judgments and use this information to evaluate the credit risk of SMEs. The framework is illustrated in Fig. 1. To mine information from judgments, we use text mining methods that convert judgment documents into structured information. To identify which judgments are effective in predicting credit risk, we consider two aspects: time and judgment category. Regarding time, we analyze the time span between the date of the judgment and the date of the loan application, to decide on the best observation period. Regarding category, we categorize the judgments based on the lawsuit status of the loan applicant and the results of the judgment and identify the categories of judgment that are

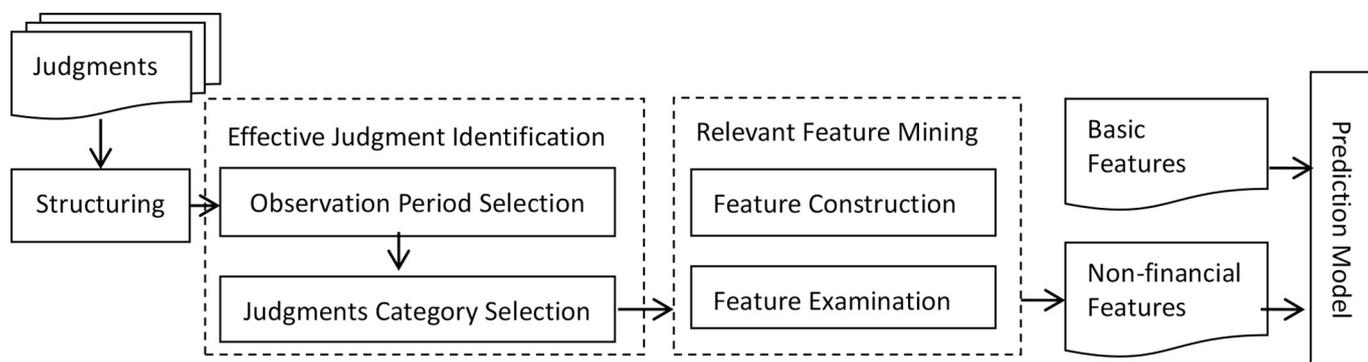


Fig. 1. Framework for incorporating judgment information into credit risk evaluation of SMEs.

relevant for evaluating credit risk. Finally, using a chi-squared test, we extract non-financial features from effective judgments and examine their effect on the prediction model. The steps of the framework are discussed below.

3.1. Extracting structured information from judgments

In judgment documents, the structure of the paragraphs is generally fixed, and professional terms are frequently used. To analyze text in a judgment document, we use the legal lexicon along with regular expressions. We develop several extraction rules that use trigger words and the structure of paragraphs to extract structured information. Appendix A shows an example judgment document to illustrate the extraction process.

3.1.1. Judgment code

The judgment code is a unique identifier of a judgment; its role is to help people find the final judgment of a trial. When either party is dissatisfied with the results of a judgment, the party has the right to appeal. Thus, a trial usually has more than one judgment. Since the judgment code of a previous judgment would be written in the current judgment, we use the judgment code to link all judgments in a trial and to find the final judgment. We extract the code using the special structure of the code and its position in the text.

3.1.2. Date of judgment

The effect of a judgment on predicting the credit risk of SMEs is influenced by the length of the period between the date of the judgment and the date of the loan application. If the period is too long, the judgment may not have much effect on loan default. To select the best observation period, in which the judgments are effective in predicting credit risk, we explore the relationship between the date of judgment (and how much time passes before the loan application) and the company's risk of default. We extract the date of judgment using its position in the document and its unique format.

3.1.3. Lawsuit status

Lawsuit status indicates whether the loan applicant was tried in a court as a defendant or brought a legal case against someone else. In most cases, the judgment has a less negative effect on the loan applicant if the applicant was a plaintiff rather than a defendant. Therefore, we consider that status should affect the ability of a judgment to impact credit risk. We extract status information based on trigger words and the structure of the paragraph, and divide the lawsuit status of loan applicants into two groups: negative (i.e., trigger words of *defendant*, *appellee*, *respondent*, *person subjected to enforcement*, *party against whom execution is filed*) and non-negative (the other trigger words).

3.1.4. Cause of action

The cause of action is the brief summary of the nature and the

process of a case. In this paper, we use a special dispute type as a keyword to represent the cause of action. Some disputes are closely related to default probability, such as private lending disputes and operational disputes. Thus, we explore these disputes that have a strong correlation with the loan default probability. We designed a dictionary based on more than four hundred types of disputes in the law. Based on the dictionary, we extracted the specific dispute type from text using a pattern-based approach and matching phrases.

3.1.5. Judgment result

The judgment result has several parts; we focus on only the parts in which the loan applicant is involved. Since a negative judgment result requires the applicant to pay the award amount and cost, the judgment result affects the applicant and their credit risk. From the result part of the judgment, we extract relevant keywords. For example, “payment” means the applicant should pay money to others and “receiving money” means the applicant should receive money from others. As with lawsuit status, we categorize the results into two types, negative (i.e., *payment*, *joint and several liability*, *freeze assets*, *detain assets*, *sell asset by auction*, *seal up goods*, *seal up assets*, *paying the price of goods*, *returning goods*) and non-negative, based on whether the result of the judgment is adverse to the applicant.

3.1.6. Amount

The amount specified in the judgment is the total money that an applicant should receive or pay in a judgment. If an SME is disciplined by the court and ordered to pay a large amount, it may cause financial distress to the company or may affect day-to-day production and operation. Therefore, we explore the relationship between the amount and the company's probability of default. Due to various writing styles, we use trigger words along with regular expressions to extract the amount from the judgment.

3.2. Selecting an observation period

The effect of a judgment on credit risk will decrease over time. So, we select an observation period that is relevant to credit risk and obtain information on all judgments in that period. To select the observation period, we first calculate the number of days between the date of judgment and the date of the loan application. Based on the distribution of these values, we find the longest time between a judgment and a loan application (6 years in our case). We define six binary variables, denoting whether an applicant has judgments within 1 to 6 years.

For example, if an SME has two or more judgments in 1 year, and the difference between each judgment date and the date of the loan application is within 4 years, the values of the last three variables (having judgment within 4 years, having judgment within 5 years, and having judgment within 6 years) are set to one, and the values of the others are set to zero. We select an observation period for judgments by looking at the relationship between these variables and loan default.

Table 1
The criteria of judgment categorization.

Lawsuit status	Judgment result	
	Negative	Non-negative
Negative	C4	C2
Non-negative	C3	C1

3.3. Categorizing judgments

Since lawsuit status and the results of a judgment have different effects on credit risk, we categorize judgment documents based on both factors. The lawsuit status and results of a judgment are divided into two groups, namely non-negative and negative. Based on this division, the selected judgments are divided into four categories (C1–C4). The criteria of categorization are shown in Table 1.

3.4. Extracting relevant features

3.4.1. Constructing the feature matrix

Before constructing the feature matrix, we represent effective judgments as vector j_i ,

$$j_i = (d_{i1}, d_{i2}, \dots, d_{im}, amount_i), m = 1, 2, \dots, 11 \quad (1)$$

where i is the index of the judgment, m is the index of words representing the cause of action, and $amount$ is the judgment amount. The variable d is a zero-one variable representing whether the term appears or not. If the term appears in the judgment, the value of d is one; otherwise it is zero.

Many enterprises may have more than one judgment. In this case, the judgments of an SME are represented by vector e_p :

$$e_p = \sum j_i \quad (2)$$

where p is the index of the enterprise, and j_i is the judgments belonging to enterprise p .

The amount of each judgment and their accumulated amount may cause the company to default. Various enterprises may have different tolerances for judgment amount. Considering the relationship between judgment amounts and enterprises' yearly income, we construct a feature, $ADYI$ (judgment amount divided by yearly income), by dividing the value of the judgment $amount$ of vector e_p by the enterprise's yearly income. If judgments are published in different years, the enterprise's yearly income equals the average of the yearly incomes of the years in which the judgments were published. In practice, banks are also interested in knowing the critical value of $ADYI$, beyond which the SMEs are more likely to default. To further study the relationship between $ADYI$ and default risk, we discretize $ADYI$ into two intervals and find the critical value by the chi-squared method.

3.4.2. Feature selection

The chi-squared test is a popular feature selection method that can remove irrelevant and redundant attributes to improve the accuracy of a prediction model [17]. In our study, we chose the chi-squared test to measure the correlation between features and loan default probability, and to select features that could improve the performance of the prediction model. Higher values of chi-squared represent a greater correlation. Other feature selection methods could also be used for this purpose. The Correlation-based Feature Selection (CFS) method was used in our experiment, and the results (see Appendix B) are consistent with those of the chi-squared method. Chi-squared can be computed as follows:

$$\chi^2(t, C) = \sum_{t \in \{0,1\}} \sum_{C \in \{0,1\}} (N_{t,C} - E_{t,C})^2 / E_{t,C} \quad (3)$$

where N is the observed frequency and E is the expected frequency for each state of feature t and class C [36].

3.5. Predictive performance

From the model perspective, we demonstrate that judgment information can improve the prediction ability of the loan default prediction model by comparing the discrimination performance of the models. From the application perspective, we discuss the economic effect of using the judgment information by comparing the granting performance of models with and without judgment information.

3.5.1. Discrimination performance

Discrimination performance refers to the ability to distinguish bad loans from good loans. To measure the discrimination performance of the prediction model, we chose two standard measures: the area under the receiver operating characteristic curve (AUC) and the Kolmogorov-Smirnov statistic (KS). The AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive case higher than a randomly chosen negative case [10]. The KS is the maximum difference between the cumulative score distributions of positive and negative cases, and measures the accuracy relative to a single reference point [20]. The higher the AUC and KS values, the greater the discrimination performance of a model.

To estimate the discrimination performance (in terms of AUC and KS) of each prediction model, we performed ten independent ten-fold cross validations, resulting in 100 values of performance estimates. Further, we used full pairwise comparisons to prove that adding features extracted from judgments into models significantly improved the discrimination performance over the baseline model with basic features.

3.5.2. Granting performance

The granting performance refers to the number of defaults under different granting ratios. We first estimated the default probability of each loan application in our sample by using the prediction model through a ten-fold cross validation and thus ranked them. We selected different cut-off values of the percentage of applications approved in our dataset based on the ranked list and calculated the number of defaults (i.e., the granting performance). For example, assume that the bank decides to lend money to the top 50% of the loan applicants in our sample. The number of actual defaults of these selected loans is the granting performance of the model with judgment information, with a 50% cut-off value.

4. Empirical evaluation

4.1. Data

We evaluated the proposed framework on a dataset collected from a commercial bank in the Anhui province of China, and from the websites "China Judgments Online" and "www.qcc.com." To obtain the entire loan profile, we collected the credit loan records and financial data of SMEs that applied for a 12-month loan between 2015 and 2017 (ending between 2016 and 2018). We calculated ten financial features (see Table 2, No. 1 to No. 10) from one year of financial data before the loan application date. As discussed earlier, the financial systems of SMEs are not perfect, so the likelihood of missing data is quite high. We removed the enterprises whose values for all financial factors were null. As a result, the dataset contained 1091 loan observations in the manufacturing sector. To deal with missing values, we used the k-nearest-neighbors algorithm (KNN) to find the k nearest neighbors and filled in the missing values using the neighbors' mean value. The website www.qcc.com is an institution for enterprise credit inquiry. We crawled it for firm-specific information and selected six non-financial features. The statistics of non-financial features are shown in Table 2 (No. 11 to

Table 2
Statistics of basic features.

No.	Feature	Summary statistics			
		Min.	Mean	Max.	S.D.
1	Current ratio (%)	0	100.460	416.590	0.538
2	Debt asset ratio (%)	0	59.300	91.930	0.175
3	Quick ratio (%)	0.080	74.549	160.645	0.231
4	Receivables turnover ratio (years)	0	3.272	10.277	1.974
5	Inventory turnover ratio (years)	0	3.561	9.510	2.103
6	Total assets turnover (years)	0	0.792	4.318	0.809
7	Operating profit ratio (%)	-36.209	-5.586	27.113	0.082
8	Rate of return on common stockholders' equity (%)	-15.873	-1.289	17.288	0.040
9	Return on total assets ratio (%)	-15.294	-0.437	14.383	0.042
10	Missing ratio (%) _a	0	31.450	90.000	0.259
11	Age (years)	1	9.264	37	4.641
12	Registered capital (ten thousand RMB)	3	2266	55,880	4386.557
13	The number of insured	1	30.01	768	44.701
14	The number of patents	0	9.652	237	20.605
15	City	1	8.037	16	4.623
16	The number of shareholders	1	2.467	46	2.464

* *Missing ratio* denotes a ratio calculated using the number of missing values in No. 1 to No. 9, divided by 9; S.D. refers to standard deviation; *The number of insured* refers to the number of employees who got insurance from their employer; there are 16 cities in the Anhui province of China, and the value of *City* is 1 to 16 according to the GDP rank of those cities in 2015.

No. 16).

The China Judgments Online website was used to collect judgments for the 1091 enterprises in our sample. As of January 2019, this website had more than 62.17 million judgments from all of China and had 212.84 million visits. For the 1091 enterprises, we crawled 1920 judgments from the website for the period 2010 to 2017. Forty-four percent of the enterprises in the dataset had at least one judgment. We evaluated these judgments using the proposed framework.

The dependent variable is a binary variable whose value is one if the SME defaulted, and zero if the SME did not default. Following the rules adopted by the bank, we defined a default event as occurring when the payment of a loan is more than 90 days late. There were 1011 non-default loan observations and 80 default loan observations in our sample (7.33% default rate). To analyze the effect of imbalanced data, we used the Synthetic Minority Over-sampling Technique (SMOTE) to examine whether over-sampling will improve model performance. The results are provided in Appendix C.

4.2. Selection of observation period

To select the best observation period before extracting the judgment information, we tested the correlation between the observation period and loan default using the chi-squared method. We then used the logistic regression method to examine the predictive power of each observation period.

Table 3 shows the number of judgments in different periods and the results of the chi-squared test. The variable named “*Having judgment within 2 years*” has the largest chi-squared value. This result indicates that *Having judgment within 2 years* has the highest degree of correlation with defaulting on a loan. In addition, the discriminant of the observation period “*Having judgment within 1 year*” is the lowest. There may be two reasons: first, the effect of a judgment takes some time to appear; second, the number of judgments within 1 year in our dataset is small, which might have influenced the results of the chi-squared test. The results for the other four observation periods show a downward trend. This demonstrates that the effect of judgments on credit risk evaluation gradually diminishes over time.

Table 3
Results of chi-squared test for six observation periods.

Observation period	Number of judgments	Chi-value	p-value
Having judgment within 1 year	887	0.000	0.998
Having judgment within 2 years	1558	0.971	0.325
Having judgment within 3 years	1848	0.196	0.658
Having judgment within 4 years	1898	0.115	0.734
Having judgment within 5 years	1913	0.104	0.747
Having judgment within 6 years	1920	0.093	0.761

Note: Please note that having a judgment within *N* year(s) includes judgments in all years less than *N*. For example, the variable *Having judgment within 4 years* will include judgments in years 1, 2, 3, and 4.

We then used logistic regression for prediction and created six prediction models (H1 to H6) that add each of the above variables to the basic features of the model. Table 4 shows the AUC and KS values of models H1 to H6. The values of AUC and KS of model H2 are the largest, which implies that that variable (*Having judgment within 2 years*) has the strongest predictive power. In addition, only the coefficient for the variable *Having judgment within 2 years* is significant in the results of the prediction models. Thus, we selected two years before the loan application date as the best observation period.

4.3. Examination of judgment category

We divided the judgments in the selected observation period into four categories based on the lawsuit status and the results of the judgment (details about how we categorized are in section 3.3). Table 5 shows the number of judgments in each category, the number of SMEs, and the probability of default. The default probability of SMEs that have judgments in category C4 is the highest, and is almost twice that of the sample average. This indicates that judgments categorized as C4 may have a higher effect on credit risk.

We tested two aspects of each category: whether an enterprise has judgment(s) in a category (*CW1*, *CW2*, *CW3*, *CW4*) and the number of judgments in each category that an enterprise has (*CN1*, *CN2*, *CN3*, *CN4*). We used three classification methods (logistic regression (LR), random forest (RF), and eXtreme Gradient Boosting (XGB)) to develop the prediction model of the loan default risk.

Table 6 shows the AUCs and KSs of the prediction models of the three classification methods (LR, RF, and XGB) using different feature sets (basic features, *CW1*, *CW2*, *CW3*, *CW4*, and their combinations). For example, the combination variable “*CW1234*” denotes that an applicant has at least one judgment in each of the categories C1, C2, C3, and C4. Only the feature set “*B + CW4*” shows significant improvement in the discrimination performance of the models using the three methods. Thus, the variable *CW4* is the best predictor. Therefore, we can conclude the following: first, only some judgments influence credit risk evaluation; and second, adding judgments whose result and applicant status are negative into the model improves the discrimination performance.

To ensure robustness, we tested the effect of judgments on credit risk evaluation using continuous variables also. Table 7 shows the discrimination performance of prediction models of the three classification methods using feature sets (basic features, *CN1*, *CN2*, *CN3*, *CN4*, and their combinations). These variables denote the number of judgments an applicant has in different categories. As with the results shown in Table 6, the feature set “*B + CN4*” shows significant improvement in the discrimination performance of models using the LR and XGB methods. Thus, the feature set “*B + CN4*” is superior to the other feature sets. This conclusion is consistent with that from Table 6.

In summary, these results demonstrate that judgments in category C4 affect the credit risk evaluation of SMEs. Based on this, in our follow-up study, we only retained the judgments with a negative result and negative applicant status.

Table 4
Discrimination performance of prediction models for six observation periods.

Model	Observation period	AUC	KS
H1	Having judgment within 1 year	0.702 (0.681–0.723)	0.455 (0.424–0.486)
H2	Having judgment within 2 years	0.709 (0.689–0.730)	0.461 (0.431–0.492)
H3	Having judgment within 3 years	0.707 (0.686–0.727)	0.459 (0.429–0.489)
H4	Having judgment within 4 years	0.706 (0.686–0.727)	0.458 (0.428–0.489)
H5	Having judgment within 5 years	0.706 (0.686–0.727)	0.459 (0.428–0.489)
H6	Having judgment within 6 years	0.706 (0.686–0.727)	0.459 (0.428–0.489)

Notes: 95% confidence interval in the parentheses. The highest values of AUC and KS are in bold.

Table 5
Statistics of the four judgment categories.

Judgment category	C1	C2	C3	C4
Number of judgments	905	412	32	209
Number of SMEs	300	239	28	125
Default probability of SMEs	5.33%	9.62%	3.57%	14.4%

4.4. Examination of relevant features extracted from legal judgments

To evaluate whether other features of a judgment have prediction power, we examined all 1091 loan applications, integrating their basic features with the features extracted from judgments in category C4. If an applicant didn't have any judgments in category C4, the applicant was treated as having no judgment (the judgment feature extracted was set to zero). The features extracted from judgments are *amount* (the sum of the amounts of each judgment an SME has had to pay within two years), *ADYI* (which denotes dividing the amount by the enterprise's yearly income, as discussed in section 3.4.1), and the disputes summarized from the cause of action. We measured the relationship between these features and loan default by the chi-squared method. In addition, for a better explanation, we chose the logistic regression method to estimate the effect of these features on the applicant's probability of default.

Table 8 shows the results of the chi-squared tests with eleven disputes. The values of *Dispute about loan contracts (D_{LC})*, *Dispute about production and business operation (D_{PBO})*, and *Dispute about business contracts (D_{BC})* are the top three features, and the *D_{LC}* has the largest impact. These results imply that *D_{LC}*, *D_{PBO}*, and *D_{BC}* have a strong correlation with loan default, with the variable *D_{LC}* being the strongest one.

To further estimate the effect of these three variables on the applicant's probability of default, we added them into the prediction models. The AUCs and KSs are shown in Table 9. Model D1 with variable *Dispute about loan contracts (D_{LC})* has the highest value of AUC and KS and is significantly ($p < .01$) better than others. Thus, the variable *D_{LC}* is a stronger predictor to evaluate the credit risk of SMEs than the other two variables (*D_{PBO}*, *D_{BC}*).

Table 6
Discrimination performance of models with features CW1 to CW4 and their combinations.

Feature set	LR		RF		XGB	
	AUC	KS	AUC	KS	AUC	KS
<i>B</i>	0.703(0.682–0.724)	0.458(0.427–0.489)	0.701(0.683–0.719)	0.402(0.367–0.437)	0.700(0.680–0.720)	0.449(0.420–0.478)
<i>B + CW1</i>	0.700(0.680–0.719)	0.445(0.415–0.474)	0.698(0.680–0.716)	0.396(0.361–0.432)	0.698(0.679–0.717)	0.444(0.417–0.470)
<i>B + CW2</i>	0.708(0.688–0.728)	0.463(0.433–0.493)	0.702(0.684–0.720)	0.403(0.367–0.439)	0.703(0.683–0.723)	0.453(0.424–0.483)
<i>B + CW3</i>	0.703(0.682–0.724)	0.459(0.428–0.490)	0.695(0.678–0.713)	0.391(0.357–0.425)	0.701(0.680–0.721)	0.451(0.422–0.480)
<i>B + CW4</i>	0.719(0.699–0.738)	0.476(0.446–0.506)	0.716(0.698–0.734)	0.432(0.395–0.468)	0.718(0.698–0.737)	0.475(0.445–0.504)
<i>B + CW12</i>	0.702(0.681–0.723)	0.458(0.427–0.488)	0.706(0.689–0.723)	0.412(0.378–0.446)	0.698(0.678–0.718)	0.446(0.417–0.475)
<i>B + CW123</i>	0.702(0.681–0.723)	0.457(0.427–0.488)	0.708(0.691–0.725)	0.416(0.382–0.450)	0.698(0.678–0.718)	0.446(0.417–0.475)
<i>B + CW1234</i>	0.709(0.688–0.729)	0.461(0.430–0.491)	0.701(0.683–0.719)	0.402(0.366–0.438)	0.704(0.684–0.725)	0.450(0.421–0.480)

Notes: “*B*” refers to the basic features; based on paired *t*-tests, values of AUC and KS that are significantly improved over the basic features only are represented in bold (p -value $< .05$); 95% confidence interval in the parentheses.

The amount of a judgment is a punishment to an SME, and it may impact the day-to-day operation and production of the SME. Therefore, we considered the amount as significant information to explore. We first added two continuous variables—*amount* and *ADYI*—into the prediction models, but they were not significant. We then used the chi-squared method to find the best critical value, which divided the *ADYI* into two intervals. When *ADYI* was beyond the critical value, the SMEs were more likely to default. Fig. 2 shows the value of chi-squared for different critical values of *ADYI*; the best critical value of *ADYI* is 12.15%, whose chi-squared value is 26.882 ($p < .000$). The result indicates that when *ADYI* is greater than 12.15%, the effect on default probability is highest. Therefore, we created a feature called “*ADYI* > 12.15%” and added it into the prediction model. The results are shown in Table 10 and indicate that the feature *ADYI* > 12.15% is a stronger predictor than are the other two continuous variables.

Based on the above analysis, we can conclude that the dispute about loan contracts (*D_{LC}*) and having an *ADYI* value greater than 12.15% (*ADYI* > 12.15%) are the predictors that can most improve the performance of our prediction model. We added the two predictors into the prediction model and the values of AUC and KS are 0.737 (0.709–0.751) and 0.504 (0.473–0.535), respectively. Based on the baseline model, which uses only basic features, the values of AUC and KS are 0.703 and 0.458, respectively. Thus, the results illustrate the usefulness of judgments as non-financial information in loan default prediction.

4.5. The economic benefit of the proposed prediction model

To evaluate the economic benefit of the proposed model, we analyzed the effect of judgment information from an economic perspective. We used the estimated loan performance specifically known as granting performance to demonstrate the value added by using our model with judgment information. The granting performance represents the amount lost by a bank due to default on loans made to the borrowers. To compare the granting performance of the two prediction models (the model with judgment information and the baseline model), we simulated real credit scenarios and selected different cut-off values of percentage of loan applications approved. The results are shown in Fig. 3.

Table 7
Discrimination performance of models with features *CN1* to *CN4* and their combinations.

Feature set	LR		RF		XGB	
	AUC	KS	AUC	KS	AUC	KS
<i>B</i>	0.703(0.682–0.724)	0.458(0.427–0.489)	0.701(0.683–0.719)	0.402(0.367–0.437)	0.700(0.680–0.720)	0.449(0.420–0.478)
<i>B</i> + <i>CN1</i>	0.696(0.677–0.715)	0.438(0.410–0.466)	0.699(0.683–0.716)	0.398(0.365–0.431)	0.695(0.677–0.714)	0.444(0.417–0.470)
<i>B</i> + <i>CN2</i>	0.710(0.690–0.730)	0.464(0.433–0.494)	0.704(0.686–0.721)	0.407(0.371–0.443)	0.704(0.684–0.724)	0.452(0.423–0.481)
<i>B</i> + <i>CN3</i>	0.703(0.682–0.724)	0.460(0.429–0.491)	0.704(0.686–0.722)	0.408(0.372–0.443)	0.701(0.681–0.722)	0.452(0.422–0.481)
<i>B</i> + <i>CN4</i>	0.723(0.702–0.743)	0.486(0.455–0.517)	0.709(0.691–0.726)	0.418(0.383–0.452)	0.722(0.702–0.742)	0.484(0.454–0.515)
<i>B</i> + <i>CN12</i>	0.701(0.680–0.722)	0.455(0.423–0.486)	0.698(0.679–0.716)	0.395(0.359–0.432)	0.696(0.676–0.716)	0.442(0.414–0.470)
<i>B</i> + <i>CN123</i>	0.701(0.680–0.722)	0.455(0.424–0.486)	0.695(0.677–0.713)	0.389(0.353–0.425)	0.697(0.677–0.716)	0.442(0.414–0.470)
<i>B</i> + <i>CN1234</i>	0.706(0.685–0.726)	0.463(0.431–0.494)	0.688(0.671–0.705)	0.376(0.343–0.409)	0.701(0.680–0.721)	0.451(0.421–0.481)

Notes: “*B*” refers to basic features; based on paired *t*-tests, values of AUC and KS that are significantly improved over the basic features only are represented in bold (*p*-value < .05); 95% confidence interval in the parentheses.

Table 8
Results of chi-squared method for eleven types of disputes.

No.	Variable	Chi-value	p-value
1	Dispute about loan contracts (D_LC)	17.031	0.000
2	Dispute about production and business operation (D_PBO)	2.741	0.098
3	Dispute about business contracts (D_BC)	1.178	0.278
4	Dispute about creditor's rights (D_CR)	0.873	0.350
5	Dispute about labor (D_L)	0.127	0.722
6	Dispute about personality rights (D_PR)	0.042	0.837
7	Dispute about construction projects (D_CP)	0.000	1
8	Others	0.000	1
9	Dispute about guarantee (D_G)	0.000	1
10	Dispute about real right (D_RR)	0.000	1
11	Dispute about shareholders (D_S)	0.000	1

Note: “*Others*” refers to those disputes that are not written clearly in the judgment. For example, the dispute is not mentioned in the body of a judgment.

Table 9
Discrimination performance of models with the top three dispute variables.

Model	Feature set	AUC	KS
D1	<i>B</i> + <i>D_LC</i>	0.725(0.704–0.745)	0.488(0.455–0.521)
D2	<i>B</i> + <i>D_PBO</i>	0.703(0.683–0.724)	0.457(0.425–0.488)
D3	<i>B</i> + <i>D_BC</i>	0.703(0.683–0.724)	0.458(0.427–0.490)

Notes: “*B*” refers to basic features; 95% confidence interval in the parentheses.

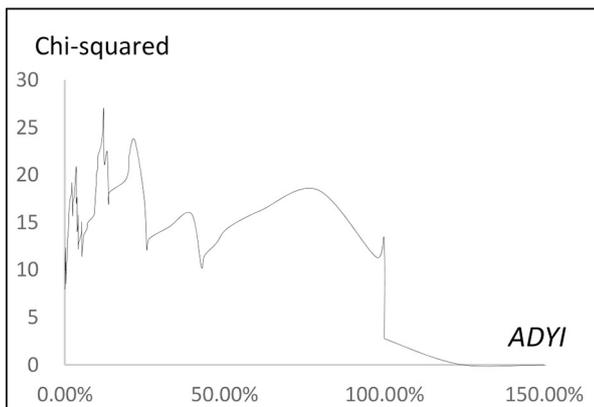


Fig. 2. The value of chi-squared for different critical values of *ADYI*.

As the figure shows, the performance of the proposed model is better for all percentages of loans approved (30% to 100%). Less than 30% approval was not considered feasible by the bank. Thus, when loan decisions are made using the proposed model with judgment information, the granting performance is superior compared to when the loan

Table 10
Discrimination performance of models with features constructed by amount.

Model	Feature set	AUC	KS
A1	<i>B</i> + <i>Amount</i>	0.705 (0.685–0.726)	0.461 (0.430–0.492)
A2	<i>B</i> + <i>ADYI</i>	0.707 (0.686–0.727)	0.460 (0.430–0.491)
A3	<i>B</i> + <i>ADYI</i> > 12.15%	0.734 (0.714–0.755)	0.497 (0.466–0.528)

Notes: “*B*” refers to financial features; 95% confidence interval in the parentheses. The highest values of AUC and KS are in bold.

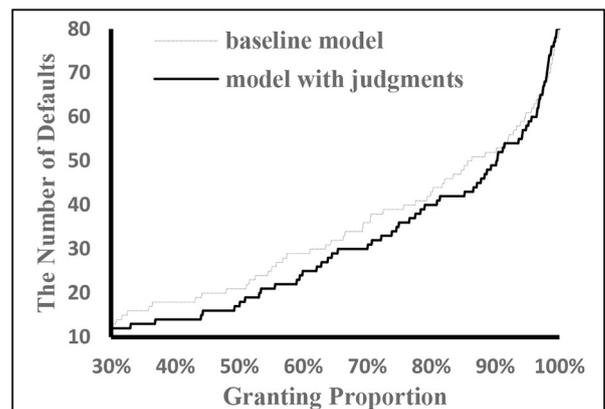


Fig. 3. The granting performance of the two models.

decisions are made using the baseline model. From an economic perspective, using the model with judgment information can help banks reduce financial losses due to loan defaults.

We use the following example to demonstrate the value added by using our model. Let us assume that a bank decides to lend money to 50% of the loan applicants (546 SMEs) in our sample. Since the exact loan amount requested by each applicant is not available (due to privacy reasons), we estimate the total granting (loan) amount by multiplying the average granting (loan) amount (RMB 1,800,000; this estimate was provided by the bank) by the total number of loans approved (546 SMEs), that is, RMB 982,800,000. If we select the 50% top-ranked loan applications based on our model, the number of defaults is expected to be 17. On the other hand, if we select the 50% top-ranked loan applications based on the baseline model, the number of defaults is expected to be 21. Thus, the use of our model improves the loan performance by 19% (i.e., (21–17)/21) compared to use of the baseline model. We assume that when a customer defaults on a loan, not all of the loan amount is lost but on average 30% is lost (based on the business experience of the bank we cooperate with). Considering this, the savings experienced by the bank due to the reduced defaults is RMB 2,160,000.

5. Conclusion and future directions

In this article, we study how legal judgments related to an enterprise can be used to complement basic financial and other information in evaluating the credit risk of SMEs. We propose a framework for mining valuable information from judgments and incorporate this information into our loan default prediction model to help with decision making. Specifically, we examine the effect of four categories of judgments on the default prediction. We extract relevant features from selected judgments and demonstrate that these features can significantly improve the performance of the prediction model.

Our results show that judgments that are less than two years old from the date of the loan application influence predicting the credit risk of SMEs when the judgment status and judgment results are negative (category C4). Additionally, we find that legal judgments related to loan contract disputes strongly correlate with loan default. Also, when ADYI (representing the ratio of judgment amount to annual income of an SME) is greater than 12.15%, the effect on default probability is highest.

Our work contributes to both research and practice. From the research perspective, first, we examine the role of legal judgment information in predicting the probability of the default of SMEs on credit loans. Thus, we extend the literature by identifying a new type of non-financial information (judgments) that can be used in prediction, and examine which features extracted from judgments improve the performance of the prediction model. This is significant because authentic information on legal judgments is easily available in almost all countries. Second, the proposed framework can be used to analyze information contained in legal judgments in other financial-service scenarios, such as P2P lending. Third, the features identified in our

proposed framework also have implications. For example, the feature “absolute value of judgment amount” was not significantly important; however, when we constructed the feature ADYI, defined as “the ratio of the judgment amount to the SME’s yearly income,” and found the best critical value, the ADYI feature became significantly important. Thus, the conversion from absolute value to relative value may make features more useful, as demonstrated by the empirical analysis.

From the practical perspective, judgments and our proposed method can be used in real-world practice by banks and other lending institutions. Exploiting the valuable information contained in judgments can help them mitigate the problem of the information asymmetry of SMEs, thus allowing them to make better decisions and reduce the financial losses caused by default.

Like most research this work has some limitations that may be addressed in future research. First, in this work, we only used a very small part of the judgments. As an extension, it would be interesting to investigate the interaction between the amount awarded in a judgment and the loan amount of the applicant. Second, due to the differences in legal systems in different countries, we chose data only from China. Third, due to a lack of available data, we did not consider fraud and corruption in our evaluation, and we did not apply the framework we provided in other industries, although we believe the framework is a general method.

Acknowledgements

This work was funded by the National Natural Science Foundation of China (Grant Nos. 71731005, 71571059) and the Fundamental Research Funds for the Central Universities (Grant No. JZ2020HGQA0173).

Appendix A. The process of extracting structured information from a judgment

The information extraction from judgments was performed automatically and manually checked for accuracy. We use an example to illustrate the extraction process of structured information (discussed in section 3.1) from a judgment document. To protect privacy, we hid some information such as the name of the parties, and used * to replace it. Before extracting structured information, we focus on the structure of the paragraphs, which is usually fixed in a judgment document. For discussion, we split the judgment text into four parts (shown in Fig. A1 to Fig. A4).

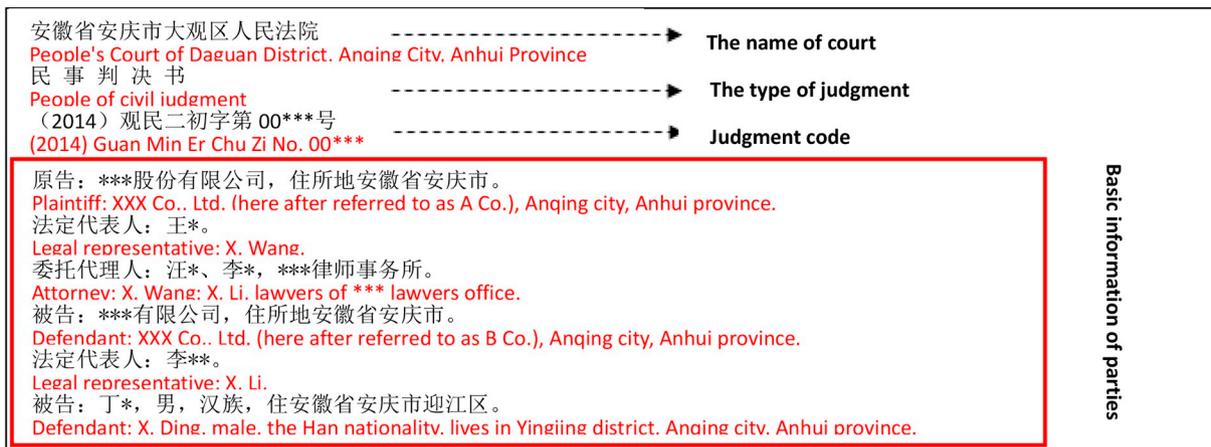


Fig. A1. The first part of a judgment.

Cause of action

原告***股份有限公司诉被告***有限公司、丁*金融借款合同纠纷一案，本院受理后，依法组成合议庭，公开开庭进行了审理。原告***股份有限公司委托代理人汪*、李*到庭参加诉讼。被告***有限公司、丁*经本院合法传唤，无正当理由，拒不到庭。本案现已审理终结。

The financial loan contract dispute case with the plaintiff, A Co., accusing B Co. and X, Ding, the defendant, was formed as collegiate bench legally and had been tried in opened court publicly after being accepted and heard by our court. Plaintiff, A Co., entrusted the agent X, Wang and X, Li to be on the court to participate in the lawsuit. The defendant, B Co. and X, Ding, had received legal calling from our court, and refused to present in the court with no justified reasons. This case trial has been finalized now.

The process of judgment

***股份有限公司诉称：……（具体内容）

A Co. claimed that ... (Note: here we omitted the details about the course of financial loan contact dispute.)

原告***股份有限公司为证明其主张，举出如下证据：……（证据内容）

To support of its claim, the plaintiff A Co. cited the following evidences: ... (Note: here we omitted the description of evidences.)

针对原告***股份有限公司所举证认定如下：

原告***股份有限公司所举证据因被告***有限公司、丁*未提出书面意见，故予以认定。

The court considers these evidences provided by plaintiff A Co. should be admitted because the defendants did not provide the objection(s) or other comment(s) in written.

根据以上认定的证据及当事人的当庭陈述，本院确认如下事实：……（事实描述）

According to the aforementioned evidences which has been admitted and statements of the parties in court, our court identified facts as follows: ... (Note: here we omitted the details of description of facts.)

Fig. A2. The second part of a judgment.

本院认为：……（判决依据），判决如下：

一、判令被告***有限公司自本判决生效后十日内清偿原告***股份有限公司借款本金 1000000 元及利息（利息以借款本金 2000000 元，按月利率 2%为计算标准，自 2014 年 1 月 22 日起至 2014 年 7 月 14 日至；以 1000000 元为本金，按月利率 2%为计算标准，自 2014 年 7 月 15 日起至本判决书确定还款之日止）。二、被告***有限公司逾期不能归还上述借款被告丁*承担保证责任。三、驳回原告***股份有限公司其他诉讼请求。如果未按本判决指定的期间履行给付金钱义务，应当依照《中华人民共和国民事诉讼法》第二百五十三条之规定，加倍支付迟延履行期间的债务利息。

Our court believes that: ... (Note: here we omitted the details about the ground of decision made by the court). The judgment is as follows:

1. The defendant, B Co., is required to pay off the borrowed capital of RMB 1,000,000 and corresponding interests, which is calculated according to borrowing principal RMB 2,000,000 with standard monthly interest rate (2%) from January 22nd, 2014 to July 14th, 2014, and is calculated according to the borrowing principal RMB 2,000,000 under standard monthly interest rate (2%) from July 15th, 2014 to the confirmed repayment date declared in this judgment) to the plaintiff, A Co. within 10 days after the judgment becomes effective. 2. If defendant B Co. fails to repay the borrowed money mentioned above, defendant X, Ding will take the guaranteed responsibility about overdue. 3. All other lawsuit requests by the plaintiff A Co will be rejected... If the plaintiff fails to carry out the obligation of paying the money according to the appointed duration by this judgment, in accordance with the 253rd regulation of Civil Procedure Law of the People's Republic of China, the one double pays the interest of the debt in the duration of delayed performance.

本案诉讼费 14700 元、保全费 5000 元，共计 19700 元，由被告***有限公司承担。

The total fee, RMB 19,700 in which the lawsuit fee is RMB 14,700 and the insurance fee is RMB 5,000, shall be borne by the defendant B Co...

如不服本判决，可在判决书送达之日起十五日内，向本院递交上诉状，并按对方当事人的人数提出副本，上诉于安徽省安庆市中级人民法院。

If not satisfied with this judgment, within 15 days after the date of delivery of this written judgment, submit the appellate petition to our court, and appeal to the intermediate people's court in Anqing, Anhui Province with the transcript according to the number of people from opposite party.

Fig. A3. The third part of a judgment (results of the judgment).

审判长 李* 审判员 杜* 人民陪审员 王*	Presiding judge: X, Li Judicial Officer: X, Du People's assessor: X, Wang	Officers' Information
二〇一四年九月二日 书记员 江*	September 2, 2014 Clerk: X. Jiang	Date of judgment

Fig. A4. The fourth part of a judgment.

A.1. Judgment code

We extract the *judgment code* from the first part of the judgment based on its position and naming structure. It usually appears after the type of judgment (see Fig. A1). We collected all judgment types and created a bag of words to match judgment type. In addition, the *judgment code* has a special structure (i.e., (a particular year) XXX Zi No. XXX). We scan the document line by line from the beginning; when we find a word group in a line that conforms to the naming structure and find a word group in the previous line that matches a judgment type, we extract the words in the line as the *judgment code*.

A.2. Lawsuit status

We select the keywords (shown in Table A1) associated with lawsuit status from the legal lexicon to extract the applicant's *lawsuit status*, which appears in the basic information of the parties (see Fig. A1). We scan the document from the judgment code line onward and find the first place

where the name of the applicant appears. If a word, written before the applicant's name, is found in our bag of words, we consider the keyword to represent the *lawsuit status*.

Table A1

Keywords associated with lawsuit status of loan applicant.

Lawsuit status	Keywords
Negative	Defendant (449), appellee (145), respondent (52), person subjected to enforcement (130), party against whom execution is filed (0)
Non-negative	Plaintiff (702), appellant (92), applicant (86), execution applicant (257), retrial applicant (4), third party (3), claimant (0)

Note: The frequency of occurrence is in parentheses.

A.3. Cause of action

We extract the keyword of *cause of action* from the second part of the judgment (see Fig. A2) by a pattern-based approach. The first sentence of the *cause of action* is usually written as follows: “the XXX dispute case with the plaintiff XXX accusing the defendant XXX...” We consider the specific dispute as the keyword of the *cause of action*. Therefore, we use the regular expressions and the legal lexicon to find the specific dispute. In the predefined dictionary, there are 1960 terms related to disputes. Due to the space limitation of the article, we uploaded the word list to a website (<https://doi.org/10.7910/DVN/LWFPDS>). In our example, the specific dispute is “the financial loan contract dispute.”

A.4. Judgment result

This structured information was extracted from the third part of the judgment (see Fig. A3). We extract the sentences between the two phrases “the sentence shown below:” and “the total fee ... the lawsuit fee ... the insurance fee ...”. Then we split these sentences by the sequence number and only retained the items involving the applicant. After analyzing the content of the judgment results in our sample, we constructed 29 regular expressions to match each item in the judgment results. The keywords shown in Table A2 are part of these regular expressions, and each keyword belongs to only one regular expression and vice versa. If an item matches one of these regular expressions, we use the keyword of that regular expression to represent the item.

Table A2

The keywords used in the description of the results of a judgment.

Judgment result	Keywords
Negative	Payment (136), joint and several liability (14), frozen assets (119), detained assets (14), sold asset by auction (7), sealed up assets (32), returning goods (1), compulsory execution (4), detained the income from the creditor's rights (2), issuing false invoice (1), the notice of discharging obligation (12)
Non-negative	Receiving money (250), suspension of payment (1), no compensation shall be paid (1), preserved assets (2), release of frozen assets (35), release of the sealed assets (15), release of the detained assets (7), allowed to withdraw a lawsuit (414), allowed to withdraw an appeal (6), revoked a case placed on file (1), no labor relation (1), confirmed the labor relation (2), execution termination (99), lawsuit termination (3), revised the mistake in the judgment (13), null (584), affirmed the original judgment (166), the notice of assisting in enforcement (6)

Notes: The frequency of occurrence in parentheses. A “null” means that the loan applicant was not involved in any items of the judgment results.

In our example, the B Co. is the loan applicant. The phrases “B Co., pays off... to the plaintiff, A Co. ...” in the first item indicate that B Co. should pay money to another party in this case. Therefore, we used “payment” to represent this item. Regarding the second item, since the main object is defendant X. Ding, this item was discarded. We also discarded the third item for the same reason as the second.

A.5. Amount

We extract amount from each item of the judgment results that relate to the applicant and are written in the form “... pays ... RMB XXX”. We add them together to give the value of the *amount*. There are various ways to express an amount in Chinese. In our example (see Fig. A3), the information about *amount* is the first item of the results, and the value of *amount* is “RMB 1,000,000”.

The extraction of the *amount* is complicated and there are many ways to write it. Fig. A5 gives another example. The applicant is the plaintiff A Co., and we split the results of the judgment into three items according to the sequence number. We retain the first two items, which relate to the applicant. From the first item, we extract “RMB 1,290,000” and from the second item, we extract “RMB 185,760.” We then add these amounts together. Thus, the value of the *amount* is RMB 1,475,760.

判决如下：
 一、任**于本判决生效之日起十日内向 *** (A) 有限公司支付混凝土货款本金 129 万元；二、任**于本判决生效之日起十日内向*** (A) 有限公司支付混凝土货款逾期付款违约金 185760 万元（计算至 2014 年 5 月 4 日，之后损失以实际未付货款为基数，按每日万分之六计算至付清货款时止）；三、*** (B) 有限公司对本判决第一、二项内容承担连带清偿责任。如果未按本判决指定的期间履行给付金钱义务，应按照《中华人民共和国民事诉讼法》第二百五十三条的规定，加倍支付迟延履行期间的债务利息。
 案件受理费 18082 元，由任**、*** (B) 有限公司负担（于本判决生效之日起七日内向本院缴纳）。
 The sentence has shown below:
 1. The defendant, X, Ren pays off the payment for concrete of 1,290,000 Yuan to the plaintiff, A Co. within 10 days after the judgment goes into effect. 2. The defendant, X, Ren pays the penalty 185,760 Yuan to the plaintiff, A Co. within 10 days after the judgment goes into effect (this penalty is calculated until May 4, 2014, and the subsequent loss is based on the actual unpaid payment, calculated at 0.06% per day until the payment is paid off). 3. Defendant B Co. takes the guaranteed responsibility for the 1, 2 items. If one does not carry out the obligation of paying the money according to the appointed duration by this judgment, in accordance with the 253rd regulation of Civil Procedure Law of the People's Republic of China, the one double pays the interest of the debt in the duration of delayed performance.
 The total fee, 18,082 Yuan, shall be borne by the defendant X, Ren and B Co. (It shall be paid to this court within seven days after the effective date of this judgement).

Fig. A5. An example of extracting an amount.

A.6. Date of judgment

The date of a judgment is written at the end of the document (the fourth part of a judgment) and includes the “year,” “month,” and “day” in Chinese characters. We go through the document from the bottom up and find the line in which the words conform to the naming structure (XXXX Nian XX Yue XX Ri, in Chinese characters). We extract the words in this line and change them into a standard form (like “0902, 2014” for September 2, 2014, in our example; see Fig. A4).

Appendix B. Results of feature selection methods

The chi-squared test is a traditional feature selection method and often used to examine the independence of two events. We chose this test to measure the correlation between features and default probability. To make a sufficient validation, the Correlation-based Feature Selection (CFS) method was used in our experiment. The results (shown in Tables B1 and B2) are consistent with the chi-squared method.

Table B1
 The results of feature selection methods for six observation periods.

No	Feature	Chi-squared	CFS merits
1	Having judgment within 1 year	0.000 (0.998)	0.004
2	Having judgment within 2 years	0.971 (0.325)	0.033
3	Having judgment within 3 years	0.196 (0.658)	0.017
4	Having judgment within 4 years	0.115 (0.734)	0.014
5	Having judgment within 5 years	0.104 (0.747)	0.013
6	Having judgment within 6 years	0.093 (0.761)	0.013

Note: *p*-values are in parentheses.

Table B2
 The results of feature selection methods for eleven types of disputes.

No.	Feature	Chi-squared	CFS merits
1	Dispute about loan contracts (D_LC)	17.031 (0.000)	0.151
2	Dispute about production and business operation (D_PBO)	2.741 (0.098)	0.058
3	Dispute about business contracts (D_BC)	1.178 (0.278)	0.042
4	Dispute about creditor's rights (D_CR)	0.873 (0.350)	0.039
5	Dispute about labor (D_L)	0.127 (0.722)	-0.027
6	Dispute about personality rights (D_PR)	0.042 (0.837)	-0.026
7	Dispute about construction projects (D_CP)	0.000 (1)	-0.003
8	Others	0.000 (1)	-0.012
9	Dispute about guarantee (D_G)	0.000 (1)	-0.009
10	Dispute about real right (D_RR)	0.000 (1)	-0.009
11	Dispute about shareholders (D_S)	0.000 (1)	-0.009

Note: *p*-values are in parentheses; the value of CFS merits of other feature subsets are lower than the value of feature D_LC; (D_LC, D_PBO) = 0.149, (D_LC, D_BC) = 0.135, (D_LC, D_CR) = 0.129, (D_LC, D_L) = 0.088, (D_LC, D_PR) = 0.089, (D_LC, D_CP) = 0.104, (D_LC, Others) = 0.099, (D_LC, D_G) = 0.101, (D_LC, D_RR) = 0.101, (D_LC, D_S) = 0.101.

Appendix C. Sensitive analysis with SMOTE method

Table C1 shows the discrimination performance of models “using” and “not using” the SMOTE method. We tested the statistical significance of discrimination performance using a non-parametric test (full pairwise). The results of the two methods (LR and RF) are statistically significant ($p < .05$), and the result of the XGB method is not significant ($p > .1$).

Table C1
Discrimination performance of models “using” and “not using” SMOTE.

	Feature set	Not using SMOTE		Using SMOTE	
		AUC	KS	AUC	KS
LR	<i>B</i>	0.703(0.682–0.724)	0.458(0.427–0.489)	0.694(0.674–0.714)	0.450(0.421–0.480)
	<i>B</i> + <i>Having judgment within 2 years</i>	0.709(0.689–0.730)	0.461(0.431–0.492)	0.703(0.683–0.724)	0.454(0.424–0.484)
	<i>B</i> + <i>CW4</i>	0.719(0.699–0.738)	0.476(0.446–0.506)	0.708(0.691–0.726)	0.456(0.430–0.483)
	<i>B</i> + <i>CN4</i>	0.723(0.702–0.743)	0.486(0.455–0.517)	0.714(0.694–0.734)	0.478(0.447–0.508)
RF	<i>B</i> + <i>D_{LC}</i> + <i>ADYI</i> > 12.15%	0.737(0.717–0.758)	0.504(0.473–0.535)	0.730(0.709–0.750)	0.490(0.459–0.522)
	<i>B</i>	0.701(0.683–0.719)	0.402(0.367–0.437)	0.716(0.699–0.732)	0.431(0.398–0.465)
	<i>B</i> + <i>Having judgment within 2 years</i>	0.708(0.691–0.726)	0.417(0.383–0.451)	0.722(0.706–0.738)	0.444(0.411–0.476)
	<i>B</i> + <i>CW4</i>	0.716(0.698–0.734)	0.432(0.395–0.468)	0.728(0.710–0.747)	0.456(0.420–0.493)
XGB	<i>B</i> + <i>CN4</i>	0.709(0.691–0.726)	0.418(0.383–0.452)	0.721(0.705–0.737)	0.442(0.410–0.475)
	<i>B</i> + <i>D_{LC}</i> + <i>ADYI</i> > 12.15%	0.728(0.710–0.746)	0.457(0.421–0.492)	0.736(0.720–0.752)	0.472(0.441–0.504)
	<i>B</i>	0.700(0.680–0.720)	0.449(0.420–0.478)	0.696(0.676–0.715)	0.450(0.422–0.479)
	<i>B</i> + <i>Having judgment within 2 years</i>	0.705(0.684–0.725)	0.450(0.421–0.480)	0.706(0.685–0.726)	0.456(0.427–0.485)
	<i>B</i> + <i>CW4</i>	0.718(0.698–0.737)	0.475(0.445–0.504)	0.716(0.696–0.736)	0.477(0.447–0.506)
	<i>B</i> + <i>CN4</i>	0.722(0.702–0.742)	0.484(0.454–0.515)	0.716(0.696–0.737)	0.477(0.447–0.507)
	<i>B</i> + <i>D_{LC}</i> + <i>ADYI</i> > 12.15%	0.733(0.713–0.752)	0.498(0.469–0.527)	0.727(0.707–0.747)	0.494(0.464–0.524)

Notes: “*B*” refers to the basic features; 95% confidence interval in the parentheses.

References

- S. Angilella, S. Mazzù, The financing of innovative SMEs: a multicriteria credit rating model, *Eur. J. Oper. Res.* 244 (2) (2015) 540–554.
- J.K. Bae, J. Kim, Combining models from neural networks and inductive learning algorithms, *Expert Syst. Appl.* 38 (5) (2011) 4839–4850.
- R. Calabrese, G. Marra, S.A. Osmetti, Bankruptcy prediction of small and medium enterprises using a flexible binary generalized extreme value model, *J. Oper. Res. Soc.* 67 (4) (2016) 604–615.
- P. Casin, Categorical multiblock linear discriminant analysis, *J. Appl. Stat.* 45 (8) (2017) 1–14.
- X. Chen, X. Wang, D.D. Wu, Credit risk measurement and early warning of SMEs: an empirical study of listed SMEs in China, *Decis. Support. Syst.* 49 (3) (2010) 301–310.
- Y. Chen, R.J. Huang, J. Tsai, L.Y. Tzeng, Soft information and small business lending, *J. Financ. Serv. Res.* 47 (1) (2015) 115–133.
- L. Cultrera, X. Brédart, Bankruptcy prediction: the case of Belgian SMEs, *Rev. Acc. Financ.* 15 (1) (2016) 101–119.
- L. D’Aurizio, T. Oliviero, L. Romano, Family firms, soft information and bank lending in a financial crisis, *J. Corp. Finan.* 33 (2015) 279–292.
- A. Dietrich, Explaining loan rate differentials between small and large companies: evidence from Switzerland, *Small Bus. Econ.* 38 (4) (2012) 481–494.
- T. Fawcett, An introduction to ROC analysis, *Pattern Recogn. Lett.* 27 (8) (2006) 861–874.
- R. Ge, J. Feng, B. Gu, P. Zhang, Predicting and deterring default with social media information in peer-to-peer lending, *J. Manag. Inf. Syst.* 34 (2) (2017) 401–424.
- J. Grunert, L. Norden, M. Weber, The role of non-financial factors in internal credit ratings, *J. Bank. Financ.* 29 (2) (2005) 509–531.
- J. Grunert, L. Norden, Bargaining power and information in SME lending, *Small Bus. Econ.* 39 (2012) 401–417.
- J. Gupta, N. Wilson, A. Gregoriou, J. Healy, The value of operating cash flow in modelling credit risk for SMEs, *Appl. Financ. Econ.* 24 (9) (2014) 649–660.
- P. Hajek, Municipal credit rating modelling by neural networks, *Decis. Support. Syst.* 51 (1) (2011) 108–118.
- U. Hamzani, D. Achmad, Bankruptcy Prediction: SMEs Case Study in Pontianak, Social Science Electronic Publishing, Indonesia, 2018.
- X. Jin, A. Xu, R. Bie, P. Guo, Machine learning techniques and Chi-Square feature selection for cancer classification using SAGE gene expression profiles, *Data Mining for Biomedical Applications*, April. DBLP, Pakdd Workshop, Biodm, Singapore, 2006.
- H.S. Kim, S.Y. Sohn, Support vector machines for default prediction of SMEs based on technology credit, *Eur. J. Oper. Res.* 201 (3) (2010) 838–846.
- V. Kuntchev, R. Ramalho, J. Rodriguezmeza, J.S. Yang, What Have we Learned from the Enterprise Surveys Regarding Access to Credit by SMEs? (2013) (Policy Research Working Paper).
- S. Lessmann, B. Baesens, H.V. Seow, L.C. Thomas, Benchmarking state-of-the-art classification algorithms for credit scoring: an update of research, *Eur. J. Oper. Res.* 247 (1) (2015) 124–136.
- S.M. Lin, J. Ansell, G. Andreeva, Predicting default of a small business using different definitions of financial distress, *J. Oper. Res. Soc.* 63 (4) (2012) 539–548.
- L. Lugovskaya, Predicting default of Russian SMEs on the basis of financial and non-financial variables, *J. Financ. Serv. Mark.* 14 (4) (2010) 301–313.
- Y. Ma, J. Ansell, G. Andreeva, Exploring management capability in SMEs using transactional data, *J. Oper. Res. Soc.* 67 (1) (2014) 1–8.
- A. Moro, M. Fink, Loan managers’ trust and credit access for SMEs, *J. Bank. Financ.* 37 (3) (2013) 927–936.
- A. Moro, M. Fink, D. Maresch, Reduction in information asymmetry and credit access for small and medium-sized enterprises, *J. Financ. Res.* 38 (1) (2015) 121–143.
- D.L. Olson, D. Delen, Y. Meng, Comparative analysis of data mining methods for bankruptcy prediction, *Decis. Support. Syst.* 52 (2) (2012) 464–473.
- M. Papouskova, P. Hajek, Two-stage consumer credit risk modelling using heterogeneous ensemble learning, *Decis. Support. Syst.* 118 (2019) 33–45.
- K. Pirzada, D. Wickramasinghe, Gabriël A Moens, Hamid, A. F. A., Viverita, & Lubis, A. W., Foreign bank entry and credit allocation to SMEs: evidence from ASEAN countries, *Procedia Soc. Behav. Sci.* 211 (2015) 1049–1056.
- M. Psillaki, I.E. Tsolas, D. Margaritis, Evaluation of credit risk based on firm performance, *Eur. J. Oper. Res.* 201 (3) (2010) 873–881.
- Carlos Piñero Sánchez, P.D.L. Monelos, Manuel Rodríguez López, A parsimonious model to forecast financial distress, based on audit evidence, *Contaduría y Administración* 58 (4) (2013) 151–173.
- S.Y. Sohn, J.W. Kim, Decision tree-based technology credit scoring for start-up firms: Korean case, *Expert Syst. Appl.* 39 (4) (2012) 4007–4012.
- J. Sun, J. Lang, H. Fujita, H. Li, Imbalanced enterprise credit evaluation with DTE-SBD: decision tree ensemble based on SMOTE and bagging with differentiated sampling rates, *Inf. Sci.* 425 (C) (2018) 76–91.
- E. Tობბაკ, T. Bellotti, J. Moeyersoms, M. Stankova, D. Martens, Bankruptcy prediction for SMEs using relational data, *Decis. Support. Syst.* 102 (2017) 69–81.
- M.F. Tsai, C.J. Wang, On the risk prediction and analysis of soft information in finance reports, *Eur. J. Oper. Res.* 257 (1) (2017) 243–250.
- I.E. Tsolas, Firm credit risk evaluation: a series two-stage DEA modeling framework, *Ann. Oper. Res.* 233 (1) (2015) 483–500.
- A.K. Uysal, S. Gunal, A novel probabilistic feature selection method for text classification, *Knowl.-Based Syst.* 36 (2012) 226–235.
- F. Voulgaris, D.C. Zopounidis, On the evaluation of Greek industrial SMEs’ performance via multicriteria analysis of financial ratios, *Small Bus. Econ.* 15 (2) (2000) 127–136.
- N.H. Wellalage, S. Locke, Factors affecting the probability of SME bankruptcy: a case study on New Zealand unlisted firms, *Bus. J. Entrepreneurs* 2012 Issue 2 (2012) 109–126.
- Y. Zhu, C. Xie, G.J. Wang, X.G. Yan, Comparison of individual, ensemble and integrated ensemble machine learning methods to predict china’s SME credit risk in supply chain finance, *Neural Comput. & Applic.* 28 (1) (2017) 41–50.

Chang Yin is a doctoral student at the School of Management, Hefei University of Technology, China. Her research interests include credit risk evaluation of SMEs and data mining.

Cuiqing Jiang is a Professor at School of Management, Hefei University of Technology. He received his PhD degree in 2007 from Hefei University of Technology. His research interests include big data analytics and business intelligence, data mining and knowledge discovery, information systems, and financial technology (Fintech). He has published in such journals as *Journal of Management Information Systems*, *European Journal of Operational Research*, *Information Sciences*, *Decision Support Systems*, and *International Journal of Production Research*.

Hemant K. Jain is W. Max Finely Chair in Business, Free Enterprise and Capitalism and Professor of Data Analytics, in Gary W. Rollins College of Business at University of Tennessee Chattanooga. He is internationally acclaimed for his pioneering work on Effectiveness of Presentation of Product Information in E-Business Systems. His work has appeared in *Information Systems Research*, *MIS Quarterly*, *IEEE Transactions on Software Engineering*, *Journal of MIS*, *IEEE Transactions on Systems Man and Cybernetics*, *Naval*

Research Quarterly, Decision Sciences, Decision Support Systems, Communications of ACM, and Information & Management. He served as Associate Editor-in-Chief of IEEE Transactions on Services Computing and as Associate Editor of Journal of AIS. Recently he served as Program Chair, of IEEE International Conference on Big Data. He received his Ph. D. in information system from Lehigh University, a M. Tech. from IIT Kharagpur, and B. E. University of Indore, India.

Zhao Wang is an Assistant Professor at the School of Management, Hefei University of Technology. He received his PhD degree in management science and engineering from that university. His research interests include data mining and credit scoring. He has published in such journals as *Journal of Management Information Systems*, *European Journal of Operational Research*, *Annals of Operations Research*, *Electronic Commerce Research and Applications*, and many others.