Decision Support Systems 140 (2021) 113429

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection

Gang Kou [a], Yong Xu [a], Yi Peng [b,*], Feng Shen [c], Yang Chen [a,*], Kun Chang [d], Shaomin Kou [d]

[a] School of Business Administration, Southwestern University of Finance and Economics, Chengdu 611130, China
[b] School of Economics and Management, University of Electronic Science and Technology of China, Chengdu 611731, China
[c] School of Finance, Southwestern University of Finance and Economics, Chengdu 611130, China
[d] Data Application Department, Shandong City Commercial Banks Alliance Co., Ltd, Jinan 250104, China

ARTICLE INFO

Keywords:
Bankruptcy prediction
Payment and transactional data
Expected maximum profit
Data imbalance
Feature selection

ABSTRACT

Many bankruptcy prediction models for small and medium-sized enterprises (SMEs) are built using accounting-based financial ratios. This study proposes a bankruptcy prediction model for SMEs that uses transactional data and payment network–based variables under a scenario where no financial (accounting) data are required. Offline and online test results both confirmed the predictive capability and economic benefit of transactional data–based variables. However, incorporating those features in predictive models produces high dimensional problems, which deteriorates model interpretability and increases feature acquisition costs. Thus, we propose a two-stage multiobjective feature-selection method that optimizes the number of features as well as model classification performance. The results showed that the proposed model achieved similar classification performance while greatly reducing the cardinality of the feature subset. Finally, the feature importance evaluation for features in the optimal subset confirmed the importance of transactional data and payment network-based variables for bankruptcy prediction.

1. Introduction

Small and medium-sized enterprises (SMEs) are an important part of the global economy [33]. In China, SMEs are the foundation of economic growth and have become important drivers of social progress [28]. The Chinese government has implemented various policies to support the growth of private companies, including credit increases and reductions in taxes and fees. Nevertheless, access to financing remains a challenge for SMEs. For large enterprises, certified audited financial statements are used to evaluate credit risk and support financial decision-making such as granting loans [3,33]. For SMEs, however, due to a lack of reliable data and other factors, evaluating credit risk is complicated and costly for banks [33]. On the one hand, banks use relationship lending to accumulate soft information over time to deal with credit-data scarcity; on the other, SMEs often face high costs when accessing financing due to information opacity and high bankruptcy risk [26,27,33]. Moreover, compared with large firms, SMEs are more vulnerable to changes in the external environment [5]. In 2018, the Financial System Survey Report of Chinese SMEs, issued by the People's Bank of China, noted that SMEs face various macroeconomic constraints—including changes in the domestic economic structure and international trade friction—which have also made it more difficult for SMEs to access financing.

Most studies use financial ratio–based variables constructed from financial statements as the main variables for bankruptcy prediction [39,44]. In reality, however, the annual account information of SMEs is not available for banks and is unreliable due to the lack of an internal control system [5,44]; this makes it difficult for banks to evaluate the credit risk of SMEs and make financial decisions [51]. Moreover, financial statements only provide a snapshot of a firm's financial situation in the past year and do not reflect its latest operational status [44]. In a dynamically changing economic environment, banks prefer timely credit-risk models that reflect the latest operating status of SMEs. New credit-data sources that are constantly updated are preferred in this scenario. For banks, SMEs payment and transaction records are registered daily; these records contain information related to daily transactions, wages, tax payments, and other information. In the era of big

* Corresponding authors.
E-mail addresses: pengyi@uestc.edu.cn (Y. Peng), chenyang@swufe.edu.cn (Y. Chen).

https://doi.org/10.1016/j.dss.2020.113429
Received 23 April 2020; Received in revised form 20 October 2020; Accepted 21 October 2020
Available online 3 November 2020
0167-9236/

data, these data are easily stored by banks and provide useful information for evaluating the credit risk of SMEs.

Using bank-provided data,[1] this study constructed a bankruptcy prediction model using transactional data–based features as predictors for SMEs under a scenario where no financial (accounting) data are required. Compared with financial statements, payment and transactional data have the following advantages: (1) The data are reliable and of high quality. (2) For banks, bank reconciliation data are only available if the company has an account with the bank. (3) Unlike annual account data, which provide static information about a company's financial situation, transactional data are registered in a timely manner and reflect the latest operational conditions of SMEs. Meanwhile, relational data for daily interactions, including common managers/shareholders and transaction interactions, were also considered in our model. To evaluate model performance, in addition to commonly used statistical measures, net present value (NPV) is also an important indicator for banks in financial decision-making [18]. Based on NPV, we propose a profit-based measure to better estimate economic profit under a data-imbalance scenario. Finally, a two-stage multiobjective feature-selection method is proposed to select the most relevant features in bankruptcy prediction. Irrelevant and noisy features were removed using feature-ranking and optimum-seeking methods. Meanwhile, compact feature subsets [30] were found using a multiobjective wrapper-based feature-selection method in the second stage.

The contributions of this study are as follows. First, it contributes to research on decision-support systems (DSS) by proposing a novel bankruptcy prediction architecture for SMEs and providing an innovative DSS for banks in loan granting that incorporates payment and transactional data into feature extraction. Both offline and online tests provided by banks confirmed the practical value of the model. Second, we show that transactional data can significantly improve bankruptcy prediction when there is a scarcity of financial-statement data. This has theoretical value for bankruptcy-prediction research; it also has practical value for banks by providing a credit-risk evaluation that considers the latest SME information and improves the precision of financial decision-making. Finally, we propose a two-stage multiobjective feature-selection method for bankruptcy prediction, which contributes to finding key bankruptcy predictors for research and creating rule-based credit-risk models to support financial decision-making for banks. Further, the empirical results confirm the effectiveness of the proposed method.

The rest of this paper is organized as follows. The next section provides a brief review of the literature. Section 3 describes the proposed model architecture and the details of our proposed feature-selection method. Model preprocessing is presented in Section 4, including the raw data, the transformation from relational data into the SME network, and the model variables. Sections 5 and 6 provide the experimental setup and the results, respectively. Section 7 discusses the importance of payment and transactional data in bankruptcy prediction as well as the contributions and limitations of the study. Finally, Section 8 concludes the paper.

2. Literature review

2.1. Bankruptcy prediction for SMEs

Research on corporate bankruptcy prediction can be traced to the 1960s. Altman et al. [1] used a set of financial ratios extracted from financial statements to gauge bankruptcy risk. Since then, accounting-based financial ratios and other nonfinancial information have been incorporated to improve the accuracy of bankruptcy prediction and default prediction [2,3,8].

Given the high default risk of SMEs compared with large firms, a credit model specific to SMEs is preferred for credit-granting [22,27]. Researchers and financial institutions use not only financial (accounting)-based variables but also non-financial-based variables in credit-risk assessment for SMEs [4,5]. These nonfinancial data are mainly divided into three groups: firm-based nonfinancial data (e.g., business sector, age, company patents, corporate governance indicators), soft information (e.g., management quality, preferences of domain experts), and external information (e.g., market information). Table 1 shows the different types of variables used in bankruptcy prediction and credit-risk evaluation for SMEs.

With the development of data-storage technology, banks have accumulated a large amount of payment and transactional data. These data are collected constantly and are more reliable and flexible than financial statements. Effectively incorporating such data into a credit-risk model has theoretical and practical importance. In this study, therefore, we used transactional data to predict bankruptcies and investigate the predictive power of payment and transactional data–based variables.

Daily interactions between firms—including payments, business partnerships, and mutual ownership relations—are rarely considered when assessing the credit risk of SMEs [16], even though such relationships, especially the characteristics of payments issued, are important for credit-risk evaluation [14,15]. Researchers do, however, consider the networks created by common shareholders and directors when determining SME credit risk [17,28]. Letizia et al. [16] constructed predictive variables based on payment networks to calculate enterprises' credit ratings. Using data provided by banks, we extend this research by considering the correlation between SMEs' transaction interactions and bankruptcy risk. Relational data for SMEs, including manager, shareholder, and payment networks, are incorporated into bankruptcy prediction.

2.2. Feature selection

Feature selection is a dimension-reduction method that reduces dataset dimensionality and maintains model performance by removing irrelevant and redundant variables [23,24]. Effective feature-selection methods reduce the cost of feature acquisition in the classification model and are conducive to model interpretability [20,34]. In a credit-risk model, comprehensibility and profitability are two contradictory objectives of concern for banks; feature selection is considered a multiobjective problem [20]. We extend the framework proposed in [20] by introducing a two-stage multiobjective feature-selection method for credit scoring. In reality, some features extracted from new types of datasets are noisy and redundant; it is reasonable to remove those

Algorithm 1
First stage feature selection algorithm.

Input: Train dataset D_{train}; Feature rank list R; Classifier P; Allowed optimal region threshold ε;
Output: optimal feature subset S^*;
$t \leftarrow 0$, $I_t \leftarrow [a_t, b_t]$, $X \leftarrow \inf\{I_t\}$;
while $|I_t| > \varepsilon$ **do**
  $f_1 \leftarrow m(S_{x(t+1)})$, $x(t+1) \leftarrow a_t + \lfloor 0.382 \cdot (b_t - a_t) \rfloor$
  $f_2 \leftarrow m(S_{y(t+1)})$, $y(t+1) \leftarrow a_t + \lceil 0.618 \cdot (b_t - a_t) \rceil$
  if $f_1 \geq f_2$ **then**
  $|I_{t+1}| = [a_t, y_{t+1}]$
else
$|I_{t+1}| = [x_{t+1}, b_t]$
end if
$t \leftarrow t + 1$
  $X \leftarrow X \cup \{\inf\{I_t\}\}$
end while
Update I_{T+1} according to Eq. (1). // t^* in Eq. (1) equals the last second value in X.
Find the optimal feature subset S^* in final candidate feature subset $\{S_j | j \in I_{T+1}\}$.

[1] Data were provided by Shandong City Commercial Bank Alliance Co., Ltd. To comply with privacy regulations, all records were anonymized, and figures such as transaction amounts were normalized.

features in the first stage. First, those features may be selected by wrapper-based feature selection if the inclusion of those features in the model does not affect classification performance. Second, removing those features reduces the search space of the wrapper-based method. Thus, in the first stage, we remove those features using feature-relevance and optimum-seeking methods. In the second stage, a multiobjective wrapper-based feature selection that optimizes both the feature set size and model performance is used to find an optimal solution [20,31].

3. Model architecture and two-stage multiobjective feature selection for bankruptcy prediction

3.1. Conceptual framework

We aimed to predict bankruptcy likelihood for SMEs under a scenario where no financial statements are available. Fig. 1 presents the general framework, including the datasets, networks, feature extraction and selection, model-performance comparison, and evaluation of feature importance. First, we constructed three SME networks and extracted firm-, network-, and transaction-based variables. Then, we evaluated the predictive power of payment and transactional data–based variables, including transaction-based variables and payment network–based variables, for bankruptcy prediction by comparing the classification performance of five different models. Finally, a two-stage multiobjective feature-selection method is proposed to select the optimal feature subset, and we evaluate the importance of each feature in the optimal feature subset.

3.2. Proposed two-stage multiobjective feature selection

We treat feature selection as a two-stage multiobjective optimization problem. In the first stage, features are ranked in descending order by feature relevance, and the top k features are selected. The optimal k value is found using the optimum-seeking method. In the second stage, multiobjective wrapper-based feature-selection methods [20,31] are used to find the compact subset [30] from the feature subset selected in the first stage. Compared to a feature selection consisting only of the second stage, the first stage reduces the feasible region of the second stage and the risk of finding a suboptimal solution.

3.2.1. First-stage feature selection

In the first stage, we measure the feature relevance between features and labels by integrating four commonly used methods: (1) chi-squared, (2) ReliefF, (3) information gain, and (4) gain ratio [37]. All features are

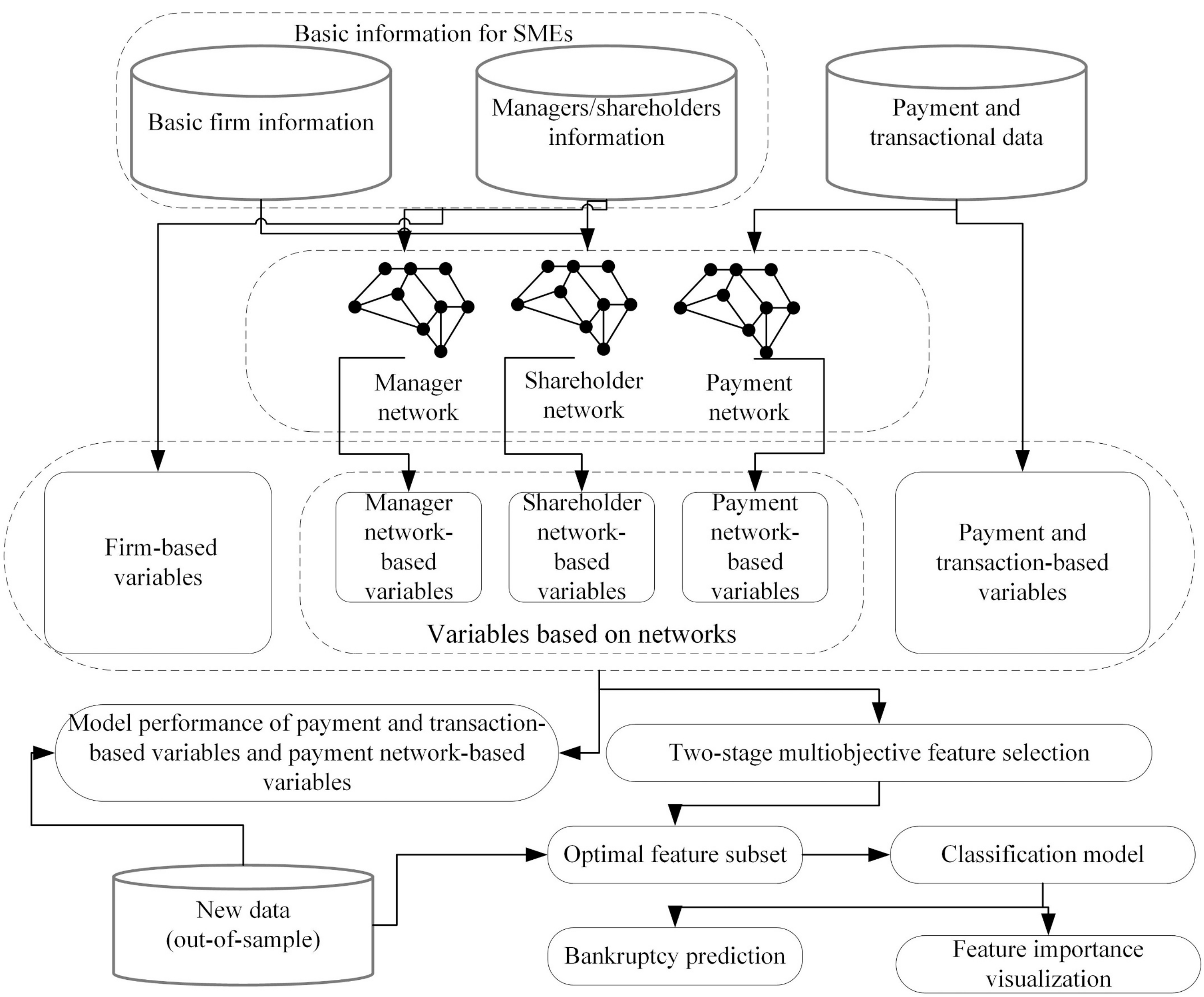

Fig. 1. Proposed model architecture for bankruptcy prediction of SMEs.

then ranked based on feature relevance. First, four feature-relevance indexes are calculated for each feature and are normalized to internals [0,1]. Then, the final relevance index for each feature is calculated by averaging the normalized feature-relevance indexes [37]. Finally, a ranked feature list $R = [R_1R_2, \ldots, R_N]$ is obtained by ranking all features in descending order according to the final relevance index.

Based on the ranked feature list, a predictor is applied for each feature subset$S_i = [R_1R_2, \ldots, R_i]$, $i = 1, 2, \ldots, N$ and the subset with the highest accuracy is selected as the optimal feature subset [38]. However, this method is not feasible when dealing with a highly dimensional dataset. In this study, we used the optimum-seeking method to find the optimal feature subset in the first stage. Below, we present some definitions and assumptions.

Definition 1. Evaluation value $m(S_i)$. The evaluation value $m(S_i)$ for subset S_i is a classification performance measure of a predictor in a training dataset using feature subset S_i only.

Definition 2. An evaluation curve f is a curve that illustrates variation in the evaluation value $m(S_i)$ according to feature subset sizes $|S_i|$.

We assume an evaluation curve has four types of shapes, as shown in Fig. 2. This assumption is reasonable since all features are ranked in descending order according to the feature-relevance index. The addition of feature R_i to the previous subset S_{i-1} results a nondecrease in the evaluation value when i is smaller than a certain cutoff point x. Meanwhile, for $i > x$, the addition of irrelevant features does not improve model performance and can even deteriorate model-classification ability.

Under the above assumptions of the evaluation curves, the optimum-seeking method is guaranteed to find the global optimal solution. However, for type b and type c evaluation curves, all points $x \in [x_1, x_2]$ are optimal solutions. In this case, optimal solution x_1 is preferred. To find the optimal solution with the smallest feature subset size, a modified 0.618 method is proposed, as in Algorithm 1.

Algorithm 1 has two parts. In the first part, the optimum-seeking method is applied to find the optimal value $m^* := \max_i m(S_i)$ and reduce the feasible region. After the iteration of the first part, a reduced feasible region $I_T = [a(T), b(T)]$ is produced, and the optimal value satisfies $m^* \geq \max\{m(S_{m(a(T))}), m(S_{m(b(T))})\}$.

In the second step, we evaluate whether the algorithm has found the optimal solution with the minimum number of features i^*. If $m(S_{m(a(T))}) < m(S_{m(b(T))})$, then $i^* \in [a(T), b(T)]$, otherwise, $i^* \in [\inf\{I_t^*\}, b(T)]$, where $t^* = \underset{j \in \{i \mid \inf\{I_i\} \neq a(T)\}}{\arg\max} \left(a(T) - \inf\{I_j\}\right)$ is the last-second updated value for the left end point of I_t:

$$i^* \in I_{T+1} := \begin{cases} [a(T), b(T)], m(S_{m(a(T))}) < m(S_{m(b(T))}) \\ [\inf\{I_{t^*}\}, b(T)], m(S_{m(a(T))}) \geq m(S_{m(b(T))}) \end{cases} \quad (1)$$

Then a sequential backward search is applied in the final feasible region I_{T+1} to find the optimal feature subset.

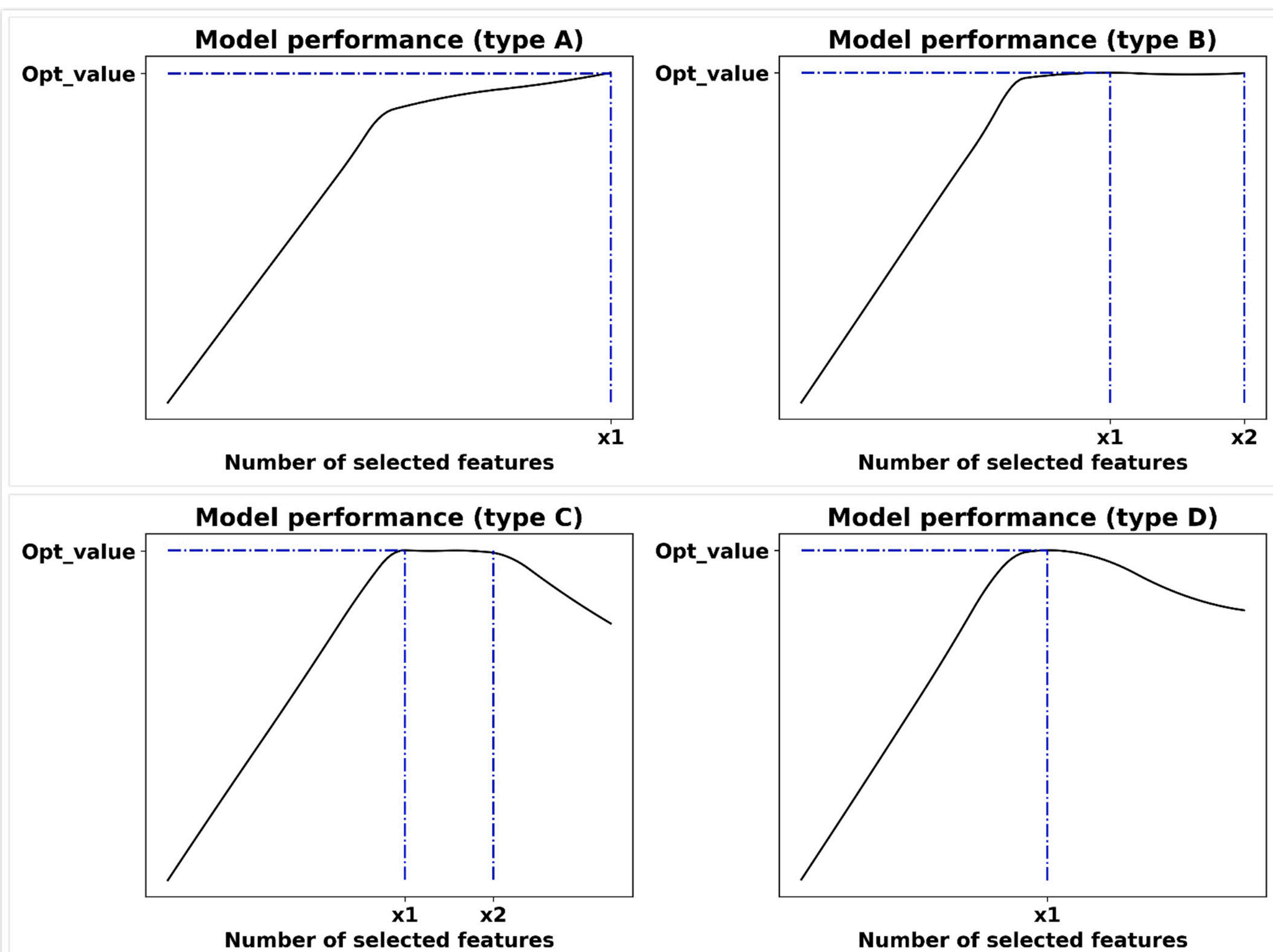

Fig. 2. Four types of evaluation curves. The horizontal axis is the cardinality of the feature subset while the vertical axis represents classification performance. x_1 is the optimal number of features while x_2 is the feature subset size of one optimal solution to problem $\max_i m(S_i)$.

For a given ε, the convergence of the first part of Algorithm 1 is guaranteed when $T \geq \frac{\ln\varepsilon - \ln N}{\ln 0.618} + 1$. For the second part, the minimum feature subset can be found within $\lfloor \varepsilon \rfloor - 1$ steps when $I_{T+1} = [a(T), b(T)]$ and $b(T) - \inf\{I_t^*\} - 1$ when $I_{T+1} = [\inf\{I_t^*\}, b(T)]$. Even in the worst case, the optimal solution is guaranteed to be found within $\lfloor 0.382 \cdot N \rfloor + \lfloor \varepsilon \rfloor - 1$ iterations.

3.2.2. Second-stage feature selection

Here, we aim to find a compact subset that optimizes model comprehensibility and performance from the optimal subset of the first stage. In terms of model performance, our case considers AUC and model profit. Another objective is to minimize feature subset sizes. Similar to Kozodoi et al. [20], NSGA-II is applied in the second-stage feature selection because it has the advantages of low time complicity in elite strategy sorting to improve search efficiency and fewer hyper-parameters [25].

4. Data, networks, and features

4.1. Datasets

This study used two datasets to predict bankruptcy. One dataset contains basic information for 3,500,000+ SMEs, including enterprise type, industry, operational status (bankrupt or active), shareholders (senior managers), and so on (see Appendix A for the field description).

The other dataset contains transaction and payment information for more than 170,000 SMEs. These records are registered in a timely manner, including more than 240 million transactions for the period January 1, 2016, to June 30, 2018. Information such as daily transactions, wages, tax payments, and other records related to daily operations are included in this dataset.

During the data preprocessing, we removed samples with no payment and transactional data for two reasons. First, it is reasonable for a bank to require bank reconciliation statements before applying for a loan if the bank does not have payment information for a firm. Second, features such as transaction-based variables are unavailable for firms without transactional data. Filling in those missing values is impossible and will result in poor model performance.

4.2. Networks

In this study, the networks of the SMEs are established through relationships among common managers/shareholders and transaction interactions between two SMEs. Three types of SME networks were established based on those interactions: manager network, shareholder network, and payment network. The constructed networks were then used for downstream feature generation and prediction.

4.2.1. Manager and shareholder networks of SMEs

The manager and shareholder networks depict the relationships of common managers/shareholders between two SMEs. In a manager network of SMEs, a node represents one SME; two nodes are connected by an edge if they have common managers, and the edge weight is the number of common managers. Similarly, the shareholder network of SMEs is created based on the interactions of common shareholders, where each edge is assigned a weight according to the number of common shareholders. Manager and shareholder networks are undirected weighted networks.

4.2.2. SME payment networks

A network is an SME payment network if nodes in the network represent SMEs, and edges represent payment interactions between two SMEs. Unlike manager and shareholder relationships, payment interactions have specific characteristics: (1) transactional data are registered with daily granularity, (2) two firms may have at least one transaction in a period of time, and (3) payment is directed—that is, a payment (or transfer) from firm A to firm B is different from a payment (or transfer) from firm B to firm A. We created a payment network based on the above characteristics and assigned a link between two firms (firm A and firm B) if there was at least one payment from firm A to B in the last 180 days; the edge weight is the summation of all payments from firm A to B in the last 180 days. Different from manager and shareholder networks, a payment network is a directed weighted network. Appendix A presents the summary statistics for SME networks.

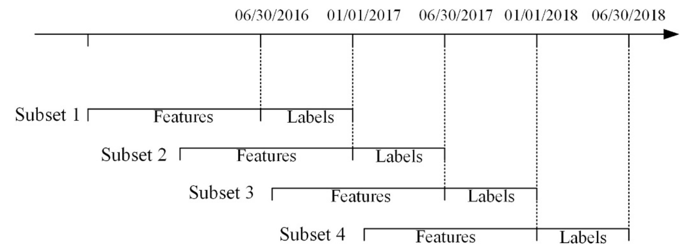

Fig. 3. Information about the four subsets of data, including the forecast horizon, time period for generating features, and labels for each dataset. For the feature-generation period, the start time is not provided since different features were generated in time periods with different start times. For example, features reflecting transactional information of the last 30 days were generated in a time period within 30 days while features related to transactional information of the last 180 days were constructed in the time period with a start time 180 days before the start time of the label-generation period.

4.3. Features

We used three subsets of features: basic firm information variables, SME network– based variables and transaction-based variables. First, we extracted basic firm information variables, including business sector and business type, as well as the historical bankruptcy rates of different sectors and geographic locations [5]. Second, the network-based variables included node centrality measure, community detection[2] [16], and risk propagation [17] for the three types of SME networks. Finally, transaction-based variables included recent changes in a firm's daily business activities (see Appendix A for specific feature descriptions and summary statistics of selected features).

5. Experimental setup

5.1. Data collection for training and testing

Fig. 3 depicts the data-collection process. A six-month forecast horizon was considered in this study [3]; that is, data were collected to predict bankruptcy events for the next six months. Four subsets of data were collected, referred to as Subset 1, Subset 2, Subset 3, and Subset 4; labels were generated from sequential nonoverlapping time intervals. Then, those subsets were amalgamated into three datasets, referred to as Dataset 1, Dataset 2, and Dataset 3, for rolling evaluation. For example, Subset 1 and Subset 2 were considered the training set and test set of Dataset 1, respectively. The test set of Dataset 1 was used as the training set for Dataset 2 and so on.

Further, we split Datasets 1–3 into offline and online tests. We tested the model offline prior to July 1, 2017; after that date, the model was tested online (i.e., Datasets 2–3 were tested in the online test environment). Table 2 reports the detailed information for each dataset.

[2] We use the fast unfolding algorithm (Louvain method) [35] to detect modular structure.

Table 1
Analysis of datasets (information) used in bankruptcy prediction and credit-risk evaluation for SMEs.

Bankruptcy prediction and credit-risk study (in chronological order)	Financial statements	Other datasets	Task
Edmister (1972) [21]	Financial ratios		Bankruptcy prediction
Grunert et al.(2005) [9]	Financial ratios	Management quality and market position	Default prediction
Altman et al.(2007) [4]	Financial ratios		Default prediction
Altman et al.(2010) [6]	Financial ratios	Regulatory compliance and event data	Bankruptcy prediction
Chen et al.(2010) [12]	Financial statements	Closing stock price	Credit-risk measurement
Altman et al.(2013) [19]	Financial ratios	Credit relationship information	Default prediction
Ciampi et al.(2013) [5]	Financial ratios	Firm size, geographical area, and business category	Bankruptcy prediction
Ciampi (2015) [10]	Financial ratios	Corporate governance variables.	Default prediction
Angilella et al.(2015) [11]	Financial ratios	Soft information including development risk, technological risk, market risk and production risk	Credit risk evaluation
Fernandes et al. (2016) [7]	Financial ratios	Geographical location	Default prediction
Tobback et al.(2017) [17]	Financial ratios	Network of SMEs	Bankruptcy prediction
Figini et al. (2017) [29]	Financial ratios	Local outlier factor	Default prediction
Yuan et al.(2019) [28]		Shareholder structure and board-of-directors structure.	Credit-risk evaluation
Ptak-Chmielewska (2019) [13]	Financial ratios	Sector of the company's activity; company's legal form, region, age; employment	Bankruptcy prediction

5.2. Performance measure

In this study, unless otherwise specified, we used AUC to compare model classification performance. We also evaluated the economic profit of each model based on NPV, as defined in [18]. However, NPV considers not only the net profit value of the model but also the profit when no classification model is used, which has no correlation with model performance, as described in [18]. Since we are interested in a model's economic performance, it is preferable to consider the expected (average) profit of the classification model, defined as

$$EP_{model}(s) := -c \cdot FPR(s) + b \cdot TPR(s) \tag{2}$$

where $c = (1-p)\cdot(C_{FP}+B_{TN})$, and $b = p\cdot(C_{FN}+B_{TP})$ are the cost [50] (benefit) related to the true positive rate tpr and false positive rate fpr for a specific cutoff s.

For a specific model, we maximize the expected profit by optimizing the threshold s:

$$EMP := \max_s EP_{model}(s) \tag{3}$$

The following inequality holds for EMP, $0 \leq EMP \leq b$, and $EMP = b$ for a perfect model. The supremum of the economic profit of a classification model is relatively small when the dataset is highly imbalanced. In this case, EMP is a flawed measure since the difference in EMP between a powerful model and a weaker one is limited. To unify dimensions, we propose the following profit-based classification measure.

Definition 3. The increment ratio of expected maximum profit (IEMP) of a classification model measures the economic benefit increments of a model compared to those of a perfect model, or, equivalently, $IEMP_i = 100 \times \frac{EMP_i}{b}$, where EMP_i is the expected maximum profit of model i.

Table 2
Data characteristics for each dataset. N_{train} is the number of samples in the training dataset, P_{train} and p_{train} are the number of positive samples and bankruptcy ratios in the training dataset. Similarly, N_{test}, P_{test} and p_{test} are the values in the test dataset. The last column indicates the test environment used for each dataset.

Dataset ID	N_{train}	P_{train}	p_{train}	N_{test}	P_{test}	p_{test}	Environment
Dataset 1	169,759	1234	0.0073	168,466	1072	0.0064	Offline
Dataset 2	168,466	1072	0.0064	167,364	797	0.0048	Online
Dataset 3	167,364	797	0.0048	166,527	431	0.0026	Online

To determine IEMP, we assume that the costs and benefits are supposed to be known and consider a deterministic approach to calculate IEMP.[3] Table 3 presents our assumptions about figures related to calculating costs and benefits. The cost and benefit parameters are calculated in the same manner as [18]. Then, the IEMP of each model is calculated according to Definition 3.

Further, hypervolume and the spacing metric were used evaluate the performance of the multiobjective feature-selection method. Since we do not know the optimal population for the problem, other metrics, such as spread and maximum spread, were not considered. The spacing metric evaluates the diversity of nondominated solutions [25]. A smaller spacing value means the solutions are more uniformly distributed in the obtained front [25]. The hypervolume measure, proposed by Zitzler et al. (1998) [43], is defined as the volume of the objective space dominated by the final solutions in the front [43]. It considers diversity, pays more attention to the convergence of solutions, and is a commonly used measure in comparative studies [31,42]. The diversity of the results provides a variety of flexible choices to the decision-maker, and the optimality of solutions indicates better model performance compared to other solutions with lower hypervolume.

5.3. Model evaluation

First, to evaluate the predictive power of payment and transactional data in bankruptcy prediction, we combined all variables into five different models, which are distinctive based on the feature sets they use as input. Table 4 shows the specific features used in the five models. Model 1 is the baseline model, which only uses basic firm information features. Model 2 combines the features used in Model 1 and the manager/shareholder networks of SME-based variables. Models 3 and 4 add, respectively, transaction-based features and payment network–based variables to Model 2. Finally, Model 5 is a hybrid of all abovementioned features. Seven commonly used classifiers (i.e., linear discriminant analysis (LDA), logical regression (LR), support vector machine (SVM), decision tree (DT), random forest (RF), XGBoost (XGB), and neural network (NN)) [52] were used on these five models. The meta-parameters of these classifiers were tuned through grid search; Appendix B lists the detailed meta-parameter information.

Second, we tested the model performance of our proposed feature-selection method. The four subsets of data in Fig. 3 were amalgamated into one dataset.[4] Twofold crossvalidation was used to evaluate model performance, and the process was repeated five times to get the final results.

In the first-stage feature selection, the optimal region threshold ε was set to 1. For the second stage, we tuned the meta-parameter of NSGA-II, including the number of generations and population size, on a subset of

[3] All cost (profit) figures were determined after discussion with the product manager of the data provider.

[4] We also tested the model performance on each dataset, the results were similar to that on the amalgamated dataset.

Table 3
Assumptions regarding costs (benefits) related to false positive rate and true positive rate.

	Variable	Baseline value
1	Interest spread (per annum)	1.25%
2	Underwriting fees (up front)	0.50%
3	Workout fees (on default)	2.0%
4	Loss given default (on default)	75.0%
5	Risk free rate	4.0%

Table 4
Description of Models 1–5 and the feature subset used in each model. The last row is the number of features for each model.

Features	Model 1	Model 2	Model 3	Model 4	Model 5
Basic firm information–based features	√	√	√	√	√
Manager/shareholder network–based features	×	√	√	√	√
Transaction-based features	×	×	√	×	√
Payment network–based features	×	×	×	√	√
Total number of features	29	43	85	56	98

training data. Three pairs of metaparameters (generation number and population size were equal to (50, 25), (100, 50), and (200,100), respectively) were compared using twofold cross-validation with five independent runs and the results are listed in panel A of Table E.1. Based on statistical tests of hypervolume and the spacing metric of each model, the number of generations and population size were set to 100 and 50, respectively. For all stages, XGBoost was used as the base classifier. Fourfold cross-validation was used to calculate the evaluation value for the first stage and fitness values in the second stage. To elaborate the importance of first-stage feature selection, we compared our proposed algorithms—that is, two-stage $MOFS_{AUC}$ ($TSMOFS_{AUC}$) and two-stage $MOFS_{IEMP}$ ($TSMOFS_{IEMP}$)— with $MOFS_{AUC}$ and $MOFS_{IEMP}$ as shown in Table 5. Further, two more two-stage feature-selection methods—referred to as $TSGA_{AUC}$ (or $TSGA_{IEMP}$) and $TSPSO_{AUC}$ (or $TSPSO_{IEMP}$) in Table 5—were included for comparison [40,41]. For fair comparison, the same objectives (AUC or IEMP) were used in the second stage, and the setup of meta-parameters was identical to [40,41] respectively. Finally, two traditional wrapperbased methods, sequential forward selection (SFS) and sequential backward selection (SBS) [24], were included in our experiments for comparison.

6. Results and analysis

6.1. Model performance of payment and transactional data–based variables

We first conducted comparative experiments during offline tests. To evaluate the robustness of the model in different prediction periods, we also tested model performance on online test datasets and used rolling evaluation to evaluate model performance. Appendix C provides the detailed results of the online tests.

To investigate whether data-imbalance issues jeopardized model performance, we tested whether undersampling or oversampling improved the predictive capacity. Three commonly used methods—random undersampling, random oversampling, and SMOTE—were compared with the original imbalanced dataset. The results presented in Appendix D indicate that the model trained on the original imbalanced dataset outperformed the results of the sampling methods. We also investigated the sensitivity of the imbalanced ratios after sampling for each sampling technique; the results did not indicate that sampling methods improved predictive power. This is reasonable, since our datasets are relatively imbalanced datasets; that is, the bankruptcy samples are not rare in an absolute sense but are rare relative to non-bankruptcy objects (see Table 2) [36,48]. The predictive results were thus unaffected by data-imbalance issues [36].

In the following, we report the results from the original imbalanced datasets. Panel A of Table 6 shows the AUC values of the five models for different classifiers. The out-of-sample performance of Models 3–5 was better than the performance of Models 1 and 2 in almost all classifiers in terms of AUC values. This indicates that payment information plays an important role in improving model performance for bankruptcy prediction. We conclude that using SME payment information in bankruptcy prediction improves the model's predictive power. In terms of classifier performance, similar to many empirical results in bankruptcy prediction problems [45,46], ensemble models such as XGB outperform individual classifiers. DT was found to perform the worst among the seven classifiers, which is similar to practical investigations of credit-risk problems [47].

Panel B in Table 6 shows the IEMP results of different models. First, the model with payment and transactional data–based variables outperforms Models 1 and 2 in terms of economic benefit for most classifiers. This confirms the economic benefit of using payment and transactional data–based variables. Furthermore, ensemble models outperform single classifiers in profit-based classification measures. Some classifiers, such as DT, have an IEMP equal to 0 in most of the five models. Although the discriminative power of those classifiers is better than random guesses, the optimal threshold of those classifiers is $s^* = 0$ (equivalently, $tpr = fpr = 0$) due to bad classification performance.

The result confirms that incorporating payment and transactional data–based features into a prediction model improves model performance in terms of AUC and profit. This is reasonable since transactional data reflect the latest operational conditions of SMEs; thus, incorporating such information improves the prediction ability of the classification model. Furthermore, the recency, frequency, and monetary value of transactional data are positively correlated with firm growth in the economic sense [39]. For example, a potentially nonbankrupt enterprise generates higher cash flows compared to enterprises going bankrupt [39]. Finally, it is reasonable that incorporating payment networks also increases model performance. Payment network–based features, such as degree, reflect the amount of customers and trading volumes of a firm. Meanwhile, the bankruptcy rate of a detected community, or neighboring nodes in a payment network, affect the operation of SMEs since the bankruptcy of these firms decreases the business size of SMEs.

Incorporating SME payment information in bankruptcy prediction improves model performance but increases the cardinality of the feature set, making the model more opaque. It also increases the cost related to feature extraction. Therefore, we propose a two-stage feature-selection method that not only reduces the cardinality of feature subset but also directly optimizes model performance, thus increasing the economic benefit brought by the model.

Table 5
Information about two-stage feature-selection methods.

Model abbreviation	First-stage feature selection	Optimization methods (second stage)	Objectives (second stage)
$TSMOFS_{AUC}$	√	NSGA II	AUC, no. of features
$TSMOFS_{IEMP}$	√	NSGA II	IEMP, no. of features
$MOFS_{AUC}$	×	NSGA II	AUC, no. of features
$MOFS_{IEMP}$	×	NSGA II	IEMP, no. of features
$TSGA_{AUC}$ [41]	CHI2	Genetic algorithms (GA)	AUC
$TSGA_{IEMP}$ [41]	CHI2	GA	IEMP
$TSPSO_{AUC}$ [40]	Information gain	Particle swarm optimization (PSO)	AUC
$TSPSO_{IEMP}$ [40]	Information gain	PSO	IEMP

Table 6
Out-of-sample prediction performance of Dataset 1(AUC and IEMP).

Model	LR	LDA	DT	SVM	RF	XGB	NN
Panel A: Model performance in terms of AUC							
Model 1	0.623	0.625	0.592	0.624	0.637	0.653	0.627
Model 2	0.622	0.626	0.596	0.623	0.63	0.655	0.623
Model 3	0.700	0.675	**0.684**	0.678	0.725	0.767	0.696
Model 4	0.684	0.675	0.653	0.677	0.693	0.721	0.686
Model 5	**0.718**	**0.726**	0.683	**0.717**	**0.744**	**0.785**	**0.723**
Panel B: Model performance in terms of IEMP							
Model 1	0.44	0.23	0.00	0.36	0.79	1.36	0.06
Model 2	0.36	0.25	0.00	0.03	1.03	1.39	0.21
Model 3	2.47	3.49	**3.61**	**3.66**	6.39	9.92	2.05
Model 4	0.36	0.30	0.00	0.09	0.35	1.84	0.18
Model 5	**3.93**	**4.60**	0.00	3.10	**8.02**	**11.58**	**3.27**

Table 7
Hypervolume indicators and spacing metrics in terms of AUC and IEMP for each multiobjective feature-selection method.

Method	Hypervolume (AUC)	Hypervolume (IEMP)	Spacing (AUC)	Spacing (IEMP)
Panel A: $TSMOFS_{AUC}$ vs $MOFS_{AUC}$				
$TSMOFS_{AUC}$	**30.65** ± **0.81**	**12.06** ± **1.01**	0.11 ± 0.04*	**0.12** ± **0.03**
$MOFS_{AUC}$	27.43 ± 0.56	10.93 ± 0.74	**0.10** ± **0.05**	0.12 ± 0.06*
Panel B: $TSMOFS_{IEMP}$ vs $MOFS_{IEMP}$				
$TSMOFS_{IEMP}$	**29.79** ± **0.80**	**11.16** ± **0.58**	**0.11** ± **0.05**	0.12 ± 0.07*
$MOFS_{IEMP}$	26.72 ± 0.45	10.81 ± 0.86*	0.15 ± 0.08*	**0.10** ± **0.08**

Note: The mean values ± standard deviation are reported, and the best results are in bold.
* Indicates significant similarity to the best results at $p < 0.05$.

6.2. *Model performance of two-stage multiobjective feature selection*

First, we analyzed the model performance of four multiobjective feature-selection methods. Table 7 reports the hypervolume indicator and spacing metric results in terms of AUC and IEMP. As seen in panel A of Table 7, $TSMOFS_{AUC}$ outperforms $MOFS_{AUC}$ in terms of the hypervolume indicator. In terms of the spacing metric, $TSMOFS_{AUC}$ achieves similar or better results compared with $MOFS_{AUC}$. Further, similar results are observed in panel B of Table 7 when model profit and the number of features are used as objectives.

To test whether increasing generation number and population size improves the performance of single-stage multiobjective feature-selection methods, we compared $TSMOFS_{AUC}$ to $MOFS_{AUC}$ in Subset 1, as depicted in Fig. 3, when generation number and population size were set to 200 and 100, respectively. The results are presented in panel B of Table E.1. The results suggest that two-stage methods generally have similar or significantly better performance compared to single-stage methods.

The experiment results confirm that two-stage multiobjective feature-selection methods generally have better performance than $MOFS_{AUC}$ and $MOFS_{IEMP}$. This result is reasonable since we removed irrelevant features in the first stage, which is helpful for finding a more compact feature subset in the second stage. Finally, in terms of IEMP-based hypervolume and spacing metrics, $TSMOFS_{IEMP}$ does not perform significantly better than $TSMOFS_{AUC}$ and $MOFS_{AUC}$. This is because AUC-based feature-selection methods directly optimize the area under the curve; one model is guaranteed to have a better performance for profit than the other as long as the ROC curve of the model dominates the ROC of the others. Although a model with a larger AUC is not a sufficient condition for a model with better profit, empirically, an AUC-optimized model often performs better than a profit-optimized one.

Furthermore, we compared the proposed two-stage multiobjective methods and other feature-selection benchmarks. Similar to Hancer et al. (2018) [38], we collected the 10 set of solutions into a union set for each method. The best results among solutions with the same number of features were selected to form a new set of solutions and were compared. Fig. 4 shows the results when AUC was used as a fitness value for each single-objective method. The horizontal axis in each chart represents the number of selected features, and the vertical axis represents AUC (left-hand-side charts) or IEMP (right-hand-side charts). Fig. 4 shows that $TSMOFS_{AUC}$ generally includes a smaller number of features while achieving better classification performance in terms of AUC and profit compared to other methods. Specifically, solutions identified by $TSMOFS_{AUC}$ use a fewer number of features and provide a trade-off between model comprehensibility and profitability [20] compared to the solutions of $MOFS_{AUC}$. Further, most of the solutions of other benchmarks are dominated by solutions on the nondominated front of $TSMOFS_{AUC}$.

Finally, we conducted similar experiments when the IEMP metric was used as an objective for single-objective feature-selection methods and compared the performance of $TSMOFS_{IEMP}$ with other benchmarks. The results were similar when AUC was used as the model performance measure for optimization (Fig. 5). In summary, compared to other single-stage multiobjective methods and single-objective wrapper-based methods, two-stage multiobjective methods provide multiple choices for decision-makers and generate solutions with a fewer number of features while achieving higher or comparable model performance.

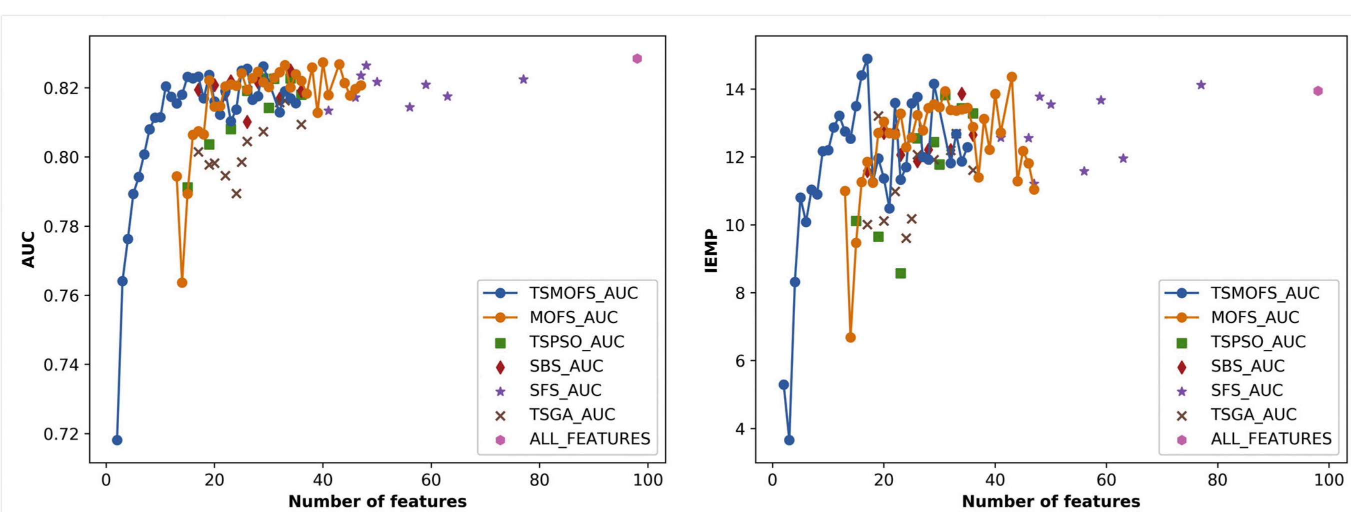

Fig. 4. Evaluation results of feature-selection methods with AUC as the optimization objective.

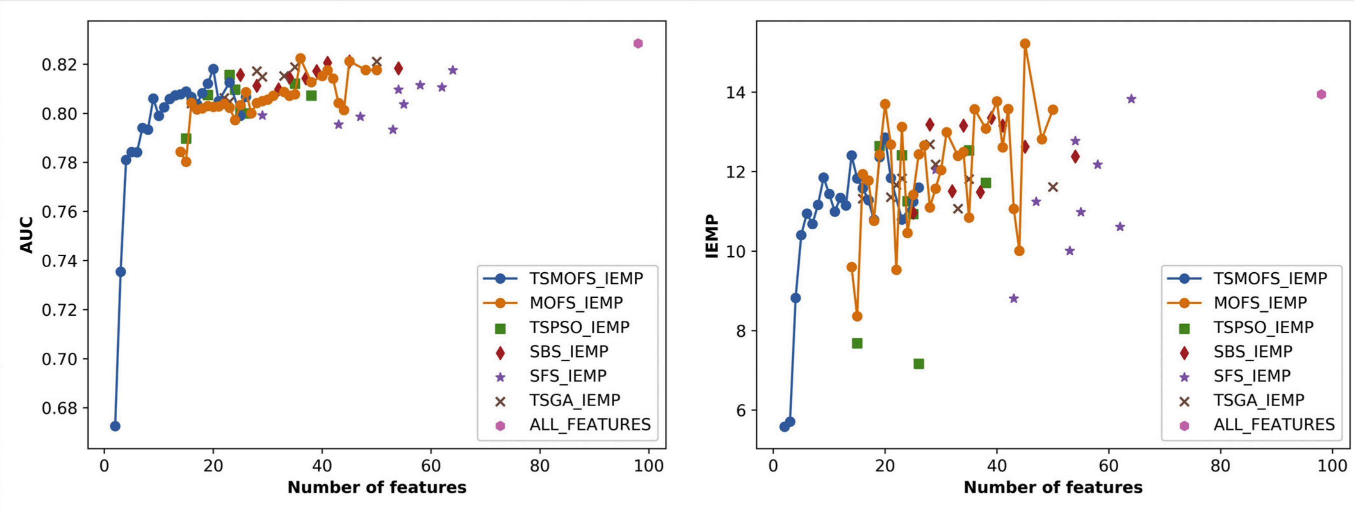

Fig. 5. Evaluation results of feature-selection methods with IEMP as the optimization objective.

7. Discussion

We created a bankruptcy prediction model based on payment and transactional data provided by banks. Both offline and online tests indicated that features extracted from transactional data improve the predictive power of bankruptcy prediction for SMEs. Further, a compact feature subset was derived from the proposed two-stage multiobjective feature-selection method.

7.1. Theoretical implications

Before summarizing the theoretical contributions, let us take a closer look at the importance of features in a compact feature subset. First, the importance of a feature is computed as the (normalized) total reduction of node impurity brought by that feature as measured by the Gini index; the higher it is, the more important the feature. Fig. 6 (left-hand-side chart) depicts the feature importance of the 15 most important features. Payment and transactional data–based variables were found to reach high feature importance, confirming that payment and transactional data are important for bankruptcy prediction and credit-risk evaluation.

The economic value brought by a variable has more practical value for a credit-risk model. We proposed a profit-based feature-importance measure based on permutation importance [32] to measure the economic benefit of a specific variable. Instead of evaluating accuracy-based permutation importance, we evaluated profit-based permutation importance. Fig. 6 (right-hand-side chart) shows the profit-based feature importance. Again, we conclude that payment and transactional data–based variables play an important role in economic value increments.

This study contributes to the bankruptcy-prediction literature by showing the additional predictive power of payment and transactional data–based variables where there is a scarcity of financial-statement data. This is of theoretical importance for finding the key predictors of

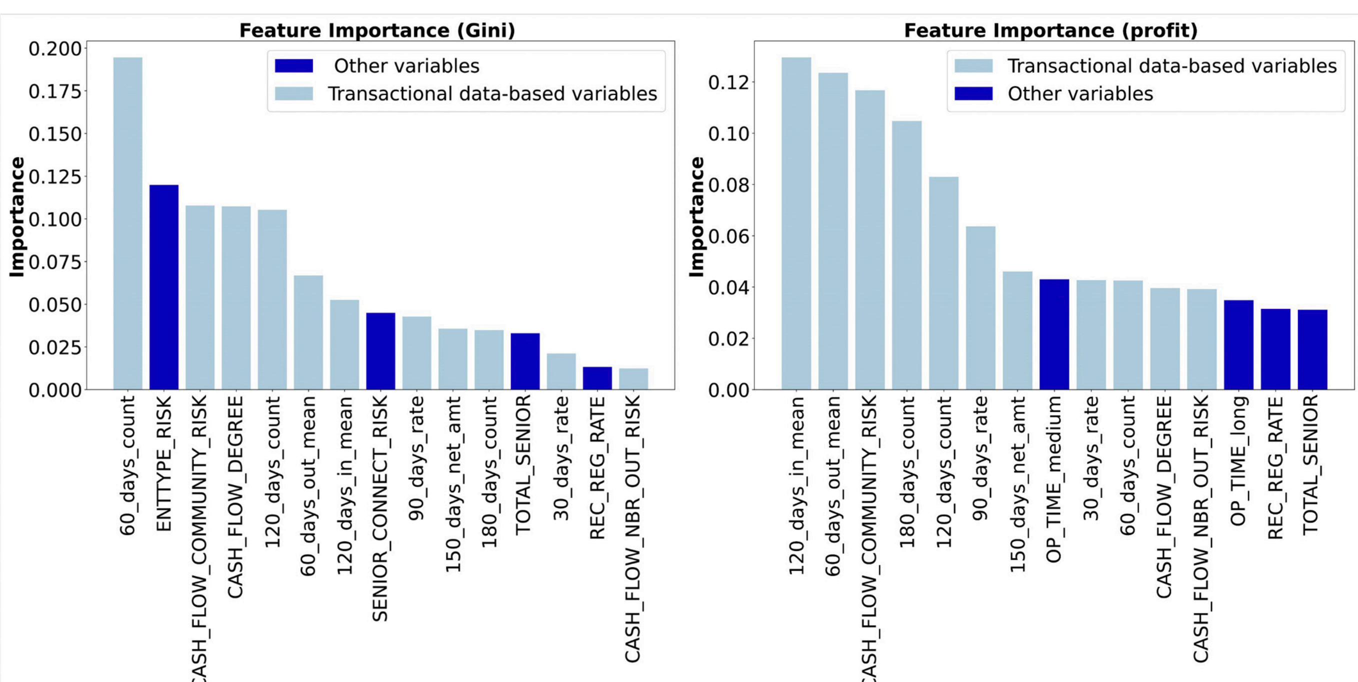

Fig. 6. Feature importance for features in an optimal feature subset.

SME bankruptcy. We also evaluated the feature importance of selected features in terms of the Gini index and profit; both feature importance measures confirm the importance of transactional data–based variables for bankruptcy forecasting. Finally, we proposed a two-stage multi-objective feature-selection method, which has theoretical implications for finding key predictors in credit-risk research.

7.2. Practical implications

This study proposed a bankruptcy prediction architecture under a scenario where no financial (accounting) data are available. This has practical value for banks in supporting financial decision-making and can be easily extended to different applications, such as default prediction and credit ratings for SMEs [49]. The results of the offline and online tests also confirmed the practical value of transactional data, which suggest that banks should be aware of the importance of collecting and storing transactional data and that banks should fully utilize the potential value of transactional data when establishing the credit scoring model for SMEs.

Finally, our proposed two-stage feature-selection method can help practitioners find compact feature subsets and construct rule-based credit models that are more propitious to supervision. For instance, two-stage filter-wrapper feature-selection method is a good choice when banks focus on model interpretability, since the empirical results indicate that it generally achieves similar or sometimes better model performance while uses fewer number of features. Furthermore, multi-objective methods give a trade-off between model interpretability and model performance, which provide flexible options for decision-makers [20].

7.3. Limitations

One limitation of this study is the availability of financial statements. Future research can comparatively analyze transactional data-based variables and financial ratio variables for bankruptcy prediction when accounting data are available. Moreover, although we measured the feature importance of transactional data–based variables, we do not know the causal relationship between these variables and bankruptcy. Future research can focus on the interpretability of bankruptcy models and variables.

8. Conclusion

We studied the potential of payment and transactional data for bankruptcy prediction using data provided by banks. Through a comparative analysis of offline and online tests, we found that payment and transactional data–based variables improve SME bankruptcy prediction. Moreover, we proposed a two-stage multiobjective feature-selection method. In the first stage, it removes irrelevant and noisy features using the optimum-seeking method; then, it finds an optimal feature subset that optimizes model performance and feature subset size through multiobjective wrapper-based feature selection. The empirical results showed that the proposed model achieved a similar and sometimes better performance while using far fewer features compared to multiobjective wrapper-based feature-selection methods and other benchmarks. Finally, we evaluated the feature importance of selected features in terms of the Gini index and profit. Both feature importance measures confirmed the importance of payment and transactional data–based variables for bankruptcy forecasting.

CRediT authorship contribution statement

Gang Kou: Conceptualization, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Software, Supervision. **Yong Xu:** Data curation, Formal analysis, Investigation, Methodology, Writing - original draft. **Yi Peng:** Validation, Writing - original draft, Writing - review & editing. **Feng Shen:** Investigation, Methodology, Project administration. **Yang Chen:** Formal analysis, Validation, Writing - review & editing. **Kun Chang:** Methodology, Project administration. **Shaomin Kou:** Funding acquisition, Investigation, Validation.

Declaration of Competing Interest

None.

Acknowledgements

This research has been partially supported by grants from the National Natural Science Foundation of China (#U1811462, #71725001, and #71910107002).

Appendix A. Raw data and information on networks and features

Table A.1
Field descriptions for basic firm information and payment information provided by our dataset.

Field name	Description
Firm information	
PRIPID	Entity key (all unique identifiers are anonymous so that companies cannot be identified by the account ID provided by banks)
ENTTYPE	Company's legal form
INDUSTRYPHY	Business' sector
OPFROM	Date of last annual inspection
APPRDATE	Date that information was last updated
REGSTATE	Operation status (bankruptcy, non-bankruptcy)
REGCAP	Registered capital
RECCAP	Paid-in capital
DOMDISTRICT	Zip code of enterprise location
Manager information	
PRIPID	Entity key (unique identification)
NAMEID	Manager ID
Shareholder information	
PRIPID	Entity key (unique identification)
NAMEID	Shareholder ID
Payment and transactional information	

(continued on next page)

Table A.1 (*continued*)

Field name	Description
SA CUST_NO	CustomerID (Unique identification)
SA_CR_AMT	The quantity (value) of the payment (to comply with privacy regulation, all figures are normalized)
SA_OP_CUST_NAME	Reciprocal account ID (ID of reciprocal enterprises or individuals)
SA_RMRK	Transaction remarks (transfer details, including the payment of taxes, payment of goods, loans, etc.)
SA_TX_DT	Time stamp of the operation

Table A.2
Summary statistics for SMEs' network.

	Manager network	Shareholder network	Payment network
Network type	Weighted undirected network	Weighted undirected network	Weighted directed network
Number of nodes	118,356	897,537	171,203
Number of edges	72,517	73,824	354,960
Average degree	1.225	1.767	2.073
No. of communities	49,484	24,863	1566

Table A.3
Features and descriptions. This table lists features extracted from the last 30 days for transaction-based features due to space constraints. Additionally, features from the last 60, 90, 120, 150, and 180 days are extracted.

Variables	Description
Basic firm features	
INDUSTRY	Business sector
OP_TIME	Age of company
RECCAP_IS_NULL	Missing paid-in capital
RECCAP_IS_ABNORMAL	Abnormal paid-in capital
REG_REC_RATE	Paid-in capital/registered capital
TOTAL_SENIOR	Total number of mangers
TOTAL_SHAREHOLDER	Total number of shareholders
COUNTY_RISK	Historical bankruptcy rate of firm's geographical location
INDUSTRY_RISK	Historical bankruptcy rate of firm's business sector
ENTYPE_RISK	Historical bankruptcy rate of entity's legal form
Manager and shareholder networks-based features	
SENIOR_IN_GRAPH	Whether an entity is in manager network
SENIOR_DEGREE	Degree of nodes in manager network
SENIOR_WEIGHTED_DEGREE	Weighted degree of nodes in manager network
SENIOR_PAGERANK	PageRank of nodes in manager network
SENIOR_NBR_RISK	Bankruptcy rate of neighborhood nodes in manager network
SENIOR_CONNECT_RISK	Bankruptcy rate in connected component of manager network
SENIOR_COMMUNITY_RISK	Bankruptcy rate in community of manager network
SHAREHOLDER_IN_GRAPH	Whether an entity is in shareholder network
SHAREHOLDER_DEGREE	Degree of nodes in shareholder network
SHAREHOLDER_WEIGHTED_DEGREE	Weighted degree of nodes in shareholder network
SHAREHOLDER_PAGERANK	PageRank of nodes in shareholder network
SHAREHOLDER_NBR_RISK	Bankruptcy rate of neighborhood nodes in shareholder network
SHAREHOLDER_CONNECT_RISK	Bankruptcy rate in connected component of shareholder network
SHAREHOLDER_COMMUNITY_RISK	Bankruptcy rate in community of shareholder network
Payment network-based features	
CASH_FLOW_IN_GRAPH	Whether an entity is in payment network
CASH_FLOW_ DEGREE	Degree of nodes in payment network
CASH_FLOW_ IN_DEGREE	In degree of nodes in payment network
CASH_FLOW_ OUT_DEGREE	Out degree of nodes in payment network
CASH_FLOW_ IN_TOTAL_AMT	Total cash inflow from neighbors
CASH_FLOW_ OUT_TOTAL_AMT	Total cash outflow to neighbors
CASH_FLOW_ IN_TOTAL_AVG	Average cash inflow from neighbors
CASH_FLOW_ OUT_TOTAL_AVG	Average cash outflow to neighbors
CASH_FLOW_ PAGERANK	PageRank of nodes in payment network
CASH_FLOW_ NBR_IN_RISK	Bankruptcy rate of inflow neighborhood nodes in payment network
CASH_FLOW_ NBR_OUT_RISK	Bankruptcy rate of outflow neighborhood nodes in payment network
CASH_FLOW_ CONNECT_RISK	Bankruptcy rate in connected component of payment network
CASH_FLOW_ COMMUNITY_RISK	Bankruptcy rate in community of payment network
Transaction-based features	
30_days_in_amt	Total cash inflow of last 30 days
30_days_out_amt	Total cash outflow of last 30 days
30_days_count	Number of records of last 30 days
30_days_in_mean	Average cash inflow of last 30 days
30_days_out_mean	Average cash outflow of last 30 days
30_days_net_amt	Total cash inflow minus total cash outflow of last 30 days
30_days_rate	Total cash inflow/(total cash outflow plus total cash inflow) of last 30 days

Table A.4
Summary statistics of variables selected by the proposed feature-selection method.

Variable	mean	P25	P50	P75
30_days_rate	0.4517	0.4411	0.4539	0.4865
60_days_count	0.0217	0.0047	0.0071	0.0106
60_days_out_mean	0.0056	0.0003	0.0005	0.0007
90_days_rate	0.4586	0.3878	0.4613	0.5000
120_days_count	0.0210	0.0031	0.0056	0.0106
120_days_in_mean	0.0056	0.0002	0.0004	0.0007
150_days_net_amt	0.7364	0.7402	0.7404	0.7404
180_days_count	0.0210	0.0022	0.0048	0.0114
CASH_FLOW_DEGREE	0.0254	0.0000	0.0041	0.0207
CASH_FLOW_COMMUNITY_RISK	0.4201	0.0154	0.0461	0.9863
CASH_FLOW_NBR_OUT_RISK	0.0277	0.0000	0.0000	0.0000
SENIOR_CONNECT_RISK	0.3908	0.4041	0.4052	0.4060
TOTAL_SENIOR	0.0985	0.0769	0.0769	0.0769
ENTTYPE_RISK	0.1866	0.1215	0.1617	0.1641
REC_REG_RATE	0.6272	0.0000	1.0000	1.0000

Note: To comply with privacy regulation, all variables were normalized by min-max scaling.

Appendix B. Meta-parameters for each classifier

Table B.1
Meta-parameters for each classifier.

Classifier	Meta-parameters
LR	L1 penalty with penalty parameter C = 1/1.5
LDA	Least squares solution
DT	CART with maximum depth of trees = 12, Min. total weight of instances in a leaf = 1.0
SVM	Linear SVM with l1 penalty and penalty parameter C = 10
RF	Base learner = CART with maximum depth of trees = 5, no. of member classifiers = 100, no. of features to consider when looking for the best split = sqrt(n features)
XGB	Maximum depth of trees = 5, nrounds = 100, learning rate = 0.1
NN	No. of hidden layers =1, neurons in hidden layers =100, early stopping = False, adam solver with initial learning rate = 0.001, no. of iterations = 300

Appendix C. Results of the online test

This Appendix shows the out-of-sample model performance of online test datasets. Due to the limitations of coverage, we report the predicted performance of out-of-sample model for XGBoost only. AUC and IEMP are reported in Table C.1.

Table C.1
Out-of-sample model performance (AUC and IEMP) of online tests datasets.

Model	Dataset 2		Dataset 3	
	AUC	IEMP	AUC	IEMP
Model 1	0.672	0.31	0.755	3.25
Model 2	0.675	0.35	0.753	3.35
Model 3	0.774	9.01	0.801	6.97
Model 4	0.737	4.01	0.778	4.14
Model 5	**0.787**	**11.18**	**0.817**	**7.26**

Appendix D. Model results using different sampling methods

Table D.1
Model results using different sampling methods on the offline dataset.

Sampling methods	Training data		Test data	
	AUC	IEMP	AUC	IEMP
Random undersampling	0.762 ± 0.21*	12.15 ± 4.74*	0.769	9.38
Random oversampling	0.759 ± 0.21	13.53 ± 5.30	0.748	9.85
SMOTE	0.715 ± 0.20	6.87 ± 3.32	0.665	0.09
No sampling used	$\mathbf{0.782 \pm 0.22}$	$\mathbf{15.74 \pm 5.35}$	**0.785**	**12.05**

Note: XGB is used as the base classifier. The results of tenfold cross-validation of the training data and the predict result on test data are reported. The best results are in bold, while * indicates significant similarity to the best results at $p < 0.05$.

Appendix E. Model results using different generation number and population size

Table E.1
Model results using different generation number and population size.

Model ID	Population size	No. of generation	Hypervolume (AUC)	Hypervolume (IEMP)	Spacing (AUC)	Spacing (IEMP)
Panel A: Meta-parameter tuning results for a subset of training data						
$MOFS_{AUC}$	25	50	22.17 ± 1.10	14.32 ± 1.77*	0.16 ± 0.05*	0.193 ± 0.09*
$MOFS_{AUC}$	50	100	23.44 ± 0.74*	14.17 ± 0.85*	**0.14 ± 0.04**	**0.13 ± 0.05**
$MOFS_{AUC}$	100	200	**23.96 ± 0.50**	**14.55 ± 1.40**	0.14 ± 0.05*	0.16 ± 0.07*
Panel B: Comparison results for $TSMOFS_{AUC}$ and $MOFS_{AUC}$						
$TSMOFS_{AUC}$	100	200	**26.05 ± 0.82**	**16.17 ± 1.43**	**0.12 ± 0.05**	**0.12 ± 0.06**
$MOFS_{AUC}$	100	200	24.25 ± 0.36	15.15 ± 1.20*	0.17 ± 0.09*	0.15 ± 0.10*

Note: The mean values ± standard deviation are reported, and the best results are in bold.
* Indicates significant similarity to the best results at $p < 0.05$.

References

[1] E.I. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, J. Financ. 23 (4) (1968) 589–609, https://doi.org/10.2307/2978933.
[2] E.I. Altman, A. Saunders, Credit risk measurement: developments over the last 20 years, J. Bank. Financ. 21 (11–12) (1997) 1721–1742, https://doi.org/10.1016/S0378-4266(97)00036-8.
[3] S. Tian, Y. Yu, H. Guo, Variable selection and corporate bankruptcy forecasts, J. Bank. Financ. 52 (2015) 89–100, https://doi.org/10.1016/j.jbankfin.2014.12.003.
[4] E.I. Altman, G. Sabato, Modelling credit risk for SMEs: evidence from the U.S. market, Abacus 43 (43) (2007) 332–357, https://doi.org/10.1111/j.1467-6281.2007.00234.x.
[5] F. Ciampi, N. Gordini, Small enterprise default prediction modeling through artificial neural networks: an empirical analysis of Italian small enterprises, J. Small Bus. Manag. 51 (2013) 23–45, https://doi.org/10.1111/j.1540-627X.2012.00376.x.
[6] E.I. Altman, G. Sabato, N. Wilson, The value of non-financial information in small and medium-sized enterprise risk management, J. Credit Risk 6 (2) (2010) 1–33, https://doi.org/10.21314/JCR.2010.110.
[7] G.B. Fernandes, R. Artes, Spatial dependence in credit risk and its improvement in credit scoring, Eur. J. Oper. Res. 249 (2016) 517–524, https://doi.org/10.1016/j.ejor.2015.07.013.
[8] D. Liang, C.C. Lu, C.F. Tsai, G.A. Shih, Financial ratios and corporate governance indicators in bankruptcy prediction: a comprehensive study, Eur. J. Oper. Res. 252 (2) (2016) 561–572, https://doi.org/10.1016/j.ejor.2016.01.012.
[9] J. Grunert, L. Norden, M. Weber, The role of non-financial factors in internal credit ratings, J. Bank. Financ. 29 (2) (2005) 509–531, https://doi.org/10.1016/j.jbankfin.2004.05.017.
[10] F. Ciampi, Corporate governance characteristics and default prediction modeling for small enterprises. An empirical analysis of Italian firms, J. Bus. Res. 68 (5) (2015) 1012–1025, https://doi.org/10.1016/j.jbusres.2014.10.003.
[11] S. Angilella, S. Mazz, The financing of innovative SMEs: a multicriteria credit rating model, Eur. J. Oper. Res. 244 (2) (2015) 540–554, https://doi.org/10.1016/j.ejor.2015.01.033.
[12] 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, https://doi.org/10.1016/j.dss.2010.03.005.
[13] A. Ptak-Chmielewska, Predicting micro-enterprise failures using data mining techniques, J. Risk Financ. Manag. 12 (2019) 30, https://doi.org/10.3390/jrfm12010030.
[14] S. Battiston, D. Delli Gatti, M. Gallegati, B. Greenwald, J.E. Stiglitz, Credit chains and bankruptcy propagation in production networks, J. Econ. Dyn. Control. 31 (6) (2007) 2061–2084, https://doi.org/10.1016/j.jedc.2007.01.004.
[15] S.A. Yang, J.R. Birge, R.P. Parker, The supply chain effects of bankruptcy, Manag. Sci. 61 (10) (2015) 2320–2338, https://doi.org/10.1287/mnsc.2014.2079.
[16] E. Letizia, F. Lillo, Corporate payments networks and credit risk rating, EPJ Data Sci. 8 (2019) 8–21, https://doi.org/10.1140/epjds/s13688-019-0197-5.
[17] E. Tobback, T. Bellotti, J. Moeyersoms, M. Stankova, D. Martens, Bankruptcy prediction for SMEs using relational data, Decis. Support. Syst. 102 (2017) 69–81, https://doi.org/10.1016/j. dss.2017.07.004.
[18] R.M. Stein, The relationship between default prediction and lending profits: integrating ROC analysis and loan pricing, J. Bank. Financ. 29 (5) (2005) 1213–1236, https://doi.org/10.1016/j. jbankfin.2004.04.008.
[19] E.I. Altman, A. Giannozzi, O. Roggi, G. Sabato, Building Sme rating: is it necessary for lenders to monitor financial statements of the borrowers? Bancaria 10 (2013) 54–71.
[20] N. Kozodoi, S. Lessmann, K. Papakonstantinou, Y. Gatsoulis, B. Baesens, A multi-objective approach for profit-driven feature selection in credit scoring, Decis. Support. Syst. 120 (2019) 106–117, https://doi.org/10.1016/j.dss.2019.03.011.
[21] R.O. Edmister, An empirical test of financial ratio analysis for small business failure prediction, J. Financ. Quant. Anal. 7 (2) (1972) 1477–1493, https://doi.org/10.2307/2329929.
[22] J. Saurina, C. Trucharte, The impact of Basel II on lending to small and medium-sized firms: a regulatory policy assessment based on Spanish credit register data, J. Financ. Serv. Res. 26 (2004) 121–144, https://doi.org/10.1023/B:FINA.0000037542.16864.11.
[23] I. Guyon, A. Elisseeff, An introduction to feature selection, Appl. Predict. Model. 3 (1) (2003) 1157–1182, https://doi.org/10.1007/978-1-4614-6849-3_19.
[24] I. Guyon, S. Gunn, M. Nikravesh, et al., Feature Extraction, Foundations and Applications, Springer, Berlin, 2006.
[25] K. Deb, S. Agrawal, A. Pratap, T. Meyarivan, A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II, Lect. Notes Comput. Sci 1917 (2000) 849–858, https://doi.org/10.1007/3-540-45356-3_83 (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics).
[26] E.I. Altman, G. Sabato, Modeling credit risk for SMEs: evidence from the US market, ABACUS. 43 (2008) 332–357, https://doi.org/10.1111/j.1467-6281.2007.00234.x.
[27] M. Dietsch, J. Petey, Should SME exposures be treated as retail or corporate exposures? A comparative analysis of default probabilities and asset correlations in French and German SMEs, J. Bank. Financ. 28 (2004) 773–788, https://doi.org/10.1016/j.jbankfin.2003.10.006.
[28] G.X. Yuan, H. Wang, T. Zeng, T. Wu, S. Gao, H. Ma, The dynamical mechanism for SMEs evolution under the hologram approach, SSRN Electron. J. 3325013 (2019) 1–51, https://doi.org/10.2139/ssrn.3325013.
[29] S. Figini, F. Bonelli, E. Giovannini, Solvency prediction for small and medium enterprises in banking, Decis. Support. Syst. 102 (2017) 91–97, https://doi.org/10.1016/j.dss.2017.08.001.
[30] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: criteria of MaxDependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238, https://doi.org/10.1109/TPAMI.2005.159.
[31] F. Jimnez, G. Snchez, J.M. Garca, G. Sciavicco, L. Miralles, Multi-objective evolutionary feature selection for online sales forecasting, Neurocomputing. 234 (2017) 75–92, https://doi.org/10.1016/j.neucom.2016.12.045.
[32] L. Breiman, Random Forest, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1017/CBO9781107415324.004.
[33] A.N. Berger, W.S. Frame, Small business credit scoring and credit availability, J. Small Bus. Manag. 45 (2007) 5–22, https://doi.org/10.1111/j.1540-627X.2007.00195.x.
[34] J. Lpez, S. Maldonado, Profit-based credit scoring based on robust optimization and feature selection, Inf. Sci. 500 (2019) 190–202, https://doi.org/10.1016/j.ins.2019.05.093.
[35] V.D. Blondel, J.L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 2008 (2008), https://doi.org/10.1088/1742-5468/2008/10/P10008.
[36] H. He, E.A. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (2009) 1263–1284, https://doi.org/10.1007/978-3-030-04663-7_4.
[37] Y.S. Chen, Classifying credit ratings for Asian banks using integrating feature selection and the CPDA-based rough sets approach, Knowledge-Based Syst. 26 (2012) 259–270, https://doi.org/10.1016/j.knosys.2011.08.021.
[38] E. Hancer, B. Xue, M. Zhang, Differential evolution for filter feature selection based on information theory and feature ranking, Knowledge-Based Syst. 140 (2018) 103–119, https://doi.org/10.1016/j.knosys.2017.10.028.
[39] M. Kregar, Cash Flow Based Bankruptcy Risk and Stock Returns in the Us Computer and Elec-tronics Industry, 2011, pp. 1–154.
[40] A.K. Uysal, On two-stage feature selection methods for text classification, IEEE Access. 6 (2018) 43233–43251, https://doi.org/10.1109/ACCESS.2018.2863547.
[41] L.Y. Chuang, C.H. Ke, H.W. Chang, C.H. Yang, A two-stage feature selection method for gene expression data, Omi. A J. Integr. Biol. 13 (2009) 127–137, https://doi.org/10.1089/omi.2008.0083.
[42] E. Zitzler, L. Thiele, M. Laumanns, C.M. Fonseca, V.G. Da Fonseca, Performance assessment of multiobjective optimizers: an analysis and review, IEEE Trans. Evol. Comput. 7 (2003) 117–132, https://doi.org/10.1109/TEVC.2003.810758.
[43] E. Zitzler, L. Thiele, Multiobjective optimization using evolutionary algorithms a comparative case study. In International conference on parallel problem solving from nature, 1998, pp. 292–301, https://doi.org/10.1007/BFb0056872.

[44] S. Balcaen, H. Ooghe, 35 years of studies on business failure: an overview of the classical statistical methodologies and their related problems, Br. Account. Rev. 38 (2006) 63–93, https://doi.org/10.1016/j.bar.2005.09.001.
[45] V. Garca, A.I. Marqus, J.S. Snchez, Exploring the synergetic effects of sample types on the performance of ensembles for credit risk and corporate bankruptcy prediction, Inf. Fusion. 47 (2019) 88–101, https://doi.org/10.1016/j.inffus.2018.07.004.
[46] E. Alfaro, N. Garca, M. Gmez, D. Elizondo, Bankruptcy forecasting: an empirical comparison of AdaBoost and neural networks, Decis. Support. Syst. 45 (2008) 110–122, https://doi.org/10.1016/j.dss.2007.12.002.
[47] 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 (2015) 124–136, https://doi.org/10.1016/j.ejor.2015.05.030.
[48] G.M. Weiss, Mining with rarity: a unifying framework, SIGKDD Explor. 6 (2004) 7–19, https://doi.org/10.1145/1007730.1007734.
[49] F Shen, X. Zhao, G. Kou, Three-stage reject inference learning framework for credit scoring using unsupervised transfer learning and three-way decision theory, Decis. Support. Syst. 137 (2020) 1–10, https://doi.org/10.1016/j.dss.2020.113366.
[50] H. Wang, G. Kou, Y. Peng, Multi-class misclassification cost matrix for credit ratings in peer-to-peer lending, J. Oper. Res. Soc. 05 (2020) 1–12, https://doi.org/10.1080/01605682.2019.1705193.
[51] G. Kou, Y. Peng, G. Wang, Evaluation of Clustering Algorithms for Financial Risk Analysis using MCDM Methods, Inf. Sci. 275 (11) (2014) 1–12, https://doi.org/10.1016/j.ins.2014.02.137.
[52] G. Kou, X. Chao, Y. Peng, F.E. Alsaadi, E. Herrera-Viedma, Machine learning methods for systemic risk analysis in financial sectors, Technol. Econ. Dev. Eco. 25 (5) (2019) 716–742, https://doi.org/10.3846/tede.2019.8740.

Gang Kou is a Distinguished Professor of Chang Jiang Scholars Program in Southwestern University of Finance and Economics, managing editor of International Journal of Information Technology & Decision Making (SCI) and managing editor-in-chief of Financial Innovation (SSCI). He is also editors for the following journals: European Journal of Operational Research, and **Decision Support Systems**. Previously, he was a professor of School of Management and Economics, University of Electronic Science and Technology of China, and a research scientist in Thomson Co., R&D. He received his Ph.D. in Information Technology from the College of Information Science & Technology, Univ. of Nebraska at Omaha; Master degree in Dept of Computer Science, Univ. of Nebraska at Omaha; and B.S. degree in Department of Physics, Tsinghua University, China. He has published more than 100 papers in various peer-reviewed journals.

Yong Xu is currently a Ph.D. candidate in School of Business Administration, Southwestern University of Finance and Economics. His research interests include credit scoring and machine learning, particularly cost-sensitive learning.

Yi Peng is a Chang Jiang Scholar Chair Professor of School of Management and Economics, University of Electronic Science and Technology of China. Previously, she worked as Senior Analyst for West Co., USA. Dr. Peng received her Ph.D. in Information Technology from the College of Information Science & Technology, Univ. of Nebraska at Omaha and got her Master degree in Dept of Info. Science & Quality Assurance, Univ. of Nebraska at Omaha and B.S. degree in Department of Management Information Systems, Sichuan University, China. Dr. Peng's research interests cover Knowledge Discover in Database and data mining, multi-criteria decision making, data mining methods and modeling, knowledge discovery in real-life applications. She has published more than eighty papers, which have been cited for several thousand times, in various peer-reviewed journals such as Omega. Yi Peng is listed as the Highly Cited Researcher by Clarivate Analytics (Web of Science).

Feng Shen is a professor of School of Finance, Southwestern University of Finance and Economics. He received his Ph.D. from the School of Business, Sichuan University in 2015. His current research interests include financial risk management, credit scoring, machine learning and data mining.

Yang Chen is a Professor in School of Business Administration, Southwestern University of Finance and Economics in China. He received his Ph.D. in Finance and Decision Sciences from Hong Kong Baptist University. His current research interests include the dark side of social media, human resource management, and so on. He has published research papers in journals such as Decision Sciences, Decision Support Systems, European Journal of Information Systems, Human Resource Management, Information & Management, Journal of Business Ethics, Journal of Information Technology, and Journal of Product Innovation Management. He is currently serving as the Associate Editor for Internet Research.

Kun Chang, General Manager of Department of Data Application, Shandong City Commercial Banks Alliance Co., LTD, Jinan, 250104, China.

Shaomin Kou, Assistant General Manager of Department of Data Application, Shandong City Commercial Banks Alliance Co., LTD, Jinan, 250104, China.