sustainability

MDPI

Article

Predicting the Insolvency of SMEs Using Technological Feasibility Assessment Information and Data Mining Techniques

Sanghoon Lee [1], Keunho Choi [2] and Donghee Yoo [1,*]

[1] Department of Management Information Systems, BERI, Gyeongsang National University, 501 Jinju-daero, Jinju 52828, Korea; lshoon@kosmes.or.kr

[2] Department of Business and Accounting, Hanbat National University, 125 Dongseo-daero, Yuseong-gu, Daejeon 34158, Korea; keunho@hanbat.ac.kr

* Correspondence: dhyoo@gnu.ac.kr; Tel.: +82-55-772-1536

Received: 17 October 2020; Accepted: 20 November 2020; Published: 24 November 2020

Abstract: The government makes great efforts to maintain the soundness of policy funds raised by the national budget and lent to corporate. In general, previous research on the prediction of company insolvency has dealt with large and listed companies using financial information with conventional statistical techniques. However, small- and medium-sized enterprises (SMEs) do not have to undergo mandatory external audits, and the quality of accounting information is low due to weak internal control. To overcome this problem, we developed an insolvency prediction model for SMEs using data mining techniques and technological feasibility assessment information as non-financial information. We divided the dataset into two types of data based on three years of corporate age. The synthetic minority over-sampling technique (SMOTE) was used to solve the data imbalance that occurred at this time. Six insolvency prediction models were created using logistic regression, a decision tree, an artificial neural network, and an ensemble (i.e., boosting) of each algorithm. By applying a boosted decision tree, the best accuracies of 69.1% and 82.7% were derived, and by applying a decision tree, nine and seven influential factors affected the insolvency of SMEs established for fewer than three years and more than three years, respectively. In addition, we derived several insolvency rules for the two types of SMEs from the decision tree-based prediction model and proposed ways to enhance the health of loans given to potentially insolvent companies using these derived rules. The results of this study show that it is possible to predict SMEs' insolvency using data mining techniques with technological feasibility assessment information and find meaningful rules related to insolvency.

Keywords: policy funds; SMEs; technological feasibility assessment; insolvency prediction model; SMOTE; logistic regression; decision tree; artificial neural network; boosting

---

1. Introduction

A company bankrupted due to its insolvency faces a very important event causing economic losses for various stakeholders, including shareholders, investors, and creditors. In case of a bankruptcy, it is hard to receive full legal compensation, and accumulated resources such as manufactured capital, human capital (e.g., education, knowledge, and science) and social capital (e.g., administration, social trust, and networks) by the company disappear. For these reasons, sustainability in business operations is very important to prevent social losses due to legal procedures such as liquidation proceedings and restructuring proceedings [1].

Thus, predicting insolvent companies and normalizing their business before becoming insolvent companies are one of the ways to minimize economic and social losses. As financial information provides useful information for predicting a bankrupt company, it has been used by previous researchers to predict company insolvency [2,3] and for analyzing a company's credit risk.

According to a survey conducted in South Korea in 2019, the main reasons for the closure were failure in sales marketing, failure in financing, lack of experience, difficulty in marketing environment, failure in financial management, and failure in Research and Development (R&D) [4]. Research conducted in South Korea to predict company insolvency has focused primarily on companies listed on the Korea Exchange (KRX) and Korea Securities Dealers Automated Quotation (KOSDAQ). The reason for this is that it is easy to collect information about these companies from financial statements provided by the Financial Supervisory Service's Data Analysis Retrieval and Transfer System (dart.fss.or.kr), news provided by various media outlets, industry trends report, and social media. However, it is relatively hard to collect financial information of small- and medium-sized enterprises (SMEs), especially SMEs established for fewer than three years and therefore having limited financial information. In addition, as financial statements are prepared once a year, they are not timely enough to reflect the difference between the time we collect financial information and the time we analyze the data in rapidly changing environments. In other words, previous research on the prediction of company insolvency has dealt with large and listed companies using financial statements with conventional statistical techniques. However, it is difficult to have sufficient and accurate financial information for SMEs because of their vulnerability and difficulties in management. This situation has resulted in a lack of research on the prediction of insolvency focusing on SMEs. Therefore, it is less appropriate to use financial information to predict SMEs' insolvency, and there is a need to explore new information that enables SMEs to make insolvency predictions based on their latest activities.

The Korea SMEs and Startups Agency (hereinafter KOSME) is a quasi-government institution that supplies approximately KRW 4 trillion (USD 3.4 billion) as policy funds to the market every year. KOSME divides various types of funds into policy support areas and implements loans by assessment. At each stage of assessment, appropriate technological feasibility assessment and financial assessment models are built and used according to the types of SMEs. As this information is timely and includes various aspects of SMEs, it can be used to predict SMEs' insolvency. Therefore, this study intends to use this information as a new indicator for insolvency prediction of SMEs instead of using financial information.

This study examines whether it is possible for SMEs to predict insolvency with the technological feasibility assessment information from KOSME. To this end, we first divide collected companies into those established for fewer than three years and those for more than three years. We then use logistic regression, a decision tree, an artificial neural network algorithm, and an ensemble (i.e., boosting) of each of them to build six insolvency prediction models for each type of SME in order to determine which predictive model has the highest accuracy. We also seek to determine whether the prediction model built solely using items of technological feasibility assessment information without financial information can be used to predict insolvency by comparing it with prediction models proposed by previous research using only financial information. Finally, the decision tree-based prediction model is used to derive the rules of insolvency of SMEs established for fewer than three years and more than three years. Based on the rules derived, we propose a method to increase the soundness of policy funds in the corporate review and assessment stage.

The rest of the paper is organized as follows: Section 2 reviews policy funds and previous research closely related to company insolvency. Section 3 presents our research method and experimental data. Section 4 shows the experimental results for prediction models and insolvency rules of SMEs. Finally, Section 5 provides the conclusion to this paper, including research implications, possible limitations of our approach, and potential directions for future work.

2. Literature Review

2.1. Policy Funds

Giacosa and Mazzoleni [5] used financial debt ratios and profitability to classify companies into six types: excellent companies, mature companies, development-stage companies, declining companies, restructuring companies, and crisis companies. They presented the sources of financing from each type of company as private equity funds, mini bond, traditional banks, institutional guarantees, and investment funds. Among them, most SMEs generally raise their funds by financial firms.

When financial firms make decisions to provide funds to companies, they rely on the credit rating of the companies in their decision-making process. As financial firms are in business areas that can legally differentiate companies and there is no disadvantage even if they do not lend to companies with low credit ratings, they target companies that have received investment grade or higher in credit ratings. Therefore, if funding for companies is only assigned to the financial market, companies with low credit ratings such as venture companies and companies in the early stages of growth will not be able to access the financial market despite their potential.

From the financial firms' perspective, SMEs have high credit risks due to their lack of management system, human resources, and marketing capabilities. The lack of SMEs' management stability makes it difficult to calculate their accurate credit risk. In addition, as the amount of SMEs' loans tends to be low and the handling costs tend to be high, the profit of the financial firms is low. Meanwhile, from the SMEs' perspective, the interest rates and collateral are relatively unfavorable for SMEs compared to large companies [6]. To lessen the problems, the government supplies policy funds to them in the way of loans, guarantees, and investments so that they can raise funds more smoothly. The policy fund is a system that supplies public credit support funds in order to foster the SMEs in a policy manner. It is raised by government ministries, local governments, and their affiliated agencies through government budgets, bond issuances, and so on [7]. In other words, policy funds are defined as "credit that the government provides limited funds to specific sectors preferentially in terms of interest rates and repayment conditions in order to achieve its policy goals" [8].

As such, one of the devices to compensate for failure in the financial market is the policy fund of KOSME. KOSME operates policy funds by differentiating itself from market financing in all aspects of financial conditions, including credit ratings, company size, company age, usage of loan, collateral, and interest rates. Companies that are shunned by market financing are eligible [6]. The policy fund strengthens the financial accessibility of SMEs and contributes to resolving the financial difficulties while it inevitably leads to a higher rate of insolvency than banks [9].

A financial assessment and technological feasibility assessment are conducted to screen companies seeking to be beneficiaries of policy funds. In Lee and Kim's [10] research, when the weight of the credit score and technical score was 7 to 3, the accuracy of the insolvency prediction was the highest. This result shows the possibility that technology credit information can be substituted for credit rating information. Lim [11] showed that better accuracy of insolvency prediction was achieved when considering financial statement and technological feasibility assessment information at the same time, instead of using only financial statement information. In terms of the technology assessment information, it was estimated that variables related to management and business feasibility contribute the most to insolvency prediction. In addition, some studies have argued that the ability to select and discover promising companies based on growth and technological prowess is needed to differentiate policy funds from private financing [7].

Kim and Han [12] found that the higher the technology rating, the better the financial performance based on a comparison of the validity of the technological feasibility assessment information rating to the financial performance variables, such as the ratio of ordinary income to sales, asset turnover ratio, R&D investment ratio, and debt ratio.

As KOSME does not use the financial assessment model when evaluating companies established for fewer than three years, it is necessary to make efforts to achieve the policy's purpose of supplying funds to companies in the early stages of their start-ups and enhance the stability of the funds by utilizing the technological feasibility assessment information.

2.2. Research on Predicting Company Insolvency

Predicting company insolvency is a long-discussed topic in the accounting and financial domains and a rare event affecting a number of stakeholders in economies connected by various supply chains. Most research on insolvency prediction has used static or dynamic techniques for the analyses and used samples without considering the size of the companies. Research focusing on SMEs has mainly used financial ratios.

Altman's z-score model predicted the insolvency of the listed companies using five ratios: working capital/total assets, retained earnings/total assets, Earnings Before Interest and Taxes (EBIT)/total assets, market value equity/book value of total debt, and sales/total assets [3]. Kolte et al. [13] analyzed insolvency by calculating Altman's z-scores for an airline and a financial company. Ohlson's research [14] used nine variables for insolvency prediction. Finally, Altman and Sabato [15] made the predictions using cash/total assets, short-term borrowing, current-growth period liabilities/equity, Earnings Before Interest, Taxes, Depreciation and Amortization (EBITDA)/total assets and EBITDA/interest expenses. In addition, Podviezko et al. [16] recognized that the financial ratio is the most important indicator of the company's performance, but it is lacking data. Thus, they tried to analyze the weights of criteria before and after bankruptcy by combining two categories, soft and hard criteria, to identify the SME's crisis.

Starting with the univariate analysis of Beaver [2], the insolvency prediction model was developed using various statistical techniques, such as multiple linear regression analysis [17] and logistic regression analysis [18]. These traditional statistical techniques showed an accuracy of about 80% to 90%, even though they have shortcomings, including strict assumptions such as linearity and normality [3,19,20]. Although previous research in South Korea has utilized a variety of more advanced algorithms, it has shown the accuracy of about 70% to 80%, which is lower than that of foreign research [21].

Financial information has the advantage of being able to access the status of companies with the most standardized form of data. But it has a limitation in that it takes a certain period of time to be disclosed after a quarterly or annual basis of settlement, external audit, etc. Although the financial information of large companies and listed companies is easy to access because they are open to the public, the financial information of SMEs is limited in their acquisition. For these reasons, there are few studies on the insolvency of SMEs and start-up companies established for fewer than three years.

Recently, in addition to research utilizing conventional statistical techniques, research utilizing artificial intelligence (AI) has been active in developing prediction models for company insolvency. This is to overcome the limitations of the conventional statistical approach, and there have been various studies applying data mining techniques, such as artificial neural networks, decision trees, Bayesian networks, support vector machines, ensembles, and mixed techniques [19,22,23].

Another study attempted to predict the insolvency of companies using a decision tree and an association algorithm by analyzing news information, which is unstructured information, with the exception of financial information [24]. Lee and Han [25] used non-financial information, such as type of business, number of employees, number of exports, and technical manpower ratings, to build a prediction model along with financial information. In addition to the financial ratio, the types of non-financial information used for insolvent prediction were based on research focusing on industry sector, region, and company age [26,27]; the quality of accounting information, firm owners' personal credit performance, and management quality [28]; context-based feature set designing industry-representative models [19]; and indicators such as value-added rate and labor productivity [29]. The prediction rates for these studies ranged from 74.1% to 93%. Kim et al. [21] built an insolvency prediction model using Statistics Korea's economic trends, the volatility of financial status derived from the financial ratios, and the volatility of industrial conditions. In addition, another study improved the accuracy of traditional insolvency prediction models using a sentiment analysis based on a sentiment dictionary built from economic news [30].

In general, bankruptcy can be caused by poor management, improper sales forecasting, inexperienced management, rapid technological advances, preference changes, and inability [31]. In addition, company age is an important factor to determine company bankruptcy. Pompe and Bilderbeek [32] found that bankruptcy among young firms is more difficult to predict than among established firms. Makropoulos et al. [33] also showed that the age of a company was a determinant of failure for SMEs in the UK. In the case of companies in Austria, the lack of equity is a core factor in company bankruptcy, which is related to the age of the company [34]. Akbar et al. [35] divided company life cycle, not company age, into five stages including introduction, growth, mature, shake-out, and decline, and revealed the risk of bankruptcy at each of the stages.

Table 1 summarizes the main research on predicting company insolvency.

Table 1. Research on insolvency prediction.

Reference	Number of Features	Method	Data Type	Number of Companies		Accuracy (%)
				Insolvency	Healthy	
[2]	6	Univariate Analysis	Finance	79	79	78–87
[3]	5	Multiple Discriminant Analysis	Finance	33	33	74–95
[14]	9	Logit Model	Finance	105	2058	92.84–96.12
[20]	26	Case-based Reasoning	Finance	42	279	96.90
[22]	11	Artificial Neural Network	Finance	944	944	80.89
[23]	24	FS-Boosting	Finance	66	66	86.79
[32]	45	Multiple Discriminant Analysis, Neural Network	Finance	1500	1500	72–79
[36]	5	Artificial Neural Network, Multivariate Adaptive Regression Splines	Finance	256	63,107	89.58
[37]	16	Genetic Algorithm	Finance	1570	1570	77.3–82.5
[38]	11	Logistic Regression Discriminant Analysis	Finance	8220	67,432	90.4–93.8
[39]	11	Logistic Regression	Finance	46	46	82.6
[24]	7	Decision Tree	Non-Finance	177	160	56.2–83.1
[19]	12	Stacking	Finance, Non-Finance	815	8191	93.10
[25]	9	Artificial Neural Network	Finance, Non-Finance	20	221	75
[26]	16	Logistic Regression	Finance, Non-Finance	3576	3576	83
[27]	25	Linear Discriminant Analysis	Finance, Non-Finance	584	8393	79
[28]	7	Logistic Regression	Finance, Non-Finance	127	698	88.1
[29]	4	Logistic Regression	Finance, Non-Finance	112	112	74.1–79

In summary, many previous studies have attempted to predict the insolvency of companies using financial information or non-financial information, which SMEs and start-up companies rarely have. Thus, previous prediction models using this information are less appropriate for predicting SMEs' insolvency.

On the other hand, a technological feasibility assessment has been conducted to screen companies seeking to be beneficiaries of policy funds, and some studies have argued that the ability to identify promising companies based on growth and technological prowess, which the technical feasibility assessment information includes, is needed to differentiate policy funds from private financing.

Therefore, to explore new information enabling SMEs to make insolvency predictions based on their latest activities, we developed the following hypothesis:

Hypothesis 1 (H1). *Technical feasibility assessment information could be a substitute for financial information when predicting the insolvency of SMEs.*

3. Research Method

3.1. Research Framework

Figure 1 shows our research framework. First, we collected data on the technological feasibility assessment conducted when lending to SMEs. The pre-processing process extracts technological feasibility assessment information of the manufacturing industry to be used for analysis. Based on company age, we organized two datasets: companies established for fewer than three years and those established for more than three years. The over-sampling technique was then applied to mitigate the data imbalance problem, which occurs in the samples of insolvent and healthy companies.

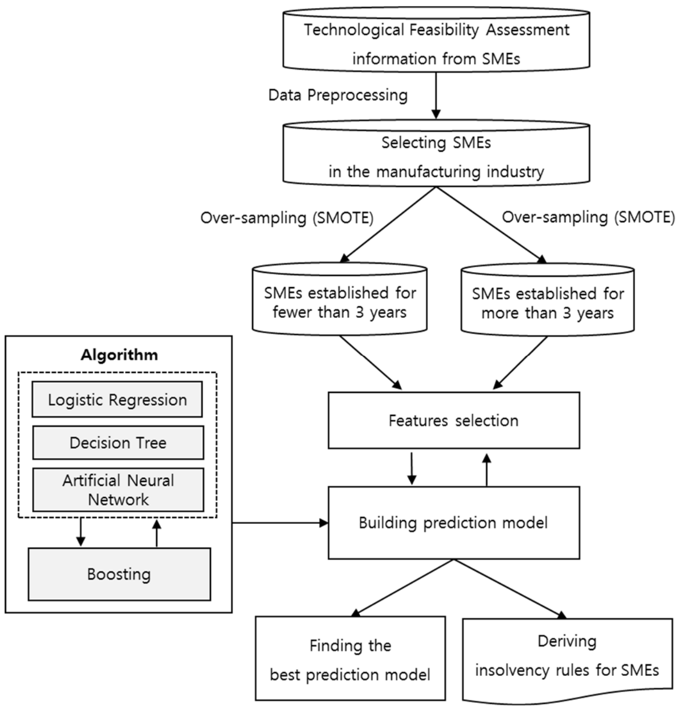

Figure 1. Research framework.

Next, the technological feasibility assessment variables to be used as the candidates of the independent variable in the analysis were determined, and the variables that play an important role in predicting the target variable among the candidates were finally selected. Then six prediction models based on the logistic regression, the decision tree, the artificial neural network, and an ensemble of each algorithm were built in each dataset.

Among the prediction models, we would like to find the best predictable model and derive rules for the insolvency of companies established for fewer than three years and more than three years from the results of decision tree-based prediction model. Finally, this study attempted to derive the variables of technological feasibility assessment, which were highly related to insolvency, and to suggest strategies for reducing the insolvency.

3.2. Data

Company insolvency is a rare event, and the data imbalance that appears in collecting data may cause problems by potentially reducing the accuracy of classification, thereby producing inaccurate results when building a company insolvency prediction model if we use the entire population as learning data without balancing through sampling [40]. To mitigate the data imbalance problem, under-sampling and over-sampling methods can be used [41]. Under-sampling is a method of fitting the number of samples of a major class to that of a minor class by reducing the number of samples of the major class according to a set of rules. This method has benefits in that it reduces learning times and has a low possibility of an over-fitting problem. However, it has problems in that the sample might not represent the characteristics of the population and the number of samples might not be enough for learning the prediction model. Meanwhile, over-sampling is a method of fitting the number of samples of a minor class to that of a major class by increasing the number of samples of the minor class according to a set of rules.

This study used the synthetic minority over-sampling technique (SMOTE) proposed by Chawla et al. [42] to mitigate the data imbalance problem. SMOTE is a method of creating new samples by synthesizing k-nearest neighbors around the samples of a minor class. First, we selected a random x_i in samples of the minor class and selected a random neighbor $\hat{x}_i$ among k-nearest neighbors. Then we generated a new synthetic sample x_{new} using Equation (1):

$$x_{new} = x_i + \lambda(\hat{x}_i - x_i), \quad \lambda \in [0, 1] \tag{1}$$

Figure 2a shows that the nearest neighbor $\hat{x}_i$ is selected for x_i, and Figure 2b shows that a new synthetic sample is generated by generating a random number between x_i and $\hat{x}_i$. This method is performed on a randomly selected sample repeatedly until the desired equilibrium ratio is reached. As this method does not simply replicate samples in a minor class, it has the advantage of minimizing the over-fitting problem [43].

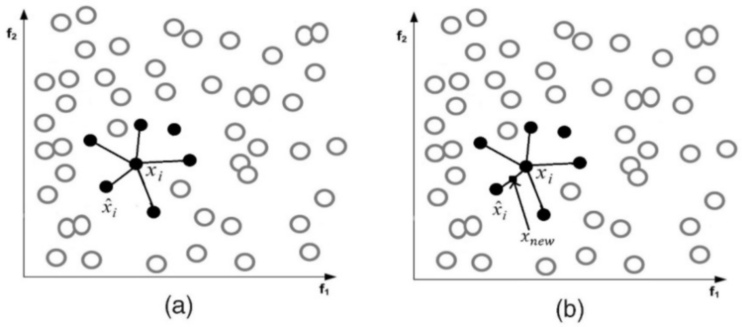

Figure 2. Synthetic sample generation using SMOTE (**a**) The K-nearest neighbors-based over-sampling technique. (**b**) The synthetic minority over-sampling technique (SMOTE).

The data set used in this study was obtained from KOSME. After receiving applications for policy funds, a total of 5118 cases were financed after assessment. Among them, 4358 cases were ultimately analyzed after removing cases of small asset companies with limited data. Companies were then divided into those established for fewer than three years and those established for more than three years.

As the indicators in the assessment model for companies established for fewer than three years are different from those for companies established for more than three years, there is no direct comparison between the indicators. The loan period was generally around five years, so we used the data for companies financed in 2014, which could confirm the insolvency within the principal repayment period.

The number of samples for each type of company is shown in Table 2. SMOTE was applied to solve the data imbalance problem in which healthy companies occur 2.7 times more often than insolvent companies. Finally, 2342 companies established for fewer than three years and 4036 companies established for more than three years were used in our experiments.

Table 2. Number of samples.

Type	Initial Collected Data			Over-Sampling Data		
	Healthy	Insolvency	Sum	Healthy	Insolvency	Sum
Fewer than three years	1171	618	1789	1171	1171	2342
More than three years	2018	549	2569	2018	2018	4036
Sum	3189	1167	4358	3189	3189	6378

3.3. Target Variable

Insolvency means a loss arising from the failure of a financed company to perform its obligations normally, and it can be calculated monthly, quarterly, or yearly in terms of the period. In this case, as the insolvency rate can be changed according to the period, the annual insolvency rate is the general standard for financial firms and credit rating firms. In addition, the insolvency rate can be calculated differently depending on the number of companies and the amount of insolvency. The insolvency rate based on the number of companies can be calculated by dividing the number of companies with outstanding loans for a certain period of time by that at the beginning of the year. Similarly, the insolvency rate based on the amount of insolvency can be calculated by dividing the amount of insolvency for a certain period of time by that at the beginning of the year.

This study defined an insolvent company as a company that terminates a contract or has arrearages for more than three months among the companies financed from KOSME; it used a variable representing whether a company is an insolvent one or not as a target variable.

3.4. Independent Variable

The technological feasibility assessment is designed to evaluate indicators of management ability, business feasibility, and technical ability depending on the company age, the industry to which the company belongs, and the type of technology characteristics. In this study, a total of 37 assessment indicators of manufacturing companies evaluated for the policy funds in 2014 were used as independent variables. The summary of assessment indicators is shown in Table 3.

The technological feasibility assessment indicator consists of three main categories: management ability, business feasibility, and technical ability. Management ability evaluates the business promotion ability, CEO's reliability and expertise, and the business management ability such as financing ability, adequacy of the financing plan, and management stability. Business feasibility evaluates the market environment and future profitability, such as market competitiveness, marketing capability, and future growth potential. Technical ability evaluates the factors related to product development and production, such as the company's core technology, production technology, and technology development environment.

Table 3. Description of variables.

No	Features	Definition	Category
1	KA1	Business management ability	Management ability
2	KA2	Management stability	
3	KA3	Internal control	
4	KA4	Labor relations	
5	KA5	Appropriateness of financing plan	
6	KA6	Financing ability	
7	KB10	Credit status	
8	KC17	Business propulsion	
9	KC18	CEO's reliability	
10	KC19	CEO's professionalism	
11	K-Score	Sum of management ability assessment	
12	SA2	Transaction stability	Business feasibility
13	SA9	Sales management	
14	SB11	Sales growth	
15	SB12	Future profitability	
16	SC16	Market entry/expansion possibility	
17	SC18	Market position	
18	SC22	Product competitiveness	
19	SD23	Competitive strength	
20	SD24	Market growth	
21	SD25	Market environment	
22	S-Score	Sum of business feasibility assessment	
23	TA2	Technology development manpower	Technical ability
24	TA3	Technology development environment	
25	TA7	R&D investment	
26	TG24	Process improvement	
27	TG26	Production efficiency	
28	TG28	Facility adequacy	
29	TG32	Quality and process improvement	
30	TG33	Quality innovation	
31	TO67	Technology development performance	
32	TO68	Technical application capacity	
33	TO69	Possibility of technology expansion	
34	TO72	Core technology superiority/discrimination	
35	T-Score	Sum of technical assessment	
36	Total Score	Total score	Other
37	Usage	Usage of loan (working or facility capital)	

3.5. Feature Selection

The process of selecting only variables that play an important role in the prediction of a target variable among the independent variables is called feature selection. Among the independent variables selected after preprocessing, we removed those independent variables not helpful to the prediction model as they might have negative effects on the accuracy of the prediction [44].

In order to select influential variables to the prediction model, we used the gain ratio as a criterion for evaluating the importance of independent variables. Using the backward elimination among wrapper methods, the ideal combination of variables was found and the prediction model was built with the combination.

4. Experimental Results

4.1. Evaluation of Prediction Models

In this study, a model was developed to predict the insolvency of SMEs using data mining techniques. Weka ver. 3.8.3, an open-source software for data mining, was used as a tool for building prediction models. The experiment was conducted by dividing the dataset into the training data and the test data at a ratio of 7:3.

Logistic regression, a decision tree, and an artificial neural network are used as an algorithm for building prediction models, and the ensemble technique, boosting, was applied to these three algorithms, thereby resulting in a total of six predictive models. Boosting initially provides samples in training data with the same weight and increases the misclassified samples' probability of being included in new training data by giving high weights to them by repeatedly building prediction models. Several prediction models are built as a result of boosting, and the prediction results of each model are aggregated using the weighted average. The final prediction is made based on the weighted average.

Table 4 shows the experimental results of our prediction models. The results of insolvency prediction for companies established for fewer than three years versus more than three years show that, among the single algorithms, the decision tree model yielded the highest accuracies of 68.1% and 80.6%, respectively. Similarly, among the ensemble algorithms, the decision tree models achieved the highest accuracies of 69.1% and 82.7%, respectively.

Table 4. Experimental results.

Algorithm		**Accuracy (%)**	
		Fewer Than Three Years	**More Than Three Years**
Single	Logistic Regression	59.3	62.8
	Decision Tree	68.1	80.6
	Artificial Neural Network	61.4	70.1
Ensemble	Boosting (Logistic Regression)	59.3	62.8
	Boosting (Decision Tree)	69.1	82.7
	Boosting (Artificial Neural Network)	63.8	76.3

Thus, in both single and ensemble algorithms, the decision tree model showed the best accuracy, and the boosting model based on the decision tree yielded the highest accuracy both for companies established for fewer than three years and more than three years.

Considering that previous prediction models using financial information of the companies listed on the KRX or KOSDAQ achieved an accuracy of 80% to 90%, our proposed prediction model using only the technological feasibility assessment information showing an accuracy of 82.7% has the possibility of being applicable to the prediction of insolvency of SMEs in the future. Thus, this result implies that our hypothesis denoting that the technical feasibility assessment information could be a substitute for financial information when predicting the insolvency of companies could be reasonably confirmed.

In addition, our proposed model also has an advantage in that it can be applicable to the companies established for fewer than three years, whose financial information is hard to obtain and thus have been much less analyzed in previous research.

4.2. Insolvency Rules of SMEs

With regard to the purpose of the policy funds, the monies should be given to many companies in need. However, the policy funds are raised from taxes, so to ensure the soundness of the policy fund, KOSME should seek ways to reduce loans to companies that are likely to become insolvent or fail to strengthen management after financing [45].

In general, the results of a prediction model built based on a decision tree is easy to interpret as the classification rules are created in easy-to-understand forms [46]. Thus, it needs to derive meaningful rules by analyzing the results of the decision tree model. Based on the derived rules, it is possible to supplement the indicators of technological feasibility assessment and consult companies likely to be insolvent, thereby improving the viability of the companies.

In previous experiments, although the boosting model based on the decision tree showed the highest accuracy, it is difficult to understand the relationships between the rules as the results of several different prediction models are combined to produce a final prediction. Therefore, this study utilized the result from a single decision tree model to derive meaningful rules of insolvency.

As shown in Table 5, for the companies established for fewer than three years, the highest accuracy was achieved by a decision tree model with nine variables. The nine variables include three features in management ability, four in business feasibility, one in technical ability, and one in another ability. The maximum points for these variables were between 4 and 7.

Table 5. Variables used in a decision tree model for companies established for fewer than three years.

Category	Features	Definition	Max Points
Management ability	KA6	Financing ability	5
	KC18	CEO's reliability	7
	KC19	CEO's professionalism	5
Business feasibility	SA2	Transaction stability	6
	SA9	Sales management	4
	SB12	Future profitability	5
	SC16	Market entry/expansion possibility	4
Technical ability	TO72	Core technology superiority/discrimination	7
Other	Usage	Usage of loan (working or facility capital)	-

Figure 3 shows the insolvency rules of the decision tree model for companies established for fewer than three years. In order to give a brief representation of the decision tree and improve the readability of the article, we present only the top six rules with an accuracy of more than 80% among the derived insolvency rules. Table 6 explains the details of these six rules.

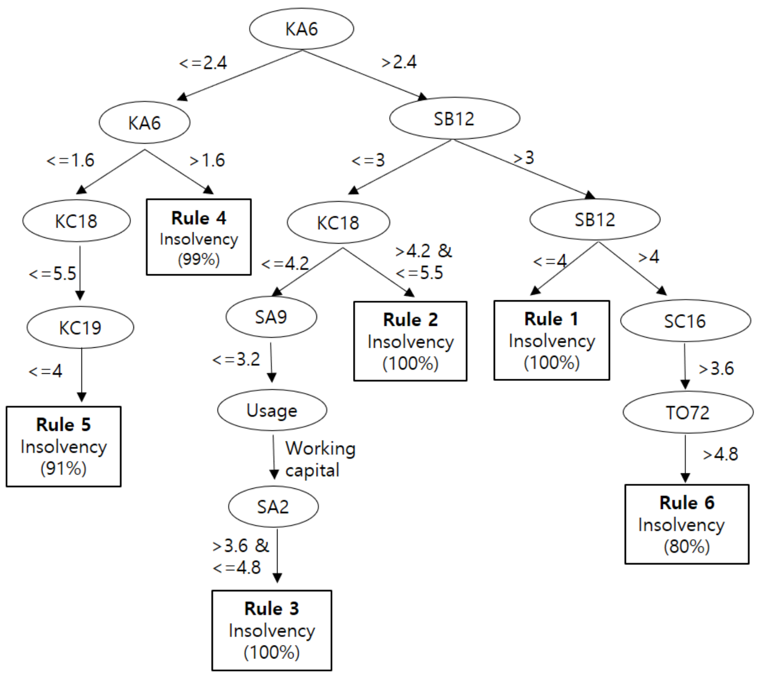

Figure 3. Insolvency rules for companies established for fewer than three years.

Table 6. Description of the insolvency rules of companies established for fewer than three years.

Rule	Description	Target Variable	Prediction		Probability (%)
			True	False	
1	(KA6 > 2.4) and (3 < SB12 <= 4)	Insolvency	77	0	100
2	(KA6 > 2.4), (SB12 <= 3), and (4.2 < KC18 <= 5.5)	Insolvency	32	0	100
3	(KA6 > 2.4), (SB12 <= 3), (KC18 <= 4.2), (SA9 <= 3.2), (3.6 < SA2 <= 4.8), and (Usage = "Working capital")	Insolvency	20	0	100
4	(1.6 < KA6 <= 2.4)	Insolvency	118	1	99
5	(KA6 <= 1.6), (KC18 <= 5.5), and (KC19 <= 4)	Insolvency	20	2	91
6	(KA6 > 2.4), (SB12 > 4), (SC16 > 3.6), and (TO72 > 4.8)	Insolvency	24	6	80

Rule 1 shows that financing ability (KA6) and future profitability (SB12) are the most important variables for companies established for fewer than three years when predicting whether the companies are insolvent or not. After starting up, the company spends a lot of capital on product development, the making of a production base, and the implementation of a business network, which we think explains why financing ability (KA6) had a great impact on the sustainability of companies established for fewer than three years.

Rule 3 shows that when the company receives a low evaluation score on its business plan as well as future profitability (SB12), which evaluates the estimated ratio of operating profit to the net sales based on internal and external environments; the CEO's reliability (KC18), which evaluates the Credit Bureau score; the sales management (SA9), which evaluates the recovery rate of sales; and the transaction stability (SA2), which evaluates the diversification of customers, it is likely to be insolvent. Furthermore, companies that invest their finances in expanding production facilities and purchasing business sites are less insolvent than those that invest their finances in operating capital.

Rule 4 shows that the most insolvent companies were observed when they received an assessment score between 1.6 and 2.4 in the financing ability (KA6) indicator. Financing ability is closely related to the CEO's reliability (KC18) and business propulsion (KC17), and this result implies that the sustainability of a company depends on the ability to utilize the necessary finances in a timely manner.

Rule 5 shows that the insolvency rate is higher if financing ability (KA6), which evaluates the amount of debt and the exchangeability of assets, is less than 1.6; the reliability (KC18), which evaluates the management authority and provisional payment, is less than 5.5; and the industry's experience or the understanding of the market and technology is lower than 4. However, even if financing ability (KA6) is lower than 1.6, if the CEO's reliability (KC18), which evaluates the Credit Bureau (CB) Score, is higher than 5.5, it is predicted to be a healthy company, implying that the CEO assessment is also important. SMEs confronted with a higher level of uncertainty, but whose CEOs are closely involved in different networks, make greater use of sophisticated management [47].

This study suggests strategies to predict whether a company established for fewer than three years is insolvent or not using a technological feasibility assessment as follows: Among the assessment indicators, financing ability (KA6), the CEO's reliability (KC18), and future profitability (SB12) were identified as important variables affecting insolvency. We think the reason is that companies in their early stages need a lot of capital for the opening costs, such as start-up costs and company establishment costs; research and development costs such as commercialization costs, labor costs, and production facilities; and costs related to commercialization necessary for production and sales. However, the maximum points of each indicator are large—in descending order, core technology superiority/discrimination (TO72), the CEO's reliability (KC18), and transaction stability (SA2). It seems that the impact of the size of the maximum points on insolvency is not significant.

Next, as shown in Table 7, companies established for more than three years achieved the highest accuracy when building a decision tree model using seven variables, including four variables related to management ability and three variables related to business feasibility. The maximum points allocated to the variables ranged from 3 to 5, and the highest 5-point variables included credit status (KB10), which evaluates finance procurement rates and whether there is a delinquency or not, and asset procurement ability, which evaluates the current level of debt and whether there is unused debt or not.

Table 7. Variables used in a decision tree model for companies established for more than three years.

Category	Features	Definition	Max Points
Management ability	KA3	Internal control	4
	KA6	Financing ability	5
	KB10	Credit status	5
	KC17	Business propulsion	3
Business feasibility	SC18	Market position	4
	SD23	Competitive strength	3
	SD25	Market environment	3

Figure 4 shows the insolvency rules of the decision tree model for companies established for more than three years. In order to give a brief representation of the decision tree and improve the readability of the article, we present only the top six rules with an accuracy of more than 80% among the derived insolvency rules. Table 8 explains the details of these six rules.

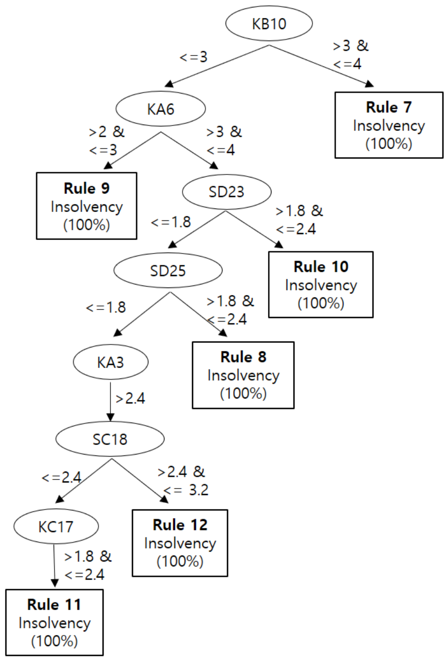

Figure 4. Insolvency rules for companies established for more than three years.

Table 8. Description of the insolvency rules of companies established for more than three years.

Rule	Description	Target Variable	Predict		Probability (%)
			True	False	
7	(3 < KB10 <= 4)	Insolvency	250	0	100
8	(KB10 <= 3), (3 < KA6 <= 4), (SD23 <= 1.8), and (1.8 < SD25 <= 2.4)	Insolvency	232	0	100
9	(KB10 <= 3) and (2 < KA6 <= 3)	Insolvency	185	0	100
10	(KB10 <= 3), (3 < KA6 <= 4), and (1.8 < SD23 <= 2.4)	Insolvency	130	0	100
11	(KB10 <= 3), (3 < KA6 <= 4), (SD23 <= 1.8), (SD25 <= 1.8), (KA3 > 2.4), (SC18 <= 2.4), and (1.8 < KC17 <= 2.4)	Insolvency	76	0	100
12	(KB10 <= 3), (3 < KA6 <= 4), (SD23 <= 1.8), (SD25 <= 1.8), (KA3 > 2.4), and (2.4 < SC18 <= 3.2)	Insolvency	54	0	100

Rule 7 shows that the credit status (KB10) assessment mainly evaluates the average interest rate on financing, whether there are arrears or not, and whether there is infringement of real estate rights or not, and insolvency occurs mainly among companies that received fewer than 4 points on these indicators. On the other hand, a company that received 5 points on the credit status (KB10) assessment indicator has a stable business foundation after its establishment. In other words, it can be seen that the company obtained an excellent rating from the credit rating agency and stably managed its credit status through the appropriate finance limits and preferential interest rates from financial firms. Thus, credit status (KB10) appeared to be an important assessment indicator affecting insolvency.

It turns out that companies are likely to be insolvent when they receive fewer than 4 points on the credit status (KB10) assessment, especially when receiving 3 to 4 points. However, if companies receive fewer than 3 points on credit status (KB10) and more than 4 points on financing ability (KA6), they are classified as healthy companies. This implies that financing ability is one of the influential determinants of insolvency.

Rule 8 shows that companies are likely to be insolvent according to changes in policies or regulations when they receive fewer than 1.8 points on competitive strength (SD23) of the industry, 3 to 4 points (i.e., above average capacity) on financing ability (KA6), and fewer than 3 points on credit status (KB10).

Rule 9 shows that companies are likely to be insolvent when they receive fewer than 3 points on credit status (KB10) and 2 to 3 points on financing ability (KA6).

Rule 10 shows that companies are likely to be insolvent when they receive 1.8 to 2.4 points on competitive strength (SD23), which evaluates whether they are monopolistic or fully competitive, whether there are entry barriers or not, and whether there is the possibility of the emergence of substitutes. However, if companies receive more than 2.4 points on competitive strength (SD23), they are likely to be healthy companies even if they receive 3 to 4 points on financing ability (KA6). This implies that competitive strength is also an influential determinant of insolvency.

Rule 11 shows that the business propulsion (KC17) assessment evaluates organizational composition, leadership, and crisis response capability, and the possibility of insolvency decreases when companies have high capabilities in market position (SC18) or business propulsion (KC17).

Rule 12 shows that companies are likely to be insolvent when they receive fewer than 3 points on credit status and fewer than 1.8 points on competition strength (SD23) and market environment (SD25).

This study suggests strategies to predict whether a company established for more than three years is insolvent or not using the following technological feasibility assessment. Among the assessment indicators, credit status (KB10), financing ability (KA6), and competitive strength (SD23) were identified as important variables affecting insolvency. Thus, financing ability, which is needed for companies in a growth phase, such as the ability to manage finances and financing capacity, entry barriers, and competition level of the industry, significantly affects insolvency.

Core technology superiority/discrimination (TO72, 6 points), product competitiveness (SC22, 6 points), and transaction stability (SA2, 6 points), which have relatively high scores, did not appear in the derived insolvency rules, indicating that their abilities to predict insolvency are not significant.

Based on the results of our experiments, in the case of SMEs established for fewer than three years, KOSME can upgrade the prediction performance of the current assessment system by developing detailed sub-assessment items of financing ability, the CEO's reliability, and future profitability. In SMEs established for more than three years, KOSME can make a better assessment system by developing detailed sub-items of credit status, financing ability, and competitive strength and reducing scores with less impact on predictions, such as core technology superiority/discrimination, product competitiveness, and transaction stability assessment items.

5. Conclusions

This study attempted to predict the insolvency of companies using the technological feasibility assessment information of SMEs established for fewer than three years versus more than three years. Technological feasibility assessment information from 4358 manufacturing companies that received a financial assessment in 2014 was collected and used to build prediction models. We defined company insolvency as the annulment of contracts, comparing the experimental results of companies established for fewer than three years versus more than three years. Using the backward elimination feature selection method, we identified a combination of nine variables from companies established for fewer than three years and a combination of seven variables from those established for more than three years, showing the highest accuracy.

This study's results have the following implications: First, we developed a decision tree model showing the highest accuracy of 68.1% for companies established for fewer than three years and 80.6% for those established for more than three years. We improved the decision tree model using boosting, showing the highest accuracy of 69.1% and 82.7%. Second, we identified important variables affecting the prediction of insolvency in each dataset. Among the assessment indicators, management ability and business feasibility were identified as important variables that determine the insolvency of a company. In particular, financing ability, the CEO's reliability, and future profitability were identified as important variables for companies established for fewer than three years. On the other hand, credit status, financing ability, and competitive strength were found to be important variables for companies established for more than three years. Third, the results of this study contribute to improving the soundness of the policy funds raised by the government budget and bond issuance and to converting companies predicted to be insolvent into healthy companies through financial management and comprehensive support.

This study proposed insolvency prediction models for SMEs in the field of policy funds, which has different characteristics from the general financial market, without using financial statements. Based on the results of this study, the government's budget officer can use the indicators of the prediction models for ensuring the soundness of policy funds and predicting the insolvency of companies in advance. A company's manager can prevent the company's insolvency by managing the critical indicators regularly. Although this study focused on the business environment in South Korea, it is possible to use the research methodology from this study to determine company insolvency in other countries because the main factors for evaluating companies are similar.

This study has limitations in that there seems to be a special situation in domestic and international economic conditions during the five-year repayment period, as this study was conducted on companies financed in 2014. Also, to address the small number of insolvent companies (i.e., the data imbalance problem), this study applied the over-sampling method to the dataset. However, if sufficient data are secured and analyzed by extending the evaluation period in the future, the accuracy could be increased.

As this research used limited technological feasibility assessment information, future research could explore how the accuracy of the prediction model increases when technological feasibility assessment information is combined with commonly used financial information. Future research could also divide the manufacturing sector according to technological characteristics or the revised Pavitt taxonomy proposed in Bogliacino and Pianta [48]. Finally, bankruptcy of SMEs has recently been increasing in many countries due to the COVID-19 pandemic, so future research should examined the relationship between the role of government policy funds and the rate of insolvency of SMEs in the COVID-19 environment.

Author Contributions: Conceptualization, S.L. and D.Y.; methodology, S.L.; validation, S.L., K.C. and D.Y.; formal analysis, S.L. and K.C.; writing—original draft preparation, S.L.; writing—review and editing, K.C. and D.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A5A8016415).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Linna, T. Insolvency proceedings from a sustainability perspective. *Int. Insolv. Rev.* **2019**, *28*, 210–232.
2. Beaver, W.H. Financial ratios as predictors of failure. *J. Account. Res.* **1966**, *4*, 71–111.
3. Altman, E.I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J. Financ.* **1968**, *23*, 589–609.
4. Kim, K.H. *A Detailed Analysis of Start-Up Support Companies*; Korea Institute of Startup & Entrepreneurship Development: Daejeon, Korea, 2019; pp. 67–95.
5. Giacosa, E.; Mazzoleni, A. A decision model for the suitable financing for small and medium enterprises. *Int. J. Manag. Financ. Account.* **2016**, *8*, 39–74.
6. Cho, Y.S. *Major Issues and Policy Tasks of SME Policy Fund*; ISSUE PAPER 2008-232; KIET: Seoul, Korea, 2008.
7. Kim, H.O.; Yoon, B.S. The Current State of Policy Fund Support: Differentiation Measures of Small and Medium Business Corporation (SBC). *J. Decis. Sci.* **2015**, *23*, 113–133.
8. Kim, J.K. *Improvement of Financing in Policy Fund*; Korea Development Institute: Seoul, Korea, 1993; pp. 113–176.
9. Lim, H.J. A study on the effect of Macroeconomic variables on credit guarantee performance. *J. SME Financ.* **2009**, *Autumn*, 39–67.
10. Lee, J.S.; Kim, J.J. A Study on the Effective Combining Technology and Credit Appraisal Information in the Innovation Financing Market. *J. Digit. Converg.* **2017**, *15*, 199–208.
11. Lim, H.J. Firm Characteristics and Default Predictability: Relationship-Banking, Age and Size. *Korean Econ. Anal.* **2016**, *22*, 81–142.
12. Kim, T.H.; Han, B.H. Association between Technology Evaluation Grades and Financial Performance for Small and Medium-Sized Enterprises. *Korean J. Bus. Adm.* **2009**, *22*, 2789–2808.
13. Kolte, A.; Capasso, A.; Rossi, M. Critical analysis of failure of Kingfisher Airlines. *Int. J. Manag. Financ. Account.* **2018**, *10*, 391–409.
14. Ohlson, J.A. Financial ratios and the probabilistic prediction of bankruptcy. *J. Account. Res.* **1980**, *18*, 109–131.
15. Altman, E.I.; Sabato, G. Modelling credit risk for SMEs: Evidence from the U.S. market. *Abacus* **2007**, *43*, 332–357.
16. Podviezko, A.; Kurschus, R.; Lapinskiene, G. Eliciting weights of significance of criteria for a monitoring model of performance of SMEs for successful insolvency administrator's intervention. *Sustainability* **2019**, *11*, 5667.
17. Meyer, P.A.; Pifer, H.W. Prediction of bank failures. *J. Financ.* **1970**, *25*, 853–868.
18. Dimitras, A.I.; Zanakis, S.H.; Zopounidis, C. A survey of business failures with an emphasis on prediction methods and industrial applications. *Eur. J. Oper. Res.* **1996**, *90*, 487–513.
19. Abbasi, A.; Albrecht, C.; Vance, A.; Hansen, J. Metafraud: A meta-learning framework for detecting financial fraud. *MIS Q.* **2012**, *36*, 1293–1327.
20. Chuang, C.L. Application of hybrid case-based reasoning for enhanced performance in bankruptcy prediction. *Inf. Sci.* **2013**, *236*, 174–185.
21. Kim, R.H.; Yoo, D.H.; Kim, G.W. Development of Prediction Model of Financial Distress and Improvement of Prediction Performance Using Data Mining Techniques. *Inf. Syst. Rev.* **2016**, *18*, 173–198.
22. Bae, J.K. An integrated approach to predict company bankruptcy with voting algorithms and neural networks. *J. Corp. Innov.* **2010**, *3*, 79–101.
23. Wang, G.; Ma, J.; Yang, S. An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Syst. Appl.* **2014**, *41*, 2353–2361.
24. Choi, J.W.; Han, H.S.; Lee, M.Y.; Ahn, J.M. The Prediction of Company Bankruptcy Using Text-mining Methodology. *Product. Rev.* **2015**, *29*, 201–228.
25. Lee, J.S.; Han, J.H. Usability test of non-financial information in bankruptcy prediction using artificial neural network—The case of small and medium-sized firms. *J. Intell. Inf. Syst.* **1995**, *1*, 123–134.
26. Cultrera, L.; Brédart, X. Bankruptcy prediction: The case of Belgian SMEs. *Rev. Account. Financ.* **2016**, *15*, 101–119.
27. Lugovskaya, L. Predicting default of Russian SMEs on the basis of financial and non-financial variables. *J. Financ. Serv. Mark.* **2010**, *14*, 301–313.
28. Pervan, I.; Kuvek, T. The relative importance of financial ratios and nonfinancial variables in predicting of insolvency. *Croat. Oper. Res. Rev.* **2013**, *4*, 187–197.

29. Shin, D.R. A Study on the Usefulness of Productivity Indicators in Company Financial Distress Prediction. *Product. Rev.* **2006**, *20*, 1–24.
30. Jo, N.O.; Shin, K.S. Bankruptcy Prediction Modeling Using Qualitative Information Based on Big Data Analytics. *J. Intell. Inf. Syst.* **2016**, *22*, 33–56.
31. Bărbuță-Mișu, N.; Madaleno, M. Assessment of bankruptcy risk of large companies: European countries evolution analysis. *J. Risk Financ. Manag.* **2020**, *13*, 58.
32. Pompe, P.P.; Bilderbeek, J. The prediction of bankruptcy of small-and medium-sized industrial firms. *J. Bus. Ventur.* **2005**, *20*, 847–868.
33. Makropoulos, A.; Weir, C.; Zhang, X. An analysis of the determinants of failure processes in UK SMEs. *J. Small Bus. Enterp. Dev.* **2020**, *27*, 405–426.
34. Mayr, S.; Mitter, C.; Kücher, A.; Duller, C. Entrepreneur characteristics and differences in reasons for business failure: Evidence from bankrupt Austrian SMEs. *J. Small Bus. Entrep.* **2020**, in press.
35. Akbar, M.; Akbar, A.; Maresova, P.; Yang, M.; Arshad, H.M. Unraveling the Bankruptcy Risk-Return Paradox across the Corporate Life Cycle. *Sustainability* **2020**, *12*, 3547.
36. Sánchez-Lasheras, F.; de Andrés, J.; Lorca, P.; de Cos Juez, F.J. A hybrid device for the solution of sampling bias problems in the forecasting of firms' bankruptcy. *Expert Syst. Appl.* **2012**, *39*, 7512–7523.
37. Ok, J.K.; Kim, K.J. Integrated Company Bankruptcy Prediction Model Using Genetic Algorithms. *J. Intell. Inf. Syst.* **2009**, *15*, 99–120.
38. Svabova, L.; Michalkova, L.; Durica, M.; Nica, E. Business Failure Prediction for Slovak Small and Medium-Sized Companies. *Sustainability* **2020**, *12*, 4572.
39. Ahmeti, L.; Zubanovic, A. The Predictive Power of Financial Ratios on Bankruptcy: A Quantitative Study of Non-Listed Limited Liability SMEs Companies in Sweden. Master's Thesis, Jönköping University, Jönköping, Sweden, 30 May 2020.
40. Kim, M.J. Ensemble Learning for Solving Data Imbalance in Bankruptcy Prediction. *J. Intell. Inf. Syst.* **2009**, *15*, 1–15.
41. Zhou, B.; Zhang, X.; Zhang, S.; Li, Z.; Liu, X. Analysis of Factors Affecting Real-Time Ridesharing Vehicle Crash Severity. *Sustainability* **2019**, *11*, 3334.
42. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357.
43. He, H.; Garcia, E.A. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284.
44. Dash, M.; Liu, H. Feature selection for classification. *Intell. Data Anal.* **1997**, *1*, 131–156.
45. Wi, P.R. Changes in SBC's SME Policy Fund Support and Its Implications (2001–2012). *Econ. Reform Rep.* **2014**, *1*, 1–29.
46. Witten, I.H.; Frank, E. Data mining: Practical machine learning tools and techniques with Java implementations. *ACM Sigmod Rec.* **2002**, *31*, 76–77.
47. Kalkhouran, A.A.; Nedaei, B.H.; Rasid, S.Z.A. An exploratory investigation of an integrated model of costing practices in small and medium-sized enterprises. *Int. J. Manag. Financ. Account.* **2017**, *9*, 338–360.
48. Bogliacino, F.; Pianta, M. The Pavitt Taxonomy, revisited: Patterns of innovation in manufacturing and services. *Econ. Polit.* **2016**, *33*, 153–180.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.