Iranian Journal of Management Studies (IJMS) 2021, 14(4): 817-834
RESEARCH PAPER

Identification of the Most Critical Factors in Bankruptcy Prediction and Credit Classification of Companies

Gholamreza Jandaghi[1], Alireza Saranj[2*], Reza Rajaei[3], Ahmadreza Ghasemi[4], Reza Tehrani[5]

1. Professor, Faculty of Management and Accounting, College of Farabi, University of Tehran, Qom, Iran
2. Assistant Professor, Department of Finance and Accounting, Faculty of Management and Accounting, College of Farabi, University of Tehran, Qom, Iran
3. PhD in Financial Management, Faculty of Management and Accounting, College of Farabi, University of Tehran, Qom, Iran
4. Assistant Professor, Department of Industrial Management, Faculty of Management and Accounting, College of Farabi, University of Tehran, Qom, Iran
5. Professor, Department of Financial Management, Faculty of Management, University of Tehran, Tehran, Iran

(Received: July 14, 2019; Revised: May 22, 2021; Accepted: May 25, 2021)

Abstract
Banks and financial institutions strive to develop and improve their credit risk evaluation methods to reduce financial loss resulting from borrowers' financial default. Although in previous studies, many variables obtained from financial statements – such as financial ratios – have been used as the input to the bankruptcy prediction process, seldom a machine learning method based on computing intelligence has been applied to select the most critical of them. In this research, the data from companies that are were listed in Tehran's Stock Exchange and OTC market during 26 years since 1992 to 2017 has been investigated, with 218 companies selected as the study sample. The ant colony optimization algorithm with k-nearest neighbor has been used to feature the selection and classification of the companies. In this study, the problem of the imbalanced dataset has been solved with the under-sampling technique. The results have shown that variables such as EBIT to total sales, equity ratio, current ratio, cash ratio, and debt ratio are the most effective factors in predicting the health status of companies. The accuracy of final research model is estimated that the bankruptcy prediction ranges between 75.5% to 78.7% for the training and testing sample.

Keywords: Credit risk, Probability of default (PD), Bankruptcy prediction (BP), K-nearest neighbor (KNN), Ant colony algorithm, Imbalanced dataset

1. Introduction

Bankruptcy prediction is a topic that has been studied extensively due to its importance for the banking sector (Tobback et al., 2017). The 2008-2010 financial crisis (FC) showed the vulnerability of firms due to the complicated relationship they have in business and economy, which affected the firm's financial health difficulties that in some cases led to bankruptcy (Boratyńska & Grzegorzewska, 2018; Marcinkevičius & Kanapickienė, 2014; Veganzones & Eric Séverin, 2018; Wang et al., 2017; Zoričák et al., 2020). Although many studies have been done in the corporation bankruptcy field, the FC increased the importance of credit risk (CR) (Barboza et al., 2017). Due to the global pandemic of Corona Virus Disease (COVID-19), economic activities have been disrupted in many companies (Abdullah & Achsani, 2020) and

[*] Corresponding Author Email: alisaranj@ut.ac.ir

the stability of economic is in high risk. According to Kiss and Österholm (2020), this instability is near the 2008-2010 financial crisis risk level. The negative consequences of FC have a deep influence on companies' shareholders (Chen et al., 2019; Qu et al., 2019, Veganzones & Eric Séverin, 2018; Uthayakumar et al, 2017), employees (Chen et al., 2019; Veganzones & Eric Séverin, 2018), customers (Chen et al., 2019; Qu et al., 2019; Veganzones & Eric Séverin, 2018), investors (Qu et al., 2019) partners (Qu et al., 2019; Uthayakumar et al., 2017), managers (Chen et al., 2019; Veganzones & Eric Séverin, 2018; Uthayakumar et al., 2017) and the economy (Liu et al., 2019; Mai et al., 2019) in large scale. Therefore, company bankruptcy predication (BP) has increasingly gained attention for banks and financial institutions (Wang et al., 2017). Due to the consequences of FC, it has proven that CR is one of most important factors in bank and financial institutions that can affect the survival of them (Djebali & Zaghdoudi, 2020), since the importance of banks in economy in terms of their vital role in financing is remarkable (Djebali & Zaghdoudi, 2020). Because of the importance of CR and BP on their stability, CR and BP study fields have long been valuable issues to practitioners and academic researchers (Mai et al., 2019; Qu et al., 2019; Tobback et al., 2017; Veganzones & Eric Séverin, 2018). An appropriate framework for CR and BP analysis can prevent the negative effect of a firm's bankruptcy (Uthayakumar et al., 2017). Providing a BP analyzing tool for screening the firms is among the main priorities of banks, and having an analytical and trustworthy early warning system that can accurately predict firm bankruptcy is required for them (Antunes et al., 2017, Wang et al., 2017), as it empowers them to recognize the signs of bankruptcy at an early stage (Chou et al., 2017; Hosaka, 2019). Thus, banks strive to develop their prediction system toward managing the CR, financial distress, and firm's bankruptcy (García et al., 2019). Like the stock market, investors that are using information provided by credit rating agencies or banks also use their own credit ratings to assess borrowers' credit (Grunert et al., 2005).

Since 2004 and following the global implementation of Basel Committee Rules on Banking Supervision and the Basel II Agreement, banks have been encouraged to provide and improve the framework of their internal risk models. To achieve this, they have to demonstrate that they have the ability to accurately assess their risk according to Basel II requirements and their own internal risk models. Banks with an advanced internal grade rating have more advantages than others because they are allowed to use a lower weighting coefficient for credit risk backing (Sousa et al., 2016), and so they can benefit from less requirement for capital adequacy. Therefore, credit risk assessment is essential for their risk control. A system that can provide the possibility of credit risk assessment should be science-based, capable, and accurate in order to properly measure the borrower's ability to repay the commitments. Such a system has an advantage of early detection of credit risk and optimization of the lender's investment portfolio (Zhang et al., 2016).

Most credit risk assessments (Barboza et al., 2017; García et al., 2019; Liu et al., 2019; Mai et al., 2019; Nedumparambil & Bhandari, 2020; Uthayakumar et al., 2017) are based on credit rating models, and that is why it is called probability of default or bankruptcy models. The result driven out of the models is a number that indicates how likely an entity, person, or company may be bankrupt in future during a specific time. Therefore, these models are considered as the core in the banking industry, in deciding on credit, pricing, and determining the cost of capital. Therefore, credit rating models are important tools in the process of granting credit (Fernandes & Artes, 2016). These models measure the credit risk of a client based on specific variables and economic factors. These variables mostly include financial ratios such as leverage ratios, liquidity ratios, operational ratios, profitability ratios, and solvency ratios, as well as other variables that are derived from the firm's annual report such as balance sheets and profit-and-loss statements that will be discussed in more detail below.

Choosing a classification algorithm and the input variables are two of the major tasks of BP (Volkov et al., 2017; Chou et al., 2017). Although there has been many research on BP field, many of them have used pre-selected variables of previous studies (Barboza et al., 2017; Son et al., 2019; Zięba et al., 2016), few of them (Chou et al., 2017; Hosaka, 2019; Kim et al., 2016; Son et al., 2019; Tobback et al., 2017; Volkov et al., 2017) have used a method to feature selection. It means they lost one significant part of BP model development. In this study, ant colony algorithm – as a learning machine method – has been used to feature selection. It was one of the main contributions of this paper in using a learning machine method to feature selection. The learning machine based methods such as neural network, support vector machines (SVM), k-nearest neighbor (KNN), ant colony algorithm, decision trees, and rough sets do not have some limitations of statistical techniques, make fewer assumptions in comparison with them, and are more accurate (Barboza et al., 2017; Chou et al., 2017; García et al., 2019; Hosaka, 2019; Liu et al., 2019; Mai et al., 2019; Son et al., 2019; Zoričák et al., 2020).

In many cases (e.g., García et al., 2019; Hosaka, 2019; Nedumparambil & Bhandari, 2020; Shi & li, 2019; Tobback et al., 2017), researchers have used the ultimate failure (bankruptcy) as an indicator when they distinguish failure from non- failure companies (Qu et al., 2019; Shi & li, 2019). Many studies point out that the bankrupt companies are far less than non-bankrupt companies in their dataset (Chou et al., 2017; Du Jardin, 2017; Kim et al., 2016; Liu et al., 2019; Volkov et al., 2017; Zięba et al., 2016; Zoričák et al., 2020). Some of these studies (Kim et al., 2016; Liu et al., 2019; Zoričák et al., 2020) mention that this problem makes more misclassification errors and they have used a method to deal with this problem. In this study, the under-sampling method has been applied to gain a balanced distribution between classes. The two basic methods of resampling include under-sampling and over-sampling (Kim et al., 2016). In this study, due to the advantages of the under-sampling method, this method is preferred to the over-sampling method. In over-sampling method, there is the probability of created artificial data to be noise data that reduces the classification performance. In under-sampling method, due to the required data volume, the probability of deleting useful data is low and therefore the under-sampling method generally has a better performance (Kim et al., 2016) than over-sampling method.

One requirement of using this technique is to have a huge dataset to choose the suitable numbers of data that should be defined as the machine learning (ML) method input. In this paper, the population was inspected for 26 years to deal with the imbalance data and provide the appropriate number of bankrupt companies.

Another contribution of this paper is broadening the investigated forecasting horizon to 4 years and its ability to provide good mid-term BP. In studies (e.g., Barboza et al., 2017) that BP sample is defined in two years (bankrupt or none bankrupt year known as event year (t) and a year prior it (t-1)), the forecasting horizon is limited to two years. Nonetheless, bankruptcy is a long-term process (Boratyńska & Grzegorzewska, 2018) the symptoms of which can be exploited from financial situation of firms some years prior to bankruptcy. Using an accurate mid-term horizon, BP model is essential for banks because they want to decide whether a mid-term loan should be granted to a firm or not due to the financial situation that it would have in the coming 3 or 4 years.

In this paper, accuracy analysis was also based on type I and type II errors. Few studies (Chou et al., 2017; Uthayakumar et al., 2017) analyze the BP bases on type I and type II errors. Bases on this method in analyzing BP, not only the whole accuracy of model is analyzed but also the accuracy of BP model bases on these two error types is investigated. Therefore, the analyzing output becomes more comprehensive.

The main contribution of this study is summarized as follows:

- Using a learning machine method for feature selection instead of using pre-selected variables of previous studies;
- Broadening the investigated forecasting horizon to 4 years and the ability to provide good mid-term BP;
- Using a resample technique to deal with imbalanced dataset;
- Using computing intelligence for both steps of BP process;
- Analyzing the BP based on type I and type II errors.

The purpose of this study was to identify the most important variables in bankruptcy prediction, classify the companies based on the identified factors, and determine the predictive accuracy of the model. In this research, feature selection was done through ant colony algorithm (ACO). As it was mentioned, there are many studies that point out the superiority of computing intelligence and learning machine method to statistical methods for the development of BP models. Although ACO is a ML method, it is not a black box. It has two advantages that make it suitable for this study. On the one hand, the interpretability of results is more than the black-box method. This advantage is significant in this study because the feature selection has been applied to it and not only the output is consisted of the most critical factors but also it is made up of the weight of them. On the other hand, the comprehensibility of the BP black-box models is less than the none-black-box models (Tobback et al., 2017). Thus, to simultaneously gain the BP model accuracy, comprehensibility, and interpretability, ACO has been used in this study.

According to Chen et al. (2011), the advantage of KNN method compared to the neural network is its simplicity and easiness to interpret and yet being able to predict with high precision. Using machine-learning method in BP indicates that the influence of a firm's financial criteria such as statement ratios on determining the status of it in terms of financial health is not linear. Therefore, using the black-box and non-linear method makes the BP models less comprehensive (Tobback et al., 2017). In addition, according to some studies (e.g., Du Jardin, 2017; Kim et al., 2016; Volkov et al., 2017; Zięba et al., 2016; Zoričák et al., 2020), the dataset in the BP problems is imbalanced. Imbalanced dataset is the property of a dataset in which the sample group numbers are not the same and the dataset is skewed (Kim et al., 2016). As the imbalanced dataset decreases the accuracy of many learning machine methods, the KNN method is suitable for handling the imbalanced dataset (García et al., 2019).

Financial distress occurs when a company is unable to fulfill its financial obligations. Bankruptcy is the final phase of financial distress and occurs when company business is terminated. Nonetheless, BP problems are two-class classifications (García et al., 2019; Hosaka, 2019; Wang et al., 2017). In such problems, samples (in this study companies) are labeled as healthy and unhealthy, and by investigating the financial ratios and other indices in a specific period of time, the likelihood of their belonging to each group is determined (Antunes et al., 2017). Various studies use different driving lines to distinguish two groups. In some studies (e.g., Tobback et al., 2017; Wang et al., 2017), bankruptcy is chosen as the driving line. The status of bankruptcy was also different in studies, and situations such as the declaration of bankruptcy, inability to repay the loans, and the necessity of legal reorganization used to classify the companies as bankrupt or non-bankrupt (Chou et al., 2017). In this study, the article 141 of the Commercial Code of Iran has been taken as the driving line to distinguish bankrupt and non-bankrupt companies.

Bankruptcy models are usually designed using financial ratios calculated with data from balance sheets and income tax returns. The amount of using ratios depends on their predictive power, availability, and standardization. However, the major weakness of these models is that they are estimated using a small number of variables, and the timeframe for each company is limited to just a year under review and a year before it (Du Jardin, 2016). As a result, the output model would have an error and cannot predict bankruptcy with high accuracy. In order to address this issue, it is necessary to select companies as model inputs for bankrupt and healthy companies whose economic status had been monitored not only in one year but also in several years prior to the review year with the same experienced conditions in terms of bankruptcy and non-bankruptcy. Therefore, the time horizon for reviewing companies should be more widespread and the number of reviewed companies should increase as much as needed in the statistical community. In this research, one main contribution is to increase bankruptcy prediction accuracy by expanding the analysis time horizon. To this end, the data of companies listed in Tehran Stock Exchange and OTC market have been investigated from 1992 to 2017 for 26 years. Another contribution of this research is to select the features and classifications simultaneously. Although much research has been done in the field of BP, the focus of this research has been on improving the method (Kim et al., 2016; Liang & He, 2020) of BP classification (Mai et al., 2019; Qu et al., 2019; Shi & li, 2019; Volkov et al., 2017). A very small number of articles (Bayat et al., 2019; Hosaka, 2019) in the field of BP have addressed the systematic identification of factors affecting BP. In most of these studies (e.g., Barboza et al, 2017; Zięba et al., 2016), the previous literature is used to identify some variables as the more fundamental variables and to conduct categorization based on them. In this study, feature selection was performed using ACO, which is a highly accurate and effective method for selecting features, and classification is done by k-nearest neighbor method using the data of selected variables that were the outcome of a machine learning method.

In the following section, the literature and background of research are discussed and the research methodology, which includes the methods of ant colony and k- nearest neighbor, is introduced. In the next section, an analytical and statistical analysis of data is presented and the results of the research and five selected ratios are reported, categorization is done with selected variables, and their accuracy is measured. Finally, the conclusions derived from the obtained outputs of the research are presented.

In this study, bankruptcy in article 141 of the Commercial Code of Iran has been defined as the driving line to distinguish bankrupt and non-bankrupt companies. According to this article, if more than half of the company's capital is wasted by losses, the company has to decide to stop the business or take the required legal action.

2. Theoretical Foundations and Research Background

2.1. Theoretical Foundations and Research

Measuring the credit risk of borrowers is complex and difficult. Banks and financial centers are trying to secure the company's bankruptcy damages by managing contracts and contracting efficiently. The necessity of considering credit risk assessment as a key factor and increasing the use of computer technology while developing new risk and statistical models lead to the improvement of available information quality for decision-making. These enhancements have become even more important with the possibility of increasing ranking accuracy in Basel II content (Zhang et al., 2016). Many mathematical and statistical models have been proposed to identify the probability of default for companies and contracts. These models are referred to as credit rating models that often measure the probability of default risk by some external indicators (Fernandes & Artes, 2016). In the credit rating method, each customer is given a credit according to his credit profile. This means that a customer with a low credit rating has a higher risk of default and non-fulfillment of obligations. These privileges help institutions make more informed decisions in the decision-making process to provide companies with facilities, and therefore using these methods has been greatly welcomed by banks and financial centers (Sousa et al., 2016).

In the field of financial analysis, credit risk assessment in terms of historical record is categorized as a very important and difficult method. However, by utilizing the knowledge of information technology, researchers were able to analyze corporate information and credit status to model credit risk assessments. This way, their models' outputs became much better than the old methods of credit risk assessment that was performed by banking professionals, since it included many advantages such as increased objectivity and reliability and reduced costs and labor in assessing credit risk (Hayashi, 2016). In this regard, different models were used, each one suggesting a kind of learning algorithm. However, the number of these algorithms is high and various algorithms have been proposed (Sousa et al., 2016) to deal with real-world issues. Some models were introduced and suggested by combining two or more learning algorithms.

2.2. Ant Colony Algorithm

The ant algorithm method that previously investigated by Chen et al. (2010), Kanan and Faez (2008), and Tabakhi et al. (2014) in feature selection, inspired by the behavior of real ants, suggests an artificial colony of ants' algorithm in order to solve hybrid optimization problems. The ant colony algorithm is created by imitating the behavior of ants and using simulated ants. In nature, ants roam for food, and their chosen paths are random at first. These ants leave behind a path called a pheromone. Once an ant reaches a food source, it takes food as much as it can and returns to the nest. The return of such an ant to the nest strengthens the frontier in its path. By reinforcing this route, other ants are more likely to use this route to access the source. In fact, routing of other ants is done using the positive feedback that they receive, and ultimately, the ants use this feedback to be directed along the same path. In the ant colony algorithm, which is a superstructure-based algorithm, ants collaborate to navigate a vast space. The ants' path has one starting point and one ending point. When an ant finds a better route, it uses that path, too, to return to the nest, which will reinforce the pheromones of that path and cause other ants to choose it. However, the pheromones' paths that the ant does not cross gradually disappears (Chen et al., 2010). This gradual fading prevents rapid convergence of the algorithm.

One of the main contributions of this study is that the two steps in BP have been done with machine learning algorithms. As it mentioned in the introduction part of this article, the variables in the majority of studies in BP are often selected based on the research literature. In some cases, the variables are directly selected based on previous studies, and the researchers use the variables that have been suggested by other researchers (Barboza et al., 2017; Berg, 2007; Gepp & Kumar, 2008; Pompe & Bilderbeek, 2005). In some other cases, the variables are selected based on mathematical and statistical methods (Kim et al., 2016; Nam & Jinn, 2000). Finally, in very rare number of studies (Chen et al., 2019; Chou et al., 2017; Liang et al., 2016; Wang et al., 2014), the researchers use computing intelligence or machine learning methods to this end. Although the feature selection plays a major role in designing the BP model (Chou et al, 2017), using an accurate method on feature selection is a requirement. As computing intelligence (CI) methods are more suitable than mathematical and statistical ones in addressing BP problems (Mai et al., 2019; Son et al., 2019), the ant colony algorithms have been used as a machine learning technique in selecting the features in this study.

The main advantage of ACO than other CI methods such as artificial intelligence, support vector machine, and genetic algorithm is that unlike other methods, ACO is not a black-box method. This makes the possibility of simultaneously using the accuracy and interpretability of results. In the case of feature selection, the coefficient of each feature can be exploited. The capability of processing a large amount of data (Uthayakumar et al., 2017) is another advantage of

ACO for this paper, as there were 34 primitive variables that should be analyzed in 1-, 2-, 3-, and 4-years periods during 26 years and for 218 companies to determine the most effective variables.

To use ant colony method in this study, initial values are given to the ant colony parameters, and then the features are defined as graph nodes. In the next step, some artificial ants are created and each one is placed randomly on the node so that each ant starts with a random path and adjusts its path by gaining feedback from other ants. This way, all ants make a solution. The ant A is randomly placed to the node i, and it will most likely choose its path in the next node j, according to Formula 1 (Chen et al., 2010; Kanan & Faez, 2008; Liu et al., 2016; Tabakhi et al., 2014).

$$P_A(i.j)=\begin{cases}\dfrac{[\tau_{ij}]^{\alpha}[\eta_{ij}]^{\beta}}{\sum_{u\epsilon J_A(i)}[\tau_{iu}]^{\alpha}[\eta_{iu}]^{\beta}} & if\, j\epsilon J_A(i)\\ 0 & otherwise\end{cases} \tag{1}$$

In this formula, η_{ij} shows the exploratory information and $J_A(i)$ is a set of neighboring nodes of node i that has not yet been visited by the ant A. α and β are two parameters that determine the relative importance of pheromone with respect to exploratory information. Exploratory information is also calculated by Formula 2.

$$\eta_{ij}=\frac{\sum_{n=1}^{N}x_{ni}x_{nj}}{\sqrt{\sum_{n=1}^{N}x_{ni}x_{nj}}} \tag{2}$$

N is the total number of educational samples. The next node j is calculated by Formula 3.

$$j=\begin{cases} argmax_{u\in J_A(i)}\{[\tau_{iu}]^{\alpha}[\eta_{iu}]^{\beta}\} & if\, q\le q_0\\ J & if\, q>q_0\end{cases} \tag{3}$$

Since q is an identical random number in the range of zero to one, q_0 is a threshold parameter, and $J\epsilon J_A(i)$ is a node that is randomly added with respect to probabilities. After selecting the next node, a subset of a new feature is obtained and the ant calculates the optimal path. If the stopping criterion is met, the crossing is stopped, and the stopping criterion is achieved when adding a new feature is either not improving the path or producing a very small impact. In the next step, the ant with the least error is chosen as a solution and pheromones are corrected according to its paths. In this case, only those paths will be strengthened that have found by ants that have found the best solution.

2.3. K-Nearest Neighbor's Algorithm

In this study, the k-nearest neighbor's (KNN) algorithm has been used for companies' classification. The KNN is a non-parametric pattern of classification. In the KNN algorithm, a class (level) is assigned according to the most common class among its closest neighbors. First, samples are assigned to different categories based on Formula 4,

$$u_i(x)=\frac{\sum_j^k u_{ij}(1/x-x_j^{2/(m-1)})}{\sum_j^k(1/x-x_j^{2/(m-1)})} \tag{4}$$

in which (i=1, 2, …, C) and (j=1, 2, …, K), C is the number of classes, and k is the number of nearest neighbors. The parameter m is used to determine the weight proportional to the distance and its value is usually chosen between 1 and infinity. $\|x - x_j\|$ is the distance between x and j, the nearest neighbor of xj, whose membership is assigned to Formula 5.

$$u_{ij}(x_k) = \begin{cases} 0.51 + \left(\frac{n_j}{K}\right) * 0.49 & if\, j = i \\ \left(\frac{n_j}{K}\right) * 0.49 & if\, j \neq i \end{cases} \tag{5}$$

The value of nj is the number of neighbors found belonging to the j class. It should be noted that the membership calculated through the above formula should be applied using formulas 8 and 9.

$$\sum_{I=1}^{C} \mu_{ij} = 1.\ \ j = 1.2.\ldots.n \tag{6}$$

$$0 < \sum_{j=1}^{n} u_{ij} < n \qquad u_{ij} \in [0.1] \tag{7}$$

After calculating all membership items for the sample, values are allocated to the class, which according to Formula 8, has the highest membership value.

$$C(x) = argmax_{i=1}^{C}(u_i(x)) \tag{8}$$

2.4. Prior Research

In recent years, much attention has been paid to rating systems in financial and credit markets. Nowadays, the role of rating institutions has become more important in the area of credit risk (Sironi & Resti, 2007), and internal rating systems in banks are recognized as one of the main sources of interest rates for customers.

Credit rating is a review of borrower’s company in terms of its ability to comply with its obligations based on financial and economic data of the borrower as well as its qualitative aspects such as competitive position, product portfolio, management quality, and the economic outlook of the industry that the borrower company relies on.

Most of the literature review (Mai et al., 2019; Son et al., 2019; Veganzones & Eric Séverin, 2018; Zięba et al., 2016; Zoričák et al., 2020) emphasizes the use of financial indicators such as liquidity, profitability, and capital structure in risk assessment. The discriminant analysis was the first method used to develop credit rating systems. Altman used it in 1968 to predict the bankruptcy of institutions. The most popular credit rating methods at the time included logistic regression and linear programming (Crook et al., 2007). Then, researchers became more tended to use artificial intelligence techniques including neural network and decision tree that were taken from statistical learning theory. The support vector machines that were also derived from learning theory were welcomed by the researchers, and this method was used to create credit rating systems for financing customers and for predicting bankruptcy. In a general classification, predictive methods can be found in titles such as neural network, support vector machines (SVM), multiple discriminant analysis (MDA), k-nearest neighbor (KNN), decision tree, case-based reasoning, genetic algorithm, rough sets, and logistic regression (Chou et al, 2017; Du Jardin, 2017; Qu et al., 2019; Veganzones & Eric Séverin, 2018; Zoričák et al, 2020).

In a study that Du Jardin (2015) performed by using the neural network, multiple audit analysis, and logistic regression in order to improve the accuracy of the prediction model of bankruptcy, he assessed the accuracy of these models as 80.8, 80.1, and 80.6%, respectively. Virág and Nyitrai (2014) conducted a review of 156 companies with the aim of comparing the rough sets method with the two methods of neural network and decision vector machines. They estimated the accuracy of support vector machines and rough sets the same, while the neural network method accuracy was estimated to be 89.32%. Tsai et al. (2014), reviewing the data of 690 companies, estimated the accuracy of the support vector machines and decision tree to be 86.37% and the neural network method to be 84.38%. Heo and Yang (2014) surveyed 2762 companies from Korea, and estimated the precision of 73.3, 77.1, 73.1, and 51.3 percent, respectively, for vector machines, neural network, decision tree, and multiple auditory analysis. Gordini (2014) aimed at comparing the method of genetic algorithm with other techniques by evaluating 3,100 companies and came to the result that support vector machines, genetic algorithm, and logistic regression have the precision of 69.5, 71.5, and 66.8 percent, respectively. In Table 1, a summary of the above-mentioned research and other studies in which prediction has been estimated by various methods along with the estimated values and their methods are presented.

Table 1. Summary of Other Research PD Accuracy Measurements

		Measured accuracy						
Researcher	**Year**	**Neural network**	**Support vector machines**	**Logistic regression**	**Decision tree**	**Linear/Multiple discriminant analysis**	**k-nearest neighbor**	**Other methods**
Du Jardin	2017	80%	80%	78%		75%		
Kim et al.	2016	85%						
Barboza et al.	2017	73%	71%	76%	87%	52%		
Zięba, et al.	2016		50%	62%		64%		
García et al.	2019						91%	
Antunes et al.	2017		84%	86%				
Uthayakumar et al.	2017	80%		85%	84%			
Bayat et al.	2019	95%						
Nyitrai & Virág	2019			76%	82%	72%		
Mai et al.	2019		60%	54%	63%			
Veganzones & Eric Séverin	2018	85%	85%	83%	83%	84%		
Son et al.	2019	82%		67%	77%			
Huang et al.	2018	83%						
Jabeur	2017			94%				
Iturriaga & Sanz	2015	93%	89%					
Pirayesh et al.	2017							91% (novel multiple combination method)
Ghazizadeh et al.	2019	78%	86%	67%		63%	57%	
Ghasemi et al.	2018	97%	77%		83%			70% (fuzzy systems)
Nazemi Ardakani & Zare Mehrjerdi	2017					83%		
Nazemi Ardakani et al.	2018				95%			
Du Jardin	2015		89%	81%		80%		
Virág & Nyitrai	2014	88%	89%					89% (rough sets)
Tsai et al.	2014	84%	86%		86%			
Heo & Yang	2014	77%	73%		73%	51%		
Gordini	2014		69%	67%				71% (genetic algorithm)
Tserng et al.	2014			79%				
Wang et al.	2014	76%	80%	74%	76%			
Zhou et al.	2014	76%		74%	51%	72%	61%	
Arieshanti et al.	2013	71%	70%				75%	
Tsai & Hsu	2013	77%	79%	79%				
Xiong et al.	2013		71%					
Zhou et al.	2012	68%	71%	54%	76%	64%	64%	
De Andrés et al.	2012	76%				75%		
Du Jardin & Séverin	2012		81%	82%		81%		
Jeong et al.	2012	81%	79%	76%	76%	73%		73% (case base reasoning)
Shie et al.	2012	76%	82%	73%	78%			
Divsalar et al.	2011	79%		76%				
Chen et al.	2011	80%	77%				79%	83% (genetic algorithm)
Yang et al.	2011	78%	79%					
Yoon & Kwon	2010		74%	70%		70%		73% (case base reasoning)
Kim & Kang	2010							
Cho et al.	2010			72%	66%			

2.5. Feature Selection

The general assumption of BP processes is that the companies' financial status can be exploited from financial statements. The number of variables suggested by researchers in the BP field is high, although there is a consensus about some more significant ones (Chou et al, 2017). In a general classification, financial ratios can be classified into six groups, including activity ratios, profitability, financial structure, liquidity, solvency, and turnover (Du Jardin, 2015; Veganzones & Eric Séverin, 2018). Although the ratios of these groups were widely used from researchers in BP but the financial ratios are not limited to them. Some most frequently used and suggested financial ratios that have been selected as initial variables of this study are shown in Table 2.

Table 2. Initial Research Variables

		Researcher									
Name of the ratio	**Formula**	**Son *et al.*, 2019**	**Veganzones & Eric Séverin, 2018**	**Mai et al., 2019**	**Zoričák et al., 2020**	**Nyitrai & Virág, 2019**	**Antunes et al., 2017**	**Volkov et al., 2017**	**Tobback et al., 2017**	**Wang et al., 2017**	**Kim et al., 2016**
Current ratio	Current assets/current liabilities			**	**	**	**	**	**	**	
Quick ratio	(Current assets-inventories)/ current liabilities	**				**	**	**			
Cash ratio	Cash/current liabilities		**	**	**	**				**	
Debt ratio	Total liabilities/total assets	**	**	**	**	**			**	**	
	Working capital/total assets	**	**	**		**				**	**
Net profit margin	Net profit/sales		**			**	**	**		**	
	EBIT/total sales			**			**	**			**
	Sales/total assets			**	**	**				**	
Return of assets (ROA)	Net profit/total assets				**	**	**		**		**
Accounts receivable turnover	Accounts receivable/ net sales	**			**	**	**	**	**	**	
Asset turnover ratio	Net sales/average total assets				**		**	**			
Current assets ratios	Current assets/total assets		**	**		**			**		
Equity ratio	Equity/total assets			**		**					**
	EBIT/total assets			**						**	
	Gross profit/total sales			**		**	**				
Shareholders' liquidity ratio	Shareholders' funds/non-current liabilities							**		**	
Total capital turnover	Turnover/total assets	**					**				
Inventory turnover	Net sales/inventories				**	**					
Return of Equity	Net Income/shareholders' equity	**			**				**	**	
	Current liabilities/total liabilities			**							

Table 2 ratios constituted the initial variables of this study. The final variables that formed the basis of the companies' classification were exploited from the output of ACO analysis of the companies' initial variables data. Then, the most critical variables in BP have been defined.

3. Methodology of Research

3.1. Research Method

At this stage, the key variables in the field of BP have been identified through library studies and review of local (Fallahpour et al., 2017; Fallahpour et al., 2018; Nazemi Ardakani et al., 2018) and international research literature. Financial data including the balance sheet, income statement, cash flow, and financial ratios of companies that participated in Tehran Stock Exchange or OTC from 1992 to 2017 through Rahavard Novin software were collected. This data is related to 759 companies. Subsequently, a preliminary screening was carried out, and companies and monetary and financial institutions, insurers, and banks were removed from the investigated companies. Then, the sample was consisted of two categories (according to the article 141 of the Commercial Code of Iran), namely the healthy and bankrupt companies. Since, in the present study, the sample of bankrupt companies includes companies that were not subject to bankruptcy in the first three years of the four-year period under review and were subject to bankruptcy in the last year, 108 companies were found meeting these conditions, all of which were selected as the sample companies. The healthy sample also included companies with four years of continuous activity and no eligibility for bankruptcy definition. Regarding the higher number of healthy samples compared to the bankrupt instances, random sampling was done among healthy companies, and 110 companies were selected. Most frequency and suggested variables from research literature are also introduced, and the ant colony algorithm is used to extract the most effective features. Then, the extracted variables were classified in k-nearest neighbor, the classification was done in MATLAB software, and ultimately its accuracy was examined. Figure 1 is a schematic illustration of the present study.

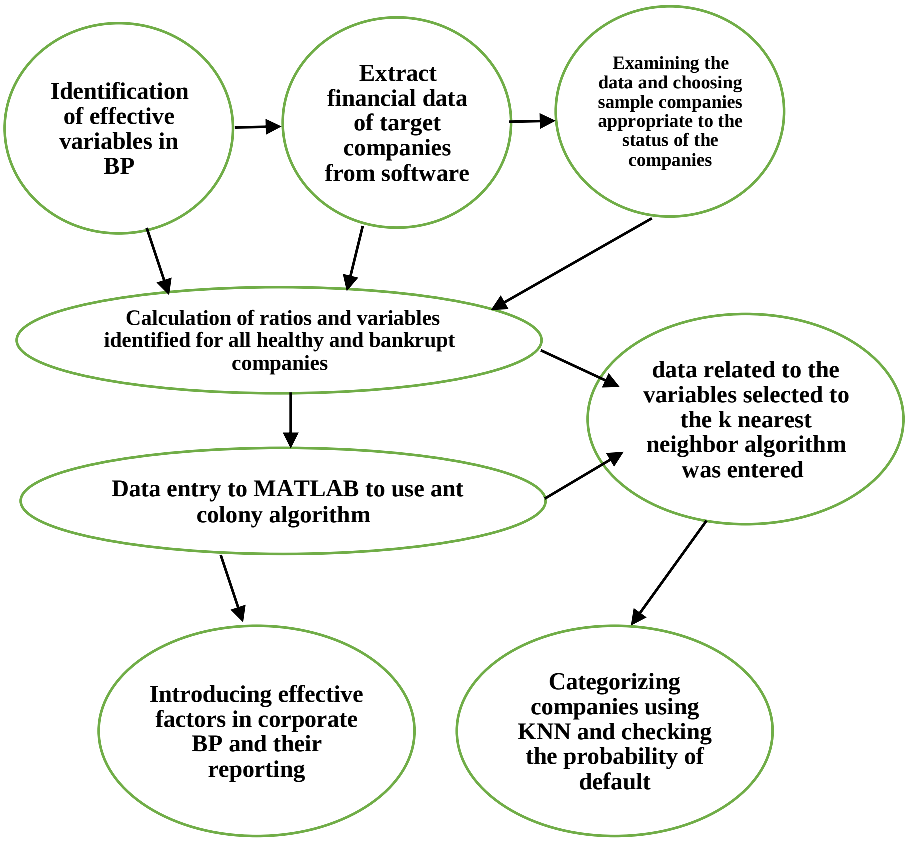

Fig. 1. Schematic Image of Research Methodology

In order to identify the factors affecting companies' bankruptcy, the data of the companies listed in Tehran Stock Exchange and OTC market were collected from 1992 to 2017 for 26 years. This data included financial statements' data containing a balance sheet, income statement, cash flow statements, and financial ratios. Subsequently, these companies were investigated to extract two categories of bankrupt and healthy companies. Considering that 759 companies in the 26-year period were investigated, the number of cases in which companies was healthy for two sequential years, or healthy in one year and bankrupt for next year, is high. In this study, we call the base year t and the years before it t-1, t-2, and t-3 respectively. As mentioned earlier, in this research and in order to increase the precision, companies that have been healthy for three sequential years and then become bankrupt in the fourth year were considered as bankrupt companies and companies that have been healthy in four sequential years were classified as healthy. Therefore, there was a significant amount of data needed to extract companies with these particular circumstances. In total, 108 companies were found in the examined data that had been healthy for three sequential years and were bankrupt in the fourth year. Among the companies that have been healthy for four sequential years, 110 companies were randomly selected as examples of healthy firms.

3.2. Selected Features

In order to select the best features using ACO, 20 variables with their data from 218 companies including 110 healthy firms (being active in four consecutive years and not subjected to bankruptcy) and 108 bankrupt companies (being active in three consecutive years and without being bankrupt but became bankrupt in 4^{th} year) were entered in ACO to choose the best features among 20 variables. The analysis was done using MATLAB software. Five variables including EBIT to total sales, equity ratio, current ratio, cash ratio, and debt ratio were selected as the best features.

4. Results

The classification was done via the application of the k-nearest neighbor (KNN) algorithm to the sample set data. Results showed that five selected variables with high accuracy percentages had the ability to predict bankruptcy. These five variables, categorized by k- nearest neighbor, were able to identify 83 of the 110 healthy companies. Nonetheless, 27 healthy companies were classified as bankrupt firms; i.e., for 27 firms, the first type of error is occurred. The forecast accuracy of healthy firms was 75.5 percent. Among 108 bankrupt companies, the model was able to identify 85 ones correctly, while 23 were classified into the healthy category. Therefore, the measurement accuracy of bankrupt companies was 77.8%. That is to say, the second type of error occurred for 23 companies. In total, the model was able to correctly classify 168 companies out of 218 cases, thus the total output accuracy is 77.7%.

This analysis was also carried out for t-2 and t-3 years. In the year t-2, the model was able to identify 73 companies out of 110 healthy firms, while 37 healthy firms were ranked in the category of bankrupt companies. Therefore, the measurement accuracy of healthy companies was 66.4%. Among 108 bankrupt companies, the model was able to correctly identify 75 cases and classified 33 ones into healthy categories. Therefore, the measurement accuracy of bankrupt firms was 71.3%, and the second type of error occurred for 31 companies. In total, in t-2, the model was able to correctly classify 150 out of 218 companies, and thus the total output accuracy was 68.8%. In t-3, the model was able to identify 73 companies out of 110 healthy firms, and 37 healthy firms were categorized as bankrupt companies. Therefore, the forecast accuracy of healthy firms was 66.4%. Among 108 bankrupt companies, the model was able to correctly identify 75 firms, while it classified 33 of them into healthy categories. Therefore, the measurement accuracy of the bankrupt companies was 69.4%, and the second type of error occurred for 33 companies. In total, in t-3 year, the model was able to correctly classify 148 companies out of 218 ones, and thus the total output accuracy was 67.9%. Results are shown in Table 3.

Table 3. The Accuracy Estimation According to Year's "t", "t-1", "t-2" and "t-3" Data

Status of company		According to year's "t", "t-1"data		According to year's "t", "t-2"data		According to year's "t", "t-3"data	
		Num.	Per.	Num.	Per.	Num.	Per.
Healthy		110		110		110	
	Correct estimate	83	75.5%	73	66.4%	73	66.4%
	Wrong estimate (type I error)	27	24.5%	37	33.6%	37	33.6%
Bankrupt		108		108		108	
	Correct estimate	85	78.7%	77	71.3%	75	69.4%
	Wrong estimate (type II error)	23	21.3%	31	28.7%	33	30.6%
All		218		218		218	
	Correct estimate	167	77.7%	150	68.8%	148	67.9%
	Wrong estimate	50	22.9%	68	31.2%	70	32.1%

5. Discussion and Conclusion

The purpose of this study was to identify the most important variables affecting BP and to offer a classification based on the identified factors. ACO was used to select the feature and k-nearest neighbor for classification. The results of feature selection by ant colony algorithm showed that five variables from initial variables that are shown in Table 2 – including EBIT to total sales, equity ratio, current ratio, cash ratio, and debt ratio – were the best features that were selected. Using data of these features, the classification of sample companies was performed. The results showed that k- nearest neighbor approach could forecast the company's status as bankrupt or not for the target year using pre-year data with an accuracy of 77.7%. The accuracy was measured using data from two to three years before the target year. As the obtained results corroborate, a model's accuracy has usually an inverse relationship with years that the target year is forecasted with them. This means that as the prediction time horizon increases, the estimated accuracy of current data decreases. The outcome accuracy derived from forecasting the company's healthy status using the data of three years indicates that the selected variables are able to accurately forecast the company's credit status in the current year or the following years.

The research results obtained from different classification methods used show that the use of accurate methods can significantly increase the forecast level of banks and financial and credit centers in relation to the company's credit situation in the future, thus reducing losses due to possible default by the borrower. The need for this is especially evident in times of economic prosperity than in times of economic downturn, when companies are mostly struggling to meet their short-term and long-term commitments. The economic crisis of 2008-2010 and the crisis caused by the pandemic of the Coronavirus led to the emergence of multiple bankruptcies of companies that increased the risk of banks and financial and credit centers that came to be directly at risk of default due to the inability of borrowers to pay debts.

The main contributions of this study are summarized as follows,

- Using a learning machine method for feature selection instead of using pre-selected variables of previous studies;
- Broadening the investigated forecasting horizon to 4 years and enhancing the ability to provide a good mid-term BP;
- Using a resampling technique to deal with the imbalanced dataset;
- Using computing intelligence for both steps of the BP process, and
- Analyzing the BP based on type I and type II errors.

The main implications of this study for banks and financial centers are as follows,

- Using accurate models and methods of BP according to their accuracy in predicting the credit status of the borrowing companies;
- Paying attention to the issue of BP due to its importance in the field of banking;
- Applying more sensitivity to the credit risk of companies during the economic recession than during the economic boom, considering that during the economic recession, the possibility of the company's inability and unwillingness to repay debts and fulfill short-term and long-term obligations decreases;
- Exercising more rigorous management of borrowers' collateral commensurate with their credit risk status rather than tightening or easing without scientific and intelligence-based backing, and
- Reducing the taste- and judgment-based view in the field of credit risk, examining companies using credit risk models according to the accuracy of the output of these models, and considering the possibility of errors in human judgments that increase the chances of type I and II errors.

The main limitations of this study are summarized as follows,

- The missing data of the companies and the necessity of calculating all financial ratios led us to drop some cases.
- The lack of data centers and the shortage of classified data in the field of non-financial variables of companies made the researchers to consider only the financial factors while non-financial factors are regarded as important in BP.
- The lack of customer credit information systems in banks made it difficult to compare the results of this study with the situation of bank customers.

In this paper, we investigated the companies that were listed in Tehran Stock Exchange and OTC market, which means they have more reliable financial documents and more stable financial status. Since many of bank customers do not necessarily have these qualifications, it is recommended that future studies focus on BP models that is suitable for SMEs and companies with less financial scale than companies that are listed in stock exchange and OTC market.

References

Abdullah, A., & Achsani, N. A. (2020). Bankruptcy analysis of national airline companies in regional Asia after COVID-19 pandemic. *Jurnal Aplikasi Bisnis dan Manajemen* (JABM), *6*(3), 691-691.

Antunes, F., Ribeiro, B., & Pereira, F. (2017). Probabilistic modeling and visualization for bankruptcy prediction. *Applied Soft Computing, 60,* 831-843.

Arieshanti, I., Purwananto, Y., Ramadhani, A., Nuha, M. U., & Ulinnuha, N. (2013). Comparative study of bankruptcy prediction models. *Telkomnika, 11*(3), 591-596.

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, 405-417.

Bayat, A., Ahmadi, S., & Mohamadi, M. (2019). The Bankruptcy prediction of Tehran stock exchange using firefly algorithm (FA). *Financial Engineering and Portfolio Management*, 9(37), 234-262.

Berg, D. (2007). Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry, 23*(2), 129-143.

Boratyńska, K., & Grzegorzewska, E. (2018). Bankruptcy prediction in the agribusiness sector: Lessons from quantitative and qualitative approaches. *Journal of Business Research*, *89*, 175-181.

Chen, H. L., Yang, B., Wang, G., Liu, J., Xu, X., Wang, S. J., & Liu, D. Y. (2011). A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method. *Knowledge-Based Systems, 24*(8), 1348-1359.

Chen, Y., Miao, D., & Wang, R. (2010). A rough set approach to feature selection based on ant colony optimization. *Pattern Recognition Letters, 31*(3), 226-233.

Chen, Z., Chen, W., & Shi, Y. (2019). Ensemble learning with label proportions for bankruptcy prediction. *Expert Systems with Applications*, 113155.

Cho, S., Hong, H., & Ha, B. C. (2010). A hybrid approach based on the combination of variable selection using decision trees and case-based reasoning using the Mahalanobis distance: For bankruptcy prediction. *Expert Systems with Applications, 37*(4), 3482-3488.

Chou, C. H., Hsieh, S. C., & Qiu, C. J. (2017). Hybrid genetic algorithm and fuzzy clustering for bankruptcy prediction. *Applied Soft Computing*, *56*, 298-316.

Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research, 183*(3), 1447-1465.

de Andrés, J., Landajo, M., & Lorca, P. (2012). Bankruptcy prediction models based on multinorm analysis: An alternative to accounting ratios. *Knowledge-Based Systems, 30*, 67-77.

Divsalar, M., Firouzabadi, A.K., Sadeghi, M., Behrooz, A.H., & Alavi, A.H. (2011). Towards the prediction of business failure via computational intelligence techniques. *Expert Systems, 28*(3), 209-226.

Djebali, N., & Zaghdoudi, K. (2020). Threshold effects of liquidity risk and credit risk on bank stability in the MENA region. *Journal of Policy Modeling, 42*(5), 1049-1063.

Du Jardin, P., & Séverin, E. (2012). Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time. *European Journal of Operational Research*, *221*(2), 378-396.

Du Jardin, P. (2015). Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research, 242*(1), 286-303.

Du Jardin, P. (2016). A two-stage classification technique for bankruptcy prediction. *European Journal of Operational Research*, *254*(1), 236-252.

Du Jardin, P. (2017). Dynamics of firm financial evolution and bankruptcy prediction. *Expert Systems with Applications, 75,* 25-43.

Fallahpour, S., Norouzian Lakvan, E., & Hendijani Zadeh, M. (2017). Use of combined approach of support vector machine and feature selection for financial distress prediction of listed companies in Tehran Stock Exchange Market. *Financial Research Journal*, *19*(1), 139-156.

Fallahpour, S., Raei, R., & Norouzian, E. (2018). Applying combined approach of sequential floating forward selection and support vector machine to predict financial distress of listed companies in Tehran Stock Exchange Market. *Financial Research Journal*, *20*(3), 289-304.

Fernandes, G. B., & Artes, R. (2016). Spatial dependence in credit risk and its improvement in credit scoring. *European Journal of Operational Research, 249*(2), 517-524.

García, V., Marqués, A. I., & Sánchez, J. S. (2019). Exploring the synergetic effects of sample types on the performance of ensembles for credit risk and corporate bankruptcy prediction. *Information Fusion, 47*, 88-101.

Gepp, A., & Kumar, K. (2008). The role of survival analysis in financial distress prediction. *International research journal of finance and economics, 16*(16), 13-34.

Ghasemi, A. A., Seyghalib, M., & Moradi, M. (2018). Prediction of financial distress, using metaheuristic models. *Financial and Credit Activity: Problems of Theory and Practice, 1*(24), 238-249.

Ghazizadeh, N., Abbaszadeh, M. R., Salehi, M., & Jabbari Nooghabi, M. (2019). Comparison of new technologies statistical models and machine learning models to predicting banks failure [paper presentation]. 17th national conference of Iran accounting, Qom, Iran .

Gordini, N. (2014). A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy. *Expert Systems with Applications, 41*(14), 6433-6445.

Grunert, J., Norden, L., & Weber, M. (2005). The role of non-financial factors in internal credit ratings. *Journal of Banking & Finance, 29*(2), 509-531.

Hayashi, Y. (2016). Application of a rule extraction algorithm family based on the Re-RX algorithm to financial credit risk assessment from a Pareto optimal perspective. *Operations Research Perspectives, 3*, 32-42.

Heo, J., & Yang, J. Y. (2014). AdaBoost based bankruptcy forecasting of Korean construction companies. *Applied Soft Computing, 24*, 494-499.

Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications, 117*, 287-299.

Huang, X., Liu, X., & Ren, Y. (2018). Enterprise credit risk evaluation based on neural network algorithm. *Cognitive Systems Research, 52*, 317-324.

Iturriaga, F. J. L., & Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of US commercial banks. *Expert Systems with applications, 42*(6), 2857-2869.

Jabeur, S. B. (2017). Bankruptcy prediction using partial least squares logistic regression. *Journal of Retailing and Consumer Services, 36*, 197-202.

Jeong, C., Min, J. H., & Kim, M. S. (2012). A tuning method for the architecture of neural network models incorporating GAM and GA as applied to bankruptcy prediction. *Expert Systems with Applications, 39*(3), 3650-3658.

Kanan, H. R., & Faez, K. (2008). An improved feature selection method based on ant colony optimization (ACO) evaluated on face recognition system. *Applied Mathematics and Computation, 205*(2), 716-725.

Kim, H. J., Jo, N. O., & Shin, K. S. (2016). Optimization of cluster-based evolutionary undersampling for the artificial neural networks in corporate bankruptcy prediction. *Expert Systems with Applications, 59*, 226-234.

Kim, M. J., & Kang, D. K. (2010). Ensemble with neural networks for bankruptcy prediction. *Expert Systems with Applications, 37*(4), 3373-3379.

Kiss, T., & Österholm, P. (2021). Corona, crisis and conditional heteroscedasticity. *Applied Economics Letters, 28*(9), 755-759.

Liang, K., & He, J. (2020). Analyzing credit risk among Chinese P2P-lending businesses by integrating text-related soft information. *Electronic Commerce Research and Applications, 40*, 100947.

Liu, C., Xie, J., Zhao, Q., Xie, Q., & Liu, C. (2019). Novel evolutionary multi-objective soft subspace clustering algorithm for credit risk assessment. *Expert Systems with Applications, 138*, 112827.

Liu, M., Zhang, F., Ma, Y., Pota, H. R., & Shen, W. (2016). Evacuation path optimization based on quantum ant colony algorithm. *Advanced Engineering Informatics, 30*(3), 259-267.

Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, *274*(2), 743-758.

Marcinkevičius, R., & Kanapickienė, R. (2014). Bankruptcy prediction in the sector of construction in Lithuania. *Procedia-Social and Behavioral Sciences, 156*, 553-557.

Nam, J. H., & Jinn, T. (2000). Bankruptcy prediction: Evidence from Korean listed companies during the IMF crisis. *Journal of International Financial Management & Accounting*, *11*(3), 178-197.

Nazemi Ardakani, M., & Zare Mehrjerdi, V. (2017). Prediction of firms bankruptcy based on industry characteristics. *Quarterly Journal of Asset Management and Financing, 7*(2), 122-139.

Nazemi Ardakani, M., Zare Mehrjerdi, V., & Mohammadi Nodooshan, A. (2018). A firms' bankruptcy prediction model based on selected industries by using decision trees model. *Accounting Research*, *6*(2), 121-138.

Nedumparambil, E., & Bhandari, A. K. (2020). Credit risk–return puzzle: Evidence from India. *Economic Modelling*, *92*, 195-206.

Nyitrai, T., & Virág, M. (2019). The effects of handling outliers on the performance of bankruptcy prediction models. *Socio-Economic Planning Sciences*, *67*, 34-42.

Pirayesh, R., Dadashi Arani, H., Barzegar, M. (2017). Mathematical model design for predicting bankruptcy of companies accepted in the Tehran Stock Exchange. *Financial Engineering and Portfolio Management*, *8*(31), 187-200.

Pompe, P. P., & Bilderbeek, J. (2005). The prediction of bankruptcy of small-and medium-sized industrial firms. *Journal of Business venturing, 20*(6), 847-868.

Qu, Y., Quan, P., Lei, M., & Shi, Y. (2019). Review of bankruptcy prediction using machine learning and deep learning techniques. *Procedia Computer Science, 162*, 895-899.

Shi, Y., & Li, X. (2019). A bibliometric study on intelligent techniques of bankruptcy prediction for corporate firms. *Heliyon*, *5*(12), e02997.

Shie, F. S., Chen, M. Y., & Liu, Y. S. (2012). Prediction of corporate financial distress: An application of the America banking industry. *Neural Computing and Applications, 21*(7), 1687-1696.

Sironi, A., & Resti, A. (2007). *Risk management and shareholders' value in banking: from risk measurement models to capital allocation policies* (Vol. 417). John Wiley & Sons.

Sironi, A., & Resti, A. (2007). Risk management and shareholders' value in banking: From risk measurement models to capital allocation policies (Vol. 417) pp.369-370. *John Wiley & Sons*.

Son, H., Hyun, C., Phan, D., & Hwang, H. J. (2019). Data analytic approach for bankruptcy prediction. *Expert Systems with Applications*, *138*, 112816.

Sousa, M., Gama, J., & Brandão, E. (2016). Dynamic credit score modeling with short-term and long-term memories: The case of Freddie Mac's database. *The Journal of Risk Model Validation, 10*(1), 59–80.

Tabakhi, S., Moradi, P., & Akhlaghian, F. (2014). An unsupervised feature selection algorithm based on ant colony optimization. *Engineering Applications of Artificial Intelligence, 32*, 112-123.

Tobback, E., Bellotti, T., Moeyersoms, J., Stankova, M., & Martens, D. (2017). Bankruptcy prediction for SMEs using relational data. *Decision Support Systems, 102,* 69-81.

Tsai, C. F., & Hsu, Y. F. (2013). A meta-learning framework for bankruptcy prediction. *Journal of Forecasting, 32*(2), 167-179.

Tsai, C. F., Hsu, Y. F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing, 24,* 977-984.

Tserng, H. P., Chen, P. C., Huang, W. H., Lei, M. C., & Tran, Q. H. (2014). Prediction of default probability for construction firms using the logit model. *Journal of Civil Engineering and Management*, *20*(2), 247-255.

Uthayakumar, J., Vengattaraman, T., & Dhavachelvan, P. (2020). Swarm intelligence based classification rule induction (CRI) framework for qualitative and quantitative approach: An application of bankruptcy prediction and credit risk analysis. *Journal of King Saud University-Computer and Information Sciences, 32*(6), 647-657.

Veganzones, D., & Séverin, E. (2018). An investigation of bankruptcy prediction in imbalanced datasets. *Decision Support Systems*, *112*, 111-124.

Virág, M., & Nyitrai, T. (2014). Is there a trade-off between the predictive power and the interpretability of bankruptcy models? The case of the first Hungarian bankruptcy prediction model. *Acta Oeconomica, 64*(4), 419-440.

Volkov, A., Benoit, D. F., & Van den Poel, D. (2017). Incorporating sequential information in bankruptcy prediction with predictors based on Markov for discrimination. *Decision Support Systems*, *98*, 59-68.

Wang, G., Ma, J., & Yang, S. (2014). An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications, 41*(5), 2353-2361.

Wang, M., Chen, H., Li, H., Cai, Z., Zhao, X., Tong, C., Li, J & Xu, X. (2017). Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction. *Engineering Applications of Artificial Intelligence*, *63*, 54-68.

Xiong, T., Wang, S., Mayers, A., & Monga, E. (2013). Personal bankruptcy prediction by mining credit card data. *Expert Systems with Applications, 40*(2), 665-676.

Yang, Z., You, W., & Ji, G. (2011). Using partial least squares and support vector machines for bankruptcy prediction. *Expert Systems with Applications, 38*(7), 8336-8342.

Yoon, J. S., & Kwon, Y. S. (2010). A practical approach to bankruptcy prediction for small businesses: Substituting the unavailable financial data for credit card sales information. *Expert Systems with Applications*, *37*(5), 3624-3629.

Zhang, F., Tadikamalla, P. R., & Shang, J. (2016). Corporate credit-risk evaluation system: Integrating explicit and implicit financial performances. *International Journal of Production Economics, 177*, 77-100.

Zięba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications, 58*, 93-101.

Zhou, L., Lai, K. K., & Yen, J. (2012). Empirical models based on features ranking techniques for corporate financial distress prediction. *Computers & Mathematics with Applications, 64*(8), 2484-2496.

Zhou, L., Lai, K. K., & Yen, J. (2014). Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimisation. *International Journal of Systems Science, 45*(3), 241-253.

Zoričák, M., Gnip, P., Drotár, P., & Gazda, V. (2020). Bankruptcy prediction for small-and medium-sized companies using severely imbalanced datasets. *Economic Modelling, 84*, 165-176.