# Financial Distress Prediction Using Support Vector Machines and Logistic Regression

**Seyyide Doğan, Deniz Koçak, and Murat Atan**

**Abstract** Financial distress and bankruptcies are highly costly and devastating processes for all parts of the economy. Prediction of distress is notable both for the functioning of the general economy and for the firm's partners, investors, and lenders at the micro-level. This study aims to develop an effective prediction model with Support Vector Machine and Logistic Regression Analysis. As the field of the study, 172 firms that are traded in Borsa İstanbul, have been chosen. Besides, two basic prediction methods, LRA was also used as a feature selection method and the results of this model were compared. The empirical results show us, both methods achieve a good prediction model. However, the SVM model in which the feature selection phase is applied shows the best performance.

**Keywords** Financial distress · Support vector machine · Logistic regression analysis

## 1 Introduction

Financial distress, by the simplest definition, is a specific type of financial difficulty that a company faces due to internal or external reasons and tries to overcome. Financial difficulties are the obstacles that the company faces in meeting its obligations. These obstacles are lack of liquidity, lack of owner's equity, failing to pay the debts, and lack of capital (Sun et al., 2014a). Companies face a legally binding bankruptcy if they cannot overcome these obstacles for a long time. Given all these, financial

S. Doğan (✉)
Department of Econometrics, Karamanoğlu Mehmetbey University, Karaman, Turkey
e-mail: dogans@kmu.edu.tr

D. Koçak
Department of Econometrics, Osmaniye Korkut Ata University, Osmaniye, Turkey
e-mail: deniz.kocakerturk@yok.gov.tr

M. Atan
Department of Econometrics, Ankara Hacı Bayram Veli University, Ankara, Turkey
e-mail: murat.atan@hbv.edu.tr

distress can be identified as a long and difficult process that starts with a firm's inability to meet its obligations and extending to bankruptcy.

Classical literature limited the financial failure only to the event of bankruptcy. However, as some authors pointed out, financial failure may not always result in bankruptcy. A company can avoid bankruptcy even in a troubled process by accelerating cash flows through selling assets, downsizing, closure of loss-making transactions (Hashi, 1997). On the other hand, a company can unexpectedly come to the brink of bankruptcy due to unpredictable external shocks such as natural disasters, badly ending cases, global economic and financial crises, even if the company did not face financial difficulties previously (Meyer, 1982). Therefore, a typical commercial distress measurement like bankruptcy cannot refer to financial distress on its own. It is more realistic to assess financial distress as a process, rather than a specific incident, even though it makes it more complex to define and classify exactly. As a process, financial distress corresponds to steps that come sequentially to each other, rather than just one event (Agostini, 2013).

Financial distress can occur in different sizes in companies; while its results can affect an entire economy with a domino effect. The distress of companies can leave states, all stakeholders with whom the firm is connected in the finance and public industries in a difficult situation. Therefore, predicting the distress, by developing a good prediction model will give the company and its stakeholders, creditor institutions an opportunity to decrease the costs that will arise in case of a distress, manage, and monitor the process (Zhou et al., 2015).

*Financial distress prediction* (FDP), which is an important research topic in finance, economy, accounting, and engineering fields, is also called *bankruptcy prediction* and *prediction of company distress.* In general, FDP is the prediction of whether the firm will fail or not based on current financial data of the firms through mathematical, statistical, and artificial intelligence techniques. It is accepted that financial distress often remains under the surface but bankruptcy becomes open and obvious to all upon its declaration, therefore it requires an in-depth analysis (Pindado and Rodrigues, 2005; Doğan, 2020: 13).

In recent years, the academic and industrial interest in this topic has increased because of a growing number of firm bankruptcy with the impact of economic crises. The researchers used classical statistical techniques despite some disadvantages in the first years while they went into the effort of developing early warning models convenient for FDP with the machine learning methods in the recent years. This study used Support Vector Machine, which is a powerful machine learning method. There are many successful FDP studies performed with SVM. This study aimed to contribute to the literature by the selection and parameter optimization phase, whose importance for SVM was recently revealed.

## 2　Theoretical Background

The concept of financial distress emerges as a very important concept in financial research. There are many different solutions for this subject, from univariate ratio analysis to multivariate prediction methods, from traditional statistical methods to artificial intelligence-based machine learning methods, from a single classifier method to hybrid classifier methods designed to combine different classifiers (Sun et al., 2014b; Kumar and Ravi, 2007; Lin et al., 2012). Making financial distress prediction (FDP) through statistical models dates back to the 1960s. The first of those studies was Beaver's (1966) study, which proposed a model with a single variable, and which tried to present the financial distress of an enterprise by dealing with financial ratios individually and thus obtaining a general idea about the financial risk of the enterprise. The study was considered a pioneer study in the finance literature. But in the following years, it was criticized since financial distress or business performance cannot be measured based on a single financial ratio and the prediction capacity would be very low. Following those criticisms, Altman (1968) used statistical methods with multiple variables for the first time through the Z-score model he developed. According to the results of the study, more reliable and consistent findings were obtained by evaluating different financial ratios together with their weights. After Altman's success, some examples such as the multiple-regression analysis introduced by Meyer and Pifer (1970), the logistic regression analysis (LRA) introduced by Ohlson (1980), and the probit model introduced by Zmijewski (1984) were applied in the related field. However, some necessity of traditional methods such as linearity, normality, independent variables of prediction, and the functional form already existing between dependent and independent variables cannot quite be ensured in real-life problems. Today, there are alternative methods, which are less sensitive to the above-mentioned assumptions and which are developed based on artificial intelligence techniques.

Decision Trees (DT) are frequently used in artificial intelligence-based studies carried out on FDP due to their easy understanding and interpretation. Gepp and Kumar (2008), Gepp et al. (2010), and Li et al. (2010) proposed DT, classification and regression trees (CART), C5.0 algorithms for FDP, and showed that they yield better results than multidimensional analysis (MDA). Chen (2011) used the C5.0, CART, and CHAID and LRA methods in his FDP study on businesses registered on the Taiwanese stock exchange. In the findings of the study, it was concluded that the predictive power of decision trees increases even more as the financial distress approaches the year. Genetic programming (GP), which is one of the meta-heuristic methods, was used by Etemadi et al. (2009) for bankruptcy estimation and has been shown to perform better than MDA.

Artificial neural network (ANN) in pattern recognition and classification problems is a highly powerful instrument due to its non-linear non-parametric adaptive learning properties. ANN can very effectively represent and define the non-linear relationship in a data set. ANN was first applied to bankruptcy prediction by Odom and Sharda (1990). They also applied the multiple-variable discriminant analysis (MVDA) to

their sampling of 129 enterprises, 65 of which went bankrupt. As a result, the correct classification rate for MVDA was 74.28%, whereas the rate reached 81.81% for ANN. Many similar studies have emphasized that ANN performs better than statistical methods (Tam, 1991; Tam and Kiang, 1992; Fletcher and Goss, 1993; Zhang et al., 1999; Liang and Wu, 2005).

SVM, developed by Vapnik (1995), has also been of interest to many researchers since they provide considerable results. The most fundamental difference between SVM and ANN is that the structure of SVM is based on structural risk minimization. Because it is aimed to minimize the empirical risk to minimize the training set error in ANN. Conversely, SVM adopts the principle of structural risk minimization, which has been shown to yield better performance than empirical risk, using quadratic programming to predict a single and optimal separator plane in the hidden feature space (Min et al., 2006; Zhongsheng et al., 2007). Fan and Palaniswami (2000), for the first time, applied SVM on three different datasets using the financial ratios suggested by the three models (Altman, 1993; Lincoln, 1984; Ohlson, 1980) best known in the literature. Besides, MDA has tested SVM's success by developing financial failure prediction models with a multi-layer perceptron (MLP) and learning vector quantization (LVQ). Min and Lee (2005) also applied SVM to bankruptcy prediction problems. The results of the study show that when compared to ANN, SVM both gives better results and learning is made possible with a smaller number of training sets. To validate the high classification rate, ANN with backpropagation is compared with multiple-variable discriminant and Logit models, and according to the empirical results, SVM provided better results than all other methods. Shin et al. (2005) compared SVM to ANN to show the effectiveness of SVM, and SVM yielded better empirical results. The study also emphasizes these two important points: first, SVM reaches a better generalization capacity with fewer training sets since it tries to understand the geometric structure of the feature space without reproducing the weights of training samples; second, it makes SVM more useful than ANN, as ANN has certain limitations regarding classification problems. Similarly, Shin et al. (2005) made financial distress predictions for Chinese firms and they compared SVM to the other methods used in the above-mentioned study and reached the same conclusion.

Wu et al. (2007) presented a very comprehensive study in the financial failure estimation study using MDA, logit model, probit model, ANN, SVM methods. In this study, it is aimed to enhance the predictive performance of SVM. For this, researchers have optimized SVM parameters using the Genetic Algorithm (GA). Liang et al. (2016) presented a comprehensive study in which the main classifying method was SVM on 239 successful and 239 failed companies operating in Taiwan Stock Exchange from 1999 to 2009, and SVM inputs were investigated. Machine learning has tested the success of SVM with four methods that have been proven and used in the literature. These methods are k-NN, Naïve Bayes (NB), CART, and MLP. According to the experimental results, SVM has been found as the best prediction model. The reasons why SVM is preferred as a method over other data mining techniques in the present study are that SVM yield equivalent or better results can work with fewer training samples, and has fewer parameters to adjust. For this reason, the main estimation method of the study is determined as SVM. The contribution of the

study to the literature is that it is aimed to try new ways to increase the predictive accuracy rate of SVM. Different processes affect the predictive accuracy of SVM. One of these processes is the determination of the optimal feature set (or variables) that provides quality information to the classifier. The learner may encounter redundant, irrelevant, or interrelated data while understanding the geometric structure of the classifier property space. When too much and unnecessary information is given to the model as input, a lot of time and cost will be spent and even the model's suitability rate will decrease slightly (Piramuthu, 2004; Huang and Wang, 2006). However, it is not an easy way to interpret or exclude unnecessary information. For this reason, it is an important issue to filter large amounts of data and intensify it to provide more information especially in financial failure estimation (Tsai, 2008). In most of the current studies, the financial ratios that provide information were chosen from the financial ratios produced by the prediction models previously made. The classification ability of these models will largely depend on studies in which the selected financial ratios are taken (Wu et al., 2006).

In the first studies in the FDP literature (Beaver, 1966; Altman, 1968), the feature selection process was generally carried out using a qualitative approach such as the popularity of features (financial ratios), good results in past studies, or based on expert opinion. This approach has been replaced by quantitative selection techniques over time. Jo et al. (1997), Atiya (2001), Park and Han (2002), Shin and Lee (2002), Min and Lee (2005), Ding et al. (2008), Chen (2011), Li and Sun (2012) selected the features using statistical methods such as progressive regression, t-test, and correlation matrix, factor, and principal component analysis, which are examined in the filter methods category. Min et al. (2006) and Wu et al. (2007) preferred GAs examined in the wrapper feature selection methods category. In these studies, it was emphasized that the power of the prediction model depends on the selected prediction method and feature set. However, another important situation that increases the estimation performance is to investigate the optimal parameter set. SVM has two important parameters called "C" and "gamma". There are lots of studies that emphasized the parameter optimization improves the performance of SVM (Wu et al., 2007; Shin et al., 2005). But there are a limited number of studies investigating both the optimal parameter pair and the optimal feature set. In this study, both parameter optimization and feature selection methods are used for SVM. The preferred feature selection method in the study is LRA. Despite some limitations, LRA is a multivariate statistical method that is frequently preferred in the studies of financial failure estimation. For this reason, it will be used as an alternative method in the study to test the success of SVM. For parameter optimization, the Grid search technique, which is one of the easy and effective methods, is preferred and this technique is presented in Sect. 3. In the next section, the empirical results are summarized. The final section presents a general summary of the study.

# 3    Proposed Methods for the Prediction Model

This chapter presents the working principle of SVM for a typical two-class classification problem and explains the LRA, which is a multiple-variable statistical technique. For detailed explanations about SVM, please refer to Gunn (1998), Smola and Schölkopf (1997), and Cristianini and Shawe-Taylor (2000).

## 3.1   SVM Classifier

The sample-class labels pair, $(x_i, y_i)$, $i = 1, 2, ..., m$, $x_i \in \Re^n$, and $y_i \in \{+1, -1\}$ which has $p$ number of feature (attributes) and which comprises the training set as the linear hyperplane which will separate S training set to represent the class that output samples represent is formulated as follows:

$$\text{w.x} + \text{b} \tag{1}$$

There can be many linear planes that separate the problem linearly. This can be seen in Fig. 1:

However, it is aimed to find the most suitable separator hyperplane. This hyperplane maximizes the distance between support vectors from different classes, which is called the margin. The distance between $\langle w, x \rangle + b = 0$ separator plane and the

**Fig. 1**  Linear classification
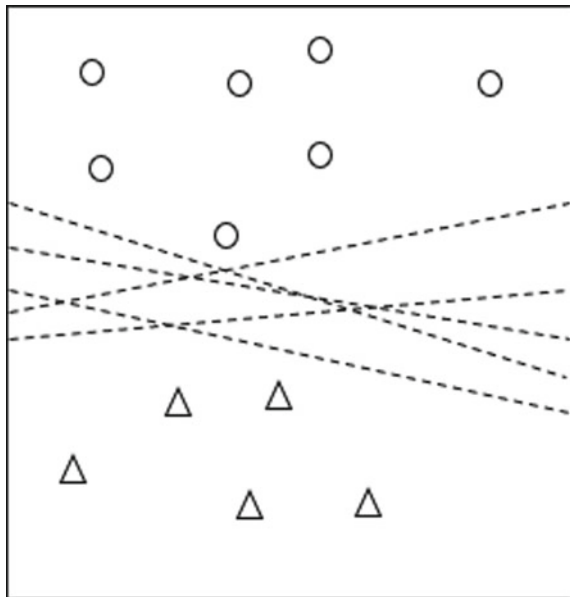
newly observed $x'$ pattern is determined by $\left| \langle w, x' \rangle + b \right| / \|w\|$. Each training pattern is at least $\Delta$ distant from decision boundary and the distance of each training sample from the hyperplane for $y_i \in \{+1, -1\}$ is determined on condition that

$$\frac{y_i [\langle w, x_i \rangle + b]}{\|w\|} \geq \Delta, \quad i = 1, ..., n \tag{2}$$

by the equality in the limit value Eq. (3).

$$\frac{1}{\|w\|} (\min_{x:y=1} |\langle w, x_i \rangle + b| + \min_{x:y=+1} |\langle w, x_i \rangle + b|) = \frac{2}{\|w\|} \tag{3}$$

The hyperplane that best separates the training samples is the plane that minimizes the equation $\eta(w) = \frac{1}{2}\|w\|^2$. Finding the optimum hyperplane for separable data is a quadratic optimization problem defined by linear limits. The problem is modeled as follows:

$$\begin{aligned} &\underset{w,b}{Min}\ \frac{1}{2} w^T w \\ &subject\ to : y_i(\langle w \cdot x_i \rangle + b) - 1 \geq 0 \end{aligned} \tag{4}$$

If the problem has a very large data space, then it is not practical to look for a solution through the primal model. Therefore, it will be beneficial to construct the dual of the problem. For that Khun-Tucker theorem is used (Srang 1986: 538–540) and it is of two steps. In the first step, an unrestricted optimization problem is formed using the Lagrange function:

$$L_D(w, b, \alpha) = \frac{1}{2} w^T \cdot w - \sum_{i=1}^{m} \alpha_i y_i (\langle w \cdot x_i \rangle + b) - 1) \tag{5}$$

In the above-mentioned equation, $\alpha_i$ is the dual Lagrange multipliers and this multipliers should be maximized by the condition, $\alpha_i \geq 0$. On the other hand, when $w$ and $b$ are taken into consideration, the Lagrange function should be minimized. Therefore, the optimal value point of the Lagrange function is required. When Karush Khun-Tucker (KKT) conditions are to be satisfied in order to find the derivation of the function according to $w$ and $b$, and to express it only according to $\alpha_i$ parameter, the restricted optimum function is rewritten. That is the second step of forming the dual model. To form the dual model, the Lagrange function is rearranged using KKT conditions. Thus, the formulation of the dual problem is determined by:

$$\underset{\alpha}{Max} \ L_D(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle$$

$$subject \ to: \ \sum_{i=1}^{l} \alpha_i y_i = 0 \ , \quad \alpha_i \geq 0, \quad i = 1, ..., m \tag{6}$$

The Lagrange function should be maximized based on the non-negative variable $\alpha_i$ with the aim of finding the optimal separator hyperplane. In the dual optimization problem, the $w^*$ and $b^*$ hyperplane parameters determine $\alpha_i$. Thus, the optimal separator decision function $f(x) = \text{sgn}(\langle w^* \cdot x \rangle + b^*)$ is rewritten:

$$f(x) = y = sig \left( \sum_{i=1}^{m} y_i \alpha_i^* \langle x_i, x \rangle + b^* \right) \tag{7}$$

In a typical classification problem, $\alpha_i$ smallest sub-set of the Lagrange multipliers tends to be larger than zero. Besides, these non-negative training vectors are geometrically very close to the optimal separator plane. These vectors are termed support vectors and the optimal separator hyperplane is defined only on these support vectors.

If the problem is complex and non-linear, the margin could have a negative value and the appropriate solution area of the problem is empty. In order to overcome this situation, which makes the solution impossible, either you need to relax the strict inequalities, which is called "soft margin optimization", or the problem is made linear using kernel trick. The soft margin optimization can be applied to make a small change in the solution explained above for linearly inseparable data.

In Fig. 2 below, (a) is an example to data that is linearly separated by the maximal margin, and (b) is an example to data that cannot be separated linearly.

In the second situation, the data can be linearly separated by assuming that a specific error is assigned for misclassified samples. In this case, the problem aims to find the hyperplane that minimizes the training errors by means of slack variables:
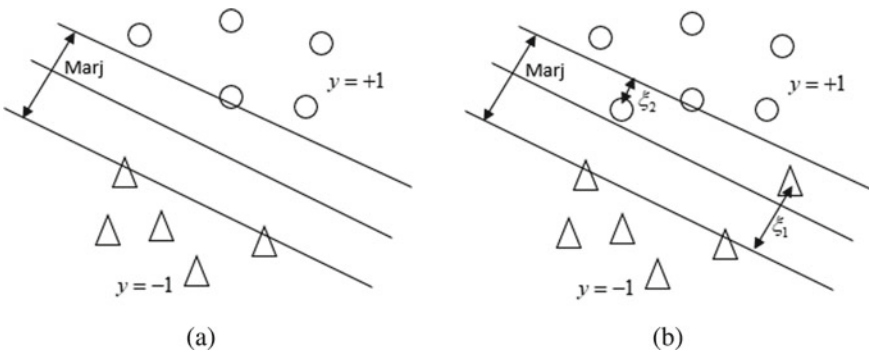


(a)                                                                                (b)

**Fig. 2** Linearly separable data (a), and Linearly inseparable data (b)

$$\underset{w,b,\xi}{Min} \ \frac{1}{2}w^T w + C \sum_{i=1}^{m} \xi_i \tag{8}$$

$$subject\ to: y_i(\langle w \cdot x_i \rangle + b) \geq 1 - \xi_i \ and \ \xi_i \succ 0, \quad i = 1, ..., l$$

In the above-mentioned model, the penalty parameter on training errors is represented by C, and the non-negative slack variable is represented by $\xi_i$. This optimization problem can be solved via the Lagrange multipliers technique. The solution of problem is furthered almost in the same way as in the linear learning case. The Dual model is given below:

$$\underset{\alpha}{Max}\ L_d(\alpha) = \sum_{i=1}^{m} \alpha_i + \frac{1}{2} \sum_{i=1}^{m} \sum_{i,j=1}^{m} y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle \tag{9}$$

$$subject\ to: \ \sum_{i=1}^{n} y_i \alpha_i = 0 \ , \ 0 \leq \alpha_i \leq C, \quad i = 1, ..., m$$

In model (9), the majorant of the Lagrange variable is represented by penalty parameter, C, and this parameter is predetermined by the user. Besides, the optimal separator hyperplane function is the same as Eq. (7). The mapping function $\phi$ is applied for training samples in the non-linear SVM. Using the appropriate kernel function defines dot product (inner product) in feature space, the classifier could separate non-linear data. The Kernel function given in Eq. (10) uses the space of the inner product that we have used in the objective function in the Dual model (9).

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \tag{10}$$

$$\underset{\alpha}{Max}\ L_d(\alpha) = \sum_{i=1}^{m} \alpha_i + \frac{1}{2} \sum_{i=1}^{m} \sum_{i,j=1}^{m} y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle \tag{11}$$

$$subject\ to: \ \sum_{i=1}^{n} y_i \alpha_i = 0 \ , \ 0 \leq \alpha_i \leq C, \quad i = 1, ..., m$$

When we follow the solution stage in the linearly separable case, the decision function is derived from $f(x) = y = sig\left(\sum_{i=1}^{m} \alpha_i^* y_i \langle K(x_i, x_j) \rangle + b^*\right)$. Besides, it must be said that there are lots of kernel functions that enhanced SVM to get the optimal result. The most commonly used of those functions are polynomial (12), radial basis (13), and sigmoid (14) kernels (Burges, 1998; Liao et al., 2004).

$$K(x_i, x_j) = (1 + x_i \cdot x_j)^d \tag{12}$$

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \tag{13}$$

$$K(x_i, x_j) = tanh(K(x_i \times x_j) - d) \tag{14}$$

## 3.2 Logistic Regression Analysis

Logistic regression is a regression analysis used to predict a dependent variable with two categories. The categories of the dependent variable here are formed by using a coding scheme as zero or one to signify that an event has occurred or has not occurred. LRA aims to find the most appropriate model to determine the relationship between a two-category dependent variable and a number of independent variables (Caesarendra, Widodo and Yang, 2010). In this manner, the logistic function with $p$ number of independent variables is expressed as in (15):

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p)}} = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p)}} \tag{15}$$

where, the statement $P(Y = 1)$ represents the probability of the relevant event of the dependent variable to occur, whereas, $\beta_0$, $\beta_1$, ..., $\beta_p$ represent regression coefficients. In the case that the dependent variable represents the probability of the relevant event to occur, the output variables comprise of a number of responses restricted between 0 and 1. Logistic regression also provides a linear model, the natural logarithm of the rate of $P(Y = 1)$ to $1 - P(Y = 1)$ in the logistic regression model:

$$g(x) = \ln\left(\frac{P(Y = 1)}{(1 - P(Y = 1))}\right) = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p \tag{16}$$

$g(x)$ in the Eq. (16) has several features desired in a linear regression model. The independent variables here can be integrated in the model as a combination of continuous and categorical variables. In the analysis, to predict $\beta_0$, $\beta_1$, ..., $\beta_p$ parameters, the maximum probability prediction is applied after the transformation of the dependent variable to logit variable (Dreiseitl and Ohno-Machado, 2002; Kurt, Ture and Kurum, 2008; Yilmaz, 2009).

## 4 Experimental Study

In the SVM literature, many different model suggestions have been made within the scope of testing and strengthening the success of the method. One of these models is LRA, which is one of the multi-variable statistical techniques. The results of the analysis, which we call the logit model, have been compared to the results

obtained by SVM. In another model, the logit model is used as a feature selection technique and with the variables which have been found significant and which would increase its prediction performance, another analysis was done by SVM. The obtained results from the proposed models have been discussed, and the comparisons are visualized through graphs. In this study, developed SVM model has been designed via MATLAB 9.4 (R2018a)—The Language of Technical Computing program and LIBSVM software system (Chang and Lin, 2011). Besides, the IBM SPSS Statistic-21 package program has been used for LRA.

## 4.1 Datasets

The firms that will be used for financial distress prediction operate in the manufacturing industry and sub-sectors of this industry. Besides, these firms are traded on the BIST stock exchange. Within the scope of these given, 172 of the firms constitute the datasets of the research. Considering that the firms which are subject to Capital Market Law (CML) and traded in Borsa Istanbul (BIST or Stock Market) have prepared their financial statements in accordance with the international financial reporting standards since 2007, the period between 2010 and 2017 has been determined as the "Research Period". Besides, 24 financial ratios in 6 groups were used in the research. These ratios have been obtained from the firms' annual balance sheets which are updated through footnotes. Using financial ratios makes it possible to control any potential problem that might occur due to the size of the enterprise and sector differences, and to minimize the impacts of those factors. Therefore, financial ratios, which are frequently used and considered important for firm distress predictions in the literature and which are statistically effective predictors, have been preferred. The financial ratios are given in Table 1. The balance sheets and income statements of the firms whose shares are traded in the Stock Market during the whole or part of the Investigation Period have been obtained by using Finnet Analysis Program.[1]

The "success" or "distress" situations of the firms were used as classifying variables in this research. Based on the definitions regarding the concepts of financial distress in the literature reviewed within the framework of the study, the financial distress criteria have been determined. According to Beaver (1966), Deakin (1972), Aktaş (1993), Altman, Zhang and Yen (2007), Özdemir (2011), these criteria are as follows:

1.  That the enterprise has filed for bankruptcy or has gone bankrupt,
2.  That the enterprise has made a loss in the last 3 years,
3.  That the enterprise has been delisted from stock exchange,
4.  That the enterprise has a negative equity,
5.  That the enterprise has been on the watchlist firms market for over a year,

---

[1] Finnet: **F**inancial **I**nformation **N**ews **Net**work. Web: https://www.finnet.com.tr/FinnetStore/Tr/Urun/Fta40.

**Table 1** Financial ratios

| Definitions | Codes | Financial ratios |
|---|---|---|
| Growth rates | X1 | Asset Growth (%) |
| | X2 | Share Equity Growth (%) |
| | X3 | Net Sales Growth (%) (Annual) |
| Valuation ratios | X4 | Market Value / Net Sales |
| | X5 | Market Value / Book Value |
| Operating ratios | X6 | Accounts Receivable Turnover (Annual) |
| | X7 | Stock Turnover (Annual) |
| | X8 | Asset Turnover (Annual) |
| Financial structure ratios | X9 | Fixed Assets / Assets |
| | X10 | Short Term Loans / Share Equity |
| | X11 | Short Term Loans / Assets |
| | X12 | Share Equity / Assets |
| | X13 | Short Term Loans / Total Loans |
| | X14 | Share Equity / Real Assets |
| | X15 | Loan Capital Ratio (%) |
| | X16 | Total Loan Growth (%) |
| Profitability ratios | X17 | Net Profit Margin (Annual) |
| | X18 | Return on Assets (%) (Annual) |
| | X19 | Real Operating Profit Margin (Annual) |
| | X20 | Profit Capital (%) (Annual) |
| | X21 | Gross Real Operating Profit Margin (Annual) |
| Liquidity ratios | X22 | Quick Ratio |
| | X23 | Current Ratio |
| | X24 | Current Assets / Total Assets |

6. That the enterprise has lost 10% of its total assets, and
7. That the enterprise has restructured its debts.

The enterprises that comply with at least one of the above criteria have been considered "distressed", and all of those that do not as "non-distressed". The distressed or non-distressed situations of all 172 firms in our data set have been identified. There are firms that were distressed all through the sampling period or firms which suffered financial distress for only one year and were non-distressed for the rest of the years. The exact opposite situation is also available. Many FDP researchers have used a balanced sample in which class frequencies are distributed as 50–50% (Altman, 1968; Park and Han, 2002; Shin et al., 2005; Sun and Li, 2011). However, most real-life problems have unbalanced class distribution (Liu et al., 2009). According to Zmijewski (1984), if the proportions of distressed and non-distressed classes differ clearly from the real-world stack, the prediction ability of the model is distorted. So

the choice covers the whole spectrum in order to avoid any selection bias, firms have been randomly selected with their financial ratios for the years in question and added to the sampling. In the entire data, it was observed that 71 of the firms are classified as distressed firms, and 101 of the firms as successful firms. İt was divided into two groups. Since there is a consensus in the literature, the data set has been randomly split into two: training and testing set (%80–%20).

## *4.2 Study Design and Experiments*

The outline of the process that has been proposed for the application part of the study is presented in Fig. 3. The detailed explanations are as follows:

### 4.2.1 Kernel Function

Different kernel functions promote SVM in finding the optimal result. Also, it is possible for the user to write their own kernel function based on the structure of the problem. The polynomial, radial basis, and sigmoid kernel are the most used kernel functions (Liao et al., 2004). Since Radial basis function (RBF) can classify multidimensional data, it is the most widely used kernel. When compared to the polynomial kernel, it is known that RBF has fewer parameters. In several studies, RBF is compared to other kernel functions and no significant difference is observed.

In this study, the radial-based kernel function is used. Because RBF for SVM has been accepted as an effective choice in finding the most suitable result.

There are two significant parameters used in SVM that are called C and gamma. The selection of the value of C, which is called the penalty parameter, affects the classification output. If we assign a very high value to C, the classification accuracy rate during the training will be very high. However, the accepted model will most probably have a very low accuracy rate on the test data. If we select C to be very small, it is known that the classification accuracy rate will not be satisfactory. Therefore, the model is impractical. Gamma parameter, on the other hand, has a higher impact on the classification output than does C, because the value of gamma affects the separation output in the feature space. Assigning very high values to gamma leads to over-fitting and very low values to under-fitting (Pardo and Sberveglieri, 2005).

### 4.2.2 Parameter Optimization

The easiest way to adjust C and gamma parameter is the Grid search technique (Hsu et al., 2003). In this technique, the identification of the appropriate parameter to ensure a high classification accuracy rate is done by trying all different combinations between the lower limit and the upper limit determined for gamma and C. As can be seen in Fig. 4, the limits for C range from $2^{-5}$ to $2^{15}$. Besides, the limits for gamma
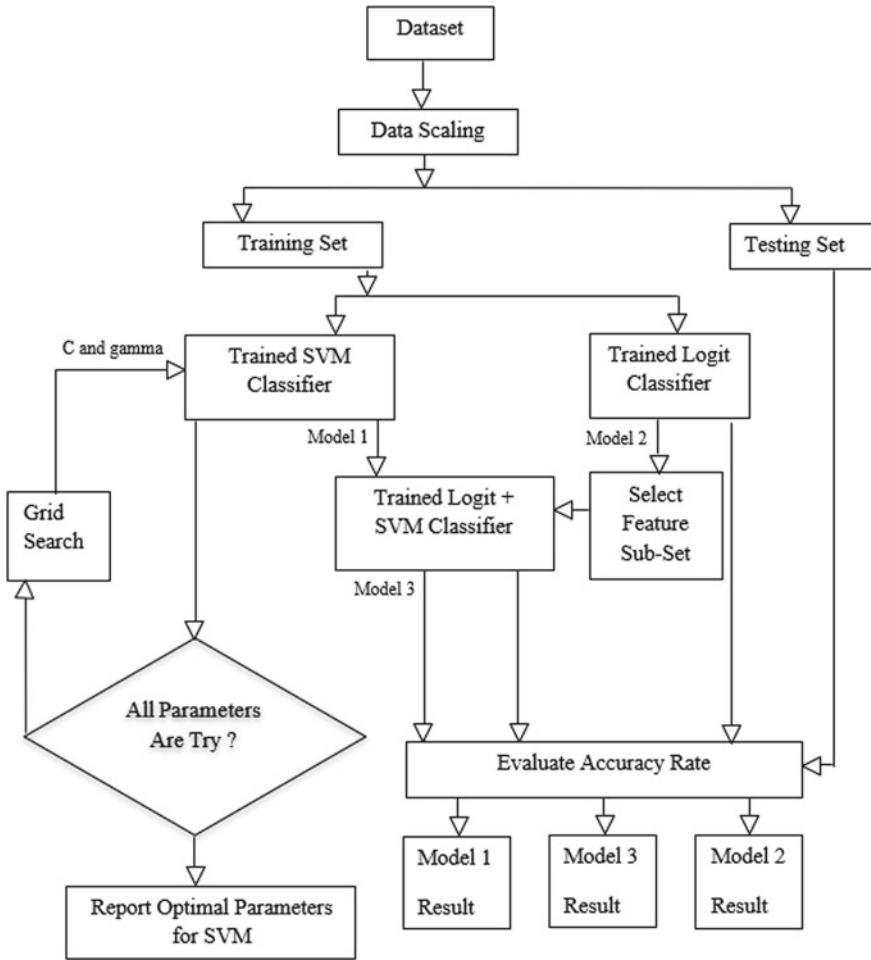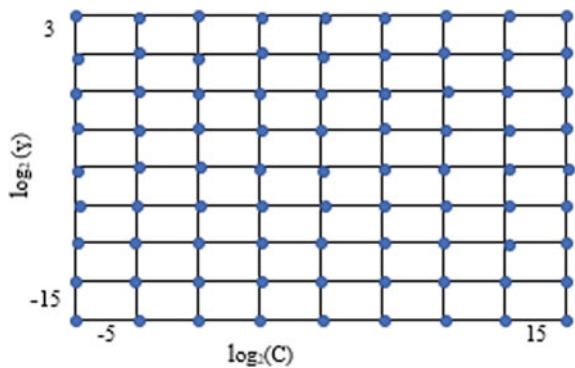
**Fig. 3** The proposed analysis process for financial distress prediction

**Fig. 4** Grid search

range from $2^{-15}$ to $2^3$. Here, 110 different results are tried and the cross-validation rate for each parameter is calculated. Then SVM training process is initiated with the parameter pair that yields the best cross-validation rate.

In this technique, which is a local search technique, the interval determined for the parameter values should be well adjusted (Lin et al., 2008). A very wide interval means wasted calculation time and determining a narrow interval might indicate that the satisfactory results are left out of the search space, or in other words, that good results are sacrificed. Determining an appropriate parameter for SVM is a separate area of study in itself and it is yet to be developed.

### 4.2.3 Feature Selection

The accuracy rate of SVM is not only affected by C and gamma parameters; the quality of the data set also effect this rate. For instance, a high correlation between features influences the solution results. Excluding an important feature from the model may reduce the accuracy rate. Conversely, some features included in the data set may not affect results or may contain noise.

Feature selection methods are analyzed under three categories as filter and wrapper (Liu and Motoda, 1998), and embedded (Saeys et al., 2007). As filter methods, factor analysis (FA), the principal components analysis (PCA), independent components analysis (ICA), and discriminant analysis (DA) are mostly used. As for wrapper methods, mostly meta-intuitive techniques with a road map which are based on the exploration of the optimal sub-set are used. In embedded techniques, random forest walk, the vector weights of SVM, and logistic model weights are used. Filter methods are fast, but they do not guarantee to give the optimal sub-set; wrapper methods work slowly and give the best *approximate* optimal solution. Embedded methods require more complicated calculations than wrapper methods since they work interactively with the classifier. While the outputs of the filter and wrapper methods are estimators, in embedded methods the output is an estimator and a feature sub-set. Based on Min and Lee (2005), LRA was used in the feature selection phase in the present study.

### 4.2.4 Data Pre-processing

Data pre-processing is applied not to have numeric difficulty during calculations and also to ensure that the large values of the variables are not affected by small values. Moreover, pre-processing appears to be a requirement for many machine learning techniques. The raw data is transformed using the formula given in Eq. (17).

$$Z_{score} = \frac{X_i - X_{mean}}{S} \qquad (17)$$

where $X_i$ is the raw value that each variable takes, $X_{mean}$ is the average of variable values, and $S$ is standard deviation. Thus, raw financial ratios are normalized, with their average as zero and standard deviation as unit across samples.

### 4.2.5 Cross Validation (*k*-fold)

In order to make sure that we have developed a model that would assign the newly added data in the sample to the correct class, the model must have an acceptable accuracy rate on the test data set which was kept out of the analysis independently. The most reliable way to do so is to divide the data into $k$ parts and to keep each time 1 part aside independently as the test set, and then train the model on the remaining $k$-1 parts. This method is called cross-validation. The advantage of cross-validation is that the test data set kept aside for each time is independent and increases the reliability of the results (Huang and Wang, 2006). $k$-fold cross-validation method was first applied in Salzberg's study in 1997 taking $k = 10$ (Salzberg, 1997).

The parameters of the method we are going to use in the application stage are optimized by the Grid search technique. The parameter pairs, and therefore, the conformity rates will change in each iteration. For that reason, in the evaluation of prediction results, $k$-fold ($k = 10$) cross-validation rate is taken into consideration.

### 4.2.6 Performance Evaluation

The confusion matrix is used with the aim of comparing the predictions of the model with actual results. The $2 \times 2$ confusion matrix to be used for a two-class example is presented in Table 2. On the left column of the table are the estimated class values of the samples kept aside as the test data set, and on the upper line are the actual class values.

In some cases, an example in the positive class might also be classified as positive in the prediction, which is called true positive (TP) separation; on the other hand, it is also possible that an example in a positive class might have been predicted to be placed in a negative class (false negative (FN) separation), which is called Type 2 error. In the exact opposite case, an example in a negative class might have been predicted to be in a negative class (true negative (TN) separation), or in a positive class (false positive (FP) separation). This is an indication of Type 1 error. The sensitivity which is called the true positive rate and specificity which is called true negative rate provides significant information about how the classifier separates the positive and negative limits. To

**Table 2** Confusion Matrix

|  |  | Actual |  |
| --- | --- | --- | --- |
|  |  | (+) | (−) |
| Prediction | (+) | (TP) | (FP) |
|  | (−) | (FN) | (TN) |

evaluate the performances of the models, some performances criteria in the related literature are used criteria. The formulas of these performance criteria (accuracy, sensitivity, specificity, certainty, and Matthews correlation coefficient (MCC)) are as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{18}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{19}$$

$$Specificity = \frac{TN}{TN + FP} \tag{20}$$

$$Precision = \frac{TP}{TP + FP} \tag{21}$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{22}$$

### 4.2.7 Model Propositions

In order to obtain a powerful and useful prediction model, three different models have been proposed. Explanations about the models are presented under the titles below; the results and interpretations are discussed in the sect. 4.3 "*Empirical Results and Discussion*".

**Model 1:** The Analysis by the Support Vector Machines. In Model 1, all variables (Table 1) are used. These variables are the financial ratios which are most commonly encountered in the literature and which provide in many studies significant information regarding explaining financial distress. For the dependent variables of the sampling of 172 firms, only SVM, the support vector machine, the parameters of which have been optimized, has been applied in Model 1. This model has been named Grid SVM.

**Model 2:** The Analysis by the Logistic Regression. In Model 2, all variables are used to do LRA. This model, which we have called Logit, has been used to be informed about the performance of SVM.

**Model 3:** The Analysis with Feature Selection. In Model 3, LRA is used as the feature selection technique. Thanks to this analysis, the sub-set of features that will provide useful information was determined and SVM model was used. This model has been named Logit + Grid SVM.

## 4.3   *Empirical Results and Discussion*

Empirical results are analyzed under three main sections: The titles are: *(1) Logistic Regression Model Output, (2) SVM Models Output (3) The Performances of Models.*

### 4.3.1   Logistic Regression Model Outputs

LRA takes the cumulative logistic function as the basis. This function, when the financial characteristics of the firms are given, gives the probability of whether the firm will be included in the distressed or non-distressed class. The empirical results of this model are presented in Table 3.

$x_1$ : asset growth, $x_{19}$ : real operating profit margin, $x_{17}$ : net profit margin, $x_{21}$ : gross real operating profit margin, $x_{23}$ : current ratio, and $x_{22}$ : quick ratio in the model have been found to be significant at the 95% confidence level. The B value in the table indicates the coefficients of the logit model. The obtained logit model according to these results can be written as follows:

$$L_i = -1.031 - 1.761x_1 - 1.947x_{19} - 1.750x_{17} + 0.746x_{21} + 1.465x_{23} - 2.728x_{22}$$

It is seen that the prediction model is completely meaningful according to the statistical results *(-2 Log Likelihood = 86.949; $\chi - Squared = 12.493$; degrees of freedom (d. f.) = 8; p value = 0.131)*. From the statistical results of the coefficients *($\chi - Squared\ 100,654$; degrees of freedom (d. f.) 6; p value = 0.000)*, it is concluded that the coefficients are significant. For the obtained Logit model, it is interpreted that the independent variables can explain 69.5% of the variability *(Nagelkerke R-Square = 0.695)* in the financial situations of the firms.

To calculate the probability of whether a firm is financially non-distressed, the relevant financial ratios of the firm are placed in the $L_i$ function. The probability

**Table 3**  Logistic model outputs

|          | $\beta$  | Standard error | Wald   | d. f | p-value | Exp ($\beta$) | 95% confidence interval | |
|----------|----------|----------------|--------|------|---------|---------------|-------------------------|-------------|
|          |          |                |        |      |         |               | Lower bound | Upper bound |
| $x_1$    | −1.761   | 0.408          | 18.664 | 1    | 0.000   | 0.172         | 0.077       | 0.382       |
| $x_{19}$ | −1.947   | 0.767          | 6.453  | 1    | 0.011   | 0.143         | 0.032       | 0.641       |
| $x_{17}$ | −1.750   | 0.772          | 5.136  | 1    | 0.023   | 0.174         | 0.038       | 0.789       |
| $x_{21}$ | 0.746    | 0.377          | 3.916  | 1    | 0.048   | 2.108         | 1.007       | 4.412       |
| $x_{23}$ | 1.465    | 0.705          | 4.310  | 1    | 0.038   | 4.326         | 1.085       | 17.242      |
| $x_{22}$ | −2.728   | 1.096          | 6.197  | 1    | 0.013   | 0.065         | 0.008       | 0.560       |
| Constant | −1.031   | 0.355          | 8.455  | 1    | 0.004   | 0.357         | –           | –           |

value corresponding to these numbers is calculated using $P(L_i) = \frac{1}{1+e^{-L_i}}$ equation. When this value is higher than 0.5, it is decided that the firm will be non-distressed; otherwise, it will be distressed.

### 4.3.2 SVM Models Output

Under this title, the classification performances of Logit + Grid SVM models in which we have applied the LRA as the feature selection technique are compared. In addition to the optimization of SVM parameters, it is concluded that the optimal feature sub-set selection affects the classification success of SVM. The analysis outputs presented in Fig. 5 show that the parameters of SVM can affect the results. As has been mentioned in previous sections, when C and gamma values are set to be very high causes over-fitting error. In the analyses done on the test data, the classification success of the method decreases. When the constant value for C is determined to be $2^5$ and when we look at the cross-validity rate that it takes for all values in the determined interval for gamma value, the cross-validation rate decreases to around 60%—the values shown on the blue line—at too high or too low values. This situation applies to both model propositions.

When we look at Fig. 5a, which shows Grid SVM results, the highest accuracy rate is 87.21%. It is seen that this accuracy rate is achieved when 2048 values for C and 1.2207e-04 values for gamma are assigned. In Fig. 5b, the impact of C and gamma on classification success in Logit + Grid SVM model is seen. Here, the highest cross-validation rate is 90.06% and this rate has been obtained at 256 value for C and 0.002 value for gamma. Another noteworthy point here is that the addition of feature selection stage to the analysis has increased the maximum value of cross-validation from 87.21 to 90.06%. In Table 4, a brief assessment of the effects of feature selection on SVM results was made. The values on the table are the values obtained by running both models 100 times on the test data set. As is indicated by
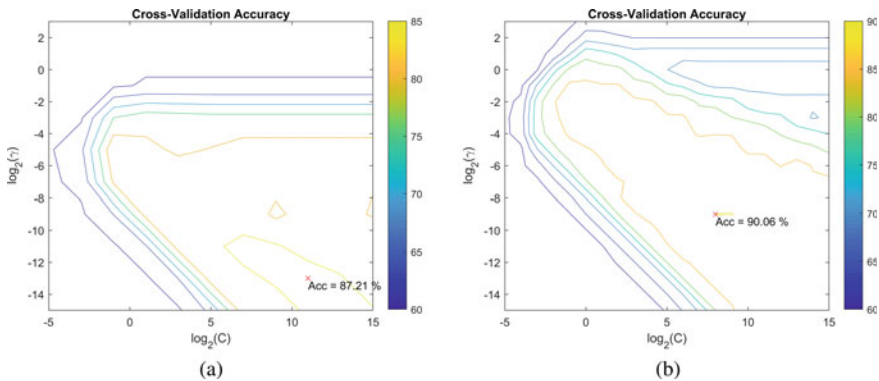


**Fig. 5** Grid SVM (a) and Logit + Grid SVM Models Cross Validation Rates Graphs (b)

**Table 4** Empirical results regarding models

| Models | Accuracy rate | | | Cross-validation rate | | |
|---|---|---|---|---|---|---|
| | Mean | Standard Deviation | Max | Mean | Standard deviation | Max |
| Grid SVM | 83.28% | (±0.0597) | 90.63% | 70.39% | (±11.25) | 87.21% |
| Logit + Grid SVM | 85.44% | (± 0.5520) | 93.75% | 74.80% | (±11.63) | 90.12% |

the results, the accuracy rate for SVM after feature selection increased from 83.28 to 85.44%. The cross-validation rate, which is a more reliable rate, increased from 70.39 to 74.80%.

### 4.3.3 Performances of the Proposed Models

It is seen that some different performance criteria are used in comparing the classification performances of the proposed models. Table 5 presents the results for the selected performance criteria. The accuracy rate of Logit + Grid SVM for training and the test sets are 94.24% and 93.75%, respectively. It can also be said that this model has a remarkably high sensitivity for both the training and the test set at a rate of 93.75% and 94.44%, respectively. The highest value of the specificity rate indicates the accuracy of the classifying model has been given by the logit model. It is the certainty rate which gives information about how many of the estimations of financial distress are real. The highest certainty value, too, has been obtained through Logit + Grid SVM. MCC value, which we have preferred for the situations in which the values in the confusion matrix are not distributed evenly, also provides information about the quality of the classifier. The highest MCC value again belongs to Logit + Grid SVM. It can be said that all three models are useful and produce classifiers with considerably high performances. As for the generalization capacity of the models, the relatively higher difference between the accuracy rates of the Logit model on the training data set and test data set indicates that its generalization performance is low.

**Table 5** Performance of the proposed models

| | Grid SVM | | Logit | | Logit + Grid SVM | |
|---|---|---|---|---|---|---|
| | Training | Test | Training | Test | Training | Test |
| Accuracy | 0.9282 | 0.9063 | 0.9000 | 0.8000 | 0.9424 | 0.9375 |
| Sensitivity | 0.9310 | 0.9000 | 0.8545 | 0.6153 | 0.9375 | 0.9444 |
| Specificity | 0.9268 | 0.9091 | 0.9294 | 0.9411 | 0.9452 | 0.9286 |
| MCC | 0.8539 | 0.7896 | 0.7893 | 0.6018 | 0.8831 | 0.8730 |
| Precision | 0.9000 | 0.8182 | 0.8867 | 0.8888 | 0.9615 | 0.9444 |

For precision, the Grid SVM yielded the lowest rate for the test set. This value is lower than the Logit has. Although it is shown in this study that logistic regression provides significant information with regard to the selection of the new feature sub-set, it is also seen that the performance of SVM operated by this new feature sub-set has increased.

## 5 Conclusion and Future Work

Since the financial distress of firms does not only affect the firm but also has an impact on the whole economy, financial distress prediction is a critically important subject, which has been frequently studied. In recent years, SVM has been commonly used in financial distress prediction studies. The financial distress model with SVM has been compared to other machine learning methods, it has been shown to yield good results. In the present study, it is aimed to make distress prediction by SVM. C and gamma parameters, which are considered as two significant parameters of SVM, are optimized by using grid search technique. It is shown to what extent the results are affected as a result of defining the relevant parameter pair correctly. Besides, it was seen that feature selection for SVM is another factor that significantly affects the results. To understand how feature selection affects classifying performance, the logistic regression analysis has been done. There are two reasons why this method has been chosen in the study: the first is that LRA does not require strict assumptions as in multiple-variable statistical techniques and it can be used as a feature selection technique; the second reason is that we wanted to compare the results of the logistic regression analysis to those of SVM.

Financial distress prediction is made based on a real data set of firms (172 firms) traded in the BIST share market between 2010 and 2017. The proposed models are compared based on this real data set. When the results of these proposed models are compared, it is concluded that SVM, which allows parameter optimization and feature selection, has a better success. As a consequence, a useful early warning model in financial distress prediction problem through SVM, a newly developed technique, is presented in the study.

## References

Agostini M (2018) Corporate financial distress: going concern evaluation in both international and US contexts. Springer

Akkaya GC, Demireli E, Yakut ÜH (2009) İşletmelerde Finansal Başarısızlık Tahminlemesi: Yapay Sinir Ağları Modeli ile IMKB Üzerine Bir Uygulama. Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi 10(2):187–216

Aktaş R (1993) Endüstri İşletmeleri Için Mali Başarısızlık Tahmini: Çok Boyutlu Model Uygulaması, T. İş Bankası Kültür Yayınları, Genel Yayin No 323, Ankara

Alifiah MN (2014) Prediction of financial distress companies in the trading and services sector in Malaysia using macroeconomic variables. Procedia Soc Behav Sci 129:90–98

Altman EI, Zhang L, Yen J (2007) Corporate financial distress diagnosis in China. New York University Salomon Center Working Paper

Altman EI (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J Financ 23(4):589–609

Altman EI (1993) Corporate financial distress and bankruptcy. Wiley, New York

Atiya AF (2001) Bankruptcy prediction for credit risk using neural networks: a survey and new results. IEEE Trans Neural Netw 12(4):929–935

Beaver WH (1966) Financial ratios as predictors of failure. J Account Res 4:71–111

Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 2(2):121–167

Caesarendra W, Widodo A, Yang BS (2010) Application of relevance vector machine and logistic regression for machine degradation assessment. Mech Syst Signal Process 24(4):1161–1171

Chang CC, Lin CL (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol (TIST) 2(3):1–27

Chen MY (2011) Predicting corporate financial distress based on integration of decision tree classification and logistic regression. Expert Syst Appl 38(9):11261–11272

Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press

Dambolena IG, Khoury SJ (1980) Ratio stability and corporate failure. J Financ 35(4):1017–1018

Deakin EB (1972) A discriminant analysis of predictors of business failure. J Account Res 10(1):167–179

Ding Y, Song X, Zeng Y (2008) Forecasting financial condition of Chinese listed companies based on support vector machine. Expert Syst Appl 34:3081–3089

Doğan S, Koçak D, Atan M (2019) Support vector machines and logistic regression analysis on predicting financial distress model. In: International Conference on Data Science, Machine Learning and Statistics. pp 292–295

Doğan S (2020) Optimal Parametre ve Özellik Seçimi ile Destek Vektör Makinesi Kullanılarak Finansal Başarısızlık Tahmini (Doktora Tezi), Gazi Üniversitesi, Ankara

Dreiseitl S, Ohno-Machado L (2002) Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 35(5–6):352–359

Etemadi H, Rostamy A, Dehkordi H (2009) A genetic programming model for bankruptcy prediction: empirical evidence from Iran. Expert Syst Appl 36:3199–3207

Fan A, Palaniswami M (2000) Selecting bankruptcy predictors using a support vector machine approach. In:Proceeding of the International Joint Conference on Neural Network vol 6, pp 354–359

Fletcher D, Goss E (1993) Forecasting with neural networks: an application using bankruptcy data. Inf Manag 24(3):159–167

Gepp A, Kumar K, Bhattacharya S (2010) Business failure prediction using decision trees. J Forecast 29:536–555

Gepp A, Kumar K (2008) The role of survival analysis in financial distress prediction. Int Res J Financ Econ 16:1450–2887

Gunn SR (1998) Support vector machines for classification and regression. ISIS Tech Rep 14(1):5–16

Hashi I (1997) The economics of bankruptcy, reorganization, and liquidation: lessons for East European Transition Economies. Russ East Eur Financ Trade 33(4):6–34

Hsu CW, Chang CC, Li CJ (2003) A practical guide to support vector classification. Available from http://www.csie.ntu.edu.tw/~cjlin/paper/guide/guide.pdf

Huang CL, Wang CJ (2006) A GA-based feature selection and parameters optimization for support vector machines. Expert Syst Appl 31(2):231–240

Jo H, Han I, Lee H (1997) Bankruptcy prediction using case-based reasoning, neural network and discriminant analysis for bankruptcy prediction. Expert Syst Appl 13(2):97–108

Kumar P, Ravi V (2007) Bankruptcy prediction in banks and firms via statistical and intelligent techniques. A Rev Eur J Oper Res 180:1–28

Kurt I, Ture M, Kurum AT (2008) Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. Expert Syst Appl 34(1):366–374

Li H, Sun J, Wu J (2010) Predicting business failure using classification and regression tree: an empirical comparison with popular classical statistical methods and top classification mining methods. Expert Syst Appl 37(8):5895–5904

Li H, Sun J (2012) Forecasting business failure: the use of nearest-neighbour support vectors and correcting imbalanced samples: evidence from the chinese hotel industry. Tour Manage 33(3):622–634

Liang D, Lu CC, Tsai CF, Shih GA (2016) Financial ratios and corporate governance indicators in bankruptcy prediction: a comprehensive study. Eur J Oper Res 252:561–572

Liang L, Wu D (2005) An application of pattern recognition on scoring chinese corporations financial conditions based on backpropagation neural network. Comput Oper Res 32(5):1115–1129

Liao Y, Fang SC, Nuttle HLW (2004) A neural network model with bounded-weights for pattern classification. Comput Oper Res 31:1411–1426

Lin SW, Lee ZJ, Chen SC, Tseng TY (2008) Parameter determination of support vector machine and feature selection using simulated annealing approach. Appl Soft Comput 8:1505–1512

Lin WY, Hu YH, Tsai CF (2012) Machine learning in financial crisis prediction: a survey. IEEE Trans Syst Man Cybern Part C Appl Rev 42(4):421–436

Lincoln M (1984) An empirical study of the usefulness of accounting ratios to describe levels of insolvency risk. J Bank Finance 8(2):321–340

Liu H, Motoda H (1998) Feature extraction, construction and selection: a data mining perspective. Springer Science & Business Media

Liu XY, Wu J, Zhou ZH (2009) Exploratory undersampling for class-imbalance learning. IEEE Trans Syst, Man, Cybern Part B (Cybern) 39(2):539–550

Meyer AD (1982) Adapting to environmental jolts. Adm Sci Q 515–537

Meyer PA, Pifer H (1970) Prediction of bank failures. J Financ 25:853–868

Min JH, Lee YC (2005) Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert Syst Appl 28(4):603–614

Min SH, Lee J, Han I (2006) Hybrid genetic algorithms and support vector machines for bankruptcy prediction. Expert Syst Appl 31:652–660

Odom M, Sharda R (1990) "A neural networks model for bankruptcy prediction". In: Proceedings of The IEEE International Conference on Neural Network 2:163–168

Ohlson J (1980) Financial ratios and the probabilistic prediction of bankruptcy. J Account Res 18(1):109–131

Özdemir FS (2011) Finansal Başarısızlık ve Finansal Tablolara Dayalı Tahmin Yöntemleri. Ank: Siyasal Kitapevi 82(33–37):106–108

Pardo M, Sbervaglieri G (2005) Classification of electronic nose data with support vector machines. SensS Actuators 107:730–737

Park C-S, Han I (2002) A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction. Expert Syst Appl 23:255–264

Pindado J, Rodrigues L (2005) Determinants of financial distress costs. Fin Markets Portfolio Mgmt 19(4):343–359

Piramuthu S (2004) Evaluating feature selection methods for learning in data mining application. Eur J Oper Res 156:483–494

Saeys Y, Inza I, Larranaga P (2007) A review of feature selection techniques in bioinformatics. Bioinformatics 23(19):2507–2517

Salzberg SL (1997) On comparing classifiers: pitfalls to avoid and a recommended approach. Data Min Knowl Discov 1:317–327

Shin KS, Lee TS, Kim HJ (2005) An applications support vector machines in bankruptcy prediction model. Expert Syst Appl 28:127–135

Shin KS, Lee YJ (2002) A genetic algorithm application in bankruptcy prediction modeling. Expert Syst Appl 23:321–328

Smola A, Schölkopf B (1997) On kernel-based method for pattern recognition, regression, approximation and operator inversion. Algorithmica 22:211–231

Srang G (1986) Introduction to applied mathematics. Wellesley-Cambridge Press, Wellesley, MA

Sun J, Li H, Huang QH, He KY (2014a) Predicting financial distress and corporate failure: a review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. Knowl Based Syst 57:41–56

Sun J, Li H (2011) Dynamic financial distress prediction using instance selection for the disposal of concept drift. Expert Syst Appl 38:2566–2576

Sun J, Li H, Huang QH, He KY (2014b) Predicting financial distress and corporate failure: a review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. Knowl Based Syst 57:41–56

Tam K, Kiang M (1992) Managerial applications of neural networks: the case of bank failure predictions. Manage Sci 38(7):926–947

Tam K (1991) Neural network models and the prediction of bank bankruptcy. Omega 19(5):429–445

Tsai CF (2008) Financial decision support using neural networks and support vector machines. Expert Syst J Knowl Eng 25(4):380–393

Vapnik VN (1995) The nature of statistical learning theory. Springer-Verlag, New York

Woods K, Bowyer KW (1997) Generating ROC curves for artificial neural networks. IEEE Trans Med Imaging 16(3):329–337

Wu CH, Tzeng GH, Goo YJ, Fang WC (2007) A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. Expert Syst Appl 32(2):397–408

Wu W, Cheng V, Lee S, Tan TY (2006) Data preprocessing and data parsimony in corporate failure forecast models: evidence from Australian materials industry. Account Financ 46:327–345

Yilmaz I (2009) Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat Landslides (Tokat-Turkey). Comput Geosci 35(6):1125–1138

Zhang G, Hu MY, Patuwo BE, Indro DC (1999) Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis. Eur J Oper Reseach 116:16–32

Zhongsheng H, Yu W, Xiaoyan X, Bin Z, Liang L (2007) Predicting corporate financial distress based on integration of support vector machine and logistic regression. Expert Syst Appl 33(2):434–440

Zhou L, Lu D, Fujita H (2015) The performance of corporate financial distress prediction models with features selection guided by domain knowledge and data mining approaches. Knowl-Based Syst 85:52–61

Zmijewski ME (1984) Methodological issues related to the estimation of financial distress prediction models. J Account Res 22:59–82