EGE AKADEMİK BAKIŞ / *EGE ACADEMIC REVIEW*

Cilt 23 • Sayı 2 • Nisan 2023
SS. 145/162
Doi: 10.21121/eab.1027084
Başvuru Tarihi: 23.11.2021 • Kabul Tarihi: 25.01.2023

Article Type: *Research Article*

Using Machine Learning Methods in Financial Distress Prediction: Sample of Small and Medium Sized Enterprises Operating in Turkey

Yusuf AKER[1], Alper KARAVARDAR[2]

ABSTRACT

Financial distress has become one of the main topics on which lots of research has been done in the recent finance literature. This paper aims to predict the financial distress of Turkish small and medium firms using Logistic Regression, Decision Tree, Random Forest, Support Vector Machines, K-Nearest Neighbor and Naive Bayes model. Empirical results indicate that decision tree model is the best classifier with overall accuracy of %90 and %97 respectively for 1 and 2 years prior to financial distress. Three years prior to financial distress, Naive Bayes outperform other models with an overall accuracy of 92.86%. Furthermore, this study finds that distressed firms have more bank loans and lower equity. In the Turkish economy, where cyclical fluctuations are high in the last decade, distressed firms grew rapidly with high bank loans and gained higher operating profits than non-distressed firms. After a while, distressed firms that cannot manage their financial expenses get into financial trouble and go bankrupt. This article can be useful for managers, investors and creditors as well as its contribution to academic research.

Keywords: Financial Distress Prediction, Decision Tree, Naive Bayes, Support Vector Machine, Random Forest, Logistic Regression.

JEL Classification Codes: G32, G33, C52

Referencing Style: APA 7

INTRODUCTION

Financial distress prediction describes any situation where a company has certain kind of financial difficulties, especially loan payments due to creditors, making it unable to meet its financial obligations. If necessary measures are not taken on time, prolonged financial distress may eventually lead to bankruptcy. After the 2008 global financial crisis, this issue has never lost its importance due to 15 July 2016 coup attempt and the 2018-2021 currency and debt crisis in Turkey. Notwithstanding being financially supported by the government recently, company bankruptcies continued to increase. According to the report of the International Finance Institute (2021), Turkey ranks first in the world in the number of bankruptcies in 2020 and second in the increase in debt ratio among developing countries. Highly volatile flow of foreign exchange, increase in interest rates and soaring inflation caused a recession in Turkish economy, as a result of this, there has been a great increase in the number of companies going bankrupt recently. Additionally, according to the news that Bloomberg (2018) based on a consultancy company report, the fact that 80 percent of newly established companies in Turkey go bankrupt within the first five years shows us how important this issue can be for Turkey.

The main purpose of this study is to determine both which machine learning model has higher prediction accuracy and which financial ratios are more successful in distinguishing distressed and non-distressed firms. In the Turkish economy, where cyclical fluctuations are high, we wonder which financial ratios maximized the prediction accuracy that occurred 1, 2 and 3 years before the failure. Additionally, using selected ratios, this paper aims to make a comparison of forecast accuracy for six models, namely Logistic regression (LR), Decision Tree (DT), k-Nearest-Neighbours (kNN), Support Vector Machine (SVM), Naive Bayes (NB) and Random Forest (RF).

According to the report of the Banking Regulation and Supervision Agency (BDDK, 2019) the SME sector, which has a very important place in the Turkish economy, is the sector most affected by the negativities or crises that may occur in market conditions. The share of SME loans in all loans in Turkey is 24%, and these loans are more profitable loans for financiers due to the high spread. But we see that in the literature review conducted, it was

[1] Dr., Türkiye Finans Katılım Bankası, yusuf_aker@yahoo.com
[2] Doç. Dr., Giresun Üniversitesi, İşletme Bölümü, akaravardar@yahoo.com

seen that 88.5% of the studies on financial distress were conducted on commercial companies originating from Borsa Istanbul (BIST). 102 studies published on financial distress were analyzed and it was determined that only 2 of these studies examined SMEs. This study also aims to contribute to the literature on financial distress in SMEs in Turkey.

The second contribution of this article to the literature is related to the number of SMEs examined. In the literature review in the Turkish context, an average of 106 company data per survey was analyzed. In this study, the financial statements of 392 companies were examined and studied with much more data than other studies. According to Öğündür (2020), it is critical to work with enough data to avoid overfitting and underfitting in machine learning. The third and most important contribution of the article is related to the distress criteria of distressed firms. In the literature review conducted in the context of Turkey, it is seen that the distress criteria in most studies consist of make loss two years in a row, make loss three years in a row, 2/3 decrease in equity, Altman Z score ratio, previous year's loss exceeding 10% of the asset. These criteria are not definitive criteria for bankruptcy and companies with these characteristics can survive. In this study, the companies that were given bankruptcy or concordat decisions by the courts were deemed financially distressed, so the probability of making a mistake in this matter was reduced to zero. This issue is very important because error in the selection of distress criteria will make the research results controversial.

When examining the financial statements of companies in Turkey, two types of balance sheet structures are encountered. Inasmuch as analyzes should be made with similar types of financial statements, the research was continued with the financial statements of the companies that keep records on a calendar year basis. According to the official announcement website of the government, ilan.gov.tr, in 2018, 814 SMEs consisting of the same type of financial statements declared concordat or bankruptcy. Data on 219 SMEs, which constitute 27% of these unsuccessful SMEs, were reached.

This paper consists of 6 sections. Section 2 contains a brief review of the literature. In section 3, we present a study that demonstrates the working techniques of the machine learning models used in this study. Chapter 4 illustrates the samples and feature selection process. Section 5 presents the empirical results. In this section, a specific research methodology was applied while analyzing the data. Section 7 shows the results.

LITERATURE REVIEW

After Patrick (1931) who was one of the first to work in this field, discovered that there is a difference between distressed and non-distressed firms, many researchers (Beaver, 1966; Altman, 1968; Blum, 1972; Deakin, 1972; Libby, 1975; Springate, 1978; Fulmer, 1984) established new models or developed existing models to predict financial distress by means of financial ratios (Tuncay, 1998). Especially in the 2000s, financial markets developed, and with the increase in the software and hardware features of computers, the big data era was entered. Instead of theoretical and statistical models, machine learning models have started to be used more. Aziz & Dar (2006) analyzed 46 articles included 89 empirical studies. He ascertained that financial ratios were used as explanatory variables in more than 60% of the studies, and that machine learning models had a higher accuracy rate than statistical and theoretical models.

Table 1. Methodology of data analysis

<table>
<tr><th colspan="7">Data pre-processing</th></tr>
<tr><td>Process</td><td colspan="6">Obtaining 2015-2016-2017 distressed and non-distressed SMEs data
Missing data detection - Replacing missing values with median
Outlier detection - Replacing outliers with Tukey method
Data normalization
Splitting the data for training and testing (%70 train - %30 test)
Feature selection with Random Forest - Recursive feature elimination (RF-RFE)
Correlation and multicollinearity (VIF) analysis of independent variables</td></tr>
<tr><th colspan="7">Analysis methods</th></tr>
<tr><td>Methods</td><td>LR</td><td>DT</td><td>kNN</td><td>RF</td><td>NB</td><td>SVM</td></tr>
</table>

Both traditional statistical methods and machine learning methods were used in 13 studies (İçerli, 2005; Aksoy, 2018; Ural, 2020; Çelik, 2009; Torun, 2007; Ergin, 1999; Paket, 2014; Şengören, 2019; Civan & Dayı, 2014; Hesarı, 2018; Aktaş, Doğanay & Yıldız, 2003; Yazıcı, 2018; Yakut & Elmas, 2013) examined in the context of Turkey. While the predictive accuracy of these models is compared, it is noticed that machine learning models give 92% more successful result. It has been observed that the machine learning models used in our research are either not used at all or used very little in studies in Turkey. One of the 2 studies with the DT model was conducted by Hesarı (2018), and it was concluded that the DT model has higher prediction power than the artificial neural networks and discriminant model. Yakut & Elmas (2013), who conducted the other study, found that DT was more successful than discriminant analysis. Apart from Selcik (2019) in the RF model and Aksoy & Boztosun (2019) in the kNN model, no other financial distress studies used these models were found. One of the 2 studies with the SVM model was done by Ceran & Bülbül (2019) that mentioned the lack of studies with the SVM model in Turkey. Şengören (2019), who conducted the other study, used the LR and SVM models and found that the SVM model had higher prediction power than the LR model. No bankruptcy prediction studies have been conducted in Turkey with the NB method before, and this article is the first study in this field.

FINANCIAL DISTRESS PREDICTION MODELS

Beaver (1966), who made the first statistical study on the prediction of financial distress, examined the effect of accounting data on the prediction of bankruptcy. Beaver, with his univariate analysis, reached the end of that financial ratios have different trends in predicting bankruptcy. Afterwards, one of the most important studies in the bankruptcy literature was fulfilled by Altman (1968). Altman formed two groups with 33 successful and 33 unsuccesful firms and examined these firms using the multiple discriminant analysis method and then developed the Z score model consisting of 5 important financial ratios. The accuracies of this model were found that equal to 95%, and 72% for one and two years prior to bankruptcy, respectively (Mselmi, Lahiani & Hamza, 2017).

Deakin (1972), on the other hand, tried to bring together the good sides of both models by using the model developed by Beaver and Altman. Deakin matched 32 distressed firms that went bankrupt between 1964 and 1970 with 32 non-distressed firms, taking into account asset size and financial data periods. The findings obtained as a result of the study showed great similarities with the results obtained by Beaver in 1967. In both studies, it has been determined that the ratio of cash flow to total debt is the most effective method in estimating financial failure. In recent years, with the quickly progress of computer technology, machine learning algorithms and data mining have begun to be used successfully in estimating banktruptcy (Tuncay,1998).

In this paper, six different classification methods are used to predict bankruptcy of Turkish small and medium sized firms, namely LR, DT, RF, SVM, kNN, NB. In the context of Turkey, it has been observed that there are not enough financial distress prediction research with machine learning models except for LR and artificial neural networks (ANN). We only included the LR model in the study because we want to compare this model with other models.

Logistic Regression

The logistic regression model emerges as a model suitable for situations where the dependent variable is categorical or classified. It is a method used to determine the cause-effect relationship with explanatory variables in cases where the dependent variable is observed in binary, triple and multiple categories (Özdamar, 2002). In logistic regression analysis, predictions are made using equations. By drawing the linear line that best fits the relationship between the dependent and independent variables, it is tried to find the equation that will minimize the difference between the actual value of the dependent variable and the predicted value of the independent variable (Shannon & Davenport, 2001). In the logistic regression model, π represents probability, β1 represents the regression coefficient and Xi represents the independent variables. The β0 and β1s represents in the model are estimated using the maximum probability method. The logistic regression model is written as follows (Kalaycı, 2009);

$$\text{Log}\,\frac{\pi}{1-\pi} = \beta0 + \beta1X1 + \beta2X2 + \beta3X3 + \ldots.\ \beta mXm$$

Since the main purpose here is to put the model in training or testing, a Z value is obtained after X1, X2 and all the following variables are multiplied by their weights.

$$Z = b\ (\text{bias}) + X1w1 + X2w2 + X3w3 + \ldots\ldots. + Xnwn$$

The resulting Z value is inserted into the Sigmoid function and equalized to a value between 0 and 1. Since the sigmoid function is a derivative function, the values and weights of X will be constantly updated as data is entered, and the Z value will take a new value each time (udemy.com).

Decision Tree

The decision tree model is a classification model in the tree structure consisting of decision nodes and leaf nodes. In this method, the data set is developed by dividing it into small pieces. In decision trees, which can consist of both categorical and numerical data, decision nodes can contain more than one branch. The first node is called the root node (Uzun, 2020).

The decision tree model can be defined as "the iterative model obtained by dividing n sets of statistical units into subgroups". The purpose of dividing the unit set into subsets is to increase the purity of the group by creating more subsets than the other group. The algorithm terminates when homogeneity is achieved in the subgroup and no more pure groups can be obtained (Soo & Upneja, 2014).

Random Forest

The random forest model, which is one of the types of supervised algorithms, is an algorithm that creates a random forest, as the name suggests, and is used in both classification and regression problems. Although there is no direct relationship between the number of trees and the result, the result is achieved as the number of trees increases. The difference with the decision tree algorithm is that the root nodes in the random forest algorithm are chosen randomly. The random forests algorithm is easy to understand and interpret. Its main advantages are that tree structures are visualizable and do not require extensive data preparation for analysis. However, if it produces overly complex trees that do not explain the data well, the probability of yielding positive results decreases. At the same time, in case of overfitting the system, the algorithm may lose its flexibility and give a result close to memorization and go into unnecessary details. Algorithm is more successful at classification than regression (Çebi, 2020).

k-Nearest Neighbor

The k-Nearest Neighbor algorithm is a type of algorithm that classifies the data to be classified according to the proximity relationship to the previous data. It is a simpler learning model compared to the complex structure in other algorithms. In this model, which does not have a training phase, training and testing mean almost the same thing. In this model, the closest points to the new point are sought. The closest points are represented by k. The k value represents the distance of the new data to the old data points to be measured. If we choose the k value as 3, the distance of 3 data points close to the new point is measured. The classification decision is made by looking at which class the 3 closest points belong to. The algorithm calculates these distances according to the Euclidean distance rule.

$$\sqrt{\sum_{i=1}^{n}(\boldsymbol{q1}-\boldsymbol{p1})^2}$$

The kNN algorithm has some limitations. The choice of parameter k is critical to the success of the algorithm. Choosing the k parameter too small makes the algorithm sensitive to noisy data. The second important point is the selection of the appropriate distance criterion. Data need to be standardized without applying the distance criterion. It is known that the Euclidean distance criterion, which is suitable if the independent variables are continuous, cannot classify well in multivariate high-dimensional data. This algorithm, which needs high memory in high-dimensional data, is recommended to make classify with a small number of variables (Lantz, 2019).

Naive Bayes

Naive Bayes classifier is a model named after Thomas Bayes and based on Bayes Theorem. It determines the category of data with a set of calculation methods defined according to probability principles. Bayes' Theorem is a graphical model that shows the relationships between variables using conditional probabilities. In the model, variables are represented as nodes, while possible relationships are represented as lines between nodes. An unconnected node means that it is independent of variables. Bayes' theorem is expressed with the following equation (Haltaş & Alkan, 2013);

$$P(X|Y)=(\mathbf{P}(\mathbf{Y} \mid \mathbf{X})\mathbf{P}(\mathbf{X}))/P(Y)$$

P(X|Y) represents the probability that event X will occur when event Y occurs. P(Y|X) represents the probability of event Y occurring when event X occurs, while P(X) and P(Y) are apriori probabilities of events X and Y. In the Naive Bayes algorithm, it is accepted that it performs well since each feature is considered independent from each other. With its simple and easy structure, very good works can be done with little data, and it can also give very positive results in high-dimensional data. It can be used with continuous and intermittent data as well as unbalanced data.

Support Vector Machine

Support vector machines are one of the supervised learning algorithms generally used in classification problems. SVM developed by Vapnik (1995) has attracted the attention of many researchers recently because it has produced remarkable results. The superiority of this method is due to its ability to generalize better than other models (Min, Lee, & Han, 2006). Tay & Cao (2002) used SVM for estimating the wholesale price index. Fan & Palaniswami (2000) also used this method for bankruptcy prediction. And then Huang et al. (2004) used this method for credit rating.

In the financial distress prediction literature, SVM is considered to be a very powerful and effective method due to its high prediction accuracy. SVM has different processes compared to other models. Its most important feature is its success in variable selection. Because while trying to comprehend the geometric structure of the sample space in the data subject to the research, unnecessary, irrelevant and unrelated data will be encountered. In such a case, more time will be spent and the fit rate of the model will decrease somewhat (Piramuthu, 2004).

SAMPLE AND DATA SELECTION

Data of this paper consists of annual financial statements of Turkish small and medium sized companies operating in Turkey from 2015 up to 2019. The firms in the study were selected from different sectors such as education, energy, furniture, transportation, mining, automotive, textile, tourism. However, construction and contracting companies were not included in the sample. Because it is quite possible that the works in the balance sheets of these firms are not annual or spread over more than one year. We continued to this sudy with companies in sectors whose financial statements have the same time (annual basis) period. All statistical analyzes were made using the python programming language.

Firms for which bankruptcy or concordat decisions were made by the commercial courts in the 2018-2019 period were accepted as distressed. The data collected for distressed firms includes annual data one (2017), two (2016) and three years (2015) before the the judgment date.

Our data consists of 173 non-distressed and 219 distressed firms. Data was randomly subsampled as 70% training set and 30% test set. The training set is used to train the prediction model. The model calculates the "prediction" values from the training result. The model is evaluated by comparing it with the test set, which consists of data that is not included in the training set.

In this study, 47 initial financial ratios were used and the selection of these ratios is based on previous studies (Mselmi, Lahiani & Hamza, 2017; Aksoy & Boztosun, 2018; Yürük & Ekşi, 2019; Kısakürek, Arslan & Bircan, 2018; Ertan & Ersan, 2018; Yazıcı, 2018). These ratios are selected from among the liquidity ratios, financial structure ratios, profitability ratios and turnover ratios.

Table 2 illustrates financial ratios before feature selection, also called initial ratios. From the primary financial ratios, the best subset that could represent the data set for each year before the failure was selected by the feature selection method. Feature selection is widely used in machine learning and statistics. Reducing the number of features has many benefits in data analysis. Reasons such as decreasing the size of the data, increasing the speed of the algorithm, eliminating the irrelevant and noisy data, making the data simpler and more visualizable, reducing the memory required for data storage are some of them. Random Forest - Recursive Feature Elimination (RF-RFE) method was used for feature selection in this research. RF-RFE method, one of the most popular feature selection approaches, is effective in reducing data size and increasing efficiency (Chen, Meng, Liu, Jin & Su, 2018). According to Voyle et al. (2016), RF-RFE has proven to be more effective compared to other methods and can use fewer features to achieve a higher classification accuracy. The RFE method can be also used with other classification algorithms (SVM, DT, etc.) other than RF. With RF-RFE, 8 financial ratios were selected for the first year prior to failure and 9 for the second and third years. The selected financial ratios are summed in table 3.

For one, two and three years before bankruptcy, the most important financial ratios selected by RF-RFE are Stock dependency ratio (R06), net tangible assets/long-term liabilities (R24) and financial expense/net sales (R32). R06 shows how dependent a firm is on inventories to pay for its short-term liabilities. It is seen that the stock dependency ratio of distressed firms is higher than that of non-distressed firms. In other words, this ratio shows that distressed firms are more dependent on stocks for the payment of short-term liabilities. The second most important ratio, R24, shows that the ratio of net tangible fixed assets to long-term debts of distressed firms is low, and it is two times higher in non-distressed firms than in unsuccessful firms. The third most important ratio, R32, shows the ratio of finance expense to net sales. It is seen that distressed firms have financing expenses of approximately 4 percent of their net sales in all three years. In non-distressed firms, it is seen that this rate is between 5 per thousand and 7 per thousand. As expected, it is seen that distressed firms pay higher financial expenses as a result of using more foreign resources.

For one and two years before bankruptcy, the most important financial ratios selected by RF-RFE are long-term liabilities/constant capital (R18) and fixed asset turnover (R46). R18 shows that the ratio of long-term foreign resources to constant capital of successful firms is lower than that of distressed firms. It has been observed that non-distressed firms have stronger capital, while distressed firms use much more foreign resources than their own resources. For two and three years prior to failure, the most discriminant financial ratios selected by RF-RFE are Equities/Total foreign assets (R11) and (Financial expenses+İncome before tax)/Financial expenses (R35). R11 shows that distressed firms have 5 times more foreign resources compared to their own resources, while this rate is about 2 times higher in non-distressed firms. R35, on the other hand, shows that non-distressed firms operate with higher pre-tax profits and lower financing costs than distressed firms.

One year before bankruptcy, the most important financial ratios selected by RF-RFE are bank loans/total foreign assets (R15), operating profits/net sales (R30) and (financial expenses+profit after tax)/financial expenses (R34). R15 shows that the share of bank loans in foreign resources used by successful firms is lower than that of non-successful firms. When we analyze the R30 ratio, we encounter a different situation.

Contrary to what is known the results show that the operating profit margin of distressed firms is higher than that of non-distressed firms. However, in this case, the covering power of net financial debts should be considered. It has been observed that distressed companies grow rapidly with high bank debt, have high operating profits, but after a while they fail to manage their financing expenses. It has been observed that R34 gives similar results to the previously described and closely related R35.

Table 2. Selected initial financial ratios

Variable	Meaning	Variable	Meaning
Liquidity		**Profitability**	
R01	Current ratio	R26	Net income after tax/Net sales
R02	Liquidity ratio	R27	Cost of goods sold/Net sales
R03	Cash ratio	R28	Gross sales margin/Net sales
R04	Stocks/Current assets	R29	Operational expenses/Net sales
R05	Stocks/Total assets	R30	Operating profits/Net sales
R06	Stock dependency ratio	R31	Oper. profits/(Total assets-Fin.tangible assets)
R07	Short-term trade receivables/Current assets	R32	Financial expense/Net sales
R08	Short-term trade receivables/Total assets	R33	(Finan. exp.+Netincome beforetax)/Total Liabilities
Financial structure ratios		R34	(Financial exp.+Profit after tax)/Financial Expenses
R09	Total foreign assets	R35	(Fin.expenses+İncome before tax)/Fin. Expenses
R10	Debt Ratio	R36	Net profit after tax/Equities
R11	Equities/Total foreign assets	R37	Profit before tax/Equities
R12	Short term liabilities/Foreign assets	R38	Net profit after tax/Total assets
R13	Short term liabilities/Total liabilities	R39	(Retained earn.+Reserves)/Tot.assets
R14	Bank loans/Total assets	**Turnover Rates**	
R15	Bank loans/Total foreign assets	R40	Equity turnover
R16	Short term bank loans/Short term liabilities	R41	Working capital turnover
R17	Long-term liabilities/Total liabilities	R42	Net working capital Turnover
R18	Long-term liabilities/Constant capital	R43	Asset turnover
R19	Current assets/Total assets	R44	Accounts receivable turnover
R20	Fixed assets/Equities	R45	Stock turnover
R21	Fixed assets/Total foreign assets	R46	Fixed asset turnover
R22	Fixed assets/Constant capital	R47	Net tangible asset Turnover
R23	Net tangible assets/Equities		
R24	Net tangible assets/Long-term liabilities		
R25	Net tangible assets/Total assets		

Note: This table illustrates the 41 financial variables. These variables represent the initial financial ratios applied to all firms.

Table 3. Financial ratios selected by Random Forest – Recursive Feature Elimination (RF-RFE)

Discriminant Ratios	1-year ahead	2-year ahead	3-year ahead
R6:Stock dependency ratio	✓	✓	✓
R11:Equities/Total foreign assets		✓	✓
R15:Bank loans/Total foreign assets	✓		
R16:Short term bank loans/Short term liabilities			✓
R17:Long-term liabilities/Total liabilities		✓	
R18:Long-term liabilities/Constant capital	✓	✓	
R20:Fixed assets/Equities			✓
R24:Net tangible assets/Long-term liabilities	✓	✓	✓
R30:Operating profits/Net sales	✓		
R32:Financial expense/Net sales	✓	✓	✓
R34:(Financial expenses+Profit after tax)/Financial expenses	✓		
R35:(Financial expenses+İncome before tax)/Financial expenses		✓	✓
R37:Profit before tax/Equities			✓
R42:Net working capital turnover			✓
R44:Accounts receivable turnover		✓	
R46:Fixed asset turnover	✓	✓	

Note: This table presents the financial variables selected by RF-RFE to be used in the prediction of financial distress.

Two years before bankruptcy, the most important financial ratios selected by RF-RFE are long-term liabilities/total liabilities (R17) and accounts receivable turnover (R44). R17 shows that the share of long-term liabilities of distressed firms in total liabilities is quite high. R44, on the other hand, is a financial ratio that shows how many times receivables are collected in an accounting period, and it has been observed that this ratio is high in non-distressed companies. Three years before bankruptcy, the most discriminant financial ratios selected by RF-RFE are fixed assets/equities (R20), profit before tax/equities (R37), net working capital turnover (R42) and short term bank loans/short term liabilities (R16). R20 shows that the fixed assets of the distressed firms are higher than the equities, while the non-distressed firms have fixed assets at the rate of about half of their equity. R37 financial ratio, on the other hand, shows that non-distressed firms have higher pre-tax profits than distressed firms compared to equity. The R42 financial ratio reflects the extent to which businesses achieve sales volume success with their Net Working Capital. As expected, successful companies have a better sales volume. Finally, R16 shows that distressed firms have much more short term bank loans than non-distressed firms.

EXPERIMENTAL RESULTS

Descriptive statistic

In this section, whether the data are normally distributed or not, the comparison of the variable means with the non-parametric (distribution-independent) Mann Whitney U test, the multicollinearity effect (VIF) and other descriptive statistics of the variables are examined. The normality assumption control of the data was done with the Shapiro-Wilk test. As it can be seen in table 4, for one, two and three *years prior* to the financial *distress,* since the p-value is less than 0.05 in all independent variables, it is understood that the distribution is not normal. It has been observed that the data are generally not normally distributed in studies conducted in the field of financial failure (Mselmi, Lahiana & Hamza, 2017; Selimoğlu & Orhan, 2015; Gör, 2016; Toraman & Karaca, 2016; Karadeniz & Öcek, 2019).

In the analysis performed with the Mann-Whitney U-test, which is a statistical comparison of the mean of independently selected variables, it was concluded that there was a statistically significant difference for all variables 1 and 2 years before the failure. 3 *years prior* to the financial *distress,* all independent variables except R37 and R42 are statistically significant.

The majority of correlations between independent variables are low. There is a high correlation (0.95) between R34 and R35. Therefore, R34 was excluded from the variables. Multicollinearity, with its simple definition, is the situation where there is a very high correlation between at least two variables that predict a variable. Our findings show that all independent variables have VIF values less than 10. Therefore, there is no multicollinearity problem between the variables.

Missing values and outliers

One year prior to failure, it has been determined that there are 280 missing values out of 7,990 data in non-distressed firms. The highest missing values were in the independent variables R34 with 41.76%, R35 with 41.76%, R24 with 41.18% and R44 with 25.88%, respectively. There is less missing value in distressed firms. Out of 9776 data, 152 are missing values. The highest loss is in the R24 variable with 13%. There is 2.43% missing data in total (distress and non-distress firms). The dataset was updated by filling in the missing data with the median values of the relevant variables.

Two years prior to failure, the data of distressed and non-distressed companies were examined and it was determined that there were 389 missing values out of 7,943 data in non-distressed companies (independent variables). The highest missing values were in the independent variables R24 with 48.52%, R34 with 44.37%, R35 with 44.37% and R44 with 26.04%. There is less missing value in distressed firms. Out of 9917 data, 127 are missing values. The highest loss is in the R24 variable with 18.48%. There is a total of 2.89% missing value in the data groups of companies that are distressed and non-distressed. The data set was updated by filling in the missing values with the median values of the relevant variables.

Three years before the failure, the data of successful and non-successful firms were examined and it was determined that there were 686 missing value out of 7,802 data in non-distressed firms. The highest missing value is in the independent variables R24 with 59.04%, R34 with 50%, R35 with 50%, R44 with 33% and R45 with 10.24%, respectively. There is less missing value in distressed firms. Of 10,293 data, 415 are missing value. The highest missing values are in R24 with 25.57%, R34 with 12.78% and R35 with 12.78%, respectively. In the data groups of distressed and non-distressed firms, there is a total of 6.08% missing data. The dataset was updated by filling in the missing data with the median values of the relevant variables.

After the missing value analysis, the outliers in the data were taken to their normal limits according to the Tukey method. In this method, when calculating outliers, the median value is found first. For this, the data is sorted from lowest to highest. The second step is to find the median of the data. Then the lower quartile (Q1) and upper quartile (Q3) values are found. In the following process, the values of Q1 and Q3 are multiplied by 1.5 to find the gap of quarters. Thus, the lower and upper bounds of the data are found. All values outside these limits are considered outliers.

MODEL PERFORMANCE COMPARISON AND DISCUSSION

When the models were evaluated according to their 2017 accuracy rates, the most successful prediction model was DT with 90%, RF with 89%, kNN with 87%, SVM and LR with 84%, and NB with 82%.When the models were evaluated according to their 2016 accuracy rates, the most successful prediction model was DT with 97%, RF with 96%, kNN and SVM with 89%, NB with 85%, and LR with 80%, respectively. When the models were evaluated according to their 2015 accuracy rates, the most successful prediction model was NB with 97%, DT and RF with 94%, kNN and SVM with 92%, and LR with 81%, respectively.

DT: We can say that the decision tree model is the most successful model in classification with a general average of 93.67. It is seen that the model with the best prediction in year t-1 and t-2 is the second model with the best prediction in year t-3. The model made its best prediction at t-2 and had a lower accuracy rate at t-1 than the others.

RF: With in overall average correct classification success, RF was the second best model with 93%. In this model, which has the advantages of processing with missing data and working with as many tree structures as desired, the success of correct classification in t-2 and t-3 years is higher. The model achieved the highest correct classification in t2.

SVM: It ranked third with 89.33% in overall classification success. He made the most successful classification in the year t-3. The rate of correct classification of model increased with the distance from the failed period. It showed the lowest performance in t-1 year. Successful companies have a higher rate of correct classification.

kNN: It is in the third rank with 89.33% correct classification success together with SVM in general classification. It has similar accuracy rates as SVM. It

Table 4. Summary statistics

One year priot to distress

Variables	R6	R15	R18	R24	R30	R32	R34	R46
Panel A: Entire data set								
Number of firms	378	378	378	378	378	378	378	378
Mean	1.849.557	0,604377	0,372624	0,918207	0,046092	0,027699	2.500.073	12.164.229
Std. Deviation	1.714.372	0,701298	0,324820	0,736784	0,050256	0,034877	1.733.570	14.300.647
Min	-1.968.473	0,000000	-0,567549	-1.232.655	-0,072043	0,000000	-0,975730	0,000000
Max	6.390.963	2.724.057	1.081.260	2.638.308	0,177656	0,141963	6.542.847	55.452.146
Shapiro-Wilk (p value)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Mann-Whitney U (p-value)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Panel B:Distressed firms								
Number of firms	208	208	208	208	208	208	208	208
Mean	2.332.664	0,860544	0,511647	0,713795	0,060167	0,044018	1.764.553	7.918.617
Std. Deviation	2.052.956	0,793298	0,296164	0,777527	0,055165	0,038967	1.252.516	7.746.015
Min	-1.968.473	0,000000	-0,181926	-1.232.655	-0,072043	0,000000	-0,975730	0,000000
Max	6.390.963	2.724.057	1.081.260	2.638.308	0,177656	0,141963	4.626.431	23.645.576
Panel C: Non-distressed firms								
Number of firms	170	170	170	170	170	170	170	170
Mean	1.258.461	0,290949	0,202524	1.168.313	0,028871	0,007731	3.400.003	17.358.861
Std. Deviation	0,869124	0,383047	0,273927	0,596777	0,036940	0,011171	1.817.144	18.258.983
Min	-0,529976	0	-0,567549	0,166867	-0,050417	0	0,086331	0,32531
Max	2.904.297	1.100.631	0,894771	2.060.690	0,102070	0,032298	6.542.847	55.452.146

Two years priot to distress

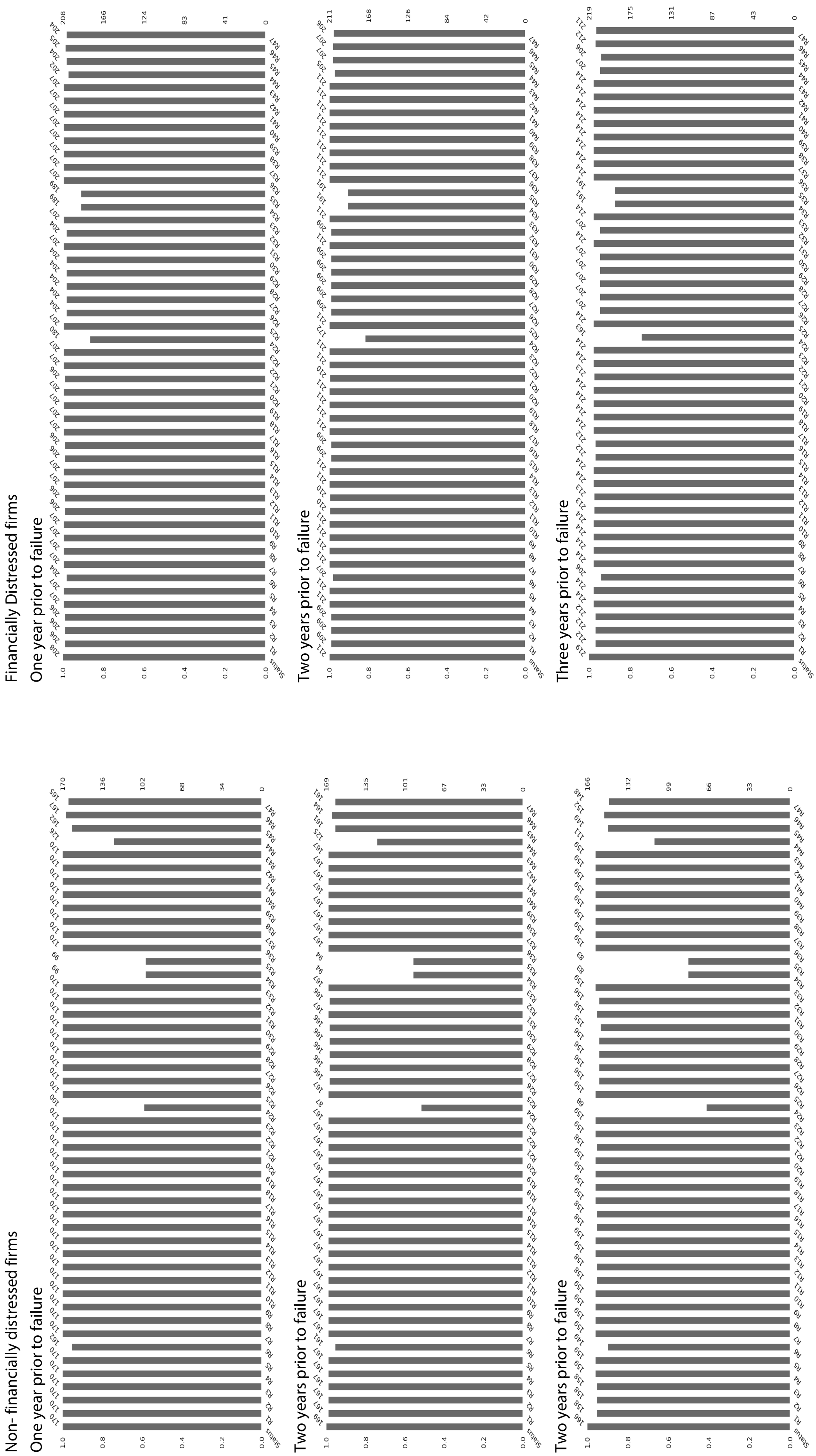
Variables	R6	R15	R17	R18	R24	R32	R35	R44	R46
Panel A: Entire data set									
Number of firms	380	380	380	380	380	380	380	380	380
Mean	1.902.626	0,460397	0,146860	0,319295	0,797053	0,026858	1.891.485	8.305.613	13.586.545
Std. Deviation	1.811.481	0,503204	0,185932	0,340008	0,514782	0,034739	0,958350	7.628.059	16.158.007
Min	-2.418.927	0,000000	0,000000	-0,831633	-0,621116	0,000000	-0,530658	0,000000	0
Max	6.792.352	1.933.519	0,780676	1.610.463	1.867.806	0,146562	3.757.784	27.898.704	65.860.889
Shapiro-Wilk (p value)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Mann-Whitney U (p-value)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Panel B:Distressed firms									
Number of firms	211	211	211	211	211	211	211	211	211
Mean	2.392.867	0,631280	0,210324	0,438052	0,739323	0,042394	1.624.414	6.065.722	8.469.488
Std. Deviation	2.072.580	0,546349	0,212740	0,353286	0,659067	0,039259	1.120.664	5.547.003	8.366.145
Min	-2.418.927	0,000000	0,000000	-0,831633	-0,621116	0,000000	-0,530658	0,000000	0,000000
Max	6.792.352	1.933.519	0,780676	1.610.463	1.867.806	0,146562	3.757.784	17.833.136	25.022.325
Panel C: Non-distressed firms									
Number of firms	169	169	169	169	169	169	169	169	169
Mean	1.290.549	0,247046	0,067623	0,171024	0,869130	0,007462	2.224.928	11.102.164	19.975.297
Std. Deviation	1.163.882	0,340077	0,100213	0,254871	0,213203	0,010688	0,548321	8.867.789	20.677.305
Min	-1.089.725	0,000000	0,000000	-0,253363	0,532658	0,000000	1.398.079	0,830515	0,206249
Max	3.540.690	1.063.317	0,283595	0,775582	1.156.743	0,031778	3.072.351	27.898.704	65.860.889

Three years priot to distress

Variables	R6	R11	R16	R20	R24	R32	R35	R37	R42
Panel A: Entire data set									
Number of firms	385	385	385	385	385	385	385	385	385
Mean	1.664.436	0,377779	0,217862	1.069.187	0,903652	0,024855	2.152.028	0,110036	4.369.192
Std. Deviation	1.434.573	0,408893	0,258234	1.288.109	0,513968	0,033625	1.143.455	0,152928	9.371.304
Min	-1.719.263	-0,701137	0,000000	-2.077.135	-0,717077	0,000000	-1.080.795	-0,195678	-1.330.962
Max	5.377.166	1.673.011	0,963857	4.565.448	2.067.443	0,131258	4.886.129	0,406570	22.859.310
Shapiro-Wilk (p value)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Mann-Whitney U (p-value)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,877	0,527
Panel B:Distressed firms									
Number of firms	219	219	219	219	219	219	219	219	219
Mean	2.075.591	0,232468	0,306650	1.404.374	0,834385	0,039280	1.775.682	0,106594	4.147.905
Std. Deviation	1.621.150	0,196115	0,284628	1.494.190	0,673897	0,038166	1.404.670	0,163007	9.809.513
Min	-1.719.263	-0,231713	0,000000	-2.077.135	-0,717077	0,000000	-1.080.795	-0,195678	-13.309.623
Max	5.377.166	0,649309	0,963857	4.565.448	2.067.443	0,131258	4.886.129	0,406570	22.859.310
Panel C: Non-distressed firms									
Number of firms	166	166	166	166	166	166	166	166	166
Mean	1.122.009	0,569485	0,100725	0,626982	9,950347	0,005824	2.648.532	0,114578	4.661.132
Std. Deviation	0,892650	0,522771	0,154568	0,752247	0,000000	0,008138	0,000000	0,138871	8.780.536
Min	-0,793944	-0,701137	0,000000	-1.102.812	9,950347	0,000000	2.648.532	-0,177186	-11.446.018
Max	2.964.440	1.673.011	0,417816	2.265.504	9,950347	0,021405	2.648.532	0,393072	21.303.075

Note: This table illustrates descriptive statistics of the 16 selected variables.

Graph 1. Visualizing Missing Values



Note: The missing values in this graph are visualized with using the missingno library. Missingno is a Python library that provides the ability to understand the distribution of missing values.

Table 5. Classification performance

	LR			RF			DT			SVM			NB			kNN		
	T-1	T-2	T-3	T-1	T-2	T-3	T-1	T-2	T-3	T-1	T-2	T-3	T-1	T-2	T-3	T-1	T-2	T-3
Accuracy	87	80	81	89	96	94	90	97	94	84	89	92	82	85	97	87	89	92
Precision	87	80	82	89	96	94	91	97	94	84	90	92	82	85	96	87	89	89
Sensitivity	87	79	81	88	97	94	90	97	94	84	89	92	82	85	97	87	89	88
F-measure	87	80	81	89	96	94	90	97	94	84	89	92	82	85	97	87	89	87

Note: This table illustrates the prediction accuracy of LR, RF, DT, SVM, NB and kNN.

reached the highest accuracy at t-3 and the lowest accuracy at t-1. Successful companies have a higher rate of correct classification.

LR: It was the model with the lowest accuracy rate with 81.67% in overall performance. It reached the highest accuracy rate in t-1 year with 84%. It achieved 81% correct classification success in t-3 and 80% in t-2 year.

While the highest correct classification success was achieved in t-3 with 91.67%, this rate was 89.33% in t-2 year and 86% in t-1 year. The models in the t-2 and t-3 have higher accuracy. Although the t-1 year is the closest period to failure, its lower accuracy has pushed us to investigate the dynamics of the Turkish economy in which businesses operate in more detail. It is observed that interest, inflation and dollar/tl exchange rates did not change much in t-2 and t-3 years in Turkey. However, in the year t-1, despite the Central Bank interest rates remaining the same, it is observed that there is a serious movement in inflation and dollar/tl rate. In 2018, the year of failure, these three variables increased dramatically.

In t-1, it is seen that the idle capacity increased more than the other two years (R46), and the share of bank loans in liabilities increased (R15). According to Güngen (2018), reasons such as the 20.5% depreciation of the Turkish Lira from 15 July 2016 to the end of 2017 and the financing of the current account deficit with hot money adversely affected the economic dynamics of the country.

According to Çetin (1996) & Erez (1994), the most important problem of SMEs is financing. The increasing inflation since 2017, the dollar/TL exchange rate and the dramatically increasing interest rates in mid-2018 made it difficult for SMEs to access credit and increased the cost of the credit they used. Compared to 2015 and 2016, SME sector bank loans increased in 2017 (R15). SMEs were most affected by this situation and the NPL ratio of SME loans increased rapidly compared to retail and commercial loans.According to the results of this study, it is understood that cyclical fluctuations have an effect on financial failure. The rapid increase in interest rates, exchange rates and inflation in 2017 and 2018, which had a stable trend in 2015 and 2016, led to a rapid increase in bank loans in distressed firms and reduced their capacity utilization. Contrary to the commercial and consumer customer segments of these businesses, which use loans with weak collateral and high rates, it is observed that NPL ratios have increased rapidly since mid-2017. We believe that it would be more appropriate to consider macroeconomic dynamics when making evaluations about this differentiated group.

CONCLUSION

The financial failure of the enterprises negatively affects the groups related to the enterprise, especially the environment of the enterprise, and the general economic structure. Predicting financial failure has become one of the important issues in the field of finance. The purpose of financial failure forecasting is to develop a forecasting model that will enable to predict the financial condition of a business by using various econometric indicators. Knowing the possibility of financial failure of a business for the creditors and investors of the business has become a very important issue before the decisions to be taken.

Since the 1960s, models that predict financial failure have been tried to be developed using different methods. Especially today, with the dizzyingly developing technological innovations, hardware and software developments of computers, larger and more complex data masses, big data have begun to emerge. With this increasing volumetric growth, traditional statistical estimation models measuring financial failure began to fail to meet the need, and modern methods were needed.

Table 6. Correlation matrix of selected independent variables

2017		*R6*	*R15*	*R18*	*R24*	*R30*	*R32*	*R34*	*R46*
R6		1							
R15		-0,04713	1						
R18		0,023742	0,561236	1					
R24		0,077526	-0,25591	-0,51959	1				
R30		-0,10478	0,21522	0,197285	-0,08493	1			
R32		-0,02582	0,446509	0,407102	-0,1501	0,694608	1		
R34		-0,08637	-0,29287	-0,40681	0,20231	-0,1226	-0,46496	1	
R46		-0,15284	-0,2829	-0,28045	-0,00095	-0,185	-0,31208	0,250055	1
VIF		2,049434	2,847239	3,447713	2,606682	4,050813	4,899684	3,408806	1,737293
2016	*R6*	*R15*	*R17*	*R18*	*R24*	*R32*	*R35*	*R44*	*R46*
R6	1								
R15	-0,0699	1							
R17	-0,10277	0,727466	1						
R18	-0,04129	0,533566	0,794596	1					
R24	0,01099	-0,04646	-0,32634	-0,32257	1				
R32	0,145187	0,502446	0,414569	0,373735	-0,11977	1			
R35	-0,11874	-0,30743	-0,30065	-0,31493	0,113634	-0,48274	1		
R44	-0,3053	-0,09051	-0,07302	-0,11357	0,116626	-0,23219	0,111954	1	
R46	-0,13269	-0,33351	-0,30983	-0,25432	0,008867	-0,32956	0,182065	0,18696	1
VIF	2,061043	4,91414	7,42647	5,08798	3,59912	2,39152	3,79763	2,32642	1,8558
2015	R6	R11	R16	R20	R24	R32	R35	R37	R42
R6	1								
R11	-0,325	1							
R16	0,142	-0,1177	1						
R20	0,0135	-0,2674	0,1102	1					
R24	0,0565	0,1027	0,0147	-0,0559	1				
R32	0,1366	-0,2663	0,3735	0,2213	-0,2251	1			
R35	-0,0648	0,28	-0,2191	-0,2624	0,0922	-0,5298	1		
R37	0,1325	-0,1097	-0,0968	-0,0615	-0,0147	-0,1712	0,3492	1	
R42	-0,1218	-0,0035	-0,0379	-0,1558	-0,013	-0,0821	0,0802	0,0425	1
VIF	2,5499	2,2643	2,0261	1,7035	3,5718	1,9551	4,9539	1,8521	1,2282

Note: This table shows the correlation matrix of the selected independent financial ratios.

Table 7. Confusion Matrix of DT, RF, kNN, SVM, NB and LR

	1 year prior to distress						2 years prior to distress						3 years prior to distress					
	Training sample			Test sample			Training sample			Test sample			Training sample			Test sample		
	0	**1**	**Sum**	**0**	**1**	**Sum**	**0**	**1**	**Sum**	**0**	**1**	**Sum**	**0**	**1**	**Sum**	**0**	**1**	**Sum**
	Random Forest																	
0	100	0	100	90,48	9,52	100	100	0	100	95,08	4,92	100	100	0	100	95,31	4,69	100
	(145/145)	(0/145)		(57/63)	(6/63)		(150/150)	(0/150)		(58/61)	(3/61)		(155/155)	(0/155)		(61/64)	(3/64)	
1	0	100	100	13,73	86,27	100	0	100	100	1,89	98,11	100	2,63	97,37	100	7,69	92,31	100
	(0/119)	(119/119)		(7/51)	(44/51)		(0/116)	(116/116)		(1/53)	(52/53)		(3/114)	(111/114)		(4/52)	(48/52)	
	Decision Tree																	
0	94,48	5,52	100	87,30	12,70	100	100	0	100	98,36	1,64	100	100	0	100	95,31	4,69	100
	(137/145)	(8/145)		(55/63)	(8/63)		(150/150)	(0/150)		(60/61)	(1/61)		(155/155)	(0/155)		(61/64)	(3/64)	
	0	100	100	5,88	94,12		0	100	100	3,77	96,23	100	2,63	97,37	100	7,69	92,31	100
1	(0/119)	(119/119)		(3/51)	(48/51)	100	(0/116)	(116/116)		(2/53)	(51/53)		(3/114)	(111/114)		(4/52)	(48/52)	
	LR																	
0	88,29	11,71	100	87,30	12,70	100	81,33	18,67	100	85,25	14,75	100	84,52	15,48	100	90,63	9,37	100
	(128/145)	(17/145)		(55/63)	(8/63)		(122/150)	(28/150)		(52/61)	(9/61)		(131/155)	(24/155)		(58/64)	(6/64)	
	15,13	84,87	100	19,60	80,40		25,86	74,14	100	26,42	73,58	100	23,68	76,32	100	30,77	69,23	100
1	(18/119)	(101/119)		(10/51)	(41/51)	100	(30/116)	(86/116)		(14/53)	(39/53)		(27/114)	(87/114)		(16/52)	(36/52)	
	Naive Bayes																	
0	85,52	14,48	100	85,71	14,29	100	89,33	10,67	100	85,25	14,75	100	90,97	9,03	100	93,75	6,25	100
	(124/145)	(21/145)		(54/63)	(9/63)		(134/150)	(16/150)		(52/61)	(9/61)		(141/155)	(14/155)		(60/64)	(4/64)	
1	15,13	84,87	100	21,57	78,43	100	18,10	81,90	100	15,09	84,91	100	0	100	100	0	100	100
	(18/119)	(101/119)		(11/51)	(40/51)		(21/116)	(95/116)		(8/53)	(45/53)		(0/114)	(114/114)		(0/52)	(52/52)	
	DVM																	
0	88,28	11,72	100	84,13	15,87	100	89,33	10,67	100	86,89	13,11	100	86,45	13,55	100	90,63	9,37	100
	(128/145)	(17/145)		(53/63)	(10/63)		(134/150)	(16/150)		(53/61)	(8/61)		(134/155)	(21/155)		(58/64)	(6/64)	
1	10,92	89,08	100	15,69	84,31	100	2,59	97,41	100	7,55	92,45	100	0	100	100	5,77	94,23	100
	(13/119)	(106/119)		(8/51)	(43/51)		(3/116)	(113/116)		(4/53)	(49/53)		(0/114)	(114/114)		(3/52)	(49/52)	
	kNN																	
0	86,90	13,10	100	87,30	12,70	100	90,66	9,34	100	88,52	11,48	100	78,06	21,94	100	85,94	14,06	100
	(126/145)	(19/145)		(55/63)	(8/63)		(136/150)	(14/150)		(54/61)	(7/61)		(121/155)	(34/155)		(55/64)	(9/64)	
1	10,92	89,08	100	13,73	86,27	100	3,45	96,55	100	11,32	88,68	100	1,75	98,25	100	0	100	100
	(13/119)	(106/119)		(7/51)	(44/51)		(4/116)	(112/116)		(6/53)	(47/53)		(2/114)	(112/114)		(0/52)	(52/52)	

Note: This table illustrates confusion matrix of LR, DT, RF, SVM, kNN, and NB method. The numbers in brackets represent the number of companies. The results are given in the form of training and testing. 0 illustrates non-successful firms and 1 illustrates successful firms.

Financial distress estimations have always been important in economies such as Turkey, where periodic cyclical fluctuations are high, and it has turned into a field on which intensive studies are conducted. Mistakes that can be made in this area may cause the creditors to incur losses. This will adversely affect the shareholders, depositors, employees, the public and the general economic structure in financial institutions.

In this study, the financial statements of 392 SMEs operating in Turkey, 173 of which are non-distressed and 219 of which are distressed, between the years 2015-2019 were examined. It is aimed to find the model with the highest correct prediction power that can be valid 1, 2 and 3 years before the failure by using logistic regression, decision tree, random forests, support vector machines and k nearest neighbor methods to predict failure. SMEs for which bankruptcy or concordat decisions were made by the competent courts in 2018 were deemed unsuccessful.

Based on the annual Income and Corporate Tax returns approved by the Ministry of Finance, 47 financial ratios for all businesses were calculated separately for each year. The study was continued with the data pre-processing step and missing data analysis was performed. A total of 2.43%, 2.89% and 6.08% missing data were encountered for the years 2017, 2016 and 2015, respectively. Missing data are filled in with the median value of each independent variable. Afterwards, the extreme value analysis of each independent variable was made using the box-plot method, and the negative effect of the extreme data on the analysis was tried to be reduced.

As a result of the study, the highest percentage of correct classification 1 and 2 years before the failure belonged to the decision tree model with an average of 90% and 97%, respectively. Three years before the failure, the Naive Bayes has the highest level of average accuracy of 97%. It has been observed that the models categorize distressed firms better than non-distressed firms in every 3 years examined.

When the accuracies of all models were examined, the highest correct prediction were obtained in the year t-3, which is the year farthest from the failure year. Although the t-1 year was the closest to failure, its lower accuracy prompted us to further investigate the dynamics of the Turkish economy in which businesses operate. Macroeconomic indicators (interest, inflation, exchange rate), which acted similarly in 2015 and 2016, increased dramatically in 2017 and 2018. In 2017, it is seen that the share of bank loans in foreign resources (R15) and idle capacities of distressed firms increased (R46). It has been observed that the loan repayment capacity of SMEs, which have unique disadvantages such as weak collateral structures and high rate of borrowing, is not as durable as commercial firms and individual consumers.

To summarize, it was observed that there was no problem in the dynamics of the general economic structure of Turkey in 2015 and 2016. The deterioration that started in 2017 reached its peak in 2018. Distressed firms grew rapidly with high bank loans and gained high operating profits. However, these firms with low equity could not manage their financial expenses after a while. Distressed firms, which could not make a profit with high bank indebtedness and high financing expenses, failed in 2018 because they could not financially translate themselves.

In this study, it was tried to predict failure by considering only financial ratios. However, management skills in businesses, education and age of business owners or shareholders, business management information systems, short and long-term strategies, business efficiency, concentration on customers, supplier reliability, relations with banks, customers' check and promissory notes, credit occupancy, credit record Many variables such as bureau scores, external economic environment, unemployment rates, political and economic stability of the country, increasing the sample size used, country interest, exchange rate and inflation rates, sector return averages, exchange rates and growth rates also have an effect on failure. It is thought that including these variables in the model can be beneficial in terms of obtaining more precise and reliable results.

REFERENCES

Aksoy, B. (2018). A comparison of data mining methods in financial failure prediction of businesses: An application in BIST. *(Doctoral thesis, Social Sciences Institute)*. University of Erciyes, Kayseri.

Aksoy, B., and Boztosun, D. (2018). Financial Failure Prediction by using Discriminant and Logistics Regression Methods: Evidence From BIST Manufacturing Sector. *The Journal of Finans Politik & Ekonomik Yorumlar*, (646), p. 9–32.

Aksoy, B., & Boztosun, D. (2019). *Comparison of Financial Failure Estimation and Classification Performance Using Machine Learning Methods in Manufacturing Companies: The Example of Borsa İstanbul.* Proceedings of the 2nd International Banking Congress. Çorum province, p. 11–18. Isbn:9786055244156.

Aktaş, R., Doğanay, M., & Yıldız, B. (2003). Prediction of Financial Failure: Comparison of Statistical Methods and Artificial Neural Network. *The journal of Ankara University Social Sciences Enstitute.* 58(4), 3–24. Ankara https://doi.org/10.1501/sbfder_0000001691

Altman, E. I. (1968). The Prediction of Corporate Bankruptcy: A Discriminant Analysis. *The Journal of Finance*, 23(1), 193. https://doi.org/10.2307/2325319.

Aziz, M. A., & Dar, H. A. (2006). Predicting corporate bankruptcy : where we stand ? *Corporate Governance*, 6(1), 18–33. https://doi.org/10.1108/14720700610649436

Bddk (2019). Türk Bankacılık Sektörü Temel Göstergeleri Mart 2019, Erişim adresi: https://www.bddk.org.tr/ContentBddk/dokuman/veri_0014_40.pdf, Erişim Tarihi: 03.03.2021

Beaver, W., H. (1966). Financial Ratios as Predictors of Failure, *Journal of Accounting Research,* (4):71-102.

Bloomberg (2018), retrieved 01.08.2021, from https://businessht.bloomberght.com/ekonomi/haber/1866855-turkiye-de-yeni-sirketin-omru-5-yil date

Ceran, M. & Ergün, B., S. (2019). Scientific approach to the determination of problem loans in Banking industry. *Journal of Research in Economics* , 3 (1) , 1-18 . Retrieved from https://dergipark.org.tr/tr/pub/jore/issue/44898/559000

Chen, Q., Meng, Z., Liu, X., Jin, Q., & Su, R. (2018). Decision Variants for the Automatic Determination of Optimal Feature Subset in RF-RFE, *Genes* 2018, *9*(6), 301; https://doi.org/10.3390/genes9060301

Civan, M. & Dayı, F. (2014). Financial Failure Prediction In Health Care Organization Using Altman Z-Score And Artifical Neural Network Model. *Akademik Bakış Uluslararası Hakemli Sosyal Bilimler Dergisi* , (41) , 0-0 . Retrieved from https://dergipark.org.tr/en/pub/abuhsbd/issue/32979/366612

Civan, M., & Dayı, F. (2014). Altman Z Skoru ve Yapay Sinir Ağı Modeli ile Sağlık İşletmelerinde Finansal Başarısızlık. *Akademik Bakış Dergisi*, 41.

Çebi, C. B., (2020). Rastgele orman algoritması, Erişim adresi: https://medium.com/@cemthecebi/rastgele-orman-algoritmas%C4%B1-1600ca4f4784, Erişim Tarihi: 02.01.2021.

Çelik, M. K. (2009). Comparative Analysis of Financial Failure Forecast Models for Firms in the IMKB. *(Doctoral thesis, Social Sciences Institute)*. Black Sea Technical University, Trabzon.

Çetin, C. (1996). Yeniden Yapılanma, Girişimcilik, Küçük ve Orta Boy İşletmeler ve Bunların Özendirilmesi. s. 165, İstanbul: Der Yayınları.

Erez, Y. (1994). Orta ve Küçük İşletmeler. *Ekonomik Forum Dergisi*, TOBB, 11, 3-6.

Ergin, H. (1999). *The Use of Artificial Neural Networks in the Prediction of Financial Failure and an Empirical Application.* (Doctoral thesis, Social Sciences Institute). University of Dumlupınar, *Kütahya*.

Ertan, A. S., & Ersan, Ö. (2019). Determinants of Financial Default: The Case Of Manufacturing Industry In Turkey. *Marmara University Journal of Economics and Administrative Sciences.* 40(2), 181–207.

Fan, A. & Palaniswawİ, M. (2000). Selecting Bankruptcy Predictors using A Support Vector Machine Approach, Proceeding of the international joint conference on neural network, 6, 354–359.

Gör, Y. (2016). A Research On The Place Of Corporate Governance Financial Failure Prevention. *The Eurasian Journal of Researches in Social and Economics*, 5(12), p: 689–697.

Güngen, A. R. (2018). Türkiye'nin 2018 Krizi: Nereden Nereye? *Mülkiye Dergisi*, 42 (3), 449-452.

Haltaş, A. & Alkan, A. (2013). İmmunohistokimyasal Boyalar ile Tiroid Tümörü Teşhisinde Naive Bayes Algoritması Kullanılması, Ankara Bilişim Konferansı, Ankara.

Hesarı, S. (2018). *Financial Failure Prediction: A Review on Artificial Neural Network and Decision Tree Methods* (Master thesis, Social Sciences Institute). University of Dokuz Eylül, İzmir.

Huang, Z., Chen, H., C., Hsu, C., J., Chen, W., H., & Soushan, W. (2004). Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study. Decision Support Systems, 37, 543–558.

Institue Of International Finance (2021), Global debt monitör COVID drives debt surge-stabilization ahead ?, World Bank. Retrieved, 03.08.2021, from https://www.iif.com/Portals/0/Files/content/Global%20Debt%20Monitor_Feb2021_vf.pdf

İçerli, M.Y., (2005). Prediction of Financial Failure in Businesses and an Application. (Doctoral thesis, Social Sciences Institute). University of Dokuz Eylül, İzmir.

Kalaycı, Ş. (2009). SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri., s. 273, Asil yayınlar, Ankara.

Karadeniz, E., & Öcek, C. (2019b). Finansal Başarısızlık Riski Taşıyan ile Taşımayan İşletmelerin Finansal Oranlarının Karşılaştırmalı Analizi: Borsa İstanbul Turizm İşletmelerinde Bir Araştırma. *Seyahat ve Otel İşletmeciliği Dergisi*, 16(2), 191–206.

Kısakürek, M. M., Arslan, Ö., & Bircan, H. (2018). Model Proposal For Financial Failure of the Business Forecasting: An Application Inthe BIST Engaged In Manufacturing Bussiness. *Journal of Kahramanmaraş Sütçü İmam University Faculty of Economics and Administrative Sciences*, 8(1), 99–114.

Lantz B., (2019). Machine Learning with R: Expert Techniques for Predictive Modeling. 3rd ed. Birmingham, UK, Packt Publishing Ltd.

Min, S. H., Lee, J. & Han, I. (2006). Hybrid Genetic Algorithms and Support Vector Machines for Bankruptcy Prediction, Expert Systems with Application, (31), S.652-660.

Mselmi, N., Lahiani, A. & Hamza, T. (2017). Financial distress prediction: The case of French small and medium-sized firms, *İnternational Review of Financial Analysis,(50), 67-80.*

Ögündür, G., (2020). Overfitting (Aşırı Öğrenme), Underfitting(EksikÖğrenme)veBias-VarianceÇelişkisi. Erişimadresi:https://medium.com/@gulcanogundur/overfitting-a%C5%9F%C4%B1r%C4%B1-%C3%B6%C4%9Frenme-underfitting-eksik-%C3%B6%C4%9Frenme-ve-bias-variance-%C3%A7eli%C5%9Fkisi-b92bef2f770d Erişim tarihi: 23.12.2020.

Özdamar, K. (2002). Paket Programlarla İstatistiksel Veri Analizi, 4. Baskı, s.475, Eskişehir, Kaan Yayınları.

Paket, H. (2014). *Prediction of Financial Failures of Businesses Traded in Borsa Istanbul: A Comparative Application with Artificial Neural Networks and Discriminant Analysis Methods.* (Master thesis, Social Sciences Institute). University of Süleyman Demirel, Isparta.

Yayımlanmamış Yüksel Lisans Tezi, Süleyman Demirel Üniversitesi/Sosyal Bilimler Enstitüsü, Isparta. https://doi.org/10.1017/CBO9781107415324.004

Patrick, P.J.F. (1932). A Comparison of Ratios of Successful Industrial Enterprises with Those of Failed Firms, *Certified Public Accountant*, p. 656- 662.

Piramuthu, S. (2004). Evaluating Feature Selection Methods for Learning in Data Mining Application, *European Journal of Operational Research, 156,483-494.*

Selcik, S. (2019). *Financial bankruptcy prediction: An application in the BIST* (Master thesis, Social Sciences Institute). İstanbul University, İstanbul.

Selimoğlu, S., & Orhan, A. (2015). Measuring Business Failure by Using Ratio Analysis and Discriminant Analysis: A Research on Textile, Clothes and Leather Firms Listed In The Istanbul Stock Exchange. *The Journal of Accounting anf Finance*, 66, 21–40. https://doi.org/10.25095/mufad.396529

Shannon, M., D. & Davenport, A., M. (2001). Using SPSS to Solve Statistical Problems: A Self-Instruction Guide, s. 287-288, Prentice Hall, ABD.

Soo, Y., K. & Upneja, A. (2014). Predicting Restaurant Financial Distress Using Decision Tree and AdaBoosted Decision Tree Models, Economic Modelling, (36), s.356.

Şengören, F. (2019). *Predicting the financial success and failure: Comparison of the logit regression and support vector machine*. (Master thesis, Social Sciences Institute). TOBB University, Ankara.

Tay, F., E., H. & Cao, L. (2001). Apllication of Support Vector Machines in Financial Time Series Forecasting, Omega, (29),309-317.

Toraman, C., and Karaca, C. (2016). Financial Failure Prediction Of The Firms Operating In Chemical Industry: Evidence From Istanbul Stock Exchange. *The Journal of Accounting and Finance,* (70), p:111–128.

Torun, T. (2007). *Comparison of Artificial Neural Networks with Traditional Statistical Methods in Financial Failure Prediction and Application on Industrial Enterprises*. (Doctora thesis, Social Sciences Institute). University of Dokuz Eylül, İzmir.

Tuncay, Y. (1998). Mali Oranlar ve Diskriminant Modeli ile Aracı Kurumların Mali Açıdan Başarılı-Başarısız Olarak Sınıflandırılması, *Yeterlilik Etüdü*, Sermaye Piyasası Kurulu, s.38-40. Ankara.

Ural, K. (2020). *Comparison of Bankruptcy Forecasting Methods: Application in Borsa İstanbul Manufacturing Companies.* (Doctora thesis, Social Sciences Institute). University of Yaşar, İzmir.

Uzun, E. (2020). Decision Tree (Karar Ağacı): ID3 Algoritması - Classification (Sınıflama) – Erişim adresi: https://erdincuzun.com/makine_ogrenmesi/decision-tree-karar-agaci-id3-algoritmasi-classification-siniflama/ *Erişim Tarihi: 01.01.2021*

Vapnik, V., N. (1995). The Nature of Statistical Learning Theory. New York: SpringerVerlag.

Voyle, N., Keohane, A., Newhouse, S., Lunnon, K., Johnson, C., Soininen, H., Kloszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., et al. A pathway based classification method for analyzing gene expression for Alzheimer's disease diagnosis, Journal of Alzheimer's Disease, 2016, 49, 659–669.

Yakut, E., & Elmas, B. (2013). Estimating Financial Failure of Enterprises with data Mining and Discriminant Analysis. *Afyon Kocatepe University Journal of Economics and Administrative Sciences*, XV(I), 261–280.

Yazıcı, M. (2018). Comparison of Discriminant Analysis, Logistic Regression and Artificial Neural Networks in Credit Risk Analysis. *Journal of Finance Letters*, 1(109), 91–106.

Yürük, M. F., & Ekşi, H. İ. (2019). Financial Failure Prediction of Companies Using Artificial Intelligence Methods: An Application in BIST Manufacturing Sector *Mukaddime*, 10 (1), p:393–422.

www.Udemy.Com