A research on the comparison of classification algorithm in finance

Una investigación sobre la comparación del algoritmo de clasificación en finanzas

Serkan Eti*[1] , Mehmet Nuri İnel

Marmara University, Turkey

Abstract

Nowadays, data is developing with technology. With developments in the Internet and technology, it makes a warehouse that changes the data more rapidly and in volume. This situation increases the importance of data mining every day. The aim of this study is to examine classification algorithms which are the sub-subject of data mining. In the analysis part of the study, it is aimed to compare the algorithms with the accuracy ratio by using the financial data. As a result of the study conducted on stock marketing, the classification of enterprises of decision tree models with artificial neural networks has also been found to be more successful than other methods.

JEL code: C88, G17
Keywords: Data mining; Machine learning; Profitability; Financial ratio; Compare of methods

---

* Corresponding author.
E-mail address: : serkaneti@yandex.com (S. Eti).
Peer Review under the responsibility of Universidad Nacional Autónoma de México.

Resumen

Hoy en día, los datos se están desarrollando con tecnología. Con los desarrollos en Internet y la tecnología, se crea un almacén que cambia los datos más rápidamente y en volumen. Esta situación aumenta la importancia de los datos de la minería todos los días. El objetivo de este estudio es examinar los algoritmos de clasificación que son el sub -objetivo de datos de la minería. En la parte de análisis del estudio, el objetivo es comparar los algoritmos con la relación de precisión utilizando los datos financieros. Como resultado del estudio realizado sobre la comercialización de acciones, también se ha encontrado que la clasificación de empresas con decisión modelos de árbol con redes neuronales artificiales es más exitosa que otros métodos.

Código JEL: C88; G17
Palabras clave: Minería de datos; Aprendizaje automático; Rentabilidad; Ratio financiero; Comparar de métodos

Introduction

One of the main objectives of the business is to profit. For this purpose, the company conducts many activities and situations you want to know the sources of profit. Intensive and detailed study of this subject is made in both research and academic work in the real sector. Investors use many methods to make decisions (Eti, 2019). Profit status of the business dealt with theoretical and empirical studies have tried to reveal the scientific method, various methods of scientific results stemming from differences were obtained. In this study, profit conditions of enterprises other than holding and finance companies in BIST100 have been examined by classification algorithms by taking advantage of machine learning techniques which have become more important with the development of computer and technology in recent years. Looking at the financial ratios, it was aimed to investigate whether it is possible to classify whether or not an entity is profitable. Thus, the financial ratios obtained from the financial statements of the companies which are divided into two classes as profit and non-profit companies have been tried to be classified.

The relationship between financial ratios and profit has been shown in different previous studies. In the literature, the relationship between profitability and profitability of an enterprise with financial statements is tried to be explained with financial ratios. Nissim and Penman (2003) examined the impact of commercial and financial debts on profitability with 38-year data from manufacturing enterprises traded on the New York exchange. Omran and Ragab

(2004) examined the relationship between return on equity and return on equity by using the financial ratios of the 46 enterprises in Egypt in 1996-2000. Chen and Zhau (2005) investigated the effect of total sales on profitability by taking the natural logarithm of total sales and found a positive effect. Solano and Teruel (2006) showed that the 7-year data obtained from enterprises will have a positive effect on profitability, inventory turnover and cash conversion period. Albayrak and Akbulut (2012), on the service of ISE service and industrial enterprises also looked at the effect of the 18 financial ratio profitability. Büyükşalvarcı (2010) examined the effect of liquidity ratios, activity rates, financial structure ratios, and stock market performance rates on profitability ratios in the study carried out on manufacturing enterprises which are traded on the Istanbul Stock Exchange. Oruç (2010) analyzed the effect of financial indicators on stock returns in ISE100 indexed firms. The objective of this study is to determine the stock returns of the next period with the help of asset turnover, equity total asset ratio, equity profitability, sales size, asset growth, and market value book value ratios. Karadeniz and İskenderoğlu (2011) used integrated regression analysis to determine the variables affecting the active profitability of tourism enterprises. The effect of asset size, net working capital, receivable turnover rate, inventory turnover rate, a market share of the enterprise in the sector and the active turnover on asset profitability was investigated. Akbulut (2011) examined the relationship between capital management and profitability in the ISE-traded manufacturing sector enterprises, and in the study of asset profitability as a dependent variable, the average collection period of receivables, stock turnover, cash cycle, asset size, growth rate, and leverage ratio were considered as independent variables.

Data mining and machine learning

Data mining is the technique of extracting valid information that is not known before from large datasets. With the development of the computer and the Internet, the rapid increase in data sets played an important role in the development of data mining. The feature of this large data is defined by the so-called 5V concepts: Volume, Velocity, Variety, Verification, and Value. Nowadays, the concept of reality, volatility, validity, precision and variability is added to the big data concept and a wider definition is made. (Atalay & Çelik, 2017). For the analysis of large data sets, artificial intelligence methods and machine learning algorithms have been developed as well as statistical analyzes. Machine learning is divided into two categories as supervised and unsupervised learning techniques. Basically, algorithms are being developed on association, clustering, classification, prediction models. (Silahtaroğlu & Ergül, 2016).

Classification is based on the estimation or prediction of the class of one of the categorical variables in the variables in the dataset. Classification algorithms are controlled learning models where the dependent variable is categorical. Classification algorithms, statistical-based algorithms, logistic regression, separation analysis, artificial neural networks, genetic algorithms, support vector machines, fuzzy rule (logic), random forest algorithms and decision trees are methods. (Silahtaroğlu & Ergül, 2016; Atalay & Çelik, 2017). These methods and their use in literature such as profitability and bankruptcy are summarized in the following section.

Logistic regression is a generalized linear model that accepts the logarithm of a chance of an event as a dependent variable (Agresti, 2018; Yüksel et. al., 2016). Π is including the possibility of belonging to a class, a general mathematical equation of the logistic regression model is,

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon$$

where β_i is the coefficient of regression and shows the effect of independent variables on the logarithm of the chance of belonging to the class. If the calculated probability of the general mathematical expression is higher than the determined cutoff value, it is assigned to class 1 and below class 0. Usually, this cut point is taken as 0.5 (Eti & Ínel, 2016). Separation analysis is the classification analysis by using the linear combinations of groups with more than one variable to make the difference between the grade averages. These two methods are based on statistics and are based on probability theory. In the analysis of separation, the covariance matrix is homogeneous, there are no multiple correlations, the relationships between the independent variables are linear, there are no contradictory values and there is a multivariate normality assumption and these assumptions are more flexible in logistic regression. (Alpar, 2017). Fisher allocation function including a number of arguments p and class k,

$$C_j = c_{jo} + c_{j1} x_1 + c_{j2} x_2 + \cdots + c_{jp} x_p \ (j = 1,2, \ldots, k)$$

and, assigned to the class according to the maximum Cj value. Altman used financial ratios for insolvency by using separation analysis in his study (Altman, 1968). Kumaş and partners have benefited from the logistic regression but they do try to demonstrate the theory of the relationship between firm size and labor market segmentation with microdata from Turkey were modeled profit and non-profit companies (Kumaş et. al., 2014). Jabeur used logistic regression to predict the bankruptcy of companies using financial ratios. In the study, partial least squares were used to estimate the logistic regression parameters (Jabeur, 2017). Dieguez

et al. compared the decision trees and logistic regression models for failure models in their study with financial ratios. They showed that the CART algorithm was more advantageous and better in performance (Irimia-Dieguez, 2015). Han and his partners in the analysis of online workforce while working on separation analysis (Han, 2018). Rodrigues and Rodrigues in Brazil have been established classification model with the help of financial ratios, debt and profitability using clustering and separation analysis in the economic financial performance of the sugarcane energy industry (Rodrigues & Rodrigues, 2018).

Artificial neural networks are the learning of a machine consisting of input, hidden and output layers inspired by the work of the human brain. Two types of artificial neural network models can be established as feeder and feedback. Statistical models do not require a pre-analysis assumption. Basic neural structure and basic artificial neural network model is given in Figure 1.
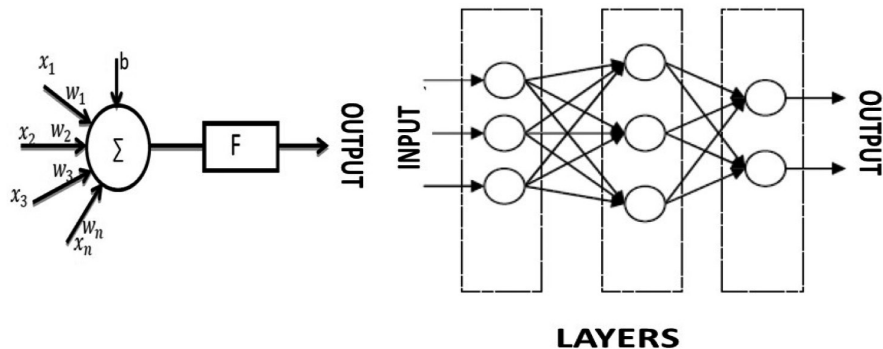

Figure 1. Basic neural structure and basic artificial neural network model

Hosaka has developed a model for predicting bankruptcy by looking at the financial ratios of companies with an artificial neural network (Hosaka, 2019). Jafarian and colleagues studied neural networks with new cost functions and developed a general neural network approach for a fractional order problem in their study (Jafarian et. al., 2018). Genetic algorithms are another machine learning algorithm which is developed by inspiration from artificial neural networks. In this algorithm, each individual is trying to obtain optimal individuals by passing mutations. In a study that used genetic algorithms for the bankruptcy of companies, over 99% could be predicted correctly (Zelenkov et. al., 2017). In another bankruptcy study, fuzzy clus-

tering was used with hybrid genetic algorithms and bankruptcy was predicted with financial ratios and profitability (Chou et. al., 2017).

Support vector machines are a classification algorithm developed with the help of math based vector spaces. Using linear and non-linear structures, it is a supervised learning technique that allows dividing the vector space into the most appropriate classes with the help of the available data. For support vector systems, two categorical dependent variables are determined, two are negative and one is positive. Thus, the output of the model has a negative level of -1 and a positive output of -1 results in +1. It allows space between -1 and +1. A support vector machine is,

$$M_{a,w_0}(q) = \sum_{i=1}^{s}(t_i a[i](d_i q) + w_0)$$

Where ti is the dependent variable, q is a set of descriptive properties, w0 is the first weight of the decision limit, and a set of parameters determined in the learning process. This method is considered a constrained quadratic optimization problem (Kelleher et. al., 2015). In a study on the efficiency of Chinese banks in the global financial crisis, RIA and support vector machines were used as estimators, thus allowing the separation of banks as low/high efficiency. For the classification, financial ratios, profit, number of employees, debts, etc. were obtained and the result was a successful performance with 85% accuracy and 84% sensitivity (Chen et. al., 2018). Elkano and his friends made use of fuzzy mathematics in his study. Thus, a high performance positive and negative classification is obtained by using a set of fuzzy rules for large data sets (Elkano et. al., 2018).

Random Forest algorithm is a median computational combination of subspace sampling, decision trees and bagging methods (Kelleher et. al., 2015). This method is a learning algorithm that makes classification from their prediction by producing multiple classifiers (Pınar et. al., 2017). Ye and colleagues have combined the genetic algorithms with the random forest algorithm to evaluate the credit score for P2P credits. For this purpose, financial data was used and variables were formed through debts and derivatives (Ye et. al., 2018).

Decision trees are the algorithms for creating a sequence of rules with the highest knowledge gain. It provides the possibility of classification by building a model consisting of branches and leaves. There are many algorithms according to a data structure and tree structure such as CART, ID3, C4.5, and C5. The financial hardship of a restaurant business can be estimated by looking at the financial ratios by using decision tree models (Kim & Upneja, 2014). In another similar study, decision tree and survival analysis techniques were compared to predict financial distress. The conditions of these two methods have been examined, different financial

hazard estimates have been found to have different benefits. While nonparametric decision trees perform well inaccurate estimation, survival analysis has performed better to make estimates of varying lengths and to analyze financial hardship over time (Gepp & Kumar, 2015). Roy and his friends in their study in 2019 decision tree model for classification were built (Roy et. al., 2019). Emir et. al. (2012) studied artificial neural networks and support vector machines. As a result of the study, it was seen that support vector machines performed better.

In addition, classification algorithms not included in the study can be analyzed. An example of this is the study of migration with the Logit model (Yüksel et. Al., 2016). Apart from the classification, it has been seen in the literature that these models are also used for other purposes. In addition, integrated versions of these algorithms are included in the studies. Nageswari et. al. (2019) studied student performance. They used decision tree, artificial neural networks, naive bayes, support vector machine and k-nearest neighbor algorithms. The best result has yielded decision trees and neural networks. The performance of these two algorithms has been reported in the literature. Therefore, recent studies have begun to use the hybrid structure of the two algorithms (Pu et. al., 2019; Maji & Arora, 2019).

Ratio analysis

The ratio analysis involves the establishment, measurement, and interpretation of the relationship between the items in the financial statements. By analyzing the performance of an enterprise in the past and the current period by analyzing, it can provide forecasting for the future and information that will shed light on the future planning studies (Aydın et. al., 2012). The primary objective of financial reporting is to accurately measure the financial position and financial position of a company through financial statements. The purpose of financial reporting is to obtain cheap capital (Wallace, 2008).

Basically, four rate types can be calculated from financial statements, including operating rates, liquidity ratios, financial ratios, and financial structure ratios. Activity ratios, sales size, asset turnover, capital/total resource ratio, average collection time of receivables, business size and net sales / fixed asset ratios are covered current ratio and acid-test ratio are considered as liquidity ratios, while leverage ratio is considered as short-term leverage ratio, long-term leverage ratio, and equity turnover rate. Financial structure ratios are debt and financial expense ratio (Eti, 2006).

The financial ratio refers to the mathematical relationship between the two items, or the transformed mathematical relationship, and these ratios are used instead of the raw financial data in the analysis. In order to measure the effect of sales size on the profitability of enterprises, the natural logarithm of sales is used (Oruç, 2010). The acid-to-test ratio is the ratio

of the assets that are reduced to short-term liabilities. This ratio is expected to be one but in Turkey, an acceptable tolerance limit is between 0,80 and 1,20 (İnel & Armutlulu, 2017). The high rate of leverage, which indicates the rate at which business assets are earned from foreign sources, indicates that the entity is financed in a risky manner (Nikolas et. al., 2002). The current ratio, which is used to measure the relationship between current assets and short-term liabilities, is widely used to measure short-term solvency and shows how much the entity has a rotating asset against its short-term foreign source of 1 TL. For the liquidity to be sufficient, this ratio is 2. Turkey's tolerance limits for this ratio from 1.60 to 2.40 are acceptable. (İnel & Armutlulu, 2017). The financial expense ratio is a weight that shows the weight of financing expenses in total foreign resources (Kısakürek & Aydın, 2013). The Company plays an important role in ratio analysis, indicating the ratio of the assets to which they are financed by short-term foreign resources and by which amount they are financed by long-term foreign sources. The rate of active turnover is a ratio of the ratio of sales to total assets and it is a rate that is desired to be high (Karadeniz & İskenderoğlu, 2011; Omran & Ragab, 2004). A proportion of how much of the total resources are provided by the business owners is another positive rate that is desired to be higher than the financial strength of the enterprise for creditors (Aktan & Bodur, 2006). Another ratio is the average collection period of the receivables and explains the relationship between the receivables in the balance sheet and sales in the income statement. A low turnover rate indicates that the entity has difficulty collecting (Akbulut, 2011). Another ratio showing the effect of debt on profitability is the debt ratio (Okuyan, 2013). Net sales / fixed assets ratio, which is mentioned as the turnover rate of fixed assets in some sources, is a ratio used to measure the level of investments in fixed asset (Omran & Ragab, 2004). The rate of equity turnover rate is calculated as the ratio of the sales to equity capital (Karaca & Başçı, 2011). One of the ratios showing the size of the enterprise is taking the logarithm of the assets (Akhtar, 2004; Chen & Zhao, 2005).

Research methods

Turkey is a developing country. Thus, Turkey is the focus of attention of foreign investors. Therefore, investors come to the country to invest. These investors evaluate their investments by weight in the stock market. The 100 companies with the highest shares are listed on the BIST100. BİST100, Turkey's economic performance is the most important indicator built and the financial system. Therefore, those who want to invest in Turkey to consider the index BİST100. For this reason, companies in BIST100 are taken as the sample of the study.

The 2017 financial statements of companies listed in BIST100 were collected from the Public Disclosure Platform (KAP). Financial ratios from the financial statements of 70 companies except for holding and finance companies were calculated by using the formulas given in Table 1 and the variables were obtained. These variables were selected from the literature review mentioned in the first section and shown to have an effect on profitability or profit. The aim of the study is to show the applicability of classification algorithms in the literature in the field of finance. In addition, it is aimed to compare the applied methods over the accuracy rate. The variables used for analysis and calculation formulas are given in Table 1. According to the profitability of the companies, the class variable is coded as 1 and 0 to obtain the binary variable. Thus, two classes were obtained as profitable companies and non-profitable companies.

Table 1

Variables and Their Formulas

Variables	Formulas
Sales Size	$Ln(Sales)$
Liquidity Ratio	$\dfrac{Current\ Assets\ -\ Stocks}{Short-Timeterm\ Foreign\ Resources}$
Leverage Ratio	$\dfrac{Total\ Foreign\ Resources}{Total\ of\ Liabilities\ (Assets)}$
Current Rate	$\dfrac{Current\ Assets}{Short-TimetermForeign\ Resources}$
Short-term Leverage Ratio	$\dfrac{Short-Term\ Foreign\ Resources}{Total\ Liabilities\ (Assets)}$
Financial Expense Rate	$\dfrac{Financing\ Expenses}{Total\ Liabilities}$
Net Sales / Fixed Assets Ratio	$\dfrac{Net\ Sales}{Fixed\ Assets\ Ratio}$
Asset Turnover	$\dfrac{Net\ Sales}{Total\ of\ Assets}$
Debt Ratio	$\dfrac{Total\ Debt}{Total\ Assets}$
Average Collection Period Receivables	$365x\ \dfrac{Short-Term\ Trade\ Receivables}{Net\ Sales}$
Equity Turnover	$\dfrac{Net\ Sales}{Equity}$
Capital / Total Resource Ratio	$\dfrac{Capital}{Total\ Resource\ Ratio}$
Long-term Leverage Ratio	$\dfrac{Long-Term\ Foreign\ Resources}{Total\ Liabilities\ (Assets)}$
Business size	$Ln(Assets)$
Profitability Status	$0-1$

When the previous studies are examined, profitability is tried to be explained through the methods used or it is seen that these methods are compared in pairs. We did not find any study that used these methods together for profitability. For this purpose, in order to determine whether a company is a profitable enterprise with the help of the financial ratios which are frequently included in the literature, classification has been made with data mining techniques. In order to implement data mining techniques, data set was divided into two parts, 50% of the data for learning and 50% for the test. The analyzes were performed in the KNIME and MATLAB programs. Models developed with learning data were applied to the test data. The classification of models and actual profit conditions were compared. Thus, the models are tried to be compared. For this purpose, accuracy rates showing the correct classification possibilities of the models are given in Table 2.

Table 2

Accuracy Ratios of models

Methods	Accuracy Ratios(%)
Support Vector Machine	85,714
Decision Trees	94,286
Logistic Regression	91,429
Artificial Neural Networks	94,286
Fuzzy Rules (Logic)	91,429
Random Forest Algorithm	88,571
Discriminant Analysis	90,000

When Table 2 is examined, the accuracy rates of the installed models can be seen. Accuracy ratio is the ratio of the true states to the total state in the cross table created between the prediction of the model and the actual state. Accuracy gives the possibility to know accurately whether a business is profitable with the help of established models. For example, a model established with decision trees knows whether 94,286% of a company is profitable. Thus, it can be said that an enterprise knows artificial neural networks with the highest accuracy with the help of independent variables.

Multivariate normality, homogeneity of the covariance matrix, multiple correlation and linear relationship were investigated. It was found that the financial ratios provided multiple normalities, the covariant matrix was homogeneous according to the Box's M statistic and

the VIF values were less than 10, so there were no multiple correlations. The accuracy rate was evaluated as assumptions were made on the implementation of the separation analysis model. Since there is no assumption of machine learning other than statistics-based methods, it is compared with the correct classification rates obtained from the test data. Artificial neural networks are one of the methods that make him go away. In recent studies, many artificial neural network models are fed forward in this computer. The artificial neural network model is a multi-layer feedforward network model.

According to the results of the table, it was determined that the most accurate classification belongs to artificial neural networks with decision trees. It is seen that these two models have the highest classification successes and therefore make the most accurate class estimation. The classification of the two methods is given in Table 3.

Table 3

Classification with model results

		Prediction	
		Profitable	Not-Profitable
Variable	Profitable	31	0
	Not-Profitable	2	2

As can be seen in the table, only two classes predicted 35 predictions and two models were mistaken. Accordingly, it has classified it as profitable to 2 non-profitable companies. 33 of the data in the test data were classified correctly and models were obtained with 94,286% accuracy.

Conclusions

Both managers and investors are interested in the profitability of a business. Both academic and sectoral studies are carried out to identify factors affecting profitability, as well as to predict profitability. With empirical studies, models have been tried to be formed. In the literature,

profit or profitability is generally considered as a dependent variable and financial ratios are calculated by using the items in the financial statements and independent variables are formed.

In this study, the 14 most common financial ratios are considered in the literature. The relationship between these 14 variables and profit status is tried to be explained with the help of classification techniques of data mining. For this purpose, the financial statements of 70 companies, excluding the holding and finance companies in the BIST100, were utilized. The mentioned 14 variable and net profit conditions are calculated from the financial statements.

Machine learning is in two forms: supervised and unsupervised learning models. Classification algorithms are controlled learning models and models have been formed in order to take into consideration the profit conditions of enterprises. For machine learning, the data were randomly divided into two as 50% learning data and 50% test data. Logistic regression, separation analysis, artificial neural networks, support vector machines, fuzzy rule (logic), random forest algorithms and decision tree models were formed from the data selected as learning data. Then, these models were applied to the test data and their results were compared with their actual situations. In order to determine the success of the classification algorithms, accurate prediction rates were taken into account.

The two models that among the models were found the highest accuracy rate. Artificial neural networks and decision trees applied the correct classification of 94.286% on the test data. As the test data, 33 of the 35 randomly chosen companies correctly predicted the profit status of the company and 2 classified the non-profitable business as profitable.

The best performance was obtained by decision tree and artificial neural networks. Nageswari et. al. in parallel w0ith their study. Two algorithms with the highest accuracy were determined. Pu et. al. and Maji & Arora 's work in hybrit models will increase accuracy. In this way, decision makers can make more accurate decisions. Unlike the study performed by Emir et al., artificial neural networks showed better results than support vector machines.

Under these conditions, it can be said that the use of artificial neural networks or decision trees would make the least misleading predictions when it is desired to examine whether a business is profitable. An investor or business owner is advised to choose these two methods when they want to predict whether a business is profitable or not.

Before making an investment decision, it is important for investors whether companies will profit. Whether the investor wants to invest in stock, project or capital partnership, it is important that the investor determines the profit situation accurately. For this purpose, the methods mentioned in the study can be used.

In this study, BIST100 was preferred as a sample. In future studies, the implementation of these methods in other countries or index and is recommended to test the consistency of the results. In addition, future studies, taking advantage of the results of different algorithms at the method compared with the results of this study. Higher accuracy can be achieved with

integrated algorithms. In future studies, it is recommended that hybrit models be tested in the field of finance. Furthermore, classification algorithms other than this study can be used in the field of finance. The obtained accuracy rates can be compared with this study. It is recommended that integrated studies with algorithms that give higher rate are performed.

References

Agresti, A. (2018). "An Introduction to Categorical Data Analysis". *Wiley.*

Akbulut, R. (2011). "İMKB'de İmalat Sektöründeki İşletmelerde İşletme Sermayesi Yönetiminin Karlılık Üzerindeki Etkisini Ölçmeye Yönelik Bir Araştırma. Istanbul University Journal of the School of Business Administration, 40(2).

Akbulut, R. (2011). İMKB'de İmalat Sektöründeki İşletmelerde İşletme Sermayesi Yönetiminin Karlılık Üzerindeki Etkisini Ölçmeye Yönelik Bir Araştırma". *Istanbul University Journal of the School of Business Administration*, 40(2).

Akhtar, S. (2004). "Australian Multinational and Domestic Corporations Capital Structure Determinants".

Albayrak, A. S., & Akbulut, R. (2012). "Karlılığı Etkileyen Faktörler: İMKB Sanayi ve Hizmet Sektörlerinde İşlem Gören İşletmeler Üzerine Bir İnceleme." *Uluslararası Yönetim İktisat ve İşletme Dergisi*, 4(7), 55-82.

Alpar, R. (2017). "Çok Değişkenli İstatistiksel Yöntemler", Ankara: *Detay Yayıncılık*.

Altman, E. I. (1968). "Financial Ratios, Discriminant Analysis and The Prediction of Corporate Bankruptcy". *The Journal of Finance,* 23(4), 589-609.

Atalay, M., & Çelik, E. (2017). "Büyük Veri Analizinde Yapay Zekâ ve Makine Öğrenmesi Uygulamaları-Artificial Intelligence and Machine Learning Applıcatıons in Big Data Analysis". *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(22), 155-172.

Aydın N., Şen Mehmet & Berk N., "Finansal Yönetim-1", 1.Baskı, Eskişehir, Haziran 2012, S:85

Büyükşalvarcı, A. (2010). "Finansal Oranlar ile Hisse Senedi Getirileri Arasındaki İlişkinin Analizi: İMKB İmalat Sektörü Üzerine Bir Araştırma". *Muhasebe ve Finansman Dergisi*, (48), 130-141.

Chen, L., & Zhao, X. S. (2005). "Profitability, Mean Reversion of Leverage Ratios, and Capital Structure Choices".

Chen, Z., Matousek, R., & Wanke, P. (2018). "Chinese Bank Efficiency During The Global Financial Crisis: A Combined Approach Using Satisficing Dea And Support Vector Machines". *The North American Journal of Economics and Finance*, 43, 71-86.

Chou, C. H., Hsieh, S. C., & Qiu, C. J. (2017). "Hybrid Genetic Algorithm and Fuzzy Clustering for Bankruptcy Prediction". *Applied Soft Computing,* 56, 298-316.

Elkano, M., Galar, M., Sanz, J., & Bustince, H. (2018). Chi-Bd: "A Fuzzy Rule-Based Classification System for Big Data Classification Problems." *Fuzzy Sets and Systems*, 348, 75-101.

Emir, S., Dinçer, H., & Timor, M. (2012). "A Stock Selection Model Based on Fundamental and Technical Analysis Variables by Using Artificial Neural Networks and Support Vector Machines".

Eti, S. (2016). "Finansal Verilerle Kârlılığı Açıklayan Modellerin Karşılaştırılması Üzerine Bir Araştırma", Yüksek Lisans Tezi, *Marmara Üniversitesi Sosyal Bilimler Enstitüsü*.

Eti, S. (2019). The Use of Quantitative Methods in Investment Decisions: A Literature Review. In Handbook of Research on Global Issues in Financial Communication and Investment Decision Making (pp. 256-275). IGI Global.

Eti, S., & İnel, M. N. (2016). "A Research on Comparison of Regression Models Explaining The Profitability Base on Financial Data". *International Journal of Business and Management*, 4(10), 470-475.

Gepp, A., & Kumar, K. (2015). "Predicting Financial Distress: A Comparison of Survival Analysis and Decision Tree Techniques." *Procedia Computer Science*, 54, 396-404.

Goel, P. M. (2019). Comparison of Classification Techniques on Data Mining.

Han, X., Gong, X., & Zhang, P. (2018). "Online Labor Service Crowdsourcing Analysis Based on Linear Discriminant Regression". *Cognitive Systems Research*, 52, 168-173.

Hosaka, T. (2019). "Bankruptcy Prediction Using Imaged Financial Ratios and Convolutional Neural Networks." *Expert Systems with Applications*, 117, 287-299.

Irimia-Dieguez, A. I., Blanco-Oliver, A., & Vazquez-Cueto, M. J. (2015). "A Comparison of Classification/Regression Trees and Logistic Regression in Failure Models". *Procedia Economics and Finance*, 26, 23-28.

Ínel, M. N., & Armutlulu, İ. H. (2017). "Belirsizlik Ortamında Fuzzy Finansal Oranlarla Karar Verme".

Ínel, M., Eti, S. & Yıldırım H. (2016), «A Comparison Of Artificial Neural Network And Decision Tree For Profitability In Technology Sector», *International Journal of Development Research*, 6(7), 8417-8421

Jabeur, S. B. (2017). "Bankruptcy Prediction Using Partial Least Squares Logistic Regression." *Journal of Retailing and Consumer Services*, 36, 197-202.

Jafarian, A., Nia, S. M., Golmankhaneh, A. K., & Baleanu, D. (2018). "On Artificial Neural Networks Approach with New Cost Functions." *Applied Mathematics and Computation*, 339, 546-555.

Karaca, S. S., & Başci, E. S. (2011). "Hisse Senedi Performansını Etkileyen Rasyolar ve İmkb 30 Endeksinde 2001-2009 Dönemi Panel Veri Analizi." *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 16(3).

Karadeniz, E., & İskenderoğlu, Ö. (2011). "İstanbul Menkul Kıymetler Borsası'nda İşlem Gören Turizm İşletmelerinin Aktif Kârlılığını Etkileyen Değişkenlerin Analizi." *Anatolia: Turizm Araştırmaları Dergisi*, 22(1).

Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2015). "Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies." *MIT Press*.

Kısakürek, M., & Aydın, Y. (2013). "İşletmelerde Sermaye Yapısı ile Kârlılık Arasındaki İlişkinin Analizi: 1992-2011 Yılları Arası Finansal Krizler Odaklı Bist'te Bir Uygulama". *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 14(2), 97-121.

Kim, S. Y., & Upneja, A. (2014). "Predicting Restaurant Financial Distress Using Decision Tree and Adaboosted Decision Tree Models". *Economic Modelling*, 36, 354-362.

Kumaş, H., Çağlar, A., & Karaalp, H. S. (2014). "Firm Size and Labour Market Segmentation Theory: Evidence From Turkish Micro Data." *Procedia-Social and Behavioral Sciences*, 150, 360-373.

Maji, S., & Arora, S. (2019). Decision Tree Algorithms for Prediction of Heart Disease. In Information and Communication Technology for Competitive Strategies (pp. 447-454). *Springer, Singapore*.

Nissim, D., & Penman, S. H. (2003). "Financial Statement Analysis of Leverage and How it Informs about Profitability and Price-to-book Ratios." *Review of Accounting Studies*, 8(4), 531-560.

Okuyan, H. A. (2013). "Türkiye'deki En Büyük 1000 Sanayi İşletmesinin Kârlılık Analizi*/The Profitability Analysis of The Largest 1000 Industrial Firms in Turkey". *Business and Economics Research Journal*, 4(2), 23.

Omran, M., & Ragab, A. (2004). "Linear Versus Non-Linear Relationships Between Financial Ratios and Stock Returns: Empirical Evidence From Egyptian Firms." *Review of Accounting and Finance*, 3(2), 84-102.

Oruç, E. (2010). "İmkb'de İşlem Gören İşletmelerin Hissi Senedi Getirileri ile Çeşitli Finansal Göstergeleri Arasındaki İlişki." *Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 3(1-2), 33-43.

Pınar, M., Okumuş, O., Turgut, U. O., Kalıpsız, O., & Aktaş, M. S. "Büyük Veri İçeren Öneri Sistemleri İçin Hiperparametre Optimizasyonu."

Pu, Q., Li, Y., Zhang, H., Yao, H., Zhang, B., Hou, B., ... & Zhao, L. (2019). "Screen efficiency comparisons of decision tree and neural network algorithms in machine learning assisted drug design". *Science China Chemistry*, 62(4), 506-514.

Rodrigues, L., & Rodrigues, L. (2018). "Economic-Financial Performance of The Brazilian Sugarcane Energy Industry: An Empirical Evaluation Using Financial Ratio, Cluster and Discriminant Analysis." *Biomass and Bioenergy*, 108, 289-296.

Roy, S., Mondal, S., Ekbal, A., & Desarkar, M. S. (2019). "Dispersion Ratio Based Decision Tree Model for Classification." *Expert Systems with Applications,* 116, 1-9.

Silahtaroğlu, G., & Ergül, H. (2016). "Şehirleşme, Mekân–İnsan Etkileşiminin Birey Algısına Yansıması: Bir Veri Madenciliği Analizi." *Beykent Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 9(2).

Ye, X., Dong, L. A., & Ma, D. (2018). "Loan Evaluation in P2p Lending Based on Random Forest Optimized by Genetic Algorithm with Profit Score." *Electronic Commerce Research and Applications*, 32, 23-36.

Yuksel, S., Eroglu, S., & Ozsari, M. (2016). "An Analysis of The Reasons of Internal Migration in Turkey with Logit Method." *Business and Management Horizons*, 4(2), 34-45.

Zelenkov, Y., Fedorova, E., & Chekrizov, D. (2017). "Two-Step Classification Method Based on Genetic Algorithm for Bankruptcy Forecasting." *Expert Systems with Applications*, 88, 393-401.