Muhasebe ve Vergi Uygulamaları Dergisi
Journal of Accounting and Taxation Studies

Muhasebe ve Vergi Uygulamaları Dergisi, 2025
18(2), 239-259

Submitted: 27.08.2024
Accepted: 25.02.2025
Published Online: 21.07.2025
Similarity Rate: %20

DOI: https://doi.org/10.29067/muvu.1539635
Web: https://dergipark.org.tr/tr/pub/muvu

Araştırma Makalesi (Research Article)

Prediction of Independent Audit Firm Switching By Using Machine Learning Methods: The Case of Türkiye

Ahmet ÇANKAL[1]
Erdem KÜRKLÜ[2]

Abstract

This study aimed to predict independent audit firm switching of the companies traded in Borsa Istanbul Star Market (BIST STARS) in Türkiye by using financial ratios and machine learning algorithms. In this context, 13 financial datasets of 158 companies traded in BIST STARS in the 2019-2021 period were used as input variables. First, the significance values of the input variables were found by using the Mutual Information (MI) method. Then, input variables were grouped sequentially in order of importance to select the most accurate subset representing the data. Among the machine learning algorithms, Support Vector Machine, Decision Tree, Random Forest, Naive Bayes,K-Nearest Neighbors, and XGBoost algorithm methods were used for group selection. GridSearchCV technique was applied to optimize the initial parameters of the methods. As a result of the experiments, the XGBoost algorithm was found to be the most successful method in predicting the change of independent audit firm with an accuracy value of 88.4%. It was sufficient for the method to use 8 attributes selected from 13 financial datasets. On the other hand, the Return on Assets (ROA) was determined as the most important attribute.

Keywords: Audit Firm Switch, Financial Ratios, Machine Learning, Mutual Information.

JEL Codes: M49, M42, E37, C38, C53.

Makine Öğrenmesi Yöntemlerini Kullanarak Bağımsız Denetim Firma Değişikliğinin Tahmini: Türkiye Örneği

Öz

Bu çalışmanın amacı, Türkiye'de Borsa İstanbul Yıldız Pazar'da (BIST Yıldız Pazar) işlem gören işletmelerin bağımsız denetim firması değişikliğini, finansal oranlar ve makine öğrenmesi algoritmaları kullanarak tahmin etmektir. Bu kapsamda, 2019-2021 döneminde Borsa İstanbul Yıldız Pazar'da işlem gören 158 işletmeye ait 13 finansal veri kümesi girdi değişkenleri olarak kullanılmıştır. Öncelikle Mutual Information (MI) yöntemi kullanılarak girdi değişkenlerinin önem değerleri bulunmuştur. Daha sonra girdi değişkenleri önem sırasına göre gruplandırılarak veriyi en doğru şekilde temsil eden alt küme seçilmiştir. Makine öğrenmesi algoritmaları arasında Destek Vektör Makinesi, Karar Ağacı, Rastgele Orman, Naive Bayes, K-En Yakın Komşu ve XGBoost yöntemleri grup seçiminde kullanılmıştır. Yöntemlerin başlangıç parametrelerini optimize etmek için GridSearchCV tekniği uygulanmıştır. Yapılan deneyler sonucunda XGBoost algoritmasının %88,4 doğruluk değeri ile bağımsız denetim şirketi değişikliğini tahminlemede en başarılı yöntem olduğu bulunmuştur. Yöntem için 13 finansal veri setinden seçilen 8 niteliğin kullanılması yeterli olmuştur. Öte yandan Aktif Kârlılık Oranı (ROA) en önemli nitelik olarak belirlenmiştir.

Anahtar Sözcükler: Denetim Firma Değişikliği, Finansal Oranlar, Makine Öğrenmesi, Karşılıklı Bilgi.

JEL Kodları: M49, M42, E37, C38, C53.

[1] Sorumlu Yazar (Corresponding Author): Ahmet ÇANKAL, (Dr. Öğr. Üyesi), Osmaniye Korkut Ata Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Öğretim Üyesi, Osmaniye/Türkiye, ahmet.cankal@osmaniye.edu.tr, ORCID ID: 0000-0002-3639-8861.
[2] Erdem KÜRKLÜ, (Öğr. Gör.), Osmaniye Korkut Ata Üniversitesi, Osmaniye Meslek Yüksekokulu, Ulaştırma Hizmetleri Daire Başkanlığı Öğretim Elemanı, Osmaniye/Türkiye, erdemkurklu@osmaniye.edu.tr, ORCID ID: 0000-0001-7075-6995.
APA 6 Stili Kaynak Gösterimi: (To Cite This Article)
Çankal, A. & Kürklü, E. (2025). Prediction of independent audit firm switching by using machine learning methods: The case of Türkiye. *Journal of Accounting and Taxation Studies, 18*(2), 239-259. doi: https://doi.org/10.29067/muvu.1539635

1. INTRODUCTION

Companies often tend to switch with a new audit firm due to reasons such as financial difficulties, disagreements with the audit firm, changes in company management, etc. These changes can be perceived by the market as a signal that the company is facing difficulties. In this case, the company's reputation and market value may be damaged (Black, 2013). Therefore, the transition of companies from the Big Four audit firms to other audit firms or from other audit firms to the Big Four audit firm can directly affect the quality of the companies' financial reports, investor confidence, and market value (Jain and Agarwalla, 2023). Choosing the wrong audit firm can both jeopardize the accuracy of financial reports and damage the company's reputation (Guo et al., 2023).

The study was conducted to develop a model that can predict audit firm switching in public companies other than the mandatory rotation. Therefore, it is thought that developing a data-driven model that predicts audit firm switching with financial ratios and machine learning methods will eliminate the uncertainties in audit firm changes. Thus, it is aimed at adding a new dimension to the existing studies on audit firm switching.

Using the model created, company managers can more clearly assess whether they can afford the costs of working with a large audit firm and whether this will contribute to the company. Companies can evaluate whether it is appropriate to work with the Big Four according to their growth targets and international expansion plans. The model developed using machine learning can reinforce the perception among investors, shareholders, and creditor financial institutions that companies working with the Big Four are under strict supervision and have more reliable financial reports. Furthermore, demonstrating the impact of financial ratios on audit firm switch through the model could open up opportunities for new academic research.

It is thought that estimating the audit firm switching of public companies in Türkiye with the help of financial ratios will be effective. When studies in literature are evaluated in general, it has been determined that the most used statistical methods are surveys, questionnaires, logistic regression and discriminant analysis. However, the number of studies predicting the switching of audit firms by using machine learning algorithms and financial ratios is limited. It is thought that estimating the switching of independent audit firms with financial ratios and machine learning algorithms such as SVM (Support Vector Machine), DT (Decision Tree), RF (Random Forest), NB (Naive Bayes), KNN (K-Nearest Neighbors) and XGBoost will fill the gap in the literature on this subject.

The research consists of 6 sections. In the introduction section, the problem determination, the importance of the subject and its contribution to the solution of the problem are explained. Then, the literature is reviewed and previous studies, their methods and findings are summarized. In the methodology section, information about the methods used in the research is given. In the dataset section, sample selection, characteristics and formulas of dependent and independent variables used in the research are explained. In the experiments section, application stages and results are reported. According to these results, the comparison of models and the importance levels of variables are determined. In the last section, the findings are interpreted and compared with the literature.

2. LITERATURE REVIEW

Audit firm switching is carried out on two bases: mandatory and voluntary. While mandatory audit firm switching is carried out due to legal regulations, voluntary audit firm switching is carried out in line with the wishes of the management (Susanto, 2018). There are many factors behind the voluntary audit firm switching decision (Aisyah and Faridah, 2023). Some studies in the literature on these factors are listed below:

Kwak et al. (2011) used 13 financial ratios and multi-criteria linear programming and data mining approaches to predict audit firm switching. The data set consists of 790 company-years that switched audit firms and 1,132 company-years that did not switch audit firms between 2007-2008 from Compustat database. Data mining techniques such as MCLP, BayesNet, CART and linear logistic regression methods were used. As a result of the research, the general accuracy rate of the models predicting audit firm switching was found to be around 60% on average. It was also shown that companies that switched audit firms had lower liquidity, lower profitability, lower asset turnover and tended to pay less dividends than companies that did not switch audit firms.

Nazri et al. (2012) investigated the factors affecting audit firm switching using 18 years of data from 400 companies listed in Bursa Malaysia between 1990 and 2008. They found that management changes, firm size, audit opinion and change in sales were effective in audit firm switching.

Kirkos (2012) tried to develop models that predict audit firm switching using financial variables and data mining techniques. Using the FAME (Financial Analysis Made Easy) database, he determined the firms that changed their audit firms between 2003-2005 from the firms belonging to the UK and Ireland. The dataset consists of 338 firm-year observations, 169 firms that switched their audit firms and 169 firms that did not switch their audit firms. He selected 6 variables from 59 financial variables of the firms by applying correlation-based feature selection. Then, he compared the success of the models using 6 different data mining methods. Bayesian Network was the most successful method with an accuracy rate of 84.6%. According to the study results, data mining techniques outperform logistic regression.

Black et al. (2013) analyzed whether audit firm switch can be predicted by applying discriminant analysis and logistic regression to companies traded on the Athens Stock Exchange. The dataset of the study consists of 268 companies traded between 2008-2009. As a result of the analysis, they determined the accuracy of audit firm switch prediction as 92.5% using four financial ratios and logistic regression. They stated that these financial ratios are very good predictive variables in predicting audit firm switch. The financial ratios used in the model are Working Capital/Total Assets, Return on Assets, Market Value of Equity/Book Value of Total Debt, Sales/Total Assets.

Ünal and Altay (2015) investigated the link between corporate governance and audit firm switch using logistic regression analysis on 882 firm-year observations from companies in the manufacturing sector of Borsa Istanbul between 2005 and 2013. Their findings showed that several factors influence audit firm switching, including the composition of the board of directors and ownership structure, the market-to-book value ratio, the logarithm of total assets, the current ratio, and the status of the CEO within the company.

Chung et al. (2021) examined the impact of audit firm switching on audit opinion and quality using 8136 firm-year observations from public companies obtained from the Korean Information Service (KIS) between 2011 and 2016. They found that audit quality deteriorated and investors' trust in companies decreased after audit opinion shopping. They also found that audit firm switch harmed auditor independence and distorted the company's financial information.

Kamarudin et al. (2022) examined 28,073 observations across five Asian countries (Indonesia, Malaysia, Philippines, South Korea, and Thailand) from 2007 to 2016. They explored how auditor switching impacts conditional conservatism and how moving between Big Four auditors and industry specialists influences earnings quality. They found that transitions from firms other than the Big Four to the Big Four will result in high conservatism. They showed that transitions from the Big Four to others will lead to low conditional conservatism.

Ding et al. (2022) examined the impact of industry specialization on audit quality by analyzing 2,364 audit firm switches made by publicly listed companies in China between 2006 and 2015. Their study found that switching audit firms is often associated with poor financial prospects and low audit quality. Aisyah and Faridah (2023) examined the factors affecting audit firm switching using logistic regression method on a total of 252 samples using 6-year data of 42 companies listed on Indonesia Stock Exchange (BEI). As a result of the study, they showed that audit delay, going concern, audit opinion, management changes, auditor reputation and company size have no effect on auditor switching.

Altass (2023) used 168 company-year data from 2015 to 2018 for construction materials companies listed in the Tadawul All Share (TASI) index. Based on these data, the relationship between audit firm switching, auditor tenure and business performance was examined with a regression model. As a result of the study, it was found that audit firm switching did not affect business performance. In addition, no relationship was found between auditor tenure and business performance.

Saluy et al. (2024) surveyed with 175 auditors in Indonesia and found that factors such as audit fees and internal control had a significant impact on auditor turnover.

It has been determined that the studies examining audit firm switching in the literature generally address variables such as audit opinion, management change, firm size, audit quality, audit fee and financial failure. However, the number of studies on predicting audit firm switching by using machine learning algorithms and financial ratios is limited. Therefore, the use of machine learning algorithms such as SVM (Support Vector Machines), DT (Decision Trees), RF (Random Forests), NB (Naive Bayes), KNN (K-Nearest Neighbors) and XGBoost has the potential to fill the gap in the literature. Such algorithms, especially by making predictions on financial ratios, can help develop more sensitive and predictive models regarding audit firm switching. This will contribute to the existing literature and fill an important gap in both academic and practical terms.

3. METHODOLOGY

The learning process includes processes such as classification, regression, clustering, and attribute extraction. Learning from data is defined as the creation of a learning machine or algorithm that allows the estimation function to be determined based on a limited number of training data. Owing to these characteristics, learning with data-based methods (data mining) is used in many areas like marketing, finance, management, engineering, production, industry, banking, health, tourism, and medicine (Chunhong and Licheng, 2004). In the process of separating these data sets into certain classes according to their common characteristics, various classification methods are needed. Various algorithms have been developed for this purpose. The main algorithms used in this regard can be listed as entropy-based classification (C4.5 algorithm, C5.0 algorithm), Regression and Decision Trees (Gini algorithm, Twoing algorithm), Memory-Based Algorithms (K-nearest neighbor algorithm), Bayesian Classifiers, Regression Trees, Support Vector Machines, and Random Forest (Okun, 2011).

Among these algorithms, to determine whether an effective estimation of the audit firm switch would be made and to identify the significance levels of the determined data sets (financial ratios), the correlation between the dependent variable (audit firm switch) and the input variables (financial ratios) has been explained using Support Vector Machines, Decision Trees, Random Forest, Naive Bayes, K-Nearest Neighbor, and XGBoost classification algorithms. The purpose here is to establish an estimation model that predicts the correlation between dependent and input variables to have the best fit. Descriptive information about the machine learning system used in the research are presented below.

3.1. Support Vector Machines (SVM)

A support vector machine (SVM) is a supervised machine learning method that can produce high accuracy with a low number of data in both regression and classification. SVM estimates the appropriate function to classify the data. (Ghosh et al., 2019). SVM method also directs many current developments in kernel-based methods. SVM is similar to neural networks and radial-based artificial neural networks, but it is generally indicated that it performs better than these algorithms. This learning algorithm is seen as an alternative training technique for Polynomial, Radial Basic Function and Multilayer Perceptron (Artificial neural network) classifiers. The underlying idea in the technique is to create an optimal hyperplane that maximizes the margin between classes of data (Pradhan, 2012). The number of samples to be used in SVMs is not important. SVM also has the ability to generalize data that has not been seen during training, as it classifies them as problem-free. Its ability to generalize makes SVM a good alternative to other techniques. SVMs can be run on classification and regression problems. Thus, it has been observed that they are more successful in comparison to statistical methods in various areas such as predicting stock market index movements, predicting financial information manipulations, or creating early warning systems against financial failures (Cao and Tay, 2003).

3.2. Decision Tree (DT)

A decision tree is a type of supervised learning algorithm that does not rely on parametric assumptions and is used for both regression and classification tasks. It is structured hierarchically like a tree, with key components including a root node, branches, internal nodes, and leaf nodes. Each internal node represents a decision point based on a feature, while the branches represent possible outcomes, and the leaf nodes indicate final classifications or predictions (Song and Lu, 2015). The first step taken in creating a decision tree is to determine according to which criterion the branching in the tree will be made or according to which attribute values the tree structure will be created (Seethapathy and Babu, 2021). For this reason, a variety of algorithms have been developed in the literature such as C4.5, C5.0, CHAID, CART, and ID3 (Patel and Prajapati, 2018). Decision trees are commonly used as they can be easily implemented in integration with databases (Nassif et al., 2013). DT is based on investigating all possible relationships between dependent and independent variables to obtain the best prediction (Seethapathy and Babu, 2021, p.14-15). When the independent variable with the highest relationship is determined, the data set is divided into two according to the values of this independent variable. This process continues until all separations are completed. The decision rules reflected by the logical model presented by decision trees are clear enough to be easily understood by humans. This method has a wide range of applications since it possesses features such as having high classification accuracy rate and producing simple rules (Barros et al., 2012, p.1237).

3.3. Random Forest Algorithm (RF)

Random forest is proposed by Leo Breiman (2001), a machine learning algorithm widely used in classification and regression problems, based on combining the output of multiple decision trees to reach a single result. (Boulesteix et al., 2012; Liaw and Wiener, 2022). On the basis of the RF algorithm is a single decision tree, a model that uses binary splits on variables to predict outcome. In this method, random feature selection is made to divide the nodes into branches and certain weight values are assigned to their decision trees (Breiman, 2001; Siva et al., 2012). At each node of the decision tree, the feature and corresponding split value that best optimize the splitting criterion are selected. The prediction errors for all decision trees are calculated, with higher weights assigned to trees that achieve the lowest errors and lower weights to those with the highest errors. These weights play a crucial role in the voting process during class estimation. The votes from all trees are then combined, and the final decision is determined based on this collective voting process (Chen et al., 2009).

3.4. Naive Bayes Algorithm (Naive Bayes-NB)

The Naive Bayes classifier is a widely used supervised machine learning classifier known for its simplicity. When this classifier is given the likelihood of an event that has already occurred, it finds the probability of another event occurring (Al-Garadi et al., 2020). This approach allows the algorithm to make predictions quickly and accurately (Srinivasa et al., 2020). Therefore, it has been successfully used in many applications such as weather prediction services, customer credit evaluations, health condition categorizations and so on (Yang, 2018).

3.5. XGBoost Algorithm (XGBoost)

The XGBoost algorithm, introduced by Chen and Guestrin (2016), is a powerful machine learning framework for tree boosting and has gained wide support in academic research (Liu, 2021). It enhances model performance by combining weak base learners into a stronger model through iterative training. In each iteration of gradient boosting, the residuals from the previous model are used to adjust and optimize the next model, allowing the indicated loss function to be minimized more effectively with each step. This iterative process strengthens the overall predictive accuracy of the final model. The algorithm has attracted attention as it has won many Kaggle competitions. XGboost is used in structured or tabular data sets in regression and classification problems (Zhou et al., 2021, p.28). XGBoost is an improved implementation of the basic Gradient Boosting Machines framework through system optimization and algorithmic improvements (Chen and Guestrin, 2016).

3.6. K-Nearest Neighbors Algorithm (KNN)

The KNN algorithm is one of the algorithms widely used for classification and regression problems that make predictions based on the distances of observations to each other. Many functions are used for similarity and distance measurements. It is classified according to the proportion of similar examples among nearest neighbors. It has been applied in many fields such as pattern recognition, object recognition, and text recognition (Han et al., 2011; Hu et al., 2016; Abu Alfeilat et al., 2019).

4. DATASET

In this environment, where the full application of the principle of financial transparency, which is of great importance for information users such as investors, lenders, and regulatory bodies, in the markets is discussed, there is no information sharing about the factors according to which the independent audit firm switching of the enterprises are made. Due to this gap, a literature review was conducted on which factors businesses are affected in their independent audit firm switching, although many factors were identified, it was observed that the weighted point of the studies was audit fee and financial performance factors (Nasser et al., 2006; Wan Mohamed et al., 2007; Ettredge et al., 2007; Ismail et al., 2008; Calderon and Ofobike, 2008; Chen et al., 2008; Kwak et al., 2011; Eldridge et al., 2012; Huang and Scholz, 2012; Suyono et al., 2013; Black et al., 2013). However, as the data on audit fees in Türkiye only show the revenue generated by the audit firms from the total audit of that year and the audit fee is not reflected on the client business basis, the present study focused on the financial performance variable. In this respect, the latest research aimed to estimate the factors affecting the independent audit firm switching of the companies traded in BIST STARS (Borsa Istanbul Stock Exchange Star Market) by using Machine Learning algorithms. Companies with market capitalization of shares in actual free float > 1 Billion TL and meeting other requirements are listed on BIST STARS. Information about the calculation of 1 dependent and 13 input variables created for use in machine learning algorithms is presented in Table 1.

Table 1. Variables Used in the Model

Dependent Variable	Input Variables	Calculation
	Current Ratio (CUR)	Current Assets / Short Term Liabilities
	Liquidity Ratio (LIR)	(Current Assets-Inventories) / Short Term Liabilities
	Cash Ratio (CAR)	(Current Assets-Inventories-Receivables) /Short Term Liabilities
	Financial Debt Ratio (FDR)	Financial Liabilities / Total Assets
Audit Firm Switching (AFS)	Leverage Ratio (LER)	Total Liabilities / Total Assets
	Asset Turnover Ratio (ATR)	Net Sales /Total Assets

Inventory Turnover Ratio (ITR)	Cost Of Sales/ Average Stock Amount
Return on Assets (ROA)	Net Profit /Total Assets
Gross Profit Margin (GPM)	Gross Sales Profit / Revenue
EBITDA Margin (EBM)	EBITDA / Revenue
Net Profit Margin (NPM)	Net Profit / Revenue
Return on Equity (ROE)	Net Profit / Equity
Return on Invested Capital ROIC)	Net Operating Profit After Tax / Invested Capital

Source: Saalem and Rehman, 2011; Adjirackor et al., 2017

While the dependent variable used in the study was defined as Audit Firm Switching (AFS), the input variables were determined as Return on Assets (ROA), Financial Debt Ratio (FDR), Net Profit Margin (NPM), Return on Invested Capital (ROIC), EBITDA Margin (EBM), Leverage Ratio (LER), Inventory Turnover Ratio (ITR), Current Ratio (CUR), Gross Profit Margin (GPM), Cash Ratio (CAR), Liquidity Ratio (LIR), Asset Turnover Ratio (ATR), and Return on Equity (ROE).

The data related to these input variables, which were thought to best represent the financial capabilities of the enterprises, were obtained from the website of Fintables Informatics Technologies Inc. which included the financial data analyses of the enterprises, covering the years 2019-2021. Since these input variables, which were determined in order to reveal the correlation between financial ratios and independent audit firm switching, were taken as indicators that could directly affect financial performance in many studies (Cheng et al., 2005; Lei and Liu, 2021; Yim and Mitcbell, 2005; Lin, 2009), it was deemed appropriate to use them in this study as well.

While classifying the dependent variable of the research, if the audit firm switching (AFS) of the said enterprises was one of the four major audit firms (Ernst and Young-Guney Independent Audit, PWC - Basaran Nas Independent Audit, Deloitte - DRT Independent Audit, KPMG-Akis Independent Audit) it was given the value of 1, while it was given the value of 0 for an audit firm other than the four major firms.

In this study, data from companies operating continuously in the Borsa Istanbul Star Market (BIST STARS) between 2019 and 2021 were analyzed. In the following years, changes in market categories and the number of companies listed may occur due to factors such as the periodic performance of companies, sector conditions, new public offerings, and other requirements. Accordingly, the number of companies included in BIST STARS is likely to change over time. The data set of the research consisted of 474 data sets - 158 companies each year - between 2019-2021. Since classical statistical methods can be partially successful in the estimation phase, it was preferred to use SVM, RF, DT, KNN, NB, and XGBoost algorithms, which are among the machine learning methods that can learn from samples in recent years, can give a higher accuracy rate with fewer data sets, have generalization ability, and have fewer assumptions compared to statistical methods.

5. EXPERIMENTS

The study was coded using the Python 3.6 scikit-learn library. 80% of the data belonging to the variable group were used as training data and 20% as test data.

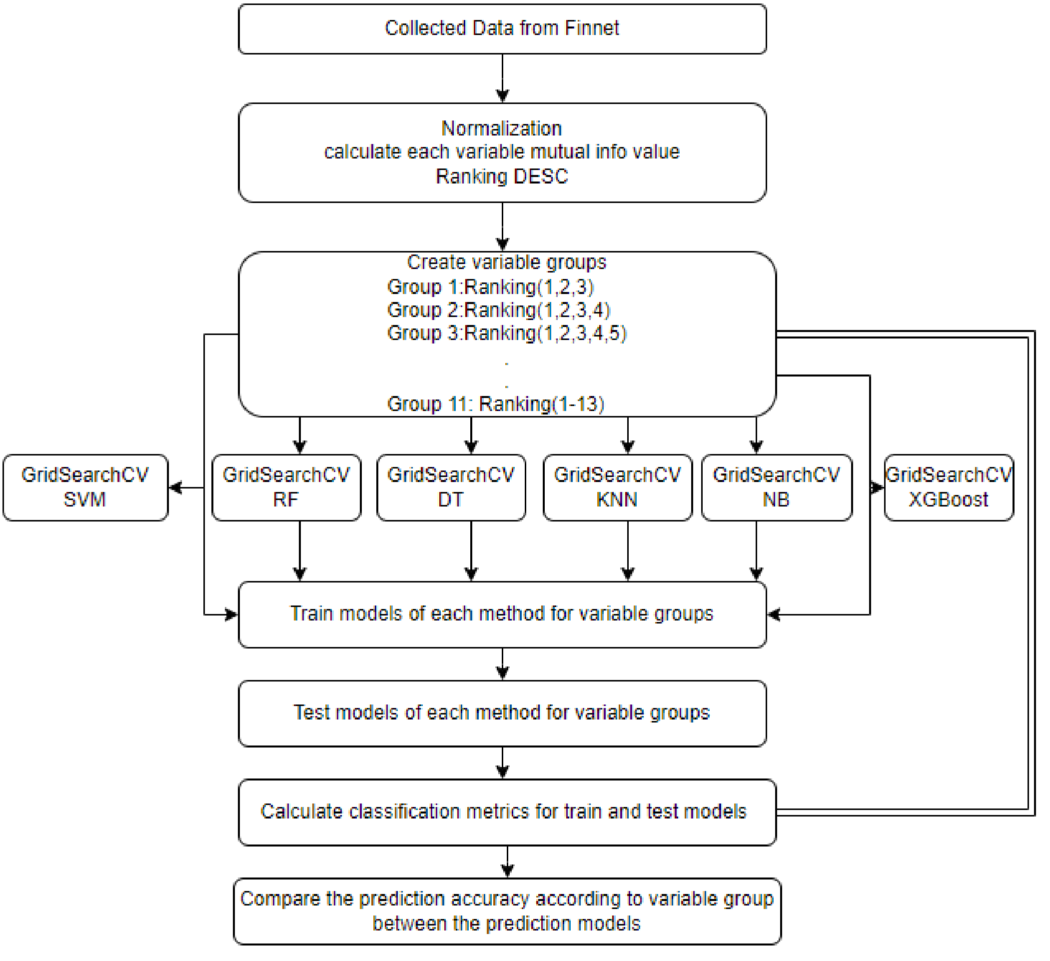

Figure 1. The Process of the Experiment

The process steps performed for the research and the information about these processes are presented in Figure 1.

Table 2. Initial Data Set Descriptive Statistics

	CUR	LIR	CAR	FDR	LER	ATR	ITR	ROA	GPM	EBM	NPM	ROE	ROIC
Number	474	474	474	474	474	474	474	474	474	474	474	474	474
Average	1.9	1.4	0.7	26.7	53.8	0.9	9.9	8.9	25.5	16.0	11.2	19.8	22.5
Std.	1.9	1.8	1.6	20.4	25.2	0.7	15.6	12.5	20	19.2	25.4	29.3	36.6
Min	0.1	0.09	0	0	0.9	0	-0.2	-38.1	-37.9	-84.5	-116	-118	-103.2
25%	1	0.6	0.09	9.03	35.8	0.4	3.1	1.7	13.8	7.9	1.7	4.2	4.5
50%	1.3	0.9	0.2	25.5	57.0	0.8	5.3	6.3	22.2	15.2	7.6	18.5	21
75%	2.1	1.5	0.6	39.6	72.7	1.2	9.8	14.4	32.5	22.3	16.9	32.2	42
Max	14.6	14.5	17.3	158.8	260.5	4.3	124.6	75	118.4	94.9	177.1	339.5	199

The Mutual Information method was used to identify attributes with high importance levels in the data set. Mutual Information (MI) measures the amount of information that a random variable can obtain about the other random variable (Witten, 2016). Mathematically, mutual information is expressed for the two abstract random variables of x and y as (Peng et al., 2005):

$$I(x;y) = \iint p(x,y) log \frac{p(x,y)}{p(x)p(y)} dxdy$$

Here, p (x,y) shows relative density function,

for x and y, p(x) and p(y) represent marginal probability density function.

The application of Mutual Information (MI) as an attribute evaluation method offers significant advantages due to its independence from any specific classifier and its lack of reliance on parameter tuning. This characteristic ensures ease of implementation and robust generalization across various datasets. Table 3 presents the ranked importance of variables determined through MI. The variables were categorized into 11 distinct Input Variable Groups (IVGs). For instance, the first three variables ranked by importance—Return on Assets (ROA), Net Profit Margin (NPM), and Financial Debt Ratio (FDR)—constituted IVG (1)-(3). Similarly, as additional variables were incorporated, groups expanded sequentially, such as IVG (1)-(4) consisting of ROA, NPM, FDR, and Return on Invested Capital (ROIC), and so forth up to IVG (1)-(13).

Table 3. Mutual Information Importance Ranking of the Variables Used in the Study

Input Variable	Ranking	Mutual Information Value	Input Variable Group
ROA	1	0.0580	
NPM	2	0.0362	
FDR	3	0.0300	IVG (1)-(3)
ROIC	4	0.0284	IVG (1)-(4)
EBM	5	0.0256	IVG (1)-(5)
LER	6	0.0211	IVG (1)-(6)
CAR	7	0.0210	IVG (1)-(7)
ITR	8	0.0202	IVG (1)-(8)
ATR	9	0.0063	IVG (1)-(9)
GPM	10	0.0053	IVG (1)-(10)
ROE	11	0.0023	IVG (1)-(11)
CUR	12	0.0000	IVG (1)-(12)
LIR	13	0.0000	IVG (1)-(13)

The Mutual Information (MI) values presented in Table 3 determine the effect of the variables used in the study on independent audit firm switch. The variable with the highest MI value is ROA (0.0580), which stands out as the most important factor in independent audit firm switch. This is followed by NPM (0.0362) and FDR (0.0300). This finding suggests that profitability and financial leverage ratios are key determinants of independent audit firm switching.

On the other hand, the zero MI values of LIR (0.0000) and CUR (0.0000) indicate that these variables do not provide significant information gain in relation to audit firm switching. Variables such as ROE, GPM, and ATR also have low MI values and make limited contributions to the predictive power of the model.

These input variable groups served as the basis for applying machine learning classification algorithms, specifically Support Vector Machines (SVM), Random Forest (RF), Decision Trees (DT), k-Nearest Neighbors (KNN), Naïve Bayes (NB), and XGBoost. The objective was to estimate audit firm switching, designated as the dependent variable. This methodological approach ensures consistency with established literature and highlights the practical utility of MI in variable selection for machine learning tasks.

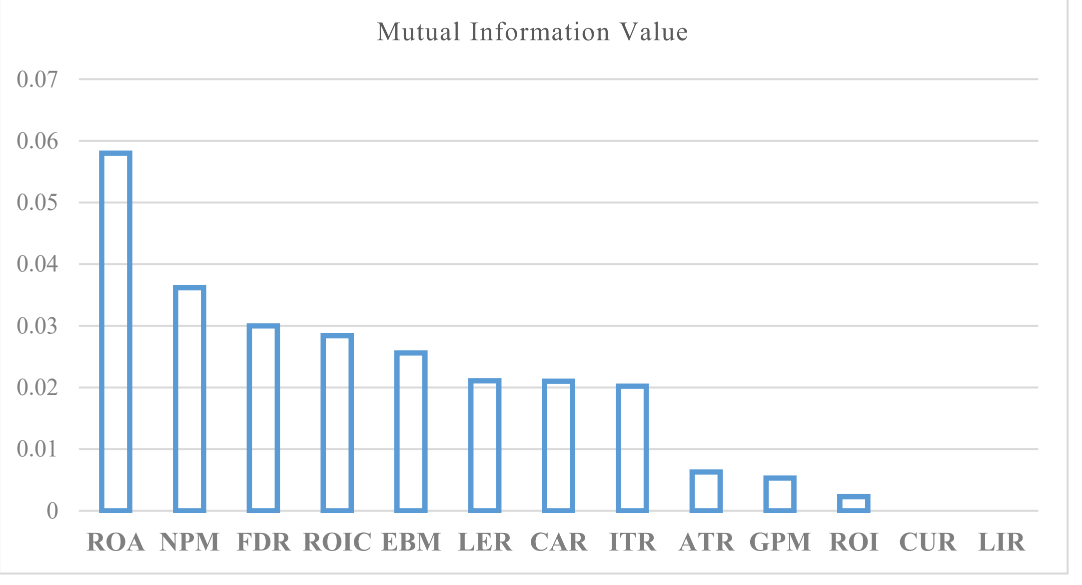

Figure 2. Feature importance values

Considering that there are many hyperparameters for machine learning algorithms and many values that these hyperparameters can take, it is necessary to choose the best parameter combination to be used in SVM, RF, DT, KNN, NB, and XGBoost methods. In this respect, especially when working with small data sets, GridSearchCV (GridSearch+Cross-Validation) hyperparameter optimization method, which gives successful results in determining the hyperparameter set with the best performance, was preferred (Shuai et al., 2018). In this method, separate models are set with all combinations for the hyperparameters and values desired to be tested, and the most successful hyperparameter set is determined according to the metrics established. The parameters and their potential values used in the GridSearchCV method in the study are presented in Table 4.

Table 4. Values Used in Parameter Optimization

Models	Parameter	Parameter's Value
SVM	C	0.1,1,5,10,15,20,100,1000
	kernel	rbf, linear, poly, sigmoid
	degree	3,8
RF	n estimators	1,5,10,50,100
	max depth	4,5,6,7,8,9,10
	min samples leaf	2,10,20
	min samples split	10,15,20
	criterion	gini, entropy

	max features	auto, sqrt
DT	max leaf nodes	2-100
	min samples leaf	1-5
	min sample split	2,3,4
	criterion	gini, entropy
	max depth	2,3,4,5,6,7,8,9,10,11,12
	max features	auto, sqrt, log2
KNN	n-neighbors	range (1,30)
	leaf_size	20,40,1
	p	1,2
	weights	'uniform', 'distance'
	algorithm	'auto', 'ball_tree', 'kd_tree', 'brute'
	metric	'Minkowski', 'chebyshev'
NB	Priors	[None, [0.1,]* len (n_classes),]
	var_smoothing	[1e-9, 1e-6, 1e-12]
XGBoost	Objective	['binary: logistic']
	learning_rate	[0.001, 0.01, 0.1, 0.20, 0.25, 0.30]
	max_depth	[3,4,5,6,8,10,12,15]
	min_child_weight	[1,5,10,11]
	subsample	[0.8]
	colsample_bytree	[0.7]
	n_estimators	[5,100,500,1000]

5.1. Findings

After the parameters required for the models were determined with the GridSearchCV method, with each of the SVM, RF, DT, KNN, NB, and XGBoost machine learning methods, 66 estimation models for 11 different variable groups were established. In order to measure the success rate of the estimation models, the performance values used for classification like accuracy, precision, recall, f1-score, and Cohen's Kappa values were calculated and presented in Table 5.

Table 5. Prediction Result of Models According to the Variants of Test Input Variable

Model	Input Variable Group	Accuracy	Precision	Recall	F1 Score	Kappa
	IVG (1)-(3)	0.621	0.521	0.621	0.501	-0.006
	IVG (1)-(4)	0.642	0.616	0.642	0.568	0.090
	IVG (1)-(5)	**0.747**	**0.750**	**0.747**	**0.729**	**0.408**
	IVG (1)-(6)	0.716	0.717	0.716	0.688	0.321

SVM	IVG (1)-(7)	0.747	0.742	0.747	0.739	0.430
	IVG (1)-(8)	0.705	0.696	0.705	0.695	0.335
	IVG (1)-(9)	0.737	0.768	0.737	0.700	0.354
	IVG (1)-(10)	0.695	0.684	0.695	0.683	0.307
	IVG (1)-(11)	0.684	0.672	0.684	0.670	0.278
	IVG (1)-(12)	0.684	0.673	0.684	0.656	0.250
	IVG (1)-(13)	0.695	0.689	0.695	0.665	0.270
RF	IVG (1)-(3)	0.653	0.632	0.653	0.625	0.180
	IVG (1)-(4)	0.726	0.735	0.726	0.697	0.341
	IVG (1)-(5)	0.716	0.711	0.716	0.713	0.378
	IVG (1)-(6)	0.758	0.753	0.758	0.751	0.457
	IVG (1)-(7)	0.758	0.757	0.758	0.745	0.443
	IVG (1)-(8)	0.768	0.765	0.768	0.761	0.478
	IVG (1)-(9)	0.789	0.791	0.789	0.780	0.519
	IVG (1)-(10)	0.789	0.787	0.789	0.786	0.537
	IVG (1)-(11)	0.705	0.698	0.705	0.698	0.343
	IVG (1)-(12)	**0.811**	**0.808**	**0.811**	**0.808**	**0.583**
	IVG (1)-(13)	0.779	0.776	0.779	0.777	0.516
DT	IVG (1)-(3)	0.632	0.608	0.632	0.608	0.142
	IVG (1)-(4)	0.653	0.647	0.653	0.649	0.240
	IVG (1)-(5)	0.653	0.689	0.653	0.659	0.307
	IVG (1)-(6)	0.705	0.696	0.705	0.695	0.335
	IVG (1)-(7)	0.684	0.698	0.684	0.688	0.345
	IVG (1)-(8)	0.663	0.659	0.663	0.661	0.267
	IVG (1)-(9)	**0.768**	**0.766**	**0.768**	**0.767**	**0.496**
	IVG (1)-(10)	0.674	0.710	0.674	0.679	0.349
	IVG (1)-(11)	0.737	0.759	0.737	0.741	0.463
	IVG (1)-(12)	0.737	0.752	0.737	0.741	0.457
	IVG (1)-(13)	0.705	0.714	0.705	0.708	0.381
NB	IVG (1)-(3)	0.642	0.614	0.642	0.588	0.115
	IVG (1)-(4)	0.663	0.647	0.663	0.620	0.178
	IVG (1)-(5)	0.684	0.673	0.684	0.656	0.250
	IVG (1)-(6)	**0.737**	**0.733**	**0.737**	**0.734**	**0.424**
	IVG (1)-(7)	0.579	0.619	0.579	0.586	0.165
	IVG (1)-(8)	0.579	0.671	0.579	0.578	0.216

	IVG (1)-(9)	0.568	0.653	0.568	0.568	0.193
	IVG (1)-(10)	0.579	0.66	0.579	0.580	0.208
	IVG (1)-(11)	0.568	0.664	0.568	0.566	0.201
	IVG (1)-(12)	0.558	0.657	0.558	0.554	0.186
	IVG (1)-(13)	0.547	0.650	0.547	0.542	0.171
KNN	IVG (1)-(3)	0.663	0.653	0.663	0.655	0.249
	IVG (1)-(4)	0.716	0.708	0.716	0.708	0.363
	IVG (1)-(5)	0.726	0.721	0.726	0.722	0.398
	IVG (1)-(6)	0.747	0.743	0.747	0.744	0.444
	IVG (1)-(7)	0.726	0.721	0.726	0.722	0.398
	IVG (1)-(8)	0.653	0.664	0.653	0.657	0.275
	IVG (1)-(9)	0.789	0.788	0.789	0.782	0.525
	IVG (1)-(10)	**0.811**	**0.811**	**0.811**	**0.804**	**0.572**
	IVG (1)-(11)	0.779	0.775	0.779	0.775	0.510
	IVG (1)-(12)	0.779	0.776	0.779	0.773	0.504
	IVG (1)-(13)	0.768	0.764	0.768	0.763	0.484
XGBoost	IVG (1)-(3)	0.684	0.684	0.684	0.684	0.321
	IVG (1)-(4)	0.716	0.748	0.716	0.669	0.292
	IVG (1)-(5)	0.737	0.735	0.737	0.720	0.387
	IVG (1)-(6)	0.768	0.771	0.768	0.755	0.464
	IVG (1)-(7)	0.758	0.753	0.758	0.751	0.457
	IVG (1)-(8)	**0.884**	**0.884**	**0.884**	**0.884**	**0.750**
	IVG (1)-(9)	0.874	0.874	0.874	0.872	0.722
	IVG (1)-(10)	0.853	0.853	0.853	0.853	0.683
	IVG (1)-(11)	0.821	0.819	0.821	0.819	0.608
	IVG (1)-(12)	0.863	0.863	0.863	0.863	0.704
	IVG (1)-(13)	0.811	0.808	0.811	0.808	0.583

As seen in Table 5, the highest accuracy rate was achieved with different numbers of variables (IVG) for each method in the prediction models created with machine learning methods. The XGBoost model showed the highest performance with an Accuracy value of 0.884, followed by the Random Forest (RF) and K-Nearest Neighbors (KNN) models with Accuracy values of 0.811. The Decision Tree (DT) model showed lower performance with Accuracy values of 0.768, the Support Vector Machine (SVM) model with Accuracy values of 0.747, and the Naive Bayes (NB) model with Accuracy values of 0.737. These results show that XGBoost is more successful than other models in predicting the independent audit firm switch. However, it is thought that this high accuracy rate is not only due to the XGBoost algorithm, but also the features in our dataset and the MI feature importance value method make significant contributions to the accuracy. The fact that different input variable groups change the model performance shows that feature selection is a critical factor that directly affects the model success.

The formulas used in the calculation of the performance metrics in Table 5 are presented below (Briliani, 2019).

The following formula was used to calculate the accuracy value:

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

The following formula was used to calculate the precision value:

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

The following formula was used to calculate the recall value:

Recall=$\frac{TP}{TP+FN}$

The following formula was used to calculate the F1-score value:

F1-score=2* $\frac{Precision * Recall}{Precision + Recall}$

Cohen's Kappa $P = \frac{P_0 - P_e}{1 - P_e}$

P_0 =Relative observed agreement among accuracy.

P_e = Hypothetical probability of chance agreement

In the formulas, TP represents the real positive number, TN the real negative number, FP the incorrect positive number, and FN the incorrect negative number. In Figure 3, where the performance metrics of the machine learning algorithms for the test results are presented, it was observed that the XGBoost algorithm yielded a more successful estimation result compared with other machine learning systems.

Figure 3. Comparison of ML Algorithms According to Performance Metrics

When ranked according to performance metrics, the best performance was exhibited by the 8-variable XGBoost algorithm - ROA, NPM, FDR, ROIC, EBM, LER, CAR, ITR - with an accuracy rate of 88.4% (0.884). It was followed by the 12-variable RF algorithm —ROA, NPM, FDR, ROIC, EM, LER, CAR, ITR, ATR, GPM, ROE, CUR - and the 10-variable KNN algorithm - ROA, NPM, FDR, ROIC, EBM, LER, CUR, ITR, ATR, GPM - .

The eight variables in which XGBoost exhibited the highest performance (ROA, NPM, FDR, ROIC, EBM, LER, CAR, and ITR) play a significant role in predicting independent audit firm switches. The model's ability to achieve high performance with a limited number of variables underscores the impact of feature selection on predictive accuracy. Therefore, these variables should be carefully considered when developing audit firm switch forecasting models. Furthermore, incorporating additional financial metrics during the feature selection process may enhance the overall predictive performance of the model. This approach can contribute to the development of more accurate and robust predictive models.

Table 6. Comparison of the Models According to the Most Successful Variable Groups

Model	Accuracy	Precision	Recall	F1-Score	Cohen's Kappa	Ranking Group
XGBoost	**0.884**	**0.884**	**0.884**	**0.884**	**0.750**	**IVG (1)-(8)**
RF	0.811	0.808	0.811	0.808	0.583	IVG (1)-(12)
KNN	0.811	0.811	0.811	0.804	0.572	IVG (1)-(10)
SVM	0.747	0.75	0.747	0.729	0.408	IVG (1)-(5)
DT	0.768	0.766	0.768	0.767	0.496	IVG (1)-(9)
NB	0.737	0.733	0.737	0.734	0.424	IVG (1)-(6)

According to the performance metrics—precision, recall, F1-score, and Cohen's Kappa—the XGBoost algorithm was identified as the most successful machine learning method, achieving the highest accuracy rate of 88.4%.

The RF and KNN algorithms showed the same performance with an estimation accuracy rate of 81.1%. The SVM, DT, and NB algorithms yielded close values with estimation accuracy values of 74.7%, 76.80%, and 73.7%, respectively. The best accuracy values with SVM, DT, and NB algorithms were obtained in the 5-9-6-variable estimation models, respectively. The DT and NB algorithms showed the lowest performances.

In this study, it was shown through performance metrics that machine learning algorithms can predict audit firm switches with high accuracy. The analysis performed on the dataset used revealed that the XGBoost algorithm can successfully predict independent audit firm switches with lower prediction errors compared to other models such as Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT) and Naive Bayes (NB).

Our study shows that machine learning algorithms are a promising tool for researchers in independent audit firm switch prediction. However, it should not be forgotten that these algorithms have their own limitations. In particular, the selection of features that researchers will include in the models is of great importance. Incorrect or incomplete feature selection can significantly reduce model performance.

6. CONCLUSION

This study aimed to predict independent audit firm switching of the enterprises traded in BIST STARS in Türkiye by using financial ratios and machine learning algorithms. In this study, data from companies operating continuously in BIST STARS between 2019 and 2021 were analyzed. In the following years, changes in market categories and the number of companies listed may occur due to factors such as the periodic performance of companies, sector conditions, new public offerings, and other requirements. Accordingly, the number of companies included in BIST STARS is likely to change over time. The data set of the research consisted of 474 data sets - 158 companies each year - between 2019-2021.

In these estimations, machine learning methods like SVM, RF, DT, KNN, NB, and XGBoost were used. The purpose here is to establish an estimation model that predicts the correlation between dependent and input variables to have the best fit. While the dependent variable in the data sets of the models was the audit firm switch, the input variables were determined as financial ratios - ROA, NPM, FDR, ROIC, EBM, LER, CAR, ITR, ATR, GPM, ROI, CUR, LIR - . Among the 13 attributes used in the study, the highest prediction value was obtained by using the first 8 attributes, respectively. In selecting the attributes, the Mutual Information (MI) method was applied to find importance values. The ranking of importance of the attributes was determined as follows: ROA (Return on Assets), NPM (Net Profit Margin), FDR (Financial Debt Ratio), ROIC (Return on Investment Capital), EBM (EBITDA Margin), LER (Leverage Ratio), CAR (Cash Ratio), and ITR (Inventory Turnover Rate). It was determined that ROA (Return on Assets) data was the most important attribute.

ROA is a critical ratio that shows how effectively and efficiently a company uses its assets. This ratio reflects how well the company's investments in its assets are managed and the company's general financial situation. Companies with high ROA values are generally companies with strong financial structures and growth potential. Therefore, the high ROA has been seen as an important indicator for audit firm switching.

Each of the financial ratios used in the model provides information about the company's financial performance and financial structure from different perspectives. While ratios that express profitability, such as NPM, reveal the company's financial success and operational efficiency, debt ratios, such as FDR and LER, reflect the company's financial risks. Companies with high profitability and liquidity tend to prefer to work with more prestigious and large auditing firms. Companies with high financial debt ratios, on the other hand, need to work with more competent and reliable auditing firms during the auditing process as their financial risks increase. This may lead to a switch in the auditing firm.

When compared with literature, our study has achieved higher overall prediction accuracy than the studies of Kirkos (2012) and Kwak et al. (2011) who tried to predict audit firm switching with financial ratios using data mining methods. While Kirkos (2012) achieved 84.6% overall prediction accuracy with the Bayesian Network method, Kwak et al. (2011) achieved 60% overall prediction accuracy with all the methods used, in our study 88.4% overall prediction accuracy was achieved with the XGBoost method. Black et al. (2013) found 92.5% overall accuracy rate in predicting audit firm switching using the logistic regression method. This rate is higher than the prediction accuracy rate we obtained in our study. The reason for the high prediction accuracy is thought to be due to the different countries and time period of the data set.

In our study, in comparison with the models that showed the highest success in predicting audit firm switching, it was seen that financial ratios were effective. For example, in Kirkos (2012) study, our common variable was Net Profit Margin (NPM). Kwak et al. (2011), Leverage Ratio (LER) and EBITDA Margin (EBM), which were found to be important for audit firm switching, also came to the fore in our model. Similarly, in Black et al. (2013) study, a common finding was obtained on Return on Assets (ROA). These variables are consistent with different studies, supporting the strength and validity of our model.

66 estimation models were established with each of the SVM, RF, DT, KNN, NB, and XGBoost machine learning methods for 11 different variable groups. To measure the success rate of the estimation models, the performance values used for classification like accuracy, precision, recall, f1-score, and Cohen's Kappa values were calculated. When ranked according to performance metrics, the best performance was displayed by the 8-variable XGBoost algorithm - ROA, NPM, FDR, ROIC, EBM, LER, CAR, ITR - with an accuracy rate of 88.4%.

With these and similar artificial intelligence-supported estimation studies, it is thought that public companies will be able to make accurate estimations regarding their audit firm switching. In future studies, it is aimed to carry out estimation studies with different machine learning methods by creating larger data sets.

Hakem Değerlendirmesi: Dış Bağımsız

Çıkar Çatışması: Yazar(lar) çıkar çatışması bildirmemiştir.

Finansal Destek: Yazar(lar) bu çalışma için finansal destek almadığını belirtmiştir.

Etik Onay: Bu makale, insan veya hayvanlar ile ilgili etik onay gerektiren herhangi bir araştırma içermemektedir.

Yazar(lar) Katkısı: Ahmet ÇANKAL (%50), Erdem KÜRKLÜ (%50)

Peer-review: Externally peer-reviewed.

Conflict of Interest: The author(s) declares that there is no conflict of interest.

Funding: The author(s) received no financial support for the research, authorship and/or publication of this article.

Ethical Approval: This article does not contain any studies with human participants or animals performed by the authors.

Author(s) Contributions: Ahmet ÇANKAL (%50), Erdem KÜRKLÜ (%50)

REFERENCES

Abu Alfeilat, H. A., Hassanat, A. B., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., Eyal Salman, H. S., and Prasath, V. S. (2019). Effects of distance measure choice on k-nearest neighbor classifier performance: a review. *Big data*, 7(4), 221-248. doi: 10.1089/big.2018.0175

Adjirackor, T., Asare, D. D., Asare, F. D., and Gagakuma, W. (2017). Financial ratios as a tool for profitability in Aryton drugs. *Research Journal of Finance and Accounting, 8*(14).

Aisyah, L. and Faridah, R. (2023). Voluntary Auditor Switching in Listed Companies: What Influences It?. Asian Journal of Islamic Economics and Business, 1(1), 42-63.

Al-Garadi, M. A., Mohamed, A. K., Al-Ali, X. Du, I. Ali and M. Guizani, (2020). "A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security," in IEEE Communications Surveys and Tutorials, vol. 22, no. 3, pp. 1646-1685, doi: 10.1109/COMST.2020.2988293

Altass, S. (2023). Auditor Switching, Tenure, and Corporate Performance: The Saudi Evidence. Quality-Access to Success, 24(192).

Barros, R. C., Basgalupp, M. P., Carvalho, A. C., and Freitas, A. A. (2012). A hyper-heuristic evolutionary algorithm for automatically designing decision tree algorithms. *Gecco'12,* 1237-1244. doi: https://doi.org/10.1145/2330163.2330335

Black, E. L., Burton, F. G., and Maggina, A. G. (2013). Auditor switching in the economic crisis: The case in Greece'', *International Journal of Accounting and Economic Studies*, 1(2), 39-46.

Boulesteix, A. L., Janitza, S., Kruppa, J., and König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(6), 493-507. doi: https://doi.org/10.1002/widm.1072

Briliani, A., Irawan, B., and Setianingsih, C. (2019). Hate speech detection in Indonesian language on Instagram comment section using K-nearest neighbor classification method'', *2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS).* doi:

10.1109/IoTaIS47347.2019.8980398

Calderon, T. G., and Ofobike, E. (2008). Determinants of client-initiated and auditor-initiated auditor changes. *Managerial Auditing Journal*, *23*(1), 4-25.

Cao, L. J. and Tay, F. E. H. (2003). Support vector machine with adaptive parameters in financial times series forecasting. *IEEE Transactions on Neural Networks*, 14(6).

Chen, C. L., Chang F. H. and Yen G. (2008). The information contents of auditor change infinancial distress prediction-empirical findings from the TALEX-listed firms. Draft of paper retrieved from www.google.com, at January 12.

Chen, J., Wang, X., and Zhai, J. (2009). Pruning decision tree using genetic algorithms. *IEEE Computer Society,* doi: 10.1109/AICI.2009.351

Chen, T., and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acmsigkdd international conference on knowledge discovery and data mining,* San Francisco, California, USA, p.785-794.

Cheng, C. B., Chen, C. L., and Fu, C. J. (2005). Financial distress prediction by a radial basis function network with logit analysis learning. *An International Journal Computers and Mathematics with Applications,* doi: 10.1016/j.camwa.2005.07.016

Chung, H., Kim, Y. and Sunwoo, H. Y. (2021). Korean evidence on auditor switching for opinion shopping and capital market perceptions of audit quality. Asia-Pacific Journal of Accounting & Economics, 28(1), 71-93.

Chunhong, Z., and Licheng, J. (2004). Automatic parameters selection for SVM based on GA. *Proceedings of the 5th World Congress on Intelligent Control and Automation*, June 15-19.

Ding, F., Qiao, Z., Hu, M. and Lu, M. (2022). Industry specialization and audit quality: Evidence from audit firm switches in China. Asia-Pacific Journal of Financial Studies, 51(5), 657-681.

Eldridge, S., Kwak, W., Venkatesh, R., Shi, Y. and Kou, G. (2012). Predicting auditor changes with financial distress variables: discriminant analysis and problems with data mining approaches. *The Journal of Applied Business Research*, 28(6), 1357-1372.

Ettredge, M. L., Li, C., and Scholz, S. (2007). Audit fees and auditor dismissals in the Sarbanes-Oxley era. *Accounting Horizons,* 21(4), 371-386.

Ghosh, S., Dasgupta, A. and Swetapadma, A. (2019). A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification. 2019 International Conference on Intelligent Sustainable Systems (ICISS), doi:10.1109/ISS1.2019.8908018.

Guo, Q., Koch, C. and Zhu, A. (2023). Switching Costs and Market Power in Auditing: Evidence from a Structural Approach. Available at SSRN 4166004.

Han, J., Kamber, M., and Pei, J. (2011). *Data mining concepts and techniques*. Third Edition, Morgan Kaufmann, Massachusetts.

Hu, L.Y., Huang, M.W., Ke, S.W., and Tsai, C.F. (2016). The distance function effect on k-nearest neighbor classification for medical datasets. *Springer Plus,* 5(1), 1-9.

Huang, Y. and Scholz, S. (2012). Evidence of the association between financial restatement and auditor resignations. *Accounting Horizon*, 26(3), 439-464.

Ismail, S. H., Aliahmed H., Nassir, A., and Hamid, M. A. (2008). Why Malaysian second board companies switch auditors: Evidence of Bursa Malaysia. *International Research Journal of Finance and Economics*, 13, 123-130.

Jain, S. and Agarwalla, S. K. (2023). Big-4 auditors and audit quality: A novel firm life-cycle approach. Meditari Accountancy Research, 31(5), 1436-1452. https://doi.org/10.1108/MEDAR-06-2021-1344

Kamarudin, K. A., Islam, A., Habib, A. and Wan Ismail, W. A. (2022). Auditor switching, lowballing and conditional conservatism: evidence from selected Asian countries. Managerial Auditing Journal, 37(2), 224-254.

Kirkos, E. (2012). Predicting auditor switches by applying data mining, Journal of Applied Economic Sciences, 7(3), 344-347.

Kwak, W., Eldridge, S., Shi, Y., and Kou, G. (2011). Predicting auditor changes using financial distress variables and the multiple criteria linear programming (MCLP) and other data mining approaches. *The Journal of Applied Business Research*, 27(5), 73-84.

Lei, R., and Liu, H. (2021). Financial distress prediction using GA-BP neural network model. *International Journal of Economics and Finance,* 13(3).

Liaw, A., and Wiener, M. (2022). Classification and regression by random forest. *R News*, 2(3).

Lin, T. H. (2009). A cross model study of corporate financial distress prediction in Taiwan: multiple discriminant analysis, logit, probit and neural networks models. *Neurocomputing,* 72(16), 3507-3516.

Liu, Y. (2021, March). Analysis on Bilibili Video Categorical Segmentation Using XGBoost. In *2021 2nd International Conference on E-Commerce and Internet Technology (ECIT)* (pp. 259-263). IEEE

Nasser, A. T., Wahid, E. A., Nazri, S. N., and Hudaib, M. (2006). Auditor client relationship: The case of audit tenure and auditor switching in Malaysia. *Managerial Auditing Journal*, 21(7), 724-737.

Nassif, A. B., Azzeh, M., Capretz, L. F., and Ho, D. (2013). A Comparison between Decision Trees and Decision Tree Forest Models for Software Development Effort Estimation. *Computational Intelligence Applications in Software Engineering (CIASE),* Beirut.

Nazri, S. N., Smith, M., and Ismail, Z. (2012). Factors influencing auditor change: Evidence from Malaysia. *Asian Review of Accounting*, *20* (3), 22-44.

Okun, O. (2011). *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations*. Information Science Reference - Imprint of: IGI Publishing, Hershey, PA.

Patel, H. H., and Prajapati, P. (2018). Study and analysis of decision tree-based classification algorithms. *International Journal of Computer Sciences and Engineering*, 6(10).

Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8).

Pradhan, A. (2012). Support vector machine a survey. *International Journal of Emerging Technology and Advanced Engineering*, 2(8).

Saalem, Q. and Rehman, R. U. (2011). Impacts of Liquidity Ratios on Profitability. *Interdisciplinary Journal of Research in Business,* 1(7), 95-98.

Saluy, A. B., Kemalsari, N., Handiman, U. T., Arwiya, P., Faridi, A., Caya, B. A. and Machmud, H. (2024). Human Resources Perspective: Audit Fee, Internal Control, and Audit Materiality Affect Auditor Switching. WSEAS Transactions on Business and Economics, 21, 21-34.

Seethapathy, S. K. and Babu, C. N. (2021). Enhanced approach for soil classification using boosted c5.0 decision tree algorithm. *BSSS Journal of Computer:* XII(I),11-21.

Shuai, Y., Zheng, Y., and Huang, H. (2018). Hybrid software obsolescence evaluation model based on PCA-SVM-GridSearchCV,in:*2018 IEEE 9th international conference on software engineering and service science (ICSESS), IEEE (2018) 449–53.* Doi: 10.1109/ICSESS.2018.8663753

Siva, S. S., Geetha, S., and Kannan, A. (2012). Decision tree based light weight intrusion detection using a wrapper approach. *Expert Systems with Applications,* 39, 129-141.

Song, Y. Y., and Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2).

Srinivasa, R., Yashashwini, S., Venkatesh, K. andYaswanth, S. P. (2020). Prediction of Diabetes Using Machine Learning, *International Journal of Advanced Science and Technology*, 29(06), p.7593-7601.

Susanto, Y. K. (2018). Auditor switching: management turnover, qualified opinion, audit delay, financial distress. International Journal of Business, Economics and Law, 15(5), 125-132.

Suyono, E., Feng, Y., and Riswan, M. (2013). Determinant factors affecting the auditor switching: An Indonesian case. *Global Review of Accounting and Finance*, 4(2), 103-116.

Ünal, M. and Altay, A. (2015). Kurumsal yönetimin bağımsız dış denetime etkisi ve denetim firması seçimindeki rolü: BIST imalat sektöründe bir uygulama. *Journal of Accounting and Taxation Studies*, 8(2), 91-106.

Wan Mohamed, W. A., Hussain. W. S., and Mohd Rodzi, N. K. (2007). *Characteristics' of companies that change and do not change and do not change auditor- an empirical investigation of Malaysian public listed companies.* Unpublished manuscript, University of Teknology MARA, Shah Alam, Malaysia.

Witten H., Frank E., Hall M., and Pal C. (2016). *Data mining: practical machine learning tools and techniques*, (4th ed.). Morgan Kaufmann.

Yang, F. J. (2018). An Implementation of Naive Bayes Classifier. *2018 International Conference on Computational Science and Computational Intelligence (CSCI), Doi* 10.1109/CSCI46756.2018.00065.

Yim, J., and Mitcbell, H. (2005). A comparison of corporate distress prediction models in Brazil: hybrid neural networks, logit models and discriminant analysis. *Nova Economia*, 15(1), 73-93.

Zhou, J., Qiu, Y., Khandelwal, M., Zhu, S., and Zhang, X. (2021). Developing a hybrid model of Jaya algorithm-based extreme gradient boosting machine to estimate blast-induced ground vibrations. *International Journal of Rock Mechanics and Mining Sciences,* 145, 21-34.