El Cezeri
Journal of Science and Engineering
ISSN: 2148-3736

Research Article

Audit Opinion Prediction with Data Mining Methods: Evidence from Borsa Istanbul (BIST)

Zafer Kardeş[1a], Tuğrul Kandemir[2b]

[1] Emirdag Vocational School, Afyon Kocatepe University, Afyonkarahisar, 03600, Türkiye
[2]Faculty of Economics and Administrative Sciences, Afyon Kocatepe University, Afyonkarahisar, 03030, Türkiye

zaferkardes@aku.edu.tr

DOI : 10.31202/ecjse.1679862
Received: 19.04.2025 Accepted: 30.06.2025
How to cite this article:
Zafer Kardeş and Tuğrul Kandemir, "Audit Opinion Prediction with Data Mining Methods: Evidence From Borsa Istanbul (BIST)", El-Cezeri Journal of Science and Engineering, Vol: 12, Iss: 3, (2025), pp.(395-409).
ORCID: 0000-0002-5719-8551, 0000-0002-3544-7422.

Abstract This study compares the performance of 12 different data mining methods in predicting audit opinions for companies' financial statements. The study dataset consists of 2,093 firm-year observations from 161 companies listed on Borsa Istanbul for 2010-2022. The independent audit opinion types were classified using a set of 28 independent financial and non-financial variables. The following prediction models were used in the study: Bayesian Networks, Naive Bayes, Logistic Regression, Artificial Neural Networks, Radial Basis Function, Support Vector Machines, K-Nearest Neighbor, AdaBoost.M1 Algorithm, Decision Trees (J48), Random Forest, Decision Stump, and Classification and Regression Tree (CART). According to the results of the analysis, the Random Forest model demonstrated the best performance with a prediction accuracy rate of 96.68% for predicting audit opinions. The statistical results of the models were compared based on prediction accuracy, confusion matrix, detailed accuracy results, Type I error rate, Type II error rate, and performance measures. This study pioneers developing models capable of accurately predicting audit opinions using 2,093 firm-year observations based on financial and non-financial variables. This contributes likely to the audit literature by predicting audit opinion types through data mining classification methods. The framework design used in the study is anticipated to serve as a decision support tool for internal and external auditors, accountants, shareholders, company executives, tax authorities, other public institutions, individual and institutional investors, stock exchanges, law firms, financial analysts, credit rating agencies, and the banking system when making decisions.

Keywords: Audit Opinion Prediction, Auditing, Classification, Data Mining, Machine Learning.

1. INTRODUCTION

According to International Auditing Standards, auditors are required to consider a range of factors related to the risk of material misstatements in their clients' financial statements when designing and implementing audit procedures. Specifically, auditors must "discuss the susceptibility of the entity's financial statements to material misstatements" and "identify and assess the risks of material misstatement (a) at the financial statement level, and (b) at the level of classes of transactions, account balances, and disclosures" [1]. In this context, auditors must develop and use models that predict the type of audit opinion and provide timely and reliable assessments of the risk of material misstatements, given the data derived from the financial statements, the audited entity, and other available company information. By utilizing such models, auditors can scan their client portfolios and focus on clients who are more likely to receive adverse audit opinions. This approach enables auditors to save time and resources and helps mitigate potential risks.

The fundamental concept of auditing is to verify the information presented in financial statements, as this information serves as the basis for decision-making by various groups such as shareholders, potential investors, intermediaries, managers, financial advisors, financial analysts, creditors, and government authorities. From these users' perspective, an audit is effective when auditors confirm that financial statements are free from errors, omissions, fraud, and manipulation. The development and application of new technologies across various disciplines also impact the accounting and auditing fields. One way to enhance the effectiveness of audits is to use audit opinion prediction methods and employ a model tailored to audit conditions and the environment [2].

Data mining is an emerging method that is gaining increasing attention because of its ability to enhance the effectiveness of financial statement audits. Audit opinion models developed by researchers using data mining analytics enable auditors to utilize large datasets during the audit process. Traditional data analysis methods assume that the data are already structured, and that the reliability of the data analysis process has been verified. Therefore, these methods cannot handle large datasets obtained

from diverse sources. However, data mining makes this possible in the audit process by automatically extracting patterns from the data and testing hypotheses [3].

Data mining has become widely applied in various business practices in the fields of accounting and finance. Recently, [4] highlighted that 82% of data mining applications in accounting are focused on predictive studies. Additionally, they noted that the most commonly used data mining techniques in accounting are artificial neural networks, regression, decision trees, and support vector machines.

The primary aim of this study is to explore the capabilities of data mining by comparing the predictive performance of 12 different data mining classification methods for predicting audit opinions. In this context, the study problem, that is, the main study question to be answered, is as follows:

- Which data mining method will demonstrate the best performance in predicting independent audit opinions in Turkey?

In this study, 12 different data mining methods were analyzed, and models were developed using WEKA software based on data related to the type of independent audit opinion. The performance of the models was presented based on the prediction accuracy, confusion matrix, detailed accuracy metrics (TP Rate, FP Rate, Precision, Recall, F-Measure, and ROC Area), Type I and Type II error rates, and performance metrics (Kappa statistic, MAE, RMSE, RAE, and RRSE).

The study sample included 161 companies from 19 sectors. The dataset comprises 2,093 firm-year observations for the period 2010 and 2022. In the models, 24 financial variables and four non-financial variables were used. The results indicate that the independent variables have high explanatory power in explaining the type of audit opinion.

Numerous studies [5], [6], [7] have examined the prediction of audit opinion types. The varying findings of these studies have motivated subsequent study. Most of these studies have been conducted in developed markets, with relatively few analyzing the prediction of audit opinions in emerging markets. This study, which explores data mining capabilities by comparing 12 different classification methods using a Turkey sample, is expected to contribute to the relevant literature. The findings of this study may be valuable for auditors and researchers, and could be a decision-support tool for predicting audit opinions.

Section 2 reviews previous studies on audit opinion prediction and data mining. Section 3 outlines the study's methodology. The findings of this study are discussed in Section 4. Finally, Section 5 present the main conclusions, limitations, and recommendations for future study.

2. LITERATURE REVIEW

In recent years, data mining methods have gained significant attention in the international literature, particularly regarding decision-making processes within various accounting, management, finance, and auditing applications to improve audit opinion prediction and decision support. For instance, auditors have employed data mining techniques to predict audit opinions [5], [6], [7], [8]. Researchers have also used data mining to identify factors influencing audit opinions [9], [10], [11], predict fraudulent financial reporting [12], [13], [14], predict management fraud [15], predict bankruptcy [16], [17], detect financial statement restatements [18], assist in auditor selection [19], evaluate risk [20], apply analytical review procedures [21], and perform continuous auditing [22].

Data mining methods are also gaining attention in Turkey, especially in decision-making processes within various accounting, finance, and auditing applications. For example, data mining methods have been applied to predict audit opinions [23], [24], [25]. Additionally, studies conducted with Turkey samples have used data mining techniques to identify factors influencing audit opinions [26], [27], [28], [29], predict fraudulent financial reporting [30], [31], [32], [33], [34], [35], [36], detect manipulations in financial statements [37], assess audit quality [38], and detect fraud in vehicle insurance claims [39]. Studies that use data mining methods for audit opinion prediction are summarized chronologically in Table 1.

Table 1 presents a comparative overview of the studies based on the periods they covered, the countries in which they were conducted, the size of the datasets used, the independent variables employed, and the data mining methods applied.

Early studies primarily relied on classical statistical modelling techniques, such as probit and logistic regression models [9]. However, more recent studies have shown an increasing preference for modern machine learning techniques, such as decision trees and artificial neural networks [5], [6], [7], [23]. Artificial neural networks appear to be the most frequently used data mining methods in the literature. These are followed by logistic regression, decision trees, discriminant analysis, support vector machines, K-nearest neighbor, naive Bayes, and probit regression. Some studies [9], [40], [41], [43], [44], [49] have focused on a single method, whereas others have conducted comparative analyzes of multiple methods. This study presents a comparative analysis of 12 data mining methods.

Studies present varying levels of accuracy depending on the methodology used, sample characteristics, and country of application. The primary motivation for conducting this study is to compare the performance of data mining methods in predicting audit opinions within a Turkey sample using financial and non-financial ratios that are crucial for the audit process.

Table 1. Summary of Studies on Audit Opinion Prediction

Author(s)	Data Set	Method(s) Used	Findings
[9]	Period: 1969-1980 275 modified and 441 unmodified opinions Country: ABD	Probit Regression	A probit regression model is developed using financial and market variables to predict qualified audit opinions.
[40]	Period: 1992-1994 eight modified and 103 unmodified opinions Country: Finland	Kruskal-Wallis Test and Logistic Regression	The study's results suggest that an effective model can explain the qualifications in the audit reports of publicly traded Finnish companies. The model achieved a 62% prediction accuracy.
[41]	Period: 1997-1999 50 modified and 50 unmodified opinions Country: Greece	Ordinary Least Squares (OLS) and Logistic Regression	The developed model achieved a 78% prediction accuracy.
[42]	Period: 1997-1999 50 modified and 50 unmodified opinions Country: Greece	UTADIS, Discriminant Analysis, and Logistic Regression	The developed UTADIS model achieved a 78.83% prediction accuracy.
[43]	Period: 1998-2003 859 modified and 5.189 unmodified opinions Countries: England and Ireland	Support Vector Machines	The results demonstrate a correct classification performance for 73% of the companies in the sample, with "Credit Rating" highlighted as the most crucial variable.
[44]	Period: 2001 162 modified and 23 unmodified opinions Country: Greece	Ordinary Least Squares (OLS) and Logistic Regression	The developed model achieved a 90% prediction accuracy.
[45]	Period: 1997-2004 264 modified and 3.069 unmodified opinions Country: England	Artificial Neural Network (ANN), Probabilistic Neural Network (PNN), and Logistic Regression (LR)	The probabilistic neural network outperformed traditional artificial neural networks and logistic regression with an 84% classification accuracy.
[46]	Period: 1998-2003 980 modified and 4.296 unmodified opinions Country: England	K-Nearest Neighbor (K-NN), Discriminant Analysis, and Logistic Regression	The K-NN model's overall classification accuracy (66-72.2%) was higher than that of discriminant analysis (60.4-62.4%) and logistic regression (61.1-65.9%).
[8]	Period: 1995-2004 225 modified and 225 unmodified opinions Countries: England and Ireland	Artificial Neural Network, C4.5 Decision Trees, and Bayesian Belief Network	The highest classification performance was shown by Bayesian belief networks at 82.22% followed by the artificial neural network model at 81.11% and the decision tree model at 77.69%.
[6]	Period: 2001-2007 708 modified and 310 unmodified opinions Country: Iran	Support Vector Machines and Decision Trees	A new method combining support vector machines and decision trees showed better performance.
[5]	Period: 2001-2007 347 modified and 671 unmodified opinions Country: Iran	Probabilistic Neural Network (PNN), Multilayer Perceptron (MLP), Radial Basis Function (RBF), and Logistic Regression (LR)	The results show that MLP achieved 88% accuracy in predicting audit opinions and demonstrated the best performance among the methods considered.
[23]	Period: 2010-2013 55 modified and 55 unmodified opinions Country: Turkey	C5.0 Decision Trees, Discriminant Analysis, and Logistic Regression	The classification results of the models revealed that the C5.0 decision tree algorithm had the highest classification accuracy at 98.2% outperforming discriminant and logistic regression models in explaining audit opinions.
[24]	Period: 2011-2014 58 modified and 58 unmodified opinions Country: Turkey	GRI Algorithm, C5.0, and CART Decision Trees	The C5.0 decision tree model showed the highest performance in predicting audit opinion types with an overall correct classification rate of 88.8%.
[7]	Period: 2008-2010 78 modified and 369 unmodified opinions Country: Spain	Probabilistic Neural Network (PNN) and Multilayer Perceptron (MLP)	The study showed accuracy performances of 98.10% and 81.70% for the probabilistic neural network and the multilayer perceptron, respectively, on the test sample.

Author(s)	Data Set	Method(s) Used	Findings
[47]	Period: 2010-2013 4.421 modified and 9.140 unmodified opinions Country: Serbia	C5.0, Random Forest, Regularized Random Forest, Stochastic Gradient Boosting, Extreme Gradient Boosting, K-Nearest Neighbor, Multilayer Perceptron, Support Vector Machine, Linear Discriminant Analysis, Logistic Regression, Probit Regression, and Mixed-Effects Logistic Regression	The results show that several methods from both fields achieved comparable prediction performance at approximately 89% in the first scenario. However, in the second scenario, machine learning algorithms, particularly tree-based ones like Random Forest, performed significantly better reaching 79%.
[25]	Period: 2009-2019 868 modified and 6.259 unmodified opinions Country: Turkey	K-Nearest Neighbor, Decision Trees, Logistic Regression, Discriminant Analysis, Multilayer Perceptron, Naive Bayes, Random Gradient, Random Forest, AdaBoost, and XGBoost	The Random Forest (94.7%) and XGBoost (95.0%) algorithms provided the best results in the study. It is stated that information such as financial ratios, previous period opinion type, previous period independent audit firm class, and audit report delays effectively predict future period results.
[3]	Period: Five annual data 7.199 modified and 23.000 unmodified opinions Countries: England and Ireland	Artificial Neural Network, Support Vector Machines, and K-Nearest Neighbor	The empirical results reveal that the support vector machine and artificial neural network models achieved a higher average accuracy rate (95.9%) and outperformed the K-Nearest Neighbor method (93.8%).
[48]	Period: 2018-2019 9.363 modified and 16.644 unmodified opinions Country: England and Ireland	Linear Discriminant Analysis, Logistic Regression, Naive Bayes, and Decision Trees	The results show that the decision tree model achieved a correct classification performance of 96.3%. The decision tree model outperformed the logistic regression, linear discriminant, and naive Bayes network models on the 2018 and 2019 datasets.
[49]	Period: 2012-2016 251 modified and 229 unmodified opinions Country: Iran	Probit Model	The model shows a 72.9% accuracy in correctly classifying the overall sample to predict the type of auditor's opinion.
[50]	Period: 2018 year 25 modified and 898 unmodified opinions Country: China	BP Neural Network and Adam Optimizer Algorithm	The results illustrate that the Adam optimizer model achieved the highest prediction accuracy at 94.05%.
[51]	Period: 2001-2017 25.943 unqualified, 10.217 unqualified with explanatory language and 1.165 going-concern opinions Country: ABD	Support Vector Machines, Decision Trees, K-Nearest Neighbor (K-NN), and Rough Clusters	This study compares the performance of four data mining techniques in predicting audit opinions on companies' financial statements. Support Vector Machines demonstrated the highest overall prediction accuracy (75.6%), Type I, and Type II error rates.

3. METHODOLOGY

4.1. Study Purpose

This study aims to compare the prediction performance of data mining classification methods using these methods to predict audit opinions.

4.2. Data Set

The statistical population comprises all the companies listed on the BIST between 2009 and 2022. The sample for this study included companies that met the following criteria:

- Companies continuously traded on BIST from 2009 to 2022 were included. Independent audit reports must be available for year t-1 and year t, and all companies must have their financial statements available for year t. Company data has been accessible on the KAP (Public Disclosure Platform) website since 2009. The independent variables used in this study include the opinion of the previous year's audit. Therefore, the 2009 audit report data were used for the 2010 data. Based on this, the starting year for the dataset was set to 2010.
- Banks, financial institutions, investment firms, financial intermediaries, and holding companies were excluded from the sample because of their distinct reporting structures.

After these restrictions were applied, the annual sample size was 161 companies. Table 2 presents the selection of companies included in the study sample.

The selection of sample companies and applied restrictions is presented in Table 2. During the data collection process, the total number of companies operating on BIST was 651. A total of 250 companies whose financial statements and independent auditor reports were unavailable between 2009 and 2022 were excluded from the sample. Additionally, 240 companies comprising

financial institutions, brokerage firms, banks, holding companies, and investment firms were excluded because of their distinct reporting structures.

Table 2. Selection of Companies in the Study Sample

Explanation	Number of Companies
Total Number of Companies Listed on BIST	651
Exclusions:	
1. Companies without Financial Statements and Independent Auditor Reports for 2009-2022	(250)
2. Financial Institutions, Brokerage Firms, Banks, Holdings, and Investment Firms	(240)
Final Sample Size	161

Companies listed under these restrictions were excluded from the sample to ensure more consistent and meaningful results. This approach helps prevent study findings from being influenced by the structural differences between companies. The final sample comprised of 161 companies.

The dataset used in the study consists of 13 years of data (2010-2022) from 161 companies comprising 2,093 firm-year observations. Table 3 presents the distribution of audit opinion types for the 2,093 observations.

Table 3. Distribution of Dataset by Audit Opinion Types

Total Observations	Unmodified Opinion Observations (n)	Modified Opinion Observations (n)
2.093	1.855	238

Table 3 illustrates the distribution of the dataset by audit opinion types. Modified audit opinions include qualified, disclaimer, and adverse opinion types. Table 4 shows the breakdown of modified audit opinions which accounted for 11.37% (238 observations) of the dataset.

Table 4. Distribution of Modified Audit Opinions

Qualified Opinion (n)	Disclaimer of Opinion (n)	Adverse Opinion (n)
223	15	-

When examining the distribution of modified audit opinions, 93.70% (223 observations) are qualified opinions, while 6.30% (15 observations) are disclaimers. There were no audit reports with adverse opinions from the observations included in the study.

4.3. Dependent and Independent Variables

This study utilized both financial and non-financial variables to predict audit opinions. The independent variables consist of a new set of variables inspired by those commonly used in the literature and are thought to influence audit opinions. The Stockeys Pro database was used to construct a dataset for financial variables. The variable set includes 28 independent financial and non-financial variables derived from each company's annual financial statements and audit reports, along with a dependent variable representing modified and unmodified audit opinions. Information on the dependent variable, the four non-financial independent variables, and the 24 financial independent variables used in the study are presented in Table 5.

Table 5 presents the variables used in this study. The set of independent variables consists of both financial and non-financial variables. Financial variables are divided into six categories: (1) liquidity ratios, (2) profitability ratios, (3) valuation ratios, (4) growth ratios, (5) financial structure ratios, and (6) activity ratios. The non-financial variables include auditor change, auditor size, previous year's audit opinion, and audit report delay. Among the independent variables used in the study, three variables - auditor change, auditor size, and the previous year's audit opinion (X1-X3) - are nominal and binary (1/0). The remaining 25 independent variables (X4-X28) are numeric.

4.4. Data Preprocessing Process

The dataset prepared for this study contained 38 attributes and 2,094 rows containing 79,572 data cells to predict audit opinions. The final dataset included 29 attributes consisting of one dependent variable and 28 independent variables. Among the financial variables, 446 instances of missing data were missing from the company-year observations. Missing values in the dataset were replaced with the mean values of the respective attributes.

The models used in this study were analyzed using WEKA software. After loading the dataset into WEKA, the "Interquartile Range" filter is applied using a filtering method. The "Interquartile Range" filter detects outliers and extreme values based on interquartile ranges in the observations. Interquartile Range, a filter for detecting outliers and extreme values based on the interquartile range. The filter skips the class attribute.

Outliers: Q3 + OF*IQR < x <= Q3 + EVF*IQR or Q1 - EVF*IQR <= x < Q1 - OF*IQR

Extreme values: x > Q3 + EVF*IQR or x < Q1 - EVF*IQR

Key:

Q1 = 25% quartile

Q3 = 75% quartile
IQR = Interquartile Range, difference between Q1 and Q3
OF = Outlier Factor
EVF = Extreme Value Factor

Following the application of this filter, two nominal variables, "Outlier" and "Extreme Value", were added to the dataset. Table 6 provides the information on these variables.

Table 5. Variables Used in the Study

Variable Group	Variables
Dependent Variable:	
Y	Audit Opinion Type (1: Unmodified audit opinion; 0: Modified audit opinion)
Non-Financial Independent Variables:	
X_1	Did the Auditor Change? (1: Auditor has not changed 0: Auditor has changed)
X_2	Auditor Size (1: Big Four Audit Firms; 0: Big Four Non-audit Firms)
X_3	Prior Year Audit Opinion (1: Unmodified audit opinion; 0: Modified audit opinion)
X_4	Audit Report Delay (Number of days between the end of the fiscal year and the audit report announcement date)
Financial Independent Variables:	
X_5 Liquidity Ratios	Current Ratio
X_6 Liquidity Ratios	Liquid Ratio
X_7 Liquidity Ratios	Cash Rate
X_8 Profitability Ratios	Return on Assets (ROA) (%)
X_9 Profitability Ratios	Gross Margin (%)
X_{10} Profitability Ratios	Operating Margin (%)
X_{11} Profitability Ratios	EBITDA Margin (%)
X_{12} Profitability Ratios	Net Profit Margin (%)
X_{13} Profitability Ratios	Return on Equity (ROE) (%)
X_{14} Profitability Ratios	EBIT Margin (%)
X_{15} Valuation Rates	Market Value/Book Value
X_{16} Growth Rates	Active Growth (%)
X_{17} Growth Rates	Net Sales Growth (%)
X_{18} Growth Rates	Equity Growth (%)
X_{19} Financial Structure	Debt to Capital Ratio (%)
X_{20} Financial Structure	Fixed Assets / Assets (%)
X_{21} Financial Structure	Short Term Debt / Assets (%)
X_{22} Financial Structure	Short Term Debt / Total Debt (%)
X_{23} Financial Structure	Equity / Assets (%)
X_{24} Financial Structure	Debt to Equity Ratio (%)
X_{25} Activity Ratios	Asset Turnover Ratio
X_{26} Activity Ratios	Receivables Turnover Ratio
X_{27} Activity Ratios	Stock Turnover Ratio
X_{28} Activity Ratios	Trade Payables Turnover Ratio

Table 6. Information on Outliers and Extreme Values

Variable Name	Label	Number	Percent (%)
Outliers	No	1.487	71.05
Outliers	Yes	606	28.95
Extreme Values	No	1.625	77.64
Extreme Values	Yes	468	22.36

In the dataset, if an observation is classified as an outliers or extreme value, it is labelled with "Yes." If it is not, it is labelled with "No." In Table 6, regarding the outliers variable, 71.05% (1,487 observations) of the 2,093 observations in the dataset are labelled as "No," while 28.95% (606 observations) are labelled as "Yes." For the extreme values variable, 77.64% (1,625 observations) are labelled as "No," and 22.36% (468 observations) are labelled as "Yes." When considering the outliers and extreme values variables in general, it was observed that there were not too many outliers or extreme values.

Table 7. Dataset Information after Removing Outliers and Extreme Values

Variable Name	Label	Number	Percent (%)
Outliers	No	1.286	100.00
Outliers	Yes	-	-
Extreme Values	No	1.286	-
Extreme Values	Yes	-	-
Audit Opinion Type	1	1.189	92.46
Audit Opinion Type	0	97	7.54

In total, 606 observations are labelled "Yes" for the outliers variable, and 468 observations are labelled "Yes" for the extreme values variable. These values were excluded from the dataset to ensure meaningful results. The "Remove With Values" filter removes the outliers and extreme values from the dataset. After this process, information regarding the dataset, including outliers, extreme values, and dependent variables, is presented in Table 7.

After removing outliers and extreme value, no observations are labelled "Yes." That is, none of the 1,286 remaining observations contained outliers or extreme values. Following the removal of these values, the dependent variable consisted of 92.46% (1,189 observations) unmodified audit opinions and 7.54% (97 observations) modified audit opinions.

To address the issue of dataset imbalance, we apply the SMOTE (Synthetic Minority Over-sampling Technique) method increasing the number of non-unqualified audit opinion observations from 97 to 1,188. After the SMOTE application, the distribution of the dependent variable became 1,189 (50.02%) unmodified audit opinions and 1,188 (49.98%) modified audit opinions. We resolved the data imbalance issue by applying the SMOTE method, achieving a more even distribution between the two categories.

4.5. Data Mining Classification Methods Used in the Study

We utilized 12 different data mining classification methods to predict independent audit opinions. Table 8 lists the methods used in this study.

Table 8. Data Mining Classification Methods Used in the Study

Serial No.	Method
1	Bayesian Belief Network (BBN)
2	Naive Bayes (NB)
3	Logistic Regression (LR)
4	Artificial Neural Networks (ANN)
5	Radial Basis Function (RBF)
6	Support Vector Machines (SVM)
7	K-Nearest Neighbor (K-NN)
8	AdaBoost.M1 Algorithm
9	Decision Trees J48 (DT)
10	Random Forests (RF)
11	Decision Stump (DS)
12	Classification and Regression Tree (CART)

The 12 different data mining classification methods listed in Table 8 were used to establish the models and perform the analyses. To ensure that the results obtained are reliable and generalizable, 12 different methods were preferred instead of just a few methods. The methods considered were diversified and included probabilistic, statistical, rule-based, machine learning-based, and ensemble methods. This made it possible to objectively compare which model would perform better. As the methods analyzed are common data mining and machine learning algorithms, this study also offered the possibility of comparison with other studies in the literature.

The data mining classification methods used in the study were analyzed using WEKA software. The default parameter settings of the algorithms were used in all analyses. This made the model comparisons repeatable and more objective. The default parameters of the relevant algorithms are accessible via the WEKA written interface.

4. FINDINGS

This section presents the results of the models that were established to predict independent audit opinions. First, we provided descriptive statistics for the variables, followed by the model results displayed in separate tables based on the classification matrix and detailed accuracy metrics.

4.1. Descriptive Statistics on Variables

Table 9 presents the descriptive statistics.

Table 9. Descriptive Statistics for Study Variables

Panel A. Descriptive Statistics for Nominal Data				
Variable Name	Label	Explanation	Number	Percent (%)
Opinion Type	1	Unmodified Opinion Type	1.189	50.02
	0	Modified Opinion Type	1.188	49.98
Did the Auditor Change?	1	Auditor Unchanged	1.930	81.19
	0	Auditor Changed	447	18.81
Auditor Size	1	Big Four Audit Firms	1.376	57.89
	0	Big Four Non-audit Firms	1.001	42.11
Prior Year Audit Opinion	1	Unmodified Opinion Type	1.343	56.50
	0	Modified Opinion Type	1.034	43.50
Panel B. Descriptive Statistics for Numerical Data				

Variable Name	Minimum	Maximum	Mean	StdDev
Audit Report Delay	25	104	65.819	13.651
Current Ratio	0.15	4.93	1.58	0.777
Liquid Ratio	0.04	3.54	0.959	0.573
Return on Assets (ROA) (%)	0	1.671	0.229	0.283
Gross Margin (%)	-0.27	0.407	0.033	0.077
Operating Margin (%)	-0.128	0.575	0.194	0.119
EBITDA Margin (%)	-0.255	0.412	0.055	0.097
Net Profit Margin (%)	-0.203	0.49	0.097	0.094
Return on Equity (ROE) (%)	-0.345	0.42	0.026	0.113
EBIT Margin (%)	-0.622	0.836	0.073	0.171
Market Value/Book Value	-0.261	0.732	0.135	0.111
Active Growth (%)	0.17	11.08	1.942	1.503
Net Sales Growth (%)	-0.507	1.304	0.18	0.222
Equity Growth (%)	-0.835	1.455	0.193	0.285
Debt to Capital Ratio (%)	-0.833	1.552	0.169	0.275
Fixed Assets / Assets (%)	0.052	1.733	0.553	0.239
Short Term Debt / Assets (%)	0.002	0.979	0.477	0.194
Short Term Debt / Total Debt (%)	0.042	1.257	0.394	0.188
Equity / Assets (%)	0.121	0.997	0.717	0.167
Debt to Equity Ratio (%)	-0.736	0.948	0.434	0.236
Asset Turnover Ratio	-5.371	8.248	1.37	1.697
Receivables Turnover Ratio	0.03	3.44	0.9	0.501
Stock Turnover Ratio	0.36	21.05	6.041	3.652
Trade Payables Turnover Ratio	0.19	36.08	6.327	6.806

Panel C. Descriptive Statistics for Outliers and Extreme Value

Variable Name	Label	Data Type	Number	Percent (%)
Outlier	No	Nominal	2.377	100.00
	Yes	Nominal	-	-
Extreme Value	No	Nominal	2.377	100.00
	Yes	Nominal	-	-

Descriptive statistics for the study variables are presented in three panels: nominal data in Panel A, numerical data (including minimum, maximum, mean, and standard deviation values) in Panel B, and outliers and extreme values statistics in Panel C. Panel A, which contains descriptive statistics for nominal data, shows the distributions of variables, such as audit opinion, auditor change, auditor size, and the previous year's audit opinion. Panel B, which presents the descriptive statistics for the numerical data, displays the numerical variables' minimum, maximum, mean, and standard deviation values. Panel C provides descriptive statistics for outliers and extreme values, showing the distributions of these variables.

A regression model was established to determine which of the variables had a significant influence on the dependent variable. The result of the regression model showed that the variables trade payables turnover, stock turnover, prior year audit opinion, liquid ratio, ebitda margin, debt to capital ratio, receivables turnover, return on assets, and asset turnover were more important than the other variables. The R-squared value, which was calculated to test the significance of the model created, was found to be 0.9058. This value is quite high and shows that the role of the variables in explaining the dependent variable is strong.

4.2. Comparison of Data Mining Classification Methods Based on Prediction Performances

This study analyzes data related to the independent audit opinion type using 12 different data mining methods in the WEKA software, and models were developed. This section presents a comparison of the statistical results of the models based on prediction accuracy, confusion matrix, detailed accuracy metrics (TP Rate, FP Rate, Precision, Recall, F-Measure, and ROC Area), Type I and Type II error rates, and performance metrics (Kappa statistic, MAE, RMSE, RAE, and RRSE).

The classification models used in this study were divided into five categories: Bayes, Functions, Lazy, Meta, and Trees. Bayesian Networks and Naive Bayes were used for the Bayes category. For the functions category, Logistic Regression, Artificial Neural Networks, Radial Basis Function, and Support Vector Machines were applied. K-Nearest Neighbor was employed in the Lazy category, while AdaBoost.M1 was used for the Meta category. For the Trees category, the J48 Decision Tree, Random Forest, Classification and Regression Tree (CART), and Decision Stump were used. Each of these classifiers has its own advantages and disadvantages. This study aims to determine the method that performs better in the context of audit opinion prediction. The prediction accuracy, performance, and confusion matrices of the models are listed in Table 10.

Table 10 presents the prediction accuracy rates and confusion matrices of the data mining models. The BBN, NB, LR, ANN, RBF, SVM, K-NN, AdaBoost.M1, DT, RF, DS, and CART models show overall prediction performances of 90.66%, 85.36%, 91.21%, 96.05%, 91.88%, 91.33%, 96.42%, 91.17%, 95.62%, 96.68%, 91.33%, and 94.11%, respectively. As seen in Table 10, the best performance in building the prediction model was achieved by Random Forest (RF), with an accuracy rate of 96.68%. K-Nearest Neighbor (K-NN) was the second-best model, with an accuracy of 96.42%. Naive Bayes (NB) shows the lowest performance among the models, with an accuracy rate of 85.36%. This result suggests that the NB model is weaker than the other

models in predicting the correct audit opinion type. When comparing the models in terms of prediction accuracy, it is clear that the RF model outperformed the others. In summary, the RF model is more effective in predicting audit opinion types than the other models. Some studies in the literature [25], [47] also indicate that the Random Forest model performs better than other models.

When examining the confusion matrices in Table 10, the support vector machines (SVM) and decision stump (DS) models show the highest performance in predicting unmodified opinions, with an accuracy of 97.81%. The model with the lowest performance in predicting unmodified opinions was naive Bayes (NB), with 84.44%. For predicting modified opinions, the K-nearest neighbor (K-NN) demonstrated the highest performance at 97.31%, followed by the Random Forest (RF) model at 96.55%. The models with the lowest performance in predicting modified opinions were support vector machines (SVM) and decision stump (DS), at 84.85%.

Table 10. Data Mining Classification Methods Used in the Study

Model	Prediction Accuracy (%)	Confusion Matrix		
BBN	90.66	a	b	
		1073	116	a=1
		106	1082	b=0
NB	85.36	a	b	
		1004	185	a=1
		163	1025	b=0
LR	91.21	a	b	
		1133	56	a=1
		153	1035	b=0
ANN	96.05	a	b	
		1137	52	a=1
		42	1146	b=0
RBF	91.88	a	b	
		1140	49	a=1
		144	1044	b=0
SVM	91.33	a	b	
		1163	26	a=1
		180	1008	b=0
K-NN	96.42	a	b	
		1136	53	a=1
		32	1156	b=0
AdaBoost.M1	91.17	a	b	
		1154	35	a=1
		175	1013	b=0
DT J48	95.62	a	b	
		1133	56	a=1
		48	1140	b=0
RF	**96.68**	a	b	
		1151	38	a=1
		41	1147	b=0
DS	91.33	a	b	
		1163	26	a=1
		180	1008	b=0
CART	94.11	a	b	
		1119	70	a=1
		70	1118	b=0

Figure 1 shows the prediction accuracy rates of the models arranged from the highest to the lowest.

When examining the prediction accuracies of the models, it is evident that they are effective in accurately predicting audit opinion type. Therefore, the results suggest that the development of an accurate classification model for audit opinion prediction in Turkey could significantly contribute to the field.

Detailed accuracy results based on the weighted averages of the models are presented in Table 11. These results include the weighted averages for the TP Rate, FP Rate, precision, recall, F-measure, and ROC area for each model.

In Table 11, the detailed accuracy results based on the weighted averages of all models are compared. The TP rate represents the ratio of correctly predicted positive instances. The TP rate aligns with the prediction accuracy results shown in Table 10; the higher the TP rate, the better the model performance. The TP rates for the models ranged between 85.4% and 96.7%, indicating that the models can accurately classify audit opinions. The FP rate represents the ratio of incorrectly predicted negative instances. The lower the FP rate, the better the model performance. FP rates for the models ranged from 14.6% to 3.3%. The naive Bayes (NB) model had the highest FP rate at 14.6%, whereas the Random Forest (RF) model had the lowest FP rate at 3.3%. The

precision rates of the models, similar to the TP rates, range from 85.4% to 96.7%. Because recall measures the percentage of the entire dataset, the recall results are parallel to the TP rates of the models.

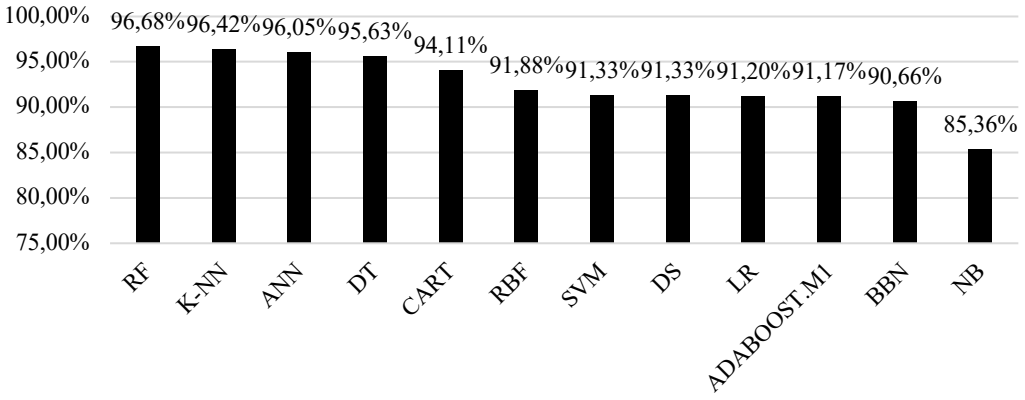

Figure1. Prediction accuracy graph of models

Table 11. Detailed Accuracy Results Based on Weighted Averages of Models

Model	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
BBN	0.907	0.093	0.907	0.907	0.907	0.961
NB	0.854	0.146	0.854	0.854	0.854	0.919
LR	0.912	0.088	0.915	0.912	0.912	0.967
ANN	0.960	0.040	0.960	0.960	0.960	0.984
RBF	0.919	0.081	0.921	0.919	0.919	0.967
SVM	0.913	0.087	0.920	0.913	0.913	0.913
K-NN	0.964	0.036	0.964	0.964	0.964	0.965
AdaBoost.M1	0.912	0.088	0.917	0.912	0.911	0.964
DT J48	0.956	0.044	0.956	0.956	0.956	0.951
RF	**0.967**	**0.033**	**0.967**	**0.967**	**0.967**	**0.992**
DS	0.913	0.087	0.920	0.913	0.913	0.903
CART	0.941	0.059	0.941	0.941	0.941	0.955

The F-measure addresses the issue of using a single metric instead of two by combining precision and recall into a single measurement. Specifically, the F-measure is the weighted harmonic mean of precision and recall [52]. The F-measure results for the models range from 85.4% to 96.7%, consistent with the overall prediction accuracy results.

The final evaluation metric presented in Table 11 was the ROC area. Based on the ROC area results obtained using the WEKA software, the RF model demonstrated the best performance of 99.2%. The ANN model followed this trend with 98.4%, the RBF and LR models with 96.7%, the K-NN model with 96.5%, the AdaBoost.M1 model with 96.4%, the BBN model with 96.1%, the CART model with 95.5%, the DT (J48) model with 95.1%, the NB model with 91.9%, the SVM model with 91.3%, and the DS model with 90.3%. The RF model showed the best prediction performance for the ROC area values.

Overall, when assessing the detailed accuracy results of the models, the Random Forest (RF) model consistently delivered the most meaningful results in terms of TP rate, FP rate, precision, recall, F-measure, and ROC area. The FP rate is the most significant at its lowest value, whereas the other metrics are most meaningful at their highest values. The RF model achieved the lowest FP rate at 3.3%, whereas its TP rate, precision, recall, and F-measure were 96.7%, and it achieved the highest ROC area of 99.2%. Among the tested models, the RF model outperformed the others by a considerable margin.

The models were also evaluated based on the Type I and Type II error rates, and the results are presented in Table 12. The Type I and Type II error rates were critical when evaluating the performance of the methods. A Type I error occurs when a modified opinion is classified as unmodified. Conversely, a Type II error occurs when an unmodified opinion is classified as modified. The costs of Type I and Type II errors differed. Classifying a modified opinion as an unmodified opinion can lead to incorrect decisions and severe economic losses, whereas classifying an unmodified opinion as a modified opinion can waste time on additional investigations. Although each model aims to reduce both the Type I and Type II error rates, a model with a Type I error rate lower than the Type II error rate is generally preferred [12]. The K-NN model exhibited the lowest Type I error rate of 2.7%. This was followed by the RF and ANN models, with a Type I error rate of 3.5%. The SVM and DS models showed the lowest Type II error rate of 2.2%, followed by the AdaBoost.M1 model with 2.9%. Overall, the results indicate that the Type I and Type II error classification rates are below 15.2% and 15.6%, respectively, across all models. The RF model had the lowest overall error rate of 3.3%, followed by the K-NN model at 3.6% and the ANN model at 4%.

Table 12. Type I and Type II Error Rates of Models

Model	Type I Error Rate	Type II Error Rate	Overall Error Rate
BBN	0.089	0.098	0.093
NB	0.137	0.156	0.146
LR	0.129	0.047	0.088
ANN	0.035	0.044	0.040
RBF	0.121	0.041	0.081
SVM	0.152	0.022	0.087
K-NN	**0.027**	0.045	0.036
AdaBoost.M1	0.147	0.029	0.088
DT J48	0.040	0.047	0.044
RF	0.035	0.032	**0.033**
DS	0.152	**0.022**	0.087
CART	0.059	0.059	0.059

Although the K-NN classification model performed best with respect to the Type I error rate, and the SVM and DS classification models performed best in terms of the Type II error rate, the RF classification model demonstrated the best overall classification result regarding the total error rate.

In the study, the SMOTE method was applied due to the unbalanced classes, and a balanced data set was obtained. As the model was trained on a dataset after removing outliers and extreme values, it was ensured to be more robust and generalizable than the synthetic data created with SMOTE. The high accuracy rates of the RF, K-NN, and ANN models indicate a possible overfitting of the model to the training data. To evaluate this possibility, not only the accuracy rates but also the type I and type II error rates were analyzed. The obtained low error rates indicate that the models are successful and acceptable in terms of generalizability. Additionally, the comprehensive evaluation using metrics such as Precision, Recall, F-Measure, and ROC Area supports that the models do not overfit to the training data.

The models are evaluated based on their performance results presented in Table 13.

Table 13. Performance Results of Models

Model	Kappa Statistics	MAE	RMSE	RAE (%)	RRSE (%)
BBN	0.813	0.105	0.277	21.00	55.37
NB	0.707	0.164	0.345	32.80	66.03
LR	0.824	0.125	0.253	25.01	50.57
ANN	0.921	0.043	0.187	8.69	37.42
RBF	0.838	0.141	0.250	28.19	50.00
SVM	**0.934**	0.103	0.179	20.54	35.89
K-NN	0.929	**0.036**	0.189	**7.24**	37.80
AdaBoost.M1	0.823	0.128	0.252	25.69	50.32
DT J48	0.913	0.057	0.205	11.42	40.94
RF	**0.934**	0.103	**0.179**	20.54	**35.89**
DS	0.827	0.153	0.276	30.50	55.25
CART	0.882	0.081	0.231	16.19	46.27

Table 13 provides the kappa statistic, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), and Root Relative Squared Error (RRSE) results for the models. Regarding the Kappa statistic, the RF and SVM models demonstrated the highest performance with a value of 0.934. This was followed by K-NN model at 0.929, ANN model at 0.921, DT (J48) model at 0.913, CART model at 0.882, RBF model at 0.838, DS model at 0.827, LR model at 0.824, AdaBoost.M1 model at 0.823, BBN model at 0.813, and NB model at 0.707. Landis and Koch provided ranges for interpreting the kappa statistic presented in Table 14 [53]. These ranges assist in understanding the performance levels of the models based on their kappa values.

Table 14. Interpretation of Kappa Statistics Results

Kappa Statistics Value	Strength of Agreement
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Based on the values provided in the table, when interpreting the results, it is observed that the BBN model demonstrates good performance, while the other 11 models perform very well. A review of the literature reveals that, apart from the study by [47], this study has yet to consider the kappa statistic. This study also considered the kappa statistic when presenting the results.

The MAE, RMSE, RAE, and RRSE values indicate that the closer they are to zero, the better the performance of the model. According to the MAE values, the K-NN model performed best at 3.6%. Based on the RMSE values, the RF and SVM models exhibited the best performances (17.9%). Regarding the RAE values, the K-NN model performed best at 7.24%. Finally, for the RRSE values, the RF and SVM models performed the best, with 35.89%.

5. CONCLUSION

The primary objective of this study is to predict the type of audit opinion using 12 different data mining methods and to compare the performance of these methods. During the data collection process, 651 companies listed on BIST were considered. A total of 250 companies without financial statements and independent audit reports from 2009 to 2022, and 240 companies comprising financial institutions, brokerage firms, banks, holding companies, and investment firms were excluded from the sample to obtain more meaningful results because of their distinct reporting structures. The final sample includes 161 companies from 19 sectors. The dataset consists of 2,093 firm-year observations from companies listed on BIST from 2010 to 2022, including 1,855 unmodified and 238 modified opinions. Among the modified audit opinions, 223 were qualified opinions and 15 were disclaimers of opinion, with no audit reports containing adverse opinions.

Data were collected from the Public Disclosure Platform website, company websites, financial reports, audit reports, and the Finnet Stokeys Pro database. A total of 28 independent financial and non-financial variables were used. The development and application of the classification model were performed using WEKA software. The training and testing phases were conducted using 10-fold cross-validation, in which the dataset was split into layers. In each iteration, ten training data samples were selected, trained with the chosen classifier, and tested on the selected test data.

The “Interquartile Range” filter was applied to detect outliers and extreme values in the dataset. As a result, 606 outliers and 468 extreme values were removed from the dataset, resulting in 1,286 firm-year observations consisting of 1,189 unmodified opinions and 97 modified opinions. In the subsequent phase, a synthetic minority oversampling technique (SMOTE) filter was applied to address the data imbalance issue. After applying this filter, the number of modified opinion observations increased to 1,188, resulting in a final dataset of 2,377 firm-year observations, consisting of 1,189 unmodified and 1,188 modified opinions.

In this study, 12 different data mining classification methods were used: Bayesian networks, naive Bayes, logistic regression, artificial neural networks, radial basis function, support vector machines, K-nearest neighbor, AdaBoost.M1 algorithm, decision trees (J48), random forest, decision stump, and classification and regression tree (CART). The statistical results of the models were compared based on prediction accuracy, confusion matrix, detailed accuracy results (TP Rate, FP Rate, Precision, Recall, F-Measure, and ROC Area), Type I and Type II error rates, and performance metrics (kappa statistic, MAE, RMSE, RAE, and RRSE).

In terms of prediction accuracy, the Random Forest model demonstrated the best performance at 96.68%, followed by the K-nearest neighbor at 96.42%, Artificial Neural Networks at 96.05%, Decision Trees (J48) at 95.62%, Classification and Regression Tree (CART) at 94.11%, Radial Basis Function at 91.88%, Support Vector Machines at 91.33%, Decision Stump at 91.33%, Logistic Regression at 91.21%, AdaBoost.M1 Algorithm at 91.17%, Bayesian Networks at 90.66%, and Naive Bayes at 85.36%. All 12 models showed a prediction accuracy of over 85%. These results indicate the potential of using public financial statements and independent audit report data to accurately classify audit opinions.

Among the methods reported in the literature for predicting audit opinions, the highest prediction accuracy has been achieved by Decision Trees ([23] (98.2%); [24] (88.8%); [48] (96.3%)), Support Vector Machines ([3] (95.9%); [43] (73%); [51] (75.6%)), Logistic Regression ([40] (62%); [41] (78%); [44] (90%)), and Neural Networks ([5] (88%); [7] (98.1%); [45] (84%)).

According to the findings obtained in this study, the Random Forest method, which belongs to the decision tree category, showed the highest prediction performance for audit opinion classification at 96.68%. This result aligns with the studies by [23], [24], [48], who also found that decision tree-based methods perform best. Based on these results, similar to many previous studies, it is evident that the 12 models used in this study can successfully predict audit opinions.

When reviewing the performance of previous models used to predict audit opinions ([8] (82.22%); [40] (62%); [41] (78%); [42] (78.83%); [43] (73%); [45] (84%); [46] (72.20%); [49] (72.90%); [51] (75.60%)), most of these models have achieved classification accuracy below 85%. In contrast, the models used in this study provided overall classification performances ranging from 85.36% to 96.68% showing that they surpassed the 85% threshold. This high classification accuracy further enhances the practical utility of the models used in this study for predicting audit opinion types.

The models developed in this study can be used as a decision support tool for auditors, as expressed by [54], when assessing potential clients and evaluating potential audit risks. For instance, the model can be applied to a potential client’s financial statements to evaluate the likelihood of receiving a non-unqualified audit opinion. A potential audit opinion can be considered part of the client assessment process. The model can also be used to identify current clients likely to receive a non-unqualified

audit opinion during the planning phase of the auditor's work. It can also be utilized in the examination of company audit reports by institutions such as BIST, the Capital Markets Board (SPK), and the Public Oversight Accounting and Auditing Standards Authority (KGK). Companies that receive an audit opinion different from what the model predicts could be targeted for fair reporting investigations. In other words, the models can be used as a complementary "supervisory tool" by regulatory authorities in conducting "audits of audits".

This study is unique because it represents a comprehensive data mining classification investigation to predict independent audit opinion types. By analyzing 12 different data mining methods using WEKA software and comparing their predictive performance, this study is expected to contribute to the literature on independent auditing. Additionally, it is anticipated that this study will provide a more precise definition of data mining, which will assist in solving current business problems and development. This study is expected to shed light on the future study in this area.

Future study could focus on alternative data mining or deep learning methods to predict audit opinion type. Hybrid models can be developed by integrating multiple models into a single framework. Second, the financial and non-financial variables can be expanded, and the results can be compared with those of this study. Finally, the results were evaluated using different data mining software (such as Python, R, Orange, RapidMiner, and Knime).

Acknowledgments
This study developed in the context of the first author's Doctoral Thesis.

Authors' Contributions
Zafer KARDEŞ: %75, Tuğrul KANDEMİR: %25.

Competing Interests
The authors declare that they have no conflict of interest.

References

[1] "Independent auditing standard 315." Accessed: Dec. 18, 2024. [Online]. Available: https://www.kgk.gov.tr/Portalv2Uploads/files/Duyurular/v2/BDS/bdsyeni25.12.2017/BDS%20315-Site.pdf
[2] H. Valipour, F. Salehi, and M. Bahrami, "Predicting audit reports using meta-heuristic algorithms," *J. Distrib. Sci.*, vol. 11, no. 6, pp. 13–19, 2013, doi: 10.13106/jds.2013.vol11.no6.13.
[3] A. K. Nawaiseh, M. F. Abbod, and T. Itagaki, "Financial statement audit using support vector machines, artificial neural networks and k-nearest neighbor: empirical study of UK and Ireland," *Int. J. Simul. Syst. Sci. Technol.*, Mar. 2020, doi: 10.5013/IJSSST.a.21.02.07.
[4] F. A. Amani and A. M. Fadlalla, "Data mining applications in accounting: A review of the literature and organizing framework," *Int. J. Account. Inf. Syst.*, vol. 24, pp. 32–58, 2017, doi: https://doi.org/10.1016/j.accinf.2016.12.004.
[5] O. Pourheydari, H. Nezamabadi-pour, and Z. Aazami, "Identifying qualified audit opinions by artificial neural networks," *Afr. J. Bus. Manag.*, vol. 6, no. 44, pp. 11077–11087, Nov. 2012, doi: 10.5897/AJBM12.855.
[6] S. M. Saif, M. Sarikhani, and F. Ebrahimi, "Finding rules for audit opinions prediction through data mining methods," *SSRN Electron. J.*, 2012, doi: 10.2139/ssrn.2185919.
[7] M. A. Fernández-Gámez, F. García-Lagos, and J. R. Sánchez-Serrano, "Integrating corporate governance and financial variables for the identification of qualified audit opinions with neural networks," *Neural Comput. Appl.*, vol. 27, no. 5, pp. 1427–1444, Jul. 2016, doi: 10.1007/s00521-015-1944-6.
[8] E. Kirkos, C. Spathis, A. Nanopoulos, and Y. Manolopoulos, "Identifying qualified auditors' opinions: a data mining approach," *J. Emerg. Technol. Account.*, vol. 4, no. 1, pp. 183–197, Jan. 2007, doi: 10.2308/jeta.2007.4.1.183.
[9] N. Dopuch, R. W. Holthausen, and R. W. Leftwich, "Predicting audit qualifications with financial and market variables," *Account. Rev.*, pp. 431–454, 1987.
[10] K. Keasey, R. Watson, and P. Wynarczyk, "The small company audit qualification: a preliminary investigation," *Account. Bus. Res.*, vol. 18, no. 72, pp. 323–334, Sep. 1988, doi: 10.1080/00014788.1988.9729379.
[11] D. Zdolšek, T. Jagrič, and M. Odar, "Identification of auditor's report qualifications: an empirical analysis for Slovenia," *Econ. Res.-Ekon. Istraživanja*, vol. 28, no. 1, pp. 994–1005, Jan. 2015, doi: 10.1080/1331677X.2015.1101960.
[12] E. Kirkos, C. Spathis, and Y. Manolopoulos, "Data mining techniques for the detection of fraudulent financial statements," *Expert Syst. Appl.*, vol. 32, no. 4, pp. 995–1003, 2007, doi: 10.1016/j.eswa.2006.02.016.
[13] P. Ravisankar, V. Ravi, G. Raghava Rao, and I. Bose, "Detection of financial statement fraud and feature selection using data mining techniques," *Decis. Support Syst.*, vol. 50, no. 2, pp. 491–500, Jan. 2011, doi: 10.1016/j.dss.2010.11.006.
[14] G. L. Gray and R. S. Debreceny, "A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits," *Int. J. Account. Inf. Syst.*, vol. 15, no. 4, pp. 357–380, Dec. 2014, doi: 10.1016/j.accinf.2014.05.006.

[15] M. Cecchini, H. Aytug, G. J. Koehler, and P. Pathak, "Detecting management fraud in public companies," *Manag. Sci.*, vol. 56, no. 7, pp. 1146–1160, Jul. 2010, doi: 10.1287/mnsc.1100.1174.
[16] K. Takahashi, Y. Kurokawa, and K. Watase, "Corporate bankruptcy prediction in Japan," *J. Bank. Finance*, vol. 8, no. 2, pp. 229–247, Jun. 1984, doi: 10.1016/0378-4266(84)90005-0.
[17] W. Kwak, Y. Shi, and C. F. Lee, "Data mining applications in accounting and finance context," in *Handbook of Financial Econometrics, Mathematics, Statistics, and Machine Learning*, World Scientific, 2020, pp. 823–857. doi: 10.1142/9789811202391_0021.
[18] I. Dutta, S. Dutta, and B. Raahemi, "Detecting financial restatements using data mining techniques," *Expert Syst. Appl.*, vol. 90, pp. 374–393, Dec. 2017, doi: 10.1016/j.eswa.2017.08.030.
[19] E. Kirkos, C. Spathis, and Y. Manolopoulos, "Applying data mining methodologies for auditor selection," in *Proceedings 11th Pan-Hellenic conference in informatics (PCI)*, 2007, pp. 165–178. [Online]. Available: https://datalab-old.csd.auth.gr/wp-content/uploads/publications/PCI07ksm.pdf
[20] T. G. Calderon and J. J. Cheh, "A roadmap for future neural networks research in auditing and risk assessment," *Int. J. Account. Inf. Syst.*, vol. 3, no. 4, pp. 203–236, Dec. 2002, doi: 10.1016/S1467-0895(02)00068-4.
[21] E. Koskivaara, "Artificial neural networks in analytical review procedures," *Manag. Audit. J.*, vol. 19, no. 2, pp. 191–223, Feb. 2004, doi: 10.1108/02686900410517821.
[22] E. Koskivaara and B. Back, "Artificial neural network assistant (anna) for continuous auditing and monitoring of financial data," *J. Emerg. Technol. Account.*, vol. 4, no. 1, pp. 29–45, Jan. 2007, doi: 10.2308/jeta.2007.4.1.29.
[23] A. Yaşar, E. Yakut, and M. M. Gutnu, "Predicting qualified audit opinions using financial ratios: Evidence from the Istanbul Stock Exchange," *Int. J. Bus. Soc. Sci.*, vol. 6, no. 8, pp. 57–67, 2015.
[24] A. Yaşar, "Olumlu görüş dışındaki denetim görüşlerinin veri madenciliği yöntemleriyle tahminine ilişkin karar ve birliktelik kuralları," *Financ. Anal. Cozum Derg.*, vol. 26, no. 133, 2016, Accessed: Jan. 07, 2025. [Online]. Available: https://search.ebscohost.com
[25] T. Büyüktanir and T. Toraman, "Auditor's opinions prediction with machine learning algorithms," in *2020 28th Signal Processing and Communications Applications Conference (SIU)*, IEEE, 2020, pp. 1–4. Accessed: Dec. 18, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9302441/
[26] A. Özcan, "Determining factors affecting audit opinion: evidence from Turkey," *Int. J. Account. Financ. Report.*, vol. 6, no. 2, Art. no. 2, Aug. 2016, doi: 10.5296/ijafr.v6i2.9775.
[27] B. Adiloglu and B. Vuran, "Identification of key performance ındicators of auditor's reports: evidence from Borsa Istanbul (BIST)," *Press. Procedia*, vol. 3, no. 1, pp. 854–859, 2017, doi: https://doi.org/10.17261/Pressacademia.2017.665.
[28] M. Mat and S. Onal, "Bağımsız denetim raporlarında denetim görüşünü etkileyen faktörlerin belirlenmesi: Borsa İstanbul'da bir uygulama," *Muhasebe Bilim Dünya. Derg.*, vol. 21, no. 3, pp. 733–760, 2019, doi: 10.31460/mbdd.533127.
[29] İ. Sakin, "Bağımsız denetim görüşünü etkileyen faktörlerin tespiti: Borsa İstanbula kayıtlı işletmeler üzerine bir uygulama," Master's Thesis, Sosyal Bilimler Enstitüsü, 2019. Accessed: Jan. 06, 2025. [Online]. Available: https://acikbilim.yok.gov.tr/handle/20.500.12812/124427
[30] H. A. Ata and I. H. Seyrek, "The use of data mining techniques in detecting fraudulent financial statements: an application on manufacturing firms," *Suleyman Demirel Univ. J. Fac. Econ. Adm. Sci.*, vol. 14, no. 2, 2009, Accessed: Dec. 18, 2024. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=27ff227919838c7c0991b061751f0e246557788d
[31] S. Terzi, "Hile ve usulsüzlüklerin tespitinde veri madenciliğinin kullanımı," *Muhasebe Ve Finans. Derg.*, no. 54, Art. no. 54, Apr. 2012.
[32] S. Terzi and İ. Şen, "Finansal tablo hilelerinin veri madenciliği yardımıyla tespit edilmesi: üretim sektöründe bir araştırma," *J. Account. Tax. Stud.*, vol. 5, no. 2, Art. no. 2, Jul. 2012.
[33] M. C. Sorkun and T. Toraman, "Fraud detection on financial statements using data mining techniques," *Int. J. Intell. Syst. Appl. Eng.*, vol. 3, no. 5, pp. 132–134, Sep. 2017, doi: 10.18201/ijisae.2017531428.
[34] K. Kirda and M. Katkat Özcelik, "Finansal tablo hilesi riski taşıyan şirketlerin veri madenciliği ile belirlenmesi," *Muhasebe Ve Vergi Uygulamaları Derg.*, vol. 14, no. 2, pp. 609–639, Jul. 2021, doi: 10.29067/muvu.802703.
[35] B. Tatar and H. Kiymik, "Finansal tablolarda hile riskinin tespit edilmesinde veri madenciliği yöntemlerinin kullanılmasına yönelik bir araştırma," *J. Yaşar Univ.*, vol. 16, no. 64, pp. 1700–1719, Oct. 2021, doi: 10.19168/jyasar.956623.
[36] I. Kilic and S. Onal, "Finansal hilelerin tespit edilmesinde kullanılan veri madenciliği yöntemleri ve Borsa Istanbul'da bir uygulama," *Muhasebe Ve Denetime Bakış*, vol. 22, no. 67, pp. 181–208, Sep. 2022, doi: 10.55322/mdbakis.1068503.
[37] G. Ozdagoglu, A. Ozdagoglu, Y. Gumus, and G. Kurt Gumus, "The application of data mining techniques in manipulated financial statement classification: the case of Turkey," *J. AI Data Min.*, no. Online First, 2017, doi: 10.22044/jadm.2016.664.

[38] I. F. Ceyhan, "Bağımsız denetim kalitesini artırıcı bir yöntem olarak veri madenciliği: Borsa Istanbul uygulaması," PhD Thesis, Sakarya Universitesi (Turkey), 2014. Accessed: Dec. 18, 2024. [Online]. Available: https://search.proquest.com/openview/692e0761f5f5ea36a5a8559ac464da35/1?pq-origsite=gscholar&cbl=2026366&diss=y&casa_token=zh2KnS5lw3sAAAAA:PdmSn3AqhdFbuV9omDXFsQ-lyOuy1_dnp9Q0HpkNiebBNVEMyW_TafJoFqO4F1QbkgcFl3Zozw

[39] N. Cömert and M. Kaymaz, "Araç sigortası hilelerinde veri madenciliğinin kullanımı," *Marmara Üniversitesi İktisadi Ve İdari Bilim. Derg.*, vol. 41, no. 2, pp. 364–390, Jan. 2020, doi: 10.14780/muiibd.665058.

[40] E. K. Laitinen and T. Laitinen, "Qualified audit reports in Finland: evidence from large companies," *Eur. Account. Rev.*, vol. 7, no. 4, pp. 639–653, Dec. 1998, doi: 10.1080/096381898336231.

[41] C. T. Spathis, "Audit qualification, firm litigation, and financial ınformation: an empirical analysis in greece," *Int. J. Audit.*, vol. 7, no. 1, pp. 71–85, Mar. 2003, doi: 10.1111/1099-1123.00006.

[42] C. Spathis, M. Doumpos, and C. Zopounidis, "Using client performance measures to identify pre-engagement factors associated with qualified audit reports in Greece," *Int. J. Account.*, vol. 38, no. 3, pp. 267–284, Sep. 2003, doi: 10.1016/S0020-7063(03)00047-5.

[43] M. Doumpos, C. Gaganis, and F. Pasiouras, "Explaining qualifications in audit reports using a support vector machine methodology," *Intell. Syst. Account. Finance Manag.*, vol. 13, no. 4, pp. 197–215, Dec. 2005, doi: 10.1002/isaf.268.

[44] C. Caramanis and C. Spathis, "Auditee and audit firm characteristics as determinants of audit qualifications: Evidence from the Athens stock exchange," *Manag. Audit. J.*, vol. 21, no. 9, pp. 905–920, Dec. 2006, doi: 10.1108/02686900610705000.

[45] C. Gaganis, F. Pasiouras, and M. Doumpos, "Probabilistic neural networks for the identification of qualified audit opinions," *Expert Syst. Appl.*, vol. 32, no. 1, pp. 114–124, Jan. 2007, doi: 10.1016/j.eswa.2005.11.003.

[46] C. Gaganis, F. Pasiouras, C. Spathis, and C. Zopounidis, "A comparison of nearest neighbours, discriminant and logit models for auditing decisions," *Intell. Syst. Account. Finance Manag.*, vol. 15, no. 1–2, pp. 23–40, Jan. 2007, doi: 10.1002/isaf.283.

[47] N. Stanišić, T. Radojević, and N. Stanić, "Predicting the type of auditor opinion: Statistics, machine learning, or a combination of the two?," *Eur. J. Appl. Econ.*, vol. 16, no. 2, pp. 1–58, 2019, doi: 10.5937/EJAE16-21832.

[48] A. K. Nawaiseh and M. F. Abbod, "Financial statement audit utilising naive bayes networks, decision trees, linear discriminant analysis and logistic regression," in *The ımportance of new technologies and entrepreneurship in business development: in the context of economic diversity in developing countries*, B. Alareeni, A. Hamdan, and I. Elgedawy, Eds., Cham: Springer International Publishing, 2021, pp. 1305–1320. doi: 10.1007/978-3-030-69221-6_97.

[49] H. Zarei, H. Yazdifar, M. Dahmarde Ghaleno, and R. Azhmaneh, "Predicting auditors' opinions using financial ratios and non-financial metrics: evidence from Iran," *J. Account. Emerg. Econ.*, vol. 10, no. 3, pp. 425–446, Jun. 2020, doi: 10.1108/JAEE-03-2018-0027.

[50] M. I. H.-P. Wu and L. Li, "The BP neural network with adam optimizer for predicting audit opinions of listed companies," 2021. Accessed: Jan. 07, 2025. [Online]. Available: https://www.semanticscholar.org/paper/The-BP-Neural-Network-with-Adam-Optimizer-for-Audit-Wu-Li/a08f165c1ad8b530c0a1a356d9233639e23bb52f

[51] A. Saeedi, "Audit opinion prediction: a comparison of data mining techniques," *J. Emerg. Technol. Account.*, vol. 18, no. 2, pp. 125–147, Sep. 2021, doi: 10.2308/JETA-19-10-02-40.

[52] N. Japkowicz and M. Shah, *Evaluating learning algorithms: a classification perspective*. Cambridge University Press, 2011.

[53] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data.," *Biometrics*, vol. 33, no. 1, p. 159, Mar. 1977, doi: 10.2307/2529310.

[54] G. S. Monroe and S. T. Teh, "Predicting uncertainty audit qualifications in Australia using publicly available information," *Account. Finance*, vol. 33, no. 2, pp. 79–106, Nov. 1993, doi: 10.1111/j.1467-629X.1993.tb00200.x.