Journal of AI and Data Mining
Vol 5, No 1, 2017, 67-77

DOI: 10.22044/jadm.2016.664

The application of data mining techniques in manipulated financial statement classification: The case of turkey

G. Özdağoğlu[1*], A. Özdağoğlu[2], Y. Gümüş[3] and G. Kurt-Gümüş[4]

1. Dept. of Business Administration, Faculty of Business, Dokuz Eylül University, Tınaztepe Campus, Buca, İzmir, Turkey.
2. Dept. of Business Administration, Faculty of Business, Dokuz Eylül University, Tınaztepe Campus, Buca, İzmir, Turkey.
3. Dept. of Tourism Management, Reha Midilli Foça Tourism Faculty, Dokuz Eylul University, Foça, İzmir, Turkey.
4. Dept. of International Business and Trade, Faculty of Business, Dokuz Eylül University, Tınaztepe Campus, Buca, İzmir, Turkey.

Received; Accepted
**Corresponding author: guzin.kavrukkoca@deu.edu.tr (G.Özdağoğlu).*

Abstract
Predicting financially false statements to detect frauds in companies has an increasing trend in recent studies. The manipulations in financial statements can be discovered by auditors when related financial records and indicators are analyzed in depth together with the experience of auditors in order to create knowledge to develop a decision support system to classify firms. Auditors may annotate the firms' statements as "correct" or "incorrect" to add their experience, and then these annotations with related indicators can be used for the learning process to generate a model. Once the model is learned and tested for validation, it can be used for new firms to predict their class values. In this research, we attempted to reveal this benefit in the framework of Turkish firms. In this regard, the study aims at classifying financially correct and false statements of Turkish firms listed on Borsa İstanbul, using their particular financial ratios as indicators of a success or a manipulation. The dataset was selected from a particular period after the crisis (2009 to 2013). Commonly used three classification methods in data mining were employed for the classification: decision tree, logistic regression, and artificial neural network, respectively. According to the results, although all three methods are performed well, the latter had the best performance, and it outperforms other two classical methods. The common ground of the selected methods is that they pointed out the Z-score as the first distinctive indicator for classifying financial statements under consideration.

Keywords: *Classification, Data Mining, Manipulated Financial Statements, Audit Opinion, Borsa İstanbul.*

Introduction

Financial statement fraud is recently an important issue in the related research. Particularly, Enron crisis emerged the financial statement fraud and earnings management concepts. Fraud causes falsifications in elements of financial statements, and gradually financially false statements are generated by manipulating assets, liabilities, revenue, expenses, profit or losses [2]. Period shifting, changing accounting methods, fiddling with managerial estimates of costs, manipulating documents, changing test documents and preparing false work reports are the other common techniques for fraud and manipulation of profits [3,4]. In this regard, American Institute of Certified Public Accountants (AICPA) Auditing Standards group the red flags, i.e. undesired and risky indicators in financial activities, into three categories: management characteristics, industry conditions, and operating characteristics and financial stability [5]. Recently, frauds over falsified financial statements (manipulated financial statements) got firms into a scrape during auditing processes. This kind of institutional behavior and the undesired financial activities, i.e. the red flags, increased the importance of financial audit process and directly affected the auditor's opinion at the end of the process. Audit opinions are also categorized with respect to the criteria defined in red flags. According to the International Standards on Auditing [1] there are four types of an audit opinion: unmodified opinion, adverse opinion, disclaimer of opinion, and qualified opinion according to the completeness and the correctness of the financial records. All types of the opinions,

except unmodified, carry some suspicious situations and risks from the auditor's perspective.
In auditing, a number of articles have been published in order to provide decision support to detect suspicious situations through financial indicators such as ratios. Data mining techniques which have made a significant contribution to the field of decision science, are also used to develop this kind of decision support tools with the help of information technologies. Hence, the manipulations originated from red flags can be discovered by auditors when related financial statements and indicators are analyzed in depth together with their audit experience. This knowledge can be used to develop a decision support system in order to classify financial statements as true or false through financial indicators. Once auditors annotate the firms' statements as "true" or "false", this knowledge with the other indicators can be used for the supervised learning process of classification techniques in data mining. After training and testing for validation, the learned model can be beneficially used as a decision support tool or system to predict the status of new statements.
In this research, we attempted to reveal this benefit in the framework of Turkish firms. In this regard, the study aims at classifying financially correct and false statements of Turkish firms listed on Borsa İstanbul, using their particular financial ratios as indicators of a success or a manipulation. An annotated term dataset was selected from a particular period after the crisis (2009 to 2013). Commonly used three classification methods in data mining were employed for the classification: decision tree, logistic regression, and artificial neural network. According to the results, although all of three methods performed well, the latter had the highest performance and it outperforms other two classical methods. The common ground of the selected methods is that they pointed out the Z-score as the first distinctive indicator for classifying financial statements under consideration.
The rest of the paper is organized as follows. First of all, a theoretical framework is discussed, and then the methodology is explained. Thirdly, the analysis results are given, and the final section concludes.

1. Related work

Data mining methods are frequently implemented for financial forecasting to identify market trends [6]. Dhar [7] discussed the real value of data mining, particularly in the field of finance, and a framework was constructed to decide when the results of a data mining effort, the patterns are usable for decision support.
Koyuncugil and Ozgulbas [8] proposed an early warning system model based on data mining for financial risk detection through Turkish central bank database. Financial data from balance sheets were used to calculate financial indicators. Zhou et al. [9] investigated the performance of different financial distress prediction models with feature selection approaches based on domain knowledge or data mining techniques. Zhang et al. [10] used the information fusion technique to build a finance early-warning model based on data mining methods. In the paper, the respective strengths of different data mining methods were integrated to improve the prediction accuracy rate. The model based on Support Vector Machines, and Logistic and Dempster-Shafer theory was used for firm's financial risk prediction. Liu [11] constructed a data mining process with discriminant analysis, logistic regression and neural network to predict financial distress.
Predicting financially false statements to detect frauds in companies has an increasing trend in recent studies. Some studies discuss the theoretical framework and indicators or metrics of fraud where others focus on obtaining significant prediction models. Rezaee [12] explained causes and consequences of financial statement and discussed fraud prevention and detection strategies theoretically. Phua et al. [13] analyzed the fraud detection studies for defining the adversary, the types and subtypes of fraud, the technical nature of data, performance metrics, and the methods and techniques. Omar et al. [14] proposed M-Score, Beneish Model, and Z-Score based ratio analysis for detecting fraud in small market cap companies.
Several learning algorithms and other data mining techniques are used to develop classification models; the learned patterns are then used to predict unlabeled testing data sets. Bai et al. [15] used Classification and Regression Tree (CART) statistical technique for identifying and predicting the impacts of false financial statements in China stock market and the findings obtained from CART were also compared with Logit regression. They concluded that CART model achieved better accuracy in identifying fraud cases and making predictions. Pai et al. [16] integrated sequential forward selection (SFS), support vector machines (SVM), and CART for reducing unnecessary information, detecting fraudulent financial statements and providing optimum resource allocation. The features used in the study were

leverage, liquidity, efficiency, corporate governance and probability features which were also divided into several related sub features. Amara et al. [17] analyzed the theoretical foundations of fraud in the financial statements and the impact of the elements of "fraud triangle" on the detection of fraud in the financial statements. Logistic regression was used for the empirical model. The variables of the logistic regression model in this study were ratios of current assets to current liabilities, income before extraordinary items to total assets and the number of outside directors to the total number of directors. Wuerges and Borba [18] used logit and probit models for estimating frauds in US companies. Factor analysis was performed for classifying the companies accused of fraud by the Securities and Exchange Commission (SEC). Kotsiantis et al. [19] designed an artificial neural network (ANN) model) to predict fraudulent financial statements and corporate bankruptcy. Fraud and non-fraud Greek firms in the recent period 2001-2002 were used for training ANNs in the first part. Failed and solvent Greek firms in the recent period 2003-2004 were used for training ANNs in the second part. The main research variable categories in the study are profitability variables, liquidity/leverage variables, efficiency variables, growth variables and the size variable.

Based on these related research, this paper presents such a research aiming at classifying financially true and false statements of Turkish firms with the help of financial ratios which are important for the audit process. Audit opinions are used as the basic criteria to divide financial statements as financially false or correct. Decision tree, logistic regression, and artificial neural network are employed for the detection process.

2. Methodology

2.1. Term dataset

The term dataset covers firms listed on Borsa İstanbul and the period from 2009 to 2013. The dataset for classification was constructed by categorizing the firms into two, as the firms with financially correct statements which have received unmodified opinion (coded as 1) and the firms with financially false statements which have not received unmodified opinion (coded as 0)[1]. For the selected term to collect data, only 110 records could be classified in the risky class coded as 1 (Table 1), then to make a balanced dataset for classification 114 additional records having successful audit reports were added to the dataset to form the entire dataset and conduct the analyses. Financial statements were collected from Finnet (an online platform providing financial data), and related indicators were then calculated.

Table 1. Frequency of audit opinions of firms with financially false statements.

Type of Audit Opinion	Frequency
Qualified Opinion	97
Disclaimer of Opinion	13
Adverse Opinion	0
Total	**110**

2.2. Variables

The below-mentioned ratios are a new mix inspired by the previous studies [2, 20] which emphasized the distinctive characteristics of these ratios to detect a potentially falsified statement. Referring to those studies, 13 variables are chosen as possible indicators of financially false statements, i.e. Net Working Capital/Total Assets, Retained Earnings/Total Assets, Earnings Before Interest and Taxes/Total Assets, Net Sales/Total Assets, Total Debt/Total Equity, Net Profits After Taxes/Net Sales, Receivables/Net Sales, Net Profits After Taxes/Total Assets, Gross Profit/Total Assets, Inventory/Total Assets, Total Debt/Total Assets, Market Value of Equity/Book Value of Debt, Z-Score.

Net working capital/Total assets (R1): Net Working Capital is calculated by taking the difference between current assets and current liabilities. The difference is divided by Total Assets.

Retained earnings/Total assets (R2): Retained Earnings is not shown as one specific amount in Turkish financial statements, contrarily it is shown by two indicators: previous year’s losses if the firm had a loss, previous year’s profits if the firm had profit. The related amount is divided by total asset value of the companies.

Earnings before Interest and taxes/Total assets (R3): Since Turkish financial statements are having some differences with international financial statements; operating profit value is used as an indicator of Earnings before Interest and Taxes. This amount is divided by total asset value.

Net sales/Total assets (R4): Net sales values is used instead of gross sales value to show the exact sales amount of the companies. Net sales ratio is calculated by subtracting the sales returns and

[1] Although according to the definitions of International Standards on Auditing, adverse opinion is the best indicator of financially false statements, none of the firms received adverse opinion during the analysis period. Thus, the firms which received qualified opinion and disclaimer of opinion were included into the dataset.

allowances from gross sales. The calculated amount is divided by total assets.

Total debt/Total equity (R5): It is a capital structure indicator. Total debt value (current liabilities and long-term liabilities) is divided by total equity value.

Net profits after taxes/Net sales (R6): It is a profitability indicator. Net Profit or loss is divided by net sales value.

Receivables/Net sales (R7): It shows sales in credit. Only trade sales are considered since they are directly related to operations of the firm. Both short and long-term trade receivables are taken into consideration. The total is divided by net sales.

Net profits after taxes/Total assets (R8): It is a profitability indicator. Also referred as Return on Assets. Net profit or loss is divided by total assets.

Gross profit/Total assets (R9): This profitability indicator also gives information about asset profitability. Gross profit is divided by total asset value.

Inventory/Total assets (R10): This indicator shows the ability of the firm to convert its inventory into sales. Inventory is divided by total asset value.

Total debt/Total assets (R11): It is another indicator of debt/capital structure and shows how much of the assets are financed by external funds. Total debt amount is divided by total asset value.

Market value of equity/Book value of total debt (R12): It shows how much the firm's assets can decline in value before the liabilities exceed the assets and the firm faces with insolvency.

Z-Score: It is an insolvency indicator used firstly by [21].

Z=1.2*(Net Working Capital/Total Assets) +1.4*(Retained Earnings/Total Assets) + 3.3*(Earnings Before Interest and Taxes/Total Assets+0.06*(Market Value of Equity/Book Value of Total Debt) +1.0*(Net Sales/Total Assets).

2.3. Classification algorithms used in analysis

Classification techniques are used to predict a particular output based on defined set of input variables or attributes. The attribute that is supposed to be predicted is defined as a class label or attribute. Data set covers the class labels and other attributes that are assumed to have an impact on the selected class label.

There exist many classification algorithms adopting supervised learning that have been developed for different data types and purposes. In general, a classification algorithm processes a training set including a set of attributes and the corresponding output, i.e., the class attribute, generally called prediction attribute. The selected algorithm tries to find out relationships between the attributes that would provide predictions for the outcome through the training set. In the next phase, the learned relationships are applied a test set that includes the same attributes, except for the class attribute, and therefore, predictions are generated [22]. Finally, those predictions are compared with the real class values to analyze the performance of the algorithm through a test set. Performance levels depend on both the structure of algorithms and their parameter values. Performances are measured on the confusion matrix by many formulations, e.g., accuracy, precision, recall, f-measure, Kappa statistics, and so forth [23].

From the financial perspective, classification techniques can classify a firm as low, medium, or high risky category based on a set of attributes related to financial statements and movements [20]. In this study, similarly, particular classification algorithms are used for predicting financially false statements through financial ratios. Financial statements of the firms in the data set are classified as true or false and then this attribute is labeled as a class attribute for the analysis phase. Among the common classification algorithms, artificial neural network, logistic regression, and decision tree are selected for classifying the firm's financial statement.

Decision tree algorithms work based on a divide-and-conquer approach to classifying a target attribute by seeking an attribute at each stage to split on that best separates the classes; then recursively processing the following branches that result from the split. This algorithm generates a decision tree, which can be converted into a set of classification rules. Decision tree classification method is an effective method due to its simplicity in understanding and interpreting. ID3 and C 4.5 or 5.0 are the most common algorithms of decision tree induction [24].

Logistic regression builds a linear model based on a transformed output variable, i.e. target class or class attribute. Logistic regression is used for binary classification problems where the target class consists of only two values such as yes/no, true/false, 0/1. It is also possible to obtain multiple classes by executing this model for each class. Suppose that there are only two classes. Logistic regression replaces the original target variable which cannot be approximated accurately using a linear function, with a log transformation of odd ratio. The resulting values are no longer constrained to the interval from 0 to 1 but can lie

anywhere between negative infinity and positive infinity. These values are converted into the desired interval using logit transformation. The final model is obtained by using maximization of log-likelihood via a standard optimization approach. Once the parameters have been learned, then these parameters are used to predict the class values in testing dataset [25].
Artificial Neural Networks (ANNs) are computing technology whose fundamental purpose is to recognize patterns in data. Based on a computing model, ANNs use the simulated brain's ability to learn or adapt in response to external inputs. When exposed to a stream of training data, neural networks can discover previously, unknown relationships and learn complex nonlinear mappings in the data. ANNs behave like a black box, in other words, it is difficult to understand the inside of the network. Each arc has a weight, and these weights are summed up in each layer considering a particular policy for propagation. These networks are iteratively improved until they reach a state where the error term is minimum [26]. Parameters are assigned to the inputs of the neurons; the output of the neuron is a nonlinear combination of the inputs, weighted by the parameters similar to the synaptic weights of biological neurons. Calculations are often performed based on the weighted sum of inputs and parameters plus a constant called bias. Execution of a neural network is initialized by activation, and this can only happen when the other neurons are activated through the edges that are connected to it. The neurons within a neural network are usually arranged in layers. The number of layers, number of neurons in each layer, learning rate, momentum values are important parameters for the design of such networks. Neural networks and related data mining algorithms are widely used for financial applications. Predicting fraudulent credit transactions, interest rates, and exchange rate fluctuations in currency markets, bankruptcy; managing portfolios; assessing risks [27].

2.4. The application platform

The classification model development and implementation were performed using RapidMiner Studio 6.4, i.e. an open-source data mining tool, based on the data and variable set. The selected parameters of each classifier were optimized by embedding the model into "optimize parameters" operator. This operator is a collapsed process in which a set of parameters in a wide range can be tested iteratively to discover the best ones. Figure 1 explains the general flow of the modeling process.
Decision tree method embedded into RapidMiner is similar to C 4.5. In each node, the split variable is selected by iterating all variables for finding the best split for each variable with respect to the splitting criterion. Finally, the method uses the variable that maximizes the criterion and continues until all branches end up with a class decision. For nominal variables, one branch for each value is created whereas for numerical attributes a binary split is performed to achieve the best split value by trying all possible values in the training set. Pre-pruning conditions can be used and considered during the splitting period, and then optionally post-pruning can also be added to improve the structure of the tree.
In this research, the decision tree was executed at different pruning levels, and the tree that has given the best performance was selected. Then linear and quadratic logistic regression models were run comparatively, and the linear model was found to have higher performance. As the last classification algorithm, the artificial neural network model was executed based on the optimized values for its two important parameters, i.e., learning rate and momentum. Thus, the predictions having the highest performance were tried to obtain based on the given input set. Finally, these results were compared to indicate the method that gave the better results. For all of the learning algorithms, training, and testing datasets were determined based on 10-fold cross validation.

3. Analysis and results

The current status of the firm categories in the data set was compared regarding their ratios to see the preliminary differences between them. Table 2 illustrates the mean values of the variables both for firms with financially false and correct statements separately. It is remarkable that all profit related variables of the statements that could not have an unmodified opinion (coded as 0) are negative except gross profit/total asset ratio, conversely all have positive values for the statements having an unmodified opinion (coded as 1).
The other striking point is debt-related ratios: Net working capital/total assets, total debt/total equity and market value of equity/book value of debt.

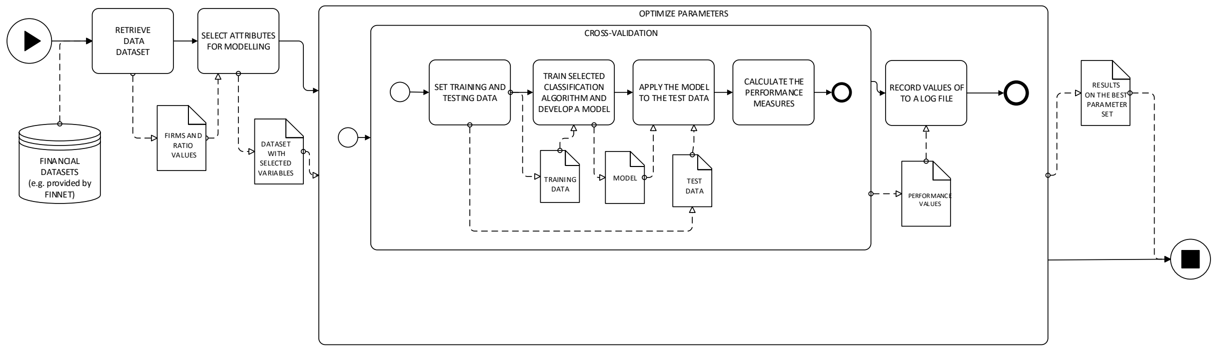


Figure 1. General flow of modelling process.

Table 2. Mean values of the variables.

Ratio Definition		Class-0 (mean)	Class-1 (mean)
Net Working Capital/Total Assets	R1	-0,34418	0,200973
Retained Earnings/Total Assets	R2	-1,19817	0,037748
Earnings Before Interest and Taxes/Total Assets	R3	-0,02042	0,056646
Net Sales/Total Assets	R4	0,603963	0,953397
Total Debt/Total Equity	R5	3,191075	1,835819
Net Profits After Taxes/Net Sales	R6	-0,25488	0,162205
Receivables/Net Sales	R7	0,496203	0,249811
Net Profits After Taxes/Total Assets	R8	0,006841	0,032273
Gross Profit/Total Assets	R9	0,098115	0,183753
Inventory/Total Assets	R10	0,093032	0,14989
Total Debt/Total Assets	R11	0,922455	0,472218
Market Value of Equity/Book Value of Debt	R12	1,823368	5,322425
Z-Score		-1,44449	1,753689

The first one is negative for the class-0 by contrast with the class-1, showing current liabilities exceed current assets; according to the second one both classes are predominantly financed with debt instead of equity, but the class-0 relies on debt more; because of the second debt related ratio and most probably high stock prices of the companies, the class-1 has extremely high market value of equity/book value of debt ratio. Finally, the Z-score is negative for the class-0 and on the other hand positive for the class-1.

In the light of these distinctive indicators, the selected firms were classified with the help of above-mentioned classification techniques in data mining and a general classification modeling scheme was developed for this purpose. The classification process started with a training set to learn the pattern in the dataset in the first partition within a validation operator, therefore, the learned pattern was applied to the testing set and the selected performance indicators, i.e. accuracy, f-measure, area under ROC curve, were calculated using the related performance indicators (Figure 1). Training and testing steps were embedded in 10-fold cross-validation, and for the training part, the classifier related to the classification method was placed where each training group was selected with respect to stratified sampling.

In this regard, the decision tree algorithm was firstly executed at different pruning levels, and the tree that has given the best performance was selected. The overall accuracy rate of this classifier was calculated as 82.5% with a tolerance value where the f-measure is 78.22%. The AUC gave values that are close to “1”. These performance measures (see Table 3) indicate a good prediction model. Especially the accuracy values close to one may imply an overestimation where low values, generally less than fifty percent, imply underestimation. The differences in performances of the classes were also measured by using recall and precision measures. These measures in class-1 were greater than the same values in class 0, indicating that the decision tree model could predict true class-1 better than class-0.

Appendix 1 presents the decision tree indicating the important attributes that have higher impact on the false statements. According to this tree, Z-score is the most important attribute to distinct true and false statements. In the second level, from the top edge of the tree, R11 (Total Debt/Total Assets), then R2 (Retained Earnings/ Total Assets), R8 (Net Profits After Taxes/ Total Assets), R1 (Net Working Capital/ Total Assets) and R7 (Receivables/Net Sales) are the important factors on the decomposition at the successive levels of the tree as well, respectively. Some conditional statements can be generated by following the branches given in the decision tree[2].

For the same dataset, linear and quadratic logistic regression models were executed and reported comparatively. Parameter optimization process was performed based on kernel-type and kernel-gamma parameters, i.e., important parameters of logistic regression.

---

[2] For example, “if z-score < 0.368 and R11 < 0.198 and R2 > -1.261, then the statement is true” (Appendix 1).

Table 3. Results based on the selected performance measures.

Classification Model		true Class-1	true Class-0	class precision	Accuracy	AUC	f_measure
Decision Tree	pred. Class-1	102	30	77.27%	82.25% +/- 9.78%	0.984	78.22% +/- 12.03%
	pred. Class-0	8	74	90.24%			
	class recall	92.73%	71.15%				
Logistic Regression	pred. Class-1	102	30	77.27%	82.34% +/- 7.45%	0.870 +/- 0.066	79.41% +/- 8.90%
	pred. Class-0	8	74	90.24%			
	class recall	92.73%	69.23%				
Artificial Neural Network with one hidden layer	pred. Class-1	99	25	79.84%	83.29% +/- 7.89%	0.874 +/- 0.089	80.96% +/- 10.34%
	pred. Class-0	11	79	87.78%			
	class recall	90.00%	75.96%				

Finally, the linear model was found to have higher performance. As implemented in the model based on the decision tree, training, and testing steps were embedded in the cross-validation operator in which the logistic regression operator was placed as a classifier in the training partition (Figure 1). Similar results were obtained from the linear logistic regression model as indicated in Table 3Error! Reference source not found.. The accuracy of the model was obtained as 82.34% with some tolerance and f-measure was 79.41% which is also close the results obtained from decision tree, however, AUC on logistic regression is worse than AUC calculated on the decision tree. These values could be observed a bit above the values of the decision tree model. Logistic regression could detect two more false statements. However, logistic regression model underestimated false statements (class-0) if compared with the performance of the true (class-1) statements. Appendix 3Error! Reference source not found. gives the weights of the variables in the logistic regression model to predict future values. As seen from this list, z-score has the highest effect on the classification; R2, R1, and R8 are the other attributes following the Z-score.

As the last classifier, ANNs model was executed based on the values for its three important parameters (learning rate, the number of learning cycles, and momentum). The infrastructure of the model was developed in a similar manner to the models presented in the previous sections, only the classifier operator and related parameters change (Figure 1). The related ANN algorithm was executed on the normalized dataset via the optimized number of training cycles, learning rate, and momentum with sigmoid function and one hidden layer to use for predictions. This classifier with one layer ended up with a better performance than multiple hidden layers. Thus, only the findings of the one-layer model were presented. The best performance was obtained via learning rate=0.14; momentum=0.22 with 170 training cycles. Table 3Error! Reference source not found. shows the performance level and parameter set of improved neural network with the same metrics. It can be inferred from the table that both class recall and precision values were improved and balanced. The ANNs model which predicted "false" statements through the selected attributes or variables resulted in more than 80% accuracy which means that a correct prediction with a relatively higher probability is possible by adopting this classifier. The details about the neural network model, network scheme, and numerical values on neurons and edges are given in Appendix 2 for prediction**.** Each column of this table corresponds to weight list of the arcs incoming from the input layer to the hidden layer or outgoing from the hidden layer to the output layer. Weighted sums plus bias value activates the further phases until the network achieves a class value.

Gray and Deprecency [20] investigated the use of data mining techniques to classify firms with respect to their statements and proposed a taxonomy to detect manipulations. Data mining targets in the audit environment were tabulated according to data classes, target datasets, signaling, data types, semantic representations, and score. Data mining applications were classified according to account scheme and evidence scheme combinations. At the last phase of that study, fraud and evidence schemes and application of data mining were integrated into a relationship matrix. In the light of Gray and Deprecency [20], our study performed particular classification methods in data mining to the term dataset gathered from Borsa Istanbul by evaluating their suggestion and also considering additional ratios used in the previous studies.

Spatis [2] applied only logistic regression on the relatively smaller dataset and analyzed firms with and without Z-score to show the effect of this indicator. The study also compared the impact of the ratios with the previous studies. Our analyses produced the results compatible with this study providing that z-score is very distinctive to detect falsified financial record. Furthermore, our study applied additional classification methods with additional ratios in order to present the comparative performances of different methods. Bai et al. [15] compared CART and Logit Regression methods and found that CART gave better results for fraud detection. Pai et al. [16] compared over data mining techniques on eighteen financial ratios and obtained different model accuracies lying between 73% and 92%. Our findings and the findings of the previous studies revealed that the performance of the model is very dependent on the data set, the selected algorithm, and the variables, i.e. selected financial ratios.

4. Conclusion and policy implications

Fraud on falsified financial statements got firms into a scrape during auditing processes. This potential negative situation has a critical impact on the auditing process and increases the stress on auditors because of the responsibility to reveal the situation over the financial records and statements. Although the related auditing standards try to make the process easier by defining undesired manipulations (red flags) that can be performed by firms, additional decision support tools are necessary to detect these negative situations efficiently and immediately. There exist many studies as mentioned in "the related work" section proposing various approaches and case studies which benefitted from data mining techniques in order to detect false statements over the particular financial indicators such as ratios through the selected classification techniques.

In this context, this paper presented research that aims at classifying financially true and false statements with the help of financial ratios which are important indicators for the auditors through particular classification techniques in data mining. The analysis was conducted on the group of 214 records gathered from 136 different Turkish firms listed on Borsa İstanbul for the period between 2009 and 2013. The dataset for classification was constructed by categorizing the firms into two, as the firms with financially correct statements which have received unmodified opinion (coded as 1) and the firms with financially false statements which have not received unmodified opinion (coded as 0). For the selected term to collect data, only 110 records could be classified in the risky class coded as 1, then to make a balanced dataset for classification 114 additional records having successful audit reports were added to the dataset in order to form the entire dataset. Although the adverse opinion is the best indicator of financially false statements, none of the firms received adverse opinion during the analysis period. Thus, the firms which received qualified opinion and disclaimer of opinion were included in the dataset.

This study originally combined distinctive financial ratios that were handled in the previous studies and analyzed as a whole through the selected classification methods. Twelve financial ratios and one combination of financial ratios were calculated from the firms' balance sheets and income statements to construct the dataset for the classification process: Net Working Capital/Total Assets, Retained Earnings/Total Assets, Earnings before Interest and Taxes/Total Assets, Net Sales/Total Assets, Total Debt/Total Equity, Net Profits after Taxes/Net Sales, Receivables/Net Sales, Net Profits after Taxes/Total Assets, Gross Profit/Total Assets, Inventory/Total Assets, Total Debt/Total Assets, Market Value of Equity/Book Value of Total Debt, and Z-Score.

For the purpose of indicating the difference between the record groups in terms of the selected ratios, class means were calculated and based on these preliminary results, a valuable finding was obtained over debt ratios, i.e. net working-capital/total assets, total debt/total equity and market value of equity/book value of debt. The first one was negative for the class-0 by contrast with the class-1, showing current liabilities exceeded current assets; according to the second one both classes were predominantly financed with debt instead of equity, but the class-0 relied on debt more. Because of the second debt related ratio and most probably high stock prices of the companies, the class-1 had an extremely high market value of equity/book value of debt ratio. In addition to the findings on the debt ratios, another important point was that the Z-score was negative for the class-0 and positive for the class-1. These results were the preliminary signs of the distinctive characteristics of the selected ratios used for classification.

The classification model development and implementation were performed by using RapidMiner 6.4. The parameters of each classifier were optimized through its "optimize parameters" operator to obtain the best prediction model on the classifier for the given dataset. The classification process started with a training set to learn the pattern in the dataset; therefore, the learned pattern was applied to the testing set and the selected performance indicators. Training and testing stages were carried out over 10-fold cross validation where ten training data were selected in each iteration with respect to stratified sampling and trained through the selected classifier and then tested on the selected test data. In this regard, decision tree, logistic regression, and artificial neural networks were employed for the classification process. The results indicated that the artificial neural network had the highest prediction accuracy. Furthermore, Z-score which is a combination of different financial ratios is a better indicator to classify financially false or correct statements than the individual ratios. Thus, auditors might start using this kind of ratio combinations while they are auditing the financial statements of the companies.

Consecutively, this paper presented a study that developed a classification modeling framework for categorizing the selected Turkish firms' financial statements as "correct" or "incorrect" regarding their financial indicators in order to use the framework for predicting the status of new firms or upcoming statements of the existing firms. The optimized values of the parameters and the performance of the selected classifiers may change depending on the data set and the selected financial indicators. The main idea is developing such a decision support system for auditors to provide some clues for a potential manipulation before inspecting the other accounting data in detail. Various classifiers can easily be added to the modeling framework to extend the scope of classification process, and even non-experts can use it as a reference by only changing the dataset.

References

[1] IAASB. (2015). Handbook of International Quality Control, Auditing, Review, Other Assurance, and Related Services Pronouncements, The International Auditing and Assurance Standards Board (IAASB).

[2] Spathis, C. T. (2002). Detecting false financial statements using published data: some evidence from Greece. *Managerial Auditing Journal, 17*(4), 179-191.

[3] Worthy, F. S. (1984). Manipulating profits: how is it done? Fortune, 25, 50-54.

[4] Comer, J. M. (1998). Corporate Fraud (3 ed.). Aldershot: Gower.

[5] AICPA. (2016, 01 15). Statements on Auditing Standards. Retrieved from AICPA: www.aicpa.org/Research/Standards/AuditAttest/Pages/SAS.aspx

[6] Kovalerchuck, B., & Vityaev, E. (2000). Data Mining in Finance: Advances in Relational and Hybrid Methods. Norwell, Massachusetts: Kluwer Academic Publishers, ISBN 0-7923- 7804-0.

[7] Dhar, V. (1998). Data Mining In Finance: Using Counterfactuals To Generate Knowledge From Organizational Information Systems. Information Systems, 23(7), 423-437.

[8] Koyuncugil, A. S., & Ozgulbas, N. (2012). Financial Early Warning System Model and data mining application for risk detection. Expert Systems with Applications, 39, 6238–6253.

[9] Zhou, L., Lu, D., & Fujita, H. (2015). The performance of corporate financial distress prediction models with feature selection guided by domain knowledge and data mining approaches. Knowledge-Based Systems, 85, 52–61

[10] Zhang, L., Zhang, L., Teng, W., & Chen, Y. (2013). Based on information fusion technique with data mining in the application of finance early-warning. Procedia Computer Science, 17, 695 – 703. doi:10.1016/j.procs.2013.05.090

[11] Liu, M. (2004). A data mining method for forecasting financial distress of Chinese listed companies. Proceedings of The 2004 International Conference on Management Science & Engineering (pp. 370-374). , Harbin, Peoples R China, Vol: 1- 2.

[12] Rezaee, Z. (2005). Causes, consequences, and deterrence of financial statement fraud. Critical Perspectives on Accounting, 16(3), 277-298.

[13] Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. arXiv preprint arXiv:1009.6119.

[14] Omar, N., Sanusi, Z. M., Johari, A., & Mohamed, I. S. (2014). Predicting financial stress and earning management using ratio analysis. Predicting financial stress and earning management using ratio analysis, 8(8), Predicting financial stress and earning management using ratio analysis.

[15] Bai, B., Yen, J., & Yang, X. (2008). False financial statements: characteristics of China's listed companies and CART detecting approach. International Journal of Information Technology & Decision making, 7(2), 339-359.

[16] Pai, P. F., Hsu, M. F., & Wang, M. C. (2011). A support vector machine-based model for detecting top management fraud. Knowledge-Based System, 24(2), 314-321.

[17] Amara, I., Amar, A. B., & Jarboui, A. (2013). Detection of fraud in financial statements: french companies as a case study. International Journal of Academic Research in Accounting, Finance and Management Sciences, 3(3), 40-51.

[18] Wuerges, A. F., & Borba, J. A. (2011). Accounting Fraud: an estimation of detection probability.

[19] Kotsiantis, S., Koumanakos, E., Tzelepis, D., & Tampakas, V. (2006). Financial application of neural networks: two case studies in Greece. Artificial Neural Networks–ICANN 2006 (pp. 672-681). Springer Berlin Heidelberg.

[20] Gray, G. L., & Debreceny, R. S. (2014). A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits.

International Journal of Accounting Information Systems, 15(4), 357-380.

[21] Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance, 23(7), 423-437.

[22] Witten, I. H., & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

[23] Han, J., & Kamber, M. (2006). Data Mining: Concepts and Techniques. Data Mining: Concepts and Techniques.

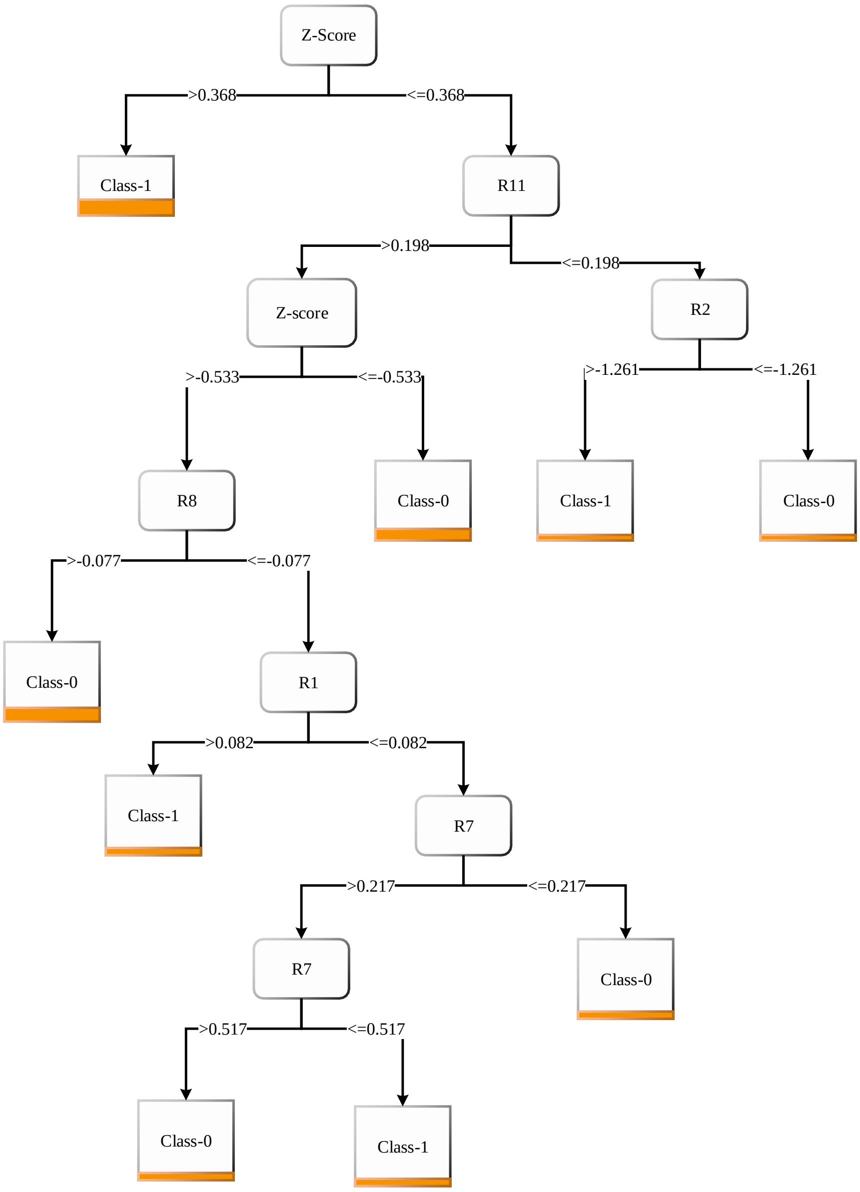
[24] Siciliano, R., & Conversano, C. (2006). Decision tree induction. In J. Wang, Encyclopedia of Data Warehousing and Mining (p. 353). Idea Group, USA.

[25] Panik, M. (2009). Regression Modeling Methods, Theory, and Computation with SAS. CRC Press.

[26] Dreyfus, G. (2005). Neural Networks: An Overview. In Neural Networks (pp. 1-83). Springer Berlin Heidelberg.

[27] Bigus, J. P. (1996). Data Mining with Neural Networks: Solving Business Problems from Application Development to Decision Support. McGraw-Hill, Inc.

Appendix 1: Decision Tree Representation



Appendix 2: Weights on the Arcs of the ANN

HIDDEN LAYER				
Node 1 (Sigmoid)	Node 2 (Sigmoid)	Node 3 (Sigmoid)	Node 4 (Sigmoid)	Node 5 (Sigmoid)
R1: -0.902	R1: 1.665	R1: 0.666	R1: 0.369	R1: -1.333
R2: -1.431	R2: 3.214	R2: 1.253	R2: 0.792	R2: -5.081
R3: 0.545	R3: -0.052	R3: 0.082	R3: 0.062	R3: -3.527
R4: 0.064	R4: 0.273	R4: 0.375	R4: 0.342	R4: -1.590
R5: -0.139	R5: 0.694	R5: 0.353	R5: 0.243	R5: -0.357
R6: -0.279	R6: 0.932	R6: 0.591	R6: 0.446	R6: -1.214
R7: -0.321	R7: 0.761	R7: 0.571	R7: 0.429	R7: 0.894
R8: 0.053	R8: 0.282	R8: 0.299	R8: 0.248	R8: 0.244
R9: 0.066	R9: 0.480	R9: 0.239	R9: 0.176	R9: -2.091
R10: 0.273	R10: 0.182	R10: -0.444	R10: -0.110	R10: -2.599
R11: 0.626	R11: -1.031	R11: -0.379	R11: -0.240	R11: 0.724
R12: -0.599	R12: 1.825	R12: 1.044	R12: 0.739	R12: -2.182
Z Score: -1.326	Z Score: 3.185	Z Score: 1.420	Z Score: 0.955	Z Score: -4.727
Bias: 0.472	Bias: -1.506	Bias: -0.784	Bias: -0.581	Bias: 1.884

HIDDEN LAYER				OUTPUT LAYER	
Node 6 (Sigmoid)	Node 7 (Sigmoid)	Node 8 (Sigmoid)	Node 9 (Sigmoid)	Class '1.0' (Sigmoid)	Class '0.0' (Sigmoid)
R1: -1.534	R1: -0.077	R1: 1.601	R1: -0.988	Node 1: -1.287	Node 1: 1.245
R2: -2.797	R2: 0.029	R2: 3.066	R2: -1.614	Node 2: 1.870	Node 2: -1.867
R3: 0.438	R3: 0.088	R3: -0.026	R3: 0.408	Node 3: 0.828	Node 3: -0.793
R4: 0.026	R4: 0.335	R4: 0.271	R4: -0.016	Node 4: 0.529	Node 4: -0.498
R5: -0.614	R5: 0.196	R5: 0.731	R5: -0.167	Node 5: -4.113	Node 5: 4.100
R6: -0.681	R6: 0.218	R6: 0.936	R6: -0.367	Node 6: -2.082	Node 6: 2.129
R7: -0.536	R7: 0.313	R7: 0.757	R7: -0.226	Node 7: -0.027	Node 7: 0.020
R8: -0.150	R8: 0.217	R8: 0.297	R8: -0.048	Node 8: 1.772	Node 8: -1.787
R9: -0.191	R9: 0.184	R9: 0.469	R9: -0.036	Node 9: -1.322	Node 9: 1.379
R10: -0.013	R10: 0.176	R10: 0.171	R10: 0.161	Threshold: -0.606	Threshold: 0.583
R11: 0.934	R11: 0.149	R11: -0.999	R11: 0.620		
R12: -1.344	R12: 0.402	R12: 1.688	R12: -0.717		
Z Score: -2.641	Z Score: 0.229	Z Score: 3.089	Z Score: -1.500		

Appendix 3: Coefficients of the Linear Logistic Regression Model

Bias (offset): 0.779			
Weight	**Value**	**Weight**	**Value**
w[R1]	-1.163	w[R7]	0.163
w[R2]	-1.522	w[R8]	0.846
w[R3]	-0.446	w[R9]	-0.160
w[R4]	-0.138	w[R10]	0.265
w[R5]	-0.012	w[R11]	-0.029
w[R6]	-0.234	w[R12]	-1.576