

10

INDEPENDIENT EDUCATING OPTIMUM METHOD OF INDEPENDENT AUDITOR REPORT TYPE PREDICTION

EDUCANDO EL MÉTODO ÓPTIMO DE LA PREDICCIÓN DEL TIPO DE INFORME DE AUDITOR

Ali Bakhshi¹

E-mail: ali.bakhshi75@yahoo.com

ORCID: <https://orcid.org/0000-0002-9781-2800>

Shohre Yazdani²

E-mail: yazdani_82@yahoo.com

ORCID: <https://orcid.org/0000-0003-4193-0397>

Ali Maleki³

E-mail: maleki@iaufb.ac.ir

ORCID: <https://orcid.org/0000-0002-9026-0273>

¹ Ph.D. Student, Department of Accounting. Islamic Azad University, Damavand. Iran.

² Assistant Professor, Department of Accounting. Islamic Azad University. Damavand. Iran.

³ Assistant Professor, Department of Statistics. Azad University. Firoozkuh. Iran.

Cita sugerida (APA, séptima edición)

Bakhsh, A., Yazdani, S., & Maleki, A. (2020). Independent educating optimum method of independent auditor report type prediction. *Revista Conrado*, 16(74), 79-92.

ABSTRACT

Accountability requires the existence of reliable and valid information and auditing is one of fundamental bases for the accountability process. Thus educating the optimum method is of great importance. With the presence of a great volume of information, using the prediction methods can contribute the auditors in this respect. This research aims to compare a variety of methods of teaching of that. In current research, J48, random forest, vector machine and neural network have been used. Research population involves the corporates accepted by Tehran Stock Exchange in 2008-2017. Here, 19 financial and non-financial independent variables have been applied in two groups of test and training. Also, the independent auditor report has been classified into two groups of acceptable and conditional. Comparing the above-mentioned methods has indicated that random forest algorithm with the prediction accuracy average as 78.83% was the most optimum model to predict the report type in the both groups and other models involving J48, support vector machine, CART decision tree and finally, artificial neural network were of the most accuracy.

Keywords: Auditor report type, J48 algorithm, random forest, support vector machine, CART decision tree, artificial neural network.

RESUMEN

La rendición de cuentas requiere la existencia de información confiable y válida y la auditoría es una de las bases fundamentales para el proceso de rendición de cuentas. Por lo tanto, educar el método óptimo es de gran importancia. Con la presencia de un gran volumen de información, el uso de los métodos de predicción puede contribuir a los auditores a este respecto. Esta investigación tiene como objetivo comparar una variedad de métodos de enseñanza de eso. En la investigación actual, se han utilizado J48, bosque aleatorio, máquina de vectores y red neuronal. La población de investigación involucra a las empresas aceptadas por la Bolsa de Teherán en 2008-2017. Aquí, se han aplicado 19 variables independientes financieras y no financieras en dos grupos de prueba y capacitación. Además, el informe del auditor independiente se ha clasificado en dos grupos de aceptable y condicional. La comparación de los métodos mencionados anteriormente ha indicado que el algoritmo de bosque aleatorio con un promedio de precisión de predicción de 78.83% fue el modelo más óptimo para predecir el tipo de informe en ambos grupos y otros modelos que involucran J48, máquina de vectores de soporte, árbol de decisión CART y finalmente, La red neuronal artificial fue de la mayor precisión.

Palabras clave: Tipo de informe de auditor, algoritmo J48, bosque aleatorio, máquina de vectores de soporte, árbol de decisión CART, red neuronal artificial.

INTRODUCTION

Considerable increase and variety of economic transactions necessity the financial and economic information and it can be stated that with more data, everybody is more likely to be successful. Information users are investors, creditors, government and others who apply the information to make reasonable decisions. On the other hand, the separation of management and ownership has caused that the goals of information providers vary with those of users or they may be contrary; here, there may be the risk of immoral management and the recruitment of independent auditors in order to validate the financial statements. Feloropous (2010) believed that managers tend to report unreal financial status leading to asymmetric data between the corporate and information users; auditors play a key role in assuring the financial statements and reducing the representative costs. Boshen & Smith (2009), stated that the audited statements are the information of corporates which are regulated for the public because the users of unaudited ones are affected by cheating and scandals of corporate. Researches demonstrated that the auditing reports content affected the stock price of corporates in Tehran Stock Exchange and capital markets reacted to the release of conditional audit reports and the elimination of condition terms (Alikhani Dehghi, 2006). Considering the information content of audit reports, up-to-date and efficient auditing is required in order to enhance the quality of financial reports so that the advances in theoretical audit concepts, wide application of IT in business and the appearance of novel technologies and knowledge have created new challenges in audit methods (Efsthathios, et al., 2007). Also, given the information content of audit reports, it can be concluded that the prediction of comment type is of information content which influences the decisions of users so that the results can be used in evaluating the potential employers, reviewing the tradesmen, investigating the quality control, predicting the auditor comments in similar circumstances as defense against the lawsuits (Gaganis, et al., 2007). Rapid technological changes and application of various sciences have led to encourage the auditors to use them in increasing the effectiveness and efficiency. One of them is data mining which is a general term involving a series of methods to extract human intelligence from data. It assumes that hypotheses are extracted from data automatically. Data mining exploits a hidden rule or knowledge in data in order to develop various models of data analysis (Efsthathios, et al., 2007; Tambunan, 2019; Abolfathi & Taeabi, 2019).

In another definition, data mining is a process with a novel attitude towards the information extraction from bulky data and precisely recognizes the patterns and relations

in data using a set of statistical methods and modelling as soon as possible. It aims to seek valuable information in a database and predict the future trends and behavior of financial markets (Bagherpor Valashani, et al., 2012). Data mining techniques can be divided into two classes of direct (top-down approach) and indirect (bottom-up approach). Using direct data mining, specific variables may be identified and it finds relationships between a variable and desired population. Indirect data mining targets no specific variable (dependent variable) but it aims to find relationships between variables in a wealth of information. Another description for these classes is to examine specific hypotheses by the top-down approach and bottom-up approach produces new hypotheses.

Few data mining techniques include decision tree, busy networks, random forest, CART decision tree, neural networks, support vector machine and genetic algorithm. This research seeks to compare J48 algorithm, random forest, artificial neural network, CART decision tree and support vector machine in terms of predicting comment types of independent auditors and develop an optimum model. Main research issue is to determine the best prediction model of auditor comment types; afterwards, research variables, methodology, definitions, population and data mining methods are discussed and finally, the research findings and results are analyzed.

Khajavi, et al. (2018), addressed the suitability of different methods in terms of selecting comment type prediction variables and found out the positive impact and suitability of using the variable selection method on the prediction of comment type and a significant difference between the suitability degrees of desired methods. Abaszade, et al. (2017), investigated the accuracy of heuristic algorithms and liner Logit regression in predicting the auditor comment types and concluded that the changes of auditor comments, audit report type of previous year, investment return, current cash ratio, debt ratio, earnings to price ratio, net profit and losses in the corporate are more likely to have the most impact on the prediction of auditor comment type. Hasas Yegane, Taqavifard & Mohammadpor (2014), addressed the prediction of independent auditor report and compared two approaches of neural network and probabilistic neural network in Iran. Therefore, data related to the corporates in Tehran Stock Exchange were used in 2003-2010. Results indicated that the accuracy of probabilistic neural network is more than the neural one.

Bagherpor Valashani, et al. (2011), investigated the prediction of independent auditor report using data mining approach in Iran and applied two data mining techniques including C.5 decision tree and artificial neural networks. Thus, data related to the corporates in Tehran Stock

Exchange were used in 2003-2009. Results indicated that the accuracy of C.5 decision tree was more than the other one.

Barkhordarian, Hashemi & Hosseini (2011), discussed the prediction of conditional auditor comments using multilayer perceptron neural network and decision tree. In this research, two mentioned models were addressed to predict the comments and findings indicated that multilayer perceptron neural network was able to predict the auditor comments with the validity as 22% and CART could predict the auditor comments with the accuracy as 22%; in other words, the both modes were indicative of high power of the model so that H_1 and H_2 have been confirmed and the beneficiaries were enabled to use these model for predicting the comments. Comparing two models showed no significant difference and H_3 was rejected.

Porheydari & Azami (2010), investigated the prediction of independent auditor report and compared two approaches including perceptron neural network and logistic regression. Results indicated that the accuracy of neural network was more than logistic regression. Logistic regression had a weaker performance in predicting the conditional comment and an unbalanced pattern was seen in predicting the auditor comments. Concerning the effective factors on the audit comment type, lots of researches have been conducted and a variety of factors have been identified. Few elements are discussed in the following section. The mentioned factors specify a general framework of effective factors on the independent auditor comment types.

Gaya, et al. (2017), discussed the effect of audit report quality on the relationship between family ownership and tax avoidance in 11 corporates during 2008-2013. Results indicated that there is a significant direct relationship between family ownership and tax avoidance and also, the audit report quality is reversely related to family ownership and tax avoidance so that family ownership tendency towards tax avoidance is reduced.

Kirkos, Spathis & Yannis (2007), used three techniques of data mining classification to develop models in order to identify the conditional reports and found out that the indices of financial distress and profitability ratio were the most important factors.

Giger et al. (2006) in a research on making the decision and judgments of auditors in various circumstances studied 694 bankrupt corporates in 1991-2001. Results indicated that the auditors gave No Comment when faced with the corporates in financial distress regardless the acceptable risk level.

Carcello, Hermanson & Neal (2003), used artificial neural network to predict the auditor comments and proposed that artificial neural network may predict the audit report type and guide the investors, creditors, and beneficiaries in stock transactions.

Ireland (2003), assessed the released audit reports and the observed features of corporates such as corporate size in Britain and suggested that the corporates with lack of liquidity and high financial risk are more likely to receive conditional reports as compared to the others.

Isatis, (2003), studied the prediction ability of auditor comment type using financial and non-financial information and reported that financial statements are able to predict the conditional comments.

DEVELOPMENT

Research method is quasi-experimental and post-event and population involves all the corporates in Tehran Stock Exchange in 2008-2017; finally, 48 corporates were selected as the research sample due to the following constraints:

- Corporates should not be in financial dealing groups.
- Audit report and financial statements should be available.
- Corporates should be selected from different industries.
- Fiscal year should be ended at the end of year.
- Corporates should be profitable with operational profits.
- Fiscal year should not be changed.
- Transactions should not be stopped more than 3 months.

In this study, 19 financial and non-financial variables have been identified in terms of audit comment type. Considering the constraints in achieving the required data, the variables were selected by experts and questionnaires. Afterwards, using data mining techniques, the most important elements were specified with respect to audit reports.

Research Variables

X1: Auditor's report (1 for acceptable and 0 for unacceptable)

X2: Size of board of directors. Board of directors with more managers cannot be useful for the corporate and is accompanied with lots of costs. It seems that larger board of directors leads to improve the supervisory and

effectiveness but it may affect the quality of relationships in the corporate.

X3: Unemployed members. An unemployed manager is a part-time member in board of directors and has no executive responsibilities. According to Article 1 in Corporate Governance Regulations draft, most members should be unemployed in the stock corporates.

X4: Independence of board of directors. It refers to number of employed and unemployed members in board of directors.

X5: Auditor type (1 for government auditor and 0 for non-government auditor).

X6: Cash in banks. Cash to debts ratio indicates the corporate ability to pay back the short-term debts by cash assets.

X7: Current debt; Debts to total assets ratio.

X8: Current assets; market assets to net earnings ratio.

X9: Net fixed assets.

X10: Sum of total assets.

X11: Profit and loss after tax deduction; profit after tax and interests to assets ratio.

X12: Sum of earnings.

X13: Incurrent debts.

X14: Equity; equity to assets ratio.

X15: Financial costs.

X16: Profit and loss before tax deduction; profit before interests and tax deduction to assets ratio.

X17: Operational profit and loss.

X18: Cash balance.

X19: Corporate size; natural logarithm of corporate assets.

Research Models

Validity and Reliability of research method

Since the presented method should be assessed in terms of validity, data were divided in two sets of training and test data to achieve knowledge through training data and algorithm but the results validity should be examined by new data and algorithm prediction power concerning the data which are not faced so far. Thus, test data as supervisors are given to the algorithm and results can evaluate the model accuracy. Data are divided by data mining software randomly. Number of training data should be more than number of test data.

J48 algorithm

One of classification methods is J48 algorithm which is C4.5 decision tree written by Java. It is one of generalizations of ID3 algorithm which uses gain ratio criterion to choose a specific property. It applies post-pruning technique and accepts numerical data. It can be utilized for incomplete data with few changes. It selects a trait with maximum separation degree among classes and accordingly, it makes the decision tree. Creating primary decision tree by a set of data is the most important part. Finally, the algorithm produces a classification in the form of a tree with two types of nodes. One node is a leaf specifying a class and one node is a decision testing a trait to produce a branch or sub-tree for an output. To make a similar tree, there is a regression to a subset from samples. This trend continues to include samples belonging to the same class. It stops when number of samples is less than a specific limit (Tabatabaee et al., 2014). Weka (version 8) software is utilized to implement J48 algorithm.

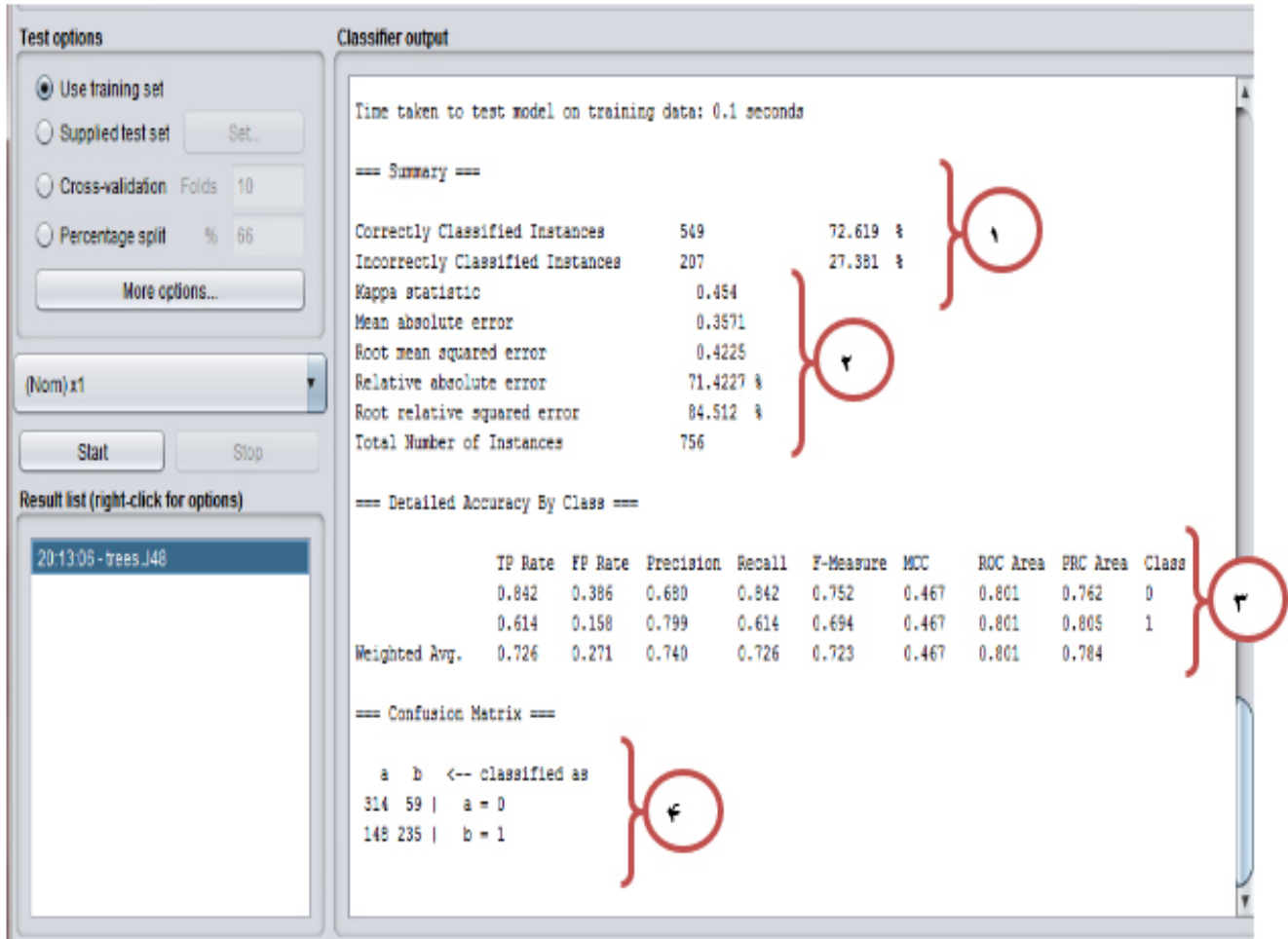


Figure 1. Output of J48 algorithm in training section.

Evaluation variables in training section of J48 algorithm

No.1: Out of 756 samples, 549 samples were correctly classified and 207 samples were wrongly classified. The model accuracy is 72.62% and the error was 27.38% (Figure 1).

No.2: A series of evaluation variables can be used to assess the algorithm and performance when the desired variable is continuous.

TP Rate: It stands for true positive indicating correct classification of data.

FT Rate: It stands for false positive indicating wrong classification of samples.

No.4: It shows how to classify data in different classes.

Class A: 314 and 59 samples were classified in correct and wrong manners, respectively.

Class B: 235 and 148 samples were classified in correct and wrong manners, respectively.

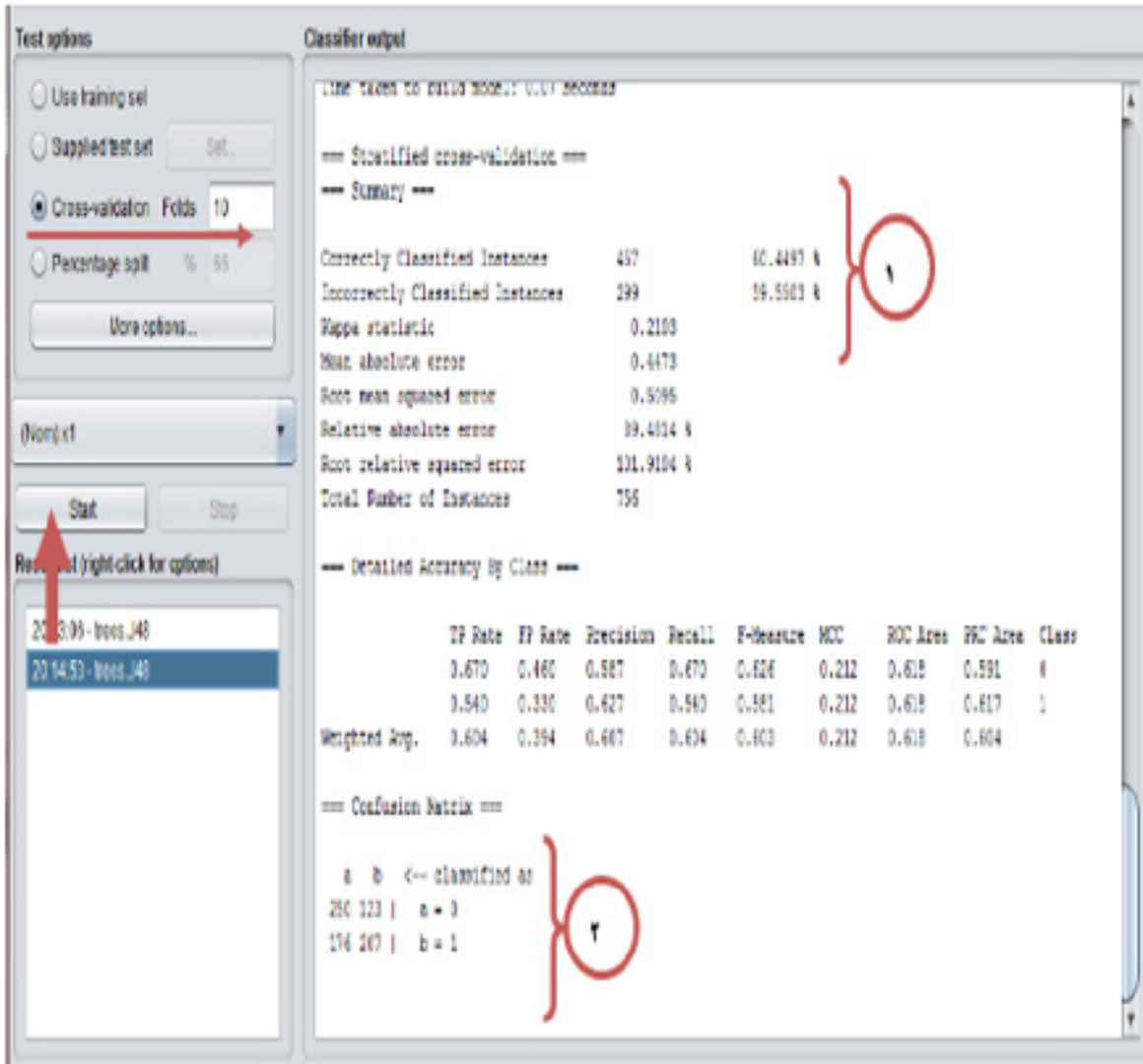


Figure 2. Output of J48 algorithm in test section.

Evaluation variables in test section of J48 algorithm

No.1: Out of 756 samples, 457 samples were correctly classified and 299 samples were wrongly classified. The model accuracy is 60.44% and the error was 39.55% (Figure 2).

No.2:

Class A: 250 and 123 samples were classified in correct and wrong manners, respectively.

Class B: 207 and 176 samples were classified in correct and wrong manners, respectively.

Random forest algorithm

Random forest algorithm is a group algorithm with a set of decision trees. Classification accuracy of random forest was considerable while developing a set of trees and voting among them to achieve a category with the most votes. The model involves several single-tree models.

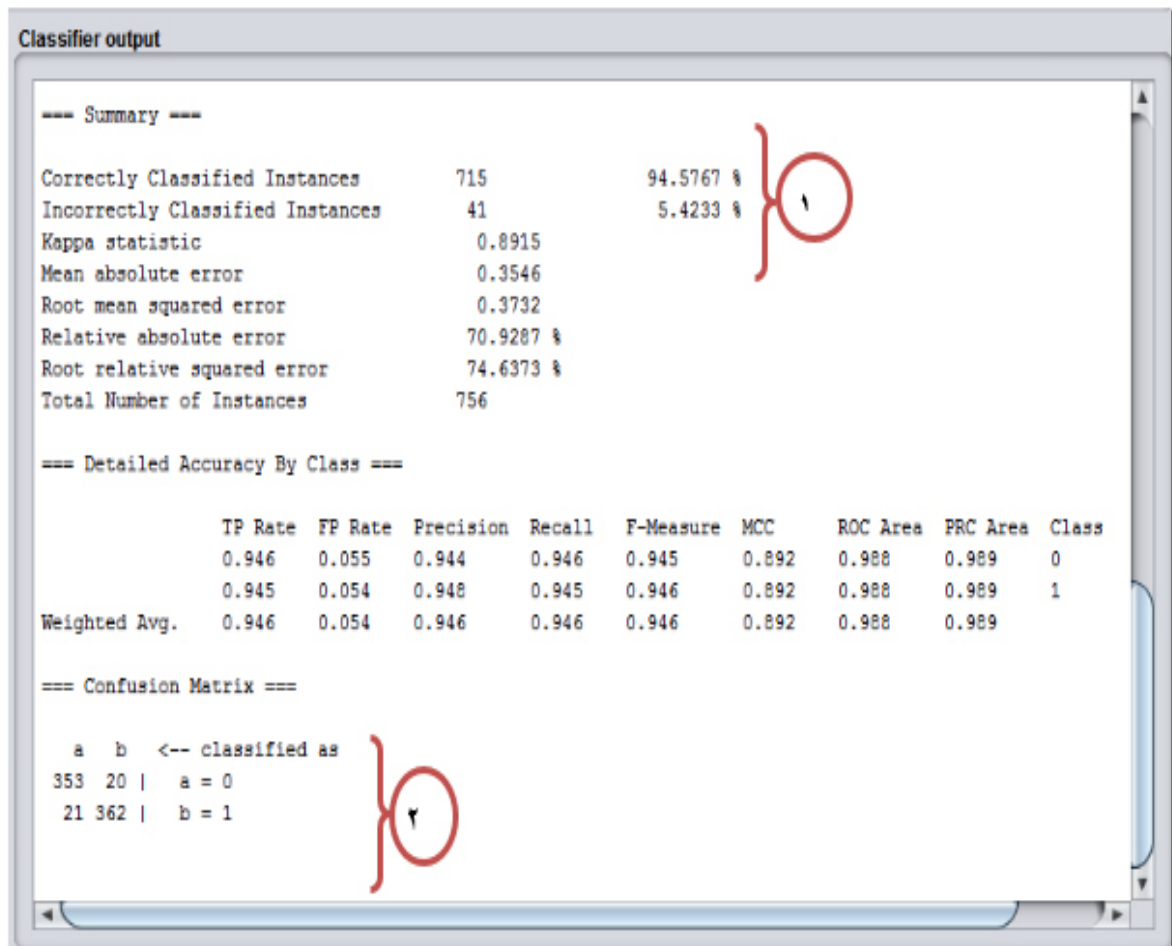


Figure 3. output of random forest in training section.

No.1: Out of 756 samples, 715 samples were correctly classified and 41 samples were wrongly classified. The model accuracy is 94.57% and the error was 5.42% (Figure 3).

No.2:

Class A: 353 and 20 samples were classified in correct and wrong manners, respectively.

Class B: 362 and 21 samples were classified in correct and wrong manners, respectively.

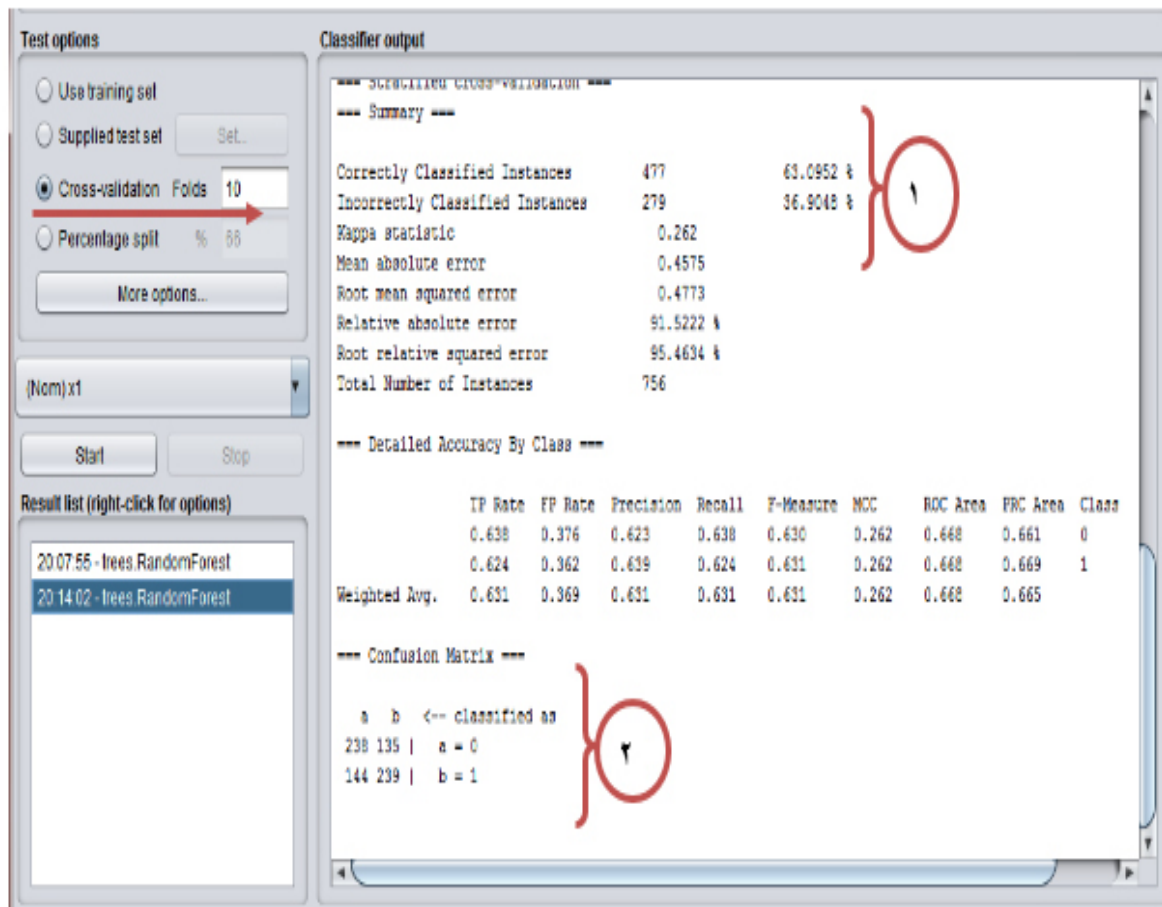


Figure 4. output of random forest in test section.

No.1: Out of 756 samples, 477 samples were correctly classified and 279 samples were wrongly classified. The model accuracy is 63.09% and the error was 36.90% (Figure 4).

No.2: It is different to classify data in different classes.

Class A: 238 and 135 samples were classified in correct and wrong manners, respectively.

Class B: 239 and 144 samples were classified in correct and wrong manners, respectively.

Support Vector Machine Algorithm

To solve the problem of classification, one of the effective and widely used methods is support vector machine algorithm which was first presented by Winik and could decrease the experimental error and avoid over fitting. It seeks to find a hyper plane with maximum margin between two classes and converts the problem to a convex quadratic optimization one; afterwards, it will be solved by quadratic programming technique. Also, support vector machine can classify non-linear samples using kernel functions (Kerabona et al., 2010).

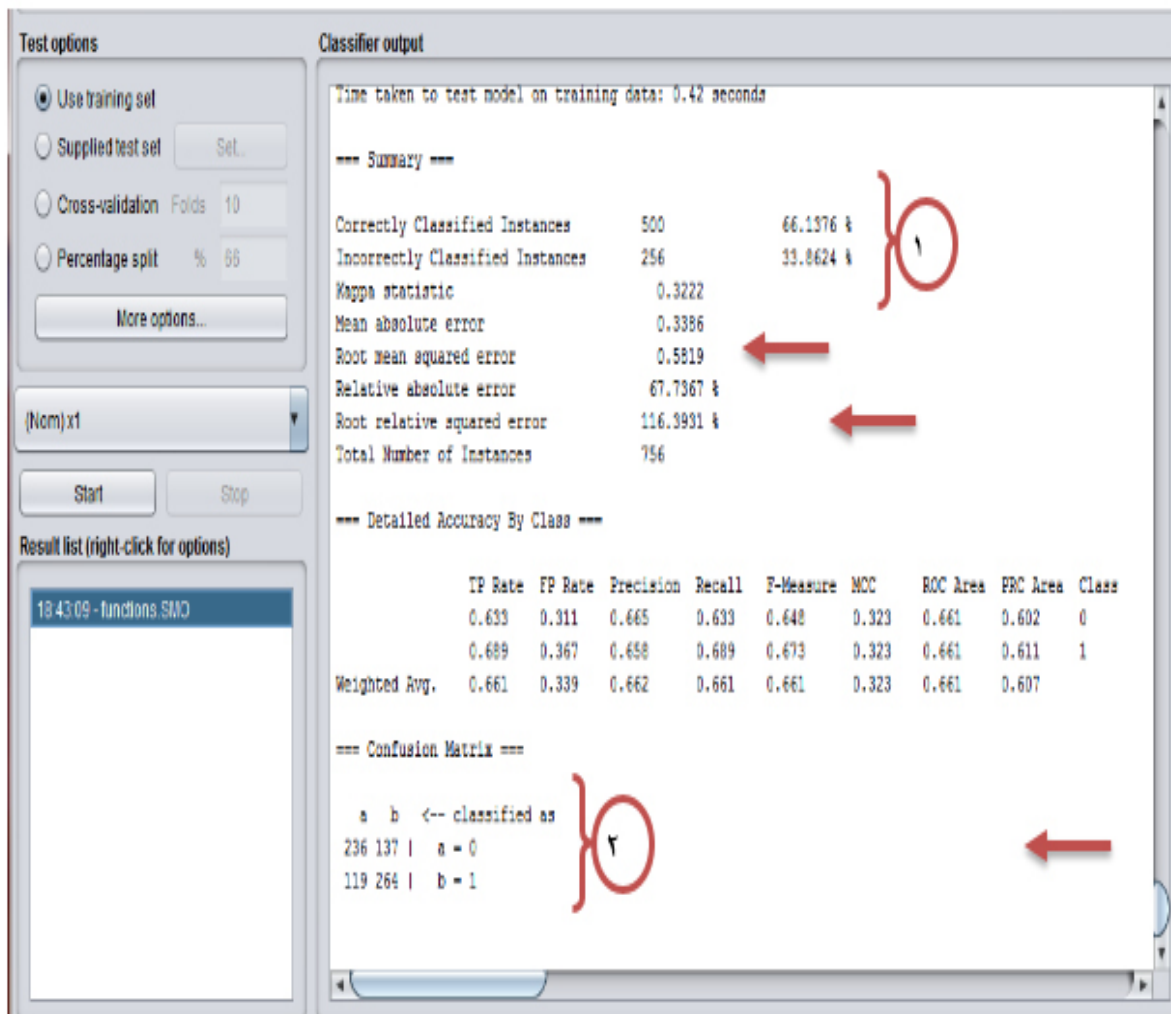


Figure 5. output of support vector machine algorithm in training section.

No.1: Out of 756 samples, 500 samples were correctly classified and 256 samples were wrongly classified. The model accuracy is 66.13% and the error was 33.86% (Figure 5).

No.2: It is different to classify data in different classes.

Class A: 236 and 137 samples were classified in correct and wrong manners, respectively.

Class B: 264 and 119 samples were classified in correct and wrong manners, respectively.

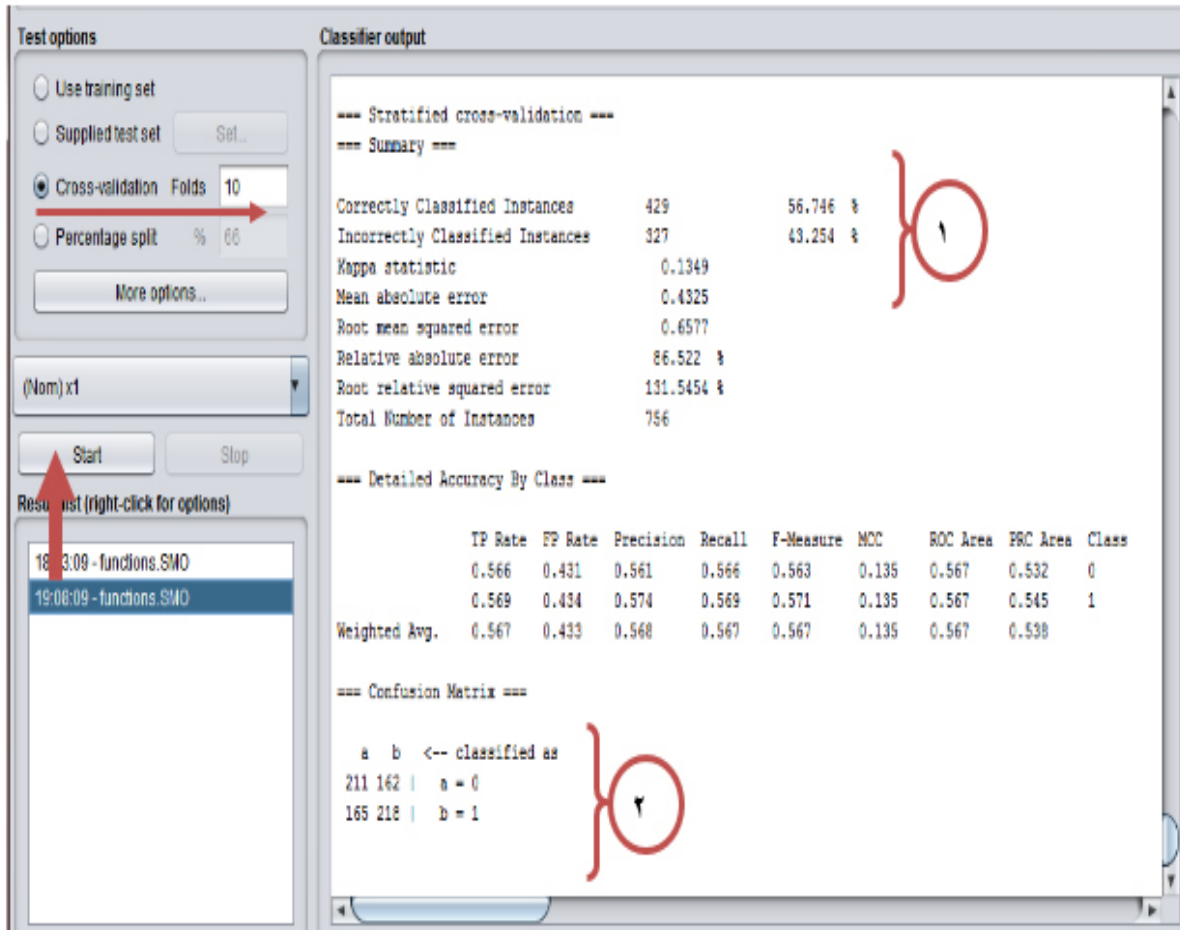


Figure 6. output of support vector machine algorithm in test section.

No.1: Out of 756 samples, 429 samples were correctly classified and 327 samples were wrongly classified. The model accuracy is 56.74% and the error was 43.25% (Figure 6).

No.2: It is different to classify data in different classes.

Class A: 211 and 162 samples were classified in correct and wrong manners, respectively.

Class B: 218 and 165 samples were classified in correct and wrong manners, respectively.

Artificial neural network

Artificial neural network is a data processing system which was taken from human brain and processes data by lots of small CPUs which are connected as a network and behave in parallel in order to solve a problem (Russell & Nurvich, 2008). Figures 7 and 8, present the output of artificial neural network.

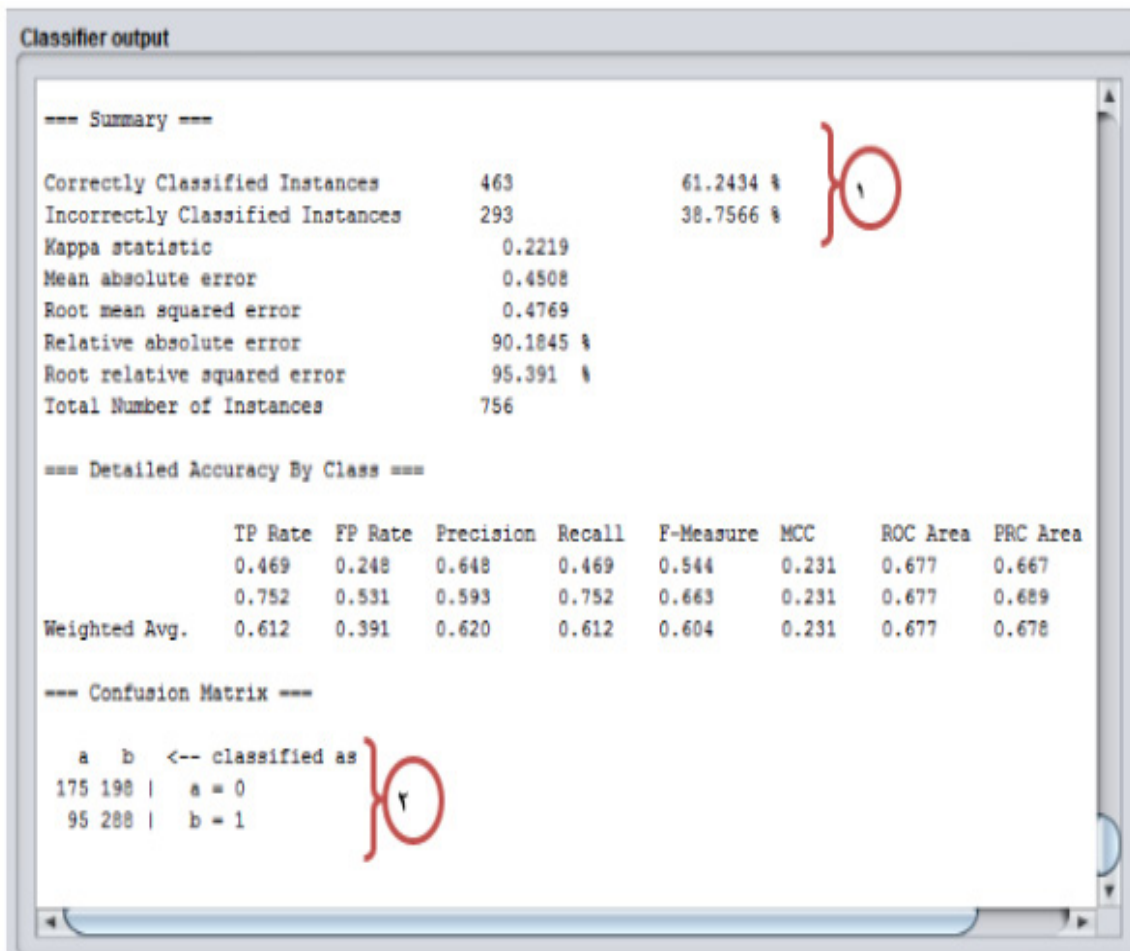


Figure 7. output of artificial neural network in training section.

No.1: Out of 756 samples, 463 samples were correctly classified and 293 samples were wrongly classified. The model accuracy is 61.24% and the error was 38.75% (Figure 7).

No.2: It is different to classify data in different classes.

Class A: 175 and 198 samples were classified in correct and wrong manners, respectively.

Class B: 288 and 95 samples were classified in correct and wrong manners, respectively.

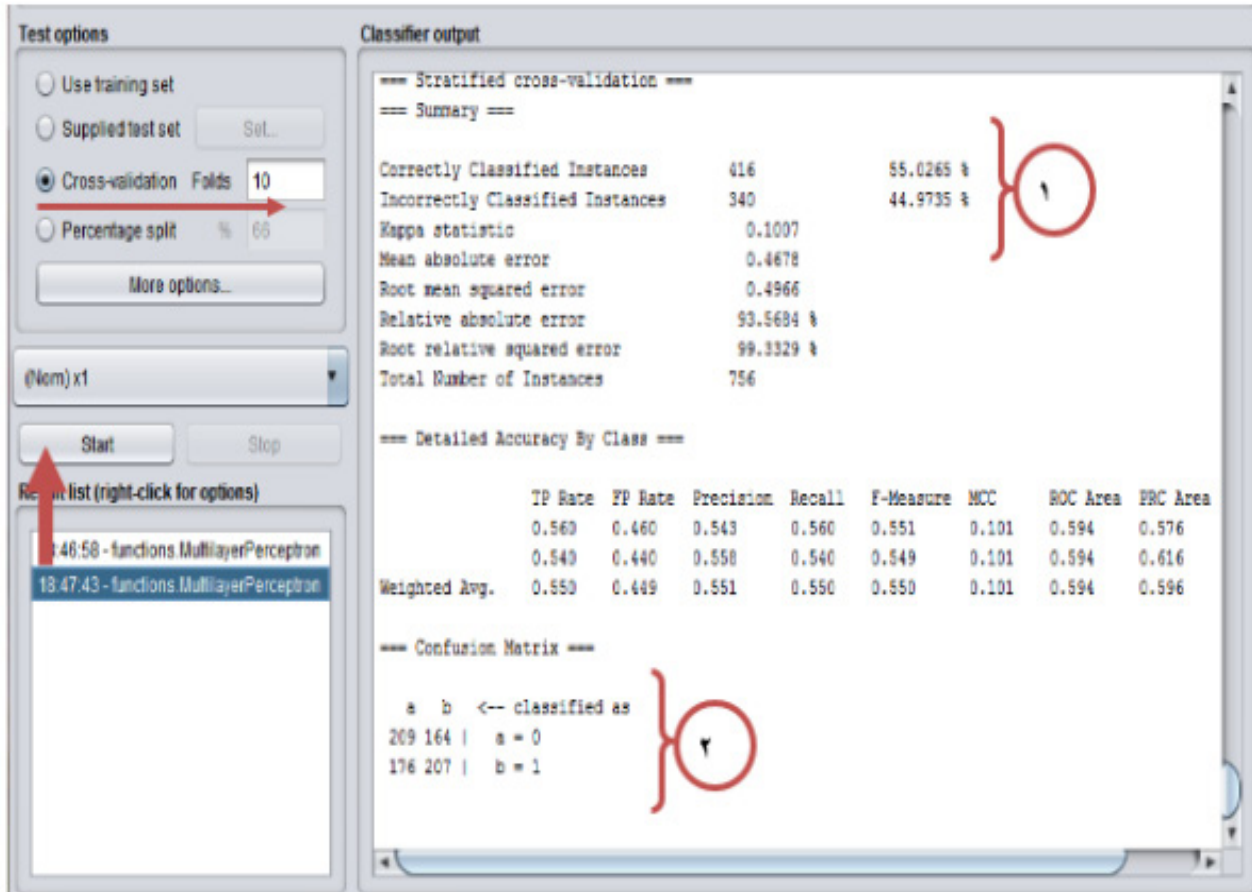


Figure 8. output of artificial neural network in test section.

No.1: Out of 756 samples, 416 samples were correctly classified and 340 samples were wrongly classified. The model accuracy is 55.02% and the error was 44.97% (Figure 8).

No.2: It is different to classify data in different classes.

Class A: 209 and 164 samples were classified in correct and wrong manners, respectively.

Class B: 207 and 176 samples were classified in correct and wrong manners, respectively.

C&R algorithm

It minimizes the impurities in each class. When a node is free from impurities with the elements belonging to one field as a target, all the classifications will be binary; namely, only two subgroups from one node will be divided.

Requirements: To train the C&R tree model, one or more input fields and one output field are required. Target and predictive fields can be intervals or classes.

Strengths: These models act against such problems as missing data very well and usually do not take a long time for training. As well, the understanding of C&R tree models is easier.

Accuracy of CART decision tree in test and training sections

The screenshot shows a software interface with 'Collapse All' and 'Expand All' buttons. Below them, a window titled 'Results for output field x1' contains a sub-window 'Comparing \$R-x1 with x1'. This sub-window displays a confusion matrix table with the following data:

'Partition'	1_Training		2_Testing	
Correct	346	64.92%	128	57.4%
Wrong	187	35.08%	95	42.6%
Total	533		223	

Figure 9. Output of CART tree.

The model accuracy is as follows:

Model accuracy in training section: 64.92% (Figure 9).

Model accuracy in test section: 57.40%

Model error in training section: 35.08%

Model error in test section: 42.60%

CONCLUSIONS

This paper aims to present the best prediction of audit comment type using artificial neural network, support vector machine, random forest algorithm, J48 decision tree, and CART decision tree and compare the performance of mentioned methods. All the used methods involved two fundamental steps. First, using the training sample, the audit comment type is identified and trained in order to achieve knowledge on giving comments about financial reports. Second, the model performance is shown. It should be noted that costs of two error types vary so that putting a corporate with a conditional report in the acceptable class wrongly may cause an incorrect image of corporate whereas putting a corporate with an acceptable report in the conditional class may cause economic problems and lack of investment opportunities; these risks are called Alpha and Beta risks, respectively. Research results indicate that random forest model with the accuracy of 78.835 is the most optimum model to predict the independent auditor report type. Other models accuracy averages are as follows: J48 algorithm 66.31%, support vector machine 61.43%, CART decision tree 61.16%, and artificial neural network 53.13%. Also, the most important independent variables are auditor type, equity, liquidity, financing cost, earnings and loss and profit before tax. This research can be used by independent auditors, internal auditors, investors, creditors, financial analysts and tax authorities.

Future researchers are recommended to investigate the following issues:

- Presenting an optimum model to predict independent auditor report type using Hemming neural network

- Investigating methods of financing and their effect on audit report type in stock corporates
- Investigating a relationship between non-audit services wage and audit report type
- Investigating the reasons of delay in independent auditor report condition articles

Investigating privatization process of governmental corporates and its effect on audit report type.

BIBLIOGRAPHIC REFERENCES

Abaszade, M., Maftonian, M., Babae, M., & Fadaee, M. (2017). Investigating accuracy of heuristic algorithms and linear logit regression in predicting auditor comment type. *Modern Researches in Accounting Journal*, 4, 39-73.

Abolfathi, E., & Taebi, P. (2019). Providing transfer function for the population model. *Journal of Social Sciences and Humanities Research*, 7(01), 64-69.

Alikhani Dehaghi, Ho. (2006). Investigating effective elements in longevity of audit report release. (Master Thesis). Shahid Beheshti University.

Bagherpor Valashani, M. A., Sae, M. J., Meshkani, A., & Bagheri, M. (2012). Predicting independent audit report in Iran: data mining approach. (Paper). 10th National Accounting Conference. Iran.

Barkhordarian, B., Hashemi, S. A., & Hosseini, S. M. (2011). Predicting conditional audit comment using multilayer perceptron neural network and decision tree. (Master Thesis). Isfahan University.

Boshen, J. K., & Smith, J. R. (2009). Does size matter? The influence of large clients on office-level auditor reporting decisions. *Journal of Accounting and Economics*, 30, 375-400.

Carcello, J. V., Hermanson, D. R., & Neal, T. L. (2003). Auditor reporting behavior when GAAS lack specificity: the case of SAS No.59. *Journal of Accounting and Public Policy*, 22, 63-81.

Efstathios, K., Spathis, Ch., Nanopoulos, A., & Manolopoulos, Y. (2007). Identifying Qualified Auditors Opinions: A Data Mining Approach. *Journal of Emerging Technologies in Accounting*, 4, 183-197.

Gaganis, Ch., Pasiouras, F., & Doumpos, M. (2007). Probabilistic Neural Networks for the Identification of Qualified Audit Opinions. *Expert Systems with Applications*, 32, 114-124.

- Hasas Yegane, Y., Taqavifard, M. T., & Mohammadpor, F. (2014). Using probabilistic neural network to identify audit comment type. *Theory and Practice Journal*, 1, 131-159.
- Ireland, J. (2003). An Empirical Investigation of Determinants of Audit Reports in the UK. *Journal of Business Finance and Accounting*, 30(78), 975-1015.
- Khajavi, Sh., Kazemnejad, M., Dehghani Saadi, A. A., & Momtazian, A. (2018). Investigating different methods of selecting predictive variables in predicting auditor comment type. *Accounting Journal*, 7, 81-102.
- Kirkos, E. Spathis, C. N., & Yannis, A. M. (2007). Identifying Qualified Auditors opinion: A Data Mining Approach. *Journal of Emerging technologies in Accounting*, 4, 183-197.
- Porheydari, O., & Azami, Z. (2010). Identifying auditor comment type using neural networks. *Accounting Knowledge Journal*, 3, 77-97.
- Russell, S., & Nurvich, P. (2008). Artificial Intelligence Modern Approach. Naghous Publishing.
- Tambunan, H. (2019). The Effectiveness of the Problem Solving Strategy and the Scientific Approach to Students' Mathematical Capabilities in High Order Thinking Skills. *International Electronic Journal of Mathematics Education*, 14(2), 293-302.