Int. J. Nonlinear Anal. Appl. 14 (2023) 1, 1717–1725
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.26506.3336

Selection of optimal method to predict report type of independent auditor: Comparison of two approaches of support vector machine and neural network

Ali Bakhshi[a], Shohreh Yazdani[a,*], Mohammadhamed Khanmohammadi[a], Ali Maleki[b]

[a] *Department of Accounting, Damavand Branch, Islamic Azad University, Damavand, Iran*

[b] *Department of Statistics, Firoozkooh Branch, Islamic Azad University, Firoozkooh, Iran*

(Communicated by Madjid Eshaghi Gordji)

Abstract

Investors, creditors, government and other users of financial statements rely on financial information given by the managers of firms to make logical and reasonable decisions. In many cases, the purposes of providers are contradictory to the users' ones. Therefore, auditing is a tool to enhance the reliability of financial statements presented by firms. In the current research, the selection of an optimal method to predict the report type of independent auditor has been addressed and two approaches of vector machine and neural network have been compared. It was conducted during 2008-2017. 84 firms were reviewed. To train and test the research variables, Voka software has been implemented. The dependent variable is the report type of auditor. Results indicated that the accuracy of the support vector machine algorithm was computed as 66.13% and 56.74% for the training and testing sections, respectively. As well, the accuracy of the neural network model was 61.24% and 55.02% in the training and testing sections, respectively. The support vector machine model was more effective than the neural network.

Keywords: Auditor report type, Support vector machine, Neural network
2020 MSC: 68T07, 92B20

1 Introduction

The existence of reliable and transparent information is the product of a comprehensive and suitable reporting system and one of the basic components in assessing the status and performance of a specific firm while making decisions on the exchange of securities released by the desired firm. Nowadays, in the professional communities, from the viewpoint of users, the information is reliable when an independent firm supervises the process of reporting and financial statements. Auditory institutions are the example of independent firms, which investigate and supervise the structure of reporting unit as internal control, namely financial statements in business units [1].

*Corresponding author
Email addresses: first.author@email.address (Ali Bakhshi), yazdani_82@yahoo.com (Shohreh Yazdani)

Received: January 2022 *Accepted:* March 2022

2 Research Background

Zdolsek and Jagric [11] developed a model in order to determine the conditional auditing comments by a sample involving 530 English and Irish firms and evaluated the model using a curve of receiver operational traits and wrong classification costs. They concluded that using mental preferences of auditors is sufficient for evaluating the model performance based on performance criteria.

Efstathios et al. [5] used three data analysis methods such as multilayer perceptron, decision tree and Bayesian network to classify the comments of auditors. Research results indicated better performance of Bayesian network as compared to other methods. Gaganis et al. [6] investigated the potential of probable neural network in developing a model for predicting the comments type of auditor through a sample of 881 English firms during 1997-2004. Results indicated the model power using the probable neural network. Also, findings showed that this method is more suitable than artificial neural network and logistic regression. Gaganis et al. [7] studied the efficiency of nearest neighbor algorithm to develop some models for estimating the auditor comment type as compared to logistic regression and linear discriminant analysis. The desired sample consists of 5276 observations. Comparing a variety of methods demonstrated that the nearest neighbor method can be more effective than the others.

Hasas Yeganeh et al. [8] addressed the prediction of independent auditor report type in Iran through comparing two approaches of neural network and probable neural network. Thus, data related to the firms in Tehran Stock Exchange during 2003-2010. Research results displayed that the accuracy of probable neural network is more than neural network.

Baqerpoor Velashani et al. [3] studied the prediction of independent auditor report type in Iran and addressed the data analysis approach. In order to predict the independent auditor report, two approaches including C.5 decision tree and artificial neural networks have been applied. Data in relation to the firms accepted by Tehran Stock Exchange during 2003-2009 were used. Results indicated more accuracy of C.5 decision tree than the other technique. Barkhordarian et al. [4] reviewed the prediction of conditional auditor comments using multilayer perceptron neural network and decision tree. In the study, the power of two neural network and decision tree has been considered with respect to the prediction of auditor comments. With regard to the research goals, three hypotheses have been presented. The first hypothesis investigates the power of multilayer perceptron neural network, the second one refers to the power of CART decision tree and finally, the third one compares the results achieved by two desired models. Statistical population involves the data of financial statements presented by the firms in Tehran Stock Exchange during 2005-2009. Sampling has been conducted by the systematic deletion method and financial ratios have been used as research variables. Finally, by implementing the mentioned models, the results were analyzed. Research findings indicated that multilayer perceptron neural network is capable of predicting the auditor comments with the validity of 70% showing high capability of the model. Also, CART decision tree could predict the auditor comments with the accuracy of 70%. In each mode, the results refer to high power of model so that the first and second hypotheses have been confirmed: the possibility was provided for the beneficiaries to use these models for prediction. But comparing two models showed no significant difference and the third hypothesis has been rejected.

3 Case Study

Auditing aims to remove the information pollution of financial statements and provide a reliable and suitable bed for the users outside the organizations; the final product is to give comments on appropriateness of financial statements according to the accounting standards. Thus, considering recent events in terms of bankruptcy, liquidation and scandal of firms, the auditor plays an important role in preventing these events.Results achieved inside and outside Iran showed that the auditor report has information which is useful to predict the report type and make decisions for the next year. As well, the conditions of auditing report have negative impact on stock and rewards received by the managers. Given that the auditors extract the reports based on auditing methods and standards, not necessarily scientific models, the main problem is the explanation of effective elements scientifically in the form of exploratory research and in relation to the comments type, which are obtained by the auditors in reality. However, the effective elements include financial leverage m, profitability, capital structure, etc. In this regard, the following factors are addressed:

A. Effective elements in the auditor report type

Financial leverage. Getting loans is somehow a double-edged sword; if the loan is received timely and sufficiently, it can increase the firm sale; otherwise, the firm will face a serious risk and financial leverage is regarded as a firm status monitor for repaying the loans. Basically, if the firm cannot repay the debts and the financial leverage is high, the auditor report may not be accepted.

Profitability. Profitability criteria express the effects of liquidity, debts and assets management in the firms so that all the strategies are demonstrated. Also, continuity of a firm activity is based on profitability power of firm assets; consequently, it may be a suitable index to predict financial crisis and bankruptcy or the issue of unaccepted auditing report.

Capital structure. Considering the increase and expansion of markets followed by the firm sale growth and for developing the business unit activities, new financial resources are required, which are not necessarily provided by stockholders or internal sources and should be supplied by external sources like loans and stock. Low equity to assets ratio means excessive reliance on financial resources supplied by the debts, which have advantages involving tax ones and risks such as financial risk increase and bankruptcy. It directly affects the auditor report type.

Performance. Performance criteria indicate the application of assets in making income and profit for the firms with good performance and high profit stability. It leads to improve the financial reporting quality followed by the issue of unacceptable reports. On the other hand, the firms with weak performance try to conceal their results causing the increased probable issue of unacceptable reports.

Liquidity. Liquidity criteria refer to the firm ability to do short-time commitments. In other words, liquidity displays a relationship between cash given to the firm in short-time and cash the firm will need [2].

Firm size. As the firm is larger, number of supervision and regulatory contracts increases; therefore, the auditors apply more accuracy and precision to present the reports.

B. Other elements related to auditor report

It has been specified that financial ratios of a firm in a specific industry tend towards the averages due to competitive forces; in other words, the industry averages are indicative of optimal operational structure. Thus, the type of industry as an effective element in firm performance can be applied to predict the auditing report type. Firms with more experience in preparing financial statements and firms in stock markets because of stock regulations and supervision will provide high quality financial reporting.

This research aims to collect effective elements through studying plenty of articles comprehensively and gathering the views of experts at national and international levels and to determine the most effective elements and suitable method for predicting the report type.

4 Research Methodology

Researchers should choose a method after determining the research topic. Research method is data-driven with regard to the nature of data analysis. In order to direct data analysis systematically, a general process has to be followed. Such standard processes as CRISP-DM are mentioned.

CRISIP-DM: A standard industrial process involves six stages in relation to data analysis and is widely used in the industries. Six stages are as follows: 1- knowledge of business, 2- data recognition, 3- data preparation, 4- modeling, 5- model evaluation and 6- model development.

4.1 Research Population

Research population includes all the elements and people in a specific geographical scale with one or more traits in common. Here, research population involves all the firms in Tehran Stock Exchange during 2008-2017.

4.2 Sampling Method

To take samples, systematic deletion or screening has been done. Considering the limitations mentioned in the below Table, the sample volume is 84 corporates.

Table 1: Research statistical sample selection

	No	Description
Total	435	Firms in Tehran Stock Exchange at the end of 2017
Filter	122	Number of firms exit Tehran Stock Exchange
Filter	32	Number of firms without the 12/29 fiscal year
Filter	48	Firms classified as investors, financial intermediaries
Filter	12	Number of firms with changes in fiscal year
Filter	9	Number of firms with financial pause
Filter	25	Number of firms without sufficient data
Remains	84	Sum of firms in statistical population

5 Research Variables

X1: Auditor report (dependent variable): if it is 1, it will be accepted; otherwise, it is 0.

X2: Board size: the boards with lots of managers can be useless for the corporate and may be accompanied by lots of costs. It seems that larger board leads to improve the supervisory role and as a result, more efficacy. On the other hand, when the board gets bigger, the quality of relationships gets affected (Tehran Stock Exchange, 2006).

X3: Non-mandatory members: Non-mandatory manager is a member who is a part-time manager on the board and has no executive responsibilities in the firm. According to the article 1 of leadership principles draft, the most members of board should be non-mandatory in the stock firms(Tehran Stock Exchange, 2006).

X4: Board independence: Number of board members/ number of non-mandatory members.

X5: Auditor type: If the auditor is from government sectors, the number is 1; otherwise, it will be 0.

X6: Cash by the banks: the cash to current liabilities ratio indicates the firm ability to repay the short-term debts by the cash assets.

X7: Current debts: Debts to total assets ratio.

X8: Current assets ratio: Average ratio of market assets to net income.

X9: Net fixed assets

X10: Total assets

X11: Profit and loss after tax: profit after interest and tax to assets ratio.

X12: Total incomes

X13: Non-current liabilities.

X14: Equity: equity to assets ratio.

X15: Financial costs.

X16: Profit and loss before tax: profit before interest and tax to assets ratio.

X17: Operational profit and loss.

X18: Cash balance at the beginning of the year.

X19: Firm size: natural logarithm of firm assets.

5.1 Descriptive statistics

Descriptive statistics are those which help the researcher to classify, summarize, describe and interpret data through collecting the required information. One of the most important advantages in terms of descriptive statistics is to summarize a great volume of information. To present a general overview of important traits regarding the computed variables and using SPSS25 software, Table 2 has demonstrated some variable concepts of descriptive statistics like the biggest, and smallest values, mean, standard deviation and skewness.

5.2 Inferential statistics

One of the methods which is widely used to solve the problem of classification is support vector machine. It was first developed by Wepnik and could be generally expanded through decreasing the experimental error and avoiding

Table 2: Descriptive statistics of research variables

Variable	Min	Max	Mean	SD	skewness
Comment type	0	1	0.51	0.5	−0.02
Number of board member	3	5	4.20	0.001	0.14
Non-mandatory member	3	4	3.45	0.005	0.25
Board independence	0.6	0.8	0.65	0.23	−0.49
Auditor	0	1	0.26	0.44	1.08
Cash by banks	147	6783317	185400.51	619987.13	6.01
Total current debts	7978	150752864	3203526.86	12203969.71	7.21
Total current assets	17026	116426401	2920766.01	116426401	7.33
Net fixed assets	3238	36378676	1087440.12	3559384.79	5.94
Total assets	36481	190731126	5187222.62	16890113.12	6.47
Profit/loss after tax	−7204976	15760512	457982.31	1747132.17	4.51
Sum of incomes	7330	257851151	4112204.97	16052440.24	9.40
Non-current debts	0	23888735	347482.64	1649665.22	9.90
Equity	−11623136	49326619	1636213.12	4855343.26	5.32
Costs	0	17194689	307021.11	1357512.79	7.54
Profit/loss before tax	−7204976	15760512	499152.91	1809831.85	4.43
Operational profit/loss	−5052189	16098952	617844.29	1931865.97	5.02
Cash balance	222	5466988	160166.38	510862.37	5.67
Firm size	10.10	18.45	13.82	1.38	0.70

excessive fitness. Support vector machine finds a super plane with the maximum margin between two classes to solve the classification problem; it can be converted into a second-grade optimization problem and afterwards, it can be solved by second-grade programming technique. As well, SVM is able to classify the nonlinear samples using Kernel functions [9]. Output of support vector machine algorithm in the training section is presented in Fig. 1.

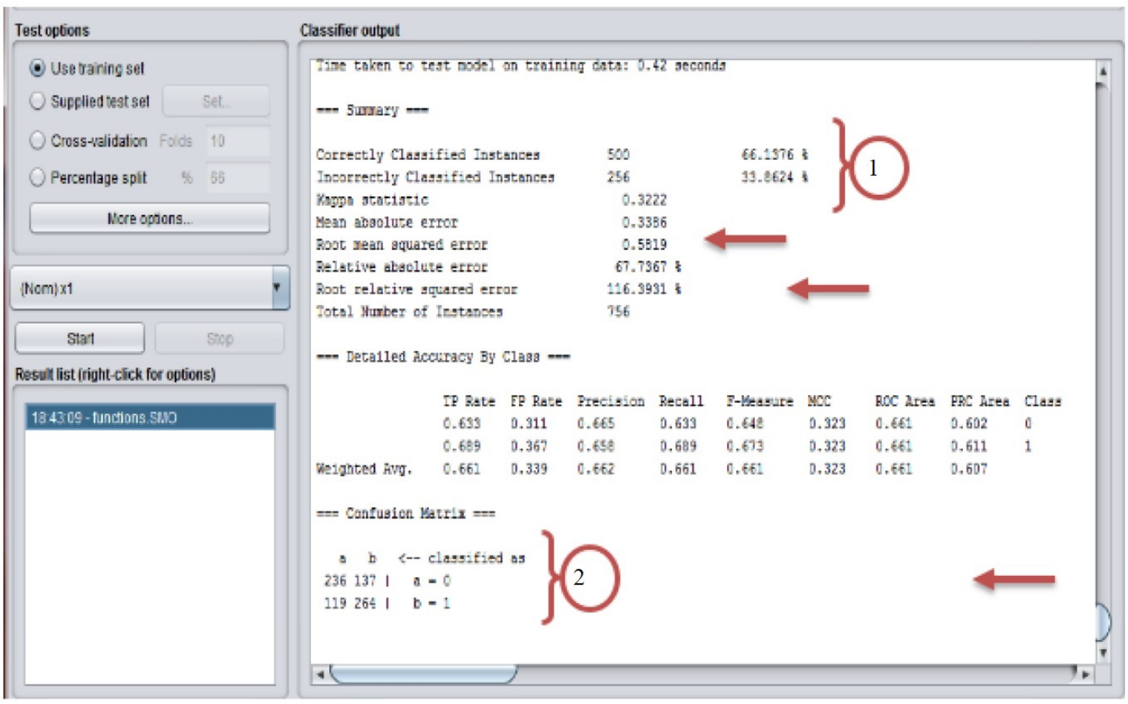


Figure 1: Output of support vector machine algorithm in training section

No. 1: Out of 756 samples, 500 samples have been classified correctly and 256 samples have been classified wrongly in Fig. 1. The accuracy and error of model are 66.13% and 33.86%, respectively.

No. 2: Data classification is different in different classes.

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

5.2.1 Evaluation criteria in training section

Evaluation criteria of support vector machine algorithm in the training section has been presented in Tab. 3.

Table 3: Evaluation criteria training section

Class	0	1
Accuracy	66.13	66.13
Precision	0.665	0.658
Recall	0.633	0.689
F-Measure	0.648	0.673
MCC	0.323	0.323
ROC	0.661	0.661

Evaluation of classification algorithms

K-Fold and cross-validation

K=10

Output of support vector machine algorithm in the training section has been demonstrated in Fig. 2.

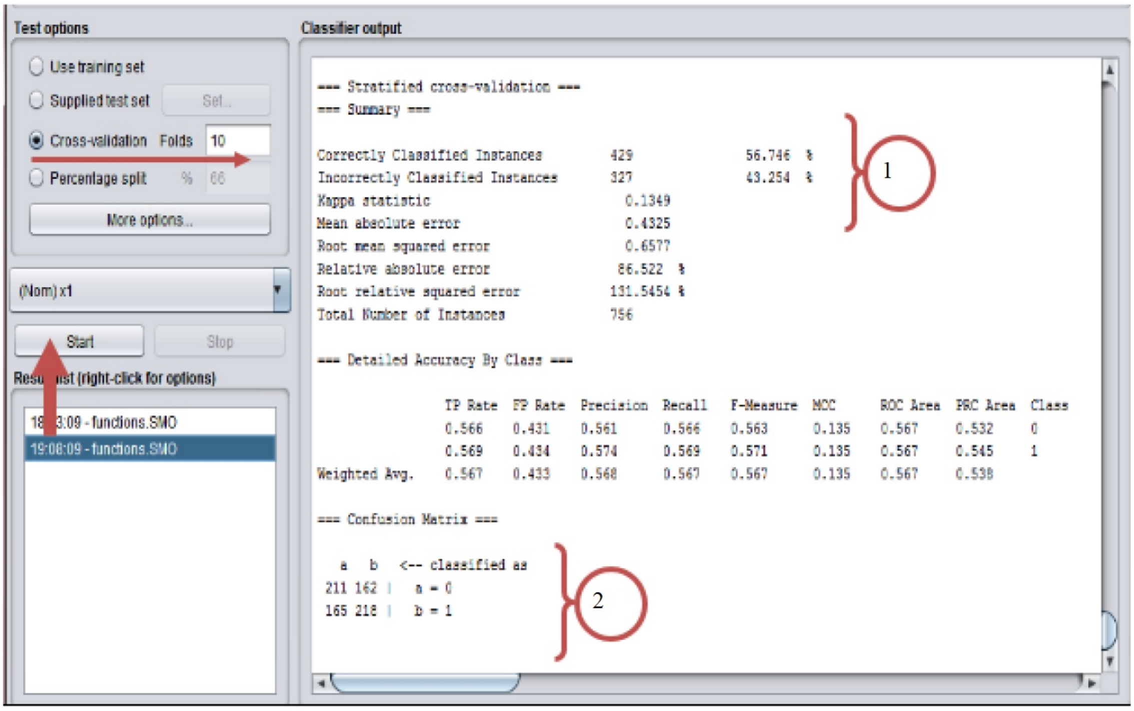


Figure 2: Output of support vector machine algorithm in training section

No. 1: In Fig. 2, it can be seen that out of 756 samples, 429 samples have been correctly classified and 327 samples were classified wrongly. The model accuracy and error are 56.74 and 43.25%, respectively.

No. 2: It shows the data classification in different classes.

Class A: 211 samples were correctly classified and 162 samples were classified in a wrong manner.

Class B: 218 samples were correctly classified and 165 samples were classified wrongly.

5.2.2 Evaluation criteria in test section

Evaluation criteria of support vector machine algorithm in the test section have been presented in Table 4.

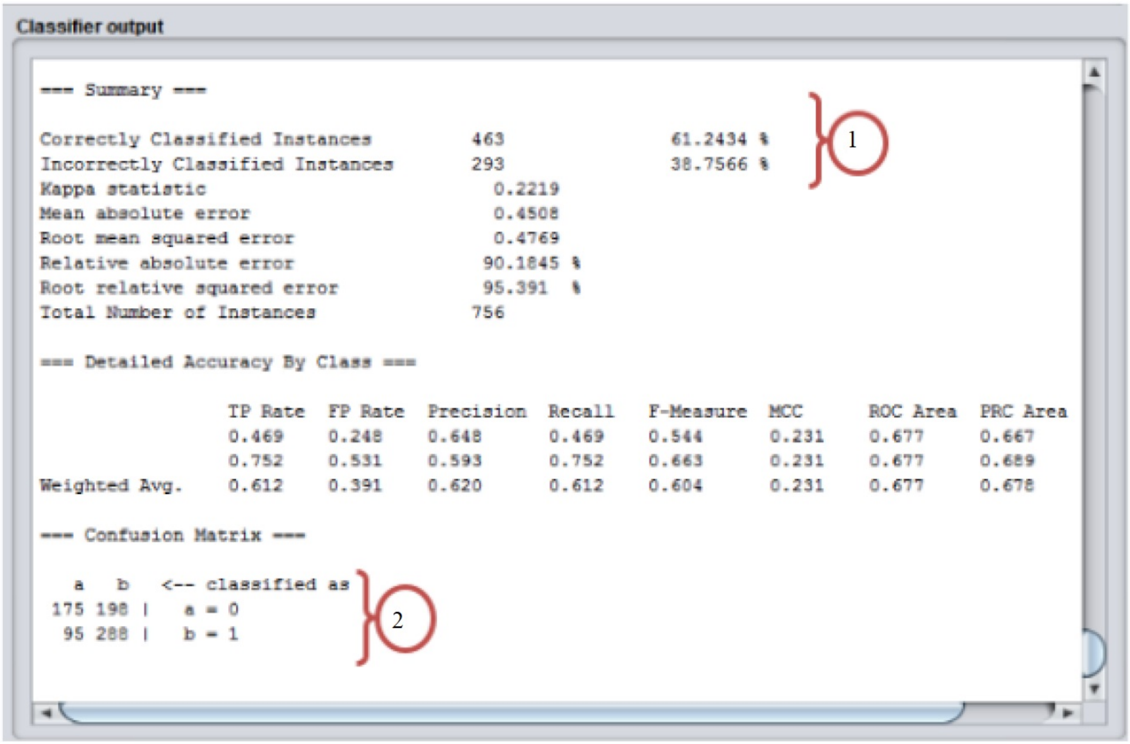
Table 4: Evaluation criteria in test section

Class	0	1
Accuracy	56.74	56.74
Precision	0.561	0.574
Recall	0.566	0.569
F-Measure	0.563	0.571
MCC	0.135	0.135
ROC	0.567	0.567

5.3 Artificial Neural Network

Artificial neural network is a system of data processing, which was taken from human brain and makes lots of tiny processors responsible to solve a problem as a parallel and connected network [10].

Output of neural network algorithm in test section has been presented in Fig. 3.


```
Classifier output

=== Summary ===

Correctly Classified Instances        463               61.2434 %
Incorrectly Classified Instances      293               38.7566 %
Kappa statistic                         0.2219
Mean absolute error                     0.4508
Root mean squared error                 0.4769
Relative absolute error                90.1845 %
Root relative squared error            95.391  %
Total Number of Instances             756

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area
                 0.469    0.248    0.648      0.469    0.544      0.231    0.677     0.667
                 0.752    0.531    0.593      0.752    0.663      0.231    0.677     0.689
Weighted Avg.    0.612    0.391    0.620      0.612    0.604      0.231    0.677     0.678

=== Confusion Matrix ===

   a   b   <-- classified as
 175 198 |   a = 0
  95 288 |   b = 1
```


Figure 3: Output of neural network algorithm in test section

No. 1: In Fig. 3, it can be seen that out of 756 samples, 463 samples have been classified correctly and 293 samples were classified wrongly. The model accuracy and error are 61.24 and 38.75%, respectively.

No. 2: It shows data classification in different classes.

Class A: 175 samples were classified correctly and 198 samples were classified wrongly.

Class B: 288 samples were classified correctly and 95 samples were classified wrongly.

5.3.1 Evaluation criteria in training section

Evaluation criteria of neural network algorithm in training section have been presented in Table 5.

Evaluation of classification algorithms

K-Fold and cross-validation

K=10

Table 5: Evaluation criteria in training section

Class	0	1
Accuracy	61.24	61.24
Precision	0.648	0.593
Recall	0.469	0.752
F-Measure	0.544	0.544
MCC	0.231	0.231
ROC	0.677	0.677

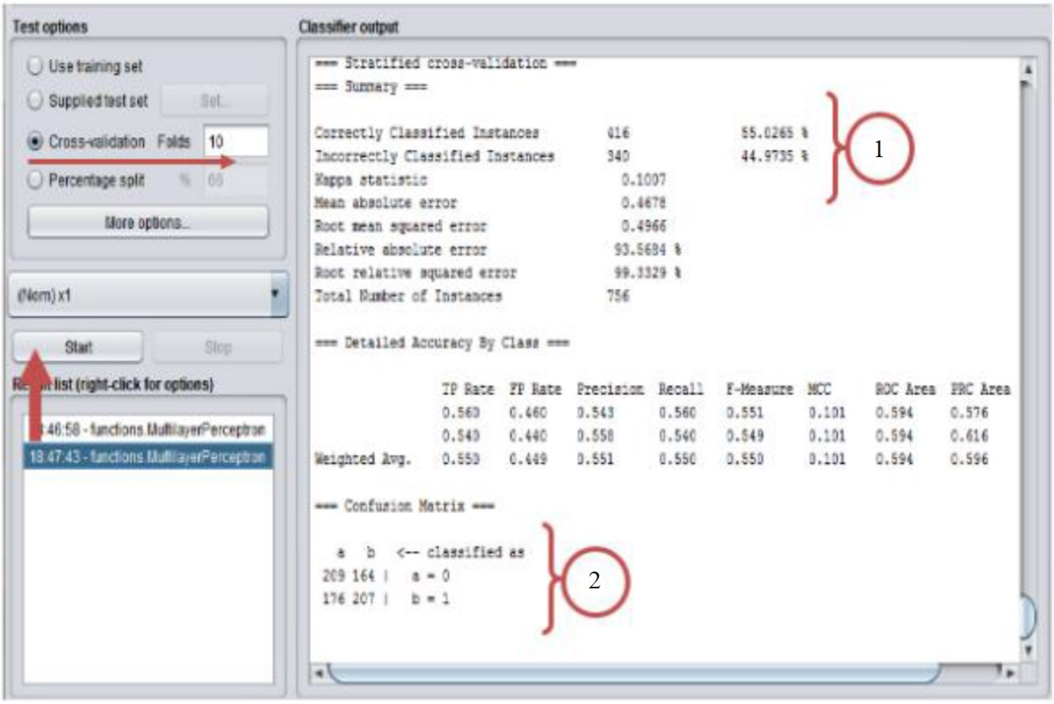


Figure 4: Output of neural network algorithm in training section

Output of neural network algorithm in training section have been presented in Fig. 4.

No. 1: Fig. 4 shows that out of 756 samples, 416 samples were classified correctly and 340 samples were classified wrongly. The model accuracy and error are 55.02 and 44.97%, respectively.

No. 2: It shows the data classification in different classes.

Class A: 209 samples were classified correctly and 164 samples were classified wrongly.

Class B: 207 samples were classified correctly and 176 samples were classified wrongly.

5.3.2 Evaluation criteria in test section

Evaluation criteria of neural network algorithm in test section have been presented in Table 6.

Table 6: Evaluation criteria in test section

Class	0	1
Accuracy	55.02	55.02
Precision	0.543	0.558
Recall	0.560	0.540
F-Measure	0.551	0.549
MCC	0.101	0.101
ROC	0.594	0.594

6 Conclusion

According to the results achieved by research models using Voka software, the accuracy of support vector machine algorithm in training and test sections is 66.13 and 56.74%, respectively. The neural network model accuracy in training and testing sections is 61.24 and 55.02%, respectively. Results indicated the efficiency of support vector machine algorithm as compared to neural network algorithm.

Research Limitations

1. The first and most important research limitation is to generalize the results to other conditions and periods based on inferential statistics. It is possible that the relationships which have been confirmed in this research do not exist in different periods and conditions or the relationships are different. Therefore, the generalization of results should be done cautiously.
2. Lack of easy and fast access to exact classified information which can be converted into the necessary formats.
3. The data extracted from financial statements have not been adjusted because of inflation. If the data are adjusted, different results may be achieved.
4. Lack of control over some effective elements in research results such as political conditions, Barjam, regulations and industry type, which may affect the study of relationships.

References

[1] A. Abdolahzade and M. Hemmatfar, *Effect of auditing report on relation of information of firms accepted by Tehran Stock Exchange*, MA thesis, Management and Accounting College, 2017.

[2] Q. Asadi and S. Azizibasir, *Studying relationships between profitability and liquidity in firms and the impact on dividends*, Manag. Message J. **2008** (2008), no. 28, 133–155.

[3] M.A. Baqerpoor Velashani, M.J. Saee, A. Meshkani and M. Baqeri, *Prediction of independent auditor report in Iran: A data analysis approach*, J. Account. Audit. Res. **5** (2013), no. 19, 134–150.

[4] B. Barkhordarian, A. Hashemi and M. Hosseini, *Prediction of conditional auditing comments using multilayer perceptron neural network and decision tree*, MA Thesis, Isfahan University, Economics and Administrative Sciences College, 2011.

[5] K. Efstathios, C. Spathis, A. Nanopoulos and Y. Manolopoulos, *Identifying qualified auditors' opinions: A data mining approach*, J. Emerg. Technol. Account. **4** (2007), 183–197.

[6] C. Gaganis, F. Pasiouras and M. Doumpos, *Probabilistic neural networks for the identification of qualified audit opinions*, Expert Syst. Appl. **32** (2007), no. 1, 114–124.

[7] C. Gaganis, F. Pasiouras, C. Spathis and C. Zopounidis, *A comparison of nearest neighbors, discriminant and logit models for auditing decisions*, Intell. Syst. Account. Finance Manag. **15** (2007), 23–40.

[8] Y. Hasas Yeganeh, M.T. Taqavifard and F. Mohammadpoor, *Using probable neural networks to identify auditors comments type: Hypothesis and practice*, Audit.: Theory Practice J. **1** (2014), no. 1, 131–159.

[9] F. Orabona, C. Castellini, B. Caputo, L. Jie and G. Sandini, *On-line independent support vector machines*, Pattern Recogn. **43** (2010).

[10] S. Russell and P. Nurvich, *Artificial intelligence: A modern approach*, CreateSpace Independent Publishing Platform, 2016.

[11] D. Zdolšek and T. Jagrič, *Audit opinion identification using accounting ratios: experience of United Kingdom and Ireland*, Aktual'ni Prob. Ekon. **115** (2011), no. 1, 285–310.