

Nataša Spahić¹, Bojana Vuković²¹ University of Novi Sad, Faculty of Sciences, Serbia² University of Novi Sad, Faculty of Economics, Subotica, Serbia

The Possibility of Predictions in Auditor's Opinion: The Case of the Serbian Tobacco Industry

DOI: 10.7595/management.fon.2018.0015

Abstract: **Research Questions:** This paper investigated whether there is a possibility for predictions of an auditor's opinion that can be used to predict, in an extremely accurate way, future developments in one company. **Motivation:** The research of Dopuch, Holhausen and Leftwich (1987); Kirkos, Spathis and Manolopoulos (2007) or Kirskos (2012) and Kim and Upneja (2014) open space for new challenges for using auditing methods. The most trying task is to find a technique that will be able to timely, accurately and with the least waste of resources respond to the challenge. The fact that auditors are forced to expand the scope and purpose of the audit work, respecting new risks that are continually changing represents the primary inspiration for this paper. **Idea:** Our goal was to explore whether one of the possible techniques for prediction the auditor's opinion – multivariate discriminant analysis – can precisely predict a correct future audit opinion and whether this analysis is useful for finding solutions to performing predictions. **Data:** The analysis was conducted using data from financial statements of 4 Serbian tobacco companies of years 2011, 2012, 2013, 2014 and 2015 published by the Serbian Business Registers Agency. **Tools:** The presented research, based on theoretical and mathematical support, uses statistical software tools Statistica. **Findings:** The application of discriminant analysis in Serbian tobacco companies showed statistically major variables of the balance sheet, manely "Intangible assets", "Supplies" and "Liabilities". Following these variables, we obtained results which we used as the predictors. The outcome of our preliminary investigation presented accurate and correct prediction which is also confirmed by historical data. The result of this investigation can be used for further more complex investigations when using some variables that will lead to discriminatory analysis for more classification groups to mark and rank the most significant variables for expressing the audit opinion. **Contribution:** Provided information is important for every business, because every entity that is listed on the business market aims to be as better as possible, and find out and exploit the possibility of avoiding a negative result.

Keywords: multivariate discriminant analysis, financial statements, auditing, the tobacco industry

JEL classification: M42, C35

1. Introduction

The need to predict the auditor's opinion is primarily focused on taking care of the company. Based on the accurately predicted audit opinions it is possible to act preventively, but also to react on possible future sequences, which is the starting premise of this paper.

In today's modern business era, auditing becomes an ever more demanding task. Auditors meet some regulatory requirements, face new data analysis, transform the audit process, the balance between own activities, and, due to dynamic and changing environments, they are forced to expand the scope and purpose of the audit work, respecting new risks that are usually changing. The space for new challenges for using auditing methods is open as presented by the research of Dopuch, Holhausen and Leftwich (1987); Kirkos, Spathis and Manolopoulos (2007) or Kirskos (2012) and Kim and Upneja (2014). The toughest task is to find a technique that will be able to timely, accurately and with the least waste of resources, respond to the challenge.

* Corresponding author: Nataša Spahić, e-mail: natasa.spahic@dmi.uns.ac.rs

The papers' goal is to present one of the possible techniques for predicting the auditor's opinion – multivariate discriminant analysis, explaining the use of Mahalanobis distance, as well as Decision Trees.

The paper presents the assumptions of discriminatory analysis and problems in its implementation. Also, the paper shows the application of multivariate discriminative analysis itself in the process of auditing the financial statements of tobacco companies in the Republic of Serbia. The purpose of the investigation is to present the results which can be interpreted as a regular basis for further research and predicting outcomes. Also, on the basis of presented results, there is a possibility to choose the manner for future proceeding to mitigate the consequences and mistakes in business decisions and reduce them to a minimum.

2. Method

According to Tabachnick and Fidell (2007), the application of multivariate discriminant methods began during the 30s of the last century as a statistical category. In time, it came to be used in economics as showed by Holdsworth, Tan and Chung (2008) or Zhou, Lu and Fujita (2015), and in the analysis of bankrupt corporations as showed by Altman (1968) or Sudjianto, Yuan, and Cela-Diaz (2010). Also it was used in determining creditworthy clients as showed by Maddala G. S. (1992) or Sun, Huang and He (2014). Similarly, Drury (2005), Debreceenyb and Gray (2010), Albashrawi (2016) or Amania and Fadlallab (2017) came to the same conclusions during the classification of potential accounting errors.

The techniques that make up this analysis can be divided into two groups: dependence and interdependence techniques. Using dependence techniques we can see the relationship between variables, where, on one side, we have dependent variables influenced by independent variables. Multiple regression and correlation, logit analysis, discriminant analysis, canonical analysis, and conjoint analysis belong to this group. In contrast, interdependence techniques are characterised as defining all phenomena observed in the same way to determine the pattern of behaviour of the variables, i.e., to determine which variables are similar to each other and which are not. The factor analysis, cluster analysis, correspondence analysis, multidimensional scaling, structural equations belong to this group.

From all the above-mentioned dependence techniques, we highlight the application of discriminant analysis, because according to Sharma (1996) it "is the most stable method" from the group of dependence techniques. This technique allows for the identification of precisely those variables which contribute the most to separating groups and predicting the probability that the object belongs to one of these groups by a set of values for independent variables. According to Kovacic (1994) "the multivariate analysis method which deals with the separation of the different groups and the allocation of observations in predefined groups is called discriminant analysis."

It is very important to point out that there are set problems which will arise in case of assumptions be abandoned. With each statistical process, it is necessary to determine whether given parameters are statistically significant. With significance testing procedures, we define how many discriminant functions are required for further analysis and how many variables will turn up. Naturally, the question arises whether all the variables are necessary to perform the most optimal separation of groups or observation allocation. It is also required to consider whether these variables are qualitative or quantitative.

One way to determine the degree of contributions of a variable in group separation is to view the discriminant coefficient of the variable. Therefore, if the number of a specific variable is close to zero, then we can leave this variable out from the discriminant function. Apart from this method, according to Kovacic (1994), this can also be approached from the standpoint of the "separation of two groups where on the basis of Fisher's discriminative approach a result obtained that the maximum value of Fisher's criterion is equal to the square of Mahalanobis distance between the two groups." If by leaving out particular variables, the new value of Mahalanobis distance is unchanged or is only slightly different from the previous, then we conclude that the omitted variables do not contribute to the separation of groups. However, just as we check how an existing variable is enclosed, we can also introduce an additional variable in addition to the previously included variables. This procedure is implemented using the Wilk's lambda.

In addition, since it is well-known that the discriminant analysis method is, in fact, a method of dependence, and we know that a dependent variable is of a qualitative type, the thing which is important to us is what if a qualitative variable included in an independent variable, or what if all independent variables are of that type. In these cases the application of the discriminative logistic analysis is suggested. If it happens that all independent variables are qualitative, or if there is a mixed case where some of the independent variables

are qualitative, then it is not reasonable to assume the normalcy. In this case, Fisher’s discriminative function gives poor classification results. Therefore, if a normal distribution may not be supposed, the use of discriminative logistic analysis is recommended.

Taking into account available techniques, we consider that is useful to use them to examine whether it is possible to determine the auditor’s opinion.

3. Application of Multivariate Discriminative Analysis in the Process of Auditing the Financial Statements of Tobacco Companies in the Republic of Serbia

The tobacco industry is one of the older productions which, in addition to growing and processing tobacco, also manufactured tobacco products. In the Republic of Serbia, cigarettes are produced by:

- Niš Tobacco Industry - Philip Morris Company, (1.)
- Vranje Tobacco Industry - British American Tobacco, (2.)
- Japan Tobacco International, Senta, (3.)
- Coka Tobacco Industry (4.) and
- Monus Indjija. (5.)

Monus has not been included in this analysis as it is a small legal entity and thus is not required to conduct an audit of its financial statements.

The data used for this analysis refer to 2011, 2012 and 2013; they were officially published and are publicly available by the law the Republic of Serbia on the website of the Serbian Business Registers Agency.

The balance sheet positions analysed are: “Intangible Assets”, “Supplies”, “Receivables”, “Capital” and “Liabilities”.

Taking into account the five account balances from the balance sheets of the above-mentioned tobacco industries, by previous years (2013, 2012 and 2011), we will try to find out which of these positions affects the auditor’s opinion the most.

Table 1: Data from the balance sheet (in 000 dinars) **Note:** The auditor’s opinion has been replaced by numbers to make managing data easier. 1=positive (unqualified opinion); 5=qualified opinion

Year	Firm	Intangible Assets	Supplies	Receivable	Capital	Liabilities	Audits' Opinion *
2013	1	689,511	4,942,207	3,251,774	13,572,140	1,900,303	1
	2	0	1,793,333	4,066,425	0	306,898	5
	3	1,174	2,638,659	1,957,554	2,844,475	1,411,948	5
	4	589	165,77	15,946	252,948	25,755	1
2012	1	682,472	4,691,889	2,990,171	11,741,260	1,585,287	1
	2	0	825,242	8,624,124	0	596,446	1
	3	1,549	2,071,759	2,397,732	3,407,903	407,498	5
	4	595	175,684	24,1	224,923	11,332	1
2011	1	697,514	3,546,548	2,704,902	11,509,543	1,060,225	1
	2	25,414	512,89	5,999,750	0	279,154	1
	3	1,806	2,010,706	2,492,189	1,190,490	423,179	5
	4	601	175,684	45,766	214,785	201,696	1

Source: Authors’ calculation

In this paper, multivariate discriminant analysis is applied during the audit process, by using the auditor’s opinion for grouping observations which are based on historical data. In the case of the tobacco industry in the Republic of Serbia, this is either “unqualified” or “qualified”, while the variables used for separation of observations are positions from the balance sheets taken by the highest value. The discriminant function is a linear combination of variables, in this case, a linear combination conditioned by the data obtained from the balance sheet. Specifically, by using the specified data, the following results are obtained:

Table 2: Summary of Discriminant analysis (calculated in the programming tool STATISTICA)

N=12	Discriminant Funcion Analysis Summary					
	No. of vars in model: 5; Grouping: Opinion (2grps)					
	Wilks' Lamda: .06984 approx. F85.6)=15.982 p<.0021					
	Wilks' Lamda	Partial Lamda	F-remove (1.6)	p-value	Toler.	1-Toler (R-Sqr.)
Intangible Assets	0.153882	0.453849	7.22027	0.036194	0.018753	0.981247
Inventory	0.479963	0.145509	35.23446	0.00102	0.01093	0.98907
Supplies	0.075966	0.919343	0.5264	0.495439	0.64315	0.35685
Equity	0.071074	0.982631	0.10606	0.755736	0.01844	0.98156
Liabilities	0.146855	0.475566	6.61654	0.042206	0.0661	0.9339

Source: Authors' calculation

Statistically significant variables are shown in bold; in fact, these are the variables that will contribute most to the discrimination between groups. It can be seen that for "Intangible Assets", "Supplies" and „Liabilities“ Partial Lambda has the lowest values compared to the others. Also, the p-values are less than 0.05 which means that they really are statistically significant (there is a statistical significance for group separation and it is evident that the expected average of all independent variables is not equal). The F value of a certain variable indicates its statistical significance in the discrimination between groups. It is a measure of a variable's unique contribution in the process of predicting the belonging to a group or an object, where it can be observed that the highest value(s) is/are for "Supplies", "Intangible Assets" and "Liabilities".

Tolerance is the coefficient of determination between the dependent variable and the other variables included in the model. That is, therefore, a measure of redundancy of a given variable. If the Tolerance of some of the variables entered into the equation which describes the model is lower than the default tolerance which is 0.01, this means that the variable is 99% superfluous. In this example, this is not the case, given that they all have a Tolerance higher than 0.01.

The prediction of belonging to a group conducted with the help of the Mahalanobis distance. This distance is quite similar to the squared Euclidean distance of the respective case (object) from the centroids for each group. A classification is performed so that the object is assigned to a group according to Mahalanobis distance (object from centroid). In this example, there are two groups G_1:1 and G_2:5 where the first represents a case when the auditor's opinion will be unqualified, and the other instance will be a qualified opinion.

Table 3: Squared Mahalanobis distance

Case	Squared Mahalanobis Distances from Group Centroids Sign * will be use for incorrect classifications		
	Observed Classification	G_1:1 p=.66667	G_2:5 p=.33333
1	G 1:1	4.29595	44.42499
2	G 2:5	52.09452	3.46909
3	G 2:5	44.87213	6.72167
4	G 1:1	2.12224	46.68526
5	G 1:1	5.24856	36.87013
6	G 1:1	5.70828	53.72985
7	G 2:5	62.69846	6.33225
8	G 1:1	2.22645	45.13151
9	G 1:1	8.36075	93.33558
10	G 1:1	2.13320	54.87442
11	G 2:5	57.65649	1.01939
12	G 1:1	2.36217	56.96424

Source: Authors' calculation

For further analysis, we need to consider a correlation matrix which shows to what extent two individual variables are in correlation, i.e., which variables in their mutual relationship have a high, and which show a low correlation coefficient. The correlations between independent variables (with outliers which appeared in these correlations) are represented graphically in Figure 1.

The highest coefficient of correlation is found between the variables “Capital” and “Intangible assets”, and immediately after that, between the “Supplies” and “Capital” as well as “Supplies” and “Intangible assets”. Finding the correlation between each of these two variables helps us predict with greater certainty how the value of another variable will change if the first one changes. Contrary to this, the correlation for variables “Capital” and “Receivables” is negative.

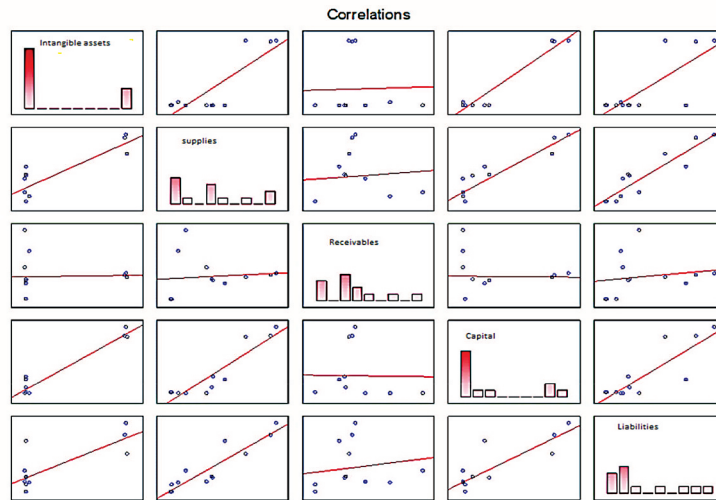


Figure 1: Correlation of all variables

In case of the value analysis of the five balance sheet positions, we receive a statistically significant variable which poses the question for further data categorisation. The following Decision Tree is obtained by using these data:

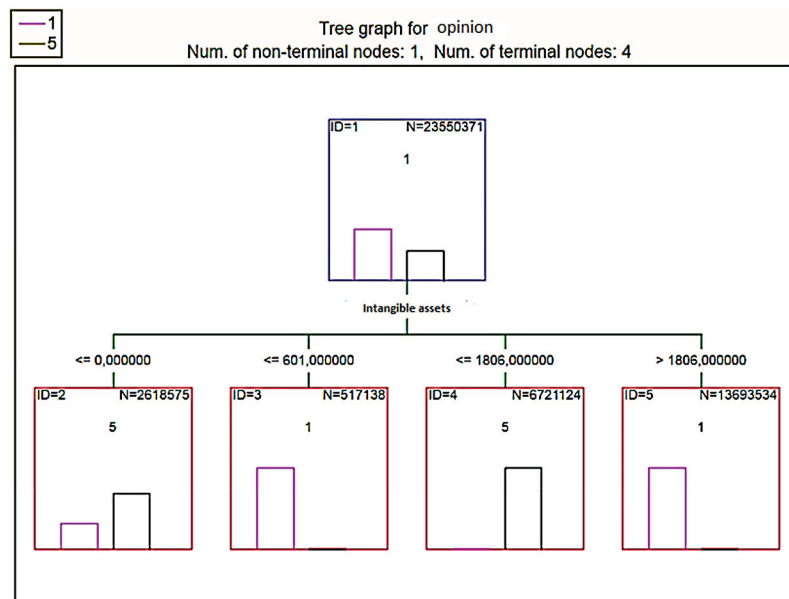


Figure 2: Decision Tree based on Intangible Assets

We can see that there are one major node and four child nodes. Depending on the number of Intangible Assets, the rectangles contain a “1” if the opinion is unqualified and a “5” if the opinion is qualified. Therefore, if Intangible Assets equal zero or are between 601 and 1806, then the opinion will be qualified, while in other cases it will be unqualified. What can also be noticed is that on the first node, within the given node, there are rectangles that represent a percentage of the groups of dependent variables. Colours separate groups related to different opinions, while the height of the rectangle notes how often a particular group occurs. Therefore, via these rectangles, each subsequent node (“answer node”) points out to which group the observed object will belong in comparison to the observed independent variables.

The Decision Tree provides a geometric framework for organising decisions. Every Decision Tree provides us with an alternative to the next step. In addition to “Intangible Assets”, the Decision Tree is used to present positions under “Supplies” and “Liabilities”.

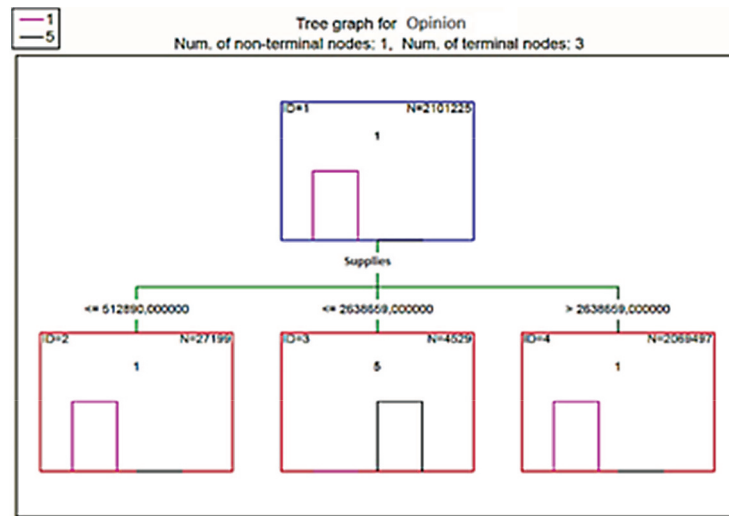


Figure 3: Decision Tree for Supplies

The amounts of supplies smaller than 2.638.659 and higher than 512.890 (in thousands of dinars) reflect the qualified audit opinion.

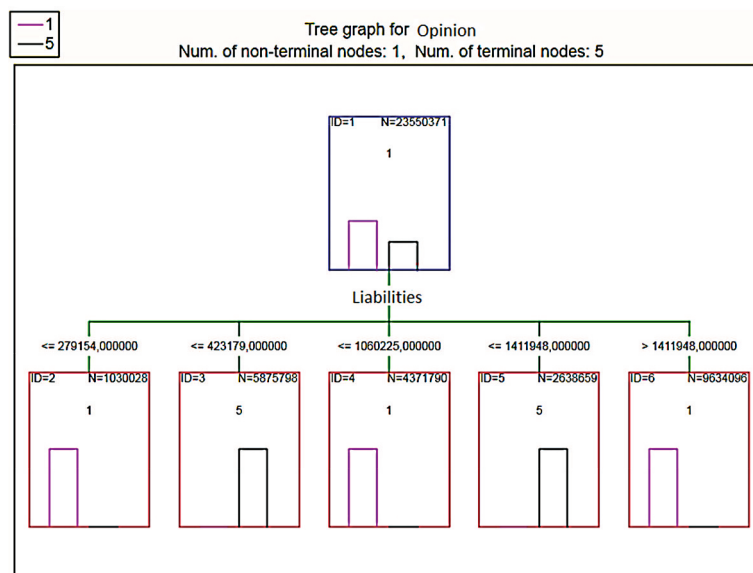


Figure 4: Decision Tree for Liabilities

The interpretation of the tree graph “Opinion” for Liabilities shows that there exists a total of six nodes, more precisely one root node and the other five nodes of the response. It means that all amounts larger than 279.154 and smaller than 423.179 (thousands of dinars), as well as higher than 1.060.225 and smaller than 1.411.948 (thousands of dinars), show a qualified audit opinion. All others will show an unqualified audit opinion.

The CHAID method (Chi-square Automatic Interaction Detector) was used for the previous three Decision Trees. Components used in this method can be continuous, sequential and nominal. The CHAID Tree is a Decision Tree built by dividing the subsets into two or more answer nodes, on many occasions, starting with the entire data set.

It is important to note that the CHAID is not the only method that can deliver results. There is also a possibility for using the C&RT (Classification and Regression Tree) method as presented in Huberty and Olejnik (2006), James and Witten (2013) or Shaikh, Adi and Logrippa (2017). Its purpose is to use the Tree construction algorithm to determine a set of *If-then* logical conditions that allow precise estimates for the classification of

the observed object (statement). It creates interactive regression Trees for continuous and categorical predictors; building an optimal Tree structure and also providing incessant dependent variables, as presented in detail by Wang (2007), Gepp (2010) and Clemen and Reilly (2014).

However, it should be noted that if some of the discriminatory analysis conditions is broken, then it cannot be applied, and some other techniques have to be used. These are the lack of multipolarity between independent variables, the presence of an independent multivariate normal distribution; the matrices of variance and covariance of independent variables between groups are equal and similar.

4. Results and Discussion

The primary conclusion arising from this paper is that there is a real basis for further and extensive research that will consider the use of multivariate discriminant analysis to predict the auditor’s opinion regarding the business of a particular company.

Through the specific analysis and computer data processing, we concluded that the statistically most material positions in the balance sheet of the tobacco producing companies are “Intangible Assets”, “Supplies” and “Liabilities”. With further processing, we arrived at the discriminant function coefficients through which we obtained additional information about the significance of independent variables. Given that the discriminant analysis seeks to predict future developments through data clustering, a variable with greater discriminant abilities accompanied by a coefficient with higher values issupporting that.

With the help of centroids, representing the matrix discriminant scores for each group, we managed to, making use of the value of the “cutting score”, predict what the character of the auditor’s opinion will be in the analysed tobacco industry companies. The results showed that in a situation when the value of the obtained discriminant functions was below the “cutting score” the auditor’s opinion was qualified and if the value was above them, it was unqualified.

Based on the received values and the presence of “concentration of tobacco market” as quoted in Vukovic, Mijic, Spahic (2015), the auditor’s opinion from previous years is undisputed in predicting future outcomes, in addition to the values of significant balance sheet positions. In this manner predictions, with some speculations, can be made for the following year. But, would we be mistaken if we were to expect that all auditors’ opinions for 2013 would be same in 2014?

We assessed that the results from the application of multivariate discriminant analysis could best be viewed by applying the results for the upcoming 2014 and on the balance sheet of the above-mentioned companies.

If we look at it from the company point of view, with the help of discriminant functions that have already been calculated by historical data,

$$f = 0,00001887 * x_1 - 0,00000511 * x_2 + 0,00000014 * x_3 + 0,00000019 * x_4 + 0,00000434 * x_5 + 2,6494 \quad (1)$$

variables x_1, x_2, x_3, x_4 and x_5 are independent variables, and they represent the corresponding amounts from the 2014 balance sheet more accurately.

Table 4: Values of independent variables (for 2014)

Firm	X ₁ Intangible Assets	X ₂ Supplies	X ₃ Receivable	X ₄ Capital	X ₅ Liabilities
1	594,581	5,604,441	4,234,617	13,395,080	2,615,566
2	0	1,167,752	1,303,318	0	142,209
3	728	2,257,507	1,718,557	2,381,255	608,17
4	583	260,392	15,946	252,473	26,751

Source: Authors’ calculation

The calculated value of the discriminatory function of the Tobacco Industry (1) is equal to -0.28008202 , for the Tobacco Industry (2) is -2.51816114 , for the Tobacco Industry (3) is -5.54022918 , and for the Tobacco Industry (4) is 1.49609974

After calculating the value of the discriminant functions for each company, the resulting value is compared with the previously determined “cutting score”. If the value for a specific company is below -1.178 , then the opinion will be a qualified opinion and if it is above those values then the opinion will be unqualified.

The results are as follows:



Figure 5: Values of discriminant functions presented on a number line

Therefore, the tobacco industry (2) and the tobacco industry (3) have a value of the function below -1.178 , so they can expect a qualified opinion while the tobacco industry (1) and tobacco industry (4) have a value above -1.178 and can expect an unqualified opinion.

After the conducted audit for 2014, the results publicly disclosed during 2015 confirm the results predicted by discriminant analysis. According to the data from the website of the Serbian Business Register Agency, we had access to the audit reports of tobacco companies and confirmed that tobacco companies (2) and (3) acquired a qualified audit opinion, while the tobacco companies (1) and (4) actually obtained an unqualified audit opinion.

Conclusion

In today's world that is continuously and persistently progressing in technological terms, feeding on a daily basis we are forced into finding methods to take better advantage of the time and place for making timely business decisions and acting adequately. This claim has a stronghold in a multitude of published papers such as Pasiouras; Gaganis and Zopoundidis (2007); Li, Huang and Shi (2008); Mojtaba, Sarikhani and Ebrahimi (2012) and Saif, Sarikhani and Brahim (2012). It is very important to be able to, by perfecting techniques and methods, predict the future outcome with particular precision. One of the answers to predicting outcomes and the manner of proceeding, so that the consequences and mistakes in decisions should be reduced, is the presented analysis where we use historical data to look for some consistency in behaviour, in order that we be able to see what we can expect.

In particular, the focus of this paper was the application of multivariate discriminant analysis as a method for data manipulation, the essence of which is the proper classification of observations. By relying on discriminant analysis, during the auditing process, we obtained some significant results. However, naturally, we cannot ignore the fact that it is hard to avoid deviation in this analysis so that every result and outcome should be taken with a grain of salt – seeking new, better and more applicable techniques.

The presented research, based on theoretical and mathematical support, using statistical software tools, managed to apply discriminant analysis on the data from the financial statements of the tobacco industry in the Republic of Serbia. The aim was to predict the success of a business in the future and what kind of opinion will be given by the auditor. Concretely, we used the following positions from the balance sheet for the predictors: “Intangible assets”, “Supplies”, “Receivables”, “Capital” and “Liabilities”. The application of discriminant analysis showed that major variables are “Intangible assets”, “Supplies” and “Liabilities”. Following these variables, we obtained certain results on which we were able to make the predictions. Also, the given variables were entered into the discriminatory function, so we calculated the values of the given function and precisely located the membership of a particular group. Groups of dependent variables were the opinions of the auditor. According to the data from the website of the Serbian Business Register Agency, we had access to the financial statements of tobacco companies through the officially published data. It should be noted here that the auditor's opinions varied between “positive (unqualified)” and “positive with distraction (qualified)”, so the classification was established on the basis of that existing situation. The result of our preliminary investigation presented accurate and correct prediction and can be used for further complex investigations.

The presented results allow us to continue our research to make the discriminatory analysis as credible as possible. The study will be continued in a way that the number of cases in the smallest group will be superior to the number of independent variables. It means that we will have some variables that will lead to discriminatory analysis for more

classification groups. All of this implies that the future work will be based on several types of auditors' opinions. Using the Wilks' Lambda and gaining more discriminating functions by inserting a largest number possible of classification groups (divided over more different auditor's opinion), the research will be based on a wider Decision Tree to mark and rank the variables for expressing the audit opinion.

We consider purposeful to draw attention to a processed example of the so-called supervised learning via a Decision Tree. Via the Decision Tree we reach a logical-consequential connection that leads to grouping. The Decision Tree provides a geometric framework for organising decisions and ensures some alternatives as to the next step. What connects multivariate discriminant analysis and the Decision Tree is the ability to find hidden knowledge. That is, it is what every participant on the market is seeking – that they will be able to, by reading publicly available data, see something before others do, or predict something seemingly unexpected.

REFERENCES

- [1] Albashrawi M. (2016). Detecting Financial Fraud Using Data Mining Techniques: A Decade Review from 2004 to 2015, *Journal of Data Science*, 14, 553-570.
- [2] Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance*, 23 (4), 589-609. DOI: 10.2307/2978933
- [3] (3)Amania F.A. and Fadlallab A.M. (2017). Data mining applications in accounting: A review of the literature and organizing framework, *International Journal of Accounting Information Systems*, 24, 32-58. <http://dx.doi.org/10.1016/j.accinf.2016.12.004>
- [4] Clemen, R. T., & Reilly, T. (2014). *Making Hard Decisions with DecisionTools*, South-Western Cengage Learning; 3rd edition, 1-816
- [5] Debrecenyb R.S. and Gray G.L. and (2010).Data mining journal entries for fraud detection: An exploratory study, *International Journal of Accounting Information Systems*, 11, 157-181. DOI:10.1016/j.accinf.2010.08.001
- [6] Dopuch N., Holthausen R.W. &Leftwich R.W (1987). Predicting Audit Qualifications with Financial and Market Variables, *Accounting Review*, 62 (3), 431-454.
- [7] Drury C. (2005). *Management Accounting for Business*, Thomson Learning, London, 47-159.
- [8] Gepp, A., K. Kumar, & S. Bhattacharya. (2010). Business failure prediction using decision trees, *Journal of Forecasting*, 29, 536-555. DOI:10.1002/for.1153
- [9] Gray G.L. &Debrecenyb R.S. (2014). 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, <http://dx.doi.org/10.1016/j.accinf.2014.05.006>
- [10] Holdsworth K. D., Tan S. S. & Chung K.,(2008). Insolvency Prediction Model Using Multivariate Discriminant Analysis and Artificial Neural Network for the Finance Industry in New Zealand.*International Journal of Business and Management*, 3 (1),19-29.
- [11] Huang S.Y., Tsaih R.H. & Yu F. (2014). Topological Pattern Discovery and Feature Extraction for Fraudulent Financial Reporting, *Expert Systems with Applications*, 41(9), 4360–4372. DOI: 10.1016/j.eswa.2014.01.012
- [12] Huberty C. &Olejnik S. (2006). *Applied MANOVA and Discriminant Analysis*, John Wiley & Sons, Inc., Hoboken, New Jersey, 3-33
- [13] James G., Witten D., Hastie T. &Tibshirani R. (2013). *An Introduction to Statistical Learning with Applications in R*, Springer, New York, 127-175. doi 10.1007/978-1-4614-7138-7
- [14] Kim S.Y. &Upneja A. (2014). Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models, *Economic Modelling*, 36, 354-362. DOI:10.1016/j.econmod.2013.10.005
- [15] Kirkos E., Spathis C. &Manolopoulos Y. (2007). Data Mining for the detection of fraudulent financial statements, *Expert Systems with Applications*, 32(4),995-1003.DOI: 10.1016/j.eswa.2006.02.016
- [16] Kirkos E., Spathis C. ,Nenopoulos A. &Manolopoulos Y. (2007). Identifying Qualified Auditors' Opinions: A Data Mining Approach, *Journal of Emerging Technologies in Accounting* 4(1), 183-197 DOI:10.2308/jeta.2007.4.1.183.
- [17] Kirkos E. (2012) Predicting auditor switches by applying data mining, *Journal of Applied Economic Sciences* 7 (3), 246-261
- [18] Koskivaara E. (2003). Artificial neural in Auditing: State of the Art, *Turku School of Economics and Business Administration*, Turku, Finland, .2-18.
- [19] (19)Kovacic, Z. (1994). *Multivariate Analysis*. Faculty of Economy, University Belgrade, 95-188.
- [20] (20) Kwak, W., S. Eldridge, Y. Shi, & G. Kou. (2011). Predicting Auditor Changes Using Financial Distress Variables and the Multiple Criteria Linear Programming (MCLP) and Other Data Mining Approaches, *Journal of Applied Business Research*, 27(5), 73 - 84. DOI: 10.19030/jabr.v27i5.5597

- [21] (21) Li, J., Huang, K. Y., Jin, J. & Shi, J. (2008). A survey on statistical methods for health care fraud detection. *Health Care Management Science* 11, 275-287. DOI: 10.1007/s10729-007-9045-4
- [22] (22) Maddala G. S. (1992). *Introduction to Econometrics*, Macmillan, Second Edition, 127-201.
- [23] (23) Mojtaba S., Sarikhani M. & Ebrahimi F. (2012). Finding rules for audit opinions prediction through data mining methods, *European Online Journal of Natural and Social Sciences* 1(2), 28-36.
- [24] (24) Pasiouras F., Gaganis Ch. & Zopounidis C., (2007). Multicriteria decision support methodologies for auditing decisions: the case of qualified audit reports in the UK. *European Journal of Operational Research* 180 (3), 1317-1330. DOI: 10.1016/j.ejor.2006.04.039
- [25] (25) Saif S.M., Sarikhani M. and Ebrahimi F. (2012). Finding rules for audit opinions prediction
- [26] through data mining methods, *European Online Journal of Natural and Social Sciences* 1(2), 28-36.
- [27] (26) Shaikh, R.A., Adi, K. & Logrippio L. (2017). A Data Classification Method for Inconsistency and Incompleteness Detection in Access Control Policy Sets, *International Journal of Information Security* 16, 91-113. DOI: 10.1007/s10207-016-0317-1
- [28] (27) Sharma, S. (1996). *Applied Multivariate Techniques*, University of South Carolina, John Wiley & Sons, Inc., 1-468.
- [29] (28) Sudjianto, A., Yuan, M., Kern, D., Nair, S., Zhang, A. & Cela-Diaz, F. (2010). Statistical methods for fighting financial crimes. *Technometrics* 52, 5-19. DOI: 10.1190/tech.2010.07032
- [30] (29) Sun J, Li H., Huang Q.H. & He K.Y (2014). Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches, *Knowledge-Based Systems*, 57, 41-56. DOI: 10.1016/j.knsys.2013.12.006
- [31] (30) Tabachnick B. & Fidell L. (2007). *Using Multivariate Statistics*, Pearson Education, 249-268.
- [32] (31) Vukovic, B., Mijic K. & Spahic N. (2015). Concentration of tobacco market: Evidence from Serbia, *Economics of Agriculture*, 62(2), 385-398. DOI: 10.5937/ekoPolj1502385V
- [33] (32) Wang W., Guan X. & Zhang X. (2008). Processing of massive audit data streams for real-time anomaly intrusion detection, *Computer Communications* 31 (1), 58-72.
- [34] (33) Wang X, Nauck D.D., Spott M & Kruse R. (2007). Intelligent data analysis with fuzzy decision trees, *Soft Computing* 11(5), 439-457. DOI: 10.1007/s00500-006-0108-0
- [35] (34) Zhou L, Lu D & Fujita H. (2015). The performance of corporate financial distress prediction models with features selection guided by domain knowledge and data mining approaches, *Knowledge-Based Systems*, 85, 52-61. DOI: 10.1016/j.knsys.2015.04.017

Received: 2017-09-07

Accepted: 2018-05-30



About the Author

Nataša Spahić

University of Novi Sad, Faculty of Sciences, Serbia
 natasa.spahic@dmi.uns.ac.rs

Ph.D. Nataša Spahić is an associate professor at the Faculty of Sciences, Department of Mathematics and Informatics, University of Novi Sad. She teaches at the basic and masters studies in subjects closely related to her scientific interests: Audit, Accounting and Analysis of balance sheet.

Besides to the professors' engagement she performs the function of Director of internal audit in OTP Bank Serbia and Novi Sad.



Bojana Vuković

University of Novi Sad, Faculty of Economics, Subotica, Serbia
 bojanavuk@ef.uns.ac.rs

Bojana Vuković, Ph.D is an assistant professor at the Faculty of Economics, Subotica, Department of Finance, Banking, Accounting and Auditing. She teaches Auditing, Analysis of Financial Statements and Auditing of Public Sector Entities. The teaching activity is combined with the applied activity working auditing of financial statements and providing consulting services in different projects.

