A METHOD TO IMPROVE ECONOMIC PERFORMANCE EVALUATION USING CLASIFICATION TREE MODELS

Madalina Ecaterina POPESCU
(Corresponding Author)
The Bucharest University of Economic Studies,
6 Romana Square, 010374, Bucharest, Romania
The National Scientific Research Institute for Labour and Social Protection,
6-8 Povernei Str., District 1, Bucharest, 010643, Romania;
madalina.andreica@gmail.com; +40722827889

Marin ANDREICA

The Bucharest University of Economic Studies,
6 Romana Square, 010374, Bucharest, Romania,
marinandreica@gmail.com

Dragos MICU

The Bucharest University of Economic Studies,
6 Romana Square, 010374, Bucharest, Romania, micuvdragos@gmail.com

ABSTRACT

In this paper we propose a method for improving the evaluation process of economic performance applied for the case of the 28 European Union countries based on macroeconomic indicators, Hierarchical Cluster Analysis and CHAID decision trees. Using indicators of economic growth, current account, labour productivity, unemployment rate and real net earnings for the year 2013 we first applied a Hierarchical Cluster Analysis and identified two main E.U. country clusters— one corresponding to high economic performance countries and the second one corresponding to the E.U. countries with lower economic performances in 2013. We then build several CHAID decisional trees and tested their prediction ability in order to determine the most efficient model that can correctly classify the E.U. countries into high and low economic performance countries.

Keywords: *economic performance, evaluation method, CHAID tree, classification, efficiency*

1. Introduction

In this paper we propose a method for improving the evaluation process of economic performance that was applied for the case of the 28 European Union countries, using indicators of economic growth, current account, labour productivity, unemployment rate and real net earnings for the year 2013. The choice of the macroeconomic indicators used in the study in order to describe the economic performance of a country was based on the results of previous studies concerning aspects of competitiveness, adjustment and macroeconomic risk management (Matei, Zamfir and Lungu, 2014; Rahman, 2008; Spahn, 2013) and real convergence within the European Union (Lazar, 2010; 2012).

Having into consideration macroeconomic performances, we decided to apply a Hierarchical Cluster Analysis in order to identify two main E.U. country clusters, one corresponding to high economic performance countries and the second one corresponding to the E.U. countries with lower economic performances registered in 2013.

Having this classification in mind, we further on propose a method to improve performance evaluation based on decision trees.

However, the international literature concerning methods for efficient classification is quite vast. For instance, Altman (1968) used a multivariate technique, known as Multivariate Discriminant Analysis (MDA), Ohlson (1980) proposed the logistic model and Shumway (2001) used the hazard model, which is actually a multi-period logit model because the likelihood functions of the two models are identical.

In more recent years many types of heuristic algorithms such as neural networks and decision trees have also been applied as classification models. For example the studies made by Tam and Kiang (1992), and by Jain and Nag (1998) provided evidence to suggest that neural networks outperform conventional statistical models such as discriminant analysis, logit models in financial applications involving classification and prediction.

Soon after that, hybrid Artificial Neural Network methods (ANN) were proposed. For example, Yim and Mitchell (2005) tested the ability of a new technique, hybrid ANN’s to predict corporate distress in Brazil. The models used in their study were compared with the traditional statistical techniques and conventional ANN models and the results suggested that hybrid neural networks outperform all other models.

According to Zheng and Yanhui (2007) and to Koyuncugil and Ozgulbas (2007), however, who used CHAID decision tree methodologies for corporate financial distress prediction, the main disadvantage of neural network models consists in the difficulty of building up a neural network model, the required time to accomplish iterative process and the difficulty of model interpretation. On the other hand, compared to neural networks, decision tree is not only a non-linear architecture, which is able to discriminate patterns that are not linearly separable and allow data to follow any specific probability distribution, but also plain to interpret its results, require little preparation of the initial data and perform well with large data in a short time. Thus, the CHAID classification tree model has more advantages in comparison to a neural network model or to a statistic model such as logistic regression (Andreica, 2013), and multivariate discriminate regression, where the patterns need to be linearly separable and samples are assumed to follow a multivariate normal distribution. Taking these arguments into consideration, we decided to base our study on CHAID decision tree models. Therefore, several CHAID decision tree models will be built in this study and tested for their prediction ability in order to determine the most efficient model that can correctly classify the 28 E.U. countries into high and low economic performance countries.

The paper is structured as follows: section 2 is dedicated to the data description, section 3 presents the evaluation method based on CHAID decision tree methodology, while the conclusions are presented in the last section.

2. Data description

In order to evaluate the economic performances of the 28 European Union member states for the year 2013, we took into consideration the following variables, based on the Eurostat database: economic growth (%), current account balance (% GDP), unemployment rate (%), labour productivity (calculated as a ratio between real GDP and the employed population) and the annual net earnings (expressed in euros and deflated using the Harmonised Indices of Consumer Prices (HICP)).

European countries have been heavily challenged during the last years. There are important differences among them, as in 2013, only 11 countries registered economic growth below E.U. 28 average, while in the other 17 countries, GDP growth rate exceeded E.U. 28 average (0.1%) and varied from 0.2 to 4.2% (see fig. 1). Most of the E.U. member states registered only slight improvements in the GDP growth rate as compared to the previous year, while Romania is one of the five countries that registered the highest growth among Latvia, Lithuania, Malta and Luxemburg. However, this annual GDP growth of about 3.5% is not based on sustainable economic grounds, but rather on a favourable agricultural year.

When considering the current account levels for the year 2013, the European Union countries can once again be separated in two main groups, since there are ten E.U. member states that registered a current account deficit, out of which United Kingdom is by far in the top, with a -4.4% current account balance as percentage of GDP. Next are Cyprus (-1.9%), Belgium (-1.6%) and Czech Republic (-1.2%). At the opposite pole are the Netherlands, with a current account surplus of 10.4%, followed by Germany (7.5%), Denmark (7.3%) and Ireland (6.6%).

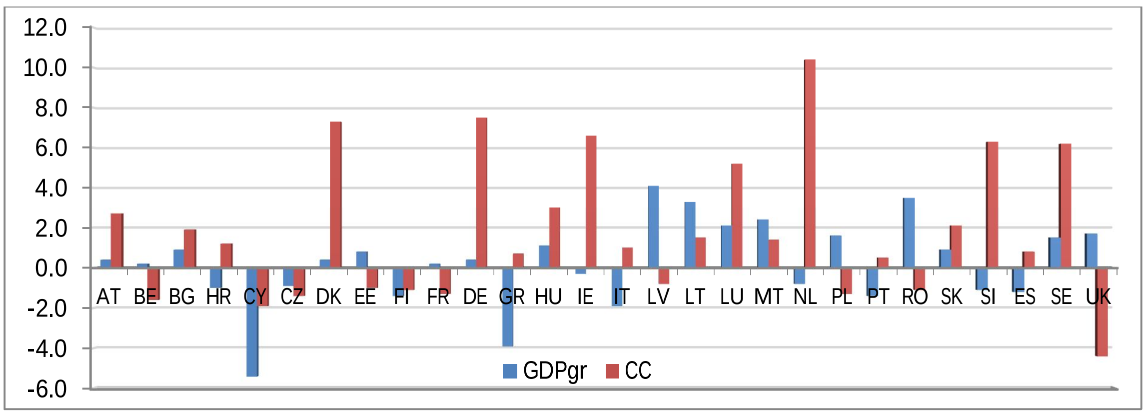

Fig. 1. GDP growth and current account levels for the E.U. countries in 2013

Regarding labour productivity, as presented in fig. 2, Bulgaria, Romania, Latvia and Poland are the four European Union countries that registered the lowest levels in this respect, with ratios of real GDP to employed population below 20. Spain, United Kingdom, Belgium, Italy, Sweden, Netherlands and France were above E.U. 28 average in terms of labour productivity, while Luxemburg is by far the most productive country of the European Union, with a ratio of around 155, based on its high GDP level and low number of employed population.

When considering the unemployment rate levels, Austria (4.9%), Germany (5.3%), Luxemburg (5.8%), Malta (6.5%) and the Netherlands (6.8%) are in top 5 with the lowest unemployment rates, while at the opposite pole are Croatia (17.2%), Spain (26%) and Greece (27.3%). Romania is in top 10 countries, with an unemployment rate of only 7.3% in 2013.

European Journal of Business and Social Sciences, Vol. 3, No. 4 , pp 249-256, **July 2014.** **P.P. 249 - 256**
URL: *http://www.ejbss.com/recent.aspx*
ISSN: 2235 -767X

In order to check for the differences in the economic performance of the 28 E.U. countries, a Hierarchical cluster analysis was applied, using an unsupervised learning method that assigns a set of observations into subsets (called clusters) based on their similarities. The cluster technique was built on the between groups linkage cluster method, whereas the intervals were calculated using the squared Euclidean distance. Based on it we were able to classify the E.U. members into two distinct groups using the values of the year 2013 of the following variables: economic growth, current account, labour productivity, unemployment rate and real net earnings.

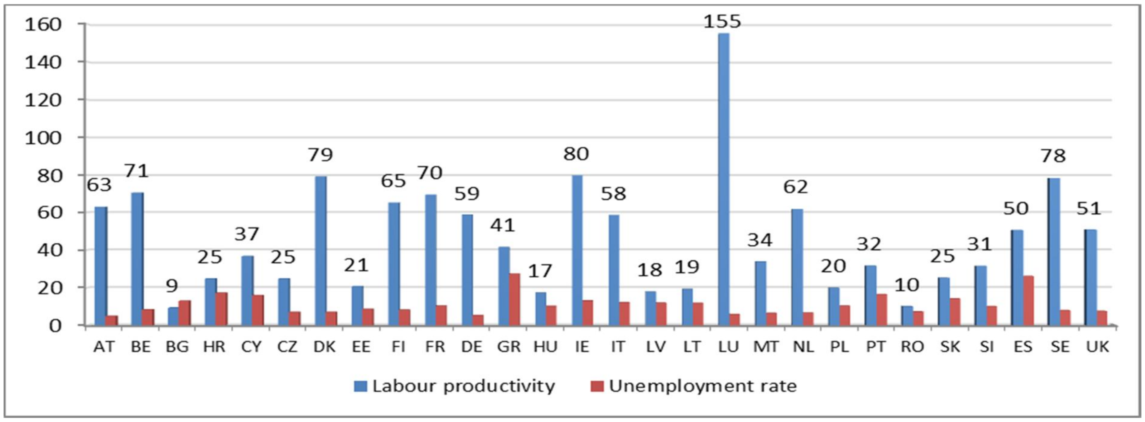

Fig. 2. Unemployment rate and labour productivity for the E.U. countries in 2013

After applying a Hierarchical cluster analysis, two main E.U. country clusters were obtained, according to the dendrogram presented in fig.3:

- *A cluster with high economic performance countries:* Austria, Denmark, Netherlands, Germany, Finland, Belgium, Luxemburg, Sweden, France, United Kingdom, Ireland, Italy and Spain.
- *A cluster with lower economic performance countries:* Latvia, Lithuania, Bulgaria, Romania, Croatia, Cyprus, Malta, Hungary, Poland, Estonia, Slovakia, Czech Republic, Portugal, Slovenia and Greece.

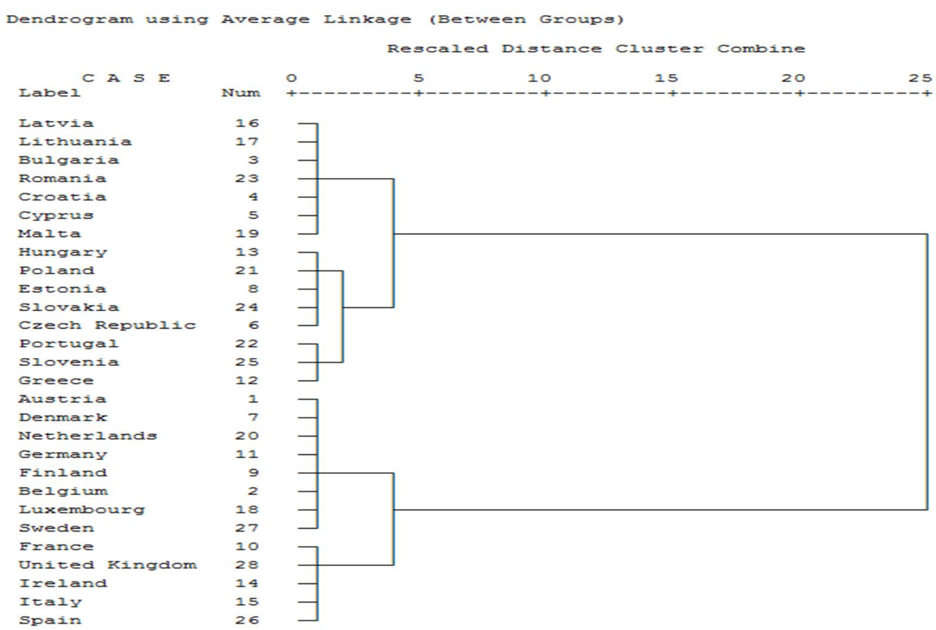

Fig. 3. Dendrogram

3. The evaluation method based on CHAID decision tree models

Based on the two clusters, we then focused on building a classification model in order to correctly classify the 28 E.U. countries into high and lower economic performance countries.

For that we built several decision trees, which are predictive models build in the process of learning from instances and can be viewed as a tree (Andreica and Tapus, 2008). Each branch of the tree is a classification question and the leaves of the tree are partitions of the dataset with their classification.

Because of their tree structure and ability to easily generate consistent rules for segmentation of the original database, decision trees can become an efficient method for classification.

There are a lot of useful decision tree algorithms, out of which the Chi-square Automatic Interaction Detector (CHAID) has the advantage of generating non-binary trees (Andreica, 2009).

CHAID model finds the pair of values that is least significantly different with respect to the target attribute. The significant difference is measured by the p-value obtained from a Pearson chi-square test. For each selected pair, CHAID checks if p-value obtained is greater than a certain merge threshold. If the answer is positive, it merges the values and searches for an additional potential. In this case, the two alpha levels: α_{merge} and α_{split} values were set at a 10% level.

Three CHAID models were built in this study, based on the macroeconomic indicators, not only in order to determine the variables that can be best used as predictors of high versus lower economic performance countries, but also in order to determine consistent classification rules.

The results indicated that the three decision trees built in this study have similar characteristics. More precisely, they all have two layers and have split just one time, indicating that there is only one variable that is relevant to classify the 28 E.U. countries into high and lower economic performance countries. This can be justified by the low number of observations that were used in order to build the decision trees.

The first CHAID decision tree model is presented in fig. 4 and was built assuming the GDP growth rate to be the best predictor. Since a decision tree generates a rule for each of its leaves, there are two classification rules based on the values of the GDP growth rate.

More precisely, the decision tree assigns an E.U. member state to the high economic performance group in case the GDP growth rate is higher than 0.8%. In the other case, the country is considered to have low economic performances. It is therefore obvious that these rules are very sensitive to the initial data set.

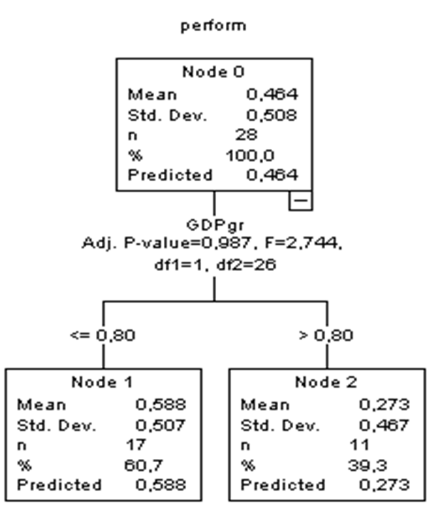

Fig. 4. CHAID model using GDP growth

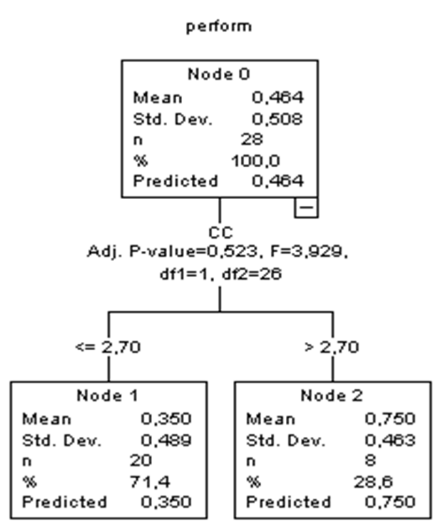

Fig. 5. CHAID model using Current Account

European Journal of Business and Social Sciences, Vol. 3, No. 4 , pp 249-256, **July 2014.** **P.P. 249 - 256**
URL: *http://www.ejbss.com/recent.aspx*
ISSN: 2235 -767X

For the second CHAID decision tree model, presented in fig. 5, it was assumed that the Current account balance would be the best predictor of economic performance. This time, the classification rules indicated that an E.U. member state is considered to have high economic performance in case the current account balance is higher than 2.7%, while for a value of the current account balance below 2.7% the country would be considered to have low economic performances.

The last CHAID decision tree model was built assuming that labour productivity would be the best predictor of economic performance for the 28 E.U. countries (see fig. 6). This decision tree model has the particularity that although it also has two layers and has split one time, in comparison to the other two decision tree models this one has 5 terminal nodes, indicating that the classification rules are more rigorous and sensitive.

More precisely, the classification tree indicates that if an E.U. country has a labour productivity of 36.7 or higher, then the country will be considered to have high economic performances. On the contrary, in case of a lower labour productivity, the country will be considered less productive. However, the presence of three more terminal nodes of the decision tree indicates that out of the 14 countries considered to have lower economic performances, there are four subgroups that can be identified, based on their similarities and on the level of the labour productivity (less than 19.2; between 19.2 and 24.8; between 24.8 and 31.5 and between 31.5 and 36.7). This information can be of real use when determining the prediction ability of the decision tree, since in general prediction errors are to be expected for these exact cases, where the classification is less obvious.

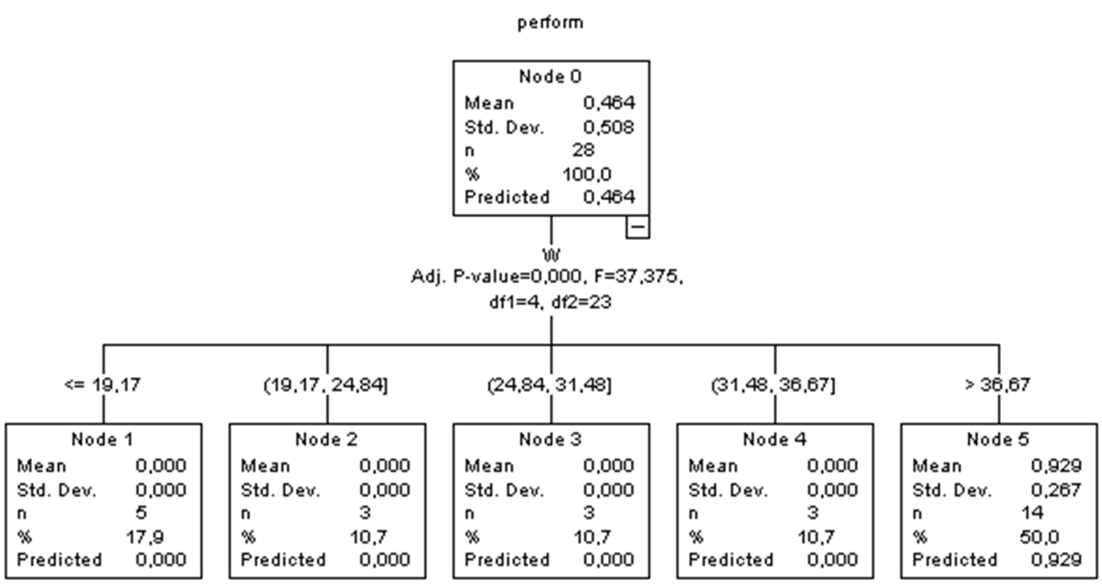

Fig. 6. CHAID model using Labour Productivity

In order to measure the decision tree models' efficiency, the prediction values for each country were calculated and then compared to the original ones. We counted the number of correct and incorrect matches separately for both the high and low performance country clusters and determined the general prediction ability of the three decision tree models, as presented in table 1.

The decision tree model based on the GDP growth rate was the least efficient one, with a prediction ability of only 64.3% after having incorrectly predicted 47% of the low performance countries.

The CHAID model based on current account balance had a similar prediction ability of only 67.9%, but has incorrectly predicted 54% of the high performance countries.

The most efficient CHAID model turned out to be the one based on the labour productivity indicator, for which the prediction ability was of 96.4%.

Table 1. Results of CHAID models performances

	CHAID MODEL based on GDP GROWTH			CHAID MODEL based on CURRENT ACCOUNT			CHAID MODEL based on PRODUCTIVITY		
	high performance	**lower performance**	**TOTAL**	**high performance**	**lower performance**	**TOTAL**	**high performance**	**lower performance**	**TOTAL**
Total	13	15	*28*	13	15	*28*	13	15	*28*
incorect	3	7	*10*	7	2	*9*	0	1	*1*
corect	10	8	*18*	6	13	*19*	13	14	*27*
% incorect	23.1	46.7	*35.7*	53.8	13.3	*32.1*	0.0	6.7	*3.6*
% corect	76.9	53.3	**64.3**	46.2	86.7	**67.9**	100.0	93.3	**96.4**

Source: authors' calculations

4. Conclusions

In this paper we propose a method for improving the evaluation process of economic performance that was applied for the case of the 28 E.U. countries, using indicators of economic growth, current account, labour productivity, unemployment rate and real net earnings for the year 2013. We also applied a Hierarchical Cluster Analysis and identified two main E.U. country clusters, one corresponding to high economic performance countries and the second one corresponding to the E.U. countries with lower economic performances registered in 2013. Based on this classification, we proposed a method based on CHAID classification trees. The decision tree models were tested for their prediction ability in order to determine the most efficient CHAID model that can correctly classify the 28 E.U. countries into high and lower economic performances countries.

Based on the models' prediction ability, we concluded that the most efficient CHAID model was the one based on the labour productivity indicator, with a prediction ability of 96.4%, as compared to the decision models based on GDP growth rate or on current account balance.

5. Acknowledgment

This paper was co-financed from the European Social Fund, through the Sectoral Operational Programme Human Resources Development 2007-2013, project number POSDRU/159/1.5/S/138907 "Excellence in scientific interdisciplinary research, doctoral and postdoctoral, in the economic, social and medical fields - EXCELIS", coordinator The Bucharest University of Economic Studies.

References

1. Altman, E I. (1968), "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", *Journal of Finance*, 23, pp. 589-609.
2. Andreica, M.E. (2013) “Early warning models of financial distress. Case study of the Romanian firms listed on RASDAQ”, *Theoretical and Applied Economics*, Vol. XX (2013), No. 5(582), pp. 4-12
3. Andreica, M.E. (2009) "Predicting Romanian Financial Distressed Companies", *Advances in Economic and Financial Research - DOFIN Working Paper Series 37*, Bucharest University of Economics, Centre for Advanced Research in Finance and Banking - CARFIB
4. Andreica M.I., Tapus, N. (2008) "Reliability Analysis of Tree Networks Applied to Balanced Content Replication", *Proceedings of the IEEE Int. Conf. on Automation, Quality and Testing, Robotics*, pp. 79-84.
5. Jain, B.A., Nag B.N. (1998), „A neural network model to predict long-run operating performance of new ventures”, *Annals of Operations Research* 78, pp.83 – 110
6. Koyuncugil, A. S. Ozgulbas, N. (2007), “Detecting financial early warning signs in Istanbul Stock Exchange by data mining”, International Journal of Business Research
7. Lazar, I. (2010) “Estimation of Real Convergence within the European Union using Sigma and Beta Convergences”, *Revista Tinerilor Economişti*, VII, vol. 1, no. 14 S, pp. 165-175.
8. Lazar, I., Velciu, M., Drăgoiu, C. (2012) “Distribution of Tax Burdens in the Member States of Euro Area”, *Calitatea - Acces la Succes*, S3, pp. 916 - 921;
9. Matei, M., Zamfir, A., Lungu, E. (2014) “A Nonparametric Approach for Wage Differentials of Higher Education Graduates”, *Procedia - Social and Behavioral Sciences*, Vol. 109, pp. 188–192.
10. Ohlson, J.A. (1980), "Financial ratios and the probabilistic prediction of bankruptcy", *Journal of Accounting Research*, 18, pp. 109-131.
11. Rahman, J., “Current Account Developments in New Member States of the European Union: Equilibrium, Excess and EU-Phoria”, *IMF Working Paper* WP/08/92, 2008.
12. Shumway, T. (2001), "Forecasting bankruptcy more accurately: A simple hazard model", *Journal of Business,* 74 (1), pp. 101-124.
13. Spahn, P.,(2013) “Competitiveness, Adjustment and Macroeconomic Risk Management in the Eurozone”, *ROME Discussion Paper Series*, No 2013-16
14. Tam, K.Y., Kiang, M. Y. (1992), „Managerial applications of neural networks: The case of bank failure predictions”, *Management Science* 38926 – 947.
15. Yim J., Mitchell, H. (2005), “ A comparison of corporate distress prediction models in Brazil: hybrid neural networks, logit models and discriminant analysis”, Nova Economia Belo Horizonte 15, 73-93
16. Zheng, Q., Yanhui, J., (2007) “Financial Distress Prediction on Decision Tree Models”, *IEEE*.