DECISION SUPPORT SOLUTION TO BUSINESS FAILURE PREDICTION

Madalina Ecaterina POPESCU[1]
Marin ANDREICA[2]
Ion-Petru POPESCU[3]

ABSTRACT
This paper aims to develop a practical decision support solution to business failure prediction, as early warning signals of potential financial distress could become a true asset in the decision making process of a firm. Several prediction models, such as decision trees and neural networks are built on a sample of Romanian firms and tested for their prediction ability. In order to try to improve the prediction ability of the tree model, we propose a method based on principal component analysis. The high prediction accuracy of the models suggests that the proposed decision support solution can become a practical tool for any decision maker.

KEYWORDS: *decision support solution, financial distress, prediction, CHAID trees, neural networks*

JEL CLASSIFICATION: *G32, C61*

1. INTRODUCTION

Under economic instability, more and more companies struggle with financial difficulties and require a sound strategic planning and an efficient management system. The growing interest in the recent years for elaborating strategic management in a company has generated an increased focus on the development of effective decision support solutions (Tudor et al., 2015; Popescu, 2015; Delcea et al. 2013a, Delcea et al. 2013b) that could turn out to be extremely useful for a company in overcoming the challenges of the economic environment.
In this context, a decision support solution to provide early warning signals of potential financial distress with at least one year before actually turning to insolvency or bankruptcy could become a true asset in the decision making process of a firm.
Therefore, this paper aims to develop a practical decision support solution for business failure prediction. The case of Romanian firms is taken into consideration, as financial ratios for the year 2013 were collected for 346 Romanian firms that were randomly selected from the AMADEUS database of Bureau Van Dijk.
The selected financial ratios reflect the company's size, profitability, solvency and asset utilization. Based on them, several prediction models such as decision trees and neural networks will be built and tested for their prediction ability. In order to try to improve the prediction ability of the tree model, we will propose a method that can be taken into consideration when trying to improve the efficiency of a classification model in the process of decision making. This will imply conducting a Principal Component Analysis and replacing the initial data sample with the principal components of the initial data matrix.

[1] Bucharest University of Economic Studies; The National Scientific Research Institute for Labour and Social Protection, Romania, madalina.andreica@gmail.com
[2] Bucharest University of Economic Studies, Romania, marinandreica@gmail.com
[3] Bucharest University of Economic Studies, Romania, ion.petru.popescu@gmail.com

The paper is structured as follows. Section 2 presents a short literature review in the field of financial distress prediction, while Section 3 is dedicated to the description of the architecture of the decision support solution. Finally, last section concludes.

2. LITERATURE REVIEW

In this section we present a short literature review of the main findings in financial distress prediction. Since the 60s, there has been a growing interest on the topic of bankruptcy and financial distress prediction, especially under conditions of economic instability that can generate serious financial difficulties to a high number of companies, out of which thousands eventually turned into bankruptcy.

The issue of bankruptcy prediction was first tackled by Beaver (1966), who applied a dichotomous classification test through a univariate framework and found that the ratio between Cash flow and Total Debt is the best predictor of bankruptcy. Soon after that Altman (1968) proposed the use of a Multivariate Discriminant Analysis model that was later on intensively used in financial distress prediction. However, several inadequacies of the model were then found through other empirical studies (Eisenbeis, 1977), arguing with respect to the assumptions of normality and group dispersion.

Hence many new approaches have been proposed ever since, in order to predict financial distress more efficiently. For instance, several econometric models were introduced, such as: the logit model (Ohlson, 1980), as well as the hazard models (Shumway, 2001) which are actually multi-period logit models, with identical likelihood functions to a logit model.

Recently, some other heuristic algorithms, such as neural networks and decision trees have also been successfully used to predict bankruptcy. For example, Jain & Nag (1998) provided empirical evidence in favor of neural networks when compared to conventional statistical models (such as discriminant analysis and dichotomous models) in financial distress prediction.

Moreover, hybrid Artificial Neural Network models were introduced in financial distress prediction problems, proving to be very useful in early warning systems for firm failure prediction (Yim & Mitchell, 2005).

However, Koyuncugil & Ozgulbas (2007) and Zheng & Yanhui (2007) argued in favor of CHAID decision trees when compared to neural networks, that are difficult to build and interpret and require more time to accomplish iterative process. On the other hand, CHAID decision trees are easy to build and to interpret and generate clear classification rules.

Based on the literature review in the field, the distress prediction problem remains a true challenge, especially in times of economic instability when each company's surviving skills become crucial. Thus, early warning signals could turn out to be of great help in preventing financial distress or even bankruptcy.

3. DECISION SUPPORT SOLUTION

In this study we draw on previous main empirically based studies on financial distress to decide upon using classification trees and neural networks models for building effective business failure prediction methods. A decision support solution to business failure prediction problem is therefore presented in this paper. The solution was empirically tested for Romanian firms, based on a sample of 173 distressed firms and 173 healthy firms.

The architecture of the proposed decision support solution is further on described, by considering the following main components:

- The database
- The model-based component
- The knowledge management component
- The user interface

Regarding **the database,** financial ratios for the year 2013 were collected for 346 Romanian firms that were randomly selected from the AMADEUS database of Bureau Van Dijk.
Since no standard definition is currently provided by the literature review for distressed firms, we followed the classification criteria provided by the AMADEUS database, where a distressed company could either be an inactive, dissolved, in liquidation or even bankrupt.
Fifteen financial ratios were collected from the AMADEUS database, reflecting the company's profitability, solvency, asset utilization and size, as presented in table 1.

Table 1. The financial ratios

CATEGORY	CODE	FINANCIAL RATIOS	DEFINITION
Profitability	**I1**	Profit Margin	(Profit before tax / Operating revenue)*100
	I2	Return on Assets	(Profit before tax / Total assets)*100
	I3	Return on Equity	(Profit before tax /Shareholders funds)*100
	I4	Profit per employee	Profit before tax / Employees
	I5	Operating revenue per employee	Operating revenue / Employees
	I6	Cash flow / Operating revenue	(Cash flow / Operating revenue) * 100
Solvency	**I7**	Current ratio	Current assets / Current liabilities
	I8	Liquidity ratio	(Current assets - Stocks)/Current liabilities
	I9	Solvency Asset based	(Shareholders funds/Total assets) * 100
Asset utilization	**I10**	Shareholders funds per employee	Shareholders funds / Employees
	I11	Total Assets per employee	Total Assets / Employees
	I12	Net assets turnover	Operating revenue / (Shareholders funds + Non current liabilities)
Size	**I13**	Company size	ln (Total Assets)
	I14	Average cost per employee	Cost of employees / Employees
	I15	Cost of employee / Operating revenue	Cost of employees/Operating revenue)*100

Source: authors own computation

Regarding **the model-based component**, both classification trees and neural network models were implemented in order to predict business failure.
A classification tree is a predictive model viewed as a tree and built in the process of learning from instances. Each tree's branch is associated with a classification question, whereas each tree's leave represents a partition of the dataset according to the classification (Zamfir et al., 2017).
Among the types of decision tree algorithms, the Chi-square Automatic Interaction Detector (CHAID) was selected for this empirical study, as it has the advantage of generating non-binary trees. The CHAID algorithm implies finding the pair of values that is least significantly different with respect to the target attribute based on a Pearson chi-square test p-value. For each selected pair, CHAID checks if the obtained p-value is greater than a certain merge threshold. If the answer is positive, it merges the values and searches for an additional potential. The two alpha levels α_{merge} and α_{split} values were set at a 5% level.

The initial sample was divided into a 70% training sample and a 30% test sample. In order to measure the decision tree model's efficiency, both the in-sample and the out-of-sample performances were then calculated and summarized in table 2.

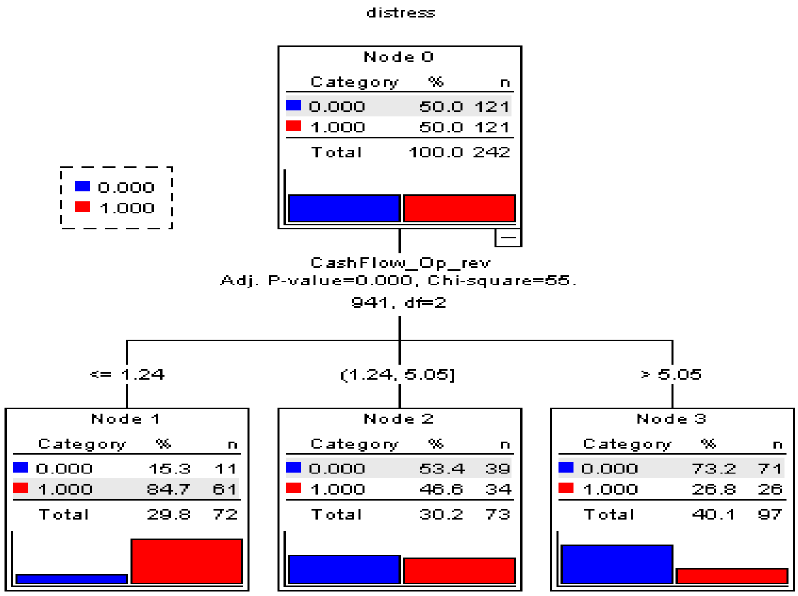

Figure 1. The CHAID tree model based on training sample
Source: authors own computation

The resulted CHAID model has two layers and one split, indicating that there is just one best predictor of financial distress. More precisely, according to the CHAID tree model presented in figure 1, the best way to classify the initial sample into "healthy" and "unhealthy" firms is based on the *Cash flow/ Operating revenue* (I6) indicator.

As noticing, the results indicated a profitability financial ratio to be the best predictor of financial distress, which is consistent with the international literature on the field (Zheng & Yanhui, 2007). The I6 indicator actually compares a firm's operating cash flow to its net revenues, which gives investors an idea of the firm's ability to turn sales into cash. *Cash flow/ Operating revenue* (I6) indicator could thus indicate a warning signal if a firm's sales grow without a parallel growth in operating cash flow.

Next an Artificial Neural Network (ANN) was built by considering as input data all available 15 firms specific financial ratios. The same 70:30 sub-sampling technique was applied to the initial sample in order to train and test the network.

A Hyperbolic tangent activation function was used when building the ANN model, along with a cross-entropy error function. Before feeding the data into the neural network, some variable transformations were also required. For instance, all positive values of the financial indicators were rescaled to the [0,1] range, while all negative values were rescaled between [-1,0]. The ANN was built based on the following structure: one input layer, 1 hidden layer (with 3 neurons) and one output layer, as described in figure 2.

The prediction accuracy of the neural network was then computed for both in-sample and out-of-sample datasets and summarized in table 2, were the total number of correct and incorrect matches was also included.

When analyzing the weights of the 15 financial indicators in the neural network, our findings suggest that the most relevant predictors of business failure have proven to be the following, in this precise order: *Cash flow/Operating revenue* ***(I6)***, *Profit margin* ***(I1)***, *Return on Equity* ***(I3)***, *Operating revenue per employee* ***(I5)***, *Liquidity ratio* ***(I8)*** and *Shareholders funds per employee* ***(I10)***.

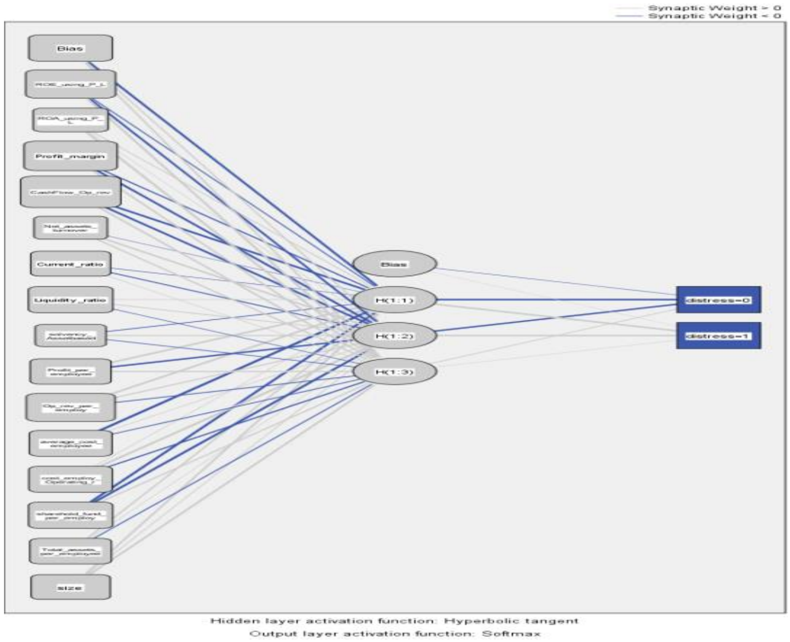

Figure 2. The neural network structure
Source: authors own computation

The financial ratios hierarchy drawn on the ANN weights (see figure 3) suggests once again just how relevant profitability indicators are, followed by solvency and asset utilization indicators in business failure prediction. Moreover, the presence of *Cash flow/Operating revenue* indicator on top of all selected financial ratios that were inputs for the neural network model brings new evidence on just how relevant this indicator truly is in business failure prediction for this empirical study. These results are consistent to previous findings yielded from the CHAID tree model and confirm the indicator's prediction ability.

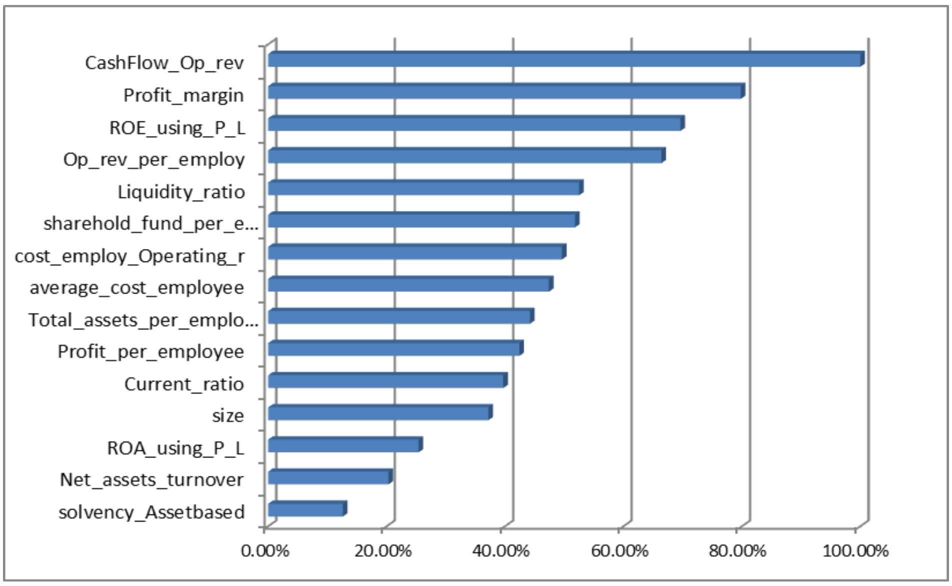

Figure 3. Financial ratios hierarchy based on ANN weights
Source: authors own computation

When computing the prediction ability of the two financial distress prediction models based on both in-sample and out-of-sample data-sets, we notice that the ANN model outperforms the CHAID model. More precisely, the ANN model has a prediction accuracy of 82% in the learning phase and a slight increase to 83.2% in the testing phase, while the CHAID tree model only reaches a modest Our results actually suggest that when using principal components of the initial data matrix, the prediction performance of a CHAID model can improve. However, we are aware that our empirical findings are sensitive to the initial data set and even though the results might not apply to a more general case, we consider that we proposed a method that can be taken into consideration when trying to improve the efficiency of a classification model in the process of decision making.
70.7% prediction accuracy in the training sample and a smooth increase to 71.2% in the testing phase. The results are summarized in Table 2, according to company type and the percentage of correct matches and miss-matches.

In order to try to improve the prediction ability of the CHAID model, we then decided to check whether the prediction ability of the model can improve when replacing the initial data sample with the main principal components (PCs).

Having this in mind, a Principal Component Analysis (PCA) was performed to cumulate the relevance of all of the initially considered financial indicators, based on a reduced dimensionality of the original data space with minimum loss of information.

The first five principal components that have eigenvalues higher than 1 are the following: λ_1=3.1, λ_2=2.5, λ_3=2.1, λ_4=1.5 and λ_5=1.2, with a minimum loss of information of approximately 31%.

According to the Rotated Component Matrix based on the Varimax with Kaiser Normalization we can conclude that:

- First principal component is highly correlated to the following indicators: Cash Flow over Operating revenue, ROE, ROA, Profit margin and Profit per employee and describes a *profitability indicator*;
- 2nd PC is mostly correlated to: Total assets per employee, company's size and Shareholders fund per employee, describing a *size indicator*;
- 3rd PC is a *solvability indicator*, as it is mostly correlated to: Current ratio and Liquidity ratio;
- 4th PC describes an *assets utilization indicator*, as it is highly correlated to: Net assets turnover and Operating revenue per employee
- 5th PC is mostly an *employee costs indicator*, as it is highly correlated to: Average cost of employee, Cost of employee over Operating revenue and Solvency Asset-based.

Once the principal components were computed, a second CHAID model was built having as input data the first principal components.

The CHAID classification tree built on PCs only identified one predictor for financial distress, as only the first principal component describing profitability indicators was selected in the model. We notice that once again, the *Cash flow/ Operating revenue* indicator is part of the 1st PC and confirms just how relevant it is in business failure prediction for the case of Romania.

When comparing the out-of-sample prediction accuracy of the two CHAID models, we found that when replacing the initial data sample with the principal components of the initial data matrix, the prediction performance of the new CHAID model improved and reached an out-sample accuracy of prediction of over 73.1% (see table 2).

Regarding **the knowledge management component**, decision trees turned out to be extremely effective not only in defining best predictors of financial distress, but also in developing consistent classification rules.

Due to their tree structure it is extremely easy to generate classification rules for data units. As decision trees generate one rule for each of their leaves, the first CHAID model indicates there are three classification rules, based on the values of *Cash flow/Operating revenue* indicator. More precisely, the decision tree classifies a firm as being distressed if I6 is less than 1.24%, whereas if I6 is higher than 5.05%, the company is marked as healthy. However, if I6 is between the interval

(1.24 – 5.50], then the matching has to be analyzed with cautious, as the CHAID model tends to overestimate when classifying a company as being healthy under these circumstances.
The second CHAID model that relies only on the first principal component assigns a firm to the distressed group in case the *profitability indicator* is less than -0.093. In the other case, the firm is considered to be healthy. It is however obvious that these rules are very sensitive to the initial data set.

Table 2. The prediction ability of the financial distress models

Sample	Observed	**CHAID tree prediction**			**Neural network prediction**			**CHAID tree prediction using PCs**		
		0	1	% Correct	0	1	% Correct	0	1	% Correct
Training	0	110	11	90.9%	113	8	93.4%	98	23	81.0%
	1	60	61	50.4%	34	78	69.6%	47	74	61.2%
	Total %	70.2%	29.8%	**70.7%**	63.1%	36.9%	**82.0%**	59.9%	40.1%	**71.1%**
Testing	0	47	5	90.4%	48	4	92.3%	41	11	78.8%
	1	25	27	51.9%	12	31	72.1%	17	35	67.3%
	Total %	69.2%	30.8%	**71.2%**	63.2%	36.8%	**83.2**%	55.8%	44.2%	**73.1%**

Source: authors own computation

Regarding **the user interface**, the SPSS statistical software was used in order to compute the financial ratios and to build the decision trees and the neural network models, since it has a friendly user interface and can offer support to solving various decision problems.

4. CONCLUSIONS

Under economic instability, more and more companies struggle with financial difficulties and require a sound strategic planning and an efficient management system. In this context, a decision support solution to provide early warning signals of potential financial distress before actually turning to insolvency or bankruptcy could become a true asset in the decision making process of a firm.
The aim of this paper consisted in developing a practical decision support solution for business failure prediction for the case of Romanian firms. Therefore, financial data were collected for a randomly selected sample of Romanian firms and the effectiveness of several prediction models such as decision trees and neural networks were then tested.
Out of the two prediction methods tested, best out-of-sample results were obtained by the neural network model as compared to the CHAID decision tree model. The prediction accuracy of the neural network was quite high, reaching over 83% in the testing phase, as compared to the CHAID model (71.2%). In order to try to improve the prediction ability of the CHAID model, we applied a Principal Component Analysis and found that when replacing the initial data sample with the principal components of the initial data matrix, the prediction performance of the new CHAID model improved. However, the neural network model still kept the highest prediction accuracy among the tested models.
Our results actually suggest that when using principal components of the initial data matrix, the prediction performance of a CHAID tree model can improve. However, we are aware that our empirical findings are sensitive to the data used in the analysis and even though the results might not apply to a more general case, we consider that we proposed a method that can be taken into consideration when trying to improve the efficiency of a classification model in the process of decision making

Regarding the top best predictors of financial distress, profitability ratios turned out to perform best. More precisely, both models highlighted the importance of the *Cash flow/Operating revenue* indicator when predicting business failure for this empirical study focused on Romanian firms.
Thus, we conclude that our results are consistent with the economic theory and the literature review and the high prediction accuracy of the models suggest that the proposed decision support solution can become a practical tool for any decision maker.

REFERENCES

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, *23*, 589-609

Beaver, W. (1966). Financial ratios as predictors of failure, *Journal of Accounting Research (Supl), 4*, 71-102.

Delcea, C., Bradea, I., Maracine, V., Scarlat, E. & Cotfas, L. (2013a). GM(1,1) in bankruptcy forecasting, *Grey Systems: Theory and Application*, Emerald, 3, 250-265, ISSN 2043-9377.

Delcea, C., Bradea, I. & Scarlat, E. (2013b). A Computational Grey Based Model for Companies Risk Forecasting, *The Journal of Grey System*, 3/2013, vol. 25, 70- 83.

Eisenbeis, R. (1977). Pitfalls in the application of discriminant analysis in business, finance and economics, *Journal of Finance*, *32*, 875-900.

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*, 83-110.

Koyuncugil, A. S. & Ozgulbas, N. (2007). Detecting financial early warning signs in Istanbul Stock Exchange by data mining, *International Journal of Business Research.*

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research, 18*, 109-131.

Popescu, M. E. (2015). Proposal for a decision support system to predict financial distress, *Review of International Comparative Management*, *16* (1), 112-118

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model, *Journal of Business*, *74* (1), 101-124.

Tudor, L., Popescu, M. E. & Andreica, M. (2015). A Decision Support System to Predict Financial Distress. The Case of Romania, *Romanian Journal of Economic Forecasting, 18* (4), 170-179

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

Zamfir, A. M., Mocanu, C. & Grigorescu, A. (2017). Circular Economy and Decision Models among European SMEs, *Sustainability*, 9, 1507, doi:10.3390/su9091507

Zheng, Q. & Yanhui, J. (2007). Financial Distress Prediction on Decision Tree Models, *IEEE.*