

ORIGINAL ARTICLE


Citation: Durica., M, Frnda, J. & Svabova, L. (2019). Decision tree based model of business failure prediction for Polish companies. *Oeconomia Copernicana*, 10(3), 453–469. doi: 10.24136/oc.2019.022

Contact to corresponding author: marek.durica@fpedas.uniza.sk; Department of Quantitative Methods and Economic Informatics, Faculty of Operation and Economics of Transport and Communications, University of Zilina, Univerzitna 1, SK-01026 Zilina, Slovak Republic

Received: 3.01.2019; Revised: 12.08.2019; Accepted: 27.08.2019; Published online: 28.09.2019


Marek Durica

University of Zilina, Slovakia

 orcid.org/0000-0002-2118-166X


Jaroslav Frnda

University of Zilina, Slovakia

 orcid.org/0000-0001-6065-3087

Lucia Svabova

University of Zilina, Slovakia

 orcid.org/0000-0002-4722-4103

Decision tree based model of business failure prediction for Polish companies

JEL Classification: C38; G33

Keywords: *decision trees; prediction model; financial ratios; business failure; Polish companies*

Abstract

Research background: The issue of predicting the financial situation of companies is a relatively young field of economic research. Its origin dates back to the 30's of the 20th century, but constant research in this area proves the currentness of this topic even today. The issue of predicting the financial situation of a company is up to date not only for the company itself, but also for all stakeholders.

Purpose of the article: The main purpose of this study is to create new prediction models by using the method of decision trees, in achieving sufficient prediction power of the generated model with a large database of real data on Polish companies obtained from the Amadeus database.

Methods: As a result of the development of artificial intelligence, new methods for predicting financial failure of the company have been introduced into financial prediction analysis. One of

the most widely used data mining techniques in this field is the method of decision trees. In the paper, we applied the CART and CHAID approach to create a model of predicting the financial difficulties of Polish companies.

Findings & Value added: For the creation of the prediction model, a total of 37 financial and economic indicators of Polish companies were used. The resulting decision trees based prediction models for Polish companies reach a prediction power of more than 98%. The success of the classification for non-prosperous companies is more than 83%. The created decision tree-based prediction models are useful mainly for predicting the financial difficulties of Polish companies, but can also be used for companies in another country.

Introduction

Business failure prediction has been an interesting topic over the recent decades because of its great importance for companies, interested stakeholders and even for the economy of a country. If the prediction of business failure is reliable, managers of companies can initiate remedial measures to avoid financial distress or even bankruptcy. Also, investors can make the company profitable and adjust their investment strategies to reduce anticipated losses. However, the rapid development of the capital market and the globalisation has increased the number of companies that over the years, suffer from financial distress.

The main aim of our research is to create a business failure prediction model for Polish companies using CART and CHAID growing algorithm. For model development, real data of nearly 29,000 Polish companies were used. Therefore, it can be expected that the model reflects the actual specifications of the Polish economy. There is a potential of this model to become a commonly used tool for predicting the business failure of Polish companies. Our model achieves pretty high prediction ability.

The purpose of the paper is to solve the research problem by the formation of a model of the business failure identification, based on the significant financial indicators identified by the growing algorithm CART and CHAID. The main contribution of the paper is the identification of the most significant predictors and modelling of business failure of Polish companies one year in advance. The originality of the article lies in the identification of the significant determinants in actual Polish conditions by modelling of business failure. Models are constructed without regard to any sector, thus eliminating potential financial risks threatening the company, which can be useful not only for companies themselves, but also for all market subjects.

The paper is divided into four main parts. Literature review focuses on the development in the field of business failure prediction models. The Research section depicts a brief description of data on Polish companies — financial indicators used as potential predictors in the prediction model.

This section also specifies the methodology of growing algorithm CART and CHAID. Results section is focused on the description of the created models. Discussion compares and analyses the studies of other authors in the field of decision tree prediction models and compares the prediction ability of our models with other models based on the analysis of classification table and the AUC value.

Literature review

Since the first developed prediction model (Fitzpatrick, 1932, pp. 727–731), there have been numerous pieces of research done. Various predictors have been identified to predict the future situation of companies. As the pioneer model can be denoted Altman model constructed by Multidimensional discriminant analysis (MDA) (Altman, 1968, pp. 589–609), Ohlson LOGIT-type model (Ohlson, 1980, pp. 109–131), and Zmijewski model developed by Probit analysis (Zmijewski, 1984, pp. 59–82; Valaskova, 2018).

These classical statistical techniques are still widely used to create prediction models despite their unrealistic assumptions. Klietnik *et al.* (2018) created MDA-based model for individual Visegrad Group countries and a complex Visegrad model. In 2017, Kovacova and Klietnik created a Slovak national model using LOGIT and Probit application. The authors concluded that their LOGIT-based model outperforms other models constructed in the conditions of Slovakia.

Other statistical methods are not so popular in the field of business failure prediction, because they require more computations or do not act so accurate. Karas *et al.* (2017) studied the possibilities of business failure prediction of agricultural companies. They validated various MDA and LOGIT-based models. The models based on logistic regression exhibited better results than traditional Z-score models based on the MDA approach.

Among the most used data mining techniques, only models based on decision trees (DT) are commonly used by economists and company managers (Lin *et al.*, 2017, pp. 1158–1170). Frydman *et al.* (1985, pp. 269–291) were the first who used DT to predict business failure. They found their DT to be a better predictor of business failure compared with MDA. Since then, many models have been created using these techniques in this field. Karas and Reznakova (2017, pp. 145–154) developed the model based on 29 financial ratios of companies operating in the construction industry. A non-parametric method of CART was used to derive this model. In the study of Chen (2011, pp. 11261–11272), the DT classification methods (C5.0,

CART, and CHAID) and logistic regression (LR) techniques were used to implement the financial distress prediction model for Taiwan listed companies. Brozyna *et al.* (2016, pp. 93–114) used the data about 2006–2012 annual series of 25 financial ratios of 155 banks in the Eurozone to create bank financial distress prediction model for Poland and Slovak banks. Irimia-Dieguez *et al.* (2015, pp. 23–28) dealt with a comparison of CART a Logit financial distress company models. Combination of CART and LOGIT was used in Brezigar-Masten and Masten (2012, pp. 10153–10159). Prediction model generated by the CHAID algorithm was also developed for Romanian companies (Andreica, 2012, pp. 196–200). Business failure prediction of Japanese companies was studied in Aoki and Hosonuma (2004, pp. 299–302). This model was created by CHAID algorithm and achieved the accuracy rate of 91.2%. Many studies compare decision tree-based models with classical statistical methods (mainly MDA and LOGIT) based models and with other data mining techniques based models (Chen, 2012, pp. 65–73; Misankova & Bartosova, 2016, pp. 1260–1269).

In Poland, the issue of business failure models began to develop in the mid-1990s, after the transformation of the economy. Pioneering studies were focused on the use of foreign models for business failure prediction of Polish companies (Prusak, 2018). The first Polish national models were developed by MDA (Holda, 2001; Maczynska, 2004; Hamrol *et al.*, 2004). Then, LOGIT method began to be used (Pisula *et al.*, 2013, pp. 113–133). Of course, many datamining techniques have also been used (Brozyna *et al.*, 2016, pp. 93–114; Zieba *et al.*, 2016, pp. 93–101; Berent *et al.*, 2017, pp. 753–773). Several studies provide an overview and comparison of existing prediction models (Pawelek *et al.*, 2017, pp. 29–42; Prusak, 2018; Tokarski, 2018; Wyrobek & Kluza, 2018, pp. 24–35; Pociecha *et al.*, 2018, pp. 163–172).

Most of the mentioned models were based on a database as large as our model. Moreover, there are often created only for selected economic sectors. Also, several studies were aimed at verifying the functionality of existing models. They proved that these models lose their accuracy when used in another country, or at another time. For these reasons, we would like to create a new prediction model based on a large sample of Polish companies from the recent post-crisis years.

Research methodology

This article presents a business failure prediction model created in specific post-crisis economic conditions of Poland. DTs technique was chosen be-

cause of its good empirical results in previous studies (Karas & Reznakova, 2017, pp. 145–154; Kumar & Ravi, 2007, pp. 1–28). DTs have the following advantages over the two popular classical statistical methods (MDA and LOGIT) and the other popular data mining methods used in this field. DTs work without strong model assumptions about the distribution of variables. During the training process, no parameters have to be selected and optimised. Algorithms generate straightforward decision rules, so, the resulting models are straightforward to interpret and to implement (in comparison with the artificial neural network (ANN) models, etc.). Also, these models achieve at least the same overall accuracy as conventional MDA or LOGIT models (Li *et al.*, 2010, pp. 5895–5904).

This study aims to form a DT-based model in the Polish economic environment to predict business failure. For this reason, we used the Amadeus database; we chose the accounting and financial records of accounting entities operating in Poland in the years 2016 and 2017. The database contains nearly 29,000 Polish companies. The values of 37 financial indicators (not only ratios) were used as predictors (Table 1). Of course, these indicators are not only the most frequently used (Korol, 2013, pp. 22–30), but also those less, but still commonly used, which may take into account the specificities of the Polish economy.

Since the 1st January 2016, insolvency in Poland has been regulated by the texts of the Bankruptcy Law of 28 February 2003 (until 1st January 2016 known as the Bankruptcy and Reorganization Law) and the Restructuring Law of the 15th May 2015 (which is an entirely new legal act published in the Journal of Laws dated 14th July 2015, item 978). These two Acts regulate the situation of companies that are struggling with insolvency — both at an early stage (the threat of liquidity loss) and at its very advanced stage (bankruptcy). The fundamental changes of the Law, include, among other things, changing the definition of debtor insolvency, and threatened with insolvency (Niewczas & Mientkiewicz, 2017).

Under the current wording of the law, a Polish company is insolvent or threatened with insolvency if at least one of the following tests is true (Szymanska-Rutkowska & Galkowski, 2017):

- The balance sheet test: A debtor will be deemed insolvent when the sum of its financial liabilities exceeds the value of its assets, and this situation continues for longer than 24 months. Insolvency will be presumed if, according to the balance sheet, the debtor's obligations exceed the value of its assets, and this situation continues for longer than 24 months.

- The liquidity test: The debtor will be deemed insolvent if it is unable to perform its due pecuniary liabilities. The insolvency will be presumed if a delay in payments exceeds three months.

Using the criteria of this Law, a sample of 28,908 Polish companies was divided into 26,210 non-bankrupt companies and 2,698 bankrupt companies.

For the creation of a business prediction model, two DT generating algorithms were used. The first used algorithm is CART (Classification and Regression Trees) algorithm, developed by Breiman *et al.* (1984). CART creates DT by choosing the variable which provides the best separation of the population (parental node) into two sub-populations (child nodes). Each child node contains the most significant possible proportion of individuals in a single class. This operation is then repeated until no further separation of the companies is possible or desirable (according to stop criterion). These terminal nodes (or leaves) and the set of splitting rules for all the leaves forms the classification model (Kliestik *et al.*, 2018). As a function of impurity, the Gini index was used.

The importance of a variable in a CART tree can be measured for each node of the tree by calculating the impurity reduction (improvement of purity) of the split created by the variable (when this variable has been selected for the split), and then adding up these impurity reductions for all the nodes of the tree. Splitting of nodes was stopped (Kliestik *et al.*, 2018)

- if the depth of the tree has reached a fixed limit of 5 levels of splitting,
- or if the node is pure (all companies belong to the group of bankrupt companies or the non-bankrupt companies group),
- or if the numbers of companies contained in node are less than 100,
- or if the further division of a node would result in the creation of a child with a number of individuals below 50 companies,
- or if the quality of the tree is no longer increasing significantly (minimum change in purity improvement 0.0001 was achieved).

The second use algorithm is CHAID (Chi-square Automatic Interaction Detection). This algorithm was developed by Kass (1980). Unlike CART, CHAID is not binary, and therefore produces trees that tend to be wider rather than deeper. It does not have a pruning function. It could, in some cases, be considered as a weakness of this method, as the resulting tree could be more complicated. When the maximum tree grows, and the stop criteria are reached, the growing stops. When generating a tree, the same stopping criterion as for the CART algorithm was used. The only change was to reduce the maximum tree depth to 3 levels of splitting.

There are many advantages of the DT technique. The first one is that the final tree (or model) is easy for understanding, interpretation and imple-

mentation. A great advantage is also that this technique is fully non-parametric, which means that independent variables may not have a specific probability distribution. Independent variables may be collinear. Decision trees have no problems with outliers that are isolated in small nodes, without a relevant effect on the overall classification. Also, there is no problem with missing values (Kliestik *et al.*, 2018).

On the other hand, the main disadvantage of decision trees is the "divide and conquer" rule used to create the tree. Variables that appear in the first division conditions have much higher weight and affect the impact of other variables in the tree. Even a small change within these variables may, but does not have to, lead to a significant change of the tree itself, and, therefore, its prediction capability. As a disadvantage, we can also mention the fact that a problem may occur with preferring multi-category variables. Another disadvantage is the fact that a relatively wide data sample is needed to create a tree, because otherwise it is threatened by relatively quick overfitting (Weissova *et al.*, 2016).

We suppose that the application of the two techniques mentioned above will develop models that reach the prediction power of at least 80% in the test sample of companies. Furthermore, we expect that the model created by the CHAID technique achieve better prediction. But its structure will be much more complicated than the one of the model constructed using the CART technique.

We evaluated the quality of discriminatory ability of the prediction model based on the classification table analysis and on the ROC curve (Receiver Operating curve). Overall accuracy can be overestimated in the case of data that have been used to model creation. Therefore, we divided the dataset into two samples: the training sample, consisting of 80% of all data, used for generating the model, and the testing sample (remaining 20% of all data) used for calculating of prediction ability or accuracy.

The ROC curve illustrates the behaviour of the created model. The vertical axis shows the percentage of bankrupt companies that have been correctly classified in the bankrupt group, called a true positive rate or sensitivity. The horizontal axis shows the percentage of non-bankrupt companies that have been incorrectly classified in the bankrupt group, which is also called a false positive rate or 1-specificity. If the size of the AUC is close to 1, then the created model has an excellent predictive ability (Kliestik *et al.*, 2018).

Results

Korol (2013, pp. 22–30) claims that DT-based models usually outperform similar models based on conventional tools, including the MDA method or LOGIT or sophisticated ANNs. According to this and many abovementioned advantages of this technique, the failure prediction models for Polish companies in our study were created by DT.

Models were created based on a sample of real 28,908 Polish companies, among them 26,210 non-bankrupt companies and 2,698 bankrupt companies (according to current Polish legislative). To verify the classification accuracy of a model, the whole sample was randomly divided into a training sample (80% of the total number of companies), and the test sample (other 20%). On the training sample, models were created, and then the test sample served to verify the classification ability of the established models.

A growing CART algorithm was set up for the maximum tree depth of 5; the minimum number of cases in parental nodes was set up to 100 and 50 in the child nodes. According to this, the algorithm grows a maximum tree. After that, the algorithm prunes the tree to avoid overfitting concerning achieved overall accuracy. As the impurity function, Gini index with minimum change in improvement 0.0001 was used. As predictors, 37 financial indicators calculated using real financial statements from 2016 (see Table 1) together with company size and NACE indicators were used.

Final CART-based model with three levels of nodes, four non-terminal nodes, and five terminal nodes, was grown (see Figure 1). Algorithm used the values of 23 financial indicators (of all 37) and NACE indicator. However, after the pruning process of the maximum tree that was grown as a first, the final model uses values of 3 predictors only. These three ratios are *X10* (Total Liabilities to Total Assets), *X28* (Return on Equity) and *X30* (Solvency Ratio (Liability based)).

For the practical use of the model, however, it is required that it has sufficient discrimination ability or accuracy. The classification table (Table 2) implies that the created model has pretty high overall accuracy of 97.9% (in the test sample). The model has correctly classified more than 99% of non-bankrupt companies and more than 84% of bankrupt companies. Figure 2 shows a ROC curve whose shape confirms the high predictive ability of the created model. The same results from the AUC value of 0.936.

The relatively high prediction capability is also achieved by the second created model grown by a CHAID algorithm. The same predictors and the same settings were used to create this model. The only change is to set the maximum tree depth to 3 levels of splitting. It is because a CHAID algo-

rithm generates a non-binary tree and the generated tree was not pruned. The model generated by the CHAID algorithm uses the values of 12 financial indicators (X_{30} , X_{28} , X_{03} , X_{37} , X_{21} , X_{20} , X_{36} , X_{09} , X_{35} , X_{18} , X_{24} , X_{01}) and company size and NACE indicators. The final tree is relatively huge, because it contains a total of 57 nodes, of which 36 are terminal nodes.

Table 3 illustrates the accuracy of the model on both the training and the test sample. For the test sample, the overall accuracy is more than 98%. Nearly all of the non-bankrupt companies in test sample are classified correctly. In the group of bankrupt companies, classification ability is more than 83%. Even the shape of the ROC curve illustrates the excellent prediction ability of the created model. The value of AUC is 0.986, which confirms this fact.

Discussion

Comparing created models, we can claim that both of them have high overall accuracy of about 98%. Especially, the correct classification of non-bankrupt companies is excellent (more than 99%). The correct classification of bankrupt companies is about 1% better for the CART-based model (84.4%) than for the CHAID-based model (83.3%). On the other hand, the CHAID-based model achieved 0.04 (or 4%) higher AUC value. However, both AUC values (0.936 vs 0.986) are very high, and both generated models can be considered as really good. Also, the final CART-based model is straightforward, which makes its real implementation very easy. The CHAID-based model is more complicated, but still usable in practical application.

We can compare the models created in this study with similar models created in by other authors. For example, Wyrobek and Kluza (2018) used the technique of Gradient Boosted Decision Trees algorithm to predict bankruptcy of Polish companies one year in advance. Their database covers the years 2008–2017. The resulting model contains 20 variables and achieved the prediction accuracy of 99.11%. These results are a little better, but comparable to our model, which gained 98.3% overall prediction accuracy. Due to the number of variables used, our model is more straightforward, making it easier to apply.

Pociecha *et al.* (2018), in their study, focused on the comparative analysis of the most frequent bankruptcy prediction models. Their database consisted of companies operating in the manufacturing sector in Poland during the years 2005–2009. Very similarly to our research, 35 financial ratios of

the companies were included in the study. For the creation of the prediction model, the authors have applied linear discriminant function, logit model, CART method of classification tree, and MPL neural network. The resulting models, created by the classification tree method, achieve a maximum prediction ability of 83.3%. From this point of view, we consider the models of classification trees created by us to be a little better.

Zieba *et al.* (2016) presented a novel approach for bankruptcy prediction that uses Extreme Gradient Boosting for CART decision trees. To evaluate the quality of the method, the authors used data about Polish companies operating in the manufacturing sector. In the classification process, a set of 64 features of the companies was used. Using the CART classification tree method, they created a model whose AUC is 0.717. By applying boosted trees, they achieved the best prediction results, with these models having an AUC of 0.935 to 0.959. Our CART model presented in this study achieved very similar AUC values of 0.936 and CHAID model has a little better AUC of 0.986.

Brozyna *et al.* (2016) in their paper applied classic linear discriminant analysis, logistic regression, classification trees and the method of nearest neighbours for predicting bankruptcy of logistics sector Polish companies. As predictors, authors used 28 financial indicators of the companies. The CART model for prediction bankruptcy one year in advance achieved a prediction power of 84.2%, and its AUC was 0.83. Our model created using the CART method has a very similar prediction ability, but its AUC is more significant, so we consider it from this point of view to be a stronger model.

Pawelek *et al.* (2017) devoted their study to prediction of bankruptcy using the logit model and the classification tree based on the CART algorithm. The created CART models contain 2 to 4 financial ratios and have 100% sensitivity, i.e. the correct classification of bankrupt companies. Their AUC ranges from 0.951 to 1. However, it is the classification of the training sample of companies, as the authors did not verify the model on the test sample. Our CART model also contains three variables, has a sensitivity of 84.4%, and its AUC is 0.936, but this is a test sample of companies.

Conclusions

Although in recent decades the issue of financial distress prediction has been discussed worldwide, so far, there has no generally accepted business failure prediction model considering the specifics of the Polish economic

conditions and legislation. To fill this gap, new business failure prediction models based on CART and CHAID decision tree algorithm were designed in this study. The proposed models were developed for prediction business failure of Polish companies one year in advance. For creation of these models, real data of nearly 29,000 Polish companies covering the years 2016 and 2017 was used. In 2017, if the financial ratios of the company met the conditions of insolvency identified by the actual Polish legislation, it was considered as a bankrupt company.

The final models achieved a high overall accuracy. Besides, the CART-based model is straightforward, so it is straightforward to interpret. Therefore, it is very easily applicable for business failure prediction also for companies for which we do not have complete accounting data. Models have been evaluated by the analysis of the classification table and by the AUC value. The final models provide excellent prediction ability of 97.9% and 98.2%, respectively.

On the other hand, the results could differ based on the data set. Besides, the proposed models should be tested in the following years to find out the possibilities for construction of the business failure prediction model generally accepted in the Polish economic conditions.

References

- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4). doi: 10.1111/j.1540-6261.1968.tb00843.x.
- Aoki S., & Hosonuma, Y. (2004). Bankruptcy prediction using decision tree. In H. Takayasu (Ed.). *The application of econophysics*. Tokyo: Springer. doi: 10.1007/978-4-431-53947-6_43.
- Berent, T., Blawat, B., Dietl, M., Krzyk, P., & Rejman, R. (2017). Firm's default - new methodological approach and preliminary evidence from Poland. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(4). doi: 10.24136/eq.v12i4.39.
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. New York: Routledge. doi: 10.1201/9781315139470.
- Brezigar-Masten, A., & Masten, I. (2012). CART-based selection of bankruptcy predictors for the logit model. *Expert Systems with Applications*, 39(11). doi: 10.1016/j.eswa.2012.02.125.
- Brozyna, J., Mentel, G., & Pisula, T. (2016). Statistical methods of the bankruptcy prediction in the logistics sector in Poland and Slovakia. *Transformations in Business & Economics*, 15(1).

- Chen, M. (2011). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications*, 38(9). doi: 10.1016/j.eswa.2011.02.173.
- Chen, M. (2012). Comparing traditional statistics, decision tree classification and support vector machine techniques for financial bankruptcy prediction. *Intelligent Automation & Soft Computing*, 18(1). doi: 10.1080/10798587.2012.10643227.
- Fitzpatrick, F. (1932). A comparison of ratios of successful industrial enterprises with those of failed firm. *Certified Public Accountant*, 6.
- Frydman, H., Altman, E. I., & Kao, D. (1983). Introducing recursive partitioning for financial classification: the case of financial distress. *Journal of Finance*, 40(1). doi: 10.1111/j.1540-6261.1985.tb04949.x
- Irimia-Dieguez, A. I., Blanco-Oliver, A., & Vazquez-Cueto, M. J. (2015). A comparison of classification/regression trees and logistic regression in failure models. *Procedia Economics and Finance*, 26. doi: 10.1016/s2212-5671(15)00797-2.
- Karas, M., & Reznakova, M. (2017). Predicting the bankruptcy of construction companies: a CART-based model. *Engineering Economics*, 28(2). doi: 10.5755/j01.ee.28.2.16353.
- Karas, M., Reznakova, M., & Pokorny, p. (2017). Predicting bankruptcy of agriculture companies: validating selected models. *Polish Journal of Management Studies*, 15(1). doi: 10.17512/pjms.2017.15.1.11.
- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29(2). doi: 10.2307/2986296.
- Kliestik, T., Kliestikova, J., Kovacova, M., Svabova, L., Valaskova, K., Vochozka, M., & Olah, J. (2018). *Prediction of financial health of business entities in transition economies*. New York: Addleton Academic Publishers.
- Kliestik, T., Vrbka, J., & Rowland, Z. (2018). Bankruptcy prediction in Visegrad group countries using multiple discriminant analysis. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 13(3). doi: 10.24136/eq.2018.028.
- Korol, T. (2013). Early warning models against bankruptcy risk for Central European and Latin American enterprises. *Economic Modelling*, 31. doi: 10.1016/j.econmod.2012.11.017.
- Kovacova, M., & Kliestik, T. (2017). Logit and Probit application for the prediction of bankruptcy in Slovak companies. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(4). doi: 10.24136/eq.v12i4.40.
- Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – a review. *European Journal of Operational Research*, 180(1). doi: 10.1016/j.ejor.2006.08.043.
- Li, H., Sun, J., & Wu, J. (2010). Predicting business failure using classification and regression tree: an empirical comparison with popular classical statistical methods and top classification mining methods. *Expert Systems with Applications*, 37(8). doi: 10.1016/j.eswa.2010.02.016.
- Lin, W., Ke, S., & Tsai, C. (2017). Top 10 data mining techniques in business applications: a brief survey. *Kybernetes*, 46(7). doi: 10.1108/k-10-2016-0302.

- Misankova, M., & Bartosova, V. (2016). Comparison of selected statistical methods for the prediction of bankruptcy. In T. Loster & T. Pavelka (Eds.). *Conference proceedings of 10th international days of statistics and economics*. Prague: Melandrium.
- Niewczas, B., & Mientkiewicz, D. (2017). Poland. In D. S. Bernstein (Ed.). *The insolvency review*. Retrieved from <https://thelawreviews.co.uk/edition/the-insolvency-review-edition-5/1149945/poland> (29.11.2018).
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1). doi: 10.2307/2490395.
- Pawelek, B., Galuszka, K., Kostrzewska, J., & Kostrzewski, M. (2017). Classification methods in the research on the financial standing of construction enterprises after bankruptcy in Poland. In F. Palumbo, A. Montanari & M. Vichi (Eds.). *Data science studies in classification, data analysis, and knowledge organization*. Springer. doi: 10.1007/978-3-319-55723-6_3.
- Pisula, T., Mentel, G., & Brozyna, J. (2013). Predicting bankruptcy of companies from the logistics sector operating in the Podkarpackie region. *Modern Management Review*, 18. doi: 10.7862/rz.2013.mmr.33.
- Pociecha, J., Pawelek, B., Baryla, M., & Augustyn, S. (2018). Classification models as tools of bankruptcy prediction – Polish experience. In W. Gaul, M. Vichi & C. Weihs (Eds.). *Studies in classification, data analysis, and knowledge organization*. Springer. doi: 10.1007/978-3-319-55708-3_18.
- Popescu, M. E. (2012). Financial distress prediction of the Romanian companies using CHAID models. *Metalurgia International*, 17(12).
- Prusak, B. (2018). Review of research into enterprise bankruptcy prediction in selected central and eastern European countries. *International Journal of Financial Studies*, 6(3). doi: 10.3390/ijfs6030060.
- Szymanska-Rutkowska, K., & Galkowski, S. (2017). The new Polish restructuring law: a “second chance” for businesses. *Emerging Markets Restructuring Journal*, 4.
- Tokarski, A. (2018). The phenomenon of bankruptcy of enterprises in the Polish economy in the years 2008-2015. In E. Lotko, U. K. Zawadzka-Pak, & M. Radvan (Eds.). *Optimization of organization and legal solutions concerning public revenues and expenditures in public interest (Conference proceedings)*. doi: 10.15290/ooolscrepi.2018.30.
- Valaskova, K., Klietnik, T., Svabova, L., & Adamko, P. (2018). Financial risk measurement and prediction modelling for sustainable development of business entities using regression analysis. *Sustainability*, 10(7). doi: 10.3390/su10072144.
- Weissova, I., Siekelova, A., & Kramarova, K. (2016). Modeling of company’s default probability in relation to its credit risk. *Global Journal of Business, Economics and Management: Current Issues*, 6(2). doi: 10.18844/gjbem.v6i2.1378.

- Wyrobek, J., & Kluza, K. (2018). Efficiency of gradient boosting decision trees technique in Polish companies' bankruptcy prediction. In L. Borzemski, J. Świętek & Z. Wilimowska (Eds.). *Advances in intelligent systems and computing information systems architecture and technology: proceedings of 39th international conference on information systems architecture and technology – ISAT 2018*. doi: 10.1007/978-3-319-99993-7_3.
- Zieba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications*, 58. doi: 10.1016/j.eswa.2016.04.001.
- Zmijewski, M. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22. doi: 10.2307/2490859.

Acknowledgements

This research was financially supported by the Slovak Research and Development Agency – Grant NO. APVV-14-0841: Comprehensive Prediction Model of the Financial Health of Slovak Companies, and Faculty of Operation and Economics of Transport and Communications, University of Zilina, Slovakia – Institutional Grant No. 11/PEDAS/2019: Using of selected data mining methods in business failure prediction.

Annex

Table 1. Financial indicators used as predictors

| ID | Method for calculation | ID | Method for calculation |
|-----|--------------------------------------|-----|--|
| X1 | Sales / Total Assets | X20 | Net Income / Sales |
| X2 | Current Assets / Current Liabilities | X21 | Non-current Liabilities / Total Assets |
| X3 | Gross Profit / Total Assets | X22 | Cash / Current Liabilities |
| X4 | Net Income / Equity | X23 | Cash-flow / Current Liabilities |
| X5 | EBITDA / Sales | X24 | Working Capital / Sales |
| X6 | Liabilities / EBITDA | X25 | Current ratio |
| X7 | Net Income/ Total Assets | X26 | Liquidity ratio |
| X8 | Working Capital / Total Assets | X27 | Return on Assets |
| X9 | Operating Profit / Total Assets | X28 | Return on Equity |
| X10 | Total Liabilities / Total Assets | X29 | Shareholder Liquidity Ratio |
| X11 | Current Assets / Total assets | X30 | Solvency ratio (Liability based) |
| X12 | Cash / Total Assets | X31 | Cash-flow / Operating Revenue |
| X13 | Cash-flow / Total Assets | X32 | Net Assets Turnover |
| X14 | Cash-flow / Total Liabilities | X33 | Interest Paid |
| X15 | Current Liabilities / Total Assets | X34 | Gross Margin |
| X16 | Current Assets / Sales | X35 | Profit Margin |
| X17 | Operating Profit / Interest Paid | X36 | Net Current Assets |
| X18 | Stock / Sales | X37 | Working Capital |
| X19 | Cash-flow / Sales | | |

Table 2. Classification table of CART-based business failure prediction model

| Sample | Observed | Predicted | | Percent Correct |
|----------|--------------------|-----------|------|-----------------|
| | | no | yes | |
| Training | no | 20867 | 113 | 99.5% |
| | yes | 353 | 1762 | 83.3% |
| | Overall Percentage | 91.9% | 8.1% | 98.0% |
| Test | no | 5197 | 33 | 99.4% |
| | yes | 91 | 492 | 84.4% |
| | Overall Percentage | 91.0% | 9.0% | 97.9% |

Table 3. Classification table of CHAID-based business failure prediction model

| Sample | Observed | Predicted | | Percent Correct |
|----------|--------------------|-----------|------|-----------------|
| | | no | yes | |
| Training | No | 20893 | 63 | 99.7% |
| | yes | 377 | 1787 | 82.6% |
| | Overall Percentage | 92.0% | 8.0% | 98.1% |
| Test | No | 5240 | 14 | 99.7% |
| | yes | 89 | 445 | 83.3% |
| | Overall Percentage | 92.1% | 7.9% | 98.2% |

Figure 1. Decision tree of the CART-based business failure prediction model

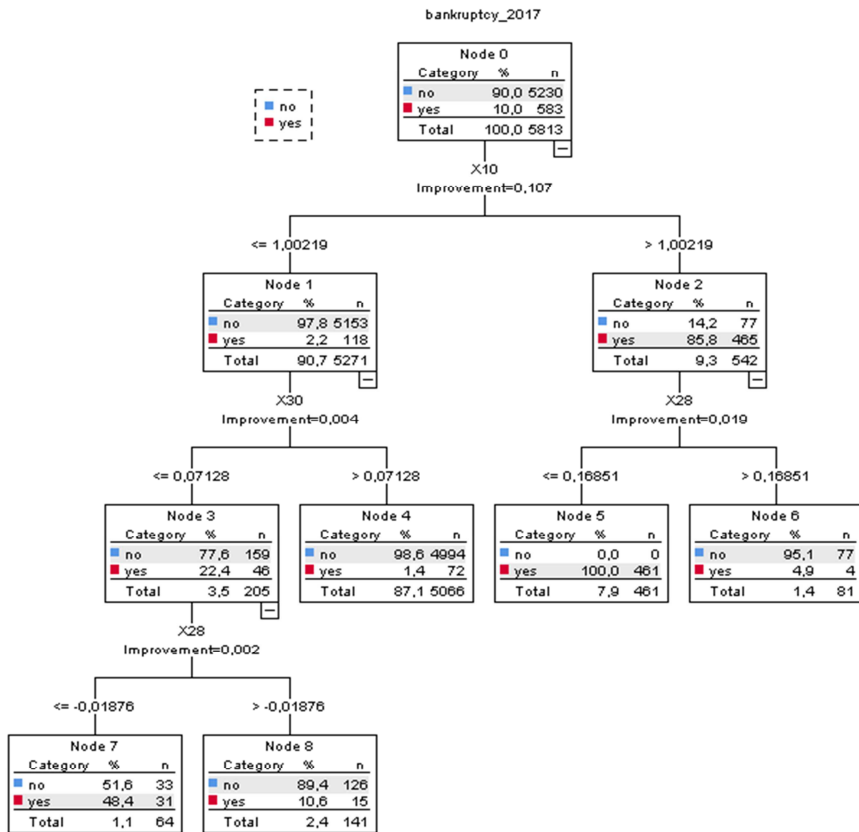


Figure 2. ROC curve of the CART-based business failure prediction model

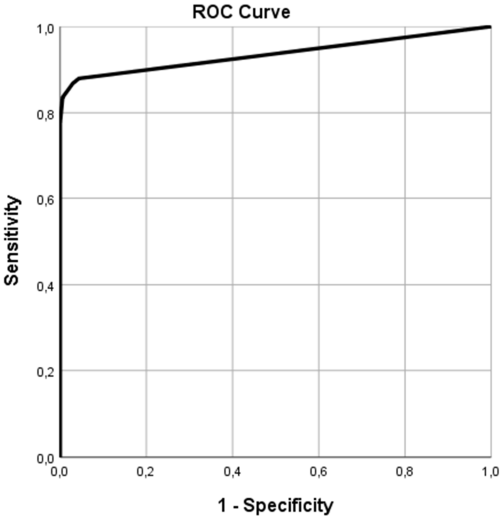


Figure 3. ROC curve of the CHAID-based business failure prediction model

