Ivana Podhorska,
Ph.D., University of Zilina, Slovak Republic
 ORCID ID, 0000-0001-5281-9024
email: ivana.podhorska@fpedas.uniza.sk
Jaromir Vrbka,
Ph.D., Institute of Technology and Business in Ceske Budejovice, Czech Republic
 ORCID ID, 0000-0002-6356-4810
email: vrbka@mail.vstecb.cz
George Lazaroiu,
Ph.D., Associate Professor, The Cognitive Labor Institute, USA
Spiru Haret University, Romania
 ORCID ID, 0000-0002-3422-6310
email: lazaroiu@aa-er.org
Maria Kovacova,
Ph.D., University of Zilina, Slovak Republic
 ORCID ID, 0000-0003-2081-6835
email: maria.kovacova@fpedas.uniza.sk

Correspondence author: ivana.podhorska@fpedas.uniza.sk

INNOVATIONS IN FINANCIAL MANAGEMENT: RECURSIVE PREDICTION MODEL BASED ON DECISION TREES

Abstract. *Issue of enterprise financial distress represents the actual and interdisciplinary topic for the economic community. The bankrupt is thus one of the major externalities of today's modern economies, which cannot be avoided even with every effort. Where there are investment opportunities, there are individuals and businesses that are willing to assume their financial obligations and the resulting risks to maintain and develop their standard of living or their economic activities. The decision tree algorithm is one of the most intuitive methods of data mining which can be used for financial distress prediction. Systematization literary sources and approaches prove that decision trees represent the part of the innovations in financial management. The main propose of the research is a possibility of application of a decision tree algorithm for the creation of the prediction model, which can be used in economy practice. Paper main aim is to create a comprehensive prediction model of enterprise financial distress based on decision trees, under the conditions of emerging markets. Paper methods are based on the decision tree, with emphasis on algorithm CART. Emerging markets included 17 countries: Slovak Republic, Czech Republic, Poland, Hungary, Romania, Bulgaria, Lithuania, Latvia, Estonia, Slovenia, Croatia, Serbia, Russia, Ukraine, Belarus, Montenegro and Macedonia. Paper research is focused on the possibilities of implementation of decision tree algorithm for creation of prediction model in the condition of emerging markets. Used data contained 2,359,731 enterprises from emerging markets (30% of total amount); divided into prosperous enterprises (1,802,027) and non-prosperous enterprises (557,704); obtained from Amadeus database. Input variables for model represented 24 financial indicators, 3 dummy variables and countries GDP data, in the years 2015 and 2016. The 80% of enterprises represented training sample and 20% test sample, for model creation. The model correctly classified 93.2% of enterprises from both the training and test sample. Correctly classification of non-prosperous enterprises was 83.5% in both samples. The result of the research brings the new model for identification of bankrupt of enterprises. The created prediction model can be considered sufficiently suitable for classifying enterprises in emerging markets.*

Keywords: prediction model, decision tree, emerging markets.

Introduction. Nowadays, in a dynamic economic environment and increasing competitive struggle, it is more difficult to maintain its market position and achieve prosperity for enterprises (Konigova, et al.,

2012; Rajnoha and Lorincova, 2015; Afonina, 2015; Zyka and Drahotsky, 2019). For this reason, it is desirable for an enterprise to continually investigate its financial health; to know its strengths and weaknesses; to know the direction to prevent the enterprise from failing. It is precisely the prediction of bankruptcy that has been the subject of many economists around the world for many years. The results of their works are predictive models, by means of which it is possible to predict the financial health of the company for several years in advance. The bankrupt is thus one of the major externalities of today's modern economies, which cannot be avoided even with every effort. Where there are investment opportunities, there are individuals and businesses that are willing to assume their financial obligations and the resulting risks in order to maintain and develop their standard of living or their economic activities. (Fialova and Folvarcna, 2020)

Nowadays, the problem with the failure of business entities is a very actual topic in the economy. One of the biggest business risks is credit risk (Belas et al., 2012; Cipovova and Belas, 2012; Kljucnikov et al., 2016), which relates with secondary insolvency of business entity. The failure of the business entity has a negative influence on all subject with a relationship with this business entity. Possibility of evaluating and predicting credit risk and financial situation of business entities is an advantage for creditors, investors and business owners. Based on the assessment and forecasts of credit risk and financial health of business entities, it was possible to take necessary corrective action, in time. Economists are able to evaluate and predict credit risk and financial health of business entities due to ex-ante financial analysis - through prediction models. Prediction models can early predict the probability of failure of the business entity. The problem which is necessary to solve, for application of prediction models, is the selection of prediction model which could provide the best evaluation of the probability of failure of analyzed business entity and of course, the model which could provide exact and relevant results. Over the years, five basic groups of prediction models have been developed: 1) Statistical models, e.g. one-dimensional analysis, multidimensional discrimination analysis, logit, probit; 2) Mathematical programming, e.g. linear programming, multi-criteria decision-making; 3) Artificial Intelligence, e.g. fuzzy models, decision trees, neural networks; 4) Credit risk models, e.g. Merton's model, KMV model; and 5) Alternative theoretical models, e.g. multi-logit models, cash-flow models, chaos theory and the like. The paper assumption is a possibility of application of a decision tree algorithm for the creation of the prediction model, which can be used in economy practice. The decision tree algorithm is one of the most intuitive and most commonly used methods of data mining, especially since this approach provides explicit classification rules. These rules are very simple and easy to interpret, leading to a quick evaluation of the results. Decision trees deal fairly well with varied or missing data, as well as non-linear effects between variables. When creating predictive models, decision trees are (not only) currently the only major competitor of classical approaches, namely multidimensional discriminant analysis and logistic regression. It should be stressed that decision trees are at the interface between predictive and descriptive method because they create the classification structure of the data set to which they apply. Therefore, they belong to the class of hierarchical grouping methods. A decision tree is a graph structure in the form of a tree that is used to divide a relatively large heterogeneous data set into smaller, more homogeneous subsets by successively applying simple decision rules. This tree structure contains a root node, internal nodes, and terminal nodes. Paper main aim is to create a comprehensive prediction model of enterprise financial distress based on decision trees, under the conditions of emerging markets. Emerging markets included 17 countries: Slovak Republic, Czech Republic, Poland, Hungary, Romania, Bulgaria, Lithuania, Latvia, Estonia, Slovenia, Croatia, Serbia, Russia, Ukraine, Belarus, Montenegro and Macedonia. The countries from emerging markets were chosen for their common economic conditions and similar bankrupt and financial distress reasons, which are described in the literature review. Paper data were obtained from the Amadeus database for 2015 (for all independent variables) and 2016 (for the dependent variable). Data overall contained 2,359,731enterprises from emerging markets; divided into prosperous enterprises (1,802,027) and non-

prosperous enterprises (557,704). They presented the dependent variable. Input variables for model represented 24 financial indicators, 3 dummy variables and countries GDP data. They represented independent variables. The decision tree was constructed in IBM SPSS Statistics 24 using algorithm CART with the top-down method. The tree has always been pruned to prevent it overfitting. The 80% of enterprises represented training sample and 20% test sample, for model creation.

Paper consists of five chapters, namely: 1) literature review which includes the brief historical development of decision trees algorithm; 2) methodology which includes the methodology of decision trees creation and methods which were used in the paper; 3) empirical results which include results of empirical research and prediction model of enterprise financial distress based on decision trees; 4) discussion; 5) conclusions.

Literature Review. The first studies dealing with enterprise bankrupt has been arising since the 30s of the 20th century, during the Great Depression. According to Altman (1993), Smith and Winakor (1935), they were the first to deal with bankruptcy prediction in their studies seriously. Their findings were followed by Merwin (1942). Both studies pointed out that failing firms report significantly different values of selected financial ratios than successful firms. This fundamental principle was a huge breakthrough and offered considerable prospects for further research. Among the studies focused on the issue is also the work of Fitzpatricka (1932) dealing with significant differences between successful and failed businesses. This work has inspired many applied studies that began in the mid-1960s. Among the most important are the studies: Beaver (1966), Altman (1968), Edmister (1972), Tamarih (1976), Springat (1978), Ohlson (1980), Beerman (1982), Fulmer (1984), Zmijewski (1984), Bathory (1984), Argenti (1985), Zavgrenen (1985), Jones (1987), Salchenberger (1992), Shirat (2002), and others. The studies are described in detail in, for example, research articles. (Dimitras et al., 1996; Kumar and Ravi 2006; Bellovary et al., 2007; Balcaen and Ooghe 2004). The issue of enterprise default, financial distress, insolvency and bankrupt is still a managerial challenge to an interdisciplinary problem. Many authors have tried to find basic reasons for this negative economic situation. Enterprise is in bankrupt if it is insolvent. It means it is impossible to pay its two commitment at least to one creditor 30 days after their maturity. Slattery and Lovett in 1999 described several basic reasons for enterprise bankrupt in emerging markets. They created two basic groups, namely: endogenous and exogenous. Endogenous reasons include poor management and its failures, insufficient financial control, poor working capital management, high expenses, insufficient marketing, inappropriate financial policy of the company. Exogenous reasons include negative changes in market demand, competition, a change in prices of input commodities in an unfavourable direction. Mitroff (2011) defined eight basic reasons for worsening financial health of an enterprise in Emerging markets – economic reasons, information reasons, physical reasons, human resources, image, natural disaster, crime. Newton (2005) considers the causes of the enterprise financial distress in emerging markets to be: the inability to manage cash flow; the low value of equity; absence of a good business plan; determination of unrealistic goals; excessive optimism; the inexperience of management; organizational arteriosclerosis. The first studies devoted to bankruptcy prediction were based on one-dimensional analysis of the financial ratios. These studies simply analyzed financial ratios and compared the results of these indicators in creditworthy enterprises and enterprises in bankrupt. In 1930, the Bureau of Business Research presented a study that analyzed the development of 24 financial ratios from 29 industrial enterprises in bankrupt (Fitzpatrick 1932; Malin 2017).

In 1935, R. F. Smith and A. Winakor verified the BBR study results. They analyzed financial indicators of 183 enterprises in bankrupt, and their results confirmed BBR study (Bellovary et al., 2007). W. H. Beaver was the first economist who used statistical methods for prediction of the financial health of enterprises. In the study «Financial Ratios of Prediction of Failure» (1966), under the certain number of financial indicators the author divided enterprises into the two categories of creditworthy enterprises or bankrupt enterprises (Beaver, 1966; Jones, 1987). In 1968, E. I. Altman created one of the most famous and also

the first bankruptcy prediction model, which is known as «Z-score». This model interconnected the explanatory power of several variables. This model is the basic stone of multiple discrimination analysis (Altman, 1968). Since this time, the number of bankruptcy models has risen. In the 70s of 20ty century, were published 28 studies about prediction models, in 80s of 20ty century were published 53 studies and in 90s of 20ty century were published 70 studies about prediction models. Other researchers who have tried to improve Altman´s multiple discrimination analysis are Deakin (1972), Taffler (1974), Loris (1976), Springate (1983), Fernandez (1988), Neumaier and Neumaierova (1995, 1999, 2000, 2005), Gajdka and Stos (1996), Virag and Hajdu (1996), Chrastinova (1998), Binkert (2000), Gurcik (2002), Sharita (2003) and so on (Virag and Kristof, 2005; Mousavi et al., 2015; Gurcik, 2002; Agarwal and Taffler, 2007; Kubickova, 2015; Zavrgen, 1985). In the 70s of 20ty century, also raised prediction models based on the logistic regression - logit and probit models. The first authors who used logistic regression for prediction the bankrupt of the enterprise were Santomero and Vinso (1977) and Martin (1977). They only analyzed bankrupt of American banks. In 1980, Ohlson analyzed bankrupt of enterprise by logistic regression in general. The result of his model is one value which directly determining the probability of bankrupt of the enterprise. Logit analysis was also analyzed in works of Casey and Bartczak (1985), Zavgren (1985), Pantalone and Platt (1987), Jakubik and Teply (2006), Sajter (2008), Hurtosova (2009), Bredar (2014), Gulka (2016). (Zavrgen, 1985, Ohlson, 1990, Poddig, 1995, Siekelova, 2017, Hiadlovsky and Kral (2014).

A pioneer in the area of probit regression was M. Zmijewski (1984) whit his work «Methodological Issues Related to the Estimation of Financial Distress Prediction Models». Other authors followed up his results, for example, Gloubos and Gramamatikos (1988), Skogsvik (1990), Theodossiou (1991), Boritz and Kennedy (1995), Lennox (1999), Zmijewski (1984), Skogsvik (1990), Gulka (2016), Svabova and Kral (2016), Durica and Adamko (2016). Since the 90s of 20ty century have been created, studies focused on the development of prediction models by neural network methods. The basic stone of this method put Odoma and Sharda in 1990. Later Coats and Fant (1992), Altman and Varetto (1994), Wilson and Shardy (1994), Rudorfer (1995), Alici (1996), Serrano-Cinca (1996), Kivilluoto (1998), Sung et al. (1999), Zhang (1999) Lee (2001), Witkowska (2002), Bellovary et al. (2007). Table 1 captures the historic development of individual methods for prediction model creations. It shows the number of prediction models which were created by individual methods in studies about prediction models creation.

Table 1. Historical development of methods for prediction models creation

	Methods				
	Discrimination analysis	Logit analysis	Probit analysis	Neural networks	Different
1960	2	0	0	0	1
1970	22	1	1	0	4
1980	28	16	3	1	7
1990	9	16	3	35	11
2000 - 2004	2	3	0	4	3
2010 - 2017	5	10	0	15	2
Total	68	46	7	55	28

Source: developed by the authors.

According to data in previous Table 1, the most commonly used method for prediction model creation is discrimination analysis. At the second place are neural networks. However, between neural networks and logit analysis is very thin border – only nine studies. Logit analysis was used in 46 studies about prediction models, and neural networks were used in 55 studies about prediction models. Paper research is focused on the possibilities of implementation of decision tree algorithm for creation of prediction model in the condition of emerging markets. The first computer-implemented algorithm of the decision tree was

founded in 1963 by J. Sonquist and J. Morgan developed the Automatic Interaction Detection (AID) algorithm. It is the first computer-implemented algorithm for generating a decision tree, according to Morgan and Sonquist (1963). The extension of this algorithm is CHAID (Chi-squared Automatic Interaction Detection), which is still one of the most widespread algorithms. CHAID uses both nominal and ordinal variables to model the nominal output variable. Another is Exhaustive CHAID algorithm. The motivation for the creation of the Exhaustive CHAID algorithm was, in particular, shortcomings of the CHAID algorithms, following Biggs et al. (1991).

The ID3 (Interactive Dichotomizer) algorithm is the classic case of an algorithm that generates a tree from top to bottom (TDIDT). These trees are classification trees, i.e. the output and input variables are categorical, stated by Quinlan (1986). The ID4, ID5, and finally by ID5R algorithms are the incremental modification of the ID3 algorithm, stated by Utgoff (1989). The algorithm C4.5 is derived from the ID3 algorithm by the same author. Input variables can be discrete but also continuous, Quinlan(1993). In 1998 Quinlan created a newer version of this algorithm implemented as C5.0 (for Unix and Linus) or See5 (for Windows). Over time, other algorithms have been created, e.g. CART, RECPAM, FACT, GUIDE, BART etc. Finally, the total evolution of decision trees algorithms could be divided into four basic categories – generations. This paper works with algorithm CART, which will be explained detail in methodology. A comprehensive review of decision trees algorithms evolution captures Table 2

Table 2. Review of decision trees algorithms evolution

1st generation	AID THAID CHAID	Morgan and Sonquist (1963) Messenger and Mandell (1972) Kass (1980)	3rd generation	QUEST CRUISE Beyesian CART	Loh and Shih (1997) Kim and Loh (2001, 2003) Chipman et al. (1998) Denison et al. (1998)
2nd generation	CART	Breiman et al. (1984)	4th generation	GUIDE	Loh (2002, 2009) Loh and Zheng (2013) Loh et al. (2015)
	RECPAM	Cianysi et al. (1988) Segal (1988, 1992) LeBlanc and Crowley (1992) Alexander and Grimshaw (1996) Zhang (1998)		CTREE	Hothorn et al. (2006)
	MVPART	De´ath (2002) Su et al. (2004)		MOB	Zeileis et al. (2008)
	ID3 M5	Quinlan (1986) Quinlan (1992)		Random forest TARGET	Breiman (2001) Fan and Gray (2005) Gray and Fan (2008)
	C4.5 FACT	Quinlan (1993) Loh and Vanichsetakul (1988)		BART	Chipman et al. (2010)

Source: developed by the authors.

Methodology and research methods. This chapter provides brief characteristics of research methodology which were used in this paper. The chapter is divided four parts: 1) a general introduction to the methodology of decision tree creation, with emphasis on algorithm CART, which was used in this paper; 2) determination of splitting criterion, with emphasis on Gini index which was used in this paper; 3) methodology of pruning of decision tree; 4) data collection.

The methodology of decision tree creation. For creating a decision tree to classify cases into m classes, it is necessary to define the criterion for selecting the variable and condition that will best divide the cases of individual classes in the given node. Then one can be apply this repeatedly in the individual

«parent» nodes of the tree to create the «child» notes. These nodes are repeatedly divided until the stopping condition is met. The stopping condition of tree generation is usually a combination of multiple rules:

1. Nodes are clean, i.e. they only contain the cases with the same value of the output variable.
2. A predetermined depth limit of a tree (the number of levels of tree divisions) was reached.
3. A fixed maximum number of tree leaves was reached.
4. A minimum number of cases in each node (usually between 75 to 100) was reached, where it is assumed that the nodes cannot be divided anymore.
5. Further division of any node would lead to the creation of child nodes with a lower frequency than a predetermined value.

The decision tree model describes a tree graph consisting of nodes and directed edges. The tree generation begins with a root node, which is sequentially divided into (non-terminal) nodes and (terminal) nodes, the so-called leaves. Each node (apart from the leaves) is linked from top to bottom by directed edges that originate from it, with at least two child nodes. Leaf nodes do not have child nodes. The decision tree was created by CART algorithm. The growth algorithm CART (Classification and Regression Trees), according to Breiman (1984), is one of the most efficient and globally-used algorithms generating decision trees. The CART algorithm is a registered trademark of the company Salford Systems. However, similar names such as CRT and CART are often used for naming. The CART generates binary trees. To find the best splitting of nodes, it uses the Gini index. In addition to this option, CART algorithm makers provided several technical solutions that bring two main benefits: universality and performance. Universality is based primarily on the fact that the number of categories of a dependent variable can be either final or even infinite. While the CART can be used to create classification trees (categorical output variable) as well as regression trees (continuous output variable). For each type of output variable, there is an appropriate criterion for node splitting. Universality also lies in the possibility of processing the missing values by replacing each variable with the same splitting variable or the same pruning variable. The same splitting variable is such a variable that provides approximately the same node purity as the original variable. The same pruning variable is such a variable that distributes subjects in approximately the same way as the original variable. These variables could be used as «substitute» variables, but it is best to use the same pruning variables to ensure tree consistency. The performance of the CART algorithm is primarily secured by its pruning mechanism, which is more sophisticated than in the case of the CHAID algorithm. The maximum tree is designed to continue with the node splitting process until it is possible. The algorithm then derives several nested subtrees by sequential pruning operations, compares them, and then selects the one that has the lowest possible error rate measured by cross-validation. The performance is also related to the absence of arbitrary but fixed thresholds. The examples of that are the significance threshold or the X^2 significance test in case of the CHAID algorithm. Determining this threshold is always difficult because it is necessary to make the best choice between the threshold giving a too bushy tree that does not have enough robustness because it depends too much on the specimen and the threshold giving such a small tree with a lower prediction power. The performance of the CART algorithm also lies in its exhaustive search for all possible divisions of a given node, which guarantees optimal splitting. This search may take a long time, especially when it comes to qualitative variables with a large number of categories because it is necessary to test too many potential divisions (Svabova et al., 2018; Dvorsky et al., 2020).

Splitting criterion. The basic principle of tree growth is to increase the purity of the child nodes. When assessing this purity, the purity rates based on entropy are the most commonly used. In particular, the information gain and relative information gain are used. The selection of the splitting variable in the case of a categorical output variable is carried out by various methods, e.g. entropy, Gini index, information gain or x^2 independence test. This research worked with a categorical output variable representing the

prosperity or non-prosperity of an enterprise and paper worked with CART algorithm. For this reason, Gini index as a splitting variable was used. The use of the Gini index when choosing the optimal splitting variable is basically very similar to the use of entropy. The Gini index of the output variable Y is defined by the relationship:

$$G(Y) = 1 - \sum_{j=1}^{m} p_j^z = 1 - \sum_{j=1}^{m} (\frac{n_j}{n})^z \qquad (1)$$

where p_j- is again the probability of the occurrence of the class; y_i - estimated as the relative frequency of the jth class $p_j = \frac{n_j}{n}$.

Furthermore, analogously to entropy, we determine the expected Gini indices for individual input variables X_i:

$$G(X_i) = \sum_{t=1}^{l} \frac{n(x_{it})}{n} G(x_{it}) \qquad (2)$$

where $G(x_{it})$ - is the Gini index on the set of the cases in the given node for which the variable; X_i - takes the value x_{it}:

$$G(x_{it}) = 1 - \sum_{j=1}^{m} (\frac{n_j(x_{it})}{n(x_{it})})^2 \qquad (3)$$

The variable X_i with the smallest value of the Gini index $G(X_i)$ or with the highest reduction of the given index impurity will be used for branching of the given node:

$$Z_{Gini}(X_i) = G(Y) - G(X_i) \qquad (4)$$

Pruning of decision trees. When creating a decision tree, it is necessary to achieve a detailed description of the relationships within the training set. It leads to the creation of a so-called maximal tree that faultlessly explains the relationships within the training set. However, these relationships are often not generally valid, and therefore, when applying the tree to other data, a high error rate can occur. We say that such a tree is overfitting and it is mostly graphically displayed with an excessive number of branches that are a thin and small number of leaf nodes.

The solution to the problem of overfitting of a decision tree is pruning. One option is to trim a tree already during the creation. In this case, the growth of some branches is prematurely terminated if there is a sufficiently high probability that the cases within this branch belong to the same classification class.

In practice, pruning of an existing maximum tree is more often used. In this case, when pruning a tree, it is assessed how its classification capability will deteriorate. When a tree grows, and when it is pruned, it is possible to use a training set, but more often a training set is used to generate a tree, and a testing set

is used to trim it. From the set of all subtrees, the one that achieves the smallest degree of incorrect classification is selected (Alexander and Grimshaw, 1996; Reitano, 2015).

For creating prediction models, it was necessary to create the database for which were used the financial and statistical indicators from the Amadeus database for 2015 (for all independent variables) and 2016 (for the dependent variable). A comprehensive prediction model based on decision trees was created under the conditions of emerging markets, which include: Slovak Republic, Czech Republic, Poland, Hungary, Romania, Bulgaria, Lithuania, Latvia, Estonia, Slovenia, Croatia, Serbia, Russia, Ukraine, Belarus, Montenegro and Macedonia. However, data from Belarus was removed from the database due to a large amount of missing data. Overall, the database contained 2,359,731 enterprises, which were divided into prosperous enterprises (1,802,027) and non-prosperous enterprises (557,704). For verification of the model, classification ability was enterprises in database divided into a training and test sample in the ratio of 80%/20%. The training sample was used to derive the discriminatory function, and the test sample was then used to verify the ability of the model to be classified.

Results. Paper main aim was to create a comprehensive prediction model of enterprise financial distress based on decision trees, under the conditions of emerging markets. This chapter provides characteristics of research results which were obtained in this paper. The chapter is divided into two parts: (I) model indicators; (II) data analysis.

Model indicators. Model input indicators (independent variables) represented quantitative and qualitative variables. Quantitative variables included 24 financial-economic indicators for every enterprises and country GDP. Qualitative variables (in model marked as dummy variables) included country data, the size of enterprises (small, medium and large according to Amadeus database criteria), and sector according to NACE classification. Model output indicator (independent variable) was defined by prosperity and non-prosperity of the enterprise.

For the determination of the independent variables to be used for the creation of the predictive bankruptcy model, was focused on indicators determined by the prominent author as key predictors of financial health. For this purpose, were used studies and research of Sharifabadi et al. (2017), Tian et al. (2015), Bellovary et al. (2007), Ravi Kumar and Ravi (2007), Dimitras et al. (1996), Agarwal et al. (2018) and Ciszewski and Nowakowski (2018). However, two studies were mainly considered, namely Bellovary et al. (2007) «A Review of Bankruptcy Prediction Studies» 1930 to present, and Ravi Kumar and Ravi (2007) «Bankruptcy Prediction in Banks and Firms through Statistical and Intelligent Techniques – A Review». The authors of the first study analyzed 165 prediction models until 2004. They state that 752 different variables were used in the models, with up to 674 of these variables being used in only one or two models. After the study, they present 42 variables which were used in more than five models. The authors of the second study followed the previously published studies by Calderon, et al. (2002), Dimitras et al. (1996), O´Leary (1998) and Scott (1981), Afonina (2015). They completed these studies by their research. Together they analyzed 62 prediction models. Based on the completed analysis, Table 3 presents indicators selected for the model. Given the lack of required data in individual variables of individual countries, some variables from further investigation had to be discarded. These are financial indicators X03, X05, X06, X13, X14, X17, X19, X23, X28, X29, X31, X32, X33, X34.

Besides, for the construction of the model, data from Belarus were not used, since most of them were not available in the database, and, therefore, it was not possible to include this country in the creation of this model. Subsequently, enterprises for which the value of the dependent variable was not listed, also had to be excluded. Because it was not possible to determine whether an enterprise is or is not prospering. Subsequently, enterprises for which the value of the dependent variable was not listed, also had to be excluded. The reason was the impossibility to determine whether an enterprise is or is not prospering.

The decision tree was constructed in IBM SPSS Statistics 24 using the CART algorithm. The created tree has been pruned to prevent it learned and to ensure clarity and easy to the interpretation of its results.

Table 3. Model input variables (independent variables)

Indicator	Calculation Algorithm
X01	Sales/Total Assets
X02	Short-term Assets/Short-term Liabilities
X03	Gross profit/Total Assets
X04	Net income/equity
X05	EBITDA/Sales
X06	(Long-term + Short-term Liabilities)/EBITDA
X07	Net Income/Total Assets
X08	Working Capital/Total Assets
X09	Operating profit/Total assets
X10	(Long-term + Short-term Liabilities)/Total Assets
X11	Short-term assets/Total assets
X12	Cash and Cash Equivalents/Total Assets
X13	Cash flow/Total Assets
X14	Cash flow/(Long-term + Short-term Liabilities)
X15	Short-term Liabilities/Total Assets
X16	Short-term Assets/Sales
X17	Operating Profit/Interest Paid
X18	Inventories/Sales
X19	Cash flow/Sales
X20	Net income/Sales
X21	Long-term Liabilities/Total Assets
X22	Cash and cash Equivalents/Short-term Liabilities
X23	Cash flow/Short-term Liabilities
X24	Working Capital/Sales
X25	Current ratio
X26	(Short-term Assets - Inventory)/Short-term Liabilities
X27	Return on Assets (profit before tax/total assets)
X28	Return on equity (profit before tax/equity)
X29	Equity/Long-term Liabilities
X30	Equity/(Long-term + Short-term Liabilities) * 100
X31	Cash flow/Operating Revenues
X32	Turnover of Net Assets
X33	Interest Paid
X34	Gross Profit/Operating Income
X35	Profit before Tax/Operating Income
X36	Short-term Assets - Short-term Liabilities
X37	Working Capital

Source: developed by the authors.

The maximum number of tree levels was 7, which proved to be the optimal choice in this case. The maximum size of the parent node was 100 cases. It means, if the tree node contained more than 100 enterprises, it was further divided into child nodes if the maximum number of tree levels has not been reached yet. If the tree node contained lower than 100 enterprises, was no longer divided into child nodes. The minimum size of the child node was 50 cases. It means, the division of the parent node into two child nodes would indicate that one of them would contain lower than 50 enterprises, this division would not take place. For selection and detection of splitting variable was used Gini index.

A minimum node reduction was 0.0001, which was used to the branching of each node. The classification ability of the created decision tree was quantified using a classification table. For verification of the model, classification ability was enterprises in database divided into a training and test sample in

the ratio of 80%/20%. The training sample was used to derive the discriminatory function, and the test sample was then used to verify the ability of the model to be classified.

Variable X10 ((long-term+short-term liabilities)/total assets) was used like the most significant classification (splitting) variable. The breakpoint was 1. In this case, enterprises with the value of X10 lower or equal than 1, were classified as non-prosperous. The classification error at this tree point was 0.5%. This branch of the tree had another splitting variable X28 (return on equity), with breakpoint -0.47830. Enterprises with the value of X28 lower or equal than -0.47830 were definitively classified as non-prosperous. In turn, enterprises with the value of X28 higher than 1 were classified as prosperous with further splitting according to variable X28 again. In this tree branch, are two another splitting variables X30 (equity/(long-term + short-term liabilities)*100) and X36 (short-term assets-short-term liabilities).

Splitting criterion X10>1 also created the second branch of a tree which leads to the classification of an enterprise into a group of non-prosperous enterprises. This classification branch is easier. The classification error in this tree node is also 0.5%. Subsequently, the splitting variable was X4 (net income/equity). The breakpoint is -0.66663. Enterprises with the value of X4 higher than -0.66663 were definitively classified as non-prosperous. On the other side, enterprises with the value of X4 lower or equal than -0.66663 were classified as prosperous with further splitting according to same variable X4 in breakpoint -1.74686.

The classification ability of the created decision tree was expressed using a classification table. The classification table provides the number of correctly and incorrectly ranked enterprises into a group of prosperous or non-prosperous enterprises.

The model correctly classifies 93.2% of enterprises in training sample as well as in the test sample. The correct classification of non-prosperous enterprises is 83.5 % in the training sample as well as in the test sample. Thus, type I error is 16.5% in this model, which can be considered a sufficient classification model ability.

The ROC curve also expresses the classification ability of the created model. The area of AUC under the curve is 0.933. From this point of view, the created prediction model can be considered sufficiently suitable for classification of enterprises to prosperous or non-prosperous in emerging markets.

The decision tree technique has several advantages as well as disadvantages. The first advantage is that the results are expressed as simple and, therefore, very easy to understand explicit conditions for input variables. The resulting model could be easily programmed. Then, the classification of new cases is very fast. A great advantage of this method is also that the technique of the decision tree creation is non-parametric, which means that independent variables may not have a certain probability distribution. These variables may even be collinear. If one of the variables is not suitable for the classification, it will simply not be used in a tree. Besides, the dependence of the output variable on the input variables may be non-linear or even nonmonotonic. The preparation and data selection phases are considerably easier compared to other methods. It is because decision trees have no problems with extreme cases that are isolated in small nodes without a relevant effect on the overall classification.
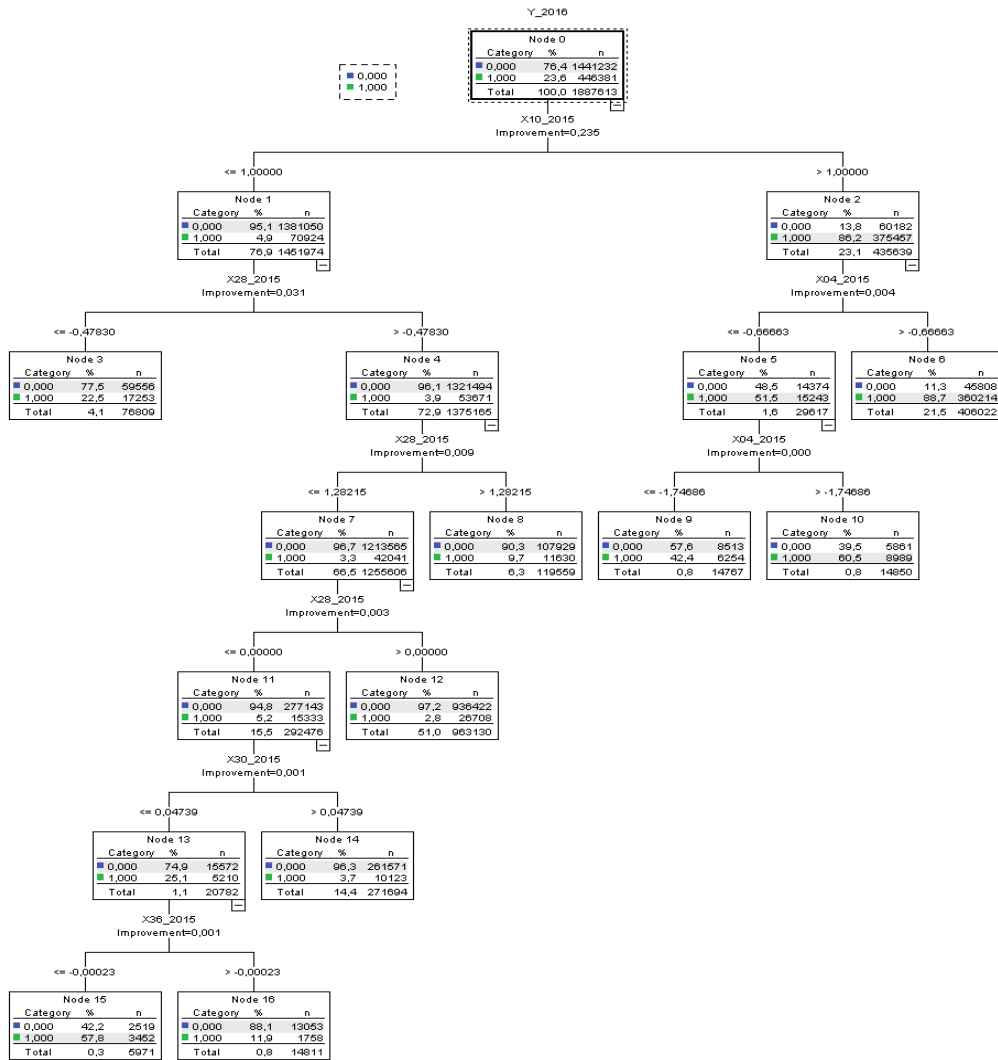
Figure 1. A comprehensive prediction model based on a decision tree model for a training sample of emerging markets
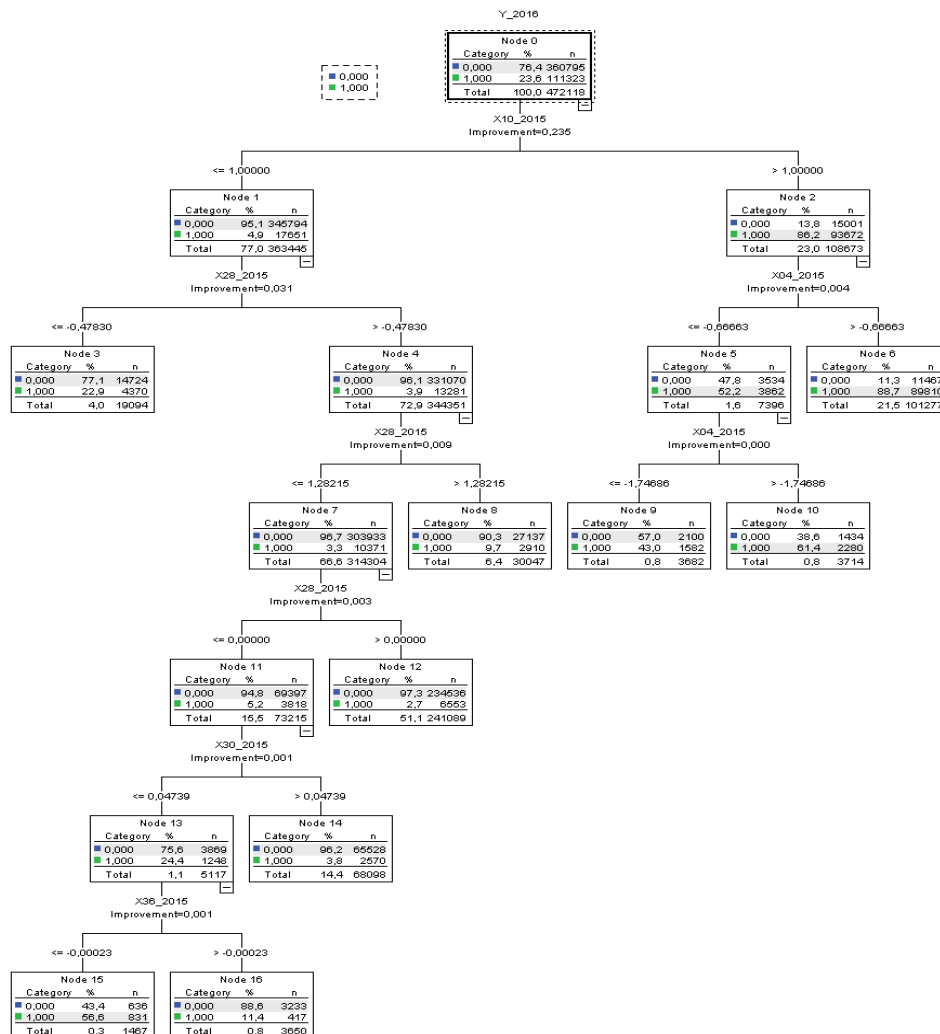
Source: developed by the authors.

Figure 2. A comprehensive prediction model based on a decision tree model for a test sample of emerging markets

Source: developed by the authors.

Most tree classifiers, such as CART and C5.0, can handle any type of variable (continuous, discrete, or qualitative). They also can reasonably cope with missing values, either by creating a new category of a variable with missing values or by using a substitute variable. This research was also addressed by Siekelova et al. (2017), Valaskova et al. (2019) and Kliestikova et al. (2017).

On the other hand, the disadvantage of decision trees is the «divide and conquer» rule used to create the tree itself. Variables that appear in the first division conditions have much greater weight (they separate

a lot more cases) and affect the impact of other variables in the tree. Even a small change within these variables may, but may not, lead to a large change of the tree itself, and, therefore, its prediction capability.

Table 4. Classification table of a prediction model based on a decision tree model for emerging markets

Classification				
Sample		Predicted		
		0	1	Percent Correct
Training	0	1,38,7044	54,188	96.2%
	1	73,726	372,655	83.5%
	Overall Percentage	77.4%	22.6%	93.2%
Test	0	347,258	13,537	96.2%
	1	18,402	92,921	83.5%
	Overall Percentage	77.5%	22.5%	93.2%

Source: developed by the authors.
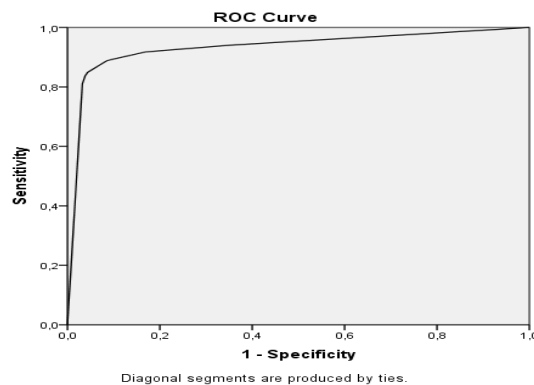
Figure 3. ROC curve of a prediction model based on a decision tree model for emerging markets
Source: developed by the authors.

As a further disadvantage, we can mention the fact that trees detect the actual impact of the variable gradually and not simultaneously. The selection of the splitting criterion is not checked as feedback. It may be a problem, as some trees prefer multi-category variables. This lack of robustness of the decision tree can sometimes be unacceptable and can be overcome by resampling, but that means losing its simplicity. The resulting decision tree models form rectangular areas in the space of variables, which do not have to correspond to the division into cases. The cases that do not have a rectangular distribution are especially difficult to classify. 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. This research was also addressed by Balcerzak et al. (2018), Salaga et al. (2015) and Huxley and Mouwafac (2018).

Conclusion. Issue of enterprise financial distress represents the actual and interdisciplinary topic for the economic community. Nowadays, the problem with the failure of business entities is a very actual topic in the market economy. One of the biggest business risks is credit risk, which relates to secondary insolvency of the business entity. The bankrupt is thus one of the major externalities of today's modern economies, which cannot be avoided even with every effort. Economists are able to evaluate and predict credit risk and financial health of business entities due to ex-ante financial analysis - through prediction models. Prediction models can early predict the probability of failure of the business entity. The paper

assumption is a possibility of application of a decision tree algorithm for the creation of the prediction model, which can be used in economy practice. Paper main aim is to create a comprehensive prediction model of enterprise financial distress based on decision trees, under the conditions of emerging markets.

The decision tree algorithm is one of the most intuitive methods and the most commonly used data mining method. Moreover, that approach provides explicit classification rules. These rules are very simple and easy to interpret, leading to a quick evaluation of the results. The decision tree algorithm could be used as a prediction model for enterprise bankrupt. Paper main aim was to create a comprehensive prediction model of enterprise financial distress based on decision trees, under the conditions of emerging markets. Emerging markets included 17 countries: Slovak Republic, Czech Republic, Poland, Hungary, Romania, Bulgaria, Lithuania, Latvia, Estonia, Slovenia, Croatia, Serbia, Russia, Ukraine, Belarus, Montenegro and Macedonia. However, data from Belarus was removed from the database due to a large amount of missing data. For creating prediction models, it was necessary to create the database for which were used the financial and statistical indicators from the Amadeus database for 2015 (for all independent variables) and 2016 (for the dependent variable). Used data overall contained 2,359,731 enterprises from emerging markets; divided into prosperous enterprises (1,802,027) and non-prosperous enterprises (557,704). They presented the dependent variable. Input variables for model represented 24 financial indicators, 3 dummy variables (country, size of enterprise and NACE classification) and countries GDP data. They represented independent variables. The decision tree was constructed in IBM SPSS Statistics 24 using algorithm CART with the top-down method. The tree has always been pruned to prevent it overfitting. The maximum tree levels were 7. The minimum size of «parent» nodes was 100 cases, and «child» nodes 50 cases. The Gini index was used to selection and calculation of the optimal splitting variable. Model classification ability was quantified using a classification table. The 80% of enterprises represented training sample and 20% test sample, for model creation. The model correctly classified 93.2% of enterprises from both the training and test sample. Correctly classification of non-prosperous enterprises was 83.5% in both samples. The created prediction model can be considered sufficiently suitable for classifying enterprises in emerging markets. Paper findings bring new possibilities for bankrupt enterprise prediction with emphasis on the conditions in Emerging markets. Research aim and purpose have been fulfilled

Author Contributions: conceptualization, I. P.; methodology, I. P. and M. K.; software, J. V and G. L.; validation, I. P. and M. K.; formal analysis, J. V. and G. L.; writing-original draft preparation, I. P. and M. K.; writing-review and editing, M. K.; visualization, I. P.; supervision, G. L.; project administration, I. P.; funding acquisition, M. K.

References

Afonina, A. (2015). Strategic management tools and techniques and organizational performance: Findings from the Czech Republic. *Journal of Competitiveness*, *7*(3). [Google Scholar] [CrossRef]

Agarwal, N., Kwan, P., & Paul, D. (2018). Merger and acquisition pricing using agent based modelling. *Economics, management, and financial markets*, *13*(1), 84-99. [Google Scholar]

Agarwal, V., & Taffler, R. J. (2007). Twenty-five years of the Taffler z-score model: Does it really have predictive ability?. *Accounting and Business Research*, *37*(4), 285-300. [Google Scholar] [CrossRef]

Alexander, W. P., & Grimshaw, S. D. (1996). Treed regression. *Journal of Computational and Graphical Statistics*, *5*(2), 156-175. [Google Scholar]

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

Altman, E. I. (1993). *Corporate Financial Distress and Bankruptcy.* 2nd ed. NewYork: John Wiley & Sons, Inc.

Arvidson, M. (2017). Operationalizing Transparency: Perspectives from the Third Sector in a Mixed Economy of Welfare. *Journal of Self-Governance and Management Economics*, *5*(1), 7-24. [Google Scholar]

Balcaen, S., & Ooghe, H. (2004). *35 Years of Studies on Business Failure: An Overview of the Classical Statistical Methodologiesand Their Related Problems.* Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium. [Google Scholar]

Balcerzak, A. P., Kliestik, T., Streimikiene, D., & Smrcka, L. (2017). Non-parametric approach to measuring the efficiency of banking sectors in European Union Countries. *Acta Polytechnica Hungarica*, *14*(7), 51-70. [Google Scholar]

Baranovskyi O. I., Khutorna M. E. Methodology of forming the system of ensuring financial stability of credit institutions. *Financial and credit activities: problems of theory and practice.* 2018. Vol. 4. No 27. P. 4-13. [Google Scholar] [CroosRef]

Baranovskyi O. I. (2018). Quality of the transformational processes in the financial sector of the national economy: vectors of the measuremen. *Financial and credit activities: problems of theory and practice*. 2018. Vol. 4. No 27. P. 4-13. [Google Scholar] [CrossRef]

Beaver, W. (1966). FinancialRatios as Predictors of Failure. *Journal of AccountingResearch*, *4*, 71-102. [Google Scholar] [CrossRef]

Belas, J., Cipovova, E., Novak, P., & Polach, J. (2012). Impacts of the foundation internal ratings based approach usage on financial performance of commercial bank. *E+ M Ekonomie a Management*. [Google Scholar]

Bellovary, J. L., Giacomino, D. E., & Akers, M. D. (2007). A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial education*, 1-42. [Google Scholar]

Biggs, D., De Ville, B., & Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of applied statistics*, *18*(1), 49-62. [Google Scholar] [CroosRef]

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32. [Google Scholar]

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Calderon, T. G., & Cheh, J. J. (2002). A roadmap for future neural networks research in auditing and risk assessment. *International Journal of Accounting Information Systems*, *3*(4), 203-236. [Google Scholar] [CrossRef]

Chipman, H. A., George, E. I., & McCulloch, R. E. (1998). Bayesian CART model search. *Journal of the American Statistical Association*, *93*(443), 935-948. [Google Scholar] [CrossRef]

Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, *4*(1), 266-298. [Google Scholar]

Chrastinova, Z. (1998). *Metodyhodnoteniaekonomickejbonity a predikciefinancnejsituaciepoľnohospodarskychpodnikov.* Bratislava: Vyskumnyústaveconomikypoľnohospodarstva a potravinarstva. [Google Scholar]

Ciampi, A. (1991). Generalized regression trees. *Computational Statistics & Data Analysis*, *12*(1), 57-78. [Google Scholar] [CrossRef]

Ciampi, A., Hogg, S. A., McKinney, S., & Thiffault, J. (1988). RECPAM: A Computer Program for Recursive Partition and Amalgamation for Censored Survival Data and Other Situations Frequently Occurring in Biostatistics. *Computer Methods and Programs in Biomedicine*, *26*, 239-256.

Cipovova, E., & Belas, J. (2012). Assessment of credit risk approaches in relation with competitiveness increase of the banking sector. *Journal of Competitiveness*. [Google Scholar] [CrossRef]

Ciszewski, T., & Nowakowski, W. (2018). Economic analysis of the life-cycle cost structure for railway traffic control systems. *Ekonomicko-manazerske spektrum*, *12*(1), 30-43. [Google Scholar]

Davis, R. B., & Anderson, J. R. (1989). Exponential survival trees. *Statistics in Medicine, 8*(8), 947-961. [Google Scholar]

Dimitras, A. I., Zanakis, S. H., & Zopounidis, C. (1996). A survey of business failures with an emphasis on prediction methods and industrial applications. *European Journal of Operational Research*, *90*(3), 487-513. Google Scholar] [CrossRef]

Durica, M., & Adamko, P. (2016). Verification of MDA bankruptcy prediction models for enterprises in Slovak Republic. *Proceedings of the 10th international days of statistics and economics. Praha: Melandrium.* [Google Scholar]

Dvorsky, J., Petrakova, Z., Khan, K. A., Formanek, I., Milolas, Z., & Danko, L. (2020). Selected aspects of strategic management in the service sector. *Journal of Tourism and Services, 11*(20), 109-123. [CrossRef]

Fan, G., & Gray, J. B. (2005). Regression tree analysis using TARGET. *Journal of Computational and Graphical Statistics*, *14*(1), 206-218. [Google Scholar] [CrossRef]

Fialova, V., & Folvarcna, A. (2020). Default Prediction Using Neural Networks for Enterprises from the Post-Soviet Country. *Ekonomicko-manazerske spektrum*, *14*(1), 43-51. [Google Scholar]

Fitzpatrick, P. (1932). A Comparison of the Ratios of Successful Industrial Enterprises with Those of Failed Companies. *Certified Public Accountant*, (6), 727-731. [Google Scholar]

Gordon, L., & Olshen, R. A. (1985). Tree-structured survival analysis. *Cancer treatment reports*, *69*(10), 1065-1069. [Google Scholar]

Gray, J. B., & Fan, G. (2008). Classification tree analysis using TARGET. *Computational Statistics & Data Analysis*, *52*(3), 1362-1372. [Google Scholar] [CrossRef]

Gurcík, Ľ. (2002). G-index–metoda predikcie finančného stavu poľnohospodarskych podnikov. *Agricultural economics*, *48*(8), 373-378.

Hrytsenko L., Roienko V., Boiarko I. (2018). Institutional background of the role of state in investment processes activation. Financial and credit activities: problems of theory and practice, 1, 24, 338-344. [Google Scholar] [CrossRef]

Hiadlovsky, V., & Kral, P. (2014). A Few Notes to Business FinancialHealthPrediction. In: *7ᵗʰ International ScientificConferenceManaging and Modeling of FinancialRisks*. Conferenceproceedings, 248-255.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics*, *15*(3), 651-674. [Google Scholar] [CrossRef]

Huxley, S. J., & Sidaoui, M. (2018). Gaining Market Share in Emerging Markets Portfolios by Moderating Extreme Returns: The Case of Peru. *Economics, Management & Financial Markets*, *13*(3). [Google Scholar]

Jones, F. L. (1987). Current Techniques in Bankruptcy Prediction. *Journal of Accounting Literature*, *6*, 131-164. [Google Scholar]

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *29*(2), 119-127. [Google Scholar]

Kim, H., & Loh, W. Y. (2001). Classification trees with unbiased multiway splits. *Journal of the American Statistical Association*, *96*(454), 589-604. [Google Scholar] [CrossRef]

Kim, H., & Loh, W. Y. (2003). Classification trees with bivariate linear discriminant node models. *Journal of Computational and Graphical Statistics*, *12*(3), 512-530. [Google Scholar] [CrossRef]

Kliestikova, J., Misankova, M., & Kliestik, T. (2017). Bankruptcy in Slovakia: international comparison of the creditor´ s position. *Oeconomia Copernicana*, *8*(2). [Google Scholar]

Kljucnikov, A., Belas, J., & Smrcka, L. (2016). Risk-taking and Aggressiveness as the Significant Part of the Entrepreneurial Orientation of SMEs: Case of the Czech Republic. *Polish Journal of Management Studies*, *14*(1), 129-139. [Google Scholar]

Konigova, M., Urbancova, H., & Fejfar, J. (2012). Identification of Managerial Competencies in Knowledge-based Organizations. *Journal of Competitiveness*, *4*(1). [Google Scholar] [CrossRef]

Kubickova, D. (2015). Bankruptcy Prediction and Qualitative Parametres: The Ohlson's Model and its Variants. In: *7th International Scientific Conference on Finance and Performance of Firms in Science, Education and Practice*. Conference proceedings, 805-818. [Google Scholar]

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), 1-28. [Google Scholar] [CrossRef]

Kuznetsova A., Kalynets K., Kozmuk N. Innovative management in global financial csr governance. *Marketing and management of innovations*. 2018. No 2. P. 262-269.. [Google Scholar] [CrossRef]

LeBlanc, M., & Crowley, J. (1992). Relative risk trees for censored survival data. *Biometrics*, 411-425. [Google Scholar] [CrossRef]

Lee, S. K. (2005). On generalized multivariate decision tree by using GEE. *Computational Statistics & Data Analysis*, *49*(4), 1105-1119. [Google Scholar] [CrossRef]

Loh, W. Y. (2002). Regression tress with unbiased variable selection and interaction detection. *Statistica sinica*, 361-386. [Google Scholar]

Loh, W. Y. (2009). Improving the precision of classification trees. *The Annals of Applied Statistics*, 1710-1737. [Google Scholar]

Loh, W. Y., & Shih, Y. S. (1997). Split selection methods for classification trees. *Statistica sinica*, 815-840. [Google Scholar]

Loh, W. Y., & Vanichsetakul, N. (1988). Tree-structured classification via generalized discriminant analysis. *Journal of the American Statistical Association*, *83*(403), 715-725. [Google Scholar]

Loh, W. Y., & Zheng, W. (2013). Regression trees for longitudinal and multiresponse data. *The Annals of Applied Statistics*, *7*(1), 495-522. [Google Scholar]

Messenger, R., & Mandell, L. (1972). A modal search technique for predictive nominal scale multivariate analysis. *Journal of the American statistical association*, *67*(340), 768-772. [Google Scholar] [CrossRef]

Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American statistical association*, *58*(302), 415-434. [Google Scholar] []

Mousavi, M. M., Ouenniche, J., & Xu, B. (2015). Performance evaluation of bankruptcy prediction models: An orientation-free super-efficiency DEA-based framework. *International Review of Financial Analysis*, *42*, 64-75. [Google Scholar] [CrossRef]

Newton, W. G. (2005). *Bankruptcy and Insolvency Accounting*. John Wiley & Sons. Canada. [Google Scholar]

O'leary, D. E. (1998). Using neural networks to predict corporate failure. *Intelligent Systems in Accounting, Finance & Management*, *7*(3), 187-197. [Google Scholar] [CrossRef]

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, *1*(1), 81-106. [Google Scholar] [CrossRef]

Quinlan, J. R. (1992, November). Learning with continuous classes. In *5th Australian joint conference on artificial intelligence* (Vol. 92, pp. 343-348). [Google Scholar] [Google Scholar]

Quinlan, J. R. (1993). C4. 5: Programs for machine learning Morgan Kaufmann San Francisco. *CA, USA*. [Google Scholar]

Rajnoha, R., & Lorincova, S. (2015). Strategic management of business performance based on innovations and information support in specific conditions of Slovakia. *Journal of Competitiveness*. [Google Scholar] [CrossRef]

Reitano, V. (2015). Decision Trees for Analytics Using SAS Enterprise Miner. *Social Science Computer Review*, *33*(3), 415-417. [Google Scholar] [CrossRef]

Salaga, J., Bartosova, V., & Kicova, E. (2015). Economic Value Added as a Measurement Tool of Financial Performance. *Procedia Economics and Finance*, (26), 484-489. [Google Scholar] [CrossRef]

Scott, J. (1981). The probability of bankruptcy: a comparison of empirical predictions and theoretical models. *Journal of Banking & Finance*, *5*(3), 317-344. [Google Scholar] [CrossRef]

Segal, M. R. (1988). Regression trees for censored data. *Biometrics*, 35-47. [Google Scholar] [CrossRef]

Segal, M. R. (1992). Tree-structured methods for longitudinal data. *Journal of the American Statistical Association*, 87(418), 407-418. [Google Scholar]

Sela, R. J., & Simonoff, J. S. (2012). RE-EM trees: a data mining approach for longitudinal and clustered data. *Machine learning*, 86(2), 169-207. [Google Scholar] [CrossRef]

Sharifabadi, M. R., Mirhaj, M., & Izadinia, N. (2017). The impact of financial ratios on the prediction of bankruptcy of small and medium companies. *QUID: Investigacion, Ciencia y Tecnología*, (1), 164-173. [Google Scholar]

Slatter, S. S. P., & Lovett, D. (1999). *Corporate recovery: Managing companies in distress*. Beard Books. [Google Scholar]

Su, X. G., Wang, M., & Fan, J. J. (2004). Maximum Likelihood Regression Trees. *Journal of Computational and Graphical Statistics*, 13, 586–598.

Svabova, L., Kramarova, K., & Durica, M. (2018). Prediction model of firms financial distress. *Ekonomicko-manazerske spektrum*, 12(1), 16-29. [Google Scholar]

Utgoff, P. E. (1989, January). Improved training via incremental learning. In *Proceedings of the sixth international workshop on Machine learning* (pp. 362-365). Morgan Kaufmann. [Google Scholar] [CrossRef]

Valaskova, K., Bartosova, V., & Kubala, P. (2019). Behavioural aspects of the financial decision-making. *Organizacija*, 52(1), 22-31. [Google Scholar] [CrossRef]

Valaskova, K., Siekelova, A., & Weissova, I. (2017). Credit Risk Measurement Using VaR Methodology. In *Advances in Applied Economic Research* (pp. 289-302). Springer, Cham. [Google Scholar]

Virag, M., & Kristof, T. (2005). Neural networks in bankruptcy prediction-A comparative study on the basis of the first Hungarian bankruptcy model. *Acta Oeconomica*, 55(4), 403-426. [Google Scholar]

Zavgren, C. V. (1985). Assessing the vulnerability to failure of American industrial firms: a logistic analysis. *Journal of Business Finance & Accounting*, 12(1), 19-45. [Google Scholar]

Zhang, H. (1998). Classification trees for multiple binary responses. *Journal of the American Statistical Association*, 93(441), 180-193. [Google Scholar]

Zyka, J., & Drahotsky, I. (2019). Methodology for Assessing the Impact of Workplace Ergonomic Factors on Airport Security Screener s Reliability and Performance. *Journal of Tourism and Services*, 10(18), 104-116. [Google Scholar] [CrossRef]

Івана Подгорська, *Ph.D., Жилінський університет, Словацька Республіка*
Яромир Врбка, *Ph.D., Інститут технологій і бізнесу в Чеських Будейовицях, Чеська Республіка*
Георгій Лазарою, *Ph.D., доцент, Інститут когнітивної науки, США, Університет Спіру Харет, Румунія*
Марія Ковацова, *Ph.D., Жилінський університет, Словацька Республіка*

Інновації у фінансовому менеджменті: модель рекурсивного прогнозу на основі алгоритму дерева рішень

У статті проаналізовано ефективність інновацій у фінансовому менеджменті підприємства. Авторами зазначено, що банкрутство є одним із наслідків не ефективного фінансового менеджменту компанії. При цьому менеджмент компанії між час розроблення системи фінансових інтсрументів повинен враховувати наявні зовнішні та внутрішні ризики діяльності ккомпанії. У статті розглянуто алгоритм дерева рішень, як один із найбільш інтуїтивних методів збору даних, який можна використовувати для прогнозування ймовірності настання фінансових ризиків. Результати систематизації наукового доробку за даним напрямом засвідчили, що експерти використовують дерево рішень у якості інноваційного інструменту фінансового менеджменту. Статтю присвячено аналізу можливостей застосування алгоритму дерева рішень для створення моделі прогнозування банкрутсва компанії. Авторами розроблено комплексну модель прогнозування фінансової бідності підприємств на основі дерева рішень з використанням алгоритму CART. Для формування моделі прогнозування використано дані 2 359 731 підприємств (30% від загальної суми) із 17 країн, а саме: Словаччини, Чехії, Польщі, Угорщини, Румунії, Болгарії, Литви, Латвії, Естонії, Словенії, Хорватії, Сербії, Росії, України, Білорусії, Чорногорії та Македонії. При цьому зазначено, що 1 802 027 компанфії є процвітаючими та 557 704 – не прибутковими. Статистичні дані згенеровано з бази даних Amadeus. Вхідними змінними моделі обрано 24 фінансові показники, 3 допоміжні змінні та дані про ВВП країн у 2015 та 2016 роках. Вибірку для розроблення моделі сформовано на основі даних 80% підприємств, тоді як дані 20% – для її тестування. Отримана модель дозволила класифікувати 93,2% підприємств як процвітаючі та 83,5% – не прибуткові. Авторами наголошено, що запропонована модель прогнозування є придатною для класифікації підприємств за рівнем ефекткивності їх діяльності на зростаючих ринках.

Ключові слова: модель прогнозу, дерево рішень, зростаючі ринки.