Tomasz PISULA[1]
Grzegorz MENTEL[2]
Jacek BROŻYNA[3]

PREDICTING BANKRUPTCY OF COMPANIES FROM THE LOGISTICS SECTOR OPERATING IN THE PODKARPACIE REGION

Research on effectiveness of various concepts for modelling the bankruptcy of companies from the logistics sector is described in this article. In order to present this issue more completely the above-mentioned prediction of possible negative effects for the conducted business activity was conducted for all companies operating in that sector in the Podkarpacie region. The study was supported by the data from the database EMIS (Emerging Markets Information Service). A wide range of 28 financial indicators was grouped into five groups i.e. liquidity ratios, profitability, debt, performance, and financial respectively. The above mentioned research trial was divided into a group of companies – so-called ill - in relation to which the bankruptcy was declared and healthy ones (of good financial condition).

Such an approach allows for a better and right assessment of the methods in modeling bankruptcy. The purpose of this publication was to find factors (models) describing the risk of bankruptcy of enterprises in terms of their effectiveness prediction in one - and two year- horizon. The logistics regression models, classification trees and two lunatics artificial neural networks were applied. A full evaluation of the models application were made in the validation process. The primary tool used in this case to study the effectiveness of models classification are matrices of correct classification. It was made an estimation of the correct and wrong indications in both the above mentioned models. Finally, an assessment of the method was done as well as the overall condition of the logistics sector in the Podkarpacie region.

Keywords: bankruptcy, logistic sector, modeling, financial indexes.

---

[1] Tomasz Pisula, PhD, Department of Quantitative Methods, Faculty of Management, The Rzeszow University of Technology, Rzeszow, al. Powstańców Warszawy 8, 35-959 Rzeszów, tel. 17 8651894, e-mail: tpisula@prz.edu.pl. (Corresponding Author).

[2] Grzegorz Mentel, PhD, Department of Quantitative Methods, Faculty of Management, The Rzeszow University of Technology, Rzeszow, al. Powstańców Warszawy 8, 35-959 Rzeszów, tel. 17 8651894, e-mail: gmentel@prz.edu.pl.

[3] Jacek Brożyna, PhD, Eng., Department of Quantitative Methods, Faculty of Management, The Rzeszow University of Technology, Rzeszow, al. Powstańców Warszawy 8, 35-959 Rzeszów, tel. 17 8651894, e-mail: Jacek.Brozyna@prz.edu.pl.

1. INTRODUCTION

Insolvency, also called bankruptcy, constitutes a fundamental element of enterprises in the economic sense. Frequency of its appearance has on the one hand a deep influence on the result of economic growth and unemployment and on the other hand on the so-called financial stability, both in relation towards banks and towards broadly defined financial markets. Thus on the micro level insolvency can be perceived as the main credit risk factor, posing primary problem for the mentioned banks and investors, whereas on the macroscale as a negative element of possible recession.

The current period of economic recession on the global markets has lately increased discussions on effective way of predicting the companies' bankruptcy. Emerging solution concepts in that scope do not always allow to predict possible risk of bankruptcy effectively, thus do not always give signal for possible countermeasures. Multitude of solutions in scope of bankruptcy risk modelling does not always goes hand in hand with their „quality".

The content of this publication is the issue of effectiveness evaluation of widely used ways of predicting companies' bankruptcy. In order to present a more comprehensive opinion about commonly used mechanisms in that scope a modelling for one- and two year prediction periods has been conducted. Research sample is made of logistics companies from the Podkarpacie region. Companies have been divided into two groups, i.e. group of the so-called healthy companies, not threatened with bankruptcy and ones that in a given period declared bankruptcy. Such approach was aimed at verification of researched methods in relation to the prediction horizon. Analyzed methods have been divided into three categories, namely models of logistic regression, artificial neural networks and method of classification and regression trees.

2. ANALYSIS OF CHOSEN LITERATURE CONCERNING RESEARCH ON COMPANIES' BANKRUPTCY RISK

An exhaustive analysis of works concerning the issue of predicting companies' bankruptcy risk can be found in the work (Kumar & Ravi, 2007). The authors analyzed 128 publications concerning the issue of predicting companies' bankruptcy risk, which were published in a period from 1968 to 2005. Analysis of publications has been conducted from a perspective of usage of statistical methods and artificial intelligence methods to solve problems related with predicting companies' and banks' bankruptcy risk. Most publications concerned research of bankruptcy risk for companies (both listed and non-listed). Out of 128 analyzed publications only a dozen concerned research on bank bankruptcy risk. In some publications both companies and banks bankruptcy models have been researched.

Volume of research sample used by various authors in their research is very diverse (it ranges from 24 up to even 8977). Similar time periods for used financial data were diversified and included different time periods from 1997 to 2003 with a time horizon from one year up to even a couple of years.

Usage of research techniques by various authors is also very diverse. In research of this type one successfully uses both statistical methods, such as: discriminant analysis, models of logistic regression, decision trees and methods of nearest neighbours, as well as methods based on the optimization algorithms of operational research or methods of

artificial intelligence, such as: neural networks, theory of rough sets, mathematical programming, genetic algorithms, etc.).

The most commonly used statistical techniques to research companies' bankruptcy are based on discriminant analysis, logit models and decision trees. Nowadays they are very rarely used as sole and only research methods. They are used rather as a comparison model, in relation to other non-statistical models, or as component models in hybrid approaches.

One vital issue in regards to application of statistical methods to predict companies' bankruptcy is the work of Altman (Altman & Haldeman & Narayanan, 1977), where for the first time authors introduced a new model of classification of bankrupted companies, which they named „Zeta analysis". Ohlson (Ohlson, 1980) proposed a model of logistic regression to research the risk of companies' bankruptcy, while in 2002 Kolari and others (Kolari & Glennon & Shin & Caputo, 2002) introduced the so-called bankruptcy risk early warning system for large banks in the USA, also based on the logistic regression model. Effectiveness of correct classifications for data collected one year before bankruptcy amounted up to more than 96%, whereas for data two years before bankruptcy up to more than 95%. Among non-statistical methods used in classification of companies bankrupted because of bankruptcy the most often used ones are neural networks models. Tam and Kiang (Tam & Kiang, 1992) compared classifying quality of bankrupted companies for LDA models, logistic regression, k-nearest neighbours with neural networks models. Neural network models estimated by them had the best classifying statistics for prediction horizon of 1 year. In the case of data for 2 years before bankruptcy period the best model turned out to be an LDA model. Fletcher and Goss (Fletcher & Goss, 1993) used neural networks models to predict bankruptcy of a company in relation to logit model. They used a technique of V-times cross-validation to choose the best model and used 3 independent indicators (data was taken for 33 researched enterprises) as potential bankruptcy predictors in created and trained neural networks. Classifying quality of estimated neural network model amounted to 82%, whereas for an alternative logit model only to 77%.

More about usage of different types of artificial neural networks in predicting companies' bankruptcy can be additionally found in works (Lee & Booth & Alam, 2005), (Lam, 2004), (Leshno & Spector, 1996), (Wilson & Sharda, 1994), (Kiviluoto, 1998).

3. CHARACTERISTICS OF FINANCIAL FACTORS AND RESEARCH SAMPLES USED TO PREDICT BANKRUPTCY OF LOGISTICS COMPANIES

Information about bankruptcies of Polish companies were taken from bankruptcy database of Polish companies - Corporate Database EMIS information system (Emerging Markets Information Service).

To predict bankruptcy of logistics sector companies 28 financial indicators characterizing financial condition and managing effectiveness of researched companies have been chosen as bankruptcy predictors. Indicators have been divided into 5 groups: financial liquidity indicators, profitability indicators (return on sales), indebtedness indicators and

financial leverage, operating effectiveness (proficiency) and other indicators of capital-material structure of a company.

Statistical data of financial indicators for Polish companies were taken from financial reports of companies. The following financial indicators were chosen for research:

- *Liquidity indicators (*100%):* X1 - CURRENT LIQUIDITY INDICATOR: Current assets / Short-term liabilities, X2 - FAST LIQUIDITY INDICATOR: (Current assets − Stock) / Short-term liabilities, X3 - LIQUIDITY INDICATOR (KO/SB): Circulating capital (working capital) / Balance sheet total = (Current assets − Short-term prepayments and accruals - Short-term liabilities) / Balance sheet total, X4 - IMMEDIATELY DUE INDICATOR: (Current assets − Stock − Short-term receivables) / Short-term liabilities, X5 − CASH LIQUIDITY INDICATOR: Cash and cash equivalents / Short-term liabilities

- *Profitability indicators (*100%):* X6 - OPERATING PROFIT MARGIN: Operating result (profit-operating loss) / Net sales income, X7 − Profitability: Net profit / (Equity capital − Net profit), X8 - RETURN ON ASSETS (Asset profitability) (ROA): Net profit / Balance sheet total, X9 − RETURN ON EQUITY (profitability of equity capital) (ROE): Net profit / Equity capital, X10 − RETURN ON CAPITAL: Net profit / (Assets in total − Short-term liabilities), X11 − NET SALES PROFITABILITY (ROS): Net profit / Net sales income, X12 − GROSS PROFIT MARGIN: (Net income from sales of goods and products and equal to them − Operating expenses) / Net income from sales of goods and products and equal to them

- *Indebtedness indicators and financial leverage effect (*100%):* X13 − GENERAL DEBT: (Short-term liabilities + Long-term liabilities) / Balance sheet total, X14 - DEBT ON EQUITY: Total liabilities / Equity capital, X15 − DEBT (Equity capital + Long-term liabilities) / Fixed assets, X16 − ASSETS DEBT: Short-term liabilities / Balance sheet total, X17 − DEBT Gross profit / Short-term liabilities, X18 − DEBT (Net profit + Depreciation) / Total liabilities, X19 − LONG-TERM DEBT: Long-term liabilities / Equity capital, X20 − FINANCIAL LEVERAGE: Assets total / Equity capital, X21 − LEVERAGE (DEBT/COMPANMY VALUE): Total liabilities / (Equity capital + Total liabilities − Cash and its equivalents)

- *Operating effectiveness indicators:* X22 − RECEIVABLES TURNOVER [in days]: Average short-term receivables/ Net sales income *360, X23 - OBRÓT AKTYWAMI: Net sales income / Assets *100%, X24 − STOCK TURNOVER [in days]: Stock / Net sales income * 360, X25 − CASH CYCLE: Short-term receivables / Net sales income * 365 + Stock / Operating expenses * 365 − Short-term liabilities (without special funds and other short-term financial liabilities) / Operating expenses (without other operating expenses) * 365

- *Financial indicators − characterizing the companies' capital and material structure (*100%):* X26 − Equity capital / Balance sheet total, X27 − Fixed assets (without long-term prepayments and accruals) / Balance sheet total, X28 − Fixed assets / Current assets

Research samples were created on the basis of collected statistical data. Dependent variable was a qualitative dichotomous dependent variable Y defining whether a company is a company which declared bankruptcy (Y=1 – bankrupt), or a company not threatened with bankruptcy (Y=0 – non-bankrupt). 28 previously characterized financial indicators were chosen as a set of entry variables (bankruptcy predictors).

Two research samples were created. The first one included these bankrupted companies from the logistics sector and healthy companies corresponding to them, for which statistical data for one year before bankruptcy period was available (1-year prediction horizon). The second research sample included these bankrupted and healthy companies for which statistical data for two years before bankruptcy period was available (2-year prediction horizon). For each of research samples one corresponding healthy company not threatened with bankruptcy was chosen for one bankrupted company. Selection of healthy companies was preceded by a thorough indicator analysis and the only chosen companies from logistics sector were the ones which indicators pointed at good financial condition and ability to pay their liabilities.

Research sample for data one year before bankruptcy period included: 33 bankrupted companies and 33 healthy companies (statistical data for one year before bankruptcy was available for only that number of companies), whereas in the case of data for 2 years before bankruptcy period there were 57 healthy companies and 57 bankrupted ones. Research samples were divided randomly into two samples: the learning sample, on the basis of which the prediction model parameters were estimated, and test sample researching the effectiveness of correct classifications. The learning sample for one year prediction horizon included: 47 companies (23 bankrupts and 24 non-bankrupts), whereas the test sample included: 19 companies (10 bankrupts and 9 non-bankrupts). For two year prediction horizon the learning sample included 86 logistic companies (43 bankrupts and non-bankrupts), whereas the test sample included: 28 companies (14 bankrupts and non-bankrupts).

In order to scrutinize influence of chosen variables explanatory variable on explained variable – identifying the companies' bankruptcy a ranking analysis of predictors was conducted. A vital issue when choosing proper predictors is also posed by a necessity to choose only such predictors which have the best prognostic properties in scope of separation, i.e. distinguishing between bankrupt and healthy companies. When preparing a ranking of predictors depending on their classifying power one can use in practice the following factors: Information Value (IV), Gini factor and Cramer's V factor.

IV factor – information value of a predictor is expressed by the formula:

$$IV = \sum_{i=1}^{k} \left(\frac{n_i^{NB}}{n_{NB}} - \frac{n_i^{B}}{n_B} \right) \cdot \ln \left(\frac{n_i^{NB} / n_{NB}}{n_i^{B} / n_B} \right) \qquad (1)$$

where: k - the number of attributes (variability intervals) of the examined predictor, n_i^{NB} - the number of healthy companies for i-variability interval of predictor's value, n_i^{B} - the number of bankrupted companies for i-variability interval of predictor's value, n_{NB} - the total number of healthy companies, n_B - the total number of bankrupted companies.

The higher the values of IV factor, the higher the predictive power of the explanatory variable in scope of differentiation between healthy and bankrupted companies. It is assumed that IV values above 0.3 point out to a strong predictive power, while values below 0.02 show complete lack of such predictive power.

Gini factor is based on Lorenz curve factor (for the so-called ROC curve - Receiver Operating Characteristic). It expresses a ratio of fields on the graph of ROC curve (see fig. 1) which is expressed by the formula:

$$Gini = \frac{A}{A+B} = \frac{A}{0.5} = 2 \cdot A = 2 \cdot (0.5 - B) = 1 - 2 \cdot B = 1 - \sum_{i=1}^{k-1} \left(y_{i+1} - y_i \right) \cdot \left(x_{i+1} + x_i \right)$$

(2)

where: k - the number of attributes (variability intervals) of the examined diagnostic variable, $y_i = \sum_{j=1}^{i} \frac{n_j^B}{n_B}$ - cumulated percent of bankrupts, for i-attribute value of variable, $x_i = \sum_{j=1}^{i} \frac{n_j^{NB}}{n_{NB}}$ - corresponding cumulated percent of healthy companies.

It is assumed that values of the Gini factor below 0.35 point out that predictor does not have a sufficient classifying ability to correctly distinguish between healthy and bankrupted companies.
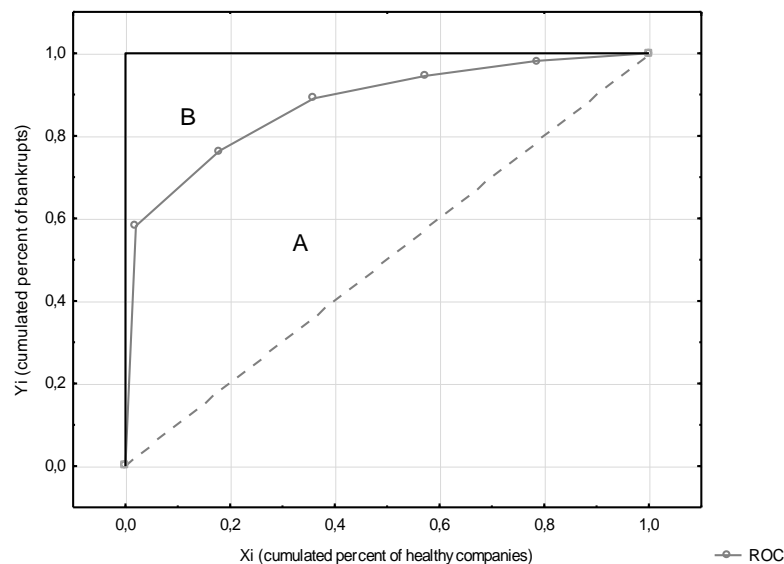

Fig. 1. Example of ROC curve

Source: own study.

Cramer's V factor measures dependence power between values of dichotomous dependent variable 0-1 defining company's bankruptcy and values of the given diagnostic variable. Values of this factor are contained in interval between 0 and 1. It is based on Chi-square independence measure and calculated with the formula:

$$V = \sqrt{\frac{\chi^2}{n}} \qquad (3)$$

where: n – the number of statistical observations, and χ^2 - statistic value for Chi-square independence test, between variable 0-1 defining company's bankruptcy and examined indicator (predictor) of bankruptcy.

The higher the V-Cramer's factor values (closer to 1), the better predictive power of the examined indicator in predicting companies' bankruptcy.

3. CHARACTERISTICS OF MODELS USED IN PREDICTING BANKRUPTCY OF LOGISTICS COMPANIES

To predict bankruptcy risk of companies from the logistics sector in this work the following statistical models of bankruptcy classification were used: logistic regression and CRT classification trees. However, from non-statistical methods (using artificial intelligence methods) neural networks models based on multilayer perceptrons (MLP) were used.

Models of logistic regression - Logit

General form of two-state model of logistic regression describing dependence of the possibility of bankruptcy of examined companies depending on a set of factors influencing the occurrence of this event is expressed by function:

$$P(Y = 1) = \frac{1}{1 + e^{-(\alpha_0 + \alpha_1 X_1 + ... + \alpha_k X_k)}} \qquad (4)$$

In order to choose potential variables for a logit model a factor analysis was used as well as values of ranking statistics for importance of predictors (Tab. 1 and Tab. 2). For prediction horizon of 1 year the X_{23} and X_{28} variable were discarded from the list of potential variables, because they had low value of ranking measures, whereas for a model with prediction horizon of 2 years the following variables were discarded: X_{12}, X_{19}, X_{22}, X_{23}, X_{24}, X_{28}.

After implementing factor analysis the following variables were chosen for estimating model for a one year prediction horizon: X_{26}, X_{18}, X_{20}, X_{11}, X_{22}, X_5, X_{10} as well as other variables (weakly correlated with factors and between themselves): X_1, X_2, X_7, X_{15}, X_{24}, X_{25}, X_{27}

A list of potential diagnostic indicators for a model with two year prediction horizon, including variables: X_3, X_5, X_7, X_8, X_9, X_{13}, X_{17}, X_{21}, X_{25} was selected in a similar way.

To estimate parameters of logistic regression model a module of generalized linear and non-linear models was used (generalized logit model).

In the estimated models there were only these diagnostic variables, for which the *Wald* statistics value was statistically relevant on the level of p<0.05.

The table below (Tab. 1) presents estimated coefficients and values of *Wald* statistics for both logit models with 1 year and 2 year prediction horizon.

Table 1. Estimation of parameters for logistic regression models.

Predictor	Evaluation of parameter	Estimation error	Wald statistics value	Test probability (*p-value*)
Model of logistic regression – 1 year to bankruptcy				
absolute term	6.16642	2.142386	8.284	0.004
X_1	-0.04938	0.015481	10.174	0.001
X_{11}	-0.11751	0.061335	3,670	0.050
X_{27}	-0.04283	0.021563	3.945	0.047
Model of logistic regression – 2 years to bankruptcy				
absolute term	1,4645	0.569849	6.605	0.0102
X_5	-0.05544	0.023063	5.779	0.0162
X_{11}	-0.0722	0.036052	4.011	0.0452
X_{27}	-0.02084	0.008869	5.522	0.0188

Source: own study.

Classification trees – C&RT

C&RT (Classification and Regression Trees) is a tool for statistical analysis of data used to create classification and regression models. Tree is a kind of a graphic model, created as a result of recurrent division of a set of output observations into numerous subsets. The aim of such division is to gain subsets as homogenous as possible in regards to dependent variable value. Algorithm of recurrent division (so-called Recursive Partitioning) can use different independent variable on each stage of division. All independent variables (predictors) are always taken into account and the chosen variable guarantees the best division of node, namely one receives division into the most homogenous subsets.

Algorithms of decision trees can be divided into 3 basic types:

- CLS (Concept Learning System)
- AID (Automatic Interaction Detection) – an example of this type of algorithms are CHAID type trees
- C&RT (Classification and Regression Trees)

More about methods and trees algorithms in classifying and regression use can be found in (Breiman & Friedman & Olshen & Stone, 1993).

Table 2. Classification trees for bankruptcy models

Node number	Left node branch	Right node branch	Number of nodes	Size of NB class	Size of B class	Chosen class	Division variable	Division constant

Node number	Left node branch	Right node branch	Number of nodes	Size of NB class	Size of B class	Chosen class	Division variable	Division constant
Prediction horizon: 1 year to bankruptcy								
Selection rule B: ($X_{15}<=75.4$ AND $X_{21}>63.1$)								
Selection rule NB: ($X_{15}>75.4$) OR ($X_{15}<=75.4$ AND $X_{21}<=63.1$)								
Effectiveness of correct classification: learning sample = 93.6 [%], test sample=84.2 [%]								
1	2	3	47	24	23	Non-bankrupt	X_{15}	78,4
2	4	5	26	4	22	Bankrupt	X_{21}	63,1
4			2	2	0	Non-bankrupt		
5			24	2	22	Bankrupt		
3			21	20	1	Non-bankrupt		
Prediction horizon: 2 years to bankruptcy								
Selection rule B:								
($X_{13}>89.4$) OR ($X_{13}<=89.4$ AND $X_{24}>13.3$) OR ($X_{13}<=89.4$ AND $X_{24}<=13.3$ AND $X_7<=-51.0$)								
Selection rule NB:								
($X_{13}<=89.4$ AND $X_{24}<=13.3$ AND $X_7>-51.0$)								
Effectiveness of correct classification: learning sample = 84.9 [%], test sample=71.4 [%]								
1	2	3	86	43	43	Non-bankrupt	X_{13}	89.4
2	4	5	56	39	17	Non-bankrupt	X_{24}	13.3
4	6	7	49	38	11	Non-bankrupt	X_7	-51.0
6			3	0	3	Bankrupt		
7			46	38	8	Non-bankrupt		
5			7	1	6	Bankrupt		
3			30	4	26	Bankrupt		

Source: own study.

C&RT trees algorithms were used in this publication to analyze bankruptcy of logistics companies. A Statistica package module – General models of classification and regression trees was used. All 28 financial indicators were chosen as entry variables. Gini measure was used as a method of trees division, whereas to choose the best trunked tree – a V-times cross-validation, as a rule of one standard error. Minimization of average costs of incorrect classification was used as a criterion of optimal tree trunking (the same costs of incorrect classification, equal to 1, were set for bankrupts and non-bankrupts).

Structure of the best classification trees for one year and two year prediction horizon is presented in Table (tab. 2). There are also rules of tree division and node creation, as well

as classification effectiveness of trees given in the table. For a classification tree for a one year prediction the average costs of incorrect classification amounted to 0.106 for a learning sample and 0.162 for a test sample. For a two year prediction these costs amounted to 0.256 and 0.258, respectively.

The figure below (fig. 2) presents a graphic illustration of a classification tree to classify logistics companies threatened with bankruptcy in one year period horizon.
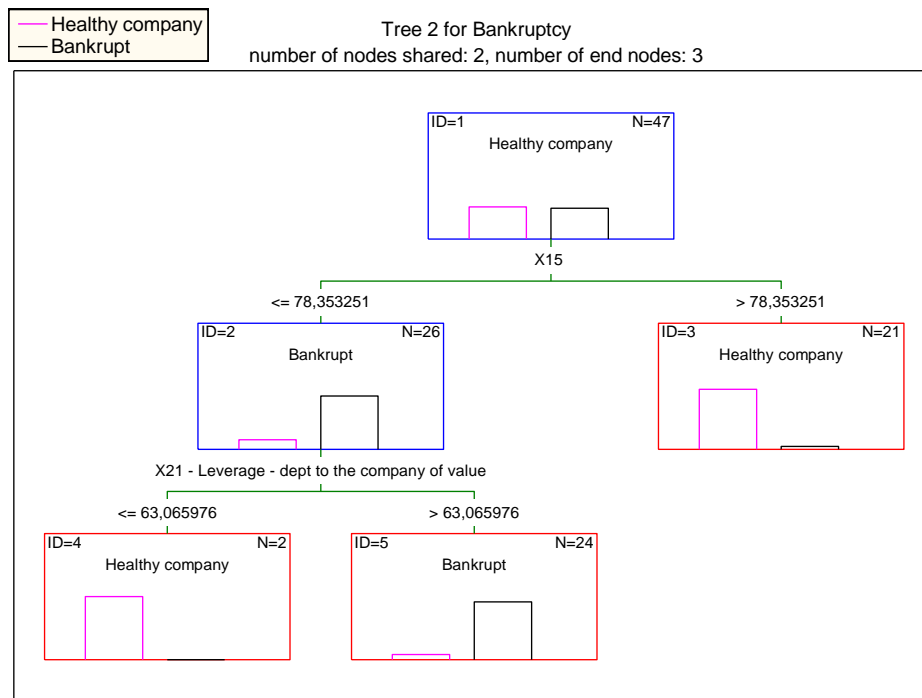

Fig. 2. Graphic illustration of tree structure for classifying logistics companies for a one year period horizon

Source: own study.

Artificial neural networks – MLP

Artificial neural networks are one of the most commonly used techniques to solve problems of correct classification of companies threatened with bankruptcy. The structure of an artificial neural network is modelled after human brain models. Neural networks comprise of many elements processing pieces of information - so-called neurons. Functioning schematic of an artificial network neuron is presented in the figure below (see fig. 3a). Each neuron processes entry signals, marked in figure with x_k $(k = 1, ..., n)$ into one output signal y. Weights play a very significant role in this model, they are marked by w_k $(k = 1, ..., n)$, which on the one hand define the importance of information provided

by $i$channel upon entry, on the other hand they record relations occurring between entry signals and output signal. On the basis of data appearing upon entry of a neuron a linear combination of weight vector and entry data is calculated. This combination defines the so-called total neuron stimulation e. Signal occurring at the output of a neuron is a function of neuron activation $y = f(e)$, dependent on the total neuron stimulation.

In practical applications various forms of neuron activation function are assumed. The simplest form is identity function of neuron activation $e = f(e)$. More complex activation functions are also used (Witkowska, 2002), such as: threshold function, linear function, logistic function, exponential function, hyperbolic tangent, sinus function, Gauss function and others.
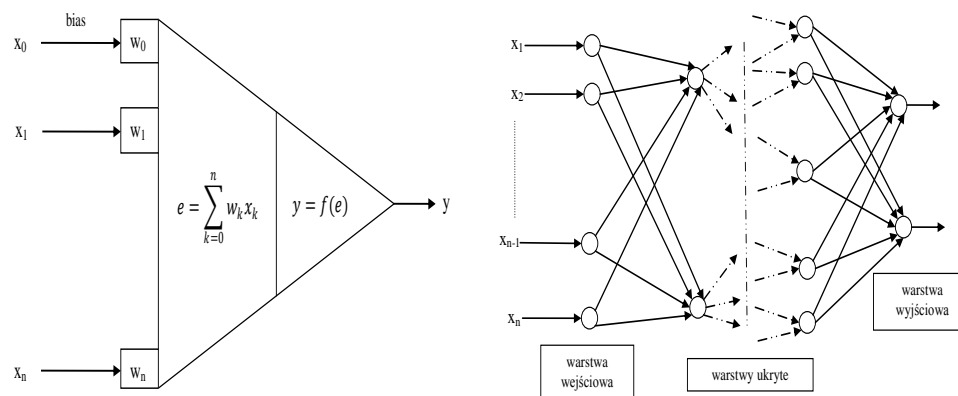

Fig. 3a) Model of an artificial cell

3b) Structure of a one-way multilayer neural network; in a neural network

Source: own study, based on: Witkowska D., *Sztuczne sieci …*, 2002, pp.4, pp.10.

Artificial neural networks are built from many layers of neurons connected with many structural and topological interrelations. Neurons that belong to the first layer create entries to the network and create the so-called network entry layer. The neurons belonging to the last layer are the network exits and create an exit layer of a topological structure of an artificial neural network. Between outer layers of network (entry and exit ones) there are often intermediate layers of neurons creating the so-called hidden layers of network (neurons that create them are called hidden neurons).

Usage of neural networks to predict bankruptcy of logistics companies was conducted by using an *Automatic Neural Networks* module from *Statistica 10.0* package. Artificial neural networks module from Statistica package has embedded algorithms of automatic analysis of neural networks, enabling automatic search for the best neural networks, which have the best classifying properties for a given set of diagnostic variables. Constructed neural networks were *MLP* (multilayer perceptron)-type networks, which had only one layer of hidden neurons. Two versions of entry data were used (both for one year and two

year prediction horizon). In the first version all indicators were used as entry variables, for which values of ranking indicators (Tab. 1 and Tab. 2) were on a properly high level (at least one indicator of ranking measure higher than 0.4), whereas in the second version as entry variables only those were chosen which were chosen as significant for logit and discriminative models. This is why the number of neurons in a hidden layer depended on the number of entry variables in a network and it changed depending on the chosen version in a range between 3 and 21 hidden neurons.

In the process of training (learning) of neural networks a BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm was used – it uses quasi Newton's method and two versions of error function were used: sum of squares ERR_{SoS} and cross entropy (CE – $Cross$ $Entropy$) ERR_{CE}.

Usage of Artificial Neural Networks module in the Statistica package allowed for an automatic choice of the best networks for a given set of entry variables and research sample (learning and test ones), for which the total error of classification correctness is the lowest. Table 3 presents a summary of learning results for 4 best networks (one for each version of entry variables and chosen prediction horizon).

Table 3. The best neural networks models obtained for different versions of entry variables and bankruptcy prediction horizon of one year and two years.

Network type (network id)	Function of learning error	Function of hidden neuron activation	Function of output neurons activation	Percent of correct classification learning sample [%]	Percent of correct classification test sample [%]
Prediction horizon - 1 year to bankruptcy					
Version 1: entry variables: X_1-X_{28} with exclusion: X_{23} i X_{28}					
MLP 26-8-2	Cross entropy ERR_{SoS}	Linear	Linear	91.5	94.7
Version 2: entry variables: X_1, X_2, X_6, X_{11}, X_{18}, X_{27}					
MLP 6-3-2	Cross entropy ERR_{CE}	Tanh	Softmax	87.2	89.5
Version horizon - 2 years to bankruptcy					
Version 1: entry variables: X_1-X_{28} wih exclusion: X_{12}, X_{19}, X_{22}, X_{23}, X_{24}, X_{28}					
MLP 22-17-2	Sum of squares ERR_{CE}	Logistic	Softmax	86.1	92.9
Version 2: entry variables: X_2, X_5, X_{13}, X_{26}					
MLP 4-8-2	Cross entropy ERR_{SoS}	Exponential	Tanh	73.3	82.1

Source: own study.

4. VALIDATION OF ESTIMATED BANKRUPTCY MODELS

In order to choose the best models for practical applications, which will be used to predict bankruptcy of logistics companies from the Podkarpacie region, estimated models of bankruptcy prediction were subject to thorough validation analysis. Usefulness of models

in scope of their best classifying properties in correct recognition of companies threatened with bankruptcy and healthy companies, as well as proper model calibration to data from learning samples was examined.

The fundamental tool used to scrutiny classifying effectiveness of classification models are proper classification matrices (see Tab. 4). *TN* (True Negative) number in the table denotes the number of healthy companies properly qualified by the model. Similarly *TP* (True Positive) number denotes the number of bankrupted companies properly qualified by the model. If healthy companies are classified by the model as bankrupts, then such classification error is called I-type error, and *FP* (False Positive) means the number of these incorrect classifications. Much more serious is a II-type classification error, which is made when model qualifies bankrupts as not threatened with bankruptcy, and *FN* denotes the number of such incorrect classifications.

Table 4. Matrix of correct classifications for bankruptcy prediction model.

True affiliation of company	Predicted affiliation of company	
	NB	B
NB	TN (True Negative)	FP (False Positive) I type error
B	FN (False Negative) II type error	TP (True Positive)

Source: own study.

I-type error, namely the percentage of incorrectly qualified healthy companies (also often denoted as: 1-specificity) is expressed with the formula:

$$Err_1 = \frac{FP}{FP + TN} = 1 - \frac{TN}{FP + TN} = 1 - Eff_1 \qquad (5)$$

where: Eff_1 (I-type effectiveness or the so-called specificity) – the percentage of correctly qualified companies as not threatened with bankruptcy.

II-type error, namely the percentage of incorrectly qualified bankrupts (often also denoted as: 1-sensitivity) is expressed with the formula:

$$Err_2 = \frac{FN}{FN + TP} = 1 - \frac{TP}{FN + TP} = 1 - Eff_2 \qquad (6)$$

where: Eff_2 (II-type effectiveness or the so-called sensitivity) – the percentage of correctly qualified bankrupted companies.

In the process of validation of models for classification of companies threatened with bankruptcy the most commonly used are the following model validation measures: Information Value (IV) factor, Gini factor and Divergence factor, as well as Kolmogorov-Smirnov statistics and Hosmer-Lemeshow statistics.

Information value factor (IV) for model expresses the ability of the model to separate division of results for a population of bankrupts and non-bankrupts. It is calculated according to the formula (1) by previously putting objects from the sample in order, sorting them in decreasing order in relation to estimated values of probability of objects affiliation with negative class on the basis of a model (probability of company's bankruptcy).

Gini factor is used to examine superiority of estimated model over random model – randomly made decisions. It is calculated using the formula (2), however, one should firstly put objects in order in research samples in relation to decreasing values of bankruptcy probability. k index present in formulas (1) and (2) means in this case the number of different attributes or categories of variability for values of bankruptcy probability in research samples.

Validation values of IV and Gini factors are interpreted as follows: the higher (closer to 1) the values of these factors, the better the model's ability to correctly classify bankrupts and non-bankrupts. Whereas for models with a strong predictive power they should take values of at least 0.35 or higher.

Kolmogorov-Smirnov statistics value (KS statistics) defines the maximal distance between distribution functions for conditional distributions in population of healthy companies (NB) and bankrupts (B) and is calculated using the formula (Thomas, 2009):

$$KS = \max_{x} \left| F(\mathrm{x} \mid B) - F(\mathrm{x} \mid NB) \right| \tag{7}$$

Divergence also expresses a unit of measure of distance between the scrutinized conditional distributions of bankruptcy probability for both company classes and it is described with the following formula (Thomas, 2009):

$$D = \frac{1}{2}\left(\frac{1}{\sigma_G^{\;2}} + \frac{1}{\sigma_B^{\;2}} \right)\left(\mu_G - \mu_B \right)^2 + \frac{\left(\sigma_G^{\;2} - \sigma_B^{\;2} \right)^2}{2\sigma_G^{\;2}\sigma_B^{\;2}} \tag{8}$$

where: $\mu_{NB} = \sum_{x} x \cdot f(\mathrm{x} \mid NB)$ - the average value of bankruptcy probability in a population of healthy companies (NB), $\mu_B = \sum_{x} x \cdot f(\mathrm{x} \mid B)$ - the average value of bankruptcy probability in a population of bankrupts (B), $\sigma_{NB}^2 = \sum_{x} \left(x - \mu_{NB} \right)^2 \cdot f(\mathrm{x} \mid NB)$, $\sigma_B^2 = \sum_{x} \left(x - \mu_B \right)^2 \cdot f(\mathrm{x} \mid B)$ - variance of bankruptcy probability distribution respectively for the population of healthy companies and bankrupts, $f(\mathrm{x} \mid NB), f(\mathrm{x} \mid B)$ - percentage of healthy and bankrupt companies for a given category of bankruptcy probability.

It is assumed that divergence should take values above 0.5, in order for the scrutinized distributions to lay far enough from each other and the scrutinized model to have

acceptable ability to properly separate bankrupts from companies not threatened with bankruptcy.

Hosmer-Lemeshow statistics is based on Chi-squared statistics and it is calculated using the following formula (Thomas, 2009):

$$HL = \sum_{i=1}^{N} \frac{(n_i p_i - NB_i)^2}{n_i p_i (1 - p_i)}$$
(9)

where: p_i - average probability of affiliation with non-bankrupt class for the given i rating category, NB_i - the number of healthy companies in a given rating category, N - set number of rating categories, into which the range of bankruptcy probability fluctuation has been divided. Hosmer-Lemeshow statistics has a distribution χ^2 with $df = N - 2$ degrees of freedom. The higher the values of H-L statistics, the better the model's ability to differentiate distribution in both populations (B and NB) and better classifying abilities of the model.

ROC concentration curve is a graphic way of presenting classification power of models in correct separation of bankrupted and healthy companies in comparison with the perfect model (having an effectiveness of 100% correct classification) and random model (completely random classification). Measure of conformity with the perfect model is the measure of field under ROC curve $(AUROC = 0.5(Gini + 1))$. The higher (closer to 1) are the values of field under ROC curve, the better predictive ability of the evaluated model.

The previously characterized measures measure the discriminative quality of models. To examine both discriminative quality and calibration precision of models to learning data and test data one uses brier factor (Brier Score) and LL factor (Likelihood of the model).

Brier factor BS is calculated using the following formula (Löffler, Posch, 2007):

$$BS = \frac{1}{n} \sum_{i=1}^{n} (d_i - PD_i)^2$$
(10)

where: n – the number of observations in sample, d_i - dummy variable with value of 1, when company is considered bankrupt and ones with a value of 0 otherwise, PD_i - bankruptcy probability estimated on the basis of a model.

The lower the Brier factor value, the better calibrated is the model for data and it should have better prediction properties.

LL model reliability factor (LL) is defined with the following formula (Prusak, 2005):

$$LL = \prod_{i=1}^{n} P(Y_i \mid X_i) = \prod_{i=1}^{n} PD_i(X_i)^{Y_i} \cdot (1 - PD_i(X_i))^{1-Y_i}$$
(11)

where: n – the number of observations, $PD_i(X_i)$ - estimated bankruptcy probability at given values of entry variables (independent) in a model, Y_i - dummy variable defining (Y=1 – bankrupts) and (Y=1 – non-bankrupts).

The higher the values of classification model reliability for a learning sample, the better it is calibrated on the basis of entry data. High values of reliability indicator for the test sample should point to good classifying value of the model also for new, unknown cases.

The table below (Tab. 5) presents validation statistics for all examined models of predicting bankruptcy of logistics companies

Table 5. Validation parameters of estimated models for a prediction horizon of 1 year

Model	Eff_1 NB	Eff_2 B	IV	K-S	Gini	Divergence	H-L	AUROC	Brier Score	LL (model)
Logit	learning sample									
	88%	96%	3.6	0.83	0.89	8.8	11.2	0.95	0.081	$1,8 \cdot 10^{-6}$
	test sample									
	78%	90%	2.8	0.80	0.91	5.7	3,3	0.95	0.108	$3,0 \cdot 10^{-3}$
Network MLP 26-8-2	learning sample									
	92%	91%	4.0	0.83	0.89	5.3	17.9	0.95	0.152	$1,3 \cdot 10^{-10}$
	test sample									
	89%	100%	2.4	0.89	0.82	2.9	48.2	0.91	0.162	$2,1 \cdot 10^{-5}$
Network MLP 6-3-2	learning sample									
	92%	83%	2.6	0.75	0.86	4.2	13.1	0.93	0.135	$1,5 \cdot 10^{-9}$
	test sample									
	89%	90%	2.8	0.79	0.91	7.0	3.6	0.96	0.111	$8,6 \cdot 10^{-4}$
C&RT Tree	learning sample									
	92%	96%	5.7	0.96	0.93	14.4	7.2	0.97	0.059	$1,8 \cdot 10^{-5}$
	test sample									
	78%	90%	1.3	0.68	0.67	4.2	19.0	0.83	0.140	$1,1 \cdot 10^{-4}$

Source: own study.

Table 6. Validation parameters of estimated models for a prediction horizon of 2 years

Model	Eff$_1$ NB	Eff$_2$ B	IV	K-S	Gini	Divergence	H-L	AUROC	Brier Score	LL (model)
					learning sample					
Logit	74%	81%	1.9	0.58	0.65	2.2	6.0	0.82	0.172	$1,1 \cdot 10^{-19}$
					test sample					
	79%	79%	1.8	0.57	0.70	2.9	4.9	0.85	0.153	$2,8 \cdot 10^{-6}$
					learning sample					
Network MLP 22-17-2	88%	84%	3.6	0.74	0.87	5.9	9.6	0.94	0.103	$4,8 \cdot 10^{-13}$
					test sample					
	100%	86%	3.7	0.86	0.92	10.2	2.4	0.96	0.087	$3,7 \cdot 10^{-4}$
					learning sample					
Network MLP 4-8-2	67%	79%	2.4	0.56	0.70	2.4	14.8	0.85	0.184	$1,8 \cdot 10^{-21}$
					test sample					
	100%	86%	3.7	0.86	0.94	6.8	11.4	0.97	0.167	$4,6 \cdot 10^{-7}$
					learning sample					
C&RT tree	88%	81%	2.8	0.70	0.75	4.2	8.8	0.88	0.127	$2,6 \cdot 10^{-16}$
					test sample					
	71%	71%	1.3	0.50	0.56	0.8	19.0	0.78	0.229	0

Source: own study.

5. CONCLUSIONS

When setting predictions of possible bankruptcy with help of examined models it is worth introducing a separation into two groups, like it was done previously. One of them comprises of predictions made by models estimated on the basis of data for the period of one year before bankruptcy, the second one includes predictions for the same group of models, however estimated on the basis of data for two years until the bankruptcy moment (Tab. 7). In the first case we consider a sample of 82 "healthy" logistics entities from the Podkarpacie region. In the second version the total number of companies amounts up to 61.

Table 7. Average value of predictions in section of examined models for Podkarpacie.

estimation based on data from one year before bankruptcy period	
logit model	0.316692
C&RT model	0.234901
MLP 26-8-2 model	0.456908
MLP 6-3-2 model	0.410757
Average one-year prediction	**0.454873**
Average two-year prediction	**0.596491**
Average three-year prediction	**0.683806**
estimation based on data from one year before bankruptcy period	
logit model	0.354433
C&RT model	0.439643
MLP 4-8-2 model	0.407781
MLP 22-17-2 model	0.298116
Average one-year prediction	**0.374993**
Average two-year prediction	**0.562758**
Average three-year prediction	**0.671097**

Source: own study.

Dividing the above bankruptcy probabilities into two categories, namely one up to a value of 0.5 and the other more than 0.5(non-bankrupt or bankrupt), one needs to underline the fact that among operating logistics companies from Podkarpacie there are no negative indicators for the whole sector. the worst negative premises concerning whole groups of companies are given with help of an artificial neural network *MLP* models, because they exceed value of 0.4 and they are much higher than the two others. In the case of estimations on the basis of data for two years before bankruptcy period one can observe improvement in indications of artificial neural network models and deterioration of classification trees' indications.

When evaluating average predictions for a period from one up to three years it can be generally said that only average two year indications show signals of possible bankruptcy. Of course, the longer the prediction horizon, the higher the bankruptcy possibility.

An interesting comparison in case of prediction values in section of four examined models can be the illustration of a number of signs for possible companies' bankruptcy (Tab. 8).

Table 8. Evaluation of bankruptcy threat scale in a survey of models.

	Number of bankruptcies				
	0	1	2	3	4
	estimation based on data from one year before bankruptcy period				
Podkarpacie	38 (46.34%)	17 (20.73%)	5 (6.10%)	9 (10.98%)	13 (15.85%)
	estimation based on data from two years before bankruptcy period				
Podkarpacie	28 (45.90%)	13 (21.31%)	7 (11.48%)	8 (13.11%)	5 (8.20%)
Scale of bankruptcy risk	Small threat		Average threat	High probability	

Source: own study.

Large number of companies was in the group of so-called small threat in regard to possible predictions concerning bankruptcy. Entities for which none or one of the examined models showed a potential danger of bankruptcy can be included in this class. One needs to underline the fact that high number of companies in a survey of estimated methods does not have any negative indications. It may prove a really solid position of these entities. This group makes more than a half of the researched ones.

When comparing estimations on the basis of data for one year and two years before the bankruptcy period, the second group fares better. In this case one can observe a distinct declining trend for a number of companies in relation to the number of possible negative indications.

An interesting juxtaposition can be made of signals for being threatened with bankruptcy for each researched model (Tab. 9). Here one can conduct a real evaluation of the quality of their indications. It seems that the most stable, without regard to the examined version, is the logit model. The most sensitive are models of artificial neural networks.

Table 9. Enumeration of threats for given models.

	Bankruptcy	
	Yes	No
	estimation based on data from one year before bankruptcy period	
logit model	25 (30.5%)	57 (69,5%)
C&RT model	18 (22.0%)	64 (78.0%)
MLP 26-8-2 model	34 (41.5%)	48 (58,5%)
MLP 6-3-2 model	29 (35.4%)	53 (64.6%)
	estimation based on data from two years before bankruptcy period	
logit model	21 (34.4%)	40 (65.6%)
C&RT model	23 (37,7%)	38 (62,3%)
MLP 4-8-2 model	12 (19.7%)	49 (80.3%)
MLP 22-17-2 model	15 (24.6%)	46 (75.4%)

Source: own study.

When performing a general evaluation of the examined group of models one has to implement the previously used division. So for estimations on the basis of data for one year before bankruptcy period of a given entity the best results are produced by a model of the *C&RT* classification trees, mainly for the so-called learning sample and models of artificial neural networks for the test sample. As for the second group, model of artificial neural networks *MLP 22-17-2* for both test and learning groups can be used.

REFERENCES

[1]   Altman E. I., Haldeman R. G., Narayanan P. (1977), *ZETA ANALYSIS, a new model to identify bankruptcy risk of corporations*, "Journal of Banking and Finance 1, pp. 29–54.

[2]   Atiya A.F. (2001), *Bankruptcy prediction for credit risk using neural networks: A survey and new results*, IEEE Transactions on Neural Networks 12 (4), pp. 929–935.

[3]   Breiman L., Friedman J.H., Olshen R.A., Stone C.J. (1993); *Classification and Regression Trees*, Chapman and Hall.

[4]   Fletcher D., Goss E. (1993), *Application forecasting with neural networks an application using bankruptcy data,* Information and Management 24, pp. 159–167.

[5]   Frydman H., Altman E.I., Kao D. (1985), *Introducing recursive partitioning for financial classification: The case of financial distress*, Journal of Finance 40 (1), pp. 269–291.

[6]   Jones S., Hensher D.A. (2004), *Predicting firm financial distress: A mixed logit model*, Accounting Review 79 (4), pp. 1011–1038.

[7]   Karels G.V., Prakash A.J. (1987), *Multivariate normality and forecasting of business bankruptcy*, Journal of Business Finance and Accounting 14 (4) (1987).

[8]   Kaski S., Sinkkonen J., Peltonen J. (2001), *Bankruptcy analysis with self-organizing maps in learning metrics*, IEEE Transactions on Neural Networks 12 (4), pp. 936-947.

[9]   Kiviluoto K. (1998), *Predicting bankruptcies with self-organizing map*, Neurocomputing 21, pp. 191–201.

[10]   Kolari J., Glennon D., Shin H., Caputo M. (2002), *Predicting large US commercial bank failures*, Journal of Economics and Business 54 (32 1), pp. 361–387.

[11]   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 (2007), pp. 1-28.

[12]   Lam M. (2004), *Neural networks techniques for financial performance prediction: integrating fundamental and technical analysis*, Decision Support Systems 37, pp. 567–581.

[13]   Lee K., Booth D., Alam P. (2005), *A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms*, Expert Systems with Applications 29, pp. 1–16.

[14]   Leshno M., Spector Y. (1996), *Neural network prediction analysis: The bankruptcy case*, Neurocomputing 10, pp. 125–147.

[15]   Löffler G., Posch P., N. (2007), *Credit risk modeling using Excel and VBA*, Wydawnictwo Wiley, Chichester, West Sussex, pp. 156.

[16]   Marais M.L., Patel J., Wolfson M. (1984), *The experimental design of classification models: An application of recursive partitioning and bootstrapping to commercial bank loan classifications*, Journal of Accounting Research 22, pp. 87–113.

[17]   Martin D. (1977), *Early warning of bank failure: A logit regression approach*, "Journal of Banking and Finance", 1, pp. 249–276.

[18]   Matuszyk A. (2004), *Credit scoring – metoda zarządzania ryzykiem kredytowym*, Wydawnictwo CeDeWu, Warszawa, pp. 119-122.

[19]   Ohlson J.A. (1980), *Financial rations and the probabilistic prediction of bankruptcy*, Journal of Accounting Research 18, pp. 109–131

[20]   Prusak B. (2005), *Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw*, Wydawnictwo Difin, Warszawa, pp. 50.

[21]   Serrano-Cinca C. (1996), *Self-organizing neural networks for financial diagnosis*, Decision Support Systems 17, pp. 227–238.

[22]   Tam K.Y., Kiang M. (1992), *Predicting bank failures: A neural network approach*, Decision Sciences 23, pp. 926–947.

[23]   Thomas L. C. (2009), *Consumer credit models. Pricing, Profit and Portfolios*, Oxford University Press, Oxford, pp. 111.

[24]   Wilson R.L., Sharda R. (1994), *Bankruptcy prediction using neural networks*, Decision Support Systems 11, pp. 545–557.

[25]   Witkowska D. (2002), *Sztuczne sieci neuronowe i metody statystyczne. Wybrane zagadnienia finansowe*, C.H. Beck, Warszawa, pp. 86-87.

[26]   Yu L., Wang S., Lai K. K., Zhou L. (2008), *Bio-Inspired Credit Risk Analysis. Computational Intelligence with Support Vector Machines*, Springer-Verlag, Berlin Heidelberg, pp. 14-15.

Notes

Note 1. site.securities.com

PROGNOZOWANIE UPADŁOŚCI FIRM Z SEKTORA LOGISTYCZNEGO DZIAŁAJĄCYCH W REGIONIE PODKARPACIA

W artykule przeprowadzono badanie skuteczności różnych koncepcji modelowania upadłości przedsiębiorstw z sektora logistycznego. W celu pełniejszego zobrazowania zagadnienia ww. prognozowanie ewentualnych negatywnych skutków prowadzonej działalności przeprowadzono dla wszystkich firm ww. sektora działających w regionie Podkarpacia. Analiza została poparta danymi pochodzącymi z bazy danych *EMIS* (*Emerging Markets Information Service*). Szeroka gama 28 wskaźników finansowych została pogrupowana na pięć grup wskaźników tj. odpowiednio wskaźniki płynności, zyskowności, zadłużenia, sprawności działania oraz finansowe. Wyżej wspomnianą próbę badawczą podzielono na grupę przedsiębiorstw chorych – w stosunku co do których ogłoszono upadłość – oraz grupę tzw. firm zdrowych (sprawie działających, o dobrej kondycji finansowej. Podejście takowe pozwala na lepszą, obiektywną ocenę stosowanych metod w zakresie modelowania upadłości. Celem niniejszej publikacji była zatem chęć znalezienia czynników (modeli) opisujących ryzyko upadłości przedsiębiorstw w kontekście ich skuteczności przewidywań w horyzoncie jedno- i dwu letnim. Zastosowano w tym wypadku modele regresji logistycznej, drzew klasyfikacyjnych oraz dwóch wariatów sztucznych sieci neuronowych. Pełnej oceny zastosowanych modeli dokonano w procesie walidacji. Podstawowym narzędziem stosowanym w tym wypadku do badania efektywności klasyfikacyjnej modeli klasyfikacyjnych są macierze poprawnych klasyfikacji. Dokonano zatem oszacowania poprawnych oraz błędnych odsetków wskazań modeli zarówno w grupie wskazanych wcześniej przedsiębiorstw zdrowych jak i chorych. Ostatecznie przeprowadzono ocenę poruszanych na łamach artykułu metod oraz ogólnej kondycji sektora logistycznego w rejonie Podkarpacia.

Słowa kluczowe: upadłość, sektor logistyczny, modelowanie, wskaźniki finansowe.