A parametric or nonparametric approach for creating a new bankruptcy prediction model: The Evidence from the Czech Republic

Michal Karas, Mária Režňáková

***Abstract*—** For many years now the development of models capable of predicting company bankruptcy has aimed at increasing their accuracy. Among the decisive factors determining the accuracy of the bankruptcy model have been the choice of variable models and applied classification algorithms. The prevailing opinion in literature is that the accuracy of bankruptcy models cannot be appreciably improved by the choice of classification algorithm. A reflection of this assertion is the frequent usage of parametric methods. In particular this involves the method of linear discrimination analysis. This method formed the basis of the first bankruptcy model and continues to be the most frequently applied classification algorithm. However, it requires the fulfilment of assumptions which financial data does not provide and therefore limits the improvement of the model's predictive capabilities. This led the authors to the idea of testing the possibility of improving the bankruptcy model's predictive capabilities by using non-traditional approaches. Using the example of companies from the Czech Republic it was discovered that a nonparametric method, when used for the selection of model variables as well as the actual classification, can yield significantly better results than the traditional parametric approach.

Keywords— *bankruptcy prediction models, boosted trees, stepwise discriminant analysis.*

I. INTRODUCTION

THE bankruptcy of a company represents a sizeable economic loss not only for the owners of the company and its creditors, but for society as a whole. In an effort to identify the risk of bankruptcy, many researchers have investigated bankruptcy modelling and tried to improve the accuracy with which the threat of bankruptcy can be detected. The accuracy of the model is a critical feature of the model.

This is due to the fact that even smaller improvements can lead to significant economic savings when applied to a larger portfolio.

Building such a model involves two phases: finding suitable variables (called predictors), and choosing a classification algorithm that can effectively utilize those predictors.

The objective of this paper is to determine how the effectiveness of a bankruptcy model is influenced by the choice of a method, specifically a Linear Discriminant Analysis Method (hereinafter LDA method) and a Boosted Trees Method (hereinafter BT method).

II. LITERATURE REVIEW

Historically, various algorithms have been employed to devise models of bankruptcy. The first was the LDA method [2], developed by Fisher [20]. In reaction to its shortcomings, other algorithms were applied. A number of parametric methods exist, such as the probit model [47] the logistic regression or logit model [33, 36] and the Cox model [26, 39]. The probit and the logit model are applications of the inverse density function of the normal or logistic distribution. The probit model can be written in the form, see [24]:

$$P_i = \int_{-\infty}^{\alpha+\beta x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt \qquad (1)$$

α, β are estimated parameters,
x is the vector of independent predictors (here financial indicators),
P_i is the probability of default (bankruptcy),

The logit model can be written in the form, see [24]:

$$P_i = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \qquad (2)$$

The Cox hazard model can be generally written in the form [39],

$$\phi(t, x; \theta) = \frac{f(j, x; \theta)}{S(t, x; \theta)} \qquad (3)$$

M. Karas is with the Department of Finance, Faculty of Business and Management, Brno University of Technology, Kolejní 2906/4, 620 00, Brno, Czech Republic, (corresponding author, phone: 00420 54114 3708, email: Karas@fbm.vutbr.cz)

M. Režňáková is with the Department of Finance, Faculty of Business and Management, Brno University of Technology, Kolejní 2906/4, 620 00 Brno, Czech Republic, (email: Reznakova@fbm.vutbr.cz)

where

θ represents the vector of parameters,
x_i is the vector of explanatory variables used in a bankruptcy prognosis,
t_i is the moment when the i-th firm goes bankrupt,
f is the density of probability,

S represents the so-called Survival Function, which expresses the probability that in the interval <1,2,3…t> leads to bankruptcy. On the assumption of discretely defined time, the Survival Function can be written in the form:

$$S(t,x;\theta)=1-\sum_{j<t} f(j,x;\theta) \quad (4)$$

Over time, nonparametric methods were also tried, such as artificial neural networks (ANN) [1, 5, 32], Data Envelopment Analysis (DEA) methods [13, 18], and even some combinations of parametric and non-parametric methods [16, 27].

Nevertheless, the LDA method remains the most widely used classification algorithm [3].

One of the reasons may be the prevalent opinion that the choice of classification method does not offer much potential to improve the bankruptcy model [34], i.e. that there is not a statistically significant difference in accuracy between the individual methods [3].

III. Sample and Methodology

In the course of this research, a total of 44 financial indicators were tested. They had been used in previous bankruptcy studies, especially those of [2, 6, 7, 17, 36, 18, 42, 34, 40, 38], see table II. The sample consisted of 1,908 enterprises from manufacturing industries (1,500 active and 408 bankrupt) that operated within the Czech Republic in the period 2004-2011. The data was obtained from Amadeus Database and the calculations utilized Statistical 10 statistical software.

Table I Descriptive statistics on concerns investigated (in thousands of CZK)

	n	Mean	Median	Quant. (1%)	Quant. (99%)	Std. Dev.
TA (A)	1434	693077	233526	18488	7497559	2092870
TA (B)	190	32396	11600	0	374979	60753
S(A)	1434	1017713	306901	50109	9908315	3974954
S (B)	190	47150	13406	0	725220	100096

Source: Our own analysis of data from the Amadeus database,
note: TA – total assets, S – sales, A – active company, B – bankrupt company.

Table II List of researched variables

Ratio	**Shortcut**	**Ratio**	**Shortcut**
Cash flow/sales	CF/S	Ohlson's change of NI	NI-change
Cash flow/total asset	CF/TA	Operation cost/operation revenue	OC/OR
Cash flow/total liabilities	CF/TL	Operation profit (loss)/ average capital (3-year average)	OP/AC
Current liabilities/sales	CL/S	Operation profit (lost)/operation revenue	OP/OR
Current ratio	CR	Operation revenue/current assets	OR/CA
Debt-Equity ratio	DR	Operation revenue/current liabilities	OR/CL
EBIT (5-years volatility)	EBIT (5-vol)	Operation revenue/fixed assets	OR/FA
EBIT/interest	EBIT/Int.	Operation revenue/long-term liabilities	OR/LTL
EBIT/total asset (=ROI)	EBIT/TA	Operation revenue/total assets	OR/TA
EBITDA/interest	EBITDA/Int.	Operation revenue/total liabilities	OR/TL
EBITDA/total liabilities	EBITDA/TL	Profit margin (3-year average)	PM
Fixed assets/long-term liabilities	FA/LTL	Quick asset/sales	QA/S
Income (loss) before tax/operation revenue	EBT/OR	Retained earnings/total asset	RE/TA
Intangible assets/total assets	Int. A/Tot. A	Sales/Total Assets	S/TA
Log of equity	EQ	Tangible assets/total assets	Tan. A/Tot. A
Log of sales	S	Total liabilities /EBITDA	TL/EBITDA
Log of total assets	TA	Total liabilities/total assets	TL/TA
Net income (loss)/average capital (3-year average)	NI/AC	Working capital/operating cost	WC/OC
Net income/current assets	NI/CA	Working capital/sales	WC/S
Net income/fixed assets	NI/FA	Working capital/total assets	WC/TA
Net income/operation revenue	NI/OR	1 if net income was negative for past two years, 0 otherwise	NI < 0
Net income/total asset (= ROA)	NI/TA	1 if total liabilities > total assets, 0 otherwise	TL>TA

Source: Our own work based on the literature [2, 6, 7, 17, 36, 18, 42, 34, 40, 38]

Due to strong correlation with the other indicators, the following 10 indicators were omitted: *CF/TL*, *NI/TA*, *NI/FA*, *NI/CA*, *WC/S*, *WC/OC*, *EBT/OR*, *OP/OR*, *OR/TA* and *EBIT/Int.*

Strong correlation is considered to be a Spearman coefficient correlation higher than 0.9. Overall, a total of 34 indicators were analysed.

A. *LDA Method*

The objective of this method is, according to [25], *"to find a **linear** combination of p monitored predictors, i.e. Y = bTx, where $b^T = [b_1, b_2, \dots, b_p]$ is a vector of parameters that would segregate better than any other linear combination the H groups under consideration, so that its variability within the groups would be minimal and its variability between the groups maximal."*

The LDA method produces a discriminatory rule (function) which according to calculated predictors assigns each company to a group of enterprises either threatened or not threatened by bankruptcy.

Discriminant analysis works with the assumption of multivariate normal distribution of data. The density of probability of multivariate normal distribution of a variable x can be written as follows [15, p. 108]:

$$f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right] \quad (5)$$

where
x is the vector of independent predictors, where $x = (x_1, x_2, \dots, x_p)$,
μ_k is the vector of middle values of the quantity x k-th group,
Σ_k is the covariance matrix of the k-th group,

The Linear discriminant analysis (LDA) is a special kind of discriminant analysis which adds the assumption of identical covariance matrices (Σ_k). Under these assumptions the discriminant rule, based on the Mahalanobis distance, can be written as follows [25]:

$$x^T \Sigma^{-1}(\mu_1 - \mu_2) > 1/2(\mu_1 + \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) \quad (6)$$
→ group 1 (e.g. active)

$$x^T \Sigma^{-1}(\mu_1 - \mu_2) < 1/2(\mu_1 + \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) \quad (7)$$
→ group 2 (e.g. bankrupt)

where
π_1 or. π_2 is apriori the probability of units belonging to the group corresponding to the range group 1 or 2,

Where the assumptions are not fulfilled for the identical covariant matrices (Σ_k) represents a quadratic form of discriminant analysis (QDA), a more suitable discriminant rule. However, the disadvantage of QDA is its significant sensitivity to deviations from the normality and for this reason LDA is more frequently applied.

The factors beneficial for the accuracy of the LDA method are: at least roughly normal distribution of data [37], negatively correlated indicators [2, 12], and the absence of extreme values [44, 45, 46].

B. *Boosted Trees Method*

The method of Boosted Trees (BT) is a combination of the classification and regression trees method (CART) [11], with a boosting algorithm introduced by J. Friedman [22]. Using the boosting algorithm raises the accuracy of the classification algorithm, to which it is applied by progressively reducing the error term [4, 11, 22]. The resultant classification rule represents a set of many "weak" learners. The boosting algorithm is most often applied to CART, but an ANN application may be encountered as well [32].

Classification and Regression Trees (CART)

The basic idea behind the Trees is the division of a complex problem of feature space in a set of smaller parts known as regions (R), which is possible to describe through simpler models (for example, constants). For a two-dimensional classification problem it is possible to describe the approach of such a division using the following schemata (see Fig. 1 and 2). These schemata document the division of two-dimensional *feature space* in the mentioned regions using the constant *t*.

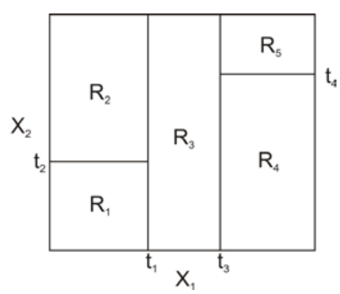

Fig. 1, The division of two-dimensional space into regions
Source: Own modification according to [15,p. 306]

Alternatively, the same division can be shown using trees, as in the following schema.

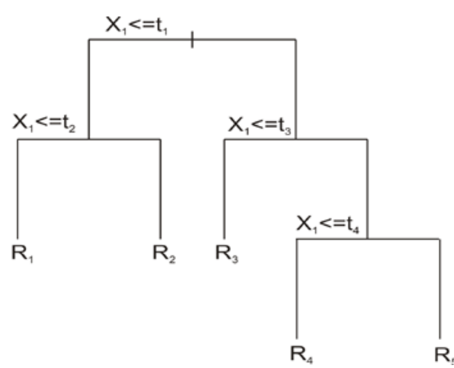

Fig. 2 Division of two-dimensional space into regions using trees
Source: Own modification according to [15,p. 306]

The central problem of the method of using trees is establishing the optimal divisional boundaries *t* between those regions *R*. The boundaries are established in such a way that the demarcated regions, or the trees, fulfilled specific defined properties.
This property of the regions, or the trees, is defined as a *node impurity* and the aim of the method is its minimalization.

For classification purposes, where the output can take the value *1, 2, …, K*, it is possible to describe *node impurity* in the following way, see [15,p. 306].

In the *m*-th node, representing the *m*-th region R_m with N_m, the number observed is a proportion of the group *k* in the node *m*, given by the relation:

$$\hat{p}_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = k) \tag{8}$$

Then it is necessary to define the majority of observed elements of the *k*-th group in the node m as:

$$k(m) = \arg\max x_k\, \hat{p}_{mk} \tag{9}$$

Node impurity of the tree T or $Q_m(T)$ can be defined using several standards, the following is used:

Misclassification error

$$\frac{1}{N_m} \sum_{i \in R_m} I(y_i \neq k(m)) = 1 - \hat{p}_{mk(m)} \tag{10}$$

Gini index

$$\sum_{k \neq k'} \hat{p}_{mk} \hat{p}_{mk'} = \sum_{k=1}^{K} \hat{p}_{mk} \cdot (1 - \hat{p}_{mk}) \tag{11}$$

Cross-entropy or deviance

$$-\sum_{k=1}^{K} \hat{p}_{mk} \log \hat{p}_{mk} \tag{12}$$

Deviance as a level of node impurity was used here as part of the presented research.

Boosting

Boosting is a general approach for making the final deciding rules as a set of several "weak" rules or classifiers. Amongst the boosting algorithms AdaBoost.M1 is one most frequently applied, see [21], the principle of which will be described further.

Let us consider a classification problem with a dichotomous dependent variable *Y*, i.e. $Y \in \{-1;1\}$ and a vector of independent predictors *X* and a classifier *G(X)*, which can only take the values -1 and 1, i.e. $G(X) \in \{-1;1\}$. Error rate for the training sample is given by the relationship, see [15, p. 337]:

$$\overline{err} = \frac{1}{N} \sum_{i=1}^{N} I(y_i \neq G(x_i)) \tag{13}$$

The basis of boosting is the gradual application of the classifier *G(X)* to the repeatedly modified version of data and thus gradually produce other *M* "weak" classifiers $G_m(X)$, $m = 1, 2, \ldots, M$. It is possible to describe the method of boosting algorithms in the following schemata, see [15, p. 338].

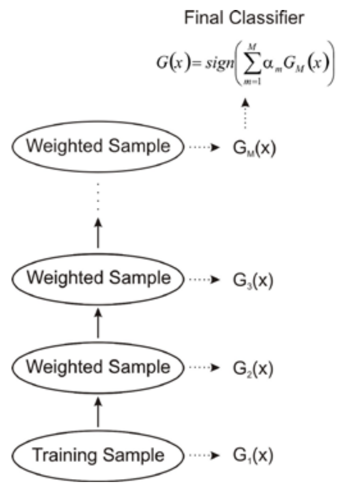

Fig. 3, AdaBoost algorithm method
Source: Own modification according to [15, p. 338]

The resulting classifier $G_{final}(X)$ is then made up of the individual partial rules $G_m(X)$, which are given the weights α_m. The output is standardized to attain a value of only *-1* or *1*, see [15, p. 338].

$$G_{final}(x) = sign\left(\sum_{m=1}^{M} \alpha_m G_M(x)\right) \tag{14}$$

The weights $\alpha_1, \alpha_2, \ldots, \alpha_M$ are calculated using a boosting algorithm, representing the partial contribution of each classifier $G_m(X)$. The modification of data in each step of the boosting algorithm is the application of the weights $w_1, w_2, \ldots, w_N$ for each pair of training data (x_i, y_i), where $i = 1, 2, \ldots, N$. At the start of the algorithm the weights are set at the value $w_i = 1/N$. In every other iteration $m = 2, 3, \ldots, M$ the weights of individual observations are adjusted. In the *m*-th iteration the weights of those observations which had been wrongly classified in the previous step are increased by the classifier $G_{m-1}(X)$, while the weights of those which were successful are lowered. By this method the wrongly classified observation is given more attention in order to increase the accuracy of the whole rule. The algorithm *Adaboost.M1.* can be described as follows, see [15, p. 338-339]:

1. Set the observation weights to the default value $w_i = 1/N$, where $i = 1, 2, \ldots, N$.
2. From $m = 1$ to $m = M$:

   a) Fit a classifier $G_m(X)$ for the training data using the weights w_i.
   b) Calculate

$$err_m = \frac{\sum_{i=1}^{N} w_i I(y_i \neq G_m(x_i))}{\sum_{i=1}^{N} w_i} \tag{15}$$

   c) Calculate $\alpha_m = \log((1 - err_m)/err_m)$.
   (16)

d) Set $w_i \leftarrow w_i \cdot \exp[\alpha_m \cdot I(y_i \neq G_m(x))], i = 1,2,...,N.$ (17)

3. Output

$$G(x) = sign\left(\sum_{m=1}^{M} \alpha_m G_M(x)\right) \quad (18)$$

A useful feature of this method is that it allows the sorting out of the variables x_j according to their relative influence I_j on the variability of the approximation function $\hat{G}(x)$ across the entire division of input predictors, this measurement can be described as follows, see [22]:

$$I_j = \left(E_x\left[\frac{\partial \hat{G}(x)}{\partial x_j}\right] \cdot \mathrm{var}_x[x_j]\right)^{1/2} \quad (19)$$

Among the advantages of the BT method, aside from its nonparametric nature (the data need not be normally distributed), is its tolerance for outliers in the input variable space [41]. In addition, the method can even capture non-linear relationships between the variables [23]. Since the lack of normality and the presence of outliers tend to be commonplace in financial data [8, 9, 39, 43], it can be expected that a method which is immune to these aspects will deliver higher classification accuracy. In other words, we assumed that the BT method would produce better results than the LDA method. To better assess the potential of these methods, we will use them for the selection of suitable variables as well as for the classification algorithm itself.

C. Variance Inflation Factor

Each variable included in the model should bring with it unique information. However, the situation may arise when the information which the variable contains can be explained with a high degree of accuracy using a combination of other variables.

From a statistical viewpoint this problem is termed multicollinearity. One of the measurements for evaluating the degree (i.e. severity) of the multicollinearity problem is the *Variance Inflation Factor* (VIF) method. The VIF method is one of the tools of the diagnostics of the general linear regression model. The VIF value tells you how much the variance of the independent variable can be explained by a combination of other variables [14]. The general linear regression model can be written as follows:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \ldots + \beta_k \cdot X_k + \varepsilon \quad (20)$$

where

Y is a dependent (response) variable,

$X_1, X_2, \ldots, X_k$ are independent (explanatory variables),

ε is the error model.

The VIF value for a given indicator, for example X_1, is calculated in two steps:

1) Form a regression model with X_1 as a dependent variable and $X_2, X_3, \ldots, X_k$ as independent variables or:

$$X_1 = \alpha_0 + \alpha_1 \cdot X_2 + \alpha_2 \cdot X_3 + \ldots + \alpha_k \cdot X_k + \varepsilon_2 \quad (21)$$

2) To determine the value of the index determination (R^2) of the *i*-th VIF model then represents the transformation of this value as follows:

$$VIF = \frac{1}{1 - R_i^2} \quad (22)$$

Alternatively there can be used so-called *tolerance*, where *tolerance* = *1- R*^{2}. Tolerance represents the unique contribution of the given variable towards the overall explanatory power of the model.

For values of the VIF coefficient lower than 10 or 4, the multicollinearity can be considered to be insignificant [32]. The value of the VIF coefficient greater than 10 or 4 points to a situation where it is possible to express a partial independent variable, for example, X_1, through the linear combination of the other independent variables ($X_2, X_3, \ldots, X_k$), which explain 90% or 75% of the value of variable X_1, i.e. a tolerance of 0,1 or 0,25. If omitting the variable X_1, which reaches VIF higher than 10 or 4, from the original model, then there remains within the model at least 90, or 75% of the value of its information.

IV. RESULTS

We derived two differing bankruptcy models in order to fulfil the aim of our study. The first model uses a parametric algorithm (i.e. the LDA method), while the second model uses a non-parametric algorithm (i.e. the BT method). We employed the investigated methods both to find significant variables (model predictors) and for the purposes of classification itself.

The data sample under investigation was modified fifty times to enable a statistical assessment of which method leads to the derivation of a more precise model. The given sample modification consisted of dividing the sample of data on active and bankrupt companies into 50 parts. 50 differing data samples were obtained by the progressive omission of 1/50 of the observations on active or bankrupt companies. The models created were again applied to each of these data samples and their accuracy studied. In this way, we obtained 50 observations of the accuracy of models applying the LDA method and 50 observations of models applying the BT method. A non-parametric Mann-Whitney U-test was applied to compare the difference between the classification accuracy of the given methods.

We now turn to details on the drawing up of the individual models.

A. Bankruptcy Model Derived from the LDA Method (Model LDA)

Finding the suitable predictors was done with a stepwise **forward** variant of the LDA method (see Model LDA 1), where only the variables with sufficient discriminatory ability are included in the model [5, 28]. The model thus contained 22 out of 34 analyzed variables. Only 13 variables were statistically significant by the F-test at 5% level. See the following table III.

Table III Model LDA 1

Variable	Wilk. lambda	Parc. Lambda	F to rem.	p-val.	Toler.
EQ	0.7299	0.9919	7.88	0.005104	0.079
DR	0.7392	0.9795	20.30	0.000007	0.562
NI/OR	0.7483	0.9676	32.50	0.000000	0.875
S/TA	0.7241	1.0000	0.04	0.850833	0.165
EBIT(5-vol)	0.7445	0.9726	27.37	0.000000	0.545
NI-change	0.7292	0.9929	6.89	0.008807	0.741
PM	0.7265	0.9966	3.28	0.070244	0.679
WC/TA	0.7445	0.9725	27.45	0.000000	0.233
OR/CA	0.7268	0.9962	3.72	0.054059	0.286
OR/CL	0.7352	0.9848	14.95	0.000118	0.396
OR/FA	0.7276	0.9950	4.83	0.028231	0.653
EBITDA/TL	0.7347	0.9855	14.31	0.000165	0.374
CF/TA	0.7308	0.9907	9.10	0.002617	0.360
RE/TA	0.7267	0.9963	3.57	0.059034	0.483
S	0.7260	0.9973	2.64	0.104816	0.117
QA/S	0.7277	0.9950	4.85	0.027860	0.451
CD/S	0.7283	0.9942	5.66	0.017552	0.369
IntA/TotA	0.7280	0.9946	5.25	0.022142	0.921
TanA/TotA	0.7265	0.9966	3.31	0.069144	0.240

Source: Our own analysis of data from the Amadeus database

As a whole, the model LDA 1 resulting from forward discriminant is statistically significant to 1% of the F-test level (see the general characteristics of the model – Wilks' lambda 0.72404, F (22,970) = 16.805, p<0.0000).

The model variables, which are statistically significant to at least 5% of the F-test level, were tested for severe multicollinearity. According to the tolerance of values it is possible to discard as redundant the variable EQ because the other variables explain 92.1% of the information which this variable contains. No other variable in the model can be labelled redundant (i.e. exhibit a higher tolerance than 0.1, which corresponds to a VIF value lower than 10).

For comparison the LDA model was derived from **backward** stepwise discriminant analysis (see Model LDA 2). The result is a model of 10 variables. All 10 variables have a statistical significance at a level of 1%. See the following table IV.

Table IV Model LDA 2

Variable	Wilk. lambda	Parc. lambda	F to rem.	p-val.	Toler.
CF/TA	0.773	0.978	22.348	0.000003	0.557
DR	0.795	0.951	50.743	0.000000	0.923
WC/TA	0.770	0.982	17.747	0.000028	0.443
RE/TA	0.769	0.983	17.016	0.000040	0.652
NI/OR	0.778	0.972	28.415	0.000000	0.938
OR/CL	0.773	0.978	22.447	0.000002	0.575
TA	0.836	0.905	103.565	0.000000	0.516
EBIT(5-vol)	0.778	0.972	27.869	0.000000	0.566
EBITDA/TL	0.766	0.988	12.192	0.000501	0.415
TanA/TotA	0.770	0.982	18.087	0.000023	0.618

Source: Our own analysis of data from the Amadeus database

Model LDA 2 derived from backward discrimination has an overall statistical significance of 1% of the F-test level. Its overall discriminatory ability is slightly lower compared to the previous version, though when connected to less than half of the variables (see the overall characteristics of the Wilks' lambda model: 0.75619 approximately F (10,982) = 31,661 p<0,0000). According to the tolerance values the model contains one redundant variable - *DR*. The other variables are not considered to be redundant because they exhibit a tolerance higher than 0.1% which corresponds to a VIF value lower than 10.

B. Bankruptcy Model Derived from the BT Method (Model BT)

When setting up the model it was first of all necessary to derive a model with the use of all variables (model BT 1). Calculating this model (through boosting algorithm) is documented by the following Fig. 4.

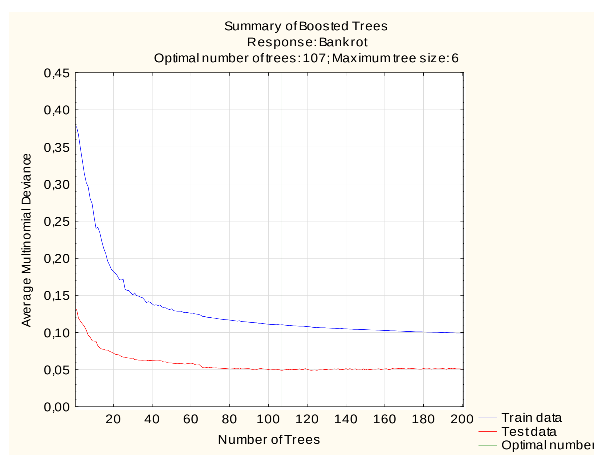

Fig. 4 Calculating model BT 1

In total Model BT 1 is made up of 107 individual trees (CART). It is the number at which the minimum *node impurity* or *deviance* is reached. The following graphs 2 and 3 present examples from individual trees (CART) for the group of active companies (category: 0), and for the group of bankrupt businesses (category: 1).

The example of the tree for the group of active companies (Fig. 5) shows a significantly simpler structure than the example of the graph for the category of bankrupt companies (Fig. 6).

The aforementioned simplicity of the structure lies in the higher number of nodes with zero variance. These nodes are now homogenous and it is not necessary to reduce them further, unlike with the example from the bankrupt companies. In accordance with the literature, see [15, p. 363], the overall number of terminal nodes was limited up to 8. The parameter of the number of terminal nodes determines the maximum number of iterations between variables.

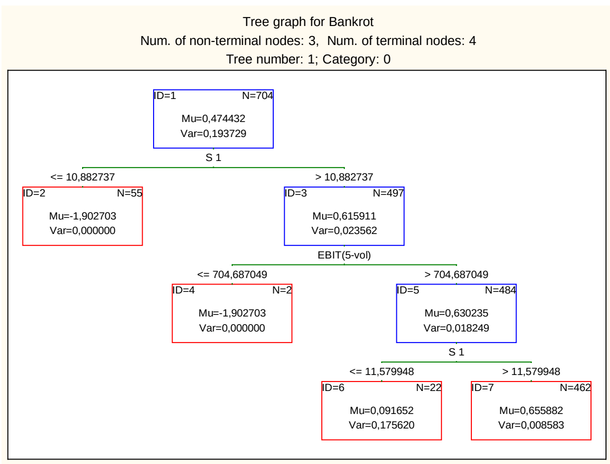

Fig. 5 Example of an individual tree (classifier) model BT 1 for active companies

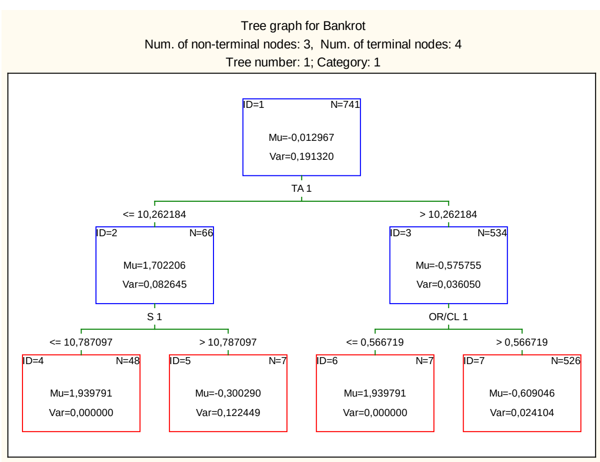

Fig. 6 Example of the individual tree (classifier) model 1 for bankrupt companies.

The BT method allows the ranking of predictors by their relative significance (the degree of their contribution to final classification capability).

Table V Relative importance of variable model BT 1

Variable	RI [%]	Variable	RI [%]
S	100	EBITDA/TL	55.38
TA	97,48	OC/OR	55.32
NI/OR	85,06	OR/FA	54.9
OR/CL	81,93	CF/S	53.27
TL/TA	78,05	OR/LTL	52.79
EBIT(5-vol)	75,8	NI/AC	47.62
DR	75,66	CL/S	47.55
OR/TL	74,88	OP/AC	42.98
S/TA	73,57	IntA/TotA	42.97
EQ	71,43	EBITDA/Int.	39.82
OR/CA	64,97	TanA/TotA	38.74
CR	64,46	PM	37.13
TL/EBITDA	60,58	TL>TA	35.05
WC/TA	58,55	FA/LTL	29.3
RE/TA	58,29	QA/S	27.49
CF/TA	58,13	NI-change	25.6
EBIT/TA	56,66	NI<0	19.32

Source: Our own analysis of data from the Amadeus database

An analysis showing the individual variables' representation in the intervals of their significance showed that their distribution was rather uneven. The relative importance of variables in a bankruptcy assessment differs greatly. A significance higher than 60% is achieved only by 38.24% of predictors or 13 out of 34, which appear in the following table. Those were the only predictors used to build Model BT 2.

Table VI Relative importance of variables for model BT 2

Variable	RI*[%]	Variable	RI*[%]
TA	100.00	TL/TA	60.58
S	99.26	EQ	59.39
NI/OR	77.00	OR/CA	54.85
EBIT(5-vol)	71.43	CR	54.79
OR/CL	65.47	DR	54.00
OR/TL	63.86	TL/EBITDA	36.53
S/TA	63.44		

*RI = relative importance of given variables.
Source: Our own analysis of data from the Amadeus database

Next was an analysis for the presence of multicollinearity in the model. The Variance Inflation Factor (VIF) method was again employed for this purpose. The indicators with a value greater than 4 were removed from the model [32]. Eight additional variables were removed in this manner.

Table VII VIF values for variable model BT 2

Variable	Tolerance	VIF	R^2
CR	0.21476	4.65636	0.78524
DR	0.437784	2.28423	0.562216
TL/TA	0.102836	9.42	0.897164
NI/OR	0.96393	1.03742	0.03607
OR/CA	0.421457	2.37272	0.578543
OR/CL	0.179076	5.58423	0.820924
OR/TL	0.175182	5.70834	0.824818
TA	0.028281	35.35924	0.971719
EQ	0.031405	31.84257	0.968595
S	0.046411	21.54665	0.953589
TL/EBITDA	0.917513	1.0899	0.082487
EBIT(5-vol)	0.581655	1.71923	0.418345
S/TA	0.145163	6.88879	0.854837

Source: Our own analysis of data from the Amadeus database

After removing the less important variables and those that were redundant (because of multicollinearity), the Model BT emerged in its final form with only 7 variables (see Model BT 3).

Table VIII Relative importance of the variable model BT 3

Variable	RI [%]
DR	100.00
NI/OR	97.14
EBIT(5-vol)	77.23
TL/EBITDA	69.35
OR/CA	61.08

Source: Our own analysis of data from the Amadeus database

The following table contains the minimal values for the loss function attained. This value demonstrates the informative quality of the model, referred to as the "goodness of fit".

Table IX Level of goodness of fit with BT models

Model	Train		Test	
	Risk estimate	Standard error	Risk estimate	Standard error
Model BT 1 (34 var.)	0.04271	0.00507	0.01582	0.00702
Model BT 2 (13 var.)	0.0589	0.00616	0.02232	0.00698
Model BT 3 (5 var.)	0.06528	0.00651	0.03633	0.00865

Source: Our own analysis of data from the Amadeus database

C. *Accuracy Comparison of Model 1 and Model 2*

The models were evaluated on the given sample. The accuracy of the BT models was calculated as the weighted average between the accuracy of the training and test sample, where the weights are the number of observations in the sample. The results of testing (percentage of correctly identified cases) are shown in the following table.

Table X Comparing the accuracy of the LDA and BT models

Model	Active [%]	Bankrupt [%]	Type I error [%]	Type II error [%]
LDA 1 (22 var.)	98.36	45.61	45.83	2.30
LDA 2 (10 var.)	98.92	51.19	25.86	2.89
Average	**98.64**	**48.40**	**35.85**	**2.59**
BT 1 (34 var.)	94.39	99.75	14.47	1.32
BT 2 (13 var.)	93.36	96.16	15.88	1.48
BT 3 (5 var.)	92.30	96.65	17.76	1.32
Average	**93.35**	**97.52**	**16.4**	**1.37**

Source: Our own analysis of data from the Amadeus database

A type I error occurs when a bankruptcy-prone company is assessed as financially stable. A type II error describes the opposite situation, i.e. perceiving a financially stable company as facing bankruptcy. According to [48], the type I error is 2 to 20 times more serious (thus more costly) than the type II error.

The test results indicate that Model 2, generated by the nonparametric BT method, is clearly superior in its ability to identify the risk of bankruptcy, relative to Model 1 based on the parametric LDA method.

D. *Testing the LDA and BT models on modified versions of data*

The content of the following table are descriptive statistics of accuracy (in the form of an index) of the LDA and BT models. The abbreviations LDA and BT indicate the method applied; (Act), (Bankr) and (Total) indicate accuracy on a sample of active and bankrupt companies and total accuracy.

Table XI Descriptive statistics on the accuracy of the BT and LDA models on modified versions of data (values in the form of an index)

	Mean	Median	Min.	Max.	Std. Dev.
BT (Act)	0.939701	0.938776	0.933288	0.952381	0.003863
BT (Bankr)	0.928350	0.928750	0.890000	0.967500	0.023061
BT (Total)	0.941970	0.942108	0.915809	0.964957	0.013785
LDA (Act)	0.996284	0.996340	0.994920	0.997080	0.000462
LDA (Bankr)	0.510446	0.511905	0.487180	0.544304	0.009529
LDA (Total)	0.968692	0.968287	0.967564	0.972242	0.001110
type. I. Err (LDA).	0.489554	0.488095	0.455696	0.512821	0.009529
type. II. Err (LDA).	0.003716	0.003660	0.002920	0.005080	0.000462
type. I. Err (BT).	0.042577	0.043416	0.005450	0.087193	0.022425
type. II. Err (BT).	0.078480	0.078488	0.058651	0.087125	0.004777

Source: Our own analysis of data from the Amadeus database

The results of the Mann-Whitey U-test which tests the null hypothesis that two populations are the same. Accuracy was tested for active companies (Act), bankrupt companies (Bankr) and total accuracy (Total). Type I and type II error was also tested.

Table XII The results of the Mann-Whitney test

	Sum of ranks (1)	Sum of ranks (2)	M-W (U)	Z	p-value
Act.	1275	3775	0,00000	-8,61383	0,00000
Bankr.	3775	1275	0,00000	8,61383	0,00000
Total	1275	3775	0,00000	-8,61383	0,00000
type. I.	1275	3775	0,00000	-8,61383	0,00000
type II.	3775	1275	0,00000	8,61383	0,00000

Source: Our own analysis of data from the Amadeus database

According to the results of the Mann-Whitney U-test, a statistically significant difference exists between the accuracy of the model applying the BT method and the model applying the LDA method. The model based on the non-parametric BT method achieves significantly greater accuracy for bankrupt companies in comparison with the LDA model applying a parametric method. The LDA model, however, achieves greater accuracy for active companies. As the total accuracy (Total) is calculated as a weighted average, the LDA model also achieves a higher total accuracy. The accuracy with which the bankruptcy model proves capable of predicting bankruptcy and the type I error shown can, however, be considered the most important thing, and in this respect the BT model clearly dominates the LDA model.

V. Discussion

Increasing the classification accuracy of predictive bankruptcy models can be achieved in two ways - in the choice of suitable variables for the models and in the choice of methods of model selection. The opinion that the choice of method does not offer much room for improvement in bankruptcy modelling predominates in contemporary literature [3, 34].

The consequence of this notion is the frequent usage of the LDA method which, to be effective, requires compliance with some specific conditions, most notably normality of data and the absence of outliers. However, the nature of financial indicators used to build these bankruptcy models is very often quite different: the data deviates from normality and contains outliers [30]. However, the lack of normality and the presence of outliers in financial ratios may in fact be viewed as natural, because they often stem from their very definition [37].

The fact that the classification accuracy of LDA is affected by the natural properties of inputted data, to which BT is immune, means that the choice of method used to create a predictive model can surely influence the classification accuracy of that model.

Unlike the previous approaches (especially [3]), the methods evaluated in our research were used for classification purposes as well as for finding some suitable predictors. This caused the two models to feature different predictors, the selection of which is the consequence of applying a certain method. The potential of the LDA method can be enhanced by an appropriate data transformation, particularly the Box-Cox transformation [10, 35].

A bankruptcy model can then be derived from the transformed indicators that exhibit normal distribution [29, 31].

Our research did not resort to data transformation except for a logarithmic transformation of indicators *TA*, *S*, and *EQ*, for two reasons:

1. Transformation of a monotone function has no effect on the conclusions with the BT method.
2. In the case of LDA, the Box-Cox transformation was not done because the combination of transformation and the method itself would have affected the model accuracy.

In the research presented it is impossible to separate the crucial problem of forming bankruptcy models based on the choice of appropriate methods and the choice of appropriate indicators, as the choice of variables was influenced by the method used. Another option for increasing the classification accuracy of bankruptcy models is the construction of new ratios that form the basic set of potential factors.

VI. Conclusion

The research presented follows on from research in the area of predictive model creation. The authors have attempted to increase the accuracy of the predictive bankruptcy models; firstly, in the manner in which variables are chosen, and secondly, in the manner in which new methods for model creation are tested. The research presented herein examined the effectiveness of creating a bankruptcy model by taking two different approaches, namely the commonly used LDA method and the newer BT nonparametric method.

Both methods were applied to the same sample of companies and the same initial set of indicators. In the process of creating the models there was a reduction in the original amount of data in connection to the method used. The resulting model created by the MDA model contains 8 variables, which are significant for a 1% level of the F-test. In the case of creating a model using the BT method, the resulting model contains only 7 variables. Some of the predictors in the resultant models were identical (*NI/OR, EBIT (5-vol)* and *WC/TA*), others were different. From the viewpoint of the classification accuracy of a model created using the nonparametric BT method, much better results were achieved than with the model using the MDA parametric method.

References

[1] Q. K, Al-Shayea, G. A. El-Refae, “Evaluation of Banks Insolvency Using Artificial Neural Networks”, Proceedings of the 11th WSEAS international conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED'12), Cambridge, United Kingdom, February 22-24. pp. 113-118 2012.

[2] E. I. Altman, “Financial Ratios, “Discriminant Analysis and the Prediction of Corporate Bankruptcy”, The Journal of Finance, vol. 23, no.4, pp. 589-609, 1968.

[3] M. Aziz, H. Dar, ”Predicting corporate bankruptcy: where we stand?” Corporate Governance, vol. 6, pp. 18-33, 2006.

[4] I. Braun, Ch. Mues, “An experimental comparison of classification algorithms for imbalanced credit scoring data sets” Expert Systems with Applications, vol. 39, pp. 3446–3453, 2012.

[5] B. Back, T. Laitinen, K. Sere, “Neural Networks and Genetic Algorithms for Bankruptcy Predictions,” Expert Systems With Applications, vol.11, no.4, pp. 407-413, 1996.

[6] W. H. Beaver, “Financial Ratios as predictors of Failure”, Journal of Accounting Research, vol.4, pp. 71-111, 1966.

[7] W. H. Beaver, M. F. McNichols, and J. W. Rhie, “Have Financial Statements Become Less Informative? Evidence from the Ability of Financial Ratios to Predict Bankruptcy,” Review of Accounting Studies, vol.10, pp. 93-122, 2005.

[8] P. Barnes, “Methodological implications of non-normally distributed financial ratios”, Journal of Business Finance and Accounting, vol.9, no.1, pp. 51-62, 1982.

[9] P. Barnes, “The analysis and use of financial ratios: A review article”, Journal of Business Finance and Accounting, vol.14, no.4, pp. 449-461, 1987.

[10] G. E. P. Box, D. R. Cox, “An Analysis of Transformations”, Journal of the Royal Statistical Society. Series B (Methodological), vol.26, no.2, pp. 211-252, 1964.

[11] L. Breiman, J. H. Friedman, R. Olshen and C. Stone,: ”Classification and Regression Trees”, Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software, 1983.

[12] W. G. Cochran, “On the Performance of the Linear Discriminant Function”, Technometrics, vol.6, no.2, pp. 179-190, May, 1964.

[13] A. Cielen, L. Peeters and K. Vanhoof, “Bankruptcy prediction using a data envelopment analysis”, European Journal of Operational Research, vol. 154, pp. 526–532, 2004.

[14] T. A. Craney, , J. G. Surles, 2002, “Model-Dependent Variance Inflation Factor Cutoff Values”, Quality Engineering, vol 14, no. 3, pp. 391–403. 2002,

[15] T. Hastie, R. Tibshirani, J. Friedman, “The Elements of Statistical Learning: Data Mining, Inference, and Prediction“, 2nd ed. Springer. 2009.

[16] J., De Andres, P. Lorca, F. J. De Cos Juez, and F. Sanchez-Lasheras, “Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS)”. Expert Systems with Applications. vol. 38, pp. 1866–1875, 2011.

[17] E. B. Deakin, “A Discriminant Analysis of Predictors of Business Failure”, Journal of Accounting Research, vol. 10, no. 1, pp. 167-179, 1972.

[18] Y. Ding, X. Song and Y. Zen, “Forecasting financial condition of Chinese listed companies based on support vector machine” Expert Systems with Applications, vol. 34, pp. 3081–3089, 2008.

[19] B. A. Emel, M. Oral, A. Reisman and R. Yolalan, “A credit scoring approach for the commercial banking sector”, Socio-Economic Planning Sciences, vol. 37, pp. 103–123, 2003.

[20] R. A. Fisher, “The use of multiple measurements in taxonomic problems*”,* Annals of Human *Genetics,* vol. 7, no. 2, pp. 179–188, 1936.

[21] Y. Freund, R. Schapire, “A decision-theoretic generalization of online learning and an application to boosting”, Journal of Computer and System Sciences, vol. 55, pp. 119–139, 1997.

[22] J. H. Friedman, “Greedy Function Approximation: A Gradient Boosting Machine”, Annals of Statistics, vol. 29, no 5, pp 1189-1232, 2001.

[23] L. Guelman, “Gradient boosting trees for auto insurance loss cost modeling and prediction”, Expert Systems with Applications, vol. 39, pp. 3659–3667, 2012.

[24] P. Gurný, M. Gurný (2010, September) Logit vs probit model při determinaci souhrnnýchukazatelů výkonnosti bank. [Logit vs. probit model in the determination of aggregate performance indicators of banks]. [online] Available: http://www.ekf.vsb.cz/miranda2/export/sites-root/ekf/konference/cs/okruhy/rmfr/prispevky/dokumenty/Gurny.Petr_2.pdf

[25] P. Hebák, J. Hustopecký, E. Jarošová, and I. Pecáková, *Vícerozměrné statistické metody* [Multivariate Statistical Methods] (1), Informatorium, 2004.

[26] K. L. Henerby, “Do Cash Flows Variables Improve the Prediction Accuracy of a Cox Proportional Hazards Model for Bank Failure?” The Quarterly Review of Economics and Finance, vol. 36, no. 3, pp. 395-409, 1996.

[27] L. H. Chen, H. D. Hsiao, “Feature selection to diagnose a business crisis by using a real GA-based support vector machine: An empirical study,” Expert Systems With Applications, vol. 35, pp. 1145-1155, 2008.

[28] C. Hung, J. H. Chen, “A selective ensemble based on expected probabilities for bankruptcy prediction,” Expert Systems With Applications, vol.36, pp. 5297-5303, 2009.

[29] M. Karas, M., Režňáková, ”Financial Ratios as Bankruptcy Predictors: The Czech Republic Case” In: Advances in Finance & Accounting: *Proceeding of the 1st WSEAS International Conference on Finance, Accounting and Auditing (FAA 12).* WSEAS Press, pp. 86-91, 2012.

[30] M. Karas, M., Režňáková, “The distributional properties of financial ratios: The case of Czech bankruptcy data” Trends of Economics and Management, in press.

[31] M. Karas, M., Režňáková, “Bankruptcy Prediction Model of Industrial Enterprises in the Czech Republic”, International Journal of Mathematical Models and Methods in Applied Sciences, vol. 7, no. 5, p. 519-531, 2013.

[32] M. J. Kim, D. K.. Kang, “Ensemble with neural networks for bankruptcy prediction”, Expert Systems with Applications, 2010: ol. 37, pp. 3373–3379. 2010.

[33] D. Martin, “Early warning of bank failure: A logit regression approach”, Journal of Banking & Finance, vol. 1, no. 3, pp. 249–276. 1977.

[34] M. Niemann, J. H. Schmidt, and M. Neukirchen, “Improving performance of corporate rating prediction models by reducing financial ratio heterogeneity*”, Journal of Banking & Finance,* vol.32, pp. 434-446, 2008.

[35] J. Nikkinen, P. Sahlstrőm, “Distributional properties and transformation of financial ratios: The impact of the accounting environment”, Advances in International Accounting, vol.17, pp. 85-101, 2004.

[36] J. A. Ohlson, “Financial Ratios and the Probabilistic Prediction of Bankruptcy”, Journal of Accounting Research, vol.18, no.1, pp. 109-131, 1980.

[37] A. Omar, S. McLeay, “The sensitivity of prediction models to the non-normality of bounded and unbounded financial ratios”, British Accounting Review, vol. 32, pp. 213–230, 2000.

[38] M. Psillaki, I. T. Tsolas and M. Margaritis, “Evaluation of credit risk based on firm performance, European Journal of Operational Research, vol. 201, pp. 873–881, 2010.

[39] T. Shumway, “Forecasting Bankruptcy More Accurately: A Simple Hazard Model”, Journal of Business, vol.74, no.1, pp. 101-24, 1999.

[40] F. M. Tseng, Y. C. Hu, “Comparing four bankruptcy prediction models: Logit, quadratic interval logit, neural and fuzzy neural networks”, Expert Systems with Applications, vol. 37, pp. 1846–1853, 2010.

[41] B. Twala, “Multiple classifier application to credit risk assessment”, Expert Systems with Applications, vol. 37, pp. 3326–3336, 2010.

[42] Y. J. Wang, H. S. Lee, “A clustering method to identify representative financial ratios,” *Information Sciences,* vol. 178, pp. 1087-1097, 2008.

[43] Y. Wu, C. Gaunt and S. Gray, “A comparison of alternative bankruptcy prediction models”, Journal of Contemporary Accounting & Economics, vol.6, pp. 34-45, 2010.

[44] D. W. Zimmerman, “A note on the influence of outliers on parametric and nonparametric tests”, The journal of general psychology, vol.121, no.4, pp. 391-401, 1994.

[45] D. W. Zimmerman, “Increasing the power of nonparametric tests by detecting and down weighting outliers”, Journal of experimental education, vol.64, no.1, pp.71-79, 1995.

[46] D. W. Zimmerman, “Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions,” *Journal of experimental education*, vol.67, no.1, pp. 55-69, 1998.

[47] M. E. Zmijewski, “Methodological issues related to the estimation of financial distress prediction models*”*, Journal of Accounting Research, vol.22, pp. 59-82, 1984.

[48] Y. Zhou, T. M. S. Elhag, “Apply Logit analysis in Bankruptcy Prediction,” Proceedings of the 7th WSEAS International Conference on Simulation, Modelling and Optimization, Beijing, China, September 15-17. pp. 301-308, 2007.