BANKRUPTCY PREDICTION OF ENGINEERING COMPANIES IN THE EU USING CLASSIFICATION METHODS

Michaela Staňková[1], David Hampel[1]

[1]Department of Statistics and Operation Analysis, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

Abstract

This article focuses on the problem of binary classification of 902 small- and medium-sized engineering companies active in the EU, together with additional 51 companies which went bankrupt in 2014. For classification purposes, the basic statistical method of logistic regression has been selected, together with a representative of machine learning (support vector machines and classification trees method) to construct models for bankruptcy prediction. Different settings have been tested for each method. Furthermore, the models were estimated based on complete data and also using identified artificial factors. To evaluate the quality of prediction we observe not only the total accuracy with the type I and II errors but also the area under ROC curve criterion. The results clearly show that increasing distance to bankruptcy decreases the predictive ability of all models. The classification tree method leads us to rather simple models. The best classification results were achieved through logistic regression based on artificial factors. Moreover, this procedure provides good and stable results regardless of other settings. Artificial factors also seem to be a suitable variable for support vector machines models, but classification trees achieved better results using original data.

Keywords: bankruptcy prediction, binary classification, classification trees, logistic regression, support vector machines

INTRODUCTION

Prognosis of the future development of a company is being researched by a considerable number of scientists, who – over time – have compiled more or less successful prediction models. We are able to predict different future statuses of a company, but the prediction of bankruptcy is the most common. Bankruptcy is a clearly defined situation (as opposed to detection of financial distress or other risks). All prediction models are based on the premise that companies display bankruptcy symptoms for some time prior to actual bankruptcy. Such symptoms include in particular liquidity problems, problems with the amount of net working capital and problems with returns on invested capital, etc.

The first worldwide known models used for bankruptcy prediction (or generally for classification) were based on multiple discriminant analysis (MDA). This technique was used by Beaver (1966), Beerman (1976) and Altman (1968). Further developments were made by Ohlson (1980) with the application of the logistic regression model (Logit) and Zmijewski (1984) with the probit model. Clearly statistical methods are not the only possibility for bankruptcy prediction. Artificial intelligence methods have been used

by Tsukada and Baba (1994), Chen and Du (2009) and Tsakonas *et al.* (2006), who employed neural networks (NN) for the construction of bankruptcy models. Shin *et al.* (2005) tested the accuracy of prediction using the relatively new method of support vector machines (SVM) originally presented by Vapnik (1995) and NN and proved that classification based on SVM has a higher level of accuracy than NN. They admitted that further research is required, especially to examine the influence of various parameter settings (for example the choice of the kernel) on the performance of classification or prediction. SVM has become a popular method in the field of prediction because it does not require such assumptions as normality of data or large data sets.

Min and Lee (2005) have compared the performance of the bankruptcy prediction model using MDA, Logit, NN and SVM with a radial basis function (RBF) kernel. They used a sample of 1,888 companies, half of them in bankruptcy. In the cases of SVM, MDA and Logit, each data set was split into two subsets: a training set of 80% and a holdout set of 20% of companies. The holdout dataset was used to test the results, which are not utilized to develop the model. In the case of NN, they used three subsets: a training set of 60%, a validation set of 20% and a holdout set of 20% companies. The results of their work show that SVM outperforms other techniques in training as well as holdout data. For the holdout data, the prediction accuracy of SVM is better than NN, MDA and Logit by 0.5, 4.8, and 3.9 percentage points respectively. Authors report that another topic for future work should be the choice of the appropriate kernel for SVM, because they believed that this choice will depend individually on the problem and dataset.

Niknya *et al.*, (2013) also compared the predictive ability of models of financial distress using the SVM, MDA and Logit models on the data of 660 Tehran stock exchange companies (over the 2007–2013 period). The SVM method achieved the highest accuracy (93%). The Logit model achieved an accuracy of 85% and the MDA method had the lowest accuracy, of less than 82% Ding, Song and Zen (2008); Min, Lee and Han (2006); and Huarng *et al.* (2005) have also achieved a higher accuracy with SVM in comparison to other approaches.

Sun and Li (2012) tested the accuracy of individual SVM classifiers compared to an ensemble method for financial distress prediction. In an empirical experiment based on 270 listed Chinese companies they applied a SVM classifier with different kernel types such as linear, polynomial (degrees two and three), RBF and sigmoid. As an extraction method they applied stepwise MDA, stepwise Logit and principal component analysis (PCA). Their results demonstrate that a correctly composed SVM ensemble is much better than an individual SVM classifier. In the case of prediction using a single SVM classifier they recommend RBF-SVM with features selected by stepwise MDA.

In business failure prediction, Li *et al.* (2010) proposed classification and regression trees (CART) as an alternative method to the commonly used methods (i.e. SVM, *k*-nearest neighbour etc.). Among the advantages of using CART they include in particular the simplicity of the results and easy implementation. For the data of 153 companies in distress and in health, they showed that there is no significant difference between the two methods of CART and SVM on predictive performance for short-term business failure prediction of listed Chinese companies. The CART model outperformed the SVM model by less than one percentage point. The mean accuracy of CART reached 90.30%, SVM achieved 89.41% accuracy, k-nearest neighbour 88.82%, MDA 88.00% and the Logit model only 86.89%. The appropriateness of using CART models for predicting financial distress has been proved, for example, by the research of Kim and Upneja (2014) and Gepp and Kumar (2015).

Lin and McClean (2001) compared two feature selections for statistical approach and machine learning methods on a set of 1133 companies in the 1980–1999 period. The first one was human judgement (based on financial theory, which has been used by many researchers before), the second one was the ANOVA statistical method. Prediction accuracy was most improved with ANOVA feature selection versus selection using human judgement: Logit from 84.0% to 84.6%, NN from 87.5% to 88.1% and CART from 86.6% to 88.7% Only the MDA model had higher prediction accuracy when we use human judgement (with 78.6% accuracy) than with ANOVA selection (just 77.4%).

The influence of feature selection on the prediction ability of the model was demonstrated in Klepáč and Hampel (2018). They focused on the comparison of different machine learning algorithms such as SVM, CART (including Random forest) and Adaptive boosting. The difference between the best and the worst model (based on total accuracy values) is: 16.8 percentage points for the SVM method, 16 percentage points for the CART method, 29.7 percentage points for the random forest method and 37.5 percentage points for the Adaptive boosting method.

The main aim of this paper is to evaluate the accuracy of models for bankruptcy prediction of engineering companies in the EU. The effort is to compare the accuracy of the different methods together with the appropriate setting within the individual methods. In contrast to other papers, the individual classification methods will be combined with other methods (principal component analysis together with factor analysis), which could increase the predictive ability of the models. This point builds on previous work by Staňková and Hampel (2017), where the set of companies' financial data was examined more closely in order to identify bankruptcy factors. These new artificial factors can be used as variables for bankruptcy models.

Besides the comparison of predictive accuracy, this article also focuses on whether and with what degree of accuracy it is possible to predict the bankruptcy of a company for up to three years in advance.

MATERIALS AND METHODS

With regards to the goal, a total of three methods were selected. The first of them is a well-known logistic regression, as a representative of the classical statistical approach. In addition to this, modern methods of machine learning are applied. From this area, the methods of SVM and CART were selected.

Financial (annual accounting) data on companies from the years 2011 to 2013 are obtained from the Amadeus database. To achieve a more homogeneous data set, we include only small- and medium-sized engineering companies (NACE code 28). In order to obtain the necessary number of bankrupt companies, we focus on the entire area of the European Union, even though there are some differences within the economies of the individual states. The data set includes 902 active companies and 51 companies which went bankrupt in 2014. With regards to financial theory, we choose a set of 19 characteristics representing the four basic groups of financial indicators (i.e. solvency ratios, profitability ratios, liquidity ratios and turnover ratios), see Tab. I. In this table, medians are presented which show the development of financial variables during the reference time period.

Financial ratios are often correlated with each other. For this reason, all models will be made based on all 19 ratios and also using identified artificial factors. For this purpose, the dataset was pre-processed using factor analysis and principal component analysis. Based on this, interpretable factors were constructed that could be used to compile a model for bankruptcy prediction of a company. This procedure has already been successfully verified by cluster analysis in Staňková and Hampel (2017). Within each method (including the various settings described below), six model variants were constructed. All the models were constructed based on all financial indicators, as well as using artificial factors. In the case of artificial factors, a variant with two, three, four, five and six factors was selected. All these steps are repeated for the data coming from one, two and three years before bankruptcy.

Some studies use a different procedure based on feature selection. For example Shin *et al.* (2005) and Sun and Li (2012) used the MDA stepwise method to reduce dimensionality in order to obtain a simplified model. Wang *et al.* (2014) choose information gain for feature selection. In this process variables are selected according to their predictive information. In general, feature selection is used to reduce the dimensionality of the data set, and removes redundant, irrelevant, or noisy data. We choose to use sequential feature selection as developed by Kohavi and John (1997). This procedure selects

I: *Medians of financial indicators in individual years (for active and bankrupt companies)*

Variable	Active companies			Bankrupt companies		
	2011	2012	2013	2011	2012	2013
Current ratio	1.52	1.54	1.57	1.09	1.01	0.75
Quick ratio	0.98	1.01	1.05	0.67	0.59	0.41
Cash ratio	0.42	0.44	0.48	0.18	0.13	0.10
Cash flow liquidity	0.14	0.15	0.15	0.03	−0.03	−0.03
Net working capital (mln EUR)	15.03	15.76	16.19	0.15	0.01	−0.43
Working capital per employee (thous. EUR)	59.00	59.56	59.28	71.71	61.64	36.00
Return on assets (%)	3.95	4.18	3.82	−0.13	−3.64	−38.03
Profit margin (%)	4.66	4.39	4.62	0.82	−3.99	−65.77
EBIT Margin (%)	5.07	4.99	5.17	2.76	−1.73	−57.68
Net assets turnover	2.52	2.52	2.41	3.33	2.68	1.52
Stock turnover	5.61	5.75	5.86	3.16	2.75	2.57
Collection period (days)	66.00	63.89	63.66	111.73	142.57	151.82
Credit period (days)	46.86	42.93	43.03	97.94	114.01	230.94
Debt ratio (%)	66.88	64.90	63.64	93.14	97.45	138.58
Solvency ratio (%)	33.12	35.10	36.36	6.86	2.55	−18.42
Debt to equity ratio	1.96	1.76	1.69	9.14	7.69	−2.07
Interest cover	6.90	6.97	7.82	1.11	−0.78	−9.49
Debt repayment period	6.15	6.10	5.50	19.38	−2.26	−3.57
Fixed asset to equity capital ratio	1.28	1.32	1.30	0.32	0.14	−1.92

a subset of features from the origin dataset that best predict this data by sequentially selecting features until there is no improvement in prediction. The CART and Logit model requires no additional filtration, because this method itself is used to select variables. In these cases, which variables were selected by the model was monitored. Subsequently, the significance of individual variables for the prediction was measured for each estimated model.

The selected financial data of the 953 companies are used for classification in several stages – we use a training (together with validation) and a testing subset. The proportion of these sub-groups (random samples) is 60 % for training (and validation) and 40 % for testing. In line with Gaudioso *et al.* (2017) we use ten-fold cross validation to find the most appropriate setting for each method. We randomly split the training sample into 10 folds. For each k-fold in dataset, model on k–1 folds of the dataset was build and test. The average of k recorded errors is called the cross-validation error and serves as performance metric for the model. It is a widely recommended procedure because it generally results in a less biased or less optimistic estimate of the model skill than simple train/test split procedure.

All the procedures described below are performed in the MATLAB 2017b computing system with Statistics and Machine Learning Toolbox.

Models used for binary classification

For each method we set different setup options. For the SVM model, the kernel function is an important issue for the construction. As there is no completely unambiguous opinion on the ideal kernel type, we decide to construct three models based on SVM, each with a different type of kernel. Using the most frequently mentioned kernels, we choose the linear kernel, the polynomial kernel (second degree) and the radial basis function (Gaussian) kernel.

In addition to choosing a different kernel type, attention was also drawn to the probability calculation. Models (based on the above-mentioned kernel types) were estimated by default but also based on the transformation function which computes the posterior probability that an observation is classified into the positive class. These SVM classifiers contain the optimal score-to-posterior-probability transformation function for two-class learning by 10-fold cross validation as outlined in Platt (2000). The last change in setting the model was to standardize the predictors using their corresponding weighted means and weighted standard deviations. In total, 12 SVM models were defined due to all these different default settings. Technical details about the SVM algorithm can be found in Vapnik (1995).

For each year of the reporting period a CART model is also constructed. The CART method is a widespread and useful tool for a machine learning task intended for classification and prediction. In order to obtain a suitable tree and avoid the overfitting problem, we used the pruning method, which consists of setting a minimum number of leaf node observations. On closer examination, the best approach is to set this rule to 2 % of the original data size. Values of 3 % and higher resulted in a tree built with only one variable. Conversely, values of 1 % and below have led to a complicated and long decision tree.

In the case of the CART method, nine models were selected for each monitored period due to the different setting options. The first setting related to the split criterion. Gini's diversity index, twoing rule and the rule for maximum deviance reduction (also known as cross entropy) were used. Another change in the setting was the algorithm used to select the best split predictor. One option selects the split predictor that maximizes the split-criterion gain over all possible splits of all predictors as in Breiman *et al.* (1984). The curvature test could be used for this purpose too. In this case models use the split predictor that minimizes the p-value of chi-square tests of independence between each predictor and the response as in Loh and Shih (1997). Another option is the interaction test which chooses the split predictor that minimizes the p-value of chi-square tests of independence between each predictor and the response, and that minimizes the p-value of a chi-square test of independence between each pair of predictors and response as in Loh (2002). Technical details about the CART model can be found in Breiman *et al.* (1984). In total, nine CART models were defined as a result of all the aforementioned different default settings.

For each year of the reporting period the Logit model is constructed as a benchmark. It is a nonlinear model which can be used to construct a model for a prediction based on our dataset. The stepwise regression to add or remove predictors was used. In the case of the Logit model, two models were defined based on a different character vector specifying the type of the starting model. The first model (Logit 1) contains only a constant (intercept) term. The second model (Logit 2) contains an intercept and linear terms for each predictor at the beginning. Technical details about the Logit model can be found in Hosmer and Lemeshow (2000).

Measuring prediction quality

Since the proportion of active companies to companies on the brink of bankruptcy is not balanced in the data set, it is advisable to highlight total accuracy as well as overall error rates for both active and bankrupt companies in order to prevent the loss of error margin classification of the less frequent companies that went bankrupt during observed period, similarly to Klepáč and Hampel (2017). Conversely, thanks to this, a data set with a predominance of functional companies can be described as more realistic because it better reflects the situation on today's market which is populated far more densely by active companies than by those

that are on the brink of bankruptcy. A confusion matrix is used to evaluate the level of successful classification (in this case we will focus only on the test set). In line with Klepáč and Hampel (2017) the values of total accuracy with the type I error (incorrect classification of the active companies) and the type II error (misclassification of bankrupted companies) were calculated.

Based on these values we can construct the ROC (Receiver Operating Characteristic) curve as well, as in Zhang *et al.* (2015). ROC curves are useful tools for evaluating binary classifiers based on their performance. The area under the ROC curve (AUC) has been proposed as an alternative single-number measure for evaluating the predictive ability of learning algorithms. It was proved that a precisely defined AUC is a better measure than accuracy when evaluating and comparing classifiers. The resulting AUC value is between 0.5 and 1, where higher values indicate a more successful predictive ability for a model. Another possibility for comparison of methods could be for example Kolmogorov-Smirnov statistics as outlined in Gong and Huang (2012) or Brier score as in Rufibach (2010), but don't investigate them.

In order to objectively evaluate the predictive ability of the individual methods, the model was not estimated only once. Within each model, a thousand repetitions were performed. The results of the prediction ability (i.e. total accuracy with the type I and II errors and the area under ROC curve) are then given as average values. Similarly, the frequency and importance of individual variables in CART and Logit models are averaged.

Research steps

To clarify and summarize our working routine, we state research steps briefly in this section. Similarly to Li *et al.* (2010) we process with the following steps:

1. Dividing dataset onto training group together with validation group and testing (prediction) group. In this case we want to test accuracy for 1000 random samples with 60/40 as for training (with validation)/test partitioning.
2. In the case of SVM methods, feature selection for totally five the most important variables to get simplified model.
3. Pre-processing of all variables using factor analysis and principal component analysis.
4. Training (and validation) and prediction on all 19 financial ratios (in the case of SVM also on 5 selected variables) and also using identified artificial factors.
5. Measuring prediction quality based on overall accuracy, type I and II errors and AUC.

RESULTS

First, we identify variables playing an important role in classification. Tab. II represents an overview of the most commonly used variables employed in the classification models in individual years. These are variables that have been used within the given individual model settings in over 40% of cases. Financial ratios are sorted in descending order according to relative frequency. In the case of the SVM method, these are variables selected using feature selection in alphabetical order. In the case of CART methods, these are the aggregated results of all nine models.

Unfortunately, there was no complete match in the selection of variables in the individual methods in individual years. There are variables such as the debt ratio, which is common to the CART and Logit methods in all three mentioned periods. However, for the SVM method, this variable was not selected at all. Differences in the choice of variables are also apparent in models based on the same

II: *Overview of the most commonly used variables used in the classification models in individual years*

Year	Model	Variables
2013	SVM	EBIT margin, Net working capital, Solvency ratio
	CART	EBIT Margin, Debt ratio, Solvency ratio
	Logit 1	Debt ratio, Collection period, Return on assets
	Logit 2	Cash ratio, Working capital per employee, Collection period, Debt ratio, Return on assets, Quick ratio
2012	SVM	Collection period, EBIT Margin, Net working capital
	CART	Debt ratio, Solvency ratio, Net working capital, Credit period, Collection period
	Logit 1	Debt ratio, Collection period, Credit period, EBIT Margin, Return on assets
	Logit 2	Collection period, Debt ratio, EBIT Margin, Quick ratio, Current ratio, Debt to equity ratio, Working capital per employee
2011	SVM	Collection period, Credit period, EBIT margin, Net assets turnover, Net working capital, Profit margin, Return on assets, Solvency ratio
	CART	Net working capital, Debt ratio, Current ratio, Solvency ratio
	Logit 1	Solvency ratio, Credit period, Collection period, Cash ratio
	Logit 2	Collection period, Debt ratio, Credit period, Cash ratio, Solvency ratio

III: *Overview of the most important variables used in the classification models in individual years*

Year	Model	Variables
2013	SVM	EBIT margin, Net working capital, Solvency ratio
	CART	EBIT Margin, Debt ratio, Solvency ratio, Return on assets, Debt to equity ratio
	Logit 1	Cash ratio, Debt ratio, Current ratio
	Logit 2	Debt ratio, Cash ratio, Quick ratio, Current ratio
2012	SVM	Collection period, EBIT Margin, Net working capital
	CART	Debt ratio, Solvency ratio, Net working capital
	Logit 1	Debt ratio, Solvency ratio
	Logit 2	Debt ratio, Cash ratio, Stock turnover, Working capital per employee
2011	SVM	Collection period, Credit period, EBIT margin, Net assets turnover, Net working capital, Profit margin, Return on assets, Solvency ratio
	CART	Net working capital, Debt ratio
	Logit 1	Debt ratio, Solvency ratio
	Logit 2	Debt ratio, Solvency ratio

method. It is noteworthy that between individual CART (as well as Logit) models via data coming from the three years before bankruptcy, there is a greater match than for periods closer to bankruptcy. Furthermore, individual CART models were most often constructed using three variables. From this point of view, these models can be considered easier to interpret against both Logit and SVM models. It was also found that Logit 2 models were composed of a greater number of variables than Logit 1 models.

In addition to the frequency of use of individual variables, attention was also paid to their importance in calculating the probability used in the classification. Tab. III represents an overview of the most important variables used in the classification models in individual years. These are variables with more than 40% importance arranged in descending order according to their individual importance. In the case of the SVM method, importance is assigned directly during the process of selecting variables. In the case of CART methods, these are the aggregated results.

Even in the case of the most important variables, there is no complete agreement between the models. There are also variables that are common to two of the three methods. This is, for example, the previously mentioned debt ratio variable for the CART and Logit models. But there are also great differences on the importance of some variables. For example, the financial indicator EBIT margin (2013) or net working capital (2011) was the most important for all CART models. For both Logit models, these indicators did not exceed 40% of importance. Besides, in the case of individual CART models the most commonly used and the most important variables are more consistent.

Based on the general results coming from Tabs. II and III, it is obvious that only a few variables have a major influence on the probability calculation for Logit models. Variables such as the collection period or the cash ratio, etc. were very often included in the models, but did not provide vital information for prediction.

In order to evaluate the appropriate settings of the SVM method, attention was also paid to the variable selection procedure. For this purpose, the predictive ability of the models using all 19 ratios and of models based on few financial ratios selected by feature selection was compared. Looking at the overall accuracy and AUC values, it is not easy to tell which procedure is more appropriate. Some models have improved these values (especially within the AUC), but elsewhere they have deteriorated. It was not possible to find a clear formula in these results over the three monitored periods. However, in general for the period of three years before bankruptcy, the values of total accuracy and AUC are better (or similar) when using feature selection regardless of the other model settings. Looking at the error rate of the model, there are stronger differences between models. Values of type II errors increased significantly with just a few variables instead of all financial indicators in the case of models with a linear kernel. From this point of view, the use of this pre-selection method does not appear to be the most appropriate procedure for models with a linear kernel, although it greatly reduced the time demands of model estimation. On the other hand, the results of the type II errors are usually better when we use feature selection for more complex functions of the kernel (especially the RBF kernel).

Within all methods, a comparison of the quality of artificial factor models and of models via all 19 ratios was performed. There were no significant differences in the results of total accuracy and type I errors of the models within each method. The main differences between the different methods and within the different models of the same method were found in the area of type II error and AUC values. In Fig. 1, only these

results of the abilities of all used methods can be seen in all mentioned periods. The circle symbol represents the SVM models, the square symbol represents the CART models and the triangle symbol denotes the Logit models.

Generally good results in this area were achieved by the Logit models through factors. In this case, the model has very good classification ability (based on the values of AUC). Compared to the competing models, it achieves relatively low error rates in the classification of the less frequent (bankrupt) companies. The SVM method had the greatest variability in the results of the individual models. Based on the values of AUC and total accuracy, it can generally be said that this method achieves better results when we use principal component analysis and factor analysis for pre-processing the original dataset. In particular, models with an RBF kernel without the standardization process based on all 19 financial ratios are not applicable. With regard to the type I errors of the SVM models, the linear and RBF kernel models are suitable.

In the case of type II errors, the situation is more complicated. For a period longer than one year before bankruptcy, the lowest values of the type II errors were achieved with the polynomial kernel model with standardized values of predictors based on all 19 ratios–but at the expense of an increased type I error rate. In 2013, the lowest value of type II error was recorded in the RBF kernel model with probability transformation based on five factors–but similar results were achieved with the linear kernel model with probability transformation via five factors. Generally, models with an RBF kernel (regardless of the other settings) are not the ideal choice for this dataset. Promising results (based on values of AUC and total accuracy) are given by models with a linear kernel, especially when using the four factors.

In the case of the CART method, it can be generally said that this method is suitable for use on the original dataset. Promising results are given by the CART model using the maximum deviance reduction as split criterion together with all splits as predictor selection. In 2011, as in 2012, this model reached the highest AUC value (in 2012 the highest total accuracy, too) and the lowest value of type II error via the original dataset. In the period of one year before bankruptcy, the best results (i.e. The highest value of AUC, the highest value of total accuracy and lowest value of type II error) were achieved with the same model's settings, but via two factors. However, the results of the same set model based on all 19 financial ratios were very close to same set model via two factors.

Unlike CART models, better results were generally achieved for Logit models on pre-processed data. Furthermore, it was found that in all monitored seasons a Logit 2 model was more successful. For example, in 2012, using two factors, there is a clear fall in the values of AUC of both Logit models. Even together, two factors in this period do not explain 80% of the original variability. In contrast, in 2013, two factors explain over 85% of the original variability, and both models were very successful here. Also noteworthy is that in the period of one year before bankruptcy, according to values of AUC and type II error, a Logit 2 model based on six factors was the most successful. Nevertheless, the sixth factor was composed of financial ratios that have only a weak correlation to them. However, similar predictive ability results were achieved through three factors in this period.

Choosing the best model was not an easy task. The overall evaluation was performed primarily using AUC values, which include sensitivity and specificity (which reflect type I and II errors). Fig. 2 shows all the observed prediction quality
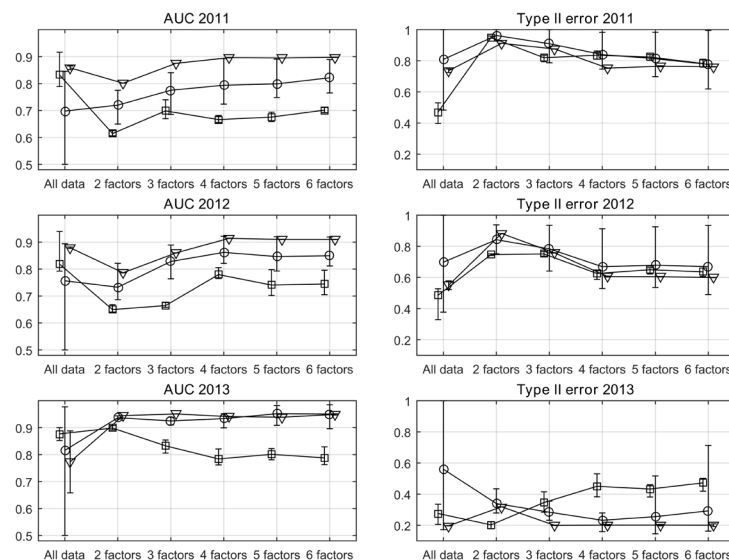
1: *Results of the AUC (left) and the type II errors (right) of all methods in individual years*

characteristics for models selected according to their AUC values. In this Fig. the results can be seen for the best, average and the worst models of the CART (green), SVM (red) and Logit (blue) methods via all financial ratios and four factors in all three periods mentioned. Initially, the CART method had the highest specificity (i.e. lowest type II error) values. With approaching bankruptcy, the Logit models using artificial factors outperform this method. Moreover, the Logit method provides very good and stable results regardless of other settings. SVM models are on average relatively good. However, the choice of the ideal setting is individual for each reference period. Unfortunately, the highest AUC values often come from models that fail in the area of specificity.

DISCUSSION

The results of this empirical research are completely opposite than for example in the studies of Sun a Li (2012), where the SVM model with an RBF kernel is generally recommended. In our case, models with a linear kernel have proven themselves. We agree with Min and Lee (2005) that the kernel choice will depend individually on the problem and data set and pre-processing process. At this point a reference can be made to the articles of Klepáč and Hampel (2016) or Klepáč and Hampel (2018), which demonstrate the effects of different types of kernels in combination with different types of feature selection regarding the error rate of each model.

The combination of principal component analysis and factor analysis, together with the Logit method achieves significantly better results than the other processes selected. However,

other methods have achieved interesting results. The success of the constructed models is apparent when comparing the total accuracy of these bankruptcy models with other (already mentioned) studies. For example, when we compare the average accuracy of our models and models by Li *et al.* (2010), we find differences of several percentage points. Our best prediction mean accuracy is better than their models by 9.40 in the case of SVM, 8.10 in the case of CART, and 11.81 percentage points in the case of the Logit model.

For all the mentioned periods, the best model of all from each of the three selected methods achieved similar results in terms of overall accuracy. In other criteria, the SVM models lagged behind. In this respect, these results are consistent with the work of Li *et al.* (2010), where classification trees are better than the SVM method. But in the case of SVM models, there is still the possibility of examining the results at various settings. There is an opportunity to focus on the types of selection or kernels to determine the effect on the variables in the model and on accuracy.

The results might be even more precise using a different division of the training, validation and testing subsets. We chose the proportion of these subsets as 60 % for training and validation and 40 % for testing but for example Min and Lee (2005) gave a larger number of observations to learning itself (80 %). The accuracy of such models could be compared with other methods (neural networks, *k*-nearest neighbour etc.) just as Ding Song and Zen (2008) and Min and Lee (2005) did in their studies. Another benefit could be the results obtained during the application of these methods within the classification of several statuses of company.
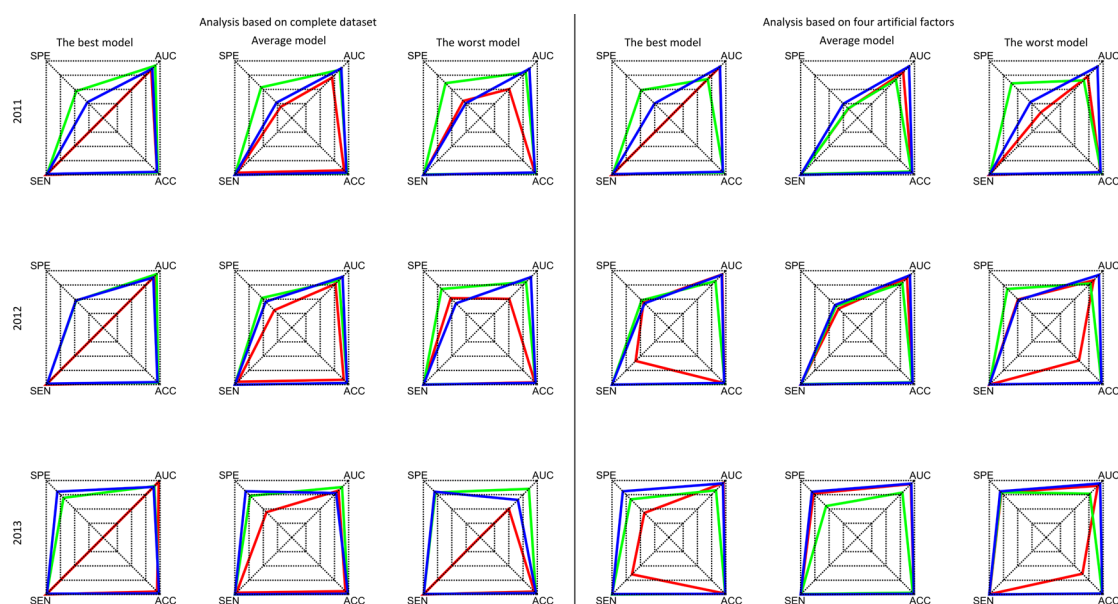
2: *Results of the AUC, total accuracy (ACC), specificity (SPE) and sensitivity (SEN) of the best, average and the worst model of all methods via all financial ratios and four factors in individual years*

CONCLUSION

The results clearly show that with increasing distance from bankruptcy the predictive ability of these models decreases. The resulting AUC values are higher for data coming from one year before bankruptcy. Conversely, type II errors are reduced over time. The success of each method depends on how the original dataset is pre-processed. CART models achieved better results via original data. On the other hand, the SVM and Logit models based on artificial factors have achieved better results. All the methods used have achieved similar results in total accuracy, especially for the period one year ahead of bankruptcy. However, the models constructed differ greatly especially in the values of type II errors. CART models excelled in their simplicity. Nevertheless, the best results were achieved using the Logit models using artificial factors.

REFERENCES

ALTMAN, E. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4): 589–609.

BEAVER, W. 1966. Financial Ratios as Predictors of Failure. *The Journal of Accounting Research*, 4: 71–102.

BEERMAN, K. 1976. *Possible Ways to Predict Capital Losses with Annual Financial Statements*. Düsseldorf, Germany.

BREIMAN, L., FRIEDMAN, J., OLSHEN, R. and STONE, C. 1984. *Classification and Regression Trees*. Boca Raton, FL: CRC Press.

DING, Y., SONG, X. and ZEN, Y. 2008. Forecasting financial condition of Chinese listed companies based on support vector machine. *Expert Systems with Applications*, 34(4): 3081–3089.

GEPP, A. and KUMAR, K. 2015. Predicting Financial Distress: A Comparison of Survival Analysis and Decision Tree Techniques. *Procedia Computer Science*, 54: 396–404.

GAUSIOSO, M., GORGONE, E., LABBÉ, M. and RODRÍGUEZ-CHÍA, A. M. 2017. Lagrangian relaxation for SVM feature selection. *Computer and Operations Research*, 87: 137–145.

HOSMER, D. W. and LEMESHOW, S. 2000. *Applied Logistic Regression*. 2nd Edition. New York: Wiley.

GONG, R. and HUANG, S. H. 2012. A Kolmogorov–Smirnov statistic based segmentation approach to learning from imbalanced datasets: With application in property refinance prediction. *Expert Systems with Applications*, 39(6): 6192–6200.

RUFIBACH, K. 2010. Use of Brier score to assess binary prediction. *Journal of Clinical Epidemiology*, 63(8): 938–939.

HUARNG, K., YU, H. K. and CHEN, C. J. 2005. The application of decision trees to forecast financial distressed companies. In: *International Conference on Intelligent Technologies and Applied Statistics Proceeding 2005*. Taipei, Taiwan.

CHEN, W. S. and DU, Y. K. 2009. *Using neural* networks and data mining techniques for the financial distress prediction model. *Expert Systems with Applications*, 36(2): 4075–4086.

KIM, Y. S. and UPNEJA, A. 2014. Predicting restaurant financial distress using decision tree and Ada Boosted decision tree models. *Economic Modelling*, 36: 354–362.

KLEPÁČ, V. and HAMPEL, D. 2016. Prediction of bankruptcy with SVM classifiers among retail business companies in EU. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 64(2): 627–634.

KLEPÁČ, V. and HAMPEL, D. 2017. Prediction of financial distress of agriculture companies in EU. *Agricultural Economics*, 63(8): 347–355.

KLEPÁČ, V. and HAMPEL, D. 2018. Predicting Bankruptcy of Manufacturing Companies in EU. *E+M: Ekonomie a Management*, 21(1): 159–174.

KOHAVI, P. and JOHN, G. 1997. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1–2): 272–324.

LI, H., SUN, J. and Wu, J. 2010. Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods. *Expert Systems with Applications*, 37(8): 5895–5904.

LIN, Y. F. and McCLEAN, S. 2001. A data mining approach to the prediction of corporate failure. *Knowledge–Based Systems*, 14(3–4): 189–195.

LOH, W. Y. 2002. Regression Trees with Unbiased Variable Selection and Interaction Detection. *Statistica Sinica*, 12: 361–386.

LOH, W. Y. and SHIH, Y. S. 1997. Split Selection Methods for Classification Trees. *Statistica Sinica*. 7: 815–840.

MIN, H. J. and LEE, C. Y. 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28(4): 603–614.

MIN, H. J., LEE, J. and HAN, I. 2006. Hybrid genetic algorithms and support vector machines for bankruptcy prediction. *Expert Systems with Applications*, 31(3): 652–660.

NIKNYA, A., DARABI, R. and FARD, V. H. R. 2013. Financial Distress Prediction of Tehran Stock Exchange Companies Using Support Vector Machines. *European Online Journal of Natural and Social Science*, 2(3): 2878–2893.

OHLSON, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1): 109–131.

PLATT, J. 2000. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, 10: 61–74.

SHIN, K. S., LEE, S. T. and KIM, J. H. 2005. An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, 28(1): 127–135.

SUN, J. and LI, H. 2012. Financial distress prediction using support vector machines: Ensemble vs. individual. *Applied Soft Computing*, 12(8): 2254–2265.

STAŇKOVÁ, M. and HAMPEL, D. 2017. Identification of bankruptcy factors for engineering companies in the EU. *In: Mathematical Methods in Economics: Conference Proceedings*. Hradec Králové: Univerzita Hradec Králové, pp. 714–719.

TSAKONAS, A., DOUNIAS, G., DOUMPOS, M. and ZOPOUNIDIS, C. 2006. Bankruptcy prediction with neurallogic networks by means of grammar–guided genetic programming. *Expert Systems with Applications*, 30(3): 449–461.

TSUKUDA, J. and BABA, I. S. 1994. Predicting Japanese corporate bankruptcy in terms of financial data using neural network. *Computers and Industrial Engineering*, 27(1–4): 445–448.

VAPNIK, V. N. 1995. *The nature of statistical learning theory.* New York: Springer.

WANG, G., MA, J. and YANG, S. 2014. An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications*, 41(5): 2353–2361.

ZHANG, X., LI, X., FENG, Y. and LIU, Z. 2015. The use of ROC and AUC in the validation of objective image fusion evaluation metrics. *Signal Processing*, 115: 38–48.

ZMIJEWSKI, M. E. 1984. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 22: 59–86.

Contact information

Michaela Staňková: xstanko7@node.mendelu.cz