TECHNOLOGICAL AND ECONOMIC DEVELOPMENT OF ECONOMY
ISSN 2029-4913 / eISSN 2029-4921

2014 Volume 20(4): 721–738
doi:10.3846/20294913.2014.979456

FORECASTING CORPORATE FINANCIAL PERFORMANCE USING SENTIMENT IN ANNUAL REPORTS FOR STAKEHOLDERS' DECISION-MAKING

Petr HAJEK[a], Vladimir OLEJ[a], Renata MYSKOVA[b]

[a]*Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Studentská 84, Pardubice, Czech Republic*
[b]*Institute of Business Economics and Management, Faculty of Economics and Administration, University of Pardubice, Studentská 84, Pardubice, Czech Republic*

Received 02 June 2014; accepted 25 July 2014

Abstract. This paper is aimed at examining the role of annual reports' sentiment in forecasting financial performance. The sentiment (tone, opinion) is assessed using several categorization schemes in order to explore various aspects of language used in the annual reports of U.S. companies. Further, we employ machine learning methods and neural networks to predict financial performance expressed in terms of the *Z*-score bankruptcy model. Eleven categories of sentiment (ranging from negative and positive to active and common) are used as the inputs of the prediction models. Support vector machines provide the highest forecasting accuracy. This evidence suggests that there exist non-linear relationships between the sentiment and financial performance. The results indicate that the sentiment information is an important forecasting determinant of financial performance and, thus, can be used to support decision-making process of corporate stakeholders.

Keywords: financial performance, financial distress, bankruptcy forecasting, annual reports, sentiment analysis, opinion mining.

JEL Classification: C45, G33.

Introduction

When deciding on the current and future activities of any company, a great role is played by the amount and quality of information available to the firm's managers and owners. Bazerman and Moore (2009) state that most organizations have the opportunity to increase the effectiveness of their decision-making processes significantly. This is true especially for the financial sphere, as the evaluation of financial source availability and amount influences

Corresponding author Petr Hajek
E-mail: *petr.hajek@upce.cz*

practically all short- and long-term decisions. Financial managers monitor and optimize property and financial structure, costs and income as well as profit distribution. In particular, the monitoring of financial performance utilizes accounting information, which is available to all stakeholders, namely both external and internal users. Quantitative financial information, however, comprises only approximately 20% of all the information contained in annual reports (Beattie *et al.* 2004). Therefore, understanding how the financial and non-financial information in annual reports is connected and how the latter may be used to predict corporate financial performance is of great importance to the stakeholders of companies.

Much research has been aimed at elucidating the mechanisms of corporate financial performance prediction using accounting information. More specifically, historical financial ratios (profitability, liquidity, debt ratios, etc.) have been applied to predict future financial distress. These approaches have evolved from the use of univariate and multivariate statistical models (Altman 1968) to recent use of artificial intelligence (AI) methods such as neural networks (NNs) (Wilson, Sharda 1994; Hajek 2011; Cimpoeru 2011), support vector machines (SVMs) (Huang *et al.* 2004; Hajek, Olej 2011), decision trees (Hajek, Michalak 2013), fuzzy rule-based systems (Chen *et al.* 2011; Hajek 2012), or evolutionary algorithms (Varetto 1998). Comparative reviews of the modes can be found in Bellovary *et al.* (2007), Ravi Kumar and Ravi (2007), or Kirkos (2012). However, these prediction models have not been capable to fully explain the relationships between previous and future financial performance yet. Recent efforts have shown that this fact may be caused by omitting important qualitative information which can be substantially extracted from corporate annual reports and other text documents (Gajzler 2010).

Because every company is different, it is necessary to know as much as possible about its activities and results – and these details are usually stated in verbal rather than in numerical form. We assume that knowledge of these qualitative aspects helps not only evaluate the current economic situation of the company but also improve company management and establish firm goals. Verbal comments are also important for investors, which use them as a basis to evaluate investment success, either realized or planned. Some information is, therefore, published by the company on a voluntary basis to increase its reputation. Therefore, to obtain a complete picture of the financial situation, we assume that it is necessary to use the context-sensitive knowledge provided in annual reports in addition to the financial information. This assumption has been supported by a growing body of literature that indicates that the tone of the documents (sentiment) is significantly correlated with profitability, trading volume, and unexpected earnings (Loughran, McDonald 2011). However, far too little attention has been paid to the use of sentiment analysis in financial performance forecasting (Hajek, Olej 2013; Qiu *et al.* 2014). The objective of our work is to employ sentiment analysis to predict both future financial performance and future change in financial performance expressed as the Z-score. We hypothesize that the sentiment analysis is a decision support tool that improves the accuracy of the forecasting of financial performance and test this hypothesis by using several statistical and AI methods.

The rest of this paper is organized as follows. First, we provide an overview of the information communicated in annual reports and, thus, we explain the motivation behind our research. Next, we briefly review previous literature on using text analysis in forecasting corporate financial performance. Section 3 presents the methods used for: (1) financial performance evaluation, (2) sentiment analysis, and (3) the prediction of financial performance. Section 4 describes the data and, finally, the results of experiments are provided and discussed.

1. Information provided by annual reports

Stakeholders typically utilize accounting information both to evaluate the past and the current financial situation and to serve as a base for future decision-making. External users include potential investors, capital market authorities, state institutions, competitors, business partners and finance analysts. Potential investors, for instance, rely on the information presented in accounting reports when searching for a suitable subject to increase the value of their free funds. Similarly, competitors and business partners utilize this information to compare their companies with the monitored subject and estimate their own strengths and weaknesses. On the contrary, internal users include owners (stock-holders, partners, share holders), company management and, to a certain extent, employees. Accounting information that has a long-term view is important for managers, especially in regards to decision-making about potential investments, as investment activities are considered to be the key element of accelerating the economic development of a company in all its forms (Jac 2012).

Experts understand the accounting system as providing an insight into the economic nature of the company (Biondi *et al.* 2012). The explanatory power of accounting information depends on the quality of its assessment and presentation. Biddle *et al.* (2009), for example, report that firms that have higher financial reporting quality show less sensitivity to macro-economic conditions. In order to correctly assess the financial situation of a company, it is necessary to fulfill certain requirements regarding accounting information. First, accounting information must have sufficient explanatory power and it must not be distorted. Second, the assessment of a firm utilizes data from accounting reports; however, these are accompanied by information about the quality of the company. Therefore, the assessment is the result of a complex approach to the obtained data. Finally, this assessment does not ignore the surroundings of the company. In other words, the obtained data are also evaluated in regards to the micro- and macro-environment of the company, ensuring that feedback is received. Ensuring the understandability of accounting information requires a clear, non-complicated form of presentation.

Accounting information is taken mostly from the basic accounting reports (i.e. from balance sheet, P&L, cash-flow report) as well as from the report on equity capital changes. Supplementary information is included in the notes/comments, which are part of the accounting closure and which explain and add information to the items in previous accounting reports. The need to explain some items of these accounting reports further is connected with the methods used in the accounting unit.

The information obtained from the accounting reports is evaluated through financial analysis based on a variety of methods and indicators (Boguslauskas *et al.* 2011; Stankevičienė, Mencaitė 2012). Their selection is influenced by the reason they are used for. Apart from the oft-used ratio indicators, a number of measures represent analytic systems modified to the needs of the finance managers in companies of a particular branch or field. This financial analysis also takes into account information source and/or level of algorithmization. Another sorting criterion might be whether one evaluates the current situation or predicts future development. This reason might lead a user to choose a set of hierarchically ordered indicators or a dedicated selection of bankruptcy prediction models, for example (Apergis *et al.* 2011).

Companies listed on the stock market and those defined by the legislation of a particular country publish an annual report in addition to their accounting reports. The information contained in the annual report, which is structured into several parts, qualitatively supplements the data derived from the accounting systems and thus it may be highly significant for stakeholders. The annual report aims to provide supplementary information about the company's situation, its development and the circumstances that have influenced these significantly. Further, this report outlines the expected development of the firm. Annual reports are one of the most important external documents that reflect organizational performance and strategy. They describe not only the company, but also its managerial priorities, most important risks and operations. Companies may also discuss how they manage during an economic downturn.

At the economic level, certain criteria that can have a social character exist because company effectiveness depends on the human factor (Myskova 2009). In order to fully evaluate the economic situation of a company, it is important to assess the quantitative information in accounting reports by using financial analysis as well as complement the assessment by understanding the qualitative aspects.

2. Forecasting financial performance using text information – a literature review

Kohut and Segars (1992) provide one of the first studies recognizing the importance of text information in forecasting corporate financial performance. More specifically, their study suggests that annual reports differ in terms of the subjects emphasized when the financial performance worsens. Magnusson *et al.* (2005) employ self-organizing maps to visualize the changes in the writing style of the annual reports of telecommunication companies. Their findings suggest that when a company is expected to perform well, the tone of the report remains positive with extensive use of optimistic vocabulary, active verbs, and clause constructions. On the contrary, a less optimistic and more conservative tone was observed for the companies expecting worsening of their financial performance.

The most recent efforts to understand the relationship between text information and corporate financial performance have focused on sentiment analysis (opinion mining). This approach aims to evaluate people's opinions, appraisals, attitudes and emotions toward entities, individuals, topics and their attributes (Liu, Zhang 2012). Feldman *et al.* (2010), for example, use two word categories, positive and negative, to assess the change in sentiment (tone) in corporate annual reports. Their results suggest that stock market significantly reacts to the change in the sentiment. The same fact has been reported by several other studies, such as those by Li (2010), Huang *et al.* (2010), Schumaker *et al.* (2012), and Hajek *et al.* (2013). However, word categorization (also known as a rule-based approach) requires available dictionary of terms and their categorization according to their sentiment (pre-defined rules). The main issue to be addressed is that such a dictionary is context-sensitive. Thus, a specific dictionary has been developed for financial domain recently (Loughran, McDonald 2011), introducing six sentiment categories. The comparative advantage of the financial dictionary over the traditional Harvard's General Inquirer has been demonstrated on the predictions of returns, trading volume, return volatility, fraud, material weakness and unexpected earnings (Loughran, McDonald 2011). Demers and Vega (2010) use two sentiment categories, net optimism and certainty, to predict future earnings.

Cecchini *et al.* (2010) extract concept scores from annual reports in order to develop a financial ontology discriminating between bankrupt/non-bankrupt companies. Lu *et al.* (2012) employ sentiment analysis on the database of news articles. Their results show that topic-specific negative sentiment is more important for future credit rating changes compared with positive sentiment. Similarly, further studies report that credit rating prediction can be improved with sentiment information extracted from news articles (Lu *et al.* 2013) and annual reports (Hajek, Olej 2013), respectively. Finally, Qiu *et al.* (2014) use machine learning methods employed with the results of text analysis to predict next year's earnings per share.

3. Research methodology

The Altman's model of financial health (Altman's *Z*-score) was selected for the quantitative evaluation of the assessed companies. The reason of the selection was that this model was created for assessing company financial health in industrial branch with shares tradeable on the American publicly regulated market and exactly such companies were selected for the case study.

The scope of this so called bankruptcy model is to predict the probability of survival or bankruptcy of the analyzed company. The nearer to the bankruptcy a company is, the better Altman's index works as a predictor of financial health. It predicts bankruptcies reliably about two years in advance. Altman's *Z*-score model uses the following relation to define the value of a company in industrial branch with shares publicly tradeable on stock market:

$$Z_i = 1.2\, x_{1,i} + 1.4\, x_{2,i} + 3.3\, x_{3,i} + 0.6\, x_{4,i} + 1.0\, x_{5,i}, \tag{1}$$

where i denotes the i-th company, x_1 is working capital/total assets, x_2 is retained earnings/total assets, x_3 is earnings before interest and tax/total assets, x_4 is market value of equity/book value of total liabilities, and x_5 is sales/total assets. The Z_i value is in range –4 to +8. The higher the value, the higher financial health of a company. It holds true that if (1) $Z_i \geq 3$, the company is in the "safe zone" (a company with high probability to survive – financially strong company); (2) $1.80 \leq Z_i \leq 2.99$ "grey zone" (the future of the company cannot be determined clearly – a company with certain financial difficulties); and (3) $Z_i < 1.80$ "distress zone" (the company has serious financial problems – the company is endangered by bankruptcy).

Altman's *Z*-score from 1968 has been modified several times and examined in regards to the changes which were a result of "new" information about stock price (Altman, Brenner 1981). Altman *et al.* (1998) have modified the Altman *Z*-Score model to create the emerging market scoring. Further modifications were created e.g. for private firms, non-manufacturers, etc. (Altman 2000, 2002). Also now the Altman's model belongs to the best know and most used models; it has its critics and supporters. For example, Shumway (2001) criticizes the static type of analysis. Biondi *et al.* (2012) assess key conceptual dilemmas in financial reporting, which are reflected in financial indicators. Kubenka and Kralova (2013), on the other hand, consider Altman's *Z*-score to be the most utilized and spread model, similarly as Pitrova (2011), just with the comment that its reliability for prediction of bankruptcies in the far future decreases significantly.

Figure 1 depicts our research methodology. First, we collected annual reports (10-Ks) in order to extract both the financial performance from financial statements and narrative text for sentiment analysis. Then, the text documents were linguistically pre-processed using (1) tokenization and lemmatization (Feldman *et al.* 1998). The tagged lemmas (potential term candidates) were compared with the following two dictionaries: (1) financial dictionary developed by Loughran and McDonald (2011); and (2) Diction 7.0 word categorizations (Hart 2001). The categorizations include word lists according to the sentiment in the text. In the next stage, the *tf.idf* term weighting scheme was applied to obtain the importance of terms in the corpus of annual reports. An average weight was then calculated for each sentiment category. The weights of the sentiment categories were used as inputs to the statistical methods, machine learning and NNs classifiers. In this section, we present the basic notions of the used methods.

The main issue with the sentiment analysis of textual documents is the right choice of positive (neutral, negative, etc.) terms. Obviously, the categorization of words is not always unambiguous and requires context knowledge. This is due to the various meanings of words and domain specific tone of the words, respectively.

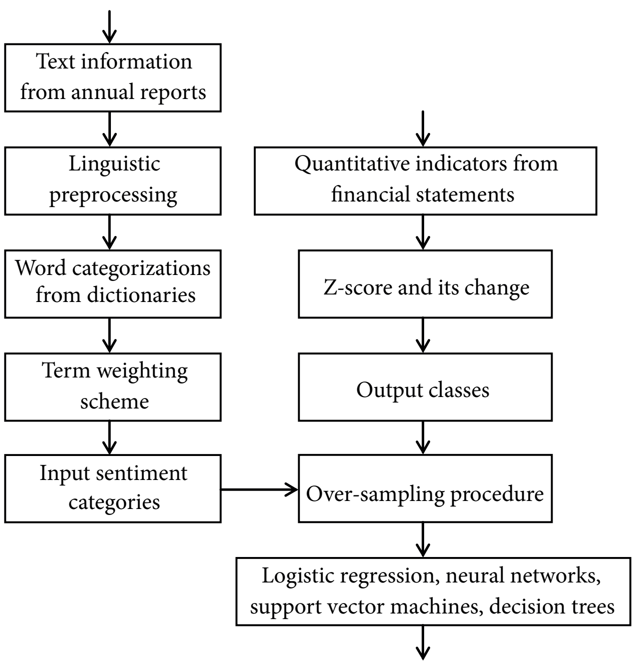

Fig. 1. Research methodology

Therefore, we employed the financial dictionary considering the following categories of terms: (1) Negative (e.g. loss, bankruptcy, problem, suffer, unable, weak); (2) Positive (e.g. achieve, effective, gain, progress, strong, succeed); (3) Uncertainty (e.g. ambiguity, assume, risk, unknown, variable); (4) Litigious (e.g. allege, amend, bail, contract, indict, legal, sue); (5) Modal strong (e.g. always, definitely, strongly, undoubtedly); and (6) Modal weak (e.g. nearly, seldom, sometimes, suggest). The frequency of net positive words was calculated as the positive term count minus the count for negation (positive terms are easily qualified or compromised). Note that ambiguous words (such as increase) are not included in the categories. The other dictionary provided 5 additional categories: (7) Certainty (resoluteness, inflexibility, and completeness); (8) Activity (movement, change, etc.); (9) Optimism (highlighting the positive entailments of persons or events); (10) Realism (tangible, immediate, recognizable matters that affect people's everyday lives); and (11) Commonality (the agreed-upon values of a group). We assumed that well-performing companies use different tone of language in the annual reports, being more positive and optimistic, less uncertain, etc. On the contrary, we expected a more active language in the case of poorly performing companies that need to take actions to improve their financial situation.

We used the most common *tf.idf* (term frequency-inverse document frequency) term weighting scheme, where weights are defined as follows:

$$w_{i,j} = \begin{cases} \dfrac{(1+\log(tf_{i,j}))}{(1+\log(a))}\log\dfrac{N}{df_i} & \text{if } tf_{i,j} \geq 1 \\ 0 & \text{otherwise} \end{cases}, \qquad (2)$$

where N represents the total number of documents in the sample, df_i denotes the number of documents with at least one occurrence of the i-th term, $tf_{i,j}$ is the frequency of the i-th term in the j-th document, and a denotes the average term count in the document.

The outputs of the forecasting models were represented by the classes of financial performance obtained using the *Z*-score bankruptcy model, namely classes "safe zone", "grey zone", and "distress zone". In addition, we also considered the change in financial performance over the monitored period and, thus, we obtained additional output classes (increase, no change, and decrease). Given the fact that the classes were imbalanced in the dataset, we use the Synthetic Minority Oversampling Technique (SMOTE) algorithm (Chawla *et al.* 2002) to modify the training dataset. The algorithm over-samples the minority classes so that all classes are represented equally in the training dataset.

In the next step, we employed several categories of methods to examine the forecasting accuracy based on the sentiment information, namely logistic regression (LR) (Le Cessie, van Houwelingen 1992), artificial immune recognition systems 2-parallel (AIRSs2-p) (Watkins *et al.* 2004), radial basis function neural networks (RBF NNs) (Broomhead, Lowe 1988; Park, Sandberg 1993; Byatt *et al.* 2004), multilayer perceptron neural networks (MLP NNs) (Haykin 1999), SVMs (Cristianini, Shawe-Taylor 2000) with sequential minimal optimization (SMO) (Platt 1998; Uestuen *et al.* 2006), decision tree with naive Bayes classifiers (NBDT) (Kohavi 1996), best-first decision tree classifier (BFDT) (Friedman *et al.* 2000), and decision tree that considers *K* randomly chosen attributes at each node (KRDT) (Breiman 2001).

4. Data description

The set of 11 sentiment attributes presented in the previous section was drawn from annual reports available at U.S. Securities and Exchange Commission EDGAR System. The input attributes were collected for U.S. companies in the year 2008, while the output financial performance (*Z*-score) was evaluated for the year 2010, and the change in the financial performance was measured as *Z*-score in 2010 related to its value in 2008. Following previous studies (Hajek 2012), we excluded the companies from the mining and financial industries to prevent problems with both industry-specific attributes and different financial performance evaluation. As a result, we were able to collect data for 448 U.S. companies, 199 of them classified into "safe zone", 172 classified into the "grey zone", and 44 into the "distress zone" category. Additionally, the change in the financial situation from the year 2008 to 2010 was as follows: 101 companies improved, 304 was without change, and 20 companies worsened.

The financially distressed companies showed the following sentiment characteristic: a more negative, less positive, more uncertain, less litigious and more modal tone (Fig. 2a). The similar tone of language was also used by companies with deteriorating performance (Fig. 3a).

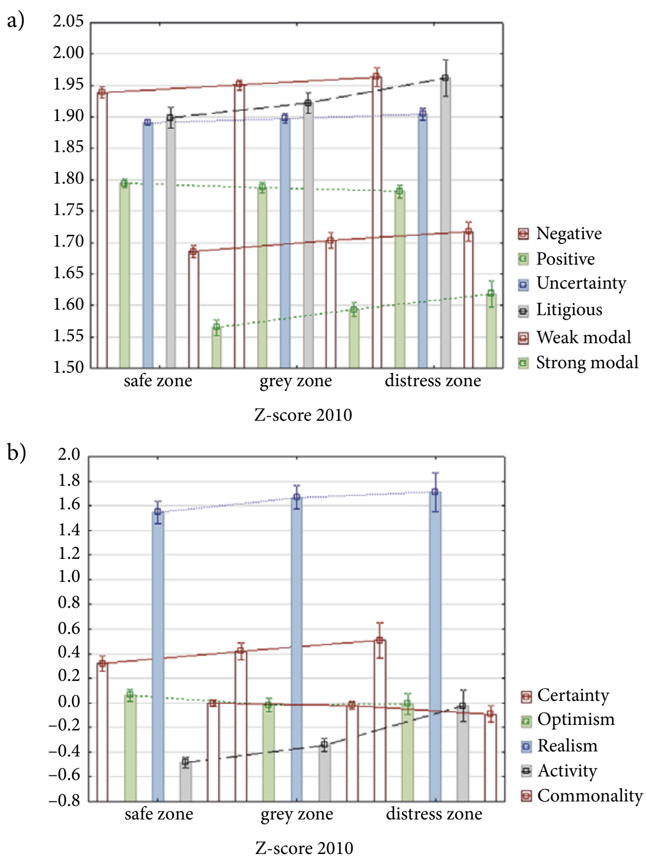

Fig. 2. Sentiment in annual reports and financial performance
Legend: a) sentiment categories (1) to (6), b) sentiment categories (7) to (11)

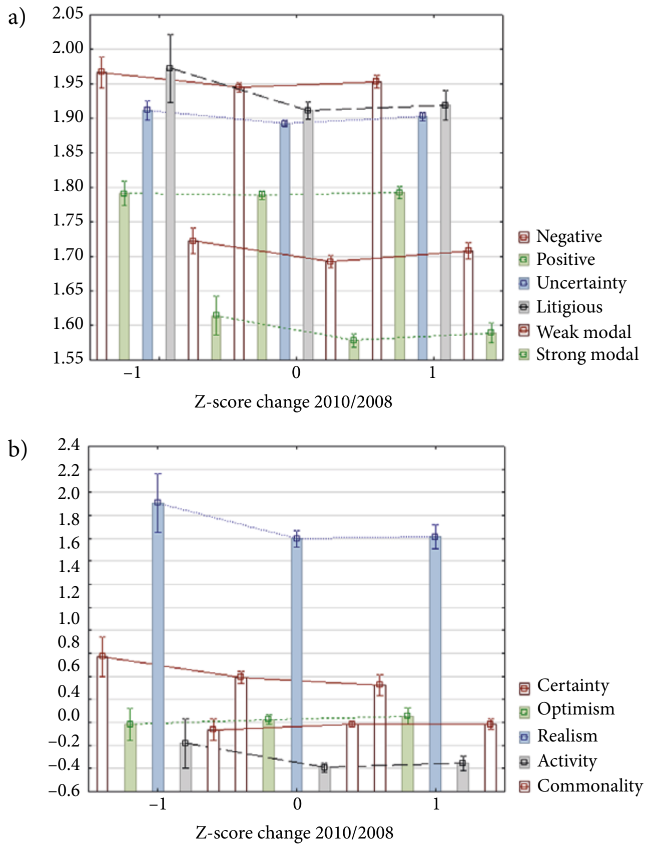

Fig. 3. Sentiment in annual reports and the change in financial performance
Legend: a) sentiment categories (1) to (6), b) sentiment categories (7) to (11)

The remaining categories suggest that both the distressed and worsening companies use less optimistic, more realistic, more active, and less common language in their annual reports (Fig. 2b and Fig. 3b). Interestingly, a more certain and, at the same time, more uncertain tone is reported for the distressed companies. This may be explained by connecting uncertain terms with financial situation while, on the other hand, the certain terms are used together with the concepts (strategies, restructuring, etc.) handling the poor financial situation.

5. Experimental results

We designed various structures of the forecasting methods, i.e. LR, AIRS2-p, RBF and MLP NNs, machine learning (SVM with SMO), and decision trees NBDT, BFDT, and KRDT. Apart from LR, all methods can process non-linear data. We used the LR model with a ridge estimator defined by Le Cessie and van Houwelingen (1992). The classification performance of the LR depends on the number of iterations and ridge factor, respectively.

The AIRS2-p algorithms represents artificial immune systems (AISs) using populations. The algorithm is based on the principle of RB or ARB (Recognition Ball, Artificial Recognition Ball), which can be described as a recognition area or artificial recognition area that combines feature vector (antibody) and vector class. The principle solves the issue of the completeness of AISs. Each antibody is surrounded (in the sense of antibody representation in the state space) by a small area called RB, in which the antibody recognizes all antigens (training dataset). Further, the AIRS2-p algorithms use the principle of limited resource. Each ARB area competes for limited resource according to its stimulation level. The classification performance of the AIRS2-p algorithm proposed by (Watkins *et al.* 2004) depends on several user-defined parameters (Table 1).

RBF NNs (Park, Sandberg 1993; Broomhead, Lowe 1988) was trained with the Broyden-Fletcher-Goldfarb-Shanno method (Byatt *et al.* 2004) and the Gaussian initial centers. The initial centres for the Gaussian RBFs were found using a *k*-means algorithm. The initial sigma values were set to the maximum distance between any center and its nearest neighbor in the set of centers (Table 1).

MLP NNs (Haykin 1999) was trained using the backpropagation algorithm with momentum. The following parameters of the MLP NN were examined to achieve the best classification performance: the number of neurons in the hidden layer, learning rate, momentum, and the number of iteration. Table 1 presents the best settings.

SVMs (Cristianini, Shawe-Taylor 2000; Tian *et al.* 2012) represent an essential kernel-based method with many modifications proposed recently. We used Pearson VII function-based universal kernel function (Uestuen *et al.* 2006) to separate the hyperplane between classes by maximizing the margin between the closest data points. This is done in a higher-dimensional space where the data become linearly separable. We used the SVM trained by the SMO (Platt 1998) algorithm. The classification performance of the SVM was tested for the user-defined parameters as presented in Table 1.

Table 1. Parameters' setting of the used forecasting methods

Methods	Parameters for the best classification performance
LR	The number of iterations = 500, ridge factor = 1.0E-8
AIRS2p	Affinity threshold scalar = 0.1, clonal rate = 128, hypermutation rate = 64, number of k nearest neighbors = 2, initial memory cell pool size = 100, number of instances to compute the affinity threshold = 5, stimulation threshold = 0.9, total resources = 500
RBF NN	The number of iteration = 100, minimum standard deviation for the clusters = 0.1, the number of clusters for *k*-means = 128, and ridge factor = 1.0E-8
MLP NN	The number of neurons in the hidden layer = not limited, learning rate = 0.3, momentum = 0.5, the number of iteration = 1000, validation threshold = 20
SVM	Complexity parameter C = 24, round-off error ε = 1.0E-4, tolerance parameter = 0.1
NBDT	Debug = true
BFDT	Heuristic search, minimal number of instances at the terminal nodes = 2, number of folds in internal cross-validation = 5, pruning strategy = post pruning, the percentage of the training set size = 32
KRDT	Sets the number of K randomly chosen attributes = 60, the maximum depth of the tree = 100, the minimum total weight of the instances in a leaf = 1, number folds = 0

The group of NBDT, BFDT, and KRDT algorithms uses a tree representation assigning a class to an object based on its attributes, which can be continuous or discrete. An attribute with the best value of the splitting criterion is assigned to each root and intermediate node. The classification performance of the DTs depends on the setting of parameters (Table 1).

The quality of forecasting measured by the accuracy of correctly classified instances Acc [%] depends on the settings of parameters (Table 1) of the given methods. Therefore, we tested many settings, and provide only the best results. In order to avoid overfitting, the experiments were realized using 10-fold cross-validation. The measures of classification performance were, in addition to Acc, represented by the averages of standard statistics applied in classification tasks (Powers 2011): true positives rate (TP), false positives rate (FP), precision (Pre) and recall (Re), F-measure (F-m), and the area under the receiver operating characteristic (ROC) curve. F-m is the weighted harmonic mean of Pre and Re. ROC is a graphical plot which illustrates the performance of a binary classifier system, and represents a standard technique for summarization classifier performance over a range of tradeoffs between TP and FP error rates. The best results in terms of the accuracy of correctly classified instances Acc [%] are presented in Table 2 (for the change in financial performance classes) and Table 3 (for the financial performance classes).

Table 2. Average accuracy of the analyzed methods and weighted average TP, FP, Pre, Re, F-m and ROC for classification of change in financial performance (growth, no change, decrease)

Methods	Acc [%]	TP	FP	Pre	Re	F-m	ROC
LR	59.67	0.597	0.204	0.589	0.597	0.591	0.762
AIRS2p	81.03	0.810	0.095	0.808	0.810	0.806	0.858
RBF NN	82.87	0.829	0.087	0.826	0.829	0.826	0.905
MLP NN	83.79	0.838	0.081	0.837	0.838	0.835	0.913
SVM	**86.92**	0.869	0.068	0.869	0.869	0.869	0.916
NBDT	84.16	0.842	0.080	0.843	0.842	0.839	0.939
BFDT	**84.89**	0.849	0.077	0.846	0.849	0.847	0.910
KRDT	**86.01**	0.860	0.072	0.858	0.860	0.859	0.914

Note: significantly higher accuracy at $P < 0.05$ (paired t-test) is marked in bold.

Table 3. Average accuracy of the analyzed methods and weighted average TP, FP, Pre, Re, F-m and ROC for classification of financial performance (safe zone, grey zone, and distress zone)

Methods	Acc [%]	TP	FP	Pre	Re	F-m	ROC
LR	54.92	0.549	0.228	0.544	0.549	0.546	0.749
AIRS2p	75.19	0.752	0.126	0.751	0.752	0.747	0.813
RBF NN	80.49	0.805	0.101	0.805	0.805	0.804	0.837
MLP NN	79.17	0.792	0.106	0.789	0.792	0.789	0.862
SVM	**83.79**	0.838	0.081	0.837	0.838	0.834	0.891
NBDT	80.30	0.803	0.103	0.808	0.803	0.804	0.899
BFDT	78.98	0.790	0.107	0.789	0.790	0.789	0.882
KRDT	79.54	0.795	0.105	0.795	0.795	0.795	0.860

Note: significantly higher accuracy at $P < 0.05$ (paired t-test) is marked in bold.

The results obtained by modeling for the change in financial performance classes (increase, no change, and decrease) and for the financial performance classes (safe zone, grey zone, and distress zone) point to the fact that the accuracy of correctly classified instances Acc [%] is the lowest for LR. This fact confirms that linear methods, however optimized, are not suitable for modeling such heterogeneous, inconsistent, and uncertain data.

Next, from Table 2 and Table 3 it is visible that the accuracy of correctly classified instances Acc [%] AIRS2-p is much higher than for LR, nevertheless, it does not reach the values of RBF NN, MLP NN and SVM. This is given by the fact that the group of AIS algorithms requires a larger set of data for modeling.

The machine learning method – SVM reaches in both cases the highest accuracy of correctly classified instances Acc [%] which confirms the fact that SVM is a suitable choice for modeling the presented data. The results listed in Table 2 and Table 3 point also out that SVM processes well strongly non-linear data with sentiment.

The group NBDT, BFDT, and KRDT tree shows much better results for the change in financial performance classes (increase, no change, and decrease) than AIS and NNs, for the financial performance classes (safe zone, grey zone, and distress zone) the listed group of trees matches NNs. The statistical characteristics in each used method further characterize the classification into three classes.

Due to the fact that weighted averages of the ROC curve in Table 2 and Table 3 for NBDT tree reach the highest values, it is convenient to show detailed characteristics in Table 4 and Table 5 which characterize the classification into three classes. Further, these values are compared to machine learning – SVM, which shows the best accuracy of correctly classified instances Acc [%].

The standard statistics of the classification tasks (Tables 4, 5) for method SVM (NBDT) and change in financial performance (classification performance) lead to the order of correctly classified instances in the given class, i.e. increase, decrease and no change (safe zone, distress zone, and grey zone). For the classification by SVM (NBDT) methods it is possible to state that the classes "increase", and "safe zone" significantly influence the average values of the statistics. Further, the classes "no change" and "grey zone" exhibit the lowest number of correctly classified instances.

Table 4. TP, FP, Pre, Re, F-m and ROC for classification of the change in financial performance classes (increase, no change, and decrease)

Methods	Class	TP	FP	Pre	Re	F-m	ROC
SVM	increase	1	0	1	1	1	1
	no change	0.820	0.103	0.815	0.820	0.817	0.866
	decrease	0.797	0.096	0.801	0.797	0.799	0.891
NBDT	increase	0.994	0.030	0.940	0.994	0.966	0.997
	no change	0.711	0.077	0.836	0.711	0.769	0.909
	decrease	0.836	0.131	0.755	0.836	0.794	0.916

Table 5. TP, FP, Pre, Re, F-m and ROC for classification of financial performance (safe zone, grey zone, and distress zone)

Methods	Class	TP	FP	Pre	Re	F-m	ROC
SVM	safe zone	1	0.065	0.878	1	0.935	0.981
	grey zone	0.711	0.072	0.847	0.711	0.773	0.825
	distress zone	0.819	0.107	0.778	0.819	0.803	0.877
NBDT	safe zone	0.892	0.025	0.943	0.892	0.916	0.946
	grey zone	0.804	0.171	0.724	0.804	0.762	0.868
	distress zone	0.717	0.104	0.770	0.717	0.743	0.887

We used Error/Baseline (E/B) to evaluate the prediction contribution of sentiment categories. This procedure measures the sensitivity of model performance when the corresponding input attribute is left out. The results of the sensitivity analysis for the SVM model are presented in Table 6. All sentiment categories were significant (E/B > 1). In other words, the use of the attributes leads to the reduction of Root Mean Squared Error.

Table 6. Sensitivity analysis of the SVM model

Sentiment category	*Z*-score change		*Z*-score	
	E/B	Rank	E/B	Rank
Negative	1.114	4	1.106	6
Positive	1.080	8	1.107	5
Uncertainty	1.813	1	1.058	10
Litigious	1.088	6	1.141	2
Modal strong	1.130	3	1.176	1
Modal weak	1.111	5	1.135	4
Certainty	1.060	11	1.136	3
Activity	1.323	2	1.069	8
Optimism	1.063	10	1.061	9
Realism	1.082	7	1.073	7
Commonality	1.073	9	1.057	11

When sentiment was included to the evaluation, it was found that in the annual reports which include both, positive and negative changes in the development, the management is much more focused not only on the explanation of this situation but especially on the activities which support further development of the company and its position on the market. When a company is in a difficult economic situation, the annual report contains a large quantity of legal formulation. Characteristic features were also recognized for the companies in the grey zone – the annual report uses indeterminate words. When assessing the differences between the static values result (*Z*-score values) and the dynamic values (*Z*-score changes), it arose that the methods show better results as the performance changes are better commented. The reason is the effort to present the company in such a way that the trust of stakeholders is not jeopardized.

Conclusions

Information about financial position and company performance is designated for all stakeholders, who use it to make rational economic decisions. As this article used data on American companies, we emphasize financial reporting by US GAAP. In the US, there is a sophisticated set of requirements on the content, form and publishing mode of accounting reports, as it is crucial to provide information that is important for shareholders, potential investors and creditors.

From the viewpoint of sentiment, the additional information included, especially in the addendum to the accounting closure (footnotes) and annual report, is most important. The appendix in the annual report should be prepared in great detail including, for example, a complete list of long-term obligations. The need to explain some of the items in accounting reports further is related to the methods used for accounting for long-term orders, sales and so on. The annual report is considered to be not only a compulsory document but also proof of informational openness, responsiveness and transparency in relation to other subjects.

If we used any indicators of financial analysis to evaluate the economic situation of a company, this quantitative assessment would not suffice. Through the application of Altman's model, the resulting Z-score would place the company into one of the defined zones; however, based on this result, we are not able to recognize all the circumstances that led to this result. The company might also indicate a condition that makes it impossible to predict future development (i.e. the output is a numerical value that places the company into the so-called grey zone). The need to diagnose the company in the best way requires finding the reasons for the current status and predicting future development. Hence, it is not significant whether we use Altman's or another model, the important point is the influence of sentiment on the perception of the company by its stakeholders.

The presented research has its limitations, which lie in the possible utilization of creative accounting; however, it may be assumed that company management has learned from the results of the financial crisis caused, among other drivers, by dishonest practices. In further research, it would be possible to focus on how macroeconomic indicators (e.g. inflation) influence the financial condition of companies and consider their connection to sentiment.

Acknowledgments

We gratefully acknowledge the help provided by constructive comments of the anonymous referees. This work was supported by the scientific research project of the Czech Sciences Foundation Grant No: 13-10331S.

References

Altman, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance* 23(4): 589–609. http://dx.doi.org/10.1111/j.1540-6261.1968.tb00843.x

Altman, E. I. 2000. *Predicting financial distress of companies: revisiting the Z-Score and Zeta Models* [online], [cited 16 February 2014]. New York University, New York. Available from Internet: http://pages.stern.nyu.edu/~ealtman/Zscores.pdf

Altman, E. I. 2002. *Corporate distress prediction models in a turbulent economic and Basel II environment* [online], [cited 15 February 2014]. New York University, New York. Available from Internet: http://pages.stern.nyu.edu/~ealtman/Corp-Distress.pdf

Altman, E. I.; Hartzell, J.; Peck, M. 1998. Emerging markets corporate bonds: a scoring system, in R. M. Levich (Ed.). *Proc. of a Conference held at the Stern School of Business*, 23–24 May 1996, New York, 391–400.

Altman, T. I.; Brenner, M. 1981. Information effects and stock market response to signs of firm deterioration, *Journal of Financial and Quantitative Analysis* 16(1): 35–51. http://dx.doi.org/10.2307/2330665

Apergis, N.; Sorros, J.; Artikis, P.; Zisis, V. 2011. Bankruptcy probability and stock prices: the effect of Altman Z-score information on stock prices through panel data, *Journal of Modern Accounting and Auditing* 7(7): 689–696. http://dx.doi.org/10.2304/csee.2011.10.1.3

Bazerman, M. H.; Moore, D. A. 2009. *Judgment in managerial decision making.* 7th ed. Westford: John Wiley & Sons. 230 p.

Beattie, V.; McInnes, W.; Fearnley, S. 2004. A methodology for analysing and evaluating narratives in annual reports: a comprehensive descriptive profile and metrics for disclosure quality attributes, *Accounting Forum* 28(3): 205–236. http://dx.doi.org/10.1016/j.accfor.2004.07.001

Bellovary, J.; Giacomino, D.; Akers, M. 2007. A review of bankruptcy prediction studies: 1930 to present, *Journal of Financial Education* 33: 1–42.

Biddle, G. C.; Hilary, G.; Verdi, R. S. 2009. How does financial reporting quality relate to investment efficiency?, *Journal of Accounting & Economics* 48(2–3): 112–131. http://dx.doi.org/10.1016/j.jacceco.2009.09.001

Biondi, Y.; Glover, J.; Jamal, K.; Ohlson, K. J.; Stephen H.; Penman, S. H.; Sunder, S.; Tsujiyama, E. 2012. Some conceptual tensions in financial reporting, *Accounting Horizons* 26(1): 125–133. http://dx.doi.org/10.2308/acch-50087

Boguslauskas, V.; Mileris, R.; Adlytė, R. 2011. New internal rating approach for credit risk assessment, *Technological and Economic Development of Economy* 17(2): 369–381. http://dx.doi.org/10.3846/20294913.2011.583721

Breiman, L. 2001. Random forests, *Machine Learning* 45: 5–32. http://dx.doi.org/10.1023/A:1010933404324

Broomhead, D. S.; Lowe, D. 1988. Multivariate functional interpolation and adaptive networks, *Complex Systems* 2: 321–355.

Byatt, D.; Coope, I. D.; Price, C. J. 2004. Effect of limited precision on the BFGS quasi-Newton algo-rithm, *ANZIAM Journal* 45 E: C283–C295 [online], [cited 16 February 2014]. Available from Internet: http://anziamj.austms.org.au/V45/CTAC2003/Byat

Cecchini, M.; Aytug, H.; Koehler, G. J.; Pathak, P. 2010. Making words work: using financial text as a predictor of financial events, *Decision Support Systems* 50(1): 164–175. http://dx.doi.org/10.1016/j.dss.2010.07.012

Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; Kegelmeyer, W. P. 2002. SMOTE: Synthetic minority over-sampling technique, *Journal of Artificial Intelligence Research* 16: 321–357. http://dx.doi.org/10.1613/jair.953

Chen, H. L.; Yang, B.; Wang, G.; Liu, J.; Xu, X.; Wang, S. J.; Liu, D. Y. 2011. A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbour method, *Knowledge-Based Systems* 24(8): 1348–1359. http://dx.doi.org/10.1016/j.knosys.2011.06.008

Cimpoeru, S. S. 2011. Neural networks and their application in credit risk assessment. Evidence from the Romanian market, *Technological and Economic Development of Economy* 17(3): 519–534. http://dx.doi.org/10.3846/20294913.2011.606339

Cristianini, N.; Shawe-Taylor, J. 2000. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge: Cambridge University Press. 198 p. http://dx.doi.org/10.1017/CBO9780511801389

Demers, E. A.; Vega, C. 2010. *Soft information in earnings announcements: news or noise?*, Working paper, INSEAD [online], [cited 15 February 2014]. Available from Internet: http://www.federalreserve.gov/pubs/ifdp/2008/951/ifdp951.pdf

Feldman, R.; Fresko, M.; Kinar, Y.; Lindell, Y.; Liphstat, O.; Rajman, M.; Schler, Y.; Zamir, O. 1998. Text mining at the term level, in J. M. Zytkow, M. Quafafou (Eds.). *Principles of data mining and knowledge discovery*. Springer-Verlag, 65–73.

Feldman, R.; Govindaraj, S.; Livnat, J.; Segal, B. 2010. Management's tone change, post earnings announcement drift and accruals, *Review of Accounting Studies* 15: 915–953. http://dx.doi.org/10.1007/s11142-009-9111-x

Friedman, J.; Hastie, T.; Tibshirani, R. 2000. Additive logistic regression: a statistical view of boosting, *Annals of Statistics* 28(2): 337–407. http://dx.doi.org/10.1214/aos/1016218223

Gajzler, M. 2010. Text and data mining techniques in aspect of knowledge acquisition for decision support system in construction industry, *Technological and Economic Development of Economy* 16(2): 219–232. http://dx.doi.org/10.3846/tede.2010.14

Hajek, P. 2011. Municipal credit rating modelling by neural networks, *Decision Support Systems* 51(1): 108–118. http://dx.doi.org/10.1016/j.dss.2010.11.033

Hajek, P. 2012. Credit rating analysis using adaptive fuzzy rule-based systems: An industry-specific approach, *Central European Journal of Operations Research* 20(3): 421–434. http://dx.doi.org/10.1007/s10100-011-0229-0

Hajek, P.; Michalak, K. 2013. Feature selection in corporate credit rating prediction, *Knowledge-Based Systems* 51: 72–84. http://dx.doi.org/10.1016/j.knosys.2013.07.008

Hajek, P.; Olej, V. 2011. Credit rating modelling by kernel-based approaches with supervised and semi-supervised learning, *Neural Computing and Applications* 20(6): 761–773. http://dx.doi.org/10.1007/s00521-010-0495-0

Hajek, P.; Olej, V.; Myskova, R. 2013. Forecasting stock prices using sentiment information in annual reports – a neural network and support vector regression approach, *WSEAS Transactions on Systems* 10(4): 293–305 [online], [cited 15 February 2014]. Available from Internet: http://www.wseas.org/multimedia/journals/economics/2013/235702-202.pdf

Hajek, P.; Olej, V. 2013. Evaluating sentiment in annual reports for financial distress prediction using neural networks and support vector machines, *Communications in Computer and Information Science* 384: 1–10.

Hart, R. P. 2001. Redeveloping DICTION: theoretical considerations, in M. D. West (Ed.). *Theory, method, and practice in computer content analysis.* Greenwood Publishing Group, 43–60.

Haykin, S. 1999. *Neural networks: a comprehensive foundation*. 2nd ed. New Jersey: Prentice-Hall Inc. 842 p.

Huang, Z.; Chen, H.; Hsu, C. J.; Chen, W. H.; Wu, S. 2004. Credit rating analysis with support vector machines and neural networks: a market comparative study, *Decision Support Systems* 37(4): 543–558. http://dx.doi.org/10.1016/S0167-9236(03)00086-1

Huang, A.; Zang, A.; Zheng, R. 2010. *Informativeness of Text in Analyst Reports: A Naïve Bayesian Machine Learning Approach*, Working paper [online], [cited 15 February 2014]. The Hong Kong University of Science and Technology. Available from Internet: http://ifas.xmu.edu.cn/pic/201011/1/201011110517633.pdf

Jac, I. 2012. Risk estimation successful of investment by alternative script of development economy, *E&M Ekonomie a Management* 15(2): 38–53. (in Czech).

Kirkos, E. 2012. Assessing methodologies for intelligent bankruptcy prediction, *Artificial Intelligence Review* 11: 1–41.

Kohavi, R. 1996. Scaling up the accuracy of Naive-Bayes classifiers: a decision-tree hybrid, in *Proc. of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 2–4 August 1996, Portland, Oregon, 202–207.

Kohut, G. F.; Segars, A. H. 1992. The president's letter to stockholders: an examination of corporate communication strategy, *Journal of Business Communication* 29(1): 7–21. http://dx.doi.org/10.1177/002194369202900101

Kubenka, M.; Kralova, V. 2013. "Z" score in assessing the financial health in the construction sector, *E&M Ekonomie a Management* 16(1): 101–112. (in Czech).

Le Cessie, S.; van Houwelingen, J. C. 1992. Ridge estimators in logistic regression, *Applied Statistics* 41(1): 191–201. http://dx.doi.org/10.2307/2347628

Li, F. 2010. The information content of forward-looking statements in corporate filings – A Naïve Bayesian machine learning approach, *Journal of Accounting Research* 48(5): 1049–1102. http://dx.doi.org/10.1111/j.1475-679X.2010.00382.x

Liu, B.; Zhang, L. 2012. A survey of opinion mining and sentiment analysis, in *Mining Text Data*. New York: Springer, 415–463. http://dx.doi.org/10.1007/978-1-4614-3223-4_13

Loughran, T.; McDonald, B. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *The Journal of Finance* 66(1): 35–65. http://dx.doi.org/10.2469/dig.v41.n2.20

Lu, Y. C.; Shen, C. H.; Wei, Y. C. 2013. Revisiting early warning signals of corporate credit default using linguistic analysis, *Pacifin-Basin Finance Journal* 24: 1–21. http://dx.doi.org/10.1016/j.pacfin.2013.02.002

Lu, H. M.; Tsai, F. T.; Chen, H.; Hung, M. W.; Li, S. H. 2012. Credit rating change modeling using news and financial ratios, *ACM Transactions on Management Information Systems* 3(3): 14. http://dx.doi.org/10.1145/2361256.2361259

Magnusson, C.; Arppe, A.; Eklund, T.; Back, B.; Vanharanta, H.; Visa, A. 2005. The language of quarterly reports as an indicator of change in the company's financial status, *Information & Management* 42(4): 561–574. http://dx.doi.org/10.1016/j.im.2004.02.008

Myskova, R. 2009. Company information systems used in decision-making processes of financial managers, *WSEAS Transactions on Systems* 8(4): 532–542 [online], [cited 15 February 2014]. Available from Internet: http://www.wseas.us/e-library/transactions/systems/2009/29-269.pdf

Park, J.; Sandberg, I. W. 1993. Approximation and radial basis function network, *Neural Computing* 5: 305–316. http://dx.doi.org/10.1162/neco.1993.5.2.305

Pitrova, K. 2011. Possibilities of the Altman ZETA Model application to Czech firms, *E&M Ekonomie a Management* 14(3): 66–76.

Platt, J. C. 1998. Fast training of support vector machines using sequential minimal optimization, in B. Schoelkopf, C. Burges, A. Smola (Eds.). *Advances in Kernel methods – support vector learning.* MIT Press, 185–208.

Powers, D. M. W. 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, *Journal of Machine Learning Technologies* 1(2): 37–63.

Qiu, X. Y.; Srinivasan, P.; Hu, Y. 2014. Supervised learning models to predict firm performance with annual reports: an empirical study, *Journal of the Association for Information Science and Technology* 65(2): 400–413. http://dx.doi.org/10.1002/asi.22983

Ravi Kumar, P.; Ravi, V. 2007. Bankruptcy prediction in banks and firms via statistical and intelligent techniques – a review, *European Journal of Operational Research* 180(1): 1–28. http://dx.doi.org/10.1016/j.ejor.2006.08.043

Schumaker, R. P.; Zhang, Y.; Huang, C. N.; Chen, H. 2012. Evaluating sentiment in financial news articles, *Decision Support Systems* 53(3): 458–464. http://dx.doi.org/10.1016/j.dss.2012.03.001

Shumway, T. 2001. Forecasting bankruptcy more accurately: a simple hazard model, *The Journal of Business* 74(1): 101–124. http://dx.doi.org/10.2139/ssrn.171436

Stankevičienė, J.; Mencaitė, E. 2012. The evaluation of bank performance using a multicriteria decision making model: a case study on Lithuanian commercial banks, *Technological and Economic Development of Economy* 18(1): 189–205. http://dx.doi.org/10.3846/20294913.2012.668373

Tian, Y.; Shi, Y.; Liu, X. 2012. Recent advances on support vector machines research, *Technological and Economic Development of Economy* 18(1): 5–33. http://dx.doi.org/10.3846/20294913.2012.661205

Uestuen, B.; Melssen, W. J.; Buydens, L. M. C. 2006. Facilitating the application of support vector regression by using a universal Pearson VII function based kernel, *Chemometrics and Intelligent Laboratory Systems* 81: 29–40. http://dx.doi.org/10.1016/j.chemolab.2005.09.003

Varetto, F. 1998. Genetic algorithms applications in the analysis of insolvency risk, *Journal of Banking and Finance* 22(10–11): 1421–1439. http://dx.doi.org/10.1016/S0378-4266(98)00059-4

Watkins, B. A.; Timmis, J.; Boggess, L. 2004. Artificial immune recognition system (AIRS): an immune inspired supervised learning algorithm, *Genetic Programming and Evolvable Machines* 5: 291–317. http://dx.doi.org/10.1023/B:GENP.0000030197.83685.94

Wilson, R. L.; Sharda, R. 1994. Bankruptcy prediction using neural networks, *Decision Support Systems* 11(5): 545–557. http://dx.doi.org/10.1016/0167-9236(94)90024-8

Petr HAJEK (assoc. prof. MSc. Ph.D.) is an associate professor with the Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Czech Republic. He has been working with the modelling of economic and financial processes using soft computing methods. He has published his research in leading journals such as Knowledge-Based Systems, Decision Support Systems, etc.

Vladimir OLEJ (prof. MSc. Ph.D.) is a professor with the Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Czech Republic. He has been working with the modelling of economic and environmental processes on the basis of soft computing methods. He has published a number of papers concerning fuzzy logic, neural networks and genetic algorithms.

Renata MYSKOVA (assoc. prof. MSc. Ph.D.) is an associate professor with the Institute of Economics and Business Management, Faculty of Economics and Administration, University of Pardubice, Czech Republic. She has been working with the strategic management, management analysis, financial reporting and financial management. She has published a number of papers concerning economics and finance.