



PREDICTING FINANCIAL DISTRESS FOR ROMANIAN COMPANIES

Gheorghe RUXANDA, Cătălina ZAMFIR*, Andreea MURARU

*The Faculty of Economic Cybernetics, Statistics and Informatics,
The Bucharest Academy of Economic Studies, Bucharest, Romania*

Received 16 March 2018; accepted 03 July 2018

Abstract. Using a moderately large number of financial ratios, we tried to build models for classifying the companies listed on the Bucharest Stock Exchange into low and high risk classes of financial distress. We considered four classification techniques: Support Vector Machines, Decision Trees, Bayesian logistic regression and Fisher linear classifier, out of which the first two proved to have the highest prediction accuracy. Classifiers were trained and tested on randomly drawn samples and on four different databases built starting from the initial financial indicators. As the literature related to the topic on Romanian data is very scarce, our study, by using a variety of methods and combining feature selection and principal components analysis, brings a new approach to predicting financial distress for Romanian companies.

Keywords: Support Vector Machines, Bayesian Analysis, Financial distress prediction, data mining, discriminant analysis, logistic regression.

JEL Classification: C58, C45, C11

Introduction

Companies' performance is a widely spread research subject and is of interest for researchers and practitioners alike. The possibility to predict companies' bankruptcy using their financial indicators was intensively studied starting half a century ago and it still represents an intriguing research topic. The methodologies approached in this respect range from very simple indicator-like scores and go to sophisticated supervised learning data mining techniques and neural networks.

Most studies in this area of research are focused on identifying a pattern in financial data which can predict future distress for companies by considering indicators with one to five years prior to the bankruptcy date, consequently, the financial indicators are calculated taking into account the bankruptcy date for companies.

*Corresponding author. E-mail: ionela.zamfir@csie.ase.ro

In our analysis we used financial ratios calculated from the balance sheets of end-2015 and we split the companies into two classes of risk of financial distress (low and high), according to the problems they encountered either in 2015 or in the subsequent two years.

The main objective of our paper can be seen as two-sided: on one hand we try to identify the classification method that best fits our data (in terms of accuracy rate) by comparing a relatively large number of techniques (discriminant analysis, logistic regression, support vector machine – SVM and decision trees – DT), using data selection and synthesizing variables (principal components analysis), and, on the other hand, we try to determine the most relevant economic indicators (out of the 30 financial ratios most frequently used in the literature) that could act as early warning indicators.

In this paper we took a novel approach to the insolvency problem, as, in order to avoid a false low risk classification of distressed companies, we considered as risky all companies that encounter problems either at the time of the balance sheet or in the subsequent two years. Also, in order to obtain a result that is not affected by the random selection of firms for the training and test sample, we repeated the experiment for 1000 draws and then averaged the results, obtaining therefore a more robust outcome. In terms of methodology we opted for a wide variety of techniques as there is a small number of studies in this field on Romanian data. Furthermore, we considered an extensive set of indicators based on balance sheets data, so as not to ignore any possible information.

The choice of the indicators was based on the existent literature and also on statistical tests, as criteria for their inclusion in models. Subsequently, the models were applied on four datasets: one containing all variables, one including the variables indicated by the selection criteria and other two databases consisting of the principal components extracted from the previous two data sets.

The results show good performance of all methods with average accuracy rates (for 1000 random samples) higher than 70%, reaching even 90% for decision trees. This method outperforms all classifiers for the training datasets, feature selection (Levene test) and principal components datasets, while support vector machines perform better for test datasets and feature selection models. From an economic perspective, accuracies over 85% in average for test datasets (30% of total companies) show that the financial indicators we used provide quality information about the risk of distress. This could be a starting point for the development of a general model that may be applied on Romanian companies from different markets.

The paper is structured as follows: literature review, methodological aspects, data and preliminary analysis, results and discussions and conclusions.

1. Literature review

The topic of bankruptcy prediction started to be approached by researchers since Altman's model in 1968, a model that still provides high accuracy rates and continues to be used on different datasets. The problems of Altman's classifier (normal distribution of variables and equal covariance matrices) were surpassed by Ohlson (1980) who was the first to introduce the logistic regression in predicting bankruptcy. Since then, a great number of studies treated

this subject and a wide variety of classification techniques were used to predict companies' failure. A comprehensive review of all aspects of methodology (models, samples, model integration, research directions) can be found in Sun, Li, Huang and He (2014).

Most studies in the field, especially recent ones, use multiple methods for predicting bankruptcy, focusing either on their independent use or consider ensemble models. In 2017, Barboza, Kimura and Altman use a database composed of more than 10000 companies from North America to identify bankruptcy one year prior to the event. Authors use methods like support vector machines, random forest, boosting compare them with artificial neural networks, logistic regression and discriminant analysis and conclude that the accuracy for machine learning techniques are, on average, higher with 10% than "traditional models". Methods like neural networks, support vector machines and decision trees are also used by Wang (2017) in order to identify the model with the highest prediction rate. The author shows that these relatively new methods have higher accuracy rates than genetic algorithms or logistic regression. In the same year, Wagle, Yang and Benslimane (2017) study this issue by taking into account 240 companies and 33 attributes. Using logistic regression, Bayesian network, decision tree and support vector machines as methodologies, and different methods to identify the proper financial indicators to use, the authors obtain accuracies between 60% and 70% on average and conclude that, among all methodologies approached, neural networks provide the most accurate results.

In 2015, Dellepiane, Marcantonio, Laghi, and Renzi start from the idea that even though the model proposed by Altman in 1968 still provides relevant information, there are country-specific macroeconomic indicators that may influence the accuracy of supervised learning models, like "gross sovereign rates with maturity of 1 and 10 years" and "the sovereign rating spread". Using a database of over 6900 companies, 28 financial ratios and 3 country specific indicators along with methods like Altman's Z-Score, features selection and support vector machines, authors conclude that macroeconomic variables are relevant for bankruptcy prediction, considering that the accuracy rates are high for most models (even over 80%). However, Alaminos, del Castillo, and Fernández (2016) develop a global model for predicting bankruptcy, by using logistic regression, and conclude that a global model provides much better results than regional ones.

In 2014, Burda and Pancarz analyze a database of over 3500 companies and 32 financial and economic indicators from 2000 to 2007 in order to identify a self-organizing map model that predicts bankruptcy with the highest accuracy. Their findings reveal that, using the Kohonen map, over 62% of non-bankrupt companies and over 72% of bankrupt companies were identified. On the other hand, Lopez-Iturriaga and Sanz (2015) use a combined model of multilayer perceptrons and self-organizing maps to identify bankruptcy on U.S. banks using data between 2002 and 2012. Their model shows a high accuracy rate, over 96% in detecting failed companies and authors conclude that this combination of methodologies provides better results than traditional models. Accuracies of over 90% were also obtained by Koklu and Tutuncu (2014), who considered a database of 175 companies and 6 attributes and tested classifiers like Naive Bayes Classifier, Multilayer Perceptron and Classification via Regression, and obtained the highest correct classification rates for the Bayes classifier. Pena, Martinez, and Abudu (2009) use, besides the traditional discriminant analysis and logistic regression,

Gaussian processes for classification, and make use of automatic relevance determination for ranking the ratios that enter the models according to their importance. The performance of logistic lasso and ridge regression are investigated by Pereira, Basto, and da Silva (2015) and compared to the stepwise method in equal and unequal sample sizes, the results being dependent on the type of sample used. Gepp and Kumar (2015) develop semi-parametric Cox survival analysis and non-parametric CART decisions trees and compare result with the conventional methods, showing that, on their dataset, data mining techniques led to better results. Kostopoulos, Karlos, Kotsiantis, and Tampakas (2017) use a dataset of Greek companies to analyze the active learning effectiveness in bankruptcy prediction. Their results show that the accuracy rate of this new method is higher than other supervised methods. Geng, Bose, and Chen (2015) predict financial distress for listed Chinese companies by using a wide variety of data mining techniques on a set of 31 financial ratios and various test and train samples. They find that neural networks offer the best results and also provide a ranking of indicators according to their importance in predicting distress.

Besides the modelling techniques, another important issue in predicting bankruptcy is the selection of indicators. A review of the ways in which indicators were selected in the bankruptcy prediction literature can be found in du Jardin (2009). In some studies, aggregate indicators are developed starting from raw data. Succuro (2017) builds a model for predicting firm bankruptcy in European countries and finds that indebtedness, an aggregate indicator built by the author, is relevant for prediction in countries that are more bank-oriented whereas in economies with more developed financial markets it is less important. The problem of data selection in predicting financial distress was also a topic of research in Tian, Yu, and Zhou (2015). They conclude that for binary choice models the best suited are complete data sampling techniques whereas when different classification costs are considered it is better to establish a cut-off value for the probability. The literature dedicated to the stock exchange is very vast and covers a wide variety of topics, see for example Barak, Dahooie, and Tichý (2015) who predict stock prices and obtained high accuracy rates, of 87%, proving that combining several methodologies can lead to a novel model to model a specific economic problem. Studying the same problem, Stádník, Raudeliūnienė, and Davidavičienė (2016) use Fourier analysis to forecast stock market price for US stocks price series.

There are few studies concentrated on forecasting financial distress in emerging markets. For Romania, Prodan-Palade (2017) used financial ratios (from 2010 to 2014) for 55 companies from manufacturing industry that are listed at Bucharest Stock Exchange and two backpropagation neural network models. The author build neural network models to predict the overall liquidity ratio and overall solvency ratio in order to identify the risk of bankruptcy based on these indicators. Another study on financial distress in emerging markets is that of I. Pervan, M. Pervan, and Vukoja (2011), who confirm that there is information content in public financial statements for bankruptcy prediction of Croatian firms by the use of logistic regression and discriminant analysis. In 2018, Erdogan and Akyüz use longitudinal data and support vector machines to analyze a dataset that contains 41 Turkish commercial banks. Their findings reveal a good accuracy rate, especially if an appropriate normalization technique is used.

2. Methodological approach

The literature we reviewed shows a high interest in and a wide use of methodologies like support vector machines, logistic regression, decision trees, discriminant analysis and neural networks. We selected four methods for predicting companies' financial distress: discriminant analysis (Fisher linear classifier), logistic regression (variational Bayesian inference logistic regression), support vector machines and decision trees. The choice of these methods was justified by their popularity on account of their high accuracy rates on both training and testing samples, and on both small and large datasets, which proved to be true also on our data. Alternative techniques consist of neural networks methods like Kohonen self-organizing maps, backpropagation neural network models or multilayer perceptrons, and other methods like CART, K-means algorithm, Naive Bayes Classifier.

The methodology we decided to apply covers a wide spectrum as we considered it necessary to apply multiple algorithms in order to provide a consistent conclusion, adequate for the economic reality in our data. In fact, most authors use multiple techniques for predicting financial distress, as their outcome could be different from one dataset to another (for example du Jardin, 2017; Barboza, et al., 2017; Geng et al., 2015). In what concerns the logistic regression, we used three specifications of the Bayesian form with the goal of identifying the most accurate of them, as in recent studies the relative performance of this traditional method against other techniques proved to be weaker (Barboza et al., 2017). The selected techniques complement each other in the sense that two of them, as it will be seen in the following parts, with the best fit are accompanied by two methods that allow the identification of indicators' coefficients and therefore permit a ranking of warning indicator variables.

We applied the before mentioned techniques on training samples and then validated them on test samples. These samples were randomly selected from the 283 firms, 70% of them being used for training the classifiers and 30% for validating them. In both the training and test samples, the real proportion of low and high risk companies was preserved. We repeatedly extracted samples (1000 successive draws), performed each analysis on all samples and then averaged the results. We used two datasets, one consisting of all variables and the other containing only the variables indicated as significant by the statistical tests applied (details in the data section). Furthermore, we also tested the classification abilities of the methods on principal components in both data sets. Principal component analysis has not only the advantage that it offers the possibility to reduce the number of variables without significant loss of information, but it also offers the benefit of obtaining independent components, therefore it eliminates informational overlap.

2.1. Fisher linear classifier

Linear discriminant functions are amongst the first techniques used in studies predicting bankruptcy. The Fisher linear classifier provides the greatest separability between the class means while ensuring minimum within-class variance and maximum between-class variance. If we consider the linear discriminant function as $z = w^t x$, the criterion can be written as:

$$J(w) = \frac{w^t S_B w}{w^t S_w w}, \tag{1}$$

where S_B is the between groups covariance matrix and S_w is the within groups covariance matrix and by differentiating with respect to w , we obtain the solution (Bishop, 2006). It result that:

$$\tilde{w} = \arg \max_w \left\{ \frac{w^t S_b w}{w^t S_w w} \right\}, \tag{2}$$

\tilde{w} is an eigenvector of $S_w^{-1} S_b$, associated with the highest (positive) eigenvalue. At the same time, \tilde{w} is associated with the axis along which the objects' coordinates have minimum within- and maximum between-class variability.

After the score is calculated ($z_i = \tilde{w}_1 x_{1i} + \tilde{w}_2 x_{2i} + \dots + \tilde{w}_p x_{pi}$, p – number of explanatory variables, and i – number of the object, $i = 1, \dots, n$), it is compared with a threshold value, and the object consequently classified into one of the classes. Even though the best accuracy of this method is obtained when the two classes have the same variance-covariance matrix, it continues to be frequently used in predicting firm bankruptcy as it proved to offer fairly robust results.

2.2. Variational Bayesian inference logistic regression

Logistic regression is considered as one of the traditional methods for classifying firms in distress. In our study, we used the variational Bayesian inference logistic regression (VBILR) for predicting the risk of bankruptcy. For performing the analysis we used the MATLAB code made public by Jan Drugowitsch¹ and the accompanying theoretical background. The method was applied (i) with hyper-priors on parameters and automatic relevance determination (ARD); (ii) with hyper-priors and without ARD and (iii) without hyper-priors and without ARD.

As in any Bayesian estimation, the goal is to obtain the posterior distribution. For the logistic regression

$$P(y = 1 / X, \theta) = \frac{1}{1 + \exp(-\theta^T X)} = \sigma(\theta^T X), \tag{3}$$

(where y is the binary response variable, $X = \{X_1, X_2, \dots, X_n\}$ the n explanatory variables, θ the vector of parameters) the sigmoid $\sigma(\cdot)$ doesn't admit a conjugate prior, so an analytic expression for posterior $P(\theta / X, Y)$ can be found through approximations and has the form of the conjugate Gaussian prior (the model described below is the one in Drugowitsch (2017)):

$$P(\theta / \alpha) = N(\theta / 0, \alpha^{-1} I) = \left(\frac{\alpha}{2\pi} \right)^{1/2} \exp\left(-\frac{\alpha}{2} \theta^T \theta \right) \tag{4}$$

and the α hyper-parameter is modelled by a conjugate Gamma distribution ($P(\alpha) = \text{Gam}(\alpha; a_0, b_0)$).

The variational bound inference falls in the category of deterministic approximation schemes for posteriors in a Bayesian model, being, therefore, in the opposite category of stochastic or sampling methods (as for example MCMC). They do not offer an exact

¹ The code can be found at GitHub (n.d.) website https://github.com/DrugowitschLab/vb_logit.

solution, but their advantage over sampling methods could be the monotonically improving approximations they offer (Jaakkola & Jordan, 2000). The evaluation of the posterior distributions by the use of variational bound inference leads to approximate analytical expressions for the posterior. The variational inference is a generalization of the EM algorithm. According to this, the log-likelihood can be decomposed into a functional $\mathcal{L}(Q)$, which is called the lower bound ($Q(z)$ is the variational distribution) and the Kullback-Leibler divergence. By maximizing the lower bound, and implicitly minimizing the KL divergence, the variational distribution $Q(z)$ (in our case $Q(\theta, \alpha)$) will approximate the posterior ($P(z/X)$ (in our case $P(\theta, \alpha / X, Y)$)². For the logistic model the marginal log-likelihood has the form:

$$\begin{aligned} \ln P(Y / X) &= \iint P(Y / X, \theta) P(\theta / \alpha) P(\alpha) d\theta d\alpha = \\ & \prod_n P(y_n / X_n, \theta) P(\theta / \alpha) P(\alpha) d\theta d\alpha = \prod_n \sigma(y_n \theta^T X_n), \end{aligned} \tag{5}$$

while the lower bound has the expression:

$$\ln P(Y / X) \geq \mathcal{L}(Q) = \iint Q(\theta, \alpha) \ln \frac{P(Y / X, \theta) P(\theta / \alpha) P(\alpha)}{Q(\theta, \alpha)} d\theta d\alpha \tag{6}$$

and factorization of Q is assumed: $Q(\theta, \alpha) = Q(\theta) Q(\alpha)$.

The lower bound on the sigmoid function is (for computational details see Bishop, 2006):

$$\sigma(z) \geq \sigma(\xi) \exp\left(\frac{z - \xi}{2} - \lambda(\xi)(z^2 - \xi^2)\right), \tag{7}$$

where $\lambda(\xi) = \frac{1}{2\xi} \left(\sigma(\xi) - \frac{1}{2}\right)$.

By substituting (7) into (5), the data log-likelihood can be expressed as lower-bounded by:

$$\ln P(Y / X, \theta) \geq \ln h(\theta, \xi). \tag{8}$$

The expression of $\ln h(\theta, \xi)$ contains a local variational parameter ξ_n . Therefore, a new variational bound is constructed:

$$\tilde{\mathcal{L}}(Q, \xi) = \iint Q(\theta, \alpha) \ln \frac{\ln h(\omega, \xi) P(\theta / \alpha) P(\alpha)}{Q(\theta, \alpha)} d\theta d\alpha, \tag{9}$$

which is $\tilde{\mathcal{L}}(Q, \xi) \leq \mathcal{L}(Q)$. By applying standard variational methods for factorized distribution, variational posteriors can be inferred (see Drugowitsch, 2017; Bishop, 2006).

In the case of automatic relevance determination the difference is that each element of θ has its own different prior, so the posterior can be written as

$$P(\theta / \alpha) = N(\theta / 0, A^{-1}) = \frac{|A|^{1/2}}{\sqrt{2\pi}^D} \exp\left(-\frac{1}{2} \theta^T A \theta\right). \tag{10}$$

A being a diagonal matrix with the $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_D)^t$ vector on its diagonal, and α follows the conjugate hyper-prior $P(\alpha) = \prod_i \text{Gam}(\alpha_i / a_0, b_0)$. Such a framework automatically determines the relevance of variables in predicting y , as α_i sets the precision of θ_i .

² We used the models as described by Drugowitsch (2017), therefore all computational details can be found in his paper. Further extensive treatment of variational inference can be found in Bishop (2006).

2.3. Decision trees

Decision trees are often used as decision support systems and many researchers use this methodology in supervised learning for problems like: character recognition, medical diagnosis, speech recognition (Safavian & Landgrebe, 1991). Starting with the definition of a graph, a path in a graph, edge, node, depth of a node and decision rule, Safavian and Landgrebe (1991) present the reasons for using decision trees classifiers (DTC) and the drawbacks for avoiding this methodology. Among the advantages for DTC, the most important is that decision trees use less computation by choosing the variables that optimally discriminate between classes and uses, in this way, less features at each node, without decreasing the performance of classifier. The main disadvantage of decision trees is that, for large amount of data and classes, the resources used for training increase and for large trees, errors may occur and it is difficult to optimize a tree by keeping both performance and efficiency at a certain accepted level.

2.4. Support vector machines

Support vector machines is a methodology developed by Vapnik (1998) that uses a minimization problem in order to identify support vectors that define the optimal hyperplan. Vapnik proposed several problems depending on the type of separability: linear or non-linear (cases of hard-margin or soft-margin). For each type of problem there is another Lagrange function that provide the support vectors.

The definition of support vector machines starts from a dataset of T observations, n descriptive variables and an $n + 1$ variable that represents the class and could be $+1$ or -1 (for the example of two classes). Supposing that classes are linearly separable, the separation hyperplane is given by the function: $h(x) = w^t x + w_0$ and the set of points

$H(w, w_0) = \{x \in \mathbb{R}^n \mid w^t x + w_0 = 0\}$. If we consider a vector $w^{(0)} = \left(0, 0, \dots, 0, -\frac{w_0}{w_n}\right)^t$, then

the hyperplan $H(w, w_0)$ is a translation of subspace $L(w)$ with the vector $w^{(0)}$, where $L(w) = \{x \in \mathbb{R}^n \mid w^t x = 0\}$ and the vector w is orthogonal both on subspace $L(w)$ and hyperplan $H(w, w_0)$. Considering that weight vector w and the number w_0 are scaled such that the following normalization conditions are true:

$$\min_{x \in \omega_-} |h(x)| = \min_{x \in \omega_+} |h(x)| = 1 \Rightarrow \min_{x \in \omega_-} |w^t x + w_0| = \min_{x \in \omega_+} |w^t x + w_0| = 1, \tag{11}$$

where ω_+ and ω_- represents the two classes defined by variable $n + 1$ (positive and negative class). These conditions have several consequences: the hyperplan defined like this is at the middle of distance between the closest neighbours from classes (distance represented by the inverse of the length of the vector w). The distance (Euclidian distance) from a point $x^{(i)}$ to hyperplan $H(w, w_0)$ is:

$$d(x^{(i)}, H(w, w_0)) = \frac{1}{w} w^t (x^{(i)} - w^{(0)}) = \frac{1}{w} w^t x^{(i)} - w^t w^{(0)} = \frac{1}{w} w^t x^{(i)} - w_0, \tag{12}$$

while the distance from origin to hyperplan is:

$$d(0, H(w, w_0)) = \frac{1}{w} w^t 0_v - w_0 = \frac{|w_0|}{w}. \tag{13}$$

Also, for each observation $x^- \in \omega_-$ (taking into account the normalization from above and the fact that the hyperplan $H(w, w_0)$ is defined by $w^t x + w_0 = 0$), the following inequality is true: $w^t x^- + w_0 \leq -1$, while for each $x^+ \in \omega_+$, the following inequality is true: $w^t x^+ + w_0 \geq 1$. This leads to $y_-(w^t x^- + w_0) \geq 1$. Another consequence given by pervious normalization conditions is the possibility of defining another two hyperplans (considered to be margins for classes separation) that are parallel to hyperplan $H(w, w_0)$: $H_-(w, w_0) = \{x \in \mathbb{R}^n \mid w^t x + w_0 = -1\}$ and $H_+(w, w_0) = \{x \in \mathbb{R}^n \mid w^t x + w_0 = +1\}$. At last, the most important consequence is the possibility of writing the optimization problem that represents the main idea behind support vector machines technology: finding the hyperplan $H(w, w_0)$ such that the distance between hyperplans $H_-(w, w_0)$ and $H_+(w, w_0)$ is maxim. Each distance from a margin hyperplan to hyperplan $H(w, w_0)$ is $\frac{1}{w}$, so that the equivalent of maximization the distance between hyperplans $H_-(w, w_0)$ and $H_+(w, w_0)$ is to minimize the length of vector w . In this way, the optimization problem is:

$$\begin{cases} \min \frac{1}{2} w^2 \\ y_i (w^t x^{(i)} + w_0) \geq 1; i = 1, 2, \dots, T \end{cases} \tag{14}$$

and the Lagrange function associated to this problem is:

$$\mathcal{L}(w, w_0, a_1, a_2, \dots, a_T) = \frac{1}{2} w^2 - \sum_{i=1}^T a_i [y_i (w^t x^{(i)} + w_0) - 1]. \tag{15}$$

Necessary conditions are:

$$\begin{cases} \frac{\partial \mathcal{L}(w, w_0, a_1, a_2, \dots, a_T)}{\partial w^t} = w - \sum_{i=1}^T a_i y_i x^{(i)} = 0; \\ \frac{\partial \mathcal{L}(w, w_0, a_1, a_2, \dots, a_T)}{\partial w_0} = \sum_{i=1}^T a_i y_i = 0. \end{cases} \Rightarrow \begin{cases} \sum_{i=1}^T a_i y_i = 0; \\ w = \sum_{i=1}^T a_i y_i x^{(i)}. \end{cases} \tag{16}$$

Because Lagrange multipliers associated with restrictions (that are verified with equality in optimal solution) are strictly positive and because this happens only for observations that belong to margins hyperplans ($H_-(w, w_0)$ and $H_+(w, w_0)$), these vectors are named support vectors.

On the other hand, the complementary conditions of Karush-Kuhn-Tucker conditions gives the following relation: $a_i [y_i (w^t x^{(i)} + w_0) - 1] = 0, i = 1, 2, \dots, T$, that leads to:

- if $y_i (w^t x^{(i)} + w_0) > 1$, then $a_i = 0$ and $x^{(i)}$ is just on observation;
- if $y_i (w^t x^{(i)} + w_0) = 1$, then $a_i > 0$ and $x^{(i)}$ is a support vector.

In these condition, function $h(\cdot)$ is:

$$h(x) = \sum_{i=1}^T a_i y_i \left[(x^{(i)})^t x \right] + w_0, \tag{17}$$

while the dual Lagrange function is:

$$\mathcal{L}(w, w_0, a_1, a_2, \dots, a_T) = \frac{1}{2} w^2 - w^t \sum_{i=1}^T a_i y_i x^{(i)} - w_0 \sum_{i=1}^T a_i y_i + \sum_{i=1}^T a_i. \tag{18}$$

Using the primal optimization problem conditions and replacing w from optimal conditions in Lagrange function, the result is dual form (Wolf) of Lagrange function:

$$\mathcal{L}_D(a_1, a_2, \dots, a_T) = \sum_{i=1}^T a_i - \frac{1}{2} \sum_{i=1}^T \sum_{j=1}^T a_i a_j y_i y_j \left(x^{(i)}\right)^t x^{(j)}. \tag{19}$$

Making the notations: $z^{(i)} = y_i x^{(i)}$; $z^{(j)} = y_j x^{(j)}$, considering $Z_{T \times n}$ the matrix of normalized forms (observations) and $a_{T \times 1}$ the vector of Karush-Kuhn-Tucker multipliers, the Lagrange function becomes

$$\mathcal{L}_D(a) = a^t 1_T - \frac{1}{2} a^t Z Z^t a. \tag{20}$$

The multipliers vector (a) is the optimal solution for the problem:

$$\left\{ \begin{array}{l} \max_a \mathcal{L}_D(a) = \max_a \left\{ a^t 1_T - \frac{1}{2} a^t Z Z^t a \right\} \\ a_i \geq 0; \quad i = 1, 2, \dots, T \end{array} \right\}. \tag{21}$$

The solution for this problem is:

$$b = \sum_{i=1}^{n_{sv}} \left(\tilde{w}^t \tilde{x}^{(i)} - y_i \right), \tag{22}$$

where sv means support vectors.

3. Data description and preliminary analyses

The dataset contains 283 companies, all listed on the Bucharest Stock Exchange, 41 distressed companies, that have or had insolvency problems, and 242 safe companies, with low risk of bankruptcy– and its sources are The Bucharest Stock Exchange (n.d.) and the National Trade Register Office (n.d.) websites³. We considered as distressed all companies that have files with resolutions like: opening of the general judicial reorganization procedure or opening of the general insolvency/bankruptcy procedure. In the same category we included the companies whose financial problems have started in the past. This categorization provides the insurance that we don't ignore any of the firms with financial distress, in order to avoid a false-negative classification, but doesn't say that bankruptcy necessarily follows. Financial indicators are all collected and computed for the end of 2015. Using financial data, we computed 30 ratios, most of them reflect the liquidity, solvability and profitability for analyzed companies and are used in many research studies. The ratios are: I1 = long term debt/equity, I2 = turnover/fixed assets, I3 = turnover/total assets, I4 = gross profit/(long term debt + equity), I5 = operating result/turnover, I6 = turnover/number of employees, I7 = equity/total assets, I8 = current assets/short term debt, I9 = (inventory + current assets)/short term debt, I10 = (cash + short term investments)/short term debt, I11 = total debt/total assets, I12 = total assets/short term debt, I13 = turnover/current assets, I14 = net income/total assets, I15 = net income/equity, I16 = net income/turnover, I17 = net income/total expenses, I18 = equity/fixed assets, I19 = total assets/(short term debt + long term debt), I20 = net income/number of issued shares,

³ Retrieved from <http://www.bvb.ro/>, accessed on February 2018 and from <https://www.onrc.ro/index.php/ro/>, accessed on February 2018, respectively.

I22 = gross income/turnover, I23 = total debt/equity, I24 = current assets/total assets, I25 = (current assets – short term debt)/total assets, I26 = operating result/total assets, I27 = operating cash flow/number of employees, I28 = operating cash flow/total assets, I29 = gross income/turnover, I30 = total assets/gross income. We tried to use as many variables as possible starting from the available information in the balance sheets, in order to let the data speak for itself. There is a very limited number of studies on financial distress prediction on Romanian data, therefore there isn't yet a traditional core of variables on which we could rely a priori as early warning indicators.

Feature selection represents one of the most important issues in this type of problems, because a large number of indicators may decrease the performance of classifiers and may increase the training time. This is the reason why there are many methods to reduce the number of indicators, some taking into account the discriminative power for indicators (methods that test if the variance or mean of a variable is the same in both classes, like Levene test, t-test, F-test, Wilk's lambda, Bartlett test), while some consider size reduction methods (like principal components analysis, factor analysis). The number of financial indicators computed in this case is high and may influence the final results.

The results of the Kolmogorov-Smirnov test show a big distance from normality (only indicators I3, I7, I11, I24 and I25 can be compared to normal distribution, but even these variables are far from the normal distribution). From this point of view, in order to test the power to discriminate between bankrupt and safe companies, Levene test (that is less sensitive to normality deviation than Bartlett test) was applied to each financial ratio. A p-value lower than 0.05 (Table 1) leads to the rejection of the null hypothesis that states that the indicator has equal variance in both classes (high and low risk of bankruptcy), hypothesis that implies a low discriminative power. Therefore, only 12 of the 30 indicators may be used for further analysis, according to this test. To analyze the influence of feature selection over accuracy rates of classifiers, we consider both methods: a dataset of 12 indicators (named *std_12*, due to standardizing the original variables) and another dataset of all 30 indicators (named *std_30*).

Table 1. Levene test p-value

Variable	prob. Levene	Variable	prob. Levene	Variable	prob. Levene
I1	0.00	I11	0.00	I21	0.71
I2	0.67	I12	0.27	I22	0.02
I3	0.87	I13	0.26	I23	0.00
I4	0.00	I14	0.01	I24	0.41
I5	0.96	I15	0.00	I25	0.00
I6	0.58	I16	0.86	I26	0.00
I7	0.00	I17	0.00	I27	0.73
I8	0.34	I18	0.48	I28	0.79
I9	0.32	I19	0.31	I29	0.91
I10	0.01	I20	0.26	I30	0.75

Table 2. Correlation matrix for 12 indicators database

	I1	I4	I7	I10	I11	I14	I15	I17	I22	I23	I25	I26
I1	1.00	-0.02	-0.09	-0.02	0.10	-0.02	-0.79	-0.02	-0.02	0.98	-0.02	-0.03
I4	-0.02	1.00	-0.01	0.00	0.01	0.66	0.29	0.29	0.68	-0.03	0.07	0.46
I7	-0.09	-0.01	1.00	0.18	-0.98	0.03	0.10	0.01	0.04	-0.12	0.67	-0.01
I10	-0.02	0.00	0.18	1.00	-0.22	0.01	0.02	0.09	0.01	-0.04	0.20	-0.02
I11	0.10	0.01	-0.98	-0.22	1.00	-0.03	-0.10	-0.01	-0.04	0.12	-0.66	0.01
I14	-0.02	0.66	0.03	0.01	-0.03	1.00	0.13	0.53	0.97	-0.02	0.06	0.78
I15	-0.79	0.29	0.10	0.02	-0.10	0.13	1.00	-0.04	0.13	-0.78	0.07	-0.04
I17	-0.02	0.29	0.01	0.09	-0.01	0.53	-0.04	1.00	0.50	-0.02	0.16	0.59
I22	-0.02	0.68	0.04	0.01	-0.04	0.97	0.13	0.50	1.00	-0.02	0.10	0.87
I23	0.98	-0.03	-0.12	-0.04	0.12	-0.02	-0.78	-0.02	-0.02	1.00	-0.01	-0.04
I25	-0.02	0.07	0.67	0.20	-0.66	0.06	0.07	0.16	0.10	-0.01	1.00	0.12
I26	-0.03	0.46	-0.01	-0.02	0.01	0.78	-0.04	0.59	0.87	-0.04	0.12	1.00

The correlation matrix (Table 2, above) shows high correlations between some variables, due to the fact that almost all ratios are computed using the balance sheet and cash flow statement that contain highly correlated financial indicators. The high correlation between variables like I1 and I15, I1 and I23, I4 and I22, I7 and I11, I14 and I22, I22 and I26 reveal large informational redundancy that may affect classifiers results. A method of reducing the redundant information is principal components analysis.

Table 3. PCA results

12 variables	Eigen values	Explained (%)	Explained cum. (%)
1	3.68	30.67	30.67
2	2.85	23.78	54.45
3	2.43	20.21	74.67
4	0.96	7.98	82.64
5	0.82	6.85	89.49
6	0.51	4.28	93.77
7	0.38	3.19	96.97
8	0.20	1.69	98.66
9	0.11	0.94	99.60
10	0.02	0.18	99.78
11	0.01	0.12	99.90
12	0.01	0.10	100.00

30 variables	Eigen values	Explained (%)	Explained cum. (%)
1	5.72	19.08	19.08
2	4.39	14.63	33.71
3	3.30	11.01	44.72
4	2.91	9.69	54.41
5	2.36	7.86	62.27
6	2.11	7.03	69.30
7	1.38	4.59	73.88
8	1.34	4.45	78.34
9	1.05	3.48	81.82
10	1.01	3.37	85.19
11	0.94	3.15	88.34
12	0.81	2.70	91.04

Principal components analysis is also a method of size reduction. Performing this analysis on both original standardized datasets, we obtain 2 new datasets, w_12 and w_30, each corresponding to an original number of variables (12 or 30). Table 3 shows the results of principal components analysis for both datasets. Considering 12 indicators (the ones that passed the Levene test), we can compute and retain 3 principal components (according to Kaiser criterion) that preserve 74.66% of information from all 12 variables. On the other hand, when taking into account all 30 financial ratios we can reduce them into 10 principal components, which cumulate 85.19% of total information.

Table 4. Factor matrix

	I1	I4	I7	I10	I11	I14	I15	I17	I22	I23	I25	I26
1	-0.22	0.72	0.19	0.08	-0.19	0.91	0.30	0.64	0.94	-0.22	0.26	0.84
2	0.74	0.14	-0.63	-0.21	0.64	0.25	-0.67	0.19	0.24	0.75	-0.45	0.28
3	0.59	-0.10	0.69	0.25	-0.69	-0.02	-0.54	0.08	-0.01	0.57	0.65	0.01

Using the correlation matrix from above (Table 4), we can define and name the retained principal components for the dataset that has 12 financial ratios, as follows: the first principal component is the profitability component because it highly correlates with ratios that represent profitability, the second principal component is the component that reflects the equity state, while the third component represents the indebtedness and assets component.

Figure 1 shows the graphical representation of classes in the principal plan, composed by the first two principal components (that contain over 50% of total relevant information for financial distress prediction). Companies with high bankruptcy risk have big values for component 1, which represent the profitability and is highly negatively correlated with indicators for profitability.

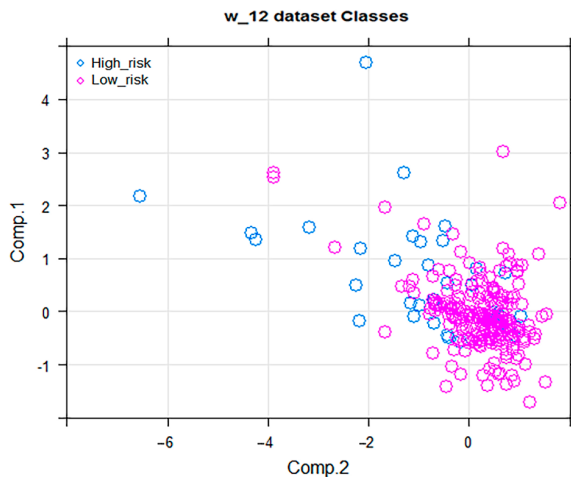


Figure 1. Graphical visualization of classes

4. Results and discussions

As mentioned previously in the paper, the classifiers were applied on the 4 datasets that resulted from the preliminary analyses presented above: 2 datasets of original standardized variables and 2 datasets of principal components.

For training the classifiers, we used the proportion 70–30, which means 70% of each dataset is used for training, and 30% for testing. This method is applied for 1000 times, because each time, the datasets are split randomly. This way, the distribution of classes (on average) for training sets is presented in the Table 5, below. About 14–15% of companies in training sets have a high bankruptcy risk, while 85–86% of companies have low bankruptcy risk.

Figure 2 is an example of the 1000 decision trees build for 1000 samples of train_w_30 dataset. Only 4 of the 10 principal components were used in building this decision tree and each node presents the decision rule for a specific class, the probability of each class at the node and the percentage of companies used for each node.

Table 5. Class distribution on average for training set

Dataset	Average % High_risk	Average % Low_risk
train_std_12	14%	86%
train_std_30	15%	85%
train_w_12	14%	86%
train_w_30	14%	86%

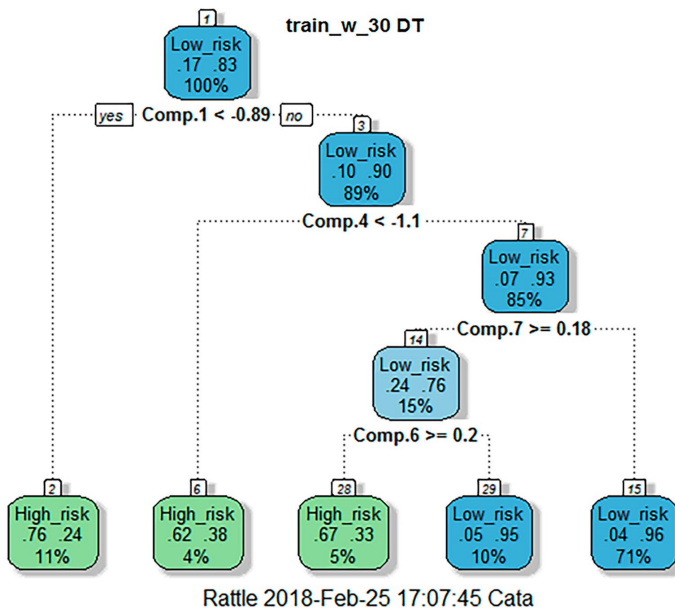


Figure 2. Decision tree example for train_w_30 dataset

We obtained the classifiers using R and Matlab. In order to identify the effectiveness of each classifier, an average accuracy rate (for all 1000 times of randomly split datasets) was computed for each dataset used. In this way, methods like cross-validation and resubstitution were applied. The table below (Table 6) shows the average accuracy rates of models (highest figures for samples are in bold). For decision trees, the correct classification rates if we substitute the training data are over 90%, while using the test datasets, the accuracies decrease to 86% on average. The same happens for support vector machines classifier, but the decrease of accuracy rates is about 1–2%. Moreover, comparing the original datasets (that have 100% of information) with principal components datasets, we notice that, on average, there is no significant difference of the correct classification rates, that means that, the amount of lost information (25% for 12 variables and 15% for 30 variables) has almost no impact on final results. The fact that DT and SVM provide the best classification is in line with the expectations we had from the literature (like Dellepiane et al., 2015; Burda & Pancerz, 2014; Wagle et al., 2017). The fact that machine learning techniques offers better results than traditional models was also shown by Barboza et al. (2017).

Table 6. Average accuracy rates of models

Dataset	Average accuracy rates				Average error rates			
	DT	SVM	LD	VBILR ⁴	DT	SVM	LD	VBILR ⁵
test_std_12	86%	87%	78%	69%	14%	13%	22%	31%
test_std_30	86%	86%	76%	75%	14%	14%	24%	22%
test_w_12	86%	87%	86%	71%	14%	13%	14%	29%
test_w_30	85%	87%	88%	68%	15%	13%	12%	32%
train_std_12	91%	89%	81%	71%	9%	11%	19%	29%
train_std_30	91%	90%	85%	73%	9%	10%	15%	22%
train_w_12	90%	88%	83%	70%	10%	12%	17%	30%
train_w_30	91%	89%	84%	66%	9%	11%	16%	34%

The forecasting ability of a simple Fisher linear classifier proved to be weaker than the previous two methods, but better than that of any form of logistic regression we tried to fit on our data, on almost all cases. Its performance decreases marginally when applied on the test sample, on average to 82% from 84%.

Variational Bayesian inference logistic regression, even though has an accuracy of more than 70% on average both on training and test sample, seems to be the worst performer in the group. Because there were groups of highly correlated variables both in the large (30 variables) and small (12 variables) sample, we eliminated some of the variables when estimating the logistic regression on standardized variables, so that evident multicollinearity

⁴ Maximum from the three specifications.

⁵ Minimum from the three specifications.

would be avoided⁶. The decision regarding which indicators to keep from a group of highly correlated variables was the result of the Levene test, Wilk’s lambda, the estimation of a stepwise regression model or the impact that variables have in the Fisher classifier⁷. Still, the removal of variables had no impact on the classification accuracy.

As mentioned in the methodology section, the logistic regression has been estimated⁸ with three different specifications: with hyper-priors, with hyper-priors and automatic relevance determination and without hyper-priors and ARD⁹, in the above table we entered the figures that represent the maximum forecasting ability of the three. Amongst them the variant with ARD generally seems to be the best classifier; its predictive ability is marginally exceeded by that of the other two VBILR specifications in 3 out of the 8 cases (Tables 7 and 8).

Another advantage of the VBILR with ARD is that it provides a clearer differentiation between probabilities, as it also provides shrinkage, therefore it is mostly efficient when a larger number of variables is used. All three models have fair to good accuracy in terms of the AUC of the ROC curve (Table 9 and Figure 3).

Table 7. VBILR predictive ability – training sample

	train_w_12	train_std_12	train_std_30	train_w_30
HyP, no ARD	0.69	0.70	0.70	0.65
No HyP, no ARD	0.70	0.70	0.72	0.66
HyP, ARD	0.69	0.71	0.73	0.65

Table 8. VBILR predictive ability – test sample

	test_w_12	test_std_12	test_std_30	test_w_30
HyP, no ARD	0.69	0.69	0.73	0.65
No HyP, no ARD	0.69	0.69	0.74	0.64
HyP, ARD	0.71	0.68	0.75	0.68

Table 9. Area under curve of ROC curves

	train_w_12	train_std_12	train_std_30	train_w_30
HyP, no ARD	0.79	0.80	0.84	0.79
No HyP, no ARD	0.80	0.80	0.85	0.79
HyP, ARD	0.75	0.81	0.86	0.76

⁶ There still remained moderately correlated variables.

⁷ In the large sample, we identified three groups of highly correlated variables ($\rho > 0.7$) and we removed variables I9, I12, I19, I20, I16, I5, I21, I27, I22, I14, I26 from the logit. In the small sample, we eliminated I1, I14 and I26.

⁸ Uninformative priors were used.

⁹ The last variant uses a conjugate Gaussian prior and not a hyper-prior on α . It increases shrinkage as dimension grows (Drugowitsch, 2017).

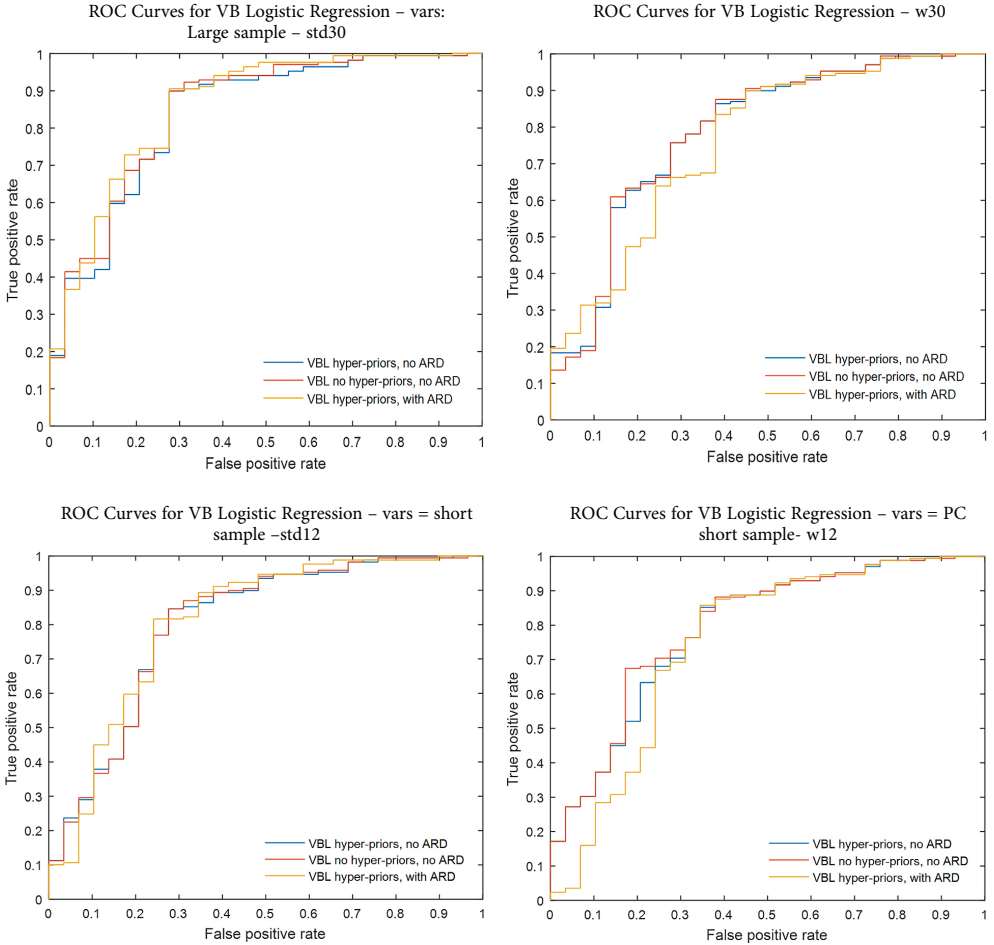


Figure 3. Area under curve of ROC curves

Interpreting the coefficients of the variables (both in the logistic regression and Fisher linear classifier) there are 4 that seem the most important in discriminating among the two groups, and these are total debt/total assets (I11), equity to assets (I7), gross profit/(long term debt + equity) (I4) and gross income/turnover (I22). What is striking about these indicators is that liquidity ratios, which usually in the literature are a good early warning indicators, are missing from the top 4 list. Still, the ratios that proved to be highly discriminative are important indicators of the health of balance sheets (most of them being leverage ratios).

Conclusions

Even though our approach to the problem of financial distress is different than that of other researchers, as we considered the high risk group as containing firms which either were in financial distress in 2015 or faced the problems in the subsequent two years, the results show

high accuracy rates for all classifiers. As there is a very limited number of studies on this topic on Romanian data, our study, by the use of multiple methods and combining feature selection and principal components analysis brings a new avenue to predicting financial distress for Romanian companies.

The application of feature selection methods (Levene and principal components analysis) led to surprising results. The minor loss of information induced by the use of principal components and by selecting only the most relevant financial indicators has, on average, almost no impact on the average accuracy rate for decision trees, support vector machines and linear discriminant classifiers. But, for variational Bayesian inference logistic regression model, taking into account 12 out of the 30 financial ratios produces a decrease of accuracy of about 6% for training sets (containing all indicators), and 2% for test sets (12 indicators). Also, their accuracy for the datasets that contain principal components is lower. Of all the methods that we used for classification, decision trees and support vector machines have the best performance on our datasets, with average accuracy rates of 90% in the training sample and 86–87% in the test sample. Concerning the variables with the highest discriminant power, our analyses identified the following factors: total debt/total assets, equity to assets, gross profit/(long term debt + equity) and gross income/turnover. As they proved to have the highest loadings in the fisher discriminant function and in the logistic regression, they could be considered as important indicators for the firms' health.

Considering the satisfactory economic results and the easy implementation of the methods, this study might be a starting point for developing a decision support system both for the management of listed companies and for investors. High accuracy rates for principal components and feature selection datasets show the importance of choosing the right financial indicators who could discriminate better between safe and distressed companies. Also, the high correlations between the financial indicators considered are relevant when analyzing bankruptcy because solvability, profitability, liquidity and activity indicators – who differentiate between companies with low insolvency risk and companies with high risk of bankruptcy –, are interconnected. All in all, we are aware that the database used is not large enough to create a model generally valid for predicting financial distress, but these methods proved to be highly efficient and could be successfully applied on different datasets. Moreover, further improvements could be done in future research both from the perspective of using various years of data, and also from the point of view of integrating time and enlarging the range of methodologies.

Author contributions

Cătălina Zamfir was responsible for data collection and description. All other sections are the result of all the authors' work.

Disclosure statement

Authors do not have any competing financial, professional, or personal interests from other parties.

References

- Alaminos, D., del Castillo, A., & Fernández, M. A. (2016). A global model for bankruptcy prediction. *PLoS ONE*, *11*(11), Article ID e0166693. <https://doi.org/10.1371/journal.pone.0166693>
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 589-609. <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>
- Barak, S., Dahooie, J. H., & Tichý, T. (2015). Wrapper ANFIS-ICA method to do stock market timing and feature selection on the basis of Japanese Candlestick. *Expert Systems with Applications* *42*(23), 9221-9235. <https://doi.org/10.1016/j.eswa.2015.08.010>
- Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, 405-417. <https://doi.org/10.1016/j.eswa.2017.04.006>
- Bishop, C. (2006). *Pattern recognition and machine learning*. New York: Springer-Verlag.
- Bucharest Stock Exchange. (n.d.). Retrieved from <http://www.bvb.ro/>
- Burda, A., & Pancerz, K. (2014). Clustering and visualization of bankruptcy patterns using the self-organizing maps. *Barometr Regionalny, Analizy i Prognozy*, *13*(3), 133-138.
- Dellepiane, U., Marcantonio, M. D., Laghi, E., & Renzi, S. (2015). Bankruptcy prediction using support vector machines and feature selection during the recent financial crisis. *International Journal of Economics and Finance*, *7*(8), 182-195.
- Drugowitsch, J. (2017). *Variational Bayesian inference for linear and logistic regression*. Retrieved from <https://arxiv.org/pdf/1310.5438v3.pdf>
- du Jardin, P. (2009). *Bankruptcy prediction models: How to choose the most relevant variables?* (MPRA Paper No. 44380). Munich Personal RePEc Archive.
- du Jardin, P. (2017). Dynamics of firm financial evolution and bankruptcy prediction. *Expert Systems With Applications*, *75*, 25-43. <https://doi.org/10.1016/j.eswa.2017.01.016>
- Erdogan, B. E., & Akyüz, S. Ö. (2018). A weighted ensemble learning by SVM for longitudinal data: Turkish Bank bankruptcy. In M. Tez & D. von Rosen D. (Eds.), *Trends and perspectives in linear statistical inference, contributions to statistics*. Cham: Springer.
- Geng, R., Bose, I., & Chen, X. (2015). Prediction of financial distress: An empirical study of listed Chinese companies using data mining. *European Journal of Operational Research*, *241*(1), 236-247. <https://doi.org/10.1016/j.ejor.2014.08.016>
- Gepp, A., & Kumar, K. (2015). Predicting financial distress: A Comparison of survival analysis and decision tree techniques. *Procedia Computer Science*, *54*, 396-404. <https://doi.org/10.1016/j.procs.2015.06.046>
- GitHub. (n.d.). *MATLAB code by Jan Drugowitsch*. Retrieved from https://github.com/DrugowitschLab/vb_logit
- Jaakkola, T. S., & Jordan, M. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing*, *10*, 25-37. <https://doi.org/10.1023/A:1008932416310>
- Koklu, M., & Tutuncu, K. (2014, July 17-18). *Qualitative bankruptcy prediction rules using artificial intelligence techniques*. Paper presented at International Conference on challenges in IT, Engineering and Technology (ICCIET2014), Phuket, Thailand.
- Kostopoulos, G., Karlos, S., Kotsiantis, S., & Tampakas, V. (2017). Evaluating active learning methods for bankruptcy prediction. In C. Frasson, G. Kostopoulos. (Eds.), *Brain Function Assessment in Learning. BFAL 2017. Lecture Notes in Computer Science*, (vol. 10512, pp. 57-66). Cham: Springer. https://doi.org/10.1007/978-3-319-67615-9_5
- Lopez-Iturriaga, F., & Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks. *Expert Systems with Applications*, *42*, 2857-2869. <https://doi.org/10.1016/j.eswa.2014.11.025>

- National Trade Register Office. (n.d.). Retrieved from <https://www.onrc.ro/index.php/ro/>
- Ohlson, J. A. (1980). Financial ratios and probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109-131. <https://doi.org/10.2307/2490395>
- Pena, T., Martinez, S., & Abudu, B. (2009). *Bankruptcy prediction: A comparison of some statistical and machine learning techniques*. (Banco de Mexico Working Papers, 2009-18). Banco de Mexico.
- Pereira, J., M., Basto, M., & da Silva, A. F. (2016). The logistic lasso and ridge regression in predicting corporate failure. *Procedia Economics and Finance* 39, 634-641. [https://doi.org/10.1016/S2212-5671\(16\)30310-0](https://doi.org/10.1016/S2212-5671(16)30310-0)
- Pervan, I., Pervan, M., & Vukoja, B. (2011). Prediction of company bankruptcy using statistical techniques – case of Croatia. *Croatian Operational Research Review (CRORR)*, 2, 158-167.
- Prodan-Palade, D. (2017). Bankruptcy risk prediction models based on artificial neural networks. *Audit Financiar*, XV(3), 418-429. <https://doi.org/10.20869/AUDITF/2017/147/418>
- Safavian, S., R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(3), 660-674. <https://doi.org/10.1109/21.97458>
- Stádník, B., Raudeliūnienė, J., & Davidavičienė, V. (2016). Fourier analysis for stock price forecasting: assumption and evidence. *Journal of Business Economics and Management*, 17(3), 365-380. <https://doi.org/10.3846/16111699.2016.1184180>
- Succuro, M. (2017). Financial bankruptcy across European Countries. *International Journal of Economics and Finance*, 9(7), 132-146. <https://doi.org/10.5539/ijef.v9n7p132>
- Sun, J., Li, H., Huang, Q., & He, K. (2014). Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. *Knowledge-Based Systems*, 57, 41-56. <https://doi.org/10.1016/j.knosys.2013.12.006>
- Tian, S., Yu, Y., & Zhou, M. (2015). Data sample selection issues for bankruptcy prediction. *Risk, Hazards & Crisis in Public Policy*, 6(1), 91-116. <https://doi.org/10.1002/rhc3.12071>
- Vapnik, V. N. (1998). *Statistical Learning Theory*. John Wiley & Sons.
- Wagle, M., Yang, Z., & Benslimane, Y. (2017, May 7-9). Bankruptcy prediction using data mining techniques. In *8th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES)*, Chonburi, Thailand (pp. 1-4). IEEE. <https://doi.org/10.1109/ICTEmSys.2017.7958771>
- Wang, N. (2017). Bankruptcy prediction using machine learning. *Journal of Mathematical Finance*, 7, 908-918. <https://doi.org/10.4236/jmf.2017.74049>