Expert Systems with Applications 39 (2012) 10140–10152

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Enhanced default risk models with SVM+

Bernardete Ribeiro [a,*], Catarina Silva [a,b], Ning Chen [c], Armando Vieira [c], João Carvalho das Neves [d]

[a] CISUC, Department of Informatics Engineering, University of Coimbra, Portugal
[b] ESTG, School of Technology and Management, Polytechinc Institute of Leiria, Portugal
[c] GECAD, Polytechnic Institute of Porto, Porto, Portugal
[d] ISEG-School of Economics and Management, Technical University of Lisbon, Portugal

ARTICLE INFO

Keywords:
Bankruptcy prediction
Default risk model
Support vector machines
Multi-task learning

ABSTRACT

Default risk models have lately raised a great interest due to the recent world economic crisis. In spite of many advanced techniques that have extensively been proposed, no comprehensive method incorporating a holistic perspective has hitherto been considered. Thus, the existing models for bankruptcy prediction lack the whole coverage of contextual knowledge which may prevent the decision makers such as investors and financial analysts to take the right decisions. Recently, SVM+ provides a formal way to incorporate additional information (not only training data) onto the learning models improving generalization. In financial settings examples of such non-financial (though relevant) information are marketing reports, competitors landscape, economic environment, customers screening, industry trends, etc. By exploiting additional information able to improve classical inductive learning we propose a prediction model where data is naturally separated into several structured groups clustered by the size and annual turnover of the firms. Experimental results in the setting of a heterogeneous data set of French companies demonstrated that the proposed default risk model showed better predictability performance than the baseline SVM and multi-task learning with SVM.

1. Introduction

The world is observing one of the most severe financial crisis ever observed. While in the past the SME (Small and Medium Enterprises) and micro companies had higher propensity of bankruptcies in the recent past an increasing number of large bankruptcies is systematically announced and the financial distress of all type of firms across all industries is in danger. Aided by technology and lower barriers to international capital flows, these crisis have shown a greater tendency to spread to markets through out the world, severely affecting the global economic activity. At the heart of the present global recession is an inappropriate evaluation of credit risk and most of governments were forced to implement rescue plans for the banking systems, including the Portuguese Government.

Given the devastating effects of the financial distress of firms, it is urgent that management and regulators are able to anticipate this kind of issues. Although credit loss is a normal cost of doing business in the case of banks, the excess of losses can be sufficiently severe to threaten their own existence. The evidence in the present situation is the need for banks to revise their models which evaluate the risk of each loan and the default rates of portfolio's loans. International rating agencies, like Moodys and Standard & Poor's, are also criticized for their models and inadequacy of quantifying and predicting the risk of insolvency of firms and banks. Moreover, these local and international rating agencies tend to analyze the risk of large companies while the financial system and banks, in particular, also need models for analyzing the risk of SME's. Banks have their own internal rating models to quantify the risk of loans but they are still in their infancy and rely on relatively simple mathematical methods with inadequate assumptions.

As a consequence, there is an ever-increasing need for fast automated recognition systems for bankruptcy prediction. The extensive recent literature shows that at the core of the business failure problem is the asymmetric information between banks and firms. Additionally, the development of analytical tools to determine which financial information is more relevant to predict financial distress has gained popularity along with the design of early warning systems that predict bankruptcy (Pena, Martinez, & Abudu, 2009).

The health of firm in a highly competitive business environment is dependent upon its capability to yield profitability and financial solvency. This means that a firm becomes unhealthy when it loses its competence to maintain profitability and financial solvency (Wu, 2010). Business failure is not only common with new start-ups but also with listed companies, and it can easily happen to firms of any and all sizes.

* Corresponding author. Tel.: +351 239 790 000; fax: +351 239 701 266.
E-mail addresses: bribeiro@dei.uc.pt (B. Ribeiro), catarina@dei.uc.pt (C. Silva), ningchen74@gmail.com (N. Chen), asv@isep.ipp.pt (A. Vieira), jcn@iseg.utl.pt (J. Carvalho das Neves).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.02.142

In Portugal, according to the Annual Survey of Insolvency and Constitutions Company Coface,[1] 4519 companies were declared insolvent in 2011, 1192 more than in the previous year (for the same period) representing an increase by 35.8%, while the number of start-ups raised only by 13.2%.

However, for this study we used a large database of French companies. This database is very detailed containing information on a wide set of financial ratios spawning over a period of several years. It contains up to three thousands distressed companies and about sixty thousand healthy ones. The financial Coface Data set (French credit risk provider) is strongly heterogenous with regards to the type of companies and their financial statuses. A great deal of research has been pursued disregarding this aspect. In this paper we focus on the improvement of financial distress decision-making by including structured information into heterogeneous groups of companies. We investigate the use of an advanced SVM+ approach by Vapnik and Vashist (2009) whose rationale is to take into account additional information in a financial setting of French companies. The firms are grouped by their category of large, medium and small sizes thus clustered by the number of employees and annual turnover. The properties resulting from these well-defined profiles unveil decisive correlations among firms. Our study shows that (a) SVM+ not only yields better prediction model than baseline SVM but also a better model as compared to a similar approach of Multi-Task Learning (MTL) and (b) the most salient data parameters per group both in the kernel decision space and in kernel correcting space are optimized, whereby the parameters and parameter ranges that shape the various firms profiles are exposed. The classification results demonstrate the prediction capability and robustness of the proposed method.

The rest of the paper is organized as follows. Next section describes relevant background knowledge on bankruptcy prediction and related work. In Section 3 we introduce SVM+ algorithm and give its mathematical foundations. In Section 4 the proposed approach is described, and further details on comparable settings are schematically illustrated for model comparison. In Section 5 we describe the database with information on healthy (and distressed) firms appropriately labeled for bankruptcy prediction model design in a case study of the French Market. We also describe the preprocessing preparation phase and the performance metrics. The experimental results including model and parameter selection, discussion and statistical hypothesis tests are performed in Section 6. Finally, in Section 7 we present the conclusions and draw further lines of work.

2. Related work

The prediction of bankruptcy is a well-researched area in finance analysis and attracts much interest to creditors, auditors and bank managers. The accurate prediction and early warning of bankrupt events has critical impact on economics to control the risk associated with wrong decisions, decrease the cost of monitoring solvency, and shorten the time of credit assessment. Bankruptcy prediction solves the important decision-making problem that identifies the potential bankrupt company based on the analysis of historical finance characteristics.

The problem is stated as follows: given a set of parameters (mainly of financial nature) that describe the situation of a company over a given period, predict the probability that the company may become bankrupt during the following year. During the years, this problem has been approached by various methods ranging from statistics to machine learning. A review of the topic of bankruptcy prediction with emphasis on neural networks (NN) is given in Atiya (2001). Also, in Ravi Kumar and Ravi (2007) there is a broad coverage of a wide range of other intelligent techniques such as fuzzy set theory (FS), decision trees (DT), rough sets, case-based reasoning (CBR), and support vector machines (SVM). More recently, in Verikas, Kalsyte, Bacauskiene, and Gelzinis (2010) a comprehensive review of hybrid and ensemble-based soft computing techniques applied to bankruptcy prediction is presented. Despite the numerous papers dealing with the problem, it is often difficult to compare the techniques due to possible differences in assumptions, data sets, time periods and failure definitions.

Fig. 1 illustrates the general framework of bankruptcy prediction using machine learning methods, composed of feature selection, dimensionality reduction using linear (e.g., PCA/KPCA) or nonlinear (e.g., ISOMAP/NMF) projection methods, followed by a machine learning (through NN, SVM etc.) process. It may be noticed that the additional information is helpful to attain better generalization of default risk models, which can be, for example in SVM+, the structured information in the data.

2.1. Neural networks

Neural Networks (NNs) are particularly suited for predicting the bankrupt probability, thus they are a strategic choice among other methods. Likewise, their properties make them often used in financial applications because of their excellent performances of treating non-linear data with self-learning capability (Fu-yuan, 2008). As a competitive learning neural network, self-organizing map (SOM) is used to determine the credit class through a visual exploration (Merkevicius, Garsva, & Simutis, 2004). Learning Vector Quantization (LVQ) is a supervised variant of SOM useful for non-linear separation problems (Kohonen, 2001). The network is composed of two levels, in which the input level is fully connected with the output level. The modeling technique is based on the neurons representing prototype vectors and the nearest neighbor classification rule. The goal of learning is to determine the weights that best represent the classes. LVQ has been employed to detect the distressed companies with satisfactory performance as in Chen and Vieira (2009) and Boyacioglu, Kara, and Baykan (2009). Carvalho das Neves and Vieira (2006) show that an enhanced version of Hidden Layer Learning Vector Quantization can enhance the performance of a multi-layer perceptron (MLP). In recent research efforts, combined techniques have been studied to optimize the learning models by evolutionary algorithms, in particular, Genetic algorithm (GA) is used by Sai, Zhong, and Qu (2007) and Huang, Kuo, and Yeh (2008) to optimize the parameters and connected weights of back-propagation neural networks.

2.2. Support vector machines

Support Vector Machines (SVMs) transform the input vectors nonlinearly into a high-dimensional feature space through a kernel function so that the data can be separated by linear models. In the literature there is an endless list of articles with SVM approaches. Min and Lee (2005) applied a grid-search technique to find out the optimal parameter settings of both polynomial and RBF kernel functions and showed that SVM outperforms techniques such as multiple discriminant analysis (MDA), logistic regression analysis (Logit), and three-layer fully connected back-propagation neural networks (BPNs).

More often, evolutionary algorithms including genetic algorithm, annealing simulation, particle swarm optimization, ant colony optimization are widely used in hybrid classification to significantly advance both the SVM single prediction model and feature selection (Lin, Shiue, Chen, & Cheng, 2009). Research efforts have been directed to combine SVM with other soft computing

[1] http://www.cofaceportugal.pt/CofacePortal/PT/pt_PT/pages/home/noticias/Estudos.

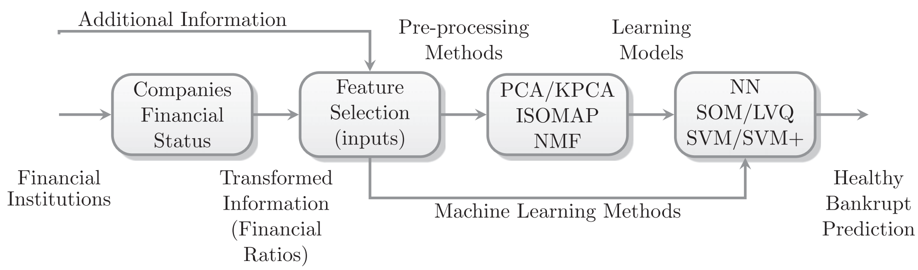


Fig. 1. Bankruptcy machine learning methods pipeline.

tools for better performance. For example, rough set theory (RST) serves as a useful preprocessor of SVM to reduce the redundant variables (Yeh, Chi, & Hsu, 2010). The introduction of fuzzy membership into SVM is capable to find optimal feature subset and parameters, and hence increase the accuracy of SVM prediction (Chaudhuri & De, 2011).

2.3. Feature selection and projection models

Most of the prediction models use financial ratios as predictor variables, by employing the selection of only a few financial ratios according to a choice based criterion. Model selection of corporate distress prediction is advisable for reducing problem complexity and saving computational costs. Rekba Pai, Annapoorani, and Vijayalakshmi Pai (2004) test a linear pre-processing stage using principal component analysis (PCA) for dimensionality reduction purposes. However, nonlinear projection methods (e.g. ISOMAP) have been successfully used by Ribeiro et al. (2009b) making them more suitable for this problem. Recently, Lin, Yeh, and Lee (2011) also uses a hybrid manifold learning approach model (ISOMAP-SVM) and shows the approach is better capable to predict the failure of firms. With the same goal, non-negative matrix factorization (NMF) is performed by Ribeiro, Silva, Vieira, and Carvalho das Neves (2009a) for extracting the most discriminative features with promising results.

2.4. Probability models

While the forecast of bankruptcy is of paramount importance to all stakeholders, to estimate the probability of a corporate failure can prevent the adverse effects that such event can provoke. Ribeiro, Vieira, and Carvalho das Neves (2006) apply a probabilistic framework for bankruptcy detection based on a Relevance Vector Machine (RVM). It is shown therein that the classifier yields a decision function leading to significant reduction in the computational complexity with regards to SVM while the prediction accuracy compares favorably. Pena et al. (2009) use a Gaussian Process to estimate bankruptcy probabilities.

2.5. Linear models

Although machine learning models have been widely used in the last decades, still the pioneer statistical techniques are worth mentioning in the modeling of corporate bankruptcy prediction such as univariate and multivariate discriminant analysis (Altman, 1968, 1993). Classification algorithms like linear discriminant analysis (LDA) and logistic regression (LR) are also popular linear approaches. All these techniques aim at finding an optimal linear combination of explanatory input variables, such as, e.g., solvency and liquidity ratios, in order to analyze, model and predict corporate default risk. Unfortunately the financial ratios violate the assumptions of (i) linear separability, (ii) multivariate normality and (iii) independence of the predictive variables. Therefore, the models overlook the complex nature, boundaries and interrelationships of the financial ratios.

In this paper we look at a new learning paradigm SVM+ invented by Vapnik and Vashist (2009) and Vapnik, Vashist, and Pavlovitch (2009) propose a financial distress prediction model where privileged information is taken into account by extending our earlier work (Ribeiro, Silva, Vieira, Cunha, & Carvalho das Neves, 2010). The additional information regarding heterogenous financial ratios grouped by the type of firms according to the number of employees and annual turnover or global balance is incorporated in the model. In this regard the approach takes a holistic view of the overall process enhancing the learning inductive process by improving generalization.

3. Learning models with large margin classifiers

The rationale behind large margin classifiers intuitively lies on how the classifier with the largest margin will give low expected risk, and hence better generalization. Among many different approaches to the classification of data, support vector machines use the concept of margin, a confidence parameter rather than a raw training error, which allows the design of better algorithms (Smola et al., 2000).

Vapnik and Vashist (2009) and Vapnik et al. (2009) discuss in detail the extension of the new formulation of the SVM algorithm presented formerly in Vapnik (1982, 2006) and demonstrate its superiority toward other machine learning techniques. The new paradigm SVM+ while upholding the main principles of SVM extends its concept, by incorporating the essence of 'untold' information often not handled in a learning problem. In the scope of many practical problems, it is showed that in terms of the capability of transmitting privileged or hidden information, the role of a supervisor (or even of an oracle) leverages the classification (or regression) tasks. It is a new step in machine learning paradigms which had never been put before. Liang and Cherkassky (2008) present a learning paradigm for multi-task learning (MTL) with several models to deal with heterogeneous data and lateral information. The SVM+ paradigm is then adapted to the multi-task learning framework (SVM + MTL). The authors compare in several papers (Cai & Cherkassky, 2009; Liang & Cherkassky, 2007, 2008) SVM+ and MTL approaches demonstrating their similarities and differences. In particular, it is therein emphasized that Vapnik's "Learning with structured data" formulation is similar to the multi-task learning in the sense that both of them try to exploit the group information. However, they point out that in the former only one model is set up while in MTL t models are considered. Another difference lies on group membership of new testing data which is not required in SVM+ whereas under MTL test inputs are assumed to have group labels (Cai & Cherkassky, 2009).

3.1. Support vector machines

Support Vector Machines (SVMs) are maximum margin classifiers with low capacity and good generalization. The SVM trains a classifier by finding an optimal separating hyperplane which maximizes the margin between two classes of data in the kernel induced feature space.

Suppose we are given l instances of training data. Each instance consists of a $(\mathbf{x}_i, y_i)$ pair where $\mathbf{x}_i \in \mathbb{R}^N$ is a vector containing N attributes of the instance i, and $y_i \in \{+1, -1\}$ is the correspondent class label. The method uses input–output training pairs from $\mathcal{D} = \{(\mathbf{x}_i, y_i) \in \mathcal{X} \subseteq \mathbb{R}^N \times \mathcal{Y} : 1 \leqslant i \leqslant l\}$ such that the SVM classifies correctly unobserved data $(\mathbf{x}, y)$.

Each $\mathbf{x}$ in $\mathcal{X}$ is then mapped to a $\phi(\mathbf{x})$ in the kernel-induced feature space, which is related to the kernel function K by $\phi(\mathbf{x})^T\phi(\mathbf{x}') = K(\mathbf{x}, \mathbf{x}')$ for any $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$. SVM tries to find the optimal separating hyperplane $\mathbf{w}^T\phi(\mathbf{x}) + b$ that has large margin and small training error.

The quadratic programing optimization problem originally proposed by Cortes and Vapnik (1995) is:

$$\min_{\mathbf{w}, b \in \mathbb{R}, \xi} \Phi(\mathbf{w}, b, \xi) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l} \xi_i \tag{1}$$

subject to constraints

$$\begin{aligned} y_i((\mathbf{w}^T\phi(\mathbf{x})) + b) &\geqslant 1 - \xi_i \quad i = 1, \ldots, l \\ \xi_i &\geqslant 0 \qquad i = 1, \ldots, l. \end{aligned} \tag{2}$$

Here $\xi = [\xi_1, \ldots, \xi_l]^T$ is the vector of slack variables upholding the errors, and C is a user-defined regularization parameter that trades-off the margin error. This problem is solved by maximizing the function:

$$R(\boldsymbol{\alpha}) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{3}$$

with respect to α_i, under the constraints $\sum_{i=1}^{l} \alpha_i y_i = 0$ and $0 \leqslant \alpha_i \leqslant C$, $i = 1, \ldots, l$. The solution is a linear combination of the input data points $\mathbf{x}_i$ for which α_i is different zero (the so-called the support vectors (SVs)) and is given by the decision function with the form:

$$f(\mathbf{x}, \boldsymbol{\alpha}) = \sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \tag{4}$$

with $\alpha_i, b \in \mathbb{R}$. The SVM finds the class of a given test point $\mathbf{x}_j$ by computing $f(\mathbf{x}_j, \alpha)$ and then checks which side of the hyperplane it falls on.

3.2. SVM+

Recently, a generalization of SVM, designated by support vector machine plus (SVM+), was proposed by Vapnik (2006) and coworkers in Vapnik and Vashist (2009). The technique formerly known as Learning With Structured Data (LWSD), or even Learning Using Privileged Information (LUPI) has an SVM-based optimization formulation whose goal is to find the best mapping function f such that the expected loss:

$$R_{LWSD} = \int L(f(\mathbf{x}, \mathbf{w}), y)P(\mathbf{x}, y)d\mathbf{x}dy$$

is minimized. The SVM+ approach is designed to take advantage of the structure in the training data (for example, noise present in data, or invariants in the data). By leveraging this structure, the SVM+ technique can lower the overall system's VC-dimension and hence attain better generalization.

In SVM+ the slack variables are restricted by the correcting functions, and the correcting functions represent additional information about the data. Vapnik uses this concept to control the learning machine by establishing a privileged information setting which leads to a holistic view of the whole process.

In this subsection we closely follow Liang, Cai, and Cherkassky (2009). Suppose that training data are the union of $t > 1$ disjoint groups $\mathcal{D}_r$. If we denote the indice set of samples from: $\mathcal{D}_r$ by T_r then

$$\mathcal{D}_r = \{(\mathbf{x}_{r_i}, y_{r_i}) \in \mathcal{X} \subseteq \mathbb{R}^N \times \mathcal{Y} : i \in T_r\}.$$

To account for the group information, Vapnik (2006) proposed to define the slack variables within each group by the so-called 'correcting function':

$$\xi_i^r = \xi_r(\mathbf{x}_i) = \varphi_r(\mathbf{x}_i, \mathbf{w}_r), \quad i \in T_r, \; r = 1, \ldots, t. \tag{5}$$

To define the correcting function $\xi_i^r = \xi_r(\mathbf{x}_i) = \varphi_r(\mathbf{x}_i, \mathbf{w}_r)$ for group T_r the input training vectors $\mathbf{x}_i$, $i \in T_r$ are mapped into two different Hilbert spaces (i) the decision space and (ii) the correction space. The former is defined by the decision functions and the latter by the correcting functions for a given group r.

A dual-optimization technique similar to the one used before is performed in SVM+. The main idea is to find the hyperplane in the feature space that solves the minimization problem:

$$\min_{\mathbf{w}, b \in \mathbb{R}, \xi^r} \Phi(\mathbf{w}, b, \xi^r) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{\gamma}{2}\|\mathbf{w}_r\|^2 + C\sum_{r=1}^{t}\sum_{i \in T_r} \xi_i^r \tag{6}$$

subject to constraints

$$\begin{aligned} & y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geqslant 1 - \xi_i^r, \quad i \in T_r, \quad r = 1, \ldots, t \\ & \xi_i^r \geqslant 0, \quad i \in T_r, \quad r = 1, \ldots, t \\ & \xi_i^r = \left(\mathbf{w}_r^T\phi_r(\mathbf{x}_i) + d_r\right), \quad i \in T_r, \quad r = 1, \ldots, t. \end{aligned} \tag{7}$$

It may be noticed that the capacity of a set of decision functions is reflected by $\|\mathbf{w}\|$ and the capacity of a set of correcting functions for group r is $\|\mathbf{w}_r\|$.

The decision function has again the form:

$$f(\mathbf{x}, \boldsymbol{\alpha}) = \sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \tag{8}$$

where $\alpha_i \geqslant 0$, $i = 1, \ldots, l$ are values that maximize the function:

$$\begin{aligned} R(\alpha, \beta) = &\sum_{i=1}^{l} \alpha_i - \frac{1}{2}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ &- \frac{1}{2\gamma}\sum_{r=1}^{t}\sum_{i,j \in T_r}(\alpha_i + \beta_i - C)(\alpha_j + \beta_j - C)K^r\left(\mathbf{x}_i^r, \mathbf{x}_j^r\right) \end{aligned} \tag{9}$$

subject to constraints

$$\begin{aligned} & \sum_{i}^{l} \alpha_i y_i = 0 \\ & \sum_{i \in T_r}(\alpha_i + \beta_i - C) = 0, \quad r = 1, \ldots, t \\ & \alpha_i \geqslant 0, \quad \beta_i \geqslant 0, \quad i = 1, \ldots, l. \end{aligned} \tag{10}$$

SVM+ has two kernels which in different spaces define similarity measures between two data points. Thus it allows to control the learning capacity of both the decision functions and the correcting functions. The parameters γ and C have a crucial role over the learning machine that combines regularization and control of the margin, in which γ adjusts the relative weight of these two capacities, and C controls the trade-off between complexity and proportion of sample errors. Fig. 2 illustrates the SVM+ learning procedure and indicates in the most left block diagram the parameters with respect to the four (different) models M_1, M_2, M_3, M_4 featured by the type of kernel in each space.

From a practical point of view the SVM classifier has only one free parameter, the parameter C, in the case of the linear SVM; and two parameters, C and σ, in the case of the RBF kernel. Regarding the SVM+ classifier, the four identified models (M_1, M_2, M_3 and

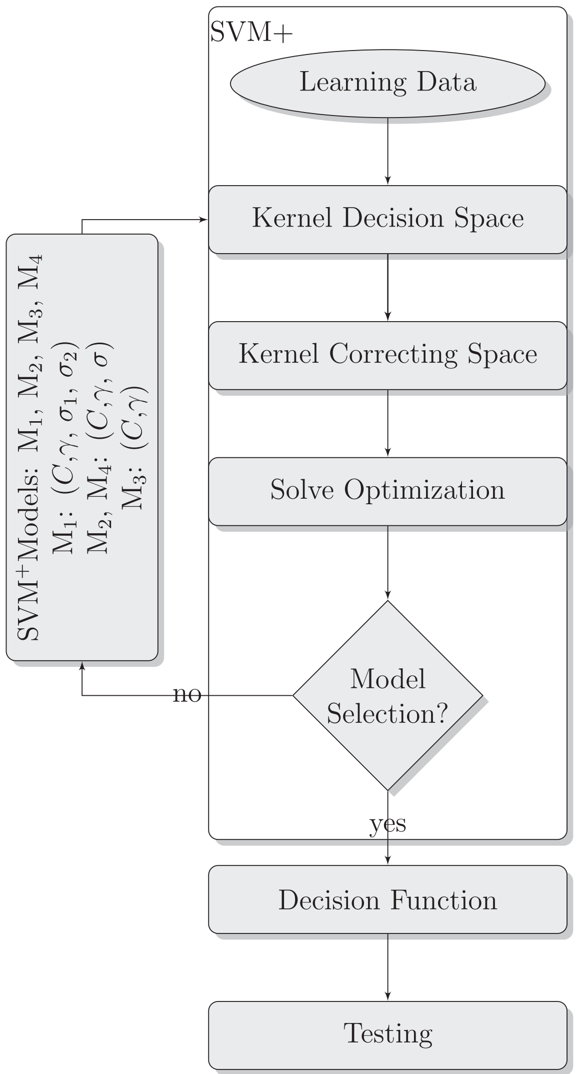


Fig. 2. SVM+ learning diagram model. Models M_1, M_2, M_3 and M_4 are selected according to the kernels parameters for the decision space and the correcting space.

M_4) are characterized by the type of kernel in both spaces as illustrated in Table 1. In the case of M_1 where the RBF kernel is used for both the decision space and the correcting space, the SVM+ classifier needs four parameters (C, γ, σ_1 and σ_2). Meanwhile, in the case of M_3 which requires little parameter specification both kernels are linear. In the case of M_2 (or M_4) where the linear kernel (or RBF kernel) is chosen for the decision space and the RBF kernel (or linear kernel) is used for the correcting space only three parameters are required: C (as in standard linear SVM), γ and σ (RBF kernel width).

3.3. Multi-task learning with SVM (SVM + MTL)

In multi-task learning (MTL) framework (Liang et al., 2009) the setting is similar to the previous learning methodology using

Table 1
Models selection.

	Decision space		Correcting space	
	Kernels type			
M1	–	RBF	–	RBF
M2	Linear	–	–	RBF
M3	Linear	–	Linear	–
M4	–	RBF	Linear	–

SVM+. Likewise, the training data can be naturally separated into several structured groups, whereby one model estimation is realized for each group as a separate task, while in SVM+ the objective is to estimate a single predictive model for all groups. The adaptation of the SVM+ to MTL entails the specification of decision functions for different groups and modeling the interrelationship among the groups. It gives rise to the method SVM + MTL where the input vectors $\mathbf{x}_i \in T_r$ are simultaneously mapped into two different Hilbert spaces: the decision space $\mathbf{z}_i = \phi_z(\mathbf{x}_i) \in Z$ and the correcting space $\mathbf{z}_i^r = \phi_{z_r}(\mathbf{x}_i) \in Z_r$ for a given group r. The goal is to find the t decision functions:

$$f_r(\mathbf{x}) = (\mathbf{w}^T\phi_z(\mathbf{x})) + b + (\mathbf{w}_r^T\phi_{z_r}(\mathbf{x})) + d_r, \quad r = 1, \dots, t \tag{11}$$

where each decision function above involves two spaces (the decision space and the correcting space) contrasting with SVM+ which only defines a function in the decision space. Formally, SVM + MTL solves the following optimization problem:

$$\min_{\mathbf{w}, b \in \mathbb{R}, \xi^r} \Phi(\mathbf{w}, b, \xi^r) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{\gamma}{2}\|\mathbf{w}_r\|^2 + C\sum_{r=1}^{t}\sum_{i \in T_r}\xi_i^r \tag{12}$$

subject to constraints

$$y_i^r\left((\mathbf{w}^T\phi_z(\mathbf{x}_i)) + b + (\mathbf{w}_r^T\phi_{z_r}(\mathbf{x}_i)) + d_r\right) \geqslant 1 - \xi_i^r, \quad i \in T_r,$$
$$r = 1, \dots, t\xi_i^r \geqslant 0, \quad i \in T_r, \quad r = 1, \dots, t \tag{13}$$

Likewise in SVM+ the parameters C and γ control, respectively, the trade-off between the complexity and the proportion of non-separable examples, and the relative weight of the decision function and the correcting function capacity. The slack variables ξ_i^r measure the error that each of the final model makes on the data.

4. Proposed default risk model with SVM+

Fig. 3 presents the overall design of the proposed approach based on SVM+ computational algorithm for an enhanced default risk model. As indicated, the main computational stages include data preparation and re-sampling, model selection task using the SVM, SVM+ and SVM + MTL approaches, and finally the resulting default risk models are statistically compared for significance testing.

Ooghe and De Prijcker (2006) come up with a failure conceptual model of the possible causes of bankrupt. In an earlier study the authors group the causes of bankruptcy into five interactive facets including general environment (economics, technology, foreign countries, politics, and social factors), immediate environment (customers, suppliers, competitors, banks and credit institutions, stockholders, and misadventure), management/entrepreneur characteristics (motivation, qualities, skills, and personal characteristics), corporate policy (strategy and investments, commercial, operational, personnel, finance and administration, corporate governance), and company characteristics (size, maturity, industry, and flexibility).

Herein, in Fig. 4 we hierarchically represent a main two-level informative analysis involving, respectively, the firms key financial ratios and non-financial information. In each of the next four-levels, we include, in the former, financial indicators, such as, operational performance, financial liquidity, risk and return, and sustainable growth. In the latter, non-financial (though relevant) information comprise government policy, economic environment, marketing reports, customers screening, industry trends, competitors landscape, etc. By exploiting additional information able to improve classical inductive learning we propose a prediction model where data is naturally separated into several structured groups clustered by the size and annual turnover of the firms. This is achieved by using SVM+ where beyond control over the separation

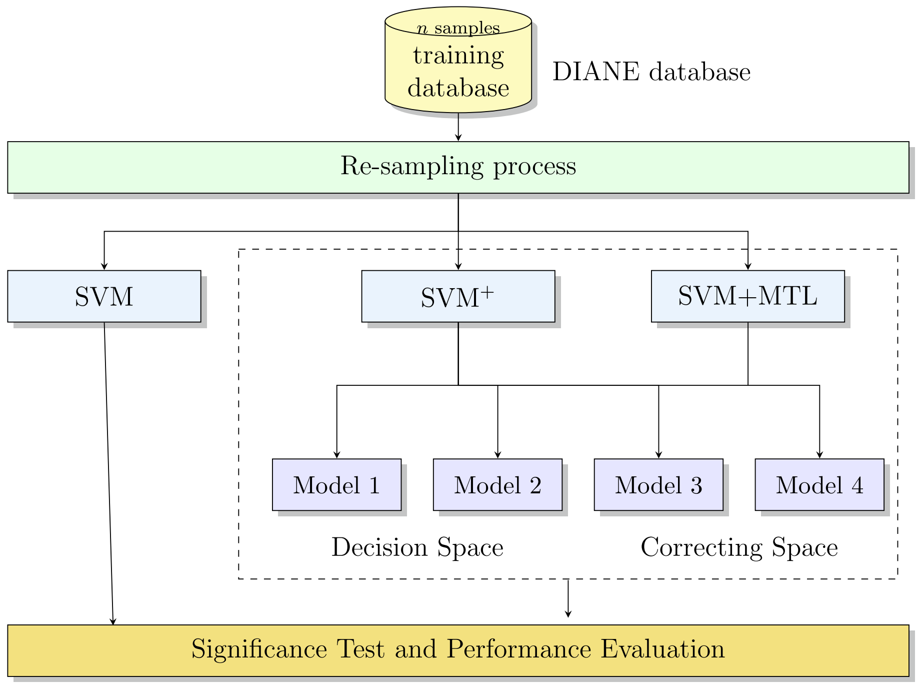


Fig. 3. General model of SVM+ default risk model.

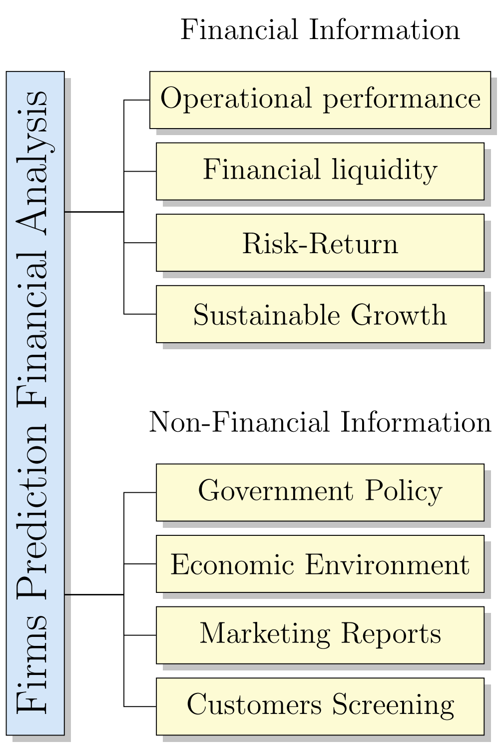


Fig. 4. Key financial ratios and non-financial information.

margin, the algorithm uses appropriate correcting functions which are able to provide supplementary information not accounted for in the Key Performance Indicators (KPI) variables.

5. Experimental setup

Based on the proposed prediction model in this section we describe the data set, present the preprocessing procedure, detail the key performance indicators (KPI), and indicate the performance evaluation metrics for problem assessment.

5.1. Data description

We used the Diane database[2] which contains financial statements of French companies for the years of 2002 to 2006. The database is very complete containing information about the ratio of distressed and healthy firms across all the years. The data set is also very diversified by companies sector such as: construction firms, real estate firms, manufacturing, IT firms, etc. Fig. 5 illustrates the composition of the main industries of the French Market, distributed among several sectors, included in the database.

The initial sample consisted of financial ratios of about 60 000 industrial French companies (for the years of 2002 to 2006) with at least 10 employees. From these companies, about 3000 were declared bankrupted in 2007 or presented a restructuring plan to the court for approval by the creditors. Fig. 6(a) depicts the distribution of distressed firms in 2006 and 2007 under various conditions, namely, large and small bankruptcies, friendly liquidation and legal regulation.

Regarding their size, it varies according the number of employees: >=10 and <50 (Small), >=50 and <500 (Medium), and >=500 (Large) and includes structured and heterogeneous information with respect to the company financial statuses. In particular, the capital turnover which measures the firm's efficiency in using the capital employed to generate revenue is also taken into account. Specifically, for the years of 2007 and 2006, in Fig. 6(b) we schematically represent the number of distressed companies comprising above clustered information. In this paper, we worked with 30 financial ratios extracted from the database which are detailed in Table 2. These financial indicators allow to describe the firms in terms of financial strength, liquidity, solvability, productivity of labor and capital, margins, net profitability and return on investment. With linear statistical models (some of) these variables have

[2] COFACE financial French risk provider.

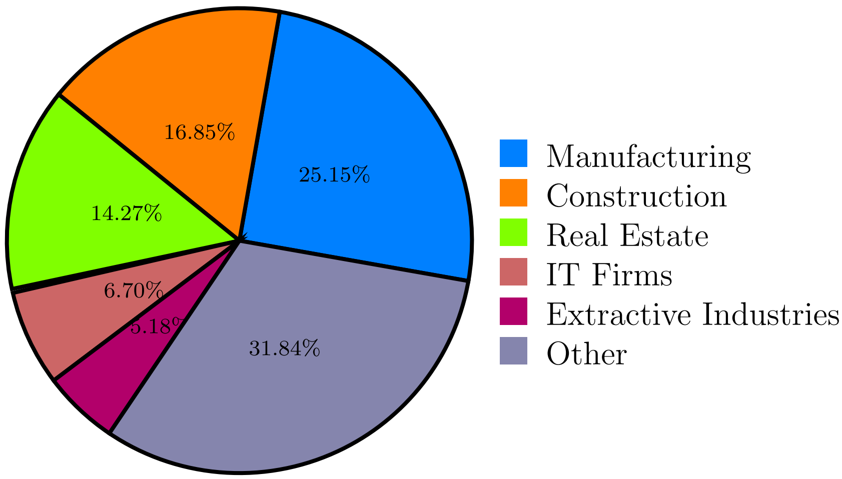
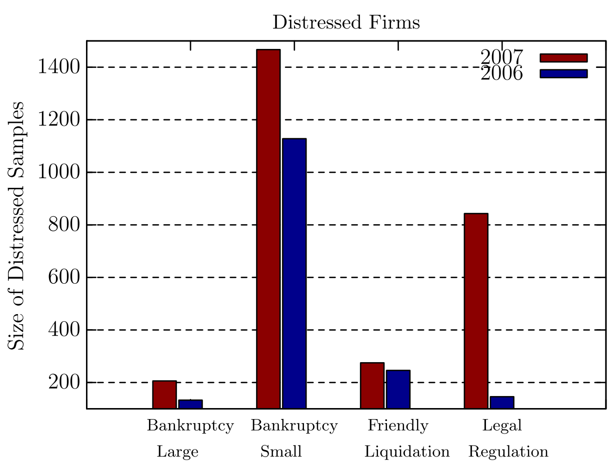
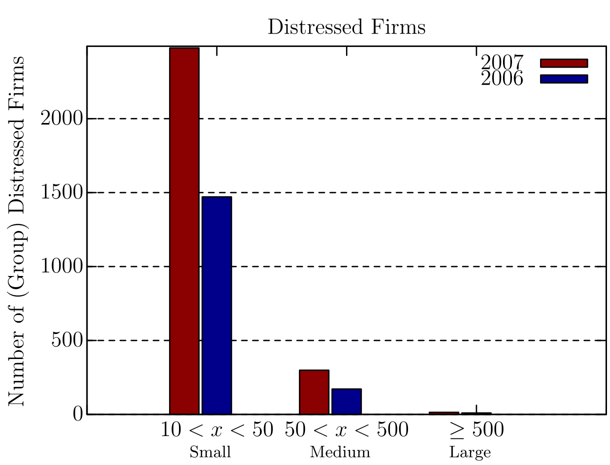


Fig. 5. Industry sectors in DIANE.



(a) Distribution of Distressed Firms w.r.t. French Market in Figure 5



(b) Heterogeneous Clustered Distressed Firms

Fig. 6. Diane firms type samples.

small discriminatory capabilities for default prediction, thus the rationale with non-linear approaches is to improve the classification accuracy without compromising generalization. The ultimate goal is to predict the class (Distressed, Healthy) by extracting relevant information contained in the financial ratios. In order to attain better predictability group information should be included in the model. This goal is achieved by adjusting a good (model) fitting of the (heterogeneous) data.

Table 2
DIANE database financial ratios.

Variable description	
x_1 – Number of Employees last year	x_{16} – Cashflow/Turnover
x_2 – Capital Employed/Fixed Assets	x_{17} – Working Capital/Turnover days
x_3 – Financial Debt/Capital Employed	x_{18} – Net Current Assets/Turnover days
x_4 – Depreciation of Tangible Assets	x_{19} – Working Capital Needs/Turnover
x_5 – Working Capital/Current Assets	x_{20} – Export
x_6 – Current Ratio	x_{21} – Added Value per Employee EUR
x_7 – Liquidity Ratio	x_{22} – Total Assets Turnover
x_8 – Stock Turnover days	x_{23} – Operating Profit Margin
x_9 – Collection Period days	x_{24} – Net Profit Margin
x_{10} – Credit Period days	x_{25} – Added Value Margin
x_{11} – Turnover per Employee EUR	x_{26} – Part of Employees
x_{12} – Interest/Turnover	x_{27} – Return on Capital Employed
x_{13} – Debt Period days	x_{28} – Return on Total Assets
x_{14} – Financial Debt/Equity	x_{29} – EBIT Margin
x_{15} – Financial Debt/Cashflow	x_{30} – EBITDA Margin

5.2. Data preprocessing

In order to obtain a balanced dataset we randomly selected 600 non-default examples resulting in a set of 1200 samples of default and non-default companies. An appropriate treatment of the database to eliminate firms with missing values was pursued.

The following strategy is pursued from original financial database to appropriately set up a prediction model.

- A set of 600 default companies are selected with at most 10 missing data;
- A set of 600 non-default companies is sampled randomly to obtain a balanced data set;
- The missing values are replaced by the value of the closest available year;
- The ratios are preprocessed by logarithmized operation to decrease the scatter of data distribution:

$$y = \begin{cases} log(x+1) & \text{if } x > 0 \\ -log(1-x) & \text{otherwise} \end{cases} \tag{14}$$

- The features are then normalized for the purpose of equal influence on classification. We use the linear normalization which transforms the maximum value to 1 and the minimum value to 0:

$$y = \frac{x - min(x)}{max(x) - min(x)} \tag{15}$$

- The companies are grouped by their category of large, medium and small regarding their size according to the number of employees and annual turnover or global balance.

5.3. Key performance indicators analysis

Fig. 7 illustrates the mean values of the financial key performance indicators (KPI) in terms of the financial ratios presented in Table 2 with respect to default firms and non-default firms, respectively. It is interesting to notice that a close look into the KPI mean values along all the indicators might heighten our understanding regarding the firms' behavior. For instance, if we consider the five important ratios such as x_6 current ratio, x_{14} Financial Debt/Equity, x_{16} Cashflow/Turnover, x_{20} Export, and x_{29} EBIT (Earns Before Interest Rates) Margin it may be observed an opposite behavior between Default and Non-Default firms. Indeed, according to Altman (1968, 1993) these KPI indicators seem to have a prominent role into the bankruptcy prediction problem.

5.4. Evaluation metrics

The performance metrics were evaluated based on the classification contingency matrix defined in Table 3. Here *tp*, *fp*, *tn*, *fn* represent the usual notation for the confusion matrix in terms of true (or false) and positive (or negative) results from the classifier.

In this classification problem, two types of misclassification carry different weights. This is due to the fact that if a potentially distressed ('bad') company is classified as financially healthy ('good') then the amount of loss incurred by a stakeholder is entirely different from the other type of misclassification. In our next definitions we consider that the positive class is the distressed firm whereas the negative class is the healthy firm. Then the two possible types of misclassification errors are defined as follows. A 'Type I

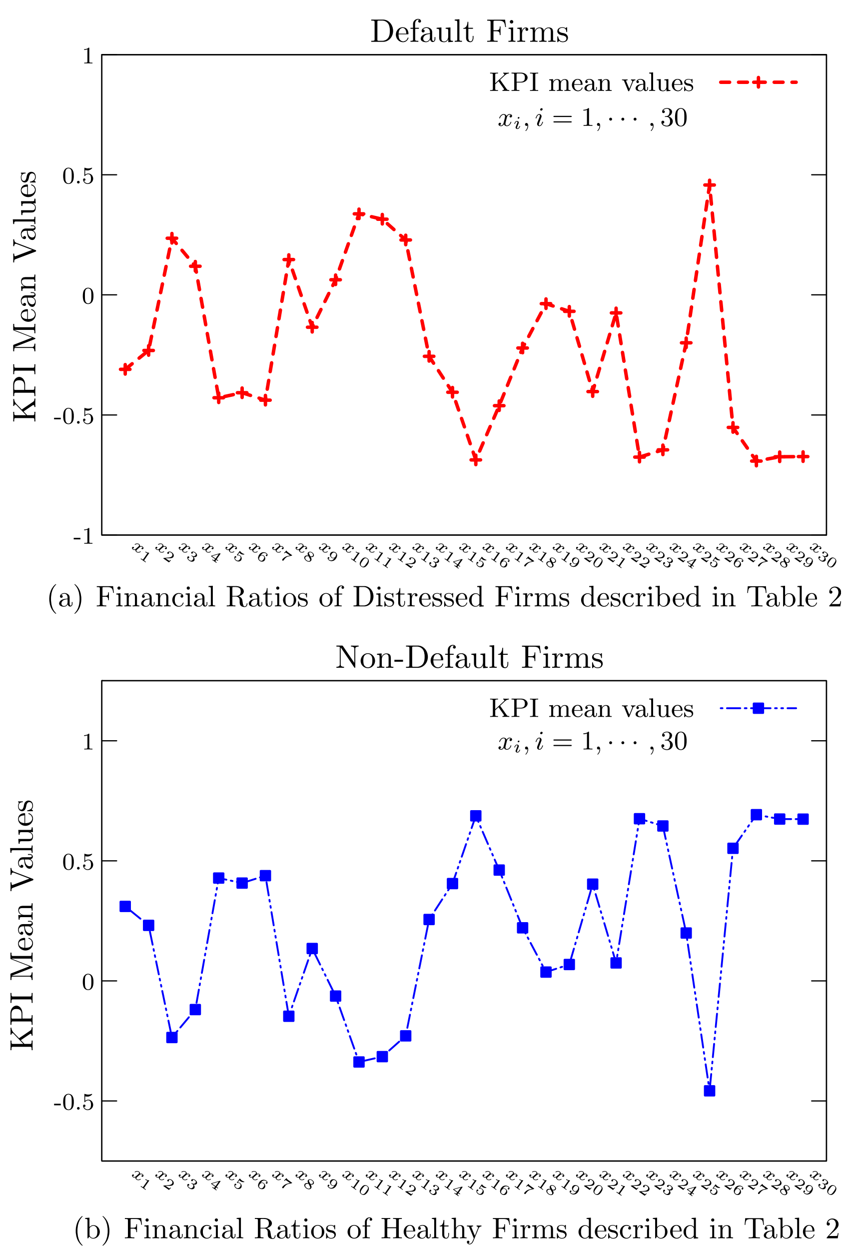


Fig. 7. Diane KPI mean values.

Table 3
Contingency matrix.

Real class	Predicted class		
	Bankrupt	Healthy	Total
Bankrupt	*tp*	*fn*	pos
Healthy	*fp*	*tn*	neg

error' (or false positive rate (*fpr*), i.e. $\frac{fp}{fp+tn}$) indicates the misclassification of a healthy firm as distressed. Conversely, a 'Type II error' (or false negative rate (*fnr*), i.e. $\frac{fn}{fn+tp}$) is the one in which a distressed firm is misclassified by the predictor as viable. This error is very important since the predictor should not make a mistake preventing the decision maker to take a wrong decision. According to Ooghe and Spaenjers (2006) the former corresponds to a 'commercial risk' meanwhile the latter corresponds to a 'credit risk'. An "overall hit" refers to the total correct classifications for the set $\left(\frac{tp+tn}{tp+fp+fn+tn}\right)$ regardless of type. We also illustrate the results with F1-score which quantifies the trade-off between Recall $\left(\frac{tp}{tp+fn}\right)$ and Precision $\left(\frac{tp}{tp+fp}\right)$ and is fairly indicative of the performance of the overall algorithm. F1-score is defined as $2\frac{Precision*Recall}{Precision+Recall}$, which reaches the best value at 1 and worst score at 0. All the results represent mean values obtained in test financial data.

Another performance metric is ROC (Receiver Operating Characteristic) curve (Fawcett, 2006) which is obtained by plotting *tpr* versus *fpr*. The curve depicts the tradeoffs between *tp* and *fp*. ROC provides an easy and natural way to compare the performance of different classifiers independently from the cost context and the class distribution, able to observe if one classifier dominates another, and therefore identify the optimal model.

6. Results and discussion

In this section we present and discuss the results. Further comparison with SVM baseline and multi-task learning approach SVM + MTL demonstrate the easiness of the proposed model. The experimental design follows two sets of experiments, namely, model selection and parameters selection. Next, statistical significance tests are performed for models assessment. Further discussion is given for the bankruptcy prediction model.

The entire data set is divided randomly into five folds for cross-validation, in which 4 folds are used for model training, and the remaining is used for testing the generalization capability of the built model. In each run the SVM, SVM+ and SVM + MTL are applied to the learning dataset. For validation, each sample of the test data set is input to the resultant model and the class is predicted. After the experiment is repeated 5 times, the confusion matrix is calculated by comparing the real class and predicted class for the entire data. Then the evaluation criteria are obtained from the confusion matrix.

6.1. Model selection

In a first set of experiments we performed the runs according to the design indicated in Fig. 3. As said above we used four models (M_1, M_2, M_3 and M_4) according to the type of kernel chosen for the decision and correcting spaces. We decided to run the experiments with two types of kernels: the RBF kernel, since it was shown to present the best behavior in previous empirical results run in the same data set (Ribeiro et al., 2009b, 2009a), and the linear kernel.

The results in Fig. 8 with respect to the F1-score clearly indicate that the best approach is SVM+ and the best models are M_1 and M_4

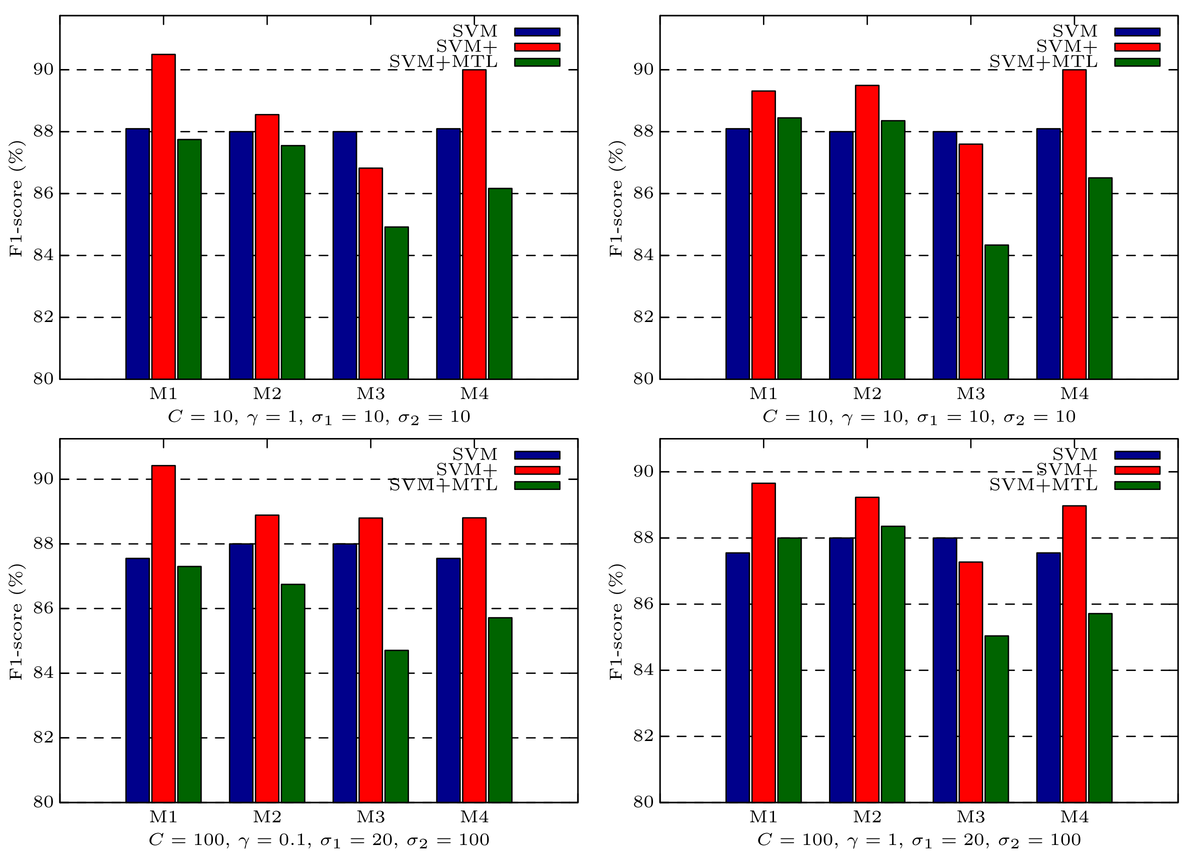


Fig. 8. F1-score versus model selection.

for runs where the parameters were set as illustrated in the plots. It is interesting to notice that both models uphold an RBF kernel for the decision space while the correcting space kernel is either an RBF (M_1) or linear (M_4). As observed in Section 3.2 the slack variables in Eq. (5) are restricted by the correcting functions which contain useful information about the data. In particular, the characteristics of each firm in terms of its size and annual turnover might convey additional capability to the learning machine facilitating the discriminative task between failing and non-failing cases.

In Fig. 8 the parameters at the bottom of each plotted graph correspond to the complete range of parameters used when the selected model is M_1. For instance in the top left graph, the whole list of parameters is ($C = 10, \gamma = 1, \sigma_1 = 10, \sigma_2 = 10$). As indicated in Fig. 2, the configuration of parameters for the other models M_1, M_2 and M_3 is trivially obtained.

A more extensive analysis of the effect of the variation of model parameters in the performance measures will be presented later.

In Table 4 several performance measures are presented, namely, the overall hit of firms financial status, the two types of misclassification errors and the F1-score measure which is a trade-off of precision and recall and represents fairly well the quality of the classifier. We indicate the mean values (and standard deviations) in percentage while the best results are highlighted in bold. We observe that F1-score for the SVM+ w.r.t. to the SVM + MTL approach improved by 1.55% (M_1), 0.43% (M_2), 2.81% (M_3) and 2.57% (M_4) while w.r.t. the baseline SVM the highest increase in F1-score was observed in M_1 (1.72%).

We observe in Table 4 that the significant measures F1-score (89.35 ± 1.48), predictability accuracy (89.40 ± 1.71) and Error Type II (12.59 ± 0.74) (in bold) have the highest values w.r.t. the same measures in both counterpart approaches, i.e., the SVM baseline and the SVM + MTL. As expected, the errors of type I and type II in the SVM baseline do not change for models M_2 and M_3 since the correcting functions are not taken into account in this approach.

Since the misclassification cost on the 'bankrupt' class is higher (in this study as said above corresponds to type II error) the classifier achieving less error type II is preferred in practice.

In Fig. 9 we illustrate in (a) Type I and Type II errors bar plots for the SVM+ which confirm that the best model is M_1 whereby the expensive cost presents the lowest value. In (b) the increase of errors of Type II in the SVM and SVM + MTL approaches w.r.t. SVM+ is depicted. It may be observed that regardless of kernel functions, SVM+ consistently performs better than SVM + MTL. It may be observed that the highest increase in the misclassification error corresponds to models M_3 and M_4 w.r.t. the SVM + MTL approach. Interestingly the kernel is linear for the correcting space, therefore it indicates that in multi-task learning the correcting functions with RBF kernel seem to have a prominent influence in the learning model regarding the 'credit risk'.

6.2. Parameters selection

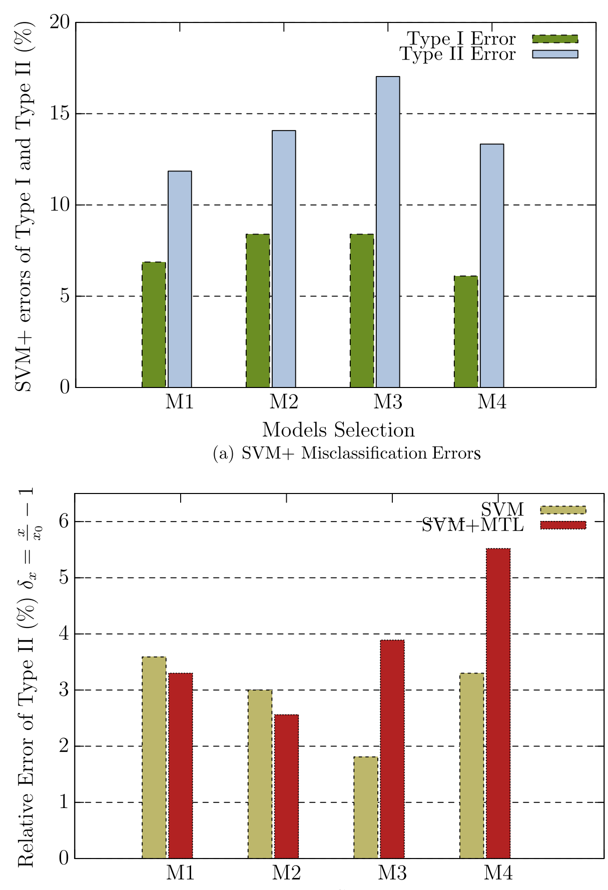
In the second run of experiments, we study how the model parameters affect the learning model M_1 and compare its variation in all of the tested approaches. The strategy was to analyze how the parameters C and γ whose combined action is in control of the margin and regularization influence the measures behavior.

By keeping fixed both the parameters of the kernel decision space σ_1 and of the kernel correcting space σ_2, and parameter γ, in Fig. 10(a) we graphically illustrate with varying C, how $F1$-score performs consistently better for SVM+ than the other two approaches, while baseline SVM performs better than SVM + MTL.

In a similar way, the decision (and correcting) space kernel parameters are kept constant as well as the trade-off parameter C. We vary γ in the interval (1 → 1000), whereby the logscale is herein used for better visualization. It may be also noticed in the Figure F1-score is constant for the baseline SVM since its formulation does not include γ. Fig. 10(b) shows the plots of $F1$-score where also better performance of SVM+ is obtained in most cases as compared to the other methods. The parameter γ adjusts the rel-

Table 4
Mean (and standard deviations) of Performance results (a) Model 1 (M_1), (b) Model 2 (M_2), (c) Model 3 (M_3), (d) Model 4 (M_4). Three learning approaches are considered: SVM baseline, SVM+ and SVM + MTL.

Models	Performance measures			
	Accuracy	Type I Error	Type II Error	F1-score
	SVM			
M_1	88.12 ± 0.94	6.41 ± 4.96	17.19 ± 3.07	87.63 ± 0.51
M_2	88.72 ± 0.00	3.82 ± 0.00	18.52 ± 0.00	88.00 ± 0.00
M_3	**88.72 ± 0.00**	**3.82 ±0.00**	18.52 ± 0.00	**88.00 ±0.00**
M_4	88.12 ± 0.94	6.41 ± 4.96	17.19 ± 3.07	87.63 ± 0.51
	SVM+			
M_1	**89.40 ± 1.71**	8.55 ± 3.79	**12.59 ± 0.74**	**89.35 ± 1.48**
M_2	**89.10 ± 0.53**	7.18 ± 2.94	**14.52 ± 2.65**	**88.83 ± 0.59**
M_3	87.82 ± 1.02	8.55 ± 4.37	**15.70 ± 2.64**	87.55 ± 0.75
M_4	**88.87 ± 1.69**	9.31 ± 3.83	**12.89 ± 0.66**	**88.85 ± 1.45**
	SVM + MTL			
M_1	88.27 ± 0.72	**6.41 ± 4.17**	16.89 ± 2.84	87.80 ± 0.45
M_2	89.10 ± 1.41	**3.51 ± 0.68**	18.07 ± 2.32	88.40 ± 1.59
M_3	85.49 ± 0.21	8.24 ± 1.00	20.59 ± 0.97	84.74 ± 0. 27
M_4	86.99 ± 0.63	**6.41 ± 0.68**	19.41 ± 0.62	86.28 ± 0.66



(a) SVM+ Misclassification Errors

(b) Relative Error of Type II Error

Fig. 9. (a) SVM+ Errors of Type I and II versus Models design. (b) Increase in the relative error of Type II errors of SVM and SVM + MTL approaches w.r.t. SVM+.

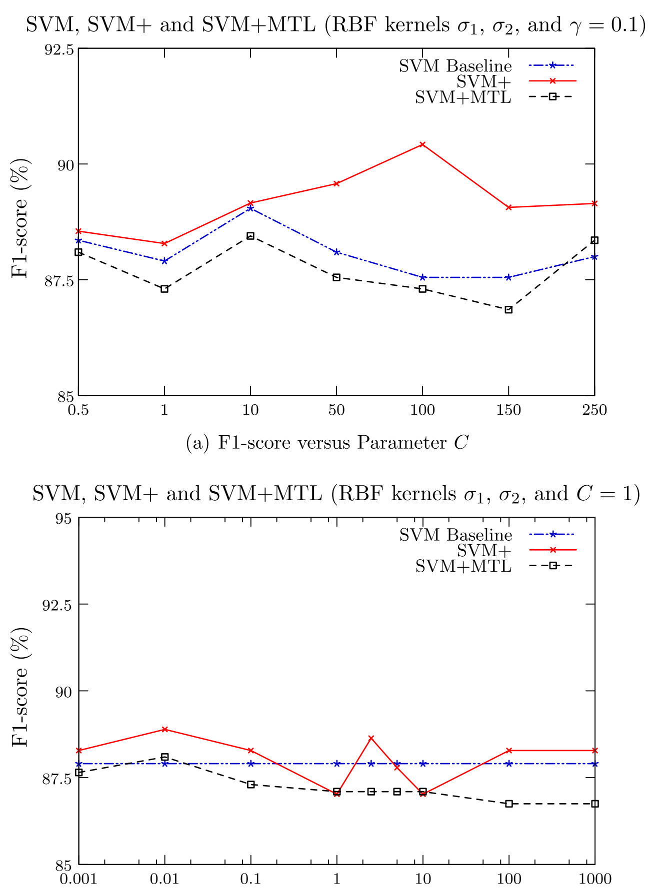
ative weights of the two machine capacities above referred. By analyzing the results it is possible to observe that for $\gamma = 1$ and $\gamma = 10$ the F1-score is identical for the SVM+ and SVM + MTL approaches, which seems to indicate that under these values SVM+ does compare favorably to SVM + MTL, although it is not better.

Considering the intended purpose of SVM+ and, moreover, the nature of the importance of the credit risk, it has been concluded that SVM+ has substantial advantages in comparison with other structures. In Fig. 11 the two types of misclassification errors are plotted. We run the experiments for the M_1 model and changed the parameter C across the indicated range. The results are fairly indicative of the better performance of SVM+ when the default risk cost is a concern. It may be observed that the type II error falls in SVM+ with respect to the tested models. Thus, the likelihood of not missing to detect a distressed company and enhancing the default risk model significantly increases.

An overall view of the classifiers performance is observed in Fig. 12 for the three studied methods. The ROC curves depicted provide for each cut-off value the proportion of observations incorrectly classified as default by the model against the proportion correctly classified as default. The plotted results demonstrate the superiority of the SVM+ over the SVM baseline and SVM + MTL approaches.

6.3. Significance hypothesis statistical tests

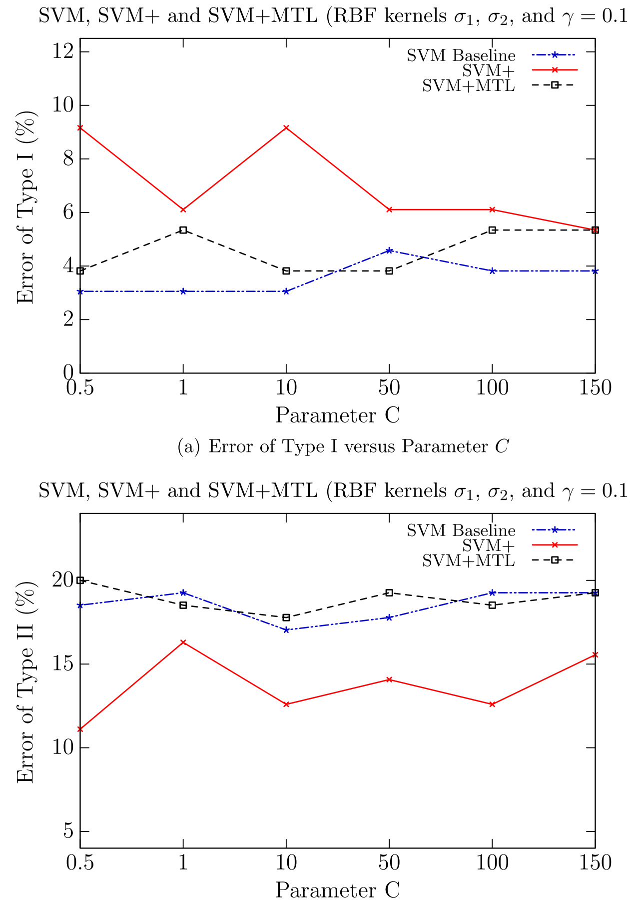
There is a heightened need for implementing effective financial models, therefore two statistical tests were performed, first,



(a) F1-score versus Parameter C

(b) F1-score versus Parameter γ

Fig. 10. Results of parameters variation for model M_1 in SVM baseline, SVM+ and SVM + MTL approaches.



(a) Error of Type I versus Parameter C

(b) Error of Type II versus Parameter C

Fig. 11. Results of misclassification errors for model M_1 in SVM baseline, SVM+ and SVM + MTL approaches.

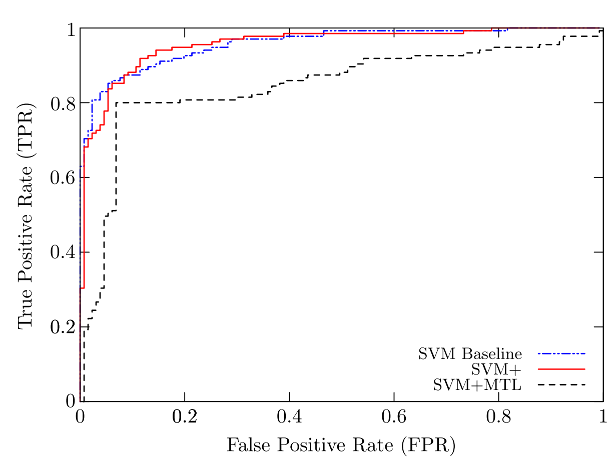


Fig. 12. ROC curves.

Table 5
Significance Tests with statistical variable F1-score: (a) SVM, (b) SVM+, (c) SVM + MTL.

	SVM+	SVM + MTL
SVM	0.022151*	0.102722
SVM+		0.002192**

* Significance at 5% level.
** Significance at 1% level.

Table 6
Significance Tests with statistical variable Error of Type II: (a) SVM, (b) SVM+, (c) SVM + MTL.

	SVM+	SVM + MTL
SVM	0.0412*	0.252237
SVM+		0.024533*

* Significance at 5% level.

regarding the model approach and, second, with respect to the model selection under the chosen model.

A statistical t-test was performed for the three approaches used: SVM, SVM+ and SVM + MTL. The statistical variables chosen were the F1-score measure and error of type II which is the expensive cost in the model design.

6.3.1. Performance evaluation

Table 5 summarizes the statistical significance of the difference between three methods by means of t-test. The significance level is set as 5%, so that the p-value less than 5% indicates that the two underlying methods are significantly different in the mean. As it may be observed, the method SVM+ significantly outperforms the SVM + MTL method in terms of F1-score, as shown by the p-value 0.002192 indicating the t-test is highly significant at significance levels of 1%. The p-value of 0.02151 shows that SVM+ is significantly different in the mean at the level of 5% w.r.t. the SVM baseline.

The significance tests above are verified by the results in Table 6 whereby the statistical variable is the expensive cost error of type II. This cost identifies the risk of missing a positive case, that is, to detect successfully a bankrupt company. As shown, the p-value of 0.0412 indicates that at a significance level of 5% the results are statistically significant for the SVM and SVM+, meanwhile, the p-value of 0.024533 shows that the difference between the means of SVM+ and SVM + MTL is statistically significant too.

6.3.2. Model selection

We observe in Table 7 that Models M_1 and M_4 are statistically significant at a significance level of 1% as compared with M_3. It is interesting to notice that this occurs when the kernel in the decision space is RBF. Moreover, we observe that the better model (M_1) corresponds also to the RBF kernel for the correcting space.

7. Conclusion and future work

In response to the recent growth of the credit industry and to the world economic crisis early planning for declaring bankruptcy is of great importance to various stakeholders. The health of a firm in a highly competitive business environment is dependent upon its capability of achieving profitability and financial solvency. Therefore, the unhealthy status come to play when the firm loses profitability and financial solvency. As one of the most developed techniques in bankruptcy prediction, SVMs have shown excellent generalization performance due to many attractive features. In this study we investigate the predictive capability of SVM+, which handles privileged information in the model. Business failure can easily happen to firms of any and all sizes. We use 30 financial ratios as inputs to the corporate failure prediction model using structured and heterogeneous information grouped by the category of large, medium, or small size according to the number of employees and annual turnover of companies. By taking a holistic perspective it was possible to incorporate firms additional (and useful) information into the model. As a consequence of this approach, different optimized parameters both in the kernel decision space and kernel correcting space are selected, resulting in better overall-predictability performance. As expected, the SVM+ model yields improvement of F1-score performance measure while decreasing type II error which quantifies the cost of missing a company in a bad status. On one hand, the comparison with the baseline SVM and SVM + MTL shows that regardless of the kernel functions employed, SVM+ always produces the best performance. It suggests that by leveraging the structured information in the training data, the model can attain better generalization and robustness prediction than other approaches. On the other hand, comparable results on different models of SVM+ show that the model using RBF kernel function is significantly better than that of linear kernel function in decision space, which gives some insight to select optimal models in practical applications.

Future work will extend this study. It would be interesting when faced with the prospect of losing solvency to take a holistic perspective based on the differentiated activity of firms. In particular our following approach will also consider the different branches of production connected to the industrial sectors. Although we have compared SVM+ with the baseline SVM and the multi-task learning approach SVM + MTL in the present work, extensive comparisons will be conducted using more state-of-the-art classification methods and other financial data set in the future study.

Table 7
Significance Tests with statistical variable Error of Type II: (a) Model 1 (M_1), (b) Model 2 (M_2), (c) Model 3 (M_3), (d) Model 4 (M_4).

	(b)	(c)	(d)
(a)	0.278571	0.003437**	0.495025
(b)		0.313561	0.287190
(c)			0.007897**

** Significance at 1% level.

References

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

Altman, E. I. (1993). *Corporate financial distress and bankruptcy: A complete guide to predicting and avoiding distress and profiting from Bankruptcy* (2nd ed.). John Wiley & Sons.

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

Boyacioglu, M. A., Kara, Y., & Baykan, O. K. (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. *Expert Systems with Applications, 36*, 3355–3366.

Cai, F., Cherkassky, V. (2009). SVM+ regression and multi-task learning. In *International joint conference on neural networks. Atlanta, Georgia* (pp. 418–424).

Carvalho das Neves, J., & Vieira, A. (2006). Improving bankruptcy prediction with hidden layer learning vector quantization. *European Accounting Review, 15*(2), 253–271.

Chaudhuri, A., & De, K. (2011). Fuzzy support vector machine for bankruptcy prediction. *Applied Soft Computing, 11*(2), 2472–2486.

Chen, N., Vieira, A. (2009). Bankruptcy prediction based on independent component analysis. In *First international conference on agents and artificial intelligence (ICAART09)* (pp. 150–155).

Cortes, C., Vapnik, V. (1995). Support-vector networks. In *Machine Learning* (pp. 273–297).

Fawcett, T. E. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861–874.

Fu-yuan, H. (2008). A genetic fuzzy neural network for bankruptcy prediction in chinese corporations. In *International conference on risk management and engineering management (ICRMEM '08)* (pp. 542–546).

Huang, K. C., Kuo, Y., Yeh, I. C. (2008). A novel fitness function in genetic algorithm to optimize neural networks for imbalanced data sets. In *Eighth international conference on intelligent systems design and applications* (pp. 647–650).

Kohonen, T. (2001). *Self-organizing maps* (3rd ed.). Springer-Verlag.

Liang, L., Cherkassky, V. (2007). Learning using structured data: Application to fmri data analysis. In *International joint conference on neural networks, Orlando, FL* (pp. 495–499).

Liang, L., Cherkassky, V. (2008). Connection between SVM+ and multi-task learning. In *International joint conference on neural networks. Hong-Kong* (pp. 2048–2054).

Liang, L., Cai, F., & Cherkassky, V. (2009). Predictive learning with structured (grouped) data. *Neural Networks*, 766–773.

Lin, S., Shiue, Y. R., Chen, S. C., & Cheng, H. (2009). Applying enhanced data mining approaches in predicting bank performance: A case of taiwanese commercial banks. *Expert Systems with Applications, 36*, 11543–11551.

Lin, F., Yeh, C., & Lee, M. (2011). The use of hybrid manifold learning and support vector machines in the prediction of business failure. *Knowledge-Based Systems, 24*(1), 95–101.

Merkevicius, E., Garsva, G., & Simutis, R. (2004). Forecasting of credit classes with the self-organizing maps. *Informaciens Technologijos Ir Valsymas, 4*(33), 61–66.

Min, J. H., & Lee, Y. C. (2005). Bankruptcy prediction using support vector machine (SVM) with optimal choice of kernel function parameters. *Expert Systems with Applications, 28*, 603–614.

Ooghe, H., De Prijcker, S. (2006). Failure processes and causes of company bankruptcy: A typology. Vlerick Leuven Gent Management School Working Paper Series 2006–21, Vlerick Leuven Gent Management School.

Ooghe, H., Spaenjers, C. (2006). A note on performance measures for failure prediction models. Vlerick Leuven Gent Management School Working Paper Series 2006–29, Vlerick Leuven Gent Management School.

Pena, T., Martinez, S., Abudu, B. (2009). Bankruptcy prediction: A comparison of some statistical and machine learning techniques. Technical Report, TR-N 2009-18.

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.

Rekba Pai, G. A., Annapoorani, R., Vijayalakshmi Pai, G. A. (2004). Performance analysis of a statistical and an evolutionary neural network based classifier for the prediction of industrial bankruptcy. In *IEEE Conference on cybernetics and intelligent systems* (pp. 1033–1038).

Ribeiro, B., Vieira, A., Carvalho das Neves, J. (2006). Sparse Bayesian models: Bankruptcy-predictors of choice? In *IEEE International joint conference on neural networks* (pp. 3377–3381).

Ribeiro, B., Silva, C., Vieira, A., Cunha, A., Carvalho das Neves, J. (2010). Financial distress model prediction using SVM+. In *IEEE International joint conference on neural networks* (pp. 1–7).

Ribeiro, B., Silva, C., Vieira, A., & Carvalho das Neves, J. (2009a). Extracting discriminative features using non-negative matrix factorization in financial distress data. In M. Kolehmainen et al. (Eds.), *International conference on adaptive and natural computing algorithms. Lecture notes in computer science (LNCS)* (Vol. 4432, pp. 537–547). Springer.

Ribeiro, B., Vieira, A., Duarte, J., Silva, C., Carvalho das Neves, J., Liu, Q., et al. (2009b). Learning manifolds for bankruptcy analysis. In M. Köppen et al. (Eds.), *International conference on neural information processing. Lecture Notes in Computer Science (LNCS)* (Vol. 5506, pp. 722–729). Springer.

Sai, Y., Zhong, C., Qu, L. (2007). A hybrid GA-BP model for bankruptcy prediction. In *Eighth international symposium on autonomous decentralized systems (ISADS '07)* (pp. 473–477).

Smola, A. J., Bartlett, P., & Schölkopf, B. (Eds.). (2000). *Advances in large margin classifiers*. The MIT Press.

Vapnik, V. (1982). *Estimation of dependences based on empirical data*. New York: Springer-Verlag.

Vapnik, V. (2006). *Empirical inference science afterword of 2006*. Springer.

Vapnik, V., Vashist, A., Pavlovitch, N. (2009). Learning using hidden information (learning with teacher). In *International joint conference on neural networks* (pp. 3188–3195).

Vapnik, V., & Vashist, A. (2009). A new learning paradigm: Learning using privileged information. *Neural Networks, 22*(5–6), 544–557. URL http://dx.doi.org/10.1016/j.neunet.2009.06.042.

Verikas, A., Kalsyte, Z., Bacauskiene, M., & Gelzinis, A. (2010). Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: A survey. *Soft Computing – A Fusion of Foundations, Methodologies and Applications, 14*(9), 995–1010.

Wu, W. W. (2010). Beyond business failure prediction. *Expert Systems with Applications, 37*, 2371–2376.

Yeh, C., Chi, D., & Hsu, M. (2010). A hybrid approach of DEA, rough set and support vector machines for business failure prediction. *Expert Systems with Applications, 37*(2), 1535–1541.