# Financial Distress Prediction of Iranian Companies Using Data Mining Techniques

Mahdi Moradi[1], Mahdi Salehi[1*], Mohammad Ebrahim Ghorgani[2], Hadi Sadoghi Yazdi[1]

[1]Ferdowsi University of Mashhad, Iran
[2]East Oil and Gas Company, NIOC, Iran

Decision-making problems in the area of financial status evaluation are considered very important. Making incorrect decisions in firms is very likely to cause financial crises and distress. Predicting financial distress of factories and manufacturing companies is the desire of managers and investors, auditors, financial analysts, governmental officials, employees. Therefore, the current study aims to predict financial distress of Iranian Companies. The current study applies support vector data description (SVDD) to the financial distress prediction problem in an attempt to suggest a new model with better explanatory power and stability. To serve this purpose, we use a grid-search technique using 3-fold cross-validation to find out the optimal parameter values of kernel function of SVDD. To evaluate the prediction accuracy of SVDD, we compare its performance with fuzzy c-means (FCM). The experiment results show that SVDD outperforms the other method in years before financial distress occurrence. The data used in this research were obtained from Iran Stock Market and Accounting Research Database. According to the data between 2000 and 2009, 70 pairs of companies listed in Tehran Stock Exchange are selected as initial data set.

Keywords: financial distress prediction; Support vector data description; Fuzzy c-mean.

## 1 Introduction

The empirical literature of financial distress prediction has gained considerable attention in the post 2007-2009 global financial crises. Policymakers (Dodd-Frank Act of 2010) and regulators (SEC, Basel III) emphasize about failure of many banks in the aftermath of the global financial crisis and are seeking the best way to predict business failures. Prior studies have addressed two major research trends in financial distress prediction. One is investigating the situation of failure to find the symptoms (Dambolena & Khoury, 1980; Gombola & Ketz, 1983; Jo & Han, 1997; Scott, 1981). The other is comparing the prediction accuracy of the diverse classification methods (Tam & Kiang, 1992; Jo & Han, 1997). This study belongs to the second group of research. The primary purpose of this study is to apply support vector data description (SVDD) to the financial distress prediction problem in an attempt to suggest a new model with better explanatory power and stability. We use a grid-search technique using 3-fold cross-validation to examine the optimal parameter values of kernel function of SVDD. In addition, to evaluate the prediction accuracy of SVDD, we compare its performance with fuzzy c-means

(FCM). Using the data from Iran Stock Market and Accounting Research Database for 70 couples of companies listed in Tehran Stock Exchange during 2000 and 2009; we find that SVDD outperforms the other method.

## 2 Literature review

The empirical literature of financial distress prediction has recently gained further momentum and attention from financial institutions. Academicians and practitioners realize that the problem of asymmetric information between banks and firms lies at the heart of important market crashes such as credit rationing and that improvement in monitoring techniques represents a valuable alternative to any incomplete contractual arrangement aimed at reducing the borrowers' moral hazard (Becchetti & Sierra, 2003; Stiglitz & Weiss, 1981; Xu, 2000). Among financial distress forecasting methods, discriminant analysis was the dominant method for predicting corporate failure from 1966 until the early part of the 1980s (Altman, 1968, 1983; Back et al., 1996b). It gained wide popularity due to its ease of use and interpretation. However,

both linear and quadratic discriminant analyses are sensitive to deviation from multivariate normality (Karels & Prakash, 1987; Laitinen & Laitinen, 2000). During the 1980s, the probit (Zmijewski, 1984) and, especially, the logit methods (logistic regression model) (Back et al., 1996b; Ohlson, 1980) used the discriminant method.

These two models do give a crisp relationship between explanatory and response variables of the given data from a statistical viewpoint and do not assume multivariate normality, but the probit model assumed that the cumulative probability distribution must be standardized normal distribution, while the logit model assumed that the cumulative probability distribution must be logistic distribution. Since the 1990s, neural networks have been the most widely used techniques in developing quantitative financial distress prediction (Back et al., 1996b; Tam & Kiang, 1992; Wilson & Sharda, 1994), in particular, the approximation or classification powers of the MLP trained by the backpropagation algorithm (Hassoun, 1995; Hertz, Krogh, & Palmer, 1991). Many studies compared the neural networks backpropagation algorithm with the statistical methods and found neural networks backpropagation outperforms the other statistic methods, such as multivariate discriminant analysis (MDA), probit and logit methods (Back et al., 1996a; Shin, Lee & Kim, 2005; Wilson & Sharda, 1994). Neural networks have recently been employed to extract rules for solving fuzzy classification problems (Kim et al., 2003). A number of fields use the radial basis function network (i.e., RBFN), for classification problems (Jang, Sun, & Mizutani, 1997; Surendra & Krishnamurthy, 1997), function approximations (Chuang, Jeng, & Lin, 2004; Hertz et al., 1991; Jang, 1993; Jang et al., 1997; Nam & Thanh, 2003) and management sciences (Stam, Sun, & Haines, 1996; Vythoulkas & Koutsopoulos, 2003).

The approximation or classification powers of the MLP trained by the back propagation algorithm (Hassoun, 1995; Hertz et al., 1991) and RBFN are determined by the number of hidden nodes. In fact, the number of hidden layers influences the performance of back propagation MLP. Additionally, an RBFN is functionally equivalent to a zero-order Sugeno fuzzy inference system under some conditions (Jang et al., 1997). In addition, it was proven that the zero-order Sugeno fuzzy inference system could approximate any nonlinear function on a compact set to an arbitrary degree of accuracy under certain conditions (Jang, 1993). However, if a phenomenon under consideration does not have stochastic variability but is also uncertain in some sense, it is more natural to seek a fuzzy functional relationship for the given data, which may be either fuzzy or crisp.

Sun and Li (2008) use weighted majority voting combination of multiple classifiers for FDP, Chen and Du (2009) introduced an integration strategy with subject weight based on neural network for financial distress prediction. They all generated diverse classifiers by applying different learning algorithms (with heterogeneous model representations) to a single data set, and concluded that, to some degree, FDP based on combination of multiple classifiers was superior to single classifiers according to accuracy rate or stability. The most used machine learning technique is the neural network model (Haykin, 1999), trained by the back-propagation learning

algorithm (Wong et al., 1997; Wong and Selvi, 1998) whose prediction accuracy outperforms statistical models including logistic regression (LR), linear discriminant analysis (LDA), multiple discriminant analysis (MDA) and other machine learning models, such as k-nearest neighbor (k-NN) and decision trees. In addition, the back-propagation neural network (BPN) model can be used as the benchmark for financial decision support models. Chen and Du (2009) found that prediction performance for the clustering approach is more aggressively influenced than the BPN model and the BPN approach obtains better prediction accuracy than the data mining (DM) clustering approach in developing a financial distress prediction model. classifiers which were diversified by using neural networks on different data sets for financial distress prediction, and their experimental results showed that multiple neural network classifiers did not outperform a single best neural network classifier, based on which they considered that the proposed multiple classifiers system may be not suitable for the binary classification problem as financial distress prediction. Song et al, (2010) presented genetic algorithm (GA) based approach and statistical filter approaches are applied to identify the best features for the support vector machine (SVM). The proposed GA-based approach is carefully designed in order to have the capability of simultaneously optimizing the features and parameters of the SVM. Experimental results on the data from Chinese companies showed that the GA-based approach could extract fewer features with a higher accuracy compared with statistical filter approaches.

Recent studies in Artificial Intelligence (AI) approach, such as ANN (Ravisankar & Ravi 2010), SVM (Lin et al. 2011; Min & Lee 2005; Bao et al., 2012) have also been successfully applied to financial distress prediction.

The purpose of this paper is to apply fuzzy clustering means and support vector data description (SVDD) in financial distress prediction model. Fuzzy c-means (FCM) clustering is one of well-known unsupervised clustering techniques, which allows one piece of data were two or more clusters. SVDD is known as the algorithm that finds a special kind of linear model with the maximum margin hyperplane. The maximum margin hyperplane gives the maximum separation between decision classes. The training examples that are closest to the maximum margin hyperplane are called support vectors. The SVDD classifier will be trained by different kernel functions in order to compare it with the benchmark of the neural network model. In SVDD, Using different kernel functions and the determination of optimal values of the parameters to train SVMs will lead to different results.

Therefore, the current study aims to compare the accuracy of these two forecasting techniques in predicting financial distress of companies. The original classification accuracy indicates that SVDD outperforms the FCM model.

# 3 Technical background

## 3.1 Support vector data description

SVDD, inspired by the idea of support vector machine by Vapnik (1995), is a method of one-class classification, for

which not the optimal separating hyperplane but the sphere with minimal volume containing all or most objects has to be found; its sketch in two-dimensional spaces is shown in Figure 1. It is often used as a method of novelty detection.

Novelty detection based on boundary essentially is to find a sphere with minimum volume containing all (or most of) the normal data objects. For a data set containing $N$ normal data objects, when one or a few very remote objects are in it, a very large sphere is obtained, which will not represent the data very well. Therefore, some data points outside the sphere are allowed with introducing slack variable $\xi i$ , then, the sphere can be described by centre $a$ and radius $R$ as follows

$$\min L(R) = R^2 + C \sum_{t=1}^{N} \xi t$$
$$s.t. (x_t - a)^T (x_t - a) \le R^2 + \xi t \tag{1}$$
$$\xi t \ge 0 \quad (i = 1, 2, \dots N)$$

Where the variable $C$ gives the trade-off between simplicity (or volume of the sphere) and the number of errors (number of target objects rejected). We construct the Lagrange

$$L(R, a, \alpha t, \xi t)$$
$$= R^2 + C \sum_t \xi t - \sum_t \alpha_t [R^2 + \xi^2 - (x_t^2 - 2ax_t + a^2 \tag{2}$$
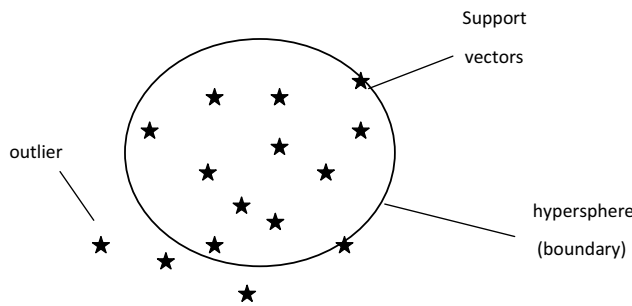$$- \sum_t \gamma_t \xi t \quad (\alpha_t \ge 0, \gamma \ge 0)$$



*Figure.1 The sketch of SVDD in two-dimensional space*

Setting the partial derivatives to 0, new constraints are obtained

$$\sum_t \alpha t = 1, \qquad a = \frac{\sum_t \alpha_t x_t}{\sum_t \alpha_t} = \sum_t \alpha_t x_t \tag{3}$$
$$c - \alpha t - \gamma t = 0 \quad \forall i$$

Then, new optimal equation can be obtained

$$\max L = \sum_{t=1}^{N} \alpha_t (x_t, x_t) - \sum_{t=1}^{N} \sum_{f=1}^{N} \alpha_t \alpha_f (x_t, x_f) \tag{4}$$
$$s.t. \ 0 \le \alpha_t \le C_t \qquad \sum_{t=1}^{N} \alpha t = 1$$

From the above process, it is observed that the centre of the sphere is a linear combination of data objects with weight factors $\alpha_i$. Only for a small set of objects the equality of the second equation in equation (1) is satisfied, which are the objects on the boundary of the sphere. This means that SVDD holds sparseness, which determines its excellent computation.

The above method computes a sphere around the data in the original input space. Normally, data are not spherically distributed, and we cannot expect to obtain a very tight description. Therefore, kernel function $K(x_i, x_j)$ is introduced to replace inner products $(x_i \cdot x_j)$, which implicitly maps the objects $xi$ into some feature space. A better, tighter description can be obtained when a suitable feature space is chosen. Different kernel functions result in different description boundaries in the original input space. Gaussian kernel is the most commonly used function, its expression is as follows:

$$k_G(x_t, x_f) = \exp \frac{-(x_t - x_f)^2}{\sigma^2} \tag{5}$$

where $\sigma$ is the width parameter, also called extension constant. This function can suppress the growing distances for large feature spaces.

## 3.2 Fuzzy c-means

FCM theory is the most perfect one among many fuzzy clustering analysis methods that are effective for pattern recognition; details can be seen in reference. Considering a sample set $X = \{x_1, x_2, \dots, x_N\}$, $xi$, $R$s, which is required to be divided into $C$ categories; the aim of FCM is to obtain each category's clustering centre $vc = \{vc_1, vc_2, \dots, vc_S\}$ $(1 \_ c \_ C)$ by minimizing the weighed square sum of inner-cluster error.

Therefore its objective function is as follows

$$J_m(U, V) = \sum_{c=1}^{C} \sum_{n=1}^{N} (\mu_{cn})^m (d_{cn})^2, \quad m \in [1, \infty) \tag{6}$$

With constraints

$$s.t. \begin{cases} 0 \le \mu_{cn} \le 1, & 1 \le c \le C, 1 \le n \le N \\ 0 \langle \sum_{n=1}^{N} \mu_{cn} \langle N, & 1 \le c \le C, \\ \sum_{c=1}^{C} \mu_{cn} = 1, & 1 \le n \le N \end{cases} \tag{7}$$

where $m$ is the smoothing parameter, which makes it effective from hard c-means to FCM. This parameter controls the sharing degree among each fuzzy categories, bigger $m$ will result in more fuzzy division, or results in more definitive division. Its experimental range is [1.1, 5]; $\mu cn$ is subjection of $xn$ to the $c$th category; $dcn$ represents the distance between $xn$ and $vc$, which often is measured in Euclidean space

$$(d_{cn})^2 = \|x_n - v_c\| = (x_n - v_c)^T (x_n - v_c) \tag{8}$$

**U** and **V** can be optimized by performing a number of iterative computations using following equations (9) to (11), whose astringency has been proved

$$\mu_{\ln} = \begin{cases} \dfrac{1}{\sum_{c=1}^{C} (d_{tn}/d_{cn})^{2/(m-1)}} & I_n = \Phi \\ 0, & \forall c \in \overline{I_n} \\ 1, & \forall c \in In \ne \Phi \end{cases} \tag{9}$$

where

$$I_n = \{ c \mid 1 \le c \le C, d_{cn} = 0 \} \tag{10}$$
$$\overline{I_n} = \{ 1, 2, ..., C \} - I_n$$

$$v_c = \frac{1}{\sum_{n=1}^{N} (\mu_{cn})^m} \sum_{n=1}^{N} (\mu_{cn})^{m_{x_n}} \tag{11}$$

# 4 Research method

## 4.1 Data collection and preprocessing

The data used in this study obtained from Tehran Stock Exchange. Based on the background of Iranian listed companies, the criteria whether the listed company is specially treated (ST) by the Teheran Stock Exchange categorizes companies into two classes based on their financial condition: normal or distressed. Distressed companies are referred to ST (specially treated) companies and are classified as such if their accumulated losses are more than 50% of stockholder equity (Iran Business Law Article 141). Companies were chosen so that the Ln total assets are almost equal distressed companies. In this analysis, we use financial data from two years before a company is classified as ST and denote it as year (t-2).

The data used in this study is obtained from the Iran Stock Market during 2000 and 2009 and Accounting Research Database and includes 70 pairs of companies listed on the Teheran Stock Exchange. Firms with missing financial ratios or ratios that are more than 3 Standard Deviations from the mean are excluded. After eliminating companies with missing and outlier data, the final number of sample companies is 120.

## 4.2 Feature Selection

This study uses more variables than other authors, which usually do not use more than 20. The ratios initially selected allow for a very comprehensive financial analysis of the firms including financial strength, liquidity, solvability, productivity of labour and capital, various kinds of margins and profitability and returns. Although, in the context of linear models, some of these variables have small discriminatory capabilities for default prediction, the non-linear approaches used here can extract relevant information contained in these ratios to improve the classification accuracy without compromising generalization. Feature selection is an important issue in financial distress prediction, as in other problems where a large set of attributes is available, since elimination of useless features may enhance the accuracy of detection while reducing the amount of time for processing the data. Due to the lack of an analytical model, the relative importance of the input variables can only be estimated through empirical methods. A complete analysis would require examination of all possibilities, for example, taking two variables at a time to analyze their dependence or correlation, then taking three at a time, etc. This, however, is both infeasible and not error free since the available data may be of poor quality in sampling the full input space. 24 financial ratios covering profitability, activity ability, debt ability and growth ability are selected as the initial features (see Table 1).

# 5 Results and analysis

To investigate the effectiveness of the SVDD approach trained by small data set size in the context of the corporate financial distress classification problem, we utilize a grid-search tech-

*Table 1. Definition of predictor variables*

| Variable | Financial ratios Description | Variable | Financial ratios Description |
|---|---|---|---|
| X1 | Funds provided by operations to Stockholders' equity | X13 | Accumulated earnings to total assets |
| X2 | Funds provided by operations to total liabilities | X14 | Current ratio |
| X3 | Net working capital to total assets | X15 | Interest expenses to total expenses |
| X4 | Total assets turnover | X16 | Debt ratio |
| X5 | Monetary asset to current assets | X17 | Inventory stock turnover |
| X6 | Monetary asset to current liabilities | X18 | Gross income to sales |
| X7 | Earnings before interest and taxes to Interest expenses | X19 | Net income to Stockholders' equity |
| X8 | Net interest expenses to total liabilities | X20 | Net income to sales |
| X9 | Funds provided by operations to Net working capital | X21 | Net working capital to sales |
| X10 | Earnings before interest and taxes to total assets | X22 | Interest expenses to sales |
| X11 | Natural logarithm total assets | X23 | Interest expenses to Net working capital |
| X12 | Inventory stock to current assets | X24 | Market value stockholders' equity to total assets |

nique using 3-fold cross-validation in order to choose optimal values of the upper bound C and the kernel parameter g that are most important in SVDD model selection. Since the aim is to find the most discriminative features, classification accuracy is the key criterion to evaluate the fitness function. For median-sized problems, cross-validation might be the most reliable way to choose the model parameters. Hence, the fitness is defined as the *3*-fold cross-validation accuracy on the training set. In *3*-fold cross-validation, the training set is divided into *3* subsets of equal size. Sequentially, each subset is tested by the classifier trained on the remaining *3*– 1 subsets (i.e. valida-tion sets). Thus, the cross-validation accuracy is the average accuracy across *3*-fold subsets. Note that the cross-validation method can prevent the over-fitting problem. We conducted the experiment with respect to various kernel parameters and the upper bound C, and compared the prediction performance of SVDD with various parameters as the training set size got smaller. We set an appropriate range of parameters as follows: a range for kernel parameter is between 1 and 100 and a range for C is between 0 and 1.

## 5.1 Findings from planning the support vector data description model:

Results from algorithm test based on support vector data description by using data in the financial distress occurrence year (t year) are in Table 2.

*Table 2. Results from algorithm based on support vector data description in the financial distress occurrence year - educational sample*

| sample | Classifying type | year | Correct classifying percent | incorrect classifying percent | sum |
|---|---|---|---|---|---|
| educa-tional | financial distress | t | 97.10 | 2.9 | 100 |

As in seen in the Table No. 2, the model recognize in t-year, 97.10 percent of financial distress firm in educational sample correctly. On the other hand, classification error is 2.90 percent.

*Table 3. Results from algorithm test based on support vector doctor data description in the financial distress year- experimen-tal sample*

| sample | Classifying type | year | Correct classifying percent | incorrect classifying percent | sum |
|---|---|---|---|---|---|
| experi-mental | financial distress | t | 91.90 | 8.10 | 100 |

Table 3 shows, the model in t-year, 91.90 percent of finan-cial distress in experimental sample correctly (classification error is 8.1 percent). Further, results from model testing show that general accuracy for this model in classifying educational sample and experimental sample in financial distress year is 97.10% and 91.90% respectively. Noticeable point about this

model is non-falling (lack of falling) of predicting accuracy in experimental sample relative to educational sample, which could represent relevant general ability of the model. Results from testing basic on support vector data description in each one of t-l year (one year before financial distress) and t-2 year (two years before financial distress) based on entire samples are presented in Table 4.

*Table 4. Results from testing support vector data description algorithm in one and two year before financial distress occur-rence*

| sample | Classifying type | year | Correct classifying percent | incorrect classifying percent | sum |
|---|---|---|---|---|---|
| Total data | financial distress | t-1 | 85 | 15 | 100 |
| Total data | financial distress | t-2 | 78 | 22 | 100 |

As it is seen in Table No. 4 the model classifies financial distress firms in accuracy 85 and 78 percent respectively in t-1 and t-2 year. Noticeable point about this model is intense non-falling in predicting accuracy financial distress which could represent relevant general ability of this model.

## 5.2 Findings from designing fuzzy c-means model:

Designing fuzzy c-means algorithm was performed so that in the first step it divides the data into two selected clusters based and their features. Summary results from testing fuzzy c-means algorithm by using data in the financial distress occurrence year (t year) are presented in Table 5.

*Table 5. Clustering data in the financial distress occurrence year*

| Number of feature(K) | Selected features | Degree of Unconformity |
|---|---|---|
| 24 | All features | 0.8966 |

Table 6 shows that if we use all features in designing a fuzzy c-means algorithm, then the model is able to separate two clusters with an accuracy of 89.66%. In our application, so we document unconformity of these two clusters with an accu-racy of close to 100%. In other words, our results imply that the clusters have been separated better and there is maximum asymmetry between these two clusters. In the second step, another test is performed to determine degree of symmetry for each data (firms) with Iran Business law Article 141 (criteria for classifying two classes). In this step, percentage of sym-metry between two generated clusters by fuzzy c-means with two clusters which have been classified under Article 141 to going concern. With this advantage that it could be determined data percentage to every class. By analyzing the finding is became clear that in going concern cluster based FCM, there was 100% of financial distress data (classified with criteria for Article 141 in Iran Business Law) but data which was classi-

fied in financial distress cluster under FCM 96.67% at them are the same data which were attributed to financial distress class under Article 141 in Iran Business law (implying that 3.33% error and all errors is related to financial distress class).

Summary results of the study are presented in Table 6.

*Table 6. Degree of conformity FCM clusters with clusters of Iran Business law 141 Article in the financial distress occurrence year*

| Selected feature | $\alpha_1$ | $\alpha_2$ | $\beta_1$ | $\beta_2$ |
|---|---|---|---|---|
| All features | .96670 | 1 | 1 | 0. |

Where:

$\alpha_1$: number of data financial distress is correct clustering to total number of data financial distress.

$\alpha_2$: number of data going concern is correct clustering to total number of data going concern.

$\beta_1$: number of data financial distress is incorrect clustering to total number of data incorrect clustering.

$\beta_2$: number of data going concern is incorrect clustering to total number of data incorrect clustering.
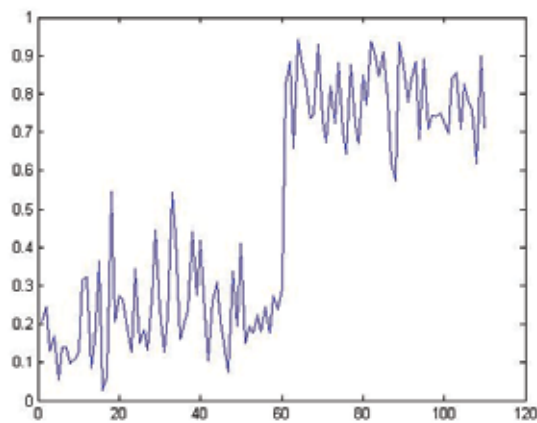
Figure 2 shows the amount of data to its class.



*Figure 2. Data in financial distress occurrence year*

Horizontal axis show the number of firms and vertical axis show the percent of belong for data to its class. Financial distress data are in the right side and going concern data are in the left side in this axis. The closer the data in its class to top horizontal axis (financial distress) or down (going concern) are, their percent belong to its class is greater. Results from testing fuzzy c-means algorithm by using data in one year before financial distress (t-1 year are provided in Table 7.

*Table 7. Clustering data in year before financial distress*

| Number of feature(K) | Selected features | Degree of Unconformity |
|---|---|---|
| 24 | All features | 0.8519 |

Another test was performed to determine the amount conformity between clusters based FCM and Iran Business law 141 Article. Summary results of this research have been provided in Table 8 based on selected feature.

*Table 8. Degree of conformity FCM clusters with clusters of Iran Business law 141 Article in one year before the financial distress occurrence*

| Selected features | $\alpha_1$ | $\alpha_2$ | $\beta_1$ | $\beta_2$ |
|---|---|---|---|---|
| All features | 0.83440 | 1 | 1 | 0 |

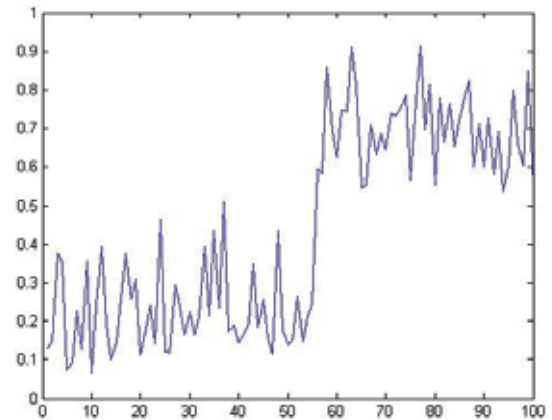Figure 3 shows the amount of belong for data to its class



*Figure 3. Data in one year before financial distress occurrence*

Results from testing fuzzy c-means algorithm by using data in two year before financial distress (t-2 year) are provided in Table 9.

*Table 9. Clustering data in two years before financial distress*

| Number of feature(K) | Selected features | Degree of Unconformity |
|---|---|---|
| 24 | Features | .7538 |

Another test was performed to determine the amount conformity between clusters based FCM and Iran Business law 141 Article. Summary results of this research have been provided in Table 10 based on selected feature.

*Table 10. Degree of conformity FCM clusters with clusters of Iran Business law 141 Article in two year before the financial distress occurrence*

| Selected features | $\alpha_1$ | $\alpha_2$ | $\beta_1$ | $\beta_2$ |
|---|---|---|---|---|
| All features | 0.7734 | 0.9832 | 0.96 | 0.04 |

Figure 4 shows the amount of data to its class in two year before the financial distress.

As we showed, in two years before financial distress belong percent for going concern data to their class have become lower. On the other hand, behavior at going concern firms in further years have had not so many stability (constant) and some they have had tendency toward financial distress. Generally, results from testing fuzzy c-means algorithm show that the model cluster financial distress firms in accuracy 96.67, 83.44 and 77.34 percent respectively by using data in financial distress year, one and two years before financial distress, thus in classifying going concern firms it could cluster in accuracy 100, 100 & 98.32 percent respectively by using
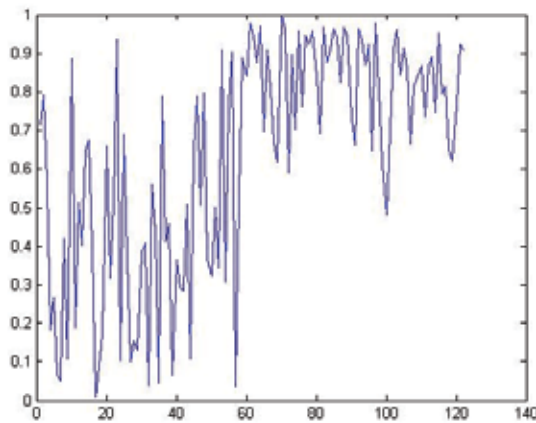
*Figure 4. Percent data in two years before financial distress occurrence*

data for financial distress occurrence year, one and two years before it. Obtained results from both algorithms have been summarized in Table 11.

*Table 11. Algorithm test based fuzzy c-means*

| Model | SVDD | | | FCM | | |
|---|---|---|---|---|---|---|
| Period of financial distress | t | t-1 | t-2 | t | t-1 | t-2 |
| financial distress prediction | 0.92 | 0.85 | 0.78 | 0.97 | 0.83 | 0.77 |

# 6    Conclusion

To the best of our knowledge, this paper is the first to model financial distress using SVDD. We show the flexibly of the proposed measure with noise rejection capability. Mapping input vectors into a high-dimensional feature space, SVDD transforms complex problems (with complex decision surfaces) into simpler problems that can use linear discriminated functions, and it has been successfully introduced in several financial applications recently. Particularly in this study, we utilize a grid-search technique using 3-fold cross-validation in order to choose optimal values of the upper bound C and the kernel parameter g that are most important in SVDD model selection. Selecting the optimal parameter values through the grid-search, we could build a going concern prediction model with high stability and prediction power. Results from algorithm test based on support vector data description show that model accuracy in classifying financial distress samples in the financial distress occurrence year, One and two years before it, is 91.9%, 85% and 78% respectively. Noticeable point about this model is lack of falling in predicting accuracy in experimental sample relative to educational sample, which could represent relevant general ability of this model.

Results from algorithm test based fuzzy c-means indicated that model accuracy in classifying financial distress samples in the financial distress occurrence year, One and two years before it, is 96.67%, 83.44% and 77.34% respectively.

Our experimental results showed that SVDD approach obtains better prediction accuracy than the FCM approach in developing a financial distress prediction model.

# Acknowledgements

# Literature

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4), 589-609, http://dx.doi.org/10.1111%2Fj.1540-6261.1968.tb00843.x

Altman, E. I. (1983). Multidimensional graphics and bankruptcy prediction: a comment. *Journal of Accounting Research*, 21(Spring), 297-299, http://dx.doi.org/10.2307%2F2490950

Back, B., Laitinen, T., Sere, K. & van Wezel, M. (1996a). Choosing Bankruptcy Predictors using Discriminant Analysis, Logit Analysis and Genetic Algorithms', Technical Report no. 40, Turku Centre for Computer Science, Turku School of Economics and Business Administration.

Back, B., Laitinen, T. & Sere, K. (1996b). Neural networks and genetic algorithms for bankruptcy predictions. *Expert Systems with Applications*, 11(4), 407-413, http://dx.doi.org/10.1016%2FS0957-4174%2896%2900055-3

Becchetti, L. J. & Sierra (2003). Bankruptcy Risk and Productive Efficiency in Manufacturing Firms, *Journal of Banking and Finance*, 27, 2099-2120.

Chen M.Y. & Du Y.K. (2009). Using neural networks and data mining techniques for the financial distress prediction model. *Expert Systems with Applications*, 36(2), 4075–4086, http://dx.doi.org/10.1016/j.eswa.2008.03.020

Chuang, C. C, Jeng, J. T. & Lin, P. T. (2004). Annealing robust radial basis function networks for function approximation with outliers, *Neurocomputing*, 56, 123-139, http://dx.doi.org/10.1016/S0925-2312(03)00436-3

Dambolena, I.G. & S.J. Khoury. (1980). Ratio Stability and Corporate Failure, *The Journal of Finance*, 35:1017-1026, http://dx.doi.org/10.1111%2Fj.1540-6261.1980.tb03517.x

Gombola, M.J. & Ketz, J.E. (1983). A Note on Cash Flow and Classification Patterns of Financial Ratios. *The Accounting Review*, 58(1), 105-114, http://www.jstor.org/stable/246645

Hassoun, M. H. (1995). *Fundamentals of Artificial Neural Networks*, MIT Press, Cambridge.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. London: Prentice-Hall.

Hong-Bao, W., Wang Fu-Sheng, W. & Xian-Fei, Y. (2012). Financial Distress Prediction Based on Cost Sensitive Learning. *Information Technology Journal*, 11(2), 294-300, http://dx.doi.org/10.3923/itj.2012.294.300

Hertz, J., Krogh, A. & Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*, Addison-Wesley, Reading

Jang, J. S. R. (1993). ANFIS: Adaptive-network-based fuzzy inference systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), 665-685, http://dx.doi.org/10.1109%2F21.256541

Jang, J. S. R., Sun, C. T, & Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice-Hall, NJ.

Jo, H. & Han, I. (1997). Bankruptcy Prediction Using Case-Based Reasoning, Neural Networks, and Discriminant Analysis.

*Expert Systems with Applications,* 13, 97–108 http://dx.doi.org/10.1016/S0957-4174(97)00011-0

Karels, G.V. & Prakash A.J. (1987). Multivariate Normality and Forecasting of Corporate Bankruptcy. *Journal of Business Finance and Accounting*, 14 (4), 573-592, http://dx.doi.org/10.1111/j.1468-5957.1987.tb00113.x

Kim, M. T., Han, H. R., & Phillips, L. (2003). Metric equivalence assessment in cross-cultural research: Using an example of the center for epidemiological studies-depression scale (CES-D). *Journal of Nursing Measurement*, 11, 5–18

Laitinen, E. K. & Laitinen, T. (2000). Bankruptcy Prediction: Application of the Taylor's Expansion in Logistic Regression, *International Review of Financial Analysis*, 9(4), 327-349, http://dx.doi.org/10.1016/S1057-5219(00)00039-9

Lin, F., Yeh, C.C. & Lee, M.Y. (2011). The use of hybrid manifold learning and support vector machines in the prediction of business failure, *Knowledge-Based Systems*, 24(1), 95-101, http://dx.doi.org/10.1016/j.knosys.2010.07.009,

Min, J.H. & Lee, Y.C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters, *Expert Systems with Applications*, 28(4): 603-614, http://dx.doi.org/10.1016/j.eswa.2004.12.008

Nam, M. D. & Thanh, T. C. (2003). Approximation of function and its derivatives using radial basis function networks. *Applied Mathematical Modeling*, 27(3), 197-221. http://dx.doi.org/10.1016/j.eswa.2009.07.081

Ohlson, J. (1980).Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109-131, http://www.jstor.org/stable/2490395

Ravisankar, P. & Ravi, V. (2010). Financial distress prediction in banks using Group Method of Data Handling neural network, counter propagation neural network and fuzzy ARTMAP. *Knowledge-Based Systems*, 23(8), 823-831, http://dx.doi.org/10.1016/j.knosys.2010.05.007

Scott, J. (1981). The probability of bankruptcy: A comparison of empirical prediction and theoretical model. *Journal of Banking and Finance*, 5, 317-344, http://dx.doi.org/10.1016/0378-4266(81)90029-7

Shin, K., Lee, T.S. & Kim, H. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, 28: 127-135. http://dx.doi.org/10.1016/j.eswa.2004.08.009

Song et al. (2010). Feature selection for support vector machine in financial crisis prediction: a case study in China. *Expert Systems*, 27(4), 299–310, http://dx.doi.org/10.1111/j.1468-0394.2010.00546.x

Stam, A., Sun, M. & Haines, M. (1996). Artificial neural network representations for hierarchical preference structures. *Computers and Operations Research*, 23(12), 1191-1201, http://dx.doi.org/10.1016/S0305-0548(96)00021-4

Stiglitz, J. E., & Weiss A. (1981). Credit rationing in markets with imperfect information. *American Economic Review*, 71(June), 393-410. http://www.jstor.org/stable/1802787

Sun, J., & Li, H. (2008). Data mining method for listed companies' financial distress prediction. *Knowledge-Based Systems*, 21(1), 1–5, http://dx.doi.org/10.1016/j.knosys.2006.11.003,

Surendra, R. & Krishnamurthy, A. (1997). Face recognition using transform features and neural networks, *Neural Networks*, 30(10), 1615-1622. http://dx.doi.org/10.1016/S0031-3203(96)00184-7

Tam, K. Y. & Kiang, M. (1992). Managerial Applications of Neural Networks: The Case of Bank Failure Predictions, *Management Science*, 38(7), 926-947.

Vapnik V. (1995). *The Nature of Statistical Learning Theory.* Springer-Verlag, New York.

Vythoulkas, P. C. & Koutsopoulos, H. N. (2003). Modeling discrete choice behavior using concepts from fuzzy set theory, approximate reasoning and neural networks. *Transportation Research*, 11, 51-73, http://dx.doi.org/10.1016%2FS0968-090X%2802%2900021-9

Wilson, R. L. & Sharda, R. (1994). Bankruptcy Prediction Using Neural Networks, *Decision Support Systems*, 11, 545-557. http://dx.doi.org/10.1016/0167-9236(94)90024-8

Wong B. K., Bodnovich T.A. & Selvi Y. (1997). Neural network applications in business: a review and analysis of the literature (1988–95). *Decision Support Systems*, 19, 301–320, http://www.jstor.org/stable/254047

Wong, B. & Selvi, Y. (1998). Neural network applications in finance: A review and analysis of literature (1990-1996), *Information & Management*, 34, 129-139. http://dx.doi.org/10.1016/S0378-7206(98)00050-0

Xu, B. (2000). The welfare implications of costly monitoring in the credit market: A note. *The Economic Journal*, 110(463), 576–580, http://dx.doi.org/10.1111%2F1468-0297.00538

Zmijewski, M. E. (1984). Methodological issues related to the estimated of financial distress prediction models. *Journal of Accounting Research*, 22(1), 59–82, **http://www.jstor.org/stable/2490859**

**Mahdi Moradi** is an Associate Professor of Accounting, Ferdowsi University of Mashhad, Iran, He has more than 20 years academic experiences in Iran. So far he has published more than 25 papers in international journals and 12 papers in national journals in Persian. His research interests include financial reporting, ERP and Auditing.

**Mahdi Salehi** is an Assistant Professor of Accounting, Ferdowsi University of Mashhad, Iran. He has more than 8 years of academic experience. So far, he has published more than 161 papers in international journals, his interests include: Auditing, Audit expectation gap, and financial distress prediction.

**Mohammad Ebrahim Ghorgani** has obtained his M.A in Accounting from the Ferdowsi University of Mashhad, Iran. Currently he is an Expert of the East Oil and Gas Company, NIOC, Iran.

**Hadi Sadoghi Yazdi** is an Associate Professor of Computer Science, Ferdowsi University of Mashhad, Iran. He has more than 18 years academic experiences in Iranian Universities. So far, he has published more than 120 papers in different journals.

**Predvidevanje finančnih pretresov v iranskih podjetjih z uporabo rudarjenja podatkov**

Odločanje na področju evaluacije finančnega statusa podjetij so zelo pomembne: napačne odločitve zelo verjetno povzročijo pretres in finančno krizo podjetja. Predvidevanje finančnih kriz in pretresov v proizvodnih podjetjih je pomembno za manager-je, investitorje, revizorje, finančne analitike, državne uradnike in zaposlene. Cilj tega članka je analizirati predvidevanje finanč-nih pretresov v iranskih podjetjih. Naša študija uporablja metodo SVDD (Support Vector Data Description) za predvidevanje finančnih pretresov in predlaga nov bolj stabilen  model predvidevanja z večjo močjo razlage. V ta namen smo uporabili tehni-ko preiskovanja mreže in uporabili 3-kratno prečno validacijo, da smo poiskali parametre jedrne funkcije SVDD. Da bi ocenili natančnost predvidevanja SVDD, smo jo primerjali z metodo FCM (fuzzy c-means). Rezultati eksperimenta so pokazali, da je SVDD uspešnejša od drugih metod v letih pred pojavam finančnega pretresa. Podatke, ki smo jih uporabili v naši študiji, smo dobili s teheranske borze in baze podatkov računovodskih raziskav.  V skladu s podatki iz obdobja 2000 in 2009 smo izbrali 70 parov družb, ki so bile uvrščene na teheransko borzo.

**Ključne besede:** predvidevanje finančnega pretresa, SVDD, FCM