Comparison of SVM and FSVM for predicting bank failures using chi-square feature selection

Z Rustam¹,³, F Nadhifa¹,⁴ and M Acar²,⁵

¹ Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok 16424, Indonesia
² Department of Business Administration, Faculty of Economics and Administrative Sciences, Sulcuk University, Campus 42031, Konya, Turkey

E-mail: ³rustam@ui.ac.id, ⁴farah.nadhifa@sci.ui.ac.id, ⁵melekacar@selcuk.edu.tr

Abstract. Bankruptcy doesn’t happen suddenly, but there are early indications that can be seen by investigating the financial statement of a bank. In this research, we aim to find the best bankruptcy prediction model to give an early warning for regulators so that it can help them to prevent or lessen the negative effects on economic systems. We will be performing SVM and modification of SVM by adding fuzzy membership function called FSVM to analyze bank’s health. We chose machine learning for bankruptcy prediction because it can give faster result rather than traditional statistical method. The prediction accuracy will be measured by using the dataset that consists of 65 Turkish banks of which each of them has an information of 20 financial ratios. Furthermore, to improve the accuracy prediction, we also perform chi-square feature selection (CSFS) to filter any irrelevant features of total 20 features in our dataset. CSFS can sort all 20 features based on chi-square score from the most relevant feature to the least one. After that, we will choose 5, 10, and 15 best features, so that we have four datasets to be classified into healthy and non-healthy banks. We found that using 5 features and SVM classifier gives the highest accuracy prediction, which scores 98.28%. For most cases, SVM gives better performance compared to FSVM.

1. Introduction
The bank plays a big role on economic system as they significantly contribute through the facilitation of business. Hence, the collapse of several banks can cause a huge damage to financial systems not only in a country but also globally as it had happened in the late 1990s and around 2008 [1]. There are so many caused that can triggered banks to go bankrupt such as interest rate volatility, high-risk taking, poor management, and inadequate accounting. As a consequence, the study of bankruptcy prediction of banks is becoming crucial for investors/creditors, borrowing firms, and governments as it can decrease the negative effects that can harm financial stability [2]. Constructing bankruptcy prediction model could identify banks with potential problems before they face solvency and liquidity crisis so that the danger of cascading failures of banks will be significantly reduced [3]. In other words, bankruptcy prediction model can act as an early warning system for regulators so that they can take actions before a financial crisis occurs.

To predict the bank’s health, we need to investigate bank performance which can be influenced by internal factors (bank specific) and macroeconomic (external) [4]. There are various rating systems that can be used to determine bank’s healthiness through analyzing their balance sheet reports. Therefore, in this research, we will be using dataset containing 20 financial ratios from 65 Turkish banks in the period of 1997—2004 issued by Banks Association of Turkey (BAT) as the variable predictor based on
CAMELS rating system [5]. In order to achieve a faster result, we chose to implement a machine learning technique rather than the traditional statistical technique. Moreover, machine learning capability for pattern recognition is very essential for this case. With two supervised machine learning approaches executed that is Support Vector Machines (SVM) and modified SVM called Fuzzy Support Vector Machines (FSVM) to construct the prediction. In previous studies, SVM has been applied by I.P.A Wardana and Z Rustam for decision-making in stock investment [6], Rustam and Frederica for insolvency prediction in insurance companies [7], Z. Rustam, D.F. Vibranti, and D. Widya for predicting the direction of Indonesian stock price movement [8]. Therefore, we know that SVM could be a great tool for building a prediction model, especially in the financial field. Furthermore, we will also be modifying SVM using fuzzy membership and in some cases, according to [9], adding fuzzy membership to SVM will minimize the error classification. Furthermore, using chi-square feature selection (CSFS) we aim to filter any irrelevant features and deduct 20 features to 15, 10, and 5 best features. Thereupon, there will be four datasets to be processed using two classifiers as mention before. As a result, we can compare each accuracy prediction and find the best model for our problem.

A brief explanation about theoretical background will be discussed in section 2. After that, section 3 will give an explanation about experimental designs for our research. Then, experimental results will be presented in section 4. And lastly, we will conclude and discuss the outcome of our research in section 5.

2. Theoretical background

2.1. Support vector machines (SVM)

Support vector machines (SVM) is a powerful method that gives high accuracy in terms of prediction because it holds structural risk minimization (SRM) principle. Moreover, according to [5] the application of SVM is relatively rare compared to neural networks and statistical method. Hence, we chose SVM as the classifier to construct a model for bankruptcy prediction. The idea of SVM is to find the best discriminant boundaries of two classes called hyperplane. In this case, the hyperplane is discriminant boundaries with the largest margin. We use [10], [11], and [12] as our main references for this section. In linearly separable case, given sample of several banks as training dataset $x_i \in \mathbb{R}^n$ with corresponding label $y_i \in \{-1, +1\}$ ($i = 1, 2, ..., N$), we can formulate a separator as $f(x) = w \cdot x + b = 0$. Thus, maximizing hyperplane can be done by solving optimization primal problem (1) respected to constraint (2) as follows:

$$\min \left( \frac{1}{2} ||w||^2 \right)$$ \hspace{1cm} (1)

$$s.t. \ y_i(w \cdot x_i + b) \geq 1, \forall i = 1, ..., N$$ \hspace{1cm} (2)

Where $w$ is the parameter vector and $b$ is the bias. Therefore, the equation above can be qualified as a quadratic programming (QP) problem. We will solve the optimization problem by introducing Lagrange multipliers $\alpha_i \geq 0$ for each constraint in (2). Using Lagrange multipliers will obtain inner product form which is very useful for non-separable case later on. Hence, we have primal Lagrangian form as follows:

$$L(w, b, \alpha) = \frac{1}{2} ||w||^2 - \sum_{i=1}^{N} \alpha_i (y_i ((w \cdot x_i) + b) - 1))$$ \hspace{1cm} (3)

Setting the derivatives $L(w, b, \alpha)$ respected to $w$ and $b$ both equal to zero, we will get

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{N} \alpha_i y_i x_i = 0$$ \hspace{1cm} (4)

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{N} \alpha_i y_i = 0$$ \hspace{1cm} (5)
Substitute both (4) and (5) to (3) will eliminate $w$ and $b$ so that we have dual Lagrangian form as follows:

$$\max \left( -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i x_j) + \sum_{i=1}^{N} \alpha_i \right)$$  \hspace{1cm} (6)

s.t. $\sum_{i=1}^{N} \alpha_i y_i = 0, \alpha_i \geq 0 \ (i = 1, ..., N)$  \hspace{1cm} (7)

In real life cases, it’s very rare to find linearly separable data as we assume in the previous case. Hence, we use the kernel trick method to couple with SVM to cope with non-linear separable data. Kernel trick follows the projection idea, where the dataset is mapped into an inner product space. Furthermore, we will be mapping data from the original space to another higher space using kernel function defined as $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$. Therefore, the data will be linearly separable without even knowing $\phi$ explicitly [10]. Hence substituting $K$ to (6) we will have:

$$\max \left( -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) + \sum_{i=1}^{N} \alpha_i \right)$$  \hspace{1cm} (8)

We can find the value of $\alpha_i$ using SMO algorithm and assume there are $\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_2, ..., \bar{\alpha}_N)$ as the solution of Lagrange multipliers so that we can easily obtain $\bar{w}, \bar{b},$ and $f(x)$ in Eq. (9), (10), (11) consecutively

$$\bar{w} = \sum_{i=1}^{N} \bar{\alpha}_i y_i x_i$$  \hspace{1cm} (9)

$$\bar{b} = y_u - \sum_{i \in S} \bar{\alpha}_i y_i K(x_i, x_u)$$  \hspace{1cm} (10)

$$f(x) = sgn \left( \sum_{i \in S} \bar{\alpha}_i y_i K(x_i, x_u) + \bar{b} \right)$$  \hspace{1cm} (11)

It is stated in [11] that a low polynomial or RBF kernel give satisfactory results and to outperform conventional classifiers. Hence, we’ll try to use both kernels with the following formulation respectively:

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^d$$  \hspace{1cm} (12)

$$K(x_i, x_j) = exp \left( -\frac{||x_i - x_j||^2}{2\sigma^2} \right)$$  \hspace{1cm} (13)

2.2. Fuzzy support vector machines (FSVM)

We will modify SVM by adding fuzzy membership to each data in order to have weighted data attributed to each corresponding class [9]. We formulate fuzzy membership for each data in (1) and (2) [13]

$$\mu_{i y_0} = e^{-\frac{||x_i||^2}{\sigma^2}} \text{ for } x_i \in y_0 = 0$$  \hspace{1cm} (14)

$$\mu_{i y_1} = 1 - e^{-\frac{||x_i||^2}{\sigma^2}} \text{ for } x_i \in y_1 = 1$$  \hspace{1cm} (15)
where $y_l (l = 0, 1)$ is observed result or class of healthy ($y_0 = 0$) or non-healthy banks ($y_1 = 1$),

$$\|x_l\|^2 = \sum_{i=1}^{N} x_l^2, \sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_l - \bar{x})^2,$$

and $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_l$ with $N$ as sample size. The idea is to have weighted data by multiplying every data with corresponding fuzzy membership value.

2.3. **Chi-square feature selection**

Feature selection is one of many ways to improve data quality. This research will propose chi-square feature selection (CSFS) to filter any redundant ratios by investigating chi-square score of every ratio. CSFS can measure how dependent a ratio respected to the corresponding class. In this case, the smaller chi-square score indicates that it gives a bigger impact for determining bank’s health. First, we will construct a contingency table that provides the foundation for statistical inference to see the relationship between observed variables.

| $C_1$ | $n_{10}(\mu_{10})$ | $n_{11}(\mu_{11})$ | $n_{1*}$ |
|------|-----------------|-----------------|-------|
| ...  | ...             | ...             | ...   |
| $C_r$| $n_{r0}(\mu_{r0})$ | $n_{r1}(\mu_{r1})$ | $n_{r*}$ |
|      | $n_{1*}$        | $n_{*1}$        | $n$ |

Chi-square score of every ratio will be measured using the following formulation [14]:

$$\chi^2 = \sum_{i=1}^{r} \left( \frac{(n_{i1} - \mu_{i1})^2}{\mu_{i1}} + \left( \frac{n_{i0} - \mu_{i0}}{\mu_{i0}} \right)^2 \right)$$

(16)

where $n_{ij}$ is the number of instances that have a value of $C_i$ (feature label) with $i = 1, \ldots, r$ and class $j$, $n_{*j} = \sum_{i=1}^{r} n_{ij}$, $n_{*1} = n_{i0} + n_{i1}$, expected value $\mu_{ij} = \frac{n_{ij}n_{*j}}{n}$, and $n = n_{*0} + n_{*1}$ with $n$ is the number of instances in the labeled data.

3. **Experimental designs**

3.1. **Dataset and variables**

Using data provided by Prof. Melek Acar [5] that she collected from the annual publication “Banks in Turkey” by the Banks Association of Turkey (BAT). This dataset has 21 solvent and 44 insolvent banks, which is 65 Turkish banks in total. Each of this bank has 20 financial ratios grouped in six components of CAMELS rating system including Capital adequacy, Asset quality, Management quality, Earnings, Liquidity, and Sensitivity to market. All 20 ratios from six components will be used as features in this study shows in Table 2 below.

| Ratio | Description | Ratio | Description |
|-------|-------------|-------|-------------|
| CA₁   | Shareholder’s equity/total assets | E₂    | Net profit/average shareholder’s equity |
| CA₂   | Shareholder’s equity/total loans | E₃    | Income before taxes/average assets |
| CA₃   | Shareholder’s equity+net profit/total assets + off-balance sheet commitments | E₄    | Interest income/total operating income |
| AQ₁   | Specific provision/loans under follow-up | E₅    | Non-interest expenses/total operating income |
| AQ₂   | Total loans/total assets | L₁    | Liquid assets/total assets |
In the preprocessing phase, we will be labeling the continuous data to qualitative form so that we can filter the features using CSFS. Based on [4] each ratio can be categorized as strong, satisfactory, fair, marginal, and unsatisfactory. Hence, we labeled all ratios with the numerical value of 1, 2, 3, 4, and 5 by dividing the original domain into five subintervals equally.

3.2. Measure for performance evaluation
The performance of the classifier will be evaluated using the following metrics:

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100\%
\]  (17)

where TP is the number of true positive cases, which represents instances which are correctly categorized in the ‘positive’ class, FN is the number of false negatives, which represents ‘positive’ class instances that are wrongly classified, TN is the number of true negatives, which represents instances that are correctly categorized in the ‘negative’ class, and FP is the number of false positive, which represents ‘negative’ class cases that are wrongly classified. A higher accuracy value indicates better classifier performance.

4. Experimental results
4.1. Variable selection
Based on chi-square score for every ratio that we’ve measured using Eq. (16), we will be choosing 5, 10, and 15 best subset features for our prediction model. Consequently, we will have a total of four datasets including all features dataset. Table 2 below shows chi-square score for every feature by order which number 1 shows smallest chi-square score (most relevant feature) to largest chi-square score (least relevant feature).

| Var   | Chi  | Var   | Chi  | Var   | Chi  | Var   | Chi  |
|-------|------|-------|------|-------|------|-------|------|
| AQ4   | 0.299| AQ1   | 3.0499| SMR4  | 5.1933| AQ3   | 8.8697|
| CA1   | 0.9373| L1    | 3.3093| E5    | 6.2718| E3    | 9.1636|
| E1    | 1.3207| CA3   | 3.4552| AQ5   | 6.8119| M1    | 9.2849|
| L2    | 1.6967| CA2   | 4.144 | E3    | 7.8713| E4    | 11.8446|
| AQ2   | 2.5892| SMR1  | 4.8388| SMR3  | 8.55 | SMR2  | 21.2597|

4.2. Classifier accuracy
Table 3 and Table 4 below have summarized the result of our experiments. We have tried to used several parameter values and found out that highest accuracy prediction achieved while using kernel RBF with \( \sigma = 0.05 \) and polynomial kernel of degree 4 \((d = 4)\). We split SVM and FSVM into four categories correspond to features subsets that we are using while performing two different kernels for each category.

| Data    | Accuracy (%) |
|---------|--------------|
| training|              |
| 20 features| 15 features| 10 features| 5 features|
|         | RBF | Polynom | RBF | Polynom | RBF | Polynom | RBF | Polynom |
|---------|-----|---------|-----|---------|-----|---------|-----|---------|
| 10%     | 79.31 | 65.52 | 84.48 | 68.97 | 84.48 | 74.14 |
| 20%     | 82   | 51.92 | 84.62 | 59.62 | 48.08 | 65.38 |
| 30%     | 80   | 68.89 | 84.44 | 71.11 | 77.78 | 75.56 |
| 40%     | 76.92 | 87.18 | 82.05 | 82.05 | 76.92 | 92.31 |
| 50%     | 75   | 84.38 | 81.25 | 84.38 | 71.88 | 90.63 |
| 60%     | 88.46 | 69.23 | 88.46 | 69.23 | 88.46 | 84.62 |
| 70%     | 89.47 | 73.68 | 89.47 | 73.68 | 94.74 | 94.74 |
| 80%     | 84.62 | 84.62 | 85.71 | 85.71 | 85.71 | 92.86 |
| 90%     | 83.33 | 71.43 | 83.33 | 66.67 | 83.33 | 83    |

Table 5. The accuracy of FSVM from four different datasets using RBF kernel ($\sigma^2 = 0.05$) and polynomial kernel $d = 4$

| Data training | 20 features | 15 features | 10 features | 5 features |
|---------------|-------------|-------------|-------------|------------|
|               | SVM         | FSVM        | SVM         | FSVM       |
| 10%           |             |             |             |            |
| 10%           | 77.59       | 82.76       | 67.24       | 82.76      |
| 20%           | 73.08       | 80.77       | 69.23       | 80.77      |
| 30%           | 80          | 80.00       | 80          | 62.22      |
| 40%           | 82.05       | 82.05       | 74.36       | 84.62      |
| 50%           | 84.38       | 84.38       | 84.38       | 87.50      |
| 60%           | 80.77       | 80.77       | 84.62       | 84.62      |
| 70%           | 78.95       | 78.95       | 78.95       | 84.21      |
| 80%           | 76.92       | 84.62       | 71.43       | 78.57      |
| 90%           | 83.33       | 83.33       | 83.33       | 83.33     |

It can be seen from Table 4 that the best accuracy prediction generated by SVM is 98.28% while using 5 subset features and RBF kernel. Whereas, from Table 5 we know that FSVM gives the best accuracy prediction at 97.44% while using a polynomial kernel. We also can conclude that the accuracy prediction of both classifiers increase as we reduce the irrelevant features. In other words, selecting features using CSFS successfully increase our dataset quality so that we have a better bank failures prediction model. Furthermore, Table 6 below will be giving information for running time and memory usage of each algorithm to give a brief comparison between the two classifiers.

Table 6. Running time for each experiment with best accuracy prediction based on feature subset using SVM and FSVM

| Features | SVM          | FSVM         |
|----------|--------------|--------------|
|          | Running time (s) | Memory (Kb)  | Running time (s) | Memory (Kb)  |
| 5        | 0.874        | 14666.06     | 0.844           | 14146.82     |
| 10       | 0.917        | 21127.53     | 0.996           | 15736.51     |
| 15       | 1.484        | 26548.1      | 0.971           | 22380.82     |
| 20       | 3.837        | 110136.42    | 1.25            | 30640.19     |
|          | 1.778        | 43119.5275   | 1.01525         | 20726.085    |

From Table 6, we know that SVM has longer running time and two times bigger memory consumption than FSVM. However, the gap between the two classifiers in terms of running time isn’t too big.
5. Conclusion
This research aimed to apply SVM and modified SVM, called FSVM to construct a bankruptcy prediction model. As we can see from our experimental results, SVM gives higher accuracy with the highest value being 98.28% but while using 10 features, FSVM achieved better accuracy prediction with the value of 96.88%. Although SVM algorithm tends to spend more time and memory usage we still conclude that SVM will be a better bank failures predictor compared to FSVM because there isn’t a significant difference between the two cases. This may occur because FSVM tends to give a maximal performance in multiclass cases. For the next research, we suggest to use another modification approach, also try to find another bigger and more update dataset in order to have an even more reliable bankruptcy prediction model. We also suggest using another feature selection approach that can be used for continuous data. Then, the result of feature selection can be compared to our results by using CSFS.

6. References
[1] Hong H and Jean-Laurent V 2016 Predicting bank failure statistical technique versus intelligent technique VEAM Seminar Paper
[2] Sungjae L and Wu S C 2013 A multi-industry bankruptcy prediction model using the back-propagation neural network and multivariate discriminant analysis Expert Systems with Applications 40 2941–2946
[3] Michael N 2017 Identification of failing banks using clustering with self-organizing neural networks NTELLI 2013: The Second International Conference on Intelligent Systems and Applications 108C 1327–1333
[4] Desta T S 2016 Financial performance of “best african banks”: Comparative analysis through CAMEL rating Journal of Accounting and Management 6 1–20.
[5] Melek A, Kara Y and Baykan Ö 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.
[6] [6] I P A Wardana and Z Rustam 2017 Application of support vector machines on decision-making in stock investment Proceeding of 3rd International Symposium on Current Progress in Mathematics and Sciences
[7] Z Rustam and Frederica Y 2017 Insolvency prediction in insurance companies using support vector machines and fuzzy kernel c-means Journal of Physics: Conference Series 1028 012118
[8] Z Rustam, D F Vibranri and D Widya 2017 Predicting the direction of Indonesia stock price movement using support vector machines and fuzzy kernel c-means Proceeding of 3rd International Symposium on Current Progress in Mathematics and Sciences
[9] Shigeo A and Takuya I 2002 Fuzzy support vector machines for multiclass problems European Symposium on Artificial Neural Networks (Bruges: Belgium) p 113–118
[10] Burges C J C 1998 A tutorial on support vector machines for pattern recognition Data Mining and Knowledge Discovery 2 121–167
[11] Hofmann M 2006 Support vector machines–kernels and kernel trick Houptseminar Report
[12] Bishop C M 2006 Pattern Recognition and Machine Learning (New York: Springer)
[13] Z Rustam 2008 Face recognition using fuzzy supportvectorr machines National Mathematics Conference XIV
[14] Shaomin W and Flach P A 2002 Feature selection withlabeledd and unlabelled data Workshop on Integrating Aspects of Data Mining, Decision Support, and Meta-Learning p 156–167