

PAPER • OPEN ACCESS

Application of Support Vector Machines in Evaluating the Internationalization Success of Companies

To cite this article: Z Rustam *et al* 2018 *J. Phys.: Conf. Ser.* **1108** 012038

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Application of Support Vector Machines in Evaluating the Internationalization Success of Companies

Z Rustam*¹, F Yaurita*¹ and M J Segovia-Vergas²

¹ Department of Mathematics, Universitas Indonesia, Depok, 16424, Indonesia
(E-mail: rustam@ui.ac.id, frederica.yaurita@ui.ac.id)

² Departement of Financial Economy and Accounting I, Facultad de Ciencias Económicas y Empresariales, Universidad Complutense de Madrid 28223, Spain.

Abstract. The internationalization started to be seen as an opportunity for many companies. This is one of the most crucial growth strategies for companies. Internationalization can be defined as a corporative strategy for growing through foreign markets. It can enhance the product lifetime and improve productivity and business efficiency. However, there is no general model for a successful international company. Therefore, the success of an internationalization procedure must be estimated based on different variables such as the status, strategy, and market characteristics of the company. In this paper, we try to build a model in evaluating the internationalization success of a company based on existing past data by using Support Vector Machines. The results are very encouraging and show that Support Vector Machines can be a useful tool in this sector. We found that Support Vector Machines achieved 81.36% accuracy rate with RBF Kernel, 80% training set, and $\sigma = 0.05$

1. Introduction

Internationalization started to be seen as an opportunity for many companies. It is an important strategic choice to face economic globalization challenge [1]. Internationalization is defined as the process of involvement increase in international operations [2, 3]. It refers to the degree of sale incomes or operations which firms direct in foreign market [4]. Thomson et al in 2015 said there are five reasons why a company may opt to enter the foreign market. There are: to gain access to new customers, to gain access to resources in foreign market, to gain access to low-cost inputs of production, to spread the business risk, and the last is to achieve lower costs through the increase of purchasing power [5]. However, internationalization is a high risky business strategy. The firms should prepare for political and economic risks, they will face international market change, technology, market and financial risks.

There is no general model for a successful international company. The success of an internationalization procedure must be estimated based on different variables such as the status, strategy, and market characteristics of the company [6]. In this study, we will build a model by using existing past data that have faced an internationalization process before. This model will be able to classify internationalization success or failure of a company. It will help a company to predict the internationalization success of the company base on the current company's condition.

We used machine learning techniques to build the model. Machine learning is an application of artificial intelligence that provides systems the ability to automatically learn and improve from experience without being explicitly programmed (Arthur Samuel, 1959). Support Vector Machines (SVM) is known as the powerful machine learning tool for classification. Application of SVM has been used by Rustam Z and Yaurita F [7], Rustam Z and Zahras D [8], Rustam Z and Ariantari N P A [9], Wardana I P A and Rustam Z [10], Puspitasari D A and Rustam Z [11], Rustam Z, Vibranti D F, and



Widya D [12], respectively for insolvency prediction, classifier for intrusion detection system, classifying policyholders satisfactorily, decision-making in stock investment, stock analysis of Indonesian stock exchange, and predicting the direction of Indonesian stock price movement.

Since a lot of financial problems could be solved by SVM and the results are trustworthy, in this study we try to conduct the insolvency prediction model used SVM algorithms. We built the model by using 595 companies which have faced an internationalization process before. The results are very encouraging and show that SVM can be a useful tool to predict company's internationalization success. The rest of paper is structured as follows. In the next section, there are brief introduction to SVM algorithm and kernel function. The third section describes the data set that we used and steps of the experiment. And finally, the last section contains some concluding remarks.

2. Support Vector Machines and Kernel Function

2.1. Support Vector Machines (SVM)

SVM has received much attention in classification problems. This method first proposed by Vapnik in 1998. In a few years later, Nello Cristianini (Professor of Artificial Intelligence at the University of Bristol) did a research about SVM based on Vapnik's theory before and have written the result in [13]. After that in 2002, Bernhard Scholkopf developed SVM theory and published it in a book which is about kernels in SVM as we can see in [14]. We used [15], [16], [17] and [18] as our main reference in this SVM section.

Let $D = \{(x_i, y_i)\}_{i=1}^N$ be a classification data set, with N points in d -dimensional space. Further, let us assume that there are two class labels, $y_i \in \{+1, -1\}$ denoting the positive and negative classes. The main purpose of SVM is to find the best hyperplane that is defined as follows:

$$f(x) = \mathbf{w}^T \cdot \mathbf{x} + b \quad (1)$$

Here, \mathbf{w} is a d dimensional weight vector, and b is a scalar, called bias. For points that lie on the hyperplane, we have:

$$\mathbf{w}^T \cdot \mathbf{x} + b = 0 \quad (2)$$

Figure 1 shows the illustration of SVM. Let the red point belongs to the positive class and the blue point belongs to the negative class. The H line is called hyperplane, H_- and H_+ are called canonical hyperplane. Support vectors are the data points that lie closest to the hyperplane. There are many hyperplanes which able to separate two classes, but there is only one the optimal hyperplane [15]. It should be the hyperplane that maximizes the margin between to class, so that the distance from the hyperplane to the nearest data point on each side is maximized.

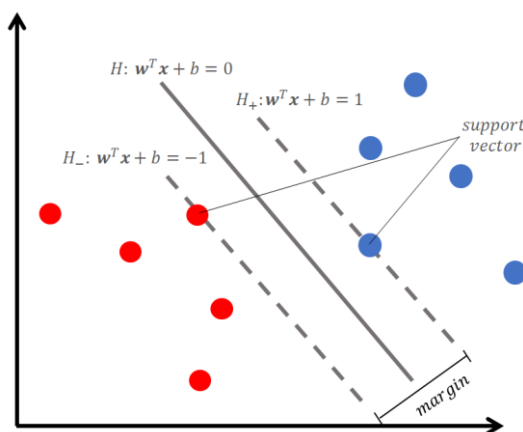


Figure 1 The Illustration of Support Vector Machines

The formula of margin is $\frac{2}{\|\mathbf{w}\|}$ where \mathbf{w} is a vector that perpendicular to the hyperplane. Consider that instead of maximizing the margin $\frac{2}{\|\mathbf{w}\|}$, we get an equivalent formula if we minimize $\|\mathbf{w}\|$. Then we can obtain an equivalent minimization formula given as follows:

$$\text{minimize } \frac{1}{2} \|\mathbf{w}\|^2 \quad (3)$$

$$\text{subject to } y_i(\mathbf{w}^T \cdot x_i + b) \geq 1, \forall i = 1, \dots, N \quad (4)$$

The **Eq. (3)** can be solved directly by using standard optimization algorithms. However, it is more common to solve the dual problem which is obtained by the use of Lagrange function. We introduce Lagrange multipliers α_i for each of the constraints in (4), thus we get:

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i \{y_i(\mathbf{w} \cdot x_i + b) - 1\} \quad (5)$$

where $a = (a_1, a_2, \dots, a_n)^T$. Setting the derivatives of $L(\mathbf{w}, b, \alpha)$ with respect to \mathbf{w} and b equal to zero, the following two conditions were obtained:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^N \alpha_i y_i x_i = 0 \rightarrow \mathbf{w} = \sum_{i=1}^N \alpha_i y_i x_i \quad (6)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^N \alpha_i y_i = 0 \rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \quad (7)$$

The **Eq. (3)** can be transformed into a dual problem by substituting the derivatives of Lagrangian (**Eq. (6), (7)**) back into the Lagrangian function [12], therefore we obtained the dual form as follows:

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

$$\text{subject to } \left\{ \sum_{i=1}^N y_i \alpha_i = 0, \alpha_i \geq 0 \right\} \quad (9)$$

There is a case that the data set is not linearly separable, therefore the data set should transform into a higher dimension using kernel function $\kappa(x_i, x_j) = \phi(x_i)^T \phi(x_j)$. In this paper, we used RBF and polynomial kernel as described in **Sec 2.2**, and by changing the inner product of two vectors with $\kappa(x_i, x_j)$. The **Eq. (8)** can be rewritten as:

$$L(\alpha) = \max \left\{ -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j K(x_i, x_j) + \sum_{i=1}^N \alpha_i \right\} \quad (10)$$

Solving **Eq. (10)** with constraints in **Eq. (9)** will determine the Lagrange multipliers (α_i). By using KKT condition the formula of \mathbf{w} and b are finally obtained as follows:

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i x_i \quad (11)$$

and

$$b = \frac{1}{N_s} \sum_{i \in S} \left(y_i - \sum_{m \in S} \alpha_m y_m x_m \right) \quad (12)$$

where N_s is the number of support vectors.

2.2. Kernel Function

Based on the prior experiments, most of training set are not linearly separable. To solve this problem, each vector \mathbf{x}_i is mapped to a higher dimensional space which is called feature space. It is expected that the transformed data can approach the linearly separable data, so that classification accuracy can be improved. A kernel is a function κ that for all $\mathbf{x}_i, \mathbf{x}_j \in X$ satisfies [14]:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle \quad (13)$$

where ϕ is a mapping from X to an (inner product) feature space F .

$$\phi: \mathbf{x}_i \rightarrow \phi(\mathbf{x}_i) \in F \quad (14)$$

In our research we used RBF and polynomial kernel that defined as follows:

$$RBF \quad \kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (15)$$

$$Polynomial \quad \kappa(\mathbf{x}_i, \mathbf{x}_j) = (\langle \mathbf{x}_i, \mathbf{x}_j \rangle + 1)^d \quad (16)$$

where d and σ are the adjustable parameters.

3. Experiment and Results

3.1. Data set

In this study, we used SVM algorithm to build the model that can predict internationalization's success of companies. The data set has got from Prof. M J Segovia [19]. This data set used is yearly elaborated by the Spanish Foundation of Statal Industrial Participation (SEPI), by investigated more than ten employees in manufacturing companies. This database is elaborated for the whole country and all variables involved are annual. In our study, there are 595 Spanish companies data including the success or failure of the company in its internationalization process. Each firm is described by 30 variables (V1, V2, ..., V30). All the respective variables describing the companies themselves and their internationalization strategy.

3.2. Results

The aim of this research is to build the model that able to predict the internationalization success of a company. We solved the problem by using SVM algorithm. In machine learning, before using the algorithm, we should split the data set into training set and test set. Training set is implemented to build up the model while the test set use to validate the model built (to check the accuracy). Therefore, the first step in our research is, we randomized all the data and split it into training set and test set. Let the training set is $p\%$ data from each class, then it is obvious the test data will be $(100 - p)\%$ data from each class.

Assume that $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ are the given training set, where $\{\mathbf{x}_i\}_{i=1}^N$ are the set of companies that have faced internationalization process with N points in d -dimensional space, and $y_i \in \{+1, -1\}$ denotes label of each company. Label $+1$ belongs to companies that succeeded in internationalization process, and -1 belongs to companies that failed in internationalization process. Therefore, we had two class in the experiment, the first class is the set of successful companies and the second class is the set of unsuccessful companies.

After randomized the data set, and determined several data to be training set, SVM algorithm worked by solving the quadratic problem in **Eq. (10)**. In short, the algorithm got the weight vector (\mathbf{w}) and bias (b). Substituting \mathbf{w} and b into **Eq. (1)** brought the algorithm to get the best function of hyperplane. The next step was to see how accurate the hyperplane for classifying the successful and unsuccessful companies. In this process, the algorithm used $(100 - p)\%$ data from each class. Every test

set is substituted into function $f(x)$. $f(x) > 1$ belongs to the success class, and $f(x) < -1$ belongs to unsuccess class. If the input data is a successful company and $f(x) > 0$, this indicates that the model does a true prediction. Same as if the input data is an unsuccessful company and $f(x) < 0$, this indicates that the model does a true prediction. Other than these means that our prediction is incorrect. Then the accuracy calculated by using this form:

$$Accuracy = \frac{\text{Number of true prediction}}{\text{Number of test set}} \times 100\% \quad (17)$$

4. Results

Table 1. Accuracy and running time of SVM algorithm with polynomial kernel (degree=2) and RBF kernel ($\sigma=0.5$)

| Training Set | Polynomial | | RBF | |
|--------------|---------------|------------------|---------------|------------------|
| | % Accuracy | Running Time (s) | % Accuracy | Running Time (s) |
| 20% | 60.92% | 0.0051 | 50.09% | 0.0051 |
| 40% | 70.23% | 0.0137 | 60.96% | 0.0091 |
| 60% | 73.53% | 0.0218 | 71.85% | 0.0133 |
| 80% | 79.66% | 0.0357 | 81.36% | 0.0174 |

Table 1 shows the summary of our experiments. In the first column, there is % training set, it is the percent data set (from each class) that implemented to build up a model. The second and third column is the accuracy and running time of experiments that used SVM algorithm with polynomial kernel while the fourth and fifth column is the accuracy and running time of experiments that used SVM algorithm with RBF kernel. Accuracy shows the percentage of true prediction that the model did, and the running time shows how long is the program can provide the result.

In the first experiment, we used polynomial kernel with degree = 2. It can be seen from **Table 1** that the best accuracy achieved when we used 80% data training, it is 79.66%, and the worst accuracy achieved when we used 20% training set. In the second experiment, we used RBF kernel with $\sigma = 0.05$. From **Table 1**, we know that the best accuracy achieved when we used 80% training set, it is 81.36%, and the worst accuracy achieved when we used 20% training set.

5. Conclusions

In this research, we have made a model for evaluating the internationalization success of companies by using 595 companies which have faced an internationalization process before. This new model can help a company to predict its internationalization success. The highest accuracy is obtained by the experiment that use RBF kernel with $\sigma = 0.05$, it is 81.36%. In future studies, we interested to try the other machine learning algorithms such as AdaBoost, kNN, Naïve Bayes, etc.

6. References

- [1] Jiang M and Xu M 2011 Chinese Firms 2011 *International Conference of Information Technology, Computer Engineering and Management Sciences* pp 186-189 IEEE. Behyan M 2016 *Investigating the Influence of External Networks on Intenationalization Success of Firms* IEEE 978-1-5090-2172-7/16, 2016.
- [2] Welch, L S and Luostarinen R 1988 *Internationalisation: evolution of a concept* Journal of General Management, Vol. 14 No. 2, pp. 34-55.

- [3] Holmlund M, Kock S and Vanyushyn V 2007 Small and medium-sized enterprises' internationalisation and the influence of importing on exporting *International Small Business Journal* **25** pp 459-477.
- [4] Thompson A A, Strickland A J, Gamble J E, and Zeng'an Gao 2008 Crafting and executing strategy: The quest for competitive advantage: Concepts and cases. Elango B and Pattnaik C *Building capabilities for international operations through networks: a study of Indian firms Journal of International Business Studies* **38** pp 541-555
- [5] Rustam Z and Yaurita F 2018 *Insolvency Prediction in Insurance Companies Using Support Vector Machines and Fuzzy Kernel C-Means* Journal of Physics: Conference Series J. Phys. 1028 012118.
- [6] Rustam Z and Durrabida Z 2018 *Comparison between Support Vector Machine and Fuzzy C-Means as Classifier for Intrusion Detection System* Journal of Physics: Conference Series J. Phys. 1028 012118 1028 012227.
- [7] Rustam Z and Ariantari NPA 2018 *Comparison between Support Vector Machine and Fuzzy C-Means as Classifier for Intrusion Detection System* Journal of Physics: Conference Series J. Phys. 1028 012005.
- [8] Wardana I P A and Rustam Z 2017 Application of Support Vector Machines on Decision-Making in Stock Investment *The 3rd ISCPMS AIP Proceeding Conference*.
- [9] Rustam, Z, Vibranti D F and Widya D 2017 Predicting The Direction of Indonesian Stock Price Movement using Support Vector Machines and Fuzzy Kernel C-Means *The 3rd ISCPMS AIP Proceeding Conference*.
- [10] Rustam Z and Yaurita 2017 Application of Support Vector Machines for Reject Inference in Credit Scoring *The 3rd ISCPMS AIP Proceeding Conference*.
- [11] Cristianini, N and Taylor J S 2000 *Support Vector Machines and other kernel-based learning methods* (New York: Cambridge University Press)
- [12] Taylor J S and Cristianini N 2004 *Kernel Methods for Pattern Analysis* (New York: Cambridge University Press)
- [13] Vapnik V N 1998 *Statistical Learning Theory* (USA: Wiley)
- [14] Scholkopf and Smola 2002 *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (USA: MIT)
- [15] Campbell C and Ying Y 2011 Learning with support vector machines *Synthesis lectures on artificial intelligence and machine learning* **5**(1) pp1-95
- [16] Zaki M J and Meira W 2013 *Data Mining and Analysis: Fundamental Concepts and Algorithms* (New York: Cambridge University Press)
- [17] Landa-Torres I, Ortiz-Garcia E G, Salcedo-Sanz S, Segovia-Vargas M J, Gil-Lopez S, Miranda M, ... & Del Ser J 2012 Evaluating the internationalization success of companies through a hybrid grouping harmony search—extreme learning machine approach. *IEEE Journal of Selected Topics in Signal Processing* **6**(4) pp 388-398
- [18] Stoean, C. & Stoean R. 2014 *Support Vector Machines and Evolutionary Algorithms for Classification* (London: Springer)