Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022 e-ISSN : 2541-2019
DOI : https://doi.org/ 10.33395/sinkron.v7i3. 11597 p-ISSN : 2541-044X

Support Vector Machine Using A Classification Algorithm

Nurul Huda Ovirianti[1)], Muhammad Zarlis[2)], Herman Mawengkang[3)]
University of Sumatera Utara, Medan, Indonesia
nhudaovirianti@gmail.com

Submitted : July 31, 2022 | **Accepted** : June 25, 2022 | **Published** : August 13, 2022

Abstract. Support vector machine is a part of machine learning approach based on statistical learning theory. Due to the higher accuracy of values, Support vector machines have become a focus for frequent machine learning users. This paper will introduce the basic theory of the Support vector machine, the basic idea of classification and the classification algorithm for the support vector machine that will be used. Solving the problem will use an algorithm and prove the effectiveness of the algorithm on the data that has been used. In this study, the support vector machine has obtained very good accuracy results in its completion. The SVM classification uses kernel RBF with optimum parameters Cost = 5 and gamma = 2 is 88%.

Keywords: Support Vector Machine, Classification Algorithm

INTRODUCTION

Support Vector Machines (SVM) have been recently developed in the framework of statistical learning theory and have been successfully applied to a number of applications, ranging from time series prediction (Müller et al., 1997), to face recognition (Guo et al., 2000), to biological data processing for medical diagnosis (Widodo & Yang, 2007). Their theoretical foundations and their experimental success encourage further research on their characteristics, as well as their further use (Ma & Guo, 2014).

Basically, vector machines support for class classification problems, but with the rapid development of computer technology, network technology, database technology, for the classification and management of large amounts of information, class classification problems can no longer meet the needs of society. This method is extended to the multi-class classification problem. It is currently a hot research topic. Multi-class classification problems in the existing treatment methods have the following two categories: First, build a series of second-class classification problems in some waly, the last class classification problem to solve multi-class classification problems, called based on two basically multi-class SVM classification method. Second, directly build multi-class SVM, one time to solve the multi-class classification problem called multi-class SVM (Deng et al., 2012).

By example, the complexity of the algorithm will be minimized. Based on the classification of support vector machines using the RBF kernel function and the choice of parameters that have been determined. This method can achieve very high classification accuracy.

LITERATURE REVIEW

Machine Learning

Machine learning approach based on statistical learning theory that is widely used to replace or imitate human behaviour to solve problems or perform automation (Ding et al., 2010). As the name implies, Machine Learning tries to imitate how humans or intelligent creatures learn and generalize processes. There are at least two main applications in Machine Learning, namely, classification and prediction. The hallmark of Machine Learning is the existence of a training, learning, or training

*Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022 e-ISSN : 2541-2019
DOI : https://doi.org/ 10.33395/sinkron.v7i3. 11597 p-ISSN : 2541-044X

process. Therefore, Machine Learning requires data to be learned which is known as training data. Based on the data that has been studied in the training. The most popular Machine Learning methods are Decision Making Systems, Support Vector Machines (SVM) and Neural Networks.

Machine learning algorithms are divided into two categories, namely supervised learning and unsupervised learning. Supervised learning is when we know what the results or targets of a data are, but we don't know how to calculate or how to get those targets (Li & Cui, 2011). Unsupervised learning does not have training data because it does not have results or targets called labels. He will learn by himself from the existing data to get a certain target. So there is no need for a training process (Zhang et al., 2012).

SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a guided learning system whose classification uses a hypothetical space in the form of linear functions in al high-dimensional features space that is trained using al learning algorithm based on optimization theory by implementing al learning bias. SVM was developed to solve classification problems because SVM has al belter ability to generalize data when compared to previously existing techniques. The approach using SVM has many benefits such as the model being built has an explicit dependence on al subset of datapoints, as well as support vectors that help in model interpretation (Jensen & Meckling, 2019).

Basically, SVM has al linear principle, but SVM has evolved so that it can work on non-linear problems. The waly SVM works on non-linear problems is to include the kernel concept in al high-dimensional space. In this dimensional space, al separator will bel sought later or often called al hyperplane. Hyperplanes can maximize the distance or margin between data classes (Inekwe et al., 2018). The best hyperplane between the two classes can bel found by measuring the margins and finding the maximum point. Efforts to find the best hyperplane as al class separator is the core of the process in the SVM method (Kiani et al., 2016).

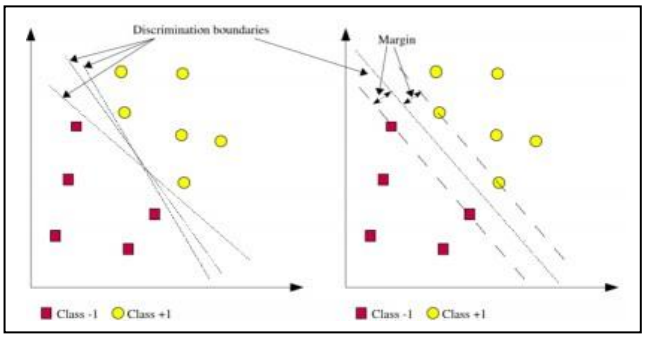

Figure 1. Support Vector Machine

In pattern recognition, the linear discriminate function in the n-dimensional space: g(x) = $\omega.x$ + b, the classification hyperplane equation can be written g(x) = ($\omega.x$) + b = 0. Linear separable case, the discriminate function g(x) was normalized. So that all training samples are met $|g(x)| \geq 1$ even to meet away from the classification of the surface of the sample $|g(x)| \geq 1$, so the class interval is equivalent to $\frac{2}{\|w\|}$, thus making the interval on the equivalent to $\|w\|$ or $\|w\|^2$. Make a classification of the surface of all samples correctly classified, it is necessary to meet:

$$y, [(w.x_i) + b] - 1 \geq 0, \qquad i = 1,2, \dots, n \tag{1}$$

Satisfy the above equation (1) and make $\|w\|^2$ the smallest classification surface is the optimal classification surface. Point on the hyper plane are called support vectors (support vector), they support the optimal classification surface (Jandik & Makhija, 2005).

SUPPORT Vector Machine Classification Algorithm

*Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022 e-ISSN : 2541-2019
DOI : https://doi.org/ 10.33395/sinkron.v7i3. 11597 p-ISSN : 2541-044X

Multi-class classification problem in mathematical language is described as follows: Training set $T = \{(x_1, y_1), \dots (x_1, y_1)\} \in (XxY)^1$, where $x_i \in x = R^n, y_1 \in Y = \{1,2, \dots, k\}, i = 1, \dots l$

Find al decision Function $f(x): X = R^n \rightarrow Y$

Thus, solving the multi-class classification problem is to find the point of n R into the part of k the rules. If k = 2, Compared to second-class classification problem, directly using SVM to solve the problem; if k > 2 for the multi-class classification problem. Handle multi-class support vector machine algorithm: One-against-rest algorithms, one- against -one algorithm, al directed acyclic graph algorithms, error correction coding algorithm, the binary tree algorithm. The following describes the One-against-rest algorithm, one-against-one algorithm, al directed acyclic graph algorithms (Chen & Du, 2009; Gregova et al., 2020).

Financial Distress

Financial distress is the stage of downturn in financial condition that occurs in al company before bankruptcy or liquidation. Al company can bel categorized as experiencing financial difficulties if the company shows negatively figures on operating profit, net profit and equity book value and the company merges. Another phenomenon of financial difficulties that companies tend to experience liquidity difficulties indicated by the company's declining ability to fulfil its obligations to creditors. Information Financial difficulties are indispensable in the framework of checking the condition of an enterprise (Wu et al., 2020).

METHOD

This study uses a literature study as al reference to find theories related to support vector machines (SVM) to predict financial difficulties contained in books, journals, articles, and previous research that are relevant to the case or problem in this study. This study uses al classification algorithm in classifying companies that are included in Financial Distress and not Financial Distress. In research, a research flow or flowchart is needed which is used as al reference in the research process. This research was conducted in several steps, including:

1. Determine research variables and collect data to bel used.
2. Input data that has been taken from the Indonesia Stock Exchange website. The data inputted is data on Profitability, Leverage, Liquidity and Company Assets.
3. Classify companies that are predicted to experience financial distress (Financial Distress).
4. Divide the data into 2 parts, namely training data and testing data. The proportion of training data is greater than the testing data. The training data is used to train the data in forming the model, while the testing data is used to predict and see the accuracy of the model formed.
5. From the training data, it will bel done to determine the kernel on the Support Vector Machine.
6. From the determination of the kernel, the Support Vector Machine model will bel obtained.
7. Perform al confusion matrix to determine the accuracy value of the Support Vector Machine method.
8. Calculating the value of accuracy using data testing to conduct research on financial distress predictions.
9. Gelt the results of the classification that has been done

RESULT AND DISCUSSION

The data used is data on property and real estate companies as many as 60 companies that are divided into financial distress and non-financial distress companies. Companies that experience financial distress or consist of 45 companies or 75% and companies that do not experience financial difficulties (non-financial distress) consist of 15 companies or 25%.

The variables used in this study consisted of dependent variables (Y) and independent variables (X). The dependent variables used in this study are the condition of financial distress of al company while the independent variables used are Profitability, Leverage, Liquidity and Assets which are used to find how much influence it has on financial distress in property and real estate companies on the Indonesia Stock Exchange in 2017-2020.

*Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022 e-ISSN : 2541-2019
DOI : https://doi.org/ 10.33395/sinkron.v7i3. 11597 p-ISSN : 2541-044X

The data will bel classified into two, namely training data and testing data. Training data is used to gelt the model, while testing data is used to test the model obtained. This is useful for finding good modelling accuracy. The distribution of training data and testing data was carried out randomly using al comparison of 70% for training data and 30% for testing data.

RBF kernel is a kernel function used when data separates linearly. In carrying out analysis with the RBF function, optimization of cost (C) and gamma (γ) parameters is carried out. In determining the best parameters, trial and error are carried out. Researchers used the C value trial four times, namely 1, 5, 10, and 50, and the gamma value trial four times, namely 1, 2, 3, and 4. Here is the result of the accuracy value on the RBF kernel as follows:

Table 1. Test Results

Parameters	accuracy			
	γ = 1	γ = 2	γ = 2	γ = 4
C = 1	0.94	0.94	0.94	0.94
C = 5	0.94	0.96	0.94	0.94
C = 10	0.94	0.94	0.94	0.94
C = 50	0.94	0.94	0.94	0.94

From the cost and gamma parameters above, al confusion matrix can bel created to obtain accuracy values from the SVM model of the RBF kernel using the testing dataset.

Table 2. Confusion Matrix

Predictions	Current	
	Non-FD	FD
Non-FD	7	2
FD	0	7

Based on the table above, the results are obtained that there are two possible company status, namely financial distress or non-financial distress. In the classifier table above, there are 16 predictions obtained. From the sample above, the classifier predicts the "non-FD" option is 9 times and the "FD" prediction is 7 times. After obtaining the confusion matrix, it can bel continued by looking for accuracy values from testing data using the following calculations:

$$\text{accuracy} = \frac{TP+TN}{TP+TN+FN+FP}$$

$$\text{accuracy} = \frac{7+7}{7+2+0+7} = \frac{14}{16}$$

$$\text{accuracy} = 0{,}875$$

So the classification accuracy value of 0.875 or 88% was obtained.

CONCLUSION

Based on the results of the research that has been described in the discussion chapter, it can be concluded that the general description of the financial ratio of property and real estate companies in 2017-2020 shows that there are 45 companies that do not experience financial distress and those that experience financial distress as many as 15 companies out of 60 companies. The SVM classification uses kernel RBF with optimum parameters Cost = 5 and gamma = 2 is 88%.

REFERENCES

Chen, W.-S. & 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.

Deng, N., Tian, Y. & Zhang, C. (2012). *Support vector machines: optimization based theory, algorithms, and extensions*. CRC press.

Ding, Y., Qin, X. & He, H. hui. (2010). Parameter optimizing of support vector machine and

*Corresponding author

application in text classification. *Computer Simulation*, *27*(11), 187–190.
Gregova, E., Valaskova, K., Adamko, P., Tumpach, M. & Jaros, J. (2020). Predicting financial distress of slovak enterprises: Comparison of selected traditional and learning algorithms methods. *Sustainability*, *12*(10), 3954.
Guo, G., Li, S. Z. & Chan, K. (2000). Face recognition by support vector machines. *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, 196–201.
Inekwe, J. N., Jin, Y. & Valenzuela, M. R. (2018). The effects of financial distress: Evidence from US GDP growth. *Economic Modelling*, *72*, 8–21.
Jandik, T. & Makhija, A. K. (2005). Debt, debt structure and corporate performance after unsuccessful takeovers: evidence from targets that remain independent. *Journal of Corporate Finance*, *11*(5), 882–914.
Jensen, M. C. & Meckling, W. H. (2019). Theory of the firm: Managerial behavior, agency costs and ownership structure. In *Corporate Governance* (pp. 77–132). Gower.
Kiani, S. H., Mahmood, K., Khattak, U. F. & Munir, M. (2016). U patch antenna using variable substrates for wireless communication systems. *International Journal of Advanced Computer Science and Applications*, *7*(12).
Li, G. & Cui, G. (2011). A Improved Algorithms of Fuzzy Support Vector Machines. *Computer Measurement & Control*, *19*(4).
Ma, Y. & Guo, G. (2014). *Support vector machines applications* (Vol. 649). Springer.
Müller, K.-R., Smola, A. J., Rätsch, G., Schölkopf, B., Kohlmorgen, J. & Vapnik, V. (1997). Predicting time series with support vector machines. *International Conference on Artificial Neural Networks*, 999–1004.
Widodo, A. & Yang, B.-S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, *21*(6), 2560–2574.
Wu, L., Shao, Z., Yang, C., Ding, T. & Zhang, W. (2020). The impact of CSR and financial distress on financial performance—evidence from Chinese listed companies of the manufacturing industry. *Sustainability*, *12*(17), 6799.
Zhang, Z.-C., Wang, S.-T., Deng, Z.-H. & Chung, F.-L. (2012). A fast decision algorithm of support vector machine. *Control and Decision*, *27*(3), 459–463.

*Corresponding author