Fuzzy Support Vector Machine Using Function Linear Membership and Exponential with Mahanalobis Distance

Wiwi Widia Sukeiti[1], Sugiyarto Surono[2]

[1,2]Departement of Mathematics, Ahmad Dahlan University, Yogyakarta, Indonesia
wiwidiasukaeiti@gmail.com[1], sugiyarto@math.uad.ac.id[2]

ABSTRACT

Support vector machine (SVM) is one of effective biner classification technic with structural risk minimization (SRM) principle. SVM method is known as one of successful method in classification technic. But the real-life data problem lies in the occurrence of noise and outlier. Noise will create confusion for the SVM when the data is being processed. On this research, SVM is being developed by adding its fuzzy membership function to lessen the noise and outlier effect in data when trying to figure out the hyperplane solution. Distance calculation is also being considered while determining fuzzy value because it is a basic thing in determining the proximity between data elements, which in general is built depending on the distance between the point into the real class mass center. Fuzzy support *vector machine* (FSVM) uses Mahalanobis distances with the goal of finding the best hyperplane by separating data between defined classes. The data used will be going over trial for several dividing partition percentage transforming into training set and testing set. Although theoretically FSVM is able to overcome noise and outliers, the results show that the accuracy of FSVM, namely 0.017170689 and 0.018668421, is lower than the accuracy of the classical SVM method, which is 0.018838348. The existence of fuzzy membership function is extremely influential in deciding the best hyperplane. Based on that, determining the correct fuzzy membership is critical in FSVM problem.

——————————— ◆ ———————————

A. INTRODUCTION

Support vector machine (SVM) is a machine learning method developed by Vapnik in 1995 to predict, both classification and regression. SVM is included in the supervised learning technique. In addition, there are also unsupervised learning techniques, including clustering techniques. Several studies related to clustering were conducted by (Haryati et al., 2021; Surono, Haryati, et al., 2021) using one type of clustering, namely Fuzzy Subtractive Clustering. SVM theory is based on the principle of structural risk minimization (Vapnik, 1995). In many of its applications, SVM has proven to perform better than conventional machine learning and has built a reputation for its distinct approach to classification. That is the reason why this algorithm was introduced as an effective tool for solving classification problems (Lin & Wang, 2002).

Aside of mathematical foundation in theoretic statistic learning, SVM has shown a very competitive performance in a lot of real world application, for the example in text mining

(Manochandar & Punniyamoorthy, 2018), bioinformatic (Battineni et al., 2019)(Viloria et al., 2020), face recognition (Anzid et al., 2019), and graphic processing that has established SVM as one of the tool in machine learning and data mining, together with the other software and computation technic, for the example neural network and fuzzy system (Ladwani, 2018). SVM based algorithm, just like the other algorithm, has a drawback that is sensitive to noise and outlier (Mohammadi & Sarmad, 2019), (J. Liu, 2020), (Xiaokang et al., 2016b).

Uncertainty can be defined as the absence in class membership criteria that is defined sharply rather than random variable (J.ROSS, 2010). Theory and fuzzy group method can solve the uncertainty and the vagueness. This is extremely useful to tackle a complex real-world problem, including the problem in introducing prediction inference and control, and decision making. The combination of fuzzy group theory and classification technic in determining a class come in a handy to improve the ability of classifier generalization (Lin & Wang, 2002)(Mohammadi & Sarmad, 2019)(Inoue & Abe, 2001; Jiang et al., 2006; W. Liu et al., 2020; Richhariya & Tanveer, 2018; Wu, 2011).

According to (Lin & Wang, 2002), one of the criteria that is widely chosen to consider one point as outlier is the appropriate distance to the center of class. Euclidean distance is a general common distance. Euclidean distance has been applied in FSVM by (Surono, Nursofiyani, et al., 2021) for classification. But this distance is not an appropriate metric for data with correlation feature. As alternative method, Mahalanobis distance is often being used to measure the distance of every point to class center while also taking covariant into consideration (Mohammadi & Sarmad, 2019).

To lessen the noise and outlier effect, fuzzy membership function combined with SVM can give a lighter mass to the outlier data point. A new fuzzy membership function is being proposed to be formed using mixed kernel function on future space. Based on research (Lu et al., 2009), the results of the smallest training error and testing error are found in mixed kernels, where the training error is 0.00457893 and the testing error is 0.01598367. Three FSVM models were created and normal SVM was selected as a reference for comparison and it can effectively reduce noise impact and improve classification accuracy. The results showed that FSVM had the best classification performance compared to SVM (Xiaokang et al., 2016b). To tackle the outlier, An dan Liang (An & Liang, 2013) proposed a fuzzy support vector machine based on class scattering to solve classification problem. FSVM Interval type-2 is proposed to be able to reach the level of accurate classifier (Ekong et al., 2016).

To solve this problem, we propose the fuzzy support vector machine (FSVM) with different fuzzy membership value and using Mahalanobis distance to measure the distance of each point to the class center. The method used is especially targeted to improve SVM in lessening noise and outlier effect in data point.

B. METHODS
1. Support Vector Machine

Support Vector Machine (SVM) is a biner classification method to find the best hyperplane by separating defined data between classes. Support vector machine is a biner classification algorithm that maximizes margin around the hyperplane separator and minimizes training fault at the same time. For the example S is in n group training sample (Lu et al., 2009),

$$S = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}, \tag{1}$$

where $x_i \in R^n$ is a matrix with n dimension and $y_i \in \{-1, +1\}$. To find the separator area (symbolized by H) for data that can be separated in linier is

$$H = w^T x_i + b = 0 \tag{2}$$

with w defined as normal vector from hyperplane separator that is sized as big as p-dimension and b is scalar. If H_1 and H_2 are first and second classes barrier area, so:

$$H_1 = w^T x_i + b = +1$$
$$H_2 = w^T x_i + b = -1 \tag{3}$$

Every point on barrier area class (support vector) has the equation of:

$$y_i = w^T x_i + b \tag{4}$$

So that for every data point, applies:

$$w^T x_i + b = +1$$
$$w^T x_i + b = -1 \tag{5}$$

with requirement

$$y_i(w^T x_i + b) \geq 1, \ i = 1, 2, \dots, n \tag{6}$$

Hyperplane optimal can be obtained by followed primal optimization problem (J. Liu, 2020):

$$\min \frac{1}{2} \|w\|^2 \tag{7}$$

If the data training is unable to be separated by linier, slack variable ξ_i can be added as misclassification from difficult data training or noise. Slack variable addition will change the formula into (Xiaokang et al., 2016a):

$$min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i \tag{8}$$

with the requirement,

$$y_i(w^T x_i + b) \geq 1 - \xi_i; \ \ \xi_i \geq 0; \ 1 \leq i \leq n \tag{9}$$

Optimal hyperplane in (2) can be obtained by transforming the primal shape into the double shape of quadratic programming (QP). Equivalent double problem can be solved based on the condition of Kuhn-Tucker (KKT) Lagrange multiplier (Mohammadi & Sarmad, 2019),

$$max \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \tag{10}$$

with the requirement,

$$\sum_{i=0}^{n} \alpha_i y_i = 0; \ 0 \leq \alpha_i \leq C; \ 1 \leq i \leq n, \tag{11}$$

where $\alpha = (\alpha_i, \dots, \alpha_n)$ is a vector multiplier Lagrange. Finding α_i by solving the quadratic optimization problem so can found the result as follows:

$$w^* = \sum_{i=1}^{n} \alpha_i y_i x_i \qquad (12)$$

According to KKT condition, bias definition can be solved by (Mohammadi & Sarmad, 2019):

$$b^* = y_i - \sum_{i=1}^{n} \alpha_i^* x_i y_i \qquad (13)$$

also, can be solved for every supporting vector (based on the observation that the suitable α_i is bigger than zero). Sample point x_i is classified based on function sign in which explained as follows (Mohammadi & Sarmad, 2019).

$$f(x) = sign(w^T(x_i) + b) \qquad (14)$$

For non-linier data can be separated in feature room, kernel function $K(x_i, x_j) = \phi(x_i)\phi(x_j)$ is used to find the optimal hyperplane in higher dimension room, in which $\phi(x_i)$ is a non-linier mapping function. In research conducted by (Ningrum, 2018), of the three kernels used, the kernel with a polynomial function provides the highest classification accuracy value, with an accuracy value of 95.45%. Therefore, this research uses a polynomial function kernel, which can be stated as follows in Figure 1.

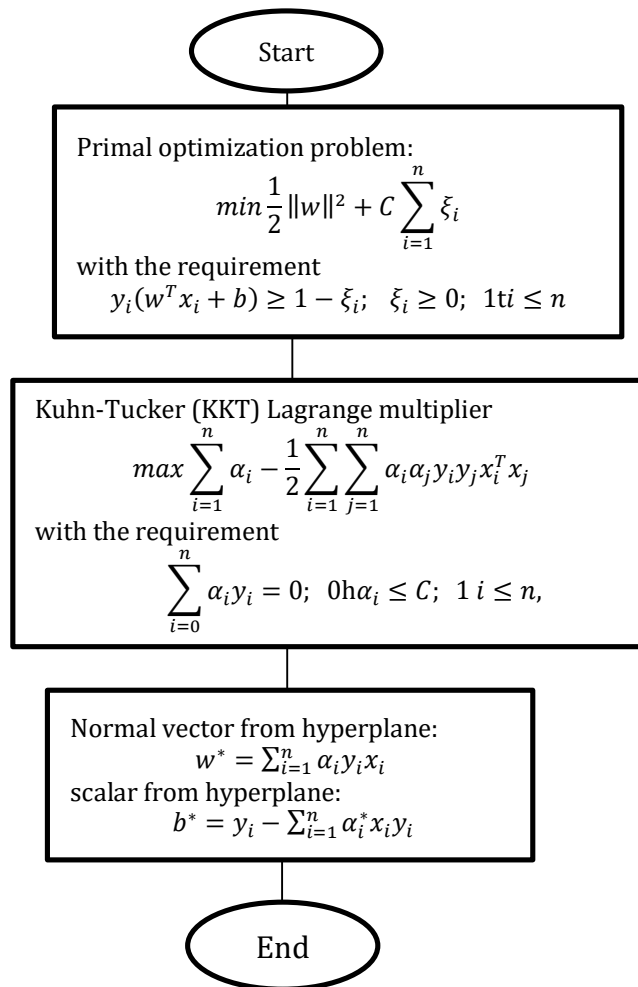$$K(x_i, x_j) = \left(x_i^T x_j + t\right)^l \qquad (15)$$

Figure 1. Flowchart SVM

2. Fuzzy Support Vector Machine

Even though SVM is an effective biner classification method, it is sensitive towards noise and outlier. Similar to SVM, the purpose of FSVM is to find the optimal separating hyperplane that is separating sample point into two classes with maximum margin. In the year of 2002, (Lin & Wang, 2002) extension method from SVM called Fuzzy Support Vector Machine (FSVM) was proposed. In FSVM, every data training point is given a fuzzy membership value symbolized by μ_i. The effect of outlier data point over classification accuracy can be lessened by giving a value to fuzzy membership because the outlier data point is unable to show the correct ownership on the defined class. Outlier data point is going to have a small fuzzy membership value. Usually, fuzzy membership function is designed based on the distance from suitable center class. These will minimalize the impact on error and on the forming of dividing area. For the example a S data set filled with training points (Mohammadi & Sarmad, 2019):

$$S = \{(x_1, y_1, \mu_1), (x_2, y_2, \mu_2), \dots, (x_n, y_n, \mu_n)\}$$

where every data training point $x_i \in R_n$ is a class in the label of $Y_i \in \{1, -1\}$ and fuzzy membership value, where $0 < \mu_i \leq 1$ for $i = 1, \ldots, n$. The optimal formula of hyperplane for this problem is

$$min \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} \xi_i \mu_i \tag{16}$$

with the requirement,

$$y_i \left(w^T x_i + b\right) \geq 1 - \xi_i; \quad \xi_i \geq 0; \quad 1 \leq i \leq n. \tag{17}$$

To solve the FSVM optimal problem, (4) is being transformed into following double problem by introducing Langrangian multiplier:

$$max \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \tag{18}$$

with the requirement,

$$\sum_{i=0}^{n} \alpha_i y_i = 0; \quad 0 \leq \alpha_i \leq \mu_i C; \quad 1 \leq i \leq n, \tag{19}$$

The only difference between fuzzy SVM with classic SVM is the upper limit α_i, which is obtained from multiplying both from parameter C and μ_i.

3. Mahalanobis Distance

Mahalanobis distance is a statistic method used to obtaining a data with certain distance over the data average so that the data distribution with patter over average value is obtained. Mahalanobis distance can be defined as the distance of two points involving covariant or correlation between the changer. Mahalanobis distance between two objects are being presented in the shape of vector and matrix. This distance is taking account of correlation between variable, because it is considered as the opposite of covariant matrix in calculation. It can be defined as (Mohammadi & Sarmad, 2019):

$$d = \sqrt{(x - \bar{x})^T S^{-1} (x - \bar{x})} \tag{20}$$

where x is a vector of p–dimension in R^p, $\bar{x}$ is a vector of a sample average and S^{-1} shows and inversion of covariant matrix. For the example, the used data is formed based on the amount of p variable and the amount of observation N. So that it is obtained the matrix form p×p written as follows:

$$S = \begin{bmatrix} cov(x_1, x_1) & cov(x_1, x_2) & \ldots & cov(x_1, x_p) \\ cov(x_2, x_1) & cov(x_2, x_2) & \ldots & cov(x_2, x_p) \\ cov(x_3, x_1) & cov(x_3, x_2) & \ldots & cov(x_3, x_p) \\ \vdots & \vdots & \ddots & \vdots \\ cov(x_p, x_1) & cov(x_p, x_2) & \ldots & cov(x_p, x_p) \end{bmatrix} \tag{21}$$

$$S = \begin{bmatrix} cov(x_1,x_1) & cov(x_1,x_2) & \dots & cov(x_1,x_p) \\ cov(x_2,x_1) & cov(x_2,x_2) & \dots & cov(x_2,x_p) \\ cov(x_3,x_1) & cov(x_3,x_2) & \dots & cov(x_3,x_p) \\ \vdots & \vdots & \ddots & \vdots \\ cov(x_p,x_1) & cov(x_p,x_2) & \dots & cov(x_p,x_p) \end{bmatrix}$$

where to the covariant value can be found by using the formula below:

$$S_{ij} = \frac{1}{N-1}\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)(x_{ij} - \bar{x}_j) \tag{22}$$

with $i = 1,2,\dots,N$, dan $j = 1,2,\dots,p$.

4. Membership Function

Fuzzy membership function in linier case. In FSVM, fuzzy membership is used to lessen the effect of outlier or noise and a different fuzzy membership function will give another result in fuzzy algorithm. Basically, the rule to determine the correct membership value to data point can depend on relative importance of date point to their own class. On this research, we consider two fuzzy membership functions given in (Lu et al., 2009).

Definition 1. μ_{lin} as FSVM 1 is referred as fuzzy linier membership and μ_{lin} can be defined as follows

$$\mu_{lin} = \begin{cases} \dfrac{1 - d_i}{(max(d_{i+}) + \delta)} & if\ y_i = 1 \\ \dfrac{1 - d_i}{(max(d_{i-}) + \delta)} & if\ y_i = -1 \end{cases} \tag{23}$$

in which δ is a small positive value used to avoid the value of μ_{lin} becoming zero.

Definition 2. μ_{exp} as FSVM 2 is referred as exponential fuzzy membership and μ_{exp} can be defined as

$$\mu_{exp} = \begin{cases} \dfrac{2}{1 + exp(\lambda d_{i+})} & if\ y_i = 1 \\ \dfrac{2}{1 + exp(\lambda d_{i-})} & if\ y_i = -1 \end{cases} \tag{24}$$

in which the parameter of $\lambda \in [0, 1]$ decides the steepness of disintegration.

$$d_i = \begin{cases} \sqrt{(x_i - \bar{x}_+)^T S^{-1}(x_i - \bar{x}_+)}, & if\ y_i = 1 \\ \sqrt{(x_i - \bar{x}_-)^T S^{-1}(x_i - \bar{x}_-)}, & if\ y_i = -1 \end{cases} \tag{25}$$

in which d_i is the calculation of Mahalanobis distance based on its class while also considering the inside covariant.

5. Dataset

The experiment is done with medical data extracted from UCI database repository in the form of diabetes risk. It is very probable that this real dataset will have several noise and outlier in a different amount. The group of data is composed of 520 people (328 males and 192 females, aged between 16 to 90 years old). The data is obtained from questionnaire that is directly given to the people who just diagnosed with diabetes or who haven't but showing symptoms. The data variables are x_1(Age), x_2(Sex), x_3(Polyuria), x_4(Polydipsia), x_5(Sudden weight loss), x_6 (Weaknes), x_7 (Polyphagia), x_8 (Genital thrush), x_9 (Visual blurring),

x_{10}(Itching), x_{11}(Irritability), x_{12}(Delayed healing), x_{13}(Partial presis), x_{14}(Muscle stiffness), x_{15}(Alopecia), x_{16}(Obesity), and y(Class).

C. RESULT AND DISCUSSION

In this section we will examine the performance of the membership function based on the Mahalanobis distance. To test the performance of our algorithm, FSVM using Mahalanobis distance will be compared with classic SVM. Diabetes data with the amount of data as much as 520 and the number of variables as many as 16 variables. The data is processed several times by modifying the amount of training data. The amount of training data used in this study is 70%, 80% and 90% where the number of positive class data and negative class data is shown in Table 1. While data that is not included in data training is defined as data testing.

Table 1. Data Group Composition

Data Training (%)	Data Training Positive Class	Data Training Negative Class	Data Testing (%)	Data Testing Positive Class	Data Testing Negative Class
70	138	226	70	62	94
80	160	256	80	40	64
90	183	285	90	17	35

Algorithm 1. **Fuzzy support vector machine using Mahalanobis distance**
Input:
Data training sample $\{(x_i, y_i), i = 1, 2, 3, \ldots, m\}$
Data testing sample $\{(x_j, y_j), j = 1, 2, 3, \ldots, n\}$
Output:
Label prediction of y_j from data $\{x_j, j = 1, 2, 3, \ldots, n\}$
Process:
1) Calculating the distance using (12) for data $\{(x_i, y_i), i = 1, 2, 3, \ldots, m\}$
2) Calculating fuzzy membership function using (10) or (11)
3) Determining the C parameter to calculate α^*(9), w^*(4) dan b^*(5) by using QP solver for data $\{(x_i, y_i), i = 1, 2, 3, \ldots, m\}$ that has been changed using polynomial kernel (7)
4) Use the decision function (6) with sample x_j and resulting in the last class label

Calculating membership function is the most important thing in FSVM. Before calculating the fuzzy membership function, measuring the distance is done first, the reason for this is to determine the approach size for each sample data over class data distribution center. The data used is processed using Python software. By using (12) so the distance is obtained as Shown in Table 2.

Table 2. Result of Mahalanobis Distance with Data Training Value Consecutively: 70%, 80% and 90%

x_i	Training Sample 70%	x_i	Training Sample 80%	x_i	Training Sample 90%
x_1	2.639358425	x_1	2.055891072	x_1	4.308982858
x_2	3.845592898	x_2	5.340175795	x_2	3.248246951
x_3	4.742628051	x_3	3.56063521	x_3	2.881856185
x_4	4.613615806	x_4	3.974708829	x_4	4.002542938
⋮	⋮	⋮	⋮	⋮	⋮
x_{364}	5.172826746	x_{416}	5.276751408	x_{468}	5.18299005

After obtaining the distance value, the next step is to calculate the fuzzy membership function using (10) to determine linier membership function and (11) to determine exponential membership function. The calculation result is showed in Table 3, Table 4 and Table 5 below:

Table 3. Calculation result using linier membership function

x_i	Training Sample 70%	x_i	Training Sample 80%	x_i	Training Sample 90%
x_1	0.664967575	x_1	0.751512912	x_1	0.457729376
x_2	0.511851705	x_2	0.354554942	x_2	0.6176725
x_3	0.425916232	x_3	0.54837909	x_3	0.637328343
x_4	0.414361127	x_4	0.519593312	x_4	0.496293782
⋮	⋮	⋮	⋮	⋮	⋮
x_{364}	0.343376529	x_{416}	0.33071176	x_{468}	0.347738586

**Table 4.** Calculation result using exponential membership function

x_i	Training Sample 70%	x_i	Training Sample 80%	x_i	Training Sample 90%
x_1	0.170132299	x_1	0.271674893	x_1	0.040542005
x_2	0.060883023	x_2	0.016225731	x_2	0.102015577
x_3	0.027622474	x_3	0.077988088	x_3	0.139093769
x_4	0.030970704	x_4	0.054385374	x_4	0.053075611
⋮	⋮	⋮	⋮	⋮	⋮
x_{364}	0.018838348	x_{416}	0.017170689	x_{468}	0.018668421

Table 5. Accuracy result of SVM, FSVM-1, and FSVM-2 by modifying data training and different parameter value of C

x_i	Training Sample 70%	x_i	Training Sample 80%	x_i	Training Sample 90%
x_1	0.170132299	x_1	0.271674893	x_1	0.040542005
x_2	0.060883023	x_2	0.016225731	x_2	0.102015577
x_3	0.027622474	x_3	0.077988088	x_3	0.139093769
x_4	0.030970704	x_4	0.054385374	x_4	0.053075611
⋮	⋮	⋮	⋮	⋮	⋮
x_{364}	0.018838348	x_{416}	0.017170689	x_{468}	0.018668421

Accuration of classification in Table 2 show that the result of SVM classic. The accuracy classification result shown in Table 2 explain that the performance of SVM classic has a better accuracy on every data training comparison than FVSM that has been modified with Mahalanobis distance. However, the research conducted by (Mohammadi & Sarmad, 2019)

showed the best performance was with FSVM, but the data used in this study was different from the data used in this study. Thus, it is safe to say that the method of giving fuzzy membership value for data training will play an important role when FSVM method is applied in a group of data. In another word, an incorrect method in determining membership value can worsen the performance of the model compared to SVM classic model. The accuracy result in Table 2 is presented in a form of graphic line as shown in Figture 2.
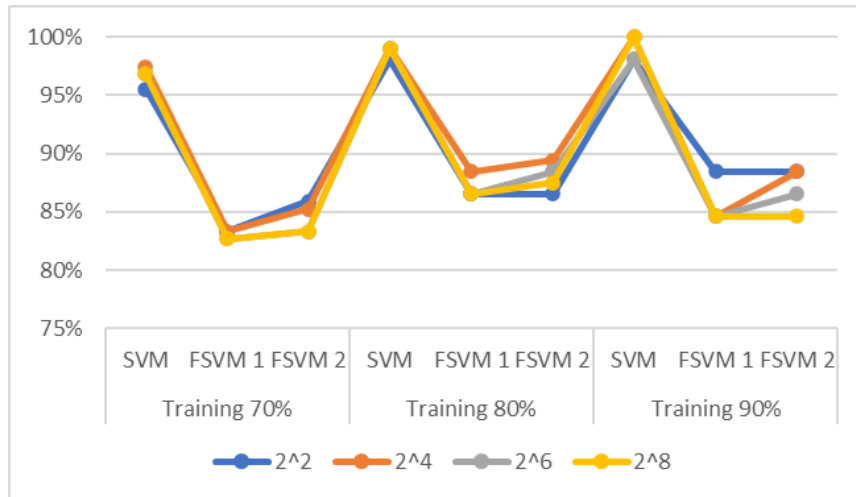
Figure 2 Accuracy result of SVM, FSVM-1, dan FSVM-2 by modifying
data training and different parameter value of C

Figure 2 is a comparison of the accuracy of SVM, FSVM-1, and FSVM-2. Accuracy results using training data with 70%, 80%, and 90%. Based on Figure 1, the accuracy results for SVM have better accuracy results compared to FSVM modified with Mahalanobis distance.

D. CONCLUSION AND SUGGESTIONS

In this essay, we suggest FSVM method with consideration of Mahalanobis distances. On this research, we propose the FSVM method while also considering Mahalanobis distance. The proposed distance has the goal to characterize the distance of different data point to the real group of class. FSVM will determine the fuzzy membership value for every input point so that another input point can give a different contribution in decision making. By determining different fuzzy, that is linier membership function and exponential membership function, we are able to observe the FSVM performance in solving different kind of problem easily. This will enrich our knowledge on SVM application that is developed with fuzzy. In this study, the data used were 520 data. The results of data accuracy from the FSVM method are 0.017170689 and 0.018668421. While the results of the SVM accuracy of 0.018838348. This shows that adding fuzzy membership does not always guarantee better results than classical SVM. This also means that the act of adding the correct fuzzy membership is very important in generalizing the performance of FSVM in classification problems.

REFERENCES

An, W., & Liang, M. (2013). Fuzzy support vector machine based on within-class scatter for classification problems with outliers or noises. *Neurocomputing*, *110*, 101–110. https://doi.org/10.1016/j.neucom.2012.11.023

Anzid, H., Goic, G. Le, Bekkari, A., Mansouri, A., & Mammass, D. (2019). Multimodal Images Classification using Dense SURF, Spectral Information and Support Vector Machine. *Procedia Computer Science*, *148*, 107–115. https://doi.org/10.1016/j.procs.2019.01.014

Battineni, G., Chintalapudi, N., & Amenta, F. (2019). Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). *Informatics in Medicine Unlocked*, *16*(May), 100200. https://doi.org/10.1016/j.imu.2019.100200

Ekong, U., Lam, H. K., Xiao, B., Ouyang, G., Liu, H., Chan, K. Y., & Ling, S. H. (2016). Classification of epilepsy seizure phase using interval type-2 fuzzy support vector machines. *Neurocomputing*, *199*, 66–76. https://doi.org/10.1016/j.neucom.2016.03.033

Haryati, A. E., Surono, S., & Suparman, S. (2021). Implementation of Minkowski-Chebyshev Distance in Fuzzy Subtractive Clustering. *EKSAKTA: Journal of Sciences and Data Analysis*, *2*(2), 1–7. https://doi.org/10.20885/EKSAKTA.vol2.iss1.art

Inoue, T., & Abe, S. (2001). Fuzzy support vector machines for pattern classification. *Proceedings of the International Joint Conference on Neural Networks*. https://doi.org/10.1109/ijcnn.2001.939575

J.ROSS, T. (2010). *Fuzzy Logic With Engineering Application*.

Jiang, X., Yi, Z., & Lv, J. C. (2006). Fuzzy SVM with a new fuzzy membership function. *Neural Computing and Applications*, *15*(3–4), 268–276. https://doi.org/10.1007/s00521-006-0028-z

Ladwani, V. M. (2018). Support vector machines and applications. *Computer Vision: Concepts, Methodologies, Tools, and Applications*, 1381–1390. https://doi.org/10.4018/978-1-5225-5204-8.ch057

Lin, C. F., & Wang, S. De. (2002). Fuzzy support vector machines. *IEEE Transactions on Neural Networks*, *13*(2), 464–471. https://doi.org/10.1109/72.991432

Liu, J. (2020). Fuzzy support vector machine for imbalanced data with borderline noise. *Fuzzy Sets and Systems*, *1*, 1–10. https://doi.org/10.1016/j.fss.2020.07.018

Liu, W., Ci, L. L., & Liu, L. P. (2020). A new method of fuzzy support vector machine algorithm for intrusion detection. *Applied Sciences (Switzerland)*, *10*(3). https://doi.org/10.3390/app10031065

Lu, Y. L., Li, L., Zhou, M. M., & Tian, G. L. (2009). A new fuzzy support vector machine based on mixed kernel function. *Proceedings of the 2009 International Conference on Machine Learning and Cybernetics*, *1*(July), 526–531. https://doi.org/10.1109/ICMLC.2009.5212552

Manochandar, S., & Punniyamoorthy, M. (2018). Scaling feature selection method for enhancing the classification performance of Support Vector Machines in text mining. *Computers and Industrial Engineering*, *124*(July), 139–156. https://doi.org/10.1016/j.cie.2018.07.008

Mohammadi, M., & Sarmad, M. (2019). Robustified distance based fuzzy membership function for support vector machine classification. *Iranian Journal of Fuzzy Systems*, *16*(6), 191–204. https://doi.org/10.22111/ijfs.2019.5028

Ningrum, H. C. S. (2018). Perbandingan Metode Support Vector Machine (SVM) Linear, Radial Basis Function (RBF), dan Polinomial Kernel dalam Klasifikasi Bidang Studi Lanjut Pilihan Alumni UII. *Tugas Akhir Statistika Universitas Islam Indonesia*, 1–90.

Richhariya, B., & Tanveer, M. (2018). A robust fuzzy least squares twin support vector machine for class imbalance learning. *Applied Soft Computing Journal*, *71*, 418–432. https://doi.org/10.1016/j.asoc.2018.07.003

Surono, S., Haryati, A. E., & Eliyanto, J. (2021). *An Optimization of Several Distance Function on Fuzzy Subtractive Clustering* (Antonio J. Talloon-Ballestores (ed.)). IOS Press.

Surono, S., Nursofiyani, T., & Haryati, A. E. (2021). Optimization of Fuzzy Support Vector Machine (FSVM) Performance by Distance-Based Similarity Measure Classification. *HighTech and Innovation Journal*, *2*(4), 285–292. https://doi.org/10.28991/hij-2021-02-04-02

Vapnik, V. (1995). *The Nature of Statistical Learning*. https://ci.nii.ac.jp/naid/10020951890

Viloria, A., Herazo-Beltran, Y., Cabrera, D., & Pineda, O. B. (2020). Diabetes Diagnostic Prediction Using Vector Support Machines. *Procedia Computer Science*, *170*, 376–381. https://doi.org/10.1016/j.procs.2020.03.065

Wu, Q. (2011). Fuzzy robust ν-support vector machine with penalizing hybrid noises on symmetric triangular fuzzy number space. *Expert Systems with Applications*, *38*(1), 39–46. https://doi.org/10.1016/j.eswa.2010.06.003

Xiaokang, D., Lei, Y., Jianping, Y., & Zhaozhong, Z. (2016a). Optimization and analysis on fuzzy SVM for object classification. *Open Cybernetics and Systemics Journal*, *10*(6), 155–162. https://doi.org/10.2174/1874110X01610010155

Xiaokang, D., Lei, Y., Jianping, Y., & Zhaozhong, Z. (2016b). Optimization and analysis on Fuzzy SVM for targets classification in forest. *Open Cybernetics and Systemics Journal*, *10*(6), 155–162. https://doi.org/10.2174/1874110X01610010155