C-52

by Armin Lawi

Proceeding of International Conference on Mathematics, Statistics, Computer Sciences, and Mathematics Education (*ICMSCSME*) 2015
ISBN 978-602-72198-2-3

THE COMPARISON MULTICLASS CLASSIFICATION USING SUPPORT VECTOR MACHINE

MS Rahayu, A Lawi, SA Thamrin
Mathematics Department Hasanuddin University
monikasrahayu@gmail.com

Abstract

Support Vector Machine (SVM) is a very popular large data classifications techinique in the field of data mining. The basic idea of SVM is to find the linear separation that divides data into two classes. An interesting further research is develop an algorithm to classify SVM data more than two classes (multiclass). This research reviewed about the SVM multiclass comparison between One Against All method and One Against One method. Experiment in assessing the accuracy of the two methods is performed on the two data sets. The research result obtained that the accuracy of One Against One method is higher than One Against All method.

Keyword: Data mining, multiclass, one against all method, one against one method, accuracy level.

1. Introduction

Data mining is a term used to discover hidden knowledge in in a database .Data mining is spring otomatik process that uses statistical techniques , mathematics , artificial intelligence , and machine learning for extracting identify information and knowledge potential of useful and beneficial stored in large databases [8].

Support vector machines (svm) is a technique classifications algorithm very popular in the field of mining data. SVM developed by boser, guyon, vapnik, and first presented in 1992 in annual workshop on computational learning theory. SVM aims to minimize upper limit generalization through maximize error margin between hyperplane to separate data. Which is in trouble linear problems. Because the initial concept of svm only to deal with classifications the two so svm cannot be applied to issue multiclass thus developed metode-metode the decision to solve this problem. These metode include one against all, one against one, directed acyclic graph support vecctor machine, and others.

The research classifying data with multiclass in svm. Experiment in assessing the accuracy of classifying data in a large data with multiclass svm.

2. Literature

2.1. Big Data

Technique data mining often applied to data on a large scale [6]. Big data is data enormous and diverse in processing will find it difficult if they use the method or application statistika [7].

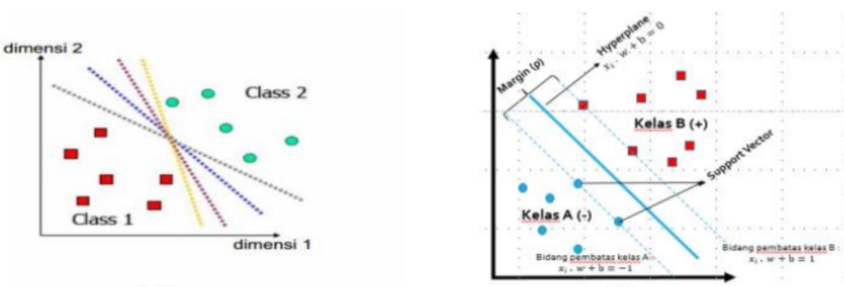
2.2 Data Mining

Data mining is a term used to discover hidden knowledge in in a database .Data mining involves the spring otomatik who uses the technique statistics , mathematics , artificial intelligence , and machine learning for extracting and identify information potential knowledge and useful been stored in large database [8]. Some techniques in literature data mining include the association rule mining, clustering, classification, neural network, genetic algorithm and others [6].

2.3. Support Vector Machine

The basic idea of of the svm is to find the barrier the separate data into the two class in a room features (features space) [5]. SVM developed to issue classifications while to solve problems regression known as support factor regression (SVR). The characteristic in general svm [5]:

1. In principle svm is linear classifier.
2. Pattern recognition
3. Applying the structural risk minimization (srm).
4. The principle of svm basically only unable to handle classifications two classes.



(a) (b)

Picture 2.1. (a) The alternative of fields (b) the best alternative of field with the big margin (m) [5]

[1] said the similarities fields to every class is:

$$x_i.\boldsymbol{w} + b \geq +1, \quad \text{untuk } y_i = +1$$

$$x_i.\boldsymbol{w} + b \leq -1, \quad \text{untuk } y_i = -1. \qquad (2.1)$$

Meanwhile, margin (m) or the distance from both the field delimiter in equation (2.1) can be written as follows:

$$m = \frac{|x_i.w + b|}{||\boldsymbol{w}||} + \frac{|x_j.w + b|}{||\boldsymbol{w}||} = \frac{1}{||\boldsymbol{w}||} + \frac{1}{||\boldsymbol{w}||} = \frac{2}{||\boldsymbol{w}||} \ . \qquad (2.2)$$

Best interface is the interface that has the greatest margin. Therefore, the margin (m) in equation (2.2) should be maximized but still consider the equation (2.1). Equation (2.1) and (2.2) can digeneraliasasi be an optimization problem with the objective function as follows:

minimize $\frac{1}{2}||\boldsymbol{w}||^2$

ISBN 978-602-72198-2-3

with constraints $y_i(x_i.\boldsymbol{w}) + b \geq 1$, for each $i = 1, \dots, m$ (2.3)

Equation (2.3) will be easier if it is brought into function Lagrange. Lagrange function is formed by multiplying the constraint functions by Lagrange multipliers (α) is then deducted the origin function . Lagrange functional form of equation (2.3) are :

$$L\,(w, b, \alpha) = \frac{1}{2}\,\|\boldsymbol{w}\|^2 - \sum_{i=1}^{m} \alpha_i\,[y_i(x_i.\boldsymbol{w} + b) - 1]. \qquad (2.4)$$

α_i variable called Lagrange multipliers , which is a constant that must be searched to satisfy the equation (2.4). Equation (2.4) should be minimized to w and b and should be maximized to variable α by finding the first derivative of the function L (w , b , α) to variable w and b and equated to 0. This process is described in the following mathematical conditions :

1. Conditions 1

$$\frac{\partial L(w, b, \alpha)}{\partial b} = 0$$

2. Conditions 2

$$\frac{\partial L(w, b, \alpha)}{\partial w} = 0$$

The application of condition 1 in equation (2.4) will produce

$$\sum_{i=1}^{m} \alpha_i y_i = 0 \qquad (2.5)$$

The application of condition 2 in equation (2.4) will produce

$$w = \sum_{i=1}^{m} \alpha_i y_i x_i \qquad (2.6)$$

Equation (2.4) is difficult to know the solution because there are elements w and b are unknown. This causes primal problem of equation (2.4) should be brought into the form of a dual.

Dual form of equation (2.9) are :

Max $\sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i\, \alpha_j\, y_i y_j x_i^T x_j$

with constraints

$$\sum_{i=1}^{m} \alpha_i y_i = 0 \qquad (2.7)$$

$$\alpha_i \geq 0, i = 1, \dots, m.$$

Equation (2.17) is used in the training process to get the value of α . After getting the value of α , according to equation (2.10) it will be found the value of w. After getting the value w, the value of b (Boswell , 2002) are :

$$b = \frac{-((x_i.\boldsymbol{w})^- + (x_i.\boldsymbol{w})^+)}{2}\ .$$

Because the values of b and w have known then dividing function (f (x)) to determine the class of a data x are :

$$f(x) = x.\boldsymbol{w} + b, \qquad \text{where } x = \begin{cases} \text{class} - 1 & \text{if } f(x) < 0 \\ \text{class} + 1 & \text{if } f(x) > 0 \end{cases}. \qquad (2.8)$$

2.4. Kernel Method

Linear SVM can be converted into a non - linear SVM using kernel . This method works by mapping the input data into a feature space whose dimensions are higher. Expected results of the mapping input data to a separate space will feature a linear manner so as to look for the optimal hyperplane. Kernel for SVM method was introduced in 1992 by Boser , Guyon , and Vapnik [2] . In kernel methods , the data is mapped by the function $\varphi: x \rightarrow \varphi(x)$ to the space features a higher dimension.

Because the transformation φ is generally unknown , and very difficult to understand it easily , then the calculation of the dot product can be replaced with the kernel function $K(x_i . x_j)$ which implicitly defines the transformation of φ. This is called Kernel Trick, which formulated

$$K(x_i . x_j) = \varphi(x_i) . \varphi(x_j)$$

$$f(\varphi(x)) = \varphi\ (x)\ .\ \boldsymbol{w} + b$$

$$= \sum_{i=1}^{m} a_i y_i \langle \varphi(x_i) . \varphi(x_j) \rangle + b$$

$$= \sum_{i=1}^{m} a_i y_i K\langle x_i . x_j \rangle + b$$

There are some kernel functions that are often used in the literature , among others, the following SVM [3]:

a. Linear Kernel $K(x_i . x_j) = x_i^T x_j$

b. Radial Basis Function (RBF) Kernel $K(x_i . x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$

c. Polynomial Kernel $K(x_i . x_j) = (\gamma x_i^T x_j + r)^2, \gamma > 0$

d. Sigmoid Kernel $K(x_i . x_j) = \tanh(\rho x_i^T x_j + r)^2$

2.5. Multiclass Support Vector Machine

SVM the first time introduced by vapnik , can only classifying data in two a class (classifications binary). But, further research to develop svm so that it can classifying data over two classes, continued. There are two options for implementing multiclass SVM is to combine several binary SVM or combine all of the data that consists of multiple classes into a form of optimization problems . However , the latter approach optimization problems to be solved much more complicated.

2.6. Metode One Against All

SVM models built binary k (k is the number of classes). Each model of the i-th classification trained to use all the data, to seek solutions to problems. For example, there are problems with the classification of grade 4 pieces . For training to use 4 pieces of binary SVM as shown in Table 2.1 and their use in classifying new data can be viewed at Picture 2.5 [3].

$$\min_{w^i, b^i, \xi_j^i} \frac{1}{2}(w^i)^T w^i + C \sum_{j=1}^{l} \xi_j^i ,$$

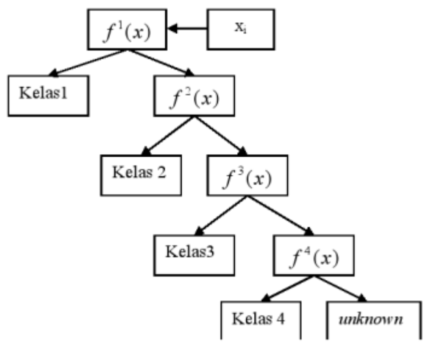
$$(w^i)^T \Phi(x_j) + b^i \geq 1 - \xi_j^i , \quad \text{jika } y_j = \text{i}$$
$$(w^i)^T \Phi(x_j) + b^i \leq 1 - \xi_j^i , \quad \text{jika } y_j \neq \text{i}$$

$$\xi_j^i \geq 0, j = 1, \dots, l$$

Proceeding of International Conference on Mathematics, Statistics, Computer Sciences, and Mathematics Education (*ICMSCSME*) *2015*
ISBN 978-602-72198-2-3

Table 2.1. Example of classification method one against all for 4 classes binner

$y_i = 1$	$y_i = -1$	Hipotesis
Kelas 1	Bukan kelas 1	$f^1(x) = (w^1)x + b^1$
Kelas 2	Bukan kelas 2	$f^2(x) = (w^2)x + b^2$
Kelas 3	Bukan kelas 3	$f^3(x) = (w^3)x + b^3$
Kelas 4	Bukan kelas 4	$f^4(x) = (w^4)x + b^4$



Picture 2.5 Example of a classification method one against all for 4 classes

2.7. Metode One Against One

Built a number $\frac{k(k-1)}{2}$ pieces of binary classification model (k is the number of classes). Each classification models trained on the data from the two classes. For classroom training data from the i-th and j-th class , made the search for solutions for the problem of binary classification as follows [3]:

$$\min_{w^{ij}, b^{ij}, \xi_t^{ij}} \frac{1}{2}(w^{i,j})^T w^{i,j} + C \sum_{j=1}^{t} \xi_t^{ij},$$

$$(w^{i,j})^T \Phi(x_j) + b^{i,j} \geq 1 - \xi_t^{ij}, \quad \text{jika } y_t = \text{i}$$

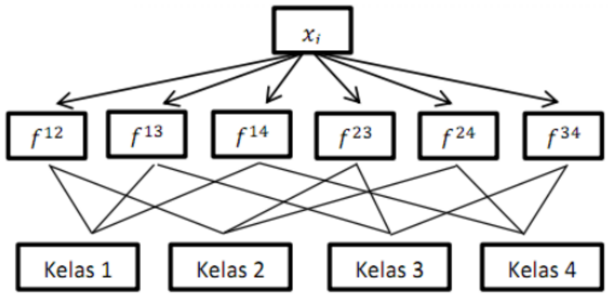
$$(w^{i,j})^T \Phi(x_j) + b^{i,j} \leq -1 + \xi_t^{ij}, \quad \text{jika } y_t \neq \text{i}$$

$$\xi_t^{ij} \geq 0$$

Tabel 2.2 Example of classification method one against one for 4 classes binner

$y_i = 1$	$y_i = -1$	Hipotesis
Kelas 1	Kelas 2	$f^{12}(x) = (w^{12})x + b^{12}$
Kelas 1	Kelas 3	$f^{13}(x) = (w^{13})x + b^{13}$
Kelas 1	Kelas 4	$f^{14}(x) = (w^{14})x + b^{14}$
Kelas 2	Kelas 3	$f^{23}(x) = (w^{23})x + b^{23}$
Kelas 2	Kelas 4	$f^{24}(x) = (w^{24})x + b^{24}$
Kelas 3	Kelas 4	$f^{34}(x) = (w^{34})x + b^{34}$

Proceeding of International Conference on Mathematics, Statistics, Computer Sciences, and Mathematics Education (*ICMSCSME*) 2015
ISBN 978-602-72198-2-3



Picture 2.6 Example of a classification method one against one for 4 classes

2.8. Performance Evaluation Methods of Classification

Accuration : Confusion Matrix

Performance methods can be seen from the level error is generated . The level of error in the classification can be seen in the confusion matrix .

Confusion Matrix for two class		Prediction	
		Positif	Negatif
Aktual	Positif	True Positif (TP)	False Negatif (FN)
	Negatif	False Positif (FP)	True Negatif (TN)

3. Results And Discussion

3.1 Sources and Types of Data

The data used in this study consists of two datasets with the following details:

1. http://archive.ics.uci.edu/ml/datasets/Iris for the data type of plant iris
2. http://www.grappa.univlille3.fr/~torre/Recherche/Experiments/Datasets/#cmc for the data about the method of choice for contraception

Dataset 1 consists of three classes in the form of the types of plants that iris setosa class , class versicolor , and virginica class . Attributes for dataset 1 totaled four sepals are long , wide sepals, petal length , petal width is measured in centimeters . Dataset 2 is the preferred method of contraception consists of three classes in the form of the first class is no method is used , the second class is a long-term method and the third class is a short-term method . Samples for dataset 1 totaled 150 items while the second dataset totaling 1473 items.

3.2 Multiclass classification method on Support Vector Machine

SVM classification methods require training and testing process . According to [4] , a comparison between the data used for training and testing is not binding or suit of researchers , but commonly used ratio of 70 % -30 % . As many as 70 % of the total data used for the training process , while the remaining 30 % used in the testing process . This leads to as many as 105 items of datasets 1 and 1032 items from dataset 2 is used for the training of the SVM method , while the remaining 45 items of

ISBN 978-602-72198-2-3

datasets 1 and 441 items from dataset 2 used in the process of testing the accuracy of the SVM method using One Against All and One Against One. The accuracy (λ) is measured by the following formula:

$$\lambda = \frac{C}{N} * 100\%$$

where C is the total prediction right and N is the total data to be tested . Corresponding workflow, SVM classification process begins with the training process to obtain the value of α . Matlab 2008 used in the training process of SVM . After the value of α is known, then the next process looking for value w and b . The value of w and b are used to compile the function of dividing the data classification with SVM method . W and b values are different for each kernel, the parameters used W and b values based on the type of kernel and its parameters for datasets 1 and 2 can be seen in Table 3.1 and Table 3.2:

Table 3.1. w and b values based on the type of kernel in dataset 1 and 2 for methods One Against All

ISBN 978-602-72198-2-3

The type of kernel		w		b	
		Dataset 1	*Dataset* 2	*Dataset* 1	*Dataset* 2
RBF	α = 0.05	5.6093 -0.2886 5.9066 -3.3099 -2.5168 -0.7939 6.5532 -0.6965 7.2372 6.0469 -0.8230 6.8458	8.1309 8.6092 3.6954 -0.4558 0.5172 -1.2433 5.2611 2.5008 2.2866	0.2574 -0.4847 -0.2607	-0.1133 -0.6255 -0.5674
Polynomia l	Orde 3	0.0247 0.0744 -0.0325 -0.0244 -0.0214 -0.0289	0.6379 0.2363 -0.0030 -0.1137 0.0362 0.0091	0.9849 1.4855	-0.8790 -1.1987
		0.0408 -0.0606 0.1350 0.0314 -0.1062 0.1639	0.0393 -0.0620 -0.0608	-1.4540	-0.0726
Linear		0.3072 -0.1928 -0.0343 -0.3295 -0.4853 -0.2711 0.6115 0.7255 1.4822 0.5865 -0.5614 1.4197	0.0500 0.0455 0.0109 0.0796 0.0244 0.0243 0.1473 0.1533 -0.0052	0.8704 -0.3763 -1.900	-0.5413 -1.1998 -0.3043

Table 3.2. w and b values based on the type of kernel in dataset 1 and 2 for methods One Against One

The type of kernel		w		b	
		Dataset 1	*Dataset* 2	*Dataset* 1	*Dataset* 2
RBF	α = 0.05	-2.6216 -2.7233 -2.2710 2.4807 2.2910 -1.1437 -3.3493 -3.2780 -4.4062 -3.3238 -3.1681 -4.4267	-3.2046 -4.1771 -0.1342 0.2242 -0.3431 -0.7419 -3.2527 -1.9649 2.0887	-0.0501 -0.0377 -0.0742	0.2781 0.2229 -0.0489
Polinomial	Orde 3	-0.0350 -0.0489 0.1060 0.0929 0.0529 0.0043 -0.1189 -0.1087 -0.3362 -0.1110 -0.1079 -0.4751	-0.0672 -0.0217 0.0592 0.0019 -0.0091 -0.0100 -0.1152 -0.0234 0.1103	-0.2626 -0.4173 -0.1419	-0.2450 -0.3913 -0.1040
Linear		-0.1902 -0.2060 0.1548 0.2926 0.2011 0.1326 -0.3732 -0.3983 -0.7199 -0.3909 -0.4107 -0.6881	-0.1495 -0.0880 0.0431 0.0204 -0.0534 -0.0511 -0.2788 -0.1396 0.1378	-0.0647 -0.0665 -0.0496	0.0281 0.0030 -0.0051

(Sumber: Data olahan peneliti)

3.3 Classification error

Classification error is a state where an item of data categorized into a particular class but SVM classification model identifies the data item into another class.

ISBN 978-602-72198-2-3

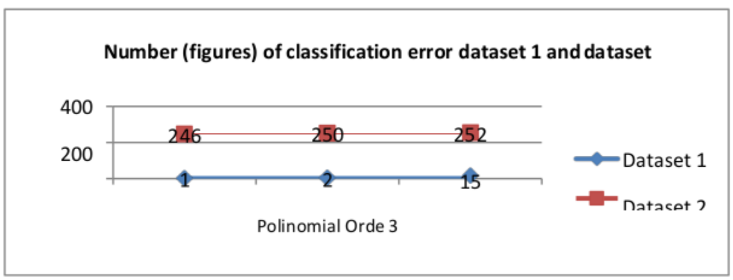
Classification error occurs in the testing process. Number of classification error for testing in this study varies according to the type of kernel used .

Table 3.1 Number (figures) of classification error dataset 1 and dataset 2 for each kernel using One Against All.

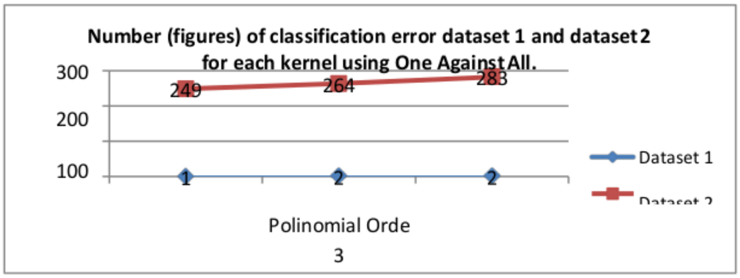
Numb.	The type of kernel		classification error	
			Dataset 1	*Dataset 2*
1	RBF	α = 0.05	**1**	**246**
2	Polynomial	Orde 3	2	250
3	Linear		15	252

Table 3.2 Number (figures) of classification error dataset 1 and dataset 2 for each kernel using One Against One.

Numb.	The type of kernel		classification error	
			Dataset 1	*Dataset 2*
1	RBF	α = 0.05	**1**	**249**
2	Polynomial	Orde 3	2	264
3	Linear		2	283



Picture 3.3 Graph amount of misclassification of datasets 1 and 2 for each kernel using One Against All



Picture 3.4 Graph amount of misclassification of datasets 1 and 2 for each kernel using One Against One.(Source : Data processed researchers)

3.4 Accuracy levels

The level of accuracy obtained from the amount of data to the appropriate class predicted by the model SVM. Accuracy rate in this study is different for every kernel used . The level of accuracy of the method One Against All can be seen in Table 3.5 and Figure 3.3 , while the number of misclassified One Against One method is shown in Table 3.6 and Figure 3.4 below :

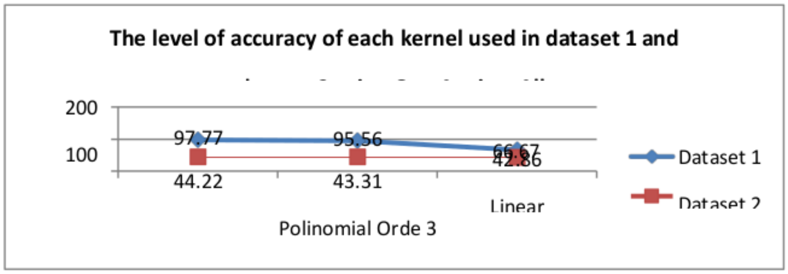
Table 3.5. The level of accuracy of each kernel used in dataset 1 and dataset 2 using One Against All.

Numb	The type of kernel		accurations (%)	
			Dataset 1	*Dataset 2*
1	RBF	α = 0.05	**97.77**	**44.22**
2	Polinomial	Orde 3	95.56	43.31
3	Linear		66.67	42.86

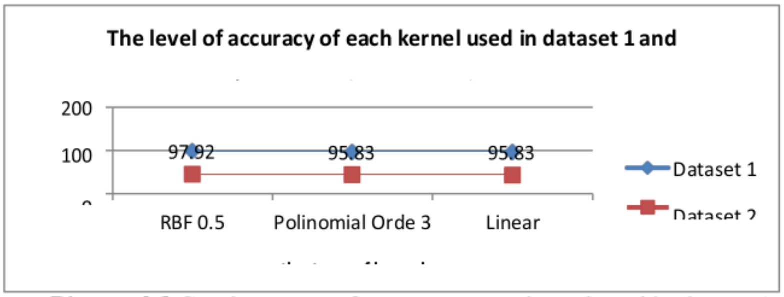
Table 3.6. The level of accuracy of each kernel used in dataset 1 and dataset 2 using One Against One

Numb	The type of kernel		accurations (%)	
			Dataset 1	*Dataset 2*
1	RBF	α = 0.05	**97.92**	**49.18**
2	Polynomial	Orde 3	95.83	46.12
3	Linear		95.83	42.24

ISBN 978-602-72198-2-3



Picture 3.7 Graph amount of accuracy every kernel used in dataset 1 and dataset 2 using One Against All



Picture 3.8 Graph amount of accuracy every kernel used in dataset 1 and dataset 2 using One Against One (Source : Data processed researchers).

4. Conclusions and Suggestions

4.1 Conclusion

Based on the findings in the previous chapter , it can be deduced as follows:

1. Classification with SVM multiclass method begins with training to obtain a function separator that divides the data into each class . After finding the dividing function, the next process is to classify the test data to see the level of SVM model accuracy.
2. Based on the results in Tables 3.5 and 3.6 , methods One Against One is a higher degree of accuracy than methods One Against All . Matters affecting the level of misclassification depending on the type of data and the selection of the kernel and its parameters . Selection of the proper parameters will result in fewer misclassification.
3. Gaussian kernel has a level of accuracy that is best among other kernels such as polynomial or linear for the data slice and data usage methods of contraception.

4.2 Suggestion

Some of the things research for the continuation of this study was to test the multiclass another method to compare the level of accuracy of the method and the method of One Against All and One Against One.

ISBN 978-602-72198-2-3

REFERENCES

[1] Burges, Christopher J.C. 1998. *A Tutorial on Support Vector Machine for Pattern Recognition*. Kluwer Academic Publisher, Boston, 1-43.

[2] Christianini, N, dan Taylor. 2000. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.Cambridge : University Press.

[3] Hsu, Chih-Wei dkk. 2010. A Practical Guide to Support Vector Classification. Taipei: National Taiwan University.

[4] Keller, F. 2013. Connectionist and Statistical Language Processing. Saarbrücken : Saarland University.

[5] Nugroho, A.M. 2008. Pengantar Support Vector machine. Pusat Teknologi Informasi dan komunikasi BPP Teknologi.

[6] Santoso, H. 2012. Analisis dan Prediksi pada Perilaku Mahasiswa Diploma untuk Melanjutkan Studi ke Jenjang Sarjana Menggunakan Teknik Decision Tree dan Support Vektor Machine. Medan: Universitas Sumatera Utara.

[7] Snijders, C. 2012. “Big data”: Big Gaps of Knowledge in the Field of Internet Science.International Journal of Internet Science, 7(1):1-5.

[8] Turban, dkk. 2005. Decision Support Sistems and Intelligent Sistems (Sistem Pendukung Keputusan dan Sistem Cerdas).Yogyakarta : Penerbit Andi.

C-52

ORIGINALITY REPORT

12%	4%	9%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Armin Lawi, Yudhi Adhitya. "Classifying Physical Morphology of Cocoa Beans Digital Images using Multiclass Ensemble Least-Squares Support Vector Machine", Journal of Physics: Conference Series, 2018 Publication	1%
2	Nova Ayu Rizanti, Arini, Anif Hanifah Setyaningrum. "Colon detection using Principal Component Analysis (PCA) and Support Vector Machine (SVM)", 2016 4th International Conference on Cyber and IT Service Management, 2016 Publication	1%
3	www.coursehero.com Internet Source	1%
4	Marsujitullah, Zahir Zainuddin, Salama Manjang, Aksan Surya Wijaya. "Rice Farming Age Detection Use Drone Based on SVM Histogram Image Classification", Journal of Physics: Conference Series, 2019 Publication	1%

5 Nuranisah, Syahril Efendi, Poltak Sihombing. "Analysis of algorithm support vector machine learning and k-nearest neighbor in data accuracy", IOP Conference Series: Materials Science and Engineering, 2020
Publication
1%

---

6 Chairunissa Atimas Nurahmad, Mirna Adriani. "Identifying traditional music instruments on polyphonic Indonesian folksong using mel-frequency cepstral coefficients (MFCC)", Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia - MoMM '12, 2012
Publication
1%

---

7 Submitted to Universiti Teknologi Malaysia
Student Paper
1%

---

8 dbkit.bibliothek.kit.edu
Internet Source
1%

---

9 Submitted to Higher Education Commission Pakistan
Student Paper
1%

---

10 Submitted to University of Edinburgh
Student Paper
1%

---

11 repository.uinsu.ac.id
Internet Source
1%

---

12 repository.unhas.ac.id
Internet Source

		1%
13	eprints.umm.ac.id Internet Source	1%
14	www.ijcttjournal.org Internet Source	1%
15	Tarjo, , and Nurul Herawati. "Application of Beneish M-Score Models and Data Mining to Detect Financial Fraud", Procedia - Social and Behavioral Sciences, 2015. Publication	1%
16	jeecs.ubhara.ac.id Internet Source	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		