Jurnal Matematika MANTIK
Volume 5, Nomor 2, October 2019, pp. 90-99

ISSN: 2527-3159 (*print*) 2527-3167 (*online*)

Comparison of Kernel Function on Support Vector Machine in Classification of Childbirth

Putroue Keumala Intan

UIN Sunan Ampel Surabaya, *putroue@uinsby.ac.id*

doi: https://doi.org/10.15642/mantik.2019.5.2.90-99

Abstrak: Angka kematian ibu hamil saat proses melahirkan dapat diturunkan melalui upaya ketepatan tim medis dalam menentukan proses persalinan yang harus dijalani. Pembelajaran menggunakan mesin dalam hal mengklasifikasikan proses persalinan bisa menjadi solusi bagi tim medis dalam menentukan proses persalinan. Salah satu metode klasifikasi yang dapat digunakan adalah metode *Support Vector Machine* (SVM) yang mampu menentukan *hyperplane* yang akan membentuk *decision boundary* yang baik sehingga mampu mengklasifikasikan data dengan tepat. Pada SVM terdapat fungsi kernel yang berguna untuk menyelesaikan kasus klasifikasi non linier dengan cara mentransformasi data ke dimensi yang lebih tinggi. Pada penelitian ini akan digunakan empat fungsi kernel; Linier, Radial Basis Function (RBF), Polinomial, dan Sigmoid pada proses klasifikasi proses persalinan guna mengetahui fungsi kernel yang mampu menghasilkan nilai akurasi tertinggi. Berdasarkan penelitian yang telah dilakukan diperoleh bahwasanya nilai akurasi yang dihasilkan oleh SVM dengan fungsi kernel linier lebih tinggi dibandingkan dengan tiga fungsi kernel yang lainnya.

Kata kunci*: persalinan, SVM, fungsi kernel*

Abstract: The maternal mortality rate during childbirth can be reduced through the efforts of the medical team in determining the childbirth process that must be undertaken immediately. Machine learning in terms of classifying childbirth can be a solution for the medical team in determining the childbirth process. One of the classification methods that can be used is the Support Vector Machine (SVM) method which is able to determine a hyperplane that will form a good decision boundary so that it is able to classify data appropriately. In SVM, there is a kernel function that is useful for solving non-linear classification cases by transforming data to a higher dimension. In this study, four kernel functions will be used; Linear, Radial Basis Function (RBF), Polynomial, and Sigmoid in the classification process of childbirth in order to determine the kernel function that is capable of producing the highest accuracy value. Based on research that has been done, it is obtained that the accuracy value generated by SVM with linear kernel functions is higher than the other kernel functions.

Keywords*: childbirth, SVM, kernel functions*

How to cite: *P. K. Intan, "Comparison of Kernel Function on Support Vector Machine in Classification of Childbirth", Mantik, vol. 5, no. 2, pp. 90-99, October 2019.*

1. Introduction

Childbirth is a period of high risk for mother and baby, either by vaginal birth or cesarean section. Some risks caused by childbirth are postpartum risks of cardiac arrest, wound hematoma hysterectomy, major puerperal infection, anaesthetic complications, venous thromboembolism and haemorrhage requiring hysterectomy[1]. Cesarean section risks include the morbidity associated with any major abdominal surgical procedure such as anaesthesia accidents, damage to blood vessels, an accidental extension of the uterine incision, damage to the urinary bladder and other organs. The cesarean section procedure is a potent risk factor for respiratory distress syndrome (RDS) in preterm infants and for other forms of respiratory distress in mature infants. RDS are major causes of neonatal morbidity and mortality.[2]

In order to reduce the risks caused by childbirth, several solutions are needed. One of the solutions that can be implemented is to use machine learning in the childbirth classification process, such as Amin and Ali's research on the performance evaluation of supervised machine learning classifiers for predicting healthcare operational decisions [3]. The most popular classification method is Support Vector Machine (SVM) which is capable of producing good accuracy values, as in the study conducted by Qiong Li, Qinglin Meng, et al. which compares SVM models and different artificial neural network models to predict hourly cooling loads in buildings and SVM results can achieve better accuracy and generalization [4]. In other research about Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points, these results indicated that SVM's had superior generalization capability, particularly with respect to small training sample sizes and the overall accuracies for the SVM algorithm were 91% (Kappa = 0.77) and 64% (Kappa = 0.34) for homogeneous and heterogeneous pixels [5], and in research on Comparison of Vector Engine Support and Artificial Neural Network Systems for Drug / Nondrug Classification, SVM yields 82% correct predictions and ANN produces 80% correct predictions[6]. SVM can also be applied in the field of medical science such as research conducted by Rajani and Selvi who applied the SVM classifier for early detection of breast cancer and research on SVM for diagnosis of Diabetes Mellitus.[7]

SVM method is able to determine the optimal hyperplane that will form a good decision boundary so that it can classify data appropriately. In SVM on nonlinearly separable data, there are two solutions, the first is to make soft margin that is particularly adapted to noised data, and the second is to use a kernel function. The kernel function is used to protect the data points to higher dimensional space for better classification.[8][9]

In this paper, we have developed a framework for the classification of childbirth using SVM classifier and studied the effect of various kernel functions on classification accuracy and performance.

2. Antecedents

2.1 Kernel Function

The Kernel function is used to protect the data points to higher dimensional space to improve its ability to find the best hyperplane to separate the data points of different classes.[9]

Definition (kernel function) [10]

A kernel is a function of κ for all $x, z \in X$ will meet the condition

$$\kappa(x,z) = \langle \phi(x).\phi(z)\rangle,$$

where ϕ is the mapping of inner product from X to space with higher dimension F

$$\phi: x \mapsto \phi(x) \in F.$$

Some common kernel function:[11]

a. Linear

The linear kernel function is defined as:

$$\kappa(\boldsymbol{x_i},\ \boldsymbol{x_j}) = \boldsymbol{x_i}^T\boldsymbol{x_j} \tag{1}$$

The linear kernel function is the simplest kernel function which is a dot product of two vectors.

b. Radial Basis Function (RBF)

RBF can also be called a Gaussian kernel function. RBF is defined as:

$$\kappa(\boldsymbol{x_i}, \boldsymbol{x_j}) = \exp\left(-\gamma\|\boldsymbol{x_i} - \boldsymbol{x_j}\|^2\right), \qquad \gamma > 0 \tag{2}$$

where γ is a positive parameter to set the distance

c. Polynomial

The polynomial kernel function with has a degree$d$, where r and d are the parameters defined as follows:

$$\kappa(\boldsymbol{x_i}, \boldsymbol{x_j}) = \left(\gamma\, \boldsymbol{x_i}^T\boldsymbol{x_j} + r\right)^d, \quad \gamma > 0 \tag{3}$$

d. Sigmoid

The sigmoid kernel function is defined as:

$$\kappa(\boldsymbol{x_i}, \boldsymbol{x_j}) = \tanh(\gamma\boldsymbol{x_i}, \boldsymbol{x_j} + \boldsymbol{r}) \tag{4}$$

where $\tanh(a) = 2\sigma(a) - 1$, and $\sigma(a) = \frac{1}{1+\exp(a)}$

2.2 Support Vector Machine (SVM)

SVM is a data mining method developed by Boser, Guyon, and Vapnik and presented for the first time in 1992 [12]. The basic idea of this SVM is to determine a hyperplane function in the form of a linear model that will form a decision boundary (DB) by maximizing margins. Margin is the distance between the hyperplane and the nearest data. The SVM method is not only able to solve linear classification problems but also able to solve non-linear classification problems using the kernel trick concept.

a. SVM on Linearly Separable Data

For N data set:

$$\{\boldsymbol{x_i}, t_i\}, \qquad i = 1 \dots N \tag{5}$$

with $\boldsymbol{x_i} = [x_1, x_2, \dots, x_n]$ is a line vector with dimensions n and $t_i = \{-1,1\}$ is target value on each row vector. The data will be classified into two class, i.e. class R_1 for target value $t_i = +1$ and class R_2 for target value $t_i = -1$. SVM uses a linear model as a hyperplane with a general form:

$$y(\boldsymbol{w}) = \boldsymbol{w}^T\boldsymbol{x} + b \tag{6}$$

where **x** is the input vector, **w** is the weight parameter, and b is a bias. So with the hyperplane SVM will classify the data into two classes R_1 and R_2 with each class will have a delimiter field parallel to the hyperplane as:

$$\boldsymbol{w}^T\boldsymbol{x_i} + b \geq 1, for\ t_i = +1 \tag{7}$$
$$\boldsymbol{w}^T\boldsymbol{x_i} + b \leq -1, for\ t_i = -1 \tag{8}$$

Or both of the delimiter fields are written in the following inequality:

$$t_i(\boldsymbol{w}^T\boldsymbol{x_i} + b) - 1 \geq 0 \tag{9}$$

The search for the best hyperplane in the SVM method is done by maximizing margins, which is maximizing the value of $\frac{1}{||\boldsymbol{w}||}$ which is the same as minimizing the value of $||\boldsymbol{w}||^2$ can be formulated into the following optimization problem: [13]

$$\underset{w,b}{\arg min}\ \frac{1}{2}||w||^2 \tag{10}$$
$$s.t\ \ t_i(\boldsymbol{w}^T\boldsymbol{x_i} + b) - 1 \geq 0, \quad i = 1, \dots, N \tag{11}$$

This problem will be more easily solved if it is changed into the Lagrange function (primal problem), thus the optimization problem can be changed to:

$$L(\boldsymbol{w}, b, a) = \frac{1}{2}||\boldsymbol{w}||^2 - \sum_{i=1}^{N} a_i t_i(\boldsymbol{w}^T\boldsymbol{x_i} + b) + \sum_{i-1}^{N} a_i \tag{12}$$

With the addition of the lagrange multiplier $a_i \geq 0$.

The dual problem is obtained as follows:

$$\underset{a}{\arg max}\ L(a) = \sum_{i-1}^{N} a_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} a_i t_i a_j t_j \left(\boldsymbol{x_i^T}\boldsymbol{x_j}\right) \tag{13}$$

$$s.t\ \ \sum_{i-1}^{N} a_i\, t_i = 0, \qquad a_i \geq 0 \tag{14}$$

The optimization form above must meet the following The Karush-Kuhn-Tucker (KKT) conditions [13]:

$$a_i \geq 0 \tag{15}$$

$$t_i(\boldsymbol{w^T}\boldsymbol{x}_i + b) - 1 \geq 0 \tag{16}$$

$$a_i(t_i(\boldsymbol{w^T}\boldsymbol{x}_i + b) - 1) = 0 \tag{17}$$

From the KKT conditions, the optimization problem solution (13) is $(a^*, \boldsymbol{w}^*, b^*)$ that satisfies $a_i^*(t_i(\boldsymbol{w}^{*\boldsymbol{T}}\boldsymbol{x}_i + b^*) - 1) = 0$. So if there is training data that has the value $a_i > 0$ then $t_i(\boldsymbol{w^T}\boldsymbol{x}_i + b) = 1$, it means that the data is a support vector while the rest have the value $a_i = 0$. Thus the resulting decision function is only influenced by support vectors. After the problem solution is found $(a^*, \boldsymbol{w}^*, b^*)$ then the class of the $\boldsymbol{x}$ test data can be determined based on value of the decision function:

$$y(\boldsymbol{x}) = \sum_{i-1}^{Ns} a_i{}^* \, t_i \boldsymbol{x_i^T} \boldsymbol{x} + b^* \tag{18}$$

where $\boldsymbol{x_i}$ is a support vector, Ns = number of support vector.

b. SVM on Nonlinearly Separable Data

For data that is not classified correctly, the SVM model must be modified by adding the slack variable ξ_i. Hyperplane search by adding slack variables is also called a hyperplane soft margin as follows:

$$\underset{w,b,\xi}{\arg min} \; \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{N} \xi_i\right) \tag{19}$$

$$s.t \quad t_i(\boldsymbol{w}^T\boldsymbol{x_i} + b) \geq 1 - \xi_i, \quad i = 1, \dots, N \tag{20}$$

$$\xi_i \geq 0 \qquad , i = 1, 2, \dots, n \tag{21}$$

Where C is the parameter that determines the amount of penalty due to errors in classification and the value of C is not obtained in the learning process but must be determined before learning.

In addition to adding slack variables to deal with data that is not linearly classified, it is also necessary to transform the data into higher dimensions with kernel functions so that it can be linearly separated at higher dimensions. Using the kernel trick, the $\boldsymbol{x_i}$ data will be mapped with the function $\phi(\boldsymbol{x}_i)$, and each product $(\boldsymbol{x}_i . \boldsymbol{x}_j)$ will be calculated using $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

Thus the linear model used as a hyperplane is:

$$y(\boldsymbol{x}) = \boldsymbol{w}^T\phi(\boldsymbol{x}) + b \tag{22}$$

The form of optimization issues from soft margin to:

$$\underset{\boldsymbol{w}, b, \xi}{\arg\min} \; \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n} \xi_i \tag{23}$$

$$s.t \;\; t_i(\boldsymbol{w}^T\phi(\boldsymbol{x}_i) + b) \geq 1 - \xi_i \quad , i = 1, 2, \dots, N \tag{24}$$

$$\xi_i \geq 0 \qquad , i = 1, 2, \dots, N \tag{25}$$

By multiplying the Lagrange multiplier $a_i \geq 0$ and $\mu_i \geq 0$ to the primal form of the optimization problem, so the Lagrange function is obtained as follows:

$$L(\boldsymbol{w}, b, \xi, a, \mu) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n} \xi_i + \sum_{i=1}^{n} a_i\{1 - \xi_i - t_i(\boldsymbol{w}^T\phi(\boldsymbol{x}_i) + b)\} - \sum_{i=1}^{n} \mu_i\xi_i$$

The dual problem form is obtained as follows:

$$\underset{a}{\max} \; \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i t_i a_j t_j \left(\phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j)\right) \tag{26}$$

$$s.t \quad \sum_{i=1}^{n} a_i t_i = 0$$

$$0 \leq a_i \leq C \qquad ,i = 1, 2, \dots, n$$

The dual form in equation (3.35) satisfies the KKT conditions as follows [12]:

$$a_i \geq 0 \tag{27}$$

$$\mu_i \geq 0 \tag{28}$$

$$t_i(\boldsymbol{w^T}\phi(\boldsymbol{x}_i) + b) - 1 + \xi_i \geq 0 \tag{29}$$

$$\xi_i \geq 0 \tag{30}$$

$$a_i\{t_i(\boldsymbol{w^T}\phi(\boldsymbol{x}_i) + b) - 1 + \xi_i\} = 0 \tag{31}$$

$$\mu_i\xi_i = 0 \tag{32}$$

From the KKT conditions, the optimization problem solution (26) is $(a^*, \boldsymbol{w}^*, b^*)$ that satisfies $a_i^*(t_i(\boldsymbol{w}^{*\boldsymbol{T}}\phi(\boldsymbol{x}_i) + b^*) - 1 + \xi_i) = 0$ and $\mu_i\xi_i = 0$. So if there is training data that has the value $a_i = 0$, then $\mu_i = C > 0$ and $\xi_i = 0$ results in $t_i(\boldsymbol{w^T}\phi(\boldsymbol{x}_i) + b) \geq 0$ then the data is correctly classified and not support vector. if there is training data that has the value $0 < a_i < C$, then $\mu_i > 0$ and $\xi_i = 0$ results in $t_i(\boldsymbol{w^T}\phi(\boldsymbol{x}_i) + b) \geq 0$ then the data is correctly classified and not support vector. Whereas if there is training data that has the value $a_i = C$, then $\mu_i = 0$ and $\xi_i > 0$ results in $\xi_i \leq 1$ or $\xi_i > 1$, so $t_i(\boldsymbol{w^T}\phi(\boldsymbol{x}_i) + b) \geq 0$ then the data is not properly classified. After the problem solution is found $(a^*, \boldsymbol{w}^*, b^*)$ then the class of the x test data can be determined based on the value of the decision function:

$$y(\boldsymbol{x}) = \sum_{i=1}^{Ns} a_i^* t_i \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) + b^* \tag{33}$$

where $\boldsymbol{x_i}$ is a support vector, Ns = number of support vectors.

3. Methodology

3.1 Tools Used

The main tools used for this analysis and study is python programming language, which is a free open source platform for machine learning. There are many packages available with standard implementations for various machine learning algorithms. The scikit-learn packages (reference) used in this study for the implementation of preprocessing, model selection and SVM classification with four kernel functions [14].

3.2 Preprocessing

Normalization or scaling of each feature data highly recommended before being processed by SVM, i.e. in the interval [-1, +1] [15]. Preprocessing is done to avoid the domination of features with large values over features with small values and to avoid numerical difficulties during the calculation process.

3.3 Model Selection

In almost all data mining methods, there will be parameters that cannot be determined in the learning process. Determining the value of these parameters must be done before the learning process. Determination of these parameters is called model selection. Model selection aims to tune the hyperparameters of SVM classification (the penalty parameter C and any kernel parameters) in order to achieve the lowest test error, *i.e.* the lowest probability of misclassification of unseen test examples [16].

In SVM, there two parameters that need to be determined parameters C and γ. Rows of parameter pairs $(C,\ \gamma)$ are recommended as candidates is $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$ dan $\gamma = 2^{15}, 2^{13}, \dots, 2^{3}$ [11]. Random search methods will be used in this study. According to Bergstra and Bengio (2012), random search method is a more efficient method than the grid search method, due to determine the value of the parameter that can produce the same accuracy value even better without having to try all possible parameter values in the range of a specified value [17].

3.4 Classification

The childbirth data will be classified using SVM with different kernel functions after obtaining the parameter values (C, γ) using the scikit learn package in python programming language.

3.5 Classification Evaluation

a. Evaluation Procedure

K - Fold Cross Validation is one procedure that is commonly used to estimate the performance of a model.[15] This product consists of 3 stages:

- Divide data into k parts of the same size.
- k-1 part is used as training data, and one part is used as testing data.
- This process is done as k times as repetition for each different combination of testing data and training data so that the whole section will become testing data.

The accuracy of each iteration is averaged to get an estimate of the final accuracy of the model. This study uses k = 5 or 5-fold cross-validation, so that for each experiment will use four subsets for training data, and one subset for data testing conducted five times trial for all possibilities.

b. Unit of Evaluation Size

Confusion matrix

The confusion matrix is used to present the results of the K-Fold Cross Validation as follows:[18]

Table 1. *Confusion matrix*

		Prediction	
		-1	1
actual	-1	True Negative (TN)	False Positive (FP)
	1	False Negative (FN)	True Positive (TP)

One unit of performance measurement based on confusion matrix is Accuracy Value which can be calculated by:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{34}$$

ROC Curve

To visualize the comparison results of two or more classification models, you can use the ROC Curve (Receiver Operating Characteristics). The ROC curve is a two-dimensional graph with False Positive (FP) as a horizontal line and True Positive (TP) as a vertical line[18]. Point (0.1) states that a perfect classification of all positive and negative cases with no FP value or FP = 0 and a high TP value or TP = 1. Point (0,0) states that the classification predicts each case to be -1. Point (1,1) states that the classification that predicts each case becomes 1. To evaluate the classification can be seen from the AUC (Area Under the Curve) value. The accuracy level of the AUC (Area Under the Curve) value in the classification is divided into five groups expressed in Table 2:[18]

Table 2. Accuracy Levels of AUC Value in Classifications

AUC Interval Value	Accuracy Level
0.90 – 1.00	excellent classification
0.80 – 0.89	good classification
0.70 – 0.79	fair classification
0.60 – 0.69	poor classification
0.50 – 0.59	failure classification

4. Experimental Analysis

This experiment used 304 data childbirth which is divided into two classes; vaginal birth 104 data and caesarean section 200 data. The data consists of 10 features; age, hypertension history, glucose disease, first pregnancy, fetal position, parturition history, number of fetuses, hip size, another disease, and ruptured amniotic fluid history.

The results of model selection we got the parameters value $C = 2048{,}0$ and $\gamma = 8{,}63167457503109383e - 0.5$, that will be used to whole kernel functions on support vector classification and specifically the polynomial kernel function will be used in 3rd degree.

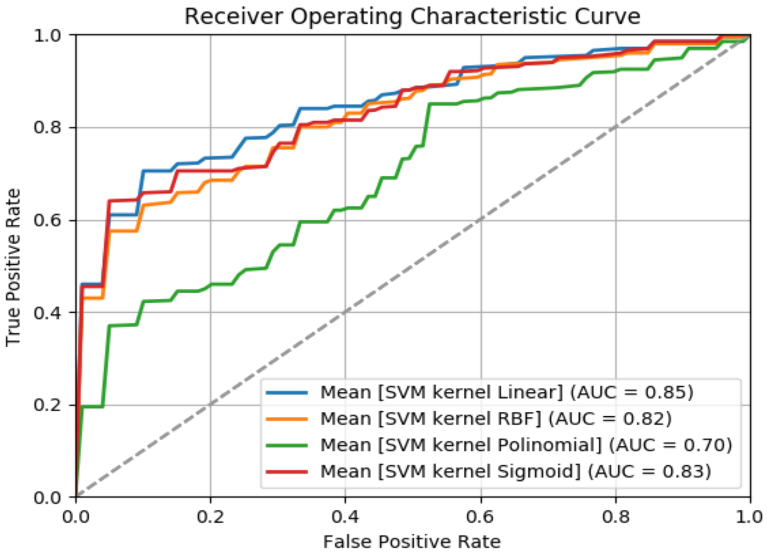

Figure 1. ROC Curve for various kernel functions

Figure 1. shows the mean AUC from 5fold validation for SVM with four different kernel function. SVM-Linear, SVM-RBF, and SVM-Sigmoid are categorized in good classification because the AUC values are at intervals of 0.80 - 0.89, but SVM-Polynomial is only categorized as a fair classification because its AUC value is 0.70. Based on the results of the AUC value, it can be seen that SVM-Linear is the best model because its AUC value is greater than the others. To confirm the conclusion that we need to see the results of the accuracy of the model. Accuracy value of each SVM model with different kernel functions can be seen below:

Table 3. Accuracy value of each model

No	Model	TN	FP	FN	TP	Accuracy Value
1	SVM-Linear	75	29	24	176	0,83
2	SVM-RBF	67	37	18	182	0,82
3	SVM-Polynomial	0	104	0	200	0,66
4	SVM-Sigmoid	78	26	43	157	0,77

Table 4. shows the whole results of the confusion matrix from each classification with a different kernel function. SVM-Polynomial fails to predict class -1 data properly because the result of TN is 0, but it succeeds in predicting class 1 data perfectly. Other models produce different class 1 and -1 data predictions so that different accuracy values are obtained.

Based on the results of the accuracy value of each model, we can find out the best model for classifying the childbirth. SVM-Linear can produce an accuracy value of 0,83, and this result is greater than the SVM-RBF result of 0,82, SVM-Polynomial of 0,66, and SVM-Sigmoid of 0,77. It can be concluded that SVM-Linear method is the best model in classifying the childbirth. So the classification model that is most suitable for labour classification is SVM which uses linear kernel functions.

5. Conclusion

Based on experiments that have been done to see the effect of kernel function selection on the SVM method on the accuracy of the analysis of the childbirth classification, it can be concluded that the SVM-linear method is the best model in classifying the childbirth.

This can be seen from the accuracy value produced by the SVM-Linear method that is 0.83, and this value is higher than the accuracy value generated by SVM-RBF which is 0.82, SVM-Polynomial which is 0.66, and SVM-Sigmoid which is 0,77. In addition, the AUC value produced by SVM-Linear is 0.85 and is greater than other methods, which shows good classification.

References

[1] S. Liu *et al.*, “Maternal mortality and severe morbidity associated with low-risk planned cesarean delivery versus planned vaginal delivery at term.,” *CMAJ*, vol. 176, no. 4, pp. 455–60, Feb. 2007.

[2] M. Wagner, “Choosing caesarean section.,” *Lancet (London, England)*, vol. 356, no. 9242, pp. 1677–80, Nov. 2000.

[3] A. Amin, Muhammad & Ali, “Performance Evaluation of Supervised

Machine Learning Classifiers for Predicting Healthcare Operational Decisions," 2018.

[4] Q. Li, Q. Meng, J. Cai, H. Yoshino, and A. Mochida, "Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks," *Energy Convers. Manag.*, vol. 50, no. 1, pp. 90–96, Jan. 2009.

[5] Y. Shao and R. S. Lunetta, "Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points," *ISPRS J. Photogramm. Remote Sens.*, vol. 70, pp. 78–87, Jun. 2012.

[6] E. Byvatov, U. Fechner, J. Sadowski, and G. Schneider, "Comparison of Support Vector Machine and Artificial Neural Network Systems for Drug/Nondrug Classification," *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 6, pp. 1882–1889, Nov. 2003.

[7] Y. I. A. Rejani and S. T. Selvi, "Early Detection of Breast Cancer using SVM Classifier Technique," Dec. 2009.

[8] D. Novitasari, "Klasifikasi Alzheimer dan Non Alzheimer Menggunakan Fuzzy C-Mean, Gray Level Co-Occurence Matrix dan Support Vector Machine", *mantik*, vol. 4, no. 2, pp. 83-89, Oct. 2018.

[9] S. Pahwa and D. Sinwar, "Comparison Of Various Kernels Of Support Vector Machine," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 3, no. VII, pp. 532–536, 2015.

[10] John_Shawe-Taylor_&_Nello_Christianini, *Kernel Methods For Pattern Analysis*, vol. 111, no. 479. New York: Cambridge university Press, 1965.

[11] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A Practical Guide to Support Vector Classification," 2003.

[12] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "Training algorithm for optimal margin classifiers," *Proc. Fifth Annu. ACM Work. Comput. Learn. Theory*, pp. 144–152, 1992.

[13] Christopher M Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Heidelberg: Springer-Verlag Berlin, 2006.

[14] "1.4. Support Vector Machines — scikit-learn 0.21.3 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/svm.html#svm-classification. [Accessed: 29-Aug-2019].

[15] F. Pedregosa Fabianpedregosa *et al.*, "Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos Pedregosa, Varoquaux, Gramfort Et Al. Matthieu Perrot," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[16] C. Gold and P. Sollich, "Model selection for support vector machine classification," *Neurocomputing*, vol. 55, no. 1–2, pp. 221–249, 2003.

[17] J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Optimization," *J. Mach. Learn. Res.*, vol. 13, no. Feb, pp. 281–305, 2012.

[18] F. Gorunescu, *Dana Mining: Concepts, Models, and Techniques*. New York: Springer US, 2011.