

# Comparison of Accuracy Level of Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) Algorithms in Predicting Heart Disease

Dimas Aryo Anggoro<sup>1</sup>, Naqshauliza Devi Kurnia<sup>2</sup>

<sup>1</sup>Informatics Department, Universitas Muhammadiyah Surakarta, Indonesia, [dimas.a.anggoro@ums.ac.id](mailto:dimas.a.anggoro@ums.ac.id)

<sup>2</sup>Informatics Department, Universitas Muhammadiyah Surakarta, Indonesia, [L200160087@student.ums.ac.id](mailto:L200160087@student.ums.ac.id)

## ABSTRACT

Heart disease receives a lot of attention in medical research because of its great effect on the human health state. Based on cases in 2008, it was estimated that more than 3 million deaths were caused due to heart disease. This study aims to compare two algorithms and find which algorithm can be utilized appropriately in predicting the accuracy of heart disease data and has the benefit to consider the health problems regarding the percentage of heart disease as well as being an accurate information material. This study used a comparison of the Support Vector Machine (SVM) algorithm and the K-Nearest Neighbor (KNN) algorithm. The method used is the Classification method, one of the techniques in Machine Learning, to find parameters in a linear equation that can map inputs and outputs. The results show that the SVM algorithm testing with normalization had better accuracy results compared to KNN algorithm either with or without normalization continued to produce poor accuracy. The SVM classification results without normalization were 84.61% and with normalization was 90.10%, while the KNN algorithm showed an accuracy of 64.83% and 81.31% with normalization.

**Key words :** K-Nearest Neighbor, Heart Disease, Support Vector Machine.

## 1. INTRODUCTION

Heart is one of the most vital organs in human body because it is located in the center of the circulatory system. The heart functions as a pump that transports blood to all parts of the human body through the network of blood vessels and supplies oxygen and other important components. Problems in parts of the body and even death may occur without good cardiac activity[1].

According to the latest statistics from the World Health Organization (WHO), heart disease gains a lot of attention in medical research because of its great impact on human health [2] and is the number one disease with high mortality rates worldwide[3]. Based on cases in 2008, it was estimated that more than 3 million deaths were caused by heart disease. 4% of premature deaths occur in high-income countries and 42%

in low-income countries, to this date, heart disease is one of the most worrying diseases[4]. Heart disease occurs due to impaired function of the heart[5]. As in the case of coronary heart disease occurs due to the accumulation of cholesterol, therefore, it forms plaque in the walls of the coronary arteries for a long period. Another disorder of coronary heart case is caused by the blockage of blood flow to the heart, this disorder is known as atherosclerosis. Over time, coronary heart disease can cause heart muscle to weaken, and increase the risk of heart attacks and complications such as heart failure and arrhythmia or heart rhythm disorders[6].

Data mining is widely used as an adequate and precise predictor and diagnosis of many diseases. The technique used is competent for designing clinical support systems because of its ability to find patterns and relationships hidden in medical data[7]. One of the most important uses is in heart disease diagnosis cases because it is one of the highest causes of death worldwide[8]. Almost every system in heart disease prediction uses a dataset obtained from a laboratory. Heart disease diagnose is conducted through testing heart disease, including chest X-Ray, coronary angiograms, electrocardiograms (ECG or EKG), and other several tests for stress relief [9].

This study aims to compare two algorithms and find which algorithm can be used to predict the highest accuracy of heart disease data and has the benefit of solving health problems based on the percentage of heart disease as well as being an accurate information material. Classification is a technique of mining data in a dataset [10], which can be conducted automatically or semi-automatically to find patterns from datasets in which the results can be validated or identified and utilized in generating predictions, detection as well as recommendations in decision making.

Several prominent algorithms with good reliability in making data classification are Support Vector Machine (SVM) algorithm and K-Nearest Neighbor (KNN) algorithm. The Support Vector Machine algorithm was first introduced by Vapnik in 1992[11] as a series that excels in pattern recognition milieu and is used for classification by applying statistical methods using data mining technique. Support Vector Machine method is a classification method that is reliable in a large amount of data processing and eases error

detection [12]. The Support Vector Machine method was chosen because it can be administered analytically and has good generalization capability compared to the NN (Neural Network) model because it has proven superior performance, and is identified as the best predictor[13] because it has been able to provide good accuracy results of approximately 95.4% [14]. Through Support Vector Machine algorithm, it can measure how accurate the model is to classify the value of results from previous data and is expected to find the accuracy level of the number of people with heart disease.

K-Nearest Neighbor (KNN) algorithm is an approach used to classify data into existing data simply and efficiently. The concept of K-Nearest Neighbor algorithm is to find data with the closest distance between the data evaluated and the closest number of K (neighbors) to the training data[15]. The K-Nearest Neighbor algorithm has a goal, that is, new objects are classified based on attributes and training samples, then the results of the test are classified according to the majority of the categories in the KNN. This research has a high accuracy of 91%. The advantage of K-Nearest Neighbor algorithm is the use of training data which has a lot of noise and is compelling on a large amount of training data. However, the K-Nearest Neighbor algorithm also has shortcomings, such as the need to determine the value of k in choosing good attributes[16] and the determination of the optimal k parameter is too high. This will reduce the effect of noise on the classification, thus making the boundaries between the classifications bias. If the value of k is low, it will cause the sampling of the comparative sample value becomes little and the accuracy is reduced[17].

## 2. METHOD

### 2.1 Data Collection

The first step is to analyze and select the entire data to obtain relevant outputs in the data mining process. The heart disease data has a high dimension which includes 304 data and 14 attributes. Following is a description of the attributes and variables used as presented in Table 1.

**Table 1:** Attributes and variables of heart disease

Variable	Attributes	Description
X1	Age	Age in year
X2	Sex	1 = male;0= female
X3	Cp	1 = Typical Angina; 2 =Atypical Angina; 3 =Non-anginal chest pain;4 = Asymptomatic
X4	Trestbps	Blood pressure at rest in mm Hg
X5	Chol	Cholesterol serum in mg/dl
X6	Fbs	Fasting blood glucose> 120 mg / dl (1 = correct; 0 = incorrect)
X7	Restecg	Electrocardiography results: 0 = Normal 1 = Contains ST-T

		waveabnormality (wave inversion T and elevation or ST depression> 0.05 mV) 2 = Shows possible or fixed left ventricular hypertrophy with estes criteria.
X8	Thalach	Target of maximum heart rates
X9	Exang	Exercise induced by angina (1 = yes; 0 = no)
X10	Oldpeak	ST depression due to relative exercise against resting
X11	Slope	The slope of ST segment relative peak 1 = upward; 2 = flat; 3 = downward
X12	Ca	Number of <i>major vessels</i> (0-3) colored by flourosopy
X13	Thal	3 = normal ; 6 = permanent defect; 7 = fixable defect
Y	Target	0 = painless ; 1 = pain

Attributes and variables can be obtained from valid data on Kaggle regarding heart disease. Data on this heart disease was made by Andras Janosi, M.D from the Hungarian Institute of Cardiology Budapest, William Steinbrunn, M.D from the University Hospital, Zurich, Switzerland, Matthias Pfisterer, M.D from the University Hospital, Basel, Switzerland, and Robert Detrano, MD, Ph.D. from VA Medical Center, Long Beach Foundation and Cleveland Clinic.

### 2.2 Data preprocessing

#### A. Data Sampling

Before the data mining process was carried out, the heart disease dataset must be sampled previously by dividing it into training sample data and testing sample data. The heart dataset is divided into 70% training data and 30% testing data to determine the effect of the amount of training data on the level of accuracy generated by neglecting the testing data[18].

#### B. Normalization Data

The next step is the normalization process of the min-max method on the dataset by transforming the original data[19], by changing the scale of attribute values, it can balance the comparative value of data when before and after being processed, and reduce errors in data mining processing[20] as shown in equation 1.

$$\text{Normalization}(x) = \frac{\text{min Range} + (x - \text{min Value})(\text{max Range} - \text{min Range})}{\text{max Value} - \text{min Value}} \quad (1)$$

X = Original value

Min Range = The minimum normalized value

Max Range = The maximum normalized values

Min Value = The lowest value of the dataset

Max Value = The highest value of the dataset

### 2.3 Data Processing

#### A. Support Vector Machine (SVM) Algorithm

Support Vector Machine (SVM) algorithm is a part of the machine learning technique first introduced in 1992 by

Vapnik, Boser, and Guyon. The Support Vector Machine algorithm learning system used a hypothetical space in the form of linear functions in a high-dimensional feature and was trained with an algorithm based on optimization theory. Problem encountered in the Support Vector Machine algorithm are how to separate the two classes with a function obtained from available training data and the concept of classification using the Support Vector Machine algorithm is simply an attempt to find the best hyperplane that functions as a separator of two classes of data in the input space [21]. Hyperplane can be found by measuring margins and finding the maximum point[22].

Margin is the distance between the hyperplane and the closest pattern of each class where the closest pattern is called support vector[23]. The effort to find the location of a hyperplane is the core of the learning process of the Support Vector Machine algorithm. Each data (example) is denoted as  $x_i \in \mathbb{R}^D$ ,  $i = 1, 2, \dots, N$ .  $N$  is the amount of data, with each data and class label notified as

$$y_i \in \{-1, +1\}. \tag{1}$$

It was assumed that the two classes could be completely separated by a hyperplane in the  $D$ -dimensional feature space. The hyperplane is defined as in equation 2.

$$w \cdot x_i + b = 0 \tag{2}$$

Where,  $w$  = normal plane and  $b$  = plane position relative to the center. Data  $x_i$  classified as a negative class were those that fulfilled the inequality as in equation 3.

$$w \cdot x_i + b \leq -1 \tag{3}$$

While classified as a positive class is shown in equation 4.

$$w \cdot x_i + b \geq 1 \tag{4}$$

Optimal margins were obtained by maximizing the distance between the closest hyperplane and pattern. This distance is formulated in equation 5.

$$\frac{1}{\|w\|} \tag{5}$$

The minimum point could be found using equation 6 by taking the constraints of equation 7 into account.

$$y_i(w \cdot x_i + b) - 1 \geq 0, \forall i \tag{6}$$

$$\text{Constraint } \alpha_i \geq 0, \sum_{i=1}^N \alpha_i y_i = 0 \tag{7}$$

Optimization could be solved by various computational techniques, for instance, Lagrange multiplier where the technique was used to determine the minimum or maximum relative value of a function limited by a constrain condition as in equation 8.

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i y_i (w \cdot x_i + b - 1) \tag{8}$$

Where  $\alpha_i$  was Langrange multiplier that corresponded to  $x_i$  and the value of  $\alpha_i$  obtained was mostly zero or positive. To earn the optimal value, one must minimize  $L$  with respect to  $w$  and  $b$ , and maximize  $L$  with respect to  $\alpha_i$  while minding the nature of optimal gradient point  $L = 0$ .

The maximum point could be found using equation 9 by considering the constraints of equation 10.

$$\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1} \alpha_i \alpha_j y_i y_j x_i x_j \tag{9}$$

$$\text{Constraint } \alpha_i \geq 0, \sum_{i=1}^N \alpha_i y_i = 0 \tag{10}$$

The solution to this problem will produce an amount of  $\alpha_i$  with positive values. Data correspondings to non-zero  $\alpha_i$  were support vectors, which were data with the closest distance to the hyperplane.

*B. K-Nearest Neighbor (KNN) Algorithm.*

K-Nearest Neighbor (KNN) algorithm is the simplest algorithm compared to other algorithms to predict a class in a sample with the unsupervised algorithm method in which the results were classified based on the majority of  $k$ -NN categories. The main objective is to research the number of sufferers of heart disease in the coming year based on data that previously used the K-Nearest Neighbor (KNN) method. To classify it, the K-Nearest Neighbor algorithm will work based on the closest distance between objects in the following circumstances[24]:

1. Determine the parameter  $k$  (number of nearest neighbors)
2. Calculate the distance between data to be evaluated with training data
3. Sort the distance from high value to low value
4. Determine the closest distance to the  $k$  order
5. Pair the appropriate class

The condition of the  $k$  value itself cannot be greater than the amount of training data, and the  $k$  value must be odd and more than one[25]. The  $k$  value is good based on the number of data. The higher the  $k$  value, the lower the effect of noise in classification [26]. In calculating the distance, the method used is the Euclidean distance method. The advantage of the K-Nearest Neighbor algorithm is its resilience to training data which contains a lot of noise and will be sufficient if the training data is large [16]. In measuring the distance neighbors can be calculated using Euclidean distance as in equation 11.

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \tag{11}$$

Where:

$x_{1i}$  = Data Samples

$x_{2i}$  = Testing Data

$i$  = Data Variable

$d$  = Distance

$p$  = Data Dimension

**2.4 Evaluation Model**

*A. Accuracy*

Accuracy is a true prediction ratio (positive and negative) by testing the entire data into the model to obtain the target value [27]. Accuracy is used to find the amount of data classified correctly [7]. Algorithm performance in terms of accuracy can be evaluated using the formula in equation 12 following.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + FN + TN)} * 100\% \tag{12}$$

Where,

TP = True Positive

TN = True Negative

FP = False Positive

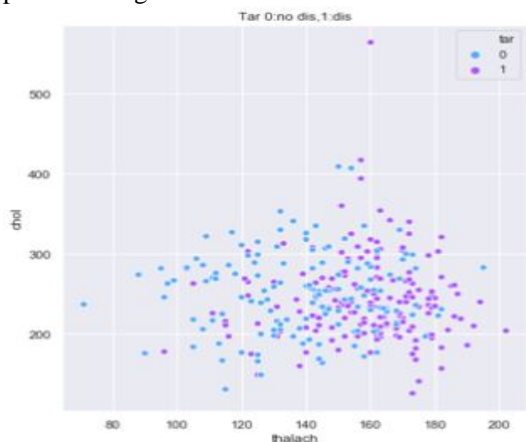
FN = False Negative

### 3. RESULT AND DISCUSSION

A comparative analysis of various classification algorithms in the heart disease dataset had been carried out. The dataset used in this study is heart disease data with 304 data with 14 variables consisting of 13 input variables and 1 target variable. Based on the heart disease dataset, the percentage of people with heart disease is higher than those with no heart disease. 138 patients did not have heart disease with a percentage of 45.54% and 165 people with heart disease with a percentage of 54.46%. Based on the dominant data the majority of sufferers of heart disease were men with a percentage of 68.32% and women of 31.68%.

Early stages of the process of collecting and processing datasets. Referring to the heart dataset above, the comparison was used to find which algorithm could produce the best accuracy, by comparing the Support Vector Machine algorithm and the K-Nearest Neighbor algorithm. This study used 70% of training data and 30% of testing data from the heart disease dataset in the testing phase, the heart dataset was tested using normalization and without normalization.

Normalization plays an important role in the iteration process, because of the addition of the normalization formula, thus, an increase in the value of accuracy occurs resulting from the testing process. Normalization also plays a role so that the value of each feature is in the same interval so the classification will be adequate and accurate [28]. The normalized data points of chol and thalach variables are plotted in Figure 1.



**Figure 1:** Normalization Graph

Normalization is also used to produce higher modeling accuracy because normalization transforms values at the time of modeling, to obtain higher accuracy than before normalization[29].

The next stage is the processing of dataset using a python programming language on the Anaconda Navigator application by including the Support Vector Machine algorithm in it. Testing the Support Vector Machine algorithm was conducted one by one by initial initiation for the values  $\alpha = 0$ ,  $c = 1$ ,  $\epsilon = 0.001$ ,  $\gamma = 0.5$ , and  $\lambda = 0.5$  then input the test data where nis the class of the heart

dataset separated by hyperplane with dimension D and weight is +1.

Classification of Support Vector Machine algorithms obtained kernel models with weights from each attribute as in Table 2 where the results of the optimization of the margin used to maximize the distance between hyperplane with the closest pattern had a bias value (offset) of 0.946.

**Table 2:** Weight value in each attribute

No	Weight [ ]	Value
1.	Age	0.060
2.	Cp	0.261
3.	Chol	-0.045
4.	Restecg	0.261
5.	Exang	0.206
6.	Slope	0.143
7.	Thal	-0.089
8.	Sex	0.222
9.	Trestbps	-0.118
10.	Fbs	0.127
11.	Thalach	0.047
12.	Oldpeak	0.043
13.	Ca	0.146

The next step after initiating and obtaining the value of w is to add an optimization method called Lagrange multipliers which can transform problem constraints into no constraints and is used to perform nonlinear functions in solving optimization problems [30] which then produces positive values derived from hyperplane from the closest distance called support vector. After obtaining the support vector results, then input normalization in the Support Vector Machine algorithm testing so that it produced an accuracy of 90.10% with the results of the confusion matrix testing as in Table 3.

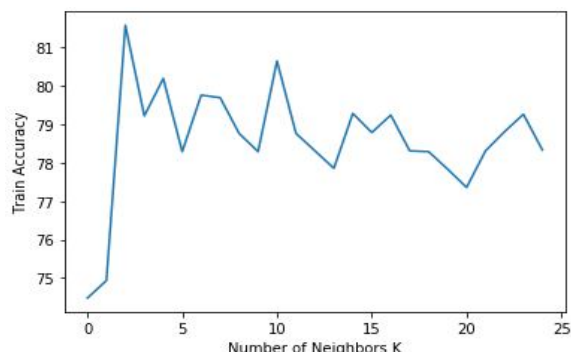
**Table 3:** Testing Results of Support Vector Machine Algorithms

Dataset	Accuracy
Normalization	90.10%
No Normalization	84.61%

The results of using normalization were higher than those that without normalization, because through the normalization, the value of each feature was transformed into intervals of 0 to 1 so the classification results were better and more accurate, so testing the Support Vector Machine algorithm with no normalization only produced an accuracy of 84.61%. As a conclusion, normalization in the Support Vector Machine algorithm affected accuracy because it resulted in higher modeling accuracy while data without normalization resulted in lower accuracy.

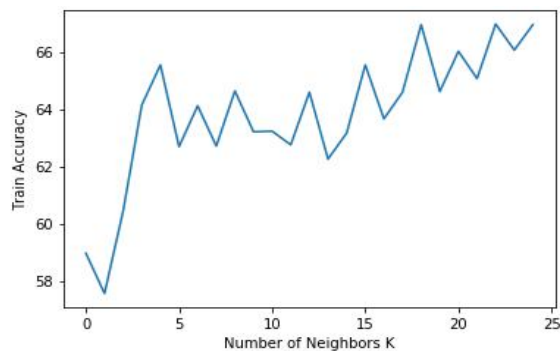
The next test was testing the heart disease dataset using the K-Nearest Neighbor algorithm testing using a cross-validation technique by finding the k value using K-fold

cross-validation. The k value in the k-fold used was 10, 10-fold cross-validation. Tests on a heart dataset of 123 records, the value of k neighbors used  $1 \leq K \leq 25$  and must use k which has an odd value so it used the value (K = 11) of 80.65 (+/- 10.68) as shown in Figure 2.



**Figure 2:** Results of the optimum k graph with normalization

K-Nearest Neighbor algorithm produced high accuracy value due to the influence of normalization which was included in the calculation, so it could produce an accuracy value of 81.31%. If there is no effect of normalization, then the graph obtained will show as in Figure 3.



**Figure 3:** Results of the optimum k without normalization

All things considered, the calculation using the K-Nearest Neighbor algorithm without any normalization resulted in a lower accuracy value with only 64.83% which means normalization on the K-Nearest Neighbor algorithm affected the accuracy results. The test results are shown in Table 4.

**Table 4:** Test Results for the K-Nearest Neighbor Algorithm

Dataset	Accuracy
Normalization	81.31%
No Normalization	64.83%

Based on the SVM and KNN test results table, the accuracy of the Support Vector Machine algorithm is superior to the K-Nearest Neighbor algorithm due to the ability of the Support Vector Machine algorithm to solve problems better related to large dimensions[31] and how it works in finding the best hyperplane parameters to produce optimal accuracy[32] that cannot be overcome by the K-Nearest Neighbor algorithm.

#### 4. CONCLUSION

Based on the results of the conducted research, testing the Support Vector Machine algorithm by using normalization showed better accuracy results, while the K-Nearest Neighbor algorithm with and without normalization still resulted in lower accuracy. The results of the classification of Support Vector Machine algorithm without normalization were 84.61% and 90.10% with normalization, while the K-Nearest Neighbor algorithm showed an accuracy of 64.83% and 81.31% using normalization. Thus, the classification of patients with heart disease by testing the Support Vector Machine algorithm that used normalization showed better accuracy results, while the K-Nearest Neighbor algorithm with and without normalization still resulted in lower accuracy.

Based on research that has been carried out there are many shortcomings and the need for further development. As for suggestions for further research, it is recommended to use other algorithms to increase the value of better accuracy in cases of heart disease.

#### REFERENCES

- [1] P. Dembla and T. Bhatia, **Multiclass diagnosis model for heart disease using pso based svm**, July, 2016.
- [2] M. Abdar, S. R. N. Kalhori, T. Sutikno, I. M. I. Subroto and G. Arji, **Comparing performance of data mining algorithms in prediction heart diseases**, vol. 5, no. 6, pp. 1569–1576, 2015. <https://doi.org/10.11591/ijece.v5i6.pp1569-1576>
- [3] M. S. Amin, Y. K. Chiam and K. D. Varathan, **Identification of significant features and data mining techniques in predicting heart disease**, *Telemat. Informatics*, 2018. <https://doi.org/10.1016/j.tele.2018.11.007>
- [4] A. Rohman, V. Suhartono and C. Supriyanto, **Penerapan algoritma c4.5 berbasis adaboost untuk prediksi penyakit jantung**, vol. 13, pp. 13–19, 2017.
- [5] K. Mathan, P. M. Kumar, P. Panchatcharam, G. Manogaran and R. Varadharajan, **A novel gini index decision tree data mining method with neural network classifiers for prediction of heart disease**, *Des. Autom. Embed. Syst.*, vol. 22, no. 3, pp. 225–242, 2018. <https://doi.org/10.1007/s10617-018-9205-4>
- [6] E. Wahyudi and S. Hartati, **Case-based reasoning untuk diagnosis penyakit jantung**, vol. 11, no. 1, 2017.
- [7] A. Joshi, E. J. Dangra and D. M. K. Rawat, **A decision tree based classification technique for accurate heart disease classification & prediction**, vol. 3, no. 11, pp. 1–4, 2016.
- [8] A. Mustaqeem, S. M. Anwar, A. R. Khan, and M. Majid, **A statistical analysis based recommender model for heart disease patients**, *Int. J. Med. Inform.*, vol. 108, October, pp. 134–145, 2017. <https://doi.org/10.1016/j.ijmedinf.2017.10.008>

- [9] A. N. Safriandono, **Algoritma k-nearest neighbor berbasis forward selection untuk mendiagnosis penyakit jantung koroner**, vol. 3, no. 1, 2017.
- [10] K. Shaikat, N. Masood, S. Mehreen and U. Azmeen, **Data mining in genomics & proteomics dengue fever prediction : a data mining problem**, vol. 6, no. 3, 2015.
- [11] I. Fitriati and M. Ghazali, **Pemodelan pengaruh rata-rata lama sekolah terhadap indeks kedalaman kemiskinan di indonesia menggunakan support vector regression**, April, pp. 100–105, 2017.
- [12] Y. Zhao, C. Zhang, Y. Zang, Z. Wang and J. Li, **A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis**, 2019.
- [13] C. B. C. Latha and S. C. Jeeva, **Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques**, *Informatics Med. Unlocked*, vol. 16, June, p. 100203, 2019. <https://doi.org/10.1016/j.imu.2019.100203>
- [14] H. Situmorang, **Klasifikasi wilayah demam berdarah dengan metode svm di kota medan**, 2015.
- [15] D. A. Anggoro and P. I. Rahmatullah, **The implementation of subspace outlier detection in k-nearest neighbors to improve accuracy in bank marketing data**. *International Journal of Emerging Trends in Engineering Research*, vol. 8, pp. 545–550, February 2020. <http://www.warse.org/IJETER/static/pdf/file/ijeter44822020.pdf> <https://doi.org/10.30534/ijeter/2020/44822020>
- [16] M. Ary and D. A. F. Rismiati, **SATIN – Sains dan teknologi informasi ukuran akurasi klasifikasi penyakit mesothelioma menggunakan algoritma k-nearest neighbor dan backward elimination**, vol. 5, no. 1, 2019.
- [17] J. A. Putra and A. L. Akbar, **Klasifikasi pengidap diabetes pada perempuan menggunakan penggabungan metode support vector machine dan k-nearest neighbor**, vol. 1, no. 2, pp. 47–52, 2016.
- [18] E. Siswanto, Suprapedi and Purwanto, **Metode sample bootstrapping pada k-nearest neighbor untuk klasifikasi status desa**, vol. 14, pp. 13–23, 2018.
- [19] U. Rofiqoh, R. S. Perdana and M. A. Fauzi, **Analisis sentimen tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler indonesia pada twitter dengan metode support vector machine dan lexicon based features**, October, 2017.
- [20] D. A. Nasution, H. H. Khotimah and N. Chamidah, **perbandingan normalisasi data untuk klasifikasi wine menggunakan algoritma k-nn**, vol. 4, no. 1, pp. 78–82, 2019. <https://doi.org/10.24114/cess.v4i1.11458>
- [21] C. A. Hasibuan, M. A. Mukid and A. Prahutama, **Klasifikasi diagnosa penyakit demam berdarah dengue (dbd) menggunakan support vector machine (svm) berbasis gui matlab**, vol. 6, pp. 171–180, 2017.
- [22] T. Setiyorini and R. T. Asmono, **Komparasi metode neural network , support vector machine dan linear regression pada estimasi kuat tekan**, vol. 15, no. 1, pp. 51–56, 2018.
- [23] A. Fouad, H. M. Moftah and H. A. Hefny, **MRI Brain cancer diagnosis approach using gabor filter and support vector machine**, vol. 7, no. 11, pp. 2–6, 2019.
- [24] S. Aulia, S. Hadiyoso and D. N. Ramadhan, **Analisis perbandingan knn dengan svm untuk klasifikasi penyakit diabetes retinopati berdasarkan citra eksudat dan mikroaneurisma**, vol. 3, no. 1, pp. 75–90, 2015. <https://doi.org/10.26760/elkomika.v3i1.75>
- [25] M. Rivki and A. M. Bachtiar, **Implementasi algoritma k-nearest neighbor dalam pengklasifikasian follower twitter yang menggunakan bahasa indonesia**, no. 112, 2017.
- [26] I. A. A. Angreni, S. A. Adisasmita, M. I. Ramli and S. Hamid, **Pengaruh nilai k pada metode k-nearest neighbor (knn) terhadap tingkat akurasi identifikasi kerusakan jalan**, vol. 7, no. 2, pp. 63–70, 2018.
- [27] D. A. Anggoro and W. Supriyanti, **Improving accuracy by applying z-score normalization in linear regression and polynomial regression model for real estate data**, pp. 1–7. <https://doi.org/10.30534/ijeter/2019/247112019>
- [28] S. Jain, S. Shukla and R. Wadhvani, **Dynamic selection of normalization techniques using data complexity measures**, *Expert Syst. Appl.*, 2018.
- [29] H. Gao, Y. -F. Wu and C. -Q. Li, **Performance of normalization method for steel with different strain hardening levels and effective yield strengths**, *Eng. Fract. Mech.*, vol. 218, July, p. 106594, 2019.
- [30] T. M. Manik, **Analisis karakteristik fungsi lagrange dalam menyelesaikan permasalahan optimasi berkendala**, vol. 1, no. 1, pp. 0–7, 2018. <https://doi.org/10.32734/st.v1i1.187>
- [31] A. J. Latipah, G. Ariyanto, R. Hasudungan and N. N. Fatihah, **Classification of butterfly species based on venasi using support vector machine**, vol. 8, pp. 173–176, 2019.
- [32] X. Wang, K. An, L. Tang and X. Chen, **Short term prediction of freeway exiting volume based on svm and knn**, *Int. J. Transp. Sci. Technol.*, vol. 4, no. 3, pp. 337–352, 2015. <https://doi.org/10.1260/2046-0430.4.3.337>