ICIC Express Letters
Volume **15**, Number **1**, January **2021**
 ISSN 1881-803X
pp. **9–18**

COMPARISON OF ACCURACY LEVEL OF SUPPORT VECTOR MACHINE (SVM) AND ARTIFICIAL NEURAL NETWORK (ANN) ALGORITHMS IN PREDICTING DIABETES MELLITUS DISEASE

Dimas Aryo Anggoro and Dian Novitaningrum

Informatics Department
Universitas Muhammadiyah Surakarta
Jl. Ahmad Yani, Pabelan, Kartasura, Surakarta 57162, Jawa Tengah, Indonesia
dimas.a.anggoro@ums.ac.id; L200160139@student.ums.ac.id

Received April 2020; accepted July 2020

Abstract. *Diabetes Mellitus is a chronic disease that occurs because the pancreas does not produce enough insulin or when the body is ineffective in using the insulin. Diabetes causes terrible complications of the disease. As a result, these complications can cause death. Therefore, the right algorithm is needed to predict the occurrence of the disease. The algorithms are expected to be able to work accurately in the classification method. Among them, two algorithms used were Artificial Neural Network (ANN) and Support Vector Machine (SVM). The purpose of this research is to compare the two algorithms, which algorithm can be used precisely in seeing the accuracy value of the data in Diabetes Mellitus. Also, this research has benefits in solving health problems that are used as reference material regarding the percentage of Diabetes Mellitus and can be accurate information material. The results of this study are to get an algorithm that has a high level of accuracy. With the results of the accuracy, it is expected to help someone in knowing the percentage of Diabetes Mellitus. The results of the normalization accuracy obtained were that the SVM method produced 83.54%, while the ANN method was 85.20%.*
Keywords: Accuracy, ANN, Diabetes Mellitus, SVM

1. **Introduction.** Changes in body condition are caused by consuming a diet containing sugar, high fat, protein, and lack of water consumption, plus, more and more people are obese due to lack of exercise. Obesity is one of the few health problems in the world that still becomes a research topic. Many studies have stated that obesity is one of the risk factors that cause type 2 Diabetes Mellitus [8].

Diabetes is a disease that occurs when the body is ineffective in using insulin and is caused by the pancreas not producing insulin properly. Diabetes is also included in several health problems, which are the priority of the four non-communicable diseases and are included in the target of leaders in the world for further action. It was done because of an increase in the number of cases of diabetics during the last few years.

Diabetes Mellitus is a degenerative disease that can be treated with the four pillars of management. Diet is one of the relevant things to be a preventive effort to prevent an increase in blood sugar, which is expected to help control blood sugar in the body effectively [24]. Besides the need for control of diet, it is done, so that blood sugar levels remain well-controlled [26]. Several risks of disease that can be caused due to the influence of fasting blood sugar levels are not controlled, namely hypertension, coronary heart disease, and kidney failure [18].

World Population Data Sheet 2017 revealed that there were 7.5 billion people in the world. In 2017, the International Diabetes Federation (IDF) explained that diabetes was experienced by 425 million people aged 20-79 years. The highest number of people with

DOI: 10.24507/icicel.15.01.9

diabetes was between the ages of 40 and 59 years. Diabetes caused a total of 4 million people to die in 2017.

Diabetes Mellitus is one of the three major groups of non-communicable diseases (PTM) that can cause death [3]. Therefore, we need an algorithm in data mining that precisely predicts the occurrence of the disease. Data mining is a process that is added to look for information that aims to manipulate, change, and recognize patterns of data contained in a particular database [25]. It is expected that the algorithm in data mining can work accurately in the classification method. Among them, two algorithms used are Artificial Neural Networks (ANN) and Support Vector Machine (SVM).

In the problems related to handling non-linear data, the neural network algorithm has an effective capacity but has limitations in overcoming high data disruption [20]. The form of an algorithm that has a complex relationship pattern is used to find patterns in the data between input and output [7]. SVM is a method used to analyze classification and regression patterns that have been entered to predict specific data [13]. The advantage of SVM is the learning process that is carried out quickly. Meanwhile, the drawback is that when a problem occurs, it will be difficult to use if large amounts of data are used [6].

The study was conducted to compare the performance of the classification among SVM, neural network, and Naive Bayes to detect the quality of credit applications in savings and loan cooperatives. It resulted in an accuracy rate of 86.81%, which was higher than the SVM of 85.62%, and Naive Bayes of 83.24% [17].

The subsequent research was conducted to compare the classification of algorithms in neural network disease to detect hepatitis. From this research, the final result in the form of accuracy in SVM was higher when compared to the neural network (multilayer perceptron) algorithm, which was 90.64% from 84.38% [23].

Therefore, the purpose of this research is to compare the two algorithms, which algorithm can be used precisely in seeing the accuracy value of the data in Diabetes Mellitus. In addition, it is expected to be able to develop algorithms and methods that were not previously used to detect the disease. Also, the objective achieved is to get good accuracy values between the two algorithms, including ANN and SVM. The benefits obtained are for patients to be able to prevent the occurrence of Diabetes Mellitus. The output to be developed is to be able to produce a better classification algorithm between ANN and SVM.

2. Method.

2.1. Data collection.

Attributes and variables used in the data mining process were based on the objectives of the research to be achieved. The attributes and variables used are shown in Table 1.

The data used were Diabetes Mellitus data from Kaggle Datasets, whose originality came from the National Institute of Diabetes and Digestive and Kidney Diseases.

2.2. Data pre-processing.

2.2.1. *Data splitting.* Before proceeding to the next stage, the process of splitting should be carried out on the Diabetes Mellitus disease dataset. It was divided into a training dataset and a testing dataset. The training dataset contains label data used to produce another specific model. Meanwhile, the testing dataset is the removal of the label value and then stored with the correct values separately [2]. In this study, splitting was conducted, which was divided into 70% of training data and 30% of testing data.

2.2.2. *Normalization data.* Normalization is a file that consists of elements that are repeated and are applied to being used again [15]. The advantage of this method is that it is more easily processed and includes one method that does not eliminate content by using complex data to change it [1]. The method chosen was the min-max method. This

Table 1. Attributes and variables of patients with Diabetes Mellitus

Variables	Attributes
X1	*Pregnancies*, a presentation of how many times a woman gets pregnant during her life.
X2	*Glucose*, plasma glucose concentration.
X3	*Blood Pressure*, blood pressure from the patient.
X4	*Skin Thickness*, the value used to estimate body fat (mm) measured in the right arm halfway between the olecranon process from the elbow and the acromial process of the scapula.
X5	*Insulin*, a hormone formed in the pancreas that controls blood sugar levels.
X6	*Body Mass Index (BMI)*, a number that determines whether the patient's weight is included in ideal, less, or excessive categories.
X7	*Diabetes Pedigree Function*, an indicator of a family history of diabetes.
X8	*Age*, age of the patient (years).
Y	*Outcome*, class variable (0 or 1). 0 is an indicator of the value of patients who are not sick with diabetes. Meanwhile, 1 is an indicator of the value of patients who are diagnosed with diabetes.

method would produce balanced data with a comparison between the two, before or after the normalization process took place [16]. This method was used to rearrange data from the old range to the new range. The method used is presented in Equation (1).

$$x_{normal} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

where x is the value to be normalized; x_{normal} is the normalized value results; $x_{\min}$ is the minimum value of the entire value to be normalized; and $x_{\max}$ is the maximum value of the entire value to be normalized.

2.3. **Data processing.**

2.3.1. *Support Vector Machine (SVM).* SVM is a part of the machine learning technique that was first introduced in 1992 by Vapnik, Boser, and Guyon, which is used for classification [9]. Simply, the concept of SVM is an attempt to find the best line (hyperplane) contained in the input space and serves to separate the two classes. The best hyperplane could be found by measuring margins and finding the maximum point. Margin is the distance between the hyperplane and the closest pattern of each class. The closest pattern is called a support vector. The essence of learning in SVM is the best effort in finding hyperplane locations. Data denoted as $\vec{x_i} \in R^D$, $i = 1, 2, \ldots, N$. N is the amount of data. Positive class is denoted as $+1$ and negative class as -1. Each data and label are denoted as: $y_i \in \{-1, +1\}$.

Suppose the class can be completely separated by a hyperplane of dimension D. The hyperplane is defined and shown in Equation (2).

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

Data $\vec{x_i}$ classified in negative class (-1) are the data shown in Equation (3).

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

Meanwhile, the data $\vec{x_i}$ classified as positive class $(+1)$ are the data shown in Equation (4).

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

The best margin was calculated by maximizing the distance between the closest hyperplane and pattern. The distance was formulated as $1/\left\|\vec{w}\right\|$ ($\left\|\vec{w}\right\|$ is the normalized value of the weight vector $\vec{w}$). Thus, it could be formulated in Quadratic Programming (QP) used to find the minimum point in Equation (5), by taking account of the constraints in Equation (6).

$$\min_{\vec{w}} \tau(w) = \frac{1}{2} \|w\|^2 \tag{5}$$

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

This optimization could be completed with Lagrange multipliers. The Lagrange method was advanced by Joseph Louis Lagrange (1736-1813). The method works to transform the optimization problem into optimization without constraints. Besides, the method was used to solve the problem with the equation or inequality function contained in the constraint function [14]. The method is presented in Equation (7).

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{N} \alpha_i (y_i ((\vec{x_i} \cdot \vec{w} + b) - 1)) \tag{7}$$

α_i is Lagrange multipliers related to $\vec{x_i}$. α_i is zero or positive. To minimize the value of L to $\vec{w}$ and b and maximize L to α_i, Equation (6) can be modified, and present it in Equation (8) and Equation (9), as follows:

Optimum:

$$\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \cdot \alpha_j \cdot y_i \cdot y_j \cdot \vec{x_i} \cdot \vec{x_j} \tag{8}$$

$$\alpha_i \geq 0 \quad (i = 1, 2, \ldots, l) \quad \sum_{i=1}^{N} \alpha_i \cdot y_i = 0 \tag{9}$$

It can be concluded that the results obtained from α_i are mostly positive. Positive data are called a support vector. Support vector means data that has the closest distance to the hyperplane.

2.3.2. *Artificial Neural Network (ANN).* ANN is one of the networks that represent artificial human neural networks. ANN is a system that can solve problems by changing the structure owned by external and internal information adaptively [7]. The main objective is to research the accuracy of the number of people with Diabetes Mellitus in the coming years based on past data using the ANN method.

The neurons in the neural network are arranged in groups, called layers. Basically, ANN consists of three layers, which have been described in Figure 1, namely:

1) *An input layer*, containing neurons that are used to receive input from outside.

2) *The hidden layer*, consisting of neurons that receive the results of the input layer and then forward it to the next layer.

3) *The output layer*, consisting of neurons that contain the output of the data processing in the hidden layer; the results are then used by the user.

The procedures of the ANN algorithm are:

1) Initialization of weights and biases

Initialize values for weights and biases between 0 and 1.

2) Spread the input value

The input values entered into the front layer are calculated by multiplying the weights contained in the appropriate neurons, then adding them up to produce the appropriate output. This value is explained in Equation (10):

$$I_j = \sum_{i=1}^{N} W_{ij} O_i + \theta_j \tag{10}$$

where W_{ij} is the weight connected from unit i in the previous layer to unit j, O_i becomes the output of unit i before entering the next layer, and θ_j is the bias value of the unit bias that functions to vary the unit activity.

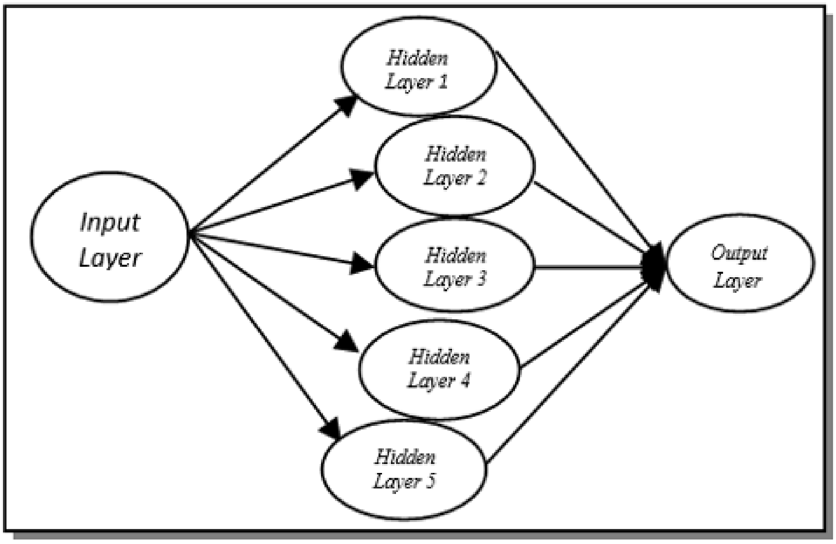

FIGURE 1. Base layers in ANN

Each unit that has entered value then applies the activation function. The activation function is used for the activation of neurons represented by each unit. The Rectified Linear Unit (ReLU) function is used by input and hidden layers. The function produces 0 when output $< x$ and then continues with a value which is below the slope of 1 when $x > 0$. The condition uses the *if* function; when an element is negative, the value is 0, and there is no calculation operation in it. The sigmoid function used by the output in unit j is calculated and shown in Equation (11), namely:

$$O_j = \frac{1}{1 + e^{-I_j}} \tag{11}$$

3) Spread the error value

It increases errors by updating weights and biases to reflect predictions obtained. It optimizes the values to be able to produce new optimization values to do the iteration process. It uses an optimization method called Adaptive Moment Estimation (Adam). This method is a method that is often used for optimization models [27]. Methods that consider the average decay of the quadratic gradient are called the initial and second moments. The gradient is calculated when the bias is towards 0. The bias is corrected by the initial gradient, then calculated, and the weights are updated [21]. For the value of j in the output layer, it is calculated and shown in Equation (12), namely:

$$Error_j = O_j(1 - O_j)(T_j - O_j) \tag{12}$$

where the value of O_j is the output value of unit j, and T_j becomes the target value given. To calculate the error contained in layer j, the weighted number of errors in unit j is calculated. The errors in layer j are shown in Equation (13), which are

$$Error_j = O_j(1 - O_j) \sum_k Error_k W_{jk} \tag{13}$$

where W_{jk} becomes the weight connecting unit j to unit k in the next layer, while $Error_k$ is the error contained in unit k. Next, the weights and biases are updated to reflect the errors caused. The weight value is updated and shown in Equation (14), which is

$$W_{ij} = W_{ij} + (l) Error_j O_i \tag{14}$$

where variable l is a constant value that has a value between 0 to 1. Updating the bias is shown in Equation (15), as follows:

$$\theta_j = \theta_j + (l)Error_j \tag{15}$$

4) Dismissal condition

It is said to stop if it produces a higher accuracy value and is considered to have a good value, when: (a) The initialization of the weights and biases is precisely within range, (b) All weights in the previous training can spread the input to the layer in front of it utilizing output calculated using a sigmoid threshold to a certain extent, (c) The results from the previous classification are wrong, and then spread the error to the previous layer by dividing them into each unit according to the number of errors they have, and (d) End the condition when an error occurs or when the iteration process is finished, if the previous percentage has exceeded the limit.

The ANN method can solve the problem and in accordance with the example. The results in the form of quality and quantity depend on performance for modeling the estimated cost of the sample; the more the number of inputs contained in the trial process, the less error predicted [19]. In ANN testing, it is generally necessary to use normalization for the effectiveness of the developing training model.

2.4. **Evaluation model.** The accuracy value of the algorithm that has been made is obtained from the evaluation process [28]. One of the things that need to be considered in an evaluation is accuracy [4]. In this research, we would explain the comparison of the accuracy values of the two algorithms.

Accuracy. It is a comparison of the size of predictions that are considered true to the overall amount of data that has been evaluated [11]. The formula of accuracy can be seen in Equation (16), namely

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \tag{16}$$

where TP is True Positive; TN is True Negative; FP is False Positive; and FN is False Negative.

3. **Results and Discussion.** The dataset used in this study was data on Diabetes Mellitus, with a total of 768 data consisting of 8 variables and one target. The initial stage done was the process of collecting and processing existing datasets. Referring to the dataset, a comparison was made using the SVM algorithm and the ANN algorithm. The dataset was divided into 70% as a training data of 538 data and 30% as a testing data of 230 data. The next stage was the processing of data that has been processed using the Python programming language in the Anaconda Navigator application, resulting in the percentage of patients suffering from diabetes 34.20% as many as 263 patients and patients who did not suffer from diabetes 65.80% as many as 505 patients. Diabetes Mellitus data were tested using the normalization method and without using the normalization method.

In this dataset, many data had different value intervals. Therefore, the normalization method was needed. The method had an essential role in this study because it could do the processing of minimum and maximum values of each attribute using a range of values from 0 to 1 [16]. Min-max normalization was chosen because it is one of the simplest methods, among others. Besides, min-max can also transform the original data by changing the value of a variable to produce a comparison between the value of data before normalized with the value of data that has been normalized [22]. Before carrying out the normalization process, the following was a comparative analysis through graphs that have been described in Figure 2. The figure compares two variables: Pregnancies, and Outcome.

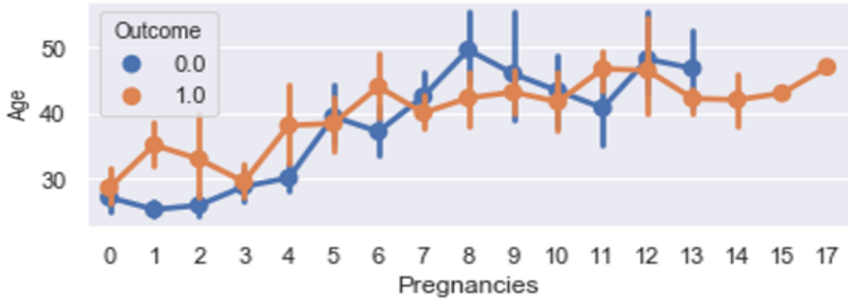

Figure 2. Before normalizing

The following are the accuracy values generated from the SVM algorithm before and after using the normalization used. The accuracy score of the dataset without normalization is 77.92% while the accuracy from dataset using normalization is 83.54%. It implies that the accuracy value of SVM using normalization was higher than the value without normalization, which was 83.54% from 77.92%. This value is generated after processing normalization so that the end result is balanced [16]. These results were obtained by testing the SVM algorithm, starting with the initial values for $\alpha = 0$, $c = 1$, $\varepsilon = 0.001$, $\gamma = 0.5$, and $\lambda = 0.5$. Then, the value of N from the Diabetes Mellitus dataset was entered, in which some classes had a D dimension and were separated through a hyperplane that had a value of +1. From this, the results obtained in the form of margin optimization values were used to maximize the distance between hyperplane with the closest pattern in the form of a kernel model, with a bias of −0.008 and the weight of each attribute used, as shown in Table 2.

Table 2. Weight values in each attribute

No	Weight [ ]	Value
1	Pregnancies	1.362
2	Glucose	1.744
3	Blood Pressure	0.949
4	Skin Thickness	1.355
5	Insulin	1.280
6	BMI	1.265
7	Diabetes Pedigree Function	0.451
8	Age	0.965

It was followed by adding an optimization method named Lagrange multipliers, which has the characteristics of being able to transform problems with constraint into problems without constraint and was used to carry out nonlinear functions in solving optimization problems [14]. Then, it could produce positive values. The positive value came from the hyperplane from the closest distance, which was given the name support-vector.

The next algorithm was the ANN. The results of the algorithm used are shown in Table 3.

For the explanation from Table 3, the results were obtained using the Anaconda application assisted by Tensorflow, which was employed to help the calculation of data and also the Hard layer that runs above Tensorflow, which aimed to speed up the calculation. The calculation started with initiating an initial weight of random and biased values starting at 0. Then, each value entered in the layer in front of it was calculated and multiplied by the weight to produce an output. Each output then applied the activation function that was entered. In this study, the activation function used was ReLU (Rectifier Linear Unit). It was chosen because it has a simple calculation function and can reduce data

Table 3. Accuracy value of ANN

Hidden layer to-	*The number of epochs 300*	*Number of batches_sizes 8*		*The number of epochs 900*	*Number of batch_sizes 9*
	Without normalization	Normalization		Without normalization	Normalization
1	76.02%	77.30%		77.55%	80.36%
2	76.79%	78.06%		78.32%	80.61%
3	77.81%	78.57%		79.08%	80.87%
4	78.32%	78.83%		79.34%	81.12%
5	78.32%	79.34%		79.59%	81.63%
6	78.57%	80.61%		80.87%	81.89%
7	79.85%	80.61%		82.14%	82.14%
8	80.36%	81.12%		82.14%	82.65%
9	81.12%	81.12%		82.91%	83.93%
10	81.38%	81.63%		83.67%	85.20%

processing time significantly when it encounters a large number of neurons. It generated a classification value between 0 to 1. The next process to do was value optimization. The value optimization performed was assisted with an algorithm named Adam. The method was chosen because it is easy to apply, efficient, and appropriate to use in large amounts of data [29]. The results of the optimization obtained weight values and then produced a number of weights that were updated in the iteration. The iteration process was said to stop when the weights and biases had reached the optimum value.

These conditions were interconnected and produced ANN accuracy value at the 10th layer, obtained using the normalization model, which was 85.20%, higher than those without normalization, which was 83.67%. Besides, the more layers, epochs, and batch_size added in the ANN algorithm, the accuracy values obtained for each layer would be increased. It was evidenced in this diabetes dataset that the performance of the ANN algorithm before normalization at the 10th layer was better than the normalized SVM, namely, 83.67% better than 83.54%. In addition, the ANN accuracy value after normalization at the 10th layer was 85.20% higher than SVM after normalizing 83.54% and better than ANN at the 9th layer, which was 83.93%. It indicated that normalization could produce large and accurate values due to the distribution of data, resulting in higher accuracy values compared to iterations that did not add normalization methods.

With the processing steps that have been carried out, the neural network algorithm works with an iterative process that will cause improvements in each weight and bias during the optimization process in each neuron continuously and finally will find the best accuracy of the dataset [17]. The iteration process repeatedly aimed to get the maximum weight value.

The appropriate epoch value in this study could not be known. Therefore, some epoch values are presented to achieve optimum accuracy [29]. Each batch is processed sequentially; samples contained in one batch can be processed in parallel [5]. Small batch sizes tend to produce smaller accuracy results [12]. Whereas, this study required a higher accuracy value so that it is necessary to have a large batch size. The existence of this large batch offers more parallel data, which can be used to improve the efficiency of the iteration process [10]. The large number of layers used in the ANN algorithm can also increase and influence the high accuracy value generated.

4. **Conclusion.** This research was used to compare the accuracy values between the two algorithms, such as SVM and ANN. Both algorithms were used to predict Diabetes Mellitus. The final result in the form of normalization accuracy from the ANN algorithm

had the highest accuracy, which was 85.20%, compared to the SVM algorithm, which had a value of 83.54%. Normalization also played an essential role in the iteration process, because with the addition of data distribution, an accuracy value was generated from the testing process. Besides, in the ANN algorithm, the large number of layers used and added also affected increasing the accuracy value.

Based on this research, there were still shortcomings and additional needs in the development process. The suggestion that can be used for further research is to develop algorithms on other types of datasets using different algorithms, such as K-Nearest Neighbors, Bayesian Classification, as well as, other similar algorithms. Then, from the results obtained, it can be developed to form an information system in predicting Diabetes Mellitus by using other additional programming languages based on websites and desktops. Also, the processing time in the iteration process of the ANN algorithm with a high number of epochs required time long enough. It made ANN had a weakness in processing time. Thus, for further researchers, it is expected to find an appropriate algorithm and time-efficient in terms of data processing.

REFERENCES

[1] A. Adeyemo, H. Wimmer and L. Powell, Effects of normalization techniques on logistic regression in data science, *Journal of Information Systems Applied Research*, vol.12, no.2, p.37, 2019.

[2] D. A. Anggoro and W. Supriyanti, Improving accuracy by applying z-score normalization in linear regression and polynomial regression model for real estate data, *International Journal of Emerging Trends in Engineering Research*, vol.7, no.11, pp.549-555, https://doi.org/10.30534/ijeter/2019/247112019, 2019.

[3] N. C. Aquarista, Differences in characteristics of type 2 diabetes mellitus patients with and without coronary heart disease (Perbedaan karakteristik penderita diabetes melitus tipe 2 dengan dan tanpa penyakit jantung koroner), *Jurnal Berkala Epidemiologi*, vol.5, no.1, pp.37-47, 2017 (in Indonesian).

[4] F. Attal, S. Mohammed, M. Dedebrishvili, F. Chamroukhi, L. Oukhellou and Y. Amirat, Physical human activity recognition using wearable sensors, *Sensors*, vol.15, no.12, pp.31314-31338, 2015.

[5] L. Bottou, F. E. Curtis and J. Nocedal, Optimization methods for large-scale machine learning, *Siam Review*, vol.60, no.2, pp.223-311, 2018.

[6] A. Budianto, D. Maryono and R. Ariyuana, Comparison of k-nearest neighbor (KNN) and support vector machine (SVM) in recognizing the character of motor vehicle plate (Perbandingan k-nearest neighbor (KNN) dan support vector machine (SVM) dalam pengenalan karakter plat kendaraan bermotor), *Jurnal Ilmiah Pendidikan Teknik dan Kejuruan*, vol.11, no.1, pp.27-35, 2018 (in Indonesian).

[7] S. Dewi, Comparison of 5 data mining classification algorithm methods in predicting the success of marketing banking product services (Komparasi 5 metode algoritma klasifikasi data mining pada prediksi keberhasilan pemasaran produk layanan perbankan), *Jurnal Techno Nusa Mandiri*, vol.13, no.1, pp.60-65, 2016 (in Indonesian).

[8] R. N. Fatimah, Type 2 diabetes mellitus (Diabetes melitus tipe 2), *Jurnal Majority*, vol.4, no.5, pp.93-101, 2015 (in Indonesian).

[9] H. Herlina and D. Y. Rakhmawati, Comparison of cross entropy method application and particle swarm optimization in support vector machine (Perbandingan aplikasi metode cross entropy dan particle swarm optimization pada support vector machine), *HEURISTIC: Jurnal Teknik Industri*, vol.15, no.1, pp.17-24, 2018 (in Indonesian).

[10] P. Goyal, P. Dollar, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia and K. He, Accurate, large minibatch SGD: Training imageNet in 1 hour, *arXiv Preprint*, arXiv:1706.02677, 2018.

[11] M. Hossin and M. N. Sulaiman, A review on evaluation metrics for data classification evaluations, *International Journal of Data Mining and Knowledge Management Process*, vol.5, no.2, 2015.

[12] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy and P. T. P. Tang, On large-batch training for deep learning: Generalization gap and sharp minima, *arXiv Preprint*, arXiv:1609.04836, 2017.

[13] M. Kusmira, Sentiment analysis of SIM card re-registration on twitter using SVM and KNN algorithms (Analisis sentimen registrasi ulang kartu SIM pada twitter menggunakan algoritma SVM dan K-NN), *INTI Nusa Mandiri*, vol.14, no.1, pp.105-110, 2019 (in Indonesian).

[14] T. M. Manik, P. Gultom and E. Nababan, Analysis of lagrange function characteristics in resolving problem-based optimization problems (Analisis karakteristik fungsi lagrange dalam menyelesaikan

permasalahan optimasi berkendala), *Talenta Conference Series: Science and Technology (ST)*, vol.1, no.1, pp.37-43, 2018 (in Indonesian).

[15] A. P. Nanda, Management of education and training data for Bank Rakyat Indonesia employees (Pengelolaan data pendidikan dan pelatihan untuk pegawai bank rakyat Indonesia), *Jurnal J-Click*, vol.3, no.2, pp.6-10, 2016 (in Indonesian).

[16] D. A. Nasution, H. H. Khotimah and N. Chamidah, Comparison of data normalization for wine classification using K-NN algorithm (Perbandingan normalisasi data untuk klasifikasi wine menggunakan algoritma K-NN), *Computer Engineering, Science and System Journal*, vol.4, no.1, pp.78-82, 2019 (in Indonesian).

[17] I. T. A. Nur, N. Y. Setiawan and F. A. Bachtiar, Comparison of performance of SVM, neural network, and Naive Bayes classification methods to detect quality of credit application in savings and credit cooperatives (Perbandingan performa metode klasifikasi SVM, neural network, dan naive bayes untuk mendeteksi kualitas pengajuan kredit di koperasi simpan pinjam), *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol.6, no.4, pp.445-450, 2019 (in Indonesian).

[18] L. Nurayati and M. Adriani, Relationship between physical activity and fasting blood sugar levels of type 2 diabetes mellitus patients (Hubungan aktifitas fisik dengan kadar gula darah puasa penderita diabetes melitus tipe 2), *Amerta Nutrition*, vol.1, no.2, pp.80-87, 2017 (in Indonesian).

[19] A. Y. Prathama, A. Aminullah and A. Saputra, ANN (Artificial Neural Network) approach for determining percentage of job weights and estimating the value of structural work at Pratama hospital (Pendekatan ANN (Artificial Neural Network) untuk penentuan prosentase bobot pekerjaan dan estimasi nilai pekerjaan struktur pada rumah sakit pratama), *Jurnal Teknosains*, vol.7, no.1, pp.14-25, 2017 (in Indonesian).

[20] Y. Ramdhani, S. Susanti, M. F. Adiwisastra and S. Topiq, Application of neural network algorithms for cardiotocography classification (Penerapan algoritma neural network untuk klasifikasi kardiotokografi), *Jurnal Informatika*, vol.5, no.1, pp.43-49, 2018 (in Indonesian).

[21] S. V. G. Reddy, K. T. Reddy and V. V. Kumari, Optimization of deep learning using various optimizers, loss functions and dropout, *International Journal of Recent Technology and Engineering*, vol.7, pp.448-455, 2018.

[22] U. Rofiqoh, R. S. Perdana and M. A. Fauzi, Sentiment analysis of user satisfaction levels of Indonesian cellular telecommunications service providers on Twitter with the support vector machine and lexicon based features method (Analisis sentimen tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler Indonesia pada twitter dengan metode support vector machine dan lexicon based features), *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol.1, no.12, pp.1725-1732, 2017 (in Indonesian).

[23] S. Saputra, Comparison of neural network based algorithms in detecting hepatitis (Komparasi algoritma berbasis neural network dalam mendeteksi penyakit hepatitis), *Faktor Exacta*, vol.10, no.1, pp.40-49, 2017 (in Indonesian).

[24] S. Soegondo, *Integrated Diabetes Mellitus Management for Doctors and Diabetes Educators*, Balai Penerbit FKUI, Jakarta, 2015.

[25] W. Supriyanti and N. Puspitasari, Implementation of forward selection feature selection technique in data mining classification algorithm for prediction of study period of Indonusa Surakarta polytechnic students (Implementasi teknik seleksi fitur forward selection pada algoritma klasifikasi data mining untuk prediksi masa studi mahasiswa politeknik Indonusa surakarta), *Jurnal Informatika*, vol.4, no.2, pp.49-54, 2018 (in Indonesian).

[26] S. Susanti and D. N. Bistara, Relationship of diet and blood sugar levels in patients with diabetes mellitus (Hubungan pola makan dengan kadar gula darah pada penderita diabetes mellitus), *Jurnal Kesehatan Vokasional*, vol.3, no.1, pp.29-34, 2018 (in Indonesian).

[27] M. A. D. Suyudi, E. C. Djamal and A. Maspupah, Stock price prediction using the recurrent neural network method (Prediksi harga saham menggunakan metode recurrent neural network), *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, pp.A-33-A-38, 2019 (in Indonesian).

[28] L. A. Utami, Sentiment analysis public opinion forest fire news through comparison of support vector machine and k-nearest neighbor based particle swarm optimization (Analisis sentimen opini publik berita kebakaran hutan melalui komparasi algoritma support vector machine dan k-nearest neighbor berbasis particle swarm optimization), *Jurnal Pilar Nusa Mandiri*, vol.13, no.1, pp.103-112, 2017 (in Indonesian).

[29] M. S. Wibawa, Effect of activation, optimization and number of epoch functions on artificial neural network performance (Pengaruh fungsi aktivasi, optimisasi dan jumlah epoch terhadap performa jaringan saraf tiruan), *Jurnal Sistem dan Informatika (JSI)*, vol.11, no.2, pp.167-174, 2017 (in Indonesian).