Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 3 (2023) 652 - 656 ISSN Media Electronic: 2580-0760

Comparative Analysis of Support Vector Machine and Perceptron In The Classification of Subsidized Fuel Receipts

Jaka Tirta Samudra[1], Rika Rosnelly[2], Zakarias Situmorang[3]
[1,2,3]Computer Science, Potensi Utama University
[1]jakatirta135@gmail.com, [2]rikarosnelly@gmail.com, [3]zakarias65@yahoo.com

Abstract

Currently, fuel oil is one of the important factors for the community and even a country on this earth to utilize this natural gas fuel for daily use as the main use and also by increasing the community's need for fuel oil, starting from the type of pertalite, Pertamax, to diesel for various types of private vehicles to tourists who use it for vehicles and even daily staples used. But there are several factors that cause this fuel problem, there is a factor of time and usage time, which is certain that one day it will expire and its capacity in a country, even if the country runs out of fuel, will make requests to other countries and also obstacles to supplying this fuel oil to the public. which is the main fuel from the Pertamina government agency which has begun to limit purchases for this fuel oil to certain circles by marking the types of subsidies or not subsidies that must be controlled by the government in limiting purchases for the public. In dealing with solving problems from the perspective of ownership or even utilization, there are limits to owning fuel, and not everyone has to have a lot or even too much. That way, to get everyone in the community who deserves to receive this fuel limit to the maximum by designing subsidized fuel oil revenues. In solving the problem of dividing fuel revenue, which is good for filling revenue, it can be solved by using machine learning, namely data mining itself can help in completing subsidized fuel receipts without being excessive for the community so that they can be controlled and managed for their purchases. In building a fuel oil reception design, it can be grouped into a classification model that uses SVM and perceptron which uses the activation function of the sigmoid to get the final result of accuracy where getting the average value of 5-fold, 10-fold, 20-fold is accuracy. is 90.0%, the F1 value is 85.6%, the precision value is 87.6%, and the recall value is 90.0%.

Keywords: data mining; SVM; perceptron; classification; fuel

1. Introduction

Indonesia is one of the countries in the world where most of the domestic demand for natural gas is still supported by the strength of oil imports. Oil imported by Indonesia itself is used by the public as fuel which is commonly referred to as fuel oil or the term BBM. Fuel itself as a commodity that plays an important role for the economy in Indonesia to support the nation's future economic growth. From the existing wealth and oil resources that were owned by Indonesia itself during the New Order government and also the desire of the government from all over so that more and more people could benefit from the abundance of petroleum, the fuel subsidy policy. set. Subsidized oil products themselves could cause oil prices to be cheaper than economic prices, even though of all these prices the production costs incurred were very high but at that time the Indonesian government itself was still financially capable and also phenomena like this would continue until the end of the year. the oil deficit approached Indonesia as it is today.

With the increasing production of two-wheeled vehicles to become more for everyday vehicles such as motorcycles or cars, this has not stopped large vehicle manufacturers such as Honda, Suzuki or Yamaha from always increasing sales productivity in these vehicles. sector. [1]. However, in the increasing use of motorbikes or cars, it is also important to identify the fuel used [2] that is needed to continue refueling [3], [4] so that the obligation as a vehicle owner is to refuel. in locations that already exist, for example in Pertamina which has agency rights for the government, there are also others, such as cells, AKA, or fuel oil that are sold on the roadside [5].

How about consumers having to get a supply of available fuel oil for vehicle needs, especially when fuel oil is scarce [6]. Thus, the government's policy is to issue appropriate fuel needs from the public to get subsidized fuel, not all of which get this fuel

Accepted: 30-11-2022 | Received in revised: 25-03-2023 | Published: 02-06-2022

opportunity. In contrast to owning a vehicle that supplies electric vehicles, it does not require consumption of fuel oil, therefore with fuel inflation the government limits the use of subsidized fuel in purchasing from predetermined aspects, for example from a high CC vehicle or from a license plate. vehicle owner accordingly. grouping for specific BBM beneficiary analysis for various subsidies [7].

Data mining itself is one of the sciences in the field of computer science which can be in the form of grouping to be used in solving problems such as predictions, such as classification or regression [8] [9]. According to him, data mining itself can be used for grouping in analyzing income from various types of subsidized fuel income classifications that can be accepted by comparing new data obtained from refueling queues in Pertamina's area or commonly known as gas stations. [10] also from the origin of new data (data from the results of a refueling survey) and also classifies it in the form of subsidized fuel oil income classification [11].

Machine learning itself can also be defined as a type of application on a computer and also a mathematics learning algorithm that is adopted by following learning concepts that originate from the form of data and also to produce predictions in the future with certainty. Machine learning itself is also associated as a branch of data mining learning which has been widely used to solve a problem related to classification by utilizing neural network models [12], [13]. Based on the number of hidden layers, ANN models are divided into two types, namely single layer ANN which has one hidden layer and multi-layer ANN which has more than one hidden layer [14], [15].

The learning process in question is an effort to gain intelligence through two stages, namely training and testing. The field of machine learning is concerned with the question of how to build computer programs to improve automatically based on experience [16].

The Support Vector machine (SVM) is a method that follows to perform self-classification which was first introduced by Vapnik in 1998. Where on the basis of this concept for modeling the method so that it can work properly in defining the boundary between two different classes. different from the maximum distance from the closest data [17]. To be able to carry out the execution of obtaining by getting the maximum limit between two different classes, it must be formed into a hyperplane or the best dividing line in the input space or space obtained by measuring the hyperplane margins and also to find the maximum point. The margin itself is the distance between the hyperplane and the closest points of each class [18]. For the closest point itself, this is what is referred to as a support vector machine (SVM). Support vector machine or SVM can also classify separate data in linear and non-linear ways [19].

Single layer perceptron is a network that has one layer of connected weighted layers. It also has advantages, among others, being able to obtain necessary knowledge from uncertainties in the data used to carry out tests of these data which can also be made through generalization of results to extraction [20]. The single layer perceptron method itself is the most basic method for machine learning cases and also the simplest at the case level. The single layer perceptron itself is a feedforward type, which is a type of NN where neurons in one layer can only connect to neurons in a different layer. Therefore in this case the Single Layer perceptron network method is included in supervised learning because the learning method is carried out by studying examples that are known to have input and output [21]. The network will be trained with a set of data examples that are known input and output. During the learning process, the network will adjust the weight values to produce the desired output [22].

In the process of increasing the results of the desired accuracy to be analyzed for the results is how to use the model from the SVM and perceptron by increasing the use of the function of the sigmoid [23]. Where is the activation function that plays a very important role in activating each neuron in the artificial neural network and also determines the output of the artificial neural network. Choice of activation functions that can be used in SVM and perceptron models such as binary sigmoid function and hyperbolic tangent. the selection of the right activation function will certainly affect the performance of the results, such as in terms of speed of data processing or also the level of accuracy of the results obtained [24].

This study compares the classification results of subsidized fuel oil receipts using the SVM and perceptron algorithms between sigmoid functions by measuring the values of accuracy, F1, precision, and recall using cross validation. with the number of folds of 5, 10, and 20, the average value of each value is calculated to see which modeling algorithm is better.

2. Research Methods

The research method that can be used to classify refueling is by using dataset management from SPPBU by using two algorithms, namely support vector machine (SVM) and perceptron which uses the sigmoid function shown in Figure 1.

This study compares the algorithm with the SVM and Perceptron models using the sigmoid function on the Pertamina refueling queue dataset using the Orange 3.30 application, with the shape of the model as shown in Figure 2

DOI: https://doi.org/10.29207/resti.v7i3.4731

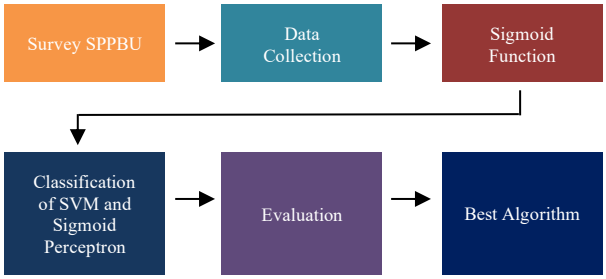

Figure 1. System Design

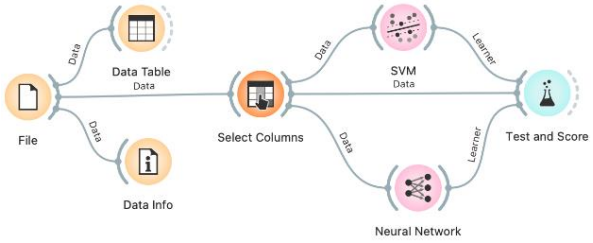

Figure 2. Classification Model

The data used in this study is primary data obtained directly from records that have filled one of Pertamina's fuel oil in the form of data on 255 oil fillers from motor vehicles, cars, buses or trucks which are processed into the features needed in this study such as vehicle, vehicle type, vehicle brand, vehicle cc, type of refueling, amount of filling, vehicle plate series, and filling which is shown in Table 1 where there are 10 sample data sets used.

Table 1. Sampel Data Set

No	F-1	F-2	F-3	F-4
1	Car	Personal	Honda	1200 Cc
2	Car	Personal	Suzuki	1300 Cc
3	Motorcycle	Personal	Honda	150 Cc
4	Motorcycle	Personal	Yamaha	125 Cc
5	Motorcycle	Personal	Yamaha	110 Cc
6	Car	Tourism	Yamaha	3000 Cc
7	Car	Tourism	Yamaha	3500 Cc
8	Car	Personal	Hyundai	3500 Cc
9	Car	Personal	Hyundai	1500 Cc
10	Car	Tourism	Hyundai	3500 Cc
11	Car	Tourism	Toyota	3500 Cc
12	Car	Personal	Daihatsu	1400 Cc
13	Car	Personal	Mitsubishi	3000 Cc
14	Motorcycle	Personal	Honda	110 Cc
15	Motorcycle	Personal	Honda	125 Cc
16	Motorcycle	Personal	Honda	150 Cc
17	Motorcycle	Personal	Honda	150 Cc
18	Car	Personal	Toyota	3800 Cc
19	Car	Personal	Honda	1500 Cc
20	Car	Personal	Toyota	1550 Cc

Table 1. Sample Data Sets - Connections

No	F-5	F-6	F-7	T
1	Pertalite	11 Liter	BK	Subsidy
2	Pertamax	10 Liter	BB	Not Subsidized
3	Pertalite	2 Liter	BK	Subsidy
4	Pertalite	1 Liter	BK	Subsidy
5	Pertalite	1.5 Liter	BK	Subsidy
6	Pertalite	18 Liter	BK	Subsidy
7	Pertalite	15 Liter	BB	Subsidy
8	Pertalite	10 Liter	BK	Subsidy
9	Pertalite	5 Liter	BB	Subsidy
10	Pertalite	20 Liter	BB	Subsidy
11	Pertalite	10 Liter	BK	Subsidy
12	Pertalite	20 Liter	BK	Subsidy
13	Pertamax	17 Liter	BB	Not Subsidized
14	Pertalite	1.5 Liter	BK	Subsidy
15	Pertamax	2 Liter	BK	Subsidy
16	Pertalite	3 Liter	BK	Subsidy
17	Pertalite	1 Liter	BL	Subsidy
18	Pertamax	14 Liter	BK	Subsidy
19	Pertalite	16 Liter	BK	Subsidy
20	Pertalite	4 Liter	BK	Subsidy

F-1 : The type of vehicle to distinguish the attributes of the type of motorbike is only used 2, namely cars and trains; F-2 : Type of vehicle to differentiate in the ownership vehicle type section from private (self-owned) and private (public) types; F-3 : Vehicle brands to show existing brands such as Honda, Suzuki, Yamaha and others; F-4 : Cc or available vehicle engine capacity, which varies from the smallest 110 cc to more than 3000 cc; F-5 : Filling fuel with its type based on several categories such as pertalite, Pertamax and diesel fuel; F-6 : The amount of oil refueling is based on filling at gas stations with various fillings from 1 liter to more than 5 liters; F-7 : The vehicle plate series is based on those in the province of North Sumatra, namely BK, BB, BL, and even from outside North Sumatra, such as D; and T : The target for refueling is the main factor in determining whether subsidies are received or not.

By having a target for each feature is as follows:

F-1 = Car (2), Motorcycle (1)
F-2 = Personal (2), Tourism (1)
F-3 = Honda (5), Yamaha (4), Suzuki (3), Toyota (2), Lainnya (1)
F-4 = > 4000 (5), > 2500 (2), > 1000 (3), > 250 (2), > 0 (1)
F-5 = Pertalite (3), Pertamax (2), Solar (1)
F-6 = > 10 (5), > 6 (4), > 4 (3), > 2 (2), > 0 (1)
F-7 = BL (3), BB (2), Other (1)
T = Subsidy (1), Not Subsidized (0)

The classification process with the SVM and perceptron algorithms uses a number of folds of 5, 10, 20 for cross validation, sigmoid activation function. The evaluation of the classification can be measured by going through

DOI: https://doi.org/10.29207/resti.v7i3.4731

the values of classification accuracy, F1, precision and recall which are calculated using equations (1) to equations (4) as follows [25]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FN} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 = 2.\frac{Precision.Recall}{Precision+Recall} \quad (4)$$

TP is the True Positive, TN is the True Negative, FP is the False Positive, and FN is the False Negative.

Regarding the data to obtain evaluation results that can be used, one of the types used is the confusion matrix which can be obtained independently from the results of the types of accuracy, F-score, precision, and recall as well as from the values on the ROC curve to obtain the size of the AUC value. That way, the greater the area under the curve (AUC) of the results, the more results are obtained, meaning that the prediction results are better in the Confusion Matrix as shown in Table 2.

Table 2. Confusion Matrix

Actually	Prediction	
	True	False
True	TP	FN
False	FP	TN

3. Results and Discussions

From each classification parameter that has been trained and can be inputted into the SVM and perceptron models in Figure 2 with a maximum number of epochs of 100. The accuracy, F1, precision and recall values obtained from the SVM and perceptron classification results using a variation of the sigmoid function using 5-Fold, 10-Fold, and 20-Fold Cross Validation as shown in Table 3

Table 3. Cross Validation Results

Algoritma	Fold	Cross Validation			
		Accuracy	F1	*Precision*	*Recall*
SVM	5	88,3	84,3	80,6	88,3
	10	88,7	85,1	82,9	88,7
	20	87,9	84,6	82,3	87,9
	Average	88,3	84,6	82,0	88,3
Perceptron	5	90,2	86,0	91,2	90,2
	10	89,8	85,0	80,7	89,8
	20	90,2	86,0	91,2	90,2
	Average	90,0	85,6	87,6	90,0

From Table 3 it can be seen that the SVM and perceptron algorithm models for classification with the sigmoid function, at the highest evaluation the 5-fold cross validation produces an accuracy value of 90.2% for the perceptron, an F1 value of 86.0% for the perceptron, a precision value of 91.2% for the perceptron, and a recall value of 90.2% for the peceptron. In the highest evaluation of the 10-fold cross validation the sigmoid function produces an accuracy value of 89.8% for the perceptron, an F1 value of 85.1% for SVM, a precision value of 82.9% for SVM, and a recall value of 89.8% for perceptrons. In the highest evaluation of 20-fold cross validation, the sigmoid function produces an accuracy value of 90.2% for the perceptron, an F1 value of 86.0% for the perceptron, a precision value of 91.2% for the perceptron, and a recall value of 90.2 % for perceptrons. From all these cross validation evaluations, the highest average accuracy value was 90.0%, the F1 value was 85.6%, the precision value was 87.6%, and the recall value was 90.0%.

The following results of the classification that have been tested can be seen in the Confusion Matrix display along with the 5-Fold validation for the SVM and Perceptron models shown in Figure 3.

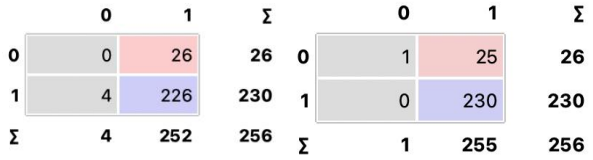

Figure 3. Confusion Matrix 5-Fold Model SVM and Perceptron

The following results of the classification that have been tested can be seen in the Confusion Matrix display along with the 10-Fold validation for the SVM and Perceptron models shown in Figure 4.

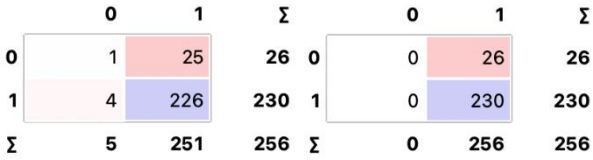

Figure 4. Confusion Matrix 10-Fold Model SVM and Perceptron

The following results of the classification that have been tested can be seen in the Confusion Matrix display along with the 20-Fold validation for the SVM and Perceptron models shown in Figure 5.

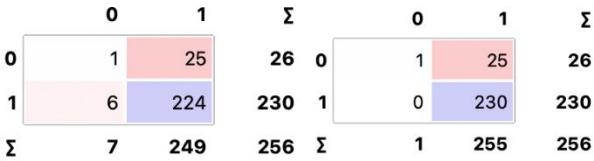

Figure 5. Confusion Matrix 20-Fold Model SVM and Perceptron

4. Conclusion

From the results of the conclusion when getting the results of the evaluation that has been carried out on the classification test using the SVM and perceptron algorithms with the use of the sigmoid function itself, it can be concluded that using the model from the perceptron algorithm gets the best classification to classify subsidized fuel oil reception datasets where it can be seen that the use of 5-fold, 10-fold, and 20-fold

DOI: https://doi.org/10.29207/resti.v7i3.4731

cross validation for the perceptron model is better for the overall average value of the combined values of accuracy, F1, precision, recall is 88.3% compared to the average value of the SVM model -the average is 85.8%, for the largest value of the classification results is 90.0% in the perceptron model and the smallest value is 82.0% in the SVM model so to carry out this classification the perceptron gets a good level of value.

From the results of observations from this study, it was lifted for acceptance from the results of testing the use of the perceptron algorithm modeling itself, it did get good and dominant results for the end result, rather than using the SVM algorithm model itself in research that was able to train data in acceptance of subsidized fuel results or not on relative data.

However, from this research to improve the quality and development of suggestions that are referred to new research in development, it is hoped that further research can create new models to test accuracy other than using the SVM and perceptron models or provide a larger number in the k-fold, and test comparison of these two algorithms in the implementation of different algorithms.

References

[1] A. A. Nur, D. Wiryawan, and M. Al Amrie, "Kepuasan Konsumen Astra Motor Honda Tanjung Selor Terkait Pelayanan Showroom," *Plakat (Pelayanan Kpd. Masyarakat)*, vol. 2, no. 2, p. 109, 2020.

[2] B. Hermawan and A. Faris Indriya Himawan, "Pembuktian Isu Kehilangan BBM Head Truck Depo Transhipment Pada PT. Berkah Multi Cargo Logistic," *Etn. J. Ekon. dan Tek.*, vol. 1, no. 5, pp. 357–363, 2022.

[3] F. A. Akbar, "Pembuatan Aplikasi Berbasis Android Untuk Pengambilan Bbm (Bahan Bakar Minyak) Dengan Menggunakan Fitur Qr Code," *J. Inform. dan Sist. Inf.*, vol. 1, no. 2, pp. 422–431, 2020.

[4] R. M. Kosanke, "Analisis Kelangkaan Bbm Premium Di Padangsidimpuan Perspektif Hukum Positif Dan Hukum Ekonomi Syariah," *J. Huk. Ekon.*, vol. 7, no. 1, pp. 56–76, 2019.

[5] D. Putu *et al.*, "Implementasi Pasal 23 Ayat (1) Undang-Undangnomor 22 Tahun 2001 Tentang Minyak Dan Gas Bumiterkait Penjualan Bahan Bakar Eceran Ilegal Di Desa Sangsit Kabupaten Buleleng," *Komunitas Yust. Univ. Pendidik. Ganesha*, vol. 4, no. 2, pp. 312–321, 2021.

[6] G. R. Hrp and N. Aslami, "Analisis Dampak Kebijakan Perubahan Publik Harga BBM terhadap Perekonomian Rakyat Indonesia," *J. Ilmu Komputer, Ekon. dan Manaj.*, vol. 2, no. 1, pp. 1464–1474, 2022.

[7] P. Inflasi, S. Bbm, and D. A. N. Anggaran, "Pengaruh Inflasi, Subsidi Bbm, Dan Anggaran Infrastruktur Terhadap Pertumbuhan Ekonomi," *Manaj. Kewirausahaan*, vol. 18, no. 01, pp. 29–38, 2021.

[8] B. Yanto, J. Jufri, A. Lubis, B. H. Hayadi, and E. Armita, NST, "Klarifikasi Kematangan Buah Nanas Dengan Ruang Warna Hue Saturation Intensity (Hsi)," *INOVTEK Polbeng - Seri Inform.*, vol. 6, no. 1, p. 135, 2021.

[9] K. Hulliyah, "Predicting Airline Passenger Satisfaction with Classification Algorithms," *IJIIS Int. J. Informatics Inf. Syst.*, vol. 4, pp. 82–94, Mar. 2021.

[10] K. S. Prawiro and D. Agfazar, "Analisis Antrian Sepeda Motor pada SPBU Tanah Merdeka Menggunakan Simulasi Promodel," *Bull. Appl. Ind. Eng. Theory*, vol. 2, no. 1, pp. 28–31, 2020.

[11] R. I. Liperda, I. K. Hardianti, I. N. Widyah, N. A. Fadjri, and R. R. Agustin, "Simulasi-Optimasi Sistem Transportasi Penentuan Kebutuhan Truk Tangki Pada Proses Distribusi BBM : Studi Kasus TBBM Plumpang," vol. 9, no. 2, 2022.

[12] A. Wanto *et al.*, "Forecasting the Export and Import Volume of Crude Oil, Oil Products and Gas Using ANN," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019.

[13] B. H. Hayadi, I. G. I. Sudipa, and A. P. Windarto, "Model Peramalan Artificial Neural Network pada Peserta KB Aktif Jalur Pemerintahan menggunakan Artificial Neural Network Back-Propagation," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 11–20, 2021.

[14] H. A. G. Al-kaf, K. S. Chia, and N. A. M. Alduais, "A comparison between single layer and multilayer artificial neural networks in predicting diesel fuel properties using near infrared spectrum," *Pet. Sci. Technol.*, vol. 36, no. 6, pp. 411–418, 2018.

[15] B. Yanto, B. -, J. -, and B. H. Hayadi, "Indentifikasi Pola Aksara Arab Melayu Dengan Jaringan Syaraf Tiruan Convolutional Neural Network (Cnn)," *JSAI (Journal Sci. Appl. Informatics)*, vol. 3, no. 3, pp. 106–114, 2020.

[16] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian J. Comput. Inf. Technol.*, vol. 5, no. 1, pp. 75–82, 2020.

[17] I. C. R. Drajana, "Metode Support Vector Machine Dan Forward Selection Prediksi Pembayaran Pembelian Bahan Baku Kopra," *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 116–123, 2017.

[18] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, "Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB," *J. Inform. Upgris*, vol. 7, no. 1, pp. 36–40, 2021.

[19] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm)," *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020.

[20] A. N. Iffah'da and Anita Desiani, "Implementasi Algoritma K-Nearest Neighbor (K-NN) dan Single Layer Perceptron (SLP) Dalam Prediksi Penyakit Sirosis Biliari Primer," *J. Ilm. Inform.*, vol. 7, no. 1, pp. 65–74, 2022.

[21] E. Yuliani, A. N. Aini, and C. U. Khasanah, "Perbandingan Jumlah Epoch Dan Steps Per Epoch Pada Convolutional Neural Network Untuk Meningkatkan Akurasi Dalam Klasifikasi Gambar," *J. Inf. Politek. Indonusa Surakarta*, vol. 5, no. 2442–7942, pp. 2–6, 2019.

[22] Y. U. Budiman *et al.*, "Identifikasi Kata Benda Dan Bukan Kata Benda Menggunakan," vol. 01, no. 5, pp. 759–768, 2022.

[23] E. Supriyadi, "Prediksi Parameter Cuaca Menggunakan Deep Learning Long-Short Term Memory (Lstm)," *J. Meteorol. dan Geofis.*, vol. 21, no. 2, p. 55, 2021.

[24] S. Susilawati and M. Muhathir, "Analisis Pengaruh Fungsi Aktivasi, Learning Rate Dan Momentum Dalam Menentukan Mean Square Error (MSE) Pada Jaringan Saraf Restricted Boltzmann Machines (RBM)," *J. Informatics Telecommun. Eng.*, vol. 2, no. 2, p. 77, 2019.

[25] H. Dalianis, "Evaluation Metrics and Evaluation," *Clin. Text Min.*, no. 1967, pp. 45–53, 2018.

DOI: https://doi.org/10.29207/resti.v7i3.4731