

## MACHINE LEARNING PREDICTION OF TIME SERIES COVID-19 DATA IN WEST JAVA, INDONESIA

Intan Nurma Yulita<sup>1</sup>, Afrida Helen<sup>2</sup>, Mira Suryani<sup>3</sup>

<sup>1,2,3</sup>Department of Computer Science, Universitas Padjadjaran

email: intan.nurma@unpad.ac.id<sup>1</sup>, helen@unpad.ac.id<sup>2</sup>, mira.suryani@unpad.ac.id<sup>3</sup>

### Abstract

In 2019, the COVID-19 pandemic appeared. There have been several efforts to curb the spread of this virus. West Java, Indonesia, employs social restrictions to prevent the spread of this disease. However, this method destroyed the economy of the people. If no instances were detected in the region, the World Health Organization (WHO) authorized the social restrictions to be relaxed. If the government lifts the social limitation, the decision must also consider the potential of future confirmed instances. By utilizing machine learning, it is possible to forecast future data. This work utilized the following algorithms: linear regression (LR), locally weighted learning (LWL), multi-layer perceptron (MLP), radial basis function regression (RBF), and support vector machine (SVM). The study investigated daily new instances of COVID-19 in West Java, Indonesia, from March 2, 2020, to October 15, 2020. The RBF algorithm was the best in this investigation. Mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and relative absolute error (RAE) were 48.85, 89.73, 88.67, 62.99, and 60.88, respectively. The RBF prediction model may be proposed to the government of West Java for assessing data on COVID-19 instances, particularly in social restriction management. It is anticipated that West Java would have a minimum of 275 new cases every day for the following 30 days beginning on October 16, 2020. Consequently, the easing of societal limitations requires careful consideration.

**Keywords :** COVID-19, Machine Learning, Indonesia, Artificial Intelligence, Artificial Neural Networks

---

**Received:** 10-02-2023 | **Revised:** 09-05-2023 | **Accepted:** 10-05-2023

DOI: <https://doi.org/10.23887/janapati.v12i2.58505>

---

### INTRODUCTION

In China, severe acute respiratory syndrome (SARS) was identified for the first time in 2002 [1]. After 18 years, the SARS virus reappeared. In China, the novel coronavirus resurfaced in December 2019. SARS-CoV-2, a coronavirus with a novel form, has affected several countries throughout the world. The spread of COVID-19 has rendered several nations defenseless [2]. Like in 2002, SARS-CoV-2 induced fear and panic [3]. The coronavirus has been defeated through a variety of initiatives and tactics. Nevertheless, the pathogen continues to spread. COVID-19 has claimed a substantial number of lives. In addition to its consequences on health, the COVID-19 epidemic has wrecked the economics of several nations. Since a lockdown policy was implemented, the global economy has stopped [4]. The scenario might result in a worldwide crisis. As a result, several nations have eased their lockdown restrictions and allowed companies to reopen in response to the CsocialOVID-19 outbreak. Indonesia announced

its first occurrence in West Java on March 2, 2020 [5]. Indonesia employs extensive societal restrictions to prevent the development of this disease. Similar to other locations, this limitation led to a decline in economic activity [6]. As a result, West Java loosened regulations. However, the World Health Organization (WHO) stipulates conditions under which a region may relax its limits [7]. One of them is the absence of instances in the region.

According to the present data trend, West Java had zero instances on several occasions in May but reported an increase in July 2020 [8]. If the government intends to loosen limits, it must consider the possibility of future instances. Using machine learning, it is possible to forecast future data [9]. This is a component of artificial intelligence [10], [11]. Its intelligence can provide more accurate projections of the number of new COVID-19 cases based on previously supplied input data. Time series prediction is a frequent use of linear regression. Yip et al. employed this technique to detect energy theft [12]. Consumers

purposefully destroyed the electrical meter, resulting in a theft. Using linear regression, power usage trends may be observed. If there is a deviation from the normal pattern of energy usage, this is considered stealing. Similar research is conducted to predict the demand for heat [13] and the stock market [14]. In contrast, locally weighted learning (LWL) uses linear regression to give surrounding data weight. It has been widely implemented for time series prediction, such as temperature prediction [15] and short-term forecasting of urban traffic flow [16]. There are hardly many real-world issues where the link between independent and dependent variables is evident. It demonstrates that linear regression might have problems recognizing it, resulting in a huge prediction error. The MLP algorithm can be used to tackle this problem. It is an artificial neural network (ANN) [17]. Kumar et al. examined it for the prediction of nanofluid viscosity. It resulted in a root mean square error of 0.018 [18]. Utilizing six hidden layers decreased the inaccuracy. Another ANN technique is the radial basis function (RBF). This approach is capable of modeling nonlinear data. Another ANN technique is the radial basis function (RBF). This approach is capable of modeling nonlinear data [19]. MLP and RBF are the most prevalent ANN models. These approaches can readily identify patterns in complicated data, but not in high-dimensional ones. In addition, it is tough to deal with overfitting situations. This deficiency can be remedied with a support vector machine (SVM).

Multiple studies have anticipated COVID-19 time series data. Ardabili et al.

anticipated cumulative statistics for the United States, Italy, China, Iran, and Germany [20]. The data was collected from January through March 2020. Predictions were made utilizing MLP and an adaptive network-based fuzzy inference system (ANFIS) in this study. This study suggested comparing many algorithms within each nation. This is due to the fact that each nation has distinct epidemic patterns. There are no areas that are identical. Additionally, Wang et al. forecasted cumulative statistics for a number of nations [21], including Indonesia. This study utilized data from January to June 2020, despite the fact that Indonesia reported its instances only in March 2020. A logistic model was used to make the prediction [21]. Ahmad et al. reached a different conclusion than Ardabili et al. [22]. They thought that the first epidemic curve of one nation might be used as a benchmark for other countries, despite the fact that the pandemic stages in each country are distinct. This research examines the conditions necessary for loosening societal restraints. This study examines the daily number of new cases rather than the total number. The social border policy of Indonesia is province-based, not national. The study only looks at provinces. This study focuses on the province of West Java. This province is the first in Indonesia to report instances of COVID-19. Several machine learning techniques, including linear regression (LR), locally weighted learning (LWL), multi-layer perceptrons (MLP), radial basis function regression (RBF), and support vector machines (SVM), are employed for prediction in this work.

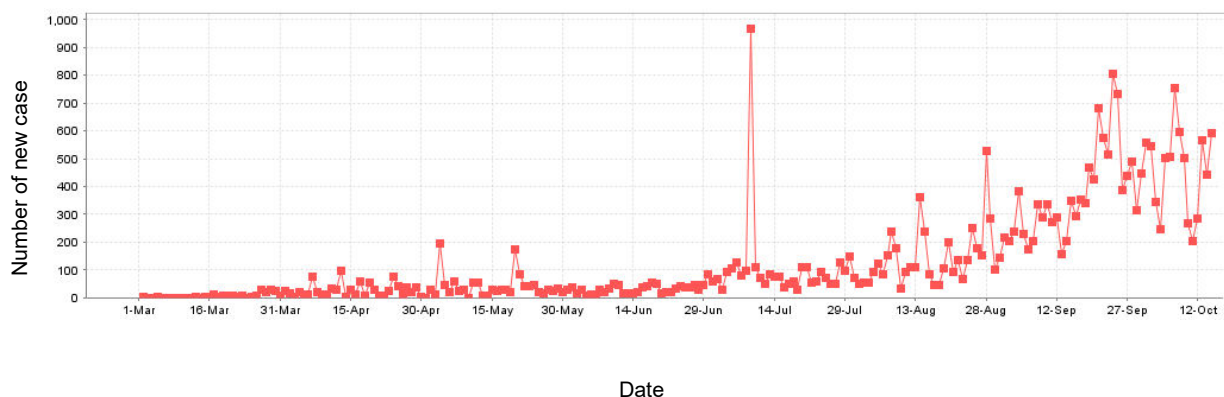


Figure 1. Confirmed New Cases of COVID-19 in West Java from March 2nd to October 15th, 2020

**METHOD**

This research collected information from the website <https://pikobar.jabarprov.go.id/>. It

only assessed new instances daily between March 2 and October 15, 2020. The daily data chart is depicted in Figure 1. The highest point

was reached on July 9, 2020. On that date, there were 965 confirmed cases. It increased tenfold from the previous day. Figure 2 depicts the study's approach. This study included just two characteristics. They were the date and daily total of new instances. These two properties were retrieved using a mix of remapping, lag selection, and both. This processing produced 21 types of attributes, namely:

- caseNew,
- DayOfWeek,
- Weekend,
- Timestamp-remapped,
- Lag\_caseNew-1,
- Lag\_caseNew-2,
- Lag\_caseNew-3,
- Lag\_caseNew-4,
- Lag\_caseNew-5,
- Lag\_caseNew-6,
- Lag\_caseNew-7,
- Timestamp-remapped^2,
- Timestamp-remapped^3,
- Timestamp-remapped\*Lag\_caseNew-1,
- Timestamp-remapped\*Lag\_caseNew-2,
- Timestamp-remapped\*Lag\_caseNew-3,
- Timestamp-remapped\*Lag\_caseNew-4,
- Timestamp-remapped\*Lag\_caseNew-5,
- Timestamp-remapped\*Lag\_caseNew-6,
- Timestamp-remapped\*Lag\_caseNew-7.

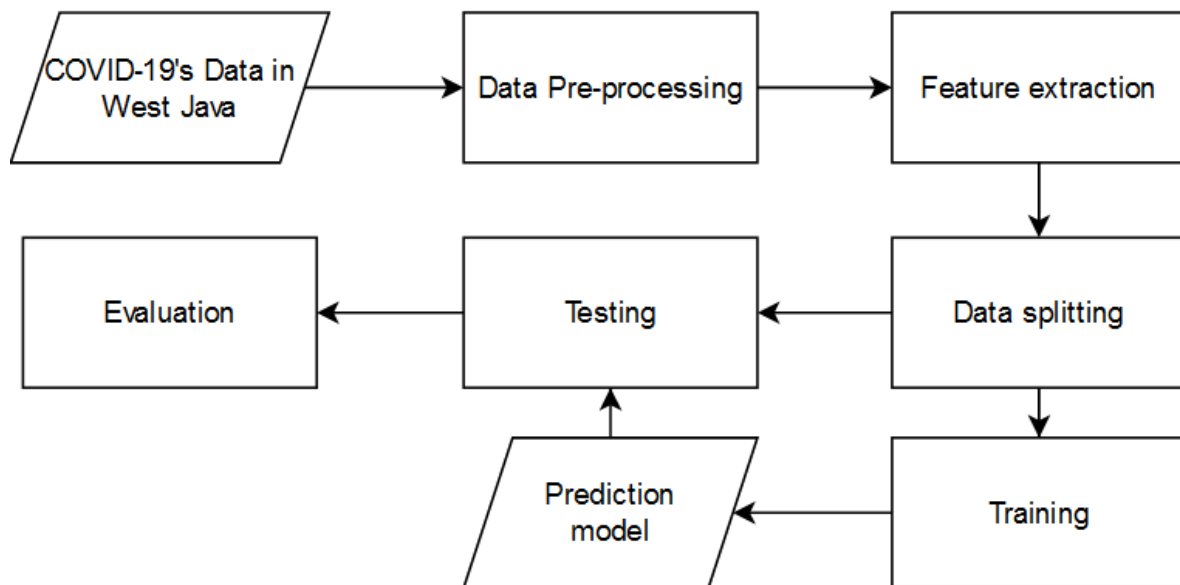


Figure 2. The methodology of Prediction of COVID-19 Data Time Series in West Java

The data set was divided into 80% training data, while the remaining 20% was test data. The training data was used to develop models, which were then applied to the test data to get the prediction. This study constructed models utilizing the following algorithms: LR, LWL, MLP, RBF, and SVM. In this part, each algorithm is explained. This study determined the number of statistical metrics to assess the performance of machine learning models. Mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), relative absolute error (RAE), and root relative squared error (RRSE) were the five parameters. They assisted in the ranking of

algorithms and subsequently determined the algorithm with the highest performance for this study (Moayed et al., 2019). Equations 1 through 5 represent MAE, RMSE, MAPE, RAE, and RRSE for n observations.

$$MAE = \frac{1}{n} \sum_{i=1}^n |f(x)_i - x_i| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f(x)_i - x_i)^2} \quad (2)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|f(x)_i - x_i|}{f(x)_i} \quad (3)$$

$$RAE = \frac{\sum_{i=1}^n |f(x)_i - x_i|}{\sum_{i=1}^n |f(x)_i - \bar{x}_i|} \quad (4)$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^n |f(x)_i - x_i|}{\sum_{i=1}^n |f(x)_i - \bar{x}_i|}} \quad (5)$$

Where  $x_i$  represents the  $i$ th data and  $f(x_i)$  represents the  $i$ th data prediction.

### Linear Regression (LR)

The study of the connection between the dependent and independent variables is known as regression analysis. The goal is to anticipate the dependent variable's value given the value of the independent variable. The linear regression approach employs a straight line to represent the relationship between these two variables [12], [15]. Creating a function that approximates the data points is required. Assume that the measurement result data is  $(x_i, y_i)$ . To approach these sites, a straight line is formed, and the error must be as little as possible. Equation 6 represents the regression equation model for linear regression.

$$y = bx + ba \quad (6)$$

$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (7)$$

$$a = \bar{y} - b\bar{x} \quad (8)$$

where  $y$ ,  $x$ ,  $a$ , and  $b$  represent the dependent, independent, constant, and regression coefficient.

### Locally Weighted Learning (LWL)

LWL predicts a-data based on other data that are geographically nearest to it. This is an example of lazy learning [15]. No models were constructed. The forecast solely considers the target data's immediate neighbors. The LWL's predicted steps include

1. Calculate the value of  $k$  (the number of closest neighbors).
2. Using the Euclidean distance to calculate the distance between the target data and all training data.
3. Identify the weight of the  $k$  training data with the shortest distance. Calculation of weight based on the specified kernel type.
4. Calculating forecasts using Equation 9.

$$\hat{y} = \frac{\sum w_i y_i}{\sum w_i} \quad (9)$$

### Multi-Layer Perceptron

Artificial neural networks (ANN) mimic the structure and function of the human brain. ANN is made up of several neurons. A neuron may create connections with other neurons, and each connection has a weight. The learning process is carried out by ANN by varying the weights of neural connections. The back-propagation technique is the most often used learning algorithm for training MLP models [17]. It makes use of an output error to change the value of its weights in reverse. Before an error may occur, the feed-forward stage must be finished. Neurons are activated during forward propagation using the sigmoid activation function. In general, back-propagation MLP training consists of the following steps:

1. Determine the total number of inputs, hidden, and output layers.
2. Give all weights between the input-hidden and output-hidden layers starting random values.
3. Each input unit receives an input signal in a feed-forward mode.
4. All buried layer units get the signal. Each unit in the succeeding layer adds weighted input signals. Furthermore, the activation function computes each output signal.
5. Performing back-propagation
6. A pattern target corresponding to the learning input pattern is delivered to each unit in the output layer. After then, the system computes error information. Weight and bias adjustments are computed, and the results are used to change the weight values in the bottom layer.
7. If the minimum error or maximum iteration limit is reached, the loop should be terminated.

### Radial Basis Function (RBF)

RBF maps a multidimensional nonlinear function based on the distance between the input and center vectors [23]. It excels at modeling nonlinear data. It has a network architecture, much like other neural networks. Its structure includes input, hidden, and output layer units. It is very useful for dealing with noisy input data. There are three major differences between MLP and RBF:

1. The link between the input layer and the hidden layer is unweighted.
2. The buried layer node has a radially symmetrical activation function.
3. Unlike MLP, RBF employs only one hidden layer. As a result, the RBF technique is expected to learn faster than the MLP network.

### Support Vector Machine

It aims to find a regression function of a hyperplane (borderline) in a time series prediction situation [24]. This regression function must be as small as possible while matching every piece of input data with an error. For example, there are  $n$  data  $(x_i, y_i)$  with  $i=1$  to  $n$ . It will find the function  $f(x)$  that deviates the most from the real target across all training data. If the variance is zero, the regression is perfect. Equation 10 represents the regression function.

$$f(x) = w^T \varphi(x) + b \quad (10)$$

Where  $\varphi(x)$  is a feature space point in higher dimensions. It is the result of mapping the vector  $x$  in the lower-dimensional input space. To estimate weight  $w$  and bias  $b$ , minimize the risk function shown in equation 11.

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{n} \sum_{i=1}^n (\xi_i, \xi_i^*) \quad (11)$$

which fulfills Equation 12 and 13.

$$y_i - w^T \varphi(x_i) - b - \xi_i \leq \Sigma \quad (12)$$

$$y_i - w^T \varphi(x_i) - b - \xi_i \leq \Sigma \quad (13)$$

$$\xi_i, \xi_i^* \geq 0 \quad (14)$$

Where  $\xi_i$  and  $\xi_i^*$  are slack variables used to solve a viable constraint issue. Constant  $C$  sets the trade-off between the function's thinness and the deviation's allowable upper limit. Any deviation larger than this value incurs a  $C$  penalty.

### RESULT AND DISCUSSION

This study utilized five machine learning methods. Each algorithm's batch size was 100.

The algorithm of LR performed feature selection. Using the Akaike information criterion, each property was evaluated. If an attribute had the lowest standardization coefficient, it was eliminated. This process ceased once all mistake corrections had been made. The crest on LR measured  $1.0E-8$ . The model of LWL considered all of its neighbors. Ranking of neighbors depending on their proximity to the target data. This search was conducted using brute force and the Euclidean distance. In addition, its weighting employed a linear kernel.

In this investigation, the MLP activation function was an approximation sigmoid with a squared error loss function. The architecture had one concealed layer. There were two fundamental functions in RBF. A thread is handled in a pool for both MLP and RBF. In addition, both methods utilized the Ridge penalty factor and tolerance values of 0.01 and  $1.0E-6$ , respectively. In this work, SVM utilized a poly kernel as its kernel. Its optimizer utilized epsilon, and the tolerance value was set at 0.001. The optimizer utilized a specifically designed implementation of Shevade et al work for regression [25].

These five algorithms were run on a system powered by an Intel Core i7-9700K CPU running at 3.60GHz. 8 GB of RAM was put on a PC with a 64-bit operating system, an x64-based CPU, and Windows 10. Figures 3 through 7 show the efficacy of each algorithm. Using previous data, each algorithm forecasted data for the next 30 days. Each graph has two lines. The red line depicts actual data, whereas the blue line depicts expected data. The situation has improved as the blue line approaches the red line. The bigger the difference between the blue and red lines, the greater the difference between predicted and actual circumstances.

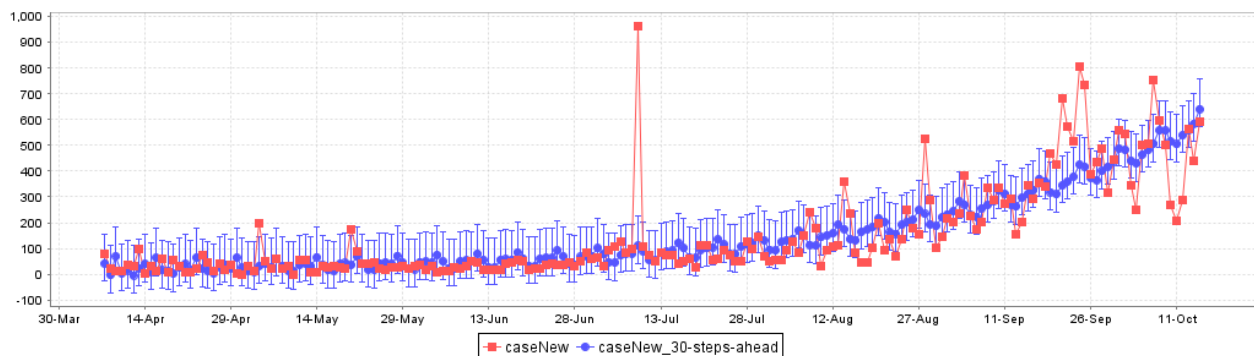


Figure 3. 30 Step-Ahead Forecasts For The Number of New Cases Based on LR

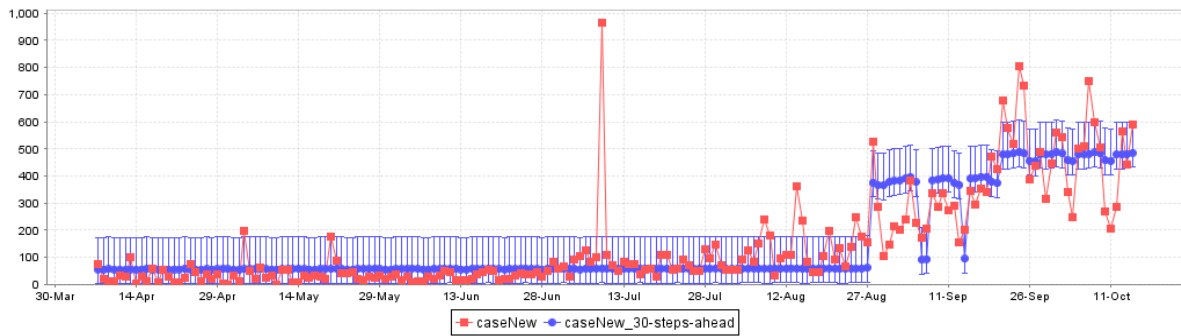


Figure 4. 30 Step-Ahead Forecasts For The Number of New Cases Based on LWL

Based on Figure 4, LWL had the greatest number of incorrect predictions compared to other algorithms. Until the end of August, the blue line formed a straight horizontal line, although the red line did not reveal this. From early September to mid-October, the

number of forecast mistakes increased. As seen in Figure 3, the LR's blue line indicated a state that was more similar to the red line. LR formed a somewhat curved line until the end of August, following which it continued to rise.

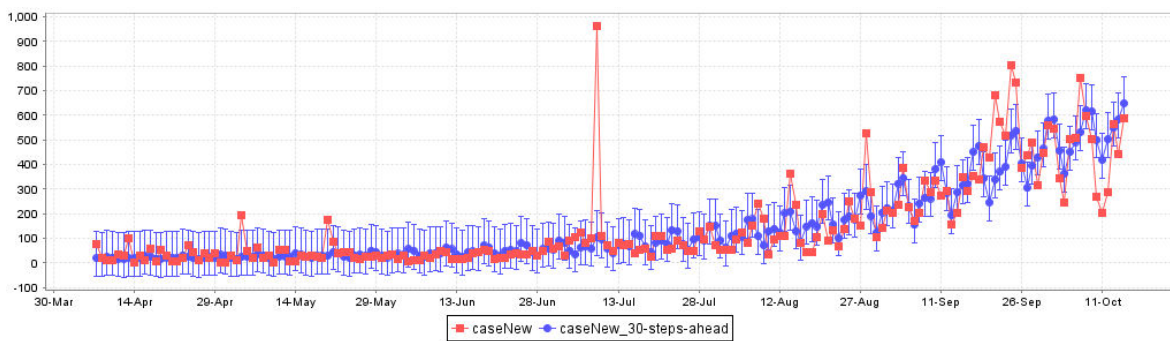


Figure 5. 30 Step-Ahead Forecasts For The Number of New Cases Based on MLP

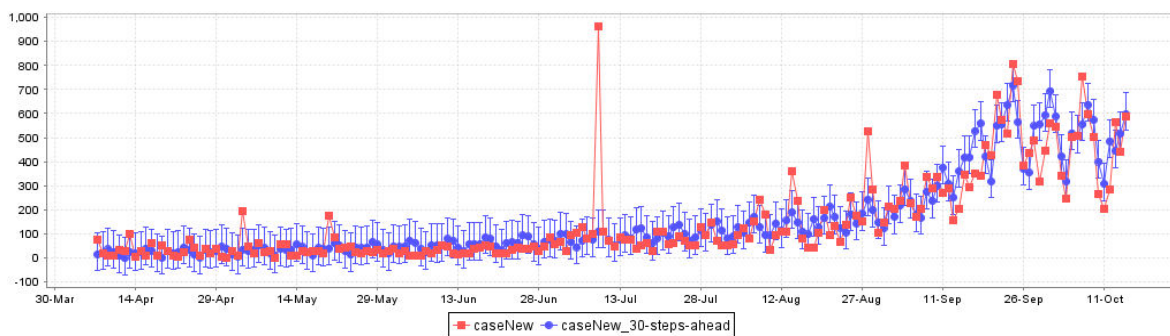


Figure 6. 30 Step-Ahead Forecasts For The Number of New Cases Based on RBF

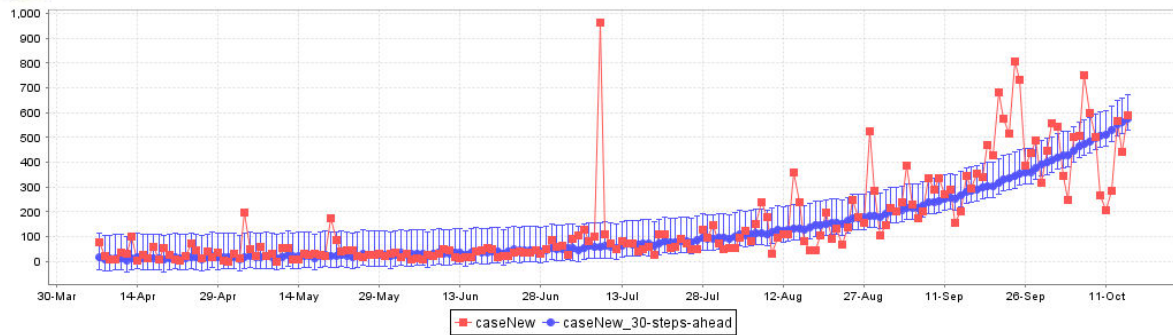


Figure 7. 30 Step-Ahead Forecasts For The Number of New Cases Based on SVM

Table 1. The Performance of Five Machine Learning in This Study

Methods	Models error					Models ranking					Total ranking	Final ranking
	MAE	RMSE	MAPE	RAE	RRSE	MAE	RMS E	MAPE	RAE	RRSE		
LR	54.03	99.73	99.03	69.68	67.63	4	3	4	4	3	18	4
LWL	61.77	104.8	163.4	79.93	71.19	5	4	5	5	4	23	5
MLP	49.6	93.9	78.43	63.99	63.66	2	2	2	2	2	10	2
RBF	48.85	89.73	88.67	62.99	60.88	1	1	3	1	1	7	1
SVM	51.36	106.3	55.61	66.07	72.05	3	5	1	3	5	17	3

Figure 7 depicts the SVM prediction error that was most prevalent between mid-August and October. It was very unlike MLP and RBF. Figure 5's MLP demonstrates that the majority of the real line was close to the projected line. The data from the middle of September to the beginning of October, however, reveals a distinct change in number. This is where RBF's expertise is compared to MLP's. Figure 6 demonstrates that RBF more accurately anticipated the data for this time.

Table 1 displays the performance of the five studied methods. In this work, MAE, RMSE, MAPE, RAE, and RRSE were used as performance assessment measures. The algorithm was ranked based on the error value per parameter; the less the error, the more efficient the algorithm. Except for MAPE, RBF has always been ranked #1. In contrast, LWL was always ranked last, except for RMSE.

Figure 3-7 demonstrates that none of the algorithms can forecast the July data surge. Such data surges are highly probable in the real world. Numerous reasons can induce this syndrome. Even though RBF is the greatest algorithm studied, it cannot anticipate this spike. It implies that the number of new cases in West

Java is influenced by a variety of variables, not just the time series trend shown in this study. Therefore, this study cannot be considered the major resource to anticipate future COVID-19 data in West Java.

In contrast, RBF could effectively model the data surges from August to October. The predicted line was approaching the real line. It indicates that RBF was able to simulate the anticipated time series for this period using historical data. It can be recommended as a tool to assist the government of West Java in assessing data on COVID-19 cases, particularly in managing social limitations. Figure 8 and Table 2 depict the RBF forecast for the following 30 days beginning on the 16th of October, 2020. According to RBF's model, West Java is approaching/has reached the height of the epidemic. Increasing social restraints is necessary. If WHO requires relaxing if the number of cases is less than one, yet the prediction data indicate that West Java has at least 275 new cases every day, what must the government do? Obviously, the easing of societal limitations requires serious consideration.

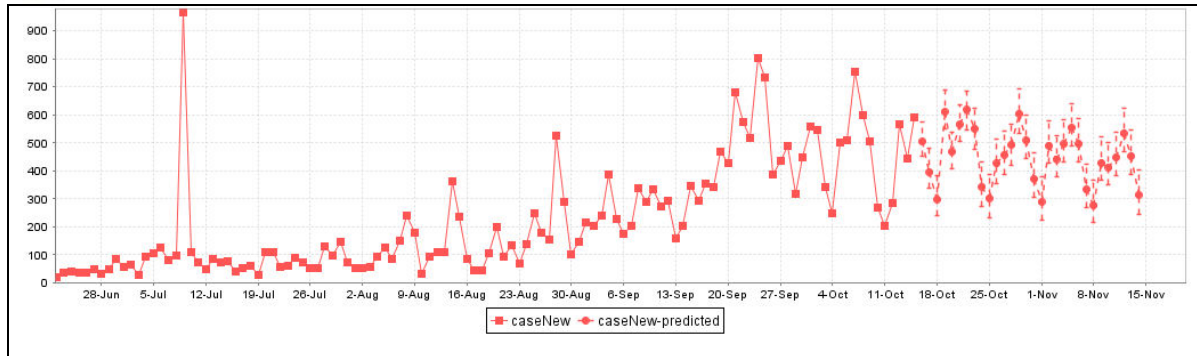


Figure 8. Forecast For The Number of New Cases From October 16th to November 14th, 2020

Table 2. Number of New Cases Expected in The Next 30 Days

Date	Estimated number of new cases
10-16-2020*	504
10-17-2020*	395
10-18-2020*	299
10-19-2020*	611
10-20-2020*	467
10-21-2020*	565
10-22-2020*	618
10-23-2020*	551
10-24-2020*	344
10-25-2020*	303
10-26-2020*	426
10-27-2020*	457
10-28-2020*	492
10-29-2020*	605
10-30-2020*	508
10-31-2020*	372
11-01-2020*	289
11-02-2020*	490
11-03-2020*	442
11-04-2020*	495
11-05-2020*	554
11-06-2020*	495
11-07-2020*	333
11-08-2020*	275
11-09-2020*	428
11-10-2020*	412
11-11-2020*	448
11-12-2020*	534
11-13-2020*	453
11-14-2020*	314

**CONCLUSION**

The COVID-19 pandemic first surfaced in the year 2019. There have been a number of initiatives taken to slow the progression of this infection. The province of West Java in Indonesia has implemented stringent social regulations in an effort to halt the development of this illness. Nevertheless, the people's

economy was completely ruined as a result of using this strategy. In the event that there were no cases found in the area, the World Health Organization (WHO) gave its approval for the social restrictions to be loosened. If the government decides to remove the social restriction, the decision must take into account the possibility of further confirmed incidents in



the future. It is feasible to make accurate projections of future data by employing machine learning. In this particular piece of work, the following sets of algorithms were utilized: linear regression (LR), locally weighted learning (LWL), multi-layer perceptron (MLP), radial basis function regression (RBF), and support vector machine (SVM) (SVM). During the course of the study, which began on March 2, 2020 and lasted until October 15, 2020, researchers tracked daily new cases of COVID-19 in West Java, Indonesia. This analysis found that the RBF algorithm performed the best. The relative absolute error (RAE), the mean absolute percentage error (MAPE), the root mean squared error (RMSE), and the mean absolute error (MAE) were respectively 48.85, 89.73, 88.67, and 60.88. It is possible to suggest to the government of West Java that they use the RBF prediction model in order to evaluate data on COVID-19 occurrences, particularly in relation to social restriction management. Beginning on October 16, 2020, it is projected that West Java would have a daily average of at least 275 new cases over the subsequent 30 days until the end of the outbreak. As a consequence of this, giving more leeway to societal restrictions calls for cautious analysis.

#### ACKNOWLEDGMENTS

The authors warmly appreciate the Rector and Directorate of Research and Community Service (DRPM). This study was funded by the Library and Online Data Research Grant, Universitas Padjadjaran, No. Contract: 1959/UN6.3.1/PT.00/2021.

#### REFERENCES

- [1] D. Kang, H. Choi, J. H. Kim, and J. Choi, "Spatial epidemic dynamics of the COVID-19 outbreak in China," *International Journal of Infectious Diseases*, vol. 94, 2020, doi: 10.1016/j.ijid.2020.03.076.
- [2] F. Reyes-Santias, I. Barrachina-Martinez, and D. Vivas-Consuelo, "Predictions of FluSurge 2.0 methodology on hospital utilization during the Covid-19 outbreaks in several countries," *International Journal of Engineering Business Management*, vol. 13, 2021, doi: 10.1177/18479790211020530.
- [3] İ. Yalçın, N. Can, Ö. Mançe Çalışır, S. Yalçın, and B. Çolak, "Latent profile analysis of COVID-19 fear, depression, anxiety, stress, mindfulness, and resilience," *Current Psychology*, vol. 41, no. 1, 2022, doi: 10.1007/s12144-021-01667-x.
- [4] J. R. Hargreaves and C. H. Logie, "Lifting lockdown policies: A critical moment for COVID-19 stigma," *Global Public Health*, vol. 15, no. 12, 2020, doi: 10.1080/17441692.2020.1825771.
- [5] J. Adhani, I. N. Yulita, A. Sholahuddin, M. N. Ardisasmita, and D. Agustian, "COVID-19 Social Safety Nets Sentiment Analysis on Twitter Using Gated Recurrent Unit (GRU) Method," in *2021 International Conference on Artificial Intelligence and Big Data Analytics, ICAIBDA 2021*, 2021, doi: 10.1109/ICAIBDA53487.2021.9689737.
- [6] N. Prawoto, E. P. Purnomo, and A. A. Zahra, "The impacts of Covid-19 pandemic on socio-economic mobility in Indonesia," *International Journal of Economics and Business Administration*, vol. 8, no. 3, 2020, doi: 10.35808/ijeba/486.
- [7] R. Singh and K. Sharma, "Impact of lockdown measures in abating the COVID-19 pandemic: Lessons from India's phased approach," *Eur J Public Health*, vol. 30, no. Supplement\_5, 2020, doi: 10.1093/eurpub/ckaa165.430.
- [8] I. N. Yulita, A. S. Abdullah, A. Helen, S. Hadi, A. Sholahuddin, and J. Rejito, "Comparison multi-layer perceptron and linear regression for time series prediction of novel coronavirus covid-19 data in West Java," in *Journal of Physics: Conference Series*, 2021, vol. 1722, no. 1, doi: 10.1088/1742-6596/1722/1/012021.
- [9] I. Rodríguez-Rodríguez, J. V. Rodríguez, D. J. Pardo-Quiles, P. Heras-González, and I. Chatzigiannakis, "Modeling and forecasting gender-based violence through machine learning techniques," *Applied Sciences (Switzerland)*, vol. 10, no. 22, 2020, doi: 10.3390/app10228244.
- [10] C. Matava, E. Pankiv, L. Ahumada, B. Weingarten, and A. Simpao, "Artificial intelligence, machine learning and the pediatric airway," *Paediatric Anaesthesia*, vol. 30, no. 3, 2020, doi: 10.1111/pan.13792.
- [11] M. Woschank, E. Rauch, and H. Zsifkovits, "A review of further directions for artificial intelligence, machine learning, and deep learning in smart logistics," *Sustainability (Switzerland)*, vol. 12, no. 9, 2020, doi: 10.3390/su12093760.
- [12] S. C. Yip, K. S. Wong, W. P. Hew, M. T. Gan, R. C. W. Phan, and S. W. Tan, "Detection of energy theft and defective smart meters in smart grids using linear

- regression," *International Journal of Electrical Power and Energy Systems*, vol. 91, 2017, doi: 10.1016/j.ijepes.2017.04.005.
- [13] T. Fang and R. Lahdelma, "Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system," *Appl Energy*, vol. 179, 2016, doi: 10.1016/j.apenergy.2016.06.133.
- [14] D. Bhuriya, G. Kaushal, A. Sharma, and U. Singh, "Stock market predication using a linear regression," in *Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017*, 2017, vol. 2017-January, doi: 10.1109/ICECA.2017.8212716.
- [15] N. Chen, J. Dai, X. Yuan, W. Gui, W. Ren, and H. N. Koivo, "Temperature Prediction Model for Roller Kiln by ALD-Based Double Locally Weighted Kernel Principal Component Regression," *IEEE Trans Instrum Meas*, vol. 67, no. 8, 2018, doi: 10.1109/TIM.2018.2810678.
- [16] A. Raza and M. Zhong, "Hybrid artificial neural network and locally weighted regression models for lane-based short-term urban traffic flow forecasting," *Transportation Planning and Technology*, vol. 41, no. 8, 2018, doi: 10.1080/03081060.2018.1526988.
- [17] I. N. Yulita, R. Rosadi, S. Purwani, and M. Suryani, "Multi-Layer Perceptron for Sleep Stage Classification," in *Journal of Physics: Conference Series*, 2018, vol. 1028, no. 1, doi: 10.1088/1742-6596/1028/1/012212.
- [18] P. C. Mukesh Kumar and R. Kavitha, "Prediction of nanofluid viscosity using multilayer perceptron and Gaussian process regression," *J Therm Anal Calorim*, vol. 144, no. 4, 2021, doi: 10.1007/s10973-020-09990-4.
- [19] M. A. F. B. Lima *et al.*, "Radial basis function for solar irradiance forecasting in equatorial areas," *Renewable Energy and Power Quality Journal*, vol. 17, 2019, doi: 10.24084/repqj17.288.
- [20] S. F. Ardabili *et al.*, "COVID-19 outbreak prediction with machine learning," *Algorithms*, vol. 13, no. 10, 2020, doi: 10.3390/a13100249.
- [21] P. Wang, X. Zheng, J. Li, and B. Zhu, "Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics," *Chaos Solitons Fractals*, vol. 139, 2020, doi: 10.1016/j.chaos.2020.110058.
- [22] A. Ahmad, S. Garhwal, S. K. Ray, G. Kumar, S. J. Malebary, and O. M. Barukab, "The Number of Confirmed Cases of Covid-19 by using Machine Learning: Methods and Challenges," *Archives of Computational Methods in Engineering*, vol. 28, no. 4, 2021, doi: 10.1007/s11831-020-09472-8.
- [23] C. Sbarufatti, M. Corbetta, M. Giglio, and F. Cadini, "Adaptive prognosis of lithium-ion batteries based on the combination of particle filters and radial basis function neural networks," *J Power Sources*, vol. 344, 2017, doi: 10.1016/j.jpowsour.2017.01.105.
- [24] B. M. Henrique, V. A. Sobreiro, and H. Kimura, "Stock price prediction using support vector regression on daily and up to the minute prices," *Journal of Finance and Data Science*, vol. 4, no. 3, 2018, doi: 10.1016/j.jfds.2018.04.003.
- [25] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to the SMO algorithm for SVM regression," *IEEE Trans Neural Netw*, vol. 11, no. 5, 2000, doi: 10.1109/72.870050.