

Predicting the weekly COVID-19 new cases using multilayer perceptron: An evidence from west Java, Indonesia

Yuyun Hidayat^a, Dhika Surya Pangestu^a, Subiyanto^b, Titi Purwandari^a, Sukono^c and Jumadil Saputra^{d*}

^aDepartment of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, 45363 Indonesia

^bDepartment of Marine Science, Faculty of Fishery and Marine Science, Universitas Padjadjaran, Sumedang, 45363 Indonesia

^cDepartment of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, 45363 Indonesia

^dFaculty of Business, Economics and Social Development, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia

CHRONICLE

Article history:

Received January 15, 2022

Received in revised format:

February 29, 2022

Accepted March 19 2022

Available online

March 19, 2022

Keywords:

Feedforwards Neural Networks

Multilayer Perceptron

Weekly COVID-19 new case

Forecasting and Indonesia

context

ABSTRACT

COVID-19 is a contagious disease caused by the coronavirus (SARS-CoV-2) that attacks the respiratory tract. On August 14th, 2021, 653,741 persons had been proven positive for COVID-19. The number of patients tends to increase as the number of COVID-19 cases grows. The more infected people, the more cases of COVID-19 there will be. The Bed Occupancy Ratio (BOR) in West Java reached an all-time high of 91.6 percent in June 2021, far exceeding the WHO recommendation of 60 percent, before gradually declining to 30.69 percent in August. Because of the new cases mentioned, the rate of spread of COVID-19 in West Java, the forecast of new cases is very strategic. The number of new cases in this study was predicted using a Multilayer Perceptron (MLP). The data used in this study were sourced from the COVID-19 Task Force. The data is the number of positive and new cases from 34 provinces in Indonesia from March 2nd, 2020, to August 14th, 2021. The results of the evaluation using test data on the number of active cases in the last 19 weeks, namely April 10th - August 14th, 2021, The MLP is accurate in forecasting the number of new cases 18 times for both forecast periods with APE < 15%, with the value MAPE, RMSE and MAE obtained were 5.52%, 1157.61, and 706.811. The results of this study can be helpful for the government as a reference in conditioning hospital bed capacity to deal with active COVID-19 cases in West Java in the next two weeks so that the hospital rejects no COVID-19 patients because the hospital is full.

© 2022 by the authors; licensee Growing Science, Canada.

1. Introduction

Indonesia is now one of the nations affected by the COVID-19 pandemic. COVID-19 was confirmed to have been first seen on March 2nd, 2020, in Indonesia. Due to contact with Japanese citizens, two people were exposed to COVID-19. It was discovered after a Japanese citizen was diagnosed with the coronavirus shortly after leaving Indonesia and landing in West Java (Subiyanto, Hidayat, Afrianto, & Supian, 2021). Since the virus's first appearance, the number of COVID-19 cases in Indonesia has steadily increased, with 3,833,541 people affected as of August 14th, 2021. According to the Worldometer, Indonesia is ranked 13th in the world and 4th in Asia for COVID-19 positive cases (Worldometer, 2021). West Java is one of Indonesia's provinces. West Java reported that as of August 14th, 2021, there were 653,741 confirmed cases of COVID-19 throughout the province, with a total of 65,000 active cases. Based on these results, West Java is rated second in Indonesia for positive COVID-19 instances (Subiyanto et al., 2021). Fig. 1 depicts the weekly number of new COVID-19 cases in West Java from March 2nd, 2020, to August 14th, 2021.

* Corresponding author.

E-mail address: jumadil.saputra@umt.edu.my (J. Saputra)

© 2022 by the authors; licensee Growing Science, Canada.

doi: 10.5267/dsl.2022.3.002

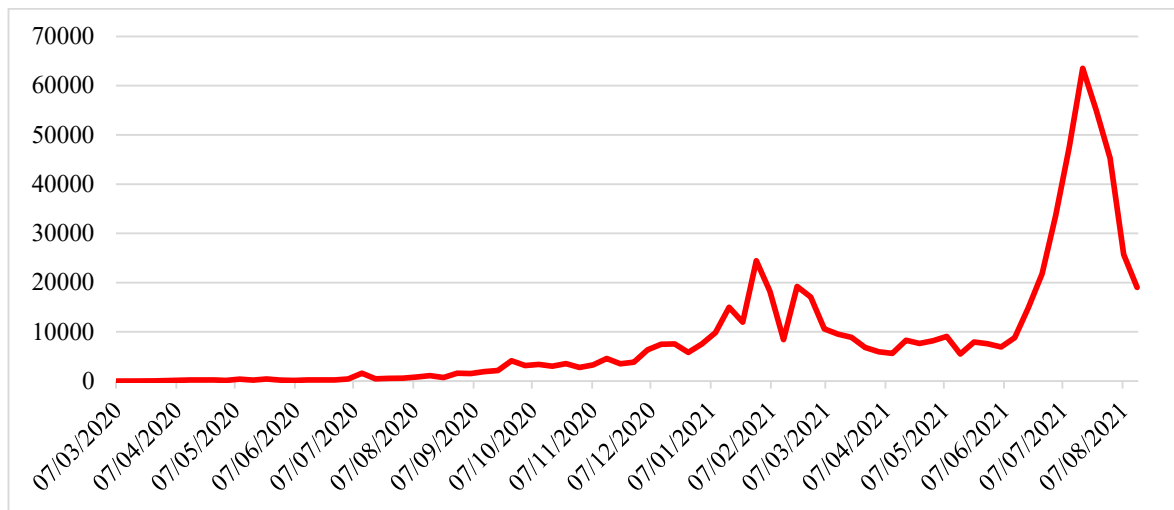


Fig. 1. The Number of Weekly COVID-19 New Case in West Java on March 2020 – August 2021

Fig. 1 demonstrates the number of weekly new cases of COVID-19 in West Java increased from its first appearance on March 2nd, 2020, to July 17th, 2021, before falling until August 14th, 2021. On the basis of available data, there has been a considerable increase in weekly new cases in June 2021, the number of weekly new cases of COVID-19 in West Java was 6,898 on June 5th, 2021, and this number grew to 63,511 on July 17th, 2021. The number of new cases indicates the COVID-19 in West Java spread speedily. Therefore, if calculated within a month, the acceleration of the spread of the COVID-19 pandemic is 56,613 cases. In other words, an acceleration of the spread of the COVID-19 pandemic by 820 percent during that period. It is one of the worst circumstances that has ever occurred in West Java during the COVID-19 pandemic. The new instances describe an increase in the number of people infected with COVID-19 and serve as a barometer for the effectiveness of the government's containment strategies in dealing with the COVID-19 epidemic. The number of new cases that continue to climb every week indicates good treatment. However, the number of new cases that continue to increase also indicates poor treatment. It is significant because the government must regulate the dissemination rate of COVID-19 by various strategies to accommodate it by existing health facilities.

The Bed Occupancy Rate (BOR) is a metric that evaluates a hospital's ability to offer adequate patient care. The BOR of the isolation chamber for COVID-19 patients in West Java reached its peak on June 28th, 2021, with a value of 91.6 percent (West Java Provincial Government, 2021), significantly exceeding the WHO standard of 60 percent. Over 85 percent occupancy negatively influences the safety and efficiency of hospital services (Forero & Hillman, 2008). The BOR of the COVID-19 patient isolation room is related to the hospital's ability to provide care for COVID-19 patients, which means that if the BOR of the COVID-19 patient isolation room reaches 100 percent, the hospital is full and cannot accept COVID-19 patients, and patient rejection will occur. COVID-19 patients have been rejected multiple times in Indonesia. From the end of December 2020 to January 21st, 2021, 34 complaints of COVID-19 patients were denied by hospitals in Indonesia, according to Laporan Covid-19 (Zulfikar, 2021). Refusal of COVID-19 patients happened in various regions around West Java, including Depok and Pangandaran. The refusal of patients by hospitals to be fatal for the safety of the lives of persons infected with COVID-19 and seeking medical care was particularly concerning. Ridwan Kamil, Governor of West Java, emphasized that all hospitals in West Java, including private hospitals, must be willing to serve COVID-19 patients and not let anyone be rejected.

The government's challenges in the COVID-19 pandemic include (i) how to deal with a prospective rise in COVID-19 patients and (ii) whether the containment measures implemented thus far have been effective or not. Hospital BOR numbers in West Java have exceeded WHO guidelines, and there has also been a refusal of COVID-19 patients in the province. It is a critical issue that the government must address. The authorities must be informed about the accelerated spread of COVID-19. The acceleration that occurs when the BOR rate is high is a disaster in the making. Therefore, monitoring the level of BOR and the rate of spread is essential in handling the COVID-19 pandemic. It makes predicting the number of new COVID-19 cases in West Java a strategic thing to do. This prediction can be an early alarm for the government. The rising prediction is an alarm for evaluating the effectiveness of ongoing containment measures and as input for adjusting existing health facilities. Appropriate policies that can contain the spread of COVID-19 and health facilities ready to serve various possible surges in patients are the keys to successfully handling the COVID-19 pandemic. The accurate prediction of the number of new cases is beneficial for the government as one of the inputs in evaluating and taking policies to deal with the COVID-19 pandemic in West Java. So that way, the containment measures taken can stop the spread of COVID-19 and nothing is rejected because the hospital is full.

The number of new cases of COVID-19 can be predicted with various methods. What is the best method to use to serve as an early warning system and input for the government? That necessitates real-time short-term forecasting. The use of

conventional statistical forecasting methods such as ARIMA and Exponential Smoothing is one method that can be used. There is a compartmental forecasting approach in addition to the conventional forecasting method. The compartmental model divides the population into multiple compartments, with people moving from one compartment to another (Brauer, 2008). Compartmental models that have been studied for their application in the COVID-19 pandemic include SIR model (Cooper, Mondal, & Antonopoulos, 2020), SIRD model (Faruk & Kar, 2021) and SPIR model (Hasan & Nasution, 2021). Forecasting methods based on artificial intelligence are currently being developed (Artificial intelligence). Machine learning (ML) methods or neural networks develop forecasting results. A significant amount of research has been done in this regard, using machine learning (ML) methods, especially neural networks, which can be used to develop forecasting results. Hundreds of articles have been published suggesting novel machine learning algorithms to develop methodologies and improving prediction accuracy (Makridakis, Spiliotis, & Assimakopoulos, 2018). The advantage of utilizing machine learning over conventional statistical approaches and compartmental models is that it reduces the usage of assumptions in existing data, hence avoiding errors caused by improper assumption selection.

2. Literature Review

2.1. Corona Virus Disease 2019 (COVID-19)

COVID-19 is a contagious disease caused by the coronavirus (SARS-CoV-2), a pathogen that attacks the respiratory tract. WHO first became aware of the new virus in Wuhan, the People's Republic of China, on December 31st, 2019 (Roosa et al., 2020). Coronaviruses are viruses that circulate between animals, with some infecting humans. Bats are considered the natural hosts of these viruses, and several other animal species are also known as sources. For example, Middle East Respiratory Syndrome Coronavirus (MERS-CoV) is transmitted to humans from camels, whereas Severe Acute Respiratory Syndrome Coronavirus-1 (SARS-CoV-1) is transmitted to humans from civets (Liang, 2020). A person can become infected with COVID-19 by getting into close contact with someone coughing and sneezing (Kumar, Malviya, & Sharma, 2020). COVID-19 can infect people of all ages.

The virus is conveyed through droplets produced by infected people while coughing and sneezing, although a person can also be infected through someone who does not have symptoms (Rothe et al., 2020). People who have tested positive for COVID-19 have reported various symptoms - from mild symptoms to severe illness. Symptoms can appear 2-14 days after exposure to the virus. The most common symptoms are having fever and cough, but there are other possible symptoms. On March 11th, 2020, WHO declared that COVID-19 was a pandemic. At that time, data from China showed that adults, especially those with congenital diseases, had a higher risk of developing severe cases of COVID-19 and a higher mortality rate than younger people (Chintalapudi, Battineni, & Amenta, 2020). Data from the EU/ European Economic Area (from countries for which data is available) shows that around 20-30% of diagnosed COVID-19 cases are hospitalized, and 2% of them suffer from severe disease. However, it's important to note that people with more severe symptoms are more likely to be tested than people with less severe symptoms. Therefore, the actual proportion of people requiring hospitalization out of the number of infected persons is lower than this figure indicates. Hospitalization rates are higher for those aged 60 years and over and those with underlying health conditions (Garg et al., 2020).

2.2. Time Series Clustering

Clustering is a technique for finding groups in a data set to get the data in one closely similar group and has clear differences from other groups (Kaufman & Rousseeuw, 1990). Time-series clustering is a unique sort of clustering. A temporal sequence comprises a series of nominal symbols from a specific alphabet, while a time series comprises continuous, real-valued elements (Shahnawaz, Ranjan, & Danish, 2011). A time series is characterized as dynamic data because its feature values change as a function of time, implying that each point's value(s) is/are one or more chronologically made observations. Time-series data is a sort of temporal data that is naturally huge in data size and has a high dimensionality (Liao, 2005; Lin, Vlachos, Keogh, & Gunopulos, 2004; Rani & Sikka, 2012). The most common use of time-series data clustering is detecting intriguing patterns in datasets (Das, Lin, Mannila, Renganathan, & Smyth, 1998; Shokoohi-Yekta et al., 2015). This task can be divided into two types: The first group identifies patterns that frequently emerge in the dataset. The second category includes identifying patterns that have unexpectedly appeared in datasets (Keogh, Lonardi, & Chiu, 2002).

To summarize, discovering time-series clusters can be beneficial in various disciplines to solve the following real-world problems: 1- Recognizing dynamic changes in time-series: detecting time-series correlation (He et al., 2012; Sfetsos & Siriopoulos, 2004). For example, it can locate companies with similar stock price movements in financial datasets. 2- Prediction and recommendation: a hybrid technique that combines clustering and per-cluster function approximation can assist users in predicting and recommending (Pavlidis, Plagianakos, Tasoulis, & Vrahatis, 2006; Sfetsos & Siriopoulos, 2004). For example, scientific databases can solve difficulties like predicting today's pattern by discovering patterns of solar magnetic wind. 3- Pattern discovery: locating and analyzing intriguing patterns in databases. For example, different daily patterns of sales of a specific product in a store can be discovered in the marketing database. So to get the provinces that have similar COVID-19 new cases dynamic changes to West Java, this research will use time-series clustering.

2.3. Artificial Neural Networks

Artificial Neural networks (ANN) is a set of computational units or nodes based on the function of neurons in animals. The ability to process ANN is found in the relationship between neurons, or weights, which is obtained by adapting to learning a set of patterns obtained from training data. ANN is commonly used for statistical analysis and data modelling (Cheng & Titterington, 1994). Besides that, ANN is also commonly used in classification or forecasting (Gurney, 2004). ANN has three layers: the input layer, the output layer, and the hidden layer. ANN are divided into two types: Feed Forward Neural Networks and Recurrent Neural Networks. Feed Forward Neural Networks (FFNN) are networks where the connections between neurons in the layer do not form a cycle, which means that the input only propagates forward from the input layer to the output layer. If there is no hidden network between the two layers, it is called a perceptron, whereas a hidden layer is called a multi-layer perceptron. When a feed-forward neural network is extended to include a feedback connection, the network is called a Recurrent Neural Network (RNN). Because the neuron layer has connections, RNN is considered a network with memory (Kandiran & Hacinliyan, 2019).

2.4. Multi-layer Perceptron

Feed Forwards Neural Networks (FFNN) are networks where the connections between neurons in the layer do not form a cycle, which means that the input only propagates forward from the input layer to the output layer. If there is no hidden network between the two layers, it is called a perceptron, whereas a hidden layer is called a multi-layer perceptron. MLP is a universal approximator. The ability of this universal approximation comes from the nonlinearity of the computational unit (neuron) (Du & Swamy, 2013). When the network starts running, each neuron in the hidden layer carries the computational result of the input and produces the result according to the layer of the existing nodes. MLP has proven its superiority in forecasting in the field of epidemiology. Many predictive epidemiological studies use MLP (Al-Qaness, Ewees, Fan, & Aziz, 2020; Manliura Datilo, Ismail, & Dare, 2019; Sujatha, Chatterjee, & Hassanien, 2020). MLP was used for daily case forecasting in the COVID-19 pandemic in West Java (Pontoh, Toharudin, Zahroh, & Supartini, 2020). MLP provides more accurate prediction results than other machine learning methods in that study. So far, research on COVID-19 in Indonesia, covering the Province of West Java, has frequently used daily data such as Tosepu et al. (2020) and Wibowo (2021). The use of daily data will undoubtedly increase training data for machine learning. Still, it is important to note that daily data in Indonesia is very volatile and prone to human error, as evidenced by the numbers that frequently differ between the central government and the provincial governments and lab performance, which varies by region. It is something on which the paper would like to expand. The weekly observation time was chosen to account for the 14-day incubation theory and reduce data volatility. This time horizon also aims to address, among other things, the instability of the Covid-19 examining laboratories' performance, which are spread throughout West Java.

3. Materials and Methods

The study uses a quantitative approach for measuring the development of the number of positive, recovered, dead, and active cases of COVID-19. The data is used to forecast the number of active cases and new cases of COVID-19 in West Java. The model is continuously validated every week to ensure that the resulting forecasting model is accurate and reliable. This study carried out the computational process using Python software, using package keras, tensorflow, and numpy. The forecasting steps, as seen in Fig. 2 below:

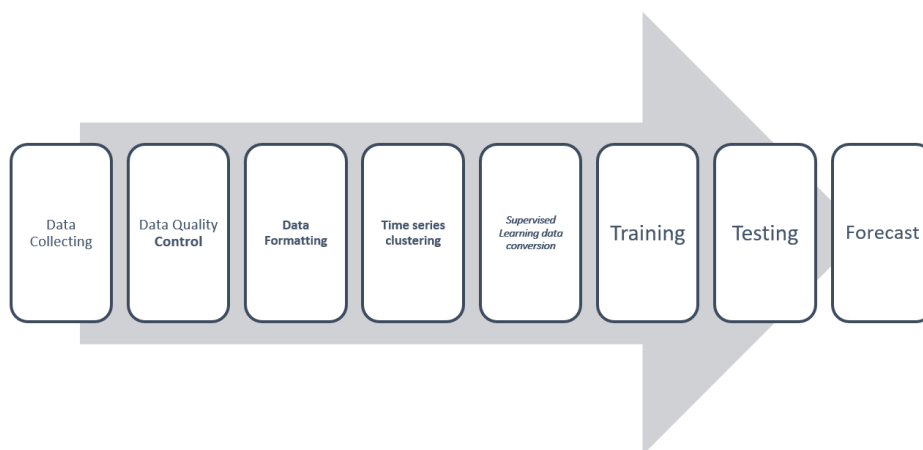


Fig. 2. Forecasting steps in determining the weekly new cases of COVID- in West Java using MLP

3.1. Data Sources

The data collection locations are at the West Java COVID-19 Information and Coordination Center, the COVID-19 Task Force, and the Ministry of Health. Data collection time starts from March 2nd, 2020, to August 14th, 2021, data is recorded

weekly, so there are approximately 76 data entries per province. The data is time-series with five variables: confirm, recover, death, new case, and active case from 34 Indonesian provinces. Confirm is a person who has been confirmed to have been exposed to COVID-19. Recover is a person who has been declared cured of COVID-19. Death is a person who died as a result of COVID-19. Active case is a person who is currently infected with COVID-19, and new cases are people who have been infected with COVID-19 for the first time. The data is transformed into Covid Weekly Data (CWD), which comprises four variables: confirm, recover, death, and new case. The following is the definition of a new case:

$$\text{New Case} = \text{Confirm}_{(t+1)} - \text{Confirm}_{(t)}, \quad (1)$$

3.2. Data Quality Control

The data used in this study is secondary data obtained from various sources such as Pikobar, the COVID-19 Task Force, the Ministry of Health, and others. The accuracy of web data is checked by using 30% acceptance sampling. Acceptance sampling involves collecting and analyzing several units of measurement to make an "accept or reject" decision about a relatively large number of units (Allen, 2010). 30% of the data was obtained using the Shewhart Control chart method. Then the data would be checked according to the out-of-control action plan (OCAP). OCAP is a flow chart or description of a series of activities that must be carried out when the data is outside the control limits of the control chart (Montgomery, 2019). Then the outlier data is checked by comparing it with the values at the source. If there is no difference, then the data is considered valid. If there is a difference, correction is made by replacing the data value with the data value on the web as the actual value.

3.3. Selection of Provinces with Similar Dynamic Changes to West Java

K-Medoids Clustering or Partitioning Around Medoids (PAM) is similar to K-Means. The algorithm used in K-Medoids is based on the search for k representative objects among data set objects. Clustering has a representative object, and it is often called the centroid. In the K-Medoids, the representative object is also called the medoid of the group (Kaufman & Rousseeuw, 2005). To overcome the problem of using K-Means, K-medoids can be used on the object with a substantial value that may deviate from the data distribution. This method can be chosen because it is more robust than most non-hierarchical clustering methods based on its sum of the squared estimate of errors (SSE) minimum value. The first step in K-Medoids is to compute the distance measure based on the cross-correlation between a pair of numeric time series. The cross-correlation based distance between two numeric time series is calculated as follows:

$$d_{i,j} = \sqrt{\frac{1 - \rho_{i,j,0}^2}{\sum_{k=1}^{\max} \rho_{i,j,k}}}, \quad (2)$$

where $\rho_{i,j,k}^2$ denote the cross-correlation between two time series x_i and y_j at lag k and \max is the maximum lag. After that, we use elbow methods in this research to decide the number of Clusters. This method is useful for determining the optimal number of clusters. The elbow method is the sum of squares at each number of clusters. It is calculated and graphed, and the user looks for a slope change to determine the optimal number of clusters. The Elbow method is a method that looks at the percentage of variance explained as a function of the number of clusters (Bholowalia & Kumar, 2014). The intuitive idea is to choose a point where diminishing returns are no longer worth the additional cost (Thorndike, 1953). It is a visual method. It starts with $k = 2$ and keeps increasing it in each step by 1, calculating the clusters and the cost of the training. At some value for K , the cost drops dramatically, and after that, it reaches a plateau when you increase it further. K is the value you want. The rationale is that you increase the number of clusters after this, but the new cluster is very near some existing ones (Kodinariya & Makwana, 2013).

The last step is the grouping of K-Medoids. The step of grouping using the K-Medoids method is as follows:

Calculate the distance of each object using Cross-Correlation Based distance with Eq. (2).

Calculate v_j for each object j with $d_i = \sum_{j=1}^n d_{i,j}$:

$$v_j = \sum_{i=1}^n \frac{d_{i,j}}{d_i}, \quad j = 1, \dots, n, \quad (3)$$

$d_{i,j}$: Distance matrix elements Cross Correlation
 v_j : Standardization of the number of rows for each column j

Arrange v_j from smallest to largest. Select k cluster with the first smallest v_j as the center (medoid).

Allocate objects that are non-medoid to the nearest medoid based on the distance of Cross Correlation Based Distance.

Calculate the total distance from non-medoid cluster to the center.

Define a new medoid for each cluster, an object that minimizes the total distance to other objects in the cluster.

Update the current medoid in each cluster by replacing it with a new medoid obtained from the existing cluster.

Allocate objects that are non-medoid to the nearest medoid based on the distance of Cross Correlation.

Calculate the total distance from non-medoid cluster to the center.

If the total of the new center differs from the total distance center of the first cluster, change the center (medoid).

Otherwise, the iteration stops and that result becomes the final clustering or grouping.

The number of groups (k) in K-Medoids is selected based on The Elbow Method.

3.4. Supervised Learning Data Conversion Using Sliding Window Method

The forecasting method used in this study is the Multilayer Perceptron Feedforwards Neural Networks. The data used for training is time-series data of the provinces in the same cluster as West Java. Backpropagation is used to train the neural networks in this study. It comprises forecasting based on the model's current state, comparing the prediction to the expected values, and using the difference to estimate the error information. This mistake information is then used to adjust neuron weights, and the process is repeated. The incorrect data is a statistical guess—the more training instances used in the estimate, the more accurate the estimate. The more likely it is that the network weights will be changed to enhance the model's performance, we use data from provinces in the same cluster as West Java. The improved error estimate comes at the cost of requiring the model to generate many more predictions before the estimate can be derived, and hence the neuron weights must be modified.

In the development process of neural networks, the procedure used is the determination of hyperparameters, the model training process, and the evaluation of the model. In this research, we use Spyder software in Python language for modelling.

Parameter Setting

Before training neural networks, the parameters are set first. It is done for optimum results, which can be defined randomly or by using an algorithm. Determination of the parameters is needed to get the optimum model, namely:

Input Neurons

The neurons in the input layer are called input neurons. Input neurons receive input patterns from the outside that describe a problem. The number of nodes or neurons in the input layer depends on the number of inputs in the model, and each input determines one neuron.

Learning rate

The learning rate is set during the training process to update the weights on the neurons until they reach the smallest local error values. The learning rate determines how fast the network learns. The learning rate value interval is between 0 to 1. If the learning rate value is close to 0 it will take a long-time during training to reach the smallest error, but if the value is close to 1 it will result in being stuck at a point that is not the smallest error. Learning rate is a parameter of the optimizer. The optimizer used in this research is Adaptive Moment Estimation (Adam). Adam is used because this optimizer can efficiently solve regression and deep learning problems (Kingma & Ba, 2015). There is no definite analytical method to determine the best learning rate. To get a reasonable learning rate, trial and error are usually used. Learning rates with a log scale is an initial recommendation in trial and error with the grid search technique.

Hidden Layer

To get the optimum network, the number of hidden layers can be determined by trial and error. The more layers added do not necessarily produce the best model and cause the model overfitting. In general, one hidden layer is sufficient to solve the problem. Using two hidden layers increases the risk for convergence at a point that is not a local minimum.

Hidden Neurons

Determining the number of neurons in the hidden layer is done by trial and error. Increasing the number of neurons can increase the capacity of the model to represent an event. However, it can also increase the time and memory used in modelling. Too many neurons can also cause overfitting, which is a condition where the model is only good for training data. Meanwhile, reducing the number of neurons can reduce the ability of a network to carry out the training and testing process (Panchal & Panchal, 2014).

Maximum Epoch

Epoch is a condition where all data has gone through the training process on the neural network until it returns to the beginning in one round. Each epoch can be partitioned into multiple batches. It is also an efficiency optimization. Table 1 shows the parameters tested in the study to predict the number of active cases and weekly new cases of COVID in West Java in the next two weeks.

Table 1

MLP parameter setting

Hyper-parameter	Weekly New Case
Learning Rate	Adam (0.002)
Input Neurons	6
Hidden Layer	2
Hidden Neurons	(100,20)
Maximum Epochs	3000

The dataset utilized in this study was quite limited, and we employed backpropagation, an iterative approach. Therefore we tried to increase the number of epochs to maximize learning. The number of times the weight in the neural network are adjusted increases as the number of epochs increases. There is no clear answer to how many epochs should be used in neural network training. The epoch differs for each dataset, but we may assume that the number of epochs is proportional to the data's divergence. The number of epochs used in this study was relatively large due to data usage from clusters.

Training Process

The backpropagation algorithm was used in this study as the training algorithm. The sequence of backpropagation training stages begins with the feed-forward stage at the input then performs an error calculation by comparing the output to the predetermined target. If the error is still above tolerance, the weight value on the neuron is updated, and the process is repeated backwards. The training procedure is known as the backpropagation algorithm, and its steps are as follows:

- Initialize the weights with the smallest possible random value. Each input unit ($x_i, i = 1, 2, \dots, n$) receives the input signal x_i and forwards it to all units in the next layer, namely the hidden layer. Every hidden unit ($z_j, j = 1, 2, \dots, n$) calculating the weighted input signal by,

$$z_{in_j} = w_{0j} + \sum_{i=1}^n x_i w_{ij} \quad (4)$$

use the activation function to calculate the output signal,

$$z_j = f(z_{in_j}) \quad (5)$$

- The signal is sent to all units in the next layer (output layer). It process is repeated as many times as the number of hidden layers. Each unit of output ($y_k, k = 1, 2, \dots, m$) sums the weighted input signal with formula:

$$y_{in_k} = w_{0k} + \sum_{j=1}^p z_j w_{jk} \quad (6)$$

use the activation function to calculate the output signal,

$$y_k = f(y_{in_k}) \quad (7)$$

- Each output unit ($y_k, k = 1, 2, \dots, m$) receives the target pattern according to the input training pattern, calculates the error information,

$$\delta_k = (t_k - y_k) f'(y_{in_k}) \quad (8)$$

- Calculate the weight correction (after this will be used to update w_{jk}) with the learning rate α ,

$$\Delta w_{jk} = \alpha \delta_k z_j \quad (9)$$

- Calculate the bias correction (will be used later to update w_{0k}),

$$\Delta w_{0k} = \alpha \delta_k \quad (10)$$

send δ_k to the next layer layer

- Each hidden unit ($Z_j, j = 1, 2, \dots, n$) add up the input delta (from the unit in the layer that is after it),

$$\delta_{in_j} = \sum_{k=1}^m \delta_k w_{jk} \quad (11)$$

Multiply this value by the derivative of the activation function to calculate the error information,

$$\delta_j = \delta_{in_j} f'(z_{in_j}) \quad (12)$$

- Calculate weight correction (will be used to update later),

$$\Delta v_{ij} = \alpha \delta_j x_i \quad (13)$$

and calculate the bias correction (after this will be used to update),

$$\Delta v_{0j} = \alpha \delta_j \quad (14)$$

- Each output unit ($y_k, k = 1, 2, \dots, m$) corrects the bias and weights ($j = 0, 1, \dots, p$):

$$w_{jk} (new) = w_{jk} (old) + \Delta w_{jk} \quad (15)$$

- Each hidden unit ($Z_j, j = 1, 2, \dots, p$) corrects for bias and weights ($i = 0, 1, \dots, n$):

$$v_{ij} (new) = v_{ij} (old) + \Delta v_{ij} \quad (16)$$

The training procedure will be repeated as long as the resulting error value decreases. The training procedure will be repeated as long as the resulting error value decreases. When the error value begins to rise, the network has memorized the training data pattern too well, and its ability to generalize suffers as a result. The training is terminated (Fausett, 2006). The best architecture has the lowest error value during the model testing process.

3.5. Neural Networks Evaluation

After obtaining various neural networks for prediction from the training process, the networks will be evaluated using testing data to measure the accuracy of these networks. The model obtained will be evaluated using the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). When comparing forecasting methods in one or several time-series with the same unit, RMSE is widely used. Meanwhile, MAE and MAPE are used as comparisons. It aims to find out the ideal lag and the most appropriate architecture to use in forecasting active cases of COVID-19 in West Java using MLP.

$$APE = \frac{|x_t - \hat{x}_t|}{x_t} \times 100\%, \quad (17)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (x_t - \hat{x}_t)^2}, \quad (18)$$

$$MAE = \frac{\sum_{t=1}^n |x_t - \hat{x}_t|}{n}, \quad (19)$$

$$MAPE = \frac{\sum_{t=1}^n \frac{|X_t - \hat{X}_t|}{X_t}}{n} \times 100\% \tag{20}$$

As stated, this research aims to create real-time short-term forecasts; hence, the Walkforwards validation approach will be used to validate each forecasting result. The testing system utilized in this investigation is depicted in Fig. 3.

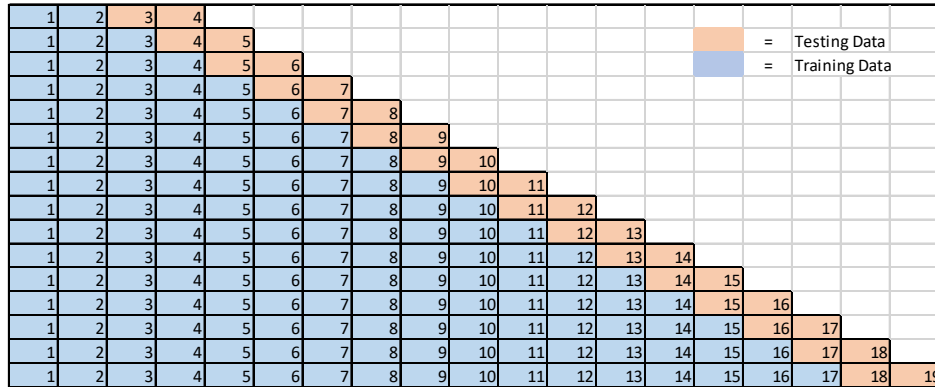


Fig. 3. Result of Walkforwards Validation

After evaluating the model, active cases and new cases of COVID-19 in West Java are predicted using the network with the best settings based on the evaluation results carried out.

4. Results and Discussion

This analysis examined new case data from 34 provinces from March 5th, 2020, to August 14th, 2021. The best clusters for COVID-19 new cases from 34 provinces in three clusters were determined using the Elbow Method.

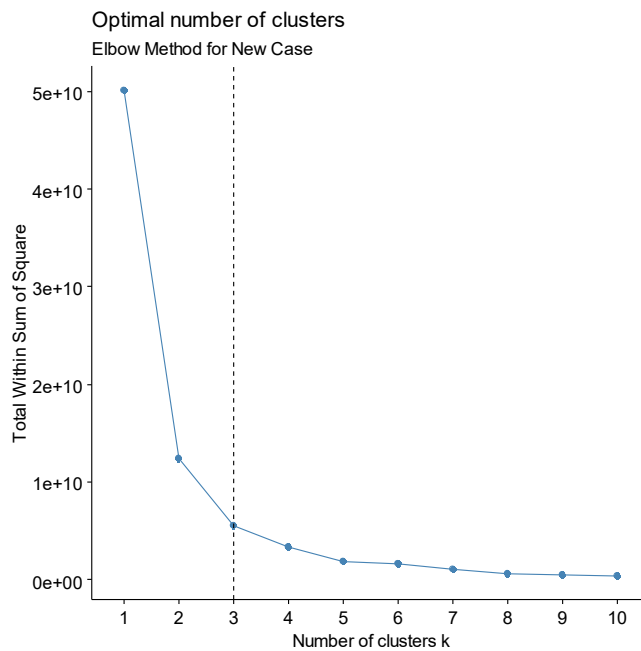


Fig. 4. Result of optimal number of clusters using Elbow methods.

Fig. 4 shows the overall number of squares in the cluster does not change considerably, as shown in the graph. As a result, the number of clusters formed in this study is 3. After obtaining the optimal number of clusters, weekly new cases were clustered in 34 provinces in Indonesia. Using $k = 3$, Table 2 shows clusters based on the number of new cases of COVID-19 Provinces in Indonesia.

Table 2
Clusters of weekly new cases of COVID-19 Provinces in Indonesia

Cluster	Province
1	Aceh, Bangka Belitung, Jambi, Riau, Sumatera Barat
2	Bali, Gorontalo, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Lampung, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Utara, Sumatera Utara
3	Banten, Bengkulu, Daerah Istimewa Yogyakarta, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Tengah, Kepulauan Riau, Maluku, Maluku Utara, Papua Barat, Sulawesi Tenggara, Sumatera Selatan

Table 2 shows temporary clusters for weekly new cases of COVID-19 provinces in Indonesia. To ensure the consistency of the cluster, iterations were carried out 3 times. The result is that the cluster does not change, so table 2 shows the final cluster for the weekly number of new cases of COVID-19 in the Provinces in Indonesia. The characteristics of each cluster are observed based on a graphic plot of the development of the number of new COVID-19 cases in each province. Fig. 5 displays weekly number of new cases of COVID-19 in Cluster 1.

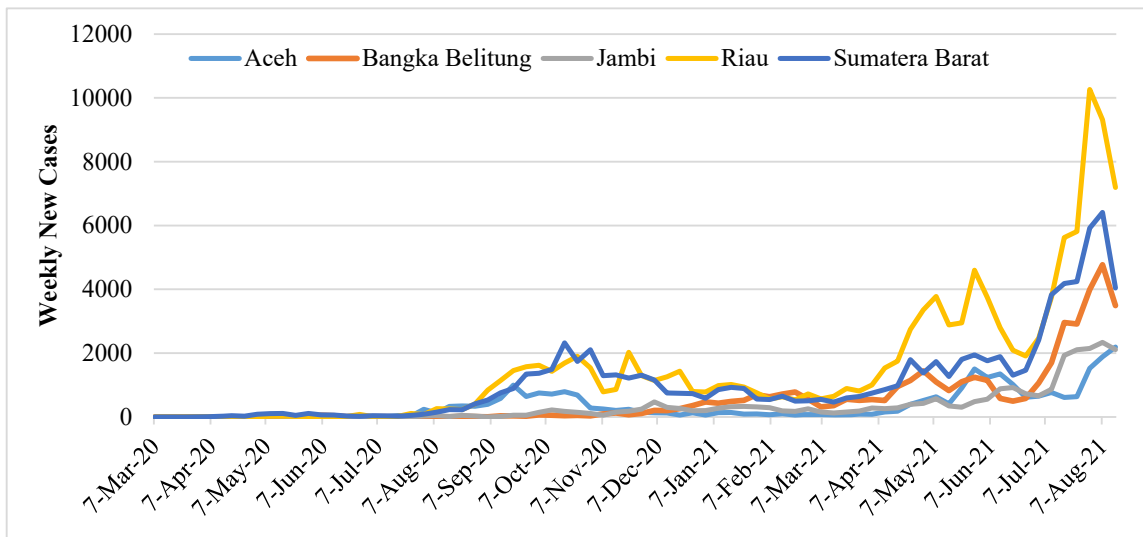


Fig. 5. Result of Weekly new cases of COVID-19 in Cluster 1

This cluster has the fewest members, which is only 5 Provinces. The characteristics of members in this cluster are found in cluster members who have 3 peaks during the COVID-19 pandemic in Indonesia. In general, members in this cluster experience a peak in October 2020, May 2021, and finally July 2021. For cluster 2, the development of the number of new events from cluster members is shown in Fig. 6.

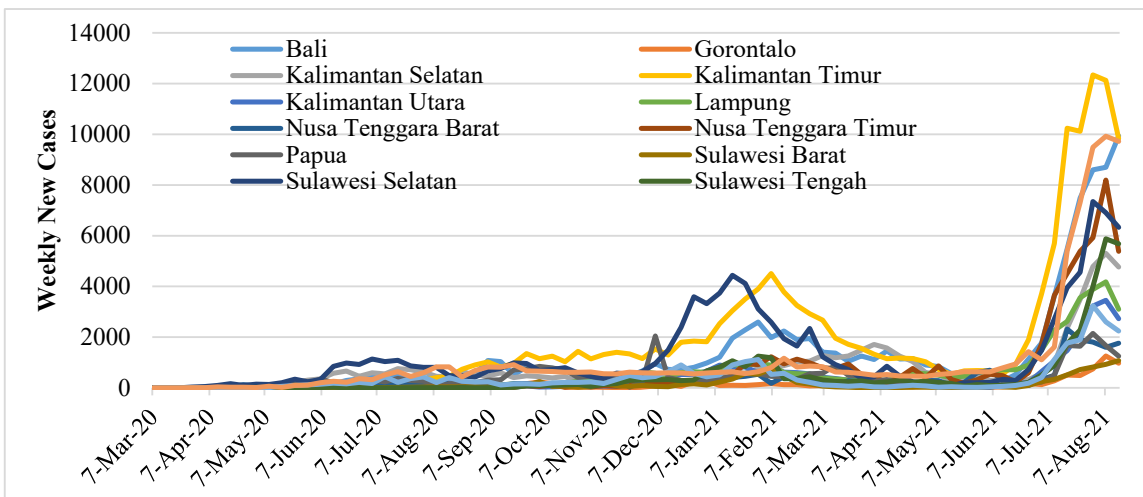
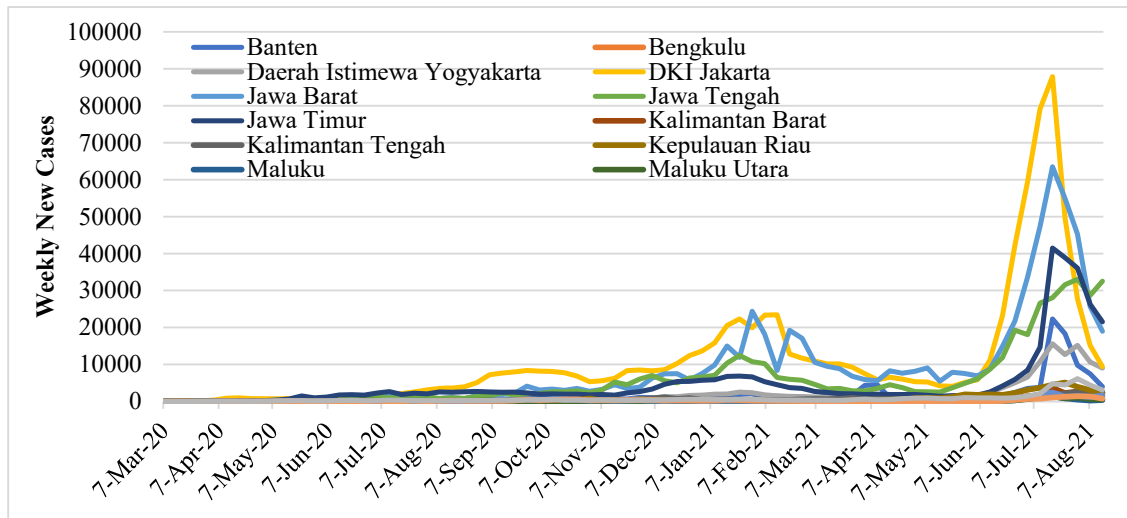


Fig. 6. Result of Weekly new cases of COVID-19 in Cluster 2

Fig. 6 shows the number of new cases in Cluster 2. This cluster consists of 14 Provinces. In general, members of this cluster experience two peaks, the first occurring between January and February 2021, and the second peak occurring in July 2021. Another observed characteristic of this cluster is the fluctuation of the number of new cases, which is quite sharp from time to time. It can be seen from the graph, which tends to be rough. Next, Fig. 7 shows the number of new cases in cluster 3.

**Fig. 7.** Result of weekly new cases of COVID-19 in Cluster 3

Cluster 3 has the most members, as many as 16 provinces. West Java is a member of this cluster. The characteristics in cluster 3 are found in cluster members who tend to have 2 peaks. The first peak occurs between January and February 2021. Compared to cluster 2, the fluctuations of members of this cluster are not too sharp. Cluster 3 is a cluster used for the training and testing process in MLP. Before proceeding to the training and testing process, we tried to observe the characteristics of the training and testing data used in this study. Table 3 shows the statistical characteristics of the training and testing data used in this study.

Table 3
Statistical characteristics of training and testing data

	Data	
	Training	Testing
Mean	822.998	3569.701
Standard Error	50.348	340.692
Median	201.5	942.5
Standard Deviation	2169.087	8659.224
Sample Variance	4704939.952	74982164.52
Range	24440	87879
Minimum	0	0
Maximum	24440	87879

After the training and testing process is carried out on the MLP using data obtained from the results of time series clustering. Fig. 8 shows the data plot between the actual data and the forecast results of new COVID-19 cases in West Java using MLP. Fig. 8 shows the data plot between the MLP forecast results and the actual new cases every week. Based on the graph, it can be observed that the forecast results for the next one week and the next two weeks have excellent accuracy. It can be seen from the graphic form that is close to the graphic form of the actual new cases of COVID-19. Between the two forecasting periods, namely for the next one week and the next two weeks, the most striking difference occurs on May 15th 2021. In the period May 15th 2021, actual data shows a decrease in the number of weekly COVID-19 cases from the previous week, which was 9071 cases, to 5482 cases, while the forecast results do not show this decline. Overall, there are not many significant differences in forecasts for the next one or two weeks. After plotting the data between the actual and predicted values, forecasting the number of new cases is evaluated empirically using absolute percentage error (APE).

In this section, the accuracy of forecasting the number of active cases and weekly new cases of COVID-19 is evaluated using APE in each period. Table 4 shows the absolute percentage error for forecasting the number of active COVID-19 cases in West Java from April 10th – August 14th, 2021.

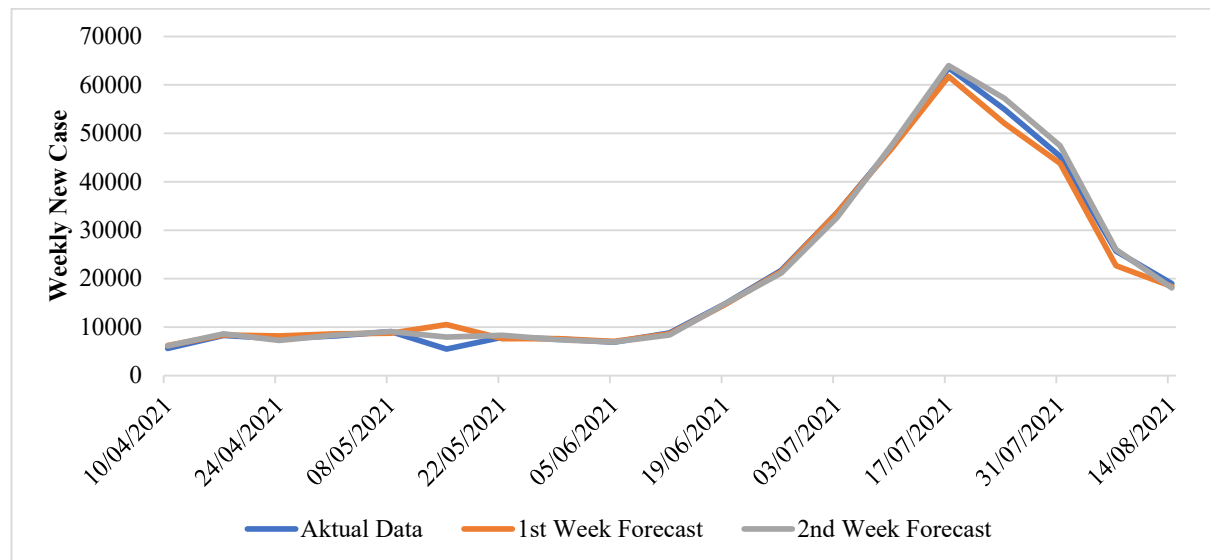


Fig. 8. Result of MLP forecast data and actual COVID-19 cases in West Java.

Table 4

Absolute percentage error COVID-19 weekly new case forecast

Week	From Date	To Date	1st Week APE	2nd Week APE
58	04/04/2021	10/04/2021	4,28%	8,57%
59	11/04/2021	17/04/2021	4,88%	7,01%
60	18/04/2021	24/04/2021	1,37%	0,79%
61	25/04/2021	01/05/2021	1,22%	4,37%
62	02/05/2021	08/05/2021	1,27%	0,14%
63	09/05/2021	15/05/2021	61,92%	41,55%
64	16/05/2021	22/05/2021	0,06%	3,45%
65	23/05/2021	29/05/2021	1,43%	2,20%
66	30/05/2021	05/06/2021	6,23%	3,45%
67	06/06/2021	12/06/2021	1,84%	0,14%
68	6/13/2021	6/19/2021	1,99%	1,41%
69	6/20/2021	6/26/2021	0,31%	1,80%
70	6/27/2021	7/3/2021	1,37%	1,70%
71	7/4/2021	7/10/2021	2,08%	1,92%
72	11/07/2021	17/07/2021	2,34%	0,30%
73	7/18/2021	7/24/2021	6,58%	0,98%
74	7/25/2021	7/31/2021	2,83%	0,26%
75	8/1/2021	8/7/2021	10,04%	3,76%
76	08/08/2021	14/08/2021	1,95%	12,01%

The APE value indicates forecasting accuracy concerning the actual number of new COVID-19 cases in West Java per timeframe; the lower the APE value, the more accurate the forecast results can be, and vice versa. According to table 4, the highest APE value in forecasting the number of new cases in West Java for the next week and the next two weeks was found on May 15th, 2021, with an APE value of 61.92 percent and 41.55%, respectively. Using the APE = 15% as accuracy limit, MLP is estimated to accurately estimate the number of new COVID-19 cases 18 times for forecasting 1 and 2 weeks from the 19 weeks testing period. The performance of MLP in forecasting is shown in Table 4. The table shows the MAPE, RMSE, and MAE values of each setting used in MLP to predict the number of new cases of COVID-19 in West Java.

Table 5

Result of MAPE, RMSE and MAE using Multilayer Perceptron

MAPE	RMSE	MAE
5,52%	1157.6101	706,811

The MAPE, RMSE, and MAE values indicate the network's overall performance for forecasting; the smaller these values, the more accurate the network, and the larger the value, the less accurate the network. Table 5 shows the testing results carried out in weeks 58 - 76 (April 10th – August 14th 2021). The MAPE, RMSE, and MAE values were obtained for forecasting the number of new COVID-19 cases in West Java were 5.52%, 1157,61 and 706.811. A comparison with

previous works on the same subject (Wibowo, 2021) was also made. It is conducted by comparing prediction results for the same time period as the study, i.e. September-October 2020. According to our findings, the RMSE value generated by the model in this study is lower than the previous work's results. For positive COVID-19 cases in West Java, we found an RMSE of 23.913, much lower than the RMSE of 129.014 found in the study we mentioned. As a result, our West Java forecasting is more accurate than previous research.

5. Conclusions

The new cases discuss the increase in the number of people infected with COVID-19 and will assess the effectiveness of the government's control efforts in dealing with the COVID-19 pandemic. A weekly drop in the number of new cases suggests effective treatment, whereas an increase shows ineffective treatment. The management must be aware that their actions will speed COVID-19's spread. When the BOR rate is high, the acceleration is a cause for disaster. As a result, forecasting new COVID-19 cases in real time is a strategic move, as it can serve as an early warning and feedback to the government. The number of COVID-19 cases investigated in Indonesia and affecting West Java has been determined using daily data. Indonesia's daily data is highly changeable and susceptible to human error, as evidenced by statistics that frequently contradict federal and local governments and lab performance that varies by region. It is a problem we are looking to address in this study. This time horizon also attempts to resolve, among other things, the inconsistency of the performance of the Covid-19 examination laboratories, which are dispersed over West Java.

Overall forecasting of new cases for the next two weeks using MLP produces a really good accuracy, based on the findings obtained, with MAPE, RMSE, and MAE values of 5.52%, 1157.61, and 706.812. MLP has been accurate in estimating the number of active cases for one and two weeks to come as much as 18 times with an absolute percentage error (APE) < 15%. Although this MLP provides forecast results with very good accuracy, it should also be noted that using MLP in forecasting has a limitation in that it does not produce a model that can be interpreted to describe how the COVID-19 phenomenon occurs. This method only generates predictive numbers based on the number of new cases that the government can use as a reference. Furthermore, the loss function used in this study is the symmetric loss function, which means that the consequences of under and over forecasting of the network used are considered the same.

In contrast, in the COVID-19 pandemic, this may need to be reconsidered. Whether the prediction of the number of lower new cases of COVID-19 have the same consequence is higher? This prediction could be used for various policies, such as quarantine or adjusting hospital bed capacity. The next stage in this research uses machine learning approaches with an asymmetric loss function to provide distinct treatments for under forecast and forecast. Hopefully, the government will utilize the outcomes of this study as a guide to forecast accurate weekly new cases of COVID-19 over the next two weeks. As a result, the government warned of the COVID-19 pandemic and can adjust hospital bed allocations based on forecasts. Consequently, the government may conduct a comprehensive assessment of the effectiveness of policies that are being implemented, and no patient will be refused entry due to hospital capacity.

Funding

The Lecturers Competence Research Grant of Universitas Padjadjaran (RKDU), with a contract number: 1959/UN6.3.1/PT.00/2021.

Acknowledgments

The authors would like to thank the Dean of the Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, and the Directorate of Research and Community Service (DRPM-Unpad). Also, we would like to thank Universiti Malaysia Terengganu, which has supported this research publication and excellent collaboration work.

Conflicts of Interest

The authors declare no conflict of interest

References

- Al-Qaness, M. A. A., Ewees, A. A., Fan, H., & Aziz, M. A. El. (2020). Optimization method for forecasting confirmed cases of COVID-19 in China. *Journal of Clinical Medicine*, 9(3), 1–15.
- Allen, T. (2010). *Introduction to Engineering Statistics and Lean Sigma*. London, UK: Springer-Verlag.
- Bholowalia, P., & Kumar, A. (2014). EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. *International Journal of Computer Applications*, 105(9), 17–24.
- Brauer, F. (2008). Compartmental models in epidemiology. In *Mathematical epidemiology* (pp. 19–79). Springer.
- Cheng, B., & Titterton, D. M. (1994). Neural Networks: A Review from Statistical Perspective. *Statistical Science*, 9(1), 2–30.
- Chintalapudi, N., Battineni, G., & Amenta, F. (2020). COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach. *Journal of Microbiology, Immunology and*

Infection.

- Cooper, I., Mondal, A., & Antonopoulos, C. G. (2020). A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals*, 139, 110057.
- Das, G., Lin, K.-I., Mannila, H., Renganathan, G., & Smyth, P. (1998). Rule Discovery from Time Series. In *KDD* (Vol. 98, pp. 16–22).
- Du, K. L., & Swamy, M. N. S. (2013). *Neural Networks and Statistical Learning*. London, UK: Springer.
- Faruk, O., & Kar, S. (2021). A Data Driven Analysis and Forecast of COVID-19 Dynamics during the Third Wave Using SIRD Model in Bangladesh. *COVID*, 1(2), 503–517.
- Fausett, L. V. (2006). *Fundamentals of neural networks: architectures, algorithms and applications*. Pearson Education India.
- Forero, R., & Hillman, K. (2008). Access block and overcrowding: a literature review. *Prepared for the Australasian College for Emergency Medicine. Sydney: University of New South Wales*.
- Garg, S., Kim, L., Whitaker, M., O'Halloran, A., Cummings, C., Holstein, R., ... Alden, N. B. (2020). Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019—COVID-NET, 14 States, March 1–30, 2020. *Morbidity and Mortality Weekly Report*, 69(15), 458.
- Gurney, K. (2004). *Introduction to neural networks*. London, UK: Taylor & Francis e-Library.
- Hasan, A., & Nasution, Y. (2021). A compartmental epidemic model incorporating probable cases to model COVID-19 outbreak in regions with limited testing capacity. *ISA Transactions*.
- He, W., Feng, G., Wu, Q., He, T., Wan, S., & Chou, J. (2012). A new method for abrupt dynamic change detection of correlated time series. *International Journal of Climatology*, 32(10), 1604–1614.
- Kandiran, E., & Hacinliyan, A. (2019). Comparison of Feedforward and Recurrent Neural Network in Forecasting Chaotic Dynamical System. *AJIT-e: Online Academic Journal of Information Technology*, 10(37), 31–44.
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. New York, NY, USA: John Wiley & Sons, Inc.
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. New York, NY, USA: John Wiley & Sons, Inc.
- Keogh, E., Lonardi, S., & Chiu, B. (2002). Finding surprising patterns in a time series database in linear time and space. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 550–556).
- Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (pp. 1–15). San Diego, CA, USA.
- Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining of cluster in K-means. *International Journal of Advance Research in Computer Science and Management Studies*, 1(6), 90–95.
- Kumar, D., Malviya, R., & Sharma, P. K. (2020). Corona virus: a review of COVID-19. *EJMO*, 4(1), 8–25.
- Liang, K. (2020). Mathematical model of infection kinetics and its analysis for COVID-19, SARS and MERS. *Infection, Genetics and Evolution*, 82(April), 104306.
- Liao, T. W. (2005). Clustering of time series data—a survey. *Pattern Recognition*, 38(11), 1857–1874.
- Lin, J., Vlachos, M., Keogh, E., & Gunopulos, D. (2004). Iterative incremental clustering of time series. In *International Conference on Extending Database Technology* (pp. 106–122). Springer.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS One*, 13(3), e0194889.
- Manliura Datilo, P., Ismail, Z., & Dare, J. (2019). A Review of Epidemic Forecasting Using Artificial Neural Networks. *International Journal of Epidemiologic Research*, 6(3), 132–143.
- Montgomery, D. C. (2019). *Introduction to Statistical Quality Control* (8th Editio). New York, NY, USA: John Wiley & Sons, Inc.
- Panchal, F. S., & Panchal, M. (2014). International Journal of Computer Science and Mobile Computing Review on Methods of Selecting Number of Hidden Nodes in Artificial Neural Network. *International Journal of Computer Science and Mobile Computing*, 3(11), 455–464.
- Pavlidis, N. G., Plagianakos, V. P., Tasoulis, D. K., & Vrahatis, M. N. (2006). Financial forecasting through unsupervised clustering and neural networks. *Operational Research*, 6(2), 103–127.
- Pontoh, R. S., Toharudin, T., Zahroh, S., & Supartini, E. (2020). Effectiveness of the public health measures to prevent the spread of covid-19. *Commun. Math. Biol. Neurosci.*, 2020, Article-ID.
- Rani, S., & Sikka, G. (2012). Recent techniques of clustering of time series data: a survey. *International Journal of Computer Applications*, 52(15).
- Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J. M., ... Chowell, G. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling*, 5, 256–263.
- Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., ... Guggemos, W. (2020). Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *New England Journal of Medicine*, 382(10), 970–971.
- Sfetsos, A., & Siriopoulos, C. (2004). Time series forecasting with a hybrid clustering scheme and pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 34(3), 399–405.
- Shahnawaz, M., Ranjan, A., & Danish, M. (2011). Temporal Data Mining: An Overview. *International Journal of Engineering and Advanced Technology*, 1(1), 20–24.

- Shokoohi-Yekta, M., Chen, Y., Campana, B., Hu, B., Zakaria, J., & Keogh, E. (2015). Discovery of meaningful rules in time series. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1085–1094).
- Subiyanto, Hidayat, Y., Afrianto, E., & Supian, S. (2021). Numerical Estimation of the Final Size of the Spread of COVID-19 in West Java Province, Indonesia. *Engineering Letters*, 29(3), 1015–1019.
- Sujatha, R., Chatterjee, J. M., & Hassanien, A. E. (2020). Correction to: A machine learning forecasting model for COVID-19 pandemic in India. *Stochastic Environmental Research and Risk Assessment*, 34, 959–972.
- Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, 18(4), 267–276.
- Tosepu, R., Effendy, D. S., Lestari, H., Bahar, H., Asfian, P., & Sakka, A. (2020). Vulnerability of weather on COVID-19 pandemic in West Java, Indonesia. *Public Health of Indonesia*, 6(4), 123–128.
- West Java Provincial Government. (2021). Incidence Rate in West Java.
- Wibowo, F. W. (2021). Prediction modelling of COVID-19 on provinces in Indonesia using long short-term memory machine learning. In *Journal of Physics: Conference Series* (Vol. 1844, p. 12006). IOP Publishing.
- Worldometer. (2021). Covid-19 Coronavirus Pandemic.
- Zulfikar, M. (2021). 19 received 34 reports of COVID-19 patients being rejected by the hospital.



© 2022 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).