



Early Information on Active Cases in Zero Rejection Efforts for COVID-19 Patients in West Java Province 2021 Using the Feedforwards Multilayer Perceptron Neural Network

Yuyun Hidayat^{1,*}, Subiyanto², Titi Purwandari³, Dhika Surya Pangestu⁴, Sukono⁵

^{1,3,4} *Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jatinangor, West Java, Indonesia*

² *Department of Marine Science, Faculty of Fishery and Marine Science, Universitas Padjadjaran, West Java, Indonesia*

³ *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jatinangor, West Java, Indonesia*

**Corresponding author email: yuyun.hidayat@unpad.ac.id*

Abstract

West Java noted, as of August 14, 2021, 653,741 people were confirmed positive for COVID-19. On the same date, the number of active COVID-19 cases in West Java was 65,000. There is a significant increase in active cases of COVID-19 in 2021 in West Java. In the period 5 June – 17 July 2021, there was an increase in the number of active cases by 95,532. In that period, active cases increased by 484%, and the Bed Occupancy Ratio (BOR) in West Java had jumped in June 2021 with the highest number of 91.6%, this figure far exceeded the WHO recommendation of 60% before finally continuing to decline and finally in August was at 30.69%. This has an impact on the incidence of patient rejection at the COVID-19 referral hospital. Active cases talk about COVID-19 patients who need medical treatment and new cases talk about the rate of spread of COVID-19 in West Java, so these two things are very strategic to study. In this study, active cases and new case were predicted using Multilayer Perceptron (MLP). The data used in this study were sourced from the COVID-19 Task Force. The data is the number of positive cases, recovered and died of COVID-19 sufferers in 34 provinces in Indonesia in the period 2 March 2020 - 14 August 2021. The results of the study found, from the results of the evaluation using data testing the number of active cases in the last 19 weeks, namely April 10 – August 14, 2021, MLP is accurate in predicting the number of active cases for the first coming week 17 times, and the next two weeks for the second week 12 times with an absolute percentage error (APE) < 20%. As for weekly new cases, MLP has been accurate 10 times for the next one week and 9 times for the next two weeks. It is hoped that the results of this study can be useful for the government as a reference in conditioning the hospital bed capacity to deal with active cases of COVID-19 in West Java in the next two weeks so that no COVID-19 patients are rejected by the hospital because the hospital is full.

Keywords: COVID-19, active case, bed capacity, new case

1. Introduction

Indonesia is one of the countries currently hit by the COVID-19 pandemic. COVID-19 was confirmed to have first appeared in Indonesia on March 2, 2020. At that time there were two people who were exposed to COVID-19 due to contact with Japanese citizens. This was known after a Japanese citizen was declared infected with the corona virus after leaving Indonesia and arriving in West Java [1]. Since the beginning of its emergence, the number of COVID-19 cases in Indonesia has continued to increase, until August 14, 2021, there have been 3,833,541 people infected by COVID-19 [2]. Based on the Worldometer, Indonesia is ranked 13th in the world and 4th in Asia for positive cases of COVID-19 [3].

West Java is one of the provinces in Indonesia. West Java noted that as of August 14, 2021, there have been 653,741 confirmed cases of West COVID-19 with a total of 65,000 active cases. So based on these figures, West Java is ranked 2nd for positive cases and 1st for active COVID-19 cases in Indonesia [4]. In the COVID-19 pandemic, active cases speak of COVID-19 patients who require medical care and are directly related to hospital capacity. Graph 1. Shows the development of the number of active cases and weekly new cases of COVID-19 in West Java in the period 2 March 2020 – 14 August 2021.

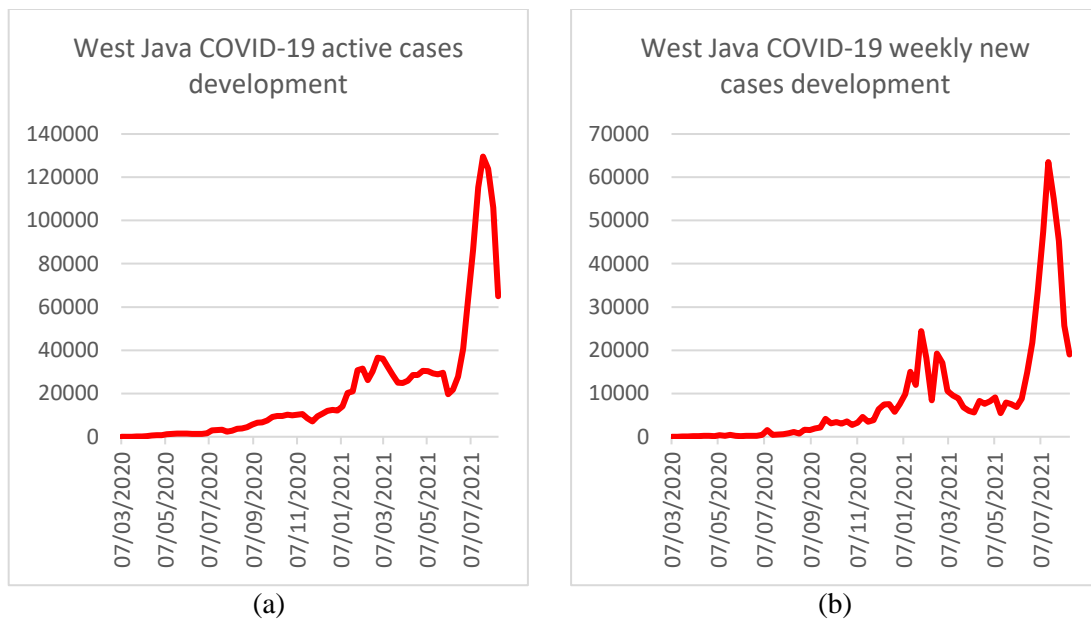


Figure 1: Graph of the development of the number of active cases (a) and new cases (b) in West Java 3 March 2020 – 14 August 2021

Figure 1. shows that since its initial appearance on March 2, 2020, until July 17, 2021, the development of the number of active cases and weekly new cases of COVID-19 in West Java has continued to increase, before decreasing further until August 14, 2021. Based on available data, there was a significant increase in active cases and weekly new cases in June 2021, on 5 June 2021 the number of active cases of COVID-19 in West Java was 19,725 cases and on 17 July 2021, this number had increased to 115,257 cases. In the span of 1 month, there was an increase in the number of active cases as many as 95,532 cases, or during that period there was an increase in the number of active cases by 484%. This situation is one of the worst situations that has ever occurred during the COVID-19 pandemic in West Java. The new cases speak of the growing number of people infected by COVID-19. The new cases are a benchmark for the efficiency of containment measures implemented by the government in dealing with the COVID-19 pandemic. While active cases talk about COVID-19 patients who need medical care and are directly related to hospital capacity. Bed Occupancy Rate (BOR) is a measure that reflects the hospital's ability to provide proper care to patients. The BOR of the isolation room for COVID-19 patients in West Java was at its highest on June 28, 2021, with a value of 91.6% [5] this number far exceeds the recommendation from WHO, which is 60%. BOR is related to the hospital's ability to provide care for COVID-19 patients, meaning that if the BOR of the COVID-19 patient isolation room reaches 100%, the hospital is full and can no longer accept COVID-19 patients, and patient rejection will occur.

The rejection of COVID-19 patients has occurred several times in Indonesia. Based on Laporan Covid-19 from the end of December 2020 to January 21, 2021, 34 reports of COVID-19 patients were rejected by hospitals in Indonesia [6]. For West Java, refusal of COVID-19 patients occurred in several areas such as in Depok [7] and Pangandaran [8]. The incident was very worrying, the refusal of patients by hospitals to be fatal for the safety of the lives of people infected with COVID-19 and requiring medical treatment. 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 do not let anyone be rejected [9].

In the COVID-19 pandemic, the challenges faced by the government include (i) how to deal with a possible surge in COVID-19 patients (ii) whether the containment measures that have been carried out so far have been effective or not. Hospital BOR figures in West Java have been above the limits set by WHO and there has also been a refusal of COVID-19 patients in West Java. This is something the government really needs to be aware of. If there are active cases that exceed the capacity of the hospital, it is feared that there will be a spike in the death rate due to patients not receiving proper medical treatment. In overcoming the refusal of COVID-19 patients, appropriate policies are needed. One of them is that the government can condition its health facilities to deal with a surge in active cases in the future. Therefore, predicting the number of active cases and new cases of COVID-19 is a strategic thing to do. This prediction can be used as a reference for the conditioning of existing health facilities and also a reference for the government to ensure that its resources are sufficient to deal with the number of COVID-19 patients in the future and also evaluate the results of the policies that have been taken so far. Accurate predictions really help the government determine steps and take policies to deal with the COVID-19 pandemic in West Java. So that way, COVID-19 positive patients can be given good care, and no one is rejected by the hospital because the hospital is full.

2. Literature Review

2.1. Corona Virus Disease 2019 (COVID-19)

COVID-19 is a contagious disease caused by the coronavirus, SARS-CoV-2, which is 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 31, 2019 [10]. Coronaviruses are viruses that circulate between animals, with some infecting humans. Bats are considered to be 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 [11]. People who have tested positive for COVID-19 have reported a variety of 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, cough, but there are other possible symptoms as well [12]. On March 11, 2020, WHO declared that COVID-19 was a pandemic [13]. 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 also a higher mortality rate than younger people [14]. 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 total number of infected persons is lower than this figure indicates. Hospitalization rates are higher for those aged 60 years and over, and for those with underlying health conditions [15].

2.2. Time Series Clustering

Clustering is a technique for finding groups in a data set in order to get the data in one group are closely similar, and have clear differences with other groups [16]. A special type of clustering is time-series clustering. A sequence composed of a series of nominal symbols from a particular alphabet is usually called a temporal sequence, and a sequence of continuous, real-valued elements, is known as a time-series [17]. A time-series is essentially classified as dynamic data because its feature values change as a function of time, which means that the value(s) of each point of a time-series is/are one or more observations that are made chronologically. Time-series data is a type of temporal data which is naturally high dimensional and large in data size [18][19][20]. Clustering of time-series data is mostly utilized for discovery of interesting patterns in time-series datasets [21] [22]. This task itself, fall into two categories: The first group is the one which is used to find patterns that frequently appears in the dataset [23]. The second group are methods to discover patterns which happened in datasets surprisingly [24]. Briefly, finding the clusters of time-series can be advantageous in different domains to answer following real world problems : 1- Recognizing dynamic changes in time-series: detection of correlation between time-series [25] [26]. For example, in financial databases, it can be used to find the companies with similar stock price move. 2- Prediction and recommendation: a hybrid technique combining clustering and function approximation per cluster can help user to predict and recommend [26][27]. For example, in scientific databases, it can address problems such as finding the patterns of solar magnetic wind to predict today's pattern. 3- Pattern discovery: to discover the interesting patterns in databases. For example, in marketing database, different daily patterns of sales of a specific product in a store can be discovered. So in order to get the provinces which has similar COVID-19 dynamic changes to West Java, this research will use time-series clustering.

2.3. Artificial Neural Networks

Artificial Neural networks (ANN) are a set of computational units or nodes, which are based on the function of neurons in animals. The ability to process ANN is found in the relationship between neurons, or what is called 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 modeling [28]. Besides that, ANN is also commonly used in classification or forecasting [29]. ANN has three types of layers, namely the input layer, the output layer, and the hidden layer. ANN are divided into two types, namely 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 if there is a hidden layer it 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 its own connections, RNN is considered a network with memory [30].

2.4. Multilayer 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 if there is a hidden layer it 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) [31]. 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 of the predictive epidemiological studies use MLP [32][33] [34].

3. Research Methodology

The research methodology carried out is quantitative research by 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. The following is a flowchart of this research:

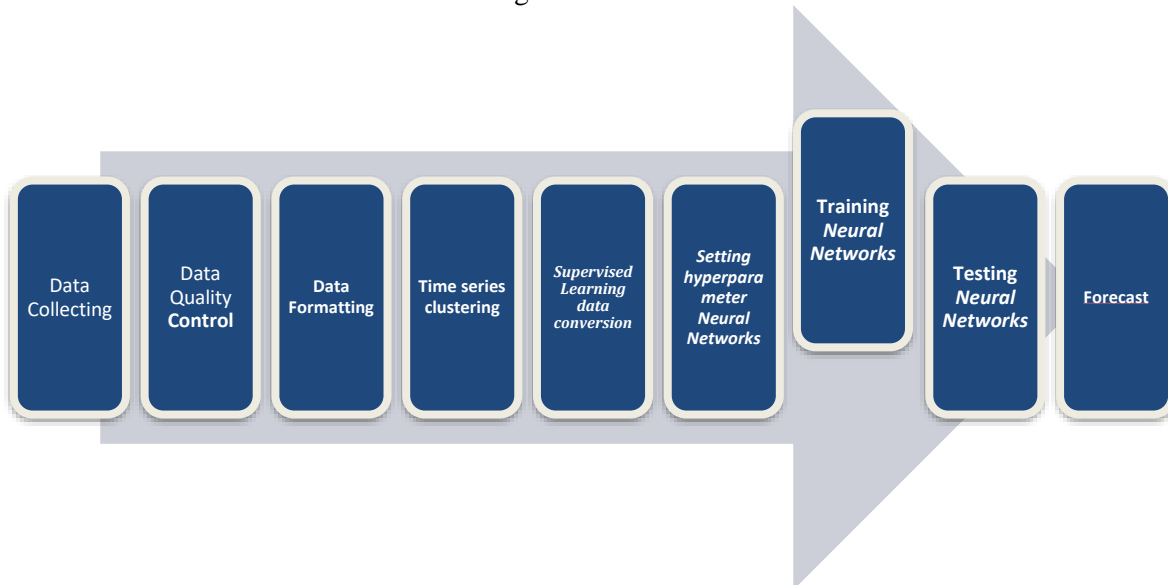


Figure 2. Flowchart of forecasting the number of active case and weekly new cases of COVID-19 in West Java using MLP

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 2, 2020, to August 14, 2021. Data from the source is converted into Covid Weekly Data (CWD), which consists of 4 variables, namely confirm, recover, death, new case, and active case from COVID-19 sufferers, with active cases as follows:

$$Active\ Case = Confirm - Recover - Death \quad (1)$$

$$New\ Case = Confirm_{(t+1)} - Confirm_{(t)} \quad (2)$$

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 [35]. 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 [36]. 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 have an representative

object, it is often called the centroid. In the K-Medoids, the representative object is also called the medoid of the group [37]. To overcome the problem of using K-Means, K-medoids can be used to the object with very large value which 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 squared estimate of errors (SSE) minimum value. The first step in K-Medoids is to computes 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 :

$$d_{i,j} = \sqrt{\frac{1 - \rho_{i,j,0}^2}{\sum_{k=1}^{max} \rho_{i,j,k}^2}} \quad (3)$$

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 to decide the number of Cluster, in this research we use elbow methods. This method is useful for determining the optimal number of clusters. The elbow method, in which the sum of squares at each number of clusters is calculated and graphed, and the user looks for a change of slope to determine the optimal number of clusters. Elbow method is a method which looks at the percentage of variance explained as a function of the number of clusters [38]. The intuitive idea is to choose a point where diminishing returns are no longer worth the additional cost [39]. It is a visual method. It starts with $k=2$, and keep increasing it in each step by 1, calculating the clusters and the cost that comes with the training. At some value for K , the cost drops dramatically, and after that it reaches a plateau when you increase it further This is the K value you want. The rationale is that after this, you increase the number of clusters but the new cluster is very near some of the existing [40].

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

1. Calculate the distance of each object using Cross Correlation Based distance with equation (3).
2. Calculate v_j for each object j with $d_i = \sum_{j=1}^n d_{ij}$:

$$v_j = \sum_{i=1}^n \frac{d_{ij}}{d_i}, j = 1, \dots, n \quad (4)$$

d_{ij} : Distance matrix elements Cross Correlation

v_j : Standardization of the number of rows for each column j

3. Arrange v_j from smallest to largest. Select k cluster which have the first smallest v_j as the center (medoid).
4. Allocate objects that are non-medoid to the nearest medoid based on the distance of Cross Correlation Based Distance.
5. Calculate the total distance from non-medoid cluster to the center.
6. Define a new medoid for each cluster which is 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 which is obtained from the existing cluster.
7. Allocate objects that are non-medoid to the nearest medoid based on the distance of Cross Correlation.
8. Calculate the total distance from non-medoid cluster to the center.
9. If the total of new center is differs from the total distance center of the first cluster, change the center (medoid). Otherwise, the iteration is stop and that result become 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

Time-series data from CSSE JHU are converted into supervised learning data, using the sliding window method. This method converts the existing time-series data into several hw windows classifier to predict the individual output of y . To be specific, windows $(x_{i,t-d}, x_{i,t-d+1}, \dots, x_{i,t}, \dots, x_{i,t+d-1}, x_{i,t+d})$ will be used to predict each $y_{i,t}$. This method converts sequential supervised learning problems into classical supervised learning problems so that making algorithms for classical supervised learning such as backpropagation can be done [41].

3.5. Training Multilayer Perceptron Feedforwards Neural Networks

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 that in the same cluster as west Java. 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 [42] for modelling.

Parameter Setting

Before training neural networks, the parameters are set first. This is done for optimum results which can be defined randomly or 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 value. 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 problems and deep learning problems [43]. There is no definite analytical method that can determine what the best learning rate is. To get a good learning rate, trial and error are usually used. The use of learning rates with a log scale is an initial recommendation in trial and error with the grid search technique [44].

Hidden Layer

Determining the number of hidden layers can be done by trial and error to get the optimum network. The more layers added do not necessarily produce the best model. Because it can also cause overfitting on the model. 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. By increasing the number of neurons, it can increase the capacity of the model to represent an event. However, it can also increase the time and memory used in modeling. 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 [45].

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. below 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

Hyperparameter	Active Case	Weekly New Case
Learning Rate	Adam (0.1)	Adam (0.05)

Input Neurons	3	3
Hidden Layer	1	1
Hidden Neurons	10	10
Maximum Epochs	500	1000

3.6. Neural Networks Evaluation

After obtaining various neural networks for prediction from the training process, then 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 [46]. Meanwhile, MAE and MAPE are used as comparisons. We do this to find out what is the ideal lag and also the most appropriate architecture to use in forecasting active cases of COVID-19 in West Java using MLP.

$$APE = \frac{|X_t - \hat{X}_t|}{X_t} \times 100\% \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (X_t - \hat{X}_t)^2} \quad (6)$$

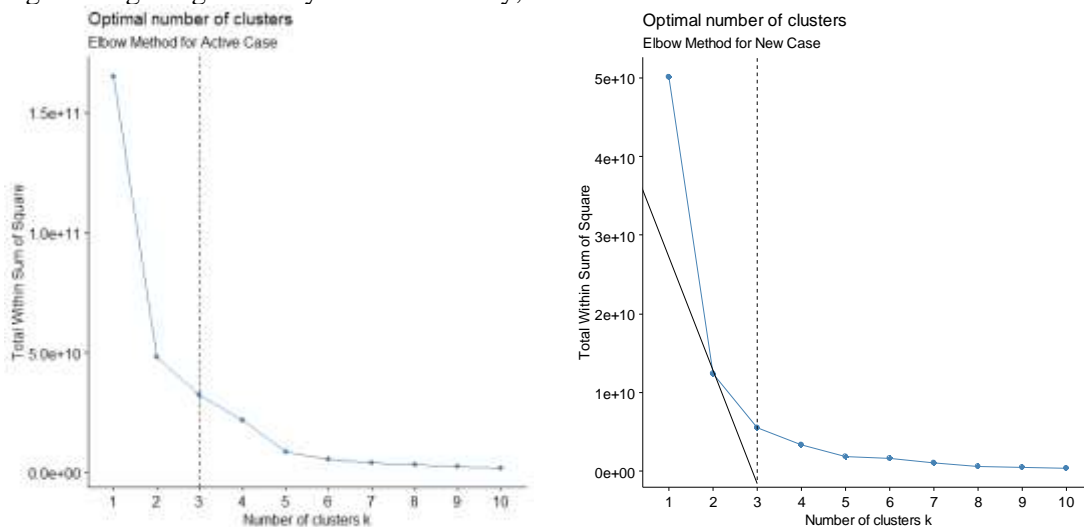
$$MAE = \frac{\sum_{t=1}^n |X_t - \hat{X}_t|}{n} \quad (7)$$

$$MAPE = \frac{\sum_{t=1}^n \frac{|X_t - \hat{X}_t|}{X_t}}{n} \times 100\% \quad (8)$$

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 results of the evaluation carried out.

4. Results and Discussion

The data used in this study is weekly active case data from 34 provinces for the period 5 March 2020 – 14 August 2021. Using the Elbow Methods, the optimal number of clusters for active case data and new COVID-19 cases from 34 provinces is 3 clusters. This can be seen in Figure 3. The figure shows that after 3 clusters, the total within the sum of squares no longer changes significantly. So in this study, the number of clusters to be formed is 3.



(a)

(b)

Figure 3. Optimal number of cluster according to Elbow Methods (a) Active cases (b) Weekly new case

Using K-Medoid Clustering with K=3. Table 2 shows the clusters of the number of active COVID-19 cases in 34 provinces in Indonesia.

Table 2. COVID-19 active cases clusters of Provinces in Indonesia

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

After obtaining the temporary clusters in Table 2, to ensure the consistency of the clusters, three iterations were performed. The result until the 3rd iteration of the cluster does not change anymore. Therefore, table 2 shows the final classification for the number of active COVID-19 cases in 34 provinces in Indonesia. After clustering active cases, the same thing was done for new cases of COVID-19 in 34 provinces in Indonesia. Table 3 shows clusters based on the number of new cases of COVID-19 Provinces in Indonesia.

Table 3. COVID-19 weekly new cases clusters of 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 3. Shows temporary clusters for weekly new cases of COVID-19 provinces in Indonesia. To ensure consistency, it is repeated 3 times. The result is that the cluster has not changed, so table 3 shows the final cluster for the number of new COVID-19 cases in the provinces in Indonesia. Active case data in cluster 2 in table 2 will be used as training data for forecasting the number of active cases with historical data from West Java used for training and testing. While the new case data in cluster 3 in table 3 is used as training data for forecasting the number of new cases

with historical data from West Java used for training and testing. Furthermore, the process of training and testing on MLP using data obtained from the results of time series clustering. Figure 4 shows the actual value and forecast results for the number of active COVID-19 cases in West Java using MLP.

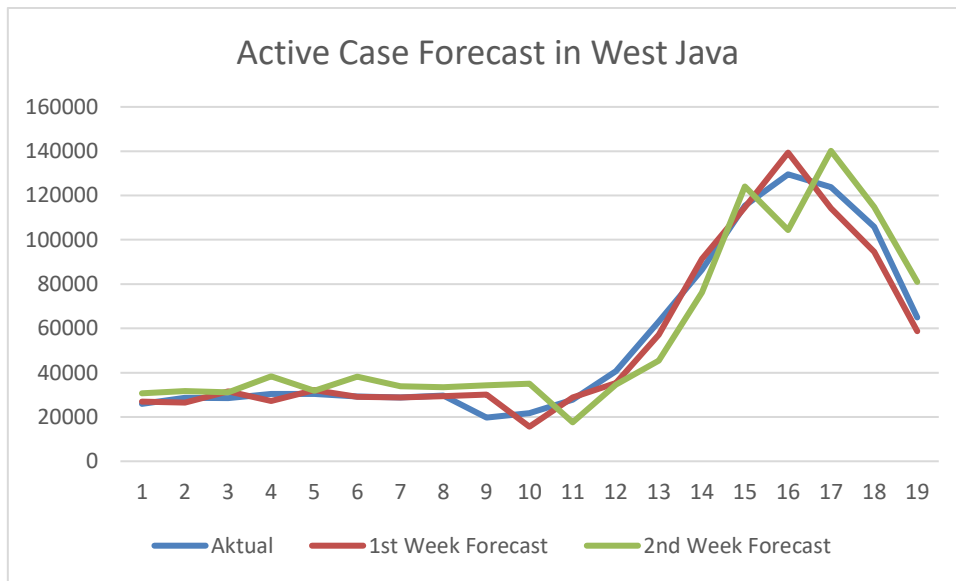


Figure 4. Plot the actual and predicted number of active cases of COVID-19 West Java

Based on Figure 4, it can be observed that forecasting for the next one week gives relatively more accurate results than forecasting for the next two weeks. It is known based on the data plot. The plot of the forecast results for the next week has a shape that is closer to the graph of the actual active cases that occur compared to the plot of the forecast results in the second week. This plot between the actual and predicted values shows the accuracy of the prediction results intuitively. To check the accuracy of the forecast results on weekly new cases, the same thing is done. Figure 5 shows the actual value and predicted results of the weekly number of new cases of COVID-19 in West Java MLP.

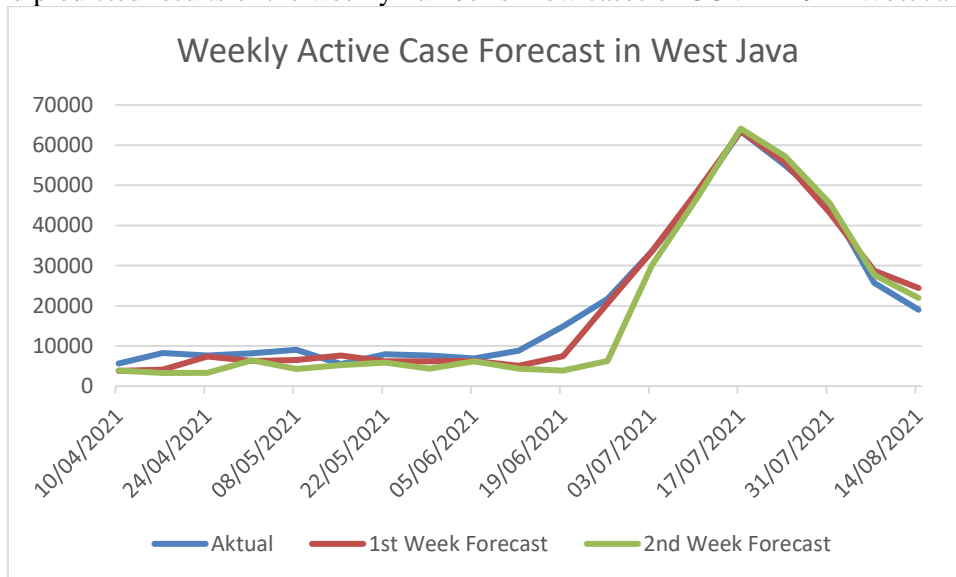


Figure 5. Plot the actual and predicted number of active cases of COVID-19 West Java for the second week

Figure 5 shows the MLP forecast results for weekly new cases. MLP provides forecast results for the coming week which are relatively closer to the actual value compared to those for the next two weeks. This is known based on the plot of the data presented. The plot of the forecast for the number of weekly new cases for the next week has a shape that is closer to the graph of the actual value occurring in West Java compared to the plot of the forecast for the next two weeks. A significant difference is found in the weekly forecast of new cases on June 26, 2021, the graph shows the forecast results for the second week have a value that is relatively far from the actual value compared to the forecast results for the first week. After plotting the data between the actual value and the predicted value, forecasting the number of active cases and 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 10 – August 14, 2021.

Table 4. Absolute percentage error of COVID-19 active case forecast

Week	From Date	To Date	1st Week APE	2nd Week APE
58	04/04/2021	10/04/2021	4.15%	18.74%
59	11/04/2021	17/04/2021	7.37%	10.80%
60	18/04/2021	24/04/2021	10.30%	8.83%
61	25/04/2021	01/05/2021	10.67%	25.85%
62	02/05/2021	08/05/2021	5.77%	4.81%
63	09/05/2021	15/05/2021	0.41%	30.46%
64	16/05/2021	22/05/2021	0.27%	17.31%
65	23/05/2021	29/05/2021	0.84%	12.80%
66	30/05/2021	05/06/2021	52.49%	73.81%
67	06/06/2021	12/06/2021	28.15%	61.21%
68	6/13/2021	6/19/2021	3.32%	36.81%
69	6/20/2021	6/26/2021	13.20%	15.05%
70	6/27/2021	7/3/2021	9.34%	28.04%
71	7/4/2021	7/10/2021	5.63%	11.81%
72	11/07/2021	17/07/2021	0.58%	7.61%
73	7/18/2021	7/24/2021	7.57%	19.40%
74	7/25/2021	7/31/2021	7.80%	13.20%
75	8/1/2021	8/7/2021	10.77%	8.44%
76	08/08/2021	14/08/2021	9.65%	24.60%

Based on table 4, it is found that the APE value for forecasting the number of weekly new cases for the coming week is relatively smaller than for forecasting the next two weeks. This is in line with what is shown in Figure 3. In addition, the largest APE value in forecasting the number of active COVID-19 cases in West Java was found in the forecasting period of June 5, 2021, with an APE value of 52.49% for the number of active cases in the coming week. and 61.21% for the number of active cases in the next two weeks. Using the APE limit of $\leq 20\%$, it was found that the use of MLP was successful in predicting the number of active COVID-19 cases accurately 17 times for forecasting the next 1 week and 12 times for forecasting the number of active cases in the next 2 weeks. To assess the accuracy of forecasting the number of new COVID-19 cases weekly in West Java, the same thing was done. Table 5 shows the absolute percentage error for forecasting the number of weekly new COVID-19 cases in West Java from April 10 – August 14, 2021.

Table 5. Absolute percentage error COVID-19 new case forecast

Week	From Date	To Date	1st Week APE	2nd Week APE
58	04/04/2021	10/04/2021	32.60%	31.83%
59	11/04/2021	17/04/2021	50.64%	60.52%
60	18/04/2021	24/04/2021	3.33%	56.72%
61	25/04/2021	01/05/2021	23.37%	21.97%
62	02/05/2021	08/05/2021	28.75%	53.20%
63	09/05/2021	15/05/2021	38.64%	4.48%
64	16/05/2021	22/05/2021	22.32%	25.78%
65	23/05/2021	29/05/2021	18.72%	42.24%
66	30/05/2021	05/06/2021	8.87%	10.80%
67	06/06/2021	12/06/2021	42.39%	50.33%
68	6/13/2021	6/19/2021	49.78%	74.14%
69	6/20/2021	6/26/2021	5.09%	71.07%
70	6/27/2021	7/3/2021	0.08%	10.21%
71	7/4/2021	7/10/2021	1.87%	1.82%

72	11/07/2021	17/07/2021	0.05%	1.01%
73	7/18/2021	7/24/2021	1.39%	4.15%
74	7/25/2021	7/31/2021	4.65%	0.63%
75	8/1/2021	8/7/2021	11.68%	7.69%
76	08/08/2021	14/08/2021	28.70%	15.57%

Table 5 shows that the APE value for forecasting the number of new cases in the next one week is relatively smaller than the APE value for forecasting the next two weeks. The largest APE value found in the forecast for the next one week was 50.64% in the forecasting period of April 17, 2021, for the second week of forecasting, the largest APE value was found in the period June 19, 2021, with an APE value of 74.14%. In addition, when compared with forecasting active cases, the APE value generated in weekly new case forecasting is relatively larger, this is also reflected in the accuracy of forecasting. Using the APE limit of 20%, it was found that the forecast for the weekly number of new cases of COVID-19 in West Java was accurate 10 times for the coming week and 9 times for the next two weeks. The overall performance of MLP in forecasting is shown in Table 6. The table shows the MAPE, RMSE, and MAE values for each setting used in MLP to predict the number of active cases and new cases of COVID-19 in West Java.

Table 6. MAPE, RMSE, and MAE of Multilayer Perceptron

	Active Case	New Case
MAPE	16.26%	24.13%
RMSE	9033	4154
MAE	7045	2857

Based on the results of the testing carried out in weeks 58-76 (10 April – 14 August 2021) it was found that the MAPE value for forecasting the number of active COVID-19 cases was smaller than the MAPE value for forecasting the number of new COVID-19 cases. As for the MSE, RMSE, and MAE values, the value for forecasting the number of new cases is relatively smaller than for forecasting active COVID-19 cases.

5. Conclusion

Active cases talk about COVID-19 patients who need medical care so that it is directly related to hospital capacity and weekly new cases are related to the efficiency of containment measures carried out by the government. These two things are determinants in the successful handling of the COVID-19 pandemic. Therefore, predicting the number of active cases and new cases is a very strategic thing to do. Based on the research results obtained, overall forecasting of active COVID-19 cases for the next two weeks using MLP provides relatively good accuracy compared to the use of other settings, namely with MAPE, RMSE, and MAE values of 16.26%, 9033, and 7045. MLP has been accurate in predicting the number of active cases for the next one-week period as many as 17 times, and the next two weeks as many as 12 times with an absolute percentage error (APE) < 20%. As for forecasting the number of new cases weekly, the results obtained are not as good as in forecasting active cases, with the forecasting results being accurate 10 times for the next week and 9 times for the next two weeks, with MAPE, RMSE, and MAE values of 24.13%, 4154 and 2857. Although MLP gives smaller RMSE and MAE values in forecasting the number of new cases weekly compared to forecasting the number of active cases, however, the percentage of errors that occur when compared between actual and predicted values still has a greater value than forecasting the number of active cases. active case.

In addition, it was also found that when there was a sudden acceleration in the number of active COVID-19 cases, the APE value jumped to a very high level, this was observed at weeks 66 and 67. The resulting APE value was the highest APE value for all settings used both for forecasting the first week or the second week. This can be a starting point for modeling that can overcome sudden acceleration in time series data. The right setting on the MLP can increase the accuracy of forecasting active cases of COVID-19 in West Java. Hopefully, the results of this study can be used as a reference by the government to forecast accurate COVID-19 active cases for the next two weeks. So, the government can condition the number of hospital beds by the existing prediction results. So that all COVID-19 patients can be treated properly, and no patient is rejected by the hospital because the hospital is full.

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) who have provided scheme funding of the Universitas Padjadjaran Lecturer Competency Research (RKDU), with a contract number:

1960/UN6.3.1/PM.00/2021. Any attempt at any level could not have been satisfactorily completed without the support and guidance of my wife and friends.

References

- [1] Nuraini, T. N., 2020. *Merdeka*. [Online] Available at: <https://www.merdeka.com/trending/kronologi-munculnya-covid-19-di-indonesia-hingga-terbit-keppres-darurat-kesehatan-klh.html> [Accessed 22 May 2021].
- [2] Gugus Tugas Percepatan Penanganan Covid-19, 2021. *Gugus Tugas Percepatan Penanganan Covid-19*. [Online] Available at: <https://covid19.go.id/peta-sebaran>. [Accessed 22 May 2021].
- [3] Worldometer, 2020. *Worldometer*. [Online] Available at: <https://www.worldometers.info/coronavirus/#countries> [Accessed 22 May 2021].
- [4] Wikipedia. (2021, 2 27). Wikipedia. Diambil kembali dari Wikipedia: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Indonesia
- [5] Pusat Informasi & Koordinasi COVID-19 Provinsi Jawa Barat. Available at: <https://pikobar.jabarprov.go.id/>.
- [6] Zulfikar, M. (2021, 1 25). Laporan Covid-19 terima 34 laporan pasien COVID-19 ditolak rumah sakit . Diambil kembali dari Antara News: <https://www.antaraneews.com/berita/1966256/laporcovid-19-terima-34-laporan-pasien-covid-19-ditolak-rumah-sakit>
- [7] Wijaya, L. (2021, 1 16). Ditolak 10 Rumah Sakit, Pasien Covid-19 Asal Depok Meninggal di Taksi Daring. Diambil kembali dari Tempo: <https://metro.tempo.co/read/1423712/ditolak-10-rumah-sakit-pasien-covid-19-asal-depok-meninggal-di-taksi-daring>
- [8] Amiruddin, F. (2020, 3 20). Kisah Pilu PDP Corona Pangandaran Ditolak 5 Rumah Sakit. Diambil kembali dari Detiknews: <https://news.detik.com/berita-jawa-barat/d-4946953/kisah-pilu-pdp-corona-pangandaran-ditolak-5-rumah-sakit>
- [9] Fikri, A. (2021, 2 1). Tegur RS Swasta di Jabar, RK: Ini Lagi Perang, Tak Boleh Tolak Pasien Covid. Diambil kembali dari Suara.com: <https://www.suara.com/news/2021/02/01/111145/tegur-rs-swasta-di-jabar-rk-ini-lagi-perang-tak-boleh-tolak-pasien-covid>
- [10] World Health Organization. Coronavirus disease 2019 (COVID-19) situation report–57. Geneva, Switzerland: World Health Organization; 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19>
- [11] "Q & A on COVID-19: Basic facts". www.ecdc.europa.eu. 25 September 2020. Retrieved 8 October 2020.
- [12] "CDC FAQ on COVID-19" <https://www.cdc.gov/coronavirus/2019-ncov/faq.html#Symptoms-&Emergency-Warning-Signs>
- [13] World Health Organization. Coronavirus disease 2019 (COVID-19) situation report–51. Geneva, Switzerland: World Health Organization; 2020. https://www.who.int/docs/default-source/coronaviruse/situationreports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10
- [14] Novel Coronavirus Pneumonia Emergency Response Epidemiology Team.(2020). The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China [Chinese]. *Chinese Center for Disease Control and Prevention Weekly* 2020;41:145.
- [15] "COVID-19 situation update for the EU/EEA and the UK". www.ecdc.europa.eu. 25 September 2020. Retrieved 8 October 2020.
- [16] Kaufman, L. Rousseeuw PJ., *Finding Groups in Data: An Introduction to Clustering Analysis*, NewYork : John Wiley, 1990.
- [17] C. Antunes, A.L. Oliveira, Temporal data mining: an overview, in: *KDD Workshop on Temporal Data Mining*, 2001, pp. 1–13
- [18] T. Warrenliao, Clustering of time series data—a survey, *Pattern Recognit.* 38 (11) (2005) 1857–1874.
- [19] S. Rani, G. Sikka, Recent techniques of clustering of time series data: a survey, *Int. J. Comput. Appl* 52 (15) (2012) 1–9.
- [20] J. Lin, M. Vlachos, E. Keogh, D. Gunopulos, Iterative incremental clustering of time series, *Adv. Database Technol* 2004 (2004) 521–522.
- [21] H. Wang, W. Wang, J. Yang, P.P.S. Yu, Clustering by pattern similarity in large data sets, in: *Proceedings of 2002 ACM SIGMOD International Conference Management data – SIGMOD '02*, vol. 2, 2002, p. 394.
- [22] G. Das, K.I. Lin, H. Mannila, G. Renganathan, P. Smyth, Rule discovery from time series., *Knowl. Discov. Data Min* 98 (1998) 16–22
- [23] T.C. Fu, F.L. Chung, V. Ng, R. Luk, Pattern discovery from stock time series using self-organizing maps, in: *Workshop Notes of KDD2001 Workshop on Temporal Data Mining*, 2001, pp. 26–29.
- [24] E. Keogh, S. Lonardi, B.Y. Chiu, Finding surprising patterns in a time series database in linear time and space, in: *Proceedings of the Eighth ACM SIGKDD*, 2002, pp. 550–556

- [25] W. He, G. Feng, Q. Wu, T. He, S. Wan, J. Chou, A new method for abrupt dynamic change detection of correlated time series, *Int. J. Climatol.* 32 (10) (2011) 1604–1614.
- [26] A. Sfetsos, C. Siriopoulos, Time series forecasting with a hybrid clustering scheme and pattern recognition, *IEEE Trans. Syst. Man Cybern* 34 (3) (2004) 399–405.
- [27] N. Pavlidis, V.P. Plagianakos, D.K. Tasoulis, M.N. Vrahatis, Financial forecasting through unsupervised clustering and neural networks, *Oper. Res.* 6 (2) (2006) 103–127.
- [28] Cheng, B. & D.M. Titterton, (1994). *Neural networks: a review from a statistical perspective*. s.l.: Statistical Science.
- [29] Gurney, K. (1997). *An Introduction To Neural Networks*. London: UCL Press.
- [30] Kandiran, E. & Hacinliyan, A., (2019). Comparison of Feedforward and Recurrent Neural Network in Forecasting Chaotic Dynamical System. *AJIT-e: Bilişim Teknolojileri Online Dergisi*, 10(37), pp. 31-44.
- [31] K. L. Du and M. N. S. Swamy, (2013). “*Neural Network and Statistical Learning*,” Montreal: Concordia University, pp. 86-87.
- [32] Datilo, P. M., Ismail, Z. & Dare, J., (2019). A Review of Epidemic Forecasting Using Artificial Neural Networks. *International Journal of Epidemiologic Research*, 6(3), pp. 132-143.
- [33] Al-qaness, M.A.A., Ewees, A.A., Fan, H., Abd El Aziz, M. (2020). Optimization Method for Forecasting Confirmed Cases of COVID-19 in China. *J. Clin. Med.* 2020, 9, 674.
- [34] Sujath, R., Chatterjee, J.M. & Hassanien, (2020). A.E. A machine learning forecasting model for COVID-19 pandemic in India. *Stoch Environ Res Risk Assess* 34, 959–972. <https://doi.org/10.1007/s00477-020-01827-8>
- [35] Allen, T. T. (2006). *Introduction to Engineering Statistics and Six Sigma*. London: Springer.
- [36] Montgomery, D. C. (2009). *Introduction to Statistical Quality Control* (6th ed.). NJ: Wiley: Hoboken.
- [37] Kaufman, L. Rousseeuw P.J., *Finding Groups in Data: An Introduction to Cluster Analysis*, New York : John Wiley, 1990.
- [38] 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).
- [39] Thorndike, R. L. (1953). Who belongs in the family?. *Psychometrika*, 18(4), 267-276.
- [40] Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, 1(6), 90-95.
- [41] Dietterich, T. G., (2002). Machine Learning for Sequential Data: A Review. *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, Volume 2396, pp. 15-30.
- [42] Python Software Foundation. Python Language Reference, version 2.7. Available at <http://www.python.org>
- [43] Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. CoRR.
- [44] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning* (1st ed.). London: MIT Press.
- [45] Panchal, Foram, S., & Panchal, M. (2014). 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.
- [46] Hyndman, R. J. & Athanopoulos, G., (2018). *Forecasting: principles and practice*. s.l.: OTexts