Forecasting the Number of COVID-19 Active Cases in West Java Using the Multilayer Perceptron Feedforwards Neural Networks

Yuyun Hidayat, Dhika Surya Pangestu, Titi Purwandari, Enny Supartini
Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia
yuyun.hidayat@unpad.ac.id; dhikasurya.surya7@gmail.com; titipurwandari@yahoo.com; enny.supartini@unpad.ac.id

Sukono
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia
sukono@unpad.ac.id

Abdul Talib Bon
Department of Production and Operations, University Tun Hussein Onn Malaysia, Malaysia
talibon@gmail.com

Abstract

Abstract. West Java reported that as of July 17, 2021, 508,814 persons had been proven positive for COVID-19. On the same day, there were 115,257 active COVID-19 cases in West Java. In West Java, there will be a large increase in active COVID-19 cases in 2021. The number of active cases increased by 95,532 between June 5 and July 17, 2021. During that time, active cases grew by 484 percent. On July 17, 2021, the Bed Occupancy Ratio (BOR) in West Java was at 80.48 percent, exceeding the WHO standard of 60 percent. This has an effect on the number of patients who are rejected at the COVID-19 referral hospital. COVID-19 patients in need of medical attention are the focus of active cases. Active cases were predicted in this study using a Multilayer Perceptron (MLP). The COVID-19 Task Force provided the data for this study. The data show the number of COVID-19 positive cases, recovered, and deaths in 34 Indonesian provinces from 2 March 2020 to 17 July 2021. The results of the study found, from the results of the evaluation using data testing the number of active cases in the last 15 weeks, namely April 10 – July 17, 2021, the use of adam 0.1 provides relatively good accuracy compared to the use of other settings, namely with MAPE, RMSE and MAE values of 17.21%, 7092.25, and 5470. From 15 testing periods, MLP is accurate in predicting the number of active cases the first week 13 times, and the second week 9 times with an absolute percentage error (APE) <20%. It is hoped that the findings of this study would be useful to the government as a reference in conditioning hospital bed capacity to deal with active COVID-19 cases in West Java during the following two weeks, so that no COVID-19 patients are turned away because the hospital is full.

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

I. INTRODUCTION

Indonesia is now one among the nations affected by the COVID-19 pandemic. COVID-19 was confirmed to have first seen on March 2, 2020 in Indonesia. At the time, two people had been exposed to COVID-19 through interaction with Japanese citizens. This was discovered after a Japanese citizen was diagnosed with the coronavirus shortly after leaving Indonesia and landing in West Java (Nuraini, 2020). Since the virus's first appearance, the number of COVID-19 cases in Indonesia has steadily increased, with 2,832,755 people affected as of July 17, 2021 (Gugus Tugas Percepatan Penanganan Covid-19, 2021). According to the Worldometer, Indonesia is ranked 14th 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 508,814 confirmed cases of West COVID-19 as of July 17, 2021, with a total of 115,257 active cases. Based on these figures, West Java was ranked second in Indonesia for positive cases and first for active COVID-19 cases, with 737,144 positive cases and 111,582 active cases. DKI Jakarta was ranked first for positive cases and second for active cases, with 737,144 positive cases and 111,582 active cases (Wikipedia, 2021). Active cases in the COVID-19 pandemic refer to COVID-19 patients who require medical attention and are directly tied to hospital capacity. Graph 1 depicts the development of the number of active COVID-19 cases in West Java from 2 March 2020 to 17 July 2021.

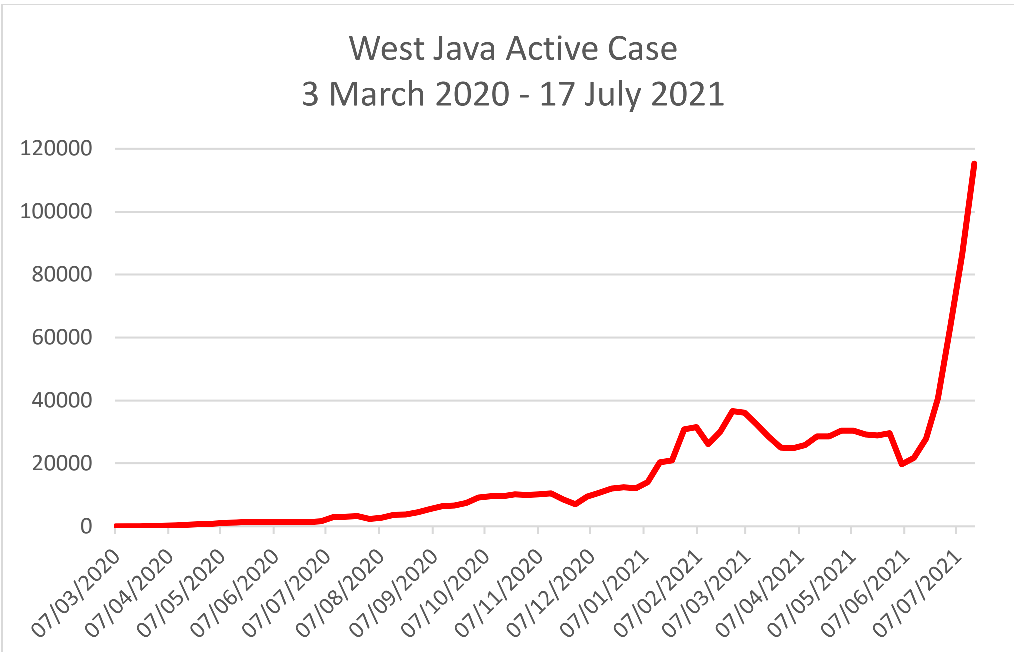


Figure 1. Graph of the development of the number of active cases in West Java Province 3 March 2020 – 17 July 2021

Figure 1 demonstrates that the number of active cases of COVID-19 in West Java has increased steadily from its first appearance on March 2, 2020 to July 17, 2021. It was also discovered that there was a considerable increase in June 2021, with the number of active COVID-19 cases in West Java increasing from 19,725 on 5 June 2021 to 115,257 on 17 July 2021. There was a rise in the number of active cases of up to 95,532 cases in one month, or a 484 percent increase in the number of active cases during that time period. This circumstance demonstrates the COVID-19 pandemic's emergency in West Java.

Active cases involve COVID-19 individuals who require medical attention and are directly tied to hospital capacity. 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 till July 17, 2021 was 80.48 percent (PIKOBAR, 2021), which is significantly higher than the WHO standard of 60 percent. 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 21, 2021, there were 34 complaints of COVID-19 patients who were denied by hospitals in Indonesia, according to LaporCovid-19 (Zulfikar, 2021). Refusal of COVID-19 patients happened in various regions around West Java, including Depok (Wijaya, 2021) and Pangandaran (Amiruddin, 2020). 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 do not let anyone be rejected (Fikri, 2021).

The trend of increasing the number of active COVID-19 cases to date continues; this has to be addressed. One of the issues that the government has is dealing with a likely increase in active COVID-19 cases. Hospital BOR numbers in West Java have exceeded WHO guidelines, and there has also been a refusal of COVID-19 patients in the province. This is something that the government should be aware of. If there are active cases that surpass the hospital's capacity, it is believed that the fatality rate may rise as a result of patients not obtaining sufficient medical treatment. Appropriate policies are required to overcome COVID-19 patients' rejection. One of them is that the government can prepare its health-care facilities to deal with a future rise in active cases. Predicting the number of active COVID-19 cases is thus a strategic move. Prediction of active cases can be used to condition existing health facilities as well as to ensure that the government's resources are adequate to deal with the number of active cases that will exist in the future. Accurate forecasts will substantially aid the government in identifying next steps and plans to combat the COVID-19 outbreak in West Java. As a result, COVID-19 positive patients can receive quality care, and no one is turned away because the hospital is full.

II. LITERATURE REVIEW

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 (WHO, 2020). 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 (ECDC, 2020). 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 (CDC, 2020). On March 11, 2020, WHO declared that COVID-19 was a pandemic (WHO, 2020). 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 . (Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, 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 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 (ECDC, 2020).

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 (Kaufman & Rousseeuw , 1990). 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 (Antunes & Oliveira, 2001). 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 (Warrenliao, 2005; Rani & Sikka, 2012; Lin et al, 2004). Clustering of time-series data is mostly utilized for discovery of interesting patterns in time-series datasets (Wang et al, 2002; Das et al, 1998). 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 (Fu et al, 2001). The second group are methods to discover patterns which happened in datasets surprisingly (Keogh et al, 2002). 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 (He et al, 2011; Sfetsos & Siriopoulos, 2004). 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 (Pavlidis et al, 2006). 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.

Artificial Neural Networks

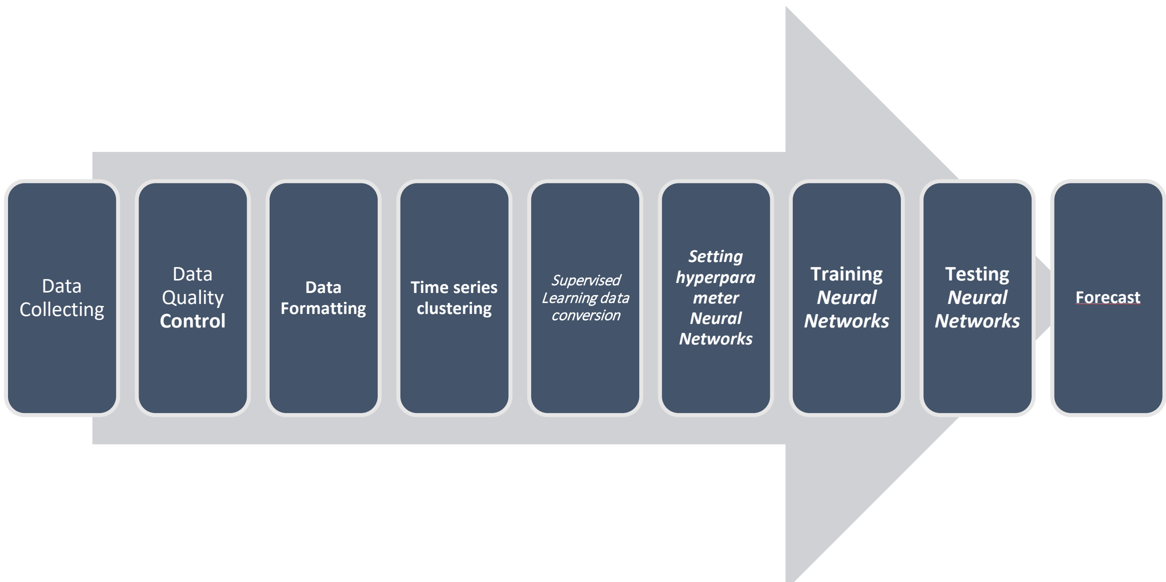
Artificial Neural networks (ANN) are a set of computational units or nodes, which are based on the function of neurons in animals. The ability to process ANN is found in the relationship between neurons, or what is called weights, which is obtained by adapting to learning a set of patterns obtained from training data. ANN is commonly used for statistical analysis and data modeling (Cheng & Titterington, 1994). Besides that, ANN is also commonly used in classification or forecasting (Gurney, 1997). 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 (Kandiran & Hacinliyan, 2019).

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) (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 of the predictive epidemiological studies use MLP (Datilo et al, 2019; Al-qaness et al, 2020; Sujath et al, 2020)

III. RESEARCH MEHODOLOGY

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 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 of COVID- in West Java using MLP

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 July 17, 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:

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

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, 2006). 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, 2009). 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.

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 (Kaufman & Rosseuw, 1990). 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 :

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

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 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 (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 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 (Kodinariya& Makwana, 2013).

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 (2).
2. Calculate v_j for each object j with $d_i. = \sum_{j=1}^{n} d_{ij}$:

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

   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.

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 ($x_{i,t-d}$, $x_{i,t-d+1}$,...,$x_{i,t}$,...,$x_{i,t+d-1}$, $x_{i,t+di}$) 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 (Dietterich, 2002).

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 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 (Kingma & Ba, 2015). 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 (Goodfellow et al, 2016).

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 (Panchal et al, 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. below shows the parameters tested in the study to predict the number of active cases of COVID in West Java in the next two weeks.

Table 1. MLP parameter setting

Hyperparameter	**Setting**
Learning Rate	Adam (0.1, 0.01, 0.001)
Input Neurons	3
Hidden Layer	1
Hidden Neurons	10
Maximum Epochs	1000

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 (Hyndman & Athanasopoulos, 2018). 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.

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

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

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

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

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

V. RESULT AND DISCUSSION

Forecasting COVID-19 cases in West Java using the MLP method is carried out with the following procedure. The data used in this study is weekly active case data from 34 provinces for the period 5 March 2020 – 17 July 2021. Using the Elbow Methods, the optimal number of clusters for data on active COVID-19 cases from 34 provinces is 3 clusters. This can be seen from Figure 3. The figure shows that after 3 clusters, the total within sum of squares no longer changes significantly. So in this study the number of clusters to be formed is 3.

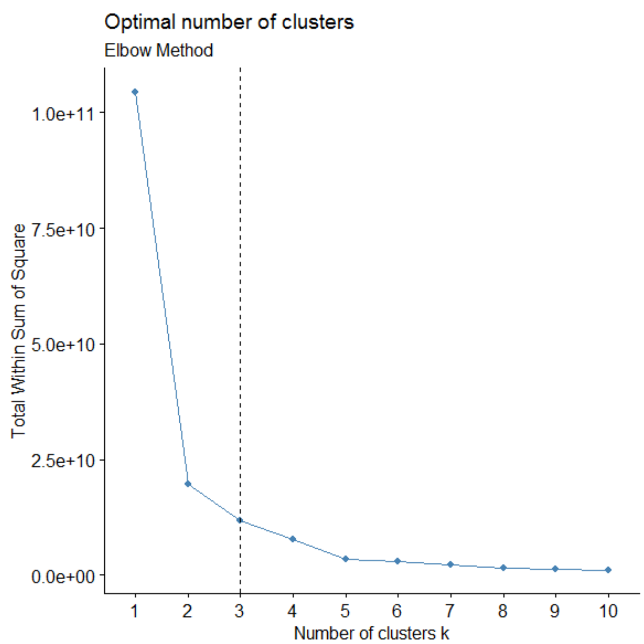


Figure 3. Determining the optimal number of clusters using Elbow Methods

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. Clusters of active cases of COVID-19 Provinces in Indonesia

Cluster	Provinces

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

After getting the temporary cluster in Table 2. then iterated 3 times. The result until the 3rd iteration of the cluster does not change anymore. Therefore, Table 2 shows the final cluster for the number of active COVID-19 cases in 34 provinces in Indonesia.

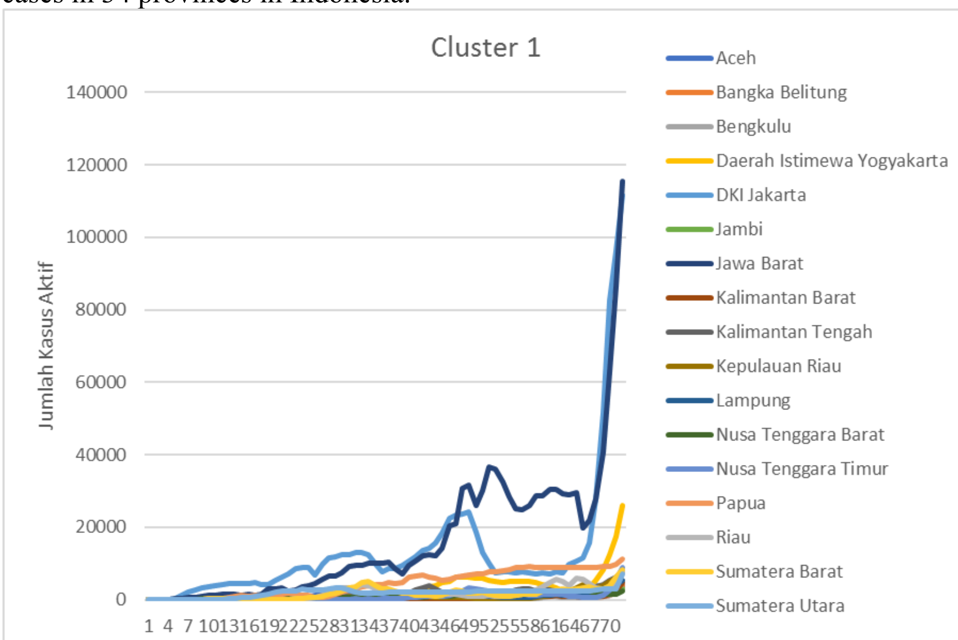


Figure 4. Graph of the development of the number of active COVID-19 cases in Cluster 1

Figure 4. shows the development of the number of active COVID-19 cases in Cluster 1. This cluster is the cluster with the most members, namely 17 Provinces. West Java is included in this cluster, and this cluster will be used for the training and testing process on MLP. The similarity in this cluster is in the tendency to continue to rise, with fluctuations occurring between weeks 49-66, before experiencing an increase again until the last week of observation, which is week 72.

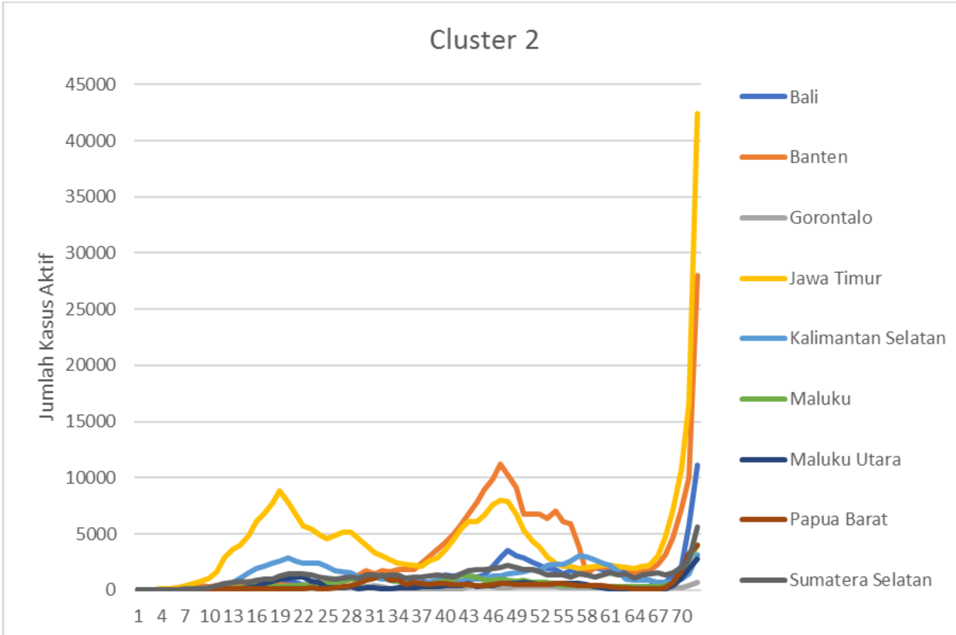


Figure 5. Graph of the development of the number of active COVID-19 cases in Cluster 2

Figure 5 shows the development of the number of active cases in Cluster 2. This cluster consists of 9 Provinces. In general, members of this cluster experienced two peaks, the first between weeks 16 and 23, and the second between weeks 43 and 49. This cluster also experienced an increase from weeks 66 to 72.

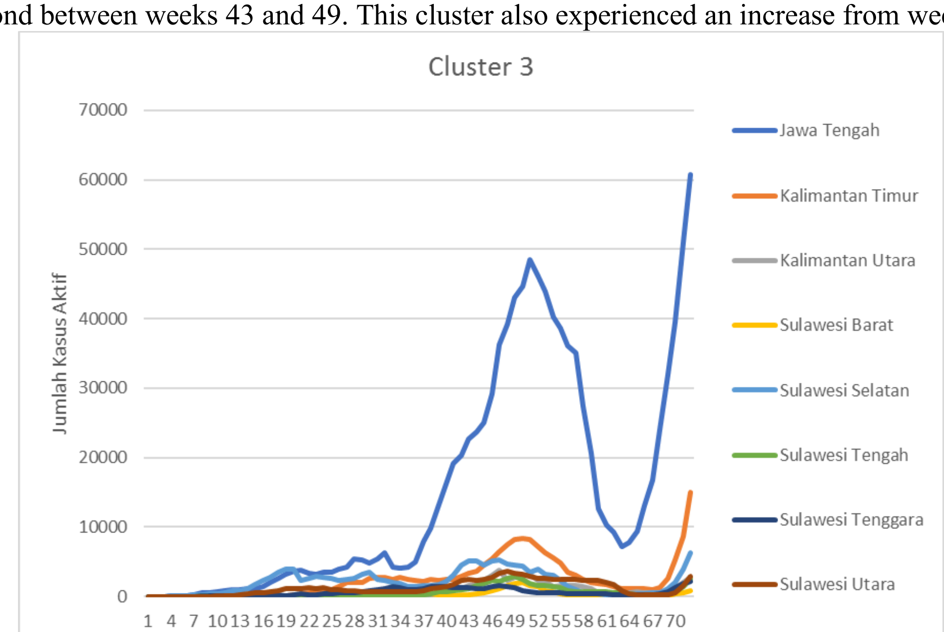


Figure 6. Graph of the development of the number of active COVID-19 cases in Cluster 2

Figure 6 shows the development of the number of active COVID-19 cases in Cluster 3. This cluster is a cluster with at least 8 provinces. The similarity of this cluster is seen at the peak that occurred between weeks 47-57, before further decreasing, and starting to rise again at weeks 63 to 72. In all the bound clusters, after week 66, there was a significant increase that continued until week 72, this is in line with national data showing a similar trend. Active case data in cluster 1 will be used as training data, while

West Java historical data will be used for training and testing. Figure 7 shows the actual value and the estimated results of the number of active COVID-19 cases in West Java for the first week using the MLP.

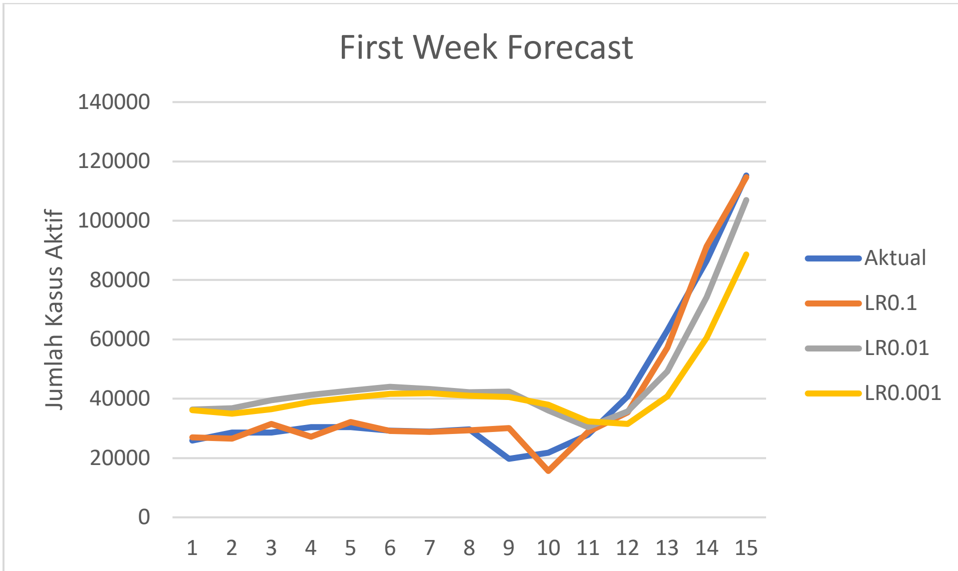


Figure 7. Plot the actual and predicted number of active cases of COVID-19 West Java for the first week

Based on Figure 7, it can be observed that the use of adam 0.1 gives a value that is closest to the actual case compared to the use of adam 0.01 and 0.001. This can be known based on the data plot that is close to the actual value. Figure 8 shows the actual value and predicted results of the number of active COVID-19 cases in West Java for the second week using MLP.

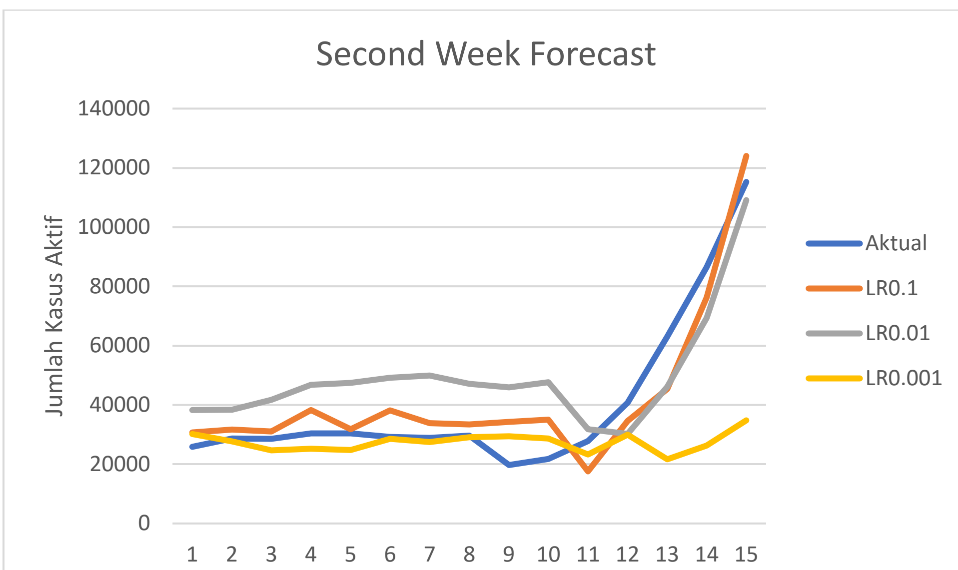


Figure 8. Plot the actual and predicted number of active cases of COVID-19 West Java for the second week

Figure 8 shows the use of adam 0.1 provides forecast results that are closest to the actual case compared to the use of adam 0.01 and 0.001. This is similar to the napa shown in Figure 7. In addition to observations on the comparison of forecast results and actual cases, forecasting accuracy is also evaluated based on the accuracy of forecasting per period. This is shown in Figure 9 which shows the absolute percentage error (APE) value in forecasting the number of active COVID-19 cases in West Java using MLP.

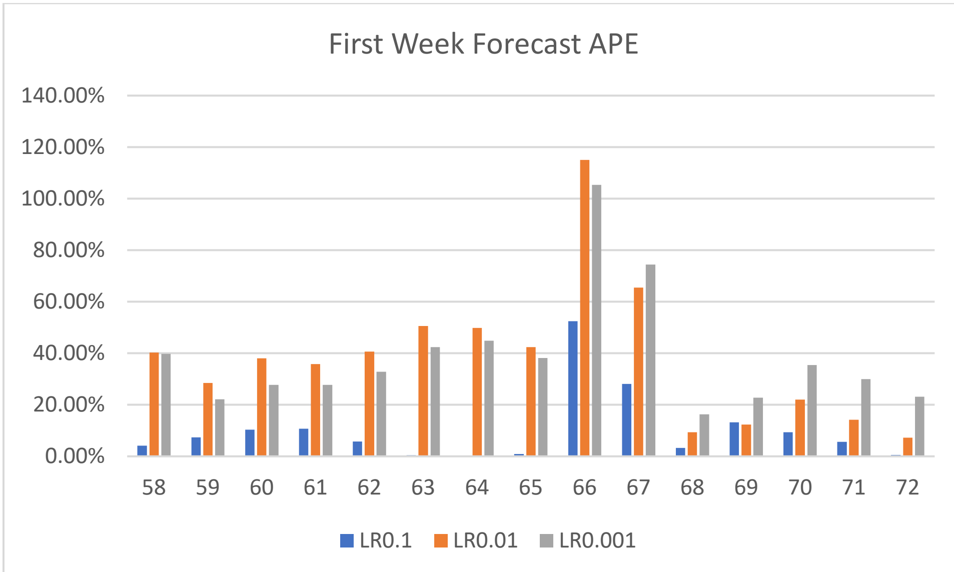


Figure 9. APE value for the first week active cases forecast

Figure 9 shows that the use of adam 0.1 gave the smallest APE value in the entire testing period (58th week to 72nd week). This shows that in the first week of forecasting, the use of adam 0.1 provides the most accurate forecast value compared to the use of adam 0.01 and 0.001. In the first week of forecasting, the largest APE value was achieved in the 66th-week forecast, with all APE values above 40%. During this period there was a very sharp increase in the number of active cases and continued until the 72nd week. Furthermore, the APE Forecast value for the second week is shown in Figure 10.

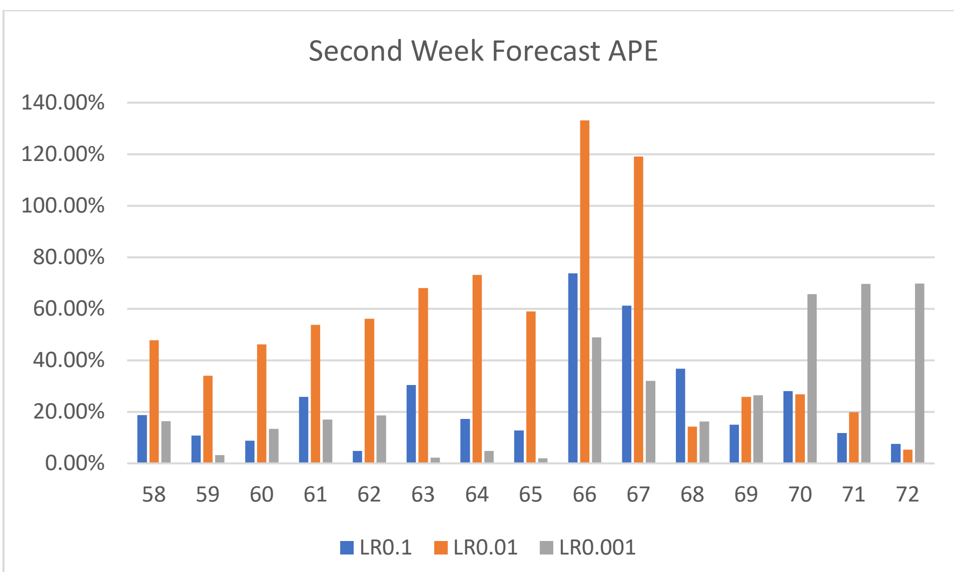


Figure 10. APE value for the second week active cases forecast

Figure 10 shows that in general the APE value obtained is greater than the forecast for the number of active cases in the first week. In this graph, it can be seen that the use of adam 0.1 and 0.001 gives a smaller APE value than the use of adam 0.01 except for weeks 68-72. At week 68 the greatest APE value was obtained by adam 0.1 with a value of 36.81% while for weeks 69-72 the largest APE value was obtained by using adam 0.001. The overall performance of MLP in forecasting is shown in Table 3. The table shows the MAPE, RMSE, and MAE values for each setting used in MLP to predict the number of active COVID-19 cases in West Java.

Table 3. MAPE, RMSE, and MAE of Multilayer Perceptron

	Adam 0.1	**Adam 0.01**	**Adam 0.001**
MAPE	**17.21%**	45.13%	32.98%
RMSE	**7092.251908**	14769.62202	22873.96893
MAE	**5470**	13585	14724

Based on the results of the testing carried out at weeks 58-72, the largest MAPE, MSE, RMSE and MAE values obtained using a learning rate of 0.01. The smallest MAPE, RMSE and MAE values are found in the use of a learning rate of 0.1 , with MAPE, RMSE and MAE values of 17.21%, 7092.25, and 5470.

VI. CONCLUSION

Active cases involve COVID-19 individuals who require medical attention and are directly tied to hospital capacity. As a result, forecasting the number of active COVID-19 cases is a very strategic move. According to the research findings, overall forecasting of active COVID-19 cases for the next two weeks using MLP with adam 0.1 provides rather good accuracy compared to the usage of other settings, with MAPE, RMSE, and MAE values of 17.21 percent, 7092.25, and 5470, respectively. MLP with this setting predicted the number of active cases 13 times in the first week and 9 times in the second week with an absolute percentage error (APE) of 20%. Furthermore, it was discovered that when there was a rapid increase in the number of active COVID-19 cases, the APE value increased to a very high level, as seen during weeks 66 and 67. The resulting APE number was the highest APE value for all forecasted settings employed for either the first or second week. This can be used as a starting point for modeling that can deal with rapid acceleration in time series data. The correct MLP settings can improve the accuracy of forecasting active COVID-19 cases in West Java. Hopefully, the findings of this study will be used by the government as a reference to estimate accurate COVID-19 active cases for the next two weeks. As a result, the government can condition the number of hospital beds based on the existing projection findings. So that all COVID-19 patients can be adequately treated and no patient is turned away because the hospital is full.

ACKNOWLEDGMENT

Acknowledgments are conveyed to the Director of General of Higher Education of the Republic of Indonesia (Deputy for Strengthening and Development of the Ministry of Research and Technology/National Research and Innovation Agency), and Chancellor, Director of the Directorate of Research, Community Engagement, and Innovation, and the Dean of the Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, who have provided the Applied Research Grants (HPT) with the Contract Number: **1207/UN6.3.1/PT.00/2021**.

REFERENCE

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.

Allen, T. T. (2006). Introduction to Engineering Statistics and Six Sigma. London: Springer.

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.

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.

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).

C. Antunes, A.L. Oliveira, Temporal data mining: an overview, in: KDD Workshop on Temporal Data Mining, 2001, pp. 1–13

CDC FAQ on COVID-19. https://www.cdc.gov/coronavirus/2019 ncov/faq.html#Symptoms-&-Emergency-Warning-Signs

Cheng, B. & D.M.Titterington ,. (1994) . Neural networks: a review from a statistical perspective. s.l.:Statistical Science.

COVID-19 situation update for the EU/EEA and the UK. www.ecdc.europa.eu. 25 September 2020. Retrieved 8 October 2020.

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.

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.

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

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.

G. Das, K.I. Lin, H. Mannila, G. Renganathan, P. Smyth, Rule discovery from time series,, Knowl. Discov. Data Min 98 (1998) 16–22

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning (1st ed.). London: MIT Press.

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].

Gurney, K. (1997) . An Introduction To Neural Networks. London: UCL Press.

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.

Hyndman , R. J. & Athanasopoulos, G.,. (2018) . Forecasting: principles and practice. s.l.: OTexts.

J. Lin, M. Vlachos, E. Keogh, D. Gunopulos, Iterative incremental clustering of time series, Adv. Database Technol 2004 (2004) 521–522.

K. L. Du and M. N. S. Swamy,.(2013). "Neural Network and Statistical Learning," Montreal: Concordia University, pp. 86-87.

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.

Kaufman, L. Rousseeuw PJ., Finding Groups in Data: An Introduction to Cluster Analysis,NewYork : John Wiley, 1990.

Kaufman, L. Rousseeuw PJ., Finding Groups in Data: An Introduction to Clustering Analysis, NewYork : John Wiley, 1990.

Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. CoRR.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. International Journal, 1(6), 90-95.

Montgomery, D. C. (2009). Introduction to Statistical Quality Control (6th ed.). NJ: Wiley: Hoboken.

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.

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.

Nuraini, T. N., 2020. Merdeka. [Online] Available at: https://www.merdeka.com/trending/kronologi-munculnya-covid-19-di-indonesia-hingga-terbit-keppres-darurat-kesehatan-kln.html [Accessed 22 May 2021].

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.

Pusat Informasi & Koordinasi COVID-19 Provinsi Jawa Barat. Available at: https://pikobar.jabarprov.go.id/. (PIKOBAR, 2021)

Python Software Foundation. Python Language Reference, version 2.7. Available at http://www.python.org

Q & A on COVID-19: Basic facts. www.ecdc.europa.eu. 25 September 2020. Retrieved 8 October 2020.

S. Rani, G. Sikka, Recent techniques of clustering of time series data: a survey, Int. J. Comput. Appl 52 (15) (2012) 1–9.

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

T. Warrenliao, Clustering of time series data—a survey, Pattern Recognit. 38 (11) (2005) 1857–1874.

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.

Thorndike, R. L. (1953). Who belongs in the family?. Psychometrika, 18(4), 267-276.

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.

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

Wikipedia. (2021, 2 27). Wikipedia. Diambil kembali dari Wikipedia: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Indonesia.

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.

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

Worldometer, 2020. Worldometer. [Online] Available at: https://www.worldometers.info/coronavirus/#countries [Accessed 22 May 2021].

Zulfikar, M. (2021, 1 25). LaporCovid-19 terima 34 laporan pasien COVID-19 ditolak rumah sakit . Diambil kembali dari Antara News: https://www.antaranews.com/berita/1966256/laporcovid-19-terima-34-laporan-pasien-covid-19-ditolak-rumah-sakit.

Biographies

Yuyun Hidayat is a lecturer in the Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Pdjadjaran, Indonesia. He received his PhD in the University of Malaysia, Terengganu, Malaysia. He has published more then 20 papers in national and international journals and participated in many national and international conference. Currently initiated accurate predictions about Covid-19 case data in Indonesia

Dhika Surya Pangestu, is a graduate student in the Department of Mathematics of Universitas Padjadjaran. He received his bachelor degree in Statistics, in the Departement of Statistics Faculty of Mathematics and Natural Science, Universitas Padjadjaran, Indonesia.

Titi Purwandari, is a lecturer in the Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Pdjadjaran, the field of applied statistics, with a field of economic statistics.

Enny Supartini, is a lecturer in the Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Pdjadjaran, the field of applied statistics, with a field of quality control statistics.

Sukono is a lecturer in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran. Currently as Chair of the Research Collaboration Community (RCC), the field of applied mathematics, with a field of concentration of financial mathematics and actuarial sciences.

Abdul Talib Bon is a professor of Production and Operations Management in the Faculty of Technology Management and Business at the Universiti Tun Hussein Onn Malaysia since 1999. He has a PhD in Computer Science, which he obtained from the Universite de La Rochelle, France in the year 2008. His doctoral thesis was on topic Process Quality Improvement on Beltline Moulding Manufacturing. He studied Business Administration in the Universiti Kebangsaan Malaysia for which he was awarded the MBA in the year 1998. He’s bachelor degree and diploma in Mechanical Engineering which his obtained from the Universiti Teknologi Malaysia. He received his postgraduate certificate in Mechatronics and Robotics from Carlisle, United Kingdom in 1997. He had published more 150 International Proceedings and International Journals and 8 books. He is a member of MSORSM, IIF, IEOM, IIE, INFORMS, TAM and MIM.