Forecasting Covid-19 Active Cases in Bandung City Using the Long Short Term Memory

Yuyun Hidayat, Mohammad Niks Iksan, Titi Purwandari, Enny Supartini
Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia
yuyun.hidayat@unpad.ac.id; niksiksan3221@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

COVID-19 in Indonesia was first confirmed on March 2, 2020. On May 8, 2021, according Pusicov, Bandung City had reached 707 active cases. The increase in the number of active cases indicates whether the government can not make good decisions in order to decrease the number of new cases or the service of hospital is not good in dealing with COVID-19 patients so the number of recovered is not likely increase. Therefore, forecasting the number of active COVID-19 cases in Bandung can be used to evaluate the government and hospital in facing COVID-19. To overcome the problem of analysis using conventional methods, this research will use Long Short Term Memory (LSTM) method. This method doesn't require parametric assumptions and can be used for data with long time periods, as there is a cell state to overcome vanishing gradients in the Recurrent Neural Network method. Data is the number of positive cases, recovered, and death in 27 cities in West Java in the period March 02, 2020 – May 08, 2021 which were obtained from the pusicov and pikobar. Based on the results, the Mean Absolute Percentage Error (MAPE) is 31,46% for data testing.

Keywords
Active Cases, COVID-19, Long Short Term Memory, Mean Absolute Percentage Error

Introduction

COVID-19 is a new type of Coronavirus that has never been previously identified in humans. COVID-19 in Indonesia was first confirmed on March 2, 2020. Each individual can be infected with the virus COVID-19 if inhaling spark from an infected person. When someone is confirmed positive COVID-19 with symptoms, then that person can be hospitalized COVID-19 which has appropriate facilities that service standards have been determined (Keputusan Menteri Kesehatan Nomor HK.91.07/MENKES/413/2020). On May 08, 2021, Bandung City was the fifth-largest city with cumulative confirmed cases of COVID-19 in West Java (Pusicov and Pikobar, 2021). While the number of Active Case in Bandung City was 707 active cases.

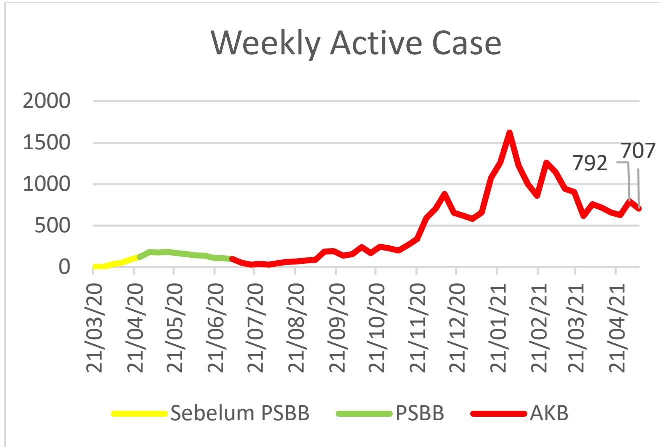

Figure 1. Weekly Active Case in Bandung City

From Figure 1, it can be seen that the number of active cases during PSBB (Pembatasan Soial Berskala Besar) is decreased. Howefer, after the PSBB was replaced by AKB (Adaptasi Kebiasaan Baru), the number of active cases is higher than before. Therefore, weekly forecasting the number of active COVID-19 cases in Bandung City can be used to evaluate whether the government cannot make good decisions in order to decrease the number of new cases or the service of hospital is not good in dealing with COVID-19 patients so the number of recovered is not likely increase. But before doing the predictive, 27 cities in West Java will be clustered in order to know which are the city that have trend of active COVID-19 cases similar to Bandung City.

Literature Review

Clustering

Cluster analysis 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 dan Rousseuw, 1990). Clustering is done by similarities or dissimilarities distance. There are two type of Clustering: Hierarchical Clustering and Non Hierarchical Clustering. The hierarchical method is forming a certain level such as in a tree structure because the grouping process is carried out in stages. In contrast to Hierarchical Method, the Non Hierarchical Clustering do not involve the treelike construction process. Instead, they assign objects into clusters once the number of clusters is specified. The grouping process in hierarchical analysis is not suitable for data with a large number of observations with a limit of no more than 1000 data, if the sample is greater than 1000, this can complicate the grouping process and determining the number of groups that are subjective based on dendrogram (Hair Jr et al., 1995). Therefore, this research will use Non Hierarchical Clustering to grouping 27 City in West Java because this type of clustering can process large data in a short time. The examples of Non Hierarchical Clustering are K-Means, K-Modes, and K-Medoids.

One of the simplest non-hierarchical methods is K-Means clustering. K-Means is the method that dividing data into several partitions with the mean or average into centroids, each data that is close to the average of each partition will be a group, after that iteration is done to find optimal results (Johnson and Wichern, 2002). This method can not apply to this research because the data is time series, if forecasting done by K-means, there will be shape averaging. Shape averaging is the average of time series is not the real average so it can not be made the centroid (Niennattrakul and Ratanamahatana, 2007). Beside that, K-means is very sensitive to the outlier.

The concept of K-Modes clustering is grouping the object based on their proximity to the centroid, namely mode. This method is applied for large categorical data (Huang, 1998). K-Modes is not suitable for this research because of numerical data, while K-Modes works on categorical data. Mode is not appropriate to use in this research because the COVID-19 active cases in several cities have the same value at the beginning of its appearance so that the mode only represents data in the initial period.

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, 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. So in order to get the data which has similarity trend to Bandung City, this research will use K-Medoids method.

Forecasting

The common method in forecasting is ARIMA (Autoregressive Integrated Moving Average). But in ARIMA, overfitting may occur if too many variables are included in the selected model or complex nonlinear models are used to estimate linear relationships or simple non-linear relationships (Liu, 2000). To overcome this problem, Artificial Neural Networks (ANN) is believed to get high accuracy. ANN is divided into two types, Feed Forward Neural Networdk (FFNN) and Recurrent Neural Network (RNN) (Puspitaningrum, 2006). FFNN with backpropagation network offers good performance, but this performance could be improved by using recurrence or reusing past inputs and outputs (Lawrence, 1997). But in the RNN, there is a vanishing gradient problem during the backpropagation process as training data periods increase (Ashrovy, 2017). Therefore, there is a development of the RNN method, namely Long Short Term Memory (LSTM) that can be overcome the vanishing gradient. Before doing the analysis, LSTM needs a lot of data training in order to get better accuracy or lower error in data testing. The data of Active Case in Bandung City only have 60 weeks (period) can be formed into more training data by taking data on active cases of COVID-19 in city of West Java which is similar to the Bandung City to be trained by LSTM.

Methods

K-Medoids Clustering

The first step in K-Medoids is to get Dynamic Time Warping (DTW) which calculated by D* matrix n x m for distance d_{ij}^*.

$$d_{ij}^* = (s_i - t_j)^2, i = 1,2,3, \dots, m \text{ and } j = 1,2,3, \dots, n \tag{1}$$

Where s and t are two time series data which has m x n dimension. After that, calculate matrix E by adding up d_{ij}^* with the minimum value of the three closest elements with the following formula

$$e_{ij} = d_{ij}^* + \min \{e_{(i-1)(j-1),} e_{(i-1),} e_{(j-1),}\} \tag{2}$$

E Matrix has various possible paths which will become the distance from DTW. The distance DTW is the minimum distance between two time series S and T which is calculated by the square root of the smallest cumulative distance in the cell (i, j) (Niennattrakul and Ratanamahatana, 2007).

$$D_{DTW}(S,T) = min_{\forall W} \left[\sqrt{\sum_{l=1}^{L} e_{wl}} \right], w = 1,2,3, \dots, W \tag{3}$$

Where W is a bunch of all possible warping paths, w is the element of W which come from E matrix with $w \in i$ so $W \leq m$, L is the length of warping path, l is the element of L which come from E matrix with $l \in j$ so $L \leq n$.

The next step is determine the number of groups. Pseudo F is the method to get the number of groups. The highest value of Pseudo F indicate that the optimal of number of groups, where the diversity within groups is homogeneous while between groups is heterogeneous (Suherman et al., 2021). To get Pseudo F-statistic value with the following formula (Calinski and Harabaz, 1974).

$$\textbf{Pseudo F} = \frac{(R^2/_{k-1})}{(1 - R^2/_{n-k})} \tag{4}$$

With:

$$R^2 = \frac{(SST - SSW)}{SST}, SST = \sum_{i=1}^{n_k}\sum_{j=1}^{k}\sum_{v=1}^{p}(x_{ijv} - \bar{x}_v)^2, SSW = \sum_{i=1}^{n_k}\sum_{j=1}^{k}\sum_{v=1}^{p}(x_{ijv} - \bar{x}_{jv})^2$$

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

a) Calculate the distance of each object using DTW distance with equation (X).
b) Calculate v_j for each object j with $d_i. = \sum_{j=1}^{n} d_{ij}$:

$$v_j = \sum_{i=1}^{n} \frac{d_{ij}}{d_i}, j = 1, \ldots, n \qquad (5)$$

d_{ij}: Distance matrix elements DTW
v_j : Standardization of the number of rows for each column j

c) Arrange v_j from smallest to largest. Select k cluster which have the first smallest v_j as the center (medoid).
d) Allocate objects that are non-medoid to the nearest medoid based on the distance of DTW.
e) Calculate the total distance from non-medoid cluster to the center.
f) 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.
g) Allocate objects that are non-medoid to the nearest medoid based on the distance of DTW.
h) Calculate the total distance from non-medoid cluster to the center.
i) 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 smallest pseudo-F with formula (4).

Long Short Term Memory

The LSTM method was introduced by Hochreiter and Schmidhuber. LSTM is a variant of RNN that can solved problems in RNN by adding cell states or memory cells with constant errors. LSTM consists of three gates, namely an input gate, a forget gate, and an output gate. Before modelling, data for Bandung City must be splitted into training and testing and the others city that similar to Bandung City will be trained. The size of the test data is usually about 20% of the total data, although this value depends on the amount of data and the size of the predictions (Hyndmann and Anthanasopoulos, 2018). After that, the data is normalized using the min-max scaling technique.

$$X_i' = \frac{(x_i - x_{min})}{x_{max} - x_{min}} \qquad (6)$$

Where X_i' is input data after normalization, x_i is input data, x_{min} is minimum value of the data, and x_{max} is maximum value of the data.

Training data is used for modeling and testing data for model evaluation. The determination of parameters, such as epoch, batch size, neuron size in hidden layer, and learning rate may be random. Training in the cell is done until the learning process stops according to predetermined conditions or error target is reached.

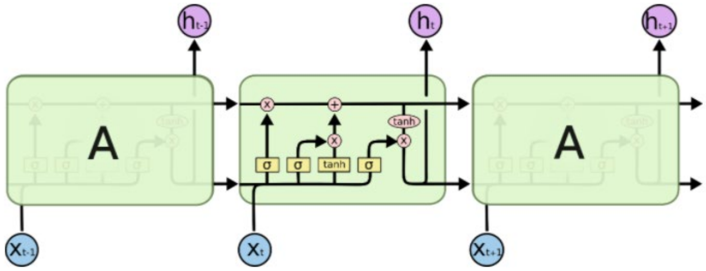

Figure 2. Long Short Term Memory's Architecture. Reproduced from Olah, 2015

Figure 2 shows the structure of LSTM. The first gate is a forget gate with sigmoid function that defines whether information is should be remembered or not.

$$f_t = \sigma(W_f.[h_{t-1} + x_t] + b_f) \quad (7)$$

where σ is sigmoid activation function, h_{t-1} is value of the previous hidden state, x_t is the current input, and b_f is the bias. The using of sigmoid affected the output in forget is a number between 0 and 1. If a value close to 0 it will be forgotten, whereas if it is close to 1 then there will continue to be remembered.
After the forget gate, there is an input gate (i_t) which is using sigmoid and tanh. This gate determines new information that will be store in the cell state.

$$i_t = \sigma(W_i.[h_{t-1} + x_t] + b_i) \quad (8)$$

Output from this gate will perform an elementwise multiplication operation with the output of the candidate state cell ($\tilde{C}_t$) which has the following equation.

$$\tilde{C}_t = tanh(W_c.[h_{t-1} + x_t] + b_c) \quad (9)$$

where tanh is an activation that produces number between -1 and 1. After that, a new cell state (c_t) is obtained by adding the output of forget gate and input gate.

$$c_t = (f_t * c_{t-1} + i_i * \tilde{C}_t) \quad (10)$$

The last gate is output gate (o_t) that determines how much information that is output by the cell.

$$o_t = \sigma(W_o.[h_{t-1} + x_t] + b_i) \quad (11)$$

The value of a new cell state is combined with the output gate by the multiplication operation that will become the hidden state output.

$$h_t = o_t * tanh(c_t) \quad (12)$$

Input gate, output gate, and forget gate has independent weight and bias (Olah, 2015).

Model Evaluation Criteria

The model is evaluated using data testing of Bandung City to find out how precisely the model can predict data it has never seen before. The evaluation of the model in this study used the Mean Absolute Percentage Error (MAPE) (Sirait et al.,2020).

$$MAPE = \frac{\sum_{t=1}^{n} \frac{|X_t - \hat{X}_t|}{X_t}}{n} \times 100\% \quad (13)$$

where X_t is an actual value periode t, meanwhile $\hat{X}_t$ is an predicted value periode t. The best model to predict active cases of COVID-19 in Bandung City is the model with the smallest MAPE.

Prediction

After getting the best model, then performed a prediction one period (a week) ahead. The normalization of data caused the output of prediction is between 0 and 1, so data denormalization is needed to get the real results of the prediction of active COVID-19 cases.

$$y_i = Y_i'(x_{max} - x_{min}) + x_{min} \quad (14)$$

where y_i is an output, Y_i' is output data after denormalization is done, x_{min} is a minimum value from the real active cases in each city, and x_{max} is a maximum value from the real active cases in each city.

Data Collection

The data used in this research is obtained from the Pikobar for each city in West Java, while the data for Bandung City is obtained from the Pusicov. Data was collected starting from the first COVID-19 case in Indonesia (March 02, 2020 - May 08, 2021) which was converted to every week and became Covid Weekly Data (CWD) which consists of 4 variables, namely confirm, recovered, death, and active case COVID-19 patients in 27 Cities in West Java.

$$Active\ Case = Confirm - Recovered - Death \quad (15)$$

Results and Discussion

The forecasting of active COVID-19 cases in Bandung City uses the Long Short Term Memory (LSTM) method, The data are selected using the K-Medoids Clustering to select Cities in West Java that has an active cases trend similar to Bandung City.

Numerical Results

The optimal number of groups is determined by the highest value of Pseudo F.

Table 1. Pseudo F

The number of Groups	Pseudo F
2	11,73
3	6,87
4	4,80
5	3,45
6	2,69
7	2,16

Based on Table 1. Two number of Groups has the highest value of Pseudo F. Therefore, the data will be clustered into two groups (k=2). K-Medoids clustering starts with initiating the center (medoid) based on the smallest vj value.

Table 2. The value of each v_j

v_j	value	v_j	value
7	0,480859	5	0,622838
17	0,48267	14	0,653741
11	0,48929	4	0,691556
10	0,511213	21	0,708204
18	0,534248	2	0,736291
25	0,55745	6	0,75741
26	0,559651	12	0,830959
27	0,561436	19	1,372979

9	0,571419	15	1,518049
20	0,581334	1	1,874478
3	0,581967	16	2,259299
8	0,599703	23	2,499785
22	0,603397	24	4,755726
13	0,604048		

Based on Table 2. The two smallest are $\boldsymbol{v}_7$ and $\boldsymbol{v}_{17}$ or column 7 and 17, so Ciamis Regency and West Bandung Regency will be the temporary center (medoid).

Table 3. Clustering City/Regency

Cluster	City/Regency
1	Bogor Regency, Bandung Regency, Garut Regency, Cirebon Regency, Indramayu Regency, Subang Regency, Karawang Regency, Bekasi Regency, **West Bandung Regency**, Bogor City, Bekasi City, Depok City
2	Sukabumi Regency, Cianjur Regency, Tasikmalaya Regency, **Ciamis Regency**, Kuningan Regency, Majalengka Regency, Sumedang Regency, Purwakarta Regency, Pangandaran Regency, Sukabumi City, Bandung City, Cirebon City, Cimahi City, Tasikmalaya City, Banjar City

After getting the temporary clusters on Table 3. Then iterate the result until the result is not change again. Based on data that used, The iteration is only done once because when the second iteration is done, the result is not changing. Therefore, the final Cluster is on Table 3.

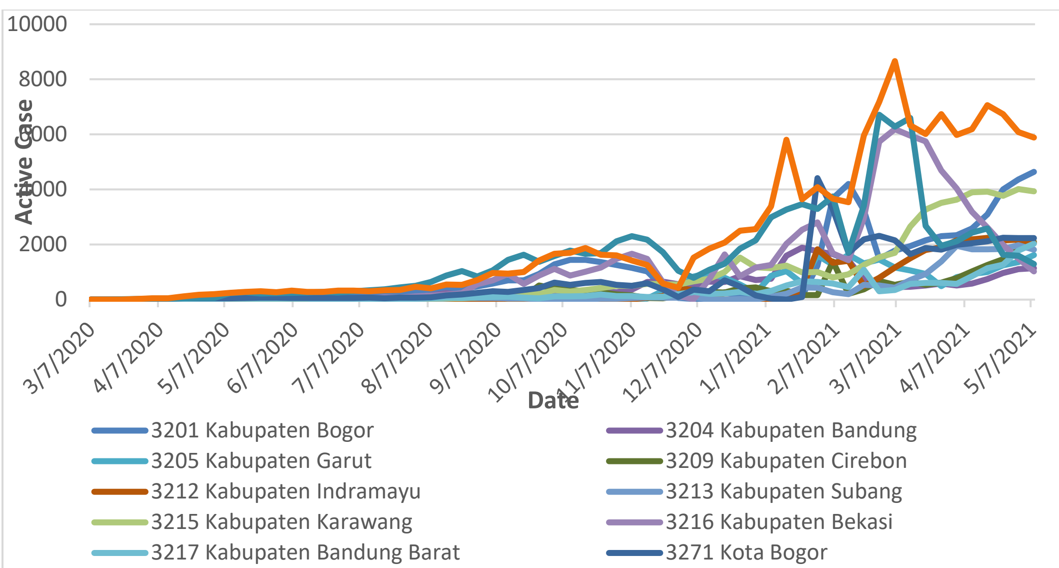

Figure 3. Graphic of Cluster 1

Figure 3 shows that Cluster 1 contains 12 City/Regency. The similarity can be seen around January 16, 2021 and March 06, 2021, at that time there is very high number of active case.

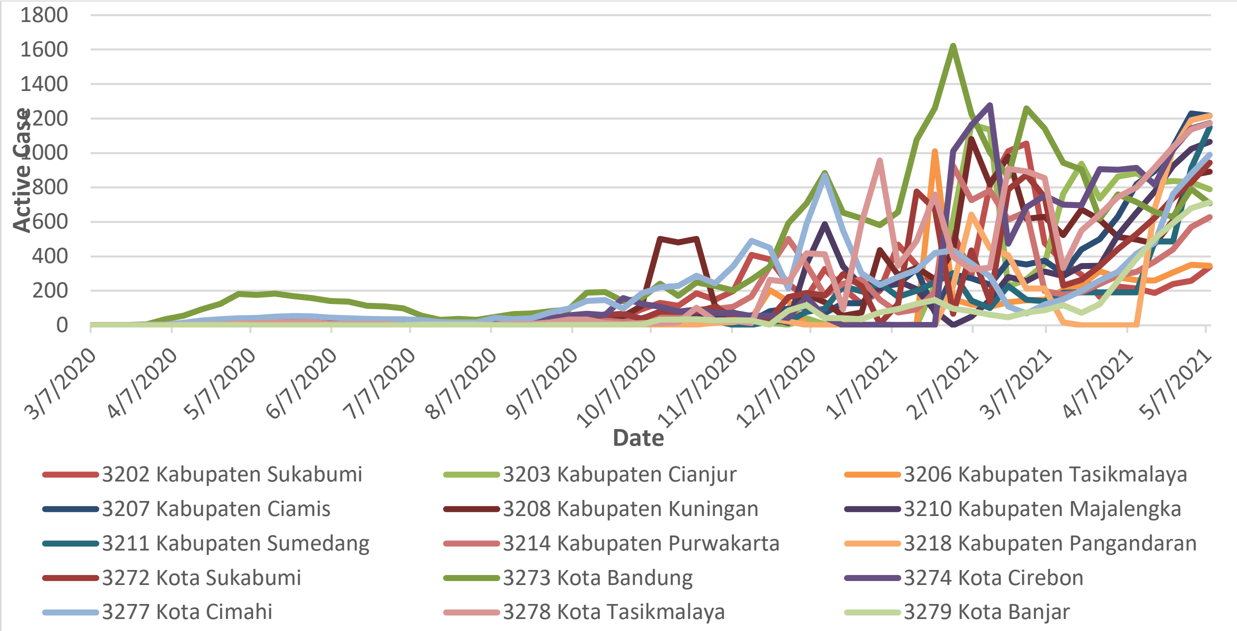

Figure 4. Graphic of Cluster 2

Figure 4. shows that Cluster 2 contains 15 City/Regency. The similarity can be seen around December 12, 2020 and February 27, 2021, at that time 15 City/Regency have the same active cases fluctuation. Those historical data in Figure 4. except Bandung City will be used as data training, meanwhile historical data of Bandung City will be divided into 75% training and 25% testing. Then, normalization is carried out to get value of active case in each city to be between 0 and 1.

This study uses the Python software in modeling and imported Numpy, Matplotlib, Pandas, Tensorflow, and Keras. Before modeling, the dimension of data should be reformed to be three dimensions [samples, timesteps, features] for predictor, and [samples, timesteps] for target. In this study, using 6 timesteps. For example, predictor of Bandung City has dimension [39, 6, 1] for training and [9, 4, 1] for testing. Meanwhile, target of Bandung City has dimension [39, 1] for training and [9, 1] for testing.
Several training options and parameters of the model were changed to select the best results. The parameters such as weights and bias were also updated depending on a loss function of an iteration step.

Table 4. LSTM Model Setup

Parameter	**Setting**
Learning rate	Adam (0.001)
Batch size	1
Hidden Layer	1
Neuron of Hidden Layer	3
Epoch	250

MAPE is used to evaluate the model in data testing. Using Formula (13), the value of MAPE is 31,46%.

Graphical Results

The loss function using parameter on Table 4 can be seen on Figure 5.

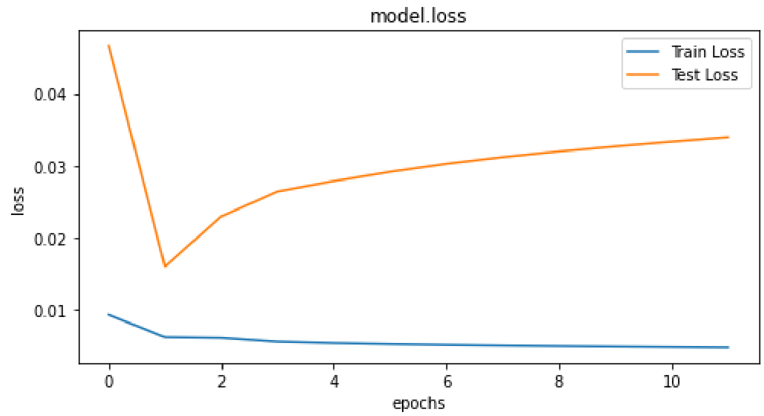

Figure 5. Lost Function of Best LSTM Model.

Figure 5.show a graph of the loss function using the best LSTM model that has decreased errors in the test data up to the 12th epoch. This is because using an earlystopping configuration with 10 patience, which means that the training process will stop if the error results from the test data is not decrease again after 10 epochs.

Validation

After determining the best parameter for LSTM. Furthermore, predictions will be made for the next week (Saturday) on May 15, 2021 using the best parameters.

Table 5. Prediction Results Using LSTM

Period			Active Cases	
Week	From	Until	Interval Prediction	Actual
29	27/09/2020	03/10/2020	139 - 189	158
30	04/10/2020	10/10/2020	217 - 295	241
31	11/10/2020	17/10/2020	136 - 186	171
32	18/10/2020	24/10/2020	191 - 259	248
33	25/10/2020	31/10/2020	225 - 305	228
34	01/11/2020	07/11/2020	161 - 219	202
35	08/11/2020	14/11/2020	218 - 296	266
36	15/11/2020	21/11/2020	277 - 377	339
37	22/11/2020	28/11/2020	487 - 659	592
38	29/11/2020	05/12/2020	565 - 765	707
39	06/12/2020	12/12/2020	714 - 966	884
40	13/12/2020	19/12/2020	609 - 825	654
41	20/12/2020	26/12/2020	580 - 786	620
42	27/12/2020	02/01/2021	468 - 634	582
43	03/01/2021	09/01/2021	498 - 674	657
44	10/01/2021	16/01/2021	833 - 1129	1077
45	17/01/2021	23/01/2021	1004 - 1360	1261
46	24/01/2021	30/01/2021	1212 - 1642	1622
47	31/01/2021	06/02/2021	1163 - 1575	1227

48	07/02/2021	13/02/2021	993 - 1345	1003
49	14/02/2021	20/02/2021	813 - 1101	861
50	21/02/2021	27/02/2021	919 - 1245	1260
51	28/02/2021	06/03/2021	1108 - 1500	1143
52	07/03/2021	13/03/2021	932 - 1262	943
53	14/03/2021	20/03/2021	812 - 1100	905
54	21/03/2021	27/03/2021	592 - 802	618
55	28/03/2021	03/04/2021	568 - 770	759
56	04/04/2021	10/04/2021	554 - 750	715
57	11/04/2021	17/04/2021	602 - 816	657
58	18/04/2021	24/04/2021	574 - 777	629
59	25/04/2021	01/05/2021	617 - 834	792
60	02/05/2021	08/05/2021	629 - 851	707
61	09/05/2021	15/05/2021	606 - 820	-

Table 5 shows the result of weekly prediction active cases in Bandung City using the best parameters in every week is never exceed 15% error. The prediction result for May 15, 2021 using the best parameter is 713 actives cases and if using 15% error, then the prediction for week 62th or Saturday, May 15, 2021 will be in the interval 606 - 820 active cases.

Conclusion

The Active cases in Bandung City after PSBB is increasing. Therefore, weekly forecasting the number of active COVID-19 cases in Bandung City can be used to evaluate whether the government can not make good decisions in order to decrease the number of new cases or the service of hospital is not good in dealing with COVID-19 patients so the number of recovered is not likely increase. In this study, to get better result prediction, using K-Medoid Custering obtained a cluster consisting of the data from Sukabumi Regency, Cianjur Regency, Tasikmalaya Regency, Ciamis Regency, Kuningan Regency, Majalengka Regency, Sumedang Regency, Purwakarta Regency, Pangandaran Regency, Sukabumi City, Bandung City, Cirebon City, Cimahi City, Tasikmalaya City, and Banjar City which are used as data training. After that, the trial and error was done for every parameter and obtained the optimum results with the lowest MAPE. One hidden layer, three neurons of hidden layer, adam optimization, and 250 epochs with the stochastic gradient descent algorithm can deliver MAPE of 31,46% in Bandung City. The prediction result for May 15, 2021 using the best parameter is 713 actives cases and if using 15% error, then the prediction will be in the interval 606 – 820 active cases. It is to be hoped that this prediction can be used by government and hospital to pay attention in facing COVID-19.

References

Ashrovy, R., Recurrent Neural Networ-Part One, Available: https://medium.com/@ashrovy/recurrentneuralnetwork-part-one822f1341fec. 2017.

Calinski T., Harabasz J., *A Dendrite Method for Cluster Analysis* [J]. Poland: Communications in tatistics,3(1):127, 1974.

Hair JF Jr., Anderson RE., Tatham RL. Black WC, *Multivariate Data Analysis 7th Edition*, New Jersey :Prentice Hall, 1996.

Hochreiter, S., Schmidhuber, J., Long short term memory. *Neural Computation, pp. 132,* 1997.

Huang, Z., *Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values*, Australia: Data Mining and Knowledge Discovery 2, 283–304, 1998.

Hyndmann, RJ., Athanasopoulos, G., *Forecasting: Principles and Practice, 2nd Edition*, OTexts, Australia, 2018.

Johnson, RA., Wichern, D., *Applied Multivariate Statistical Analysis Six Edition*, New Jersey: Pearson Prentice Hall, 2002.

K.K.R.Indonesia,Keputusan Menteri Kesehatan Republik IndonesiaNomorHK.01.07/MENKES/413/2020. Tentang Pedoman Pencegahan dan Pengendalian Coronavirus Disease 2019 (COVID-19), 2020.

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

Lawrence, R., Using Neural Networks to Forecast Stock Market Prices. *Department of Computer Science, University of Manitoba,Wannipeg, Canada,* 1997.

Liu, Y., Overfitting and forecasting: linear versus non-linear time series models, *Iowa State University Digital Repository, pp. 1-103,* 2020.

Meesrikamolkul, W., Niennattrakul, V., & Ratanamahatana, C. A., *Shape-based Clustering for time series data. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp.530-541)*, Canada: John Wiley an Sons Inc, 2017.

Olah, C., Understanding LSTM Networks, Available: colah.github.io/posts/2015-08-Understanding-LSTMs, 2015.

Pusicov and Pikobar, Pusat Informasi COVID-19 Kota Bandung and Pusat Informasi dan Koordinasi COVID-19 Jawa Barat, Available:https://covid19.bandung.go.id and https://pikobar.jabarprov.go.id/, 2021.

Puspitaningrum, D., *Pengantar Jaringan Saraf Tiruan*, ANDI, Yogyakarta, 2006.

Sirait, H., Sukono, Sundari, S., and Kalfin. (2020). Ratio Estimator of Population Mean using Quartile and Skewness Coefficient. *International Journal of Advanced Science and Technology*, 29(6), pp. 3289 – 3295

Suherman, Ginanjar, I., J. Purwandari, T., *Pengelompokan Kabupaten dan Kota di Jawa Barat Berdasarkan Kasus Aktif COVID-19 Menggunakan K-Medoids Clustering dengan Jarak Dynamic Time Warping*, Sumedang, 2021.

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

Mohammad Niks Iksan, is an undergraduate student. Currently, he is studying in the Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran.

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.