The Generalized STAR Modelling with Minimum Spanning Tree Approach of Weight Matrix for COVID-19 Case in Java Island

Utriweni Mukhairyar¹,*, Bayu Imadul Bilad², and Udjianna Sekteria Pasaribu¹

¹Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia,
²Student of Graduate Program in Mathematics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia

*utriweni@math.itb.ac.id

Abstract. The ongoing global Coronavirus 2019 (COVID-19) pandemic poses a major threat. The spread of the COVID-19 virus is likely to occur from one location to another location due to the mobility of people. Many efforts and policies have been made by each country to reduce the spread of the COVID-19 outbreak. The imposition of lockdown and large-scale social restrictions or social distancing has been widely applied to limit the transmission of this virus among the community and provincial levels. Both policies have proven effective in reducing the spread of the COVID-19 virus. To obtain the overview of this case, many researchers were conducted. Here, the Generalized STAR (GSTAR) model was applied to model the increasing number of COVID-19 positive cases per day in six provinces in Java Island. The data was recorded simultaneously in six locations, namely in the Provinces of Banten, Jakarta, West Java, Central Java, Yogyakarta Special Region, and East Java. This paper proposes a new approach in constructing the weight matrix required to build the GSTAR model, namely Minimum Spanning Tree (MST). The weight matrix represents the relationship among observed locations. By using the MST, a topological (undirected graph) network model could be created to show the correlation, centrality, and relationship on the increase of COVID-19 positive cases among the provinces in Java Island. The GSTAR(1;1) with the inverse distance weight matrix using MST presents a good ability to predict the COVID-19 increasing cases of Java island. This is indicated by the final MAPE average score of 19.55.

Keywords: Centrality, correlation, COVID-19, GSTAR, Minimum Spanning Tree, weight matrix

1. Introduction

The continuing global pandemic, Coronavirus 2019 (COVID-19), poses a significant danger to public health, health infrastructure, and economy throughout the world. To reduce the spread of the COVID-19 epidemic, each country considered various initiatives and strategies. To restrict COVID-19 transmission at the community level and safeguard susceptible populations, Lockdown and large-scale
social distancing or Social Distancing have been widely used. These two strategies have successfully contained the increase in positive cases of COVID-19. The World Health Organization's (WHO) China office received the first reports of a previously unknown virus causing pneumonia cases in Wuhan, a city in Eastern China, on December 31st, 2019 [1]. Then this case evolved into a worldwide pandemic which still happen until now. The COVID-19 is a novel coronavirus that causes infectious sickness [1]. When symptomatic patients speak, sneeze, or cough, the virus could transfer through respiratory droplets [2]. At close range, droplets can land on people's mouths, noses, and eyes [3]. Virus particles that are small enough to be classified as aerosols or fine particles can float in the air for hours and travel hundreds of feet with water currents. In addition, COVID-19 is more stable in plastic and stainless steel, as well as copper and cardboard, where it can survive for up to 72 hours [4].

In Indonesia, the first positive case of COVID-19 was officially announced by the government on March 2, 2020. It is known that there were two Indonesian citizens who made direct contact with Japanese citizens who visited Indonesia. For the first time, on March 11, 2020, it was discovered that a 59-year-old man from the city of Solo died as a result of this COVID-19 Virus. Until now, the spread of the COVID-19 in Indonesia has spread to 34 provinces. As of April 9, 2021, in Indonesia, today more than 1.5 million people have been confirmed to be infected with this virus. The total recovered and died from the COVID-19 were 1.4 million and 42 thousand people, respectively. The Java island has become the epicentre of the spread of COVID-19 in Indonesia. The confirmed cases of COVID-19 reached 1,004,436, which made the Java island became the largest contributor of COVID-19 cases. Many factors contribute to the rise of the positive cases, including (1) high mobility between provinces in Java island, (2) Jakarta is the capital city of Indonesia, and (3) many migrants come from other provinces outside Jakarta, allowing people to move from one province to another. The COVID-19 virus has a high chance of being transmitted during this movement. In addition, it is also known that Java, whose number of provinces is six, has several large cities such as Jakarta, Tangerang, Bandung, Yogyakarta, Solo, and Surabaya. The mobility inside those cities is very high, resulting the spread of the COVID-19 virus to be accelerated. Thus the research regarding covid cases in Indonesia be rapidly developed in many sectors, such as an early epidemiological assessment of covid transmission [5], the COVID-19 cases prediction in many provinces simultaneously [6], and the economic impact [7].

The Generalized Space-Time Autoregressive or GSTAR (1; 1) model can be used to examine the distribution of COVID-19 cases across provinces in Java. The model implies that the past occurrences at a site are influenced not just by previous events at that place, but also by previous events at nearby locations [8]. The weight matrix expressing the spatial dependence of an event in a location with its surroundings is one of the model's unique features. In order to analyse the transmission of the COVID-19 in Java island, a modified Inverse Distance Weight, abbreviated as IDW, is proposed. The development of the weight matrix on the GSTAR model has been carried out by many previous researchers, such as using the kernel approach [9], involving the geographical dependence of train distances between observed provinces [6], and minimum spanning tree approach [10] Here another development has been carried out to construct the inverse distance weight matrix on the modified GSTAR model using the Minimum Spanning Tree (MST) weight matrix approach through its forest graph. MST weight matrix is less subjective compared to others since it is constructed based on the observations. It is possible to obtain a random weight matrix, whose realizations are possible to be different by the times. It depends on how the relationship among observations of locations is. Meanwhile, common weight matrix, such as uniform, binary, and Euclidean distance are based on real distance, which is fixed all the time. Daily positive cases of COVID-19 in six provinces of Java Island, namely Banten, Jakarta, West Java, Central Java, DI Yogyakarta (DIY), and East Java, were utilized as the objects. The goal, therefore, is to use the Spatio-temporal GSTAR (1; 1) model to anticipate COVID-19 cases in Java.

The MST model has become an indispensable tool in the field of econophysics to filter important information contained in the complex structure of the correlation matrix between stocks in a particular portfolio. MST research has been widely used by researchers, as in [11] regarding network analysis by conducting robust filters and in [12, 13] regarding network topology analysis of LQ45. LQ45 is a stock
market index for the Indonesia Stock Exchange (IDX) (formerly known as the Jakarta Stock Exchange). In this study, using MST, a topological network model (not directed) shows the correlation and centrality between provinces in Java related to the increase in positive cases of COVID-19.

2. GSTAR with Modified Inverse Distance-Spatial Weight of MST Approach

The GSTAR model is an extension of the STAR\((p_r\lambda_1, ..., \lambda_p)\) model [14], which assumes that the model parameters change for each location, so that the GSTAR model tends to be more flexible than the STAR model. With these different parameters, the GSTAR model indicates the heterogeneous of locations. The simplest model of GSTAR\((p; \lambda_1, ..., \lambda_p)\) is GSTAR(1;1), which has formula as,

\[ Y_{i,t} = \phi_{0i}Y_{i,t-1} + \phi_{1i} \sum_{j=1}^{N} w_{ij} Y_{j,t-1} + e_{i,t} \]

where \(Y_{i,t}\) is stationary observation at time \(t\) and in location \(i\), \(\phi_{0i}\) and \(\phi_{1i}\) are the element of diagonal matrix of autoregressive parameters, consecutively for the first lag of time and zero lag of spatial order and the first lag of time and first lag of spatial order, \(e_{i,t}\) is a noise process at time \(t\), and \(w_{ij}\) is the spatial weight matrix for location \(j\) to \(i\) and be the element of the weight matrix. This matrix is the beauty of the GSTAR model since it represents the spatial dependency between locations.

Many researchers have developed the GSTAR model both in theory and application. Mukhaiyar and Pasaribu [15] developed a procedure for the GSTAR model using the IAcM (Inverse Autocovariance Matrix) method. Nugraha [16] used the weighted average of the fuzzy method to create a weight matrix for the GSTAR model meanwhile Yundari [17] also conducted research on the assumption of the GSTAR model error which be continued to the Monte Carlo simulation of error assumptions. GSTAR modellling was developed by involving outlier factors and exogenous variable [18] and used kernel weight matrix [19]. Recently, the discrete case of GSTAR modelling was developed [15]. This model had been applied to chili prices in the Bandung market [20], tea plantations production [21], crime [22], dengue fever [23], and COVID-19 case [6].

The focus in this paper is to introduce a new approach in constructing the weight matrix. Generally, researchers use uniform weights, binary, or non-uniform weights based on distance. This weight selection process is still subjective. The proposed weight matrix also assumes the closer the distance, the stronger the relationship between the provinces. This matrix is a combination of the ordinary IDW matrix with the MST model. The MST model provides an undirected graph in which nodes connecting one location to another location express the reciprocal influence of one location with another. In addition, MST has been used to determine the number of neighbours that affect each location, so that the correlation and centrality between provinces can be seen regarding the increase in positive cases of COVID-19 in Java island.

MST is one of the network topology analysis models. The network topology itself consists of nodes and a line which is the correlation relationship of each variable that exists and is formed in a diagram. To filter all the information on the network, MST is often used using Prim and Kruskal algorithms. But the two algorithms are unable to detect the uniqueness of the MST, where with MST, it is still possible for different network topologies to occur. Otherwise, if the entire MST proceeds, only one network topology will be formed and could produce more accurate information than only one MST. Therefore, in this study, an inter-provincial network topology analysis was carried out regarding the increase in positive cases of COVID-19 on the island of Java using the Forest of all MST model proposed by Djauhari [11, 24, 25]. From this model, the correlation between provinces and which areas need attention were identified in order to reduce the number of COVID-19 positive cases.
The first step is to compute the correlation values among variables or locations. To calculate the correlation value $c(i,j)$, it is defined by \[4, 11\],
\[
c(i,j) = \frac{(\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle)}{\sqrt{(\langle x_i^2 \rangle - \langle x_i \rangle^2)(\langle x_j^2 \rangle - \langle x_j \rangle^2)}}
\] (2)
where $\langle x_i \rangle$ is the average of $x_i(t)$ for all $t$. The relationship between these variables will be formed into a correlation matrix. Formation of correlation matrix $C = c(i,j)$ size $(n \times n)$, where $c(i,j)$ is the correlation coefficient between the $i$-th and $j$-th variables.

After obtaining the correlation values, a distance matrix $D$ is made and from it the neighbouring (adjacency) matrix will be searched to compose a forest of all minimum spanning trees. The distance value is defined by \[11\]:
\[
D = d(i,j) = \sqrt{\sum_k (x_{ik} - x_{jk})^2}
\] (3)
for all $i, j = 1, 2, ..., n$. Then the network between locations is analyzed through the network topology created with MST $D$. If matrix $D$ contains a unique MST, the filtered information provided by the MST will be strong in the sense that no other filtered information source can be used. However, in practice, matrix $D$ often contains more than one MST. Thus, the development of using the forest of all MSTs be an alternative powerful filter. For this reason, an algorithm was built to build the forest using a fuzzy relation approach. This approach supports us to look further into the properties of matrix D. The adjacency matrix is constructed using the distance matrix, $D$. The results of the neighbouring matrix are 1 and 0. The value is 1 if a location has a neighbouring relationship with other locations and is 0 if it is not neighbours.

Here, the weight values are defined as unchanged spatial weight matrix. The MST be used to obtain the spatial configuration which is used to modify the inverse distance weight matrix. Originally, the inverse distance weight for each location can be written as \[6\]:
\[
w_{ij}^{(0)} = \frac{1}{d_{ij}}
\] (4)
where $i \neq j$, the total weight for each location is 1 and for all locations found $N$, $d_{ij}$ is the distance between provinces ($i$ and $j$) based on the distance between the governor's offices.

3. Model Evaluation

The best model criteria for out-sample data are seen based on the Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) values. There are four classifications of MAPE values to see how well the model used predicts the data \[26\]. Those are “very good” if MAPE is less than 10%, “well” if MAPE between 10% and 20%, “feasible” if MAPE between 20% and 50%, and “poor” if MAPE is more than 50%. Mathematically, the MAPE value can be expressed by the following formula \[26\]:
\[
MAPE = \left| \frac{\sum_{t=1}^{T} \left( \frac{Z_t - \hat{Z}_t}{Z_t} \right) \times 100\%}{T} \right|
\] (5)
where \( t \) is the number of data to be predicted, \( \hat{Z}_t \) is the result of the \( t \)-th prediction and \( Z_t \) is the actual data of the \( t \)-th time.

Then RMSE can be expressed by the following formula [26]:

\[
RMSE = \sqrt{\frac{1}{n} \sum (Y_t - \hat{Y}_t)^2}
\]  

where \( Y_t \) is the original response data, \( \hat{Y}_t \) is the predicted result, and \( n \) is the amount of data. The selected RMSE value is the smallest value. Figure 1 shows the procedure of GSTAR modeling based on modified IDW using the MST weight matrix.

![Flowchart of computing the modified IDW matrix with MST](image)

**Figure 1.** Flowchart of computing the modified IDW matrix with MST
4. Data Analysis and Discussion

4.1. Descriptive Statistics

The data used is a secondary data, which is about the positive cases of COVID-19 per day in six provinces on the Java island. The locations used in this study are Banten, Jakarta abbreviated as DKI, West Java abbreviated as JABAR, Central Java abbreviated as JATENG, Yogyakarta Special Region abbreviated as DIY, and East Java abbreviated as JATIM. The distances between the six locations were obtained from the results of calculations using the Google Maps application which is shown in Table 1. The furthest distance between locations is 878 km, which is between Banten Province and East Java Province. While the closest distance is given by Banten Province and Jakarta Province, which is as far as 95 km.

| Table 1. Distance between Provinces in Java Island (Km) |
|--------------------------------------------------------|
| Banten     | DKI | JABAR | JATENG | DIY | JATIM |
| Banten     | 0   | 95    | 234    | 528 | 640   | 878   |
| DKI        | 94  | 0     | 165    | 475 | 551   | 785   |
| JABAR      | 234 | 162   | 0      | 361 | 394   | 768   |
| JATENG     | 551 | 457   | 361    | 0   | 120   | 313   |
| DIY        | 619 | 531   | 394    | 115 | 0     | 332   |
| JATIM      | 878 | 768   | 673    | 317 | 327   | 0     |

The data from the COVID-19 Task Force (SATGAS) in Indonesia, namely covid19.go.id. The sample size used is 203 data points from July 1, 2020, to January 18, 2021. There are 198 time of observations which were treated as in-sample data and the remainder were for out-sample data and be used to validate the prediction results.

The descriptive statistics from data are shown in Table 2. From the table, it was found that the daily average of positive cases increasing in the six locations have significant differences from each other. The Jakarta Province occupies the first position with the highest average of active positive case, that is 1033 positive cases per day and DIY has the lowest average of 78 positive cases per day. Meanwhile, the maximum and minimum values of the increase in positive cases per day in the six locations above are 3476 cases and 0 cases, respectively. This may be due to differences in demographics, lifestyle, and community compliance in each province.

Table 2. Descriptive Statistics of Data on the Increase in positive cases of COVID-19 per day in Six Provinces on the Island of Java

|               | DKI   | JATIM | JABAR | BANTEN | JATENG | DIY  |
|---------------|-------|-------|-------|--------|--------|------|
| MIN           | 137   | 168   | 24    | 0      | 36     | 1    |
| 1st Q         | 608   | 288   | 151   | 32     | 198    | 19   |
| Median        | 1022  | 344   | 395   | 100    | 339    | 41   |
| Mean          | 1033  | 422   | 502   | 99     | 465    | 78   |
| 3rd Q         | 1258  | 480   | 731   | 145    | 706    | 103  |
| MAX           | 3476  | 1025  | 1824  | 364    | 2036   | 379  |

Figure 2 is a data plot of the increase in positive cases of COVID-19 per day in six provinces on the island of Java from July 1, 2020 to January 12, 2021. It can be seen that from the six locations there is an upward trend over time. Jakarta Province had a significant increasing of positive cases during the last month of observations. The average cases per day in Jakarta was below 500 people per day in 2020 and then changed to more than 2000 people per day in early January, 2021. This was followed by...
other provinces which also experienced a case increasing. Thus, it is assumed that the increasing of COVID-19 positive cases in one province will affect its surrounding provinces.

Figure 2. Data Plot of Increase in positive cases of COVID-19 per day in Six Provinces in Java Island from 01 July 2020 to 13 January 2021

4.2. Correlation between Locations

| Banten | DKI   | JABAR | JATENG | DIY   | JATIM |
|--------|-------|-------|--------|-------|-------|
| Banten | 1     |       |        |       |       |
| DKI    | 0.769897 | 1     |        |       |       |
| JABAR  | 0.743229 | 0.796423 | 1     |       |       |
| JATENG | 0.699558 | 0.730484 | 0.723218 | 1     |       |
| DIY    | 0.71556 | 0.826296 | 0.845468 | 0.753499 | 1     |
| JATIM  | 0.583519 | 0.711724 | 0.733982 | 0.600379 | 0.873071 | 1     |

The correlation of the increasing of COVID-19 positive cases per day in six locations be presented in table 3. It shows that the average of correlation values are above 0.5. It means that the COVID-19 positive cases among provinces have strong linear relationship. In addition, the highest correlation value is owned by the Province of East Java and the Special Region of Yogyakarta with correlation value is 0.87. As for the geographical location, the locations of both provinces are close. Then followed by West Java Province and Yogyakarta Special Region which have correlation values of 0.84. On the other hand, the lowest correlation value is owned by Banten Province and East Java Province, which are the provinces with the farthest distance, with a correlation value of 0.58.

4.3. Minimum Spanning Tree Modelling

After calculating the correlation matrix among provinces, the distance matrix between provinces in Java Island was obtained using the distance matrix equation (3) with the results as below.

\[
D = \begin{bmatrix}
0  & 0.37 & 0.42 & 0.43 & 0.51 & 0.62 \\
0.37 & 0 & 0.29 & 0.42 & 0.30 & 0.47 \\
0.42 & 0.29 & 0 & 0.43 & 0.26 & 0.44 \\
0.43 & 0.42 & 0.43 & 0 & 0.47 & 0.59 \\
0.51 & 0.30 & 0.26 & 0.47 & 0 & 0.31 \\
0.62 & 0.47 & 0.44 & 0.59 & 0.31 & 0
\end{bmatrix}
\]
Then after getting the distance matrix D, calculations are carried out to obtain the adjacency matrix by looking for the subdominant-ultra (SDU) matrix first. The SDU matrix is obtained by applying the fuzzy relation approach, min-max transitive. Based on the flowchart in figure 1, then the adjacency matrix is obtained as follows,

\[
\Delta = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 \\
\end{bmatrix}
\]

If the resulting adjacency matrix is plotted into a graph, then the forest of all MST from each province is as figure 3:

Figure 3. Forest of All possible MST

Figure 3 shows all possible MST obtained from six provinces. It shows the high correlation among the six provinces. The smaller the distance between provinces is, the closer the relationship between the COVID-19 positive case in those provinces. For example, DKI relates to Banten, JABAR, JATENG, and DIY. It can be concluded that the increasing of COVID-19 positive cases in DKI has a close relationship with the increasing case in those four provinces.

The last step, using the Kruskal algorithm to obtain a Minimum Spanning Tree with the minimum number of weights, compared to other MSTs formed as shown in figure 4.

Figure 4. MST using the Kruskal Algorithm.

4.4. Formation of a modified Spatial Weight Matrix

The weight matrix used in this study is based on the results of the MST modelling in Section 4.3. The MST model produces an undirected topological network where the edges show a correlation relationship among provinces. The more a location is connected to other locations, the more influential location neighbours are. The weight value used in this modelling was derived from the inverse distance value in general. Before building a spatial weight matrix, a distance matrix is built based on the MST, that is,
Then by using the Equation (4), the modified inverse distance matrix using MST is obtained as,

\[ W_{MST}^{(1)} = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 \\
0.34 & 0 & 0.27 & 0.39 & 0 & 0 \\
0.52 & 0 & 0 & 0.48 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.46 & 0 & 0 & 0.54 \\
0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix} \]

4.5. GSTAR (1;1) Model Establishment

The GSTAR order used is GSTAR(1;1), which consider the involvement of most influential neighbours. The estimation of the GSTAR(1;1) model parameters using the Ordinary Least Square (OLS) method with the help of GSTAR package on R software, thus the GSTAR(1;1) model parameters are obtained at each location for each model.

4.5.1 GSTAR(1;1) with Minimum Spanning Tree. The estimated GSTAR(1;1) model with MST for each location is as follows:

\[
\hat{Y}_1,t = -0.29Y_{1,t-1} + 0.05Y_{2,t-1} \\
\hat{Y}_2,t = -0.33Y_{2,t-1} + 0.02Y_{1,t-1} + 0.01Y_{3,t-1} + 0.003Y_{4,t-1} \\
\hat{Y}_3,t = -0.38Y_{3,t-1} + 0.05Y_{2,t-1} + 0.02Y_{5,t-1} \\
\hat{Y}_4,t = -0.29Y_{4,t-1} - 0.62Y_{2,t-1} \\
\hat{Y}_5,t = -0.36Y_{5,t-1} + 0.01Y_{3,t-1} + 0.02Y_{6,t-1} \\
\hat{Y}_6,t = -0.29Y_{6,t-1} + 0.4Y_{5,t-1}
\]

with 1: Banten Province 2: Jakarta Province, 3: West Java Province, 4: Central Java Province, 5: Yogyakarta Special Region Province and 6: East Java Province.

From the results of the GSTAR(1;1) model with the modified inverse distance of weight matrix using MST, it could interpret that the pattern of COVID-19 positive cases in Banten province is more influenced by the pattern in Banten province itself one day earlier. The same pattern occurred in the provinces of Jakarta, West Java and Yogyakarta. Furthermore, the increasing of COVID-19 cases in Central Java province was more influenced by the cases in Jakarta province one day earlier. Finally, the cases in East Java province are obviously influenced by the pattern of the COVID-19 positive cases in Yogyakarta Special Region.

Figure 5 shows the plot of the actual and the GSTAR (1;1) estimated observations of COVID-19 cases per day in six provinces of Java island. The RMSE of the model can be calculated as shown in table 4. It shows a small value for each location, except in West Java Province, which is 86.31.

| Table 4. Root Mean Square Error (RMSE) of GSTAR (1;1) model with Inverse Distance Weight Matrix modified by MST |
|---------------------------------------------------------------|
| Model            | Banten | Jakarta | West Java | Central Java | Yogyakarta | East Java |
|------------------|--------|---------|-----------|--------------|------------|-----------|
| GSTAR(1;1) with MST | 14     | 63      | 86        | 58           | 12         | 25        |
4.5.2 Model Diagnostic Test. Statistically, residual checking was carried out by using the Ljung-Box test to detect the correlation of residuals and the Kolmogorov-Smirnov (K-S) test to examine the residuals normality assumptions. Based on the results of the Ljung-Box test, it was found that only East Java Province satisfied the white noise assumption while the remaining five unsatisfying due to the $p$-value was less than 0.05. For the Kolmogorov-Smirnov test results, it is obtained that only Banten and Jakarta provinces satisfy the assumption of normality, meanwhile the rest of provinces are unsatisfied, such as West Java, Central Java, and the Special Region of Yogyakarta.

Table 5. Ljung-Box Test for Residual of GSTAR(1;1) with IDW Matrix modified by MST

| Model | Location  | $p$-value of Ljung-Box Test |
|-------|-----------|-----------------------------|
| MST   | Banten    | 0.0004                      |
|       | DKI       | 0.0034                      |
|       | JABAR     | $3e^{-09}$                  |
|       | JATENG    | 0.0001                      |
|       | DIY       | $4e^{-14}$                  |
|       | JATIM     | 0.0643*                     |

*null hypotheses (uncorrelated residuals) is not rejected.
4.6. Prediction

The predictions were made for the next five days. Figure 6 shows the comparison of the actual and the predicted observations using GSTAR(1;1) model with the MST-modified IDW matrix. Table 7 shows the MAPE value for each province in Java island and the average for all provinces. The smallest MAPE is obtained by Special Region of Yogyakarta, which is 8.16. This shows that the GSTAR(1;1) model with the MST-modified IDW matrix has a very good ability to predict the COVID-19 increasing cases per day in this province. Meanwhile, the largest MAPE value is in the West Java province, which is 29.49. It is possible caused by the existence of many factors which influence the increasing of COVID-19 cases in West Java province.

As a whole, the average MAPE using this model is 19.55. Based on the classification of MAPE in Section 3, it can be concluded that this model has a well ability to represent and predict the increase in COVID-19 cases in Java.
Table 7. Mean Absolute Error (MAPE) Prediction Results

| Model | BANTEN | DKI   | JABAR  | JATENG | DIY   | JATIM   | Mean   |
|-------|--------|-------|--------|--------|-------|---------|--------|
| MST   | 21.40  | 15.99 | 29.49  | 24.14  | 8.16  | 18.15   | 19.555 |

5. Conclusion
The parameter of GSTAR(1;1) model with the MST-modified of IDW matrix, be estimated using the Least Square (LS) method. The development of this weight matrix unchanged the linearization structure of the LS method. Therefore, the LS method is still appropriate to use in estimating the parameters. A case study conducted to the COVID-19 positive cases in six provinces of the Java island showed that the GSTAR(1 ;1) with the MST-modified of IDW matrix, has a good ability to predict the increasing of COVID-19 positive cases in Java island. It is indicated by the final MAPE average score of 19.55. There is still space for improving this quality of model in predicting. The deeper analysis of this approach should be explored more. It is suspected, this new approach will be excellent for prediction for some appropriate data. It can be researched more and be compared with other approaches to show its goodness.

Acknowledgement
The authors would like to thank the RISTEK/BRIN grant, "Penelitian Dasar Unggulan Perguruan Tinggi (PDUPT)" 2020-2021 for supporting funds. Our gratitude for the reviewers of this manuscript.

References
[1] World Health Organization 2020 Coronavirus Disease (COVID-19) Pandemic https://www.who.int/ Accessed 21 May 2020
[2] Ningthoujam R 2020 Covid-19 can spread through breathing, talking, study estimates. Curr. Med. Res. Pract. 10(3) 132–3
[3] Centers for Disease Control and Prevention 2020 Infection Control Guidance https://cdc.gov/ Accessed 8 June 2020
[4] Doramelen N, Morris D H, Holbrook M G 2020 Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1 N Engl. J. Med. 382 1564–67
[5] Hasan A, Susanto H, Kasim M F, Nuraini N, Lestari B, Triany D, and Widyastuti W 2020 Superspreading in early transmissions of Covid-19 in Indonesia Scientific reports 10(1) 1-4
[6] Pasaribu U S, Mukhaiyar U, Huda N M, Sari K N and Indratno S W 2021 Modelling COVID-19 growth cases of provinces in Java island by modified spatial weight matrix GSTAR through railroad passenger’s mobility Heliyon 7(2) e06025
[7] Mukhaiyar U, Widyangi D and Vantika S 2021a The Time Series Regression Analysis in Evaluating the Economic Impact of Covid-19 Cases in Indonesia Journal of Model Assisted Statistics and Applications 16(3)
[8] Mukhaiyar U, Pasaribu U S 2012 A New Procedure for Generalized STAR Modeling using IAcM Approach ITB J. Sci. 44A(2) 179–92
[9] Yundari, Huda N M, Mukhaiyar U, Pasaribu U S and Sari K N 2020 Stationary Process in GSTAR(1;1) through Kernel Function Approach AIP Conference Proceedings 2268 020010
[10] Mukhaiyar U, Mahdiyasa A W, Sari K N and Noviana N T 2021b A new weight matrix approach of Generalized Space-Time Autoregressive in modeling the number of vehicles entering the Bandung city through the Purbaleunyi toll Heliyon. submitted
[11] Djauhari M A 2012 A robust filter in stock networks analysis Physica A: Statistical Mechanics and its Applications Elsevier
[12] Putra T, Umbara R F and Indwiarti 2018 Analisis Topologi Jaringan Saham-Saham LQ45 Dengan Menggunakan Minimum Spanning Tree (MST) e-Proceeding of Engineering 5(3) Univeristas Telkom.

[13] Utami R, Tegar, Umbara R F and Indwiarti 2018 Analisis Topologi Jaringan Saham-saham LQ45 dengan Menggunakan Forest of All Minimum Spanning Tress e-Proceeding of Engineering 5(3) Universitas Telkom.

[14] Pfeifer P E and Deutsch S J 1980 A Three-Stage Iterative Approach for Space-time Modelling, Technometrics 22 397-408.

[15] Huda N M, Mukhaiyar U and Pasaribu U S 2021 The approximation of GSTAR model for discrete cases through INAR model Journal of Physics: Conference Series 1722(1) 012100

[16] Nugraha RF, Setyowati S, Mukhaiyar U 2015 Prediction of Oil Palm Production using The Weighted Average of Fuzzy Sets Concept Approach. AIP Conf. Proc. 1692 (020008)

[17] Yundari, Pasaribu U S, Mukhaiyar U 2017 Error Assumptions on Generalized STAR Model Journal of Mathematical and Fundamental Sciences 49(2) 136–55

[18] Mukhaiyar U, Huda N M, Sari K N and Pasaribu U S 2020 Analysis of Generalized Space Time Autoregressive with Exogenous Variable (GSTARX) Model with Outlier Factor Journal of Physics: Conference Series 1496(1) 012004

[19] Yundari, Pasaribu U S, Mukhaiyar U, Heriawan M N 2018 Spatial Weight Determination of GSTAR(1;1) Model by Using Kernel Function Journal of Physics: Conference Series 1028 (012223) 1-8

[20] Fadlilah N F I, Mukhaiyar U and Fahmi F 2015 The Generalized STAR(1;1) Modeling with Time Correlated Errors to Red-Chili Weekly Prices of some Traditional Markets in Bandung West Java AIP Conf. Proc. 1692(020014)

[21] Mukhaiyar U 2015 The Goodness of generalized STAR in spatial dependency observations modeling AIP Conf. Proc. 1692 (020012)

[22] Masteriana D, Riani M I and Mukhaiyar U 2019 Generalized STAR (1;1) Model with Outlier–Case Study of Begal in Medan, North Sumatera Journal of Physics: Conference Series 12456(1) 012046

[23] Mukhaiyar U, Huda N M, Sari K N and Pasaribu U S 2019 Modeling Dengue Fever Cases by Using GSTAR(1;1) Model with Outlier Factor Journal of Physics: Conference Series 1366(1) 012122.

[24] Sharif S and Djauhari M A 2012 A Proposed Centrality Measure: The Case of Stocks Traded at Bursa Malaysia Modern Applied Science 6(10) 62

[25] Lee G S and Djauhari M A 2012 Network topology of Indonesian stock market In Cloud Computing and Social Networking (ICCCSN) International Conference IEEE 1-4

[26] Chang P C 2007 The evelopment of a weighted evolving fuzzy neural network for PCB sales forcasteng Expert Systems with Applications 32 86–96