BAREKENG: Journal of Mathematics and Its Applications
March 2023 Volume 17 Issue 1 Page 0263–0274
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss1pp0263-0274

MODIFIED WEIGHT MATRIX USING PRIM'S ALGORITHM IN MINIMUM SPANNING TREE (MST) APROACH FOR GSTAR(1;1) MODEL

Nur'ainul Miftahul Huda[1*], Fransiskus Fran[2], Yundari[3], Lisa Fikadila[4], Fauziah Safitri[5]

[1,2,3,4,5]Mathematics Department, Universitas Tanjungpura
Jl. Prof. Dr. H. Hadari Nawawi, Pontianak, 78124, Indonesia

Corresponding author's e-mail: * ***nur'ainul@fmipa.untan.ac.id***

ABSTRACT

Article History:
Received: 30th September 2022
Revised: 20th December 2022
Accepted: 29th January 2023

Keywords:
Modified weight;
Prim;
Spatial.

The Generalized Space-Time Autoregressive (GSTAR) model is able to utilize modeling of both space and time simultaneously. The existence of a weight matrix is one of the aspects that established this model. The matrix illustrates the spatial impact that occurs between locations. In this research, a modified weight matrix is presented using the Minimum Spanning Tree approach of graph theory. Prim's algorithm is utilized for calculation here. Not only does the modified weight matrix depend on distance, but it also highlights the correlation. It makes the modified weight matrix unique. Before starting Prim's algorithm, the correlation is utilized as an input in forming the initial graph. Following that, find the graph with the least MST weight. Afterward, the graph is described utilizing a weight matrix, which is applied to the normalization process. Following this, the GSTAR(1;1) modeling process begins with estimating the parameters and then forecasting. The case study is Covid-19 cases on Java Island between July 2020 (when early Covid-19 entered Indonesia) and the beginning of January 2021. The research aims to model the Covid-19 cases using modified weights and to predict the following five times. The outcome is a GSTAR(1;1) model with modified weights that can capture both temporal and spatial patterns. The accuracy of the model is achieved for both the training data and the testing data by the MAPE computations, which yielded of 11.40% and 21.57%, respectively. Predictions are also obtained for each province in the following five times.

How to cite this article:

N. M. Huda, F. Fran, Yundari, L. Fikadila and F. Safitri., "MODIFIED WEIGHT MATRIX USING PRIM'S ALGORITHM IN MINIMUM SPANNING TREE (MST) APROACH FOR GSTAR(1;1) MODEL," *BAREKENG: J. Math. & App.*, vol. 17, iss. 1, pp. 0263-0274, March 2023.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article • Open Access

1. INTRODUCTION

In everyday life, it's hardly rare to come into situations that are correlated to events that happened in the past. Many individuals have concerns concerning the probability that these events will occur in the future or the historical relevance of their events in the past. This issue is on the basis of the recent proliferation of research on time series analysis. The growth of time series analysis has brought to light the possibility that certain data derived from an event not only has a link with data derived from prior events but also has a relationship with data derived from locations that are immediately next to the event. This is also one of the fundamental reasons behind the development of studies on space-time analysis. The Generalized Space Time Autoregressive (GSTAR) model is one example of a type of model that is applicable for use in spaces that display a variety of characteristics. The STAR model, which was initially presented by Pfeifer and Deutsch, was further developed into what is now known as the GSTAR model **[1]**. In the space- time analysis, the GSTAR model is one of the models that are utilized the most frequently, specifically to model and predict space-time data. Within the context of a multivariate time series, this model integrates the concepts of space and time. The space-time concept has as its defining feature the dependence on both the location and the time itself. Within the scope of the GSTAR(1;1) model, the location dependence is captured by a spatial weight matrix **[2]**. According to the findings of research conducted on the GSTAR model, it is assumed that the first-order spatial weight matrix, $\boldsymbol{W}$, was already determined before the modeling process started. There is no one method that is universally accepted for determining this weight matrix. In most cases, a weighting system that is uniform and is based on the number of locations that are the closest is utilized, in addition to a weighting system that is based on the distance between locations **[3]**. The works of Lai **[4]**, Nurhayati **[5]**, and Mukhaiyar **[6]** provide a more in-depth look at studies on the determination of nearest neighbors and distance weighting schemes.

In the research, a method based on graph theory is utilized to compute the weight matrix. The Minimum Spanning Tree (MST) with Prim's Algorithm is the approach that is taken from the field of graph theory. Mukhaiyar's work, which involved the modification of weight matrices with MST and the use of the Kruskal algorithm, was also part of the research that was carried out **[7]**. The path taken by the graph is determined by considering both the distance between locations and the correlation between them. At the initiation stage, it is presumed that all of the locations are connected together. The correlation between various locations is the factor that is used to determine the initial graph that is used as an input in Prim's algorithm. It is broadly accepted that two locations are connected if their correlation value is greater than 75%. After obtaining the initial graph, the process of the Prim algorithm is applied in order to acquire an MST graph with the smallest amount of weight.

In this particular research project, the case study that was looked at is the daily addition of positive cases of Covid-19. The provinces of Java, including Banten, DKI Jakarta, West Java, Central Java, and East Java, as well as DI Yogyakarta, are the locations that serve as the objects in this study. The period considered is from the middle of the year 2020 (when Covid-19 first passed in Indonesia) through the beginning of the year 2021. Due to the spread of Covid-19, which not only has an effect between places but also has an effect over time, the Covid-19 example was selected as a case study for the GSTAR(1;1) modeling with a modified weight. This was done because Covid-19 has both a spatial effect between locations and an effect over time. Because of this, once one province on the island of Java experiences an increase in the number of cases of Covid-19, it is not impossible that this will also have an impact on the provinces that surround it. This is because the provinces on the island of Java are interconnected with one another. Meanwhile, concerning the correlation between time, the Covid-19 case that occurred one day was greatly influenced by the Covid-19 case that occurred in the previous days. This is because the incubation period of the virus and the pattern of transmission of the virus are both factors involved when deciding about this. The purpose of this research is to develop a model for the number of Covid-19 cases that occur each day in Java by modifying the space-time model, GSTAR(1;1), with modified weights by applying the MST Graph theory technique with Prim's Algorithm.

2. RESEARCH METHODS

2.1 Minimum Spanning Tree (MST)

One type of tree that represents a graph is called a spanning tree. The graph is a linked, undirected, but non-treelike graph (has a cycle). While the tree is an example of a tree graph, which may be generated by removing a cycle from within a graph. Therefore, a spanning tree is a type of tree graph in which the set of all vertices is a significant subset of a set of vertices that is contained in graph, and the set of all edges is an appropriate subset of the set of all edges that is contained in graph.

If it is a weighted graph, then the definition of a spanning tree for can be thought of as the accumulation of all the various weights. The term "minimum spanning tree" refers to the tree that, out of all the trees that span, has the fewest leaves on it and, thus, the least amount of weight [8]. The shortest route problem has several permutations, one of which being the least spanning tree problem. The difference between the two problems is the path that is being sought. In the shortest route, we look for a path from source to destination that gives the minimum total distance, whereas in the minimum spanning tree, the question is how to determine the arcs connecting the nodes in the network in such a way that the minimum total arc length is obtained. This is in contrast to the search for the shortest route, which looks for a path from source to destination that gives the minimum total distance [9].

2.2 Prim Algorithm

For a weighted linked graph, Prim's algorithm in graph theory finds for the minimum spanning tree. This entails identifying the subset of edges in a tree that include each vertex and have the least overall weight of all the edges in the tree. The lowest spanning forest (minimum spanning tree for each connected component) is found if the graph is not connected [10]. The fundamental idea behind Prim's algorithm is to select an edge from the graph G at each step that has the lowest weight but does not form a circuit in T. The following is Prim's algorithm:

i. Begin the process of sorting each edge of the graph G by beginning with the side that has the least weight.
ii. Pick the side (u, v) that has the least amount of weight while it is unformed, that is, when the circuit is at the T position. Add (u, v) into T.
iii. Keep repeating step 2 until the tree spans a minimum, which is shown by the fact that there are $n - 1$ edges in the tree that spans T (n is the number of vertices of the graph G)

2.3 Generalized Space-Time Autoregressive (GSTAR) Model

A generalization of the Space Time Autoregressive (STAR) model is presented here in the form of the GSTAR model. Because all of the locations in the STAR model have the same autoregression parameters, it is possible to make the assumption that the locations that were utilized were homogenous. This means that the locations that were used were considered to be the same. As a result, this restricts the use of the STAR model to just uniform locations. As a result, the STAR model was transformed into the GSTAR model by adding more adaptability to the original STAR model [11]. The GSTAR general model with the time order p and the spatial order k is as follows:

$$Y_{i,t} = \left(\sum_{k=\ell}^{p} \sum_{\ell=0}^{\lambda_k} \phi_{k\ell} \boldsymbol{W}^{(\ell)} Y_{i,t-k} \right) + e_{i,t}$$

where ϕ_{kl} is GSTAR's parameter, $\boldsymbol{W}$ is weight matrix, and $e_{i,t}$ is noise term (normally distributed with zero mean and constant variance).

One of the challenges of GSTAR modeling is determining the location weights, which can be challenging because the location weights that are chosen have to be appropriate for use with the data that is being examined [12]. Within the framework of the GSTAR model, Suhartono and Atok present a variety of different weighting strategies [13]. Uniform weights, binary distances, and inverse distances are some of the methods that are typically utilized for determining weights. A uniform weight assigns the same value for the weight to each individual location. As a result, this location weight is frequently utilized for sets of data whose locations are either identical to one another or have the same distance between them. On all of the other elements besides the main diagonal, the binary weight values are 0 and 1, respectively. The significance of the weight is determined by the connection that exists between the various locations. The equation $w_{ij} = 1$ describes the connection that exists between two cities that are located relatively close to one another. In the meantime, it is defined as $w_{ij} = 0$ if the locations are geographically far away. The actual distance that exists in the field between each of these sites is what will be used to calculate this weight. The normalization of the actual distance inverse result is what provides the basis for the calculation of the weight.

2.4 Modified Weight Matrix

A theoretical graph strategy is applied to modify the weight matrix presented in this paper. Each location is believed to be interconnected in the weight matrix that is typically employed. This makes it possible to look for the weights that connect the various places in a straightforward manner. When approaching the problem using a strategy based on graph theory, the starting assumption is that every place is linked to every other site. The weight difference between the places is proportional to the number of kilometers that separate them. This weight adjustment makes use of both the distance assumption and correlation as the first stage in the process of assessing whether or not the points used are connected to one another. The distance assumption is used. If there is a substantial correlation between the points (more than 75% of the time), then we say the points are related. As a direct consequence of this, a brand-new graph that serves as a depiction of the connection between points has been produced. The initial graph formed is used as the initial graph in determining MST. Prim's algorithm is applied in finding MST. After the graph has been constructed, the connected edges that comprise the MST graph will serve as the foundation around which the weight matrix will be based. The distance that exists between the points in the updated distance matrix is the value that is used to represent the linked edge of the MST graph. While the value of the edge that is not connected is 0, The following stage, which comes after the formation of the updated distance matrix, is to normalize each row of the matrix so that the sum of the values in each row is equal to 1. Following the previous sentence, the normalized matrix will be called the modified weight matrix.

3. RESULTS AND DISCUSSION

3.1. Descriptive Statistics

The National Disaster Management Agency provided the information that was used, which was statistics on the growth of Covid-19 instances per day for the period beginning on July 1st, 2020, and ending on January 18th, 2021 (202 observations) [14].In this investigation, the information is broken down into two categories: training data (197 observations that span from July 1st, 2020 to January 13rd, 2021) and testing data (January 14th – 18th, 2021). The provinces of Java, including Banten (V_1), DKI Jakarta (V_2), West Java (V_3), Central Java (V_4), DI Yogyakarta (V_5), and East Java (V_6), served as the locations for this study. Table 1 contains statistical information that is illustrative of each location.

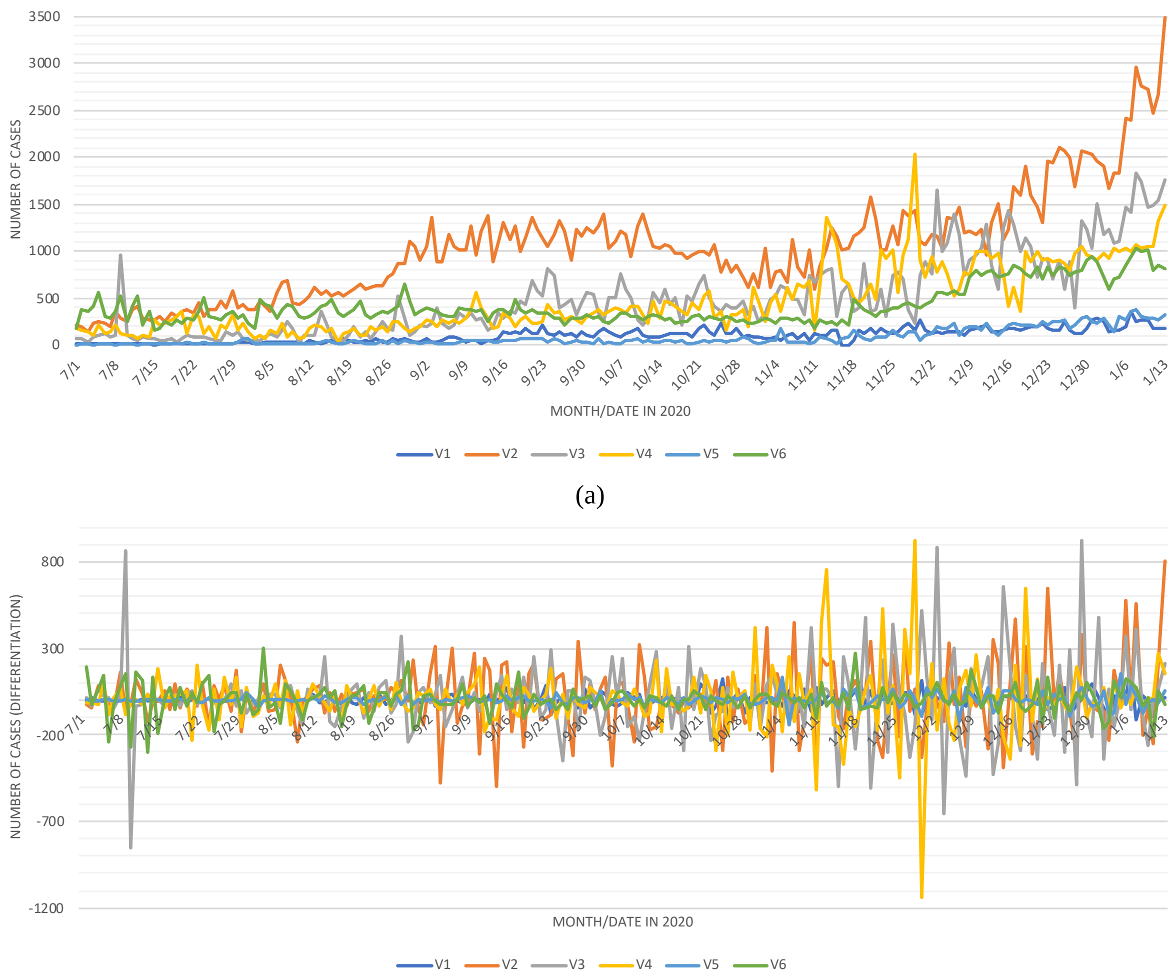
Table 1. Distance between Edges

	V_1	V_2	V_3	V_4	V_5	V_6
Min.	0.00	138.00	27.00	36.00	1.00	165.00
Mean	103.48	1080.70	548.78	498.17	83.74	436.08
Q_1	32.25	608.00	152.50	198.00	19.25	283.25
Q_2	104.50	1032.50	403.00	351.00	41.50	350.00
Q_3	154.00	1273.75	746.25	762.50	119.25	511.75
Variance	5881.02	423201.24	273746.89	163158.77	8576.29	49757.96
Dev. Std.	76.69	650.54	523.21	403.93	92.61	223.06
Max.	364.00	3536.00	3460.00	2036.00	379.00	1198.00

Figure 1.a additionally depicts the overall data plot for each location. The statistics exhibit a significant amount of fluctuation throughout the period of time, as shown by the graphic. Considering that stationarity of the data on the mean and variance is one of the most important assumptions in space-time modeling, then the data have to be differentiated in the first order if it is going to be used [15]. It would appear that the data

already satisfy the assumption of stationarity based on the plot that is shown in **Figure 1.b**, which shows the results of the first differentiation.

Additionally, the distance between edges is determined by taking into account the actual distance covered in the area with the assistance of Google Maps in the calculation process. The determined distance is expressed in kilometers. The edges that were taken into consideration were those of the province's capital city. **Table 2** has a comprehensive listing of the distances that separate each edge.



(b)

Figure 1. Time Series Plot of (a) Number of Daily Cases and (b) Differentiation of Covid-19 for Six Locations.

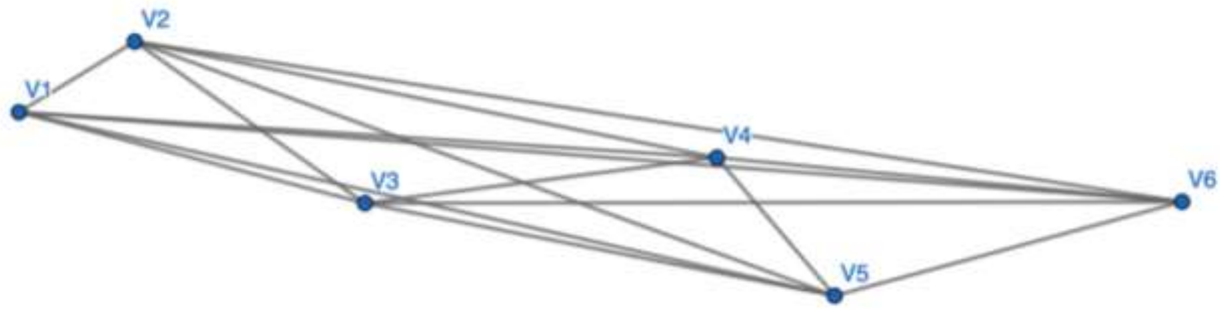
Table 2. Distance between Edges

$Dist.$	V_1	V_2	V_3	V_4	V_5	V_6
V_1	0.00	102.33	189.94	465.80	511.29	848.50
V_2	102.33	0.00	115.34	374.71	425.00	755.49
V_3	189.94	115.34	0.00	278.33	321.73	661.06
V_4	465.80	374.71	278.33	0.00	96.46	382.90
V_5	511.29	425.00	321.73	96.46	0.00	360.00
V_6	848.50	755.49	661.06	382.90	360.00	0.00

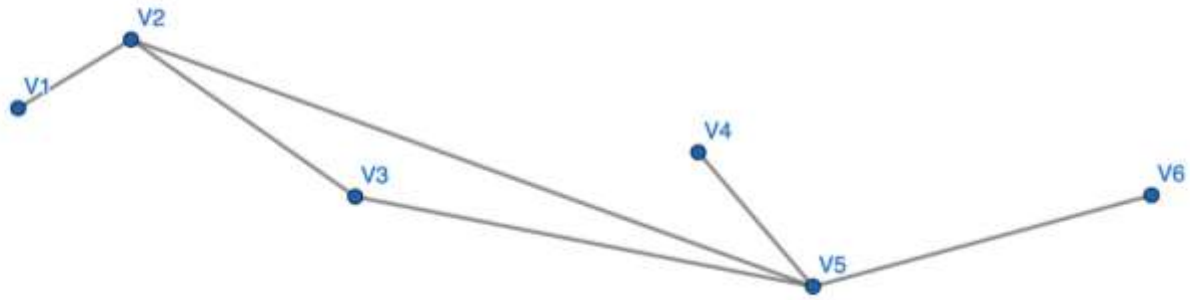
3.2 Modified Weight Matrix using Prim's Algorithm

Assuming that all of the vertices are connected is the first thing that needs to be done in order to figure out the adjusted weight matrix before starting Prim's algorithm (see **Figure 2.a**). The distance that exists between the vertices, which is denoted by the matrix of distance (denoted by D), serves as a representation of the edge that connects the vertices.

$$D = \begin{bmatrix} 0 & 102.33 & 189.94 & 465.80 & 511.29 & 848.50 \\ 102.33 & 0 & 115.34 & 374.71 & 425.00 & 755.49 \\ 189.94 & 115.34 & 0 & 278.33 & 321.73 & 661.06 \\ 465.80 & 374.71 & 278.33 & 0 & 96.46 & 382.90 \\ 511.29 & 425.00 & 321.73 & 96.46 & 0 & 360.00 \\ 848.50 & 755.49 & 661.06 & 382.90 & 360.00 & 0 \end{bmatrix}$$



(a)



(b)

Figure 2. (a) Connected and (b) Initial Graph

Observe that all of the vertices in **Figure 2.a** are connected if only the distance between them is considered. In fact, however, especially for Covid-19 instances, the correlation of daily cases between points must also be considered. The correlation between vertices of the increase in Covid-19 is displayed in **Table 3**.

Table 3. Correlation of Daily New Cases Covid-19 between Edges

$Corr$	V_1	V_2	V_3	V_4	V_5	V_6
V_1	1.00	0.77	0.74	0.70	0.72	0.59
V_2	0.77	1.00	0.80	0.73	0.83	0.72
V_3	0.74	0.80	1.00	0.72	0.85	0.74
V_4	0.70	0.73	0.72	1.00	0.75	0.60
V_5	0.72	0.83	0.85	0.75	1.00	0.87
V_6	0.59	0.72	0.74	0.60	0.87	1.00

The text in red in **Table 3** indicates that there is a strong correlation between the vertices (more than 75% of the cases). According to **Table 3**, the following pairs of points exhibit a correlation of greater than 75%: $V_1V_2, V_2V_3, V_2V_5, V_3V_5, V_4V_5$, and V_5V_6. In addition, these points are utilized in the formation of a new graph, which will afterward be utilized by Prim's method in order to find the MST graph. The new graph, shown in **Figure 2.b**, is based on the correlation value between the points, and as a result, a modified distance matrix, D^*, is formed, which could be written as:

$$D^* = \begin{bmatrix} 0 & 102.33 & 0 & 0 & 0 & 0 \\ 102.33 & 0 & 115.34 & 0 & 425.00 & 0 \\ 0 & 115.34 & 0 & 0 & 321.73 & 0 \\ 0 & 0 & 0 & 0 & 96.46 & 0 \\ 0 & 425.00 & 321.73 & 96.46 & 0 & 360.00 \\ 0 & 0 & 0 & 0 & 360.00 & 0 \end{bmatrix}$$

Applying Prim's technique in order to locate the MST graph is the following step in the process. The first step of this procedure is to pick one of the nodes. Because there were only six nodes available for this research, the researcher decided to make all possible combinations out of the initial node options. Suppose that node V1 is chosen; the next step in Prim's Algorithm is to select the node with the lowest weight that is connected to V1; this will be done by selecting node with the lowest weight. The weight at concern here is the weight assigned to the entry for the modified distance matrix. As a consequence of this, the node V_2 with a weight of 102.33 has the least amount of weight when combined with V_1 (distance in km). Next, select a node that is connected to V_2 (other than V_1) and has minimal weight. In this example, V_3 is chosen since it has a weight of 115.34, therefore this step is complete. The following step is to select a node that has a minimum weight and is connected to V_3 in addition to V_1 and V_2. After this, V_5 is chosen and given a weight of 321.73 as the final step. In the same manner, V_4 is chosen, which is linked to V_5 and has a weight of 96.46. Since V_4 is not connected to any other node besides V_5 (which means that V_5 cannot be re-selected in this scenario), it reverts to the prior node, which is V_5. Search for other nodes besides V_4 that are connected to V_5 (besides V_1, V_2, and V_3); in this case, the last node V_6 has a weight of 360, so look for that. As a consequence of this, the total spanning tree is equal to the sum of the weights that were chosen prior, namely

$$\begin{aligned} T_{V_1} &= D^*_{V_1V_2} + D^*_{V_2V_3} + D^*_{V_3V_5} + D^*_{V_5V_4} + D^*_{V_5V_6} \\ &= 102.33 + 115.34 + 321.73 + 96.46 + 360.00 \\ &= 995.86 \end{aligned}$$

In addition, the Prim's Algorithm is applied, in the same way, to a number of other starting nodes, specifically V_2, V_3, V_4, V_5, and V_6. In **Figure 3**, each initial vertex is represented by its own spanning tree graph. **Figures 3.a, 3.b, 3.d, 3.e**, and **3.f** all have the same spanning tree graph, but their starting vertices are different. On the other hand, **Figure 3.c** has a completely different spanning tree graph. As a consequence of this, the total weight of the spanning tree with the beginning vertices V_1, V_2, V_4, V_5, and V_6 has the same total weight, in contrast to the total weight with the initial node V_3, which has a different value. **Table 4** provides additional information that can be used in the calculation of the total weight of the spanning tree. It is possible to make the following conclusion from **Table 4**: there are two minimum spanning trees constructed, each of which has a total weight of 995.86 and 1099.13 correspondingly. In this particular instance, the MST graph that has a total weight that is equal to or more than 995.86 is selected. This indicates that a graph with initial vertices V_1, V_2, V_3, V_4, V_5, or V_6 may be chosen to function as a representation of the modified weight matrix.

$V_1V_2 - V_2V_3 - V_3V_5 - V_5V_4 - V_5V_6$

(a)

$V_2V_1 - V_2V_3 - V_3V_5 - V_5V_4 - V_5V_6$

(b)

$V_3V_2 - V_2V_1 - V_2V_5 - V_5V_4 - V_5V_6$

(c)

$V_4V_5 - V_5V_3 - V_3V_2 - V_2V_1 - V_5V_6$

(d)

$V_5V_4 - V_5V_3 - V_3V_2 - V_2V_1 - V_5V_6$

(e)

$V_6V_5 - V_5V_4 - V_5V_3 - V_2V_1 - V_5V_6$

(f)

Figure 3. MST Graph with Initial Vertices (a) V_1, (b) V_2, (c) V_3, (d) V_4, (e) V_5, and (f) V_6

Table 4. MST Weight Calculation for Each Initial Node

Edge	Weight	Edge	Weight	Edge	Weight	Edge	Weight	Edge	Weight	Edge	Weight
V_1V_2	102.33	V_2V_1	102.33	V_3V_2	115.34	V_4V_5	96.46	V_5V_4	96.46	V_6V_5	360.00
V_2V_3	115.34	V_2V_3	115.34	V_2V_1	102.33	V_5V_3	321.73	V_5V_3	321.73	V_5V_4	96.46
V_3V_5	321.73	V_3V_5	321.73	V_2V_5	425.00	V_3V_2	115.34	V_3V_2	115.34	V_5V_3	321.72
V_5V_4	96.46	V_5V_4	96.46	V_5V_4	96.46	V_2V_1	102.33	V_2V_1	102.33	V_2V_1	102.33
V_5V_6	360.00	V_5V_6	360.00	V_5V_6	360.00	V_5V_6	360.00	V_5V_6	360.00	V_5V_6	360.00
T_{V_1}	**995.86**	T_{V_2}	**995.86**	T_{V_3}	1099.13	T_{V_4}	**995.86**	T_{V_5}	**995.86**	T_{V_6}	**995.86**

In this context, the graph $V_1V_2 - V_2V_3 - V_3V_5 - V_5V_4 - V_5V_6$with a minimum total weight of 995.86 is chosen as input for the modified weight matrix, so that a new modified distance matrix is formed based on the MST graph, namely D^{**},

$$\boldsymbol{D}^{**} = [d_{ij}^{**}] = \begin{bmatrix} 0 & 102.33 & 0 & 0 & 0 & 0 \\ 102.33 & 0 & 115.34 & 0 & 0 & 0 \\ 0 & 115.34 & 0 & 0 & 321.73 & 0 \\ 0 & 0 & 0 & 0 & 96.46 & 0 \\ 0 & 0 & 321.73 & 96.46 & 0 & 360.00 \\ 0 & 0 & 0 & 0 & 360.00 & 0 \end{bmatrix}$$

The following is a step-by-step guide on how to obtain a modified weight matrix by normalizing the rows on the $D\ **$ matrix as

$$w_{ij}^* = \frac{d_{ij}^{**}}{\sum_{j=1}^{6} d_{ij}^{**}}$$

Hence,

$$\boldsymbol{W}^* = [w_{ij}^*] = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0.53 & 0 & 0.47 & 0 & 0 & 0 \\ 0 & 0.74 & 0 & 0 & 0.26 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.19 & 0.63 & 0 & 0.17 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

3.3 GSTAR (1;1) Model

The following step, which comes after obtaining the modified weight matrix, is to estimate the parameters of the GSTAR (1;1) model by utilizing the Least Squares approach. The parameters that were acquired, such as the following results.

$$\boldsymbol{\Phi}_{10} = [-0.29; -0.33; -0.38; -0.31; -0.34; -0.29]$$
$$\boldsymbol{\Phi}_{11} = [0.03; -0.01; -0.01; 0.43; -0.01; 0.21]$$

so that

$$\begin{aligned} Z_{1,t} &= -0.29Z_{1,t-1} + 0.03Z_{2,t-1} \\ Z_{2,t} &= -0.01Z_{1,t-1} - 0.33Z_{2,t-1} - 0.01Z_{3,t-1} \\ Z_{3,t} &= -0.01Z_{2,t-1} - 0.38Z_{3,t-1} - 0.01Z_{5,t-1} \\ Z_{4,t} &= -0.31Z_{4,t-1} + 0.43Z_{5,t-1} \\ Z_{5,t} &= -0.01Z_{3,t-1} - 0.01Z_{4,t-1} - 0.34Z_{5,t-1} - 0.01Z_{6,t-1} \end{aligned}$$

$$Z_{6,t} = 0.21Z_{5,t-1} - 0.29Z_{6,t-1}$$

where $Z_{i,t} = Y_{i,t} - Y_{i,t-1}$

Therefore,

$$\boldsymbol{Y_t} = (\boldsymbol{I} + \boldsymbol{\Phi_{10}} + \boldsymbol{\Phi_{11}}\boldsymbol{W})\boldsymbol{Y_{t-1}} - (\boldsymbol{\Phi_{10}} + \boldsymbol{\Phi_{11}}\boldsymbol{W})\boldsymbol{Y_{t-2}}$$

so,

$$\begin{aligned}
Y_{1,t} &= 0.71Y_{1,t-1} + 0.03Y_{2,t-1} \\
Y_{2,t} &= -0.01Y_{1,t-1} + 0.67Y_{2,t-1} - 0.01Y_{3,t-1} \\
Y_{3,t} &= -0.01Y_{2,t-1} + 0.62Y_{3,t-1} - 0.01Y_{5,t-1} \\
Y_{4,t} &= 0.69Y_{4,t-1} + 0.43Y_{5,t-1} \\
Y_{5,t} &= -0.01Y_{3,t-1} - 0.01Y_{4,t-1} + 0.66Y_{5,t-1} - 0.01Y_{6,t-1} \\
Y_{6,t} &= 0.21Y_{5,t-1} + 0.71Y_{6,t-1}
\end{aligned}$$

After obtaining a model, the next step is to fit the data to the model. This is done to determine how accurately the model captures the patterns found in the data. **Figure 4** presents the results of fitted values based on the GSTAR (1;1) model with modified weights utilizing the MST graph theory approach for each location. These values were determined by applying the model to the data.

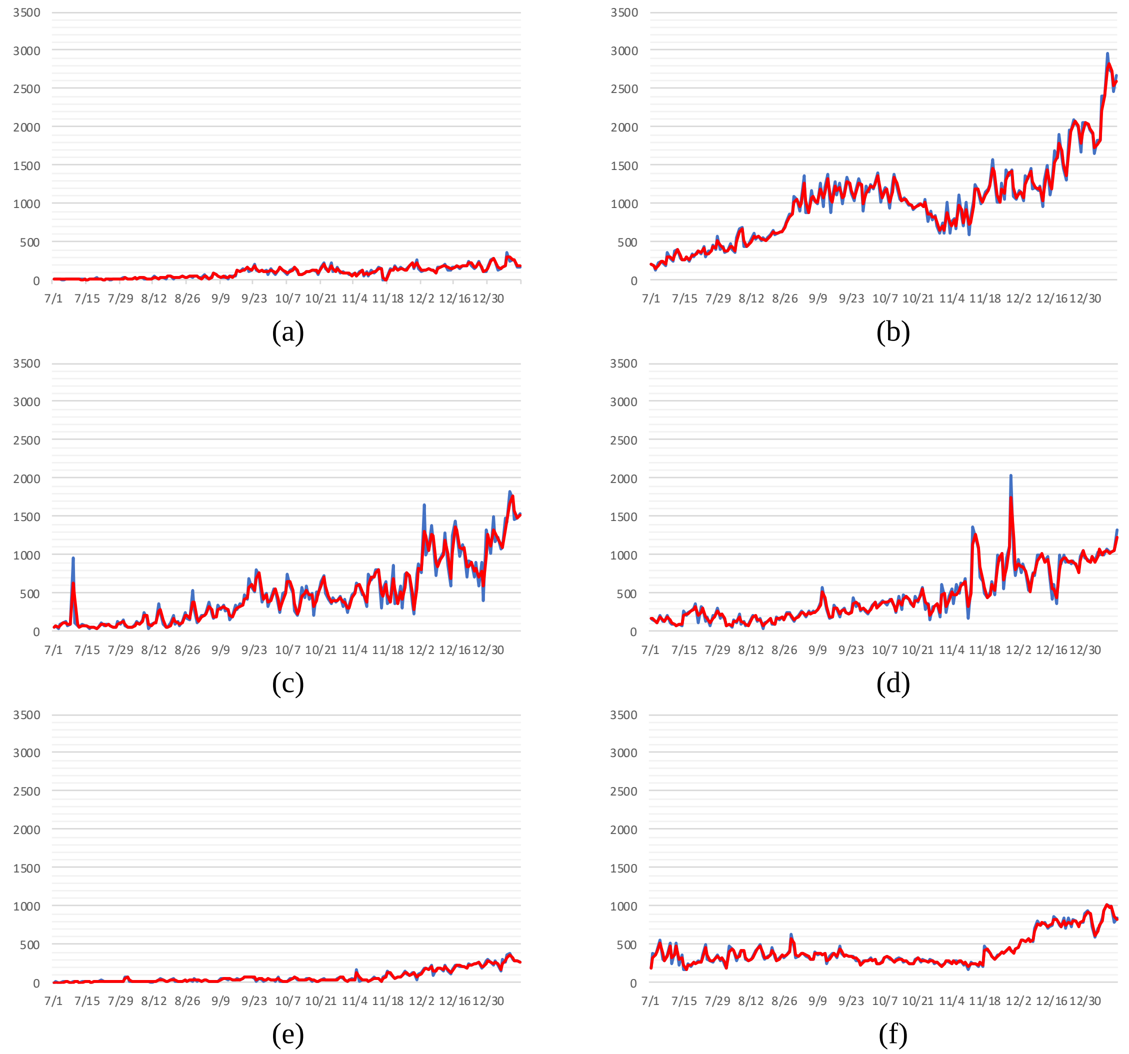


Figure 4. Fitted Values vs Actual for (a) $\boldsymbol{V_1}$, (b) $\boldsymbol{V_2}$, (c) $\boldsymbol{V_3}$, (d) $\boldsymbol{V_4}$, (e) $\boldsymbol{V_5}$, and (f) $\boldsymbol{V_6}$

3.4 Forecasting

Predictions are made on the testing data and for the next five times. When testing a model, prediction is performed to determine how well the model can be utilized to make predictions. A comparison of the test data with the actual data is shown in **Figure 5**. The results of determining the accuracy of the model for testing data and training data at each location using Mean Absolute Percentage Error (MAPE) are shown in Table 5. As a direct consequence of this, the average MAPE value for the training data is identified to be lower than the value obtained for the testing data. This indicates that the GSTAR (1;1) model with modified weights that offers the Graph theory approach and Prim's algorithm is more effective at estimating than it is at forecasting. In addition, the outcomes of the predictions made for the next five times using the GSTAR (1;1) model with modified weights are presented in **Table 6** below. These results can be found below.

Table 5. MAPE for Training and Testing Data

MAPE	V_1	V_2	V_3	V_4	V_5	V_6	Average
Training	13.70	5.08	18.13	9.74	17.21	4.55	**11.40**
Testing	28.62	16.10	28.38	32.57	5.51	18.23	21.57

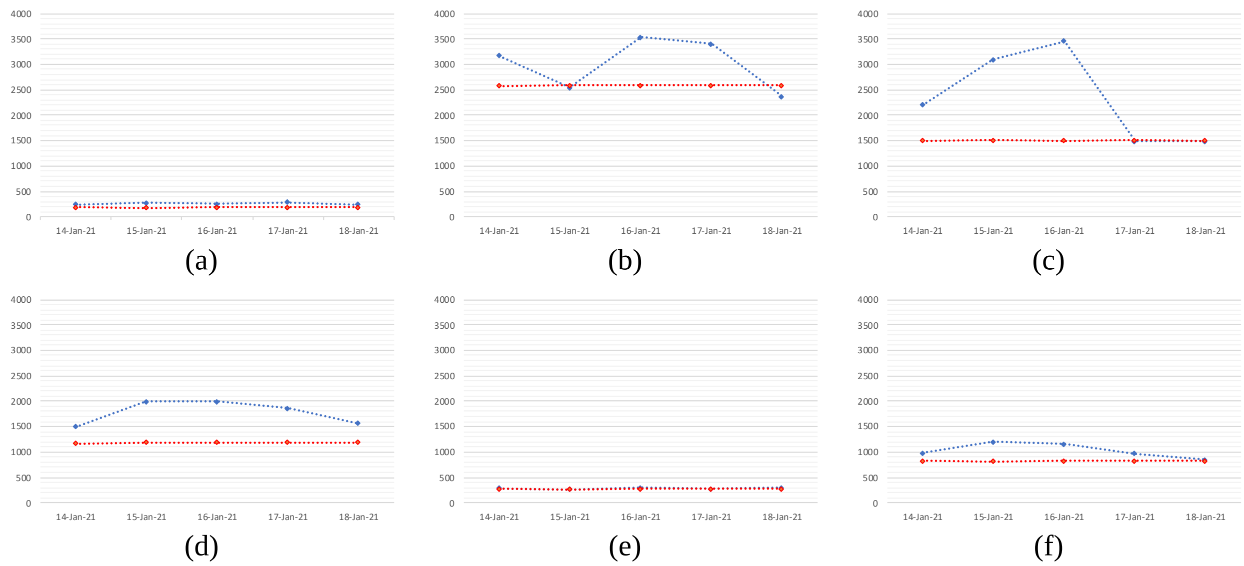


Figure 5. Forecasting for Testing Data (a) V_1, (b) V_2, (c) V_3, (d) V_4, (e) V_5, and (f) V_6

Table 6. Forecasting for Five Days Ahead

Period	V_1	V_2	V_3	V_4	V_5	V_6
Jan 19th	227	2706	1494	1657	291	888
Jan 20th	218	2665	1497	1520	282	875
Jan 21st	219	2679	1496	1558	286	877
Jan 22nd	220	2674	1496	1548	284	877
Jan 23rd	219	2676	1496	1551	285	877

4. CONCLUSIONS

Modification of the weight matrix can be applied to the GSTAR (1;1) model by utilizing the MST Graph theory technique with Prim's algorithm. As a direct consequence of this, the provinces of Banten and Jakarta, Jakarta and West Java, West Java and Central Java, DIY and East Java, and DIY and East Java are all connected to one another. The GSTAR (1;1) model, which represents the spatial and time patterns of Covid-19 cases on the island of Java, is obtained once the adjusted weights have been calculated. When applied to the training data, the accuracy of the model has an average error of 11.4% across all of the locations. While the overall accuracy of the predictions is only 21.6% of the time. This indicates that the GSTAR (1;1) model with modified weights should be utilized for data estimate purposes for this particular scenario. In addition, forecasts for the occurrence of more Covid-19 cases on the island of Java are carried out for the subsequent five times, specifically from the 19th to the 23rd of January in 2021.

AKNOWLEDGEMENT

The authors thank to DIPA FMIPA Universitas Tanjungpura Pontianak 2022 for supporting funds.

REFERENCES

[1] P. Pfiefer and S. Deutsch, "A Three-Stage Iterative Procedure for Space-Time Modeling," *Technometrics*, vol. 22, no. 1, pp. 35–47, Feb. 1980.

[2] Y. Yundari, U. S. Pasaribu, and U. Mukhaiyar, "Error Assumptions on Generalized STAR Model," *Journal of Mathematical and Fundamental Sciences*, vol. 49, no. 2, p. 136, Oct. 2017, doi: 10.5614/j.math.fund.sci.2017.49.2.4.

[3] U. Mukhaiyar, U. S. Pasaribu, W. S. Budhi, and K. Syuhada, "The stationarity of generalized STAR(2;λ1,λ2) process through the invers of autocovariance matrix," 2014, pp. 484–487. doi: 10.1063/1.4868849.

[4] P.-C. Lai, F.-M. So, and K.-W. Chan, *Spatial Epidemiological Approaches in Disease Mapping and Analysis*. CRC Press, 2008. doi: 10.1201/9781420045536.

[5] N. Nurhayati, U. S. Pasaribu, and O. Neswan, "Application of Generalized Space-Time Autoregressive Model on GDP Data in West European Countries," *J Probab Stat*, vol. 2012, pp. 1–16, 2012, doi: 10.1155/2012/867056.

[6] U. Mukhaiyar and U. S. Pasaribu, "A New Procedure of Generalized STAR Modeling using IAcM Approach," *ITB Journal of Sciences*, vol. 44, no. 2, pp. 179–192, 2012, doi: 10.5614/itbj.sci.2012.44.2.7.

[7] U. Mukhaiyar, B. I. Bilad, and U. S. Pasaribu, "The Generalized STAR Modelling with Minimum Spanning Tree Approach of Weight Matrix for COVID-19 Case in Java Island," *J Phys Conf Ser*, vol. 2084, no. 1, p. 012003, Nov. 2021, doi: 10.1088/1742-6596/2084/1/012003.

[8] R. Munir, *Matematika Diskrit*, Third. Bandung, 2005.

[9] T. T. Dimyati and A. Dimyati, *Operations Research Model-Model Pengambilan Keputusan*. Bandung: Sinar Baru Algensindo, 1999.

[10] N. Subadra, "Directed Graph Algorithms for Tours - A Case Study," *Journal of Emerging Trends in Engineering and Applied Sciences (JETEAS)*, vol. 2, no. 4, pp. 615–618, 2011.

[11] S. Borovkova, H. Lopuhaa, and B. Nurani, "Generalized STAR model with experimental weights," in *17th International Workshop on Statistical Modelling*, 2002, pp. 139–147.

[12] Yundari, U. S. Pasaribu, U. Mukhaiyar, and M. N. Heriawan, "Spatial Weight Determination of GSTAR(1;1) Model by Using Kernel Function," *J Phys Conf Ser*, vol. 1028, p. 012223, Jun. 2018, doi: 10.1088/1742-6596/1028/1/012223.

[13] Suhartono and R. Atok, "Pemilihan Bobot Lokasi yang Optimal pada Model GSTAR."

[14] Badan Nasional Penanggulangan Bencana, "Update Percepatan Penanganan Covid-19 di Indonesia," *BNPB*, Jan. 2021.

[15] D. J. Cryer and K.-S. Chan, *Time Series Analysis: With Application in R*. USA: Springer Science and Bussiness Media, 2008.