Modeling Human Mobility by Train on the Spread of COVID-19 in East Java Province Using Distance-Decay PageRank Algorithm

Rizha Al-Fajri, Medria Kusuma Dewi Hardhienata, Yeni Herdiyeni
Computer Science, Institut Pertanian Bogor, Bogor 16880, Indonesia

ARTICLE INFO

Article history:

Received September 24, 2023
Revised October 04, 2023
Published October 07, 2023

Keywords:

Modeling Human Mobility;
COVID-19;
Distance-decay PageRank;
Spatial Network;
East Java

ABSTRACT

Since early 2020, the world has been dealing with the COVID-19 outbreak. A person who has been infected with COVID-19 has the potential to transmit the virus to others. This study aims to model human mobility by train using the spatial network in East Java Province. This research examines the relationship between human mobility by train and the spread of COVID-19 in East Java Province. The spatial network is formed based on train stations and train trips, and the model was created using the Distance-decay PageRank algorithm. This research has modeled human mobility using the train in East Java Province. The result shows that human mobility by train is highly correlated with the spread of COVID-19 in East Java Province, with a correlation coefficient of 0.7 (r = 0.7).

Corresponding Author:

Rizha Al-Fajri, Computer Science, Institut Pertanian Bogor, Bogor 16880, Indonesia
Email: rizhaalfajri@gmail.com

1. INTRODUCTION

Since early 2020, the world has been dealing with an outbreak of a highly deadly disease called COVID-19 [1]–[4]. COVID-19 is an infectious disease that attacks the human respiratory tract (https://www.who.int/emergencies/diseases /novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19). This disease is caused by the Coronavirus, SARS-CoV-2 (https://www.who.int/emergencies/diseases /novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19). COVID-19 was first identified in Wuhan, the capital of Hubei Province, China, and reported by the Chinese government in late December 2019 before spreading globally and becoming a worldwide outbreak [5]–[9]. As of April 9, 2021, 134 million cases and 2.9 million deaths were reported globally (https://www.worldometers.info/coronavirus). COVID-19 is a highly contagious virus [10], [11] also state that viral infection can occur when breathing in droplets or touching the nose, mouth, or eyes after touching the contaminated surface [12]. Komplas *et al.* (2020) and [13] believe that COVID-19 also transmits through aerosols.

Indonesia is one of the countries affected by the COVID-19 outbreak [14]–[17]. As of April 9, 2021, 1,552,880 positive cases were reported, and 42,227 died (https://www.worldometers.info/coronavirus). According to [18], East Java Province is one of the hot spots for spreading the COVID-19 disease in Indonesia. Even on May 5, 2020, East Java Province was the area with the highest additional cases of COVID-19 in Indonesia (https://www.cnnindonesia.com/nasional/20200510011524-20-501685/kasus-corona-melonjak-jatim-tertinggi-per-9-mei) As of April 9, 2021, 142,028 cases were confirmed in the region (http://infocovid19.jatimprov.go.id/#peta).

Human mobility is one contributing factor to the spread of an infectious virus [19]–[21]. [22] have directly detected human mobility using mobile phone data, then analyzed it to identify the spread of COVID-19 in China. According to them, this virus quickly spreads when one moves, resulting in a possibility of wider spreading of the virus [23].

The geographic approach is essential in modeling the concentration of human mobility [24], [25]. One of the approaches that are widely used is the distance approach. Distance-decay successfully depicts spatial interaction patterns more succinctly, in this case, migration or population movement [26]. The distance-decay effect derives from the first law of geography, "Everything is related, but the closer the distance, the closer the relationship is" [27].

The identification of human mobility against the spread of the epidemic can be modeled using the PageRank algorithm [21]. The PageRank algorithm is utilized to rank a web page according to its quality [28]. [28] represent web pages as a network that forms a graph to make an analogy. The PageRank algorithm will set as if a web user randomly visits a web page on a graph or a random surfer. Then, a score will be calculated based on the visit, indicating the quality of each web page. A web page with more visits will rank higher.

In the previous study, [29] adopted the PageRank algorithm to identify and predict human mobility in geospatial networks. It was carried out by creating a network graph from a map, in which nodes represented web pages and links denoted hyperlinks of a web page. Chin & Wen (2015) modified the PageRank algorithm by adding distance parameters as a spatial consideration in identifying human mobility in geospatial networks. This algorithm is called Distance-Decay PageRank. This study indicated that the Distance-Decay PageRank algorithm that considers the distance parameters is more effective in identifying human mobility with a correlation coefficient value of 0.746 if compared to algorithms that use other parameters.

According to President Joko Widodo, the train is one of the transport modes that remains a preferable choice among Indonesian. This argument was supported by a high number of passengers at DAOP VIII Surabaya in 2019, that is, on average, 1 million passengers per month (https://jatim.bps.go.id). During the pandemic, rail transportation kept operating according to the Train Travel Schedule 2021 issued by PT. KAI (https://cloud.kai.id/s/Ga76SQwamHXwEaj?fbclid=IwAR3l50DTiaTzgdwnuCy8UBuH5pv0f7BtBYg8CUQyyGZtXi-Xj2Oxvj96azU#pdfviewe). In the current study, a graph was formed with nodes representing stations and links signifying railroads between stations. The graph was formed based on the Train Travel Schedule 2021 (https://cloud.kai.id/s/Ga76SQwamHXwEaj?fbclid=IwAR3l50DTiaTzgdwnuCy8UBuH5pv0f7BtBYg8CUQyyGZtXi-Xj2Oxvj96azU#pdfviewer).

On the other hand, research related to the geospatial networks of COVID-19 in Indonesia remains scarce [30] have modeled population density and temperature on COVID-19 cases in East Java. However, no studies count aspects of human mobility using rail transportation and geospatial networks in the model. The human mobility by rail transportation in this study focuses on East Java Province, especially around Surabaya, as it is one of the hot spots for spreading COVID-19 in Indonesia [18], [31], with the second-highest number of cases in Indonesia.

Therefore, this study aims to (1) model human mobility by the rail transportation mode in East Java using the Distance-Decay PageRank algorithm and (2) determine the relationship between human mobility using rail transportation and the spread of COVID-19 in East Java Province.

2. METHODS

2.1. Data

All Data used in this study were classified into two: modeling and evaluation data. The modeling data comprised data from the East Java Province map, train travel schedules, the number of train passengers, and the distance between train stations in East Java Province. Meanwhile, the evaluation data included data on the spread of COVID-19 cases in districts/cities in East Java Province.

2.2. Procedure

This study consisted of six primary stages: data collection, node sampling, graph formation, modeling using Distance-decay PageRank, analysis, and implementation. Data were collected at the data collection stage. Then, the sampling was conducted on the nodes used for the graph formation. The graph formation employed data on the distance between train stations and the distribution of train trips in East Java Province collected previously. After the graph was created, modeling using the Distance-Decay PageRank algorithm was performed. After obtaining the results from the model, an analysis of confirmed COVID-19 cases in East Java was carried out. In detail, the stages applied in this research are depicted in Fig. 1.

2.2.1. Data Collection

Data used in this study included data on the spread of COVID-19 cases in districts/cities of East Java Province, Indonesian Topographical Maps (ITM), the distribution of train trips, and the distance between train stations in East Java Province. Data on the spread of confirmed COVID-19 cases referred to numerical data taken from the government of East Java Province http://infocovid19.jatimprov.go.id/#peta Data on the map of East Java Province was downloaded from the website "*Tanah air*" https://tanahair.indonesia.go.id/portal-web

While national road distribution data were retrieved from the Train Travel Schedule 2021 (KAI 2021). Moreover, data on the number of train passengers per station was taken from "*Provinsi Jawa Timur Dalam Angka 2021*" (translated: East Java Province in 2021) (https://jatim.bps.go.id). Meanwhile, data on the distance between train stations constituted numerical data derived from the ArcGis website (https://www.Arcgis.com/apps/View/index.html?appid=27b7119dc6754d3e9e584a4fa71e5744) and Google Maps (https://www.google.com/maps/?hl=id). The research procedure show in Fig. 1.

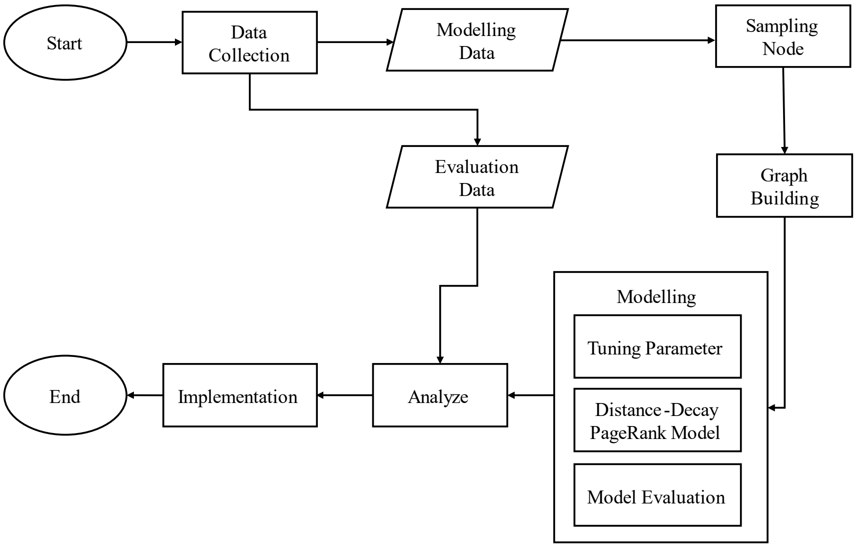

Fig. 1. The research procedure

2.2.2. Node Sampling

Node sampling was conducted to take a sample of stations used as nodes in the graph formation stage. This stage was carried out on the QGIS application by creating a buffer. The buffer was made with a radius of 0.75 Degrees or 83.25 KM from the centroid point of Surabaya City.

2.2.3. Graph formation

The graph was formed by generating the Origin-Destination Matrix according to the Train Travel Schedule 2021. The graph was created by employing data on active train stations and the distribution of train trips in East Java Province. A graph consists of nodes and links. In this study, a node represents a train station in East Java Province, while a link denotes the railroad in East Java Province. The nodes involved stations included in the Buffer radius at the sampling node stage. The link used in this study was a two-way system. The completed graph was then implemented using Distance-Decay PageRank modeling at the subsequent stage.

2.2.4. Parameter Tuning

The greater the value of the distance factor (β), the greater the distance-decay effect is [32]. In other words, the greater the value of β, the more influential the value of the distance between stations on the calculation of the DPRR score will be. In this study, the determination of the optimal distance factor (β) value will be influenced by data on the average number of passengers per station in East Java Province. The data will be utilized to capture the spatial concentration of the population, indicating the intensity of human mobility at each station.

The parameter tuning on the distance factor (β) was carried out at this stage. Trial and error were used from the value of 0 to 10 on β. The modeling and evaluation were performed using the value of β that had been previously attained. The value of β having the highest correlation with the number of train passengers is the most optimal value. This value will be employed in the modeling and evaluation stages.

2.2.5. Modelling using Distance-Decay PageRank

The model formation was conducted by processing graph data generated using the Distance-Decay PageRank algorithm [32]–[35]. The result of this modeling is the DDRR score used to determine the ranking of each node. The workflow of the model is illustrated in Fig. 2. DDRt(a) is the PR score of node a at iteration t. This score will be used when ranking the nodes. B is the distance factor, and distance(i, j) is the distance between node i and node j. The distance-decay in (2) represents the calculation scale of the distance between stations. The greater the distance factor, the more significant the effect of distance on the calculation of the distance-decay effect will be. When the distance is 0, the calculation result will be the same as the initial algorithm, the PageRank algorithm.

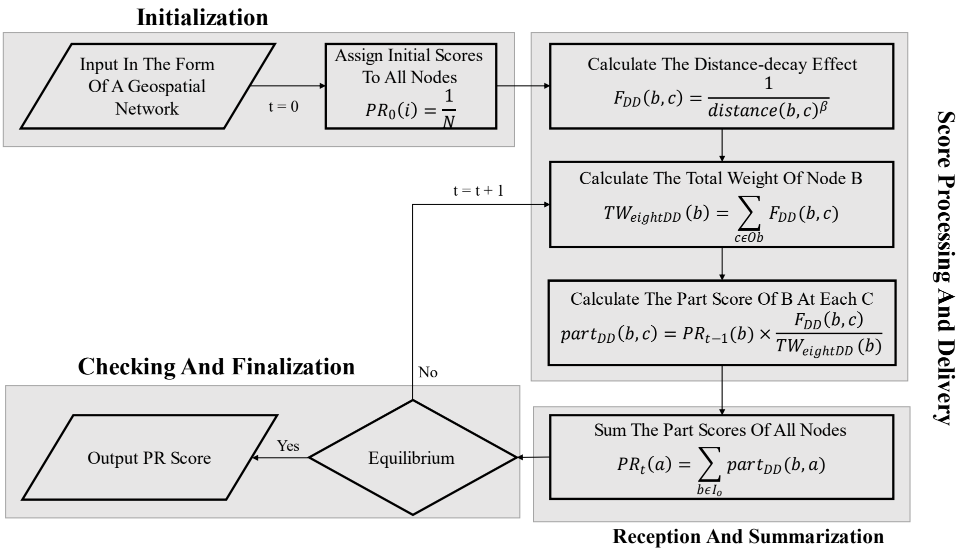

Fig. 2. The workflow of the DDPR model [36]

The model workflow consists of four stages: initialization, processing, and sending of scores, receiving, and summing up scores, and checking and finalizing. At the initialization stage, all nodes in the graph were assigned the same initial value, namely $1/N$, where N is the number of nodes in the graph. At the stage of processing and sending scores, the initial scores were sent to their respective nodes for processing. For each starting node (b), the distance was calculated from the starting node to the nodes c (c∈Ob), the node associated with the starting node. After that, the distance value was used to calculate the weight. This weight was then entered into the calculation, producing the DDRR score from each node. Afterward, the DDRR scores of each node (b) at node (a)) (b∈Ia) were added up to get the DDRR scores of node (a). After yielding the score of nodes (a), the score will be checked at the equilibrium stage. If the score from each node is the same as the previous one, it has reached the state of equilibrium, which will be used as the final score. Otherwise, it will return to the stage of processing and sending scores and subsequently begin the next iteration. The magnitude of the DDRR score indicates the level of human mobility at that node. The greater the DDRR score on a node, the higher the level of human mobility at the node will be.

2.2.6. Model Evaluation

The model evaluation was carried out by calculating the Pearson's correlation coefficient between the DDPR scores yielded by the model of data on the average number of train passengers per station in East Java Province. This stage aims to determine the relationship level between the DDRR scores and the actual concentration of human mobility, in this case, the number of train passengers, by finding the highest correlation coefficient of the two. The higher the correlation value, the closer the relationship between the DDRR scores and the number of passengers per station in East Java Province.

$$r = \frac{\sum_i (x_i - \underline{x})(y_i - \underline{y})}{\sqrt{\sum_i (x_i - \underline{x})^2 \sum_i (y_i - \underline{y})^2}} \tag{1}$$

where, r is Pearson's correlation coefficient of x and y, x is DDPR score from each district/city, y is the number of confirmed COVID-19 cases in East Java, $\underline{x}$ is Mean of PageRank score, $\underline{y}$ is Mean of confirmed COVID-19 cases.

2.2.7. Analysis

The analysis was performed by calculating the Pearson's correlation coefficient (r) and the coefficient of determination (R^2) between the DDPR scores generated by the model and the data on the spread of COVID-19 cases in East Java Province. The value of the determination coefficient signifies the ability of independent variables in the model to explain the variation of dependent variables [37], in this case, the distance variable towards the spread of COVID-19. A low value of R^2 indicates that the ability of the distance variable to explain the spread of COVID-19 is limited. If the value of R^2 is high, the distance variable has a significant influence in predicting the spread of COVID-19.

$$R^2 = \frac{Explained\ Variation}{Total\ Variation} \tag{2}$$

where, R^2 is Coefficient of determination, *Explained variation* is the proportion of variation described by the model, *Total variation* is the proportion of total variation.

2.2.8. Implementation

This study employed a web-based implementation using the Django framework by applying the DDRR modeling function and its analysis of COVID-19 in East Java. The system received the input of an OD Matrix file with the distance parameters from train stations in East Java Province. After the input data were stored, the data were modeled, and the DDRR score for each station was obtained. Afterward, the system showed the categories of scores previously obtained, but this process was not included in this study. Finally, the system displayed the graph in the form of a map, resulting from modeling that had been previously made.

2.3. Development Environment

The specifications of hardware and software used in this research are presented as follows: (1) The hardware is a laptop with specifications: Intel Core i5 Dual-Core 1.8 GHz processor, 8 GB RAM, and VGA Intel HD Graphics 6000; and (2) Software consisting of operating system macOS Catalina version 10.15.7, Python version 3.7.4, Pycharm Edu version 2019.2, QGIS Desktop 3.10, Google Colab [38], [39], and Visual Studio Code [40], [41].

3. RESULTS AND DISCUSSION

3.1. Data Collection

The data used in this study is divided into evaluation and modeling data. Evaluation data is used to evaluate the results at the evaluation stage. Evaluation data was the cumulative data on COVID-19 cases up to June 22, 2021, as shown in Fig. 3.

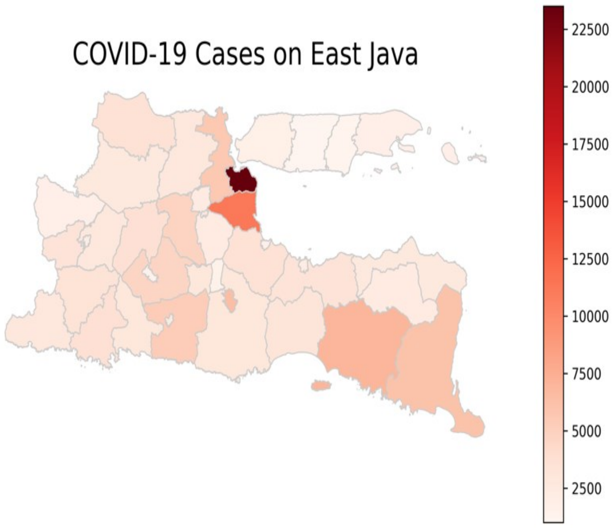

Fig. 3. Cumulative COVID-19 cases per Regency/City

Furthermore, modeling data is used to model human movement at the modeling stage. Modeling data were the number of train passengers, the distribution of train trips, distances between train stations, and the

Indonesian Topographical Map (RBI) of East Java Province. The map of East Java Province can be seen in Fig. 4. These data are used at the stage of graph formation and modeling.

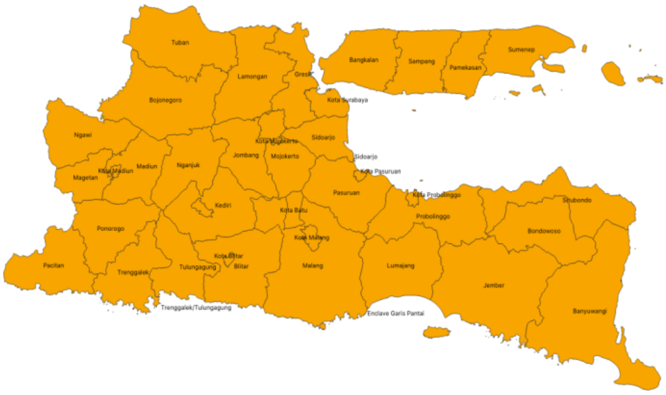

Fig. 4. Topographical Map of East Java Province

3.2. Sampling Node

The data was analyzed using the Buffer feature in the QGIS application. Fig. 5 shows the sampling results of 17 train stations included in the reach of Buffer. Table 1 presents the train stations covered.

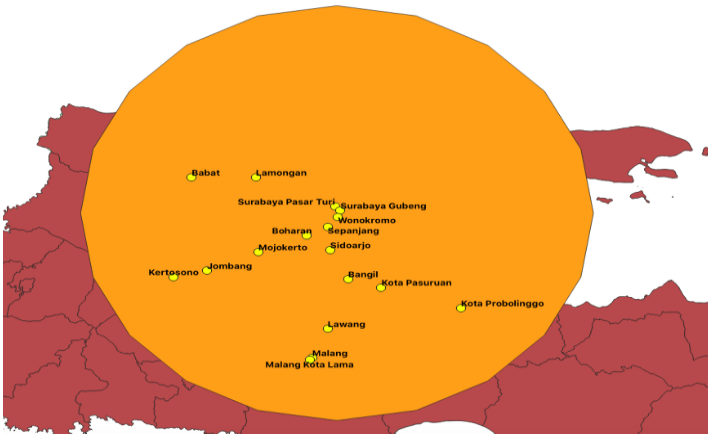

Fig. 5. Results of node sampling in the QGIS application

Table 1. Train stations within the reach of Buffer

Station	Regency/City
Surabaya Gubeng	Surabaya City
Surabaya Pasar Turi	Surabaya City
Wonokromo	Surabaya City
Sidoarjo	Sidoarjo Regency
Sepanjang	Sidoarjo Regency
Boharan	Sidoarjo Regency
Bangil	Pasuruan Regency
Pasuruan	Pasuruan Regency
Mojokerto	Mojokerto Regency
Jombang	Jombang Regency
Kertosono	Nganjuk Regency
Lawang	Malang Regency
Malang	Malang Regency
Malang Kota Lama	Malang Regency
Lamongan	Lamongan Regency
Babat	Lamongan Regency
Probolinggo	Probolinggo Regency

3.3. Graph Formation

After obtaining the stations included in the Buffer region in the previous stage, a graph was formed using the Origin-Destination Matrix with nodes based on the station data and links based on the travel distribution data. While the weight represents the distance between the train stations. Table 2 and Fig. 6 present the Illustrations of the OD Matrix and graphs.

Table 2. Illustration of OD Matrix

Origin	Destination	Distance
Surabaya Gubeng	Wonokromo	4
Surabaya Gubeng	Sepanjang	11
Wonokromo	Sepanjang	7
Wonokromo	Surabaya Gubeng	4
Sepanjang	Surabaya Gubeng	11
Sepanjang	Wonokromo	7

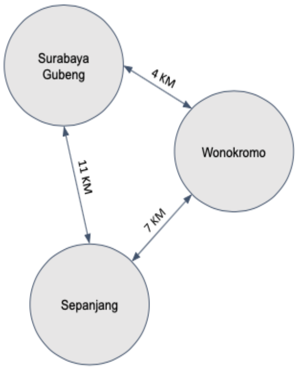

Fig. 6. Graph Illustration

3.4. Parameter Tuning

Parameter tuning is carried out at this stage by modeling and evaluating the model using values 0 to 10 with multiples of 0.1 in the distance factor (β). The modeling results showing the highest correlation with the number of train passengers were analyzed further at the analysis stage.

3.5. Distance-decay PageRank Modelling

The model formation was done by processing the graph data formed using the Distance-Decay PageRank algorithm. This modeling was done using the values assigned previously in the stage of parameter tuning. The correlation coefficient was calculated for the modeling results and human movement at the model evaluation stage.

3.6. Model Evaluation

The model evaluation was conducted by looking for the correlation coefficient (r) between the DDPR score and the number of passengers per station in East Java. The DDRR score used was the modeling result using the values assigned previously at the parameter tuning stage. Fig. 7 presents the distribution of r from each modeling result.

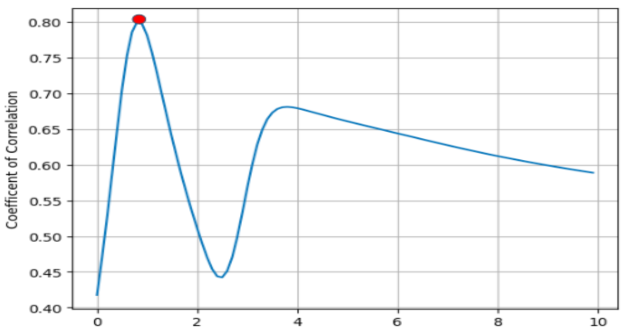

Fig. 7. Correlation coefficient (r) of the modeling with different β

The evaluation results in Fig. 8 illustrate that the highest correlation coefficient (r) was 0.8001, indicating that distance has a strong positive correlation with the average number of train passengers. The correlation is stronger than the research of [36]. Table 3 and Fig. 8 show the results of the DPRR scores per station.

Table 3. Score results from the DPRR modeling

No	Station	DPRR Score
1	Surabaya Gubeng	0.144
2	Wonokromo	0.099
3	Malang	0.09
4	Sidoarjo	0.071
5	Bangil	0.071
6	Malang Kota Lama	0.062
7	Surabaya Pasar Turi	0.058
8	Mojokerto	0.056
9	Sepanjang	0.055
10	Jombang	0.05
11	Kertosono	0.046
12	Kota Pasuruan	0.044
13	Boharan	0.44
14	Lamongan	0.029
15	Kota Probolinggo	0.028
16	Lawang	0.028
17	Babat	0.024

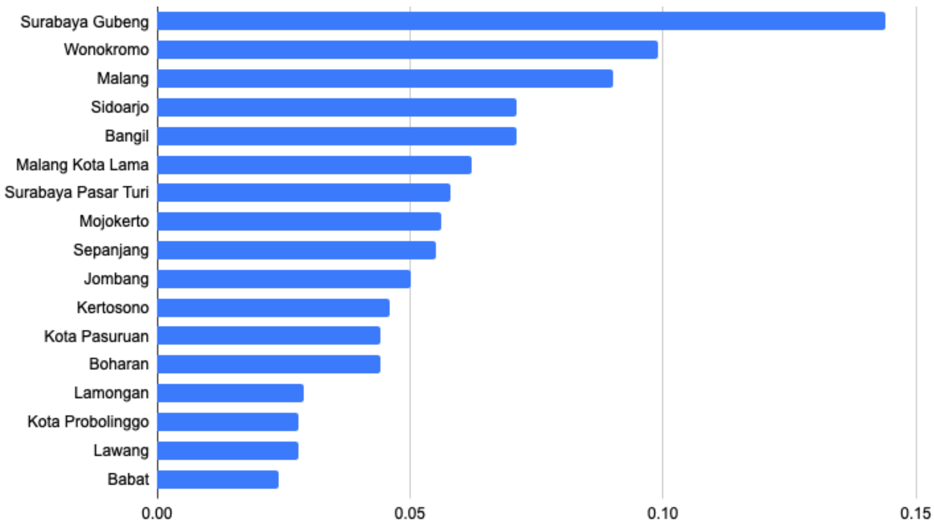

Fig. 8. Modeling results of DPRR scores per station

The modeling results show two stations from the Surabaya area, namely Surabaya Gubeng and Wonokromo Stations, have the highest DPRR scores. This result indicates that Surabaya Gubeng and Wonokromo stations are stations with the highest levels of human movement. Gubeng Station is connected to 15 existing stations. Fig. 9 shows that Wonokromo Station is the closest station to Surabaya Gubeng Station. Moreover, Surabaya is the capital of East Java province.

Fig. 10 shows the distribution of the average number of passengers per station in East Java Province on the DPRR score. The blue dots represent the modeled stations. Fig. 10 reveals that several stations have a low average number of passengers but have a high DDRR score, and vice versa. Wonokromo Station is the second highest GDPR after Surabaya Gubeng Station. However, the average passenger from Wonokromo Station is not too high, being 56,832 passengers per month. It is not far from the average passenger from Mojokerto Station, being 55,731 passengers per month despite Mojokerto Station with a relatively low score. This circumstance is because Wonokromo Station is the closest station to Surabaya Gubeng Station, the station with the highest score in the modeling results, as seen in Fig. 10. Therefore, Wonokromo Station gets a very high score.

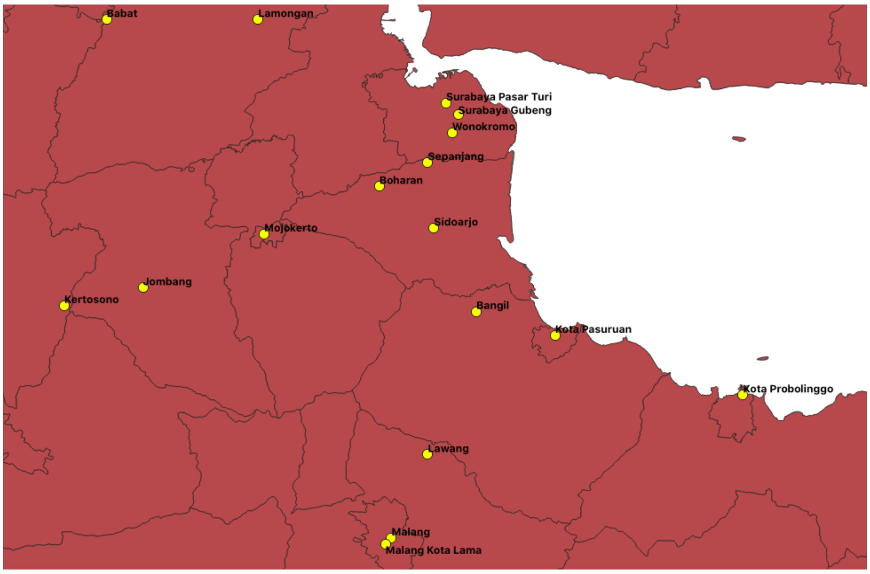

Fig. 9. Station points on the East Java RBI map

Malang Kota Lama Station has the fourth highest GDPR score after Malang Station, and this station has a low average number of passengers (10,305 passengers per month). This is supported by data from https://www. Arcgis.com/apps/View/index.html?appid=27b7119dc6754d3e9e584a4fa71e5744, stating that Malang Kota Lama Station is classified as the medium group station, but it is the closest station to Malang Station. Malang Station has a fairly high DDPR score, the third highest after Surabaya Gubeng and Wonokromo stations. Therefore, Malang Kota Lama Station has a relatively high score.

Fig. 11 shows the mean distance distribution from stations that visit other stations (indegree). The smaller the distance value, the closer the distance between stations. The distribution in Fig. 12 indicates several stations have a close indegree distance but low DDRR score and vice versa. Panjang Station has the closest average indegree distance from other stations. However, the modeling results of the Panjang Station show a score that is not too high (0.055). Looking at the distribution in Fig. 11, we can see that the mean of the number of passengers from the Panjang Station is the third-lowest, higher than Pasuruan and Boharan Stations. It is also supported by the relatively small number of stations connected to the Panjang Station, only five stations, and three of them have a fairly low score: Mojokerto, Jombang, and Kertosono stations. Thus, the Panjang Station obtains a fairly low score.

Malang Station has a fairly high average indegree distance and a high DDRR score. Based on the distribution in Fig. 11, it can be seen that Malang Station has the second-highest average number of passengers per month after Surabaya Gubeng Station. It indicates that Malang Station is a busy station frequently visited by the public. Thus, Malang Station has a high DPRR score.

Fig. 10. The distribution of the average number of passengers on the DDPR score

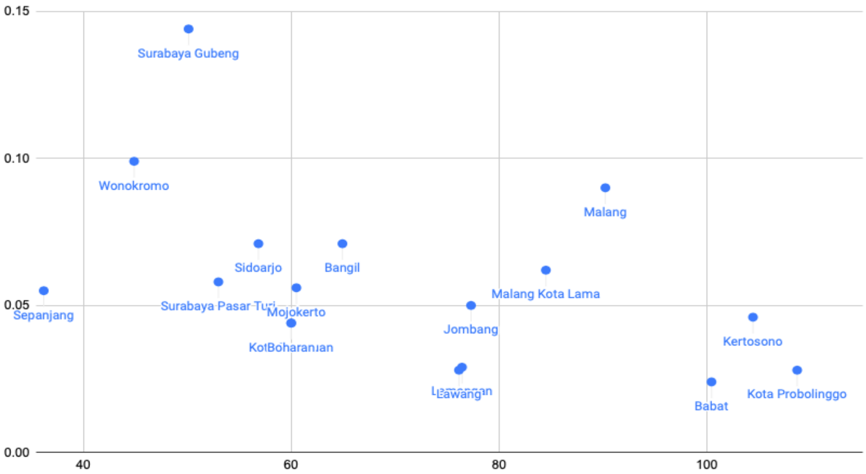

Fig. 11. The mean distribution of indegree distance per station to the DDPR score

3.7. Analysis

The analysis was carried out by calculating the Pearson's correlation coefficient and the coefficient of determination between the Distance Decay PageRank score generated by the model and the data on the spread of confirmed COVID-19 cases in East Java Province. The results can be seen in Table 4.

Table 4. The correlation coefficient and the coefficient of determination

Coefficient	
Pearson's correlation coefficient (r)	0.7
Coefficient of determination (R2)	0.49

Table 4 shows that the correlation coefficient between the Distance Decay PageRank score and confirmed cases of COVID-19 is strongly positive ($r = 0.7$). It indicates that human movement and the spread of COVID-19 in East Java Province are closely related. In other words, the higher the movement of people, the greater the spread of COVID-19 in the area. Meanwhile, the coefficient of determination shows that human movement affects the spread of COVID-19 by 49% in East Java. In other words, there are 51% of other factors affecting the spread of COVID-19 in East Java.

The distribution of data on confirmed cases of COVID-19 in East Java Province against the DPRR score can be seen in Fig. 12.

Fig. 12 shows that several stations have an equal number of COVID-19 cases but have different DPRR scores because data of confirmed cases of COVID-19 is taken by the district/city from each station. Hence, stations located in the same district/city will have the same data on confirmed cases of COVID-19. Meanwhile, the DDPR score at each station is calculated based on the influencing parameters.

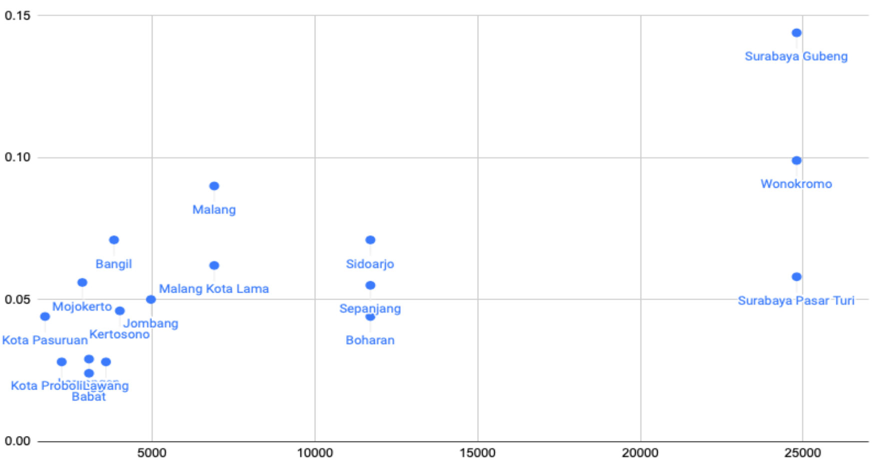

Fig. 12. The distribution of confirmed cases of COVID-19 on the DPRR score

Pasar Turi Surabaya Station is located in Surabaya, so confirmed cases from this station will be equal to other stations located in Surabaya. Surabaya has the highest number of confirmed cases in East Java Province, but the DPRR score from Surabaya Pasar Turi Station is low. Based on the 2021 Train Travel Schedule (KAI 2021), Surabaya Pasar Turi Station is only connected to stations in the northern part of East Java Province and those in other provinces.

Sepanjang, Sidoarjo, and Boharan stations are located in Sidoarjo Regency, a regency with the second-highest number of COVID-19 cases after Surabaya. However, two of these three stations have lower scores than stations in the Malang area: Malang Station and Malang Kota Lama Station, because Malang Station and Malang Kota Lama Station have the closest distance compared to other stations. With this close distance score, both stations get high scores.

3.8. Implementation

Users can model human movement on trains against the spread of COVID-19 in East Java Province. Users must upload an OD Matrix file in excel format (.xlsx) and a confirmed case file for COVID-19 in East Java Province. Fig. 13 presents the form page to upload data. Once the input data is uploaded, the modeling will be done using the DDPR algorithm to get the DDRR score from each station. The DDPR scores per station are displayed on the results page (Fig. 14). Next, the results are also visualized in Fig. 15.

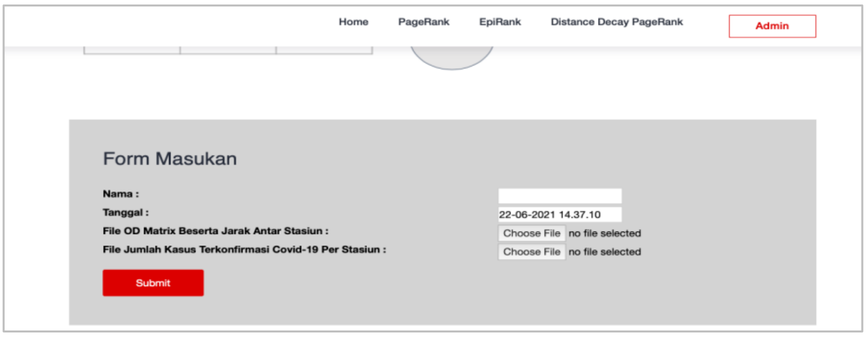

Fig. 13. The form page for uploading input data, Form Masukan (Entry Form), Nama (Name), File OD Matrix Beserta Jarak Antar Stasiun (OD Matrix File with Distance Between Stations), File Jumlah Kasus Terkonfirmasi Covid-19 Per Stasiun (File Number of Covid-19 Confirmed Cases for Each Station)

Halaman Hasil Prediksi Menggunakan Algoritme Distance Decay PageRank

Nama Prediksi : coba sidang 1

Peringkat	Stasiun	Skor DDPR	Prediksi Status
1	Surabaya Gubeng	0,144	Risiko Tinggi
2	Wonokromo	0,099	Risiko Tinggi
3	Malang	0,09	Risiko Tinggi
4	Sidoarjo	0,071	Risiko Sedang
5	Bangil	0,071	Risiko Sedang
6	Malang Kota Lama	0,062	Risiko Sedang
7	Surabaya Pasar Turi	0,058	Risiko Rendah
8	Sepanjang	0,056	Risiko Rendah
9	Mojokerto	0,056	Risiko Rendah
10	Jombang	0,05	Risiko Rendah
11	Kertosono	0,046	Risiko Rendah

Fig. 14. The DDRR modeling results page, Halaman Hasil Prediksi Menggunakan Algoritme Distance Decay PageRank (Prediction Result Page Using Distance Decay PageRank Algorithm), Nama Prediksi (Prediction Name), Peringkat (Ratings), Stasiun (Station), Skor DDPR (DDPR Score), Prediksi Status (Status Prediction), Resiko Tinggi (High Risk), Resiko Sedang (Medium Risk), Resiko Rendah (Low Risk)

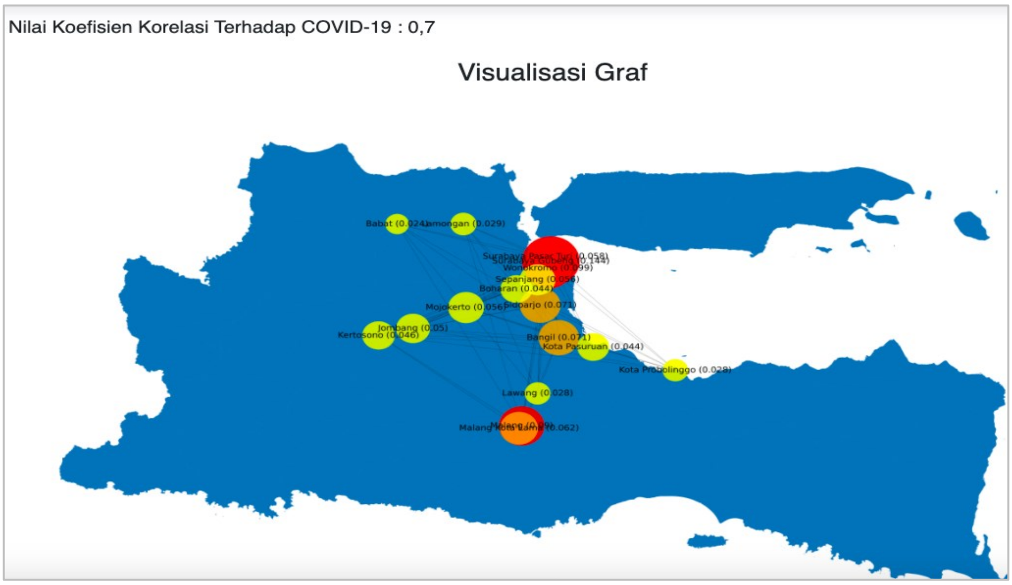

Fig. 15. Display of the visualization page of the DDRR modeling results, Nilai Koefisien Korelassi Terhadap COVID-19 : 0,7 (Correlation Coefficient Value Against COVID-19: 0.7), Visualisasi Graf (Graph Visualization)

4. CONCLUSION

Author This study has modeled human movement using a graph formed based on the distance between train stations and the distribution of the journey. This study implements the Distance-Decay Page rank algorithm to identify human movement on rail transportation in the East Java Province, Indonesia. The results reveal several suitable cases and some anomalies or results that are not in accordance with the actual case. For example, Surabaya Gubeng Station has the highest DDPR score because it has a low average indegree distance and a high average number of passengers. However, Malang Station, with a fairly high average distance or is far from other stations, has a high DDRR score.

In addition, this study also shows that the correlation between human movement and the spread of COVID-19 in East Java is high, with a correlation coefficient (r) = 0.7. The more people travel via rail transportation, the greater the spread of COVID-19 in the area. The coefficient of determination shows that human movement affects the spread of COVID-19 by 49% in East Java. In other words, there are other factors (51%) that may affect the spread of COVID-19 in East Java.

It can be concluded that 51% of other factors influence the spread of COVID-19 in East Java besides the distance factor. Therefore, it is recommended that further researchers expand this research by adding several other parameters to the calculation of the DDPR score per station. Parameters such as travel time, number of passengers, and other parameters on rail transportation may cause the spread of COVID-19 in East Java. Finally, it is recommended that the East Java Provincial government continue to limit human movement, especially in rail transportation, to reduce the spread of COVID-19 in East Java Province, such as by reducing the number of train trips continuing to apply health protocols on every trip.

REFERENCES

[1] K. K. Agustiningsih, E. Utami, and O. M. A. Alsyaibani, "Sentiment Analysis and Topic Modelling of The COVID-19 Vaccine in Indonesia on Twitter Social Media Using Word Embedding," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 8, no. 1, pp. 64–75, 2022, https://doi.org/10.26555/JITEKI.V8I1.23009.

[2] H. Ouassou *et al.*, "The Pathogenesis of Coronavirus Disease 2019 (COVID-19): Evaluation and Prevention," *J Immunol Res*, pp. 1–7, 2020, https://doi.org/10.1155/2020/1357983.

[3] M. Khan *et al.*, "COVID-19: A Global Challenge with Old History, Epidemiology and Progress So Far," *Molecules*, vol. 26, no. 1, p. 39, 2020, https://doi.org/10.3390/molecules26010039.

[4] M. Yamin, "Counting the cost of COVID-19," *International Journal of Information Technology*, vol. 12, no. 2, pp. 311–317, 2020, https://doi.org/10.1007/s41870-020-00466-0.

[5] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus outbreak of global health concern," *The Lancet*, vol. 395, no. 10223, pp. 470–473, 2020, https://doi.org/10.1016/S0140-6736(20)30185-9.

[6] D. Kang, H. Choi, J. H. Kim, and J. Choi, "Spatial epidemic dynamics of the COVID-19 outbreak in China," *International Journal of Infectious Diseases*, vol. 94, pp. 96–102, 2020, https://doi.org/10.1016/J.IJID.2020.03.076.

[7] C. Sohrabi *et al.*, "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *International Journal of Surgery*, vol. 76, pp. 71–76, 2020, https://doi.org/10.1016/j.ijsu.2020.02.034.
[8] I. Chakraborty and P. Maity, "COVID-19 outbreak: Migration, effects on society, global environment and prevention," *Science of The Total Environment*, vol. 728, p. 138882, 2020, https://doi.org/10.1016/j.scitotenv.2020.138882.
[9] M. N. Kamel Boulos and E. M. Geraghty, "Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbr," *Int J Health Geogr*, vol. 19, no. 1, p. 8, 2020, https://doi.org/10.1186/s12942-020-00202-8.
[10] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J Pediatr*, vol. 87, no. 4, pp. 281–286, 2020, https://doi.org/10.1007/S12098-020-03263-6.
[11] A. Hafeez, S. Ahmad, S. Siddqui, M. Ahmad, S. M.- Ejmo, and undefined 2020, "A review of COVID-19 (Coronavirus Disease-2019) diagnosis, treatments and prevention," *aarc.gov.et*, 2019, https://doi.org/10.14744/ejmo.2020.90853.
[12] A. F. Lubis and B. Basari, "Development of Proximity-Based COVID-19 Contact Tracing System Devices for Locally Virus Spread Prevention," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 8, no. 1, pp. 144–156, 2022, https://doi.org/10.26555/JITEKI.V8I1.23697.
[13] E. L. Anderson, P. Turnham, J. R. Griffin, and C. C. Clarke, "Consideration of the Aerosol Transmission for COVID-19 and Public Health," *Risk Analysis*, vol. 40, no. 5, pp. 902–907, 2020, https://doi.org/10.1111/RISA.13500.
[14] M. E. Atmojo and H. D. Fridayani, "An Assessment of Covid-19 Pandemic Impact on Indonesian Tourism Sector," *Journal of Governance and Public Policy*, vol. 8, no. 1, pp. 1–9, 2021, https://doi.org/10.18196/jgpp.811338.
[15] S. Olivia, J. Gibson, and R. Nasrudin, "Indonesia in the Time of Covid-19," *Bull Indones Econ Stud*, vol. 56, no. 2, pp. 143–174, 2020, https://doi.org/10.1080/00074918.2020.1798581.
[16] Md. M. Alam, A. M. Fawzi, Md. M. Islam, and J. Said, "Impacts of COVID-19 pandemic on national security issues: Indonesia as a case study," *Security Journal*, vol. 35, no. 4, pp. 1067–1086, 2022, https://doi.org/10.1057/s41284-021-00314-1.
[17] A. Asyary and M. Veruswati, "Sunlight exposure increased Covid-19 recovery rates: A study in the central pandemic area of Indonesia," *Science of The Total Environment*, vol. 729, p. 139016, 2020, https://doi.org/10.1016/j.scitotenv.2020.139016.
[18] T. Eryando, T. Sipahutar, and S. Rahardiantoro, "The Risk Distribution of COVID-19 in Indonesia: A Spatial Analysis," *Asia Pac J Public Health*, vol. 32, no. 8, pp. 450–452, 2020, https://doi.org/10.1177/1010539520962940.
[19] B. Adams and D. D. Kapan, "Man bites mosquito: Understanding the contribution of human movement to vector-borne disease dynamics," *PLoS One*, vol. 4, no. 8, 2009, https://doi.org/10.1371/JOURNAL.PONE.0006763.
[20] S. T. Stoddard *et al.*, "The role of human movement in the transmission of vector-borne pathogens," *PLoS Negl Trop Dis*, vol. 3, no. 7, 2009, https://doi.org/10.1371/JOURNAL.PNTD.0000481.
[21] C. Y. Huang, W. C. B. Chin, T. H. Wen, Y. H. Fu, and Y. S. Tsai, "EpiRank: Modeling Bidirectional Disease Spread in Asymmetric Commuting Networks," *Scientific Reports*, vol. 9, no. 1, pp. 1–15, 2019, https://doi.org/10.1038/s41598-019-41719-8.
[22] J. S. Jia, X. Lu, Y. Yuan, G. Xu, J. Jia, and N. A. Christakis, "Population flow drives spatio-temporal distribution of COVID-19 in China," *Nature*, vol. 582, no. 7812, pp. 389–394, 2020, https://doi.org/10.1038/s41586-020-2284-y.
[23] C. Deniz and G. Gökmen, "A New Robotic Application for COVID-19 Specimen Collection Process," *Journal of Robotics and Control (JRC)*, vol. 3, no. 1, pp. 73–77, 2022, https://doi.org/10.18196/JRC.V3I1.11659.
[24] A. S. Fotheringham, "Spatial structure and distance-decay parameters," *Annals of the Association of American Geographers*, vol. 71, no. 3, pp. 425–436, 1981, https://doi.org/10.1111/J.1467-8306.1981.TB01367.X.
[25] C. R. Browning, N. P. Pinchak, and C. A. Calder, "Human Mobility and Crime: Theoretical Approaches and Novel Data Collection Strategies," *Annu Rev Criminol*, vol. 4, no. 1, pp. 99–123, 2021, https://doi.org/10.1146/annurev-criminol-061020-021551.
[26] P. J. Taylor, "Distance Decay Models in Spatial Interactions (Concepts and techniques in modern geography)," p. 35, 1975, Accessed: Jan. 17, 2023. [Online]. Available: http://www.amazon.com/Distance-Interactions-Concepts-techniques-geography/dp/0902246445.
[27] Z. C. Yin, Z. H. N. Jin, S. Ying, H. Liu, S. J. Li, and J. Q. Xiao, "Distance-Decay Effect in Probabilistic Time Geography for Random Encounter," *ISPRS International Journal of Geo-Information 2019, Vol. 8, Page 177*, vol. 8, no. 4, p. 177, 2019, https://doi.org/10.3390/IJGI8040177.
[28] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web.," *Proceedings of ASIS*, vol. 98, 161-172, 1998, https://cir.nii.ac.jp/crid/1571417125844972032.
[29] B. Jiang, "Ranking spaces for predicting human movement in an urban environment," *International Journal of Geographical Information Science*, vol. 23, no. 7, pp. 823–837, 2009, https://doi.org/10.1080/13658810802022822.
[30] R. Fitriani, I. J.-Commun. Math. Biol. Neurosci., and undefined 2020, "Spatial modeling of confirmed COVID-19 pandemic in East Java province by geographically weighted negative binomial regression," *scik.org*, Accessed: Jan. 17, 2023. [Online]. Available: http://scik.org/index.php/cmbn/article/view/4874.
[31] D. M. Harfina, Z. Zaini, and W. J. Wulung, "Disinfectant Spraying System with Quadcopter Type Unmanned Aerial Vehicle (UAV) Technology as an Effort to Break the Chain of the COVID-19 Virus," *Journal of Robotics and Control (JRC)*, vol. 2, no. 6, pp. 502–507, 2021, https://doi.org/10.18196/JRC.26129.

[32] W.-C.-B. Chin and T.-H. Wen, "Geographically Modified PageRank Algorithms: Identifying the Spatial Concentration of Human Movement in a Geospatial Network," *PLoS One*, vol. 10, no. 10, p. e0139509, 2015, https://doi.org/10.1371/journal.pone.0139509.

[33] L. Zhang, P. Chen, and F. Hui, "Refining the accessibility evaluation of urban green spaces with multiple sources of mobility data: A case study in Shenzhen, China," *Urban For Urban Green*, vol. 70, p. 127550, 2022, https://doi.org/10.1016/j.ufug.2022.127550.

[34] J. Wang, D. Du, and J. Huang, "Inter-city connections in China: High-speed train vs. inter-city coach," *J Transp Geogr*, vol. 82, p. 102619, 2020, https://doi.org/10.1016/j.jtrangeo.2019.102619.

[35] J. Day, "Mapping the cultural divides of England and Wales: Did the geographies of 'Belonging' act as a brake on British Urbanisation, 1851–1911?," *PLoS One*, vol. 18, no. 5, p. e0286244, 2023, https://doi.org/10.1371/journal.pone.0286244.

[36] W. C. B. Chin and T. H. Wen, "Geographically modified PageRank algorithms: Identifying the spatial concentration of human movement in a geospatial network," *PLoS One*, vol. 10, no. 10, 2015, https://doi.org/10.1371/JOURNAL.PONE.0139509.

[37] I. Ghozali, "Aplikasi analisis multivariate dengan program IBM SPSS 25," *Badan Penerbit Universitas Diponegoro*, 2018, Accessed: Jan. 17, 2023. [Online]. Available: http://repo.unikadelasalle.ac.id/index.php?p=show_detail&id=13099.

[38] K. M. Sudar, P. Nagaraj, S. Nithisaa, R. Aishwarya, M. Aakash, and S. I. Lakshmi, "Alzheimer's Disease Analysis using Explainable Artificial Intelligence (XAI)," in *International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, pp. 419–423, 2022, https://doi.org/10.1109/ICSCDS53736.2022.9760858.

[39] S. Abdel-Salam and A. Rafea, "Performance Study on Extractive Text Summarization Using BERT Models," *Information*, vol. 13, no. 2, p. 67, 2022, https://doi.org/10.3390/info13020067.

[40] A. Del Sole, "Customizing Visual Studio Code," in *Visual Studio Code Distilled*, pp. 103–125, 2023, https://doi.org/10.1007/978-1-4842-9484-0_5.

[41] S. bin Uzayr, *Mastering Visual Studio Code*, no. 15018. Boca Raton: CRC Press, 2022. https://doi.org/10.1201/9781003311973.