Journal of Computers for Society 5(2) (2024) 109-132

Journal of Computers for Society

Journal homepage: https://ejournal.upi.edu/index.php/JCS

Correlation Analysis of Open Street Map, Demography, and Vaccination on the Number of Covid-19 Cases Using Multiple Linear Regression and Pearson Correlation Product Moment

*Aqhbar Habib, Erna Piantari, Lala Septem Riza**

Department of Computer Science Education, Universitas Pendidikan Indonesia, Indonesia
*Correspondence: E-mail: lala.s.riza@upi.edu

ABSTRACT

At the beginning of 2020, the world was shocked by the spread of Coronavirus Disease 2019 (Covid-19). The resulting losses cover various areas. This research aims to analyze the correlation between spatial data, demographic data, and vaccination data on the spread of Covid-19 in Bandung City using Multiple Linear Regression (MLR) and Pearson Correlation Product Moment (Pearson's r). The results show that there are only 3 variables that are significantly correlated with Covid-19 cases. The lowest variables are Residential, Population Density, and Healthy Homes. Has a significant simultaneous correlation with Covid-19 cases with a coefficient of determination (R^2) of 0.55404. The model built also passed the 3 Classical Assumptions test so that the results can be trusted for their level of truth and feasibility. The results of experiments using the Pearson's r model involving 5 vaccination periods show that out of 30 sub-districts in Bandung City, there are 20 sub-districts that have a significant correlation between vaccination and the addition of Covid-19 cases and have a negative correlation direction of 80.54%. The results of the Pearson's r model experiment involving 6 vaccination periods show that there are 9 sub-districts that have a relationship. With a negative correlation direction of 72.93%.

ARTICLE INFO

Article History:
Submitted/Received 19 Feb 2024
First Revised 28 Apr 2024
Accepted 06 Jun 2024
First Available Online 07 Jun 2024
Publication Date 15 Sep 2024

Keyword:
Arcgis,
Covid-19,
Multiple linear regression,
Open street map,
Pearson correlation product moment.

1. INTRODUCTION

The world is currently on high alert due to the spread of a deadly virus. The 2019 coronavirus (Covid-19) is a highly dangerous infectious disease that first appeared in Wuhan, the capital of China's Hubei Province, in 2019 and has since spread globally. Coronaviruses belong to the Orthocronavirinae subfamily within the Coronaviridae family and the Nidovirales order. These viruses can infect birds and mammals, including humans, causing respiratory infections that range from mild colds to severe illnesses such as SARS, MERS, and Covid-19. On March 11, 2020, the World Health Organization (WHO) declared Covid-19 a pandemic, signifying its rapid spread worldwide and the inability of any country to guarantee complete protection from the virus. Transmission occurs easily through contact with infected individuals. In Indonesia, the Covid-19 pandemic has significantly impacted not only the health sector but also the economy and education.

The economic sector in Indonesia has been significantly impacted by the Covid-19 pandemic, particularly in the trade and tourism industries. Over 1.5 million workers were either laid off or dismissed, 12,703 flights were canceled between January and March 2020, and daily tourist visits plummeted by up to 6,800. In the education sector, the government encouraged people to work, study, and worship from home to curb the spread of Covid-19. However, online learning presented numerous challenges. Not all children could participate due to lack of smartphones, internet access, or technological literacy. Additionally, the country lacked a culture of distance learning, having relied predominantly on face-to-face education. These issues highlight that online learning has not been fully effective in Indonesia. Addressing this problem is crucial, and understanding the factors that influence the spread of Covid-19 in Indonesia is vital. Identifying these factors can help the government and relevant authorities prioritize and tailor their responses to the pandemic across different regions. By focusing on these factors, effective and timely policy decisions can be made to break the chain of Covid-19 transmission.

Numerous studies have been conducted worldwide to analyze the spread of infectious diseases. One study examined the spatial relationship between Covid-19 spread and influencing factors in Iran using the Spatial Index Moran I and LISA (Local Indicator of Spatial Association) methods (Athena *et al*., 2020; Dewi, 2020; Ramirez-aldana *et al*., 2020). The research identified that provinces with higher Covid-19 cases typically had more urbanization, an aging population, a higher number of physicians, efficient communications, and higher average temperatures, whereas provinces with higher literacy rates had fewer cases. Another study Haq *et al*. (2019) conducted a spatial analysis of pulmonary tuberculosis in Pariaman, Bukittinggi, and Dumai from 2010-2016. This study aimed to determine the correlation between regional altitude, population density, and the prevalence of healthy homes with the proportion of positive acid-fast bacilli (BTA) pulmonary TB cases. Researchers used the Spearman Correlation test for bivariate statistical analysis and overlay techniques for spatial analysis. Additionally, a study Han *et al*. (2021) focused on the spatial analysis of Covid-19 distribution patterns and environmental factors in Beijing, China, using Moran's I Index, Geographically Weighted Regression (GWR), and the Spearman Correlation Coefficient. It found that population density and proximity to the Xinfadi market (a Covid-19 hotspot) were the most critical factors influencing the pandemic's spread. These findings provide valuable insights for the global effort to combat Covid-19.

The Covid-19 phenomenon and previous research have prompted the author to conduct a study related to Covid-19. This research includes several main stages: collecting Open Street Map (OSM) data (Haklay & Weber, 2008) using ArcGIS software (Yagoub, 2017), normalizing

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

the OSM data, grouping and weighting the OSM data, and combining the OSM data with demographic data and Covid-19 case data to create dataset I. Additionally, Covid-19 vaccination data will be combined with Covid-19 case data to form dataset II. The correlation relationships will be analyzed using Multiple Linear Regression (MLR) (Hall *et al.*, 2011) and Pearson Correlation Product Moment (Pearson's r) (Chee, 2015). The study aims to identify environmental, demographic, and vaccination factors that correlate with the spread of Covid-19, understand the relationships between these variables, and determine the direction of these relationships.

2. METHODS

Researchers create a computational model design that will be used as a guideline in conducting research. The computational model used can be seen in **Figure 1**.

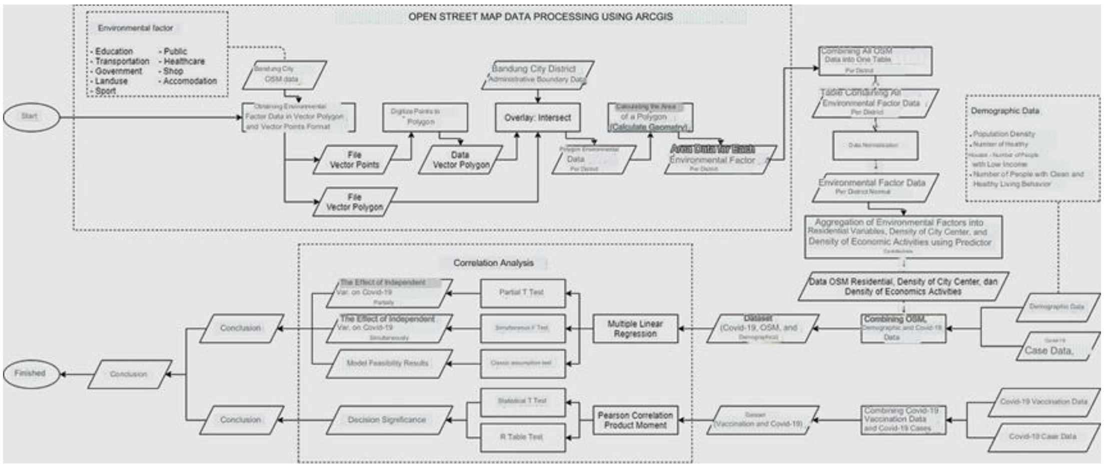


Figure 1. Computational Model for Analysis of OSM Data, Demography, and Covid-19 Vaccination Against the Spread of Covid-19.

2.1. OSM Data Processing Using ArcGIS

The first stage is the Open Street Map (OSM) data collection stage which consists of four steps. At this stage, OSM data is processed using ArcGIS software to then calculate the area of features (environmental factors) in each sub-district in Bandung City.

2.2. Acquisition of Environmental Factor Data (OSM Data)

Data on the surface features of Bandung City can be obtained from the Open Street Map (OSM). OSM data can be accessed through the Hot Export Tool website by first registering an account. After registering an account, users can immediately access the required OSM data, such as data about Buildings, Commercial, LandUse, and many other features that can be accessed. In this study, the author downloaded data with the scope of Bandung City with the last update time on April 7, 2021. The total data downloaded is as many as 27 files that will be processed using the ArcGIS image analysis application (Yagoub, 2017).

2.3. Digitasi Shapefile Vector Point to Polygon

In general, OSM environmental factor data downloaded on Hot Export Tool has 2 types of shapefile data, namely shapefile vector point data and polygon vector shapefile data. These two data cannot be separated from each other because both data complement each other. Because the calculation of the area of environmental factors can only be calculated if the data

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

is in polygon vector shapefile format, the shapefile point format data must first be digitized to vector polygon form.

Digitizing vectors point to polygon shapefiles can be done using ArcGIS software (Yagoub, 2017). To digitize vector point data to vector polygon, the author must first create a new shapefile with polygon type, then the author adjusts the location and name of the vector point attribute to a new shapefile through the OSM basemap to create features according to the attribute data of the vector point data. Do that process for all vector points attributes until all attributes are digitized to a new polygon shapefile. **Figure 2** is data before and after digitization from vector points to vector polygons.

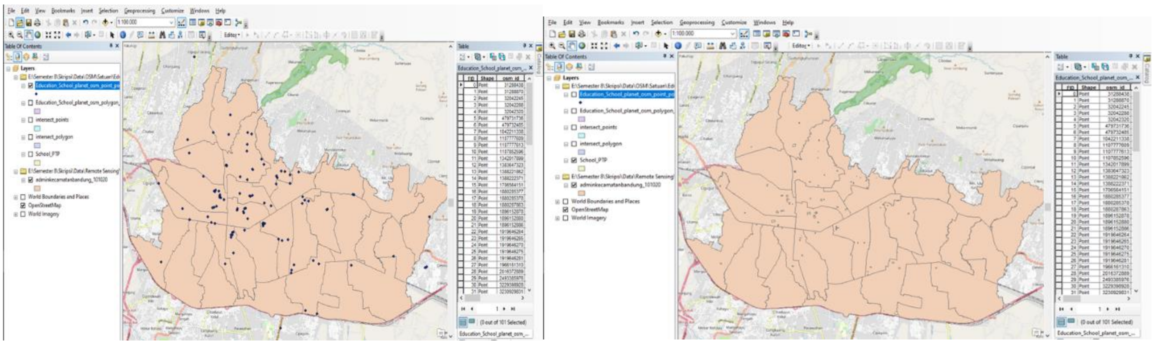
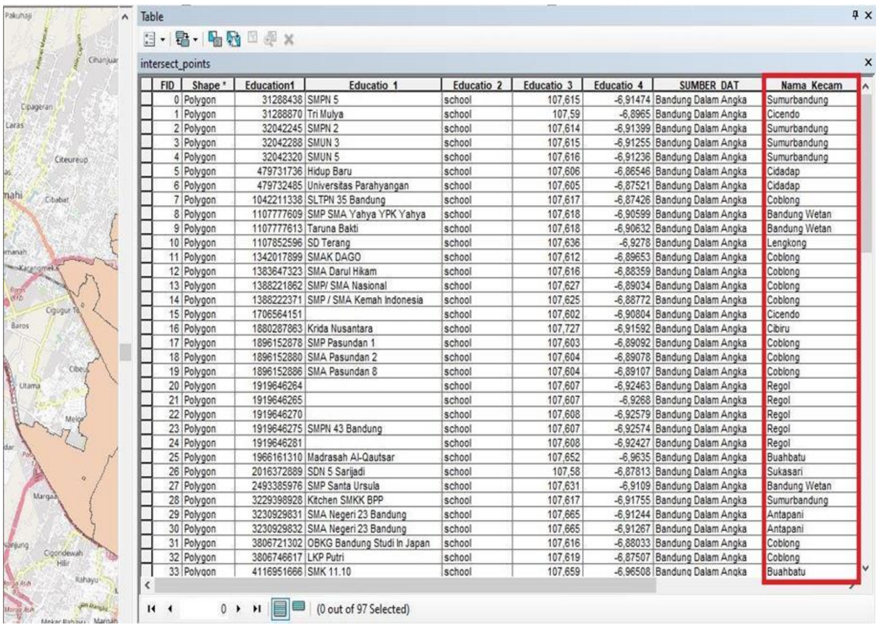


Figure 2. Data before and after digitization to vector polygons

2.4. Overlay: Intersect

After digitizing the point vector shapefile data into a polygon vector shapefile for 27 environmental factors from OSM, the next step is to group the features in the shapefile into each sub-district using the Overlay Intersect technique in ArcGIS software (Yagoub, 2017). This technique is used on Bandung City Government Boundary shp data which has been downloaded from the Indonesian Geospatial Portal page and OSM data which has been digitized on vector point data. The Intersect Overlay process in ArcGIS can be done using the intersect tool on the Geoprocessing menu toolbar by selecting shp format data for the Bandung City Government area boundaries and digitized OSM data (see **Figure 3**).



Table

intersect_points

FID	Shape *	Education1	Educatio 1	Educatio 2	Educatio 3	Educatio 4	SUMBER DAT	Nama Kecam
0	Polygon	31288438	SMPN 5	school	107,615	-6,91474	Bandung Dalam Angka	Sumurbandung
1	Polygon	31288870	Tri Mulya	school	107,59	-6,8965	Bandung Dalam Angka	Cicendo
2	Polygon	32042245	SMPN 2	school	107,614	-6,91399	Bandung Dalam Angka	Sumurbandung
3	Polygon	32042288	SMUN 3	school	107,615	-6,91255	Bandung Dalam Angka	Sumurbandung
4	Polygon	32042320	SMUN 5	school	107,616	-6,91236	Bandung Dalam Angka	Sumurbandung
5	Polygon	479731736	Hidup Baru	school	107,606	-6,86546	Bandung Dalam Angka	Cidadap
6	Polygon	479732485	Universitas Parahyangan	school	107,605	-6,87521	Bandung Dalam Angka	Cidadap
7	Polygon	1042211338	SLTPN 35 Bandung	school	107,617	-6,87426	Bandung Dalam Angka	Coblong
8	Polygon	1107777609	SMP SMA Yahya YPK Yahya	school	107,618	-6,90599	Bandung Dalam Angka	Bandung Wetan
9	Polygon	1107777613	Taruna Bakti	school	107,618	-6,90632	Bandung Dalam Angka	Bandung Wetan
10	Polygon	1107852596	SD Terang	school	107,636	-6,9278	Bandung Dalam Angka	Lengkong
11	Polygon	1342017899	SMAK DAGO	school	107,612	-6,89653	Bandung Dalam Angka	Coblong
12	Polygon	1383647323	SMA Darul Hikam	school	107,616	-6,88359	Bandung Dalam Angka	Coblong
13	Polygon	1388221862	SMP/ SMA Nasional	school	107,627	-6,89034	Bandung Dalam Angka	Coblong
14	Polygon	1388222371	SMP / SMA Kemah Indonesia	school	107,625	-6,88772	Bandung Dalam Angka	Coblong
15	Polygon	1706564151		school	107,602	-6,90804	Bandung Dalam Angka	Cicendo
16	Polygon	1880287863	Krida Nusantara	school	107,727	-6,91592	Bandung Dalam Angka	Cibiru
17	Polygon	1896152878	SMP Pasundan 1	school	107,603	-6,89092	Bandung Dalam Angka	Coblong
18	Polygon	1896152880	SMA Pasundan 2	school	107,604	-6,89078	Bandung Dalam Angka	Coblong
19	Polygon	1896152886	SMA Pasundan 8	school	107,604	-6,89107	Bandung Dalam Angka	Coblong
20	Polygon	1919646264		school	107,607	-6,92463	Bandung Dalam Angka	Regol
21	Polygon	1919646265		school	107,607	-6,9268	Bandung Dalam Angka	Regol
22	Polygon	1919646270		school	107,608	-6,92579	Bandung Dalam Angka	Regol
23	Polygon	1919646275	SMPN 43 Bandung	school	107,607	-6,92574	Bandung Dalam Angka	Regol
24	Polygon	1919646281		school	107,608	-6,92427	Bandung Dalam Angka	Regol
25	Polygon	1966161310	Madrasah Al-Qautsar	school	107,652	-6,9635	Bandung Dalam Angka	Buahbatu
26	Polygon	2016372889	SDN 5 Sarijadi	school	107,58	-6,87813	Bandung Dalam Angka	Sukasari
27	Polygon	2493385976	SMP Santa Ursula	school	107,631	-6,9109	Bandung Dalam Angka	Bandung Wetan
28	Polygon	3229398928	Kitchen SMKK BPP	school	107,617	-6,91755	Bandung Dalam Angka	Sumurbandung
29	Polygon	3230929831	SMA Negeri 23 Bandung	school	107,665	-6,91244	Bandung Dalam Angka	Antapani
30	Polygon	3230929832	SMA Negeri 23 Bandung	school	107,665	-6,91267	Bandung Dalam Angka	Antapani
31	Polygon	3806721302	OBKG Bandung Studi In Japan	school	107,616	-6,88033	Bandung Dalam Angka	Coblong
32	Polygon	3806746617	LKP Putri	school	107,619	-6,87507	Bandung Dalam Angka	Coblong
33	Polygon	4116951666	SMK 11.10	school	107,659	-6,96508	Bandung Dalam Angka	Buahbatu

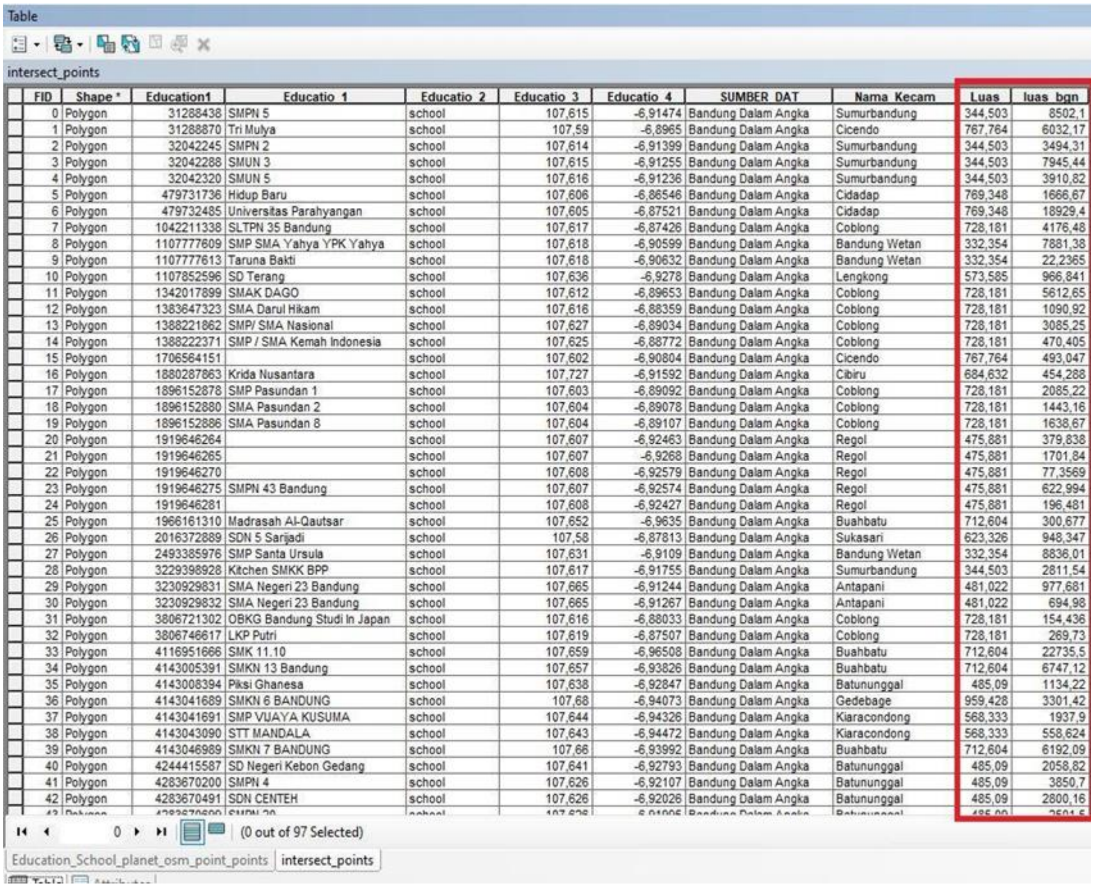
(0 out of 97 Selected)

Figure 3. Intersect overlay result data attribute.

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

2.5. Calculate Geometry

After grouping each attribute of the sub-district OSM data, the next step is to calculate the area of each attribute using ArcGIS. Calculate Geometry Tools can be used to calculate the area of polygon objects from vector polygon shapefile data, this tool can calculate the area in hectares (ha), square meters (m^2), and other units of area (see **Figure 4**).



FID	Shape *	Education1	Educatio 1	Educatio 2	Educatio 3	Educatio 4	SUMBER DAT	Nama Kecam	Luas	luas bgn
0	Polygon	31288438	SMPN 5	school	107,615	-6,91474	Bandung Dalam Angka	Sumurbandung	344,503	8502,1
1	Polygon	31288870	Tri Mulya	school	107,59	-6,8965	Bandung Dalam Angka	Cicendo	767,764	6032,17
2	Polygon	32042245	SMPN 2	school	107,614	-6,91399	Bandung Dalam Angka	Sumurbandung	344,503	3494,31
3	Polygon	32042288	SMUN 3	school	107,615	-6,91255	Bandung Dalam Angka	Sumurbandung	344,503	7945,44
4	Polygon	32042320	SMUN 5	school	107,616	-6,91236	Bandung Dalam Angka	Sumurbandung	344,503	3910,82
5	Polygon	479731736	Hidup Baru	school	107,606	-6,86546	Bandung Dalam Angka	Cidadap	769,348	1666,67
6	Polygon	479732485	Universitas Parahyangan	school	107,605	-6,87521	Bandung Dalam Angka	Cidadap	769,348	18929,4
7	Polygon	1042211338	SLTPN 35 Bandung	school	107,617	-6,87426	Bandung Dalam Angka	Coblong	728,181	4176,48
8	Polygon	1107777609	SMP SMA Yahya YPK Yahya	school	107,618	-6,90599	Bandung Dalam Angka	Bandung Wetan	332,354	7881,38
9	Polygon	1107777613	Taruna Bakti	school	107,618	-6,90632	Bandung Dalam Angka	Bandung Wetan	332,354	22,2365
10	Polygon	1107852596	SD Terang	school	107,636	-6,9278	Bandung Dalam Angka	Lengkong	573,585	966,841
11	Polygon	1342017899	SMAK DAGO	school	107,612	-6,89653	Bandung Dalam Angka	Coblong	728,181	5612,65
12	Polygon	1383647323	SMA Darul Hikam	school	107,616	-6,88359	Bandung Dalam Angka	Coblong	728,181	1090,92
13	Polygon	1388221862	SMP/ SMA Nasional	school	107,627	-6,89034	Bandung Dalam Angka	Coblong	728,181	3085,25
14	Polygon	1388222371	SMP / SMA Kemah Indonesia	school	107,625	-6,88772	Bandung Dalam Angka	Coblong	728,181	470,405
15	Polygon	1706564151		school	107,602	-6,90804	Bandung Dalam Angka	Cicendo	767,764	493,047
16	Polygon	1880287863	Krida Nusantara	school	107,727	-6,91592	Bandung Dalam Angka	Cibiru	684,632	454,288
17	Polygon	1896152878	SMP Pasundan 1	school	107,603	-6,89092	Bandung Dalam Angka	Coblong	728,181	2085,22
18	Polygon	1896152880	SMA Pasundan 2	school	107,604	-6,89078	Bandung Dalam Angka	Coblong	728,181	1443,16
19	Polygon	1896152886	SMA Pasundan 8	school	107,604	-6,89107	Bandung Dalam Angka	Coblong	728,181	1638,67
20	Polygon	1919646264		school	107,607	-6,92463	Bandung Dalam Angka	Regol	475,881	379,838
21	Polygon	1919646265		school	107,607	-6,9268	Bandung Dalam Angka	Regol	475,881	1701,84
22	Polygon	1919646270		school	107,608	-6,92579	Bandung Dalam Angka	Regol	475,881	77,3569
23	Polygon	1919646275	SMPN 43 Bandung	school	107,607	-6,92574	Bandung Dalam Angka	Regol	475,881	622,994
24	Polygon	1919646281		school	107,608	-6,92427	Bandung Dalam Angka	Regol	475,881	196,481
25	Polygon	1966161310	Madrasah Al-Qautsar	school	107,652	-6,9635	Bandung Dalam Angka	Buahbatu	712,604	300,677
26	Polygon	2016372889	SDN 5 Sarijadi	school	107,58	-6,87813	Bandung Dalam Angka	Sukasari	623,326	948,347
27	Polygon	2493385976	SMP Santa Ursula	school	107,631	-6,9109	Bandung Dalam Angka	Bandung Wetan	332,354	8836,01
28	Polygon	3229398928	Kitchen SMKK BPP	school	107,617	-6,91755	Bandung Dalam Angka	Sumurbandung	344,503	2811,54
29	Polygon	3230929831	SMA Negeri 23 Bandung	school	107,665	-6,91244	Bandung Dalam Angka	Antapani	481,022	977,681
30	Polygon	3230929832	SMA Negeri 23 Bandung	school	107,665	-6,91267	Bandung Dalam Angka	Antapani	481,022	694,98
31	Polygon	3806721302	OBKG Bandung Studi In Japan	school	107,616	-6,88033	Bandung Dalam Angka	Coblong	728,181	154,436
32	Polygon	3806746617	LKP Putri	school	107,619	-6,87507	Bandung Dalam Angka	Coblong	728,181	269,73
33	Polygon	4116951666	SMK 11.10	school	107,659	-6,96508	Bandung Dalam Angka	Buahbatu	712,604	22735,5
34	Polygon	4143005391	SMKN 13 Bandung	school	107,657	-6,93826	Bandung Dalam Angka	Buahbatu	712,604	6747,12
35	Polygon	4143008394	Piksi Ghanesa	school	107,638	-6,92847	Bandung Dalam Angka	Batununggal	485,09	1134,22
36	Polygon	4143041689	SMKN 6 BANDUNG	school	107,68	-6,94073	Bandung Dalam Angka	Gedebage	959,428	3301,42
37	Polygon	4143041691	SMP VIJAYA KUSUMA	school	107,644	-6,94326	Bandung Dalam Angka	Kiaracondong	568,333	1937,9
38	Polygon	4143043090	STT MANDALA	school	107,643	-6,94472	Bandung Dalam Angka	Kiaracondong	568,333	558,624
39	Polygon	4143046989	SMKN 7 BANDUNG	school	107,66	-6,93992	Bandung Dalam Angka	Buahbatu	712,604	6192,09
40	Polygon	4244415587	SD Negeri Kebon Gedang	school	107,641	-6,92793	Bandung Dalam Angka	Batununggal	485,09	2058,82
41	Polygon	4283670200	SMPN 4	school	107,626	-6,92107	Bandung Dalam Angka	Batununggal	485,09	3850,7
42	Polygon	4283670491	SDN CENTEH	school	107,626	-6,92026	Bandung Dalam Angka	Batununggal	485,09	2800,16

Figure 4. Data attributes of the calculate geometry process.

2.6. Combining all OSM data per sub-district into one table

After each of the 27 OSM data is processed and obtained the cumulative area for each sub-district in Bandung City, the next step is to combine all these data into one unit in the form of a table. The unification of OSM data environmental features is done to facilitate data management at a later stage.

2.7. OSM Data Normalization

After each of the 27 processed OSM data is combined in one unit in the form of a table, the next step is to normalize the data in the table by dividing it by the area of each sub-district so that OSM data (environmental factors) will be obtained with a range of values between 0-1. The units for the area of environmental factors of the sub-district and the area of the sub-district use units of hectares (ha).

2.8. Aggregation of OSM data into 3 variables

After data normalization is carried out for each of the 27 OSM data, the next step is to perform data aggregation. From 27 OSM data downloaded from the Hot Export Tool, it will be aggregated into 3 research variables only, namely Residential (housing), Density of City

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

Center (city center density), and Density of Economics Activities (density of economic activities) (see **Table 1**).

Table 1. OSM data grouping.

No	Variable Name	Data OSM
1	Residential	Residential
2	Density of City Center	Guest House
		Hostel
		Motel
		Hotel
		University
		College
		School
		Kindergarten
		Stadium
		Pitch
		Swimming Pool
		Airport
		Bus Terminal & Train Station
		Park
		Hospital
		Clinic
		Pharmacy
		Government Office
		Military
		Library
		Community Centre
		Place Of Worship
3	Density of Economics Activities	Mall
		Traditional Market
		Convenience
		Supermarket

2.9. Combining all OSM data per sub-district into one table

The aggregation method used to obtain the value of the 3 variables in this study is the predictor contribution method consisting of effective contribution (SE) and relative contribution (SR). to calculate SE and SR can be seen in the formula **Equation 1** and **Equation 2**:

$$\text{SE(X)\%} = \text{Beta}_X \text{ x } r_{xy} \text{ x } 100\% \quad (1)$$

$$\text{SR(X)\%} = \frac{\text{XE(X)}}{R^2} \text{ x } 100\% \quad (2)$$

Description:
SE = Effective Donation
SR = Relative Contribution
Beta_X = Xth Beta Regression Coefficient
r_{xy} = Correlation coefficient
R^2 = Coefficient of Determination

After the weighting of the variables Density of City Center and Density of Economics Activities is carried out, the next step is to multiply the value of the Relative Contribution (SR) weight by the normalized value of OSM sample data. The result of this process is to get 3 variables with their respective values. The Residential variable is filled directly with the value of LandUse_Residential data, the value of the Density of City Center and Density of Economics Activities variables is filled with the final value of the sum of the weight values multiplied by the value of the sample data for each data respectively.

2.10. Combining OSM, Demographic, and Covid-19 Case Data

Data collected from the OSM data collection stage, Demographic data, and data on the Number of Covid-19 Cases in Bandung City were collected in one file with Comma-separated values (.csv) format as dataset I. This dataset I will be further processed in model design using Multiple Linear Regression (MLR) to analyze the correlation between independent and dependent variables. The total data that has been collected in dataset I is 30 rows (according to the number of sub-districts in Bandung City) and 8 columns consisting of 7 independent variables and 1 dependent variable. **Table 2** shows some of the final rows of data from dataset I.

From **Table 2**, there are 8 variables in the dataset. The Residential, Density_of_city_center, and Density_of_economics_activities variables come from OSM data that has been processed in the previous stage, the variables Density_of_city_center, and Density_of_economics_activities come from OSM data that has been processed in the previous stage. Coronavirus case.

Table 2. Some final data dataset I.

No	Residential	Density of City Center	Density of Economics Activities	Population Density	Rumah Sehat	MBR	NON_PHBS	Covid-19 Cases
1	0.0000	0.435590817	0.12189238	2.373647472	0.850869201	0.492582349	0.637355396	712
2	0.324454467	0.170996911	0.038883624	2.259115844	0.811090748	0.643572484	0.582941729	900
3	0.197013831	0.1735688	0.004528734	1.988558957	0.885322024	0.361832061	0.674775557	820
4	0.002827213	0.038448188	0.009008768	2.439332694	0.699741714	0.493481095	0.596060704	392

2.11. Merging Covid-19 Vaccination Data and Covid-19 Cases

Data collected from the Covid-19 Vaccination data collection stage and data on the Number of Covid-19 Cases in Bandung City were collected in one file with Comma-separated values (.csv) format as dataset II. In dataset II, the vaccination data obtained in each sub-district is divided into 6 vaccination periods, starting from 21/01/2021 – 18/02/2021, 25/02/2021 – 13/03/2021, 15/03/2021 – 20/03/2021, 22/03/2021 – 27/03/2021, 29/03/2021 – 03/04/2021, and 05/04/2021 – 17/04/2021. Meanwhile, the time period for the Covid-19 Case Increase data to be used is not aligned with the time period of vaccination data. According to the World Health Organization (WHO), the time needed in the human body to form antibodies that are effective against exposure to Covid-19 is approximately 28 days since the first dose of vaccination is carried out. The period used in the Covid-19 Case Addition data must use 28 days of vaccination data. Therefore, the period of Additional Covid-19 Cases

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

starts from 18/02/2021 – 17/03/2021, 24/03/2021 – 09/04/2021, 11/04/2021 – 17/04/2021, 18/04/2021 – 23/04/2021, 25/04/2021 – 30/04/2021, and 02/05/2021 – 14/05/2021.

This dataset II will be further processed in model design using Pearson Correlation Product Moment to analyze the correlation between independent and dependent variables. **Table 3** shows some of the final rows of data from dataset II.

From **Table 3**, there are 4 variables in the dataset. The Vaccination variable for each period is data on the cumulative number of current and previous periods, and the Case Addition variable is data on the number of additional Covid-19 cases 28 days after that period.

Table 3. Some final data datasets II.

No	Kecamatan	Period	Vaccination (soul)	Case Addition (soul)
1	Andir	1	3174	143
2	Andir	2	5922	92
3	Andir	3	7194	16
4	Andir	4	11121	11
5	Andir	5	14531	2
6	Andir	6	14531	27
7	Antapani	1	1200	211
8	Antapani	2	2626	74
9	Antapani	3	3422	10
10	Antapani	4	5311	11
11	Antapani	5	5311	18
12	Antapani	6	8068	49

2.12. Correlation Analysis

After getting dataset I and dataset II that are ready for use, the next step is to analyze the correlation relationship between the independent variable and the dependent variable using Multiple Linear Regression for dataset I and Pearson Correlation Product Moment] for dataset II.

2.12.1. Correlation Analysis using Multiple Linear Regression (MLR)

MLR is a method that can be used to model the relationship between one dependent variable and more than one independent variable. The relationship is described by **the following Equation 3**.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_n X_{ni} + e_i \quad (3)$$

Information:
Y_i : Dependent variable
β_0 : Constant (intercept)
$\beta_{1\ldots n}$: Regression coefficient/estimate
$X_{1\ldots n}$: Independent variable
e_i : error

To analyze whether the factors in the independent variable affect the dependent variable, several stages are carried out as follows:

2.12.1.1. Calculating the Coefficient Regression value

Equation 4 below can be formally obtained from **Equation 3** by multiplying them by 1, , ..., respectively and summing them $X_1 X_n$ (Putri & Handayani, 2016; Park *et al.*, 2015).

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

$$\sum_{Y} = n\alpha + \beta_1 \sum_{X1} + \cdots + \beta_n \sum_{Xn} \quad (4)$$

$$\sum_{YX1} = \alpha \sum_{X1} + \beta_1 \sum_{X1^2} + \cdots + \beta_n \sum_{XnX1}$$

$$\sum_{Y\ldots} = \alpha \ldots + \beta_1 \sum_{X1\ldots} + \cdots + \beta_n \sum_{Xn\ldots}$$

$$\sum_{YXn} = \alpha \sum_{Xn} + \beta_1 \sum_{X1Xn} + \cdots + \beta_n \sum_{Xn^2}$$

The regression coefficient can be calculated by solving the above system of equations to obtain the values α, , ,..., as regression constant ($\beta_1\beta_2\beta_n$intercept) and as regression coefficient (estimate coefficient). The method used to solve the linear equation is the Cramer method by using the determinant of the matrix formed from the coefficients and constants of each equation. The system **of Equation 4** can also be written in the form of matrix multiplication as in the matrix below (Wang & Xu, 2005):

$$[\text{Matriks A}][\text{Matriks }\beta] = [\text{Matriks H}]$$

$$[n \sum_{X1} \ldots \sum_{Xn} \; \sum_{X1} \ldots \sum_{Xn} \; \sum_{X1^2} \ldots \sum_{Xn\,X1} \; \ldots \ldots \ldots \; \sum_{X1\,Xn} \ldots \sum_{Xn\,Xn} \;]$$

$$[\alpha\ \beta_1\ \beta_{\ldots}\ \beta_n\ ] \quad = \quad [\sum_{Y} \sum_{X_1 Y} \sum_{X_{\ldots} Y} \sum_{X_n Y}\ ]$$

2.12.1.1.1. Partial T Test

To ascertain which independent variables correlate with the dependent variable, a Partial T-Test is conducted. This test involves comparing the T-Calculate with the T-Table value. The T-Table value is determined based on the research conditions, considering the significance level (α) and degrees of freedom (n – k-1), where n represents the number of samples/districts, and k denotes the number of independent variables. The significance value used in this study was 5%. T-Count can be calculated using the following Equation 5 (Alita *et al*., 2021).

$$\text{T Hitung} = \frac{\beta_i}{S\beta_i} \quad (5)$$

$$S\beta_i = \frac{S_{Y.X}}{\sqrt{\{(\sum X_i^2 - n\underline{X}_i^2)(1-r_i^2)\}}} \quad (6)$$

$$S_{Y.X} = \sqrt{\frac{\sum (Y_i - \hat{Y})^2}{n-(k+1)}} \quad (7)$$

Description:

Description:	β_i	: Regression coefficient of i-th independent variable
	i	: i= 1, ..., k
	k	: Number of independent variables X
	$S\beta_i$	: Standard error penduga β_i
	$S_{Y.X}$	: Standard error variable Y based on all variable X
	X_i	: i-th independent variable
	n	: Number of sample data
	$\underline{X}_i$	: Average value of the i-th independent variable
	r_i	: The correlation value of the i-th independent variable
	$\hat{Y}_i$	: Y Value Prediction
	Y_i	: Actual Y Value

If the absolute value of T-Count is greater than the absolute value of T-Table then the variable can be said to be correlated, if vice versa then the variable is said to be uncorrelated. If the result is that there are still independent variables that are not correlated, then recalculation can be done by removing the least significant variables, this process is carried out until obtaining the results of all independent variables used in the correlated MLR model (Alita *et al*., 2021).

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

2.12.1.1.2. Simultaneous F Test

To determine whether the previously identified significant independent variables collectively affect the dependent variable, the F-Simultaneous Test is utilized. This test involves comparing the F-Calculate value with the F-Table values. The F-Table values are derived from the Table-F based on the research conditions, considering the significance level (α), the degrees of freedom of the numerator (Numerator, df) = k – 1, and the degrees of freedom of the denominator (Denominator, df) = n – k. Here, n represents the number of data samples/subdistricts, and k represents the number of independent variables. The significance value used in this study was 5%. F-Calculate can be calculated using the following Equation 8 (Gregoire, 2014):

$$\text{F Count} = \frac{MSR}{MSE} \quad (8)$$

$$\text{F Calculate} = \frac{\Sigma(\hat{Y}_i - \underline{Y})^2/k}{\Sigma(Y_i - \hat{Y}_i)^2/(n-k-1)} \quad (9)$$

Keterangan:
- MSR : Mean of Square Regression
- MSE : Mean of Square Tota
- $\hat{Y}_i$: Y Value Prediction
- Y_i : Actual Y Value
- $\underline{Y}$: Y Average Value
- n : Number of Research Samples
- k : Number of Independent Variables

If the absolute value of F-Calculate is greater than the absolute value of F-Table then the independent variables can simultaneously be said to be correlated, if vice versa then the independent variables are simultaneously said to be uncorrelated (Gregoire, 2014). If the result turns out that the independent variable is simultaneously correlated with the dependent variable, then to see how much influence it simultaneously can be known through the Coefficient of Determination () in R^2Equation 10 (Ningrum & Suwandi, 2023).

$$R_2 = \frac{SSR}{SST} \quad (10)$$

$$R_2 = \frac{\Sigma(\hat{Y}_i - \underline{Y})^2}{\Sigma(Y_i - \underline{Y})^2}, i=1,2, \dots, n \quad (11)$$

From the equation above, $\hat{Y}_i$ it is the predicted Y value, is the actual Y value, and is the average value of Y. To find the value can use linear equation 4.3 by replacing the value of the constant () and regression coefficient () with the value of the constant and regression coefficient that was obtained previously $Y_i \underline{Y} \hat{Y}_i \alpha \beta$ (Ningrum & Suwandi, 2023).

2.12.1.1.3. Classical Assumption Test

To find out whether the MLR model produces regression coefficients, partial T test results, and simultaneous F test results that are valid (correct and acceptable), it is necessary to test for possible violations of classical assumptions. In conducting analysis using MLR, there are three classical assumptions that must be met, namely the classical assumptions of Normality, Heteroscedasticity, and Multicollinearity. Manually, in performing the classical assumption test of linear regression, you must first obtain residual data. Keep in mind that classical assumption testing uses residual data, not observational data, except for multicollinearity assumption tests. The normality test was performed using the Shapiro Wilk method, the

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

Heteroscedasticity Test was performed by the Glejser method, and the Multicollinearity Test was performed using the Variance Inflation Factor (VIF) method (Ningrum & Suwandi, 2023).

2.12.1.2. Correlation Analysis using Pearson Correlation Product Moment (Pearson's r)

Pearson's r merupakan metode yang dapat mengukur kekuatan, arah dan probabilitas hubungan korelasi antara dua variabel interval atau rasio dengan rentang nilai antara -1 sampai 1.

2.12.1.3. Calculating the value of Pearson Correlation

Pearson Correlation Product Moment can be calculated the value of the correlation coefficient using the following Equation 12.

$$r_{xy} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{(n\sum x_i{}^2 - (\sum x_i)^2)\,(n\sum y_i{}^2 - (\sum y_i)^2)}} \tag{12}$$

Description: r_{xy} : Pearson correlation coefficient
x_i : x-i value
y_i : I-th Y Value
N : Number of sample data

To find out the level of the relationship, it can be seen in **Table 4** below.

Table 4. Degree of correlation relationship.

Pearson correlation coefficient	Degree of Relationship
0.00 – 0.19	Very weak
0.20 – 0.39	Weak
0.40 – 0.59	Keep
0.60 – 0.79	Strong
0.80 – 1.00	Very Powerful

2.12.1.3.1. Significance Test

After getting the value of the correlation coefficient, then to find out whether the relationship between the independent variable and the dependent variable is significant can be done using the Statistical T Test. The Statistical T Test is performed by comparing T-count values and T-Table values. The T-Table is obtained by finding its value in the T-Table according to the conditions of the research conducted by considering the significance value (α) and degrees of freedom (n – 2). The significance value (α) used in this study was 10% because the study sample was small (<10). To calculate the value of T-Calculate can use the following Equation 13.

$$\text{t-count} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{13}$$

Description: r : pearson correlation coefficient
N : Number of data samples

If the value of T-Count is greater than the value of T-Table then the independent variable and the dependent variable have a significant correlation relationship, otherwise the independent variable and the dependent variable do not have a significant correlation relationship.

In addition to employing the Statistical T Test, determining whether the independent variable and the dependent variable exhibit a significant correlation relationship can be

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

achieved through the Table r Test. This test involves comparing the Pearson correlation coefficient (r) value with the r Table value. The r Table value is determined from the R-Table based on the research conditions, considering the significance level (α) and the degrees of freedom (n). In this study, a significance level (α) of 10% is utilized due to the small data sample size (<10). If the calculated r value exceeds the r Table value, it indicates a significant correlation between the independent variable and the dependent variable. Conversely, if the result is the opposite, it suggests that the independent variable and the dependent variable do not exhibit a significant correlation relationship.

3. RESULTS AND DISCUSSION

3.1. Experiment Design

Because in this study there were 2 models built, there were 2 experimental designs designed to analyze the correlation relationship between the independent variable and the dependent variable of dataset I and dataset II.

3.2. MLR model Experiment Design

Experiments conducted on the MLR model used dataset I am containing OSM, Demographic, and Covid-19 Case data. Experiment.

3.2.1. Partial correlation analysis

The experiment was conducted to identify independent variables with a significant influence on the dependent variable and to determine the direction of the correlation relationship between them. To achieve this, the constant value and regression coefficient of each independent variable were calculated, resulting in a linear regression equation derived from dataset I. Subsequently, predicted Y values were calculated for use in the Partial T Test stage. This process enables understanding of which independent variables impact the dependent variable significantly and in which direction.

3.2.2. Simultaneous correlation analysis

The experiment aimed to determine whether the correlation relationship between all independent variables and the dependent variable collectively had a significant effect, conducted through the Simultaneous Ui F. In addition, this scenario will also calculate the value of the coefficient of determination () to find out how much influence all independent variables have on the dependent variableR^2.

3.2.3. Model qualification analysis

The experiment aimed to assess the suitability of the obtained model for use. This assessment was divided into three parts: the Normality Test, Heteroscedasticity Test, and Multicollinearity Test. If the model is found to be free from the symptoms of these three classical assumptions, then the results from both the partial correlation analysis and the simultaneous correlation analysis scenarios can be trusted, as they would produce unbiased results.

3.3. Pearson's model Experimental Design r

The experiment conducted on Pearson's r model used dataset II containing data on Covid-19 Vaccination and Covid-19 Cases. Experiments conducted include:

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

3.3.1. Correlation value experiment

This experiment aimed to assess the correlation between the variables of Vaccination and the Addition of Covid-19 Cases in each sub-district of Bandung City. Additionally, it sought to determine the direction of the correlation relationship between these two variables. The calculation of correlation values was performed using the Pearson's r method.

3.3.2. Correlation value significance experiment

This experiment was conducted to find out whether the pearson correlation (r) value obtained has a significant relationship or not. This scenario is done in 2 ways, namely doing T Statistics Test and R Table Test.

3.4. Experimental Results

The following are the results of experiments on the MLR model and Pearson's r model that the author has done in accordance with the experimental design that has been designed.

3.4.1. MLR model experiment results

The results of the experiments carried out are in accordance with the experimental design that has been designed as follows:

3.4.1.1. Partial correlation

Of the 7 independent variables (**Table 5**) analyzed, there are only 3 independent variables that have a significant correlation relationship with the dependent variable Number of Covid-19 Cases (see **Table 6**).

Table 5. MLR regression coefficient.

No	Variable	Notation	Regression Coefficient
1	Constanta	α	-1088.91358
2	Residential	β_1	1221.74656
3	Population Density	β_2	439.18181
4	Healthy Home	β_3	853.41117

Table 6. Partial T Test Results.

No	Variable	T-Count (T1)	T-Table (T2)	Information	Conclusion
1	Residential	3.26456	2.05553	T1 > T2	Significant
2	Population Density	3.04659	2.05553	T1 > T2	Significant
3	Healthy Home	2.11501	2.05553	T1 > T2	Significant

3.4.1.2. Simultaneous Correlation

Of the 3 variables that have a partial significant correlation relationship, it also has a significant correlation relationship together (simultaneously). Judging from the value of R Square, the variables Residential, Population Density, and Number of Healthy Houses have an influence on the Number of Covid-19 Cases by 55.404% (see **Table 7**).

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

Table 7. Simultaneous F Test Results.

R Square (R^2)	F-Count (F1)	F-Table (F2)	Information	Conclusion
0.55404	10.76719	2.96035	F1 > F2	Significant

3.4.1.2. Model Eligibility

The MLR model that has been obtained will be tested through 3 tests, namely the Normality Test, Heteroscedasticity Test, and Multicollinearity Test. Tables 4.12, 4.13, and 4.14 are experimental results of normality, heteroscedasticity, and multicollinearity testing (see **Tables 8-10**).

Table 8 shows that through the Shapiro Wilk Test the model is declared to be normally distributed and passes the Classical Assumption Test of Normality. **Table 9** shows that all independent variables are insignificant. This means that heteroscedasticity does not occur in the model. And **Table 10** shows that all variables independent has a VIF value of less than 10 so that the three variables do not occur multicollinearity including: Residential, Population density, and healthy home.

Table 8. Normality Test Results.

Shapiro Wilk Count (SW1)	Shapiro Wilk Table (SW2)	Information	Conclusion
0.97082	0.92700	SW1 > SW2	Residual Normal Distributed

Table 9. Heteroscedasticity Test Results.

No	Variable Name	T-Count (T1)	T-Table (T2)	Information	Significance
1	Residential	2.01924	2.05553	T1 < T2	Insignificant
2	Population Density	0.01705	2.05553	T1 < T2	Insignificant
3	Healthy Home	0.87385	2.05553	T1 < T2	Insignificant

Table 10. Multicollinearity Test Results.

No	Variable Name	VIF value	< 10
1	Residential	1.20502	Meet
2	Population Density	1.19059	Meet
3	Healthy Home	1.36010	Meet

3.4.2. Experiments Model Pearson's r Result

The results of the experiments carried out are in accordance with the experimental design that has been designed as follows:

3.4.3. Correlation Value

The value of the correlation relationship is calculated using **Equation 12**. The results of this experiment are as follows **Tables 11 and 12**.

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

Table 11. Results of Pearson Correlation (r) Value Calculation Involving 5 Time Periods Between Covid-19 Vaccination Data and Covid-19 Case Addition Data in Bandung City.

No	Kecamatan	Pearson Correalation (r)	Degree of Relationship
1	Andir	-0.719	Strong
2	Antapani	-0.818	Very Powerful
3	Arcamanik	-0.689	Strong
4	Astana Anyar	-0.928	Very Powerful
5	Babakan Ciparay	-0.612	Strong
6	Bandung Kidul	-0.677	Strong
7	Bandung Kulon	-0.883	Very Powerful
8	Bandung Wetan	-0.962	Very Powerful
9	Batununggal	-0.850	Very Powerful
10	Bojongloa Kaler	-0.782	Strong
11	Bojongloa Kidul	-0.801	Very Powerful
12	Buahbatu	-0.887	Very Powerful
13	Cibeunying Kaler	-0.894	Very Powerful
14	Cibeunying Kidul	-0.890	Very Powerful
15	Cibiru	-0.933	Very Powerful
16	Cicendo	-0.741	Strong
17	Cidadap	-0.976	Very Powerful
18	Cinambo	-0.697	Strong
19	Coblong	-0.937	Very Powerful
20	Gedebage	-0.894	Very Powerful
21	Kiaracondong	-0.930	Very Powerful
22	Lengkong	-0.949	Very Powerful
23	Mandalajati	-0.910	Very Powerful
24	Panyileukan	-0.871	Very Powerful
25	Rancasari	-0.913	Very Powerful
26	Regol	-0.823	Very Powerful
27	Sukajadi	-0.773	Strong
28	Sukasari	-0.896	Very Powerful
29	Sumur Bandung	-0.106	Very Powerful
30	Ujung Berung	-0.827	Very Powerful

Table 12. Results of Pearson Correlation (r) Value Calculation Involving 6 Time Periods Between Covid-19 Vaccination Data and Covid-19 Case Addition Data in Bandung City.

No	Subdistrict	Pearson Correlation (r)	Degree of Relationship
1	Andir	-0.719	Strong
2	Antapani	-0.561	Keep
3	Arcamanik	-0.618	Strong
4	Astana Anyar	-0.925	Very Powerful
5	Babakan Ciparay	-0.377	Weak
6	Bandung Kidul	-0.592	Keep
7	Bandung Kulon	-0.728	Strong
8	Bandung Wetan	-0.657	Strong
9	Batununggal	-0.549	Keep
10	Bojongloa Kaler	-0.617	Strong
11	Bojongloa Kidul	-0.737	Strong
12	Buahbatu	-0.789	Strong
13	Cibeunying Kaler	-0.793	Strong
14	Cibeunying Kidul	-0.654	Strong
15	Cibiru	-0.749	Strong

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

Table 12 (continue). Results of Pearson Correlation (r) Value Calculation Involving 6 Time Periods Between Covid-19 Vaccination Data and Covid-19 Case Addition Data in Bandung City.

No	Subdistrict	Pearson Correlation (r)	Degree of Relationship
16	Cicendo	-0.471	Keep
17	Cidadap	-0.884	Very Powerful
18	Cinambo	-0.530	Keep
19	Coblong	-0.704	Strong
20	Gedebage	-0.571	Keep
21	Kiaracondong	-0.908	Very Powerful
22	Lengkong	-0.912	Very Powerful
23	Mandalajati	-0.713	Strong
24	Panyileukan	-0.636	Strong
25	Rancasari	-0.527	Keep
26	Regol	-0.717	Strong
27	Sukajadi	-0.772	Strong
28	Sukasari	-0.659	Strong
29	Sumur Bandung	-0.177	Very weak
30	Ujung Berung	-0.489	Keep

3.4.3. Significance of Correlation Value

In this experiment, there are 2 test methods used, namely the T Statistics Test and the Table R Test. **Table 13** and **Table 14** are the experimental results of the T Statistics Test and R-Table of the correlation relationship between Vaccination and the addition of Covid-19 for each sub-district in Bandung City involving 5 vaccination periods and 6 vaccination periods. The T-Table value in the experiment using 5 vaccination periods was 2.3534 and the R-Table value was 0.8054. While the T-Table value in the experiment using 6 vaccination periods was 2.13185 and the R-Table value was 0.72930.

Table 13. Results of T Test and R Test Pearson Correlation values with 5 Time Periods Between Covid-19 Vaccination Data and Covid-19 Case Addition Data in Bandung City.

No	Subdistrict	T-Count	R-Count	Conclusion
1	Andir	1.7942	0.7195	Insignificant
2	Antapani	2.4637	0.8181	Significant
3	Arcamanik	1.6487	0.6895	Insignificant
4	Astana Anyar	4.3009	0.9276	Significant
5	Babakan Ciparay	1.3401	0.6119	Insignificant
6	Bandung Kidul	1.5915	0.6766	Insignificant
7	Bandung Kulon	3.2571	0.8829	Significant
8	Bandung Wetan	6.0841	0.9618	Significant
9	Batununggal	2.7951	0.85	Significant
10	Bojongloa Kaler	2.1739	0.7821	Insignificant
11	Bojongloa Kidul	2.3174	0.801	Insignificant
12	Buahbatu	3.3219	0.8867	Significant
13	Cibeunying Kaler	3.4482	0.8936	Significant
14	Cibeunying Kidul	3.3734	0.8896	Significant
15	Cibiru	4.4899	0.933	Significant
16	Cicendo	1.912	0.7411	Insignificant
17	Cidadap	7.8426	0.9765	Significant
18	Cinambo	1.6816	0.6966	Insignificant
19	Coblong	4.6463	0.937	Significant

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

Table 13 (continue). Results of T Test and R Test Pearson Correlation values with 5 Time Periods Between Covid-19 Vaccination Data and Covid-19 Case Addition Data in Bandung City.

No	Subdistrict	T-Count	R-Count	Conclusion
1	Andir	1.7942	0.7195	Insignificant
2	Antapani	2.4637	0.8181	Significant
3	Arcamanik	1.6487	0.6895	Insignificant
4	Astana Anyar	4.3009	0.9276	Significant
5	Babakan Ciparay	1.3401	0.6119	Insignificant
6	Bandung Kidul	1.5915	0.6766	Insignificant
7	Bandung Kulon	3.2571	0.8829	Significant
8	Bandung Wetan	6.0841	0.9618	Significant
9	Batununggal	2.7951	0.85	Significant
10	Bojongloa Kaler	2.1739	0.7821	Insignificant
11	Bojongloa Kidul	2.3174	0.801	Insignificant
12	Buahbatu	3.3219	0.8867	Significant
13	Cibeunying Kaler	3.4482	0.8936	Significant
14	Cibeunying Kidul	3.3734	0.8896	Significant
15	Cibiru	4.4899	0.933	Significant
16	Cicendo	1.912	0.7411	Insignificant
17	Cidadap	7.8426	0.9765	Significant
18	Cinambo	1.6816	0.6966	Insignificant
19	Coblong	4.6463	0.937	Significant
20	Gedebage	3.4504	0.8937	Significant
21	Kiaracondong	4.369	0.9296	Significant
22	Lengkong	5.2089	0.9489	Significant
23	Mandalajati	3.7903	0.9095	Significant
24	Panyileukan	3.0731	0.8712	Significant
25	Rancasari	3.8819	0.9132	Significant
26	Regol	2.5133	0.8234	Significant
27	Sukajadi	2.112	0.7732	Insignificant
28	Sukasari	3.4954	0.896	Significant
29	Sumur Bandung	0.1855	0.1065	Insignificant
30	Ujung Berung	2.5445	0.8267	Significant

Table 14. Results of T Test and R Test Pearson Correlation values with 6 Time Periods Between Covid-19 Vaccination Data and Covid-19 Case Addition Data in Bandung City.

No	Subdistrict	T-Count	R-Count	Conclusion
1	Andir	2.0696	0.7191	Insignificant
2	Antapani	1.3568	0.5614	Insignificant
3	Arcamanik	1.5718	0.6179	Insignificant
4	Astana Anyar	4.8652	0.9249	Significant
5	Babakan Ciparay	0.814	0.377	Insignificant
6	Bandung Kidul	1.4689	0.592	Insignificant
7	Bandung Kulon	2.1227	0.7278	Insignificant
8	Bandung Wetan	1.7412	0.6566	Insignificant
9	Batununggal	1.3127	0.5487	Insignificant
10	Bojongloa Kaler	1.567	0.6167	Insignificant
11	Bojongloa Kidul	2.1814	0.7371	Significant
12	Buahbatu	2.5712	0.7893	Significant
13	Cibeunying Kaler	2.6063	0.7933	Significant
14	Cibeunying Kidul	1.7293	0.6541	Insignificant

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

Table 14 (continue). Results of T Test and R Test Pearson Correlation values with 6 Time Periods Between Covid-19 Vaccination Data and Covid-19 Case Addition Data in Bandung City.

No	Subdistrict	T-Count	R-Count	Conclusion
15	Cibiru	2.2589	0.7487	Significant
16	Cicendo	1.0685	0.4712	Insignificant
17	Cidadap	3.7788	0.8838	Significant
18	Cinambo	1.2495	0.5299	Insignificant
19	Coblong	1.9834	0.7042	Insignificant
20	Gedebage	1.3894	0.5705	Insignificant
21	Kiaracondong	4.3326	0.9079	Significant
22	Lengkong	4.4608	0.9125	Significant
23	Mandalajati	2.0312	0.7126	Insignificant
24	Panyileukan	1.6483	0.636	Insignificant
25	Rancasari	1.239	0.5266	Insignificant
26	Regol	2.0593	0.7174	Insignificant
27	Sukajadi	2.4315	0.7723	Significant
28	Sukasari	1.7538	0.6593	Insignificant
29	Sumur Bandung	0.36	0.1771	Insignificant
30	Ujung Berung	1.1199	0.4886	Insignificant

4.2. Discussion

From the experimental results obtained, the author can do some discussion which will be described in the following sub-sections.

4.2.1. Multiple Linear Regression Model (MLR)

MLR model experiments are carried out with 3 main stages, starting from partial correlation analysis experiments, simultaneous correlation analysis, and model feasibility analysis. The calculation of constants and regression coefficients is carried out using MLR with the Backward method. The Backward method is good enough to use because it removes one independent variable that has an insignificant relationship to the dependent variable in each iteration, resulting in a constant and regression coefficient of the independent variable that has a significant relationship only in the last iteration. **Table 5** above is a table of experimental results that produce constants and regression coefficients from dataset I, from these constants and coefficients can be determined the regression equation model of research data based on the previous Equation 3:

$$Y_i = -1088.91358 + 1221.74656X_{1i} + 439.18181X_{2i} + 853.41117X_{3i} + e_i$$

The regression equation above can be interpreted as follows: (1) Every addition of 1 unit of value of the variable X_1(Residential) can increase the value of the dependent variable (Covid-19 Cases) by 1221.74656 cases assuming other independent variables have fixed values, (2) Every addition of 1 unit of value of the variable LogX_2(Population Density Log) can increase the value of the dependent variable (Covid-19 Cases) by 439.18181 cases assuming other independent variables has a fixed value, and (3) Every addition of 1 unit of value of the variable X_3(Rumah Sehat) can increase the value of the dependent variable (Covid-19 Cases) by 853.41117 cases assuming other independent variables have a fixed value.

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

Table 6 is the experimental result of the Partial T Test on the regression model formed. Of the 7 independent variables analyzed, there are 3 independent variables that have a significant correlation relationship, as seen from the T-Calculate value of the three variables is smaller than the T-Table value. From the table it can also be seen that the variables Residential, Population Density, and Healthy Homes have a positive T-Count value, which means that the 3 variables have a positive correlation relationship. A positive correlation relationship means that, if the value of the variables Residential, Population Density, and Healthy Houses increases, the value of the dependent variable (Covid-19 Cases) will increase cases as well, on the other hand, if the value of the 3 variables decreases, the dependent variable (Covid-19 Cases) will also decrease. The results of the T-Count calculation show that, the greater the value of T-Calculation, the greater the influence of the independent variable on the dependent variable. From the table of Partial T Test Results, it can be concluded that the Residential variable has the highest influence, followed by the Population Density variable, then the Healthy House variable.

After knowing that only the variables Residential, Population Density, and Healthy Houses have a significant influence, then the next step is to conduct a Simultaneous F Test to find out whether the 3 independent variables together have a significant effect on the dependent variable (Covid-19 Cases). **Table 7** shows that the F-Count value of the 3 variables is smaller than the F-Table value so that it can be concluded that the variables Residential, Population Density, and Healthy Houses together (simultaneously) have a significant correlation relationship with the dependent variable. In addition to calculating F-Calculate and F-Table, this experiment also calculated the value of the coefficient of determination / R Square (), the 3 independent variables have a value of 55.404% which shows that the three variables have a joint influence on the dependent variable by 55%, the rest are influenced by variables outside the scope of this study.R^2R^2.

After analyzing the correlation relationship partially and simultaneously, to ensure that the model that has been obtained is feasible for use, Classical Assumptions will be tested. The classical Assumption Test consists of 3 tests, namely (1) Normality Test, (2) Heteroscedasticity Test, and (3) Multicollinearity Test. If the model passes the 3 tests, then the decisions obtained from partial correlation analysis and simultaneous correlation analysis performed previously can be trusted for the level of truth.

4.2.1.1. Normality Test

In this study, the author used the shapiro wilk method because of the small research sample data (30 data samples). The Normality Test is performed by comparing the Shapiro Wilk Count value with the Shapiro Wilk Table value. **Table 8** is the experimental result of calculating the value of Shapiro Wilk Calculate and Shapiro Wilk Table. From the table the value of Shapiro Wilk Calculate is greater than the value of Shapiro Wilk Table so it can be said that the residual data is normally distributed. If the residual data is not normally distributed, the regression model will be calculated again by transforming variables or other actions. The consequence of allowing this problem is that the value of the Partial T Test to see the significance of the independent variable to the dependent variable cannot be applied or in other words the T-Calculate value produced by the model is biased or cannot be trusted to the level of truth.

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

4.2.1.2. Heteroscedasticity Test

The authors utilized the statistical method of the Glejser test to identify whether the regression model exhibited symptoms of heteroscedasticity. The Glejser test involves regressing all independent variables against residual absolute variables, followed by a Partial T test, as conducted in the previous stage. Table 9 presents the results of experiments testing for symptoms of heteroscedasticity. From the table, it is evident that none of the independent variables are significant in the T test. This indicates that heteroscedasticity is not present in the model. If a regression model contains heteroscedasticity, it can have serious consequences for the Ordinary Least Squares (OLS) method estimator. The increased variance of the estimator compromises the reliability of calculating the standard error of the OLS method, rendering estimates and hypothesis tests based on the distribution of T and F unreliable for evaluating regression results.

4.2.1.3. Multicollinearity Test

This test is conducted to determine whether there is a high correlation between independent variables in a model. If a high correlation exists between independent variables, it can disrupt the relationship between the independent variable and the dependent variable. From **Table 10**, all independent variables have a VIF value of less than 10 so it can be said that there are no symptoms of multicollinearity between the variables of Residential, Population Density, and Healthy Homes. Regression models that experience symptoms of multicollinearity will as a result, the value of the regression coefficient (β) of each independent variable and the standard value of error (Sβ) tend to be widened and biased because it is difficult to get a precise estimate because the estimator has a large variance and covariance. Another consequence is that the statistical calculated value of the t-test will be small to make the independent variable statistically insignificant.

Of the 3 Classical Assumption tests that have been carried out above, the Multiple Linear Regression model obtained escapes all Classical Assumptions, so that decision making on partial correlation relationship analysis and simultaneous correlation relationship analysis is reliable and the model is feasible to use.

4.2.3. Pearson Correlation (Pearson's r) Model

This experiment was carried out with 2 main stages, starting from the experiment of calculating the correlation value of pearson (r), and the experiment testing the significance of the value of r. The value of r is calculated using Equation 12. **Table 11** is the experimental result of r calculation for 30 sub-districts in Bandung City along with the degree of correlation relationship. The degree of correlation relationship is determined based on **Table 4**. From **Table 12** above, there are 4 sub-districts with a moderate degree of correlation, 19 sub-districts with a strong degree of correlation, and 7 sub-districts with a very strong degree of correlation. **Table 12** Additionally, the results reveal that nearly all r values are negative, except for Sumur Bandung sub-district, which is positive. The negative r values for 29 sub-districts suggest a negative correlation between Covid-19 Vaccination and the Increase of Covid-19 Cases. In other words, an increase in the number of patients vaccinated is associated with a decrease in the number of additional Covid-19 Cases. In Sumur Bandung District, the opposite applies (positive correlation relationship), the more patients vaccinated will still increase the number of additional Covid-19 cases. From the results of this experiment, the vaccination steps carried out by the Bandung City Health Office have mostly been successfully proven by the decrease in the number of additional Covid-19 cases after vaccination.

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

After knowing the value of r, the next step is to find out the value of r of any sub-district that has a significant correlation relationship. This experiment was carried out using 2 approaches, namely the T Statistics Test approach and the Table R Test. The T Statistics test is performed by comparing the T-count value and the T-Table value, while the r Table Test is performed by comparing the r value (pearson correlation coefficient) with r Table. From **Table 12** above, Both the T Statistics Test and the R Table Test yield identical results. Of the 30 sub-districts in Bandung City that have been vaccinated, 14 sub-districts show a significant correlation between vaccination and the increase in Covid-19 cases, while the remaining 16 sub-districts do not demonstrate a significant correlation. According to the R Table Test, a sub-district is considered to have a significant correlation between vaccination and the increase in Covid-19 cases if the r value exceeds 0.72930, or in other words, if the r value is greater than 73%. From the results of this experiment, it can be said that vaccination carried out by the Bandung City Health Office has been effective in 14 sub-districts with a correlation value between vaccination and the addition of covid-19 cases above 73%, 12 sub-districts already have a correlation relationship with a strong degree but not yet significant (60%-73%), and 4 other sub-districts have a moderate and insignificant correlation relationship.

This experiment was carried out by conducting analysis twice, namely analysis with 5 vaccination time periods and analysis with 6 vaccination time periods. Each analysis will be carried out with 2 main stages, namely experiments in calculating the Pearson correlation value (r), and testing experiments on the significance of the r value.

The first analysis involving 5 vaccination periods alone showed that there was 1 sub-district with a very weak degree of correlation, 8 sub-districts with a strong degree of correlation, and 21 sub-districts with a very strong degree of correlation. The experimental results of calculating the pearson (r) correlation value for 30 sub-districts in Bandung City along with the degree of correlation relationship can be seen in **Table 11**. The degree of correlation relationship is determined based on **Table 4**. **Table 11** also shows that the overall value of pearson correlation (r) obtained is negative, which means that there is a negative correlation between Covid-19 Vaccination and the Increase of Covid-19 Cases or in other words the more the number of patients vaccinated, the number of Additional Covid-19 Cases will decrease. After knowing the value of correlation pearson (r), the next step is to find out the value of correlation pearson (r) what districts have a significant correlation relationship. This experiment was carried out using 2 approaches, namely the T Statistics Test approach and the Table R Test. The T Statistics test is performed by comparing the T-count value and the T-Table value, while the r Table Test is performed by comparing the r value (pearson correlation coefficient) with r Table. From **Table 13**, the T Statistics Test and the Table R Test have the same result. Both methods produce results that, of the 30 sub-districts in Bandung City that have been vaccinated, there are 20 sub-districts that have a significant correlation relationship between Vaccination and the addition of Covid-19 cases, while the other 10 sub-districts produce results of a correlation relationship that has not been significant. From the results of the R Table experiment, a sub-district is said to have a significant correlation relationship between Vaccination and the Increase of Covid-19 Cases if the Pearson correlation value (r) is more than 0.8054 or in other words the correlation value (r) must be more than 80.54%.

The second analysis involving 6 vaccination periods showed that there was 1 sub-district with a very weak degree of correlation, 1 sub-district with a weak degree of correlation, 8 sub-districts with a moderate degree of correlation, 16 sub-districts with a strong degree of correlation, and 4 sub-districts with a very strong degree of correlation. The experimental results of calculating the pearson (r) correlation value for 30 sub-districts in Bandung City

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

along with the degree of correlation relationship can be seen in **Table 12**. **Table 12** also shows that the overall pearson correlation value (r) obtained is negative, which means that there is a negative correlation between Covid-19 Vaccination and the Increase of Covid-19 Cases. After knowing the value of correlation pearson (r), the next step is to find out the value of correlation pearson (r) what districts have a significant correlation relationship. This experiment was carried out using 2 approaches, namely the T Statistics Test approach and the Table R Test. From **Table 14**,

The Statistical T Test and the Table R Test yielded identical results. Both methods indicated that, out of the 30 vaccinated sub-districts in Bandung City, 9 sub-districts exhibited a significant correlation between Vaccination and the increase in Covid-19 cases, while the remaining 21 sub-districts showed insignificant correlation results. According to the R Table experiment, a sub-district is considered to have a significant correlation relationship between Vaccination and the increase in Covid-19 Cases if the Pearson correlation value (r) exceeds 0.7293, or in other words, if the correlation value (r) is greater than 72.93%.

The decline in significance was attributed to another surge in the number of Covid-19 cases during the 6th period. Nonetheless, the results of this experiment indicate that the steps taken by the Bandung City Health Office in providing vaccinations have largely been successful. This is evidenced by the decrease in the number of additional Covid-19 cases in 20 districts after the vaccination periods 1 to 5, although there was an increase in Covid-19 cases again during period 6.

4. CONCLUSION

After conducting research on the correlation between Open Street Map (OSM) data, demographic data, and vaccination data with Covid-19 cases, it can be concluded that this study successfully designed a model to analyze these relationships using Multiple Linear Regression (MLR) and Pearson Correlation Product Moment (Pearson's r) methods. The system developed through this research involved several stages: collecting OSM data using ArcGIS, merging all OSM data into one table, normalizing OSM data, grouping and weighting OSM data, integrating OSM data with demographic and Covid-19 case data, combining Covid-19 vaccination data with Covid-19 case data, and conducting correlation analysis. The MLR model indicates that out of the 7 independent variables in dataset I, only 3 variables show a significant correlation with Covid-19 cases. These variables, listed from the highest to the lowest correlation, are residential areas, population density, and healthy homes, all of which have a positive correlation with Covid-19 cases. Together, these three variables significantly correlate with Covid-19 cases, accounting for 55.4% of the variance. The model also passed the test of 3 Classical Assumptions, ensuring that the results are reliable.

The system built using the Pearson's r model indicates that, from the two data analysis experiments conducted, the first analysis involving 5 vaccination periods yielded better results than the second analysis involving 6 vaccination periods. The first analysis, using vaccination data from periods 1 to 5, identified 20 sub-districts with a significant correlation between Covid-19 vaccination and the increase in Covid-19 cases, with a correlation value exceeding 80.54%. In contrast, the second analysis, which used vaccination data from periods 1 to 6, identified only 9 sub-districts with a significant correlation, with a correlation value exceeding 72.93%.

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

5. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. The authors confirmed that the paper was free of plagiarism.

6. REFERENCES

Athena, A., Laelasari, E., and Puspita, T. (2020). Pelaksanaan disinfeksi dalam pencegahan penularan covid-19 dan potensi risiko terhadap kesehatan di Indonesia. *Jurnal Ekologi Kesehatan, 19*(1), 1-20.

Dewi, W. A. F. (2020). Dampak Covid-19 terhadap implementasi pembelajaran daring di Sekolah Dasar. *Edukatif: Jurnal Ilmu Pendidikan*, *2*(1), 55-61.

Ramírez-Aldana, R., Gomez-Verjan, J. C., and Bello-Chavolla, O. Y. (2020). Spatial analysis of COVID-19 spread in Iran: Insights into geographical and structural transmission determinants at a province level. *PLoS neglected tropical diseases*, *14*(11), e0008875.

Haq, A., Achmadi, U. F., and Susanna, D. (2019). Analisis Spasial (Topografi) Tuberkulosis Paru di Kota Pariaman, Bukittinggi, dan Dumai Tahun 2010-2016. *Jurnal Ekologi Kesehatan, 18*(3), 149-158.

Han, Y., Yang, L., Jia, K., Li, J., Feng, S., Chen, W., Zhao, W. and Pereira, P. (2021). Spatial distribution characteristics of the COVID-19 pandemic in Beijing and its relationship with environmental factors. *Science of the Total Environment*, *761*, 144257.

Haklay, M., and Weber, P. (2008). Openstreetmap: User-generated street maps. *IEEE Pervasive computing, 7*(4), 12-18.

Yagoub, M. M. (2017). Assessment of OpenStreetMap (OSM) Data: The Case of Abu Dhabi City, United Arab Emirates. *Journal of Map and Geography Libraries, 13*(3), 300-319.

Hall, R., Fienberg, S. E., and Nardi, Y. (2011). Secure multiple linear regression based on homomorphic encryption. *Journal of Official Statistics, 27*(4), 669-691.

Chee, J. D. (2015). Pearson's product-moment correlation: Sample analysis. *Research Gate*, *4*(1), 4-90.

Putri, M. K., and Handayani, M. N. (2016). Faktor-faktor Internal yang Mempengaruhi Mahasiswa Pendidikan Teknologi Agroindustri dalam Pemilihan Karir. *EDUFORTECH, 2*(1), 24-32.

Park, B. U., Mammen, E., Lee, Y. K., and Lee, E. R. (2015). Varying coefficient regression models: a review and new developments. *International Statistical Review, 83*(1), 36-64.

Wang, G., and Xu, Z. (2005). Solving a kind of restricted matrix equations and Cramer rule. *Applied mathematics and computation*, *162*(1), 329-338.

Alita, D., Putra, A. D., and Darwis, D. (2021). Analysis of classic assumption test and multiple linear regression coefficient test for employee structural office recommendation. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 15*(3), 295-306.

Grégoire, G. (2014). Multiple linear regression. *European Astronomical Society Publications Series*, 66, 45-72.

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088

Ningrum, Y. P., and Suwandi, D. (2023). Pengaruh lingkungan keluarga dan fasilitas belajar terhadap prestasi belajar pada siswa kelas X SMA Negeri 5 Surakarta. *Jurnal Ilmiah Mandala Education, 9*(3), 1894-1903.

DOI: https://doi.org/10.17509/jcs.v5i2.70798
e- ISSN 2723-4088