The 4th International Seminar of Science and Technology
ISST 2024 Vol 4 (2025) 01 027
Innovations in Science and Technology to Realize Sustainable Development Goals
Faculty of Science and Technology
Universitas Terbuka

The Use of ST-DBSCAN in the Analysis of Covid-19 Spread Patterns Based on Spatio-Temporal Data

Muhammad Syahid Pebriadi[1][1], and Muhammad Arfan[2]

[1]Politeknik Negeri Banjarmasin, Computerized Accounting, 70123, Banjarmasin, South Kalimantan, Indonesia
[2]Universitas Halu Oleo, Computer Science, 93232, Kendari, Southeast Sulawesi, Indonesia

Abstract. The spread of Covid-19 poses significant risks, necessitating strict policies and specialized plans. One of the measures that can be taken to help control the spread is accurately identifying areas with a high number of Covid-19 cases and areas with lower case numbers. This study aims to analyze the spread of Covid-19 using the ST-DBSCAN algorithm. The ST-DBSCAN algorithm was applied to Covid-19 spread data in Makassar city in August and September 2021, using spatial aspect parameters (Eps1) = 0.002, temporal aspect (Eps2) = 14, and minimum cluster members (MinPts) = 10, resulting in 78 clusters and 1639 noise points. The clusters formed through the application of the ST-DBSCAN algorithm were used to analyze patterns based on spatio-temporal aspects. The spatio- temporal patterns identified include occasional patterns and stationary patterns. The analysis results indicate that the area with the highest Covid-19 spread is in the southeastern part of Makassar city, specifically in Rappocini district, with 921 cases.

Keywords: Covid-19, Spatio-temporal Pattern, ST-DBSCAN.

1 Introduction

The Covid-19 pandemic has presented unprecedented challenges to public health systems worldwide. Understanding the spatial and temporal dynamics of the virus's spread is crucial for devising effective containment strategies. In recent years, advanced data mining techniques, such as clustering algorithms, have been employed to analyze complex datasets, revealing patterns that are not immediately apparent through traditional statistical methods [1]. Among these techniques, the Spatial-Temporal Density-Based Spatial Clustering of Applications with Noise (ST-DBSCAN) has gained attention for its ability to detect clusters in data that exhibit both spatial and temporal dimensions [2].

ST-DBSCAN is an extension of the well-known DBSCAN algorithm, designed to handle data that varies over both space and time. Unlike conventional clustering methods, ST-DBSCAN can identify regions with high concentrations of events, such as Covid-19 cases, while accounting for

[1]Corresponding author: m.syahid@poliban.ac.id

the time dimension, making it particularly useful for analyzing the spread of infectious diseases [3]. By applying ST-DBSCAN, researchers can identify clusters of Covid-19 cases that are not only geographically close but also occur within a specific time frame, providing insights into the spatio-temporal patterns of the virus's transmission.

Previous studies have demonstrated the effectiveness of ST-DBSCAN in various fields, including environmental monitoring, transportation, and epidemiology [4]–[6]. Chile's high seismic activity, driven by the Nazca plate subduction, poses challenges in classifying seismic events. This study introduces ST-DBSCAN-EV, a density-based clustering method with a variable radius, to classify precursors, main events, and aftershocks. Tested on three major Chilean earthquakes (>8.0 Mw), ST-DBSCAN-EV outperformed DBSCAN, ST-DBSCAN, and K-means, achieving F1-Scores >0.8. The method's accuracy was further validated against a declustering approach, proving its effectiveness in seismic event classification [4]. Trajectory anomaly detection is crucial for enhancing transportation safety and efficiency. This study introduces TS- DBSCAN, a novel density-based clustering model for detecting outliers in time-series trajectory data, which analyzes time correlations and uses distance density distribution to determine clustering parameters. Experimental results on real vehicle trajectory data demonstrate that TS- DBSCAN effectively and accurately identifies abnormal trajectories [5]. Spatio-temporal data mining is increasingly vital due to the abundance of data from diverse sources, enabling the exploration of disease patterns in epidemiology using methods like spatio-temporal clustering. This paper implements ST-DBSCAN on a public health dataset, identifies its limitations, and proposes a fuzzy version, successfully applying it to analyze Chicago West Nile Virus (WNV) data from 2007 to 2017, while suggesting improvements to the original method [6]. However, its application to the analysis of Covid-19 spread patterns remains underexplored, particularly in the context of urban environments where the virus's transmission dynamics are complex and multifaceted.

This study aims to apply the ST-DBSCAN algorithm to analyze Covid-19 spread patterns in Makassar, Indonesia, focusing on data from August and September 2021. By identifying spatio-temporal clusters of Covid-19 cases, this research seeks to provide valuable insights into the areas with the highest risk of transmission and inform targeted interventions to mitigate the spread of the virus.

2 Theory

2.1 Covid-19

Covid-19 is a type of pneumonia virus caused by Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2). This virus is the third highly pathogenic type of coronavirus, following the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and the Middle East Respiratory Syndrome Coronavirus (MERS-CoV). Covid-19 was first reported in Wuhan, Hubei province, China, in December 2019 [7]. The most common symptoms of Covid-19 are fever, fatigue, and dry cough. These symptoms are mild and occur gradually. Covid-19 is zoonotic, but as we know, SARS-CoV-2 can be transmitted from human to human [8]. Covid-19 is a new type of disease that had not previously been identified in humans. The spread of this virus occurs very rapidly.

2.2 ST-DBSCAN

The ST-DBSCAN algorithm has four parameters: Eps1 is used to measure the distance parameter in spatial data, Eps2 to measure the distance in temporal data, MinPts is the minimum number of members within Eps1 and Eps2, and cluster homogeneity ($\Delta\in$) is used to prevent the discovery of merged clusters due to small differences in the temporal values of neighboring locations [2].

The flow of the ST-DBSCAN algorithm is as follows:

a. Determine the values of Eps and MinPts.

b. Calculate all Euclidean distances between objects based on spatial and temporal aspects.

c. Form a distance matrix for all pairs of n objects based on spatial and temporal aspects.

d. Starting from the first point, then take all points in the spatial and temporal aspects with the conditions:

   $A = \{x \mid x \leq Eps1, x \in \text{spatial aspect matrix}\}$

   $B = \{x \mid x \leq Eps2, x \in \text{temporal aspect matrix}\}$

e. Take all intersections of spatial and temporal aspects with the condition:

   $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$

f. If the number of objects in the intersection is less than the value of MinPts, then the point is considered noise.

g. A cluster is formed if the point meets the Eps1, Eps2, and MinPts parameters.

h. If point p is a border point and there are no other points in the intersection, proceed to the next point, and a new cluster is formed.

i. Repeat steps 4 - 8 until all points are processed.

j. If two clusters, C1 and C2, are close to each other, a point q may belong to both clusters. However, this algorithm will designate point q as part of the cluster that first discovered it.

2.3 Spatio-temporal Pattern

Spatio-temporal patterns refer to the process of grouping objects based on spatial and temporal similarities. The temporal aspect describes changes in the object's data over time, while the spatial aspect describes the object's location. There are four types of spatio-temporal patterns: Stationary, Reappearing, Occasional, and Tracks [9]. An illustration of spatio-temporal patterns can be seen in Figure 1 [10].

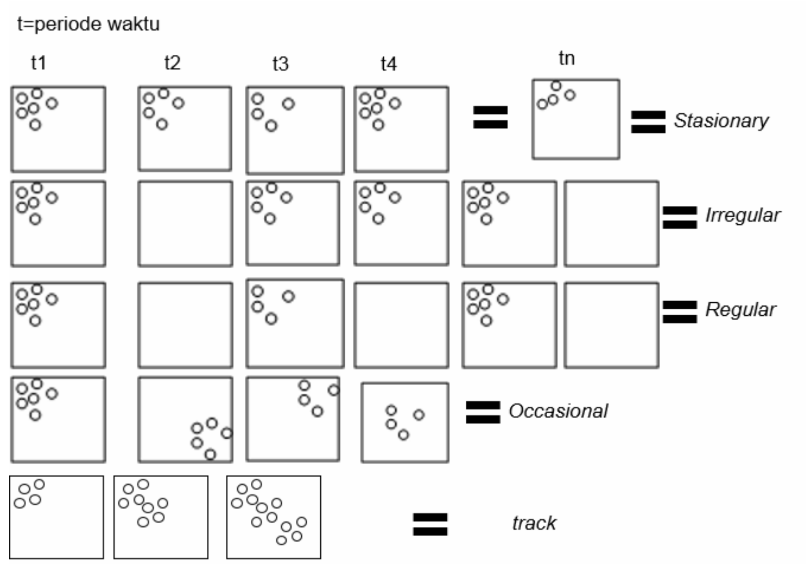

Fig 1. Illustration of Spatio-temporal Patterns

a. Stationary: refers to clustering that is spatially confined and temporally extended over the research period.

b. Reappearing: refers to temporal clustering occurring in the same place, separated by time intervals during which few or no events occur in that place. There are two types of reappearing patterns: regular (periodic) and irregular. Regular patterns are temporal clusters separated by intervals of approximately equal length, while irregular patterns are temporal clusters separated by intervals of unequal length.

c. Occasional: refers to clustering that changes location as time changes.

d. Track: refers to dense temporal clustering of events that then move spatially relative to previous events.

2.4 Previous Research

The first study implemented and analysed ST-DBSCAN on a public health dataset. This method was successfully applied to identify spatio-temporal clusters in West Nile Virus (WNV) data from the period 2007 to 2017 [6].

The second study utilised spatial analysis using the Global Moran's Index and LISA (Local Indicators of Spatial Autocorrelation). The results of the spatial pattern analysis indicated areas with high concentrations of cases, particularly in North Jakarta, with other hotspots scattered in West Jakarta and Central Jakarta [11].

The third study employed DBSCAN to perform clustering of Covid-19 spread in Palopo City. This clustering process provided information on the total number of clusters, noise data, and areas most affected by the Covid-19 spread. The visualisation results showed that the highest spread of

Covid-19 occurred in the central part of Palopo City, with a concentration in the Wara District [12].

3 Method

The research was conducted in several stages. Figure 2 shows the stages of the research method.

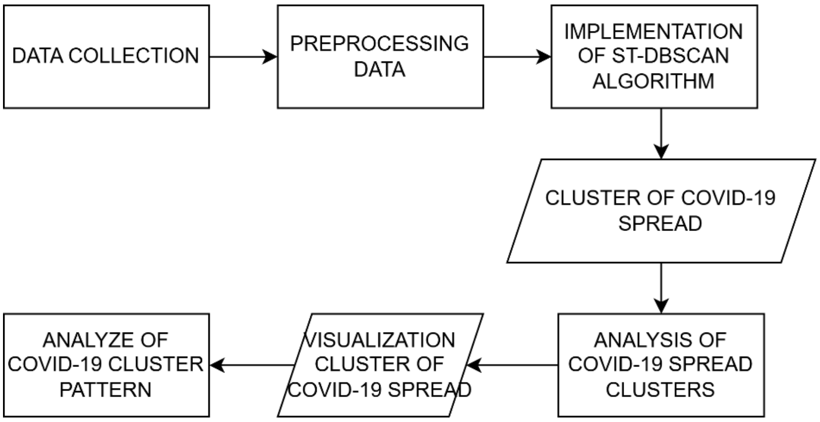

Fig 2. Stages of research

The research began with data collection related to the spread of COVID-19 and the ST-DBSCAN algorithm, both from literature sources and through surveys at the South Sulawesi Provincial Health Office. Next, the collected COVID-19 data was processed to obtain the attributes that would be used in the clustering process. Subsequently, the ST-DBSCAN algorithm was implemented to perform clustering of the COVID-19 spread. In this stage, the necessary parameters for the ST-DBSCAN algorithm, namely Eps1, Eps2, and MinPts, were set. After that, clusters of the COVID-19 spread were formed based on predetermined measurement scenarios, with the results presented in cluster plots.

4 Results and Discussion

4.1 Data

The data used in this research is the spread of Covid-19 cases in Makassar City, South Sulawesi Province in August and September 2021. In the August data, there were 6,917 Covid-19 case data, and in September there were 890 Covid-19 case data. This data was obtained by surveying the P2P Division of the South Sulawesi Provincial Health Office. The attributes of the Covid-19 case data obtained from the South Sulawesi Provincial Health Office can be seen in Table 1.

Table 1. Attributes of Covid-19 Data in Makassar City

Attribute	Description
Report Date	The date the Covid-19 patient data was entered into the Health Office
Regency	The regency/city of origin of the Covid-19 patient
Gender	The gender of the Covid-19 patient
Age	The age of the Covid-19 patient
Residential Address	The address of the confirmed Covid-19 patient

Lab Result Date	The date the patient was confirmed Covid-19 positive
Status	The final status of the Covid-19 patient

Table 1 shows that there are 7 attributes in the Covid-19 case data for Makassar City, namely report date, regency, gender, age, residential address, lab result date, and status. Examples of Covid-19 case data for Makassar City can be seen in Table 2.

Table 2. Example Data of Covid-19 in Makassar City

Report Date	Regency	Gender	Age	Residential Address	Lab Result Date	Status
01/08/2021	Makassar	L	54	Jl. Pengayoman B1 RT 001 RW 004 Kel. Pandang Kec. Panakkukang	30/07/2021	Healed
01/08/2021	Makassar	L	67	Jl. Nikel I Blok A.22 / 26 B	30/07/2021	Healed
01/08/2021	Makassar	L	54	Jl. Permata Hijau No.3 Kel. Gunung Sari Kec. Rappocini	30/07/2021	Healed
01/08/2021	Makassar	P	65	Jl. Hati Suci No. 21	31/07/2021	Die
01/08/2021	Makassar	L	26	TN Wessabbe Blok D No.61	30/07/2021	Healed

Table 2 shows data on several confirmed COVID-19 patients in Makassar City. During August and September 2021, there were 7,427 recovered cases, 155 cases treated in hospitals, 15 cases of self-isolation, and 210 deaths.

4.2 Clustering

Clustering of the COVID-19 spread data is conducted to identify clusters of COVID-19 distribution. Areas with high concentrations of COVID-19 cases will indicate regions that are at higher risk of transmission. Figure 3 presents a plot of the COVID-19 spread in Makassar City, with the distribution data derived from residential address information that has been converted into longitude and latitude coordinates.

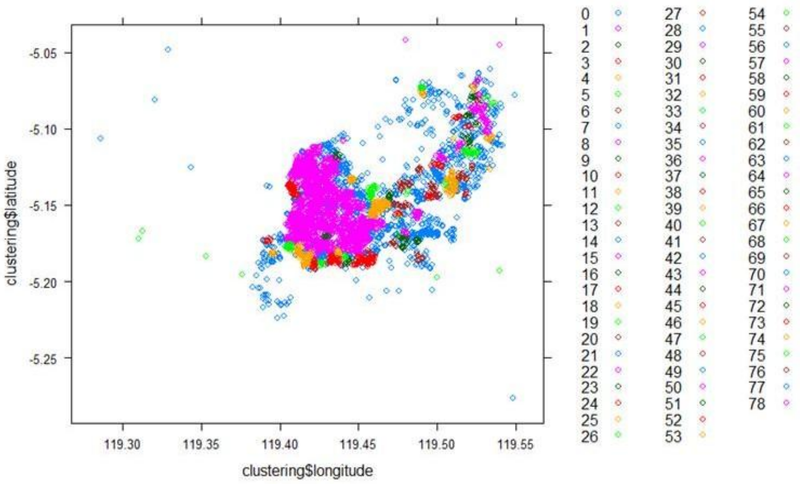

Fig 3. Plot of Covid-19 Spread using ST-DBSCAN

Figure 3 shows the clustering results using the ST-DBSCAN algorithm on COVID-19 data in Makassar City, with the parameters set as Eps1=0.002, Eps2=14, and MinPts=10. The results indicate that there are 78 clusters and 1,639 noise cases, which are marked with a label of 0 during the clustering process. Cluster 1 contains the highest number of COVID-19 cases, totaling 3,556, while the smallest clusters have only 10 cases each. The clusters with exactly 10 cases are clusters 24, 55, 56, 60, 64, 66, 67, 69, 70, 73, 77, and 78.

Clusters that can be analyzed are those with at least 30 cases [9]. Clusters with a minimum of 30 cases are referred to as large clusters, while those with fewer than 30 cases are considered small clusters. Out of the 78 clusters formed, there are 24 large clusters and 54 small clusters. The clusters classified as large clusters are shown in Figure 4.

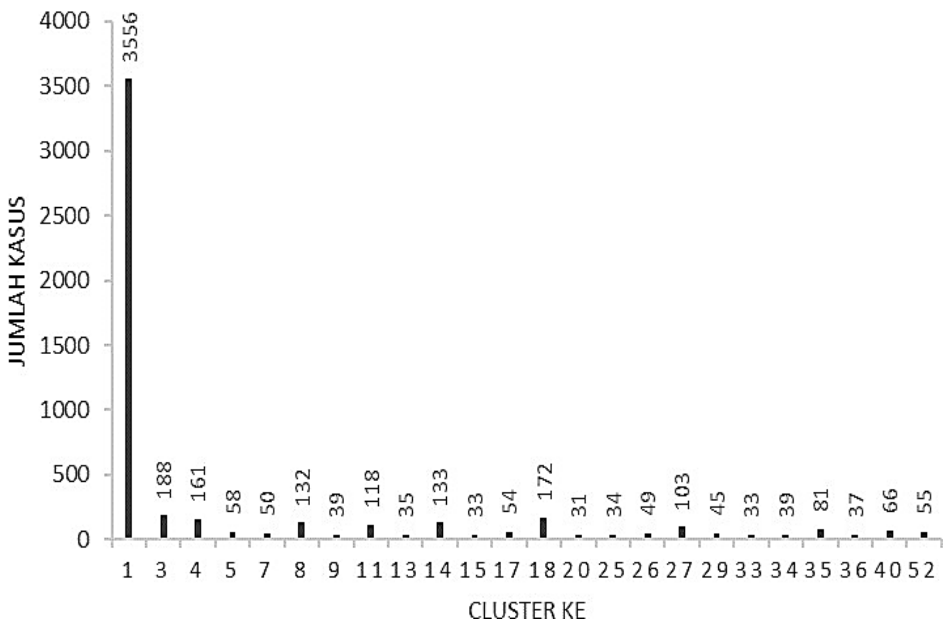

Fig 4. Number of Cases in Large Clusters

Figure 4 illustrates that the highest number of cases in the large clusters is found in Cluster 1, with a total of 3,556 cases. Meanwhile, Cluster 20 is the large cluster with the fewest cases, having only 31 cases.

4.3 Cluster Patterns

The large clusters are then analysed based on their spatio-temporal patterns [9]. The patterns are examined by visualising each time period within each large cluster. In this study, each period within a cluster covers a duration of 14 days (Eps2). An example of Cluster 1 divided by time periods can be seen in Figure 5.

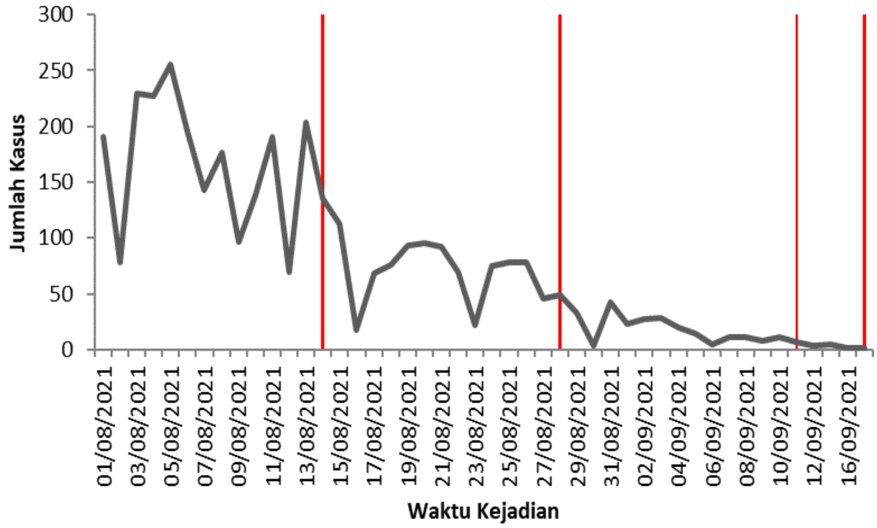

Fig 5. Cluster 1 Event Time Period

Figure 5 indicates that Cluster 1 consists of 4 distinct time periods. The boundaries of these periods within Cluster 1 are marked by red horizontal lines. The highest number of cases in Cluster 1 occurred during the first period, specifically on August 5, 2021.

The pattern analysis reveals that the number of time periods within large clusters varies, each with a different number of cases. There are 6 clusters with 4 time periods: clusters 1, 3, 4, 5, 18, and 29. Cluster 1, which has 4 time periods, is the cluster with the highest number of cases in a single time period, specifically on August 5, 2021, with a total of 255 cases. Cluster 40 has the fewest time periods, occurring only in one time period, which took place on August 3 and 6, during the first period. On August 3, there were 65 cases, while on August 6, there was just 1 case.

The visualisation of the cluster 1 pattern based on the time period of the incident can be seen in Figures 6 to 9. Figure 6 is the result of the visualisation of period 1 in cluster 1.

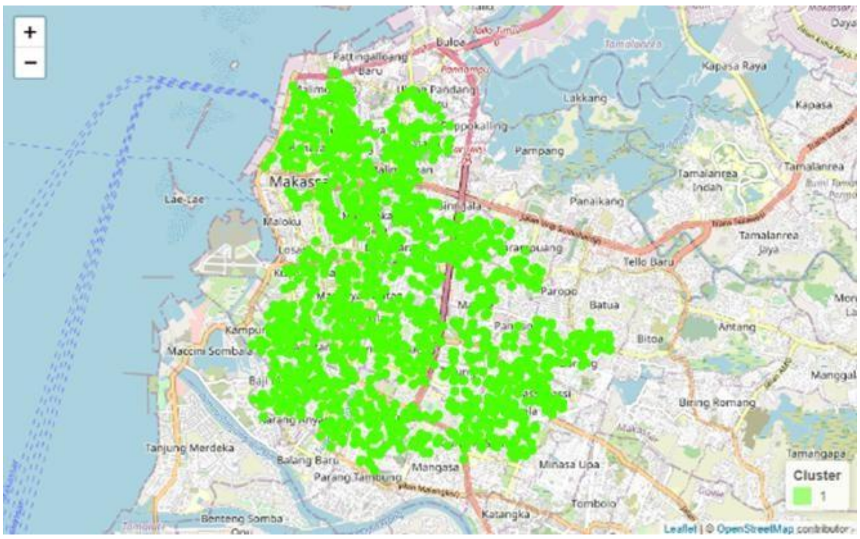

Fig 6. Period 1 Cluster 1

Figure 6 shows that the cases are spread almost throughout the Makassar city area, except in the northeast. The number of cases in period 1 cluster 1 is 2,328 cases. Figure 7 is the result of the visualisation of period 2 in cluster 1.

Fig 7. Period 2 Cluster 1

Figure 7 shows that the distribution of cases is the same as what happened in period 1, the difference is the number of cases. The number of cases in period 2 cluster 1 is 972 cases. Figure 8 is the result of the visualisation of period 3 in cluster 1.

Fig 8. Period 3 Cluster 1

Figure 8 shows that the distribution of cases is still the same as what happened in periods 1 and

2, but there is a decrease in cases. The number of cases in period 3 cluster 1 is 245 cases. Figure 9 is the result of the visualisation of period 4 in cluster 1.

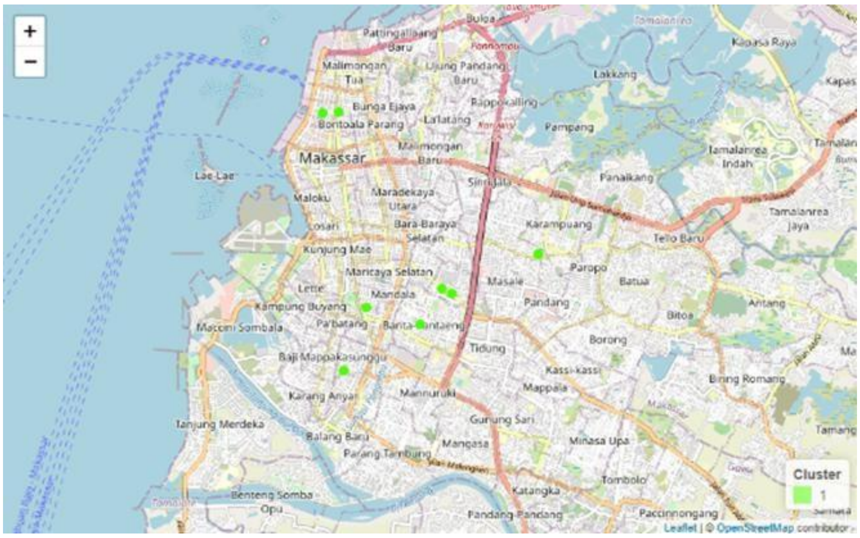

Fig 9. Period 4 Cluster 1

Figure 9 shows that the cases are spread only in a few areas of Makassar city, namely the northern and central parts. The number of cases in period 4 cluster 1 is 11 cases.

Based on the visualization results of cluster 1 in each period, it can be stated that cluster 1 has an occasional pattern. Examples of occasional patterns found based on large clusters from the visualization results in this study can be seen in Table 3.

Table 3. Example of the Distribution of Covid-19 in Makassar City, Occasional Pattern

Cluster	Period	Region (District)	Number of Cases	Information
1	1	Bontoala	131	The most cases are in Rappocini sub-district with a total of 478 cases
		Makassar	256	
		Mamajang	177	
		Manggala	67	
		Mariso	136	
		Panakukkang	282	
		Rappocini	478	
		Tallo	200	
		Tamalate	368	
		Ujung Pandang	96	
		Ujung Tanah	38	
		Wajo	99	

Cluster	Period	Region (District)	Number of Cases	Information
	2	Bontoala	60	The most cases are in Rappocini sub-district with a total of 234 cases
		Makassar	107	
		Mamajang	100	
		Manggala	37	
		Mariso	48	
		Panakukkang	129	
		Rappocini	234	
		Tallo	54	
		Tamalate	90	
		Ujung Pandang	28	
		Ujung Tanah	16	
		Wajo	69	
	3	Bontoala	6	The most cases are in Rappocini sub-district with a total of 73 cases
		Makassar	29	
		Mamajang	34	
		Manggala	14	
		Mariso	17	
		Panakukkang	15	
		Rappocini	73	
		Tallo	9	
		Tamalate	15	
		Ujung Pandang	3	
		Ujung Tanah	3	
		Wajo	27	
	4	Mamajang	1	The most cases are found in Panakukkang, Rappocini and Wajo districts with 3 cases each.
		Panakukkang	3	
		Rappocini	3	
		Tamalate	1	
		Wajo	3	

Table 3 shows that Cluster 1 has an occasional pattern because there is a change in the distribution of regions in one of the time periods. In periods 1 to 3 it occurred in 12 regions, but in period 4 it was only present in 5 regions. A cluster is considered to have a stationary pattern if there is no change in the distribution of regions. Examples of stationary patterns found based on large clusters from the visualization results in this study can be seen in Table 4.

Table 4. Example of the Distribution of Covid-19 in Makassar City, Stationary Pattern

Cluster	Period	Region (District)	Number of Cases
	1	Rappocini	102
	2	Rappocini	64
3	3	Rappocini	19
	4	Rappocini	3

Table 4 shows an example of a stationary pattern found in cluster 3, where in periods 1 to 4 it only occurred in 1 region, namely the Rappocini sub-district. The analysis of 24 large clusters in this study shows that the most common type of spatio-temporal pattern is the stationary pattern. There are 16 stationary patterns and 8 occasional patterns.

5 Conclusions and Suggestions

This study analyses the distribution of COVID-19 cases in the city of Makassar using the Spatio-Temporal Density Based Clustering Algorithm with Noise (ST-DBSCAN). The analysis results show that COVID-19 cases occur almost throughout the city of Makassar, except in the Kepulauan Sangkarrang sub-district. The most frequent cases occurred in the northern part of Makassar, precisely in the Biringkanaya sub-district with 8 clusters, while the largest number of cases was in the southeastern part of Makassar, namely in the Rappocini sub-district with 921 cases. The application of the ST-DBSCAN algorithm resulted in 78 clusters and 1639 noises. From the formed clusters, two spatio-temporal patterns were found, namely the occasional pattern (8 clusters) and the stationary pattern (16 clusters).

For further research, it is recommended to perform more optimal parameter measurements and better cluster evaluation. In addition, the application of this algorithm can also be implemented in a web-based application system to be more dynamic in displaying the distribution of COVID-19 cases and facilitate access to information for the government and the public.

References

[1] A. Kantarci, K. P. Moloney, and S. He, “Big Data Analytics for Covid-19 Pandemic: A Literature Review,” *J. Med. Syst.*, vol. 45, no. 4, pp. 1–12, 2021.
[2] D. Birant and A. Kut, “ST-DBSCAN: An algorithm for clustering spatial-temporal data,” *Data \& Knowl. Eng.*, vol. 60, no. 1, pp. 208–221, 2007.
[3] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu, “Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications,” *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 169–194, 1998.
[4] O. Nicolis, L. Delgado, B. Peralta, M. Díaz, and M. Chiodi, “Space-time clustering of seismic events in Chile using ST-DBSCAN-EV algorithm,” *Environ. Ecol. Stat.*, vol. 31, no. 2, pp. 509–536, 2024, doi: 10.1007/s10651-023-00594-3.
[5] X. Wu, L. Liao, F. Zou, J. Liu, B. Chen, and Y. Zheng, “TS-DBSCAN: To Detect Trajectory Anomaly for Transportation Vehicles BT - Genetic and Evolutionary Computing,”J.-S. Pan, J. C.-W. Lin, Y. Liang, and S.-C. Chu, Eds., Singapore: Springer Singapore, 2020, pp. 151–160.
[6] K. B. Chimwayi and J. Anuradha, “Clustering West Nile Virus Spatiotemporal data using ST-DBSCAN,” *Procedia Comput. Sci.*, vol. 132, pp. 1218–1227, 2018, doi: 10.1016/j.procs.2018.05.037.
[7] X. Li, M. Geng, Y. Peng, L. Meng, and S. Lu, “Molecular immune pathogenesis and diagnosis of COVID-19,” *J. Pharm. Anal.*, vol. 10, no. 2, pp. 102–108, 2020, doi: 10.1016/j.jpha.2020.03.001.
[8] J. F.-W. Chan *et al.*, “A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster,” *Lancet*, vol. 395, no. 10223, pp. 514–523, 2020, doi: 10.1016/S0140-6736(20)30154-9.
[9] C. Pölitz and G. N. Andrienko, “Finding arbitrary shaped clusters with related extents in space and time,” in *1st*

International Symposium on Visual Analytics Science and Technology (EuroVAST) at EuroVis 2010, 2010, pp. 19–25. doi: 10.2312/PE/EuroVAST/EuroVAST10/019-025.

[10] N. A. Indrawan and H. A. Adrianto, "Spatio-Temporal Clustering Hotspot di Sumatera Selatan Tahun 2002-2003 Menggunakan Algoritme ST-DBSCAN dan Bahasa Pemrograman R," *J. Ilmu Komput. dan Agri-Informatika*, vol. 3, no. 2, pp. 112–121, 2016, doi: 10.29244/jika.3.2.112-121.

[11] I. R. Ginting, M. R. Makful, and M. Muhtar, "Pola Penyebaran COVID-19 di DKI Jakarta pada Bulan Maret-Juli Tahun 2020 Secara Spasial," *J. Kedokt. dan Kesehat.*, vol. 17, no. 2, pp. 161–169, 2020.

[12] M. Arfan and M. S. Pebriadi, "Analysis of Covid-19 Spread in Palopo City Using DBSCAN Algorithms," *Inspir. J. Teknol. Inf. Dan Komun.*, vol. 12, no. 2, pp. 1–7, 2022, doi: https://doi.org/10.35585/inspir.v12i2.21.