

Implementasi Metode ST-DBSCAN untuk Pengelompokan Pola Penyebaran Petir di Kota Malang

Hanifatul Mufidah*, Hisyam Fahmi

Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim
Malang, Indonesia

hanifatulmufidah6@gmail.com, hisyam.fahmi@uin-malang.ac.id

Abstrak

Petir merupakan fenomena alam tak terelakkan yang terjadi di atmosfer bumi. Terjadinya fenomena petir ini sangat berbahaya karena energi yang dikeluarkan sangat besar hingga mencapai tiga jutaan volt. Sulit untuk memperkirakan waktu, lokasi, dan intensitasnya, sehingga sambaran petir dapat menimbulkan kerugian fisik karena sering mengakibatkan jatuhnya korban jiwa. Salah satu metode yang dapat digunakan untuk mengidentifikasi suatu kawasan dan waktu yang rawan terjadinya petir adalah dengan teknik *clustering*. Dalam penelitian ini metode *clustering* yang digunakan adalah algoritma ST-DBSCAN (*Spatio Temporal-Density Based Spatial Clustering Application with Noise*), yaitu algoritma *clustering* yang mengelompokkan data berdasarkan aspek spasial dan temporal. Dataset yang digunakan dalam penelitian ini adalah titik-titik petir di Kota Malang pada periode 1 Januari hingga 31 Desember 2022 dengan jumlah 16.800 data. Penelitian ini menghasilkan empat *cluster* dan terdapat 26 terdapat noise, memberikan nilai *Koefisien Silhouette* sebesar 0,104 yang menggunakan parameter seperti jarak spasial ($Eps1 = 0,2$), jarak temporal ($Eps2 = 7$), dan titik minimum anggota dalam kelompok ($MinPts = 7$). Sambaran Petir di Kota Malang pada tahun 2022 ini banyak terjadi di bulan Januari hingga Juli pada *cluster* pertama dengan jumlah 13.337 titik dan paling sedikit terjadi pada bulan Agustus dengan jumlah 110 titik.

Kata kunci: Petir; *Data Mining*; *Clustering*; ST-DBSCAN

Abstract

Lightning is an inescapable natural occurrence in the Earth's atmosphere. Lightning is extremely harmful since the energy released can reach up to three million volts. Lightning strikes are difficult to forecast in terms of time, position, and intensity, therefore they may result in physical losses as they often result in fatalities. One method that can be used to identify an area and time that is prone to lightning is the clustering technique. The clustering approach utilized in this study is the ST-DBSCAN algorithm (*Spatio Temporal-Density Based Spatial Clustering Application with Noise*), which groups data based on spatial and temporal aspects. The dataset used in this study is lightning spots in Malang City from 1 January to 31 December 2022, with a total of 16,800 data. The most accurate analysis findings revealed four clusters and 26 contained noise, giving a Silhouette Coefficient value of 0.104 which employs parameters such as spatial distance ($Eps1 = 0.2$), temporal distance ($Eps2 = 7$), and minimum parts of spots within the group ($MinPts = 7$). Lightning strikes in Malang City in 2022 are anticipated to be frequently encountered between January and July in the first cluster, totaling 13.337 spots and the least occurred in August with 110 spots.

Keywords: Lightning; *Data Mining*; *Clustering*; ST-DBSCAN

PENDAHULUAN

Data mining merupakan pengumpulan dan pemrosesan data dengan tujuan mengekstraksi informasi penting yang terdapat pada data [11]. Pada umumnya terdapat beberapa metode untuk melakukan *data mining* yang terdiri dari *association*, *classification*, *regression*, dan *clustering* [10]. Pentingnya *data mining* semakin terasa dalam era *modern* ini, khususnya seiring peningkatan penggunaan teknologi informasi dan pertumbuhan signifikan dalam pengumpulan data. Berkembangnya teknologi dan perkembangan di bidang *data mining* memberikan peluang yang terus berkembang untuk memperoleh manfaat dari analisis data. Salah satu tujuan *data mining* adalah mengelompokkan data tanpa label berdasarkan kemiripan karakteristiknya, yang disebut sebagai *clustering* [5].

Dalam *data science*, pengelompokan adalah alat yang berguna. *Clustering* adalah teknik untuk menemukan struktur *cluster* dalam kumpulan data yang dibedakan berdasarkan tingkat kesamaan tertinggi di dalam suatu *cluster* dan tingkat ketidaksamaan tertinggi di seluruh *cluster*. Teknik pengelompokan awal yang digunakan oleh para ahli biologi dan ilmuwan sosial adalah *hierarchical clustering*, sedangkan analisis *cluster* berkembang menjadi subbidang *statistical multivariate analysis* [9]. Salah satu metode yang digunakan dalam *clustering* yaitu *ST-DBSCAN* (*Spatio Temporal-Density Based Spatial Clustering Application with Noise*).

ST-DBSCAN merupakan perbaikan dari algoritma *DBSCAN* [1]. Algoritma ini adalah salah satu metode *clustering* berbasis kepadatan dengan kerapatan yang berbeda. *ST-DBSCAN* digunakan untuk mengelompokkan data dalam tiga dimensi, yaitu data spasial dan temporal [6]. Pengelompokan spasial dan temporal adalah bidang studi baru. Banyaknya data geografis yang ditawarkan oleh GPS, satelit, teknologi nirkabel, jaringan sensor, dan perangkat lain yang dapat mengkomunikasikan data posisi dan waktu memungkinkannya menentukan lokasi dan waktu secara tepat [8].

Salah satu data geografis yang tepat untuk pengelompokan menggunakan jarak spasial dan temporal adalah petir. Petir adalah peristiwa kelistrikan udara berupa pelepasan muatan positif dan negatif yang diakibatkan oleh perbedaan potensial antara awan dan bumi [2]. Terjadinya fenomena petir ini sangat berbahaya karena energi yang dikeluarkan sangat besar hingga mencapai jutaan *volt*. Sambaran petir dapat menimbulkan kerugian fisik maupun materi karena sering mengakibatkan jatuhnya korban jiwa.

Sangat sedikit penelitian yang dilakukan tentang petir hingga saat ini. Dibuktikan dengan penelitian sebelumnya terkait analisis kebutuhan sistem proteksi sambaran petir pada gedung bertingkat [7]. Penelitian lain terkait pengelompokan titik api kebakaran hutan Indonesia. Penelitian ini menggunakan metode *clustering* yaitu *ST-DBSCAN* dengan hasil pengujian menunjukkan bahwa terdapat 1.540 cluster di Indonesia dari tahun 2015 hingga 2020 [6]. Penelitian selanjutnya terkait analisis pola penyebaran petir di Pasuruan dengan menggunakan metode *DBSCAN* pada tahun 2020 dengan hasil terdapat 2 *cluster* dengan 15 titik *noise* [4].

Berdasarkan pemaparan di atas, penulis termotivasi untuk melakukan penelitian “Implementasi Metode *ST-DBSCAN* untuk Pengelompokan Pola Penyebaran Petir di Kota Malang”. Diharapkan hasil penelitian ini akan mendorong penelitian tambahan yang membahas topik terkait petir. Hal ini sangat bermanfaat bagi kemajuan teknologi *modern*.

METODE PENELITIAN

Data dan Sumber Data

Data yang digunakan dalam penelitian ini ialah data sekunder. Data petir yang merupakan jenis data sekunder penelitian ini bersumber dari Badan Meteorologi, Klimatologi, dan Geofisika

(BMKG) Stasiun Geofisika Klas II, Melaten Kabupaten Pasuruan. Data ini diambil dari deteksi petir yang terdapat di BMKG Stasiun Geofisika Klas II, Melaten Kabupaten Pasuruan. Deteksi petir ini mencatat secara *realtime* waktu terjadinya petir, jenis petir, koordinat petir, dan tempat terjadinya petir. Data yang didapatkan adalah data petir di seluruh wilayah Jawa Timur, kemudian dipilih titik objek yang ada pada koordinat kota Malang saja dengan batas garis bujur utara 112.5702° hingga 112.7816° dan garis lintang selatan 7.8623° hingga 8.1128° dari tanggal 1 Januari hingga 31 Desember pada tahun 2022 dengan jumlah 16.800 data.

Tahapan Penelitian

1. Mengumpulkan data sambaran petir beserta parameter terjadinya petir dengan 13 fitur.
2. *Exploratory data analysis* dilakukan deskripsi fitur pada data untuk mengetahui gambaran umum terkait faktor yang berpengaruh pada pola persebaran terjadinya petir.
3. *Preprocessing data* yang akan dilakukan pada penelitian ini melalui beberapa tahapan sebagai berikut:
 - a. *Data Reduction*
 - b. *Data Transformation*
4. Penentuan nilai parameter Epsilon 1, Epsilon 2, dan *Minimum Points*.
5. Implementasi algoritma *ST-DBSCAN* menggunakan *tools python*.
6. Evaluasi hasil *cluster* menggunakan nilai *silhouette coefficient*.

HASIL DAN PEMBAHASAN

Exploratory Data Analysis

Exploratory Data Analysis adalah langkah awal dalam *data science* yang digunakan untuk memahami data. Proses ini menggunakan teknik visual dan statistik untuk menemukan pola dan tren yang dapat digunakan untuk membuat keputusan. Pada penelitian ini hasil perhitungan *statistics descriptive data* diperoleh dengan program *python* seperti Tabel 1 berikut.

Tabel 1 Hasil *Statistics Descriptive Data*

Fitur	Minimum	Maximum	Mean	Std. Deviation	Variance
Bujur	112,57026	112,78157	112,69354	0,056302	0,00317
Lintang	7,86230	8,11278	7,95951	0,06978	0,00487
Datetime	44562,624	44923,766	44712,030	91,7938	8426,108

Pada Tabel 1 didapatkan hasil pada 1 Januari hingga 31 Desember tahun 2022, di wilayah Kota Malang terjadinya petir diambil sampel sejumlah 16.800. Simpangan Baku (*Standart Deviation*) pada Garis Bujur adalah 0,056302 dan pada Garis Lintang adalah 0,06978 yang menandakan bahwa sangat besar persebaran datanya. Sedangkan *variance* yang diperoleh dari data tersebut adalah 0,00317 pada Garis Bujur dan 0.00487 pada Garis Lintang yang menunjukkan bahwa dari data tersebut memiliki keragaman yang sangat besar.

Data Preprocessing

1. Data Transformation

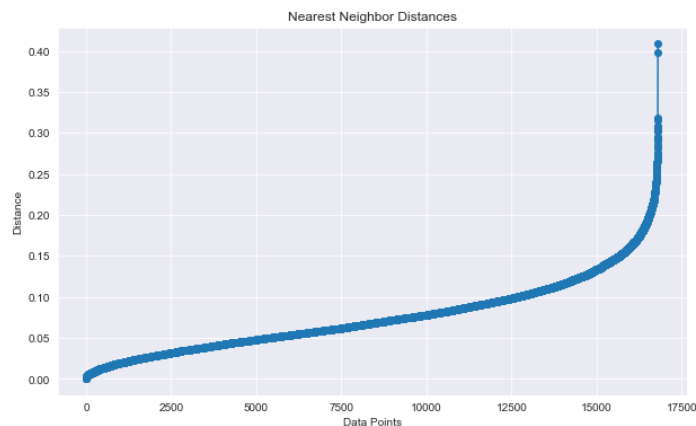
Data transformation dilakukan untuk merubah format pada kolom waktu terjadinya petir menjadi numerik. Tahapan ini dilakukan penggabungan data tanggal, bulan, dan tahun menjadi *Date*. Kemudian penggabungan data jam, menit, dan detik menjadi *Time*. Dari kedua data *Date* dan *Time* tersebut akan digabungkan menjadi data *Datetime* dalam format numerik.

2. Data Reduction

Pada tahap ini dilakukan pengurangan jumlah fitur yang tidak diperlukan dalam data yang telah ada. Adapun data yang tidak diperlukan dalam penelitian ini adalah pada fitur FID_1, *Source*, Tanggal, Jenis, tg, bl, th, jm, mn, det, NAME_1, *Date*, *Time*. Fitur yang tersisa dalam data ini adalah Bujur, Lintang dan *Datetime*.

Penentuan Parameter

Pada penelitian yang menggunakan metode *ST-DBSCAN*, diperlukan beberapa parameter, yaitu *Eps1*, *Eps2*, dan *MinPts*. Untuk menentukan nilai *Eps1*, digunakan menggunakan nilai spasial yaitu pada fitur garis bujur dan lintang. Dalam penentuan nilai *Eps1* yang optimal, dapat digunakan grafik *k-dist* yang menggunakan KNN. Grafik *k-dist* menunjukkan hubungan antara sumbu x yang merupakan ketetangaan pada setiap titik data dengan sumbu y merupakan nilai jarak ketetangaan masing-masing titik. Untuk memperoleh nilai *Eps1*, setiap titik tersebut diurutkan berdasarkan jaraknya. Pada penelitian ini nilai *k* yang diambil adalah 7, yang artinya menghitung rata-rata pada 7 titik ketetangaan pada sumbu x dan akan dihasilkan jarak KNN pada sumbu y. Tahap ini dilakukan hingga perhitungan semua titik selesai. Kemudian untuk menentukan nilai *Eps1* yang didapatkan secara optimal, maka dengan melihat perbedaan jarak yang signifikan yang dapat dilihat dari garis kelengkungannya. Sedemikian sehingga diperoleh grafik berikut pada Gambar 1.



Gambar 1 Grafik *k-dist* Untuk Penentuan Parameter *Eps1*

Pada Gambar 1 tersebut dengan mengambil nilai $k=7/\text{MinPts}=7$, nilai *Eps1* yang diperoleh pada garis dengan kelengkungan yang signifikan adalah pada rentang 0,15 – 0,25. Sedangkan nilai *Eps2* yang ditentukan yaitu 1 untuk mendapatkan pola *clustering* dalam runtun waktu 1 hari dan $\text{Eps2} = 7$ untuk mendapatkan pola *clustering* dalam runtun waktu 7 hari. Parameter Δ_ϵ merupakan

aspek non spasial yang digunakan untuk meningkatkan *cohesion* intra-*cluster* dan separation antar-*cluster* tidak digunakan dalam penelitian ini karena tidak tercantum fitur pada *dataset*nya.

Implementasi Metode ST-DBSCAN

Setelah dilakukan eksperimen terhadap data petir di Kota Malang dengan parameter yang telah ditetapkan, maka hasil yang diperoleh dari eksperimen tersebut akan dievaluasi menggunakan *silhouette coefficient*. Sehingga diperoleh hasil *cluster* pada eksperimen data petir di Kota Malang pada Tabel 2 berikut.

Tabel 2 Hasil *Cluster* Pada Eksperimen Data Petir

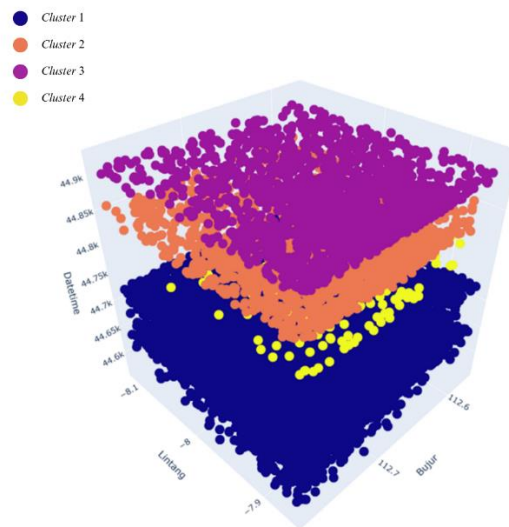
Eps1	Eps2	MinPts	Jumlah <i>Cluster</i>	Hasil	
				<i>Noise</i>	<i>Silhouette Coefficient</i>
0,15	1	5	10	41	-0,196
0,15	7	5	7	29	-0,089
0,15	1	7	5	73	-0,004
0,15	7	7	5	42	0,007
0,15	1	11	4	151	0,066
0,15	7	11	4	75	0,090
0,175	1	5	6	23	-0,072
0,175	7	5	7	10	-0,089
0,175	1	7	5	58	-0,012
0,175	7	7	5	32	0,006
0,175	1	11	6	90	-0,042
0,175	7	11	6	41	-0,068
0,2	1	5	9	16	-0,093
0,2	7	5	7	9	-0,042
0,2	1	7	5	38	-0,005
0,2	7	7	4	26	0,104
0,2	1	11	5	62	-0,029
0,2	7	11	5	33	0,008
0,225	1	5	5	10	0,030
0,225	7	5	4	4	0,093
0,225	1	7	5	25	-0,024

Eps1	Eps2	MinPts	Jumlah Cluster	Hasil	
				Noise	Silhouette Coefficient
0,225	7	7	4	5	0,100
0,225	1	11	4	51	0,057
0,225	7	11	4	17	0,091
0,25	1	5	4	6	0,078
0,25	7	5	4	3	0,084
0,25	1	7	4	13	0,059
0,25	7	7	4	5	0,100
0,25	1	11	4	35	0,049
0,25	7	11	4	11	0,076

Pada Tabel 2 dapat menunjukkan bahwa parameter terbaik dengan hasil nilai *silhouette coefficient* tertinggi yaitu senilai 0,104 dan memiliki 26 *noise* dengan 4 jumlah *cluster* adalah Eps1 = 0,2, Eps2 = 7, dan MinPts = 7 yang memiliki hasil yang sama. Sehingga parameter tersebut merupakan hasil parameter terbaik yang diperoleh pada penelitian ini.

Analisis Hasil Cluster

Dari Tabel 2, hasil eksperimen terbaik menggunakan algoritma *ST-DBSCAN Clustering* pada data petir yang terdapat di Kota Malang dapat divisualisasikan seperti pada Gambar 2 berikut.

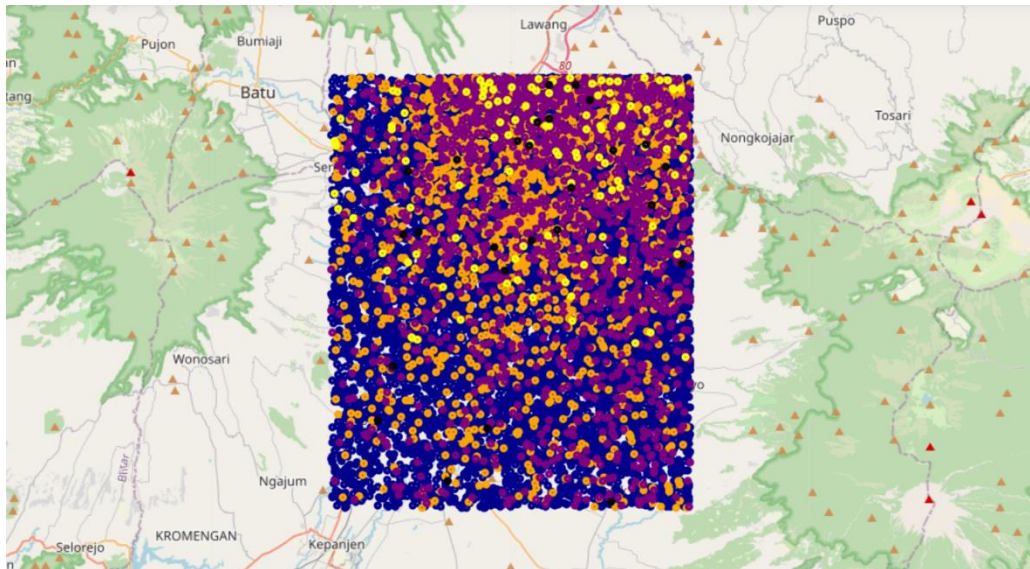


Gambar 2 Hasil Cluster Terbaik

Gambar 2 menunjukkan hasil pengelompokan dari *ST-DBSCAN Clustering* dengan $Eps_1 = 0,2$, $Eps_2 = 7$, dan $MinPts = 7$. Dari eksperimen yang telah dilakukan tersebut mendapatkan hasil dengan 4 jumlah *cluster* sebagai berikut:

1. *Cluster* pertama yang diwakili dengan tanda berwarna biru dengan jumlah 13.337 titik yang terjadi pada bulan Januari hingga Juli.
2. *Cluster* kedua dengan tanda berwarna oranye yang berjumlah 1.717 titik yang terjadi pada bulan September hingga November.
3. *Cluster* ketiga dengan tanda berwarna ungu memiliki anggota 1.610 titik yang terjadi pada bulan Desember.
4. *Cluster* keempat dengan tanda berwarna kuning memiliki anggota 110 titik yang terjadi pada bulan Agustus.

Jumlah *noise* pada penelitian ini adalah 26 titik dan diperoleh nilai *silhouette coefficient* tertinggi yaitu senilai 0,104. Dapat dilihat juga bahwa hasil tersebut terkelompok secara temporal, sedangkan jika dilihat secara spasial tidak terkelompok dengan jelas. Visualisasi tersebut akan dipaparkan sebagaimana pada Gambar 3 berikut.



Gambar 3 Hasil *Cluster* Dilihat Secara Spasial Pada Semua *Cluster*

Gambar 5 menunjukkan hasil visualisasi dari pengelompokan dari *ST-DBSCAN Clustering* dengan $Eps_1 = 0,2$, $Eps_2 = 7$, dan $MinPts = 7$ yang dilihat dari pengelompokan secara spasial. Gambar tersebut terlihat bahwa hasil dari pengelompokan yang dilihat secara spasial tidak terkelompokkan secara jelas. Hal tersebut disebabkan oleh volume data petir yang besar.

Berdasarkan data Badan Meteorologi, Klimatologi, dan Geofisika Karangploso 2022, jumlah curah hujan di Kota Malang pada tahun 2022 terbanyak yaitu pada bulan Oktober dengan jumlah 496,1 milimeter. Sedangkan paling sedikit pada bulan Juli dengan jumlah 20,3 milimeter. Hal tersebut tidak sama dengan hasil *cluster* pada penelitian ini yang mengelompokkan kejadian petir di Kota Malang pada tahun 2022 [3].

KESIMPULAN

Berdasarkan pembahasan, maka dapat disimpulkan bahwa hasil dari *clustering* data petir yang berbentuk *spatio-temporal* pada tanggal 1 Januari hingga 31 Desember pada tahun 2022 di Kota Malang menggunakan algoritma *ST-DBSCAN* didapatkan parameter terbaik yakni Eps 1 bernilai 0,104, Eps 2 bernilai 7, dan MinPts bernilai 7 yang menghasilkan pengelompokan terlihat jelas pada aspek temporal dan tidak secara jelas pada aspek spasial. Nilai *silhouette coefficient* tertinggi yang diperoleh yaitu senilai 0,104 dan memiliki 26 *noise* dengan jumlah *cluster* sebanyak 4 *cluster*. Kejadian Petir di Kota Malang pada tahun 2022 banyak terjadi di bulan Januari hingga Juli dengan jumlah 13.337 titik, sedangkan paling sedikit terjadi pada bulan Agustus dengan jumlah 110 titik.

DAFTAR PUSTAKA

- [1] Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data & knowledge engineering*, 60(1), 208-221.
- [2] BMKG. (2021). KAJIAN GEOFISIKA. https://statgeof_bandung.bmkg.go.id/. Diakses pada tanggal 2 Oktober 2023.
- [3] BPS Kota Malang. (2023). Jumlah Curah hujan di Kota Malang (milimeter(mm)), 2022. <https://malangkota.beta.bps.go.id/id/statistics-table/2/NTA4IzI=/jumlah-curah-hujan-di-kota-malang.html>. Diakses pada tanggal 29 Mei 2023.
- [4] Fahamsyah, M. (2020). *Metode DBSCAN clustering untuk analisis pola penyebaran petir di Pasuruan* (Undergraduate thesis, Universitas Islam Negeri Maulana Malik Ibrahim).
- [5] Han, J., Kamber, M., & Pei, J. (2012). Data mining: concepts and techniques, Waltham, MA. *Morgan Kaufman Publishers*, 10, 978-1.
- [6] Johar, A., Vatesia, A., & Donny, I. A. (2023). Implementasi Metode Spatio Temporal Clustering Dengan Algoritma ST-DBSCAN Pada Titik Api Kebakaran Hutan Indonesia (2015-2020). *Rekursif: Jurnal Informatika*, 11(1), 1-9.
- [7] Karta, A., Agung, A. I., & Widartono, M. (2020). Analisis Kebutuhan Sistem Proteksi Sambaran Petir Pada Gedung Bertingkat. *Jurnal Teknik Elektro*, 9(3), 773-780.
- [8] Rus, A. M. M., Othman, Z. A., Bakar, A. A., & Zainudin, S. (2022). A Hierarchical ST-DBSCAN with Three Neighborhood Boundary Clustering Algorithm for Clustering Spatio-temporal Data. *International Journal of Advanced Computer Science and Applications*, 13(12).
- [9] Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE access*, 8, 80716-80727.
- [10] Suyanto, D. (2019). Data Mining Untuk Klasifikasi dan Klasterisasi Data. Bandung: Penerbit Informatika.
- [11] Tan, P. N., Steinbach, M., & Kumar, V. (2013). Data mining cluster analysis: basic concepts and algorithms. *Introduction to data mining*, 487, 533.