

**CLUSTERING GEMPABUMI DI WILAYAH REGIONAL VII
MENGUNAKAN METODE DBSCAN DAN K-MEANS**

THESIS

**Oleh:
IHSAN BAGUS FAHAD ARAFAT
NIM. 210605220013**



**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**CLUSTERING GEMPABUMI DI WILAYAH REGIONAL VII
MENGUNAKAN METODE DBSCAN DAN K-MEANS**

THESIS

**Diajukan Kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh:
IHSAN BAGUS FAHAD ARAFAT
NIM. 210605220013**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**CLUSTERING GEMPABUMI DI WILAYAH REGIONAL VII
MENGUNAKAN METODE DBSCAN DAN K-MEANS**

THESIS

**Diajukan Kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh:
IHSAN BAGUS FAHAD ARAFAT
NIM. 210605220013**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

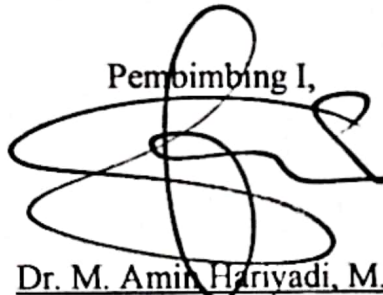
**CLUSTERING GEMPABUMI DI WILAYAH REGIONAL VII
MENGUNAKAN METODE DBSCAN DAN K-MEANS**

THESIS

**Oleh:
IHSAN BAGUS FAHAD ARAFAT
NIM. 210605220013**

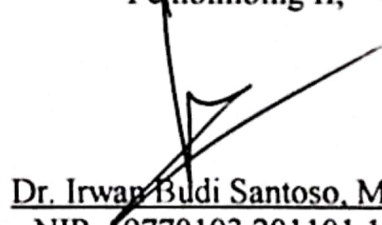
Telah diperiksa dan disetujui untuk diuji:
Tanggal: 23 Mei 2023

Pembimbing I,



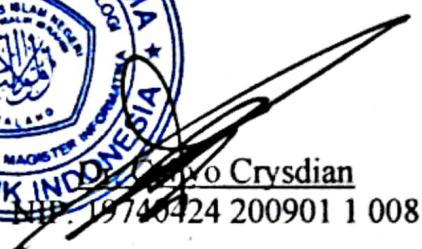
Dr. M. Amin Hariyadi, M.T
NIP. 19670118 200501 1 001

Pembimbing II,



Dr. Irwan Budi Santoso, M.Kom
NIP. 19770103 201101 1 004

Mengetahui,
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Crysdyan
NIP. 19740424 200901 1 008

**CLUSTERING GEMPABUMI DI WILAYAH REGIONAL VII
MENGUNAKAN METODE DBSCAN DAN K-MEANS**

THESIS

Oleh:
IHSAN BAGUS FAHAD ARAFAT
NIM. 210605220013

Telah Dipertahankan di Depan Dewan Penguji Thesis
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Magister Komputer (M.Kom)
Tanggal: 23 Mei 2023

Susunan Dewan Penguji

Penguji Utama : Dr. Ririen Kusumawati, M.Kom
NIP. 19720309 200501 2 002

Ketua Penguji : Dr. Fresy Nugroho, M.T
NIP. 19710722 201101 1 001

Sekretaris Penguji : Dr. M. Amin Hariyadi, M.T
NIP 19670118 200501 1 001

Anggota Penguji : Dr. Irwan Budi Santoso, M.Kom
NIP 19770103 201101 1 004

Tanda Tangan



Mengetahui dan Mengesahkan
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Ihsan Bagus Fahad Arafat
NIM : 210605220013
Program Studi : Magister Informatika
Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa Thesis yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila di kemudian hari terbukti atau dapat dibuktikan Thesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 23 Mei 2023

Yang membuat pernyataan,



Ihsan Bagus Fahad Arafat

NIM. 210605220013

MOTTO

“Fokus pada setiap prioritas adalah salah satu kunci keberhasilan”

PERSEMBAHAN

Alhamdulillah penulis ucapkan kepada Allah SWT atas nikmat-Nya sehingga penulis bisa menyelesaikan Thesis ini yang penulis persembahkan kepada:

1. Keluarga tercinta, orang tua, istri dan anak-anak serta kakak dan adik penulis yang selalu memberikan doa dan dukungannya.
2. Seluruh sivitas Universitas Islam Negeri Maulana Malik Ibrahim Malang khususnya Fakultas Sains dan Teknologi Program Studi Magister Informatika atas ilmu, masukan dan pengalamannya di bidang informatika dan agama Islam.
3. Segenap pimpinan dan rekan-rekan di Stasiun Geofisika Nganjuk yang selalu memberikan motivasi.
4. Seluruh rekan-rekan angkatan 5 Program Studi Magister Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang atas kebersamaannya.
5. Serta seluruh rekan-rekan yang tidak bisa penulis sebutkan satu per satu yang telah membantu secara materiil maupun moril dalam penyelesaian Thesis ini.

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Alhamdulillah penulis sampaikan kepada Allah SWT yang telah memberikan Nikmat, Rahmat dan Hidayah-Nya, sehingga penulis dapat menyelesaikan studi Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan Thesis ini dengan baik.

Selanjutnya, penulis sampaikan terima kasih *Alhamdulillah jaza kumullahu khaira* kepada semua pihak yang telah membantu terselesaikannya Thesis ini. Ucapan terima kasih penulis sampaikan kepada:

1. Bapak Dr. M. Amin Hariyadi, M.T dan Bapak Dr. Irwan Budi Santoso, M.Kom selaku dosen pembimbing Thesis, yang telah banyak memberikan pengarahan, saran, masukan dan pengalaman yang berharga.
2. Segenap sivitas akademika Program Studi Magister Informatika, khususnya seluruh Bapak/Ibu dosen, atas ilmu dan bimbingannya.
3. Seluruh teman seperjuangan angkatan 5 Program Studi Magister Informatika, yang senantiasa memberikan dukungan dan semangat.
4. Keluarga tercinta, orang tua, istri dan anak-anak serta kakak dan adik penulis yang selalu memberikan doa dan dukungannya.
5. Segenap pimpinan dan rekan-rekan di Stasiun Geofisika Nganjuk yang selalu memberikan motivasi.
6. Semua pihak yang ikut membantu dalam menyelesaikan Thesis ini.

Penulis menyadari bahwa dalam penyusunan Thesis ini masih terdapat kekurangan dan penulis berharap semoga Thesis ini dapat memberikan manfaat kepada para pembaca dan bagi penulis secara pribadi.

Wassalamu'alaikum Wr. Wb.

Malang, Mei 2023

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiii
DAFTAR ALGORITMA	xv
ABSTRAK	xvi
ABSTRACT	xvii
الخلاصة	xviii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Pernyataan Masalah	4
1.3. Tujuan Penelitian	5
1.4. Manfaat Penelitian	5
1.5. Ruang Lingkup Penelitian	6
BAB II STUDI PUSTAKA	7
2.1. <i>Clustering</i> Gempabumi	7
2.2. Kerangka Teori	14
BAB III METODOLOGI PENELITIAN	20
3.1. Desain Penelitian	20

3.1.1. Pengumpulan Data.....	20
3.1.2. Desain Sistem	24
3.1.3. Eksperimen	29
3.2. Instrumen Penelitian	43
BAB IV METODE DBSCAN	44
4.1. Desain	44
4.2. Uji Coba	46
4.3. Hasil <i>Clustering</i>	52
4.4. Kesimpulan	58
BAB V METODE K-MEANS	60
5.1. Desain	60
5.2. Uji Coba	62
5.2.1. <i>Clustering K-Means k</i> Optimal dengan <i>Euclidean Distance</i>	62
5.2.2. <i>Clustering K-Means k</i> Optimal dengan <i>Manhattan Distance</i>	63
5.3. Uji <i>k Cluster</i>	64
5.4. Hasil <i>Clustering</i>	64
5.5. Kesimpulan	69
BAB VI PEMBAHASAN.....	71
6.1. Perbandingan Performa Algoritma	71
6.2. <i>Clustering</i> Gempabumi dalam Pandangan Al-Qur'an.....	79
BAB VII KESIMPULAN	81
7.1. Kesimpulan	81
7.2. Saran	82
DAFTAR PUSTAKA	83

DAFTAR GAMBAR

Gambar 2.1 Kerangka Teori Penelitian.....	14
Gambar 3.1 Desain Penelitian.....	20
Gambar 3.2 Peta Seismisitas Indonesia hasil relokasi <i>teletomoDD</i> (Puslitbang PUPR, 2017)	23
Gambar 3.3 Sebaran episenter gempabumi hasil relokasi <i>teletomoDD</i> di Pulau Jawa dan sekitarnya untuk event gempa dengan kedalaman ≤ 100 km dan magnitudo ≥ 4 (Puslitbang PUPR, 2017)	24
Gambar 3.4 Desain Sistem Penelitian.....	25
Gambar 3.5 Pseudocode Algoritma DBSCAN (Schubert et al., 2017)	26
Gambar 3.6 Pseudocode Algoritma <i>K-Means</i> (Yedla, Madhu, Srinivasa Rao Pathakota, 2010)	28
Gambar 3.7 Kurva <i>k-dist</i> perhitungan nilai epsilon (ϵ) <i>Euclidean Distance</i>	32
Gambar 3.8 Kurva <i>k-dist</i> perhitungan nilai epsilon (ϵ) <i>Manhattan Distance</i>	35
Gambar 3.9 Kurva nilai <i>silhouette</i> terhadap nilai <i>k</i> optimal dengan <i>Euclidean Distance</i>	39
Gambar 4.1 Diagram Alir Algoritma DBSCAN.....	44
Gambar 4.2 Peta <i>Clustering</i> Gempabumi Tiap Periode dengan <i>K</i> Optimal menggunakan <i>Euclidean Distance</i>	55
Gambar 4.3 Peta <i>Clustering</i> Gempabumi Tiap Periode dengan <i>K</i> Optimal menggunakan <i>Manhattan Distance</i>	58
Gambar 5.1 Diagram Alir Algoritma <i>K-Means</i>	61
Gambar 5.2 Peta <i>Clustering</i> Gempabumi Tiap Periode dengan <i>K</i> Optimal menggunakan <i>Euclidean Distance</i>	66
Gambar 5.3 Peta <i>Clustering</i> Gempabumi Tiap Periode dengan <i>K</i> Optimal menggunakan <i>Manhattan Distance</i>	69

Gambar 6.1 Peta sebaran episenter gempabumi Pulau Jawa hasil relokasi <i>teletomoDD</i> katalog PuSGeN 2017	77
Gambar 6.2 Hasil <i>Clustering</i> gempabumi DBSCAN dgn <i>Euclid. Distance</i> $T_5_{2017-2021}$	77

DAFTAR TABEL

Tabel 2.1 Daftar Literatur	15
Tabel 3.1 Data Parameter Gempabumi	21
Tabel 3.2 Hasil <i>clustering</i> DBSCAN <i>Euclidean Distance</i> dengan ϵ dan MinPts optimal	33
Tabel 3.3 Hasil <i>clustering</i> DBSCAN <i>Manhattan Distance</i> dengan ϵ dan MinPts optimal	36
Tabel 3.4 Hasil <i>clustering</i> <i>K-Means Euclidean Distance</i> dengan k optimal.....	40
Tabel 4.1 Jumlah Kejadian Gempabumi Berdasarkan Periode Waktu	47
Tabel 4.2 Nilai Epsilon (ϵ) dan MinPts Optimal Berdasarkan Periode Waktu dengan <i>Euclidean Distance</i>	47
Tabel 4.3 Hasil <i>Clustering</i> DBSCAN Berdasarkan Periode Waktu	48
Tabel 4.4 Nilai <i>Silhouette Coefficient</i> DBSCAN Berdasarkan Periode Waktu	49
Tabel 4.5 Nilai Epsilon (ϵ) dan MinPts Optimal Berdasarkan Periode Waktu dengan <i>Manhattan Distance</i>	50
Tabel 4.6 Hasil <i>Clustering</i> DBSCAN Berdasarkan Periode Waktu	51
Tabel 4.7 Nilai <i>Silhouette Coefficient</i> DBSCAN Berdasarkan Periode Waktu	52
Tabel 4.8 Hasil <i>Clustering</i> Algoritma DBSCAN <i>Euclidean Distance</i>	52
Tabel 4.9 Hasil <i>Clustering</i> Algoritma DBSCAN <i>Manhattan Distance</i>	53
Tabel 5.1 Nilai Optimal k <i>cluster</i> Berdasarkan Periode Waktu dengan <i>Euclidean Distance</i>	63
Tabel 5.2 Nilai Optimal k <i>cluster</i> Berdasarkan Periode Waktu dengan <i>Manhattan Distance</i>	63
Tabel 5.3 Nilai <i>Silhouette Coefficient</i> Berdasarkan Periode Waktu	64
Tabel 6.1 Perbandingan hasil <i>clustering</i> DBSCAN <i>Euclidean</i>	

dan <i>Manhattan Distance</i>	73
Tabel 6.2 Perbandingan hasil clustering <i>K-Means Euclidean</i> dan <i>Manhattan Distance</i>	74
Tabel 6.3 Perbandingan hasil <i>clustering</i> DBSCAN dan <i>K-Means Euclidean</i> dan <i>Manhattan Distance</i>	75
Tabel 6.4 Hasil <i>clustering</i> DBSCAN <i>Euclidean Dist</i> T ₅ ₂₀₁₇₋₂₀₂₁ Per Wilayah di Regional VII.....	78

DAFTAR ALGORITMA

Algoritma 1 Library <i>Python</i> yang digunakan pada <i>clustering</i> DBSCAN.....	31
Algoritma 2 Perintah untuk menentukan jumlah N terdekat dengan KNN <i>Euclidean Distance</i>	31
Algoritma 3 Perintah untuk menentukan nilai epsilon (ϵ), MinPts dan <i>Silhouette Score</i> terbaik	32
Algoritma 4 Perintah <i>clustering</i> DBSCAN dengan eps dan MinPts optimal	33
Algoritma 5 Perintah perhitungan <i>Silhouette Coefficient</i>	34
Algoritma 6 Perintah untuk menentukan jumlah N terdekat dengan KNN <i>Manhattan Distance</i>	35
Algoritma 7 Perintah untuk menentukan nilai epsilon (ϵ), MinPts dan <i>Silhouette Score</i>	36
Algoritma 8 Library <i>Python</i> yang digunakan pada <i>clustering</i> <i>K-Means</i>	38
Algoritma 9 Perintah untuk menentukan jumlah k optimal dengan <i>Euclidean Distance</i>	38
Algoritma 10 Perintah untuk menentukan nilai k optimal berdasarkan <i>Silhouette Score</i> terbaik	39
Algoritma 11 Perintah <i>clustering</i> <i>K-Means</i> dengan k optimal	40
Algoritma 12 Perintah perhitungan <i>Silhouette Coefficient</i>	41
Algoritma 13 Library <i>Python</i> yang digunakan pada <i>clustering</i> <i>K-Means</i> dengan <i>Manhattan Distance</i>	42
Algoritma 14 Perintah untuk menentukan nilai k optimal berdasarkan <i>maximum correct centroid</i>	42

ABSTRAK

Arafat, Ihsan B. F. 2023. **Clustering Gempabumi di Wilayah Regional VII Menggunakan Metode DBSCAN dan K-Means**. Thesis. Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. M. Amin Hariyadi, M.T.(II) Dr. Irwan Budi Santoso, M.Kom.

Kata Kunci: *unsupervised learning, clustering, gempabumi, DBSCAN, K-Means, silhouette coefficient*.

Wilayah Regional VII meliputi Provinsi Jawa Tengah, Yogyakarta dan Jawa Timur merupakan wilayah tektonik aktif karena terletak di wilayah zona subduksi lempeng Indo-Australia dan Eurasia serta terdapat beberapa patahan aktif di daratan. Oleh karena itu, perlu dilakukan *clustering* gempabumi untuk memetakan zona potensi gempabumi di wilayah Regional VII berdasarkan *cluster* gempabumi yang terbentuk. Penelitian ini menggunakan pendekatan algoritma *Unsupervised Learning* berbasis kepadatan yaitu, *Density Based Spatial Clustering of Application with Noise* atau DBSCAN. Algoritma ini memerlukan parameter input epsilon (ϵ) dan MinPts. Hasil yang terbentuk selanjutnya dibandingkan dengan hasil *clustering* menggunakan algoritma *K-Means* dan dievaluasi menggunakan *Silhouette Coefficient* serta dibandingkan dengan *golden standart* peta sebaran episenter gempabumi Pulau Jawa hasil relokasi *teletomoDD* dari katalog PuSGeN 2017. Hasil *clustering* menggunakan DBSCAN diperoleh jumlah *cluster* sebanyak 3 hingga 6 *cluster* dengan nilai *Silhouette Coefficient* terendah 0,208 untuk *cluster* gempabumi periode $T_{5_2017-2021}$ dan tertinggi sebesar 0,499 untuk periode T_{1_2020} . Sedangkan algoritma *K-Means* menghasilkan *cluster* sebanyak 3 hingga 7 *cluster* dengan nilai *Silhouette Coefficient* terendah 0,475 untuk *cluster* gempabumi periode T_{1_2019} dan tertinggi sebesar 0,570 untuk periode T_{1_2020} . Hasil *clustering* gempabumi DBSCAN periode $T_{5_2017-2021}$ mendekati peta sebaran episenter gempabumi Pulau Jawa hasil relokasi *teletomoDD* dari katalog PuSGeN 2017 yang menunjukkan wilayah Provinsi Jawa Timur memiliki potensi gempabumi tertinggi dibandingkan dengan wilayah Provinsi Jawa Tengah dan Yogyakarta.

ABSTRACT

Arafat, Ihsan B. F. 2023. **Earthquake Clustering in Regional VII Using DBSCAN and K-Means Method**. Thesis. Master Program in Computer Science Faculty of Science Universitas Islam Negeri Maulana Malik Ibrahim Malang. Supervisor: (I) Dr. M. Amin Hariyadi, M.T.(II) Dr. Irwan Budi Santoso, M.Kom.

Keywords: unsupervised learning, clustering, earthquake, DBSCAN, K-Means, silhouette coefficient.

Regional VII, covering the provinces of Central Java, Yogyakarta and East Java is a tectonic active area because it is located in the subduction zone of the Indo-Australia and Eurasian plates and there are several active faults on land. Therefore, it is necessary to do earthquake clustering to map earthquake-prone zones in Regional VII based on the earthquake clusters formed. This study uses a density-based Unsupervised Learning algorithm approach, namely Density Based Spatial Clustering of Applications with Noise or DBSCAN. This algorithm requires the input parameters epsilon (ϵ) and MinPts. The results formed were then compared with the results of clustering using the K-Means algorithm and evaluated using the Silhouette Coefficient and compared with the golden standard map of the Java Island earthquake epicenter distribution resulting from the tele-tomoDD relocation from the 2017 PuSGeN catalog. The results using DBSCAN obtained a number of clusters of 3 to 6 clusters with the lowest Silhouette Coefficient value 0.208 for $T_{5_2017-2021}$ period and the highest is 0.499 for T_{1_2020} period. While the K-Means algorithm produces clusters of 3 to 7 clusters with a Silhouette Coefficient value of 0.475 for earthquake clusters for T_{1_2019} period and the highest is 0.570 for T_{1_2020} period. The results of DBSCAN earthquake clustering for $T_{5_2017-2021}$ period approximate the distribution map of the earthquake epicenters of Java Island as a result of the relocation of the tomoDD from the 2017 PuSGeN catalog which shows that the East Java Province has the highest earthquake potential compared to the Central Java and Yogyakarta Provinces.

الخلاصة

عرفات، إحسان. ٢٠٢٣. تجميع الزلازل في المنطقة الإقليمية السابعة باستخدام طريقة DBSCAN و *K-Means*. أطروحة. برنامج دراسة ماجستير المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف: (أ) الدكتور محمد أمين هريادي الماجستير، (ب) الدكتور إيروان بودي سانتوسو الماجستير.

الكلمات المفتاحية: *unsupervised learning, clustering, زلازل, DBSCAN, K-Means, silhouette coefficient*.

المنطقة الإقليمية السابعة التي تتكون من جاوة الوسطى ويوجياكارتا وجاوة الشرقية هي مناطق تكتونية حية لأنها تقع في منطقة الاندساس للصفائح الهندية الأسترالية والأوراسية وهناك العديد من الصدوع الحية في الأرض. لذلك، من الضروري القيام بتجميع الزلازل لرسم خريطة لمناطق الزلازل المحتملة في المنطقة الإقليمية السابعة بناءً على مجموعات الزلازل المتكونة. يستخدم هذا البحث منهج خوارزمية التعلم غير الخاضع للإشراف (*Unsupervised Learning*) القائم على الكثافة، وهي التجميع المكاني القائم على الكثافة للتطبيقات مع الضوضاء (*Density Based Spatial Clustering of Application with Noise*) أو DBSCAN. تحتاج هذه الخوارزمية معلمة إدخال إبسيلون (ϵ) و *MinPts*. ثم تتم مقارنة النتائج التي تم تكوينها مع نتائج التجميع باستخدام خوارزمية *K-Means*. وتم تقييمها باستخدام طريقة معامل الصورة الظلية (*Silhouette Coefficient*)، ومقارنتها بالخريطة القياسية الذهبية (*golden standart*) لتوزيع بؤر الزلازل في جزيرة جاوا والتي نتجت عن نقل *teletomoDD* من كتالوج PuSGeN لعام ٢٠١٧. حصلت نتائج التجميع باستخدام DBSCAN على عدد من المجموعات من ٣ إلى ٦ مجموعات مع أدنى قيمة لمعامل صورة ظلية (*Silhouette Coefficient*) قدرها ٠,٢٠٨. مجموعات الزلازل للفترة T₅ 2017-2021 وأعلى ٠,٤٩٩ للفترة T₁ 2020. وأنتجت خوارزمية *K-Means* من ٣ إلى ٧ مجموعات بأقل قيمة لمعامل صورة ظلية تبلغ ٠,٤٧٥ لمجموعات الزلازل لفترة T₁ 2019 وأعلى ٠,٥٧٠ لفترة T₁ 2020. نتائج تجميع زلازل DBSCAN للفترة T₅ 2017-2021 تقريبا لخريطة توزيع بؤر الزلازل في جزيرة جاوا والتي هي نتيجة نقل *teletomoDD* من كتالوج PuSGeN لعام ٢٠١٧ الذي يوضح أن جاوة الشرقية أعلى إمكانية للزلازل مقارنة بجاوة الوسطى ويوجياكارتا.

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Provinsi Jawa Tengah, Yogyakarta dan Jawa Timur terletak di dalam wilayah tektonik zona subduksi lempeng Indo-Australia dan Eurasia (Supendi *et al.*, 2021). Selain dari zona subduksi, sumber gempabumi di daratan juga sering menyebabkan gempabumi dangkal di Jawa dalam beberapa puluh tahun terakhir (Pustlitbang PUPR, 2017). Kedua hal tersebut menyebabkan wilayah Regional VII yaitu, Provinsi Jawa Tengah, Yogyakarta dan Jawa Timur memiliki risiko gempabumi tinggi. Ditambah dengan faktor demografi yang menunjukkan padatnya pemukiman penduduk di Regional VII, gempa-gempa tersebut menimbulkan dampak yang cukup merusak (Pustlitbang PUPR, 2017). Sehingga kajian analisis seismik sangat penting dilakukan sebagai upaya mitigasi terhadap bencana gempabumi salah satunya dengan melakukan pengelompokan kejadian gempabumi atau *clustering*. Hal ini merujuk pada catatan BMKG hampir setiap hari terjadi gempabumi di wilayah Regional VII meskipun tidak semuanya dirasakan.

Metode *clustering* yang populer digunakan salah satunya adalah *Density Based Spatial Clustering of Application with Noise* atau DBSCAN. Metode ini pertama kali dikenalkan oleh Ester *et al.* (1996), yang menerapkan nilai minimum parameter input dan untuk mengelompokkan database spasial yang besar (Daszykowski & Walczak, 2009). Metode ini merupakan metode pengelompokan berdasarkan kepadatan dan sangat cocok diterapkan pada data yang besar (Kazemi-

Beydokhti *et al.*, 2017). Algoritma DBSCAN dapat digunakan untuk mendeteksi data *outlier* atau *noise* dan tidak perlu menentukan jumlah *cluster* terlebih dahulu. Algoritma ini membutuhkan parameter input Epsilon (ϵ) yang merupakan radius maksimum antara satu titik terhadap titik lainnya, serta MinPts yang merupakan titik minimum dalam sebuah *cluster* yang terbentuk.

Pada penelitian ini mencoba melakukan *clustering* DBSCAN menggunakan metode *k-Nearest Neighbors* (KNN) dengan perhitungan jarak menggunakan pendekatan *Euclidean Distance* dan *Manhattan Distance* dengan tujuan untuk mengetahui perbedaan hasil *clustering* dengan kedua pendekatan tersebut berdasarkan nilai *Silhouette Coefficient* dan kemiripan *clustering* berdasarkan *golden standart* yang telah ada di wilayah Regional VII. Sehingga hasil yang terbaik dari kedua pendekatan tersebut dapat dijadikan rekomendasi dalam penentuan jumlah *cluster*. KNN adalah salah satu algoritma paling populer dalam teknologi klasifikasi *data mining* (Gao & Li, 2020). KNN menggunakan prinsip jumlah tetangga terdekat, di mana jika suatu objek memiliki kesamaan dengan *k*-tetangga terdekat yang sebelumnya telah ditentukan, maka objek tersebut termasuk ke dalam kelompoknya.

Bencana gempabumi telah diterangkan di dalam Al-Qur'an seperti di dalam Surat Az-Zalzalah ayat 1-5 Allah berfirman:

إِذَا زُلْزِلَتِ الْأَرْضُ زِلْزَالَهَا وَأَخْرَجَتِ الْأَرْضُ أَثْقَالَهَا وَقَالَ الْإِنْسَانُ مَا لَهَا
يَوْمَئِذٍ نُحَدِّثُ أَخْبَارَهَا إِنَّ رَبَّكَ أَوْحَىٰ لَهَا

Artinya:

“Apabila bumi diguncangkan dengan guncangan yang dahsyat (1), dan bumi telah mengeluarkan isi perutnya (2), dan manusia bertanya, “Apa yang terjadi dengannya (bumi)?” (3). Pada hari itu (bumi) menyampaikan berita (tentang apa yang diperbuat manusia di atasnya) (4), karena sesungguhnya Tuhanmu telah memerintahkan (yang sedemikian itu) kepadanya (5)” (QS. Al-Zalzalah[99:1-5]).

Penjelasan menurut Tafsir Tahlili bahwa sebenarnya ayat 1-5 berkaitan dengan hari kiamat (Kementerian Agama RI, 2012). Namun, dalam skala lebih kecil ayat-ayat tersebut berkaitan dengan proses terjadinya gempa, tentunya jauh lebih kecil dibanding kejadian kiamat. Sebagaimana dijelaskan sebelumnya berdasarkan kajian ilmiah bahwa lempeng-lempeng bumi bergerak dan saling berinteraksi satu sama lain. Pertemuan lempeng ini yang bisa menimbulkan gempabumi. Sebagai contoh adalah selatan Pulau Jawa merupakan tempat pertemuan antara lempeng Eurasia dan Indo-Australia. Bila dua lempeng bertemu, maka terjadi tekanan yang terus-menerus, dan bila lempengan tidak tahan lagi menahan tekanan tersebut, maka beban yang telah terkumpul selama ratusan tahun itu lepas, dan dikeluarkan dalam bentuk gempabumi. Pada hari itu bumi menceritakan beritanya. Beban berat berjalan bertahun-tahun. Seterusnya seperti demikian itu, setiap selesai beban dilepaskan, kembali proses pengumpulan beban terjadi. Proses geologi ini dapat direkam baik secara alami maupun dengan menggunakan peralatan geofisika.

Secara singkat dijelaskan bahwa gempabumi adalah peristiwa berguncangnya bumi yang disebabkan pelepasan energi di dalam bumi. Hal ini sesuai dengan teori terjadinya gempabumi yang dijelaskan di dalam ilmu kegempaan atau seismologi.

Berdasarkan uraian yang telah disampaikan, gempa bumi merupakan suatu bencana alam yang dapat menimbulkan kerusakan. Sehingga upaya-upaya untuk mengurangi risiko terhadap bencana gempa bumi penting untuk dilakukan sebagai bentuk usaha manusia untuk meminimalkan dampak akibat bencana gempa bumi. *Clustering* gempa bumi merupakan salah satu wujud usaha untuk mengurangi risiko akibat bencana gempa bumi. Dengan melakukan *clustering* gempa bumi diharapkan mampu memetakan zona potensi gempa bumi di wilayah Regional VII tiap provinsi sehingga perencanaan pembangunan untuk daerah dengan potensi gempa bumi tinggi dapat diperkuat. Hal ini merupakan salah satu wujud upaya mitigasi terhadap bencana gempa bumi.

1.2 Pernyataan Masalah

Telah banyak penelitian mengenai *clustering* gempa bumi dilakukan dengan berbagai macam metode. Masing-masing metode tentunya memiliki kekurangan dan kelebihan untuk menentukan *cluster* gempa bumi. Pernyataan masalah pada penelitian ini adalah:

- a. Metode *clustering* mana di antara DBSCAN dan *K-Means* yang dapat direkomendasikan untuk menentukan *cluster* gempa bumi berdasarkan nilai *Silhouette Coefficient* dan kemiripan hasil *clustering* berdasarkan *golden standart* telah ada?
- b. Wilayah manakah di antara Provinsi Jawa Tengah, Yogyakarta dan Jawa Timur yang memiliki potensi gempa bumi tinggi berdasarkan jumlah gempa bumi dalam *cluster* yang terbentuk?

1.3 Tujuan Penelitian

Mengacu pada latar belakang dan pernyataan masalah yang telah disampaikan, maka tujuan dari penelitian ini adalah:

- a. Mengetahui metode *clustering* di antara DBSCAN dan *K-Means* yang dapat direkomendasikan untuk menentukan *cluster* gempa bumi berdasarkan nilai *Silhouette Coefficient* dan kemiripan hasil *clustering* berdasarkan *golden standart*.
- b. Mengetahui potensi gempa bumi di wilayah Provinsi Jawa Tengah, Yogyakarta dan Jawa Timur berdasarkan jumlah gempa bumi dalam *cluster* yang terbentuk.

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

- a. Dapat memberikan informasi mengenai *cluster* gempa bumi di wilayah Regional VII bagi instansi BMKG dalam penentuan zona potensi gempa bumi berdasarkan sumber gempanya.
- b. Dapat digunakan sebagai acuan untuk mitigasi bencana gempa bumi bagi pemerintah daerah dan masyarakat di wilayah Regional VII sehingga dapat meminimalkan korban jiwa.
- c. Dapat digunakan sebagai acuan untuk perencanaan pembangunan bagi instansi PUPR atau swasta di wilayah Regional VII sehingga untuk daerah dengan potensi gempa bumi yang tinggi dapat diperhitungkan konstruksinya.

1.5 Ruang Lingkup Penelitian

Berdasarkan penjelasan pada uraian yang telah disampaikan sebelumnya dan untuk memberikan batasan masalah yang jelas pada penelitian ini, maka ruang lingkup pada penelitian ini hanya meliputi:

- a. Penelitian ini menggunakan data gempa bumi wilayah Regional VII yaitu, Provinsi Jawa Tengah, Yogyakarta dan Jawa Timur yang bersumber dari BMKG dengan batasan koordinat 108°BT - 114°BT dan 6°LS - 14°LS dari 1 Januari 2017 hingga 31 Desember 2021.
- b. Penelitian ini menggunakan variabel bebas yang terdiri dari lintang, bujur, magnitudo, dan kedalaman sebagai input

BAB II

STUDI PUSTAKA

2.1 Clustering Gempabumi

Penelitian terkait *clustering* gempabumi telah dilakukan oleh beberapa peneliti terdahulu menggunakan berbagai metode. Beberapa penelitian tersebut di antaranya:

Karmenova *et al.* (2022), memperkenalkan sebuah pendekatan efektif untuk mendeteksi wilayah dengan peningkatan kerapatan spasial dari peristiwa seismik dan pengelompokkan data seismik dengan algoritma DBSCAN. Pemilihan algoritma DBSCAN didasarkan pada fakta bahwa algoritma DBSCAN dapat mendeteksi *cluster* yang memiliki bentuk kompleks termasuk koordinat geografis. Pada penelitian ini mereka menggunakan data seismik dari katalog seismik Republik Kazakhstan dari tahun 2011 hingga 2021. Hasilnya, dengan menggunakan pendekatan ini dapat mendeteksi *cluster* berdasarkan model periode spasial dari distribusi gempabumi dan juga dapat mendeteksi wilayah dengan peningkatan kepadatan spasial pada peta. Secara umum, hasil penelitian dibandingkan dan dicocokkan dengan peta zona seismik Republik Kazakhstan menunjukkan hasil yang dapat diandalkan dari pengelompokkan berdasarkan kepadatan.

Jufriansah *et al.* (2021), melakukan pengelompokkan gempabumi di Indonesia menggunakan algoritma *K-Means* dan DBSCAN dengan tujuan mengembangkan klasifikasi sebaran gempabumi di Indonesia tahun 2019 berdasarkan magnitudo, kedalaman, dan lokasi gempabumi. Kelebihan metode ini

adalah lebih efisien daripada metode lain karena metode ini mampu menghilangkan variabel redundan menggunakan korelasi dan mengabaikan variabel target. Hasil *clustering* menunjukkan algoritma *K-Means* lebih baik daripada DBSCAN dengan nilai optimal *cluster* $k=3$ diperoleh dengan menggunakan *Elbow Method* dan nilai *silhouette index* 0.837 untuk algoritma *K-Means* dan 0.730 untuk algoritma DBSCAN.

Amalia *et al.* (2021), melakukan klasifikasi data gempabumi di wilayah Indonesia periode tahun 2018-2020 yang diperoleh dari BMKG menggunakan algoritma DBSCAN. Tujuan penelitian ini adalah untuk mengelompokkan data gempabumi berdasarkan koordinat yang terdiri dari variabel data frekuensi kejadian, kedalaman, dan kekuatan gempa. Tahapan *clustering* pertama kali dilakukan dengan menentukan variabel data yang digunakan yaitu lintang dan bujur, kedalaman, kekuatan, dan frekuensi kejadian gempabumi dari 30.302 data. Kemudian data dikelompokkan berdasarkan wilayah kejadian gempabumi sehingga diperoleh 47 wilayah gempabumi. Selanjutnya dilakukan penentuan parameter epsilon dan MinPts optimal menggunakan grafik *k-dist*. Terakhir dilakukan validasi jumlah *cluster* menggunakan *Silhouette Coefficient*. Nilai *Silhouette Coefficient* terbaik diperoleh dengan melakukan *trial and error* nilai epsilon optimal dan MinPts. Sehingga diperoleh nilai *Silhouette Coefficient* terbaik 0.35 untuk *clustering* berdasarkan koordinat dan 0.39 untuk *clustering* berdasarkan karakteristik gempa. Kelebihan metode ini adalah mampu mendeteksi data yang termasuk *outlier* atau *noise*.

Cesca (2020), melakukan pengelompokan gempabumi menggunakan perangkat lunak *open-source Python* berbasis kepadatan bernama *seiscloud*, menggunakan *library pyrocko* untuk seismologi. Tujuannya adalah untuk mengklasifikasikan gempabumi secara otomatis berdasarkan kesamaan atribut, untuk memperoleh informasi tentang proses kegempaan dan pola patahan dari kumpulan data seismik yang besar. *Seiscloud* bekerja berdasarkan algoritma DBSCAN dan mengandalkan matriks jarak untuk mengidentifikasi kumpulan kejadian yang membagikan fitur tidak langsung seperti *waveform* yang serupa. Kelebihan metode ini adalah mampu menjelajah dan memvisualisasikan katalog gempabumi yang besar dan dapat diterapkan untuk berbagai aplikasi, seperti sebagai identifikasi cepat anomali kegempaan, dengan hiposenter lokal, atau gempa *swarm*, untuk memetakan geometri patahan aktif dan merekonstruksi kompleks patahan dengan memetakan daerah dengan lokasi dan *focal mechanism* yang serupa, dan untuk melacak pola makroseismisitas yang diinduksi oleh operasi antropogenik seperti injeksi cairan, operasi penambangan atau reservoir air.

Avip *et al.* (2020), melakukan analisis *cluster* gempabumi di Indonesia menggunakan metode *K-medoids* kemudian divisualisasikan dalam peta menggunakan ArcGIS. Penelitian ini menghasilkan 9 *cluster* gempabumi berdasarkan area kejadian gempabumi dengan karakteristik kedalaman gempabumi yang berbeda pada masing-masing *cluster*. Berdasarkan area kejadian gempabumi selama 1 Januari hingga 19 Oktober 2020 gempabumi terjadi hampir di seluruh wilayah di Indonesia kecuali Kalimantan. Berdasarkan karakteristik kedalaman menunjukkan dari *cluster* 1 hingga 7 termasuk ke dalam *shallow earthquake*,

cluster 8 termasuk ke dalam *medium earthquake* dan *cluster 9* termasuk ke dalam *deep earthquake*. Kelebihan dari metode ini adalah mampu mengelompokkan data gempa bumi berdasarkan karakteristiknya dalam penelitian ini adalah kedalaman dan area kejadian gempa bumi. Hal ini dapat digunakan untuk mengetahui kerawanan gempa bumi suatu wilayah.

Fan and Xu (2019), menerapkan algoritma clustering *K-means* dan DBSCAN untuk melakukan analisis data seismik. Kinerja kesesuaian pola gempa bumi dibandingkan dengan data gempa bumi. Tahapannya adalah pertama, memetakan posisi parameter data gempa ke dalam koordinat dua dimensi dan kemudian mengelompokkannya menggunakan algoritma DBSCAN. Selanjutnya mengombinasikan antara magnitudo dan kedalaman gempa bumi, menentukan nilai *k* terbaik menggunakan *Elbow Method*, kemudian mengelompokkan data dengan algoritma *K-means*. Hasilnya menunjukkan bahwa algoritma DBSCAN lebih baik daripada *K-means* dalam kesesuaian pola gempa bumi.

Harini *et al.* (2020), memperkenalkan pemetaan zona potensi gempa bumi di wilayah Pusat Gempabumi Regional 3 (Bali dan Nusa Tenggara) menggunakan metode *clustering* DBSCAN. Metode tersebut cocok digunakan untuk mengelompokkan data spasial dan tidak cembung seperti data gempa bumi. Mereka melakukan modifikasi parameter epsilon menggunakan magnitudo karena dianggap dapat mewakili dampak dari kejadian gempa bumi sehingga nilai epsilon berbeda untuk setiap titik. Sebelum melakukan *clustering*, magnitudo dikonversi ke dalam kilometer (km) untuk menghitung wilayah terdampak gempa bumi, yang disebut dengan epsilon. Dengan menggunakan persamaan Radius Persepsi

Gempabumi *McCue* wilayah terdampak gempabumi diperoleh dengan menghitung radius gempabumi. Pendekatan ini memberikan hasil yang baik untuk pemetaan zona potensi gempabumi diperkuat dengan indeks validasi *cluster* yang memberikan sebuah rata-rata lebih tinggi daripada 0 menggunakan *Silhouette index*.

Karri *et al.* (2019), melakukan pengelompokan data gempabumi menggunakan aplikasi pengelompokan data spasial dan *noise* berbasis kepadatan (DBSCAN) pada data gempabumi di India untuk identifikasi zona rawan gempa berdasarkan aktivitas seismik di wilayah tersebut. Hasilnya menunjukkan bahwa DBSCAN dapat digunakan untuk analisis zona kegempaan dan mengidentifikasi batas *cluster* secara efektif dengan kualitas *cluster* yang diperoleh mewakili zona berbagai zona seismik sesuai dengan peta seismik *India Bureau of India Standards IS 1893:2022*.

Kazemi-Beydokhti *et al.* (2017), melakukan pengelompokan gempabumi di Iran menggunakan algoritma DBSCAN dengan tujuan untuk mengembangkan model kegempaan secara *spatio-temporal* di Iran berdasarkan kemiripan perilaku seismik dalam bentuk *cluster* terpisah. Penelitian dilakukan dengan membagi data menjadi tiga jendela periode yang terdiri dari rentang 2-, 5- dan 10- tahun. Sehingga diperoleh pola kegempaan yang berbeda untuk masing-masing periode. Hasilnya menunjukkan bahwa batas-batas beberapa daerah sesimik yang diperoleh hampir sama di berbagai jendela periode, seperti tenggara dan barat laut Zagros, lokasi patahan Mosha di utara Alboz, dan lokasi sesar Tabriz di Azerbaijan. Penelitian ini menunjukkan hasil mirip dengan penelitian sebelumnya yang dilakukan oleh

Mojarab *et al.* (2014) dan Ansari *et al.* (2009) pada daerah yang disajikan dalam model. Kelebihan dari metode ini adalah mampu mengelompokkan gempa bumi berdasarkan tingkat kerapatan kejadian gempa bumi meskipun bentuknya tidak beraturan.

Novianti *et al.* (2017), melakukan pengelompokan episenter gempa bumi di wilayah Bengkulu menggunakan *K-Means clustering* dengan tujuan untuk mengetahui karakteristik gempa bumi di wilayah tersebut. Variabel yang digunakan dalam perhitungan adalah latitude, longitude, dan magnitude. Tahapan pada *clustering* ini adalah dengan menentukan pusat *cluster* atau *centroid* terlebih dahulu dengan menghitung *mean* dari data. Kemudian menentukan anggota *cluster* dengan menghitung jarak antara objek terhadap pusat *cluster* dengan persamaan *Euclidean distance*. Kemudian menghitung ulang *centroid* untuk mendapatkan objek baru dan langkah ini dilakukan berulang kali hingga nilai *centroid* tidak berubah. Pada penelitian ini untuk menentukan nilai optimum *cluster* menggunakan metode *Krzanowski and Lai criteria* atau *KL index*. Kelebihan metode ini adalah dapat mengelompokkan kejadian gempa bumi berdasarkan lokasi episenternya, yaitu diperoleh 5 *cluster* gempa bumi dengan lokasi episenter di laut dan daratan Bengkulu, 1 *cluster* gempa bumi dengan lokasi episenter di sekitar Kepulauan Mentawai Provinsi Sumatera Barat dan 1 *cluster* dengan lokasi episenter di laut Provinsi Lampung. Hasil clustering pada penelitian ini adalah fitur geometri menggunakan *Euclidean distance* sehingga karakteristik *clustering* yang diperoleh terbatas pada sifat geometris. Oleh karena itu, perlu dilakukan kajian ulang secara seismologi dan karakteristik geologi untuk menjadikan hasil analisis pada penelitian

tersebut ke dalam zona gempa bumi di Provinsi Bengkulu atau juga bisa dibandingkan dengan model analisis bahaya seismik.

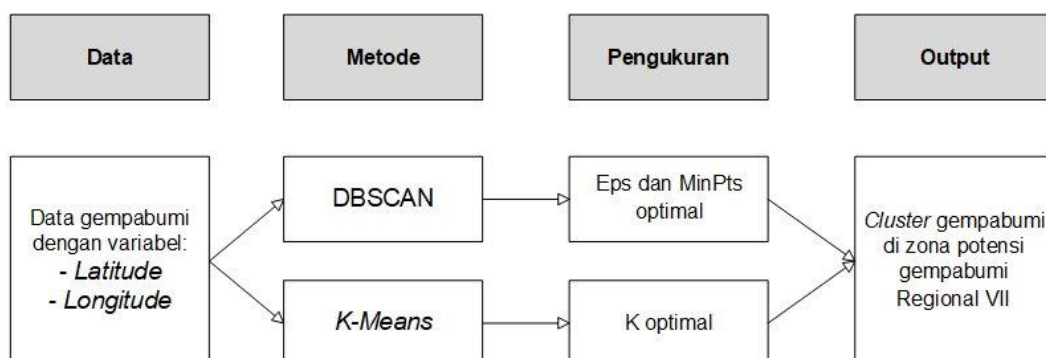
Georgoulas *et al.* (2013), mengusulkan pendekatan baru untuk *clustering* gempa bumi secara *spatio-temporal*. Pendekatan tersebut dibangun berdasarkan implementasi yang dimodifikasi dari algoritma DBSCAN yang memperkenalkan konsep akumulasi massa seismik untuk isolasi *cluster* peristiwa seismik dalam ruang dan waktu yang disebut SM-DBSCAN dan kemudian menggunakan *single linkage agglomerative hierarchical clustering* atau pengelompokan agglomeratif hierarki hubungan tunggal untuk pengelompokan spasial tingkat kedua. Hasil yang diperoleh menunjukkan bahwa metode ini mampu menemukan *cluster* dengan bentuk yang tidak teratur yang merupakan fitur penting karena zona patahan seismik jarang berbentuk teratur. Selain itu algoritma ini mampu mengenali beberapa wilayah seismik antara lain bahwa wilayah Cretan selatan dan Ionia sebagai wilayah seismik sesuai dengan hasil empiris pada zona seismik yang khusus.

Mojarab *et al.* (2014), melakukan klasifikasi wilayah seismik di Iran ke dalam zona seismik dengan karakteristik seismik yang serupa menggunakan algoritma pintar yang dinamakan *Self Organizing Map* (SOM). Algoritma ini memerlukan parameter gempa bumi seperti koordinat, distribusi seismik fault, topografi, gravitasi, dan magnet bumi sebagai input layer. Berdasarkan penelitian sebelumnya *cluster* yang dipilih antara 5-25. Pada penelitian ini untuk validasi *cluster* menggunakan metode validasi pengelompokan dan penilaian ahli dengan membandingkan hasil clustering dengan peta yang telah ada sebelumnya.

Ansari *et al.* (2009), melakukan identifikasi pola tersembunyi katalog historis gempabumi dan instrumental di Iran menggunakan *unsupervised fuzzy clustering*. Metode *clustering* yang digunakan pada penelitian ini berdasarkan modifikasi *fuzzy estimasi maximum likelihood*. Selain itu, *fuzzy hyper-volume* dan indeks kerapatan partisi digunakan sebagai indeks kinerja untuk memilih *cluster* terbaik. Hasilnya menunjukkan bahwa hasil analisis *clustering* dan model seismotectonik mengungkapkan bahwa beberapa informasi tersembunyi dapat diekstraksi dengan cara analisis *fuzzy clustering*. Kelebihan dari metode ini adalah mampu mendeteksi *cluster ellips* dengan ukuran yang berbeda.

2.2 Kerangka Teori

Pada tahapan ini membahas tentang metode yang telah dibahas pada subBab sebelumnya dan membandingkannya berdasarkan studi literatur pada penelitian ini disertai dengan penjelasan alur desain penelitian yang terdiri dari input, proses, dan output seperti dijelaskan Gambar 2.1.



Gambar 2.1 Kerangka Teori Penelitian

Pada penelitian ini penulis menggunakan data gempabumi dengan variabel *latitude* dan *laongitude* sebagai parameter input. Kemudian diolah menggunakan metode DBSCAN dan *K-Means* untuk melakukan *clustering* gempabumi. Pada metode DBSCAN dilakukan pengukuran nilai epsilon (ϵ) dan MinPts optimal untuk mendapatkan *cluster* terbaik. sedangkan pada metode *K-Means* dilakukan pengukuran *k* optimal untuk mendapatkan *cluster* terbaik. Sehingga hasil yang diperoleh adalah jumlah *cluster* gempabumi di wilayah Regional VII yang menunjukkan zona potensi gempabumi.

Penulis menggunakan beberapa artikel dari jurnal-jurnal ilmiah penelitian yang dilakukan oleh peneliti-peneliti sebelumnya mengenai *clustering* gempabumi dengan menggunakan berbagai metode. Adapun penelitian tersebut dirangkum dalam Tabel 2.1.

Tabel 2.1 Daftar Literatur

No.	Judul Artikel	Penulis	Metode	Variabel Input	Hasil
1	An Approach for Clustering of Seismic Events using Unsupervised Machine Learning	Karmenova <i>et al.</i> (2022)	DBSCAN	<ul style="list-style-type: none"> ▪Latitude ▪Longitude 	<i>Cluster</i> yang terbentuk pada area dengan kepadatan spasial tinggi yang merupakan daerah <i>seismic active</i> di Kazakhstan memiliki kemiripan dengan peta zonasi seismik Kazakhstan.
2	Analysis of Earthquake Activity in Indonesia by Clustering Method	Jufriansah <i>et al.</i> (2021)	<i>K-Means</i> and DBSCAN	<ul style="list-style-type: none"> ▪Latitude ▪Longitude ▪Magnitude ▪Depth 	<i>K-Means clustering</i> lebih baik daripada DBSCAN dengan <i>k</i> optimal = 3 nilai <i>silhouette</i> 0,867 untuk <i>K-Means</i> dan 0,730 untuk DBSCAN dan distribusi <i>cluster</i> gempabumi mendekati distribusi katalog gempabumi Indonesia 1900-2016 (Permana, 2021).

No.	Judul Artikel	Penulis	Metode	Variabel Input	Hasil
3	Clustering of Seismicity in the Indonesian Region for the 2018-2020 Period using the DBSCAN Algorithm	Amalia <i>et al.</i> (2021)	DBSCAN	<ul style="list-style-type: none"> ▪Latitude ▪Longitude ▪Magnitude ▪Depth ▪Frequency 	<i>Clustering</i> dengan variabel koordinat dan karakteristik gempabumi menghasilkan <i>silhouette</i> terbaik sebesar 0,35 dan 0,39. Distribusi <i>cluster</i> menunjukkan Indonesia bagian barat memiliki frekuensi kegempaan tinggi dibandingkan bagian tengah dan timur.
4	Seiscloud, a tool for density-based seismicity clustering and visualization	Cesca (2020)	DBSCAN	<ul style="list-style-type: none"> ▪Latitude ▪Longitude ▪Magnitude ▪Depth ▪Focal Mechanism 	<i>Seiscloud</i> dapat menghasilkan <i>cluster</i> gempabumi berdasarkan atribut pada katalog seperti kedalaman, magnitudo, <i>focal mechanism</i> dan momen tensor.
5	Analysis of Characteristic of Earthquake Area in Indonesia in 2020 with Cluster Analysis as Natural Disaster	Avip <i>et al.</i> (2020)	<i>K-Medoids</i>	<ul style="list-style-type: none"> ▪Latitude ▪Longitude ▪Magnitude ▪Depth 	Menghasilkan <i>cluster</i> gempabumi berdasarkan karakteristik kedalaman gempabumi dan area sesuai dengan potensi gempabumi.
6	Application and visualization of typical clustering algorithms in seismic data analysis	Fan and Xu (2019)	<i>K-Means</i> and DBSCAN	<ul style="list-style-type: none"> ▪Latitude ▪Longitude ▪Magnitude 	Algoritma DBSCAN menghasilkan <i>cluster</i> gempabumi lebih baik dibandingkan <i>K-Means</i> karena distribusi gempabumi dapat menyesuaikan busur seismik Pasific Barat dan Pasific Timur sedangkan <i>K-Means</i> kurang sesuai.
7	The earthquake events and impacts mapping in Bali and Nusa Tenggara using a	Harini <i>et al.</i> (2020)	DBSCAN	<ul style="list-style-type: none"> ▪Latitude ▪Longitude ▪Magnitude 	Diperoleh pemetaan daerah potensi gempabumi di wilayah Regional 3 per bulan dengan hasil baik berdasarkan hasil validasi <i>cluster</i>

No.	Judul Artikel	Penulis	Metode	Variabel Input	Hasil
	clustering method				<i>silhouette index</i> tidak kurang dari 0 dengan rata-rata waktu <i>clustering</i> 1,25 s.
8	Identification of Seismic Zones of India using DBSCAN	Karri <i>et al.</i> (2019)	DBSCAN	<ul style="list-style-type: none"> ▪Latitude ▪Longitude ▪Magnitude ▪Depth 	Diperoleh kemiripan antara <i>clustering</i> DBSCAN dengan peta seismistas standar India, <i>cluster</i> dengan densitas tinggi di area Andaman dan Nicobar.
9	Spatio-Temporal Modeling of Seismic Provinces of Iran Using DBSCAN Algorithm	Kazemi-Beydokhti <i>et al.</i> (2017)	DBSCAN	<ul style="list-style-type: none"> ▪Latitude ▪Longitude 	Hasil <i>clustering</i> menunjukkan kemiripan dengan batas-batas beberapa daerah seismik di Iran pada berbagai jendela periode data, seperti tenggara dan barat laut Zagros, lokasi patahan Moshā di utara Alborz dan lokasi sesat Tabriz di Azerbaijan.
10	K-Means cluster analysis in earthquake epicenter clustering	Novianti <i>et al.</i> (2017)	<i>K-Means</i>	<ul style="list-style-type: none"> ▪Latitude ▪Longitude ▪Magnitude 	<i>Cluster</i> yang terbentuk terbatas sifat geometris sehingga terdapat <i>cluster</i> di luar wilayah yang menjadi fokus penelitian.
11	“Seismic-mass” density-based algorithm for spatio-temporal clustering	Georgoulas <i>et al.</i> (2013)	SM-DBSCAN dan <i>Agglomerative Hierarchical Clustering</i>	<ul style="list-style-type: none"> ▪Latitude ▪Longitude ▪Magnitude 	Hasil yang diperoleh menunjukkan metode ini mampu menemukan <i>cluster</i> dengan bentuk yang tidak teratur sesuai zona patahan seismik dan mampu mengenali beberapa wilayah seismik seperti wilayah Cretan selatan dan Ionia sebagai wilayah seismik sesuai dengan hasil empiris pada zona seismik yang khusus.
12	Modeling of the seismotectonic	Mojarab <i>et al.</i> (2014)	<i>Self Organizing Map</i> (SOM)	<ul style="list-style-type: none"> ▪Latitude ▪Longitude ▪Fault 	<i>Cluster</i> yang dihasilkan sebagian besar memiliki

No.	Judul Artikel	Penulis	Metode	Variabel Input	Hasil
	provinces of Iran using the self-organizing map algorithm			<ul style="list-style-type: none"> ▪ Earthquake events ▪ Topography ▪ Gravity ▪ Magnetic 	kemiripan dengan wilayah seismotektonik Iran.
13	Clustering analysis of the seismic catalog of Iran	Ansari <i>et al.</i> (2009)	<i>Fuzzy Clustering</i>	<ul style="list-style-type: none"> ▪ Latitude ▪ Longitude ▪ Magnitude 	Hasil <i>clustering</i> memiliki kemiripan dengan kondisi seismotektonik Iran berupa partisi <i>cluster</i> .

Berdasarkan Tabel 2.1 diperoleh beberapa metode *clustering* yang dapat digunakan untuk melakukan pengelompokkan data gempa bumi yang bersumber dari berbagai artikel ilmiah pada beberapa penelitian terdahulu. Pada beberapa penelitian yang dilakukan oleh peneliti-peneliti terdahulu terdapat perbedaan mengenai data yang digunakan sebagai input dan metode yang digunakan dalam tahap *processing* data meskipun *output* yang diperoleh adalah hampir sama yaitu berupa *cluster* gempa bumi di wilayah seismotektonik atau wilayah dengan aktivitas seismik aktif. Adapun langkah-langkah dari input, proses, hingga memperoleh output pada penelitian sebelumnya diuraikan sebagai berikut:

Berdasarkan Tabel 2.1 dan uraian di atas dalam penentuan *cluster* gempa bumi secara umum dapat diperoleh kesimpulan sebagai berikut:

- a. Dalam penentuan cluster gempa bumi dengan beberapa metode berdasarkan penelitian terkait sebelumnya, variabel input yang digunakan untuk proses clustering terdiri dari berbagai macam, di antaranya: *latitude*, *longitude*, *magnitude*, *depth*, *earthquake frequency*, *focal mechanism*, *earthquake event*, *topography*, *magnetic*, dan *gravity*.

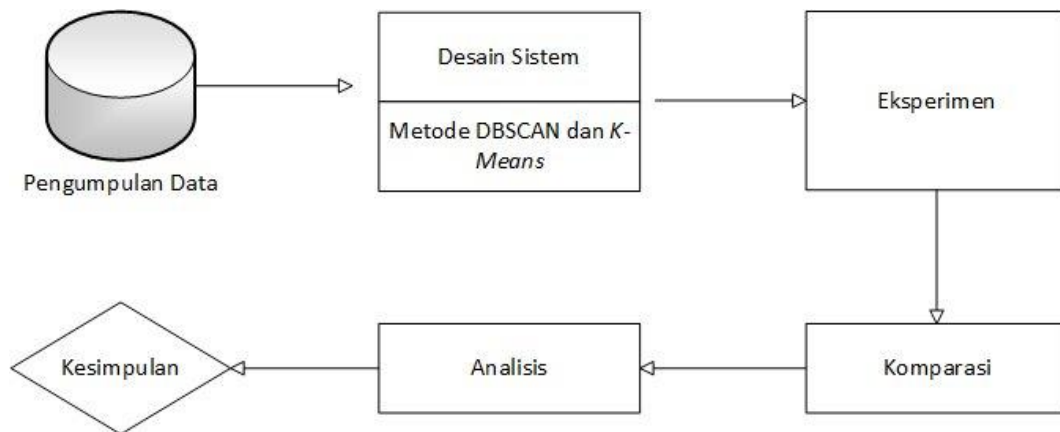
- b. Dari berbagai macam variabel input tersebut, output yang dihasilkan pada umumnya adalah *cluster* gempa bumi di berbagai wilayah penelitian yang menunjukkan kesesuaian dengan zona potensi gempa bumi atau zona seismik yang telah ada yang menjadi standar.

BAB III

METODOLOGI PENELITIAN

3.1 Desain Penelitian

Pada bagian ini menampilkan desain yang digunakan pada penelitian yang akan dilakukan. Adapun desainnya bisa dilihat pada Gambar 3.1.



Gambar 3.1 Desain Penelitian

Gambar 3.1 menjelaskan tentang alur dan proses penelitian. Dimulai dengan proses pengumpulan data, desain sistem yang terdiri desain algoritma DBSCAN dan *K-Means*, eksperimen, komparasi algoritma DBSCAN dan *K-Means*, analisis hasil dan terakhir kesimpulan.

3.1.1 Pengumpulan Data

Pengumpulan data penting dilakukan dalam sebuah penelitian. Kualitas data yang baik menentukan hasil dari sebuah penelitian. Penggunaan data dalam penelitian ini melalui proses literasi berdasarkan jurnal-jurnal penelitian ilmiah

yang memiliki tema yang sama. Penelitian ini menggunakan data gempabumi wilayah Regional VII meliputi Provinsi Jawa Tengah, Yogyakarta, dan Jawa Timur tahun 2017-2021 dengan batasan koordinat 6°S - 14°S dan 108°E - 114°E yang diperoleh dari website BMKG repogempa.bmkg.go.id/repo_new/ yang diakses pada tanggal 18 Mei 2022 sebanyak 1.827 data parameter kejadian gempabumi. Data yang dipakai dalam penelitian ini adalah data sekunder karena diperoleh dengan tidak langsung melalui pengambilan data di lapangan atau dari sebuah alat melainkan sudah melalui tahap pengolahan. Pengumpulan data dalam penelitian ini adalah sebagai berikut:

Data sekunder yang digunakan dalam penelitian ini diperoleh dengan cara memasukkan nilai parameter data dengan rincian sebagai berikut:

- a. Menentukan tanggal awal dan akhir,
- b. Menentukan batas wilayah geografis objek penelitian yaitu, dengan memasukkan batas lintang dan bujur,
- c. Menentukan rentang nilai magnitudo, dan
- d. Menentukan rentang nilai kedalaman.

Sehingga data yang diperoleh seperti yang ditunjukkan pada Tabel 3.1.

Tabel 3.1 Data Parameter Gempabumi

Date	Time	Lat	Lon	Dep	M	MT	Region
01/01/2019	25:34.4	10.48 S	108.50 E	10	4.7	Yes	South of Java, Indonesia (Status: confirmed)
01/01/2019	51:07.6	8.25 S	108.48 E	32	2.9	-	Java, Indonesia (Status: confirmed)
01/01/2019	22:48.1	10.52 S	108.48 E	32	4.8	-	South of Java, Indonesia (Status: confirmed)
02/01/2019	51:26.6	8.82 S	112.38 E	92	3.5	-	Java, Indonesia (Status: confirmed)
02/01/2019	52:05.9	10.38 S	108.49 E	10	4.3	Yes	South of Java, Indonesia (Status: confirmed)

Date	Time	Lat	Lon	Dep	M	MT	Region
03/01/2019	42:29.0	9.29 S	109.64 E	10	4.4	Yes	South of Java, Indonesia (Status: confirmed)
03/01/2019	44:22.0	9.22 S	113.15 E	10	3.1	-	South of Java, Indonesia (Status: confirmed)

Sumber: repogempa.bmkg.go.id/repo_new/ (2022)

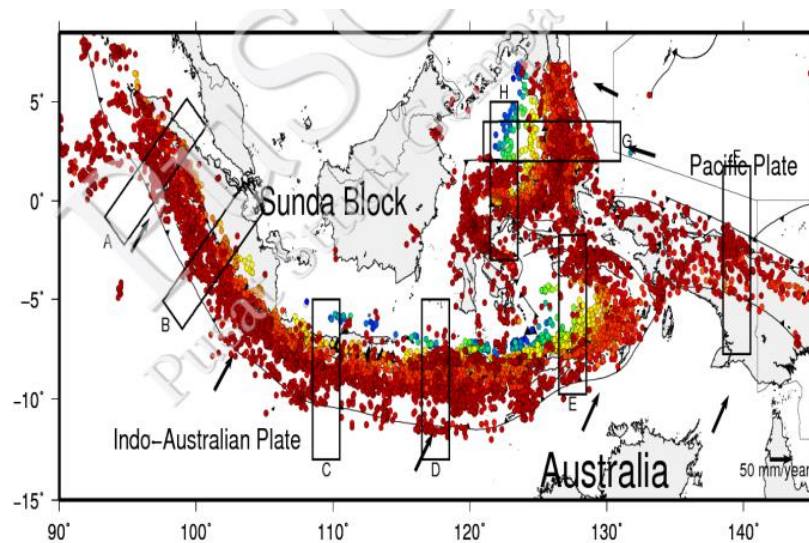
Berdasarkan Tabel 3.1 data yang diperoleh terdiri dari variabel *Date*, *Time*, *Latitude*, *Longitude*, *Depth*, *Magnitude*, *Momen Tensor*, dan *Region*.

Berikutnya, selain data gempabumi yang merupakan data utama yang digunakan untuk proses *clustering* gempabumi, penulis juga menggunakan data berupa peta seismisitas hasil penelitian beberapa ahli kegempaan dari berbagai institusi pemerintah yaitu, Kementerian Pekerjaan Umum dan Perumahan Rakyat, Kementerian Energi dan Sumber Daya Mineral, Kementerian Ristek dan Pendidikan Tinggi, Kementerian Perhubungan, Badan Meteorologi Klimatologi dan Geofisika, Badan Nasional Penanggulangan Bencana, Lembaga Ilmu Pengetahuan Indonesia, Badan Informasi Geospasial, Institut Teknologi Bandung dan unsur asosiasi/praktisi, serta Geoscience Australia yang tergabung dalam “Tim Pemutakhiran Peta Bahaya Gempa Indonesia Tahun 2007 dan Penyiapan Pusat Studi Gempa Nasional (PuSGeN)” yang hasilnya tertuang dalam buku “Peta Sumber dan Bahaya Gempa Indonesia Tahun 2017”.

Penelitian yang dilakukan oleh (Pustlitbang PUPR, 2017) dalam mengidentifikasi lokasi kelompok gempabumi menggunakan metode *Teleseismic Tomography Double Difference* atau *teletomoDD*. Metode ini memanfaatkan waktu kedatangan absolut dan relatif dari gelombang seismik (Zhang & Thurber, 2006) untuk mendapatkan lokasi kejadian gempabumi yang lebih akurat dan struktur

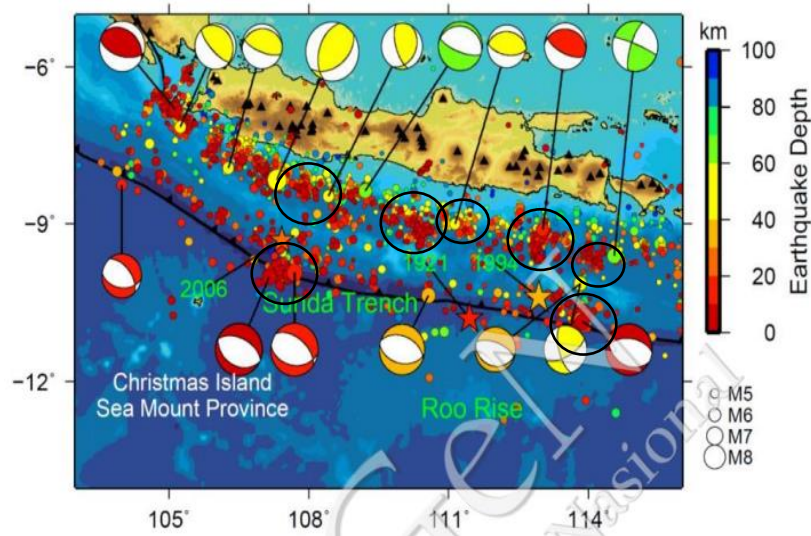
kecepatan dekat dengan sumber gempa bumi daripada tomografi standar, yang hanya menggunakan waktu kedatangan absolut.

Adapun hasil relokasi gempa bumi menggunakan *tomoDD* untuk data gempa bumi di wilayah Indonesia periode April 2009-Juli 2016 sebanyak ~40.000 *event* yang didapatkan dari katalog BMKG adalah sebagai berikut.



Gambar 3.2 Peta Seismisitas Indonesia hasil relokasi *teletomoDD*
Sumber: (Pustlitbang PUPR, 2017)

Setelah diperbesar Gambar 3.2 di atas diperoleh sebaran episenter gempa bumi untuk wilayah Jawa sebagai berikut.

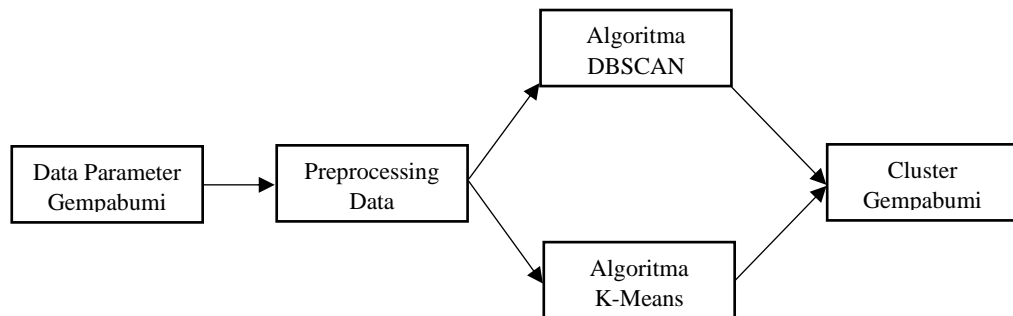


Gambar 3.3 Sebaran episenter gempa bumi hasil relokasi *teletomoDD* di Pulau Jawa dan sekitarnya untuk *event* gempa dengan kedalaman ≤ 100 km dan magnitudo ≥ 4
 Sumber: (Pustlitbang PUPR, 2017)

Berdasarkan Gambar 3.3 hasil relokasi gempa bumi di wilayah Pulau Jawa menggunakan *teletomoDD* menunjukkan distribusi episenter gempa bumi terbentuk dalam *cluster*. Provinsi Jawa Tengah, Yogyakarta dan Jawa Timur termasuk ke dalam area Regional VII terdapat 7 *cluster* gempa bumi yang merupakan zona potensi gempa bumi yang berada di sekitar zona pertemuan lempeng atau zona *megathrust* yang menjadi sumber utama gempa bumi di selatan Pulau Jawa. Kemudian, hasil ini yang digunakan oleh penulis sebagai *golden standart* pada penelitian ini.

3.1.2 Desain Sistem

Tahapan desain sistem menjelaskan mengenai alur pengolahan data input, proses, hingga menghasilkan output ditampilkan pada Gambar 3.4.



Gambar 3.4 Desain Sistem Penelitian

3.1.2.1 Preprocessing

Tahapan berikutnya adalah *preprocessing* yaitu, menentukan variabel data yang akan digunakan. Pada penelitian ini variabel data yang digunakan sebagai input dalam penentuan *cluster* adalah *latitude* dan *longitude*. Sehingga data pada penelitian ini sebanyak 1.827 data.

3.1.2.2 Algoritma DBSCAN

Penelitian ini menggunakan algoritma DBSCAN untuk proses clustering gempabumi. Algoritma DBSCAN merupakan salah satu pendekatan algoritma *machine learning* berbasis kepadatan untuk analisis *cluster* yang diusulkan pertama kali oleh Ester *et al.*, 1996 (Daszykowski & Walczak, 2009). Tujuannya adalah untuk menentukan *cluster* dengan bentuk tidak beraturan. Dalam pendekatan ini kepadatan suatu kumpulan data dari titik-titik tertentu direpresentasikan dengan jumlah titik yang ada dalam radius tertentu (ϵ) dari titik tersebut dihitung (Rahman & Wijayanto, 2021). Algoritma ini memandang *cluster* sebagai area dengan kepadatan tinggi yang dipisahkan oleh area dengan kepadatan rendah. *Cluster* digambarkan berupa daerah dengan kepadatan titik yang tinggi sedangkan *noise*

ditunjukkan berupa daerah dengan kepadatan yang rendah (Omran *et al.*, 2007). Algoritma ini membutuhkan dua parameter input yaitu, epsilon (ϵ) yang merupakan radius maksimum antara satu titik ke titik lain dan jumlah minimum objek (MinPts) yang merupakan titik minimum dalam sebuah *cluster* yang terbentuk. Proses clustering dengan algoritma DBSCAN dapat dilihat pada Gambar 3.5.

```

Input: DB: Database
Input:  $\epsilon$ : Radius
Input: minPts: Density threshold
Input: dist: Distance function
Data: label: Point labels, initially unvisited
1 foreach point  $p$  in database DB do // Iterate over every point
2     if label( $p$ )  $\neq$  unvisited then continue // Skip processed points
3     Neighbors  $N \leftarrow$  RangeQuery(DB, dist, $p,\epsilon$ ) // Find initial neighbors
4     if  $|N| < minPts$  then // Non-core points are noise
5         label( $p$ )  $\leftarrow$  Noise
6         continue
7      $c \leftarrow$  next cluster label // Start a new cluster
8     label( $p$ )  $\leftarrow$   $c$ 
9     Seed set  $S \leftarrow N \setminus \{p\}$  // Expand neighborhood
10    foreach  $q$  in  $S$  do
11        if label( $q$ ) = Noise then label( $q$ )  $\leftarrow$   $c$ 
12        if label( $q$ )  $\neq$  unvisited then continue
13        Neighbors  $N \leftarrow$  RangeQuery(DB, dist, $q,\epsilon$ )
14        label( $q$ )  $\leftarrow$   $c$ 
15        if  $|N| < minPts$  then continue // Core-point check
16         $S \leftarrow S \cup N$ 

```

Gambar 3.5 Pseudocode Algoritma DBSCAN (Schubert et al., 2017)

Berdasarkan Gambar 3.5 menjelaskan proses *clustering* gempabumi mulai dari input data hingga proses *clustering*. Adapun proses *clustering* algoritma DBSCAN dijelaskan sebagai berikut:

1. Menentukan nilai epsilon (ϵ) dan MinPts. Nilai epsilon yang optimal sangat penting karena mempengaruhi hasil *cluster* yang baik. Penentuan nilai epsilon optimal menggunakan metode jarak KNN (Karmenova et al., 2022). Hal utama dalam perhitungan nilai epsilon optimal dengan metode KNN adalah dengan menghitung jarak dalam matriks titik ke KNN. Jarak rata-rata dari setiap titik

KNN dihitung. Nilai k diatur sesuai dengan nilai MinPts dan jarak k ditampilkan dalam kurva naik di mana nilai yang sesuai dengan *threshold* di mana ada perubahan tajam dalam kurva merupakan jarak k yang optimal. Terdapat beberapa persamaan perhitungan jarak pada KNN salah di antaranya adalah *Euclidean Distance* dan *Manhattan Distance*. Perhitungan jarak dengan *Euclidean Distance* dijelaskan pada persamaan 3.1.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.1)$$

di mana:

x_i : titik x ke i
 y_i : titik y ke i

Sedangkan perhitungan jarak menggunakan *Manhattan Distance* dijelaskan pada persamaan 3.2.

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (3.2)$$

di mana:

x_i : titik x ke i
 y_i : titik y ke i

2. Melakukan *clustering* dengan menggunakan epsilon (ϵ) dan MinPts sebagai parameter input.
3. Melakukan uji *cluster* dengan *Silhouette Coefficient* dengan persamaan sebagai berikut:

$$si = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3.3)$$

di mana:

s_i : nilai *Silhouette Coefficient*

$a(i)$: rata-rata jarak antara titik I dengan seluruh titik dalam A
(*cluster* di mana titik I berada).

$b(i)$: rata-rata jarak antara titik I ke seluruh titik dalam *cluster* selain A.

4. Melakukan visualisasi *cluster* yang diperoleh.

3.1.2.3 Algoritma *K-Means*

K-Means merupakan algoritma pengelompokan data berdasarkan karakteristik yang banyak digunakan untuk mengelompokkan data dalam jumlah besar yang diusulkan pertama kali oleh *MacQueen* pada tahun 1967 (Shi *et al.*, 2010). *K-Means* juga salah satu algoritma *clustering* non hirarki yang bekerja berdasarkan pemisahan data dan kriteria kesalahan kuadrat atau *squared error* yang berbasis titik pusat atau *centroid* (Wagstaff, 2012). Algoritma *K-Means* mempunyai kemampuan dalam mengelompokkan data dalam jumlah cukup besar dengan waktu komputasi yang cepat dan efisien (Reviantika *et al.*, 2020). Proses *clustering* dengan algoritma *K-Means* dapat dilihat pada Gambar 3.6.

```
K-Means clustering algorithm

Require:  $D = \{d_1, d_2, d_3, \dots, d_i, \dots, d_n\}$  // Set of n data points
// Number of desired clusters

Ensure: A set of  $k$  clusters

Steps:
1. Arbitrarily choose  $k$  data points from  $D$  as initial centroids;
2. Repeat
   Assign each point  $d_i$  to the cluster which has the closest centroid;
   Calculate the new mean for each cluster;
Until convergence criteria is met
```

Gambar 3.6 Pseudocode Algoritma *K-Means* (Yedla, Madhu, Srinivasa Rao Pathakota, 2010)

Gambar 3.6 menjelaskan proses *clustering* gempa bumi mulai dari input data hingga proses *clustering*. Proses *clustering* algoritma *K-Means* dijelaskan sebagai berikut:

1. Menentukan jumlah k *cluster*.
2. Menentukan pusat *cluster* secara acak.
3. Menghitung jarak antara setiap objek data d_i ($1 \leq i \leq n$) dengan semua k pusat *cluster* c_j ($1 \leq j \leq k$) dan menetapkan objek data d_i ke *cluster* terdekat.
4. Untuk setiap *cluster* j ($1 \leq j \leq k$), hitung ulang pusat *cluster* hingga tidak ada perubahan pada pusat *cluster*.
5. Melakukan uji *cluster* menggunakan *Elbow Method* yang merupakan selisih kuadrat dari nilai k yang berbeda dengan perhitungan sebagai berikut:

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} |x - C_i|^2 \quad (3.4)$$

di mana:

SSE : *Sum of Squares Error*
 x : bagian dari *cluster* C_i
 k : jumlah *cluster* C_i

3.1.3 Eksperimen

Tahap ini dilakukan proses *clustering* mulai dari input data, proses *clustering* dengan algoritma DBSCAN dan *K-Means* sehingga diperoleh output berupa *cluster* gempa bumi. Peneliti menggunakan *software Jupyter Notebook* dengan *Python* versi 3.9.12 untuk melakukan pengolahan data pada penelitian ini.

Pertama, sebelum melakukan proses *clustering* dengan algoritma DBSCAN terlebih dahulu harus ditentukan nilai epsilon (ϵ) yang merupakan radius

maksimum antara satu titik ke titik lain dan jumlah minimum objek (MinPts) yang merupakan titik minimum dalam *cluster* yang terbentuk. Penentuan nilai epsilon (ϵ) dan MinPts yang optimal berpengaruh terhadap performa hasil *clustering*. Pada penelitian ini untuk menentukan nilai epsilon (ϵ) dan MinPts menggunakan metode KNN (*k-Nearest Neighbors*).

Eksperimen kedua adalah melakukan *clustering* gempa bumi menggunakan algoritma *K-Means*. Sebelum melakukan *clustering*, terlebih dahulu menentukan jumlah *k* optimal dengan menggunakan KNN dengan 2 metode perhitungan jarak yaitu *Euclidean Distance* dan *Manhattan Distance*. Adapun langkah-langkah dalam menentukan epsilon (ϵ) dan MinPts dengan KNN dan menentukan jumlah *k* optimal menggunakan sampel data adalah sebagai berikut:

3.1.3.1 Clustering DBSCAN dengan Epsilon dan MinPts menggunakan Metode *Euclidean Distance*

Perhitungan jarak untuk menentukan nilai Epsilon dan MinPts dengan metode KNN dapat dilakukan dengan beberapa persamaan, salah satunya adalah dengan *Euclidean Distance* sebagai berikut:

1. Menentukan nilai *k* untuk menentukan jumlah tetangga terdekat pada perhitungan epsilon (ϵ). Persamaan perhitungan jarak pada KNN menggunakan *Euclidean Distance*.

Implementasi pada penelitian ini untuk data gempa bumi tahun 2017 menggunakan *Jupyter Notebook* dengan *Python* versi 3.9.12 ditampilkan pada penjelasan berikut ini.

- Menyiapkan *library* yang akan digunakan sebagai berikut:

```
#library yang digunakan untuk DBSCAN clustering
import pandas as pd
from sklearn.cluster import DBSCAN
import matplotlib.pyplot as plt
import numpy as np
from sklearn.metrics import silhouette_samples, silhouette_score
from collections import Counter
from sklearn.preprocessing import StandardScaler
from pylab import rcParams
import geopandas as gpd
import seaborn as sns
from descartes import PolygonPatch
```

Algoritma 1. Library *Python* yang digunakan pada clustering DBSCAN

- Menyiapkan dataset yang akan digunakan
- Melakukan perhitungan jarak dengan KNN *Euclidean Distance* untuk menentukan epsilon (ϵ) dan MinPts optimal dengan perintah sebagai berikut.

```
#menentukan jumlah N

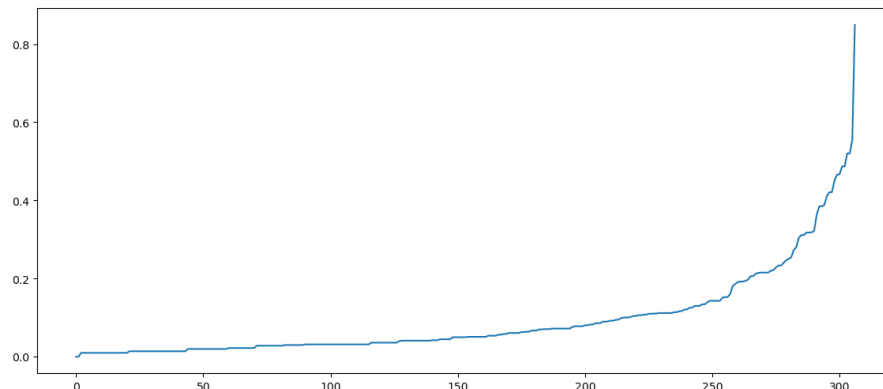
from sklearn.neighbors import NearestNeighbors // untuk menghitung KNN

nn = NearestNeighbors(n_neighbors=20) // menentukan jumlah N
y_nn = nn.fit(features)
distances, indices = y_nn.kneighbors(features)

distances_sorted = np.sort(distances, axis=0)
plt.plot(distances_sorted[:,1])
plt.show()
```

Algoritma 2. Perintah untuk menentukan jumlah *N* terdekat dengan KNN *Euclidean Distance*

Sehingga diperoleh kurva *k-dist* yang menunjukkan rentang nilai epsilon (ϵ) yang selanjutnya digunakan untuk menentukan nilai epsilon (ϵ) dan MinPts optimal.



Gambar 3.7 Kurva *k-dist* perhitungan nilai epsilon (ϵ) *Euclidean Distance*
 Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan kurva *k-dist* di atas diperoleh nilai epsilon (ϵ) dengan rentang 0,2-0,5 dan selanjutnya dari nilai tersebut dapat dihitung nilai epsilon (ϵ) dan MinPts optimal yang digunakan untuk clustering gempa bumi DBSCAN.

- Melakukan perhitungan epsilon (ϵ), MinPts dan *Silhouette Score*.

```
#menentukan epsilon, MinPts dan Silhouette Score terbaik

min_samples = range(10,21)
eps = np.arange(0.2,0.5, 0.01) # returns array of ranging from 0.2 to 0.5 with step of 0.01, that
range value is get from above plot's elbow point

output = []

for ms in min_samples:
    for ep in eps:
        labels = DBSCAN(min_samples=ms, eps = ep).fit(features).labels_
        score = silhouette_score(features, labels)
        output.append((ms, ep, score))

min_samples_opt, eps_opt, score_opt = sorted(output, key=lambda x:x[-1])[-1]
eps_opt_round = round(eps_opt,2)
print(f"Best silhouette_score: {score_opt}")
print(f"min_samples: {min_samples_opt}")
print(f"eps: {eps_opt_round}")

Best silhouette_score value: 0.41338732571989667
min_samples value: 18
eps value: 0.49
```

Algoritma 3. Perintah untuk menentukan nilai epsilon (ϵ), MinPts dan *Silhouette Score* terbaik

Berdasarkan perhitungan pada algoritma 3, diperoleh nilai epsilon (ϵ), MinPts dan *Silhouette Score* terbaik adalah 0,49, 18 dan 0,4133.

- Melakukan perhitungan *cluster* dengan menggunakan nilai epsilon (ϵ) dan MinPts optimal dengan perintah sebagai berikut.

```
df_opt = df.copy()

# create DBSCAN model/object
dbscan = DBSCAN(eps=eps_opt_round, min_samples=min_samples_opt)

# do clustering
y = dbscan.fit(features)
# save results
labels = y.labels_

# send back into dataframe and display it
df_opt['Cluster'] = labels
# display the number of member each clustering
df_opt['Cluster'].value_counts()

2  134
-1   61
0   59
1   53
```

Algoritma 4. Perintah *clustering* DBSCAN dengan eps dan MinPts optimal

Berdasarkan perintah *clustering* pada Algoritma 4, diperoleh jumlah *cluster* sebanyak 3 *cluster* yaitu *cluster* 0 = 59 anggota, *cluster* 1 = 53 anggota dan *cluster* 2 = 134 anggota serta *noise* yang ditandai dengan angka -1 = 61 anggota. Adapun hasil *clustering* disajikan seperti tabel berikut ini.

Tabel 3. 2 Hasil *clustering* DBSCAN *Euclidean Distance* dengan eps dan MinPts optimal

No.	Latitude	Longitude	Magnitudo	Depth	Cluster
0	-8.96	110.00	4.0	10	0
1	-11.73	113.26	4.9	34	-1
2	-8.18	108.28	4.3	31	1
3	-8.73	110.61	3.3	35	0
4	-8.50	108.89	3.9	11	1
...
302	-9.73	112.88	4.8	58	2
303	-8.79	112.47	3.1	27	2

No.	Latitude	Longitude	Magnitudo	Depth	Cluster
304	-9.04	111.18	4.1	10	0
305	-9.77	112.86	4.2	10	2
306	-9.80	112.87	4.0	10	2

Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan Tabel 3.2 dari 307 data yang digunakan setelah dilakukan *clustering* diperoleh 3 *cluster* yaitu 0, 1 dan 2 serta 1 *noise* yang ditandai dengan angka -1.

- Melakukan uji performa hasil *clustering* menggunakan *Silhouette Coefficient*.

Adapun implementasinya dengan *Python* adalah sebagai berikut.

```
#perhitungan silhouette coefficient
from sklearn.metrics import silhouette_score as ss
ss(features, df_opt['Cluster'])
0.41338732571989667
```

Algoritma 5. Perintah Perhitungan *Silhouette Coefficient*

Berdasarkan Algoritma 5 diperoleh nilai *Silhouette Coefficient* sebesar 0.413387.

3.1.3.2 Clustering DBSCAN dengan Epsilon dan MinPts Menggunakan Metode *Manhattan Distance*

Perhitungan jarak untuk menentukan nilai Epsilon dan MinPts dengan metode KNN dapat dilakukan dengan beberapa persamaan, salah satunya adalah dengan *Manhattan Distance* sebagai berikut:

- Menentukan nilai k untuk menentukan jumlah tetangga terdekat pada perhitungan epsilon (ϵ). Persamaan perhitungan jarak pada KNN menggunakan *Manhattan Distance* dijelaskan pada persamaan 3.4.

Implementasi pada penelitian ini untuk data gempa bumi tahun 2017 *Jupyter Notebook* dengan *Python* versi 3.9.12 ditampilkan pada penjelasan berikut ini.

- Menyiapkan library yang akan digunakan seperti pada gambar 3.5 di atas.
- Menyiapkan dataset yang akan digunakan.
- Melakukan perhitungan jarak dengan KNN *Manhattan Distance* untuk menentukan epsilon (ϵ) dan MinPts optimal dengan perintah sebagai berikut.

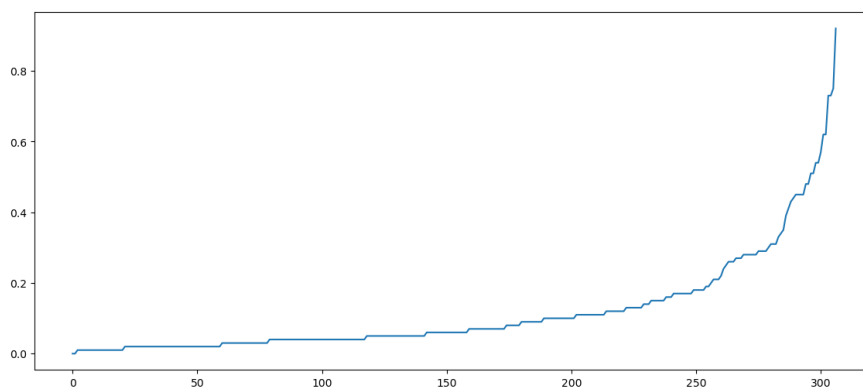
```
#menentukan jumlah N terdekat dengan KNN Manhattan Distance
from sklearn.neighbors import NearestNeighbors // menentukan NN

nn = NearestNeighbors(n_neighbors=20, metric='manhattan') // menghitung NN
y_nn = nn.fit(features)
distances, indices = y_nn.kneighbors(features)

distances_sorted = np.sort(distances, axis=0)
plt.plot(distances_sorted[:,1])
plt.show()
```

Algoritma 6. Perintah untuk menentukan jumlah N terdekat dengan KNN *Manhattan Distance*

Sehingga diperoleh kurva *k-dist* yang menunjukkan rentang nilai epsilon (ϵ) yang selanjutnya digunakan untuk menentukan nilai epsilon (ϵ) dan MinPts optimal.



Gambar 3.8 Kurva *k-dist* perhitungan nilai epsilon (ϵ) *Manhattan Distance*
Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan kurva *k-dist* gambar 3.14 diperoleh nilai epsilon (ϵ) dengan rentang 0,2-0,6 dan selanjutnya dari nilai tersebut dapat dihitung nilai epsilon

(ϵ) dan MinPts optimal yang digunakan untuk *clustering* gempa bumi DBSCAN.

- Melakukan perhitungan epsilon (ϵ), MinPts dan *Silhouette Score*.

```
#menentukan nilai epsilon, MinPts dan Silhouette Score

min_samples = range(10,21)
eps = np.arange(0.2,0.6, 0.01) # returns array of ranging from 0.2 to 0.6 with step of
0.01, that range value is get from above plot's elbow point

output = []

for ms in min_samples:
    for ep in eps:
        labels = DBSCAN(min_samples=ms, eps = ep).fit(features).labels_
        score = silhouette_score(features, labels)
        output.append((ms, ep, score))

min_samples_opt, eps_opt, score_opt = sorted(output, key=lambda x:x[-1])[-1]
eps_opt_round = round(eps_opt,2)
print(f"Best silhouette_score: {score_opt}")
print(f"min_samples: {min_samples_opt}")
print(f"eps: {eps_opt_round}")

Best silhouette_score value: 0.41771788579712227
min_samples value: 20
eps value: 0.51
```

Algoritma 7. Perintah untuk menentukan nilai epsilon (ϵ), MinPts dan *Silhouette Score*

Berdasarkan perhitungan Algoritma 7, diperoleh nilai epsilon (ϵ), MinPts dan *Silhouette Score* terbaik adalah 0,51, 20 dan 0,417717.

- Melakukan perhitungan *cluster* dengan menggunakan nilai epsilon (ϵ) dan MinPts optimal dengan perintah seperti pada Algoritma 7. Sehingga diperoleh hasil *clustering* sebagai berikut.

Tabel 3.3 Hasil *clustering* DBSCAN *Manhattan Distance* dengan eps dan MinPts optimal

No.	Latitude	Longitude	Magnitudo	Depth	Cluster
0	-8.96	110.00	4.0	10	0
1	-11.73	113.26	4.9	34	-1
2	-8.18	108.28	4.3	31	1

No.	Latitude	Longitude	Magnitudo	Depth	Cluster
3	-8.73	110.61	3.3	35	0
4	-8.50	108.89	3.9	11	1
...
302	-9.73	112.88	4.8	58	2
303	-8.79	112.47	3.1	27	2
304	-9.04	111.18	4.1	10	0
305	-9.77	112.86	4.2	10	2
306	-9.80	112.87	4.0	10	2

Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan Tabel 3.3 dari 307 data yang digunakan setelah dilakukan *clustering* diperoleh 3 *cluster* yaitu 0, 1 dan 2 serta 1 *noise* yang ditandai dengan angka -1.

- Melakukan uji performa hasil *clustering* menggunakan *Silhouette Coefficient*. Adapun persamaan perhitungannya seperti pada Algoritma 5 sehingga diperoleh nilai *Silhouette Coefficient* sebesar 0,417717.

Setelah dilakukan *clustering* dengan algoritma DBSCAN dengan perhitungan jarak menggunakan persamaan *Euclidean Distance* dan *Manhattan Distance*, selanjutnya melakukan *clustering* dengan algoritma *K-Means* sebagai pembandingan untuk mendapatkan hasil *clustering* terbaik.

3.1.3.3 Clustering *K-Means* dengan *K* Optimal Menggunakan Metode *Euclidean Distance*

Perhitungan *k* optimal dengan metode KNN dengan perhitungan jarak menggunakan *Euclidean Distance* pada *Jupyter Notebook* dengan *Python* versi 3.9.12 ditampilkan pada penjelasan berikut ini.

1. Menyiapkan library *Python* yang digunakan pada perhitungan *clustering* sebagai berikut.

```
#library K-Means clustering

import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn.metrics import silhouette_samples, silhouette_score
from collections import Counter
from sklearn.preprocessing import StandardScaler
from pylab import rcParams
import geopandas as gpd
import seaborn as sns
from descartes import PolygonPatch
from sklearn.cluster import Kmeans           // clustering k-means
from sklearn.metrics import silhouette_score // menentukan silhouette
```

Algoritma 8. Library *Python* yang digunakan pada clustering *K-Means*

2. Menyiapkan dataset yang akan digunakan.
3. Melakukan perhitungan jarak dengan KNN *Euclidean Distance* untuk menentukan *k* optimal dengan perintah sebagai berikut.

```
#menentukan jumlah k optimal dengan Euclidean Distance

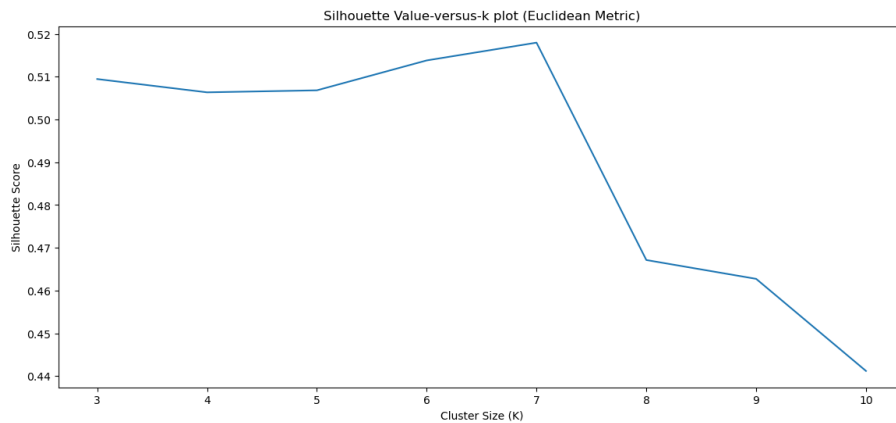
sil = []
k_size = []
kmax = 10

# dissimilarity would not be defined for a single cluster, thus, minimum number of clusters should be 2
for k in range(3, kmax+1):
    kmeans = KMeans(
        init="random",
        n_clusters=k,
        n_init=10,
        max_iter=300,
        random_state=42
    ).fit(features)
    labels = kmeans.labels_
    sil.append(silhouette_score(features, labels, metric = 'euclidean'))
    k_size.append(k)

plt.plot(k_size,sil)
plt.xlabel('Cluster Size (K)'); plt.ylabel('Silhouette Score'); plt.title('Silhouette Value-versus-k plot (Euclidean Metric)');
plt.show()                                     // plot silhouette
```

Algoritma 9. Perintah untuk menentukan jumlah *k* optimal dengan *Euclidean Distance*

Sehingga diperoleh kurva *maximum Silhouette Score* yang menunjukkan nilai *Silhouette* untuk menentukan *k* optimal seperti yang ditunjukkan pada gambar berikut ini.



Gambar 3.9 Kurva nilai *silhouette* terhadap nilai *k* optimal dengan *Euclidean Distance*

Sumber: Diolah penulis dari hasil penelitian (2023)

- Melakukan perhitungan *k* optimal berdasarkan *Silhouette Score* dengan perintah sebagai berikut:

```
#menentukan nilai k optimal

df_k_e = pd.DataFrame({'K': k_size, 'Silhouette Score':sil })
opt_df_e = df_k_e[df_k_e['Silhouette Score'] == df_k_e['Silhouette Score'].max()]
optimized_k_e = opt_df_e['K']
print(optimized_k_e)
4 7
Name: K, dtype: int64
```

Algoritma 10. Perintah untuk menentukan nilai *k* optimal berdasarkan *Silhouette Score* terbaik

Berdasarkan Algoritma 10, diperoleh nilai *k* optimal sebanyak 7 *cluster* yang selanjutnya digunakan untuk perhitungan jumlah dan anggota *cluster*.

- Melakukan *clustering* menggunakan *k* optimal dengan perintah sebagai berikut.

```

#K-Means clustering dengan k optimal
df_opt_e = df.copy()

kmeans_opt_e = KMeans(
    init="random",
    n_clusters=optimized_k_e.iloc[0],
    n_init=10,
    max_iter=300,
    random_state=42
)

# do clustering
y_opt_e = kmeans_opt_e.fit(features)
# save results
labels_opt_e = y_opt_e.labels_
df_opt_e['Cluster'] = labels_opt_e

df_opt_e['Cluster'] = labels_opt_e
df

df_opt_e['Cluster'].value_counts()
1    94
3    62
4    50
0    50
6    22
5    20
2     9
Name: Cluster, dtype: int64

```

Algoritma 11. Perintah *clustering K-Means* dengan *k* optimal

Berdasarkan perintah clustering pada Algoritma 11, diperoleh jumlah *cluster* sebanyak 7 *cluster* yaitu *cluster* 0 = 50 anggota, *cluster* 1 = 94 anggota, *cluster* 2 = 9 anggota, *cluster* 3 = 62 anggota, *cluster* 4 = 50 anggota, *cluster* 5 = 20 anggota dan *cluster* 6 = 22 anggota. Adapun hasil *clustering* ditampilkan pada Tabel 3.4.

Tabel 3.4 Hasil *clustering K-Means Euclidean Distance* dengan *k* optimal

No.	Latitude	Longitude	Magnitudo	Depth	Cluster
0	-8.96	110.00	4.0	10	5
1	-11.73	113.26	4.9	34	6
2	-8.18	108.28	4.3	31	7
3	-8.73	110.61	3.3	35	5
4	-8.50	108.89	3.9	11	7
...

No.	Latitude	Longitude	Magnitudo	Depth	Cluster
302	-9.73	112.88	4.8	58	4
303	-8.79	112.47	3.1	27	4
304	-9.04	111.18	4.1	10	3
305	-9.77	112.86	4.2	10	4
306	-9.80	112.87	4.0	10	4

Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan Tabel 3.4 hasil *clustering* dari 307 data gempabumi diperoleh jumlah *cluster* sebanyak 7 *cluster* yaitu *cluster* 0, 1, 2, 3, 4, 5 dan 6.

- Melakukan uji performa hasil *clustering* menggunakan *Silhouette Coefficient* dengan perintah sebagai berikut.

```
from sklearn.metrics import silhouette_score as ss
ss(features, df_opt_e['Cluster'])

0.5179622304906369
```

Algoritma 12. Perintah Perhitungan *Silhouette Coefficient*

Berdasarkan Algoritma 12 diperoleh nilai *Silhouette Coefficient* sebesar 0.517962230.

3.1.3.4 Clustering *K-Means* dengan *K* Optimal Menggunakan Metode *Manhattan Distance*

Setelah dilakukan clustering gempabumi menggunakan *K-Means* dengan *Euclidean Distance*, selanjutnya pada eksteperimen kedua penulis melakukan *clustering* gempabumi menggunakan *Manhattan Distance*. Adapun implementasi pada *Python* langkah-langkahnya sebagai berikut:

- Menyiapkan library *Python* yang akan digunakan sebagai berikut.

```

#library K-Means clustering dengan Manhattan Distance

import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn.metrics import silhouette_samples, silhouette_score
from collections import Counter
from sklearn.preprocessing import StandardScaler
from pylab import rcParams
import geopandas as gpd
import seaborn as sns
from descartes import PolygonPatch
from sklearn.cluster import Kmeans // clustering k-means
from sklearn.metrics import silhouette_score // menentukan silhouette

# pyclustering kmeans
import random

from pyclustering.samples.definitions import SIMPLE_SAMPLES, FCPS_SAMPLES,
FAMOUS_SAMPLES

from pyclustering.cluster import cluster_visualizer_multidim
from pyclustering.cluster.center_initializer import kmeans_plusplus_initializer
from pyclustering.cluster.kmeans import kmeans, kmeans_observer, kmeans_visualizer

from pyclustering.utils import read_sample
from pyclustering.utils import timedcall
from pyclustering.utils.metric import distance_metric, type_metric

```

Algoritma 13. Library *Python* yang digunakan pada *clustering K-Means* dengan *Manhattan Distance*

2. Menyiapkan dataset.
3. Melakukan perhitungan jarak dengan *KNN Manhattan Distance* untuk menentukan *k* optimal berdasarkan *maximum correct centroid value*.
4. Melakukan perhitungan *k* optimal berdasarkan *maximum correct centroid value* dengan perintah sebagai berikut.

```

#menentukan k optimal berdasarkan maximum correct value
df_k_m = pd.DataFrame({'K': k_size, 'Total CC':cc })
opt_df = df_k_m[df_k_m['Total CC'] == df_k_m['Total CC'].max()]
optimized_k = opt_df['K']

print(opt_df)
K Total CC
1 4 65

```

Algoritma 14. Perintah untuk menentukan nilai *k* optimal berdasarkan *maximum correct centroid*

5. Melakukan *clustering* menggunakan k optimal dengan perintah seperti pada Algoritma 11.

Sehingga diperoleh hasil *clustering* sebanyak 4 *cluster* yaitu, *cluster* 0 = 65 anggota, *cluster* 1 = 127 anggota, *cluster* 2 = 98 anggota dan *cluster* 3 = 17 anggota.

6. Melakukan uji performa hasil *clustering* menggunakan *Silhouette Coefficient* dengan perintah seperti pada Algoritma 12 sehingga diperoleh nilai *Silhouette Coefficient* 0,506346411.

3.2 Instrumen Penelitian

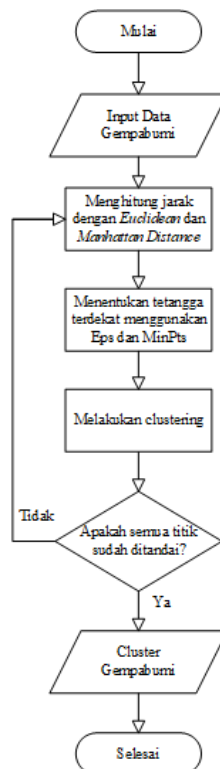
Instrumen penelitian merupakan bagian terpenting dalam penelitian karena menentukan baik atau tidaknya hasil yang diperoleh. Pada penelitian ini menggunakan instrumen penelitian berupa data gempa bumi wilayah Regional VII meliputi Jawa Tengah, Yogyakarta, dan Jawa Timur tahun 2017-2021 dengan batasan koordinat 6°S - 14°S dan 108°E - 114°E . Penulis menggunakan variabel independen berupa *latitude* dan *longitude* dan variabel dependen adalah *cluster* gempa bumi.

BAB IV

METODE DBSCAN

4.1 Desain

Peneliti menggunakan metode *Density Based Spatial Clustering of Application with Noise* atau DBSCAN untuk melakukan *clustering* data gempabumi di wilayah Regional VII sesuai dengan Alur Algoritma DBSCAN pada Gambar 3.3. Selanjutnya, peneliti melakukan *clustering* data gempa menggunakan metode DBSCAN untuk menghitung jumlah *cluster* berdasarkan kepadatan. Algoritma DBSCAN merupakan salah satu pendekatan algoritma *machine learning* berbasis kepadatan untuk analisis *cluster*. Proses *clustering* dijelaskan pada Gambar 4.1.



Gambar 4.1 Diagram Alir Algoritma DBSCAN
Sumber: (Pakuani & Kurniawan, 2021)

Gambar 4.1 menjelaskan tentang proses *clustering* gempa bumi menggunakan algoritma DBSCAN mulai dari menggunakan data gempa bumi untuk menentukan nilai k jumlah tetangga terdekat pada perhitungan epsilon (ϵ) dengan KNN. Sebelumnya, dari data gempa bumi yang digunakan sebanyak 1.827 data gempa bumi di wilayah Regional VII selama 5 tahun, yaitu tahun 2017 hingga 2021 yang bersumber dari BMKG yang terdiri dari variabel *Longitude*, *Latitude*, *Magnitudo* dan *Depth*. Kemudian penulis membagi data tersebut menjadi 2 berdasarkan periode tahun kejadian, yaitu periode tahunan dan periode 5 tahun. Hal ini dilakukan bertujuan untuk mengetahui pola sebaran *cluster* gempa bumi berdasarkan periode tahun kejadian. Penjelasan mengenai periode tahunan dan 5 tahun sebagai berikut.

T_{1_2017} : Periode untuk data gempa bumi tahun 2017.

T_{1_2018} : Periode untuk data gempa bumi tahun 2018

T_{1_2019} : Periode untuk data gempa bumi tahun 2019

T_{1_2020} : Periode untuk data gempa bumi tahun 2020

T_{1_2021} : Periode untuk data gempa bumi tahun 2021

T_{5_2017-2021} : Periode untuk data gempa bumi tahun 2017-2021

Pertama, menentukan nilai k jumlah tetangga terdekat untuk menentukan nilai epsilon (ϵ) untuk data gempa bumi periode waktu tahunan menggunakan persamaan *Euclidean Distance* dengan persamaan sebagai berikut.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4.1)$$

di mana:

x_i : titik x ke i

y_i : titik y ke i

Implementasi perhitungannya penulis menggunakan Jupyter Notebook dengan bahasa pemrograman *Python* dengan memperhatikan grafik *Elbow Method* yang merupakan selisih kuadrat dari nilai k yang berbeda (Cui, 2020) sebagai acuan untuk menentukan rentang nilai epsilon (ϵ) dan MinPts dengan perhitungan sebagai berikut:

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} |x - C_i|^2 \quad (4.2)$$

di mana:

SSE : *Sum of Squares Error*

x : bagian dari *cluster* C_i

k : jumlah *cluster* C_i

Adapun langkah selengkapnya dijelaskan pada bagian selanjutnya berikut ini.

4.2 Uji Coba

Langkah selanjutnya adalah melakukan implementasi algoritma DBSCAN dengan menerapkan 2 metode perhitungan jarak untuk menentukan nilai epsilon (ϵ) dan MinPts optimal yaitu, metode *Euclidean Distance* dan *Manhattan Distance*. Adapun langkah-langkah 2 metode perhitungan jarak tersebut dijelaskan pada bagian berikut ini.

4.2.1 *Clustering* DBSCAN dengan Optimasi Epsilon (ϵ) dan MinPts Menggunakan *Euclidean Distance*

Clustering gempa bumi dibagi menjadi 2 bagian yaitu, *clustering* berdasarkan periode tahunan (T_{1_2017} , T_{1_2018} , T_{1_2019} , T_{1_2020} , T_{1_2021}) dan periode 5 tahun ($T_{1_2017-2021}$) berdasarkan pada penelitian ini. Dari data gempa bumi selama rentang waktu 5 tahun (2017-2021) terdapat sebanyak 1.827 data gempa bumi di wilayah Regional VII, kemudian data tersebut dibagi berdasarkan periode tahun kejadian sehingga diperoleh jumlah data untuk masing-masing tahun mulai tahun 2017 hingga 2021 sebagai berikut.

Tabel 4.1 Jumlah Kejadian Gempabumi Berdasarkan Periode Waktu

Periode	T_{1_2017}	T_{1_2018}	T_{1_2019}	T_{1_2020}	T_{1_2021}	$T_{1_2017-2021}$
Jumlah	307	464	417	289	350	1827

Sumber: Diolah penulis dari hasil penelitian (2023)

Selanjutnya, melakukan perhitungan nilai epsilon (ϵ) dan MinPts optimal untuk mendapatkan hasil *clustering* terbaik. Berdasarkan perhitungan epsilon (ϵ) dan MinPts optimal menggunakan persamaan *Euclidean Distance* pada pembahasan sebelumnya, maka diperoleh nilai epsilon (ϵ) dan MinPts untuk data gempa bumi tiap periode adalah sebagai berikut:

Tabel 4.2 Nilai Epsilon (ϵ) dan MinPts Optimal Berdasarkan Periode Waktu dengan *Euclidean Distance*

Periode	T_{1_2017}	T_{1_2018}	T_{1_2019}	T_{1_2020}	T_{1_2021}	$T_{1_2017-2021}$
Epsilon (ϵ) opt	0,49	0,39	0,47	0,49	0,33	0,19
MinPts opt	18	20	20	14	12	18

Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan hasil nilai epsilon (ϵ) dan MinPts optimal pada tabel di atas kemudian dilakukan *clustering* gempa bumi menggunakan algoritma DBSCAN untuk data gempa bumi tiap periode waktu. Adapun hasilnya diperoleh seperti Tabel 4.3.

Tabel 4.3 Hasil *Clustering* DBSCAN Berdasarkan Periode Waktu

Periode	T _{1_2017}	T _{1_2018}	T _{1_2019}	T _{1_2020}	T _{1_2021}	T _{1_2017-2021}
Jumlah cluster	3 cluster dan 1 noise	3 cluster dan 1 noise	4 cluster dan 1 noise	3 cluster dan 1 noise	4 cluster dan 1 noise	6 cluster dan 1 noise

Sumber: Diolah penulis dari hasil penelitian (2023)

Setelah diperoleh jumlah *cluster* berdasarkan nilai epsilon (ϵ) dan MinPts optimal menggunakan matriks jarak *Euclidean Distance*, langkah berikutnya adalah melakukan uji coba *cluster* menggunakan metode *Silhouette Coefficient*. Adapun langkah selengkapnya dijelaskan pada bagian berikutnya.

4.2.1.1 Uji *Cluster* DBSCAN dengan Optimasi Epsilon (ϵ) dan MinPts

Menggunakan *Euclidean Distance*

Setelah diperoleh nilai epsilon (ϵ) dan MinPts optimal serta jumlah *cluster* dari masing-masing data gempa bumi berdasarkan periode waktu tahunan dan 5 tahun, selanjutnya untuk mengetahui kualitas *cluster* yang terbentuk maka dilakukan evaluasi terhadap *clustering* masing-masing periode dengan menghitung perbandingan antara selisih rata-rata jarak antara titik I ke seluruh titik selain *cluster* dengan rata-rata jarak antara titik I dengan seluruh titik di dalam satu *cluster* terhadap nilai maksimal rata-rata jarak keduanya atau biasa disebut dengan *Silhouette Coefficient*. Rata-rata dari semua *Silhouette* dalam dataset merupakan

Silhouette rata-rata untuk semua sampel dalam dataset (Zhihong & Jie, 2019).

Perhitungan *Silhouette Coefficient* dijelaskan pada persamaan berikut ini:

$$si = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3.3)$$

di mana:

si : nilai *Silhouette Coefficient*

$a(i)$: rata-rata jarak antara titik I dengan seluruh titik dalam A
(*cluster* di mana titik I berada).

$b(i)$: rata-rata jarak antara titik I ke seluruh titik dalam *cluster* selain A.

Implementasi perhitungannya dengan *Python* yang dijalankan menggunakan Jupyter Notebook dijelaskan pada bagian berikut ini.

4.2.1.2 Perhitungan *Silhouette Coefficient*

Silhouette Coefficient diperoleh dengan menghitung selisih rata-rata jarak antara titik I dengan seluruh titik dalam *cluster* dengan rata-rata jarak antara titik I ke seluruh titik selain *cluster* dilakukan terhadap seluruh *cluster* yang terbentuk dibagi dengan rata-rata jarak maksimal antara keduanya kemudian dirata-rata.

Langkah tersebut dilakukan terhadap data *cluster* gempa bumi yang terbentuk untuk tiap-tiap periode waktu. Sehingga diperoleh nilai *Silhouette Coefficient* untuk cluster pada periode waktu tahunan dan 5 tahun ditampilkan pada Tabel 4.4.

Tabel 4.4 Nilai *Silhouette Coefficient* DBSCAN Berdasarkan Periode Waktu

Periode	T _{1_2017}	T _{1_2018}	T _{1_2019}	T _{1_2020}	T _{1_2021}	T _{1_2017-2021}
<i>Silhouette Coefficient</i>	0,413	0,376	0,403	0,486	0,368	0,208

Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan hasil perhitungan *Silhouette Coefficient* pada Tabel 4.4 diperoleh nilai *Silhouette Coefficient* tertinggi adalah *cluster* gempabumi untuk periode T_{1_2020} sebesar 0.486 dan terendah pada *cluster* gempabumi untuk periode $T_{1_2017-2021}$ sebesar 0.208.

4.2.2 Clustering DBSCAN dengan Optimasi Epsilon (ϵ) dan MinPts Menggunakan *Manhattan Distance*

Penentuan nilai epsilon (ϵ) dan MinPts optimal selanjutnya adalah menggunakan perhitungan *Manhattan Distance*. Adapun langkah-langkah *clustering* pada bagian ini hampir sama dengan *clustering* dengan menggunakan *Euclidean Distance* dalam penentuan epsilon (ϵ) dan MinPts pada bagian sebelumnya, hanya saja matriks perhitungan epsilon (ϵ) dan MinPts nya yang berbeda. *Clustering* gempabumi dibagi menjadi 2 bagian yaitu, *clustering* berdasarkan periode tahunan (T_{1_2017} , T_{1_2018} , T_{1_2019} , T_{1_2020} , T_{1_2021}) dan periode 5 tahun ($T_{1_2017-2021}$). Sehingga diperoleh nilai epsilon (ϵ) dan MinPts optimal untuk tiap periode adalah sebagai berikut:

Tabel 4.5 Nilai Epsilon (ϵ) dan MinPts Optimal Berdasarkan Periode Waktu dengan *Manhattan Distance*

Periode	T_{1_2017}	T_{1_2018}	T_{1_2019}	T_{1_2020}	T_{1_2021}	$T_{1_2017-2021}$
Epsilon (ϵ) opt	0,51	0,41	0,47	0,57	0,33	0,24
MinPts opt	20	20	20	20	12	20

Sumber: Diolah penulis dari hasil penelitian (2023)

Selanjutnya, berdasarkan hasil nilai epsilon (ϵ) dan MinPts optimal pada Tabel 4.5 kemudian dilakukan *clustering* gempabumi menggunakan algoritma

DBSCAN untuk data gempa bumi tiap periode waktu. Adapun hasilnya diperoleh seperti pada Tabel 4.6.

Tabel 4.6 Hasil *Clustering* DBSCAN Berdasarkan Periode Waktu

Periode	T _{1_2017}	T _{1_2018}	T _{1_2019}	T _{1_2020}	T _{1_2021}	T _{1_2017-2021}
Jumlah cluster	3 cluster dan 1 noise	3 cluster dan 1 noise	4 cluster dan 1 noise	3 cluster dan 1 noise	4 cluster dan 1 noise	4 cluster dan 1 noise

Sumber: Diolah penulis dari hasil penelitian (2023)

Setelah diperoleh jumlah *cluster* berdasarkan nilai epsilon (ϵ) dan MinPts optimal menggunakan matriks jarak *Manhattan Distance*, langkah berikutnya adalah melakukan uji coba *cluster* menggunakan metode *Silhouette Coefficient*. Adapun langkah selengkapannya akan dijelaskan pada bagian selanjutnya.

4.2.2.1 Uji *Cluster* DBSCAN dengan Optimasi Epsilon (ϵ) dan MinPts

Menggunakan *Manhattan Distance*

Langkah berikutnya adalah melakukan uji *cluster* hasil *clustering* DBSCAN dengan menghitung selisih rata-rata jarak antara titik I dengan seluruh titik dalam *cluster* dengan rata-rata jarak antara titik I ke seluruh titik selain *cluster* dilakukan terhadap seluruh *cluster* yang terbentuk dibagi dengan jarak rata-rata maksimal antara keduanya kemudian dirata-rata atau dikenal dengan nama *Silhouette Coefficient*.

Adapun hasil yang diperoleh untuk *cluster* tiap periode disajikan dalam Tabel 4.7.

Tabel 4.7 Nilai *Silhouette Coefficient* DBSCAN Berdasarkan Periode Waktu

Periode	T _{1_2017}	T _{1_2018}	T _{1_2019}	T _{1_2020}	T _{1_2021}	T _{1_2017-2021}
Silhouette Coefficient	0,417	0,390	0,403	0,499	0,368	0,372

Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan hasil perhitungan *Silhouette Coefficient* pada Tabel 4.7 diperoleh nilai *Silhouette Coefficient* tertinggi adalah *cluster* gempa bumi untuk periode T_{1_2020} sebesar 0,499 dan terendah pada *cluster* gempa bumi untuk periode T_{1_2021} sebesar 0,368.

4.3 Hasil Clustering

Berdasarkan hasil implementasi *clustering* data gempa bumi berdasarkan periode waktu tahunan dan 5 tahun dengan menggunakan 2 uji coba perhitungan nilai epsilon (ϵ) dan MinPts optimal menggunakan *Euclidean Distance* dan *Manhattan Distance* dan setelah dilakukan uji coba validasi *cluster* menggunakan perhitungan *Silhouette Coefficient* maka diperoleh hasil sebagai berikut:

Tabel 4.8 Hasil *Clustering* Algoritma DBSCAN *Euclidean Distance*

Periode	Eps Optimal	MinPts Optimal	Jumlah Cluster	Silhouette Coefficient
T _{1_2017}	0,49	18	3 cluster dan 1 noise	0,413
T _{1_2018}	0,39	20	3 cluster dan 1 noise	0,376
T _{1_2019}	0,47	20	4 cluster dan 1 noise	0,403
T _{1_2020}	0,49	14	3 cluster dan 1 noise	0,486
T _{1_2021}	0,33	12	4 cluster dan 1 noise	0,368
T _{5_2017-2021}	0,19	18	6 cluster dan 1 noise	0,208

Sumber: Diolah penulis dari hasil penelitian (2023)

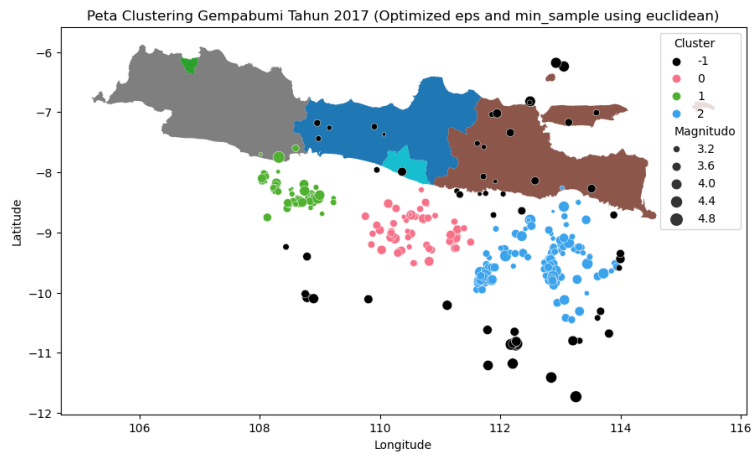
Tabel 4.9 Hasil *Clustering* Algoritma DBSCAN *Manhattan Distance*

Periode	Eps Optimal	MinPts Optimal	Jumlah Cluster	Silhouette Coefficient
T _{1_2017}	0,51	20	3 cluster dan 1 noise	0,417
T _{1_2018}	0,41	20	3 cluster dan 1 noise	0,390
T _{1_2019}	0,47	20	4 cluster dan 1 noise	0,403
T _{1_2020}	0,57	20	3 cluster dan 1 noise	0,499
T _{1_2021}	0,33	12	4 cluster dan 1 noise	0,368
T _{5_2017-2021}	0,24	20	4 cluster dan 1 noise	0,372

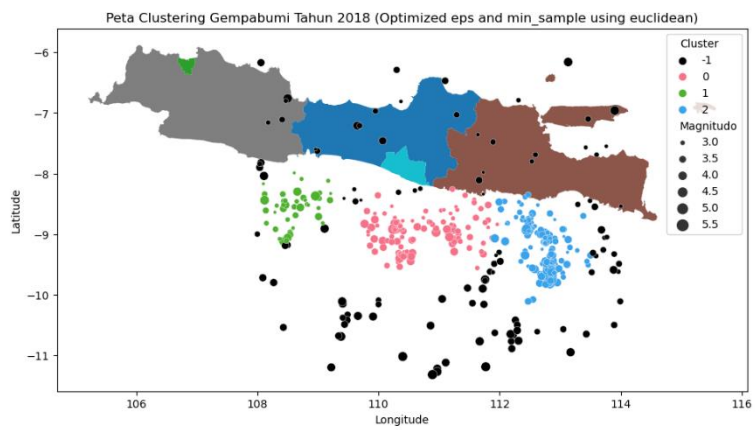
Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan Tabel 4.8 dan 4.9 didapatkan nilai *Silhouette Coefficient* tertinggi adalah 0,499 untuk *cluster* periode T_{1_2020} untuk perhitungan epsilon (ϵ) dan MinPts menggunakan *Manhattan Distance* dengan jumlah *cluster* yang terbentuk sebanyak 3 *cluster*. Sedangkan untuk nilai *Silhouette Coefficient* terendah adalah 0,208 untuk *cluster* periode T_{5_2020-2021} untuk perhitungan epsilon (ϵ) dan MinPts menggunakan *Euclidean Distance* dengan jumlah *cluster* yang terbentuk sebanyak 6 *cluster*.

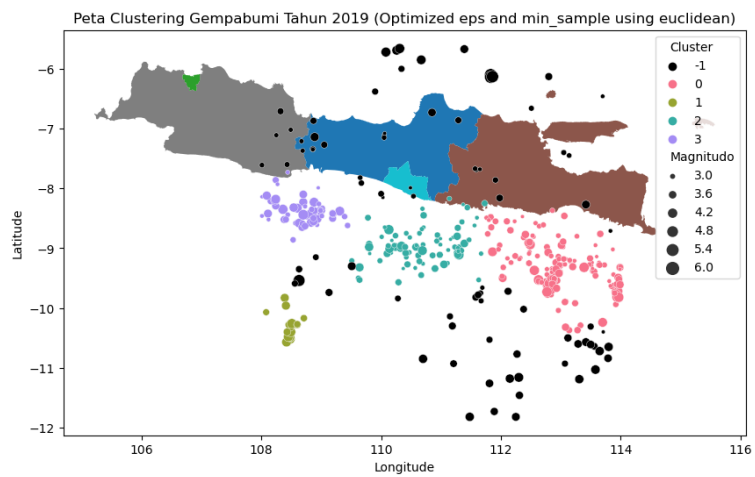
Setelah diperoleh jumlah *cluster* untuk tiap periode gempa bumi dengan perhitungan jarak menggunakan *Euclidean Distance* maupun *Manhattan Distance*, selanjutnya untuk mempermudah analisis *cluster* disandingkan dengan *golden standart* peta seismisitas yang telah ada sebelumnya, maka dilakukan *ploting* data dalam bentuk peta. Adapun visualisasi data hasil DBSCAN *clustering* ditampilkan pada Gambar 4.2.



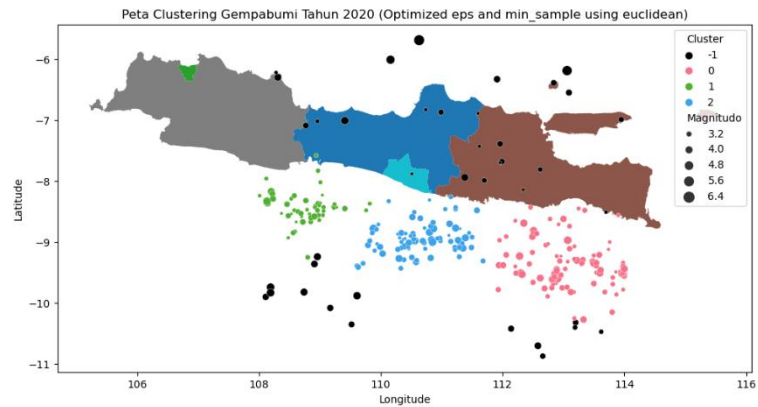
(a)



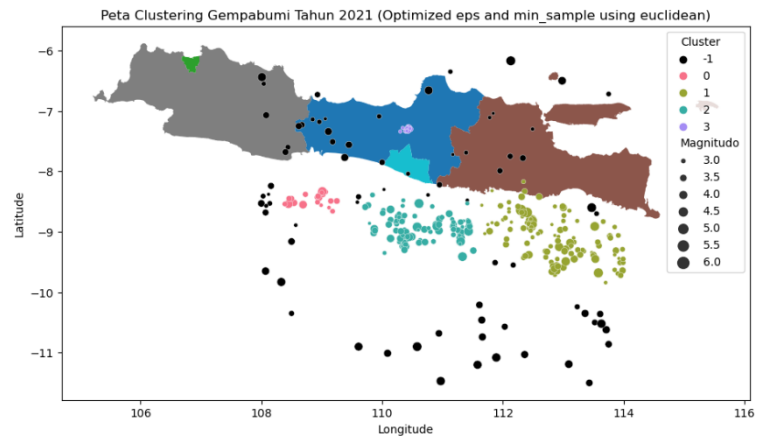
(b)



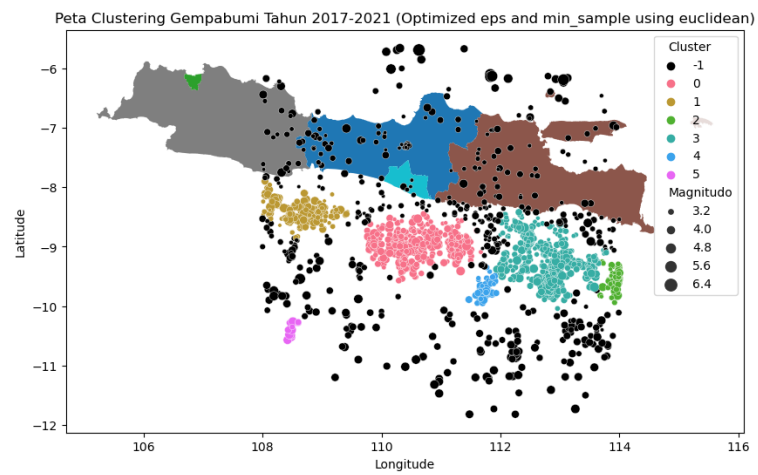
(c)



(d)



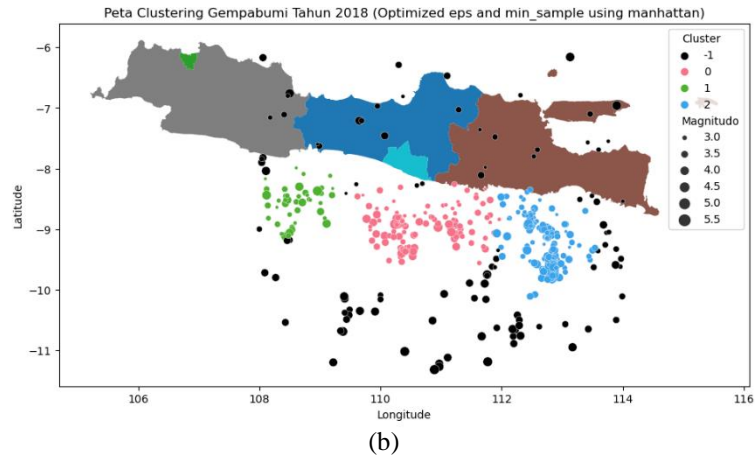
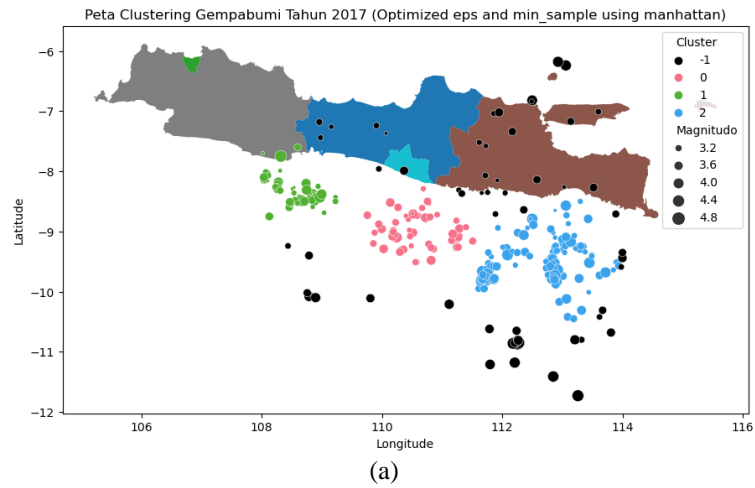
(e)

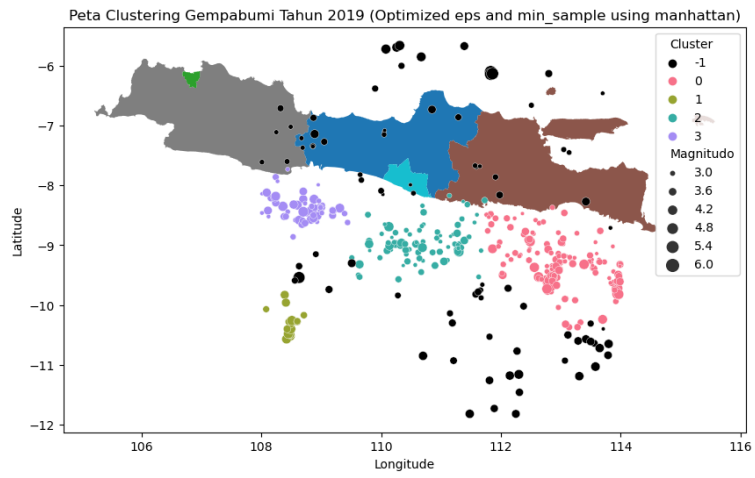


(f)

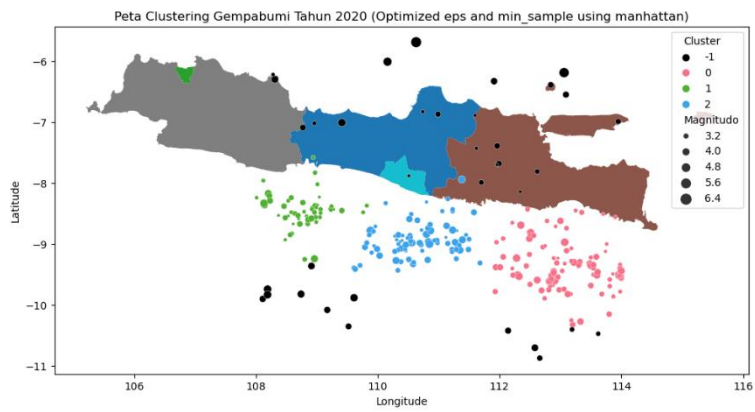
Gambar 4.2 Peta *Clustering* Gempabumi Tiap Periode dengan K Optimal menggunakan *Euclidean Distance*
 Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan Gambar 4.2 a, b, c, d, e dan f hasil *clustering* gempa bumi tiap periode dengan nilai epsilon (ϵ) dan MinPts optimal menggunakan *Euclidean Distance* diperoleh jumlah cluster 3, 4 dan 6 cluster serta 1 noise yang ditandai dengan warna hitam.

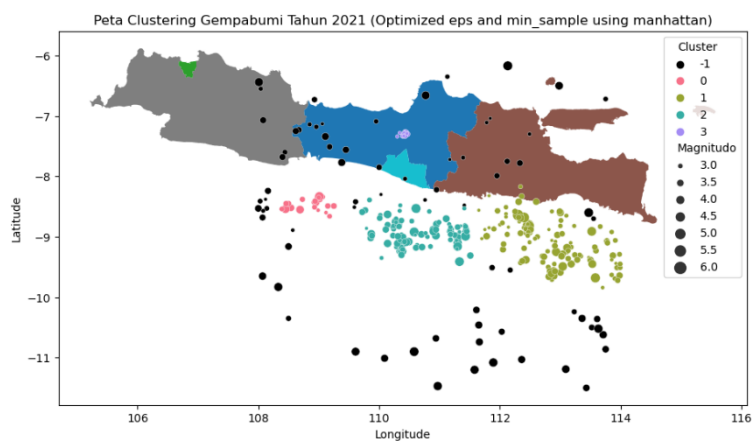




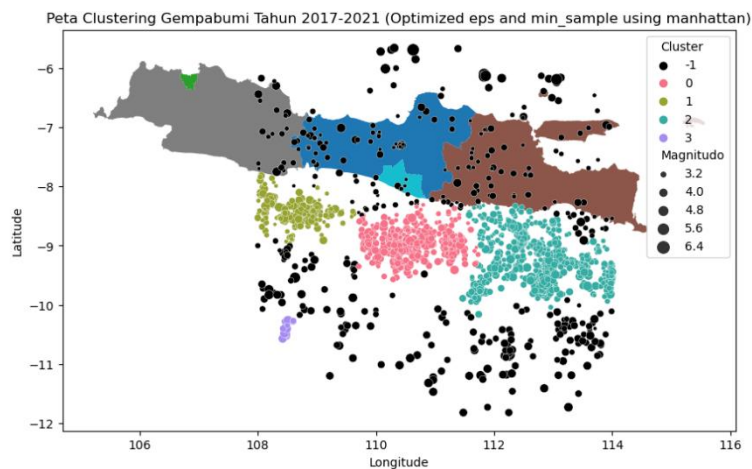
(c)



(d)



(e)



(f)

Gambar 4.3 Peta *Clustering* Gempabumi Tiap Periode dengan K Optimal menggunakan *Manhattan Distance*
 Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan gambar 4.3 a, b, c, d, e dan f hasil *clustering* menunjukkan jumlah *cluster* gempabumi yang terbentuk untuk tiap periode waktu sebanyak 3 dan 4 *cluster* dengan 1 *noise*.

4.4 Kesimpulan

Hasil analisis clustering menggunakan metode DBSCAN dengan perhitungan jarak menggunakan 2 metode yaitu, *Euclidean Distance* dan *Manhattan Distance* menghasilkan nilai epsilon (ϵ) dan MinPts optimal, jumlah *cluster* dan *Silhouette Coefficient* yang berbeda. *Clustering* DBSCAN dengan *Euclidean Distance* cenderung menghasilkan nilai epsilon (ϵ) dan MinPts optimal lebih kecil dibandingkan dengan *clustering* DBSCAN dengan *Manhattan Distance*. Hal ini berpengaruh terhadap jumlah *cluster* yang terbentuk DBSCAN dengan *Euclidean Distance* menghasilkan jumlah *cluster* maksimal lebih banyak dibanding

dengan DBSCAN dengan *Manhattan Distance*. Namun untuk nilai *Silhouette Coefficient* DBSCAN dengan *Manhattan Distance* menghasilkan nilai *Silhouette Coefficient* maksimum lebih besar dibandingkan dengan DBSCAN dengan *Euclidean Distance*. Meskipun demikian, hasil tersebut akan disandingkan dengan peta seismisitas hasil penelitian sebelumnya yang dijadikan *golden standart* dalam bidang gempabumi untuk mengetahui *clustering* DBSCAN mana yang mendekati *golden standart* yang telah ada.

BAB V

METODE K-MEANS

5.1 Desain

K-Means merupakan salah satu algoritma *clustering* non hirarki yang bekerja berdasarkan pemisahan data dan kriteria kesalahan kuadrat atau *squared error* (Wagstaff, 2012) yang berbasis titik pusat atau *centroid*. Peneliti menggunakan metode *K-Means* karena mempunyai kemampuan mengelompokkan data dalam jumlah cukup besar dengan waktu komputasi yang cepat dan efisien (Reviantika et al., 2020) untuk melakukan *clustering* data gempabumi di wilayah Regional VII sesuai dengan Alur Algoritma *K-Means* pada Gambar 3.6. Algoritma *K-Means* terdiri dari dua fase terpisah yaitu, fase pertama memilih k pusat secara acak, di mana nilai k ditetapkan terlebih dahulu.

Tahap selanjutnya, adalah membawa setiap objek data ke pusat terdekat atau *centroid*. Algoritma ini mengelompokkan semua objek data ke dalam beberapa *cluster*. Selanjutnya, rata-rata *cluster* awal yang telah terbentuk dihitung ulang, di mana proses ini berlanjut berulang kali hingga fungsi kriteria menjadi minimum atau pusat *cluster* tidak berubah lagi. Untuk menentukan jarak antara setiap objek data dengan pusat *cluster* menggunakan perhitungan *Euclidean Distance*. Jika vektor $x = (x_1, x_2, \dots, x_n)$ dan vektor $y = (y_1, y_2, \dots, y_n)$, maka *Euclidean distance* $d(x_i, y_i)$ dapat diperoleh sebagai berikut:

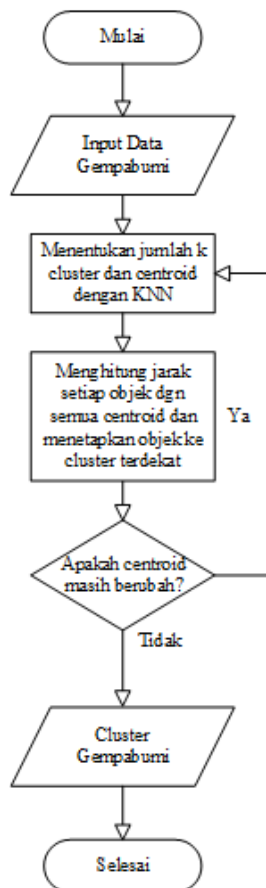
$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5.1)$$

di mana:

x_i : titik x ke i

y_i : titik y ke i

Untuk setiap cluster j ($1 \leq j \leq k$), hitung ulang pusat *cluster* (*centroid*) hingga tidak ada perubahan pada pusat *cluster*. Adapun proses clustering dengan algoritma *K-Means* dijelaskan pada Gambar 5.1 tentang diagram alir algoritma *K-Means* berikut ini.



Gambar 5.1 Diagram Alir Algoritma *K-Means*
Sumber: (Reviantika et al., 2020)

Gambar 5.1 menjelaskan tentang proses *clustering* gempabumi menggunakan algoritma *K-Means* mulai dari menentukan jumlah k *cluster* dengan

KNN dari data gempa bumi di wilayah Regional VII yang meliputi wilayah Provinsi Jawa Tengah, Yogyakarta dan Jawa Timur selama 5 tahun sebanyak 1.827 data mulai dari tahun 2017 hingga 2021 yang bersumber dari BMKG.

Implementasi perhitungannya penulis menggunakan Jupyter Notebook dengan bahasa pemrograman *Python*. Adapun langkah selengkapnya dijelaskan pada bagian selanjutnya berikut ini.

5.2 Uji Coba

Tahap selanjutnya adalah melakukan implementasi algoritma *K-Means* dengan menerapkan 2 metode perhitungan jarak untuk menentukan k optimal yaitu, metode *Euclidean Distance* dan *Manhattan Distance*. Adapun langkah-langkah 2 metode perhitungan jarak tersebut dijelaskan pada bagian berikut ini.

5.2.1 *Clustering K-Means Optimasi k dengan Euclidean Distance*

Clustering gempa bumi dibagi menjadi 2 bagian yaitu, *clustering* berdasarkan periode tahunan (T_{1_2017} , T_{1_2018} , T_{1_2019} , T_{1_2020} , T_{1_2021}) dan periode 5 tahun ($T_{1_2017-2021}$). Kemudian menentukan nilai k cluster dan pusat cluster atau *centroid* secara acak dan selanjutnya menghitung jarak antar cluster menggunakan persamaan matriks jarak *Euclidean Distance* untuk menentukan anggota cluster

Sehingga diperoleh nilai optimal k cluster untuk data gempa bumi tiap periode baik tahunan maupun lima tahunan adalah sebagai berikut:

Tabel 5.1 Nilai Optimal k cluster Berdasarkan Periode Waktu dengan *Euclidean Distance*

Periode	T _{1_2017}	T _{1_2018}	T _{1_2019}	T _{1_2020}	T _{1_2021}	T _{5_2017-2021}
Nilai k optimal	7	3	5	4	4	4

Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan nilai k optimal pada tabel 5.1 di atas diperoleh nilai k yang berbeda untuk tiap periode waktu.

5.2.2 Clustering K-Means Optimasi k dengan *Manhattan Distance*

Setelah dilakukan *clustering K-Means* dengan optimasi k menggunakan *Euclidean Distance*, selanjutnya melakukan perhitungan k optimal menggunakan *Manhattan Distance*. Jika *Euclidean Distance* merupakan akar jumlah selisih kuadrat antar titik atau koordinat maka *Manhattan Distance* merupakan jumlah nilai mutlak selisih jarak antara dua titik.

Pada penentuan nilai k optimal dengan *Manhattan Distance* mengacu pada nilai maksimum pusat *cluster* yang sesuai atau *maximum correct centroid value*. Sehingga diperoleh nilai optimal k cluster untuk data gempabumi tiap periode tahunan dan 5 tahun seperti dijelaskan pada Tabel 5.2.

Tabel 5.2 Nilai Optimal k cluster Berdasarkan Periode Waktu dengan *Manhattan Distance*

Periode	T _{1_2017}	T _{1_2018}	T _{1_2019}	T _{1_2020}	T _{1_2021}	T _{5_2017-2021}
Nilai k optimal	4	3	3	3	4	3

Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan Tabel 5.2 diperoleh nilai k optimal untuk *cluster* gempabumi tiap periode waktu menggunakan perhitungan jarak *Manhattan Distance* dengan rentang antara 3 dan 4.

5.3 Uji k Cluster

Setelah diperoleh nilai k optimal pada langkah sebelumnya, untuk mengetahui apakah nilai k cluster yang terbentuk merupakan cluster terbaik atau bukan maka perlu dilakukan pengujian cluster. Salah satu metode pengujian cluster pada algoritma K -Means adalah menggunakan *Maximum Silhouette Score* yang merupakan perbandingan antara selisih rata-rata jarak antara titik I ke seluruh titik selain cluster dengan rata-rata jarak antara titik I dengan seluruh titik di dalam satu cluster terhadap nilai maksimal rata-rata jarak keduanya. Sehingga diperoleh nilai *Silhouette Coefficient* untuk cluster tiap periode waktu baik untuk *Euclidean Distance* maupun *Manhattan Distance* adalah sebagai berikut:

Tabel 5.3 Nilai *Silhouette Coefficient* Berdasarkan Periode Waktu

Periode		T _{1_2017}	T _{1_2018}	T _{1_2019}	T _{1_2020}	T _{1_2021}	T _{5_2017-2021}
Silhouette Coefficient	Euclidean Distance	0,517	0,516	0,518	0,570	0,508	0,511
	Manhattan Distance	0,506	0,516	0,475	0,526	0,508	0,485

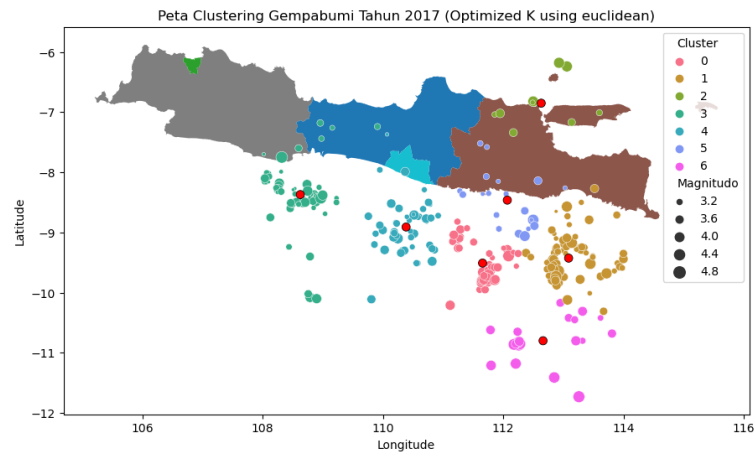
Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan nilai tabel 5.3 di atas diperoleh nilai *Silhouette Coefficient* dengan nilai terendah sebesar 0,475 untuk cluster gempabumi periode T_{1_2019} dengan perhitungan jarak menggunakan *Manhattan Distance* dan nilai terendah sebesar 0,570 untuk cluster gempabumi periode T_{1_2020} dengan perhitungan jarak menggunakan *Euclidean Distance*.

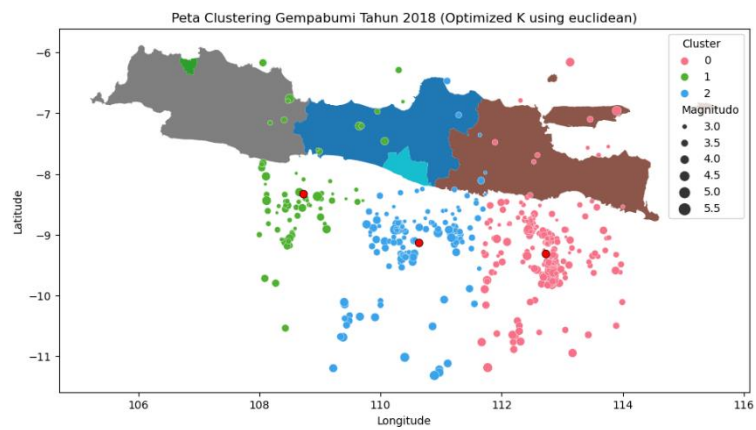
5.4 Hasil Clustering

Berdasarkan perhitungan nilai k cluster dan penentuan pusat cluster serta pengujian k cluster terbaik menggunakan *Maximum Silhouette Score* pada

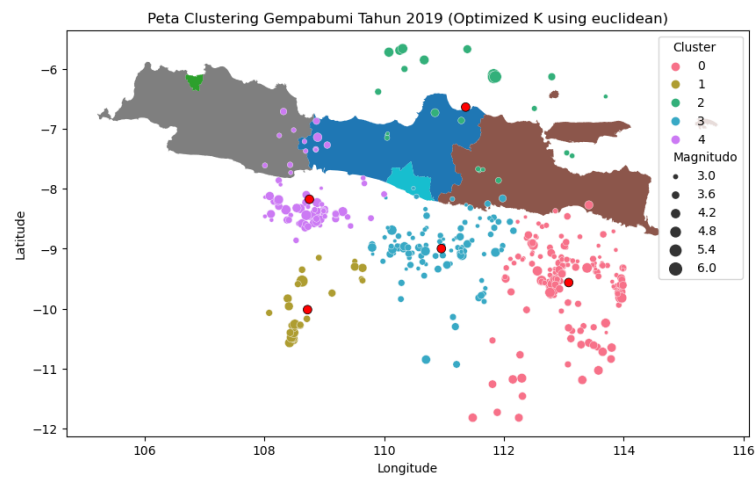
pembahasan sebelumnya, maka dapat diperoleh visualisasi data *cluster* gempa bumi tiap periode waktu sebagai berikut.



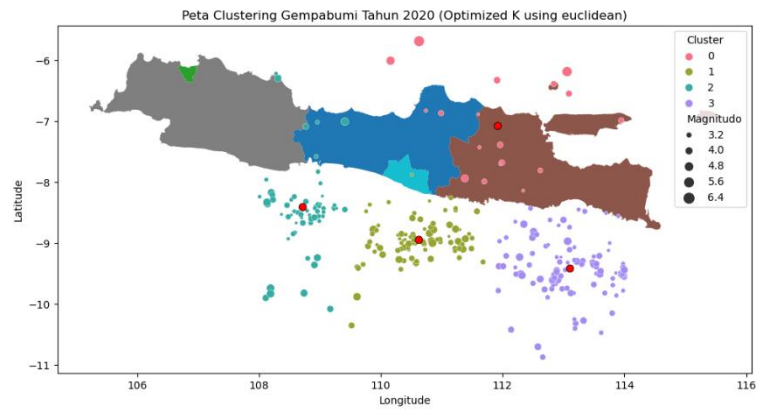
(a)



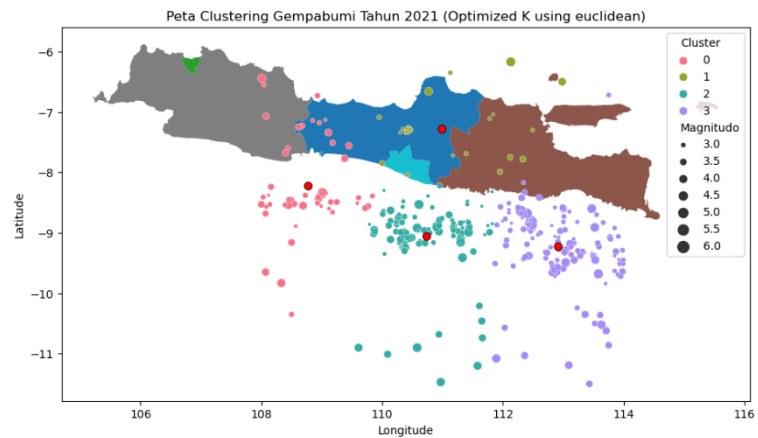
(b)



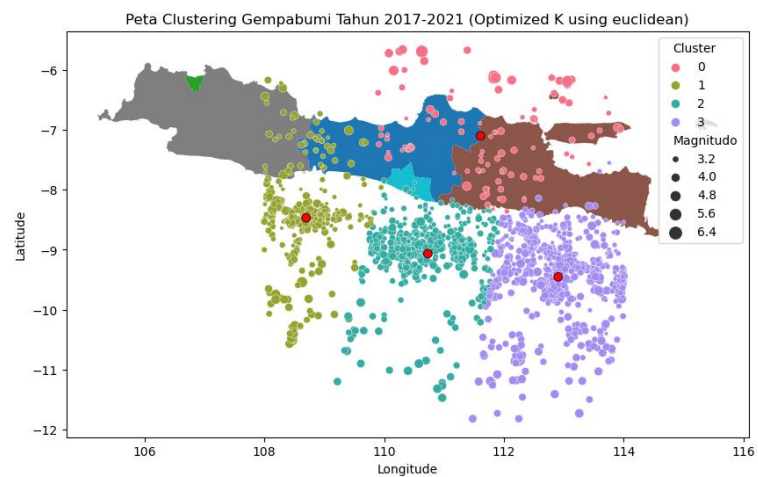
(c)



(d)



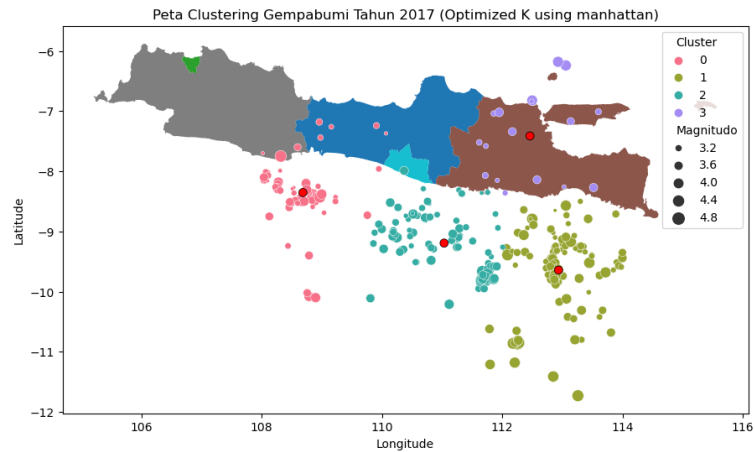
(e)



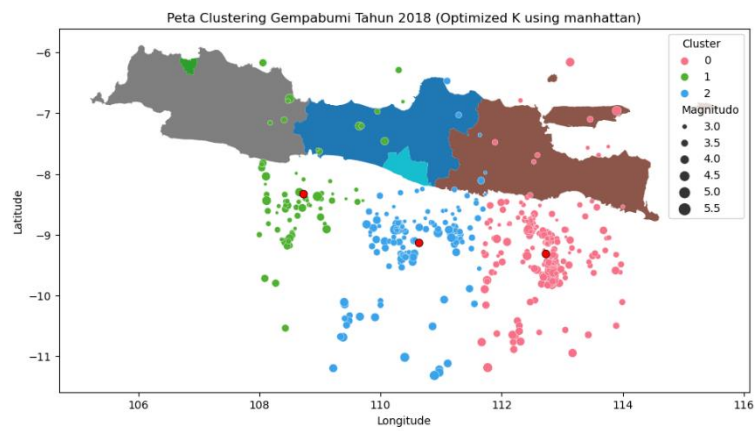
(f)

Gambar 5.2 Peta *Clustering* Gempabumi Tiap Periode dengan *K* Optimal menggunakan *Euclidean Distance*
 Sumber: Diolah penulis dari hasil penelitian (2023)

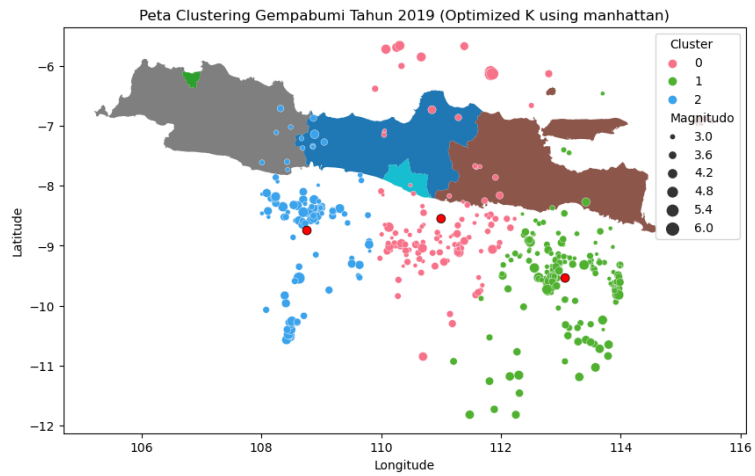
Berdasarkan Gambar 5.2 a, b, c, d, e dan f dapat dilihat bahwa hasil *clustering* gempabumi tiap periode waktu dengan perhitungan jarak menggunakan *Euclidean Distance* memiliki jumlah *cluster* yang terbentuk beragam yaitu, mulai dari 3 *cluster* hingga terbanyak 7 *cluster*.



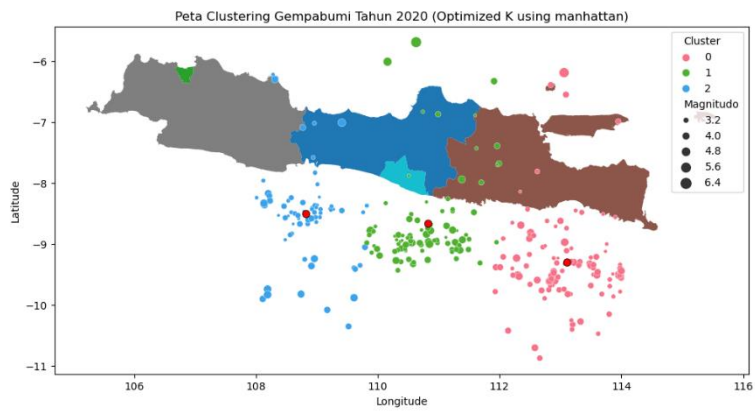
(a)



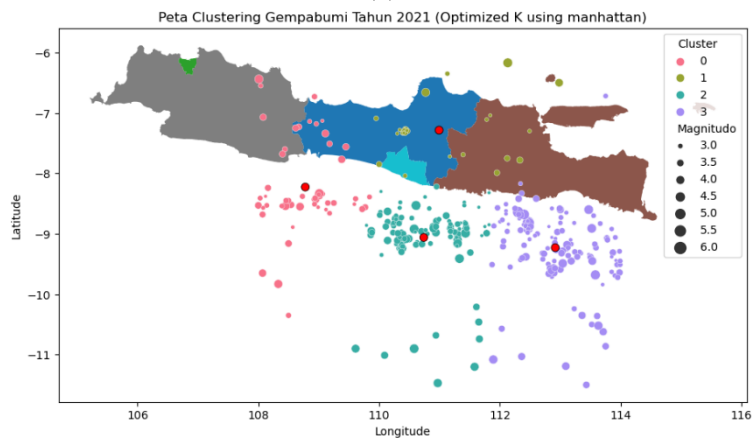
(b)



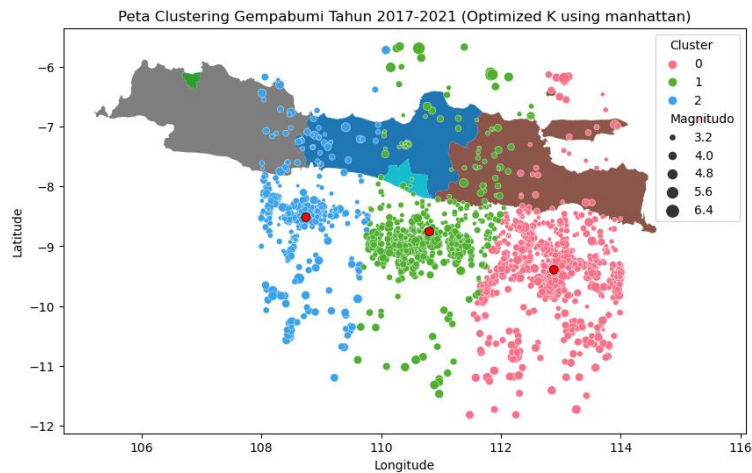
(c)



(d)



(e)



(f)

Gambar 5.3 Peta *Clustering* Gempabumi Tiap Periode dengan *K* Optimal menggunakan *Manhattan Distance*

Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan Gambar 5.3 a, b, c, d, e dan f dapat dilihat bahwa hasil *clustering* gempabumi tiap periode waktu dengan perhitungan jarak menggunakan *Manhattan Distance* memiliki jumlah *cluster* yang terbentuk adalah 3 dan 4 *cluster*.

5.5 Kesimpulan

Hasil analisis *clustering* menggunakan metode *K-Means* dengan perhitungan jarak menggunakan 2 metode yaitu, *Euclidean Distance* dan *Manhattan Distance* menghasilkan nilai *k* optimal dan *Silhouette Coefficient* yang berbeda. *K-Means* dengan *Euclidean Distance* menghasilkan nilai *k* optimal lebih yaitu, 3, 4, 5 dan 7 *cluster*. Sedangkan untuk *K-Means* dengan *Manhattan Distance* menghasilkan nilai *k* optimal sebanyak 3 dan 4 *cluster*. Sebaliknya untuk *K-Means* dengan *Euclidean Distance* menghasilkan nilai *Silhouette Coefficient* dengan selisih antara nilai tertinggi dan terendah sebesar 0,062 dengan nilai tertinggi sebesar 0,570 untuk *cluster* gempabumi dengan periode T_{1_2020} dan nilai terendah

sebesar 0,508 untuk *cluster* gempa bumi dengan periode T_{1_2021} . Sedangkan untuk *K-Means* dengan *Manhattan Distance* menghasilkan selisih nilai *Silhouette Coefficient* sebesar 0,041 dengan nilai tertinggi sebesar 0,516 untuk *cluster* gempa bumi dengan periode T_{1_2018} dan nilai terendah sebesar 0,475 untuk *cluster* gempa bumi dengan periode T_{1_2019} .

Hal ini menjelaskan bahwa *clustering K-Means* dengan *Euclidean Distance* lebih unggul dibanding dengan *clustering K-Means* dengan *Manhattan Distance* jika berdasarkan nilai *Silhouette Coefficient*nya. Namun, untuk mengetahui seberapa mirip hasil *clustering K-Means* dengan 2 metode perhitungan jarak tersebut dibandingkan dengan peta seismisitas yang menjadi *golden standart* yang telah ada, maka perlu dibandingkan hasil pemetaan dengan metode *clustering* tersebut.

BAB VI

PEMBAHASAN

6.1 Perbandingan Performa Algoritma

Pada bab-bab sebelumnya telah membahas tentang *clustering* gempa bumi menggunakan 2 algoritma yaitu algoritma DBSCAN dan *K-Means* dengan eksperimen masing-masing algoritma menggunakan 2 metode perhitungan jarak yaitu *Euclidean Distance* dan *Manhattan Distance*. Selanjutnya dilakukan perbandingan performa kedua algoritma tersebut berdasarkan jumlah *cluster* yang terbentuk, nilai *Silhouette Coefficient* dan kecocokan hasil *clustering* dengan *golden standart* yang telah ada yaitu peta sebaran episenter gempa bumi Pulau Jawa hasil relokasi *teletomoDD* dari katalog PuSGeN 2017.

Berdasarkan hasil implementasi algoritma DBSCAN dengan 2 eksperimen dengan menggunakan data gempa bumi selama rentang 5 tahun (2017-2021) sebanyak 1.827 data gempa bumi di wilayah Regional VII yang dibagi berdasarkan periode waktu kejadian yaitu periode tahunan (T_{1_2017} , T_{1_2018} , T_{1_2019} , T_{1_2020} , T_{1_2021}) dan periode 5 tahun ($T_{5_2017-2021}$) bahwa eksperimen pertama *clustering* gempa bumi DBSCAN dengan *Euclidean Distance* diperoleh hasil epsilon (ϵ) optimal untuk periode T_{1_2017} , T_{1_2018} , T_{1_2019} , T_{1_2020} , T_{1_2021} dan $T_{5_2017-2021}$ secara urut adalah 0,49, 0,39, 0,47, 0,49, 0,33 dan 0,19 dan MinPts optimal secara urut adalah 18, 20, 20, 14, 12 dan 18. Selanjutnya, dari hasil epsilon (ϵ) dan MinPts optimal tersebut diimplementasikan ke dalam *clustering* menghasilkan jumlah *cluster* gempa bumi untuk tiap periode T_{1_2017} , T_{1_2018} , T_{1_2019} , T_{1_2020} , T_{1_2021} dan

$T_{5_2017-2021}$ adalah 3 *cluster* dan 1 *noise* untuk periode T_{1_2017} , T_{1_2018} , dan T_{1_2020} , 4 *cluster* dan 1 *noise* untuk periode T_{1_2019} dan T_{1_2021} serta 6 *cluster* dan 1 *noise* untuk periode $T_{5_2017-2021}$. Berdasarkan hasil *clustering* tersebut jumlah *cluster* tertinggi adalah untuk periode $T_{5_2017-2021}$ dengan nilai epsilon (ϵ) dan MinPts optimal adalah 0,19 dan 18. Bila dibandingkan dengan nilai epsilon (ϵ) optimal periode lainnya periode tersebut memiliki nilai epsilon (ϵ) optimal terendah. Berdasarkan hasil uji *cluster* menggunakan *Silhouette Coefficient* diperoleh nilai tiap periode T_{1_2017} , T_{1_2018} , T_{1_2019} , T_{1_2020} , T_{1_2021} dan $T_{5_2017-2021}$ secara urut adalah 0,413, 0,376, 0,403, 0,486, 0,368 dan 0,208.

Pada eksperimen kedua *clustering* gempa bumi DBSCAN dengan *Manhattan Distance* diperoleh hasil epsilon (ϵ) optimal untuk periode T_{1_2017} , T_{1_2018} , T_{1_2019} , T_{1_2020} , T_{1_2021} dan $T_{5_2017-2021}$ secara urut adalah 0,51, 0,41, 0,47, 0,57, 0,33 dan 0,24 dan MinPts optimal secara urut adalah 20, 20, 20, 20, 12 dan 20. Selanjutnya, dari hasil epsilon (ϵ) dan MinPts optimal tersebut diimplementasikan ke dalam *clustering* menghasilkan jumlah *cluster* gempa bumi untuk tiap periode T_{1_2017} , T_{1_2018} , T_{1_2019} , T_{1_2020} , T_{1_2021} dan $T_{5_2017-2021}$ adalah 3 *cluster* dan 1 *noise* untuk periode T_{1_2017} , T_{1_2018} , dan T_{1_2020} , 4 *cluster* dan 1 *noise* untuk periode T_{1_2019} , T_{1_2021} , dan $T_{5_2017-2021}$. Berdasarkan hasil uji *cluster* menggunakan *Silhouette Coefficient* diperoleh nilai tiap periode T_{1_2017} , T_{1_2018} , T_{1_2019} , T_{1_2020} , T_{1_2021} dan $T_{5_2017-2021}$ secara urut adalah 0,417, 0,390, 0,403, 0,499, 0,368 dan 0,372. Dengan demikian dapat disimpulkan bahwa *clustering* DBSCAN dengan *Euclidean Distance* dengan nilai epsilon (ϵ) optimal dan *Silhouette Coefficient* kecil dapat menghasilkan jumlah *cluster* yang lebih banyak

dibandingkan dengan nilai epsilon (ϵ) optimal yang lebih besar. Sedangkan *clustering* DBSCAN dengan *Manhattan Distance* diperoleh jumlah *cluster* hampir seragam yaitu 3 dan 4 *cluster* dengan 1 *noise*.

Clustering gempabumi telah diimplementasikan menggunakan DBSCAN dengan 2 eksperimen perhitungan jarak *Euclidean* dan *Manhattan Distance* telah dilakukan dan diperoleh hasil seperti pada penjelasan di atas. Selanjutnya dilakukan *clustering* menggunakan *K-Means* dengan 2 eksperimen yaitu *K-Means* dengan *Euclidean Distance* dan *K-Means* dengan *Manhattan Distance*. Eksperimen pertama *clustering* menggunakan *K-Means* dengan *Euclidean Distance* diperoleh hasil *k* optimal sebanyak 7 *cluster* untuk periode T_{1_2017} , 3 *cluster* untuk periode T_{1_2018} , 5 *cluster* untuk periode T_{1_2019} , dan 4 *cluster* untuk periode T_{1_2020} , T_{1_2021} dan $T_{5_2017-2021}$. Berdasarkan hasil tersebut kemudian dilakukan uji *cluster* menggunakan *Maximum Silhouette Score* sehingga diperoleh nilai untuk T_{1_2017} , T_{1_2018} , T_{1_2019} , T_{1_2020} , T_{1_2021} dan $T_{5_2017-2021}$ secara urut adalah 0,517, 0,516, 0,518, 0,570, 0,508 dan 0,511.

Perbandingan hasil *clustering* 2 algoritma DBSCAN dengan 2 metode perhitungan jarak yaitu *Euclidean Distance* dan *Manhattan Distance* dapat dilihat pada tabel berikut ini.

Tabel 6.1 Perbandingan hasil *clustering* DBSCAN *Euclidean* dan *Manhattan Distance*

Periode	Epsilon Opt		MinPts Opt		Jumlah Cluster		Silhouette Coef	
	Euclid.	Manhatt.	Euclid.	Manhatt.	Euclid.	Manhatt.	Euclid.	Manhatt.
T_{1_2017}	0,49	0,51	18	20	3	3	0,413	0,417
T_{1_2018}	0,39	0,41	20	20	3	3	0,376	0,390
T_{1_2019}	0,47	0,47	20	20	4	4	0,403	0,403
T_{1_2020}	0,49	0,57	14	20	3	3	0,486	0,499
T_{1_2021}	0,33	0,33	12	12	4	4	0,368	0,368
$T_{5_2017-2021}$	0,19	0,24	18	20	6	4	0,208	0,372

Sumber: Diolah penulis dari hasil penelitian (2023)

Pada eksperimen kedua *clustering* gempabumi menggunakan *K-Means* dengan *Manhattan Distance* diperoleh hasil *k* optimal sebanyak 4 *cluster* untuk periode T_{1_2017} dan T_{1_2021} dan 3 *cluster* untuk periode T_{1_2018} , T_{1_2019} , T_{1_2020} dan $T_{5_2017-2021}$. Setelah dilakukan uji *cluster* menggunakan *Maximum Silhouette Score* sehingga diperoleh nilai untuk T_{1_2017} , T_{1_2018} , T_{1_2019} , T_{1_2020} , T_{1_2021} dan $T_{5_2017-2021}$ secara urut adalah 0,506, 0,516, 0,475, 0,526, 0,508 dan 0,485. Dengan demikian berdasarkan hasil 2 eksperimen *clustering* menggunakan *K-Means* dengan *Euclidean* dan *Manhattan Distance* dapat disimpulkan bahwa *clustering K-Means* dengan *Euclidean Distance* menghasilkan jumlah *cluster* lebih beragam mulai dari 3, 4, 5 *cluster* dengan *cluster* terbanyak adalah 7 *cluster* dan apabila dibandingkan dengan *clustering K-Means* dengan *Manhattan Distance* hanya menghasilkan jumlah *cluster* sebanyak 3 dan 4 *cluster*. Berdasarkan hasil uji *cluster* menggunakan *Maximum Silhouette Score clustering* gempabumi menggunakan *K-Means* dengan *Euclidean Distance* diperoleh nilai *Silhouette Score* lebih baik untuk nilai tertinggi yaitu sebesar 0,570 untuk *cluster* periode T_{1_2020} bila dibandingkan dengan nilai *Silhouette Score* tertinggi untuk *clustering* gempabumi menggunakan *K-Means* dengan *Manhattan Distance* yaitu sebesar 0,516 untuk *cluster* periode T_{1_2018} .

Perbandingan hasil *clustering* 2 algoritma *K-Means* dengan 2 metode perhitungan jarak yaitu *Euclidean Distance* dan *Manhattan Distance* dapat dilihat pada Tabel 6.2.

Tabel 6.2 Perbandingan hasil *clustering K-Means Euclidean* dan *Manhattan Distance*

Periode	Jumlah k Cluster Opt		Silhouette Coef	
	Euclidean Dist.	Manhattan Dist	Euclidean Dist.	Manhattan Dist
T_{1_2017}	7	4	0,517	0,506
T_{1_2018}	3	3	0,516	0,516
T_{1_2019}	5	3	0,518	0,475

Periode	Jumlah k Cluster Opt		Silhouette Coef	
	Euclidean Dist.	Manhattan Dist	Euclidean Dist.	Manhattan Dist
T ₁ _2020	4	3	0,570	0,526
T ₁ _2021	4	4	0,508	0,508
T ₅ _2017-2021	4	3	0,511	0,485

Sumber: Diolah penulis dari hasil penelitian (2023)

Secara umum hasil *clustering* gempabumi menggunakan 2 metode yaitu DBSCAN dan *K-Means* diperoleh hasil yang berbeda, hal ini karena metode DBSCAN melakukan pengelompokan data berdasarkan kepadatan, data dengan kepadatan tinggi dikelompokkan menjadi 1 *cluster* dan sebaliknya untuk data dengan kepadatan rendah dianggap *noise*. Sehingga untuk data yang berada di area kepadatan tinggi atau menyebar tidak termasuk ke dalam *cluster*. Berbeda dengan *clustering* menggunakan metode *K-Means* yang melakukan pengelompokan data berdasarkan jumlah pusat *cluster* atau *centroid* sesuai jumlah *k* optimal yang terbentuk. Sehingga data yang memiliki jarak terdekat dengan pusat *cluster* maka termasuk ke dalam *cluster* tersebut.

Adapun perbedaan hasil *clustering* 2 metode DBSCAN dan *K-Means* dapat dilihat pada Tabel 6.3.

Tabel 6.3 Perbandingan hasil *clustering* DBSCAN dan *K-Means* Euclidean dan Manhattan Distance

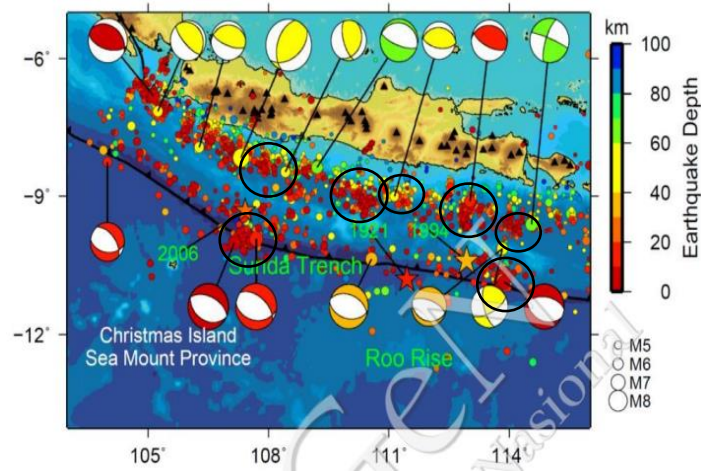
Periode	Jumlah Cluster			
	DBSCAN Euclid.	DBSCAN Manhatt.	K-Means Euclid.	K-Means Manhatt.
T ₁ _2017	3	3	7	4
T ₁ _2018	3	3	3	3
T ₁ _2019	4	4	5	3
T ₁ _2020	3	3	4	3
T ₁ _2021	4	4	4	4
T ₅ _2017-2021	6	4	4	3

Sumber: Diolah penulis dari hasil penelitian (2023)

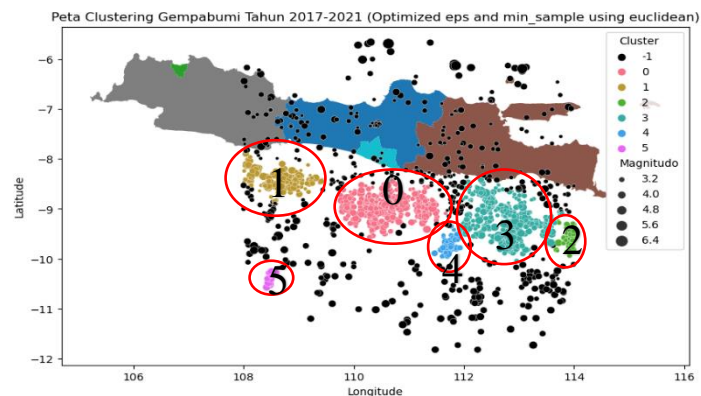
Berdasarkan Tabel 6.3 diperoleh jumlah *cluster* yang berbeda dari hasil *clustering* menggunakan 2 metode yang berbeda. Dapat dilihat hasil *clustering*

dengan perhitungan jarak menggunakan *Euclidean Distance* menghasilkan jumlah *cluster* yang beragam dengan jumlah *cluster* tertinggi mendekati *golden standart*, yaitu DBSCAN menghasilkan *cluster* tertinggi sebanyak 6 *cluster* dan *K-Means* menghasilkan *cluster* tertinggi sebanyak 7 *cluster*. Perbedaannya terdapat pada periode data gempa bumi yang digunakan. Jika DBSCAN *cluster* maksimal diperoleh dari data gempa bumi untuk periode T_{5_2017-2021} sedangkan *K-Means cluster* maksimal diperoleh dari data gempa bumi untuk periode T_{5_2017}. Artinya, semakin banyak data yang digunakan dalam *clustering* metode DBSCAN menghasilkan *cluster* lebih baik karena algoritma DBSCAN membentuk *cluster* berdasarkan kepadatan data. Sebaliknya, semakin banyak data yang digunakan dalam *clustering* dengan metode *K-Means* menghasilkan *cluster* lebih sedikit karena algoritma *K-Means* membentuk *cluster* berdasarkan *centroid* atau pusat *cluster*.

Setelah diperoleh nilai *Silhouette Coefficient* untuk *clustering* tiap periode waktu dari 2 metode dengan 2 eksperimen tersebut di atas, selanjutnya dilakukan visualisasi hasil *clustering* ke dalam peta seperti yang telah dilakukan pada Bab IV dan Bab V. Hasilnya kemudian akan disandingkan dengan peta sebaran episenter gempa bumi Pulau Jawa hasil relokasi *teletomoDD* dari katalog PuSGeN 2017 yang menjadi *golden standart* pada penelitian ini. Adapun perbandingan antara hasil *clustering* dengan *golden standart* dapat dilihat pada Gambar 6.1 dan Gambar 6.2.



Gambar 6.1 Peta sebaran episenter gempa bumi Pulau Jawa hasil relokasi *teletomoDD* katalog PuSGeN (2017)
 Sumber: (Pustlitbang PUPR, 2017)



Gambar 6.2 Hasil *clustering* gempabumi DBSCAN dgn *Euclid. Distance* $T_{5,2017-2021}$
 Sumber: Diolah penulis dari hasil penelitian (2023)

Hasil *clustering* gempabumi untuk data periode $T_{1,2017-2021}$ menggunakan metode DBSCAN dengan perhitungan jarak menggunakan *Euclidean Distance* memiliki hasil yang paling mendekati dengan *golden standart* yang ada. *Cluster* yang terbentuk berada di area zona pertemuan lempeng Indo-Australia dan Eurasia atau zona *megathrust* yang berada di selatan Pulau Jawa tepatnya di selatan wilayah

Regional VII. Hal ini membuktikan bahwa pada zona tersebut memiliki potensi gempabumi yang tinggi dibandingkan dengan wilayah daratan di Regional VII.

Jumlah anggota *cluster* dapat memberikan informasi mengenai potensi kegempaan di wilayah tersebut, dengan asumsi semakin banyak jumlah gempabumi dalam *cluster* tersebut maka semakin tinggi potensi gempabumi di wilayah tersebut. Adapun hasil *clustering* DBSCAN dengan *Euclidean Distance* untuk periode T_{1_2017-2021} tiap wilayah di Regional VII disajikan dalam Tabel 6.4.

Tabel 6.4 Hasil *clustering* DBSCAN *Euclid. Dist* T_{1_2017-2021}
Per Wilayah di Regional VII

Nama Wilayah	Nama Cluster						Total
	0	1	2	3	4	5	
Jawa Tengah	41	214	-	-	-	20	275
DIY	228	-	-	-	-	-	228
Jawa Timur	134	-	77	564	46	-	821

Sumber: Diolah penulis dari hasil penelitian (2023)

Berdasarkan Tabel 6.4 Provinsi Jawa Timur merupakan wilayah yang memiliki potensi gempabumi tertinggi di antara wilayah Provinsi Jawa Tengah dan Yogyakarta. Hal ini ditunjukkan dengan jumlah *cluster* gempabumi yang terbentuk sebanyak 4 *cluster* dengan total kejadian gempabumi sebanyak 821.

Selain itu, secara umum peta hasil *clustering* gempabumi menggunakan DBSCAN dengan perhitungan jarak menggunakan *Euclidean Distance* Gambar 6.2 hasilnya mempunyai kemiripan dengan peta sebaran episenter gempabumi Pulau Jawa hasil relokasi *teletomoDD* dari katalog PuSGeN tahun 2017. Hal ini menyatakan bahwa *clustering* menggunakan algoritma DBSCAN dengan perhitungan jarak menggunakan *Euclidean Distance* menunjukkan hasil yang lebih baik daripada *clustering K-Means*.

6.2 Clustering Gempabumi dalam Pandangan Al-Qur'an

Berdasarkan hasil pada pembahasan di atas bahwa *cluster* gempabumi hasil pemodelan memiliki kemiripan dengan *golden standart* yang telah ada, artinya bahwa gempabumi akan mengelompok pada suatu lokasi yang menjadi sumber gempabumi tersebut. Sehingga dengan *clustering* gempabumi kita dapat mengetahui zona potensi gempabumi yang dapat digunakan sebagai salah satu upaya manusia untuk meminimalkan dampak akibat bencana gempabumi. Hal ini termasuk ikhtiar manusia untuk mengubah suatu keadaan yang bisa menimbulkan kerusakan. Adapun ayat Al-Qur'an yang menjelaskan tentang hal tersebut adalah seperti di dalam QS. Ar-Ra'd ayat 11 Allah berfirman:

لَهُ مُعَقِّبَاتٌ مِّنْ بَيْنِ يَدَيْهِ وَمِنْ خَلْفِهِ يَحْفَظُونَهُ مِنْ أَمْرِ اللَّهِ ۗ إِنَّ اللَّهَ لَا يُغَيِّرُ مَا
ۚ بِقَوْمٍ حَتَّىٰ يُغَيِّرُوا مَا بِأَنفُسِهِمْ ۗ وَإِذَا أَرَادَ اللَّهُ بِقَوْمٍ سُوءًا فَلَا مَرَدَّ لَهُ
وَمَا لَهُمْ مِّنْ دُونِهِ مِّنْ وَّالٍ

Artinya:

“Untuk setiap orang ada malaikta berurutan di belakangnya: mereka menjaganya dengan perintah Allah. Sesungguhnya Allah tidak akan merubah keadaan suatu kaum sampai mereka merubahnya sendiri. Tetapi ketika Allah menghendaki azab pada suatu kaum, tidak ada yang dapat menolaknya dan tidak akan mereka temukan selain Dia untuk berlindung” (QS. Ar-Ra'd[13:11]).

Ayat tersebut menjelaskan bahwa setiap manusia diikuti oleh beberapa malaikat yang bertugas mencatat amal perbuatan baik dan amal perbuatan buruk. Oleh karena itu, kita harus senantiasa ingat kepada Allah dalam setiap melakukan suatu perbuatan. Ayat tersebut juga menjelaskan Allah tidak akan merubah keadaan pada manusia melainkan mereka sendiri yang melakukan perubahan tersebut. Namun, jika Allah berkehendak lain semua usaha yang telah dilakukan oleh

manusia tidak akan bisa mengalahkan kehendak-Nya. Jika dikaitkan dengan *clustering* gempa bumi artinya, bahwa manusia diperbolehkan melakukan usaha untuk mencegah dari keburukan yaitu melakukan *clustering* gempa bumi untuk mengetahui wilayah yang memiliki potensi gempa bumi. Sehingga untuk wilayah dengan potensi gempa bumi tinggi dapat dihindari untuk dibangun pemukiman atau jika ingin dibangun, perlu diperhatikan upaya-upaya mitigasinya seperti membangun jalur dan tempat evakuasi, memperkuat struktur bangunan, melakukan pelatihan dalam menghadapi bencana dan sebagainya.

BAB VII

KESIMPULAN

7.1 Kesimpulan

Berdasarkan hasil yang diperoleh pada pembahasan di atas maka dapat diambil kesimpulan sebagai berikut:

- a. Berdasarkan hasil uji *cluster* menggunakan *Silhouette Coefficient* terbaik, *clustering* gempa bumi menggunakan *K-Means* dengan *Euclidean Distance* memiliki performa terbaik jika dibandingkan dengan metode *clustering K-Means* dengan *Manhattan Distance* dan DBSCAN, dengan nilai *Silhouette Coefficient* di atas 0,5 dengan nilai tertinggi sebesar 0,570 untuk data periode T_{1_2020}. Namun setelah diuji dengan kemiripan peta sebaran episenter gempa bumi Pulau Jawa hasil relokasi dari katalog PuSGeN 2017 yang menjadi *golden standart* dari penelitian ini diperoleh hasil yang paling mendekati adalah hasil *clustering* gempa bumi DBSCAN dengan *Euclidean Distance* untuk periode T_{5_2017-2021} dengan distribusi *cluster* di sekitar sumber gempa bumi zona subduksi atau *megathrust* sebanyak 6 *cluster*.
- b. Berdasarkan jumlah *cluster* gempa bumi dan total kejadian gempa bumi di wilayah Regional VII, wilayah Provinsi Jawa Timur memiliki potensi gempa bumi tertinggi dibandingkan wilayah Provinsi Jawa Tengah dan Yogyakarta dengan jumlah *cluster* sebanyak 4 *cluster* yaitu *cluster* 0, 2, 3 dan 4 dengan jumlah total kejadian gempa bumi sebanyak 821 kejadian. Hal ini menunjukkan bahwa wilayah Jawa Timur merupakan wilayah dengan potensi

gempabumi tinggi. Sehingga perlu diperkuat upaya-upaya mitigasi menghadapi bencana gempabumi tersebut di antaranya dengan melakukan perencanaan pembangunan meliputi sarana dan prasarana yang matang khususnya bagi pemerintah setempat melalui dinas-dinas terkait seperti dinas PUPR dan dinas Pariwisata maupun sektor swasta.

7.2 Saran

Hasil dari penelitian yang telah dilakukan oleh penulis perlu dilakukan pengembangan di antaranya adalah:

1. Perlu dilakukan *clustering* gempabumi dengan metode lainnya seperti *K-Medoids*, *Fuzzy Clustering*, *Agolomerative Hierrachical Clustering*, *Self Organizing Map (SOM)*, atau gabungan dari kedua metode di atas K-DBSCAN untuk mendapatkan hasil yang dapat digunakan sebagai pembandingan.
2. Perlu ditambah dataset yang lebih banyak dengan periode waktu berbeda agar dapat terlihat perbedaan hasil *clustering* dengan periode waktu lebih panjang dan dataset yang lebih banyak.

DAFTAR PUSTAKA

- Algorithm, D. (2021). *Journal of Physics and Its Applications Clustering of Seismicity in the Indonesian Region for the 2018-2020 Period using the.* 4(1), 1–6.
- Ansari, A., Noorzad, A., & Zafarani, H. (2009). Clustering analysis of the seismic catalog of Iran. *Computers and Geosciences*, 35(3), 475–486. <https://doi.org/10.1016/j.cageo.2008.01.010>
- Avip, B., Martadiputra, P., Rachmatin, D., & Hidayat, A. S. (2020). *Analysis of Characteristics of Earthquake Area in Indonesia in 2020 with Cluster Analysis as Natural Disaster.* 9(11), 1243–1250. <https://doi.org/10.21275/SR201122121148>
- Cesca, S. (2020). Seiscloud, a tool for density-based seismicity clustering and visualization. *Journal of Seismology*, 24(3), 443–457. <https://doi.org/10.1007/s10950-020-09921-8>
- Cui, M. (2020). *on the Elbow Method.* 5–8. <https://doi.org/10.23977/accaf.2020.010102>
- Daszykowski, M., & Walczak, B. (2009). Density-Based Clustering Methods. *Comprehensive Chemometrics*, 2, 635–654. <https://doi.org/10.1016/B978-044452701-1.00067-3>
- Fan, Z., & Xu, X. (2019). Application and visualization of typical clustering algorithms in seismic data analysis. *Procedia Computer Science*, 151(2018), 171–178. <https://doi.org/10.1016/j.procs.2019.04.026>
- Gao, X., & Li, G. (2020). A KNN Model Based on Manhattan Distance to Identify the SNARE Proteins. *IEEE Access*, 8, 112922–112931. <https://doi.org/10.1109/ACCESS.2020.3003086>
- Georgoulas, G., Konstantaras, A., Katsifarakis, E., Stylios, C. D., Maravelakis, E., & Vachtsevanos, G. J. (2013). “Seismic-mass” density-based algorithm for spatio-temporal clustering. *Expert Systems with Applications*, 40(10), 4183–4189. <https://doi.org/10.1016/j.eswa.2013.01.028>
- Harini, S., Fahmi, H., Mulyanto, A. D., & Khudzaifah, M. (2020). The earthquake events and impacts mapping in Bali and Nusa Tenggara using a clustering method. *IOP Conference Series: Earth and Environmental Science*, 456(1). <https://doi.org/10.1088/1755-1315/456/1/012087>
- Jufriansah, A., Pramudya, Y., Khusnani, A., & Saputra, S. (2021). Analysis of

Earthquake Activity in Indonesia by Clustering Method. *Journal of Physics: Theories and Applications*, 5(2), 92.
<https://doi.org/10.20961/jphys theor-appl.v5i2.59133>

Karmenova, M., Tlebaldinova, A., Krak, I., Denissova, N., Popova, G., Zhantassova, Z., Ponkina, E., & Györök, G. (2022). An Approach for Clustering of Seismic Events using Unsupervised Machine Learning. *Acta Polytechnica Hungarica*, 19(5), 7–22.
<https://doi.org/10.12700/APH.19.5.2022.5.1>

Karri, N. A., Yousuf Ansari, M., & Pathak, A. (2019). Identification of seismic zones of India using DBSCAN. *2018 International Conference on Computing, Power and Communication Technologies, GUCON 2018*, 65–69. <https://doi.org/10.1109/GUCON.2018.8674964>

Kazemi-Beydokhti, M., Ali Abbaspour, R., & Mojarab, M. (2017). Spatio-Temporal Modeling of Seismic Provinces of Iran Using DBSCAN Algorithm. *Pure and Applied Geophysics*, 174(5), 1937–1952.
<https://doi.org/10.1007/s00024-017-1507-0>

Kementerian Agama RI. (2012). *Lajnah Pentashihan Mushaf Alquran*.
<https://quran.kemenag.go.id/surah/99>

Mojarab, M., Memarian, H., Zare, M., Hossein Morshedy, A., & Hossein Pishahang, M. (2014). Modeling of the seismotectonic provinces of Iran using the self-organizing map algorithm. *Computers and Geosciences*, 67, 150–162. <https://doi.org/10.1016/j.cageo.2013.12.007>

Novianti, P., Setyorini, D., & Rafflesia, U. (2017). K-means cluster analysis in earthquake epicenter clustering. *International Journal of Advances in Intelligent Informatics*, 3(2), 81–89.
<https://doi.org/10.26555/ijain.v3i2.100>

Omran, M. G. H., Engelbrecht, A. P., & Salman, A. (2007). An overview of clustering methods. *Intelligent Data Analysis*, 11(6), 583–605.
<https://doi.org/10.3233/ida-2007-11602>

Pakuani, K. W., & Kurniawan, R. (2021). Kajian Penentuan Nilai Epsilon Optimal Pada Algoritma DMDBSCAN Dan Pemetaan Daerah Rawan Gempa Bumi Di Indonesia Tahun 2014-2020. *Seminar Nasional Official Statistics*, 2021(1), 991–1000.
<https://doi.org/10.34123/semnasoffstat.v2021i1.847>

Pustlitbang PUPR. (2017). *Buku Peta Gempa 2017*.

Rahman, R. R. A., & Wijayanto, A. W. (2021). Pengelompokan Data Gempa Bumi

Menggunakan Algoritma DbSCAN. *Jurnal Meteorologi Dan Geofisika*, 22(1), 31. <https://doi.org/10.31172/jmg.v22i1.738>

Reviantika, F., Harahap, C. N., & Azhar, Y. (2020). Analisis Gempa Bumi Pada Pulau Jawa Menggunakan Clustering Algoritma K-Means. *Jurnal Dinamika Informatika*, 9(1), 51–60.

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN Revisited, Revisited. *ACM Transactions on Database Systems*, 42(3), 1–21. <https://doi.org/10.1145/3068335>

Shi, N., Liu, X., & Guan, Y. (2010). Research on k-means clustering algorithm: An improved k-means clustering algorithm. *3rd International Symposium on Intelligent Information Technology and Security Informatics, IITSI 2010*, 63–67. <https://doi.org/10.1109/IITSI.2010.74>

Supendi, P., Triyono, R., Permana, D., & Karnawati, D. (2021). *The 10 April 2021 (Mw 6 . 1) South of East Java Destructive Earthquake , Indonesia. 2021(2)*.

Wagstaff, K. L. (2012). Data Clustering. *Advances in Machine Learning and Data Mining for Astronomy*, 31(3), 543–561. <https://doi.org/10.1201/b11822-19>

Yedla, Madhu, Srinivasa Rao Pathakota, T. M. S. (2010). Enhancing K-means Clustering Algorithm with Improved Initial Center. *International Journal of Computer Science and Information Technologies*, 1(2), 121–125.

Zhang, H., & Thurber, C. (2006). Development and applications of double-difference seismic tomography. *Pure and Applied Geophysics*, 163(2–3), 373–403. <https://doi.org/10.1007/s00024-005-0021-y>

Zhihong, L., & Jie, Z. (2019). Online Knowledge Community Governance Based on Blockchain Token Incentives. In *Communications in Computer and Information Science: Vol. 1103 CCIS*. https://doi.org/10.1007/978-981-15-1209-4_5