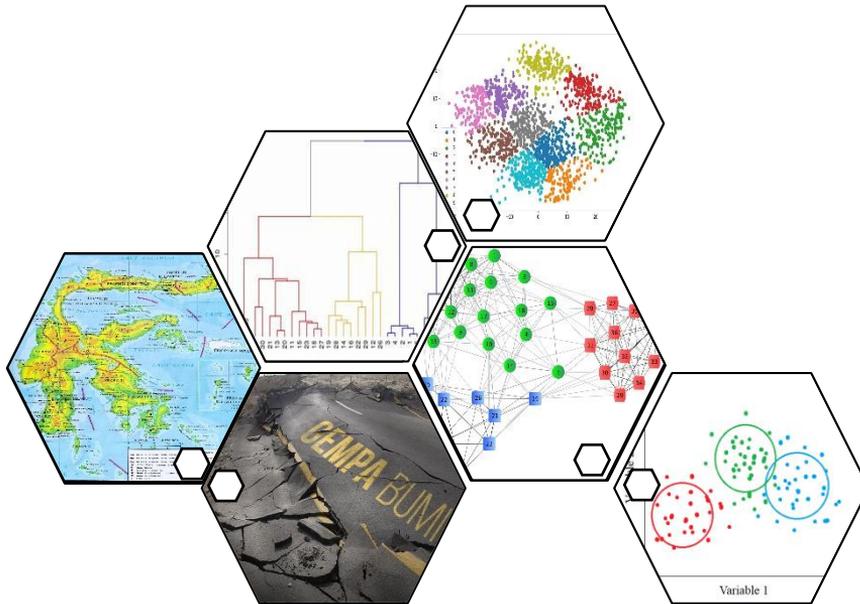


**ALGORITMA *K*-NEAREST NEIGHBOR SPASIO TEMPORAL DENSITY  
BASED CLUSTERING KASUS GEMPA BUMI DI PULAU SULAWESI**

**SPATIAL TEMPORAL DENSITY *K*-NEAREST NEIGHBOUR BASED  
CLUSTERING ALGORITHM FOR EARTHQUAKE CASES IN SULAWESI  
ISLAND**



**MAHRANI**

**H062221013**



**PROGRAM STUDI MAGISTER STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2024**

**ALGORITMA *K-NEAREST NEIGHBOR SPASIO TEMPORAL DENSITY*  
BASED CLUSTERING KASUS GEMPA BUMI DI PULAU SULAWESI**

**MAHRANI**

**H062221013**



**PROGRAM STUDI MAGISTER STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2024**

**ALGORITMA *K*-NEAREST NEIGHBOR SPASIO TEMPORAL DENSITY  
BASED CLUSTERING KASUS GEMPA BUMI DI PULAU SULAWESI**

Tesis  
sebagai salah satu syarat untuk mencapai gelar magister

Program Studi Magister Statistika

Disusun dan diajukan oleh

MAHRANI  
H062221013

Kepada

**PROGRAM STUDI MAGISTER STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2024**

**TESIS**

**ALGORITMA K-NEAREST NEIGHBOR SPASIO TEMPORAL DENSITY  
BASED KLASTERING KASUS GEMPA BUMI DI PULAU SULAWESI**

**MAHRANI  
H062221013**

telah dipertahankan di depan Panitia Ujian Magister pada 15 Agustus 2024  
dan dinyatakan telah memenuhi syarat kelulusan

pada

Program Studi Magister Statistika  
Departemen Statistika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Hasanuddin  
Makassar

Mengesahkan:

Pembimbing Utama

Pembimbing Pendamping



Dr. Anna Isjamiyati, S.Si., M.Si  
NIP. 197708082005012002



Prof. Dr. Dr. Geordina Maria Tinungki M.Si.  
NIP. 196209261987022001



Dr. Enia N. Hirdiani, S.Si., M.Si.  
NIP. 19750429 200003 2 001



Dr. Eng. Amiruddin, M.Si.  
NIP. 19720515 199702 1 002

**PERNYATAAN KEASLIAN TESIS  
DAN PELIMPAHAN HAK CIPTA**

Dengan ini saya menyatakan bahwa, tesis berjudul " *Algoritma K-Nearest Neighbor Spasio Temporal Density Based Klustering Kasus Gempa Bumi Di Pulau Sulawesi*" adalah benar karya saya dengan arahan dari tim pembimbing (Dr. Anna Islamiyati, S.Si.,M.Si. dan Prof Dr. Dr Georgina Maria Tinungki M.Si.). Karya Ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apapun kepada perguruan tinggi manapun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari tesis ini akan dipublikasikan di Vita E Pensiero sebagai artikel dengan judul "*A Method Density Based Spatial Clustering Of Application With Noise On Tuberculosis Cases In Indonesia*".

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 15 Agustus 2024



MAHRANI

NIM. H062221013

## UCAPAN TERIMA KASIH

*Bismillahirrohmanirrohim,*

Segala puji hanya milik Allah *Subhanallahu Wa Ta'ala* atas limpahan rahmat dan hidayah-Nya kepada penulis. Shalawat dan salam tercurahkan kepada Rasulullah *Shallallahu 'Alaihi Wa sallam*, keluarganya, *tabi'in, tabi'ut tabi'in*, serta orang-orang sholeh yang haq hingga kadar Allah berlaku atas diri mereka. *Alhamdulillahirobbil'aalamiin*, berkat rahmat dan kemudahan dari Allah *Subhanallahu Wa Ta'ala*, penulis dapat menyelesaikan tesis berjudul " *Algoritma K-Nearest Neighbor Spasio Temporal Density Based Clustering* Kasus Gempa Bumi Di Pulau Sulawesi " sebagai salah satu syarat memperoleh gelar magister pada Program Studi Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Terima Kasih yang tak terhingga kepada keempat orang tuaku tercinta Ayah (**Sudiarso & Syamsuddin**) dan Ibu (**Badaria & Hasna Gali**) yang selalu mendengarkan keluh kesahku, memberikan kasih sayang tak terhingga, materi, semangat, motivasi dan doa yang tak pernah putus, juga kepada saudara-saudaraku, kakak (**Subiyanto, Zulfikar, Zulqaedah, Zulhijrah, Sukmawati, Sriwahyuni, Sumarni, Mardatillah dan Fahrul**) dan adikku (**Fausan, Firda, Fatir dan Mahendra**) yang selalu memberikan dukungan, semangat dan dalam penyelesaian tesis ini . Ucapan rasa hormat dan juga terima kasih yang tulus kepada:

1. Yth. **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc** selaku Rektor Universitas Hasanuddin
2. Yth. **Dr. Eng. Amiruddin, M.Si** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam beserta seluruhnya jajarannya.
3. Yth. **Dr. Anna Islamiyati, S.Si., M.Si** selaku Ketua Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam dan sekaligus pembimbing utama yang senantiasa meluangkan waktu, tenaga, pemikiran dalam membimbing dan mengarahkan penulis dalam menyelesaikan tesis.
4. Yth. **Dr. Erna Tri Hardiani, S.Si., M.Si** selaku Ketua Program Studi Magister Statistika dan sekaligus sebagai penguji yang telah bersedia menguji serta memberikan masukan-masukan dan arahan dalam penyusunan tesis
5. Yth. **Prof. Dr. Dr Georgina Maria Tinungki, M.Si** sebagai pembimbing yang senantiasa meluangkan waktu, tenaga, pemikiran dalam membimbing dan mengarahkan penulis dalam menyelesaikan tesis.
6. Yth. **Prof.Dr. Nurtiti Sunusi, S.Si., M.Si** selaku penguji yang telah bersedia menguji serta memberikan masukan-masukan dan arahan dalam penyusunan tesis
7. Yth. **Dr. Nirwan, M.Si** selaku penguji yang telah bersedia menguji serta memberikan masukan-masukan dan arahan dalam penyusunan tesis
8. Sahabat terbaik penulis yang bernama **Rachmad** yang telah dengan tulus membantu dan mendukung saya untuk terus berjuang menyelesaikan tesis

ini. Kehadiranmu menjadi sumber kekuatan yang luar biasa bagi saya,  
Terima Kasih Terkasih

9. Sahabatku tercinta **Nurfadiah Mutmainnah, Ivaldi A. Prawinata, Nursalfa UI Fahira, dan Sriwahyuni Astuti Ahmad** yang memberi semangat bertubi-tubi bagi penulis, si pemberi pelangi setelah hujan.
10. Teman-teman seperjuangan **Nur Ikhwana, Irwan Usman, Taufiqurrahman, Khairunnisa Abdullah dan Aqilah Salsabilah.**
11. Teman-teman seperjuangan ruang diskusi yang selalu memberikan semangat dan bantuan
12. Semua pihak yang telah membantu penulis yang tidak bisa disebutkan satu per satu terimakasih atas doa serta dukungannya

Semoga Tuhan Yang Maha Esa memberikan balasan yang berlipat ganda, terimakasih dan hikmat-Nya atas segala kebaikan yang telah diberikan kepada penulis. Penulis menyadari bahwa masih banyak kekurangan dalam tesis ini, untuk itu dengan segala kerendahan hati, penulis memohon maaf.

Makassar, 15 Agustus 2024

Mahrani

## ABSTRAK

MAHRANI. **Algoritma *K-Nearest Neighbor Spasio Temporal Density Based Clustering* Kasus Gempa Bumi Di Pulau Sulawesi** (dibimbing oleh Dr. Anna Islamiyati, S.Si.,M.Si. dan Prof Dr. Dr Georgina Maria Tinungki M.Si.)

**Latar Belakang.** Pulau Sulawesi merupakan salah satu kawasan seismik aktif di Indonesia yang mengakibatkan wilayah ini tergolong rawan terjadi gempa bumi. Pulau Sulawesi memiliki aktivitas kegempaan yang tinggi dikarenakan lokasinya terletak pada zona benturan ketiga lempeng tektonik dunia yaitu lempeng Indo-Australia, lempeng Eurasia dan lempeng Pasifik. Ketiga lempeng ini memiliki pertemuan yang konvergen dan bertumbukan secara relative sehingga menyebabkan intensitas gempa yang tinggi sehingga perlu dilakukan kluster untuk melihat wilayah yang rentan terjadi gempa. Salah satu metode *clustering* dalam kasus kepadatan yang dapat mendeteksi *outlier* yaitu DBSCAN yang dikembangkan menjadi metode ST-DBSCAN. Dalam melakukan *clustering* dengan metode ST-DBSCAN penentuan parameter sangatlah krusial karena mempengaruhi hasil kluster yang terbentuk, sehingga metode ini akan di optimasi menggunakan algoritma *K-Nearest Neighbor*. **Tujuan.** Penelitian ini bertujuan untuk melakukan clustering serta menganalisis pola spasio temporal pada data kejadian gempa bumi di Pulau Sulawesi tahun 2021-2023. **Metode.** Adapun metode yang digunakan pada penelitian ini adalah Algoritma *K-Nearest Neighbor Spasio Temporal Density Based Clustering*. **Hasil.** Hasil penelitian menunjukkan bahwa Algoritma *K-Nearest Neighbor Spasio Temporal Density Based Clustering* memiliki nilai parameter terbaik yang didapatkan yaitu  $\varepsilon_1=10$ ,  $\varepsilon_2=7$  dan MinPts=3 mendapatkan nilai dengan nilai *silhouette coefficient* =0.440, DBI = 1,303, *calinski harabaz* = 2.155,773 dan *average score* = 0,952 menghasilkan 12 kluster, 6 diantaranya kluster besar meliputi kluster ke-1 dengan 74 jumlah titik gempa, kluster ke-2 dengan 55 jumlah titik gempa, kluster ke-6 dengan 93 jumlah titik gempa, kluster ke-7 dengan 42 jumlah titik gempa, kluster ke-9 dengan 30 jumlah titik gempa dan kluster ke-11 dengan 53 jumlah titik gempa **Kesimpulan.** Dari hasil penelitian, penggunaan Algoritma *K-Nearest Neighbor Spasio Temporal Density Based Clustering* mampu melakukan clustering serta mendeteksi adanya *outlier*.

Kata Kunci: DBSCAN; ST-DBSCAN; K-Nearest Neighbor; *clustering*; *outlier*

## ABSTRACT

MAHRANI. **Algoritma K-Nearest Neighbor Spatio Temporal Density Based Clustering of Earthquake Cases in Sulawesi Island** (supervised by Dr. Anna Islamiyati, S.Si., M.Si. and Prof. Dr. Georgina Maria Tinungki M.Si.).

**Background.** Sulawesi Island is one of the seismically active regions in Indonesia, which makes this region prone to earthquakes. Sulawesi Island has high seismic activity due to its location in the collision zone of the three world tectonic plates, namely the Indo-Australian plate, Eurasian plate and Pacific plate. These three plates have converged and collided relatively, causing high earthquake intensity, so it is necessary to cluster to see areas that are prone to earthquakes. One of the clustering methods in the case of density that can detect outliers is DBSCAN which was developed into the ST-DBSCAN method. In clustering with the ST-DBSCAN method, the determination of parameters is crucial because it affects the results of the clusters formed, so this method will be optimised using the K-Nearest Neighbour algorithm.

**Objective.** This research aims to cluster and analyse spatio-temporal patterns in earthquake occurrence data on Sulawesi Island for 2021-2023. **Method.** The method used in this research is the K-Nearest Neighbour Spatio Temporal Density Based Clustering Algorithm. **Results.** The results showed that the K-Nearest Neighbour Spatio Temporal Density Based Clustering Algorithm has the best parameter values obtained, namely  $\epsilon_1=10$ ,  $\epsilon_2=7$  and  $\text{MinPts}=3$  getting a value with a silhouette coefficient = 0.440,  $\text{DBI} = 1.303$ ,  $\text{Calinski Harabaz} = 2.155.773$  and average score = 0.952 resulting in 12 clusters, 6 of which are large clusters including the 1st cluster with 74 earthquake points, the 2nd cluster with 55 earthquake points, the 6th cluster with 93 earthquake points, the 7th cluster with 42 earthquake points, the 9th cluster with 30 earthquake points and the 11th cluster with 53 earthquake points.

**Conclusion.** From the research results, the use of the K-Nearest Neighbour Spatio Temporal Density Based Clustering Algorithm is able to perform clustering and detect outliers.

Keywords: DBSCAN; ST-DBSCAN; K-Nearest Neighbour; clustering; outlier

## DAFTAR ISI

	<b>Halaman</b>
HALAMAN JUDUL.....	i
PERNYATAAN PENGAJUAN .....	ii
HALAMAN PENGESAHAN .....	<b>Error! Bookmark not defined.</b>
PERNYATAAN KEASLIAN TESIS .....	<b>Error! Bookmark not defined.</b>
UCAPAN TERIMA KASIH .....	v
ABSTRAK .....	vii
ABSTRACT .....	viii
DAFTAR ISI .....	x
DAFTAR TABEL .....	xii
DAFTAR GAMBAR .....	xiii
DAFTAR LAMPIRAN .....	xiv
BAB I .....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah .....	2
1.4 Tujuan 3	
1.5 Manfaat .....	3
1.6 Kajian Teori .....	3
BAB II .....	16
METODE PENELITIAN .....	16
2.1 Sumber Data .....	16
2.2 Variabel Penelitian.....	16
2.3 Metode Analisis .....	16

BAB III .....	18
HASIL DAN PEMBAHASAN.....	18
3.1 Analisis Deskriptif .....	18
3.2 Deteksi Pola Persebaran Titik .....	22
3.3 Clustering Gempa Bumi di Pulau Sulawesi Menggunakan Algoritma K-Nearest Neighbor ST-DBSCAN.....	23
3.4 Jarak Euclidean .....	24
3.5 Optimasi Penentuan Nilai Eps dan MinPts .....	25
3.6 Pemilihan Hasil Kluster Terbaik .....	30
3.7 Hasil Kluster Terbaik.....	31
3.8 Analisis Pola Spasio Temporal.....	36
BAB IV .....	42
PENUTUP .....	42
4.1 Kesimpulan .....	42
4.2 Saran .....	42
DAFTAR PUSTAKA .....	43
LAMPIRAN.....	45

## DAFTAR TABEL

Nomor urut	Halaman
1. Interpretasi Nilai Silhouette Coefficient .....	10
2. Penggolongan Kedalaman Gempa .....	13
3. Penggolongan Magnitudo Gempa .....	13
4. Variabel Penelitian .....	16
5. Kekuatan Gempa Bumi .....	19
6. Data Magnitude Terbesar .....	20
7. <i>Depth</i> Gempa Bumi .....	21
8. <i>Depth</i> Terdangkal .....	22
9. Jarak Euclidean .....	25
10. Nilai $\epsilon_1$ yang Diteliti .....	27
11. Hasil Klaster 112 Pasangan Parameter Eps1 , Eps2, dan MinPts .....	27
12. Parameter Terbaik dari Tiap MinPts .....	30
13. Klaster Gempa Bumi .....	31
14. Karakteristik Gempa Bumi Masing-Masing Klaster .....	33
15. Klaster Besar .....	36
16. Wilayah Persebaran Gempa Bumi Tiap Periode Pada Klaster 6 .....	40

**DAFTAR GAMBAR**

Nomor urut	Halaman
1. Kerangka Pikir.....	15
2. Persebaran Titik Gempa Bumi di Pulau Sulawesi.....	18
3. Kekuatan Gempa .....	19
4. Kedalaman Gempa .....	21
5. Grafik KNN Tiap K .....	26
6. Plot Silhouette Coefficient Setiap MinPts.....	28
7. Plot DBI Setiap MinPts.....	29
8. Plot Calinski Harabaz Setiap MinPts.....	29
9. Persebaran Klaster Gempa Bumi Hasil Eksplorasi Data .....	35
10. Pola Titik Gempa Klaster ke-6 periode 1 hingga 27 .....	39

**DAFTAR LAMPIRAN**

Nomor urut	Halaman
1. Data Gempa Bumi di Pulau Sulawesi Tahun 2021-2023 .....	45
2. Jarak Euclidean Spasial.....	46
3. Jarak Euclidean Temporal .....	47
4. Jarak Rata- Rata Terdekat Observasi .....	48
5. Parameter Hasil Simulasi.....	49

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi mengakibatkan melimpahnya sumber data. Di era digital, data telah banyak tersedia dari berbagai media. Data yang tersedia memiliki karakteristik yang berbeda sehingga pengolahan data yang dilakukan juga menggunakan metode yang berbeda. Tipe data yang sering ditemukan dalam data mining adalah data spatio-temporal. Spatio temporal berasal dari kata spasial dan temporal. Data spasial merupakan data mengenai koordinat geografis (latitude dan longitude) sedangkan data temporal merupakan data yang memuat kejadian sementara waktu seperti tahun, bulan, menit dan detik. Data spasio temporal (ST) dapat dibagi menjadi lima kategori yaitu peristiwa, referensi geografis variabel, referensi geografis deret waktu, titik bergerak, dan lintasan (Dahmouni et al., 2018). Pengolahan data spatio-temporal menjadi informasi dapat dilakukan dengan menggunakan metode clustering. Clustering merupakan salah satu metode data mining yang digunakan untuk mengetahui struktur alami dataset. Secara prinsip kluster adalah kumpulan dari objek data yang mempunyai kemiripan berdasarkan karakteristik tertentu kemudian melakukan pengelompokkan jika dianggap mirip (Huang et al., 2023).

Permasalahan yang muncul ketika melakukan analisis clustering adalah adanya bentuk, ukuran, kepadatan kluster yang tidak beraturan serta mengandung *outlier* (Liu et al., 2012). Permasalahan ini dapat diatasi dengan menggunakan analisis clustering berbasis kepadatan. Salah satu metode yang tepat digunakan adalah DBSCAN (*Density Based Spatial Clustering Applications With Noise*). Algoritma DBSCAN dapat mendeteksi outlier atau noise dan menghasilkan kluster lebih akurat serta tidak perlu menentukan jumlah kluster di awal, algoritma ini lebih efektif dan lebih baik dalam menentukan parameter dari pada algoritma lainnya, juga dapat mengenali bentuk kluster yang sulit tidak beraturan (Gaonkar & Sawant, 2013).

Perluasan algoritma dari DBSCAN disebut dengan ST-DBSCAN (*Spatio Temporal- Density Based Spatial Clustering Applications With Noise*) yang dapat mengkalukulasi kepadatan berdasarkan jarak dan waktu (Andraud et al., 2021). Algoritma ST-DBSCAN memiliki dua parameter jarak yaitu epsilon1 ( $\epsilon_1$ ) dan epsilon2 ( $\epsilon_2$ ) untuk menentukan kesamaan dengan konjungsi dari dua tes kepadatan.  $\epsilon_1$  digunakan untuk nilai spasial dalam mengukur kedekatan dua titik secara geografis.  $\epsilon_2$  digunakan untuk mengukur kesamaan nilai non-spasial (Birant & Kut, 2007). ST-DBSCAN memiliki kelebihan lebih fleksibel dengan ukuran data yang besar karena merupakan teknik *clustering* yang berbasis kepadatan. Algoritma ST-DBSCAN termasuk algoritma dalam *unsupervised learning* sehingga tidak memerlukan asumsi dalam pengerjaannya dan merupakan salah satu metode yang paling baik untuk menentukan kluster dari database spasial yang besar (Faraouk et al., 2023).

Parameter dalam algoritma DBSCAN mempunyai peran yang signifikan untuk menentukan jumlah klasternya. Pemilihan parameter perlu dilakukan secara tepat

agar metode digunakan secara optimal sehingga perlu mengoptimasi parameter menggunakan metode tertentu. Optimasi parameter epsilon ( $\epsilon$ ) sebagai parameter ditentukan menggunakan K-Nearest Neighbor dengan minPts sebagai k (Sharma & Sharma, 2017). Metode tersebut akan diterapkan pada kasus kepadatan yang erat kaitannya dengan data spasial dan data temporal seperti fenomena alam yaitu kejadian gempa bumi.

Gempa bumi merupakan bencana alam yang cukup serius di Indonesia dikarenakan tidak dapat diperkirakan kapan datangnya, dimana tempatnya, seberapa dalam gempanya dan seberapa kuat getarannya begitupun di pulau Sulawesi. Pulau Sulawesi merupakan salah satu kawasan seismik aktif di Indonesia yang mengakibatkan wilayah ini tergolong rawan terjadi gempa bumi. Pulau Sulawesi memiliki aktivitas kegempaan yang tinggi dikarenakan lokasinya terletak pada zona benturan ketiga lempeng tektonik dunia yaitu lempeng Indo-Australia, lempeng Eurasia dan lempeng Pasifik. Ketiga lempeng ini memiliki pertemuan yang konvergen dan bertumbukan secara relative (Ihda et al., 2015) . Badan Meterologi, Klimatologi dan Geofisika (BMKG) menerbitkan kaleidoskop gempa bumi berkekuatan  $\geq 4$  SR pada tahun 2021- 2023 berjumlah 448 kejadian dengan jumlah kejadian yang bertambah setiap tahunnya di Pulau Sulawesi. Hal tersebut disebabkan kondisi geografis di Pulau Sulawesi.

Penelitian sebelumnya telah menggunakan optimasi K-Nearest Neighbor pada metode DBSCAN dengan nilai parameter yang optimal dan menghilangkan unsur subjektivitas dalam penelitian (Sharma & Sharma, 2017). Selanjutnya peneliti akan menggunakan *K- Nearest Neighbor* dalam mengoptimasi parameter algoritma Spasio Temporal Density Based Clustering sebagai pengembangan dari metode DBSCAN. Berdasarkan uraian tersebut, peneliti ingin mengambil judul “Algoritma K-Nearest Neighbors Spasio Temporal Density Based Clustering Kasus Gempa Bumi di Pulau Sulawesi”

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana algoritma K-Nearest Neighbors dalam optimasi parameter pada Spasio Temporal Density Based Clustering kasus gempa bumi di Pulau Sulawesi?
2. Bagaimana pola clustering kasus gempa bumi di Pulau Sulawesi menggunakan algoritma K-Nearest Neighbors Spasio Temporal Density Based Clustering?

## 1.3 Batasan Masalah

Berdasarkan permasalahan yang telah diuraikan, batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan yaitu data spasial temporal ( waktu, latitude dan longitude) dan nonspasial (kedalaman dan kekuatan gempa)
2. Menggunakan data gempa tektonik Tahun 2021-2023 dengan kekuatan  $\geq 4$  SR

## 1.4 Tujuan

Berdasarkan rumusan masalah yang telah diuraikan, maka tujuan penelitian ini adalah:

1. Menentukan clustering menggunakan optimasi parameter K-Nearest Neighbors Spasio Temporal Density Based Clustering pada data gempa bumi di Pulau Sulawesi.
2. Menentukan pola clustering kasus gempa bumi di Pulau Sulawesi menggunakan algoritma K-Nearest Neighbors Spasio Temporal Density Based Clustering

## 1.5 Manfaat

Manfaat yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Memberikan pengetahuan dan wawasan dalam bidang statistika khususnya yang berhubungan dengan metode optimasi parameter K-Nearest Neighbors Spasio Temporal Density Based Clustering
2. Memberikan informasi bagi Pemerintah di wilayah Pulau Sulawesi serta dinas terkait sehingga dapat dijadikan sebagai bahan acuan dalam pengambilan setiap kebijakan

## 1.6 Kajian Teori

### 1.6.1. Data Mining

Data mining adalah proses menggunakan teknik statistic, matematika dan kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database yang besar (Handoko & Lesmana, 2018). Data mining memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (artificial intelligent), machine learning, statistic dan basis data. Beberapa teknik yang sering disebut dalam literatur data mining antara lain: *clustering*, *classification*, *association rule mining*, *neural network*, dan *genetic algorithm* (Manamperi & Manamperi, 2023)

### 1.6.2. Clustering

Clustering adalah suatu proses membagi sekumpulan data menjadi beberapa kelompok atau klaster. Semua objek yang berada di satu *klaster* memiliki kemiripan (*similar*) antara satu dengan lainnya dan berbeda (*dissimilar*) dengan objek *klaster* lain (Dahmouni et al., 2018). *Clustering* disebut juga *unsupervised learning* atau tanpa arahan. Clustering menggunakan pendekatan observasi untuk melakukan proses pengelompokan sehingga tidak ditentukan oleh label-label kelas yang diberikan. Proses pengelompokan tersebut tidak dilakukan oleh manusia, melainkan dengan menggunakan *clustering algorithm*, teknik ini merupakan salah satu algoritma di dalam *machine learning* yang paling umum dipergunakan pada data mining (Handoko & Lesmana, 2018). Karena itu, clustering dapat digunakan untuk mencari tahu kelompok yang sebelumnya tidak diketahui di data, sebagai fungsi data mining, analisis klaster dapat digunakan sebagai sarana independent (standalone tool) untuk mendapatkan pengetahuan dari distribusi data, meneliti berbagai karakteristik setiap klaster, dan memfokuskan pada sekelompok untuk analisis lebih lanjut.

Clustering dapat digunakan sebagai langkah preprocessing untuk algoritma lain, seperti penentuan label untuk proses klasifikasi (Latifi-Pakdehi & Daneshpour, 2021). Dalam data mining ada dua jenis metode *clustering* yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering* (Huang et al., 2023). Metode *hierarchical clustering* merupakan metode *clustering* dimana jumlah kelompok yang digunakan diketahui pada akhir proses clustering, sedangkan metode *non-hierarchical clustering* merupakan kebalikfan dari metode sebelumnya dimana jumlah kelompok yang digunakan ditentukan pada awal proses clustering.

### 1.6.3. Dissimilarity Matrix Jarak Euclidean

Kemiripan atau ketidakmiripan antara satu objek dengan objek lainnya digambarkan oleh suatu ukuran jarak. Apabila ukuran jarak antara dua objek itu dekat maka dapat dianggap memiliki karakteristik yang mirip. Sedangkan apabila jaraknya jauh, maka kedua objek tersebut dianggap memiliki karakteristik yang berbeda (Trisminingsih & Shaztika, 2017). Dissimilarity Matrix Jarak Euclidean adalah sebuah matriks jarak yang digunakan untuk mengukur jarak antara dua vektor dengan cara menghitung akar kuadrat dari jumlah selisih kuadrat antara keduanya. Matriks perbedaan ini mencatat kedekatan atau jarak untuk semua pasangan dari  $n$  objek. Jarak Euclidean merupakan jarak antara dua objek data  $(i, j)$  yang memiliki  $n$  atribut bernilai numerik, yang dinyatakan sebagai  $i = X_{i1}, X_{i2}, \dots, X_{in}$  dan  $j = X_{j1}, X_{j2}, \dots, X_{jn}$  pada ruang dimensi  $n$  ( $R^n$ ).

$$d(i, j) = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_n - Y_n)^2} \quad (1)$$

Dikarenakan hanya melibatkan /2 dimensi , maka persamaan (1) menjadi persamaan berikut:

$$d_s(i, j) = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \quad (2)$$

Keterangan,

$i$  dan  $j$  = Objek dua dimensi

$n$  = Banyaknya objek

$X$  = *Longitude* data ke-  $i$ ,  $i = 1, 2, \dots, n$

$Y$  = *Latitude* data ke-  $i$ ,  $i = 1, 2, \dots, n$

Aspek temporal persamaan jarak Euclidean dimodifikasi menjadi persamaan:

$$d_t(i, j) = |X_{tanggal\ i} - X_{tanggal\ j}| \quad (3)$$

Keterangan,

$X_{tanggal}$  = Objek satu dimensi

Dikarenakan data yang digunakan adalah *longitude* ( $^\circ$ ) dan *latitude* ( $^\circ$ ), maka  $1^\circ = 111,322$  KM

### 1.6.4. Nearest- Neighbor Analysis

*Nearest- neighbor analysis* merupakan metode yang dirancang untuk melihat pola dalam sebuah data titik dalam dua atau tiga dimensi. Metode ini melibatkan

perhitungan jarak titik rata-rata antara semua titik dengan tetangga terdekatnya (Ebdon, 1985). *Nearest-neighbor index* dinyatakan sebagai rasio jarak observasi yang dibagi dengan jarak ekspektasi. Jika  $R$  dinyatakan sebagai *nearest-neighbor index*, berikut merupakan persamaannya:

$$R = \frac{\bar{d}_{obs}}{\bar{d}_{ran}} \quad (4)$$

dengan  $\bar{d}_{obs}$  adalah *observed mean nearest-neighbor* atau jarak rata-rata observasi antara setiap titik dengan tetangga, dengan persamaan:

$$\bar{d}_{obs} = \frac{\sum_{i=1}^n d_i}{n} \quad (5)$$

dan  $\bar{d}_{ran}$  adalah *expected mean nearest-neighbor* atau jarak rata-rata ekspektasi untuk titik-titik yang diberikan dalam pola acak, dengan persamaan:

$$\bar{d}_{ran} = \frac{1}{2\sqrt{p}} = \frac{1}{2\sqrt{\frac{n}{A}}} = \frac{1}{2}\sqrt{\frac{A}{n}} \quad (6)$$

dengan,  $d_i$  merupakan jarak antara  $i$  dengan tetangga terdekatnya,  $n$  merupakan banyak titik, dan  $A$  adalah area persegi panjang minimum di sekitar semua titik, atau nilai area yang ditentukan. Jika  $SE_{\bar{d}}$  merupakan *standard error of the mean nearest-neighbor* dan  $c$  merupakan test statistics, maka berikut merupakan persamaan untuk test statistics  $c$ :

$$c(Z_{score}) = \frac{\bar{d}_{obs} - \bar{d}_{ran}}{SE_{\bar{d}}} \quad (7)$$

dengan

$$SE_{\bar{d}} = \frac{k}{\sqrt{np}} = \frac{k}{\sqrt{\frac{n^2}{A}}} \quad (8)$$

Distribusi sampling dari test statistics  $c$  adalah distribusi normal, Dengan kata lain  $c$  merupakan deviasi standar normal (Ebdon, 1985). Adapun hipotesis pada test *statistics*  $c(Z_{score})$  ditunjukkan sebagai berikut (Boots & Getis, 1998).

a. Hipotesis

$H_0$  : Data berdistribusi secara acak (*random*)

$H_1$  : Data berdistribusi secara mengelompok (klaster)

b. Tingkat signifikansi

$\alpha = 5\%$

c. Statistik uji

$$c(Z_{score}) = \frac{\bar{d}_{obs} - \bar{d}_{ran}}{SE_{\bar{d}}}$$

- d. Daerah kritis  
Tolak  $H_0$  jika nilai dari  $[Z_{score}] \geq Z_{tabel}$
- e. Keputusan  
Dikarenakan nilai dari  $[Z_{score}] \geq 1,96 = Z_{tabel}$ , maka tolak  $H_0$
- f. Kesimpulan  
Dengan tingkat kepercayaan sebesar 95% data yang ada mendukung untuk tolak  $H_0$

### 1.6.5. Density- Based Spatial Clustering Algorithm With Noise (DBSCAN)

*Density- Based Spatial Clustering Algorithm with Noise* (DBSCAN) adalah metode clustering yang menggunakan kepadatan data sebagai dasar pengelompokan untuk menemukan noise pada sebuah data spasial. DBSCAN digunakan untuk mengelompokkan data numerik. Algoritma ini menggunakan dua parameter utama yaitu epsilon ( $\epsilon$ ) yang menentukan radius dari suatu klaster dan MinPts yang merupakan jumlah minimum titik yang harus berada dalam radius  $\epsilon$  agar suatu titik dianggap sebagai *core point* (Creţulescu et al., 2019). Berikut adalah beberapa definisi dan rumus matematika terkait dengan DBSCAN:

#### 1. Core Point (titik inti)

Sebuah titik  $p$  diklasifikasikan sebagai core point jika jumlah dalam jarak  $\epsilon$  dari titik  $p$  setidaknya sama dengan MinPts dengan persamaan:

$$\text{Core}(p) = \{p \in D \mid |N_{\epsilon}(p)| \geq \text{MinPts}\} \quad (9)$$

Dimana,  $|N_{\epsilon}(p)|$  adalah himpunan tetangga  $p$  dalam jarak  $\epsilon$ , dan MinPts adalah jumlah minimum tetangga yang dibutuhkan untuk *core point*.

#### 2. Border point (titik batas)

Sebuah titik  $p$  diklasifikasikan sebagai border point jika jumlah tetangga dalam jarak  $\epsilon$  lebih sedikit dari MinPts, namun tetap memiliki setidaknya satu tetangga dalam jarak  $\epsilon$  yang merupakan titik core dengan persamaan:

$$\text{Border}(p) = \{p \in D \mid |N_{\epsilon}(p)| \geq \text{MinPts} \wedge \exists q \in N_{\epsilon}(p)\} \quad (10)$$

Dimana,  $|N_{\epsilon}(p)|$  adalah himpunan tetangga  $p$  dalam jarak  $\epsilon$ , dan MinPts adalah jumlah minimum tetangga yang dibutuhkan untuk core point.

3. *Epsilon* merupakan parameter jarak.  $\epsilon$  secara umum dicari menggunakan KNN dengan  $k$  yang digunakan adalah minpts

4. *Directly density-reachable* adalah observasi  $q$  berhubungan langsung dengan  $p$ , jika  $p$  adalah core point dan  $q$  merupakan tetangga dari  $p$  dalam jangkauan epsilon.

5. *Density-reachable* adalah observasi  $q$  dan  $x$  dalam satu klaster, namun  $x$  bukan tetangga dari  $q$  yang berada dalam jangkauan epsilon

6. *Noise* merupakan sebuah titik  $p$  dianggap sebagai noise jika tidak ada tetangga dalam jarak  $\epsilon$  dari  $p$  yang memenuhi kriteria sebagai titik core, dengan kata lain tidak

ada tetangga dalam jarak  $\varepsilon$  yang memiliki jumlah tetangga lebih dari atau sama dengan MinPts. Secara matematis sebuah titik  $p$  dikategorikan noise jika:

$$|N_\varepsilon(p)| < \text{MinPts} \quad (11)$$

dengan  $|N_\varepsilon(p)|$  adalah himpunan tetangga  $p$  dalam jarak  $\varepsilon$ , dan MinPts adalah jumlah minimum tetangga yang dibutuhkan untuk core point. Jika jumlah tetangga dalam jarak  $\varepsilon$  dari  $p$  kurang dari MinPts, maka titik  $p$  dianggap sebagai noise dalam algoritma DBSCAN.

### 1.6.6. Spatial- Temporal Density-Based Spatial Clustering Of Application With Noise (ST-DBSCAN)

ST- DBSCAN merupakan singkatan dari Spatial-Temporal Density Based Spatial Clustering of Application with Noise, istilah ini merupakan bagian dari algoritma yang mampu melakukan pengolahan data spasial dan temporal. Metode Algoritma ST-DBSCAN termasuk salah satu metode *clustering* yang mampu mengolah data temporal dengan menggunakan nilai parameter jarak pada aspek spasial dan temporal. Metode algoritma ini memiliki sifat yang lebih fleksibel terhadap ukuran data yang besar. Algoritma ST-DBSCAN merupakan pengembangan dari algoritma DBSCAN (Indrawan & Adrianto, 2016). Dimana ST-DBSCAN ini berbeda dengan algoritma clustering lainnya yang berbasis kepadatan, algoritma ST-DBSCAN memiliki kemampuan untuk menemukan *klaster* yang berkaitan dengan nilai-nilai spasial dan temporal dari objek. Adapun modifikasi terhadap algoritma DBSCAN adalah sebagai berikut (Birant & Kut, 2007):

1. ST-DBSCAN dapat mengelompokkan data spasial temporal menurut atribut non spasial, spasial, dan temporal.
2. *Density Based Spatial Clustering of Application with Noise* (DBSCAN) tidak dapat mendeteksi beberapa titik noise ketika ada klaster dengan kepadatan berbeda, algoritma ST-DBSCAN mampu memecahkan masalah dengan menetapkan factor kepadatan ke setiap kluster
3. Algoritma ST-DBSCAN memecahkan masalah ini dengan membandingkan nilai rata-rata klaster dengan nilai baru yang akan datang

Analisis data spatio-temporal yang menggunakan metode ST-DBSCAN menghasilkan klaster mengikuti pola tertentu, yaitu pola stasionary, reappearing, occasional dan track (Politz & Andrienko, 2010). Pola stationary menunjukkan klaster berada pada tempat terbatas dan waktu berlangsung pada setiap waktu atau pada saat periode data dikumpulkan. Pola reappearing menunjukkan kejadian spatio-temporal muncul pada lokasi yang sama tetapi dipisahkan oleh rentang waktu kejadian tersebut muncul tidak banyak bahkan tidak ada. Pola reappearing dapat dibagi menjadi dua tipe, yaitu tipe regular dan irregular. Pola regular menunjukkan tentang waktu yang sama, sedangkan pola irregular menunjukkan rentang waktu yang tidak beraturan. Pola occasional menunjukkan kejadian spatio-temporal yang muncul pada periode tertentu tidak berada pada lokasi yang sama. Pola tracks menunjukkan terjadi perubahan ruang berdasarkan kejadian sebelumnya.

### 1.6.7. K- Nearest Neighbor ST-DBSCAN

K-Nearest Neighbor ST-DBSCAN adalah variasi dari algoritma ST-DBSCAN (Spasial Clustering Density Based Spatial Clustering of Applications With Noise) yang menggunakan KNN (K- Nearest Neighbors) untuk menentukan parameter epsilon yang adaptif. Algoritma ini berguna untuk mengelompokkan data spasial atau temporal berdasarkan kedekatannya dan mengidentifikasi titik-titik noise. ST- DBSCAN merupakan algoritma clustering yang populer karena mampu mengatasi data yang memiliki kepadatan yang bervariasi dan dapat menangani pencilan (outlier) dengan baik, namun membutuhkan parameter epsilon dan minimum points yang optimal. KNN- ST-DBSCAN mencoba memperbaiki masalah ini dengan mengadaptasi parameter epsilon menggunakan metode KNN. K- Nearest neighbor merupakan algoritma yang melakukan pengelompokan berdasarkan kedekatan jarak suatu data dengan data yang lain. Tahapan dari KNN secara umum adalah:

1. Tentukan jumlah tetangga (K) yang akan digunakan untuk pertimbangan penentuan kelas.
2. Hitung jarak dari data baru ke masing-masing data point di dataset
3. Ambil sejumlah K dengan jarak terdekat, kemudian tentukan kelas dari data baru tersebut

K-Nearest Neighbor sebuah data mempunyai tetangga sejumlah  $k$ , yang dapat mencari  $k$  data terdekat untuk tiap data. Untuk mengetahui jarak antar tiap data, dapat menggunakan Euclidean Distance. Untuk setiap data akan dilakukan sorting berdasarkan data yang paling dekat sampai yang paling jauh. Kemudian akan diambil sebanyak  $k$ -neighbor yang terdekat atau mempunyai jarak terkecil dari data tersebut. Formula pencarian jarak dijelaskan pada formula berikut:

$$dist(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (12)$$

dengan

$p_i$	= Fitur ke- $i$ pada titik $p$
$q_i$	= Fitur ke- $i$ pada titik $q$
$N$	= Banyaknya fitur
$dist(p, q)$	= jarak dari titik $p$ ke titik $q$

Menghitung jarak pada k-nearest neighbor suatu data, dapat dilakukan dengan mencari jarak terjauh dari pada tetangga pada suatu data. Dengan mencari jarak terjauh ini, didapatkan jarak k-nearest neighbor yang nantinya akan digunakan sebagai acuan untuk mencari nilai parameter densitas. Formula pencarian jarak k-nearest neighbor dijelaskan pada formula:

$$Knn(p) = \max\{dist(p, q) | q \in kNeighbor(p)\} \quad (13)$$

dengan:

$dist(p, q)$	= Jarak dari titik $p$ ke titik $q$
$kNeighbor(p)$	= Tetangga dari titik $p$

$Knn(p)$  = Jarak tetangga terjauh dari titik  $p$

Hasil dari *k-nearest neighbor* tidak hanya jarak, tetapi juga tetangga dari sebuah data. Setiap data dicari tetangganya kemudian diurutkan berdasarkan jarak. Hal ini dilakukan supaya pada saat melakukan pelebaran kluster pada ST-DBSCAN menjadi semakin lebih mudah. Formula untuk pengurutan tetangga pada titik  $p$  terdapat pada formula berikut:

$$N_k(p) = \{q \in D, d(p, q_{i-1}) \leq d(p, q_i), i = 1, \dots, k\} \quad (14)$$

dengan

$N_k(p)$  = Tetangga dari titik  $p$  dengan jumlah tetangga sebanyak  $k$  titik

$D$  = Dataset yang digunakan

$k$  = Banyaknya tetangga pada suatu titik

$i$  = Nilai iterasi dari  $k$

$d(p, q_{i-1})$  = Jarak dari titik  $p$  ke titik  $q_{i-1}$

$d(p, q_i)$  = Jarak dari titik  $p$  ke titik  $q_i$

### 1.6.8. Metode Analisis Kluster

Analisis kluster merupakan metode *unsupervised learning* sehingga data yang digunakan untuk *clustering* dianggap tidak mempunyai label asli. Pengecekan performa pada metode *clustering* tidak dapat dilakukan dengan perhitungan jumlah *error* seperti pada klasifikasi (Indrawan & Adrianto, 2014). Pengecekan kualitas metode kluster dilakukan dengan menghitung similaritas dan disimilaritas antar *kluster point*. Jika jarak antar kluster menunjukkan disimilaritas dan jarak dalam kluster menunjukkan similaritas, performa kluster dapat dikategorikan baik.

#### a Indeks *Silhouette Coefficient*

Metode validasi *silhouette index* merupakan salah satu validasi yang berbasis kriteria internal. Secara umum, koefisien *silhouette* menggunakan perbandingan jarak rata-rata objek dalam kluster dengan jarak rata-rata semua objek yang berbeda di kluster yang berbeda (Nahdliyah et al., 2019). Menghitung koefisien *silhouette* yang didefinisikan dengan rata-rata  $s(i)$ . Langkah dalam menghitung *Silhouette coefficient* dimulai dengan mencari jarak rata-rata ke-  $i$  dengan semua data di kluster yang sama, diasumsikan data ke- $i$  berada di kluster A. Rumus dari  $a(i)$  ditulis dalam persamaan :

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (15)$$

dengan  $A$  merupakan banyaknya data di kluster A

Selanjutnya menghitung nilai  $b(i)$  yang merupakan nilai minimum dari jarak rata-rata data ke-  $i$  dengan semua data di kluster berbeda. Diasumsikan kluster berbeda selain A dengan kluster C. Maka perhitungan jarak rata-rata data ke- $i$  dengan semua data di kluster C ditulis sebagai berikut:

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (16)$$

Dengan C merupakan banyaknya data di kluster C

Setelah menghitung nilai  $d(i, C)$  untuk semua kluster  $C \neq A$ , selanjutnya memilih nilai jarak paling minimum sebagai nilai  $b(i)$

$$b(i) = \min_{C \neq A} d(i, j) \quad (17)$$

Jika kluster B memiliki nilai jarak minimum, maka  $d(i, B) = b(i)$  yang disebut sebagai tetangga dari data ke-i dan merupakan kluster terbaik kedua untuk data ke-i setelah kluster A. Setelah  $a(i)$  dan  $b(i)$  diketahui, maka proses terakhir menghitung Silhouette Coefficient sebagai berikut:

$$s(i) = \frac{b(i) - a(i)}{\max a(i), b(i)} \quad (18)$$

Keterangan:

$a(i)$  : Rata-rata jarak di dalam kluster

$b(i)$  : Rata-rata jarak dengan kluster lain

Nilai  $s(i)$  berada diantara -1 dan 1, di mana setiap nilai diinterpretasi sebagai berikut:

$s(i) \approx 1 \Rightarrow$  data ke-i digolongkan dengan baik (dalam A)

$s(i) \approx 0 \Rightarrow$  data ke-i berada di tengah antara dua kluster (A dan B)

$s(i) \approx -1 \Rightarrow$  data ke-i digolongkan dengan lemah baik (dekat ke kluster B daripada A). Penafsiran nilai *silhouette coefficient* ditunjukkan dalam Tabel 1.

Tabel 1. Interpretasi Nilai Silhouette Coefficient

<i>Silhouette Coefficient</i>	Interpretasi
0.71 - 1.00	Struktur yang dihasilkan kuat
0.41 - 0.70	Struktur yang dihasilkan baik
0.1 - 0.4	Struktur yang dihasilkan lemah
< 0.1	Tidak terstruktur

Sumber : (Chimwayi & Anuradha, 2018)

#### b Indeks Davis- Bouldin (DBI)

Indeks Davis- Bouldin merupakan salah satu metode evaluasi internal yang mengukur kluster pada suatu metode pengelompokan yang didasarkan pada nilai korelasi dan separasi (Sakti, 2017). Dalam suatu pengelompokan, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap centroid dari kluster yang diikuti. Sedangkan separasi pada jarak antara centroid dari kluster-nya. *Sum of square kluster* (SSW) merupakan persamaan yang digunakan untuk mengetahui matrik kohesi dalam sebuah kluster ke-i yang dirumuskan sebagai berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (19)$$

dengan

$m_i$  : Jumlah data dalam kluster ke- $i$

$c_i$  : Centroid kluster ke- $i$

$d()$  : Jarak setiap data ke centroid

*Sum of square between kluster* (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar kluster yang dihitung menggunakan persamaan:

$$SSB_{i,j} = d(c_i, c_j) \quad (20)$$

dengan

$d()$  : Jarak setiap data ke centroid

$c_i, c_j$  : Matriks pembobot

Setelah nilai kohesi (jumlah dari kedekatan data terhadap centroid dari kluster yang diikuti) dan separasi (jarak antar centroid dari klasternya) diperoleh, kemudian pengukuran rasio ( $R_{i,j}$ ) untuk mengetahui nilai perbandingan antara kluster ke- $i$  dan kluster ke- $j$ . Kluster yang baik adalah kluster yang memiliki nilai kohesi sekecil mungkin dan separasi yang sebesar mungkin. Nilai rasio dihitung menggunakan persamaan sebagai berikut:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (21)$$

Sifat-sifat yang dimiliki  $R_{i,j}$  sebagai berikut:

- a.  $R_{i,j} > 0$
- b.  $R_{i,j} = R_{j,i}$
- c. Jika  $SSW_j > r$  dan  $SSB_{i,j} = SSB_{i,r}$  maka  $R_{i,j} = R_{i,r}$
- d. Jika  $SSW_j = r$  dan  $SSB_{i,j} = SSB_{i,r}$  maka  $R_{i,j} > R_{i,r}$

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai Davies-Bouldin Indeks (DBI) dari persamaan berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (22)$$

Dari persamaan tersebut,  $k$  merupakan jumlah kluster yang digunakan. Jadi nilai DBI didapat dari nilai rata-rata  $R_{i,j}$ . Dari syarat-syarat perhitungan yang didefinisikan di atas, dapat diamati bahwa semakin kecil SSW maka hasil clustering yang didapat juga semakin baik. Secara esensial, DBI menginginkan nilai sekecil (non-negatif  $\geq 0$ ) mungkin untuk menilai baiknya kluster yang didapat.

c Indeks Calinski-Harabasz (CH)

Indeks Calinski-Harabasz bertujuan untuk mengevaluasi validitas kluster berdasarkan perhitungan between sum of square dan within sum of square (Central, 2017), Indeks CH dapat dihitung menggunakan persamaan sebagai berikut:

$$CH(K) = \frac{(SSW_i / K - 1)}{(SSB_{i,j} / N - K)} \text{ untuk } K \in N \quad (23)$$

Dimana :

K : Kelompok

N : Obajek yang dikelompokkan

SSW : Jumlah kuadrat error antar kelompok dalam kluster ke-i.

### 1.6.9. Pola Spatio- Temporal

Pola sebaran titik yang terdiri dari 4 tipe kluster berikut merupakan tipe kluster (Pölitz & Andrienko, 2010)

- a. *Reappearing* merupakan tipe kluster temporal yang terjadi dalam lokasi yang sama dan dipisahkan oleh interval waktu. Tipe kluster *reappearing* dibagi menjadi dua yaitu regular (periodik) kluster yang memiliki interval waktu serta tempat kemunculan titik yang sama dan irregular kluster yang memiliki interval waktu yang berbeda namun tempat kemunculan titik sama.
- b. Stationary merupakan kluster diperluas secara temporal pada seluruh rentang waktu penelitian atau dari waktu tertentu sampai akhir waktu tertentu dalam rentang waktu penelitian namun terbatas secara spasial.
- c. Occasional merupakan kluster yang memiliki perpindahan spasial dan perubahan temporal secara bersamaan.
- d. Tracks merupakan kluster yang mengalami pergerakan secara spasial dan memiliki temporal yang sangat padat.

Kluster besar yang dapat dianalisis polanya adalah kluster yang memiliki setidaknya 30 titik, sementara kluster dengan jumlah titik kurang dari 30 dianggap kecil dan polanya tidak dapat terdeteksi.

### 1.6.10. Gempa Bumi

Gempa bumi merupakan getaran bumi yang terjadi akibat pelepasan energi di dalam bumi secara tiba-tiba sehingga efeknya dapat dirasakan sampai ke permukaan bumi. Penyebab gempa bumi dapat berupa dinamika bumi (tektonik), aktivitas gunung api, akibat meteor jatuh, longsor (di bawah muka air laut), atau ledakan bom di bawah permukaan. Gempa bumi terjadi apabila penumpukan energi pada batas lempeng yang bersifat konvergen (bertumbukan), divergen (saling menjauh) dan transform (berpapasan) atau pada sesar (patahan) dan blok batuan tersebut tidak mampu lagi menahan batas elastisitasnya, sehingga akan dilepaskan sejumlah energi dalam bentuk rangkaian gelombang seismic yang dikenal sebagai gempa bumi (Susanta et al., 2019).

Gempa bumi merupakan salah satu dari berbagai macam bencana alam yang ada di Indonesia terutama di Pulau Sulawesi. Pulau Sulawesi secara geologis diyakini oleh para ahli kebumihantoran terletak pada pertemuan lempeng (*complex junction*) utama dunia yaitu lempeng Eurasia, lempeng Indo-Australia, lempeng Pasifik dan satu lempeng mikro yaitu lempeng Laut Filipina. Akibat tekanan dari pergerakan lempengan-lempengan tersebut menyebabkan interior lempengan bumi dari Kepulauan Sulawesi terpecah-pecah menjadi empat lengan yakni lengan Selatan, lengan Tenggara, lengan timur dan lengan utara yang menyerupai huruf K. Pulau Sulawesi merupakan pusat benturan keempat lempeng kerak bumi (Baeda & Husain, 2012). Dampak dari

benturan antar lempeng kerak bumi yang berbeda jenis tersebut menimbulkan terjadinya penimbunan energi sehingga dalam kurun waktu tertentu akan dilepaskan secara tiba-tiba dalam bentuk gempa bumi yang beragam (Pasau et al., 2017).

Badan Meterologi, Klimatologi dan Geofisika (BMKG) pusat gempa Regional Wilayah IV Makassar menerbitkan kaleidoskop gempabumi tahun 2023 di wilayah provinsi Sulawesi Selatan, Sulawesi Tengah, Sulawesi Tenggara dan sebagian pulau Kalimantan menyampaikan bahwa selama periode tahun 2023 periode 1 Januari 2023 sampai 27 Desember 2023 tercatat sebanyak 3.214 kali kejadian gempa bumi baik yang terjadi didarat maupun dilaut, dari 3.214 kejadian gempa bumi tersebut jika dibandingkan dengan kejadian gempa bumi pada tahun 2022 yang berjumlah 2.862 terjadi peningkatan jumlah gempa sebanyak 352 kali kejadian atau peningkatan sebanyak 1.26% (Humas BMKG, 2023).

Informasi yang berkaitan dengan kejadian gempa bumi meliputi waktu kejadian lokasi episenter, kedalaman sumber gempa bumi, dan magnitudo. Waktu kejadian gempa bumi adalah waktu terlepasnya akumulasi tegangan (*stress*) yang berbentuk penjaralan gelombang gempa bumi dan dinyatakan dalam hari, tanggal, bulan, tahun, jam dan detik. Episenter adalah titik dipermukaan bumi yang merupakan refleksi tegak lurus dan hiposenter atau focus gempa bumi Dimana Lokasi episenter dibuat dalam sistem koordinat kartesian bola bumi atau system koordinat geografis dan dinyatakan dalam derajat lintang dan bujur. Kedalaman sumber gempa bumi adalah jarak hiposenter dihitung tegak lurus dari permukaan bumi yang dinyatakan oleh besaran jarak dalam satuan kilometer(km). Badan Meteorologi, Klimatologi dan Geofisika telah menetapkan penggolongan (kedalaman) dan kekuatan (magnitudo) gempa bumi dengan beberapa jenis, dapat dilihat pada Tabel 2.

**Tabel 2.** Penggolongan Kedalaman Gempa

<b>Penggolongan</b>	<b>Kedalaman</b>	<b>Deskripsi</b>
Gempa Dangkal	0 – 70 km	Gempa dangkal seringkali paling merusak pusat gempa berada dekat dengan permukaan bumi
Gempa Menengah	70 – 300 km	Gempa ini biasanya kurang merusak, tetapi getarannya dirasakan oleh manusia dan dapat menyebabkan kerusakan tergantung pada magnitudonya
Gempa Dalam	300–700 km	Gempa dalam biasanya memiliki intensitas yang lebih rendah di permukaan, getarannya lemah dan lebih jarang menyebabkan kerusakan besar.

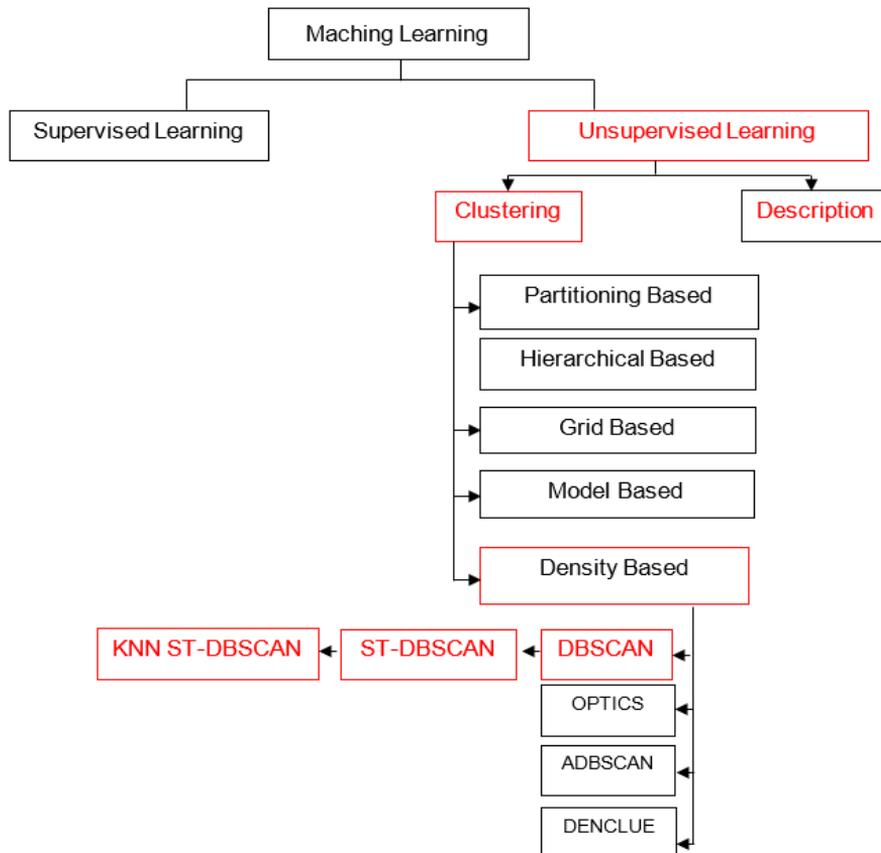
Sumber: (BMKG, 2018)

**Tabel 3.** Penggolongan Magnitudo Gempa

<b>Penggolongan</b>	<b>Magnitudo</b>	<b>Deskripsi</b>
Gempa Mikro	< 2,0	Gempa ini biasanya tidak dirasakan oleh manusia tapi direkam oleh seismograf
Gempa Minor	2,0 – 3,9	Gempa ini biasanya terasa tetapi jarang menyebabkan kerusakan
Gempa Ringan	4,0 – 4,9	Gempa ini dapat dirasakan oleh orang banyak tetapi biasanya hanya menyebabkan kerusakan minor
Gempa Sedang	5,0 – 5,9	Gempa ini dapat menyebabkan kerusakan ringan hingga sedang pada bangunan dan struktur
Gempa Kuat	6,0 – 6,9	Gempa ini menyebabkan kerusakan parah di daerah padat penduduk
Gempa Sangat Kuat	7,0 – 7,9	Gempa ini menyebabkan kerusakan besar dan meluas, terutama di daerah yang dekat dengan pusat gempa
Gempa Dahsyat	≥ 8,0	Gempa ini dapat menyebabkan kerusakan yang sangat besar dan meluas, mempengaruhi daerah yang sangat luas

Sumber: (BMKG, 2018)

### 1.6.11. Kerangka Konsep



Gambar 1. Kerangka Pikir

## BAB II

### METODE PENELITIAN

#### 2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data gempa bumi di Pulau Sulawesi pada Tahun 2021- 2023. Data diperoleh dari BMKG pada situs [earthquake.usgs.gov](http://earthquake.usgs.gov) dengan atribut tanggal, longitude latitude, kedalaman gempa (depth) dan kekuatan gempa (magnitudo)  $\geq 4,0$  SR.

#### 2.2 Variabel Penelitian

Variabel penelitian yang digunakan terdiri dari tanggal kejadian (time), longitude, latitude, kedalaman gempa (depth), dan kekuatan gempa (mag).

**Tabel 4.** Variabel Penelitian

Variabel	Defenisi Operasional Variabel	Satuan
Waktu kejadian	Waktu terjadinya gempa bumi di Pulau Sulawesi	Tanggal/Bulan/Tahun
Longitude	Koordinat lintang titik yang terjadi gempa bumi di Pulau Sulawesi	Derajat
Latitude	Koordinat bujur titik yang terjadi gempa bumi di Pulau Sulawesi	Derajat
Magnitudo	Kekuatan gempa bumi yang terjadi di Pulau Sulawesi	SR
Depth	Kedalaman titik pusat gempa bumi yang terjadi di Pulau Sulawesi	Km

#### 2.3 Metode Analisis

Penelitian ini melakukan analisis dengan metode algoritma K- Nearest Neighbor Spasio Temporal Density Based Clustering. Adapun langkah-langkah yang dilakukan untuk menganalisis data dalam penelitian tersebut sebagai berikut:

1. Melakukan processing data
2. Mengidentifikasi pola sebaran data menggunakan analisis jarak terdekat yaitu Nearest – Neighbor Analysis menggunakan persamaan 4.
3. Memeriksa apakah hasil distribusi data menunjukkan adanya kelompok atau tidak dengan menggunakan persamaan 8.
4. Menentukan nilai  $\varepsilon_1$  dan MinPts dilakukan dengan menggunakan algoritma K-

Nearest Neighbor (KNN) distance, kemudian  $\varepsilon_2$  ditentukan sesuai dengan periode yang akan dianalisis

5. Menghitung jarak Euclidean antar objek berdasarkan aspek spasial (longitude latitude) dan berdasarkan aspek temporal (tanggal terjadinya gempa).

$$Euclidean(p, q) = \sqrt{((p_1 - q_1)^2) + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

6. Membentuk matriks jarak untuk semua pasangan dari n objek berdasarkan spasial dan temporal.
7. Mulai dari titik pertama kemudian mengambil semua titik pada aspek spasial dan aspek temporal dengan ketentuan:

$$A = \{x | x \leq \varepsilon_1, x \in \text{matriks jarak spasial}\}$$

$$B = \{x | x \leq \varepsilon_1, x \in \text{matriks jarak temporal}\}$$

8. Mengambil semua irisan dari aspek spasial dan aspek temporal dengan ketentuan:

$$A \cap B = \{x | x \in A \text{ dan } x \in B\}$$

9. Jika jumlah obojek pada irisan lebih kecil dari MinPts, maka titik dinilai sebagai noise
10. Kluster terbentuk jika titik memenuhi parameter  $\varepsilon_1$ ,  $\varepsilon_2$ , dan MinPts
11. Lakukan pengujian Indeks Silhouette Coefficient, Indeks DBI dan Indeks CH untuk menguji optimal atau ketepatan sebuah kluster yang telah terbentuk dari proses *clustering*
12. Melakukan visualisai dan interpretasi hasil kluster yang terbentuk dari KNN ST-DBSCAN
13. Menganalisa pola spatio-temporal.