

**METODE *DBSCAN CLUSTERING* UNTUK ANALISIS POLA
PENYEBARAN HUJAN DI SUMENEP**

SKRIPSI

**OLEH:
MUHAMMAD AMIN CHERID
NIM. 19610011**



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

**METODE *DBSCAN CLUSTERING* UNTUK ANALISIS POLA
PENYEBARAN HUJAN DI SUMENEP**

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan
dalam Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh
Muhammad Amin Cherid
NIM. 19610011**

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

**METODE DBSCAN CLUSTERING UNTUK ANALISIS POLA
PENYEBARAN HUJAN DI SUMENEP**

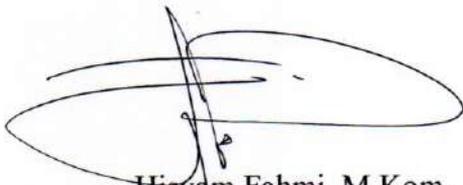
SKRIPSI

Oleh
Muhammad Amin Cherid
NIM. 19610011

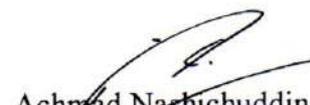
Telah Disetujui untuk Diuji
Malang, 22 Desember 2023

Dosen Pembimbing I,

Dosen Pembimbing II,



Hisyam Fahmi, M.Kom.
NIP. 19890727 201903 1 018



Achmad Nashichuddin, M.A.
NIP. 19730705 200003 1 002

Mengetahui,
Ketua Program Studi Matematika



Dr. Ely Susanti, M.Sc.
NIP. 19741129 200012 2 005

METODE *DBSCAN CLUSTERING* UNTUK ANALISIS POLA PENYEBARAN HUJAN DI SUMENEP

SKRIPSI

Oleh
Muhammad Amin Cherid
NIM. 19610011

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

Tanggal 27 Desember 2023

Ketua Penguji : Dr. Hairur Rahman, S.Pd., M.Si

Anggota Penguji 1 : Muhammad Khudzaifah, M.Si

Anggota Penguji 2 : Hisyam Fahmi, M.Kom.

Anggota Penguji 3 : Achmad Nashichuddin, M.A.

Mengetahui,
Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc.
NIP. 19741129 200012 2 005

HALAMAN PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Amin Cherid
NIM : 19610011
Program Studi : Matematika
Fakultas : Sains dan Teknologi
Judul Skripsi : Metode DBSCAN Clustering Untuk Analisis Pola
Penyebaran Hujan Di Sumenep

Menyatakan dengan sebenar-benarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan dan pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 27 Desember 2023

Yang membuat pernyataan,



Munhammad Amin Cherid
NIM. 19610011

HALAMAN MOTO

“Maka, sesungguhnya beserta kesulitan ada kemudahan. Sesungguhnya beserta kesulitan ada kemudahan.”

-Q.S Al Insyirah:5-6

HALAMAN PERSEMBAHAN

Bismillahirrahmanirrahim

Alhamdulillahillobbil alamin, Allahumma sholli ala sayyidina Muhammad wa ala ali sayyidina Muhammad.

Dengan rasa syukur yang mendalam, penulis ingin mengucapkan terima kasih kepada Allah SWT, Tuhan Yang Maha Esa, atas segala rahmat dan hidayah-Nya yang senantiasa membimbing langkah-langkah penulis dalam menyelesaikan perjalanan akademis ini.

Kepada almarhum ayahanda tercinta Syd Idrus Cherid, Ibunda tercinta Srf Fatimah Al Hinduan, yang tak pernah lelah memberikan dukungan, cinta, dan doa restu selama ini. Terima kasih tak terhingga atas pengorbanan dan kasih sayang yang telah diberikan. Skripsi ini adalah buah dari perjuangan kita bersama. Kepada keluarga tercinta, yang senantiasa memberikan semangat dan dukungan tanpa batas. Terima kasih karena telah menjadi sumber inspirasi dan kekuatan bagi penulis.

Tidak lupa kepada sahabat-sahabat terbaik yang selalu hadir dalam setiap lika-liku perjalanan ini. Terima kasih atas tawa, dukungan, dan motivasi yang kalian berikan. Kalian adalah bagian tak terpisahkan dari kisah ini.

Penulis juga ingin menyampaikan terima kasih kepada dosen pembimbing yang telah memberikan arahan, bimbingan, dan inspirasi dalam menyelesaikan skripsi ini. Ilmu yang diberikan menjadi cahaya yang menerangi perjalanan akademis penulis.

Akhir kata, penulis menyadari bahwa perjalanan ini tak mungkin terwujud tanpa dukungan dan bantuan dari berbagai pihak. Semoga skripsi ini dapat menjadi kontribusi kecil yang bermanfaat bagi ilmu pengetahuan dan masyarakat. Semoga Allah SWT senantiasa melimpahkan rahmat dan keberkahan kepada kita semua.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji bagi Allah SWT yang selalu melimpahkan rahmat, taufik dan hidayah-Nya, sehingga penulis dapat menyelesaikan penelitian yang berjudul “METODE *DBSCAN CLUSTERING* UNTUK ANALISIS POLA PENYEBARAN HUJAN DI SUMENEP” sebagai salah satu syarat untuk memperoleh gelar sarjana dalam bidang Matematika di Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Sholawat serta salam senantiasa tercurahkan kepada Nabi Muhammad SAW yang telah menuntun manusia dari jalan kegelapan menuju ke jalan yang terang benderang yaitu Islam. Dalam proses penyelesaian skripsi ini, peneliti banyak mendapat bimbingan, dukungan, dan arahan serta sumbangsi dari berbagai pihak. Oleh karena itu peneliti mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. M. Zainuddin, M.A., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Sri Harini, M.Si., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Elly Susanti, M.Sc., selaku Ketua Program Studi Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Hisyam Fahmi, M.Kom., selaku dosen pembimbing I yang telah memberikan berbagai pengetahuan, pengalaman, arahan, nasihat, serta motivasi kepada penulis.
5. Achmad Nashichuddin, M.A., selaku dosen pembimbing II yang telah memberikan ilmu, nasihat, bimbingan, pengalaman, serta motivasi kepada penulis.
6. Seluruh dosen Program Studi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
7. Kedua orang tua serta seluruh keluarga yang senantiasa mendoakan, memberikan semangat, dukungan, nasihat, serta kasih sayang sehingga penulis dapat menyelesaikan tugas akhir.

8. Rekan penulis yang berjuang bersama dalam mengerjakan skripsi yang senantiasa memberikan dukungan dan semangat kepada penulis.
9. Seluruh teman-teman Program Studi Matematika angkatan 2019 yang senantiasa memberikan bantuan, dukungan dan semangat kepada peneliti dalam berbagai kondisi.

Semoga Allah SWT melimpahkan rahmat dan karunia-Nya kepada kita semua. Selain itu, penulis juga berharap semoga penelitian ini dapat bermanfaat khususnya bagi penulis dan pembaca pada umumnya. Aamiin.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Malang, 27 Desember 2023

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN TULISAN	v
HALAMAN MOTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR PERSAMAAN	xiv
DAFTAR LAMPIRAN	xv
ABSTRAK	xvi
ABSTRACT	xvii
مستخلص البحث	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	5
1.6 Daftar Istilah	5
BAB II KAJIAN PUSTAKA	7
2.1 Data Mining	7
2.1.1 <i>Data Cleaning</i> (Pembersihan Data).....	8
2.1.2 <i>Data Selection</i> (Seleksi Data).....	9
2.1.3 <i>Data Transformation</i> (Transformasi Data)	9
2.1.4 <i>Data Mining</i> (Proses Mining).....	9
2.1.5 <i>Pattern Evaluation</i> (Evaluasi Pola)	10
2.2 <i>Clustering</i> (Pengklastran)	11
2.3 DBSCAN.....	11
2.4 <i>Cluster Validity</i>	14
2.5 Hujan	15
2.6 Hujan Dalam Al Qur'an	16
BAB III METODE PENELITIAN	19
3.1 Pendekatan Penelitian	19
3.2 Data dan Sumber Data	19
3.3 Variabel Penelitian.....	19
3.4 Tahapan Penelitian.....	20
BAB IV PEMBAHASAN	23
4.1 Deskripsi Data	23
4.2 Perhitungan <i>DBSCAN Clustering</i>	26
4.3 Hasil Eksperimen <i>DBSCAN Clustering</i>	28
4.4 Analisis Pola Penyebaran	32

4.5 Konsep <i>DBSCAN Clustering</i> Dalam Penyebaran Agama Islam	33
BAB V PENUTUP	36
5.1 Kesimpulan.....	36
5.2 Saran	37
DAFTAR PUSTAKA	38
LAMPIRAN	40
RIWAYAT HIDUP	42

DAFTAR GAMBAR

Gambar 3.1	<i>Flowchart DBSCAN Clustering</i>	21
Gambar 4.1	Hasil <i>DBSCAN Clustering</i> Dengan $\epsilon = 4$ dan <i>MinPts</i> = 6.....	30
Gambar 4.2	Hasil <i>DBSCAN Clustering</i> Dengan $\epsilon = 10$ dan <i>MinPts</i> = 6.....	31
Gambar 4.3	Persebaran Hujan Di Sumenep Pada Tahun 2022	32

DAFTAR TABEL

Tabel 4.1	Deskriptif Data Hujan di Sumenep Tahun 2022.....	26
Tabel 4.2	Eksperimen <i>DBSCAN Clustering</i>	29
Tabel 4.3	Tabel Hasil Eksperimen Dengan $\epsilon = 4$ dan <i>MinPts</i> = 6	30
Tabel 4.4	Tabel Hasil Eksperimen Dengan $\epsilon = 10$ dan <i>MinPts</i> = 6	31
Tabel 4.5	Analisis Hasil <i>DBSCAN Clustering</i>	33

DAFTAR PERSAMAAN

(2. 1)	<i>Euclidean distance</i>	14
(2. 2)	Rata-rata jarak i terhadap semua objek di klaster A	14
(2. 3)	Rata-rata jarak i terhadap semua objek di klaster C.....	15
(2. 4)	Nilai minimum rata-rata jarak i terhadap semua objek C	15
(2. 5)	Nilai <i>Silhouette score</i>	15

DAFTAR LAMPIRAN

Lampiran 1 Source Code

ABSTRAK

Cherid, Muhammad Amin. 2023. **Metode DBSCAN Clustering Untuk Analisis Pola Penyebaran Hujan Di Sumenep**. Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Hisyam Fahmi, M.Kom., (II) Achmad Nashichuddin, M.A.

Kata kunci: Hujan, Data Mining, DBSCAN, Pengklasteran

Hujan merupakan jatuhan berupa partikel-partikel air (hydrometeor) dengan diameter 0,5 mm atau lebih. Fokus penelitian ini adalah melakukan *data mining* untuk menggali pengetahuan baru dari data hujan. DBSCAN merupakan algoritma yang digunakan pada penelitian ini untuk menemukan jumlah kluster serta *noise* pada data petir berdasarkan kepadatan atau kerapatannya. Nilai epsilon (ϵ) dan *minimum points* (*MinPts*) diperoleh dari hasil eksperimen dengan memilih *Silhouette Score* yang terbaik. Tujuan penelitian ini adalah menganalisis hasil dari DBSCAN *Clustering* pada data hujan dengan keadaan geografis di wilayah Sumenep dan mengevaluasi hasil *DBSCAN clustering*.

Dari hasil penelitian ini, dengan menggunakan ϵ yang optimum, yaitu 10 dan *MinPts* sebanyak 6, diperoleh jumlah *cluster* sebanyak 3 dan *noise* sebanyak 13. Setelah dilakukan analisis, *cluster* pertama adalah lokasi dengan jumlah air hujan yang besar, yaitu dengan curah hujan lebih dari 300 mm. *Cluster* kedua adalah lokasi dengan jumlah air hujan sedang, yaitu dengan curah hujan sekitar 200-300 mm. *cluster* ketiga adalah lokasi dengan jumlah air hujan kecil, yaitu dengan curah hujan kurang dari 100 mm. Sedangkan untuk *noise* adalah hujan yang sangat jarang terjadi dan jumlah curah hujan yang sangat besar.

ABSTRACT

Cherid, Muhammad Amin. 2023. **DBSCAN Clustering Method for Pattern Analysis the he Spreading of Rain in Sumenep**. Undergraduate Thesis. Mathematics Department, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors: (I) Hisyam Fahmi, M.Kom. (II) Achmad Nashichuddin, M.A.

Keywords: Rain, Data Mining, DBSCAN, Clustering

Rain is the droplets of water particles (hydrometeor) with a diameter of 0.5 mm or bigger. The research focuses on data mining to explore a new understanding of rainfall data. The research employed the DBSCAN algorithm to determine the number of clusters and noise of lightning data based on its density. The epsilon (ϵ) and the minimum points (MinPts) were selected from the experiment data using the best Silhouette Score. The research aims to analyze the result of rainfall data clustering of Sumenep and evaluate the result of DBSCAN clustering.

From the research results, using the most optimum ϵ of 10 and MinPts of 6 achieves three clusters and 13 noises. After the analysis, the first cluster has the highest rainfall volume, more than 300 mm. The second cluster is the location with a moderate rainfall volume of 200-300 mm. The third cluster is the location with the lowest rainfall volume, less than 100 mm. Meanwhile, the noise is the rare rainfall and extremely high rainfall volume.

مستخلص البحث

خرد، محمد أمين. ٢٠٢٣. خوارزمية *DBSCAN* للتجميع في تحليل أنماط توزيع الأمطار في سومنب. البحث الجامعي. قسم الرياضيات، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: هشام فهمي، الماجستير. المشرف الثاني: أحمد ناصح الدين، الماجستير.

الكلمات الرئيسية: مطر، تنقيب عن بيانات، *DBSCAN*، تجميع.

المطر هو شكل من أشكال قطرات الماء المتساقطة من السحاب في السماء (الظواهر الجوية المائية) التي يبلغ قطرها ٠,٥ مم أو أكثر. يركز هذا البحث على إجراء التنقيب عن البيانات لاستكشاف المعرفة الجديدة من بيانات المطر. *DBSCAN* هي خوارزمية مستخدمة في هذا البحث للعثور على عدد المجموعات والضوضاء في بيانات البرق بناء على الكثافة أو المتانة. يتم الحصول على قيمة إبسيلون (ϵ) والحد الأدنى من النقاط (*MinPts*) من النتائج التجريبية عن طريق اختيار أفضل درجة صورة ظليلة. الهدف من هذا البحث هو تحليل نتائج خوارزمية *DBSCAN* للتجميع على بيانات الأمطار مع الظروف الجغرافية في منطقة سومنب وتقييم نتائج خوارزمية *DBSCAN* للتجميع.

من نتائج هذا البحث، باستخدام ϵ الأمثل، أي ١٠ و *MinPts* بقدر ٦، تم الحصول على عدد المجموعات ما يصل إلى ٣ و *noise* بقدر ١٣. بعد التحليل، المجموعة الأولى هي موقع به كمية كبيرة من مياه الأمطار، والتي يزيد هطول الأمطار فيها عن ٣٠٠ ملم. المجموعة الثانية هي موقع به كمية معتدلة من مياه الأمطار، والتي يبلغ هطول الأمطار حوالي ٢٠٠-٣٠٠ ملم. المجموعة الثالثة هي موقع به كمية صغيرة من مياه الأمطار، أي مع هطول أمطار أقل من ١٠٠ ملم. أما بالنسبة ل *noise* فهي أمطار نادرة جدا وكمية كبيرة جدا من الأمطار.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Curah hujan merupakan salah satu faktor penting dalam menentukan tipe atau variasi iklim, yang pada gilirannya memengaruhi segala aspek kehidupan seperti keselamatan masyarakat, aspek sosial-ekonomi, produksi pertanian, perkebunan, perikanan, penerbangan, dan lain sebagainya di suatu daerah. Oleh karena itu, informasi mengenai curah hujan memiliki dampak besar dan termasuk salah satu unsur yang sangat penting.

Hujan merupakan jatuhan berupa partikel-partikel air (hydrometeor) dengan diameter 0,5 mm atau lebih (Florence et al., 2014). Saat musim hujan, wilayah Indonesia yang dikelilingi oleh pulau-pulau dan garis khatulistiwa seringkali mengalami hujan dengan intensitas tinggi. Kondisi ini biasa terjadi mengingat cuaca di Indonesia yang tidak menentu, terutama dalam hal curah hujan. Namun, hal tersebut dapat meningkatkan risiko kecelakaan di berbagai jenis transportasi, baik laut, udara, maupun darat.

Di Indonesia, curah hujan diukur menggunakan satuan milimeter (mm). Intensitas hujan mengacu pada jumlah curah hujan dalam suatu periode waktu tertentu. Jika intensitas hujan dianggap tinggi, artinya hujan sangat deras, dan ini sangat berbahaya karena dapat menyebabkan banjir, longsor, dan berdampak buruk pada pertumbuhan tanaman (Maulidani S et al., 2015). Curah hujan dapat direkam menggunakan alat pengukur hujan seperti rain gauge secara manual atau otomatis. Untuk pengukuran manual, alat yang paling sering digunakan dalam

waktu 24 jam adalah alat pengukur hujan tipe Observatorium. Sedangkan untuk pengukuran otomatis, dapat digunakan alat pengukur hujan tipe Hilmann (Aldrian, 2003).

Hujan adalah salah satu sumber air yang sangat penting bagi manusia dan ekosistem. Hujan memberikan sumber air segar yang dapat diolah menjadi air minum. Ketersediaan air minum yang bersih sangat penting untuk mempertahankan kesehatan manusia. Hujan penting bagi kehidupan satwa liar dan ekosistem, karena memberikan air untuk minum dan membantu menjaga kelembaban udara. Pentingnya hujan bahkan telah diungkapkan oleh Allah SWT. Allah SWT telah berfirman di dalam Al-Qur'an surah Al-An'am Ayat 99 yang artinya:

“Dan Dialah yang menurunkan air hujan dari langit, lalu kami tumbuhkan dengan air itu segala macam tumbuh-tumbuhan.”

Berdasarkan tafsir Ibnu Katsir, artinya dengan kadar tertentu, air hujan adalah sebuah anugerah dan rezeki dari Allah kepada para hamba-Nya. Air hujan diperlukan untuk menghidupi dan menyuburkan berbagai makhluk, dan juga sebagai tanda kasih sayang Allah kepada seluruh makhluk-Nya. Setelah itu, kami menciptakan tanaman dan pohon yang hijau serta memberikan biji-bijian dan buah-buahan pada mereka (Abdul Ghofar, 2003).

Sebagian besar penelitian terbaru pada data spasial menggunakan teknik *clustering* dikarenakan oleh sifat dari data tersebut. *Clustering* merupakan proses pengelompokan sejumlah besar data menjadi beberapa kelas sesuai dengan ciri khasnya masing-masing. Di antara berbagai jenis algoritma *clustering*, *density based clustering* lebih efisien untuk menentukan *cluster* pada data dengan kepadatan yang berbeda. *Density-Based Spatial Clustering of Application with*

Noise (DBSCAN) adalah salah satu contoh pelopor perkembangan teknik pengelompokan berdasarkan kepadatan atau yang biasa dikenal dengan sebutan *density based clustering* (Pratama, 2016).

Hingga saat ini, jumlah penelitian yang mengkaji hujan di Sumenep tergolong minim. Salah satu contohnya ialah penelitian yang dilakukan oleh Muhandhis, yang membahas mengenai Implementasi Metode Inferensi Fuzzy Tsukamoto Untuk Memprediksi Curah Hujan Dasarian Di Sumenep. Metode inferensi fuzzy Tsukamoto memberikan hasil peramalan curah hujan yang akurat dengan MAPE sebesar 10,64%. Metode Fuzzy Time Series Cheng memiliki nilai MAPE terkecil sebesar 5,6%, menjadikannya sebagai metode fuzzy terbaik dibandingkan dengan metode-metode fuzzy lain yang digunakan sebelumnya. Metode Average Based Fuzzy Time Series Markov-Chain juga menghasilkan peramalan yang baik dan menjadi pilihan kedua terbaik. Faktor ini memiliki pengaruh besar terhadap akurasi peramalan ketika membentuk rentang nilai linguistik pada himpunan fuzzy. Pada tahun 2020, Metode Fuzzy Inference System Tsukamoto mampu memprediksi awal musim kemarau dan awal musim hujan di Dasarian 2 pada bulan November dengan akurasi yang tinggi (Muhandhis et al., 2021).

Sangat penting untuk melakukan pengolahan informasi curah hujan dengan metode yang tepat, sehingga data tersebut dapat dikategorikan menjadi tiga kategori yaitu rendah, sedang, dan tinggi dengan akurasi yang tinggi. Hal ini akan memungkinkan pemanfaatan informasi mengenai pola curah hujan yang terjadi di suatu wilayah sesuai dengan kebutuhan. Penelitian ini melakukan pengolahan data dengan metode statistik yang bertujuan untuk mendapatkan

kelompok curah hujan dengan karakteristik yang sama. Dalam lingkup klimatologi, metode pengelompokan objek yang memiliki karakteristik serupa adalah dengan menggunakan analisis kluster (Mimmack et al., 2000). Dari analisis kluster, didapat kelompok-kelompok yang memiliki tingkat kemiripan yang tinggi dalam satu kluster, namun memiliki tingkat ketidakmiripan yang tinggi antara satu kluster dengan kluster yang lain. Metode *Density-Based Spatial Clustering of Application with Noise* (DBSCAN) cocok untuk diterapkan dalam penelitian ini.

1.2 Rumusan Masalah

1. Bagaimana hasil *clustering* dari data hujan di Sumenep menggunakan algoritma DBSCAN?
2. Bagaimana hasil evaluasi *cluster* sebaran hujan dengan algoritma DBSCAN?

1.3 Tujuan Penelitian

1. Mengetahui hasil *clustering* dari data hujan di Sumenep menggunakan algoritma DBSCAN.
2. Mengetahui hasil evaluasi *cluster* sebaran hujan dengan algoritma DBSCAN.

1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini antara lain:

1. Teoritis
Memberikan informasi terkait metode *DBSCAN Clustering* untuk analisis pola penyebaran hujan di Sumenep serta dapat dijadikan bahan referensi

untuk penelitian selanjutnya.

2. Praktis

- a. Membantu dalam perencanaan dan pengambilan keputusan dalam pengelolaan sumber daya air dan irigasi di Sumenep, terutama dalam hal distribusi air pada saat musim hujan.
- b. Mengetahui daerah rentan hujan beserta ciri-cirinya, sehingga dapat menghindari terjadinya dampak negatif yang akan timbul akibat hujan di daerah tersebut.

1.5 Batasan Masalah

Untuk memastikan penelitian ini terarah sesuai dengan tujuan yang diinginkan, maka penting untuk menetapkan batasan pada masalah yang akan dikaji. Mengabaikan aspek temporal pada data pada saat melakukan *clustering*, hanya dikelompokkan berdasarkan aspek spasial/lokasi.

1.6 Daftar Istilah

1. *Clustering*

Sebuah proses pengelompokan data atau objek-objek yang mempunyai kesamaan berdasarkan kriteria tertentu.

2. *Density-based Clustering*

Algoritma pengelompokan berbasis kepadatan, seperti DBSCAN (Density-Based Spatial Clustering of Applications with Noise), menggunakan kepadatan data untuk mengidentifikasi kluster.

3. *Noise*

Titik-titik yang bukan titik inti dan juga tidak memiliki tetangga di dalam

radius tertentu.

4. *Core-point*

Titik-titik data yang memiliki jumlah minimum tetangga dalam radius tertentu.

5. *Border-point*

Titik-titik yang bukan titik inti tetapi memiliki setidaknya satu tetangga di dalam radius tertentu.

6. *Minimum point*

Parameter yang menentukan jumlah minimum tetangga dalam radius tersebut untuk suatu titik dianggap sebagai titik inti.

7. *Epsilon*

Parameter yang menentukan radius di sekitar setiap titik inti.

BAB II

KAJIAN PUSTAKA

2.1 Data Mining

Data Mining telah lama dikenal sebagai bidang studi yang tidak baru. Namun, mendefinisikan *Data Mining* menjadi tantangan karena terdapat beragam aspek dan teknik yang berasal dari berbagai bidang ilmu sebelumnya. Dari beberapa disiplin ilmu, *Data Mining* memiliki tujuan untuk meningkatkan teknik-teknik tradisional agar dapat mengatasi berbagai masalah seperti jumlah data yang sangat besar, dimensi data yang tinggi, dan keberagaman jenis data (Maulana & Fajrin, 2018).

Data Mining terbagi ke dalam beberapa kelompok yaitu:

1. Deskripsi

Deskripsi adalah suatu teknik untuk menjelaskan pola dan tren yang ada dalam data yang tersedia. Terdapat banyak cara yang dapat digunakan dalam proses deskripsi, salah satunya adalah melalui penggunaan statistik deskriptif..

2. Estimasi

Estimasi memiliki kemiripan dengan klasifikasi, namun perbedaannya terletak pada fokus variabel target. Estimasi lebih berorientasi pada variabel target yang bersifat numerik daripada kategori. Model yang dibuat dengan teknik estimasi menggunakan seluruh data yang tersedia dan menggunakan nilai variabel target sebagai nilai prediksi.

3. Prediksi

Prediksi sering digunakan untuk memproyeksikan nilai pada waktu yang akan datang. Proses prediksi membutuhkan data masa lalu dan teknik tertentu yang digunakan untuk menganalisis tren dan pola dalam data tersebut, sehingga memungkinkan untuk memperkirakan nilai yang mungkin terjadi di masa depan.

4. Klasifikasi

Klasifikasi melibatkan penggunaan kategori sebagai variabel target, seperti contohnya pengelompokan pendapatan ke dalam tiga kategori, yakni tinggi, sedang, dan rendah. Teknik klasifikasi juga bisa digunakan untuk memprediksi kategori yang akan ditempati oleh suatu data.

5. Pengklasteran

Pengklasteran adalah teknik untuk mengelompokkan data, pengamatan, atau objek-objek berdasarkan kesamaan karakteristik yang dimilikinya. Umumnya, pengklasteran melibatkan perhitungan jarak antara objek-objek tersebut untuk mengetahui tingkat kesamaannya dalam data (Fahamsah, 2020).

2.1.1 Data Cleaning (Pembersihan Data)

(Eska, 2018) menyatakan bahwa sebelum melakukan *data mining*, langkah awal yang perlu dilakukan adalah membersihkan data (*cleaning*) yang akan digunakan dalam *Knowledge Discovery in Data* (KDD). Proses penghilangan duplikasi data, verifikasi konsistensi data, dan perbaikan kesalahan pada data termasuk dalam proses pembersihan data. Selain itu, terdapat proses *enrichment* yang bertujuan untuk meningkatkan kualitas data dengan menambahkan informasi

atau data eksternal yang relevan untuk KDD.

2.1.2 Data Selection (Seleksi Data)

Hanya data yang relevan yang diambil dari *data base* untuk dianalisis, tidak semuanya digunakan. Sebagai contoh, ketika melakukan analisis faktor kecenderungan pembeli barang dalam *market basket analysis*, informasi pelanggan seperti nama tidak perlu digunakan, tetapi hanya *id* pelanggan yang dibutuhkan (Eska, 2018).

Seleksi data adalah proses untuk mengurangi jumlah data yang digunakan dalam proses *mining* tanpa mengurangi representasi data asli. Teknik *data selection* dapat berupa pengambilan sampel (*sampling*), pembersihan noise (*denoising*), dan ekstraksi fitur (*feature extraction*) (Sulastri & Gufroni, 2017).

2.1.3 Data Transformation (Transformasi Data)

Proses transformasi data merupakan tindakan merubah atau mengkombinasikan data ke dalam bentuk yang sesuai untuk diproses oleh teknik *data mining*. Sebelum dapat diaplikasikan, beberapa teknik *data mining* memerlukan format data yang spesifik. Contohnya, beberapa metode standar seperti analisis asosiasi dan pengklasteran hanya menerima *input* data yang bersifat kategorikal. Sehingga, untuk dapat digunakan pada teknik-teknik tersebut, data numerik yang kontinu perlu dipecah menjadi beberapa interval (Eska, 2018).

2.1.4 Data Mining (Proses Mining)

Proses mining merupakan langkah utama dalam menerapkan metode untuk

mengungkapkan pengetahuan yang berharga dan tersembunyi dari data. Pada langkah ini, algoritma *Machine Learning* mulai diaplikasikan. Algoritma yang telah ditentukan digunakan dalam proses *mining* untuk menghasilkan pola baru pada data yang telah diproses (Eska, 2018).

Setelah data terkumpul, maka data tersebut harus dianalisis, diproses, dan diubah menjadi informasi yang dapat memberikan jawaban, instruksi, pemahaman, atau keputusan. Kemampuan untuk mengekstraksi informasi ini sangatlah penting, terutama dalam dunia bisnis yang sangat kompetitif saat ini. Contohnya, ketika data telah digunakan untuk melakukan prediksi, maka hasil prediksi tersebut dapat digunakan untuk mengambil keputusan yang diinginkan (Stubbe & Coleman, 2014).

Metode sains digunakan dalam *data mining* dengan melakukan eksplorasi dan aplikasi pada data yang ada. Apabila data yang dianalisis terlalu besar, maka dapat dilakukan pemilihan sebagian data untuk dianalisis. Hasil analisis pada data sampel dapat diterapkan pada populasi secara umum, yang artinya hasil tersebut dapat digeneralisasi dan relevan dengan keadaan populasi secara keseluruhan (Stubbe & Coleman, 2014).

2.1.5 *Pattern Evaluation* (Evaluasi Pola)

Tahap evaluasi pola ini dipergunakan untuk menemukan pola-pola menarik dalam *knowledge base* yang ditemukan dari teknik *data mining*. Tahap ini melibatkan evaluasi terhadap hasil pola khas atau model prediksi yang ditemukan, dengan tujuan untuk menilai apakah hipotesa yang telah dibuat dapat tercapai (Eska, 2018).

2.2 *Clustering* (Pengklasteran)

Menurut (Maulana & Fajrin, 2018), *clustering* adalah sebuah proses pengelompokan data atau objek-objek yang mempunyai kesamaan berdasarkan kriteria tertentu. Proses *clustering* bertujuan untuk mengelompokkan data dengan karakteristik serupa ke dalam satu *cluster* dan memisahkan data dengan karakteristik yang berbeda ke dalam kelompok yang berbeda pula (Sani et al., 2018).

Clustering merupakan salah satu jenis *unsupervised learning* yang sering digunakan dalam *data mining*. Tujuan dari *clustering* adalah untuk mengelompokkan data ke dalam kumpulan yang terbatas berdasarkan kesamaan karakteristik secara semantik dengan menggunakan perhitungan kemiripan (Kim et al., 2019).

2.3 DBSCAN

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) adalah salah satu contoh metode *clustering* berdasarkan kepadatan yang paling populer dalam *data mining*. Algoritma DBSCAN digunakan untuk mengelompokkan data berdasarkan jumlah minimum data dalam radius tertentu, yang disebut epsilon (ϵ), untuk menentukan apakah data tersebut termasuk dalam kategori kepadatan yang diinginkan. Konsep kepadatan ini menghasilkan tiga status data, yaitu inti (*core*), batas (*border*), dan *noise*. Sebuah data akan dikategorikan sebagai inti jika jumlah data tetangga dan dirinya sendiri dalam radius ϵ berjumlah lebih besar atau sama dengan nilai *Minimum Points* (*MinPts*)

yang ditentukan. Sedangkan, data yang memiliki tetangga yang termasuk inti namun jumlah tetangga dan dirinya sendiri dalam radius ϵ kurang dari *MinPts* akan dianggap sebagai batas. Sedangkan, data yang tidak memenuhi kriteria tersebut dikategorikan sebagai *noise* (Safitri et al., 2017).

Kelebihan dari DBSCAN antara lain:

1. DBSCAN tidak memerlukan informasi sebelumnya tentang jumlah kelompok yang akan dibentuk.
2. DBSCAN mampu mengidentifikasi kelompok dengan bentuk yang bervariasi, bahkan dapat mengidentifikasi kelompok yang sepenuhnya terisolasi (namun tidak terhubung) dalam cluster yang berbeda.
3. DBSCAN juga dapat menemukan data yang tidak masuk ke dalam kelompok tertentu (*noise*).
4. Hanya dua parameter yang diperlukan oleh DBSCAN dan urutan titik dalam basis data tidak terlalu berpengaruh terhadapnya (Safitri et al., 2017).

Berikut ini adalah prinsip-prinsip dasar dari metode *density based clustering* (Silitonga, 2016):

1. *Neighborhood* yang terletak didalam radius disebut ϵ -*neighborhood* dari objek data.
2. Apabila ϵ -*neighbourhood* dari suatu objek mengandung minimal jumlah objek yang ditetapkan, yaitu *MinPts*, maka objek tersebut disebut sebagai objek inti (*core object*).
3. Suatu objek *p* merupakan *density reachable* dari objek *q* apabila memenuhi ϵ dan *MinPts* dalam suatu set objek *D* jika terdapat suatu rantai

objek p_1, p_2, \dots, p_n , dimana $p_1 = q$ dan $p_n = p$, dimana p_{i+1} *density reachable* secara langsung dari p_i yang memenuhi ϵ dan *MinPts*, untuk $1 \leq i \leq n$, p_i anggota D .

4. Sebuah objek p dikatakan *density connected* dengan objek q jika objek p dan q memenuhi kriteria ϵ dan *MinPts* di dalam himpunan objek D , dan terdapat objek o di dalam himpunan D dimana kedua p dan q adalah *density reachable* dari o yang memenuhi ϵ dan *MinPts*.

Metode clustering DBSCAN memproses pembentukan cluster-cluster dengan mengikuti langkah-langkah berikut (Silitonga, 2016):

1. *DBSCAN* akan memeriksa semua titik data dalam *database* dan menelusuri *cluster-cluster* dengan mengevaluasi wilayah tetangga ϵ . Jika wilayah tetangga ϵ dari suatu titik data p memiliki jumlah titik data yang lebih besar dari *MinPts*, maka sebuah *cluster* baru akan terbentuk dengan p sebagai titik pusat (*core object*).
2. Selanjutnya, *DBSCAN* akan mengumpulkan secara iteratif objek-objek yang dapat dicapai dengan kepadatan (*density reachable*) langsung dari titik pusat tersebut. Proses ini mungkin melibatkan penggabungan beberapa *cluster-cluster* yang dapat dicapai dengan kepadatan.

Berikut adalah cara komputasi Algoritma *DBSCAN* (Devi et al., 2015):

1. Lakukan inisialisasi parameter *MinPts* dan ϵ
2. Pilih titik awal atau p secara acak
3. Hitung semua jarak titik yang terhubung dengan p menggunakan persamaan *euclidean distance*, yaitu

$$Distance = \sqrt{(X_i - X_c)^2 + (Y_i - Y_c)^2} \quad (2.1)$$

Keterangan:

X_i dan Y_i : Koordinat dari titik yang sedang diproses.

X_c dan Y_c : Koordinat dari titik awal atau pusat *cluster*

4. Jika jumlah titik yang memiliki jarak kurang dari atau sama dengan ϵ dengan titik p lebih dari *MinPts*, maka titik p dianggap sebagai titik inti (*core point*) dan membentuk sebuah kelompok (*cluster*)
5. Lakukan langkah 3–4 secara berulang hingga semua titik diproses

2.4 Cluster Validity

Agar hasil *clustering* sesuai dengan populasi dan akurat, perlu dilakukan validasi. Salah satu cara validasi yang bisa dilakukan adalah dengan membagi data awal menjadi dua kelompok secara acak, lalu melakukan *clustering* pada setiap kelompok dan membandingkan profilnya menggunakan metode penjumlahan. Hal ini dilakukan untuk menghindari hasil yang tidak cocok dan memastikan hasil *clustering* merepresentasikan populasi dengan baik (Larose & Larose, 2014).

Silhouette Score adalah salah satu metode untuk menguji validitas hasil *clustering*. *Silhouette Score* dapat diperoleh dengan cara (Sri Fastaf & Yamasari, 2022):

1. Cari nilai rata-rata jarak antara anggota *cluster* i dan semua anggota lainnya di dalam *cluster* yang sama seperti pada persamaan (2.2)

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad (2.2)$$

Dengan j merupakan anggota lain dari *cluster* A, dan $d(i,j)$ merupakan jarak antara anggota i dan j .

2. Cari nilai jarak rata-rata antara anggota cluster i dengan semua anggota di cluster lain, dan ambil nilai minimumnya.

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (2.3)$$

Di mana $d(i,C)$ merupakan jarak rata-rata antara i dengan semua anggota pada *cluster* lain yaitu C, dan $A \neq C$. Sehingga diperoleh seperti persamaan (2.3):

$$b(i) = \min_{C \neq A} d(i, j) \quad (2.4)$$

3. Setelah memperoleh nilai $a(i)$ dan $b(i)$, *Silhouette Score* didapatkan dengan rumus seperti persamaan (2.5):

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.5)$$

Keterangan :

$a(i)$: Rata-rata jarak i terhadap semua objek di klaster A

$b(i)$: Rata-rata jarak i terhadap semua objek pada klaster lain

Nilai silhouette berada pada interval $-1 \leq s(i) \leq 1$.

2.5 Hujan

Curah hujan (mm) adalah ketinggian air hujan yang jatuh pada permukaan datar tanpa ada pengaruh penguapan, infiltrasi, atau aliran. Satu mm curah hujan dapat diartikan sebagai jumlah air hujan sebanyak satu mm yang menyebar pada luas permukaan datar seluas satu meter persegi, dengan asumsi tidak ada

penguapan, infiltrasi, atau aliran.

Pada wilayah tropik, curah hujan yang tinggi disetiap tahun dapat ditemukan di kepulauan maritim Indonesia, terutama di daerah pegunungan. Curah hujan yang tinggi terjadi karena adanya konveksi dan pembentukan awan hujan panas. Gerakan massa udara lembab ke atas adalah faktor utama yang menyebabkan terjadinya curah hujan. Gerakan massa udara lembab yang bergerak ke atas disebabkan oleh keadaan atmosfer yang tidak stabil, dimana kondisi ini terjadi ketika udara yang naik mengandung kelembaban dan tingkat penurunan suhu udara lingkungan berada di antara tingkat penurunan suhu adiabatik kering dan adiabatik jenuh.

Kondisi kelembaban mempengaruhi kestabilan udara, sehingga pola curah hujan seperti jumlah, intensitas, durasi, frekuensi, dan distribusinya dapat sangat berbeda-beda baik dalam ruang maupun waktu. Di wilayah tropis, intensitas curah hujan biasanya tinggi karena proses konveksi. Curah hujan yang diterima di Indonesia bisa berbeda-beda antara 8 hingga 37 %, dengan rata-rata sekitar 22 %. Sedangkan di Bavaria, Jerman, nilai tertinggi hanya sebesar 3.7%. Lebih dari 80% dari curah hujan di Bogor terjadi saat curah hujan minimal mencapai 20 mm (Mulyono, 2016).

2.6 Hujan Dalam Al Qur'an

Salah satu cara untuk berbuat baik kepada makhluk non-hidup adalah dengan memanfaatkan air secara optimal, karena air adalah sumber kehidupan yang mendasar. Pada bagian tertentu dalam ayat suci Surah An-Nur Ayat 45, Allah berfirman:

”Dan Allah telah menciptakan semua hewan dan air.”

Ibnu Katsir dalam penafsirannya menjelaskan bahwa dengan menciptakan beragam jenis makhluk yang berasal dari unsur yang sama, yaitu air, Allah menunjukkan kekuasaan-Nya yang luar biasa dan kerajaan-Nya yang sangat besar, yang tercermin pada perbedaan bentuk, warna, rupa, dan gerak-gerik yang dimiliki oleh makhluk-makhluk tersebut. Dalam ayat lain yaitu Surah Al-Anbiya Ayat 30 disebutkan:

“Dan apakah orang-orang kafir tidak mengetahui bahwasanya langit dan bumi itu keduanya dahulu adalah suatu yang padu, kemudian kami pisahkan antara keduanya. Dan dari air kami jadikan segala sesuatu yang hidup. Maka mengapakah mereka tiada juga beriman?”

Yang mana dijelaskan dalam tafsir Ibnu Katsir yaitu, dalam firman-Nya, Allah SWT memberikan peringatan mengenai kekuasaan-Nya yang tak terbatas dan kerajaan-Nya yang besar. yaitu orang-orang yang tidak beriman, yang menyembah selain Allah dan mengingkari ke-Esaan-Nya, apakah mereka tidak menyadari bahwa Allah adalah satu-satunya Rabb yang Maha Esa dalam penciptaan dan penataan? Bagaimana mungkin mereka menyembah atau menyekutukan-Nya dengan yang lain? Mereka juga tidak menyadari bahwa langit dan bumi dulunya bersatu dan kemudian terpisah dan terbentuk menjadi tujuh langit dan tujuh lapisan bumi, dipisahkan oleh udara dan memungkinkan hujan turun dan tanaman tumbuh. Allah telah menciptakan segala sesuatu dari air, yang menjadi bukti nyata akan kebesaran-Nya. Namun, mengapa mereka masih tidak beriman? (Abdul Ghofar, 2003)

Sejak zaman dahulu, Al-Quran telah menekankan bahwa air merupakan sumber kehidupan, bahkan sebelum ilmu biologi modern berkembang. Manusia sejak dahulu telah menyadari pentingnya air sebagai faktor utama yang

mempengaruhi pemukiman dan kemajuan peradaban mereka. Keberadaan tanda-tanda kehidupan selalu dikaitkan dengan keberadaan sumber air (Saputro, 2016).

Dalam kesempatan lain, Al-Quran Surah Al-Anfal Ayat 11 juga menunjukkan betapa pentingnya air hujan.

“(Ingatlah) ketika Allah membuat kamu mengantuk untuk memberikan ketenteraman dari-Nya, dan Allah menurunkan air (hujan) dari langit kepadamu untuk menyucikan kamu dengan (hujan) itu dan menghilangkan gangguan-gangguan setan dari dirimu dan untuk menguatkan hatimu serta memperteguh telapak kakimu (teguh pendirian).”

Buya Hamka menjelaskan kata hujan dalam surat al anfal pada ayat 11 ialah, hujan yang turun mengakibatkan sumur-sumur terisi, tempat penampungan air menjadi penuh, dan pasir yang biasanya mudah menempel pada kaki ketika perjalanan menjadi keras sehingga lebih mudah dipijak. Buya Hamka juga memaparkan beberapa keuntungan yang dapat diraih oleh para hamba Allah ketika turun hujan. Yang pertama, mereka dapat membersihkan diri dan pikiran mereka. Kedua, semua kotoran yang mungkin berasal dari pengaruh setan akan hilang, karena kurangnya air bisa menyebabkan pengaruh setan muncul di hati. Ketiga, kebahagiaan karena air membuat orang-orang menjadi bersatu dalam sukacita. Keempat, kondisi bumi yang keras saat diinjak akan membuat hati orang menjadi lebih bulat dan lebih siap menghadapi musuh (Khairullah et al., 2023).

Air hujan memainkan peran penting dalam perkembangan tumbuhan, yang kemudian menjadi sumber pangan bagi hewan dan manusia. Nabi Muhammad SAW bersabda: “Tidak ada seorang pun yang tahu kapan hujan turun kecuali hanya Allah”.

BAB III

METODE PENELITIAN

3.1 Pendekatan Penelitian

Penelitian ini menggunakan pendekatan eksperimen deskriptif kuantitatif. Pendekatan eksperimen dilakukan dengan mengontrol dan memodifikasi variabel bebas, yaitu nilai *Epsilon* dan *MinimumPoints*, kemudian mengamati variabel terikatnya, yaitu jumlah *cluster* dan *Outliers/Noise*. Sementara itu, pendekatan deskriptif kuantitatif digunakan untuk menganalisis data dan mengatur data yang sudah ada sesuai dengan kebutuhan penelitian.

3.2 Data dan Sumber Data

Dalam penelitian ini, data yang dipakai berasal dari Badan Meteorologi Klimatologi dan Geofisika (BMKG) Stasiun Klimatologi Kelas II Jawa Timur – Karangploso, Malang dan berfokus pada data curah hujan di wilayah Sumenep selama tahun 2022. Data curah hujan tahun 2022 berjumlah 204 dengan 3 atribut yaitu *longitude*, *latitude*, dan curah hujan.

3.3 Variabel Penelitian

Dalam penelitian ini, ada dua jenis variabel yaitu, variabel bebas yang terdiri dari nilai *Epsilon* (ϵ) dan *Minimum Points* (*MinPts*), dan variabel terikat yang terdiri dari jumlah klaster (*n_cluster*), jumlah *outlier* (*n_outlier*), dan nilai *Silhouette Score*.

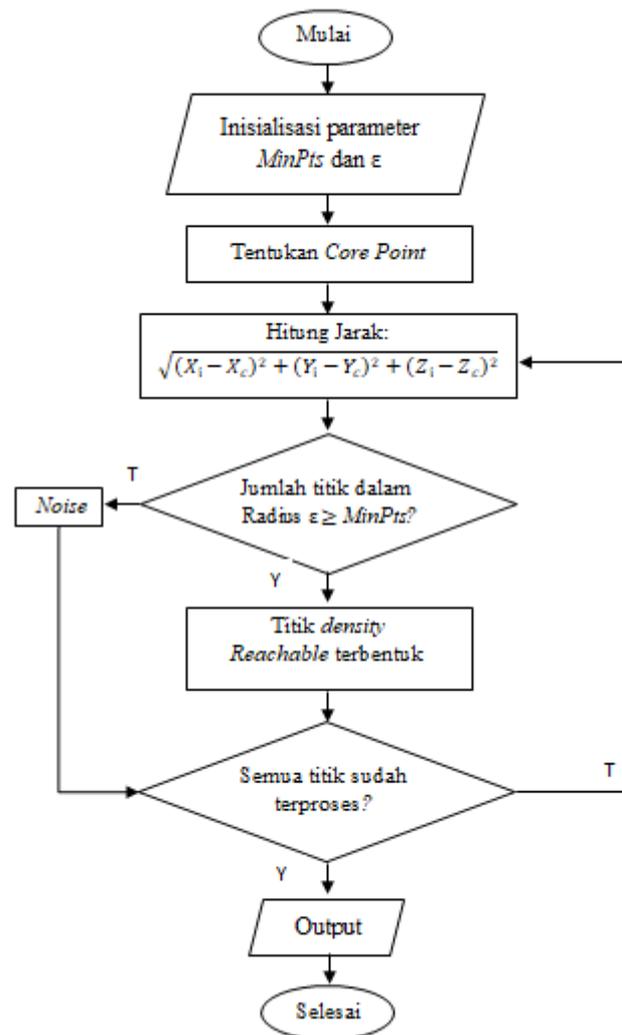
3.4 Tahapan Penelitian

1. Analisis deskriptif statistik data untuk mengetahui gambaran awal persebaran hujan di wilayah Sumenep.

Analisis deskriptif statistik data untuk mengetahui gambaran awal persebaran hujan di wilayah Sumenep dilakukan untuk memperoleh informasi dasar tentang data hujan seperti rata-rata, median, standar deviasi, dan sebagainya. Tujuan dari analisis ini adalah untuk memahami karakteristik data hujan di wilayah Sumenep.

2. Melakukan eksperimen *DBSCAN Clustering* dengan *Python*.

Eksperimen *DBSCAN Clustering* dengan *Python* dilakukan untuk melakukan pengelompokan data hujan menjadi beberapa kelompok atau cluster berdasarkan kemiripan karakteristiknya. *DBSCAN* (*Density-Based Spatial Clustering of Applications with Noise*) adalah algoritma *clustering* yang membagi data ke dalam kelompok berdasarkan kepadatan dan jarak antara data. Dalam eksperimen ini, penggunaan bahasa pemrograman Python membantu dalam memproses data hujan secara efisien dan menghasilkan kelompok-kelompok yang berbeda. Algoritma *DBSCAN Clustering* adalah seperti pada gambar 3.1.



Gambar 3.1 Flowchart DBSCAN Clustering

3. Evaluasi Cluster

Evaluasi Cluster adalah tahapan untuk mengevaluasi kualitas dan kegunaan dari hasil *DBSCAN Clustering*. Evaluasi dilakukan dengan menguji sejauh mana hasil *clustering* yang dihasilkan sesuai dengan data asli dan juga mengukur seberapa baik *cluster* yang dihasilkan dapat membantu dalam mengambil keputusan atau menarik kesimpulan dari data hujan di wilayah Sumenep. Metode evaluasi *clustering* yang digunakan adalah *Silhouette Score*. Evaluasi yang baik dapat membantu

meningkatkan keakuratan hasil clustering dan memastikan bahwa hasil clustering dapat dipercaya dan berguna bagi pengambil keputusan.

4. Menganalisis data dari hasil *DBSCAN Clustering*.

Analisis data dari hasil *DBSCAN Clustering* dilakukan untuk mengetahui informasi lebih lanjut tentang setiap kelompok data hujan yang terbentuk. Hal ini dapat membantu dalam mengidentifikasi faktor-faktor yang mempengaruhi karakteristik hujan di wilayah Sumenep serta memberikan informasi yang berguna untuk pengambilan keputusan dan perencanaan di masa depan.

BAB IV

PEMBAHASAN

4.1 Deskripsi Data

Sebelum melakukan proses *mining*, dilakukan deskripsi data terlebih dahulu untuk melihat gambaran umum faktor yang mempengaruhi persebaran hujan di wilayah Sumenep.

1) Rata-rata (*Mean*)

Rata-rata untuk data *Longitude*:

$$\begin{aligned}\tilde{x} &= \frac{\sum_{i=1}^n x_i}{n} \\ \tilde{x} &= \frac{113,7425 + 113,997777777777 + \dots + 113,564}{204} \\ &= \frac{23195,39}{204} \\ &= 113,7028921569\end{aligned}$$

Rata-rata untuk data *Latitude*:

$$\begin{aligned}\tilde{y} &= \frac{\sum_{i=1}^n y_i}{n} \\ \tilde{y} &= \frac{-6,8927778 + (-6,9466667) + \dots + (-7,02001)}{204} \\ &= \frac{-1424,9}{204} \\ &= -6,9848039216\end{aligned}$$

Rata-rata untuk data Curah Hujan:

$$\tilde{z} = \frac{\sum_{i=1}^n z_i}{n}$$

$$\begin{aligned}\tilde{z} &= \frac{367 + 263 + \dots + 90}{204} \\ &= \frac{38777}{204} \\ &= 190,0833333333\end{aligned}$$

2) Ragam (*Variance*)Ragam untuk data *Longitude*:

$$\begin{aligned}S^2 &= \frac{\sum_{i=1}^n (x_i - \tilde{x})^2}{n - 1} \\ S^2 &= \frac{(113,7425 - 113,7029)^2 + (113,9978 - 113,7029)^2 + \dots + (113,564 - 113,7029)^2}{203} \\ &= \frac{17,211730956}{203} \\ &= 0,084786852\end{aligned}$$

Ragam untuk data *Latitude*:

$$\begin{aligned}S^2 &= \frac{\sum_{i=1}^n (y_i - \tilde{y})^2}{n - 1} \\ S^2 &= \frac{(-6,8928 - (-6,9848))^2 + (-6,9467 - (-6,9848))^2 + \dots + (-7,02001 - (-6,9848))^2}{203} \\ &= \frac{2,079003591}{203} \\ &= 0,010241397\end{aligned}$$

Ragam untuk data *Curah Hujan*:

$$\begin{aligned}S^2 &= \frac{\sum_{i=1}^n (z_i - \tilde{z})^2}{n - 1} \\ S^2 &= \frac{(367 - 190,0833)^2 + (263 - 190,0833)^2 + \dots + (90 - 190,0833)^2}{203} \\ &= \frac{3425429,58391}{203} \\ &= 16874,52997\end{aligned}$$

3) Simpangan Baku (*Standart Deviation*)

Simpangan baku untuk data *Longitude*:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \text{ atau } S = \sqrt{S^2}$$

Karena S^2 dari data *Longitude* telah diketahui melalui perhitungan sebelumnya, maka:

$$S = \sqrt{0,084786852}$$

$$S = 0,2911818195$$

Simpangan baku untuk data *Latitude*:

$$S = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}} \text{ atau } S = \sqrt{S^2}$$

Karena S^2 dari data *Longitude* telah diketahui melalui perhitungan sebelumnya, maka:

$$S = \sqrt{0,010241397}$$

$$S = 0,1011997875$$

Simpangan baku untuk data Curah Hujan:

$$S = \sqrt{\frac{\sum_{i=1}^n (z_i - \bar{z})^2}{n-1}} \text{ atau } S = \sqrt{S^2}$$

Karena S^2 dari data Curah Hujan telah diketahui melalui perhitungan sebelumnya, maka:

$$S = \sqrt{16874,52997}$$

$$S = 129,9020014088$$

Perhitungan deskriptif data juga diperoleh dengan menggunakan program *SPSS* seperti pda Tabel 4.1

Tabel 4.1 Deskriptif Data Hujan di Sumenep Tahun 2022

Descriptive Statistics							
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic
LON	204	1,83083	113,0116	114,8425	113,70289	0,2912	0,085
LAN	204	0,83399	-7,396	-6,56201	-6,98479	0,1012	0,010
Curah Hujan	204	589	2	589	190,08	129,902	16874,530

Pada Tabel 4.1 Didapatkan hasil pada Tahun 2022, di wilayah Sumenep terjadi hujan sejumlah 204 yang diwakili oleh sejumlah koordinat *Longitude* dan *Latitude* dalam jangkauan *Longitude* sebesar 1,83083 derajat dan jangkauan *Latitude* sebesar 0,83399 derajat. Standard deviasi sebesar 0,2912 pada *Longitude* dan 0,1012 pada *Latitude* menunjukkan bahwa persebaran datanya sangatlah kecil, yang berarti hujan satu dengan yang lainnya jaraknya berdekatan. Dalam hal varians, data *Longitude* memiliki nilai varians sebesar 0,085 dan data *Latitude* memiliki nilai varians sebesar 0,010. Hal ini menunjukkan bahwa keragaman data keduanya sangatlah kecil, hal ini dapat dikatakan bahwa koordinat hujan satu dengan yang lainnya memiliki perbedaan yang sangat kecil.

4.2 Perhitungan *DBSCAN Clustering*

1. Inisialisasi *core point* menggunakan data koordinat yang pertama, yaitu koordinat (113,7425, -6,892778, 367).
2. Pembentukan radius ϵ di sekitar *core point*. Misalkan, radius ϵ yang digunakan adalah 10 dan *MinPts* adalah 6

3. Perhitungan jarak dari *core point* dengan semua titik lain. Untuk perhitungan jarak tersebut, digunakan rumus *Euclidean Distance* sebagai berikut:

$$\sqrt{(X_i - X_c)^2 + (Y_i - Y_c)^2 + (Z_i - Z_c)^2}$$

Dengan X_i adalah koordinat *Longitude* pada data ke-i, X_c adalah koordinat *Longitude* pada data yang dijadikan *core point*, Y_i adalah koordinat *Latitude* pada data ke-i, Y_c adalah koordinat *Latitude* pada data yang dijadikan *core point*, Z_i adalah koordinat Curah Hujan pada data ke-i, dan Z_c adalah koordinat *Latitude* pada data yang dijadikan *core point*.

Untuk perhitungan jarak dengan *core point* yang pertama adalah sebagai berikut:

- 1) Jarak titik kedua pada data dengan titik *core point*:

$$\begin{aligned} & \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2 + (Z_2 - Z_1)^2} \\ & \sqrt{(113,9589 - 113,7425)^2 + (-6,9975 - (-6,892778))^2 + (361 - 367)^2} \\ & \sqrt{(0,2164)^2 + (-0,104722)^2 + (-6)^2} \\ & \sqrt{0,04682896 + 0,0109666973 + 36} = 6,0048143733 \end{aligned}$$

- 2) Jarak titik ketiga pada data dengan titik *core point*:

$$\begin{aligned} & \sqrt{(X_3 - X_1)^2 + (Y_3 - Y_1)^2 + (Z_3 - Z_1)^2} \\ & \sqrt{(113,8686 - 113,7425)^2 + (-7,00861 - (-6,892778))^2 + (364 - 367)^2} \\ & \sqrt{(0,1261)^2 + (-0,115832)^2 + (-3)^2} \\ & \sqrt{0,01590121 + 0,0134170522 + 9} = 3,0048824041 \end{aligned}$$

- 3) Penelusuran dilakukan hingga semua titik telah dihitung jaraknya dengan *core point*.

4. Setelah diperoleh semua jaraknya, dipilih titik-titik yang *density reachable* atau yang jarak dari titik tersebut menuju *core point* $\leq \epsilon$ dan bandingkan titik-titik yang *density reachable* $\geq MinPts$.
5. Setelah diperoleh titik-titik yang *density reachable* terhadap *core point* pertama, dilakukan tahap 2 hingga 4 untuk *core point* selanjutnya. Penentuan *core point* yang selanjutnya dipilih dari titik-titik yang *density reachable* terhadap *core point* pertama.
6. Setelah semua titik *density reachable* sudah terproses, maka akan terlihat *noise* dari data tersebut. *Noise* dalam kasus ini adalah koordinat hujan yang tidak biasa daripada titik yang lain, artinya titik *noise* ini berada jauh diluar jangkauan ϵ .
7. Setelah semua titik telah terproses, maka akan terlihat *cluster* yang terbentuk beserta *noise* nya. Keanggotaan *cluster* terbentuk dari titik-titik yang *density reachable*, sedangkan *noise* terbentuk dari titik-titik yang bukan termasuk *density reachable*.

4.3 Hasil Eksperimen DBSCAN Clustering

Pada uji coba yang dilakukan menggunakan ϵ 3 sampai 12 dan *MinPts* 3 sampai 7 sehingga ada 50 kali uji coba *clustering*. Pada setiap uji coba dilihat banyak *cluster* yang terbentuk ($n_cluster$), *noise* (n_noise), dan dihitung *Silhouette Score*. Dari hasil berbagai eksperimen tersebut, *Silhouette Score* terbaik dipilih, yang ditentukan oleh nilai yang mendekati 1. Hasilnya menunjukkan bahwa nilai optimum untuk ϵ adalah 10 dan *MinPts* yang optimal adalah 6, menghasilkan *Silhouette Score* sebesar 0,521. Dan diketahui juga untuk nilai

Silhouette Score adalah -0,25 pada percobaan dengan menggunakan ϵ adalah 4 dan *MinPts* adalah 6. Hasil pada eksperimen tersebut ditampilkan pada Tabel 4.2.

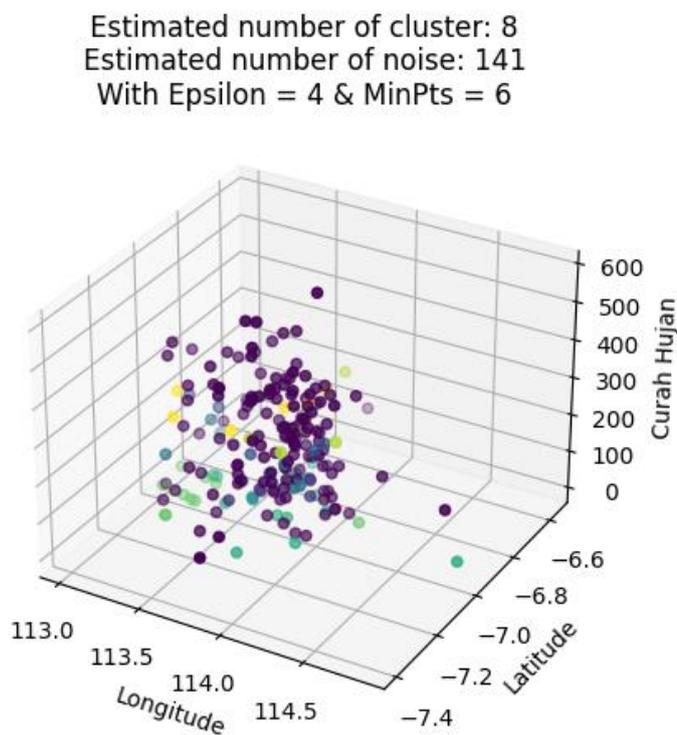
Tabel 4.2 Eksperimen *DBSCAN Clustering*

$\epsilon \backslash$ <i>MinPts</i>	3		4		5		6		7	
3	27	53	22	84	10	142	3	179	2	186
	0,276		0,077		-0,273		0,001		-0,059	
4	21	35	18	51	17	77	8	141	3	178
	0,392		0,312		0,134		-0,25		0,003	
5	15	21	13	34	15	40	16	56	10	118
	0,436		0,384		0,364		0,27		0,27	
6	12	13	12	23	11	25	12	44	13	56
	0,315		0,393		0,384		0,345		0,291	
7	9	13	7	19	7	19	7	30	13	34
	0,382		0,363		0,363		0,41		0,355	
8	7	10	6	13	6	15	6	25	8	32
	0,327		0,354		0,316		0,388		0,371	
9	3	10	2	13	2	13	5	14	6	17
	0,38		0,423		0,423		0,411		0,46	
10	2	8	1	11	1	11	3	13	5	16
	0,289		0,46		0,46		0,521		0,413	
11	2	8	1	11	1	11	1	11	5	11
	0,289		0,46		0,46		0,46		0,358	
12	2	8	1	11	1	11	1	11	3	11
	0,289		0,46		0,46		0,46		0,226	

Ket:

	: <i>n_cluster</i>
	: <i>n_noise</i>
	: <i>Silhouette Score</i>

Setelah semua data sudah membentuk *cluster* dan menemukan *noise*, maka langkah terakhir adalah melakukan visualisasi data hasil *DBSCAN Clustering* menggunakan *Python* seperti pada Gambar 4.1 dan Gambar 4.2.

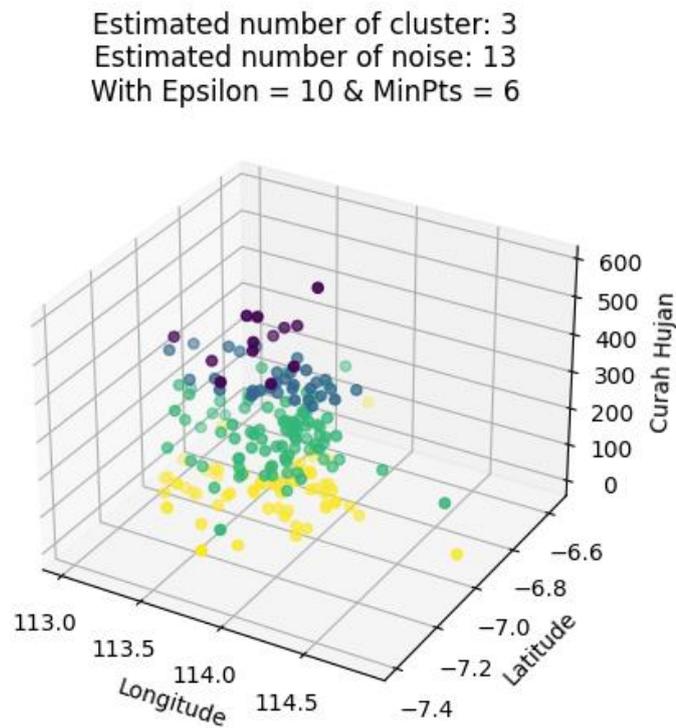


Gambar 4.1 Hasil *DBSCAN Clustering* Dengan $\epsilon = 4$ dan *MinPts* = 6

Tabel 4.3 Tabel Hasil Eksperimen Dengan $\epsilon = 4$ dan *MinPts* = 6

ϵ	<i>MinPts</i>	Hasil		
		<i>n_cluster</i>	<i>n_noise</i>	<i>Silhouette Score</i>
4	6	8	141	-0,25

Gambar 4.1 dan Tabel 4.3 menjelaskan bahwasanya hasil *DBSCAN Clustering* dengan $\epsilon = 4$ dan *MinPts* = 6 merukan hasil yang terburuk dikarenakan memiliki nilai *Silhouette Score* = -0,25 dengan menghasilkan jumlah *cluster* 8 dan *noise* 141.



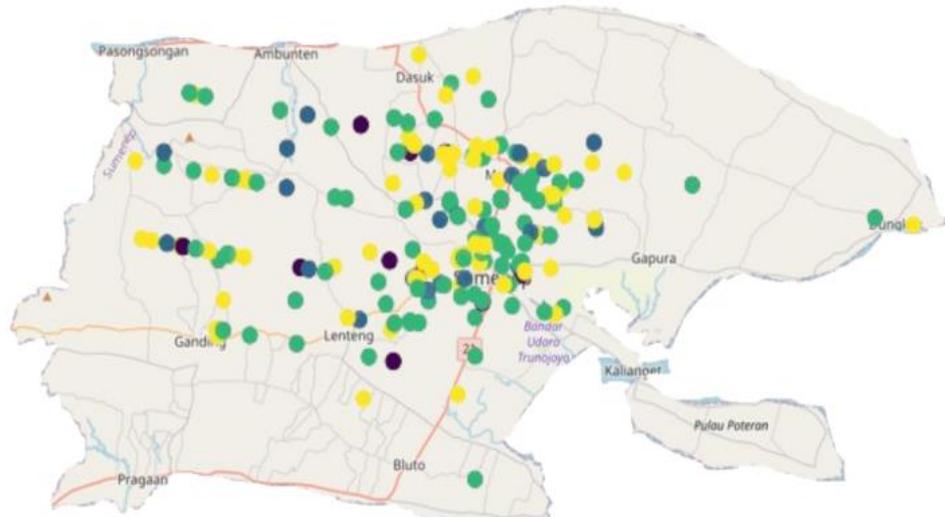
Gambar 4.2 Hasil *DBSCAN Clustering* Dengan $\epsilon = 10$ dan *MinPts* = 6

Tabel 4.4 Tabel Hasil Eksperimen Dengan $\epsilon = 10$ dan *MinPts* = 6

ϵ	<i>MinPts</i>	Hasil		
		<i>n_cluster</i>	<i>n_noise</i>	<i>Silhouette Score</i>
10	6	3	13	0,521

Gambar 4.2 dan Tabel 4.4 menjelaskan bahwasanya hasil *DBSCAN Clustering* dengan $\epsilon = 10$ dan *MinPts* = 6 merukan hasil yang terbaik dikarenakan memiliki nilai *Silhouette Score* = 0,521 dengan menghasilkan jumlah *cluster* 8 dan *noise* 141.

4.4 Analisis Pola Penyebaran



Gambar 4.3 Persebaran Hujan Di Sumenep Pada Tahun 2022

Ket:



Gambar 4.3 menjelaskan persebaran hujan di Sumenep pada tahun 2022 yang mana titik biru pada gambar adalah *cluster 1*, titik hijau adalah *cluster 2*, titik kuning adalah *cluster 3*, dan titik ungu merupakan *noise*. Peta pada penyebaran ini hanya sebagian dari wilayah sumenep yaitu untuk wilayah kepulauan seperti Sepudi, Kangean, Sapeken, dan lainnya tidak termasuk dikarenakan data yang terbatas.

Setelah model hasil *clustering* terbentuk, langkah terakhir adalah melakukan analisis berdasarkan hasil dari *DBSCAN Clustering*. Hasil yang dianalisis adalah model terbaik yaitu dengan nilai *Silhouette Score* yang paling mendekati satu yaitu hasil *DBSCAN Clustering* dengan menggunakan $\epsilon = 10$ dan

$MinPts = 6$ menghasilkan nilai *Silhouette Score* sebesar 0,521. Analisis dari hasil tersebut ditampilkan pada tabel berikut:

Tabel 4.5 Analisis Hasil *DBSCAN Clustering*

	<i>Membership</i>	Bulan Paling Sering Terjadi Hujan
Cluster 1	32	Januari dan Februari
Cluster 2	93	Oktober dan November
Cluster 3	66	Juli dan Agustus
Noise	13	Januari

Berdasarkan Tabel 4.3 terlihat bahwa hujan sering terjadi pada bulan Januari dan Februari dan terletak pada koordinat yang tergabung pada *cluster* pertama. Hujan juga sering terjadi pada bulan Oktober dan November yang terletak pada koordinat yang tergabung pada *cluster* kedua. Pada bulan Juli dan Agustus hujan juga sering terjadi pada koordinat yang tergabung pada *cluster* ketiga. *Noise* kejadian hujan di sepanjang Tahun 2022 banyak terjadi di bulan Januari.

4.5 Konsep *DBSCAN Clustering* Dalam Penyebaran Agama Islam

Pada dasarnya, konsep dari *DBSCAN Clustering* adalah mengelompokkan suatu himpunan menjadi sub-sub himpunan yang memiliki karakteristik sama dalam satu sub himpunan, namun memiliki perbedaan karakteristik antar sub himpunan yang satu dengan sub himpunan yang lainnya. Penentuan anggota dari sub himpunan tersebut ditelusuri melalui kedekatan dari satu objek dengan objek yang lainnya dalam jangkauan tertentu (ϵ) dan memiliki syarat, yaitu jumlah minimum anggota yang harus terpenuhi agar dapat masuk ke dalam suatu sub himpunan (*MinPts*).

Dalam menyebarkan agama Islam, Nabi Muhammad secara tidak langsung menerapkan konsep dari *DBSCAN Clustering*. Dalam berdakwah, Nabi Muhammad selalu memulai jihadnya dengan menyebarkan Islam di lingkungan terdekatnya terlebih dahulu, terutama pada keluarga dan para sahabatnya (*density reachable*). Kemudian dari mereka inilah yang juga ikut menyebarkan dakwah Islam yang telah diajarkan oleh Nabi Muhammad kepada orang-orang terdekatnya. Hal ini terus menerus menyebar hingga sekarang. Sedangkan orang-orang kafir yang belum mendapat hidayah untuk memeluk Islam kemungkinan adalah orang-orang yang jauh dengan Nabi Muhammad maupun umatnya (*noise*). Salah satu upaya agar mereka dapat mendapat hidayah sehingga masuk Islam adalah dengan merangkul orang kafir tersebut (*noise*) dengan orang-orang terdekat kita (*density reachable*) sekaligus memberikan pemahaman tentang Islam, tidak menutup kemungkinan orang kafir tersebut (*noise*) mendapat hidayah sehingga dapat masuk ke dalam *cluster* kita, yaitu Islam.

Penelusuran menggunakan *DBSCAN Clustering* masih dapat memungkinkan suatu *noise* bisa menjadi anggota suatu *cluster* dengan beberapa cara. Cara yang pertama adalah dengan merubah ϵ dan *MinPts* nya. Jika dikaitkan dengan dakwah Islam, merubah hal tersebut sama halnya dengan merubah sistem dalam berdakwah. Cara yang kedua adalah dengan mendekatkan titik yang *density reachable* dengan *noise*. Hal ini dapat menjadi pisau bermata dua, kemungkinan pertama adalah *noise* akan dapat bergabung dengan *cluster* jika titik *density reachable* memenuhi *MinPts*, kemungkinan kedua adalah titik yang berawal dari *density reachable* akan tergabung dalam *noise* karena titik *density reachable* yang lain tidak berada di sekitarnya. Oleh sebab itu, Nabi Muhammad dalam

berdakwah selalu bersama dengan orang-orang terdekatnya yang sangat beliau percaya.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan rumusan masalah beserta pembahasan di atas, dapat disimpulkan bahwa:

1. Hasil *DBSCAN Clustering* untuk data hujan di wilayah Sumenep pada Tahun 2022 merujuk pada Gambar 4.1. Dari perhitungan *DBSCAN Clustering*, didapatkan hasil, yaitu terdapat 3 *cluster* dan 13 titik *noise*, dengan anggota *cluster* pertama berjumlah 32 titik hujan, *cluster* kedua berjumlah 93 titik hujan yang menyebar hampir di seluruh wilayah Sumenep, *cluster* ketiga berjumlah 66 titik.

Analisis dari hasil *DBSCAN Clustering* tersebut adalah sebagai berikut:

- a) Hujan yang tergolong pada *cluster* pertama adalah lokasi dengan jumlah air hujan yang besar, yaitu dengan curah hujan lebih dari 300 mm.
- b) Hujan yang tergolong pada *cluster* kedua adalah lokasi dengan jumlah air hujan sedang, yaitu dengan curah hujan sekitar 200-300 mm dan menyebar hampir di seluruh wilayah Sumenep.
- c) Hujan yang tergolong pada *cluster* ketiga adalah lokasi dengan jumlah air hujan kecil, yaitu dengan curah hujan kurang dari 100 mm.
- d) Hujan yang tergolong *noise* adalah hujan yang sangat jarang terjadi dan jumlah curah hujan yang sangat besar.

2. Evaluasi *cluster* menggunakan *Silhouette Score* dengan parameter $\epsilon = 10$ dan *MinPts* = 6 menghasilkan nilai sebesar 0,521 menunjukkan bahwa hasil *Clustering* relatif baik.

5.2 Saran

Pada skripsi ini, penulis terfokus pada pembuatan *cluster* dengan menggunakan algoritma *DBSCAN* dan analisis persebaran hujannya berdasarkan hasil *clustering* tersebut. Untuk penelitian selanjutnya penulis menyarankan untuk menganalisis hubungan antara kejadian hujan dengan keadaan geografis, terutama lokasi sungai, sawah, dan perkebunan serta melakukan analisis data *spatio temporal* untuk melihat *trend* pada setiap waktu.

DAFTAR PUSTAKA

- Abdul Ghofar, M. (2003). *Tafsir Ibnu Katsir (terj)*. Bogor: Pustaka Imam asy-Syafi'i.
- Aldrian, E. (2003). Simulations of Indonesian Rainfall with a Hierarchy of Climate Models. *Dissertation, Universitas Hamburg, Max-Planck-Institut Für Meteorologie*, 92.
- Devi, N., Putra, I., & Sukarsa, I. (2015). Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan. *Ojs.Unud.Ac.Id*, 6(3), 2088–1541. <https://doi.org/10.24843/LKJITI.2015.v06.i03.p05>
- Eska, J. (2018). *Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5*. <https://doi.org/10.31227/OSF.IO/X6SVC>
- Fahamsah, M. (2020). *Metode DBSCAN Clustering Untuk Analisis Pola Penyebaran Petir di Pasuruan*. 21(1), 1–9.
- Florence, F., Tongkukut, S. H. J., & Klimatologi, M. (2014). *Analisis Spasial Pengaruh Dinamika Suhu Muka Laut Terhadap Distribusi Curah Hujan di Sulawesi Utara*. 3(1), 25–29.
- Khairullah, M. F., Zuhri, A., & Yuzaidi. (2023). *Al-Wasathiyah: Journal of Islamic Studies Pemanfaatan Air Hujan Dalam Perspektif Alquran: Studi terhadap Al- Al-Wasathiyah: Journal of Islamic Studies*. 2, 183–194. <https://doi.org/10.56672/alwasathiyah.v2i2.8>
- Kim, J. H., Choi, J. H., Yoo, K. H., & Nasridinov, A. (2019). AA-DBSCAN: an approximate adaptive DBSCAN for finding clusters with varying densities. *Journal of Supercomputing*, 75(1), 142–169. <https://doi.org/10.1007/s11227-018-2380-z>
- Larose, D., & Larose, C. (2014). *Discovering knowledge in data: an introduction to data mining*.
- Maulana, A., & Fajrin, A. A. (2018). PENERAPAN DATA MINING UNTUK ANALISIS POLA PEMBELIAN KONSUMEN DENGAN ALGORITMA FP-GROWTH PADA DATA TRANSAKSI PENJUALAN SPARE PART MOTOR. *KLIK - KUMPULAN JURNAL ILMU KOMPUTER*, 5(1), 27–36. <https://doi.org/10.20527/KLIK.V5I1.100>
- Maulidani S, S., Ihsan, N., & Sulistyawati. (2015). Analisis Pola Dan Intensitas Curah Hujan Berdasarkan Data Observasi Dan Satelit Tropical Rainfall Measuring Missions (Trmm) 3B42 V7 Di Makassar. *Jurnal Sains Dan Pendidikan Fisika (Jspf)*, 11(1), 99.
- Mimmack, G. M., Mason, S. J., & Galpin, J. S. (2000). *NOTES AND CORRESPONDENCE Choice of Distance Matrices in Cluster Analysis: Defining Regions*.
- Muhandhis, I., Ritonga, A. S., & Murdani, M. H. (2021). Implementasi Metode Inferensi Fuzzy Tsukamoto Untuk Memprediksi Curah Hujan Dasarian Di Sumenep. *Jurnal Ilmiah Edutic : Pendidikan Dan Informatika*, 8(1), 01–10. <https://doi.org/10.21107/edutic.v8i1.8907>
- Mulyono, D. (2016). Analisis Karakteristik Curah Hujan Di Wilayah Kabupaten Garut Selatan. *Jurnal Konstruksi*, 12(1), 1–9. <https://doi.org/10.33364/konstruksi/v.12-1.274>
- Pratama, R. (2016). *Review of Density-Based Spatial Clustering Algorithm*. 204,

3–5.

- Safitri, D., Wuryandari, T., & Rahmawati, R. (2017). Metode dbscan untuk pengelompokan kabupaten/kota di provinsi jawa tengah berdasarkan produksi padi sawah dan padi ladang. *Jurnal.Unimus.Ac.Id*.
- Sani, A., Teknologika, J., & Muningsih, E. (2018). *PENERAPAN METODE K-MEANS CLUSTERING PADA PERUSAHAAN*.
- Saputro, A. D. (2016). Keajaiban Hujan Dalam Prespektif Sains Dan Islam. *LITERASI (Jurnal Ilmu Pendidikan)*, 6(1), 23. [https://doi.org/10.21927/literasi.2015.6\(1\).23-36](https://doi.org/10.21927/literasi.2015.6(1).23-36)
- Silitonga, P. (2016). Analisis pola penyebaran penyakit pasien pengguna badan penyelenggara jaminan sosial (bpjs) kesehatan dengan menggunakan metode dbscan clustering. *Ejournal.Stmik-Time.Ac.Id*.
- Sri Fastaf, C. A., & Yamasari, Y. (2022). Analisa Pemetaan Kriminalitas Kabupaten Bangkalan Menggunakan Metode K-Means dan K-Means++. *Journal of Informatics and Computer Science (JINACS)*. <https://doi.org/10.26740/jinacs.v3n04.p534-546>
- Stubbe, A. A., & Coleman, S. (2014). *A Practical Guide to Data Mining for Business and Industry - Andrea Ahlemeyer-Stubbe, Shirley Coleman - Google Books*. United Kingdom: Wiley.
- Sulastri, H., & Gufroni, A. I. (2017). PENERAPAN DATA MINING DALAM PENGELOMPOKAN PENDERITA THALASSAEMIA. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 3(2), 299–305. <https://doi.org/10.25077/TEKNOSI.V3I2.2017.299-305>

LAMPIRAN

Lampiran 1 Source Code

```

import pandas as pd
from pandas import read_csv
from sklearn.cluster import DBSCAN
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D # Import module for
3D projection
import numpy as np
from sklearn.metrics import silhouette_samples,
silhouette_score

from google.colab import drive
drive.mount ('/content/drive')

data = read_csv("/content/drive/MyDrive/Data Hujan Di
Sumenep.csv")
data.head()

data.info()

fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.scatter(data.iloc[:, 0], data.iloc[:, 1], data.iloc[:, 2],
c='b', marker='o')

ax.set_xlabel('Longitude')
ax.set_ylabel('Latitude')
ax.set_zlabel('Curah Hujan')
ax.set_title('Grafik Data 3D')

plt.show()

dbscan = DBSCAN(eps=10, min_samples=6).fit(data.iloc[:, 0:3])
labels = dbscan.labels_

fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.scatter(data.iloc[:, 0], data.iloc[:, 1], data.iloc[:, 2],
c=labels, marker='o')

ax.set_xlabel('Longitude')
ax.set_ylabel('Latitude')
ax.set_zlabel('Curah Hujan')
ax.set_title('DBSCAN Clustering')

```

```
plt.show()

# Menampilkan jumlah kluster (tidak termasuk noise)
num_clusters = len(set(labels[np.where(labels != -1)]))
print('Jumlah kluster: {}'.format(num_clusters))

# Menampilkan nilai noise
num_noise_points = len(labels[np.where(labels == -1)])
print('Jumlah elemen noise: {}'.format(num_noise_points))

# Menampilkan label dari setiap data point
print('Labels:')
print(labels)

SI = silhouette_score(data.iloc[:, 0:3], labels)
print('Silhouette Score:', SI)
```

RIWAYAT HIDUP



Muhammad Amin Cherid, Lahir di Kabupaten Sumenep, 11 Juni 2000. Penulis merupakan anak kedua dari tiga bersaudara. Putra dari Bapak Syd Idrus dan Ibu Srf Fatimah.

Penulis memulai pendidikan di TK Pertiwi Ambunten Sumenep pada tahun 2005, pendidikan dasarnya ditempuh di SDN Ambunten Timur 1 dan lulus pada tahun 2013, kemudian penulis melanjutkan pendidikan di SMPN 1 Ambunten dan lulus pada tahun 2016, dan setelah itu penulis melanjutkan pendidikan di SMAN 1 Ambunten dan lulus pada tahun 2019.

Pada tahun yang sama penulis terdaftar sebagai Mahasiswa di program studi Matematika, program Strata Satu (S1) di Universitas Negeri Maulana Malik Ibrahim Malang Melalui jalur SNMPTN.



**KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI**

Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

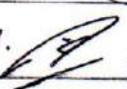
BUKTI KONSULTASI SKRIPSI

Nama : MUHAMMAD AMIN CHERID
NIM : 19610011
Fakultas / Program Studi : Sains dan Teknologi/ Matematika
Judul Skripsi : Metode DBSCAN Clustering Untuk Analisis Pola Penyebaran Hujan Di Sumenep
Pembimbing I : Hisyam Fahmi, M.Kom.
Pembimbing II : Achmad Nashichuddin, M.A.

No	Tanggal	Hal	Tanda Tangan
1.	3 Maret 2023	Konsultasi Bab I dan II	1.
2.	24 Maret 2023	Konsultasi Bab III	2.
3.	27 Maret 2023	Konsultasi Bab I dan Bab II(Kajian Keislaman)	3.
4.	31 Maret 2023	Konsultasi Revisi Bab I dan Bab II(Kajian Keislaman) dan ACC	4.
5.	6 April 2023	Konsultasi Lanjutan Revisi Bab I sampai Bab III dan ACC	5.
6.	6 September 2023	Konsultasi Bab IV	6.
7.	20 September 2023	Konsultasi Program	7.
8.	25 September 2023	Konsultasi Program	8.
9.	19 Oktober 2023	Konsultasi Bab IV dan Bab V	9.
10.	19 Oktober 2023	Konsultasi Bab IV dan Bab V	10.
11.	27 Oktober 2023	Konsultasi Revisi Bab IV dan Bab V	11.
12.	30 Oktober 2023	Konsultasi Revisi Bab IV dan Bab V	12.
13.	31 Oktober 2023	Konsultasi Lanjutan Revisi Bab IV dan Bab V	13.
14.	29 November 2023	ACC Bab IV dan Bab V	14.
15.	30 November 2023	Konsultasi Bab IV(Integrasi Agama)	15.



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

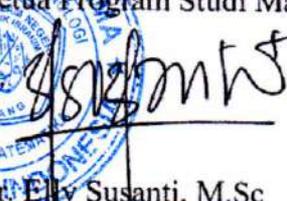
16.	27 Desember 2023	ACC Keseluruhan	16. 
17.	27 Desember 2023	ACC Keseluruhan	17. 

Malang, 27 Desember 2023

Mengetahui,

Ketua Program Studi Matematika




Dr. Elly Susanti, M.Sc

NIP. 19741129 200012 2 005