BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis klaster merupakan salah satu metode dalam analisis data yang digunakan untuk mengelompokkan objek ke dalam sejumlah kelompok berdasarkan tingkat kemiripan karakteristik tertentu, sehingga struktur dan pola yang terkandung dalam data dapat diidentifikasi secara sistematis (Jain, 2010). Metode ini banyak diterapkan pada data kejadian untuk mengidentifikasi kecenderungan pengelompokan dan keterkaitan antar objek yang tercermin dari pola kemiripan dan kedekatan dalam data (Setiyawan dan Barkah, 2025). Pada data yang memiliki dimensi ruang dan waktu, proses pengelompokan tidak hanya ditentukan oleh kemiripan atribut, tetapi juga dipengaruhi oleh kedekatan lokasi serta keterkaitan waktu terjadinya suatu peristiwa (Shi dan Cheng, 2019). Kondisi tersebut menjadikan klasterisasi spasial dan temporal sebagai pendekatan yang diperlukan untuk mengelompokkan kejadian berdasarkan kedekatan geografis dan waktu kejadian secara bersamaan (Aparna dan Idicula, 2022). Selain itu, data kejadian yang tersebar di berbagai wilayah umumnya menunjukkan pola sebaran yang tidak beraturan dan tingkat kepadatan yang bervariasi (Shen et al., 2018). Karakteristik ini menuntut penggunaan metode klasterisasi yang mampu mengidentifikasi klaster dengan bentuk arbitrer, yaitu klaster yang tidak mengikuti pola geometris tertentu, serta mampu membedakan titik-titik yang bersifat *noise* (Ester et al., 1996). Salah satu algoritma yang banyak digunakan dalam klasterisasi berbasis kepadatan dengan karakteristik tersebut adalah *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) (Ramadhan et al., 2025).

Penerapan DBSCAN dalam kajian spasial telah menunjukkan efektivitasnya dalam mengidentifikasi konsentrasi kejadian berbasis lokasi (Yolandari et al., 2025). Namun, penggunaan DBSCAN masih menghadapi kendala pada data dengan tingkat kepadatan yang bervariasi karena sensitivitas terhadap parameter global, sehingga berpotensi menghasilkan struktur klaster yang kurang stabil antarwilayah (Hahsler et al., 2019). Penentuan parameter pada DBSCAN menjadi aspek penting karena nilai parameter yang bersifat tunggal dapat membatasi kemampuan algoritma dalam menangkap struktur klaster pada tingkat kepadatan yang berbeda (Schubert et al., 2017). Kondisi ini menunjukkan bahwa diperlukan pendekatan yang mampu memetakan struktur klaster secara lebih fleksibel tanpa bergantung pada penggunaan satu nilai parameter ε dan MinPts yang diterapkan secara seragam pada seluruh data untuk menentukan kepadatan lokal dan pembentukan klaster (Azzahra, 2024). Pendekatan yang lebih adaptif diperlukan agar pola klaster tetap dapat dikenali pada data yang memiliki heterogenitas kepadatan dan perbedaan karakteristik sebaran antarwilayah (Reddy dan Reddy, 2018). Keterbatasan tersebut mendorong pengembangan algoritma klasterisasi berbasis kepadatan yang mampu menangkap struktur klaster pada berbagai tingkat kepadatan secara lebih fleksibel (Bhattacharjee dan Mitra, 2021). Pendekatan yang dikembangkan tidak hanya berfokus pada pembentukan klaster, tetapi juga pada pemetaan struktur kepadatan data secara bertahap agar pola klaster dapat diamati pada berbagai tingkat kepadatan data. Ankerst et al. (1999) memperkenalkan algoritma

Ordering Points to Identify the Clustering Structure (OPTICS) sebagai salah satu pendekatan yang dikembangkan untuk tujuan tersebut.

Algoritma *Ordering Points to Identify the Clustering Structure* (OPTICS) dikembangkan untuk memetakan struktur klaster pada berbagai tingkat kepadatan tanpa menetapkan satu nilai ambang kepadatan (Ankerst et al., 1999). Algoritma ini bekerja dengan menyusun urutan titik berdasarkan kedekatan dan keterjangkauannya sehingga struktur klaster dapat diamati melalui representasi *reachability* (Grabowski dan Kuo, 2023). Nilai *reachability distance* merepresentasikan struktur kepadatan relatif antar titik dalam proses *ordering*, sehingga variasi kepadatan dapat diidentifikasi melalui analisis *reachability plot* (Ankerst et al., 1999). Kemampuan OPTICS dalam menggambarkan struktur klaster pada berbagai tingkat kepadatan memberikan fleksibilitas dalam mengidentifikasi kelompok data yang tidak seragam, terutama pada kondisi sebaran titik yang bervariasi antarwilayah (Hajihosseinlou et al., 2024). Pendekatan ini memungkinkan pemetaan klaster tanpa bergantung pada satu parameter kepadatan, sehingga interpretasi pola kejadian dapat dilakukan pada tingkat kepadatan yang berbeda-beda (Moulavi et al., 2014).

Pengembangan OPTICS ke dalam dimensi ruang dan waktu melalui *Spatio-Temporal* OPTICS (ST-OPTICS) memungkinkan identifikasi klaster kejadian yang berdekatan secara geografis dan temporal secara simultan (Agrawal et al., 2016). ST-OPTICS memperluas mekanisme pengurutan dan perhitungan *reachability* pada OPTICS dengan mempertimbangkan kedekatan spasial dan temporal sebagai dasar pengelompokan, sehingga kejadian yang memiliki hubungan lokasi dan waktu yang berdekatan dapat dikenali dalam satu struktur klaster yang sama (Malhan, 2017). Kelebihan ST-OPTICS terletak pada kemampuannya menangkap pola kejadian yang terbentuk secara *spatio-temporal* tanpa bergantung pada satu struktur kepadatan yang seragam, sehingga lebih fleksibel pada data yang memiliki variasi kepadatan dan keterulangan kejadian pada periode tertentu (Agrawal et al., 2016). Kerangka ini memungkinkan eksplorasi pola kejadian yang membentuk konsentrasi ruang dan menunjukkan keterkaitan periode kejadian pada waktu tertentu. Oleh karena itu, penelitian ini menerapkan ST-OPTICS untuk mengelompokkan data bencana banjir berdasarkan kedekatan spasial dan temporal di Indonesia.

Banjir merupakan salah satu bencana hidrometeorologi yang paling sering terjadi di dunia dan memberikan dampak signifikan terhadap kehidupan manusia, perekonomian, serta lingkungan (Jonkman et al., 2024). United Nations Office for Disaster Risk Reduction (2025) melaporkan bahwa banjir termasuk salah satu bencana meteorologi paling dominan secara global, dengan proporsi sekitar 35-40% dari total kejadian bencana cuaca dunia. Data tersebut menunjukkan bahwa banjir menjadi salah satu penyebab utama besarnya beban bencana global, baik dilihat dari jumlah kejadian maupun luas wilayah terdampak (Yang et al., 2024). Peningkatan risiko banjir pada skala global tersebut juga tercermin pada tingkat nasional, khususnya di Indonesia. Laporan Aqueduct Global Flood Analyzer menunjukkan bahwa Indonesia termasuk salah satu negara dengan tingkat paparan banjir yang tinggi. Badan Nasional Penanggulangan Bencana (2024) mencatat bahwa jumlah kejadian bencana pada tahun 2023 lebih tinggi dibandingkan tahun 2022 dengan total 5.400 kejadian dan kondisi tersebut masih berlanjut pada data kejadian bencana tahun 2024.

Karakteristik kejadian banjir di Indonesia menunjukkan sebaran kejadian yang tidak merata antarwilayah serta variasi pola yang jelas dalam dimensi spasial dan temporal. Pada beberapa wilayah, banjir cenderung terjadi berulang pada lokasi yang relatif sama dalam periode tertentu, sementara wilayah lain memperlihatkan pola kejadian yang lebih tersebar dan tidak teratur (Albers dan Evers, 2024). Pola tersebut mengindikasikan adanya keterkaitan antara lokasi kejadian dan waktu terjadinya banjir yang belum sepenuhnya dapat dijelaskan melalui pendekatan analisis yang bersifat statis atau pemetaan berdasarkan batas administrasi wilayah (Liu et al., 2022). Keterbatasan pendekatan ini menunjukkan bahwa analisis kejadian banjir memerlukan kerangka metodologis yang mampu menangkap hubungan kejadian secara lebih mendalam (Lourenço et al., 2020). Kondisi tersebut memperlihatkan bahwa banjir tidak terjadi secara acak atau sesekali, melainkan berlangsung berulang dalam pola tertentu, sehingga diperlukan analisis berbasis data kejadian banjir yang mempertimbangkan dimensi spasial dan waktu kejadian untuk mendukung perencanaan mitigasi banjir yang lebih efektif (Awah et al., 2024).

Berdasarkan karakteristik kejadian banjir yang menunjukkan variasi spasial dan temporal, pemahaman terhadap pola sebaran dan frekuensi kejadian banjir memerlukan pendekatan analisis yang mampu menangkap keterkaitan lokasi dan waktu kejadian secara simultan. Pendekatan klasterisasi *spatio-temporal* berbasis kepadatan memberikan kerangka analisis yang lebih sesuai untuk mengidentifikasi kelompok kejadian banjir yang berdekatan secara geografis dan temporal, khususnya pada data dengan distribusi tidak beraturan dan variasi kepadatan yang tinggi. Oleh karena itu, untuk memahami pola sebaran dan frekuensi kejadian banjir secara lebih rinci, dilakukan penelitian berjudul "**Penerapan *Spatio-Temporal* Menggunakan Algoritma *Ordering Points to Identify the Clustering Structure* pada Data Bencana Banjir di Indonesia Tahun 2022-2024**". Pendekatan *Spatio-Temporal* OPTICS mampu mengidentifikasi struktur klaster kejadian banjir berdasarkan kedekatan spasial dan temporal secara simultan pada berbagai tingkat kepadatan, sehingga diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai pola kejadian banjir di Indonesia serta menjadi dasar yang lebih kuat dalam perencanaan mitigasi bencana dan pengambilan kebijakan berbasis risiko yang lebih efektif.

1.2 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang dianalisis merupakan data kejadian banjir di Indonesia pada periode tahun 2022-2024 yang direpresentasikan dalam bentuk titik *spatio-temporal* berdasarkan koordinat geografis serta waktu kejadian.
2. Nilai parameter yang digunakan dalam penelitian ini dibatasi pada kombinasi kandidat, yaitu parameter spasial (ε_1), parameter temporal (ε_2), serta MinPts yang dibentuk dari kombinasi nilai dasar dan nilai tambahan berbasis $log(n)$.
3. Evaluasi kualitas klaster dibatasi pada penggunaan metrik evaluasi internal, yaitu *Density-Based Clustering Validation* (DBCV) dan *Adjusted Rand Index* (ARI).

1.3 Tujuan dan Manfaat Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Mengidentifikasi pola klaster kejadian banjir di Indonesia pada periode 2022-2024 menggunakan algoritma ST-OPTICS.
2. Mengevaluasi kinerja klaster kejadian banjir yang terbentuk guna memperoleh gambaran kestabilan dan pemisahan klaster berdasarkan dimensi spasial dan temporal.

Hasil penelitian diharapkan dapat memberikan manfaat sebagai berikut:

1. Menambah dan mengembangkan wawasan terkait penerapan metode klasterisasi *spatio-temporal* khususnya algoritma ST-OPTICS.
2. Memberikan informasi mengenai pola sebaran dan keterulangan kejadian banjir di Indonesia yang dapat digunakan sebagai bahan pertimbangan dalam perencanaan mitigasi bencana.

1.4 Landasan Teori

1.4.1 *Data Mining*

Data mining merupakan proses penemuan pola, hubungan, dan pengetahuan yang bermakna dari kumpulan data berukuran besar dan kompleks. Menurut Han et al. (2012) data mining bertujuan mengidentifikasi informasi yang valid, baru, berpotensi berguna, serta dapat dipahami untuk mendukung analisis dan pengambilan keputusan. Sumber data pada proses data mining dapat berasal dari basis data, *data warehouse*, web, maupun repositori informasi lainnya. *Data mining* merupakan bagian inti dari kerangka *Knowledge Discovery in Databases* (KDD) yang berfokus pada tahap analitis setelah data melalui proses seleksi, pembersihan, integrasi, dan transformasi. Peran utama data mining terletak pada penerapan teknik berbasis statistika, matematika, dan *machine learning* untuk mengekstraksi pola atau struktur tersembunyi yang tidak mudah teridentifikasi melalui pendekatan analisis konvensional.

Tujuan data mining tidak hanya untuk menjelaskan karakteristik suatu data, tetapi juga untuk menemukan hubungan baru yang dapat memberikan pemahaman lebih mendalam terhadap fenomena yang diteliti. Pola yang dihasilkan dapat digunakan untuk mendukung analisis eksploratif, mengidentifikasi kecenderungan tertentu, serta menjadi dasar pengambilan keputusan berbasis data. Berbagai fungsi analisis digunakan pada data mining sesuai dengan tujuan penelitian, antara lain asosiasi untuk mengidentifikasi keterkaitan antar variabel, klasifikasi untuk memprediksi kelas suatu objek berdasarkan atribut tertentu, regresi untuk memodelkan hubungan numerik antar variabel, serta klasterisasi (*clustering*) untuk mengelompokkan objek ke dalam kelompok yang homogen berdasarkan tingkat kemiripan karakteristik tanpa memerlukan label kelas awal.

1.4.2 Analisis Klaster

Analisis klaster merupakan metode statistik multivariat yang bertujuan untuk mengelompokkan objek ke dalam klaster yang homogen berdasarkan tingkat kemiripan tertentu (Han et al., 2012). Objek dalam klaster yang sama memiliki jarak yang lebih kecil dibandingkan dengan objek pada klaster yang berbeda. Ukuran kemiripan antar

dua objek i dan j umumnya dihitung menggunakan fungsi jarak Euclidean pada Persamaan (1).

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (x_{i,k} - x_{j,k})^2} \quad (1)$$

dengan

$d(x_i, x_j)$: Jarak Euclidean antara objek ke-i dan objek ke-j
x_i : Vektor data objek ke-i
x_j : Vektor data objek ke-j
$x_{i,k}$: Nilai variabel ke-k pada objek ke-i
$x_{j,k}$: Nilai variabel ke-k pada objek ke-j
k : Indeks variabel
n : Jumlah variabel atau dimensi data

Ukuran jarak ini menjadi dasar pembentukan klaster karena objek dengan jarak lebih kecil dianggap memiliki kemiripan yang lebih tinggi. Pendekatan klasterisasi berbasis densitas mendefinisikan klaster sebagai wilayah dengan kepadatan data yang tinggi dan memisahkannya dari wilayah berkepadatan rendah yang diperlakukan sebagai *noise*. Konsep densitas umumnya ditentukan berdasarkan jumlah titik data pada radius tertentu. Pada algoritma berbasis *point-based density*, sebuah titik p dikategorikan sebagai *core point* apabila memenuhi kondisi sebagaimana dinyatakan pada Persamaan (2).

$$| N_\varepsilon(p) | \geq MinPts \quad (2)$$

dengan $N_\varepsilon(p)$ menyatakan himpunan titik pada jarak ε dari titik p. Klaster terbentuk dari kumpulan titik yang saling terhubung secara kepadatan. Bhattacharjee dan Mitra. (2021) mengklasifikasikan algoritma klasterisasi berbasis densitas ke dalam kategori *point-based*, *grid-based*, *probabilistic*, dan *data-dependent*, di mana DBSCAN dan OPTICS termasuk ke dalam pendekatan *point-based density*. Pendekatan klasterisasi berbasis densitas memiliki keunggulan dalam mengidentifikasi klaster berbentuk arbitrer serta memisahkan *noise* secara eksplisit. Keterbatasan pendekatan ini terletak pada sensitivitas terhadap parameter kepadatan dan kebutuhan memori yang relatif besar, terutama pada data dengan variasi densitas yang tinggi.

1.4.3 *Ordering Points To Identify The Clustering Structure*

Algoritma OPTICS merupakan metode klasterisasi berbasis kepadatan yang diperkenalkan oleh Ankerst et al. (1999) untuk merepresentasikan struktur klaster berbasis densitas pada berbagai tingkat kepadatan tanpa bergantung pada satu nilai parameter global. Algoritma ini dikembangkan sebagai generalisasi dari DBSCAN untuk menangani data dengan distribusi kepadatan yang tidak seragam, di mana satu nilai parameter ε sering kali tidak mampu mengungkap seluruh struktur klaster secara akurat. OPTICS tidak secara langsung menghasilkan label klaster seperti pada metode klasterisasi konvensional. Algoritma ini membangun suatu *ordering* titik data yang mencerminkan hubungan kepadatan antar objek. *Ordering* tersebut menyimpan

informasi yang ekuivalen dengan hasil klasterisasi berbasis densitas untuk berbagai nilai parameter ε hingga suatu jarak maksimum yang disebut *generating distance*.

OPTICS didasarkan pada konsep dasar klasterisasi berbasis densitas yaitu *density-reachability* dan *density-connectivity* sebagaimana didefinisikan pada DBSCAN (Reddy dan Reddy, 2018). Suatu titik p dapat dikategorikan sebagai *core point* apabila wilayah tetangga ε di sekitarnya memuat sekurang-kurangnya MinPts objek. Berdasarkan konsep tersebut, OPTICS memperkenalkan dua ukuran utama, yaitu *core distance* dan *reachability distance* yang digunakan dalam proses penyusunan struktur *ordering*. *Core distance* dari suatu titik x_i didefinisikan sebagai jarak minimum dari titik tersebut ke tetangga ke-MinPts terdekat. Penentuan *core distance* diawali dengan mendefinisikan lingkungan ε dari titik x_i, yaitu himpunan titik x_j dalam dataset D yang memiliki jarak tidak lebih dari ε terhadap titik x_i, sebagaimana ditunjukkan pada Persamaan (3).

$$N_\varepsilon(x_i) = \{ x_j \in D \mid j \neq i,\ d(x_i, x_j) \leq \varepsilon \} \tag{3}$$

dengan

$N_\varepsilon(x_i)$: Lingkungan ε (ε-*neighborhood*) dari titik x_i
x_i : Objek ke-i yang sedang diamati
x_j : Objek lain dalam dataset
D : Himpunan seluruh titik dalam dataset
$d(x_i, x_j)$: Jarak antara titik $x_i$dan titik x_j
ε : Nilai ambang jarak (radius lingkungan)
$j \neq i$: Syarat bahwa titik $x_j$tidak sama dengan titik x_i

Core distance dari titik x_i dinyatakan sebagai jarak $d_{i,(\mathrm{MinPts})}$, yaitu jarak dari titik x_i ke tetangga ke-MinPts terdekat, apabila jumlah titik dalam lingkungan ε memenuhi syarat minimum MinPts. Apabila jumlah titik pada lingkungan ε dari titik x_i kurang dari MinPts, maka *core distance* dinyatakan tidak terdefinisi. Kondisi tersebut menunjukkan bahwa titik x_i tidak berada pada wilayah dengan kepadatan yang cukup sehingga tidak memenuhi kriteria sebagai *core point*, sebagaimana ditunjukkan pada Persamaan (4).

$$CoreDist_{\varepsilon,\mathrm{MinPts}}(x_i) = \begin{cases} d_{i,(\mathrm{MinPts})}, & N_\varepsilon(x_i) \geq \mathrm{MinPts} \\ \text{tidak terdefinisi}, & N_\varepsilon(x_i) < \mathrm{MinPts} \end{cases} \tag{4}$$

Reachability distance antara titik x_i dan x_j didefinisikan sebagai nilai maksimum antara *core distance* titik x_i dan jarak aktual antara kedua titik tersebut. Secara matematis, *reachability distance* dinyatakan pada Persamaan (5).

$$ReachDist_{\varepsilon,(\mathrm{MinPts})}(x_i, x_j) = \max\{\mathrm{CoreDist}_{\varepsilon,\mathrm{MinPts}}(x_i),\ d(x_i, x_j)\} \tag{5}$$

Nilai *reachability distance* merepresentasikan jarak minimum yang diperlukan agar titik x_j dapat dianggap dapat dijangkau secara kepadatan atau dikenal dengan istilah *density-reachable* dari titik x_i, dengan asumsi bahwa titik x_i merupakan *core point*, yaitu memiliki *core distance* yang terdefinisi. Algoritma OPTICS memproses setiap titik satu kali dan menghasilkan suatu urutan titik berdasarkan nilai *reachability distance* yang meningkat. Pada proses tersebut, setiap titik disertai informasi *core distance* dan *reachability*

distance sehingga struktur klaster berbasis densitas dapat direkonstruksi untuk berbagai nilai ε yang lebih kecil dari *generating distance*.

1.4.4 *Spatio-Temporal Ordering Points To Identify The Clustering Structure*

Algoritma ST-OPTICS merupakan pengembangan dari algoritma OPTICS untuk mengidentifikasi struktur klaster pada data yang memiliki ketergantungan spasial dan temporal secara simultan (Agrawal et al., 2016). Pendekatan ini memperluas konsep klaster berbasis kepadatan dengan memasukkan dimensi ruang dan waktu ke proses penghitungan jarak, sehingga pola kejadian yang berdekatan secara geografis dan terjadi pada rentang waktu tertentu dapat diidentifikasi secara lebih representatif. Algoritma ini merepresentasikan setiap kejadian sebagai sebuah titik pada ruang fitur *spatio-temporal* yang tersusun atas koordinat lokasi dan informasi waktu kejadian. Kepadatan lokal suatu titik ditentukan berdasarkan jumlah titik lain yang berada di lingkungan *spatio-temporal* tertentu. Struktur klaster yang terbentuk merefleksikan keterhubungan kejadian secara lokasi sekaligus secara waktu, sehingga pola kejadian yang bersifat lokal dan berulang dapat dikenali dengan lebih akurat.

Perhitungan jarak pada ST-OPTICS dilakukan dengan mengombinasikan jarak spasial dan jarak temporal ke satu metrik terpadu. Jarak *spatio-temporal* antara titik pengamatan ke-i dan ke-j dirumuskan pada Persamaan (6) (Agrawal et al., 2016).

$$d(z_i, z_j) = \sqrt{\left(\frac{x_i - x_j}{\varepsilon_1}\right)^2 + \left(\frac{y_i - y_j}{\varepsilon_1}\right)^2 + \left(\frac{t_i - t_j}{\varepsilon_2}\right)^2} \tag{6}$$

dengan

$d(z_i, z_j)$: Jarak *spatio-temporal* antara titik pengamatan ke-i dan ke-j
z_i, z_j : Vektor fitur *spatio-temporal* titik ke-i dan ke-j
x_i, y_i : Koordinat spasial hasil proyeksi metrik dari titik ke-i
x_j, y_j : Koordinat spasial hasil proyeksi metrik dari titik ke-j
t_i : Waktu relatif kejadian pada titik ke-i
t_j : Waktu relatif kejadian pada titik ke-j
ε_1 : Parameter skala spasial
ε_2 : Parameter skala temporal

ST-OPTICS menggunakan tiga parameter utama, yaitu parameter spasial (ε_1), parameter temporal (ε_2), dan jumlah minimum titik ($MinPts$). Parameter ε_1 menentukan batas maksimum jarak spasial antar titik agar dapat dianggap bertetangga, sedangkan parameter ε_2 membatasi selisih waktu kejadian antar titik. Parameter $MinPts$ menyatakan jumlah minimum titik yang harus berada dalam lingkungan *spatio-temporal* tertentu agar suatu titik dapat dikategorikan sebagai titik inti (*core point*). Berdasarkan parameter tersebut, *core distance* untuk suatu titik i didefinisikan sebagai jarak *spatio-temporal* minimum ke tetangga ke-$MinPts$ dinyatakan pada Persamaan (7).

$$CoreDist^{ST}_{\varepsilon,\text{MinPts}}(x_i) = \begin{cases} d^{ST}_{i,(\text{MinPts})}, & |\ N_\varepsilon(x_i)\ | \geq \text{MinPts} \\ \text{tidak terdefinisi}, & |\ N_\varepsilon(x_i)\ | < \text{MinPts} \end{cases} \tag{7}$$

Nilai *core distance* menggambarkan tingkat kepadatan lokal di sekitar titik tersebut. Apabila jumlah tetangga tidak memenuhi syarat MinPts, maka nilai *core distance* tidak terdefinisi, yang mengindikasikan bahwa titik tersebut berada di wilayah dengan kepadatan rendah.

Reachability distance antara titik ke-i dan titik ke-j didefinisikan sebagai nilai maksimum antara *core distance* titik ke-i dan jarak *spatio-temporal* antara kedua titik tersebut. Secara matematis, reachability distance dinyatakan pada Persamaan (8).

$$ReachDist^{ST}_{\varepsilon,\mathrm{MinPts}}(x_i, x_j) = \max\{CoreDist^{ST}_{\varepsilon,\mathrm{MinPts}}(x_i),\ d^{ST}(x_i, x_j)\} \tag{8}$$

dengan

CoreDist(i) : Jarak minimum agar titik i memenuhi syarat kepadatan

ReachDist(i,j) : Jarak keterjangkauan titik j terhadap struktur kepadatan titik i

Nilai *reachability distance* mencerminkan tingkat keterhubungan kepadatan suatu titik terhadap struktur kepadatan titik sebelumnya pada proses pengurutan. Nilai yang kecil menunjukkan keterhubungan kepadatan yang kuat, sedangkan nilai yang besar mengindikasikan transisi menuju wilayah yang lebih jarang atau *noise*.

ST-OPTICS tidak secara langsung menghasilkan klaster, melainkan menyusun seluruh titik pengamatan ke dalam suatu urutan (*ordering*) berdasarkan nilai *reachability distance*. Proses *ordering* ini menghasilkan representasi struktur kepadatan data yang divisualisasikan melalui *reachability plot*. Pada grafik tersebut, sumbu horizontal menunjukkan urutan titik hasil ordering, sedangkan sumbu vertikal menunjukkan nilai *reachability distance*. Ekstraksi klaster dilakukan dengan menetapkan nilai ambang batas (*cut-off*) pada *reachability plot*. Titik-titik dengan nilai *reachability distance* yang berada di bawah ambang batas membentuk klaster *spatio-temporal*, sedangkan titik yang memiliki nilai *reachability distance* lebih besar dari ambang batas diklasifikasikan sebagai *noise*. Setiap pola lembah pada *reachability plot* merepresentasikan satu klaster dengan tingkat kepadatan *spatio-temporal* yang relatif homogen.

1.4.5 Evaluasi Klaster

Evaluasi klaster bertujuan menilai kualitas struktur klaster yang dihasilkan oleh algoritma ST-OPTICS secara kuantitatif. Proses evaluasi dilakukan untuk memastikan bahwa klaster yang terbentuk memiliki keterhubungan kepadatan internal yang baik, pemisahan antar klaster yang jelas, serta stabilitas struktur klaster yang memadai (Hassan et al., 2024). Penilaian kualitas klaster menjadi aspek penting karena algoritma ST-OPTICS tidak secara langsung menghasilkan label klaster akhir, melainkan menyajikan struktur kepadatan yang diekstraksi melalui *reachability plot* (Agrawal et al., 2016). Penelitian ini menggunakan dua pendekatan evaluasi, yaitu evaluasi internal dan evaluasi eksternal. *Density-Based Clustering Validation* (DBCV) digunakan sebagai ukuran evaluasi internal untuk menilai kualitas struktur klaster berbasis kepadatan. Kemudian, *Adjusted Rand Index* (ARI) digunakan sebagai ukuran evaluasi eksternal untuk menilai stabilitas dan konsistensi hasil klaster.

Evaluasi DBCV merupakan indeks evaluasi internal yang dirancang khusus untuk algoritma klasterisasi berbasis kepadatan (Wani, 2024). Indeks ini menilai kualitas klaster berdasarkan perbandingan antara kepadatan internal klaster dan kepadatan

pemisah antar klaster, sehingga sesuai untuk klaster berbentuk arbitrer dan data dengan variasi kepadatan yang tinggi (Moulavi et al., 2014). Tidak seperti indeks berbasis centroid, DBCV tidak mengandalkan pusat klaster, sehingga konsisten dengan karakteristik ST-OPTICS yang tidak menghasilkan centroid secara eksplisit. Secara matematis, nilai DBCV didefinisikan sebagai rata-rata berbobot dari validitas setiap klaster yang dinyatakan pada Persamaan (9).

$$DBCV = \sum_{k=1}^{K} \frac{|C_k|}{n} \cdot V(C_k) \tag{9}$$

dengan

$DBCV$: Nilai *Density-Based Clustering Validation*
K : Jumlah klaster
C_k : Klaster ke-k
$|C_k|$: Jumlah objek pada klaster ke-k
n : Jumlah total objek
$V(C_k)$: Nilai validitas klaster ke-k

Nilai validitas klaster $V(C_k)$ didefinisikan pada Persamaan (10).

$$V(C_k) = \frac{Sep(C_k) - Spar(C_k)}{\max\{Sep(C_k), Spar(C_k)\}} \tag{10}$$

Klaster yang baik memiliki nilai $Sep(C_k)$ yang tinggi dan $Spar(C_k)$ yang rendah, sehingga menghasilkan $V(C_k)$ yang mendekati 1. Besaran $Spar(C_k)$ dan $Sep(C_k)$ dihitung menggunakan konsep jarak berbasis kepadatan yang disebut *Mutual Reachability Distance* (MRD). Nilai MRD untuk pasangan objek x_i dan x_j didefinisikan pada Persamaan (11).

$$mrd(x_i, x_j) = \max\{core(x_i), core(x_j), d(x_i, x_j)\} \tag{11}$$

dengan $core(x_i)$ adalah *core distance* dari objek x_i, yaitu jarak dari x_i ke tetangga ke-$minPts$ terdekatnya, sedangkan $d(x_i, x_j)$ merupakan jarak antar dua objek. Nilai *density sparseness* dalam klaster C_k didefinisikan pada Persamaan (12) yang menunjukkan nilai *mutual reachability* terbesar pada Minimum *Spanning Tree* (MST) dari klaster C_k. Nilai ini merepresentasikan bagian paling renggang di dalam klaster.

$$Spar(C_k) = \max_{(i,j) \in MST(C_k)} mrd(x_i, x_j) \tag{12}$$

Sementara itu, nilai *density separation* klaster C_k didefinisikan pada Persamaan (13) yang menyatakan nilai *mutual reachability* minimum antara klaster C_k dengan klaster lain C_l.

$$Sep(C_k) = \min_{l \neq k} \left(\min_{x_i \in C_k, x_j \in C_l} mrd(x_i, x_j) \right) \tag{13}$$

Nilai ini menggambarkan tingkat pemisahan klaster dari klaster terdekatnya. Dengan demikian, $V(C_k)$ akan bernilai tinggi apabila klaster C_k memiliki keterhubungan internal

yang kuat dengan nilai $Spar(C_k)$ yang kecil serta pemisahan antar klaster yang jelas dengan nilai $Sep(C_k)$ yang besar.

Evaluasi ARI digunakan sebagai ukuran evaluasi eksternal untuk menilai tingkat kesesuaian antara dua hasil klasterisasi (Warrens dan Hoef, 2022). Indeks ini mengukur stabilitas struktur klaster dengan membandingkan hasil klasterisasi terhadap partisi referensi, seperti hasil klasterisasi pada subset data atau hasil klasterisasi ulang melalui *resampling* (Robert et al., 2021). Nilai ARI berada pada rentang $[-1,1]$, di mana nilai mendekati 1 menunjukkan kesesuaian dan stabilitas klaster yang tinggi, sedangkan nilai mendekati 0 menunjukkan kesesuaian yang bersifat acak. Secara matematis, ARI dirumuskan pada Persamaan (14).

$$ARI = \frac{RI - E(RI)}{\max(RI) - E(RI)} \tag{14}$$

dengan RI menyatakan *Rand Index*, $E(RI)$ menyatakan nilai ekspektasi *Rand Index*, dan $\max(RI)$ menyatakan nilai maksimum *Rand Index*. Perhitungan ARI dilakukan berdasarkan tabel kontingensi antara dua hasil klasterisasi yang dibandingkan. Misalkan n_{ij} menyatakan jumlah objek yang berada pada klaster ke-i pada partisi pertama dan klaster ke-j pada partisi kedua, maka nilai ARI dapat dinyatakan pada Persamaan (15).

$$ARI = \frac{\sum_{ij}\binom{n_{ij}}{2} - \frac{\left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]}{\binom{n}{2}}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \frac{\left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]}{\binom{n}{2}}} \tag{15}$$

dengan

n_{ij} : Jumlah objek yang berada pada klaster ke-i partisi pertama dan klaster ke-j partisi kedua

$a_i = \sum_j n_{ij}$: Jumlah objek pada klaster ke-i partisi pertama

$b_j = \sum_i n_{ij}$: Jumlah objek pada klaster ke-j partisi kedua

n : Jumlah total objek

1.4.6 Bencana Banjir

Banjir merupakan salah satu bencana hidrometeorologi yang terjadi ketika air menggenangi suatu wilayah yang seharusnya tidak tergenang, baik dalam skala kecil maupun luas (Nurhidayat, 2025). Banjir umumnya ditandai oleh meluapnya air yang melebihi kapasitas sistem alami seperti sungai dan daerah resapan, maupun sistem buatan seperti drainase dan saluran pengendali air. Kejadian banjir umumnya dipicu oleh curah hujan dengan intensitas tinggi yang menyebabkan peningkatan debit aliran air hingga melampaui kapasitas tampung sungai maupun sistem drainase, namun pada kondisi tertentu banjir juga dapat terjadi secara mendadak akibat badai ekstrem atau kegagalan infrastruktur pengendali air seperti tanggul dan bendungan yang dikenal sebagai banjir bandang (BNPB, 2014). Banjir juga dapat terjadi akibat meningkatnya limpasan permukaan ketika kemampuan tanah dalam menyerap air menurun, sehingga

volume air yang mengalir menuju saluran dan sungai menjadi lebih besar dibandingkan kapasitas yang tersedia.

Penyebab banjir bersifat multikausal dan dipengaruhi oleh faktor alam maupun aktivitas manusia. Faktor alam meliputi curah hujan tinggi, kondisi topografi, dan karakteristik hidrologis wilayah, sedangkan faktor manusia mencakup perubahan tata guna lahan, pembangunan pada kawasan bantaran sungai, penyumbatan aliran sungai oleh sampah, serta berkurangnya tutupan lahan di wilayah hulu (Arvi et al., 2025). Dampak banjir bersifat luas dan merugikan dari aspek kemanusiaan, ekonomi, dan lingkungan (Ainurrosyidah, 2022). Genangan air pada kawasan permukiman, lahan pertanian, serta pusat aktivitas perkotaan dapat menyebabkan kerusakan infrastruktur, mengganggu aktivitas sosial ekonomi, serta meningkatkan risiko kesehatan masyarakat. Banjir juga dapat menimbulkan dampak lanjutan seperti kerusakan fasilitas umum, penurunan kualitas air, gangguan transportasi, serta meningkatnya potensi penyebaran penyakit akibat lingkungan yang tercemar. Dampak tersebut menjadikan banjir sebagai salah satu bencana yang perlu memperoleh perhatian dalam upaya penanggulangan dan pengurangan risiko bencana.

Klasifikasi banjir dapat dijelaskan melalui sumber dan mekanisme terjadinya banjir. Sandink et al. (2016) menjelaskan jenis banjir yang meliputi banjir sungai, banjir drainase perkotaan, banjir akibat kegagalan susunan tanah, banjir akibat perubahan muka danau, serta banjir rob dan erosi pantai. Ragam jenis banjir tersebut menunjukkan bahwa banjir tidak hanya dipengaruhi oleh curah hujan, tetapi juga dipengaruhi oleh kondisi fisik wilayah, sistem aliran air, serta dinamika pesisir. Banjir sungai umumnya terjadi akibat luapan sungai yang dipicu peningkatan debit air. Banjir drainase perkotaan lebih sering terjadi akibat limpasan permukaan yang tidak mampu dialirkan oleh sistem drainase. Banjir rob umumnya terjadi pada wilayah pesisir akibat pasang laut dan dapat terjadi bersamaan dengan erosi pantai (Nurhidayat., 2025).

Kejadian banjir terjadi hampir setiap tahun dan sering dilaporkan di berbagai provinsi, terutama pada periode musim hujan dengan intensitas curah hujan yang tinggi. BNPB mencatat bahwa banjir merupakan bencana yang paling sering terjadi pada tahun 2022 dengan total 1.504 kejadian. Data kebencanaan nasional pada tahun 2023 mencatat total 5.400 kejadian bencana yang terjadi secara nasional (BNPB, 2023). Kejadian banjir juga menunjukkan keterulangan pada wilayah tertentu dalam rentang waktu yang relatif berdekatan, baik dalam satu musim hujan maupun pada tahun-tahun berikutnya, sehingga memperlihatkan keterkaitan antara lokasi kejadian dan waktu terjadinya bencana. Pola keterulangan tersebut menunjukkan bahwa kejadian banjir memiliki dimensi spasial dan temporal yang dapat dianalisis untuk mengidentifikasi pola kejadian banjir pada wilayah dan periode tertentu.

BAB II
METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data kejadian bencana banjir di Indonesia pada tahun 2022-2024 yang bersumber dari Badan Nasional Penanggulangan Bencana (BNPB) melalui laman resmi https://gis.bnpb.go.id/ yang disajikan secara rinci pada Lampiran 1. Data terdiri atas 4.246 titik kejadian banjir dengan atribut tanggal kejadian (*dates*) serta koordinat geografis (*latitude* dan *longitude*).

2.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri atas variabel spasial dan temporal. Variabel spasial direpresentasikan oleh *latitude* dan *longitude* yang menunjukkan lokasi terjadinya kejadian banjir, sedangkan variabel temporal direpresentasikan oleh tanggal kejadian banjir pada setiap titik pengamatan. Definisi operasional masing-masing variabel disajikan pada Tabel 1.

Tabel 1. Definisi Operasional Variabel

Variabel	Definisi
Latitude	Koordinat lintang dari lokasi terjadinya kejadian banjir di wilayah penelitian, dinyatakan dalam satuan derajat (°)
Longitude	Koordinat bujur dari lokasi terjadinya kejadian banjir di wilayah penelitian, dinyatakan dalam satuan derajat (°)
Waktu (*dates)*	Tanggal terjadinya kejadian banjir pada lokasi pengamatan di wilayah penelitian, dicatat dalam format hari/bulan/tahun

2.3 Tahapan Analisis Data

Metode analisis yang digunakan dalam penelitian ini adalah *Spatio-Temporal Ordering Points to Identify the Clustering Structure* (ST-OPTICS). Pengolahan data dilakukan menggunakan bahasa pemrograman R dengan bantuan Rstudio. Tahapan analisis data yang dilakukan adalah sebagai berikut (Agrawal et al., 2016):

1. Melakukan Pra-Pemrosesan Data
    1) Melakukan perbaikan format tanggal (*date parsing*) sehingga atribut waktu terbaca sebagai objek tanggal/waktu yang valid.
    2) Melakukan transformasi koordinat geografis ke sistem proyeksi metrik sehingga jarak spasial dapat dinyatakan dalam satuan panjang yang konsisten. Koordinat geografis dinyatakan sebagai lintang (lat) dan bujur (lon). Nilai tersebut terlebih dahulu dikonversi ke radian menggunakan Persamaan (16) dan Persamaan (17).

$$\varphi = lat \times \frac{\pi}{180} \tag{16}$$

$$\lambda = lon \times \frac{\pi}{180} \tag{17}$$

Koordinat proyeksi dihitung menggunakan persamaan proyeksi Mercator pada Persamaan (18) dan Persamaan (19).

$$x = a\lambda \tag{18}$$

$$y = a \ln\left(tan\left(\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)\right)\right) \quad (19)$$

Perhitungan koordinat menggunakan parameter $a = 6378137$ meter sebagai jari-jari bumi pada sistem referensi WGS84, kemudian hasil koordinat yang diperoleh dikonversi ke dalam satuan kilometer menggunakan Persamaan (20) dan Persamaan (21).

$$x_{km} = \frac{x}{1000} \quad (20)$$

$$y_{km} = \frac{y}{1000} \quad (21)$$

3) Membentuk waktu relatif dalam satuan hari dengan mendefinisikan waktu awal pengamatan menggunakan Persamaan (22).

$$t_0 = \min(t) \quad (22)$$

Selanjutnya, waktu relatif untuk setiap pengamatan ke-i dalam satuan hari dihitung menggunakan Persamaan (23).

$$t_i^{(\text{hari})} = \frac{t_i - t_0}{\Delta t_{\text{hari}}} \quad (23)$$

dengan

$$\Delta t_{\text{hari}} = 1 \text{ hari} = 24 \text{ jam} = 86\,400 \text{ detik} \quad (24)$$

dengan

$t_i^{(\text{hari})}$: Waktu relatif pengamatan ke-i (hari)
t_i : Waktu kejadian pada pengamatan ke-i
t_0 : Waktu kejadian paling awal pada data
Δt_{hari} : Konstanta konversi waktu untuk menyatakan selisih waktu dalam satuan hari

2. Melakukan Eksplorasi Data

Tahap eksplorasi data dilakukan untuk memahami karakteristik distribusi spasial dan temporal kejadian banjir, meliputi:

1) Menyajikan visualisasi kepadatan spasial kejadian banjir per tahun untuk mengamati pola distribusi dan konsentrasi kejadian.
2) Membuat *k-Nearest Neighbor* (k-NN) *distance plot* pada ruang spasial untuk memperoleh gambaran perubahan kepadatan data dan indikasi kandidat radius spasial.

3. Melakukan Pembentukan Ruang Fitur *Spatio-Temporal*

Implementasi ST-OPTICS pada penelitian ini menggunakan normalisasi skala spasial dan temporal melalui parameter ε_1 (km) dan ε_2 (hari), sehingga setiap titik i direpresentasikan pada Persamaan (25).

$$\mathrm{z}_i = \left(\frac{x_i}{\varepsilon_1}, \frac{y_i}{\varepsilon_1}, \frac{t_i}{\varepsilon_2}\right) \quad (25)$$

dengan

z_i : Vektor fitur *spatio-temporal* titik ke-i
x_i, y_i : Koordinat proyeksi metrik (km) titik ke-i
t_i : Waktu relatif (hari) titik ke-i
ε_1 : Parameter skala/radius spasial (km)
ε_2 : Parameter skala/radius temporal (hari)

Jarak antar titik dihitung menggunakan jarak Euclidean pada ruang fitur yang ditunjukkan pada Persamaan (6).

4. Melakukan Penentuan Parameter Menggunakan *Grid Search*
   1) Menetapkan kombinasi kandidat ε_1, ε_2, dan MinPts.
   2) Menerapkan setiap kombinasi pada proses ST-OPTICS membentuk ordering dan *reachability distance*.
   3) Menyusun kandidat nilai ambang ekstraksi ε_{cl} berdasarkan distribusi nilai *reachability distance*.
5. Melakukan pembentukan struktur klaster pada metode ST-OPTICS yang dilakukan melalui konsep kepadatan lokal yang direpresentasikan oleh *core distance* dan *reachability distance*. *Core distance* suatu titik p didefinisikan sebagai jarak minimum dari titik tersebut menuju tetangga ke-MinPts terdekat, sebagaimana dinyatakan pada Persamaan (7). *Reachability distance* titik q terhadap titik p ditentukan sebagai nilai maksimum antara *core distance* titik p dan jarak *spatio-temporal* antara p dan q, sebagaimana dinyatakan pada Persamaan (8). *Ordering* titik berdasarkan *reachability* membentuk *Reachability Plot* yang menggambarkan perubahan kepadatan dari area padat ke area renggang.
6. Melakukan ekstraksi klaster berdasarkan ε_{cl}. Ekstraksi klaster dilakukan dengan menetapkan ambang ε_{cl} pada *Reachability Plot*. Titik dengan *reachability* yang memenuhi ambang akan dipetakan menjadi klaster, sedangkan titik yang tidak memenuhi kriteria kepadatan akan dianggap sebagai *noise*.
7. Melakukan evaluasi hasil klaster dan pemilihan parameter terbaik. Evaluasi dilakukan untuk memilih kombinasi parameter yang menghasilkan klaster stabil, cakupan memadai, dan pemisahan yang baik. Kriteria yang digunakan meliputi:
   1) *Density-Based Clustering Validation* (DBCV)
      Evaluasi DBCV digunakan untuk menilai kualitas klaster berbasis kepadatan dengan mempertimbangkan kekompakan internal klaster dan keterpisahan antar klaster. Metode ini dirancang khusus untuk algoritma klasterisasi berbasis densitas seperti DBSCAN dan OPTICS. Nilai DBCV global dihitung sebagai rata-rata berbobot berdasarkan ukuran klaster, sebagaimana dinyatakan pada Persamaan (9). Nilai validitas masing-masing klaster $V(C_k)$ dihitung berdasarkan hubungan antara *density separation* dan *density sparseness*, sebagaimana ditunjukkan pada Persamaan (10).
      Nilai DBCV berada pada rentang $-1 \leq \mathrm{DBCV} \leq 1$ dengan interpretasi sebagai berikut:
      1. Nilai mendekati 1 menunjukkan klaster yang padat dan terpisah dengan baik.
      2. Nilai mendekati 0 menunjukkan struktur klaster yang lemah.

3. Nilai bernilai negatif menunjukkan struktur klaster yang buruk atau tidak valid.

2) *Adjusted Rand Index* (ARI)

Evaluasi ARI digunakan untuk menilai stabilitas klaster, yaitu konsistensi hasil pengelompokan terhadap perubahan data akibat proses *subsampling*. ARI mengukur tingkat kesesuaian antara dua partisi klaster dengan memperhitungkan kesesuaian yang dapat terjadi secara acak. Secara matematis, nilai ARI didefinisikan sebagaimana dinyatakan pada Persamaan (14). Perhitungan ARI dilakukan berdasarkan tabel kontingensi antara dua hasil klasterisasi yang dibandingkan, dengan formulasi lebih rinci sebagaimana ditunjukkan pada Persamaan (15).

Nilai ARI berada pada rentang $-1 \leq \text{ARI} \leq 1$dengan interpretasi sebagai berikut:

1. Nilai mendekati 1 menunjukkan stabilitas klaster yang sangat tinggi.
2. Nilai mendekati 0 menunjukkan kesesuaian klaster setara dengan pengelompokan acak.
3. Nilai bernilai negatif menunjukkan ketidaksesuaian antara dua hasil klasterisasi.

8. Menarik kesimpulan hasil analisis ST-OPTICS