BAB I
PENDAHULUAN

1.1 Latar Belakang

Aktivitas seismik global menunjukkan intensitas kejadian yang mengkhawatirkan. Dalam tiga bulan pertama pada tahun 2024, dua gempa besar bermagnitudo 7,60 dan 7,40 terjadi di Jepang dan Taiwan. National Centers for Environmental Information (NCEI) melaporkan bahwa kedua peristiwa tersebut menewaskan puluhan korban, merusak ribuan infrastruktur, dan menimbulkan kerugian ekonomi miliaran dolar (NCEI, 2024). Kejadian ini menunjukkan bahwa kawasan seismik Asia terus bergerak dinamis. Karakteristik yang sama juga ditemukan di Indonesia sebagai konsekuensi kondisi geografis yang berada pada jalur *Ring of Fire* dan dipengaruhi interaksi aktif Lempeng Indo-Australia, Eurasia, Pasifik, dan Filipina. Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) menyebutkan bahwa selama periode tahun 2024, tercatat lebih dari tujuh ribu gempa bumi terjadi di Indonesia, termasuk gempa bumi dengan magnitudo 7,00 pada Januari 2024 di Sulawesi Utara (Mufarida, 2024). Meskipun frekuensi kejadian relatif tinggi, pola kejadian gempa bumi tidak selalu dapat direpresentasikan secara sederhana, karena data seismik bersifat heterogen dan menunjukkan variasi spasial serta temporal yang kompleks (Jafari et al., 2024).

Data gempa bumi memiliki karakteristik statistika nonlinier (Nurtas et al., 2024), dengan distribusi magnitudo menunjukkan pola *skewed*. Gutenberg dan Richter (1944) menyebutkan adanya hubungan *log-linear* terbalik yang menghasilkan jumlah gempa bumi bermagnitudo kecil jauh lebih sering terjadi dibandingkan gempa bermagnitudo besar. Selain itu, aktivitas gempa bumi secara temporal memperlihatkan ketidakstasioneran akibat adanya rentetan kejadian seperti *foreshock* dan *aftershock* yang menyebabkan fluktuasi intensitas dalam skala waktu pendek maupun menengah (Helmstetter dan Sornette, 2003). *Foreshock* adalah gempa yang terjadi sebelum gempa utama di lokasi yang sama. Sementara, *aftershock* adalah serangkaian gempa yang terjadi setelah gempa utama, biasanya di dalam atau di sekitar area patahan yang sama, dengan magnitudo yang umumnya lebih kecil dan frekuensi yang menurun seiring waktu. Karakteristik ini menyebabkan model berbasis asumsi linier dan stasioner tidak cocok diterapkan pada data gempa bumi.

Kompleksitas temporal data gempa bumi berbanding lurus dengan keheterogenitas spasial yang sangat tinggi dan pola konsentrasi kejadian yang tidak merata (Zhao et al., 2010). Dalam analisis spasial, setiap kejadian gempa bumi direpresentasikan sebagai sebuah titik lokasi pada permukaan bumi yang dinyatakan dalam koordinat geografis dan dikenal sebagai episentrum. Berdasarkan data BMKG, distribusi titik episentrum gempa bumi tidak tersebar secara homogen, melainkan cenderung terfokus pada zona tektonik aktif seperti segmen subduksi, jalur *megathrust*, dan sistem sesar utama (McCaffrey, 2009). Sehingga menghasilkan klaster aktivitas yang hanya terbentuk pada area tertentu. Penelitian oleh Sakdiyah

dan Choiruddin (2020), menyebutkan bahwa ketidakmerataan tersebut berkaitan langsung dengan variasi struktur geologi, termasuk kompleksitas patahan, ketebalan kerak, dan kondisi pegerakan dinamis lempeng tektonik. Keragaman pola magnitudo, dinamika temporal yang tidak stasioner, dan heterogenitas spasial menuntut metode analisis yang mampu menangkap struktur nonlinier, dan tidak bergantung pada asumsi distribusi sederhana.

Struktur spasial-temporal yang kompleks dengan menggunakan pendekatan spasial klasik memiliki keterbatasan apabila diterapkan pada data seismik. Keterbatasan ini semakin terlihat pada metode berbasis kepadatan seperti *Density-Based Spatial Clustering of Applications with Noise* dan *Ordering Points to Identify the Clustering Structure*. Penelitian oleh Chen et al. (2025) menunjukkan bahwa kedua algoritma tersebut gagal mengidentifikasi klaster pada data yang memiliki variasi densitas besar sehingga menghasilkan klaster bias. Bahkan ketika kedua metode tersebut diperluas dengan penambahan dimensi waktu, ketergantungan pada ambang jarak dan sensitivitas parameter tetap tidak teratasi dan hasil klaster tetap tidak stabil (Liu et al., 2019). Keterbatasan serupa juga muncul pada statistik *hotspot* berbasis ketetanggaan seperti Getis-Ord Gi*, yang mengasumsikan struktur spasial relatif homogen sehingga tidak adaptif terhadap dinamika spasial-temporal pada data gempa bumi. Meskipun metode ini efektif untuk fenomena dengan pola yang stabil (Simanungkalit et al., 2024), penelitian lain menunjukkan bahwa pendekatan *hotspot* konvensional dapat menghasilkan bias pada wilayah dengan kontras densitas yang tinggi (Okohene dan Ozbek, 2025). Kondisi ini menegaskan perlunya pendekatan statistik yang mampu menangani densitas ekstrem dan heterogenitas spasial. *Kernel Density Estimation* (KDE) menjadi salah satu alternatif yang relevan untuk tujuan tersebut.

KDE merupakan pendekatan statistik non-parametrik yang digunakan untuk mengestimasi fungsi distribusi probabilitas dari suatu variabel acak (Silverman,1986). Menurut Nanda et al. (2019) KDE memberikan fleksibilitas dalam memodelkan distribusi spasial tanpa asumsi bentuk distribusi tertentu. Pada konteks gempa bumi, KDE terbukti efektif dalam mengestimasi dan memetakan zona kepadatan gempa dengan mempertimbangkan karakteristik penyebaran yang tidak linear (Fadli et al., 2025). Meskipun efektif untuk memetakan pola kepadatan, Fadli et al. (2025) menyebutkan bahwa KDE memiliki keterbatasan mendasar berupa *smoothing effect* yang menyebabkan kejadian menjadi berfrekuensi rendah sehingga tidak dapat membedakan suatu pola kepadatan yang termasuk aktivitas rutin. Keterbatasan tersebut menunjukkan bahwa estimasi kepadatan dengan KDE belum cukup untuk mengidentifikasi kejadian yang memiliki karakteristik berbeda dari lingkungan sekitarnya. Titik kejadian dengan perbedaan kepadatan sering disebut anomali, yaitu titik yang menunjukkan perbedaan kepadatan relatif terhadap tetangga terdekatnya, meskipun berada dalam area dengan kepadatan umum yang tinggi (Breunig et al., 2000). Oleh karena itu, diperlukan pendekatan yang memadukan KDE berdasarkan ukuran *outlier* yang membandingkan kepadatan lokal suatu titik terhadap kepadatan terhadap lingkungan sekitarnya.

Outlier factor (OF) merupakan ukuran yang menilai ketidakwajaran suatu kejadian berdasarkan perbandingan densitas lokal terhadap lingkungan terdekatnya. Konsep *outlier factor* berbasis kepadatan lokal dikembangkan untuk mengatasi keterbatasan KDE dalam membedakan area dari anomali kejadian. Qin et al. (2019) menunjukkan bahwa KDE menghasilkan *local density factor* yang lebih stabil dibandingkan pendekatan berbasis *k-Nearest Neighbors* (k-NN) pada data dengan variasi kepadatan yang besar. Liu et al. (2020) memperkuat penelitian tersebut dengan menunjukkan bahwa deteksi anomali berbasis KDE mampu menyoroti penyimpangan lokal yang tidak tampak dalam peta kepadatan klasik, terutama pada distribusi dengan bentuk klaster yang tidak beraturan. Penelitian oleh Rakhi et al. (2024) juga menegaskan bahwa pendekatan *local density score* berbasis KDE lebih efektif mengidentifikasi anomali pada lingkungan berdensitas ekstrem karena membandingkan nilai densitas titik terhadap rata-rata densitas tetangganya memberikan sensitivitas tinggi terhadap deviasi lokal. Untuk itu pendekatan *Kernel Density Estimation* berbasis *Outlier Factor* (KDE-OF) digunakan sebagai dasar analisis dalam penelitian ini. Namun, untuk memastikan bahwa pola kepadatan dan anomali yang teridentifikasi benar-benar signifikan secara spasial, diperlukan pengujian yang mampu membedakan pola sistematik dari pola acak, salah satunya melalui uji permutasi Monte Carlo.

Pengujian signifikansi permutasi Monte Carlo digunakan untuk memastikan bahwa pola kepadatan maupun anomali yang terdeteksi melalui KDE-OF benar-benar signifikan secara spasial dan bukan merupakan hasil fluktuasi acak. Penelitian oleh Liu et al. (2019) menjelaskan bahwa permutasi acak terhadap lokasi atau waktu kejadian efektif membentuk *null distribution,* yaitu distribusi acuan yang dihasilkan dari data yang diacak. Uji Permutasi bertujuan untuk mengevaluasi validitas pola signifikansi kepadatan sebagai struktur spasial (Park et al., 2009). Selain uji permutasi monte carlo, dari hasil KDE berbasis *outlier factor* diperoleh visualisasi pergeseran lintasan kejadian tahunan. Pergeseran lintasan ini merepresentasikan titik rata-rata geografis dari sekumpulan objek dan dapat digunakan untuk menelusuri perubahan posisi pusat konsentrasi fenomena dari waktu ke waktu (Dare dan Okiemute, 2020). Ketika dihitung secara periodik, pergeseran lintasan dapat menunjukkan dinamika lokasi dan arah pergerakan aktivitas spasial suatu fenomena (Fan et al., 2018), sehingga memberikan informasi perubahan kepadatan spasial. Berdasarkan hal tersebut, maka dilakukan penelitian mengenai “Penerapan *Kernel Density Estimation* Berbasis *Outlier Factor* untuk Identifikasi Kepadatan Spasial dan Pergeseran Kejadian Gempa Bumi di Indonesia dengan Uji Permutasi Monte Carlo”.

1.2 Tujuan dan Manfaat Penelitian

Tujuan dalam penelitian ini adalah:

1. Mengestimasi pola kepadatan spasial kejadian gempa bumi di Indonesia menggunakan *Kernel Density Estimation* berbasis *Outlier Factor*.
2. Mengukur signifikansi spasial dengan uji permutasi Monte Carlo pada peta kepadatan spasial dan memperoleh visualisasi pergeseran lintasan kejadian

gempa bumi berdasarkan analisis *Kernel Density Estimation* berbasis *Outlier Factor*.

Manfaat yang diharapkan dari penelitian ini adalah:

1. Menyajikan informasi mengenai kepadatan spasial gempa bumi yang diidentifikasi menggunakan *Kernel Density Estimation* berbasis *Outlier Factor dengan* uji permutasi Monte Carlo.
2. Memberikan gambaran spasial pergeseran lintasan kejadian gempa bumi tahunan yang teridentifikasi dari analisis *Kernel Density Estimation* berbasis *Outlier Factor*.
3. Menjadi referensi pendukung bagi lembaga terkait dalam memahami distribusi spasial dan karakteristik pola kejadian gempa bumi di Indonesia.

1.3 Batasan Penelitian

Batasan masalah pada penelitian ini adalah:

1. Data yang digunakan dalam penelitian adalah data kejadian gempa bumi di Indonesia pada periode 2021-2024.
2. Estimasi kepadatan spasial menggunakan tiga jenis fungsi kernel, yaitu Gaussian, Epanechnikov, dan Biweight dengan nilai *bandwidth* 0,1°; 0,3°; dan 0,5°. Sementara ukuran resolusi *grid* spasial yang digunakan adalah 0,25°; 0,5°; 1,0°; dan 5,0°.
3. Uji signifikansi spasial terhadap hasil *Kernel Density Estimation* berbasis *Outlier Factor* menggunakan uji permutasi Monte Carlo.

1.4 Teori

1.4.1 *Kernel Density Estimation*

KDE adalah metode non-parametrik yang digunakan untuk mengestimasi fungsi densitas probabilitas dari suatu variabel acak tanpa mengasumsikan bentuk distribusi tertentu (Silverman, 1986). Secara umum, KDE dimensi untuk data satu dimensi dituliskan dalam Persamaan (1):

$$\hat{f}_h(x) = \frac{1}{n\,h}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{1}$$

Sementara dalam dua dimensi bentuk yang diterapkan dituliskan dalam Persamaan (2).

$$\hat{f}_h(x,y) = \frac{1}{n\,h^2}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}, \frac{y - y_i}{h}\right) \tag{2}$$

dengan

$\hat{f}_h(x)$: Estimasi fungsi densitas di x dengan *bandwidth* h satu dimensi

$\hat{f}_h(x,y)$: Estimasi fungsi densitas di x dengan *bandwidth* h dua dimensi

n : Jumlah observasi

h : *Bandwidth*, yaitu lebar jendela *smoothing* ($h > 0$)

$K(\cdot)$: Fungsi kernel yang simetris terhadap nol dan memenuhi $\int K(u)du = 1$

x_i : Nilai observasi satu dimensi ke-i
x, y : Titik lokasi tempat densitas diestimasi
x_i, y_i : Nilai observasi dua dimens ke-i

KDE memudahkan transformasi pola titik ke permukaan kepadatan kontinu sehingga sel *high* dan *extreme* dapat divisualisasikan dengan lebih jelas (Zhang, 2022). Govorov et al. (2025) menyebutkan bahwa KDE banyak digunakan dalam analisis pemetaan kejadian spasial karena kemampuannya dalam menangkap pola non-linear dan proses distribusi yang tidak memenuhi asumsi parametris. Namun demikian, Heidenreich et al. (2013) menekankan bahwa pemilihan *bandwidth* memunculkan *trade-off* antara bias dan varians. *Bandwidth* yang terlalu kecil menghasilkan estimasi yang sangat detail tetapi penuh "*noise*" (varians tinggi), sedangkan *bandwidth* terlalu besar menghasilkan permukaan yang terlalu halus sehingga struktur asli data bisa hilang bias tinggi).

1.4.2 Fungsi Kernel

Fungsi kernel merupakan bagian utama dalam KDE, berfungsi sebagai bobot yang menentukan kontribusi setiap titik data terhadap estimasi densitas pada lokasi x. Menurut Siloko et al. (2020) fungsi kernel $K(u)$ harus memenuhi syarat dasar, sebagai berikut:

1. $\int_{-\infty}^{\infty} K(u)du = 1$
   Kernel harus memiliki total massa probabilitas sebesar 1 agar estimator $\hat{f}_h(x)$menghasilkan fungsi kepadatan yang valid dan tidak bias dalam skala global.
2. $K(u) = K(-u)$
   Kernel wajib bersifat simetris terhadap nol. Syarat ini memastikan bahwa kontribusi titik data terhadap estimasi tidak condong ke satu arah dan estimator tidak menghasilkan bias lokasi.
3. $\int u^2 K(u)du < \infty$
   Momen kedua kernel harus berhingga agar varian dari estimator kepadatan terkontrol dan tidak tak-hingga. Syarat ini menjamin bahwa KDE memiliki perilaku asimtotik yang stabil ketika ukuran sampel meningkat.

Beberapa kernel yang sering digunakan meliputi Gaussian, Epanechnikov, dan Biweight.

1.4.2.1 Kernel Gaussian

Kernel Gaussian adalah salah satu bentuk paling umum dalam KDE. Fungsi ini memiliki *unbounded support*, artinya tiap titik data memberikan kontribusi bobot tidak nol pada seluruh domain estimasi, meskipun bobotnya akan mengecil secara eksponensial bila jauh dari pusat (Oh et al., 2024). Bentuk umumnya pada Persamaan (3).

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\, u^2\right), \text{ untuk } |u| \leq 1;\ K(u) = 0 \text{ jika } |u| < \infty \qquad (3)$$

Apabila kernel Gaussian digunakan dalam konteks dua dimensi maka kontribusi titik data (x_i, y_i) terhadap estimasi pada lokasi (x, y) bergantung secara

eksponensial terhadap jarak Euclidean terstandarisasi oleh *bandwidth* h. Keunggulan kernel ini adalah menghasilkan estimasi kepadatan yang sangat mulus dan bebas dari fluktuasi tajam. Sifat ini berasal dari bentuk fungsi Gaussian yang kontinu dan tak terbatas, sehingga setiap titik data berkontribusi ke seluruh domain estimasi, meski bobotnya menurun eksponensial (Uzuazor dan Amaju, 2023).

1.4.2.2 Kernel Epanechnikov

Kernel Epanechnikov adalah kernel *compact support* sehingga nilai bobot diberikan hanya untuk $|u| \leq 1$. Kernel ini optimal dalam *mean integrated squared error* dan sering direkomendasikan dalam literatur statistik sebagai kernel yang efisien untuk estimasi kepadatan (Qin and Huang, 2024). Bentuk fungsi kernel ini dituliskan pada Persamaan (4).

$$(u) = \frac{3}{4}(1 - u^2), \quad \text{untuk } |u| \leq 1; \;\; K(u) = 0 \text{ jika } |u| > 1 \tag{4}$$

Kernel Epanechnikov optimal secara teori untuk KDE karena minimasi *mean squared error* asimptotik terkecil di antara banyak kernel lain dalam kondisi ideal. Oleh karena itu, dianggap optimal untuk estimasi densitas non-parametrik (Bijloos dan Meyers, 2021). Penggunaan kernel ini memberikan kontrol penuh terhadap batas area pembobotan sehingga estimasi densitas lebih terfokus pada lingkungan lokal tanpa kontribusi dari titik yang berada di luar radius tertentu. Keunggulan utamanya adalah pembobotan hanya pada radius satu unit terstandarisasi (*compact support*), sehingga estimasi sangat fokus pada lingkungan lokal dan tidak terpengaruh oleh titik data yang jauh.

1.4.2.3 Kernel Biweight

Kernel Biweight atau *quartic* merupakan fungsi kernel dengan *compact support* yang memiliki bentuk polinomial orde empat (Węglarczyk, 2018). Kernel ini hanya bernilai pada area terbatas dan menjadi nol di luar radius *bandwidth* yang ditetapkan. Bobot kernel menurun secara kontinu hingga mencapai nol pada batas support, sehingga menghasilkan estimasi kepadatan yang bersifat lokal dengan tingkat kehalusan yang terkendali. Fungsi matematisnya dituliskan pada Persamaan (5).

$$K(u) = \frac{15}{16}(1 - u^2)^2, \quad |u| \leq 1; \qquad K(u) = 0 \text{ untuk } |u| > 1 \tag{5}$$

Govorov et al. (2025) menyebutkan bahwa kernel ini mampu menghasilkan estimasi densitas yang stabil sekaligus peka terhadap variasi lokal karena bentuk polinomialnya memperhalus perubahan bobot di sekitar pusat observasi. Dengan *compact support*, kernel ini sangat baik untuk menghasilkan estimasi lokal yang tidak terpengaruh oleh *outlier* jauh. Sifat ini membuat kernel Biweight relevan digunakan pada data spasial yang memerlukan pembobotan lokal dengan transisi yang halus.

1.4.3 *Bandwidth*

Bandwidth adalah parameter *smoothing* dalam KDE yang mengendalikan luas pengaruh setiap titik observasi terhadap estimasi densitas di sekitarnya (Heidenreich et al., 2013). Pemilihan *bandwidth* menentukan seberapa halus atau kasar struktur yang muncul pada hasil estimasi. *Bandwidth* yang terlalu kecil menyebabkan estimasi sangat sensitif terhadap variasi lokal sehingga menghasilkan *undersmoothing*, sedangkan *bandwidth* terlalu besar menghaluskan data secara

berlebihan dapat menyebabkan *oversmoothing* (Klaichim et al., 2024). Silverman (1986) menyebutkan kerbehasilkan KDE ditentukan oleh pemilihan *bandwidth*.

Pemilihan *bandwidth* yang optimal bertujuan meminimalkan galat kuadrat. Sejumlah pendekatan telah dikembangkan untuk menentukan h. Pertama, *Rule-of-thumb methods* seperti *Silverman's Rule,* yang efektif untuk distribusi mendekati normal tetapi kurang sesuai untuk data multimodal atau sangat *skewed* (Chen, 2015). Kedua, *Cross-Validation* (CV), baik *likelihood* CV maupun *least-squares* CV, yang memilih *bandwidth* dengan meminimumkan galat (Mammen et al., 2011). Ketiga, *Plug-in methods*, yang mengestimasi turunan densitas secara *data-driven* sehingga menghasilkan *bandwidth* lebih adaptif dan akurat (Tenreiro, 2020).

Data spasial yang memiliki heterogenitas tinggi, pendekatan *fixed bandwidth* lebih umum digunakan karena memberikan kestabilan visual, memudahkan interpretasi pola geografis, dan menghindari bias berlebih akibat penggunaan *bandwidth* adaptif pada area dengan densitas rendah. *Fixed Bandwidth* menghasilkan tingkat penghalusan yang konsisten di seluruh wilayah, sehingga pola spasial yang dihasilkan lebih mudah dibandingkan antar area dan tidak terlalu dipengaruhi oleh fluktuasi lokal yang ekstrem (Davies dan Baddeley, 2017). *Bandwidth* menentukan kemampuan KDE dalam mengestimasi struktur spasial yang relevan dan menghindari distorsi akibat *oversmoothing* pada wilayah berisiko. *Bandwidth* merupakan komponen fundamental yang memengaruhi sensitivitas, stabilitas, dan interpretabilitas hasil KDE, sehingga pemilihannya harus mempertimbangkan karakteristik distribusi data, tujuan analisis, serta struktur spasial.

1.4.4 *Grid Spatial Resolution*

Grid spatial resolution adalah adalah ukuran sisi sel *grid* yang digunakan untuk merepresentasikan data spasial secara terstruktur ke dalam unit-unit tetap. Nilai resolusi ini menentukan seberapa besar wilayah di lapangan yang diwakili oleh satu sel *grid* (Skinner dan Coulthard, 2023). Semakin kecil ukuran sel, semakin halus resolusi spasialnya dan semakin tinggi detail yang dapat diungkap. Sebaliknya, ukuran sel yang besar menyebabkan representasi menjadi kasar dan mungkin mengaburkan pola spasial halus.

Resolusi *grid* berhubungan langsung dengan jumlah sel dalam wilayah analisis dan tingkat agregasi data. Misalnya, pada suatu wilayah dengan koordinat $x_{\min}, x_{\max}, y_{\min}, y_{\max}$ jika menggunakan resolusi Δx dan Δy, maka jumlah sel dapat dinyatakan dalam Persamaan (6).

$$N \approx \frac{x_{max} - x_{min}}{\Delta x} \times \frac{y_{max} - y_{min}}{\Delta y} \tag{6}$$

dengan

N : Jumlah sel *grid*
x_{max} : Batas maksimum bujur wilayah
x_{min} : Batas minimum bujur wilayah
y_{max} : Batas maksimum lintang wilayah
y_{min} : Batas minimum lintang wilayah

Δx : Resolusi *grid* bujur

Δy : Resolusi *grid* lintang

Grid spatial resolution menentukan ukuran unit analisis dalam representasi spasial berbasis *raster* atau *pixel*. Jumlah sel dalam *grid* dipengaruhi oleh luas domain pemetaan dan resolusi yang digunakan, sehingga semakin kecil ukuran sel maka semakin besar jumlah unit analisis yang terbentuk (Wasser, 2021). Secara matematis, jumlah sel dapat diperkirakan dari rasio selisih batas domain terhadap resolusi pada masing-masing sumbu. Schlosser (2024) menyebutkan bahwa resolusi berperan penting dalam kualitas analisis, resolusi yang terlalu halus menghasilkan volume data yang besar dan meningkatkan noise spasial, sedangkan resolusi yang terlalu kasar berisiko menghilangkan variasi lokal yang penting.

1.4.5 Evaluasi dan Pemilihan Parameter *Kernel Density Estimation*

Pemilihan parameter seperti jenis kernel, *bandwidth*, dan resolusi *grid* sangat mempengaruhi hasil estimasi. Oleh karena itu, diperlukan ukuran evaluasi yang dapat mengukur kinerja parameter tersebut secara objektif (Darnah et al., 2025).

1.4.5.1 *Cross-Validation Log-Likelihood*

Cross-Validation Log-Likelihood (CV-LL) mengevaluasi seberapa baik estimator kepadatan $\hat{f}(x)$ memprediksi titik yang tidak digunakan dalam proses pembentukan estimator. Nilainya dihitung berdasarkan *leave-one-out* log-likelihood Persamaan (7).

$$\mathrm{CV-LL}\ (h) = \frac{1}{n}\sum_{i=1}^{n} \log \hat{f}_{h_i}(x_i) \tag{7}$$

dengan

n : Jumlah total observasi

$\hat{f}_{h_i}(x_i)$: Estimasi kepadatan di x_i yang dihitung tanpa menggunakan titik ke-i.

Bandwidth dengan nilai CV-LL tertinggi dianggap memberikan estimasi kepadatan paling stabil dan akurat (Bolat et al., 2024).

1.4.5.2 Indeks Stabilitas Peringkat

Stabilitas hasil KDE juga dapat diukur melalui konsistensi peringkat nilai kepadatan pada *grid* ketika dilakukan *bootstrap resampling*. Stabilitas dievaluasi menggunakan korelasi Spearman antara peringkat densitas asli dan peringkat densitas hasil *bootstrap* dapat dilihat pada Persamaan (8).

$$S = Spearman\ (rank(Z), rank(Z^*)) \tag{8}$$

Indeks stabilitas tinggi menunjukkan bahwa estimasi kepadatan tidak sensitif terhadap variasi data (Li, 2017). Artinya, jika nilai S lebih rendah maka permukaan lebih kasar atau lokal lebih tinggi dan mendetail. Sementara, jika nilai S lebih tinggi maka permukaan lebih halus atau *smoothing* lebih kuat.

1.4.5.3 *Mean Integrated Squared Error*

Mean Integrated Squared Error (MISE) digunakan untuk menilai kesalahan global antara fungsi kepadatan sebenarnya $f(x)$ dan estimator kepadatan $\hat{f}_h(x)$ pada seluruh domain ruang. Secara teoretis, MISE didefinisikan dalam Persamaan (9).

$$\text{MISE}(h) = \mathbb{E}\left[\int \left(\hat{f}_h(x) - f(x)\right)^2 dx\right] \quad (9)$$

dengan h menyatakan *bandwidth* kernel dan $\mathbb{E}[\cdot]$ adalah operator ekspektasi terhadap semua probabilitas terhadap $f(x)$(Givens et al., 2013). Nilai MISE yang lebih kecil menunjukkan bahwa estimator menghasilkan permukaan kepadatan yang lebih halus dan mendekati fungsi kepadatan sebenarnya secara keseluruhan.

1.4.6 Analisis *Kernel Density Estimation* Berbasis *Outlier Factor*

1.4.6.1 *Outlier* dan Densitas Lokal

Outlier merupakan suatu observasi yang menyimpang secara abnormal dari nilai-nilai lain dalam data. Penyimpangan tersebut tidak selalu dapat diidentifikasi melalui jarak absolut, melainkan lebih tepat dinilai melalui konsistensi densitas lokal terhadap lingkungan terdekatnya. Breunig et al. (2000) menegaskan bahwa banyak anomali bersifat *local*, yaitu hanya dapat dikenali ketika densitas suatu titik dibandingkan dengan densitas titik-titik di sekitarnya.

Metode deteksi *outlier* mengalami perkembangan dari pendekatan berbasis jarak menuju pendekatan berbasis densitas, karena ukuran densitas mampu merepresentasikan tingkat kerapatan atau kekosongan lokal dalam ruang data. Zhao et al. (2019) menegaskan bahwa penilaian *outlier* perlu didasarkan pada densitas lokal agar perbedaan struktur mikro dalam distribusi data dapat diidentifikasi secara lebih sensitif. Dalam kerangka ini, suatu titik cenderung bersifat anomali apabila nilai densitas lokalnya secara signifikan lebih rendah dibandingkan densitas titik-titik di lingkungan sekitarnya. Densitas lokal pada suatu titik x dapat dinyatakan melalui estimasi densitas $\hat{f}(x)$ sebagaimana dituliskan pada Persamaan (10).

$$LD(x) = \hat{f}_h(x) \quad (10)$$

Konsep densitas lokal menggambarkan tingkat konsentrasi kejadian di sekitar suatu lokasi dan merepresentasikan kepadatan pasial dalam skala tertentu. Pada pendekatan berbasis kernel, densitas lokal suatu titik x dapat diestimasi melalui KDE. Nilai $\hat{f}_h(x)$ pada KDE merepresentasikan estimasi densitas lokal pada titik x. Namun, untuk menilai suatu titik bersifat menyimpang, densitas tersebut perlu dibandingkan dengan densitas lingkungan sekitarnya. Oleh karena itu, rata-rata densitas lokal tetangga terdekat didefinisikan dalam Persamaan (11).

$$\overline{LD}(x) = \frac{1}{k} \sum_{j \in N_k(x)} \hat{f}_h(x) \quad (11)$$

dengan $N_k(x)$ menyatakan himpunan tetangga terdekat ke-k dari titik x. Dengan perbandingan antara densitas suatu titik dan rata-rata densitas tetangganya memberikan dasar kuantitatif untuk mengukur tingkat penyimpangan lokal. Titik dengan densitas jauh lebih kecil dibandingkan lingkungannya mengindikasikan adanya ketidakkonsistenan lokal yang berpotensi sebagai anomali spasial.

Pendekatan ini digunakan sebagai dasar pembentukan berbagai ukuran *outlier* berbasis densitas. Zhao et al. (2019) menunjukkan bahwa anomali dapat dikenali melalui perbedaan mencolok antara densitas titik dan rata-rata densitas tetangganya. Penggunaan KDE menghasilkan *local density factor* yang lebih stabil

pada data dengan variasi kerapatan tinggi, sehingga penyimpangan lokal dapat diidentifikasi dengan lebih konsisten (Qin et al., 2019). Dengan demikian, konsep *outlier* dan densitas lokal digunakan sebagai fondasi bagi metode *outlier factor* yang menjadi dasar pengembangan pendekatan berikutnya.

1.4.6.2 *Kernel Density Estimation* Berbasis *Outlier Factor*

Local Outlier Factor (LOF) diperkenalkan oleh Breunig et al. (2000) sebagai ukuran yang menilai tingkat penyimpangan suatu titik dengan membandingkan kerapatan lokalnya terhadap kerapatan lokal tetangga terdekat. Meskipun efektif, LOF bergantung pada estimasi kerapatan berbasis jarak tetangga ke-k, sehingga sensitif terhadap variasi densitas ekstrem serta pemilihan parameter k. Ketergantungan ini dapat menyebabkan nilai *outlier* berubah pada data dengan struktur kepadatan yang tidak homogen. Untuk mengatasi keterbatasan tersebut, dikembangkan pendekatan *Kernel Density Estimation* berbasis *Outlier Factor*, yang bertujuan menggantikan estimasi kerapatan lokal berbasis *k-nearest neighbors* dengan estimasi kerapatan berbasis kernel. Pendekatan ini memanfaatkan keunggulan KDE dalam menghasilkan estimasi densitas yang halus dan stabil, sehingga ukuran kerapatan yang digunakan dalam perhitungan outlier factor tidak bergantung langsung pada jarak tetangga ke-k, melainkan pada kontribusi seluruh titik di sekitarnya dengan bobot yang menurun seiring bertambahnya jarak (Li et al., 2022).

Secara umum, KDE untuk titik x_i dinyatakan dalam Persamaan (1). Sementara bentuk KDE dua dimensi yang umum digunakan untuk implementasi komputasi terdapat pada Persamaan (2). Setelah estimasi densitas $\hat{f}(x_i)$, nilai densitas tersebut belum secara langsung menunjukkan adanya anomali, karena densitas kecil pada suatu titik belum tentu menyimpang apabila lingkungan sekitarnya juga memiliki densitas yang serupa. Misalkan terdapat himpunan titik kejadian.

$$\{x_1, x_2, \dots, x_n\}, \qquad x_i \in \mathbb{R}^2$$

yang masing-masing merepresentasikan koordinat episentrum dalam ruang dua dimensi. Estimasi densitas pada setiap titik x_i diperoleh KDE dua dimensi, yang dinyatakan sebagai $\hat{f}(x_i)$. Untuk setiap titik x_i, ditentukan himpunan tetangga terdekat ke-k.

$$N_k(x_i) = \{x_{i1}, x_{i2}, \dots, x_{ik}\},$$

berdasarkan jarak Euclidean antar titik episentrum. Himpunan ini merepresentasikan lingkungan lokal di sekitar titik x_i. Densitas lingkungan lokal kemudian didefinisikan dalam Persamaan (12) sebagai rata-rata estimasi densitas KDE dari tetangga terdekat.

$$\bar{f}_{N_k}(x_i) = \frac{1}{|N_k(x_i)|} \sum_{x_j \in N_k(x_i)} \hat{f}_h\left(x_j\right) \tag{12}$$

selanjutnya, KDE-OF dirumuskan sebagai rasio antara densitas lingkungan lokal dan densitas titik amatan pada Persamaan (13).

$$OF(x_i) = \frac{\bar{f}_{N_k}(x_i)}{\hat{f}(x_i)} = \frac{\frac{1}{|N_k(x_i)|}\sum_{x_j \in N_k(x_i)} \hat{f}_h\left(x_j\right)}{\hat{f}_h(x_i)} \quad (13)$$

dengan

x_i : Titik observasi ke-i yang sedang dievaluasi tingkat anomali spasialnya
x_j : Titik tetangga ke-j yang termasuk dalam himpunan tetangga $N_k(x_i)$
$\hat{f}_h(x_i)$: Estimasi densitas pada titik ke-i
$\hat{f}_h\left(x_j\right)$: Estimasi densitas KDE pada lokasi titik tetangga x_j
$\bar{f}_{N_k}(x_i)$: Rata-rata densitas lingkungan lokal pada titik ke-i
$N_k(x_i)$: Himpunan tetangga ke-k dari x_i
$|N_k(x_i)|$: Jumlah tetangga (sama dengan k)

KDE-OF mengukur tingkat penyimpangan densitas suatu titik relatif terhadap lingkungan lokalnya. Seluruh komponen densitas dalam rasio tersebut diperoleh melalui estimasi KDE, sehingga ukuran deviasi lokal sepenuhnya bergantung pada struktur densitas kernel, bukan pada konsep *local reachability density*. Secara interpretatif, nilai KDE-OF untuk suatu titik x_i dapat dijelaskan pada Tabel 1.

Tabel 1. Interpretasi Matematis Nilai KDE-OF

Nilai OF	Hubungan Densitas	Interpretasi Statistik
$\mathrm{OF}(x_i) \approx 1$	$\bar{f}_{N_k}(x_i) \approx \hat{f}_h(x_i)$	Rata-rata densitas lingkungannya sebanding dengan densitas titik artinya tidak terdapat anomali spasial.
$\mathrm{OF}(x_i) > 1$	$\bar{f}_{N_k}(x_i) > \hat{f}_h(x_i)$	Rata-rata densitas lingkungannya lebih tinggi dibandingkan densitas titik artinya mengindikasikan deviasi lokal.
$\mathrm{OF}(x_i) \gg 1$	$\bar{f}_{N_k}(x_i) \gg \hat{f}_h(x_i)$	Rata-rata densitas lingkungannya jauh lebih tinggi, terlihat perbedaan densitas yang sangat mencolok, artinya titik berpotensi sebagai anomali spasial.
$\mathrm{OF}(x_i) < 1$	$\bar{f}_{N_k}(x_i) < \hat{f}_h(x_i)$	Rata-rata densitas lingkungannya lebih rendah dibandingkan densitas titik artinya menunjukkan inti kepadatan.

Outlier Factor berfungsi sebagai ukuran kuantitatif ketidaksesuaian densitas lokal suatu titik terhadap lingkungan sekitarnya. Karena estimasi densitas $\hat{f}_h(\cdot)$ pada formulasi ini diperoleh menggunakan KDE, maka ukuran Outlier Factor yang dihasilkan selanjutnya disebut sebagai KDE-OF. Dalam konteks ini, KDE-OF bukan merupakan metode yang berdiri sendiri, melainkan kelanjutan analisis dari hasil KDE yang digunakan untuk menilai deviasi lokal berbasis perbandingan densitas.

Pendekatan KDE-OF memberikan keunggulan dibandingkan LOF konvensional karena kontribusi kernel bersifat halus dan tidak terikat pada struktur tetangga yang kaku. Qin et al. (2019) menunjukkan bahwa penggunaan estimasi densitas berbasis kernel menghasilkan ukuran deviasi lokal yang lebih stabil pada data dengan variasi kerapatan yang tinggi. Temuan tersebut diperkuat oleh Liu et al.

(2020) dan Rakhi et al. (2024), yang menegaskan bahwa *kernel-based local density score* meningkatkan ketahanan metode terhadap struktur klaster yang tidak beraturan. Oleh karena itu, KDE berbasis Outlier Factor digunakan sebagai pendekatan yang lebih adaptif untuk mengidentifikasi penyimpangan lokal pada data spasial.

1.4.7 Evaluasi *Kernel Density Estimation* Berbasis *Outlier Factor*

Pemilihan nilai k pada KDE-OF ditentukan berdasarkan bagaimana jumlah tetangga terdekat digunakan untuk membentuk estimasi densitas lokal setiap titik. Nilai k mendefinisikan jangkauan ketetanggaan melalui jarak ke tetangga ke-k, pada Persamaan (14).

$$d_k(x_i) = \left\| x_i - x_{(k)} \right\| \tag{14}$$

dengan $x_{(k)}$ menyatakan tetangga ke-k dari titik x_i. Jarak ini menjadi skala lokal yang digunakan dalam menghitung estimasi densitas berbasis ketetanggaan. Estimasi tersebut kemudian dibandingkan dengan densitas tetangga terdekat untuk memperoleh skor *outlier factor*.

Nilai k yang berbeda menghasilkan struktur skor *outlier* yang berbeda, sehingga pemilihan k dievaluasi menggunakan dua metrik indeks stabilitas peringkat dan Integrated Squared Error (ISE) (Zhou et al., 2019). Stabilitas menggunakan indeks stabilitas peringkat sebagaimana telah didefinisikan pada Subbab 1.4.5.2 pada Persamaan (8). Untuk menilai kesesuaian bentuk distribusi skor *outlier factor*, nilai ISE antara KDE-OF yang dihitung dari data asli, maka dapat dihitung dalam Persamaan (15).

$$ISE(k) = \int \left[\hat{g}_{h,k}(t) - \hat{g}_{h,0}(t) \right]^2 dt \tag{15}$$

dengan

$\hat{g}_{h,k}(t)$: Kurva kepadatan skor *outlier factor* dari data asli

$\hat{g}_{h,0}(t)$: Kurva kepadatan skor *outlier factor* dari pemilihan k tertentu

Nilai k dipilih dengan mempertimbangkan stabilitas S yang tinggi dan ISE yang rendah, karena kombinasi tersebut menunjukkan bahwa struktur anomali konsisten serta tidak mengalami distorsi ketika k berubah.

1.4.8 Uji Signifikansi Permutasi Monte Carlo

Uji permutasi Monte Carlo digunakan untuk menilai signifikansi statistik Indeks Moran sebagai ukuran autokorelasi spasial, yaitu suatu kondisi yang muncul akibat adanya interaksi antara wilayah atau karena adanya kemiripan karakteristik objek dalam suatu ruang. Interaksi ini menunjukkan bahwa nilai pengamatan di suatu wilayah dipengaruhi oleh nilai pengamatan di sekitarnya (Mailanda et al., 2022). Pemeriksaan autokorelasi spasial dapat dilakukan menggunakan Indeks Moran yang dirumuskan pada Persamaan (16) (Lee dan Wong, 2001).

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{w}_{ij} (x_i - \bar{x})(x_j - \bar{x})}{n \sum_{i=1}^{n} \sum_{j=1}^{n} n \sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{16}$$

dengan

I : Indeks Moran

n : Banyaknya pengamatan
x_i : Nilai observasi pada lokasi ke-i
x_j : Nilai observasi pada lokasi ke-j
$\bar{x}$: Rata-rata nilai pengamatan
$\widetilde{w}_{ij}$: Elemen pada pembobot spasial terstandarisasi antara daerah i dan j

Nilai dari Indeks Moran berkisar antara $-1 \leq I \leq 1$. Nilai $I < 0$ menunjukkan terjadi autokorelasi spasial negatif yang artinya wilayah yang berdekatan memiliki nilai yang cenderung berbeda. Nilai $I = 0$ menunjukkan bahwa tidak terjadi autokorelasi spasial, sedangkan nilai $I > 0$ menunjukkan terjadi autokorelasi positif yang artinya wilayah yang berdekatan memiliki kemiripan nilai (Wardana et al., 2023). Pengujian autokorelasi spasial dilakukan dengan hipotesis sebagai berikut:

$H_0: I = 0$ (tidak terdapat autokorelasi spasial)
$H_1: I \neq 0$ (terdapat autokorelasi spasial)

Statistik uji indeks moran yang digunakan dihitung menggunakan Persamaan (17).

$$Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}} \tag{17}$$

dengan

$E(I) = -\frac{1}{n-1}$

$Var(I) = \frac{n^2 S_1 - n S_2 + 3S_0^2}{(n^2-1)S_0^2} - [E(I)]^2$

$S_0 = n \sum_{i=1}^{n} \sum_{j=1}^{n} \widetilde{w}_{ij}$

$S_1 = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\widetilde{w}_{ij} + \widetilde{w}_{ji})^2$

$S_2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \widetilde{w}_{ij} + \sum_{j=1}^{n} \widetilde{w}_{ji} \right)^2$

Pengambilan keputusan pada uji ini adalah tolak H_0 pada taraf signifikansi α jika $|Z(I)| > \frac{Z_\alpha}{2}$ atau jika $p - value < \alpha$ yang berarti terdapat autokorelasi spasial dalam data. Namun, pada banyak aplikasi spasial, termasuk data dengan distribusi kompleks dan ketergantungan lokal yang kuat, pendekatan parametrik sering kali kurang andal. Oleh karena itu, signifikansi Indeks Moran umumnya diuji menggunakan pendekatan non-parametrik melalui uji permutasi Monte Carlo.

Uji permutasi Monte Carlo merupakan pendekatan non-parametrik yang digunakan untuk menilai signifikansi statistik suatu pola spasial atau nilai statistik tertentu tanpa bergantung pada asumsi distribusi parametrik. Dvořák dan Mrkvička (2024) menyebutkan prinsip dari pendekatan ini adalah membentuk distribusi nol (*null distribution*) melalui pengacakan data, sehingga nilai statistik yang diamati dapat dibandingkan secara langsung dengan distribusi acuan hasil permutasi. Menurut Liu et al. (2019), permutasi acak terhadap lokasi atau waktu pengamatan mampu menghasilkan distribusi referensi yang merepresentasikan kondisi tanpa struktur spasial, sehingga setiap pola yang terlihat setelah analisis dapat diuji apakah muncul akibat struktur sistematik atau sekadar hasil dari fluktuasi acak. Setiap pengacakan menghasilkan satu nilai Moran's I dari data hasil permutasi.

Misalkan I_{obs} menyatakan nilai Moran's I yang dihitung dari data asli, dan I_b menyatakan nilai *Moran's I* pada permutasi ke-b, dengan $b = 1,2,\dots,B$. Maka diperoleh himpunan $\{I_1, I_2, \dots, I_B\}$ yang membentuk *null distribution* Moran's I, yaitu distribusi nilai Indeks Moran yang merepresentasikan kondisi tanpa keterkaitan spasial antara nilai atribut dan lokasi. Nilai signifikansi statistik kemudian ditentukan berdasarkan proporsi hasil permutasi yang menghasilkan nilai Moran's I lebih ekstrem dibandingkan nilai observasi. Karena Indeks Moran memiliki arah ekstremitas, signifikansi statistik dalam uji permutasi Monte Carlo dapat dievaluasi menggunakan $p - value$ satu sisi maupun dua sisi. Evaluasi satu sisi digunakan untuk menilai kecenderungan autokorelasi spasial positif atau negatif, sedangkan uji dua sisi digunakan untuk menguji keberadaan autokorelasi spasial secara umum. Berdasarkan pertimbangan tersebut, untuk $p - value$ satu sisi kanan, yang digunakan untuk menilai autokorelasi spasial positif, $p - value$ permutasi dirumuskan pada Persamaan (18).

$$p_{ge} = \frac{\#(I_b \geq I_{obs}) + 1}{B + 1} \tag{18}$$

Sebaliknya, $p - value$ satu sisi kiri, yang digunakan untuk mengevaluasi autokorelasi spasial negatif, didefinisikan pada Persamaan (19).

$$p_{le} = \frac{\#(I_b \leq I_{obs}) + 1}{B + 1} \tag{19}$$

dengan

p_{ge} : $p - value$ satu sisi kanan, ge (*greater or equal*)
p_{le} : $p - value$ satu sisi kiri, le (*less or equal*)
I_b : nilai statistik uji dari permutasi ke–b
I_{obs} : nilai statistik uji dari data asli
$\#(I_b \geq I)$: jumlah permutasi dengan nilai Moran's I ≥ nilai observasi
$\#(I_b \leq I)$: jumlah permutasi dengan nilai Moran's I ≤ nilai observasi
B : jumlah total permutasi

Uji statistik permutasi Monte Carlo dengan Indeks Moran menggunakan uji dua sisi, $p - value$ dihitung menggunakan Persamaan (20).

$$p_{\text{two}} = 2\min\{p_{ge}, p_{le}\} \tag{20}$$

Maka pengambilan keputusan pada uji ini adalah tolak H_0 pada taraf signifikansi α jika $p_{two} < \alpha$. Sementara, $z - score$ berbasis permutasi menggunakan statistik uji Indeks Moran pada Persamaan (21).

$$z_b = \frac{I_{obs} - \mu_{perm}}{\sigma_{perm}} \tag{21}$$

dengan

z_b : Nilai moran dari permutasi
μ_{perm} : Rata-rata Moran's I hasil permutasi I_b
σ_{perm} : Standar deviasi hasil permutasi I_b

1.4.9 Pergeseran Lintasan Kejadian

Pergeseran kejadian merupakan lintasan titik representatif yang menggunakan posisi rata-rata suatu sekumpulan objek dalam ruang geografis sebagai titik bantu. Secara

matematis, pusat lintasan dihitung sebagai nilai tengah terukur dari koordinat lintang dan bujur seluruh titik yang dianalisis pada suatu wilayah (Dare and Okiemute, 2020). Konsep ini digunakan dalam berbagai kajian spasial untuk mengekspresikan pusat kecenderungan lokasi dan mengidentifikasi pergeseran posisi dari waktu ke waktu, terutama ketika objek mengalami dinamika spasial (Fan et al., 2018). Jika terdapat n titik dengan koordinat (x_i, y_i), maka pusat lintasan spasial pada waktu ke- t dirumuskan sebagai rata-rata aritmetika dari seluruh koordinat titik, sesuai dengan Persamaan (22).

$$\bar{x}_t = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \bar{y}_t = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{22}$$

dengan menghitung rata-rata pusat amatan secara periodik, urutan titik $(\bar{x}_t, \bar{y}_t)$ membentuk suatu lintasan yang menggambarkan arah perubahan posisi relatif distribusi spasial dari waktu ke waktu. Dalam kajian spasial dinamis, lintasan pergeseran bertujuan untuk menunjukkan indikator perubahan geometri distribusi spasial (Fan et al., 2018). Oleh karena itu, pendekatan ini bersifat deskriptif-eksploratif dan digunakan untuk mengamati pola pergeseran tanpa melibatkan pembobotan nilai atau asumsi ketergantungan spasial antar titik.

Konsep pergeseran lintasan kejadian berdasarkan hasil KDE-OF dengan uji permutasi diterapkan pada sel-sel *grid* yang memiliki nilai KDE-OF tinggi, sehingga lintasan yang terbentuk merepresentasikan perubahan posisi relatif area-area yang menunjukkan deviasi lokal signifikan menurut hasil KDE-OF. Dengan demikian, lintasan pergeseran menyajikan ringkasan spasial dinamika distribusi sel *grid* dengan tingkat penyimpangan lokal tinggi. Pendekatan ini memungkinkan integrasi hasil KDE-OF ke dalam analisis dinamika spasial secara temporal, sehingga pergeseran kejadian dapat diinterpretasikan sebagai perubahan kecenderungan area observasi dalam struktur kepadatan spasial. Pergeseran lintasan kejadian tidak hanya berfungsi sebagai informasi deskriptif, tetapi juga sebagai indikator pergeseran titik pusat lokasi spasial apabila dihitung secara periodik. Pergeseran kejadian melalui lintasan pertemuan titik pusat dapat menggambarkan pergeseran kepadatan spasial dan konsentrasi area *grid* yang memiliki deviasi lokal tinggi menurut KDE-OF.

1.4.10 Kejadian Gempa Bumi

Gempa bumi merupakan getaran atau guncangan yang terjadi di permukaan bumi akibat pelepasan energi secara tiba-tiba dari dalam litosfer (BMKG, 2023). Proses ini umumnya disebabkan oleh pergerakan lempeng tektonik, aktivitas sesar aktif, maupun deformasi batuan di zona subduksi. Energi yang dilepaskan merambat dalam bentuk gelombang seismik yang direkam dan dianalisis melalui instrumen seismometer. Secara fisik, menurut Dong dan Luo (2022) mekanisme terjadinya gempa terkait dengan akumulasi tegangan pada bidang patahan. Ketika tegangan melebihi kekuatan geser batuan, terjadi pelepasan energi yang menyebabkan pematahan dan pergeseran antarblok batuan, menghasilkan getaran yang menjalar ke segala arah. Besarnya gempa diukur menggunakan magnitudo, sedangkan dampaknya di permukaan dinilai melalui intensitas.

Gempa bumi memiliki karakteristik spasial dan temporal yang sangat bervariasi, dipengaruhi oleh kondisi tektonik regional, geometri patahan, serta kedalaman sumber gempa (Hutchings dan Mooney, 2021). Indonesia mengalami aktivitas seismik yang sangat tinggi akibat pertemuan dan subduksi Lempeng Indo-Australia, Eurasia, dan Pasifik. Kondisi ini menghasilkan pola dan intensitas gempa bumi yang tidak seragam, baik dari sisi lokasi maupun magnitudo (Muttayy et al., 2023). Variasi magnitudo tersebut secara operasional diklasifikasikan ke dalam beberapa kategori standar untuk memudahkan analisis dan interpretasi statistik. Klasifikasi ini menyajikan rentang magnitudo beserta estimasi frekuensi kejadiannya per tahun, sehingga dapat digunakan sebagai acuan ketika mengelompokkan data gempa dalam analisis eksploratif. Menurut Bolt (2025) kategori magnitudo ditunjukkan pada Tabel 2.

Tabel 2. Kategori Gempa Bumi Berdasarkan Magnitudo.

Magnitudo	Kategori	Dampak
< 3.0	*Micro*	Umumnya tidak dirasakan oleh manusia, tetapi terekam oleh instrumen seismik.
3.0–3.9	*Minor*	Dirasakan oleh sebagian masyarakat, tidak menimbulkan kerusakan.
4.0–4.9	*Light*	Dirasakan oleh banyak orang, dapat menyebabkan kerusakan ringan pada objek.
5.0–5.9	*Moderate*	Berpotensi menimbulkan kerusakan pada struktur bangunan yang lemah.
6.0–6.9	*Strong*	Menyebabkan kerusakan sedang di wilayah berpenduduk.
7.0–7.9	*Major*	Menyebabkan kerusakan serius pada area luas dan berpotensi menimbulkan korban.
≥ 8.0	*Great*	Menyebabkan kerusakan berat dan berdampak luas.

BAB II
METODOLOGI PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang mencakup kejadian gempa bumi di Indonesia selama periode 2021 hingga 2024. Data tersebut diperoleh dari BMKG dan dapat diakses melalui situs web BMKG (http://repogempa.bmkg.go.id). Dataset tersebut memuat informasi waktu kejadian gempa, meliputi tahun, bulan, tanggal, jam, menit, dan detik, serta informasi spasial berupa koordinat lokasi kejadian gempa (lintang dan bujur) dan besaran magnitudo gempa. Data disajikan dalam bentuk deret kejadian (*event-based data*), yaitu setiap baris merepresentasikan satu peristiwa gempa bumi yang bersifat independen secara pencatatan waktu. Variabel yang digunakan dalam penelitian ini beserta satuan dan definisinya disajikan pada Tabel 3.

Tabel 3. Variabel Penelitian

Variabel	Satuan	Definisi	Peran dalam Analisis
Waktu	UTC	Menunjukkan waktu terjadinya gempa bumi dalam format (YYYY-MM-DD HH:MM:SS), dicatat berdasarkan *Coordinated Universal Time* (UTC) sebagaimana disediakan dalam katalog resmi BMKG.	Segmentasi data tahunan.
Bujur	Derajat (°)	Koordinat geografis yang menunjukkan posisi geografis episentrum relatif terhadap meridian utama pada 0°, dengan nilai positif untuk bujur timur dan nilai negatif untuk bujur barat.	Variabel spasial utama.
Lintang	Derajat (°)	Koordinat geografis yang menunjukkan posisi gempa relatif terhadap garis ekuator pada 0°. Nilai positif menunjukkan lintang utara dan nilai negatif menunjukkan lintang selatan.	Variabel spasial utama.
Magnitudo	M	Ukuran besarnya energi yang dilepaskan oleh gempa bumi. Semakin besar nilai magnitudo, semakin besar energi seismik yang dilepaskan.	Variabel deskriptif.

Data gempa bumi yang digunakan disusun dalam bentuk tabel observasi, yang setiap barisnya merepresentasikan satu kejadian gempa dengan atribut waktu

kejadian, koordinat lintang-bujur, dan magnitudo. Struktur data ini menjadi *input* utama dalam proses analisis, termasuk estimasi kepadatan spasial menggunakan KDE-OF dengan uji permutasi Monte Carlo, perhitungan titik pusat dan lintasan pergeseran kejadian. Dataset yang digunakan dalam penelitian ini disajikan pada Lampiran 1.

2.2 Analisis Data

Penelitian ini menerapkan *Kernel Density Estimation* berbasis *Outlier Factor* dengan uji signifikansi permutasi Monte Carlo untuk menganalisis pola kepadatan, anomali spasial, dan pergeseran lintasan kejadian gempa bumi di Indonesia periode 2021–2024. Pengolahan data dilakukan menggunakan bahasa pemrograman *Python* pada *platform Visual Studio Code*. Tahapan analisis disusun secara sistematis sebagai berikut:

1. Melakukan eksplorasi data spasial kejadian gempa bumi. Data kejadian gempa bumi periode 2021–2024 dipetakan menggunakan koordinat bujur–lintang (x, y) dalam sistem koordinat geografis (WGS 84/EPSG:4326). Setiap kejadian dinyatakan sebagai titik $x_i = (x_i, y_i) \in \mathbb{R}^2$, dengan atribut magnitudo yang dikelompokkan ke dalam kelas tertentu. Peta dasar batas administrasi*, fault lines*, dan zona *megathrust* ditambahkan sebagai *contextual layer* untuk memberikan konteks geologi pada pola spasial sebaran titik. Luaran tahap ini berupa peta sebaran titik gempa dan ringkasan deskriptif yang menunjukkan kecenderungan konsentrasi spasial awal.
2. Menetapkan *grid* spasial untuk evaluasi kepadatan, pada penelitian ini *grid* spasial yang digunakan adalah resolusi 0,25°; 0,5°; 1,0°; dan 5,0° pada lintang dan bujur wilayah Indonesia.
3. Melakukan estimasi kepadatan spasial dengan KDE dua dimensi yang digunakan untuk menghitung kepadatan titik. Misalkan terdapat himpunan titik kejadian:

   $$\{x_1, x_2, \dots, x_n\}, x_i = (x_i, y_i) \in \mathbb{R}^2$$

   Estimator KDE dua dimensi pada lokasi (x, y) dinyatakan pada Persamaan (2). Dalam penelitian ini, estimasi dilakukan untuk beberapa kombinasi fungsi kernel, yaitu Gaussian, Epanechnikov, Biweight dan beberapa nilai *bandwidth*, $h = 0{,}1°;\ 0{,}3°; 0{,}5°$ dan *grid* dengan resolusi 0,25°; 0,5°; 1,0°; dan 5,0°. Dilanjutkan visualisasi pemetaan setiap kombinasi yang terbentuk.
4. Melakukan klasifikasi peta kepadatan. Nilai KDE bersifat kontinu, sehingga untuk interpretasi peta kepadatan digunakan klasifikasi menjadi empat kelas: *Low, Medium, High, Extreme*. Klasifikasi didasarkan pada distribusi nilai $\hat{f}_h(x)$, menggunakan ambang kuantil:

   $$Q_i = \text{kuartil ke} - i \ \text{dari} \ i \{\hat{f}_h(x)\}$$

   Kemudian kelas dibentuk berdasarkan interval kuantil yang ditetapkan, meliputi nilai ambang bawah atau *Threshold* (Thr), kuartil pertama (Q_1), kuartil kedua atau median (Q_2), kuartil ketiga (Q_3), serta nilai maksimum kepadatan ($Z_{\max}$).

5. Melakukan evaluasi KDE unruk menentukan kernel, *bandwidth*, dan *grid* optimal menggunakan CV-LL, S, dan MISE pada Persamaan (7), (8), dan (9). Parameter dengan nilai kombinasi konsisten dipilih sebagai parameter KDE optimal.
6. Menghitung KDE-OF dan evaluasi parameter k. Setelah diperoleh estimasi kepadatan $\hat{f}(x_i)$ dari KDE optimal, dilakukan perhitungan KDE-OF untuk berbagai kandidat jumlah tetangga terdekat k. Untuk setiap nilai k dalam himpunan kandidat:
$$\{k_1, k_2, \dots, k_m\}$$
KDE-OF dihitung pada seluruh titik kejadian menggunakan Persamaan (13). Hasil KDE-OF untuk setiap k belum digunakan sebagai peta akhir, melainkan dievaluasi terlebih dahulu menggunakan S dan ISE masing-masing pada Persamaan (8) dan (15). Nilai k optimal dipilih sebagai nilai yang memberikan struktur KDE-OF paling stabil dan kesalahan estimasi paling kecil, sehingga ukuran deviasi lokal tidak terlalu sensitif tetapi tetap mampu menangkap variasi kepadatan spasial yang bermakna.
7. Melakukaan pemetaan berdasarkan hasil evaluasi KDE-OF. Setelah diperoleh nilai k optimal, KDE-OF dihitung kembali menggunakan nilai k tersebut dan dijadikan hasil akhir analisis deviasi lokal. Nilai KDE-OF kemudian diagregasi ke dalam *grid* resolusi untuk membentuk peta. Setiap sel *grid* merepresentasikan tingkat deviasi lokal berdasarkan perbandingan densitas titik dengan densitas lingkungan sekitarnya..
8. Melakukan uji signifikansi spasial menggunakan uji permutasi monte carlo. Uji permutasi digunakan untuk menguji apakah pola spasial nilai KDE-OF bukan berasal dari proses acak.
   a) Indeks Moran
   Misalkan terdapat n sel *grid* valid, masing-masing memiliki nilai KDE-OF x_i, dengan rata-rata $\bar{x}$. Dengan bobot spasial terstandarisasi $\widetilde{w}_{ij}$, statistik Moran dituliskan pada Persamaan (16). Bobot $\widetilde{w}_{ij}$dibentuk dari kNN antar pusat sel *grid*, dengan parameter k_w yaitu jumlah tetangga untuk bobot Moran. Nilai k_w dipilih melalui $z - score$ paling besar atau $p - value$ paling kecil.
   b) Permutasi Monte Carlo
   Untuk menguji signifikansi, nilai KDE-OF antar sel *grid* yang dipermutasi sebanyak B kali, sementara struktur bobot spasial tetap. Untuk permutasi ke-b, dihitung:
$$I_b = 1,2, \dots, B$$
   Himpunan $\{I_1, \dots, I_B\}$ membentuk *null distribution* Moran's I. Adapun $p - value$ berbasis permutasi, menggunakan $p - value$ satu sisi kanan (autokorelasi positif) didefinisikan pada Persamaan (18), $p - value$ satu sisi kiri (autokorelasi negatif) menggunakan Persamaan (19), dan uji dua sisi uji yang digunakan adalah Persamaan (20). Adapun Keputusan pada uji permutasi Monte Carlo adalah tolak $H_0: I = 0$ pada taraf α jika $p_{two} < \alpha$.

Sementara, penentuan $z - score$ berbasis permutasi menggunakan statistik uji Indeks Moran pada Persamaan (21).

c) Konvergensi $p - value$
Stabilitas $p - value$ diperiksa dengan menghitung kekonvergenan $p - value$ hingga iterasi tertentu dan membandingkannya dengan $p - value$ akhir pada B.

9. Menghitung pusat lintasan spasial tahunan. Pusat lintasan dihitung untuk setiap tahun menggunakan Persamaan (22) bertujuan untuk menggambarkan pergeseran pusat aktivitas spasial antar tahun. Misalkan untuk setiap tahun t diperoleh himpunan sel *grid* dengan nilai KDE-OF yang termasuk dalam kategori *High* dan *Extreme*, didefinisikan:

$$\mathcal{G}_t = \{(x_i, y_i) \mid \text{KDE-OF}(x_i, y_i) \in \text{High} \cup \text{Extreme}\}$$

Sehingga, itik representatif tahunan didefinisikan sebagai rata-rata spasial pusat sel:

$$\bar{x}_t = \frac{1}{|\mathcal{G}_t|} \sum_{(x_i,y_i)\in \mathcal{G}_t} x_i\,, \qquad \bar{y}_t = \frac{1}{|\mathcal{G}_t|} \sum_{(x_i,y_i)\in \mathcal{G}_t} y_i$$

Urutan $\{(\bar{x}_{2021}, \bar{y}_{2021}), \dots, (\bar{x}_{2024}, \bar{y}_{2024})\}$ membentuk lintasan yang menggambarkan perubahan posisi relatif area-area dengan nilai KDE-OF *high* dan *extreme* antar tahun.

10. Menarik kesimpulan hasil analisis KDE-OF dengan uji permutasi Monte Carlo.