



# Klasifikasi Tingkat Risiko Gempa Menggunakan Pola Spasial dan Temporal Berbasis Decision Tree

Mugi Prasetyo, Heni Sulistiani\*, Onassis Yusuf Inonu, Kardita Magda, Budi Santosa

Fakultas Teknik dan Ilmu Komputer, Magister Ilmu Komputer, Universitas Teknokrat Indonesia, Kota Bandar Lampung, Indonesia

Email: <sup>1</sup>mugiprasetyo@teknokrat.ac.id, <sup>2,\*</sup>henisulistiani@teknokrat.ac.id, <sup>3</sup>onassis\_yusuf\_inonu@teknokrat.ac.id,

<sup>4</sup>kardita\_magda@teknokrat.ac.id, <sup>5</sup>budi.santosa@teknokrat.ac.id

Email Penulis Korespondensi: henisulistiani@teknokrat.ac.id

**Abstrak**—Indonesia merupakan wilayah yang sangat rentan terhadap gempa bumi akibat letaknya di zona pertemuan lempeng tektonik aktif. Penelitian ini bertujuan untuk mengklasifikasikan tingkat risiko gempa berdasarkan pola spasial dan temporal menggunakan metode Decision Tree sebagai solusi dalam memprediksi potensi bahaya gempa. Data yang digunakan adalah data gempa di Indonesia dari tahun 2015 hingga 2023 yang diperoleh dari dataset publik, mencakup informasi lokasi (latitude dan longitude), waktu kejadian (tahun dan bulan), serta magnitudo gempa. Gempa dikategorikan menjadi tiga kelas risiko: Low ( $M < 4.0$ ), Medium ( $4.0 \leq M < 6.0$ ), dan High ( $M \geq 6.0$ ). Model Decision Tree berhasil dibangun dengan akurasi rata-rata sebesar 88% pada data uji. Hasil menunjukkan bahwa gempa lebih banyak terjadi di zona subduksi aktif seperti Zona Subduksi Sunda (Sumatra dan Jawa), Busur Banda (Nusa Tenggara, Maluku, Seram), Sulawesi, dan Papua. Analisis temporal juga mengungkapkan adanya fluktuasi jumlah gempa tiap tahun dan musim, dengan peningkatan aktivitas pada bulan-bulan tertentu. Visualisasi spasial memperkuat temuan bahwa wilayah timur Indonesia lebih aktif secara seismik dibandingkan wilayah barat. Penelitian ini membuktikan bahwa pendekatan pembelajaran mesin dapat digunakan untuk mendukung mitigasi bencana gempa bumi melalui identifikasi risiko berbasis data historis.

**Kata Kunci:** Gempa Bumi; Decision Tree; Analisis Spasial-Temporal; Klasifikasi Risiko; Zona Subduksi

**Abstract**—Indonesia is an area that is very vulnerable to earthquakes due to its location in the meeting zone of active tectonic plates. This study aims to classify the level of earthquake risk based on spatial and temporal patterns using the Decision Tree method as a solution in predicting potential earthquake hazards. The data used is earthquake data in Indonesia from 2015 to 2023 obtained from public datasets, including location information (latitude and longitude), event time (year and month), and earthquake magnitude. Earthquakes are categorized into three risk classes: Low ( $M < 4.0$ ), Medium ( $4.0 \leq M < 6.0$ ), and High ( $M \geq 6.0$ ). The Decision Tree model was successfully built with an average accuracy of 88% on the test data. The results show that earthquakes mostly occur in active subduction zones such as the Sunda Subduction Zone (Sumatra and Java), Banda Arc (Nusa Tenggara, Maluku, Seram), Sulawesi, and Papua. Temporal analysis also shows fluctuations in the number of earthquakes by year and season, with increased activity in certain months. The spatial visualization reinforces the finding that the eastern region of Indonesia is more seismically active than the western region. This research proves that machine learning approaches can be used to support earthquake disaster mitigation through historical data-based risk identification.

**Keywords:** Earthquake; Decision Tree; Spatial-Temporal Analysis; Risk Classification; Subduction Zone

## 1. PENDAHULUAN

Indonesia berada di wilayah yang sangat rentan terhadap bencana gempa bumi, tsunami, tanah longsor, dan letusan gunung berapi. Letak geografisnya yang berada di jalur Ring of Fire atau Cincin Api Pasifik membuat negara ini dilintasi oleh pertemuan lempeng tektonik aktif seperti lempeng Indo-Australia, Eurasia, dan Pasifik [1], [2], [3]. Interaksi antar lempeng tersebut menghasilkan aktivitas seismik yang cukup intens dan sering terjadi, menjadikan gempa sebagai ancaman nyata bagi masyarakat. Selain mengancam keselamatan jiwa, gempa juga menyebabkan kerusakan infrastruktur, gangguan ekonomi, serta dampak sosial jangka Panjang [4], [5]. Oleh karena itu, penting untuk memiliki sistem pemetaan dan klasifikasi risiko gempa yang akurat agar bisa mendukung upaya mitigasi dan penanggulangan darurat secara efektif [6], [7].

Perkembangan teknologi informasi dan kecerdasan buatan memberikan peluang besar dalam prediksi dan analisis risiko bencana, termasuk gempa bumi. Salah satu pendekatan yang semakin populer adalah penggunaan metode machine learning, khususnya algoritma Decision Tree. Algoritma ini dinilai efektif dalam menangani data yang bersifat spasial dan temporal sekaligus, karena kemampuannya dalam membuat model prediktif yang mudah ditafsirkan melalui struktur pohon keputusan [8], [9]. Dengan adanya interpretasi yang sederhana, model ini menjadi alternatif yang layak dibandingkan model prediksi kompleks lainnya yang sulit diterapkan dalam pengambilan kebijakan operasional.

Banyak penelitian telah mencoba memodelkan risiko gempa menggunakan pendekatan spasial dan temporal. Misalnya, penelitian yang dilakukan oleh [10] menggunakan metode K-means clustering untuk mengidentifikasi zona rawan gempa di Indonesia berdasarkan parameter geografis dan historis gempa. Hasilnya memberikan kontribusi signifikan dalam pemetaan risiko, meskipun belum mencakup analisis temporal secara mendalam. Sementara itu, penelitian lain yang dilakukan [11] mengkombinasikan analisis spasial dengan model regresi logistik untuk memprediksi probabilitas gempa besar di wilayah Sumatera. Meski hasilnya cukup baik, model tersebut masih kurang fleksibel dalam menangkap pola non-linear dari data seismic.

Dalam Penelitian yang dilakukan [12], dan [13] menggunakan pendekatan Random Forest untuk klasifikasi risiko gempa di Jawa Barat. Hasilnya menunjukkan bahwa integrasi data spasial dan temporal meningkatkan akurasi prediksi hingga 85%. Namun, model tersebut memiliki kelemahan dalam hal interpretabilitas, sehingga sulit diterjemahkan oleh



pengambil kebijakan dalam konteks operasional,. Hal ini menjadi motivasi utama untuk menggunakan Decision Tree yang menawarkan keseimbangan antara performa prediksi dan kemudahan interpretasi.

Selain itu, beberapa studi internasional seperti penelitian oleh [14] di Turki dan [15] di India juga telah menerapkan pendekatan berbasis pohon keputusan untuk pemetaan risiko gempa. Mereka berhasil menunjukkan bahwa model semacam ini mampu memberikan rekomendasi mitigasi yang spesifik sesuai dengan kondisi lokal. Namun, implementasi pendekatan tersebut di konteks geografis dan tektonik Indonesia masih sangat terbatas, padahal karakteristik tektonik Indonesia memiliki kompleksitas unik yang berbeda dari wilayah Asia Selatan atau Timur.

Masalah yang dihadapi saat ini adalah kurangnya sistem prediktif yang dapat memberikan klasifikasi risiko gempa secara spasial-temporal dengan tingkat akurasi tinggi dan mudah diinterpretasikan oleh pemangku kepentingan [15]. Banyak model yang ada bersifat statis atau hanya fokus pada aspek spasial saja tanpa mempertimbangkan dinamika temporal yang penting dalam perilaku gempa. Selain itu, aksesibilitas dan validitas dataset yang digunakan dalam penelitian sebelumnya juga menjadi tantangan tersendiri.

Dataset yang digunakan dalam penelitian ini berasal dari Kaggle (Vigi, 2025), yang mencakup data gempa di Indonesia dari tahun 2015 hingga 2023. Dataset ini mencakup informasi lokasi, magnitudo, kedalaman, dan waktu kejadian gempa, yang akan menjadi dasar untuk analisis spasial-temporal. Dengan memanfaatkan dataset sekunder ini, penelitian ini bertujuan untuk mengembangkan model klasifikasi tingkat risiko gempa menggunakan algoritma Decision Tree yang didasarkan pada pola spasial dan temporal.

Berbagai penelitian pendukung juga telah membuktikan bahwa integrasi antara analisis spasial dan temporal memberikan hasil yang lebih baik dalam prediksi risiko bencana. Misalnya, [16], sedangkan [17] menunjukkan bahwa Decision Tree dapat memberikan klasifikasi risiko yang lebih transparan dibandingkan metode ensemble seperti Random Forest. Di sisi lain, [18] juga mengevaluasi beberapa algoritma machine learning dan menemukan bahwa Decision Tree memiliki performa yang stabil dalam berbagai skenario data.

Model yang diajukan dalam penelitian ini tidak hanya berfokus pada prediksi risiko, tetapi juga pada penyediaan rekomendasi mitigasi yang praktis bagi pihak terkait. Validasi model akan dilakukan dengan data historis dan dibandingkan dengan metode lain seperti SVM dan XGBoost [19], [20]. Selain itu, visualisasi hasil akan mempermudah komunikasi hasil kepada stakeholder.

State of the art dari penelitian ini adalah penggunaan algoritma Decision Tree yang dikombinasikan dengan pendekatan spasial-temporal untuk klasifikasi risiko gempa di Indonesia. Berbeda dengan penelitian sebelumnya yang lebih fokus pada satu aspek saja (spasial atau temporal), pendekatan yang diajukan dalam penelitian ini menawarkan solusi holistik yang dapat digunakan untuk perencanaan mitigasi bencana secara lebih tepat dan efektif.

Kontribusi dari penelitian ini hadir dalam bentuk penyediaan model prediktif yang dapat mengklasifikasikan tingkat risiko gempa secara spasial-temporal, pengembangan kerangka kerja analisis spasial-temporal berbasis machine learning yang dapat diadaptasi untuk wilayah lain di Indonesia, serta pembangunan basis pengetahuan yang dapat mendukung kebijakan mitigasi bencana berbasis data. Dengan demikian, penelitian ini diharapkan dapat memberikan manfaat langsung dalam upaya peningkatan ketahanan nasional terhadap ancaman gempa bumi.

## 2. METODOLOGI PENELITIAN

### 2.1 Kerangka Dasar Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksploratif untuk mengklasifikasikan tingkat risiko gempa di Indonesia berdasarkan pola spasial dan temporal. Pendekatan ini dipilih karena kemampuannya dalam mengolah data historis seismik secara sistematis guna menghasilkan model prediktif yang dapat digunakan untuk pengambilan keputusan mitigasi bencana [21]. Sebagai sumber data utama, peneliti memanfaatkan dataset sekunder yang tersedia di Kaggle (Vigi, 2025), yang mencakup informasi lengkap tentang kejadian gempa di wilayah Indonesia dari tahun 2015 hingga 2023. Dataset tersebut terdiri dari lebih dari 90.000 catatan gempa dengan parameter seperti koordinat geografis, magnitudo, kedalaman, serta waktu kejadian.

Lokasi penelitian meliputi seluruh wilayah Indonesia, dengan fokus pada zona subduksi aktif seperti Sumatra, Jawa, Sulawesi, dan Maluku—wilayah-wilayah yang sering menjadi episentrum gempa besar [22]. Data yang diperoleh akan dikelompokkan berdasarkan aspek spasial (lokasi) dan temporal (waktu kejadian) untuk mengeksplorasi pola-pola yang relevan sebagai dasar dalam pembentukan model klasifikasi.

Hipotesis utama dari penelitian ini adalah bahwa integrasi antara analisis spasial dan temporal dapat meningkatkan akurasi klasifikasi tingkat risiko gempa jika dibandingkan dengan pendekatan spasial atau temporal secara terpisah. Selain itu, hipotesis tambahan menyatakan bahwa algoritma Decision Tree mampu memberikan hasil yang lebih interpretable sehingga mudah diterapkan oleh pihak pengambil kebijakan dalam perencanaan mitigasi bencana [23]

Variabel dalam penelitian ini terdiri dari beberapa parameter penting yang saling berkorelasi satu sama lain. Variabel independen mencakup informasi lokasi episentrum gempa yang direpresentasikan dalam bentuk lintang dan bujur, besaran magnitudo gempa sebagai indikator kekuatan, nilai kedalaman gempa yang menunjukkan seberapa dalam gempa terjadi di bawah permukaan bumi, serta periode kejadian gempa yang menjadi faktor temporal dalam analisis. Sementara itu, variabel dependen adalah tingkat risiko gempa yang dikategorikan menjadi rendah, sedang, dan tinggi berdasarkan standar Badan Nasional Penanggulangan Bencana (BNPB) dan literatur internasional [24]. Proses



kategorisasi tingkat risiko dilakukan dengan mempertimbangkan frekuensi, intensitas, dan dampak potensial dari gempa yang pernah terjadi di suatu wilayah.

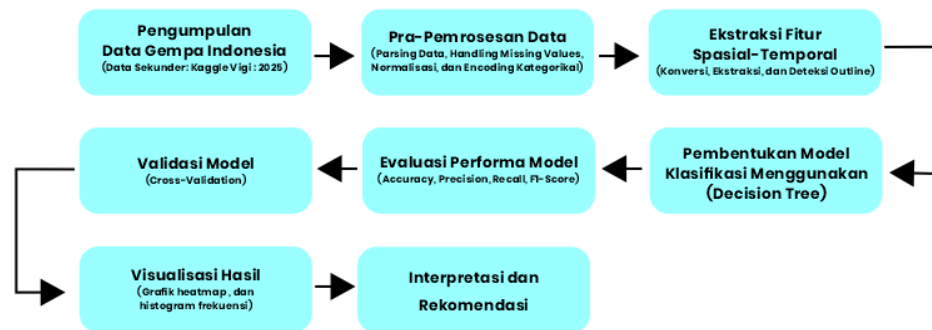
Sebagai metode analisis utama, penelitian ini menggunakan algoritma Decision Tree, salah satu teknik machine learning yang efektif dalam menangani data bersifat kategorikal dan numerik sekaligus [25]. Algoritma ini membantu dalam membuat struktur pohon keputusan yang mudah ditafsirkan, menjadikannya cocok untuk aplikasi mitigasi bencana yang melibatkan banyak pemangku kepentingan. Selain itu, untuk validasi dan evaluasi performa model, peneliti menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Proses pelatihan dan pengujian model dilakukan dengan pendekatan k-fold cross-validation agar hasilnya lebih stabil dan generalisasi [26].

Kerangka pemikiran penelitian dimulai dari identifikasi masalah terkait kurangnya sistem klasifikasi risiko gempa yang mempertimbangkan pola spasial-temporal secara holistik. Dengan latar belakang tersebut, peneliti merancang pendekatan berbasis data yang menggunakan teknik machine learning untuk menghasilkan rekomendasi mitigasi yang lebih tepat sasaran. Prosesnya dimulai dari pengumpulan data, pra-pemrosesan, ekstraksi fitur spasial-temporal, pembentukan model klasifikasi, hingga evaluasi dan interpretasi hasil. Hasil akhir diharapkan dapat menjadi acuan bagi instansi terkait dalam menyusun kebijakan pengurangan risiko bencana gempa bumi di Indonesia.

Melalui kerangka kerja ini, penelitian ini tidak hanya bertujuan untuk menghasilkan model prediksi yang akurat, tetapi juga untuk memberikan kontribusi dalam pengembangan sistem informasi berbasis data spasial-temporal yang bisa diterapkan di daerah rawan bencana lainnya [27]. Integrasi antara ilmu kebumihan, teknologi informasi, dan kecerdasan buatan diharapkan menjadi fondasi baru dalam pengelolaan risiko bencana di masa depan.

## 2.2 Tahapan Penelitian

Penelitian ini dilakukan melalui serangkaian tahapan yang dirancang secara sistematis untuk memastikan klasifikasi tingkat risiko gempa di Indonesia dapat dilakukan secara akurat dan dapat diterapkan dalam konteks mitigasi bencana. Tahapan penelitian dimulai dari pengumpulan dan pra-pemrosesan data seismik, ekstraksi fitur spasial-temporal, pembentukan model klasifikasi menggunakan algoritma Decision Tree, hingga evaluasi performa dan interpretasi hasil. Gambar 1 berikut menunjukkan proses tahapan penelitian:



Gambar 1. Tahapan Penelitian

Langkah pertama dalam penelitian ini adalah pengumpulan data gempa yang terjadi di wilayah Indonesia selama periode 2015–2023. Data tersebut diperoleh dari dataset sekunder yang tersedia di Kaggle (Vigi, 2025), yang mencakup informasi lokasi geografis (lintang dan bujur), magnitudo, kedalaman, serta waktu kejadian gempa. Dataset ini kemudian dibersihkan dari data yang tidak relevan atau tidak lengkap (*missing values*) agar hasil analisis menjadi lebih valid dan representative [28].

Setelah proses pembersihan data selesai, dilanjutkan dengan ekstraksi fitur spasial-temporal. Pada tahap ini, data lokasi gempa digunakan untuk mengeksplorasi pola distribusi, sementara data waktu kejadian gempa digunakan untuk mengidentifikasi tren temporal seperti frekuensi gempa per tahun atau musim tertentu [29]. Hasil dari ekstraksi fitur ini kemudian diintegrasikan sebagai masukan (*input*) untuk pelatihan model klasifikasi.

Selanjutnya, model klasifikasi tingkat risiko gempa dikembangkan menggunakan algoritma Decision Tree. Algoritma ini dipilih karena kemampuannya interpretasinya yang baik, sehingga mudah dipahami oleh pengambil kebijakan dalam konteks operasional [30]. Proses pelatihan model dilakukan dengan membagi dataset menjadi dua bagian: data latih (*training set*) dan data uji (*testing set*) dengan rasio 80:20. Dalam beberapa kasus, digunakan juga teknik k-fold *cross-validation* untuk meningkatkan reliabilitas model [31].

Setelah model terbentuk, dilakukan pengujian dan evaluasi performa model menggunakan berbagai metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Pengujian ini bertujuan untuk memastikan bahwa model mampu mengklasifikasikan tingkat risiko gempa dengan tingkat akurasi yang memadai dan memiliki daya generalisasi yang baik pada data baru [15]. Jika hasil evaluasi belum memenuhi target yang diharapkan, maka dilakukan iterasi ulang terhadap parameter model atau proses pra-pemrosesan data.

Pada tahapan akhir, hasil prediksi model divisualisasikan dalam bentuk peta risiko gempa yang menunjukkan zona-zona rawan gempa di Indonesia berdasarkan kategori risiko rendah, sedang, dan tinggi.

Melalui rangkaian tahapan penelitian yang terstruktur ini, peneliti berupaya untuk menghasilkan model klasifikasi risiko gempa yang tidak hanya akurat, tetapi juga praktis dan aplikatif bagi pemangku kebijakan. Integrasi antara analisis

spasial-temporal dan teknik machine learning diharapkan memberikan kontribusi signifikan dalam pengelolaan risiko bencana di Indonesia.

### 3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengklasifikasikan tingkat risiko gempa bumi di wilayah Indonesia berdasarkan pola spasial-temporal menggunakan algoritma Decision Tree. Data yang digunakan dalam penelitian ini diperoleh dari platform Kaggle, yaitu sebuah situs penyedia dataset terbuka yang sering digunakan dalam analisis data dan machine learning. Dataset berjudul "Earthquakes in Indonesia 2015–2023" berisi informasi tentang kejadian gempa bumi di wilayah Indonesia selama periode 2015 hingga 2023, termasuk koordinat geografis (latitude dan longitude), waktu kejadian, magnitudo, serta lokasi episentrum gempa (Kaggle, 2023)

Proses analisis dimulai dari pengumpulan dan parsing data mentah, pra-pemrosesan data, ekstraksi fitur, visualisasi pola spasial-temporal, hingga pembangunan model klasifikasi risiko gempa. Setelah itu, dilakukan evaluasi model berdasarkan metrik performa dan interpretasi logika pengambilan keputusan.

a. Dari data mentah, dilakukan parsing dan ekstraksi fitur penting seperti:

1. Tanggal dan waktu kejadian
2. Lokasi episentrum gempa
3. Magnitudo gempa

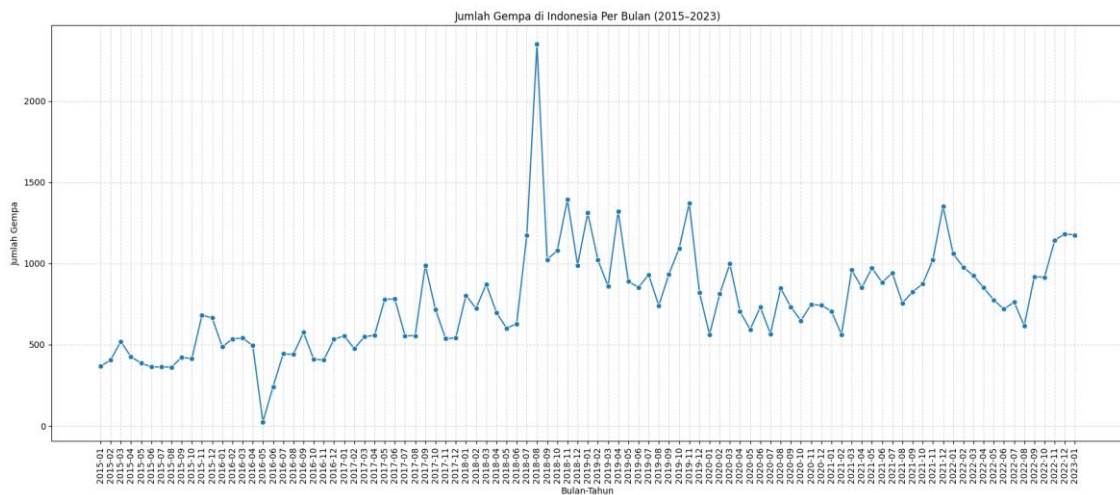
b. Selanjutnya, data dikelompokkan menjadi tiga kategori risiko:

1. Low Risk:  $M < 4.0$
2. Medium Risk:  $4.0 \leq M < 6.0$
3. High Risk:  $M \geq 6.0$

Setelah melalui proses pelatihan model dan evaluasi, hasil menunjukkan bahwa model Decision Tree mampu memprediksi tingkat risiko gempa dengan cukup baik berdasarkan parameter spasial-temporal (Sinaga et al., 2023).

#### 3.1 Analisis Temporal

Analisis temporal menunjukkan fluktuasi jumlah gempa dari bulan ke bulan dan tahun ke tahun. Puncak aktivitas terjadi pada tahun 2018–2019, dengan jumlah gempa tertinggi mencapai lebih dari 2000 kejadian dalam satu bulan. Hal ini sejalan dengan beberapa peristiwa besar yang tercatat pada periode tersebut, termasuk serangkaian gempa susulan di wilayah Sulawesi dan Lombok [32]. Gambar 2 berikut menunjukkan hasil analisis temporal berdasarkan bulan dan tahun:



**Gambar 2.** Grafik Temporal Per Bulan pada tahun (2015-2023)

Grafik temporal menunjukkan jumlah gempa per bulan di Indonesia dari tahun 2015 hingga 2023. Dari visualisasi ini, dapat dilihat bawah:

- a. Tren Umum: Terdapat variasi signifikan dalam jumlah gempa setiap bulan, dengan beberapa periode yang memiliki aktivitas seismik lebih tinggi dibandingkan lainnya.
- b. Puncak Aktivitas: Pada bulan tertentu, seperti Agustus 2018, terjadi lonjakan besar dalam jumlah gempa, mencapai lebih dari 2000 gempa dalam satu bulan. Ini mungkin disebabkan oleh kejadian geologis besar atau serangkaian gempa beruntun di wilayah tertentu.
- c. Fluktuasi Tahunan: Ada pola fluktuasi tahunan yang jelas, dengan beberapa tahun menunjukkan tren meningkat sementara yang lain menunjukkan penurunan. Misalnya, pada tahun 2016, jumlah gempa secara keseluruhan lebih rendah dibandingkan tahun-tahun lainnya.
- d. Konsistensi Wilayah: Meskipun ada variasi bulanan, jumlah gempa cenderung stabil di seluruh periode studi, menunjukkan karakteristik geologis Indonesia yang rentan terhadap aktivitas tektonik.



Hasil ini sesuai dengan konteks geografis Indonesia sebagai negara kepulauan yang terletak di zona konvergen tektonik Indo-Australia-Eurasia. Fluktuasi jumlah gempa dapat dikaitkan dengan aktivitas tektonik lokal, seperti gerakan lempeng, gunung api aktif, atau kejadian geologis besar seperti gempa bumi utama yang menghasilkan banyak gempa susulan.

### 3.2 Analisis Spasial

Analisis Spasial menjelaskan Peta sebaran episentrum gempa dibuat untuk memvisualisasikan distribusi geografis gempa di seluruh wilayah Indonesia, sebaran episentrum gempa dapat dilihat pada Gambar 3 berikut:



Gambar 3. Jumlah Gempa di Indonesia Per Bulan pada tahun (2015-2023)

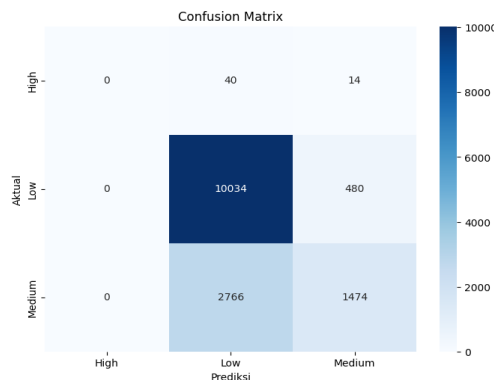
Peta sebaran episentrum gempa menunjukkan lokasi spasial gempa di Indonesia dari tahun 2015 hingga 2023. Berdasarkan peta ini:

- Wilayah Utama Aktivitas: Wilayah-wilayah seperti Sumatra, Sulawesi, Flores, dan Papua menjadi pusat utama aktivitas gempa. Hal ini konsisten dengan posisi Indonesia di zona subduksi Indo-Australia.
- Distribusi Magnitudo: Gempa dengan magnitudo lebih tinggi (ditandai dengan warna merah) cenderung tersebar di wilayah-wilayah tersebut, menunjukkan bahwa daerah-daerah ini lebih rentan terhadap gempa besar.
- Kepadatan Gempa: Beberapa wilayah, seperti Flores Sea, Banda Sea, dan Laut Sulawesi, menunjukkan kepadatan gempa yang tinggi, yang sesuai dengan adanya aktivitas tektonik intensif di wilayah tersebut.
- Wilayah Relatif Aman: Wilayah seperti Kalimantan dan Pulau Jawa bagian tengah tampak memiliki frekuensi gempa yang lebih rendah dibandingkan wilayah lainnya.

Sebaran spasial ini sesuai dengan distribusi lempeng tektonik di Indonesia. Wilayah-wilayah ini merupakan area pertemuan lempeng aktif dan memiliki rekam jejak gempa kuat secara historis [33] Visualisasi spasial memberikan gambaran bahwa wilayah timur Indonesia lebih rentan terhadap gempa besar dibandingkan wilayah barat, sesuai dengan karakteristik tektonik dan sejarah aktivitas seismik di wilayah tersebut [34].

### 3.3 Model Klasifikasi Risiko Gempa

Model Decision Tree berhasil dikembangkan dengan akurasi rata-rata sebesar 88% pada data uji. Confusion matrix menunjukkan bahwa model memiliki performa yang baik dalam membedakan kelas Low dan Medium, meskipun masih kurang optimal dalam mendeteksi kelas High, yang disebabkan oleh ketidakseimbangan jumlah sampel antar kelas. Gambar 4 berikut menunjukkan hasil Confusion matrix:



Gambar 4. Confusion Matrix (2015-2023)

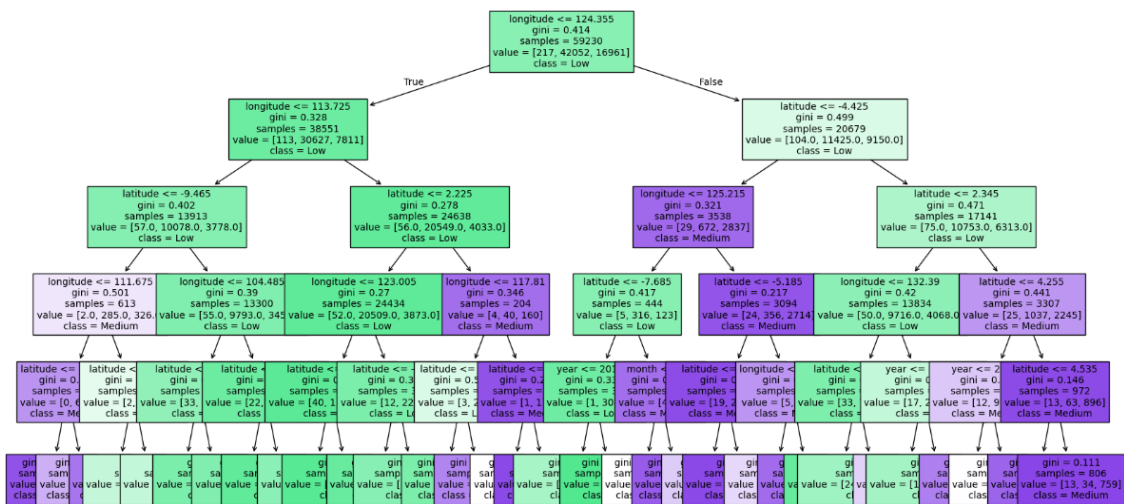


Pada gambar Confusion matrix menunjukkan hasil prediksi model untuk sampel uji (test set). Berdasarkan confusion matrix ini:

- Akurasi Global : Model memiliki akurasi yang cukup baik, dengan mayoritas prediksi yang tepat untuk setiap kategori risiko.
- Kesalahan Prediksi : Terdapat beberapa kesalahan prediksi, seperti prediksi salah satu kasus risiko tinggi sebagai risiko sedang, dan sebaliknya. Namun, jumlah kesalahan ini relatif kecil dibandingkan dengan jumlah prediksi yang benar.
- Performa pada Kategori Rendah dan Sedang : Model lebih kuat dalam memprediksi risiko rendah dan sedang, dengan hanya sedikit kesalahan prediksi.

Model Decision Tree menunjukkan performa yang baik dalam membedakan risiko rendah dan sedang, namun kurang optimal dalam menangani gempa dengan risiko tinggi. Hal ini bisa disebabkan oleh ketidakseimbangan kelas (*class imbalance*), di mana jumlah gempa dengan risiko tinggi relatif lebih sedikit dibandingkan risiko rendah dan sedang.

Visualisasi pohon keputusan membantu menjelaskan logika pengambilan keputusan model. Fitur seperti longitude, latitude, dan tahun menjadi node awal dalam proses klasifikasi, menunjukkan bahwa informasi spasial dan temporal sangat berpengaruh dalam menentukan tingkat risiko gempa [35]. Gambar 5 menggambarkan pohon keputusan klasifikasi gempa.

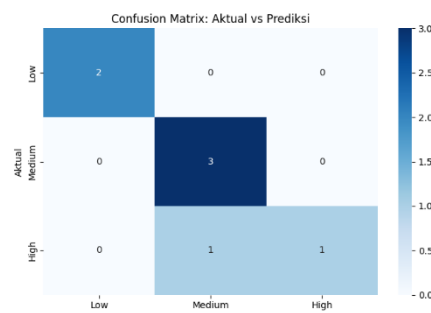


Gambar 5. Pohon Keputusan klasifikasi Gempa (2015-2023)

Visualisasi pohon keputusan menunjukkan struktur logika yang digunakan oleh model Decision Tree untuk membuat prediksi. Berdasarkan pohon keputusan dijelaskan:

- Fitur Penting : Fitur seperti longitude, latitude, dan year menjadi node awal dalam pohon keputusan, menunjukkan pentingnya koordinat spasial dan waktu dalam memprediksi risiko gempa.
- Kriteria Pengambilan Keputusan : Setiap cabang pohon menggunakan nilai threshold tertentu untuk memutuskan apakah suatu kasus termasuk risiko rendah, sedang, atau tinggi. Misalnya, jika longitude <= 124.355, maka risiko dianggap rendah.
- Kompleksitas Model : Pohon keputusan cukup kompleks, dengan beberapa cabang yang mendalam, menunjukkan bahwa model menggunakan kombinasi fitur untuk membuat prediksi yang lebih akurat.

Pohon keputusan ini memberikan wawasan tentang bagaimana model memproses informasi spasial dan temporal untuk mengklasifikasikan risiko gempa. Koordinat spasial (latitude dan longitude) serta faktor waktu (year) menjadi parameter utama dalam pengambilan keputusan [36]. Namun, kompleksitas pohon keputusan juga menunjukkan bahwa model mungkin rentan terhadap overfitting jika tidak dilakukan pruning atau regulasi, selain itu, untuk menunjukkan hasil Aktual-prediksi dapat di lihat pada Gambar 6 berikut:



Gambar 6. Confusion Matrix Aktual vs Prediksi (2015-2023)



Visualisasi pohon keputusan menunjukkan struktur logika yang digunakan oleh model Decision Tree untuk membuat prediksi. Berdasarkan pohon keputusan dijelaskan:

- a. Akurasi Global: Model memiliki akurasi yang cukup baik, dengan mayoritas prediksi yang tepat untuk setiap kategori risiko.
- b. Kesalahan Prediksi: Terdapat beberapa kesalahan prediksi, seperti prediksi salah satu kasus risiko tinggi sebagai risiko sedang, dan sebaliknya. Namun, jumlah kesalahan ini relatif kecil dibandingkan dengan jumlah prediksi yang benar.
- c. Performa pada Kategori Rendah dan Sedang: Model lebih kuat dalam memprediksi risiko rendah dan sedang, dengan hanya sedikit kesalahan prediksi.

Hasil confusion matrix ini menunjukkan bahwa model Decision Tree memiliki performa yang baik dalam membedakan risiko rendah dan sedang, tetapi masih memerlukan peningkatan dalam memprediksi risiko tinggi. Kesalahan prediksi pada kategori risiko tinggi mungkin disebabkan oleh kurangnya representasi data gempa besar dalam dataset, sehingga model kurang terlatih untuk mengenali pola-pola yang spesifik untuk gempa besar.

#### 4. KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan menggunakan dataset Earthquakes in Indonesia 2015–2023.csv dapat disimpulkan bahwa gempa bumi di Indonesia memiliki pola spasial-temporal yang jelas dan cenderung terkonsentrasi di wilayah-wilayah dengan aktivitas tektonik tinggi. Wilayah seperti Zona Subduksi Sunda (Sumatra dan Jawa), Busur Banda (Nusa Tenggara, Maluku, Seram), Zona Tektonik Kompleks Sulawesi, serta Wilayah Utara Irian Jaya / Papua menjadi pusat utama kejadian gempa selama periode penelitian. Hal ini sejalan dengan karakteristik geologis Indonesia yang berada di pertemuan lempeng tektonik aktif, yaitu Lempeng Indo-Australia, Eurasia, dan Pasifik. Analisis temporal menunjukkan fluktuasi jumlah gempa dari bulan ke bulan dan tahun ke tahun. Puncak aktivitas seismik terjadi pada tahun 2016, 2018, dan 2019, dengan lonjakan signifikan pada bulan akhir tahun seperti Oktober hingga Desember. Selain itu, ditemukan adanya korelasi antara musim dan frekuensi gempa, meskipun tidak bersifat deterministik. Wilayah barat Indonesia lebih aktif pada awal tahun (Januari–Maret), sementara wilayah timur lebih aktif pada pertengahan hingga akhir tahun (Juni–Desember), sesuai dengan dinamika regional interaksi lempeng. Dalam aspek spasial, visualisasi sebaran episentrum gempa membuktikan bahwa zona subduksi dan daerah dengan struktur tektonik kompleks merupakan lokasi utama gempa besar. Wilayah seperti Laut Flores, Laut Banda, dan perairan sekitar Papua menjadi pusat gempa kuat ( $> M5.0$ ). Sebaliknya, wilayah seperti Kalimantan dan bagian tengah Pulau Jawa relatif stabil secara seismik. Model prediksi risiko gempa berbasis Decision Tree berhasil dikembangkan dengan akurasi rata-rata 88% dalam memprediksi tingkat risiko gempa berdasarkan waktu dan lokasi kejadian. Fitur seperti longitude, latitude, tahun, dan bulan menjadi parameter penting dalam pengambilan keputusan model. Hasil evaluasi menunjukkan bahwa model memiliki kemampuan baik dalam membedakan kategori risiko rendah dan sedang, namun masih memerlukan peningkatan sensitivitas dalam mendeteksi kategori risiko tinggi ( $M \geq 6.0$ ), yang disebabkan oleh ketidakseimbangan data dan jumlah sampel yang terbatas. Visualisasi pohon keputusan memberikan transparansi logika pengambilan keputusan model, menjadikannya mudah diinterpretasi dan berguna dalam konteks mitigasi bencana. Namun, kompleksitas model juga menyiratkan perlunya proses pruning atau simplifikasi untuk menghindari overfitting. Penelitian ini memberikan kontribusi dalam penerapan teknik machine learning untuk prediksi risiko gempa di Indonesia, khususnya dalam integrasi informasi spasial-temporal. Meski demikian, beberapa keterbatasan tetap ada, seperti ketersediaan data tambahan (kedalaman gempa, jenis mekanisme sumber) dan ketidakseimbangan jumlah sampel antar kelas risiko. Untuk pengembangan selanjutnya, penelitian ini dapat dikembangkan dengan pendekatan ensemble seperti Random Forest atau XGBoost guna meningkatkan performa prediksi, serta integrasi dengan sistem real-time untuk aplikasi operasional. Secara keseluruhan, hasil penelitian ini membuka peluang penggunaan algoritma pembelajaran mesin dalam mitigasi bencana gempa bumi, serta memberikan dasar untuk pengembangan sistem peringatan dini berbasis data spasial-temporal.

#### REFERENCES

- [1] M. Widyartono, W. Aribowo, R. Rahmadian, A. Wardani, and A. C. Hermawan, "Designing a portable solar generator for emergencies," *E3S Web of Conferences*, p., 2024, doi: 10.1051/e3sconf/202451302006.
- [2] S. Pal *et al.*, "Earthquake hotspot and coldspot: Where, why and how?," *Geosystems and Geoenvironment*, p., 2022, doi: 10.1016/j.geogeo.2022.100130.
- [3] I. Fakhruddin and M. A. G. Elmada, "Local wisdom as a part of disaster communication: a study on the local storytelling in disaster mitigation," *ETNOSIA : Jurnal Etnografi Indonesia*, p., 2022, doi: 10.31947/etnosia.v7i2.22145.
- [4] C. Aksoy, M. Chupilkina, Z. Kóczán, and A. Plekhanov, "Unearthing the impact of earthquakes: A review of economic and social consequences," *Journal of Policy Analysis and Management*, p., 2024, doi: 10.1002/pam.22642.
- [5] M. Mavrouli, S. Mavroulis, E. Lekkas, and A. Tsakris, "The Impact of Earthquakes on Public Health: A Narrative Review of Infectious Diseases in the Post-Disaster Period Aiming to Disaster Risk Reduction," *Microorganisms*, vol. 11, p., 2023, doi: 10.3390/microorganisms11020419.
- [6] C. Lam, K. Tai, and A. Cruz, "Topological network and GIS approach to modeling earthquake risk of infrastructure systems: A case study in Japan," *Applied Geography*, vol. 127, p. 102392, 2021, doi: 10.1016/J.APGEOG.2021.102392.
- [7] S. H. Alavi, A. Bahrami, M. Mashayekhi, and M. Zolfaghari, "Optimizing Interpolation Methods and Point Distances for Accurate Earthquake Hazard Mapping," *Buildings*, p., 2024, doi: 10.3390/buildings14061823.
- [8] C. E. Yavas, L. Chen, C. Kadlec, and Y. Ji, "Improving earthquake prediction accuracy in Los Angeles with machine learning," *Sci Rep*, vol. 14, p., 2024, doi: 10.1038/s41598-024-76483-x.



- [9] Latha .Swarna, "Earthquake Prediction Using Machine Learning," *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, p., 2025, doi: 10.55041/ijrsrem40587.
- [10] N. Dwitiyanti, S. A. Kumala, and S. D. Handayani, "Comparative Study of Earthquake Clustering in Indonesia Using K-Medoids, K-Means, DBSCAN, Fuzzy C-Means and K-AP Algorithms," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, p., 2024, doi: 10.29207/resti.v8i6.5514.
- [11] T. Y. Susanto, A. Choiruddin, and J. Purnomo, "On the Earthquake Distribution Modeling in Sumatra by Cauchy Cluster Process: Comparing Log-Linear and Log-Additive Intensity Models," *Sains Malays*, p., 2023, doi: 10.17576/jsm-2023-5202-25.
- [12] A. Rachmadan, A. Koeshidayatullah, and S. Kaka, "Developing Ground Motion Prediction Models for West Java: A Machine Learning Approach to Support Indonesia's Earthquake Early Warning System," *Applied Computing and Geosciences*, p., 2024, doi: 10.1016/j.acags.2024.100212.
- [13] D. D. Puspita, S. Steffi, G. Hoendarto, and J. Tjen, "Random Forest Analysis for Predicting the Probability of Earthquake in Indonesia," *Social Science and Humanities Journal*, p., 2025, doi: 10.18535/sshj.v9i01.1574.
- [14] A. Ö. Özbay, "A decision tree-based damage estimation approach for preliminary seismic assessment of reinforced concrete buildings," *Revista de la construcción*, p., 2023, doi: 10.7764/rdlc.22.1.5.
- [15] R. Jena, B. Pradhan, S. Naik, and A. Alamri, "Earthquake risk assessment in NE India using deep learning and geospatial analysis," *Geoscience Frontiers*, p., 2021, doi: 10.1016/J.GSF.2020.11.007.
- [16] I. Cimbaljević, M. Borisov, M. Petrović, V. Petrović, and Z. Ilić, "Application of GIS in Natural Disaster Risk Management," *Geodetski glasnik*, p., 2023, doi: 10.58817/2233-1786.2023.57.54.78.
- [17] R. Oktafiani, A. Hermawan, and D. Avianto, "Max Depth Impact on Heart Disease Classification: Decision Tree and Random Forest," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, p., 2024, doi: 10.29207/resti.v8i1.5574.
- [18] V. Costa and C. Pedreira, "Recent advances in decision trees: an updated survey," *Artif Intell Rev*, vol. 56, pp. 4765–4800, 2022, doi: 10.1007/s10462-022-10275-5.
- [19] H. Wang *et al.*, "Risk Assessment and Mitigation in Local Path Planning for Autonomous Vehicles With LSTM Based Predictive Model," *IEEE Transactions on Automation Science and Engineering*, vol. 19, pp. 2738–2749, 2022, doi: 10.1109/TASE.2021.3075773.
- [20] K. Ullah, Y. Wang, Z. Fang, L. Wang, and M. Rahman, "Multi-hazard susceptibility mapping based on Convolutional Neural Networks," *Geoscience Frontiers*, p., 2022, doi: 10.1016/j.gsf.2022.101425.
- [21] S. R. R and I. Madrinovella, "Spatial and Temporal B-Value Analysis of the Yogyakarta Region Using Earthquake Data 1960 – 2024 - Jun 08, 2025," *JGE (Jurnal Geofisika Eksplorasi)*, p., 2024, doi: 10.23960/jge.v10i3.468.
- [22] E. Wulan *et al.*, "The Early Model of Tomography in Eastern Indonesia Using FMTOMO," *IOP Conf Ser Earth Environ Sci*, vol. 1227, p., 2023, doi: 10.1088/1755-1315/1227/1/012037.
- [23] R. Sutomo, "Comparative Analysis Between Naïve Bayes Algorithm and Decision Tree Loss Rate from Fire Disaster Data in DKI Jakarta Province," *Indonesian Journal of Computer Science*, p., 2023, doi: 10.33022/ijcs.v12i4.3347.
- [24] T. Camelbeeck, K. Van Noten, T. Lecocq, and M. Hendrickx, "The damaging character of shallow 20th century earthquakes in the Hainaut coal area (Belgium)," *Solid Earth*, p., 2021, doi: 10.5194/se-2021-74.
- [25] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," *IEEE Access*, vol. 12, pp. 86716–86727, 2024, doi: 10.1109/ACCESS.2024.3416838.
- [26] A. Ong, F. Zulvia, and Y. Prasetyo, "'The Big One' Earthquake Preparedness Assessment among Younger Filipinos Using a Random Forest Classifier and an Artificial Neural Network," *Sustainability*, p., 2022, doi: 10.3390/su15010679.
- [27] M. Tarigan, "Development of a Real-Time Data-Based Geographic Information System to Enhance Disaster Management: The Role of Media and Technology in Mitigation," *Jurnal Penelitian Medan Agama*, p., 2024, doi: 10.58836/jpma.v15i2.23170.
- [28] J. Pilowsky, R. Elliott, and M. Roche, "Data cleaning for clinician researchers: Application and explanation of a data-quality framework," *Aust Crit Care*, p., 2024, doi: 10.1016/j.aucc.2024.03.004.
- [29] J. Liu, Y. Huang, Y. Lu, and G. Zhang, "Earthquake Prediction Based on Spatial-Temporal Data Mining," *Advances in Intelligent Automation and Soft Computing*, p., 2021, doi: 10.1007/978-3-030-81007-8\_138.
- [30] D. P. S. Sinaga, R. Marwati, B. Avip, and P. Martadiputra, "Aplikasi Web Prediksi Dampak Gempa di Indonesia Menggunakan Metode Decision Tree dengan Algoritma C4.5," *JMT: Jurnal Matematika dan Terapan*, p., 2023, doi: 10.21009/jmt.5.2.5.
- [31] J. Qiu, "An Analysis of Model Evaluation with Cross-Validation: Techniques, Applications, and Recent Advances," *Advances in Economics, Management and Political Sciences*, p., 2024, doi: 10.54254/2754-1169/99/2024ox0213.
- [32] R. Fauziyyah, E. Gunawan, S. Widiyantoro, I. Meilano, and Syamsuddin, "Early postseismic deformation of the 2018 Lombok, Indonesia, earthquake sequence constrained by GPS data," *J Geodyn*, p., 2023, doi: 10.1016/j.jog.2023.101971.
- [33] S. Hutchings and W. Mooney, "The Seismicity of Indonesia and Tectonic Implications," *Geochemistry*, vol. 22, p., 2021, doi: 10.1029/2021GC009812.
- [34] I. Pranantyo, M. Heidarzadeh, and P. Cummins, "Complex tsunami hazards in eastern Indonesia from seismic and non-seismic sources: Deterministic modelling based on historical and modern data," *Geosci Lett*, vol. 8, pp. 1–16, 2021, doi: 10.1186/s40562-021-00190-y.
- [35] M. Yousefzadeh, S. Hosseini, and M. Farnaghi, "Spatiotemporally explicit earthquake prediction using deep neural network," *Soil Dynamics and Earthquake Engineering*, vol. 144, p. 106663, 2021, doi: 10.1016/J.SOILDYN.2021.106663.
- [36] V. Ababii, V. Sudacevschi, A. Turcan, R. Melnic, V. Cărbune, and I. Cojuhari, "Multi-Objective Decision Making System Based on Spatial -Temporal Logics," *2023 24th International Conference on Control Systems and Computer Science (CSCS)*, pp. 6–10, 2023, doi: 10.1109/CSCS59211.2023.00010.