

Pengaruh Optimasi *Hyperparameter* Random Forest terhadap Akurasi Prediksi Magnitudo Gempa Bumi Berdasarkan Hasil Klasterisasi DBSCAN

Rizky Dwi Prasetyo^{1*}, Nadia Anissa Maori², Akhmad Khanif Zyen³

^{1*,2,3} Program Studi Tekni Informatika, Fakultas Sains dan Teknologi, Universitas Islam Nabdlatul Ulama Jepara, Kota Jepara, Provinsi Jawa Tengah, Indonesia.

article info

Article history:

Received 10 September 2025
Received in revised form
20 September 2025
Accepted 1 November 2025
Available online January
2026.

Keywords:

Earthquake; DBSCAN;
Random Forest; Magnitude
Prediction; Machine Learning;
Hyperparameter Tuning.

Kata Kunci:

Gempa Bumi; DBSCAN;
Random Forest; Prediksi
Magnitudo; Machine Learning;
Hyperparameter Tuning.

abstract

Indonesia is a country with high seismic activity due to its location at the convergence of three major tectonic plates. This condition creates a strong need for earthquake pattern analysis and magnitude prediction to support disaster mitigation. This study aims to cluster earthquake data using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm and to predict earthquake magnitude using the Random Forest algorithm optimized through hyperparameter tuning. The Indonesian earthquake dataset was obtained from Kaggle with a total of 92,887 valid entries. The DBSCAN clustering results revealed several active seismic zones, particularly in Sumatra, Java, Sulawesi, and Papua. The comparison of R^2 between the Baseline Random Forest and the Tuned Random Forest shows a significant improvement after the parameter tuning process. The Tuned Random Forest model achieves an R^2 value of 0.478, which is higher than the Baseline Random Forest's 0.442. This indicates that the tuned model is better able to explain the variance in the data and provides more accurate predictions.

abstrak

Indonesia merupakan negara dengan tingkat aktivitas seismik yang tinggi karena berada pada pertemuan tiga lempeng besar dunia. Kondisi ini menimbulkan kebutuhan akan analisis pola gempa dan prediksi magnitudo untuk mendukung mitigasi bencana. Penelitian ini bertujuan untuk mengelompokkan data gempa menggunakan algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN) serta memprediksi magnitudo gempa dengan algoritma Random Forest yang dioptimasi melalui hyperparameter tuning. Dataset gempa Indonesia diperoleh dari Kaggle dengan jumlah 92.887 entri data valid. Hasil klasterisasi DBSCAN menunjukkan terbentuknya beberapa zona seismik aktif, terutama di Sumatra, Jawa, Sulawesi, dan Papua. Hasil perbandingan R^2 antara Baseline Random Forest dan Tuned Random Forest menunjukkan peningkatan yang signifikan setelah proses tuning parameter. Model Tuned Random Forest mencapai nilai R^2 sebesar 0.478, lebih tinggi dibandingkan Baseline Random Forest yang sebesar 0.442. Hal ini mengindikasikan bahwa model yang telah di-tuning mampu menjelaskan variasi data dengan lebih baik dan memberikan prediksi yang lebih akurat.

Corresponding Author. Email: riskydi118@gmail.com^{1}.

1. Pendahuluan

Gempa bumi merupakan salah satu bentuk bencana alam dengan dampak luas terhadap kehidupan manusia dan infrastruktur. Selain menimbulkan kerusakan fisik, peristiwa ini sering memicu bencana turunan seperti longsor dan likuifaksi (Rahayu & Purwoko, 2020). Berdasarkan data Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), ribuan gempa tercatat setiap tahun di wilayah Indonesia, terutama di jalur subduksi aktif Sumatra, Jawa, Sulawesi, dan Papua. Frekuensi tinggi kejadian tersebut menegaskan perlunya pendekatan berbasis data untuk memahami pola persebaran dan memprediksi magnitudo guna mendukung mitigasi risiko bencana. Secara geotektonik, Indonesia terletak pada pertemuan tiga lempeng utama dunia Indo-Australia, Eurasia, dan Pasifik yang menyebabkan tingginya aktivitas seismik di wilayah ini. Data dari BMKG dan dataset terbuka di Kaggle yang berisi 92.887 entri gempa menunjukkan konsistensi pola aktivitas tersebut. Analisis kuantitatif berbasis pemodelan spasial diperlukan agar potensi bahaya dapat diidentifikasi dengan presisi, sebagaimana dibuktikan oleh penelitian Pamungkas *et al.* (2024) yang menunjukkan bahwa pemetaan berbasis parameter seismik mampu memperjelas distribusi risiko.

Sejumlah penelitian terdahulu telah memanfaatkan *machine learning* untuk memprediksi dan memetakan aktivitas gempa. Kurniawan *et al.* (2024) menerapkan algoritma Random Forest dengan tingkat akurasi 87%, meskipun belum mempertimbangkan dimensi spasial. Sebaliknya, Syaifuddin & Prabowo (2022) menggunakan DBSCAN untuk menentukan zona seismik, namun tanpa model prediktif magnitudo. Budiman & Ifriza (2021) menegaskan bahwa parameter spasial seperti kedalaman dan lokasi episentrum memiliki peran penting dalam peningkatan akurasi model prediksi. Sebagian besar penelitian masih menitikberatkan pada satu pendekatan saja, baik pada klusterisasi spasial maupun prediksi magnitudo, tanpa integrasi menyeluruh antara keduanya. Karena itu, diperlukan metode yang mampu menggabungkan klusterisasi berbasis kepadatan dengan model prediksi berbasis *ensemble learning* yang dioptimalkan melalui penyesuaian parameter (*hyperparameter tuning*).

Integrasi ini berpotensi meningkatkan ketepatan model serta memberikan hasil analitik yang lebih representatif terhadap fenomena seismik. Penelitian juga memanfaatkan ekosistem Python untuk menampilkan hasil analisis secara interaktif menggunakan *Streamlit* (Rahman *et al.*, 2023). Tujuan utama penelitian adalah mengembangkan pendekatan terpadu antara DBSCAN untuk identifikasi zona seismik aktif dan Random Forest dengan optimasi *hyperparameter* untuk prediksi magnitudo. Dengan menggunakan 92.887 data gempa dari Kaggle, penelitian ini diharapkan menghasilkan model analisis seismik yang lebih presisi dan relevan bagi upaya mitigasi bencana di Indonesia.

2. Metodologi Penelitian

Penelitian ini mengadopsi pendekatan kuantitatif berbasis eksperimen komputasional dengan tujuan untuk membangun, melatih, dan mengevaluasi model *machine learning* menggunakan data historis gempa bumi. Eksperimen dilakukan dengan membandingkan performa model *Random Forest* sebelum dan sesudah dilakukan *hyperparameter tuning*, untuk menilai pengaruh optimasi terhadap akurasi prediksi magnitudo berdasarkan hasil klusterisasi yang dilakukan menggunakan algoritma DBSCAN. Pelaksanaan penelitian dilakukan secara daring dengan memanfaatkan platform Google Colab sebagai lingkungan komputasi berbasis *cloud*. Data yang digunakan dalam penelitian ini diperoleh dari dua sumber utama, yaitu portal resmi BMKG dan platform Kaggle. Penelitian ini berlangsung pada periode Mei hingga Juli 2025, dan karena sifatnya yang berbasis komputasi dan data, tidak mengharuskan lokasi fisik tertentu untuk pelaksanaannya. Populasi penelitian mencakup seluruh catatan gempa bumi yang tercatat di Indonesia antara tahun 2008 hingga 2024 yang didokumentasikan oleh BMKG. Sampel dipilih secara *purposive* dengan mempertimbangkan kelengkapan dan validitas data, yang mencakup atribut waktu, koordinat geografis, kedalaman, dan magnitudo. Data yang tidak lengkap, mengandung nilai hilang (*missing values*), atau terindikasi sebagai *outlier* dihapus pada tahap *preprocessing*. Setelah dilakukan penyaringan, diperoleh 92.887 entri data gempa yang siap digunakan untuk analisis klusterisasi dan pemodelan prediksi.

Sebagaimana yang dijelaskan oleh Rendy Swanda Narastu *et al.* (2023), analisis *exploratory data analysis (EDA)* pada data gempa Indonesia dari BMKG sangat krusial untuk memahami distribusi spasial dan temporal sebelum melanjutkan ke analisis lebih lanjut. Instrumen penelitian ini mencakup studi literatur untuk memperkuat dasar teoretis, serta analisis data berbasis komputasi untuk membangun metodologi penelitian. Studi literatur dilakukan dengan menelaah publikasi yang relevan mengenai algoritma DBSCAN, *Random Forest*, serta teknik optimasi *hyperparameter* untuk memperkuat landasan teoritis penelitian ini. Analisis data dilakukan melalui beberapa tahapan: dimulai dengan eksplorasi data seismik, dilanjutkan dengan penerapan DBSCAN untuk mengidentifikasi zona seismik aktif, hingga pembangunan model prediksi menggunakan *Random Forest* dengan dua skenario, yaitu baseline dan yang telah dioptimasi (*tuning*). Evaluasi kinerja model dilakukan dengan menggunakan metrik *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, dan *Coefficient of Determination (R^2)*. Pengumpulan data dilakukan dengan menggunakan metode dokumentasi, yaitu mengakses dan mengunduh dataset historis dari BMKG dan Kaggle. Dataset yang digunakan terdiri dari 92.887 entri data dengan 13 atribut, di antaranya adalah tanggal kejadian, waktu, koordinat geografis, kedalaman hiposentrum, dan magnitudo. Data tambahan seperti informasi mengenai mekanisme patahan (misalnya strike, dip, rake) juga tersedia dan digunakan untuk mendukung analisis lebih lanjut. Melalui pendekatan ini, penelitian bertujuan untuk menganalisis pengaruh optimasi *hyperparameter* terhadap akurasi prediksi magnitudo gempa bumi, yang didasarkan pada hasil klasterisasi yang dilakukan dengan DBSCAN. Hasil akhir penelitian akan disajikan dalam bentuk dashboard interaktif yang dibangun dengan menggunakan *Streamlit*, sehingga mempermudah interpretasi dan visualisasi hasil analisis secara real-time.



Gambar 1. Skema Penelitian

Skema tahapan penelitian yang ditunjukkan pada Gambar 1 menguraikan langkah-langkah yang dilakukan, dimulai dari pengumpulan data, *preprocessing*, klasterisasi menggunakan DBSCAN,

pemodelan prediksi dengan *Random Forest*, hingga evaluasi model. Setiap tahapan dalam skema ini akan dijelaskan secara lebih rinci pada bagian berikutnya.

Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa katalog gempa bumi Indonesia yang diperoleh dari platform Kaggle melalui tautan Kaggle Dataset. Dataset ini terdiri dari 92.887 entri dengan atribut utama berupa waktu kejadian, lokasi episenter (lintang dan bujur), kedalaman hiposentrum, serta magnitudo. Data ini mencakup gempa bumi yang tercatat di Indonesia selama periode waktu tertentu dan berfungsi sebagai dasar untuk analisis spasial dan prediksi magnitudo gempa.

Preprocessing Data

Tahap *preprocessing* dilakukan untuk memastikan bahwa data yang digunakan memiliki kualitas yang memadai sebelum dianalisis lebih lanjut. Proses ini terdiri dari beberapa langkah utama sebagai berikut:

1) Pembersihan Data (*Data Cleaning*)

Entri yang duplikat atau memiliki nilai yang tidak lengkap pada atribut penting, seperti magnitudo, kedalaman, dan koordinat geografis, dihapus untuk menghindari distorsi dalam analisis.

2) Transformasi Format Waktu

Atribut tanggal (tgl) dan waktu kejadian (ot) digabungkan menjadi satu kolom bertipe *datetime* dengan format standar (YYYY-MM-DD hh:mm:ss). Transformasi ini memungkinkan analisis temporal yang lebih mudah dan mencegah kesalahan interpretasi terkait waktu kejadian.

3) Normalisasi Data

Atribut numerik, khususnya magnitudo (*mag*) dan kedalaman (*depth*), dinormalisasi menggunakan metode Min-Max Scaling. Tujuan normalisasi ini adalah untuk memastikan semua variabel berada dalam skala yang seragam antara 0 dan 1, sehingga mencegah dominasi variabel tertentu dalam analisis.

4) Kategorisasi Magnitudo

Magnitudo dibagi ke dalam tiga kategori, yaitu ringan ($M < 4$), sedang ($4 \leq M < 6$), dan berat ($M \geq 6$). Kategorisasi ini bertujuan untuk mempermudah visualisasi distribusi gempa serta mendukung interpretasi pola yang lebih jelas pada hasil analisis.

Klasterisasi DBSCAN

Setelah tahap *preprocessing*, langkah berikutnya adalah melakukan klasterisasi dengan algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Algoritma DBSCAN dipilih karena kemampuannya dalam mengelompokkan data berdasarkan kepadatan titik serta mendeteksi data yang tidak termasuk dalam klaster manapun (noise). Berdasarkan penelitian oleh Hafid dan Hidayat (2025), DBSCAN terbukti efektif dalam mengidentifikasi zona seismik berdasarkan kepadatan wilayah tanpa memerlukan jumlah klaster yang ditentukan sebelumnya. DBSCAN beroperasi dengan dua parameter utama:

1) ϵ (epsilon)

Jarak maksimal antara dua titik untuk dianggap sebagai tetangga yang saling berhubungan.

2) minPts

Jumlah minimum titik yang diperlukan untuk membentuk sebuah klaster.

Pemilihan nilai-nilai parameter ini dilakukan secara eksperimen agar sesuai dengan distribusi spasial data gempa bumi Indonesia. Hasil klasterisasi menunjukkan terbentuknya beberapa zona seismik aktif dengan kepadatan tinggi, terutama di jalur subduksi Sumatra, Jawa, Sulawesi, dan Papua. Sementara itu, data yang tidak memenuhi kriteria kepadatan diberi label -1 yang merepresentasikan noise, atau kejadian gempa yang sporadis. Label klaster yang dihasilkan dari DBSCAN digunakan sebagai fitur tambahan dalam tahap pemodelan prediksi magnitudo. Fitur ini memungkinkan model machine learning untuk mempelajari hubungan antara zona seismik dan magnitudo gempa, yang dapat meningkatkan akurasi prediksi.

Pemodelan Prediksi dengan *Random Forest*

Tahap selanjutnya adalah pemodelan prediksi magnitudo gempa menggunakan algoritma *Random Forest*, yang merupakan metode *ensemble learning* berbasis pohon keputusan. *Random Forest* dipilih karena kemampuannya dalam menangani dataset besar, ketahanannya terhadap *noise*, serta efektivitasnya dalam memodelkan hubungan *non-linear* antar variabel. Sebagaimana ditunjukkan oleh Pratiwi & Arum (2025), *Random Forest* efektif dalam mengklasifikasikan data gempa bumi di Indonesia, dengan akurasi yang tinggi dalam membedakan

tingkat risiko gempa berdasarkan parameter spasial dan magnitudo. Dalam penelitian ini, label klaster yang dihasilkan dari DBSCAN digunakan sebagai fitur tambahan bersama atribut utama gempa bumi. Dua skenario pengujian dilakukan dalam penelitian ini:

1) Model *Baseline*

Menggunakan parameter default pada algoritma *Random Forest*.

2) Model dengan *Hyperparameter Tuning*

Menggunakan *Randomized Search* untuk mencari kombinasi *hyperparameter* terbaik, seperti jumlah pohon (*n_estimators*), kedalaman maksimum (*max_depth*), dan jumlah minimum sampel per split (*min_samples_split*).

Selain optimasi *hyperparameter*, performa model *Random Forest* juga dipengaruhi oleh pemilihan fitur yang relevan. Penelitian oleh Tantyoko *et al.* (2023) menunjukkan bahwa dengan seleksi atribut signifikan, seperti kedalaman hiposentrum, lokasi episentrum, dan zona seismik, akurasi prediksi gempa bumi dapat ditingkatkan secara substansial. Temuan ini mendukung hipotesis bahwa kombinasi klasterisasi spasial (DBSCAN) dan optimasi *hyperparameter* pada *Random Forest* akan menghasilkan model prediksi yang lebih andal dan mampu menghasilkan generalisasi yang lebih baik terhadap variasi data seismik.

Evaluasi Model

Evaluasi dilakukan untuk mengukur kinerja model *Random Forest* dalam memprediksi magnitudo gempa bumi. Pada penelitian ini digunakan tiga metrik utama, yaitu:

1) *Mean Squared Error* (MSE)

Menghitung rata-rata kuadrat selisih antara nilai aktual dan prediksi. (James *et al.*, 2021).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

2) *Root Mean Squared Error* (RMSE)

Akar kuadrat dari MSE untuk menunjukkan kesalahan dalam satuan yang sama dengan magnitudo. (Chai & Draxler, 2014)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

3) Koefisien Determinasi (R^2)

Mengukur proporsi variansi data yang dapat dijelaskan oleh model. (Draper & Smith, 1998).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

Visualisasi Hasil

Proses visualisasi hasil penelitian dilakukan melalui pembuatan dashboard interaktif yang dibangun menggunakan *Streamlit*. Dashboard ini dirancang untuk menampilkan berbagai tahapan analisis, termasuk eksplorasi data, *preprocessing*, hasil klusterisasi menggunakan DBSCAN, prediksi magnitudo gempa menggunakan model Random Forest, serta evaluasi performa model dengan menggunakan metrik *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, dan R^2 . Untuk mempermudah interpretasi hasil, dashboard ini mengintegrasikan berbagai elemen visual, seperti grafik interaktif yang dibuat dengan *Plotly* dan peta interaktif menggunakan *Folium*. Peta ini digunakan untuk menampilkan sebaran kluster gempa berdasarkan hasil klusterisasi DBSCAN, sehingga memungkinkan pengguna untuk melakukan analisis spasial pola seismik secara lebih intuitif. Pendekatan serupa telah digunakan dalam penelitian Oktaviarini *et al.* (2024), yang membangun model dashboard interaktif berbasis *Streamlit* untuk transformasi data statistik di BPS Kota Mojokerto. Meskipun fokus utama penelitian mereka adalah pada data statistik, bukan prediksi berbasis *machine learning*, pendekatan tersebut memberikan gambaran bagaimana dashboard interaktif dapat digunakan untuk memvisualisasikan data dengan cara yang mudah diakses dan dipahami.

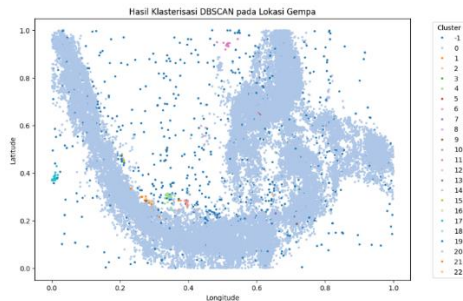
3. Hasil dan Pembahasan**Hasil****Dataset**

Dataset yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari platform Kaggle dengan judul *Earthquakes in Indonesia*. Dataset ini mencakup 92.887 entri gempa bumi yang tercatat di wilayah Indonesia, dengan cakupan spasial yang luas dan periode waktu yang panjang. Setiap entri mencakup informasi penting, antara lain waktu kejadian, koordinat episenter (lintang dan bujur), kedalaman hiposentrum, magnitudo, serta keterangan wilayah. Sebagian entri juga memuat data tambahan mengenai mekanisme sumber gempa, seperti *strike*, *dip*, dan *rake*, meskipun tidak semua entri menyertakan informasi ini. Pemilihan dataset ini didasarkan pada keberadaannya sebagai data terbuka (*open data*) yang dapat diakses oleh publik, sehingga sangat relevan untuk penelitian berbasis *machine learning*. Data historis gempa bumi yang terkandung dalam dataset ini memungkinkan pemodelan spasial dan prediksi magnitudo dengan pendekatan yang lebih akurat. Pendekatan serupa telah diterapkan dalam penelitian oleh Fansyuri *et al.* (2025), yang menggunakan dataset gempa dari Kaggle untuk klasifikasi wilayah berdasarkan potensi gempa menggunakan algoritma *Naive Bayes*. Atribut-atribut yang tersedia dalam dataset ini dirangkum pada Tabel 1 berikut.

Tabel 1. Atribut Dataset Gempa Bumi Indonesia

No	Atribut	Deskripsi
1	Tgl	Tanggal kejadian gempa bumi (format YYYY/MM/DD).
2	Ot	Waktu gempa bumi
3	Lat	Lintang lokasi pusat gempa (derajat desimal).
4	Lon	Bujur lokasi pusat gempa (derajat desimal).
5	Depth	Kedalaman pusat gempa (hiposentrum) dalam kilometer.
6	Mag	Besarnya magnitudo gempa bumi.
7	Remark	Keterangan wilayah atau lokasi administratif terdekat dengan pusat gempa.
8	Strike1	Sudut jurus (strike) bidang patahan.
9	Dip1	Sudut jurus (strike) bidang patahan pertama.
10	Rake1	Sudut kemiringan (dip) bidang patahan pertama.

11	Strike2	Sudut jurus (strike) bidang patahan kedua
12	Dip2	Sudut kemiringan (dip) bidang patahan kedua
13	Rake2	Sudut gerakan (rake) bidang patahan kedua



Gambar 2. Hasil Klasterisasi DBSCAN

Gambar ini menggambarkan hasil klasterisasi data gempa bumi di Indonesia menggunakan algoritma DBSCAN. Setiap warna yang berbeda pada peta mewakili klaster yang terbentuk berdasarkan kepadatan titik gempa, sedangkan titik berwarna biru tua yang diberi label -1 menunjukkan data yang dikategorikan sebagai noise, yakni gempa yang tidak termasuk dalam klaster manapun. Dapat dilihat bahwa sebagian besar titik gempa tergabung dalam satu klaster besar yang mengikuti jalur subduksi utama di Sumatra, Jawa, dan Nusa Tenggara. Selain itu, klaster-klaster yang lebih kecil juga terbentuk di wilayah Sulawesi, Maluku, dan Papua. Hasil ini menunjukkan efektivitas algoritma DBSCAN dalam mengidentifikasi zona seismik aktif di Indonesia berdasarkan distribusi spasial gempa bumi.

Tabel 2. Hasil Klasterisasi DBSCAN

No	Cluster	Jumlah Data	Keterangan
1	-1	1098	Noise
2	0	91552	Klaster utama
3	1	18	Klaster kecil

Tabel ini menyajikan distribusi data gempa bumi yang telah dikelompokkan menggunakan algoritma DBSCAN. Sebagian besar data tergolong dalam klaster 0, yang mewakili jalur subduksi utama di Indonesia, yakni sepanjang Sumatra, Jawa, dan Nusa Tenggara. Selain itu, klaster-klaster lainnya terbentuk dalam jumlah lebih kecil di wilayah Sulawesi, Papua, dan Maluku. Data yang tidak termasuk dalam klaster manapun diberi label -1, yang menandakan adanya *noise*, atau gempa yang tidak mengikuti pola kepadatan tertentu. Hasil ini menunjukkan

kemampuan DBSCAN dalam mengidentifikasi dan memisahkan zona seismik aktif secara spasial, sekaligus mengelompokkan kejadian gempa yang tidak terdefinisi dalam klaster yang jelas.

Pemodelan Prediksi Magnitudo dengan *Random Forest*

Pemodelan prediksi magnitudo gempa menggunakan algoritma *Random Forest* dipilih karena keandalannya dalam menangani dataset besar, ketahanannya terhadap *outliers*, serta kemampuannya dalam menangani hubungan non-linear antar variabel. *Random Forest* juga telah terbukti memiliki performa yang lebih baik dalam prediksi gempa dibandingkan dengan beberapa algoritma lain, seperti yang diungkapkan oleh Maulita dan Wahid (2024). Dalam penelitian ini, hasil klasterisasi DBSCAN digunakan sebagai fitur tambahan untuk memperkaya representasi data, yang membantu model dalam mengenali pola spasial dan temporal gempa yang lebih kompleks. Proses pemodelan dilakukan dalam dua skenario:

1) Skenario *Baseline*

Random Forest digunakan dalam konfigurasi default tanpa optimasi *hyperparameter*, yang berfungsi sebagai model dasar untuk evaluasi.

2) Skenario *Hyperparameter Tuning*

Randomized Search digunakan untuk mengoptimalkan beberapa parameter penting, termasuk jumlah pohon (*n_estimators*), kedalaman maksimum (*max_depth*), dan jumlah minimum sampel per split (*min_samples_split*). Perbandingan antara kedua skenario ini bertujuan untuk mengevaluasi sejauh mana optimasi parameter dapat meningkatkan akurasi prediksi magnitudo gempa.

Tabel 3. Konfigurasi *Hyperparameter Random Forest*

Parameter	Baseline	Setelah Tuning
<i>n_estimator</i>	100	200
<i>max_depth</i>	None	20
<i>min_samples_split</i>	2	5
<i>min_samples_leaf</i>	1	2
<i>max_features</i>	auto	Sqrt

Tabel ini menggambarkan perbedaan konfigurasi *hyperparameter* pada algoritma *Random Forest*. Model *baseline* menggunakan parameter default tanpa optimasi, sementara pada skenario kedua, dilakukan *hyperparameter tuning* menggunakan *Randomized Search* untuk mencari kombinasi parameter terbaik. Perubahan konfigurasi pada parameter utama, seperti jumlah pohon (*n_estimators*), kedalaman maksimum (*max_depth*), dan jumlah minimum sampel pada setiap *split* dan *leaf* (*min_samples_split* dan *min_samples_leaf*), terbukti meningkatkan kinerja model dalam memprediksi magnitudo gempa bumi. Optimasi parameter ini bertujuan untuk mendapatkan kombinasi yang lebih efisien dan tepat, yang pada gilirannya dapat menghasilkan prediksi yang lebih akurat.

Evaluasi Model

Evaluasi model dilakukan untuk mengukur sejauh mana algoritma *Random Forest* mampu memprediksi magnitudo gempa bumi secara akurat. Dalam penelitian ini, digunakan tiga metrik utama untuk mengukur kinerja model, sebagai berikut:

1) Mean Squared Error (MSE)

Menghitung rata-rata kuadrat selisih antara nilai aktual dan nilai prediksi. Metrik ini digunakan untuk menilai sejauh mana kesalahan prediksi yang dihasilkan oleh model.

2) Root Mean Squared Error (RMSE)

Merupakan akar kuadrat dari MSE, yang memberikan gambaran mengenai kesalahan model dalam satuan yang sama dengan data asli, sehingga memudahkan interpretasi.

3) Koefisien Determinasi (R^2)

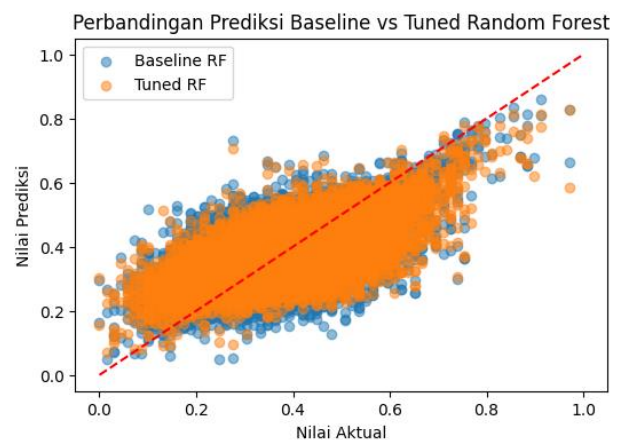
Menilai proporsi variabilitas data yang dapat dijelaskan oleh model, yang menggambarkan sejauh mana model dapat menangkap pola dalam data.

Penggunaan kombinasi metrik ini sejalan dengan praktik evaluasi yang umum digunakan dalam penelitian sebelumnya mengenai prediksi gempa menggunakan *machine learning*, sebagaimana dicontohkan oleh Nurdiansyah & Kusuma (2022). Metrik-metrik ini memberikan gambaran yang lebih komprehensif mengenai kemampuan model dalam memprediksi magnitudo gempa dan sejauh mana model dapat diandalkan.

Tabel 4. Hasil Evaluasi Model Random Forest

Model	MSE	RMSE	R^2
<i>Random Forest</i> (base)	0.0083	0.0908	0.4420
<i>Random Fores</i> (tuning)	0.0077	0.0879	0.4780

Berdasarkan hasil evaluasi, model *Random Forest baseline* menghasilkan nilai MSE sebesar 0.0083, RMSE 0.0908, dan R^2 0.4420. Setelah dilakukan *hyperparameter tuning*, performa model mengalami peningkatan, dengan nilai MSE turun menjadi 0.0077, RMSE menurun menjadi 0.0879, dan R^2 meningkat menjadi 0.4780. Hasil ini menunjukkan bahwa optimasi *hyperparameter* secara signifikan meningkatkan kemampuan model dalam memprediksi magnitudo gempa bumi. Penurunan nilai MSE dan RMSE, serta peningkatan R^2 , mengindikasikan bahwa model yang telah di-tuning mampu menjelaskan variasi data dengan lebih baik dibandingkan model baseline. Hal ini menegaskan bahwa *hyperparameter tuning* berkontribusi pada peningkatan akurasi dan efektivitas model dalam memodelkan hubungan antara variabel-variabel seismik dan magnitudo gempa.



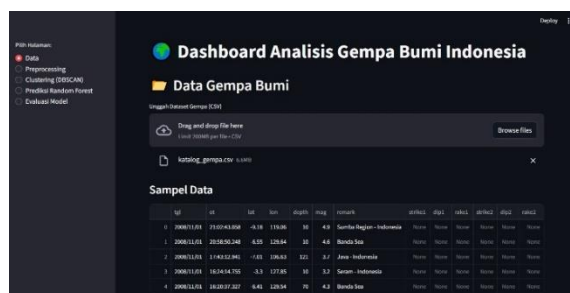
Gambar 3. Perbandingan nilai aktual dengan nilai prediksi

Gambar ini menunjukkan perbandingan antara hasil prediksi model *Random Forest baseline* dan model yang telah melalui proses *hyperparameter tuning*, dibandingkan dengan nilai aktual magnitudo gempa bumi. Titik berwarna biru menggambarkan prediksi yang dihasilkan oleh model baseline, sementara titik berwarna oranye menunjukkan hasil prediksi setelah dilakukan tuning. Secara visual, pola sebaran kedua model cenderung mengikuti garis diagonal merah

yang mewakili prediksi sempurna. Meskipun perbedaan antara keduanya relatif tipis, model hasil tuning terlihat lebih rapat terhadap garis diagonal dibandingkan model baseline. Hal ini sejalan dengan hasil evaluasi numerik, yang menunjukkan bahwa optimasi *hyperparameter* memberikan peningkatan akurasi meskipun perbaikan tersebut tidak terlalu signifikan. Tuning berhasil meningkatkan kemampuan model dalam menghasilkan prediksi yang lebih mendekati nilai aktual, meskipun peningkatan akurasi lebih jelas pada metrik evaluasi numerik seperti MSE, RMSE, dan R^2 .

Visualisasi Hasil

Untuk mempermudah interpretasi dan analisis hasil penelitian, visualisasi dilakukan melalui *dashboard* interaktif berbasis *Streamlit*. *Dashboard* ini menampilkan berbagai tahapan analisis, mulai dari eksplorasi data, *preprocessing*, klusterisasi menggunakan DBSCAN, hingga prediksi magnitudo gempa dengan *Random Forest*. Pengguna dapat menelusuri pola distribusi data gempa, melihat hasil klusterisasi pada peta interaktif, membandingkan performa model, serta memantau perbandingan prediksi dengan data aktual secara visual. Lebih rinci, visualisasi data gempa bumi Indonesia dalam proyek ini meliputi: distribusi magnitudo gempa, hasil klusterisasi DBSCAN, prediksi magnitudo gempa dengan *Random Forest*, serta evaluasi performa model berdasarkan metrik MSE, RMSE, dan R^2 . Dengan pendekatan ini, pengguna dapat mendapatkan wawasan yang lebih jelas dan mendalam mengenai pola seismik serta efektivitas model dalam memprediksi magnitudo gempa bumi.



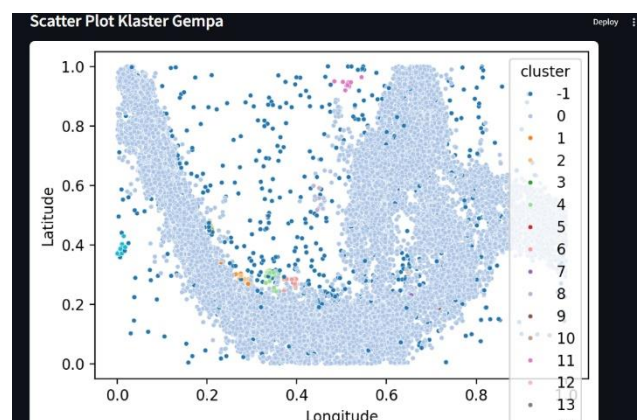
Gambar 4. Dashboard Data

Gambar menunjukkan antarmuka dashboard untuk analisis data gempa bumi di Indonesia, termasuk fitur unggah data dan pratinjau sampel data gempa dalam format tabel.



Gambar 5. Tahap Preprocessing

Gambar ini menampilkan perbandingan data gempa sebelum dan sesudah tahap preprocessing. Proses ini melibatkan pembersihan data dan pengisian nilai yang sebelumnya kosong (seperti parameter strike, dip, dan rake).



Gambar 6. Hasil Klaseterisasi DBSCAN

Gambar ini menunjukkan hasil clustering data gempa menggunakan algoritma DBSCAN, yang memetakan titik-titik gempa berdasarkan koordinat latitude dan longitude. Setiap warna merepresentasikan kluster berbeda, sedangkan kluster -1 menunjukkan data yang dianggap noise atau tidak termasuk dalam kluster manapun.



Gambar 7. Hasil Prediksi Random Forest

Gambar ini menampilkan hasil prediksi magnitudo gempa menggunakan model Random Forest. Data mencakup fitur seperti lokasi (latitude, longitude), kedalaman, dan klaster, serta membandingkan nilai magnitudo aktual dengan hasil prediksi dari model.



Gambar 8. Hasil Evaluasi Model

Evaluasi model Random Forest menunjukkan bahwa model yang telah di-tuning memberikan performa lebih baik dibanding baseline, ditunjukkan oleh penurunan nilai MSE dan RMSE serta peningkatan skor R^2 .

Pembahasan

Penelitian ini berhasil mengintegrasikan algoritma *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) untuk mengidentifikasi zona seismik aktif dan *Random Forest* untuk memprediksi magnitudo gempa di Indonesia. Hasil klusterisasi menggunakan DBSCAN menunjukkan pola distribusi spasial yang sesuai dengan jalur subduksi utama di Indonesia, yakni Sumatra, Jawa, Sulawesi, dan Papua. Ini sejalan dengan penelitian oleh Sulandari *et al.* (2024), yang juga menggunakan DBSCAN untuk mengelompokkan data gempa di Indonesia, dengan hasil yang serupa dalam identifikasi zona seismik aktif berdasarkan kepadatan gempa. Penggunaan DBSCAN dalam penelitian ini memungkinkan pemisahan antara zona seismik dengan intensitas tinggi dan kejadian gempa yang lebih sporadis, yang dapat berguna untuk memetakan risiko seismik di berbagai wilayah Indonesia. Selain itu, penggunaan *Random Forest* sebagai model prediksi magnitudo gempa sejalan dengan penelitian yang dilakukan oleh Pratiwi dan Arum (2025), yang juga menerapkan Random Forest untuk klasifikasi data gempa dengan akurasi yang tinggi. Random Forest dipilih dalam penelitian ini karena kemampuannya

untuk menangani dataset besar, serta ketahanannya terhadap noise dan variabel yang tidak relevan. Seperti yang ditemukan oleh Maulita dan Wahid (2024), penggunaan model ini dapat menghasilkan prediksi yang akurat, terutama jika dikombinasikan dengan seleksi fitur yang relevan, seperti kedalaman dan lokasi episentrum. Dengan mengoptimalkan model melalui tuning hyperparameter, penelitian ini berhasil meningkatkan performa model dengan penurunan MSE dan RMSE, serta peningkatan nilai R^2 yang menunjukkan bahwa model yang telah dioptimasi dapat menjelaskan lebih baik variabilitas data dibandingkan model baseline.

Hasil yang diperoleh juga mendukung temuan oleh Kurniawan *et al.* (2024), yang menyatakan bahwa metode machine learning, khususnya Random Forest, dapat digunakan secara efektif dalam prediksi magnitudo gempa dengan memperhatikan berbagai variabel yang berhubungan dengan kejadian gempa. Dengan menggabungkan model prediksi Random Forest dengan informasi klaster yang diperoleh dari DBSCAN, penelitian ini memperkenalkan pendekatan baru yang lebih akurat dalam memahami pola spasial dan temporal gempa bumi. Hasil ini membuka peluang untuk pengembangan lebih lanjut, terutama dalam menciptakan sistem pemantauan gempa yang berbasis pada teknologi machine learning, yang dapat memberikan prediksi lebih tepat dan berguna dalam upaya mitigasi bencana.

4. Kesimpulan dan Saran

Penelitian ini berhasil menggabungkan metode klusterisasi DBSCAN dan *Random Forest* untuk menganalisis dan memprediksi magnitudo gempa bumi di Indonesia. Melalui tahap preprocessing, data historis gempa diklusterkan menggunakan DBSCAN, yang menghasilkan zona seismik aktif di sepanjang jalur subduksi utama Indonesia. Zona-zona ini kemudian digunakan sebagai variabel tambahan dalam pemodelan prediksi. Hasil evaluasi menunjukkan bahwa model *Random Forest baseline* menghasilkan nilai MSE sebesar 0.0083, RMSE 0.0908, dan R^2 0.4420, sementara setelah dilakukan tuning hyperparameter, terdapat peningkatan performa yang signifikan. Model yang telah dioptimasi menghasilkan MSE sebesar 0.0077, RMSE 0.0879, dan R^2 0.4780. Peningkatan ini

menunjukkan bahwa tuning hyperparameter secara substansial meningkatkan akurasi prediksi serta kemampuan model dalam menjelaskan variabilitas magnitudo gempa bumi, yang juga tercermin dari visualisasi scatter plot, di mana prediksi model lebih mendekati nilai aktual. Secara keseluruhan, kombinasi DBSCAN dan Random Forest terbukti efektif dalam menganalisis data spasial-temporal gempa bumi, dengan potensi untuk mendukung sistem pemantauan seismik berbasis visualisasi interaktif melalui *Streamlit*. Untuk penelitian selanjutnya, disarankan agar model ini dikembangkan lebih lanjut dengan memasukkan sumber data yang lebih kaya, seperti data real-time dari BMKG, parameter mekanisme sumber gempa, serta fitur geofisika tambahan lainnya. Pengoptimalan parameter DBSCAN menggunakan metode sistematis seperti grid search dapat meningkatkan hasil klusterisasi, sementara penerapan algoritma lain seperti *XGBoost*, *LightGBM*, atau pendekatan deep learning berpotensi meningkatkan akurasi prediksi model. Pengembangan sistem berbasis web real-time juga sangat penting agar hasil prediksi dapat digunakan langsung dalam upaya mitigasi bencana. Penelitian ini tidak hanya memberikan kontribusi terhadap analisis data seismik secara komputasional, tetapi juga membuka peluang untuk pengembangan model prediksi yang lebih akurat dan aplikatif dalam mendukung upaya mitigasi risiko gempa bumi di Indonesia.

5. Ucapan Terima Kasih

Terima kasih kepada UNISNU Jepara, dosen pembimbing, dan rekan-rekan yang telah memberikan dukungan dan motivasi selama penelitian ini.

6. Daftar Pustaka

- Budiman, K., & Ifriza, Y. N. (2021). Analysis of earthquake forecasting using random forest. *Journal of soft computing exploration*, 2(2), 153-162.
<https://doi.org/10.52465/joscex.v2i2.51>.
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247–1250.
- Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). Wiley-Interscience.
- Fansyuri, M., Putra, R. T. S., Salsabila, A. A., Ramadhan, F., & Fiqih, M. (2025). Analisis data gempa di Indonesia menggunakan algoritma Naive Bayes. *JRIIN: Jurnal Riset Informatika dan Inovasi*, 2(10), 1914–1916.
- Hafid, H., & Hidayat, R. (2025). Implementation of DBSCAN for Earthquake Clustering in Indonesia with Potential Surface Damage. *Journal of Mathematics, Computations and Statistics*, 8(1), 124-133.
- Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.
- Hutagaol, R. M. A., Lana, V. T., Dzunnurain, Z. A., & Kurniawan, R. (2024, September). Penerapan Machine Learning dalam Prediksi Klasifikasi Big Data Kedalaman Gempa Bumi di Indonesia Tahun 2015-2024. In *PROSIDING SEMINAR NASIONAL SAINS DATA* (Vol. 4, No. 1, pp. 42-51).
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer.
- Maulita, I., & Wahid, A. (2024). Prediksi Magnitudo Gempa Menggunakan Random Forest, Support Vector Regression, XGBoost, LightGBM, dan Multi-Layer Perceptron Berdasarkan Data Kedalaman dan Geolokasi (Predicting Earthquake Magnitude Using Random Forest, Support Vector Regression, XGBoost, LightGBM, and Multi-Layer Perceptron Based on Depth and Geolocation Data). *Jurnal Pendidikan Dan Teknologi Indonesia*, 4, 221-232.

- Narastu, R. S., Pratama, R. F., & Suryani, D. (2023). Analisis eksplorasi data gempa bumi di Indonesia menggunakan data BMKG. *Jurnal Sistem dan Sains Riset (JSSR)*, 4(3), 45–55.
- Oktaviarini, K. N., Wahyuni, E. D., & Permata Sari, R. (2024). Transformasi data statistik menjadi visual interaktif menggunakan Streamlit: Studi kasus BPS Kota Mojokerto. *Jurnal Kridatama Sains dan Teknologi*, 6(2), 123–132.
- Pamungkas, T. D., Firdaus, R. A., Rohmah, N. A., Rizki, R., & Affriani, A. R. Analisis Spasial Rencana Tata Ruang Wilayah Berbasis Kerentanan Gempa Bumi (Studi Kasus: Kecamatan Cisarua Kabupaten Bandung Barat). *JGISE: Journal of Geospatial Information Science and Engineering*, 7(2), 161-170. <https://doi.org/10.22146/jgise.101944>.
- PRATAMA, M. A. W. R. (2021). *The Influence of Using Up Sampling Method in Predicting Patients' Disease Types using a Combination of Natural Language Processing, Naive Bayes Algorithm, XGBoost and Support Vector Machine* (Doctoral dissertation, Universitas Mercu Buana Jakarta).
- Pratiwi, A. P., & Arum, P. R. (2025). Implementasi Algoritma Random Forest untuk Mengklasifikasikan Data Gempa Bumi di Indonesia. *EIGEN MATHEMATICS JOURNAL*, 8(1), 25-33. <https://doi.org/10.29303/emj.v8i1.185>.
- Rahayu, T., & Purwoko, A. (2020). Pendekatan Empiris untuk Estimasi Kerugian Ekonomi dan Dampak Kerusakan Lingkungan akibat Gempabumi di Deli Serdang. *Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer*, 2, 58-62.
- Rahman, R. R. A., & Wijayanto, A. W. (2021). Pengelompokan data gempa bumi menggunakan algoritma DBSCAN. *Jurnal Meteorologi dan Geofisika*, 22(1), 31–38.
- Rahman, S., Sembiring, A., Siregar, D., Prahmana, I. G., Puspadini, R., & Zen, M. (2023). Python: Dasar Dan Pemrograman Berorientasi Objek. *Penerbit Tahta Media*.
- Sulandari, W., Wati, R. K., & Pratiwi, H. (2024). Penerapan algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN) dan Self-Organizing Map (SOM) untuk clustering data gempa bumi di Indonesia. *Statistika*, 14(1), 45–55.
- Syaifuddin, A., & Prabowo, T. (2024). Optimasi Akurasi Model Prediksi Magnitudo Gempa Bumi dengan Integrasi Clustering DBSCAN pada Ensemble Learning (Random Forest & XGBoost).
- Tantyoko, H., Sari, D. K., & Wijaya, A. R. (2023). Prediksi potensial gempa bumi Indonesia menggunakan metode Random Forest dan feature selection. *Jurnal IDEALIS*, 6(3), 1587–1597.