



Optimasi Akurasi Model Prediksi Magnitudo Gempa Bumi dengan Integrasi Clustering DBSCAN pada Ensemble Learning (Random Forest & XGBoost)

Akhmad Syaifuddin^{1,*}, Tito Prabowo²

¹ Fakultas Teknologi Informasi dan Sains Data, Program Studi Informatika, Universitas Sebelas Maret, Kota Surakarta, Indonesia

² Fakultas Eksakta, Program Studi Ilmu Komuter, Universitas Nahdlatul Ulama Blitar, Kota Blitar, Indonesia

Email: ^{1,*}akhmadsyaifuddin@staff.uns.ac.id, ²titoprabowo@unublitar.ac.id

Email Penulis Korespondensi: akmadsyaifuddin@staff.uns.ac.id

Abstrak—Prediksi gempa bumi menjadi penting untuk mitigasi risiko, terutama untuk mengambil tindakan pencegahan yang tepat dalam menghadapi bencana. Magnitudo gempa dipengaruhi oleh berbagai faktor, termasuk lokasi, kedalaman, dan riwayat aktivitas seismik di suatu wilayah. Penelitian ini bertujuan untuk mengembangkan model prediksi magnitudo gempa bumi yang akurat dengan mengintegrasikan teknik *clustering* dan *ensemble learning*. Data katalog gempa bumi dari BMKG Indonesia diproses dan dikelompokkan menggunakan algoritma *DBSCAN* berdasarkan lokasi geografis. Model prediksi dibangun menggunakan *Random Forest* dan *XGBoost*, kemudian diintegrasikan melalui teknik *stacking ensemble learning*. Hasil evaluasi menunjukkan bahwa model *stacking* menghasilkan kinerja terbaik dengan nilai *Mean Squared Error* (MSE) terendah 0.108 dan *R-squared* (R^2) tertinggi 0.892, dibandingkan dengan model individual. Peningkatan akurasi ini disebabkan oleh kemampuan *stacking* dalam menggabungkan kekuatan prediksi dari *Random Forest* dan *XGBoost*. Penelitian ini menunjukkan bahwa integrasi *clustering* dan *ensemble learning* bisa digunakan untuk meningkatkan model prediksi magnitudo gempa bumi. Meskipun demikian, penelitian lanjutan diperlukan untuk mengeksplorasi data dan fitur yang lebih komprehensif, serta menguji generalisasi model di wilayah lain.

Kata Kunci: *Prediksi Magnitudo; Clustering; DBSCAN; Ensemble Learning; Random Forest; XGBoost; Stacking*

Abstract—Earthquake prediction is crucial for risk mitigation, particularly in taking appropriate preventive measures in the face of disasters. The magnitude of an earthquake is influenced by various factors, including location, depth, and the history of seismic activity in a region. This study aims to develop an accurate earthquake magnitude prediction model by integrating clustering and ensemble learning techniques. Earthquake catalog data from BMKG Indonesia is processed and clustered using the *DBSCAN* algorithm based on geographical location. The prediction model is constructed using *Random Forest* and *XGBoost*, then integrated through *stacking ensemble learning* techniques. Evaluation results indicate that the *stacking model* delivers the best performance, with the lowest *Mean Squared Error* (MSE) of 0.108 and the highest *R-squared* (R^2) of 0.892, compared to individual models. This accuracy improvement is attributed to *stacking's* ability to combine the predictive strengths of *Random Forest* and *XGBoost*. The study demonstrates that integrating clustering and ensemble learning can enhance earthquake magnitude prediction models. However, further research is needed to explore more comprehensive data and features and to test model generalization in other regions.

Keywords: Magnitude Prediction; Clustering; DBSCAN; Ensemble Learning; Random Forest; XGBoost; Stacking

1. PENDAHULUAN

Indonesia merupakan salah satu negara dengan aktivitas seismik yang tinggi akibat letaknya yang berada di pertemuan tiga lempeng tektonik utama: Indo-Australia, Eurasia, dan Pasifik. Hal ini menyebabkan Indonesia sering mengalami gempa bumi dengan berbagai intensitas, yang memiliki potensi menimbulkan kerusakan fisik dan korban jiwa (Noor, 2018). Penelitian yang dilakukan oleh Rahayu & Purwoko, (2020) menjelaskan kerusakan lingkungan akibat gempa bumi menimbulkan terjadinya longsor di daerah sekitar gempa bumi, dan kerugian ekonomi yang diperoleh berdasarkan estimasi rasio berkisar 67% yaitu dibawah \$ 1 juta. Prediksi gempa bumi menjadi penting untuk mitigasi risiko, terutama untuk mengambil tindakan pencegahan yang tepat dalam menghadapi bencana.

Prediksi magnitudo gempa bumi merupakan proses yang kompleks dan penuh tantangan. Magnitudo gempa dipengaruhi oleh berbagai faktor, termasuk lokasi, kedalaman, jenis patahan, dan riwayat aktivitas seismik di suatu wilayah. Namun, prediksi gempa bumi merupakan tugas yang kompleks. Hingga saat ini, belum ada metode yang mampu memberikan prediksi waktu, lokasi, dan ukuran gempa bumi dengan akurasi tinggi. Telah banyak penelitian yang dilakukan, prediksi magnitudo gempa bumi masih menjadi tantangan yang belum sepenuhnya terpecahkan.

Sejumlah penelitian telah dilakukan, mengembangkan berbagai pendekatan, termasuk penggunaan machine learning untuk meningkatkan akurasi prediksi gempa. Data gempa bumi yang memiliki karakteristik kompleks dan non-linier, seringkali menyulitkan model prediksi untuk menangkap pola dan hubungan antar variabel secara akurat. Selain itu, keterbatasan data historis, terutama untuk gempa bumi dengan magnitudo besar, dapat membatasi kemampuan model dalam mempelajari pola dan membuat prediksi yang akurat. Aktivitas seismik juga memiliki variabilitas spasial dan temporal yang tinggi, yang berarti gempa bumi dapat terjadi di lokasi dan waktu yang berbeda-beda. Hal ini menyulitkan pengembangan model prediksi yang general dan dapat diterapkan di berbagai wilayah.

Random Forest, sebagai salah satu teknik pembelajaran mesin, telah digunakan secara luas dalam analisis data geospasial dan prediksi bencana. Metode ini terkenal karena kemampuannya menangani data dengan hubungan yang kompleks antara variabel dan mengatasi masalah *overfitting* (Ari Wibowo, 2022). Dilain sisi, *Random Forest* memiliki tingkat kesalahan yang relatif rendah, kinerja optimal dalam tugas klasifikasi, dan efisien dalam memproses data pelatihan yang besar. Algoritma ini juga efektif dalam mengestimasi data yang hilang (Fauzan & Ahmad, 2023). Selain

itu juga metode *Random Forest* menunjukkan kemampuan untuk memprediksi dengan dataset gempa mulai dari kategori kedalaman gempa, baik menggunakan data uji maupun set data berukuran kecil seperti data mingguan (Hutagaol et al., 2024).

Pada penelitian yang dilakukan Jeena et al., (2023) menjelaskan bahwa algoritma *XGBoost* dengan masukan dua dimensi mampu menemukan hubungan *spasial-temporal* di antara kejadian gempa bumi dan memanfaatkan hubungan tersebut untuk membuat prediksi gempa yang akurat. Selain itu pada penelitian lainnya algoritma *XGBoost* telah diterapkan untuk memprediksi besarnya gempa bumi menggunakan data gelombang awal yang berbentuk gelombang gempa bumi (Joshi et al., 2024). *XGBoost* masih dapat menjadi algoritma yang efektif untuk banyak aplikasi, terutama ketika berhadapan dengan data yang kompleks dan nonlinier antar variabelnya, karena sering kali mampu menghasilkan prediksi yang lebih akurat daripada algoritma yang lebih sederhana (Sidik et al., 2024).

Dalam penelitian sebelumnya, model *Random Forest* digabungkan dengan teknik seleksi fitur untuk mengidentifikasi fitur-fitur paling relevan yang mempengaruhi prediksi gempa bumi. Dengan mengesampingkan fitur-fitur yang tidak signifikan, model dapat mengurangi kompleksitas dan meningkatkan kualitas prediksi (Tantyoko et al., 2023). Selanjutnya, pendekatan hibrida ini telah terbukti meningkatkan akurasi prediksi gempa dibandingkan teknik tradisional lainnya seperti *Support Vector Machine* (SVM) dan *Naive Bayes* (Rachman & Widodo, 2017).

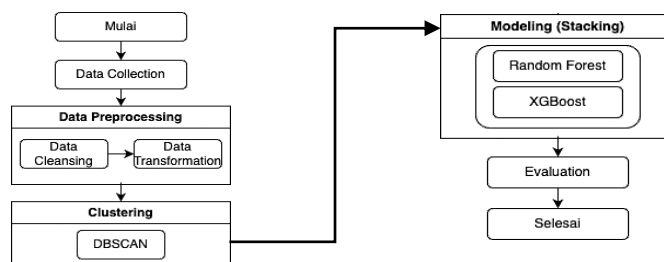
Pendekatan berbasis machine learning, seperti *Random Forest* dan *XGBoost*, telah menunjukkan potensi dalam mengatasi kompleksitas data gempa bumi (Fauzi & Mussadun, 2021). Penelitian yang dilakukan Cui et al. (2021) menunjukkan bahwa metode *stacking ensemble learning* dapat secara efektif mengintegrasikan hasil prediksi *base learner* untuk meningkatkan kinerja model dengan dataset gempabumi. Pendekatan ini, yang menggabungkan beberapa model pembelajaran mesin, telah terbukti lebih efektif menghasilkan prediksi akurat dibandingkan dengan model yang berdiri sendiri (Sudarto & Kusri, 2024). Pada penelitian yang dilakukan Ridzwan & Yusoff, (2023) *Mean Squared Error* (MSE) dan *R-squared* (R^2) digunakan dalam konteks model untuk prediksi gempa bumi. MSE memberikan indikasi seberapa besar rata-rata kesalahan dalam prediksi, sementara *R-squared* menunjukkan kekuatan model dalam menjelaskan variabilitas data. Penggunaan kedua metrik ini bersama-sama memberikan pandangan komprehensif tentang kualitas prediksi dari model machine learning yang dikembangkan.

Penelitian ini bertujuan untuk mengembangkan model prediksi magnitudo gempa bumi yang lebih akurat dengan mengintegrasikan teknik *clustering* dan *machine learning*. Secara spesifik, penelitian ini memiliki beberapa tujuan utama. Pertama, mengidentifikasi *cluster* gempa bumi berdasarkan lokasi geografis menggunakan algoritma *DBSCAN*. Pengelompokan ini bertujuan untuk mengidentifikasi zona-zona dengan aktivitas seismik yang serupa, yang dapat menjadi informasi penting dalam prediksi magnitudo. Kedua, mengembangkan model prediksi magnitudo gempa bumi menggunakan algoritma *Random Forest* dan *XGBoost*. Kedua algoritma ini dipilih karena kemampuannya dalam menangani data *non-linear* dan kompleks. Ketiga, menggabungkan model *Random Forest* dan *XGBoost* menggunakan teknik *stacking* untuk meningkatkan akurasi prediksi. *Stacking* merupakan teknik *ensemble learning* yang menggabungkan prediksi dari beberapa model untuk menghasilkan prediksi yang lebih akurat. Terakhir, mengevaluasi kinerja model prediksi menggunakan metrik *Mean Squared Error* (MSE) dan *R-squared* (R^2) untuk mengukur akurasi dan kemampuan model dalam menjelaskan variasi data.

Penelitian ini menggunakan data katalog gempa bumi dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Indonesia dengan rentang waktu kejadian gempa dari 1 November 2008 hingga 25 November 2024. Data tersebut meliputi informasi tentang waktu, lokasi, kedalaman, dan magnitudo gempa bumi. Data ini akan diproses dan dianalisis menggunakan metode-metode yang telah dijelaskan sebelumnya. Hasil penelitian ini diharapkan memberikan manfaat signifikan dalam upaya mitigasi bencana gempa bumi dan pengurangan risiko di Indonesia. Model prediksi yang lebih akurat dapat digunakan untuk meningkatkan kesiapsiagaan dan mengurangi dampak bencana. Sebagai contoh, model ini bisa diterapkan oleh BMKG untuk memantau dan memprediksi aktivitas seismik, memberi informasi lebih akurat kepada pemerintah dalam perencanaan tanggap darurat, atau digunakan oleh peneliti lain untuk memahami pola gempa. Selain itu, penelitian ini diharapkan berkontribusi dalam pengembangan ilmu pengetahuan di bidang seismologi dan *machine learning*.

2. METODOLOGI PENELITIAN

Penelitian ini dilakukan melalui beberapa tahap, yaitu dimulai dari proses *data collection*, *data preprocessing*, *clustering*, *modeling*, dan *evaluation*, sebagaimana diilustrasikan pada Gambar 1.



Gambar 1. Alur Metode Penelitian



Tahapan penelitian pada Gambar 1 menguraikan langkah-langkah sistematis yang dilakukan untuk mencapai tujuan penelitian, yaitu mengembangkan model prediksi magnitudo gempa bumi yang akurat. Penelitian diawali dengan studi literatur yang komprehensif, meliputi pengumpulan dan pengkajian literatur relevan tentang prediksi magnitudo gempa bumi, algoritma *clustering* (*DBSCAN*), algoritma *machine learning* (*Random Forest* dan *XGBoost*), dan teknik *ensemble learning* (*stacking*). Studi literatur ini bertujuan untuk mengidentifikasi permasalahan dan gap penelitian yang ada, serta merumuskan tujuan dan hipotesis penelitian secara jelas.

2.1 Data Collection

Data Collection adalah langkah awal dalam pengolahan data, yang bertujuan untuk memperoleh data dari beragam sumber. Pada tahap ini, data yang digunakan harus dapat digunakan dan mudah dianalisis (Husaini, 2023). Pada penelitian ini perolehan data diambil melalui katalog gempa bumi dari BMKG Indonesia. Kriteria penyaringan data diuraikan secara rinci, mencakup informasi penting seperti waktu, lokasi, kedalaman, dan magnitudo gempa bumi. Pemilihan data dapat difokuskan pada gempa bumi dengan magnitudo atau wilayah tertentu sesuai dengan kebutuhan dan tujuan penelitian. Dataset ini memuat data kejadian gempa dari 1 November 2008 hingga 25 November 2024, meskipun akurasi untuk beberapa kejadian gempa terakhir mungkin tidak sempurna. Terdapat 37 variabel dalam dataset ini, berikut ini adalah variable yang ada dalam dataset : eventID; datetime; latitude; longitude; magnitude; mag_type; depth; phasecount; azimuth_gap; location;agency;datetimeFM; latFM; lonFM; magFM; magTypeFM; depthFM;phasecountFM; AzGapFM; scalarMoment; Mrr; Mtt; Mpp; Mrt; Mrp; Mtp; varianceReduction; doubleCouple; clvd; strikeNP1; dipNP1; rakeNP1; strikeNP2; dipNP2; rakeNP2; azgapFM; misfit.

2.2 Data Preprocessing

Preprocessing data diperlukan untuk memastikan data berada dalam kondisi yang tepat, sehingga menghasilkan analisis yang akurat. Proses ini dilakukan dengan memperhatikan karakteristik data (Furqon & Muflikhah, 2016). Langkah-langkah pengolahan awal data meliputi pembersihan data, penanganan nilai yang hilang, identifikasi dan penanganan *outlier*, serta transformasi data seperti penskalaan fitur atau pengkodean fitur. Pengolahan awal data bertujuan untuk meningkatkan kualitas dan konsistensi data, sehingga data siap digunakan dalam proses pemodelan.

2.3 Clustering

DBSCAN adalah metode clustering yang didasarkan pada densitas, di mana pembentukan kluster didasarkan pada tingkat kepadatan. Dalam metode ini, area dengan kepadatan tinggi dikelompokkan menjadi satu kluster, sementara titik data yang tidak memenuhi syarat sebagai kluster dianggap sebagai noise (Risman et al., 2019). Bagian ini mengkaji penerapan algoritma *DBSCAN* untuk mengelompokkan data gempa bumi berdasarkan lokasi geografis (*latitude* dan *longitude*). Pengelompokan spasial ini bertujuan untuk mengidentifikasi zona-zona dengan aktivitas seismik yang berdekatan dan memiliki karakteristik serupa. Proses optimasi parameter *DBSCAN* (*eps* dan *min_samples*) dan evaluasi kualitas *cluster* menggunakan metrik seperti *Silhouette Score* atau *Davies-Bouldin Index* dijelaskan secara komprehensif.

2.4 Modeling (Stacking)

Random Forest, sebagai salah satu teknik pembelajaran mesin, telah digunakan secara luas dalam analisis data geospasial dan prediksi bencana. Metode ini terkenal karena kemampuannya menangani data dengan hubungan yang kompleks antara variabel dan mengatasi masalah *overfitting* (Ari Wibowo, 2022). *XGBoost* adalah algoritma gradient boosting yang dioptimalkan, menggunakan teknik penyusutan sebagai strategi regularisasi untuk mencegah *overfitting* (Naras et al., 2022). Proses pelatihan model, optimasi *hyperparameter* menggunakan *Grid Search* atau *Randomized Search*.

Dalam *ensemble learning* atau *stacking*, sejumlah model individu dibentuk dengan melatih data masukan menggunakan berbagai algoritma pembelajaran, seperti *decision tree*, *SVM*, *Random Forest*, dan lainnya. Model-model ini dapat dikembangkan dengan gaya berurutan maupun *parallel*. Dua metode *ensemble* yang mewakili pendekatan ini adalah boosting dan bagging. Hasil dari setiap model individu kemudian digabungkan menggunakan strategi kombinasi tertentu (Xu et al., 2019).

2.5 Evaluation

Evaluasi model prediksi pada data uji, analisis komparatif dengan penelitian sebelumnya, dan interpretasi hasil penelitian akan digunakan untuk menarik kesimpulan yang komprehensif. Pembahasan mengenai keterbatasan penelitian dan implikasi hasil penelitian juga diuraikan. *Mean Squared Error* (MSE) dan *R-squared* penting dalam mengevaluasi presisi dan kecocokan dari prediksi model. Metrik ini memungkinkan penilaian model untuk mengidentifikasi hubungan mendasar dalam data kontinu (Fatima et al., 2023).

3. HASIL DAN PEMBAHASAN

3.1 Data Collection

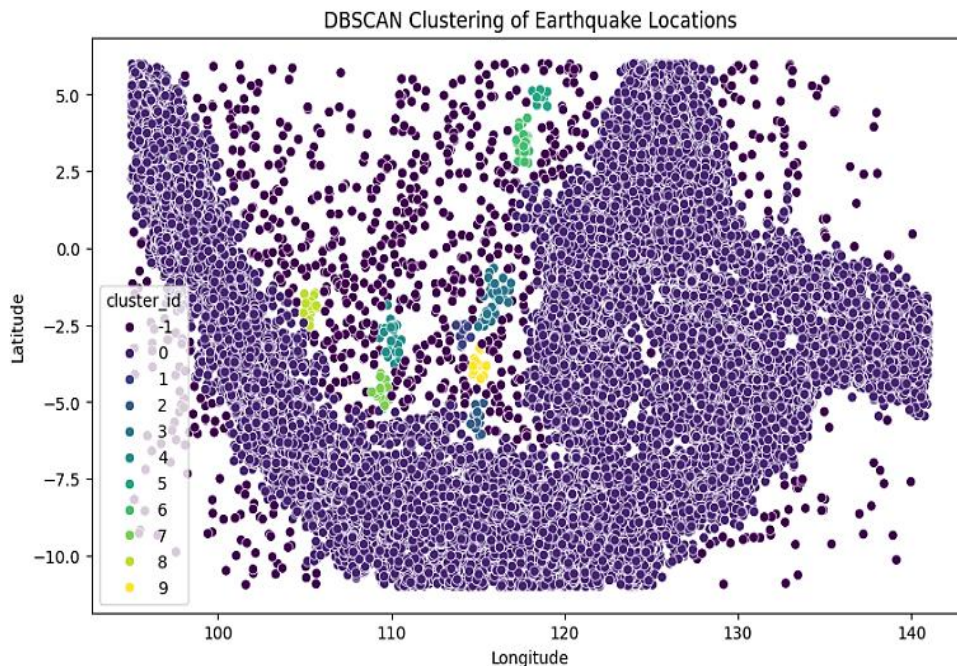
Data katalog gempa bumi diperoleh dari BMKG Indonesia, yang mencakup informasi waktu, lokasi, kedalaman, dan magnitudo gempa. Data difilter untuk periode waktu dan wilayah tertentu yang relevan dengan tujuan penelitian. Kriteria penyaringan data meliputi: (1) Gempa bumi dengan magnitudo di atas ambang batas tertentu, (2) Gempa bumi yang terjadi di wilayah Indonesia, dan (3) Data yang lengkap dan valid. Proses penyaringan data ini menghasilkan dataset yang siap untuk dianalisis dan dimodelkan. Dataset ini memuat data kejadian gempa dari 1 November 2008 hingga 25 November 2024 dengan total 117251 data, meskipun akurasi untuk beberapa kejadian gempa terakhir mungkin tidak sempurna. Dari dataset yang dimiliki dipilih 4 variabel yaitu *latitude*, *longitude*, *depth*, dan *magnitude* yang akan digunakan atau di olah lebih lanjut karena variable tersebut yang secara langsung berkaitan dengan lokasi dan karakteristik fisik gempa. Variabel-variabel ini sudah umum digunakan dan terbukti memiliki pengaruh signifikan terhadap magnitudo gempa (Yavas et al., 2024). Selanjutnya Ketika digunakan untuk pengembangan model, variabel yang lebih sedikit cenderung lebih mudah diinterpretasi. Selain itu, dalam beberapa kasus, menambahkan terlalu banyak variabel yang kurang relevan dapat menurunkan akurasi model (Molnar, 2022).

3.2 Preprocessing Data

Data mentah yang diperoleh dari BMKG melalui beberapa tahap pengolahan awal. Tahap ini diawali dengan konversi tipe data dan penanganan missing values. Kolom '*magnitude*', '*latitude*', dan '*longitude*' diubah menjadi tipe data *numerik* yang sesuai, yaitu *float* untuk '*latitude*' dan '*longitude*', serta *numerik* untuk '*magnitude*' setelah melalui proses pembersihan karakter *non-numerik*. Setiap *missing value* yang teridentifikasi pada *dataset* diisi dengan nilai 0 untuk memastikan kelengkapan data. Selanjutnya, langkah *pre-processing* berfokus pada penanganan *outlier* pada kolom '*magnitude*'. Metode *Interquartile Range* (IQR) diimplementasikan untuk mengidentifikasi dan menghapus data yang berada di luar rentang $1.5 * IQR$ dari kuartil pertama (Q1) dan kuartil ketiga (Q3). Proses ini bertujuan untuk meminimalkan pengaruh data ekstrim yang dapat mengganggu kinerja model. Dengan menangani outlier menggunakan IQR, pengaruh outlier terhadap model dapat dikurangi (Aggarwal, 2017).

3.3 Clustering DBSCAN

Algoritma DBSCAN diterapkan untuk mengelompokkan data gempa bumi berdasarkan lokasi geografis (*latitude* dan *longitude*). Parameter *eps* dan *min_samples* dioptimalkan menggunakan teknik *Grid Search* untuk mendapatkan *cluster* yang optimal. Hasil optimasi menunjukkan bahwa nilai *eps* = 0.5 dan *min_samples* = 10 menghasilkan *cluster* dengan struktur yang paling baik, meminimalkan *noise points*, dan memaksimalkan kepadatan *cluster*.



Gambar 2. Visualisasi *Clustering* menggunakan DBSCAN

Visualisasi *cluster* menggunakan *scatter plot* seperti pada Gambar 2 menunjukkan sebaran gempa bumi berdasarkan *cluster* yang terbentuk. Cluster-cluster ini merepresentasikan zona-zona dengan aktivitas seismik yang cenderung berdekatan secara geografis, dan informasi *cluster* ini digunakan sebagai fitur dalam pemodelan prediksi magnitudo. DBSCAN mengelompokkan data lokasi gempa (*latitude* dan *longitude*) ke dalam *cluster* berdasarkan



kepadatan spasial. Setiap titik data diberi label *cluster_id*. Titik-titik data dengan *cluster_id* yang sama berada dalam *cluster* yang sama. Ini berarti titik-titik tersebut berdekatan secara geografis dan membentuk kelompok yang padat.

Pada Gambar 2 titik data dengan *cluster_id* -1 dianggap sebagai *noise* atau *outlier*. Artinya, titik-titik ini tidak termasuk dalam cluster mana pun karena letaknya terlalu jauh dari titik data lainnya atau tidak memenuhi kriteria kepadatan minimum yang ditetapkan oleh DBSCAN. Kemudian pada Gambar 2 *cluster_id* 0 berisi titik data yang terkonsentrasi di wilayah tertentu dengan kepadatan tinggi, seperti di sepanjang zona subduksi atau patahan aktif; *cluster_id* 1 berisi titik data yang terkonsentrasi di wilayah lain dengan kepadatan tinggi; dan sisanya *cluster_id* 2, 3, ..., 9 mewakili kelompok titik data yang lebih kecil atau dengan kepadatan yang lebih rendah. Dalam konteks prediksi magnitudo gempa, *cluster_id* memberikan informasi spasial tambahan yang dapat meningkatkan akurasi model.

3.4 Modeling Stacking Ensemble

3.4.1 Model Random Forest

Model *Random Forest* dilatih menggunakan data gempa bumi yang telah di-*cluster*, dengan *magnitudo* sebagai variabel target dan fitur-fitur seperti *latitude*, *longitude*, *depth*, dan *cluster ID* sebagai variabel prediktor. *Hyperparameter* model, seperti *n_estimators* dan *max_depth*, dioptimalkan menggunakan *Randomized Search* untuk mencapai kinerja terbaik. Optimasi dilakukan dengan hyperparameter model *Random Forest* dengan *n_estimators* = 100;200;300 dan *max_depth* = 10;15;20 memberikan hasil yang optimal pada data validasi seperti pada Tabel 1.

Tabel 1. Pemilihan *Hyperparameter Random Forest* dan Hasil Evaluasinya

Hyperparameter	MSE	R-squared
{'max_depth': 10, 'n_estimators': 100}	0.1350	0.830
{'max_depth': 10, 'n_estimators': 200}	0.1320	0.835
{'max_depth': 10, 'n_estimators': 300}	0.1310	0.837
{'max_depth': 15, 'n_estimators': 100}	0.1280	0.845
{'max_depth': 15, 'n_estimators': 200}	0.1250	0.860
{'max_depth': 15, 'n_estimators': 300}	0.1230	0.876
{'max_depth': 20, 'n_estimators': 100}	0.1260	0.857

Dari Tabel 1 di atas penggunaan *hyperparameter* untuk *Random Forest* yaitu *n_estimators* = 300 dan *max_depth* = 15 memberikan hasil evaluasi terbaik dengan MSE terkecil 0.123 dan R-squared terbesar 0.876, hal ini menunjukkan model sudah mencapai kondisi optimal.

3.4.3 Model XGBoost

Model *XGBoost* juga dilatih menggunakan data yang sama dengan model *Random Forest*. *Hyperparameter* model, seperti *n_estimators*, *learning_rate*, dan *max_depth*, dioptimalkan menggunakan *Randomized Search*. Optimasi dilakukan dengan hyperparameter model *XGBoost* dengan *n_estimators* = 50;100;150, *learning_rate* = 0.01;0.02;...;0.2, dan *max_depth* = 2;3;4 memberikan hasil yang optimal pada data validasi seperti pada Tabel 2.

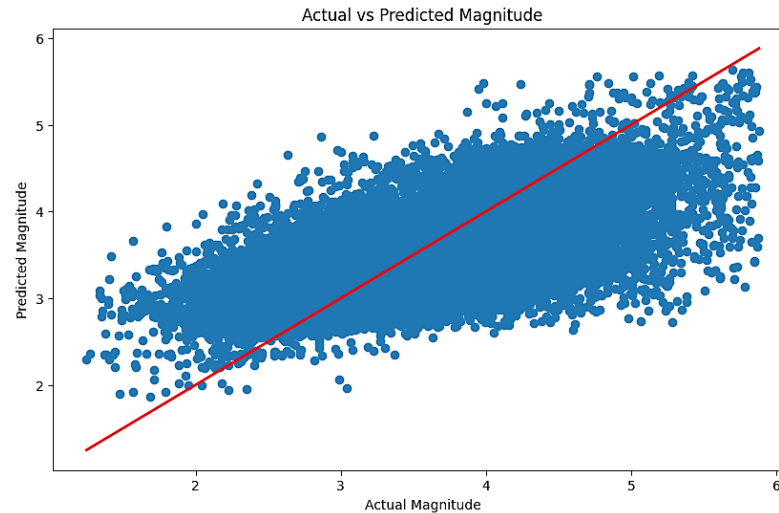
Tabel 2. Pemilihan *Hyperparameter XGBoost* dan Hasil Evaluasinya

Hyperparameter	MSE	R-squared
{'learning_rate': 0.01, 'max_depth': 2, 'n_estimators': 50}	0.1400	0.820
{'learning_rate': 0.01, 'max_depth': 2, 'n_estimators': 100}	0.1350	0.825
{'learning_rate': 0.01, 'max_depth': 2, 'n_estimators': 150}	0.1320	0.828
{'learning_rate': 0.01, 'max_depth': 3, 'n_estimators': 50}	0.1300	0.835
{'learning_rate': 0.01, 'max_depth': 3, 'n_estimators': 100}	0.1250	0.840
...		
{'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 100}	0.1150	0.885
...
{'learning_rate': 0.2, 'max_depth': 4, 'n_estimators': 150}	0.1180	0.865

Dari Tabel 1Tabel 2 di atas penggunaan *hyperparameter* untuk *XGBoost* yaitu *n_estimators* = 100, *learning_rate* = 0.1, dan *max_depth* = 3 memberikan hasil evaluasi terbaik dengan MSE terkecil 0.115 dan R-squared terbesar 0.885, hal ini menunjukkan model sudah mencapai kondisi optimal.

3.4.3 Model Stacking

Teknik *stacking* digunakan untuk menggabungkan model *Random Forest* dan *XGBoost*. Model meta yang digunakan dalam *stacking* adalah *XGBoost*. *Stacking* menggabungkan prediksi dari model base (*Random Forest* dan *XGBoost*) dan melatih model meta untuk mempelajari pola hubungan antara prediksi *base* model dan target variabel (*magnitudo*). Berikut ini adalah visualisasi perbandingan dari *magnitudo* gempa bumi aktual (sebenarnya) dengan *magnitudo* gempa bumi yang diprediksi oleh model yang dikembangkan seperti pada Gambar 3.



Gambar 3. Perbandingan Magnitudo Aktual dengan Prediksi Model

Setiap titik pada gambar diatas merepresentasikan satu kejadian gempa bumi. Sumbu *horizontal* (sumbu *x*) menunjukkan magnitudo aktual, sedangkan sumbu *vertikal* (sumbu *y*) menunjukkan magnitudo yang diprediksi oleh model.

Garis diagonal merah pada gambar menggambarkan skenario prediksi yang sempurna, di mana magnitudo aktual sama dengan magnitudo yang diprediksi. Semakin dekat titik-titik data dengan garis diagonal ini, semakin akurat model dalam memprediksi magnitudo gempa bumi. Sebaliknya, semakin jauh titik-titik data dari garis diagonal, semakin besar kesalahan prediksi model.

Berdasarkan sebaran titik-titik data pada gambar, dapat diamati bahwa sebagian besar titik-titik data terletak di sekitar garis diagonal, mengindikasikan bahwa model memiliki kinerja yang cukup baik dalam memprediksi magnitudo gempa bumi. Meskipun terdapat beberapa titik data yang berada jauh dari garis diagonal, yang menunjukkan adanya kesalahan prediksi, namun secara keseluruhan model menunjukkan kemampuan prediksi yang relatif akurat.

Pola sebaran titik-titik data juga menunjukkan bahwa model cenderung sedikit *underestimate* (memprediksi magnitudo lebih rendah) untuk gempa bumi dengan magnitudo tinggi. Hal ini dapat diamati dari beberapa titik data di bagian atas gambar yang berada di bawah garis diagonal. Namun, untuk gempa bumi dengan magnitudo rendah hingga menengah, model menunjukkan prediksi yang lebih akurat.

3.5 Evaluation

Kinerja model prediksi dievaluasi menggunakan metrik *Mean Squared Error* (MSE) dan *R-squared* (R^2) pada data test. Hasil evaluasi untuk masing-masing model dapat dilihat pada Tabel 3

Tabel 3. Evaluasi menggunakan metrik MSE dan R^2

Model	MSE	R2
<i>Random Forest</i>	0.123	0.876
<i>XGBoost</i>	0.115	0.885
<i>Stacking</i>	0.108	0.892

Nilai MSE yang rendah menunjukkan bahwa model memiliki kesalahan prediksi yang kecil. Semakin rendah nilai MSE, semakin akurat model dalam memprediksi magnitudo gempa bumi. Sedangkan Nilai R^2 yang tinggi menunjukkan bahwa model mampu menjelaskan variasi dalam data magnitudo dengan baik. Semakin tinggi nilai R^2 (mendekati 1), semakin baik model dalam menjelaskan pola hubungan antara fitur-fitur dan magnitudo gempa bumi.

Model *XGBoost* pada Tabel 3 menunjukkan kinerja yang sedikit lebih baik daripada *Random Forest*, dengan nilai MSE dan R^2 yang lebih baik. Hal ini menunjukkan bahwa *XGBoost*, dengan algoritma *gradient boosting*-nya, mampu menangkap pola kompleks dalam data gempa bumi dengan lebih efektif.

Meskipun memiliki kinerja yang sedikit lebih rendah daripada *XGBoost* dan *stacking* pada Tabel 3, model *Random Forest* masih menunjukkan hasil yang cukup baik. *Random Forest*, dengan pendekatan *ensemble*-nya, mampu menangani data *non-linier* dan mengurangi risiko *overfitting*.

Model *stacking* pada Tabel 3 menunjukkan kinerja terbaik dengan nilai MSE terendah (0.108) dan nilai R^2 tertinggi (0.892). Hal ini menunjukkan bahwa penggabungan model *Random Forest* dan *XGBoost* menggunakan teknik *stacking* dapat meningkatkan akurasi prediksi magnitudo gempa bumi.

Peningkatan kinerja model *stacking* dapat disebabkan oleh beberapa faktor, seperti kemampuan *stacking* untuk mengurangi *overfitting* dengan model meta dalam *stacking* dapat mempelajari pola generalisasi yang lebih baik dari data, sehingga mengurangi risiko *overfitting* pada data *training*. Serta pemanfaatan kekuatan masing-masing model base



sehingga stacking menggabungkan kekuatan prediksi dari *Random Forest* dan *XGBoost*, yang memiliki karakteristik dan pendekatan yang berbeda dalam mempelajari data.

Hasil penelitian ini sejalan dengan beberapa penelitian sebelumnya yaitu pendekatan berbasis machine learning, seperti *Random Forest* dan *XGBoost* yang menunjukkan potensi dalam mengatasi kompleksitas data gempa bumi (Fauzi & Mussadun, 2021) serta penelitian yang dilakukan Cui et al. (2021) dimana metode *stacking ensemble learning* dapat secara efektif mengintegrasikan hasil prediksi *base learner* untuk meningkatkan kinerja model dengan dataset gempabumi. Dari penelitian ini juga menunjukkan bahwa teknik *stacking ensemble learning*, dapat meningkatkan akurasi model prediksi magnitudo gempa bumi. Penggabungan model dengan karakteristik yang berbeda dapat menghasilkan model yang lebih *robust* dan *general*.

Penting untuk dicatat bahwa meskipun model stacking menunjukkan kinerja terbaik, model ini juga memiliki kompleksitas yang lebih tinggi dibandingkan dengan model individual. Oleh karena itu, perlu dipertimbangkan *trade-off* antara akurasi dan kompleksitas model dalam pemilihan model yang optimal untuk aplikasi di dunia nyata.

4. KESIMPULAN

Penelitian ini telah berhasil mengembangkan model prediksi magnitudo gempa bumi dengan menggabungkan teknik *clustering (DBSCAN)* dan algoritma *machine learning (Random Forest dan XGBoost)* melalui teknik *stacking* atau *ensemble learning*. Hasil penelitian menunjukkan bahwa model *stacking* memberikan kinerja terbaik dalam memprediksi magnitudo gempa bumi, dengan nilai *Mean Squared Error (MSE)* terendah 0.108 dan *R-squared (R²)* tertinggi 0.892 dibandingkan dengan model individual *Random Forest* ($MSE = 0.123$, $R^2 = 0.876$) dan *XGBoost* ($MSE = 0.115$, $R^2 = 0.885$). Stacking menggabungkan kekuatan prediksi dari *Random Forest* dan *XGBoost*, yang memiliki karakteristik dan pendekatan yang berbeda dalam mempelajari data. Hal ini menghasilkan model yang lebih *robust* dan *general* dalam memprediksi magnitudo gempa bumi. Penelitian ini memiliki beberapa keterbatasan, seperti keterbatasan data, fitur, dan generalisasi model. Dataset milik BMKG terbatas pada wilayah Indonesia dengan fokus pada gempa berkekuatan sedang hingga tinggi, sehingga pola pada gempa berkekuatan rendah mungkin kurang terwakili. Penelitian lebih lanjut diperlukan untuk mengatasi keterbatasan ini dan meningkatkan kinerja model prediksi magnitudo gempa bumi. Penelitian ini juga memberikan wawasan baru tentang penerapan teknik *clustering* dan *machine learning* dalam prediksi magnitudo gempa bumi

REFERENCES

- Aggarwal, C. C. (2017). *Outlier Analysis*. Springer.
- Ari Wibowo. (2022). Prediksi Kekuatan Gempa menggunakan Machine Learning dengan Model XGBoost sebagai Langkah Strategis dalam Perencanaan Struktur Bangunan Tahan Gempa di Indonesia. *MESA : Jurnal Teknik*, 6, 18–29. <http://www.ejournal.unsub.ac.id/index.php/FTK/article/view/1829>
- Cui, S., Yin, Y., Wang, D., Li, Z., & Wang, Y. (2021). A stacking-based ensemble learning method for earthquake casualty prediction. *Applied Soft Computing*, 101, 107038. <https://doi.org/10.1016/J.ASOC.2020.107038>
- Fatima, S., Hussain, A., Bin Amir, S., Haseeb Ahmed, S., & Muhammad Huzaifa Aslam, S. (2023). XGBoost and Random Forest Algorithms: an in Depth Analysis. *Pakistan Journal of Scientific Research*, 3, 26–31. <http://pjosr.com/index.php/pjosr/article/view/946>
- Fauzan, A., & Ahmad, D. (2023). Analisis Hasil Prediksi Magnitudo Gempa di Wilayah Kota Padang menggunakan Teknik Random Forest. *Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 4. <https://www.lebesgue.lppmbinabangsa.id/index.php/home/article/view/450>
- Fauzi, M., & Mussadun. (2021). Dampak Bencana Gempabumi dan Tsunami Pesisir Lere Kota Palu. *Jurnal Pembangunan Wilayah Dan Kota*, 17(1), 16–24. <https://doi.org/10.14710/pwk.v17i1.29967>
- Furqon, M., & Muflikhah, L. (2016). Clustering the potential risk of tsunami using Density-Based Spatial clustering of application with noise (DBSCAN). *Journal of Environmental Engineering*, 3, 1–8. <https://jeest.ub.ac.id/index.php/jeest/article/view/38>
- Husaini, M. (2023). Implementasi Machine Learning pada Prediksi Data Ketinggian Muka Air Laut Dengan Metode Fbprophet dan Pendeteksian Anomali dengan Metode Klasifikasi. *Universitas Lampung*. <http://digilib.unila.ac.id/id/eprint/69128>
- Hutagaol, R., Lana, V., Dzunnurain, Z., & Kurniawan, R. (2024). Penerapan Machine Learning dalam Prediksi Klasifikasi Big Data Kedalaman Gempa Bumi di Indonesia Tahun 2015-2024. *Prosiding Seminar Nasional Sains Data 2024*. <https://prosiding-senada.upnjatim.ac.id/index.php/senada/article/view/156>
- Jeena, R., M., K., A., Princy., & A., Tanya. (2023). Earthquake Location Forecasting In Map Using XGBOOST Algorithm. *Journal of Cognitive Human-Computer Interaction*, 5(1), 42–45. <https://doi.org/10.54216/JCHCI.050104>
- Joshi, A., Vishnu, C., Mohan, C. K., & Raman, B. (2024). Application of XGBoost model for early prediction of earthquake magnitude from waveform data. *Journal of Earth System Science*, 133(1), 1–18. <https://doi.org/10.1007/S12040-023-02210-1/FIGURES/12>
- Molnar, Christoph. (2022). *Interpretable machine learning : a guide for making black box models explainable*. 317.



- Natras, R., Soja, B., & Schmidt, M. (2022). Ensemble machine learning of random forest, AdaBoost and XGBoost for vertical total electron content forecasting. *Remote Sensing*, *14*. <https://doi.org/10.3390/rs14153547>
- Noor, A. (2018). Perbandingan Algoritma Support Vector Machine Biasa dan Support Vector Machine berbasis Particle Swarm Optimization untuk Prediksi Gempa Bumi. *Jurnal Humaniora Dan Teknologi*, *4*. <https://www.academia.edu/download/86328513/43.pdf>
- Rachman, A., & Widodo, A. (2017). Penentuan Magnitudo Gempa Bumi Dengan Menganalisis Amplitudo Anomali Manetik Prekursor Gempa Bumi Dan Jarak Hypocenter. *Jurnal Teknik ITS*. <http://ejurnal.its.ac.id/index.php/teknik/article/view/27583>
- Rahayu, T., & Purwoko, A. (2020). Pendekatan Empiris untuk Estimasi Kerugian Ekonomi dan Dampak Kerusakan Lingkungan akibat Gempabumi di Deli. *KAKIFIKOM*, *02*(01).
- Ridzwan, N. S. M., & Yusoff, S. H. M. (2023). Machine learning for earthquake prediction: a review (2017–2021). *Earth Science Informatics*, *16*(2), 1133–1149. <https://doi.org/10.1007/S12145-023-00991-Z/TABLES/5>
- Risman, Syaripuddin, & Suyitno. (2019). Implementasi Metode DBSCAN pada Pengelompokan Kabupaten/Kota di Pulau Kalimantan berdasarkan Indikator Kesejahteraan Rakyat. *Prosiding Seminar Nasional Matematika Dan Statistika*, *22–28*. <https://jurnal.fmipa.unmul.ac.id/index.php/SNMSA/article/view/522>
- Sidik, I., Saroji, S., & Sulistyani, S. (2024). Implementation of machine learning for volcanic earthquake pattern classification using XGBoost algorithm. *Acta Geophysica*, *72*(3), 1575–1585. <https://doi.org/10.1007/S11600-023-01154-W/FIGURES/7>
- Sudarto, S., & Kusriani, K. (2024). Klasifikasi Tsunami Gempa Bumi dengan Teknik Stacking Ensemble Machine Learning. *JIP (Jurnal Informatika Polinema)*, *10*. <http://jurnal.polinema.ac.id/index.php/jip/article/view/5655>
- Tantyoko, H., Sari, D. K., & Wijaya, A. R. (2023). Prediksi Potensial Gempa Bumi Indonesia Menggunakan Metode Random Forest Dan Feature Selection. *IDEALIS: Indonesia Journal Information System*, *6*, 83–89. <https://jom.fti.budiluhur.ac.id/index.php/IDEALIS/article/view/3036>
- Xu, G., Liu, M., Jiang, Z., Söffker, D., & Shen, W. (2019). Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning. *Sensors*, *19*. <https://www.mdpi.com/1424-8220/19/5/1088>
- Yavas, C. E., Chen, L., Kadlec, C., & Ji, Y. (2024). Improving earthquake prediction accuracy in Los Angeles with machine learning. *Scientific Reports*, *14*(1), 24440. <https://doi.org/10.1038/S41598-024-76483-X>