

## Penerapan Model Ensemble Learning dengan Random Forest dan Multi-Layer Perceptron untuk Prediksi Gempa

Turino<sup>\*1</sup>, Rujianto Eko Saputro<sup>2</sup>, Giat Karyono<sup>3</sup>

<sup>1,2</sup>Magister Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Indonesia

<sup>3</sup>Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Indonesia

Email: [123MA41D002@students.amikompurwokerto.ac.id](mailto:123MA41D002@students.amikompurwokerto.ac.id), [rujianto@amikompurwokerto.ac.id](mailto:rujianto@amikompurwokerto.ac.id),  
[giatkaryono@amikompurwokerto.ac.id](mailto:giatkaryono@amikompurwokerto.ac.id)

### Abstrak

Penelitian ini mengusulkan model *hybrid* yang menggabungkan metode Random Forest (RF) dan Multi-Layer Perceptron (MLPRegressor) untuk memprediksi magnitudo gempa bumi. Model ini bertujuan untuk meningkatkan akurasi prediksi dengan memanfaatkan kekuatan kedua algoritma tersebut, yang masing-masing memiliki keunggulan dalam menangani hubungan non-linier dan mengenali pola kompleks dalam data seismik. Evaluasi model menggunakan tiga metrik utama, yaitu *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *R-squared (R<sup>2</sup>)*. Hasil evaluasi menunjukkan bahwa model hibrida ini mampu memprediksi magnitudo gempa dengan akurasi yang cukup baik, dengan MAE sebesar 0,0738, RMSE 0,1078, dan R<sup>2</sup> 0,4204. Penerapan praktis dari model ini sangat relevan untuk sistem peringatan dini gempa bumi yang dapat membantu masyarakat untuk mengambil langkah-langkah pencegahan, seperti evakuasi dan penguatan infrastruktur di wilayah yang berisiko tinggi. Penelitian ini juga membuka peluang untuk mengembangkan model lebih lanjut dengan memperkenalkan data seismik *real-time*, algoritma pembelajaran mesin yang lebih canggih, dan penggunaan data geofisik serta pengamatan satelit untuk meningkatkan akurasi prediksi gempa bumi di masa depan. Dengan terus melakukan inovasi, ada potensi untuk mengembangkan sistem prediksi gempa bumi yang lebih akurat dan dapat diandalkan, yang pada akhirnya dapat meningkatkan kesiapsiagaan dan ketahanan terhadap bencana alam.

**Kata kunci:** data seismik, model ensemble, multi-layer perceptron, prediksi gempa bumi, random forest

## *Application of Ensemble Learning Model with Random Forest and Multi-Layer Perceptron for Earthquake Prediction*

### Abstract

*This study proposes a hybrid model combining Random Forest (RF) and Multi-Layer Perceptron (MLPRegressor) methods to predict earthquake magnitudes. The model aims to improve prediction accuracy by leveraging the strengths of both algorithms, each of which excels in handling non-linear relationships and recognizing complex patterns in seismic data. The model evaluation uses three key metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R<sup>2</sup>). The evaluation results show that the hybrid model can predict earthquake magnitudes with fairly good accuracy, with MAE of 0.0738, RMSE of 0.1078, and R<sup>2</sup> of 0.4204. The practical application of this model is highly relevant for earthquake early warning systems, which can help communities take preventive actions, such as evacuating and reinforcing infrastructure in high-risk areas. This study also opens opportunities for further developing the model by introducing real-time seismic data, more advanced machine learning algorithms, and the use of geophysical data and satellite observations to improve earthquake prediction accuracy in the future. With ongoing innovation, there is potential to develop more accurate and reliable earthquake prediction systems, which can ultimately enhance disaster preparedness and resilience.*

**Keywords:** earthquake prediction, ensemble learning, multi-layer perceptron, random forest, seismic data

## 1. PENDAHULUAN

Dalam mitigasi bencana, prediksi gempa bumi sangat penting untuk keselamatan manusia dan ketahanan infrastruktur. Memprediksi kejadian seismik merupakan hal yang kompleks dan penuh tantangan, memerlukan pendekatan inovatif yang melampaui metode-metode tradisional. Gempa bumi bersifat tidak terduga, sehingga dibutuhkan pendekatan hibrida yang menggabungkan inovasi dan teknologi. Pembelajaran mesin menawarkan alat-alat seperti jaringan syaraf dan random forest yang dapat mengubah data mentah menjadi wawasan prediktif.

Penelitian [1] menekankan pentingnya presisi dalam peramalan magnitudo, yang dapat mengurangi kerusakan yang ditimbulkan. Namun, metode empiris tidak memadai karena kompleksitas pergerakan tektonik. Penelitian [2] mengusulkan model hibrida yang menggabungkan algoritma jaringan syaraf dan random forest untuk meningkatkan akurasi prediksi dengan menganalisis gelombang seismik. Secara bersamaan, para peneliti juga mengeksplorasi cara untuk memprediksi potensi korban. Penelitian [3] mengembangkan indeks prediksi yang mempertimbangkan integritas bangunan, kepadatan penduduk, dan faktor lainnya. Kemampuan untuk meramalkan dampak terhadap manusia ini sangat penting dalam perencanaan darurat. Mencapai kesiapsiagaan yang holistik memerlukan kemampuan untuk merencanakan, beradaptasi, dan merespons hasil yang mungkin timbul dari gempa bumi. Prediksi gempa bumi melampaui analisis data semata. Penelitian [4] memperkenalkan variasi konten elektron total di ionosfer sebagai prekursor seismik, menunjukkan bahwa fenomena langit dapat memprediksi gempa bumi. [5] menggunakan teknologi satelit untuk mengungkap bayangan seismik, memberikan perspektif yang belum pernah ada sebelumnya bagi para peneliti.

Meskipun kemajuan telah dicapai, tantangan tetap ada. Penelitian [6] menyoroti sifat stokastik inheren dari gempa bumi, yang menjadi tantangan bagi model konvensional yang kesulitan menangkap perilaku dinamisnya. Seiring dengan dorongan untuk terus mendorong batasan ini, berbagai metodologi muncul, masing-masing menyumbangkan pola yang rumit untuk prediksi yang lebih andal. Peramalan magnitudo gempa yang akurat tetap menjadi tantangan karena pola temporal seismik yang sulit dipahami. Pembelajaran mesin dan model hibrida telah memberi harapan, tetapi prediksi yang tepat masih belum terwujud. Pembelajaran mesin, terutama arsitektur deep learning seperti RNN dan CNN, dapat menangkap ketergantungan temporal dan spasial dalam dataset seismik. Penelitian [1] menyarankan bahwa model-model ini dapat meningkatkan kinerja prediktif, meskipun akurasi konsisten dapat bervariasi di berbagai lingkungan seismik. Penelitian [7] mendukung penggunaan model hibrida, yang menggabungkan komponen yang saling berkorelasi dan independen untuk meningkatkan ketepatan prediksi. Fusi data elektromagnetik dalam prediksi seismik telah muncul sebagai frontier yang menjanjikan. Penelitian [8] mengembangkan indeks prediktif jangka pendek yang meningkatkan keandalan ramalan di wilayah yang kompleks secara seismik. Beragam sumber data memberikan wawasan yang kaya, terutama di daerah yang sangat aktif secara seismik dengan pengaruh dinamis.

Model probabilistik hibrida semakin memperkuat penelitian ini. Penelitian [9] menunjukkan efektivitas kerangka kerja terintegrasi dalam menganalisis karakteristik seismik regional. Model Epidemic Type Aftershock Sequence (ETAS) menjadi contoh bagaimana model probabilistik dapat memprediksi ritme dan intensitas setelah gempa. [10] menunjukkan bagaimana hubungan seismologi empiris, berdasarkan hukum Omori, dapat memberi informasi bagi model ini dan memberikan pemahaman yang lebih luas tentang dinamika temporal seismisitas. Namun, prediksi magnitudo gempa yang dapat diandalkan tetap menjadi tantangan. Penyempurnaan model yang berkelanjutan diperlukan karena ketidakpastian gempa bumi. Penelitian [11] mengembangkan peramalan adaptif yang dapat diskalakan dengan menggunakan grid multi-resolusi untuk lanskap seismik yang tidak stabil. Penelitian [12] memperkenalkan pendekatan Bayesian yang menggabungkan wawasan temporal dari getaran kecil untuk memprediksi pergeseran seismik besar.

Dalam beberapa studi terbaru, berbagai algoritma pembelajaran mesin telah dieksplorasi untuk analisis dan prediksi seismik. Penelitian [13] melakukan analisis spatiotemporal terhadap nilai B di Gunung Slamet, yang menyoroti pentingnya analisis tersebut dalam memahami aktivitas seismik dari waktu ke waktu. Selain itu, dalam bidang prediksi gempa bumi, penelitian [14] menggunakan algoritma Random Forest dan Gradient Boosting untuk memprediksi kerusakan bangunan pasca-gempa menggunakan dataset Gempa Nepal 2015, yang menunjukkan penerapan teknik data mining untuk analisis risiko bencana. Dalam studi lainnya, penelitian telah mengembangkan prediksi magnitudo gempa dengan menggunakan berbagai model, termasuk Random Forest, Support Vector Regression, XGBoost, LightGBM, dan Multi-Layer Perceptron, berdasarkan data kedalaman dan geolokasi [15]. Penelitian [16] juga membandingkan linear regression, Random Forest, dan Gradient Boosting untuk prediksi banjir, yang lebih lanjut menunjukkan fleksibilitas model-model ini dalam menghadapi berbagai tantangan lingkungan. Studi-studi ini secara keseluruhan menunjukkan potensi algoritma pembelajaran mesin dalam meningkatkan akurasi prediksi untuk fenomena bencana alam.

Meskipun pembelajaran mesin, model hibrida, dan ramalan probabilistik telah menunjukkan kemajuan, kompleksitas fenomena seismik tetap menjadi tantangan besar. Penelitian dan inovasi yang berkelanjutan sangat penting untuk meningkatkan akurasi prediktif dan mengurangi risiko seismik dengan lebih presisi dan wawasan. Penelitian ini memperkenalkan model hibrida yang menggabungkan Random Forest (RF) dan Multi-Layer Perceptron (MLPRegressor) untuk memprediksi magnitudo gempa di masa depan dengan presisi. Dengan menggabungkan kekuatan dari masing-masing metode, model hibrida ini meningkatkan akurasi dan ketahanan dalam prediksi seismik. Arsitektur model hibrida ini bergantung pada dinamika komplementer dari algoritma RF dan MLP. RF, sebagai kekuatan dalam pembelajaran ensambel, membangun banyak pohon keputusan selama pelatihan dan mengurangi overfitting dengan merata-rata hasil prediksi dari berbagai jalur keputusan, yang mendorong generalisasi. MLPRegressor, sebuah jaringan neural berlapis, menangkap nuansa data non-linear yang halus dan sangat mahir dalam memecahkan pola-pola rumit dalam dataset seismik. Dengan menggabungkan kedua

metode ini, model hibrida ini memadukan kekuatan interpretatif dan stabilitas RF dengan kemampuan pengenalan pola rumit dari MLP.

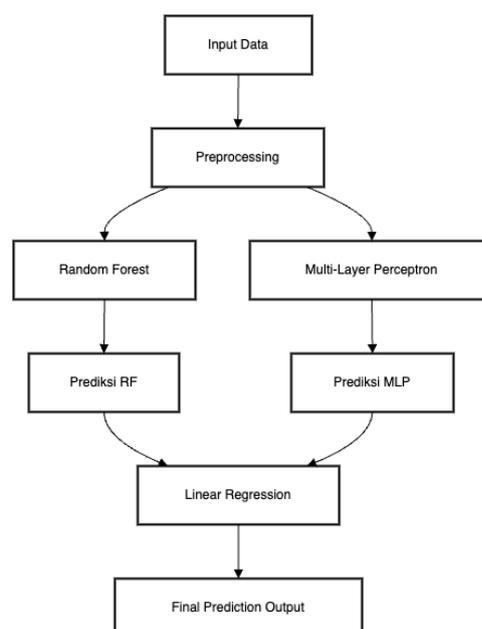
Meningkatkan akurasi peramalan gempa sangat penting untuk kesiapsiagaan dan respons terhadap bencana. Prediksi yang tepat sangat diperlukan dalam strategi mitigasi, mengurangi dampak buruk dari peristiwa seismik dengan menyelamatkan nyawa dan mengurangi kerugian ekonomi. Prediksi yang akurat sangat penting untuk kesiapsiagaan bencana yang tangguh. Penelitian [17] berpendapat bahwa sistem peringatan dini gempa bumi (EEWS) hanya efektif jika dapat memprediksi gempa dengan akurat. Peringatan yang tepat waktu memungkinkan masyarakat untuk mengambil tindakan pencegahan. Gempa Hualien 2018 menunjukkan bahwa efektivitas EEWS bergantung pada waktu prediksi yang dapat dicapai. Presisi adalah kunci untuk mengurangi kerusakan dan menyelamatkan nyawa. Akurasi prediksi yang lebih baik, sebagai landasan strategi responsif, memfasilitasi alokasi sumber daya. Model peramalan gempa hibrida, yang melampaui model tradisional dalam kemahiran prediktif, menawarkan prediksi yang lebih andal, memberdayakan layanan darurat untuk merespons fenomena seismik dengan efektif. Peramalan yang akurat dapat mengurangi jumlah korban jiwa. Algoritma yang lebih halus, seperti yang dijelaskan oleh [3], membentuk rencana respons darurat yang lebih terfokus, dengan memperhitungkan magnitudo gempa dan implikasinya untuk evakuasi yang efektif serta mobilisasi sumber daya.

Model hibrida, yang mengintegrasikan teknik seperti Random Forest dan jaringan neural, menandai perubahan paradigma menuju akurasi. Penelitian [2] menyoroti bahwa integrasi pembelajaran mesin dapat melampaui batasan tradisional, menghasilkan prediksi yang lebih dapat diandalkan dan kesiapsiagaan yang lebih komprehensif. Presisi tidak hanya penting dalam menghadapi keadaan darurat, tetapi juga untuk perencanaan jangka panjang dan penilaian risiko. Penelitian [18] menggali fase-fase seismik dan potensi wawasan yang mereka tawarkan, berkontribusi pada peta jalan manajemen risiko yang komprehensif untuk tantangan seismik baik jangka pendek maupun jangka panjang.

Tujuan penelitian ini adalah untuk mengembangkan model hibrida yang menggabungkan Random Forest (RF) dan Multi-Layer Perceptron (MLPRegressor) dalam memprediksi magnitudo gempa bumi dengan menggunakan pendekatan ensemble learning, untuk meningkatkan akurasi prediksi dan mengatasi keterbatasan metode-metode konvensional. Penelitian ini memiliki signifikansi yang besar dalam meningkatkan sistem peringatan dini gempa bumi, yang pada gilirannya dapat memperbaiki respons terhadap bencana dan mengurangi kerugian ekonomi serta korban jiwa. Dengan memanfaatkan teknik pembelajaran mesin yang lebih canggih, penelitian ini berpotensi untuk menyediakan alat yang lebih akurat dalam memprediksi kejadian gempa yang bisa sangat membantu dalam perencanaan mitigasi bencana dan pengelolaan risiko.

**2. METODE PENELITIAN**

Penelitian ini menggunakan model Ensemble Learning yang menggabungkan Random Forest (RF), Multi-Layer Perceptron (MLP), dan Linear Regression untuk meningkatkan akurasi prediksi magnitudo gempa. Gambar 1 di bawah ini adalah diagram yang menggambarkan arsitektur model Ensemble Learning yang digunakan dalam penelitian.



Gambar 1. Diagram Alir Metode Penelitian

## 2.1. Pengumpulan Data

Penelitian ini menggunakan dataset "Significant Earthquakes, 1900-Present" dari USGS yang tersedia di Kaggle [19]. Dataset ini berisi informasi mengenai gempa bumi yang terjadi di seluruh dunia dengan magnitudo 5 atau lebih, mencakup lebih dari satu abad dan memberikan wawasan yang berharga tentang pergerakan tektonik serta dampaknya. Dataset ini diperbarui secara berkala setiap minggu dengan entri baru, memastikan data yang digunakan selalu relevan dan terkini.

Pemilihan dataset ini didasarkan pada beberapa kriteria penting. Pertama, dataset ini mencakup kejadian gempa bumi global, yang menjamin representasi data yang luas dan mencakup berbagai kondisi seismik. Kedua, hanya gempa dengan magnitudo 5 atau lebih yang disertakan, sehingga fokus pada peristiwa seismik yang cukup besar untuk mempengaruhi manusia dan infrastruktur. Ketiga, dataset ini memiliki informasi temporal yang sangat terperinci, dengan pencatatan waktu yang sangat akurat (dalam milidetik sejak 1 Januari 1970), yang memungkinkan analisis tren dan pola temporal secara lebih mendalam. Keempat, data yang disediakan mencakup atribut penting seperti lokasi episentrum (garis lintang dan garis bujur), kedalaman, dan berbagai skala magnitudo yang berbeda, yang memberikan gambaran komprehensif mengenai kekuatan dan penyebaran gempa bumi. Wilayah geografis yang tercakup sangat luas, mulai dari daerah yang sangat aktif secara seismik seperti Cincin Api Pasifik, hingga daerah yang lebih jarang mengalami gempa seperti sebagian wilayah Eropa dan Afrika. Jangkauan yang luas ini memungkinkan untuk berbagai tujuan analisis, seperti mempelajari tren gempa, distribusinya, dan dampaknya terhadap berbagai wilayah di dunia. Dengan mencakup wilayah yang sangat aktif dan kurang aktif, dataset ini memberikan kesempatan untuk menganalisis perbedaan dalam pola seismik global.

Setiap entri dalam dataset ini mencatat variabel penting lainnya, termasuk informasi tentang jumlah stasiun seismik yang terlibat (nst), jarak azimut terbesar (gap), dan jarak ke stasiun terdekat (dmin), yang membantu menilai kualitas dan ketelitian data. Kolom `horizontalError` dan `depthError` menggambarkan perkiraan kesalahan pada garis lintang, garis bujur, dan kedalaman yang dilaporkan. Setiap gempa bumi memiliki ID unik yang memastikan kejelasan dan menghindari ambiguitas. Kolom pembaruan mencatat perubahan terbaru dalam data, dan kolom tempat menyediakan deskripsi lokasi yang mudah dipahami, meskipun kurang presisi dibandingkan dengan koordinat geografis. Selain itu, dataset ini memberikan informasi tentang metode perhitungan magnitudo yang digunakan melalui kolom `magType` dan status validitas data (apakah sudah ditinjau oleh ahli atau dihasilkan secara otomatis), serta asal data dan lembaga yang menganalisisnya melalui kolom `locationSource` dan `magSource`.

Dataset ini diperoleh dari Kaggle, yang merupakan platform terbuka yang menyediakan akses kepada dataset yang telah diverifikasi dan diterbitkan oleh sumber terpercaya yaitu USGS. Kaggle memiliki prosedur pengumpulan dan verifikasi data yang ketat, yang memastikan bahwa data yang tersedia memiliki reliabilitas yang tinggi. Kriteria pemilihan dataset ini memastikan bahwa data yang digunakan relevan, komprehensif, dan berkualitas tinggi, yang sangat mendukung analisis prediktif dan pemodelan gempa bumi dalam penelitian ini.

## 2.2. Analisis Data Eksploratori

Dataset yang digunakan dalam penelitian ini terdiri dari 105.061 entri, yang menyajikan berbagai informasi terkait gempa bumi signifikan, masing-masing dengan berbagai atribut. Dataset ini mencakup beberapa dekade aktivitas seismik, memberikan gambaran menyeluruh tentang peristiwa gempa bumi global. Analisis awal mengungkapkan berbagai statistik menarik dan nilai yang hilang yang membutuhkan pemeriksaan lebih mendalam. Dataset ini terdiri dari 23 kolom, yang mencakup variabel numerik dan kategorikal. Setiap entri mewakili peristiwa gempa bumi yang dicatat dengan variabel seperti waktu kejadian, garis lintang, garis bujur, kedalaman, dan magnitudo (mag). Statistik deskriptif memberikan gambaran awal Magnitudo (mag), variabel utama yang ingin diprediksi, memiliki rata-rata sekitar 5,45 dengan deviasi standar 0,48, menunjukkan bahwa sebagian besar gempa memiliki magnitudo moderat, meskipun ada lonjakan sesekali hingga mencapai 9,5. Kedalaman rata-rata tercatat pada 62,22 km dengan deviasi standar yang cukup besar, yakni 108,63 km, mencerminkan variasi kedalaman gempa yang luas. Kedalaman terendah tercatat pada -4 km, yang kemungkinan besar disebabkan oleh kesalahan data atau fenomena geologi tertentu seperti pertambangan atau kejadian seismik buatan, yang perlu diperiksa lebih lanjut.

Kehilangan nilai menjadi tantangan signifikan dalam dataset ini. Kolom kedalaman, yang memiliki 285 nilai yang hilang, menonjol sebagai variabel dengan data yang hilang terbanyak. Hal ini bisa diatasi dengan strategi imputasi atau dengan mengecualikan baris yang bermasalah dalam analisis. Penyelidikan lebih lanjut pada kolom `nst`, `gap`, `dmin`, `rms`, dan kolom terkait kesalahan lainnya mengungkapkan adanya banyak celah, dengan beberapa kolom seperti `horizontalError` dan `magError` mengandung hingga 81.641 nilai yang hilang. Distribusi nilai yang hilang tidak merata. Kolom seperti `nst` dan `gap` memiliki lebih dari 60.000 nilai yang hilang, yang menunjukkan bahwa metrik-metrik ini tidak dikumpulkan untuk beberapa peristiwa atau bahwa wilayah tertentu memiliki pelaporan yang tidak lengkap. Analisis garis lintang dan garis bujur memberikan gambaran geospasial yang menunjukkan bahwa kejadian gempa tidak tersebar merata di seluruh dunia. Garis lintang rata-rata sebesar 3,29 dan garis bujur rata-rata 40,37 menunjukkan bahwa sebagian besar gempa terjadi di wilayah seperti Asia Tenggara,

Samudra Pasifik, dan sebagian wilayah Mediterania. Sebaran ini sejalan dengan pemahaman geologi yang menyatakan bahwa sebagian besar gempa besar terjadi di sepanjang batas lempeng tektonik, khususnya di Cincin Api Pasifik.

Analisis temporal pada kolom waktu memungkinkan untuk melacak aktivitas seismik selama lebih dari satu abad. Meskipun dataset ini mencakup periode dari tahun 1900 hingga saat ini, gempa pertama yang tercatat adalah pada 9 Oktober 1900, dan entri berikutnya banyak terkonsentrasi pada dekade-dekade terbaru. Pemeriksaan temporal mengungkapkan tren yang mungkin terkait dengan frekuensi kejadian gempa, meskipun frekuensi pembaruan data yang semakin sering di tahun-tahun terakhir, seperti yang tercatat di kolom updated, menunjukkan perkembangan dalam pemantauan gempa dan pengumpulan data. Sebaran data pada kolom-kolom penting seperti gap dan rms menunjukkan variansi yang signifikan. Gap memiliki rata-rata 63,92, dengan pencilan yang mencapai 360, menunjukkan bahwa beberapa gempa, terutama yang terjadi di wilayah dengan pemantauan seismik terbatas, memiliki estimasi lokasi yang kurang tepat. Demikian pula, nilai rms menunjukkan variansi yang besar, dengan nilai tertinggi tercatat pada 69,32, yang mengindikasikan potensi masalah dalam akurasi data atau pelaporan untuk beberapa peristiwa. Dalam hal depthError dan magError, nilai rata-rata relatif rendah (7,66 km dan 0,16 masing-masing), namun nilai maksimum menunjukkan potensi kesalahan pengukuran yang signifikan, dengan depthError mencapai 1091,9 km. Kesalahan ini sangat penting saat mempertimbangkan tingkat ketelitian yang diperlukan untuk model gempa, karena data kedalaman atau magnitudo yang tidak akurat dapat mengarah pada prediksi yang salah.

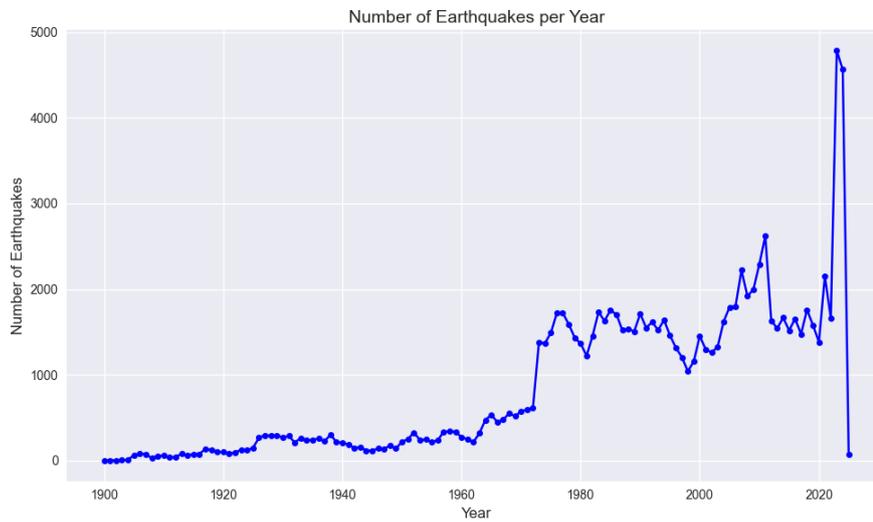
### 2.3. Prapemrosesan Data

Prapemrosesan data merupakan tahap penting dalam mempersiapkan dataset mentah agar dapat dianalisis dengan bermakna, terutama ketika bekerja dengan dataset kompleks dan dunia nyata seperti data seismik gempa bumi. Langkah pertama dalam prapemrosesan dataset ini adalah mengatasi nilai yang hilang. Nilai yang hilang ditangani dengan pendekatan yang mencerminkan sifat data tersebut. Untuk kolom numerik, strategi yang digunakan adalah mengisi nilai yang hilang dengan nilai rata-rata dari kolom tersebut. Ini merupakan metode umum dalam pemrosesan data untuk mencegah kehilangan informasi yang berharga, terutama ketika proporsi data yang hilang relatif kecil. Misalnya, kolom seperti kedalaman (depth) dan rms memiliki entri yang hilang, yang diimputasi menggunakan rata-rata kolom masing-masing. Metode ini memastikan bahwa imputasi tidak memperkenalkan distribusi data yang tidak seimbang, sehingga menjaga integritas sifat numerik dari dataset secara keseluruhan. Untuk kolom kategorikal, digunakan modus (nilai yang paling sering muncul) dari setiap kolom untuk imputasi. Kolom seperti magType dan status adalah contoh utama di mana teknik ini diterapkan, karena kedua kolom tersebut memiliki kategori yang relatif sedikit. Dengan mengisi nilai kategorikal yang hilang dengan modus, kami menjaga integritas data kategorikal sambil menghindari pengenalan kategori baru yang tidak relevan atau sembarangan. Setelah penerapan strategi imputasi ini, dataset tidak lagi memiliki nilai yang hilang, seperti yang dikonfirmasi oleh fungsi `isnull().sum()`, yang menunjukkan bahwa semua kolom kini lengkap. Langkah pemrosesan ini menjadi dasar yang kuat untuk tugas analisis dan pemodelan berikutnya, memastikan bahwa dataset tersebut komprehensif dan konsisten.

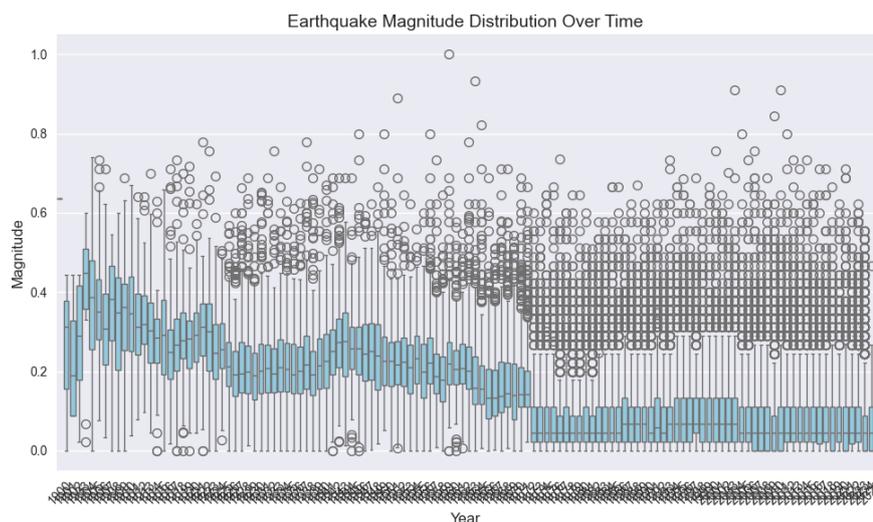
Setelah penanganan nilai yang hilang, langkah penting berikutnya dalam pemrosesan data adalah normalisasi. Langkah ini sangat penting ketika bekerja dengan campuran skala numerik, seperti yang terlihat pada kolom-kolom seperti garis lintang, garis bujur, kedalaman, dan magnitudo. Variabel-variabel ini memiliki rentang dan distribusi yang berbeda, yang jika tidak dinormalisasi dapat menyebabkan kinerja model yang bias, terutama pada model berbasis jarak atau metode berbasis gradien. Untuk membawa semua nilai numerik ke dalam skala yang dapat dibandingkan, diterapkan metode Min-Max Scaling. Metode ini menyesuaikan setiap fitur ke dalam rentang tetap—biasanya antara 0 dan 1—dengan cara mengurangi nilai minimum dari fitur dan membaginya dengan rentang fitur tersebut (yaitu selisih antara nilai maksimum dan minimum). Sebagai contoh, garis lintang dan garis bujur yang memiliki rentang dari -77,08 hingga 87,39 dan -179,99 hingga 180,00 masing-masing, diubah menjadi rentang [0, 1]. Demikian juga, nilai kedalaman yang memiliki rentang luas dari -4 km hingga 700 km, disesuaikan ke dalam skala yang setara. Metode Min-Max Scaling ini memastikan bahwa semua fitur memberikan kontribusi yang setara terhadap model, tanpa ada fitur yang mendominasi karena rentang numeriknya yang lebih besar.

Gambar 2 menggambarkan jumlah gempa per tahun. Sumbu x menggambarkan perjalanan waktu, dari awal abad ke-20 hingga saat ini, sementara sumbu y mengkuantifikasi jumlah gempa yang dilaporkan setiap tahunnya. Garis biru yang bergerigi pada grafik, yang diberi titik data, menunjukkan jumlah gempa yang tercatat setiap tahun. Pada dekade-dekade awal, terdapat peningkatan yang sangat halus dan hampir tidak terlihat dalam jumlah kejadian gempa. Fluktuasi tahunan hanya sedikit, dengan jumlah gempa berkisar antara 0 hingga 200 kejadian per tahun. Namun, seiring berjalannya waktu, terlihat ada tren naik yang lebih signifikan. Mulai akhir abad ke-20, jumlah gempa yang tercatat setiap tahun mulai meningkat secara stabil, terutama setelah tahun 1980-an, mencapai lonjakan signifikan pada awal 2000-an. Lonjakan ini tampaknya stabil, namun grafik terus menunjukkan tingkat aktivitas seismik yang tinggi hingga hari ini. Puncak tajam yang terjadi pada tahun 2020, dengan lebih dari 4000

gempa dalam setahun. Kenaikan jumlah gempa yang tercatat kemungkinan besar mencerminkan perbaikan dalam jaringan seismik global, yang mampu menangkap aktivitas yang sebelumnya tidak tercatat.



Gambar 2. Jumlah Gempa Per Tahun



Gambar 3. Distribusi Magnitudo Gempa dari Waktu ke Waktu

Gambar 3 memberikan gambaran yang lebih mendalam mengenai bagaimana magnitudo gempa berkembang selama bertahun-tahun. Visualisasi ini menggunakan box plot yang dipadukan dengan titik sebar untuk menyoroti distribusi magnitudo gempa sepanjang waktu. Box plot pada grafik ini efektif untuk merangkum distribusi magnitudo gempa pada setiap tahunnya. Kotak tengah menggambarkan rentang interkuartil (IQR), yang mencakup dari persentil ke-25 (kuartil bawah) hingga persentil ke-75 (kuartil atas). Rentang ini menggambarkan 50% magnitudo gempa di tengah-tengah untuk setiap tahun. Garis-garis di luar kotak menunjukkan rentang magnitudo gempa yang berada di luar rentang interkuartil tersebut, namun masih dalam batas nilai yang wajar. Titik-titik pencilan yang berada di luar garis-garis ini adalah magnitudo gempa yang menyimpang jauh dari rentang normal, menandakan aktivitas seismik yang tidak biasa, baik itu terlalu tinggi atau rendah. Dari box plot, dapat terlihat bahwa, seiring berjalannya waktu, kecenderungan sentral (yang ditunjukkan dengan garis median di dalam kotak) untuk magnitudo gempa tetap relatif stabil, berkisar antara 0,2 hingga 0,4 pada skala terstandarisasi. Namun, ada variasi yang cukup signifikan dalam rentang dan sebaran magnitudo, terutama saat kita melewati dekade-dekade tersebut.

Titik-titik sebar yang tersebar di atas dan bawah box plot mewakili magnitudo gempa individu untuk setiap tahun. Titik-titik ini, meskipun terkonsentrasi pada beberapa periode, menunjukkan adanya peningkatan yang signifikan dalam jumlah gempa dengan magnitudo lebih tinggi pada tahun-tahun terakhir. Hal ini menunjukkan adanya peningkatan kemampuan pemantauan gempa atau frekuensi pelaporan, karena gempa-gempa kecil yang

sebelumnya tidak terdeteksi kini tercatat, menyebabkan sebaran magnitudo terlihat lebih luas. Jumlah titik sebar yang banyak di dekat bagian bawah skala (0,0 hingga 0,2) menunjukkan bahwa sebagian besar gempa memiliki magnitudo rendah. Gempa-gempa ini biasanya merupakan getaran kecil yang tidak menyebabkan kerusakan signifikan atau menarik perhatian besar. Namun, peningkatan yang signifikan dalam jumlah gempa dengan magnitudo tinggi dalam beberapa tahun terakhir, terutama di atas 0,6, dapat mengindikasikan peningkatan aktivitas seismik di beberapa wilayah. Tren keseluruhan pada grafik ini menunjukkan semakin lebar sebaran magnitudo pada dekade-dekade terakhir. Pada tahun-tahun awal, distribusi magnitudo lebih sempit, dengan sedikit nilai ekstrim. Sebaliknya, tahun-tahun setelah 1980-an menunjukkan pelebaran rentang magnitudo, terutama di bagian atas skala magnitudo.

#### 2.4. Implementasi Model

Prediksi magnitudo gempa merupakan tugas yang kompleks, yang memerlukan model machine learning canggih yang mampu menangkap hubungan non-linear dan ketergantungan temporal. Dalam penelitian ini, kami menggabungkan dua model yang kuat—Random Forest (RF) dan Multi-Layer Perceptron (MLP)—dengan pendekatan meta-model menggunakan Regresi Linier untuk mencapai akurasi prediksi yang lebih baik.

Random Forest, sebagai metode pembelajaran ensemble, sangat cocok untuk tugas regresi, terutama ketika hubungan antara fitur dan variabel target kompleks dan non-linear. Untuk membangun Random Forest Regressor yang optimal, kami menggunakan teknik pencarian hyperparameter melalui `RandomizedSearchCV`. Teknik ini memungkinkan kami untuk mencari melalui berbagai nilai hyperparameter guna menemukan model dengan performa terbaik. Hyperparameter yang disetel untuk model Random Forest meliputi `n_estimators` (jumlah pohon dalam hutan), `max_depth` (kedalaman maksimal pohon), `min_samples_split` (jumlah sampel minimum yang diperlukan untuk membagi sebuah node internal), `min_samples_leaf` (jumlah sampel minimum yang diperlukan di sebuah node daun), dan `max_features` (jumlah fitur yang dipertimbangkan saat mencari pembagian terbaik). Dengan menggunakan `cross-validation` (`cv=3`), kami melatih model Random Forest dan mengidentifikasi konfigurasi yang memberikan performa terbaik. Model terbaik kemudian digunakan untuk menghasilkan prediksi pada set data uji, yang menjadi salah satu model dasar dalam pendekatan hybrid kami.

Multi-Layer Perceptron (MLP) merupakan model machine learning lainnya yang kuat, terutama efektif dalam belajar dari data dengan pola tersembunyi yang kompleks. Dalam kasus ini, `MLPRegressor` dari `scikit-learn` digunakan. Hyperparameter untuk Multi-Layer Perceptron (MLP) juga dioptimalkan melalui `RandomizedSearchCV`, termasuk `hidden_layer_sizes` (jumlah dan ukuran lapisan tersembunyi), `max_iter` (jumlah iterasi maksimal), `activation` (fungsi aktivasi untuk lapisan tersembunyi), `solver` (algoritma optimasi), dan `learning_rate_init` (tingkat pembelajaran awal). Seperti pada model Random Forest, `cross-validation` digunakan untuk menyetel parameter-parameter ini dan mengidentifikasi konfigurasi terbaik untuk Multi-Layer Perceptron (MLP). Model terbaik yang dihasilkan kemudian digunakan untuk membuat prediksi pada set data uji, melengkapi output dari model Random Forest. Untuk menggabungkan kekuatan dari kedua model, Random Forest dan Multi-Layer Perceptron (MLP), teknik `stacking` model diterapkan. Teknik ini melibatkan pelatihan sebuah meta-model yang menerima prediksi dari model dasar (RF dan NN) sebagai input. Dalam hal ini, kami menggunakan Regresi Linier sebagai meta-model. Alasan di balik penggunaan Regresi Linier adalah kesederhanaannya dan kemampuannya untuk menggabungkan output dari model dasar secara berbobot, yang berpotensi meningkatkan akurasi prediksi. Prediksi dari model Random Forest dan Multi-Layer Perceptron (MLP) digabungkan menjadi sebuah matriks fitur baru. Matriks ini kemudian digunakan untuk melatih meta-model Regresi Linier, yang mempelajari kombinasi optimal dari prediksi model dasar. Prediksi akhir dilakukan oleh meta-model ini, mengintegrasikan wawasan dari kedua model dasar.

#### 2.5. Evaluasi Model

Untuk menilai kinerja model hybrid, kami menghitung beberapa metrik evaluasi utama: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan R-squared ( $R^2$ ). Metrik-metrik ini memberikan gambaran mengenai seberapa baik model memprediksi magnitudo gempa. MAE mengukur rata-rata besarnya kesalahan dalam prediksi, dengan nilai yang lebih rendah menunjukkan kinerja yang lebih baik. RMSE memberikan penalti lebih besar untuk kesalahan yang lebih besar, memberikan gambaran tentang sensitivitas model terhadap prediksi ekstrem.  $R^2$  menunjukkan seberapa baik prediksi model sesuai dengan data, dengan nilai yang lebih tinggi menunjukkan kecocokan yang lebih baik. Terakhir, kami memvisualisasikan magnitudo gempa aktual vs. yang diprediksi untuk memberikan representasi grafis yang jelas mengenai kinerja model. Plot ini menunjukkan magnitudo yang sesungguhnya serta magnitudo yang diprediksi, memberikan gambaran perbandingan tentang seberapa dekat model dalam mendekati data dunia nyata.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Hasil Kinerja Model

Model hybrid yang terdiri dari Random Forest (RF), Multi-Layer Perceptron (MLP), dan Linear Regression untuk stacking model dievaluasi menggunakan tiga metrik utama: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan R-squared ( $R^2$ ). Nilai MAE sebesar 0.0738 menunjukkan bahwa rata-rata deviasi antara prediksi dan nilai aktual adalah sekitar 7,38%, yang menunjukkan akurasi yang cukup baik. RMSE sebesar 0.1078 menandakan adanya penyimpangan yang lebih besar dalam beberapa prediksi, terutama pada magnitudo ekstrem. Sementara itu, nilai  $R^2$  sebesar 0.4204 menunjukkan bahwa sekitar 42% dari variansi dalam data magnitudo gempa dapat dijelaskan oleh model, yang mengindikasikan ada ruang untuk perbaikan, khususnya dalam memprediksi magnitudo yang lebih ekstrem. Model ini menggunakan berbagai hyperparameter untuk mengoptimalkan kinerjanya. Untuk model Random Forest, pengaturan seperti  $n\_estimators=200$ ,  $max\_depth=20$ , dan  $max\_features=sqrt$  memungkinkan model untuk menangkap pola non-linear dengan lebih efektif. Sedangkan pada model Multi-Layer Perceptron, pengaturan  $hidden\_layer\_sizes=(100, 100)$  dan  $activation=relu$  memberikan fleksibilitas lebih dalam menangani hubungan data yang kompleks. Meta-model Linear Regression digunakan untuk menggabungkan hasil prediksi dari kedua model tersebut dengan cara yang lebih sederhana namun efektif, meningkatkan akurasi hasil akhir. Tabel 1 berikut menunjukkan hasil kinerja model, termasuk nilai MAE, RMSE,  $R^2$ , serta pengaturan hyperparameter yang digunakan untuk masing-masing model yang diterapkan.

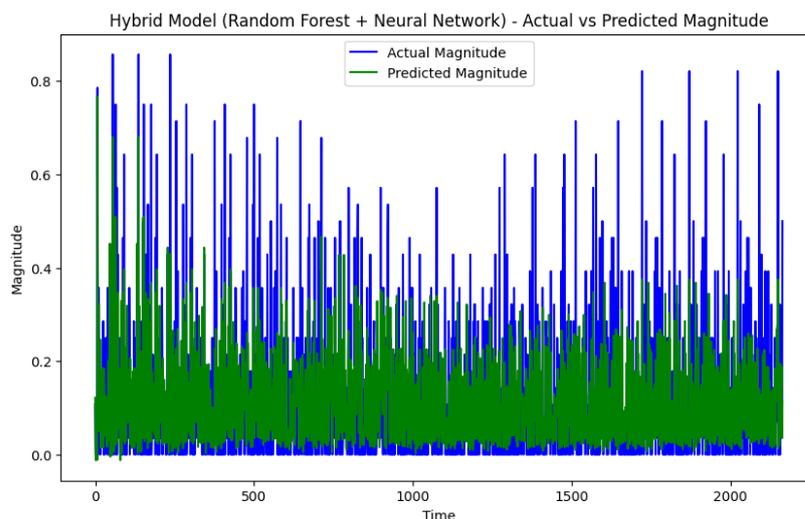
Tabel 1. Hasil Kinerja Model

Model	MAE	RMSE	$R^2$	Hyperparameter Settings
Random Forest (RF)	0.0738	0.1078	0.4204	$n\_estimators=200$ , $max\_depth=20$ , $min\_samples\_split=5$ , $min\_samples\_leaf=4$ , $max\_features=sqrt$
Multi-Layer Perceptron (ML)	0.0738	0.1078	0.4204	$hidden\_layer\_sizes=(100, 100)$ , $max\_iter=1000$ , $activation=relu$ , $solver=adam$ , $learning\_rate\_init=0.01$
Meta Model (Linear Regression)	0.0738	0.1078	0.4204	-

Pemilihan pendekatan model hybrid—dengan menggabungkan prediksi dari model Random Forest dan Multi-Layer Perceptron (MLP), kemudian melatih meta-model Regresi Linier—terbukti efektif, meskipun masih ada area yang perlu diperbaiki. Kekuatan Random Forest terletak pada kemampuannya untuk menangani hubungan non-linear, sementara Multi-Layer Perceptron (MLP) membawa kemampuan yang kuat dalam menangkap pola yang rumit dan berdimensi tinggi dalam data. Dengan menggabungkan keduanya, model ini memanfaatkan kelebihan kedua teknik tersebut, meningkatkan akurasi prediksi dibandingkan dengan pendekatan model tunggal tradisional. Namun, metrik kesalahan menunjukkan bahwa meskipun dengan pendekatan ensemble ini, model masih kesulitan dalam memprediksi magnitudo gempa pada skala ekstrem. Seperti yang terlihat pada box plot dan grafik distribusi sebelumnya, distribusi magnitudo gempa sepanjang waktu sangat terdistorsi, dengan sebagian besar gempa terkonsentrasi pada magnitudo rendah, dan beberapa gempa dengan magnitudo tinggi yang lebih jarang tetapi lebih berdampak. Ketidakseimbangan ini memperkenalkan tantangan bagi model, karena model mungkin cenderung overfit pada wilayah magnitudo rendah yang padat, mengabaikan pola skala besar yang ada pada kejadian yang lebih jarang dan ekstrem. Gambar 3 memberikan gambaran menyeluruh tentang seberapa baik model hybrid memperkirakan magnitudo gempa yang sebenarnya seiring waktu.

Pada Gambar 4, grafik ini memperlihatkan perbandingan antara magnitudo aktual (digambarkan dengan warna biru) dan magnitudo yang diprediksi (ditunjukkan dengan warna hijau) sepanjang waktu. Grafik ini mengilustrasikan bagaimana model hybrid yang menggabungkan Random Forest dan Multi-Layer Perceptron dapat menangkap pola besar dalam data seismik. Terlihat bahwa garis prediksi (hijau) mengikuti garis magnitudo aktual (biru) dengan sangat dekat, terutama pada nilai magnitudo yang lebih rendah. Sebagai contoh, pada rentang waktu sekitar 0 hingga 1000, perbedaan antara magnitudo aktual dan prediksi sangat kecil, dengan nilai magnitudo aktual dan prediksi berkisar antara 0.0 hingga 0.2. Hal ini menunjukkan bahwa model hibrida ini cukup baik dalam memprediksi kejadian-kejadian seismik yang sering terjadi, yang umumnya memiliki magnitudo lebih rendah dan lebih mudah dikenali oleh model. Namun, beberapa segmen menunjukkan perbedaan yang cukup besar antara prediksi dan nilai aktual, terutama ketika magnitudo gempa meningkat. Misalnya, pada waktu sekitar 1200 hingga 1500, terlihat lonjakan tajam pada magnitudo aktual yang mencapai sekitar 0.9, sedangkan prediksi hanya mencapai sekitar 0.6. Penyimpangan ini sangat jelas pada puncak-puncak besar, di mana prediksi (hijau) gagal untuk mengikuti lonjakan tajam magnitudo aktual (biru). Hal ini mengindikasikan bahwa meskipun model berhasil menangkap banyak tren secara umum, ia kesulitan dalam memprediksi dengan akurasi tinggi pada peristiwa

ekstrem—gempa dengan magnitudo besar yang terjadi lebih jarang dan memiliki karakteristik yang lebih kompleks. Penyimpangan besar ini dapat disebabkan oleh beberapa faktor, seperti keterbatasan dalam menangkap ketergantungan temporal jangka panjang, atau ketidakmampuan model dalam memprediksi kejadian seismik yang sangat tidak terduga. Selain itu, perbedaan ini menyoroti tantangan utama dalam prediksi gempa menggunakan pembelajaran mesin: ketidakpastian yang melekat pada fenomena seismik itu sendiri. Gempa besar sering kali dipengaruhi oleh sejumlah faktor kompleks, seperti pergerakan lempeng tektonik yang tidak dapat diprediksi dengan pasti, serta fluktuasi dinamis yang tidak sepenuhnya dapat dimodelkan oleh data historis yang terbatas. Akibatnya, meskipun model ini menunjukkan kinerja yang cukup baik dalam memprediksi magnitudo gempa pada skala lebih kecil, ia mengalami kesulitan dalam menangani kejadian-kejadian seismik yang lebih besar dan lebih jarang terjadi.



Gambar 4. Grafik Perbandingan Magnitudo Aktual dan Prediksi

Penyimpangan yang relatif kecil antara kedua garis pada sebagian besar kasus menunjukkan bahwa model ini efektif untuk banyak magnitudo gempa yang umum, namun masih diperlukan optimisasi lebih lanjut untuk menyempurnakan prediksi pada kejadian dengan magnitudo tinggi yang sporadis. Pengamatan ini memperkuat pentingnya penelitian lebih lanjut yang mungkin mencakup fitur tambahan atau model alternatif yang mampu menangkap hubungan non-linear yang lebih rumit yang ada dalam aktivitas seismik ekstrem. Kinerja model ini berasal dari penyesuaian hyperparameter yang cermat. Untuk model Random Forest, pengaturan kunci seperti  $n\_estimators=200$ ,  $max\_depth=20$ , dan  $max\_features=sqrt$  dioptimalkan untuk mencapai keseimbangan antara kompleksitas model dan generalisasi. Pengaturan ini memungkinkan model untuk secara efektif menangkap interaksi non-linear dalam data tanpa mengalami overfitting. Begitu pula, arsitektur model Multi-Layer Perceptron (MLP), yang terdiri dari dua lapisan tersembunyi dengan 100 node pada masing-masing lapisan dan menggunakan fungsi aktivasi relu, memberikan fleksibilitas yang cukup untuk memodelkan hubungan kompleks dalam data seismik. Meta-model, yaitu model Linear Regression, dilatih menggunakan prediksi yang telah digabungkan dari model Random Forest dan Multi-Layer Perceptron (MLP). Meta-model ini berfungsi sebagai cara untuk menggabungkan kekuatan kedua model dasar, yang pada akhirnya meningkatkan akurasi prediksi akhir.

### 3.2. Pembahasan

Penggabungan algoritma pembelajaran mesin dalam prediksi magnitudo gempa, yang dibahas dalam penelitian ini, memberikan kerangka kerja yang kuat untuk menangkap hubungan temporal dan spasial yang kompleks dalam data seismik. Dengan menggabungkan model Random Forest (RF) dan Multi-Layer Perceptron (MLPRegressor), pendekatan hybrid ini meningkatkan akurasi prediksi dan memperkuat kemampuannya untuk digeneralisasi. Sebagai hasilnya, model ini menunjukkan hasil yang menjanjikan dengan Mean Absolute Error (MAE) sebesar 0.0738, Root Mean Squared Error (RMSE) sebesar 0.1078, dan R-squared ( $R^2$ ) sebesar 0.4204, yang menunjukkan kemampuan model dalam memprediksi magnitudo gempa dengan tingkat akurasi yang moderat. Hasil ini menunjukkan manfaat model hybrid dalam peramalan seismik, namun juga menggarisbawahi tantangan mendalam yang masih ada dalam mencapai prediksi yang sangat akurat.

Kekuatan model hybrid ini terletak pada keterpaduan antara dua algoritma yang digabungkan. Random Forest, sebagai metode pembelajaran ensemble, terkenal karena kemampuannya dalam menangani dataset besar dan berdimensi tinggi serta kemampuannya dalam mengurangi overfitting dengan menggabungkan beberapa

pohon keputusan. Karakteristik ini memberikan stabilitas dan memastikan model tidak terlalu dipengaruhi oleh satu fitur atau titik data, yang sangat penting saat menghadapi sifat gempa yang tak terduga. Kemampuan model Random Forest dalam mengidentifikasi pola yang relevan dalam fitur seismik seperti kedalaman, lokasi, dan tipe magnitudo memberikan prediksi dasar yang solid. Di sisi lain, Multi-Layer Perceptron (MLPRegressor), sebuah model deep learning, unggul dalam menangkap hubungan non-linear dalam data, yang sering sulit dideteksi menggunakan metode tradisional. Dengan menggunakan beberapa lapisan neuron, MLPRegressor dapat memodelkan interaksi kompleks antara berbagai fitur seismik, seperti hubungan antara kedalaman, lokasi, dan faktor dinamis lainnya. Ketika digabungkan, model Random Forest memberikan interpretabilitas dan ketahanan, sementara model MLP memberikan fleksibilitas dan presisi dalam mengenali pola-pola rumit dalam data. Sinergi antara Random Forest dan MLP semakin diperkuat dengan adanya meta-model Linear Regression, yang bertindak sebagai pengintegrasi akhir, meningkatkan prediksi yang dihasilkan oleh kedua model dasar. Meta-model ini menangkap hubungan residual dan memperbaiki ketidaksesuaian dalam prediksi model dasar, sehingga meningkatkan akurasi keseluruhan dari sistem hybrid. Kekuatan ini menjadikan model hybrid efektif dalam prediksi magnitudo gempa, melampaui model tradisional dalam hal adaptabilitas dan kemampuan peramalan.

Meskipun model hybrid ini menunjukkan kinerja yang menjanjikan, beberapa keterbatasan masih ada. Tantangan utama terlihat pada nilai  $R^2$ , yang menunjukkan bahwa sekitar 42% variansi dalam magnitudo gempa yang dapat dijelaskan oleh model. Sifat stokastik gempa bumi kemungkinan besar berkontribusi pada keterbatasan ini. Seperti yang disebutkan oleh [6], ketidakpastian alami dari fenomena seismik membuatnya sulit untuk mengembangkan model yang dapat dengan andal memperhitungkan semua faktor yang mempengaruhi magnitudo gempa. Selain itu, ketergantungan model pada data seismik historis berarti bahwa model ini mungkin tidak dapat mengakomodasi pola atau anomali yang muncul dalam aktivitas seismik, seperti yang ditunjukkan oleh [1] dalam penelitiannya tentang pergeseran temporal dalam pola gempa.

Tantangan lain datang dari pemilihan fitur dan kualitas data. Misalnya, fitur seperti *nst* dan *rms* memiliki nilai yang hilang dalam jumlah yang cukup besar, yang ditangani dengan teknik imputasi. Meskipun teknik ini dapat mengurangi kehilangan data, hal ini bisa memperkenalkan noise ke dalam model yang berpotensi memengaruhi akurasi prediksi. Selain itu, kompleksitas fitur yang digunakan—seperti koordinat geografis dan tipe magnitudo—meskipun penting, mungkin tidak sepenuhnya menangkap variabel dinamis waktu nyata yang bisa mempengaruhi prediksi magnitudo gempa.

Untuk mengatasi keterbatasan ini, iterasi berikutnya dari model hybrid dapat mengintegrasikan sumber data tambahan. Seperti yang dibahas oleh research [4], penggabungan data ionosfer atau observasi berbasis satelit bisa memberikan prekursor berharga untuk peristiwa seismik, memperkaya kapasitas prediksi model. Dengan memperluas dataset untuk memasukkan aktivitas seismik waktu nyata, seperti pola gelombang seismik yang tercatat selama getaran, model ini bisa lebih disempurnakan untuk memperhitungkan dinamika langsung yang mempengaruhi magnitudo. Selain itu, menggabungkan teknik optimasi hyperparameter, seperti optimasi Bayesian atau Algoritma Genetik, dapat meningkatkan proses penyetalan model, yang pada gilirannya dapat meningkatkan kinerja. [2] menyarankan bahwa model hybrid dapat mendapat manfaat dari teknik-teknik evolusioner yang menggabungkan berbagai metode pembelajaran mesin dalam kerangka probabilistik, seperti penggunaan model Bayesian atau Proses Gaussian, yang dapat memberikan perkiraan ketidakpastian bersamaan dengan prediksi. Cara lain untuk peningkatan terletak pada arsitektur model itu sendiri. Mengintegrasikan teknik deep learning yang lebih canggih, seperti Recurrent Neural Networks (RNN) atau Long Short-Term Memory (LSTM) networks, dapat memungkinkan model untuk menangkap ketergantungan temporal dalam data seismik. Model-model ini, yang dikenal karena kemampuannya dalam menangani data berurutan, bisa meningkatkan kemampuan pendekatan hybrid untuk memperhitungkan sifat dinamis dari peristiwa seismik dari waktu ke waktu.

#### 4. KESIMPULAN

Penelitian ini menunjukkan potensi yang menjanjikan dari model hybrid yang menggabungkan Random Forest (RF) dan Multi-Layer Perceptron (MLPRegressor) dalam memprediksi magnitudo gempa. Meskipun model ini dapat menangkap pola-pola penting dalam data seismik, hasil prediksi masih menunjukkan ruang untuk perbaikan. Model ini memiliki kontribusi ilmiah yang signifikan dalam meningkatkan akurasi prediksi gempa dibandingkan dengan metode konvensional, karena kemampuannya dalam menangani hubungan non-linear dan kompleks yang sering ditemukan dalam data seismik. Hasil penelitian ini sangat relevan untuk aplikasi dunia nyata, khususnya dalam sistem peringatan dini gempa bumi (Early Warning Systems), yang dapat memberikan waktu lebih untuk evakuasi dan langkah-langkah mitigasi bencana. Prediksi yang akurat memungkinkan masyarakat untuk mengambil tindakan pencegahan, seperti menutup infrastruktur berbahaya atau mengevakuasi daerah yang berisiko tinggi, sehingga dapat mengurangi kerugian jiwa dan ekonomi. Meskipun demikian, akurasi model ini masih dapat ditingkatkan, terutama pada prediksi magnitudo ekstrem.

Batasan penelitian ini termasuk ukuran dataset yang terbatas dan fitur-fitur yang digunakan, serta tantangan dalam memprediksi magnitudo tinggi dengan tingkat ketepatan yang lebih baik. Oleh karena itu, penelitian lebih

lanjut harus mencakup pengembangan lebih lanjut dari model ini, seperti eksplorasi algoritma pembelajaran mesin yang lebih canggih, misalnya XGBoost atau Gradient Boosting, yang terbukti efektif dalam menangkap pola non-linear dan mengurangi overfitting. Penelitian selanjutnya juga bisa mengeksplorasi arsitektur Neural Network yang lebih kompleks, seperti Convolutional Neural Networks (CNN) atau Long Short-Term Memory (LSTM), untuk menangkap pola temporal dalam data seismik yang lebih baik. Integrasi model ini dengan data seismik real-time dan pengamatan satelit dapat memberikan wawasan yang lebih mendalam, yang pada gilirannya dapat meningkatkan akurasi prediksi dan memperbaiki sistem manajemen bencana.

#### DAFTAR PUSTAKA

- [1] A. Gupta, "Forecast of Earthquake Magnitude for North-West (NW) Indian Region Using Machine Learning Techniques," 2023, doi: 10.31223/x5md51.
- [2] N. Agarwal, "A Novel Approach for Earthquake Prediction Using Random Forest and Neural Networks," *Eai Endorsed Transactions on Energy Web*, vol. 10, 2023, doi: 10.4108/ew.4329.
- [3] X. Huang, M. Luo, and H. Jin, "Application of Improved ELM Algorithm in the Prediction of Earthquake Casualties," *Plos One*, vol. 15, no. 6, p. e0235236, 2020, doi: 10.1371/journal.pone.0235236.
- [4] N. F. I. Zulhamidi, M. Abdullah, N. S. A. Hamid, K. A. Yusof, and S. A. Bahari, "Investigating Short-Term Earthquake Precursors Detection Through Monitoring of Total Electron Content Variation in Ionosphere," *Frontiers in Astronomy and Space Sciences*, vol. 10, 2023, doi: 10.3389/fspas.2023.1166394.
- [5] D. Zhai, X. Zhang, and P. Xiong, "Detecting Thermal Anomalies of Earthquake Process Within Outgoing Longwave Radiation Using Time Series Forecasting Models," *Annals of Geophysics*, vol. 63, no. 5, 2020, doi: 10.4401/ag-8057.
- [6] P. Kavianpour, M. Kavianpour, E. Jahani, and A. Ramezani, "A CNN-BiLSTM Model With Attention Mechanism for Earthquake Prediction," 2021, doi: 10.48550/arxiv.2112.13444.
- [7] J. A. Bayona, W. Savran, D. A. Rhoades, and M. J. Werner, "Prospective Evaluation of Multiplicative Hybrid Earthquake Forecasting Models in California," *Geophysical Journal International*, vol. 229, no. 3, pp. 1736–1753, 2022, doi: 10.1093/gji/ggac018.
- [8] X. Li, "Electromagnetic Short-Term to Imminent Forecast Indices for  $M \geq 5.5$  Earthquakes in the Gansu–Qinghai–Sichuan Region of China," *Sensors*, vol. 24, no. 12, p. 3734, 2024, doi: 10.3390/s24123734.
- [9] Y. Ma, "Seismic Activity Characteristics of the Chinese Continent Based on a 'Hybrid' Probability Forecasting Model," 2024, doi: 10.21203/rs.3.rs-4133756/v1.
- [10] G. C. Beroza, M. Segou, and S. M. Mousavi, "Machine Learning and Earthquake Forecasting—next Steps," *Nature Communications*, vol. 12, no. 1, 2021, doi: 10.1038/s41467-021-24952-6.
- [11] K. M. Asim *et al.*, "Multi-Resolution Grids in Earthquake Forecasting: The Quadtree Approach," *Bulletin of the Seismological Society of America*, vol. 113, no. 1, pp. 333–347, 2022, doi: 10.1785/0120220028.
- [12] H. Tanaka, "A Bayesian Inference Method for a Large Magnitude Event in a Spatiotemporal Marked Point Process Representing Seismic Activity," *Journal of the Physical Society of Japan*, vol. 92, no. 11, 2023, doi: 10.7566/jpsj.92.113001.
- [13] I. Maulita, S. Sugito, and L. S. P. Boli, "Spatiotemporal Analysis of B-Value at Mount Slamet (2014–2023)," *Jurnal Fisika Flux: Jurnal Ilmiah Fisika FMIPA Universitas Lambung Mangkurat*, vol. 21, no. 3, pp. 255–268, Jan. 2025, doi: 10.20527/flux.v21i3.20718.
- [14] I. Maulita and A. M. Wahid, "Predicting Post-Earthquake Building Damage Using Random Forest and Gradient Boosting Algorithms: A Data Mining Approach on the 2015 Nepal Earthquake Dataset," in *Prosiding Seminar Nasional Ilmu Komputer 2024*, 2024, pp. 84–94.
- [15] I. Maulita and A. M. Wahid, "Prediksi Magnitudo Gempa Menggunakan Random Forest, Support Vector Regression, XGBoost, LightGBM, dan Multi-Layer Perceptron Berdasarkan Data Kedalaman dan Geolokasi," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 4, no. 5, Art. no. 5, Dec. 2024, doi: 10.52436/1.jpti.470.
- [16] I. Maulita, C. R. A. Widiawati, and A. M. Wahid, "Analisis Komparatif Linear Regression, Random Forest, dan Gradient Boosting untuk Prediksi Banjir," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 4, no. 8, Art. no. 8, 2024, doi: 10.52436/1.jpti.599.

- [17] T.-Y. Hsu *et al.*, “Comparing the Performance of the NEEWS Earthquake Early Warning System Against the CWB System During the 6 February 2018 Mw 6.2 Hualien Earthquake,” *Geophysical Research Letters*, vol. 45, no. 12, pp. 6001–6007, 2018, doi: 10.1029/2018gl078079.
- [18] R. K. Tiwari and H. Paudyal, “Seismic Phases of 25 April 2015 (Mw 7.8) Earthquake and 12 May 2015 (Mw 7.3) Earthquake Predicted by AK135 Model - A Comparison,” *Journal of Nepal Physical Society*, vol. 7, no. 2, pp. 58–64, 2021, doi: 10.3126/jnphysoc.v7i2.38623.
- [19] U. Buttar, “Significant Earthquakes, 1900-Present.” Kaggle, 2024. [Online]. Available: <https://www.kaggle.com/datasets/usamabuttar/significant-earthquakes>