Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026, Page: 12−22
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v5i1.226
https://journal.grahamitra.id/index.php/buai

Prediksi Magnitudo Gempa Indonesia Menggunakan Machine Learning Berbasis Data Seismik

Farizi Ilham*, Halili Maar, Dimas Lendesi

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia
Email: [1,*]dosen02954@unpam.ac.id , [2]dosen2957@unpam.ac.id, [3]dosen2959@unpam.ac.id
Email Penulis Korespondensi: dosen02954@unpam.ac.id

Abstrak−Indonesia merupakan wilayah dengan aktivitas seismik tinggi sehingga prediksi magnitudo gempa menjadi aspek penting dalam mitigasi bencana. Penelitian ini bertujuan membandingkan performa beberapa algoritma *machine learning* untuk memprediksi magnitudo gempa berdasarkan data seismik Indonesia periode 2016–2023 yang diperoleh dari *United States Geological Survey* (USGS). Dataset terdiri atas 17.331 kejadian gempa dengan parameter spasial, kualitas pengukuran, dan fitur temporal. Tahapan penelitian meliputi *data preprocessing*, *feature engineering*, pelatihan model, serta evaluasi menggunakan Coefficient of Determination (R^2), Mean Absolute Error (MAE), dan Root Mean Square Error (RMSE). Empat algoritma yang dibandingkan adalah Linear Regression, Support Vector Regression, Random Forest Regression, dan Gradient Boosting Regression. Hasil pengujian menunjukkan bahwa Random Forest Regression menghasilkan performa terbaik dengan nilai R^2 sebesar 0,767, MAE sebesar 0,123, dan RMSE sebesar 0,179, diikuti oleh Gradient Boosting Regression dengan nilai R^2 sebesar 0,746. Analisis *feature importance* menunjukkan bahwa variabel magError, depthError, depth, dan gap merupakan faktor yang paling berpengaruh terhadap prediksi magnitudo. Hasil penelitian menunjukkan bahwa pendekatan *ensemble learning* mampu meningkatkan akurasi prediksi serta memberikan interpretasi terhadap faktor-faktor utama yang memengaruhi magnitudo gempa sehingga berpotensi mendukung pengembangan sistem mitigasi bencana berbasis data.

Kata Kunci: Gradient Boosting Regression; Prediksi Magnitudo Gempa; Data Seismik; Feature Importance; Machine Learning

Abstract−Indonesia is one of the world's most seismically active regions, making earthquake magnitude prediction an important component of disaster mitigation. This study compares several *machine learning* regression algorithms for predicting earthquake magnitude using Indonesian seismic data collected from the *United States Geological Survey* (USGS) during 2016–2023. The dataset contains 17,331 earthquake records consisting of spatial, measurement-quality, and temporal attributes. The research process included data preprocessing, feature engineering, model training, and evaluation using the Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). Four regression algorithms were evaluated: Linear Regression, Support Vector Regression, Random Forest Regression, and Gradient Boosting Regression. Experimental results show that Random Forest Regression achieved the best performance with an R^2 of 0.767, MAE of 0.123, and RMSE of 0.179, followed by Gradient Boosting Regression with an R^2 of 0.746. Feature importance analysis indicates that magError, depthError, depth, and gap are the most influential variables in earthquake magnitude prediction. These findings demonstrate that ensemble learning provides accurate prediction performance while offering meaningful interpretation of influential seismic parameters for disaster mitigation systems..

Keywords: Gradient Boosting Regression; Earthquake Magnitude Prediction; Seismic Data; Feature Importance; Machine Learning

1. PENDAHULUAN

Gempa bumi merupakan salah satu bencana alam yang memiliki tingkat risiko tinggi karena dapat menimbulkan kerusakan infrastruktur, korban jiwa, serta kerugian ekonomi dalam skala besar. Indonesia termasuk negara dengan aktivitas seismik tertinggi di dunia karena berada pada pertemuan Lempeng Indo-Australia, Eurasia, Pasifik, dan Filipina yang membentuk kawasan *Ring of Fire*. Kondisi geologi tersebut menyebabkan ribuan kejadian gempa bumi tercatat setiap tahun dengan karakteristik magnitudo, kedalaman, dan lokasi yang beragam. Besarnya aktivitas seismik tersebut menjadikan prediksi magnitudo gempa sebagai salah satu aspek penting dalam mendukung sistem mitigasi bencana, perencanaan pembangunan infrastruktur, serta penyusunan kebijakan penanggulangan bencana berbasis data [1].

Perkembangan teknologi *machine learning* memberikan peluang baru dalam meningkatkan kemampuan analisis data seismik. Berbeda dengan pendekatan statistik konvensional yang umumnya mengasumsikan hubungan linier antarvariabel, algoritma *machine learning* mampu mempelajari hubungan kompleks dan nonlinier dari data dalam jumlah besar. Pada bidang seismologi, berbagai algoritma telah digunakan untuk memprediksi parameter gempa, seperti *Support Vector Regression[2]*, *Random Forest[3]*, *Extreme Gradient Boosting[4]*, hingga *Deep Learning[5]*. Kemampuan algoritma tersebut dalam mengenali pola tersembunyi dari parameter seismik menjadikannya sebagai pendekatan yang semakin banyak digunakan untuk mendukung sistem prediksi gempa bumi modern.

Meskipun demikian, prediksi magnitudo gempa masih menjadi permasalahan yang cukup menantang. Karakteristik data seismik yang bersifat heterogen, mengandung *noise*, memiliki hubungan nonlinier, serta dipengaruhi oleh banyak parameter menyebabkan model prediksi sering mengalami penurunan akurasi ketika diterapkan pada data nyata. Selain itu, kualitas pengukuran dari stasiun seismik yang berbeda dapat menghasilkan variasi tingkat kesalahan (*measurement error*) sehingga memengaruhi kemampuan model dalam melakukan generalisasi. Kondisi tersebut menyebabkan model prediksi yang memiliki performa baik pada suatu wilayah belum tentu memberikan hasil serupa ketika diterapkan pada wilayah lain dengan karakteristik geologi yang berbeda [6].

© 2026 The Author(s). Published by *CV. Graha Mitra Edukasi*. Page 12
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026, Page: 12−22
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v5i1.226
https://journal.grahamitra.id/index.php/buai

Beberapa penelitian telah menerapkan pendekatan *machine learning* dalam prediksi magnitudo gempa dengan berbagai metode dan dataset. Joshi *et al.* [7] mengembangkan model *stacking ensemble* yang dipadukan dengan *Conditional Tabular Generative Adversarial Network* (CTGAN) untuk meningkatkan akurasi prediksi magnitudo pada sistem *earthquake early warning*. Selanjutnya, Jang *et al.* [8] menerapkan pendekatan *Explainable Machine Learning* berbasis XGBoost dan SHAP untuk mengidentifikasi fitur-fitur yang paling berpengaruh terhadap kejadian gempa besar. Dhulipalla dan Vincent [9] mengevaluasi beberapa algoritma *machine learning* yang dikombinasikan dengan metode seleksi fitur berbasis metaheuristik dan menunjukkan bahwa Random Forest yang dioptimalkan menggunakan *Particle Swarm Optimization* menghasilkan performa terbaik. Selain itu, Zhao *et al.* [10] mengembangkan dataset benchmark DiTing yang dirancang untuk mendukung pengembangan berbagai model kecerdasan buatan pada bidang seismologi, termasuk prediksi magnitudo gempa dan sistem *early warning*.

Meskipun penelitian-penelitian tersebut menunjukkan bahwa *machine learning* mampu meningkatkan akurasi prediksi magnitudo gempa, sebagian besar masih berfokus pada pengembangan model tertentu, optimasi menggunakan metode tambahan, atau penyediaan dataset benchmark. Belum banyak penelitian yang melakukan perbandingan beberapa algoritma regresi menggunakan dataset historis USGS serta mengkaji kontribusi masing-masing parameter seismik melalui analisis *feature importance*. Oleh karena itu, penelitian ini membandingkan Linear Regression, Support Vector Regression, Random Forest Regression, dan Gradient Boosting Regression pada data seismik Indonesia untuk menentukan model yang paling efektif sekaligus mengidentifikasi variabel yang paling berpengaruh dalam proses prediksi magnitudo gempa [11].

Berdasarkan *research gap* tersebut, penelitian ini mengusulkan penerapan algoritma *Gradient Boosting Regression* untuk memprediksi magnitudo gempa bumi menggunakan data seismik Indonesia periode 2016–2023 yang diperoleh dari *United States Geological Survey* (*USGS*). Metode *Gradient Boosting Regression* dipilih karena memiliki kemampuan membangun model secara bertahap melalui proses *boosting* sehingga mampu meminimalkan kesalahan prediksi pada setiap iterasi. Selain itu, algoritma ini dikenal efektif dalam menangani hubungan nonlinier, data heterogen, serta interaksi kompleks antarfitur yang umum dijumpai pada data seismik. Penelitian ini juga memanfaatkan analisis *feature importance* untuk mengidentifikasi parameter-parameter seismik yang memberikan kontribusi terbesar terhadap prediksi magnitudo sehingga model yang dihasilkan tidak hanya memiliki tingkat akurasi yang baik, tetapi juga mudah diinterpretasikan.

Kontribusi utama penelitian ini terletak pada pengembangan model prediksi magnitudo gempa berbasis *Gradient Boosting Regression* menggunakan dataset gempa Indonesia dengan karakteristik lokal, sekaligus melakukan analisis *feature importance* untuk mengidentifikasi faktor-faktor seismik yang paling berpengaruh terhadap hasil prediksi. Berbeda dengan penelitian terdahulu yang lebih berfokus pada evaluasi performa model, penelitian ini mengombinasikan analisis akurasi dan interpretabilitas model sehingga dapat memberikan informasi yang lebih komprehensif bagi pengembangan sistem mitigasi bencana berbasis *machine learning*. Hasil penelitian diharapkan dapat menjadi referensi bagi pengembangan sistem pendukung keputusan, sistem peringatan dini, maupun penelitian lanjutan di bidang seismologi komputasional yang memanfaatkan teknologi kecerdasan buatan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk mengembangkan model prediksi magnitudo gempa berbasis *machine learning*. Data penelitian diperoleh dari United States Geological Survey (USGS) yang berisi data gempa bumi Indonesia periode 2016–2023. Dataset terdiri atas 17.331 data kejadian gempa dengan beberapa parameter seismik, antara lain *latitude*, *longitude*, *depth*, *nst*, *gap*, *dmin*, *rms*, *horizontalError*, *depthError*, dan *magError* sebagai variabel masukan, sedangkan nilai magnitudo digunakan sebagai variabel target.

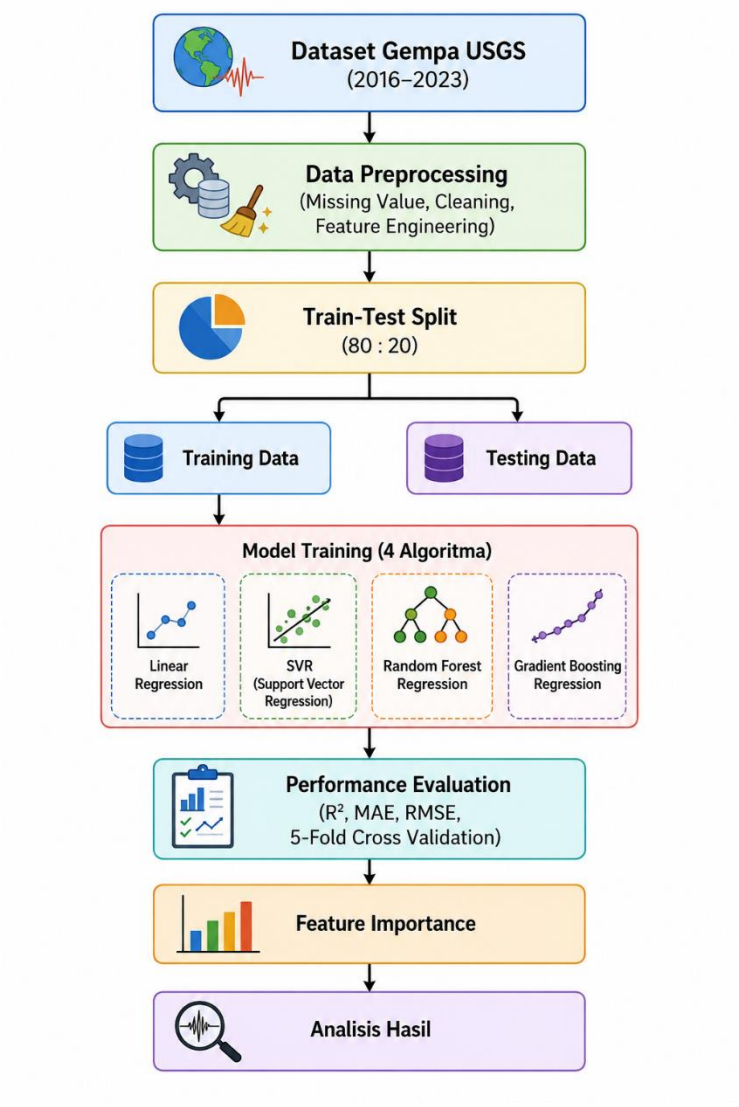
Tahapan penelitian diawali dengan proses pengumpulan data, dilanjutkan dengan *data preprocessing* untuk membersihkan data dari nilai kosong (*missing value*)[12], mengatasi data yang tidak konsisten, serta melakukan transformasi tipe data. Selanjutnya dilakukan *feature engineering* dengan menambahkan atribut temporal seperti tahun, bulan, hari, dan jam kejadian gempa. Tahapan berikutnya adalah pembagian dataset menggunakan rasio 80% data pelatihan dan 20% data pengujian untuk memperoleh model yang mampu melakukan generalisasi terhadap data baru.[13]

Model prediksi dibangun menggunakan algoritma Gradient Boosting Regression (GBR) dan dibandingkan dengan beberapa algoritma lain, yaitu Linear Regression, Support Vector Regression (SVR), dan Random Forest Regression. Seluruh model dievaluasi menggunakan nilai Coefficient of Determination (R^2), Mean Absolute Error (MAE), dan Root Mean Square Error (RMSE). Selain itu dilakukan 5-Fold Cross Validation untuk mengukur kestabilan model terhadap berbagai kombinasi data pelatihan dan pengujian. Setelah model terbaik diperoleh, dilakukan analisis Feature Importance untuk mengetahui parameter seismik yang paling berpengaruh terhadap prediksi magnitudo gempa.

Alur penelitian yang dilakukan dapat dilihat pada Gambar 1.

© 2026 The Author(s). Published by *CV. Graha Mitra Edukasi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026, Page: 12−22
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v5i1.226
https://journal.grahamitra.id/index.php/buai



Gambar 1. Tahapan Penelitian

Gambar 1 menunjukkan tahapan penelitian yang diawali dengan pengumpulan dataset gempa bumi Indonesia periode 2016–2023 yang diperoleh dari United States Geological Survey (USGS). Dataset selanjutnya melalui tahap *data preprocessing* yang meliputi pemeriksaan *missing value*, pembersihan data, dan *feature engineering*, kemudian dibagi menjadi data pelatihan dan data pengujian dengan rasio 80:20. Pada tahap pelatihan model, penelitian ini mengimplementasikan empat algoritma regresi, yaitu Linear Regression, Support Vector Regression (SVR), Random Forest Regression, dan Gradient Boosting Regression, untuk membandingkan kemampuan masing-masing dalam memprediksi magnitudo gempa. Seluruh model dievaluasi menggunakan metrik Coefficient of Determination (R^2), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), serta 5-Fold Cross Validation. Tahap akhir dilakukan melalui analisis Feature Importance untuk mengidentifikasi variabel-variabel yang memberikan kontribusi terbesar terhadap prediksi magnitudo, sehingga diperoleh model terbaik beserta interpretasi faktor-faktor yang memengaruhi hasil prediksi.

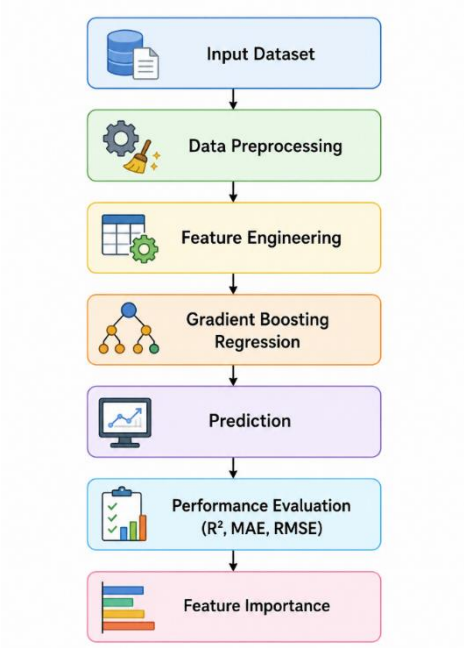
2.2 Implementasi Metode Gradient Boosting Regression

Metode Gradient Boosting Regression merupakan salah satu algoritma *ensemble learning* yang membangun model prediksi secara bertahap (*boosting*) menggunakan kumpulan *decision tree*. Setiap pohon keputusan yang dibentuk berfungsi memperbaiki kesalahan prediksi dari model sebelumnya sehingga menghasilkan model yang memiliki tingkat akurasi lebih tinggi dibandingkan metode regresi konvensional [[14]

Sebelum proses pelatihan model dilakukan, seluruh atribut numerik dinormalisasi menggunakan StandardScaler agar memiliki rentang nilai yang seragam. Selanjutnya model dilatih menggunakan parameter terbaik hasil proses *hyperparameter tuning*. Setelah proses pelatihan selesai, model diuji menggunakan data pengujian untuk memperoleh nilai R^2, MAE, dan RMSE sebagai indikator performa model. Tahapan implementasi metode ditunjukkan pada Gambar 2.

© 2026 The Author(s). Published by *CV. Graha Mitra Edukasi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026, Page: 12−22
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v5i1.226
https://journal.grahamitra.id/index.php/buai



Gambar 2. Implementasi Gradient Boosting Regression

Berdasarkan alur pada Gambar 2, hasil prediksi kemudian dibandingkan dengan nilai magnitudo aktual untuk mengetahui kemampuan model dalam melakukan estimasi terhadap data gempa bumi Indonesia. Karakteristik dataset yang digunakan pada penelitian ini ditunjukkan pada Tabel 1.

Tabel 1. Karakteristik Dataset Penelitian

Parameter	Keterangan
Sumber Data	United States Geological Survey (USGS)
Periode Data	2016–2023
Jumlah Data	17.331 kejadian gempa
Target Prediksi	Magnitude
Jumlah Data Latih	80%
Jumlah Data Uji	20%
Metode Utama	Gradient Boosting Regression
Metode Pembanding	Linear Regression, Support Vector Regression, Random Forest Regression
Evaluasi Model	R^2, MAE, RMSE, 5-Fold Cross Validation

Sebagaimana ditunjukkan pada Tabel 1, dataset yang digunakan memiliki jumlah data yang cukup besar sehingga mampu merepresentasikan karakteristik aktivitas seismik di Indonesia. Penggunaan beberapa metode pembanding bertujuan untuk mengevaluasi efektivitas algoritma Gradient Boosting Regression dalam meningkatkan akurasi prediksi dibandingkan algoritma regresi lainnya.

3. HASIL DAN PEMBAHASAN

Penelitian ini membandingkan performa empat algoritma regresi, yaitu Linear Regression, Support Vector Regression (SVR), Random Forest Regression, dan Gradient Boosting Regression, dalam memprediksi magnitudo gempa berdasarkan data seismik Indonesia periode 2016–2023. Tahapan penelitian meliputi eksplorasi dataset, *data preprocessing*, pelatihan model, evaluasi performa, hingga analisis Feature Importance untuk mengidentifikasi variabel yang memberikan kontribusi terbesar terhadap hasil prediksi. Evaluasi dilakukan menggunakan metrik Coefficient of Determination (R^2), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), serta 5-Fold Cross Validation. Hasil penelitian menunjukkan bahwa algoritma berbasis *ensemble learning* memberikan performa yang lebih baik dibandingkan metode regresi konvensional, dengan **Random Forest Regression** menghasilkan tingkat akurasi terbaik pada dataset yang digunakan. Pada bagian ini disajikan hasil eksplorasi data, proses pelatihan model, evaluasi performa, analisis **Feature Importance**, serta pembahasan yang dikaitkan dengan penelitian terdahulu.[15].

3.1 Hasil Penelitian

3.1.1 Eksplorasi Dataset

Tahap awal penelitian dilakukan dengan mengevaluasi karakteristik dataset yang diperoleh dari United States Geological Survey (USGS). Dataset terdiri atas 17.331 data gempa bumi yang terjadi di Indonesia selama periode

© 2026 The Author(s). Published by *CV. Graha Mitra Edukasi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

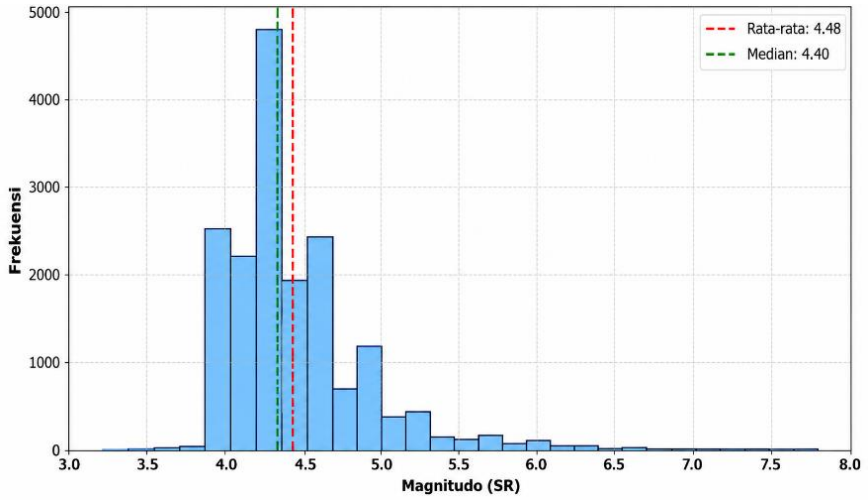
Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026, Page: 12−22
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v5i1.226
https://journal.grahamitra.id/index.php/buai

2016–2023. Setiap data memiliki informasi spasial, temporal, dan parameter kualitas pengukuran yang digunakan sebagai variabel prediktor dalam proses pembelajaran mesin. Karakteristik dataset ditunjukkan pada Tabel 2**.**

Tabel 2. Karakteristik Dataset Penelitian

Parameter	Nilai
Sumber Data	USGS
Periode	2016–2023
Jumlah Data	17.331
Jumlah Variabel	14
Target	Magnitude
Data Latih	80%
Data Uji	20%

Berdasarkan Tabel 2, jumlah data yang relatif besar memungkinkan model mempelajari hubungan antara parameter seismik dan magnitudo gempa secara lebih baik sehingga menghasilkan model prediksi yang lebih stabil. Selanjutnya dilakukan analisis distribusi nilai magnitudo untuk mengetahui karakteristik data yang digunakan.



Gambar 3. Distribusi Magnitudo Gempa

Berdasarkan Gambar 3, distribusi nilai magnitudo gempa menunjukkan bahwa sebagian besar kejadian gempa berada pada rentang 4,0 hingga 4,6 Skala Richter (SR). Histogram memperlihatkan puncak frekuensi berada di sekitar magnitudo 4,3–4,4 SR, yang menunjukkan bahwa gempa dengan kategori magnitudo sedang merupakan kejadian yang paling dominan pada dataset penelitian. Sementara itu, jumlah kejadian gempa dengan magnitudo di atas 5,5 SR relatif sedikit dan frekuensinya terus menurun seiring meningkatnya nilai magnitudo [6].

Garis vertikal berwarna merah menunjukkan nilai rata-rata (mean) sebesar 4,48 SR, sedangkan garis vertikal hijau menunjukkan nilai median sebesar 4,40 SR. Nilai rata-rata yang sedikit lebih besar dibandingkan median mengindikasikan bahwa distribusi data mengalami kemencengan ke kanan (positively skewed), yang disebabkan oleh keberadaan sejumlah kecil gempa dengan magnitudo tinggi hingga mendekati 8 SR. Meskipun jumlahnya relatif sedikit, kejadian gempa bermagnitudo besar tersebut menyebabkan nilai rata-rata bergeser ke arah kanan.

Distribusi data seperti ini merupakan karakteristik yang umum dijumpai pada data seismik, di mana gempa dengan magnitudo kecil hingga sedang terjadi jauh lebih sering dibandingkan gempa bermagnitudo besar. Kondisi tersebut juga menunjukkan bahwa dataset memiliki ketidakseimbangan distribusi (imbalanced distribution), sehingga model prediksi perlu mampu melakukan generalisasi dengan baik pada seluruh rentang magnitudo. Oleh karena itu, penelitian ini menggunakan algoritma Gradient Boosting Regression serta melakukan evaluasi menggunakan beberapa metrik, yaitu Coefficient of Determination (R^2), Mean Absolute Error (MAE), dan Root Mean Square Error (RMSE), agar performa model dapat dinilai secara lebih komprehensif.

Hasil distribusi pada Gambar 3 juga menunjukkan bahwa data yang digunakan telah mencakup variasi magnitudo yang cukup luas, mulai dari sekitar 3 SR hingga mendekati 8 SR, sehingga dapat memberikan representasi yang baik terhadap karakteristik aktivitas seismik di Indonesia. Variasi data tersebut menjadi dasar penting dalam proses pelatihan model agar mampu mengenali pola hubungan antara parameter seismik dengan besarnya magnitudo gempa secara lebih akurat.

3.1.2 Data Preprocessing

Tahap *data preprocessing* dilakukan untuk memastikan bahwa dataset yang digunakan memiliki kualitas yang baik sebelum memasuki proses pelatihan model. Tahapan ini meliputi pemeriksaan *missing value*, *feature engineering*,

© 2026 The Author(s). Published by *CV. Graha Mitra Edukasi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026, Page: 12−22
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v5i1.226
https://journal.grahamitra.id/index.php/buai

pemilihan variabel (*feature selection*), serta penghapusan data yang tidak lengkap agar model memperoleh data yang konsisten selama proses pembelajaran. Hasil pemeriksaan *missing value* pada dataset ditunjukkan pada Tabel 3.

Tabel 3. Hasil Pemeriksaan Missing Value

Variabel	Missing Value
Latitude	0
Longitude	0
Depth	0
Gap	0
RMS	1
MagError	65
DepthError	1

Berdasarkan Tabel 3, sebagian besar variabel tidak memiliki nilai kosong. Hanya tiga atribut yang mengandung *missing value*, yaitu rms sebanyak 1 data, depthError sebanyak 1 data, dan magError sebanyak 65 data. Dibandingkan dengan jumlah keseluruhan dataset sebanyak 17.331 data, jumlah nilai yang hilang tersebut relatif kecil, yaitu sekitar 0,39% dari total data. Kondisi ini menunjukkan bahwa kualitas dataset yang digunakan sudah cukup baik dan layak untuk proses pemodelan.

Setelah proses pemeriksaan kualitas data, dilakukan *feature engineering* dengan mengekstraksi informasi waktu kejadian gempa menjadi empat atribut baru, yaitu year, month, day, dan hour. Penambahan atribut tersebut bertujuan agar model dapat mempelajari kemungkinan adanya pola temporal pada aktivitas seismik. Selanjutnya dilakukan pemilihan fitur sehingga diperoleh sebelas variabel yang digunakan sebagai masukan model, yaitu latitude, longitude, depth, gap, rms, magError, depthError, year, month, day, dan hour.

Karena jumlah *missing value* sangat sedikit, penelitian ini menggunakan pendekatan penghapusan baris (*listwise deletion*) terhadap data yang tidak lengkap. Pendekatan ini dipilih agar seluruh data yang digunakan pada proses pelatihan memiliki informasi yang lengkap tanpa menimbulkan potensi bias akibat proses imputasi. Setelah proses pembersihan data dilakukan, jumlah dataset berkurang dari 17.331 menjadi 17.266 data, sehingga hanya 65 data atau sekitar 0,38% dari keseluruhan dataset yang dieliminasi. Pengurangan jumlah data tersebut tergolong sangat kecil sehingga tidak memengaruhi representativitas dataset terhadap aktivitas seismik di Indonesia.

Tahap berikutnya adalah pembagian dataset menggunakan metode train-test split dengan komposisi 80% data pelatihan dan 20% data pengujian. Berdasarkan hasil pembagian, diperoleh 13.812 data sebagai data pelatihan dan 3.454 data sebagai data pengujian. Data pelatihan digunakan untuk membangun model Gradient Boosting Regression, sedangkan data pengujian dimanfaatkan untuk mengevaluasi kemampuan model dalam melakukan prediksi terhadap data yang belum pernah dipelajari sebelumnya. Pembagian data ini bertujuan untuk mengurangi risiko *overfitting* dan menghasilkan model yang memiliki kemampuan generalisasi yang lebih baik.

Secara keseluruhan, tahapan *data preprocessing* berhasil menghasilkan dataset yang bersih, konsisten, dan siap digunakan pada proses pembelajaran mesin. Dataset hasil *preprocessing* diharapkan mampu meningkatkan stabilitas proses pelatihan serta menghasilkan model prediksi magnitudo gempa dengan tingkat akurasi yang lebih baik.

3.1.3 Hasil Pelatihan Model

Pada penelitian ini dilakukan perbandingan performa empat algoritma regresi, yaitu Linear Regression, Support Vector Regression (SVR), Random Forest Regression, dan Gradient Boosting Regression, untuk menentukan model yang paling efektif dalam memprediksi magnitudo gempa berdasarkan data seismik. Perbandingan performa masing-masing model diperlihatkan pada Tabel 4**.**

Tabel 4. Perbandingan Performa Model

Model	R^2	MAE	RMSE
Linear Regression	0,471	0,192	0,269
Support Vector Regression (SVR)	0,718	0,144	0,197
Random Forest Regression	0,767	0,123	0,179
Gradient Boosting Regression	0,746	0,137	0,187

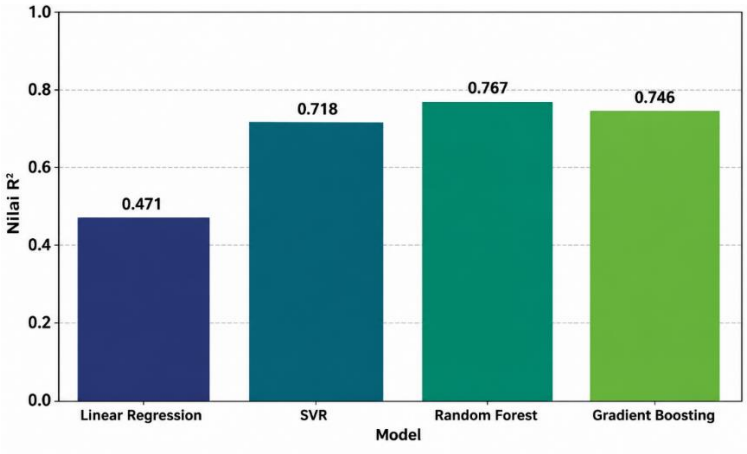
Berdasarkan Tabel 4, algoritma Random Forest Regression menghasilkan performa terbaik dengan nilai R^2 sebesar 0,767, MAE sebesar 0,123, dan RMSE sebesar 0,179. Nilai tersebut menunjukkan bahwa model mampu menjelaskan sekitar 76,7% variasi data magnitudo dengan tingkat kesalahan prediksi paling rendah dibandingkan model lainnya. Algoritma Gradient Boosting Regression menempati posisi kedua dengan nilai R^2 sebesar 0,746, MAE sebesar 0,137, dan RMSE sebesar 0,187, yang juga menunjukkan kemampuan prediksi yang baik pada data seismik Indonesia.

Sementara itu, Support Vector Regression (SVR) memperoleh nilai R^2 sebesar 0,718, dengan MAE sebesar 0,144 dan RMSE sebesar 0,197, sehingga masih mampu menghasilkan prediksi yang cukup akurat meskipun berada di bawah kedua metode berbasis *ensemble learning*. Adapun Linear Regression memberikan performa paling rendah dengan nilai R^2 sebesar 0,471, MAE sebesar 0,192, dan RMSE sebesar 0,269. Hasil ini menunjukkan bahwa hubungan antara parameter seismik dan magnitudo gempa bersifat nonlinier sehingga model regresi linier kurang mampu

© 2026 The Author(s). Published by *CV. Graha Mitra Edukasi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026, Page: 12−22
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v5i1.226
https://journal.grahamitra.id/index.php/buai

merepresentasikan pola data secara optimal. Secara keseluruhan, hasil pengujian menunjukkan bahwa algoritma berbasis *ensemble learning*, khususnya Random Forest Regression, lebih efektif dalam memprediksi magnitudo gempa dibandingkan metode regresi konvensional maupun *Support Vector Regression*. Untuk memperjelas perbedaan performa antaralgoritma, hasil evaluasi juga divisualisasikan pada Gambar 4**.**

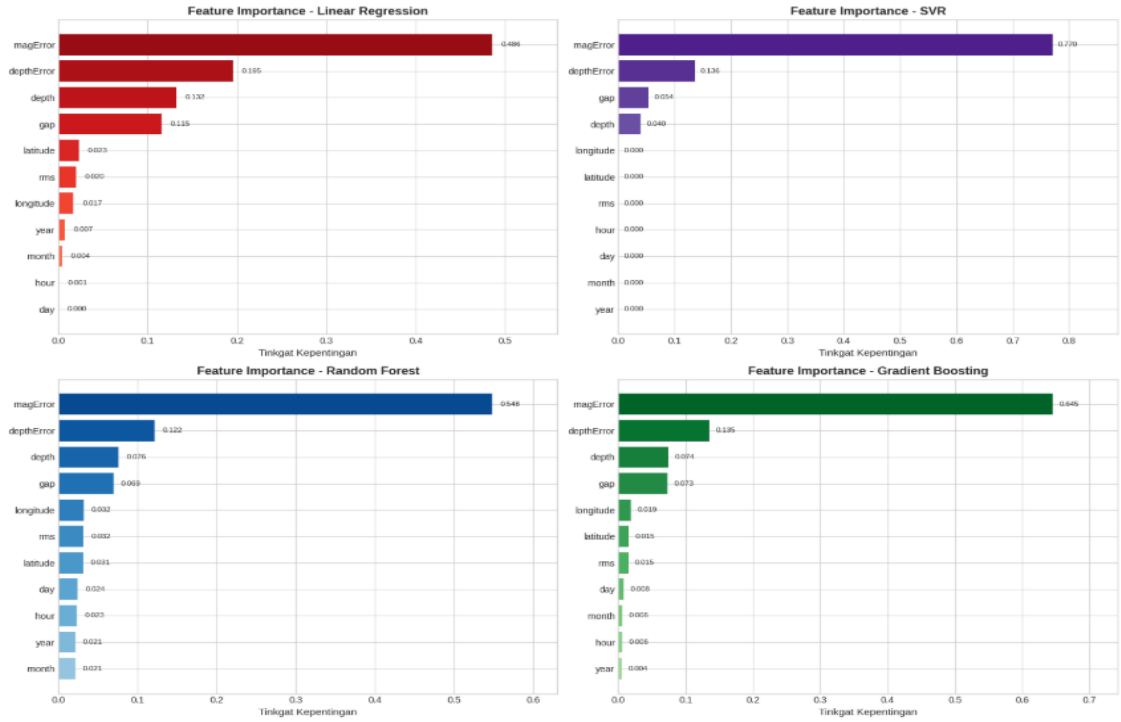


Gambar 4. Perbandingan Nilai R² Seluruh Model

Visualisasi pada Gambar 4 memperlihatkan bahwa algoritma berbasis *ensemble learning* memberikan performa yang lebih baik dibandingkan metode regresi linier.

3.1.4 Analisis Feature Importance

Setelah dilakukan evaluasi terhadap seluruh model regresi, analisis Feature Importance dilakukan pada model yang memiliki performa terbaik untuk mengidentifikasi variabel-variabel yang memberikan kontribusi terbesar dalam proses prediksi magnitudo gempa. Analisis ini bertujuan untuk memberikan interpretasi terhadap model sehingga faktor-faktor yang paling berpengaruh terhadap hasil prediksi dapat diketahui secara lebih jelas..



Gambar 5. Feature Importance

Berdasarkan Gambar 5, seluruh algoritma menunjukkan pola yang relatif konsisten dalam mengidentifikasi variabel yang paling berpengaruh terhadap prediksi magnitudo gempa. Variabel magError menjadi fitur dengan tingkat kepentingan (*feature importance*) tertinggi pada seluruh model, yaitu Linear Regression (0,498), Support Vector Regression (0,770), Random Forest Regression (0,584), dan Gradient Boosting Regression (0,646). Hasil tersebut menunjukkan bahwa tingkat kesalahan pengukuran magnitudo memiliki hubungan yang sangat kuat terhadap proses prediksi yang dilakukan oleh seluruh algoritma.

© 2026 The Author(s). Published by *CV. Graha Mitra Edukasi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026, Page: 12−22
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v5i1.226
https://journal.grahamitra.id/index.php/buai

Selain magError, variabel depthError, depth, dan gap juga secara konsisten menempati posisi empat besar pada seluruh model. Temuan ini menunjukkan bahwa parameter yang berkaitan dengan kualitas pengukuran dan karakteristik kedalaman gempa memiliki kontribusi yang lebih besar dibandingkan parameter spasial maupun temporal. Sebaliknya, variabel year, month, day, dan hour memiliki nilai *feature importance* yang relatif kecil, sehingga pengaruh faktor waktu terhadap prediksi magnitudo pada dataset ini tidak terlalu signifikan.

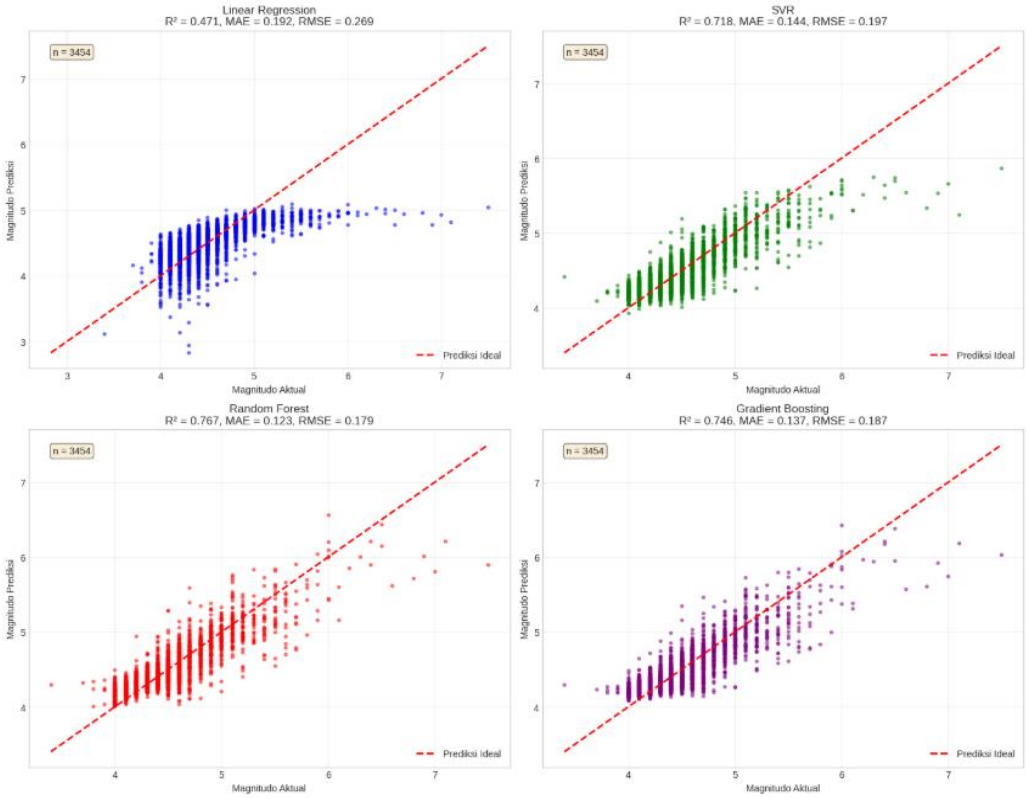
Apabila dibandingkan antaralgoritma, Support Vector Regression memberikan tingkat dominasi yang paling tinggi pada variabel magError, sedangkan algoritma Random Forest Regression dan Gradient Boosting Regression menghasilkan distribusi kepentingan fitur yang lebih seimbang. Kondisi ini menunjukkan bahwa model berbasis *ensemble learning* tidak hanya bergantung pada satu variabel utama, tetapi juga mampu memanfaatkan informasi dari beberapa parameter seismik secara bersamaan untuk meningkatkan akurasi prediksi.

Secara keseluruhan, hasil analisis *feature importance* menunjukkan bahwa parameter yang berkaitan dengan kualitas pengukuran seismik memiliki kontribusi yang lebih besar dibandingkan parameter lokasi maupun waktu kejadian gempa. Temuan ini mengindikasikan bahwa peningkatan kualitas pencatatan data oleh jaringan stasiun seismik berpotensi memberikan dampak langsung terhadap peningkatan performa model prediksi magnitudo gempa. Selain itu, konsistensi pola kepentingan fitur pada keempat algoritma menunjukkan bahwa hubungan antara variabel-variabel utama dengan magnitudo gempa bersifat stabil sehingga dapat dijadikan dasar dalam pengembangan model prediksi maupun sistem mitigasi bencana berbasis *machine learning*.

3.2 Implementasi

Implementasi model dilakukan menggunakan Kaggle Notebook sebagai lingkungan komputasi berbasis *cloud* yang mendukung pengolahan data dan eksperimen *machine learning*.[16] Seluruh proses pengembangan model dilakukan menggunakan bahasa pemrograman Python dengan bantuan pustaka Pandas untuk manipulasi data, NumPy untuk komputasi numerik, Scikit-learn untuk pembangunan dan evaluasi model, serta Matplotlib untuk visualisasi hasil. Penggunaan Kaggle memungkinkan seluruh proses eksperimen dilakukan secara terintegrasi mulai dari proses pemuatan dataset, *preprocessing*, pelatihan model, hingga evaluasi performa tanpa memerlukan konfigurasi perangkat keras tambahan. Dataset yang telah melalui proses *preprocessing* digunakan untuk melatih empat algoritma regresi, yaitu Linear Regression, Support Vector Regression (SVR), Random Forest Regression, dan Gradient Boosting Regression. Selanjutnya masing-masing model dievaluasi menggunakan metrik R^2, MAE, dan RMSE untuk menentukan model dengan performa terbaik [17].

Hasil implementasi model pada Kaggle divisualisasikan melalui grafik Actual vs Predicted Magnitude sebagaimana ditunjukkan pada Gambar 6.



Gambar 6. Perbandingan Actual dan Predicted Magnitude pada Empat Model Regresi

Berdasarkan Gambar 6, terlihat adanya perbedaan kemampuan masing-masing algoritma dalam memprediksi magnitudo gempa. Linear Regression menghasilkan penyebaran titik yang paling jauh dari garis diagonal (*prediksi*

© 2026 The Author(s). Published by *CV. Graha Mitra Edukasi*. Page 19
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026, Page: 12−22
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v5i1.226
https://journal.grahamitra.id/index.php/buai

ideal), menunjukkan bahwa model linier belum mampu menangkap hubungan kompleks antarparameter seismik. Hal ini juga tercermin dari nilai R^2 sebesar 0,471, yang merupakan nilai terendah dibandingkan model lainnya.

Model Support Vector Regression (SVR) menunjukkan peningkatan performa dengan sebagian besar titik prediksi mulai mengikuti garis diagonal. Nilai R^2 sebesar 0,718 mengindikasikan bahwa model mampu mempelajari hubungan nonlinier dengan lebih baik dibandingkan Linear Regression, meskipun masih terdapat penyimpangan pada data dengan magnitudo tinggi.

Sementara itu, Random Forest Regression menghasilkan distribusi titik yang paling mendekati garis prediksi ideal. Nilai R^2 sebesar 0,767, MAE sebesar 0,123, dan RMSE sebesar 0,179 menunjukkan bahwa model ini memiliki tingkat akurasi terbaik di antara seluruh algoritma yang diuji. Sebaran titik yang relatif rapat di sekitar garis diagonal menunjukkan kemampuan model dalam melakukan generalisasi terhadap data uji.

Model Gradient Boosting Regression juga memperlihatkan performa yang baik dengan nilai R^2 sebesar 0,746. Sebagian besar hasil prediksi berada di sekitar garis diagonal meskipun masih terdapat penyimpangan pada beberapa data dengan magnitudo besar. Kondisi tersebut menunjukkan bahwa algoritma *boosting* mampu mempelajari pola hubungan antarparameter seismik secara efektif, namun performanya pada dataset penelitian ini masih sedikit berada di bawah Random Forest Regression.

Secara keseluruhan, visualisasi pada Gambar 6 memperkuat hasil evaluasi kuantitatif pada Tabel 4, yaitu bahwa algoritma berbasis ensemble learning menghasilkan prediksi yang lebih akurat dibandingkan metode regresi konvensional. Hal ini ditunjukkan oleh semakin rapatnya distribusi titik terhadap garis prediksi ideal serta semakin kecilnya nilai kesalahan prediksi yang dihasilkan.

3.3 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma Random Forest Regression memberikan performa terbaik dalam memprediksi magnitudo gempa dibandingkan Linear Regression**,** Support Vector Regression (SVR), dan Gradient Boosting Regression. Berdasarkan hasil evaluasi, Random Forest Regression memperoleh nilai R^2 sebesar 0,767**,** MAE sebesar 0,123**,** dan RMSE sebesar 0,179, yang menunjukkan kemampuan model dalam menjelaskan variasi data magnitudo dengan tingkat kesalahan prediksi paling rendah. Sebaliknya, Linear Regression menghasilkan performa terendah dengan nilai R^2 sebesar 0,471**,** mengindikasikan bahwa pendekatan linier belum mampu merepresentasikan hubungan kompleks antarparameter seismik.

Keunggulan Random Forest Regression dipengaruhi oleh mekanisme *ensemble learning* berbasis *bagging* yang membangun sejumlah *decision tree* dari sampel data yang berbeda, kemudian menggabungkan hasil prediksi seluruh pohon melalui proses agregasi. Pendekatan tersebut mampu mengurangi variansi model, meningkatkan kemampuan generalisasi, serta lebih tahan terhadap *noise* dan *outlier* yang umum dijumpai pada data seismik. Kondisi ini menjadikan Random Forest lebih efektif dalam memodelkan hubungan nonlinier antara parameter seismik dan magnitudo gempa dibandingkan metode regresi konvensional [18].

Visualisasi pada **Gambar 6** memperkuat hasil evaluasi kuantitatif tersebut. Sebaran titik prediksi Random Forest terlihat paling mendekati garis diagonal (*prediksi ideal*), menunjukkan bahwa nilai prediksi memiliki kedekatan yang tinggi terhadap nilai magnitudo aktual. Sementara itu, Support Vector Regression dan Gradient Boosting Regression juga menunjukkan kemampuan prediksi yang baik, tetapi masih menghasilkan penyimpangan pada beberapa data dengan magnitudo tinggi. Adapun Linear Regression memperlihatkan penyebaran titik yang lebih jauh dari garis diagonal sehingga menghasilkan tingkat kesalahan prediksi yang lebih besar.

Analisis Feature Importance pada Gambar 5 menunjukkan bahwa variabel magError**,** depthError**,** depth**,** dan gap merupakan faktor yang paling berpengaruh terhadap prediksi magnitudo pada seluruh model. Konsistensi urutan variabel penting tersebut menunjukkan bahwa kualitas pengukuran seismik dan karakteristik kedalaman gempa memiliki kontribusi yang lebih besar dibandingkan atribut spasial maupun temporal. Sebaliknya, variabel **year**, **month**, **day**, dan **hour** memiliki tingkat kepentingan yang relatif rendah, sehingga faktor waktu belum memberikan pengaruh yang signifikan terhadap model prediksi pada dataset yang digunakan.

Hasil penelitian ini sejalan dengan penelitian Fikri *et al.* [19] yang melaporkan bahwa algoritma berbasis *ensemble learning* menghasilkan akurasi yang lebih baik dibandingkan metode regresi konvensional dalam prediksi magnitudo gempa. Penelitian Ariska *et al.* [20]juga menunjukkan bahwa model berbasis pohon keputusan mampu memanfaatkan hubungan nonlinier antarfitur secara lebih efektif sehingga meningkatkan kemampuan prediksi. Selain itu, penelitian Umar *et al[15].* menjelaskan bahwa analisis *feature importance* memberikan interpretasi yang lebih baik terhadap kontribusi masing-masing parameter geofisika dalam proses prediksi, sehingga hasil model tidak hanya akurat tetapi juga lebih mudah dipahami.

Kontribusi penelitian ini terletak pada penerapan dan evaluasi empat algoritma regresi menggunakan dataset gempa Indonesia periode 2016–2023 serta analisis Feature Importance untuk mengidentifikasi parameter seismik yang paling berpengaruh terhadap prediksi magnitudo. Berbeda dengan beberapa penelitian sebelumnya yang lebih berfokus pada peningkatan akurasi model, penelitian ini juga memberikan interpretasi terhadap faktor-faktor yang memengaruhi hasil prediksi sehingga dapat menjadi dasar dalam pengembangan sistem pendukung keputusan dan mitigasi bencana berbasis *machine learning*. Temuan bahwa Random Forest Regression memberikan performa terbaik menunjukkan bahwa pendekatan berbasis *ensemble learning* layak dipertimbangkan dalam pengembangan sistem prediksi magnitudo gempa pada data seismik Indonesia.

© 2026 The Author(s). Published by *CV. Graha Mitra Edukasi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026, Page: 12−22
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v5i1.226
https://journal.grahamitra.id/index.php/buai

4. KESIMPULAN

Penelitian ini berhasil membandingkan performa empat algoritma regresi dalam memprediksi magnitudo gempa berdasarkan data seismik Indonesia periode 2016–2023. Hasil evaluasi menunjukkan bahwa Random Forest Regression memberikan performa terbaik dengan nilai R^2 sebesar 0,767**,** MAE sebesar 0,123, dan RMSE sebesar 0,179**,** sehingga mampu menghasilkan prediksi yang lebih akurat dibandingkan Linear Regression**,** Support Vector Regression, dan Gradient Boosting Regression**.** Temuan ini menunjukkan bahwa pendekatan *ensemble learning* berbasis *bagging* lebih efektif dalam memodelkan hubungan nonlinier antarparameter seismik dibandingkan metode regresi konvensional. Analisis *feature importance* menunjukkan bahwa magError**,** depthError**,** depth, dan gap merupakan variabel yang memberikan kontribusi terbesar terhadap proses prediksi. Temuan tersebut mengindikasikan bahwa kualitas pengukuran data seismik memiliki pengaruh yang signifikan terhadap keberhasilan model. Penelitian ini juga memberikan kontribusi berupa evaluasi komprehensif terhadap beberapa algoritma regresi sekaligus interpretasi faktor-faktor yang memengaruhi prediksi magnitudo menggunakan analisis *feature importance*. Meskipun demikian, penelitian masih terbatas pada penggunaan dataset USGS periode 2016–2023 dan belum mengintegrasikan parameter geologi maupun model *deep learning*. Oleh karena itu, penelitian selanjutnya dapat memanfaatkan dataset yang lebih luas, menambahkan parameter geofisika lain, serta mengevaluasi algoritma berbasis *deep learning* untuk meningkatkan akurasi prediksi magnitudo gempa.

REFERENCES

[1] A. Prasetio, M. Makmun, and M. N. Dwi, "Analisis Gempa Bumi Di Indonesia Dengan Metode Clustering," *Bulletin of Information Technology (BIT)*, vol. 4, no. 2, pp. 338–343, 2023, doi: 10.47065/bit.v3i1.
[2] W. Ustyannie, E. Setyaningsih, and C. Iswahyudi, "Optimization of software defects prediction in imbalanced class using a combination of resampling methods with support vector machine and logistic regression," *Jurnal Infotel*, vol. 13, no. 4, pp. 176–184, 2021, doi: 10.20895/infotel.v13i4.726.
[3] S. T. Hamidou and A. Mehdi, "Enhancing IDS performance through a comparative analysis of Random Forest, XGBoost, and Deep Neural Networks," *Machine Learning with Applications*, vol. 22, p. 100738, Dec. 2025, doi: 10.1016/j.mlwa.2025.100738.
[4] D. T. Murdiansyah, "Prediksi Stroke Menggunakan Extreme Gradient Boosting," *JIKO (Jurnal Informatika dan Komputer)*, vol. 8, no. 2, p. 419, Sep. 2024, doi: 10.26798/jiko.v8i2.1295.
[5] K. Alkanan, Y. Zhong, X. Luo, and R. Dong, "Advanced multimodal biometric authentication for IoT-enabled smart home security: Integrating deep learning-based face recognition with behavioral patterns," *Kuwait Journal of Science*, vol. 53, no. 2, Apr. 2026, doi: 10.1016/j.kjs.2026.100548.
[6] A. Syaifuddin and T. Prabowo, "Optimasi Akurasi Model Prediksi Magnitudo Gempa Bumi dengan Integrasi Clustering DBSCAN pada Ensemble Learning (Random Forest & XGBoost)," *TIN: Terapan Informatika Nusantara*, vol. 5, no. 7, pp. 422–429, Dec. 2024, doi: 10.47065/tin.v5i7.6522.
[7] A. Belenguer, J. A. Pascual, and J. Navaridas, "GöwFed: A novel federated network intrusion detection system," *Journal of Network and Computer Applications*, vol. 217, Aug. 2023, doi: 10.1016/j.jnca.2023.103653.
[8] S. S. Shafin, "An explainable feature selection framework for web phishing detection with machine learning," *Data Science and Management*, vol. 8, no. 2, pp. 127–136, Jun. 2025, doi: 10.1016/j.dsm.2024.08.004.
[9] A. K. Rai *et al.*, "Gold prospectivity mapping in the eastern part of Mahakoshal Fold Belt, India: A comparative study of Random Forest and XGBoost leveraging knowledge-guided feature engineering," *Geosystems and Geoenvironment*, vol. 5, no. 3, Aug. 2026, doi: 10.1016/j.geogeo.2026.100532.
[10] M. Lin, D. Zhang, Q. He, J. Zhao, and Y. Yang, "VQ-Transformer: Lightweight IoT intrusion detection enhanced by optimal transport knowledge distillation," *Ain Shams Engineering Journal*, vol. 17, no. 4, Apr. 2026, doi: 10.1016/j.asej.2026.104105.
[11] A. Heriyansyah, H. A. Rohayani, and M. Jambi, "Analisis Prediksi Gempa Bumi Menggunakan Algoritma Decision Tree C4.5: Studi Literatur".
[12] R. C. Pereira, P. H. Abreu, P. P. Rodrigues, and M. A. T. Figueiredo, "Imputation of data Missing Not at Random: Artificial generation and benchmark analysis," *Expert Syst. Appl.*, vol. 249, Sep. 2024, doi: 10.1016/j.eswa.2024.123654.
[13] Y. Guo and T. E. Holy, "Recovering Missing Features in Nonnegative Matrix Factorization via Generalized Singular Value Decomposition," *iScience*, p. 114708, Feb. 2026, doi: 10.1016/j.isci.2026.114708.
[14] Abdullah-All-Tanvir, I. Ali Khandokar, A. K. M. Muzahidul Islam, S. Islam, and S. Shatabda, "A gradient boosting classifier for purchase intention prediction of online shoppers," *Heliyon*, vol. 9, no. 4, Apr. 2023, doi: 10.1016/j.heliyon.2023.e15163.
[15] M. A. Umar, Z. Chen, K. Shuaib, and Y. Liu, "Effects of feature selection and normalization on network intrusion detection," *Data Science and Management*, vol. 8, no. 1, pp. 23–39, Mar. 2025, doi: 10.1016/j.dsm.2024.08.001.
[16] M. T. Rahman, S. D. Dipto, I. J. June, A. Momin, and M. R. Al Mamun, "Machine Learning-based Disease Classification in Tomato (Solanum lycopersicum) Plants," *Jurnal Keteknikan Pertanian Tropis dan Biosistem*, vol. 12, no. 3, pp. 151–160, Dec. 2024, doi: 10.21776/ub.jkptb.2024.012.03.01.
[17] R. R. Oprasto, J. Wang, A. F. O. Pasaribu, S. Setiawansyah, R. Aryanti, and Sumanto, "An Entropy-Assisted COBRA Framework to Support Complex Bounded Rationality in Employee Recruitment," *Bulletin of Computer Science Research*, vol. 5, no. 3, pp. 207–218, Apr. 2025, doi: 10.47065/bulletincsr.v5i3.505.
[18] A. Kumar, S. Bhushan, R. Pokhrel, A. I. Al-Omari, A. R. A. Alanzi, and S. S. Alshqaq, "Imputation of missing data for domain mean estimation using simple random sampling," *Kuwait Journal of Science*, vol. 52, no. 4, pp. 1–10, Oct. 2025, doi: 10.1016/j.kjs.2025.100461.
[19] M. Fikri, R. Herteno, R. A. Nugroho, S. W. Saputro, and F. Abadi, "Multi-Criteria Decision Making Dalam Seleksi Fitur Ensemble Untuk Prediksi Cacat Perangkat Lunak Multi-Criteria Decision Making In Ensemble Feature Selection," vol. 12, no. 6, pp. 1385–1394, 2025.

© 2026 The Author(s). Published by *CV. Graha Mitra Edukasi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 5, No 1, April 2026, Page: 12−22
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v5i1.226
https://journal.grahamitra.id/index.php/buai

[20] F. Ariska, V. Sihombing, and I. Irmayani, "Student Graduation Predictions Using Comparison of C5.0 Algorithm With Linear Regression," *SinkrOn*, vol. 7, no. 1, pp. 256–266, Feb. 2022, doi: 10.33395/sinkron.v7i1.11261.

© 2026 The Author(s). Published by *CV. Graha Mitra Edukasi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.