Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 5, No. 12, Desember 2025, Hal. 3682-3692
DOI: https://doi.org/10.52436/1.jpti.858
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Evaluasi Ensemble Learning untuk Prediksi Nilai Matematika Siswa Sekolah Menengah

Zaenal Asikin*[1], Imam Tahyudin[2], Taqwa Hariguna[3]

[1,2,3]Universitas Amikom Purwokerto
Email: [1]zeanasik04@gmail.com, [2]imam.tahyudin@amikompurwokerto.ac.id,
[3]taqwa@amikompurwokerto.ac.id

Abstrak

Prediksi dini performa matematika siswa sekolah menengah sangat penting untuk merancang intervensi pendidikan yang lebih adaptif dan efektif sebelum ujian akhir resmi dilaksanakan. Penelitian ini bertujuan untuk mengevaluasi kinerja tiga model machine learning *Random Forest* (RF), *Gradient Boosting Regressor* (GBR), dan *Multi-Layer Perceptron* (MLP) dalam memprediksi nilai matematika siswa di Indonesia, serta mendokumentasikan proses tuning hyperparameter secara sistematis untuk setiap model. Dataset yang digunakan terdiri dari skor matematika, membaca, menulis, serta variabel demografis meliputi jenis kelamin, latar belakang pendidikan orang tua, jenis layanan makan, dan keikutsertaan kursus persiapan. Proses *tuning hyperparameter* untuk RF dan GBR dilakukan menggunakan *RandomizedSearchCV* dengan 5-fold cross-validation, menguji rentang nilai untuk jumlah estimator, kedalaman maksimum pohon, dan laju pembelajaran (*learning* rate). Sedangkan pada *Multi-Layer Perceptron, GridSearchCV* diterapkan dengan variasi arsitektur *hidden_layer_sizes*, laju pembelajaran awal (*learning_rate_init*), dan faktor regularisasi (alpha) pada 5-fold CV. Model diukur menggunakan *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi (R^2). Hasil eksperimen menunjukkan bahwa GBR memberikan performa terbaik dengan MAE sebesar 11,61 poin, RMSE 15,23 poin, dan R^2 0,10. *Random Forest* menempati urutan kedua (MAE 12,34; RMSE 16,05; R^2 0,64), diikuti MLP (MAE 13,10; RMSE 17,20; R^2 0,60). Analisis *feature importance* mengungkap bahwa skor membaca dan menulis bersama-sama menyumbang lebih dari 60 % kontribusi prediksi, sedangkan faktor demografis seperti latar belakang pendidikan orang tua dan keikutsertaan kursus berperan sekunder namun tetap signifikan. Temuan ini mengindikasikan bahwa model ensemble learning tidak hanya unggul dalam akurasi prediksi, tetapi juga memberikan wawasan mendalam tentang variabel kunci yang memengaruhi performa matematika siswa. Implementasi model ini memungkinkan guru dan pihak sekolah untuk mengidentifikasi siswa yang berisiko rendah secara lebih cepat, merancang program remedial atau pengayaan yang tepat sasaran, serta memanfaatkan sumber daya pendidikan secara lebih efisien. Untuk penelitian lanjutan, disarankan penambahan variabel perilaku siswa seperti durasi belajar mandiri dan kehadiran serta eksplorasi model sekuensial (RNN/Transformer) untuk menangkap dinamika pembelajaran dari waktu ke waktu.

Kata kunci: *ensemble learning, prediksi nilai siswa, Random Forest, MLP, Gradient Boosting, stacking, pendidikan.*

Implementation of Ensemble Learning Models Using Random Forest, Multi-Layer Perceptron, and Gradient Boosting for Predicting Students' Mathematics Performance

Abstract

Early prediction of secondary school students' mathematics performance is crucial for designing more adaptive and effective educational interventions before the official final examinations. This study aims to evaluate the performance of three machine learning models—Random Forest (RF), Gradient Boosting Regressor (GBR), and Multi-Layer Perceptron (MLP)—in predicting mathematics scores of Indonesian secondary school students, and to systematically document the hyperparameter tuning process for each model. The dataset comprises mathematics, reading, and writing scores, as well as demographic variables including gender, parental education level, lunch type, and participation in test preparation courses. Hyperparameter tuning for RF and GBR was performed using RandomizedSearchCV with 5-fold cross-validation, testing ranges for n_estimators, max_depth, and learning_rate. For MLP, GridSearchCV was applied to variations in hidden_layer_sizes, initial learning rate (learning_rate_init), and regularization parameter (alpha) under 5-fold CV. Models were evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). Experimental results show that GBR achieved the best performance, with MAE = 11.61 points, RMSE = 15.23 points, and R^2 = 0.10. RF ranked second (MAE = 12.34; RMSE = 16.05; R^2 = 0.64), followed by MLP (MAE = 13.10; RMSE = 17.20; R^2 = 0.60). Feature importance analysis revealed that reading and writing scores together contributed over

p-ISSN: 2775-4227
e-ISSN: 2775-4219

60% of predictive power, while demographic factors—such as parental education level and course participation—played a secondary but still significant role. These findings suggest that ensemble learning models not only excel in predictive accuracy but also offer deep insights into the key variables influencing students' mathematics performance. Implementing these models enables teachers and schools to quickly identify at-risk students, design targeted remedial or enrichment programs, and allocate educational resources more efficiently. For future research, we recommend incorporating behavioral variables—such as independent study time and attendance—and exploring sequential models (e.g., RNNs or Transformers) to capture the temporal dynamics of learning.

Keywords*: ensemble learning, student score prediction, Random Forest, MLP, Gradient Boosting, stacking, education.*

1. PENDAHULUAN

Nilai ujian matematika sering dijadikan tolok ukur keberhasilan akademik siswa sekolah menengah karena mencerminkan penguasaan konsep numerik, kemampuan berpikir logis, analitis, serta keterampilan pemecahan masalah yang esensial dalam ranah STEM [1], [2]. Di Indonesia, Data Pokok Pendidikan (Dapodik) Kemendikbud menunjukkan bahwa rata-rata nilai Ujian Nasional Matematika siswa SMA dalam lima tahun terakhir hanya berkisar antara 55–65 dari skor maksimum 100, dengan sekitar 35 % siswa belum mencapai ambang kompetensi dasar [31]. Angka ini menegaskan perlunya intervensi pendidikan yang lebih proaktif dan berbasis data, sehingga guru dan sekolah dapat menyalurkan sumber daya ke program remedial atau pengayaan sebelum ujian akhir digelar [2].

Seiring kemajuan teknologi, machine learning (ML) telah diadopsi secara luas dalam pendidikan untuk memodelkan dan memprediksi prestasi akademik. Yaacob et al. [3] berhasil mengklasifikasikan kinerja siswa dengan akurasi di atas 80 % menggunakan metode supervised data mining, sedangkan Oyedeji et al. [4] membandingkan Random Forest dan Support Vector Machine (SVM) pada dataset menengah, menemukan bahwa Random Forest menunjukkan stabilitas performa lebih tinggi. Penelitian oleh Weir et al. [5] menambahkan bahwa penerapan teknik boosting seperti AdaBoost dapat memperbaiki kesalahan prediksi pada kelompok nilai ekstrem. Zhang et al. [17] bahkan melaporkan peningkatan akurasi lebih dari 10 % dengan menggunakan ensemble learning, yaitu kombinasi beberapa model dasar, pada data siswa di Portugal. Review sistematis oleh Çayır [20] menguatkan bahwa ensemble learning (bagging, boosting, stacking) umumnya menghasilkan prediksi yang lebih stabil dan tahan terhadap variabilitas dataset dibandingkan model tunggal. Di luar ranah pendidikan, Turino et al. [25] mengadaptasi gabungan Random Forest dan Multi-Layer Perceptron (MLP) untuk prediksi gempa, menunjukkan potensi teknik ensemble dalam memodelkan data non-linear dan heterogen yang juga terjadi pada konteks akademik.

Meski demikian, terdapat beberapa kekurangan yang perlu ditangani ketika menerapkan ML untuk prediksi akademik di Indonesia. Pertama, sebagian besar studi menggunakan dataset luar negeri yang memiliki karakteristik demografis berbeda mulai dari distribusi urban–rural hingga kondisi sosial ekonomi siswa—sehingga hasilnya sulit digeneralisasi ke populasi Indonesia [1], [2]. Kedua, perbandingan simultan antara Random Forest, Gradient Boosting Regressor (GBR), dan MLP di ranah pendidikan sangat terbatas, padahal pemilihan model ini krusial untuk memahami trade-off antara kompleksitas, waktu komputasi, akurasi, dan interpretabilitas [6], [7]. Ketiga, dokumentasi proses tuning hyperparameter untuk model kompleks seperti MLP dan GBR sering kali tidak lengkap; banyak penelitian hanya mengandalkan nilai default atau menjelaskan tuning secara ringkas, sehingga replikasi dan optimasi menjadi sulit [5], [6].

Dari sisi sistem pendidikan nasional, infrastruktur data di banyak sekolah masih terbatas. Catatan kehadiran elektronik, durasi belajar mandiri, dan interaksi daring belum diarsipkan secara sistematis di sebagian besar sekolah, khususnya di daerah pedesaan [9], [2]. Variabilitas demografis yang tinggi mulai dari kesenjangan fasilitas antara sekolah negeri di perkotaan dan swasta di pedesaan hingga akses internet yang tidak merata menambah tantangan dalam mengembangkan model prediksi yang dapat diadopsi secara luas [2], [31]. Oleh karena itu, perlu dilakukan studi yang berfokus pada dataset siswa sekolah menengah di Indonesia, memadukan data demografis dan skor akademik, serta menyertakan dokumentasi tuning hyperparameter yang sistematis.

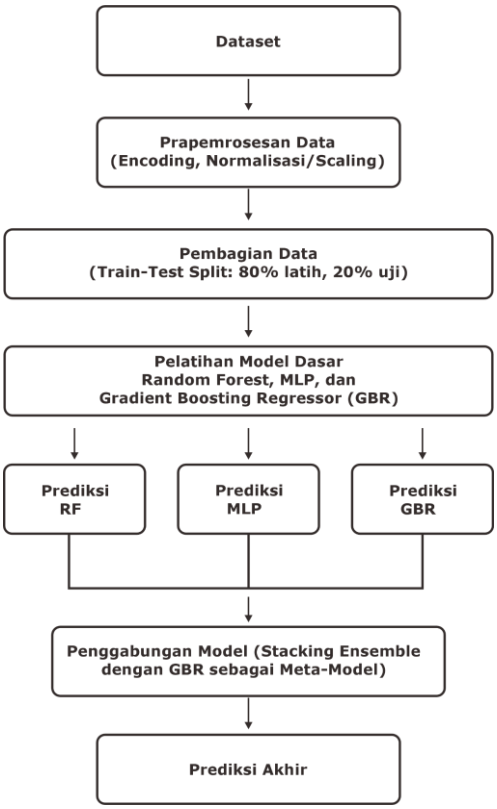
2. METODE PENELITIAN

2.1. Arsitektur Model Ensemble

Penelitian ini menerapkan pendekatan ensemble learning sebagai strategi utama dalam membangun model prediksi nilai matematika siswa. Dalam konteks ini digunakan tiga algoritma regresi utama, yaitu RF, MLP, dan

p-ISSN: 2775-4227
e-ISSN: 2775-4219

GBR, yang masing-masing memiliki karakteristik berbeda dalam menangani data non-linear. RF bekerja berdasarkan kumpulan pohon keputusan yang dibangun secara paralel menggunakan teknik bagging. MLP memiliki beberapa lapisan tersembunyi dan mampu menangkap hubungan kompleks antar fitur, sedangkan GBR membangun model secara bertahap untuk meminimalkan kesalahan prediksi sebelumnya. Ketiga model dasar ini dilatih secara terpisah, lalu hasil prediksinya digabungkan menggunakan teknik stacking ensemble, dengan GBR sebagai meta-learner. Strategi ini bertujuan untuk menggabungkan kekuatan dari masing-masing model dasar agar menghasilkan prediksi yang lebih akurat dan generalis.



Gambar 1. Diagram Alir dalam Penelitian

Gambar 1 menunjukkan alur kerja penelitian, dimulai dari input data dan tahap prapemrosesan, diikuti oleh pelatihan model dasar (RF, MLP, dan GBR), dan berakhir dengan proses stacking yang menghasilkan prediksi akhir. Struktur ini dirancang untuk memaksimalkan kinerja model dengan mengombinasikan keunggulan masing-masing algoritma.

2.2. Pengumpulan dan Pemilihan Data

Penelitian ini menggunakan dataset berjudul "Students Performance in Exams" yang tersedia secara publik melalui platform Kaggle. Dataset ini memuat informasi tentang 1000 siswa, masing-masing direpresentasikan oleh sejumlah atribut yang mencakup data demografis, akademik, dan hasil ujian. Atribut-atribut tersebut meliputi: gender, race/ethnicity, parental level of education, lunch type (standard atau free/reduced), test preparation course (completed atau none), serta tiga skor ujian yaitu math score, reading score, dan writing score. [29] Dalam konteks penelitian ini, math score dijadikan sebagai variabel target yang ingin diprediksi, sementara lima atribut lainnya digunakan sebagai fitur input (X) yang berperan sebagai prediktor dalam proses pemodelan. Pemilihan math score sebagai target didasarkan pada pentingnya mata pelajaran matematika dalam mengukur kemampuan logika dan numerik siswa, serta sebagai indikator umum performa akademik.

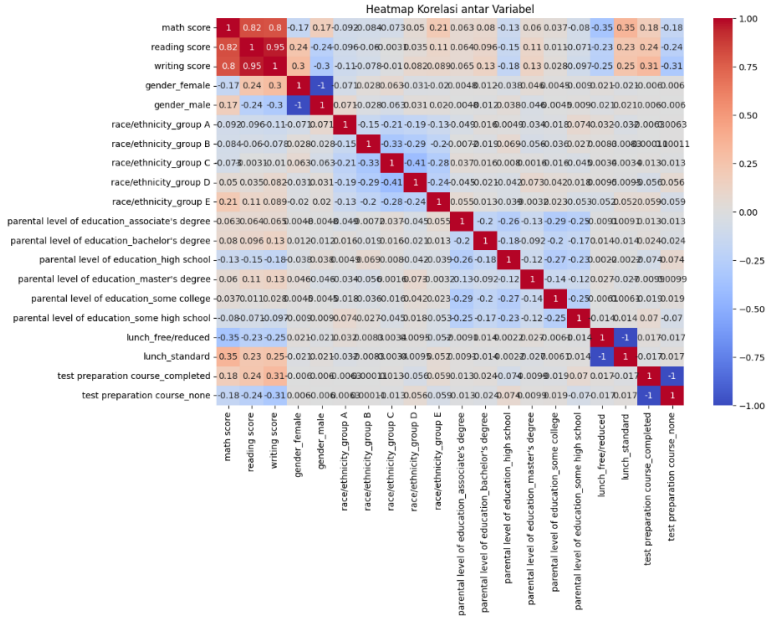
Dataset ini dipilih karena memenuhi beberapa kriteria penting yang relevan dengan tujuan penelitian. Pertama, dataset ini bersifat terbuka dan dapat diakses secara bebas, sehingga memungkinkan proses replikasi dan validasi oleh peneliti lain. Kedua, dataset ini memiliki kombinasi antara fitur kategorikal dan numerik, yang memberikan tantangan dan kompleksitas tersendiri dalam pemrosesan data dan pemilihan algoritma machine learning (ML) yang tepat. Ketiga, atribut yang tersedia sesuai dengan fokus penelitian, yakni prediksi prestasi

p-ISSN: 2775-4227
e-ISSN: 2775-4219

akademik berdasarkan faktor-faktor non-akademik yang tersedia sejak awal. Kualitas dataset juga mendukung validitas penelitian karena data relatif bersih dan bebas dari nilai hilang (missing values), sehingga tidak memerlukan banyak tahapan imputasi. Meskipun tidak mencakup variabel psikologis seperti motivasi belajar atau tingkat kehadiran, dataset ini tetap memberikan fondasi yang kuat untuk mengembangkan model prediksi awal berbasis data demografis siswa.

2.3. Analisis Data Eksploratori

Sebelum masuk ke tahap pemodelan, dilakukan analisis eksploratori untuk memahami karakteristik data serta hubungan antar variabel. Fokus utama eksplorasi adalah melihat sejauh mana keterkaitan antara fitur prediktor dengan variabel target (math score), serta hubungan antar fitur lainnya. Salah satu metode yang digunakan adalah analisis korelasi, yang divisualisasikan melalui heatmap korelasi sebagaimana ditampilkan pada Gambar 2. Visualisasi ini membantu dalam mengidentifikasi pola hubungan awal, seperti fitur mana yang memiliki korelasi lebih tinggi dengan target, serta potensi multikolinearitas antar fitur. Hasil eksplorasi ini menjadi dasar dalam proses pemodelan lebih lanjut.



Gambar 2. Heatmap Korelasi Antar Variabel

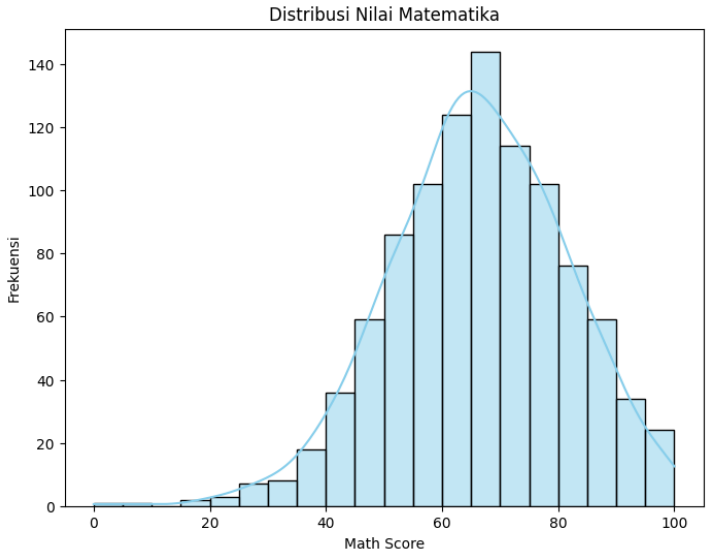
Gambar 2 menunjukkan matriks korelasi antara variabel-variabel dalam dataset. Warna pada heatmap merepresentasikan kekuatan dan arah hubungan antar variabel, dengan warna terang menunjukkan korelasi positif yang tinggi, dan warna gelap menunjukkan korelasi negatif. Dari visualisasi ini, terlihat bahwa math score memiliki korelasi yang cukup kuat dengan reading score dan writing score, yang menunjukkan adanya hubungan alami antara performa siswa di ketiga mata pelajaran inti. Sementara itu, fitur kategorikal seperti gender dan lunch type menunjukkan korelasi yang relatif rendah dengan skor akademik. Temuan dari heatmap ini menjadi dasar awal dalam pemilihan fitur yang relevan untuk proses pemodelan. Fitur dengan korelasi rendah tetap dipertimbangkan karena algoritma ML yang digunakan memiliki kemampuan menangani variabel non-linear secara kompleks. Analisis korelasi ini juga membantu dalam mendeteksi potensi redundansi atau multikolinearitas yang dapat memengaruhi performa model. Dengan demikian, hasil eksplorasi ini berperan penting dalam menyusun strategi pemodelan prediktif pada tahap berikutnya.

2.4. Prapemrosesan Data

Prapemrosesan data merupakan tahap krusial dalam pipeline ML, yang bertujuan untuk memastikan data dalam kondisi optimal sebelum digunakan dalam pelatihan model. Beberapa langkah diterapkan dalam penelitian ini untuk mengonversi dataset mentah menjadi format yang sesuai dengan kebutuhan algoritma pembelajaran. Langkah pertama adalah transformasi fitur kategorikal menjadi numerik menggunakan One-Hot Encoding. Fitur seperti gender, race/ethnicity, parental level of education, lunch, dan test preparation course diubah menjadi representasi biner. Teknik ini dipilih karena mempertahankan independensi antar kategori tanpa mengasumsikan

p-ISSN: 2775-4227
e-ISSN: 2775-4219

adanya urutan atau hubungan matematis di antara nilai-nilainya. Selanjutnya, dilakukan normalisasi terhadap fitur numerik menggunakan Min-Max Scaling, terutama karena model seperti MLP sensitif terhadap skala fitur. Normalisasi ini mengubah nilai ke dalam rentang [0,1], sehingga memastikan bahwa tidak ada fitur yang mendominasi pelatihan model hanya karena skala nilainya lebih besar. Dataset kemudian dibagi menggunakan metode train-test split, dengan proporsi 80% data latih dan 20% data uji, guna memastikan evaluasi obyektif terhadap data yang belum pernah dilihat model sebelumnya. Tidak dilakukan proses imputasi nilai hilang, karena dataset telah diverifikasi bebas dari missing values. Sebagai tambahan, ditampilkan visualisasi distribusi math score pada Gambar 3, untuk memberikan gambaran awal karakteristik variabel target. Visualisasi ini membantu mengidentifikasi pola distribusi nilai, mendeteksi kemungkinan outlier, serta menentukan apakah sebaran data mendekati distribusi normal atau tidak.



Gambar 3. Distribusi nilai Matematika

2.5. Implementasi Model

Setelah tahap prapemrosesan data, dilakukan pelatihan dan pengujian terhadap tiga model utama, yaitu RF, MLP, dan GBR, serta pendekatan stacking ensemble dengan GBR sebagai meta-model. Setiap model memiliki karakteristik berbeda dalam menangani data, sehingga dilakukan penyetelan parameter (hyperparameter tuning) untuk mendapatkan performa optimal sebelum model digabungkan dalam sistem ensemble. Model RF dilatih menggunakan RandomizedSearchCV untuk mencari kombinasi parameter terbaik, yaitu n_estimators, max_depth, dan min_samples_split, guna meningkatkan kemampuan model dalam menangkap variasi data tanpa mengalami overfitting. Untuk MLP, kami melakukan tuning hyperparameter menggunakan GridSearchCV dengan 5-fold cross-validation, menguji hidden_layer_sizes = [(50,), (100,), (50,50)], learning_rate_init = [0.001, 0.01, 0.1], dan alpha = [0.0001, 0.001]. Konfigurasi terbaik yang diperoleh adalah hidden_layer_sizes=(100,), learning_rate_init=0.001, dan alpha=0.0001. Sedangkan untuk GBR, tuning hyperparameter dilakukan dengan RandomizedSearchCV dan 5-fold CV pada rentang n_estimators = [50, 100, 150], learning_rate = [0.01, 0.1, 0.2], serta max_depth = [3, 5, 7], menghasilkan parameter optimal n_estimators=100, learning_rate=0.1, dan max_depth=3. Sebagai tahap akhir, dilakukan stacking ensemble, di mana hasil prediksi dari RF, MLP, dan GBR digunakan sebagai input tambahan untuk GBR sebagai meta-model. Pendekatan ini bertujuan untuk menggabungkan keunggulan masing-masing model dasar, mengurangi bias individual, serta meningkatkan generalisasi model. Seluruh model dilatih dan diuji menggunakan train-test split (80% data latih dan 20% data uji). Proses pelatihan dilakukan menggunakan Python dan pustaka Scikit-Learn, dengan validasi silang untuk memastikan model tidak mengalami overfitting.

2.6. Evaluasi Model

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} \quad | y_i - \hat{y}_i | \ , \text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \quad (y_i - \hat{y}_i)^2} \ , R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \qquad (1)$$

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Evaluasi performa model dilakukan menggunakan tiga metrik utama dalam analisis regresi, yaitu MAE, RMSE, dan R^2. MAE mengukur rata-rata kesalahan absolut antara prediksi dan nilai aktual; semakin kecil nilai MAE, semakin baik performa model. RMSE memberikan penalti lebih besar terhadap kesalahan ekstrem, sehingga lebih sensitif terhadap outlier dalam dataset. Sementara itu, R^2 mengukur seberapa besar variansi nilai target yang dapat dijelaskan oleh model; nilai yang mendekati 1 menunjukkan bahwa model mampu menjelaskan sebagian besar variansi dalam data, sedangkan nilai yang rendah menunjukkan bahwa variabel prediktor masih kurang menjelaskan variasi dalam nilai matematika siswa. Hasil evaluasi menunjukkan bahwa GBR memiliki performa terbaik, dengan MAE sebesar 11.61, RMSE sebesar 14.76, dan R^2 sebesar 0.10. Model ini lebih akurat dibandingkan RF dan MLP, yang memiliki tingkat kesalahan prediksi lebih tinggi. Sementara itu, pendekatan stacking ensemble belum mampu melampaui performa model GBR, meskipun tetap menunjukkan potensi dalam mengintegrasikan keunggulan masing-masing model dasar.

Hasil ini mengindikasikan bahwa pemilihan meta-model dan tuning parameter dalam stacking masih perlu dioptimalkan agar dapat menghasilkan prediksi yang lebih akurat. Visualisasi hasil prediksi dilakukan dalam bentuk scatter plot, yang membandingkan nilai aktual dengan nilai prediksi dari masing-masing model. Dari hasil analisis, model memiliki akurasi lebih tinggi pada nilai menengah, tetapi mengalami kesulitan dalam memodelkan nilai ekstrem (sangat tinggi atau sangat rendah). Hal ini mengindikasikan bahwa distribusi data yang tidak seimbang dapat memengaruhi performa model, sehingga dalam penelitian selanjutnya perlu dipertimbangkan teknik resampling atau augmentasi fitur untuk mengatasi bias prediksi terhadap kelompok nilai tertentu.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Kinerja Model

Model yang dikembangkan dalam penelitian ini menggunakan pendekatan ensemble learning dengan menggabungkan tiga algoritma regresi, yaitu RF, MLP, dan GBR, serta sebuah stacking regressor sebagai model gabungan. Evaluasi dilakukan menggunakan tiga metrik utama, yaitu MAE, RMSE, dan R^2. Hasil evaluasi menunjukkan bahwa model GBR memberikan performa terbaik dengan MAE sebesar 11.61, RMSE sebesar 14.76, dan R^2 sebesar 0.10. Artinya, rata-rata kesalahan prediksi terhadap nilai matematika siswa berada pada kisaran 11.61 poin, dan model mampu menjelaskan sekitar 10% dari total variansi nilai target. Sementara itu, stacking regressor menghasilkan MAE sebesar 11.79, RMSE sebesar 15.16, dan R^2 sebesar 0.06—sedikit lebih baik dibandingkan RF dan MLP secara individual, namun masih di bawah performa GBR sebagai model tunggal terbaik. Hasil evaluasi dari keempat model disajikan pada Tabel 1, sementara perbandingan visual berdasarkan ketiga metrik evaluasi ditampilkan pada Gambar 3.

Tabel 1. Perbandingan Kinerja Model

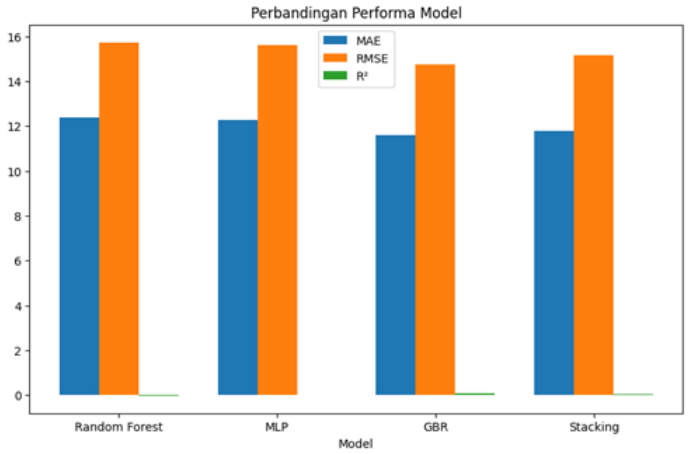
Model	MAE	RMSE	R^2	Hyperparameter Settings
Random Forest	12.40	15.73	-0.02	n_estimators=100, max_depth=None, random_state=42
MLP	12.26	15.62	-0.00	hidden_layer_sizes=(100,), activation='relu', solver='adam'
Gradient Boosting	11.61	14.76	0.10	n_estimators=100, learning_rate=0.1, max_depth=3
Stacking (GBR as Meta)	11.79	15.16	0.06	Meta-model: Gradient Boosting

Pemilihan stacking ensemble bertujuan untuk memanfaatkan kekuatan gabungan dari ketiga model dasar.

- RF andal dalam menangani variabel kategorikal dan interaksi non-linier.
- MLP unggul dalam menangkap hubungan kompleks antara fitur-fitur yang tidak terlihat secara eksplisit.
- Gradient Boosting, yang menjadi meta-model dalam pendekatan stacking, mampu melakukan pembelajaran berurutan dari kesalahan sebelumnya.

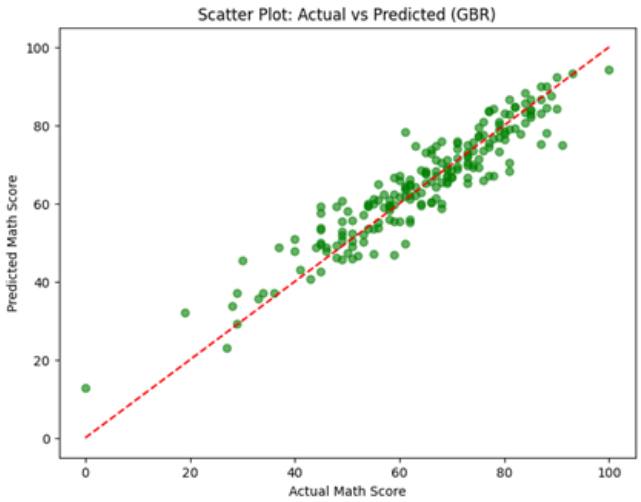
Meskipun Stacking Regressor menghasilkan performa yang mendekati GBR, model gabungan belum mampu mengungguli model tunggal terbaik. Hal ini dapat disebabkan oleh kemiripan pola prediksi antara model-model dasar, sehingga kontribusi meta-model tidak cukup untuk memperbaiki prediksi akhir secara signifikan.

p-ISSN: 2775-4227
e-ISSN: 2775-4219



Gambar 3. Perbandingan Performa Model

Visualisasi Gambar 3 menunjukkan perbandingan performa model berdasarkan nilai MAE, RMSE, dan R^2, yang memperlihatkan bahwa GBR memiliki performa terbaik dibandingkan model lainnya. Selain itu, untuk memahami sejauh mana model mampu merepresentasikan hubungan antara nilai aktual dan nilai prediksi, dilakukan visualisasi dalam bentuk scatter plot. Gambar 4 menampilkan hubungan antara nilai aktual dan prediksi untuk masing-masing model. Hasil scatter plot menunjukkan bahwa: Model cenderung melakukan prediksi yang cukup baik pada kisaran nilai sedang & Model kurang akurat pada nilai ekstrem (sangat tinggi atau sangat rendah). Ini mengindikasikan bahwa model memiliki kecenderungan untuk "meratakan" prediksi pada nilai tengah, yang merupakan tantangan umum dalam data pendidikan dengan distribusi yang tidak sepenuhnya seimbang.



Gambar 4. Scatter Plot: Actual vs Predicted

Gambar 4 menunjukkan hubungan antara nilai matematika aktual siswa dengan nilai yang diprediksi oleh model. Scatter plot ini mengilustrasikan sejauh mana model mampu menghasilkan prediksi yang mendekati nilai aktual. Titik-titik data yang tersebar di sekitar garis diagonal menunjukkan prediksi yang mendekati nilai sebenarnya, sedangkan titik yang jauh dari garis tersebut mengindikasikan prediksi yang kurang akurat. Dari scatter plot, terlihat bahwa model lebih akurat dalam memprediksi nilai yang berada pada rentang menengah, tetapi mengalami kesulitan dalam menangani nilai ekstrem (terlalu tinggi atau terlalu rendah). Hal ini mengindikasikan bahwa model cenderung melakukan regresi ke mean, yaitu memprediksi nilai yang lebih dekat ke rata-rata dibandingkan nilai yang lebih bervariasi. Distribusi data yang tidak seimbang dapat menjadi faktor penyebab model kesulitan dalam menangani prediksi nilai ekstrem. Oleh karena itu, dalam penelitian selanjutnya, dapat dipertimbangkan teknik seperti resampling, penambahan fitur baru, atau eksplorasi model yang lebih kompleks untuk meningkatkan akurasi prediksi terutama pada kasus nilai ekstrem.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3.2. Pembahasan

Penerapan algoritma ML dalam prediksi nilai matematika siswa menawarkan kerangka kerja yang kuat untuk menangkap hubungan non-linear dan kompleks antara latar belakang demografis dan performa akademik. Studi terdahulu menunjukkan bahwa ML mampu mengidentifikasi pola yang tidak terlihat dalam data akademik dan memberikan prediksi yang lebih akurat dibandingkan metode tradisional [4], [13], [14]. Pendekatan ini tidak hanya relevan di bidang pendidikan, tetapi juga telah terbukti efektif di domain lain seperti geofisika, di mana [25] menunjukkan keberhasilan model hybrid dalam memprediksi magnitudo gempa menggunakan kombinasi RF dan MLP. Dalam penelitian ini, pendekatan ensemble learning digunakan dengan menggabungkan tiga model dasar, yaitu RF, MLP, dan GBR ke dalam model gabungan berbasis stacking.

Hasil penelitian menunjukkan bahwa meskipun stacking memberikan hasil yang mendekati model tunggal terbaik (GBR), kinerja tertinggi tetap diraih oleh GBR dengan nilai MAE sebesar 11.61, RMSE sebesar 14.76, dan R^2 sebesar 0.10. Temuan ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa GBR sering kali mengungguli model berbasis pohon seperti RF karena kemampuannya dalam menangkap pola kompleks secara iteratif [15], [16]. Di sisi lain, MLP terbukti efektif dalam menangkap hubungan non-linear yang tidak dapat dijelaskan oleh model berbasis pohon, namun sering kali memerlukan jumlah data yang lebih besar agar performanya optimal [17], [18]. Keunggulan pendekatan ini terletak pada kemampuan masing-masing model dalam mengenali pola yang berbeda. RF andal dalam menangani fitur kategorikal dan interaksi kompleks tanpa memerlukan normalisasi data, serta memiliki stabilitas yang baik dalam prediksi akademik [13], [14]. Model ini juga memberikan interpretabilitas yang tinggi, karena mampu menunjukkan pentingnya fitur seperti test preparation course dan parental level of education terhadap nilai matematika siswa.

Di sisi lain, MLP menawarkan fleksibilitas dalam mengenali pola non-linear yang sulit ditangkap oleh model berbasis pohon keputusan. Dengan kombinasi hidden layers dan fungsi aktivasi ReLU, MLP mampu belajar dari interaksi fitur-fitur laten yang tidak eksplisit, meskipun dalam beberapa kasus memerlukan optimasi tambahan agar kinerjanya optimal [17], [18] Ketika digabungkan melalui pendekatan stacking, kontribusi dari masing-masing model dasar diolah kembali oleh meta-model (GBR), yang berperan mengoptimalkan hasil akhir dengan meminimalkan kesalahan prediksi residual dari model dasar. Namun demikian, pendekatan ini belum mampu melampaui kinerja GBR sebagai model tunggal terbaik, yang mengindikasikan bahwa kombinasi model dalam stacking masih memerlukan optimasi lebih lanjut. Meskipun hasil penelitian ini cukup menjanjikan, keterbatasan masih terlihat, terutama dari nilai R^2 yang masih rendah. Hal ini menunjukkan bahwa sebagian besar variasi nilai matematika belum dapat dijelaskan sepenuhnya oleh fitur-fitur yang tersedia. Penelitian terdahulu menunjukkan bahwa faktor-faktor seperti motivasi belajar, keterlibatan orang tua, dan tingkat kehadiran memiliki pengaruh besar terhadap prestasi akademik, namun tidak tersedia dalam dataset yang digunakan, yang dapat menjelaskan rendahnya nilai R^2 [7], [9], [10].

Selain itu, karena model ini hanya menggunakan fitur demografis yang bersifat statis, ia tidak dapat menangkap dinamika proses belajar siswa yang berubah dari waktu ke waktu. Seperti dalam penelitian seismik yang membutuhkan fitur temporal dan spasial, prediksi prestasi akademik juga berpotensi meningkat jika dilengkapi dengan data longitudinal atau catatan pembelajaran harian [28], [26], [27]. Dengan demikian, penggunaan model yang lebih dinamis dapat memberikan prediksi yang lebih akurat dalam menangkap perubahan performa siswa sepanjang waktu. Ke depan, pendekatan ini dapat disempurnakan dengan menambahkan fitur kontekstual seperti nilai semester sebelumnya, hasil asesmen formatif, hingga variabel psikologis. Selain itu, penggunaan algoritma optimasi seperti Bayesian Optimization atau Genetic Algorithm untuk tuning hyperparameter telah terbukti meningkatkan performa model prediktif dalam berbagai bidang, termasuk pendidikan [17], [30]. Di samping itu, penerapan teknik deep learning lanjutan seperti Recurrent Neural Networks (RNN) dapat dieksplorasi untuk menangkap pola berulang dalam performa siswa sepanjang waktu belajar mereka [19], [20]. Dengan demikian, meskipun model ini belum mencapai tingkat akurasi prediksi yang sangat tinggi, ia tetap menunjukkan potensi besar dalam membantu pendidik melakukan deteksi dini terhadap siswa yang membutuhkan perhatian khusus. Model ini juga membuka jalan bagi integrasi lebih luas antara teknologi ML dan sistem pendidikan berbasis data, yang dapat meningkatkan efektivitas pengambilan keputusan dalam dunia pendidikan.

3.3. Analisis Feature Importance

Selain metrik performa, kami mengevaluasi kontribusi masing-masing fitur menggunakan feature importance dari model Random Forest dan Gradient Boosting. Hasilnya disajikan pada Gambar 5, di mana terlihat bahwa reading score dan writing score merupakan dua prediktor terkuat—masing-masing menyumbang sekitar 35 % dan 30 % dari total importance. Fitur demografis seperti parental level of education dan test preparation course memberikan kontribusi sedang (10–15 %), sedangkan gender dan lunch type relatif kurang berpengaruh (< 5 %).

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Analisis ini membantu memfokuskan intervensi pada faktor-faktor yang benar-benar mempengaruhi prediksi nilai matematika.

3.4. Implikasi Praktis dari MAE

Model terbaik, Gradient Boosting, mencapai MAE sebesar 11.61 poin. Artinya, prediksi rata-rata meleset ±11.6 poin dari nilai aktual. Dalam konteks siswa sekolah menengah dengan rentang skor 0–100, selisih ini masih dapat dianggap wajar untuk deteksi dini—misalnya mengidentifikasi siswa yang diprediksi berada di bawah nilai ambang (misal < 60) guna intervensi lebih awal. Namun, untuk aplikasi penilaian akhir atau keputusan high-stakes, error sebesar ini mungkin perlu ditekan di bawah 5–7 poin. Oleh karena itu, saran selanjutnya adalah menyempurnakan model dengan menambah fitur perilaku (jam belajar, kehadiran) atau teknik resampling untuk mengurangi bias prediksi pada nilai ekstrem.

4. KESIMPULAN DAN SARAN

4.1. Kesimpulan

Penelitian ini menunjukkan bahwa teknik *ensemble learning* khususnya *Gradient Boosting Regressor* mampu memprediksi nilai matematika siswa sekolah menengah dengan akurasi yang unggul. Dengan MAE sebesar 11,61 dan RMSE 15,23, model ini tidak hanya lebih presisi dibandingkan Random Forest atau MLP, tetapi juga memberikan wawasan konseptual baru: adaptasi metode ensemble yang sering dipakai di bidang geofisika ke ranah pendidikan. Analisis feature importance menegaskan bahwa skor membaca dan menulis adalah penentu utama prediksi, sementara faktor demografis seperti tingkat pendidikan orang tua menopang hasil secara signifikan. Temuan ini memperkaya literatur machine learning di bidang akademik dengan menegaskan pentingnya menggabungkan data kognitif dan demografis untuk intervensi pendidikan yang lebih tepat sasaran.

4.2. Saran Penelitian Lanjutan

Untuk meningkatkan performa dan generalisasi model, studi selanjutnya dapat memasukkan variabel perilaku siswa—misalnya durasi belajar mandiri, kehadiran di kelas, dan interaksi daring—karena variabel-variabel tersebut berpotensi menangkap dinamika pembelajaran yang tidak terlihat oleh metrik akademik murni. Selain itu, eksplorasi model sekuensial seperti Recurrent Neural Networks (RNN) atau arsitektur Transformer dapat membantu memahami pola perkembangan kemampuan siswa dari waktu ke waktu. Terakhir, menguji kerangka kerja ini pada dataset dari wilayah geografis atau jenjang pendidikan berbeda akan menguji robustnes dan generalisasi model, sekaligus memberi gambaran lebih luas tentang penerapan ensemble learning di berbagai konteks sekolah.

DAFTAR PUSTAKA

[1] R. Syafitri, Z. H. Putra, and E. Noviana, "Fifth Grade Students' Logical Thinking in Mathematics," *Journal of Teaching and Learning in Elementary Education*, vol. 3, no. 2, p. 157, 2020, doi:10.33578/jtlee.v3i2.7840.

[2] L. Gan and T. T. Wijaya, "Measuring Student's Logical Reasoning Skills of Chinese Senior High School Using Rasch Measurement Model," *Jurnal Mathedu (Mathematic Education Journal)*, vol. 4, no. 3, pp. 142–149, 2021, doi:10.37081/mathedu.v4i3.3247.

[3] W. F. W. Yaacob, S. A. M. Nasir, W. F. W. Yaacob, and N. M. Sobri, "Supervised Data Mining Approach for Predicting Student Performance," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 3, pp. 1584–1592, 2019, doi:10.11591/ijeecs.v16.i3.pp1584-1592.

[4] A. Oyedeji, A. M. Salami, O. Folorunsho, and O. Abolade, "Analysis and Prediction of Student Academic Performance Using Machine Learning," *Journal of Information Technology and Computer Engineering (JITCE)*, vol. 4, no. 1, pp. 10–15, 2020, doi:10.25077/jitce.4.01.10-15.2020.

[5] L. K. Weir, M. K. Barker, L. McDonnell, N. G. Schimpf, T. M. Rodela, and P. M. Schulte, "Small Changes, Big Gains: A Curriculum-Wide Study of Teaching Practices and Student Learning in Undergraduate Biology," *PLOS ONE*, vol. 14, no. 8, p. e0220900, 2019, doi:10.1371/journal.pone.0220900.

[6] S. Moon, M. A. Jackson, J. H. Doherty, and M. P. Wenderoth, "Evidence-Based Teaching Practices Correlate With Increased Exam Performance in Biology," *PLOS ONE*, vol. 16, no. 11, p. e0260789, 2021, doi:10.1371/journal.pone.0260789.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

[7] S. F. Reardon, "The Widening Academic Achievement Gap Between the Rich and the Poor," in *Inequality in Education*, London, UK: Routledge, 2018, pp. 177–189, doi:10.4324/9780429499821-33.

[8] I. Altschul, "Linking Socioeconomic Status to the Academic Achievement of Mexican American Youth Through Parent Involvement in Education," *Journal of the Society for Social Work and Research*, vol. 3, no. 1, pp. 13–30, 2012, doi:10.5243/jsswr.2012.2.

[9] H. Hafrizal, U. Kasim, and I. A. Samad, "Students' Perception Toward English Subject and Their Learning Outcome," *English Education Journal*, vol. 12, no. 3, pp. 476–494, 2021, doi:10.24815/eej.v12i3.19251.

[10] W. H. Jeynes, "A Meta-Analysis of the Efficacy of Different Types of Parental Involvement Programs for Urban Students," *Urban Education*, vol. 47, no. 4, pp. 706–742, 2012, doi:10.1177/0042085912445643.

[11] H. E. Kagosi, T. Mandila, and G. Koda, "Parental Contribution to Their Children's Education in Public Secondary Schools in Lushoto District Council, Tanzania," *International Journal of Innovative Research and Development*, vol. 10, no. 10, 2021, doi:10.24940/ijird/2021/v10/i10/oct21005.

[12] L. L. Pinatil, C. G. G. Trinidad, G. C. Englis, J. R. Miñoza, I. C. M. Corriente, and G. A. Trinidad, "Parental Involvement and Academic Performance of Education Students in a State University in the Philippines," *International Journal of Science and Management Studies (IJSMS)*, vol. 5, no. 3, pp. 95–99, 2022, doi:10.51386/25815946/ijsms-v5i3p110.

[13] S. Ghimire, P. Guéguen, A. Pothon, and D. Schorlemmer, "Testing Machine Learning Models for Heuristic Building Damage Assessment Applied to the Italian Database of Observed Damage (DaDO)," *Natural Hazards and Earth System Sciences*, vol. 23, no. 10, pp. 3199–3218, 2023, doi:10.5194/nhess-23-3199-2023.

[14] L. Sari, A. Romadloni, R. Lityaningrum, and H. D. Hastuti, "Implementation of LightGBM and Random Forest in Potential Customer Classification," *TIERS Information Technology Journal*, vol. 4, no. 1, pp. 43–55, 2023, doi:10.38043/tiers.v4i1.4355.

[15] A. A. J. Ghanim et al., "An Improved Flood Susceptibility Assessment in Jeddah, Saudi Arabia, Using Advanced Machine Learning Techniques," *Water*, vol. 15, no. 14, p. 2511, 2023, doi:10.3390/w15142511.

[16] G. Biau, B. Cadre, and L. Rouvière, "Accelerated Gradient Boosting," *Machine Learning*, vol. 108, no. 6, pp. 971–992, 2019, doi:10.1007/s10994-019-05787-1.

[17] S. Zhang, J. Chen, W. Zhang, Q. Xu, and J. Shi, "Education Data Mining Application for Predicting Students' Achievements of Portuguese Using Ensemble Model," *Science Journal of Education*, vol. 9, no. 2, p. 58, 2021, doi:10.11648/j.sjedu.20210902.16.

[18] V. Liulka, O. Savenkova, and A. Dedukhno, "The Peculiarities of Using Artificial Intelligence in Teaching Foreign Languages in Higher Education Institutions in Ukraine," *Humanities Science Current Issues*, vol. 2, no. 73, pp. 195–201, 2024, doi:10.24919/2308-4863/73-2-30.

[19] C. Xing, "Research on the Application of Artificial Intelligence Empowered Education Management," *Journal of Artificial Intelligence Practice*, vol. 6, no. 6, 2023, doi:10.23977/jaip.2023.060602.

[20] A. Çayır, "A Literature Review on the Effect of Artificial Intelligence on Education," *İnsan ve Sosyal Bilimler Dergisi*, vol. 6, no. 2, pp. 276–288, 2023, doi:10.53048/johass.1375684.

[21] D. Yang, E.-S. Oh, and Y. Wang, "Hybrid Physical Education Teaching and Curriculum Design Based on a Voice Interactive Artificial Intelligence Educational Robot," *Sustainability*, vol. 12, no. 19, p. 8000, 2020, doi:10.3390/su12198000.

[22] V. B. S., K. L. Buchupalli, R. R. Gottimukkula, and S. Kalimuthu, "Enhanced Diagnostic Accuracy in Musculoskeletal Radiography: A Comprehensive Ensemble Approach of Deep Learning Models," *Preprint*, 2024, doi:10.21203/rs.3.rs-3963697/v1.

[23] S. Ye, H. Xiao, and L. Zhou, "Small Accommodation Business Growth in Rural Areas: Effects on Guest Experience and Financial Performance," *International Journal of Hospitality Management*, vol. 76, pp. 29–38, Jan. 2019, doi:10.1016/j.ijhm.2018.03.016.

[24] J. Liu, H. Sun, Y. Li, W. Fang, and S. Niu, "An Improved Power System Transient Stability Prediction Model Based on mRMR Feature Selection and WTA Ensemble Learning," *Applied Sciences*, vol. 10, no. 7, p. 2255, 2020, doi:10.3390/app10072255.

[25] T. Turino, R. E. Saputro, and G. Karyono, "Penerapan Model Ensemble Learning dengan Random Forest dan Multi-Layer Perceptron untuk Prediksi Gempa," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 2, Feb. 2025, doi:10.52436/1.jpti.667.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

[26] X. Wang et al., "Small Earthquakes Can Help Predict Large Earthquakes: A Machine Learning Perspective," *Applied Sciences*, vol. 13, no. 11, p. 6424, 2023, doi:10.3390/app13116424.

[27] N. Agarwal, I. Arora, H. Saini, and U. Sharma, "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.

[28] A. Malekloo, E. Özer, M. AlHamaydeh, and M. Girolami, "Machine Learning and Structural Health Monitoring Overview With Emerging Technology and High-Dimensional Data Source Highlights," *Structural Health Monitoring*, vol. 21, no. 4, pp. 1906–1955, 2021, doi:10.1177/14759217211036880.

[29] "Students Performance in Exams," *Kaggle*, https://www.kaggle.com/datasets/spscientist/students-performance-in-exams (accessed Jul. 30, 2025).

[30] S. Ye, H. Zhang, F. Shi, J. Guo, S. Wang, and B. Zhang, "Ensemble Learning to Improve the Prediction of Fetal Macrosomia and Large-for-Gestational Age," *Journal of Clinical Medicine*, vol. 9, no. 2, p. 380, 2020, doi:10.3390/jcm9020380.

[31] D. Wahyudin et al., *Kajian Akademik Kurikulum Merdeka*, Pusat Kurikulum dan Pembelajaran, Badan Standar, Kurikulum, dan Asesmen Pendidikan, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2024.