**JURIKOM (Jurnal Riset Komputer)**, Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9240
Hal 869-878
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

# Analisis Komparatif Model Regresi Machine Learning untuk Prediksi Prestasi Akademik Siswa dengan Optimasi Hyperparameter

**Fernando Hose, Robet*, Hendri**

Teknik Informatika, STMIK Time, Medan, Indonesia
Email: [1]fhose010@gmail.com, [2*]robertdetime@gmail.com, [3]h4ndr7@hotmail.com
Email Penulis Korespondensi: robertdetime@gmail.com
Submitted **27-10-2025;** Accepted **06-12-2025;** Published **31-12-2025**

**Abstrak**

Rendahnya ketepatan identifikasi dini siswa berisiko kerap menghambat intervensi akademik tepat waktu. Penelitian ini menganalisis dan membandingkan tujuh algoritma pembelajaran mesin untuk memprediksi prestasi akademik siswa, guna menyediakan dasar model peringatan dini yang andal. Dataset mencakup 2.392 siswa dengan 15 fitur demografis, perilaku belajar, dan dukungan lingkungan. Pelatihan model dilakukan dengan optimasi *GridSearchCV* yang dikombinasikan dengan *stratified cross-validation* guna menekan *overfitting*. Kinerja dievaluasi menggunakan MAE, RMSE, dan $R^2$. Hasil menunjukkan CatBoost terbaik ($R^2$ = 0,774; RMSE = 0,581; MAE = 0,306), diikuti LightGBM ($R^2$ = 0,771) dan Gradient Boosting ($R^2$ = 0,767), sementara MLP terendah. Analisis *feature importance* menempatkan GPA sebagai prediktor dominan, disusul absensi dan waktu belajar mingguan. Temuan ini menegaskan keunggulan model berbasis *boosting* dalam menangkap hubungan nonlinier yang kompleks serta memberikan kerangka praktis bagi institusi pendidikan untuk membangun sistem peringatan dini berbasis data.

**Kata kunci**: Feature Importance; Machine Learning; Optimasi Hyperparameter; Prediksi Prestasi Akademik; Regresi

**Abstract**

Low accuracy in the early identification of at-risk students often hinders timely academic intervention. This study analyzes and compares seven machine learning algorithms to predict student academic achievement, aiming to provide a foundation for a reliable early warning model. The dataset includes 2.392 students with 15 features covering demographics, learning behavior, and environmental support. Model training was performed using GridSearchCV optimization combined with stratified cross-validation to mitigate overfitting.Performance was evaluated using MAE, RMSE, and $R^2$. The results show CatBoost performed the best $R^2$ = 0,774; RMSE = 0,581; MAE = 0,306) followed by LightGBM ($R^2$ = 0,771) and Gradient Boosting ($R^2$ = 0,767), while MLP showed the lowest performance. Feature importance analysis placed GPA as the dominant predictor, followed by absenteeism and weekly study time. These findings affirm the superiority of boosting-based models in capturing complex nonlinear relationships and provide a practical framework for educational institutions to build data-driven early warning systems.

**Keywords**: Feature Importance; Machine Learning; Hyperparameter Optimization; Academic Achievement Prediction; Regression

## 1. PENDAHULUAN

Pendidikan modern di era digital menuntut institusi untuk mengoptimalkan pemanfaatan data untuk meningkatkan kualitas pembelajaran dan hasil akademik siswa. Namun, akurasi identifikasi dini siswa beresiko prestasi rendah masih belum memadai, sehingga intervensi kerap tertunda dan kurang tepat sasaran [1]. Kompleksitas faktor, dukungan orang tua, kehadiran, aktivitas, serta karakteristik personal, menciptakan pola non-linear dan interaksi fitur yang sulit ditangkap metode statistic konvensional [2].

Di konteks ini, *machine learning* menawarkan algortima yang mampu mengenali pola kompleks pada data tabular pendidikan. Model-model modern seperti *XGBoost*, *CatBoost*, dan *LightGBM* dapat menangani relasi nonlinear serta interaksi antar fitur, sementara *ensembles* dan jaringan saraf tiruan memberi alternatif arsitektur yang kompetitif. Potensi tersebut hanya optimal bila didukung optimasi *hyperparameter* yang sistematis, karena konfigurasi yang tidak tepat beresiko menimbulkan *underfitting* maupun *overfitting* [3].

Penelitian ini menguji tujuh algoritma, *XGBoost*, *LightGBM*, *CatBoost, Random Forest*, *Gradient Boosting*, *Multi-Layer Perceptron* (MLP), dan *Stacking Regressor*, yang dipilih berdasarkan kekuatan masing-masing pada data tabular: efisiensi dan penanganan *missing values* (*XGBoost*/*LightGBM*), pemrosesan *categorical features* secara *native* (*CatBoost*), interpretabilitas *feature importance* (*Random Forest*), fleksibilitas *loss function* (Gradient Boosting), pemodelan nonlinier kompleks (MLP), dan penggabungan keunggulan lintas-model (*Stacking*) [4] [5].

Literatur terkait menunjukkan kemajuan penerapan ML di beragam domain prediktif, tetapi belum menjawab kebutuhan evaluasi komprehensif lintas-algoritma untuk prediksi prestasi akademik: (i) banyak studi hanya menguji satu–dua model atau skenario klasifikasi sehingga tidak memberi pembandingan adil bagi *boosting* modern pada regresi akademik [6] [7] [8]; (ii) *tuning hyperparameter* kerap tidak sistematis (Contoh : *random search* atau pengaturan bawaan) sehingga kualitas perbandingan dan replikasi terbatas [7] [9]; dan (iii) Kurangnya analisis prediktif yang fokus pada identifikasi *feature importance*, menyulitkan derivasi kebijakan intervensi [10].

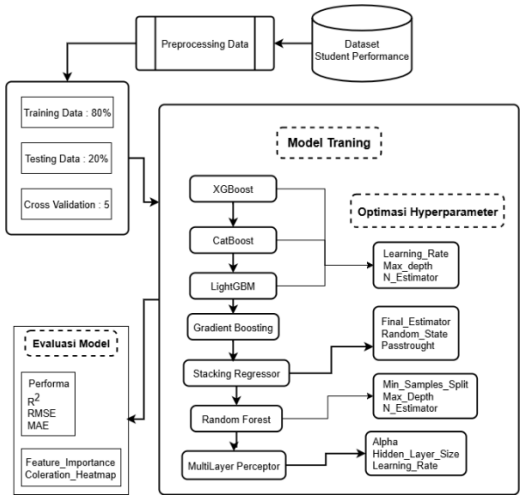
Penelitian ini mengisi celah tersebut dengan evaluasi *head-to-head* tujuh algoritma regresi mutakhir pada data siswa, *tuning hyperparameter* terstruktur berbasis *GridSearchCV* dengan stratified cross-validation, evaluasi terstandarisasi (MAE, RMSE, $R^2$), serta analisis *feature importance* untuk menafsirkan faktor dominan. Dataset mencakup 2.392 siswa dengan 15 variabel (*Age*, *Gender*, *Ethnicity*, *Parental Education*, *Study Time Weekly*, *Absences*, *Tutoring*, *Extracurricular*, *Parental Support*, *Sports*, *Music*, *Volunteering*, *Grade Class*, *GPA*, dan ID siswa). Hasil

**JURIKOM (Jurnal Riset Komputer)**, Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9240
Hal 869-878
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

penelitian diharapkan menjadi acuan institusi dalam membangun kerangka peringatan dini berbasis data yang akurat, sekaligus memberi pijakan kuat bagi penelitian lanjutan di bidang *data-driven learning*.

# 2. METODOLOGI PENELITIAN

Penelitian ini terdiri atas beberapa tahapan yang dimulai dari akuisisi dataset dan prapemrosesan, dilanjutkan pembagian data (train 80%/test 20%) dan validasi silang (k=5), pelatihan tujuh model dengan optimasi *hyperparameter*, serta evaluasi kinerja ($R^2$, RMSE, MAE) dan analisis *feature importance*, sebagaimana ditunjukkan pada Gambar 1."




**Gambar 1.** Tahapan penelitian

### 2.1 Pengumpulan Data

Data dalam penelitian ini berasal dari file *Student_Performance_Data_.csv* yang diperoleh dari website kaggle. Dataset ini mencakup berbagai aspek kehidupan siswa, baik akademik maupun non akademik. Berikut adalah deskripsi singkat dari fitur-fitur utama yang dijelaskan dalam Tabel 1.

**Tabel 1.** Fitur dan Deskripsi

| Fitur | Deskripsi |
|---|---|
| Student ID | Id unik siswa |
| Age | Usia siswa (15-18 tahun) |
| Gender | Jenis Kelamin (0 = Perempuan, 1 = laki-laki) |
| Ethnicity | Etnis (0-3) |
| ParentalEducational | Tingkat Pendidikan Orang Tua (0-4) |
| Study Time Weekly | Waktu Belajar Per minggu |
| Absences | Jumlah Absensi siswa |
| Tutoring | Mengikuti bimbingan belajar (0/1) |
| ParentalSupport | Dukungan Orang Tua (0-4) |
| Extracurricular | Kegiatan Ekstrakurikuler (0/1) |
| Sports, Music, Volunteering | Partisipasi pada aktifitas tertentu (0/1) |
| GPA | Nilai rata-rata akademik |
| Grade Class | Target prediksi atau IPK (0,0-4,0) |

Tabel 1 di atas menunjukkan 15 fitur yang akan digunakan dalam penelitian, yaitu: Etnis siswa akan dikodekan sebagai berikut (0 : Kaukasia, 1 : Afrika-Amerika, 2: Asia, 3: Lainnya), tingkat pendidikan orang tua dikodekan sebagai berikut (0 : Tidak ada, 1 : SMA/Sekolah Menengah Atas, 2 : pernah mengikuti perkuliahan tetapi tidak sampai menyelesaikan, 3 : Sarjana, 4 : Pendidikan yang lebih tinggi), ekstrakurikuler dan beberapa partisipasi keaktifan ditandai dengan (0 : tidak berpartisipasi, 1 : berpartisipasi), dan *GPA* yang dimana nilai rata-rata pada akademik.

### 2.2 *Preprocessing* Data

Dataset sebelum dilakukan pemodelan, data mengalami beberapa tahapan *preprocessing* untuk memastikan kesiapan data pada langkah selanjutnya :

a. *Data Cleansing*

Tahap pembersihan data mencakup deteksi dan penanganan nilai hilang, koreksi inkonsistensi tipe/format, serta penghapusan duplikasi, guna mengurangi bias dan kesalahan pada tahap pemodelan [11].

b. *Normalisasi Data*

**JURIKOM (Jurnal Riset Komputer)**, Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9240
Hal 869-878
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Normalisasi data tidak diterapkan pada model berbasis *boosting* (*Random Forest*, *Gradient Boosting*, *XGBoost*, *LightGBM*, *CatBoost*) karena pemilihan split bergantung pada urutan nilai dan bersifat invarian terhadap skala. *MinMaxScaler* (0–1) hanya digunakan pada jalur MLP yang sensitif terhadap skala fitur untuk menstabilkan optimisasi berbasis gradien. Persamaan normaliasai dengan *MinMaxScaler* ditunjukkan pada persamaan berikut.

$$MinMaxScaler = \frac{x_t - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Keterangan :
$x_t$ : Nilai data ke-i yang akan dinormalisasi
$x_{min}$ : Nilai minimum dari seluruh data pada fitur
$x_{max}$ : Nilai maksimum dari seluruh data pada fitur

c. Pembagian Dataset

Data dibagi menjadi data latih dan data uji dengan perbandingan 80:20. Pemisahan dilakukan secara acak terkontrol (*random_state* ditetapkan) untuk menjaga replikabilitas. Seluruh proses *preprocessing* di-*fit* hanya pada data latih, kemudian diterapkan ke data validasi/uji guna mencegah *data leakage*. Untuk mengevaluasi model secara adil, pada data latih digunakan validasi silang berlapis (k = 5) yang distratifikasi untuk regresi, sehingga distribusi nilai target pada tiap lipatan sebanding. Prosedur ini menurunkan varians estimasi performa dan mencegah optimisme berlebih saat pemilihan *hyperparameter*. Data uji 20% hanya digunakan sekali di akhir untuk konfirmasi kinerja model terbaik yang diperoleh dari proses *tuning* [12].

### 2.3 Model Training

Penelitian ini mengimplementasikan tujuh algoritma regresi untuk menentukan model dengan performa prediksi terbaik, yaitu *XGBoost*, *LightGBM*, *CatBoost*, *Random Forest*, *Gradient Boosting*, *MLP*, dan *Stacking Regressor*. Setiap model dioptimasi menggunakan *GridSearchCV* untuk menemukan *hyperparameter* terbaik. Parameter dari setiap model yang telah dikonfigurasi akan dilakukan *Cross Validation* (CV = 5) dan disesuai dengan *GridSearchCV* adalah sebagai berikut:

a. *XGBoost* adalah algoritma *gradient boosting* yang membangun model secara iteratif untuk memperbaiki kesalahan prediksi sebelumnya dengan kecepatan tinggi[13]. Parameter: *n_estimators* [100, 200], *max_depth* [5, 7], *learning_rate* [0.1, 0.05].
b. *LightGBM* merupakan *framework Microsoft* yang menggunakan strategi *leaf-wise growth* untuk efisiensi tinggi tanpa mengorbankan akurasi[14]. Parameter: *n_estimators* [100, 200], *max_depth* [5, 7], *learning_rate* [0.1, 0.05].
c. *CatBoost* adalah algoritma yang dirancang khusus menangani data kategorikal tanpa *preprocessing* tambahan menggunakan teknik *ordered boosting*[15]. Parameter: *n_estimators* [100, 200], *max_depth* [5, 7], *learning_rate* [0.1, 0.05].
d. *Random Forest* menggabungkan banyak *decision tree* dengan teknik bagging untuk meningkatkan akurasi dan mengurangi *overfitting*[16]. Parameter: *n_estimators* [100, 200], *max_depth* [None, 10], *min_samples_split* [2, 5].
e. *Gradient Boosting* membangun model secara bertahap dengan meminimalkan fungsi *loss* menggunakan *gradient descent*[17]. Parameter: *n_estimators* [100, 200], *max_depth* [5, 7], *learning_rate* [0.1, 0.05].
f. *MLP* adalah jaringan saraf tiruan dengan lapisan tersembunyi yang memerlukan normalisasi data dan tuning parameter intensif[18]. *Parameter: hidden_layer_sizes* [(50,), (100,), (100, 50)], *alpha* [0.001, 0.01], *learning_rate* ['*constant*', '*adaptive*'].
g. *Stacking Regressor* mengombinasikan prediksi dari beberapa model *base learner* melalui meta-model untuk hasil prediksi yang lebih optimal[19]. Parameter: *final_estimator* [*Random ForestRegressor, Gradient Boosting Regressor*], *passthrough* [*True, False*].

### 2.4 Evaluasi

Pada tahap akhir, evaluasi menjadi aspek penting dalam penyempurnaan model untuk mengetahui tingkat performa dari model yang telah dibangun. Penelitian ini menggunakan metrik evaluasi *Mean Absolute Error* (*MAE*), yaitu metode yang menghitung rata-rata nilai. Metrik ini dipilih karena mampu memberikan ukuran *error* yang mudah diinterpretasikan, yakni sebagai rata-rata selisih absolut antar nilai aktual dan nilai prediksi.

*MAE* lebih *robust* terhadap *outlier*, dikarenakan setiap *error* yang ada dihitung dalam bentuk *absolut* tanpa kontribusi terhadap *error* besar. Pada perhitungan *Mean Absolute Error* (*MAE*) [20], dinyatakan dengan persamaan berikut.

$$MAE_i = \frac{1}{n}\sum_{i=1}^{n}(|ai - bi|) \quad (2)$$

Keterangan dari Persamaan (1) adalah:
*n* = jumlah data uji (*sample data*) *bi* = Nilai Prediksi pada i
*ai* = Nilai aktual pada i |*ai* – *bi*| = Nilai absolut dari selisih pada nilai *pi* dan *ri*

Metrik evaluasi kedua yang digunakan adalah *Root Mean Squared Error* (*RMSE*), yaitu ukuran yang digunakan untuk menilai sejauh mana model mampu merepresentasikan data aktual. *RMSE* memberikan bobot lebih besar terhadap

**JURIKOM (Jurnal Riset Komputer)**, Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9240
Hal 869-878
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

kesalahan yang tinggi, sehingga sangat tepat dalam mengevaluasi kinerja model di mana kesalahan besar dianggap lebih serius. Karena sifatnya tersebut, *RMSE* lebih peka terhadap variasi kesalahan dan mampu memberikan penilaian yang lebih ketat, terutama ketika terdapat sejumlah *error* yang signifikan[21]. Rumus perhitungan *RMSE* dinyatakan sebagai berikut.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - d_i)^2} \quad (3)$$

Keterangan dari rumus Persamaan (2) yaitu :
n = Jumlah data (sample data) $d_i$ = nilai prediksi pada i
$a_i$ = nilai aktual terhadap i $(a_i - d_i)^2$ = kuadrat dari selisih $a_i$ dan $d_i$

*R-Square* merupakan metrik statistik yang digunakan dalam mengukur seberapa baik pada model regresi yang mampu menjelaskan variasi pada variabel dependen (target/hasil) berdasarkan satu atau lebih dalam variabel independen (fitur/prediktor) dalam hal ini $R^2$(R-Square) menunjukkan seberapa besar proporsi variasi data yang dapat dijelaskan oleh model regresi yang akan dibandingkan dengan variasi total data. $R^2$ memiliki nilai berkisaran 0 hingga 1 di mana dapat diartikan sebagai berikut.
$R^2$ = 0 (dimana model tidak mampu menjelaskan sama sekali validasi data).
$R^2$ = mendekati 1 (model semakin membaik dalam menjelaskan data).
$R^2$ = 1 (model sempurna, mampu dalam menjelaskan seluruh data yang validasi).
$R^2$ = negatif (data mengalami overfitting yang dimana prediksi lebih buruk dari rata-rata).

Pada $R^2$ terdapat rumus perhitungan sebagai persamaan berikut.

$$R^2 = 1 - \frac{SS_{Residuals}}{SS_{total}} \quad (4)$$

$SS_{Residuals}$ = *Sum of Squares of residuals* (jumlah kuadrat residual/error)
$SS_{total}$ = *Sum of Squares* (jumlah kuadrat total)

Pada $SS_{Residuals}$ adalah jumlah kuadrat dari selisih antara nilai aktual dan pada nilai prediksi, yang dimana mengukur ketidakmampuan model. Dan pada $SS_{total}$ dimana jumlah kuadrat dari selisih antara nilai yang aktual dan rata-rata nilai aktual. Mengukur kedua *Sum of Squares* dalam dua persamaan berikut.

$$SS_{Residuals} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \ dan \ SS_{total} = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2 \quad (5)$$

$y_i$ = nilai aktual pada i
$\hat{y}_i$ = nilai prediksi dari model
$\bar{y}_i$ = nilai rata-rata dari semua nilai aktual
n = jumlah data

Nilai kedua *sum* ini biasanya dalam menjumlahkan nilai aktual yang ada sebelum memasuki nilai $R^2$ yang dimana nilai tersebut akan menjadi nilai metrik statik dalam hal menentukan model ini apakah mampu membaca dataset atau tidak.

# 3. HASIL DAN PEMBAHASAN

## 3.1 Penerapan Algoritma dan Proses Optimasi Hyperparameter

Sebelum hasil prediksi akhir diperoleh, setiap model melalui proses pelatihan dan optimasi yang sistematis. Tujuannya memastikan performa dievaluasi pada konfigurasi *hyperparameter* terbaik, bukan parameter bawaan. Dataset dibagi menjadi 80% data latih dan 20% data uji (uji hanya dipakai sekali di akhir). Pada data latih, setiap algoritma dituning menggunakan *GridSearchCV* dengan skema 5-*fold cross-validation* yang distratifikasi untuk regresi melalui pembinningan kuantil target, sehingga distribusi nilai FinalScore di tiap lipatan sebanding.

Untuk setiap algoritma, kombinasi *hyperparamete*r dievaluasi berdasarkan $R^2$ rata-rata pada lipatan validasi; RMSE/MAE digunakan sebagai *tie-breaker*. Konfigurasi dengan skor tertinggi dipilih dan di-*refit* pada seluruh data latih, lalu dievaluasi pada data uji 20%. Selama tuning, MLP menunjukkan sensitivitas terhadap inisialisasi bobot dan memerlukan penskalaan ketat, sedangkan model *boosting* (*CatBoost*, *LightGBM*) relatif stabil dengan konvergensi cepat dan konsistensi skor antarlipatan. Gambar 3 (*learning curve*) menunjukkan *gap train–validation* yang kecil, terutama pada *CatBoos*t, mengindikasikan tidak terjadi *overfitting* signifikan. Tabel 2 merangkum konfigurasi akhir yang digunakan pada evaluasi.

## 3.2 Hasil Pengujian Pada Model

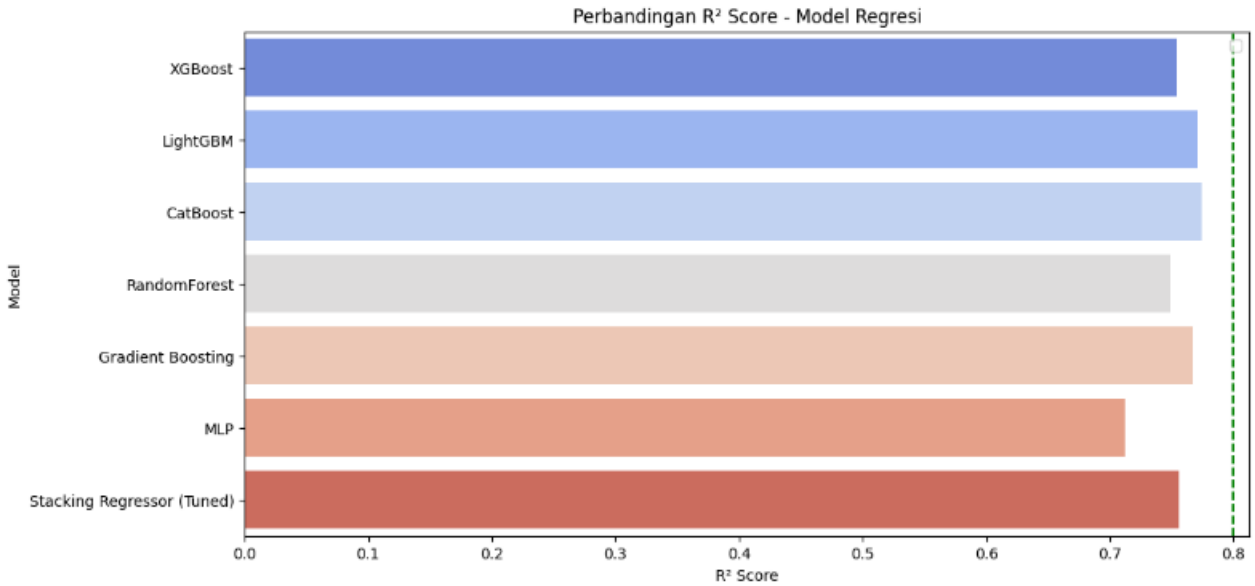
**JURIKOM (Jurnal Riset Komputer)**, Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9240
Hal 869-878
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Berdasarkan evaluasi terhadap tujuh algoritma *machine learning* yang telah melalui proses *tuning hyperparameter*, diperoleh hasil performa yang bervariasi, sebagaimana ditunjukkan pada Tabel 2, dan gambar 2.

**Tabel 2**. Performa Model Regresi Pada Hasil Prediksi *FinalScore*

| Model | Optimasi Hyperparameter | $R^2$ | RMSE | MAE |
|---|---|---|---|---|
| *CatBoost* | *Learning_Rate* : 0,05<br>*Max_depth* : 5<br>*N_Estimator* : 100 | 0.774114 | 0.581 | 0.306 |
| *LightGBM* | *Learning_Rate* : 0.05<br>*Max_depth* : 5<br>*N_Estimator* : 100 | 0.770723 | 0.585 | 0.284 |
| *Gradient Boosting* | *Learning_Rate* : 0.05<br>*Max_depth* : 5<br>*N_Estimator* : 100 | 0.767296 | 0.590 | 0.272 |
| *Stacking Regressor* | *Final_estimator : Gradient BoostingRegressor* (*Random_state* = 42)<br>*Passtrought : True* | 0.755817 | 0.604 | 0.274 |
| *XGboost* | *Learning_Rate* : 0.05<br>*Max_depth* : 5<br>*N_Estimator* : 100 | 0.753951 | 0.606 | 0.293 |
| *Random Forest* | *Min_Samples_Split* : 5<br>*Max_depth* : 10<br>*N_Estimator* : 200 | 0.748689 | 0.612 | 0.293 |
| *MLP* | *Alpha* : 0,01<br>*Hidden_layer_size* : 50<br>*Learning_Rate : constant* | 0.712129 | 0.656 | 0.411 |

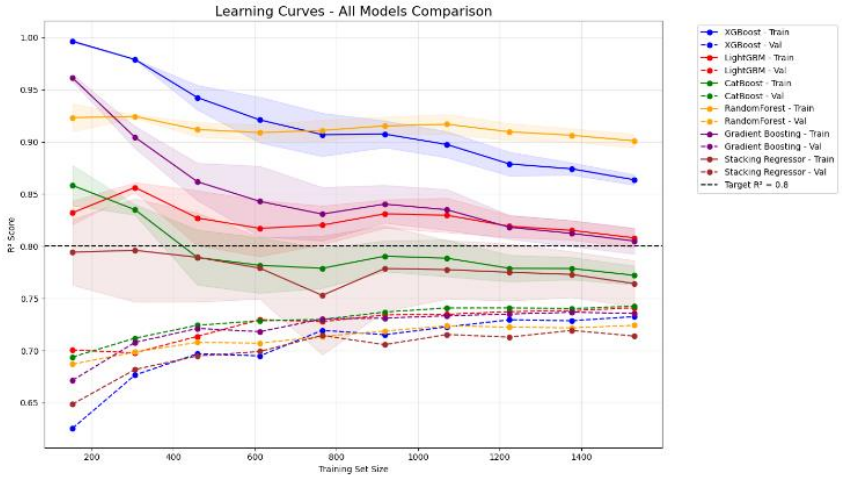
Berdasarkan evaluasi akhir (Tabel 2), CatBoost mencapai $R^2$ tertinggi (= 0.774) dengan RMSE 0.581 dan MAE 0.306. LightGBM dan Gradient Boosting mengikuti sangat dekat ($R^2$ = 0.771 dan 0.767); selisih terhadap CatBoost masing-masing $\Delta R^2 \approx 0.003$ dan 0.007. Perlu dicatat, MAE terendah justru diperoleh Gradient Boosting (0.272), disusul LightGBM (0.284), menunjukkan bahwa keduanya menghasilkan kesalahan absolut rata-rata sedikit lebih kecil, sementara CatBoost lebih unggul dalam proporsi variasi yang dijelaskan. Sesuai protokol studi ini, $R^2$ ditetapkan sebagai kriteria utama, dengan RMSE/MAE sebagai pertimbangan sekunder; oleh karena itu CatBoost ditetapkan sebagai model terbaik. Temuan ini konsisten dengan karakter data tabular pendidikan, di mana keluarga gradient boosting cenderung unggul.




**Gambar 2**. Hasil Prediksi Ketujuh Model Regresi

Gambar 2 menampilkan sebaran prediksi vs nilai aktual; garis identitas $y=x$ menunjukkan *CatBoost* memiliki penyebaran terdekat. Gambar 3 memperlihatkan *learning curve* yang stabil lintas model; *CatBoost* mencapai stabilitas tercepat tanpa indikasi *overfitting*. Keunggulan *CatBoost* selaras dengan kemampuannya menangani fitur kategorikal secara *native* dan memodelkan interaksi nonlinier secara aditif.

**JURIKOM (Jurnal Riset Komputer)**, Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9240
Hal 869-878
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



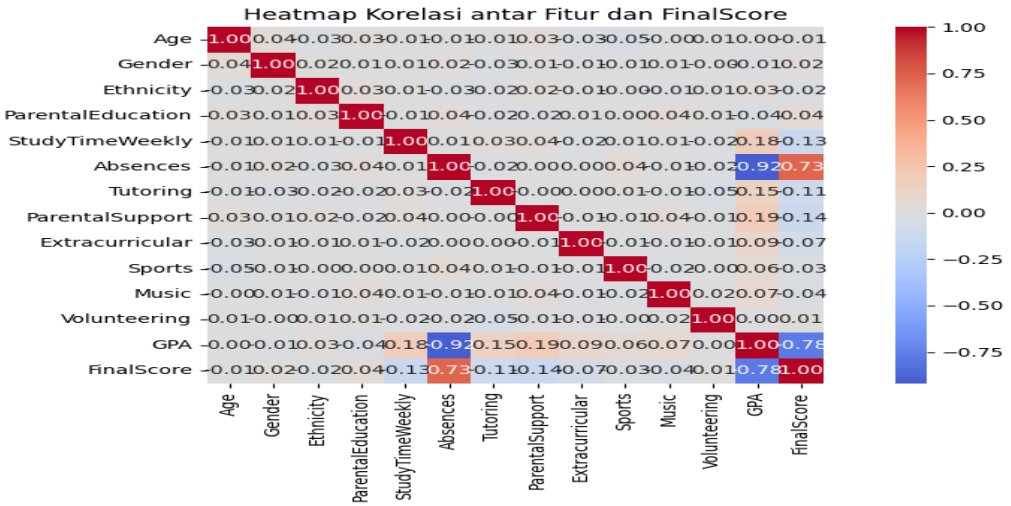

**Gambar 3**. Hasil *Learning Curve*

### 3.3 Analisis Korelasi dan *Feature Importance*

Analisis korelasi ditunjukkan pada Tabel 3, yang memperlihatkan bahwa GPA memiliki korelasi negatif yang sangat kuat dengan FinalScore (–0.783). Interpretasi ini konsisten dengan konteks sistem penilaian yang menggunakan skala inverse (nilai rendah = performa tinggi), sebagaimana dikonfirmasi oleh Gambar 6.

**Tabel** 3. Korelasi Antara Fitur Dan *FinalScore*

| Fitur | Korelasi |
|---|---|
| GPA | -0.783 |
| Absence | 0.729 |
| ParentalSupport | -0.137 |
| StudyTimeWeekly | -0.134 |
| Tutoring | -0.112 |
| Extracurricular | -0.070 |
| parentalEducational | 0.041 |
| Music | -0.036 |
| Sports | -0.027 |
| Ethnicity | -0.023 |
| Gender | 0.023 |
| Volunteering | 0.013 |
| Age | -0.006 |




**Gambar 4.** Heatmap Korelasi Antar Fitur dan *FinalScore*

Pada Gambar 4. Pola korelasi ini dapat dilihat secara visual yang menampilkan *heatmap* korelasi antar fitur dan *FinalScore*. *Heatmap* tersebut memperkuat temuan numerik pada Tabel 3, di mana GPA dan *Absences* menunjukkan intensitas warna paling gelap (korelasi kuat, baik negatif maupun positif), sedangkan sebagian besar fitur lain menunjukkan warna terang yang mencerminkan korelasi lemah. Selain itu, tidak terdapat indikasi kuat multikolinearitas

**JURIKOM (Jurnal Riset Komputer)**, Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9240
Hal 869-878
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

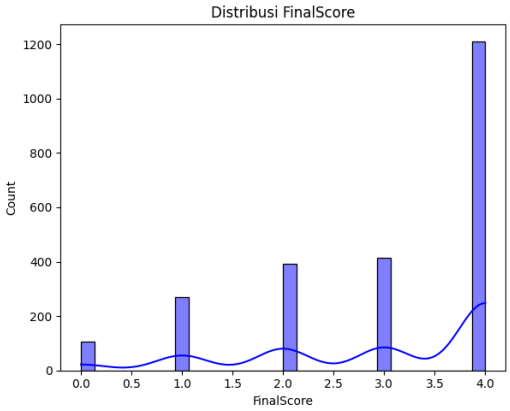
antar fitur prediktor, karena korelasi silang antar fitur (selain GPA dan *Absences*) umumnya berada di bawah 0,5. Hal ini mendukung stabilitas model regresi yang digunakan.

Setelah dilakukan evaluasi performa model, dan korelasi antar fitur, langkah berikutnya adalah menganalisis kontribusi setiap fitur terhadap prediksi. Tabel 4 menampilkan hasil *feature importance* dari ketujuh model regresi.

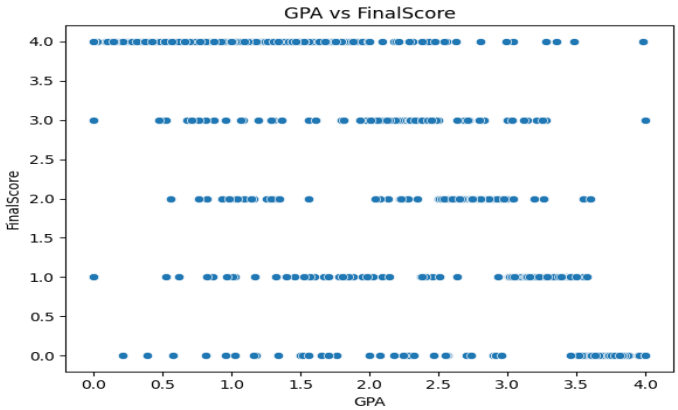
**Tabel** 4. *Feature Importance* Pada Tujuh Model

| Feature | XGBoost | LightGBM | CatBoost | R.Forest | MLP | G.Boosting | S.Regressor | Mean |
|---|---|---|---|---|---|---|---|---|
| GPA | 70.73 | 31.76 | 85.89 | 88.41 | 60.59 | 97.73 | 97.93 | 76.15 |
| Absences | 2.64 | 11.00 | 4.82 | 1.53 | 65.19 | 0.23 | 2.26 | 12.52 |
| StudyTimeWeekly | 2.85 | 20.71 | 2.36 | 5.66 | -20.74 | 1.52 | -0.44 | 1.70 |
| ParentalSupport | 3.88 | 5.19 | 1.52 | 1.14 | -1.06 | 0.15 | 0.07 | 1.56 |
| ParentalEducational | 2.60 | 6.23 | 1.06 | 0.90 | -2.10 | 0.08 | 0.04 | 1.26 |
| Ethnicity | 2.05 | 5.40 | 1.23 | 0.64 | -0.90 | 0.02 | 0.09 | 1.22 |
| Gender | 3.19 | 4.83 | 0.57 | 0.25 | -1.71 | 0.09 | -0.22 | 1.00 |
| Age | 2.29 | 5.50 | 1.07 | 0.53 | -2.56 | 0.04 | -0.04 | 0.98 |
| Tutoring | 2.11 | 2.39 | 0.49 | 0.30 | 1.30 | 0.04 | 0.01 | 0.95 |
| Music | 1.23 | 2.02 | 0.28 | 0.13 | 1.89 | 0.03 | 0.10 | 0.81 |
| Sports | 2.23 | 2.91 | 0.07 | 0.19 | -0.62 | 0.01 | 0.29 | 0.72 |
| Extracurricular | 2.54 | 0.88 | 0.46 | 0.21 | 0.49 | 0.02 | -0.01 | 0.66 |
| Volunteering | 1.67 | 1.19 | 0.19 | 0.11 | 0.22 | 0.08 | -0.07 | 0.48 |

Analisis *feature importance* pada model *CatBoos*t mengonfirmasi dominasi GPA (85.89) sebagai prediktor terkuat, diikuti oleh *Absence*s (4.82) dan *StudyTimeWeekl*y (2.36). Temuan ini konsisten dengan analisis korelasi, sekaligus menegaskan bahwa faktor akademik dan kehadiran merupakan penentu utama dalam prediksi prestasi.




Gambar 5 memperlihatkan bahwa *FinalScor*e bersifat diskrit dengan massa utama pada nilai 0, 1, 2, 3, dan 4, serta puncak dominan di 4.0. Pola ini menghasilkan sebaran multimodal dan ketidakseimbangan tingkat nilai (mode pada 4.0 jauh lebih besar dibanding tingkat lain). Karakter diskrit ini mengindikasikan skema penilaian berbasis skala 5 tingkat, bukan variabel kontinu murni. Ketidakseimbangan pada tingkat tertinggi perlu diperhatikan karena dapat memengaruhi estimasi kesalahan (MAE/RMSE cenderung didorong oleh kelompok mayoritas). Untuk menjaga evaluasi yang adil, pembentukan lipatan validasi menggunakan stratifikasi berbasis tingkat/kuantil tetap penting agar proporsi level nilai terjaga pada tiap lipatan. Mengingat variabel target bersifat diskrit, prediksi kontinu dari model regresi dapat dibulatkan ke level terdekat saat digunakan secara operasional, tanpa mengubah evaluasi ilmiah yang berbasis metrik kontinu**.**




**Gambar** 6. *GPA vs FinalScore*

**JURIKOM (Jurnal Riset Komputer)**, Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9240
Hal 869-878
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Sementara pada Gambar 6 mengonfirmasi hubungan inversi antara GPA dan *FinalScore*, yang memperkuat hipotesis bahwa *FinalScore* menggunakan skala peringkat.

### 3.4 Pembahasan dan Implikasi

Dari visualisasi learning curves tujuh model regresi yang dibandingkan, CatBoost, LightGBM, Gradient Boosting, XGBoost, Random Forest, Stacking Regressor, dan MLP, dengan metrik $R^2$ pada data latih (train) dan validasi (val) di berbagai ukuran himpunan data. Secara umum, seluruh model menunjukkan pola yang konsisten: skor train menurun seiring bertambahnya data latih (gejala normal ketika model "dikekang" oleh lebih banyak data), sementara skor validasi meningkat lalu mendatar (plateau) ketika kapasitas model dan informasi data mulai jenuh. Garis putus-putus horizontal menandai ambang target $R^2 \approx 0{,}80$; tidak ada kurva validasi yang melampaui ambang ini, meskipun CatBoost dan LightGBM mendekatinya pada ukuran data terbesar.

Dari sisi stabilitas, pita bayangan (rentang variasi antarlipatan) pada CatBoost tampak paling sempit, diikuti LightGBM dan Gradient Boosting. Hal ini mengindikasikan varians yang rendah dan estimasi kinerja yang konsisten terhadap perubahan subset data. Sebaliknya, Random Forest dan terutama Stacking Regressor memperlihatkan pita yang lebih lebar, menandakan sensitivitas terhadap pemilihan lipatan (varians lebih tinggi). MLP menunjukkan kurva validasi yang meningkat perlahan dan cenderung berada di posisi terbawah, mengindikasikan bias yang relatif tinggi pada konfigurasi yang digunakan.

Secara peringkat performa validasi, CatBoost konsisten memimpin pada hampir semua ukuran data, diikuti secara ketat oleh LightGBM. Gradient Boosting berada sedikit di bawah keduanya dan mencapai plateau lebih awal, mengindikasikan kapasitasnya cukup untuk pola utama, tetapi sedikit underfit terhadap nuansa relasi nonlinier yang lebih kompleks. XGBoost cenderung menempel di belakang LightGBM namun tidak menyusul; ini biasanya terkait kebutuhan tuning lebih agresif pada parameter regularisasi dan sampling. Random Forest menunjukkan gap train–val yang lebih besar pada ukuran data kecil, yang kemudian menyempit, indikasi varians yang berangsur terkontrol ketika data bertambah. Stacking Regressor tidak melampaui model boosting terbaik; kinerjanya tertahan oleh varians meta-learner dan potensi ketidaksempurnaan out-of-fold stacking. MLP stabil tanpa tanda overfitting berat (gap kecil), tetapi kapasitas efektifnya belum memadai untuk menyaingi varian boosting pada data tabular ini.

Dari perspektif bias, varians, ujung kanan kurva (ukuran data terbesar) menunjukkan celah train–val yang kecil pada hampir semua model, terutama CatBoost. Ini menegaskan tidak terjadi overfitting signifikan di konfigurasi akhir. Namun, karena kurva validasi belum melewati $R^2 \approx 0{,}80$, masih terdapat bias sisa, yaitu model sudah cukup stabil namun belum sepenuhnya menangkap kompleksitas hubungan fitur-target. Pada konteks seperti ini, peningkatan performa lebih efektif dicapai melalui peningkatan kapasitas terkendali (misalnya learning rate lebih kecil digabung estimators lebih banyak dan early stopping) atau rekayasa fitur yang lebih informatif, daripada sekadar menambah data dengan pola yang sama.

CatBoost unggul bukan hanya pada nilai rata-rata $R^2$, tetapi juga pada kestabilan antarlipatan. Secara teoretis, hal ini selaras dengan mekanisme ordered categorical handling dan skema boosting aditif yang efektif menangkap interaksi nonlinier tanpa rekayasa fitur ekstensif. LightGBM menyusul ketat dengan kecenderungan varians sedikit lebih tinggi, umumnya dapat dikendalikan melalui kombinasi num_leaves, min_data_in_leaf, feature_fraction, serta early stopping. Gradient Boosting (sklearn) menunjukkan MAE yang kompetitif pada hasil akhir, yang konsisten dengan kurva validasinya yang relatif mulus namun sedikit lebih high-bias; performanya biasanya naik dengan learning rate lebih kecil dan jumlah pohon lebih banyak. XGBoost berpotensi mendekati LGBM melalui penalaan subsample, colsample_bytree, min_child_weight, dan early stopping.

Untuk Random Forest, pola train tinggi, val lebih rendah di ukuran kecil menandakan kecenderungan overfit ringan yang mereda ketika data bertambah. Ini bisa diperbaiki dengan menaikkan min_samples_leaf/min_samples_split, menurunkan max_depth, serta memperbesar n_estimators agar prediksi rata-rata pohon lebih stabil. Stacking Regressor memperlihatkan pita varians paling lebar; praktik terbaiknya adalah memastikan prediksi out-of-fold benar-benar digunakan sebagai masukan meta-model, mengurangi jumlah base learners agar meta-model tidak mengalami overparameterisasi, serta memilih meta-learner yang ter-regularisasi (mis. Ridge). Pada MLP, kurva mengindikasikan underfitting: kapasitas arsitektur (ukuran/lapisan) dan skema optimisasi (pensikalaan fitur, learning rate schedule, early stopping) perlu ditingkatkan. Namun, pada data tabular menengah dengan banyak fitur kategorikal, varian boosting memang cenderung lebih unggul dibanding MLP kecuali tersedia data yang jauh lebih besar atau rekayasa fitur yang sangat kaya.

Dari sisi implikasi praktis, kurva ini mendukung penetapan CatBoost sebagai baseline operasional untuk sistem peringatan dini, dengan LightGBM sebagai alternatif ringan ketika sumber daya komputasi menjadi pertimbangan. Karena indikator seperti Absences dan StudyTimeWeekly terbukti berkontribusi besar (selaras dengan analisis korelasi dan feature importance), institusi dapat merancang pemicu intervensi berbasis ambang (mis. lonjakan absensi atau jam belajar rendah) yang diintegrasikan dengan keluaran model. Untuk memperkuat klaim perbedaan performa, disarankan melaporkan uji Wilcoxon signed-rank atas skor antarlipatan antara model terbaik dan runner-up atau interval kepercayaan 95% untuk $\Delta\text{RMSE}/\Delta R^2$.

Berdasarkan analisis korelasi dan *feature importance,* dapat diidentifikasi sejumlah faktor risiko dan protektif yang memengaruhi performa akademik siswa.

Faktor Risiko utama:

a. Tingkat absensi yang tinggi

**JURIKOM (Jurnal Riset Komputer)**, Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9240
Hal 869-878
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

b. Konsentrasi *GPA* yang rendah
c. Kurangnya dukungan orang tua

Faktor protektif:

a. Kehadiran yang konsisten
b. Maintenance *GPA* yang stabil
c. Dukungan orang tua yang kuat

Temuan ini menegaskan bahwa kombinasi antara faktor akademik dan non-akademik memiliki peran penting dalam menentukan keberhasilan belajar. Meskipun demikian, terdapat beberapa keterbatasan dalam penelitian ini. Pertama, interpretasi korelasi memerlukan kehati-hatian, terutama mengingat adanya korelasi negatif yang kuat antara GPA dan *FinalScore*, yang perlu diklarifikasi lebih lanjut terkait sistem penilaian yang digunakan. Kedua, hasil penelitian ini hanya menunjukkan hubungan asosiatif, bukan kausal, sehingga diperlukan desain penelitian longitudinal untuk membangun inferensi kausal. Ketiga, generalisasi hasil penelitian terbatas karena data yang digunakan bersifat spesifik pada dataset tertentu, sehingga belum tentu dapat diterapkan pada populasi atau konteks pendidikan yang berbeda.

## 4. KESIMPULAN

Penelitian ini menunjukkan bahwa model berbasis *gradient boosting* paling efektif untuk prediksi prestasi akademik siswa. CatBoost memberi kinerja tertinggi ($R^2 = 0{,}774$), diikuti LightGBM (0,771) dan Gradient Boosting (0,767). MLP berada terendah (0,712). Analisis kepentingan fitur menegaskan GPA sebagai prediktor dominan, disusul Absences dan StudyTimeWeekly, sehingga konsistensi akademik dan kehadiran menjadi penentu utama hasil belajar. Secara praktis, temuan ini mendukung penerapan CatBoost (atau LightGBM sebagai alternatif) sebagai dasar sistem peringatan dini berbasis data untuk mengidentifikasi siswa berisiko. Model boosting memberi akurasi dan stabilitas terbaik pada data tabular pendidikan. Ke depan, studi dapat diperluas ke analisis kausal, desain intervensi yang ditautkan ke sinyal risiko, serta validasi lintas lembaga agar generalisasi meningkat.

## REFERENCES

[1] T. Nabillah and A. P. Abadi, "Faktor penyebab rendahnya hasil belajar siswa," *Pros. Semin. Nas. Mat. dan Pendidik. Mat. Sesiomadika*, vol. 2, no. 1c, pp. 659–663, 2019, [Online]. Available: http://journal.unsika.ac.id/index.php/sesiomadika
[2] D. L. Kusumaningrini and N. Sudibjo, "Faktor-Faktor Yang Mempengaruhi Motivasi Belajar Siswa Di Era Pandemi Covid-19," *Akademika*, vol. 10, no. 01, pp. 145–161, 2021, doi: 10.34005/akademika.v10i01.1271.
[3] A. Mahmudi, J. Sulianto, and I. Listyarini, "Hubungan Perhatian Orang Tua Terhadap Hasil Belajar Kognitif Siswa," *J. Pedagog. dan Pembelajaran*, vol. 3, no. 1, p. 122, 2020, doi: 10.23887/jp2.v3i1.24435.
[4] R. I. Nurachim, "Pemilihan Model Prediksi Indeks Harga Saham Yang Dikembangkan Berdasarkan Algoritma Support Vector Machine(Svm) Atau Multilayer Perceptron(Mlp) Studi Kasus : Saham Pt Telekomunikasi Indonesia Tbk," *J. Teknol. Inform. dan Komput.*, vol. 5, no. 1, pp. 29–35, 2019, doi: 10.37012/jtik.v5i1.243.
[5] B. Prasojo and E. Haryatmi, "Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 2, pp. 79–89, 2021, doi: 10.25077/teknosi.v7i2.2021.79-89.
[6] F. A. Rangkuti, S. Sundari, T. Informatika, and U. H. Medan, "Implementasi Gradient Boosting Machines Untuk Prediksi Harga Implementation Of Gradient Boosting Machines For House Price Prediction In South Jakarta," vol. 4, no. 2, pp. 164–172, 2025.
[7] N. K. C. PRATIWI, N. IBRAHIM, and S. SAIDAH, "Prediksi Kanker Paru menggunakan Grid search untuk Optimasi Hyperparameter pada Algoritma MLP dan Logistic Regression," *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 12, no. 3, p. 556, 2024, doi: 10.26760/elkomika.v12i3.556.
[8] B. Sunarko *et al.*, "Penerapan Stacking Ensemble Learning untuk Klasifikasi Efek Kesehatan Akibat Pencemaran Udara," *Edu Komputika J.*, vol. 10, no. 1, pp. 55–63, 2023, doi: 10.15294/edukomputika.v10i1.72080.
[9] T. A. E. Putri, T. Widiharih, and R. Santoso, "Penerapan Tuning Hyperparameter Randomsearchcv Pada Adaptive Boosting Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung," *J. Gaussian*, vol. 11, no. 3, pp. 397–406, 2023, doi: 10.14710/j.gauss.11.3.397-406.
[10] A. Hadianto and W. H. Utomo, "CatBoost Optimization Using Recursive Feature Elimination," vol. 9, no. 2, pp. 169–178, 2024, doi: 10.15575/join.v9i1.1324.
[11] T. S. Wibawa, N. K. Ningrum, and A. Syahreza, "Comparison of CatBoost and LightGBM Models for Air Humidity Prediction," vol. 9, no. 3, pp. 803–809, 2025.
[12] D. Kedalaman and P. Ea, "Prediksi Magnitudo Gempa Menggunakan Random Forest , Support Vector Regression , XGBoost , LightGBM , dan Multi-Layer Perceptron Berdasarkan Prediksi Magnitudo Gempa Menggunakan Random Forest , Support Vector Regression , XGBoost , LightGBM , dan Multi-La," no. December, 2024, doi: 10.52436/1.jpti.470.
[13] D. Ananda, A. Pertiwi, and M. A. Muslim, "Prediksi Rating Aplikasi Playstore Menggunakan Xgboost Prediksi Rating Aplikasi Playstore Menggunakan Xgboost," *ResearchGate*, no. October 2020, p. 6, 2022.
[14] Y. Dasril, M. Aziz, M. F. Al, and B. Prasetiyo, "Credit Risk Assessment in P2P Lending Using LightGBM and Particle Swarm Optimization," vol. 9, no. January, pp. 18–28, 2023.
[15] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data : an interdisciplinary review," *J. Big Data*, 2020, doi: 10.1186/s40537-020-00369-8.
[16] E. Fitri, "Journal Of Applied Computer Science And Technology ( JACOST ) Analisis Perbandingan Metode Regresi Linier , Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah," vol. 4, no. 1, pp. 58–64, 2023.
[17] W. Indrasari and H. Suhendar, "Analisis Model Prediksi Cuaca Menggunakan Support Vector Machine , Gradient Boosting ,

Random Forest , DAN," vol. XII, pp. 119–128, 2024.
[18] A. S. Firdan and D. Kurniadi, "Proyek Kontruksi Jalan Menggunakan Metode Backpropagation Dengan Model MLP ( Studi Kasus : Jalan Wilayah Kabupaten PekalongaN )," vol. 2, no. 4, pp. 1–12, 2025.
[19] N. O. Syamsiah and I. Purwandani, "Penerapan Ensemble Stacking untuk Peramalan Laba Bersih Bank Syariah Indonesia (BSI)," *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 295–301, 2021, doi: 10.47065/bits.v3i3.1017.
[20] A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, p. 27, 2018, doi: 10.24014/ijaidm.v1i1.4903.
[21] G. Arther Sandag, "Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest," *Cogito Smart J.* |, vol. 6, no. 2, pp. 167–178, 2020.