JURNAL FASILKOM
Volume 16 No. 2 | Agustus 2026: 437-444

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Prediksi Keberhasilan Akademik Siswa Berbasis Fitur Kategorikal *Native* dengan *Explainable* AI (SHAP) menggunakan *CatBoost vs Light*GBM

Bayu Anugerah Putra[1], Soni[2], Rahmad Firdaus[3], Ayunda Putri[4], Anisa Dwi Sanggar Wati[5]
[1,2,3,4,5]Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau
[1]bayuanugerahputra@umri.ac.id*, [2]soni@umri.ac.id, [3]rahmadfirdaus@umri.ac.id, [4]220401003@student.umri.ac.id, [5]230401161@student.umri.ac.id

Abstract

Prediction of students academic success is important to support decision-making in education. Educational datasets are generally dominated by categorical variables that require encoding before modeling, which may cause information loss and reduce accuracy. This study applies the CatBoost algorithm, which processes categorical variables natively without additional encoding, to predict students' Exam Score on the Student Performance Factors dataset from Kaggle, with LightGBM used as a comparison model. Evaluation was carried out under three data-split schemes (70:30, 80:20, 90:10) using k-fold cross-validation and three regression metrics (R^2, MAE, RMSE), followed by model interpretation using Shapley Additive Explanations (SHAP). The results show that CatBoost consistently outperforms LightGBM across all schemes, with the best performance obtained under the 90:10 scheme (CatBoost: $R^2 = 0.851$, MAE = 0.475, RMSE = 1.414; LightGBM: $R^2 = 0.809$, MAE = 0.758, RMSE = 1.599). SHAP analysis identifies Attendance, Hours_Studied, and Previous_Scores as the most influential features in the prediction. These findings confirm that combining CatBoost with SHAP produces an academic prediction model that is both accurate and transparen.

Keywords: catboost, lightGBM, academic prediction, k-fold cross-validation, SHAP

Abstrak

Prediksi keberhasilan akademik siswa menjadi penting dalam mendukung pengambilan keputusan di bidang pendidikan. Dataset pendidikan umumnya didominasi variabel kategorikal yang memerlukan proses encoding sebelum pemodelan, sehingga berpotensi menyebabkan hilangnya informasi dan menurunkan akurasi. Penelitian ini menerapkan algoritma CatBoost yang mampu memproses variabel kategorikal secara native tanpa encoding tambahan untuk memprediksi nilai ujian siswa (*Exam Score*) pada dataset *Student Performance Factors* dari Kaggle, dengan LightGBM sebagai model pembanding. Evaluasi dilakukan pada tiga skema pembagian data (70:30, 80:20, 90:10) menggunakan *k-fold cross-validation* dan tiga metrik regresi (R^2, MAE, RMSE), dilanjutkan dengan interpretasi model menggunakan *Shapley Additive Explanations* (SHAP). Hasil pengujian menunjukkan bahwa CatBoost secara konsisten mengungguli LightGBM pada seluruh skema, dengan performa terbaik pada skema 90:10 (CatBoost: R^2 = 0,851, MAE = 0,475, RMSE = 1,414; LightGBM: R^2 = 0,809, MAE = 0,758, RMSE = 1,599). Analisis SHAP mengidentifikasi *Attendance, Hours_Studied,* dan *Previous_Scores* sebagai fitur paling berpengaruh terhadap prediksi. Temuan ini menegaskan bahwa kombinasi CatBoost dan SHAP mampu menghasilkan model prediksi akademik yang akurat sekaligus transparan.

Kata kunci: catboost, lightGBM, prediksi akademik, *k-fold cross-validation*, SHAP

1. Pendahuluan

Keberhasilan siswa merupakan hasil yang bersifat multidimensional, dipengaruhi oleh aspek akademik, perilaku individu, lingkungan belajar, serta kondisi sosial di sekitarnya [1], [2]. Perkembangan teknologi informasi memberikan peluang besar dalam pengolahan dan analisis data pendidikan, sehingga pengambilan keputusan berbasis data menjadi semakin penting untuk memahami kebutuhan siswa dan merancang strategi pembelajaran yang efektif [3]. Salah satu tantangan utama lembaga pendidikan adalah mengidentifikasi faktor-faktor yang memengaruhi capaian belajar siswa serta memprediksinya secara akurat, dan pendekatan berbasis *machine learning* terbukti mampu membantu mengenali pola tersebut secara lebih komprehensif.

Penelitian oleh [1] menerapkan algoritma *Light Gradient Boosting Machine* (LightGBM) yang didukung *Explainable Artificial Intelligence* (XAI) berbasis *Shapley Additive Explanations* (SHAP) untuk memprediksi keberhasilan akademik siswa, dan menunjukkan bahwa model tersebut mampu menghasilkan prediksi dengan akurasi tinggi (R^2 = 0,729, MAE = 0,804, RMSE = 1,957) yang dapat dijelaskan secara transparan. Namun demikian, *LightGBM* memiliki keterbatasan dalam menangani fitur kategorikal: pada proses pelatihannya, *LightGBM* mengonversi fitur kategorikal ke representasi statistik gradien dan mengelompokkan kategori berdasarkan kemiripan bobot sebagai strategi penghematan memori [4], sehingga berpotensi mengurangi kedetailan informasi dan menambah beban komputasi. Hal ini relevan pada dataset pendidikan yang umumnya memiliki banyak variabel kategorikal dengan heterogenitas tinggi, dan sejalan dengan temuan [1] bahwa distribusi prediksi *LightGBM* cenderung terkonsentrasi pada rentang nilai menengah sehingga belum optimal merepresentasikan variasi nilai ujian, terutama pada skor tinggi.

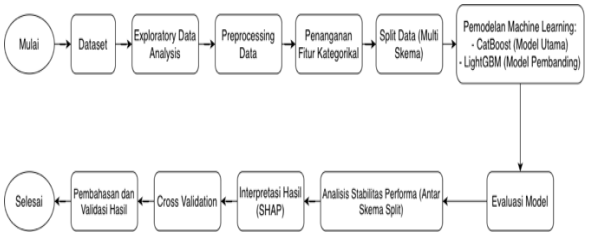
P-ISSN : 2089-3353
E-ISSN : 2808-9162

Sebagai bentuk pengembangan untuk mengatasi keterbatasan tersebut, penelitian ini menggunakan algoritma *CatBoost* (*Categorical Boosting*) yang dirancang khusus untuk memproses variabel kategorikal secara native tanpa *encoding* tambahan [5]. CatBoost menerapkan teknik *Ordered Boosting* yang mengurangi bias *overfitting* dan meningkatkan efisiensi serta akurasi model [6], serta terbukti efektif digunakan pada data pendidikan dengan banyak fitur kategorikal [7]. Keunggulan CatBoost dibandingkan *LightGBM* diperkuat oleh berbagai temuan empiris; Mawardi et al. [8] melaporkan akurasi *CatBoost* sebesar 97,37% dibandingkan *LightGBM* 89,47%, sementara Wibawa et al. [9] melaporkan *CatBoost* unggul dengan R^2 = 0,8191 dibandingkan *LightGBM* R^2 = 0,7981 pada tugas regresi lain.

Berdasarkan uraian tersebut, penelitian ini bertujuan (1) menerapkan algoritma *CatBoost* untuk membangun model prediksi keberhasilan akademik siswa pada dataset *Student Performance Factors*, (2) mengevaluasi kinerja *CatBoost* dan membandingkannya dengan *LightGBM* menggunakan skema pembagian data dan *k-fold cross-validation*, serta (3) menerapkan metode SHAP untuk menjelaskan hasil prediksi dan mengidentifikasi fitur-fitur yang paling berpengaruh terhadap keberhasilan akademik siswa. Penelitian ini diposisikan sebagai pengembangan dari studi Santana-Perera et al. [1] melalui pembaruan algoritma prediksi, tanpa mengubah kerangka interpretasi model berbasis SHAP.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan analisis komparatif terhadap performa algoritma prediksi berbasis *machine learning*. Tahapan penelitian meliputi pengumpulan data, *exploratory data analysis, preprocessing,* pembagian data dan *cross-validation*, pemodelan, evaluasi, serta interpretasi model menggunakan SHAP, sebagaimana diilustrasikan pada Gambar 1.



Gambar 1. Metode Penelitian

2.1. Dataset

Dataset yang digunakan adalah *Student Performance Factors* yang bersumber dari *platform Kaggle*, terdiri dari 6.607 data siswa dengan 20 variabel (19 variabel input dan 1 variabel target *Exam_Score*). Dari 19 variabel input, terdapat 6 variabel numerik (*Hours_Studied, Attendance, Sleep_Hours, Previous_Scores, Tutoring_Sessions, Physical_Activity*) dan 13 variabel kategorikal (di antaranya *Parental_Involvement, Access_to_Resources, Motivation_Level, Family_Income, Teacher_Quality, Peer_Influence, dan Distance_from_Home*). Dataset ini merupakan data sintetis yang dikembangkan untuk keperluan penelitian dan simulasi analisis pendidikan, sehingga hasil penelitian ini diposisikan sebagai evaluasi kinerja algoritma dalam konteks dataset sintetis.

2.2. Preprocessing Data

Pemeriksaan kelengkapan data menunjukkan missing value pada tiga variabel kategorikal, yaitu *Teacher_Quality* (78 data), *Parental_Education_Level* (90 data), dan *Distance_from_Home* (67 data), yang ditangani menggunakan imputasi modus (*most frequent*). Penanganan fitur kategorikal disesuaikan dengan karakteristik masing-masing algoritma: pada CatBoost, fitur kategorikal diproses langsung tanpa *encoding* manual dengan tetap didefinisikan sebagai variabel kategorikal [10], sedangkan pada LightGBM fitur kategorikal di-*encode* (*Label Encoding*) sebagai bagian dari *preprocessing* [11]. Tidak dilakukan *feature engineering* tambahan karena penelitian berfokus pada evaluasi kinerja dan stabilitas model.

2.3. Split Data dan Cross-Validation

Data dibagi menggunakan tiga skema proporsi, yaitu 70:30, 80:20, dan 90:10, untuk menganalisis stabilitas (*robustness*) performa model terhadap variasi proporsi data latih dan data uji. Pada setiap skema, *k-fold cross-validation* diterapkan pada data latih untuk mengurangi bias akibat satu kali pembagian data, sementara data uji tetap terpisah dari proses pelatihan [12].

2.4. Pemodelan dan Hyperparameter

Model utama yang digunakan adalah CatBoost [13], dengan LightGBM sebagai model pembanding yang mengacu pada prosedur penelitian terdahulu [1], [14]. Sebagai *baseline* sederhana tambahan, digunakan hasil regresi linear dari penelitian Aprilia et al. [3] (MSE = 3,256) tanpa diimplementasikan ulang. Konfigurasi *hyperparameter* kedua model dirancang agar sebanding secara konseptual, sebagaimana disajikan pada Tabel 1.

Tabel 1. Konfigurasi *Hyperparameter Model*

Parameter	*CatBoost*	*LightGBM*
Iterations / n_estimators	500	500
Learning Rate	0,05	0,05
Depth / Max Depth	6	6
Random State	42	42
Verbose	0	-

Parameter *iterations* pada *CatBoost* dan n_*estimators* pada *LightGBM* memiliki fungsi setara, yaitu mengatur jumlah *boosting rounds*. Parameter *learning_rate* mengatur laju pembelajaran pada kedua model, sedangkan *depth* (*CatBoost*) dan max_depth (*LightGBM*) mengontrol kompleksitas pohon keputusan. Parameter

P-ISSN : 2089-3353
E-ISSN : 2808-9162

random_state diterapkan untuk menjaga reprodusibilitas hasil eksperimen.

2.5. Evaluasi Model

Performa model dievaluasi menggunakan tiga metrik regresi: R-*Squared* (R^2) yang mengukur seberapa besar variabilitas data aktual dapat dijelaskan oleh model; *Root Mean Square Error* (RMSE) yang menghitung akar rata-rata kuadrat selisih antara nilai prediksi dan aktual serta sensitif terhadap *outlier*; dan *Mean Absolute Error* (MAE) yang menghitung rata-rata kesalahan absolut tanpa memperbesar pengaruh *outlier* [15].

2.6. Analisis Interpretabilitas dengan SHAP

Interpretabilitas model dianalisis menggunakan *Shapley Additive Explanations* (SHAP) sebagai pendekatan *Explainable* AI, yang mengukur kontribusi setiap fitur input terhadap hasil prediksi dengan membandingkan perubahan prediksi ketika suatu fitur disertakan atau dikeluarkan dari model [16]. Interpretasi dilakukan secara global menggunakan SHAP *summary plot* untuk mengidentifikasi fitur paling berpengaruh secara keseluruhan, dan secara lokal menggunakan SHAP *waterfall plot* untuk menjelaskan kontribusi fitur pada prediksi individu (siswa dengan skor rendah, sedang, dan tinggi). Analisis interpretabilitas difokuskan pada model *CatBoost* sebagai model utama penelitian

3. Hasil dan Pembahasan

Bagian ini menyajikan hasil eksperimen yang diperoleh dari proses pemodelan dan evaluasi secara bertahap. Evaluasi diawali dengan pengujian performa model pada skema *baseline* tanpa *k-fold cross-validation*, sebagai acuan perbandingan awal terhadap hasil yang diperoleh setelah penerapan *k-fold cross-validation*. Selanjutnya, evaluasi dilakukan menggunakan *k-fold cross-validation* pada beberapa skema pembagian data guna memperoleh estimasi performa model yang lebih stabil dan representatif. Setelah diperoleh skema pembagian data terbaik, dilakukan analisis lebih lanjut terhadap kemampuan prediksi model melalui perbandingan nilai aktual dan prediksi, serta interpretasi hasil model menggunakan SHAP untuk mengidentifikasi fitur-fitur yang paling berpengaruh terhadap prediksi *Exam Score*.

3.1 Evaluasi Model Tanpa *K-Fold Cross-Validation*

Sebagai *baseline* perbandingan awal yang setara dengan skema penelitian terdahulu [1], kedua model dievaluasi pada skema pembagian data 80:20 tanpa *cross-validation*. Hasil evaluasi disajikan pada Tabel 2.

Tabel 2. Hasil Evaluasi Model Tanpa *K-Fold Cross-Validation*

Model	R^2	MAE	RMSE
CatBoost	0,755	0,595	1,857
LightGBM	0,724	0,815	1,974

CatBoost menunjukkan performa lebih baik dibandingkan LightGBM pada seluruh metrik, dengan nilai kesalahan prediksi yang lebih rendah dan R^2 yang lebih tinggi. Hasil baseline ini menjadi dasar evaluasi lanjutan menggunakan *k-fold cross-validation* pada berbagai skema pembagian data.

3.2. Evaluasi Model dengan *K-Fold Cross-Validation*

Evaluasi dengan *k-fold cross-validation* dilakukan pada tiga skema pembagian data untuk memperoleh estimasi performa yang lebih stabil dan representatif. Ringkasan hasil pada ketiga skema disajikan pada Tabel 3.

Tabel 3. Hasil Performa Model pada Berbagai Skema Pembagian Data (dengan *K-Fold Cross-Validation*)

Split Data	Model	R^2	MAE	RMSE
70:30	*CatBoost*	0,765	0,564	1,795
70:30	*LightGBM*	0,721	0,853	1,955
80:20	*CatBoost*	0,755	0,595	1,857
80:20	*LightGBM*	0,724	0,815	1,974
90:10	*CatBoost*	0,851	0,475	1,414
90:10	*LightGBM*	0,809	0,758	1,599

Berdasarkan Tabel 3, CatBoost secara konsisten menghasilkan nilai R^2 yang lebih tinggi serta MAE dan RMSE yang lebih rendah dibandingkan LightGBM pada seluruh skema pembagian data. Performa kedua model relatif stabil pada skema 70:30 dan 80:20, sedangkan peningkatan performa paling besar diperoleh pada skema 90:10. Pada skema tersebut, CatBoost mencapai R^2 = 0,851, MAE = 0,475, dan RMSE = 1,414, sedangkan LightGBM mencapai R^2 = 0,809, MAE = 0,758, dan RMSE = 1,599. Dibandingkan skema 80:20, penggunaan skema 90:10 meningkatkan R^2 CatBoost sebesar 0,096 serta menurunkan MAE dan RMSE masing-masing sebesar 20,2% dan 23,9%. Pada LightGBM, peningkatan R^2 mencapai 0,085, sedangkan MAE dan RMSE menurun masing-masing sebesar 7,0% dan 19,0%. Arah perbaikan yang sama pada dua model dan tiga metrik menunjukkan bahwa pemilihan skema 90:10 tidak hanya didasarkan pada satu indikator evaluasi.

Pemilihan skema 90:10 juga mempertimbangkan karakteristik dataset yang terdiri atas 6.607 observasi. Proporsi tersebut menyediakan sekitar 5.946 data latih sehingga model memperoleh lebih banyak variasi untuk mempelajari hubungan antara fitur numerik, kategori yang heterogen, dan Exam Score, sementara sekitar 661 observasi tetap dipertahankan sebagai data uji yang tidak dilibatkan dalam pelatihan. Ketersediaan data latih yang lebih besar relevan bagi CatBoost karena mekanisme ordered boosting dan pengolahan fitur kategorikalnya memanfaatkan variasi observasi untuk mengurangi prediction shift dan bias akibat target leakage. Selain itu, penerapan 5-fold cross-validation pada data latih membantu mengurangi ketergantungan hasil terhadap satu pembagian data. Oleh karena itu, skema 90:10 ditetapkan sebagai konfigurasi dengan performa terbaik dalam ruang lingkup eksperimen ini, bukan sebagai proporsi yang selalu optimal untuk seluruh

P-ISSN : 2089-3353
E-ISSN : 2808-9162

dataset. Mengingat data yang digunakan bersifat sintetis dan data uji pada skema 90:10 lebih kecil daripada dua skema lainnya, generalisasi hasil tetap perlu dikonfirmasi menggunakan data eksternal atau pembagian acak berulang pada penelitian berikutnya.

3.3. Perbandingan Nilai Aktual dan Prediksi

Untuk mengilustrasikan kemampuan model, dibandingkan nilai aktual dan prediksi pada tiga kategori siswa (skor rendah, sedang, tinggi) menggunakan model dari skema terbaik. Hasilnya disajikan pada Tabel 4.

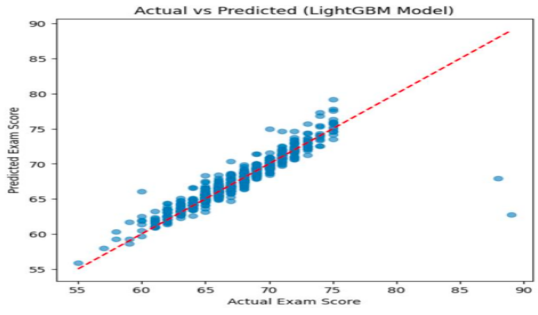
Tabel 4. Perbandingan Nilai Aktual dan Prediksi Model

Kategori Skor	Nilai Aktual	Prediksi *CatBoost*	Prediksi *LightGBM*
Rendah	65	64,71	65,18
Sedang	67	67,17	66,49
Tinggi	74	74,37	75,30

Kedua model mampu menghasilkan prediksi yang mendekati nilai aktual pada ketiga kategori skor, dengan selisih yang relatif kecil pada masing-masing kasus. Hasil ini memperkuat temuan pada evaluasi metrik regresi bahwa kedua model, terutama CatBoost, mampu mempelajari pola hubungan antara fitur input dan nilai ujian siswa dengan baik.

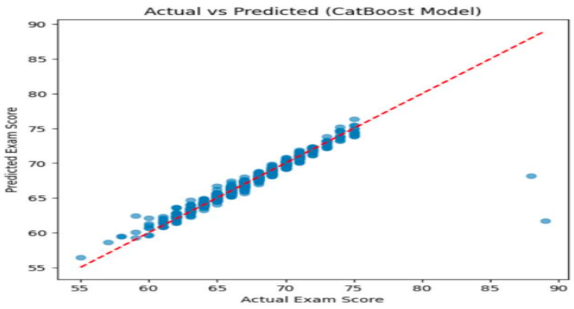
3.4 Analisis Residual dan Interpretabilitas Model

Analisis residual dilakukan untuk memeriksa pola kesalahan prediksi model secara visual melalui scatter plot nilai aktual terhadap nilai prediksi, distribusi residual error, serta rata-rata *error* berdasarkan rentang skor. Selain menunjukkan tingkat kedekatan prediksi terhadap nilai aktual, analisis ini digunakan untuk mengidentifikasi rentang nilai tertentu yang masih sulit diprediksi oleh kedua model.



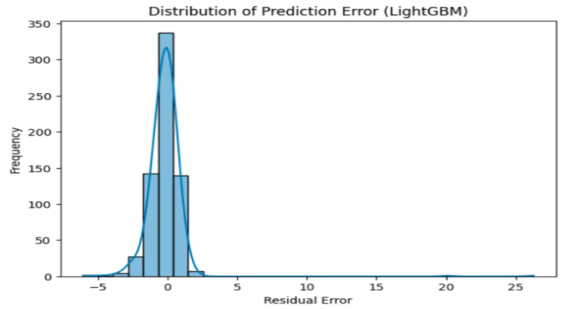
Gambar 2. Aktual dan Prediksi (*LightGBM* Model)

Scatter plot LightGBM pada Gambar 2 menunjukkan bahwa sebagian besar titik pada rentang nilai aktual sekitar 55–75 berada di sekitar garis diagonal. Kondisi tersebut mengindikasikan bahwa prediksi *LightGBM* relatif mendekati nilai aktual pada rentang skor yang paling banyak ditemukan dalam dataset. Meskipun demikian, terdapat beberapa titik yang menyimpang cukup jauh, terutama pada skor aktual yang tinggi, sehingga menunjukkan bahwa model belum sepenuhnya mampu merepresentasikan observasi ekstrem. Hasil ini merupakan performa empiris pada konfigurasi penelitian ini dan tidak menghilangkan keunggulan efisiensi *LightGBM* melalui mekanisme *Gradient-based One-Side Sampling* dan *Exclusive Feature Bundling*.



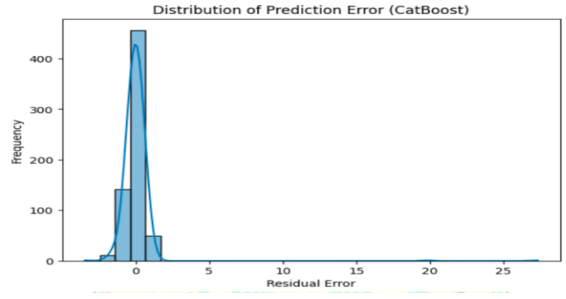
Gambar 3. Aktual dan Prediksi (*CatBoost* Model)

Pada *CatBoost*, sebagian besar titik juga mengikuti garis diagonal, tetapi sebarannya terlihat lebih rapat pada rentang skor yang dominan. Dapat dilihat pada Gambar 3, kedekatan titik terhadap garis diagonal menunjukkan bahwa selisih antara nilai aktual dan prediksi *CatBoost* cenderung lebih kecil dibandingkan *LightGBM*. Namun, beberapa penyimpangan masih muncul pada skor tinggi. Dengan demikian, *CatBoost* meningkatkan akurasi secara keseluruhan, tetapi belum sepenuhnya menghilangkan kesalahan pada observasi yang berada di bagian ekstrem distribusi.



Gambar 4. Distribusi Prediksi *Error* (*LightGBM*)

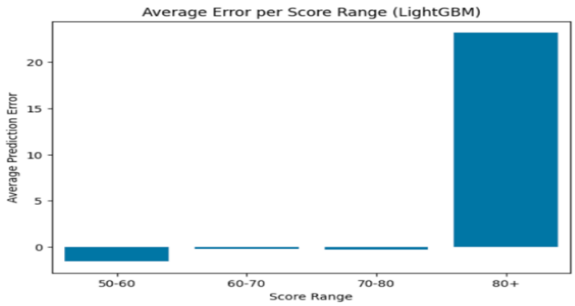
Gambar 4 menyajikan distribusi residual model LightGBM. Sebagian besar residual terkonsentrasi di sekitar nilai nol, yang menunjukkan bahwa mayoritas hasil prediksi memiliki selisih relatif kecil terhadap nilai aktual. Namun, distribusi residual masih cukup menyebar dan menunjukkan beberapa nilai ekstrem, terutama pada sisi positif. Kondisi tersebut menandakan bahwa pada sejumlah kecil observasi, *LightGBM* masih menghasilkan kesalahan prediksi yang cukup besar.



Gambar 5. Distribusi Prediksi *Error* (*CatBoost*)

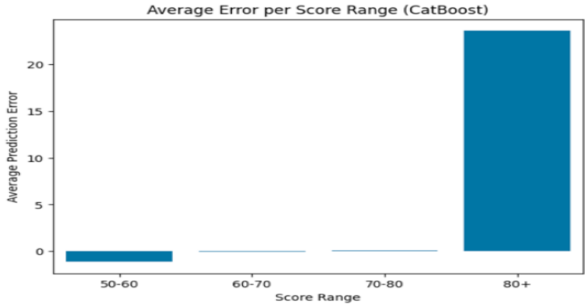
P-ISSN : 2089-3353
E-ISSN : 2808-9162

Gambar 5 memperlihatkan distribusi residual model *CatBoost* yang lebih tajam dan terkonsentrasi di sekitar nilai nol. Pola ini menunjukkan bahwa kesalahan prediksi CatBoost pada sebagian besar observasi cenderung lebih kecil dan lebih konsisten dibandingkan *LightGBM*. Temuan tersebut selaras dengan nilai *MAE* dan *RMSE CatBoost* yang lebih rendah. Meskipun masih ditemukan residual ekstrem, jumlahnya relatif terbatas sehingga tidak menggambarkan pola kesalahan model secara keseluruhan. Selanjutnya, residual dianalisis berdasarkan kelompok skor untuk mengetahui rentang nilai yang memiliki kesalahan prediksi paling besar.



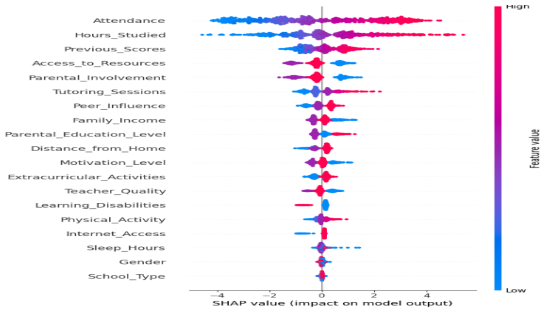
Gambar 6. Visualisasi *Error* Rentang Skor (*LightGBM*)

Pada *LightGBM*, rata-rata *error* relatif kecil pada rentang skor 60–70 dan 70–80, sedangkan besarnya error meningkat secara mencolok pada skor di atas 80. Berdasarkan Gambar 6, perbedaan tanda *error* antara kelompok 50–60 dan kelompok di atas 80 juga menunjukkan adanya arah bias prediksi yang berbeda pada kedua ujung distribusi. Hal ini mengindikasikan bahwa *LightGBM* bekerja lebih baik pada rentang skor yang memiliki lebih banyak observasi, tetapi mengalami kesulitan ketika memprediksi nilai yang jarang atau ekstrem.



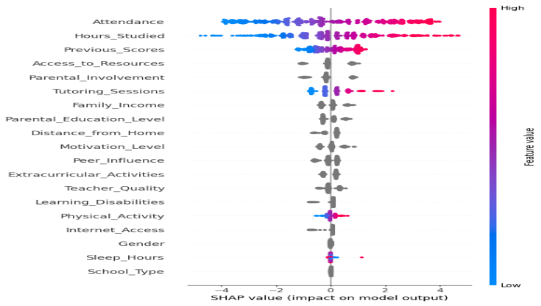
Gambar 7. Visualisasi *Error* Rentang Skor (*CatBoost*)

CatBoost memperlihatkan pola yang serupa, yaitu error yang relatif kecil pada rentang skor rendah hingga menengah dan peningkatan *error* pada kelompok skor di atas 80 sebagaimana yang dapat dilihat pada Gambar 7. Meskipun performa agregat *CatBoost* lebih baik, perbedaan *error* kedua model pada kelompok skor tertinggi relatif kecil. Temuan ini menunjukkan bahwa keterbatasan prediksi skor ekstrem kemungkinan tidak hanya dipengaruhi oleh pemilihan algoritma, tetapi juga oleh sedikitnya representasi observasi dengan skor sangat tinggi dalam dataset.



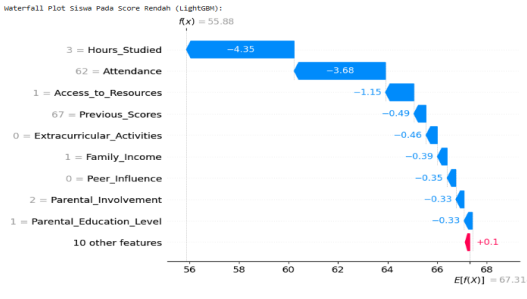
Gambar 8. SHAP *Summary Plot* (*LightGBM*)

SHAP summary plot LightGBM pada Gambar 8 menunjukkan bahwa *Attendance, Hours_Studied*, dan *Previous_Scores* memiliki sebaran nilai *SHAP* paling luas sehingga menjadi fitur yang paling berpengaruh terhadap prediksi. Nilai yang tinggi pada ketiga fitur tersebut umumnya mendorong prediksi ke arah yang lebih tinggi, sedangkan nilai yang rendah menurunkannya. Sementara itu, fitur seperti *School_Type, Gender, Internet_Access,* dan *Learning_Disabilities* memiliki nilai *SHAP* yang lebih terkonsentrasi di sekitar nol, sehingga kontribusinya terhadap perubahan prediksi relatif lebih kecil.



Gambar 9. SHAP *Summary Plot* (*CatBoost*)

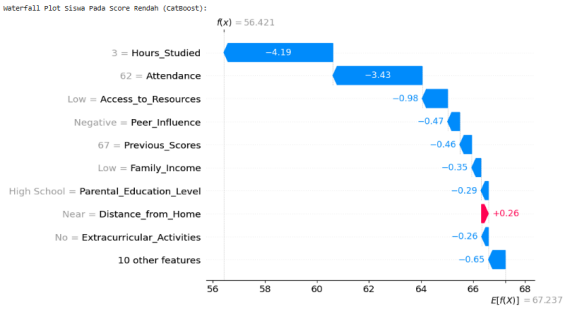
CatBoost menghasilkan urutan fitur dominan yang relatif konsisten dengan *LightGBM*, terutama pada *Attendance, Hours_Studied,* dan *Previous_Scores* sebagaimana yang dapat dilihat pada Gambar 9. Konsistensi tersebut memperkuat indikasi bahwa ketiga fitur merupakan pola utama dalam data dan bukan hanya karakteristik yang muncul pada satu algoritma. Meskipun demikian, nilai *SHAP* menjelaskan kontribusi fitur terhadap keluaran model dan tidak dengan sendirinya membuktikan hubungan sebab-akibat. Interpretasi ini sesuai dengan prinsip *SHAP* sebagai metode atribusi aditif yang membagi perbedaan antara nilai baseline dan hasil prediksi kepada setiap fitur.



Gambar 10. SHAP *Waterfall Plot Model LightGBM* – Siswa Skor Rendah

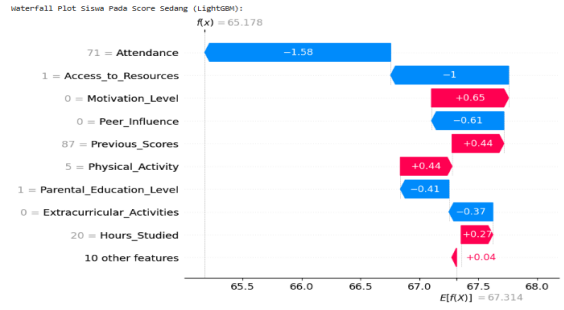
P-ISSN : 2089-3353
E-ISSN : 2808-9162

Pada Gambar 10, *waterfall plot LightGBM* untuk siswa dengan skor rendah memperlihatkan bahwa sejumlah fitur memberikan kontribusi negatif sehingga menurunkan prediksi dari nilai *baseline*. Kontribusi negatif terbesar berasal dari *Hours_Studied* dan *Attendance* yang rendah. Sementara itu, kontribusi positif dari fitur lainnya belum cukup besar untuk mengimbangi pengaruh negatif tersebut. Akumulasi kontribusi tersebut menyebabkan model menghasilkan prediksi pada kategori skor rendah.



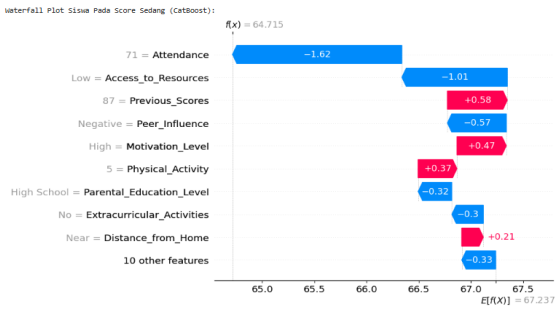
Gambar 11. SHAP *Waterfall Plot Model CatBoost* – Siswa Skor Rendah

Waterfall plot CatBoost pada Gambar 11 juga menunjukkan bahwa *Hours_Studied* dan *Attendance* yang rendah menjadi faktor utama yang menurunkan prediksi dari nilai *baseline*. Arah kontribusi fitur pada *CatBoost* konsisten dengan *LightGBM*, meskipun besarnya kontribusi setiap fitur berbeda karena kedua model memiliki struktur dan mekanisme pembelajaran yang berbeda. Dengan demikian, kedua model menangkap pola yang serupa dalam memprediksi siswa dengan skor rendah.



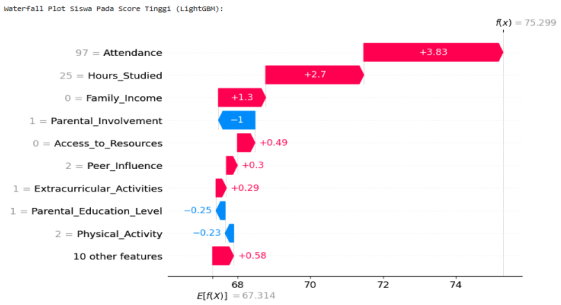
Gambar 12. SHAP *Waterfall Plot Model LightGBM* – Siswa Skor Sedang

Pada Gambar 12, *waterfall plot LightGBM* untuk siswa dengan skor sedang menunjukkan bahwa kontribusi positif dan negatif relatif lebih berimbang dibandingkan dengan siswa pada kategori skor rendah. Kontribusi negatif dari *Attendance* dan *Access_to_Resources* diimbangi oleh kontribusi positif dari *Motivation_Level, Previous_Scores, Physical_Activity,* dan *Hours_Studied*. Meskipun jumlah kontribusi negatif masih lebih besar sehingga prediksi berada sedikit di bawah nilai *baseline*, hasil akhirnya tetap berada pada kategori skor sedang.



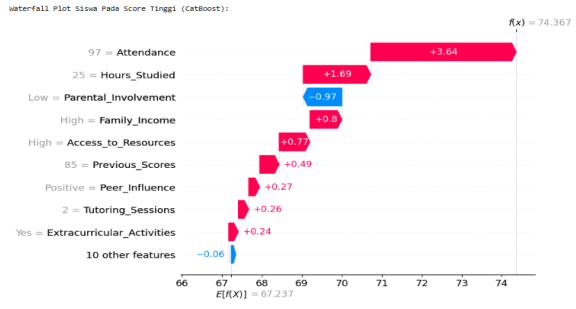
Gambar 13. SHAP *Waterfall Plot Model* CatBoost – Siswa Skor Sedang

Berdasarkan Gambar 13, *waterfall plot CatBoost* untuk siswa dengan skor sedang memperlihatkan bahwa *Attendance* dan *Access_to_Resources* memberikan kontribusi negatif terbesar. Pengaruh tersebut sebagian diimbangi oleh kontribusi positif dari *Previous_Scores, Motivation_Level,* dan *Distance_from_Home.* Kombinasi kontribusi sejumlah fitur tersebut menghasilkan prediksi yang berada sedikit di bawah nilai baseline, tetapi masih termasuk dalam kategori skor sedang.



Gambar 14. SHAP *Waterfall Plot Model LightGBM* – Siswa Skor Tinggi

Pada Gambar 14, waterfall plot *LightGBM* untuk siswa dengan skor tinggi menunjukkan bahwa *Attendance* dan *Hours_Studied* memberikan kontribusi positif yang dominan. *Family_Income* dan beberapa fitur lainnya turut meningkatkan prediksi hingga melampaui nilai baseline. Meskipun terdapat kontribusi negatif dari *Parental_Involvement, Parental_Education_Level,* dan *Physical_Activity*, pengaruhnya tidak cukup besar untuk mengimbangi kontribusi positif. Dengan demikian, tingkat kehadiran dan waktu belajar yang tinggi menjadi faktor utama yang menjelaskan prediksi pada kategori skor tinggi.



Gambar 15. SHAP *Waterfall Plot Model CatBoost* – Siswa Skor Tinggi

P-ISSN : 2089-3353
E-ISSN : 2808-9162

CatBoost pada Gambar 15 memperlihatkan bahwa *Attendance* dan *Hours_Studied* yang tinggi memberikan kontribusi positif terbesar sehingga meningkatkan prediksi dari nilai *baseline*. *Family_Income, Access_to_Resources,* dan *Previous_Scores* turut memberikan kontribusi positif, sedangkan Parental_Involvement memberikan kontribusi negatif. Namun, kontribusi negatif tersebut tidak cukup besar untuk mengimbangi akumulasi kontribusi positif sehingga hasil prediksi berada pada kategori skor tinggi. Pola ini konsisten dengan kategori siswa sebelumnya, yaitu nilai *Attendance* dan *Hours_Studied* yang rendah menurunkan prediksi, sedangkan nilai yang tinggi meningkatkan prediksi. Karena dataset penelitian ini bersifat sintetis, hasil SHAP harus dipahami sebagai pola asosiasi yang dipelajari model dan bukan sebagai bukti hubungan kausal langsung.

Secara keseluruhan, analisis global dan individual menunjukkan bahwa *CatBoost* tidak hanya menghasilkan kesalahan prediksi yang lebih rendah, tetapi juga memberikan pola interpretasi yang relatif konsisten dengan *LightGBM*. *Attendance* dan *Hours_Studied* menjadi fitur yang paling dominan dalam membedakan prediksi pada kategori skor rendah, sedang, dan tinggi, sedangkan *Previous_Scores* dan fitur lainnya memberikan kontribusi tambahan yang besarnya bergantung pada karakteristik setiap siswa. Penerapan SHAP membuat proses prediksi lebih transparan, meskipun validasi menggunakan data pendidikan nyata tetap diperlukan sebelum model digunakan untuk mendukung pengambilan keputusan.

4. Kesimpulan

Berdasarkan hasil implementasi dan analisis yang telah dilakukan, dapat disimpulkan bahwa: (1) algoritma *CatBoost* secara konsisten menghasilkan performa prediksi yang lebih baik dibandingkan *LightGBM* pada seluruh skema pembagian data yang diuji, dengan evaluasi *k-fold cross-validation* yang menunjukkan performa relatif konsisten antar fold sehingga tidak ditemukan indikasi *overfitting* yang signifikan; (2) skema pembagian data 90:10 menghasilkan performa terbaik pada kedua model dan menjadi konfigurasi paling optimal, dengan CatBoost mencapai R^2 = 0,851, MAE = 0,475, dan RMSE = 1,414, sedangkan *LightGBM* mencapai R^2 = 0,809, MAE = 0,758, dan RMSE = 1,599; serta (3) analisis interpretabilitas menggunakan SHAP menunjukkan bahwa *Attendance, Hours_Studied,* dan *Previous_Scores* merupakan fitur yang paling berpengaruh terhadap prediksi nilai ujian siswa, dengan kontribusi dominan dalam menentukan hasil prediksi model.

Kombinasi *CatBoost* dan *SHAP* terbukti mampu menghasilkan model prediksi akademik yang tidak hanya akurat, tetapi juga transparan dan dapat dijelaskan, sehingga berpotensi membantu lembaga pendidikan dalam merancang strategi pembelajaran dan intervensi akademik yang lebih tepat sasaran. Penelitian selanjutnya disarankan mengeksplorasi pendekatan deep learning serta teknik optimasi *hyperparameter* untuk meningkatkan kinerja prediksi, dan memperluas analisis interpretabilitas dengan metode *Explainable* AI tambahan guna memperoleh pemahaman yang lebih komprehensif mengenai perilaku model.

Daftar Rujukan

[1] B. Santana-Perera, C. García-Barceló, M. González Arcas, and D. Gil, "Exploring Predictive Insights on Student Success Using Explainable Machine Learning: A Synthetic Data Study," *Information*, vol. 16, no. 9, p. 763, Sep. 2025, doi: 10.3390/info16090763.

[2] M. Credé, S. G. Roch, and U. M. Kieszczynka, "Class Attendance in College," *Rev. Educ. Res.*, vol. 80, no. 2, pp. 272–295, Jun. 2010, doi: 10.3102/0034654310362998.

[3] F. Aprilia, R. A. Anggraini, and Y. D. Putri, "Prediksi Kelulusan Siswa dengan Algoritma Pembelajaran Mesin: Aplikasi Regresi Linear dan Logistik pada Faktor-Faktor Pendidikan," *ROUTERS: Jurnal Sistem dan Teknologi Informasi*, pp. 55–64, Feb. 2025, doi: 10.25181/rt.v3i1.3897.

[4] A. Joshi, P. Saggar, R. Jain, M. Sharma, D. Gupta, and A. Khanna, "CatBoost — An Ensemble Machine Learning Model for Prediction and Classification of Student Academic Performance," *Advances in Data Science and Adaptive Analysis*, vol. 13, no. 03n04, Jul. 2021, doi: 10.1142/S2424922X21410023.

[5] Zulwisli, Ambiyar, Muhammad Anwar, and Andhika Herayono, "Student's Digital Intentions Prediction Using CatBoost," *JTP - Jurnal Teknologi Pendidikan*, vol. 27, no. 1, pp. 277–290, Apr. 2025, doi: 10.21009/jtp.v27i1.54035.

[6] N. Levi Sabili, F. Rakhmat Umbara, and M. Melina, "KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA CATEGORICAL BOOSTING DENGAN FAKTOR RISIKO DIABETES," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 6, pp. 11391–11398, Nov. 2024, doi: 10.36040/jati.v8i6.11447.

[7] M. T. Syamkalla, S. Khomsah, and Y. S. R. Nur, "Implementasi Algoritma Catboost Dan Shapley Additive Explanations (SHAP) Dalam Memprediksi Popularitas Game Indie Pada Platform Steam," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 4, pp. 777–786, Aug. 2024, doi: 10.25126/jtiik.1148503.

[8] A. B. Mawardi, R. S. Pradini, and M. S. Haris, "KOMPARASI ALGORITMA BOOSTING UNTUK PREDIKSI GANGGUAN TIDUR," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3, Jul. 2025, doi: 10.23960/jitet.v13i3.7281.

[9] Tangkas Surya Wibawa, N. K. Ningrum, and Ahmad Syahreza, "Comparison of CatBoost and LightGBM Models for Air Humidity Prediction," *Journal of Applied Informatics and Computing*, vol. 9, no. 3, pp. 803–809, Jun. 2025, doi: 10.30871/jaic.v9i3.9570.

[10] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *J. Big Data*, vol. 7, no. 1, p. 94, Dec. 2020, doi: 10.1186/s40537-020-00369-8.

[11] T. Gori, A. Sunyoto, and H. Al Fatta, "Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 215–224, Feb. 2024, doi: 10.25126/jtiik.20241118074.

[12] S. Bates, T. Hastie, and R. Tibshirani, "Cross-Validation: What Does It Estimate and How Well Does It Do It?," *J. Am. Stat. Assoc.*, vol. 119, no. 546, pp. 1434–1445, Apr. 2024, doi: 10.1080/01621459.2023.2197686.

[13] T. Z. Jasman, M. A. Fadhlullah, A. L. Pratama, and R. Rismayani, "Analisis Algoritma Gradient Boosting, Adaboost dan Catboost dalam Klasifikasi Kualitas Air," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 2, Aug. 2022, doi: 10.28932/jutisi.v8i2.4906.

P-ISSN : 2089-3353
E-ISSN : 2808-9162

[14] Z. Fan, J. Gou, and S. Weng, "A Feature Importance-Based Multi-Layer CatBoost for Student Performance Prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 11, pp. 5495–5507, Nov. 2024, doi: 10.1109/TKDE.2024.3393472.

[15] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.

[16] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nat. Mach. Intell.*, vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: 10.1038/s42256-019-0138-9.