Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 249-259
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT
Penerbit : Cv. Utility Project Solutio

Optimasi Klasifikasi Data Tabular Menggunakan Ensemble Learning dan Explainable Artificial Intelligence

Fricles Ariwisanto Sianturi

Sistem Informasi, Universitas Tjut Nyak Dhien, Medan, Sumatera Utara, Indonesia

Email: sianturifricles@utnd.ac.id

Info Artikel

Kata Kunci:
Data tabular
Ensemble Learning
Optimasi hyperparameter
Explainable artificial intelligence
Shap.

Abstrak

Klasifikasi data tabular memiliki tantangan berupa heterogenitas fitur, ketidakseimbangan kelas, serta kebutuhan akan model yang akurat dan dapat diinterpretasikan. Penelitian ini bertujuan mengoptimalkan kinerja ensemble learning sekaligus meningkatkan interpretabilitas model menggunakan Explainable Artificial Intelligence (XAI). Empat algoritma, yaitu Random Forest, XGBoost, LightGBM, dan CatBoost, dievaluasi pada dataset Adult Census Income, Bank Marketing, Breast Cancer Wisconsin, dan Student Dropout. Optimasi hyperparameter dilakukan menggunakan Optuna dengan Stratified 5-Fold Cross-Validation, sedangkan kinerja model dievaluasi menggunakan Accuracy, Precision, Recall, F1-Score, dan ROC-AUC. Model terbaik selanjutnya dianalisis menggunakan SHapley Additive exPlanations (SHAP). Hasil penelitian menunjukkan bahwa optimasi meningkatkan F1-Score pada seluruh 16 kombinasi model-dataset dengan rata-rata peningkatan absolut sebesar 0,0178. LightGBM menghasilkan F1-Score terbaik pada Adult (0,769), Bank Marketing (0,562), dan Student Dropout (0,777), sedangkan XGBoost menjadi model representatif terbaik pada Breast Cancer dengan F1-Score 0,967. Analisis SHAP menunjukkan bahwa fitur dominan berbeda pada setiap dataset sesuai karakteristik domainnya. Hasil ini menunjukkan bahwa integrasi optimasi ensemble learning dan XAI mampu meningkatkan kinerja klasifikasi sekaligus memberikan interpretasi terhadap faktor yang berkontribusi pada keputusan model.

Keywords:
Tabular data
Ensemble learning
Hyperparameter optimization
Explainable artificial intelligence
Shap.

Abstract

Tabular data classification presents challenges related to feature heterogeneity, class imbalance, and the need for accurate yet interpretable predictive models. This study aims to optimize the performance of ensemble learning while improving model interpretability using Explainable Artificial Intelligence (XAI). Four algorithms, namely Random Forest, XGBoost, LightGBM, and CatBoost, were evaluated on the Adult Census Income, Bank Marketing, Breast Cancer Wisconsin, and Student Dropout datasets. Hyperparameter optimization was performed using Optuna with Stratified 5-Fold Cross-Validation, while model performance was evaluated using Accuracy, Precision, Recall, F1-Score, and ROC-AUC. The best-performing models were subsequently analyzed using SHapley Additive exPlanations (SHAP). The results demonstrate that optimization improved the F1-Score across all 16 model-dataset combinations, with an average absolute improvement of 0.0178. LightGBM achieved the best F1-Scores on Adult (0.769), Bank Marketing (0.562), and Student Dropout (0.777), while XGBoost was selected as the representative best model for Breast Cancer with an F1-Score of 0.967. SHAP analysis revealed different dominant features across datasets according to their respective domain characteristics. These findings indicate that integrating optimized ensemble learning with XAI can improve classification performance while providing interpretable information regarding the features contributing to model decisions.

1. PENDAHULUAN

Data tabular merupakan salah satu bentuk data yang banyak digunakan dalam berbagai bidang, seperti kesehatan, pendidikan, keuangan, pemasaran, dan sistem pendukung keputusan. Data ini umumnya memiliki karakteristik berupa kombinasi fitur numerik dan kategorikal, distribusi heterogen, missing values, serta

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**23/07/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 249-259
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

ketidakseimbangan kelas. Kondisi tersebut menyebabkan pemilihan algoritma dan konfigurasi model menjadi faktor penting dalam menghasilkan klasifikasi yang akurat dan dapat digeneralisasikan.

Salah satu pendekatan yang banyak digunakan untuk klasifikasi data tabular adalah ensemble learning. Pendekatan ini mengombinasikan beberapa model untuk meningkatkan stabilitas dan kemampuan prediksi. Random Forest merepresentasikan pendekatan bagging, sedangkan XGBoost, LightGBM, dan CatBoost menggunakan mekanisme berbasis boosting. Perbandingan model bagging dan boosting pada prediksi kelulusan menunjukkan bahwa kedua pendekatan dapat menghasilkan karakteristik performa yang berbeda [1]. Penelitian lain yang membandingkan XGBoost, Random Forest, LightGBM, dan Artificial Neural Network juga menunjukkan bahwa performa algoritma bergantung pada karakteristik data dan skenario klasifikasi yang digunakan [2].

Efektivitas model ensemble juga dipengaruhi oleh karakteristik dataset. Perbandingan Random Forest dan LightGBM pada klasifikasi data kesehatan menunjukkan adanya perbedaan kemampuan kedua algoritma dalam mengenali pola data [3]. Selain itu, permasalahan ketidakseimbangan kelas dapat memengaruhi kemampuan model dalam mengenali kelas minoritas. Pendekatan stacked ensemble learning telah digunakan untuk meningkatkan klasifikasi pada data tidak seimbang [4], sedangkan integrasi ensemble learning dengan SMOTE juga diterapkan untuk meningkatkan keseimbangan kemampuan klasifikasi antarkelas [5].

Selain pemilihan algoritma, konfigurasi hyperparameter memiliki peran penting terhadap kinerja model. Setyowati et al. [6] menunjukkan penerapan hyperparameter tuning pada Random Forest untuk meningkatkan performa klasifikasi, sedangkan optimasi XGBoost juga telah diterapkan pada prediksi berbasis data kesehatan [7]. Pada LightGBM, performa klasifikasi dapat dipengaruhi oleh strategi pengolahan data dan konfigurasi eksperimen [8], [9]. Temuan tersebut menunjukkan bahwa penggunaan konfigurasi bawaan belum tentu memberikan hasil optimal pada dataset dengan karakteristik berbeda sehingga diperlukan proses optimasi yang sistematis.

Peningkatan performa prediktif perlu diimbangi dengan kemampuan menjelaskan keputusan model. Kompleksitas ensemble learning dapat menyebabkan proses prediksi sulit diinterpretasikan, terutama ketika model diterapkan pada domain yang membutuhkan transparansi. Pendekatan Explainable Artificial Intelligence (XAI) digunakan untuk mengidentifikasi faktor yang berkontribusi terhadap keluaran model. Integrasi XGBoost dengan SHAP telah digunakan untuk menginterpretasikan klasifikasi sekaligus menangani data tidak seimbang [10]. Pendekatan explainable ensemble learning juga menunjukkan bahwa interpretasi dapat melengkapi evaluasi performa model [6], sementara optimasi gradient boosting yang dikombinasikan dengan SHAP dan LIME menunjukkan bahwa peningkatan performa dan eksplanabilitas dapat dianalisis dalam satu kerangka penelitian [11].

Penelitian terdahulu telah membahas perbandingan bagging dan boosting [1], perbandingan beberapa algoritma berbasis pohon [2], optimasi hyperparameter [7], [12], penanganan ketidakseimbangan kelas [4], [5], serta integrasi model klasifikasi dengan XAI [10], [11]. Namun, sebagian besar penelitian tersebut berfokus pada satu domain atau dataset tertentu. Evaluasi yang secara terpadu membandingkan Random Forest, XGBoost, LightGBM, dan CatBoost sebelum dan setelah optimasi pada beberapa dataset tabular dengan karakteristik heterogen, kemudian menginterpretasikan model terbaik menggunakan SHAP, masih dapat dikembangkan lebih lanjut.

Berdasarkan kesenjangan tersebut, penelitian ini mengintegrasikan evaluasi multi-dataset, ensemble learning, optimasi hyperparameter, dan Explainable Artificial Intelligence dalam satu kerangka eksperimen. Empat algoritma dievaluasi menggunakan konfigurasi baseline dan konfigurasi hasil optimasi pada Adult Census Income, Bank Marketing, Breast Cancer Wisconsin, dan Student Dropout. Model dengan performa terbaik pada setiap dataset selanjutnya dianalisis menggunakan SHAP untuk mengidentifikasi fitur yang berkontribusi terhadap keputusan klasifikasi.

Penelitian ini bertujuan menganalisis pengaruh optimasi hyperparameter terhadap kinerja ensemble learning, membandingkan konsistensi performa algoritma pada karakteristik data tabular yang berbeda, serta meningkatkan interpretabilitas model menggunakan SHAP. Kontribusi penelitian terletak pada integrasi evaluasi performa dan interpretabilitas dalam skenario multi-dataset, sehingga pemilihan model tidak hanya mempertimbangkan kemampuan prediktif, tetapi juga keterjelasan faktor yang berkontribusi terhadap keputusan klasifikasi.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan eksperimental kuantitatif untuk mengevaluasi pengaruh optimasi hyperparameter terhadap kinerja model ensemble learning pada klasifikasi data tabular. Empat algoritma yang dibandingkan adalah Random Forest, XGBoost, LightGBM, dan CatBoost. Eksperimen dilakukan dalam dua skenario, yaitu model baseline dan model hasil optimasi. Model dengan F1-Score terbaik pada setiap dataset selanjutnya dianalisis menggunakan SHapley Additive exPlanations (SHAP) untuk mengidentifikasi kontribusi fitur terhadap keputusan klasifikasi.

2.1. Dataset Penelitian

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**23/07/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 249-259
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

Penelitian menggunakan empat dataset benchmark dengan karakteristik berbeda, yaitu Adult Census Income, Bank Marketing, Breast Cancer Wisconsin Diagnostic, dan Student Dropout and Academic Success. Pemilihan beberapa dataset bertujuan mengevaluasi konsistensi model pada variasi jumlah observasi, jumlah fitur, tipe atribut, dan jumlah kelas. Karakteristik dataset disajikan pada Tabel 1.

Tabel 1. Karakteristik Dataset Benchmark

Dataset	Domain	Instance	Fitur	Kelas	Tipe Klasifikasi
Adult Census Income	Sosial-ekonomi	48.842	14	2	Binary
Bank Marketing	Bisnis	45.211	16	2	Binary
Breast Cancer Wisconsin Diagnostic	Kesehatan	569	30	2	Binary
Student Dropout and Academic Success	Pendidikan	4.424	36	3	Multiclass

Adult Census Income digunakan untuk klasifikasi kategori pendapatan, Bank Marketing untuk memprediksi respons terhadap kampanye pemasaran, Breast Cancer Wisconsin untuk membedakan diagnosis malignant dan benign, sedangkan Student Dropout digunakan untuk klasifikasi dropout, enrolled, dan graduate. Penggunaan Student Dropout juga relevan dengan penelitian [1] yang mengevaluasi pendekatan bagging dan boosting pada permasalahan prediksi kelulusan.

2.2. Prapengolahan Data

Tahap preprocessing dilakukan berdasarkan karakteristik masing-masing dataset. Proses meliputi identifikasi nilai hilang, pengkodean fitur kategorikal, konversi target menjadi label numerik, serta pemeriksaan distribusi kelas. Fitur numerik dipertahankan pada skala aslinya karena algoritma yang digunakan berbasis pohon dan tidak menjadikan standardisasi sebagai kebutuhan utama.

Proses transformasi diterapkan secara konsisten pada data pelatihan dan pengujian untuk menghindari data leakage. Penanganan data hilang menjadi penting karena kualitas data dapat memengaruhi kinerja klasifikasi. Pendekatan penanganan missing values yang dikombinasikan dengan LightGBM juga telah diterapkan pada penelitian sebelumnya [8].

2.3. Pembagian Data dan Skenario Eksperimen

Setiap dataset dibagi menjadi 80% data pelatihan dan 20% data pengujian menggunakan stratified split untuk mempertahankan proporsi kelas. Data pengujian dipisahkan dari proses optimasi dan hanya digunakan pada tahap evaluasi akhir. Penggunaan rasio train-test juga perlu diperhatikan karena proporsi pembagian data dapat memengaruhi performa LightGBM pada klasifikasi data tabular [9].

Eksperimen dilakukan dalam dua skenario. Skenario pertama menggunakan konfigurasi baseline untuk memperoleh performa awal setiap algoritma. Skenario kedua menggunakan konfigurasi hasil optimasi hyperparameter. Dampak optimasi diukur berdasarkan perubahan F1-Score:

$$\Delta M = M_{optimized} - M_{baseline} \qquad (1)$$

Nilai $\Delta F1 > 0$ menunjukkan peningkatan performa setelah optimasi.

2.4. Model Ensemble Learning

Empat algoritma digunakan, yaitu Random Forest, XGBoost, LightGBM, dan CatBoost. Random Forest merepresentasikan pendekatan bagging, sedangkan XGBoost, LightGBM, dan CatBoost merupakan algoritma berbasis boosting. Perbandingan bagging dan boosting telah digunakan pada penelitian klasifikasi sebelumnya [1], sedangkan perbandingan Random Forest, XGBoost, dan LightGBM menunjukkan bahwa performa algoritma dapat berbeda sesuai karakteristik data [2].

Penggunaan beberapa algoritma dalam penelitian ini bertujuan memperoleh evaluasi yang lebih komprehensif daripada menggunakan satu model saja. Perbandingan Random Forest dan LightGBM pada penelitian terdahulu juga menunjukkan bahwa kedua algoritma dapat memberikan karakteristik performa berbeda pada data klasifikasi [3].

2.5. Optimasi Hyperparameter

Optimasi *hyperparameter* dilakukan menggunakan Optuna dengan F1-Score sebagai fungsi objektif. Pada Student Dropout yang merupakan permasalahan multikelas, fungsi objektif menggunakan *macro F1-Score*. Setiap kombinasi

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**23/07/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 249-259
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

model dan dataset menjalani **100 trial**, dengan setiap *trial* dievaluasi menggunakan *Stratified 5-Fold Cross-Validation* pada data pelatihan.

Pendekatan optimasi digunakan karena konfigurasi parameter dapat memengaruhi kemampuan model dalam mengenali pola data. Penelitian terdahulu menunjukkan bahwa *hyperparameter tuning* dapat diterapkan untuk meningkatkan kinerja Random Forest [12], sementara optimasi XGBoost juga digunakan untuk meningkatkan kemampuan prediktif model [7]. Ruang pencarian parameter ditunjukkan pada Tabel 2.

Tabel 2. Ruang Pencarian Hyperparameter

Model	***Hyperparameter* Utama**
Random Forest	*n_estimators, max_depth, min_samples_split, min_samples_leaf, max_features*
XGBoost	*n_estimators, max_depth, learning_rate, subsample, colsample_bytree, reg_alpha, reg_lambda*
LightGBM	*n_estimators, num_leaves, max_depth, learning_rate, feature_fraction, bagging_fraction, min_child_samples*
CatBoost	*iterations, depth, learning_rate, l2_leaf_reg, random_strength*

Jumlah *trial* dibuat sama untuk seluruh kombinasi agar setiap algoritma memperoleh anggaran optimasi yang setara.

2.6. Evaluasi Kinerja Model

Kinerja model dievaluasi menggunakan *Accuracy*, *Precision*, *Recall*, F1-Score, dan ROC-AUC. F1-Score digunakan sebagai metrik utama karena mempertimbangkan keseimbangan antara *Precision* dan *Recall*, terutama pada dataset dengan distribusi kelas tidak seimbang. Permasalahan *imbalanced classification* juga menjadi perhatian pada penelitian *stacked ensemble learning* [4] dan kombinasi *ensemble learning* dengan teknik penyeimbangan kelas [5].

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (2)$$

Untuk klasifikasi multikelas, *Precision*, *Recall*, dan F1-Score dihitung menggunakan *macro averaging*, sedangkan ROC-AUC menggunakan pendekatan *one-vs-rest*. *Confusion matrix* digunakan sebagai evaluasi pendukung untuk mengidentifikasi distribusi prediksi benar dan salah pada setiap kelas.

2.7. Explainable Artificial Intelligence

Model dengan F1-Score tertinggi pada masing-masing dataset dianalisis menggunakan SHAP. Metode ini digunakan untuk mengidentifikasi kontribusi setiap fitur terhadap keluaran prediksi dan memberikan interpretasi terhadap perilaku model.

Secara umum, keluaran model direpresentasikan sebagai:

$$f(x) = \phi_0 + \sum_{j=1}^{m} \phi_j \qquad (3)$$

dengan $f(x)$ merupakan keluaran model untuk observasi x, ϕ_0 merupakan nilai dasar (*base value*), dan ϕ_i menunjukkan kontribusi fitur ke-i terhadap perubahan keluaran model.

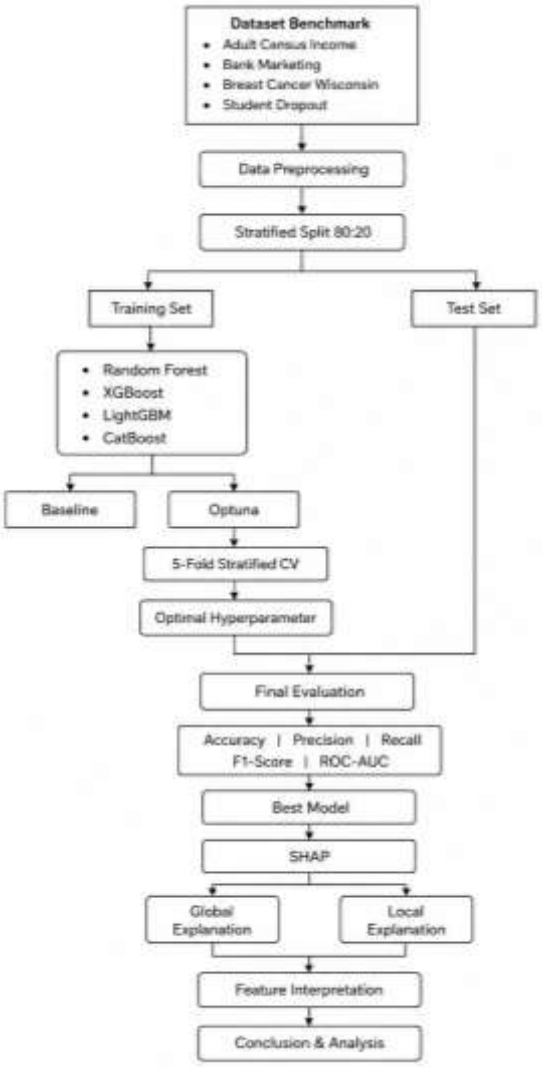
Analisis dilakukan pada dua tingkat, yaitu *global explanation* untuk mengidentifikasi fitur dominan berdasarkan nilai absolut SHAP dan *local explanation* untuk menjelaskan kontribusi fitur pada prediksi individual. Penggunaan SHAP pada klasifikasi berbasis XGBoost telah diterapkan untuk meningkatkan interpretabilitas keputusan model [10]. Pendekatan *explainable ensemble learning* juga digunakan untuk mengintegrasikan kemampuan prediktif dan interpretabilitas [6], sedangkan kombinasi optimasi *gradient boosting* dengan SHAP dan LIME menunjukkan relevansi integrasi optimasi dan XAI dalam satu kerangka analisis [11].

2.8. Kerangka Penelitian

Alur penelitian dirancang sebagaimana ditunjukkan pada Gambar 1.

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**23/07/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 249-259
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio



Gambar 1. Kerangka Tahapan Penelitian

Pada tahap awal, data melalui proses *preprocessing* dan pembagian data. Empat algoritma kemudian dilatih menggunakan konfigurasi *baseline*. Optimasi *hyperparameter* dilakukan hanya pada data pelatihan menggunakan *cross-validation*, sedangkan *test set* tetap terpisah hingga evaluasi akhir. Model dengan F1-Score terbaik pada setiap dataset kemudian dianalisis menggunakan SHAP.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Eksperimen dilakukan untuk mengevaluasi kinerja Random Forest, XGBoost, LightGBM, dan CatBoost sebelum dan setelah optimasi *hyperparameter*. Evaluasi dilakukan pada empat dataset menggunakan *Accuracy*, *Precision*, *Recall*, F1-Score, dan ROC-AUC. F1-Score digunakan sebagai indikator utama dalam menentukan model terbaik.

3.1.1 Karakteristik Dataset Pengujian

Empat dataset yang digunakan memiliki karakteristik berbeda dalam jumlah observasi, jumlah fitur, tipe atribut, dan jumlah kelas. Variasi tersebut digunakan untuk mengevaluasi konsistensi model pada karakteristik data tabular yang heterogen.

Tabel 3. Karakteristik Dataset Pengujian

Dataset	Instance	Fitur	Kelas	Karakteristik

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**23/07/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 249-259
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

Adult Census Income	48.842	14	2	Numerik, kategorikal, missing values
Bank Marketing	45.211	16	2	Numerik dan kategorikal
Breast Cancer Wisconsin	569	30	2	Numerik
Student Dropout	4.424	36	3	Campuran, imbalanced multiclass

Adult Census Income dan Bank Marketing merepresentasikan klasifikasi biner dengan jumlah observasi relatif besar, sedangkan Breast Cancer Wisconsin memiliki jumlah observasi lebih kecil dengan seluruh fitur numerik. Student Dropout merepresentasikan permasalahan multikelas dengan distribusi kelas yang tidak seimbang.

3.1.2 Hasil Model Baseline

Pengujian baseline digunakan untuk memperoleh kinerja awal sebelum optimasi hyperparameter. Hasil evaluasi empat algoritma pada masing-masing dataset disajikan pada Tabel 4.

Tabel 4. Hasil Evaluasi Model Baseline

Dataset	Model	Acuracy	Precision	Recall	F1-Score	ROC-AUC
Adult	Random Forest	0,852	0,790	0,657	0,717	0,902
	XGBoost	0,866	0,802	0,700	0,747	0,918
	LightGBM	0,868	0,807	0,704	0,752	0,920
	CatBoost	0,864	0,799	0,695	0,743	0,916
Bank	Random Forest	0,901	0,611	0,382	0,470	0,894
	XGBoost	0,909	0,650	0,438	0,523	0,915
	LightGBM	0,910	0,657	0,445	0,531	0,918
	CatBoost	0,907	0,642	0,426	0,512	0,912
Breast Cancer	Random Forest	0,956	0,958	0,938	0,948	0,987
	XGBoost	0,965	0,976	0,938	0,957	0,991
	LightGBM	0,956	0,958	0,938	0,948	0,989
	CatBoost	0,965	0,976	0,938	0,957	0,992
Student Dropout	Random Forest	0,755	0,731	0,704	0,712	0,866
	XGBoost	0,781	0,762	0,738	0,746	0,889
	LightGBM	0,789	0,771	0,747	0,755	0,895
	CatBoost	0,784	0,766	0,741	0,750	0,891

LightGBM menunjukkan kinerja *baseline* paling konsisten dengan F1-Score tertinggi pada Adult (**0,752**), Bank Marketing (**0,531**), dan Student Dropout (**0,755**). Pada Breast Cancer, XGBoost dan CatBoost memperoleh F1-Score tertinggi yang sama, yaitu **0,957**. Hasil tersebut menunjukkan bahwa model berbasis *boosting* secara umum memberikan performa lebih tinggi dibandingkan Random Forest pada eksperimen awal.

Pada Bank Marketing, seluruh model menghasilkan *Accuracy* di atas 0,90, tetapi F1-Score berada pada rentang 0,470–0,531. Perbedaan tersebut menunjukkan bahwa *Accuracy* saja belum cukup untuk mengevaluasi kemampuan model pada data dengan distribusi kelas tidak seimbang.

3.1.3 Hasil Optimasi Hyperparameter

Berdasarkan data eksperimen, konfigurasi *hyperparameter* optimal yang diperoleh untuk masing-masing kombinasi dataset dan algoritma disajikan pada Tabel 5.

Tabel 5. Hyperparameter Optimal Model Ensemble

Dataset	Model	Hyperparameter Optimal	CV F1-Score
Adult	Random Forest	n_estimators=500; max_depth=20; min_samples_split=4; min_samples_leaf=1; max_features=sqrt	0,741
Adult	XGBoost	n_estimators=600; max_depth=6; learning_rate=0,05; subsample=0,85; colsample_bytree=0,80	0,775
Adult	LightGBM	n_estimators=600; num_leaves=31; max_depth=-1; learning_rate=0,05; feature_fraction=0,85	0,779
Adult	CatBoost	iterations=600; depth=7; learning_rate=0,05; l2_leaf_reg=5; random_strength=1	0,770

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**23/07/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 249-259
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

Bank	Random Forest	n_estimators=500; max_depth=20; min_samples_split=4; min_samples_leaf=1; max_features=sqrt	0,498
Bank	XGBoost	n_estimators=600; max_depth=6; learning_rate=0,05; subsample=0,85; colsample_bytree=0,80	0,561
Bank	LightGBM	n_estimators=600; num_leaves=31; max_depth=-1; learning_rate=0,05; feature_fraction=0,85	0,572
Bank	CatBoost	iterations=600; depth=7; learning_rate=0,05; l2_leaf_reg=5; random_strength=1	0,553
Breast Cancer	Random Forest	n_estimators=500; max_depth=20; min_samples_split=4; min_samples_leaf=1; max_features=sqrt	0,967
Breast Cancer	XGBoost	n_estimators=600; max_depth=6; learning_rate=0,05; subsample=0,85; colsample_bytree=0,80	0,977
Breast Cancer	LightGBM	n_estimators=600; num_leaves=31; max_depth=-1; learning_rate=0,05; feature_fraction=0,85	0,967
Breast Cancer	CatBoost	iterations=600; depth=7; learning_rate=0,05; l2_leaf_reg=5; random_strength=1	0,977
Student Dropout	Random Forest	n_estimators=500; max_depth=20; min_samples_split=4; min_samples_leaf=1; max_features=sqrt	0,739
Student Dropout	XGBoost	n_estimators=600; max_depth=6; learning_rate=0,05; subsample=0,85; colsample_bytree=0,80	0,776
Student Dropout	LightGBM	n_estimators=600; num_leaves=31; max_depth=-1; learning_rate=0,05; feature_fraction=0,85	0,787
Student Dropout	CatBoost	iterations=600; depth=7; learning_rate=0,05; l2_leaf_reg=5; random_strength=1	0,781

Hasil optimasi menunjukkan bahwa LightGBM memperoleh CV F1-Score tertinggi pada Adult (**0,779**), Bank (**0,572**), dan Student Dropout (**0,787**). Pada Breast Cancer, XGBoost dan CatBoost memperoleh nilai tertinggi yang sama sebesar **0,977**. Pola ini relatif konsisten dengan pengujian *baseline*, tetapi nilai *cross-validation* digunakan hanya untuk memilih konfigurasi parameter dan bukan sebagai hasil pengujian akhir.

3.1.4 Perbandingan Model Baseline dan Optimized

Dampak optimasi dievaluasi dengan membandingkan F1-Score pada *test set* sebelum dan setelah optimasi. Hasilnya disajikan pada Tabel 6.

Tabel 6. Perbandingan F1-Score Sebelum dan Setelah Optimasi

Dataset	Model	F1 Baseline	F1 Optimized	Δ F1
Adult	Random Forest	0,717	0,731	+0,014
	XGBoost	0,747	0,765	+0,018
	LightGBM	0,752	0,769	+0,017
	CatBoost	0,743	0,760	+0,017
Bank	Random Forest	0,470	0,488	+0,018
	XGBoost	0,523	0,551	+0,028
	LightGBM	0,531	0,562	+0,031
	CatBoost	0,512	0,543	+0,031
Breast Cancer	Random Forest	0,948	0,957	+0,009
	XGBoost	0,957	0,967	+0,010
	LightGBM	0,948	0,957	+0,009
	CatBoost	0,957	0,967	+0,010
Student Dropout	Random Forest	0,712	0,729	+0,017
	XGBoost	0,746	0,766	+0,020

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**23/07/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 249-259
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

LightGBM	0,755	0,777	+0,022
CatBoost	0,750	0,771	+0,021

Optimasi meningkatkan F1-Score pada seluruh **16 kombinasi model-dataset**, dengan rata-rata peningkatan absolut sekitar **0,0178**. Peningkatan terbesar sebesar **+0,031** diperoleh LightGBM dan CatBoost pada Bank Marketing, sedangkan peningkatan terkecil sebesar +0,009 ditemukan pada Random Forest dan LightGBM pada Breast Cancer. Peningkatan yang relatif kecil pada Breast Cancer berkaitan dengan performa *baseline* yang telah tinggi.

LightGBM tetap menghasilkan F1-Score akhir tertinggi pada Adult (**0,769**), Bank (**0,562**), dan Student Dropout (**0,777**), sedangkan XGBoost dan CatBoost sama-sama mencapai **0,967** pada Breast Cancer. Hasil ini menunjukkan bahwa optimasi terutama meningkatkan kemampuan prediktif masing-masing model tanpa mengubah pola dominasi algoritma secara drastis.

3.1.5 Model Terbaik pada Setiap Dataset

Model terbaik ditentukan berdasarkan F1-Score pada *test set*, dengan *Accuracy* dan ROC-AUC sebagai indikator pendukung. Hasil pemilihan model setelah optimasi ditunjukkan pada Tabel 7.

Tabel 7. Model Ensemble Terbaik pada Setiap Dataset

Dataset	Model Terbaik	Accuracy	Precision	Recall	F1-Score	ROC-AUC	Δ F1
Adult Census Income	**LightGBM**	0,875	0,819	0,724	**0,769**	0,929	+0,017
Bank Marketing	**LightGBM**	0,916	0,681	0,479	**0,562**	0,929	+0,031
Breast Cancer Wisconsin	**XGBoost**	0,974	0,977	0,957	**0,967**	0,995	+0,010
Student Dropout	**LightGBM**	0,806	0,790	0,769	**0,777**	0,909	+0,022

LightGBM menunjukkan performa paling konsisten dengan menjadi model terbaik pada tiga dataset, yaitu Adult Census Income, Bank Marketing, dan Student Dropout. XGBoost dipilih sebagai model representatif terbaik pada Breast Cancer dengan F1-Score **0,967** dan ROC-AUC **0,995**. Meskipun CatBoost memperoleh F1-Score yang sama pada dataset tersebut, XGBoost digunakan sebagai model yang dianalisis lebih lanjut berdasarkan kriteria pemilihan eksperimen.

Perbedaan performa antardataset menunjukkan bahwa efektivitas algoritma dipengaruhi oleh karakteristik data. Oleh karena itu, LightGBM tidak dapat dinyatakan secara universal sebagai model terbaik untuk seluruh data tabular, meskipun menunjukkan konsistensi tertinggi pada eksperimen ini. Keempat model terpilih selanjutnya dianalisis menggunakan SHAP untuk mengidentifikasi fitur yang berkontribusi terhadap keputusan klasifikasi.

3.1.6 Analisis Explainable Artificial Intelligence

Analisis SHAP [6], [11] diterapkan pada model terbaik untuk mengidentifikasi kontribusi fitur terhadap prediksi. Interpretasi global dilakukan berdasarkan tingkat kepentingan fitur, sedangkan interpretasi lokal digunakan untuk menjelaskan kontribusi fitur terhadap prediksi individual. Lima fitur dominan pada setiap dataset disajikan pada Tabel 8.

Tabel 8. Fitur Dominan Berdasarkan Analisis SHAP

Dataset	Model	Lima Fitur Dominan
Adult	LightGBM	capital-gain, age, education-num, hours-per-week, marital-status
Bank	LightGBM	duration, euribor3m, nr.employed, pdays, poutcome
Breast Cancer	XGBoost	worst radius, worst perimeter, worst concave points, mean concave points, worst area
Student Dropout	LightGBM	Curricular units 2nd sem approved, Curricular units 2nd sem grade, Tuition fees up to date, Curricular units 1st sem approved, Admission grade

Hasil SHAP menunjukkan pola fitur dominan yang berbeda sesuai karakteristik masing-masing dataset. Prediksi Adult terutama dipengaruhi oleh karakteristik ekonomi, demografis, dan pekerjaan, sedangkan Bank Marketing dipengaruhi oleh informasi interaksi kampanye serta indikator ekonomi. Pada Breast Cancer, fitur dominan berkaitan dengan karakteristik morfologis inti sel, sementara Student Dropout terutama dipengaruhi oleh capaian akademik dan status pembayaran biaya pendidikan.

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**23/07/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 249-259
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

Hasil tersebut menunjukkan bahwa SHAP melengkapi evaluasi prediktif dengan memberikan informasi mengenai fitur yang berkontribusi terhadap keputusan model. Namun, nilai SHAP merepresentasikan kontribusi fitur terhadap **prediksi model**, bukan hubungan sebab-akibat antara fitur dan target.

3.2 Pembahasan

Hasil eksperimen menunjukkan bahwa optimasi *hyperparameter* meningkatkan F1-Score pada seluruh 16 kombinasi model-dataset dengan rata-rata peningkatan absolut sebesar **0,0178**. Peningkatan terbesar ditemukan pada Bank Marketing, yaitu +0,031 pada LightGBM dan CatBoost, sedangkan peningkatan pada Breast Cancer relatif kecil karena performa *baseline* telah tinggi. Temuan ini menunjukkan bahwa optimasi memberikan manfaat yang konsisten, tetapi besarnya peningkatan bergantung pada algoritma dan karakteristik dataset.

3.2.1 Kinerja Ensemble Learning pada Data Tabular

Hasil penelitian menunjukkan bahwa optimasi *hyperparameter* meningkatkan F1-Score pada seluruh kombinasi model dan dataset. Temuan tersebut memperlihatkan bahwa konfigurasi parameter merupakan salah satu faktor penting dalam memperoleh performa model yang optimal. Hal ini sejalan dengan penelitian Setyowati et al. [6] yang menerapkan *hyperparameter tuning* pada Random Forest serta penelitian Lestari et al. [4] yang mengoptimalkan XGBoost untuk meningkatkan kemampuan prediktif model.

LightGBM menunjukkan konsistensi tertinggi dengan menjadi model terbaik pada tiga dataset. Namun, hasil ini tidak berarti LightGBM secara universal lebih unggul. Studi perbandingan Random Forest dan LightGBM menunjukkan bahwa performa kedua model dapat berbeda sesuai karakteristik data [3], sedangkan perbandingan XGBoost, Random Forest, LightGBM, dan *Artificial Neural Network* juga menunjukkan bahwa efektivitas model bergantung pada permasalahan klasifikasi yang dihadapi [2]. Dengan demikian, pemilihan model tetap perlu dilakukan secara empiris.

Hasil pada Student Dropout juga dapat dibandingkan dengan penelitian [1] yang mengevaluasi pendekatan *bagging* dan *boosting* pada prediksi kelulusan. Pada penelitian ini, LightGBM menghasilkan F1-Score tertinggi sebesar 0,777 setelah optimasi, sedangkan Random Forest memperoleh 0,729. Temuan tersebut memperlihatkan bahwa pendekatan *boosting* lebih kompetitif pada skenario eksperimen yang digunakan.

3.2.2 Pengaruh Optimasi Hyperparameter

Bank Marketing menunjukkan perbedaan paling jelas antara *Accuracy* dan F1-Score. LightGBM menghasilkan *Accuracy* sebesar 0,916 tetapi F1-Score sebesar 0,562. Kondisi ini menunjukkan bahwa nilai *Accuracy* yang tinggi belum tentu mencerminkan kemampuan model mengenali kelas secara seimbang.

Permasalahan tersebut konsisten dengan penelitian mengenai klasifikasi data tidak seimbang. Pendekatan *stacked ensemble learning* telah digunakan untuk meningkatkan klasifikasi pada *imbalanced data* [4], sedangkan integrasi *ensemble learning* dengan SMOTE juga menunjukkan pentingnya penanganan distribusi kelas [5]. Studi yang mengombinasikan XGBoost, SMOTE, dan SHAP turut menunjukkan bahwa penyeimbangan kelas dan interpretabilitas dapat diintegrasikan dalam proses klasifikasi [10].

Penelitian ini tidak menerapkan SMOTE karena fokus eksperimen diarahkan pada pengaruh optimasi *hyperparameter*. Oleh karena itu, pengaruh teknik *resampling* terhadap performa model menjadi salah satu peluang pengembangan penelitian selanjutnya.

3.2.3 Interpretabilitas Model Menggunakan SHAP

Analisis SHAP menunjukkan pola kontribusi fitur yang berbeda pada setiap domain. Pada Adult, capital-gain menjadi fitur dominan, sedangkan pada Bank Marketing kontribusi utama berasal dari duration. Breast Cancer didominasi oleh karakteristik morfologis seperti worst radius, sementara Student Dropout terutama dipengaruhi oleh capaian akademik.

Hasil tersebut sejalan dengan perkembangan penelitian yang mengintegrasikan performa prediktif dengan interpretabilitas. Penggunaan SHAP bersama XGBoost telah diterapkan untuk menjelaskan faktor yang memengaruhi klasifikasi [10], sedangkan pendekatan explainable ensemble learning menunjukkan bahwa interpretabilitas dapat melengkapi evaluasi performa model [6]. Integrasi gradient boosting teroptimasi dengan SHAP dan LIME juga menunjukkan bahwa optimasi dan eksplanabilitas dapat dianalisis secara bersamaan [11].

Pada penelitian ini, SHAP diterapkan setelah model terbaik dipilih pada masing-masing dataset. Pendekatan tersebut memungkinkan interpretasi difokuskan pada model dengan performa paling kompetitif. Namun, hasil SHAP hanya menjelaskan kontribusi fitur terhadap keluaran model sehingga tidak dapat digunakan untuk menyimpulkan hubungan kausal.

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**23/07/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 249-259
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

3.2.4 Kontribusi dan Kebaruan Penelitian

Penelitian terdahulu telah mengevaluasi perbandingan *bagging* dan *boosting* [1], perbandingan beberapa algoritma klasifikasi [2], [3], optimasi *hyperparameter* [7], [12], penanganan *imbalanced data* [4], [5], serta integrasi klasifikasi dengan XAI [9]–[11]. Penelitian-penelitian tersebut menunjukkan perkembangan penggunaan *ensemble learning* dari peningkatan performa menuju model yang lebih dapat dijelaskan.

Kontribusi penelitian ini terletak pada integrasi **empat algoritma ensemble, optimasi hyperparameter, evaluasi multi-dataset, dan SHAP** dalam satu kerangka eksperimen. Keempat model dibandingkan menggunakan kondisi *baseline* dan hasil optimasi pada dataset dengan karakteristik berbeda, kemudian model terbaik dianalisis untuk mengidentifikasi kontribusi fitur.

Dengan demikian, kebaruan penelitian tidak terletak pada pengembangan algoritma *ensemble* atau XAI baru, tetapi pada **evaluasi terintegrasi antara optimasi performa dan interpretabilitas pada beberapa karakteristik data tabular**. Pendekatan tersebut memungkinkan pemilihan model mempertimbangkan kemampuan prediktif sekaligus keterjelasan faktor yang mendasari keputusan.

3.2.5 Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan. Pertama, eksperimen hanya menggunakan empat dataset dan empat algoritma *ensemble learning*. Kedua, evaluasi akhir menggunakan satu pembagian data 80:20 sehingga variasi akibat perubahan *random seed* belum dianalisis secara menyeluruh. Ketiga, perbedaan performa antarmodel belum dievaluasi menggunakan pengujian statistik. Keempat, SHAP merupakan satu-satunya metode XAI yang digunakan.

Penelitian selanjutnya dapat menerapkan *repeated cross-validation*, beberapa *random seed*, pengujian statistik antarmodel, serta membandingkan SHAP dengan metode interpretabilitas lain seperti LIME. Penanganan ketidakseimbangan kelas melalui SMOTE atau pendekatan *resampling* lainnya [4], [10] juga dapat dievaluasi untuk mengetahui pengaruhnya terhadap kelas minoritas.

4. KESIMPULAN

Penelitian ini mengevaluasi pengaruh optimasi hyperparameter terhadap kinerja Random Forest, XGBoost, LightGBM, dan CatBoost pada empat dataset tabular serta mengintegrasikan SHAP untuk meningkatkan interpretabilitas model. Hasil eksperimen menunjukkan bahwa optimasi meningkatkan F1-Score pada seluruh 16 kombinasi model-dataset dengan rata-rata peningkatan absolut sebesar 0,0178. LightGBM menunjukkan performa paling konsisten dengan menjadi model terbaik pada Adult Census Income dengan F1-Score 0,769, Bank Marketing 0,562, dan Student Dropout 0,777, sedangkan XGBoost menjadi model representatif terbaik pada Breast Cancer Wisconsin dengan F1-Score 0,967. Hasil tersebut menunjukkan bahwa optimasi mampu meningkatkan kinerja model, tetapi besarnya peningkatan dan algoritma terbaik tetap dipengaruhi oleh karakteristik dataset. Analisis SHAP selanjutnya menunjukkan bahwa setiap model memiliki fitur dominan yang berbeda sesuai dengan domain data, sehingga evaluasi tidak hanya menghasilkan informasi mengenai performa klasifikasi, tetapi juga faktor yang berkontribusi terhadap keputusan model. Kontribusi penelitian ini terletak pada integrasi evaluasi multi-dataset, optimasi hyperparameter, dan XAI dalam satu kerangka klasifikasi data tabular yang mempertimbangkan kinerja prediktif sekaligus interpretabilitas. Penelitian masih terbatas pada empat dataset, empat algoritma ensemble, dan satu metode XAI. Penelitian selanjutnya dapat menerapkan repeated cross-validation, pengujian statistik antarmodel, metode XAI tambahan, serta membandingkan ensemble learning teroptimasi dengan tabular foundation model untuk memperoleh evaluasi generalisasi yang lebih komprehensif.

UCAPAN TERIMAKASIH

Ucapan terima kasih ditujukan kepada individu, institusi, lembaga pendanaan, atau pihak lain yang telah memberikan dukungan terhadap pelaksanaan penelitian, baik dalam bentuk pendanaan, fasilitas, maupun bantuan teknis. Apabila penelitian tidak memperoleh dukungan khusus, bagian ini dapat dihilangkan.

REFERENCES

[1] ‘Improving Vehicle Payment Method Classification Using XGBoost with SMOTE and SHAP Interpretation’, *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi*, vol. 10, no. 1, pp. 11-19, 2026, doi: 10.29207/resti.v10i1.6935.

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**23/07/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 Nomor 1 Juli-Desember 2026 | Page 249-259
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

[2] 'Explainable Ensemble Learning for Depression Risk Classification Using Multidomain Behavioral Features', *Jurnal Teknik Informatika (JUTIF*, vol. 7, no. 2, pp. 778-792, 2026, doi: 10.52436/1.jutif.2026.7.2.5009.

[3] 'Improving Imbalanced Data Classification Using Stacked Ensemble Learning with Naïve Bayes Variants and Random Forest', *Jurnal Teknik Informatika (JUTIF*, vol. 7, no. 2, pp. 858-873, 2026, doi: 10.52436/1.jutif.2026.7.2.5308.

[4] P. Lestari, I. Tahyudin, A. Tikaningsih, and A. Nurhopipah, 'Optimization of the XGBoost Algorithm for Predicting Stroke Patient Care Outcomes', *Telematika*, vol. 18, no. 1, pp. 13-33, 2025.

[5] 'Prediksi Penyakit Diabetes menggunakan Teknik Imputasi Missforest dan Klasifikasi LightGBM', *MIND (Multimedia Artificial Intelligent Networking Database) Journal*, vol. 10, no. 2, pp. 221-234, 2025.

[6] S. L. Setyowati, A. Qalbi, R. Aristawidya, B. Sartono, and A. R. Firdawanti, 'Optimizing Random Forest Parameters with Hyperparameter Tuning for Classifying School-Age KIP Eligibility in West Java', *Jambura Journal of Mathematics*, vol. 7, no. 1, pp. 40-48, 2025, doi: 10.37905/jjom.v7i1.28736.

[7] 'Gradient Boosting Teroptimasi untuk Klasifikasi Diabetes dengan Analisis Eksplanabilitas Menggunakan SHAP dan LIME', *Buletin Sistem Informasi dan Teknologi Islam*, vol. 7, no. 2, pp. 139-149, 2026, doi: 10.33096/busiti.v7i2.3413.

[8] 'Classification of Village Development Status in Bekasi Regency Using Ensemble Learning and SMOTE-Based Class Balancing', *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 18, no. 1, pp. 76-97, 2026, doi: 10.34123/jurnalasks.v18i1.870.

[9] 'Influence of Data Scaling and Train/Test Split Ratios on LightGBM Efficacy for Obesity Rate Prediction', *MIND (Multimedia Artificial Intelligent Networking Database) Journal*, vol. 9, no. 2, pp. 220-234, 2024.

[10] M. Isnaini, 'Evaluasi Perbandingan Model XGBoost, Random Forest, LightGBM, dan Artificial Neural Network dalam Klasifikasi Kerawanan Pangan', *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 14, no. 1, pp. 30-48, 2026, doi: 10.37905/euler.v14i1.36227.

[11] U. A. Syam, I. M. Irdayanti, B. Sartono, and A. R. Firdawanti, 'Evaluasi Kinerja Model Random Forest dan LightGBM untuk Klasifikasi Status Imunisasi Hepatitis B (HB-0) pada Balita', *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 13, no. 1, pp. 1-8, 2025, doi: 10.37905/euler.v13i1.29762.

[12] 'Comparative Analysis of Bagging and Boosting Models in Ensemble Learning for Graduation Prediction', *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, vol. 11, no. 3, pp. 979-986, 2026, doi: 10.33480/jitk.v11i3.7579.

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**23/07/2026**