

Perbandingan XGBoost dan LightGBM untuk Prediksi Lama Rawat Inap di Rumah Sakit

SYARIF HIDAYATULLAH¹, FADHILAH QALBI ANNISA^{2,*}, ELLY MATUL IMAH²

¹Program Studi S1 Sains Data, Universitas Negeri Surabaya, Indonesia

²Program Studi S1 Kecerdasan Artifisial, Universitas Negeri Surabaya, Indonesia

Abstrak

Lama rawat inap (*Length of Stay*/LOS) merupakan indikator penting dalam menilai efisiensi pelayanan rumah sakit, khususnya dalam skema pembiayaan berbasis INA-CBGs. LOS yang tinggi dapat menyebabkan inefisiensi biaya serta keterbatasan kapasitas tempat tidur, terutama pada rumah sakit rujukan dengan volume pasien yang besar. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma *machine learning* berbasis *gradient boosting*, yaitu LightGBM dan XGBoost, dalam memprediksi LOS pasien rawat inap menggunakan data Rekam Medis Elektronik (RME). Data berasal dari pasien rawat inap periode 2023–2024 dan mencakup informasi demografis, diagnosis, serta konsumsi obat. Setelah melalui tahap prapemrosesan dan seleksi fitur, diperoleh dataset akhir yang terdiri dari 5.557 pasien dengan 799 fitur. Kedua model dilatih dan diuji menggunakan skema *5-fold cross-validation* dengan optimasi *hyperparameter* melalui *Bayesian Optimization*. Hasil evaluasi menunjukkan bahwa LightGBM mencapai kinerja terbaik dengan RMSE sebesar 2,31, MedAE sebesar 0,85, dan R^2 sebesar 0,79, serta waktu pelatihan 4,5 kali lebih cepat dibandingkan XGBoost. Analisis seleksi fitur mengonfirmasi bahwa pengurangan fitur dari 799 menjadi 137 (*threshold* 0,001) mampu mempertahankan akurasi LightGBM ($R^2 = 0,78$) sekaligus mereduksi waktu komputasi sebesar 74,9%. Analisis SHAP mengungkapkan bahwa fitur konsumsi obat, khususnya antibiotik injeksi dan cairan intravena, merupakan kontributor paling dominan terhadap prediksi LOS. Hasil ini menunjukkan bahwa LightGBM memiliki keseimbangan optimal antara akurasi dan efisiensi komputasi untuk mendukung perencanaan klinis berbasis data RME terstruktur.

Kata Kunci *Bayesian Optimization; efisiensi komputasi; konsumsi obat; rekam medis elektronik; seleksi fitur*

Abstract

Length of Stay (LOS) is a key indicator for assessing hospital service efficiency, particularly under the INA-CBGs reimbursement scheme. Prolonged LOS can lead to cost inefficiencies and bed capacity constraints, especially in referral hospitals with high patient volumes. This study aims to compare the performance of two gradient boosting-based machine learning algorithms, LightGBM and XGBoost, in predicting inpatient LOS using Electronic Medical Records (EMR). Data were sourced from inpatients during 2023–2024, encompassing demographics, diagnoses, and medication consumption. After preprocessing and feature engineering, the final dataset comprised 5,557 patients with 799 features. Both models were trained and evaluated using 5-fold cross-validation with Bayesian Optimization for hyperparameter tuning. Results show that LightGBM

*Corresponding author. Email: fadhilahannisa@unesa.ac.id or ellymatul@unesa.ac.id.

achieved the best performance with an RMSE of 2.31, MedAE of 0.85, and R^2 of 0.79, while being 4.5 times faster in training than XGBoost. Feature selection analysis confirmed that reducing features from 799 to 137 (threshold 0.001) maintained LightGBM accuracy ($R^2 = 0.78$) while decreasing computational time by 74.9%. SHAP analysis revealed that medication features, particularly injectable antibiotics and intravenous fluids, were the most dominant contributors to LOS prediction. These findings demonstrate that LightGBM offers an optimal balance between predictive accuracy and computational efficiency for supporting clinical planning based on structured EMR data.

Keywords *Bayesian Optimization; computational efficiency; electronic medical records; feature selection; medication consumption*

1. Pendahuluan

Lama rawat inap (*Length of Stay/LOS*) merupakan indikator fundamental yang telah lama digunakan untuk mengevaluasi pemanfaatan sumber daya, efisiensi biaya layanan kesehatan, serta luaran klinis di rumah sakit (Hirani et al., 2025). Stone et al. (2022) menunjukkan bahwa LOS pada diagnosis yang serupa dapat bervariasi secara luas antara 2 hingga lebih dari 50 hari, sehingga pengembangan model prediktif yang andal sangat krusial dalam mendukung perencanaan kapasitas dan pengelolaan alur pasien. Tantangan manajemen ini juga lazim ditemukan di negara berkembang akibat keterbatasan infrastruktur dan tingginya beban pasien, sebagaimana dilaporkan di Ethiopia (Mengistu et al., 2026). Sejalan dengan fenomena tersebut, sistem kesehatan Indonesia melalui program Jaminan Kesehatan Nasional (JKN) menghadapi tekanan keberlanjutan finansial yang memicu reformasi sistem pembayaran melalui model INA-CBG (*Indonesian Case-Based Groups*) (Alvianto and Pribadi, 2025). Dalam mekanisme tarif tetap tersebut, Fitri and Sundari (2023) melaporkan adanya selisih rata-rata sebesar 32,5% antara biaya nyata (*real cost*) rumah sakit dan tarif klaim INA-CBGs. Hal ini mempertegas bahwa prediksi LOS yang akurat bukan sekadar kebutuhan klinis, melainkan instrumen vital bagi efisiensi manajemen sumber daya dan keberlanjutan finansial rumah sakit.

Terdapat beberapa pendekatan yang dapat digunakan untuk memprediksi LOS, mulai dari metode statistik konvensional seperti regresi linier dan analisis survival, hingga pendekatan berbasis *machine learning*. *machine learning* menunjukkan keunggulan yang lebih baik karena kemampuannya menangkap hubungan non-linear dan interaksi kompleks antarvariabel klinis tanpa memerlukan asumsi distribusi yang ketat (Jain et al., 2024). Almeida et al. (2024) menyimpulkan bahwa algoritma *machine learning* sangat efektif dalam menangani data medis yang kompleks dan berdimensi tinggi.

Gradient boosting merupakan salah satu keluarga algoritma *machine learning* berbasis *ensemble* yang membangun model secara aditif dan sekuensial, yakni setiap pohon keputusan baru dilatih untuk memperbaiki kesalahan (residual) dari model sebelumnya (Friedman, 2001). Keluarga algoritma ini telah berkembang melalui beberapa varian, mulai dari AdaBoost yang memberikan bobot lebih besar pada sampel yang sulit diprediksi (Freund and Schapire, 1997), *Gradient Boosting Machine* (GBM) yang mengoptimalkan fungsi kerugian melalui pendekatan gradien (Friedman, 2001), hingga implementasi modern yang lebih efisien seperti XGBoost (Chen and Guestrin, 2016) dan LightGBM (Ke et al., 2017). Dalam prediksi LOS, Zeleke et al. (2023) menemukan bahwa *Gradient Boosting* memberikan performa terbaik di instalasi gawat darurat. Hasil serupa dilaporkan oleh Lee et al. (2024) yang menggunakan data Rekam Medis Elektronik (RME) di Korea Selatan, yang mana XGBoost dan LightGBM menunjukkan akurasi tertinggi.

Konsistensi ini menunjukkan bahwa kedua varian tersebut merupakan kandidat utama untuk dikaji lebih lanjut dalam pelayanan kesehatan di Indonesia.

Meskipun XGBoost (Chen and Guestrin, 2016) dan LightGBM (Ke et al., 2017) berbagi prinsip dasar *gradient boosting*, keduanya memiliki perbedaan arsitektural yang fundamental. XGBoost menggunakan strategi *level-wise* yang membangun pohon secara seimbang, sedangkan LightGBM mengadopsi pendekatan *leaf-wise* yang memilih daun dengan potensi penurunan *loss* terbesar. Selain itu, LightGBM mengimplementasikan *Gradient-based One-Side Sampling* (GOSS) dan *Exclusive Feature Bundling* (EFB) untuk efisiensi pada data berdimensi tinggi. Ke et al. (2017) melaporkan bahwa kombinasi kedua teknik tersebut mampu mempercepat proses pelatihan hingga lebih dari 20 kali lipat dibandingkan implementasi GBDT konvensional, dengan konsumsi memori yang lebih rendah dan akurasi yang hampir setara. Dalam prediksi LOS, Zeng (2022) juga menunjukkan bahwa LightGBM mampu memberikan performa prediktif yang kompetitif pada data pasien rawat inap. Namun, Florek and Zagdański (2023) mengingatkan bahwa tanpa *hyperparameter tuning* yang memadai, performa *gradient boosting* cenderung tidak stabil. Oleh karena itu, optimasi menggunakan *Bayesian Optimization* menjadi metode yang sangat efisien untuk mencapai performa model prediksi klinis yang optimal (Meaney et al., 2025).

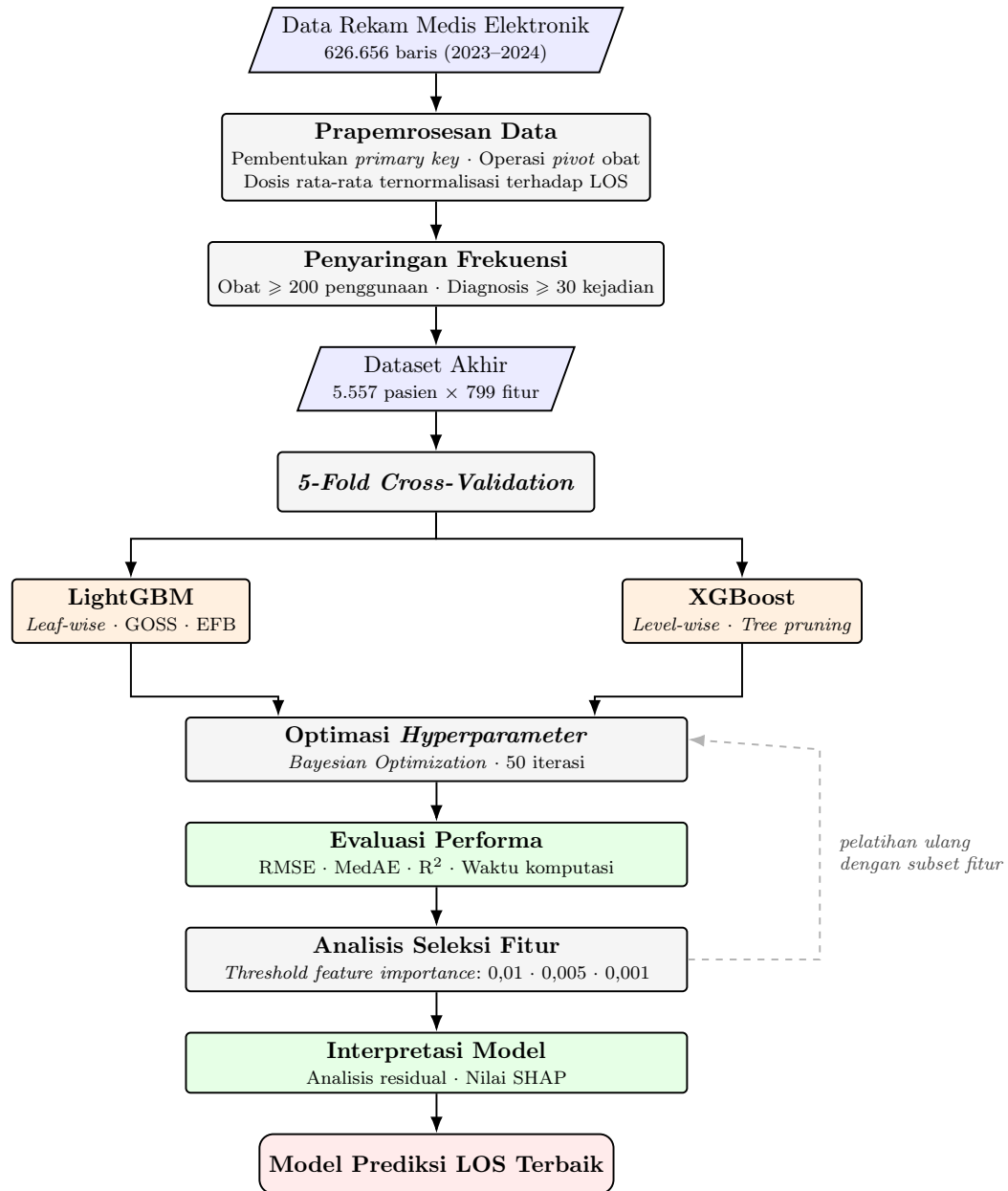
Sekalipun literatur terdahulu telah memberikan kontribusi yang bermakna, masih terdapat beberapa celah penelitian (*research gap*) yang mendasar. Pertama, sebagian besar studi LOS berbasis *gradient boosting* dilakukan di negara maju dengan karakteristik epidemiologis yang berbeda dari Indonesia. Kedua, belum terdapat studi komprehensif yang membandingkan XGBoost dan LightGBM pada data RME berdimensi tinggi yang mencakup ratusan fitur konsumsi obat, padahal pola tersebut merupakan indikator pengganti (*proxy*) yang kuat untuk durasi rawat inap (Bardak and Tan, 2021). Ketiga, penelitian prediksi LOS menggunakan *machine learning* di Indonesia masih sangat terbatas. Berdasarkan celah penelitian tersebut, penelitian ini bertujuan membandingkan kinerja XGBoost dan LightGBM dalam memprediksi LOS menggunakan data RME dari rumah sakit rujukan di Indonesia. Kedua model dioptimasi dengan *Bayesian Optimization* dan dievaluasi menggunakan metrik RMSE, R^2 , serta MedAE guna mengantisipasi variasi LOS yang luas (*outlier*), serta dilengkapi analisis SHAP untuk aspek interpretabilitas model.

2. Metode

Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1. Tahapan penelitian diawali dengan pengumpulan data Rekam Medis Elektronik (RME), dilanjutkan prapemrosesan yang mencakup pembentukan identitas unik, transformasi *pivot* data obat, serta normalisasi dosis terhadap durasi rawat inap. Data kemudian disaring berdasarkan ambang batas frekuensi untuk mengurangi *sparsity*, sehingga menghasilkan dataset akhir. Kedua model, yaitu LightGBM dan XGBoost, dilatih secara terpisah menggunakan skema *5-fold cross-validation* dengan optimasi *hyperparameter* melalui *Bayesian Optimization*. Performa kedua model dievaluasi menggunakan metrik RMSE, MedAE, R^2 , dan waktu komputasi, kemudian dilanjutkan dengan analisis seleksi fitur serta interpretasi model melalui analisis residual dan nilai SHAP.

2.1. Pengumpulan dan Prapemrosesan Data

Penelitian ini menggunakan data rekam medis elektronik (RME) yang telah dianonimkan dari sebuah rumah sakit rujukan tipe A di Indonesia selama periode dua tahun, mulai dari Januari 2023 hingga Desember 2024. Data yang dikumpulkan mencakup informasi demografis, diagnosis,



Gambar 1: Diagram Alur Penelitian

resep obat, serta lama rawat inap (*Length of Stay*/LOS). Dataset awal terdiri dari 626.656 baris data, dengan setiap entri merepresentasikan satu jenis obat yang diberikan kepada pasien selama masa perawatan.

Guna melindungi privasi pasien sekaligus membedakan kunjungan berulang, identitas unik dihasilkan melalui penggabungan kode pasien dan tanggal rawat inap. Identitas unik ini berfungsi sebagai kunci utama (*primary key*) dalam proses agregasi data. Atribut yang bersifat administratif dan tidak esensial dihapus, sementara fitur klinis tetap dipertahankan guna memberikan gambaran nyata sesuai sistem operasional rumah sakit. Operasi *pivot* diterapkan untuk mentrans-

formasi nama obat menjadi kolom-kolom fitur yang merangkum jumlah konsumsi per pasien, sehingga menghasilkan 7.568 baris dengan 852 fitur obat.

Untuk mengurangi *sparsity* dan menghindari fitur yang terlalu jarang muncul, diterapkan ambang batas frekuensi berupa minimal 200 penggunaan total untuk setiap jenis obat serta minimal 30 kejadian untuk setiap kategori diagnosis. Proses penyaringan ini menyisakan 796 fitur obat yang relevan. Setelah digabungkan dengan 3 fitur demografis dan diagnostik (jenis kelamin, usia, dan diagnosis), diperoleh dataset akhir sebanyak 5.557 pasien dengan total 799 fitur prediktor dan 1 variabel target. Untuk variabel obat, digunakan nilai dosis rata-rata per pasien yang telah dinormalisasi terhadap durasi rawat inap. Langkah ini bertujuan mencegah bias model yang memprediksi LOS semata-mata berdasarkan akumulasi jumlah atau tingginya total dosis obat yang diterima pasien. Karakteristik utama dataset disajikan pada Tabel 1.

Tabel 1: Label dan Fitur Dataset

Kategori	#	Nama	Deskripsi
Label	0	Duration	Total lama rawat inap
Demografi	1	JK	Jenis kelamin pasien
	2	Age	Usia pasien saat admisi
	3	Description	Diagnosis pasien
Obat	4	AHEP 500 ml	Dosis rata-rata AHEP 500 ml
	5	ALBUMINAR 25% 100 ml	Dosis rata-rata ALBUMINAR
	...	...	...
	798	Wida water irrigation 1000	Dosis rata-rata Wida 1000
	799	Zinc 20 tablets	Dosis rata-rata zinc 20 tab

2.2. LightGBM

LightGBM merupakan *framework gradient boosting* yang dikembangkan oleh Microsoft, dirancang untuk kecepatan pelatihan dan efisiensi yang tinggi (Ke et al., 2017). Seperti halnya metode *gradient boosting* lainnya, LightGBM membangun model secara aditif dengan menambahkan pohon keputusan secara sekuensial. Prediksi akhir untuk sampel x_i diperoleh dari penjumlahan kontribusi seluruh pohon (Chen and Guestrin, 2016):

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F} \quad (1)$$

dengan K adalah jumlah total pohon, f_k adalah fungsi prediksi pohon ke- k , dan $\mathcal{F}$ adalah ruang fungsi seluruh pohon regresi. Pada setiap iterasi ke- t , pohon baru f_t dilatih untuk meminimalkan fungsi objektif berikut (Chen and Guestrin, 2016):

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad (2)$$

yang mana l menyatakan fungsi kerugian yang dapat didiferensiasikan, $\hat{y}_i^{(t-1)}$ adalah prediksi kumulatif hingga iterasi ke- $(t-1)$, dan $\Omega(f_t)$ adalah *regularization term* yang mengontrol kompleksitas pohon.

Perbedaan fundamental LightGBM terletak pada strategi pertumbuhan pohonnya. Berbeda dengan pendekatan *level-wise* yang membagi semua node di setiap level, LightGBM menggunakan

strategi *leaf-wise* yang memilih daun dengan penurunan fungsi *loss* terbesar untuk dikembangkan lebih lanjut, sehingga pohon yang terbentuk cenderung lebih dalam dan mampu mencapai akurasi lebih tinggi dengan jumlah iterasi yang lebih sedikit (Ke et al., 2017). Untuk menentukan *split* terbaik, LightGBM menghitung *variance gain* berdasarkan gradien dari fungsi kerugian. Penurunan *loss* untuk setiap kandidat *split* dihitung sebagai (Ke et al., 2017):

$$V_j(d) = \frac{1}{n_O} \left[\frac{\left(\sum_{x_i \in O: x_{ij} \leq d} g_i \right)^2}{n_l^j(d)} + \frac{\left(\sum_{x_i \in O: x_{ij} > d} g_i \right)^2}{n_r^j(d)} \right] \quad (3)$$

dengan g_i merupakan gradien (residual negatif) dari fungsi kerugian untuk sampel ke- i , n_O adalah jumlah total sampel pada node yang bersangkutan, sedangkan $n_l^j(d)$ dan $n_r^j(d)$ masing-masing menyatakan jumlah sampel pada cabang kiri dan kanan setelah *split* pada fitur j di titik d . Pohon keputusan memilih titik *split* $d_j^* = \arg \max_d V_j(d)$ yang memaksimalkan *gain* tersebut.

Pada mekanisme GOSS, *gain* diestimasi menggunakan subset data $A \cup B$, dengan A merupakan himpunan top- $a \times 100\%$ sampel bergradien terbesar dan B adalah sampel acak sebesar $b \times 100\%$ dari sisa data. Estimasi *gain* dihitung sebagai (Ke et al., 2017):

$$\tilde{V}_j(d) = \frac{1}{n} \left[\frac{\left(\sum_{x_i \in A_l} g_i + \frac{1-a}{b} \sum_{x_i \in B_l} g_i \right)^2}{n_l^j(d)} + \frac{\left(\sum_{x_i \in A_r} g_i + \frac{1-a}{b} \sum_{x_i \in B_r} g_i \right)^2}{n_r^j(d)} \right] \quad (4)$$

yang mana koefisien $\frac{1-a}{b}$ berfungsi menormalisasi kembali kontribusi gradien dari sampel kecil agar distribusi data tetap terjaga (Ke et al., 2017).

Selain GOSS, LightGBM mengimplementasikan *Exclusive Feature Bundling* (EFB), yang menggabungkan fitur-fitur yang jarang bernilai non-nol secara bersamaan menjadi satu *bundle*, sehingga kompleksitas *histogram building* berkurang dari $O(\#data \times \#feature)$ menjadi $O(\#data \times \#bundle)$, dengan $\#bundle \ll \#feature$ (Ke et al., 2017). LightGBM juga menggunakan algoritma berbasis histogram untuk diskretisasi fitur kontinu ke dalam k bin diskret, yang secara signifikan mereduksi biaya komputasi pencarian *split* optimal (Ke et al., 2017). Kombinasi strategi *leaf-wise*, GOSS, EFB, dan algoritma histogram menjadikan LightGBM sangat efisien pada data berskala besar dan berdimensi tinggi (Seyyedattar et al., 2024).

2.3. XGBoost

XGBoost (*Extreme Gradient Boosting*) merupakan implementasi *gradient boosting* yang dikenal karena performa dan fleksibilitasnya (Chen and Guestrin, 2016). Algoritma ini menggunakan strategi pertumbuhan pohon *level-wise* dan memiliki fitur-fitur unggulan seperti *tree pruning* yang lebih cerdas serta penanganan nilai hilang secara otomatis. Fungsi objektif XGBoost terdiri dari dua komponen utama, yaitu fungsi kerugian (*loss function*) dan *regularization term* untuk mencegah *overfitting*:

$$\text{Obj} = \sum_{i=1}^m l \left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i) \right) + \Omega(f_t) \quad (5)$$

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (6)$$

dengan l merupakan fungsi kerugian, f_t adalah pohon pada iterasi ke- t , Ω mengontrol kompleksitas model melalui jumlah daun T dan bobot daun w , sementara γ dan λ adalah parameter regularisasi (Chen and Guestrin, 2016).

2.4. Bayesian Optimization

Optimasi *hyperparameter* dilakukan menggunakan *Bayesian Optimization* (BO) dengan 50 iterasi untuk setiap model. BO memanfaatkan model probabilistik *surrogate* untuk memodelkan hubungan antara konfigurasi *hyperparameter* dan performa model, kemudian menggunakan fungsi akuisisi untuk memilih titik evaluasi berikutnya secara efisien (Meaney et al., 2025). Inti metodologi BO didasarkan pada Teorema Bayes:

$$p(f | D) = \frac{p(D | f) p(f)}{p(D)} \quad (7)$$

yang mana $p(f | D)$ menyatakan probabilitas posterior, $p(D | f)$ adalah *likelihood*, dan $p(f)$ merupakan *prior belief* (Rimal and Sharma, 2024). Pencarian *hyperparameter* dilakukan dengan skema *5-fold cross-validation*. Rentang pencarian untuk setiap model disajikan pada Tabel 2.

Tabel 2: Rentang Pencarian *Hyperparameter*

Model	Hyperparameter	Deskripsi	Rentang
LightGBM	n_estimators	Jumlah iterasi <i>boosting</i>	100–500
	learning_rate	Laju pembelajaran	0,01–0,2
	max_depth	Kedalaman maksimum pohon	3–10
	num_leaves	Jumlah daun maksimum	20–150
	min_child_samples	Sampel minimum pada daun	5–30
	subsample	Rasio pengambilan sampel	0,6–1,0
	colsample_bytree	Proporsi fitur per pohon	0,6–1,0
	reg_alpha	Regularisasi L1	0,0–1,0
XGBoost	reg_lambda	Regularisasi L2	0,0–10,0
	n_estimators	Jumlah pohon keputusan	100–500
	learning_rate	Laju pembelajaran	0,01–0,2
	max_depth	Kedalaman maksimum pohon	3–10
	subsample	Fraksi sampel per iterasi	0,6–1,0
	colsample_bytree	Fraksi fitur per iterasi	0,6–1,0

2.5. Evaluasi Model

Evaluasi performa dilakukan menggunakan tiga metrik regresi. *Root Mean Squared Error* (RMSE) mengukur rata-rata besarnya kesalahan prediksi dan sensitif terhadap kesalahan besar. *Median Absolute Error* (MedAE) menghitung median selisih absolut antara prediksi dan nilai aktual, sehingga lebih tahan terhadap *outlier* (Chen et al., 2025). *R-squared* (R^2) mengukur proporsi varians target yang dapat dijelaskan oleh model (Seyed-Djawadi et al., 2026). Ketiga metrik tersebut dirumuskan sebagai berikut:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (8)$$

$$\text{MedAE} = \text{median}_{i=1}^N (|y_i - \hat{y}_i|) \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (10)$$

dengan N menyatakan jumlah total data uji, y_i adalah nilai aktual pada observasi ke- i untuk $i = 1, 2, \dots, N$, $\hat{y}_i$ adalah nilai prediksi model pada observasi ke- i , dan $\bar{y}$ merupakan rata-rata nilai aktual seluruh sampel.

2.6. SHAP

SHAP (*SHapley Additive exPlanations*) digunakan untuk analisis interpretabilitas model. SHAP menghitung kontribusi aditif setiap fitur terhadap prediksi individual berdasarkan teori nilai Shapley dari *cooperative game theory* (Alharbi, 2025):

$$\hat{y}_i = \phi_0 + \sum_{j=1}^p \phi_{ji}, \quad i = 1, 2, \dots, N \quad (11)$$

dengan p menyatakan jumlah total fitur prediktor yang digunakan dalam model, N adalah jumlah observasi, ϕ_0 merupakan nilai dasar prediksi (*base value*), yakni rata-rata prediksi model pada seluruh data latih, dan ϕ_{ji} adalah nilai SHAP fitur ke- j untuk observasi ke- i . Kontribusi fitur j dihitung berdasarkan efek marjinal penambahan fitur tersebut ke seluruh kombinasi subset fitur lainnya:

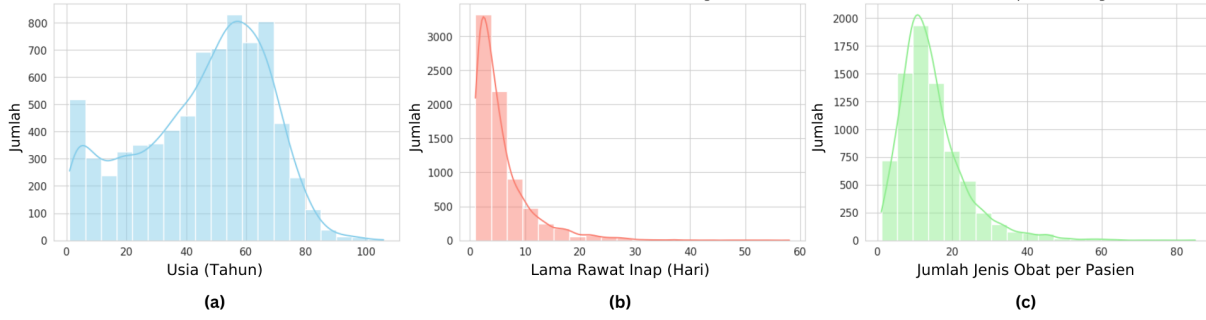
$$\Phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f_x(S \cup \{j\}) - f_x(S)] \quad (12)$$

yang mana F merupakan himpunan semua fitur dengan $|F| = p$, S adalah subset fitur tanpa fitur j , dan $f_x(\cdot)$ menyatakan fungsi prediksi model yang dievaluasi pada subset fitur tertentu.

3. Hasil dan Pembahasan

3.1. Analisis Data Eksploratif

Komposisi demografis menunjukkan keseimbangan yang cukup baik, dengan 50,98% pasien perempuan dan 49,02% laki-laki, sebagaimana ditunjukkan pada Gambar 2. Sebaran usia pasien didominasi oleh kelompok usia 55–60 tahun. Distribusi LOS menunjukkan pola *right-skewed*, dengan median sekitar 4 hari, namun terdapat sejumlah kecil pasien dengan rawat inap hingga 58 hari. Sebagian besar pasien menerima kombinasi 12 hingga 16 jenis obat selama masa rawat inap.



Gambar 2: Distribusi Demografis Pasien: (a) Usia, (b) Lama Rawat Inap (LOS), dan (c) Jumlah Obat Selama Perawatan

3.2. Evaluasi Performa Model

Setelah melalui proses optimasi *hyperparameter* menggunakan *Bayesian Optimization*, dilakukan evaluasi terhadap kedua model. Perhitungan seluruh metrik dilakukan menggunakan Persamaan (8), (9), dan (10) yang telah dijabarkan pada Subbab 2.5. Hasil evaluasi pada data latih dan data uji disajikan pada Tabel 3 dan Tabel 4.

Tabel 3: Evaluasi Model pada Data Latih (799 Fitur)

Model	RMSE	MedAE	R^2
LightGBM	1,24	0,58	0,94
XGBoost	1,25	0,62	0,94

Tabel 4: Evaluasi Model pada Data Uji (799 Fitur)

Model	RMSE	MedAE	R^2	Train (s)	Test (s)
LightGBM	2,31	0,85	0,79	1,83	0,05
XGBoost	2,43	0,83	0,77	8,23	0,08

Berdasarkan Tabel 3, kedua model menunjukkan kemampuan *fitting* yang sangat baik pada data latih dengan nilai R^2 identik sebesar 0,94. LightGBM mencatat RMSE terendah (1,24) dibandingkan XGBoost (1,25), yang mengindikasikan bahwa kedua model mampu menangkap pola hubungan antara fitur dan target secara efektif.

Namun, evaluasi pada data uji (Tabel 4) menunjukkan dinamika yang berbeda. LightGBM mempertahankan keunggulannya dengan RMSE 2,31 dan R^2 0,79, sementara XGBoost menunjukkan performa sedikit lebih rendah dengan RMSE 2,43 dan R^2 0,77. Dari sisi efisiensi komputasi, LightGBM menunjukkan keunggulan signifikan dengan waktu pelatihan 1,83 detik, sekitar 4,5 kali lebih cepat dibandingkan XGBoost (8,23 detik). Hasil ini konsisten dengan karakteristik arsitektural LightGBM yang menggunakan algoritma histogram dan strategi *leaf-wise* untuk efisiensi pada data berdimensi tinggi (Ke et al., 2017).

Penurunan performa dari data latih ke data uji pada kedua model merupakan indikasi wajar dari proses generalisasi. Meskipun XGBoost dikenal memiliki fleksibilitas tinggi dalam memod-

elkan hubungan non-linear, hasil evaluasi menunjukkan bahwa model ini menghasilkan kesalahan prediksi yang sedikit lebih besar pada data yang belum pernah dilihat sebelumnya. Dengan demikian, LightGBM dapat dianggap sebagai model yang paling seimbang dalam penelitian ini, baik dari segi akurasi maupun efisiensi komputasi.

3.3. Analisis Seleksi Fitur

Untuk mengevaluasi dampak reduksi dimensi terhadap performa kedua model, dilakukan eksperimen menggunakan subset fitur yang diseleksi berdasarkan *feature importance* dengan tiga nilai *threshold*, yaitu 0,01 (23 fitur), 0,005 (47 fitur), dan 0,001 (137 fitur). Hasil evaluasi pada data uji untuk setiap skenario disajikan pada Tabel 5.

Tabel 5: Perbandingan Performa Model pada Data Uji dengan Berbagai Skenario Seleksi Fitur

Threshold	Model	Fitur	RMSE	MedAE	R ²	Train (s)
–	LightGBM	799	2,31	0,85	0,79	1,83
	XGBoost	799	2,43	0,83	0,77	8,23
0,001	LightGBM	137	2,37	0,87	0,78	0,46
	XGBoost	137	2,53	0,88	0,76	1,22
0,005	LightGBM	47	2,47	0,90	0,76	0,59
	XGBoost	47	2,62	0,98	0,74	1,00
0,01	LightGBM	23	2,69	1,05	0,72	0,28
	XGBoost	23	2,87	1,10	0,68	0,43

Hasil menunjukkan pola yang konsisten, yakni LightGBM mengungguli XGBoost di seluruh skenario seleksi fitur. Pada *threshold* 0,01 (23 fitur), kedua model mengalami penurunan performa yang signifikan. Nilai R² LightGBM turun dari 0,79 menjadi 0,72, sedangkan XGBoost turun dari 0,77 menjadi 0,68. Kondisi ini mengindikasikan hilangnya informasi krusial akibat seleksi fitur yang terlalu agresif. Meski demikian, efisiensi waktu meningkat drastis, yaitu waktu pelatihan LightGBM turun dari 1,83 detik menjadi 0,28 detik (reduksi 84,7%).

Pada *threshold* 0,001 (137 fitur), performa mulai mendekati kondisi fitur lengkap. LightGBM mencapai R² 0,78, hanya turun 0,01 dari kondisi fitur lengkap, dengan waktu pelatihan 0,46 detik atau berkurang 74,9% dibandingkan fitur lengkap. Hasil ini mengonfirmasi bahwa seleksi fitur dengan *threshold* 0,001 merupakan titik optimal yang mampu mereduksi 82,9% fitur sambil mempertahankan akurasi prediksi. Pada seluruh skenario, LightGBM mempertahankan keunggulan atas XGBoost di setiap konfigurasi, baik dalam akurasi maupun kecepatan komputasi.

3.4. Perbandingan dengan Penelitian Terdahulu

Untuk menempatkan hasil penelitian ini dalam cakupan yang lebih luas, Tabel 6 menyajikan perbandingan dengan studi terdahulu.

Tabel 6: Perbandingan dengan Penelitian Terdahulu

Referensi	Model	Fitur	RMSE	MedAE	R ²
Penelitian ini	LightGBM	Diagnosis, demografi, obat	2,31	0,85	0,79
Jain et al. (2024)	Linear Reg.	Diagnosis, demografi	–	–	0,42
Peng and Gao (2025)	XGBoost	Klinis, demografi	3,21	–	–
Boff Medeiros et al. (2025)	Random Forest	Demografi, klinis, patologis	8,05	–	0,65

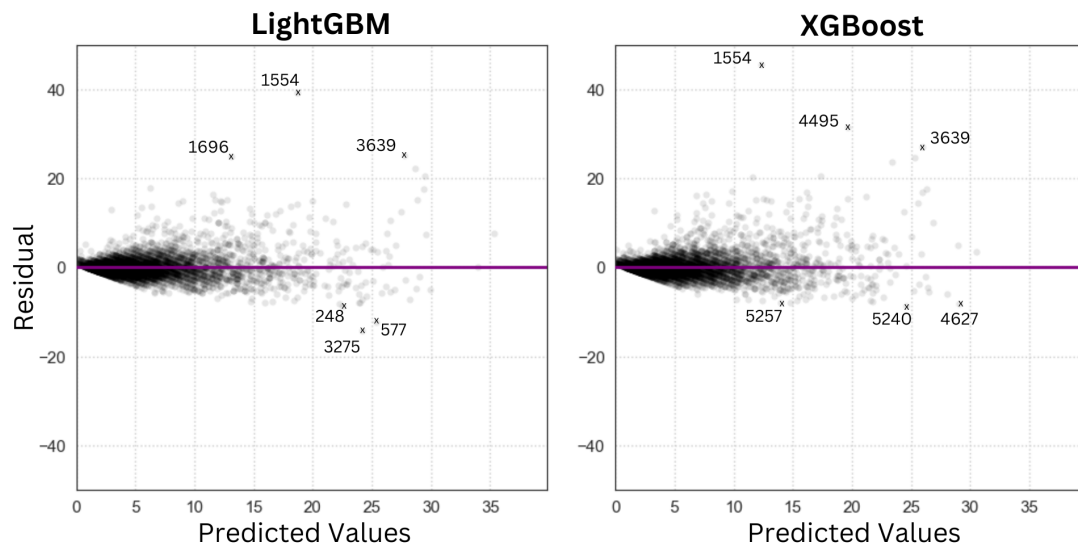
Berdasarkan penelitian [Jain et al. \(2024\)](#) yang menggunakan regresi linier pada dataset SPARCS dan memperoleh R^2 sebesar 0,42, model LightGBM dalam penelitian ini menunjukkan peningkatan substansial dengan R^2 0,79. Hal ini mengonfirmasi keunggulan pendekatan *gradient boosting* dibandingkan model linier pada data klinis yang kompleks, terutama karena kemampuannya menangkap interaksi non-linear antarfitur obat yang tidak dapat direpresentasikan oleh model linier.

Selanjutnya, [Peng and Gao \(2025\)](#) yang menggunakan XGBoost dengan seleksi fitur berbasis *mutual information* melaporkan RMSE sebesar 3,21 pada prediksi LOS. Model LightGBM dalam penelitian ini mencatat RMSE yang lebih rendah (2,31), yang mengindikasikan bahwa pemanfaatan fitur konsumsi obat sebagai indikator pengganti intensitas perawatan memberikan daya prediktif tambahan dibandingkan penggunaan fitur klinis dan demografis semata.

Keunggulan model ini semakin terlihat jika ditinjau dari penelitian [Boff Medeiros et al. \(2025\)](#), yang mengandalkan Random Forest tunggal untuk memprediksi LOS pasien pediatrik dan mencatat RMSE 8,05 serta R^2 0,65. Model LightGBM pada penelitian ini menunjukkan RMSE yang jauh lebih rendah (2,31) dan R^2 yang lebih tinggi (0,79), sehingga menegaskan bahwa pemilihan algoritma *gradient boosting* yang tepat berperan krusial dalam meningkatkan akurasi prediksi. Perlu dicatat bahwa perbandingan ini bersifat indikatif karena setiap studi menggunakan populasi pasien, rentang LOS, dan himpunan fitur yang berbeda, sehingga selisih nilai metrik tidak dapat dimaknai sebagai superioritas mutlak.

3.5. Analisis Residual

Validasi model dilakukan melalui analisis residual untuk memeriksa konsistensi kesalahan prediksi, sebagaimana ditunjukkan pada Gambar 3.

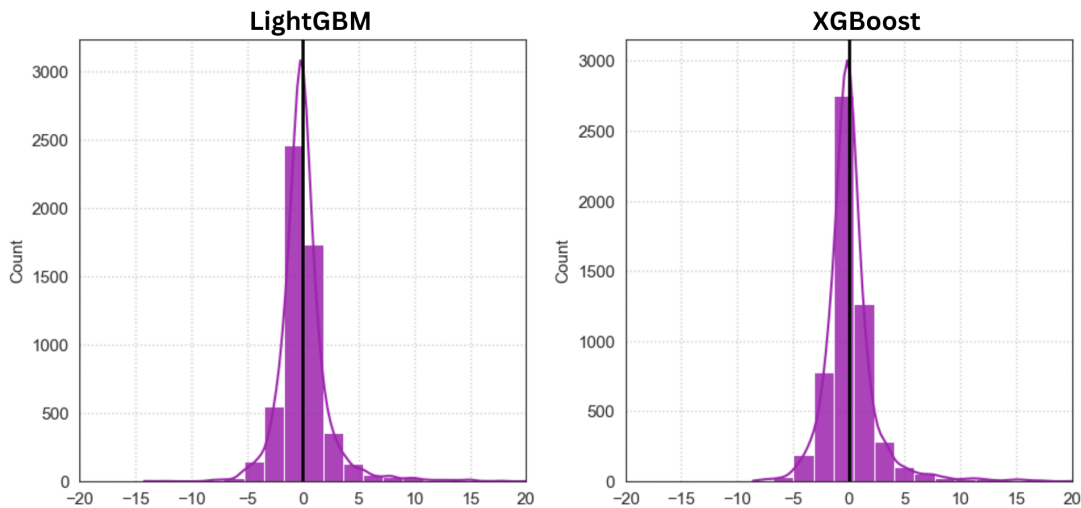


Gambar 3: Plot Residual Model LightGBM dan XGBoost terhadap Nilai Prediksi LOS

Gambar 3 menyajikan sebaran residual, yaitu selisih antara nilai aktual dan nilai prediksi, terhadap nilai prediksi LOS untuk kedua model. Secara teori, model regresi yang baik ditandai

dengan residual yang tersebar acak di sekitar garis nol tanpa membentuk pola tertentu, karena hal tersebut menunjukkan tidak adanya bias sistematis dalam prediksi. Pada LightGBM, sebaran residual tampak lebih terkonsentrasi di sekitar garis nol, terutama pada rentang prediksi 0 hingga 15 hari yang merupakan mayoritas kasus. Sebaliknya, XGBoost memperlihatkan sebaran residual yang lebih lebar dan variansi yang lebih besar, sejalan dengan nilai R^2 yang lebih rendah (0,77 dibandingkan 0,79 pada LightGBM). Kedua model juga menunjukkan pola *heteroscedasticity* ringan, yakni variansi residual cenderung membesar seiring meningkatnya nilai prediksi, yang mengindikasikan bahwa ketidakpastian prediksi meningkat pada kasus dengan durasi rawat inap yang panjang. Titik-titik ekstrem dengan label 1554, 1696, dan 3639 pada LightGBM serta 1554, 4495, dan 3639 pada XGBoost teridentifikasi sebagai residual positif tertinggi, yang menandakan terjadinya *under-prediction* pada pasien dengan LOS aktual yang sangat panjang.

Analisis histogram residual pada Gambar 4 mengonfirmasi hasil tersebut. Distribusi residual LightGBM menunjukkan kurva yang lebih ramping dan terpusat di nol, yang menandakan akurasi dan konsistensi prediksi yang lebih tinggi. Kedua model mengalami *under-prediction* pada *outlier* ekstrem, khususnya pada pasien dengan LOS aktual 53 dan 58 hari yang masing-masing hanya diwakili oleh satu pasien. Kondisi ini menegaskan tantangan model dalam mempelajari pola *rare events* akibat minimnya representasi data ekstrem selama proses pelatihan.

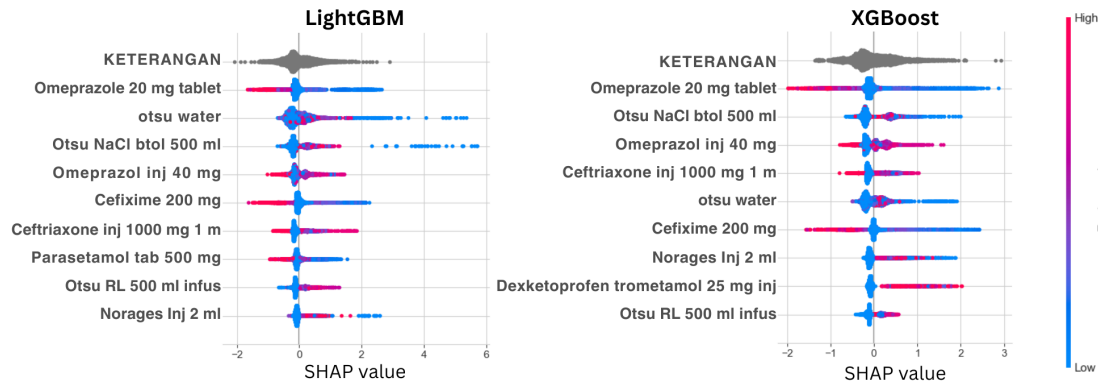


Gambar 4: Histogram Distribusi Residual Model LightGBM dan XGBoost

3.6. Analisis SHAP

Analisis SHAP mengungkapkan kontribusi setiap fitur terhadap prediksi LOS pada kedua model, sebagaimana ditunjukkan pada Gambar 5. Obat-obatan seperti Omeprazole 20 mg tablet, Omeprazole injeksi 40 mg, Otsu NaCl 500 ml, Otsu Water, Ceftriaxone injeksi 1000 mg, dan Cefixime 200 mg muncul sebagai kontributor signifikan pada kedua model dengan distribusi nilai SHAP yang luas. Sebagai inhibitor pompa proton, omeprazole digunakan untuk pencegahan atau pengobatan gangguan gastrointestinal. Nilai SHAP positifnya mengindikasikan bahwa dosis yang

lebih tinggi berkorelasi dengan LOS yang lebih lama, kemungkinan disebabkan oleh kebutuhan perawatan suportif yang lebih intensif.



Gambar 5: Plot SHAP Sepuluh Fitur Teratas terhadap Prediksi LOS pada Kedua Model

Otsu NaCl 500 ml merupakan larutan natrium klorida isotonik yang diberikan secara intravena, yaitu melalui pemasangan infus yang mengalirkan cairan langsung ke dalam pembuluh darah vena pasien. Pemberian ini bertujuan menggantikan kehilangan cairan tubuh, menjaga keseimbangan elektrolit, serta mempertahankan volume darah pada pasien yang tidak mampu memenuhi kebutuhan cairan secara oral. Fitur ini menunjukkan distribusi SHAP yang luas, yang mengindikasikan bahwa kebutuhan cairan intravena dalam jumlah tinggi berkorelasi dengan perpanjangan LOS, karena umumnya mencerminkan kondisi pasien yang memerlukan pemantauan dan dukungan hemodinamik berkelanjutan. Selanjutnya, Ceftriaxone dan Cefixime sebagai antibiotik golongan sefalosporin menunjukkan nilai SHAP positif, yang mencerminkan bahwa pasien dengan infeksi yang memerlukan terapi antibiotik intensif cenderung memiliki masa rawat yang lebih lama. Temuan ini memvalidasi hubungan yang kuat antara fitur konsumsi obat dan variabel LOS, konsisten dengan [Bardak and Tan \(2021\)](#) yang menunjukkan bahwa representasi obat klinis dapat meningkatkan akurasi prediksi LOS. Fitur diagnosis muncul pada kedua model dengan kontribusi yang bergantung pada arsitektur model, namun terbukti kurang menonjol dibandingkan fitur obat kuantitatif.

4. Kesimpulan

Penelitian ini telah berhasil mengimplementasikan dan mengevaluasi kinerja dua algoritma *machine learning* berbasis *gradient boosting*, yaitu LightGBM dan XGBoost, dalam memprediksi lama rawat inap pasien menggunakan data Rekam Medis Elektronik. Kedua model dioptimasi menggunakan *Bayesian Optimization* dan dievaluasi dengan skema yang sama sehingga perbandingan kinerja bersifat adil.

Berdasarkan hasil evaluasi pada data uji, LightGBM tercatat sebagai algoritma dengan kinerja terbaik, mencapai R^2 sebesar 0,79 dengan RMSE 2,31 dan MedAE 0,85. Selain unggul dalam akurasi, LightGBM juga menunjukkan keunggulan signifikan dari sisi efisiensi komputasi dengan waktu pelatihan 1,83 detik, sekitar 4,5 kali lebih cepat dibandingkan XGBoost (8,23 detik). Sebaliknya, meskipun XGBoost dikenal memiliki fleksibilitas tinggi, hasil evaluasi menunjukkan bahwa kompleksitas model yang lebih tinggi tidak selalu berbanding lurus dengan peningkatan kemampuan generalisasi pada data klinis yang heterogen.

Analisis seleksi fitur menunjukkan bahwa pengurangan fitur dari 799 menjadi 137 (*threshold* 0,001) mampu mempertahankan akurasi LightGBM (R^2 0,78) sekaligus mereduksi waktu komputasi sebesar 74,9%. Hasil ini menunjukkan bahwa seleksi fitur berbasis *feature importance* efektif dalam menyeimbangkan akurasi dan efisiensi pada data RME berdimensi tinggi. Analisis SHAP mengungkapkan bahwa fitur konsumsi obat, khususnya penggunaan antibiotik injeksi, cairan intravena, dan analgesik, memiliki kontribusi lebih dominan dibandingkan fitur diagnostik dalam memprediksi LOS.

Hasil penelitian menunjukkan bahwa algoritma *gradient boosting* tunggal yang dioptimasi dengan baik, khususnya LightGBM, merupakan pendekatan yang sangat efektif untuk memprediksi LOS pada data RME terstruktur berdimensi tinggi. Kombinasi antara akurasi prediksi yang baik dan efisiensi komputasi menjadikan LightGBM sebagai kandidat kuat untuk diintegrasikan ke dalam sistem pendukung keputusan klinis di rumah sakit, khususnya pada lingkungan yang membutuhkan respons cepat dan pemrosesan data berskala besar.

Penelitian selanjutnya dapat mengeksplorasi apakah penggabungan kedua algoritma melalui arsitektur *ensemble*, baik dengan pendekatan *bagging* yang melatih beberapa model secara paralel pada subset data hasil *bootstrap* maupun *stacking* yang menggunakan *meta-learner* untuk menggabungkan prediksi antarmodel, mampu memberikan peningkatan performa yang signifikan di atas model tunggal terbaik. Selain itu, integrasi data tambahan seperti hasil laboratorium dan tanda-tanda vital harian berpotensi mempertajam prediksi, terutama untuk kasus-kasus pasien dengan kondisi kompleks.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Program Studi S1 Sains Data Universitas Negeri Surabaya serta seluruh pihak yang telah memberikan dukungan dan masukan selama pelaksanaan penelitian ini.

References

- Alharbi, H. S. (2025). Interpretable and calibrated XGBoost framework for risk-informed probabilistic prediction of slope stability. *Sustainability*, 17(22):10122.
- Almeida, G., Correia, F. B., Borges, A. R., and Bernardino, J. (2024). Hospital length-of-stay prediction using machine learning algorithms—a literature review. *Applied Sciences*, 14(22):10523.
- Alvianto, M. J. and Pribadi, F. (2025). Analysis of indonesian case-based groups with global budget pilot project implementation. *Indonesian Journal of Innovation Studies*, 26(4).
- Bardak, B. and Tan, M. (2021). Using clinical drug representations for improving mortality and length of stay predictions. *arXiv preprint arXiv:2110.08918*.

- Boff Medeiros, N., Fogliatto, F. S., Karla Rocha, M., and Tortorella, G. L. (2025). Predicting the length-of-stay of pediatric patients using machine learning algorithms. *International Journal of Production Research*, 63(2):483–496.
- Chen, T. and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM.
- Chen, Y., Hossein, M., and Zhou, J. (2025). Development of the Optuna-NGBoost-SHAP model for estimating ground settlement during tunnel excavation. *Underground Space*, 24:60–78.
- Fitri, D. E. and Sundari, S. (2023). The cost difference of hospital rates and indonesia case base grup (ina cbgs) rates of inpatient with national health insurance scheme at private hospital. *Jurnal Aisyah: Jurnal Ilmu Kesehatan*, 8(2).
- Florek, P. and Zagdański, A. (2023). Benchmarking state-of-the-art gradient boosting algorithms for classification. *arXiv preprint arXiv:2305.17094*.
- Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232.
- Hirani, R., Podder, D., Stala, O., Mohebpour, R., Tiwari, R. K., and Etienne, M. (2025). Strategies to reduce hospital length of stay: Evidence and challenges. *Medicina*, 61(5):922.
- Jain, R., Singh, M., Rao, A. R., and Garg, R. (2024). Predicting hospital length of stay using machine learning on a large open health dataset. *BMC Health Services Research*, 24(1):860.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- Lee, H., Kim, S., Moon, H.-W., Lee, H.-Y., Kim, K., Jung, S. Y., and Yoo, S. (2024). Hospital length of stay prediction for planned admissions using observational medical outcomes partnership common data model: Retrospective study. *Journal of Medical Internet Research*, 26:e59260.
- Meaney, C., Wang, X., Guan, J., and Stukel, T. A. (2025). Comparison of methods for tuning machine learning model hyper-parameters: with application to predicting high-need high-cost health care users. *BMC Medical Research Methodology*, 25(1):134.
- Mengistu, A. K., Getinet, K., Alemayehu, T., Yeneakal, K. A., Gedefaw, A. E., Teym, A., Endalew, B., and Assaye, B. T. (2026). Machine learning predicts prolonged patient length of stay in a resource constrained ethiopian hospital. *Discover Artificial Intelligence*, 6(1):88.
- Peng, B. and Gao, S. (2025). Prediction of hospital length of stay: leveraging ensemble tree models and intelligent feature selection. *Journal of Hospital Management and Health Policy*, 9:13.
- Rimal, Y. and Sharma, N. (2024). Hyperparameter optimization: a comparative machine learning model analysis for enhanced heart disease prediction accuracy. *Multimedia Tools and Applications*, 83:55091–55107.
- Seyed-Djawadi, M. H., Nikoo, M. R., Shahmiri, A., Siadatmousavi, S. M., Al-Saadi, S., Sana, A., and Bruss, G. (2026). Blind-flux, bathymetry-enhanced stacking ensemble for rapid wave-energy flux prediction along Oman’s coast. *Renewable Energy*, 256.
- Seyyedattar, M., Zendeboudi, S., Ghamartale, A., and Afshar, M. (2024). Advancing hydrogen storage predictions in metal-organic frameworks: A comparative study of LightGBM and random forest models with data enhancement. *International Journal of Hydrogen Energy*,

69:158–172.

- Stone, K., Zwiggelaar, R., Jones, P., and Mac Parthaláin, N. (2022). A systematic review of the prediction of hospital length of stay: Towards a unified framework. *PLOS Digital Health*, 1(4):1–38.
- Zelege, A. J., Palumbo, P., Tubertini, P., Miglio, R., and Chiari, L. (2023). Machine learning-based prediction of hospital prolonged length of stay admission at emergency department: a Gradient Boosting algorithm analysis. *Frontiers in Artificial Intelligence*, 6:1179226.
- Zeng, X. (2022). Length of stay prediction model of indoor patients based on light gradient boosting machine. *Computational Intelligence and Neuroscience*, 2022:9517029.