

Perbandingan Kinerja Algoritma Gradient Boosting dan LightGBM pada Prediksi Diabetes Menggunakan SMOTE dan Hyperparameter Tuning Grid Search

Wawan Kurniawan^{1*}, Lisnawanty², Kartika Handayani³

^{1,2,3}Fakultas Teknik dan Informatika, Program Studi Informatika Kampus Kota Pontianak, Universitas Bina Sarana Informatika, Pontianak, Indonesia

Email: ¹15220476@bsi.ac.id, ²lisnawanty.lsy@bsi.ac.id, ³kartika.kth@bsi.ac.id

(*Email Corresponding Author: 15220476@bsi.ac.id)

Received: 10 Juni 2026 | Revision: 20 Agustus 2026 | Accepted: 3 September 2026

Abstrak

Diabetes melitus merupakan penyakit kronis dengan prevalensi yang terus meningkat secara global sehingga diperlukan sistem deteksi dini yang akurat berbasis *machine learning*. Penelitian ini membandingkan kinerja algoritma Gradient Boosting dan LightGBM dalam memprediksi diabetes menggunakan dua dataset publik, yaitu *Diabetes Prediction Dataset* (100.000 sampel) dan *Early Stage Diabetes Risk Prediction Dataset* (520 sampel). Eksperimen dilakukan dalam empat skenario, yaitu *baseline*, SMOTE, *Grid Search*, serta kombinasi SMOTE dan *Grid Search*. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *AUC-ROC* dengan *F1-score* sebagai metrik utama. Hasil penelitian menunjukkan bahwa pada Dataset 1, Gradient Boosting menghasilkan *F1-score* terbaik sebesar 97,01% pada skenario *Baseline* dan *Grid Search*, sedangkan LightGBM memperoleh performa terbaik pada skenario *Grid Search* dengan *F1-score* 97,00%. Pada Dataset 2, LightGBM mencapai *F1-score* 99,04% dan *AUC-ROC* 100,00% pada skenario *Baseline*. Selain itu, penerapan SMOTE tidak memberikan peningkatan performa yang konsisten, sedangkan efektivitas *Grid Search* bergantung pada karakteristik dataset yang digunakan.

Kata kunci: Gradient Boosting; LightGBM; SMOTE; Hyperparameter Tuning; Prediksi Diabetes

Abstract

Diabetes mellitus is a chronic disease with a continuously increasing global prevalence, highlighting the need for accurate early detection systems based on machine learning. This study compares the performance of Gradient Boosting and LightGBM algorithms for diabetes prediction using two public datasets, namely the Diabetes Prediction Dataset (100,000 samples) and the Early Stage Diabetes Risk Prediction Dataset (520 samples). Experiments were conducted under four scenarios: baseline, SMOTE, Grid Search, and a combination of SMOTE and Grid Search. Model performance was evaluated using accuracy, precision, recall, F1-score, and AUC-ROC, with F1-score as the primary evaluation metric. The results show that on Dataset 1, Gradient Boosting achieved the best F1-score of 97.01% under both the Baseline and Grid Search scenarios, while LightGBM achieved its best performance under the Grid Search scenario with an F1-score of 97.00%. On Dataset 2, LightGBM achieved an F1-score of 99.04% and an AUC-ROC of 100.00% under the Baseline scenario. Furthermore, SMOTE did not consistently improve model performance, whereas the effectiveness of Grid Search depended on the characteristics of the dataset.

Keywords: Gradient Boosting; LightGBM; SMOTE; Hyperparameter Tuning; Diabetes Prediction

1. PENDAHULUAN

Diabetes melitus merupakan salah satu penyakit kronis tidak menular yang terus mengalami peningkatan prevalensi secara signifikan di seluruh dunia. Berdasarkan laporan terbaru International Diabetes Federation (IDF) Diabetes Atlas edisi ke-11 tahun 2025, sebanyak 589 juta orang dewasa berusia 20-79 tahun atau setara dengan 1 dari 9 orang dewasa di seluruh dunia hidup dengan diabetes pada tahun 2024, dan diproyeksikan meningkat menjadi 853 juta jiwa pada tahun 2050 [1] [2]. Fakta ini mempertegas bahwa diabetes telah menjadi krisis kesehatan global yang memerlukan penanganan lebih serius dan inovatif [3].

Salah satu tantangan terbesar adalah masih tingginya angka penderita yang tidak terdiagnosis, terutama di negara berpenghasilan rendah dan menengah [3]. Indonesia termasuk negara dengan jumlah penderita diabetes terbesar, dan keterlambatan deteksi dini berpotensi mengakibatkan komplikasi serius seperti penyakit jantung, stroke, dan gagal ginjal [4]. Kondisi ini mendorong kebutuhan mendesak terhadap pengembangan sistem deteksi dini berbasis teknologi.

Perkembangan *machine learning* membuka peluang besar untuk menjawab tantangan tersebut. Berbagai penelitian menunjukkan bahwa algoritma berbasis *boosting* mampu memberikan performa yang sangat baik dalam klasifikasi diabetes. Salah satunya adalah LightGBM yang dilaporkan mampu mencapai akurasi 98,1% dan *AUC* 98,1% pada data klinis pasien diabetes, serta mengungguli beberapa algoritma lain seperti KNN, SVM, Naive Bayes, dan Random Forest [5].

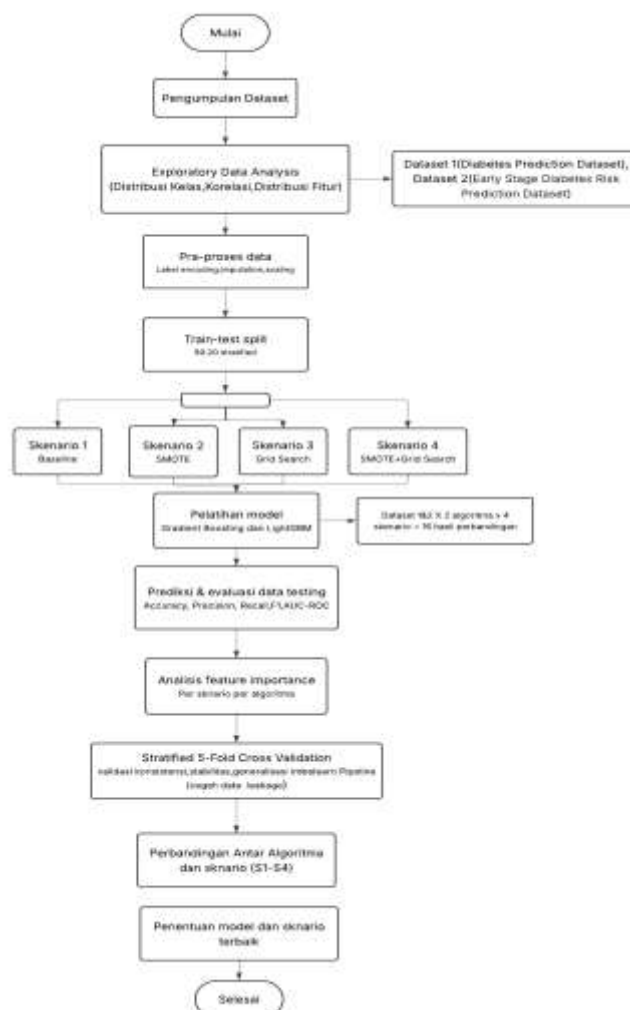
Salah satu tantangan dalam membangun model prediksi diabetes adalah ketidakseimbangan kelas pada dataset. Untuk mengatasinya, teknik Synthetic Minority Over-Sampling Technique (SMOTE) sering digunakan guna meningkatkan representasi kelas minoritas. Namun demikian, efektivitas SMOTE tidak selalu konsisten pada setiap dataset. Beberapa penelitian menunjukkan bahwa penerapan SMOTE dapat meningkatkan sensitivitas model, tetapi pada saat yang sama berpotensi menurunkan akurasi klasifikasi secara keseluruhan [6]. Di sisi lain, optimasi hyperparameter menggunakan Grid Search telah dilaporkan mampu meningkatkan performa model machine learning secara signifikan [7].

Berbagai penelitian juga telah menerapkan teknik optimasi dan ensemble learning untuk meningkatkan prediksi diabetes. Optimasi hyperparameter menggunakan Grid Search dilaporkan mampu meningkatkan performa berbagai algoritma klasifikasi [8], [9]. Pendekatan ensemble menunjukkan hasil yang kompetitif dalam prediksi risiko diabetes [10]. Selain itu, algoritma boosting seperti Gradient Boosting dan LightGBM telah menunjukkan performa yang baik pada berbagai dataset diabetes [11], [12], sedangkan SMOTE banyak digunakan untuk meningkatkan kemampuan model dalam mengenali kelas minoritas [13].

Meskipun sejumlah penelitian telah menggunakan Gradient Boosting atau LightGBM secara terpisah, studi yang secara komprehensif membandingkan keduanya head-to-head dengan mengintegrasikan SMOTE dan Grid Search secara bersamaan masih sangat terbatas. Oleh karena itu, penelitian ini bertujuan membandingkan kinerja Gradient Boosting dan LightGBM menggunakan dua dataset publik dengan empat skenario eksperimen, serta mengevaluasi performa berdasarkan akurasi, presisi, recall, F1-score, dan AUC-ROC guna menghasilkan rekomendasi algoritmik berbasis bukti empiris.

2. METODOLOGI PENELITIAN

Penelitian ini merupakan penelitian eksperimental komputasi yang bertujuan membandingkan kinerja Gradient Boosting dan LightGBM dalam memprediksi diabetes. Alur penelitian terdiri dari lima tahap: pengumpulan dataset, pra-pemrosesan data, penerapan skenario eksperimen, pelatihan dan pengujian model, serta evaluasi. Gambar 1 menyajikan alur keseluruhan penelitian.



Gambar 1. Alur Penelitian

Gambar 1 memperlihatkan tahapan penelitian yang terdiri dari pengumpulan dataset, exploratory data analysis, pra-proses data (label encoding, imputasi, dan scaling), pembagian data latih dan uji (train-test split), penerapan empat skenario eksperimen, pelatihan model, evaluasi performa, analisis feature importance, Stratified 5-Fold Cross Validation, perbandingan hasil antar algoritma dan skenario, hingga penentuan model dan skenario terbaik. Tahapan tersebut dirancang untuk mengevaluasi pengaruh SMOTE dan Hyperparameter Tuning Grid Search terhadap kinerja Gradient Boosting dan LightGBM pada dua dataset diabetes dengan karakteristik yang berbeda.

2.1 Dataset

Penelitian ini menggunakan dua dataset publik. Dataset pertama adalah Diabetes Prediction Dataset (Kaggle) yang memuat data klinis dan demografis 100.000 sampel pasien, meliputi usia, jenis kelamin, BMI, riwayat hipertensi, riwayat penyakit jantung, kebiasaan merokok, kadar HbA1c, dan kadar glukosa darah [4]. Dataset kedua adalah Early Stage Diabetes Risk Prediction Dataset (UCI) yang memuat 520 sampel berbasis gejala klinis seperti polyuria, polydipsia, penurunan berat badan mendadak, dan gejala lainnya [10]. Penggunaan dua dataset dengan ukuran dan karakteristik yang sangat berbeda bertujuan menguji konsistensi dan generalisasi performa kedua algoritma.

Dataset 1 memiliki ketidakseimbangan kelas yang sangat ekstrem pada data training, yaitu 91.5% Non-Diabetes dan hanya 8.5% Diabetes. Dataset 2 memiliki ketidakseimbangan yang lebih moderat, yaitu 61.5% Non-Diabetes dan 38.5% Diabetes. Setelah SMOTE diterapkan, distribusi kedua dataset menjadi seimbang 50:50. Gambar 2 dan 3 menyajikan visualisasi distribusi tersebut.

2.2 Pra-Pemrosesan Data

Tahap pra-pemrosesan meliputi: (1) penanganan missing values menggunakan SimpleImputer strategi median, (2) encoding fitur kategorikal menggunakan Label Encoding, (3) normalisasi fitur numerik menggunakan StandardScaler, dan (4) pembagian data 80:20 dengan stratified splitting. Dataset dibagi dengan rasio 80:20 untuk data training dan testing. Selanjutnya, SMOTE diterapkan pada data training untuk menyeimbangkan distribusi kelas sebelum proses pelatihan model.

2.3 Skenario Eksperimen

Penelitian ini merancang empat skenario eksperimen yang diterapkan pada setiap kombinasi algoritma dan dataset, seperti disajikan pada Tabel 1.

Table 1. Skenario Eksperimen

No	Skenario	SMOTE	HyperParameter Tuning Grid Search
S1	Baseline	Tidak	Tidak
S2	SMOTE	Ya	Tidak
S3	GRID SEARCH	Tidak	Ya
S4	SMOTE + Grid Search	Ya	Ya

2.4 Pelatihan Model

Kedua algoritma diimplementasikan pada lingkungan Python (Google Colab). Gradient Boosting dibangun secara sekuensial dengan mengoreksi residual error dari pohon sebelumnya, sedangkan LightGBM merupakan algoritma berbasis gradient boosting yang mendukung strategi pertumbuhan pohon secara leaf-wise dan menyediakan teknik Gradient-based One-Side Sampling (GOSS) untuk meningkatkan efisiensi komputasi[14]. Parameter yang dioptimasi Grid Search untuk Gradient Boosting mencakup `n_estimators`, `learning_rate`, dan `max_depth` dan `subsample`. Sementara itu, pada LightGBM parameter yang dioptimasi meliputi `n_estimators`, `learning_rate`, `max_depth`, dan `num_leaves`. Proses optimasi dilakukan menggunakan validasi silang Stratified 5-Fold Cross Validation. Skema Stratified 5-Fold Cross Validation yang sama juga digunakan untuk mengevaluasi stabilitas performa model pada skenario dengan hasil terbaik dari masing-masing algoritma, sebagaimana dijelaskan pada 2.5.

2.5 Evaluasi Model

Performa dievaluasi menggunakan akurasi, presisi, recall, F1-score, dan AUC-ROC. F1-score dipilih sebagai metrik utama karena menyeimbangkan presisi dan recall, yang relevan dalam konteks prediksi penyakit di mana kesalahan prediksi positif memiliki konsekuensi klinis serius. AUC-ROC mengukur kemampuan diskriminasi model terlepas dari threshold klasifikasi. Selain evaluasi pada data testing, dilakukan validasi tambahan menggunakan Stratified 5-Fold Cross Validation pada skenario dengan performa terbaik dari setiap algoritma untuk menilai konsistensi dan stabilitas model pada berbagai kombinasi data latih dan data uji. Hasil cross validation dilaporkan dalam bentuk rata-rata dan standar deviasi Accuracy, F1-score, dan AUC-ROC.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil eksperimen komparatif antara Gradient Boosting dan LightGBM pada dua dataset diabetes publik menggunakan empat skenario. Penelitian ini tidak hanya membandingkan kedua algoritma utama, tetapi juga mengevaluasi pengaruh kombinasi SMOTE dan Grid Search pada dataset dengan karakteristik berbeda. Sebagai konteks perbandingan, hasil juga dibandingkan dengan lima algoritma pembanding: AdaBoost, Random Forest, KNN, Decision Tree, dan XGBoost.

3.1 Hasil Eksperimen pada Dataset 1 (Diabetest Prediction Dataset)

Table 2. Hasil evaluasi Gradient Boosting pada dataset Diabetes Prediction Dataset

Skenario	Akurasi	Presisi	Recall	F1-Score	AUC-ROC
Baseline	97.23	97.25	97.23	97.01	97.94
SMOTE	96.67	96.52	96.67	96.54	97.66
Grid Search	97.23	97.25	97.23	97.01	97.94
SMOTE dan Grid Search	97.09	97.03	97.09	96.89	97.87

Table 2 menunjukkan hasil evaluasi Gradient Boosting pada *Diabetes Prediction Dataset*. Skenario Baseline (S1) dan Grid Search (S3) menghasilkan nilai evaluasi yang identik pada seluruh metrik, yaitu accuracy 97.23%, precision 97.25%, recall 97.23%, F1-score 97.01%, dan AUC-ROC 97.94%. Hasil tersebut menunjukkan bahwa proses *Grid Search* tidak memberikan peningkatan performa dibandingkan model baseline. Hal ini disebabkan karena kombinasi *hyperparameter* terbaik yang diperoleh (*learning_rate*=0.1, *max_depth*=3, *n_estimators*=100, *subsample*=1.0) sama dengan konfigurasi default GradientBoostingClassifier dalam ruang pencarian yang digunakan.

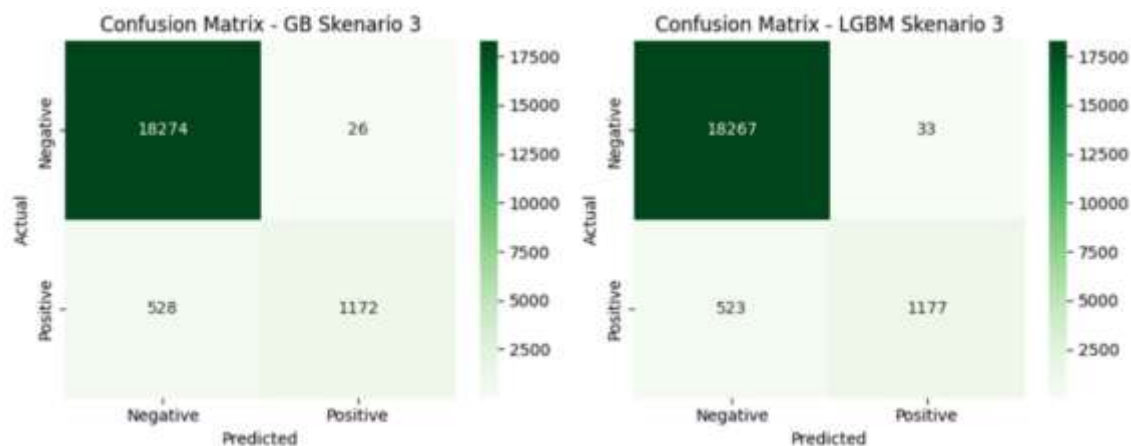
Sebaliknya, penerapan SMOTE pada S2 menurunkan performa model pada seluruh metrik evaluasi. Accuracy menurun dari 97.23% menjadi 96.67%, precision dari 97.25% menjadi 96.52%, recall dari 97.23% menjadi 96.67%, F1-score dari 97.01% menjadi 96.54%, dan AUC-ROC dari 97.94% menjadi 97.66%. Hasil ini menunjukkan bahwa meskipun Dataset 1 memiliki distribusi kelas yang tidak seimbang (91.5% berbanding 8.5%), Gradient Boosting telah mampu mempelajari pola dari data asli secara efektif, sehingga penambahan sampel sintesis melalui SMOTE tidak meningkatkan performa klasifikasi dan justru menyebabkan sedikit penurunan pada seluruh metrik evaluasi.

Table 3. Hasil evaluasi LightGBM pada dataset Diabetes Predictio Dataset

Skenario	Akurasi	Presisi	Recall	F1-Score	AUC-ROC
Baseline	97.17	97.15	97.17	96.95	97.92
SMOTE	97.20	97.19	97.20	96.99	97.87
Grid Search	97.22	97.22	97.22	97.00	97.97
SMOTE dan Grid Search	97.08	97.00	97.08	96.90	97.75

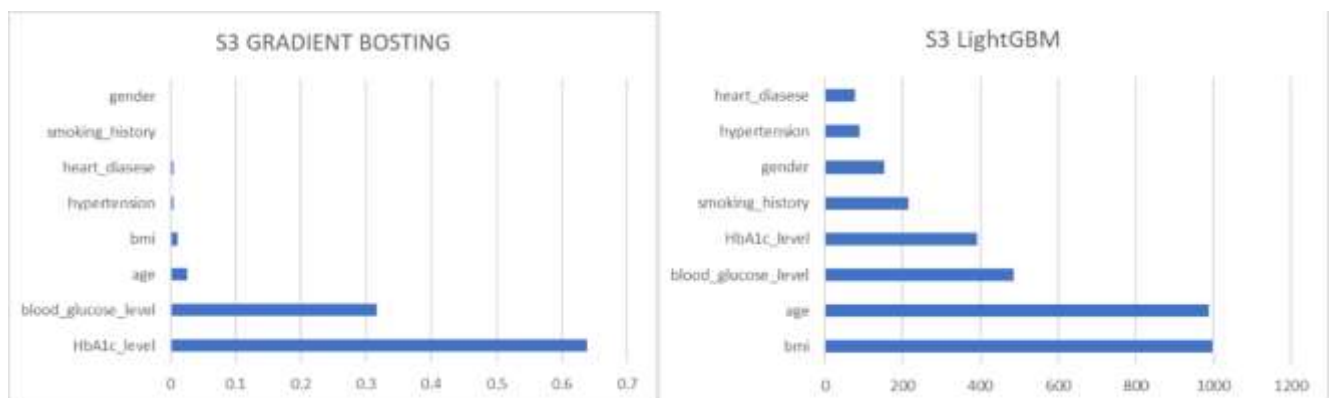
Table 3 menunjukkan bahwa LightGBM menghasilkan performa yang sangat stabil pada seluruh skenario pengujian pada *Diabetes Prediction Dataset*. Variasi performa antar skenario relatif kecil, dengan accuracy berkisar 97.08%–97.22%, precision 97.00%–97.22%, recall 97.08%–97.22%, F1-score 96.90%–97.00%, dan AUC-ROC 97.75%–97.97%. Skenario Grid Search (S3) menghasilkan performa terbaik pada seluruh metrik evaluasi, yaitu accuracy 97.22%, precision 97.22%, recall 97.22%, F1-score 97.00%, dan AUC-ROC 97.97%. Hasil ini menunjukkan bahwa optimasi *hyperparameter* melalui Grid Search mampu memberikan peningkatan performa meskipun relatif kecil dibandingkan skenario lainnya.

Jika dibandingkan secara langsung, Gradient Boosting menghasilkan F1-score sedikit lebih tinggi (97.01%) dibandingkan LightGBM (97.00%), sedangkan LightGBM memperoleh nilai accuracy dan AUC-ROC yang sedikit lebih tinggi. Namun, seluruh perbedaan tersebut hanya berkisar 0.01–0.03 poin persentase, sehingga kedua algoritma dapat dikatakan memiliki performa yang hampir setara pada Dataset 1.



Gambar 2. Confusion Matrix Terbaik Pada setiap model Dataset 1

Confusion matrix Gradient Boosting pada Skenario 3 (Grid Search) yang ditunjukkan pada Gambar 2 memperlihatkan hasil klasifikasi yang sangat baik pada Dataset 1. Model berhasil mengklasifikasikan sebanyak 18.274 data sebagai true negative dan 1.172 data sebagai true positive. Di sisi lain, masih terdapat 26 false positive dan 528 false negative. Jumlah false negative yang lebih tinggi dibandingkan false positive menunjukkan bahwa model masih lebih sering melewati beberapa kasus diabetes daripada salah mengklasifikasikan pasien non-diabetes sebagai diabetes. Kondisi ini sejalan dengan karakteristik Dataset 1 yang memiliki distribusi kelas tidak seimbang sehingga kelas diabetes lebih sulit dikenali dibandingkan kelas non-diabetes.



Gambar 3. Feature Importance Skenario Terbaik dari setiap model pada dataset 1

Berdasarkan Gambar 3, terdapat perbedaan pola feature importance antara kedua algoritma dengan skenario terbaik dari masing-masing model. Gradient Boosting lebih berfokus pada **HbA1c_level** dan **blood_glucose_level** sebagai fitur utama, sedangkan LightGBM memberikan perhatian yang lebih besar pada **BMI** dan **age**. Perbedaan ini menunjukkan bahwa masing-masing algoritma memiliki mekanisme yang berbeda dalam menentukan kontribusi fitur terhadap prediksi diabetes meskipun menggunakan dataset yang sama.

3.2 Hasil Eksperimen pada Dataset 2 (Early Stage Diabetes Risk Prediction Dataset)

Table 4. Hasil evaluasi Gradient Boosting pada dataset 2 (Early Stage Diabetes Risk Prediction Dataset)

Skenario	Akurasi	Presi	Recall	F1-Score	AUC-ROC
Baseline	99.04	99.06	99.04	99.04	99.37
SMOTE	99.04	99.06	99.04	99.04	99.30
Grid Search	98.08	98.17	98.08	98.09	99.96
SMOTE dan Grid Search	98.08	98.17	98.08	98.09	99.96

Table 4 menunjukkan bahwa Gradient Boosting menghasilkan performa yang sangat tinggi pada *Early Stage Diabetes Risk Prediction Dataset*. Skenario Baseline (S1) memberikan performa terbaik pada metrik accuracy (99.04%), precision (99.06%), recall (99.04%), dan F1-score (99.04%), dengan AUC-ROC sebesar 99.37%. Skenario SMOTE (S2)

menghasilkan nilai yang identik pada metrik berbasis klasifikasi, namun AUC-ROC sedikit menurun menjadi 99.30%, sehingga menunjukkan bahwa penerapan SMOTE tidak memberikan peningkatan performa pada dataset ini.

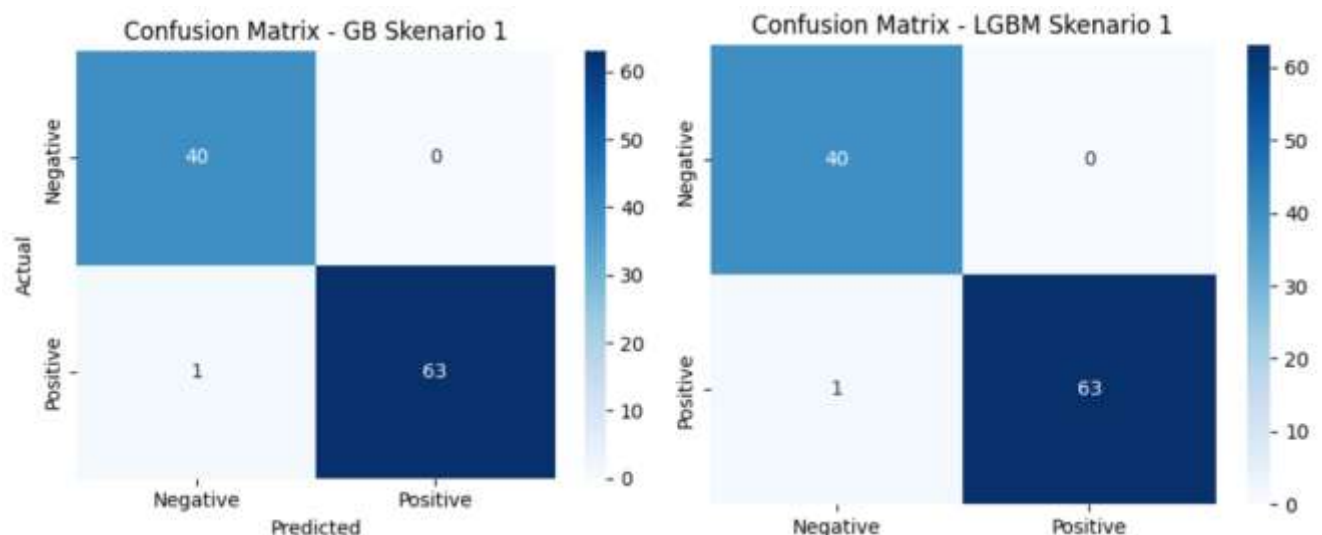
Sebaliknya, skenario Grid Search (S3) dan SMOTE + Grid Search (S4) menghasilkan nilai evaluasi yang identik, yaitu accuracy 98.08%, precision 98.17%, recall 98.08%, F1-score 98.09%, dan AUC-ROC 99.96%. Hasil tersebut menunjukkan bahwa proses *Grid Search* meningkatkan kemampuan diskriminasi model berdasarkan probabilitas, yang tercermin dari peningkatan AUC-ROC, namun diikuti oleh sedikit penurunan pada accuracy dan F1-score. Temuan ini mengindikasikan adanya *trade-off* antara kemampuan pemeringkatan probabilitas dan ketepatan klasifikasi pada dataset yang berukuran relatif kecil.

Table 5. Hasil evaluasi LightGBM pada dataset 2 (Early Stage Diabetes Risk Prediction Dataset)

Skenario	Akurasi	Presisi	Recall	F1-Score	AUC-ROC
Baseline	99.04	99.06	99.04	99.04	100.00
SMOTE	98.08	98.17	98.08	98.09	100.00
Grid Search	97.12	97.32	97.12	97.13	100.00
SMOTE dan Grid Search	98.08	98.17	98.08	98.09	100.00

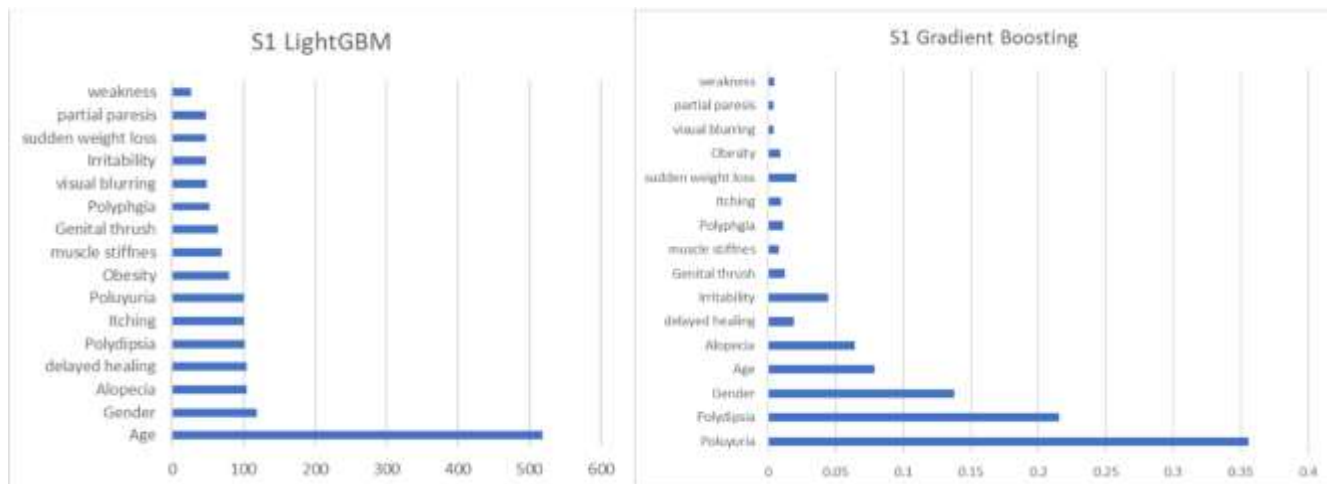
Table 5 menunjukkan bahwa LightGBM mencapai nilai AUC-ROC sempurna (100.00%) pada seluruh skenario pengujian pada *Early Stage Diabetes Risk Prediction Dataset*, yang menunjukkan kemampuan diskriminasi kelas yang sangat tinggi. Namun, performa pada metrik accuracy, precision, recall, dan F1-score masih bervariasi antar skenario. Performa terbaik diperoleh pada Baseline (S1) dengan accuracy 99.04%, precision 99.06%, recall 99.04%, dan F1-score 99.04%, sedangkan nilai terendah terjadi pada Grid Search (S3) dengan accuracy 97.12%, precision 97.32%, recall 97.12%, dan F1-score 97.13%. Hasil ini menunjukkan bahwa proses *Grid Search* tidak meningkatkan performa klasifikasi berdasarkan threshold prediksi, meskipun kemampuan diskriminasi model yang diukur menggunakan AUC-ROC tetap sempurna.

Dibandingkan dengan Gradient Boosting, LightGBM menghasilkan F1-score terbaik yang sama (99.04%), namun unggul pada metrik AUC-ROC dengan nilai 100.00% pada seluruh skenario pengujian. Sementara itu, Gradient Boosting mencapai nilai AUC-ROC tertinggi sebesar 99.96% pada skenario Grid Search (S3) dan SMOTE + Grid Search (S4). Secara keseluruhan, tingginya performa kedua algoritma menunjukkan bahwa pola antar kelas pada dataset berbasis gejala klinis relatif mudah dipisahkan, sehingga kedua model mampu membangun batas keputusan yang efektif.



Gambar 4. Confusion Matrix Terbaik Pada setiap model Dataset 2

Gambar 4 menampilkan confusion matrix skenario terbaik Dataset 2. Kedua algoritma hampir tidak membuat kesalahan klasifikasi, mencerminkan kemudahan pola pemisahan kelas pada dataset ini.



Gambar 5. Feature Importance Sknario Terbaik dari setiap model pada dataset 2

Berdasarkan Gambar 5, terdapat perbedaan pola feature importance antara LightGBM dan Gradient Boosting pada skenario terbaik masing-masing model. Pada LightGBM, fitur Age memiliki nilai feature importance tertinggi sehingga menjadi fitur yang paling berkontribusi dalam proses klasifikasi diabetes. Sementara itu, pada Gradient Boosting, fitur Polyuria merupakan fitur dengan kontribusi terbesar, diikuti oleh Polydipsia, Gender, dan Age. Perbedaan tersebut menunjukkan bahwa kedua model memberikan bobot yang berbeda terhadap fitur-fitur yang digunakan dalam penelitian ini, meskipun keduanya menghasilkan performa klasifikasi yang hampir setara.

3.3 Validasi Stabilitas Model dengan Cross Validation

Untuk memastikan bahwa hasil evaluasi pada data testing tidak hanya bergantung pada satu kali pembagian data, dilakukan validasi tambahan menggunakan Stratified 5-Fold Cross Validation pada skenario dengan performa terbaik dari masing-masing algoritma pada kedua dataset. Tabel 6 menyajikan ringkasan hasil cross validation berupa rata-rata dan standar deviasi.

Table 6. Hasil Stratified 5-Fold Cross Validation pada Skenario Terbaik

Dataset	Model	Sknario Terbaik	CV Accuracy	CV F1-Score	CV AUC-ROC
D1	Gradient Boosting	Baseline & Grid Search (S1&S3)	97,21% ± 0,11%	96,97% ± 0,13%	97,90% ± 0,12%
D1	LightGBM	Grid Search(S3)	97,19% ± 0,10%	96,97% ± 0,12%	97,90% ± 0,13%
D2	Gradient Boosting	Baseline(S1)	96,39% ± 2,16%	96,40% ± 2,14%	98,69% ± 1,38%
D2	LightGBM	Baseline(S1)	95,91% ± 1,81%	95,92% ± 1,80%	98,81% ± 0,97%

Pada Dataset 1, nilai CV Accuracy dan CV F1-Score sangat mendekati hasil pada data testing dengan standar deviasi yang rendah ($\leq 0,13\%$), yang mengonfirmasi bahwa performa tinggi Gradient Boosting maupun LightGBM bersifat stabil dan tidak bergantung pada satu kali pembagian data. Sebaliknya, pada Dataset 2 nilai CV Accuracy secara konsisten lebih rendah dibandingkan hasil pada data testing (Gradient Boosting: 96,39% berbanding 99,04%; LightGBM: 95,91% berbanding 99,04%) disertai standar deviasi yang jauh lebih besar (hingga $\pm 2,16\%$). Perbedaan ini disebabkan oleh ukuran data training yang sangat kecil (416 sampel dibagi 5 fold, ± 83 sampel per fold), sehingga performa pada satu kali pembagian data berpotensi terlalu optimistis dan kurang merepresentasikan kemampuan generalisasi model secara umum.

3.4 Pembahasan

Pertama, Gradient Boosting dan LightGBM terbukti konsisten sebagai dua algoritma teratas di kedua dataset, yang menunjukkan keunggulan pendekatan boosting dalam klasifikasi risiko diabetes. Hasil ini sejalan dengan beberapa penelitian sebelumnya yang juga melaporkan performa unggul algoritma boosting pada prediksi diabetes [5], [12].

Kedua, penerapan SMOTE tidak memberikan peningkatan performa yang konsisten meskipun Dataset 1 memiliki ketidakseimbangan kelas yang tinggi (91,5% : 8,5%). Pada penelitian ini, sebagian besar skenario tidak menunjukkan peningkatan yang berarti setelah penerapan SMOTE. Pada Gradient Boosting bahkan terjadi penurunan pada hampir seluruh metrik evaluasi, sedangkan pada LightGBM perubahan performanya relatif sangat kecil. Temuan ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa penerapan SMOTE tidak selalu meningkatkan performa klasifikasi secara keseluruhan [6]. Meskipun sensitivitas dapat meningkat akibat bertambahnya representasi kelas minoritas, akurasi dan kemampuan generalisasi model tidak selalu membaik. Namun, hasil penelitian ini berbeda dengan penelitian yang melaporkan bahwa penerapan SMOTE pada algoritma Random Forest mampu meningkatkan akurasi dari 86% menjadi 92%, serta meningkatkan F1-score kelas diabetes dari 26% menjadi 92% [15]. Perbedaan hasil tersebut menunjukkan bahwa efektivitas SMOTE sangat dipengaruhi oleh karakteristik dataset, tingkat ketidakseimbangan kelas, ukuran data, serta algoritma yang digunakan. Pada dataset tertentu, SMOTE mampu memperbaiki representasi kelas minoritas dan meningkatkan kemampuan model dalam mendeteksi kasus positif, sedangkan pada dataset lain penambahan sampel sintetis justru dapat menurunkan kemampuan generalisasi model.

Ketiga, Hyperparameter Tuning menggunakan Grid Search memberikan peningkatan performa yang relatif kecil pada Dataset 1, khususnya pada LightGBM, sedangkan pada Dataset 2 justru diikuti penurunan accuracy dan F1-score meskipun nilai AUC-ROC meningkat. Hasil tersebut menunjukkan bahwa efektivitas hyperparameter tuning dipengaruhi oleh karakteristik dataset, terutama ukuran data dan kompleksitas pola yang dipelajari model. Temuan ini mendukung penelitian sebelumnya yang menunjukkan bahwa hyperparameter tuning berperan penting dalam mengoptimalkan performa model machine learning [7].

Keempat, tingginya performa hampir seluruh model pada Dataset 2 (sebagian besar >99%, AUC-ROC mencapai 100,00%) perlu diinterpretasikan secara hati-hati. Dengan hanya 520 sampel dan fitur gejala dikotomis (ya/tidak), batas keputusan antar kelas relatif sederhana sehingga bahkan Decision Tree mampu mencapai F1-score 99,04%. Performa yang sangat tinggi ini lebih mencerminkan karakteristik dataset dibandingkan semata-mata keunggulan algoritma yang digunakan.

4. KESIMPULAN

Penelitian ini membandingkan kinerja Gradient Boosting dan LightGBM dalam memprediksi diabetes menggunakan dua dataset publik melalui empat skenario eksperimen yang mengombinasikan SMOTE dan Hyperparameter Tuning Grid Search. Hasil penelitian menunjukkan bahwa pada Dataset 1, Gradient Boosting menghasilkan performa terbaik yang identik pada skenario Baseline dan Grid Search dengan F1-score 97.01% dan AUC-ROC 97.94%, sedangkan LightGBM mencapai performa terbaik pada skenario Grid Search dengan F1-score 97.00% dan AUC-ROC 97.97%. Pada Dataset 2, LightGBM memperoleh performa terbaik pada skenario Baseline dengan F1-score 99.04% dan AUC-ROC 100.00%. Secara keseluruhan, kedua algoritma menunjukkan performa yang sangat tinggi dan relatif sebanding dalam klasifikasi risiko diabetes. Penerapan SMOTE tidak memberikan peningkatan performa yang konsisten, sedangkan Hyperparameter Tuning menggunakan Grid Search hanya memberikan peningkatan yang relatif kecil pada Dataset 1 dan menghasilkan trade-off pada Dataset 2. Hasil ini menunjukkan bahwa efektivitas kedua teknik tersebut sangat dipengaruhi oleh karakteristik dataset yang digunakan. Selain itu, analisis feature importance menunjukkan bahwa HbA1c_level dan blood_glucose_level merupakan fitur yang paling berpengaruh pada Dataset 1, sedangkan pada Dataset 2 LightGBM lebih banyak memberikan bobot pada fitur Age dan Gradient Boosting pada fitur Polyuria. Temuan ini menunjukkan bahwa kedua algoritma membangun pola pemanfaatan fitur yang berbeda meskipun menghasilkan performa klasifikasi yang hampir setara.

REFERENCES

- [1] I. Genitsaridi *et al.*, “11th edition of the IDF Diabetes Atlas : global , regional , and national diabetes prevalence estimates for 2024 and projections for 2050,” *LANCET Diabetes Endocrinol.*, vol. 14, no. 2, pp. 149–156, 2024, doi: [https://doi.org/10.1016/S2213-8587\(25\)00299-2](https://doi.org/10.1016/S2213-8587(25)00299-2).
- [2] I. D. Federation, “IDF Diabetes Atlas, 11th Edition,” Brussels, 2025. [Online]. Available: <https://diabetesatlas.org/>
- [3] O. Iparraguirre-Villanueva, K. Espinola-Linares, R. O. Flores Castañeda, and M. Cabanillas-Carbonell, “Application of Machine Learning Models for Early Detection and Accurate Classification of Type 2 Diabetes,” *Diagnostics*, vol. 13, no. 14, p. 2383, 2023, [Online]. Available: <https://doi.org/10.3390/diagnostics13142383>
- [4] A. Davinka, S. Depari, K. D. Tania, and P. E. Sevtiyuni, “Penerapan Metode Machine Learning Dan Teknik SMOTE untuk Prediksi Diabetes,” vol. 7, pp. 436–447, 2025, doi: <https://doi.org/10.30865/json.v7i2.9032>.
- [5] D. D. Rufo, T. G. Debelee, A. Ibenthal, and W. G. Negera, “Diagnosis of Diabetes Mellitus Using Gradient Boosting,” *diagnostics*, vol. 11, p. 1714, 2021.
- [6] M. Kivrak, U. Avci, H. Uzun, and C. Ardic, “The Impact of the SMOTE Method on Machine Learning and Ensemble Learning Performance Results in Addressing Class Imbalance in Data Used for Predicting Total Testosterone

- Deficiency in Type 2 Diabetes Patients and Type 2 [2]. Type 2 diabetes constitutes,” *Diagnostics*, vol. 14, no. 23, p. 2634, 2024.
- [7] Erlin, Y. N. Marlim, Junadhi, L. Suryati, and N. Agustina, “Deteksi Dini Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 11, no. 2, pp. 88–96, 2022.
- [8] W. Nugraha and A. Sasongko, “Hyperparameter Tuning pada Algoritma Klasifikasi dengan Grid Search Hyperparameter Tuning on Classification Algorithm with,” *Sist. J. Sist. Inf.*, vol. 11, no. 2, pp. 391–401, 2022.
- [9] E. Ramadanti, D. A. Dinathi, C. Sri, K. Aditya, and D. R. Chandranegara, “Diabetes Disease Detection Classification Using Light Gradient Boosting (LightGBM) With Hyperparameter Tuning,” vol. 8, no. 2, pp. 956–963, 2024.
- [10] U. Laila, K. Mahboob, A. W. Khan, F. Khan, and W. Taekeun, “An Ensemble Approach to Predict Early-Stage Diabetes Risk Using Machine Learning : An Empirical Study,” pp. 1–15, 2022.
- [11] H. Seto *et al.*, “Gradient boosting decision tree becomes more reliable than logistic regression in predicting probability for diabetes with big data,” *Sci. Rep.*, vol. 12, p. 15889, 2022, doi: <https://doi.org/10.1038/s41598-022-20149-z>.
- [12] M. R. Maulana *et al.*, “STUDI KOMPARATIF ALGORITMA RANDOM FOREST , GRADIENT BOOSTING , DAN LIGHTGBM PADA KLASIFIKASI RISIKO DIABETES MELITUS,” vol. 10, no. 2, pp. 2116–2122, 2026.
- [13] S. M. Ganie, P. K. D. Pramanik, M. B. Malik, S. Mallik, and H. Qin, “An ensemble learning approach for diabetes prediction using boosting techniques,” *Front. Genet.*, vol. 14, no. October, pp. 1–15, 2023, doi: <https://doi.org/10.3389/fgene.2023.1252159>.
- [14] F. Rahman, S. Hossain, J. Tiang, and A.-A. Nahid, “Diabetes Prediction Using Feature Selection Algorithms and Boosting-Based Machine Learning Classifiers,” *diagnostics*, vol. 15, no. 20, pp. 1–24, 2025, doi: <https://doi.org/10.3390/diagnostics15202622>.
- [15] S. K. Nisa, M. A. Barata, and P. E. Yuwita, “Optimization of Random Forest Algorithm with SMOTE Method to Improve the Accuracy of Early Diabetes Prediction,” *Sci. J. Informatics*, vol. 12, no. 3, pp. 387–396, 2025, doi: <https://doi.org/10.15294/sji.v12i3.22986>.