IMPLEMENTASI ALGORITMA LGBM MENGGUNAKAN OPTIMASI OPTUNA DALAM KLASIFIKASI PENYAKIT DIABETES

Setyobudi Utomo[1]*, Dwi Arman Prasetya[2], Muhammad Nasrudin[3]
[1,2,3]Program Studi Sains Data, Fakultas Ilmu Komputer,Universitas Pembangunan Nasional "Veteran" Jawa Timur, Surabaya, Indonesia

**Penulis korespondensi: 22083010112@student.upnjatim.ac.id*

ABSTRAK

Penyakit diabetes merupakan salah satu masalah kesehatan global dengan prevalensi yang terus meningkat termasuk di Indonesia. Deteksi dini sangat penting untuk mencegah komplikasi serius, sehingga dibutuhkan model klasifikasi yang akurat dan efisien. Perkembangan *machine learning* telah menyediakan pendekatan komputasional yang efektif dalam mendukung proses klasifikasi penyakit serta prediksi risiko. Algoritma *Light Gradient Boosting Machine* (LGBM) menjadi pilihan karena unggul dalam kecepatan pelatihan dan akurasi prediksi. Namun, performa LGBM sangat dipengaruhi oleh pemilihan *hyperparameter* yang optimal sehingga diperlukan proses optimasi untuk mendapatkan hasil terbaik. Penelitian ini bertujuan untuk membangun dan mengevaluasi model prediksi diabetes menggunakan LGBM yang dioptimalkan dengan Optuna. Pada tahap awal sebelum optimasi, model baseline menghasilkan akurasi sebesar 82.50%. Setelah dilakukan optimasi *hyperparameter* menggunakan *Optuna*, performa model meningkat signifikan dengan mencapai akurasi 96.50%. Temuan ini menunjukkan bahwa LGBM yang dioptimalkan dengan *Optuna* mampu menghasilkan model klasifikasi diabetes yang sangat akurat dan efisien. Model ini berpotensi digunakan sebagai alat pendukung dalam deteksi dini risiko diabetes untuk membantu proses pengambilan keputusan di layanan kesehatan.

Kata kunci: Diabetes, Klasifikasi, LGBM, Optuna

1 PENDAHULUAN

Di tengah perkembangan zaman dan modernisasi, diabetes muncul sebagai salah satu ancaman serius bagi kesehatan masyarakat. Diabetes adalah penyakit yang ditandai dengan terjadinya hiperglikemia dan gangguan metabolisme karbohidrat, lemak, dan akibat gangguan dalam sekresi insulin (Fatimah, 2015). Faktor utama penyebab penyakit ini adalah perubahan dalam pola makan dan cara hidup. Faktor tersebut akan mempengaruhi kondisi insulin, tekanan darah, dan kandungan kadar glukosa darah yang meningkat yaitu ≥ 200 mg/dL dan kadang-kadang disertai dengan penurunan berat badan yang tidak diketahui penyebabnya (Marta Dwi Sasmita dkk., 2021). Hal tersebut merupakan faktor utama yang mempengaruhi manusia terjangkit penyakit diabetes.

Secara global, Federasi Diabetes International (IDF) mencatat bahwa ada 537 juta orang di seluruh dunia yang mengalami diabetes pada tahun 2021. Dengan mencapai sekitar 19,5 juta orang yang memiliki diabetes pada tahun tersebut yang menempatkan Indonesia berada pada posisi ke enam (Casmuti dkk., 2025). Diabetes di Indonesia dianggap sebagai masalah kesehatan yang signifikan dan telah menjadi perhatian sejak awal tahun 1980-an. Negara Indonesia memiliki tingkat prevalensi sebesar 6,2% dengan lebih dari 10 juta orang yang hidup dengan diabetes (Ligita dkk., 2019). Angka ini menegaskan bahwa diabetes bukan sekadar persoalan individu, melainkan isu kesehatan global yang membutuhkan perhatian serius. Jika tidak ditangani dengan baik, diabetes dapat menyebabkan berbagai komplikasi serius seperti penyakit jantung, stroke, gagal

ginjal, gangguan penglihatan, hingga amputasi. Oleh karena itu, pendeteksian dini dan pemantauan kadar gula darah menjadi langkah yang sangat penting untuk mencegah komplikasi yang lebih lanjut.

Perkembangan zaman membawa kemajuan teknologi yang semakin pesat, termasuk dalam sektor kesehatan. Inovasi di bidang teknologi medis terus bermunculan dan memberikan perubahan signifikan pada proses diagnosis, perawatan, hingga pemantauan kondisi pasien. Salah satu teknologi yang berkembang adalah machine learning yang kini dimanfaatkan ke dalam praktik medis merupakan respons terhadap tantangan yang meningkat pesat di industri kesehatan (Furizal dkk., 2023). Dalam kasus diabetes, machine learning dapat digunakan untuk melakukan klasifikasi guna membantu proses identifikasi penyakit. Dengan upaya analisis dan klasifikasi tersebut, diharapkan dapat ditemukan solusi yang mampu menekan tingginya angka penderita diabetes (Mosa & Abdulazeez, 2024).

Penelitian sebelumnya yang berjudul "*Diabetes Disease Detection Classification Using Light Gradient Boosting (LightGBM) With Hyperparameter Tuning*" dalam penelitian ini menguji algoritma LGBM dengan berbagai macam optimasi *hyperparameter*. Hasil menunjukan akurasi menggunakan optimasi *GridSearchCV* mencapai 84% dan akurasi menggunakan Optimasi *RandomSearchCV* mencapai 82% (Ramadanti dkk., 2024). Penelitian lain berjudul "Klasifikasi Status Gizi pada Balita Menggunakan Metode *light Gradient Boosting Machine* (LightGBM) dengan Optimasi *Hyperparameter GridSearchCV*" juga menguji LGBM menggunakan optimasi lain. Dalam penelitian tersebut LGBM digabungkan dengan optimasi *GridSearchCV* mendapatkan akurasi sebesar 92% (Hegar dkk., 2025).

Dengan ini peneliti mengusulkan pendekatan lanjutan dengan menerapkan metode *Light Gradient Boosting Machine* (LGBM) yang dipadukan dengan optimasi *hyperparameter* menggunakan Optuna. LGBM merupakan algoritma gradient boosting yang dikenal efisien dalam mengolah data berukuran besar serta mampu membangun model prediksi dengan cepat dan akurat (Omotehinwa dkk., 2024). Untuk memaksimalkan kinerjanya, penelitian ini menggunakan Optuna yaitu *framework* optimasi *hyperparameter* berbasis *Bayesian optimization* yang mampu menemukan kombinasi parameter terbaik secara otomatis, lebih adaptif, dan lebih efisien dibandingkan metode pencarian konvensional. Melalui integrasi antara LGBM dan Optuna, model yang dihasilkan diharapkan dapat mempelajari pola kompleks pada data pasien secara lebih optimal dan memberikan prediksi risiko diabetes dengan tingkat akurasi yang lebih tinggi.

2 METODE

2.1 Data

Dataset yang digunakan dalam penelitian ini merupakan Pima Indian Diabetes Dataset, yaitu data yang berisi informasi mengenai kondisi fisiologis dan medis pasien yang berkaitan dengan risiko diabetes. Dataset ini mencakup sejumlah variabel yang menggambarkan karakteristik biologis, seperti tekanan darah, kadar glukosa, indeks massa tubuh, dan parameter klinis lainnya, serta status diabetes sebagai label target. Dataset tersebut digunakan untuk membangun dan menguji model klasifikasi guna memprediksi kemungkinan seseorang menderita diabetes berdasarkan faktor-faktor fisiologis dan medis. Deskripsi lebih rinci mengenai variabel data akan dijelaskan pada Tabel 1.

Tabel 1. Variabel Data

No	Nama Variabel	Deskripsi
1	Pregnancies	Jumlah kehamilan pasien
2	Glucose	Konsentrasi glukosa dalam plasma (mg/dL)
3	Blood Pressure	Tekanan darah diastolik (mm/Hg)
4	Skin Thickness	Ketebalan lipatan kulit triceps (mm)
5	Insulin	Kadar insulin serum 2 jam (mu U/ml)
6	BMI	Indeks massa tubuh (berat badan dalam kg dibagi kuadrat tinggi badan m^2)
7	Diabetes Pedigree Function	Fungsi silsilah diabetes (indikator risiko genetic diabetes)
8	Age	Usia pasien (dalam tahun)
9	Outcome	Status diabetes (0 = tidak diabetes, 1 = diabetes)

2.2 Light Gradient Boosting

Light Gradient Boosting Machine (LGBM) merupakan salah satu algoritma *ensemble learning* berbasis *Gradient Boosting Decision Tree* (GBDT) yang dirancang untuk memberikan performa tinggi dengan waktu komputasi yang cepat serta efisiensi memori yang lebih baik dibandingkan algoritma boosting lainnya (Deng dkk., 2025). LGBM mampu menghasilkan model dengan akurasi tinggi melalui mekanisme pembangunan pohon yang lebih efisien dengan memanfaatkan teknik *Gradient-based One-Side Sampling* (GOSS) dan *Exclusive Feature Bundling* (EFB), sehingga proses pelatihan dapat dipercepat tanpa mengurangi kualitas prediksi (Sun, 2025). Selain itu, LGBM memiliki fleksibilitas dalam menangani data berukuran besar, mendukung berbagai fungsi *loss*, serta menyediakan parameter regularisasi yang berperan dalam mengurangi risiko *overfitting*. Dari sisi arsitektur model, struktur pohon yang dibangun secara *leaf-wise* memungkinkan LGBM memfokuskan proses pemisahan pada bagian data yang paling berkontribusi terhadap penurunan nilai *loss* (Florek & Zagdański, 2023). Dukungan paralelisasi dan optimasi internal juga menjadikan LGBM memiliki skalabilitas yang baik, mampu mempertahankan performa yang stabil sekaligus mengurangi waktu pelatihan, sehingga algoritma ini banyak digunakan dalam pengembangan model klasifikasi dan regresi.

Secara umum, model *boosting* direpresentasikan sebagai berikut,

$$\square_{\square}(\square) = \square_{\square-1}(\square) + \square \cdot \square_{\square}(\square) \quad (1)$$

dimana:

$\square_{\square}(\square)$: model pada iterasi ke-t

η: learning rate

$\square_{\square}(\square)$: pohon regresi yang ditambahkan pada iterasi tersebut

Pada setiap iterasi *boosting*, LGBM menghitung *gradient* ($\square_{\square}$) dan *hessian* ($h_{\square}$) untuk setiap sampel berdasarkan fungsi *loss* yang digunakan.

Untuk kasus klasifikasi dengan *log-loss*:

$$□_□ = □_□ − □_□ \quad (2)$$

$$h_□ = □_□(1 − □_□) \quad (3)$$

dengan:

$$□_□ = □(\widehat{□}_□) = \frac{1}{1+□^{-\widehat{□}□}} \quad (4)$$

Gradient dan hessian inilah yang menentukan arah perbaikan model pada iterasi. Setelah menentukan nilai *output* dalam sebuah *leaf,* LGBM menggunakan w^* melalui persamaan,

$$□^* = -\frac{\sum□□□□□□□□_□}{\sum□□□□□□h_□+□} \quad (5)$$

dimana:
$□_□$: gradient
$h_□$:hessian
λ :regulasi L2

2.3 Optuna

Optuna merupakan *framework automated hyperparameter optimization* yang dirancang untuk memberikan fleksibilitas, efisiensi, serta kemudahan pengguna dalam menemukan *hyperparameter* terbaik (Raiaan dkk., 2024). Optuna memiliki karakteristik modularitas tinggi melalui pendekatan *define-by-run* yang memungkinkan pengguna untuk secara dinamis mengonstruksi ruang pencarian *hyperparameter* (Hanifi dkk., 2024). Selain itu, Optuna mengintegrasikan dua mekanisme inti sampling melalui *Tree-structured Parzen Estimator* (TPE) dan pruning adaptif yang membuat proses optimasi berjalan lebih efisien dibandingkan pendekatan tradisional seperti *grid search* maupun *random search* (Pravin dkk., 2022). Melalui TPE, Optuna membangun model probabilistik untuk memisahkan distribusi parameter yang menghasilkan performa baik dan buruk, sehingga proses pencarian dapat difokuskan pada wilayah parameter yang lebih menjanjikan. Di sisi lain, mekanisme *pruning* memungkinkan penghentian *trial* yang tidak memberikan indikasi performa baik pada tahap awal, sehingga mengurangi waktu komputasi secara signifikan. Kombinasi kedua teknik ini menjadikan Optuna tidak hanya cepat dan adaptif, tetapi juga mampu melakukan eksplorasi *hyperparameter* secara lebih terarah dan informatif bahkan pada ruang pencarian yang besar dan kompleks. Rumus umum optimasi *hyperparameter* yang dijalankan Optuna dapat dituliskan sebagai berikut,

$$\theta^* = \arg\min_{\theta \epsilon \Theta} □(□) \text{ atau } \theta^* = \arg\max_{\theta \epsilon \Theta} □(□) \quad (6)$$

dimana:
θ : kombinasi *hyperparameter*
ϵ : ruang pencarian
$□(□)$: nilai objektif dari model
Objektif ini dapat bersifat non-linear, tidak terdiferensiasi, atau bahkan bising dan Optuna tetap mampu menanganinya melalui metode sampling adaptif.

3 HASIL DAN PEMBAHASAN

3.1 Data dan Deskripsi Dataset

Penelitian ini menggunakan *Pima Indians Diabetes Dataset* yang dapat diakses melalui platform online kaggle. Data tersebut berisi informasi medis dari pasien perempuan keturunan Pima Indian, dengan beberapa variabel klinis seperti jumlah kehamilan, kadar glukosa, tekanan darah, ketebalan kulit, kadar insulin, BMI, nilai *Diabetes Pedigree Function*, dan usia. Seluruh atribut ini digunakan sebagai fitur input untuk membantu model mempelajari pola yang berhubungan dengan kondisi diabetes. Rincian lengkap mengenai setiap atribut dapat dilihat pada Tabel 2.

Tabel 2. Data Diabetes

Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	Outcome
2	138	63	35	0	33.6	0.127	47	1
0	84	82	31	125	38.2	0.223	23	0
0	145	0	0	0	44.2	0.630	31	1
0	135	68	42	250	42.3	0.365	24	1
1	139	62	41	480	40.7	0.536	21	0

Pada Tabel 2 menyajikan contoh beberapa atribut yang digunakan dalam penelitian ini untuk memprediksi dan klasifikasi penyakit diabetes dengan bantuan model *machine learning*. Setiap baris pada tabel menunjukkan satu data pasien, sementara kolom-kolomnya berisi berbagai fitur medis yang menjadi dasar analisis model.

3.2 Data Preprocessing

Pada tahap selanjutnya dilakukan proses *preprocessing* data untuk memastikan data siap digunakan oleh model. Tahap ini dimulai dengan menghapus beberapa kolom tertentu, yaitu *Pregnancies*, Insulin, *Diabetes Pedigree Function*, dan *Skin Thickness*. Langkah ini dilakukan untuk menyederhanakan dataset sehingga hanya fitur yang dianggap relevan saja yang dipertahankan meskipun sebenarnya beberapa kolom tersebut memiliki nilai informasi yang cukup penting dalam kasus diabetes. Selanjutnya dilakukan pengecekan *missing values* dan dari hasil pengecekan terdapat banyak nilai 0. Untuk memperbaiki kualitas data, nilai 0 tersebut kemudian diganti dengan nilai rata-rata (*mean*) dari masing-masing kolom. Proses ini dilakukan dengan menghitung mean setiap kolom terlebih dahulu, lalu mengganti nilai 0 menggunakan nilai mean tersebut.

Tabel 3. *Data Cleaning*

Glucose	*Blood Pressure*	*BMI*	*Age*	*Outcome*
138.0	62.0	33.6	37	1
84.0	82.0	38.2	23	0
145.0	69.1	44.2	31	1
135.0	68.0	42.3	24	1
139.0	62.0	40.7	21	0

Secara keseluruhan, *preprocessing* ini bertujuan membersihkan dataset dari nilai tidak valid agar model *machine learning* dapat belajar dengan lebih stabil, meskipun beberapa keputusan seperti penghapusan kolom dan metode imputasi masih bisa ditingkatkan untuk hasil yang lebih optimal.

3.3 Normalisasi Data

Untuk mengatasi ketimpangan skala antar fitur dilakukan proses normalisasi data.Hasil dari proses ini ditampilkan pada Tabel 4.

Tabel 4. Normalisasi Data

Glucose	Blood Pressure	BMI	Age	Outcome
0.606425	0.387755	0.246795	0.433333	1
0.258065	0.591837	0.320513	0.033333	0
0.651613	0.460668	0.416667	0.166667	1
0.587097	0.448980	0.386218	0.050000	1
0.612903	0.387755	0.360577	0.000000	0

Dalam tabel tersebut setiap fitur telah ditransformasi ke dalam nilai antara 0 dan 1 menggunakan metode *Min-Max Scaling.* Tujuan normalisasi ini adalah agar semua fitur memiliki kontribusi yang seimbang saat digunakan dalam pelatihan model, sehingga dapat meningkatkan akurasi dan stabilitas model.

3.4 *Data Splitting*

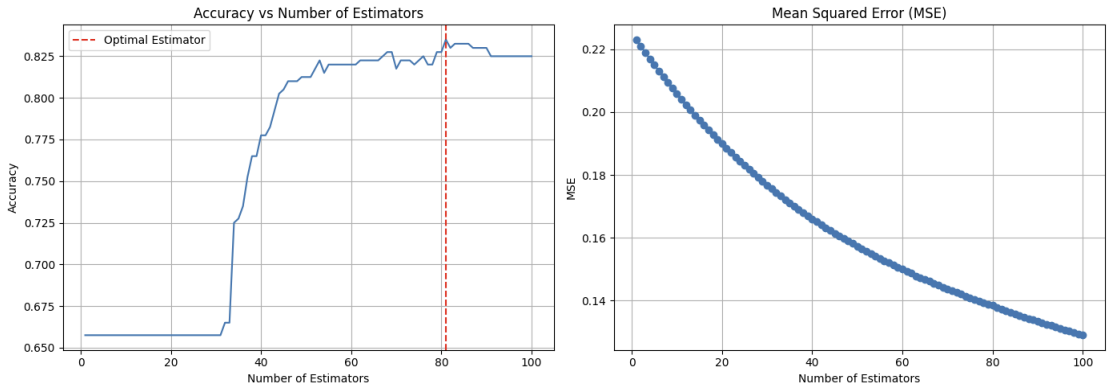
Dari hasil pembagian data terdapat 1.600 data yang digunakan untuk proses pelatihan (training) dan 400 data lainnya digunakan untuk pengujian (testing). Dapat dilihat dari Tabel 5.

Tabel 5. *Splitting Data*

Data Training	1.600
Data Testing	400

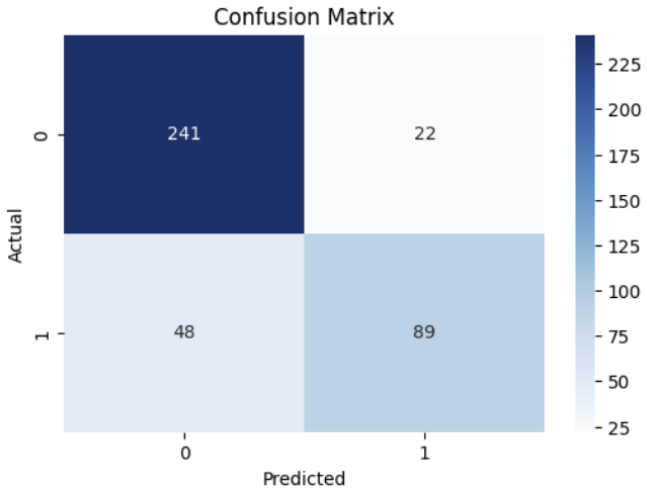
3.5 Pembuatan Model LGBM

Pada tahap ini Model LightGBM dibangun sebagai baseline karena algoritma ini efisien dan mampu menangani data numerik maupun kategorikal, tetapi konfigurasi awal yang digunakan sengaja dibuat konservatif sehingga performanya belum optimal. Inisialisasi dilakukan dengan *LGBMClassifier* menggunakan *n_estimators* = 100, *learning_rate* = 0.01, dan *objective* = 'binary', sebuah kombinasi yang cenderung menghasilkan model underfit karena learning rate kecil tidak diimbangi jumlah estimator yang memadai.



Gambar 1. Akurasi dan MSE Model LGBM sebelum Optimasi

Pada tahap selanjutnya, grafik akurasi menunjukkan bahwa peningkatan jumlah estimator awalnya menghasilkan kenaikan performa yang cukup tajam, namun setelah sekitar 40–50 estimator peningkatannya mulai melambat dan akhirnya stabil di kisaran 0.82–0.83. Titik optimal terlihat pada sekitar estimator ke-80, di mana akurasi mencapai nilai tertinggi sebelum kembali stagnan. Pada grafik MSE, error terus menurun seiring bertambahnya estimator, tetapi penurunannya semakin kecil sehingga manfaat tambahan setelah titik tertentu tidak lagi signifikan. Pola ini menunjukkan adanya *diminishing return* model memang terus belajar, tetapi peningkatan akurasi tidak lagi sebanding dengan penambahan kompleksitas.



Gambar 2. *Confusion Matrix* Model LGBM sebelum Optimasi

Hasil *confusion matrix* menunjukkan bahwa model mampu mengklasifikasikan kelas 0 dengan baik, dengan 241 prediksi benar dan hanya 22 kesalahan, namun performanya terhadap kelas 1 masih lebih lemah dengan 48 kesalahan dan hanya 89 prediksi benar. Hal ini tercermin pada recall kelas 1 yang hanya 0.65, yang mengindikasikan bahwa model masih sering gagal mendeteksi kasus positif, meskipun *precision* kelas 1 cukup baik (0.80) sehingga sebagian besar prediksi positif adalah benar. Dengan akurasi keseluruhan 82.5% dan AUC 0.916, model memiliki kemampuan diskriminatif yang kuat tetapi masih bias terhadap kelas mayoritas.

Tabel 5. Evaluasi Confusion Matrix Model LGBM sebelum Optimasi

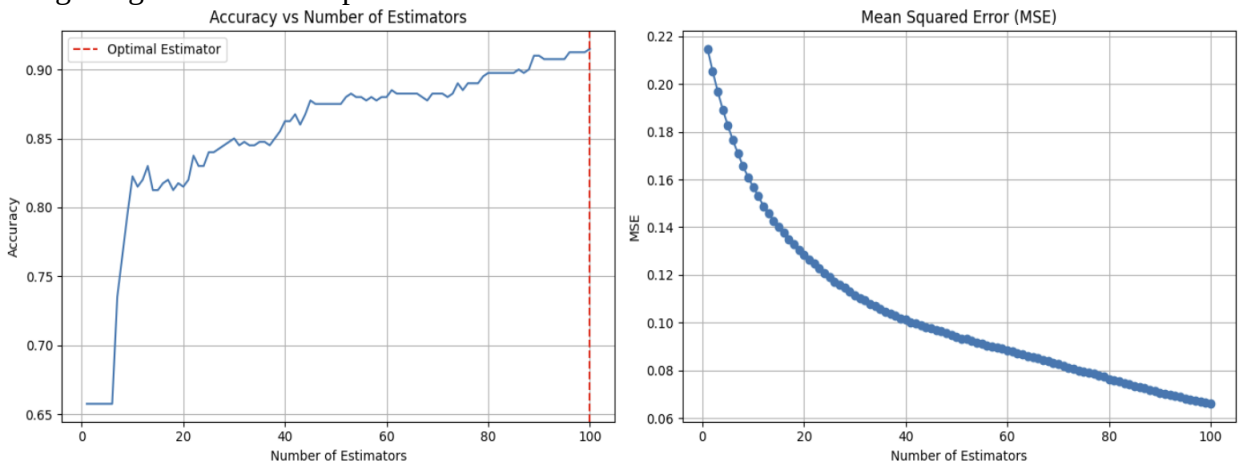
Class	Precission	Recall	F1-Score	Support
0	0.83	0.92	0.87	263
1	0.80	0.65	0.72	137
Accuracy			0.82	400
Macro Avg	0.82	0.78	0.80	400
Weighted Avg	0.82	0.82	0.82	400

Secara keseluruhan model baseline ini sudah cukup stabil namun, peningkatan pada *recall* kelas 1 misalnya melalui penyesuaian *threshold* atau penanganan ketidakseimbangan data perlu menjadi prioritas agar performa model lebih seimbang dan relevan untuk konteks prediksi yang sensitif terhadap kesalahan deteksi positif.

3.6 Pembuatan Model LGBM menggunakan Optimasi Optuna

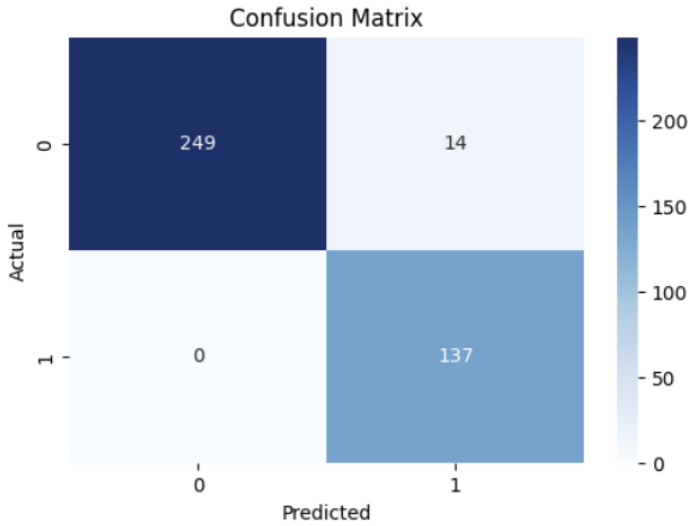
Setelah model *baseline* menunjukkan performa yang stabil namun masih memiliki kelemahan dalam mendeteksi kelas minoritas, proses optimasi hyperparameter dilakukan menggunakan Optuna untuk meningkatkan generalisasi model. Optuna dipilih karena mampu melakukan pencarian parameter secara adaptif dan efisien, dengan *AUC cross-validation* sebagai metrik utama agar evaluasi lebih tahan terhadap ketidakseimbangan data. Hasil optimasi menghasilkan *AUC* terbaik sebesar 0.9688, jauh lebih tinggi dibandingkan *baseline*, dengan kombinasi parameter

optimal seperti *num_leaves* = 31, *max_depth* = 5, *learning_rate* = 0.0995, *n_estimators* = 595. Parameter ini menunjukkan bahwa model yang optimal memerlukan struktur pohon yang relatif dangkal, jumlah estimator yang besar, serta penggunaan *subsampling* dan *colsampling* untuk mengurangi korelasi antar pohon.



Gambar 3. Akurasi dan MSE Model LGBM sesudah Optimasi

Grafik akurasi menunjukan peningkatan performa secara konsisten, dengan lonjakan signifikan pada *estimator* awal sebelum stabil di sekitar 0.91, sementara grafik *MSE* menampilkan penurunan *error* yang kuat pada tahap awal lalu melemah seiring bertambahnya *estimator* menandakan *diminishing return*.



Gambar 4. *Confusion Matrix* Model LGBM sesudah Optimasi

Evaluasi akhir model memperlihatkan performa yang sangat tinggi dengan akurasi keseluruhan 96.5% dan *AUC* 0.998 yang tetap perlu hati-hati karena berpotensi mencerminkan data yang mudah atau pola *overfitting*. Meskipun precision kelas positif berada di 0.91 dan *f1-score* masing-masing kelas mencapai 0.97 dan 0.95 menunjukkan keseimbangan performa yang baik.

Tabel 6. Evaluasi *Confusion Matrix* Model LGBM sesudah Optimasi

Class	Precission	Recall	F1-Score	Support
0	1.00	0.95	0.97	263
1	0.91	1.00	0.95	137
Accuracy			0.96	400
Macro Avg	0.95	0.97	0.96	400
Weighted Avg	0.97	0.96	0.97	400

LGBM yang dikombinasikan dengan Optuna terbukti mampu meningkatkan performa model secara signifikan, karena proses pencarian *hyperparameter* yang lebih terarah dan efisien membuat model mencapai konfigurasi yang menghasilkan generalisasi lebih baik dibandingkan *baseline*.

3.7 Perbandingan Model LGBM Sebelum dan Sesudah Optimasi

Perbandingan kinerja model LGBM sebelum dan sesudah optimasi dilakukan untuk melihat sejauh mana proses tuning mampu meningkatkan kemampuan model dalam melakukan klasifikasi. Evaluasi dilakukan menggunakan *confusion matrix* dan metrik-metrik utama seperti akurasi, presisi, *recall*, serta *AUC* pada kedua versi model. Dengan cara ini, perubahan performa dapat dilihat secara objektif dan terukur.

Tabel 7. Evaluasi *Confusion Matrix* Model LGBM sesudah Optimasi

Model	Teknik Optuna	Akurasi (%)	AUC (%)	Recall (%)	Precision (%)
LGBM	Tanpa Optuna	82.5	0.916	0.65	0.80
LGBM	Dengan Optuna	96.5	0.998	1.00	0.91

Model sebelum optimasi menunjukkan akurasi sebesar 82.50% dengan nilai AUC 0.916. Meskipun performanya sudah cukup baik, model masih mengalami jumlah *false negative* yang tinggi, terlihat dari *recall* kelas 1 yang hanya mencapai 0.65. Kondisi ini menunjukkan bahwa model masih lemah dalam mendeteksi kelas positif secara konsisten. Presisi pada kelas 1 berada di angka 0.80, menunjukkan keseimbangan prediksi yang masih moderat tetapi belum optimal. Setelah dilakukan optimasi, performa model meningkat secara signifikan. Akurasi naik menjadi 96.50% dan AUC mencapai 0.998, yang menunjukkan tingkat separabilitas model yang jauh lebih baik. *Recall* kelas 1 meningkat drastis menjadi 1.00, menandakan bahwa model tidak lagi melewatkan sampel positif. Selain itu, presisi kelas 1 juga naik menjadi 0.91, sehingga prediksi positif yang dihasilkan model menjadi lebih terpercaya. Penurunan jumlah false negative hingga nol memperlihatkan peningkatan yang sangat besar pada kemampuan deteksi kelas positif. Secara keseluruhan, hasil perbandingan ini menunjukkan bahwa proses optimasi memberikan dampak positif yang signifikan pada kinerja model LGBM. Peningkatan pada hampir seluruh metrik utama

menegaskan bahwa model menjadi lebih akurat, sensitif terhadap kelas positif, dan lebih stabil dalam melakukan prediksi.

4 KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, algoritma *Light Gradient Boosting Machine* (LGBM) yang dioptimalkan menggunakan Optuna terbukti mampu meningkatkan performa klasifikasi penyakit diabetes secara signifikan. Model yang dioptimalkan menghasilkan akurasi akhir sebesar 96,50%. *Confusion matrix* juga memperlihatkan penurunan jumlah kesalahan klasifikasi, khususnya pada kelas positif, di mana model hasil optimasi mampu mendeteksi seluruh kasus diabetes tanpa menghasilkan *false negative*. Secara praktis, model LGBM yang dioptimalkan dengan Optuna berpotensi digunakan sebagai sistem pendukung keputusan di layanan kesehatan untuk membantu proses deteksi dini dan skrining risiko diabetes secara lebih cepat dan objektif. Temuan penelitian ini dapat menjadi dasar bagi pengembangan sistem prediksi penyakit berbasis *machine learning* yang lebih adaptif dan akurat, serta membuka peluang penelitian lanjutan dengan memanfaatkan data klinis yang lebih luas dan beragam.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dalam pelaksanaan penelitian ini, mulai dari bantuan dukungan teknis, hingga kontribusi sebagai konsultan dan pendamping dalam proses pengumpulan data. Apresiasi khusus juga ditujukan kepada rekan-rekan yang turut membantu dalam analisis serta memberikan masukan berharga sehingga penelitian ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

Casmuti, Zainafree, I., Cahyati, W. H., Ningrum, D. N. A., Saefurrohim, M. Z., Hakam, A., Zaimatuddunia, I., & Prasetya, H. Y. (2025). Factors of Diabetic Retinopathy among Type 2 Diabetes Mellitus Patients in Central Java Province, Indonesia. *Unnes Journal of Public Health*, *14*(1), 57–68. https://doi.org/10.15294/ujph.v14i1.16126

Deng, L., Lu, K., & Hu, H. (2025). An Interpretable LightGBM Model for Predicting Coronary Heart Disease: Enhancing Clinical Decision-Making With Machine Learning. *PLOS ONE*, 1–26. https://doi.org/10.1371/journal.pone.0330377

Fatimah, R. N. (2015). Diabetes Melitus Tipe 2. *Jurnal Majority*, *4*(5), 93–101.

Florek, P., & Zagdański, A. (2023). Benchmarking State-of-the-art Gradient Boosting Algorithms for Classification. *Arxiv*, 1–0. http://arxiv.org/abs/2305.17094

Furizal, Ma'arif, A., & Rifaldi, D. (2023). Application of Machine Learning in Healthcare and Medicine: A Review. *Journal of Robotics and Control (JRC)*, *4*(5), 621–631. https://doi.org/10.18196/jrc.v4i5.19640

Hanifi, S., Cammarono, A., & Zare-Behtash, H. (2024). Advanced Hyperparameter Optimization of Deep Learning Models for Wind Power Prediction. *Renewable Energy*, *221*, 1–12. https://doi.org/10.1016/j.renene.2023.119700

Hegar, M. Y. L. B., Rahajoe, A. D., & Al Haromainy, M. M. (2025). Klasifikasi Status Gizi pada Balita Menggunakan Metode light Gradient Boosting Machine (LightGBM) dengan Optimasi Hyperparameter GridSearchCV. *Jurnal Ilmiah Ilmu Pendidikan*, *8*(12), 13890–13898. http://Jiip.stkipyapisdompu.ac.id

Ligita, T., Wicking, K., Francis, K., Harvey, N., & Nurjannah, I. (2019). How People Living with Diabetes in Indonesia Learn About Their Disease: A Grounded Theory Study. *PLoS ONE*, *14*(2), 1–19. https://doi.org/10.1371/journal.pone.0212019

Marta Dwi Sasmita, A., Author, C., Pendidikan Dokter, P., Kedokteran, F., & Lampung, U. (2021). Faktor-Faktor yang Mempengaruhi Tingkat Kepatuhan Berobat Pasien Diabetes Melitus. *Jurnal Medika Hutama*, *02*(04), 1105–1111. http://jurnalmedikahutama.com

Mosa, J. A., & Abdulazeez, A. M. (2024). A Review on Diabetes Classification Based on Machine Learning Algorithms. *Indonesian Journal of Computer Science (IJCS)*, *13*(2), 2445–24452473.

Omotehinwa, T. O., Oyewola, D. O., & Moung, E. G. (2024). Optimizing The Light Gradient-Boosting Machine Algorithm for an Efficient Early Detection of Coronary Heart Disease. *Journal of Informatics and Health*, *1*(2), 70–81. https://doi.org/10.1016/j.infoh.2024.06.001

Pravin, P. S., Tan, J. Z. M., Yap, K. S., & Wu, Z. (2022). Hyperparameter Optimization Strategies for Machine Learning-Based Stochastic Energy Efficient Scheduling in Cyber-Physical Production Systems. *Journal of Digital Chemical Engineering*, *4*, 1–12. https://doi.org/10.1016/j.dche.2022.100047

Raiaan, M. A. K., Sakib, S., Fahad, N. M., Mamun, A. Al, Rahman, M. A., Shatabda, S., & Mukta, M. S. H. (2024). A Systematic Review of Hyperparameter Optimization Techniques in Convolutional Neural Networks. *Decision Analytics Journal*, *11*, 1–32. https://doi.org/10.1016/j.dajour.2024.100470

Ramadanti, E., Dinathi, D. A., Aditya, C. S. K., & Chandranegara, D. R. (2024). Diabetes Disease Detection Classification Using Light Gradient Boosting (LightGBM) With Hyperparameter Tuning. *Jurnal Dan Penelitian Teknik Informatika*, *8*(2), 956–963. https://doi.org/10.33395/v8i2.13530

Sun, X. (2025). Application of an improved LightGBM hybrid integration model combining gradient harmonization and Jacobian regularization for breast cancer diagnosis. *Scientific Reports*, *15*(1), 1. https://doi.org/10.1038/s41598-025-86014-x