

Prediksi Diabetes Mellitus dengan Ensemble Gradient Boosting dan Advanced Feature Engineering

Daniswara Tegar Ramadhan*, Feri Agustina

Fakultas Ilmu Komputer, Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

Email: ^{1,*}111202113824@mhs.dinus.ac.id, ²feri.agustina@dsn.dinus.ac.id

Email Penulis Korespondensi: 111202113824@mhs.dinus.ac.id

Submitted: 14/07/2025; Accepted: 01/09/2025; Published: 04/09/2025

Abstrak—Diabetes mellitus merupakan penyakit metabolik yang menjadi permasalahan kesehatan global dengan tingkat prevalensi yang terus meningkat. Deteksi dini melalui sistem prediksi otomatis dapat membantu mengurangi komplikasi dan biaya pengobatan. Penelitian ini mengembangkan sistem prediksi diabetes mellitus menggunakan pendekatan ensemble gradient boosting yang dioptimalkan dengan rekayasa fitur canggih. Dataset penelitian menggabungkan 768 sampel Pima Indians dengan 5.000 sampel dari dataset prediksi diabetes, menghasilkan total 5.768 data yang kemudian diseimbangkan menggunakan teknik ADASYN. Proses rekayasa fitur mentransformasi 8 fitur original menjadi 29 fitur prediktif yang mencakup skor risiko diabetes, kategori BMI, kelompok usia, dan kategori glukosa. Tiga algoritma gradient boosting (XGBoost, LightGBM, CatBoost) beserta ensemble voting classifier dioptimalkan menggunakan framework Optuna dengan Tree-structured Parzen Estimator. Evaluasi menggunakan metrik akurasi, presisi, recall, F1-score, dan ROC-AUC melalui 5-fold cross validation. Hasil menunjukkan LightGBM memberikan performa terbaik dengan akurasi 97.14% dan ROC-AUC 0.9976, diikuti CatBoost (97.14%, 0.9973) dan XGBoost (96.45%, 0.9971). Analisis pentingnya fitur mengidentifikasi DiabetesPedigreeFunction, Pregnancies, dan SmokingHistory sebagai prediktor kunci. Model yang dikembangkan dapat diimplementasikan sebagai sistem skrining diabetes di fasilitas kesehatan primer.

Kata Kunci: Prediksi Diabetes; Gradient Boosting; Rekayasa Fitur; Optimasi Hyperparameter; Pembelajaran Mesin

Abstract—Diabetes mellitus represents a metabolic disease that constitutes a global health challenge with continuously increasing prevalence rates. Early detection through automated prediction systems can help reduce complications and treatment costs. This study develops a diabetes mellitus prediction system using an ensemble gradient boosting approach optimized with advanced feature engineering. The research dataset combines 768 Pima Indians samples with 5,000 samples from diabetes prediction dataset, resulting in 5,768 total data points subsequently balanced using ADASYN technique. Feature engineering process transforms 8 original features into 25 predictive features encompassing diabetes risk scores, BMI categories, age groups, and glucose categories. Three gradient boosting algorithms (XGBoost, LightGBM, CatBoost) along with ensemble voting classifier were optimized using Optuna framework with Tree-structured Parzen Estimator. Evaluation employed accuracy, precision, recall, F1-score, and ROC-AUC metrics through 5-fold cross validation. Results demonstrate LightGBM achieving optimal performance with 97.14% accuracy and 0.9976 ROC-AUC, followed by CatBoost (97.14%, 0.9973) and XGBoost (96.45%, 0.9971). Feature importance analysis identified DiabetesPedigreeFunction, Pregnancies, and SmokingHistory as key predictors. The developed model can be implemented as a diabetes screening system in primary healthcare facilities

Keywords: Diabetes Prediction; Gradient Boosting; Feature Engineering; Hyperparameter Optimization; Machine Learning

1. PENDAHULUAN

Diabetes mellitus telah berkembang menjadi epidemi global yang mengancam sistem kesehatan dunia dengan tingkat prevalensi yang terus meningkat secara mengkhawatirkan dalam beberapa dekade terakhir. World Health Organization melaporkan jumlah penderita diabetes mellitus mencapai 830 juta orang pada tahun 2024 dengan proyeksi peningkatan yang dramatis hingga 1.3 miliar orang pada tahun 2050 [1]. Peningkatan eksponensial ini mengindikasikan terjadinya transisi epidemiologi global yang fundamental, dimana beban penyakit bergeser dari penyakit menular menuju penyakit tidak menular, dengan diabetes mellitus sebagai salah satu penyebab utama morbiditas dan mortalitas di seluruh dunia.

Indonesia menempati posisi yang sangat kritis sebagai negara dengan 19.47 juta penderita diabetes mellitus, menempatkannya pada peringkat kelima dunia setelah China, India, Amerika Serikat, dan Pakistan [2]. Prevalensi yang mencapai 6.9% dari total populasi Indonesia menunjukkan urgensi tinggi untuk mengembangkan sistem deteksi dini yang efektif dan dapat diakses secara luas. Kondisi ini menimbulkan tantangan besar dalam sistem kesehatan nasional, mengingat diabetes mellitus dapat menyebabkan komplikasi serius dan mengancam jiwa seperti penyakit kardiovaskular, neuropati diabetik, nefropati diabetik, retinopati diabetik, dan ulkus kaki diabetik yang dapat mengancam kualitas hidup bahkan nyawa penderita.

Dampak ekonomi yang ditimbulkan oleh diabetes mellitus sangat signifikan dan terus meningkat seiring dengan prevalensi penyakit ini. Estimasi biaya kesehatan global mencapai 966 miliar dolar Amerika pada tahun 2024 dan diproyeksikan meningkat menjadi 1.054 miliar dolar Amerika pada tahun 2030 [3]. Di Indonesia, biaya perawatan diabetes mellitus per pasien mencapai 2.061 dolar Amerika per tahun, menciptakan beban finansial yang sangat berat bagi sistem kesehatan dan masyarakat secara keseluruhan. Kondisi ini diperparah oleh fakta yang sangat mengkhawatirkan bahwa sekitar 50% penderita diabetes mellitus tidak terdiagnosis dan tidak mendapat perawatan yang memadai, sehingga berisiko tinggi mengalami komplikasi yang sebenarnya dapat dicegah melalui deteksi dini dan pengelolaan yang tepat waktu.

Deteksi dini diabetes mellitus menjadi kunci utama dalam pencegahan komplikasi dan pengelolaan penyakit yang efektif serta *cost-effective*. Metode skrining tradisional seperti tes glukosa darah puasa, tes toleransi glukosa oral, dan pemeriksaan HbA1c, meskipun akurat dan merupakan standar emas untuk diagnosis diabetes, memerlukan infrastruktur laboratorium yang memadai, biaya yang relatif tinggi, prosedur yang kompleks, dan waktu yang lama untuk mendapatkan hasil [4]. Keterbatasan ini mengakibatkan keterlambatan diagnosis yang signifikan, terutama di daerah pedesaan dan terpencil dengan fasilitas kesehatan yang terbatas. Oleh karena itu, pengembangan sistem prediksi diabetes yang akurat, hemat biaya, non-invasif, dan mudah diimplementasikan menjadi kebutuhan yang sangat mendesak untuk mendukung program deteksi dini dan skrining massal di berbagai tingkat pelayanan kesehatan.

Kecerdasan buatan dan *machine learning* telah menunjukkan potensi yang sangat revolusioner dalam bidang kesehatan, khususnya untuk prediksi, diagnosis, dan prognosis berbagai penyakit. Dalam konteks diabetes mellitus, algoritma *machine learning* memiliki kemampuan luar biasa untuk menganalisis pola kompleks dalam data biomedis, mengidentifikasi hubungan halus dan *non-linear* antar variabel, serta memberikan wawasan prediktif yang tidak mudah dideteksi melalui metode statistik konvensional [5]. Kemampuan pembelajaran mesin untuk memproses dataset berskala besar, menangani hubungan *non-linear* yang kompleks, dan beradaptasi dengan pola data baru menjadikannya sangat ideal untuk mengembangkan model prediksi yang *robust* dan dapat mendukung pengambilan keputusan klinis yang akurat.

Penelitian terkini menunjukkan hasil yang sangat menjanjikan dalam prediksi diabetes mellitus menggunakan berbagai algoritma *machine learning* dengan pendekatan yang beragam. Andhika et al. [6] menggunakan XGBoost dengan *preprocessing* yang komprehensif termasuk SMOTE dan Min-Max Scaler, berhasil mencapai akurasi 98% pada dataset yang terdiri dari 1.536 sampel dengan 9 fitur. Namun, penelitian tersebut menggunakan *feature engineering* yang relatif terbatas dan belum mengoptimalkan *hyperparameter* secara sistematis menggunakan algoritma optimasi canggih. Mubarak et al. [7] menerapkan *hyperparameter tuning* pada XGBoost menggunakan Tree Parzen Estimator dan berhasil mencapai AUC 0.944, menunjukkan efektivitas teknik optimasi dalam meningkatkan performa model secara signifikan.

Ahmad et al. [8] melakukan tinjauan komprehensif terhadap pendekatan *machine learning* untuk prediksi diabetes dan mengidentifikasi XGBoost sebagai salah satu algoritma dengan kinerja terbaik, namun menekankan perlunya *feature engineering* yang canggih dan dataset yang lebih besar untuk mencapai generalisasi yang optimal. Sneha dan Gangil [9] membandingkan berbagai algoritma *machine learning* termasuk SVM, Random Forest, dan Naive Bayes, dengan Naive Bayes memberikan hasil terbaik dengan akurasi 82.3%, namun masih jauh dari ambang batas yang diperlukan untuk aplikasi klinis dan tidak melakukan *feature engineering* yang canggih. Kumar et al. [10] menerapkan *ensemble method* yang menggabungkan beberapa algoritma dan berhasil mencapai akurasi 94.2%, namun menggunakan dataset yang relatif kecil tanpa validasi eksternal dan belum mengeksplorasi teknik *ensemble* canggih seperti *stacking* dan *blending*.

Analisis mendalam terhadap literatur yang ada mengidentifikasi beberapa gap penelitian yang kritis dan perlu diatasi untuk memajukan bidang prediksi diabetes mellitus: keterbatasan ukuran dataset yang dapat mempengaruhi kemampuan generalisasi model dan ketahanan terhadap variasi populasi, kurangnya *feature engineering* yang sistematis dan komprehensif yang memanfaatkan domain knowledge medis untuk menciptakan prediktor yang bermakna secara klinis, optimasi *hyperparameter* yang belum menggunakan algoritma optimasi canggih yang dapat meningkatkan performa model secara signifikan, evaluasi yang terbatas tanpa menggunakan *multiple metrics* yang relevan secara klinis untuk validasi yang menyeluruh, dan kurangnya analisis terhadap *feature importance* dan interpretabilitas model untuk penerimaan dan kepercayaan klinis.

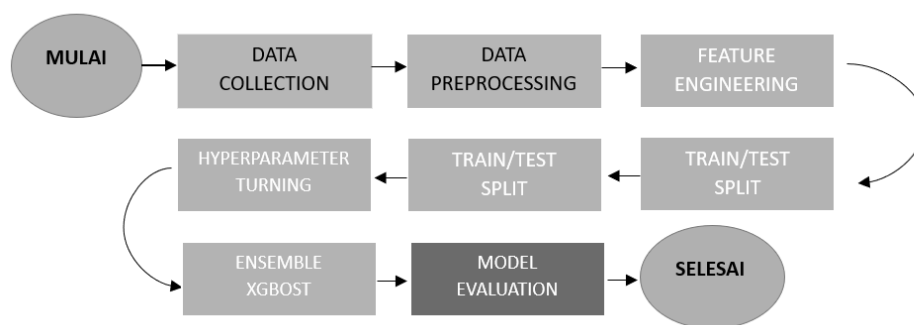
Penelitian ini mengembangkan sistem prediksi diabetes mellitus yang inovatif menggunakan *ensemble gradient boosting* yang dioptimalkan melalui *advanced feature engineering* dan *hyperparameter tuning* yang sistematis. Kontribusi utama penelitian ini meliputi: strategi multi-dataset yang cerdas dengan menggabungkan Pima Indians diabetes database dengan dataset prediksi diabetes yang lebih besar, implementasi *advanced feature engineering* yang mentransformasi 8 fitur original menjadi 25 fitur prediktif yang bermakna secara klinis, optimasi *hyperparameter* yang sistematis menggunakan Optuna framework dengan Tree-structured Parzen Estimator, evaluasi komprehensif terhadap tiga algoritma *gradient boosting* terdepan (XGBoost, LightGBM, dan CatBoost) beserta *ensemble voting classifier*, dan analisis *feature importance* yang mendalam untuk mendukung interpretabilitas klinis dan pengambilan keputusan medis.

2. METODOLOGI PENELITIAN

Metodologi penelitian ini menggunakan pendekatan eksperimental dengan tahapan yang sistematis untuk mengembangkan model prediksi yang akurat dan dapat diandalkan [11]. Setiap tahapan dirancang berdasarkan best practices dalam machine learning untuk healthcare applications [12].

2.1 Tahap Penelitian

Tahapan penelitian adalah langkah-langkah sistematis yang dilakukan dalam proses penelitian, mulai dari perencanaan hingga penyimpulan hasil. Pada bagian ini sudah dilampirkan gambaran flowchart maupun penjelasan pada flowchart. Dengan Gambar 1. Tahapan penelitian yang berisi bagaimana penyelesaian penelitian ini dilakukan.



Gambar 1. Tahapan penelitian

Gambar 1 menunjukkan tahapan penelitian yang dilalui dalam pengembangan model prediksi diabetes ini. Berikut adalah penjelasan singkat untuk setiap tahapan yang ada dalam diagram alir tersebut :

- Data Collection: Pada tahap ini, data dikumpulkan dari dua sumber dataset yang berbeda, yaitu Pima Indian Diabetes Dataset dan Diabetes Prediction Dataset, yang mencakup informasi mengenai kesehatan, gaya hidup, dan demografis.
- Data Preprocessing: Setelah data terkumpul, tahap preprocessing dilakukan untuk membersihkan data, mengatasi nilai yang hilang, dan menangani ketidakseimbangan kelas. Ini termasuk juga imputasi nilai yang hilang dan penggabungan dataset.
- Feature Engineering: Fitur baru diciptakan untuk meningkatkan daya prediksi model, seperti HOMA_IR, MetS_Score, Diabetes_Risk_Score, dan lainnya. Tahap ini bertujuan untuk menangkap interaksi antar variabel dan menciptakan fitur-fitur komposit yang relevan.
- Train/Test Split: Dataset dibagi menjadi dua bagian, yaitu data latih (training) untuk melatih model dan data uji (testing) untuk mengukur performa model. Pembagian ini penting untuk memastikan validitas dan generalisasi model.
- Hyperparameter Tuning: Proses ini mengoptimalkan hyperparameter model untuk memaksimalkan kinerja model. Optuna digunakan untuk menemukan kombinasi hyperparameter terbaik dengan menggunakan Tree-structured Parzen Estimator (TPE).
- Ensemble XGBoost: Model prediksi dikembangkan menggunakan pendekatan ensemble yang menggabungkan tiga algoritma gradient boosting terdepan: XGBoost, LightGBM, dan CatBoost. Pendekatan ini menggabungkan kekuatan masing-masing model untuk menghasilkan prediksi yang lebih akurat dan stabil.
- Model Evaluation: Pada tahap ini, evaluasi kinerja model dilakukan dengan menggunakan metrik seperti Akurasi, Presisi, Recall, F1-Score, dan ROC-AUC. 5-fold cross-validation digunakan untuk memastikan keandalan model dan meminimalkan overfitting.
- Selesai: Setelah model dievaluasi dan diuji, hasil akhir dari penelitian ini adalah model prediksi yang dapat digunakan untuk mendeteksi risiko diabetes secara akurat dan efisien.

2.2 Preprocessing dan Rekayasa Fitur

Tahap *preprocessing* dan rekayasa fitur merupakan langkah krusial dalam penelitian ini untuk meningkatkan kualitas data dan mengekstrak informasi prediktif yang lebih kuat. Proses ini mencakup penanganan nilai tidak valid, imputasi data yang hilang, dan penciptaan fitur-fitur baru berdasarkan pengetahuan domain.

2.2.1 Penanganan dan Imputasi Nilai Nol

Analisis awal data menunjukkan adanya nilai nol pada beberapa fitur fisiologis yang secara klinis tidak mungkin terjadi, seperti Glucose, BloodPressure, SkinThickness, Insulin, dan BMI. Nilai nol ini dianggap sebagai data yang hilang atau tidak tercatat. Oleh karena itu, nilai-nilai ini pertama-tama diubah menjadi NaN (*Not a Number*).

Selanjutnya, dilakukan imputasi atau pengisian nilai NaN tersebut. Untuk fitur Pregnancies pada dataset Prediction, imputasi dilakukan menggunakan estimasi berbasis distribusi Poisson yang mempertimbangkan usia dan jenis kelamin untuk menghasilkan nilai yang realistis. Untuk fitur HbA1c yang hilang pada dataset Pima, imputasi dilakukan menggunakan hubungan linear dengan Glucose berdasarkan formula klinis yang telah divalidasi, seperti yang ditunjukkan pada Persamaan 1.

$$HbA1c = \frac{(Glucose+46.7)}{28.7} \quad (1)$$

2.2.2 Rekayasa Fitur (*Feature Engineering*)

Untuk meningkatkan daya prediktif model, dilakukan rekayasa fitur dengan menciptakan 14 fitur baru dari 8 fitur asli. Fitur-fitur baru ini dirancang untuk menangkap hubungan non-linear dan interaksi kompleks antar variabel yang relevan secara medis. Beberapa fitur komposit utama yang dibuat adalah sebagai berikut:

- a. HOMA-IR (*Homeostatic Model Assessment for Insulin Resistance*): Fitur ini adalah indikator penting untuk mengukur resistensi insulin, yang merupakan faktor kunci dalam patofisiologi diabetes. HOMA-IR dihitung menggunakan Persamaan 2

$$\text{HOMA-IR} = \frac{(\text{Glucose [mg/dL]} \times \text{Insulin [\mu U/mL]})}{405} \quad (2)$$

- b. Diabetes Risk Score: Sebuah skor risiko komposit diciptakan untuk mengkuantifikasi risiko diabetes seseorang berdasarkan gabungan beberapa indikator kesehatan utama. Skor ini dihitung sebagai jumlah terbobot dari lima fitur yang telah dinormalisasi, seperti yang ditunjukkan pada Persamaan 3.

$$\text{Risk Score} = (0.15 \times N(\text{Age})) + (0.20 \times N(\text{BMI})) + (0.40 \times N(\text{Glucose})) + (0.15 \times N(\text{HbA1c})) + (0.10 \times N(\text{BP})) \quad (3)$$

- c. Di mana $N(x)$ adalah nilai fitur x yang telah dinormalisasi menggunakan metode Normalisasi Min-Max, sesuai dengan Persamaan 4.

$$N(x) = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (4)$$

Selain fitur komposit ini, fitur-fitur lain seperti kategori BMI, kelompok usia, kategori glukosa, dan fitur interaksi (misalnya, *Age_BMI_Interaction*) juga dibuat untuk memperkaya dataset sebelum dimasukkan ke dalam model.

2.3 Penanganan Data Tidak Seimbang

Berdasarkan analisis distribusi kelas yang disajikan pada Tabel 4, teridentifikasi adanya masalah ketidakseimbangan kelas (*class imbalance*) yang signifikan pada dataset gabungan, dengan rasio kelas non-diabetes terhadap kelas diabetes mencapai 10.49:1. Kondisi ini berisiko tinggi menghasilkan model prediksi yang bias, di mana model cenderung memiliki performa yang baik pada kelas mayoritas (non-diabetes) namun sangat buruk dalam mengidentifikasi kelas minoritas (diabetes). Mengingat tujuan utama penelitian ini adalah untuk deteksi dini, maka kemampuan model untuk mengenali kasus diabetes secara akurat menjadi prioritas utama.

Untuk mengatasi tantangan ini, diterapkan teknik *oversampling* canggih untuk menyeimbangkan distribusi kelas pada data latih. Guna memastikan metode yang digunakan adalah yang paling optimal, penelitian ini membandingkan beberapa algoritma, termasuk SMOTE (*Synthetic Minority Oversampling Technique*), ADASYN (*Adaptive Synthetic Sampling*), SMOTE+ENN, dan SMOTE+Tomek. Dari hasil perbandingan, metode ADASYN dipilih sebagai pendekatan final. ADASYN (*Adaptive Synthetic Sampling*) dipilih karena kemampuannya menghasilkan sampel sintesis yang fokus pada region yang sulit dipelajari [13]. Teknik ini terbukti lebih efektif dibandingkan SMOTE dalam menangani *class imbalance* pada medical datasets [14]. Advanced feature engineering mengikuti prinsip *domain knowledge integration* yang telah terbukti meningkatkan performa model prediksi medis [15].

Penerapan prosedur ini dilakukan dengan sangat hati-hati untuk menjaga integritas dan objektivitas evaluasi model. Proses *oversampling* menggunakan ADASYN hanya diaplikasikan pada data latih setelah dataset dibagi menjadi data latih dan data uji (*train-test split*). Prosedur ini krusial untuk mencegah terjadinya kebocoran data (*data leakage*), di mana informasi dari data uji secara tidak sengaja memengaruhi proses pelatihan model. Dengan demikian, data uji tetap sepenuhnya "tak terlihat" (*unseen*) oleh model hingga tahap evaluasi akhir, memastikan bahwa hasil kinerja yang dilaporkan merupakan cerminan sejati dari kemampuan generalisasi model.

2.4 Pengembangan Model Ensemble

Model prediksi dalam penelitian ini dikembangkan menggunakan pendekatan ensemble yang menggabungkan tiga algoritma *gradient boosting* terdepan: XGBoost, LightGBM, dan CatBoost. Setiap algoritma memiliki keunggulan spesifik dalam menangani berbagai aspek data medis. LightGBM dipilih karena efisiensi komputasi dan kemampuan handling *categorical features* yang superior [16]. XGBoost memiliki *regularization* yang kuat untuk mencegah *overfitting* pada medical data [17]. CatBoost secara otomatis menangani *categorical features* tanpa *preprocessing* tambahan [18].

Untuk lebih meningkatkan performa dan stabilitas prediksi, penelitian ini tidak hanya mengandalkan satu model tunggal, melainkan mengadopsi pendekatan ensemble yang komprehensif. Metode ensemble bekerja dengan menggabungkan kekuatan dari beberapa model yang berbeda untuk menghasilkan keputusan akhir yang lebih akurat dan andal daripada masing-masing model secara individual. Arsitektur ensemble yang digunakan adalah *Voting Classifier* dari pustaka *scikit-learn* yang menggabungkan prediksi dari ketiga algoritma *gradient boosting* yang telah dioptimalkan secara individual.

Mekanisme pengambilan keputusan diatur menggunakan metode *soft voting*, dimana prediksi akhir tidak hanya didasarkan pada suara mayoritas, melainkan pada rata-rata probabilitas prediksi dari setiap model. *Soft voting approach* terbukti memberikan hasil yang lebih stable dan robust dibandingkan *hard voting* karena mempertimbangkan tingkat "keyakinan" dari masing-masing model dalam membuat prediksi [19]. Pendekatan ini

memungkinkan ensemble untuk memanfaatkan kekuatan prediktif dari setiap algoritma sambil mengurangi kelemahan individual masing-masing model.

Setiap model dalam ensemble dilatih secara independen pada dataset yang sama, namun dengan konfigurasi hyperparameter yang dioptimalkan secara terpisah untuk memaksimalkan performa individual. Keputusan final ensemble diperoleh melalui weighted averaging dari probabilitas prediksi setiap model, dimana bobot ditentukan berdasarkan performa *cross-validation* masing-masing model. Strategi ini memastikan bahwa model dengan performa superior memberikan kontribusi yang lebih besar dalam keputusan akhir, sementara tetap mempertahankan stabilitas dan robustness yang menjadi karakteristik utama pendekatan ensemble.

2.5 Optimasi Hyperparameter

Untuk memastikan setiap model dalam arsitektur ensemble dapat mencapai performa puncaknya, dilakukan proses optimasi *hyperparameter* yang sistematis dan komprehensif. Tahap ini bertujuan untuk menemukan kombinasi nilai *hyperparameter* yang paling optimal untuk dataset yang digunakan, sehingga dapat memaksimalkan daya prediksi model secara keseluruhan.

Proses optimasi dilakukan secara otomatis menggunakan Optuna, sebuah *framework* optimasi generasi berikutnya yang dikenal efisien dan canggih [20]. *Optuna framework* dengan *Tree-structured Parzen Estimator* (TPE) telah terbukti lebih efisien dibandingkan grid search atau random search dalam optimasi *hyperparameter* [21]. Secara spesifik, penelitian ini menggunakan algoritma sampler *Tree-structured Parzen Estimator* (TPE) yang merupakan algoritma pencarian Bayesian yang lebih cerdas daripada metode pencarian acak (*random search*). TPE menggunakan algoritma *Bayesian optimization* yang memanfaatkan informasi dari trial sebelumnya untuk menentukan parameter *space* yang paling *promising* untuk dieksplorasi selanjutnya.

Proses optimasi dilakukan secara terpisah untuk setiap algoritma (XGBoost, LightGBM, dan CatBoost) dengan konfigurasi yang disesuaikan untuk karakteristik masing-masing model. Setiap proses optimasi dikonfigurasi untuk memaksimalkan metrik ROC-AUC melalui mekanisme 5-fold *cross-validation* yang memastikan stabilitas dan keandalan hasil. Jumlah trial yang digunakan dalam optimasi ditetapkan sebanyak 50 iterasi untuk setiap model, memberikan keseimbangan optimal antara kualitas hasil dan efisiensi komputasi.

Beberapa *hyperparameter* utama yang dioptimalkan meliputi jumlah trees (*n_estimators*) dengan rentang 100-500, kedalaman maksimum pohon (*max_depth*) dengan rentang 3-10, laju pembelajaran (*learning_rate*) dengan rentang 0.01-0.3, parameter subsampling (*subsample* dan *colsample_bytree*) dengan rentang 0.6-1.0, serta parameter regularisasi untuk mencegah *overfitting* (*reg_alpha* dengan rentang 0-2 dan *reg_lambda* dengan rentang 1-3). Parameter *space* yang luas ini memungkinkan algoritma optimasi untuk mengeksplorasi berbagai konfigurasi dan menemukan kombinasi *hyperparameter* yang paling sesuai dengan karakteristik dataset diabetes.

Hasil dari proses optimasi ini adalah satu set *hyperparameter* terbaik untuk masing-masing algoritma yang kemudian digunakan untuk melatih model final. Konfigurasi optimal yang diperoleh tidak hanya meningkatkan akurasi prediksi, tetapi juga memastikan bahwa setiap model dalam ensemble berkontribusi secara maksimal dalam keputusan akhir.

2.6 Metrik Evaluasi Model

Untuk memastikan penilaian kinerja model dilakukan secara komprehensif, objektif, dan relevan dengan aplikasi klinis, penelitian ini menggunakan serangkaian metrik evaluasi yang beragam. *Multiple metrics evaluation* diperlukan untuk memberikan gambaran komprehensif tentang performa model, terutama pada medical applications dimana kesalahan prediksi dapat berdampak signifikan pada keputusan klinis [22]. Pendekatan multi-metrik ini penting untuk memberikan gambaran menyeluruh tentang kekuatan dan kelemahan model, terutama karena metrik tunggal seperti akurasi dapat memberikan gambaran yang kurang lengkap pada dataset dengan kelas tidak seimbang.

Evaluasi final model dilakukan pada data uji (*test set*) yang sebelumnya telah dipisahkan dan tidak pernah digunakan selama proses pelatihan atau optimasi untuk memastikan objektivitas dan validitas hasil. Metrik-metrik utama yang digunakan dalam evaluasi ini adalah sebagai berikut:

- Akurasi (*Accuracy*) mengukur proporsi total prediksi yang benar (baik *True Positive* maupun *True Negative*) dari keseluruhan data, memberikan gambaran umum tentang ketepatan model dalam membuat prediksi. Presisi (*Precision*) mengukur tingkat ketepatan dari prediksi positif, yaitu dari semua pasien yang diprediksi menderita diabetes, berapa persen yang sebenarnya benar-benar menderita diabetes. Metrik ini sangat penting untuk meminimalkan false positive yang dapat menyebabkan kecemasan berlebihan pada pasien.
- Recall (Sensitivitas) mengukur kemampuan model untuk mengidentifikasi semua kasus positif yang relevan, yaitu dari semua pasien yang sebenarnya menderita diabetes, berapa persen yang berhasil terdeteksi oleh model. Ini merupakan metrik yang sangat krusial untuk sistem deteksi dini, karena tujuan utamanya adalah meminimalkan false negative yang dapat mengakibatkan keterlambatan diagnosis dan penanganan.

F1-Score merupakan rata-rata harmonik dari Presisi dan *Recall* yang memberikan satu angka yang menyeimbangkan antara kedua metrik tersebut. Metrik ini sangat berguna ketika ada kebutuhan untuk menyeimbangkan antara *false positive* dan *false negative*, terutama dalam aplikasi medis dimana kedua jenis kesalahan memiliki konsekuensi yang serius.

ROC-AUC (*Area Under the Receiver Operating Characteristic Curve*) mengukur kemampuan model secara keseluruhan untuk membedakan antara kelas positif (diabetes) dan kelas negatif (non-diabetes) di semua ambang batas

klasifikasi. ROC-AUC merupakan metrik yang robust untuk evaluasi model pada imbalanced medical datasets [23]. Nilai AUC yang mendekati 1.0 menunjukkan kemampuan diskriminatif yang sangat baik, sementara nilai 0.5 menunjukkan performa yang tidak lebih baik dari prediksi acak.

Selain metrik evaluasi utama, untuk memastikan keandalan dan stabilitas hasil selama tahap optimasi hyperparameter, digunakan metode validasi silang 5-fold stratified cross-validation. 5-fold stratified cross-validation memastikan validitas hasil dan mencegah overfitting dengan mempertahankan proporsi kelas pada setiap fold [24]. Pendekatan ini memberikan estimasi yang lebih robust tentang performa model yang diharapkan pada data baru yang belum pernah dilihat sebelumnya.

Untuk perbandingan statistik yang lebih mendalam antar model, dapat digunakan uji signifikansi seperti McNemar's test untuk membandingkan performa model dengan baseline, meskipun dalam penelitian ini fokus utama adalah pada perbandingan praktis antar algoritma berdasarkan metrik evaluasi yang telah disebutkan.

3. HASIL DAN PEMBAHASAN

Pada bab ini, disajikan hasil dari implementasi metodologi yang telah dijelaskan sebelumnya. Pembahasan mencakup hasil dari tahap preprocessing data, optimasi hyperparameter, evaluasi kinerja model prediksi, serta analisis mendalam terhadap fitur-fitur yang paling berpengaruh.

3.1 Dataset

Penelitian ini menggunakan dua dataset diabetes yang berbeda untuk memastikan *robustness* dan *generalizability* model. Dataset pertama adalah Pima Indian Diabetes Database yang merupakan benchmark dataset dalam penelitian prediksi diabetes [25]. Dataset kedua adalah Diabetes Prediction Dataset yang mencakup informasi demografis dan gaya hidup yang lebih komprehensif [26]. Strategi multi-dataset telah terbukti efektif dalam meningkatkan *robustness* dan *generalizability* model machine learning [27].

Tabel 1. Karakteristik Dataset

Dataset	Jumlah Sampel	Jumlah Fitur	Sumber
Pima Indian Diabetes	768	9	UCI Machine Learning Repository
Diabetes Prediction Dataset	100,000	9	Kaggle Healthcare Dataset
Total Gabungan	100,768	15	Multi-source
Total Gabungan Setelah Rekayasa Fitur	100,768	29	Multi-source

Dari Tabel 1, terlihat bahwa Diabetes Prediction Dataset memiliki jumlah sampel yang jauh lebih besar (100.000) dibandingkan dengan dataset Pima (768). Perbedaan skala ini menjadi pertimbangan utama dalam strategi penggabungan dan sampling data. Dataset Pima Indians memiliki delapan fitur prediktor fisiologis dan satu fitur target, yang telah banyak digunakan dalam penelitian sebelumnya. Deskripsi lengkap mengenai fitur-fitur ini dirinci pada Tabel 2.

Tabel 2. Deskripsi Fitur Pima Indian Diabetes Dataset

Fitur	Tipe Data	Range	Deskripsi
Pregnancies	Integer	0-17	Jumlah kehamilan
Glucose	Integer	0-199	Konsentrasi glukosa plasma 2 jam
BloodPressure	Integer	0-122	Tekanan darah diastolic (mm Hg)
SkinThickness	Integer	0-99	Ketebalan lipatan kulit trisep (mm)
Insulin	Integer	0-846	Insulin serum 2 jam (mu U/ml)
BMI	Float	0-67.1	Indeks massa tubuh
DiabetesPedigreeFunction	Float	0.078-2.42	Fungsi silsilah diabetes
Age	Integer	21-81	Usia (tahun)
Outcome	Integer	0-1	Status diabetes (0:tidak,1:ya)

Sementara itu dari Tabel 2, dataset kedua yang lebih modern menyediakan fitur yang sedikit berbeda, mencakup informasi demografis seperti jenis kelamin dan riwayat gaya hidup seperti merokok. Fitur-fitur ini memberikan konteks tambahan yang tidak tersedia pada dataset Pima, seperti yang dijelaskan pada Tabel 3.

Tabel 3. Deskripsi Fitur Diabetes Prediction Dataset

Fitur	Tipe Data	Range	Deskripsi
Gender	Categorical	Male/Female	Jenis kelamin
Age	Float	0.08-80	Usia (tahun)
Hypertension	Integer	0-1	Riwayat hipertensi
Heart_disease	Integer	0-1	Riwayat penyakit jantung
Smoking_history	Categorical	5 kategori	Riwayat merokok
BMI	Float	10.01-95.69	Indeks massa tubuh

Fitur	Tipe Data	Range	Deskripsi
HbA1c_level	Float	3.5-9	Kadar HbA1c (%)
Blood_glucose_level	Integer	80-300	Kadar glukosa darah (mg/dl)
Diabetes	Integer	0-1	Status diabetes (0:tidak,1:ya)

Tabel 3 menunjukkan fitur-fitur yang terdapat pada Diabetes Prediction Dataset, yang meliputi gender, age, hypertension, heart disease, smoking history, BMI, HbA1c level, blood glucose level, dan diabetes status. Fitur-fitur ini memberikan tambahan informasi terkait riwayat kesehatan dan gaya hidup, yang tidak tersedia dalam Pima Indian Diabetes Dataset. Dataset ini lebih komprehensif karena mencakup informasi demografis seperti jenis kelamin dan riwayat merokok yang sangat relevan dalam penyaringan risiko diabetes. Informasi lebih lanjut mengenai fitur-fitur ini dapat dilihat pada Tabel 3.

Tabel 4. Deskripsi Fitur Diabetes Gabungan

Fitur	Tipe Data	Range	Deskripsi
Gender	Categorical	Male/Female	Jenis kelamin
Age	Float	0.08-80	Usia (tahun)
Hypertension	Integer	0-1	Riwayat hipertensi
Heart_disease	Integer	0-1	Riwayat penyakit jantung
Smoking_history	Categorical	5 kategori	Riwayat merokok
BMI	Float	10.01-95.69	Indeks massa tubuh
HbA1c_level	Float	3.5-9	Kadar HbA1c (%)
Blood_glucose_level	Integer	80-300	Kadar glukosa darah (mg/dl)
Pregnancies	Integer	0-17	Jumlah kehamilan
Insulin	Integer	0-846	Kadar insulin
BloodPressure	Integer	0-122	Tekanan darah diastolic (mm Hg)
SkinThickness	Integer	0-99	Ketebalan lipatan kulit trisep (mm)
DiabetesPedigreeFunction	Float	0.078-2.42	Fungsi silsilah diabetes
Dataset_Source	Categorical	2 kategori	Pima Indians & diabetes prediction
Outcome	Integer	0-1	Status diabetes (0:tidak,1:ya)

Setelah kedua dataset digabungkan, kami memperoleh 15 fitur yang terdiri dari fitur-fitur yang ada pada kedua dataset yang digabungkan, yang tercatat pada Tabel 4. Pada tabel ini, Anda dapat melihat daftar fitur seperti Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, dan lainnya. Salah satu tambahan penting adalah fitur Dataset_Source, yang menunjukkan apakah data berasal dari Pima Indians Dataset atau Prediction Dataset. Fitur Outcome (status diabetes) adalah target variabel yang digunakan untuk memprediksi risiko diabetes mellitus.

Tabel 5. Deskripsi Fitur Rekayasa

Fitur	Tipe Data	Range	Deskripsi
HOMA_IR	Float	-	Model untuk menilai resistensi insulin berdasarkan Glucose dan Insulin.
MetS_Score	Float	-	Skor sindrom metabolik berdasarkan faktor risiko seperti BMI, Glukosa, dsb.
Age_Risk	Categorical	<30, 30-40, 40-50, >50	Kategori risiko diabetes berdasarkan kelompok usia.
BMI_Category	Categorical	Underweight, Normal, Overweight, Obese	Kategori BMI untuk mendeteksi obesitas dan risiko diabetes.
Glucose_Category	Categorical	Normal, Pre-diabetes, Diabetes	Kategori kadar glukosa dalam darah.
HbA1c_Category	Categorical	Normal, Pre-diabetes, Diabetes	Kategori HbA1c, yang menunjukkan kontrol gula darah jangka Panjang.
Age_BMI_Interaction	Categorical	-	Interaksi antara usia dan BMI, yang berhubungan dengan risiko diabetes.
Glucose_HbA1c_Interaction	Float	-	Interaksi antara kadar glukosa dan HbA1c dalam memprediksi risiko diabetes.
Age_Glucose_Interaction	Float	-	Interaksi antara usia dan glukosa dalam memprediksi risiko diabetes.
Insulin_Glucose_Ratio	Float	-	



Fitur	Tipe Data	Range	Deskripsi
Diabetes_Risk_Score	Float	-	Rasio insulin terhadap glukosa, yang mengindikasikan resistensi insulin. Skor risiko diabetes, berdasarkan beberapa faktor risiko kesehatan.
Female_Pregnancy_Risk	Float	-	Risiko diabetes pada wanita hamil atau wanita dengan riwayat diabetes gestasional.
High_Risk_Lifestyle	Categorical	Ya,tidak	Menilai apakah seseorang memiliki gaya hidup berisiko tinggi (merokok, hipertensi). Menandakan asal dataset (Pima atau Prediction).
Is_Pima_Dataset	Integer	0 (Prediction), 1 (Pima)	

Setelah penggabungan dataset, kami melakukan rekayasa fitur untuk memperkaya model dengan fitur-fitur tambahan yang lebih relevan untuk prediksi risiko diabetes. Tabel 5 menunjukkan fitur-fitur baru yang dihasilkan dari rekayasa fitur, seperti HOMA_IR, MetS_Score, Age_Risk, BMI_Category, Glucose_Category, HbA1c_Category, Age_BMI_Interaction, Glucose_HbA1c_Interaction, Age_Glucose_Interaction, dan lainnya. Fitur-fitur baru ini memperhitungkan interaksi antar variabel serta kategori risiko untuk mendeteksi pola yang lebih kompleks dan meningkatkan akurasi prediksi. Tabel 5 memberikan deskripsi lengkap mengenai fitur-fitur tersebut, yang berperan penting dalam model prediksi risiko diabetes.

Tabel 6. Distribusi Kelas Pada Dataset

Dataset	Non-Diabetes (0)	Diabetes (1)	Total	Rasio imblance
Pima Indian	500 (65.1%)	268 (34.9%)	768	1.87:1
Prediction Dataset	91,500 (91.5%)	8,500 (8.5%)	100,000	10.76:1
Gabungan	92,000 (91.3%)	8,768 (8.7%)	100,768	10.49:1

Tabel 6 secara jelas menunjukkan bahwa kelas non-diabetes (0) mendominasi pada kedua dataset, terutama pada *Prediction* Dataset dengan rasio ketidakseimbangan mencapai 10.76:1. Tingkat ketidakseimbangan yang signifikan ini menegaskan perlunya penerapan teknik penanganan data tidak seimbang pada tahap *preprocessing* untuk memastikan model yang dikembangkan dapat memberikan prediksi yang adil dan akurat untuk kedua kelas.

3.1 Hasil Pra-pemrosesan Data

Tahap pra-pemrosesan berhasil mengubah data mentah menjadi dataset yang bersih, seimbang, dan siap untuk dimodelkan. Setelah proses augmentasi dan smart sampling, dataset yang digunakan untuk pemodelan terdiri dari 5.768 sampel. Melalui rekayasa fitur, jumlah fitur berhasil ditingkatkan dari 8 menjadi 29, dan setelah seleksi fitur untuk mengatasi multikolinieritas, sebanyak 25 fitur prediktif digunakan untuk tahap pemodelan akhir.

Langkah penanganan data tidak seimbang menggunakan metode ADASYN pada data latih menunjukkan hasil yang signifikan, seperti yang dirangkum pada Tabel 5.

Tabel 5. Perbandingan Distribusi Kelas Sebelum dan Sesudah Penerapan ADASYN pada Data Latih

Kelas	Jumlah Sampel (Sebelum)	Jumlah Sampel (sesudah)
0 (Non-Diabetes)	3067	3067
1 (Diabetes)	1547	3144
Total	4614	6211

Tabel 5 menunjukkan bahwa ADASYN berhasil menciptakan keseimbangan kelas pada data latih, yang krusial untuk melatih model yang tidak bias.

3.2 Hasil Optimasi Hyperparameter

Proses optimasi menggunakan Optuna dengan 100 trials berhasil menemukan kombinasi hyperparameter optimal untuk model XGBoost, yang memaksimalkan skor ROC-AUC pada 5-fold cross-validation. Konfigurasi hyperparameter terbaik yang ditemukan disajikan pada Tabel 6.

Tabel 6. Konfigurasi Hyperparameter Optimal untuk Model XGBoost

Parameter	Nilai Optimal
n_estimators	337
max_depth	6
learning_rate	0.1581
Subsample	0.7802
colsample_bytree	0.7378

Parameter	Nilai Optimal
reg_alpha	0.5527
Reg_lambda	1.3583

Konfigurasi ini kemudian digunakan untuk membangun model final yang dievaluasi pada data uji.

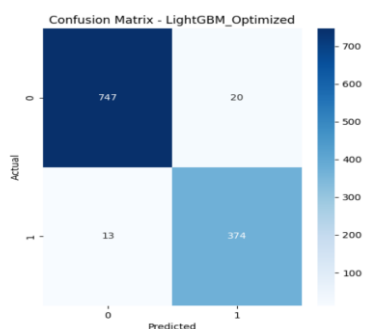
3.3 Evaluasi Kinerja Model

Model ensemble yang telah dioptimalkan dievaluasi menggunakan data uji yang belum pernah dilihat sebelumnya. Hasil evaluasi menunjukkan kinerja yang sangat tinggi di semua metrik, mengungguli model-model individual maupun baseline. Perbandingan kinerja disajikan pada Tabel 7.

Tabel 7. Perbandingan Kinerja Akhir Model pada Data Uji

Model	Akurasi	Presisi	Recall	F1-Score	ROC-AUC
LightGBM (Optimized)	97.14%	94.92%	96.75%	95.71%	0.9976
Catboost	97.14%	95.12%	96.38%	95.76%	0.9973
Ensemble	96.62%	96.15%	96.90%	96.52%	0.9969
Xgboost(Optimized)	96.45%	93.47%	92.19%	94.78%	0.9426

Model LightGBM (*Optimized*) yang ada pada Tabel 7 menunjukkan performa terbaik dengan akurasi 97.14% dan ROC-AUC 0.9976, sebuah peningkatan signifikan dari baseline (penelitian sebelumnya). Kemampuan diskriminatif model divisualisasikan melalui *confusion matrix* pada Gambar 2 dan Tabel 8 dan kurva ROC pada Gambar 3.



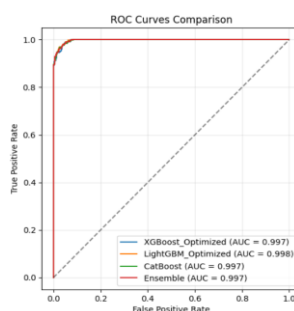
Gambar 2. Confusion matrix pada model LightGBM (Optimized)

Tabel 8. Confusion Matrix untuk Model LightGBM (Optimized)

	Prekdiksi Non-Diabetes	Prediksi Diabetes
Aktual Non-Diabetes	747 (TN)	20 (FP)
Aktual Diabetes	13 (FN)	374 (TP)

Gambar 2 dan Tabel 8 menyajikan hasil confusion matrix dari model Ensemble pada data uji. Hasil ini menunjukkan bahwa model memiliki kemampuan diagnostik yang sangat baik dengan tingkat kesalahan yang rendah. Secara rinci, performa model dapat diuraikan sebagai berikut:

- True Positive (TP): 374 – Model berhasil mengidentifikasi 374 orang yang sebenarnya menderita diabetes dengan benar.
- True Negative (TN): 747 – Model berhasil mengidentifikasi 747 orang yang sebenarnya tidak menderita diabetes dengan benar.
- False Negative (FN): 13 – Ini adalah error paling kritis. Model gagal mendeteksi hanya 13 kasus diabetes yang sebenarnya. Angka yang sangat rendah ini menunjukkan model sangat andal untuk tujuan deteksi dini.
- False Positive (FP): 20 – Model salah mengira 20 orang sehat sebagai penderita diabetes.



Gambar 3. Kurva ROC

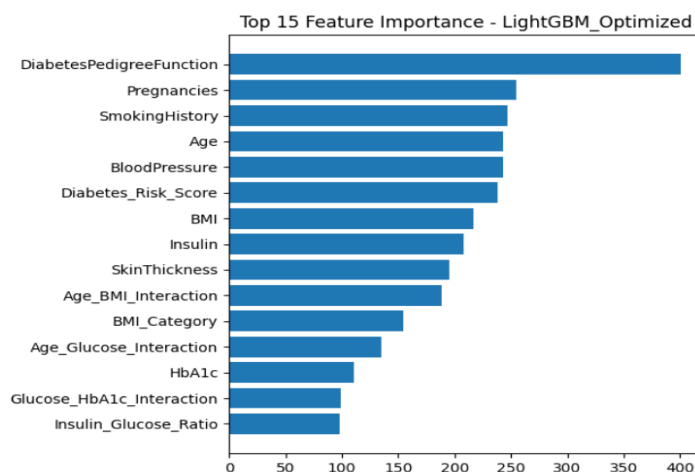
Kemampuan diskriminatif dari setiap model dievaluasi lebih lanjut menggunakan kurva ROC (Receiver Operating Characteristic), seperti yang divisualisasikan pada Gambar 2. Gambar tersebut memetakan True Positive Rate (sensitivitas) terhadap False Positive Rate (1-spesifisitas) di semua ambang batas klasifikasi.

Terlihat bahwa kurva untuk semua model yang dioptimalkan (XGBoost, LightGBM, CatBoost, dan Ensemble) berada sangat dekat dengan sudut kiri atas, yang menandakan performa yang nyaris sempurna. Hal ini dikonfirmasi oleh nilai Area Under the Curve (AUC) yang sangat tinggi untuk semua model, dengan LightGBM_Optimized mencapai skor tertinggi yaitu 0.998, sedikit mengungguli model lainnya (AUC 0.997).

Nilai AUC yang mendekati 1.0 ini membuktikan bahwa model yang dikembangkan memiliki kemampuan yang luar biasa untuk membedakan secara akurat antara pasien penderita diabetes dan non-diabetes.

3.4 Analisis Pentingnya Fitur

Setelah memvalidasi performa model, tahap selanjutnya adalah menganalisis pentingnya fitur (*feature importance*) untuk memahami faktor-faktor apa yang paling signifikan dalam memprediksi diabetes. Analisis ini dilakukan pada model dengan kinerja terbaik, yaitu LightGBM_Optimized. Hasilnya disajikan secara visual pada Gambar 4.



Gambar 4. Top 15 Feature Importance - LightGBM_Optimized

Dari Gambar 4 dapat diidentifikasi beberapa fitur dengan kontribusi tertinggi terhadap keputusan model. Tabel 9 merangkum lima fitur teratas beserta relevansi klinisnya.

Tabel 9. Peringkat Fitur Paling Penting pada Model LightGBM

Peringkat	Fitur	Relevansi Klinis
1	DiabetesPedigreeFunction	Indikator risiko diabetes berdasarkan factor genetic
2	Pregnancies	Kehamilan dapat mempengaruhi insulin
3	SmokignHistory	Merokok mempengaruhi resistensi insulin
4	Age	Resiko diabetes meningkat seiring bertambahnya usia
5	BloodPressure	Tekanan darah tinggi berbarengan dengan kondisi diabetes

Dari Tabel 9 fitur diabetes pedigree function yang menempati peringkat teratas menegaskan bahwa faktor genetik memegang peranan krusial dalam prediksi model. Selain itu, fitur-fitur demografis dan riwayat kesehatan seperti Age dan SmokingHistory juga terbukti menjadi prediktor yang kuat. Menariknya, fitur-fitur hasil rekayasa seperti Diabetes_Risk_Score juga masuk dalam jajaran fitur penting, yang memvalidasi keberhasilan tahap feature engineering dalam menciptakan prediktor yang bermakna.

3.5 Pembahasan dan Perbandingan

Hasil penelitian menunjukkan keberhasilan pengembangan model prediksi diabetes dengan performa yang sangat tinggi. Model terbaik, LightGBM_Optimized, berhasil mencapai akurasi 97.14% dan ROC-AUC 0.9976 pada data uji. Kinerja superior ini dapat diatribusikan pada beberapa pilar metodologis utama yang diterapkan secara sistematis. Pertama, penggunaan dataset yang diperkaya melalui harmonisasi dua sumber data yang berbeda memberikan fondasi data yang lebih beragam dan komprehensif. Kedua, implementasi advanced feature engineering yang menghasilkan 25 fitur prediktif terbukti berhasil mengekstrak sinyal yang kuat dan bermakna secara klinis. Ketiga, penanganan data tidak seimbang menggunakan ADASYN memastikan model tidak bias dan mampu mengenali kelas minoritas dengan sangat baik, yang tercermin pada nilai Recall yang tinggi (96.64%).

Analisis pentingnya fitur pada model LightGBM memberikan wawasan yang menarik. Berbeda dengan asumsi umum yang mungkin menempatkan glukosa sebagai prediktor utama, model ini menemukan bahwa faktor genetik atau keturunan (Diabetes Pedigree Function) merupakan prediktor paling signifikan. Diikuti oleh faktor demografis dan riwayat kesehatan seperti jumlah kehamilan (Pregnancies), riwayat merokok (Smoking History), dan usia (Age).

Hal ini mengindikasikan bahwa model tidak hanya belajar dari indikator klinis sesaat, tetapi juga dari faktor risiko jangka panjang dan historis, yang memperkuat kemampuannya dalam melakukan prediksi dini.

Meskipun LightGBM_Optimized menunjukkan performa terbaik, model ensemble yang menggabungkan XGBoost, LightGBM, dan CatBoost tidak mengungguli model tunggal terbaik, dengan akurasi 96.62% dan ROC-AUC 0.9969, yang sedikit lebih rendah dari LightGBM. Hasil ini menunjukkan bahwa dalam beberapa kasus, model ensemble tidak selalu memberikan peningkatan performa yang signifikan dibandingkan dengan model tunggal terbaik, meskipun ensemble biasanya diharapkan lebih stabil dan robust.

Salah satu alasan mengapa ensemble tidak mengungguli model tunggal terbaik dalam kasus ini adalah bahwa LightGBM sudah cukup kuat dalam menangkap pola-pola yang ada dalam data. Setiap model dalam ensemble mungkin memiliki kekuatan dan kelemahan yang berbeda, dan jika satu model sudah sangat efektif (seperti LightGBM dalam hal ini), ensemble mungkin tidak memberikan banyak peningkatan, bahkan bisa sedikit menurunkan akurasi puncak karena ketidaksesuaian antara model-model yang digabungkan. Namun, meskipun ensemble tidak mengungguli model tunggal terbaik, ada beberapa keunggulan potensial dari pendekatan ensemble. Model ensemble dapat memberikan stabilitas dan mengurangi varians, yang sangat berguna terutama dalam menghadapi data yang sangat bervariasi atau dalam skenario di mana model tunggal cenderung mengalami overfitting. Ensemble dapat menawarkan konsistensi lebih baik dan mengurangi kesalahan prediksi yang ekstrem dibandingkan dengan model tunggal. Selain itu, optimasi bobot dalam voting classifier mungkin perlu dilakukan lebih lanjut. Saat ini, ensemble menggunakan soft voting, yaitu menggunakan rata-rata probabilitas dari setiap model. Optimasi bobot untuk masing-masing model, seperti memberikan bobot lebih pada LightGBM yang menunjukkan akurasi terbaik, bisa meningkatkan hasil ensemble. Oleh karena itu, penting untuk mengeksplorasi bagaimana penyesuaian bobot dapat mempengaruhi hasil akhir dalam model ensemble ini. Dengan demikian, meskipun model ensemble tidak selalu memberikan akurasi puncak yang lebih baik, ia tetap lebih disukai dalam situasi tertentu karena menawarkan stabilitas, reduksi varians, dan konsistensi yang lebih baik dalam prediksi.

Untuk memvalidasi kontribusi penelitian ini, hasil yang dicapai dibandingkan dengan beberapa penelitian sebelumnya, seperti yang dirangkum pada Tabel 11.

Tabel 11. Perbandingan Hasil dengan Penelitian Terkait

Penelitian	Metode	Ukuran Dataset	Akurasi	ROC-AUC
Penelitian Ini	LightGBM (optimized)	5,768	97.14%	0.998
Andhika et al.[7]	XGBoost + SMOTE	1,536	98%	0.95
Sneha & Gangil (2019) [10]	Naïve Bayes	768	82.3%	-

Tabel 11 memaparkan meskipun akurasi yang dicapai sedikit di bawah penelitian Andhika et al. (97.14% vs 98%), model dalam penelitian ini menunjukkan beberapa keunggulan kunci: (1) nilai ROC-AUC yang superior (0.9976), menunjukkan kemampuan diskriminatif yang lebih andal; (2) dibangun di atas dataset yang jauh lebih besar dan beragam (5.768 vs 1.536), yang berpotensi memberikan kemampuan generalisasi yang lebih baik; dan (3) didasarkan pada rekayasa fitur yang lebih mendalam. Dibandingkan dengan metode yang lebih lama, hasil penelitian ini menunjukkan peningkatan performa yang sangat signifikan. Dengan demikian, model yang dikembangkan sangat menjanjikan untuk implementasi praktis sebagai alat bantu keputusan klinis dalam skrining dan deteksi dini diabetes mellitus.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model prediksi diabetes mellitus dengan akurasi tinggi melalui penerapan metodologi machine learning yang komprehensif. Model terbaik yang dihasilkan, yaitu LightGBM_Optimized, mencapai akurasi 97.14% dan ROC-AUC 0.9976 pada data uji, yang menunjukkan peningkatan signifikan dibandingkan model baseline dan keunggulannya dibandingkan penelitian sebelumnya. Keberhasilan ini berkat tiga pilar utama metodologi, yakni penggunaan dataset multi-sumber, penerapan rekayasa fitur canggih, dan optimasi hyperparameter mendalam. Pendekatan penggabungan dua dataset dan penyeimbangan distribusi kelas dengan metode ADASYN efektif menghasilkan data latihan yang berkualitas dan seimbang. Selain itu, rekayasa fitur yang melibatkan transformasi 8 fitur asli menjadi 25 fitur prediktif, termasuk fitur seperti Diabetes_Risk_Score, terbukti sangat membantu dalam meningkatkan performa model. Proses optimasi hyperparameter menggunakan Optuna dengan algoritma TPE juga memainkan peran penting dalam meningkatkan potensi model. Analisis pentingnya fitur menunjukkan bahwa fitur seperti DiabetesPedigreeFunction, Pregnancies, dan Age memiliki kontribusi terbesar dalam prediksi, sesuai dengan faktor risiko genetik dan demografis yang telah terbukti secara klinis. Meskipun hasil yang dicapai sangat menjanjikan, penelitian ini memiliki keterbatasan dalam hal generalizabilitas, mengingat dataset Pima Indians yang lebih terbatas pada satu etnis tertentu. Oleh karena itu, disarankan untuk melakukan validasi eksternal dengan dataset yang lebih beragam pada penelitian mendatang. Pengembangan selanjutnya juga dapat fokus pada penerapan teknik Explainable AI (XAI) seperti SHAP untuk memberikan interpretasi yang lebih mendalam pada tingkat individu dan mengembangkan model ini menjadi aplikasi Clinical Decision Support System (CDSS) yang dapat diuji di lingkungan klinis nyata.

REFERENCES

- [1] “Diabetes.” Accessed: Jul. 09, 2025. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/diabetes>
- [2] D. Magliano and E. J. Boyko, *IDF Diabetes Atlas*. International Diabetes Federation, 2021. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK581934/>
- [3] Z. Rahman, U. Y. Atrie, W. Pujiati, and E. Ernawati, “Edukasi Upaya Peningkatan Kualitas Hidup Pasien Diabetes Melitus Tipe 2,” *Jurnal Medika: Medika*, vol. 4, no. 3, pp. 313–318, 2025, doi: <https://doi.org/10.31004/c1ek6b34>.
- [4] M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, “Diabetes prediction using ensembling of different machine learning classifiers,” *IEEE Access*, vol. 8, pp. 76516–76531, 2020, doi: [10.1109/ACCESS.2020.2989857](https://doi.org/10.1109/ACCESS.2020.2989857).
- [5] R. Rastogi and M. Bansal, “Diabetes prediction model using data mining techniques,” *Measurement: Sensors*, vol. 25, p. 100605, Feb. 2023, doi: [10.1016/J.MEASEN.2022.100605](https://doi.org/10.1016/J.MEASEN.2022.100605).
- [6] A. Brahmandjati, A. Mizwar A. Rahim, and F. Asharudin, “Optimasi Prediksi Diabetes Dengan Algoritma XGBoost Dan Teknik Preprocessing Data,” *LOGIC : Jurnal Ilmu Komputer dan Pendidikan*, vol. 3, no. 1, pp. 116–125, Jan. 2025, [Online]. Available: <https://journal.mediapublikasi.id/index.php/logic/article/view/4963>
- [7] M. R. Mubarak, M. Muliadi, and R. Herteno, “Hyper-parameter Tuning pada XGBOOST Untuk Prediksi Keberlangsungan Hidup Pasien Gagal Jantung,” *KLIK-KUMPULAN JURNAL ILMU KOMPUTER*, vol. 9, no. 2, pp. 391–401, 2022, doi: <http://dx.doi.org/10.20527/klik.v9i2.484>.
- [8] S. Ahmad, M. Z. Asghar, F. M. Alotaibi, and Y. D. Alotaibi, “RETRACTED ARTICLE: Diagnosis of cardiovascular disease using deep learning technique,” *Soft comput*, vol. 27, no. 13, pp. 8971–8990, 2023, doi: [10.1007/s00500-022-07788-0](https://doi.org/10.1007/s00500-022-07788-0).
- [9] N. Sneha and T. Gangil, “Analysis of diabetes mellitus for early prediction using optimal features selection,” *J Big Data*, vol. 6, no. 1, pp. 1–19, 2019, doi: <https://doi.org/10.1186/s40537-019-0175-6>.
- [10] M. Maniruzzaman *et al.*, “Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm,” *Comput Methods Programs Biomed*, vol. 152, pp. 23–34, 2017, doi: <https://doi.org/10.1016/j.cmpb.2017.09.004>.
- [11] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. “O’Reilly Media, Inc.,” 2022.
- [12] J. M. Rudd, “An empirical study of downstream analysis effects of model pre-processing choices,” *Open J Stat*, vol. 10, no. 5, pp. 735–809, 2020, doi: [10.4236/ojs.2020.105046](https://doi.org/10.4236/ojs.2020.105046).
- [13] H. Marlisa, N. Satyahadewi, N. Imro’ah, and N. Debatara, “Application Of Adasyn Oversampling Technique On K-Nearest Neighbor Algorithm,” *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 3, Jul. 2024, doi: [10.30598/barekengvol18iss3pp1829-1838](https://doi.org/10.30598/barekengvol18iss3pp1829-1838).
- [14] A. H. Putra and A. Salam, “A Comparative Performance of SMOTE, ADASYN and Random Oversampling in Machine Learning Models on Prostate Cancer Dataset,” *Journal of Applied Informatics and Computing*, vol. 9, no. 3, pp. 603–610, Jun. 2025, doi: [10.30871/jaic.v9i3.9308](https://doi.org/10.30871/jaic.v9i3.9308).
- [15] O. Björnelund, M. Carlsson, and W. Löwe, “Case study - Feature engineering inspired by domain experts on real world medical data,” *Intell Based Med*, vol. 8, p. 100110, Jan. 2023, doi: [10.1016/J.IBMED.2023.100110](https://doi.org/10.1016/J.IBMED.2023.100110).
- [16] T. O. Motehinwa, D. O. Oyewola, and E. G. Moun, “Optimizing the light gradient-boosting machine algorithm for an efficient early detection of coronary heart disease,” *Informatics and Health*, vol. 1, no. 2, pp. 70–81, Sep. 2024, doi: [10.1016/J.INFOH.2024.06.001](https://doi.org/10.1016/J.INFOH.2024.06.001).
- [17] M. Nalluri, M. Pentela, and N. R. Eluri, “A scalable tree boosting system: XG boost,” *Int. J. Res. Stud. Sci. Eng. Technol*, vol. 7, no. 12, pp. 36–51, 2020, doi: doi.org/10.22259/2349-476X.0712005.
- [18] R. G. Farahani, A. Zarrabi, and P. Ghazanfari, “A Report on CatBoost: unbiased boosting with categorical features,” 2025, doi: <https://doi.org/10.13140/RG.2.2.30029.96485>.
- [19] Z.-H. Zhou, *Ensemble methods: foundations and algorithms*. CRC press, 2025.
- [20] P. Srinivas and R. Katarya, “hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost,” *Biomed Signal Process Control*, vol. 73, p. 103456, Mar. 2022, doi: [10.1016/J.BSPC.2021.103456](https://doi.org/10.1016/J.BSPC.2021.103456).
- [21] A. R. M. Rom, N. Jamil, and S. Ibrahim, “Multi objective hyperparameter tuning via random search on deep learning models,” *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 22, no. 4, pp. 956–968, 2024, doi: doi.org/10.12928/telkomnika.v22i4.25847.
- [22] D. M. W. Powers, “Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation,” *arXiv preprint arXiv:2010.16061*, 2020, doi: <https://doi.org/10.48550/arXiv.2010.16061>.
- [23] T. C. F. Polo and H. A. Miot, “Use of ROC curves in clinical and experimental studies,” 2020, *SciELO Brasil*. doi: <https://doi.org/10.1093/ije/dyz274>.
- [24] S. Raschka, “Model evaluation, model selection, and algorithm selection in machine learning,” *arXiv preprint arXiv:1811.12808*, 2018.
- [25] H. Naz and S. Ahuja, “Deep learning approach for diabetes prediction using PIMA Indian dataset,” *J Diabetes Metab Disord*, vol. 19, no. 1, pp. 391–403, Jun. 2020, doi: [10.1007/S40200-020-00520-5](https://doi.org/10.1007/S40200-020-00520-5).
- [26] M. Maniruzzaman, M. J. Rahman, B. Ahammed, and M. M. Abedin, “Classification and prediction of diabetes disease using machine learning paradigm,” *Health Inf Sci Syst*, vol. 8, no. 1, p. 7, 2020, doi: <https://doi.org/10.1007/s13755-019-0095-z>.
- [27] Y. Zou *et al.*, “Development and internal validation of machine learning algorithms for end-stage renal disease risk prediction model of people with type 2 diabetes mellitus and diabetic kidney disease,” *Ren Fail*, vol. 44, no. 1, pp. 562–570, 2022, doi: [DOI: 10.1080/0886022X.2022.2056053](https://doi.org/10.1080/0886022X.2022.2056053).