Implementasi *Algoritma Random Forest, Logistic Regression,* dan *Gradient Boosting* untuk Deteksi Dini Diabetes Mellitus

Larry Anthonio Ruitan[1*], Ngurah Daniel Palar[1], Andreuw Vandy Lengkong[1], Dyah Listianing Tyas[1]
[1]*Jurusan Pendidikan Teknik Informatika, Universitas Prisma, Manado, Indonesia*
Email:ruitan125@gmail.com[*]

***Abstract*— Diabetes Mellitus is a chronic metabolic disease that requires early screening to reduce complications. This study implements Random Forest, Logistic Regression, and Gradient Boosting for early diabetes risk screening using the CDC Diabetes Health Indicators dataset. The revised study strengthens the experimental section by reporting class distribution, hyperparameter configuration, confusion matrix, ROC-AUC, PR-AUC, Matthews Correlation Coefficient (MCC), Cohen Kappa, feature importance, McNemar statistical testing, and functional testing of the web application. The dataset used in this study consists of 70,692 records with 21 predictor features and a balanced class distribution (50% non-diabetes and 50% diabetes). Experimental results using a 70:30 train-test split show that Random Forest achieved the best performance with 75.64% accuracy, 73.21% precision, 80.88% recall, 76.85% F1-score, 83.51% ROC-AUC, and 81.48% PR-AUC. The most influential features based on Gradient Boosting feature importance are HighBP, GenHlth, BMI, Age, and HighChol. McNemar testing between Gradient Boosting and Random Forest produced a p-value of 0.00009, indicating a statistically significant difference. The system is implemented as a mobile-responsive web application using Flask with an alternative Streamlit interface and can be used as an early screening and educational tool, not as a substitute for medical diagnosis.**

***Keywords*— Diabetes Mellitus, Machine Learning, Gradient Boosting, Random Forest, Logistic Regression, ROC-AUC, Feature Importance, Class Imbalance, Web Application.**

***Abstrak*— Diabetes Mellitus merupakan penyakit metabolik kronis yang membutuhkan skrining dini untuk mengurangi risiko komplikasi. Penelitian ini mengimplementasikan algoritma Random Forest, Logistic Regression, dan Gradient Boosting untuk skrining risiko diabetes menggunakan dataset CDC Diabetes Health Indicators. Artikel revisi diperkuat dengan pelaporan distribusi kelas, dokumentasi hyperparameter, confusion matrix, ROC-AUC, PR-AUC, Matthews Correlation Coefficient (MCC), Cohen Kappa, feature importance, uji statistik McNemar, dan pengujian fungsional aplikasi web. Dataset yang digunakan terdiri dari 70.692 data dengan 21 fitur prediktor dan distribusi kelas seimbang (50% tidak diabetes dan 50% diabetes). Hasil eksperimen dengan pembagian data latih dan uji 70:30 menunjukkan bahwa Random Forest memperoleh performa terbaik dengan akurasi 75,64%, presisi 73,21%, recall 80,88%, F1-score 76,85%, ROC-AUC 83,51%, dan PR-AUC 81,48%. Fitur yang paling berpengaruh berdasarkan feature importance Gradient Boosting adalah HighBP, GenHlth, BMI, Age, dan HighChol. Uji McNemar antara Gradient Boosting dan Random Forest menghasilkan p-value 0,00009 yang menunjukkan adanya perbedaan performa yang signifikan. Sistem diimplementasikan dalam aplikasi web mobile-responsive menggunakan Flask dengan antarmuka alternatif Streamlit dan dapat digunakan sebagai alat bantu skrining awal serta edukasi kesehatan, bukan pengganti diagnosis medis.**

***Kata Kunci*— Diabetes Mellitus, Machine Learning, Gradient Boosting, Random Forest, Logistic Regression, ROC-AUC, Feature Importance, Class Imbalance, Aplikasi Web.**

I. PENDAHULUAN

Diabetes Mellitus merupakan salah satu penyakit metabolik kronis yang menjadi masalah kesehatan global. Peningkatan jumlah penderita diabetes berdampak pada beban biaya kesehatan, penurunan produktivitas, dan meningkatnya risiko komplikasi. Deteksi dini menjadi langkah penting agar individu berisiko dapat memperoleh edukasi dan pemeriksaan lanjutan sebelum terjadi komplikasi yang lebih berat [1].

Machine Learning dapat digunakan untuk mempelajari pola dari data kesehatan, seperti usia, indeks massa tubuh (BMI), tekanan darah, kolesterol, aktivitas fisik, dan riwayat penyakit. Beberapa penelitian sebelumnya menunjukkan bahwa algoritma ensemble learning dan model klasifikasi konvensional mampu membantu proses prediksi risiko Diabetes Mellitus [2][5].

Berbeda dari penelitian yang hanya membandingkan beberapa algoritma standar dan memilih model berdasarkan akurasi tertinggi, penelitian ini memperluas evaluasi dengan menambahkan analisis distribusi kelas, *confusion matrix, ROC-AUC, PR-AUC, Matthews Correlation Coefficient (MCC), Cohen Kappa, feature importance*, dan uji statistik McNemar. Penambahan evaluasi tersebut penting karena kasus kesehatan tidak cukup dinilai melalui accuracy saja, terutama ketika kesalahan false negative dapat berdampak pada keterlambatan pemeriksaan lanjutan.

Kontribusi penelitian ini terletak pada integrasi eksperimen machine learning dengan implementasi sistem skrining berbasis web mobile-responsive. Sistem tidak hanya menampilkan hasil prediksi, tetapi juga menyajikan

kategori risiko, faktor risiko dominan, rekomendasi kesehatan, visualisasi metrik model, konsultasi AI edukatif, dan riwayat skrining. Dengan demikian, penelitian ini memadukan aspek performa model, interpretabilitas, dan implementasi aplikasi sebagai alat bantu skrining awal.

Tujuan penelitian ini adalah: (1) membangun model klasifikasi risiko Diabetes Mellitus menggunakan Random Forest, Logistic Regression, dan Gradient Boosting; (2) melakukan evaluasi komprehensif terhadap performa model; (3) menganalisis faktor penting yang memengaruhi prediksi; dan (4) mengimplementasikan model terbaik ke dalam aplikasi web skrining mandiri.

II. METODOLOGI

A. *Jenis dan Pendekatan Penelitian*

Penelitian ini menggunakan pendekatan *Applied Machine Learning Research with Web-Based System Implementation*. Pendekatan ini dipilih karena fokus utama penelitian adalah eksperimen klasifikasi risiko Diabetes Mellitus menggunakan *machine learning*, kemudian model yang diperoleh diimplementasikan ke dalam aplikasi web. Tahapan pengembangan sistem digunakan untuk mendukung implementasi dan pengujian aplikasi, sedangkan evaluasi ilmiah difokuskan pada performa dan interpretabilitas model.

B. *Dataset dan Distribusi Kelas*

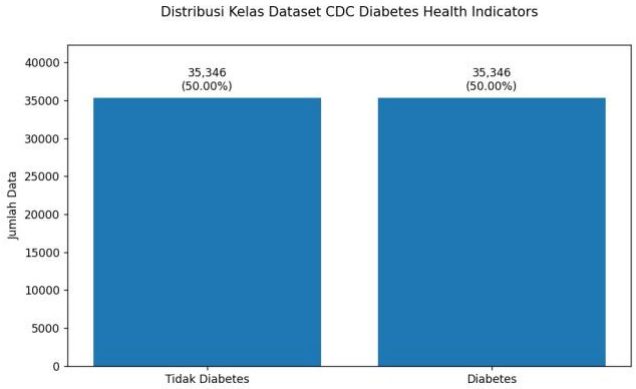
Dataset yang digunakan adalah CDC Diabetes Health Indicators versi balanced binary subset yang terdiri dari 70.692 data responden dengan 21 fitur prediktor dan satu target Diabetes_binary. Fitur yang digunakan meliputi HighBP, HighChol, CholCheck, BMI, Smoker, Stroke, HeartDiseaseorAttack, PhysActivity, Fruits, Veggies, HvyAlcoholConsump, AnyHealthcare, NoDocbcCost, GenHlth, MentHlth, PhysHlth, DiffWalk, Sex, Age, Education, dan Income.

Analisis distribusi kelas menunjukkan bahwa dataset yang digunakan dalam penelitian ini bersifat seimbang, yaitu 35.346 data kelas tidak diabetes dan 35.346 data kelas diabetes, masing-masing sebesar 50%. Dengan demikian, penelitian ini tidak memerlukan teknik penyeimbangan tambahan seperti SMOTE atau class_weight, tetapi analisis distribusi kelas tetap dilaporkan untuk memastikan kualitas eksperimen dan menjawab catatan revisi reviewer.

TABEL I. DISTRIBUSI KELAS DATASET CDC DIABETES HEALTH INDICATORS YANG DIGUNAKAN PADA PENELITIAN.

Kelas	Label	Jumlah Data	Persentase
Tidak Diabetes	0	35.346	50%
Diabetes	1	35.346	50%
Total	-	70.692	100%

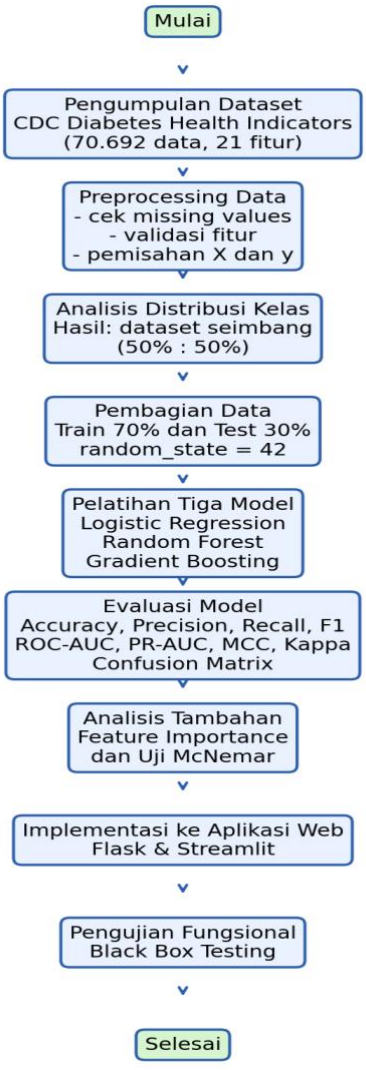
Hasil distribusi kelas juga divisualisasikan pada Gbr. 1 untuk memperjelas komposisi data yang digunakan.



Gbr. 1. Visualisasi distribusi kelas dataset.

C. *Alur Penelitian*

Alur penelitian dimulai dari pengumpulan dataset, preprocessing, analisis distribusi kelas, pembagian data latih dan uji, pelatihan model, evaluasi komprehensif, analisis feature importance, uji statistik McNemar, implementasi aplikasi web, dan pengujian fungsional sistem.



Gbr. 2. Alur metodologi penelitian

D. Preprocessing dan Rekayasa Fitur

Tahap preprocessing meliputi pemeriksaan missing values, validasi tipe data, pengecekan konsistensi nilai, dan pemisahan fitur serta target. Dataset dibagi menjadi data latih 70% dan data uji 30% menggunakan stratified train-test split dengan random_state=42 agar proporsi kelas tetap terjaga dan eksperimen dapat direproduksi. Pada aplikasi, tinggi dan berat badan diproses menjadi BMI menggunakan rumus pada rumus (1).

$$BMI=(\text{Tinggi Badan (m)})2\text{Berat Badan (kg)} \quad (1)$$

Usia asli pengguna dikonversi ke kategori usia CDC melalui fungsi age_to_cdc_bucket() agar sesuai dengan format fitur Age pada dataset.Karena distribusi kelas pada dataset penelitian ini seimbang, maka tidak dilakukan teknik balancing tambahan seperti SMOTE maupun penyesuaian class_weight. Keputusan ini diambil agar distribusi data asli tetap terjaga dan hasil evaluasi merepresentasikan performa model pada data balanced subset yang memang digunakan dalam penelitian.

Analisis imbalance tetap dilaporkan sebagai bagian dari revisi mayor agar pembaca mengetahui bahwa keputusan untuk tidak melakukan balancing didasarkan pada hasil eksplorasi data, bukan diabaikan. Dengan demikian, fokus peningkatan metodologi diarahkan pada evaluasi komprehensif, interpretasi model, dan pengujian statistik.

Hasil eksplorasi data menunjukkan proporsi kelas 50:50 sehingga baseline tanpa balancing digunakan sebagai skenario eksperimen final.

E. Hyperparameter Model

Untuk meningkatkan reproduksibilitas penelitian, revisi ini menambahkan dokumentasi hyperparameter final yang digunakan pada setiap model. Konfigurasi parameter diambil dari model .pkl yang digunakan pada aplikasi dan ditampilkan secara eksplisit sehingga eksperimen dapat direplikasi.

TABEL III. RUANG PENCARIAN HYPERPARAMETER YANG DAPAT DIGUNAKAN UNTUK REPLIKASI EKSPERIMEN.

Model	**Parameter yang Dioptimasi**	**Nilai Kandidat**
Random Forest	n_estimators, max_depth, min_samples_split, class_weight	100, 200; None, 10, 20; 2, 5; None, balanced
Gradient Boosting	n_estimators, learning_rate, max_depth, subsample	100, 200; 0,05, 0,1; 2, 3; 0,8, 1,0
Logistic Regression	solver, penalty, C, class_weight	lbfgs/liblinear; l2; 0,1, 1, 10; None, balanced

F. Pelatihan Model

Tiga algoritma yang digunakan adalah Random Forest, Logistic Regression, dan Gradient Boosting. Random Forest digunakan karena mampu menggabungkan banyak decision tree untuk meningkatkan stabilitas prediksi. Logistic Regression digunakan sebagai baseline linear yang mudah diinterpretasikan. Gradient Boosting digunakan karena membangun model secara bertahap untuk memperbaiki kesalahan model sebelumnya [6][8].

G. Evaluasi Model

Evaluasi model dilakukan menggunakan confusion matrix, accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC, Matthews Correlation Coefficient (MCC), dan Cohen Kappa. Pada kasus kesehatan, recall dan false negative menjadi perhatian penting karena pasien berisiko yang diprediksi tidak berisiko dapat menyebabkan keterlambatan pemeriksaan lanjutan.

Selain itu, uji statistik McNemar digunakan untuk membandingkan dua model terbaik pada data uji yang sama. Uji ini membantu menentukan apakah selisih performa antara Random Forest dan Gradient Boosting signifikan atau hanya terjadi karena variasi data uji.

H. Implementasi Aplikasi Web

Aplikasi dikembangkan menggunakan arsitektur client-server. Frontend dibangun menggunakan HTML, Tailwind CSS, dan JavaScript, sedangkan backend menggunakan Flask. Model machine learning disimpan dalam format .pkl dan dimuat melalui pustaka joblib. Endpoint utama meliputi GET /, GET /diabetes, POST /api/predict, GET /api/metrics, dan POST /api/chat. Aplikasi alternatif menggunakan Streamlit dengan empat tab utama, yaitu Skrining Mandiri, Analisis Model ML, Konsultasi AI, dan Riwayat Pasien. Struktur ini disesuaikan dengan rancangan sistem pada penelitian skripsi sehingga artikel jurnal tetap selaras dengan dokumen penelitian utama.

I. Pengujian Aplikasi

Pengujian aplikasi dilakukan menggunakan black box testing untuk memastikan fungsi utama berjalan sesuai kebutuhan. Pengujian meliputi tampilan halaman utama, perhitungan BMI otomatis, validasi input, prediksi risiko, visualisasi metrik, konsultasi AI, riwayat skrining, dan penanganan kesalahan ketika model atau API key tidak tersedia.

III. HASIL DAN PEMBAHASAN

A. Analisis Imbalance Dataset

Distribusi kelas menunjukkan bahwa dataset yang digunakan bersifat seimbang, yaitu 50% kelas tidak diabetes dan 50% kelas diabetes. Temuan ini menegaskan bahwa peningkatan kualitas penelitian pada revisi mayor tidak difokuskan pada teknik balancing, tetapi pada evaluasi model yang lebih komprehensif, interpretabilitas, dan validasi statistik.

B. Konfigurasi Hyperparameter Model

Tabel III menyajikan konfigurasi hyperparameter final dari ketiga model yang digunakan pada eksperimen. Dokumentasi ini penting agar penelitian dapat direplikasi secara lebih jelas.

C. Evaluasi Kinerja Model

Hasil evaluasi menunjukkan bahwa *Random Forest* memperoleh performa terbaik berdasarkan *accuracy, recall, F1-score, ROC-AUC, PR-AUC, MCC,* dan *Cohen Kappa* (dapat dilihat pda Tabel IV). Pada Tabel III dapat dilihat konfigurasi Hyperparameter model yang digunakan pada

setiap model. Sementara itu, *Logistic Regression* memberikan baseline linear yang cukup kompetitif dan Gradient Boosting tetap menunjukkan performa yang baik sebagai model ensemble boosting.

TABEL IIIII. KONFIGURASI HYPERPARAMETER MODEL YANG DIGUNAKAN.

Model	Konfigurasi Parameter	Sumber	Keterangan
Gradient Boosting	n_estimators=100; learning_rate=0,1; max_depth=3; subsample=1,0; random_state=42	model_gb.pkl	Model boosting
Random Forest	n_estimators=100; max_depth=12; criterion=gini; random_state=42	model_rf.pkl	Model dengan performa terbaik
Logistic Regression	C=1,0; solver=lbfgs; max_iter=1000; random_state=42	model_lr.pkl	Baseline linear

TABEL IV. EVALUASI KOMPREHENSIF MODEL ACCURACY, PRECISION, RECALL, F1.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Gradient Boosting	75,07	72,78	80,07	76,26
Random Forest	75,64	73,21	80,88	76,85
Logistic Regression	74,29	73,12	76,82	74,93

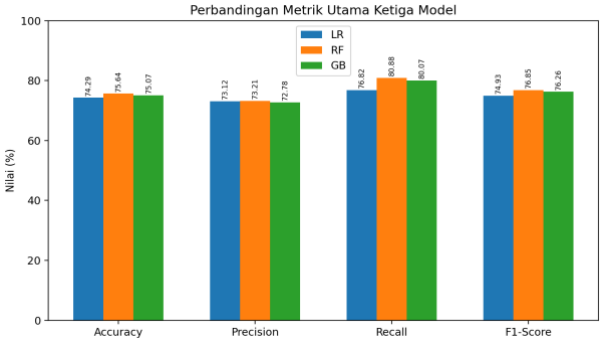
TABEL V. EVALUASI KOMPREHENSIF MODEL ROC-AUC,PR-AUC,MCC,KAPPA.

Model	ROC-AUC	PR-AUC	MCC	Kappa
Gradient Boosting	82,8	79,98	0,5039	0,5013
Random Forest	83,51	81,48	0,5157	0,5128
Logistic Regression	81,95	78,51	0,4865	0,4859

D. Evaluasi Kinerja Model

Confusion matrix seperti yang dapat dilihat pada Tabel VI, digunakan untuk melihat jumlah *True Positive, True Negative, False Positive,* dan *False Negative*. Pada Gambar 3, dapat dilihat evaluasi akurasi, *precision, recall, dan F1-score.* Pada kasus diabetes, *False Negative* perlu diperhatikan karena individu yang sebenarnya berisiko diabetes dapat dikategorikan tidak berisiko oleh sistem.

Berdasarkan *confusion matrix, Random Forest* menghasilkan jumlah false negative terendah sebanyak 2.027 kasus, lebih rendah dibandingkan *Gradient Boosting* sebanyak 2.113 kasus dan *Logistic Regression* sebanyak 2.458 kasus. Temuan ini penting karena pada konteks kesehatan, *false negative* berisiko membuat individu yang sebenarnya berisiko tidak terdeteksi lebih awal.



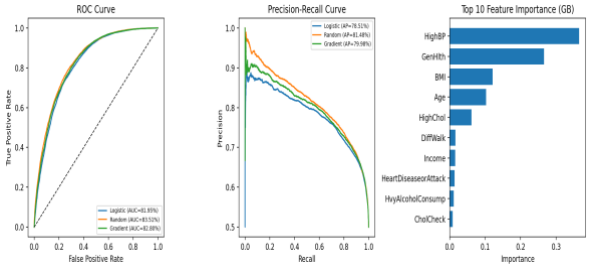
Gbr. 3. Diagram batang evaluasi accuracy, precision, recall, dan F1-score.

TABEL VI. CONFUSION MATRIX SETIAP MODEL.

Model	TN	FP	FN	TP
Gradient Boosting	7.429	3.175	2.113	8.491
Random Forest	7.465	3.139	2.027	8.577
Logistic Regression	7.610	2.994	2.458	8.146

E. Evaluasi Kinerja Model

ROC Curve seperti yang dapat dilihat pada Gbr.4, digunakan untuk melihat kemampuan model membedakan kelas positif dan negatif pada berbagai threshold, sedangkan Precision-Recall Curve digunakan sebagai pelengkap untuk mengevaluasi kualitas prediksi kelas positif. Pada revisi ini, visualisasi juga dilengkapi dengan feature importance untuk memperkuat interpretasi hasil model.



Gbr. 4. ROC Curve, Precision-Recall Curve, dan Feature Importance hasil eksperimen.

F. Feature Importance

Gradient Boosting dan *Random Forest* menyediakan *feature importance* sehingga model dapat dianalisis untuk mengetahui faktor yang paling berpengaruh terhadap prediksi risiko diabetes. Analisis ini meningkatkan interpretabilitas model dan menjawab pertanyaan faktor risiko dominan, seperti BMI, Age, HighBP, HighChol, GenHlth, atau DiffWalk.

Hasil *feature importance* pada Tabel VII menunjukkan bahwa faktor yang paling dominan dalam prediksi risiko diabetes adalah HighBP, GenHlth, BMI, Age, dan HighChol. Informasi ini mendukung interpretasi bahwa tekanan darah tinggi, kondisi kesehatan umum, indeks massa tubuh, usia, dan kolesterol tinggi merupakan indikator penting dalam skrining risiko Diabetes Mellitus.

TABEL VII. SEPULUH FITUR PALING BERPENGARUH BERDASARKAN FEATURE IMPORTANCE.

Peringkat	Fitur	Importance
1	HighBP	0,3651
2	GenHlth	0,2658
3	BMI	0,1212
4	Age	0,1024
5	HighChol	0,0617
6	DiffWalk	0,0161
7	Income	0,0157
8	HeartDiseaseorAttack	0,0141
9	HvyAlcoholConsump	0,0115
10	CholCheck	0,0084

G. Evaluasi Kinerja Model

McNemar Test (dapat dilihat pada Tabel VIII) digunakan untuk membandingkan Random Forest dan Gradient Boosting pada data uji yang sama. Pengujian ini penting karena selisih akurasi kedua model relatif kecil, yaitu 75,64% berbanding 75,07%, sehingga perlu dibuktikan apakah perbedaannya signifikan secara statistik.

Hasil uji McNemar menunjukkan nilai statistik 15,347 dengan p-value 0,00009. Karena p-value < 0,05, maka terdapat perbedaan performa yang signifikan antara Random Forest dan Gradient Boosting pada data uji yang sama.

TABEL VIII. HASIL UJI STATISTIK MCNEMAR.

Komponen	**Nilai**	**Interpretasi**
GB benar, RF salah	416	Keunggulan GB pada sampel tertentu
GB salah, RF benar	538	Keunggulan RF pada sampel tertentu
Statistik uji	15,347	Nilai statistik McNemar
p-value	0,00009	Signifikan jika < 0,05

H. Evaluasi Kinerja Model

Pengujian black box seperti yang dapat dilihat pada tabel IX, dilakukan untuk memastikan fitur aplikasi berjalan sesuai kebutuhan. Pengujian meliputi halaman utama, perhitungan BMI otomatis, pengiriman data ke endpoint prediksi, tampilan hasil skrining, pembacaan metrik model, konsultasi AI, penyimpanan riwayat skrining, dan penanganan kesalahan. Hasil pengujian menunjukkan bahwa fitur utama aplikasi berjalan sesuai rancangan fungsional.

TABEL IX. HASIL BLACK BOX TESTING APLIKASI.

No	**Fitur**	**Skenario Uji**	**Hasil**
1	Halaman utama	Membuka endpoint / dan /diabetes	Berhasil
2	BMI otomatis	Mengubah tinggi dan berat badan	Berhasil
3	Prediksi risiko	Mengirim data ke /api/predict	Berhasil
4	Analisis model	Membuka /api/metrics dan diagram metrik	Berhasil
5	Konsultasi AI	Mengirim pertanyaan ke /api/chat	Berhasil jika API key aktif
6	Riwayat pasien	Menyimpan hasil skrining selama sesi	Berhasil
7	Validasi error	Model/API key tidak tersedia	Pesan error tampil

I. Evaluasi Kinerja Model

Hasil penelitian dibandingkan dengan beberapa penelitian sebelumnya untuk memperjelas posisi kontribusi penelitian. Perbandingan dilakukan berdasarkan dataset, algoritma, metrik, dan kontribusi implementasi.

TABEL X. PERBANDINGAN PENELITIAN INI DENGAN PENELITIAN SEBELUMNYA.

Penelitian	Dataset	Algoritma	Metrik Utama	Perbedaan dengan Penelitian Ini
Kurniawan et al. [2]	Diabetes	RF, XGBoost, LightGBM	AUC	Fokus evaluasi model, tidak menekankan aplikasi web
Syaban [3]	Faktor risiko diabetes	Ensemble learning	Akurasi/ AUC	Mengevaluasi ensemble, namun tidak sama dengan implementasi sistem ini
Damayanti dan Baita [4]	Gestational diabetes	SVM, RF	Akurasi/ AUC	Berfokus pada diabetes gestasional
Ridwan dan Setyawan [5]	Diabetes	Berbagai ML	Akurasi	Membandingkan beberapa model tanpa fitur aplikasi lengkap
Penelitian ini	CDC Diabetes Health	RF, GB, LR	Accuracy, F1, ROC-AUC,	Menambahkan aplikasi web, AI edukatif,

	Indicators (balanced subset)		PR-AUC, MCC, Kappa	feature importance, dan pengujian fungsional; RF menjadi model terbaik

Hasil evaluasi menunjukkan bahwa Random Forest menjadi model terbaik pada dataset yang digunakan, dengan keunggulan terutama pada recall, F1-score, ROC-AUC, PR-AUC, MCC, dan Cohen Kappa. Hal ini menunjukkan bahwa pendekatan bagging pada Random Forest lebih stabil dalam mengenali pola risiko diabetes pada balanced subset CDC dibandingkan dua model lainnya. Selain itu, jumlah false negative yang lebih rendah pada Random Forest memperkuat alasan pemilihannya sebagai model yang lebih aman untuk skrining awal.

Analisis feature importance juga penting karena aplikasi kesehatan tidak hanya membutuhkan hasil prediksi, tetapi juga alasan faktor risiko yang memengaruhi prediksi. Hasil penelitian menunjukkan bahwa HighBP, GenHlth, BMI, Age, dan HighChol merupakan fitur paling dominan. Dengan demikian, rekomendasi kesehatan dapat diarahkan pada edukasi pengendalian tekanan darah, perbaikan kondisi kesehatan umum, pengelolaan berat badan, pemantauan usia risiko, dan pemeriksaan kolesterol secara berkala.

Dari sisi aplikasi, sistem web mobile-responsive telah sesuai dengan tujuan sebagai alat bantu skrining awal. Fitur konsultasi AI bersifat edukatif dan tidak menggantikan diagnosis dokter. Keterbatasan utama penelitian adalah penggunaan dataset CDC dari populasi Amerika Serikat dan penggunaan balanced subset, sehingga generalisasi ke populasi Indonesia masih perlu divalidasi melalui data lokal, pengujian klinis, serta pengujian usability seperti SUS atau UEQ.

IV. KESIMPULAN

Penelitian ini berhasil mengembangkan aplikasi skrining risiko Diabetes Mellitus berbasis web mobile-responsive dengan mengimplementasikan Random Forest, Logistic Regression, dan Gradient Boosting. Berdasarkan hasil evaluasi revisi, Random Forest memperoleh performa terbaik dengan akurasi 75,64%, presisi 73,21%, recall 80,88%, F1-score 76,85%, ROC-AUC 83,51%, PR-AUC 81,48%, MCC 0,5157, dan Cohen Kappa 0,5128. Evaluasi yang diperluas meliputi analisis distribusi kelas, confusion matrix, ROC-AUC, PR-AUC, MCC, Cohen Kappa, feature importance, dan McNemar Test sehingga kontribusi ilmiah penelitian menjadi lebih kuat.

Aplikasi yang dikembangkan mampu menerima input data kesehatan pengguna, menghitung BMI, memproses prediksi risiko, menampilkan kategori risiko, memberikan rekomendasi kesehatan, menyediakan visualisasi metrik, konsultasi AI, dan riwayat skrining. Sistem ini dapat digunakan sebagai alat bantu skrining awal dan edukasi kesehatan, tetapi bukan pengganti diagnosis medis. Walaupun prototipe awal aplikasi menggunakan model Gradient Boosting, hasil evaluasi revisi merekomendasikan Random Forest sebagai model yang lebih unggul untuk implementasi lanjutan. Pengembangan selanjutnya disarankan menggunakan dataset lokal Indonesia, validasi klinis, penyimpanan basis data yang aman, dan pengujian usability seperti System Usability Scale (SUS) atau User Experience Questionnaire (UEQ).

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Informatika Universitas Prisma Manado, dosen pembimbing, keluarga, dan seluruh pihak yang telah memberikan dukungan, arahan, dan masukan selama proses penyusunan penelitian ini.

DAFTAR PUSTAKA

[1] S. A. Azizah and I. Novrianti, "Pharmacotherapy of Diabetic Mellitus: A Review," Journal of Pharmacy and Science, vol. 5, no. 2, pp. 80-91, 2022.

[2] H. Kurniawan, A. D. Akbar, N. Svensons, Y. J. Antonio, S. Karnila, and E. Safitri, "Evaluasi Performa Random Forest, XGBoost, dan LightGBM dalam Diagnosis Dini Diabetes Mellitus," 2025.

[3] K. Syaban, "Evaluasi Model Ensemble Learning pada Identifikasi Faktor Risiko Diabetes Mellitus," vol. 15, no. 2, 2025, doi: 10.34010/jati.v15i2.16238.

[4] A. Damayanti and A. Baita, "Comparison of SVM and Random Forest Algorithm Performance with Random Undersampling Technique to Predict Gestational Diabetes Mellitus Risk," Journal of Applied Informatics and Computing, 2025.

[5] A. M. Ridwan and G. D. Setyawan, "Perbandingan Berbagai Model Machine Learning untuk Mendeteksi Diabetes," TEKNOKOM, vol. 6, no. 2, pp. 127-132, 2023, doi: 10.31943/teknokom.v6i2.152.

[6] N. Kumar and A. Makkar, Machine Learning in Cognitive IoT, 2020.

[7] K. Dan Implementasi and Y. Heryadi Teguh Wahyono, "Machine Learning," 2020.

[8] A. Fahri and Y. Ramdhani, "Visualisasi Data dan Penerapan Machine Learning Menggunakan Decision Tree Untuk Keputusan Layanan Kesehatan COVID-19," vol. 17, 2023.

[9] R. Ridwan, H. H. Handayani, S. A. P. Lestari, and Y. Cahyana, "Evaluasi Kinerja Algoritma Random Forest dan Gradient Boosting untuk Klasifikasi Penyakit Jantung," KOMTIKA, vol. 9, no. 1, pp. 112-124, 2025, doi: 10.31603/komtika.v9i1.13450.

[10] F. A. Rangkuti, Khairunnisa, and S. Sundari, "Implementasi Gradient Boosting Machines untuk Prediksi Harga Rumah," vol. 4, no. 2, pp. 164-172, 2025, doi: 10.69916/jkbti.v4i2.318.

[11] H. A. Abrori and E. R. Subhiyakto, "Analisis Komparatif Akurasi Prediksi Kanker Payudara Menggunakan Random Forest dan Logistic Regression," vol. 22, no. 1, pp. 300-311, 2025, doi: 10.33364/algoritma/v.22-1.2164.

[12] A. Ermillian and K. Nugroho, "Perancangan Model Deteksi Potensi Siswa Putus Sekolah Menggunakan Logistic Regression dan Decision Tree," vol. 9, no. 3, pp. 281-295, 2024, doi: 10.30591/jpit.v9i3.8007.

[13] A. Aulia, A. Andriyansyah, S. Suharjito, and S. W. Nensi, "Analisis Prediksi Stroke dengan Membandingkan Decision Tree, Naive Bayes, dan Random Forest," JIKI, vol. 3, no. 2, pp. 89-98, 2024, doi: 10.54082/jiki.90.

[14] A. R. Dana, R. V. Kristananda, M. B. S. Wibowo, and D. A. Prasetya, "Perbandingan Decision Tree dan Random Forest dengan Hyperparameter Tuning dalam Mendeteksi Penyakit Stroke," SANTIKA, vol. 4, 2024.