PERBANDINGAN ALGORITMA RANDOM FOREST DAN LOGISTIC REGRESSION DALAM PREDIKSI PENYAKIT DIABETES

Zidan Januri Zumantara[1)], Budi Sudrajat[2)], Hasta Herlan Asymar[3)]
[1,2]Prodi Informatika, Fakultas Teknik & Informatika, Universitas Bina Sarana Informatika
[3]Prodi Teknologi Komputer, Fakultas Teknik & Informatika, Universitas Bina Sarana Informatika

Correspondence author: Z.J.Zumantara, zidan15januri@gmail.com, Jakarta, Indonesia

Abstract

This study aims to compare the performance of the Random Forest and Logistic Regression machine learning algorithms in predicting diabetes using the Pima Indians Diabetes dataset from Kaggle. The dataset contains data on 768 adult female patients with eight health indicators and a target outcome variable indicating diabetes status. This quantitative study uses a comparative approach. The research stages include initial data analysis, preprocessing (zero-value cleaning), splitting the data into 70% training and 30% test, model development, evaluation using accuracy, precision, recall, and F1-score metrics, and feature analysis. The results show that Random Forest achieved 75% accuracy and Logistic Regression 74%. Random Forest also slightly outperformed Logistic Regression in precision, recall, and F1-score. This study differs from previous research in that it not only focused on evaluation metrics but also analyzed the most influential features. The analysis results show that Glucose is the most dominant indicator in Random Forest, while DiabetesPedigreeFunction is the most influential in Logistic Regression. These findings provide additional insight into the key risk factors in diabetes prediction.

Keywords: random forest, logistic regression, machine learning, diabetes prediction

Abstrak

Penelitian ini bertujuan membandingkan performa algoritma machine learning Random Forest dan Logistic Regression dalam memprediksi penyakit diabetes dengan menggunakan dataset Pima Indians Diabetes dari Kaggle. Dataset berisi 768 data pasien wanita dewasa dengan delapan indikator kesehatan serta variabel target Outcome yang menunjukkan status diabetes. Penelitian ini merupakan penelitian kuantitatif dengan pendekatan komparatif, tahapan penelitian meliputi analisis data awal, pra-pemrosesan berupa pembersihan nilai nol, pembagian data menjadi 70% training dan 30% testing, pembangunan model, evaluasi menggunakan metrik accuracy, precision, recall, dan F1-score, serta analisis fitur penting. Hasil menunjukkan bahwa Random Forest memperoleh akurasi 75% dan Logistic Regression 74%. Random Forest juga sedikit lebih unggul pada precision, recall, dan F1-score dibanding Logistic Regression. Perbedaan penelitian ini dengan penelitian terdahulu adalah tidak hanya berfokus pada metrik evaluasi, tetapi juga menambahkan analisis fitur paling berpengaruh. Hasil analisis menunjukkan bahwa Glucose merupakan indikator paling dominan pada Random Forest, sedangkan DiabetesPedigreeFunction paling berpengaruh pada Logistic Regression. Temuan

JURNAL ELEKTRO & INFORMATIKA
SWADHARMA (JEIS)
VOLUME 06 NOMOR 01 JANUARI 2026
ISSN 2774 – 5775 | eISSN 2774-5767

ini memberikan pemahaman tambahan mengenai faktor risiko utama dalam prediksi penyakit diabetes.

Kata Kunci: *random forest, logistic regression, machine learning*, prediksi diabetes

A. PENDAHULUAN

Penyakit diabetes merupakan salah satu penyakit tidak menular dengan angka kematian tinggi di dunia dan termasuk sepuluh penyebab kematian utama. Prevalensi diabetes terus meningkat setiap tahun, baik di tingkat global maupun nasional (Yuniarti et al., 2025). Data International Diabetes Federation (2024) menunjukkan bahwa Indonesia menjadi salah satu negara dengan jumlah penderita diabetes yang signifikan di kawasan Pasifik Barat. Pada tahun 2024 terdapat sekitar 20,4 juta penderita diabetes di Indonesia, dan jumlah ini diproyeksikan meningkat pada tahun 2050 (Rachmawati et al., 2025). Kondisi ini menunjukkan pentingnya deteksi dini diabetes untuk mencegah komplikasi lebih lanjut.

Penyakit diabetes adalah penyakit yang ditandai dengan kadar gula darah yang tinggi dan merupakan salah satu Penyakit Tidak Menular (PTM) yang berisiko bagi kesehatan seseorang (Oktaviana et al., 2024). kemudian gangguan metabolik yang ditandai dengan peningkatan kadar glukosa darah (hiperglikemia) akibat kerusakan pada sekresi insulin maupun gangguan kerja insulin (Lalla & Rumatiga, 2022). Hal ini disebabkan oleh gangguan yang muncul pada metabolisme yang tidak menghasilkan insulin dari pankreas. Sehingga mengakibatkan tubuh manusia tidak mampu memanfaatkan insulin dengan baik atau ketika pankreas tidak menghasilkan insulin dalam jumlah yang memadai (Fasnuari et al., 2022).

Berdasarkan paparan diatas, kemajuan teknologi informasi, khususnya dalam bidang kecerdasan buatan dan pengolahan data, memungkinkan pemanfaatan algoritma machine learning untuk memprediksi penyakit berdasarkan indikator kesehatan. Model klasifikasi seperti Random Forest dan Logistic Regression telah banyak digunakan untuk memprediksi penyakit termasuk diabetes. Namun, masih diperlukan kajian untuk mengetahui algoritma yang memiliki kinerja terbaik pada dataset tertentu.

Sejumlah penelitian sebelumnya telah membahas penerapan algoritma machine learning, khususnya Random Forest dan Logistic Regression, untuk memprediksi penyakit diabetes. Penelitian yang dilakukan oleh (Setyawan & Wakhidah, 2025) membandingkan metode Logistic Regression, Random Forest, dan Gradient Boosting. Hasil penelitian tersebut menunjukkan bahwa Logistic Regression memiliki akurasi 76%, Random Forest 77%, dan Gradient Boosting 75%. Perbedaan akurasi ketiga model relatif kecil, namun Random Forest memiliki kinerja sedikit lebih baik dibandingkan Logistic Regression dan Gradient Boosting.

Penelitian (Siridion & Siregar, 2024), melakukan analisis klasifikasi diagnosa penyakit diabetes melitus berdasarkan perbandingan berbagai algoritma supervised learning, seperti Logistic Regression, Random Forest, Artificial Neural Network (ANN), Decision Tree C4.5, dan Gradient Boosting. Hasilnya menunjukkan bahwa Random Forest memiliki tingkat akurasi paling tinggi yaitu 98,7%, diikuti ANN 94,5%, Decision Tree 93,5%, Logistic Regression 93,5%, dan Gradient Boosting 98,1%. Temuan ini memperlihatkan kemampuan Random Forest yang sangat baik dalam membedakan kelas positif dan negatif pada data diabetes.

Penelitian (Syahri et al., 2024), juga membandingkan Logistic Regression, Random Forest, dan Adaboost dalam klasifikasi diabetes mellitus. Hasil penelitian menunjukkan Logistic Regression memiliki akurasi 72%, Random Forest 79%, dan Adaboost 78%. Perbedaan akurasi antara

Random Forest dan Adaboost relatif kecil (1%), sedangkan Logistic Regression memiliki akurasi lebih rendah dengan selisih sekitar 6–7% dibandingkan dua model lainnya. Berdasarkan ketiga penelitian tersebut, Random Forest secara konsisten menunjukkan kinerja lebih baik dibandingkan Logistic Regression, meskipun selisih akurasi pada beberapa penelitian relatif kecil.

Berbeda dengan penelitian-penelitian terdahulu yang umumnya hanya berfokus pada perbandingan kinerja model berdasarkan metrik evaluasi seperti akurasi, precision, recall, dan F1-score, penelitian ini tidak hanya membandingkan performa algoritma Random Forest dan Logistic Regression, tetapi juga mencari tahu fitur-fitur kesehatan mana yang paling berpengaruh dalam prediksi penyakit diabetes. Penelitian ini bertujuan untuk membandingkan dua algoritma machine learning, yaitu Random Forest dan Logistic Regression, dalam memprediksi penyakit diabetes berdasarkan indikator kesehatan selain itu penlitian ini dilakukan untuk mengetahui pengaruh utama feature importance pada Random Forest dan koefisien regresi pada Logistic Regression, sehingga memberikan pemahaman lebih mendalam mengenai variabel-variabel yang dominan dalam mempengaruhi hasil prediksi. Dengan demikian, penelitian ini tidak hanya mengevaluasi tingkat kinerja model tetapi juga memperkaya wawasan tentang faktor-faktor risiko diabetes yang penting untuk diperhatikan.

B. METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan komparatif yang bertujuan membandingkan performa algoritma Random Forest dan Logistic Regression dalam memprediksi penyakit diabetes berdasarkan indikator kesehatan. Data yang digunakan berasal dari sumber sekunder yaitu dataset diabetes yang diperoleh dari situs Kaggle (www.kaggle.com).

Pengumpulan Data

Penelitian ini menggunakan dataset sekunder “Pima Indians Diabetes” yang diperoleh dari situs Kaggle. Dataset ini berisi 768 catatan pasien wanita dewasa dengan delapan indikator kesehatan yaitu Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, dan Age, serta satu variabel target Outcome yang menunjukkan status diabetes pasien (0 = tidak menderita diabetes, 1 = menderita diabetes). Dataset ini dipilih karena memiliki variasi indikator kesehatan yang relevan dan umum digunakan sebagai benchmark dalam penelitian prediksi diabetes.

Analisis Data Awal

Sebelum pemodelan dilakukan, peneliti melakukan analisis eksploratif terhadap dataset untuk memahami karakteristiknya. Analisis ini mencakup pemeriksaan struktur dataset yang terdiri atas 768 baris dan 9 kolom, menghitung distribusi kelas target yang terdiri dari 500 sampel non-diabetes (kelas 0) dan 268 sampel diabetes (kelas 1), serta mengidentifikasi nilai minimum, maksimum, rata-rata dan sebaran pada masing-masing fitur untuk mendeteksi data yang tidak realistis atau ekstrem. Hasil analisis awal ini menjadi dasar untuk tahap pembersihan data.

Pre-processing Data

Tahap pra-pemrosesan dilakukan agar data siap digunakan dalam proses pemodelan. Pembersihan data dilakukan dengan menangani nilai nol (0) yang tidak realistis pada fitur Glucose, BloodPressure, SkinThickness, dan Insulin yang dianggap sebagai missing value. Nilai-nilai nol ini diganti dengan nilai median masing-masing kolom untuk menjaga distribusi data. Setelah data dibersihkan, dataset dibagi menjadi data training sebesar 70% (537 sampel) untuk melatih model dan data testing sebesar 30% (231 sampel) untuk menguji performa model terhadap data baru yang belum pernah dilihat sebelumnya.

JURNAL ELEKTRO & INFORMATIKA
SWADHARMA (JEIS)
VOLUME 06 NOMOR 01 JANUARI 2026
ISSN 2774 – 5775 | eISSN 2774-5767

Pembangunan Model

Model dibangun menggunakan dua algoritma machine learning, yaitu Random Forest dan Logistic Regression. Random Forest merupakan metode ensemble berbasis pohon keputusan yang membangun banyak pohon dan menggabungkan hasilnya untuk prediksi akhir, sedangkan Logistic Regression merupakan metode regresi linier yang mengestimasi probabilitas kelas biner. Kedua model diterapkan menggunakan bahasa pemrograman Python pada platform Google Colab karena mendukung pemrosesan berbasis cloud yang stabil dan memiliki integrasi pustaka scikit-learn yang lengkap.

Evaluasi Model

Kinerja kedua algoritma dievaluasi menggunakan metrik accuracy untuk tingkat prediksi benar keseluruhan, precision untuk mengukur ketepatan prediksi positif, recall untuk mengukur kemampuan mendeteksi kasus positif sebenarnya, dan F1-score yang merupakan rata-rata harmonis precision dan recall. Selain itu digunakan confusion matrix dan classification report untuk menggambarkan jumlah prediksi benar dan salah pada masing-masing kelas secara lebih rinci sehingga interpretasi hasil menjadi lebih komprehensif.

Analisis Fitur

Sebagai pembeda dengan penelitian terdahulu yang hanya berfokus pada metrik evaluasi model, penelitian ini juga menganalisis indikator kesehatan yang paling berpengaruh dalam prediksi penyakit diabetes. Pada algoritma Random Forest digunakan penghitungan feature importance untuk menentukan fitur dengan bobot tertinggi, sedangkan pada Logistic Regression dianalisis nilai koefisien regresi untuk melihat arah dan besarnya pengaruh tiap fitur. Analisis ini memberikan pemahaman lebih mendalam mengenai faktor risiko dominan yang mempengaruhi hasil prediksi masing- masing algoritma.

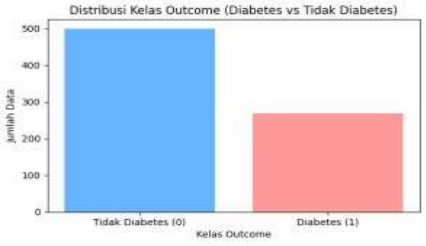
C. HASIL DAN PEMBAHASAN

Pemahaman data ini untuk memahami karakteristik dataset sebelum proses pemodelan. Tahap ini meliputi pemeriksaan struktur dataset untuk mengetahui jumlah baris dan kolom, serta mengevaluasi distribusi kelas target Outcome, yaitu jumlah data pasien yang terindikasi diabetes (kelas 1) dan yang tidak terindikasi diabetes (kelas 0). Dengan analisis ini, peneliti dapat melihat keseimbangan data dan kondisi awal variabel yang akan digunakan pada tahap pra-pemrosesan dan pemodelan

Tabel 1. Atribut Dataset

Column	Keterangan
Pregnancies	Untuk mengungkapkan kehamilan
Glucose	Untuk mengekspresikan tingkat glukosa dalam darah
BloodPresure	Untuk mengekspresikan pengukuran tekanan darah
SkinThickness	Untuk mengekspresikan ketebalan kulit
Insulin	Untuk mengekspresikan tingkat Insulin dalam darah
BMI	Untuk mengekspresikan indeks massa tubuh
Diabetes Pedigree	Untuk mengungkapkan persentase Diabetes
Age	Untuk mengekspresikan usia
Outcome	Untuk menyatakan hasil akhir 1 adalah Ya dan 0 adalah Tidak

Pada gambar 1. terlihat bahwa data pasien yang tidak menderita diabetes (kelas 0) berjumlah sekitar 500 sampel, sedangkan data pasien yang menderita diabetes (kelas 1) berjumlah sekitar 268 sampel. Ini berarti dataset sedikit tidak seimbang, dengan jumlah kasus non-diabetes lebih banyak dibanding kasus diabetes.



Gambar 1. Jumlah Data Diabetes dan tidak Diabetes

Pada gambar 2. Grafik pie chart menunjukkan proporsi kelas pada dataset utama. Sekitar dua pertiga data (65,1%) merupakan pasien tidak diabetes, sedangkan sepertiganya (34,9%) merupakan pasien diabetes. Ini menggambarkan distribusi data yang tidak seimbang antara kedua kelas.

Gambar 2. Prosentase Data Diabetes dan tidak Diabetes

Pembersihan Data

Pemeriksaan awal terhadap kelengkapan data dilakukan untuk mengetahui ada atau tidaknya nilai yang hilang pada setiap variabel. Berdasarkan hasil pengecekan, semua kolom pada dataset memiliki nilai nol untuk jumlah data yang hilang. Hal ini menunjukkan bahwa dataset yang digunakan sudah lengkap dan tidak memiliki missing value, sehingga dapat langsung diproses pada tahap pra-pemrosesan dan pemodelan.

Pada Gambar 3. menunjukkan jumlah nilai yang hilang pada setiap kolom dataset. Angka 0 di semua kolom berarti tidak ada nilai yang hilang (missing value) pada dataset, sehingga data lengkap dan siap diproses.

```
Jumlah Nilai yang Hilang per Kolom:
Pregnancies                 0
Glucose                     0
BloodPressure               0
SkinThickness               0
Insulin                     0
BMI                         0
DiabetesPedigreeFunction    0
Age                         0
Outcome                     0
dtype: int64
```

Gambar 3. Nilai yang Hilang pada Data

Pembagian Data

Pada Tabel 2. menunjukkan pembagian dataset menjadi data training dan data testing dengan proporsi 70:30. Dari total 768 data, 537 data digunakan untuk melatih model (training set) dan 231 data digunakan untuk menguji model (testing set). Pembagian ini dilakukan agar performa model dapat dievaluasi secara objektif pada data yang belum pernah dilihat sebelumnya.

Tabel 2. Pembagian Data Training dan Testing

Positif	Negatif
Training 70 %	
187	350
Testing 30%	
81	150

Hasil Random Forest

Pada Tabel 3. menunjukkan hasil evaluasi model Random Forest pada data pengujian. Model memperoleh akurasi sebesar 75% dengan nilai precision, recall, dan F1-score yang lebih tinggi pada kelas tidak diabetes (0) dibandingkan kelas diabetes (1). Nilai rata-rata (macro dan weighted) berada di kisaran 0,73–0,76, yang menunjukkan performa model cukup baik meskipun terdapat perbedaan kinerja antar kelas.

Tabel 3. Hasil Random Forest

Kelas / Metrik	**Precision**	**Recall**	**F1-score**	**Support**
0 (Tidak Diabetes)	0,82	0,80	0,81	151
1 (Diabetes)	0,64	0,66	0,65	80
Akurasi	–	–	0,75	231
Macro Avg	0,73	0,73	0,73	231
Weighted Avg	0,76	0,75	0,75	231

Hasil Logistic Regression

Pada Tabel 4. menunjukkan hasil evaluasi model Logistic Regression pada data pengujian. Model ini memperoleh akurasi sebesar 74% dengan nilai precision, recall, dan F1-score yang lebih tinggi pada kelas tidak diabetes (0) dibandingkan kelas diabetes (1). Nilai rata-rata (macro dan weighted) berada di kisaran 0,71–0,74, yang menunjukkan kinerja model cukup baik meskipun performanya pada kelas positif masih lebih rendah.

JURNAL ELEKTRO & INFORMATIKA
SWADHARMA (JEIS)
VOLUME 06 NOMOR 01 JANUARI 2026
ISSN 2774 – 5775 | eISSN 2774-5767

Tabel 4. Hasil Logistic Regression

Kelas / Metrik	Precision	Recall	F1-score	Support
0 (Tidak Diabetes)	0,80	0,79	0,80	151
1 (Diabetes)	0,62	0,62	0,62	80
Akurasi	–	–	0,74	231
Macro Avg	0,71	0,71	0,71	231
Weighted Avg	0,74	0,74	0,74	231

Hasil Perbandingan Akurasi dan Evaluasi

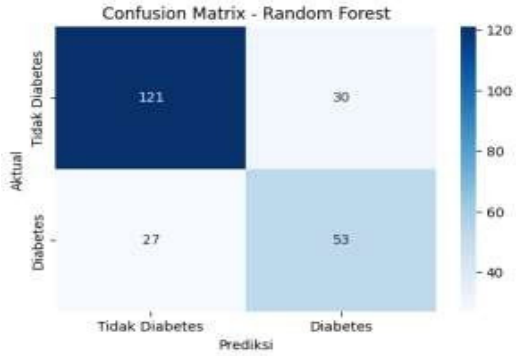
Pada Tabel 5. Menunjukan hasil perbandingkan kinerja model Random Forest (RF) dan Logistic Regression (LR) berdasarkan metrik evaluasi. Secara umum, RF menunjukkan nilai precision, recall, F1-score, dan akurasi yang sedikit lebih tinggi dibandingkan LR, dengan selisih 0,01–0,04. Kinerja pada kelas 0 tergolong baik, sedangkan pada kelas 1 tergolong cukup baik, sehingga dapat disimpulkan bahwa Random Forest memberikan hasil yang lebih konsisten dan sedikit lebih unggul dibanding Logistic Regression pada dataset ini.

Tabel 5. Hasil Perbandingan Random Forest dan Logistic Regression

Metrik	Kelas	RF	LR	Selisih (RF- LR)	Kategori
Precision	0	0.82	0.80	+0.02	Baik
Recall	0	0.80	0.79	+0.01	Baik
F1-Score	0	0.81	0.80	+0.01	Baik
Precisi on	1	0.64	0.62	+0.02	Cukup Baik
Recall	1	0.66	0.62	+0.04	Cukup Baik
F1-Score	1	0.65	0.62	+0.03	Cukup Baik
Accuracy	-	0.75	0.74	+0.01	Baik
Macro Avg	-	0.73	0.71	+0.02	Cukup Baik
Weight ed Avg	-	0.75	0.74	+0.01	Baik

Hasil Confusion Matrix Random Forest

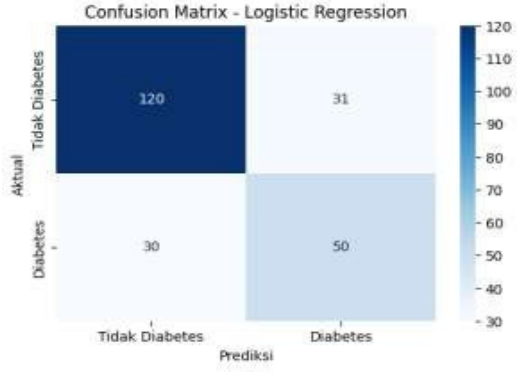
Pada Gambar 4. Berdasarkan confusion matrix pada model Random Forest, diketahui bahwa sebanyak 121 data pasien tidak diabetes berhasil diprediksi benar sebagai tidak diabetes, sedangkan 30 data pasien tidak diabetes diprediksi salah sebagai diabetes. Pada kelas diabetes, sebanyak 53 data pasien berhasil diprediksi benar sebagai diabetes dan 27 data pasien diprediksi salah sebagai tidak diabetes. Hasil ini menunjukkan bahwa model Random Forest lebih baik dalam mengenali kelas tidak diabetes dibandingkan kelas diabetes.



Gambar 4. Confusion Matrix RF

Hasil Confusion Matrix Logistic Regression

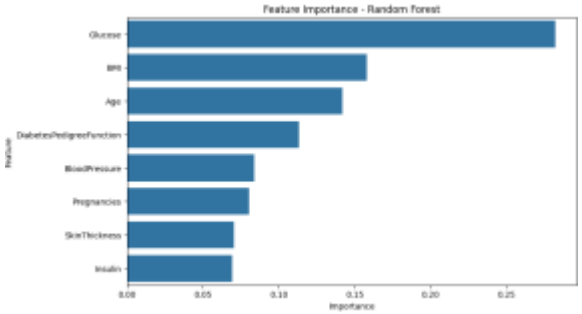
Pada gambar 5. Confusion matrix tersebut menunjukkan kinerja model Logistic Regression. Sebanyak 120 data tidak diabetes diprediksi benar sebagai tidak diabetes dan 31 data tidak diabetes diprediksi salah sebagai diabetes. Pada kelas diabetes, 50 data diprediksi benar sebagai diabetes dan 30 data diprediksi salah sebagai tidak diabetes. Model ini juga lebih baik mengenali kelas tidak diabetes dibandingkan kelas diabetes.



Gambar 5. Confusion Matrix LR

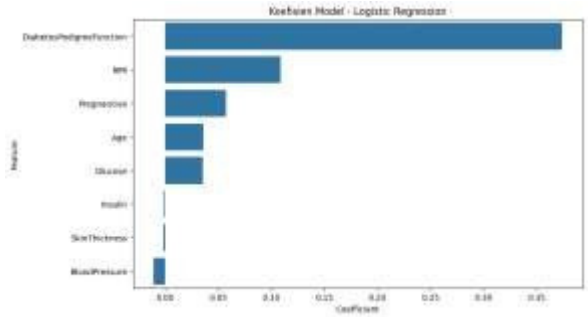
Fitur Paling Berpengaruh dalam Prediksi

Pada gambar 6. Grafik tersebut menunjukkan tingkat kepentingan fitur pada model Random Forest. Fitur Glucose memiliki pengaruh paling besar dalam prediksi diabetes, diikuti oleh BMI, Age, dan DiabetesPedigreeFunction. Sementara itu fitur Insulin, SkinThickness, dan Pregnancies memiliki pengaruh yang lebih rendah dibandingkan fitur lainnya.



Gambar 6. Grafik Random Forest Fitur Paling Berpengaruh dalam Prediksi Diabetes

Pada Gambar 7. Grafik tersebut menunjukkan nilai koefisien pada model Logistic Regression. Fitur yang memiliki pengaruh paling besar terhadap prediksi diabetes adalah DiabetesPedigreeFunction, diikuti oleh BMI, Pregnancies, Age, dan Glucose. Sementara itu, fitur seperti Insulin, SkinThickness, dan BloodPressure memiliki kontribusi yang relatif kecil terhadap hasil prediksi.



Gambar 7. Grafik Logistic Regression Fitur Paling Berpengaruh dalam Prediksi Diabetes

Hasil analisis feature importance menunjukkan bahwa variabel Glucose merupakan indikator paling dominan dalam prediksi diabetes. Sementara itu, variabel Insulin tidak terlalu berpengaruh dalam model. Hal ini kemungkinan besar disebabkan oleh keterbatasan dataset Pima Indians Diabetes dari Kaggle, di mana banyak nilai Insulin yang tercatat nol dan dianggap tidak realistis. Oleh karena itu, meskipun secara medis kadar glukosa dan insulin sama- sama penting sebagai faktor penentu diabetes, dalam penelitian ini hanya Glucose yang tampil dominan dalam model Random Forest.

D. PENUTUP

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma Random Forest dan Logistic Regression sama-sama mampu digunakan dalam klasifikasi penyakit diabetes. Random Forest memiliki performa yang lebih baik dengan akurasi 75%, sedangkan Logistic Regression memperoleh akurasi 74%. Random Forest juga memberikan nilai precision, recall, dan F1-score yang lebih tinggi pada kelas non-diabetes, meskipun pada kelas diabetes keduanya masih memiliki keterbatasan. Analisis fitur menunjukkan bahwa Glucose merupakan indikator paling dominan pada Random Forest, sedangkan Diabetes PedigreeFunction paling berpengaruh pada Logistic Regression. Hasil ini memperlihatkan bahwa selain kinerja model, pemahaman terhadap fitur yang berpengaruh juga penting untuk mendukung upaya deteksi dini diabetes. Dengan demikian, penelitian ini menegaskan bahwa Random Forest lebih konsisten dalam menghasilkan prediksi, namun Logistic Regression tetap dapat digunakan sebagai alternatif dengan kinerja yang cukup baik.

E. DAFTAR PUSTAKA

Fasnuari, H. A. D., Yuana, H., & Chulkamdi, M. T. (2022). Application of K-Nearest Neighbor Algorithm For Classification of Diabetes Mellitus. Case Study : Residents of Jatitengah Village. *Antivirus : Jurnal Ilmiah Teknik Informatika*, *16*(2), 133–142.

https://doi.org/10.35457/antivirus.v16i2.2445

Lalla, N. N., & Rumatiga, J. (2022). Type instability of Blood Glucose Levels in Type II Diabetes Mellitus Patients. *Jurnal Ilmiah Kesehatan Sandi Husada*, *11*(2), 473–479. https://doi.org/10.35816/jiskh.v11i2.816

Oktaviana, A., Wijaya, D. P., Pramuntadi, A., & Heksaputra, D. (2024). Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Algoritma K-Nearest Neighbor (K-NN). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(3), 812–818. https://doi.org/10.57152/malcom.v4i3.1268

Rachmawati, F. S., Sau, G. V. O., Syah, M. F., Effendy, D. S., Muchtar, F., Bahar, H., Lestari, H., & Tosepu, R. (2025). Edukasi Melalui Infografis Dengan Gaya Hidup dan Pola Makan Dalam Mencegah Diabetes. *IJCD : Indonesian Journal of Community Dedication*, *3*(2), 318–327. https://doi.org/10.61214/ijcd.v3i2.783

Setyawan, N. H., & Wakhidah, N. (2025). Analisis Perbandingan Metode Logistic Regression, Random Forest, Gradient Boosting Untuk Prediksi Diabetes. *JIPI : Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika*, *10*(1), 150–162. https://doi.org/10.29100/jipi.v10i1.5743

Siridion, S. T., & Siregar, B. (2024). Analisis Klasifikasi Diagnosa Penyakit Diabetes Melitus Berdasarkan Komparasi Algoritma Supervised Learning. *Mutiara: Multidiciplinary Scientifict Journal*, *2*(3), 1006–1014. https://doi.org/10.57185/mutiara.v2i2.159

Syahri, A., Fariha, U., Afandi, R., & Nurliyana, I. (2024). Comparison of Logistic Regression, Random Forest and Adaboost Algorithms for Diabetes Mellitus Classification. *IJATIS: Indonesian Journal of Applied Technology and Innovation Science*, *1*(1), 41–46. https://doi.org/10.57152/ijatis.v1i1.1116

Yuniarti, T., Haryanto, B. A. H., Kalpikawati, A. B., Aryawati, R. N., Khasanah, N. H., Annisa, A. S., Rofiah, R., & Safitri, Y. (2025). Promosi Kesehatan dan Implementasi Pemberian Kapsul Habbatusauda Untuk Mencegah Diabetes Melitus. *Jurnal Pengabdian Komunitas*, *4*(1), 1–9. https://jurnalpengabdiankomunitas.com/index.php/pengabmas/article/view/234