Seminar Nasional Informatika (SENATIKA)
Prosiding Senatika 2025 **QRSBN:62-0151-02942-8**

Prediksi Risiko Diabetes Menggunakan Model Regresi Logistik dan Random Forest

Winxy Tri Setiadi[1], Deny Jollyta[2], Erwin Budi Setiawan[3]
[1,2] Institut Bisnis Dan Teknologi Pelita Indonesia, Pekanbaru, Indonesia
[3]Universitas Telkom, Bandung, Indonesia
e-mail: [1]**winxytrisetiadi2003@gmail.com**, [2]deny.jollyta@lecturer.pelitaindonesia.ac.id,
[3]erwinbudisetiawan@telkomuniversity.ac.id

Abstrak

Diabetes melitus merupakan krisis kesehatan publik yang terus meningkat, dengan sebagian besar kasus di Indonesia tidak terdiagnosis sehingga menuntut adanya metode deteksi dini yang efektif. Penelitian ini bertujuan untuk mengembangkan dan membandingkan kinerja dua model pembelajaran mesin—Regresi Logistik dan Random Forest—dalam memprediksi risiko diabetes. Metodologi penelitian mencakup pra-pemrosesan data, pelatihan model pada 70% data, dan pengujian pada 30% data. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-Score. Hasil penelitian menunjukkan bahwa model Random Forest secara signifikan mengungguli Regresi Logistik di semua metrik, mencapai akurasi 97.03% dan recall 68.94%. Analisis kepentingan fitur mengidentifikasi kadar glukosa darah, level HbA1c, dan BMI sebagai prediktor paling signifikan. Model yang unggul kemudian diimplementasikan ke dalam aplikasi web interaktif berbasis Streamlit sebagai alat bantu skrining. Simpulan dari studi ini menegaskan bahwa model Random Forest memiliki potensi besar sebagai alat skrining non-invasif untuk deteksi dini diabetes, yang dapat membantu memprioritaskan individu untuk pengujian klinis lebih lanjut.

Kata kunci: *Diabetes Melitus, Pembelajaran Mesin, Regresi Logistik, Random Forest, Deteksi Dini, Streamlit.*

Abstract

Diabetes mellitus is an escalating public health crisis, with a large portion of cases in Indonesia remaining undiagnosed, thus requiring effective early detection methods. This study aims to develop and compare the performance of two machine learning models—Logistic Regression and Random Forest—in predicting diabetes risk. The research methodology includes data preprocessing, model training on 70% of the data, and testing on the remaining 30%. Model performance was evaluated using accuracy, precision, recall, and F1-Score metrics. The results show that the Random Forest model significantly outperformed Logistic Regression across all metrics, achieving an accuracy of 97.03% and a recall of 68.94%. Feature importance analysis identified blood glucose level, HbA1c level, and BMI as the most significant predictors. The superior model was then implemented into an interactive web application using Streamlit as a screening support tool. This study concludes that the Random Forest model holds substantial potential as a non-invasive screening tool for the early detection of diabetes, which can help prioritize individuals for further clinical testing.

Keywords: *Diabetes Mellitus, Machine Learning, Logistic Regression, Random Forest, Early Detection, Streamlit.*

---

1. Pendahuluan

Diabetes melitus merupakan salah satu krisis kesehatan global yang paling signifikan, dengan prevalensi yang terus meningkat setiap tahunnya. Indonesia menempati posisi tinggi dalam jumlah penderita diabetes di dunia, dan lebih dari 70% kasusnya belum terdiagnosis. Fenomena “gunung es” ini menyebabkan banyak individu tidak menyadari kondisi mereka

hingga timbul komplikasi serius seperti penyakit jantung, gagal ginjal, dan kebutaan. Faktor genetik dan gaya hidup, seperti pola makan, obesitas, tekanan darah tinggi, dan aktivitas fisik yang rendah, menjadi penyebab utama yang memengaruhi kadar gula darah dan risiko diabetes. Oleh karena itu, deteksi dini menjadi sangat penting untuk pencegahan serta pengelolaan penyakit secara efektif.[1], [2]

Gambar 1. Tipe Diabetes Melitus

Perkembangan ilmu komputer, khususnya dalam bidang machine learning, telah membuka peluang besar dalam membantu proses deteksi dini berbagai penyakit, termasuk diabetes. Machine learning memungkinkan komputer mempelajari pola dari data pasien dan melakukan prediksi dengan tingkat akurasi tinggi.[3] Berbagai algoritma telah digunakan dalam penelitian sebelumnya untuk mengklasifikasikan risiko diabetes. Misalnya, Pramudyantoro et al. (2024) menggabungkan algoritma K-Nearest Neighbors (KNN) dan LightGBM pada dataset Pima Indians dan memperoleh akurasi 90,6%[4]. Pratama et al. (2023) menggunakan Regresi Logistik untuk memprediksi diabetes dengan hasil yang cukup baik[5], sementara Ridwan dan Setyawan (2023) membandingkan beberapa model machine learning dan menemukan bahwa algoritma Random Forest menghasilkan akurasi tertinggi, mencapai 99%[6]. Namun, sebagian besar penelitian tersebut berfokus pada satu model tertentu atau dataset berskala kecil, serta belum memberikan perbandingan sistematis antara model yang bersifat interpretable (seperti Regresi Logistik) dan model ensemble (seperti Random Forest) pada dataset besar dan beragam.

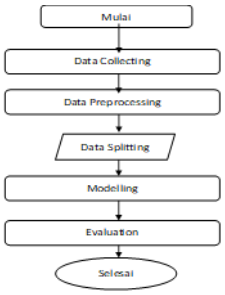
Penelitian ini berfokus pada upaya membandingkan kinerja dua algoritma machine learning, yaitu Regresi Logistik dan Random Forest, dalam memprediksi risiko diabetes berdasarkan berbagai atribut kesehatan pasien. Permasalahan utama yang diangkat dalam penelitian ini adalah bagaimana kinerja kedua model tersebut dalam mengklasifikasikan individu dengan risiko diabetes serta model mana yang memberikan hasil prediksi paling akurat dan dapat diandalkan untuk mendukung proses skrining dini penyakit ini.[7],[5],[8],[9]

Sejalan dengan perumusan masalah tersebut, tujuan penelitian ini adalah untuk mengembangkan dan mengevaluasi model Regresi Logistik dan Random Forest dalam konteks prediksi risiko diabetes, serta melakukan analisis perbandingan kinerja keduanya menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-Score. Selain itu, penelitian ini juga bertujuan mengimplementasikan model terbaik ke dalam aplikasi web interaktif berbasis Streamlit, sehingga dapat berfungsi sebagai alat bantu prediksi non-invasif yang mudah digunakan oleh masyarakat maupun tenaga kesehatan.[10],[11]

Kontribusi utama atau novelty dari penelitian ini terletak pada penyediaan perbandingan empiris yang komprehensif antara dua model pembelajaran mesin dengan pendekatan sistematis menggunakan dataset berukuran besar dan beragam. Selain itu, penelitian ini juga memberikan analisis feature importance untuk mengidentifikasi variabel paling berpengaruh dalam prediksi diabetes, seperti kadar glukosa darah, HbA1c, dan indeks massa tubuh. Hasil penelitian selanjutnya diimplementasikan dalam bentuk aplikasi berbasis web yang mengintegrasikan model terbaik, sehingga dapat digunakan secara praktis sebagai sarana deteksi dini yang efektif dan mudah diakses. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi ilmiah dan praktis dalam pengembangan sistem pendukung keputusan di bidang kesehatan, khususnya dalam membantu proses skrining awal serta pencegahan diabetes di Indonesia.

2. Metode Penelitian

Metodologi penelitian ini dirancang secara sistematis untuk memastikan pengembangan dan evaluasi model prediksi diabetes yang valid dan andal. Gambar 2 adalah penjelasan terstruktur mengenai langkah-langkah, data, algoritma, dan metode evaluasi yang digunakan.



Gambar 2. Tahapan Penelitian

2.1. Data Collecting

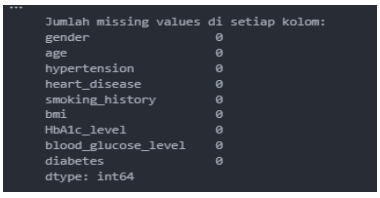
Tahap awal penelitian ini adalah pengumpulan data. Dataset yang digunakan adalah diabetes_dataset.csv, sebuah dataset publik yang berisi 100.000 catatan pasien yang diperoleh melalui situs kaggle. Setiap catatan memiliki 9 atribut yang relevan untuk prediksi diabetes, yaitu: gender, age, hypertension, heart_disease, smoking_history, bmi, HbA1c_level, blood_glucose_level, dan variabel target diabetes yang menunjukkan apakah seorang pasien menderita diabetes (1) atau tidak (0).

Tabel 1. Head Dataset

	Gender	Age	Hypertension	Heart Disease	Smoking History	BMI	HbA1c Level	Blood Glucose Level	Outcome
0	Female	80.0	0	1	never	25.19	6.6	140	0
1	Female	54.0	0	0	No Info	27.32	6.6	80	0
2	Male	28.0	0	0	never	27.32	5.7	158	0
3	Female	36.0	0	0	current	23.45	5.0	155	0
4	Male	76.0	1	1	current	20.14	4.8	155	0

2.2. Data Preprocessing

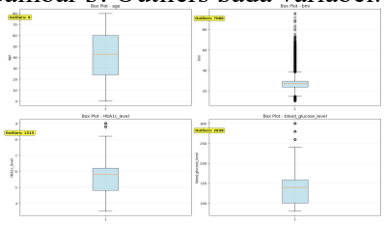
Setelah tahap data collecting, proses selanjutnya adalah data *exploration*. Data *exploration* adalah tahap yang bertujuan untuk memahami data [4] Pada proses eksplorasi ini, kumpulan dataset yang telah didapatkan melalui situs Kaggle, dilakukan *preprocessing* dengan melihat data duplikat dan memeriksa *missing value*. Pada 100,000 data ini terdapat 3,854 data duplikat serta tidak terdapat missing value secara eksplisit. Gambar 1 menunjukkan hasil pemeriksaan data yang *missing value* pada dataset.



Gambar 2. Data *Missing Value*

Preprocessing selanjutnya adalah pengecekan outlier. Outlier adalah nilai yang secara signifikan berbeda dari data lainnya dalam suatu kumpulan data. Nilai ini sering kali dianggap sebagai data yang tidak lazim atau abnormal karena berada jauh dari nilai-nilai lainnya. Berikut proses pengecekan pada data yang dimiliki:

Gambar 3. Outliers pada variabel.



Tabel 4. Data dengan Outlier yang Dihapus

	Gender	Age	Hypertension	Heart Disease	Smoking History	BMI	HbA1c Level	Blood Glucose Level	Diabetes
11	Female	54.0	0	0	former	54.70	6.0	100	0
39	Female	34.0	0	0	never	56.43	6.2	200	0
40	Male	73.0	0	0	former	25.91	9.0	160	1
55	Male	50.0	0	0	former	37.16	9.0	159	1
59	Female	67.0	0	0	never	63.48	8.8	155	1

Penghapusan outlier berhasil mengurangi standar deviasi pada semua variabel numerik, yang menunjukkan data menjadi lebih homogen dan representatif. Selanjutnya, data yang telah bersih ini siap untuk tahap pemodelan machine learning dengan terlebih dahulu melakukan encoding pada variabel kategorikal dan normalisasi data.

2.3 Data Splitting

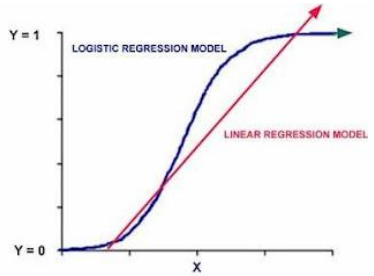
Dalam penelitian ini, setelah melalui tahap pra-pemrosesan, dataset yang berisi 100.000 sampel data pasien dipisahkan menjadi dua bagian utama: data latih (training data) dan data uji (testing data). Proses pemisahan ini menggunakan rasio 70:30. Dengan rasio tersebut, sebanyak 70% dari total data, dialokasikan sebagai data latih. Data inilah yang digunakan untuk melatih model Regresi Logistik dan Random Forest agar dapat mengenali pola-pola yang berkaitan dengan risiko diabetes. Sisa 30% dari dataset digunakan sebagai data uji. Data uji ini berfungsi sebagai data baru yang belum pernah dilihat oleh model sebelumnya, dan digunakan untuk mengevaluasi seberapa baik kinerja model dalam membuat prediksi. Semua variabel independen yang ada dalam dataset digunakan sebagai fitur prediktor untuk pemodelan ini.

2.4. Evaluation

Kinerja model dievaluasi secara kuantitatif menggunakan metrik standar yang berasal dari Matriks Kebingungan (*Confusion Matrix*). Metrik yang digunakan meliputi Akurasi untuk mengukur proporsi total prediksi yang benar, Presisi untuk mengukur proporsi prediksi positif yang akurat, Recall (Sensitivitas) untuk mengukur kemampuan model mengidentifikasi semua kasus positif yang sebenarnya, dan F1-Score yang menyeimbangkan antara presisi dan recall. Dalam konteks medis, recall menjadi metrik yang sangat penting untuk meminimalkan kasus positif yang terlewatkan.[10], [12]

2.5 Regresi Logistik

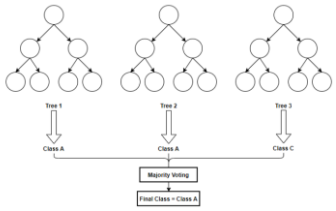
Regresi Logistik adalah algoritma klasifikasi yang digunakan untuk memprediksi probabilitas suatu data masuk ke dalam salah satu dari dua kategori (klasifikasi biner). Model ini bekerja dengan mengubah kombinasi linear dari fitur input menjadi nilai probabilitas antara 0 dan 1 menggunakan fungsi sigmoid. Sebuah nilai ambang batas, biasanya 0.5, kemudian digunakan pada probabilitas tersebut untuk menentukan prediksi kelas akhir, menjadikannya metode yang efisien dan mudah diinterpretasikan [7], [13], [14]. Gambar 4 menunjukkan konsep regresi logistik:



Gambar 4. Algoritma Regresi Logistik

2.6 Random Forest

Random Forest adalah algoritma machine learning yang bekerja dengan membangun banyak pohon keputusan (decision tree) secara independen untuk meningkatkan akurasi prediksi. Proses ini dimulai dengan teknik Bootstrap Aggregating, di mana beberapa sampel data acak diambil dari dataset awal dengan pengembalian untuk melatih setiap pohon. Selain itu, pada setiap percabangan pohon, algoritma hanya mempertimbangkan sebagian kecil fitur yang dipilih secara acak, yang dikenal sebagai metode Random Subspace. Setelah semua pohon terbentuk, prediksi akhir dihasilkan dengan menggabungkan hasil dari seluruh pohon melalui voting mayoritas untuk tugas klasifikasi atau dengan mengambil nilai rata-rata untuk tugas regresi sehingga menghasilkan model yang lebih stabil, akurat, dan tidak mudah overfitting dibandingkan pohon keputusan tunggal[6], [9], [15], [16]. Gambar 5 menunjukan cara kerja algoritmanya:

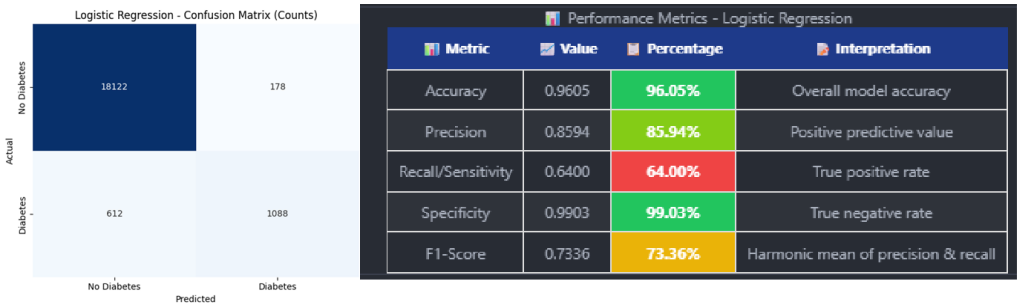


Gambar 5. Algoritma Random Forest

3. Hasil dan Pembahasan

3.1 Model dan Evaluasi Regresi Logistik

Berdasarkan hasil pengujian, model ini memperoleh nilai akurasi sebesar 96,05%, presisi 85,94%, recall 64,00%, dan F1-Score 73,36%. Nilai recall yang relatif lebih rendah menunjukkan bahwa model masih memiliki keterbatasan dalam mengenali seluruh kasus positif (pasien yang benar-benar menderita diabetes). Sementara itu, nilai specificity mencapai 99,03%, yang berarti model cukup baik dalam mengidentifikasi individu yang tidak menderita diabetes.



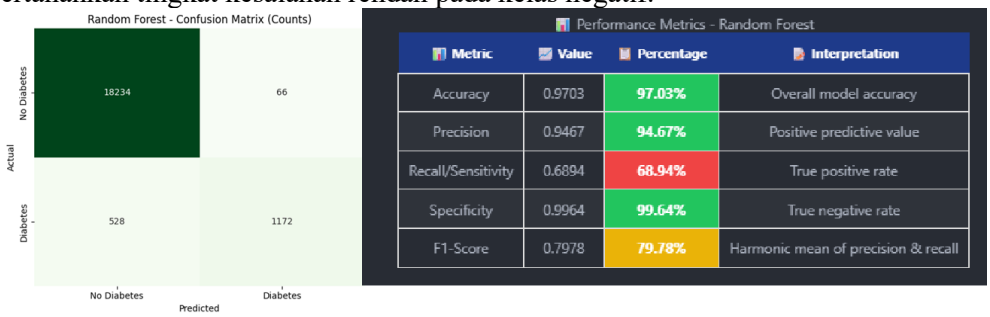
Gambar 6. Hasil evaluasi model regresi logistik.

Gambar 6 menampilkan confusion matrix dan visualisasi metrik evaluasi dari model Regresi Logistik. Terlihat bahwa jumlah prediksi benar pada kelas “No Diabetes” jauh lebih tinggi dibandingkan kelas “Diabetes”, sehingga menghasilkan ketidakseimbangan performa antar kelas. Hal ini umum terjadi pada model yang bersifat linear karena sulit menangkap hubungan non-linear antar fitur.

3.2 Model dan Evaluasi Random Forest

Model Random Forest menunjukkan hasil yang lebih unggul dibandingkan Regresi Logistik di seluruh metrik evaluasi. Berdasarkan hasil pengujian, model ini memperoleh akurasi 97,03%, presisi 94,67%, recall 68,94%, dan F1-Score 79,78%. Nilai specificity yang dicapai bahkan lebih tinggi, yaitu 99,64%. Hasil ini menunjukkan bahwa Random Forest mampu

meningkatkan keseimbangan antara kemampuan mengenali pasien berisiko (kelas positif) dan mempertahankan tingkat kesalahan rendah pada kelas negatif.

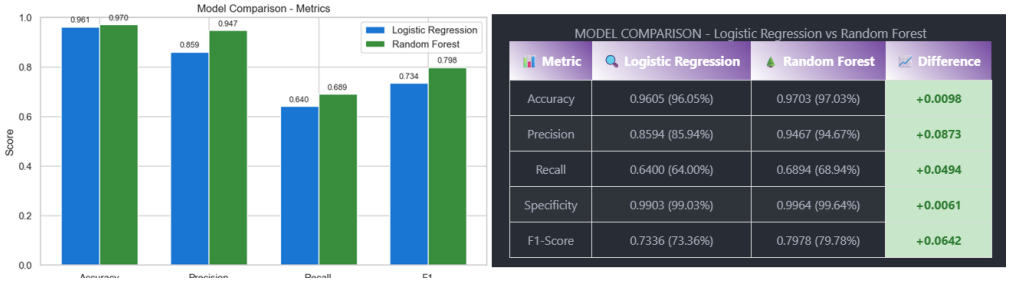


Gambar 7. Hasil evaluasi model random forest.

Gambar 7 memperlihatkan confusion matrix dan metrik performa model Random Forest. Nilai recall yang meningkat menunjukkan kemampuan model dalam mendeteksi lebih banyak kasus diabetes yang sebenarnya, sehingga sangat penting dalam konteks medis karena membantu mengurangi kemungkinan pasien berisiko tidak teridentifikasi. Selain itu, peningkatan F1-Score sebesar 6,42% dibandingkan model Regresi Logistik menegaskan bahwa Random Forest memberikan keseimbangan yang lebih baik antara presisi dan recall. Hasil ini juga menunjukkan bahwa pendekatan ensemble learning lebih efektif dalam menangkap pola kompleks antar fitur seperti kadar glukosa, BMI, dan HbA1c dibandingkan model linear tunggal.

3.3 Analisis Model

Hasil perbandingan kinerja kedua model ditunjukkan pada Gambar 8. Hasil Perbandingan Evaluasi Model. Berdasarkan gambar tersebut, algoritma Random Forest menunjukkan kinerja yang lebih baik dibandingkan Regresi Logistik pada semua metrik evaluasi. Model Random Forest memperoleh akurasi sebesar 97,03%, presisi 94,67%, recall 68,94%, dan F1-Score 79,78%, sedangkan Regresi Logistik mencatat akurasi 96,05%, presisi 85,94%, recall 64,00%, dan F1-Score 73,36%. Peningkatan paling menonjol terjadi pada metrik precision (+8,73%) dan F1-Score (+6,42%), yang menunjukkan bahwa Random Forest lebih mampu menghasilkan prediksi positif yang akurat serta menyeimbangkan antara presisi dan recall.

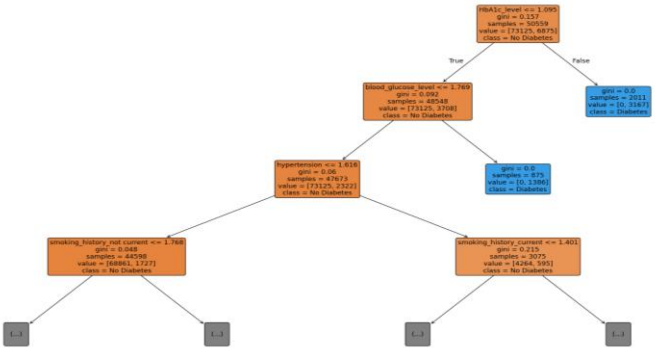


Gambar 8. Hasil Perbandingan Evaluasi Model.

Keunggulan model Random Forest disebabkan oleh mekanisme ensemble learning yang menggabungkan beberapa pohon keputusan untuk menangkap hubungan non-linear antar variabel dan mengurangi risiko overfitting. Sebaliknya, Regresi Logistik lebih unggul dari sisi interpretabilitas karena menghasilkan hubungan linier yang mudah dijelaskan. Berdasarkan hasil ini, dapat disimpulkan bahwa Random Forest merupakan model paling optimal dalam mendeteksi risiko diabetes karena mampu mencapai keseimbangan terbaik antara akurasi, presisi, dan sensitivitas, sehingga lebih efektif digunakan untuk skrining awal secara otomatis..

3.4 Visualisasi Pohon

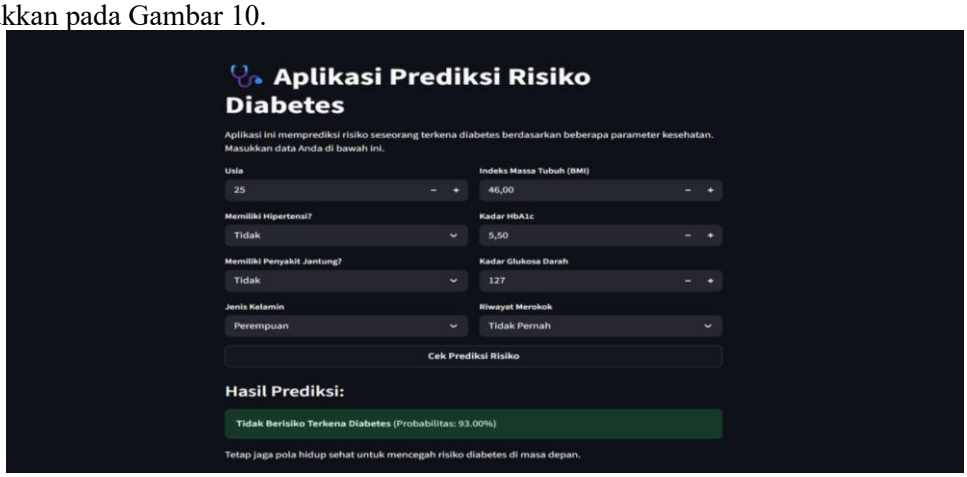
Untuk memberikan interpretasi yang lebih baik, salah satu pohon keputusan dari model Random Forest divisualisasikan seperti yang diperlihatkan pada Gambar 9. Node akar (root node) pada pohon tersebut didasarkan pada atribut HbA1c_level, yang menunjukkan bahwa variabel ini menjadi pemisah awal terpenting dalam menentukan status diabetes. Cabang-cabang berikutnya menampilkan pemisahan berdasarkan blood_glucose_level, hypertension, dan smoking_history, yang memperhalus proses klasifikasi antara "No Diabetes" dan "Diabetes".Pohon keputusan ini memiliki root node dengan kondisi HbA1c_level ≤ 1.095, gini impurity 0.157, dan 50559 sampel dengan distribusi [73125, 6875], mayoritas diklasifikasikan sebagai "No Diabetes". Jika kondisi benar, data menuju node blood_glucose_level ≤ 1.769 (gini 0.092, 48548 sampel, [73125, 3708]) yang tetap memprediksi "No Diabetes". Jika salah, data menuju node gini 0.0, 2011 sampel, [0, 3167], diklasifikasikan sebagai "Diabetes". Pemisahan berikutnya melibatkan variabel hypertension dan smoking_history, yang memperhalus klasifikasi antara "No Diabetes" dan "Diabetes".



Gambar 9. Diagram pohon random forest.

3.5 Implementasi dan Deployment Model

Model Random Forest yang terbukti memiliki performa terbaik kemudian diimplementasikan dalam aplikasi web interaktif berbasis Streamlit. Aplikasi ini dirancang agar pengguna dapat memasukkan data pasien seperti usia, BMI, kadar HbA1c, dan kadar glukosa darah untuk memperoleh hasil prediksi secara langsung. Tampilan antarmuka aplikasi ditunjukkan pada Gambar 10.



Gambar 10. Tampilan website prediksi diabetes.

Aplikasi ini diharapkan dapat menjadi alat bantu skrining non-invasif yang praktis dan mudah digunakan oleh masyarakat umum maupun tenaga medis. Integrasi hasil penelitian ke dalam sistem berbasis web ini merupakan langkah awal menuju penerapan teknologi kecerdasan buatan dalam mendukung upaya pencegahan penyakit kronis di Indonesia.

4. Kesimpulan

Berdasarkan hasil analisis dan evaluasi terhadap dua model machine learning yang digunakan, penelitian ini menyimpulkan bahwa algoritma Random Forest memiliki kinerja yang lebih unggul dibandingkan Regresi Logistik dalam memprediksi risiko diabetes. Model Random Forest menunjukkan nilai akurasi sebesar 97,03%, presisi 94,67%, recall 68,94%, dan F1-Score 79,78%, sedangkan model Regresi Logistik memperoleh akurasi 96,05%, presisi 85,94%, recall 64,00%, dan F1-Score 73,36%. Perbedaan ini menunjukkan bahwa Random Forest lebih andal dalam mengidentifikasi individu berisiko diabetes, terutama dalam meminimalkan kesalahan klasifikasi pada kasus positif (recall lebih tinggi).

Hasil analisis feature importance juga menegaskan bahwa faktor-faktor seperti kadar glukosa darah, HbA1c, BMI, dan usia merupakan variabel paling berpengaruh dalam prediksi risiko diabetes. Temuan ini sejalan dengan literatur medis yang menyatakan bahwa faktor-faktor tersebut memiliki keterkaitan kuat terhadap kemungkinan seseorang mengidap diabetes. Dengan demikian, implementasi model Random Forest dalam sistem prediksi dapat menjadi solusi yang efektif untuk membantu proses deteksi dini diabetes secara non-invasif.

Selain itu, model terbaik dari penelitian ini telah diimplementasikan dalam bentuk aplikasi berbasis web menggunakan Streamlit, yang memungkinkan pengguna melakukan prediksi risiko diabetes secara cepat, mudah, dan interaktif. Aplikasi ini berpotensi digunakan sebagai alat bantu skrining awal yang dapat diakses oleh masyarakat umum maupun tenaga medis untuk mempercepat proses identifikasi risiko.

Untuk pengembangan penelitian selanjutnya, disarankan untuk menguji model pada dataset yang berasal dari sumber berbeda agar dapat memverifikasi kemampuan generalisasi model terhadap populasi yang lebih luas. Selain itu, eksplorasi algoritma yang lebih kompleks seperti Gradient Boosting, XGBoost, atau Deep Neural Network juga dapat dilakukan guna meningkatkan akurasi dan sensitivitas prediksi. Penambahan fitur penjelasan model (model interpretability) juga perlu dipertimbangkan agar hasil prediksi dapat lebih mudah dipahami oleh pengguna di bidang kesehatan. Secara keseluruhan, penelitian ini menegaskan bahwa penerapan algoritma machine learning, khususnya Random Forest, dapat menjadi pendekatan efektif dan praktis untuk mendukung upaya deteksi dini dan pencegahan diabetes di Indonesia.

Daftar Pustaka

[1] C. A. Rahayu, “Prediksi Penderita Diabetes Menggunakan Metode Naive Bayes,” *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3, 2023, doi: 10.23960/jitet.v11i3.3055.

[2] E. Anikasari, I. E. Cantessa, F. Farmasi, I. Artikel, and D. Mellitus, “Mellitus Dengan Komplikasi Case Study of the Use of Wound Care Medication in,” pp. 28–34, 2025.

[3] M. Kholish, A. Herdianto, R. F. Setiawan, and R. Samsinar, “Perbandingan Algoritma Random Forest dan Naive Bayes dalam Memprediksi Penyakit Diabetes,” *Hubisintek*, vol. 5, no. 1, pp. 322–328, 2024, [Online]. Available: https://ojs.udb.ac.id/index.php/HUBISINTEK/article/view/4757

[4] A. Pramudyantoro, E. Utami, and D. Ariatmanto, “Penggabungan K-Nearest Neighbors Dan Lightgbm Untuk Prediksi Diabetes Pada Dataset Pima Indians: Menggunakan Pendekatan Exploratory Data Analysis,” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 3, pp. 1133–1144, 2024, doi: 10.29100/jipi.v9i3.4966.

[5] A. Pratama, A. C. Nurcahyo, and L. Firgia, “Penerapan Machine Learning dengan Algoritma Logistik Regresi untuk Memprediksi Diabetes,” *Pros. CORISINDO 2023*, pp. 116–121, 2023, [Online]. Available: https://stmikpontianak.org/ojs/index.php/corisindo/article/view/30%0Ahttps://stmikpontianak.org/ojs/index.php/corisindo/article/download/30/22

[6] 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.

[7] M. I. Gunawan, D. Sugiarto, and I. Mardianto, “Peningkatan Kinerja Akurasi Prediksi

Penyakit Diabetes Mellitus Menggunakan Metode Grid Seacrh pada Algoritma Logistic Regression," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 280, 2020, doi: 10.26418/jp.v6i3.40718.

[8] Q. R. Cahyani *et al.*, "Prediksi Risiko Penyakit Diabetes menggunakan Algoritma Regresi Logistik Diabetes Risk Prediction using Logistic Regression Algorithm Article Info ABSTRAK," *JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 2, pp. 2828–9099, 2022, doi: 10.55123/jomlai.v1i2.598.

[9] E. Safitri, D. Rofianto, N. Purwati, H. Kurniawan, and S. Karnila, "Prediksi Penyakit Diabetes Melitus Menggunakan Algoritma Machine Learning Diabetes Mellitus Disease Prediction using Machine Learning Algorithms," *JUSTIN J. Sist. dan Teknol. Inf.*, vol. 12, no. 4, pp. 760–766, 2024, doi: 10.26418/justin.v12i4.84620.

[10] R. B. Prasetyo, "Prediksi Dini Penyakit Diabetes Pada Ibu Hamil Dengan Algoritma Random Forest," *JIMUJurnal Ilm. Multidisipliner*, vol. 2, no. 04, pp. 803–812, 2024, doi: 10.70294/jimu.v2i04.440.

[11] F. T. Anggraeny and R. Mumpuni, "PREDIKSI RISIKO DIABETES MENGGUNAKAN ENSEMBLE WEIGHTED VOTING DENGAN ALGORITMA LOGISTIC REGRESSION , SVM ," 2025.

[12] M. F. Ashidiq, L. Muflikhah, and B. D. Setiawan, "Deteksi Nefropati Diabetik Pada Pasien Diabetes Melitus Menggunakan Regresi Logistik," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, pp. 2548–964, 2025, [Online]. Available: http://j-ptiik.ub.ac.id

[13] T. H. Pinem and Z. P. Putra, "Evaluasi Kinerja Algoritma Klasifikasi Deep Learning dalam Prediksi Diabetes," vol. 17, no. 1, pp. 17–28, 2025, doi: 10.22441/fifo.2025.v17i1.003.

[14] Dewi Nasien *et al.*, "Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes," *JEKIN - J. Tek. Inform.*, vol. 4, no. 1, pp. 10–17, 2024, doi: 10.58794/jekin.v4i1.640.

[15] R. Waluyo and A. S. Munir, "Optimasi Prediksi Kematian pada Gagal Jantung Analisis Perbandingan Algoritma Pembelajaran Ensemble dan Teknik Penyeimbangan Data pada Dataset," *J. Sist. dan Teknol. Inf.*, vol. 12, no. 2, p. 365, 2024, doi: 10.26418/justin.v12i2.75158.

[16] N. F. Sahamony, T. Terttiaavini, and H. Rianto, "Analisis Perbandingan Kinerja Model Machine Learning untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 413–422, 2024, doi: 10.57152/malcom.v4i2.1210.