Penerapan Model Random Forest Untuk Prediksi Pola Terjadinya Penyakit Diabetes Dengan Pendekatan *Machine Learning*

Suci Mulyani[1], Sahrul Ramadhan[2], Irfan[3]
Universitas Muhammadiyah Bima
e-mail: [1]sucimulyani410@gmail.com, [2]sahrulramadhanbinaswan@gmail.com, [3]irfanhmt05@gmail.com

Diajukan: 27 April 2026; Direvisi: 18 Mei 2026; Diterima: 20 Mei 2026

Abstrak

Penelitian ini bertujuan untuk memprediksi risiko diabetes dengan memanfaatkan rekam medis pasien di Puskesmas Ngali menggunakan algoritma Random Forest. Kasus diabetes melitus terus meningkat, tetapi deteksi dini masih kurang optimal karena gejala awal sulit dikenali dan penggunaan data kesehatan untuk prediksi risiko masih terbatas. Oleh karena itu, dibutuhkan metode pembelajaran mesin untuk membantu memprediksi risiko diabetes secara lebih akurat. Data yang digunakan terdiri dari variabel seperti kadar gula darah, tekanan darah, indeks massa tubuh (BMI), dan usia sebagai variabel prediktor, sedangkan status diabetes digunakan sebagai variabel target. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan data, pembagian data pelatihan dan pengujian, pengembangan model, dan evaluasi kinerja menggunakan akurasi, presisi,recall, dan F1-score. Hasil menunjukkan bahwa algoritma Random Forest mampu memberikan kinerja klasifikasi yang baik dan stabil dalam memprediksi risiko diabetes. Selain itu, hasil analisis menunjukkan bahwa indeks massa tubuh, kadar gula darah, dan usia merupakan variabel yang paling berpengaruh dalam prediksi diabetes. Dengan demikian, model yang dihasilkan berpotensi digunakan sebagai alat deteksi dini diabetes dan untuk mendukung pengambilan keputusan dalam pelayanan kesehatan.

Kata kunci: *Diabetes, Random Forest, Prediksi, Klasifikasi, Data Medis.*

Abstract

This study aims to predict diabetes risk by utilizing patient medical records at the Ngali Community Health Center using the Random Forest algorithm. Diabetes mellitus cases continue to increase, but early detection is still suboptimal because early symptoms are difficult to recognize and the use of health data for risk prediction is still limited. Therefore, a machine learning method is needed to help predict diabetes risk more accurately. The data used consists of variables such as blood sugar levels, blood pressure,body mass index (BMI), and age as predictor variables, while diabetes status is used as the target variable. The research stages include data collection, data pre-processing, training and testing data sharing, model development, and performance evaluation using accuracy, precision, recall, and F1-score. The results show that the Random Forest algorithm is able to provide good and stable classification performance in predicting diabetes risk. In addition, the analysis results indicate that body mass index, blood sugar levels, and age are the most influential variables in diabetes prediction. Thus, the resulting model has the potential to be used as an early detection tool for diabetes and to support decision-making in healthcare services.

Keywords*: Diabetes, Random Forest, Prediction, Classification, Medical Data*

1. Pendahuluan

Diabetes melitus merupakan penyakit kronis yang prevalensinya terus meningkat di Indonesia akibat perubahan gaya hidup penduduk. Faktor risiko utama meliputi usia, indeks massa tubuh, tekanan darah, dan kadar glukosa darah. Oleh karena itu, deteksi dini berdasarkan data klinis menggunakan analisis statistik atau pembelajaran mesin sangat diperlukan[1]. Diabetes melitus adalah penyakit kronis yang disebabkan oleh peningkatan kadar glukosa darah akibat gangguan atau resistensi insulin. Prevalensi di Indonesia meningkat dari 10,9% pada tahun 2018 menjadi 11,7% pada tahun 2023, sehingga diperlukan deteksi dini yang efektif. Karena itu, penggunaan pembelajaran mesin machine learning diperlukan untuk membantu memprediksi risiko diabetes secara lebih akurat [2].

Seiring meningkatnya kasus diabetes, teknologi telah menjadi solusi yang layak untuk membantu memprediksi risiko penyakit. Meskipun pusat kesehatan masyarakat sudah memiliki rekam medis pasien yang mencakup berbagai variabel klinis, penggunaannya masih terbatas pada tujuan administratif. Data ini

berpotensi digunakan sebagai sumber informasi untuk mendukung prediksi dan pengambilan keputusan klinis dalam perawatan kesehatan primer[3]. Namun, penggunaan pembelajaran mesin dalam perawatan kesehatan masih menghadapi beberapa tantangan, seperti kualitas data yang suboptimal, nilai yang hilang, dan variabel yang tidak relevan. Oleh karena itu, tahapan pra-pemrosesan seperti pembersihan data, ekstraksi fitur, dan pemilihan fitur diperlukan untuk meningkatkan kualitas data sebelum pemodelan. Pendahuluan sebaiknya berisi latar belakang yang jelas, permasalahan yang jelas, referensi yang relevan dengan penelitian yang dilakukan, metode atau pendekatan yang diajukan, dan kontribusi baru dari penelitian yang merupakan inovasi. Semua hal tersebut sebaiknya tidak dipisahkan menjadi sub bagian, namun merupakan satu kesatuan bagian Pendahuluan yang hanya dipisahkan oleh paragraf. Hal tersebut sebaiknya mudah dimengerti oleh rekan sesama peneliti yang terdiri dari bermacam-macam disiplin ilmu.

Penelitian sebelumnya tentang prediksi diabetes umumnya menggunakan algoritma seperti Naive Bayes dan K-Nearest Neighbor (KNN), namun berfokus pada peningkatan akurasi tanpa mempertimbangkan interpretasi hasil[4]. Selain itu, beberapa penelitian hanya menggunakan satu atau dua variabel prediktor, sehingga gagal menangkap sepenuhnya kompleksitas faktor risiko. Hal ini menunjukkan adanya kesenjangan penelitian, yaitu terbatasnya jumlah model yang mampu mengintegrasikan banyak variabel secara simultan dan memberikan interpretasi pengaruh masing-masing faktor Risiko [5].

Berdasarkan permasalahan tersebut, penelitian ini menggunakan algoritma Random Forest, sebuah metode pembelajaran mesin berbasis ensemble yang membangun beberapa pohon keputusan dan menghasilkan prediksi berdasarkan suara mayoritas atau rata-rata. Metode ini dipilih karena mampu menangani data dengan baik, memiliki tingkat akurasi yang tinggi, dan dapat mengidentifikasi pentingnya setiap variabel dalam proses prediksi[6]. Penelitian ini menggunakan lima variabel utama: usia, indeks massa tubuh, kadar glukosa, tekanan darah, dan status diabetes, sehingga memberikan analisis yang lebih komprehensif.

Urgensi penelitian ini didasarkan pada kebutuhan akan alat deteksi dini diabetes yang sederhana, akurat, dan mudah diterapkan di Puskesmas Ngali. Dengan memanfaatkan data rekam medis yang tersedia, penelitian ini diharapkan dapat menghasilkan model prediksi yang relevan dengan kondisi lokal dan dapat memberikan informasi tentang faktor-faktor yang mempengaruhi risiko diabetes. Dengan demikian, penelitian ini diharapkan dapat mendukung pengambilan keputusan oleh petugas kesehatan dan meningkatkan upaya pencegahan dan pengelolaan diabetes secara lebih efektif [7].

2. Metode Penelitian

Penelitian ini menggunakan rekam medis pasien yang diperoleh dari Puskesmas Ngali sebagai objek penelitian. Dataset berisi informasi tentang faktor risiko diabetes, termasuk usia, tekanan darah, kadar gula darah, indeks massa tubuh (ITM), dan status diabetes (ya/tidak). Data yang digunakan dikumpulkan dari tahun 2024, dikumpulkan pada tanggal 1 November 2025, melalui observasi langsung dan pencatatan oleh puskesmas. Data yang digunakan berasal dari 224 pasien yang memenuhi kriteria kelengkapan.

Ruang lingkup penelitian ini berfokus pada analisis dan prediksi risiko diabetes menggunakan variabel yang berhubungan langsung Usia, Indeks Massa Tubuh, kadar gula darah, dan tekanan darah. Penelitian ini hanya berfokus pada pengolahan dan analisis data menggunakan teknik klasifikasi, tanpa membahas aspek medis seperti diagnosis klinis atau pengobatan.

Random Forest adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan prediksi dengan membangun beberapa pohon keputusan dari data dan kemudian menggabungkan hasilnya melalui mekanisme pemungutan suara[8]. Metode ini memiliki keunggulan dalam meningkatkan akurasi dan mengurangi overfitting dibandingkan dengan pohon tunggal. Selain itu, Random Forest mampu menangani sejumlah besar data dan menunjukkan pentingnya setiap variabel, sehingga memungkinkan untuk mengidentifikasi faktor-faktor yang paling berpengaruh dalam prediksi, termasuk dalam kasus kesehatan seperti diabetes[9]. Metode ini dipilih karena mampu meningkatkan akurasi, mengurangi overfitting, dan mengidentifikasi tingkat kepentingan setiap variabel dalam model. Tahapan penelitian dilakukan secara sistematis sebagai berikut:

1) Pengumpulan Data: Memperoleh rekam medis pasien dari Puskesmas Ngali.
2) Praproses Data: Melakukan pembersihan data, menangani nilai null, dan transformasi variabel.
3) Pembagian Data: Membagi data menjadi data pelatihan dan pengujian.
4) Pemodelan: Membangun model menggunakan algoritma Random Forest.
5) Evaluasi Model: Mengukur kinerja model menggunakan metrik seperti akurasi, matriks kebingungan, dan laporan klasifikasi.

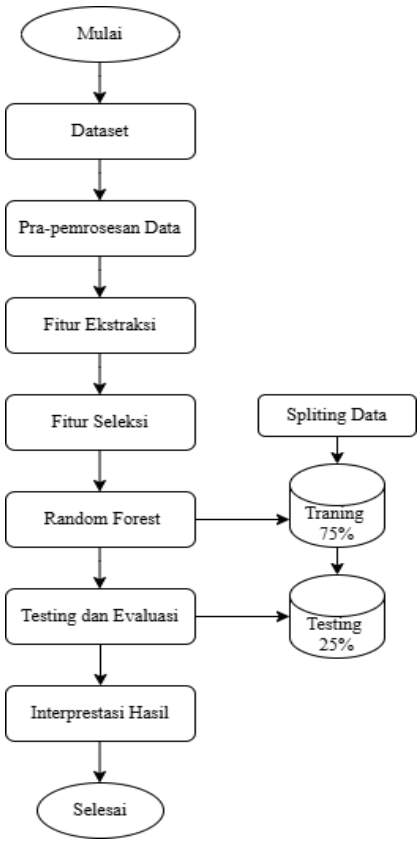
6) Analisis Hasil: Menginterpretasikan hasil prediksi dan menentukan variabel yang paling berpengaruh melalui pentingnya fitur.

Dataset yang digunakan dalam penelitian ini dipilih berdasarkan beberapa kriteria: variabel faktor risiko yang relevan, label diabetes target (1 = diabetes, 0 = bukan diabetes), kualitas data yang baik, dan data yang cukup untuk pelatihan model.

Tabel 1. Dataset Penyakit Diabetes

No	Kadar Gula	Tekanan Darah	Indeks Massa Tubuh	Umur	Ya/Tidak
1	111	72	37.1	56	0
2	180	64	34	70	1
3	133	84	40.2	80	1
4	106	92	22.7	88	1
5	171	110	45.4	67	1
6	159	64	27.4	81	1
7	180	66	42	84	1
8	146	95	29.7	72	1
9	71	70	28	35	0
10	103	66	39.1	40	0

2.2 Alur Penelitian



Gambar 1. Alur Penelitian

Penelitian ini dimulai dengan mengidentifikasi masalah yang berkaitan dengan meningkatnya jumlah kasus diabetes dan kebutuhan akan model prediksi yang akurat. Data yang digunakan adalah rekam

medis pasien yang diperoleh dari Puskesmas Ngali. Pengumpulan data melibatkan pemilihan variabel, pengecekan kualitas data, dan pengolahan data menjadi data yang siap untuk dianalisis.

Tahap selanjutnya adalah pra-pemrosesan data, yang meliputi pembersihan data, penanganan nilai yang hilang, dan normalisasi data menggunakan metode Min-Max untuk menyamakan skala antar variabel. Proses normalisasi dinyatakan dalam persamaan berikut:

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Selanjutnya, ekstraksi fitur dilakukan untuk mengubah data mentah menjadi variabel yang lebih representatif. Secara matematis, data awal dinyatakan sebagai:

$$X = (x_1, x_2, x_3, \ldots, x_n)$$

Setelah dilakukan tranformasi, diperoleh fitur baru:

$$Z = (z_1, z_2, z_3, \ldots, z_m)$$
$$\text{dengan : } m < n$$

Tranformasi tersebut dapat dinyatakan dalam bentuk fungsi:

$$Z = f(X)$$

Pada fase pemodelan, digunakan algoritma Random Forest, sebuah metode pembelajaran mesin berbasis ensemble yang membangun beberapa pohon keputusan dari data pelatihan. Prediksi diperoleh melalui mekanisme pemungutan suara mayoritas. Model kemudian menghasilkan probabilitas berdasarkan proporsi pohon yang memprediksi suatu kelas, yang dirumuskan sebagai:

$$P(Y = 1) = \frac{n1}{N}$$

Klasifikasi dilakukan secara biner, dengan kelas 1 (diabetes) dan kelas 0 (non-diabetes), dengan ambang probabilitas 0,5. Tahap selanjutnya adalah pengujian dan evaluasi model menggunakan data uji. Evaluasi dilakukan menggunakan matriks kebingungan dan beberapa metrik evaluasi, sebagai berikut:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Tahap terakhir adalah menginterpretasikan hasil, yang melibatkan analisis kinerja model berdasarkan skor evaluasi dan mengidentifikasi variabel yang paling berpengaruh melalui pentingnya fitur. Hasil interpretasi kemudian digunakan untuk menarik kesimpulan dan memberikan rekomendasi mengenai prediksi risiko diabetes.

2.3 Jenis Alat Dan Bahan Penelitian

Tabel 2. Tabel Jenis Alat

Jenis	Komponen	Spesifikasi/Fungsi
Perangkat Keras	Laptop/kompouter	Prosesor:Intel Core i3, RAM: 8gb, SSD 256, OS Windows 10

	Media Penyimpanan	Digunakan untuk pecandangan data
	Fungsi	Digunakan untuk analisis data
Perangkat Lunak	Python 3.x	Bahasa pemrograman untuk analisis data dan pemodelan
	Jupyter Notebook/Google Colab	Lingkungan kerja dan interaktif untuk coding dan visualisasi
	Pandas	Pengolahan dan pembersihan data
Phyton	Numpy	Operasi numerik dan komputasi
	Scikit-learn	Implementasi Random Forest dan evaluasi model
	Matplotlib & Seabron	Visualisasi data
	Imbalanced-learn	SMOTE
Format Data	CSV/Microsoft Excel	Pengelolaan dan pengecekan dataset

Bahan penelitian dalam penelitian ini berupa data penyakit diabetes yang merupakan data sekunder. Data tersebut diperoleh dari Puskesmas Ngali melalui pencatatan rekam medis pasien yang telah terdokumentasi, dan digunakan sebagai sumber data dalam proses analisis dan pemodelan penelitian. Dataset dipilih berdasarkan kriteria sebagai berikut:

1) Memiliki variabel dan format data faktor risiko yang relevan
2) Memiliki kelas target diabetes (0= tidak diabetes, 1= ya diabetes)
3) Kualitas data baik dan
4) Ukuran dataset sesuai untuk pelatihan model prediksi

3. Hasil dan Pembahasan

Berdasarkan isi pembahasan, jurnal ini menggunakan algoritma Random Forest sebagai metode utama dalam melakukan prediksi. Evaluasi model dilakukan secara komprehensif menggunakan beberapa metrik seperti accuracy, confusion matrix, precision, recall, dan F1-score, sehingga mampu memberikan gambaran menyeluruh terhadap kinerja model [10].

Hasil pengujian menunjukkan bahwa model Random Forest berkinerja cukup baik dengan tingkat akurasi yang wajar. Berdasarkan matriks kebingungan, model mampu mengklasifikasikan sebagian besar data dengan benar, meskipun masih terdapat beberapa kesalahan prediksi [11]. Nilai presisi, recall, dan F1-score menunjukkan bahwa model memiliki keseimbangan yang baik dalam mendeteksi kelas diabetes dan non-diabetes. Selanjutnya, hasil analisis kepentingan fitur menunjukkan bahwa indeks massa tubuh, kadar gula darah, dan usia merupakan variabel yang paling dominan dalam menentukan risiko diabetes, sedangkan tekanan darah memiliki pengaruh yang relatif lebih kecil. Hal ini menunjukkan bahwa kondisi tubuh dan kadar gula darah merupakan indikator penting dalam memprediksi diabetes [12].

Hasil penelitian disajikan dalam bentuk gambar dan grafik agar mudah dipahami. Diskusi menghubungkan hasil pengujian model dengan konsep pembelajaran mesin dan kondisi dunia nyata di sektor kesehatan, memberikan gambaran yang lebih komprehensif tentang kinerja model dan relevansinya dalam mendukung deteksi dini diabetes.

3.1. Hasil dan Analisis

3.1.1. Statistik Deskriptif

Berdasarkan analisis statistik deskriptif menggunakan Google Colab, diperoleh gambaran umum karakteristik data penelitian, termasuk nilai minimum, maksimum, dan rata-rata variabel seperti kadar gula darah, tekanan darah, indeks massa tubuh, dan usia. Hasilnya disajikan dalam bentuk tabel untuk memudahkan pemahaman distribusi data sebelum tahap pemodelan.

Tabel 3. Statistik Deskriptif

Variabel	Mean	Std	Min	25%	50%	75%	Max
Gula Darah	124.91	31.37	62	102	118	146	197
Tekanan Darah	78.11	11.96	60	68	76	88	110
Indeks Massa Tubuh	32.59	6.90	19.10	27.50	32.00	37.65	52.30
Umur	53.34	13.47	29	42	54	63	95
Status Diabetes	0.60	0.49	0	0	1	1	1

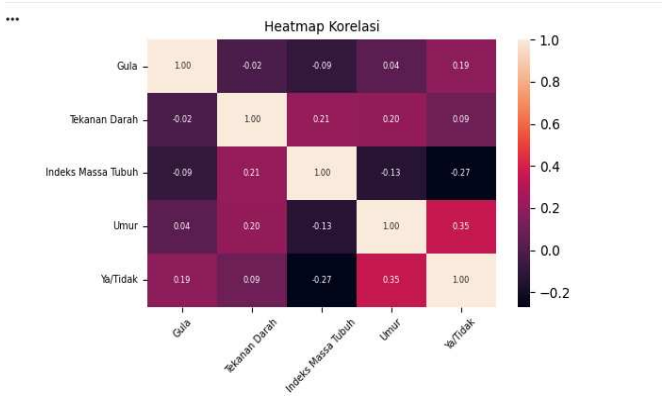
Tabel diatas menampilkan statistik deskriptif dari dataset, termasuk nilai minimum, maksimum, rata-rata, dan deviasi standar, yang dimana dapat dilihat dibawah ini :

1) Atribut Gula memiliki rata-rata 124,90 dengan nilai maksimum 197, menunjukkan variasi data yang cukup tinggi.
2) Atribut Tekanan Darah memiliki rata-rata 78,11 dengan nilai minimum 60 dan maksimum 110.
3) Atribut Indeks Massa Tubuh (ITM) memiliki rata-rata 32,58 menunjukkan bahwa sebagian besar data termasuk dalam kategori kelebihan berat badan.
4) Atribut Usia memiliki rata-rata 53,34 menunjukkan bahwa sebagian besar data berasal dari kelompok usia dewasa hingga lanjut usia.
5) Variabel Ya/Tidak memiliki nilai rata-rata 0,60 menunjukkan bahwa sebagian besar data termasuk dalam kategori positif (diabetes).

Selain itu, tidak ada nilai minimum 0 untuk atribut utama, sehingga data relatif bebas dari nilai yang tidak valid.

3.1.2. Analisis Korelasi Antar Fitur

Berdasarkan analisis korelasi fitur menggunakan Google Colab, kami mengidentifikasi hubungan antar variabel seperti kadar gula darah, tekanan darah, indeks massa tubuh, dan usia. Hasil korelasi disajikan dalam format matriks atau peta panas untuk mempermudah identifikasi hubungan antar variabel sebelum pemodelan.



Gambar 2. Korelasi

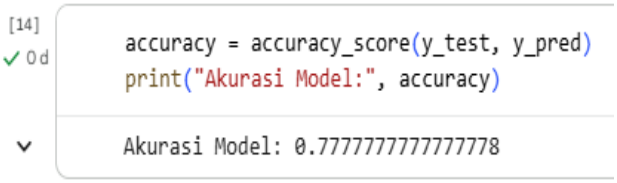
Gambar diatas menampilkan peta panas yang menunjukkan korelasi antar variabel dalam dataset. Hasil analisis menunjukkan bahwa:

1) Atribut Usia memiliki korelasi tertinggi dengan variabel target (Ya/Tidak), yaitu 0,35, yang menunjukkan bahwa risiko diabetes cenderung meningkat seiring bertambahnya usia.
2) Atribut Gula memiliki korelasi positif sebesar 0,19, yang menunjukkan bahwa kadar gula juga memengaruhi kemungkinan terkena diabetes.
3) Atribut Indeks Massa Tubuh (BMI) memiliki korelasi negatif dengan target (-0,27), yang menunjukkan hubungan terbalik, meskipun tidak terlalu kuat.
4) Atribut Tekanan Darah memiliki korelasi yang relatif rendah dengan variabel target.

Secara keseluruhan, tidak ada korelasi yang sangat kuat (>0,7), sehingga dapat disimpulkan bahwa setiap fitur memiliki kontribusi yang relatif independen terhadap proses prediksi.

3.1.3. Akurasi Model

Berdasarkan hasil pengujian menggunakan model Random Forest di Google Colab, diperoleh nilai akurasi yang menunjukkan keberhasilan model dalam memprediksi risiko diabetes. Nilai akurasi ini digunakan sebagai indikator utama untuk menilai kinerja model dalam mengklasifikasikan data dengan benar.



Gambar 3. Akurasi

Gambar diatas menampilkan Akurasi Model yang menunjukkan bahwa:
Model Random Forest yang digunakan dalam penelitian ini menunjukkan performa yang cukup baik dengan tingkat akurasi 77,78%.

3.1.4. Evaluasi Model

Berdasarkan hasil pengujian menggunakan model Random Forest di Google Colab, evaluasi model dilakukan untuk menilai kinerjanya dalam memprediksi risiko diabetes. Evaluasi menggunakan beberapa metrik, termasuk akurasi, matriks kebingungan, presisi, recall, dan F1-score. Hasil ini digunakan untuk menentukan akurasi model dan kemampuannya untuk mengklasifikasikan data dengan benar.

Tabel 4. Evaluasi Model

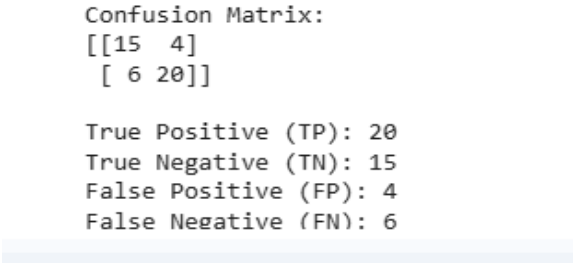
Kelas	Precision	Recall	F1-Score	Support
0	0.71	0.79	0.75	19
1	0.83	0.77	0.80	26
Accuracy			0.78	45
Macro avg	0.77	0.78	0.78	45
Weighted avg	0.78	0.78	0.78	45

Tabel 4. menampilkan Precision,Recall, f1-score, dan Support yang menunjukkan bahwa:

1) Kelas 0 (Tidak Diabetes)
   a) Presisi 0,71 yaitu dari semua kasus bukan penderita diabetes yang diprediksi, sekitar 71% benar.
   b) Recall 0,79 yaitu ari semua kasus bukan penderita diabetes, model berhasil mengidentifikasi 79% kasus.
2) Kelas 1 (Diabetes)
   a) Presisi 0,83 yaitu dari semua kasus diabetes yang diprediksi, 83% diidentifikasi dengan benar.
   b) Recall 0,77 yaitu dari semua kasus diabetes, 77% diidentifikasi dengan benar.
3) Hasil Akurasinya 78% sedangkan F1-score yaitu 78%

3.1.5. Confusion Matrix

Berdasarkan hasil pengujian menggunakan model Random Forest di Google Colab, matriks kebingungan (confusion matrix) diperoleh untuk mengevaluasi kinerja model dalam mengklasifikasikan data. Matriks kebingungan menunjukkan jumlah prediksi yang benar dan salah, yang terdiri dari True Positives, True Negatives, False Positives, dan False Negatives. Hasil ini memberikan gambaran umum tentang akurasi model dalam membedakan antara kelas diabetes dan non-diabetes.



Gambar 4. Confusion Matrix

Gambar diatas menampilkan Confusion Matrix yang menunjukkan bahwa:

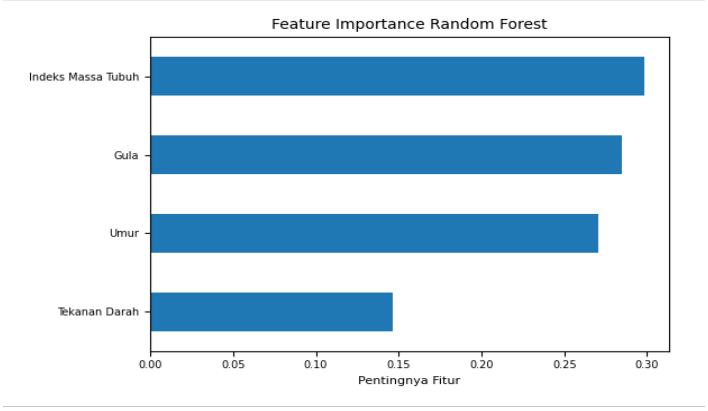
1) True Positive (TP): 20
   Data yang Tidak Diabetes diprediksi tidak diabetes (benar)
2) True Negative (TN): 15
   Data yang tidak diabetes tapi diprediksi diabetes (salah)
3) False Positive (FP): 4

Data yang diabetes tapi diprediksi tidak diabetes (salah)

4) False Negative (FN): 6
   Data yang diabetes diprediksi diabetes (benar)

3.1.6. Interpretasi Pengaruh Variavel

Berdasarkan pentingnya karakteristik, variabel indeks massa tubuh, kadar gula darah, dan usia memiliki pengaruh terbesar terhadap risiko diabetes, sedangkan tekanan darah memiliki pengaruh yang lebih rendah.



Gambar 5. Interpretasi Hasil

Gambar diatas menampilkan Interpretasi Hasil yang menunjukkan bahwa:

1) Hasil analisis pentingnya fitur menunjukkan tingkat pengaruh setiap variabel terhadap prediksi diabetes menggunakan model Random Forest. Variabel Indeks Massa Tubuh (BMI) memiliki nilai tertinggi, sekitar 0,30, menjadikannya faktor paling dominan dalam menentukan apakah seseorang berisiko terkena diabetes. Ini menunjukkan bahwa kondisi tubuh, khususnya rasio berat badan terhadap tinggi badan, secara signifikan memengaruhi model prediksi.
2) Selanjutnya, variabel Gula Darah dan Usia juga memiliki nilai yang relatif tinggi, masing-masing sekitar 0,28 dan 0,27, menunjukkan bahwa kedua faktor ini memainkan peran penting dalam memengaruhi hasil prediksi. Kadar gula darah tinggi dan peningkatan usia diketahui berhubungan erat dengan peningkatan risiko diabetes, sehingga hasil ini selaras dengan kondisi medis umum.
3) Sementara itu, variabel Tekanan Darah memiliki nilai terendah, sekitar 0,15, menunjukkan bahwa pengaruhnya terhadap prediksi diabetes relatif lebih kecil daripada variabel lain dalam model.
4) Secara keseluruhan, dapat disimpulkan bahwa model memberikan bobot yang lebih besar pada kadar gula darah dan usia terhadap risiko diabetes, sedangkan tekanan darah memiliki kontribusi yang lebih kecil.

Secara keseluruhan, hasil analisis menunjukkan bahwa model Random Forest lebih menekankan pada Indeks Massa Tubuh, kadar gula darah, dan usia dalam menentukan risiko diabetes. Hal ini menunjukkan bahwa kombinasi faktor risiko dapat memberikan hasil prediksi yang lebih komprehensif. Dengan demikian, model yang dikembangkan tidak hanya mampu melakukan klasifikasi tetapi juga memberikan informasi tentang faktor-faktor yang paling berpengaruh dalam diabetes.

4. Kesimpulan

Penelitian ini bertujuan untuk menganalisis faktor risiko dan mengembangkan model prediksi diabetes menggunakan algoritma Random Forest berdasarkan rekam medis dari Puskesmas Ngali. Hasil penelitian menunjukkan bahwa model tersebut memiliki kinerja yang cukup baik dalam memprediksi risiko diabetes, sebagaimana ditunjukkan oleh nilai akurasi dan metrik evaluasi seperti presisi, recall, dan F1-score. Lebih lanjut, hasil analisis menunjukkan bahwa indeks massa tubuh, kadar gula darah, dan usia merupakan faktor yang paling dominan dalam menentukan risiko diabetes. Dengan demikian, penelitian ini berhasil menghasilkan model prediksi yang cukup akurat dan mampu mengidentifikasi faktor-faktor penting yang mempengaruhi risiko diabetes. Penelitian lebih lanjut disarankan untuk meningkatkan jumlah data, menambahkan variabel penelitian, dan mengoptimalkan model untuk meningkatkan hasil prediksi.

Daftar Pustaka

[1] B. M. Index *Et Al.*, "Indeks Massa Tubuh , Lingkar Perut Dan Gula Darah Puasa Pada Tenaga Pendidik Wanita Poltekkes Kemenkes Bengkulu," Vol. 25, No. 3, Pp. 218–224, 2025.

[2] A. Z. Andhani, S. N. Ismanda, And R. K. Apriyani, "Penyuluhan Deteksi Dini Diabetes Melitus Sebagai Penyakit Tidak Menular Di Kelurahan Andir Kabupaten Bandung," Vol. 7, Pp. 566–576, 2026.

[3] P. Dan And D. Dini, "Jurnal Pengabdian Masyarakat," Vol. 2, No. 6, Pp. 530–541, 2025.

[4] A. Davinka Sembiring Depari, C. Cha Kirana, C. Nissa Oktariana, F. Akbar, And F. Fathoni, "Prediksi Risiko Diabetes Dengan Metode Naive Bayes: Identifikasi Faktor Risiko Utama Dan Evaluasi Akurasi Model," *Jati (Jurnal Mhs. Tek. Inform.*, Vol. 9, No. 4, Pp. 6372–6377, 2025, Doi: 10.36040/Jati.V9i4.14078.

[5] Rini Oktaviani1, "Faktor-Faktor Risiko Yang Berhubungan Dengan Kejadian Diabetes Melitus Pada Usia Dewasa Di Indonesia," *J. Kesehat. Bidkemas*, Vol. 15, No. 2, Pp. 91–99, 2024, Doi: 10.48186/F6qtd712.

[6] A. P. Argadianata *Et Al.*, "Klasifikasi Kualitas Buah Apel Menggunakan Metode Random Forest," Vol. 9, No. 2, Pp. 2016–2022, 2025.

[7] M. Anita *Et Al.*, "Klasifikasi Faktor Risiko Penyakit Jantung Menggunakan Machine," Vol. 16, No. C, Pp. 68–78, 2025.

[8] N. F. I, N. Svensons, S. A. Fatmawati, P. R. S, And K. Erviona, "Analisis Perbandingan Algoritma Klasifikasi Decision Tree , K-Nearest Neighbors , Naive Bayes , Dan Random Forest Pada Data Pemilihan Legislatif Kpu Menggunakan Kurva Roc," Vol. 01, No. 01, Pp. 7–17, 2025, Doi: 10.30873/Jodmapps.V1i1.Pp7-17.

[9] D. Manurung, B. Zealtiel, And A. H. Lubis, "Prediksi Produksi Tanaman Padi Di Indonesia Dengan Menggunakan Algoritma Random Forest Regressor," Vol. 4, No. 3, Pp. 337–345, 2025, Doi: 10.47065/Comforch.V4i3.2125.

[10] M. C. Ibrahim, "Comparison Of Diabetes Prediction Data Using Machine Learning Perbandingan Data Prediksi Diabetes Menggunakan Machine Learning," Vol. 5, No. October, Pp. 1423–1436, 2025.

[11] M. A. P. W, R. E. Saputro, And U. A. Rohmah, "Comparison Of The Accuracy Levels Of Naive Bayes , Random Forest , And Long Short-Term Memory (Lstm) Methods In Predicting Gold Jewelry Sales," Vol. 7, No. 1, Pp. 126–146, 2026.

[12] J. Teknologi, "Evaluasi Model Ensemble Learning Pada Identifikasi Faktor Risiko Diabetes Mellitus Evaluation Of Ensemble Learning Models In Identifying Risk Factors For Diabetes Mellitus," Vol. 15, No. September, Pp. 121–130, 2025, Doi: 10.34010/Jati.V15i2.16238.