



Jurnal SANTI (Sistem Informasi dan Teknologi Informasi)
Vol. 6 No.2 Tahun 2026
DOI: <https://doi.org/10.58794/santi.v5i2.2436>

Klasifikasi Risiko Kesehatan Ibu Hamil Menggunakan Random Forest dan Analisis Feature Importance

Meitrielyn¹, Yenny Desnelita^{2*}, Ramalia Noratama Putri³, Gusrianty⁴, Roni Sanjaya⁵

^{1,2,3}Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Institut Bisnis dan Teknologi Pelita Indonesia

⁴Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Institut Bisnis dan Teknologi Pelita Indonesia

e-mail: ¹meitrielyn.023@student.pelitaIndonesia.ac.id,

^{2*}yenny.desnelita@lecturer.pelitaIndonesia.ac.id,

³ramalia.noratamaputri@lecturer.pelitaIndonesia.ac.id, gusrianty@lecturer.pelitaIndonesia.ac.id,
roni.sanjaya@lecturer.pelitaIndonesia.ac.id

(Received : 15 Juli 2026; Revised: 18 Agustus 2026; Accepted: 24 Agustus 2026; Available online: 31 Agustus 2026)

Abstrak

Kualitas kesehatan ibu merefleksikan efektivitas pelayanan kesehatan, khususnya dalam mengenali potensi risiko sejak masa kehamilan untuk mengurangi kemungkinan komplikasi pada proses kehamilan dan persalinan. Berdasarkan kebutuhan tersebut, penelitian ini menerapkan algoritma Random Forest untuk mengklasifikasikan risiko kesehatan ibu dan menggunakan analisis feature importance guna mengidentifikasi faktor-faktor dominan yang berkontribusi terhadap tingkat risiko. Dataset yang digunakan adalah Maternal Health Risk dataset dari UCI Machine Learning Repository yang terdiri dari 1.014 data dengan atribut meliputi kadar gula darah, tekanan darah sistolik, usia, denyut jantung, tekanan darah diastolik, dan suhu tubuh, dan target klasifikasi terdiri atas tiga kelas risiko, yaitu risiko rendah (low risk), risiko sedang (mid risk), dan risiko tinggi (high risk). Metodologi penelitian mencakup tahap prapemrosesan data, pembagian data (80:20), pelatihan model, serta evaluasi menggunakan confusion matrix, akurasi, presisi, recall, dan F1-score. Hasil pengujian terhadap 203 data uji menunjukkan bahwa model Random Forest mencapai akurasi sebesar 83%, dengan macro-precision sebesar 0,84, macro-recall sebesar 0,84, dan macro-F1 sebesar 0,83. Analisis feature importance mengungkapkan bahwa kadar gula darah dan tekanan darah sistolik merupakan faktor yang paling berpengaruh dalam menentukan tingkat risiko kesehatan ibu. Hasil penelitian mengidentifikasi faktor prediktif dominan yang dapat menjadi dasar pengembangan sistem pendukung keputusan untuk skrining awal risiko kesehatan ibu hamil, dengan fungsi sebagai alat bantu dan bukan sebagai pengganti pemeriksaan atau keputusan klinis tenaga kesehatan.

Kata kunci: Kesehatan Ibu Hamil, Random Forest, Klasifikasi Multikelas, Feature Importance, Pembelajaran Mesin

Abstract

The quality of maternal health reflects the effectiveness of health services, especially in recognizing potential risks from the pregnancy period to reduce the possibility of complications in the pregnancy and childbirth process. Based on these needs, this study applied the Random

Forest algorithm to classify maternal health risks and used feature importance analysis to identify dominant factors that contribute to the level of risk. The dataset used is the Maternal Health Risk dataset from the UCI Machine Learning Repository which consists of 1,014 data with attributes including blood sugar levels, systolic blood pressure, age, heart rate, diastolic blood pressure, and body temperature, and the classification target consists of three risk classes, namely low risk, mid risk, and high risk. The research methodology includes the pre-data processing stage, data sharing (80:20), model training, and evaluation using confusion matrix, accuracy, precision, recall, and F1-score. The results of testing 203 test data showed that the Random Forest model achieved an accuracy of 83%, with macro-precision of 0.84, macro-recall of 0.84, and macro-F1 of 0.83. Feature importance analysis reveals that blood sugar and systolic blood pressure are the features with the highest predictive contributions. The results of the study identified the dominant predictive factors that can be the basis for the development of a decision support system for the initial screening of health risks of pregnant women, with functions as an aid and not as a substitute for examinations or clinical decisions of health workers.

Keywords: *Maternal Health Risk, Random Forest, Multiclass Classification, Feature Importance, Machine Learning*

1. Pendahuluan

Proses reproduksi manusia memiliki peranan yang sangat penting dalam kelangsungan kehidupan. Kehamilan, persalinan, dan masa nifas merupakan proses alami serta fisiologis dalam sistem reproduksi. Namun, pada kondisi tertentu, proses-proses tersebut dapat berubah menjadi situasi darurat yang membahayakan keselamatan ibu maupun bayi. Dengan demikian, peran kesehatan ibu menjadi sangat penting untuk memastikan setiap tahapan dilalui dengan aman. Kesehatan ibu mengacu pada kesehatan perempuan selama kehamilan, persalinan dan masa nifas.

Setiap tahap kehamilan, persalinan, dan masa nifas memerlukan pemantauan yang tepat untuk menjaga keselamatan ibu dan bayi. Berdasarkan data World Health Organization, sekitar 287.000 perempuan meninggal selama kehamilan, persalinan, atau masa nifas pada tahun 2020 [1]. Angka ini terlalu tinggi dan penyebab langsung paling umum dari kesehatan ibu adalah kehilangan banyak darah, infeksi, tekanan darah tinggi, aborsi yang tidak aman, dan kurangnya kesadaran dari masyarakat tentang pentingnya kesehatan ibu selama proses kehamilan. Permasalahan tersebut juga dialami oleh negara Indonesia. Tingkat kesadaran atas pentingnya kesehatan ibu selama proses kehamilan dan persalinan masih rendah di dalam masyarakat. [1]

Faktor seperti umur, tekanan darah, suhu tubuh dan gula darah dapat menyebabkan terganggunya proses persalinan yang akhirnya memengaruhi kesehatan pada ibu. [2] Peningkatan kualitas pelayanan kesehatan ibu memerlukan pengolahan data yang mampu mengidentifikasi faktor-faktor risiko selama kehamilan. Salah satu ilmu yang dapat dipakai adalah dengan menggunakan *machine learning*. *Machine learning* merupakan cabang atau aplikasi dari AI dengan berfokus untuk membuat sistem atau algoritma yang terus belajar dari data dan meningkatkan akurasi dari waktu ke waktu tanpa pemrograman tertentu. [3]

Machine learning adalah cabang kecerdasan buatan yang melatih algoritma untuk menemukan pola dan fitur tertentu dari sejumlah besar data, sehingga mampu membuat prediksi atau keputusan secara otomatis, semakin baik algoritmanya semakin tinggi pula akurasi hasil yang dihasilkan. [4] Kemampuan ini membuka potensi besar dalam dunia medis, khususnya untuk mengubah cara diagnosis dilakukan mulai dari prediksi, pengenalan pola, hingga klasifikasi karena machine learning dapat mengidentifikasi pola kompleks yang sering kali tidak terlihat oleh manusia, sehingga meningkatkan akurasi dan kecepatan diagnosis. [5]

Dalam upaya meningkatkan kualitas layanan bagi ibu hamil, teknologi machine learning telah menjadi alat yang berharga. Algoritma *Random Forest* merupakan algoritma *machine*

learning yang telah terbukti efektif dalam banyak masalah klasifikasi. *Random Forest* adalah algoritma yang digunakan untuk pengklasifikasian dataset dalam jumlah besar. Random Forest mampu menangani data dengan banyak fitur dan memiliki performa yang relatif stabil dalam berbagai tugas klasifikasi. Beberapa peneliti terdahulu menggunakan algoritma ini dengan algoritma lainnya untuk membandingkan seberapa akurat dari penggunaan algoritma ini.[6]

Berdasarkan peneliti terdahulu dengan penelitiannya yang mengkomparasi metode *K-Nearest Neighbor* dan *Random Forest* dalam klasifikasi penyakit kutil menghasilkan tingkat akurasi terhadap klasifikasinya. Hasil pengujian dengan menggunakan metode *K-Nearest Neighbor* memperoleh tingkat akurasi sebesar 76,78%, kemudian dengan metode *Random Forest* memperoleh tingkat akurasi sebesar 86,56%. [7]

Hasil penelitian tersebut sejalan dengan penelitian lain yang menunjukkan bahwa metode *Random Forest* memiliki performa klasifikasi yang lebih unggul dibandingkan *K-Nearest Neighbor*. Penelitian dalam mendeteksi penyakit liver menunjukkan bahwa *Random Forest* memperoleh akurasi sebesar 82%, lebih tinggi dibandingkan *K-Nearest Neighbor* yang hanya mencapai 72%, serta memiliki performa yang lebih stabil pada berbagai kelas data. Hal ini menunjukkan bahwa *Random Forest* mampu memberikan hasil prediksi yang lebih konsisten dalam menangani variasi data.[8]

Selain itu, penelitian dalam klasifikasi penyakit kanker payudara juga menyatakan bahwa *Random Forest* merupakan algoritma yang fleksibel dan mampu menghasilkan akurasi yang baik tanpa perlu tuning parameter yang kompleks sehingga dapat disimpulkan bahwa metode *Random Forest* memperoleh hasil yang lebih baik dan direkomendasikan untuk dataset besar dibandingkan dengan metode *K-Nearest Neighbor*. [9]

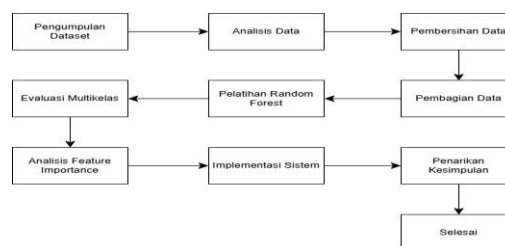
Berdasarkan penelitian terdahulu, penelitian ini dilakukan dengan tujuan untuk mengklasifikasi resiko kesehatan ibu hamil menggunakan dataset *UCI Machine Learning Repository* dengan menerapkan *machine learning* sebagai alat klasifikasi resiko kesehatan pada ibu hamil dengan menggunakan algoritma *Random Forest*. Penelitian ini diharapkan memberikan kontribusi dalam mendukung proses diagnosis medis yang lebih cepat dan akurat, serta meningkatkan kualitas hidup pasien melalui deteksi dini.

Selain menghasilkan prediksi yang akurat, interpretabilitas model juga menjadi aspek penting dalam bidang kesehatan. Oleh karena itu, penelitian ini tidak hanya berfokus pada klasifikasi menggunakan Random Forest, tetapi juga menerapkan pendekatan *feature importance* untuk mengidentifikasi variabel yang paling berpengaruh terhadap risiko kesehatan ibu hamil. Dengan mengetahui fitur yang dominan, diharapkan dapat memberikan wawasan yang lebih mendalam bagi tenaga medis dalam pengambilan keputusan.

Kontribusi utama penelitian ini terletak pada integrasi analisis *feature importance* dalam klasifikasi risiko kesehatan ibu hamil, yang tidak hanya berfokus pada performa model tetapi juga memberikan interpretasi terhadap faktor-faktor dominan yang mempengaruhi risiko.

2. Metode Penelitian

Penelitian ini dilaksanakan melalui beberapa tahapan sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Penelitian menggunakan *Maternal Health Risk Dataset* yang diperoleh dari *UCI Machine Learning Repository* sebanyak 1014 data. Dataset tersebut merupakan data sekunder yang telah dipublikasikan dan dihimpun oleh peneliti sebelumnya, yang kemudian dipublikasikan di platform *UCI Machine Learning Repository*. Contoh data dan atribut yang digunakan disajikan pada Tabel 1.

Tabel 1. Dataset Maternal Health Risk

No	Age	Systolic Blood Pressure	Diastolic Blood Pressure	Blood Sugar	Body Temperature	Heart Rate	Risk Level
1	25	130	80	15	98	86	high risk
2	35	140	90	13	98	70	high risk
3	29	90	70	8	100	80	high risk
4	30	140	85	7	98	70	high risk
5	35	120	60	6.1	98	76	low risk
...	...	...	...	...	...	...	...
1012	35	85	60	19	98	86	high risk
1013	43	120	90	18	98	70	high risk
1014	32	120	65	6	101	76	mid risk

2.2. Preprocessing Data

Preprocessing data adalah langkah atau teknik yang dilakukan untuk mengubah data mentah (*raw data*) menjadi format yang bersih, konsisten, dan berkualitas sebelum digunakan dalam proses analisis atau pelatihan model *machine learning*. Tahap ini bertujuan menghasilkan data yang bersih dan konsisten agar dapat diproses secara optimal oleh algoritma. Dalam melakukan *preprocessing data*, yang dilakukan terlebih dahulu adalah *label encoding* pada atribut *Risk Level*. Atribut tersebut memiliki tiga kategori tingkat risiko, yaitu *High Risk*, *Low Risk*, dan *Mid Risk*. Proses *label encoding* mengubah kategori tersebut menjadi nilai numerik, yaitu *High Risk* = 0, *Low Risk* = 1, dan *Mid Risk* = 2. Transformasi ini dilakukan agar data pada atribut *Risk Level* dapat direpresentasikan dalam bentuk numerik sehingga dapat digunakan dalam proses pembentukan dan pengujian model *Random Forest*. Setelah dilakukan *preprocessing data*, sebanyak 1.014 data dibagi menggunakan rasio 80:20, dengan 811 data (80%) digunakan untuk proses pelatihan model dan 203 data (20%) digunakan untuk pengujian model. Selain itu, penelitian ini tidak menerapkan standarisasi atau normalisasi pada atribut numerik. Hal tersebut dikarenakan algoritma *Random Forest* menggunakan mekanisme pemisahan (*splitting*) berdasarkan nilai fitur pada struktur *decision tree*, sehingga perbedaan skala antarfitur tidak memengaruhi proses pembentukan model. Dengan demikian, data dapat digunakan secara langsung pada tahap pelatihan dan pengujian model *Random Forest*.

2.3. Random Forest

Random Forest merupakan algoritma *ensemble* yang membangun sejumlah *Decision Tree* dan menggabungkan hasil prediksinya melalui mekanisme pemungutan suara mayoritas. Algoritma *ensemble* sendiri merupakan kombinasi dari beberapa metode *machine learning* yang digabungkan dalam satu model prediksi. Tujuan dari pendekatan ini adalah untuk mengurangi kesalahan, menurunkan bias, serta meningkatkan akurasi dalam melakukan prediksi. Rumus atau formula yang digunakan dalam *Random Forest* adalah sebagai berikut:[10]

$$\hat{C}(x) = \arg \max_{C_k} \sum_{t=1}^T I(h_t(x) = C_k) \quad (1)$$

Pada persamaan tersebut, $\hat{C}(x)$ merupakan adalah kelas hasil prediksi akhir, C_k adalah himpunan kelas, T adalah jumlah *decision tree*, $ht(x)$ adalah prediksi dari pohon ke- t terhadap data x , dan $I(ht(x)=C_k)$ merupakan fungsi indikator yang bernilai 1 apabila prediksi pohon ke- t sama dengan kelas C_k , dan bernilai 0 apabila berbeda. Kelas dengan jumlah suara terbanyak ditetapkan sebagai hasil klasifikasi akhir Random Forest.

2.4. Confusion Matrix

Confusion Matrix adalah hasil dari perhitungan dalam klasifikasi data mining yang disajikan dalam bentuk tabel. *Confusion matrix* digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan kelas aktual dan kelas prediksi. *Confusion Matrix* berfungsi sebagai alat untuk menilai kinerja model. Terdapat empat istilah yang digunakan untuk menggambarkan hasil klasifikasi dalam penerapan *confusion matrix*. Keempat istilah tersebut adalah: [11]

Tabel 2. Confusion Matrix

Kelas		Prediksi	
		True	False
Aktual	True	TP	FN
	False	FP	TN

- *False Positive* (FP) terjadi ketika data negatif salah diprediksi sebagai data positif. Sebagai contoh, data *Low Risk* atau *Mid Risk* diprediksi sebagai *High Risk*.
- *False Negative* (FN) terjadi ketika data positif salah diprediksi sebagai data negatif. Sebagai contoh, data *High Risk* diprediksi sebagai *Low Risk* atau *Mid Risk*.
- *True Positive* (TP) terjadi ketika data positif diprediksi dengan tepat. Sebagai contoh, data dengan kelas sebenarnya *High Risk* diprediksi sebagai *High Risk*.
- *True Negative* (TN) terjadi ketika data negatif diprediksi dengan benar. Sebagai contoh, data *Low Risk* atau *Mid Risk* diprediksi bukan sebagai *High Risk*.

Kinerja model dievaluasi menggunakan accuracy, precision, recall, dan F1-score.

Nilai *Accuracy* dapat diperoleh dengan persamaan 2.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2)$$

Nilai *Precision* dapat diperoleh dengan persamaan 3.

$$precision = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

Nilai *Recall* dapat diperoleh dengan persamaan 4.

$$recall = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

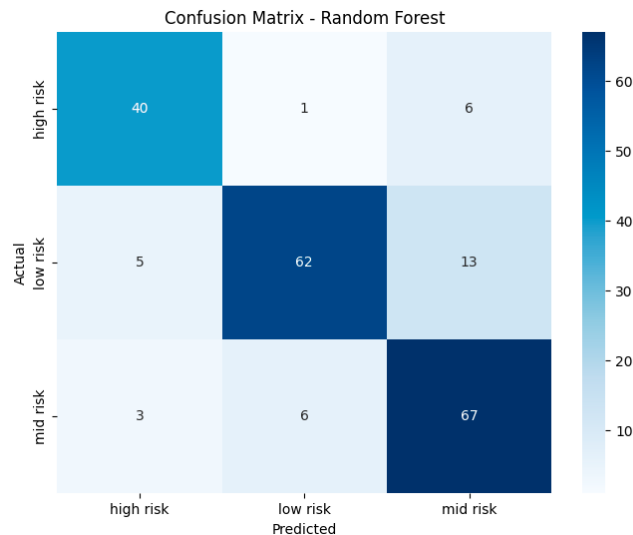
2.5. Feature Importance

Feature Importance merupakan teknik yang digunakan untuk mengukur tingkat kontribusi masing-masing fitur dalam model *Machine Learning* terhadap hasil prediksi. Pada algoritma *Random Forest*, nilai *Feature Importance* dihitung berdasarkan seberapa besar penurunan *impurity* (seperti *Gini Index*) yang dihasilkan oleh setiap fitur dalam membangun pohon keputusan. [12] Semakin tinggi nilai *Feature Importance*, semakin besar kontribusi fitur tersebut terhadap keputusan model. Dalam penelitian ini, *Feature Importance* digunakan untuk mengidentifikasi variabel yang paling berperan dalam menentukan tingkat risiko kesehatan ibu hamil, seperti tekanan darah, kadar gula darah, dan faktor fisiologis lainnya.

3. Hasil dan Pembahasan

3.1. Hasil Random Forest

Setelah melalui proses algoritma *Random Forest*, model yang dibentuk dapat divisualisasikan. Confusion matrix hasil pengujian model Random Forest disajikan pada Gambar 2.



Gambar 2. Confusion Matrix Hasil Pengujian Model Random Forest

Berdasarkan Gambar 2, model *Random Forest* mampu memprediksi dengan benar 40 data *high risk*, 62 data *low risk*, dan 67 data *mid risk* dari total 203 data uji. Pada kelas *high risk*, terdapat 1 data *high risk* yang salah diklasifikasikan sebagai *low risk*. Model menunjukkan performa yang relatif seimbang pada ketiga kelas, meskipun terdapat perbedaan nilai precision dan recall. Secara keseluruhan, *confusion matrix* menunjukkan bahwa model mampu membedakan ketiga kategori risiko dengan jumlah prediksi benar yang lebih tinggi dibandingkan prediksi yang salah.

```

--- Confusion Matrix - Random Forest ---
Accuracy: 0.83
Precision: 0.84
Recall: 0.83
F1-Score: 0.83

```

	precision	recall	f1-score	support
High Risk	0.83	0.85	0.84	47
Low Risk	0.90	0.78	0.83	80
Mid Risk	0.78	0.88	0.83	76
accuracy			0.83	203
macro avg	0.84	0.84	0.83	203
weighted avg	0.84	0.83	0.83	203

Gambar 3. Classification Report Model Random Forest

Berdasarkan Gambar 3, model *Random Forest* memperoleh akurasi sebesar 83% dari total 203 data uji. Pada kelas *high risk*, model memperoleh *precision* sebesar 83% dan *F1-score* sebesar 0,84. Sementara itu, kelas *mid risk* memiliki nilai *recall* sebesar 88%, yang

menunjukkan bahwa sebagian besar data aktual pada kelas tersebut berhasil diidentifikasi oleh model. Nilai *macro average* dan *weighted average* masing-masing berada pada kisaran 0,83–0,84. Model menunjukkan performa yang relatif konsisten pada data pengujian yang digunakan. Dengan akurasi sebesar 83%, model *Random Forest* dapat digunakan untuk melakukan klasifikasi tingkat risiko pada data yang diuji.[13]

3.2. Implementasi Sistem

Antarmuka aplikasi dikembangkan berbasis web untuk memfasilitasi pengguna dalam memasukkan parameter kesehatan dan memperoleh hasil prediksi. Tampilan halaman input aplikasi disajikan pada Gambar 4.

Gambar 4. Tampilan Aplikasi Web

Antarmuka pengguna dalam penelitian ini diimplementasikan menggunakan formulir berbasis web yang intuitif untuk memfasilitasi penginputan data klinis ibu. Halaman input terdiri dari enam parameter kesehatan utama, yaitu kadar glukosa darah, tekanan darah sistolik, usia, detak jantung, tekanan darah diastolik, dan suhu tubuh.

Hasil Prediksi

Low Risk

Selamat Ibu, risiko kesehatan Ibu tergolong rendah. Ibu dalam kondisi yang relatif aman. Tetap jaga kesehatan, konsumsi makanan bergizi, dan lakukan pemeriksaan rutin agar kondisi kehamilan tetap terpantau dengan baik.

Hasil Prediksi

Mid Risk

Perhatian Ibu, terdapat beberapa indikator kesehatan yang perlu diperhatikan. Disarankan untuk lebih rutin memeriksakan kondisi kehamilan ke tenaga kesehatan serta menjaga pola makan dan istirahat agar risiko tidak meningkat.

Hasil Prediksi

High Risk

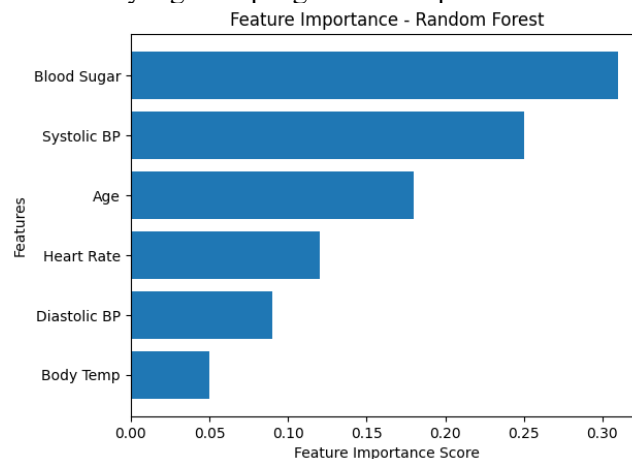
Ibu perlu waspada karena terdapat indikasi risiko tinggi pada kondisi kesehatan kehamilan. Sangat disarankan untuk segera berkonsultasi dengan dokter atau tenaga medis agar mendapatkan penanganan yang tepat demi keselamatan Ibu dan bayi.

Gambar 5. Hasil Prediksi

Gambar 5 menunjukkan tampilan hasil prediksi tingkat risiko kesehatan ibu hamil yang dihasilkan oleh sistem setelah pengguna memasukkan data parameter kesehatan. Setelah proses klasifikasi dilakukan, sistem akan menampilkan hasil prediksi dalam bentuk *pop-up* notifikasi yang berisi kategori tingkat risiko kehamilan beserta penjelasan singkat yang dapat membantu pengguna memahami kondisi kesehatan yang diprediksi oleh sistem. Hasil prediksi yang ditampilkan terdiri dari tiga kategori tingkat risiko, yaitu *Low Risk*, *Mid Risk*, dan *High Risk*.

3.3. Analisis Feature Importance

Selain evaluasi performa model, dilakukan analisis *Feature Importance* untuk mengetahui kontribusi masing-masing variabel terhadap hasil klasifikasi. Berdasarkan hasil perhitungan, fitur yang memiliki pengaruh terbesar dalam menentukan tingkat risiko kesehatan ibu hamil adalah kadar gula darah (*Blood Sugar*), diikuti oleh tekanan darah sistolik (*Systolic BP*), dan usia (*Age*). Fitur lain seperti detak jantung (*Heart Rate*), tekanan darah diastolic (*Diastolic BP*), dan suhu tubuh (*Body Temp*) memiliki kontribusi yang lebih rendah, namun tetap berperan dalam meningkatkan akurasi model secara keseluruhan. Hasil ini menunjukkan bahwa faktor kadar gula darah dan tekanan darah sistolik merupakan indikator utama dalam menentukan risiko kesehatan ibu hamil. Analisis ini memberikan nilai tambah dibandingkan model klasifikasi biasa, karena tidak hanya menghasilkan prediksi tetapi juga memberikan interpretasi terhadap faktor-faktor yang mempengaruhi hasil prediksi tersebut.



Gambar 6. Feature Importance

Berdasarkan hasil analisis *feature importance* pada Gambar 6, diketahui bahwa fitur gula darah (*Blood Sugar*) memiliki kontribusi paling besar dengan nilai sebesar 0,31, diikuti oleh tekanan darah sistolik (*Systolic BP*) sebesar 0,25 dan usia (*Age*) sebesar 0,18. Sementara itu, fitur detak jantung (*Heart Rate*), tekanan darah sistolik (*Diastolic BP*), dan suhu tubuh (*Body Temp*) memiliki kontribusi yang lebih kecil terhadap model. Hal ini menunjukkan bahwa kadar gula darah dan tekanan darah sistolik merupakan fitur dengan kontribusi terbesar dalam klasifikasi tingkat risiko kesehatan ibu hamil.

4. Kesimpulan

Penelitian ini menunjukkan bahwa algoritma *Random Forest* mampu melakukan klasifikasi risiko kesehatan ibu hamil ke dalam tiga tingkat risiko, yaitu *low risk*, *mid risk*, dan *high risk*. Berdasarkan hasil pengujian pada 203 data uji, model memperoleh akurasi sebesar 83%. Hasil analisis *feature importance* menunjukkan bahwa kadar gula darah (*Blood Sugar*) dengan nilai kepentingan sebesar 0,31 dan tekanan darah sistolik (*Systolic Blood Pressure*) sebesar 0,25 merupakan dua fitur dengan kontribusi terbesar dalam klasifikasi tingkat risiko kesehatan ibu hamil. Hasil tersebut menunjukkan bahwa algoritma *Random Forest* dapat digunakan untuk melakukan klasifikasi sekaligus memberikan informasi mengenai kontribusi masing-masing fitur terhadap hasil klasifikasi sehingga dapat mendukung interpretasi hasil penelitian.

Penelitian ini memiliki beberapa keterbatasan. Dataset yang digunakan merupakan *Maternal Health Risk Dataset* yang diperoleh dari *UCI Machine Learning Repository*, sehingga hasil penelitian masih dipengaruhi oleh karakteristik dan cakupan dataset tersebut serta belum dapat digeneralisasikan pada populasi yang lebih luas. Selain itu, penelitian ini belum melakukan validasi eksternal menggunakan dataset dari sumber atau populasi yang berbeda. Pengujian juga masih berfokus pada performa model dan belum mencakup pengujian aplikasi secara langsung oleh tenaga kesehatan. Oleh karena itu, penelitian selanjutnya dapat menggunakan dataset dari berbagai sumber atau populasi yang lebih beragam, melakukan validasi eksternal, serta melibatkan tenaga kesehatan dalam pengujian dan evaluasi model agar hasil klasifikasi dapat lebih mendukung penerapannya dalam bidang kesehatan.

Daftar Pustaka

- [1] B. Dave, "What is machine learning?," 2025. [Online]. Available: <https://www.ibm.com/think/topics/machine-learning>.
- [2] K. D. Tzamourta, M. G. Tsipouras, P. Angelidis, D. G. Tsalikakis and E. Orovou, "Maternal Health Risk Detection: Advancing Midwifery with Artificial Intelligence," *Healthcare*, vol. 13, no. 7, p. 833, 6 April 2025, doi: 10.3390/healthcare13070833
- [3] W. H. Organization, "Maternal health," WHO, 2025. [Online]. Available: https://www.who.int/health-topics/maternal-health#tab=tab_1.
- [4] B. Leo, "Random Forest," *Springer Nature*, vol. 45, pp. 5-32, 2021. Available: <https://link.springer.com/article/10.1023/A:1010933404324>
- [5] U. Erdiansyah, A. Irmansyah Lubis and K. Erwansyah, "Komparasi Metode K-Nearest Neighbor dan Random Forest Dalam Prediksi Akurasi Klasifikasi Pengobatan Penyakit Kutil," *Jurnal Media Informatika Budidarma*, vol. 6, no. 1, p. 208, 25 January 2022, doi: 10.30865/mib.v6i1.3373
- [6] N. C. Caballé, J. L. Castillo-Sequera, J. A. Gómez-Pulido, J. M. Gómez-Pulido and M. L. Polo-Luque, "Machine learning applied to diagnosis of human diseases: A systematic review," *Applied Sciences (Switzerland)*, vol. 10, no. 15, pp. 1-27, 1 August 2022, doi: 10.3390/app10155135
- [7] S. Rezazadeh, "Review of Machine Learning," *Journal of Research and Review Archives*, vol. 4, no. 2(S), pp. 174-204, 15 Juny 2025. Available: www.tmp.twistingmemoirs.com
- [8] J. J. J. Simanjuntak and M. A. I. Pakereng, "Perbandingan Metode Random Forest dan KNN Dalam Mendeteksi Penyakit Liver," *Journal of Information Technology and Computer Science(INTECOMS)*, vol. 9, no. 2, pp. 335-342, 2026, doi: <https://doi.org/10.31539/j9se3a69>
- [9] V. Angkasa and J. J. Pangaribuan, "Komparasi Tingkat Akurasi Random Forest dan KNN untuk Mendiagnosis Penyakit Kanker Payudara," *Journal Information System Development (ISD)*, vol. vii, no. 1, pp. 34-41, 2022, doi: <http://dx.doi.org/10.19166/xxxx>
- [10] D. H. Depari, Y. Widiastiwi and M. M. Santoni, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *Informatik : Jurnal Ilmu Komputer (IFTK)*, vol. 18, no. 3, pp. 239-248, 28 Desember 2022, doi: 10.52958/iftk.v18i3.4694
- [11] J. Tohka and M. Van Gils, "Evaluation of machine learning algorithms for health and wellness applications: A tutorial," *Computers in Biology and Medicine*, vol. 132, pp. 104-324, May 2021, doi: 10.1016/j.compbiomed.2021.104324
- [12] M. A. Alfarisi, B. Sumanto and I. P. F. Rachmawan, "Pengembangan Model Machine Learning untuk Deteksi Penyakit Diabetes Menggunakan Analisis Gini Importance," *Journal of Internet and Software Engineering*, vol. 6, no. 2, pp. 127-134, 29 November

2025, doi: 10.22146/jise.v6i2.16551

- [13] G. L. Simegn and M. Z. Degu, "Maternal health risk analysis, automated pregnancy risk level identification and status monitoring system based on multivariable clinical data," *Health Technology*, vol. 9, pp. 1-1, March 2025, doi: 10.21037/ht-23-10