

Perbandingan Kinerja Naïve Bayes, SVM, dan Random Forest dalam Klasifikasi Risiko Kehamilan

Reva Ekalia*, Dyah Ayu Megawaty

Fakultas Teknik dan Ilmu Komputer, Prodi Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ^{1,*}reva_ekalia@teknokrat.ac.id, ²dyahayumegawaty@teknokrat.ac.id

Email Penulis Korespondensi: reva_ekalia@teknokrat.ac.id

Submitted: 22/04/2026; Accepted: 02/06/2026; Published: 05/06/2026

Abstrak—Klasifikasi tingkat risiko kehamilan merupakan aspek penting dalam mendukung deteksi dini terhadap potensi komplikasi pada ibu hamil. Namun, sebagian besar penelitian sebelumnya masih berfokus pada penggunaan satu algoritma tertentu dan hanya mengandalkan metrik accuracy, sehingga belum memberikan gambaran kinerja model secara komprehensif pada klasifikasi multikelas. Selain itu, perbandingan kinerja antar algoritma dengan pendekatan evaluasi yang lebih lengkap masih terbatas. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma Naïve Bayes, Support Vector Machine (SVM), dan Random Forest dalam mengklasifikasikan tingkat risiko kehamilan menggunakan Maternal Health Risk Dataset dari UCI Machine Learning Repository yang terdiri dari 1.014 data dengan enam atribut kesehatan maternal. Metode yang digunakan meliputi preprocessing data, optimasi hyperparameter menggunakan GridSearchCV, serta evaluasi model menggunakan Stratified K-Fold Cross Validation dengan nilai $k = 10$. Kinerja model diukur menggunakan metrik accuracy, precision, recall, dan F1-score untuk memberikan evaluasi yang lebih komprehensif. Hasil penelitian menunjukkan bahwa algoritma Random Forest memiliki kinerja terbaik dengan nilai accuracy sebesar 0.8629, precision sebesar 0.8704, recall sebesar 0.8629, dan F1-score sebesar 0.8635, diikuti oleh SVM dan Naïve Bayes. Keunggulan Random Forest disebabkan oleh kemampuannya dalam menggabungkan beberapa pohon keputusan serta menangkap hubungan non-linear antar fitur, sehingga menghasilkan prediksi yang lebih akurat dan stabil. Dengan demikian, Random Forest direkomendasikan sebagai metode yang paling efektif dalam klasifikasi risiko kehamilan berbasis data kesehatan maternal.

Kata Kunci: Klasifikasi Risiko Kehamilan; Maternal Health; Machine Learning; Naïve Bayes; Support Vector Machine; Random Forest

Abstract—Classifying pregnancy risk levels is a crucial aspect in supporting early detection of potential complications in pregnant women. However, most previous studies have focused on a single algorithm and relied solely on accuracy metrics, thus failing to provide a comprehensive picture of model performance in multiclass classification. Furthermore, performance comparisons between algorithms using more comprehensive evaluation approaches are still limited. This study aims to analyze and compare the performance of the Naïve Bayes, Support Vector Machine (SVM), and Random Forest algorithms in classifying pregnancy risk levels using the Maternal Health Risk Dataset from the UCI Machine Learning Repository, which consists of 1,014 data sets with six maternal health attributes. The methods used include data preprocessing, hyperparameter optimization using GridSearchCV, and model evaluation using Stratified K-Fold Cross Validation with $k = 10$. Model performance was measured using accuracy, precision, recall, and F1-score metrics to provide a more comprehensive evaluation. The results showed that the Random Forest algorithm had the best performance with an accuracy value of 0.8629, precision of 0.8704, recall of 0.8629, and F1-score of 0.8635, followed by SVM and Naïve Bayes. The superiority of Random Forest is due to its ability to combine several decision trees and capture non-linear relationships between features, resulting in more accurate and stable predictions. Thus, Random Forest is recommended as the most effective method in pregnancy risk classification based on maternal health data.

Keywords: Pregnancy Risk Classification; Maternal Health; Machine Learning; Naïve Bayes; Support Vector Machines; Random Forest

1. PENDAHULUAN

Kehamilan merupakan proses fisiologis yang pada umumnya berlangsung normal, namun dalam kondisi tertentu dapat berkembang menjadi situasi berisiko yang membahayakan keselamatan ibu maupun janin. Komplikasi seperti preeklamsia, hipertensi gestasional, dan diabetes gestasional merupakan faktor utama yang berkontribusi terhadap peningkatan morbiditas dan mortalitas maternal. Organisasi Kesehatan Dunia (WHO) melaporkan bahwa sekitar 287.000 perempuan meninggal akibat komplikasi kehamilan dan persalinan pada tahun 2020, yang sebagian besar terjadi di negara berkembang dan sebenarnya dapat dicegah melalui deteksi dini serta penanganan yang tepat [1]. Di Indonesia, Angka Kematian Ibu (AKI) masih tergolong tinggi, yaitu sekitar 189 per 100.000 kelahiran hidup pada tahun 2020 berdasarkan hasil Long Form Sensus Penduduk 2020 oleh Badan Pusat Statistik [2].

Identifikasi risiko kehamilan umumnya dilakukan berdasarkan parameter klinis seperti *Age*, *SystolicBP*, *DiastolicBP*, *BS*, *BodyTemp*, dan *HeartRate*. Parameter tersebut berkaitan erat dengan kondisi kesehatan maternal yang dapat memicu terjadinya komplikasi selama kehamilan. Namun, proses identifikasi yang dilakukan secara manual masih memiliki berbagai keterbatasan, seperti potensi subjektivitas, inkonsistensi dalam pengambilan keputusan, serta ketergantungan pada pengalaman tenaga medis. Selain itu, meningkatnya jumlah data pasien dan kompleksitas variabel klinis menyebabkan proses analisis menjadi semakin sulit jika dilakukan secara konvensional. Seiring dengan perkembangan teknologi, *machine learning* menjadi salah satu pendekatan yang banyak digunakan dalam bidang kesehatan untuk membantu proses klasifikasi dan prediksi berbasis data. *Machine learning* merupakan cabang kecerdasan buatan yang memungkinkan sistem untuk belajar dari data serta meningkatkan kinerja tanpa harus diprogram secara eksplisit. Pendekatan ini mampu mengolah data dalam jumlah besar dan mengidentifikasi pola

tersembunyi yang tidak mudah dikenali secara manual oleh manusia, sehingga dapat meningkatkan akurasi dan objektivitas dalam proses pengambilan keputusan [3],[4].

Beberapa penelitian sebelumnya telah menggunakan dataset yang sama, yaitu *Maternal Health Risk Dataset* yang dapat diakses di UCI Machine Learning Repository [5], untuk mengklasifikasikan tingkat risiko selama kehamilan. Penelitian yang dilakukan oleh Purnama et al. (2024) menerapkan berbagai algoritma klasifikasi seperti Random Forest, Extra Trees, Extreme Gradient Boosting, Decision Tree, dan Light Gradient Boosting Machine, di mana algoritma Random Forest memberikan hasil terbaik dengan tingkat akurasi mencapai 82,15% [6].

Penelitian yang dilakukan Kurniawan (2025) juga memanfaatkan dataset yang sama dengan membandingkan algoritma Decision Tree (ID3 dan C4.5), Naïve Bayes, dan K-Nearest Neighbor (KNN) dengan metode optimasi parameter berbasis *Grid Search*. Temuan dari penelitian ini menunjukkan bahwa algoritma Decision Tree (ID3 dan C4.5) menghasilkan kinerja terbaik dengan akurasi mencapai 85,65% [7]. Selain itu, penelitian lain menunjukkan bahwa penerapan *machine learning* dalam data kesehatan ibu dapat meningkatkan keakuratan prediksi risiko dengan memanfaatkan data klinis yang rumit. Penelitian yang dilakukan oleh Rosyid dan Subektiningsih (2023) menggunakan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan tingkat risiko kesehatan bagi ibu hamil dan membuktikan bahwa setelah melalui proses preprocessing, akurasinya naik dari 60% menjadi 71% [8].

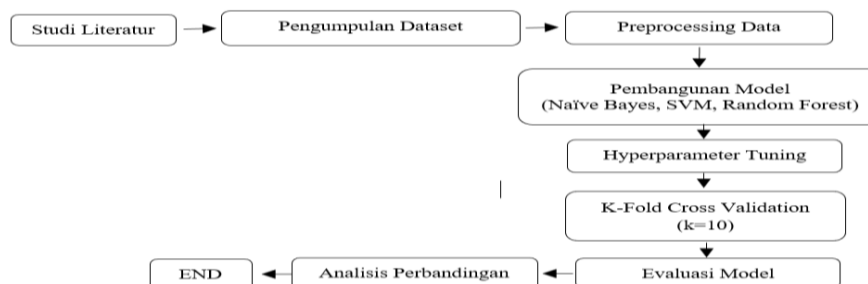
Penelitian yang dilakukan oleh Muntiari et al. (2025) menganalisis berbagai algoritma machine learning seperti Logistic Regression, K-Nearest Neighbor, Decision Tree, Random Forest, dan Support Vector Machine dalam mengklasifikasikan risiko preeklampsia pada ibu hamil. Temuan dari studi ini mengindikasikan bahwa algoritma Decision Tree dan Random Forest menunjukkan performa yang paling unggul dengan melebihi 87% [9]. Selain itu, penelitian lain yang dilakukan oleh Vasudevan et al. (2025) menunjukkan bahwa penerapan machine learning pada data kesehatan maternal mampu meningkatkan kemampuan prediksi risiko morbiditas dan mortalitas ibu berdasarkan data rekam medis elektronik [10].

Berdasarkan keterbatasan pada penelitian sebelumnya, penelitian ini melakukan analisis perbandingan kinerja algoritma Naïve Bayes, Support Vector Machine (SVM), dan Random Forest dalam klasifikasi risiko kehamilan menggunakan pendekatan evaluasi yang lebih komprehensif [6], [7], [8]. Evaluasi model tidak hanya menggunakan metrik accuracy, tetapi juga precision, recall, dan F1-score untuk memberikan gambaran performa klasifikasi multikelas secara lebih menyeluruh [11], [12]. Selain itu, penelitian ini menerapkan optimasi hyperparameter menggunakan GridSearchCV dan metode Stratified K-Fold Cross Validation guna menghasilkan evaluasi model yang lebih stabil dan representatif [13], [11]. Melalui pendekatan tersebut, penelitian ini diharapkan dapat memberikan referensi dalam menentukan algoritma machine learning yang paling efektif untuk klasifikasi risiko kehamilan berbasis data kesehatan maternal.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menerapkan pendekatan kuantitatif yang didukung oleh machine learning dengan cara eksperimen untuk menilai kinerja algoritma klasifikasi dalam mengidentifikasi tingkat risiko kehamilan. Proses penelitian dilakukan secara teratur mulai dari pengolahan data hingga penilaian model menggunakan metrik kuantitatif seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Untuk menjelaskan alur penelitian yang dilaksanakan, langkah-langkah penelitian divisualisasikan dalam skema pada Gambar 1.



Gambar 1. Skema Penelitian

Berdasarkan Gambar 1, proses penelitian ini diawali dengan studi literatur untuk menetapkan dasar teori dan metode yang akan digunakan. Selain itu, dilakukan pengumpulan dan pengolahan data dengan menggunakan dataset *Maternal Health Risk* dari *UCI Machine Learning Repository* [5]. Data yang telah dikumpulkan kemudian melalui tahap preprocessing yang meliputi pemisahan fitur dan label, label encoding, serta feature scaling menyesuaikan distribusi data agar lebih optimal dalam proses pelatihan model. Pembangunan model klasifikasi menggunakan algoritma Naïve Bayes, Support Vector Machine (SVM), dan Random Forest adalah langkah berikutnya. Untuk meningkatkan kinerja model, proses penyesuaian hyperparameter dilakukan dengan menggunakan pendekatan *GridSearchCV*. Ini dilakukan untuk mendapatkan kombinasi parameter terbaik untuk setiap algoritma. Beberapa penelitian menunjukkan bahwa *GridSearchCV* dapat meningkatkan akurasi dan efisiensi klasifikasi model.

Selanjutnya evaluasi model dilakukan menggunakan metode *K-Fold Cross Validation* dengan nilai $k=10$. Metode ini membagi dataset menjadi berbagai bagian (lipat), dan setiap bagian digunakan secara berurutan sebagai uji data. Teknik ini mampu memberikan estimasi kinerja model yang lebih stabil dan mengurangi bias dibandingkan dengan metode pembagian data biasa [14]. Kinerja model kemudian diukur menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, *F1-score*, serta *confusion matrix* untuk memberikan gambaran yang komprehensif terhadap performa klasifikasi [14]. Hasil evaluasi tersebut selanjutnya dianalisis untuk menentukan algoritma yang memiliki performa terbaik dalam mengklasifikasikan tingkat risiko kehamilan.

2.2 Studi Literatur

Studi literatur dilakukan untuk mendapatkan dasar teori serta memahami perkembangan penelitian mengenai klasifikasi risiko kehamilan dengan menggunakan pendekatan machine learning. Melalui kajian literatur, peneliti dapat mengidentifikasi metode yang telah diterapkan pada penelitian sebelumnya dan menemukan celah penelitian yang menjadi fondasi untuk studi ini [6], [8]. Pencarian literatur dilakukan melalui jurnal nasional dan internasional yang relevan dengan topik penelitian. Penelusuran dilakukan dengan kata kunci seperti “*maternal health risk*”, “*machine learning*”, dan “*classification*” melalui database seperti *Google Scholar*. Literatur yang dijadikan acuan terfokus pada publikasi dalam beberapa tahun terakhir untuk memastikan tetap terkini dan relevan. Hasil studi literatur ini digunakan sebagai acuan dalam memilih algoritma Naïve Bayes, Support Vector Machine (SVM), dan Random Forest, serta dalam menetapkan metode evaluasi yang digunakan dalam penelitian [6], [8], [10].

2.3 Dataset

Dataset yang dipakai dalam studi ini diambil dari UCI *Machine Learning Repository* dan terdiri atas 1.014 data observasi yang menggambarkan kondisi kesehatan ibu hamil di daerah pedesaan Bangladesh [5]. Atribut yang digunakan mencakup *Age*, *SystolicBP*, *DiastolicBP*, *BS (Blood Sugar)*, *BodyTemp*, dan *HeartRate* sebagai variabel independen, sedangkan *RiskLevel* berfungsi sebagai variabel dependen yang dibagi menjadi tiga kategori, yaitu *Low Risk*, *Mid Risk*, dan *High Risk*. Data ini dimanfaatkan untuk menentukan tingkat risiko kehamilan berdasarkan parameter fisiologis yang berkaitan dengan kemungkinan komplikasi selama masa kehamilan. Rincian mengenai atribut dataset disajikan dalam Tabel 1.

Tabel 1. Keterangan Dataset

No	Atribut	Keterangan
1	Age	Usia ibu hamil
2	SystolicBP	Tekanan darah sistolik
3	DiastolicBP	Tekanan darah diastolik
4	BS	Kadar gula darah
5	BodyTemp	Suhu tubuh
6	HeartRate	Detak jantung
7	RiskLevel	Tingkat risiko kehamilan

Berdasarkan Tabel 1, dataset yang digunakan dalam penelitian ini terdiri dari berbagai atribut yang menunjukkan kondisi kesehatan ibu hamil. *Age*, *SystolicBP*, *DiastolicBP*, *BS (Blood Sugar)*, *BodyTemp*, dan *HeartRate* sebagai variabel independen, serta *RiskLevel* sebagai variabel dependen. Setiap atribut memiliki peran penting dalam menentukan tingkat risiko kehamilan. Nilai masing-masing atribut digunakan sebagai dasar untuk membagi risiko menjadi tiga kategori, yaitu *Low Risk*, *Mid Risk*, dan *High Risk*.

2.4 Preprocessing Data

Tahap preprocessing data dilakukan untuk memastikan dataset berada dalam keadaan siap untuk digunakan dalam proses pembuatan model klasifikasi. Tujuan dari tahap ini adalah untuk menjaga kualitas data, memastikan format yang konsisten, serta menyesuaikan struktur data agar sesuai dengan kebutuhan algoritma *machine learning*. Preprocessing merupakan langkah penting karena secara langsung berdampak pada kinerja model yang dihasilkan [9], [10]. Dalam penelitian ini, preprocessing dilakukan melalui beberapa langkah utama, yaitu *data cleaning*, *label encoding*, dan *feature scaling*.

2.4.1 Data Cleaning

Data cleaning bertujuan untuk menemukan dan menyelesaikan masalah kualitas data, seperti nilai hilang (*missing value*), duplikasi data, serta perbedaan jenis data. Proses ini sangat penting karena kualitas data memiliki dampak langsung pada kinerja dan kemampuan model dalam generalisasi. Data yang memiliki *missing value*, *noise*, atau label yang tidak konsisten dapat mengakibatkan bias serta mengurangi akurasi model [10], [15]. Ringkasan hasil analisis *missing value* disajikan pada Tabel 2.

Tabel 2. Pemeriksaan Missing Value

Attribute	Missing Values
Age	0

Attribute	Missing Values
SystolicBP	0
DiastolicBP	0
BS	0
BodyTemp	0
HeartRate	0
RiskLevel	0

Berdasarkan Tabel 2, seluruh atribut dalam dataset tidak terdapat nilai yang hilang maupun data duplikat sehingga dataset dinyatakan valid secara struktural dan dapat digunakan pada tahap preprocessing selanjutnya.

2.4.2 Label Encoding

Variabel target dalam dataset, yaitu *RiskLevel*, terbagi menjadi tiga kelompok, yaitu *Low Risk*, *Mid Risk*, dan *High Risk*. Karena algoritma machine learning memerlukan data dalam format angka, maka dilakukan proses pengkodean label untuk mentransformasikan kelompok-kelompok ini menjadi representasi numerik. Proses ini mengubah setiap kelompok menjadi kode angka tanpa mengubah arti dari klasifikasinya.

Mengubah data kategorikal menjadi format numerik merupakan hal krusial dalam tahap persiapan agar data dapat dihitung secara matematis oleh algoritma klasifikasi [16], [17]. Melalui proses ini, kelompok *Low Risk* direpresentasikan sebagai 0, *Mid Risk* sebagai 1, dan *High Risk* sebagai 2. Representasi ini kemudian dimanfaatkan sebagai variabel target dalam tahap pelatihan dan penilaian model.

2.4.3 Feature Scaling

Feature scaling dilakukan untuk menyamakan ukuran setiap variabel numerik agar tidak ada perbedaan dampak antara variabel saat model dilatih. Langkah ini penting karena variasi rentang nilai antara fitur bisa memengaruhi performa algoritma dalam membangun model klasifikasi [15], [18]. Dalam studi ini, proses standardisasi diterapkan dengan metode *StandardScaler*, yang mengubah data sehingga memiliki rata-rata 0 dan standar deviasi standar 1. Rumusan untuk proses standardisasi adalah sebagai berikut:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

dimana x mencerminkan nilai asli dari fitur, μ adalah rata-rata dari data latih, σ adalah deviasi standar, dan z adalah nilai yang telah dinormalisasikan. Proses ini menghasilkan data dengan ukuran yang konsisten, sehingga pelatihan model menjadi lebih stabil dan efisien, terutama untuk algoritma yang berbasis jarak seperti (SVM) [18].

2.5 Pembangunan Model

Tahap pembangunan model dilakukan dengan menerapkan tiga algoritma klasifikasi, yaitu *Naïve Bayes*, *Support Vector Machine (SVM)*, dan *Random Forest*. Ketiga algoritma ini dipilih untuk dianalisis dan dibandingkan efektivitasnya dalam mengklasifikasikan tingkat risiko kehamilan berdasarkan parameter fisiologis pada dataset yang digunakan [19]. Pada tahap ini, proses pelatihan dan pengujian model dilakukan menggunakan metode Stratified K-Fold Cross Validation. Dataset dibagi menjadi beberapa subset (fold), di mana setiap fold secara bergantian berperan sebagai data uji, sementara fold lainnya digunakan sebagai data latih. Pendekatan ini memungkinkan evaluasi model yang lebih stabil dan mengurangi bias terhadap pembagian data tertentu [20], [21].

2.5.1 Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menerapkan Teorema Bayes dengan asumsi independensi fitur yang kuat. Dalam konteks data kesehatan maternal, algoritma ini dipilih karena efisiensinya yang tinggi dalam memproses parameter klinis yang bersifat kontinu dan diskret. Dalam kondisi medis, asumsi independensi antar variabel jarang terpenuhi sepenuhnya. Namun, Naïve Bayes memiliki keunggulan dalam memberikan estimasi probabilitas kelas risiko dengan cepat, yang sangat penting untuk menerapkan sistem pendukung keputusan medis pada perangkat dengan sumber daya komputasi terbatas [22].

2.5.2 Support Vector Machine (SVM)

Algoritma Support Vector Machine (SVM) adalah metode klasifikasi yang bekerja dengan menentukan hyperplane optimal untuk memisahkan data antar kelas berdasarkan margin maksimal. Karakteristik SVM sangat relevan untuk data kesehatan karena mampu menangani pola data klinis yang kompleks dan tidak linear melalui penerapan fungsi kernel. Hal ini memungkinkan model untuk melakukan pemetaan fitur kesehatan maternal ke dalam dimensi yang lebih besar, sehingga mereka dapat menemukan batas keputusan yang lebih akurat untuk membedakan kategori risiko rendah, menengah, dan tinggi [23], [24].

2.5.3 Random Forest

Algoritma Random Forest merupakan algoritma berbasis ensemble learning yang membangun sekumpulan pohon keputusan secara acak dan menentukan hasil akhir melalui mekanisme majority voting. Dalam penelitian ini, algoritma

ini relevan karena mampu meningkatkan stabilitas prediksi dan mengurangi risiko overfitting, yang sering terjadi saat menggunakan pohon keputusan tunggal. Random Forest juga berguna untuk mengklasifikasikan tingkat risiko kehamilan karena dapat menangani data kesehatan yang rumit dengan banyak variabel [25].

2.6 Optimasi Hyperparameter dan Validasi Model

Optimasi hyperparameter dan validasi model dilakukan menggunakan metode GridSearchCV yang dikombinasikan dengan *Stratified K-Fold Cross Validation* [13]. *GridSearchCV* digunakan untuk mencari kombinasi parameter terbaik dari setiap algoritma klasifikasi, yaitu Naïve Bayes, Support Vector Machine (SVM), dan Random Forest. Metode ini bekerja dengan menguji berbagai kombinasi hyperparameter secara sistematis untuk menentukan konfigurasi model yang menghasilkan kinerja terbaik [13].

Dalam proses optimasi ini, *Stratified K-Fold Cross Validation* digunakan dengan nilai $k = 10$. Pada metode ini, dataset dibagi menjadi beberapa subset (fold) dengan distribusi kelas yang seimbang di setiap fold [11]. Setiap fold digunakan secara bergantian dalam proses pelatihan dan evaluasi model, sehingga seluruh data memiliki kesempatan yang sama untuk berkontribusi dalam proses pelatihan dan evaluasi model. *Stratified K-Fold Cross Validation* bertujuan untuk meningkatkan stabilitas hasil evaluasi model serta mengurangi bias akibat pembagian data, terutama pada kasus klasifikasi multikelas [11]. Metode ini memungkinkan pengukuran kinerja model secara lebih representatif karena tidak bergantung pada satu subset data tertentu. Dalam penelitian ini, metode pembagian data train-test split konvensional tidak digunakan, karena seluruh proses validasi dilakukan menggunakan *Stratified K-Fold Cross Validation* yang terintegrasi dalam proses optimasi model.

2.7 Evaluasi Model

Evaluasi model dilakukan untuk menilai kinerja setiap algoritma dalam mengkategorikan tingkat risiko kehamilan berdasarkan data pengujian dan mengukur kemampuan model untuk menggeneralisasikan pada data yang tidak dipakai saat pelatihan. Mengingat penelitian ini menggunakan klasifikasi dengan beberapa kelas (*Low Risk*, *Mid Risk*, dan *High Risk*), maka evaluasi dilaksanakan dengan pendekatan *one-vs-rest* [20]. Kinerja model selanjutnya diukur menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* [11].

2.7.1 Confusion Matrix

Confusion matrix digunakan untuk menilai sejauh mana model berfungsi dengan membandingkan kelas yang sebenarnya dengan kelas yang diprediksi [25], [11]. Mengingat penelitian ini melibatkan tiga kategori risiko, maka confusion matrix yang digunakan berbentuk matriks 3×3 seperti pada Tabel 3.

Tabel 3. Confusion Matrix Multikelas (3×3)

Actual \ Predicted	Low Risk	Mid Risk	High Risk
Low Risk	C11	C12	C13
Mid Risk	C21	C22	C23
High Risk	C31	C32	C33

Tabel 3 memperlihatkan bahwa nilai C_{ij} menggambarkan jumlah data pada kelas aktual ke- i yang diprediksi sebagai kelas ke- j . Elemen diagonal (C11, C22, C33) menandakan prediksi yang akurat, sedangkan elemen di luar diagonal menunjukkan kesalahan dalam pengklasifikasian. Baris-baris merepresentasikan kelas aktual dan kolom-kolom menunjukkan hasil prediksi dari model. Informasi ini menjadikan landasan dalam perhitungan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*.

2.7.2 Accuracy

Accuracy digunakan untuk menilai sejauh mana model dapat dengan tepat mengklasifikasikan data pengujian secara keseluruhan [25], [11]. Rumus *accuracy* dilihat pada Persamaan (2):

$$Accuracy = \frac{C_{11} + C_{22} + C_{33}}{\sum_{i=1}^3 \sum_{j=1}^3 C_{ij}} \quad (2)$$

Dalam rumus itu, C11, C22, dan C33 menunjukkan jumlah prediksi yang tepat untuk setiap kelas, sedangkan bagian penyebut mencerminkan total keseluruhan data dalam *confusion matrix*.

2.7.3 Precision

Precision digunakan untuk mengukur ketepatan model dalam memprediksi suatu kelas tertentu. Pada klasifikasi multikelas, nilai *precision* dihitung untuk setiap kelas berdasarkan jumlah prediksi benar dibandingkan dengan seluruh data yang diprediksi sebagai kelas tersebut [25], [11].

Rumus *precision* ditunjukkan pada Persamaan (3):

$$Precision_i = \frac{C_{ii}}{\sum_{j=1}^3 C_{ji}} \quad (3)$$

Pada rumus tersebut, C_{ii} merupakan jumlah prediksi yang benar pada kelas ke- i , sedangkan penyebut menunjukkan total data yang diprediksi sebagai kelas tersebut. Nilai *precision* keseluruhan dihitung menggunakan pendekatan *macro average*.

2.7.4 Recall

Recall berfungsi untuk menilai seberapa baik model dalam mengenali semua data yang termasuk dalam kategori tertentu [25], [11]. Metrik ini menggambarkan berapa banyak data sebenarnya dalam suatu kategori yang dapat diprediksi dengan akurat oleh model. Rumus untuk menghitung *recall* terdapat dalam Persamaan (4):

$$Recall_i = \frac{C_{ii}}{\sum_{j=1}^3 C_{ji}} \quad (4)$$

Pada rumus tersebut, C_{ii} merupakan jumlah prediksi yang benar pada kelas ke- i , sedangkan penyebut menunjukkan total data aktual pada kelas tersebut. Nilai *recall* keseluruhan dihitung menggunakan pendekatan *macro average*.

2.7.5 F1-Score

F1-score adalah ukuran rata-rata harmonis antara *precision* dan *recall* yang digunakan untuk menilai kinerja dengan yang lebih seimbang [25], [11]. Rumus *F1-score* ditunjukkan pada Persamaan (5):

$$F1_i = 2 \times \frac{Precision_i \times Recall_i}{Precision_i + Recall_i} \quad (5)$$

Nilai *F1-score* keseluruhan dihitung menggunakan pendekatan *macro average*.

2.8 Analisis Perbandingan

Analisis perbandingan dilakukan untuk mengidentifikasi algoritma dengan kinerja terbaik berdasarkan hasil evaluasi model yang diperoleh. Perbandingan ini menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* yang diperoleh dari hasil validasi menggunakan *Stratified K-Fold Cross Validation*. *Accuracy* berfungsi untuk mengukur ketepatan model secara keseluruhan, sementara *precision* dan *recall* digunakan untuk mengevaluasi kemampuan model dalam mengklasifikasikan setiap kelas. *F1-score* berfungsi sebagai indikator keseimbangan antara *precision* dan *recall* dalam klasifikasi multikelas [25], [11], [12]. Dalam penelitian ini, nilai *precision*, *recall*, dan *F1-score* dihitung dengan pendekatan *macro average* agar setiap kelas (*Low Risk*, *Mid Risk*, dan *High Risk*) memiliki bobot yang setara dalam proses evaluasi. Pendekatan ini bertujuan untuk memastikan bahwa model tidak hanya berkinerja baik pada satu kelas tertentu, tetapi juga dapat mengklasifikasikan semua kategori risiko dengan seimbang. Selain membandingkan nilai metrik evaluasi, analisis terhadap *confusion matrix* juga dilakukan untuk menemukan pola kesalahan klasifikasi pada setiap algoritma. Model yang menunjukkan nilai evaluasi terbaik serta distribusi kesalahan yang lebih merata antar kelas ditetapkan sebagai model unggulan dalam penelitian ini.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Preprocessing Data

Pada tahap preprocessing, dilakukan proses data cleaning untuk memastikan kualitas data yang digunakan dalam penelitian. Hasil pemeriksaan menunjukkan bahwa dataset tidak memiliki nilai hilang (*missing value*) maupun data duplikat, sehingga tidak diperlukan penanganan tambahan pada tahap ini. Selanjutnya, dilakukan transformasi data untuk mempersiapkan data agar dapat digunakan oleh algoritma *machine learning*. Proses ini meliputi konversi data ke dalam format numerik serta normalisasi nilai fitur menggunakan metode standarisasi (*feature scaling*). Tahapan ini bertujuan untuk menyamakan skala antar fitur. Hal ini terutama berlaku untuk algoritma seperti Support Vector Machine (SVM) yang sensitif terhadap perbedaan skala data.

Metode pembagian data train-test split konvensional tidak digunakan dalam penelitian ini. Metode *Stratified K-Fold Cross Validation* digunakan untuk melakukan seluruh proses pelatihan dan evaluasi model, dengan nilai $k = 10$. Metode ini memungkinkan setiap data digunakan secara bergantian dalam proses pelatihan dan validasi, sehingga menghasilkan evaluasi yang lebih stabil dan representatif.

3.2 Hasil Pengujian Model

Pengujian model dilakukan menggunakan tiga algoritma klasifikasi, Naïve Bayes, Support Vector Machine (SVM), dan Random Forest. Proses evaluasi dilakukan menggunakan *Stratified K-Fold Cross Validation* dengan nilai $k = 10$. Penilaian kinerja model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, serta analisis *confusion matrix* untuk mengevaluasi kinerja model dalam mengklasifikasikan setiap kategori risiko.

3.2.1 Naïve Bayes

Dalam penelitian ini, algoritma Naïve Bayes digunakan untuk mengklasifikasi tingkat risiko kehamilan. Model ini didasarkan pada prinsip probabilitas dengan asumsi independensi antar fitur. Performa model Naïve Bayes ditampilkan pada Tabel 4, hasil evaluasi menggunakan *Stratified K-Fold Cross Validation* dengan nilai $k = 10$.

Tabel 4. Hasil Evaluasi Model Naïve Bayes

Metric	Nilai
Accuracy	0.6017
Precision	0.5938
Recall	0.6017
F1-score	0.5580

Performa model Naïve Bayes ditunjukkan pada Tabel 4, model Naïve Bayes menghasilkan nilai accuracy sebesar 0.6017, precision sebesar 0.5938, recall sebesar 0.6017, dan F1-score sebesar 0.5580. Hasil ini menunjukkan bahwa performa model masih relatif rendah dibandingkan dengan algoritma lainnya dalam mengklasifikasikan tingkat risiko kehamilan. Hal ini disebabkan oleh asumsi independensi antar fitur pada Naïve Bayes yang tidak sepenuhnya terpenuhi dalam dataset, sehingga mempengaruhi kemampuan model dalam menangkap hubungan antar fitur secara optimal.

3.2.2 Support Vector Machine (SVM)

Algoritma Support Vector Machine (SVM) digunakan dalam penelitian ini untuk mengklasifikasikan tingkat risiko kehamilan berdasarkan data yang tersedia. Performa model SVM ditunjukkan pada Tabel 5 berdasarkan hasil evaluasi menggunakan *Stratified K-Fold Cross Validation* dengan nilai $k = 10$.

Tabel 5. Hasil Evaluasi Model SVM

Metric	Nilai
Accuracy	0.7031
Precision	0.7048
Recall	0.7031
F1-score	0.6869

Berdasarkan Tabel 5, model SVM menghasilkan nilai accuracy sebesar 0.7031, precision sebesar 0.7048, recall sebesar 0.7031, dan F1-score sebesar 0.6869. Hasilnya menunjukkan bahwa model SVM dapat mengklasifikasikan tingkat risiko kehamilan dengan lebih baik daripada Naïve Bayes. Model ini memiliki performa yang lebih baik karena kemampuan SVM untuk memisahkan data berdasarkan hyperplane optimal.

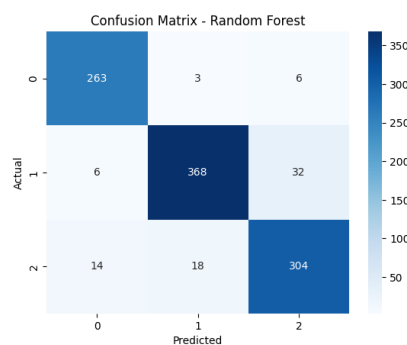
3.2.3 Random Forest

Dalam penelitian ini, algoritma Random Forest digunakan untuk mengklasifikasikan tingkat risiko berdasarkan data kehamilan yang tersedia. Random Forest adalah metode ensemble learning yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi dan mengurangi risiko overfitting. Performa model Random Forest ditunjukkan pada Tabel 6 berdasarkan hasil evaluasi menggunakan *Stratified K-Fold Cross Validation* dengan nilai $k = 10$.

Tabel 6. Hasil Evaluasi Random Forest

Metric	Nilai
Accuracy	0.8629
Precision	0.8704
Recall	0.8629
F1-score	0.8635

Berdasarkan Tabel 6, algoritma Random Forest menghasilkan nilai accuracy sebesar 0.8629, precision sebesar 0.8704, recall sebesar 0.8629, dan F1-score sebesar 0.8635. Hasil ini menunjukkan bahwa Random Forest memiliki performa terbaik dibandingkan algoritma lainnya dalam mengklasifikasikan tingkat risiko kehamilan. Untuk melihat distribusi hasil klasifikasi secara lebih rinci, digunakan confusion matrix yang ditunjukkan pada Gambar 2.



Gambar 2. Confusion Matrix Random Forest

Berdasarkan confusion matrix pada Gambar 2, model ini menunjukkan kemampuan klasifikasi yang sangat baik pada setiap kategori risiko. Pada kelas *Low Risk* (kelas 0), sebagian besar data berhasil diklasifikasikan dengan benar sebanyak 263 data, dengan kesalahan yang sangat minimal. Pada kelas *Mid Risk* (kelas 1), model mampu mengklasifikasikan 368 data dengan benar, meskipun masih terdapat beberapa kesalahan klasifikasi ke kelas *High Risk* sebanyak 32 data. Sementara itu, pada kelas *High Risk* (kelas 2), model berhasil mengklasifikasikan 304 data dengan benar, dengan beberapa kesalahan klasifikasi ke kelas *Mid Risk*. Secara keseluruhan, hasil ini menunjukkan bahwa model Random Forest memiliki kemampuan yang baik untuk membedakan antara kategori risiko, meskipun terdapat beberapa kesalahan pada kelas dengan atribut yang sebanding.

Random Forest memiliki keunggulan dalam penelitian ini karena dapat menghasilkan prediksi yang lebih akurat dan stabil dengan menggabungkan berbagai pohon keputusan. Selain itu, metode ini memiliki kemampuan untuk menangkap hubungan non-linear antar fitur, yang menghasilkan kinerja klasifikasi yang lebih baik daripada algoritma lainnya, sehingga dapat disimpulkan bahwa Random Forest merupakan model terbaik dalam penelitian ini.

3.3 Pembahasan

Berdasarkan hasil pengujian yang telah dilakukan, terdapat perbedaan kinerja antara algoritma Naïve Bayes, Support Vector Machine (SVM), dan Random Forest dalam mengklasifikasikan tingkat risiko kehamilan. Perbandingan kinerja setiap algoritma ditampilkan pada Tabel 7.

Tabel 7. Perbandingan Performa Model

Model	Accuracy	Precision	Recall	F1-score
Naïve Bayes	0.6017	0.5938	0.6017	0.5580
SVM	0.7031	0.7048	0.7031	0.6869
Random Forest	0.8629	0.8704	0.8629	0.8635

Berdasarkan Tabel 7, algoritma Random Forest menunjukkan kinerja terbaik dengan nilai accuracy sebesar 0.8629, diikuti oleh SVM dengan nilai 0.7031, dan Naïve Bayes dengan nilai 0.6017. Selain itu, nilai precision, recall, dan F1-score juga menunjukkan bahwa Random Forest memiliki performa yang lebih baik dan lebih seimbang dibandingkan algoritma lainnya dalam mengklasifikasikan setiap kategori risiko. Kinerja Naïve Bayes yang relatif rendah disebabkan oleh asumsi independensi antar fitur yang tidak sepenuhnya terpenuhi dalam data kesehatan. Variabel seperti tekanan darah, kadar gula darah, dan detak jantung cenderung saling berkaitan, sehingga mempengaruhi kemampuan model dalam membedakan kategori risiko secara optimal. Sementara itu, algoritma SVM menunjukkan performa yang lebih baik dibandingkan Naïve Bayes karena kemampuannya dalam membentuk hyperplane optimal untuk memisahkan data antar kelas. Namun, model ini masih memiliki keterbatasan dalam menangkap kompleksitas distribusi data, sehingga masih ditemukan kesalahan klasifikasi pada beberapa kategori risiko.

Random Forest menunjukkan kinerja terbaik karena kemampuan untuk menggabungkan berbagai pohon keputusan untuk meningkatkan akurasi dan stabilitas prediksi. Selain itu, metode ini memiliki kemampuan untuk menangkap hubungan non-linear antar fitur, yang membuatnya lebih efektif dalam mengidentifikasi pola data yang kompleks. Oleh karena itu, dapat disimpulkan bahwa Random Forest adalah algoritma terbaik untuk mengklasifikasikan tingkat risiko kehamilan pada dataset yang digunakan dalam penelitian ini, berdasarkan seluruh metrik evaluasi yang digunakan.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma Naïve Bayes, Support Vector Machine (SVM), dan Random Forest dalam mengklasifikasikan tingkat risiko kehamilan menggunakan Maternal Health Risk Dataset. Proses penelitian dilakukan melalui tahap preprocessing data yang meliputi data cleaning, label encoding, dan feature scaling, serta pelatihan dan evaluasi model menggunakan metode *Stratified K-Fold Cross Validation* dengan nilai $k = 10$. Evaluasi kinerja model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk memberikan gambaran performa yang komprehensif. Hasil pengujian menunjukkan bahwa algoritma Random Forest adalah yang terbaik dengan nilai 0.8629, precision sebesar 0.8704, recall sebesar 0.8629, dan F1-score sebesar 0.8635. Hasil ini menunjukkan bahwa Random Forest dapat mengklasifikasikan tingkat risiko kehamilan dengan lebih baik dan lebih stabil daripada algoritma lain. Untuk penelitian mendatang, disarankan untuk menggunakan dataset yang lebih besar, menambahkan variasi atribut kesehatan, dan melakukan optimasi *hyperparameter* yang lebih mendalam untuk meningkatkan kinerja model. Selain itu, analisis lebih lanjut seperti feature importance juga dapat dilakukan untuk mengetahui faktor-faktor yang paling berpengaruh terhadap tingkat risiko kehamilan.

REFERENCES

- [1] World Health Organization, "Trends in maternal mortality 2000–2020," Geneva, Feb. 2023. [Online]. Available: <https://www.who.int/publications/i/item/9789240068759>



- [2] Kementerian Kesehatan Republik Indonesia, “Profil Kesehatan Indonesia 2023,” Jakarta, 2023. [Online]. Available: <https://www.kemkes.go.id/id/profil-kesehatan-indonesia-2023>
- [3] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput. Sci.*, vol. 2, no. 1, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.
- [4] A. S. Nyamawe, M. M. Mjahiidi, N. E. Nnko, S. A. Diwani, G. G. Minja, and K. Malyango, *Practical Machine Learning: A Beginner’s Guide with Ethical Insights*. Boca Raton: CRC Press, 2025. doi: 10.1201/9781003486817.
- [5] M. Ahmed, “Maternal Health Risk Dataset,” UCI Machine Learning Repository. Accessed: Apr. 14, 2026. [Online]. Available: <https://archive.ics.uci.edu/dataset/451/maternal+health+risk>
- [6] J. J. Purnama, N. K. Hikmawati, and S. Rahayu, “Analisis algoritma klasifikasi untuk mengidentifikasi potensi risiko kesehatan ibu hamil,” *Journal of Applied Computer Science and Technology*, vol. 5, no. 1, pp. 120–127, 2024, doi: 10.52158/jacost.v5i1.809.
- [7] A. Kurniawan, “Perbandingan optimasi algoritma klasifikasi decision tree, naive bayes, dan knn menggunakan optimize parameter grid pada tingkat risiko ibu hamil,” *Faktor Exacta*, vol. 18, no. 2, p. 111, 2025, doi: 10.30998/faktorexacta.v18i2.28051.
- [8] M. Rosyid and Subektiningsih, “Klasifikasi tingkat risiko kesehatan ibu hamil menggunakan algoritma support vector machine,” *Indonesian Journal of Computer Science Attribution*, vol. 12, no. 5, pp. 2798–2807, 2023, doi: 10.33022/ijcs.v12i5.3372.
- [9] N. R. Muntiari, K. H. Hanif, Syamsiah, and Rokaya, “Klasifikasi penyakit preeklamsia pada ibu hamil menggunakan perbandingan algoritma machine learning,” *Jurnal Khatulistiwa Informatika*, vol. 13, no. 2, pp. 96–102, 2025, doi: 10.31294/jki.v13i2.11148.
- [10] L. Vasudevan *et al.*, “Machine Learning Models to Predict Risk of Maternal Morbidity and Mortality From Electronic Medical Record Data: Scoping Review,” *J. Med. Internet Res.*, vol. 27, p. e68225, 2025, doi: 10.2196/68225.
- [11] O. Rainio, J. Teuvo, and R. Klén, “Evaluation metrics and statistical tests for machine learning,” *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-56706-x.
- [12] S. Sathyanarayanan and B. R. Tantri, “Confusion matrix-based performance evaluation metrics,” *African Journal of Biomedical Research*, vol. 27, no. 4s, pp. 4023–4031, 2024, doi: 10.53555/ajbr.v27i4s.4345.
- [13] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, “Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis,” *Informatics*, vol. 8, no. 4, 2021, doi: 10.3390/informatics8040079.
- [14] I. Dwita, S. Tarigan, R. Habibi, R. Nuraini, and S. Fatonah, “Evaluasi algoritma klasifikasi machine learning kategori nilai akhir tunjangan kinerja pegawai,” *Jurnal Sistem Cerdas*, vol. 6, no. 3, pp. 251–261, 2023, doi: 10.37396/jsc.v6i3.246.
- [15] K. Maharana, S. Mondal, and B. Nemade, “A review: Data pre-processing and data augmentation techniques,” *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, 2022, doi: 10.1016/j.gltp.2022.04.020.
- [16] S. S. Al Mashrafi, L. Tafakori, and M. Abdollahian, “Predicting maternal risk level using machine learning models,” *BMC Pregnancy Childbirth*, vol. 24, no. 1, 2024, doi: 10.1186/s12884-024-07030-9.
- [17] S. S. Berutu, H. Budiati, J. Jatmika, and F. Gulo, “Data preprocessing approach for machine learning-based sentiment classification,” *JURNAL INFOTEL*, vol. 15, no. 4, pp. 317–325, 2023, doi: 10.20895/infotel.v15i4.1030.
- [18] H. J. Park, Y. S. Koo, H. Y. Yang, Y. S. Han, and C. S. Nam, “Study on Data Preprocessing for Machine Learning Based on Semiconductor Manufacturing Processes,” *Sensors*, vol. 24, no. 17, 2024, doi: 10.3390/s24175461.
- [19] R. Syahri and D. Puspita, “Classification of outstanding students using support vector machine (SVM) based on data mining,” *Journal of Information Technology and Engineering*, vol. 9, no. 1, pp. 13–23, 2025, doi: 10.31289/jite.v9i1.13191.
- [20] R. Oktafiani, A. Hermawan, and D. Avianto, “Pengaruh komposisi split data terhadap performa klasifikasi penyakit kanker payudara menggunakan algoritma machine learning,” *Jurnal Sains dan Informatika*, vol. 9, no. 1, pp. 19–28, 2023, doi: 10.34128/jsi.v9i1.622.
- [21] R. Alfiyan, M. Hikmatyar, and S. S. Sundari, “Implementasi metode klasifikasi C4.5 penyebab faktor risiko penyakit stroke,” *Indonesian Journal of Digital Business*, vol. 4, no. 2, pp. 37–48, 2024, doi: 10.17509/ijdb.v4i2.69221.
- [22] P. J. B. Pajila, B. G. Sheena, and A. Gayathri, “A comprehensive survey on naive bayes algorithm: advantages, limitations and applications,” in *Proc. International Conference on Smart Electronics and Communication (ICOSEC)*, 2023, pp. 1227–1233. doi: 10.1109/ICOSEC58147.2023.10276274.
- [23] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, “An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review,” *Information (Switzerland)*, vol. 15, no. 4, 2024, doi: 10.3390/info15040235.
- [24] R. Nanda, E. Haerani, S. K. Gusti, and S. Ramadhani, “Klasifikasi berita menggunakan metode support vector machine,” *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 2, pp. 269–278, 2022, doi: 10.32672/jnkti.v5i2.4193.
- [25] R. Ridwan, H. H. Handayani, S. A. P. Lestari, and Y. Cahyana, “Evaluasi kinerja algoritma random forest dan gradient boosting untuk klasifikasi penyakit jantung,” *Jurnal Komtika (Komputasi dan Informatika)*, vol. 9, no. 1, pp. 112–124, 2025, doi: 10.31603/komtika.v9i1.13450.