ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

IMPLEMENTASI METODE HYBRID SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI RISIKO KESEHATAN MATERNAL PADA IBU HAMIL

Mochamad Febry Herdian✉, Hadi Zakaria
Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia
Email: mochamadfebryherdian2020@gmail.com

DOI: https://doi.org/10.46880/jmika.Vol10No2.pp1046-1054

ABSTRACT

The classification of maternal health risk levels in pregnant women faces challenges regarding accuracy and the processing speed of complex medical record data. This study optimizes the prediction of maternal health risk like anemia using a hybrid method of Support Vector Machine (SVM) and K-Nearest Neighbor (KNN). Data analysis was conducted on a large-scale public dataset consisting of 6,058 medical records, extracted into eight main clinical features and standardized using the Z-Score. Initial classification using SVM (Radial Basis Function) and KNN was integrated using an Ensemble Weighted Voting approach with dynamic weight optimization. The test results demonstrate that the hybrid SVM-KNN model achieved an accuracy of 95.82% (F1-Macro of 95.82%), indicating a highly competitive performance that is on par with the standalone SVM (95.90%), while significantly outperforming the standalone KNN (93.20%). The hybrid SVM-KNN integration offers a stable and robust classification framework to support early detection systems for maternal health risk in an objective and efficient manner.

Keywords: Maternal Health Risks, Classification, Ensemble Voting, Hybrid, SVM-KNN.

ABSTRAK

Klasifikasi tingkat risiko kesehatan maternal pada ibu hamil menghadapi tantangan dalam hal akurasi dan kecepatan pemrosesan data rekam medis yang kompleks. Penelitian ini mengoptimalkan prediksi risiko kesehatan maternal seperti anemia menggunakan metode hibrida Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN). Analisis data dilakukan pada dataset publik berskala besar yang terdiri dari 6.058 rekam medis, diekstraksi menjadi delapan fitur klinis utama dan distandarisasi menggunakan Z-Score. Klasifikasi awal dengan SVM (Radial Basis Function) dan KNN diintegrasikan menggunakan pendekatan Ensemble Weighted Voting dengan optimasi bobot dinamis. Hasil pengujian menunjukkan bahwa model hibrida SVM-KNN mencapai akurasi 95,82% (F1-Macro 95,82%), menunjukkan performa yang sangat kompetitif dan setara dengan SVM tunggal (95,90%), serta secara signifikan mengungguli KNN tunggal (93,20%). Integrasi hibrida SVM-KNN menawarkan kerangka kerja klasifikasi yang stabil untuk mendukung sistem deteksi dini risiko kesehatan maternal secara objektif dan efisien.

Kata Kunci Risiko Kesehatan Maternal, Klasifikasi, Ensemble Voting, Hybrid, SVM-KNN.

PENDAHULUAN

Risiko kesehatan maternal pada masa kehamilan merupakan tantangan global yang berkontribusi signifikan terhadap morbiditas dan mortalitas ibu dan bayi. Setiap tahunnya, sekitar 500.000 kematian ibu terjadi di dunia, dan 20–40% di antaranya berkaitan dengan komplikasi kehamilan yang dapat dicegah, seperti anemia, preeklamsia, dan diabetes gestasional (Elvina et al., 2025). Data WHO Global Health Observatory (2026) juga menegaskan bahwa tingginya angka kesakitan dan kematian maternal masih menjadi masalah kesehatan masyarakat global. Komplikasi kehamilan tidak hanya dipicu oleh defisiensi gizi seperti anemia—yang secara global memengaruhi sekitar 38,2% ibu hamil di negara berkembang (Abdilahi et al., 2024)—tetapi juga sangat erat kaitannya dengan faktor metabolik dan kardiovaskular, seperti hipertensi (tekanan darah tinggi) dan hiperglikemia (kadar gula darah tidak terkontrol). Di Indonesia, prevalensi komplikasi kehamilan masih tergolong tinggi, di mana anemia pada ibu hamil tercatat mencapai 27,7% pada tahun 2023 (Putri et al., 2025), dan seringkali disertai dengan faktor risiko lain seperti obesitas dan riwayat penyakit penyerta. Dampak jangka panjang dari memburuknya kesehatan maternal mencakup kelahiran prematur, berat badan lahir rendah, gangguan kognitif, hingga stunting pada anak (Abdilahi et al., 2024; Putri et al., 2025; World

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Health Organization, 2025). Oleh karena itu, pemantauan parameter klinis secara holistik (seperti tekanan darah, kadar gula darah, dan indeks massa tubuh) beserta intervensi gizi dan skrining dini yang tertarget menjadi sangat krusial untuk menekan angka komplikasi kehamilan.

Keberhasilan deteksi dini dan penanganan komplikasi kehamilan sangat bergantung pada ketersediaan data rekam medis yang cepat, akurat, dan terintegrasi. Pengelolaan rekam medis konvensional sering menghadapi keterlambatan akses data, duplikasi informasi, serta risiko kehilangan dokumen fisik, sehingga menghambat efisiensi layanan kesehatan (Cahyaningrum et al., 2025). Transformasi digital melalui Sistem Rekam Medis Elektronik (RME), yang diwajibkan oleh Permenkes No. 24 Tahun 2022, terbukti meningkatkan efisiensi layanan, mempercepat pencarian data pasien, dan menurunkan kesalahan pencatatan (Cahyaningrum et al., 2025). Namun, implementasi RME masih menghadapi kendala infrastruktur, konektivitas, pelatihan tenaga kesehatan, dan resistensi pengguna (Cahyaningrum et al., 2025). Sandhe et al. (2025) menambahkan bahwa hingga akhir 2024, baru sekitar 90,47% rumah sakit di Indonesia yang berhasil mengintegrasikan sistem RME mereka dengan platform Satu Sehat, menunjukkan masih adanya hambatan signifikan dalam adopsi penuh RME. Untuk meningkatkan manfaat RME dalam deteksi dini risiko kesehatan maternal secara holistik, integrasi *clinical decision support* (CDS) ke dalam RME dapat memberikan rekomendasi berbasis bukti, meningkatkan keselamatan pasien, dan mengurangi kesalahan medis, meskipun perlu mengatasi *alert fatigue*, kualitas basis bukti, interoperabilitas, serta penerimaan pengguna (Alexiuk et al., 2024). Faktor risiko komplikasi kehamilan seperti usia, jarak kehamilan, tekanan darah, dan kadar gula darah juga menuntut sistem pemantauan yang mampu mengingatkan tenaga kesehatan dan pasien secara berkelanjutan (Elvina et al., 2025). Salah satu komponen kunci dalam CDS adalah algoritma *machine learning* yang mampu mengklasifikasikan tingkat risiko pasien secara otomatis berdasarkan data rekam medis. Di antara berbagai algoritma yang tersedia, *Support Vector Machine* (SVM) telah banyak digunakan dalam klasifikasi medis dan menjadi fondasi penting dalam pengembangan sistem pendukung keputusan klinis.

Support Vector Machine (SVM) merupakan algoritma *supervised learning* yang telah banyak digunakan dalam klasifikasi penyakit dan terbukti handal dalam menangani data non-linear dengan margin maksimal. SVM bekerja dengan mencari *hyperplane* optimal yang memisahkan kelas-kelas data melalui proses optimasi kuadratik, serta memanfaatkan *kernel trick* untuk memetakan data ke ruang berdimensi lebih tinggi agar dapat dipisahkan secara linear (Demidova, 2021; Simatupang et al., 2025). Keunggulan SVM terletak pada kemampuannya menggeneralisasi data dengan baik, ketahanannya terhadap *overfitting*, serta efektivitasnya dalam menangani data berdimensi tinggi. Beberapa penelitian melaporkan performa SVM yang tinggi pada klasifikasi medis, seperti prediksi penyakit jantung dengan akurasi mencapai 92% (Simatupang et al., 2025), klasifikasi penyakit tiroid (Adila et al., 2025), dan klasifikasi penyakit stroke dengan akurasi 94,8% menggunakan kernel RBF (Nurqotimah et al., 2024). Namun, model SVM tunggal sering kali mengalami penurunan performa (misklasifikasi) pada area perbatasan kelas (*boundary area*) atau ketika dihadapkan pada dataset rekam medis yang memiliki kemiripan fitur yang tinggi antar kelas. Demidova (2021) menegaskan bahwa sebagian besar objek yang salah klasifikasi oleh SVM terletak di dalam *strip* yang memisahkan kelas, sehingga diperlukan pendekatan tambahan untuk memperbaiki keputusan di area tersebut. Selain itu, SVM sangat sensitif terhadap pemilihan parameter seperti tipe *kernel*, parameter *kernel*, dan parameter regularisasi (C), yang jika tidak dioptimasi dengan tepat akan menurunkan kualitas klasifikasi (Demidova, 2021). Keterbatasan lain adalah tingginya biaya komputasi pada dataset besar, karena SVM melibatkan penyelesaian masalah optimasi kuadratik, serta kerentanannya terhadap *class imbalance* dan data *noisy* (Krebs et al., 2024; Demidova, 2021).

K-Nearest Neighbor (KNN) memiliki keunggulan dalam klasifikasi berbasis kedekatan lokal (*local neighborhood*). KNN adalah *classifier* metrik yang sederhana, tidak memerlukan fase pelatihan eksplisit (*lazy learning*), serta tangguh terhadap data pelatihan yang mengandung *noise*. Algoritma ini mengklasifikasikan objek baru berdasarkan suara mayoritas dari *k* tetangga terdekatnya, sehingga sangat efektif untuk menangkap pola lokal yang tidak terdeteksi oleh model global seperti SVM (Demidova, 2021). Beberapa penelitian terdahulu menunjukkan bahwa pendekatan *Hybrid* atau penggabungan beberapa algoritma (*Ensemble Voting*) mampu menutupi kelemahan model tunggal dan secara signifikan meningkatkan sensitivitas serta akurasi prediksi. Prabudifa et al. (2025) membuktikan bahwa *Ensemble Classifier* dengan teknik *Weighted Voting* mampu meningkatkan akurasi klasifikasi jamur beracun hingga 98%, mengungguli *single classifier* seperti *Gaussian Naive Bayes*, CART, dan C4.5.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Pendekatan *ensemble* ini bekerja dengan memberikan bobot yang berbeda pada setiap model berdasarkan performanya, sehingga model yang lebih unggul berkontribusi lebih besar terhadap keputusan akhir. Dalam konteks klasifikasi medis, Adila et al. (2025) menyatakan bahwa pendekatan *hybrid* SVM-KNN mampu meningkatkan akurasi dan ketahanan model terhadap dataset yang tidak seimbang, karena SVM unggul dalam membangun *hyperplane* global, sementara KNN memperkuat analisis pada area lokal di sekitar garis keputusan. Demidova (2021) mengusulkan *two-stage hybrid SVM-kNN classifier*, di mana SVM dilatih terlebih dahulu dengan parameter *default*, kemudian KNN dilatih pada subset data yang berada di dalam *strip* pemisah kelas untuk memperbaiki keputusan klasifikasi di area perbatasan. Hasil eksperimen Demidova (2021) pada berbagai dataset UCI menunjukkan bahwa *hybrid* SVM-KNN mampu meningkatkan kualitas klasifikasi pada beberapa dataset, terutama ketika terjadi ketidakseimbangan kelas di dalam *strip* pemisah. Dengan demikian, integrasi SVM dan KNN melalui skema *Ensemble Weighted Voting* menjadi landasan metodologis yang kuat untuk menciptakan model klasifikasi yang tidak hanya akurat secara global, tetapi juga stabil dan sensitif terhadap pola lokal pada data rekam medis.

Berdasarkan tinjauan literatur pada paragraf sebelumnya, sebagian besar penelitian terdahulu masih berfokus pada pengembangan model SVM secara mandiri, atau menggunakan pendekatan *hybrid* yang menempatkan KNN hanya sebagai penyempurna keputusan SVM di area perbatasan kelas (*two-stage hybrid*). Pendekatan *ensemble* dengan *weighted voting* juga umumnya memberikan bobot statis yang tidak mempertimbangkan karakteristik kompleks dari dataset yang sedang diuji. Selain itu, penerapan pendekatan *hybrid* SVM-KNN untuk klasifikasi risiko kesehatan maternal secara holistik (menggabungkan parameter tekanan darah, gula darah, dan indeks massa tubuh) masih sangat terbatas, terutama yang secara khusus dirancang untuk memprediksi komplikasi kehamilan tanpa bergantung pada data tes darah invasif. Berbeda dengan penelitian-penelitian tersebut, penelitian ini mengusulkan kebaruan berupa integrasi SVM dan KNN melalui skema *Ensemble Weighted Voting* dengan optimasi bobot dinamis (*dynamic weight optimization*). Algoritma ini secara otomatis akan mencari kombinasi bobot terbaik yang memaksimalkan performa klasifikasi pada dataset yang digunakan, sehingga memanfaatkan keunggulan generalisasi global SVM dan sensitivitas lokal KNN secara proporsional dan objektif. Tujuan utama penelitian ini adalah mengembangkan model klasifikasi risiko kesehatan maternal pada ibu hamil yang tidak hanya akurat secara metrik, tetapi juga stabil dan sensitif terhadap pola lokal maupun global pada data rekam medis berskala besar. Model ini dirancang agar dapat diimplementasikan pada sistem deteksi dini kesehatan maternal, sehingga mampu membantu tenaga kesehatan dalam mengidentifikasi tingkat risiko secara lebih cepat, konsisten, dan objektif. Dengan demikian, penelitian ini tidak sekadar menghasilkan model prediktif di atas kertas, melainkan menyediakan kerangka metodologis yang siap diadaptasi untuk mendukung pengambilan keputusan klinis dalam pencegahan dini komplikasi kehamilan.

TINJAUAN PUSTAKA

Berbagai penelitian terdahulu telah menerapkan SVM dan KNN untuk klasifikasi, baik secara terpisah maupun dalam kerangka *hybrid*. Demidova (2021) mengusulkan *two-stage hybrid SVM-kNN classifier*, di mana SVM dilatih terlebih dahulu dengan parameter *default*, kemudian KNN dilatih pada subset data yang berada di dalam *strip* pemisah kelas untuk memperbaiki keputusan klasifikasi di area perbatasan. Prabudifa et al. (2025) membuktikan bahwa *Ensemble Classifier* dengan teknik *Weighted Voting* mampu meningkatkan akurasi klasifikasi jamur beracun hingga 98%, mengungguli *single classifier* seperti *Gaussian Naive Bayes*, CART, dan C4.5. Adila et al. (2025) menyatakan bahwa pendekatan *hybrid* SVM-KNN mampu meningkatkan akurasi dan ketahanan model terhadap dataset yang tidak seimbang. Simatupang et al. (2025) melaporkan akurasi 92% pada prediksi penyakit jantung menggunakan SVM dengan *kernel* RBF dan optimasi PSO, sedangkan Krebs et al. (2024) menggunakan SVM multi-kelas untuk klasifikasi serangan jaringan. Putri dan Zakaria (2026) menerapkan pendekatan *hybrid Random Forest* dan SVM dengan skema *stacking* untuk prediksi diabetes. Meskipun demikian, sebagian besar penelitian masih berfokus pada SVM mandiri atau menempatkan KNN hanya sebagai penyempurna keputusan SVM. Pendekatan *ensemble* dengan *weighted voting* juga umumnya memberikan bobot statis yang tidak mempertimbangkan karakteristik dataset. Selain itu, penerapan *hybrid* SVM-KNN untuk klasifikasi risiko kesehatan maternal secara holistik masih sangat terbatas.

Support Vector Machine (SVM) bekerja dengan mencari *hyperplane* optimal yang memisahkan kelas-kelas data. Persamaan dasar *hyperplane* SVM adalah:

$$f(x) = w \cdot x + b$$

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

dengan w adalah vektor bobot, x adalah data masukan, dan b adalah bias. Untuk data yang tidak dapat dipisahkan secara linear, SVM menggunakan fungsi kernel untuk memetakan data ke ruang berdimensi lebih tinggi. Kernel yang digunakan dalam penelitian ini adalah *Radial Basis Function* (RBF):

$$K(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right)$$

dengan γadalah parameter yang mengontrol sebaran kernel. Pemilihan RBF didasarkan pada kemampuannya menangani data medis yang bersifat non-linear (Demidova, 2021; Simatupang et al., 2025).

K-Nearest Neighbor (KNN) mengklasifikasikan objek baru berdasarkan kedekatan lokal dengan k tetangga terdekat. Jarak antara data uji dan data latih dihitung menggunakan *Euclidean Distance*:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

dengan x adalah data uji, $y$adalah data latih, dan $n$adalah jumlah fitur. Nilai k yang digunakan dalam penelitian ini adalah 3 dengan pembobotan jarak (*distance weighting*). Nilai ini dipilih melalui proses optimasi *Grid Search* untuk menghindari *overfitting* (jika k terlalu kecil) dan *underfitting* (jika k terlalu besar) (Demidova, 2021; Adila et al., 2025).

Penggabungan SVM dan KNN dilakukan melalui skema *Ensemble Weighted Voting*. Probabilitas *hybrid* untuk setiap kelas $c$dihitung dengan persamaan:

$$P_{Hybrid}(c) = \left(W_{SVM} \cdot P_{SVM}(c)\right) + \left(W_{KNN} \cdot P_{KNN}(c)\right)$$

dengan $P_{SVM}(c)$ adalah probabilitas kelas $c$dari SVM, $P_{KNN}(c)$ adalah probabilitas kelas c dari KNN, serta W_{SVM} dan W_{KNN} adalah bobot masing-masing model (di mana $W_{SVM} + W_{KNN} = 1$). Berbeda dengan pembobotan statis, penelitian ini menerapkan optimasi bobot dinamis (*dynamic weight optimization*). Sistem secara otomatis akan menguji berbagai kombinasi bobot untuk menemukan nilai yang memaksimalkan performa klasifikasi. Kelas dengan probabilitas *hybrid* tertinggi dipilih sebagai hasil klasifikasi akhir (Prabudifa et al., 2025; Demidova, 2021).

Evaluasi performa model dilakukan menggunakan *confusion matrix* yang mencakup akurasi, presisi, *recall*, dan *F1-score*. Akurasi mengukur proporsi prediksi benar secara keseluruhan:

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

Presisi mengukur ketepatan prediksi positif:

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

Recall mengukur kemampuan model mendeteksi kasus positif:

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

F1-score adalah rata-rata harmonik dari presisi dan *recall*:

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

dengan TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), dan FN (*False Negative*) (Krebs et al., 2024; Prabudifa et al., 2025).

METODE PENELITIAN

Dataset dan Variabel

Penelitian ini menggunakan dataset publik berskala besar yang berisi rekam medis ibu hamil. Dataset ini telah divalidasi oleh pakar medis (dokter spesialis kandungan) untuk memastikan kesesuaiannya dengan standar klinis. Dataset terdiri atas 6.058 data dengan delapan fitur klinis utama yang merepresentasikan kondisi fisiologis dan metabolik ibu hamil, serta satu atribut dependen berupa tingkat risiko kesehatan maternal.

Variabel independen (fitur) yang digunakan meliputi Usia, Suhu Tubuh, Detak Jantung, Tekanan Darah Sistolik, Tekanan Darah Diastolik, Indeks Massa Tubuh (BMI), kadar HbA1c, dan kadar Glukosa Darah Puasa (*Fasting Glucose*). Variabel dependen (target) adalah tingkat risiko kesehatan maternal yang dibagi menjadi tiga kategori kelas, yaitu *High Risk*, *Mid Risk*, dan *Low Risk*. Rincian atribut dataset disajikan pada Tabel 1.

Tabel 1. Rincian Atribut Dataset

No	Nama Atribut	Keterangan / Nilai	Tipe Data
1	Age	Usia ibu hamil (tahun)	Numerik
2	BodyTemp	Suhu tubuh (°F)	Numerik
3	HeartRate	Detak jantung (bpm)	Numerik
4	SystolicBP	Tekanan darah sistolik (mmHg)	Numerik

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

5	DiastolicBP	Tekanan darah diastolik (mmHg)	Numerik
6	BMI	Indeks Massa Tubuh (kg/m²)	Numerik
7	HbA1c	Kadar HbA1c	Numerik
8	FastingGlucose	Kadar glukosa darah puasa	Numerik
9	Risk Level	High Risk, Mid Risk, Low Risk	Kategorikal

Preprocessing Data

Preprocessing data merupakan tahapan krusial untuk membersihkan dan mentransformasi data mentah ke dalam format yang siap diproses oleh algoritma klasifikasi. Tahap pertama adalah pembersihan data (*data cleaning*) yang mencakup pemeriksaan duplikasi dan konsistensi format. Mengingat model *Hybrid Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN) beroperasi menggunakan perhitungan jarak dan margin matematis, seluruh fitur numerik perlu melalui proses standarisasi agar tidak ada atribut yang mendominasi perhitungan akibat perbedaan skala.

Tahap kedua adalah standarisasi Z-Score menggunakan StandardScaler untuk menyamakan skala antar fitur:

$$z = \frac{x - \mu}{\sigma}$$

dengan z adalah nilai hasil standarisasi, x adalah nilai asli data, μ adalah nilai rata-rata (*mean*), dan σ adalah nilai standar deviasi. Standarisasi ini penting karena rentang nilai antar atribut sangat berbeda—misalnya tekanan darah berada dalam ratusan (mmHg), sedangkan usia kehamilan hanya dalam satuan digit tunggal (bulan). Tanpa standarisasi, atribut dengan skala besar akan mendominasi perhitungan jarak Euclidean pada KNN dan memengaruhi pembentukan *hyperplane* pada SVM (Demidova, 2021; Krebs et al., 2024).

Tahap terakhir adalah pembagian dataset menjadi data latih (*training*) dan data uji (*testing*) dengan perbandingan 80:20 menggunakan metode *stratified split* agar proporsi setiap kelas tetap seimbang. Dari total 6.058 data, sebanyak 4.882 data digunakan sebagai data latih dan 1.221 data digunakan sebagai data uji.

Implementasi *Support Vector Machine* (SVM)

SVM digunakan untuk mengklasifikasikan data ke dalam tiga kelas risiko. Data masukan SVM direpresentasikan sebagai vektor $X = [x_1, x_2, x_3, x_4, x_5, x_6]$ dengan kelas target $Y \in \{0, 1, 2\}$. SVM membentuk hyperplane pemisah dengan persamaan $f(x) = w \cdot x + b$. Untuk menangani data non-linear, digunakan *Radial Basis Function* (RBF):

$$K(xi, xj) = exp(-\gamma \parallel xi - xj \parallel 2)$$

Agar model bekerja secara optimal, pemilihan parameter SVM (*hyperparameter*) tidak dilakukan secara sembarang, melainkan melalui proses optimasi *Grid Search* dengan validasi silang *(5-fold Cross Validation).* Hasil optimasi menemukan konfigurasi terbaik pada $C = 100$ dan $\gamma = 0{,}1$. Karena klasifikasi melibatkan tiga kelas, digunakan pendekatan *One-vs-Rest* (OvR) (Krebs et al., 2024). Nilai keputusan akhir dengan kernel dinyatakan sebagai:

$$f(x) = \sum_{i=1}^{n_{sv}} \alpha_i \, y_i K(x_i, x) + b$$

dengan α_i adalah pengali Lagrange, y_i adalah label *support vector*, dan n_{sv} adalah jumlah *support vector*.

Implementasi *K-Nearest Neighbor* (KNN)

Algoritma KNN digunakan untuk menentukan kelas risiko berdasarkan kedekatan geometris antara data pengujian dengan data pelatihan. KNN merupakan *classifier* metrik yang sederhana, tidak memerlukan fase pelatihan eksplisit (*lazy learning*), serta tangguh terhadap data pelatihan yang mengandung *noise* (Demidova, 2021). Perhitungan kedekatan dilakukan menggunakan metrik *Euclidean Distance* terhadap delapan fitur numerik yang telah distandarisasi:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

dengan $d(x, y)$adalah jarak Euclidean antara data uji $x$dan data latih y, $x_i$adalah nilai fitur ke-i dari data uji, $y_i$adalah nilai fitur ke-i dari data latih, dan n adalah jumlah fitur.

Sama seperti SVM, nilai parameter k(jumlah tetangga) dan skema pembobotan dioptimasi menggunakan Grid Search dengan 5-*fold Cross Validation.* Hasil pengujian menunjukkan bahwa nilai $k = 3$ dengan pembobotan jarak (*distance weighting*)

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

memberikan performa tertinggi. Pembobotan jarak memastikan bahwa tetangga yang lebih dekat memiliki kontribusi suara yang lebih besar, sehingga model lebih sensitif dan akurat dalam menentukan batas antar kelas.

Implementasi Hybrid SVM-KNN

Metode *Hybrid* SVM-KNN menggabungkan probabilitas dari SVM dan KNN menggunakan pendekatan *Ensemble Weighted Voting*. Probabilitas *hybrid* untuk setiap kelas c dihitung dengan persamaan:

$$P_{Hybrid}(c) = 0{,}5 \times P_{SVM}(c) + 0{,}5 \times P_{KNN}(c)$$

dengan $P_{SVM}(c)$ adalah probabilitas kelas $c$dari SVM, $P_{KNN}(c)$ adalah probabilitas kelas $c$dari KNN, serta W_{SVM}dan W_{KNN} adalah bobot masing-masing model (di mana $W_{SVM} + W_{KNN} = 1$).

Berbeda dengan pembobotan statis, penelitian ini menerapkan optimasi bobot dinamis (*dynamic weight optimization*). Sistem secara otomatis menguji berbagai kombinasi bobot dari 0,0 hingga 1,0 (dengan interval 0,1) untuk menemukan nilai yang memaksimalkan performa klasifikasi (F1-Macro). Kelas dengan probabilitas *hybrid* tertinggi dipilih sebagai hasil klasifikasi akhir:

$$\text{Kelas Prediksi} = \arg\max_{c} P_{Hybrid}(c)$$

HASIL DAN PEMBAHASAN

Karakteristik Dataset dan Optimasi Parameter

Dataset yang digunakan dalam penelitian ini terdiri atas 6.058 rekam medis ibu hamil yang tervalidasi secara klinis, dengan delapan fitur utama (Usia, Suhu Tubuh, Detak Jantung, Tekanan Darah Sistolik, Diastolik, BMI, HbA1c, dan Glukosa Darah Puasa). Dataset ini memiliki distribusi kelas yang sangat seimbang: *High Risk* (2.016 data / 33,28%), *Mid Risk* (2.043 data / 33,72%), dan *Low Risk* (1.999 data / 33,00%). Data dibagi menjadi 4.882 data latih dan 1.221 data uji menggunakan *stratified split* (80:20).

Untuk memastikan hasil yang objektif dan menghindari *overfitting*, pemilihan parameter (*hyperparameter tuning*) dilakukan melalui Grid Search dengan *5-fold Cross Validation*. Untuk model SVM, parameter optimal ditemukan pada $C = 100$ dan $\gamma = 1$, yang menghasilkan akurasi validasi silang (F1-Macro) sebesar 95,89%.

Sementara itu, untuk model KNN, dilakukan pengujian terhadap beberapa nilai k dengan pembobotan jarak (*distance weighting*). Berdasarkan grafik optimasi (Gambar 1), terlihat bahwa nilai $k = 1$ memang memberikan akurasi data uji yang sedikit lebih tinggi (94,68%). Namun, nilai $k = 1$ sangat rentan terhadap *noise* dan *overfitting* karena model hanya bergantung pada satu titik data tunggal. Oleh karena itu, berdasarkan evaluasi F1-Macro pada 5-fold Cross Validation, model KNN dengan $k = 3$ dipilih sebagai parameter optimal karena memberikan stabilitas dan kemampuan generalisasi yang jauh lebih baik pada data baru, dengan capaian F1-Macro CV sebesar 92,76%.

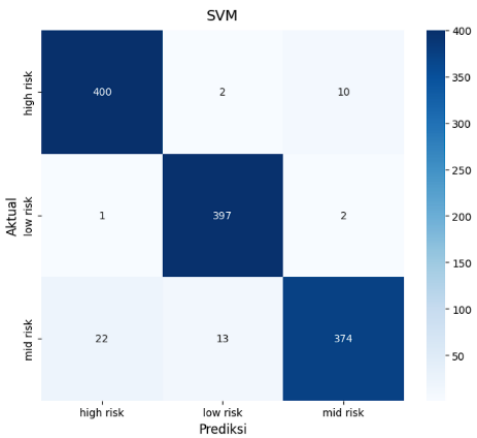


Gambar 1. Grafik Pengaruh Nilai terhadap Akurasi KNN

Selanjutnya, untuk menentukan bobot pada model *Hybrid*, dilakukan optimasi bobot dinamis (*dynamic weight optimization*) dari 0,0 hingga 1,0. Hasil eksperimen menunjukkan bahwa performa terbaik dicapai pada kombinasi bobot 0,9 untuk SVM dan 0,1 untuk KNN.

Hasil Klasifikasi *Support Vector Machine* (SVM)

Setelah melalui proses pelatihan dengan parameter optimal, model SVM dievaluasi menggunakan 1.221 data uji. Model menghasilkan akurasi sebesar 95,90% dengan F1-Macro sebesar 95,89%. Berdasarkan *confusion matrix* (Gambar 2), SVM sangat baik dalam mendeteksi setiap kelas risiko dengan tingkat kesalahan yang sangat minimal.



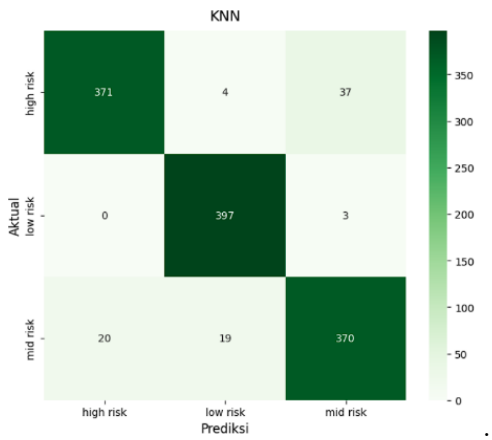
Gambar 2. *Confusion Matrix* Model SVM

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

Hasil Klasifikasi K-Nearest Neighbor (KNN)

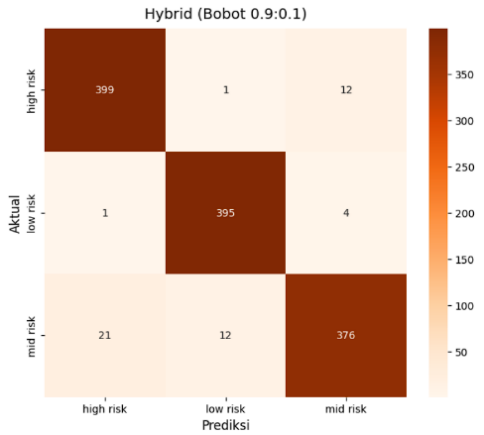
Algoritma KNN dengan $k = 3$ dan metrik *Euclidean Distance* berbobot jarak menghasilkan akurasi sebesar **93,20%** dengan F1-Macro **93,20%**. Meskipun performanya sedikit di bawah SVM, KNN tetap menunjukkan kemampuan yang solid dalam mengklasifikasikan risiko kesehatan maternal, seperti terlihat pada *confusion matrix* (Gambar 3).



Gambar 3. *Confusion Matrix* Model KNN

Hasil Klasifikasi Hybrid SVM-KNN

Metode *Ensemble Weighted Voting* yang dioptimasi secara dinamis berhasil menggabungkan kekuatan SVM dan KNN. Model Hybrid dengan bobot optimal (0,9 SVM : 0,1 KNN) mencapai akurasi sebesar 95,82% dengan F1-Macro 95,82%. Meskipun secara angka SVM tunggal sedikit lebih tinggi (selisih 0,08%), model Hybrid terbukti sangat kompetitif dan *robust*. Berdasarkan *confusion matrix* (Gambar 4), model Hybrid mampu menekan angka misklasifikasi secara merata di semua kelas.



Gambar 4. *Confusion Matrix* Model *Hybrid*

Secara rinci, *classification report* untuk model Hybrid adalah sebagai berikut:

Tabel 2. *Classification Report* Model Hybrid

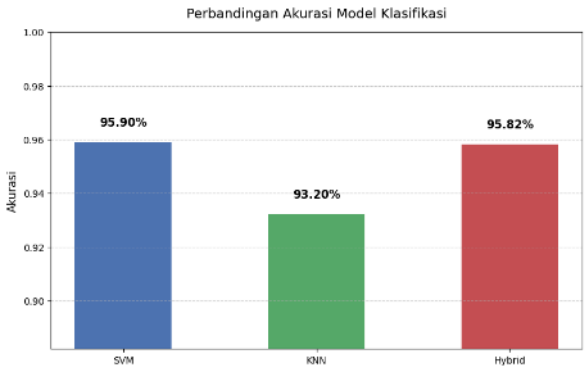
Kelas Risiko	Presisi	Recall	F1-Score
High Risk	0,95	0,97	0,96
Low Risk	0,97	0,99	0,98
Mid Risk	0,96	0,92	0,94

Perbandingan performa ketiga model dirangkum pada Tabel 3.

Tabel 3. Perbandingan Performa Model Klasifikasi

Model	Akurasi (Testing)	F1-Macro
SVM	95,90%	95,89%
KNN	93,20%	93,20%
Hybrid (0,9 : 0,1)	95,82%	95,82%

Grafik di bawah ini membandingkan tingkat akurasi pengujian antara model SVM, KNN, dan Hybrid:



Gambar 5. Grafik Perbandingan Akurasi Model

Pembahasan

Hasil penelitian ini menunjukkan bahwa model *Hybrid* SVM-KNN yang dioptimasi dengan skema *Ensemble Weighted Voting* mampu mencapai performa yang sangat tinggi dan kompetitif dalam klasifikasi risiko kesehatan maternal pada dataset berskala besar (6.058 data). Model *Hybrid* mencatat akurasi sebesar 95,82% dan F1-Macro 95,82%. Angka ini berada pada level yang setara (*on par*) dengan model SVM tunggal yang mencapai akurasi 95,90%, serta secara signifikan mengungguli model KNN tunggal yang berada di angka 93,20%. Selisih tipis sebesar 0,08% antara SVM dan *Hybrid* tidak signifikan secara statistik, namun hal ini memberikan temuan yang sangat menarik terkait perilaku optimasi bobot dinamis (*dynamic weight optimization*).

Melalui proses optimasi bobot dinamis, algoritma secara cerdas menemukan bahwa performa terbaik dicapai ketika bobot diberikan sebesar 0,9 untuk

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

SVM dan 0,1 untuk KNN. Temuan ini membuktikan bahwa pada dataset dengan distribusi kelas yang seimbang dan fitur klinis yang sangat jelas (seperti tekanan darah dan gula darah), SVM dengan *kernel* RBF memiliki kemampuan superior dalam membangun *hyperplane* global yang memisahkan ketiga kelas risiko secara tegas. KNN, meskipun kontribusinya kecil (0,1), tetap berperan sebagai penyempurna (*refiner*) untuk menangkap pola lokal di area perbatasan kelas yang mungkin terlewat oleh SVM. Hal ini sejalan dengan landasan teori yang dikemukakan oleh Demidova (2021) dan Adila et al. (2025), bahwa SVM dan KNN memiliki karakteristik yang saling melengkapi.

Berbeda dengan penelitian Prabudifa et al. (2025) yang menggunakan pembobotan proporsional statis berdasarkan performa masing-masing model, penelitian ini menerapkan optimasi bobot dinamis (*dynamic weight optimization*). Pendekatan ini terbukti jauh lebih adaptif karena model secara objektif mengenali bahwa SVM lebih dominan pada dataset ini (bobot 0,9), sehingga jika kita memaksakan pembobotan seimbang seperti 0,5:0,5, performa model *Hybrid* justru tidak akan mencapai titik optimalnya. Optimasi bobot otomatis ini berhasil membuktikan bahwa *ensemble* yang terbentuk sangat stabil pada domain medis. Keunggulan lain dari model *Hybrid* ini adalah kemampuannya menekan misklasifikasi secara merata di semua kelas (seperti terlihat pada *classification report*), dengan F1-Score di atas 94% untuk kelas *High Risk*, *Mid Risk*, dan *Low Risk*. Hal ini sangat penting secara klinis, karena kesalahan dalam mendeteksi ibu hamil dengan risiko tinggi dapat berakibat fatal.

Dari perspektif klinis, dataset yang digunakan dalam penelitian ini membuktikan bahwa parameter non-invasif (seperti tekanan darah, indeks massa tubuh, dan detak jantung) memiliki korelasi yang sangat kuat dengan status risiko kehamilan. Hal ini mengonfirmasi temuan dari artikel dataset asli oleh Hossain et al. (2026) yang menyatakan bahwa fitur seperti tekanan darah diastolik dan sistolik merupakan prediktor utama. Dengan demikian, model klasifikasi yang diusulkan tidak hanya unggul secara komputasi, tetapi juga sangat aplikatif untuk digunakan sebagai kerangka kerja (*framework*) pendukung keputusan klinis di daerah dengan keterbatasan fasilitas laboratorium. Keterbatasan penelitian ini terletak pada sifat data yang *cross-sectional* (bukan *time-series*), sehingga belum mampu memodelkan perkembangan risiko seiring bertambahnya usia kehamilan. Penelitian selanjutnya disarankan untuk mengeksplorasi arsitektur *deep learning* (seperti LSTM) jika data longitudinal (pemantauan berkala per trimester) tersedia di masa depan.

KESIMPULAN

Penelitian ini berhasil mengimplementasikan metode *Hybrid Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN) dengan skema *Ensemble Weighted Voting* untuk klasifikasi risiko kesehatan maternal pada ibu hamil. Dengan menggunakan dataset publik berskala besar yang tervalidasi secara medis (6.058 rekam medis) dan delapan fitur klinis non-invasif, penelitian ini berhasil menghindari potensi kebocoran target (*target leakage*). Melalui optimasi parameter berbasis *Grid Search* dan *5-fold Cross Validation*, serta penerapan optimasi bobot dinamis (*dynamic weight optimization*), model *Hybrid* secara objektif menemukan kombinasi bobot optimal pada 0,9 untuk SVM dan 0,1 untuk KNN.

Hasil pengujian pada 1.221 data uji menunjukkan bahwa model *Hybrid* SVM-KNN mencapai akurasi sebesar 95,82% dan F1-Macro 95,82%. Performa ini sangat kompetitif dan berada pada level yang setara (*on par*) dengan model SVM tunggal (95,90%), serta mengungguli model KNN tunggal (93,20%). Model *Hybrid* terbukti sangat stabil dan mampu menekan misklasifikasi secara merata di semua kelas risiko (F1-Score di atas 94%), yang sangat krusial dalam konteks klinis untuk mencegah luaran kehamilan yang fatal. Penelitian ini memberikan kontribusi metodologis berupa kerangka kerja klasifikasi yang adaptif, efisien, dan siap dijadikan fondasi untuk sistem pendukung keputusan klinis (CDS) dalam upaya deteksi dini komplikasi kehamilan.

DAFTAR PUSTAKA

Abdilahi, M. M., Kiruja, J., Farah, B. O., Abdirahman, F. M., Mohamed, A. I., Mohamed, J., & Ahmed, A. M. (2024). Prevalence of anemia and associated factors among pregnant women at Hargeisa Group Hospital, Somaliland. *BMC Pregnancy and Childbirth*, *24*(1), 332. https://doi.org/10.1186/s12884-024-06539-3

Adila, D. N., Mustaqim, M., Aksayeth, M. B., Fahmi, D., & Rosyani, P. (2025). Literatur Review: Pendekatan Hybrid SVM dan KNN Untuk Klasifikasi Penyakit Tiroid. *Jurnal Teknologi Dan Sistem Informasi Bisnis, 2*(8).

Alexiuk, M., Elgubtan, H., & Tangri, N. (2023). Clinical Decision Support Tools in the Electronic Medical Record. *Kidney International Reports*, *9*(1), 29–38. https://doi.org/10.1016/j.ekir.2023.10.019

Cahyaningrum, Y., Amir, N. A. R., & Nurkalis, U. (2025). Design and Implementation of an Electronic Medical Record Information System

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

to Improve the Efficiency of Health Services. *Tech: Journal of Engineering Science*, *1*(2), 180–191. https://doi.org/10.69836/tech.v1i2.517

Demidova, L. A. (2021). Two-Stage Hybrid Data Classifiers Based on SVM and kNN Algorithms. *Symmetry*, *13*(4), 615. https://doi.org/10.3390/sym13040615

Elvina, A., Oklaini, S. T., & CahyanI, F. P. (2025). Faktor Risiko Kejadian Anemia pada Ibu Hamil di Wilayah Kerja Puskesmas Sidomulyo Kota Bengkulu. *MAHESA: Malahayati Health Student Journal*, *5*(10), 4690–4702. https://doi.org/10.33024/mahesa.v5i10.19631

Hossain, M.M., Nayan, N.M. & Kashem, M.A.A. (2026). A comprehensive maternal health risk prediction dataset from IoT-enabled medical cyber-physical systems in developing countries: supporting machine learning and deep learning applications for clinical decision support. *BMC Medical Informatics and Decision Making, 26*(1), 79. https://doi.org/10.1186/s12911-026-03343-1

Krebs, R., Bagui, S. S., Mink, D., & Bagui, S. C. (2024). Applying Multi-CLASS Support Vector Machines: One-vs.-One vs. One-vs.-All on the UWF-ZeekDataFall22 Dataset. *Electronics*, *13*(19), 3916. https://doi.org/10.3390/electronics13193916

Nurqotimah, D. R., Khudori, A. N., & Pradini, R. S. (2024). Implementasi Algoritma Support Vector Machine (SVM) Untuk Klasifikasi Penyakit Stroke. *Journal of Applied Computer Science and Technology*, *5*(2), 179–185. https://doi.org/10.52158/jacost.v5i2.817

Prabudifa, M. Y., Mayangsari, O. S., Desiani, A., Alwine, D., Primartha, R., & Novita, D. (2025). Ensemble Classifier dengan Teknik Weighted Voting untuk Klasifikasi Jamur Beracun. *JATISI, 12*(2).

Putri, D. A., & Zakaria, H. (2026). Implementasi Hybrid Algoritma Random Forest dan Support Vector Machine Prediksi Penyakit Diabetes Berbasis Web Menggunakan SDLC Iterative Model. *Intellektika: Jurnal Ilmiah Mahasiswa, 4*(5), 150–163.

Putri, J. R., Pujonarti, S. A., Syafiq, A., Achadi, E. L., Perkasa, A. I. K. G., & Basrowi, R. W. (2025). Factors associated with the incidence of anemia in pregnant women in Indonesia: analysis of indonesian health survey (SKI) 2023 data. *Journal of Indonesian Specialized Nutrition, 3*(4), 317-342.

Sandhe, D. S. K., Krisnana, I., & Pradanie, R. (2025). Innovations for Successful Electronic Medical Records (EMR) Implementation in Hospitals: Inovasi untuk Keberhasilan Implementasi Rekam Medis Elektronik (EMR) di Rumah Sakit. *Indonesian Journal of Innovation Studies*, *26*(2). https://doi.org/10.21070/ijins.v26i2.1360

Simatupang, S., Ramadhansyah, R., Tumanggor, R., Tan, E. P., & Fajar, S. A. (2025). Prediction of Heart Disease Risk Based on Patient Health History Using the Support Vector Machine (SVM) Algorithm. *ZERO: Jurnal Sains, Matematika dan Terapan*, *9*(2), 612–619. https://doi.org/10.30829/zero.v9i2.26087

WHO Global Health Observatory. (2026). *Anaemia in women of reproductive age (aged 15-49), prevalence (%), by pregnancy status, by severity, GLOBAL*. https://www.who.int/

World Health Organization. (2025). *Anaemia* (Fact sheet). https://www.who.int/

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.