BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 4, June 2025 | Hal 424-433
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i4.556

Analisa Perbandingan Algorithma K-Nearest Neighbors dan Random Forest untuk Klasifikasi Tindakan Medis Persalinan pada Data Kehamilan Multi-Variabel

Alfin Mahadi, Ema Utami*

Ilmu Komputer, Magister Informatika, Universitas Amikom Yogyakarta, Sleman, Indonesia
Email: [1]alfinmahadi@students.amikom.ac.id [2,*]ema.u@amikom.ac.id
Email Penulis Korespondensi: ema.u@amikom.ac.id

Abstrak−Angka kematian ibu masih tergolong tinggi dalam aspek paling kritis yang memengaruhi kualitas hidup ibu dan bayi yang baru lahir. Urgensi yang sangat signifikan mengingat pentingnya medis yang tepat dalam prosedur persalinan. Rumah sakit islam kendal menyediakan data yang lebih komplek terhadap catatan medis maternal persalinan. Algoritma optimasi dalam klasifikasi banyak diusulkan. Banyak swarm optimasi yang berkembang, particle swarm optimasi menjadi metode optimasi yang unggul. Komparasi metode K-Nearst Neighbors dan Random Forest sering kali diterapkan tanpa optimasi. Penelitian ini membandingkan performa algoritma K-Nearest Neighbors (KNN) dan Random Forest (RF) dalam klasifikasi tindakan medis persalinan menggunakan data rekam medis pasien maternitas di RSI Kendal. Dataset multivariabel meliputi usia, berat badan, tinggi badan, dan kondisi persalinan lebih lengkap. Metode preprocessing melibatkan imputasi nilai kosong dengan KNN imputer, normalisasi data, dan oversampling kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE). KNN dan RF dioptimasi menggunakan Particle Swarm Optimization (PSO) untuk meningkatkan akurasi model. Hasil menunjukkan bahwa RF dengan akurasi 99,72% mengungguli KNN yang memiliki akurasi 97,03%. Pada kelas minoritas, RF menunjukkan keunggulan dengan precision, recall, dan F1-score mencapai 100%, sedangkan KNN lebih rentan terhadap kesalahan pada kelas minoritas. Penelitian ini menegaskan RF dalam menangani data multivariabel yang kompleks dan menyoroti pentingnya optimasi model untuk meningkatkan akurasi dalam klasifikasi tindakan medis persalinan. Temuan ini diharapkan memberikan kontribusi bagi pengembangan sistem pendukung keputusan berbasis machine learning di sektor kesehatan.

Kata Kunci: K-Nearest Neighbors; Random Forest; Particle Swarm Optimasi; Multi-Variabel; Klasifikasi Medis

Abstract−Maternal mortality rate is still high in the most critical aspect affecting the quality of life of mothers and newborns. Very significant urgency considering the importance of proper medical in childbirth procedures. Kendal Islamic Hospital provides more complex data on maternal medical records of childbirth. Many optimization algorithms in classification have been proposed. Many swarm optimizations have been developed, particle swarm optimization is a superior optimization method. Comparison of K-Nearst Neighbors and Random Forest methods is often applied without optimization. This study compares the performance of the K-Nearest Neighbors (KNN) and Random Forest (RF) algorithms in classifying medical procedures for childbirth using medical records of maternity patients at RSI Kendal. The multivariable dataset includes age, weight, height, and more complete childbirth conditions. The preprocessing method involves imputation of empty values with KNN imputer, data normalization, and class oversampling using Synthetic Minority Over-sampling Technique (SMOTE). KNN and RF are optimized using Particle Swarm Optimization (PSO) to improve model accuracy. The results show that RF with an accuracy of 99.72% outperforms KNN with an accuracy of 97.03%. In the minority class, RF shows superiority with precision, recall, and F1-score reaching 100%, while KNN is more prone to errors in the minority class. This study confirms RF in handling complex multivariate data and highlights the importance of model optimization to improve accuracy in the classification of medical labor actions. These findings are expected to contribute to the development of machine learning-based decision support systems in the health sector.

Keywords: K-Nearest Neighbors; Random Forest; Particle Swarm Optimization; Multi-Variable; Medical Classification

1. PENDAHULUAN

Kesehatan maternal menjadi salah satu aspek paling kritis dalam sistem pelayanan kesehatan, yang secara langsung memengaruhi kualitas hidup ibu dan bayi yang baru lahir. Menurut World Health Organization (WHO), komplikasi kehamilan dan persalinan masih menjadi salah satu penyebab utama kematian pada wanita usia reproduktif di seluruh dunia (WHO, 2024). Di Indonesia, Angka Kematian Ibu (AKI) masih tergolong tinggi dengan rata-rata 305 kematian per 100.000 kelahiran hidup, jauh dari target 183 per 100.000 yang diharapkan tercapai pada 2024 [1].

Urgensi dari penelitian ini sangat signifikan, mengingat persalinan dapat diprediksi [2]. Jika prediksi terhadap jenis tindakan persalinan tidak dilakukan secara akurat, maka risiko terjadinya komplikasi seperti perdarahan pasca melahirkan, kelahiran sebelum waktunya, atau pemberian tindakan medis yang tidak tepat dapat meningkat. Konsekuensi dari kegagalan dalam prediksi tindakan persalinan dapat berujung pada komplikasi serius hingga kematian ibu atau bayi, banyak peneliti telah mencoba memecahkan masalah kelahiran [3]. Para ilmuwan memandang pengalaman melahirkan sebagai pengalaman positif [4]. Oleh karena itu, diperlukan penelitian yang lebih lanjut untuk mengembangkan metode yang lebih akurat dalam memprediksi tindakan medis pada persalinan.

Algorithma optimasi dalam melakukan klasifikasi banyak diusulkan yang mendorong untuk mencoba dalam lebih mengoptimalkan algortima random forest classification [5]. Keberhasilan dalam sebuah preprocessing juga sangat penting seperti menerapkan missing value dalam ekstrasi fitur, pengoptimalan pengklasifikasi hingga prapemrosesan data [5]. Penerapan dalam peningkatan dan optimalisasi parameter algoritma melalui penggunaan particle swarm optimization (PSO) algorithm dalam hal pengabungan. Peran untuk mendukung dalam penerapan algoritma yang di usulkan tidak menutup kemungkinan memiliki kendala pada data yang digunakan. Nilai kosong yang hilang perlu dilengkapi bahkan

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 4, June 2025 | Hal 424-433
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i4.556

bisa juga untuk diabaikan hingga dihilangkan. Melengkapi nilai yang hilang dengan menggunakan KNN imputer ke dalam data kosong keaslian data [6]. Salah satu metode imputasi yang umum digunakan untuk melakukan perhitungan data yang hilang adalah metode imputasi KNN [6]. Imputer KNN memanfaatkan nilai K tetangga terdekat. Pendekatan tersebut memperkirakan angka yang hilang dengan menyimpulkan dan mengisi nilai yang hilang menggunakan titik data yang ada.

Penelitian ini mengambil data dari rekam medis pasien maternitas di Rumah Sakit Umum Islam Kendal, yang terdiri dari data kehamilan multi-variabel. Random Forest merupakan meta-classification yang mempertimbangkan teknik klasifikasi ansembel. Klasifikasi yang paling kuat digunakan dalam berbagai teknik klasifikasi machine learning untuk data dimensi yang tinggi [5]. Synthetic Minority Over-sampling Technique (SMOTE) [7] adalah metode oversampling yang diterapkan untuk mengatasi permasalahan dalam tahap pemrosesan data. Sebagian teknik oversampling menggunakan beberapa sample dari kelas minoritas dalam menduplikasi sample yang dapat menyebabkan overfiting pada SMOTE berdasarkan ke dekatan K-nearest neighbor (KNN) [5].

Eksplorasi berbagai perspektif dalam sudut pandang yang berbeda dalam penerapan algoritm k-Nearest Neighbors (kNN). Berbagai studi kasus menekankan machine learning berdasarkan issu utama seperti interpretabilitas model, parameter yang optimal. Reformasi kecerdasan buat memungkinkan analisis untuk mengubah data menjadi wawasan. Kesehatan adalah sektor integral yang menghasilkan sejumlah besar data setiap hari. Kasus yang berhubungan dengan persalinan dengan metode tradisional atau persalinan casesar dilakukan untuk menyelamatkan ibu dan janin Ketika komplikasi berhubungan dengan kelahiran vaniga muncul [8]. Tujuan dalam kasus ini adalah untuk melakukan dan mengembangkan model yang digunakan sebagai analisis system dalam mendukung keputusan perawatan bersalin baik cara persalinan sebelum melahirkan.

Dataset skaling dalam machine learning juga disebut dengan normalisasi [9]. Beberapa peneliti sebelumnya menunjukan bahwa memilih teknik yang tidak sesuai dan tidak memadai dapat merugikan kinerja klasifikasi dibandingkan dengan tidak melakukan skaling data sama sekali [9]. Sensitivitas terhadap skaling juga dapat di explorasi dalam hal ini sehingga dapat juga diterapkan dalam usulan dalam perbandingan, tujuan dari skaling yang diharapkan memberi dampak yang signifikan dalam usulan yang diajukan penelitian klasifikasi ini. Robust skaler adalah metode teknik skaling yang tidak terpengaruh oleh outlier, karena menggunakan median dan interquartile range (IQR) untuk scaling sehingga hal tersebut cocok untuk data yang banyak outlier.

Penelitian yang dilakukan untuk memprediksi jenis persalinan bersumber pada dataset yang digunakan dalam penelitian ini berasal dari rekam medis ibu hamil di 6 klinik Bidan Palupi Margiani. Hasil penelitian menunjukkan bahwa KNN mampu mengklasifikasikan proses persalinan dengan cukup baik, namun penelitian ini hanya berfokus pada satu algoritma dan tidak melakukan perbandingan dengan algoritma lain. Oleh karena itu, penelitian ini menginspirasi untuk membandingkan KNN dengan RF dalam prediksi tindakan persalinan berdasarkan data yang lebih beragam dari Rumah Sakit Islam Kendal.

Penelitian lain [2] meneliti penggunaan algoritma Random Forest (RF) dan Decision Tree dalam memprediksi pendarahan pascapersalinan pada ibu hamil. Dataset yang digunakan dalam penelitian ini berjumlah 500 data, yang terdiri dari variabel seperti paritas, status anemia, jarak kehamilan, dan komplikasi persalinan. Hasil penelitian menunjukkan bahwa algoritma RF memiliki akurasi sebesar 83%, lebih tinggi dibandingkan algoritma Decision Tree yang hanya mencapai akurasi 82%. Penelitian ini memperlihatkan keunggulan RF dalam prediksi medis. Dalam konteks penelitian ini, RF akan diuji lebih lanjut dan dibandingkan dengan KNN untuk melihat mana yang lebih efektif dalam klasifikasi tindakan persalinan.

Penelitian tentang kesehatan dengan algoritma Random Forest juga dilakukan [10] menunjukkan bahwa algoritma Random Forest (RF) memiliki performa yang sangat baik dalam mendeteksi penyakit jantung. Penelitian ini menggunakan dataset klinis yang terdiri dari beberapa variabel kunci terkait dengan kondisi jantung pasien, dan algoritma RF menghasilkan akurasi sebesar 86.9%. Meskipun tidak spesifik pada prediksi tindakan persalinan, penelitian ini menunjukkan keandalan RF dalam prediksi medis secara umum, yang dapat digunakan sebagai dasar pembanding untuk 8 melihat bagaimana algoritma ini bekerja pada data tindakan persalinan. Penelitian ini mendukung penerapan RF dalam penelitian Anda untuk memprediksi tindakan persalinan, dibandingkan dengan KNN.

Kontribusi penelitian ini adalah untuk membandingkan performa dua algoritma populer, yaitu K-Nearest Neighbors (KNN) dan Random Forest (RF), dalam konteks prediksi tindakan medis persalinan, menggunakan dataset yang di ambil pada salah satu rumah sakit RSI Kendal. Dengan membandingkan kedua algoritma ini, penelitian ini diharapkan dapat memberikan rekomendasi algoritma yang lebih akurat dan efisien, sehingga dapat mendukung tenaga medis dalam membuat keputusan yang cepat dan tepat. Selain itu, penelitian ini juga memberikan sumbangsih praktis berupa dataset yang dapat digunakan oleh peneliti lain untuk analisis lebih lanjut, serta memberikan kontribusi pada pengembangan teknologi prediksi medis berbasis machine learning di bidang kesehatan maternal. Penelitian dalam ekperiment meliputi hal penting terdiri dari dua metodologi yang membahasa tentang metodologi yang digunakan point tiga peran pembahasan dan kerangka terakhir dalam kesimpulan untuk implikasi masa depan.

2. METODOLOGI PENELITIAN

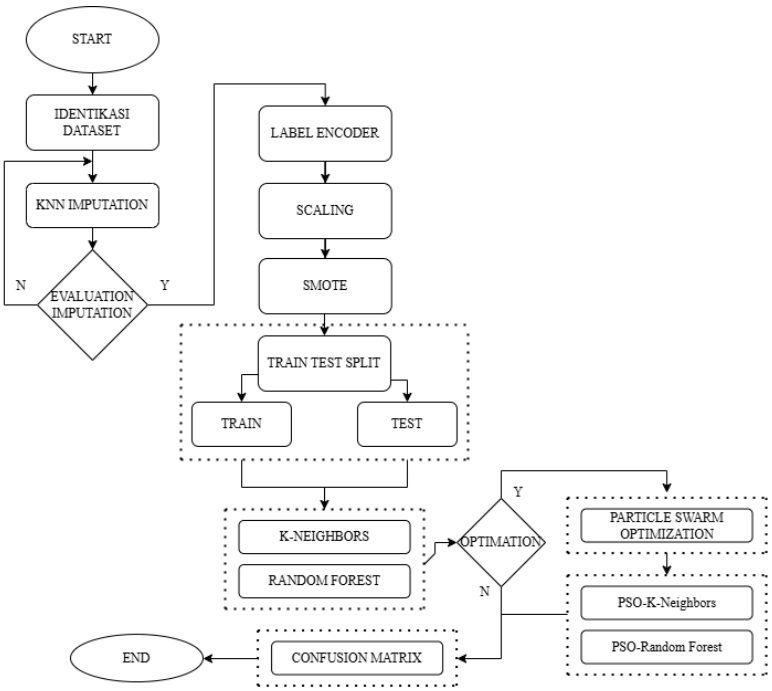
Metodologi penelitian berdasarkan point dua dalam membandingkan algoritma K-Nearest Neighbors (KNN) dan Random Forest (RF) dalam klasifikasi tindakan medis persalinan berdasarkan data rekam medis maternitas di Rumah Sakit Islam

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 4, June 2025 | Hal 424-433
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i4.556

(RSI) Kendal. Tahapan penelitian dimulai dengan pengumpulan data yang terdiri dari variabel multivariabel, seperti usia, berat badan, tinggi badan, dan hasil pemeriksaan persalinan (Leopold 1-4). Data ini kemudian melalui proses preprocessing yang mencakup imputasi nilai kosong menggunakan KNN imputer untuk melengkapi data yang hilang. Selain itu, data yang telah diimputasi dinormalisasi menggunakan teknik scaling yang sesuai untuk memastikan distribusi data seragam. Langkah penting lainnya adalah penyeimbangan dataset menggunakan metode Synthetic Minority Over-sampling Technique (SMOTE) guna mengatasi masalah ketidakseimbangan kelas pada data, sehingga meningkatkan akurasi prediksi. Langkah selanjutnya untuk meningkatkan performa model, optimasi parameter dilakukan dengan metode Particle Swarm Optimization (PSO). PSO digunakan untuk menentukan parameter optimal seperti jumlah tetangga terdekat (n_neighbors) pada KNN dan jumlah estimator (n_estimators) serta kedalaman maksimum (max_depth) pada RF. Terakhir melakukan evaluasi model yang dilakukan menggunakan matriks evaluasi seperti precision, recall, dan F1-score yang dianalisis melalui confusion matrix.

2.1 Tahap Alur Penelitian

Tahapan alur penelitian yang diusulkan menggunakan Algoritma K-Nearst Neighbor (KNN) dan Random Forest untuk klasifikasi tindakan medis persalinan pada Data Kehamilan Multi-Variabel yaitu dengan penggunakan penanganan pada missing value, normalisasi data, selanjutnya dilakukan pengujian. Proses usulan yang diajukan diharapkan dapat menjadi usulan yang baik dalam melakukan klasifikasi pada persalinan. Proses juga meliput optimasi dalam mencari nilai optimal yang dan tahap akhir yang diharapkan dapat memberi nilai evaluasi matrik yang baik.



Gambar 1. Alur Metode Penelitian

Usulan berdasarkan Gambar 1 memberikan ilustrasi metodologi secara menyeluruh. alur metodologi diawali dengan cara identifikasi dataset, identifikasi diperlukan untuk melihat penyelesaian dari data agar dapat di terapkan sehingga menjadi alasan dan penyebabkan usulan ini diajukan, terdapat banyak missing value pada dataset. Missing value yang tersedia disebabkan oleh human error atau kesalahan input data seperti kesalahan ejaan, penulisan dan sebagainya perlu ditangani. KKN imputer menjadi metode populer [6] dalam mengatasi hal tersebut. Usulan metode imputasi yang populer menjadi saran dalam penerapan usulan metode.

Imputasi yang melengkapi missing value memberikan nilai imputasi dalam pendekatanya. Penyesuaain label encoder yang digunakan untuk mengubah data kategorikal menjadi data numerik penyesuaian label encoding yang dapat disesuaikan dalam machine learning. Scaling melakukan normalisasi [9] untuk memastikan distribusi data, hal labeling yang dapat disampaikan pada pembahasan point dataset pada tabel 2. Memerlukan hal untuk menangani ketidak seimbangan data SMOTE [7][11] diusulkan untuk menagani hal tersebut. Dilanjutkan dengan alur metode seperti pembagian data dan penerapan algoritma serta optimasi dan confusion matrik dan validasi.

2.2 Multi-Variabel Dataset

Multi-variabel dataset yang digunakan bersumber dari penilaian dokter di rumah sakit RSI Kendal. Multi-variabel yang melibatkan delapan variable yang digunakan dalam pemodelan menjelaskan dalam rekam medis pasien maternitas yang terdiri pada tabel 1 sebagai berikut. Tabel tersebut memberikan beberapa indikator ke dalam dua kelas klasifikasi positif dan negative dalam indikasi kehamilan sesar dan non sesar berdasarkan parameter tertentu yang telah disebutkan sebagai variable dependen masing-masing variabel.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 4, June 2025 | Hal 424-433
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i4.556

Tabel 1. Multi-Variabel Data Kehamilan

Variabel	Indikasi Positif Sesar	Indikasi Negatif Sesar
Usia	< 16 tahun atau > 35 tahun	16-35 tahun
Tinggi Badan (TB)	< 145 cm (panggul sempit)	≥ 145 cm
Berat Badan (BB)	> 90 kg atau IMT ≥ 30 (obesitas)	IMT < 25 (BB normal)
Paritas	≥ 5 (paritas tinggi)	1 atau 2 (paritas rendah)
Leopold 1	Presentasi sungsang atau melintang	Presentasi kepala (vertex)
Leopold 2	Kepala tidak bisa masuk ke panggul	Kepala janin berada dalam panggul
Leopold 3	Presentasi sungsang atau bagian tubuh tidak jelas	Presentasi kepala dengan bagian tubuh jelas
Leopold 4	Kontraksi lemah atau janin terhalang dalam panggul	Kontraksi teratur dan janin dalam posisi optimal

Tabel 1 memberikan penjelasan parameter multi-variabel data kehamilan yang menunjukkan kriteria klinis untuk membedakan indikasi persalinan sesar (positif) dengan persalinan normal (negatif) berdasarkan parameter medis. Faktor usia ekstrem (<16 atau >35 tahun), tinggi badan <145 cm, dan obesitas (IMT ≥30) termasuk indikasi sesar, sementara kondisi sebaliknya mendukung persalinan normal. Pemeriksaan Leopold (1-4) menjadi penentu kritis - presentasi sungsang, kepala janin tidak masuk panggul, atau kontraksi lemah mengarah ke sesar, sedangkan presentasi kepala dengan kontraksi teratur mendukung persalinan normal. Paritas tinggi (≥5 kelahiran) juga menjadi pertimbangan sesar, berbeda dengan paritas rendah (1-2 kelahiran) yang lebih aman untuk persalinan normal. Kriteria ini menggambarkan kompleksitas pertimbangan medis dalam menentukan metode persalinan yang optimal untuk keselamatan ibu dan bayi [12].

2.4 Metode K-Nearest Neighbors

K-Nearest Neighbor adalah algoritma klasifikasi yang merupakan cabang dari kecerdasan buatan. Kecerdasan buatan tidak lepas dengan machine Learning (ML) yang merupakan cabang dari ilmu komputer berhubungan dengan Artificial Intelligent (AI) atau ilmu kecerdasan buatan berfokus pada pembuatan/ pengembangan dan studi suatu sistem dengan tujuan agar mampu belajar dari data-data yang diperolehnya [13]. KNN dalam metodenya memiliki tujuan untuk mengklasifikasikan objek baru dari atribut yang berasal dari data latih, dalam proses klasifikasi umumnya nilai K dapat menggunakan jumlah ganjil dengan pertimbangan perhitungan menggunakan persamaan 3 agar tidak adanya jarak yang sama, jarak tetangga terdekat dapat dihitung dengan Euclidean Distance [14].

2.5 Metode Random Forest

Random forest merupakan algoritma pembelajaran mesin yang memiliki banyak pohon keputusan [15] sebagai base classifier yang dikombinasikan. Algoritma ini merupakan kombinasi dari metode Random Sub Spaces dan Bagging, Terdapat tiga aspek penting dalam algoritma random forest, Melakukan bootstrap sampling untuk membangun pohon prediksi Melakukan bootstrap sampling untuk membangun pohon prediksi Random forest melakukan prediksi dengan mengkombinasikan hasil dari setiap pohon keputusan dengan cara majority vote untuk klasifikasi [16].

2.6 Particle Swarm Optimization

Particle swarm optimization adalah teknik optimasi yang diperkenalkan oleh kennedy dan Eberhart [17]. Particle Swarm Optimization menggunakan mekanisme sederhana yang meniru perilaku kawanan burung pada kawanan burung untuk memandu partikel untuk mencari solusi optimal global. PSO memiliki tujuan optimasi dalam menemukan solusi optimal berdasarkan posisi terbaik yang pernah dicapai. Algoritma PSO adalah bahwa suatu partikel individu mengidentifikasi solusi potensial dengan terbang di ruang batas yang didefinisikan dan menyesuaikan arah eksplorasinya untuk mendekati solusi optimum global berdasarkan informasi yang dibagikan di antara kelompok [17].

2.6 Evaluasi Matrix

Pembandingkan performa metode pada mesin pembelajaran merupakan permasalahan yang tidak mudah. Saat ini secara sederhana diasumsikan bahwa performa suatu metode diukur dari sejauh mana kemampuan metode tersebut dalam melakukan klasifikasi secara akurat terhadap data tertentu yang diuji. Confusion Matrix adalah alat yang digunakan untuk menganalisis seberapa baik classifier dapat mengenali tupel yang berbeda. Confusion matrix memberikan rincian klasifikasi, Memilih formula ukuran kinerja yang tepat untuk evaluasi algoritma adalah sebuah tahapan yang kritis [7][18]. Pengukuran ini dapat digunakan untuk mi enghitung nilai akurasi dan presisi dasarkan presentasi nilai true dan false serta nilai positif dan negatifnya dengan persamaan [19].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Accuracy menggambarkan seberapa sering model membuat prediksi yang benar secara keseluruhan, baik untuk kelas positif (persalinan sesar) maupun negatif (persalinan normal). Metrik ini dihitung dengan membagi jumlah prediksi benar (TP + TN) dengan total prediksi persamaan (2.1).

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 4, June 2025 | Hal 424-433
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i4.556

$$Precision = \frac{TP}{FP+TP} \quad (2)$$

Precision persamaan (2.2) mengukur ketepatan prediksi positif dengan menghitung berapa persen dari prediksi "sesar" yang benar-benar sesar (TP / (TP + FP)). Precision tinggi berarti sedikit pasien normal yang salah diklasifikasikan sebagai sesar (FP rendah)

$$Recall = \frac{TP}{FN+TP} \quad (3)$$

Recall menilai kemampuan model menemukan semua kasus positif yang ada, dihitung dari persentase pasien sesar yang berhasil diidentifikasi (TP / (TP + FN)). Recall tinggi berarti sedikit kasus sesar yang terlewat, krusial dalam medis karena gagal mendeteksi sesar (FN tinggi) bisa berakibat fatal persamaan (3)

$$F1 - Score = 2x\frac{precision\ x\ recall}{precision+recall} \quad (4)$$

F1-Score adalah rata-rata harmonik precision dan recall, memberikan balance antara keduanya. Metrik ini berguna ketika kelas tidak seimbang dan baik FP maupun FN perlu diminimalkan dengan persamaan (2.4).

3. HASIL DAN PEMBAHASAN

Pembahasan menyeluruh dari hasil dan pembahasan bagian ini berisi hasil dan pembahasan dari topik penelitian, penerapan metode yang digunakan dengan optimasi data yang digunakan secara transparansi dapat digunakan dan dievaluasi dalam ekperiment yang telah dilakukan pada penelitian. Bagian ini juga merepresentasikan penjelasan yang berupa penjelasan dan di visualisasikan ke dalam gambar dan berikan beberapa notasi angka tabel-tabel. Hasil dan pembahasan dalam penelitian meliputi cangkupan yang terdiri mulai dari preparing data dalam identifikasi. Korelasi hasil dari imputasi kemudian distribusi data hasil dari imputasi untuk evaluasi

3.1 Preprocessing Imputasi Dataset

Dataset yang bersumber dari RSI Kendal memiliki jumlah 3526 dataset dalam Riwayat catatan medis yang dimiliki pada rumah sakit RSI Kendal. Preprocessing imputasi dalam dataset menggunakan imputasi KNN Imputer [6]. Variabel yang digunakan meliputi usia. Usia memiliki batasan dalam kriteria multi-variabel pada tabel 1. Dimana hal tersebut digunakan sebagai parameter dalam imputasi dan langkah tersebut ditujukan untuk meminimalisir missing value dari kolom usia. Data usia yang kurang atau sama dengan tujuh belas tahun akan diseragamkan dengan usia minimum imputasi dapat direpresentasikan pada Tabel 2.

Tabel 2. Dataset Multi-variabel

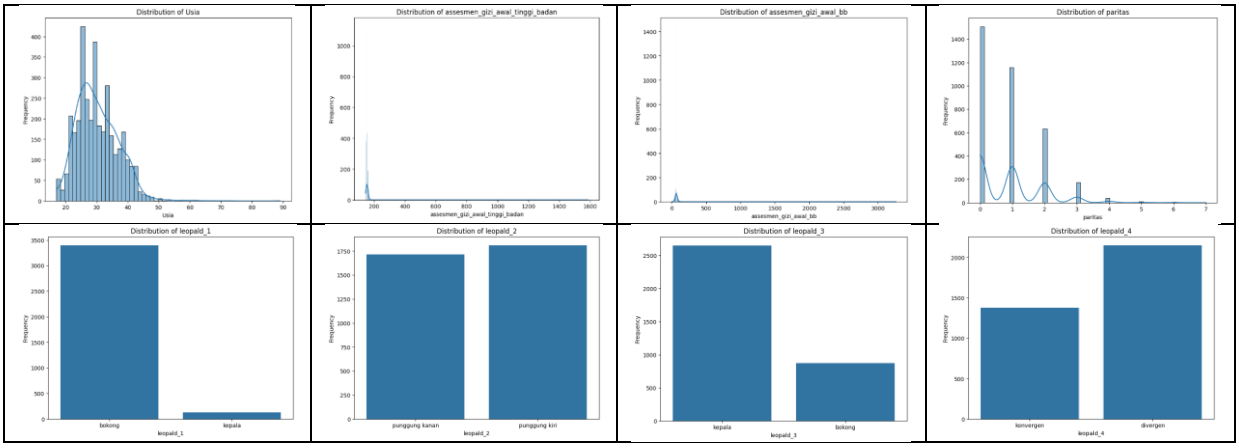
	Usia	Tinggi Badan	Berat Badan	Paritas	Leopold 1	Leopold 2	Leopold 3	Leopold 4
0	37	151	60	1	Bokong	PunggungKanan	Kepala	Konvergen
1	39	150	NaN	NaN	Kepala	PunggungKiri	Bokong	Divergen
2	NaN	NaN	53	0		NaN	Kepala	Konvergen
…..	…..	…..	…..	…..	…..	…..	…..	…..
3525	25	160	NaN	2	Kepala	NaN	Kepala	NaN
3526	22	158	68	2	NaN	PunggungKanan	Kepala	Divergen

Tinggi badan dan berat badan juga akan disesuaikan seperti batas minimum dalam proses tersebut dimana sebagai hal bentuk untuk melakukan proses pengurangan noise dan proses langkah dari tujuan mempertahankan ketersediaan data untuk memaksimalkan klasifikasi pada persalinan. Leopold satu hingga empat diambil berdasarkan pada variabel jumlah terbanyak dalam kuantitas mayoritas dan minoritas. Proses tersebut dalam imputasi, KNN imputer mengisi nilai yang hilang berdasarkan rata-rata atau nilai mayoritas K tetangga terdekat [6] Imputasi ditentukan secara arbitrer dimana pemilihan nilai K didasarkan pada rata-rata tetangga terdekat untuk mengisi nilai yang hilang, data yang telah di imputasi data akan dinormalisasikan dan diskalakan menggunakan RobustScaler untuk mengurangi dampak outlier. Data Leopold diubah menggunakan label encode dalam proses pengubahan data kategorikal ke data numerik.

3.2 Distribusi data Imputasi

Hasil imputasi dari KNNimputer memberikan hasil komputasi yang cepat meskipun hasil ini tidak dapat dipastikan angka terbaik dalam imputasi data yang dihasilkan. Terlihat dari distribusi yang dihasilkan dari beberapa data seperti usia leopald 1 pada Gambar 2 distribusi imputasi, memberikan hasil yang dominan. Hasil dominan tersebut dapat disebabkan oleh imputasi yang memiliki ketergantungan. Pemilihan K tangga terdekat yang tidak tepat dapat menyebabkan hasil imputasi yang kurang akurat. Menjadi hal yang perlu di evaluasi lebih lanjut. Distribusi secara detail dapat di representasikan ke dalam gambar point 2 sebagai hasil dari KNNimputer.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 4, June 2025 | Hal 424-433
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i4.556



Gambar 2. Distribusi Hasil Imputasi

Histogram pertama menunjukkan distribusi usia pasien, dengan puncak pada rentang usia 20–35 tahun. Histogram yang dihasilkan menunjukkan bahwa mayoritas individu dalam dataset berada pada kelompok usia reproduktif aktif, yang masuk akal mengingat konteks medis yang berkaitan dengan kehamilan atau persalinan. Kurva distribusi melandai setelah usia 35, dengan hanya sedikit individu di atas usia 50 tahun.

Histogram gambar pada kedua bahwa hasil menunjukkan distribusi nilai tinggi badan dan berat badan yang tampaknya sangat condong (skewed) ke kiri. Mayoritas nilai tinggi badan berkisar sekitar 150-170 cm, sedangkan berat badan lebih terkonsentrasi pada rentang tertentu 50-70 kg, dengan beberapa outlier yang ekstrem. Distribusi yang condong dihasilkan memberikan data yang mungkin disebabkan oleh pengaruh nilai yang tidak realistis atau outlier, yang perlu diverifikasi lebih lanjut dalam imputasi.

Histogram paritas menggambarkan jumlah kelahiran sebelumnya yang dialami oleh pasien. Mayoritas individu memiliki nilai paritas rendah (0, 1, atau 2), yang menunjukkan bahwa sebagian besar wanita dalam dataset ini adalah ibu muda atau baru memiliki satu hingga dua anak. Hanya sebagian kecil individu dengan nilai paritas lebih tinggi (di atas 3) hubungan dari nilai imputasi yang relevan dapat dipertimbangkan.

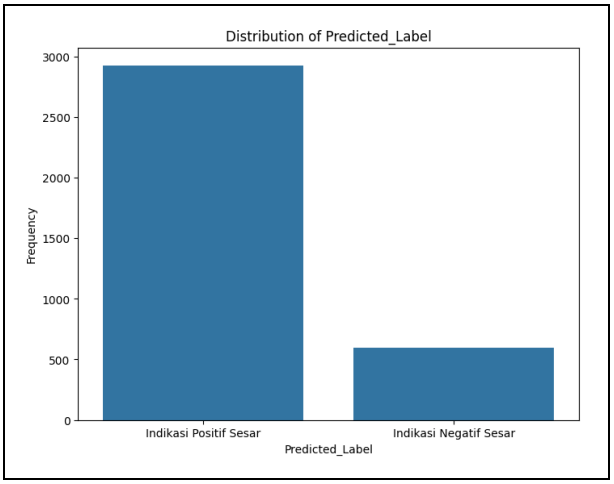
Histogram pada leopold 1-4 Mayoritas hasil pengamatan menunjukkan posisi "bokong", sementara posisi "kepala" jauh lebih sedikit. Ini menunjukkan bahwa presentasi janin cenderung lebih sering dalam posisi sungsang. Distribusi antara "punggung kanan" dan "punggung kiri" relatif seimbang, menunjukkan bahwa orientasi punggung janin tidak memberikan bias besar dalam dataset. Sedangkan leopold 3 Kebanyakan hasil menunjukkan posisi "kepala", dengan proporsi yang lebih kecil dalam posisi "bokong". Ini menunjukkan bahwa sebagian besar janin pada akhirnya berada dalam posisi kepala di bagian bawah, selain itu leopold 4 Posisi "konvergen" dan "divergen" menunjukkan distribusi yang hampir seimbang, tetapi "divergen" sedikit lebih sering diamati, yang mungkin relevan dalam pengambilan keputusan medis hasil dari imputasi.

3.3 Labeling

Proses labeling dilakukan secara manual dengan kondisi berdasarkan parameter variabel pada Tabel 1. Hasil imputasi digabungkan secara menyeluruh ke dalam 1 tabel. Maka langkah yang dilanjutkan adalah dengan langkah melakukan labeling. Hasil labeling diberikan label sebagai label prediksi dengan indikasi positif sesar dan negatif sesar. Proses klasifikasi yang di gunakan dalam klasifikasi merupakan klasifikasi dua kelas. Labeling hanya dilakukan dengan cara membuat definisi secara matematis dengan logika jika usia kurang dari dua atau lebih besar dari tiga puluh lima tahun sesuai dengan ketentuan multi-variabel maka status akan menjadi prediksi positif sesar. Langkah terebut diteruskan hal yang sama kurang atau sama dengan paritas 5 maka indikasi positif sesar dapat diterapkan maka dapat disimpulkan bahwa hasil dari imputasi data yang menyebabkan label sangat berpengaruh terhadap hasil proses labeling.

Proses menentukan label pada data multivariat apakah data termasuk ke dalam kategori "Indikasi Positif Sesar" proses tersebut akan berdasarkan logika aturan medis: jika usia kurang dari 16 tahun atau lebih dari 35 tahun, tinggi badan kurang dari 145 cm, berat badan lebih dari 90 kg, paritas 5 atau lebih, atau jika hasil pemeriksaan Leopold 1 sampai 4 menunjukkan kondisi tertentu yang tidak ideal untuk persalinan normal (seperti kepala tidak berada di bawah, janin dalam posisi bokong, atau panggul dalam kondisi divergen), maka label yang diberikan adalah Indikasi Positif Sesar. Jika tidak memenuhi satu pun dari kondisi tersebut, maka diberi label Indikasi Negatif Sesar.

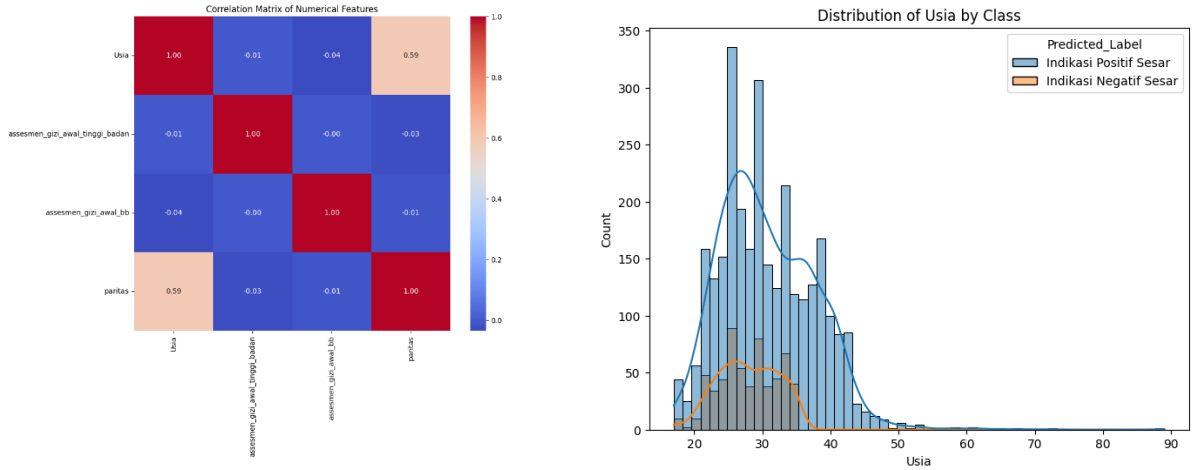
BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 4, June 2025 | Hal 424-433
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i4.556



Gambar 3. Hasil Class Manual Labeling

3.4 Distribusi Korelasi dan Prediction

Distribusi korelasi gambar 4(a) pada gambar pertama menunjukkan correlation matrix of numerical features, yang menampilkan hubungan antar fitur numerik dalam dataset. Matriks korelasi ini menggunakan skala warna untuk menunjukkan kekuatan dan arah hubungan antara dua variabel, dengan nilai korelasi berkisar antara -1 hingga 1. Pada matriks ini, variabel seperti "Usia" dan "paritas" memiliki korelasi positif sebesar 0,59, menunjukkan hubungan moderat positif dimana bertambahnya usia cenderung diikuti dengan peningkatan nilai paritas (jumlah kelahiran sebelumnya). Namun, sebagian besar pasangan variabel lain menunjukkan nilai korelasi mendekati nol, seperti "assessmen_gizi_awal_tinggi_badan" terhadap "paritas" (-0,03), yang mengindikasikan tidak adanya hubungan linear yang signifikan di antara variabel-variabel tersebut. Korelasi ini berguna untuk memahami apakah ada fitur numerik yang memiliki hubungan signifikan yang dapat memengaruhi analisis atau prediksi model.



Gambar 4. (a) Korelasi Matrix dan (b) Distribusi Class Usia

Gambar 4(b) memperlihatkan Distribusi Usia berdasarkan Kelas, di mana histogram usia dikategorikan menjadi dua label: Indikasi Positif Sesar dan Indikasi Negatif Sesar. Data distribusi ini memberikan wawasan tentang perbedaan pola usia antara kedua kategori. Pada kelas dengan indikasi positif sesar, distribusi usia memiliki puncak pada usia sekitar 25-35 tahun, yang tampaknya merupakan kelompok usia dominan untuk tindakan sesar. Sebaliknya, untuk kelas dengan indikasi negatif sesar, distribusi cenderung lebih datar dan frekuensinya menurun signifikan di luar rentang usia tersebut. Pola ini menunjukkan bahwa usia mungkin menjadi salah satu variabel penting dalam menentukan indikasi tindakan medis terkait persalinan, meskipun memerlukan analisis lebih lanjut untuk memvalidasi signifikansi pengaruhnya.

3.5 Evaluasi Matrix

Evaluasi terdiri dari kedua model yang diajukan dengan optimasi particle swarm. Evaluasi terhadap Data kategori pada fitur dan label diencode menggunakan LabelEncoder. Proses dalam evaluasi matrik yang dihasilkan selanjutnya, data diskalakan menggunakan RobustScaler untuk mengurangi dampak outlier, kemudian dibagi menjadi data latih (80%) dan data uji (20%) dengan stratifikasi untuk menjaga distribusi kelas. Untuk menangani ketidakseimbangan kelas, data latih diseimbangkan menggunakan SMOTE [5], yang menghasilkan sampel sintetik untuk kelas minoritas akurasi yang sangat

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 5, No 4, June 2025 | Hal 424-433
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i4.556

tinggi berpotensi mengindikasikan overfitting. Evaluasi peneliti telah mengantisipasi dengan penggunaan SMOTE untuk mengatasi ketidakseimbangan data [20].

Jumlah partikel (swarmsize) yang digunakan dalam proses optimasi adalah 20, dengan batas maksimum iterasi (maxiter) sebanyak 10 iterasi. Untuk KNN, ruang pencarian parameter yang dieksplorasi adalah jumlah tetangga terdekat (n_neighbors) antara 3 hingga 9, dan parameter jarak Minkowski (p) antara 1 hingga 2. Sedangkan untuk Random Forest, parameter yang dioptimalkan adalah jumlah pohon (n_estimators) dalam rentang 50 hingga 100, serta kedalaman maksimum pohon (max_depth) antara 5 hingga 15.

PSO secara default meminimalkan fungsi objektif nilai akurasi agar PSO tetap dapat mencari solusi dengan akurasi maksimum semakin kecil (lebih negatif) nilai fungsi objektif, yang menjadi tujuan optimasi. Random Forest menunjukkan performa yang lebih baik dibandingkan KNN karena kemampuannya dalam menangani data non-linear dan secara efektif menangkap interaksi kompleks antar fitur melalui ensemble pohon keputusan. Hal ini terlihat dari nilai evaluasi yang sangat tinggi pada seluruh metrik, dengan precision, recall, dan F1-score masing-masing mencapai 1.00 untuk kelas Positif Sesar dan 0.98–1.00 untuk kelas Negatif Sesar.

Model K-Nearest Neighbors (KNN) yang telah dioptimasi menggunakan Particle Swarm Optimization (PSO) untuk menemukan parameter terbaik, yaitu jumlah tetangga terdekat (n_neighbors) dan parameter jarak (p). Parameter optimal yang ditemukan adalah n_neighbors = 2 dan p = 2 (menggunakan jarak Euclidean). Model yang telah dioptimasi kemudian dilatih pada data latih yang telah diseimbangkan dan dievaluasi pada data uji, menghasilkan akurasi sebesar 97.03%.

Evaluasi model menunjukkan performa yang sangat baik. Pada kelas Indikasi Positif Sesar, precision mencapai 99%, recall 98%, dan F1-score 98%, mencerminkan kemampuan model mendeteksi kelas mayoritas dengan baik. Sementara hal tersebut, pada kelas Indikasi Negatif Sesar, precision mencapai 90% dan recall 93%, yang menunjukkan performa yang sedikit lebih rendah pada kelas minoritas. Confusion matrix mengungkap bahwa sebagian besar kesalahan terjadi pada kelas Negatif Sesar yang salah diprediksi sebagai Positif Sesar. Secara keseluruhan, model KNN dengan optimasi PSO memberikan hasil yang sangat signifikan dalam uji coba berdasarkan dari alur metodologi yang diusulkan.

Penjelasan menyeluruh tentang hasil confusion matrix dalam reportnya dapat di representasikan kedalam Gambar 5. Confusion matrix mendukung hasil dengan menunjukkan bahwa model KNN yang dioptimasi menggunakan PSO adalah pendekatan yang efektif untuk dataset ini. Perbandingan akan direpresentasikan dengan model radom forest yang juga dapat dilihat dalam evaluasinya. Komparasi dalam optimasi yang dihasilkan pada best parameter optimasi yang spesifik dapat direpesentasikan ke dalam Gambar 5.

```
Best Parameters for KNN using PSO: [1.78110821 1.67915805]

KNN Classification Report:
                      precision    recall  f1-score   support

Indikasi Negatif Sesar       0.90      0.93      0.91       120
Indikasi Positif Sesar       0.99      0.98      0.98       586

              accuracy                           0.97       706
             macro avg       0.94      0.96      0.95       706
          weighted avg       0.97      0.97      0.97       706

Metrics for KNN:
Accuracy: 97.03%
Precision: 97.09%
Recall: 97.03%
F1-Score: 97.05%
```

Gambar 5. Confusion Matrix K-Nearest Neighbors

Evaluasi matrix yang dihasilkan berdasarkan menggunakan Proses optimasi model Random Forest dilakukan menggunakan Particle Swarm Optimization (PSO) untuk menemukan parameter terbaik (n_estimators) dan kedalaman maksimum (max_depth). Parameter optimal yang ditemukan adalah n_estimators = 87.73 dan max_depth = 8.76, yang kemudian diterapkan untuk melatih model. Setelah pelatihan, model dievaluasi menggunakan data uji yang sama 80 dan 20, menghasilkan akurasi yang sangat tinggi sebesar 99.72%. akurasi yang dihasilkan dari usulan lebih tinggi dibandingkan dengan hasil penelitian sebelumnya [2].

Evaluasi model Random Forest menunjukkan performa yang baik pada kedua kelas. Untuk kelas Indikasi Negatif Sesar, precision mencapai 98%, recall 100%, dan F1-score 99%, menunjukkan kemampuan model yang hampir sempurna dalam mendeteksi kelas. Pada kelas Indikasi Positif Sesar, precision, recall, dan F1-score semuanya mencapai 100%, mencerminkan bahwa model mampu memprediksi kelas mayoritas tanpa kesalahan. Hasil ini mencerminkan kekuatan Random Forest dalam menangani data yang kompleks dan imbang.

Secara keseluruhan, metrik evaluasi seperti macro average dan weighted average memiliki nilai precision, recall, dan F1-score masing-masing 99% dan 100%, yang mendekati nilai sempurna. Dengan hasil ini, model Random Forest dengan optimasi PSO dapat dianggap unggul dan sangat andal dalam mengklasifikasikan data uji, baik untuk kelas minoritas maupun mayoritas. Particle swarm terbukti menghasilkan optimasi yang baik dalam evaluasi matrik dalam klasifikasi yang dihasilkan representasi Gambar 6.

ISSN 2774-3659 (Media Online)
Vol 5, No 4, June 2025 | Hal 424-433
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v5i4.556

```
Best Parameters for Random Forest using PSO: [87.73845074  8.76110536]

Random Forest Classification Report:
                        precision    recall  f1-score   support

Indikasi Negatif Sesar       0.98      1.00      0.99       120
Indikasi Positif Sesar       1.00      1.00      1.00       586

              accuracy                           1.00       706
             macro avg       0.99      1.00      1.00       706
          weighted avg       1.00      1.00      1.00       706

Metrics for Random Forest:
Accuracy: 99.72%
Precision: 99.72%
Recall: 99.72%
F1-Score: 99.72%
```

Gambar 6. Confusion Matrix Random Forest

Perbandingan dari komparasi ke dua model memberikan hasil yang signifikan, tingginya akurasi yang disebabkan juga menjadi penentu dalam rekomendasi medis. Namun tidak menutup kemungkinan berdasarkan hasil yang diperoleh memiliki banyak outlier terhadap data yang tidak dapat dipertanggung jawabkan dalam hasil asli dan validasi terhadap missing value yang dihasilkan berdasarkan kedekatan. Hasil evaluasi matrix random forest memiliki akurasi lebih baik dibandingkan dengan K-Nearest Neighbors.

3.6. Keterbatasan

Setiap kerangka usulan memiliki beberapa keterbatasan, sehingga data yang diusulkan dapat juga menghadapi beberapa keterbatasan seperti kinerja alur metode yang diusulkan sehingga sangat bergantung pada kualitas data yang digunakan dalam melatih model. Hal tersebut dapat diamati dimana proses imputasi KNN dan Random forest. Hasil pengambilan serta pemilihan dalam kategori kelas sangat bergantung pada kelengkapan dan keakuratan data input dan validasi oleh sumber yang memahami hal tersebut, proses labeling juga dapat menggunakan metode system rekomendasi seperti Fuzzy system dalam mengambil keputusan berdasarkan multi-variabel kategori Tabel 1.

Sensitivitas terhadap parameter KNN dalam mencari tetangga terdekat K, seperti jumlah tetangga (K) dan metrik jarak, menimbulkan tantangan dalam mengoptimalkan kualitas imputasi yang dihasilkan yang menyebabkan dapat menimbulkan hasil data yang tidak akurat. Skaling dapat juga dapat di eplorasi seperti penetapan skaling yang paling tetap dalam ketersediaan data yang lebih cocok dengan metode seperti standarlisasi. Kompleksitas teknik pengumpulan data riwayat medis persalinan yang digunakan, dapat juga membantu ataupun menghambat pemahaman tentang fitur yang berpengaruh dalam klasifikasi persalinan.

4. KESIMPULAN

Bagian penelitian membuktikan bahwa algoritma Random Forest (RF) lebih unggul dibandingkan K-Nearest Neighbors (KNN) dalam klasifikasi tindakan medis persalinan, dengan akurasi masing-masing 99,72% dan 97,03%. RF menunjukkan performa yang baik dalam mendeteksi kelas mayoritas maupun minoritas, dengan F1-score mendekati sempurna. Keunggulan RF ini didukung oleh optimasi parameter menggunakan Particle Swarm Optimization (PSO), yang juga diterapkan pada KNN untuk meningkatkan performanya. Proses preprocessing seperti imputasi nilai kosong, normalisasi, dan penyeimbangan data memainkan peran penting dalam meningkatkan akurasi model. Kelemahan KNN terutama terlihat pada kelas minoritas, yang cenderung lebih sulit untuk diklasifikasikan secara akurat. Temuan ini memberikan rekomendasi bahwa RF lebih cocok digunakan untuk aplikasi serupa di masa depan, terutama dalam konteks data multivariabel yang kompleks. Implikasi dari penelitian ini dapat membantu tenaga medis dalam membuat keputusan yang lebih cepat dan akurat, serta membuka peluang untuk pengembangan lebih lanjut dalam prediksi medis berbasis machine learning. Imputasi yang digunakan dapat di ekplore lebih baik dengan pendekatan lain. Scaling yang ditentukan penggunakan robust scaling dapat di explore dengan metode scaling yang juga dapat dilakukan komparasi. Hal secara menyeluruh dalam preprocessing yang dihasilkan perlu ditindak lebih lanjut terkait hasil dengan adanya kewajaran dari hasil imputasi yang diperoleh dapat membuat algoritma yang diterapkan meminimalisir bias pada data yang dihasilkan dalam meminimalisir rasionalitas data yang digunakan didunia nyata.

REFERENCES

[1] I. Handriani, W. Anasari, and L. O. L. Azim, “Pengaruh Faktor Intra Personal Terhadap Pelayanan Kesehatan Ibu,” *J. Kesehat. Saintika Meditory*, vol. 6, no. 1, pp. 51–57, 2022, [Online]. Available: https://jurnal.syedzasaintika.ac.id.

[2] D. P. Sinambela, H. Naparin, M. Zulfadhilah, and N. Hidayah, “Implementasi Algoritma Decision Tree dan Random Forest dalam Prediksi Perdarahan Pascasalin,” *J. Inf. dan Teknol.*, vol. 5, no. 3, pp. 58–64, 2023, doi: 10.60083/jidt.v5i3.393.

[3] T. Włodarczyk *et al.*, “Machine learning methods for preterm birth prediction: A review,” *Electron.*, vol. 10, no. 5, pp. 1–24, 2021, doi: 10.3390/electronics10050586.

[4] C. Orovas *et al.*, "Neural Networks for Early Diagnosis of Postpartum PTSD in Women after Cesarean Section," *Appl. Sci.*, vol. 12, no. 15, pp. 1–15, 2022, doi: 10.3390/app12157492.
[5] T. Wu, H. Fan, H. Zhu, C. You, H. Zhou, and X. Huang, "Intrusion detection system combined enhanced random forest with SMOTE algorithm," *EURASIP J. Adv. Signal Process.*, vol. 2022, no. 1, 2022, doi: 10.1186/s13634-022-00871-6.
[6] A. Altamimi *et al.*, "An automated approach to predict diabetic patients using KNN imputation and effective data mining techniques," *BMC Med. Res. Methodol.*, vol. 24, no. 1, p. 221, 2024, doi: 10.1186/s12874-024-02324-0.
[7] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 677–690, 2022, doi: 10.30812/matrik.v21i3.1726.
[8] Z. Ullah, F. Saleem, M. Jamjoom, and B. Fakieh, "Reliable prediction models based on enriched data for identifying the mode of childbirth by using machine learning methods: Development study," *J. Med. Internet Res.*, vol. 23, no. 6, pp. 1–12, 2021, doi: 10.2196/28856.
[9] L. B. V. de Amorim, G. D. C. Cavalcanti, and R. M. O. Cruz, "The choice of scaling technique matters for classification performance," *Appl. Soft Comput.*, vol. 133, pp. 1–37, 2023, doi: 10.1016/j.asoc.2022.109924.
[10] M. Pal and S. Parija, "Prediction of Heart Diseases using Random Forest," *J. Phys. Conf. Ser.*, vol. 1817, no. 1, 2021, doi: 10.1088/1742-6596/1817/1/012009.
[11] G. W. Jeon, Y. S. Lee, W. H. Hahn, and Y. H. Jun, "A Predictive Model for Perinatal Brain Injury Using Machine Learning Based on Early Birth Data," *Children*, vol. 11, no. 11, 2024, doi: 10.3390/children11111313.
[12] R. A. Kamel *et al.*, "Predicting cesarean delivery for failure to progress as an outcome of labor induction in term singleton pregnancy," *Am. J. Obstet. Gynecol.*, vol. 224, no. 6, pp. 609.e1-609.e11, 2021, doi: 10.1016/j.ajog.2020.12.1212.
[13] I. Santoso, Windu Gata, and Atik Budi Paryanti, "Penggunaan Feature Selection di Algoritma Support Vector Machine untuk Sentimen Analisis Komisi Pemilihan Umum," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 364–370, 2019, doi: 10.29207/resti.v3i3.1084.
[14] S. Nuraeni, S. P. A. Syam, M. F. Wajdi, B. Firmansyah, and M. Malkan, "Implementasi Metode K-NN Untuk Menentukan Jurusan Siswa di SMAN 02 Manokwari," *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, pp. 89–95, 2023, doi: 10.33379/gtech.v7i1.1905.
[15] M. Lipschuetz *et al.*, "Prediction of vaginal birth after cesarean deliveries using machine learning," *Am. J. Obstet. Gynecol.*, vol. 222, no. 6, pp. 613.e1-613.e12, 2020, doi: 10.1016/j.ajog.2019.12.267.
[16] D. M. U. Atmaja, A. R. Hakim, A. Basri, and A. Ariyanto, "Klasifikasi Metode Persalinan pada Ibu Hamil Menggunakan Algoritma Random Forest Berbasis Mobile," *JRST (Jurnal Ris. Sains dan Teknol.*, vol. 7, no. 2, p. 161, 2023, doi: 10.30595/jrst.v7i2.16705.
[17] J. Qiao *et al.*, "A hybrid particle swarm optimization algorithm for solving engineering problem," *Sci. Rep.*, vol. 14, no. 1, pp. 1–30, 2024, doi: 10.1038/s41598-024-59034-2.
[18] L. Pedersen, M. Mazur-Milecka, J. Ruminski, and S. Wagner, "A Review on Machine Learning Deployment Patterns and Key Features in the Prediction of Preeclampsia," *Mach. Learn. Knowl. Extr.*, vol. 6, no. 4, pp. 2515–2569, 2024, doi: 10.3390/make6040123.
[19] I. Nazli, E. Korbeko, S. Dogru, E. Kugu, and O. K. Sahingoz, "Early Detection of Fetal Health Conditions Using Machine Learning for Classifying Imbalanced Cardiotocographic Data," *Diagnostics*, vol. 15, no. 10, pp. 1–26, 2025, doi: 10.3390/diagnostics15101250.
[20] J. H. Wie *et al.*, "Prediction of Emergency Cesarean Section Using Machine Learning Methods: Development and External Validation of a Nationwide Multicenter Dataset in Republic of Korea," *Life*, vol. 12, no. 4, 2022, doi: 10.3390/life12040604.