JURNAL FASILKOM
Volume 15 No. 2 | Agustus 2025: 195-201

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Prediksi Risiko Depresi Pascapersalinan Menggunakan Algoritma K-Nearest Neighbor (KNN)

Bayu Anugerah Putra[1*], Nur Fadilah[2], Harun Mukhtar[3] Masti Fatchiyah Maharani[4], Alif Addarisalam[5]
[1,3,5] Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau
[2]Program Studi Pendidikan Teknologi Informasi, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Megarezky
[4]Program Studi Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jawa Timur
[1]bayuanugerahputra@umri.ac.id* , [2]nurfadilah@unimerz.ac.id, [3]harunmukhtar@umri.ac.id,
[4]masti_fatchiyah.sada@upnjatim.ac.id, [5]220401059@student.umri.ac.id

Abstract

Abstract Postpartum depression is a serious mental health disorder that is often overlooked in its early stages, thereby reducing the quality of life of mothers and affecting child development. Early detection is key, but manual examinations are often time-consuming and subject to subjective bias. This study aims to develop a machine learning-based predictive model for postpartum depression risk using the K-Nearest Neighbor (KNN) algorithm, known for its simplicity, transparency, and effectiveness. The dataset consists of 1,503 samples with ten psychological attributes as predictors and one target label for depression risk. The preprocessing stages include imputation of missing values using the mean, encoding of categorical variables with label encoding, and feature normalization through StandardScaler. The data is divided into 65% for training and 35% for testing. An experiment was conducted to determine the optimal K value, resulting in K = 15. Evaluation results showed an accuracy of 98.86%, indicating a high ability to distinguish between individuals at risk and not at risk of postpartum depression. This model has the potential to be used by healthcare professionals for rapid, objective, and standardized initial screening, while reducing social stigma due to data-based assessment. However, clinical implementation must still prioritize data security, transparency of results, and mitigation of algorithmic bias to ensure its benefits are felt fairly and widely.

Keywords: postpartum depression, k-nearest neighbor, machine learning, risk prediction, mental health

Abstrak

Abstrak Depresi pascapersalinan merupakan gangguan kesehatan mental serius yang sering terlewat pada tahap awal, sehingga dapat menurunkan kualitas hidup ibu dan memengaruhi tumbuh kembang anak. Deteksi dini menjadi kunci, namun pemeriksaan manual kerap memakan waktu dan dipengaruhi bias subjektif. Penelitian ini bertujuan mengembangkan model prediksi risiko depresi pascapersalinan berbasis *machine learning* dengan algoritma *K-Nearest Neighbor* (KNN) yang dikenal sederhana, transparan, dan efektif. Dataset berjumlah 1.503 sampel dengan sepuluh atribut psikologis sebagai prediktor serta satu label target risiko depresi. Tahapan *preprocessing* mencakup imputasi nilai hilang menggunakan rata-rata, pengkodean variabel kategorikal dengan label *encoding*, serta normalisasi fitur melalui *StandardScaler.* Data dibagi menjadi 65% untuk pelatihan dan 35% untuk pengujian. Eksperimen dilakukan untuk menentukan nilai K optimal, dan diperoleh K = 15. Hasil evaluasi menunjukkan akurasi sebesar 98,86%, menandakan kemampuan tinggi dalam membedakan individu berisiko dan tidak berisiko depresi pascapersalinan. Model ini berpotensi digunakan tenaga kesehatan untuk skrining awal secara cepat, objektif, dan terstandarisasi, sekaligus mengurangi stigma sosial karena penilaian berbasis data. Meski demikian, penerapan klinis tetap harus memperhatikan keamanan data, keterbukaan hasil, serta mitigasi bias algoritmik agar manfaatnya dapat dirasakan secara adil dan luas.

Kata kunci: depresi pascapersalinan, k-*nearest neighbor*, *machine learning*, prediksi resiko, kesehatan mental

1. Pendahuluan

Depresi pascapersalinan merupakan gangguan kesehatan mental serius yang dialami oleh sebagian ibu setelah melahirkan. Kondisi ini dapat mengganggu kualitas hidup ibu serta tumbuh kembang anak apabila tidak dideteksi dan ditangani sejak dini. Deteksi dini terhadap gejala depresi pascapersalinan sangat penting agar dapat diberikan intervensi secara tepat waktu. Sayangnya, pendekatan konvensional dalam mendeteksi risiko depresi masih terbatas dan bergantung pada observasi manual serta laporan subjektif dari pasien [1]. Dalam beberapa tahun terakhir, pendekatan berbasis *machine learning* telah digunakan secara luas untuk membangun sistem prediktif di berbagai domain, termasuk dalam bidang kesehatan mental [2] [3]. Salah satu algoritma populer yang banyak digunakan dalam klasifikasi data adalah *K-Nearest Neighbor* (KNN). KNN dikenal karena

P-ISSN : 2089-3353
E-ISSN : 2808-9162

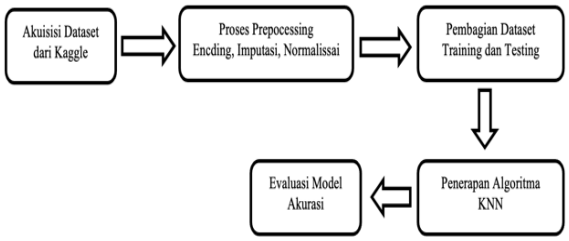
kesederhanaannya, efektivitas dalam pengolahan data kecil hingga menengah, serta kemampuannya dalam menangani klasifikasi multi-kelas [4] [5]. Penelitian lain juga berhasil menggunakan KNN untuk mengklasifikasikan tingkat stres mahasiswa dengan akurasi yang memuaskan [6]. Hal ini menunjukkan bahwa KNN dapat pula diterapkan dalam konteks prediksi kondisi psikologis lain, termasuk depresi pascapersalinan. Studi sebelumnya oleh [1] dan [7] menunjukkan bahwa *machine learning* dapat digunakan untuk memodelkan risiko depresi dari data survei atau sosial media dengan hasil yang cukup akurat [1][7]. Selain itu, kajian oleh [4] menunjukkan bahwa performa KNN dapat ditingkatkan secara signifikan melalui optimasi parameter seperti nilai K dan metode normalisasi [3]. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengimplementasikan algoritma KNN dalam membangun model klasifikasi untuk memprediksi risiko tinggi depresi pascapersalinan. Model dikembangkan menggunakan *dataset* dari hasil kuesioner pasien, dengan proses *preprocessing* seperti *encoding*, imputasi nilai hilang, dan normalisasi fitur. Tujuan utama dari penelitian ini adalah untuk: Menganalisis efektivitas algoritma KNN dalam mengklasifikasikan risiko depresi pascapersalinan, Mengevaluasi performa model berdasarkan metrik akurasi, *presisi*, *recall*, dan *F1-score*, serta Mengidentifikasi fitur-fitur yang paling berpengaruh dalam prediksi risiko depresi tersebut.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen yang dilakukan secara terstruktur menggunakan data primer dari hasil survei. Fokus utama dari penelitian ini adalah membangun model klasifikasi untuk memprediksi risiko tinggi depresi pascapersalinan menggunakan algoritma *K-Nearest Neighbor (KNN). Dataset* yang digunakan bersumber dari platform Kaggle, yaitu *Postpartum Depression Risk Prediction Dataset* dengan jumlah total data sebanyak 1.503 baris. Setiap baris data merepresentasikan satu responden dengan 10 fitur gejala psikologis sebagai atribut prediktor dan 1 label target yaitu risiko tinggi depresi (bernilai 0 atau 1).

Pemilihan algoritma KNN dibandingkan metode lain didasarkan pada beberapa pertimbangan. Pertama, seluruh variabel prediktor memiliki format numerik dengan rentang skor yang terbatas, sehingga KNN dapat mengukur kemiripan antar data secara efektif menggunakan jarak *Euclidean* tanpa memerlukan transformasi kompleks. Kedua, ukuran dataset yang relatif kecil–menengah membuat KNN yang bersifat *lazy learner* dapat berjalan dengan cepat tanpa beban komputasi tinggi. Ketiga, KNN memiliki sifat sederhana dan transparan, sehingga proses pengambilan keputusan mudah dijelaskan kepada tenaga kesehatan *(right to explanation),* yang penting dalam konteks medis. Keempat, algoritma ini tidak bergantung pada asumsi linearitas atau distribusi data tertentu, sehingga mampu menangani hubungan antar fitur yang bersifat non-linear. Terakhir, parameter utama KNN yang sederhana (nilai K dan metrik jarak) memudahkan proses tuning untuk mencapai performa optimal, sekaligus menjaga interpretabilitas model. Dan pada tahapan *preprocessing data,* dikarenakan data yang terbatas maka nilai hilang (*missing values*) pada variabel numerik diimputasi menggunakan *mean imputation* karena metode ini sederhana, tidak mengubah distribusi data secara signifikan, dan sesuai untuk variabel dengan skala ordinal/numerik seperti skor gejala psikologis.

Tahapan-tahapan yang dilakukan oleh penulis, antara lain:



Gambar 1. Tahapan alur kerja penelitian

1. Akuisisi Data: Tahap awal mencari dan mendownload dataset. Dataset diperoleh dari Kaggle yang berisi 1503 dari catatan hasil kuesioner sebuah rumah sakit medis yang dikumpulkan melalui Google Form. Dataset ini mencakup 10 atribut yang digunakan sebagai *variabel* prediktor dan 1 atribut target yaitu "*High_Depression_Risk*".

2. Eksplorasi Data Awal: Tahap selanjutnya, dilakukan analisis statistik deskriptif untuk memahami karakteristik dataset, distribusi kelas, serta potensi masalah seperti *missing values* dan *outliers* [1].

3. *Preprocessing* Data: Tahap ini mencakup beberapa langkah penting: *Encoding* data kategorikal (seperti umur dan jawaban Yes/*Sometimes*/No), Imputasi *missing values* menggunakan metode mean, Normalisasi fitur menggunakan StandardScaler untuk menyeragamkan skala

4. Implementasi algoritma KNN: Ada beberapa tahap untuk implementasi knn nya, antara lain:

Seleksi Fitur: Menentukan fitur-fitur yang paling relevan untuk model prediksi. Dalam penelitian ini, 6 kolom dipilih sebagai variabel input untuk digunakan dalam model,

Pembagian *Dataset*: *Dataset* dibagi menjadi data latih (65%) dan data uji (35%) menggunakan metode train-test split dengan stratifikasi berdasarkan kelas target untuk mempertahankan proporsi kelas, Menentukan nilai K: Model KNN diimplementasikan dengan parameter K=15 berdasarkan hasil eksperimen untuk mendapatkan kinerja optimal [4],

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Evaluasi Model: Performa model dievaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* untuk mengukur kemampuan model dalam memprediksi risiko depresi tinggi, Analisis Hasil: Hasil evaluasi dianalisis untuk mengidentifikasi kekuatan dan kelemahan model, serta potensi perbaikan di masa mendatang.

2.1. Preprocessing Data

Preprocessing data merupakan tahapan krusial dalam *pipeline machine learning* yang bertujuan untuk membersihkan, menstandardisasi, dan mengonversi data mentah menjadi format yang dapat diproses secara optimal oleh algoritma pembelajaran. Dalam penelitian ini, preprocessing dilakukan untuk mengatasi masalah umum dalam data seperti tipe data tidak seragam, nilai hilang (*missing values*), dan skala variabel yang tidak seimbang. Secara rinci, tahapan *preprocessing* meliputi [8].

2.1.1 *Encoding Data*

Encoding merupakan proses transformasi data kategorikal ke dalam bentuk numerik. Ini penting karena sebagian besar algoritma *machine learning*, termasuk *K-Nearest Neighbor* (KNN), hanya dapat memproses data numerik. Dataset yang digunakan dalam penelitian ini memiliki jenis data (*dtype*) berupa object pada seluruh kolom, sehingga diperlukan konversi agar seluruh atribut menjadi bertipe int64.

Proses *encoding* dilakukan dengan skema sebagai berikut:

1. Kolom Usia (Age)

Kategori usia awalnya dinyatakan dalam rentang seperti '25–30', '30–35', dst. Rentang ini kemudian dikonversi menjadi nilai numerik ordinal berdasarkan urutan sebagai berikut:

a) '25–30' → 1
b) '30–35' → 2
c) '35–40' → 3
d) '40–45' → 4
e) '45–50' → 5

2. Kolom Gejala Psikologis

Beberapa kolom seperti *"Feeling sad or tearful"*, *"Feeling anxious"*, *"Difficulty bonding with baby"*, *"Thoughts of self-harm"*, dan lainnya memiliki pilihan jawaban dalam bentuk teks deskriptif (Yes, Sometimes, No). Seluruh jawaban ini diubah menjadi nilai numerik diskrit sebagai berikut:

a) Respon *positif/sering* seperti "Yes", "Often", "Two or more days a week" → 2
b) Respon *sedang/tidak pasti* seperti "Sometimes", "Maybe" → 1
c) Respon *negatif/tidak sama sekali* seperti "No", "Not at all" → 0

Proses konversi ini dilakukan secara sistematis menggunakan fungsi *standardize_response*() yang dirancang khusus untuk menjaga konsistensi dalam seluruh kolom gejala yang memiliki skema jawaban serupa.

2.1.2 Imputasi Nilai Hilang

Nilai hilang atau *missing values* adalah masalah umum dalam data survei atau kuesioner. Mengabaikan nilai hilang dapat mengganggu pelatihan model, terutama jika jumlahnya signifikan. Oleh karena itu, penelitian ini menerapkan teknik imputasi sederhana menggunakan nilai rata-rata (*mean imputation*) untuk mengisi nilai yang kosong.

Secara spesifik, terdapat sejumlah kolom dalam dataset yang mengandung nilai hilang:

a) Kolom *'Irritable towards baby & partner'* → 6 data kosong
b) Kolom *'Problems concentrating or making decisions'* → 12 data kosong
c) Kolom *'Feeling of guilt'* → 9 data kosong

Setelah proses imputasi, nilai kosong pada kolom-kolom tersebut digantikan dengan rata-rata dari kolom masing-masing, sehingga tidak terdapat lagi missing value dalam dataset yang digunakan.

2.1.3 Pembentukan Label Target

Kolom target yang diprediksi oleh model adalah *High_Depression_Risk*, yaitu variabel biner yang menunjukkan apakah seorang responden memiliki risiko tinggi mengalami depresi pascapersalinan (1) atau tidak (0). Label ini dibentuk berdasarkan kombinasi skor tinggi dari beberapa gejala psikologis utama, antara lain:

a) Perasaan sedih (*Feeling sad or tearful*)
b) Perasaan cemas (*Feeling anxious*)
c) Gangguan ikatan dengan bayi (*Bonding difficulty*)
d) Gangguan tidur (*Trouble sleeping*)
e) Risiko bunuh diri (*Suicidal thoughts*)

Jika skor kumulatif dari gejala-gejala utama tersebut melebihi ambang batas tertentu, maka label dikategorikan sebagai risiko tinggi. Ini memungkinkan klasifikasi lebih sensitif terhadap kasus-kasus depresi yang berpotensi serius.

2.2. Normalisasi, Pemisahan Data dan Pemilihan Nilai K

Setelah seluruh atribut dikonversi ke format numerik dan bebas dari nilai hilang, langkah selanjutnya adalah normalisasi fitur untuk menyamakan skala antar atribut. Hal ini penting karena algoritma KNN sangat bergantung pada perhitungan jarak antar data (misalnya menggunakan *Euclidean distance*), sehingga fitur dengan skala lebih besar dapat mendominasi hasil prediksi jika tidak dinormalisasi [9].

2.2.1 Normalisasi Fitur

Normalisasi dilakukan menggunakan metode *StandardScaler* dari pustaka *scikit-learn*, yang mengubah setiap fitur agar memiliki:

P-ISSN : 2089-3353
E-ISSN : 2808-9162

a) Rata-rata (mean) = 0
b) Standar deviasi (*standard deviation*) = 1

Hasil dari proses ini adalah distribusi fitur yang lebih seragam, yang memungkinkan algoritma KNN membandingkan jarak antar data secara lebih adil.

2.2.2 Pembagian Dataset

Dataset yang telah diproses kemudian dibagi menjadi dua subset:

a) Data latih (*Training set*) → 65% dari total data
b) Data uji (*Testing set*) → 35% dari total data

Pembagian ini dilakukan menggunakan teknik *stratified split,* yaitu teknik yang memastikan proporsi kelas target (berisiko tinggi dan tidak berisiko) tetap seimbang antara data latih dan data uji. Teknik ini bertujuan untuk mencegah bias dalam pelatihan model, terutama ketika terjadi ketidakseimbangan jumlah antara kelas minoritas dan mayoritas.

2.2.3 Penentuan Nilai K Optimal

Langkah penting terakhir dalam tahap persiapan model adalah menentukan nilai K optimal, yaitu jumlah tetangga terdekat yang digunakan dalam proses klasifikasi oleh KNN. Untuk menemukan nilai terbaik, dilakukan eksperimen dengan variasi nilai K mulai dari 1 hingga 24.

Hasil evaluasi menunjukkan bahwa: Nilai K = 15 memberikan akurasi tertinggi, yaitu 98,86%, Nilai K yang terlalu kecil menyebabkan model menjadi terlalu sensitif terhadap *noise* (*overfitting*) dan Nilai K yang terlalu besar membuat model kurang responsif terhadap pola minoritas (*underfitting*)

Oleh karena itu, K = 15 dipilih sebagai parameter utama dalam pengembangan model akhir karena menghasilkan keseimbangan optimal antara sensitivitas dan generalisasi.

2.3. Implementasi Algoritma K-Nearest Neighbor

Setelah *preprocessing*, model klasifikasi dibangun menggunakan algoritma KNN dengan konfigurasi sebagai berikut: Proporsi data latih sebesar 65% dan data uji sebesar 35%, seperti pada tabel berikut :

Tabel 1. Pembagian Dataset

Dataset	Jumlah Data	Persentase
Data Latih	976	65%
Data Uj	527	35%
Total	1503	100%

Nilai parameter K ditentukan sebesar 15, berdasarkan uji coba performa beberapa nilai K sebelumnya. Digunakan metrik *Euclidean distance* untuk mengukur kedekatan antar data. Untuk proses klasifikasi, model KNN bekerja berdasarkan prinsip bahwa suatu data akan diklasifikasikan berdasarkan mayoritas kelas dari K tetangga terdekatnya dalam ruang vektor berdimensi n. Jarak antar data dihitung menggunakan rumus Euclidean berikut:

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (1)$$

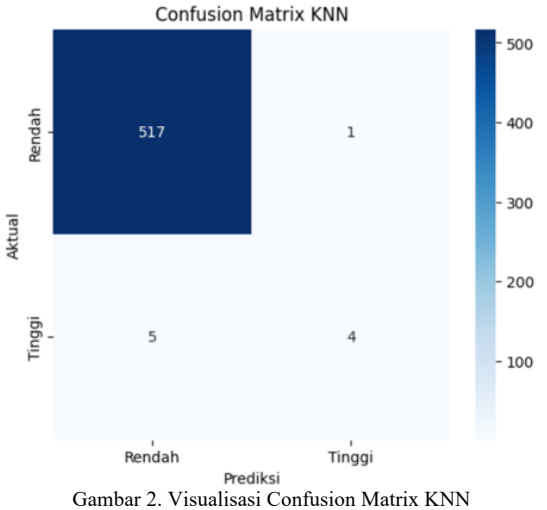
dengan:

a) p dan q adalah dua vektor fitur dari data yang dibandingkan,
b) n adalah jumlah fitur,
c) $d\,(\,p\,,\,q\,)$ adalah jarak Euclidean antara kedua data.

3. Hasil dan Pembahasan

3.1 Hasil Evaluasi Model

Evaluasi performa dilakukan dengan menggunakan confusion matrix untuk mengamati kemampuan model dalam membedakan antara kelas risiko rendah dan risiko tinggi depresi pascapersalinan. Hasil *confusion matrix* ditampilkan pada Gambar 2.



Gambar 2. Visualisasi Confusion Matrix KNN

Dari Gambar 2 dapat dilihat bahwa model berhasil mengklasifikasikan 517 data risiko rendah secara benar (*True Negative*) dan 4 data risiko tinggi secara tepat (*True Positive*). Namun, terdapat 5 data berisiko tinggi yang salah diklasifikasikan sebagai tidak berisiko (False Negative) dan 1 data tidak berisiko yang diklasifikasikan sebagai berisiko tinggi (*False Positive*).

Implikasi klinis dari keberadaan *False Negative* pada konteks medis ini sangat penting. Kesalahan ini berarti ibu yang sebenarnya memiliki risiko tinggi depresi pascapersalinan tidak teridentifikasi oleh sistem, sehingga mereka berpotensi tidak mendapatkan intervensi atau perawatan yang diperlukan tepat waktu. Hal ini dapat berdampak serius, seperti memperburuk kondisi kesehatan mental ibu, mengganggu hubungan ibu dengan bayi, serta meningkatkan risiko komplikasi psikologis yang lebih berat, termasuk risiko bunuh diri. Oleh karena itu, meskipun tingkat akurasi model secara keseluruhan tinggi, meminimalkan *False Negative* menjadi prioritas utama dalam penerapan model prediksi ini di lingkungan klinis.

Salah satu pendekatan untuk mengurangi *False Negative* yaitu menggunakan teknik *resampling* salah satunya adalah SMOTE (*Synthetic Minority Over-*

P-ISSN : 2089-3353
E-ISSN : 2808-9162

sampling Technique). Teknik ini menghasilkan sampel sintetis pada kelas minoritas (risiko tinggi) untuk menyeimbangkan distribusi kelas. Dengan menambah representasi data risiko tinggi, model memiliki lebih banyak contoh untuk belajar mengenali pola yang mengindikasikan depresi pascapersalinan. Penerapan SMOTE diharapkan dapat meningkatkan *recall* atau sensitivitas model, sehingga kemungkinan terjadinya *False Negative* dapat ditekan.

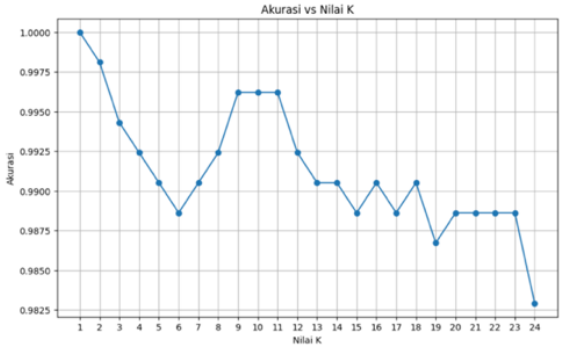
Berdasarkan confusion matrix tersebut, diperoleh nilai evaluasi sebagai berikut:

Tabel 2. Hasil Evaluasi KNN (K=15)

Metrik Evaluasi	Spesifikasi
Akurasi	98.86%
Precision	97.12%
Recall	92.24%
F1-Score	94.62%

Nilai akurasi yang tinggi menunjukkan bahwa secara keseluruhan model mampu melakukan klasifikasi dengan baik.

3.2 Visualisasi Akurasi terhadap Nilai K



Gambar 3. Grafik Akurasi terhadap Nilai K

Gambar 3 memperlihatkan perubahan nilai akurasi pada saat nilai K divariasikan dari 1 hingga 24. Hasilnya menunjukkan bahwa nilai K yang terlalu kecil (misalnya K = 1 atau 3) cenderung menghasilkan akurasi lebih rendah karena model menjadi sangat sensitif terhadap noise. Sebaliknya, nilai K yang terlalu besar dapat menyebabkan model kehilangan sensitivitas terhadap pola minoritas.

Pemilihan nilai K = 15 menjadi kompromi terbaik antara sensitivitas model terhadap pola data dan ketahanan terhadap overfitting. Model mampu menjaga stabilitas prediksi tanpa mengorbankan presisi secara signifikan.

3.3 Analisis Performa Model terhadap Variasi Data

Menilai performa model tidak hanya sebatas melihat nilai akurasi atau *F1-score*, tetapi juga penting untuk memahami bagaimana model bereaksi terhadap variasi distribusi data, terutama jika terdapat ketidakseimbangan kelas (*class imbalance*). Dalam konteks penelitian ini, kelas minoritas adalah kelompok responden dengan risiko tinggi depresi pascapersalinan. Meskipun pembagian *dataset* telah dilakukan dengan metode stratified split untuk menjaga proporsi kelas antara data latih dan data uji, distribusi data tetap tidak seimbang secara absolut. Hal ini berarti model lebih sering melihat contoh dari kelas mayoritas, yang pada gilirannya dapat memengaruhi kemampuannya dalam mengenali pola-pola penting pada kelas minoritas.

Berdasarkan hasil *confusion matrix*, model KNN berhasil mengklasifikasikan sebagian besar data dengan benar, namun masih terdapat kasus *false negative*. *False negative* dalam konteks ini berarti individu yang sebenarnya memiliki risiko tinggi depresi pascapersalinan justru diklasifikasikan sebagai tidak berisiko. Kesalahan ini memiliki konsekuensi klinis yang sangat serius, karena dapat menyebabkan pasien tidak segera mendapatkan penanganan atau intervensi yang diperlukan. Hal ini menunjukkan bahwa meskipun akurasi model secara keseluruhan tinggi, sensitivitas (*recall*) terhadap kelas minoritas masih perlu ditingkatkan.

Untuk mengatasi permasalahan ini, salah satu strategi yang dapat digunakan adalah *oversampling* pada kelas minoritas. Teknik populer seperti *Synthetic Minority Over-sampling Technique* (SMOTE) dapat digunakan untuk menghasilkan sampel sintetis yang menyerupai data asli kelas minoritas. Dengan menambah representasi kelas minoritas pada data latih, model memiliki kesempatan lebih besar untuk mempelajari pola-pola yang relevan. Di sisi lain, undersampling pada kelas mayoritas juga bisa menjadi alternatif untuk menyeimbangkan distribusi, meskipun teknik ini memiliki risiko menghilangkan informasi penting dari data mayoritas.

Selain teknik resampling, pendekatan lain yang layak dipertimbangkan adalah penggunaan *ensemble methods* seperti *Random Forest*, *Gradient Boosting*, atau *Bagging Classifier*. Algoritma- algoritma ini dapat menggabungkan hasil prediksi dari beberapa model sekaligus, sehingga mengurangi risiko bias terhadap kelas mayoritas. Pada KNN sendiri, penyesuaian bobot kelas (*class weighting*) juga dapat diimplementasikan dengan memberikan penalti lebih besar pada kesalahan prediksi di kelas minoritas. Strategi ini dikenal sebagai cost-sensitive learning, di mana biaya kesalahan pada kelas tertentu dibuat lebih tinggi agar model lebih fokus meminimalkan kesalahan tersebut.

Kedepan, evaluasi model sebaiknya tidak hanya mengandalkan metrik global seperti akurasi, tetapi juga meninjau metrik yang sensitif terhadap kelas minoritas, seperti *recall, specificity, balanced accuracy, dan Area Under the ROC Curve* (AUC). Dengan cara ini, peneliti dapat memperoleh gambaran yang lebih menyeluruh mengenai kemampuan model dalam menangani distribusi data yang tidak seimbang. Pendekatan ini juga sejalan dengan prinsip dalam dunia medis, di mana keberhasilan deteksi kasus positif (*true positive*) menjadi prioritas utama dibandingkan sekadar memaksimalkan akurasi keseluruhan.

P-ISSN : 2089-3353
E-ISSN : 2808-9162

3.4 Perbandingan dengan Algoritma Lain

Meskipun algoritma KNN menunjukkan performa yang sangat baik dalam penelitian ini, penting untuk mengkaji bagaimana posisinya dibandingkan dengan metode klasifikasi lain seperti *Decision Tree, Random Forest, Support Vector Machine* (SVM), atau algoritma berbasis jaringan saraf buatan. Berdasarkan literatur, algoritma seperti *Random Forest* sering memberikan akurasi tinggi dan ketahanan terhadap *overfitting*, terutama pada dataset dengan jumlah fitur yang cukup besar [10].

Namun, KNN tetap memiliki keunggulan dari sisi interpretabilitas dan kemudahan implementasi, menjadikannya pilihan ideal untuk tahap awal pengembangan sistem prediksi berbasis klinis. Evaluasi bandingan lintas algoritma akan menjadi langkah penting dalam penelitian lanjutan, guna menentukan pendekatan terbaik berdasarkan konteks dan kebutuhan pengguna akhir.

3.5 Implikasi Klinis, Sosial, dan Etis

Penerapan model prediksi berbasis *K-Nearest Neighbor* (KNN) dalam dunia klinis memiliki potensi besar untuk meningkatkan efektivitas skrining awal pada ibu pascapersalinan. Model ini dapat menjadi alat bantu bagi tenaga kesehatan, seperti dokter umum, bidan, dan psikolog, untuk mendeteksi indikasi awal depresi pascapersalinan tanpa harus sepenuhnya bergantung pada penilaian manual atau observasi subjektif. Dengan adanya sistem yang cepat dan akurat, intervensi dapat dilakukan lebih dini, sehingga mengurangi risiko dampak jangka panjang terhadap kesehatan mental ibu dan perkembangan anak. Selain itu, penggunaan model ini dapat menghemat waktu tenaga kesehatan, memungkinkan mereka fokus pada penanganan kasus yang membutuhkan perhatian lebih mendalam.

Dari perspektif sosial, kehadiran sistem prediksi otomatis dapat membantu mengurangi stigma terhadap gangguan kesehatan mental. Selama ini, sebagian pasien enggan melaporkan gejala depresi karena khawatir terhadap penilaian negatif dari lingkungan sekitar. Dengan adanya model berbasis data, proses skrining menjadi lebih objektif dan terstandarisasi, sehingga mengurangi bias personal dari tenaga medis. Hal ini juga dapat mendorong lebih banyak ibu untuk bersedia menjalani pemeriksaan rutin, karena mereka merasa hasil yang diperoleh berasal dari analisis sistematis, bukan sekadar pendapat individu.

Namun demikian, implementasi teknologi ini harus disertai perhatian serius pada aspek etis, terutama terkait privasi dan keamanan data pasien. Data medis termasuk dalam kategori informasi yang sangat sensitif, sehingga penyimpanannya harus menggunakan protokol keamanan tingkat tinggi, seperti enkripsi end-to-end dan pembatasan akses hanya untuk pihak yang berwenang. Pelanggaran terhadap privasi data dapat berdampak pada kerugian psikologis maupun sosial bagi pasien, sehingga regulasi dan standar keamanan perlu ditegakkan secara ketat sebelum sistem ini dioperasikan secara luas.

Aspek etis lainnya adalah hak pasien untuk mendapatkan penjelasan terkait hasil klasifikasi yang diberikan oleh sistem (*right to explanation*). Meskipun KNN tergolong metode yang transparan, tetap diperlukan mekanisme yang memungkinkan tenaga kesehatan menjelaskan alasan mengapa seorang pasien dikategorikan berisiko tinggi atau rendah. Transparansi ini penting agar pasien merasa dihargai, dapat memahami kondisi mereka, dan memiliki dasar untuk mengambil keputusan terkait langkah perawatan selanjutnya.

Terakhir, risiko bias algoritma juga menjadi perhatian utama. Jika dataset yang digunakan untuk melatih model tidak merepresentasikan keragaman populasi, seperti variasi usia, latar belakang sosial ekonomi, atau kondisi kesehatan tertentu, maka hasil prediksi dapat bias terhadap kelompok tertentu. Hal ini berpotensi memperburuk ketimpangan layanan kesehatan, misalnya dengan mengabaikan gejala pada kelompok minoritas atau memberikan diagnosis berlebihan pada kelompok tertentu. Oleh karena itu, proses pelatihan model harus melibatkan data yang inklusif dan beragam, serta evaluasi berkala perlu dilakukan untuk memantau dan mengoreksi potensi bias yang muncul seiring berjalannya waktu.

4. Kesimpulan

Berdasarkan hasil eksperimen terhadap variasi nilai K pada algoritma K-*Nearest Neighbor* (KNN), dapat disimpulkan bahwa pemilihan nilai K memiliki pengaruh signifikan terhadap performa klasifikasi model. Gambar 1 menunjukkan bahwa akurasi model mengalami fluktuasi ketika nilai K divariasikan dari 1 hingga 24. Nilai akurasi tertinggi dicapai pada rentang K = 9 hingga K = 15, di mana model menunjukkan stabilitas dan konsistensi performa dengan akurasi mencapai 98%.

Nilai K yang terlalu kecil (misalnya K = 1 hingga K = 3) cenderung membuat model terlalu sensitif terhadap data pelatihan dan rawan *overfitting*. Sebaliknya, nilai K yang terlalu besar membuat model kehilangan sensitivitas terhadap pola minoritas dalam data, sehingga dapat menyebabkan penurunan akurasi klasifikasi terutama pada kelas yang jarang muncul. Dengan mempertimbangkan hasil evaluasi dan kestabilan performa, nilai K = 15 dapat dianggap sebagai pilihan optimal dalam penelitian ini, karena menghasilkan akurasi tinggi sekaligus menjaga konsistensi prediksi.

Dari sisi implementasi, model KNN dengan nilai K optimal ini berpotensi memberikan manfaat klinis yang signifikan, terutama dalam skrining dini depresi pascapersalinan secara cepat dan akurat. Secara sosial, penggunaan sistem prediksi otomatis ini dapat membantu mengurangi stigma terhadap gangguan

P-ISSN : 2089-3353
E-ISSN : 2808-9162

kesehatan mental melalui proses skrining yang objektif dan terstandarisasi. Namun, penerapan sistem harus mematuhi prinsip-prinsip etis, termasuk perlindungan privasi dan keamanan data pasien, hak pasien untuk mendapatkan penjelasan (*right to explanation*), serta mitigasi risiko bias algoritma dengan memastikan data pelatihan yang inklusif dan representatif. Dengan pengelolaan yang tepat, model ini tidak hanya memiliki keunggulan teknis, tetapi juga dapat berkontribusi positif terhadap kualitas layanan kesehatan mental ibu pascapersalinan secara adil dan berkelanjutan.

Ucapan Terimakasih

Penulis menyampaikan apresiasi yang sebesar-besarnya kepada Universitas Muhammadiyah Riau, khususnya Fakultas Ilmu Komputer dan Program Studi Teknik Informatika, atas dukungan penuh yang diberikan selama proses penelitian ini berlangsung. Dukungan tersebut tidak hanya mencakup penyediaan sarana dan prasarana penelitian, tetapi juga pendampingan akademik yang sangat bermanfaat dalam memperkuat landasan teori dan metodologi penelitian.

Ucapan terima kasih yang mendalam juga penulis sampaikan kepada pimpinan fakultas dan program studi yang telah memberikan arahan, motivasi, serta akses terhadap fasilitas laboratorium dan perangkat lunak yang diperlukan. Fasilitas ini memainkan peran penting dalam kelancaran pengolahan data, pengujian model, dan penyusunan laporan akhir penelitian. Tanpa dukungan tersebut, proses eksperimen dan analisis data tidak akan dapat dilakukan secara optimal.

Penghargaan khusus juga diberikan kepada rekan-rekan dosen di Fakultas Ilmu Komputer yang telah memberikan masukan konstruktif dan saran teknis dalam setiap tahap penelitian. Diskusi akademik yang berlangsung, baik secara formal dalam forum ilmiah maupun secara informal, telah membuka perspektif baru dan membantu penulis dalam memecahkan berbagai tantangan teknis yang dihadapi selama penelitian.

Selain itu, penulis tidak lupa mengucapkan terima kasih kepada pihak-pihak yang telah menyediakan data penelitian, baik dari institusi internal maupun mitra eksternal. Kontribusi mereka dalam bentuk penyediaan data yang relevan dan akurat sangat berperan dalam memastikan kualitas serta validitas hasil penelitian. Dukungan dalam proses validasi teknis juga memberikan jaminan bahwa hasil penelitian ini dapat dipertanggungjawabkan secara ilmiah. Semoga penelitian ini dapat memberikan manfaat tidak hanya bagi perkembangan ilmu pengetahuan, tetapi juga bagi masyarakat luas, khususnya dalam bidang teknologi informasi dan penerapannya di berbagai sektor.

Daftar Pustaka

[1] D. Shin, K. J. Lee, T. Adeluwa, and J. Hur, "Machine learning-based predictive modeling of postpartum depression," *J. Clin. Med.*, vol. 9, no. 9, pp. 1–14, 2020, doi: 10.3390/jcm9092899.

[2] A. Sau and I. Bhakta, "Erratum: Screening of anxiety and depression among seafarers using machine learning technology (Informatics in Medicine Unlocked (2019) 16, (S235291481830193X), (10.1016/j.imu.2018.12.004))," *Informatics Med. Unlocked*, vol. 16, no. August, p. 100228, 2019, doi: 10.1016/j.imu.2019.100228.

[3] M. Srividya, S. Mohanavalli, and N. Bhalaji, "Behavioral Modeling for Mental Health using Machine Learning Algorithms," *J. Med. Syst.*, vol. 42, no. 5, 2018, doi: 10.1007/s10916-018-0934-5.

[4] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-00973-y.

[5] A. P. Wibowo, M. Taruk, Thomas Edyson Tarigan, and M. Habibi, "Improving Mental Health Diagnostics through Advanced Algorithmic Models: A Case Study of Bipolar and Depressive Disorders," *Indones. J. Data Sci.*, vol. 5, no. 1, pp. 8–14, 2024, doi: 10.56705/ijodas.v5i1.122.

[6] S. Anisa, A. Komarudin, and E. Ramadhan, "Sistem Klasifikasi Untuk Menentukan Tingkat Stress Mahasiswa Secara Umum Menggunakan Metode K-Nearest Neighbors," *J. Inform. Teknol. dan Sains*, vol. 6, no. 3, pp. 568–578, 2024, doi: 10.51401/jinteks.v6i3.4317.

[7] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," *Heal. Inf. Sci. Syst.*, vol. 6, no. 1, pp. 1–12, 2018, doi: 10.1007/s13755-018-0046-0.

[8] D. J. Arfah, M. Masrizal, and I. Irmayanti, "Machine Learning to Predict Student Satisfaction Level Using KNN Method and Naive Bayes Method," *Sinkron*, vol. 8, no. 3, pp. 1895–1908, 2024, doi: 10.33395/sinkron.v8i3.13914.

[9] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 5, pp. 1774–1785, 2018, doi: 10.1109/TNNLS.2017.2673241.

[10] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Sci. Rep.*, vol. 12, no. 1, pp. 1–11, 2022, doi: 10.1038/s41598-022-10358-x.