

Received: 17 Juli 2026; Accepted: 15 Agustus 2026; Published: 30 September 2026
DOI: 10.36350/jskom.v2i3.125

Deteksi Dini Penyakit Preeklampsia pada Ibu Hamil dalam Mendukung Penegakan Diagnosis Menggunakan Pendekatan Algoritma K-Nearest Neighbor (KNN)

Jihan Anindia Maharani^{1*}, Hardi Jamhur²

^{1,2}Program Studi Sistem Informasi, Fakultas Informatika dan Komputer, Universitas Binaniaga Indonesia

Email¹: zuyizujihan208@gmail.com

*Corresponding Author

ABSTRACT

Preeclampsia is a pregnancy complication that may increase the risk of maternal and fetal mortality if it is not detected and managed early. At the primary healthcare level, screening still relies on healthcare workers' assessment of examination results. Therefore, a data-based approach is needed to support the identification of preeclampsia risk. This study aims to develop a website-based prototype that applies the K-Nearest Neighbor (KNN) algorithm to support the early detection of preeclampsia in pregnant women. The research method used was Research and Development (R&D) with a Prototype model, while data processing followed the CRISP-DM stages. The data were obtained from Puskesmas and consisted of age, parity, height, weight, body mass index, systolic blood pressure, diastolic blood pressure, Mean Arterial Pressure (MAP), random blood glucose, and urine protein. The modeling results showed that K=3 achieved an accuracy of 99.81%, precision of 100.00%, recall of 98.97%, and F1-score of 99.48%. System testing obtained 95% in Black Box Testing, 100% in White Box Testing, and 82.8% in user testing using the Post-Study System Usability Questionnaire (PSSUQ). The developed prototype is expected to help healthcare workers conduct more objective screening and support the diagnostic process at Puskesmas.

Keywords: *Preeclampsia, Early Detection, K-Nearest Neighbor, Data Mining, Decision Support System.*

ABSTRAK

Preeklampsia merupakan salah satu komplikasi kehamilan yang menjadi penyebab utama tingginya angka kematian ibu di Indonesia. Kondisi ini sering terlambat terdeteksi sehingga meningkatkan risiko terjadinya komplikasi pada ibu maupun janin. Di tingkat puskesmas, proses skrining masih mengandalkan penilaian tenaga kesehatan berdasarkan hasil pemeriksaan, sehingga diperlukan sistem pendukung keputusan yang mampu membantu tenaga kesehatan melakukan deteksi dini risiko preeklampsia secara lebih objektif berdasarkan data hasil skrining. Penelitian ini bertujuan mengembangkan prototype sistem berbasis website untuk mendukung deteksi dini risiko preeklampsia pada ibu hamil menggunakan algoritma K-Nearest Neighbor (KNN). Metode penelitian yang digunakan adalah Research and Development (R&D) dengan model Prototype, sedangkan proses pengolahan data mengacu pada tahapan CRISP-DM. Variabel yang digunakan meliputi usia, paritas, indeks massa tubuh (IMT), tekanan darah sistolik, tekanan darah diastolik, Mean Arterial Pressure (MAP), kadar glukosa darah sewaktu (GDS), dan protein urin. Hasil pemodelan menunjukkan bahwa algoritma KNN dengan nilai k=3 menghasilkan akurasi sebesar 99,81%, precision 100,00%, recall 98,97%, dan F1-score 99,48%. Berdasarkan hasil prediksi terhadap data penelitian, model mengklasifikasikan 443 ibu hamil ke dalam kategori normal dan 96 ibu hamil ke dalam kategori preeklampsia. Hasil pengujian sistem memperoleh nilai 95% pada Black Box Testing, 100% pada White Box Testing, serta 82,8% pada uji pengguna menggunakan Post-Study System Usability Questionnaire (PSSUQ) dengan kategori Sangat Layak. Berdasarkan hasil tersebut, prototype yang dikembangkan dinyatakan layak digunakan sebagai sistem pendukung keputusan untuk membantu tenaga kesehatan dalam mendukung deteksi dini risiko preeklampsia dan penegakan diagnosis di puskesmas.

Kata kunci: Preeklampsia, Deteksi Dini, K-Nearest Neighbor (KNN), Sistem Pendukung Keputusan, Prototype

A. PENDAHULUAN

Angka Kematian Ibu (AKI) adalah jumlah kematian ibu yang disebabkan oleh kehamilan, persalinan, dan 42 hari pasca persalinan [1]. Berdasarkan artikel World Health Organization (WHO) yang berjudul Maternal Mortality, pada tahun 2023 angka kematian ibu (AKI) global mencapai 197 per 100.000 kelahiran hidup.

Melalui program Sustainable Development Goals (SDGs), WHO menetapkan target untuk menurunkan AKI global hingga di bawah 70 per 100.000 kelahiran hidup pada tahun 2030. Namun, untuk mencapai target tersebut dibutuhkan tingkat penurunan hampir 15%, yang merupakan capaian yang sulit diperoleh di tingkat nasional karena dipengaruhi oleh ketimpangan akses layanan kesehatan, kualitas penanganan medis, serta keterlambatan dalam mendeteksi komplikasi pada kehamilan [2].

Di Indonesia, angka kematian ibu pada tahun 2024 berada pada kisaran 305 per 100.000 kelahiran hidup. Angka tersebut menjadikan Indonesia sebagai salah satu negara dengan angka kematian ibu tertinggi di ASEAN [3]. Sekitar 24% kasus kematian ibu di Indonesia disebabkan oleh preeklampsia dan eklampsia [3]. Preeklampsia merupakan komplikasi kehamilan yang ditandai dengan peningkatan tekanan darah disertai adanya protein dalam urin. Kondisi ini umumnya muncul setelah usia kehamilan lebih dari 20 minggu, namun dalam kasus tertentu dapat terjadi lebih awal. Apabila tidak terdeteksi dan ditangani secara dini, preeklampsia dapat berkembang menjadi komplikasi serius yang berisiko menyebabkan kematian ibu maupun janin [4].

Terdapat dua faktor utama yang menyebabkan tingginya angka kematian ibu di Indonesia, yaitu keterlambatan dalam menegakkan diagnosis dan keterlambatan dalam mengakses fasilitas pelayanan kesehatan yang didukung oleh sarana dan prasarana yang memadai. Pada fasilitas kesehatan seperti puskesmas, deteksi preeklampsia pada ibu hamil yang tidak menunjukkan keluhan klinis umumnya dilakukan ketika usia kehamilan telah memasuki akhir trimester kedua atau ketiga. Kondisi tersebut menyebabkan risiko preeklampsia berpotensi tidak teridentifikasi sejak dini sehingga intervensi yang dapat diberikan pada tahap awal kehamilan menjadi kurang optimal dalam mencegah perkembangan komplikasi yang lebih berat.

Tabel 1 Data Klinis Ibu Hamil

No	Nama / Usia / Paritas	TB	BB	IMT	Sistol / Diastol	MAP	GDS	Protein urin	Keterangan
1	PI030/27/1	156	56.4	23.2	140/117	125	96	Positif1	Normal
2	PI031/22/1	161	88.6	34.2	120/78	92	84	Negatif	Normal
3	PI032/26/1	158	61.5	24.6	101/52	68	165	Negatif	Normal
4	PI033/30/0	150	45.2	20.1	102/62	75	108	Negatif	Normal
...	...	...	...	...	...	...	...	...	...
5	PI046/30/2	158	73.5	29.4	133/85	101	118	Positif1	Normal
6	PI047/30/2	168	66.1	23.4	141/93	109	82	Positif2	Normal
7	PI048/34/1	155	64	26.6	107/65	79	173	Negatif	Normal
8	PI049/25/1	156	51	21	100/66	77	86	Negatif	Normal
9	PI050/32/1	157	71.4	29	131/76	94	125	Negatif	Normal

Berdasarkan data klinis ibu hamil yang terdapat pada Tabel 1, terlihat adanya ketidaktepatan dalam mengelompokkan pasien ke dalam kategori preeklampsia. Contoh pada pasien dengan inisial PI047 menunjukkan hasil pemeriksaan protein urin Positif2 dengan tekanan darah sistol 141 dan diastol 93 diaktergorikan sebagai pasien normal, yang seharusnya mengindikasikan bahwa ibu hamil tersebut terdeteksi mengidap preeklampsia. Objek penelitian mengemukakan mengalami kesulitan dalam mendeteksi dini penyakit preeklampsia yang saat ini masih dilakukan dengan cara konvensional, karena penyakit ini dapat muncul sewaktu-waktu tanpa gejala yang jelas. Padahal, fasilitas pelayanan kesehatan telah memiliki data pemeriksaan ibu hamil yang cukup lengkap dan tersimpan dalam jumlah besar, seperti usia ibu, paritas, tinggi badan, berat badan, indeks massa tubuh, tekanan darah, mean arterial pressure, serta hasil pemeriksaan laboratorium. Namun data tersebut belum dimanfaatkan untuk melakukan deteksi dini preeklampsia secara sistematis. Oleh karena itu, diperlukan suatu pendekatan berbasis data mining yang mampu mengolah dan menganalisis berbagai faktor risiko untuk menghasilkan informasi yang dapat mendukung proses deteksi dini preeklampsia.

Data mining merupakan proses untuk menemukan pola, tren, dan informasi berharga dari kumpulan data yang besar dan kompleks untuk menghasilkan informasi yang dapat dimanfaatkan. Salah satu metode yang dapat diterapkan dalam proses klasifikasi adalah K-Nearest Neighbor (KNN). KNN merupakan algoritma machine learning yang mengklasifikasikan data baru berdasarkan kedekatan karakteristiknya dengan data pada set pelatihan. Pada penelitian ini, metode KNN digunakan untuk mengklasifikasikan ibu hamil ke dalam dua kategori, yaitu normal dan preeklampsia, berdasarkan karakteristik dan faktor risiko yang terdapat pada data

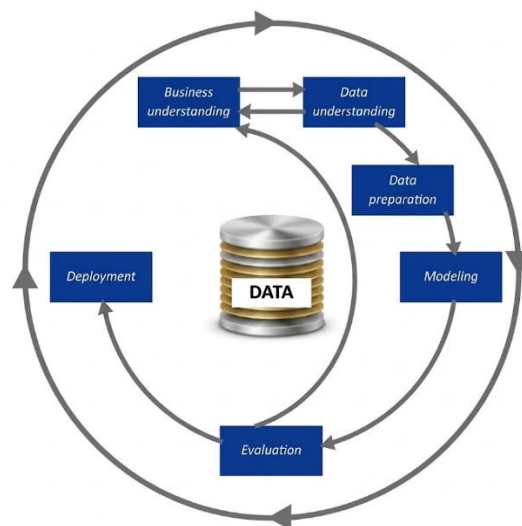
pemeriksaan. Penerapan KNN diharapkan dapat mendukung tenaga kesehatan dalam melakukan deteksi dini preeklampsia secara lebih objektif dan konsisten.

Berdasarkan permasalahan tersebut, penelitian ini dilakukan untuk menerapkan algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan kondisi ibu hamil berdasarkan faktor risiko preeklampsia. Penelitian ini juga mengembangkan prototype sistem deteksi dini yang dapat membantu proses pengolahan data dan memberikan hasil klasifikasi secara terkomputasi. Dengan demikian, hasil penelitian diharapkan dapat memberikan dukungan bagi tenaga kesehatan dalam melakukan deteksi dini preeklampsia serta memanfaatkan data pemeriksaan ibu hamil secara lebih optimal.

B. METODE PENELITIAN

1. Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi dari data dalam jumlah besar. Dalam proses data mining, ada satu atau lebih metode pembelajaran mesin komputer yang bekerja untuk mengambil dan menganalisa pengetahuan secara langsung. Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan metodologi dengan pendekatan yang terstruktur dan sistematis secara terorganisir untuk pelaksanaan proses ekstraksi pengetahuan dari data untuk menjalankan proyek data mining yang diadaptasi untuk berbagai industri dan aplikasi [5]. Model proses dalam CRISP-DM menggambarkan siklus hidup proyek data mining secara menyeluruh, yang terdiri atas sejumlah tahapan dengan tugas-tugas tertentu serta keterkaitan antar tahapan. Pada Gambar 1 menunjukkan tahapan-tahapan CRISP-DM sebagai berikut [6]:



Gambar 1 Tahapan-tahapan CRISP-DM
(sumber: Sumarauw S)

- (1) business understanding, bertujuan memahami kebutuhan dan tujuan proyek dari sudut pandang bisnis, menggambarkan masalah nyata yang ingin dipecahkan, menetapkan tujuan data mining, dan merencanakan strategi untuk mencapai tujuan tersebut;
- (2) data understanding, fokus pada pengumpulan dan eksplorasi data awal, memahami kualitas dan karakteristik data yang tersedia, termasuk mencari data yang hilang dan anomali;
- (3) data preparation, pada fase ini data dipilih, dibersihkan, dikonstruksi, dan diformat agar sesuai untuk proses pemodelan. ini adalah fase yang paling memakan waktu dalam proyek data mining;
- (4) modeling, adalah fase pemilihan dan penerapan teknik pemodelan, di mana parameter dan algoritma diatur serta model dibangun berdasarkan data yang sudah dipersiapkan;
- (5) evaluation, bertujuan memastikan model yang dihasilkan valid dan memenuhi tujuan bisnis menggunakan metrik yang sesuai dan review menyeluruh terhadap proses;
- (6) deployment, fase implementasi hasil data mining ke dalam proses bisnis, pemantauan model setelah implementasi, serta dokumentasi dan pelaporan hasil proyek.

2. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) adalah algoritma klasifikasi yang memprediksi kategori data baru dengan mengacu pada kedekatan jarak terhadap data yang sudah ada. Algoritma ini menetapkan kelas suatu data dengan melihat tingkat kesamaan atau jarak terdekatnya dengan data pada set pelatihan. KNN termasuk

dalam metode supervised learning, di mana klasifikasi data uji dilakukan dengan mempertimbangkan kelas mayoritas dari sejumlah tetangga terdekat yang dihitung berdasarkan jarak [7]. KNN dikenal sebagai algoritma machine learning yang sederhana tetapi cukup efektif untuk menyelesaikan masalah klasifikasi. Penerapan KNN bertujuan untuk menempatkan objek baru ke dalam kelas tertentu berdasarkan atribut yang dimilikinya serta pola yang terbentuk dari data latih berdasarkan jarak. Salah satu metode yang sering digunakan untuk menghitung jarak pada metode KNN adalah Euclidean distance. Berikut adalah rumus Euclidean distance:

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

keterangan:

$d(x, y)$ = jarak kedekatan antar data

x_i = data latih ke i

y_i = data uji ke i

n = jumlah atribut 1 sampai n

3. Confusion Matrix

Dalam membangun sebuah model klasifikasi, penting untuk mempertimbangkan cara menghasilkan model yang dapat memberikan hasil secara optimal dan akurat, sehingga diperlukan metode evaluasi sebelum model diterapkan. Metode evaluasi yang digunakan adalah confusion matrix. Confusion matrix adalah sebuah metode evaluasi performa model klasifikasi dengan menunjukkan jumlah prediksi yang benar dan salah dari model klasifikasi pada setiap kelas klasifikasi. Terdapat empat komponen utama, yaitu *true positive* (TP) merupakan data positif yang diprediksi dengan benar, *true negative* (TN) merupakan data negatif yang diprediksi dengan benar, *false positive* (FP) merupakan data negatif yang diprediksi sebagai positif, dan *false negative* (FN) merupakan data positif yang diprediksi sebagai negatif [8]. Kemudian empat komponen tersebut digunakan sebagai acuan untuk memperoleh nilai akurasi, presisi, recall, dan F1-Score dari hasil klasifikasi. Berikut adalah rumus untuk menghitung nilai akurasi:

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

Berikut adalah rumus untuk menghitung nilai presisi:

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

Berikut adalah rumus untuk menghitung nilai recall:

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

Berikut adalah rumus untuk menghitung nilai F1-Score:

$$F1 - score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$$

C. HASIL DAN PEMBAHASAN

1. Hasil

a. Business understanding

Tahap pertama yang dilakukan adalah menganalisis kebutuhan melalui wawancara dengan tenaga kesehatan mengenai proses skrining preeklampsia yang sedang berjalan. Berdasarkan hasil wawancara, masalah utama yang teridentifikasi adalah pengambilan keputusan diagnosis yang masih bersifat subjektif. Penentuan apakah seorang ibu hamil mengalami preeklampsia atau tidak sangat bergantung pada pengalaman dan penilaian pribadi tenaga kesehatan. Selain itu, pencatatan data yang dilakukan secara manual dan berulang meningkatkan risiko kesalahan serta kurang efisien. Kondisi tersebut

menunjukkan perlunya sistem yang dapat membantu tenaga akesehatan dalam melakukan deteksi dini preeklampsia secara lebih objektif, efekrif, dan tepat.

b. Data Understanding

Data yang digunakan dalam penelitian ini berjumlah 601 data ibu hamil yang diperoleh dari puskesmas. Dari total data tersebut, terdapat 97 data pasien dengan diagnosis preeklampsia dan 504 data pasien dengan kondisi normal. Terdapat sebelas variabel yang digunakan dalam penelitian ini, yaitu usia, paritas, tinggi badan, berat badan, indeks massa tubuh, sistol, diastol, mean arterial pressure (MAP), gula darah sewaktu (GDS), protein urin, dan keterangan (label). Berikut penjelasan mengenai variabel ditunjukkan pada Tabel 2:

Tabel 2 Variabel Penelitian

Variabel	Definisi
Usia	Usia ibu hamil
Paritas	Jumlah persalinan seorang wanita yang telah mencapai usia kehamilan layak hidup, baik persalinan hidup atau tidak
Berat Badan	Berat badan ibu hamil yang diukur dalam satuan kilogram (kg)
Tinggi Badan	Tinggi badan ibu hamil yang diukur dalam satuan sentimeter (cm)
Indeks Massa Tubuh (IMT)	Nilai yang menggambarkan kategori berat badan seseorang yang dihitung dari berat badan dan tinggi badan
Sistolik	Angka yang menunjukkan tekanan saat jantung berdetak dan memompa darah
Diastolik	Angka yang menunjukkan tekanan saat jantung beristirahat di antara detakan
Mean Arterial Pressure (MAP)	Rata-rata tekanan darah arteri yang menggambarkan aliran darah ke organ vital selama satu siklus jantung
Gula Darah Sewaktu (GDS)	Kadar glukosa dalam darah
Protein Urin	Kandungan protein di dalam urin
Keterangan	Kategori hasil klasifikasi yang terdiri dari Nomal dan Preeklampsia

c. Data Preperation

Tahap selanjutnya adalah mempersiapkan data sebeum digunakan dalam proses pemodelan menggunakan algoritma K-Nearest Neighbor. Tahapan ini meliputi pembagian data, tranformasi data, dan normalisasi data.

Pada tahapan pembagian data, data dibagi menjadi data training dan data testing. Data training digunakan pada tahap modelling dan hanya berisi data yang valid, sedangkan data testing digunakan pada tahap deployment dan berisi data yang diragukan hasil labelnya. Pembagian data dilakukan secara manual berdasarkan pedoman skrining preeklampsia pada Buku KIA serta hasil wawancara dengan narasumber. Adapun kriteria yang digunakan dalam pembagian data meliputi ibu hamil dengan tekanan darah sistolik ≥ 140 mmHg, diastolik ≥ 90 mmHg, IMT > 30 , nullipara, paritas > 3 , MAP > 90 mmHg, protein urin > 1 , gula darah sewaktu ≥ 140 mg/dL, dan usia > 35 tahun. Selanjutnya, data dibagi dengan perbandingan 90:10. Berdasarkan pembagian data tersebut, dari total 601 data diperoleh 539 data training dan 62 data testing. Berikut ini sample data training dan data testing ditunjukkan pada Tabel 3 dan Tabel 4.

Tabel 3 Sample data Training

No	Nama	Usia	Paritas	Tb	Bb	Imt	Sistolik	Diastolik	Map	Protein_urine	Gds	Status
1	PI001	29	1	149	74.5	33.6	152	106	121	Positif2	124	Preeklampsia
2	PI002	38	3	150	57.4	25.5	128	79	95	Negatif	158	Normal
3	PI006	35	3	156	76.2	31.3	126	82	97	Negatif	111	Normal
...	...	...	...	...	...	...	...	...	...	...	...	...
537	PI597	29	0	156	51.9	21.3	144	92	109	Positif1	90	Preeklampsia
538	PI598	25	1	159	90	35.6	142	108	119	Positif1	87	Preeklampsia
539	PI600	34	2	153	93	39.7	162	106	125	Positif1	101	Preeklampsia

Tabel 4 Sample Data Testing

No	Nama	Usia	Paritas	Tb	Bb	Imt	Sistolik	Diastolik	Map	Protein_urine	Gds	Status
1	PI003	30	0	160	82.1	32.1	150	91	111	Negatif	77	Normal
2	PI004	20	1	155	62	25.8	110	80	90	Positif2	152	Normal
3	PI007	29	1	150	58	25.8	103	71	82	Positif1	125	Normal
...	...	...	...	...	...	...	...	...	...	...	...	...
60	PI595	33	1	165	92.4	33.9	143	87	106	Positif1	102	Normal
61	PI599	30	1	150	72.6	32.3	141	84	103	Positif2	97	Normal
62	PI601	31	0	158	69.2	27.7	151	94	113	Positif2	129	Normal

Setelah pembagian data, dilakukan tranformasi pada variabel protein urine menggunakan metode One-hot encoding. Tranformasi ini bertujuan mengubah data kategorik menjadi numerik sehingga dapat digunakan dalam proses pemodelan. Setiap kategori pada variabel protein urine dipresentasikan dalam bentuk angka biner. Hasil tranformasi dapat dilihat pada Tabel 5 dan Tabel 6.

Tabel 5 Transformasi Data Training

No	Nama	Usia	Paritas	Tb	Bb	Imt	Sistolik	Diastolik	Map	Protein_urine_n	Protein_urine_1	Protein_urine_2	Protein_urine_3	Gds	Status
1	PI001	29	1	149	74.5	33.6	152	106	121	0	0	1	0	124	Preeklampsia
2	PI002	38	3	150	57.4	25.5	128	79	95	1	0	0	0	158	Normal
3	PI006	35	3	156	76.2	31.3	126	82	97	1	0	0	0	111	Normal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
537	PI597	29	0	156	51.9	21.3	144	92	109	0	1	0	0	90	Preeklampsia
538	PI598	25	1	159	90	35.6	142	108	119	0	1	0	0	87	Preeklampsia
539	PI600	34	2	153	93	39.7	162	106	125	0	1	0	0	101	Preeklampsia

Tabel 6 Tranformasi Data Testing

No	Nama	Usia	Paritas	Tb	Bb	Imt	Sistolik	Diastolik	Map	Protein_urine_n	Protein_urine_1	Protein_urine_2	Protein_urine_3	Gds	Status
1	PI003	30	0	160	82.1	32.1	150	91	111	1	0	0	0	124	Normal
2	PI004	20	1	155	62	25.8	110	80	90	0	0	1	0	158	Normal
3	PI007	29	1	150	58	25.8	103	71	82	0	1	0	0	111	Normal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
60	PI595	33	1	165	92.4	33.9	143	87	106	0	1	0	0	90	Normal
61	PI599	30	1	150	72.6	32.3	141	84	103	0	0	1	0	87	Normal
62	PI601	31	0	158	69.2	27.7	151	94	113	0	0	1	0	101	Normal

Selanjutnya, data dilakukan normalisasi untuk menyamakan skala setiap variabel numerik agar seluruh fitur memiliki rentang nilai yang setara. Pada penelitian ini, metode normalisasi yang digunakan adalah MinMax Scaler. Metode ini mengubah nilai data ke dalam rentang 0 sampai 1, sehingga tidak ada variabel yang mendominasi perhitungan hanya karena memiliki nilai yang lebih besar. Proses ini sangat penting pada algoritma KNN karena metode tersebut bekerja berdasarkan perhitungan jarak antar data, sehingga perbedaan skala antar variabel dapat memengaruhi hasil klasifikasi. Data training dan data testing yang telah melalui proses normalisasi dapat dilihat pada Tabel 7 dan Tabel 8.

Tabel 7 Normalisasi Data Training

No	Nama	Usia	Paritas	Tb	Bb	Imt	Sistolik	Diastolik	Map	Protein_urine_n	Protein_urine_1	Protein_urine_2	Protein_urine_3	Gds	Status
1	PI001	0.48	0.25	0.35	0.48	0.59	0.71	1	0.90	0	0	1	0	0.65	Preeklampsia
2	PI002	0.81	0.50	0.41	0.07	0.08	0.38	0.51	0.47	1	0	0	0	0.33	Normal
3	PI006	0.69	0.5	0.59	0.11	0.11	0.36	0.54	0.49	1	0	0	0	0.15	Normal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
537	PI597	0.46	0	0.59	0.06	0.06	0.50	0.61	0.60	0	1	0	0	0.07	Preeklampsia
538	PI598	0.31	0.17	0.68	0.14	0.13	0.49	0.73	0.69	0	1	0	0	0.06	Preeklampsia
539	PI600	0.67	0.50	0.50	0.75	0.84	0.83	1	0.96	0	1	0	0	0.34	Preeklampsia

Tabel 8 Normalisasi Data Testing

No	Nama	Usia	Paritas	Tb	Bb	Imt	Sistolik	Diastolik	Map	Protein_urine_n	Protein_urine_1	Protein_urine_2	Protein_urine_3	Gds	Status
1	PI003	0.43	0	0.66	0.57	0.5	0.78	0.57	0.72	1	0	0	0	0.00	Normal
2	PI004	0	0.25	0.48	0.26	0.27	0.19	0.41	0.37	0	0	1	0	0.58	Normal
3	PI007	0.39	0.25	0.31	0.20	0.27	0.09	0.28	0.23	0	0	1	0	0.37	Normal

No	Nama	Usia	Paritas	Tb	Bb	Imt	Sistolik	Diastolik	Map	Protein_urine_n	Protein_urine_1	Protein_urine_2	Protein_urine_3	Gds	Status
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
60	PI595	0.57	0.25	0.83	0.73	0.57	0.68	0.51	0.63	0	1	0	0	0.19	Normal
61	PI599	0.43	0.25	0.31	0.43	0.51	0.65	0.46	0.58	0	0	1	0	0.16	Normal
62	PI601	0.48	0	0.59	0.37	0.34	0.79	0.61	0.75	0	0	1	0	0.40	Normal

d. Modelling

Pada tahap pemodelan dilakukan pembentukan model menggunakan algoritma K-Nearest Neighbor (KNN) dengan 539 data training, yang terdiri dari 442 data pasien normal dan 97 pasien preeklampsia. Tahap awal yang dilakukan adalah dengan mencari nilai k terbaik menggunakan metode Stratified K-Fold Cross Validation. Metode ini membagi data ke dalam beberapa fold dengan tetap mempertahankan proporsi kelas pada setiap bagian, sehingga hasil evaluasi model menjadi lebih stabil dan seimbang dalam menilai kemampuan klasifikasi terhadap kedua kelas. Setelah dilakukan pengujian terhadap beberapa nilai k, diperoleh hasil dari setiap nilai k seperti yang ditunjukkan pada Tabel 9:

Tabel 9 Nilai Akurasi dari Setiap Nilai K

Nilai k	Akurasi
3	98.61
5	98.38
7	98.38
9	97.68
11	97.68
13	97.68
15	97.68
17	97.68
19	97.45
21	97.45
23	97.45

Berdasarkan hasil akurasi pada nilai k seperti yang tertera pada Tabel 9, diperoleh bahwa nilai k terbaik adalah k = 3 dengan akurasi sebesar 98.61%. Oleh karena itu, nilai k = 3 dipilih sebagai parameter terbaik dalam proses klasifikasi pada penelitian ini. Setelah nilai k terbaik diperoleh, tahap selanjutnya menguji data training dengan data training itu sendiri. Sehingga terdapat 539 data training sebagai referensi dan 539 data training yang sama digunakan sebagai data uji. Meskipun data yang digunakan memiliki isi yang sama, pada proses implementasi keduanya ditempatkan pada file yang berbeda agar sistem dapat menjalankan proses prediksi.

Pengujian ini bertujuan untuk mengetahui hasil prediksi model terhadap data yang telah dipelajari sekaligus memastikan bahwa model telah siap digunakan sebelum dilakukan pengujian menggunakan data uji yang belum pernah dipelajari. Pada Tabel 10 dan Tabel 11 merupakan sample data training yang akan digunakan pada contoh perhitungan algoritma K-Nearest Neighbor:

Tabel 10 Data Training Tahap Pemodelan

No	Nama	Usia	Paritas	Tb	Bb	Imt	Sistolik	Diastolik	Map	Protein_urine_n	Protein_urine_1	Protein_urine_2	Protein_urine_3	Gds	Status
1	PI001	0.48	0.25	0.35	0.48	0.59	0.71	1	0.90	0	0	1	0	0.65	Preeklampsia
2	PI002	0.81	0.50	0.41	0.07	0.08	0.38	0.51	0.47	1	0	0	0	0.33	Normal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
539	PI600	0.67	0.50	0.50	0.75	0.84	0.83	1	0.96	0	1	0	0	0.34	Preeklampsia

Tabel 11 Data Testing Tahap Pemodelan

No	Nama	Usia	Paritas	Tb	Bb	Imt	Sistolik	Diastolik	Map	Protein_urine_n	Protein_urine_1	Protein_urine_2	Protein_urine_3	Gds	Status
1	PI001	0.48	0.25	0.35	0.48	0.59	0.71	1	0.90	0	0	1	0	0.65	?

Penerapan perhitungan manual algoritma KNN dilakukan dengan mencari jarak antar data menggunakan euclidean distance dan menentukan hasil klasifikasi berdasarkan mayoritas label pada sejumlah tetangga terdekat. Berikut adalah contoh perhitungan manual algoritma K-Nearest Neighbor:

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

$$d(x, y) = \sqrt{(0.48 - 0.48)^2 + (0.25 - 0.25)^2 + (0.35 - 0.35)^2 + (0.48 - 0.48)^2 + (0.59 - 0.59)^2 + (0.71 - 0.71)^2 + (1 - 1)^2 + (0.90 - 0.90)^2 + (0 - 0)^2 + (0 - 0)^2 + (1 - 1)^2 + (0 - 0)^2 + (0.65 - 0.65)^2}$$

$$= 0$$

$$d(x, y) = \sqrt{(0.81 - 0.48)^2 + (0.50 - 0.25)^2 + (0.41 - 0.35)^2 + (0.07 - 0.48)^2 + (0.08 - 0.59)^2 + (0.38 - 0.71)^2 + (0.51 - 1)^2 + (0.47 - 0.90)^2 + (1 - 0)^2 + (0 - 0)^2 + (0 - 1)^2 + (0 - 0)^2 + (0.33 - 0.65)^2}$$

$$= 1,08403$$

$$d(x, y) = \sqrt{(0.67 - 0.48)^2 + (0.50 - 0.25)^2 + (0.50 - 0.35)^2 + (0.75 - 0.48)^2 + (0.84 - 0.59)^2 + (0.83 - 0.71)^2 + (1 - 1)^2 + (0.96 - 0.90)^2 + (0 - 0)^2 + (1 - 0)^2 + (0 - 1)^2 + (0 - 0)^2 + (0.34 - 0.65)^2}$$

$$= 1,5397$$

Setelah dilakukan perhitungan terhadap seluruh data, diperoleh sejumlah data tetangga terdekat dengan mayoritas memiliki label “Preeklampsia”, sehingga data pasien PI001 dikategorikan termasuk ke dalam kelompok “Preeklampsia”. Berikut sampel hasil prediksi algoritma KNN pada data training ditunjukkan pada Tabel 12.

Tabel 12 Perbandingan Label Aktual dan Label Prediksi Pada Tahap Modelling

No	Nama	Aktual	Prediksi
1	PI001	Preeklampsia	Preeklampsia
2	PI002	Normal	Normal
3	PI006	Normal	Normal
...	...	...	...
537	PI597	Preeklampsia	Preeklampsia
538	PI598	Preeklampsia	Preeklampsia
539	PI600	Preeklampsia	Preeklampsia

e. Evaluasi

Selanjutnya dilakukan evaluasi terhadap hasil prediksi pengujian data training untuk mengetahui sejauh mana model yang dibangun mampu melakukan klasifikasi dengan tepat. memastikan bahwa model KNN yang telah dibangun benar-benar mampu mempelajari pola dari data yang diberikan dengan baik. Evaluasi dilakukan dengan menghitung jumlah True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Setelah dilakukan perbandingan label aktual dan label prediksi diperoleh nilai TP=96, TN=442, FP=0, dan FN=1. Untuk memberikan gambaran yang lebih jelas mengenai distribusi hasil prediksi model, hasil tersebut disajikan dalam bentuk Confusion Matrix sebagaimana ditunjukkan pada Tabel 13.

Tabel 13 Confusion Matrix

	Prediksi Preeklampsia	Prediksi Normal
Aktual Preeklampsia	96	1
Aktual Normal	0	442

Berdasarkan hasil confusion matrix di atas, hasil pengujian pada data training menunjukkan performa yang sangat tinggi dengan akurasi 99,81%, precision 100,00%, recall 98,97%, dan F1-Score 99,48%. Tingginya nilai pada seluruh metrik evaluasi ini mengindikasikan bahwa model KNN dengan k=3 berhasil mempelajari pola dari data training dengan sangat baik dan konsisten dalam menghasilkan prediksi yang tepat.

2. Pembahasan

a. Hasil Uji Ahli (Blackbox Testing dan Whitebox Testing)

Prototype sistem yang telah dibangun selanjutnya diuji untuk menilai kelayakan dan kesesuaian fungsionalitas sistem. Pengujian dilakukan oleh dosen sebagai penguji ahli menggunakan metode blackbox testing dan whitebox testing. Kemudian diperoleh persentase kelayakan untuk pengujian blackbox testing sebesar 95% dan pengujian whitebox testing sebesar 100%. Hal ini menunjukkan bahwa sistem berada dalam kategori “Sangat Layak”.

b. Hasil Uji Pengguna (PSSUQ)

Selain itu, dilakukan pula uji pengguna untuk mengevaluasi tingkat kepuasan pengguna terhadap sistem dari aspek kemudahan penggunaan, kualitas informasi, dan kualitas sistem. Pengujian menggunakan kuesioner Post-Study System Usability Questionnaire (PSSUQ) yang diberikan kepada dua tenaga kesehatan. Berdasarkan hasil pengujian, diperoleh nilai sebesar 82,8% yang menunjukkan bahwa sistem “Sangat Layak” dalam mendukung proses deteksi ini preeklampsia pada ibu hamil.

c. Prediksi Data Testing (Deployment)

Karena model KNN telah menunjukkan kemampuan klasifikasi yang memadai, tahap selanjutnya melakukan prediksi terhadap 62 data testing. Prediksi dilakukan dengan menjadikan data training sebagai acuan dan memakai parameter yang sama yaitu k=3. Dengan memanfaatkan data training, diharapkan model KNN dapat memberikan prediksi yang lebih stabil dan konsisten, sehingga evaluasi pada data testing dapat mencerminkan kemampuan generalisasi model terhadap data baru. Sampel data training yang digunakan sebagai acuan ditunjukkan pada Tabel 14.

Tabel 14 Sampel Data Training yang Digunakan Sebagai Acuan

No	Nama	Usia	Paritas	Tb	Bb	Imt	Sistolik	Diastolik	Map	Protein_urine_n	Protein_urine_1	Protein_urine_2	Protein_urine_3	Gds	Status
1	PI001	0.48	0.25	0.35	0.48	0.59	0.71	1	0.90	0	0	1	0	0.65	Preeklampsia
2	PI002	0.81	0.50	0.41	0.07	0.08	0.38	0.51	0.47	1	0	0	0	0.33	Normal
3	PI006	0.69	0.5	0.59	0.11	0.11	0.36	0.54	0.49	1	0	0	0	0.15	Normal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
539	PI600	0.67	0.50	0.50	0.75	0.84	0.83	1	0.96	0	1	0	0	0.34	Preeklampsia

Selanjutnya, dilakukan proses prediksi terhadap data testing menggunakan model KNN. Berikut sampel data testing yang digunakan ditunjukkan pada Tabel 15:

Tabel 15 Sample Data Testing

Nama	Usia	Paritas	Tb	Bb	Imt	Sistolik	Diastolik	Map	Protein_urine_n	Protein_urine_1	Protein_urine_2	Protein_urine_3	Gds	Status
PI004	0	0.25	0.48	0.26	0.27	0.19	0.41	0.37	0	0	1	0	0.58	Normal

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

$$d(x, y) = \sqrt{(0 - 0.48)^2 + (0.25 - 0.25)^2 + (0.48 - 0.35)^2 + (0.26 - 0.48)^2 + (0.27 - 0.59)^2 + (0.19 - 0.71)^2 + (0.41 - 1)^2 + (0.37 - 0.90)^2 + (0 - 0)^2 + (0 - 0)^2 + (1 - 1)^2 + (0 - 0)^2 + (0.58 - 0.65)^2}$$

$$= 1.171537$$

Berdasarkan hasil prediksi terhadap 62 data testing, sebanyak 48 data tetap diklasifikasikan sebagai Normal, sedangkan 14 data diprediksi sebagai Preeklampsia. Perubahan tersebut menunjukkan bahwa model K-Nearest Neighbor (KNN) mengidentifikasi adanya pola pada beberapa data yang lebih mendekati karakteristik pasien preeklampsia berdasarkan pola yang dipelajari dari data valid.

D. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, algoritma K-Nearest Neighbor (KNN) dengan nilai k=3 mampu menghasilkan performa klasifikasi yang baik dengan akurasi 99,81%, precision 100%, recall 98,97%, dan F1-Score 99,48%. Algoritma KNN juga mampu mengolah data skrining menjadi informasi yang dapat mendukung tenaga kesehatan dalam melakukan penegakkan diagnosis pada proses deteksi dini preeklampsia. Prototype yang dikembangkan telah memenuhi aspek fungsional dengan hasil White Box Testing sebesar 100%, Black Box Testing sebesar 95%, serta memperoleh nilai PSSUQ sebesar 82,8% dengan kategori Sangat Layak.

Penelitian selanjutnya disarankan untuk menggunakan jumlah dataset yang lebih besar dan menambahkan variabel yang memengaruhi risiko preeklampsia, seperti riwayat preeklampsia pada ibu atau saudara Perempuan kandung, pembengkakan (edema), dan riwayat preeklampsia pada kehamilan sebelumnya.

E. DAFTAR PUSTAKA

- [1] Dinas Kesehatan Kota Bogor, “Pers Rilis Upaya Pencegahan Kematian Ibu Dan Bayi Di Kota Bogor,” Dinas Kesehatan Kota Bogor.
- [2] World Health Organization, “Maternal mortality,” World Health Organization.
- [3] Kementrian Kesehatan Republik Indonesia, “Webinar Save Mother Save The Nation,” Kementrian Kesehatan Republik Indonesia.
- [4] I. B. G. Manuaba, I. A. C. Manuaba, and I. B. G. F. Manuaba, Pengantar Kuliah Obstetri, 1st ed., vol. 8. Jakarta: Buku Kedokteran EGC , 2003.
- [5] A. Muis et al., Data Mining: Konsep, Metode, dan Aplikasi. Faaslib Serambi Media, 2025.
- [6] S. J. A. Sumarauw, Data Mining Model Self-Organizing Maps (SOMs). CV. Bintang Semesta Media, 2022.
- [7] I. Saputra, Belajar Mudah Data mining Untuk Pemula. Informatika Bandung, 2023.
- [8] H. S. S.J, G. Ridowati, and I. Binanto, Pengantar Pembelajaran Mesin Menggunakan Bahasa Pemrograman Python. Sanata Dharma University Press, 2023.