

## Klasifikasi Preeklampsia pada Ibu Hamil menggunakan Algoritma KNN

Tri Aji Tunggal Saputra<sup>\*1</sup>, Juhari<sup>2</sup>, Gandung Triyono<sup>3</sup>

<sup>1,2,3</sup>Faculty of Information Technology, Universitas Budi Luhur, Indonesia  
Email: <sup>1</sup>[triajitunggals@gmail.com](mailto:triajitunggals@gmail.com), <sup>2</sup>[simanjuntakjuhari@gmail.com](mailto:simanjuntakjuhari@gmail.com),  
<sup>3</sup>[gandung.triyono@budiluhur.ac.id](mailto:gandung.triyono@budiluhur.ac.id)

### Abstrak

World Health Organization (WHO) pada tahun 2020 mengatakan angka kematian ibu (AKI) didunia sebanyak 227.22 per 100.000 kelahiran hidup, penyebab kematian ibu hamil secara langsung salah satunya disebabkan oleh Preeklampsia. Preeklampsia adalah suatu penyakit yang dialami oleh ibu hamil yang ditandai dengan adanya tanda-tanda hipertensi (tekanan darah tinggi), *edema* (pembengkakan), dan *proteinuria* (kadar protein dalam urin meningkat). Pada salah satu rumah sakit swasta di Cikarang kasus *preeklampsia* setiap tahunnya mengalami peningkatan, Rata-rata persentase peningkatan jumlah pasien *Preeklampsia* dari tahun 2020 hingga 2022 sekitar 17,5% per tahun. Hal itu terjadi dikarenakan ibu hamil tidak menyadari ketika dirinya mengidap penyakit *preeklampsia*. Tanda-tanda awal *preeklampsia* hampir sama dengan penyakit *hipertensi*. Penelitian ini bertujuan untuk membuat model diagnosis penyakit *preeklampsia*. Harapannya model ini bisa sebagai Alat bantu *self diagnoses*, untuk deteksi awal penyakit *preeklampsia* agar bisa mencegah terjadinya penyakit *Eklampsia* yang dapat lebih membahayakan nyawa ibu dan janin. Dalam penelitian ini, beberapa skenario eksperimen diterapkan dengan mengubah jumlah tetangga terdekat (nilai K) untuk mengoptimalkan akurasi model nilai K yang berbeda beda, sehingga didapatkan hasil *klasifikasi* yang maksimal terdapat pada jumlah nilai K =3 dan K=5 dengan pemabagian data *traing* dan data *testing* 90:10 mendapatkan hasil akurasi 97%, sedangkan hasil *klasifikasi* terendah didapatkan pada pembagian data 50:50 untuk nilai K=9 sebesar 84%.

**Kata kunci:** *data mining, ibu hamil, klasifikasi, KNN, preeklampsia*

## Classification of Preeclampsia in Pregnant Women Using the KNN Algorithm

### Abstract

The World Health Organization (WHO) in 2020 reported that the global maternal mortality rate (MMR) was 227.22 per 100,000 live births. One of the direct causes of maternal death during pregnancy is preeclampsia. Preeclampsia is a condition experienced by pregnant women, characterized by signs of hypertension (high blood pressure), edema (swelling), and proteinuria (increased protein levels in urine). In one private hospital in Cikarang, the number of preeclampsia cases has been increasing every year. The average percentage increase in the number of preeclampsia patients from 2020 to 2022 is about 17.5% per year. This is because most pregnant women do not realize that they have preeclampsia. The early signs of preeclampsia are almost identical to those of hypertension. This study aims to develop a diagnostic model for preeclampsia. The hope is that this model can serve as a self-diagnosis tool for early detection of preeclampsia, to prevent the occurrence of impending eclampsia, which can be more dangerous for both the mother and the baby. In this study, several experimental scenarios were applied by changing the number of nearest neighbors (K value) to optimize the model's accuracy with different K values. The optimal classification results were obtained with K=3 and K=5, using a training and testing data split of 90:10, which achieved an accuracy of 97%. Meanwhile, the lowest classification results were obtained with a 50:50 data split for K=9, yielding an accuracy of 84%.

**Keywords:** *classification, data mining, KNN, preeclampsia, pregnant women.*

## 1. PENDAHULUAN

Menurut World Health Organization (WHO) pada tahun 2020, angka kematian ibu (AKI) global mencapai 227,22 per 100.000 kelahiran hidup. Sebagian besar kematian ibu hamil disebabkan oleh komplikasi yang terjadi selama atau setelah kehamilan dan persalinan. Sebagian besar dari komplikasi ini berkembang selama masa kehamilan. Komplikasi utama yang berkontribusi pada 80% kematian ibu adalah pendarahan, infeksi, dan tekanan darah tinggi (*preeklampsia* dan *eklampsia*[1]). Secara global, sekitar 80% penyebab langsung kematian ibu

hamil disebabkan oleh preeklamsia. Setiap tahun, sekitar 76.000 ibu hamil meninggal akibat preeklamsia. Di negara-negara berkembang, prevalensi preeklamsia dan eklamsia berkisar antara 4% hingga 18%. Sementara itu, di Indonesia, frekuensi kasus preeklamsia berada di angka sekitar 3% hingga 10%. [2] *Preeklamsia* adalah kondisi yang dialami oleh ibu hamil dan ditandai oleh gejala hipertensi (tekanan darah tinggi), *edema* (pembengkakan), serta *proteinuria* (peningkatan kadar protein dalam urin).

Banyak ibu hamil yang tidak menyadari bahwa dirinya mempunyai penyakit *Preeklamsia*, karena gejala *preeklamsia* hampir sama dengan penyakit *hipertensi*. Tidak ada tes sederhana yang tersedia untuk mendiagnosis *preeklamsia*. Diagnosis *Preeklamsia* hanya dapat dilakukan melalui kunjungan berulang kali selama kehamilan, pengukuran tekanan darah berulang, dan analisis urin, yang memerlukan biaya besar dan sangat *sensitive* [3]

*Pre-eklamsia* adalah gangguan pada ibu hamil dengan etiologi yang belum diketahui secara pasti. Penyakit ini ditandai oleh *hipertensi* dan *proteinuria* yang muncul setelah usia kehamilan mencapai 20 minggu. [4] Karena belum diketahui etiologinya pencegahan *Preeklamsia* masih sangat sulit dilakukan. Deteksi awal preeklamsia sangat penting dan dianjurkan selama trimester pertama kehamilan agar tindakan pencegahan dapat dilakukan lebih awal [5]. Jika dibiarkan penyakit *preeklamsia* dapat menjadi penyakit *eklamsia*. *Eklamsia* ialah suatu keadaan yang dimana terjadi serangan kejang tiba-tiba dan diikuti dengan penurunan kesadaran atau koma pada seorang wanita hamil [6] Oleh karena itu, diperlukan suatu metode yang dapat membantu para tenaga kesehatan untuk dapat memberikan deteksi awal penyakit *preeklamsia* yang diderita oleh ibu hamil. Salah satu metode yang dapat digunakan adalah metode data mining [7]

Metode diagnosis berbasis komputer sering kali menggunakan teknik Data Mining dengan memanfaatkan database riwayat medis pasien. *Algoritma* cerdas dalam Data Mining memungkinkan pengolahan dataset medis yang melibatkan berbagai input data untuk menyelesaikan masalah diagnosis. Data Mining dalam konteks medis memiliki potensi besar dalam mengungkap pola tersembunyi di dalam dataset medis. Pola-pola ini dapat dimanfaatkan oleh dokter untuk secara signifikan meningkatkan kualitas keputusan medis dalam menentukan ada atau tidaknya suatu penyakit. [5]

Pada penelitian sebelumnya yang dilakukan oleh Hidayat (2020) dengan melakukan klasifikasi diagnosis *Preeklamsia* menggunakan algoritma KNN dengan 120 data dan 9 atribut mendapatkan hasil *accuracy* 98,33%. [8] Penelitian lainnya dilakukan oleh Noormawati (2020) melakukan klasifikasi diagnosis dini penyakit *Preeklamsia* menggunakan algoritma KNN, dengan 100 data dan 11 atribut mendapatkan hasil *accuracy* 88%. [5]. Menurut Badriyah (2020) *preeklamsia* merupakan salah satu jenis penyakit yang berdampak buruk bagi ibu hamil dan janin, Badriyah menggunakan 6 metode yaitu *regressi logistic*, *k-nearest neighbour*, pohon Keputusan C5.0, *discriminant analysis*, *random forest*, dan SVM, dengan menggunakan 726 data dan 9 atribut, menghasilkan hasil akurasi *logistic regression* (71%), *k-nearest neighborhood* (74%), *C5.0 decision tree* (78%), *discriminant analysis* (68%), *random forest* (75%) and *support vector machine* (79%) [9]. Penelitian yang dilakukan oleh Sukmanto (2020) yaitu mengoptimalkan algoritma KNN dengan *Grid Search Algorithm* dengan menggunakan 15 atribut, berhasil meningkatkan akurasi 2,40%, dari 61,96% ke 63,45%. [10]. Penelitian yang dilakukan oleh Ilham (2024) Penelitian ini dilakukan dengan tujuan untuk mengembangkan model berdasarkan metode data mining yang dapat digunakan sebagai alat pendeteksi untuk mengidentifikasi pasien dengan *preeklamsia* dengan cara membandingkan beberapa algoritma. Algoritma yang dibandingkan dalam penelitian ini yaitu *decision tree*, *support vector machine*, *naïve bayes*, *logistic regression*, *k nearest neighbors*, dan *random forest*. Penelitian ini menggunakan 19 atribut dengan 109 data yang terdiri dari 48 data pasien *positif preeklamsia* dan 61 data pasien *negatif preeklamsia*. Hasil dari penelitian ini adalah akurasi *Decision Tree* 100%, Akurasi *Support Vector Machine* 100%, akurasi *Naïve Bayes* 73%, akurasi *Logistic Regression* 97%, Akurasi *K-Nearest Neighbors* 94%, akurasi *Random Forest* 97% [7]. Penelitian yang berjudul *Early prediction of preeclampsia via machine learning*, Penelitian ini menggunakan metode *machine learning statistical learning methods* Tujuannya untuk mengembangkan model prediksi *preeklamsia*. *Variabel* yang digunakan dalam model tersebut meliputi 16. Hasil dari penelitian ini mendapatkan akurasi 72,3%. [11]. Penelitian terdahulu yang dilakukan oleh Nurrohmah (2022) penelitian ini membahas bagaimana membuat sebuah sistem dengan menggunakan algoritma C4.5 untuk memprediksi penyakit *Preeklamsia*, Penelitian ini menggunakan 247 data. Setelah itu peneliti melakukan pemecahan data berdasarkan usia kehamilan, hingga data yang dipakai menjadi 870 dengan jumlah atribut sebanyak 12. Hasil dari penelitian ini menunjukkan bahwa menunjukkan bahwa sistem prediksi mendapatkan nilai akurasi 81,38%, *precision* 78,37%, *recall* 79,69%, dan *f1-score* 78,73%. [12]

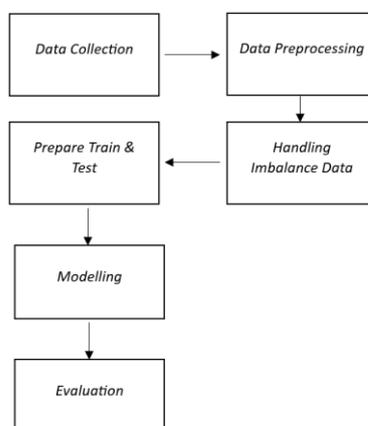
Banyak peneliti yang sudah melakukan klasifikasi pada penyakit *preeklamsia* dengan metode KNN. Tetapi jika dibandingkan dengan metode lain hasilnya masih belum ada yang mencapai akurasi maksimal. Metode algoritma KNN digunakan untuk mengklasifikasikan data dengan cara mencari jarak terdekat antara objek data. Pemilihan nilai K yang optimal dalam algoritma ini ditentukan berdasarkan data yang tersedia. Nilai K yang lebih tinggi dapat membantu mengurangi pengaruh *noise* pada *klasifikasi* serta membuat perbedaan antara setiap kelas menjadi lebih jelas. [7] Algoritma KNN umumnya digunakan untuk klasifikasi data dengan jumlah yang kecil, namun apabila dataset yang diproses cukup kompleks, algoritma ini dapat menunjukkan kelemahan

dan menjadi kurang efisien dalam hal waktu komputasi. Dalam metode KNN, data baru diklasifikasikan berdasarkan ukuran jarak yang tepat. Jumlah K tetangga terdekat dihitung, dan label kelas tetangga terdekat digunakan untuk memprediksi label kelas instance baru. Pemilihan jumlah K tetangga terdekat memiliki pengaruh besar terhadap akurasi KNN. Nilai K yang kecil dapat membuat model sensitif terhadap noise, sementara nilai K yang terlalu besar dapat menyebabkan bias pada model.[8] Sehingga Penelitian ini berfokus pada optimasi nilai K untuk meningkatkan akurasi pada *klasifikasi*, yang belum dibahas secara mendalam dalam studi terdahulu.

Berdasarkan latar belakang tersebut penelitian ini bertujuan untuk melakukan pengujian metode KNN untuk klasifikasi penyakit *preeklampsia* pada ibu hamil.

## 2. METODE PENELITIAN

Metode penelitian ini meliputi tahapan terstruktur, dimulai dari pengumpulan data, pre pemrosesan data, penyusunan dataset, pembagian data latih dan data uji, *modelling* dan *evaluation*. Dengan pendekatan ini, memastikan bahwa setiap langkah dalam proses klasifikasi dikembangkan secara sistematis dan terkoordinir untuk mencapai hasil yang akurat dan *relevan*. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

### 2.1. Data Collection

Data Collection merupakan tahapan yang terdiri dari identifikasi kebutuhan fitur data serta pengumpulan data pasien melalui data rekam medis. Data yang dibutuhkan untuk kebutuhan penelitian diidentifikasi berdasarkan *studi literatur* yang telah dilakukan dan hasil wawancara dengan bidan spesialis kehamilan. Identifikasi fitur kebutuhan data dilakukan untuk melakukan seleksi terhadap seluruh data yang didapatkan dari pihak rumah sakit menjadi beberapa bagian data yang relevan dengan kebutuhan penelitian. seleksi jumlah atribut dilakukan untuk menjaga identitas pasien seperti pengurangan atribut nama dan juga alamat dari pasien. Pengumpulan data pasien dilakukan setelah identifikasi fitur kebutuhan data. Data dikumpulkan dari data pasien yang memiliki penyakit yang berkaitan erat dengan *preeklampsia* dalam jangka waktu tertentu dari bagian rekam medis.

### 2.2. Data Preprocessing

Pada tahap ini dilakukan *pra pemrosesan* data. *Pra Pemrosesan* data dilakukan setelah dilakukan pengumpulan data pasien. Tahapan ini dilakukan agar data yang telah dikumpulkan siap untuk diolah dan memiliki format yang sesuai dengan kebutuhan pembangunan model deteksi. *Pra pemrosesan* data yang dilakukan adalah pembentukan atribut baru dan *transformasi* data text menjadi *numeric*.

### 2.3. Handling Imbalanced Data

*Imbalanced Data* dapat diartikan sebagai adanya rasio yang tidak profesional di setiap kelas, *Imbalanced Data* dapat menyebabkan hasil dari klasifikasi tidak akurat, maka dari itu diperlukan pra-pemrosesan data untuk dapat menyeimbangkan jumlah class pada data agar hasil klasifikasi lebih akurat. Salah satu cara untuk bisa menyeimbangkan jumlah data yaitu dengan menggunakan metode SMOTE. SMOTE adalah metode yang

menerapkan *oversampling* dalam mengatasi data yang kurang seimbang dengan mempertimbangkan efek pada wilayah keputusan dalam ruang lingkup.[9].

#### 2.4. Pembagian data latih dan data uji

Data penelitian akan dibagi menjadi dua bagian, yaitu data latih dan data Uji. Data latih akan digunakan untuk membuat model klasifikasi sebagai model prediksi, dan data uji akan digunakan untuk menguji model klasifikasi yang telah terbentuk.

#### 2.5. Modelling

Modelling melibatkan penggunaan model statistik atau algoritma pembelajaran mesin tertentu untuk menganalisa pola yang terdapat dalam data dan membuat prediksi atau klasifikasi. Pada bagian modelling akan digunakan metode KNN untuk proses klasifikasi. Algoritma KNN memiliki beberapa kelebihan diantaranya ketangguhan terhadap data *training* yang memiliki banyak *noise* dan data dalam jumlah yang besar. Kelemahan dari algoritma KNN adalah perlu menentukan jumlah tetangga terdekat dari target data, yang disimbolkan dengan nilai parameter K. Maka dari itu pada penelitian ini akan dicoba beberapa nilai K dengan nilai ganjil yaitu 3, 5, 7, 9. Nilai ganjil digunakan guna menghindari kemunculan hasil jumlah jarak yang sama atau seri pada proses klasifikasi[15].

#### 2.6. Evaluation

Pengujian yang dilakukan dengan menggunakan *confussion matrix*. *Confusion matrix* adalah alat untuk analisis *prediktif* dalam *machine learning*. Digunakan untuk memeriksa kinerja *model machine learning* berbasis klasifikasi. *Confusion matrix* juga bisa disebut tabel ringkasan jumlah prediksi *true* atau *false* yang dihasilkan oleh model klasifikasi. Dengan memvisualisasikan matriks konfusi, dapat menentukan keakuratan model dengan mengamati nilai diagonal untuk mengukur jumlah klasifikasi yang akurat[16]

Akurasi adalah ukuran seberapa banyak prediksi yang benar pada model.Persamaan (1) akan menjelaskan tentang rumus akurasi.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Presisi merupakan kesesuaian nilai yang didapatkan berdasarkan informasi yang kurang. Pada strata biner, presisi sama dengan nilai prediksi positif pada suatu penelitian. Persamaan (2) akan menjelaskan tentang rumus presisi.

$$Precision = \frac{TP}{TP+FN} \tag{2}$$

*Recall* adalah penghapusan data yang diambil dari data yang sesuai dengan kueri. Dalam klasifikasi biner, *recall* dikenal sebagai sensitivitas. Munculnya data relevan yang diambil adalah menyetujui dengan *query* dapat dilihat dengan *recall*. Persamaan (3) akan menjelaskan tentang rumus presisi.

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

Keterangan :

TF = adalah *true positif*, yang terprediksi secara benar pada jumlah data *positif* di sistem

TN = adalah *true negatif*, yang terprediksi benar pada jumlah data *negatif* di sistem

FN = adalah *false negatif*, yang terprediksi salah pada jumlah data *negatif* di sistem

FP = adalah *false positif*, yang terprediksi salah pada jumlah data *positif* di sistem.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Dataset

Data penelitian ini menggunakan dataset dari data Rekam Medis Rumah Sakit yang ada di daerah Cikarang, *sample* yang digunakan berjumlah 408 data. Seleksi jumlah atribut dilakukan untuk menjaga identitas pasien seperti pengurangan atribut nama dan juga alamat dari pasien, sehingga fitur yang diambil adalah Usia Ibu, Usia Kandungan, *Paritas*, *Sistol*, *Diastol*, *Protein Urine*, *Trombosite*, *Glukosa*, Riwayat Hipertensi, Riwayat Abortus, dan Keluhan. Table dataset dapat dilihat dari tabel 1.

Tabel 1. Atribut dataset

Atribut	Tipe
Usia	Numeric
Usia Kehamilan	Numeric
Paritas	Text
Sistol	Numeric
Diastol	Numeric
Protein Urine	Numeric
Trombosit	Numeric
Glukosa	Numeric
Riwayat Abortus	Numeric
Riwayat Hipertensi	Boolean
Keluhan	Categorical
Diagnosa	Label

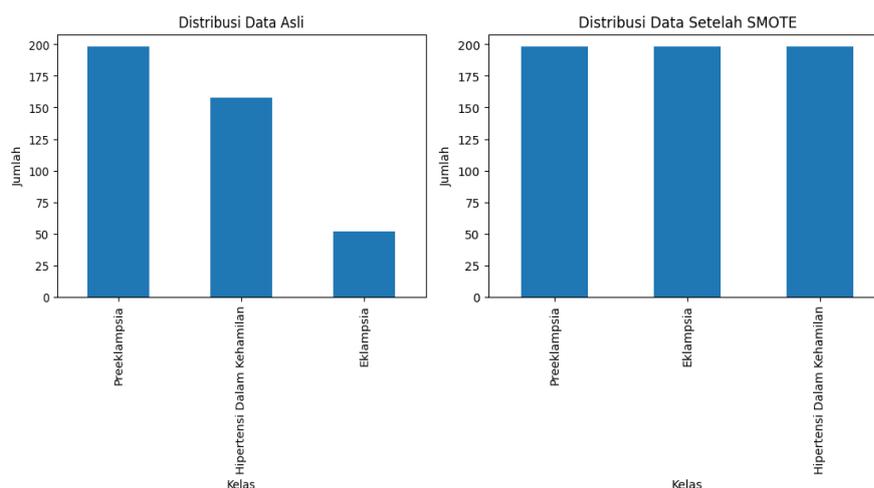
### 3.2. Preprocessing

Data preprocessing merupakan tahapan untuk mempersiapkan data agar bisa diolah dan dipakai dalam algoritma KNN. Pra pemrosesan data yang dilakukan adalah pembentukan atribut baru dan juga perubahan isi dari atribut sesuai dengan algoritma yang akan digunakan.

Pra pemrosesan data yang dilakukan adalah pembentukan atribut baru yaitu Pusing, Oedema, Pandangan Buram/Double, Nyeri Epigastric yang diambil dari data keluhan, diketahui bahwa dataset Keluhan berisi categorical sehingga pada tahap ini di pisah menjadi data numeric. 0 bahwa data tersebut tidak mengalami keluhan yang dimaksud, sedangkan 1 bahwa data tersebut mengalami keluhan yang dimaksud, perubahan data paritas yang awalnya merupakan text diubah menjadi Primipara menjadi 1, Multipara menjadi 2, dan Grandemultipara menjadi 5, Kolom Riwayat hipertensi pun diubah menjadi ya jadi 1 dan tidak menjadi 0 dan protein urine diubah menjadi bentuk bilangan bulat. Perubahan dan Pembentukan atribut baru digunakan untuk mempermudah proses pemodelan. Proses data preprocessing ini dilakukan di *microsoft excel*.

### 3.3. Handling Imbalanced Data

Imbalanced Data dapat diartikan sebagai adanya rasio yang tidak profesional di setiap kelas, Imbalanced Data dapat menyebabkan hasil dari klasifikasi tidak akurat. Proses over-sampling pada SMOTE melibatkan pemilihan instance dari kelas minoritas, kemudian mencari *k-nearest neighbors* untuk setiap instance tersebut, dan menghasilkan data sintetik. Metode ini menggantikan sekadar menggandakan data kelas minoritas, sehingga dapat mengurangi risiko *overfitting* yang berlebihan.[9].



Gambar 2. Perbedaan Distribusi Data

Jumlah data pada yang terdiagnosa *Preeklampsia*, *Eklampsia* dan hipertensi pada kehamilan pada data rekam medis kurang seimbang, jumlah yang terdiagnosa *Preeklampsia* 198 pasien, jumlah yang terdiagnosa *Eklampsia* 52 pasien dan yang terdiagnosa hipertensi pada kehamilan 152 pasien, sehingga menyebabkan ketidak

seimbangan jumlah data atau bisa disebut *Imbalanced Data*. Gambar 1 dan 2 merupakan hasil perbandingan sebelum dan sesudah menggunakan SMOTE.

Setelah dilakukan optimasi pada *imbalance data* dengan menggunakan SMOTE, data yang awalnya berjumlah 408 menjadi 594 dengan masing masing *diagnose* berjumlah 198 data. Perbedaan jumlah data dapat dilihat dari gambar 2.

### 3.4. Pembagian Data Latih dan Data Uji

Data penelitian akan dibagi menjadi dua bagian, yaitu data latih dan data Uji. Data latih akan digunakan untuk membuat model klasifikasi sebagai model prediksi, dan data uji akan digunakan untuk menguji model klasifikasi yang telah terbentuk. Adapun pembagian data latih dan data uji akan bagi menjadi 5 bagian, untuk percobaan pertama akan di coba perbandingan data latih dan data uji 80:20, untuk percobaan kedua perbandingan Jumlah data latih dan data uji 50:50, untuk percobaan ketiga perbandingan Jumlah data latih dan data uji 90:10, untuk percobaan keempat perbandingan Jumlah data latih dan data uji 70:30 dan yang kelima adalah Jumlah data latih dan data uji 60:40. Adapun table perbandingan data latih dan data uji setiap percobaan dapat dilihat dari Tabel 2

Tabel 2. Pembagian data latih dan data uji

Data Latih	Jumlah Data	Data Uji	Jumlah Data
80%	475	20%	119
50%	297	50%	297
90%	534	10%	60
70%	415	30%	179
60%	356	40%	238

### 3.5. Tahap Pengujian

Algoritma KNN (*K-Nearest Neighbors*) termasuk dalam kategori metode *supervised learning*, di mana klasifikasi *query instance* dilakukan berdasarkan mayoritas kategori dari tetangga terdekatnya. Kelebihan algoritma KNN antara lain adalah kemampuannya untuk tetap efektif meskipun data training mengandung banyak noise dan volume data yang besar. Namun, kelemahan dari KNN terletak pada kebutuhan untuk menentukan nilai parameter K (jumlah tetangga terdekat), serta tantangan dalam memilih jenis jarak yang tepat dan atribut yang relevan, yang dapat mempengaruhi akurasi perhitungan jarak dan hasil klasifikasi.[10] Perhitungan jarak yang biasa digunakan adalah perhitungan dengan metode *Euclidian distance*. Rumus perhitungan dari *Euclidian distance* dapat dilihat dari persamaan 4.

$$d(x_i, y_i) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \tag{4}$$

Keterangan :

D(x<sub>i</sub>,y<sub>i</sub>) = Jarak antar data latih

X<sub>i</sub> = Sampel data

y<sub>i</sub> = Data yang diuji

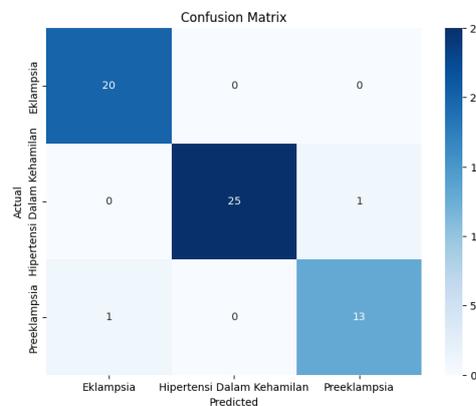
Setelah dilakukan tahap *preprocessing* dan *Handling Imbalance Data*, selanjutnya adalah tahap pengujian untuk algoritma KNN, tahap pengujian ini akan di bagi menjadi 5 kali percobaan sesuai dengan pembagian dataset yang terdapat pada table 2, dengan nilai K adalah 3,5,7 dan 9. Setelah dilakukan percobaan, dapat dilihat pada table 2 merupakan hasil dari akurasi tiap kombinasi.

Table 2. Akurasi KNN

	50:50	60:40	70:30	80:20	90:10
<b>K=3</b>	0.9057	0.9285	0.9441	0.9495	0.9666
<b>K=5</b>	0.8989	0.9159	0.9385	0.9327	0.9666
<b>K=7</b>	0.8619	0.8991	0.9106	0.9243	0.9500
<b>K=9</b>	0.8417	0.8865	0.8882	0.9159	0.9166

Berdasarkan hasil pengujian akurasi data terbaik didapatkan pada pembagian data 90:10 untuk nilai K=3 dan K=5 yaitu sebesar 97%, dan untuk nilai akurasi terendah didapatkan pada pembagian data 50:50 untuk nilai K=9 sebesar 84%. Nilai K = 3 dan K = 5 memberikan akurasi terbaik dalam model *K-Nearest Neighbors*

(KNN) karena nilai K yang lebih kecil mampu menangkap pola lokal pada dataset secara lebih akurat tanpa terlalu dipengaruhi oleh *noise*. Pada nilai K = 3 dan K = 5, model mempertimbangkan jumlah tetangga terdekat yang cukup untuk menghasilkan generalisasi yang optimal, menjaga keseimbangan antara bias dan variansi, serta meminimalkan pengaruh outlier. Selain itu, pembagian dataset 90:10 memungkinkan model untuk memiliki jumlah data latih yang cukup besar, sehingga pola-pola klasifikasi lebih jelas dapat dikenali. Sebaliknya, nilai K yang lebih besar, seperti K = 9, cenderung mempertimbangkan lebih banyak tetangga, yang dapat mengaburkan pola lokal dan mengurangi sensitivitas model terhadap karakteristik spesifik dari setiap kelas. Hasil ini menunjukkan bahwa kombinasi pembagian data 90:10 dengan nilai K = 3 atau K = 5 memberikan performa terbaik untuk klasifikasi *preeklampsia* pada ibu hamil, dengan akurasi mencapai 97%. Hasil dari penelitian ini cukup bisa dapat meningkatkan performa dari algoritma KNN dibandingkan dengan penelitian terdahulu yang masih memiliki nilai akurasi dibawah 97%. Hal ini membuktikan bahwa pembagian data dan nilai K sangat berpengaruh pada hasil akurasi klasifikasi penyakit *preeklampsia* pada ibu hamil. Semakin banyak data latih dan semakin sedikit nilai K akan mendapatkan hasil yang maksimal sedangkan semakin sedikit data latih dan semakin banyak nilai K akan mendapatkan hasil yang kurang maksimal. Untuk dapat memaksimalkan hasil pembuktian dari akurasi terbaik gambar 4 menunjukan confusion matrix dari pembagian data 90:10 dan K=5.



Gambar 4. Confusion matrix KNN 90:10

Tabel 3. Classification Report KNN 90:10

	Precision	Recall	f1-score	support
<b>Eklampsia</b>	0.95	1.00	0.98	20
<b>Hipertensi</b>	1.00	0.96	0.98	26
<b>Preeklampsia</b>	0.93	0.93	0.93	14
<b>Accuracy</b>			0.97	60
<b>macro avg</b>	0.96	0.96	0.96	60
<b>weighted avg</b>	0.97	0.97	0.97	60

#### 4. KESIMPULAN

Pada penelitian ini data yang digunakan berjumlah 408 yang diambil dari data rekam medis rumah sakit yang ada di daerah Cikarang. Proses *klasifikasi* dilakukan terhadap pembagian data latih dan data uji serta nilai K yang berbeda beda, sehingga didapatkan hasil klasifikasi yang maksimal terdapat pada jumlah nilai K=3 dan K=5 dengan pembagian data latih dan data uji 90:10 mendapatkan hasil akurasi 97%, sedangkan hasil klasifikasi terendah didapatkan pada pembagian data 50:50 untuk nilai K=9 sebesar 84%. Semakin banyak data latih dan semakin sedikit nilai K akan mendapatkan hasil yang maksimal sedangkan semakin sedikit data latih dan semakin banyak nilai K akan mendapatkan hasil yang kurang maksimal. Penelitian ini membuktikan bahwa optimasi pembagian data dan nilai K sangat berpengaruh secara signifikan pada peningkatan hasil akurasi *klasifikasi* penyakit *preeklampsia* pada ibu hamil. Dengan ini performa KNN dengan nilai akurasi tersebut diharapkan dapat menjadi acuan untuk menjadi alat *self diagnose* tahap awal pada penyakit *preeklampsia* pada ibu hamil. Saran yang dapat diberikan untuk topik *klasifikasi* penyakit *preeklampsia* adalah dengan mencoba dataset yang lebih besar dan menggunakan beberapa penyakit pada ibu hamil lainnya selain penyakit *eklampsia* dan *hipertensi* sebagai penyakit pembanding agar hasil yang didapatkan lebih akurat.

## DAFTAR PUSTAKA

- [1] D. Saputri, M. Sari, P. Fransiska, A. K. Rangga, dan H. Prabumulih, "Karakteristik Ibu Hamil Dengan Preeklampsia," *Cendekia Medika : Jurnal STIKES Al-Ma'arif Baturaja*, vol. 8, no. 1, 2023.
- [2] R. T. P. S. R. Rismayanti Mambela, "Identifikasi Kejadian Preeklampsia Pada Ibu Hamil Di Kabupaten Luwu, Sulawesi Selatan," 2020.
- [3] Z. Manoochchri, S. Manoochchri, F. Soltani, L. Tapak, dan M. Sadeghifar, "Predicting preeclampsia and related risk factors using data mining approaches: A cross-sectional study," *Int J Reprod Biomed*, vol. 19, no. 11, hlm. 959–968, Okt 2021, doi: 10.18502/ijrm.v19i11.9911.
- [4] N. Muhani dan B. Besral, "Pre-eklampsia Berat dan Kematian Ibu," *Kesmas: National Public Health Journal*, vol. 10, no. 2, hlm. 80, Nov 2015, doi: 10.21109/kesmas.v10i2.884.
- [5] S. Amelia, M. Nor Hayati, S. Prangga, P. Studi Statistika, dan J. Matematika, "Penerapan Metode Modified K-Nearest Neighbor Pada Pengklasifikasian Status Pembayaran Kredit Barang Elektronik dan Furniture," 2022.
- [6] N. Fadhilah dan R. Widyastuti, "Eklampsia," *Continuing Medical Education*, 2022.
- [7] M. Ilham, N. Luh, S. S. Adnyani, dan D. K. Suryadi, "Pembangunan Model Pendeteksi Risiko Preeklampsia pada Ibu Hamil dengan Menggunakan Metode Data Mining," vol. 23, no. 01, hlm. 50–60, 2024, [Daring]. Tersedia pada: <http://creativecommons.org/licenses/by/4.0/>
- [8] R. Hidayat dan T. Astuti, "Diagnosis Preeklampsia pada Ibu Hamil Berdasarkan Algoritme K-Nearest Neighbour," *Jurnal Ilmu-ilmu Informatika dan Manajemen STMIK*, vol. 14, no. 2, 2020.
- [9] M. Tahir, T. Badriyah, dan I. Syarif, *Neural Networks Algorithm to Inquire Previous Preeclampsia Factors in Women with Chronic Hypertension During Pregnancy in Childbirth Process*, vol. 19, Issue no. 11. IEEE, 2020.
- [10] Sukamto, Hadiyanto, dan Kurnianingsih, "KNN Optimization Using Grid Search Algorithm for Preeclampsia Imbalance Class," dalam *E3S Web of Conferences*, EDP Sciences, Nov 2023. doi: 10.1051/e3sconf/202344802057.
- [11] I. Marić *dkk.*, "Early prediction of preeclampsia via machine learning," *Am J Obstet Gynecol MFM*, vol. 2, no. 2, Mei 2020, doi: 10.1016/j.ajogmf.2020.100100.
- [12] S. Nurrohmah, D. Normawati, A. Dahlan, J. Ringroad Selatan, dan I. Yogyakarta, "Prediksi Dini Penyakit Preeklampsia Menggunakan Algoritma C4.5," vol. 10, no. 3, hlm. 120–132, 2022, doi: 10.12928/jstie.v8i3.xxx.
- [13] L. Anshori dan R. Regasari Mardi Putri, "Implementasi Metode K-Nearest Neighbor untuk Rekomendasi Keminatan Studi (Studi Kasus: Jurusan Teknik Informatika Universitas Brawijaya)," 2018. [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>
- [14] W. Puspita Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada 'Lombok Vape On,'" *Jurnal Informatika dan Teknologi*, vol. 3, no. 2, 2020.
- [15] Salsabila, S. Martha, dan W. Andani, "Komparasi Algoritma K-Nearest Neighbor Dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Stunting Balita," vol. Volume 13, No. 2, 2024.
- [16] Z. Karimi, "Confusion Matrix," 2021. [Daring]. Tersedia pada: <https://www.researchgate.net/publication/355096788>
- [17] Mustakim dan G. Oktaviani, "Algoritma K-Nearest Neighbor Classification," *Jurnal Sains, Teknologi dan Industri*, vol. 13, no. 2, hlm. 195–202, 2016, [Daring]. Tersedia pada: <http://ejournal.uin-suska.ac.id/index.php/sitekin>.