

# PERBANDINGAN ALGORITME C4.5 DAN NAÏVE BAYES DALAM KLASIFIKASI SEVERE PREEKLAMPSIA MENGGUNAKAN HEMATOLOGI

Samuel Beryl Enrico Ritonga, Ultach Enri

Informatika, Universitas Singaperbangsa Karawang

Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361

2010631170119@student.unsika.ac.id

## ABSTRAK

Penelitian ini memfokuskan perbandingan metode klasifikasi Naïve Bayes dan C4.5 dalam mengidentifikasi kasus preeklampsia berdasarkan data hematologi pada dataset Rumah Sakit Bayukarta Karawang 2021-2023. Decision Tree (C4.5) menghasilkan akurasi 75%, sedangkan Naive Bayes 71%, dengan F1-score kelas 1 masing-masing 0.83 dan 0.80. Model C4.5 di-deploy sebagai web sederhana menggunakan Flask dan TailwindCSS untuk prediksi data baru. Metodologi yang dipakai penelitian ini yaitu KDD (Knowledge Discovery in Database) yang memungkinkan membantu penambangan data yang terstruktur dan terarah dari awal hingga akhir, Saran pengembangan mencakup analisis fitur preeklampsia, peningkatan akurasi, eksplorasi metode pengolahan data, dan penggunaan dataset lebih besar. Melibatkan faktor risiko tambahan dan penerapan ensemble learning juga disarankan. Diharapkan kontribusi penelitian ini dapat meningkatkan pemahaman dan pencegahan preeklampsia serta meningkatkan ketepatan model pada tahap awal kehamilan.

**Kata kunci :** Preeklampsia, Machine Learning, Naive Bayes, C4.5, Flask.

## 1. PENDAHULUAN

Preeklampsia merupakan suatu masalah kesehatan serius dan potensial mengancam jiwa yang terjadi selama masa kehamilan. Gejala yang muncul melibatkan peningkatan tekanan darah dan kenaikan kadar protein dalam urine (proteinuria) setelah mencapai usia kehamilan 20 minggu atau trimester kedua, yang bisa berujung pada timbulnya eklampsia [1]. Preeklampsia juga dapat diartikan sebagai suatu sindrom khusus yang terjadi selama kehamilan, yang ditandai oleh kenaikan tekanan darah (tekanan darah  $\geq 140/90$  mmHg) dan kehadiran protein dalam urine (protein  $\geq 0.3$  g dalam urin). Gejala ini muncul setelah usia kehamilan mencapai 20 minggu pada wanita yang sebelumnya memiliki tekanan darah dalam rentang normal [2]. Preeklampsia yang berkembang menjadi eklampsia bisa mengakibatkan masalah pada perkembangan janin, termasuk pertumbuhan janin yang terhambat dan kelahiran prematur.

Diperlukan perhatian dan pengawasan khusus dalam upaya meningkatkan kesejahteraan ibu hamil. Menurut data statistik yang dikeluarkan oleh Kementerian Kesehatan pada tahun 2021, angka kematian ibu masih tinggi di Indonesia. Terjadi penurunan dari 4.999 kasus pada tahun 2015 menjadi 4.912 kasus pada tahun 2016. Selanjutnya, terjadi penurunan lebih lanjut menjadi 1.712 kasus pada tahun 2017. Meski demikian, angka kematian ibu mengalami lonjakan signifikan menjadi 4.197 kasus pada tahun 2019, dan mengalami peningkatan sebesar 10,25% menjadi 4.627 kasus pada tahun 2020 [3].



Gambar 1. Penyebab Kematian Ibu

Sebagian besar situasi kematian ibu, yakni sekitar 75%, disebabkan oleh komplikasi utama. Pendarahan memainkan peran besar dengan 27% dari total kasus, diikuti oleh preeklampsia dan eklampsia yang mencapai 14%, infeksi 11%, persalinan obstruktif 9%, dan komplikasi terkait aborsi sebanyak 8% [4]. Pada tahun 2019, di Indonesia, penyebab kematian ibu terbagi menjadi pendarahan sebanyak 28,29%, hipertensi 23%, dan gangguan peredaran darah 4,94% [5].

Di Indonesia, insidensi preeklampsia mencapai 128.273 kasus setiap tahun, yang setara dengan sekitar 5,3% dari seluruh ibu hamil. Sementara itu, secara global, preeklampsia terjadi pada 2-8% dari total kehamilan. Sayangnya, kondisi ini juga menjadi penyebab utama kematian ibu. Di Indonesia, hipertensi selama kehamilan, termasuk preeklampsia, menjadi penyebab utama kematian ibu dengan rasio mencapai 33% [6]. Pada tingkat global, preeklampsia menyebabkan lebih dari 70.000 kematian ibu dan 500.000 kematian janin setiap tahun. Penyebab utama terjadinya preeklampsia masih belum diketahui, sehingga pencegahannya tetap menjadi tugas yang sangat sulit [7]. Oleh karena itu, perlu adanya upaya pencegahan pada ibu hamil dengan fokus khusus pada pengelolaan dan analisis peningkatan kasus hipertensi yang berkontribusi pada tingginya angka kematian ibu di Indonesia.

Meskipun telah banyak dilakukan penelitian di bidang kedokteran sebelumnya untuk memahami dan mencegah preeklampsia, masalah ini masih menjadi fokus utama dalam bidang obstetri dan kandungan. Salah satu tantangan utama adalah mengklasifikasikan tingkat keparahan preeklampsia dengan akurat pada tahap selanjutnya dalam kehamilan, yang seringkali memerlukan evaluasi yang teliti dan kompleks. Oleh karena itu, ada kebutuhan untuk mengembangkan metode yang lebih tepat dan efisien dalam mengklasifikasikan tingkat keparahan preeklampsia, agar dapat memungkinkan diagnosis dini dan penanganan yang sesuai.

Parameter tingkat protein dan tekanan darah sudah pastinya merupakan hal penting dalam menentukan tingkat keparahan preeklampsia. Penelitian yang dilakukan Candra dkk., yang membahas perbandingan profil hematologi pada preeklampsia/eklampsia dengan kehamilan normotensi, Candra menemukan perbedaan pada jumlah eritrosit dan hemoglobin antara preeklampsia berat dengan preeklampsia ringan, serta antara preeklampsia berat dengan kehamilan normotensi. Peningkatan kadar hematokrit juga terlihat pada preeklampsia berat jika dibandingkan dengan preeklampsia ringan dan preeklampsia berat dibandingkan dengan kelompok kehamilan normotensi. Jumlah leukosit menunjukkan peningkatan pada penderita preeklampsia berat jika dibandingkan dengan kelompok kehamilan normotensi [8]. Oleh karena itu, menggunakan hematologi dapat memberikan petunjuk yang berharga dalam evaluasi tingkat parahnya preeklampsia.

Penggunaan metode klasifikasi dalam penelitian ini menjadi esensial karena penderita preeklampsia dapat tergolong dalam dua kategori, yakni *severe preeclampsia* (preeklampsia parah) dan *moderate preeclampsia* (preeklampsia sedang). Proses klasifikasi memfasilitasi penyusunan kedua kategori ini, sehingga memudahkan dalam diagnosis dan penanganan preeklampsia oleh para profesional medis.

Berdasarkan penelitian sebelumnya oleh Gusti dkk., penelitian ini fokus pada klasifikasi penyakit menular seksual menggunakan algoritme Naïve Bayes. Implementasi algoritme Naïve Bayes dalam penelitian ini mencapai tingkat optimal dalam menangani masalah klasifikasi, dengan mencapai tingkat akurasi sebesar 76,67% [9]. Meskipun demikian, penelitian ini tidak membahas tentang preeklampsia dan belum melibatkan pembuatan aplikasi web sederhana untuk mengimplementasikan model tersebut.

Penelitian lain yang dilakukan oleh Nuromah dkk., berhasil mengembangkan suatu sistem atau aplikasi web sederhana untuk prediksi awal preeklampsia, menggunakan algoritme C4.5. Hasil pengujian sistem menggunakan *confusion matrix* dan menerapkan metode *10-fold cross-validation* menunjukkan tingkat akurasi sebesar 81,38%, presisi 78,37%, recall 79,69%, dan *F1-score* 78,73%.

Penggunaan algoritme C4.5 dalam sistem prediksi dini preeklampsia dianggap oleh peneliti sebagai solusi yang sangat efektif [10]. Meski demikian, penelitian ini tidak memasukkan data hematologi dalam klasifikasi parahnya preeklampsia, dan masih diperlukan penelitian lebih lanjut untuk mengevaluasi apakah terdapat algoritme lain yang memiliki akurasi lebih baik daripada algoritme C4.5.

Pemilihan algoritme Naïve Bayes dilakukan untuk melakukan perbandingan kinerja model dengan algoritme yang telah digunakan dalam penelitian sebelumnya, yakni C4.5 yang diimplementasikan oleh Nurrohmah dkk. Algoritme Naïve Bayes dipilih juga karena banyaknya penelitian yang menggunakan algoritme tersebut dalam klasifikasi penyakit. Dalam konteks ini, Naïve Bayes dianggap terlebih dahulu memiliki potensi untuk mengklasifikasikan preeklampsia berat pada ibu hamil berdasarkan data hematologi dengan tingkat akurasi yang lebih tinggi dibandingkan C4.5.

Tujuan penelitian saat ini yaitu untuk mengetahui perbandingan mana yang lebih baik dari metode klasifikasi algoritme Naïve Bayes dan C4.5 dalam mengklasifikasikan penyakit preeklampsia berdasarkan data hematologi. Perbedaan penelitian ini dan penelitian-penelitian sebelumnya yang mencakup kedua algoritme tersebut serta penelitian-penelitian sebelumnya yang membahas preeklampsia, penelitian-penelitian sebelumnya itu belum membahas perbandingan Naïve Bayes dan C4.5 dalam klasifikasi preeklampsia. Perbedaan lainnya mencakup objek penelitian dan tahun penelitian yang berbeda, dengan penekanan pada perbandingan algoritme untuk mengklasifikasikan kasus preeklampsia yang parah. Adapun metodologi yang dipakai penelitian ini yaitu KDD (Knowledge Discovery in Database) yang memungkinkan membantu penambangan data yang terstruktur dan terarah dari awal hingga akhir dan model dari pembelajaran mesin pada penelitian ini juga akan di-*deploy* menjadi web sederhana.

## 2. TINJAUAN PUSTAKA

Preeklampsia merupakan suatu sindrom klinis akut yang sering terjadi selama masa kehamilan. Sindrom ini dicirikan oleh kehadiran hipertensi dan proteinuria, kadang-kadang disertai edema, dan biasanya muncul setelah kehamilan mencapai usia 20 minggu. Hipertensi dalam preeklampsia didefinisikan sebagai peningkatan tekanan darah, dengan nilai tekanan sistolik dan diastolik yang melebihi 140/90 mmHg, yang diukur setidaknya dua kali dengan selang waktu 6 jam. Ketika proteinuria, yaitu peningkatan jumlah protein dalam urin, terjadi bersamaan dengan hipertensi, itu menjadi tanda preeklampsia [11].

Biasanya, proteinuria muncul ketika jumlah protein yang ditemukan dalam urin selama periode 24 jam melebihi 0,3 gram atau ketika hasil pemeriksaan metode dipstik menunjukkan angka 1+ pada sampel urin acak. Jika proteinuria terjadi sebelum hipertensi,

hal ini dapat mengindikasikan adanya masalah ginjal daripada preeklamsia [12].

Naïve Bayes adalah suatu metode klasifikasi yang sederhana, berdasarkan pada teorema Bayes, yang memanfaatkan probabilitas dengan asumsi bahwa setiap kejadian adalah independen satu sama lain [13]. Pengklasifikasi Naïve Bayes merupakan produk dari implementasi metode klasifikasi Bayesian yang menggunakan teorema Bayes. Teorema ini menggambarkan relasi probabilitas bersyarat antara berbagai parameter statistik. Dalam kerangka klasifikasi Bayesian, tujuannya adalah menentukan probabilitas suatu label berdasarkan observasi fitur-fitur tertentu, yang dapat diartikan sebagai  $P(H|X)$  [14]. Teorema Bayes memberi tahu cara menyatakannya dalam kuantitas yang dapat dihitung secara lebih langsung

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (1)$$

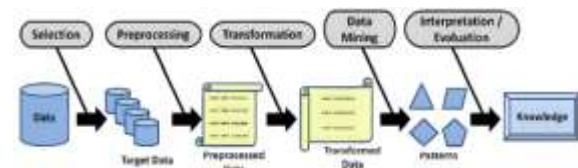
Keterangan:

- X = Sampel data yang memiliki kelas
- H = Hipotesis X pada data kelas
- P(X) = Probabilitas sebelumnya bahwa bukti tersebut benar
- P(H) = Probabilitas hipotesis H (prior)
- P(H|X) = Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)
- P(X|H) = Probabilitas data sampel X akan mempengaruhi hipotesis H (likelihood)

Algoritme C4.5 adalah algoritme pembangunan pohon keputusan, sebuah metode umum untuk klasifikasi dan prediksi data. Dikembangkan sebagai evolusi dari algoritme ID3 oleh J. Ross Quinlan, C4.5 memulai pembentukan pohon dengan memilih atribut sebagai akar, membuat cabang untuk setiap nilai atribut, dan membagi kasus ke dalam cabang tersebut. Proses ini berulang hingga semua kasus memiliki kelas yang sama. Pohon keputusan membantu menganalisis data dan mengungkapkan relasi antar atribut yang tersembunyi [15].

Metodologi KDD (Knowledge Discovery in Databases) yang diterapkan dalam penelitian ini mengacu pada pendekatan yang digunakan oleh Hidayat dkk. dalam penelitiannya, yang berfokus pada simulasi optimalisasi pengadaan barang dengan metode K-Mean Clustering. Pemilihan metode KDD ini didasarkan pada kesesuaian dengan kebutuhan penelitian saat ini, mengingat bahwa pendekatan KDD yang diaplikasikan dalam penelitian Hidayat sesuai dengan tujuan penelitian ini. KDD, atau Knowledge Discovery in Databases, merupakan proses antardisiplin yang melibatkan statistik, basis data, kecerdasan buatan (AI), visualisasi, dan komputasi paralel. Langkah-langkah utama dalam KDD mencakup pemilihan data, pembersihan data, transformasi data, penyajian, penemuan, dan evaluasi. Proses KDD digunakan untuk menghasilkan pengetahuan dari basis data, yang kemudian menjadi

landasan pengetahuan dalam pengambilan keputusan [16].



Gambar 2. Knowledge Discovery in Databases

Python, dikembangkan oleh Guido van Rossum pada tahun 1991, adalah bahasa pemrograman dinamis dan diinterpretasi dengan fokus pada keterbacaan kode. Dibandingkan dengan bahasa kompilasi, Python memungkinkan penulisan kode yang lebih singkat dan fleksibel karena tidak memerlukan deklarasi tipe variabel dan fungsi. Python mendukung modul dan paket, menekankan keterbacaan kode, serta dapat digunakan sebagai bahasa skrip atau perekat untuk menghubungkan komponen. Interpreter dan pustaka standar Python tersedia secara bebas untuk semua platform utama [17].

Flask adalah kerangka pengembangan web Python yang dikenal sebagai "kerangka mikro Python" karena menjaga inti tetap sederhana, memungkinkan perluasan sesuai kebutuhan. Berbeda dengan Django, Flask tidak memiliki fasilitas bawaan untuk berinteraksi dengan database. Sebagai kerangka prototyping, Flask fokus pada fungsionalitas dasar dan memberi keleluasaan kepada pengguna untuk menambahkan persyaratan. Fiturnya mencakup routing URL, templating menggunakan Jinja2, mode debug, dan ekstensi untuk fungsionalitas tambahan. Flask dapat diinstal dengan mudah melalui pip, dan tutorial serta sumber daya online membantu pengguna memulai pengembangan aplikasi web [18].

Tailwind CSS adalah kerangka kerja CSS yang menggunakan pendekatan "utility-first" dengan menyediakan kelas utilitas sederhana untuk menciptakan efek yang kompleks. Kerangka kerja ini memfasilitasi perancangan situs web tanpa menulis kode CSS secara manual. Berbeda dengan Bootstrap, yang menggunakan kelas berbasis komponen, Tailwind memecah gaya menjadi komponen kecil yang dapat digunakan kembali. Tailwind tidak memberlakukan spesifikasi desain, melainkan memungkinkan pengembang untuk menggabungkan komponen kecil menjadi pengalaman pengguna yang unik. Dengan menggunakan file CSS "mentah" dan memprosesnya melalui file konfigurasi, Tailwind menghasilkan desain situs web dengan mudah [19].

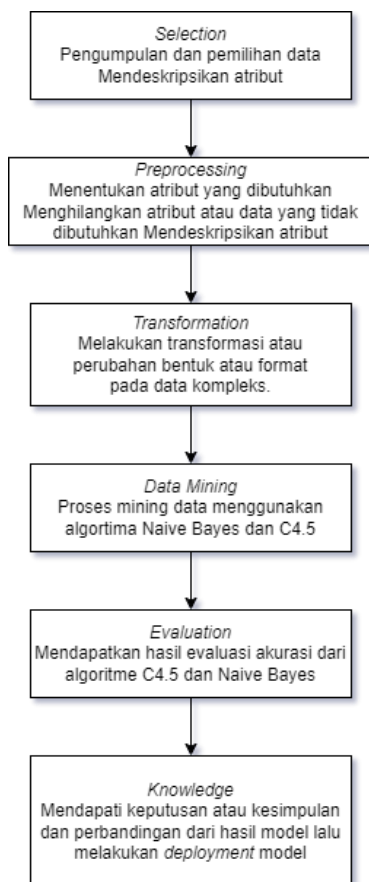
### 3. METODE PENELITIAN

#### 3.1. Tahapan Penelitian

Penelitian ini menerapkan metodologi Knowledge Discovery in Database (KDD), suatu rangkaian tindakan untuk mengidentifikasi data dan mengubahnya menjadi informasi yang mengungkapkan pola-pola tersembunyi. Pemilihan KDD dilakukan karena metode ini merupakan salah

satu teknik dalam ilmu data yang dapat diterapkan pada basis data besar yang sebelumnya belum memanfaatkan potensi informasi yang terkandung dalam data tersebut. Proses KDD melibatkan serangkaian langkah, termasuk Seleksi, *Preprocessing*, Transformasi, *Data Mining*, Evaluasi, dan *Knowledge*.

Dalam mengevaluasi perbandingan algoritme untuk mengklasifikasikan tingkat keparahan preeklampsia pada ibu hamil, setiap langkah disesuaikan dengan prosedur yang terdapat dalam model Knowledge Discovery in Database (KDD). Rancangan penelitian ini dijelaskan dengan gambaran ilustratif dalam Gambar 3.



Gambar 3. Tahapan penelitian

### 3.2. Pengumpulan Data

Penelitian ini berfokus pada analisis perbandingan algoritme C4.5 dan Naïve Bayes dalam klasifikasi parahnya preeklampsia pada ibu hamil dengan dataset yang diperoleh dari Rumah Sakit Bayukarta Karawang selama periode 2021-2023. Data yang digunakan penelitian ini merupakan data primer. Adapun untuk teknik yang digunakan untuk mengumpulkan data penelitian ini yaitu menggunakan metode observasi. Data preeklampsia yang dikoleksi atau dikumpulkan secara langsung dari bagian rekam medis Rumah Sakit Bayukarta Karawang.

Parameter dari data yang dibutuhkan pada penelitian mencakup berat badan, tinggi badan, tekanan darah, kadar protein dalam urin, dan data hematologi ibu atau selengkapnya dapat dilihat pada Tabel 1.

Tabel 1. Atribut Dataset

| Variable Name   | Unit                                      |
|-----------------|-------------------------------------------|
| Age             | year                                      |
| Height          | cm                                        |
| Weight          | kg                                        |
| Gestational Age | week                                      |
| Gravida         | number                                    |
| Parity          | number                                    |
| Abortus         | number                                    |
| Systolic        | mmHg                                      |
| Diastolic       | mmHg                                      |
| Proteinuria     | 0/1+/2+/3+/4+                             |
| Delivery Method | SC/Spontaneous/Nullipara                  |
| Creatinine      | mg/dL                                     |
| Hemoglobin      | g/dL                                      |
| Leukocytes      | 10 <sup>3</sup> /mcL                      |
| Hematocrit      | percentage                                |
| Platelets       | 10 <sup>3</sup> /mcL                      |
| Erythrocytes    | 10 <sup>6</sup> /mcL                      |
| Prognosis       | Severe Preeclampsia/Moderate Preeclampsia |

### 3.3. Evaluasi Hasil

Metrik pengujian perlu dilakukan agar mengetahui perbandingan dari kedua algoritme Naïve Bayes dan C4.5. Adapun metrik pengujian pada bagian evaluasi (*evaluation*) yang perlu diketahui dan dilakukan yaitu dengan menggunakan *confusion matrix*. Metrik tersebut digunakan untuk mengetahui nilai dari *precision*, *recall*, *f-measure* (*f1-score*), dan *accuracy*. Nilai-nilai tersebut akan dirangkum dalam bentuk tabel yang disebut tabel laporan klasifikasi. Perhitungan *precision*, *recall*, *f-measure* dan *accuracy* dilakukan untuk setiap pengujian dari masing-masing algoritme yaitu C4.5 dan Naïve Bayes dan dicari total nilai rata-rata dari setiap kelas masing-masing algoritme tersebut (*Severe Preeclampsia* atau *Moderate Preeclampsia*).

## 4. HASIL DAN PEMBAHASAN

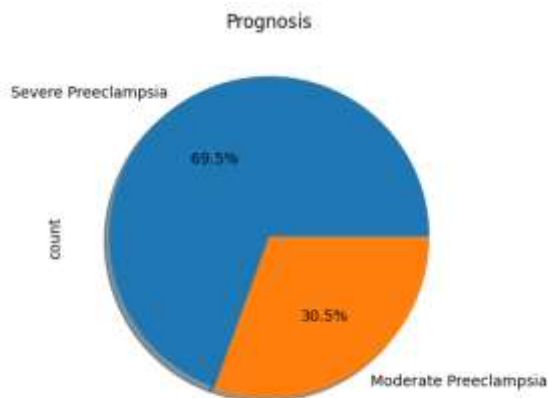
### 4.1. Selection

Pada tahap ini, dilakukan pengumpulan data yang berasal dari dataset preeklampsia berkategori tingkat keparahan pada pasien ibu hamil di RS Bayukarta Karawang selama periode 2021-2023. Data yang diambil dari rekam medis terdiri dari kurang lebih 30 atribut awal, kemudian atribut tersebut diseleksi untuk kebutuhan penelitian. Atribut yang dihilangkan seperti nama atau identitas pasien, lalu parameter-parameter rinci dari hemotologi dan urinalisa juga dihilangkan, dikarenakan hanya memerlukan beberapa parameter penting saja menurut dari penelitian sebelumnya. Setelah diseleksi dengan cermat, dataset menjadi berjumlah 18 atribut dan 131 entri. Hasil seleksi dataset dapat dilihat pada Tabel 2.

Tabel 2. Informasi Dataset

| features        | types   | non_null_counts |
|-----------------|---------|-----------------|
| id              | int64   | 131             |
| Age             | int64   | 131             |
| Height          | int64   | 131             |
| Weight          | object  | 129             |
| Gestational Age | int64   | 131             |
| Gravida         | int64   | 131             |
| Parity          | int64   | 131             |
| Abortus         | int64   | 131             |
| Systolic        | int64   | 131             |
| Diastolic       | int64   | 131             |
| Proteinuria     | object  | 131             |
| Delivery Method | object  | 131             |
| Creatinine      | object  | 129             |
| Hemoglobin      | object  | 129             |
| Leukocytes      | object  | 129             |
| Hematocrit      | object  | 129             |
| Platelets       | float64 | 129             |
| Erythrocytes    | object  | 129             |
| Prognosis       | object  | 131             |

Dapat dilihat pada Gambar 4 perbandingan prognosis atau kelas target klasifikasi yaitu *Severe Preeclampsia* dan *Moderate Preeclampsia* dalam persentase.



Gambar 4. Perbandingan Prognosis

#### 4.2. Pre-processing

Dalam tahap *Data Preprocessing*, langkah pertama yang diambil adalah membersihkan data untuk memastikan kualitasnya sebelum dilakukan proses analisis. Pada langkah ini, kolom "id" dihapus karena tidak diperlukan dalam proses pemodelan. Selanjutnya, kolom "Weight" yang awalnya bertipe objek diubah menjadi float dan nilai yang kosong diisi dengan rata-rata menggunakan fungsi *mean*.

Selanjutnya, ditemukan bahwa terdapat banyak nilai yang kosong pada indeks ke-21 dan ke-89. Oleh karena itu, dua baris tersebut dihapus dari dataset untuk memastikan integritas data yang lebih baik. Lalu hasil akhir dari *preprocessing*, entri dataset menjadi berjumlah 129 saja.

#### 4.3. Transformation

*Data Transformation* diawali pada kolom-kolom selain "Delivery Method" dan "Prognosis," yang

semestinya memiliki tipe data numerik, diubah menjadi float guna memudahkan analisis selanjutnya.

Terdapat kolom seperti "Proteinuria," "Delivery Method," dan "Prognosis" telah diubah dari tipe data objek menjadi kategorikal numerik menggunakan *Label Encoder*. Transformasi dapat dilihat pada Tabel 3, Tabel 4 dan Tabel 5.

Tabel 3. Kolom Proteinuria

| Sebelum | Sesudah |
|---------|---------|
| 0       | 0       |
| 1+      | 1       |
| 2+      | 2       |
| 3+      | 3       |
| 4+      | 4       |

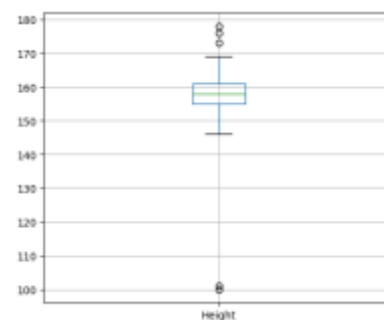
Tabel 4. Kolom Delivery Method

| Sebelum     | Sesudah |
|-------------|---------|
| Nullipara   | 0       |
| SC          | 1       |
| Spontaneous | 2       |

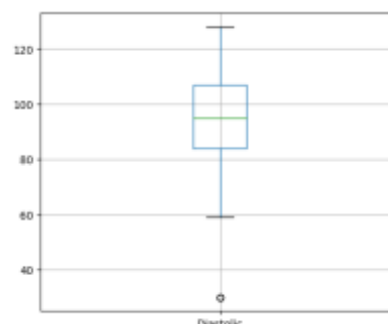
Tabel 5. Kolom Prognosis

| Sebelum               | Sesudah |
|-----------------------|---------|
| Moderate Preeclampsia | 0       |
| Severe Preeclampsia   | 1       |

Transformasi ini diperlukan untuk memudahkan mesin dalam pemrosesan model. Selanjutnya, setelah konversi ini, dilakukan kembali proses *preprocessing* data dikarenakan nilai dataset baru dapat diperiksa bila menggunakan tipe data numerik atau *float*. Setelah diperiksa, diketahui bahwa dataset ditemukan adanya outlier pada kolom "Weight" dan "Diastolic." Langkah ini diambil untuk meningkatkan integritas data sebelum melanjutkan ke tahap pemodelan.



Gambar 5. Height Outlier



Gambar 6. Diastolic Outlier

Dalam proses ini, outlier pada kolom "Weight" dan "Diastolic" diatasi dengan menggantinya dengan nilai rata-rata dari masing-masing kolom. Langkah ini membantu menghilangkan nilai yang ekstrem dan memastikan bahwa data yang digunakan untuk pemodelan adalah konsisten dan terstandarisasi.

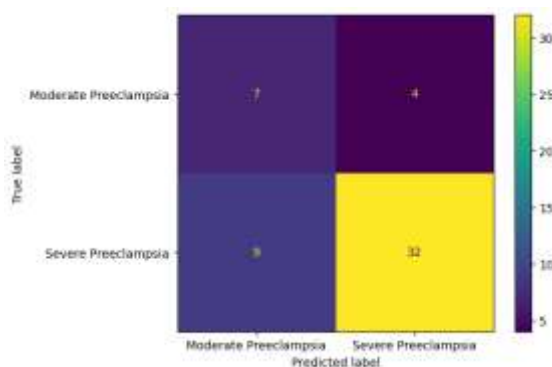
Selanjutnya, untuk mengikuti temuan penelitian sebelumnya, kolom "Weight" dan "Height" digabungkan atau diubah menjadi kolom baru yang disebut "BMI" (*Body Mass Index*). Setelah penggabungan ini, kolom "Weight" dan "Height" dihapus dari dataset. Proses ini dilakukan dengan tujuan untuk meningkatkan representasi informasi kesehatan secara lebih holistik dalam dataset preeklampsia.

#### 4.4. Data Mining

Dalam tahap Data Mining, dataset preeklampsia dibagi menjadi dua bagian: 60% untuk data testing dan 40% untuk data training, menggunakan metode *split dataset*. Penggunaan *random state shuffle* dengan nilai 42 digunakan untuk memastikan konsistensi hasil dan menghindari variasi yang tidak diinginkan pada setiap iterasi model. Selanjutnya, dilakukan proses pemodelan menggunakan dua algoritma, yaitu Decision Tree (C4.5) dan Naive Bayes. Kedua algoritma ini dipilih karena mampu menangani klasifikasi pada dataset medis seperti preeklampsia. Pada langkah ini, model diinisialisasi, dilatih menggunakan data training, dan dievaluasi menggunakan data testing untuk menilai kinerja dan generalisasi model.

#### 4.5. Evaluation

Setelah proses pemodelan, hasil evaluasi model dianalisis dengan menggunakan confusion matrix dan classification report. Untuk model Decision Tree (C4.5), *confusion matrix* menunjukkan 7 *True Negatives*, 32 *True Positives*, 4 *False Positives*, dan 9 *False Negatives*. Dengan menganalisis classification report, model ini memiliki akurasi sekitar 75%, dan F1-score yang baik pada kelas 1 (preeklampsia) sebesar 0.83. Hasil dapat dilihat pada Gambar 7 dan Tabel 5.

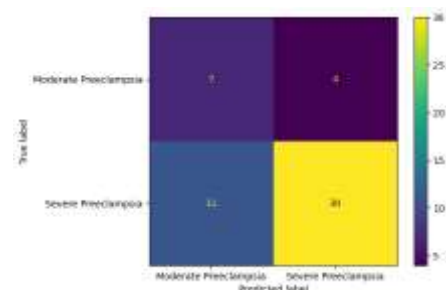


Gambar 7. C4.5 Confusion Matrix

Tabel 5. C4.5 Classification Result

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0,44      | 0,64   | 0,52     | 11      |
| 1            | 0,89      | 0,78   | 0,83     | 41      |
| accuracy     |           |        | 0,75     | 52      |
| macro avg    | 0,66      | 0,71   | 0,67     | 52      |
| weighted avg | 0,79      | 0,75   | 0,77     | 52      |

Sementara itu, model Naive Bayes memberikan *confusion matrix* dengan 7 *True Negatives*, 30 *True Positives*, 4 *False Positives*, dan 11 *False Negatives*. Dari classification report, model Naive Bayes memiliki akurasi sekitar 71%, dan F1-score yang cukup baik pada kelas 1 sebesar 0.80. Analisis ini memberikan pemahaman mendalam tentang sejauh mana kedua model dapat mengklasifikasikan preeklampsia dengan menggunakan dataset yang telah diproses dan diubah pada tahap sebelumnya. Hasil dapat dilihat pada Gambar 8 dan Tabel 6.



Gambar 8. Naive Bayes Confusion Matrix

Tabel 6. Naive Bayes Classification Result

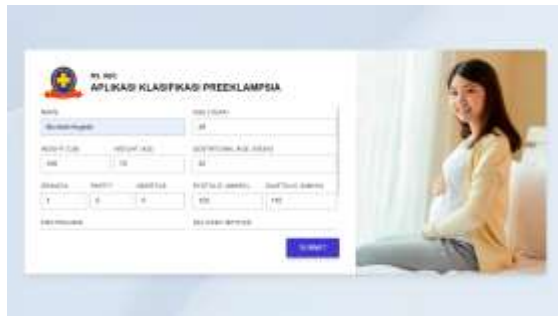
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0,39      | 0,64   | 0,48     | 11      |
| 1            | 0,88      | 0,73   | 0,80     | 41      |
| accuracy     |           |        | 0,71     | 52      |
| macro avg    | 0,64      | 0,68   | 0,64     | 52      |
| weighted avg | 0,78      | 0,71   | 0,73     | 52      |

#### 4.6. Knowledge

Setelah diidentifikasi, model Decision Tree menunjukkan tingkat ketepatan sekitar 75%, dengan nilai F1-score yang signifikan pada kelas 1 (preeklampsia) sebesar 0.83. Sebaliknya, model Naive Bayes mencapai tingkat ketepatan sekitar 71%, dengan nilai F1-score pada kelas 1 sebesar 0.80. Hasil ini memberikan gambaran menyeluruh tentang kinerja kedua model dalam klasifikasi preeklampsia, di mana Decision Tree menunjukkan tingkat ketepatan yang sedikit lebih tinggi jika dibandingkan dengan Naive Bayes.

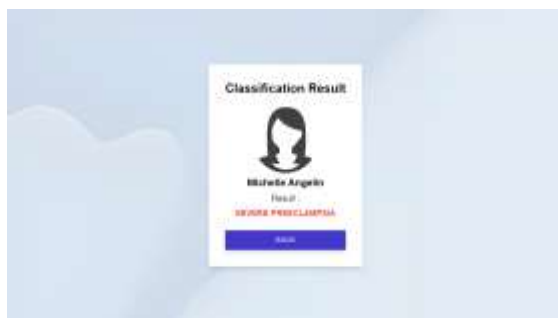
Decision Tree dipilih untuk dilakukan *deployment* karena menghasilkan tingkat akurasi yang lebih baik daripada Naive Bayes. *Deployment* yang dilakukan yaitu mengimplementasikan model menjadi halaman web sederhana dengan menggunakan Flask sebagai *web server* dan TailwindCSS sebagai *front-end*.

Halaman masukan atau halaman pertama memunculkan *input form* yang digunakan untuk masukan data yang akan diprediksi nantinya, dapat dilihat pada Gambar 9.



Gambar 9. Halaman masukan

Halaman hasil prediksi atau halaman kedua merupakan hasil keluaran prediksi model, halaman ini memunculkan nama dan hasil prediksi (*Moderate Preeclampsia/Severe Preeclampsia*), dapat dilihat pada Gambar 10.



Gambar 10. Halaman hasil prediksi

## 5. KESIMPULAN DAN SARAN

Kesimpulan yang diperoleh yaitu penelitian ini mencoba mengklasifikasikan preeklampsia dengan menggunakan Decision Tree (C4.5) dan Naive Bayes berdasarkan dataset Rumah Sakit Bayukarta Karawang 2021-2023. Analisis perbandingan menunjukkan Decision Tree memiliki akurasi lebih tinggi (75%) dibandingkan Naive Bayes (71%) dengan F1-score kelas 1 sebesar 0.83 dan 0.80 masing-masing. Model C4.5 berhasil di-deploy sebagai web sederhana dengan Flask dan TailwindCSS untuk memprediksi data baru. Terdapat beberapa saran yang dapat diambil untuk pengembangan penelitian selanjutnya yaitu perlunya mendalami analisis fitur preeklampsia, meningkatkan akurasi model, dan menjelajahi metode pengolahan data yang lebih canggih. Melibatkan dataset lebih besar, diversifikasi, dan penyelidikan faktor risiko tambahan dapat meningkatkan representativitas. Penggabungan hasil dari model lain atau penerapan ensemble learning juga dapat diterapkan untuk meningkatkan performa klasifikasi. Dengan penerapan saran-saran ini, diharapkan penelitian selanjutnya dapat lebih berkontribusi dalam pemahaman dan pencegahan preeklampsia serta meningkatkan ketepatan model pada awal kehamilan.

## DAFTAR PUSTAKA

- [1] N. Masruroh *et al.*, "Mengenal dan Mencegah Preeklampsia Pada Masa Kehamilan Di Kelurahan Wonokromo Surabaya," 2021. [Online]. Available: <https://snpm.unusa.ac.id>
- [2] A. Yunita, S. Irene Putri, and E. Yunicha Viridula, "Systematic Review: Deteksi Dini Pencegahan Preeklampsia Pada Ibu Hamil," 2021.
- [3] Kemenkes, "Pusat Data dan Informasi Kementerian Kesehatan RI."
- [4] WHO, "Maternal mortality," WHO International.
- [5] V. B. Kusnandar, "10 Provinsi dengan Angka Kematian Ibu Terbanyak pada 2020."
- [6] E. S. Wulandari, E. Ernawati, and D. Nuswantoro, "Risk Factors Of Preeclampsia With Severe Features And Its Complications," *Indonesian Midwifery and Health Sciences Journal*, vol. 5, no. 1, pp. 29–37, Jan. 2021, doi: 10.20473/imhsj.v5i1.2021.29-37.
- [7] L. C. Poon *et al.*, "The International Federation of Gynecology and Obstetrics (FIGO) initiative on pre-eclampsia: A pragmatic guide for first-trimester screening and prevention," *International Journal of Gynecology and Obstetrics*, vol. 145, no. S1, pp. 1–33, May 2019, doi: 10.1002/ijgo.12802.
- [8] C. C. Giyanto and B. A. Pramono, "MEDIA MEDIKA MUDA PERBANDINGAN PROFIL HEMATOLOGI PADA PREEKLAMPSIA/EKLAMPSIA DENGAN KEHAMILAN NORMOTENSI DI RSUP DR. KARIADI SEMARANG," 2015.
- [9] G. E. Yulastuti, C. Nurina Prabiantissa, and A. M. Rizki, "Klasifikasi penyakit Menular Seksual Menggunakan Naïve Bayes," *INTEGER: Journal of Information Technology*, vol. 7, no. 1, 2022.
- [10] S. Nurrohman, D. Normawati, A. Dahlan, J. Ringroad Selatan, and I. Yogyakarta, "Prediksi Dini Penyakit Preeklampsia Menggunakan Algoritma C4.5," vol. 10, no. 3, pp. 120–132, 2022, doi: 10.12928/jstie.v8i3.xxx.
- [11] N. Fumtidha, T. Ekasari, and M. Zakiyyah, "PENGARUH TEKNIK RELAKSASI AUTOGENIK TERHADAP TEKanan DARAH IBU HAMIL DENGAN PREEKLAMASIA DI KLINIK UTAMA PANASEA KABUPATEN LUMAJANG," *Jurnal Ilmiah Obsgin*, 2023, [Online]. Available: <https://stikes-nhm.e-journal.id/OBJ/index>
- [12] S. S. Gunawan, D. Triwidiyanti, and H. Syafrullah, "HUBUNGAN KONSUMSI ASAM FOLAT DENGAN PREEKLAMPSIA DI RSUD SUBANG TAHUN 2022," *JURNAL SEHAT MASADA*, vol. 17, no. 1, 2023.
- [13] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online JD.Id Menggunakan Metode Naïve Bayes Classifier Berbasis

- Konversi Ikon Emosi,” *Jurnal SIMETRIS*, vol. 10, no. 2, 2019.
- [14] J. VanderPlas, *Python Data Science Handbook*, 2nd ed., vol. 1. O’Reilly Media, Inc, 2019. [Online]. Available: [www.allitebooks.com](http://www.allitebooks.com)
- [15] M. W. Prihatmono and A. F. Watratan, “Implementasi Algoritma C4.5 Menggunakan Python Untuk Klasifikasi Kepuasan Konsumen,” *Jurnal Informatika Progres*, vol. 11, no. 2, 2019, doi: <https://doi.org/10.56708/progres.v11i2.146>.
- [16] I. S. Hidayat, S. Defit, and G. W. Nurcahyo, “Simulasi dalam Optimalisasi Pengadaan Barang menggunakan Metode K-Mean Clustering,” *Jurnal Sistim Informasi dan Teknologi*, vol. 3, no. 4, 2021, doi: [10.37034/jsisfotek.v3i3.79](https://doi.org/10.37034/jsisfotek.v3i3.79).
- [17] Python, “What is Python? Executive Summary.” Accessed: Nov. 07, 2023. [Online]. Available: [www.python.org](http://www.python.org)
- [18] J. Hunt, *Advanced Guide to Python 3 Programming*, 1st ed., vol. 1. Springer Nature Switzerland, 2019. doi: <https://doi.org/10.1007/978-3-030-25943-3>.
- [19] A. Yudhistira, L. Desy Pangesti, G. Isran, R. Bagus, B. Sumantri, and R. Suryani, “Perancangan dan Implementasi Sistem Informasi Perpustakaan Berbasis Web,” *Jurnal Sistem Informasi dan Komputerisasi Akuntansi*, vol. 7, no. 1, pp. 2579–4477, 2023.