

# Prediksi Dini Penyakit Preeklamsia Menggunakan Algoritma C4.5

Siti Nurrohmah<sup>a,1,\*</sup>, Dwi Normawati<sup>b,2</sup>

<sup>a,b</sup>Program Studi S1 Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan  
Jl. Ringroad Selatan, Kragilan, Tamanan, Banguntapan, Bantul, Daerah Istimewa Yogyakarta 55191  
<sup>1</sup>[siti1500018066@webmail.uad.ac.id](mailto:siti1500018066@webmail.uad.ac.id); <sup>2</sup>[dwi.normawati@tif.uad.ac.id](mailto:dwi.normawati@tif.uad.ac.id);

\*Penulis Korespondensi

## ABSTRAK

Berdasarkan data Kemenkes RI tahun 2021 menunjukkan angka kematian ibu tinggi yaitu lebih dari 4000 kasus setiap tahunnya dimana salah satu penyebabnya adalah preeklamsia. Pencegahan preeklamsia cukup sulit dikarenakan gejala utamanya belum diketahui pasti. Namun teknologi dapat digunakan untuk membantu pendeteksian preeklamsia. Penelitian ini bertujuan mendeteksi preeklamsia pada ibu hamil menggunakan algoritma C4.5. Tahapan pertama penelitian ini adalah melakukan studi literatur. Kemudian mengumpulkan data di RSKIA Sadewa Yogyakarta dan mengolahnya melalui tahapan preprocessing dengan melakukan seleksi data, transformasi data, membagi data menggunakan 10-fold cross validation. Selanjutnya data dianalisis menggunakan algoritma C4.5 dan diimplementasikan ke dalam sistem. Penelitian ini menggunakan data sebanyak 870 data dengan atribut pendidikan, pekerjaan, usia, usia kehamilan, tekanan darah, berat badan, jenis kehamilan, jumlah kelahiran, riwayat aborsi, riwayat persalinan, riwayat penyakit, dan proteinuria serta kelas klasifikasi negatif, preeklamsia ringan, dan preeklamsia berat. Penelitian ini menghasilkan sebuah sistem prediksi dini penyakit preeklamsia pada ibu hamil. Hasil pengujian menggunakan confusion matrix menunjukkan bahwa sistem prediksi mendapatkan nilai akurasi 81,38%, precision 78,37%, recall 79,69%, dan f1-score 78,73%. Hasil pengujian black box menunjukkan fungsi sistem dapat digunakan dengan baik.



### Kata Kunci

Data Mining  
C4.5  
Cross Validation  
Confusion Matrix  
Preeklamsia



This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

## 1. Pendahuluan

Kesehatan merupakan faktor yang sangat penting dalam menjamin keselamatan proses persalinan ibu hamil dan janin yang dikandungnya. Upaya peningkatan kesehatan ibu hamil perlu mendapat perhatian dan pengawasan khusus. Berdasarkan data Kementerian Kesehatan [1] tahun 2021, angka kematian ibu masih tergolong tinggi di Indonesia. Turun dari 4.999 kasus pada 2015 menjadi 4.912 kasus pada 2016. Lalu turun kembali 1.712 kasus pada 2017. Dengan cepat menjadi 4.197 kasus pada 2019 dan naik 10,25% atau 4.627 kasus pada 2020.

Komplikasi utama yang menyebabkan sekitar 75% kematian ibu adalah pendarahan 27%, preeklamsia dan eklamsia 14%, infeksi 11%, persalinan obstruktif 9%, dan komplikasi terkait aborsi 8% [2]. Penyebab kematian ibu di Indonesia pada tahun 2019 adalah pendarahan 28,29%, hipertensi 23% dan gangguan peredaran darah 4,94% [3].

Tekanan darah tinggi adalah salah satu gejala utama preeklamsia. Preeklamsia merupakan penyakit yang tidak diketahui etiologinya, terutama pada ibu hamil. Bentuk sindrom ini ditandai dengan hipertensi dan proteinuria yang terjadi setelah minggu ke-20 kehamilan [2]. Rasio preeklamsia di negara berkembang berkisar antara 3% hingga 7%. Meskipun persentase ini lebih kecil di negara maju, berkisar antara 0,05% hingga 0,1%. Di Indonesia sendiri, preeklamsia menyebabkan 1,5% hingga 25% dari kematian ibu [4]. Penyebab utama terjadinya preeklamsia belum diketahui, sehingga pencegahannya masih sangat sulit. Oleh

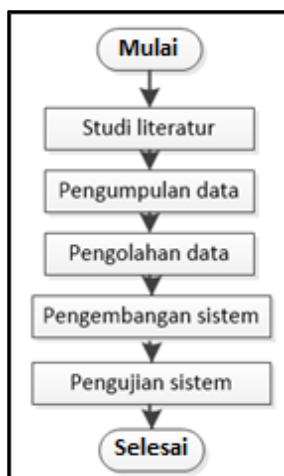
karena itu, ibu hamil harus melakukan upaya pencegahan, dengan perhatian khusus untuk mengatasi dan menganalisis peningkatan hipertensi pada kematian tahunan ibu di Indonesia.

Berdasarkan penelitian terdahulu, model data mining digunakan untuk memprediksi penyakit hipertensi kehamilan dengan metode *decision tree* [5]. Penelitian tersebut menggunakan data latih sebanyak 286 *instances* dengan kesalahan 7.3427% dan akurasi sebesar 92.6573%. Kelebihan dari metode *decision tree* ini adalah kesederhanaan dalam algoritmanya namun dapat menghasilkan klasifikasi dan memprediksi dengan baik. Kelemahan dari penelitian ini adalah belum menghasilkan sebuah sistem, melainkan berupa model *data mining* dan hasil akurasi yang dihasilkan masih dihitung menggunakan aplikasi Weka [5].

Berdasarkan pemaparan permasalahan tersebut, tingginya angka kematian ibudi Indonesia karena tekanan darah tinggi meningkat setiap tahun meningkat. Untuk itu penelitian ini akan mengembangkan model data mining dari penelitian sebelumnya menjadi sebuah sistem menggunakan algoritma *C4.5*. Penelitian ini bertujuan untuk memprediksi kemungkinan terjadinya preeklamsia pada ibu hamil. Sehingga dapat digunakan sebagai alat bantu dalam melakukan pendeteksian preeklamsia sejak dini dan untuk selanjutnya mendapatkan perhatian dan tindakan yang tepat selama masa kehamilan yang berpengaruh terhadap pengurangan tingkat angka kematian pada ibu hamil.

## 2. Metode

Tahapan penelitian dilakukan dengan proses pengembangan penelitian seperti yang dapat ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Metode Penelitian

Penelitian ini dilakukan dengan beberapa tahapan:

### 2.1. Studi Literatur

Pada tahap ini penelitian dilakukan dengan mengidentifikasi teori-teori pendukung berupa artikel jurnal, artikel konferensi, buku, e-book, karya penelitian terdahulu yang berkaitan dengan penelitian berupa topik *data mining*, pohon keputusan, *C4.5*, dan preeklamsia. Tahapan ini dilakukan untuk mempelajari, mengetahui, memahami, dan membandingkan penelitian sebelumnya dengan penelitian ini.

### 2.2. Pengumpulan Data

Tahap ini dilakukan dengan mengumpulkan data-data yang dibutuhkan yang diperoleh dari RSKIA Sadewa Yogyakarta. Data ini memiliki atribut yaitu pendidikan, pekerjaan, usia ibu, usia kehamilan, tekanan darah, berat ibu, jenis kehamilan, paritas, riwayat persalinan, riwayat abortus, riwayat penyakit, dan proteinuria. Sedangkan kelas klasifikasi yang digunakan adalah negatif, preeklamsia ringan, dan preeklamsia berat.

### 2.3. Pengolahan Data

Data yang telah diperoleh sebelum dianalisis menggunakan algoritma *C4.5* harus melalui tahapan *preprocessing* terlebih dahulu. Tahapan *preprocessing* sebagai berikut.

- Seleksi Data

Dalam proses seleksi ini, data dipilih untuk menentukan atribut yang akan digunakan dan yang akan dihapus atau dihilangkan. Data terpilih adalah data yang diperkirakan dapat menambah hasil uji validitas sistem. Dalam proses ini juga bisa terjadi penambahan data awal karena dilakukan pemecahan data.

- Transformasi Data

Data yang telah terseleksi kemudian diubah menjadi satu tipe data yaitu data numerik menurut interval yang ditentukan berdasarkan teori terkait faktor resiko preeklamsia agar dapat diproses menggunakan data mining. Aturan transformasi dapat dilihat pada Tabel 1.

**Tabel 1.** Aturan Transformasi Data

Atribut	Keterangan	Transformasi
Pendidikan [6]	Jenjang pendidikan terakhir yang pernah ditempuh ibu	0 : Dasar (SD/SLTP) 1 : Menengah (SLTA) 2 : Tinggi (Akademik / Universitas)
Pekerjaan [7]	Kegiatan yang dilakukan ibu selama proses kehamilan.	0 : Tidak Bekerja (IRT/Mahasiswa) 1 : Bekerja (Selain IRT/Mahasiswa)
Usia Ibu (Tahun) [8]	Usia ibu pada saat kehamilan	0 : Tidak Beresiko (Usia 20-35 tahun) 1 : Beresiko (Usia <20 atau >35 tahun)
Usia Kandungan (minggu) [2]	Usia janin yang dikandung pada saat pemeriksaan	0 : Trimester I (0-13 minggu) 1 : Trimester II (14-26 minggu) 2 : Trimester III (27 minggu-melahirkan)
Tekanan darah [2]	Tekanan darah ibu pada saat pemeriksaan kandungan	0 : Normal (<140/90 mmHg) 1 : Hipertensi (>=140/90 mmHg)
Jenis kehamilan [9]	Jumlah Janin yang dikandung pada saat kehamilan	0 : Tidak Beresiko (Normal) 1 : Beresiko (Kembar)
Paritas [10]	Jumlah kelahiran yang pernah dialami ibu	0 : Tidak Beresiko (2-3) 1 : Beresiko (<2 atau >3)
Riwayat abortus	Kondisi dimana ibu pernah mengalami keguguran	0 : Belum (0) 1 : Pernah (>0)
Riwayat persalinan	Proses persalinan yang pernah dialami ibu sebelum kehamilan saat ini	0 : Belum 1 : SPT (Spontan) 2 : SC (Sectio Caesar) 3 : VE (Vakum Ekstrakor)
Riwayat penyakit [9]	Penyakit yang pernah atau sedang dialami ibu	0 : Tidak Ada 1 : Ada (Hipertensi, Diabetes, dan lain-lain)
Proteinuria [2]	Kondisi protein yang terdapat di dalam urine ibu	0 : negatif 1 : Trace 2 : Positif 1 3 : positif 2 4 : Positif 3 5 : Positif 4
Kelas	Kelas yang digunakan untuk klasifikasi	0 : Negatif (N) 1 : Preeklamsia Ringan (PE) 2 : Preeklamsia Berat (PEB)

- Split Data

Pada tahapan ini data akan dibagi menjadi data latih dan data uji menggunakan *cross validation*. Jumlah *fold* yang digunakan adalah 10. Data akan dibagi menjadi 10 *subset* (bagian). Pada *fold* pertama terdapat kombinasi 9 *subset* yang berbeda digabung dan digunakan sebagai data latih, sedangkan 1 *subset* (sisa) digunakan sebagai data uji. Selanjutnya proses latih dan uji dilakukan hingga *fold* kesepuluh [11].

## 2.4. Pengembangan Sistem

Pada tahap ini dilakukan proses analisis terhadap kebutuhan sistem terlebih dahulu dan dibuat rancangan antarmukanya. Selanjutnya algoritma C4.5 juga diterapkan ke dalam sistem untuk dilakukan proses analisis data. Berikut ini proses dari pengembangan sistem penelitian ini.

### 2.4.1. Analisis Sistem

Berikut analisis informasi yang dibutuhkan sistem baik kebutuhan fungsional dan non fungsional.

- Kebutuhan Fungsional

Kebutuhan fungsional adalah jenis kebutuhan yang berisi tentang proses-proses yang akan dijalankan sistem. Berikut ini kebutuhan fungsional untuk sistem prediksi yang diusulkan:

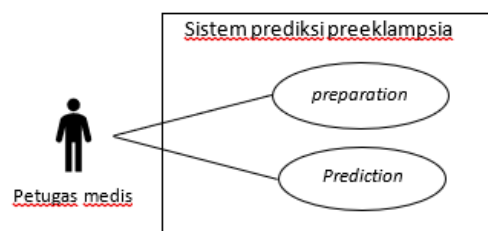
- Sistem dapat memuat file data.
  - Sistem dapat menampilkan data unggahan dalam bentuk tabel.
  - Sistem dapat menampilkan hasil akurasi dataset menggunakan  $K=10$  *cross validation* dan hasil *confusion matrix* beserta hasil rata-ratanya.
  - Sistem dapat *re-test* data unggahan tanpa harus mengunggah ulang dataset dan langsung menampilkan hasil akurasi terbaru.
  - Sistem dapat menerima inputan data rekam medis untuk dilakukan proses prediksi preeklampsia.
  - Sistem dapat menampilkan hasil estimasi dari data yang dimasukkan.
- Kebutuhan Non Fungsional

Kebutuhan non fungsional adalah analisis yang dibutuhkan untuk menentukan spesifikasi kebutuhan. Berikut ini kebutuhan non fungsional untuk sistem prediksi yang diusulkan:

- Sistem hanya dapat mengunggah *file* data rekam medis yang berekstensi *.csv*.
- Sistem hanya dapat mengunggah *file* data rekam medis dengan kapasitas maksimal 200 mb.

## 2.5. Perancangan Use Case Diagram

*Use case* diagram dirancang untuk mendeskripsikan interaksi yang terjadi antara user dengan sistem yang akan dikembangkan. Rancangannya dapat dilihat pada Gambar 2.

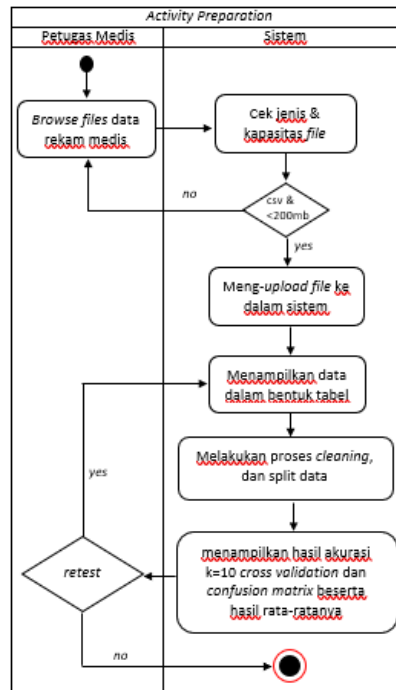


Gambar 2. Use Case Diagram

*User* atau pengguna dari sistem yang dibuat adalah petugas medis dimana dapat mengakses sistem pada menu *preparation* dan menu *prediction*.

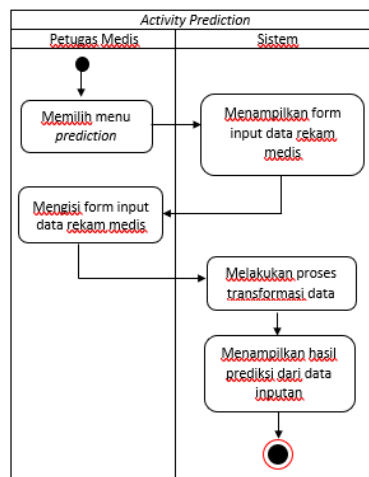
- Perancangan *Activity* Diagram

*Activity* diagram dirancang untuk menjelaskan urutan aktivitas proses yang dijalankan sistem yang akan dibuat dalam penelitian ini. Rancangannya dapat dilihat pada Gambar 3 dan Gambar 4.



**Gambar 3.** Activity Diagram Menu Preparation

Activity dimana petugas medis melakukan *upload* dataset, yaitu dengan mengklik tombol *browse files* lalu memilih *file* dataset yang telah berekstensi *csv*. Setelah *file* berhasil terbaca oleh sistem, maka akan ditampilkan dataset dalam bentuk tabel dan dibawahnya menampilkan hasil akurasi *k=10 cross validation* dan *confusion matrix* serta hasil rata-ratanya. Selain itu, petugas medis juga dapat me-*retest* dataset tanpa perlu meng-*upload* ulang data dan langsung dapat ditampilkan hasil akurasi terbaru.



**Gambar 4.** Activity Diagram Menu Prediction berikut.

Activity dimana petugas medis menginputkan data rekam medis terbaru untuk dilakukan prediksi penyakit preeklamsia yang hasilnya akan ditampilkan oleh sistem.

- Perancangan User Interface

Sistem diimplementasikan menggunakan bahasa pemrograman *Python* dan untuk tampilan antarmuka sistem atau *user interface* sistem menggunakan *framework streamlit*. Gambar 5 merupakan hasil rancangan sistem dimana terdapat dua menu yang bisa digunakan *user yaitu* rancangan menu prediction dan rancangan menu prediction pada Gambar 6.

Rata-rata peforma pengujian			
Avg accuracy	Avg precision	Avg recall	Avg F1-score

Gambar 5. Rancangan Menu Preparation

prediksi

Hasil prediksi: ...

Gambar 6. Rancangan Menu Prediction

## 2.6. Pengujian Sistem

Setelah sistem terbentuk selanjutnya dilakukan uji validitas dan fungsionalitas sistem menggunakan metode berikut.

- Black Box Testing

Selain itu untuk menguji fungsionalitas sistem, dilakukan pengujian menggunakan *black box*. Pengujian *black box* ini akan menentukan apakah sistem telah layak digunakan atau belum.

- Confusion Matrix

Peforma algoritma C4.5 yang bekerja pada sistem dapat diuji menggunakan *confusion matrix* untuk mengetahui nilai validitas berupa akurasi, presisi, *recall*, dan *f-1 score*. Karena kelas klasifikasi yang digunakan

dalam penelitian ini sebanyak tiga, berikut *confusion matrix* 3 variabel ditunjukkan pada Gambar 7 beserta cara perhitungannya ditunjukkan pada Gambar 8 [12].

Confusion Matrix		Predicted			False Negative (FN)	Recall
		Class 1	Class 2	Class 3		
Class 1	A	B	C	B + C	$A/(A + B + C)$	
Class 2	D	E	F	D + F	$E/(D + E + F)$	
Class 3	G	H	I	G + H	$I/(G + H + I)$	
Actual	False Positive (FP)	D + G	B + H	C + F	Overall Accuracy = $A + E + I / (\text{Sum of red and green squares})$	
	Precision	$A/(A + D + G)$	$E/(B + E + H)$	$I/(C + F + I)$		

■ True positive ■ True Negatives ■ Misclassified cases ■ False Positives ■ False Negatives.

Gambar 7. Confusion Matrix 3 Variabel

Performance Metric	Formula	Calculations
Accuracy	$(TP + TN) / (TP + TN + FP + FN)$	$(A + E) / (A + E + D + G + B + C)$
Precision	$TP / (TP + FP)$	$A / (A + D + G)$
Recall	$TP / (TP + FN)$	$A / (A + B + C)$
Specificity	$TN / (FP + TN)$	$(E + I) / (D + G + E + I)$
F-score	$2TP / (2TP + FP + FN)$	$(2 * A) / (2 * A + D + G + B + C)$
MCC	$TP * TN - FP * FN / \sqrt{((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN))}$	$(A * E - D * G + B * C) / \sqrt{(A + D + G) * (A + B + C) * (E + I + D + G) * (E + I + B + C)}$

Gambar 8. Rumus Confusion Matrix 3 Variabel

Keterangan :

*True Positive* (TP) = Label milik kelas dan diprediksi dengan benar

*False Positive* (FP) = Label bukan milik kelas tetapi pengklasifikasi diprediksi sebagai positif.

*True Negative* (TN) = Label bukan milik kelas dan diprediksi dengan benar

*False Positive* (FP) = Label memang milik kelas tetapi diprediksi sebagai negatif.

### 3. Hasil dan Pembahasan

#### 3.1. Pengumpulan Data

Pengumpulan data menggunakan teknik observasi langsung di RSKIA Sadewa Yogyakarta yaitu dengan melakukan dokumentasi berkas rekam medis ibu hamil secara mandiri. Diperoleh sebanyak 247 data rekam medis ibu hamil kategori penyakit Preeklampsia. Dari data tersebut 95 pasien terdiagnosis negatif preeklampsia, 120 pasien terdiagnosis positif preeklampsia ringan, dan 32 pasien lainnya terdiagnosis positif preeklampsia berat.

#### 3.2. Preprocessing Data

Tabel 2 merupakan sampel data rekam medis yang telah diperoleh dari RSKIA Sadewa Yogyakarta sebelum dilakukan *preprocessing* data.

Tabel 2. Data Rekam Medis

Pendidikan	Pekerjaan	Usia Ibu (tahun)	Usia Kandungan (minggu)	Tekanan Darah	Bb Ibu (kg)	Jenis Kehamilan	Paritas	Riwayat Abortus	Riwayat Persalinan	Riwayat Penyakit	Proteinuria	Kelas	
SLTA	IRT	37	26	108	73	69	Kembar	1	1	Caesar (SC)	Asma	Positif 1	PE
UNIVERSITAS	SWASTA	35	23	100	60	63,8	Normal	1	1	Vakum Ekstraktor (VE)	Tidak Ada	Trace	N
UNIVERSITAS	IRT	36	28	180	120	77	Normal	1	0	Caesar (SC)	Hipertensi, Asma	Positif 3	PEB
SD	IRT	18	40	120	70	84	Normal	0	0	Belum DM	Jantung	Negatif	N
SLTA	WIRASWASTA	39	38	140	90	78	Normal	0	0	Belum	Hipertensi	Positif 2	PE
AKADEMIK	WIRASWASTA	30	37	156	104	59	Normal	1	0	Spontan (SPT)	DM	Positif 1	PEB
SLTP	SWASTA	19	39	140	100	51,2	Normal	0	0	Belum	TBC	Positif 1	PE
SLTA	IRT	28	23	110	70	77	Normal	2	0	Spontan (SPT)	Tidak Ada	Negatif	N



- Seleksi Data

Pada tahap ini akan dilakukan proses penyeleksian data yang akan digunakan untuk proses prediksi. Dikarenakan terdapat kondisi dimana satu pasien memiliki beberapa data usia kandungan, oleh karena itu data setiap pasien yang memiliki beberapa data usia kandungan akan dipecah menjadi beberapa data. Hal ini dapat dilakukan karena data berat badan dari masing-masing usia kandungan memiliki bobot yang berbeda. Sehingga menurunkan kemungkinan terjadinya redudansi data karena data berat badan akan di anggap unik atau menjadi pembeda antara satu data dengan yang lainnya. Data rekam medis yang semula berjumlah 247 data, setelah dipecah maka data yang digunakan menjadi 870 data. Hasil seleksi data dapat dilihat Tabel 3 yang merujuk dari Tabel 2.

**Tabel 3 Hasil seleksi data**

Pendidikan	Pekerjaan	Usia Ibu (tahun)	Usia Kandungan (minggu)	Tekanan Darah	Bb Ibu (kg)	Jenis Kehamilan	Paritas	Riwayat Abortus	Riwayat Peralinan	Riwayat Penyakit	Proteinuria	Kelas
SLTA	IRT	37	26	108/73	69	Kembar	1	1	Caesar (SC)	Asma	Positif 1	PE
SLTA	WIRASWASTA	39	38	140/90	78	Normal	0	0	Belum	Hipertensi	Positif 2	PE
SLTA	WIRASWASTA	39	40	124/76	78.5	Normal	0	0	Belum	Hipertensi	Positif 2	PE
SLTP	SWASTA	19	39	140/100	51.2	Normal	0	0	Belum	TBC	Positif 1	PE
UNIVERSITAS	IRT	36	28	180/120	77	Normal	1	0	Caesar (SC)	Hipertensi	Positif 3	PEB
AKADEMIK	WIRASWASTA	30	37	156/104	59	Normal	1	0	Spontan (SPT)	Asma	Positif 1	PEB
AKADEMIK	WIRASWASTA	30	38	190/100	61.3	Normal	1	0	Spontan (SPT)	DM	Positif 1	PEB
SD	IRT	18	40	120/70	84	Normal	0	0	Belum	Jantung	Negatif	N
SLTA	IRT	28	23	110/70	77	Normal	2	0	Spontan (SPT)	DM	Negatif	N
UNIVERSITAS	SWASTA	35	23	100/60	63.8	Normal	1	1	Vakum Ekstraktor (VE)	Tidak Ada	Trace	N

- Transformasi Data

Pada tahap ini atribut penelitian dan kelas klasifikasi yang digunakan akan ditransformasi berdasarkan aturan yang telah di tetapkan pada Tabel 1. Aturan Transformasi Data. Atribut berat badan tidak ditransformasi karena memiliki data yang berbeda-beda, seperti ditunjukkan pada Tabel 4.

**Tabel 4. Hasil Transformasi Data**

Pendi dikan	Peker jaan	Usia ibu	Usia janin	TD	BB ibu	JK	Paritas	Abortus	Persa linan	Penya kit	Protei nuria	Kelas
1	0	1	1	0	69	1	1	1	2	1	2	1
1	1	1	2	1	78	0	1	0	0	1	3	1
1	1	1	2	0	78.5	0	1	0	0	1	3	1
0	1	1	2	1	51.2	0	1	0	0	1	2	1
2	0	1	2	1	77	0	1	0	2	1	4	2
2	1	0	2	1	59	0	1	0	1	1	2	2
2	1	0	2	1	61.3	0	1	0	1	1	2	2
0	0	1	2	0	84	0	1	0	0	1	0	0
1	0	0	1	0	77	0	0	0	1	0	0	0
2	1	0	1	0	63.8	0	1	1	3	0	1	0

- Split Data

Semula penelitian ini berhasil mengumpulkan data sebanyak 247 data rekam medis. Namun jika hanya menggunakan 247 data, sistem akan menghasilkan akurasi yang rendah. Oleh karena itu, dilakukan sebuah strategi untuk menaikkan hasil akurasi. Hasil pengamatan diperoleh bahwa pasien memiliki data usia kehamilan yang beragam dan masing-masing memiliki berat badan yang berbeda. Setelah itu dilakukan pemecahan data berdasarkan usia kehamilan. Sehingga data yang semula berjumlah 247 bertambah menjadi 870 data. Penelitian ini dilakukan menggunakan *10-fold cross validation* untuk menaikkan hasil akurasi sistem. Proses *split* data ditunjukkan pada Gambar 9 untuk menunjukkan hasil *10-fold cross validation*.



	iterasi 1	iterasi 2	iterasi 3	iterasi 4	iterasi 5	iterasi 6	iterasi 7	iterasi 8	iterasi 9	iterasi 10		
subset 1	1-87	88-174	175-261	262-348	349-435	436-522	523-609	610-696	697-783	784-870		data latih
subset 2	1-87	88-174	175-261	262-348	349-435	436-522	523-609	610-696	697-783	784-870		data uji
subset 3	1-87	88-174	175-261	262-348	349-435	436-522	523-609	610-696	697-783	784-870		
subset 4	1-87	88-174	175-261	262-348	349-435	436-522	523-609	610-696	697-783	784-870		
subset 5	1-87	88-174	175-261	262-348	349-435	436-522	523-609	610-696	697-783	784-870		
subset 6	1-87	88-174	175-261	262-348	349-435	436-522	523-609	610-696	697-783	784-870		
subset 7	1-87	88-174	175-261	262-348	349-435	436-522	523-609	610-696	697-783	784-870		
subset 8	1-87	88-174	175-261	262-348	349-435	436-522	523-609	610-696	697-783	784-870		
subset 9	1-87	88-174	175-261	262-348	349-435	436-522	523-609	610-696	697-783	784-870		
subset 10	1-87	88-174	175-261	262-348	349-435	436-522	523-609	610-696	697-783	784-870		

Gambar 9. Hasil K10-Fold Cross Validation

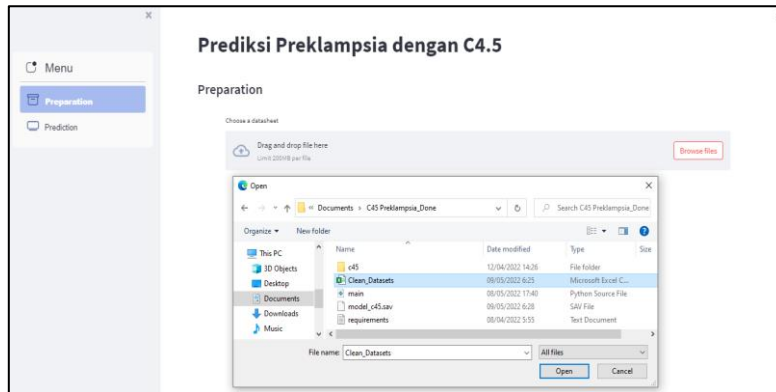
Data yang berjumlah 870 akan dibagi 10 sehingga menjadi 10 *fold* dari 1-*fold* hingga 10-*fold* bagian akan bergantian menjadi data latih dan data uji agar mendapatkan hasil validasi terhadap algoritma C4.5.

### 3.3. Implementasi Sistem

Pembuatan sistem merupakan tahapan akhir dalam penelitian ini. Berikut ini adalah beberapa hasil implementasi menu-menu yang terdapat pada sistem.

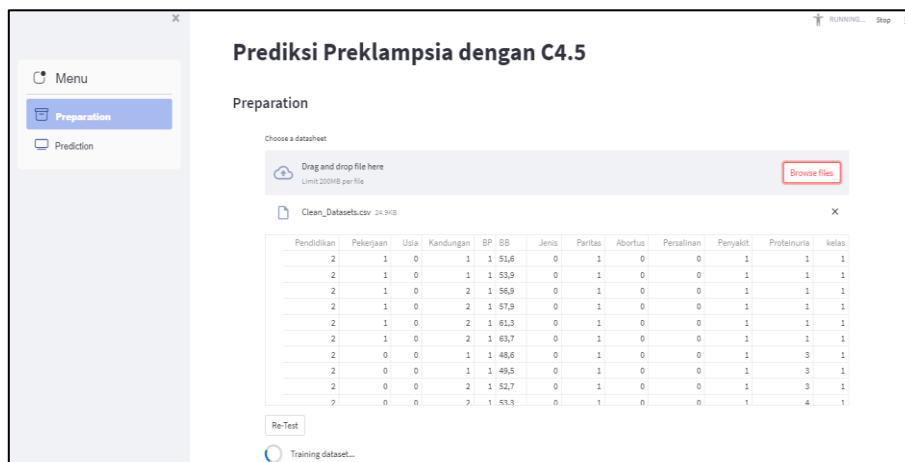
- Menu Preparation

Menu *preparation* adalah menu yang muncul pertama kali saat aplikasi diluncurkan. Menu ini berfungsi untuk melakukan pengujian validitas data rekam medis, seperti ditunjukkan pada Gambar 10.



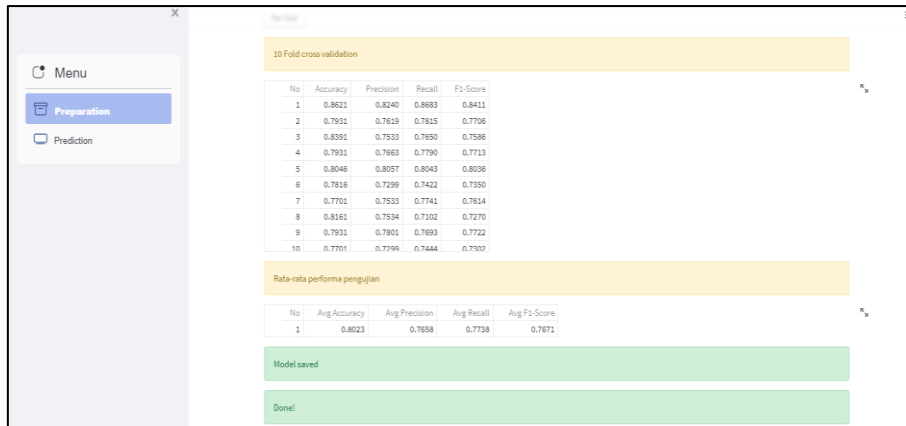
Gambar 10. Menu Preparation Upload File

Gambar 11 menunjukkan implementasi menu *preparation* berupa langkah awal yang harus dilakukan user yaitu mengunggah *file* data rekam medis yang harus berekstensi *csv* dan kapasitas *file* tidak melebihi 200 MB.



Gambar 11. Menu Preparation Hasil Upload File

Setelah *file* berhasil diunggah maka data akan ditampilkan dalam bentuk tabel berjumlah 10 data pasien yang telah ditransformasi. Kemudian sistem melakukan proses *training dataset*, seperti ditunjukkan pada Gambar 12.



The screenshot shows a web application interface with a sidebar menu containing 'Menu', 'Preparation', and 'Prediction'. The main content area is titled '10 Fold cross validation' and displays a table with 10 rows of performance metrics. Below this table is a section titled 'Rata-rata performa pengujian' (Average performance of testing) with a single row of average metrics. At the bottom, there are two green buttons labeled 'Model saved' and 'Done!'.

No	Accuracy	Precision	Recall	F1-Score
1	0.8021	0.8240	0.8583	0.8411
2	0.7931	0.7619	0.7815	0.7796
3	0.8091	0.7533	0.7650	0.7586
4	0.7931	0.7663	0.7790	0.7713
5	0.8046	0.8057	0.8543	0.8536
6	0.7816	0.7299	0.7422	0.7350
7	0.7701	0.7533	0.7741	0.7614
8	0.8161	0.7534	0.7102	0.7270
9	0.7931	0.7801	0.7693	0.7722
10	0.7701	0.7244	0.7444	0.7302

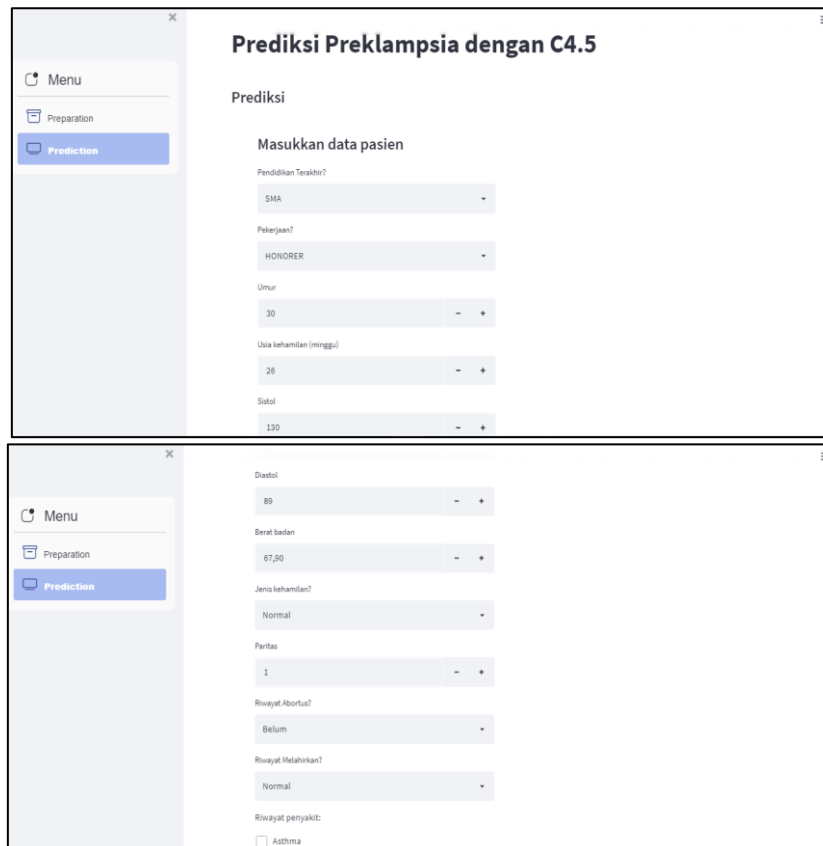
No	Avg Accuracy	Avg Precision	Avg Recall	Avg F1-Score
1	0.8023	0.7658	0.7738	0.7871

**Gambar 12.** Menu *Preparation* Hasil Pengujian Data Rekam Medis

Hasil pengujian berupa nilai akurasi *10-fold cross validation* dan *confusion matrix* berupa *precision*, *recall*, serta *f1-score*. Dan terdapat tabel hasil rata-rata dari keempat hasil tersebut. Jika pengguna mengklik tombol *Re-Test* maka hasil pengujian dan nilai rata-ratanya akan berubah sesuai dengan data yang digunakan.

- Menu *Prediction*

Halaman *prediction* digunakan untuk menerima inputan baru dari user kemudian memprediksi preklampsia dari inputan data baru tersebut. Gambar 13 menunjukkan hasil implementasi halaman *prediction*.



The screenshot shows a web application interface with a sidebar menu containing 'Menu', 'Preparation', and 'Prediction'. The main content area is titled 'Prediksi Preklampsia dengan C4.5' and contains a form for entering patient data. The form includes several input fields with labels and values, and a checkbox for 'Riwayat penyakit: Asthma'.

**Prediksi**

Masukkan data pasien

Pendidikan Terakhir?  
SMA

Pekerjaan?  
HONORER

Umur  
30

Usia kehamilan (minggal)  
26

Sistol  
130

Diastol  
80

Berat badan  
67,90

Jenis kehamilan?  
Normal

Paritas  
1

Riwayat Abortus?  
Belum

Riwayat Melahirkan?  
Normal

Riwayat penyakit:  
 Asthma

**Gambar 13.** Menu *Prediction*

Pada menu *prediction*, pengguna dapat menginputkan data baru yang akan langsung diprediksi dan ditampilkan hasilnya oleh sistem.

### 3.4. Pengujian Sistem

- Confusion Matrix

Pengujian dataset rekam medis preklampsia dilakukan menggunakan *cross validation* dengan nilai *k-fold* = 10. Dataset di uji sebanyak 10 kali dengan perbandingan rasio split data 9:1

Tabel 5. Confusion Matrix

Iterasi Ke	Akurasi	Precision	Recall	F1-Score
1	0.8851	0.8379	0.8155	0.8257
2	0.8046	0.7955	0.7955	0.7955
3	0.8276	0.7716	0.7717	0.7713
4	0.8276	0.8104	0.8279	0.8175
5	0.7241	0.6732	0.7128	0.6870
6	0.8046	0.7683	0.8262	0.7869
7	0.7931	0.7665	0.7659	0.7661
8	0.8161	0.7725	0.7723	0.7723
9	0.8736	0.8510	0.9070	0.8696
10	0.7816	0.7900	0.7738	0.7814
<b>Rata-rata</b>	<b>0.8138</b>	<b>0.7837</b>	<b>0.7969</b>	<b>0.7873</b>

Tabel 5 menunjukkan hasil pengujian sistem *confusion matrix* menggunakan *10-cross validation*. Nilai akurasi menjadi hasil tertinggi dengan hasil validitas lainnya yaitu sebesar 0.8138 atau 81.38%.

- Black Box Testing

Hasil pengujian sistem prediksi dini preklampsia menggunakan algoritma C4.5 dengan *black box testing* adalah seluruh program dan sistem sudah sesuai dengan rancangan dan berjalan tanpa adanya *bug* maupun *error*.

### 4. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dapat diketahui bahwa pengembangan sistem prediksi dini preklampsia dapat dihasilkan dengan menggunakan algoritma C4.5. Dari pengujian *confusion matrix* menggunakan *10-fold cross validation*, diperoleh hasil pengujian sistem meliputi akurasi 81,38%, presisi 78,37%, *recall* 79,69%, dan *f1-score* 78,73%. Sistem prediksi dini preklampsia menggunakan algoritma C4.5 dapat dinyatakan layak untuk digunakan karena mendapat hasil pengujian *black box* sebesar 100%.

### Daftar Pustaka

- [1] Kemenkes, "Pusat Data dan Informasi Kementerian Kesehatan RI," 2021.
- [2] N. Muhani and Besral, "Pre-eklampsia Berat dan Kematian Ibu," *Jurnal Kesehatan Masyarakat Nasional*, pp. Vol. 10, No. 2, 2015.
- [3] V. B. Kusnandar, "10 Provinsi dengan Angka Kematian Ibu Terbanyak pada 2020," 22 12 2021.
- [4] R. N. Reny, "PENERAPAN METODE CHAID ( CHI-SQUARED AUTOMATIC INTERACTION DETECTION ) DAN CART ( CLASSIFICATION AND REGRESSION TREES )," pp. 35-73, 2018.
- [5] A. Muzakir and W. R. Anisa, "Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Tree Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan teknik Decision Tree," 2016.
- [6] P. Kiondo, "Risk Factors for Pre-eclampsia in Mulago Hospital, Kampala, Uganda," *Tropical Medicine and International Health*, pp. 480-487, 2012.

- 
- [7] M. Klebanoff, P. Shiono and G. Rhoads, "Outcomes of Pregnancy in National Sample of Resident Physicians," *N Engl J Med*, p. 323, 1990.
- [8] T. Asrianti, "Faktor Risiko Kejadian Preeklampsia pada Ibu Melahirkan di RSIA Siti Fatimah Makasar," *Skripsi Universitas Hasanudin Makasar*, 2009.
- [9] N. Istifadah, Mussia and N. Riska Rahmawati, "GAMBARAN FAKTOR PENYEBAB PRE EKLAMPSIA/ EKLAMPSIA PADA IBU HAMIL DI PUSKESMAS KALISAT KABUPATEN JEMBER," *JURNAL KESEHATAN dr. SOEBANDI*, pp. 176-183, 2014.
- [10] S. Sukaesih, "Faktor-Faktor yang Berhubungan dengan Pengetahuan Ibu Hamil Mengenai Tanda Bahaya dalam Kehamilan di Puskesmas Tegal Selatan Kota Tegal Tahun 2012," *Skripsi Program Sarja Kesehatan Masyarakat Universitas Indonesia*, 2000.
- [11] P. Pitria, "ANALISIS SENTIMEN PENGGUNA TWITTER PADA AKUN RESMI SAMSUNG INDONESIA DENGAN MENGGUNAKAN NAIVE BAYES," *Sarjana Institut*, 2014.
- [12] M. Ali, D.-H. Son, K. Sang-Hee and S.-R. Nam, "An Accurate CT Saturation Classification Using a Deep Learning Approach Based on Unsupervised Feature Extraction and Supervised Fine-Tuning Strategy," *Energies*, 2017.