Diagnosis Preeklamsia pada Ibu Hamil Berdasarkan Algoritme K- Nearest Neighbour

[1]Rifki Hidayat, [2]Tri Astuti, S.Kom., M.Eng
[1,2]Progam Studi Sistem Informasi, Fakultas Ilmu Komputer / Universitas Amikom purwokerto
[1,2] Jl. Letjen Pol Sumarto, Purwanegara, Purwokerto Utara, Banyumas, Jawa Tengah, Indonesia 53123, Telp, (0281) 623321, Faks. (0281) 623196, Web : fik.amikompurwokerto.ac.id, e-mail : fik@amikompurwokerto.ac.id
email : [1]ripki09.id@gmail.com, [2]tri_astuti@amikompurwokerto.ac.id

ABSTRACT
Maternal deaths are divided into two namely direct and indirect deaths. Globally 80% of direct maternal deaths, preeclampsia are included in direct maternal deaths. Preeclampsia conditions of pregnancy with hypertension occur after the 20th week in women who previously had normal blood pressure. Preeclampsia can also be characterized by hypertension (systolic blood pressure ≥ 140 mmHg or diastolic blood pressure ≥ 90 mmHg) accompanied by proteinuria (≥ 300 mg / dl in tamping urine 24 hours). In this study, an analysis of medical records in the Purbalingga and Banyumas areas using 8 attributes, namely age, body weight, blood pressure, edema, multiple pregnancy, history of hypertension, how many children, urine protein, and preeclampsia class. From calculations using the K-NN (K-Nearest Neighbour) algorithm, the Sensitivity performance value of 98.19%, Specificity 100%, and Accuracy 98.33%.

Keywords **-** Confusion Matrix, Data Mining, KNN Algorithm, Preeclampsia

ABSTRAK
Kematian ibu dibagi menjadi dua yaitu kematian langsung dan tidak langsung. Secara global 80 % kematian ibu secara langsung, preeklamsia merupakan termasuk dalam kematian ibu secara langsung. Preeklamsia kondisi kehamilan dengan hipertensi terjadi setelah minggu ke-20 pada wanita yang sebelumnya memiliki tekanan darah normal. Preeklamsia juga dapat ditandai dengan hipertensi (tekanan darah sistol ≥ 140 mmHg atau tekanan darah diastol ≥ 90 mmHg) disertai proteinuria (≥ 300 mg/dl dalam urin tamping 24 jam). Pada penelitian ini melakukan analisis data rekam medis wilayah Purbalingga dan Banyumas dengan menggunakan 8 atribut yaitu umur, BB (Berat Badan), TD (Tekanan Darah), Edema, Kehamilan kembar, riwayat hipertensi, anak ke berapa, protein urine, dan class preeklamsia. Dari perhitungan menggunakan metode Algoritme K-NN (K-Nearest Neighbour) menghasilkan nilai performa Sensitivity 98,19 %, Specificity 100 %, dan Accuracy 98,33 %.

Kata Kunci **–** Algoritma KNN, *Confusion Matrix*, *Data Mining*, *Preeklamsia*

1. Introduction

Masalah kesehatan tentang ibu hamil adalah hal yang harus diperhatikan. Karena pada saat ibu hamil banyak merenggut nyawa. Kematian ibu hamil dibagi menjadi kematian langsung yaitu akibat komplikasi kehamilan, persalinan atau masa nifas dan segala intervensi atau penanganan tidak tepat dari komplikasi tersebut, dan tidak langsung merupakan akibat dari penyakit yang sudah ada dan atau penyakit yang timbul sewaktu kehamilan yang berpengaruh terhadap kehamilan. Secara global 80% kematian ibu tergolong pada kematian ibu langsung [6]. Yang dimana penyakit preeklamsia masuk ke dalam kematian ibu tidak langsung.

Preeklampsia merupakan suatu kondisi spesifik kehamilan dengan hipertensi terjadi setelah minggu ke-20 pada wanita yang sebelumnya memiliki tekanan darah normal. Preeklamsia merupakan suatu penyakit vasospastik, yang melibatkan banyak sistem dan ditandai oleh hemokonsentrasi, hipertensi,

dan proteinuria. Diagnosis preeklamsia secara secara tradisional didasarkan pada adanya hipertensi yang disertai proteinuria atau edema [3]. Preeklampsia ditandai dengan hipertensi (tekanan darah sistol ≥ 140 mmHg atau tekanan darah diastol ≥ 90 mmHg) disertai proteinuria (≥ 300 mg/dl dalam urin tampung24 jam) pada usia kehamilan lebih dari 20 minggu atau segera setelah persalinan. Sedangkan preeklampsia ditandai dengan hipertensi, proteinuriadan kejang [8].

Data mengenai penyakit preeklampsia di Kabupaten Banyumas tahun 2011 sebanyak 551 orang (32,1%) dari seluruh ibu hamil sejumlah 1714 orang. Sedangkan kejadiaan preeklampsia tahun 2012 sebanyak 930 orang (50,9%) dari seluruh ibu hamil sejumlah 1826 orang. Dilaporkan jumlah kematian ibu pada tahun 2013 sebanyak 33 kasus, dengan 9 kasus disebabkan karena eclampsia menurut [12].

Data mining merupakan serangkaian proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar [Han].Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data mining adalah proses yang menggunakan teknik statistic matematika, kecerdasan buatan, dan mechine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [13].

Berbagai tools komersial maupun non-komersial beredar dan digunakan untuk berbagai tujuan yang berkaitan dengan data mining. Salah satunya adalah WEKA, yang merupakan alat bantu dalam data mining, terutama dalam penerapan untuk menyelesaikan masalah klasifikasi [2].

Penelitian yang dilakukan adalah analisis data mining pada data rekam medis wilayah Banyumas dan Purbalingga menggunakan algoritma KNN (K-*Nearest Neighbor*) untuk mendiagnosis penyakit preeklamsia. Kenapa menggunakan algoritma karena KNN (K-*Nearest Neighbor*) merupakan salah satu dari banyak metode klasifikasi teks yang populer yang memberikan hasil yang akurat dan mudah dimengerti [1]. Selain itu teknik KNN (K-*Nearest Neighbor*) merupakan model klasifikasi yang memiliki beberapa kelebihan, penerapannya yang sederhana namun efektif dalam banyak kasus [9].

Tujuan dari penelitian ini mencari performa data rekam medis wilayah Banyumas dan Purbalingga sebanyak 120 data yang teridiri dari 11 data yang terkana penyakit preeklamsia dan 109 data yang tidak terkena penyakit preeklamsia. Performa data terdiri dari tiga yaitu sensitivity, specificity, dan accuracy dari perhitungan confusion matrix. Untuk mendapatkan nilai dari confusion matrix peneliti menggunakan software WEKA.

2. Research Method

Metode penelitian langkah pertama yang harus dilakukan melakukan identifikasi masalah dan selanjutnya adalah metode pengumpulan data yang digunakan sebagai berikut :

1. Studi Pustaka

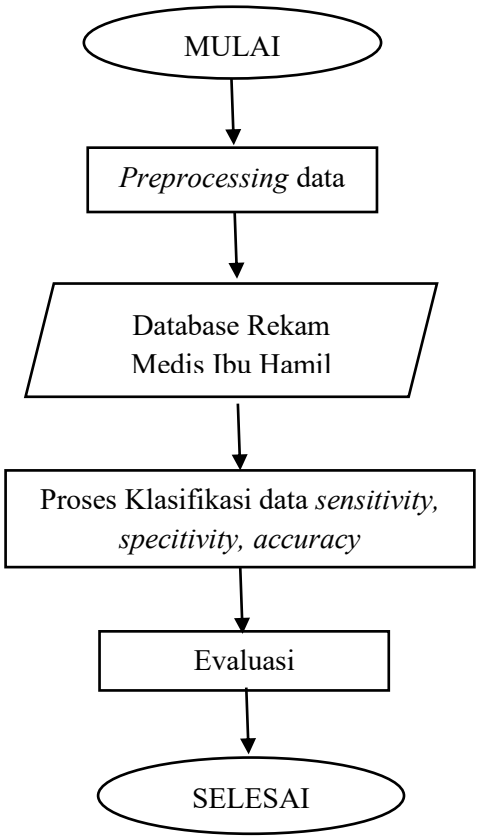
   Studi pustaka berkaitan dengan kajian teoritis dan referensi lain yang berkaitan dengan nilai, budaya dan norma yang berkembang pada situasi sosial yang diteliti, selain itu studi kepustakaan sangat penting dalam melakukan penelitian, hal ini dikarenakan penelitian tidak akan lepas dari literatur-literatur Ilmiah menurut [11]

2. Data Primer

   Pada penelitian ini menggunakan sebuah data primer yang merupakan sumber data yang diperoleh peneliti secara langsung yaitu melalui studi pustaka terhadap data yang akan diperlukan diwilayah Purbalingga dan Banyumas. Data tersebut berupa data rekam medis ibu hamil yang kemudian di salin ke dalam file excel. Dari proses pengumpulan data diperoleh sebanyak 120 data. Data yang diperoleh terdiri dari umur, berat badan <45kg, tekanan darah≥160/90 mmHg, edema, protein urine, kehamilan ke berapa, persalinan ke berapa, riwayat hipertensi, *preeklampsia*.

Dalam penelitian ini, peneliti akan melakukan sebuah konsep penelitian dalam bentuk bagan alur pada penelitian ini. Pada penelitian bahan penelitian menggunakan data rekam medis ibu hamil wilayah

purbalingga dan banyumas. Pada tahapan ini dilakukan setelah proses identifikasi masalah dan pengumpulan data, yang selanjutnya adalah tahapan-tahapan untuk memperoleh output dari metode Algoritme KNN (K-*Nearest Neighbor*) dalam melakukan diagnosis penyakit *preeklampsia*. Konsep penelitian akan dijelaskan pada gambar 1 sebagai berikut :



Gambar 1. Konsep Penelitian

Dari gambar 1 konsep penelitian yaitu tentang langkah – langkah peneliti dalam penelitian antara lain :

1. *Preprocessing* Data

   Awal konsep penelitian yang akan dilakukan pada penelitian ini adalah tahap preprocessing data pada dataset rekam medis ibu hamil. Tahapan preprocessing ini akan digunakan untuk pembangkitan aturan-aturan berdasarkan Algoritme KNN (K-*Nearest Neighbor*), dan selanjutnya dilakukan klasifikasi data dan evaluasi terhadap hasil yang diperoleh dari pengujian data.

   Preprocessing data merupakan tahapan yang sangat penting karena pada tahap ini dilakukan proses seleksi data sehingga data tetap dapat dikontrol dan tidak keluar dari jalur. Dengan demikian data akan siap digunakan sebagai bahan penelitian[9].

   Pada tahapan preprocessing hal yang pertama persiapkan data rekam medis ibu hamil wilayah purbalingga dan banyumas yang terdiri dari 120 data. Langkah awal adalah pemilihan atribut yang digunakan untuk perhitungan dan perubahan format data ke agar memudahkan dalam perhitungan pada software weka. Setelah melewati tahapan berikut maka data rekam medis wilayah purbalingga dan banyumas sudah diolah dan bisa masuk ke tahap selanjutnya.

2. Klasifikasi

   Klasifikasi Algoritme dapat menentukan prediksi pada potensi data [10]. Hasil klasifikasi akan dibandingkan dengan dataset yang sebenarnya menggunakan nilai confusion matrix antara hasil pengujian metode KNN (K-*Nearest Neighbor*) menggunakan software Weka.

Nilai confusion matrix yang akan digunakandalam penelitian ini adalah *accuracy, specificity*, dan *sensitivity*.

Confusion Matrix adalah evaluasi dari sebuah klasifikasi data mining yang direpresentasikan menjadi tabel [4]. Convussion matrix berisi informasi perbandingan label hasil klasifikasi dengan label sebenarnya. Tabel 2 menggambarkan confussion matrix dengan dua label yaitu yes dan no.

Tabel 1. Confusion Matrix (Gorunescu 2011)

Classification		Predicted class	
		Class : YES	Class : NO
Observed class	Class YES	True Positive (TP)	False Negative (FN)
	Class NO	False Positive (FP)	True Negative (TN)

Pengelompokan data pada confusion matrix ini digunakan untuk menghitung nilai *sensitivity, spesificity, recall, precision, f-measure* dan *accuracy* dapat dihitung dengan menggunakan persamaan berikut [15] :

Persamaan 2.1 :

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

Persamaan 2.2 :

$$Sensitivity = \frac{TN}{TN+FP}$$

Persamaan 2.3 :

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

Persamaan 2.4 :

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

Persamaan 2.5 :

$$F\text{-}Measure = \frac{2xPrecisionxRecal}{Precision + Recl}$$

Persamaan 2.6 :

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

Keterangan :

TP : Jumlah kasus positif yang diklasifikasi sebagai positif

FP : Jumlah kasus negatif yang diklasifikasi sebagai positif

TN : Jumlah kasus negatif yang diklasifikasi sebagai negative

FN : Jumlah kasus positif yang diklasifikasi sebagai negative

3. Evaluasi

Pada tahapan ini akan dilakukan sebuah evaluasi klasifikasi didasarkan pada pengujian pada obyek yes dan no [4] serta pengukuran sebuah keakuratan hasil yang diperoleh dengan menggunakan metode KNN (K-*Nearest Neighbor*). Pada penelitian ini akan dinilai dari nilai yang diperoleh dari confusion matrix yang akan dihasilkan menjadi sebuah parameter. Pada tahapan ini akan dievaluasi mengenai nilai *accuracy, specificity*, dan *sensitivity* yang digunakan untuk mengetahui nilai performa dari pengujian menggunakan metode KNN (K-*Nearest Neighbor*).

3. Result and Analysis

3.1 Identifikasi Masalah

Dalam tahapan ini penulis melakukan identifikasi pada penyakit preeklampsia yang merupakan sindrom kehamilan yang diderita oleh ibu hamil. Untuk mengetahui penyebab preeklampsia, penulis mempelajari literatur yang berkaitan dengan penelitian penulis yaitu untuk mendiagnosis penyakit preeklampsia. Untuk melakukan pendiagnosisan penulis memilih Algoritme K-NN (K-Nearest Neighbour) yang sesuai untuk penelitian penulis, dimana Algoritme K-NN (K-Nearest Neighbour) dapat digunakan untuk menganalisis dataset.

3.2 Pengumpulan Data

Bagian Dalam penelitian ini data yang akan digunakan yaitu dengan mengambil data ibu hamil yang ada diwilayah Purbalingga dan Banyumas, dengan jumlah 120 data dan menggunakan 9 atribut sebagai pendukungnya. Dimana atribut-atribut dan penjelasannya yang dimaksud adalah sebagai berikut :

a. Umur : variabel input ini menunjukan usia/umur pasien.
b. BB Ibu < 45 (Berat Badan Ibu <45kg) : variabel input ini menunjukan berat badan ibu hamil yang mengalami resiko terkena berada di kisaran kurang dari 45 kg.
c. TD ≥160/90 mmHg (Tekanan Darah ≥160/90 mmHg) : variabel input ini menunjukan tekanan darah pasien yang melebihi 160/90mmHg.
d. Edema yang nyata : variabel input ini menunjukan istilah yang digunakan untuk merujuk pada kondisi bengkak pada jaringan lunak seperti kulit pada ibu hamil.
e. Kehamilan kembar : variabel input ini menunjukan kondisi ibu hamil mempunyai janin lebih dari satu.
f. Riwayat hipertensi : variabel input ini menunjukan apakah pasien mempunyai riwayat hipertensi atau tekanan darah yang tinggi
g. Anak keberapa : variabel input ini menunjukan kehamilan anak keberapa pada pasien ibu hamil.
h. PU (Protein Urine) : variabel input ini menunjukan jika ibu hamil yang mengalami preotein urine yang tidak normal akan bernilai positive (Y) dan ibu hamil yang tidak mengalami protein urine tidak normal akan bernilai negative (T).
i. Class Variabel *Preeklampsia* : variabel ini adalah output terdiri dari dua kemungkinan yaitu yes dan no. Y berarti bahwa satu variabel record dari 8 variabel input akan menghasilkan instance pada variabel ke 9 (Y) bahwa pasien menderita penyakit *preeklampsia*. T mengandung arti bahwa dalam satu record dari 8 variabel input akan menghasilkan instance pada variabel ke 9 (T) bahwa pasien tidak menderita penyakit *preeklampsia*. Class Variabel ini merupakan tujuan class untuk menyatakan diagnosa penyakit *preeklampsia*.

Tabel 2. Data Rekam Medis Ibu Hamil Wilayah Purbalingga dan Banyumas.

NO	UMUR	BB<45 Kg	TD>=160/90 mmHg	Odema yang nyata	Kehamilan kembar	Riwayat hipertensi	anak ke berapa	Protein Urine	Preeklamsi
1	24	T	T	T	?	?	0	T	T

2	36	T	T	T	?	?	3	T	T
3	20	T	T	T	?	?	0	T	T
4	30	Y	T	T	?	?	1	T	T
5	34	T	T	T	?	?	1	T	T
6	35	T	T	T	?	?	0	T	T
7	28	T	T	T	?	?	1	T	T
8	29	T	T	T	?	?	1	T	T
9	26	Y	T	T	?	?	1	T	T
10	35	T	T	T	?	?	1	T	T
...	...	...	...	...	...	...	...	...	...
118	38	T	T	T	?	?	?	T	Y
119	31	T	T	T	?	?	?	Y	Y
120	33	T	T	T	?	?	?	Y	Y

Pada tabel 2. merupakan tabel data rekam medis ibu hamil dengan jumlah 120 data dan menggunakan 9 atribut sebelum diolah, penjalasan untuk masing-masing atribut yaitu atribut umur memiliki nilai terendah 17 dan tertinggi 42, atribut BB (Berat Badan) <45kg memiliki kategori Y dan T, atribut TD (Tekanan Darah) ≥160/90 mmHg memiliki kategori Y danT, atribut Edema yang nyata memiliki kategori Y dan T,atribut kehamilan kembar tidak memiliki kategori, atribut riwayat hipertensi tidak memiliki kategori, atribut anak ke berapa memiliki nilai terendah 0 dan nilai tertinggi 4, atribut Protein Urine memiliki Y dan T, atribut preeklampsia memiliki kategori instance Y dan T. Atribut preeklamsia adalah sebagai acuan klasifikasi pada penelitian ini.

Tabel 3. Penjelasan Data Rekam Medis Ibu Hamil

Data ke 1	Keterangan	Data ke 120	Keterangan
24	Umur	33	Umur
T	BB < 45kg	T	BB < 45kg
T	TD ≥ 160/90mmHg	T	TD ≥ 160/90mmHg
T	Odema	T	Odema
?	Kehamilan Kembar	?	Kehamilan Kembar
?	Riwayat Hipertensi	?	Riwayat Hipertensi
0	Anak keberapa	?	Anak keberapa
T	Protein Urine	Y	Protein Urine
T	Preeklamsia	Y	Preeklamsia

Pada tabel 3 merupakan penjelasan dua *record* dari dataset Rekam Medis Ibu Hamil yaitu data ke 1 dan 120 beserta keterangan masing-masing nilai pada atribut dan *record* berdasarkan dataset. Pada data ke 1 berisi umur : 24, BB<45kg : T, TD ≥ 160/90mmHg : T, Odema : T, Kehamilan Kembar : ?, Riwayat Hipertensi : ?, Anak keberapa : 0, Protein Urine : T, *Preeklamsia* : T dan data ke 120 berisi umur : 33, BB<45kg : T, TD ≥ 160/90mmHg : T, Odema : T, Kehamilan Kembar : ?, Riwayat Hipertensi : ?, Anak keberapa : 0, Protein Urine : Y, *Preeklamsia* : Y. Untuk penggunaan Y adalah ya, T adalah tidak, dan ? adalah data kosong.

3.3 *Preprocessing* Data

Pada tahapan *preprocessing* data harus mempersiapkan data yaitu dengan mengkonversi data rekam medis ibu hamil dari bentuk Microsoft Excel (XLSX) menjadi

CSV (Comma-Separated Values) dengan cara melakukan save as dari data set yang masih berformat xlsx dirubah menjadi csv. Sampel data rekam medis ibu hamil dalam bentuk CSV dapat dilihat pada gambar 2 dibawah ini.

```
UMUR,BB,TD,odema,Kehamilan_kembar,hipertensi,anak_ke,Protei
n_Urine,Preeklamsia
24,T,T,T,?,?,0,T,T
36,T,T,T,?,?,3,T,T
20,T,T,T,?,?,0,T,T
30,Y,T,T,?,?,1,T,T
34,T,T,T,?,?,1,T,T
Dst……
```

Gambar 2. Sampel Data Rekam Medis dalam bentuk CSV.

Pada Gambar 2 diatas yang menampilkan data rekam medis dalam bentuk CSV (*Comma Separated Values*) yang dimana file ini hanya berupa nama atribut dan data dari tabel tidak berbentuk tabel melainkan hanya berupa tulisan nama atribut dan data yang dipisahkan oleh tanda koma (,) dan tanpa spasi.

Penjelasan untuk bagian atas dari gambar 2 yang bertulisan : UMUR, BB, TD , odema, Kehamilan_kembar, hipertensi, anak_ke, Protein_Urine, *Preeklamsia* itu adalah nama atribut tabel. Sedangkan yang bagian bawahnya yang bertulisan : 24,T,T,T,?,?,0,T,T dan sampai data 120 adalah isi data dari setiap atribut. Cara membaca menggunkana data pertama 24,T,T,T,?,?,0,T,T dibacanya adalah umur (24), berat badan (tidak), tekanan darah (tidak), odema (tidak), kehamilan_kembar (?), hipertensi (?), anak_ke (0), protein_urine (tidak), preeklamsia (tidak).

Setelah merubah formatnya menjadi csv perlu ada lagi berupahan format data agar bisa diolah pada *software* WEKA. Pergantian format data bisa dilakukan dengan menggunakan *software* Notepad++ dari file csv menjadi file *arff* degan cara melakukan save as. Sebelum dirubah format datanya perlu ada tambahan *script* pada file datanya, pertama adalah *@relation* berfungsi untuk penamaan file, kedua *@attribute* berfungsi untuk menjelaskan atribut yang digunakan dan juga deklarasi tipe data attribute yang digunakan harus sesuai missal datanya angka maka tipe datanya adalah numeric, ketiga *@data* berfungsi untuk mendklarasikan data secara keseluruhan sesuai dengan atribut yang digunakan. Pada gambar 3 adalah *script* yang ada pada format *arff* dalam editor *software* Notepad++.

```
@relation data

@attribute UMUR numeric
@attribute BB {Y,T}
@attribute TD {Y,T}
@attribute odema {Y,T}
@attribute Kehamilan_kembar {Y,T}
@attribute hipertensi {Y,T}
@attribute anak_ke numeric
@attribute Protein_Urine {Y,T}
@attribute Preekalmsia {Y,T}

@data
24,T,T,T,?,?,0,T,T
36,T,T,T,?,?,3,T,T
20,T,T,T,?,?,0,T,T
…….……
31,T,T,T,?,?,?,Y,Y
33,T,T,T,?,?,?,Y,Y
```

Gambar 3. Sampel Data Rekam Medis dalam bentuk ARFF.

Pada gambar 3 merupakan data rekam medis yang sudah dalam bentuk ARFF (*Attribute-Relation File Format*). Pada bentuk *arff* ini terdapat 3 bagian yaitu @relation, @attribute dan @data yang sudah dijelaskan diatas. Pada bagian pertama pada gambar 3 yaitu @relation data adalah ini adalah file yang bernama data. Bagian kedua @attribute umur numeric sampai @attribute preeklamsia {Y,T} merupakan nama setiap atribut dan penggunaan tipe data pada setiap atribut di jabarkan. Pada gambar 3 menggunakan jenis 2 tipe data tipe data numeric dan tipe data nominal spesification. Tipe data numeric merupakan data berupa angka seperti @attribute UMUR numeric, dan @attribute anak_ke numeric. Tipe data nominal spesification merupakan tipe data yang data berdasarkan data yang sudah yang spesifik seperti {Y,T} yang harus diberi kurung kurawal dan koma sebagai pemisah spesifik data tanpa spasi. Penerapannya seperti @attribute BB {Y,T}, @attribute TD {Y,T}, @attribute odema {Y,T}, @attribute Kehamilan_kembar {Y,T}, @attribute hipertensi {Y,T}, @attribute Protein_Urine {Y,T}, dan @attribute Preekalmsia {Y,T} dari penerapan tersebut atribut mempunyai spesifikasi data Y dan T saja. Bagian ketiga @data merupakan data setiap atribut yang sudah dijabarkan pada bagian @attribute. Pada penulisannya pemisah menggunakan koma dan tanpa spasi. Penerepan seperti : @data 24,T,T,T,?,?,0,T,T sampai 33,T,T,T,?,?,?,Y,Y cara membaca datanya masih sama seperti pada bentuk *csv* yang berbeda adalah ada penamaan file yaitu @relation dan penggunaan tipe data pada setiap atribut dijabarkan.

3.4 Klasifikasi Data

Peniliti melakukan pengujian data pada software WEKA sesuai dengan penelitian mengunakan Algoritme K-NN (K-Nearest Neighbour). Pada pengujian ini akan diperoleh nilai TP (true positive), TN (true negative), FP (flase positive), dan FN (false negative). Dari pengujian tersebut akan dicari nilai confusion matrix meliputi nilai accuracy, sensitivity, dan specificity yang dihasilkan dari pengujian menggunakan metode K-NN (K-Nearest Neighbour).

Cara melakukan pengujian data pada software WEKA Algoritme, pertama masuk kedalam tab Classify karena Algoritme K-NN (K-Nearest Neighbour) merupakan kelompok klasifikasi. Langkah selanjutnya adalah masuk kedalam classifier klik tombol choose pilih folder lazy masuk didalam folder terdapat 3 file pilih IBk. Lanjut ke test options pilih cross-validation dengan folds 10. Berikutnya pastikan atribut klasifikasi adalah atribut preeklamsia karena atribut ini merupakan klasifikasi dari data ini, selanjutnya start maka akan muncul hasilnya.

Maka pengujian tadi mendapatkan hasil yaitu nilai dari confusion matrik berdasarkan Algoritme K-NN (K-Nearest Neighbour). Hasil dari pengujian ada pada gambar 4 di bawah ini.

```
=== Confusion Matrix ===

   a   b   <-- classified as
   9   2 |   a = Y
   0 109 |   b = T
```

Gambar 4. Hasil Confusion Matrix dari software WEKA.

Pada Gambar 4.4 merupakan hasil dari perhitungan dari *software* weka dengan metode algoritma K-NN (*K*-Nearest *Neighbour*). Dari hasil tersebut menghasilkan confusion matrix pada nilai TP (*true positive*) sebesar 109 berdasarkan kolom b baris kedua, TN (*true negative*) sebesar 9 berdasarkan kolom a baris pertama, FP (*flase positive*) sebesar 0 berdasarkan kolom a baris kedua, dan FN (*false negative*) sebesar 2 berdasarkan kolom b baris pertama. Agar lebih jelas baca pada tabel 4.

Tabel 4. Hasil confusion matrix metode K-NN

109 *true positive*	2 *false negative*
0 *flase positive*	9 *true negative*

Dari tabel 4 diatas merupakan hasil confusion matrix dari *software* weka dengan metode K-NN (*K-Nearest Neighbour*). Dari tabel 4.3 mempunyai hasil confusion matrik diantranya : nilai TP (*true positive*) sebesar 109, TN (*true negative*) sebesar 9, FP (*flase positive*) sebesar 0, dan FN (*false negative*) sebesar 2. Dan langkah selanjutnya adalah melakukan perhitungan performa seperti mencari nilai *sensitivity, specifity,* dan *accuracy.* Berikut adalah cara perhitungannya :

a. $Sensitivity = \frac{TP}{P} \times 100\%$
$= \frac{109}{111} \times 100\%$
$= 98{,}19\ \%$

b. $Specificity = \frac{TN}{N} \times 100\%$
$= \frac{9}{9} \times 100\%$
$= 100\%$

c. $Accuracy = \frac{TP+TN}{P+N} \times 100\%$
$= \frac{109+9}{111+9} \times 100\%$
$= \frac{118}{120} \times 100\% = 98{,}33\ \%$

Dari hasil perhitungan confusion matrix diatas, maka dapat disimpulkan bahwa pengujian metode K-NN (*K-Nearest Neighbour*) memiliki performa sebagai berikut :

Tabel 5. Hasil performa dari metode K-NN

Performa	Nilai
Sensitivity	98,19 %
Specificity	100 %
Accuracy	98,33%

Pada tabel 5 mengenai hasil performa yang diperoleh menggunakan metode K-NN (*K-Nearest Neighbour*), maka dapat disimpulkan sebagai berikut untuk nilai *sensitivity* sebesar 98,19%, nilai *specificity* sebesar 100%, dan nilai *accuracy* sebesar 98,33%. Hasil tersebut diujikan berdasarkan data rekam medis ibu hamil wilayah purbalingga dan banyumas.

3.5 Evaluasi

Hasil dari penelitian berjudul Diagnosis Preeklamsia Pada Ibu Hamil Berdasarkan Algoritme K- Nearest Neighbour. Pada penelitian hal yang dilakukan adalah melakukan pengolah data rekam medis ibu hamil wilayah purbalingga dan banyumas dengan banyak data 120 data. Untuk pengolahan data rekam medis ibu hamil menggunakan 9 atribut yaitu umur, bb (berat badan), td (tekanan darah), odema, kehamilan_kembar, hipertensi, anak_ke, protein_urine, dan *preeklamsia*. Atribut sebagai class klasifikasi adalah atribut terkhir yaitu atribut preeklamsia karena atribut sudah mempunyai hasil terkait preeklamsia berdasarkan dari atribut yang ada pada data rekama medis ibu hamil. Pengolahan data menggunakan metode data mining berdasarkan algoritme KNN (*K-Nearest Neighbor*) untuk mencari nilai performa dari data rekam medis.

Dalam pencarian nilai peforma data rekam medis harus mencari nilai *accuracy*, *specificity*,dan *sensitivity* berdasarkan nilai confusion matrik yang perhitungannya dilakukan pada *software* weka dengan algoritme KNN (*K-Nearest Neighbor*). Nilai confusion matrik dari perhitungan *software* weka adalah nilai TP (*true positive*) sebesar 109, TN (*true negative*) sebesar 9, FP (*flase positive*) sebesar 0, dan FN (*false negative*) sebesar 2. Melihat dari tahap evaluasi bahwa data rekam medis ibu hamil wilayah purbalingga dan banyumas mempunyai nilai performa data yaitu adalah nilai *accuracy* sebesar 98,33%, *specificity* sebesar 100%, *sensitivity* sebesar 98,19%. Berdasarkan dari nilai performa serta mengacu pada nilai *accuracy* pada data rekam medis ibu hamil wilayah purbalingga dan banyumas dapat melakukan diagnosis preeklamsia dengan algoritme KNN (*K-Nearest Neighbor*).

4. Conclusion

Dapat disimpulkan bahwa nilai performa dari nilai *accuracy*, *specificity*,dan *sensitivity* dari 120 data dan 9 atribut. Hasil performa dari perhitungan data rekam medis yang sudah dioalah yaitu sensitivity sebesar 98,19 %, specificity sebesar 100 %, dan accuracysebesar 98,33 %. Dari hasil tersebut dapat disimpulkan bahwa algoritme KNN (K-Nearest Neighbor) dapat digunakan untuk mendiagnosis penyakit preeklampsiadengan nilai accuracy sebesar 98,33 %. Saran untuk penelitian selanjutnya tambahkan data rekam medis agar dapat mendapat akurasi semakin akurat, gunakan algoritma lain sebagai bahan perbandingan akurasi.

References

[1]. Alhutaish, R. & Omar, N., 2015. Arabic Text Classification using K-Nearest Neighbour Algorithm. The International Arab Journal of Information Technology, 12(2), pp. 190-195.

[2]. Aswendy. (2016). Analisis data iklim indonesia menggunakan aplikasi weka dengan metode klasifiksi.*Jurnal* Teknologi *Rekayasa*, *Volume 21*, 217–228.

[3]. Emha, M. R., Hapsari, E. D., & Lismidiati, W. (2017). Pengalaman hidup ibu dengan riwayat kehamilan preeklamsia di Yogyakarta. Berita Kedokteran Masyarakat, 33(4), 193. https://doi.org/10.22146/bkm.13175.

[4]. Gorunescu, Florin. 2011. Data Mining: Concept, Models and Techniques. Vol 12. Berlin: Heidelberg: Springer Berlin Heidelberg.

[5]. Han, J., &Kamber, M. Data mining Concepts and Techniques Second Edition. San Fransisco: Morgan Kauffman. 2006.

[6]. Prawirohardjo, S. 2011.IlmuKebidanan. Jakata.

[7]. Pyle, D, Data Preparation For Data Mining, Academic Press : United States of America, 1999.

[8]. Robert JM, Pearson GD, Cutler JA, Lindheimer MD. 2003. Summary of the NHLBI Working Group on Research on Hypertension During Pregnancy. Hypertension in Pregnancy 22(2): 109-127.

[9]. SINHA, P. SINHA; P., 2015. Comparative Study of Chronic Kidney Disease Prediction using KNN and SVM. [daring] 4(12), hal.608–612.

[10]. Sitorus, 2010, Penggunaan Data Mining Dengan Metode Decision Tree Untuk Prediksi Resiko Kredit, skripsi. FMIPA UGM, Yogyakarta.

[11]. Sugiyono. 2012. Metode Penelitian Kuantitatif Kualitatif dan R&D. Bandung : Alfabeta.

[12]. Suyandari AE, Trisnawati Y. 2014. Analisis Determinan yang Mempengaruhi Bidan Desa dalam Ketepatan Rujukan pada Kasus Preeklampsia/Eklampsia di Kabupaten Banyumas. Jurnal Prada 5(2): 16-25.

[13]. Turban, E., dkk, 2005, Decicion Support Systems and Intelligent Systems, Andi Offset.

[14]. Witten, Ian H., Frank, Eibe. 2005. Data MiningPractical Machine Learning Tools and Techniques, Second Edition.Morgan Kaufmann Publishers is an imprint of Elsevier. 500 Sansome Street, Suite 400, San Francisco, CA 94111.