

## IDENTIFIKASI FAKTOR-FAKTOR YANG MEMPENGARUHI TERJADINYA PREEKLAMPSIA DENGAN METODE CHAID

(Studi Kasus pada Ibu Hamil Kategori Jampersal  
di RSUD Dr.Moewardi Surakarta)

Restu Sri Rahayu<sup>1</sup>, Moch. Abdul Mukid<sup>2</sup>, Triastuti Wuryandari<sup>3</sup>

<sup>1</sup>Mahasiswa Jurusan Statistika FSM Undip

<sup>2,3</sup>Staf Pengajar Jurusan Statistika FSM Undip

### ABSTRACT

Pre-eclampsia is a spesific pregnancy disease in which hypertency and proteinuria occurs after 20 weeks of pregnancy . This sickness is caused by many factors. To identify the factors, We lowercase a statistical analysis that can explain the characteristics of pregnant women who has pre-eclampsia. One analysis used for segmentation is CHAID (*Chi-Squared Automatic Interaction Detection*). This analysis classify and view the segmentation on nominal scale dependent variable (patient's status). CHAID analysis result indicates that the history of hypertension is the most influential independent variable. The tree diagram shows that there are seven segments of pregnant women, this study reveals that, there are three segments that need to be concerned because these segments show a high number and high index value exceeds 100% of pregnant women with pre-eclampsia. These segments need an effort to support the reduction of MMR. The three segment are segment pregnant women who has the history of hypertension; segment pregnant women of primary school degree and who are jobless, overweight, with no history of hypertension; and segment pregnant women with elementary and junior high school degree, who has jobs, and no hypertension history. Accuration of the CHAID algorithm in classifying is 78,2%.

**Keywords:** Pre-eclampsia, Classify, CHAID, Maternal Mortality Ratio, Accuration

### 1. PENDAHULUAN

Angka Kematian Ibu (AKI) merupakan salah satu indikator yang digunakan untuk melihat derajat kesehatan perempuan. Salah satu tujuan pembangunan millenium atau yang dikenal dengan Millenium Development Goals (MDG's) yaitu meningkatkan kesehatan ibu di mana target yang akan dicapai sampai tahun 2015 adalah menurunkan  $\frac{3}{4}$  resiko jumlah kematian ibu atau Angka Kematian Ibu (AKI) antara tahun 1990 dan 2015 sebesar 102 per 100.000 Kelahiran Hidup (KH) pada tahun 2015 (Bappenas, 2007). Salah satu penyebab utama kematian ibu di Indonesia di samping perdarahan adalah preeklampsia dan penyebab kematian perinatal yang tinggi. Preeklampsia adalah penyakit spesifik pada kehamilan yaitu terjadinya hipertensi dan proteinuria pada wanita hamil setelah umur kehamilan 20 minggu (Dimas, dalam Winfred *et al*, 2005).

Menurut Kim dan Loh (2001), algoritma pohon klasifikasi dibagi menjadi dua kelompok. Kelompok tersebut adalah kelompok biner contohnya CART dan QUEST dan kelompok nonbiner contohnya FACT, CRUISE dan CHAID. Algoritma CHAID merupakan algoritma pohon klasifikasi yang pertama kali diperkenalkan pada sebuah artikel yang berjudul "*An Exploratory Technique for investigating Large Quantities of Categorical Data*" oleh Dr. G. V. Kass tahun 1980 pada buku *Applied Statistics*. Menurut Gallagher (2000), CHAID merupakan suatu teknik iteratif yang menguji satu-persatu variabel independen yang digunakan dalam klasifikasi dan menyusunnya berdasarkan pada tingkat signifikansi statistik *chi-square* terhadap variabel dependennya. Oleh karena itu peneliti tertarik untuk mengetahui faktor-faktor yang mempengaruhi terjadinya preeklampsia pada ibu hamil dengan metode CHAID.

## 2. TINJAUAN PUSTAKA

### 2.1 Pengertian Preeklampsia

Preeklampsia adalah penyakit spesifik pada kehamilan yaitu terjadinya hipertensi dan proteinuria pada wanita hamil setelah umur kehamilan 20 minggu, terjadi pada sekitar 2% - 8% dari kehamilan (Dimas, dalam Winfred *et al*, 2005)

### 2.2 Data Kategorik

Data kategorik adalah data kualitatif sehingga untuk dapat dianalisis dengan menggunakan rumus statistika perlu diberi kode (coding) berupa angka. Menurut Agresti (1990), variabel data kategorik adalah salah satu dari skala pengukuran yang terdiri dari beberapa kategori data. Menurut skala pengukurannya, data dapat dibedakan atas empat, yaitu data nominal, data ordinal, data interval, dan data rasio (Hasan, 2003).

### 2.3 Chi-Squared Automatic Interaction Detector (CHAID)

Metode CHAID dikemukakan oleh Dr. G. V. Kass (1980) pada sebuah artikel yang berjudul "*An Exploratory Technique for investigating Large Quantities of Categorical Data*". CHAID merupakan suatu teknik iteratif yang menguji satu-persatu variabel independen yang digunakan dalam klasifikasi, dan menyusunnya berdasarkan pada tingkat signifikansi statistik *chi-square* terhadap variabel dependennya (Gallagher, 2000)

### 2.4 Variabel-Variabel dalam Metode CHAID

Menurut Gallagher (2000) CHAID membedakan variabel-variabel independen kategorik menjadi tiga tipe yang berbeda, yaitu:

a. Variabel Bebas

Variabel bebas adalah variabel independen yang kategori di dalamnya dapat digabungkan ketika keduanya berdekatan ataupun tidak (data nominal).

b. Variabel Monotonik

Variabel monotonik adalah variabel independen yang kategori di dalamnya dapat digabungkan oleh CHAID hanya jika keduanya berdekatan satu sama lain (data ordinal).

c. Variabel Mengambang (*floating*)

Variabel mengambang adalah variabel monotonik yang mengandung nilai kategori yang tidak jelas posisinya dalam urutan.

### 2.5 Uji Independensi $\chi^2$ (Chi-Square)

Teknik uji utama dalam metode CHAID adalah uji  $\chi^2$  (Chi-Square) yang memungkinkan untuk mengetahui kebebasan antara dua variabel pada tiap level (Kunto, 2006), dengan hipotesis sebagai berikut :

$H_0$  : Kedua kriteria klasifikasi adalah saling bebas (tidak terdapat hubungan antara variabel pertama dan variabel kedua atau independen)

$H_1$  : Kedua kriteria klasifikasi adalah tidak saling bebas (terdapat hubungan antara variabel pertama dan variabel kedua atau dependen)

Statistik Uji :

$$\chi^2_{\text{hitung}} = \sum_{i=1}^r \sum_{j=1}^c \left[ \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \right]$$

di mana :

$$E_{ij} = \frac{n_{i.} n_{.j}}{n}$$

$H_0$  ditolak jika  $\chi^2_{hitung} > \chi^2_{\alpha; (r-1)(c-1)}$  atau dengan membandingkan nilai signifikansinya dengan nilai alpha ( $\alpha$ ).

## 2.6 Koreksi Bonferroni (*Bonferroni Correction*)

Koreksi Bonferroni adalah suatu proses koreksi yang digunakan ketika beberapa uji statistik untuk kebebasan atau ketidakbebasan dilakukan secara bersamaan (Sharp *et al.*, 2002). Koreksi Bonferroni dilakukan ketika terjadi pengurangan pada tabel kontingensi. Jika tidak ada pengurangan pada tabel kontingensi asal, maka statistik uji  $\chi^2$  dapat digunakan.

Kass (1980) menyebutkan bahwa pengali Bonferroni dihitung sesuai dengan tipe variabel independen yaitu:

1. Variabel Monotonik (*Monotonic Predictors*)

$$M = \binom{c-1}{r-1}$$

2. Variabel Bebas (*Free Predictors*)

$$M = \sum_{i=0}^{r-1} (-1)^i \frac{(r-i)^c}{i!(r-i)!}$$

3. Variabel Mengambang (*Floating Predictors*)

$$M = \binom{c-2}{r-2} + r \binom{c-2}{r-1}$$

dimana :

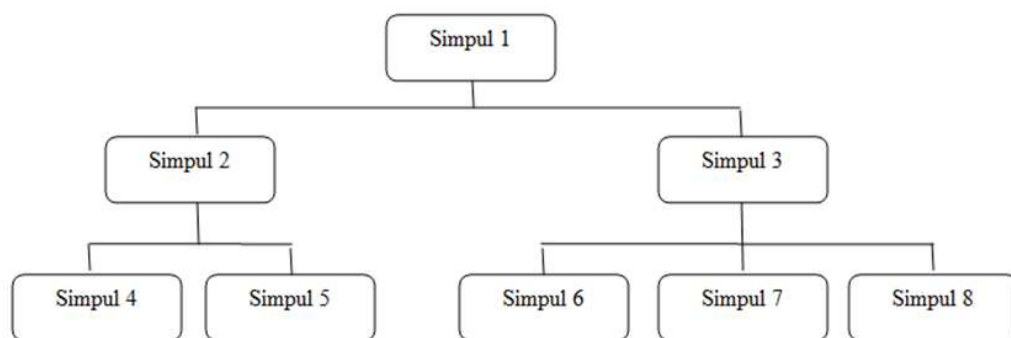
M = Pengali Bonferroni

c = banyaknya kategori variabel independen awal

r = banyaknya kategori variabel independen setelah penggabungan

## 2.7 Diagram Pohon

Hasil proses pembelahan dalam CHAID akan ditampilkan dalam sebuah diagram pohon.



Gambar 1 Diagram Pohon CHAID

Istilah dalam diagram pohon yaitu:

1. Simpul akar, pada Gambar 1 yang disebut simpul akar adalah simpul 1.
2. Simpul anak, pada Gambar 1 yang disebut simpul anak adalah simpul 2,3,4,5,6,7 dan 8.
3. Simpul induk, pada Gambar 1 yang disebut simpul induk adalah simpul 1, 2 dan 3.
4. Simpul akhir, pada Gambar 1 yang disebut simpul induk adalah simpul 4,5,6,7 dan 8.

Pada program SPSS, selain dapat menghasilkan output berupa diagram pohon, terdapat pula tabel *gain for nodes*. Tabel ini menyajikan informasi yang berkaitan dengan kategori pada variabel dependen yang menjadi fokus analisis di setiap simpul akhir. Terdapat 4 kolom

pada tabel *gain for nodes* yaitu kolom *node*, *gain*, *response* dan *index*. Pada tabel ini hal yang paling diamati adalah nilai *index*.

Nilai *index* mengindikasikan seberapa besar perbedaan persentase dari hasil observasi pada kategori variabel dependen yang menjadi fokus analisis dengan persentase hasil prediksi pada kategori variabel dependen yang menjadi fokus analisis. Nilai *index* yang lebih besar dari 100% mengindikasikan bahwa terdapat lebih banyak kasus untuk kategori variabel dependen yang menjadi fokus analisis dibandingkan dengan sampel keseluruhan sebaliknya jika nilai *index* kurang dari 100% mengindikasikan bahwa jumlah kasus untuk kategori variabel dependen yang menjadi fokus analisis lebih sedikit dari sampel keseluruhan.

## 2.8 Algoritma CHAID

### a. Penggabungan (*Merging*)

Tahap penggabungan untuk setiap variabel independen dalam menggabungkan kategori-kategori non-signifikan yaitu :

1. Membentuk tabel kontingensi dua arah untuk masing-masing variabel independen dengan variabel dependennya.
2. Menghitung statistik *chi-square* untuk setiap pasang kategori yang dapat dipilih untuk digabung menjadi satu.
3. Di antara pasangan-pasangan yang tidak signifikan, gabungkan sebuah pasangan kategori yang paling mirip (yaitu pasangan yang mempunyai nilai *chi-square* berpasangan terkecil dan *p-value* terbesar) menjadi sebuah kategori tunggal, dan kemudian dilanjutkan ke langkah 4.
4. Periksa kembali kesignifikansian kategori baru setelah digabung dengan kategori lainnya dalam variabel independen. Jika masih ada pasangan yang belum signifikan, ulangi langkah 3. Jika semua sudah signifikan lanjutkan langkah berikutnya.
5. Hitung *p-value* terkoreksi Bonferroni didasarkan pada tabel yang telah digabung.

### b. Pemisahan (*Splitting*)

Tahap pemisahan adalah tahap memilih variabel independen yang mana yang akan digunakan sebagai *split node* (pemisah simpul) yang terbaik. Langkah pemisahan adalah memilih variabel independen yang memiliki *p-value* terkecil (paling signifikan) yang akan digunakan sebagai pemisah simpul. Jika tidak ada variabel independen dengan nilai *p-value* yang signifikan, simpul tidak dipisah dan ditentukan sebagai simpul akhir.

### c. Penghentian (*Stopping*)

Peraturan penghentian proses pertumbuhan pohon yaitu :

1. Tidak ada lagi variabel independen yang signifikan.
2. Jika pohon sekarang mencapai batas nilai maksimum pohon dari spesifikasi.
3. Jika ukuran dari simpul anak kurang dari nilai ukuran simpul anak minimum yang telah ditentukan, atau berisi pengamatan-pengamatan dengan jumlah yang terlalu sedikit maka simpul tidak akan dipisah.

## 2.9 Pelabelan Kelas

Pelabelan kelas merupakan suatu tahapan menentukan setiap simpul akhir masuk ke dalam kelas kategori tertentu pada variabel dependen. Dasar yang digunakan untuk menentukan suatu simpul terakhir masuk ke dalam kelas kategori tertentu adalah persentase terbesar di antara kategori-kategori pada variabel dependen pada simpul akhir tersebut.

## 2.10 Ketepatan Pohon Klasifikasi

Evaluasi prosedur klasifikasi adalah suatu evaluasi yang melihat peluang kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi. Ukuran yang dipakai adalah Apparent Error Rate (APER). Nilai APER menyatakan nilai proporsi sampel yang salah diklasifikasikan oleh fungsi klasifikasi (Johnson dan Wichern, 2007). Menurut Prasetyo (2012), matriks konfusi merupakan tabel pencatat hasil kerja klasifikasi.

Tabel 1 Matriks Konfusi

Hasil Observasi (i)	Hasil Prediksi (j)	
	Kelas = A	Kelas = B
Kelas = A	$f_{11}$	$f_{10}$
Kelas = B	$f_{01}$	$f_{00}$

Maka dapat dihitung tingkat akurasi dan tingkat kesalahan prediksi :

$$\begin{aligned}
 1. \text{ Akurasi} &= \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \\
 2. \text{ Tingkat salah prediksi (APER)} &= \frac{f_{10} + f_{01}}{f_{11} + f_{10} + f_{01} + f_{00}}
 \end{aligned}$$

## 3. METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder dari hasil penelitian dr. Dimas Mardawati, Sp. OG, M.Kes terdiri dari 250 pasien preeklampsia dan 250 bukan pasien preeklampsia dengan responden ibu hamil usia 20-35 tahun di RSUD Dr. Moewardi Surakarta selama tahun 2013.

Variabel-variabel yang digunakan terdiri dari variabel dependen dan variabel independen. Variabel dependen adalah status pasien ibu hamil, di mana pada penelitian ini dikelompokkan ke dalam dua kategori yaitu :

1. Kelompok 0, yaitu kelompok ibu hamil dengan status pasien preeklampsia.
2. Kelompok 1, yaitu kelompok ibu hamil dengan status bukan pasien preeklampsia.

Variabel independen pada penelitian ini adalah paritas, pekerjaan, riwayat hipertensi, riwayat preeklampsia, Indeks Massa Tubuh, status pendidikan, dan pendapatan.

Langkah-langkah yang dilakukan dalam menganalisis data responden ibu hamil dengan status pasien preeklampsia dan bukan pasien preeklampsia adalah sebagai berikut:

1. Memasukkan data status pasien sebagai variabel respon dan paritas, pekerjaan, riwayat hipertensi, riwayat preeklampsia, Indeks Massa Tubuh, status pendidikan, pendapatan sebagai variabel independen.
2. Penyusunan tabulasi silang antara kategori-kategori variabel independen dengan kategori-kategori variabel dependen
3. Penggabungan kategori-kategori variabel independen yang memiliki angka uji *chi-square* terkecil.
4. Tahap pemisahan memilih variabel independen yang mana yang akan digunakan sebagai pemisah simpul yang terbaik.
5. Proses pemilihan variabel pemisah dilakukan berulang-ulang sampai proses pembentukan pohon klasifikasi CHAID berhenti dengan peraturan penghentian yang diberlakukan.
6. Interpretasi pohon klasifikasi yang terbentuk.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Tahap 1 : Penggabungan

Tahap penggabungan dilakukan dengan uji *chi-square* untuk menentukan apakah kategori-kategori dalam variabel independen bersifat seragam dan bisa digabungkan menjadi satu.

Pada penelitian ini, variabel independen yang mengalami proses penggabungan adalah variabel status pendidikan.

Dari pengujian statistik *chi-square* yang sudah dilakukan, disusun dalam tabel daftar keputusan, yaitu:

Tabel 2 Nilai statistik *Chi-Square* Status Pasien dan Status Pendidikan

Status Pasien	Jenis Kategori	<i>Chi-square</i>	<i>Sig. 2-tailed</i>	Keputusan
0 dan 1	SD (0) dan SMP (1)	2,0438	0,153	H <sub>0</sub> diterima
0 dan 1	SMP (1) dan SMA (2)	13,1795	0,000	H <sub>0</sub> ditolak

Dari Tabel 2 diperoleh hasil bahwa nilai  $\chi^2_{hitung}$  untuk uji antara variabel status pasien dan status pendidikan untuk kategori SMP (1) dan SMA (2) H<sub>0</sub> ditolak, sehingga kesimpulan yang diambil adalah kedua variabel tersebut tidak saling bebas. Kategori lainnya yaitu kategori SD (0) dan SMP (1) H<sub>0</sub> diterima, sehingga dapat disimpulkan bahwa antara variabel status pasien dan status pendidikan untuk kategori SD dan SMP saling bebas. Oleh karena itu kategori SD (0) dan SMP (1) digabung menjadi satu kategori campuran.

Langkah selanjutnya adalah penerapan uji *chi-square* antara variabel status pasien dan status pendidikan setelah penggabungan kategori SD (0) dan SMP (1). Berdasarkan uji *chi-square* yang dilakukan antara variabel status pasien dan variabel status pendidikan kategori campuran SD SMP (0,1) dan SMA (2) diperoleh hasil bahwa kedua variabel tersebut saling bebas sehingga penggabungan sudah maksimal.

Pengali Bonferroni untuk variabel monotonik adalah nilai kombinasi dari  $M = \binom{c-1}{r-1}$  dimana :

M = Pengali Bonferroni

c = banyaknya kategori variabel independen awal

r = banyaknya kategori variabel independen setelah penggabungan

sehingga nilai  $M = \binom{c-1}{r-1} = \binom{3-1}{2-1} = 2$

Setelah mendapatkan pengali Bonferroni maka nilai tingkat kesalahan untuk uji signifikansi antara variabel status pasien dan variabel status pendidikan kategori campuran SD SMP (0,1) dan SMA (2) adalah hasil kali *p-value* dan M, yaitu  $(0,000)(2) = 0,000$ .

Oleh karena itu keputusan yang diambil adalah H<sub>0</sub> ditolak karena nilai *p-value* terkoreksi tetap lebih kecil dari nilai  $\alpha = 5\%$ , yang artinya variabel status pendidikan kategori campuran SD SMP (0,1) dan SMA (2) tidak saling bebas.

### 4.2 Tahap Pemisahan

Dalam proses menentukan simpul akar dicari variabel independen yang paling signifikan yaitu memiliki nilai statistik *chi-square* yang paling besar. Hasil dari pengujian *chi-square* untuk menentukan simpul akar adalah sebagai berikut:



Tabel 3 Nilai Statistik *Chi-square* yang Signifikan Simpul Akar

Variabel Bebas	Jenis Kategori	Kategori Status Pasien	<i>Chi-square</i>	<i>Sig.2tailed</i>	Keputusan
Paritas	0 dan 1	0 dan 1	7,0202	0,008*	H <sub>0</sub> ditolak
Pekerjaan	0 dan 1	0 dan 1	21,0918	0,000*	H <sub>0</sub> ditolak
Riwayat Hipertensi	0 dan 1	0 dan 1	74,7126**	0,000*	H <sub>0</sub> ditolak
Riwayat Preeklampsia	0 dan 1	0 dan 1	37,6344	0,000*	H <sub>0</sub> ditolak
Indeks Massa Tubuh	0 dan 1	0 dan 1	6,3103	0,012*	H <sub>0</sub> ditolak
Status Pendidikan	(0,1) dan 2	0 dan 1	22,7475	0,000*	H <sub>0</sub> ditolak
Pendapatan	0 dan 1	0 dan 1	5,3625	0,021*	H <sub>0</sub> ditolak

\*variabel signifikan pada  $\alpha = 5\%$

\*\*Variabel dengan nilai *chi-square* terbesar

Tabel 3 menunjukkan bahwa variabel riwayat hipertensi memiliki nilai *chi-square* terbesar yaitu 74,7126 dengan p-value 0,000 oleh karena itu variabel riwayat hipertensi merupakan variabel pemilah terbaik di simpul akar.

Kategori pada variabel riwayat hipertensi digunakan sebagai pemisah untuk menentukan simpul anak. Proses pemisahan akan terus dilakukan pada setiap simpul selama masih ada variabel-variabel independen yang signifikan.

#### 4.3 Tahap 3 : Penghentian

Pada penelitian ini, tahap penghentian terjadi pada simpul 1, 5,7,8,9,10,dan 11. Penjelasan proses penghentian adalah sebagai berikut:

1. Simpul 5, simpul 9, dan simpul 11 mengalami proses penghentian karena jumlah responden masing-masing simpul hanya 34, 37 dan 29 sehingga jumlah kasusnya kurang dari jumlah minimum yaitu 50.
2. Simpul 1, simpul 7, simpul 8, dan simpul 10 mengalami proses penghentian karena berdasarkan uji chi square pada Lampiran 7 sudah tidak ada lagi variabel independen yang signifikan untuk membagi variabel dependen dalam keempat simpul tersebut.

#### 4.4 Analisis Hasil

Berdasarkan output yang dihasilkan dapat dilihat bahwa variabel independen yang signifikan dalam proses pembentukan diagram pohon adalah riwayat hipertensi, pekerjaan, Indeks Massa Tubuh, status pendidikan. Jumlah simpul yang dihasilkan adalah 12 dengan kedalaman pohon 4 dan memiliki 7 simpul akhir.

Dengan memperhatikan diagram pohon yang terbentuk maka segmentasi pasien dapat diklasifikasikan sebagai berikut :

1. Segmen Pasien Preeklampsia
  - a. Simpul akhir 1 : ibu hamil yang memiliki riwayat hipertensi dengan indeks 200%.
  - b. Simpul akhir 3 : ibu hamil tidak memiliki riwayat hipertensi, tidak bekerja, memiliki Indeks Massa Tubuh normal dengan status pendidikan SD dengan indeks 108,1%.
  - c. Simpul akhir 6 : ibu hamil tidak memiliki riwayat hipertensi, bekerja, dan berstatus pendidikan SD, SMP dengan indeks 200%.
2. Segmen Bukan Pasien Preeklampsia
  - a. Simpul akhir 2 : ibu hamil tidak memiliki riwayat hipertensi, tidak bekerja, dan memiliki Indeks Massa Tubuh kegemukan dengan indeks 0%.

- b. Simpul akhir 4 : ibu hamil tidak memiliki riwayat hipertensi, tidak bekerja, dan memiliki Indeks Massa Tubuh normal, dengan status pendidikan SMP dengan indeks 50,9%.
- c. Simpul akhir 5: ibu hamil tidak memiliki riwayat hipertensi, tidak bekerja, dan memiliki Indeks Massa Tubuh normal dengan status pendidikan SMA dengan indeks 0%.
- d. Simpul akhir 7 : ibu hamil tidak memiliki riwayat hipertensi, bekerja, dan memiliki status pendidikan SMA dengan indeks 85,1%.

Dalam penelitian ini terdapat tiga simpul yang memiliki index melebihi 100% yaitu :

1. Simpul 7 (simpul akhir 6) dan simpul 1 (simpul akhir 1) nilai indeksnya sebesar 200% yang artinya tingkat kemungkinan segmen ini merupakan ibu hamil berstatus pasien preeklampsia 100% lebih tinggi dari kemungkinan status ibu hamil menderita preeklampsia dalam seluruh sampel ibu hamil.
2. Simpul 9 (simpul akhir 3) nilai indeksnya sebesar 108,1% yang artinya tingkat kemungkinan segmen ini merupakan ibu hamil berstatus pasien preeklampsia 8,1% lebih tinggi dari kemungkinan status ibu hamil menderita preeklampsia dalam seluruh sampel ibu hamil.

#### 4.4 Ukuran Ketepatan Prediksi

Uji ketepatan pohon klasifikasi dalam mengklasifikasikan data dapat dilakukan menggunakan matriks konfusi pada Tabel 4.

Tabel 4 Matriks Konfusi Hasil Klasifikasi

Hasil Observasi (i)	Hasil Prediksi (j)	
	Pasien preeklampsia	Bukan pasien preeklampsia
Pasien preeklampsia	158	92
Bukan pasien preeklampsia	17	233

Uji ketepatan pohon klasifikasi tingkat kesalahan prediksi dihitung sebagai berikut :

$$\text{Akurasi} = \frac{158+233}{138+112+0+250} = 0,782$$

$$\text{Tingkat kesalahan prediksi} = \frac{17+92}{38+112+0+250} = 0,218$$

Maka tingkat akurasi pohon klasifikasi dalam mengklasifikasikan data sebesar 0,782 atau 78,2 %, dan dengan tingkat kesalahan memprediksi sebesar 0,218 atau 21,8 %.

## 5. PENUTUP

### 5.1 Kesimpulan

Dengan memperhatikan diagram pohon yang terbentuk dari analisis CHAID maka segmentasi responden pada penelitian ini dapat diklasifikasikan sebagai berikut :

1. Segmen pertama adalah ibu hamil yang memiliki riwayat hipertensi dengan jumlah responden pasien preeklampsia sebanyak 65 orang dan bukan pasien preeklampsia sebanyak 0 orang serta masuk dalam kelas kategori pasien preeklampsia.



2. Segmen kedua adalah ibu hamil tidak memiliki riwayat hipertensi, tidak bekerja dan memiliki Indeks Massa Tubuh normal dengan jumlah responden pasien preeklampsia sebanyak 0 orang dan bukan pasien preeklampsia sebanyak 34 orang serta masuk dalam kelas kategori bukan pasien preeklampsia.
3. Segmen ketiga adalah status pasien tidak memiliki riwayat hipertensi, tidak bekerja dan memiliki Indeks Massa Tubuh kegemukan dengan status pendidikan SD dengan jumlah responden pasien preeklampsia sebanyak 20 orang dan bukan pasien preeklampsia sebanyak 17 orang serta masuk dalam kelas kategori pasien preeklampsia.
4. Segmen keempat adalah status pasien tidak memiliki riwayat hipertensi, tidak bekerja dan memiliki Indeks Massa Tubuh kegemukan dengan status pendidikan SMP dengan jumlah responden pasien preeklampsia sebanyak 29 orang dan bukan pasien preeklampsia sebanyak 85 orang serta masuk dalam kelas kategori bukan pasien preeklampsia.
5. Segmen kelima adalah status pasien tidak memiliki riwayat hipertensi, tidak bekerja dan memiliki Indeks Massa Tubuh kegemukan dengan status pendidikan SMA dengan jumlah responden pasien preeklampsia sebanyak 0 orang dan bukan pasien preeklampsia sebanyak 29 orang serta masuk dalam kelas kategori bukan pasien preeklampsia.
6. Segmen keenam adalah status pasien tidak memiliki riwayat hipertensi, bekerja dan memiliki pendidikan SD,SMP dengan jumlah responden pasien preeklampsia sebanyak 73 orang dan bukan pasien preeklampsia sebanyak 0 orang serta masuk dalam kelas kategori pasien preeklampsia.
7. Segmen ketujuh adalah status pasien tidak memiliki riwayat hipertensi, bekerja dan memiliki status pendidikan SMA dengan jumlah responden pasien preeklampsia sebanyak 63 orang dan bukan pasien preeklampsia sebanyak 85 orang serta masuk dalam kelas kategori bukan pasien preeklampsia

Pada data status pasien preeklampsia dan bukan pasien preeklampsia dalam Tugas Akhir ini, akurasi algoritma CHAID dalam mengklasifikasikan data sebesar 0,782 atau 78,2 %.

Dari ketujuh segmen responden ibu hamil tersebut, segmen pertama, ketiga dan keenam merupakan segmen yang perlu ditindaklanjuti karena pada segmen ini jumlah responden yang menderita preeklampsia menunjukkan jumlah yang tinggi dan nilai indeksnya melebihi 100%.

## 5.2 Saran

1. Untuk penelitian selanjutnya bisa menggunakan data dengan variabel dependen lebih dari dua kategori
2. Studi lebih lanjut diperlukan untuk memperoleh variabel-variabel independen yang baru sebagai faktor resiko penyakit preeklampsia.

## 6. DAFTAR PUSTAKA

- Agresti, A. 1990. *Categorical Data Analysis*. New York: John Wiley & Sons. Inc.
- Badan Perencanaan Pembangunan Nasional. 2007. *Laporan Perkembangan Pencapaian Millenium Development Goals Indonesia 2007*. Jakarta : Badan Perencanaan Pembangunan Nasional.
- Gallagher, C.A. 2000. *An Iterative Approach to Classification Analysis*. [www.casact.org/library/ratemaking/90dp237.pdf](http://www.casact.org/library/ratemaking/90dp237.pdf). (diakses tanggal 15 November 2014).
- Hasan, M.I. 2003. *Pokok-Pokok Materi Statistik 1 (Statistik Deskriptif)*. Jakarta: PT. Bumi Aksara.
- Johnson, R.A. dan Winchern, D.W. *Applied Multivariate Statistical Analysis Sixth Edition*. New Jersey: Pearson.

- Kass, G.V. 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics* 29, No. 2; 119-127.
- Kim, H. dan Loh, W.Y. 2001. Classification Trees with Unbiased Multiway Splits. *Am. Statist. Assoc.* 96; 590-604.
- Kunto, Y.S. 2006. Analisis CHAID Sebagai Alat Bantu Statistika Untuk Segmentasi Pasar. *Jurnal Manajemen Pemasaran*, Vol. 1 No. 2. Surabaya: Universitas Kristen Petra.
- Prasetyo, E. 2012. *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: C.V Andi Offset.
- Sharp, A., Romaniuk, J. Cierpieki, S. 2002. *The Performance Of Segmentation Variables: A Comparative Study*. [http:// 130.195.95.7:8081 /www/ ANZMAC1998 /Cd\\_rom/ Sharp222.pdf](http://130.195.95.7:8081/www/ANZMAC1998/Cd_rom/Sharp222.pdf). (diakses tanggal 15 November 2014).
- Winfred, W. Williams, Jr., *et al*, 2005. *Case 38-2005: A 29-Year Old Pregnant Woman with the Nephrotic Syndrome and Hypertension*. New England J. Of Medicine, case records of the Massachusetts general hospital: 2593.