

KETEPATAN KLASIFIKASI PEMBERIAN KARTU KELUARGA SEJAHTERA DI KOTA SEMARANG MENGGUNAKAN METODE REGRESI LOGISTIK BINER DAN METODE CHAID

Muhammad Arif Suhendra¹, Dra. Dwi Ispriyanti, M.Si², Drs. Sudarno, M.Si³

^{1,2,3}Departemen Statistika FSM Universitas Diponegoro

Email: dwiispriyanti@gmail.com

ABSTRAK

Menurut BPS, jumlah penduduk miskin di Kota Semarang pada Maret 2018 adalah sebesar 73,65 ribu orang. Salah satu program pemerintah dalam percepatan penanggulangan kemiskinan adalah dengan mengeluarkan Kartu Keluarga Sejahtera (KKS) yang diberikan kepada masyarakat yang kurang mampu. Penelitian ini bertujuan untuk mengetahui besarnya ukuran ketepatan klasifikasi pemberian KKS di Kota Semarang. Metode klasifikasi statistik yang digunakan adalah metode Regresi Logistik Biner dan metode *Chi-Squared Automatic Interaction Detection* (CHAID). Pemberian KKS dipengaruhi oleh banyak faktor, diantaranya jumlah anggota keluarga, status perkawinan, jenis kelamin kepala keluarga, usia kepala keluarga, jenjang pendidikan kepala keluarga dan kepemilikan/penguasaan HP. Pada penelitian ini, data yang digunakan adalah data sekunder hasil Survey Sosial Ekonomi Nasional (SUSENAS) tahun 2018 yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Tengah. Perbandingan data *training* dan *testing* yang digunakan adalah 60:40. Hasil analisisnya menunjukkan bahwa dengan menggunakan Regresi Logistik Biner, faktor-faktor yang berpengaruh adalah jumlah anggota keluarga dan jenjang pendidikan kepala keluarga dengan ketepatan klasifikasi sebesar 88% dan kesalahan 12%, sedangkan dengan menggunakan CHAID, faktor-faktor yang berpengaruh adalah jumlah anggota keluarga, status perkawinan, usia kepala keluarga, jenjang pendidikan kepala keluarga dan kepemilikan/penguasaan HP dengan ketepatan klasifikasi sebesar 90,2% dan kesalahan 9,8%.

Kata kunci: Kartu Keluarga Sejahtera, Klasifikasi, Regresi Logistik Biner, CHAID

1. PENDAHULUAN

Kesejahteraan masyarakat merupakan suatu keadaan terpenuhinya kebutuhan material, spritual, dan sosial warga negara agar dapat hidup layak dan mampu mengembangkan diri, sehingga dapat melaksanakan fungsi sosialnya (UU No 11 Tahun 2009). Pada tahun 1997-1998, Indonesia dilanda sebuah krisis ekonomi terbesar yang Indonesia pernah alami sejak merdeka tahun 1945, yang dikenal dengan sebutan krisis keuangan Asia yang menyebabkan jumlah penduduk miskin di Indonesia bertambah banyak. Salah satu cara pemerintah menanggulangi kemiskinan di Indonesia yaitu dengan mengeluarkan Kartu Keluarga Sejahtera (KKS). Kartu Keluarga Sejahtera adalah salah satu program pemerintah dalam percepatan penanggulangan kemiskinan, hal ini tercantum pada Peraturan Presiden (Perpres) Nomor 166 Tahun 2014. Mengingat banyaknya kriteria dan jumlah masyarakat yang akan menerima KKS, maka akan membuat tim seleksi mengalami kesulitan dalam melakukan penyeleksian. Terlebih lagi apakah pemberian KKS kepada masyarakat sudah tepat sasaran atau belum. Maka dari itu diperlukan suatu alat bantu yang tepat untuk mengetahui ketepatan klasifikasi masyarakat yang menerima dan

tidak menerima KKS tersebut. Metode klasifikasi dalam statistik yang dapat digunakan adalah metode CHAID dan metode regresi logistik biner.

Metode *Chi-Squared Automatic Interaction Detection* (CHAID) merupakan metode yang digunakan untuk menguji satu persatu variabel independen yang digunakan dalam klasifikasi, menggabungkan pasangan yang tidak signifikan dari variabel independen yang memiliki 3 atau lebih kategori asal dan menyusunnya berdasarkan pada tingkat signifikansi statistik *chi-square* terhadap variabel dependennya (Gallagher, 2000).

Menurut Agresti (2002), metode regresi logistik biner merupakan metode yang digunakan untuk menganalisis faktor-faktor yang berpengaruh (variabel independen) terhadap KKS (variabel dependen) yang berupa data kualitatif dikotomi yang bernilai 1 untuk menyatakan keberadaan sebuah karakteristik (menerima KKS) dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karakteristik (tidak menerima KKS). Metode regresi logistik biner menghasilkan model akhir yang nantinya digunakan untuk mengetahui ketepatan klasifikasinya.

2. TINJAUAN PUSTAKA

2.1. Kartu Keluarga Sejahtera (KKS)

Kartu Keluarga Sejahtera (KKS) merupakan salah satu program pemerintah di bidang sosial ekonomi untuk mempercepat penanggulangan kemiskinan nasional. Pada awalnya program tersebut adalah hasil dari pengembangan program sebelumnya yaitu Kartu Perlindungan Sejahtera (KPS). Fungsi KKS selain menjadi penanda masyarakat kurang mampu, juga berfungsi sebagai kartu identitas untuk mendapatkan Program Simpanan Keluarga Sejahtera (PSKS).

2.2. Metode Regresi Logistik Biner

Menurut Hosmer dan Lemeshow (2000) regresi logistik merupakan metode yang menghubungkan antara variabel dependen yang bersifat kategorik dengan variabel independen. Model regresi logistik dengan variabel dependen yang terdiri dari dua kategori disebut model regresi logistik biner (dikotomus). Model regresi logistik biner digunakan untuk menganalisis hubungan antara satu variabel dependen (Y) dengan satu atau lebih variabel independen (X). Variabel dependennya berupa data kualitatif dikotomi yang bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karakteristik. Adapun bentuk model regresi logistiknya adalah sebagai berikut :

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})} \quad (1)$$

Sedangkan logit dari $\pi(x_i)$ adalah

$$g(x_i) = \ln \left[\frac{\pi(x_i)}{1 - \pi(x_i)} \right] = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} \quad (2)$$

2.2.1 Uji Signifikansi Parameter

Uji Signifikansi parameter yang digunakan adalah uji rasio Likelihood dan uji Wald.

1. Uji Rasio Likelihood

Uji Rasio Likelihood dilakukan untuk mengetahui signifikansi koefisien β terhadap variabel dependen secara serentak (Hosmer dan Lemeshow, 2000).

Hipotesis:

$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$ (tidak ada pengaruh antara variabel independen dengan variabel dependennya)

H_1 : paling sedikit ada satu $\beta_j \neq 0$ dengan $j = 1, 2, \dots, p$ (paling tidak ada satu variabel independen yang berpengaruh terhadap variabel dependennya)

Statistik Uji:

$$G = -2 \ln \left(\frac{\text{likelihood tanpa variabel bebas}}{\text{likelihood dengan variabel bebas}} \right)$$

Kriteria Penerimaan: Terima H_0 jika $G \leq \chi^2_{(p,\alpha)}$ dan terima H_1 jika $G > \chi^2_{(p,\alpha)}$.

2. Uji Wald

Uji Wald dilakukan untuk mengetahui pengaruh setiap koefisien β_i secara individual (Hosmer dan Lemeshow, 2000).

Hipotesis:

$H_0: \beta_j = 0$ dengan $j = 1, 2, \dots, p$ (tidak ada pengaruh antara variabel independen dengan variabel dependennya)

$H_1: \beta_j \neq 0$ dengan $j = 1, 2, \dots, p$ (ada pengaruh antara variabel independen dengan variabel dependennya)

Statistik Uji:

$$W_j = \left\{ \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right\}^2$$

Kriteria penerimaan: Terima H_0 jika $W_j \leq \chi^2_{(\alpha,1)}$ dan terima H_1 jika $W_j > \chi^2_{(\alpha,1)}$.

3. Uji Kesesuaian Model

Hosmer dan Lemeshow (2000) mengatakan bahwa uji kesesuaian model merupakan uji yang dilakukan untuk mengetahui apakah ada perbedaan antara prediksi dan hasil observasi (model sesuai atau tidak).

Hipotesis:

H_0 : Model sesuai (tidak ada perbedaan antara hasil observasi dengan hasil prediksi)

H_1 : Model tidak sesuai (ada perbedaan antara hasil observasi dengan hasil prediksi)

Statistik Uji:

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n_k \bar{\pi}_k)^2}{(n_k \bar{\pi}_k)(1 - \bar{\pi}_k)}$$

dengan:

o_k : Jumlah nilai variabel respon pada grup ke- k

$\bar{\pi}_k$: Rata-rata taksiran peluang pada grup ke- k

g : Banyak grup

n_k : Banyak observasi pada grup ke- k

Kriteria penerimaan: Terima H_0 jika $\hat{C} \leq \chi^2_{(\alpha,g-2)}$ dan terima H_1 jika $\hat{C} > \chi^2_{(\alpha,g-2)}$.

2.3 Metode CHAID

CHAID pertama kali diperkenalkan dalam sebuah artikel berjudul “*An Exploratory Technique for Investigating Large Quantities of Categorical Data*” oleh Dr.G.V.Kass tahun 1980. Prosedurnya merupakan bagian dari teknik terdahulu yang dikenal dengan *Automatic Interaction Detector* (AID) dan menggunakan statistik chi-square sebagai alat

utamanya. Analisis CHAID secara garis besar dibagi menjadi tiga tahap yaitu penggabungan, pemisahan dan penghentian.

2.3.1 Variabel-variabel dalam Analisis CHAID

Variabel dependen dan independen dalam analisis CHAID adalah variabel kategorik. Menurut Gallagher (2000), CHAID akan membedakan variabel-variabel independen kategorik menjadi tiga bentuk yang berbeda, yaitu:

1. Monotonik yaitu variabel independen yang kategori di dalamnya dapat dikombinasikan atau digabungkan oleh CHAID hanya jika keduanya berdekatan satu sama lain atau mengikuti urutan aslinya (data ordinal). Contoh: usia dan pendapatan.
2. Bebas yaitu variabel independen yang kategori di dalamnya dapat dikombinasikan atau digabungkan ketika keduanya berdekatan ataupun tidak (data nominal). Contoh: pekerjaan, kelompok etnik dan area geografis.
3. Mengambang (floating) yaitu variabel independen yang kategori di dalamnya dapat diperlakukan seperti monotonik kecuali untuk kategori yang missing value, yang dapat dikombinasikan dengan kategori manapun.

2.3.2 Uji Chi-Square (Independensi)

Langkah-langkah dalam melakukan uji chi-square adalah sebagai berikut:

Hipotesis:

H_0 : kedua kriteria klasifikasi bebas (independen)

H_1 : kedua kriteria klasifikasi tidak bebas (dependen)

Statistik uji:

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

$$E_{ij} = \frac{n_i \cdot n_j}{n}$$

Kriteria penerimaan: Terima H_0 jika $\chi_{hitung}^2 \leq \chi_{\alpha:(r-1)(c-1)}^2$ dan terima H_1 jika $\chi_{hitung}^2 > \chi_{\alpha:(r-1)(c-1)}^2$.

2.3.3 Koreksi Bonferroni

Pengurangan kategori pada tahap penggabungan algoritma CHAID dibutuhkan uji signifikansi. Ketika terjadi pengurangan yaitu c kategori variabel asal menjadi r kategori ($c > r$), maka tingkat kesalahan (α_i) dikalikan dengan pengali Bonferroni sesuai dengan tipe variabelnya. Gallagher (2000) menyebutkan bahwa pengali Bonferroni untuk masing-masing jenis variabel-variabel independen adalah sebagai berikut:

1. Variabel independen monotonik

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

2. Variabel independen bebas

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

3. Variabel independen mengambang (floating)

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

2.3.4 Algoritma CHAID

a. Tahap Penggabungan

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 digabungkan menjadi satu
3. Di antara pasangan-pasangan yang tidak signifikan, gabungkan sebuah pasangan kategori yang paling mirip (yaitu pasangan yang mempunyai nilai *chi-square* terkecil dan *p-value* terbesar) menjadi sebuah kategori tunggal dari variabel independen yang mempunyai 3 atau lebih kategori asal.
4. Memeriksa kembali signifikansi kategori baru setelah penggabungan dengan kategori lainnya dalam variabel independen. Jika ada pasangan yang belum signifikan, maka ulangi langkah 3. Jika semua langkah sudah signifikan, maka lanjutkan ke langkah berikutnya.
5. Menghitung *p-value* terkoreksi Bonferroni didasarkan pada tabel yang sudah digabungkan.

b. Tahap Pemisahan

Memilih variabel independen terbaik, yaitu variabel independen dengan nilai *p-value* terkecil, kemudian melakukan pembagian kelompok dengan variabel independen (yaitu gunakan masing-masing kategori-kategori variabel independen tersebut, yang telah digabungkan secara optimal, untuk menentukan sub pembagian dari kelompok induk menjadi sub kelompok yang baru). Jika tidak ada variabel independen dengan nilai *p-value* yang signifikan, jangan memulai pembagian kelompok tersebut.

c. Tahap Penghentian

Tahap penghentian dilakukan ketika pertumbuhan pohon harus dihentikan. Menurut Rokach dan Maimon (2008), penyekatan simpul dalam pembentukan pohon harus dihentikan karena:

1. Pada simpul hanya terdapat kasus yang berasal dari salah satu kelas variabel dependen.
2. Kedalaman pohon maksimal telah tercapai.
3. Jumlah kasus di simpul terminal kurang dari jumlah minimal kasus untuk menjadi simpul orang tua (parent node).
4. Jika simpul disekat, jumlah kasus dalam satu atau lebih simpul anak akan kurang dari jumlah minimal kasus untuk simpul anak, maka proses penyekatan simpul berhenti.
5. Jika semua variabel independen mempunyai nilai signifikansi lebih besar dari nilai *alpha* yang ditentukan, maka simpul tidak dapat disekat.

2.4 Ketepatan Klasifikasi

Menurut Johnson dan Wichern (1992), prosedur klasifikasi yang bisa digunakan untuk mengukur ketepatan klasifikasi adalah *Apparent Error Rate* (APER). Nilai APER menyatakan proporsi sampel yang salah diklasifikasikan oleh fungsi klasifikasi. Kesalahan klasifikasi dapat dilihat pada tabel matriks konfusi berikut:

Tabel 1. Matriks Konfusi

Hasil Observasi	Taksiran	
	y_1	y_2
y_1	n_{11}	n_{12}
y_2	n_{21}	n_{22}

Maka nilai APER dapat dihitung dengan rumus:

$$APER = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}}$$

Akurasi = $1 - APER$

3. METODE PENELITIAN

3.1. Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder yaitu data hasil Survei Sosial Ekonomi Nasional (SUSENAS) di Kota Semarang Tahun 2018. Adapun data tersebut diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Tengah.

3.2. Variabel Penelitian

Variabel pada penelitian ini terdiri dari variabel dependen (Y) dan beberapa variabel independen (X). Adapun variabel-variabel tersebut adalah sebagai berikut:

1. Variabel Dependen

Y = KKS, 0 = tidak menerima KKS, 1 = menerima KKS

2. Variabel Independen

X1 = jumlah anggota keluarga, X2 = status perkawinan, X3 = jenis kelamin kepala keluarga, X4 = usia kepala keluarga, X5 = jenjang pendidikan kepala keluarga, X6 = kepemilikan/penguasaan HP

3.3 Teknik Pengolahan Data

Pada penelitian ini, analisis yang digunakan menggunakan metode regresi logistik biner dan metode CHAID. Adapun software yang digunakan sebagai alat bantu dalam menganalisis yaitu Microsoft Excel 2007 dan IBM SPSS Statistics 23.

1. Tahapan analisis data metode regresi logistik biner adalah sebagai berikut :

- Memasukkan data KKS sebagai variabel dependen dan jumlah anggota keluarga, status perkawinan, jenis kelamin kepala keluarga, usia kepala keluarga, , jenjang pendidikan tertinggi kepala keluarga, dan kepemilikan HP sebagai variabel independen.
- Membagi data menjadi data training dan data testing.
- Menghitung estimasi parameter menggunakan data training untuk mendapatkan model awal.
- Melakukan uji likelihood untuk mengetahui apakah variabel independen secara bersama-sama mempengaruhi variabel dependennya.
- Melakukan uji Wald untuk mengetahui ada atau tidaknya pengaruh masing-masing variabel independen terhadap variabel dependennya.
- Melakukan uji kesesuaian model.
- Menentukan model akhir.
- Menghitung ketepatan klasifikasi model menggunakan data testing.

2. Tahapan analisis data menggunakan metode CHAID
 - a. Memasukkan data KKS sebagai variabel dependen dan jumlah anggota keluarga, status perkawinan, jenis kelamin kepala keluarga, usia kepala keluarga, jenjang pendidikan tertinggi kepala keluarga, dan kepemilikan HP sebagai variabel independen.
 - b. Membagi data menjadi data training dan data testing.
 - c. Melakukan tabulasi silang antara variabel independen dan variabel dependennya menggunakan data testing.
 - d. Melakukan penggabungan terhadap kategori-kategori penggabungan dalam variabel independen yang memiliki nilai chi-square terkecil.
 - e. Pemilihan variabel independen yang mempunyai nilai p-value terkecil sebagai pemisah untuk membentuk sub kelompok. Proses pemilihan variabel untuk memisah terus berjalan hingga peraturan penghentian yang diberlakukan.
 - f. Menginterpretasikan pohon klasifikasi yang terbentuk.
 - g. Menghitung nilai ketepatan klasifikasi.

4. HASIL DAN PEMBAHASAN

4.1 Analisis Data Menggunakan Regresi Logistik Biner

Berdasarkan pengolahan data didapat model regresi logistik biner. Berikut adalah model yang dihasilkan:

$$\pi(x_i) = \frac{e^{g(x_i)}}{1 + e^{g(x_i)}}$$

dengan

$$g(x_i) = -4,554 + 0,894X_{1(1)} - 17,901X_{2(1)} - 0,173X_{2(2)} - 0,142X_{3(1)} - 0,227X_{4(1)} + 2,728X_{5(1)} + 1,560X_{5(2)} + 0,227X_{6(1)}$$

4.1.1 Uji Rasio Likelihood

Hipotesis

$$H_0: \beta_1 = \beta_2 = \dots = \beta_8 = 0$$

H_1 : paling sedikit ada satu $\beta_j \neq 0$ dengan $j = 1, 2, \dots, 8$

Taraf Signifikansi $\alpha = 5\%$

Statistik Uji:

$$G = -2 \ln \left(\frac{\text{likelihood tanpa variabel bebas}}{\text{likelihood dengan variabel bebas}} \right) = 31,947$$

Nilai $G > \chi^2_{(8;0,05)}$ yaitu $31,947 > 15,507$ maka H_1 diterima sehingga paling sedikit ada satu variabel independen yang memberikan pengaruh terhadap variabel dependennya.

4.1.2 Uji Wald

Hipotesis

$$H_0: \beta_j = 0 \text{ dengan } j = 1, 2, \dots, 8$$

$H_1: \beta_j \neq 0$ dengan $j = 1, 2, \dots, 8$

Taraf signifikansi $\alpha = 5\%$

Statistik uji Wald:

$$W_j = \left\{ \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right\}^2$$

Berdasarkan output regresi logistik biner dapat dibentuk tabel 2 sebagai berikut:

Tabel 2. Hasil Uji Wald Model Awal

Variabel	Koefisien	SE Koefisien	Wald	p-value
Jumlah_AK(1)	0.894	0.382	5.477	0.019
Status(1)	-17.901	6814.356	0.000	0.998
Status(2)	-0.173	0.806	0.046	0.830
JK(1)	-0.142	0.820	0.030	0.863
Usia(1)	-0.227	0.392	0.334	0.563
Pendidikan(1)	2.728	1.038	6.904	0.009
Pendidikan(2)	1.560	1.064	2.148	0.143
HP(1)	0.227	0.428	0.280	0.597
Constant	-4.554	1.109	16.866	0.000

Berdasarkan tabel 2, nilai $W_j > \chi^2_{(0,05;1)}$ untuk variabel jumlah anggota keluarga (X_1) dan jenjang pendidikan tertinggi kepala keluarga (X_5) sehingga kedua variabel tersebut signifikan terhadap model. Karena terdapat beberapa variabel independen yang tidak signifikan maka akan dibentuk model baru dengan tidak mengikutsertakan variabel-variabel yang tidak signifikan tersebut.

4.1.3 Uji Kesesuaian Model

Hipotesis

H_0 : Model sesuai (tidak ada perbedaan antara hasil observasi dengan hasil prediksi)

H_1 : Model tidak sesuai (ada perbedaan antara hasil observasi dengan hasil prediksi)

Taraf signifikansi $\alpha = 5\%$

Statistik uji:

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n_k \pi_k)^2}{(n_k \pi_k)(1 - \pi_k)} = 2,488$$

Nilai $\hat{C} \leq \chi^2_{(0,05;4)}$ yaitu $2,488 < 9,488$ maka H_0 diterima sehingga dapat disimpulkan model sesuai (tidak ada perbedaan antara hasil observasi dengan hasil prediksi).

4.1.4 Model Akhir

$$\pi(x_i) = \frac{e^{g(x_i)}}{1 + e^{g(x_i)}}$$

dengan

$$g(x_i) = -5.051 + 0,876X_{1(1)} + 2.935X_{5(1)} + 1.736X_{5(2)}$$

4.1.5 Ketepatan Klasifikasi

Tabel 3. Matriks Konfusi Regresi Logistik Biner

Observasi	Prediksi	
	Tidak menerima KKS	Menerima KKS
Tidak menerima KKS	324	0
Menerima KKS	44	0

$$\begin{aligned} \text{APER} &= \frac{0 + 44}{324 + 0 + 44 + 0} \\ &= 0,120 \end{aligned}$$

$$\begin{aligned}
\text{Akurasi} &= 1 - \text{APER} \\
&= 1 - 0,120 \\
&= 0,88 \\
&= 88\%
\end{aligned}$$

4.2 Analisis Data Menggunakan CHAID

4.2.1 Tahap Penggabungan

Pada penelitian ini, variabel independen yang mengalami proses penggabungan adalah variabel status perkawinan dan jenjang pendidikan kepala keluarga. Adapun hasil uji statistik chi-squarenya adalah sebagai berikut:

Tabel 4. Nilai Statistik Chi-Square

Variabel	Definisi	Kategori	Chi-Square	Keputusan
X ₂	Status Perkawinan	Belum Kawin (1) dan Kawin (2)	1,411	H ₀ diterima
		Belum Kawin (1) dan Cerai (3)	5,227	H ₀ ditolak
		Kawin (2) dan Cerai (3)	7,098	H ₀ ditolak
X ₅	Pendidikan Kepala keluarga	<SMA (1) dan SMA (2)	12,751	H ₀ ditolak
		SMA (2) dan >SMA (3)	3,447	H ₀ diterima

Dari tabel 4 diperoleh hasil bahwa H₀ diterima untuk variabel status perkawinan kategori belum kawin (1) dan kawin (2) dan variabel pendidikan kepala keluarga kategori SMA (2) dan >SMA (3) sehingga dapat disimpulkan kategori-kategori tersebut saling bebas (tidak signifikan). Maka kategori-kategori tersebut perlu digabungkan menjadi sebuah kategori tunggal. Selanjutnya dilakukan pengujian yang sama untuk kategori gabungan dengan kategori lainnya dan didapatkan hasil bahwa kedua kategori tersebut sudah signifikan sehingga penggabungan telah maksimal.

- Pengali Bonferroni untuk variabel status perkawinan

$$\begin{aligned}
M &= \sum_{i=0}^{r-1} (-1)^i \frac{(r-i)^c}{i!(r-i)!} = \left[(-1)^0 \frac{(2-0)^3}{0!(2-0)!} \right] + \left[(-1)^1 \frac{(2-1)^3}{1!(2-1)!} \right] \\
&= 4 - 1 \\
&= 3
\end{aligned}$$

Uji signifikansi terkoreksi = 0,006 × 3 = 0,018. Karena 0,018 < 0,05 maka dapat disimpulkan bahwa variabel status perkawinan (belum kawin; kawin dan cerai) signifikan (tidak saling bebas) terhadap variabel KKS.

- Pengali Bonferroni untuk variabel pendidikan

$$M = \binom{c-1}{r-1} = \binom{3-1}{2-1} = \frac{2!}{1!(2-1)!} = 2$$

Uji signifikansi terkoreksi = 0,000 × 2 = 0. Karena 0,00 < 0,05 maka dapat disimpulkan bahwa variabel jenjang pendidikan tertinggi kepala keluarga (<SMA dan SMA; >SMA) signifikan (tidak saling bebas) terhadap variabel KKS.

4.2.2 Tahap Pemisahan

Tahap pemisahan adalah tahap memilih variabel independen mana yang akan digunakan sebagai simpul pemisah terbaik. Simpul pemisah terbaik merupakan variabel independen yang memiliki nilai chi-square terbesar atau nilai p-value terkecil.

Tabel 5. Uji Chi-Square Tahap Pemisahan Simpul Akar (t_0)

Variabel	Definisi	Kategori	Chi-Square	Keputusan	Kesimpulan
X_1	Jumlah Anggota Keluarga	>4 (1)	0,171	H_0 diterima	Tidak signifikan
		≤ 4 (2)			
X_2	Status Perkawinan	Belum Kawin;Kawin (1)	8,467	H_1 diterima	Signifikan
		Cerai (2)			
X_3	Jenis Kelamin Kepala Keluarga	Laki-laki (1)	0,838	H_0 diterima	Tidak signifikan
		Perempuan (2)			
X_4	Usia Kepala Rumah Tangga	>54 (1)	0,384	H_0 diterima	Tidak signifikan
		≤ 54 (2)			
X_5	Jenjang Pendidikan Tertinggi Kepala Keluarga	<SMA (1)	26,454	H_1 diterima	Signifikan
		SMA;>SMA (2)			
X_6	Kepemilikan/Penguasaan HP	Tidak memiliki/menguasai HP (1)	47,540	H_1 diterima	Signifikan
		Memiliki/menguasai HP (2)			

Tabel 5 menunjukkan bahwa variabel kepemilikan/penguasaan HP merupakan variabel pemilah terbaik pada simpul akar (t_0) karena mempunyai nilai chi-square terbesar yaitu 47,540. Proses pemilahan terus dilakukan pada setiap simpul selama masih terdapat variabel-variabel independen yang signifikan.

4.2.3 Tahap Penghentian

Tahap pemberhentian terjadi pada simpul pendidikan kepala keluarga kategori <SMA, pendidikan kepala keluarga kategori \geq SMA, jumlah anggota keluarga ≤ 4 , jumlah anggota keluarga >4, status perkawinan kategori (belum kawin;kawin) dan simpul status perkawinan kategori (cerai). Hal tersebut terjadi karena disebabkan hal berikut:

1. Proses pemberhentian terjadi pada simpul pendidikan kepala keluarga kategori <SMA, pendidikan kepala keluarga kategori \geq SMA, jumlah anggota keluarga ≤ 4 , status perkawinan kategori (belum kawin;kawin) dan simpul status perkawinan kategori (cerai) disebabkan karena sudah tidak ada lagi variabel independen yang signifikan.
2. Proses pemberhentian terjadi pada simpul jumlah anggota keluarga >4 disebabkan karena jumlah kasus di simpul tersebut kurang dari jumlah minimum untuk menjadi simpul orang tua (*parent node*) yaitu 6.

4.2.4 Ketepatan Klasifikasi

Tabel 6. Matriks Konfusi CHAID

Observasi	Prediksi	
	Tidak Menerima KKS	Menerima KKS
Tidak Menerima KKS	319	5
Menerima KKS	31	13

$$APER = \frac{31 + 5}{319 + 5 + 31 + 13} = 0,098$$

$$\begin{aligned} \text{Akurasi} &= 1 - APER \\ &= 0,902 \\ &= 90,2\% \end{aligned}$$

5 KESIMPULAN

Berdasarkan hasil dan pembahasan dapat diambil kesimpulan sebagai berikut:

1. Faktor-faktor yang mempengaruhi pemberian KKS di Kota Semarang dengan menggunakan metode regresi logistik biner adalah variabel jumlah anggota keluarga dan pendidikan kepala keluarga, sedangkan dengan menggunakan metode CHAID diperoleh hasil bahwa yang berpengaruh adalah variabel jumlah anggota keluarga, status perkawinan, usia kepala keluarga, pendidikan kepala keluarga dan kepemilikan/penguasaan HP.
2. Nilai akurasi pengklasifikasian pemberian KKS di Kota Semarang menggunakan regresi logistik biner adalah 88% dengan nilai APER 12%, sedangkan dengan menggunakan CHAID diperoleh nilai akurasi sebesar 90,2% dengan nilai APER 9,8%.
3. Dari hasil perbandingan ketepatan klasifikasi menggunakan metode regresi logistik biner dan CHAID diperoleh hasil bahwa metode CHAID lebih baik daripada metode regresi logistik biner dalam pengklasifikasian pemberian KKS di Kota Semarang.

DAFTAR PUSTAKA

- Agresti, A. (2002). *Categorical Data Analysis Second Edition*. New York: John Willey and Sons.
- Gallagher, C.A. (2000). *An Iterative Approach to Classification Analysis*.
<https://www.casact.org/pubs/dpp/dpp90/90dpp237.pdf>. (Diakses pada Juli 2019)
- Hosmer, D.W., dan Lemeshow. (2000). *Applied Logistic Regression*. USA: John Willey and Sons.
- Johnson, R. A., dan Winchern. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall International, Inc.
- Kass, G.V. (1980). *An Exploratory Technique for Invertigating Large Quantities of Categorical Data*. *Applied Statistical* 29, No. 2; 119-127.
- Rokach, L., dan Maimon, O. (2008). *Data Mining with Decision Trees "teory and Application"*. USA : World Scientific Publishing Co. Pte. Ltd.
- Undang-Undang Republik Indonesia Tahun 2009 Tentang *Kesejahteraan Sosial*.