KETEPATAN KLASIFIKASI METODE REGRESI LOGISTIK DAN CHAID DENGAN PEMBOBOTAN SAMPEL

Skripsi
disusun sebagai salah satu syarat
untuk memperoleh gelar Sarjana Sains
Program Studi Matematika

oleh

Puspa Juwita

4111413041

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI SEMARANG

2017

PERNYATAAN KEASLIAN TULISAN

Saya menyatakan bahwa skripsi ini bebas plagiat, kecuali yang secara tertulis dirujuk dalam skripsi ini dan disebutkan dalam daftar pustaka. Apabila dikemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai kenentuan perundang-undangan.

Semarang, Oktober 2017

Puspa Juwita

PENGESAHAN

Skripsi yang berjudul

Ketepatan Klasifikasi Metode Regresi Logistik dan CHAID dengan Pembobotan Sampel

disusun oleh

Puspa Juwita

4111413041

telah dipertahankan di hadapan sidang Panitia Ujian Skripsi FMIPA UNNES pada tanggal 23 Oktober 2017.

Panitia:

Ketua

Prof. Dr. Zaenuri, S.E., M.Si., Akt.
NIP. 196412231988031001

Sekretaris

Drs. Arief Agoestanto, M.Si.
NIP. 196807221993031005

Ketua Penguji

Dr. Nur Karomah Dwidayati, M.Si.
NIP. 196605041990022001

Anggota Penguji/
Pembimbing I

Drs. Sugiman, M.Si.
NIP. 196401111989011001

Anggota Penguji/
Pembimbing II

Putriaji Hendikawati, S.Si., M.Pd., M.Sc.
NIP. 198208182006042001

MOTTO DAN PERSEMBAHAN

MOTTO

1. What we learn with pleasure, we never forget (Alfred Mercier)
2. Tulislah sesuatu yang bahkan kau sendiri akan tergetar apabila membacanya (Jombang Santani Khairen)
3. Yakinlah dalam hidup maka kau akan menemukan hal-hal yang tak kau sangka

PERSEMBAHAN

1. Kedua orang tua, Bapak Akhwani dan Ibu Daryanti
2. Kakakku, Mas Cahyo
3. Adikku, Ulil
4. Keluarga besarku
5. Kawan-kawan kos, Ilmi, Nindi, Kiki S, dan Kiki J
6. Lia, Farid, Niken, Etika
7. Kawan PELANGI, Aida, Lembayung, Titi, dan Uung
8. Sahabat dan teman matematika murni 2013

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah memberikan nikmat dan karunia-Nya serta kemudahan sehingga penulis dapat menyelesaikan skripsi yang berjudul "Ketepatan Klasifikasi Metode Regresi Logistik dan CHAID dengan Pembobotan Sampel".

Penyusunan skripsi ini dapat diselesaikan berkat kerjasama, bantuan, dan dorongan dari berbagai pihak. Oleh karena itu penulis mengucapkan terima kasih kepada:

1. Prof. Dr. Fathur Rokhman, M.Hum., Rektor Universitas Negeri Semarang.
2. Prof. Dr. Zaenuri, S.E., M.Si., Akt., Dekan FMIPA Universitas Negeri Semarang.
3. Drs. Arief Agoestanto, M.Si., Ketua Jurusan Matematika FMIPA Universitas Negeri Semarang.
4. Drs. Mashuri, M.Si., Ketua Prodi Matematika FMIPA Universitas Negeri Semarang.
5. Drs. Sugiman, M.Si., selaku Dosen Pembimbing I yang telah memberikan bimbingan, pengarahan, nasehat, dan saran selama penyusunan skripsi ini.
6. Putriaji Hendikawati, S.Si., M.Pd., M.Sc., selaku Dosen Pembimbing II yang telah memberikan bimbingan, pengarahan, nasehat, dan saran selama penyusunan skripsi ini.
7. Dr. Nur Karomah Dwidayati, M.Si., selaku Dosen Penguji yang telah memberikan penilaian dan saran dalam perbaikan skripsi ini.

8. Drs. Sugiman, M.Si., selaku Dosen Wali yang telah memberikan bimbingan dan arahan.
9. Dosen-dosen Matematika Universitas Negeri Semarang yang telah membekali penulis dengan berbagai ilmu selama mengikuti perkuliahan sampai akhir penulisan skripsi ini.
10. Ibu Daryanti dan Bapak Akhwani tercinta, kakak dan adik tersayang, Mas Cahyo dan Ulil serta seluruh keluarga yang senantiasa memberikan dukungan, semangat dan doa yang tiada putusnya.
11. Ilmi, Nindi, Kiki S, dan Kiki J yang tidak pernah bosan untuk selalu ada disegala suasana.
12. Sahabat dan teman-teman seperjuangan prodi Matematika FMIPA Unnes 2013 yang selalu memberikan semangat untuk besama-sama berjuang dalam mendapat gelar S.Si ini.
13. Semua pihak yang tidak dapat disebutkan satu per satu yang telah memberikan bantuan dan semangat.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan. Oleh karena itu, penulis mengharapkan saran dan kritik yang membangun dari pembaca.

Semarang, Oktober 2017

Penulis

ABSTRAK

Juwita, Puspa. 2017. *Ketepatan Klasifikasi Metode Regresi Logistik dan CHAID dengan Pembobotan Sampel*. Skripsi, Jurusan Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Utama Drs. Sugiman, M.Si. dan Pembimbing Pendamping Putriaji Hendikawati, S.Si., M.Pd., M.Sc.

Kata Kunci: Regresi Logistik; CHAID; Pembobotan Sampel

Pada saat ini klasifikasi telah digunakan pada berbagai bidang, seperti pemerintahan, pendidikan, kesehatan, maupun teknologi. Metode regresi logistik dan CHAID merupakan dua metode klasifikasi yang dapat menangani variabel dependen kategori. Pembobotan sampel bertujuan untuk membuat sampel menjadi lebih representatif terhadap populasi dengan jumlah yang terbatas. Indonesia merupakan salah satu negara berkembang di dunia. Indonesia telah mengalami kemajuan pesat di bidang ekonomi dan sosial. Pengangguran merupakan indikator kesejahteraan masyarakat yang diakibatkan oleh pembangunan ekonomi. Pada tahun 2015, Kabupaten Temanggung mempunyai TPAK (Tingkat Partisipasi Angkatan Kerja) paling tinggi di Jawa Tengah. Penulis tertarik untuk melakukan penelitian dengan judul "Ketepatan Klasifikasi Metode Regresi Logistik dan CHAID dengan Pembobotan Sampel". Tujuan penelitian ini adalah untuk menentukan ketepatan metode regresi logistik dan CHAID dengan pembobotan sampel pada klasifikasi status angkatan kerja Kabupaten Temanggung tahun 2015.

Analisis data dilakukan menggunakan metode regresi logistik dan CHAID. Sebelum data dianalaisis, dilakukan pembobotan sampel terlebih dahulu. Analisis regresi logistik melalui beberapa tahap, yaitu estimasi parameter, uji signifikansi serentak, uji signifikansi parsial, dan uji kesesuaian model secara berturut-turut. Tahap-tahap dalam analisis CHAID secara berturut-turut adalah tahap penggabungan (*merging*), tahap pemisahan (*splitting*), dan tahap penghentian (*stopping*). Ketepatan metode regresi logistik dan CHAID akan dihitung menggunakan rumus 1 – APER (*Apparent Error Rate*).

Ketepatan metode regresi logistik dan metode CHAID dalam klasifikasi status angkatan kerja Kabupaten Temanggung tahun 2015 secara berturut-turut adalah 96,4% dan 96,6%. Berdasarkan penelitian ini, dapat ditarik kesimpulan bahwa metode CHAID mempunyai ketepatan lebih tinggi dalam klasifikasi status angkatan kerja Kabupaten Temanggung tahun 2015.

DAFTAR ISI

DAFTAR TABEL

Daftar Tabel Halaman

DAFTAR GAMBAR

Daftar Gambar Halaman

DAFTAR LAMPIRAN

Daftar Lampiran Halaman

BAB I

PENDAHULUAN

1. 1 Latar Belakang

Metode klasifikasi telah banyak digunakan dalam berbagai bidang, seperti bidang pendidikan, pemerintahan, kesehatan, teknologi, maupun sosial. Klasifikasi sendiri didefinisikan sebagai pekerjaan mengelompokkan suatu objek ke dalam kategori tertentu. Klasifikasi dapat dilakukan pada data kategorik maupun bukan, jika data bukan kategorik maka harus diubah dalam bentuk kategorik terlebih dahulu.

Regresi merupakan suatu metode statistika yang digunakan untuk menyelidiki pola hubungan antara dua atau lebih variabel, yaitu variabel dependen dan variabel independen (Suyanti *et al*, 2014). Menurut (Kleinbaum & Klein, 2010), regresi logistik adalah pendekatan matematis untuk mendeskripsikan hubungan beberapa variabel independen dengan variabel dependen dikotomi. Regresi logistik dapat digunakan untuk klasifikasi variabel dependen dikotomi. Regresi logistik sangat menarik karena beberapa hal, yaitu (1) secara konsep sederhana, (2) mudah diinterpretasikan, dan (3) telah terbukti dapat menyediakan hasil yang akurat dan baik (Antipov & Pokryshevskaya, 2009). Berdasarkan Imaslihkah *et al* (2013), regresi logistik mempunyai ketepatan klasifikasi yang akurat.

CHAID (*Chi-square Automatic Interaction Detection*) merupakan salah satu analisis pohon keputusan (*decision tree*). Keunggulan dari metode *decision tree* yaitu membutuhkan waktu yang cepat untuk membentuk diagram pohon, representasi visual, dan mudah diinterpretasikan (Swain, 2016).Metode pohon keputusan (*decision tree*) mempunyai beberapa keunggulan dibandingkan metode lainnya untuk klasifikasi atau prediksi, seperti *neural network* dan analisis diskriminan (Cha, G.W *et al*, 2017). CHAID adalah 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 *et al*, 2000). Dengan kata lain, CHAID mengklasifikasikan variabel dependen kategori ke dalam kategori tertentu berdasarkan statistik chi-square variabel independen terhadap variabel dependen. Pada setiap cabangnya, CHAID melakukan tahap penggabungan (*merging)* dan tahap pemisahan (*splitting)* (Ritschard, 2010). Dibandingkan *decision tree* yang lain CHAID memiliki beberapa keunggulan, yaitu (1) node dan cabang CHAID yang dihasilkan berdasarkan tabel kontingensi sehingga node-node saling berkaitan, (2) tidak dibatasi dengan *binary split* (tidak seperti *decision tree* CART), dan (3) lebih cepat digunakan. Menurut Rahayu *et al* (2015), metode CHAID akurat untuk klasifikasi.

Keakuratan suatu klasifikasi juga ditentukan oleh sampel penelitian. Semakin representatif sampel terhadap populasi, maka hasil klasifikasi akan semakin baik dan akurat. Oleh karena itu diperlukan pembobotan sampel, yaitu pemberian bobot pada sampel sehingga dapat mewakili/mendekati jumlah populasi sesungguhnya.

Indonesia merupakan salah satu negara berkembang di dunia. Meski begitu, Indonesia telah mengalami kemajuan pesat di bidang ekonomi dan sosial. Semakin banyak penduduk Indonesia yang menikmati standar hidup yang lebih tinggi. Indonesia juga memiliki potensi pertumbuhan yang kuat, yaitu populasi yang masih muda. Seperti negara berkembang pada umumnya, Indonesia memiliki tantangan pada bidang perekonomian. Tantangan tersebut yaitu melakukan diversifikasi ekonomi dengan memperkuat kualitas sumber daya manusia sehingga memungkinkan sektor-sektor ekonomi yang padat keterampilan dan padat tenaga kerja untuk terus berkembang. Selain itu, untuk memastikan meningkatnya standar hidup dan kesejahteraan (OECD, 2016).

Menurut Dewi (2010), pengangguran merupakan indikator kesejahteraan masyarakat yang diakibatkan oleh pembangunan ekonomi. Berdasarkan pendapat Alghofari (2010), "Jumlah pengangguran merupakan masalah yang sangat serius dan sangat mempengaruhi kondisi negara, karena jumlah pengangguran merupakan indikator majunya perekonomian suatu negara yang dapat menunjukkan tingkat distribusi pendapatan yang merata atau tidak di negara tersebut".

Kabupaten Temanggung merupakan salah satu kabupaten di Provinsi Jawa Tengah. Pada tahun 2015, Kabupaten Temanggung mempunyai Tingkat Partisipasi Angkatan Kerja (TPAK) paling tinggi di Provinsi Jawa tengah, yaitu sebesar 75,47%. Tingkat Partisipasi Angkatan Kerja (TPAK) mengindikasikan besarnya penduduk usia kerja yang aktif secara ekonomi di suatu wilayah. TPAK diukur sebagai persentase angkatan kerja (bekerja dan pengangguran) terhadap penduduk usia kerja (angkatan kerja dan bukan angkatan kerja). Indikator ini menunjukkan

besaran relatif dari pasokan tenaga kerja (*labor supply*) yang tersedia untuk memproduksi barang-barang dan jasa dalam suatu perekonomian. Dengan demikian dapat disimpulkan bahwa Kabupaten Temanggung mempunyai potensi angkatan kerja yang sangat besar.

Berdasarkan BPS (2015), pengangguran merupakan akibat dari ketidakmampuan pasar tenaga kerja menyerap angkatan kerja yang terus bertambah yang disebabkan oleh jumlah lapangan pekerjaan lebih kecil dari jumlah pencari kerja, kompetensi pencari kerja tidak sesuai dengan pasar tenaga kerja, maupun kurang efektifnya informasi pasar tenaga kerja bagi pencari kerja. Status angkatan kerja sebagai bekerja atau pengangguran ditentukan berdasarkan kriteria yang ditetapkan oleh Badan Pusat Statistik (BPS). Data ini tentu sangat berguna bagi instansi pemerintah terkait untuk mengambil suatu keputusan atau kebijakan. Dengan mengetahui jumlah penduduk yang bekerja atau menganggur, dapat mendukung pengambilan keputusan atau kebijakan instansi terkait sehingga akan semakin efektif, efisien, dan akurat.

Guna mengetahui metode klasifikasi terbaik di antara regresi logistik dan CHAID, diperlukan suatu penelitian untuk mengetahui ketepatan klasifikasi kedua metode tersebut. Penulis tertarik melakukan penelitian dengan judul "Ketepatan Klasifikasi Metode Regresi Logistik dan CHAID dengan Pembobotan Sampel" dengan studi kasus Sakernas Kabupaten Temanggung tahun 2015".

1. 2 Rumusan Masalah

Adapun rumusan masalah penelitian ini adalah sebagai berikut.

a. Berapa ketepatan metode regresi logistik pada klasifikasi status angkatan kerja Kabupaten Temanggung 2015?
b. Berapa ketepatan metode CHAID pada klasifikasi status angkatan kerja Kabupaten Temanggung 2015?
c. Dari kedua metode tersebut, manakah metode yang mempunyai ketepatan paling tinggi untuk klasifikasi status angkatan kerja Kabupaten Temanggung 2015?

1. 3 Pembatasan Masalah

Pada penelitian ini, metode yang digunakan adalah regresi logistik dan CHAID. Data penelitian ini merupakan data berbobot. Data yang digunakan diperoleh dari hasil Survei Angkatan Kerja Nasional (Sakernas) Kabupaten Temanggung 2015. Ketepatan klasifikasi kedua metode akan dihitung menggunakan 1 – APER (*Apparent Error Rate*).

1. 4 Tujuan Penelitian

Adapun tujuan penelitian ini adalah sebagai berikut.

a. Menentukan ketepatan regresi logistik pada klasifikasi status angkatan kerja Kabupaten Temanggung 2015,
b. Menentukan ketepatan CHAID pada klasifikasi status angkatan kerja Kabupaten Temanggung 2015,
c. Menentukan metode yang lebih baik untuk klasifikasi status angkatan kerja Kabupaten Temanggung 2015.

1. 5 Manfaat Penelitian

Penelitian ini memiliki manfaat tidak hanya bagi penulis, tetapi juga bagi pembaca dari kalangan akademik, pemerintahan, maupun masyarakat umum. Penulis akan semakin memahami metode statistik khususnya metode regresi logistik dan metode CHAID yang digunakan dalam penelitian ini. Selain itu, dapat memperkaya pengetahuan penulis mengenai kondisi ketenagakerjaan di Indonesia khususnya Kabupaten Temanggung. Dengan demikian, penulis menjadi semakin memahami penerapan matematika terutama bidang statistika pada bidang ketenagakerjaan.

Pembaca dapat mengetahui serta mempelajari metode klasifikasi statistik, yaitu metode regresi logistik dan CHAID. Selain itu, dapat memberikan gambaran mengenai penerapan matematika dalam bidang ketenagakerjaan, yaitu kasus klasifikasi status angkatan kerja.

BAB II

TINJAUAN PUSTAKA

2. 1 Pembobotan Sampel

Setiap subjek dalam suatu populasi harus memiliki kesempatan yang sama untuk terpilih sebagai sampel, ini dinamakan prinsip *Equal Probability of Selection Method* (EPSEM). Prinsip EPSEM terpenuhi pada sampel acak sederhana, sedangkan pada sampel kompleks belum tentu terpenuhi. Beberapa hal yang membuat sampel menjadi sampel kompleks antara lain, stratifikasi (*stratification*) dan pengambilan sampel bertahap (*multistage sampling*). Sampel kompleks memberikan beberapa keuntungan, antara lain sampel kompleks membuat anggaran menjadi efisien, meningkatkan ketepatan estimasi, dan cukup merepresentasikan populasi. Sampel kompleks dapat ditemui pada penelitian survei. Pada sampel kompleks, peluang terpilihnya subjek pada strata/kluster yang berbeda tidak sama. Pembobotan sampel bertujuan untuk menyamakan peluang terpilihnya subjek pada strata/kluster yang berbeda.

Dengan pembobotan sampel, sebuah sampel akan merepresentasikan lebih dari satu subjek pada populasi penelitian. Pembobotan sampel akan menghasilkan data berbobot, yaitu data tunggal yang jumlah data atau bobotnya lebih banyak. Data berbobot disajikan dalam tabel distribusi frekuensi tunggal. Sebagai contoh perhatikan data berikut ini.

Hasil ulangan dari 40 siswa salah satu SMA di Semarang adalah sebagai berikut.

25, 30, 45, 50, 30, 50, 75, 70, 65, 70,

50, 45, 25, 30, 70, 45, 50, 75, 80, 40,

45, 60, 70, 75, 80, 60, 35, 40, 50, 30,

25, 50, 60, 60, 75, 80, 60, 55, 60, 35.

Data tersebut kita susun menjadi data berbobot sebagai berikut.

Tabel 2.1 Distribusi Frekuensi

Nilai	25	30	35	40	45	50	55	60	65	70	75	80
Frekuensi	3	4	2	2	4	6	1	6	1	4	4	3

2. 2 Metode Regresi Logistik

Variabel dependen dalam regresi logistik adalah kategori yang biasanya dinotasikan dengan 0 (untuk kejadian gagal) atau 1 (untuk kejadian sukses). Regresi logistik memprediksi peluang Y = 1 dan bukan 0 diberikan nilai X tertentu (Josephat & Ismail, 2012). Regresi logistik sering digunakan untuk memprediksi variabel dependen biner (Koskas, 2015). Pengkodean kategori variabel dependen dapat diubah-ubah karena tidak mempunyai makna, dengan kata lain skala pengukuran datanya adalah nominal. Metode yang digunakan dalam regresi logistik mempunyai prinsip yang sama dengan metode yang digunakan dalam regresi linear (ALR). Model regresi logistik dapat dinyatakan sebagai berikut.

$$\pi(x) = \frac{e^{(\beta_0+\beta_1 x_{1i}+\cdots+\beta_j x_{ji})}}{1+e^{(\beta_0+\beta_1 x_{1i}+\cdots+\beta_j x_{ji})}} \quad (2.1)$$

Keterangan

π_i : peluang observasi ke-i

x_{ij} : variabel bebas ke-j kasus ke-i

β_j : koefisien ke-j

Tahap-tahap untuk membangun model regresi logistik secara berturut-turut adalah estimasi parameter, uji signifikansi serentak, uji signifikansi parsial, dan uji kesesuaian model.

2. 2. 1 Estimasi Parameter

Untuk memperoleh model regresi logistik dibutuhkan estimasi parameter $\boldsymbol{\beta} = \beta_0, \beta_1, \dots, \beta_p$. Estimasi parameter regresi logistik menggunakan metode *maximum likelihood*. Metode *maximum likelihood* menghasilkan nilai parameter yang memaksimalkan peluang untuk memperoleh nilai observasi. Langkah-langkah metode *maximum likelihood* adalah sebagai berikut.

1. Membentuk fungsi likelihood

   Fungsi likelihood menyatakan peluang nilai observasi sebagai fungsi parameter. Fungsi likelihood dapat dinyatakan sebagai berikut.

$$l(\beta) = \prod_{i=1}^{n} \pi(x_i)^{y_i}[1 - \pi(x_i)]^{(1-y_i)} \quad (2.2)$$

   Keterangan

   $\pi(x_i)$: peluang observasi ke-i

   y_i : nilai observasi ke-i

   $l(\beta)$: fungsi likelihood

2. Membentuk log fungsi likelihood

   Menurut Hosmer & Lemeshow (2000), prinsip *maximum likelihood* menyatakan bahwa estimasi parameter $\boldsymbol{\beta} = \beta_0, \beta_1, \dots, \beta_p$ adalah nilai yang memaksimalkan

fungsi likelihood. Untuk mempermudah perhitungan digunakan log fungsi likelihood. Log fungsi likelihood dapat dinyatakan sebagai berikut.

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^{n} \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \quad (2.3)$$

Keterangan

$L(\beta)$: log fungsi likelihood

$l(\beta)$: fungsi likelihood

$\pi(x_i)$: peluang observasi ke-i

y_i : nilai observasi ke-i

3. Menurunkan log fungsi likelihood

Nilai β diperoleh dengan menurunkan $L(\beta)$ terhadap $\beta_0, \beta_1, \dots, \beta_p$ dan menyamadengankannya dengan 0. Setelah diturunkan akan diperoleh persamaan *likelihood* sebagai berikut.

$$\sum_{i=1}^{n} [y_i - \pi(x_i)] = 0 \quad (2.4)$$

dan

$$\sum_{i=1}^{n} x_i [y_i - \pi(x_i)] = 0 \quad (2.5)$$

Keterangan

$\pi(x_i)$: peluang observasi ke-i

y_i : nilai observasi ke-i

x_i : variabel bebas ke-i

2. 2. 2 Uji Signifikansi Serentak

Uji signifikansi serentak digunakan untuk mengetahui signifikansi koefisien β terhadap variabel dependen secara serentak atau keseluruhan. Statistik uji yang digunakan adalah uji rasio likelihood atau uji G. Hipotesis yang digunakan adalah sebagai berikut.

H_0 : $\beta_1 = \beta_2 = \cdots = \beta_p = 0$

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

Statistik uji G

$$G = -2 \ln \left(\frac{likelihood\ tanpa\ variabel\ bebas}{likelihood\ dengan\ variabel\ bebas} \right) \quad (2.6)$$

Daerah penolakan : tolak H_0 jika $G > \chi^2_{(p,\alpha)}$ atau tolak H_0 jika nilai *p-value* < α.

2. 2. 3 Uji Signifikansi Parsial

Uji signifikansi parsial digunakan untuk mengetahui apakah variabel independen secara individu berpengaruh secara signifikan terhadap variabel dependen. Pada uji parsial ini menggunakan uji Wald (Hosmer & Lemeshow, 2000). Hipotesis yang digunakan adalah sebagai berikut.

H_0 : $\beta_j = 0$ dengan $j = 1,2, \ldots, p$

H_1 : $\beta_j \neq 0$ dengan $j = 1,2, \ldots, p$

Statistik uji Wald

$$W_j = \left\{ \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right\}^2 \quad (2.7)$$

Daerah penolakan : tolak H_0 jika $W_j > \chi^2_{(\alpha,1)}$ atau nilai *p-value* < α.

2. 2. 4 Uji Kesesuaian Model

Uji kesesuaian model menyelidiki apakah model sudah sesuai, yaitu tidak terdapat perbedaan yang nyata antara hasil observasi dengan prediksi. Uji Hosmer Lemeshow digunakan untuk menguji kesesuaian model regresi logistik. Hipotesis yang digunakan adalah sebagai berikut.

H_0 : tidak terdapat perbedaan antara hasil observasi dengan hasil prediksi

H_1 : terdapat perbedaan antara hasil observasi dengan hasil prediksi

Statistik uji

$$\hat{c} = \sum_{k=1}^{g} \frac{(o_k - n'_k \bar{\pi}_k)}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} \quad (2.8)$$

Keterangan

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

Daerah penolakan : tolak H_0 jika $\hat{c} > \chi^2_{(\alpha,g-2)}$ atau H_0 ditolak jika *p-value* < α.

Menurut Graubard *et al* (1997), uji Hosmer Lemeshow tidak dapat dimodifikasi untuk kasus pembobotan smapel. Maka digunakan tabel klasifikasi (ketepatan klasifikasi) untuk mengetahui kesesuaian model. Jika model memprediksi keanggotaan grup secara akurat berdasarkan beberapa kriteria (variabel independen), kemudian ini dipikirkan untuk menyediakan bukti bahwa model sesuai (Hosmer & Lemeshow, 2000). Menurut Hosmer & Lemeshow (2000),

tabel klasifikasi paling sesuai ketika klasifikasi merupakan tujuan dari analisis, jika tidak maka hanya sebagai pendukung metode kesesuaian model yang lebih sulit.

Tabel klasifikasi merupakan sebuah cara untuk merangkum hasil model regresi logistik. Tabel ini dihasilkan dari klasifikasi silang variabel dependen, y, dengan variabel dikotomi yang diperoleh dari estimasi peluang. Untuk memperolehnya, kita tentukan *cutpoint*, c, dan membandingkan setiap nilai peluang dengan c. Jika nilai estimasi lebih dari c maka y dikategorikan menjadi 1; jika tidak maka dikategorikan menjadi 0. Nilai c yang paling sering digunakan adalah 0,5 (Hosmer & Lemeshow, 2000).

2. 3 Metode CHAID

Metode CHAID (*Chi-square Automatic Interaction Detection*) diperkenalkan oleh Dr. G. V. Kass pada tahun 1980, melalui sebuah artikel yang berjudul "*An Exploratory Technique for Investigating Large Quantities of Categorical Data*". Metode CHAID merupakan pengembangan dari metode yang sudah ada sebelumnya, yaitu *Automatic Interaction Detection* (AID). CHAID adalah sebuah analisis berdasarkan variabel kategori (Perez & Cejas, 2016). 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 uji *chi-square* terhadap variabel dependen.

CHAID digunakan untuk membentuk segmentasi yang membagi data menjadi dua atau lebih kelompok yang berbeda berdasarkan sebuah kriteria (variabel independen). Pada setiap tahap, CHAID memilih variabel independen yang mempunyai interaksi paling kuat dengan variabel dependen. kategori dari

setiap variabel independen digabungkan jika mereka tidak signifikan berbeda terhadap variabel dependen (Cinca & Nieto, 2016). Hal ini kemudian diteruskan dengan membagi kelompok-kelompok tersebut menjadi kelompok yang lebih kecil berdasarkan variabel independen yang lain. Proses tersebut terus berlanjut sampai tidak ditemukan lagi variabel independen yang signifikan secara statistik (Kunto & Hasana, 2006).

2. 3. 1 Variabel-variabel Metode CHAID

Variabel yang digunakan dalam metode CHAID adalah data kategori (nominal atau ordinal), baik variabel dependen maupun variabel independen. Menurut Gallagher (2000), Variabel independen dalam metode CHAID dapat dibedakan menjadi 3 jenis. Variabel-variabel tersebut adalah sebagai berikut.

a. Variabel Monotonik

   Variabel monotonik adalah variabel independen di mana kategori-kategori di dalamnya dapat digabungkan jika berurutan (data ordinal).

b. Variabel Bebas

   Variabel bebas adalah variabel independen di mana kategori-kategori di dalamnya dapat digabungkan meskipun tidak berurutan (data nominal).

c. Variabel Mengambang

   Variabel mengambang adalah variabel independen yang dapat diperlakukan sebagai variabel monotonik, kecuali untuk kategori yang *missing value*, yang dapat dikombinasikan dengan kategori manapun.

2. 3. 2 Uji *Chi-square*

Sesuai dengan namanya, statistik uji yang digunakan dalam metode CHAID adalah statistik uji *chi-square*. Statistik uji *chi-square* dapat digunakan untuk mengetahui independensi (kebebasan) antara dua variabel.

Misalkan dua variabel akan diuji independensinya, yang mana variabel pertama mempunyai r kategori dan variabel kedua mempunyai c kategori. Maka struktur data uji *chi-square* dapat dilihat pada Tabel 2.2 (Daniel, 1989).

Tabel 2.2 Struktur Data Uji *Chi-square*

Baris	Kolom						Total
	1	2	...	j	...	c	
1	n_{11}	n_{12}	...	n_{1j}	...	n_{1c}	n_1
2	n_{21}	n_{22}	...	n_{2j}	...	n_{2c}	n_2
.	.	.	...	.	...	.	.
i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ic}	$n_{i.}$
.	.	.	...	.	...	.	.
r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rc}	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.c}$	N

Keterangan

n_{ij} : banyaknya pengamatan yang termasuk dalam kategori ke-i dari variabel pertama dan kategori ke-j dari variabel kedua

$n_{i.}$: banyaknya pengamatan yang termasuk dalam kategori ke-i dari variabel pertama

$n_{.j}$: banyaknya pengamatan yang termasuk dalam kategori ke-j dari variabel kedua

Hipotesis yang digunakan pada pengujian *chi-square* adalah 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)

Taraf signifikansi : α

Statistik Uji

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(n_{ij} - E_{ij}\right)^2}{E_{ij}} \qquad (2.9)$$

Keterangan

n_{ij} : banyaknya pengamatan yang termasuk dalam kategori ke-i dari variabel pertama dan kategori ke-j dari variabel kedua

E_{ij} : frekuensi harapan pengamatan yang termasuk dalam kategori ke-i dari variabel pertama dan kategori ke-j dari variabel kedua

r : jumlah kategori dalam variabel pertama

c : jumlah kategori dalam variabel kedua

Untuk menghitung frekuensi harapan masing-masing sel digunakan rumus (Daniel, 1989).

$$E_{ij} = \frac{n_{i.} \cdot n_{.j}}{n} \qquad (2.10)$$

Keterangan

$n_{i.}$: banyaknya pengamatan yang termasuk dalam kategori ke-i dari variabel pertama

$n_{.j}$: banyaknya pengamatan yang termasuk dalam kategori ke-j dari variabel kedua

E_{ij} : frekuensi harapan pengamatan yang termasuk dalam kategori ke-i dari variabel pertama dan kategori ke-j dari variabel kedua

n : banyaknya seluruh pengamatan

Menurut Daniel (1989), kriteria pengambilan keputusan dalam uji *chi-square* yaitu H_0 ditolak jika $\chi^2_{hitung} > \chi^2_{\alpha;(r-1)(c-1)}$ atau dengan membandingkan nilai signifikansi dengan taraf signifikansi (α).

Statistik uji *chi-square* digunakan dalam dua cara dalam analisis CHAID. Pertama, untuk menentukan apakah kategori-kategori dalam sebuah variabel independen bersifat seragam dan bisa digabungkan menjadi satu. Kedua, ketika semua variabel independen sudah diringkas menjadi bentuk yang signifikan dan tidak mungkin digabung lagi, maka statistik uji *chi-square* digunakan untuk menentukan variabel independen mana yang paling signifikan untuk membagi kategori-kategori dalam variabel dependen.

2. 3. 3 Koreksi Bonferroni

Menurut Sharp *et al* (2002), koreksi Bonferroni adalah suatu proses koreksi yang digunakan ketika beberapa uji statistik untuk kebebasan atau ketidakbebasan dilakukan secara bersamaan. Koreksi Bonferroni biasanya digunakan dalam perbandingan berganda.

Pengurangan pada tabel kontingensi pada algoritma CHAID dibutuhkan untuk uji signifikansi. Jika tidak ada pengurangan pada tabel kontingensi asal, maka statistik uji χ^2 dapat digunakan. Ketika terjadi pengurangan yaitu c kategori dari variabel asal menjadi r kategori (r < c), maka tingkat kesalahan tunggal untuk uji signifikansi antara variabel dependen dan variabel independen yang tereduksi tersebut dikalikan dengan pengali Bonferroni sesuai dengan jenis variabelnya.

Kass (1980) menyebutkan bahwa pengali Bonferroni dihitung sesuai dengan jenis variabel independen.

1. Variabel Monotonik

$$M = \binom{c-1}{r-1} \qquad (2.11)$$

Keterangan

M : pengali Bonferroni

c : banyaknya kategori variabel independen awal

r : banyaknya kategori variabel independen setelah penggabungan

2. Variabel Bebas

$$M = \sum_{i=0}^{r-1} (-1)^i \frac{(r-i)^c}{i!\,(r-i)!} \qquad (2.12)$$

3. Variabel Mengambang

$$M = \binom{c-2}{r-2} + r\binom{c-2}{r-2} \qquad (2.13)$$

2. 3. 4 Algoritma CHAID

Algoritma CHAID digunakan untuk menghasilkan diagram pohon CHAID yang dapat digunakan untuk memprediksi nilai variabel dependen. Secara garis besar algoritma ini dapat dibagi menjadi 3 tahap, yaitu tahap penggabungan (*merging*), tahap pemisahan (*splitting*), dan tahap penghentian (*stopping*). Diagram pohon diperoleh melalui tiga tahap tersebut, dimulai dari simpul akar dan dilakukan secara berulang pada setiap simpul yang terbentuk.

2. 3. 4. 1 Tahap Penggabungan (Merging)

Pada tahap ini diperiksa signifikansi dari masing-masing kategori variabel independen terhadap variabel dependen. Tahap penggabungan untuk setiap variabel

independen dalam menggabungkan kategori-kategori yang tidak signifikan adalah sebagai berikut.

1. Membentuk tabel kontingensi dua arah untuk masing-maisng variabel independen dengan variabel dependen.
2. Menghitung statistik uji *chi-square* untuk setiap pasang kategori yang dapat dipilih untuk digabung menjadi satu, untuk menguji kebebasannya dalam sebuah sub tabel kontingensi 2 x d yang dibentuk oleh sepasang kategori tersebut dengan variabel dependen yang mempunyai sebanyak d kategori.

   Misalnya, sebuah variabel independen X_i adalah variabel monotonik dengan c kategori, di mana i = 1,2,...,c. Variabel dependen Y memiliki r kategori. Untuk mengetahui kategori variabel independen mana yang tidak signifikan, maka dipasangkan masing-masing kategori pada variabel independen dengan variabel dependen. Banyaknya pasangan yang mungkin adalah kombinasi r dari c.

Tabel 2.3 Ilustrasi Penggabungan Pasangan Kategori Variabel Independen

Kategori 1	Kategori 2	*p-value*
X_1	X_2	$P_{1,2}$
X_1	X_3	$P_{1,3}$
.	.	.
.	.	.
X_c	X_1	$P_{c,1}$
.	.	.
.	.	.
X_c	X_{c-1}	$P_{c,c-1}$

3. Untuk masing-masing nilai *chi-square* berpasangan, hitung *p-value* berpasangan bersamaan. 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 nomor 4.

Dari ilustrasi Tabel 2.3, jika terdapat pasangan dengan *p-value* lebih besar dari taraf signifikasni, maka pasangan tersebut akan digabungkan. Misalnya pasangan kategori X_1 dan X_2 pada Tabel 2.3 tidak signifikan, maka pasangan tersebut akan digabungkan menjadi satu variabel baru yaitu $X_{1,2}$.

4. Periksa kembali kesignifikanan 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.

   Misalnya, pada ilustrasi sebelumnya didapat gabungan variabel baru $X_{1,2}$. Variabel tersebut akan dipasangkan dengan variabel lainnya, misalnya X_3, X_4,...,X_5 kemudian dilihat apakah pasangan tersebut sudah signifikan, ketika semua signifikan bisa dilanjutkan ke langkah 5, namun jika masih ada yang belum signifikan kembali ke langkah 3.

5. Hitung *p-value* terkoreksi Bonferroni didasarkan pada tabel yang telah digabung.

2. 3. 4. 2 Tahap Pemisahan (Splitting)

Tahap pemisahan memilih variabel independen yang mana yang akan digunakan sebagai pemisah simpul terbaik. Pemilihan dikerjakan dengan membandingkan *p-value* (dari tahap penggabungan) pada setiap variabel independen. Langkah tahap pemisahan adalah sebagai berikut.

1. Pilih variabel independen yang memiliki *p-value* terkecil (paling signifikan) yang akan digunakan sebagai pemisah simpul.

2. Jika *p-value* kurang dari atau sama dengan taraf signifikansi (α), pemisah simpul menggunakan variabel independen ini.

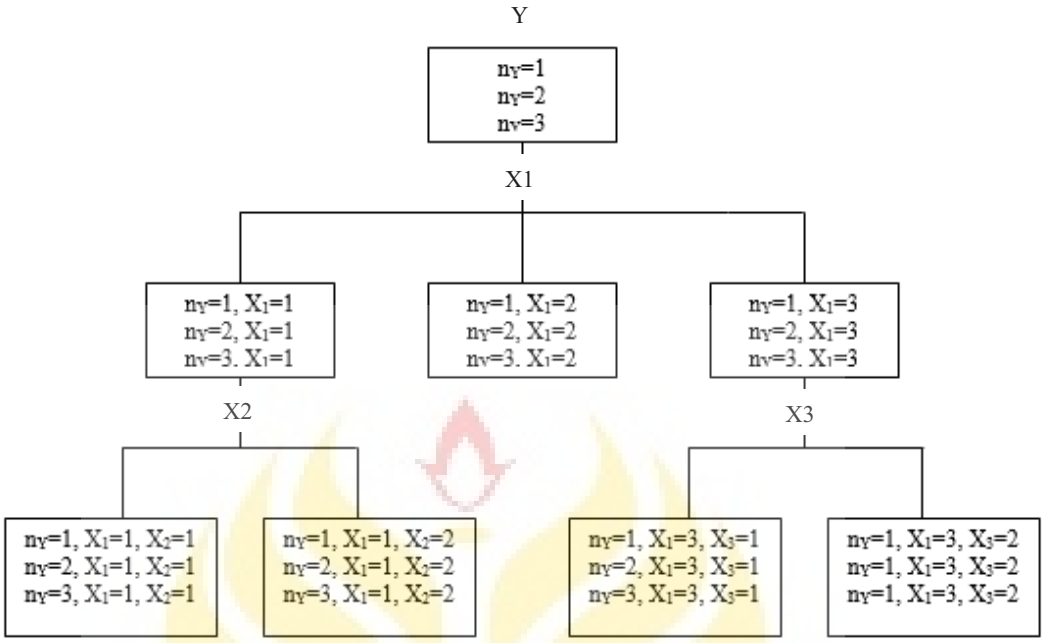
2. 3. 4. 3 Tahap Penghentian (Stopping)

Ulangi langkah penggabungan untuk subkelompok berikutnya, tahap penghentian dilakukan jika proses pertumbuhan pohon harus dihentikan sesuai dengan peraturan penghentian di bawah ini.

1. Tidak ada lagi variabel independen yang signifikan menunjukkan perbedaan terhadap variabel dependen.
2. Jika pohon sekarang mencapai batas nilai maksimum pohon dari spesifikasi maka proses pertumbuhan pohon akan berhenti. Misalnya, ditetapkan kedalaman pertumbuhan pohon klasifikasi adalah 3, ketika pertumbuhan pohon sudah mencapai kedalaman 3 maka pertumbuhan pohon klasifikasi dihentikan.
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. Misalnya, ditetapkan ukuran minimum simpul anak adalah 50, ketika pemisahan menghasilkan ukuran simpul anak kurang dari 50, maka simpul tidak akan dipecah.

2. 3. 5 Pohon Klasifikasi CHAID

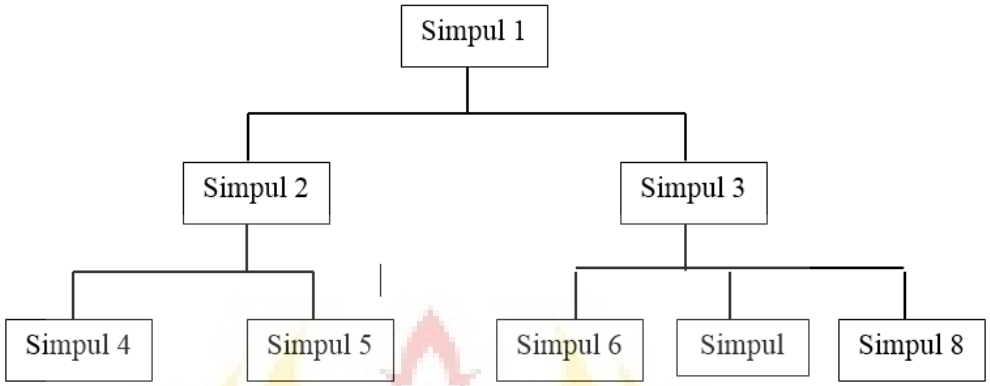
Hasil proses pembelahan dalam CHAID akan ditampilkan dalam sebuah diagram pohon. Secara umum ilustrasi diagram pohon CHAID dengan contoh kasus jumlah kategori variabel terikatnya n = 3 dapat dilihat pada Gambar 2.1.



Gambar 2.1 Diagram Pohon CHAID

Hasil proses pembelahan dalam CHAID akan ditampilkan dalam sebuah diagram pohon. Myers (1996, dalam Lehman & Eherler, 2001) mengemukakan bahwa hasil pembentukan segmen dalam CHAID akan ditampilkan dalam sebuah diagram pohon yang mengikuti aturan "dari atas ke bawah" (*top-down stopping rule*). Diagram pohon disusun mulai dari kelompok induk, berlanjut di bawahnya sub kelompok yang berturut-turut dari hasil pembagian kelompok induk berdasarkan kriteria tertentu. Tiap-tiap *node* dari diagram pohon ini menggambarkan sub kelompok dari sampel yang diteliti. Setiap *node* akan berisi keseluruhan sampel dan frekuensi absolut n_i untuk tiap kategori yang disusun di atasnya. Pada pohon klasifikasi CHAID terdapat istilah kedalaman (*depth*) yang berarti banyaknya tingkatan *node* sub kelompok sampai ke bawah pada *node* sub kelompok yang terakhir (Lehmann & Eherler, 2001).

Bagian-bagian dari diagram pohon dapat dilihat pada Gambar 2.2.



Gambar 2.2 Bagian Diagram Pohon CHAID

1. *Root node* (simpul akar) adalah sampel yang mengandung data seluruh sampel.
2. *Child node* (simpul anak) adalah simpul yang dihasilkan dari pembelahan suatu simpul yang didasarkan pada suatu variabel prediktor paling signifikan dengan kategori-kategori paling signifikan (kombinasi kategorik yang paling signifikan).
3. *Parent node* (simpul induk) adalah simpul yang dibelah berdasarkan suatu variabel prediktor, sehingga menghasilkan beberapa simpul anak.
4. *Terminal node* (simpul terminal/simpul akhir) adalah simpul yang tidak dapat dibelah lagi karena tidak ada variabel prediktor yang signifikan dalam membelah simpul tersebut. Simpul terminal mengakhiri pertumbuhan cabang pohon CHAID.

Pembelahan pada CHAID dimulai dari simpul akar menjadi beberapa simpul anak berdasarkan variabel independen yang paling signifikan dengan kategori-kategori yang signifikan. Masing-masing simpul anak yang dihasilkan dari pembelahan diperiksa secara terpisah untuk mengetahui apakah suatu simpul anak

dapat dibelah atau tidak. Pembelahan simpul anak menggunakan salah satu variabel independen paling signifikan dengan kategori-kategori signifikan yang dipilih dari variabel sisa. Proses tersebut berlanjut dengan pembelahan berurutan untuk simpul anak yang diperoleh dari tahap sebelumnya sampai dengan tidak ada lagi simpul anak yang dapat dibelah.

Pembelahan akan berhenti apabila memenuhi kriteria *stopping rule* yang ditentukan. Jika variabel dependen merupakan variabel kategorik maka variabel independen yang membelah suatu simpul akar atau simpul induk menjadi simpul anak dipilih berdasarkan uji *chi-square*.

2. 3. 6 Pelabelan Kelas

Salah satu hal yang perlu diperhatikan ketika diagram pohon sudah terbentuk adalah 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. Persentase ini merupakan jumlah kasus yang ada untuk setiap kategori pada simpul akhir tersebut dibandingkan dengan jumlah total responden pada simpul akhir terkait. Misalnya, simpul akhir ke-1 memiliki variabel dependen dengan 2 kategori yaitu A dan B. Hasil diagram pohon menunjukkan bahwa persentase untuk kategori A sebesar 55% sedangkan kategori B sebesar 45% sehingga dapat disimpulkan untuk simpul akhir ke-1 masuk ke dalam kelas kategori A.

Setiap simpul akhir yang ada akan mengalami proses pelabelan kelas. Ketika proses pelabelan kelas sudah selesai maka dapat dihitung tingkat ketepatan pohon klasifikasi yang terbentuk.

2. 4 Ketepatan Klasifikasi

Terdapat sebuah ukuran mengenai kemampuan yang tidak bergantung pada bentuk populasi awal dan dapat dihitung untuk berbagai prosedur klasifikasi. Ukuran ini dinamakan *apparent error rate* (APER), didefinisikan sebagai bagian dari observasi yang salah diklasifikasikan (*misclassified*) oleh fungsi klasifikasi. Apparent error rate (APER) dapat dihitung dengan mudah dari matriks konfusi yang menunjukkan keanggotaan observasi dan prediksi (Johnson & Winchern, 2007). Matriks konfusi dapat dilihat pada Tabel 2.4.

Tabel 2.4 Matriks Konfusi

Observasi	Prediksi	
	y_1	y_2
y_1	n_{11}	n_{12}
y_2	n_{21}	n_{22}

Keterangan

n_{11} : jumlah subjek dari y_1 tepat diklasifikasikan sebagai y_1

n_{12} : jumlah subjek dari y_1 salah diklasifikasikan sebagai y_2

n_{21} : jumlah subjek dari y_2 tepat diklasifikasikan sebagai y_2

n_{22} : jumlah subjek dari y_2 tepat diklasifikasikan sebagai y_2

Maka nilai APER dapat dihitung dengan rumus sebagai berikut.

$$APER = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}} \qquad (2.14)$$

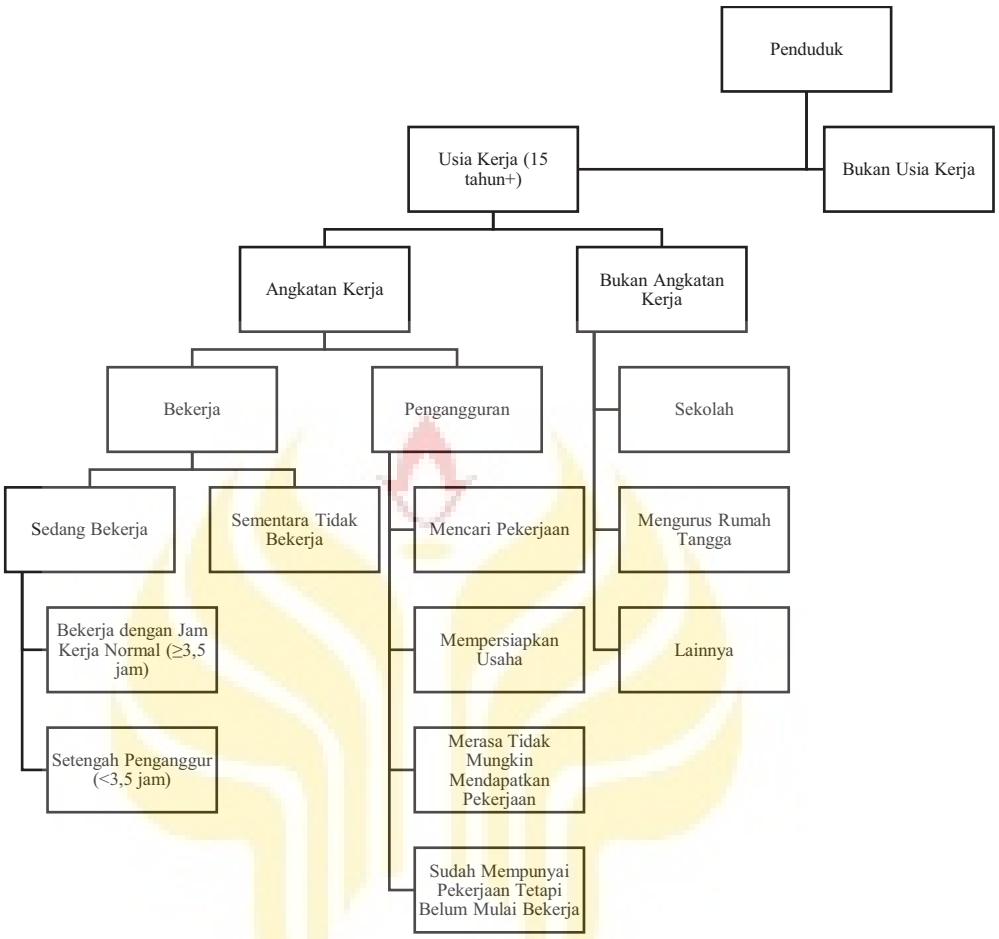
2. 5 Definisi Variabel Penelitian

Variabel merupakan sesuatu yang mempunyai nilai yang bervariasi. Pada penelitian ini, variabel penelitian terdiri dari 1 variabel dependen dan 8 variabel independen. Variabel-variabel tersebut adalah sebagai berikut.

a. Status Angkatan Kerja (Y)

Pada penelitian ini digunakan data hasil Survei Angkatan Kerja Nasional (Sakernas) Kabupaten Temanggung tahun 2015. Survei ini dilaksanakan oleh Badan Pusat Statistik (BPS) Kabupaten Temanggung. Konsep/definisi ketenagakerjaan yang digunakan oleh BPS merujuk pada rekomendasi *International Labour Organization* (ILO). Konsep dasar angkatan kerja (*Standard Labour Force Concept*) dapat dilihat pada Gambar 2.3. Penduduk dibagi menjadi usia kerja dan bukan usia kerja. Penduduk usia kerja dibagi menjadi angkatan kerja dan bukan angkatan kerja. Angkatan kerja dibagi menjadi bekerja dan pengangguran, sedangkan bukan angkatan kerja dibagi menjadi sekolah, mengurus rumah tangga, dan lainnya.

Variabel status angkatan kerja mempunyai 2 kategori, yaitu bekerja dan pengangguran. Pada data observasi, kriteria penentuan seseorang sebagai bekerja atau pengangguran ditentukan dengan kriteria-kriteria tertentu oleh BPS yang dapat dilihat pada Gambar 2.3.



Gambar 2.3 Konsep Dasar Angkatan Kerja

b. Klasifikasi Desa/Kelurahan (X1)

Dalam Sakernas Kabupaten Temanggung tahun 2015, desa/kelurahan dikelompokkan dalam daerah perkotaan dan pedesaan. Maka variabel klasifikasi desa/kelurahan mempunyai 2 kategori, yaitu perkotaan dan pedesaan.

c. Hubungan dengan Kepala Rumah Tangga (X2)

Variabel hubungan dengan kepala rumah tangga mempunyai 2 kategori, yaitu kepala rumah tangga dan bukan kepala rumah tangga. Kategori kepala rumah tangga adalah anggota rumah tangga yang bertanggung jawab atas kebutuhan sehari-hari dalam rumah tangga tersebut atau orang yang dianggap/ditunjuk

menjadi kepala rumah tangga. Sedangkan kategori bukan kepala rumah tangga adalah istri/suami, anak, menantu, cucu, orang tua/mertua, famili lainnya, pembantu rumah tangga, atau lainnya dari kepala rumah tangga.

d. Jenis Kelamin (X3)

Variabel jenis kelamin mempunyai 2 kategori, yaitu laki-laki dan perempuan.

e. Umur (X4)

Variabel umur mempunyai 3 kategori, yaitu 15 – 24 tahun, 25 – 54 tahun, dan ≥ 55 tahun.

f. Status Pernikahan (X5)

Variabel status pernikahan mempunyai 2 kategori, yaitu menikah dan tidak menikah. Kategori menikah adalah yang berada dalam status pernikahan, sedangkan kategori tidak menikah adalah yang berstatus belum menikah, cerai hidup, atau cerai mati.

g. Pendidikan (X6)

Variabel pendidikan terdiri dari 5 kategori, yaitu ≤ SD sederajat, SLTP sederajat, SLTA sederajat, DI – DIII, dan ≥ S1.

h. Pelatihan Kerja (X7)

Pelatihan kerja merupakan pendidikan/pelatihan yang diselenggarakan oleh pemerintah/swasta yang memberikan keterampilan khusus pada batas waktu tertentu dan peserta pendidikan/pelatihan tersebut memperoleh tanda lulus/sertifikat. Variabel pelatihan kerja mempunyai 2 kategori, yaitu mendapatkan pelatihan kerja (Ya) dan tidak mendapatkan pelatihan kerja (Tidak).

i. Pengalaman Kerja (X8)

Pengalaman kerja merupakan usaha/pekerjaan yang pernah dipunyai sebelumnya, sebelum akhirnya berhenti karena sesuatu hal. Variabel pengalaman kerja mempunyai 2 kategori, yaitu mempunyai pengalaman kerja (Ya) dan tidak mempunyai pengalaman kerja (Tidak).

BAB V

PENUTUP

5. 1 Simpulan

Berdasarkan penelitian ini dapat diperoleh beberapa kesimpulan sebagai berikut.

1. Ketepatan metode regresi logistik dengan pembobotan sampel pada klasifikasi status angkatan kerja Kabupaten Temanggung tahun 2015 adalah 96,4%.
2. Ketepatan metode CHAID dengan pembobotan sampel pada klasifikasi status angkatan kerja Kabupaten Temanggung tahun 2015 adalah 96,6%.
3. Ketepatan metode regresi logistik dan CHAID dengan pembobotan sampel hampir sama, akan tetapi metode CHAID mempunyai ketepatan klasifikasi yang lebih tinggi.

5. 2 Saran

Berdasarkan penelitian ini penulis dapat memberi beberapa saran, yaitu (1) metode CHAID dengan pembobotan sampel dapat dimanfaatkan oleh instansi terkait guna membantu klasifikasi data sehingga pekerjaan akan semakin efektif dan efisien, (2) metode ini dapat diaplikasikan pada data yang lain karena kemudahannya yang tidak membutuhkan asumsi tertentu, (3) apabila data dan variabel dalam jumlah yang besar, maka sebaiknya menggunakan bantuan *software* agar efektif; jika tidak maka dapat dilakukan perhitungan secara manual.

DAFTAR PUSTAKA

Alghofari, F. 2011. *Analisis Tingkat Pengangguran di Indonesia Tahun 1980-2007*. Skripsi. Semarang : Fakultas Ekonomi dan Bisnis UNDIP.

Antipov, E. & E. Pokryshevskaya. 2009. *Applying CHAID for logistic regression diagnostics and classification accuracy improvement*. Munich Personal RePEc Archieve (MPRA) No. 21499. Munich : Ludwig Maximilians Universitat Munchen.

Badan Pusat Statistik Kabupaten Temanggung. 2016. Profil Ketenagakerjaan Kabupaten Temanggung 2015. Temanggung : Badan Pusat Statistik Kabupaten Temanggung.

Cha, G.W., Y.C. Kim., H.J. Moon. & W.H. Hong. 2017. New Approach for forecasting demolition waste generation using chi-squared automatic interaction detection (CHAID) method. *Journal of Cleaner Production* 168: 375-385.

Cinca, C. S., & B. G. Nieto. 2016. The Use of Profit Scoring As An Alternative To Credit Scoring System In Peer-To-Peer (P2P) Lending. *Decision Support System* (2016).

Daniel, Wayne, W. 1989. *Statistik Nonparametrik Terapan*. Jakarta : PT Gramedia.

Dewi, A. M. C. 2010. *Analisis Tingkat Pengangguran dan Faktor-Faktor Yang Mempengaruhinya di Kota Semarang*. Skripsi. Semarang : Fakultas Ekonomi dan Bisnis UNDIP.

Gallagher, C.A., H. M. Monroe, & J.L. Fish. 2000. *An Iterative Approach to Classification Analysis*. Tersedia di https://www.casact.org/pubs/dpp/dpp90/90dpp237.pdf [diakses 10-2-2017].

Graubard, B. I., E. L. Korn. & D. Midthune. Teting Goodness of Fit for Logistic Regression with Survey Data. Prosiding American Statostical Association. ASA : Amerika.

Hosmer, D. W. & S. Lemeshow. 2000. *Applied Logistic Regression*. New York : John Wiley & Sons Inc.

Hosmer, D. W. & S. Lemeshow. 1989. *Applied Logistic Regression*. New York : John Wiley & Sons Inc.

Imaslihkah, S., M. Ratna. & V. Ratnasari. 2013. Analisis Regresi Logistik Ordinal terhadap Faktor-faktor yang Mempengaruhi Predikat Kelulusan Mahasiswa S1 di ITS Surabaya. *JURNAL SAINS DAN SENI POMITS* 2(2) : 177-182.

Johnson, R. A. dan Wichern, D. W. 2007. Applied Multivariate Statistical Analysis, 6th Edition. New Jersey: Person Prentice Hall.

Josephat, P. & A. Ismail. 2012. A Logistic Regression Model of Customer Satisfaction of Airline. *International Journal of Human Resource Studies* 2(2012) : 255-265.

Kass, G.V. 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics* 29(2), 119-127.

Kleinbaum, D.G. & M. Klein. 2010. *Logistic Regression A Self Learning Text (3rd)*. New York : Springer Science .

Koskas, M. 2015. Direct Comparison of Logistic Regression and Recursive Partitioning to Predict Lymph Node Metasets in Endometrial Cancer. *International Journal of Gynecological* 25(2015) : 1037-1043.

Kunto, Y.S. dan Hasana, S.N. 2006. Analisis CHAID sebagai Alat Bantu Statistika untuk Segmentasi Pasar. *Jurnal Manajemen*, vol. 1 No. 2. Surabaya : Universitas Kristen Petra.

Lehmann, T., dan Eherler, D. 2001. Responder Profilling with CHAID and Dependency Analysis. Tersedia di http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.8533&rep=rep1&type=pdf. [diakses pada 25-02-2017].

Lestari, Sri Puji. 2017. Pemilihan Model Regresi Linier Berganda Terbaik Pada Kasus Multikolinearitas Berdasarkan Metode Principal Component Analysis (PCA) dan Metode Stepwise. *Unnes Journal of Mathematics* 6(1)(2017).

OECD. 2016. *Survei Ekonomi OECD : Indonesia 2016*. Online. Tersedia di https://www.oecd.org/eco/surveys/indonesia-2016-OECD-economic-survey-overview-bahasa.pdf [diakses 11-02-2017].

Perez, F.M.D., & M. B. Cejas. 2016. CHAID Algorithm As An Appropriate Analytical Method for Tourism Market Segmentation. *Journal of Destination Marketing & Management* 5(2016) : 275-282.

Rahayu, R. S. 2015. Identifikasi Faktor-Faktor Yang Mempengaruhui Terjadinya Preeklampsia dengan Metode CHAID. Skripsi. Semarang : FSM UNDIP.

Ritschard, G. 2010. *CHAID and Earlier Supervised Tree Methods*. Geneva : Universitas Geneva.

Sharp, A., J. Romaniuk dan S. Cierpicki. 2002. The Performance of Segmentation Variables : A Comparative Study. Tersedia di http://130.195.95.71:8081/www/ANZMACI1998/Cd_rom/Sharp222.pdf [diakses pada 20-02-2017].

Suyanti. 2014. Deteksi Outlier Menggunakan Diagnosa Regresi Berbasis Estimator Parameter Robust. *Unnes Journal Mathematics* 3(2)(2014).

Swain, Ajaya K. 2016. Mining Big Data To Support Decision Making In Healthcare. *Journal Of Information Technology Case And Application Research* 18:3, 141 – 154.