KLASIFIKASI DATA KATEGORIK DENGAN *MISSING VALUE* MENGGUNAKAN ALGORITMA CHAID

(*Chi-square Automatic Interaction Detection*)

Skripsi
Disusun untuk melengkapi syarat-syarat guna memperoleh
gelar Sarjana Sains

TRI ASIH KUSUMASTUTI
3125051626

PROGRAM STUDI MATEMATIKA

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI JAKARTA

2011

LEMBAR PENGESAHAN

Nama	:	Tri Asih Kusumastuti
No. Registrasi	:	3125051626
Jurusan	:	Matematika
Judul	:	KLASIFIKASI DATA KATEGORIK DENGAN *MISSING VALUE* MENGGUNAKAN ALGORITMA CHAID

Skripsi ini disusun sebagai syarat untuk mendapatkan gelar sarjana sains

Menyetujui,

Dosen Pembimbing I	Dosen Pembimbing II
Dra.Widyanti Rahayu, M.Si	**Ratna Widyati, S.Si, M.Kom**
NIP: 19661103 200112 2 001	NIP: 19750925 200212 2 002

Mengetahui,

Ketua Jurusan Matematika FMIPA UNJ

Dra. Pinta Deniyanti S, M.Si

NIP : 19640731 199102 2 001

PERSETUJUAN PANITIA UJIAN SKRIPSI

KLASIFIKASI DATA KATEGORIK DENGAN MISSING VALUE
MENGGUNAKAN ALGORITMA CHAID

Nama : Tri Asih Kusumastuti

No. Registrasi : 3125051626

		Nama	Tanda Tangan	Tanggal
Penanggung Jawab				
Dekan	:	Dra. Marheni, M.Sc.	.....	.....
		NIP. 19500606 197412 2 001		
Wakil Penanggung Jawab				
Pembantu Dekan I	:	Dr. rer. Aprilia, L. F, MS., MEd.	.....	.....
		NIP. 19600408 199003 2 002		
Ketua	:	Ir. fariani Hermin, MT.	.....	.....
		NIP. 19600211 198703 2 001		
Sekretaris	:	Drs. Bambang Irawan, M.Si.	.....	.....
		NIP. 19681201 200112 1 001		
Penguji	:	Drs. Sudarwanto, M.Si,DEA	.....	.....
		NIP. 19650325 199303 1 003		
Pembimbing I	:	Dra. Widyanti Rahayu, M.Si	.....	.....
		NIP.19661103 200112 2 001		
Pembimbing II	:	Ratna Widyati, S.Si, M.Kom	.....	.....
		NIP.19740415 199803 2 002		

Dinyatakan lulus ujian skripsi tanggal: 29 Juli 2011

PERSEMBAHANKU

Untuk Ibunda tercinta yang telah melahirkan
membesarkan dengan penuh kesabaran,
dan limpahan kasih sayang,
Bunda, skripsi ini aku persembahkan khusus untuk mu;

Untuk Ayahanda tersayang yang telah mencucurkan keringatnya; demi ananda dan yang membesarkan dan memberikan kasih sayang luar biasa setelah bunda tiada
Ayahanda, skripsi ini Jiwa yang meyakini betapa besarnya karunia Allah
seperti yang telah disebutkan dalam Al-Qur'an
atau yang telah diwahyukan oleh-Nya;

Yang menyadari bahwa apa yang telah menimpanya
bukanlah untuk menyalahkan perbuatannya
sehingga dia pun tetap ridho
terhadap apa yang telah ditentukan oleh Allah
melalui ketentuan dan ketetapan-Nya;

Yang selalu berusaha sepanjang hidupnya
untuk melakukan hal-hal yang dapat mendatangkan
keridhoan Tuhannya, lalu dia tidak pernah mengharapkan
keridhoan dzat lain selain Allah;

Yang selalu berlindung dalam naungan
kabar gembira dari-Nya sehingga jiwanya selalu tenang,
kemudian dia akan naik ke sisi Tuhan Yang Maha Tinggi

dan yang menetap di taman-taman yang penuh
dengan kelembutan karena syahadat
yang telah dibacanya, lalu dia tidak pernah meminta
pertolongan, kecuali hanya kepada-Nya.

Wahai Ibuku yang sangat berarti bagiku

Semoga Allah menyayangimu dan menyatukanmu
denganku di bawah panji sang kekasih kita,
Nabi Muhammad SAW

Sesungguhnya dalam penciptaan langit dan bumi, Dan silih bergantinya malam dan siang terdapat tanda-tanda bagi orang-orang yang berakal, (yaitu) orang-orang yang mengingat Allah sambil berdiri atau duduk atau dalam keadaan berbaring dan mereka memikirkan tentang penciptaan langit dan bumi (seraya berkata): "Ya Tuhan kami, tiadalah Engkau menciptakan ini dengan sia-sia. Maha Suci Engkau, maka peliharalah kami dari siksa neraka."

(Ali Imran: 190-191).

.....Katakanlah: "Adakah sama orang-orang yang mengetahui dengan orang-orang yang tidak mengetahui? Sesungguhnya orang yang berakallah yang dapat menerima pelajaran." (Az-Zumar: 9).

ABSTRACT

Tri Asih Kusumastuti, 3125051626. Clasification of Categorical Data with Missing Value Using CHAID Algoritm. Thesis. Faculty of Mathematics and Natural Science Jakarta State University. 2011.

Discriminant and cluster analysis are familiar clasification method. Discriminant and cluster analysis need assumptions: norm distribution and interval independent variabels. In fact, independent variabels might be categorical. CHAID (chi square automatic interaction detection) is used for clasifiying data with categorical dependent variabel and does not need norm assumption. CHAID is available to data which has missing values. It is treated as single category and results homogen classes in tree chart which is used as imputation class. CHAID uses chi-square as significant test and bonferroni correction while the category reduced. Ilustration in this thesis uses data taken from UCI repository of machine learning database (http://archive.ics.uci.edu/ml/datasets/Mushroom). The purpose is clasifiying mushroom genus Agaricus and Lepiota to edible or poissonous mushroom.

Keyword: Categorical, missing value, CHAID.

ABSTRAK

Tri Asih Kusumastuti, 3125051626. Klasifikasi data kategorik dengan *missing value* menggunakan algoritma CHAID.Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Jakarta. 2011.

Analisis diskriminan dan analisis kelompok merupakan metode pengklasifikasian yang telah banyak digunakan. Pada analisis diskriminan dan analisis kelompok diperlukan beberapa asumsi yang harus dipenuhi, antara lain: data berdistribusi normal dan peubah bebasnya berskala interval. Pada kenyataannya tidak semua data berskala interval. CHAID (*Chi-square Automatic Interaction Detection*) merupakan metode klasifikasi dari *decision tree* yang dapat digunakan untuk data yang peubah bebasnya kategorik dan tidak memerlukan asumsi distribusi normal. Selain itu CHAID juga dapat digunakan pada data yang mempunyai *missing value*. CHAID memperlakukan *missing value* sebagai satu kategori tunggal dan menghasilkan diagram pohon dengan kelas-kelas yang homogen sehingga dapat digunakan sebagai kelas imputasi untuk kategori *missing value*. CHAID menggunakan uji *chi-square* sebagai uji signifikan dan koreksi bonferroni digunakan bila terjadi pengkategorian ulang. Ilustrasi pada skripsi ini menggunakan data yang diperoleh dari *UCI repository of machine learning database (http://archive.ics.uci.edu/ml/data sets/Mushroom)*. Tujuan dari penelitian ini adalah mengklasifikasikan jamur dari genus Agaricus dan Lepiota ke dalam kategori dapat dimakan atau beracun.

Kata kunci: Kategorik, *missing value*, CHAID.

KATA PENGANTAR

Bismillaahirrohmaanirrohiim

Alhamdulillahi rabbil'alamin. Puji syukur penulis panjatkan kepada Allah SWT yang telah memberikan segenap kekuatan dan kesehatan sehingga dapat menyelesaikan penyusunan skripsi yang berjudul "Klasifikasi data kategorik dengan (*Missing Value*) menggunakan algoritma CHAID". Skripsi ini ditulis guna memenuhi sebagian persyaratan untuk mendapatkan gelar Sarjana Sains. Shalawat dan salam semoga tetap dicurah limpahkan kepada Nabi Muhammad SAW.

Penulis menyadari dalam proses penulisan skripsi ini tidak lepas dari bantuan dan dorongan dari berbagai pihak. Untuk itu, pada kesempatan ini penulis ingin menyampaikan terima kasih kepada:

1. Ibu Dra. Widyanti Rahayu, M.Si selaku Dosen Pembimbing I yang telah berkenan meluangkan waktunya dalam memberikan bimbingan, saran, nasehat serta pengarahannya sehingga skripsi ini dapat terselesaikan dengan baik dan terarah.

2. Ibu Ratna Widyati, S.Si M.Kom selaku Dosen Pembimbing II yang telah berkenan meluangkan waktunya dalam memberikan bimbingan, saran, nasehat serta pengarahannya sehingga skripsi ini dapat menjadi lebih baik dan terarah.

3. Ibu Dra. Pinta Deniyanti S, M.Si selaku Ketua Jurusan Matematika.

4. Ibu Fariani Hermin, M.T selaku Ketua Program Studi Matematika.

5. Ibu Ratna Widyati, S.Si, M.Kom selaku Sekretaris Jurusan Matematika.

6. Ibu Dra. Widyanti Rahayu, M.Si selaku penasehat akademis penulis.

7. Ibunda dan Ayahanda tercinta, kakak, keponakan dan semua keluarga tercinta, yang tiada lelah dan bosan-bosannya memberikan doa, dorongan semangat, nasehat, serta bantuan secara moral maupun materil selama penulis menuntut ilmu di UNJ sampai terlaksananya penulisan skripsi ini.

8. Teman-teman seperjuangan Matematika Angkatan 2005, yaitu: Anggi S.Si, Nissa, Endah, Icha, Risma, Dewi S.Si, Dwi nita, Titin S.Si, Santi, Renti, Chiz, Retno S.Si, Dika, Pandu S.Si, Kiki S.Si, Mahe, Abdur, Budi, Pur, Nurman, Dano, Dede, Sjamsul, Rolas S.Si. Semoga persahabatan dan persaudaraan kita terjaga selamanya.

9. Teman seperjuangan meraih mimpi, Mb Iin dan Yosa. Tetap semangat.

10. Semua pihak yang turut berperan dalam penyusunan skripsi ini.

Demikianlah skripsi ini penulis susun, akhir kata penulis mohon maaf atas segala kesalahan. Semoga skripsi ini bermanfaat bagi kita semua.

Jakarta, Juli 2011

Tri Asih Kusumastuti

DAFTAR ISI

Daftar Gambar

Daftar TABEL

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Analisis multivariat merupakan metode statistik yang digunakan untuk menganalisis data dengan karakteristik lebih dari satu peubah bebas dan peubah terikat. Dengan menggunakan teknik ini dapat dianalisis pengaruh beberapa peubah terhadap peubah-peubah lainnya dalam waktu yang bersamaan, contohnya menganalisis pengaruh peubah kualitas produk, harga dan saluran distribusi terhadap kepuasan pelanggan. Contoh lain penelusuran karakteristik nasabah kredit macet dan kredit lancar melalui peubah pekerjaan, usia, pendidikan dan penghasilan. Analisis diskriminan dan analisis kelompok merupakan metode pengelompokan dalam analisis multivariat yang telah banyak digunakan. Analisis diskriminan digunakan jika peubah terikat hanya satu dan pengukurannya bersifat kategorik. Asumsi lain yang harus dipenuhi antara lain setiap peubah bebasnya berdistribusi normal dan berskala interval. Padahal dalam kenyataannya, tidak semua peubah bebas dari data berdistribusi normal dan berskala interval atau rasio. Metode pengelompokan terhadap peubah bebas yang kategorik sangat diperlukan. Apalagi jika jumlah amatan dan peubah bebasnya banyak maka diperlukan metode yang efisien untuk mengeksplorasi data.

Decision tree digunakan untuk melihat struktur keterkaitan antar peubah. *Decison tree* merupakan salah satu teknik *data mining* yang dapat digunakan

untuk melakukan klasifikasi terhadap sekumpulan objek. Hasil analisis metode ini berupa struktur pohon *dendogram* yang mudah diinterpretasikan. Umumnya terdapat lebih dari satu cara untuk mendeskripsikan data dengan struktur pohon dikarenakan ketergantungan antar peubah bebasnya. Untuk membentuk struktur pohon digunakan algoritma pembangkit antara lain: CART (*Clasification and Regression tree*), CHAID (*Chi-square Automatic Interaction Detection*), QUEST (*Quick, Unbiased, Efficient Statistical Tree*), FACT, C4.5 dan CRUISE (*Classification Rule with Unbiased Interaction Selection and Estimation*).

Metode CHAID (*Chi-square Automatic Interaction Detection*) merupakan salah satu metode *decision tree* yang digunakan untuk membagi sebuah sampel menjadi dua (biner) atau lebih dari dua (non-biner) kelompok yang berbeda berdasarkan uji *chi-square*. Berbeda dengan analisis diskriminan dan analisis kelompok, CHAID bersifat nonparametrik, nonlinear dan dapat menganalisis peubah bebas yang bersifat kategorik. Di dalam panduan penggunaan program *answeertree* dikatakan bahwa kelebihan lain dari CHAID dibanding analisis diskriminan adalah CHAID dapat menangani *missing value* dengan memperlakukannya sebagai peubah tersendiri. *Missing value* adalah data yang hilang atau tidak tersedia pada peubah tertentu. Hal ini dapat terjadi karena berbagai faktor antara lain proses input data yang tidak sempurna, kealpaan petugas survei atau responden yang memang tidak bersedia menjawab pertanyaan.

Bidang terapan yang menggunakan metode CHAID antara lain riset pemasaran (dalam hal segmentasi pasar), kedokteran (untuk diagnosis), ilmu komputer (untuk menyelidiki struktur data), botani (dalam hal klasifikasi), psikologi (teori pengambilan keputusan), dan linguistik.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, permasalahan yang akan diteliti pada skripsi ini adalah menentukan pengelompokan data yang peubah bebasnya bersifat kategorik.

1.3 Pembatasan Masalah

Dalam skripsi ini permasalahan akan dibatasi hanya menggunakan CHAID sebagai metode pengelompokan data yang peubah bebasnya kategorik dengan jumlah ≥ 500 dan mempunyai *missing value*.

1.4 Tujuan Penulisan

1. Menentukan peubah bebas dimulai dari yang paling signifikan hingga yang rendah nilai signifikansinya.
2. Menentukan kategori peubah bebas yang signifikan terhadap peubah tak bebas.
3. Menangani *missing value* menggunakan CHAID.

1.5 Manfaat Penulisan

1. Skripsi ini dapat dijadikan bahan acuan untuk penulisan selanjutnya yang berkaitan dengan metode klasifikasi berstruktur pohon .
2. Memberikan alternatif metode pada masalah pengklasifikasian data kategorik.

BAB II

LANDASAN TEORI

2.1 Data Kategorik

Dimisalkan sebuah himpunan data mempunyai beberapa peubah, yaitu $A_1, A_2, ..., A_m$. Setiap peubah A_i mempunyai domain nilai yang dinotasikan dengan $DOM(A_i)$. Domain peubah berhubungan dengan 2 tipe skala yaitu numerik dan kategorik. Sebuah domain numerik ditampilkan dengan nilai kontinu. Sebuah $DOM(A_j)$ didefinisikan sebagai kategorik jika domain ini berhingga, untuk $a, b \in DOM(A_j), a = b$ atau $a \neq b$.

Data kategorik juga dikenal sebagai data kualitatif *multi-state*. Jenis data kategorik mempunyai beberapa karakteristik yaitu:

1. Misal sebuah himpunan data D terdiri dari N objek, yang didefinisikan dengan d peubah kategorik dan A_k adalah peubah kategori ke-k dimana A_k mempunyai nilai n_k.

2. $f_k(x)$: frekuensi nilai x dari himpunan data D pada peubah A_k.

3. $\hat{p}_k(x)$: Peluang peubah A_k bernilai x pada himpunan data D.

$$\hat{p}_k(x) = \frac{f_k(x)}{N} \tag{2.1}$$

4. Distribusi $f_k(x)$ adalah distribusi frekuensi nilai x pada sebuah peubah.

Data kategorik adalah jenis data yang bukan berupa besaran numerik namun berupa kategori atau level. Karena pada statistik untuk menganalisis menggunakan teknik dan rumus matematis, maka apabila data kategorik akan diolah dengan menggunakan teknik statistik maka data tersebut harus dibuat menjadi data kuantitatif. Data kategorik terdiri atas dua jenis skala pengukuran, yaitu :

1. Skala Nominal

   Skala ini paling banyak digunakan dalam penelitian ilmu-ilmu sosial. Skala nominal adalah pemberian skala di mana skala digunakan hanya untuk membedakan suatu ukuran dari ukuran yang lain tanpa memberi atribut lebih besar atau lebih kecil. Jadi sifat skala ini adalah sejajar atau sama antara masing-masing skala. Contoh skala ini adalah pemberian skala 1 untuk jenis kelamin laki-laki dan 2 untuk jenis kelamin perempuan. Jika melihat hasil tersebut maka skala 2 tidak lebih baik dibandingkan dengan skala 1 karena kedudukannya sejajar. Angka 1 dan 2 di sini hanya berfungsi untuk membedakan antara skala untuk laki-laki dan skala untuk perempuan. Contoh lain adalah pemberian skala untuk agama yang dianut, yang tidak memberikan tanda lebih baik atau buruk.

2. Skala Ordinal

   Jika pada skala nominal tidak dibedakan urutan dan tinggi rendahnya skala tetapi hanya membedakan suatu kategori dari kategori yang lain, maka pada skala ordinal dapat dibedakan urutan dari skala. Skala ini lebih baik daripada skala nominal karena memberikan nilai lebih besar atau lebih kecil, tetapi tidak dapat dicari selisih atau perbedaan antar skala. Contoh pemberian skala ini adalah pada kuesioner di mana pendapat sangat setuju

diberi skala 5, setuju diberi skala 4, ragu-ragu diberi skala 3, tidak setuju diberi skala 2 dan sangat tidak setuju diberi angka 1. Pada skala ini dapat dilihat bahwa angka 5 lebih baik daripada angka 4 karena memberikan urutan yang lebih tinggi. Namun demikian tidak dapat dicari perbedaan eksak untuk masing-masing skala.

2.2 Teori Pengelompokan

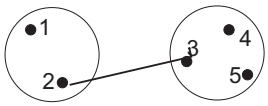
Analisis kelompok merupakan suatu analisis multivariat yang digunakan untuk mengelompokan objek pengamatan menjadi beberapa kelompok berdasarkan ukuran kemiripan atau ciri-ciri umum antar objek, sehingga objek-objek yang berada dalam satu kelompok mempunyai kemiripan yang lebih besar dibanding objek dari kelopmpok yang berbeda. Kelemahan analisis kelompok antara lain : semakin besar observasi tingkat kesalahan pengelompokan akan semakin besar. Metode yang digunakan untuk membuat kelompok dapat dipilih satu dari dua metode, yaitu metode hirarki dan non hirarki. Kedua metode tersebut dibedakan berdasarkan penentuan jumlah kelompok yang dihasilkan. Pada metode hirarki jumlah kelompok yang dihasilkan belum diketahui sedangkan metode non hirarki jumlahnya sudah ditentukan terlebih dahulu.

Lima metode yang dikembangkan untuk mengelompokan pada metode hirarki adalah :

(a) *Single Linkage*

Metode ini didasarkan pada jarak yang minimum. Artinya, mencari pasangan individu dengan jarak, terus menerus hingga diperoleh satu

kelompok individu. Selanjutnya, semakin jauh jarak adalah kelompok yang lain, dan seterusnya. Jarak antar kelompok (uv) dengan w adalah



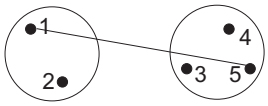
$$d_{(uv)w} = min(d_{uw}, d_{vw})$$

Gambar 2.1: Single Linkage

Hasil dari pengelompokan *single linkage* dapat digambarkan secara grafis melalui dendogram atau diagram pohon. Cabang-cabang pada pohon melambangkan kelompok (clusters). Cabang-cabang tersebut tergabung pada poros node (simpul) yang posisinya sepanjang jarak (atau kesamaan) yang menunjukkan level dimana gabungan terjadi.

(b) *Complete Linkage*

Metode ini merupakan kebalikan dari *single linkage*, yaitu mendasarkan pengelompokan dengan mencari pasangan individu dengan jarak terjauh, artinya individu dari dua atau lebih kelompok yang saling berjauhan diperoleh lebih dahulu dan seterusnya hingga individu-individu lainnya dipetakan mendekati kelompok yang mana. Jarak antar kelompok (uv) dengan w adalah

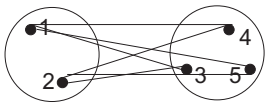


$$d_{(uv)w} = max(d_{uw}, d_{vw})$$

Gambar 2.2: Complete Linkage

(c) *Average Linkage*

Dasar pengelompokan *average linkage* ini adalah jarak rata-rata. Jarak dari dua individu adalah rata-rata dari semua kemungkinan pasangan individu. Jarak antar kelompok (uv) dengan w adalah



$$d_{(uv)w} = \frac{\sum_i \sum_k d_{ik}}{N(uv)N(w)}$$

Gambar 2.3: Average Linkage

Di mana d_{ik} adalah jarak antara objek i pada kelompok (uv) dan objek k pada kelompok W, dan N(uv) dan N(w) adalah banyaknya data pada kelompok (uv) dan w. Metode ini relatif yang terbaik dari metode hirarki yang lain (Ridho,Ali.B.2006). Namun, ini harus dibayar dengan waktu komputasi yang lebih lama.

(d) *Centroid Linkage*

Pada metode ini, jarak antara dua kelompok adalah jarak antara centroid kelompok-kelompok tersebut. *Centroid* adalah rata-rata jarak yang ada pada sebuah kelompok yang didapat dengan melakukan rata-rata pada semua anggota suatu kelompok tertentu. Dengan metode ini, setiap terjadi kelompok baru, akan terjadi perhitungan ulang centroid hingga terbentuk kelompok tetap.Jarak antar kelompok (uv) dengan w adalah

$$d_{(uv)w} = median(d_{uw}, d_{vw})$$

(e) *Ward's*

Ward adalah salah satu metode pengelompokan hirarki yang bertujuan untuk memperbaiki jarak antar kelompok. Metode ini cenderung menghasilkan kelompok dengan keragaman kecil. Dasar pengelompokan adalah jarak yang diperoleh dari deviasi jumlah kuadrat individu atas beberapa variabel.

Teknik yang digunakan dalam memperbaiki jarak antar kelompok adalah

$$d_{k(i,j)} = \frac{n_k + n_i}{n_k + n_{(i,j)}} d_{ki} + \frac{n_k + n_j}{n_k + n(i,j)} d_{kj} - \frac{n_k}{n_k + n_{(i,j)}} d_{ij}$$

keterangan :

n_i	=	jumlah objek kelompok-i
n_j	=	jumlah objek kelompok-j
n_k	=	jumlah objek kelompok-k
$n_{(ij)}$	=	$n_i + n_j$
d_{ij}	=	jarak antar kelompok ke-i dan kelompok ke-j
d_{ki}	=	jarak antar kelompok ke-i dan kelompok ke-k
d_{kj}	=	jarak antar kelompok ke-j dan kelompok ke-k

Yang termasuk dalam non hirarki adalah metode *K'means*. *K'means* memisahkan data ke k kelompok dan setiap data harus termasuk kelompok tertentu. Beberapa karakteristik dari *K-means* antara lain: memungkinkan suatu kelompok tidak mempunyai anggota dan hasil pengelompokan bersifat unik artinya memungkinkan bagi setiap data yang termasuk kelompok tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke kelompok lain. Keuntungan utama dari metode ini adalah kesederhanaan

dan kecepatan yang memungkinkan untuk pengoperasian di dataset yang besar. Tahap-tahap pengelompokan data dengan menggunakan metode ini adalah :

(a) Tentukan k buah pusat awal

(b) Tentukan jarak setiap data ke tiap pusat

(c) Lakukan pengelompokan setiap data ke pusat terdekat

(d) Tentukan nilai pusat baru sebagai rata-rata data dalam kelompok

(e) Lakukan langkah 3-5 sampai nilai pusat kelompok tak berubah lagi

2.3 Ukuran similaritas

Ukuran 'kedekatan' atau 'similaritas' dibutuhkan dalam upaya untuk menghasilkan struktur grup yang lebih sederhana dari himpunan data kompleks. Ada tiga ukuran untuk mengukur kesamaaan antarobjek, yaitu:

(a) Ukuran korelasi

Ukuran ini dapat diterapkan pada data dengan skala metrik, namun jarang digunakan karena titik beratnya pada nilai suatu pola tertentu, padahal titik berat analisis kelompok adalah besarnya objek. Kesamaan antar objek dapat dilihat dari koefisien korelasi antar pasangan objek yang diukur dengan beberapa variabel.

(b) Ukuran jarak

Ukuran jarak merupakan ukuran yang paling sering digunakan. Diterapkan untuk data berskala metrik. Sebenarnya merupakan ukuran

ketidakmiripan, di mana jarak yang besar menunjukkan sedikit kesamaan sebaliknya jarak yang pendek/kecil menunjukkan bahwa suatu objek makin mirip dengan objek lain. Bedanya dengan ukuran korelasi adalah bahwa ukuran jarak fokusnya pada besarnya nilai. Kelompok berdasarkan ukuran korelasi bisa saja tidak memiliki kesamaan nilai tapi memiliki kesamaan pola, sedangkan kelompok berdasarkan ukuran jarak lebih memiliki kesamaan nilai meskipun polanya berbeda. Ada beberapa tipe ukuran jarak antara lain:

i. Jarak *Euclidean* :

$$d(x,y) = \sqrt{\sum_{i=1}^{p}(x_i - y_i)^2} \tag{2.2}$$

ii. Matriks *Minkowski*:

$$d(x,y) = [(\sum_{i=1}^{p} |x_i - y_i|^m)]^{1/m} \tag{2.3}$$

iii. Matriks *Canberra*:

$$d(x,y) = \sum_{i=1}^{p} \frac{|x_i - y_i|}{(x_i + y_i)} \tag{2.4}$$

iv. Koefisien *Czekanowski*:

$$d(x,y) = 1 - \frac{2\sum_{i=1}^{p} min(x_i, y_i)}{\sum_{i=1}^{p}(x_i + y_i)} \tag{2.5}$$

v. Koefisien *Simple Matching*

$$d(x,y) = 1 - s_{ij} = 1 - \frac{p+s}{t} = \frac{p+q+r+s}{t} - \frac{p+s}{t} = \frac{q+r}{t} \tag{2.6}$$

dinama $s_{ij} = \frac{p+s}{t}$

p = jumlah peubah yang positif untuk kedua objek

s = jumlah peubah yang negatif ntuk kedua objek

q = jumlah peubah yang positif untuk objek ke-i dan negatif untuk objek ke-j

r = jumlah peubah yang negatif untuk objek ke-i dan positif untuk objek ke-j

(c) Ukuran Asosiasi

Ukuran asosiasi dipakai untuk mengukur data berskala kategorik (nominal atau ordinal).

	Peubah	k	
Peubah i	1	0	Total
1	a	b	a+b
0	c	d	c+d
total	a+c	b+d	a+b+c+d

Tabel 2.1: Ukuran Asosiasi

Rumus korelasi yang biasa diaplikasikan pada tabel kontingensi di atas:

$$r = \frac{ad - bc}{[(a+b)(c+d)(a+c)(b+d)]^{1/2}} \tag{2.7}$$

2.4 *Missing Value*

Missing value adalah informasi yang tidak tersedia untuk sebuah objek (kasus) akibat faktor *non sampling error*. Faktor yang dimaksud adalah *interviewer recording error, respondent inability error* dan *respondent unwillingness error*. *Interviewer recording error* terjadi akibat kealpaan petugas pengumpul data, misalnya ada sejumlah pertanyaan yang terlewatkan. *Respondent inability error* terjadi akibat ketidakmampuan responden dalam memberikan jawaban akurat, misal karena tidak memahami pertanyaan, bosan atau kelelahan akhirnya responden mengosongkan sejumlah pertanyaan atau berhenti mengisi kuesioner ditengah jalan. *Unwillingness respondent error* terjadi karena responden tidak berkenan memberikan jawaban yang akurat misalnya pertanyaan soal penghasilan, usia, berat badan atau pengalaman melakukan pelanggaran hukum.

Missing value pada dasarnya tidak bermasalah bagi keseluruhan data, apalagi jika jumlahnya hanya sedikit, misal hanya 1% dari seluruh data. Namun pada data hasil survei, yang pada umumnya berukuran besar, amatan hilang menjadi masalah yang lebih berat karena umumnya satu varibel mempunyai $30 - 90\%$ amatan hilang (Ratner, Bruce.2003). Sebagai contoh dalam pengambilan sampel acak (tabel 2.2)melalui kuesioner, 10 orang yang dijelaskan dengan 3 peubah yaitu UMUR, JENIS KELAMIN dan PENDAPATAN. Terdapat amatan hilang yang dinotasikan dengan titik (.).

Para analis data mengetahui bahwa hampir sebagian besar teknik analisis statistik membutuhkan data yang lengkap untuk mendapatkan hasil yang tepat. Teknik tersebut jika digunakan pada data yang terdapat amatan

Individu	UMUR	JENIS KELAMIN	PENDAPATAN
1	35	0	50.000
2	.	.	55.000
3	32	0	75.000
4	25	1	100.000
5	41	.	.
6	37	1	135.000
7	45	.	.
8	.	1	125.000
9	50	1	.
10	52	0	65.000
Jumlah amatan lengkap	8	7	7
persentase amatan hilang	20 %	30%	30%
0=Laki-laki 1=Perempuan			

Tabel 2.2: Contoh amatan hilang pada sampel acak

hilang akan menghasilkan hasil yang bias. Pada kenyataannya sangat jarang data yang didapat itu lengkap (tidak ada *missing value*). Maka para analis data mengusahakan berbagai macam teknik untuk mengatasi *missing value* pada kumpulan data mereka. Metode yang digunakan untuk mengatasi *missing value* diantaranya adalah *listwise deletion, pairwise deletion* dan *imputation.*

2.4.1 Metode *Listwise Deletion*

Metode pendekatan yang umum digunakan adalah menghilangkan amatan objek yang memiliki *missing value* dan hanya menganalisis amatan yang lengkap. Pendekatan ini biasa disebut *listwise deletion* atau *complete case analysis.*

Individu	UMUR	JENIS KELAMIN	PENDAPATAN
1	35	0	50.000
3	32	0	75.000
4	25	1	100.000
6	37	1	135.000
10	52	0	65.000
0=Laki-laki 1=Perempuan			

Tabel 2.3: *Listwise Deletion*

2.4.2 Metode *Pairwise deletion*

Pendekatan nilai tiap elemen matriks korelasi diduga menggunakan semua data yang ada. JIka seseorang menyebutkan pendapatan dan jenis kelamin tetapi tidak menyebutkan umurnya, maka dia termasuk kedalam korelasi pendapatan dan jenis kelamin, tetapi tidak termasuk korelasi umur.

Individu	UMUR	JENIS KELAMIN	PENDAPATAN
1	35	0	50.000
2	.	.	55.000
3	32	0	75.000
4	25	1	100.000
5	41	.	.
6	37	1	135.000
7	45	.	.
8	.	1	125.000
9	50	1	.
10	52	0	65.000
Jumlah amatan lengkap	8	7	7
0=Laki-laki 1=Perempuan			

Tabel 2.4: *Pairwise Deletion*

Permasalahannnya dari pendekatan ini adalah perhitungan korelasi didasarkan pada jumlah data yang berbeda.

2.4.3 *Imputation*

Imputation adalah proses mengisi data yang hilang untuk menghasilkan data yang lengkap. Metode *imputation* paling sederhana dan paling populer adalah *mean-value imputation*. Jika pada data yang mempunyai nilai (tidak *missing value*) pada peubah $Y_{1i}, i = 1, 2, ..., n$ maka akan diduga μ_1 menggunakan

$$\hat{\mu}_1 = n_1^{-1} \sum_{i=1}^{n_1} Y_{1i} \tag{2.8}$$

Lalu mengisi amatan hilang dengan mensubstitusi nilai μ_1.

2.5 Uji *Chi-Square*

Teknik uji ini memungkinkan untuk mengetahui independensi antara dua variabel pada tiap levelnya. Misal suatu variabel A memiliki r kategori dan variabel B memiliki c kategori. Asumsi yang data sampel adalah sampel acak dan skala pengukuran adalah nominal atau ordinal. Maka O_{ij} adalah frekuensi pengamatan pada variabel A di level i dan variabel B dilevel j, secara umum disajikan dalam tabel 2.5.

	B				
A	1	2	...	c	**total**
1	O_{11}	O_{12}	...	O_{1c}	$O_{1.}$
2	O_{21}	O_{22}	...	O_{2c}	$O_{2.}$
...	...	...	...	...	...
r	O_{r1}	O_{r2}	...	O_{rc}	$O_{r.}$
total	$O_{.1}$	$O_{.2}$	...	$O_{.c}$	n

Tabel 2.5: Struktur data uji *chi-square*

Jika $F(X)$ merupakan fungi distribusi dari X yang tidak diketahui distribusinya dan $F^*(X)$ adalah fungsi distribusi penduga. Maka hipotesis uji *chi-square*

H_0: $F(X) = F^*(X)$

H_1: $F(X) \neq F^*(X)$

atau bila dinyatakan dalam kalimat

H_0 : Fungsi disribusi pada peubah acak yang diamati adalah F^*

H_1 : Fungsi disribusi pada peubah acak yang diamati adalah berbeda dengan F^*

sedangkan uji statistiknya adalah

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \tag{2.9}$$

dimana

$$E_{ij} = \frac{n_{i.} n_{.j}}{n} \tag{2.10}$$

O_{ij} = frekuensi pengamatan pada baris ke-i dan kolom ke-j

E_{ij} = nilai harapan pengamatan pada baris ke-i dan kolom ke-j

$O_{i.}$ = total frekuensi pengamatan pada baris ke-i

$O_{.j}$ = total frekuensi pengamatan pada kolom ke-j

n = total banyaknya data.

Keputusan yang diambil, dengan derajat bebas (c-1)(r-1) dan taraf kesalahan 5%, dari uji *chi-square* ini adalah H_0 ditolak jika nilai $\chi^2_{hit} > \chi^2_{tabel}$ atau nilai-p$< \alpha$.

2.6 Koreksi Bonferroni

Koreksi Bonferonni adalah suatu proses koreksi yang digunakan ketika uji statistik untuk kebebasan atau ketidakbebasan dilakukan secara bersamaan (Sharp et al,2002). Koreksi bonferonni biasanya digunakan dalam analisis multivariat.

Ketika terdapat sebanyak M uji perbandingan yang sudah dikatakan bebas satu sama lain, peluang untuk melakukan kesalahan tipe 1 atau α (dalam satu atau lebih uji-uji tersebut),akan sama dengan 1 dikurangi peluang untuk tidak melakukan kesalahan tipe 1 dalam uji-uji tersebut, dimana nilainya akan lebih besar dari α yang telah ditentukan. Secara umum, hal tersebut dapat dirumuskan sebagai berikut (Bagozi, 1994).

$$1 - (1 - \alpha)^m > \alpha \tag{2.11}$$

dimana m=pengali Bonferroni, dan α= salah tipe 1.

Pengali Bonferroni untuk masing-masing tipe peubah bebas adalah berbeda. Gallagher menyebutkan bahwa pengali Bonferroni untuk masing-masing jenis peubah bebas adalah sebagai berikut:

(a) Peubah Monotonik

$$m = \begin{pmatrix} c-1 \\ r-1 \end{pmatrix} \tag{2.12}$$

(b) Peubah Bebas

$$m = \sum_{i=0}^{r-1} (-1)^i \frac{(r - i^i)}{i!(r-i)!} \tag{2.13}$$

(c) Peubah Mengambang (*floating*)

$$m = \begin{pmatrix} c-2 \\ r-2 \end{pmatrix} + r \begin{pmatrix} c-2 \\ r-1 \end{pmatrix} \tag{2.14}$$

dimana

m = pengali Bonferroni

c = kategori variabel dependen

r = kategori variabel independen.

Apabila terjadi penggabungan kategori-kategori sebuah peubah bebas atau pengurangan jumlah kategori dari c kategori menjadi r kategori, maka nilai-p yang digunakan dikalikan dengan pengganda Bonferroni.

BAB III

PEMBAHASAN

3.1 CHAID

CHAID (*Chi-squared Automatic Interaction Detection*)pertama kali diperkenalkan dalam sebuah artikel berjudul *"An Exploratory Technique for Investigating Large Quantities of Categorical Data"* oleh Dr. G. V. Kass pada tahun 1980. CHAID digunakan untuk membentuk kelompok yang membagi sampel menjadi dua atau lebih kelompok yang berbeda berdasarkan kriteria tertentu. CHAID secara keseluruhan bekerja untuk menduga sebuah peubah tunggal, disebut sebagai peubah terikat, yang didasarkan pada sejumlah peubah lain, disebut sebagai peubah bebas.

3.1.1 Definisi CHAID

CHAID merupakan suatu teknik iteratif yang menguji satu persatu peubah penjelas menggunakan uji *chi-square*. Kemudian diteruskan dengan membagi kelompok-kelompok tersebut menjadi kelompok yang lebih kecil berdasarkan peubah bebas yang paling signifikan. Proses pengujian berlanjut sampai tidak ditemukan lagi peubah bebas yang signifikan secara statistik (Gallagher). CHAID juga melakukan penggabungan kategori-kategori variabel bebas yang tidak memiliki pengaruh terhadap respons menjadi satu

kategori, sehingga hasil penggabungan itu menjadi kategori yang berpengaruh terhadap respons.

CHAID merupakan teknik eksplorasi nonparametrik untuk menganalisis sekumpulan data yang berukuran besar dan cukup efisien untuk menduga peubah-peubah bebas yang paling signifikan terhadap peubah terikat. CHAID tidak dapat diandalkan pada data yang berukuran kecil. CHAID sebaiknya digunakan pada data yang memiliki minimal 500 amatan dan maksimal 10 peubah penjelas. CHAID berbias pada data yang memiliki penduga berkategori lebih banyak dari 10. CHAID dapat digunakan untuk eksplorasi data dengan peubah nya berskala ukuran nominal atau ordinal.

3.1.2 Peubah-Peubah dalam CHAID

Dalam analisis CHAID, peubah yang digunakan dibedakan atas peubah terikat dan peubah bebas. Klasifikasi dalam CHAID dilakukan berdasarkan pada hubungan yang ada antara kedua peubah tersebut.

Menurut Gallagher, CHAID akan membedakan peubah bebasnya menjadi tiga,yaitu:

(a) Peubah Monotonik

Peubah yang disebut peubah monotonik oleh CHAID adalah peubah yang berskala ordinal. Kategori-kategori pada peubah ini dapat dikombinasikan atau digabungkan oleh CHAID hanya jika keduanya berdekatan satu sama lain. Misalkan pada suatu penelitian kategori umur dibagi menjadi 3 kategori, usia remaja ($\leq$23 tahun), usia 24-30 tahun, usia 31-40 tahun dan usia 41-50 tahun. Maka pada tahap penggabungan,

kategori yang dapat digabung hanya kategori yang berurutan, usia remaja ($\leq$23 tahun) dan usia 24-30 tahun dapat digabung menjadi usia $\leq$ 30 tahun, usia 31-40 tahun dan 41-50 tahun dapat digabung menjadi usia 31-50 tahun.

(b) Peubah Bebas

Data yang berskala nominal dikategorikan menjadi peubah bebas dalam CHAID. Kategori-kategori pada peubah ini dapat dikombinasikan atau digabungkan walaupun keduanya berdekatan atau tidak satu sama lain. Peubah yang termasuk peubah bebas antara lain pekerjaan, kelompok etnik, dan area geografis.

(c) Peubah Mengambang (*floating*)

Peubah yang memiliki *missing value* akan dikategorikan menjadi peubah mengambang dalam CHAID. Kategori peubah ini dapat berkombinasi dengan kategori manapun, tidak ada syarat urutan.

Pengkategorian peubah pada CHAID terkait dengan pengali Bonferroni yang digunakan.

3.1.3 Uji *chi-square* pada CHAID

Pada umumnya uji *chi-square* digunakan untuk menguji kebebasan antar peubah. CHAID menggunakan uji *chi-square* dalam dua tahap. Yang pertama, uji *chi-square* digunakan untuk menentukan apakah kategori-kategori dalam sebuah peubah bebas bersifat seragam dan bisa digabungkan menjadi satu. Kategori-kategori setiap peubah bebas di buat tabulasi silang dengan kategori-kategori peubah terikat seperti pada tabel 3.1.

	Peubah terikat				
P. bebas x	kategori 1	kategori 2	...	kategori c	**total**
kategori 1	O_{11}	O_{12}	...	O_{1c}	$O_{1.}$
kategori 2	O_{21}	O_{22}	...	O_{2c}	$O_{2.}$
total	$O_{.1}$	$O_{.2}$	...	$O_{.c}$	n

Tabel 3.1: Uji signifikan pasangan kategori peubah bebas

x = $1 \leq x \leq$ banyaknya peubah bebas

O_{ij} = frekuensi pengamatan pada peubah terikat kategori ke-i dan peubah bebas kategori ke-j

E_{ij} = nilai harapan pengamatan pada peubah terikat kategori ke-i dan peubah bebas x kategori ke-j

$O_{i.}$ = total frekuensi pengamatan pada peubah terikat kategori ke-i

$O_{.j}$ = total frekuensi pengamatan pada peubah bebas x kategori ke-j

n = total banyaknya data.

Dua kategori yang memiliki nilai χ^2 terkecil dan lebih kecil dari α yang ditentukan maka kedua kategori tersebut digabung. Begitu seterusnya hingga nilai χ^2 terkecilnya melebihi atas signifikan.

Pengurangan tabel kontingensi (penggabungan kategori) pada algoritma CHAID membutuhkan suatu uji signifikansi. Jika tidak terjadi pengurangan dari tabel kontingensi asal, maka uji *chi-square* biasa dapat digunakan. Apabila terjadi pengurangan yaitu c kategori dari peubah asal menjadi r, kategori ($r < c$), maka nilai-p dari *chi-square* yang baru dikalikan dengan pengganda Benferroni sesuai dengan tipe peubah.

Ketika semua kategori peubah bebas sudah diringkas menjadi bentuk yang signifikan dan tidak mungkin digabung lagi, kemudian uji *chi-square*

digunakan untuk menentukan peubah penjelas mana yang paling signifikan untuk membagi atau membedakan kategori-kategori dalam peubah respon (Gallagher).

	Peubah terikat (PT)				
P.bebas (PB)	PT 1	PT 2	...	PT c	**total**
PB 1	O_{11}	O_{12}	...	O_{1c}	$O_{1.}$
PB 2	O_{21}	O_{22}	...	O_{2c}	$O_{2.}$
...	...	...	...	...	...
PB r	O_{r1}	O_{r2}	...	O_{rc}	$O_{r.}$
total	$O_{.1}$	$O_{.2}$	...	$O_{.c}$	n

Tabel 3.2: Uji signifikan peubah bebas

O_{ij} = frekuensi pengamatan pada peubah terikat kategori ke-i dan peubah bebas kategori ke-j

E_{ij} = nilai harapan pengamatan pada peubah terikat kategori ke-i dan peubah bebas kategori ke-j

$O_{i.}$ = total frekuensi pengamatan pada peubah terikat kategori ke-i

$O_{.j}$ = total frekuensi pengamatan pada peubah penjelas kategori ke-j

n = total banyaknya data

Peubah bebas yang memiliki nilai-p terkecil $< \alpha$ digunakan untuk membagi atau membedakan kategori pada peubah terikat.

3.1.4 Algoritma CHAID

Tahapan-tahapan dalam analisis CHAID menurut Kass 1980 adalah sebagai berikut:

(a) Tahap 1.

Membuat tabulasi silang antara kategori-kategori peubah bebas dengan kategori-kategori peubah terikat. Hal ini dilakukan untuk setiap peubah bebas.

(b) Tahap 2.

Membuat Subtabel pasangan kategori peubah bebas berukuran $2xd$, dengan d adalah banyaknya kategori peubah terikat. Carilah nilai χ^2_{hitung} semua subtabel tersebut dengan α ditetapkan, kemudian cari χ^2_{hitung} terkecil. Jika χ^2_{hitung} terkecil $< \chi^2_{\alpha}$ maka kedua kategori peubah bebas yang memiliki χ^2_{hitung}tersebut digabungkan menjadi satu kategori gabungan.Ulangi tahap ini sehingga angka uji terkecil(χ^2_{hitung}) subtabel 2xd pasangan kategori (kategori campuran) peubah bebas melampaui nilai kritis. Pada tahap ini akan direduksi c kategori peubah bebas menjadi r kategori (r<c). Untuk peubah ordinal penggabungan hanya dapat dilakukan untuk kategori yang berurutan.

(c) Tahap 3.

Jika terdapat kategori gabungan yang terdiri dari tiga atau lebih kategori asal maka harus dilakukan pembagian biner terhadap kategori gabungan tersebut . Dari pembagian ini dicari χ^2_{hitung} terbesar, jika χ^2_{hitung} terbesar $< \chi^2_{\alpha}$ maka pembagian biner dilakukan. Selanjutnya kembali ke tahap 2. Jika $\chi^2_{hitung} > \chi^2_{\alpha}$, lanjut ke tahap 4.

(d) Tahap 4.

Dari setiap peubah bebas yang telah digabungkan secara optimal hitung nilai-p untuk masing-masing tabel yang terbentuk. Tabel yang mengalami reduksi menjadi r kategori, nilai-p nya dikalikan dengan

pengganda Bonferroni sesuai dengan tipe peubahnya. Cari nilai-p yang paling kecil dari masing-masing tabel tersebut. Jika nilai-p terkecil $< \alpha$ maka peubah bebas tersebut adalah peubah yang paling signifikan terhadap peubah terikat. Maka bagilah data menurut peubah bebas tersebut.

(e) Tahap 5.

Kembali ke tahap 1 untuk melakukan pembagian berdasarkan peubah bebas yang belum terpilih. Hentikan jika tidak ada lagi peubah bebas yang signifikan.

Yasmin Erika,2003, memperjelas algoritma CHAID sebagai berikut:

(a) Untuk setiap penduga X, cari pasangan kategori dari X yang memiliki nilai-p terbesar berdasarkan level atau kelas peubah terikat Y

(b) Untuk pasangan kategori dari X dengan nilai-p terbesar, bandingkan nilai-p nya dengan α_{merge} yang telah ditentukan sebelumnya.

- Jika nilai-p lebih besar dari α_{merge}, gabung pasangan ini kedalam satu kategori baru
- Jika nilai-p lebih kecil dari α_{merge}, lanjut ke nomor (c)

(c) Hitung nilai-p terkoreksi untuk gugus kategori X dan kategori Y dengan menggunakan koreksi Bonferroni.

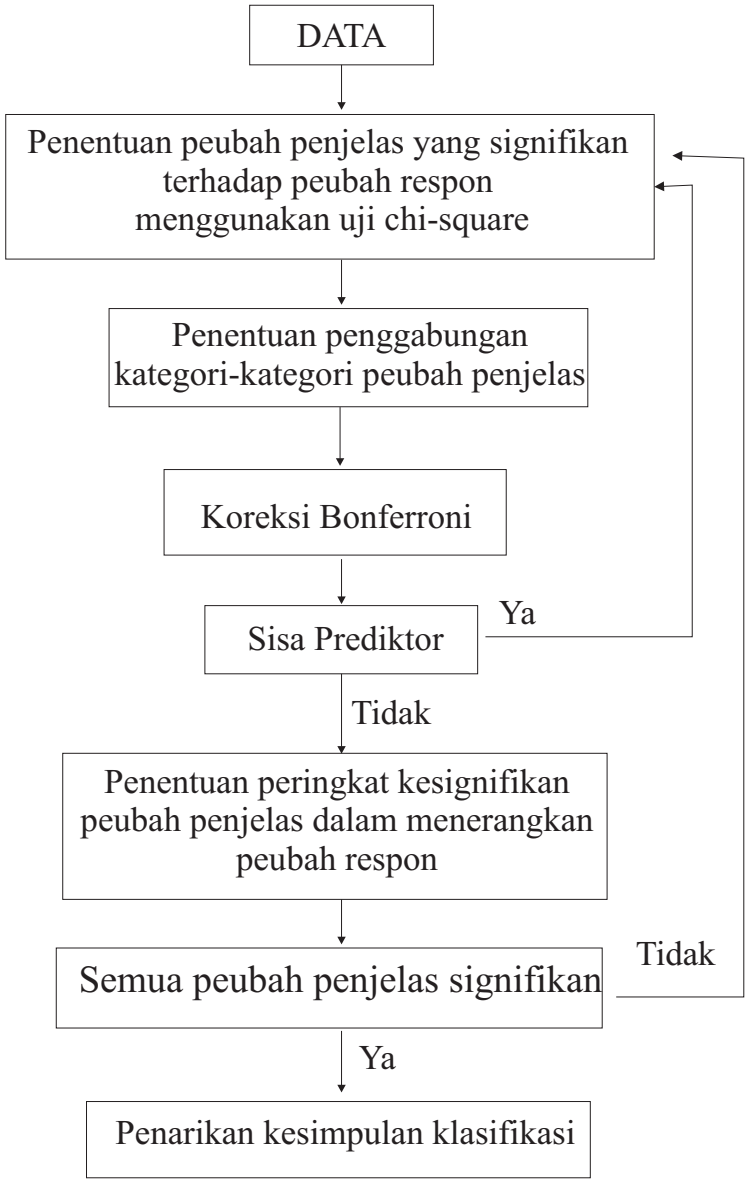
(d) Pilih penduga X yang memiliki nilai-p terkoreksi terkecil. Bandingkan dengan nilai-p tersebut dengan nilai α_{split}yang telah ditentukan sebelumnya.

- Jika nilai-p kurang atau sama dengan α_{split} maka node di split berdasarkan gugus kategori X

- Jika nilai-p lebih besar dari α_{split} maka node tidak di split. Note tersebut merupakan node akhir

(e) Kembali ke langkah (a) untuk penduga yang belum dianalisis.

Algoritma diatas dapat dijelaskan dengan diagram alir seperti pada gambar 1.



3.1.5 Amatan Hilang dalam CHAID

Allison mengatakan bahwa "***The best solution to the missing data problem is not to have any missing data***". Tidak ada yang dapat mengisi amatan hilang. Namun, Bruce Ratner (2003) menyarankan CHAID sebagai metode *imputation* yang dapat mengakomodasi amatan hilang. Alasan CHAID digunakan sebagai metode *imputation* adalah dari definisi diketahui bahwa CHAID membuat grup yang homogen secara optimal sehingga dapat digunakan sebagai kelas *imputation* yang terpercaya.

Pada panduan penggunaan program *answertree* dikatakan bahwa CHAID menangani amatan hilang dengan memperlakukannya sebagai kategori tunggal. Bila suatu peubah mempunyai amatan hilang maka peubah itu dikategorikan sebagai peubah mengambang (*floating*). Dalam proses penggabungan kategori, kategori-kategori pada peubah mengambang dapat digabungkan dengan kategori manapun, jadi berbeda dengan peubah monotonik yang proses penggabungan hanya boleh pada kategori yang terurut.

CHAID juga memiliki pengali Bonferroni untuk peubah mengambang, yang dibedakan dengan peubah monotonik dan peubah bebas, yaitu :

$$m = \begin{pmatrix} c-2 \\ r-2 \end{pmatrix} + r \begin{pmatrix} c-2 \\ r-1 \end{pmatrix} \tag{3.1}$$

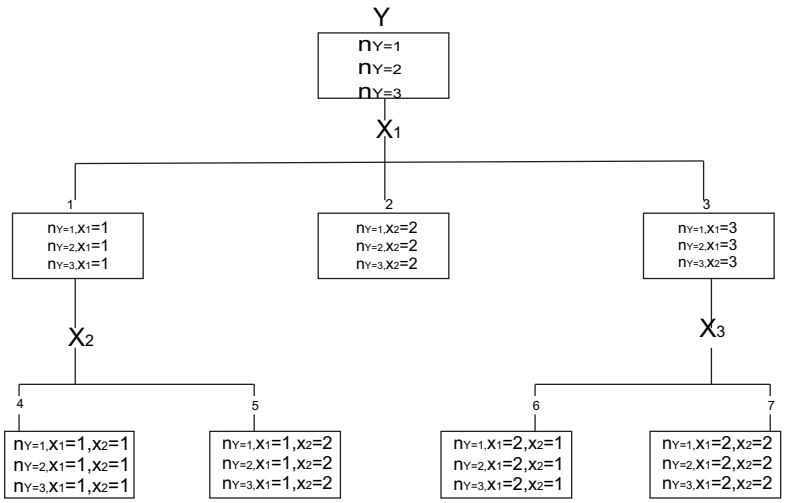
.

Berbeda dengan metode *decision tree* yang lain, CHAID memperlakukan mengikutsertakan objek dan peubah yang mempunyai amatan hilang dalam proses analisis, karena dengan menghilangkan suatu objek atau peubah maka informasi yang bisa diperoleh pun dapat hilang. Sebagai penduga im-

putasi CHAID menggunakan kategori yang paling domominan yang dilihat dari nilai proporsi pada node terakhir (Bruce, 2003).

3.1.6 Struktur Pohon

Hasil pengklasifikasian dalam CHAID akan ditampilkan dalam sebuah struktur pohon. Secara umum struktur pohon dari CHAID adalah sebagai berikut (Lehmann dan Eherler):



Gambar 3.1: Diagram pohon CHAID

Struktur pohon CHAID mengikuti aturan "dari atas ke bawah" (*Top-down stopping rule*), dimana struktur pohon disusun mulai dari kelompok induk, berlanjut dibawahnya sub kelompok yang berturut-turut dari hasil pembagian kelompok induk berdasarkan kriteria tertentu. Tiap-tiap node dari struktur 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 diatasnya. Pada pohon klasifikasi CHAID

terdapat istilah kedalaman (*depth*) yang berarti banyaknya tingkatan node-node sub kelompok sampai ke bawah pada node sub kelompok yang terakhir. Pada kedalaman pertama, sampel dibagi oleh X_1 sebagai peubah bebas terbaik untuk peubah terikat berdasarkan uji *chi-square*. Tiap node berisi informasi tentang frekuensi peubah Y, sebagai peubah bebas, yang merupakan bagian dari sub kelompok yang dihasilkan berdasarkan kategori yang disebutkan X_i. pada pembagian dari X_i (untuk node ke-1 dan ke-3). Dengan cara yang sama, sampel selanjutnya dibagi oleh peubah bebas yang lain, yaitu X_2 dan X_3, dan selanjutnya menjadi sub kelompok pada node ke-4, 5, 6, dan 7 (Lehmann dan Eherler).

Secara ringkas struktur pohon yang merupakan inti dari analisis CHAID akan berisi:

(a) Simbol yang menerangkan tentang kategori tertentu (atau kategori-kategori yang telah digabungkan.

(b) Sebuah ringkasan data dari peubah terikat dalam kelompok tersebut (misalnya persentase respon)

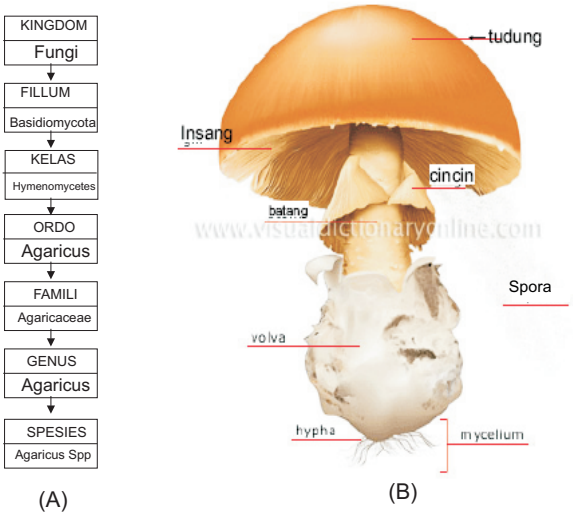
(c) Ukuran sampel untuk kelompok tersebut, atau biasa dilambangkan dengan "n".

3.2 Penerapan metode CHAID dalam pengklasifikasian jamur Agaricus dan Lepiota.

3.2.1 Pemilihan studi kasus

Dalam klasifikasi tumbuhan, kingdom Fungi dibagi kedalam empat fillum yaitu Chytridiomycota, Ascomycota, Zygomycota, dan Basidiomycota. Setiap filum dibagi ke dalam kelas, setiap kelas dibagi ke dalam ordo dan setiap ordo dibagi ke dalam famili lalu genus dan terakhir spesies. Basidiomycota adalah fillum terbesar dari Fungi karena mempunyai lebih dari 30.000 jenis (John Webster: 2007). Agaricus dan Lepiota adalah salah satu genus dari fillum Basidiomycota (gambar 3.2). Kedua jamur itu termasuk pada famili Agaricaceae. Ciri-ciri Agaricaceae yaitu memiliki sisik pada butiran-butiran kecil didaerah tudung, insang yang terpisah dari batang, memiliki tudung membran, dan kebanyakan diantaranya memiliki cincin yang melekat pada batang. Agaricus dan Lepiota secara liar hidup dialam terbuka dengan bentuk dan warna yang beraneka ragam. Beberapa jenis genus Agaricus dan Lepiota termasuk ke dalam makro fungi karena memiliki bentuk yang besar, dapat dilihat dengan mata telanjang dan dapat dipegang dengan tangan. Salah satu bentuk makro Fungi disajikan pada Gambar 3.2.

Pada umumnya jamur dari genus Agaricus dan Lepiota ini bersifat racun. Namun ada juga yang bisa dikonsumsi antara lain Agaricus Bisporus (jamur kancing). Untuk membedakan jamur Agaricus dan Lepiota yang dapat dimakan dengan jenis yang beracun ,berdasarkan bentuk dan sifatnya, sangat sulit dilakukan. Misalnya Lepiota Morgani adalah jamur



Gambar 3.2: Taksonomi Agaricus (A) dan Makro Fungi (Agaricaceae)(B)

yang beracun dan Lepiota Rachodes adalah jamur yang dapat dimakan. Kedua jamur ini terlihat mirip, namun jika seorang ahli yang mengamati akan terlihat perbedaannya, misalnya dengan melihat warna insang.

Selain dengan melihat warna insang, identifikasi bisa dilakukan dengan melihat warna cetakan spora. Selain itu jamur beracun bisa dikenali melalui aroma yang busuk dan tajam. Jamur Agaricus dan Lepiota sulit untuk diidentifikasi jika umurnya masih muda (Christensen:1972).

Tujuan dari penelitian ini adalah mengklasifikasikan jamur dari genus Agaricus dan Lepiota ke dalam kelas dapat dimakan atau beracun.

3.2.2 Data amatan

Data yang digunakan diperoleh dari *UCI repository of machine learning database (http://archive.ics.uci.edu/ml/datasets/Mushroom)*. Data ini merupakan hasil penelitian dari *The Audubon Society Field guide to North*

American Mushroom (1981) yang berisi deskripsi dari contoh jamur berinsang (*gilled mushroom*) dari genus agaricus dan lepiota. Peubah terikat dan peubah bebas yang diamati dapat dilihat di lampiran 1. Data yang didapat sudah diubah menjadi data numerik untuk masing masing atribut. Hal ini digunakan untuk mempermudah dalam memasukan (input) data dan siap untuk dianalisis menggunakan CHAID. Pada data ini terdapat kategori *unknown* yaitu kategori untuk data yang tidak diketahui nilainya (*missing value*). Semua peubah bebas berskala nominal, kecuali peubah bebas B18 (banyaknya cincin) yang berskala ordinal.

3.2.3 Analisis hasil metode CHAID

A. Statistika deskriptif

Data yang diolah berjumlah 8124 jamur dengan jamur yang tergolong dapat dimakan 4208 jamur (51,8%) dan tergolong beracun sebanyak 3916 jamur (48,2%).

(a) Tabel frekuensi

Data amatan diatas jika dibuat tabel frekuensi berdasarkan peubah-peubah yang digunakan diperoleh hasil bahwa:

- Tudungnya berbentuk cembung sebanyak 2940 jamur (36,2%)
- Tidak beraroma sebanyak 2890 jamur (35,6%)
- Bentuk tangkai lonjong sebanyak 3879 (47,7%)
- Membran pembungkus seluruhnya parsial 6696 (82,4%)
- Mempunyai satu cincin sebanyak 6557 jamur (80,7%)

- Berhabitat di kayu sebanyak 2743 jamur (33,8%)

nilai peubah lain dapat dilihat pada lampiran 2.

(b) Tabulasi silang antara peubah bebas dengan peubah terikat

Y	**1**	**2**	**3**	**4**	**5**
0	425	0	409	29	2
1	66	540	2	0	162
Total	491	540	411	29	164
	6	**7**	**8**	**9**	**Total**
0	1212	0	1565	23	3665
1	561	93	2412	0	3836
Total	1773	93	3977	23	7501

Tabel 3.3: Tabulasi silang antara peubah B15 dengan Y

Beberapa karakteristik jamur genus Agaricus dan Lepiota terlihat dari tabulasi silang peubah bebas terhadap peubah terikat Y (0: beracun, 1: dapat dimakan). Beberapa diantaranya ditampilkan di sini dan sisanya dapat dilihat pada lampiran 2. Tabel 3.3 memperlihatkan bahwa jamur genus Agaricus dan lepiota yang warna tangkai dibawah cincinnya berwarna kuning sudah pasti beracun dan dapat dimakan jika berwarna abu-abu. Berdasarkan Uji chi-square dengan nilai $\chi^2 = 1923,144$, df=8, p=0.0000.

Peubah AROMA pada tabel 3.4 juga memberikan informasi tentang jamur genus Agaricus dan lepiota. Jamur yang aroma busuk dan tajam pedas dipastikan adalah jamur beracun yang tidak dapat dimakan. Sedangkan jika jamur beraroma almond, adas dan tidak beraroma jamur tersebut bisa dimakan. Berdasarkan Uji chi-square dengan nilai $\chi^2 = 1923,144$, df=8, p=0.0000.

Y	**1**	**2**	**3**	**4**	**5**
0	2	0	158	493	1721
1	355	359	0	3	5
Total	357	359	158	496	1726
	6	**7**	**8**	**9**	**Total**
1	29	221	482	109	3510
0	1	0	6	2781	3675
Total	30	221	488	2890	6725

Tabel 3.4: Tabulasi silang antara peubah B5 dengan Y

Berdasarkan tabulasi silang antara peubah B8 (ukuran Insang) diperoleh informasi bahwa jamur yang mempunyai ukuran insang sempit dapat dikatakan jamur tersebut beracun. Sebaliknya jika ukuran insang lebar, sebagian besar jamur dapat dimakan. Berdasarkan Uji chi-square dengan nilai $\chi^2 = 1936,882$, df=1, p=0.0000.

Y	**1**	**2**	**Total**
0	1405	1869	3274
1	3195	247	3442
Total	4600	2116	6716

Tabel 3.5: Peubah B8 dengan Y

Prinsip yang sama juga digunakan untuk peubah bebas lainnya. Tabel tabulasi silang dapat dilihat pada lampiran 3.

B. Pengkategorian ulang hasil metode CHAID

Data amatan diolah menggunakan program spss answertree 2.0. Dengan nilai $\alpha = 0.05$ dan nilai kedalaman (*depth*) tiga diperoleh hasil diagram pohon 10 node. Seluruh peubah bebas yang signifikan terhadap peubah terikat (respon) mengalami pengkategorian ulang dengan metode CHAID. Pada awalnya peubah bebas berjumlah 22 peubah,

namun setelah di analisis dengan metode CHAID ternyata peubah bebas yang signifikan hanya berjumlah 3 peubah. Hasil pengkategorian ulang ditampilkan dalam tabel 3.6.

Peubah bebas	**Kategori Awal**	**Kategori baru hasil CHAID**
B5	10	Lima kategori, yaitu: 1. Tajam pedas, busuk, ter kayu, amis 2. Almond pedas 3. Tidak beraroma 4. Pedas, apak 5. Missing
B9	13	Tiga kategori, yaitu: 1. Hitam, cokelat, abu-abu, merah muda cokelat tua, ungu, missing 2. Putih, merah 3. Kekuningan
B20	10	Empat kategori, yaitu: 1. Hitam, cokelat, jingga, kekuningan 2. Cokelat tua, kuning, $< missing >$ 3. Putih 4. Hijau

Tabel 3.6: Pengkategorian ulang hasil metode CHAID

Nilai p-value dari masing-masing peubah bebas yang signifikan adalah 0,0000 (AROMA); 0,0000 (WINS)dan 0,0000 (WSPOR). Nilai p-value tersebut merupakan nilai p-value setelah dikoreksi oleh pengali Bonferroni. Semua nilai p-value kurang dari $\alpha = 0,05$ maka dapat ditarik keputusan uji *chi-square* adalah tolak H_0. Hal ini berarti bahwa terdapat hubungan antara 3 peubah bebas dengan peubah terikat, yaitu status dapat dimakan atau beracun.

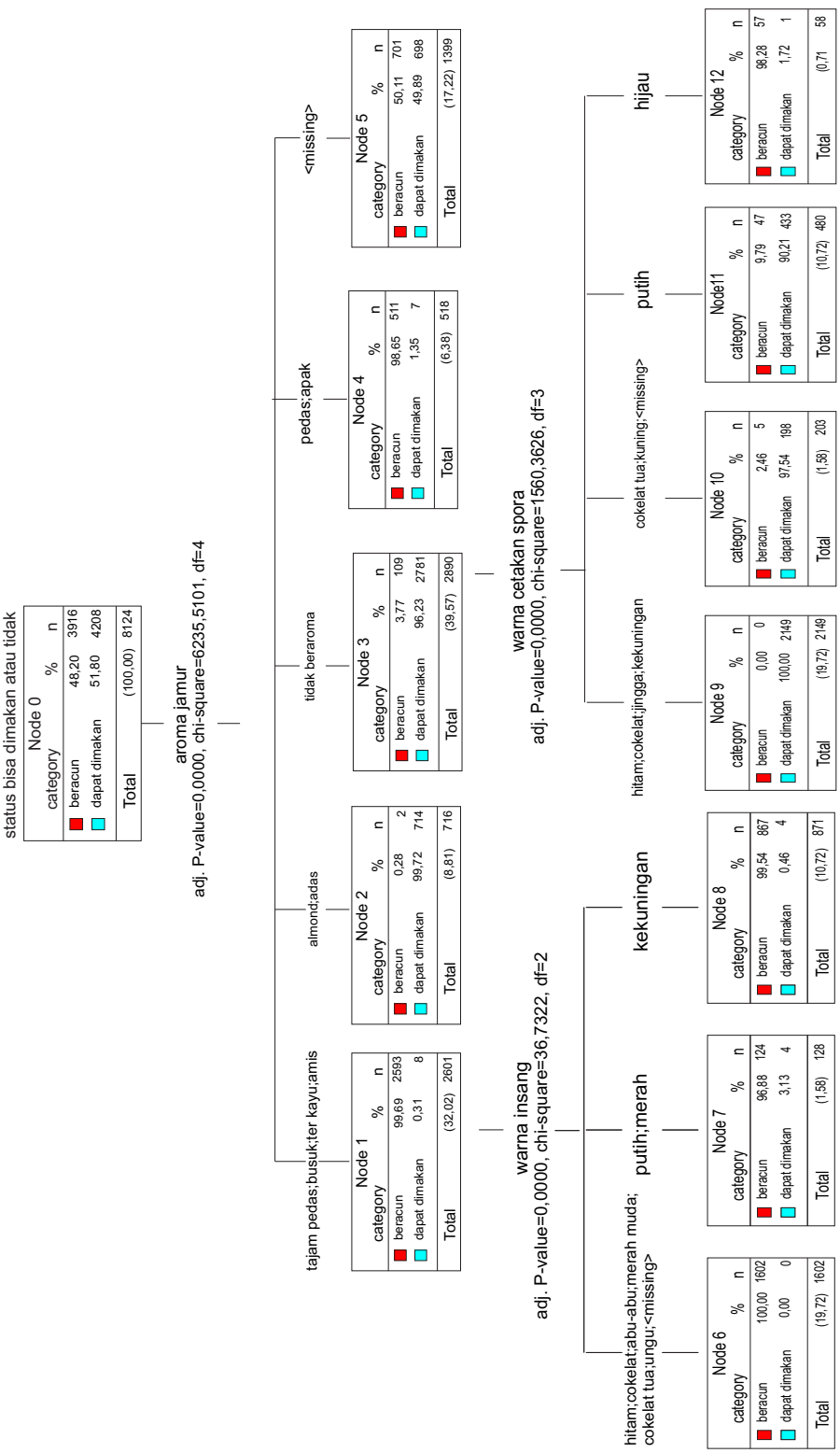
C. Interpretasi diagram pohon analisis CHAID

Dari analisis data jamur dengan metode CHAID diperoleh diagram pohon pada gambar 3.3. Nilai kritis yang ditetapkan untuk men-

dapatkan diagram pohon adalah 0,05. Peubah-peubah yang berperan dalam pemisahan adalah (1) Aroma, (2) Warna Insang , dan (3) Warna cetakan spora. Hasil pemisahan tersebut memperlihatkan peubah pertama yang terpilih dalam memisahkan status jamur adalah aroma jamur. Pada jamur yang memiliki aroma tajam pedas, busuk, terkayu, amis yang tergolong jamur beracun berjumlah 2593 jamur (99,69%), sedangkan yang beraroma almond;adas yang tergolong beracun lebih sedikit (0,28%), kategori dapat dimakan lebih besar yaitu 714 jamur (99,72%), kategori tidak beraroma 2781 jamur dapat dimakan (96,23%), sebanyak 511 jamur (98,65%) beraroma pedas ;apak adalah beracun dan aroma yang tidak diketahui (*missing*) sebanyak 701 jamur (50,11%) beracun.

Selanjutnya jamur beraroma tajam pedas;busuk;terkayu;amis dibagi menjadi tiga segmen berdasarkan warna insang. Hitam;cokelat;abu-abu;merah muda;cokelat tua;ungu;missing 100% beracun, putih;merah 96,88% beracun dan kekuningan 99,54% beracun. Sedangkan pada kelompok jamur tidak beraroma, kelompok ini dipisahkan oleh warna cetakan spora . Jamur tidak beraroma yang memiliki warna cetakan spora hitam;cokelat;jingga;kekuningan seluruhnya masuk dalam kelompok jamur dapat dimakan (100%). Jamur tidak beraroma yang memiliki warna cetakan spora cokelat tua;kuning;missing sebagian besar adalah jamur dapat dimakan (97,54%).

Diagram pohon hasil analisis CHAID menghasilkan beberapa 10 segmen. Berikut ini segmentasi hasil analisis CHAID serta frekuensi jamur beracun (p) dan jamur dapat dimakan (n).



Gambar 3.3: Diagram pohon hasil analisis CHAID

i. Segmen 1

Jamur tidak beraroma dan warna cetakan sporanya hitam;cokelat jingga;kekuningan. n= 2149 (100%), p= 0 (0%)

ii. Segmen 2

Jamur beraroma almond;adas.

n= 714 (99,72%), p= 2 (0,28%)

iii. Segmen 3

Jamur, tidak beraroma dan warna cetakan sporanya cokelat tua; kuning;*missing*. n= 198 (97,54%), p= 5 (2,46%)

iv. Segmen 4

Jamur tidak beraroma dan warna cetakan sporanya putih.

n= 433 (90,21%), p= 47 (9,79%)

v. Segmen 5

Jamur yang aromanya tidak diketahui

n= 698 (49,89%), p= 701 (50,11%)

vi. Segmen 6

Jamur yang warna insangnya putih;merah dan aromanya tajam; pedas;busuk;ter kayu;amis. n= 4 (96,87%), p= 124 (3,13%)

vii. Segmen 7

Jamur yang tidak beraroma dan warna cetakan sporanya hijau.

n= 1(1,72%), p= 57 (98,28%)

viii. Segmen 8

Jamur yang beraroma pedas;apak.

n= 7(1,35%), p= 511(98,65%)

ix. Segmen 9

Jamur yang warna insangnya kekuningan dan aromanya tajam; pedas;busuk;ter kayu;amis. n= 4 (0,46%), p= 867(99,64%)

x. Segmen 10

Jamur yang warna insangnya hitam;cokelat;jingga;kekuningan dan aromanya tajam;pedas;busuk;ter kayu;amis. n= 0 (0%), p= 1602 (100%)

Segmentasi diatas dibuat berdasarkan nilai *gain* dari yang tertinggi hingga terendah. Segmen 1, jamur yang tidak beraroma dan warna cetakan sporanya salah satunya hitam termasuk jamur yang dapat dimakan. Contoh jamur dari genus Agaricus yang cetakan sporanya hitam adalah Agaricus Arvensis (*Horse Mushroom*)dan Agaricus Campestris (*Field Mushroom*). Kedua jamur tersebut merupakan jamur yang dapat dimakan (Hall: 2003).

Segmen 4, jamur tidak beraroma dan warna cetakan sporanya putih. Sebagian besar jamur Lepiota yang memiliki spora putih dapat dimakan dan rasanya enak, namun ada pula jenis yang beracun, yaitu Lepiota Cristata. Hal ini sejalan dengan hasil diagram pemisahan yang menyatakan bahwa jamur berspora putih ada yang termasuk kategori beracun, yaitu sebanyak 9,79%. Segmen 7, jamur yang tidak beraroma dan warna cetakan sporanya hijau sebagian besar beracun. Jamur tidak beraroma yang warna cetakan sporanya berwarna hijau, sebagian besar dikategorikan menjadi jamur beracun 98,28%. Warna spora dapat menjadi petunjuk yang masuk akal. Jamur *chlorophyllum molybdites* adalah satu-satunya jamur berspora hijau, dan akan

menyebabkan gangguan pencernaan bila dimakan. Jamur yang memiliki warna cetakan spora putih sebagian besar (90,21%) dapat dimakan. Sedangkan segmen 6 jamur yang warna insangnya putih;merah dan aromanya tajam;pedas;busuk;ter kayu;amis sebagian besar adalah jamur beracun. Jamur yang memiliki aroma tajam dan busuk bisa menjadi indikasi awal yang mudah bagi orang awam bahwa jamur tersebut adalah beracun.

Gambar 3.4: Jamur genus Agaricus dan Lepiota yang dapat dimakan

D. Analisis *missing value*

Cara penanganan *missing value* oleh CHAID adalah dengan menganggap *missing value* sebagai satu kategori tunggal. Sehingga dalam diagram pohon hasil analisis CHAID akan dijumpai node dengan kategori <*missing*>. Diagram pohon hasil klasifikasi jamur Agaricus dan Lepiota juga memunculkan kategori <*missing*> pada node 5 sebagai kategori tunggal. Sedangkan pada node 6 dan 10 kategori <*missing*> bergabung dengan kategori lainnya.

Hasil diagram pohon dapat digunakan untuk penduga imputasi. Pada CHAID kategori yang paling dominan digunakan sebagai kelas imputasi (Bruner,2003). Pada Node 5 kategori yang paling dominan

adalah jamur beracun (50,1%), maka jamur dengan kategori <*missing*> atau aromanya tidak diketahui maka diduga sebagai jamur beracun. Sedangkan pada node 6 jamur dengan kategori <*missing*> atau jamur yang mempunyai aroma tajam pedas;busuk;ter kayu;amis dan warna insangnya tidak ditahui maka diduga sebagai jamur beracun karena kategori yang paling dominan adalah jamur beracun (100%). Berbeda dengan node 5 dan 6, jamur dengan kategori <*missing*> pada node 10 atau jamur yang tidak beraroma namun warna cetakan sporanya tidak diketahui maka diduga sebagai jamur yang dapat dimakan karena kategori jamur dapat dimakan adalah kategori yang paling dominan (97,54%).

BAB IV

KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan hasil penelitian dapat diambil kesimpulan sebagai berikut:

1. CHAID merupakan metode yang cukup efisien untuk menentukan peubah-peubah bebas yang signifikan terhadap peubah terikat. Hasil penelitian ini menghasilkan 3 peubah bebas yang signifikan, yaitu aroma, warna insang dan warna cetakan spora jamur. Peubah aroma merupakan peubah paling signifikan terhadap identifikasi jamur dapat dimakan atau beracun.
2. CHAID meringkas kategori pada peubah bebas. Dari diagram pohon hasil penelitian diperoleh pengkategorian ulang yaitu peubah aroma diringkas menjadi 5 kategori dari 10 kategori awal, peubah warna insang diringkas menjadi 3 kategori dari 13 kategori awal dan peubah warna cetakan spora diringkas menjadi 5 kategori dari 10 kategori awal.
3. CHAID menangani *missing value* sebagai kategori tunggal sehingga akan ditemukan node pada diagram pohon dengan kategori <*missing*>, yaitu pada node 5,6 dan 10.

4.2 Saran

Dalam skripsi ini hanya menggunakan CHAID untuk mengklasifikasikan data yang peubah bebasnya kategorik. Apabila ingin menerapkan metode lain dari *decision tree* untuk data yang peubah bebasnya kategorik maka dapat menggunakan QUEST atau CRUISE.

DAFTAR PUSTAKA

Boriah, Shyam. *Similarity Measures for Categorical Data.* Department of Computer Science and Engineering, University of Minnesota.

Conover, W.J. 1980.*Practical Non Parametric Statistic.*John Willey and Sons, Inc.

Christensen, C.M. 1972.*Common Edible Mushroom.* University of Minnesota.

Faridhan, Y. E. 2003. *Metode Klasifikasi Berstruktur Pohon dengan algoritma CRUISE, QUEST dan CHAID.* Bogor: Program Pascasarjana Institut Pertanian Bogor.

Gallagher, Cecily.A. *An Itterative Approach to Classification Analysis.*

Hall, Ian.R. 2003. *Edible and Poisonous Mushroom of The World.* The New Zealand Institute for Crop and Food Research.

Johnson, R.A dan Wichern, D.W. 1998. *Applied Multivariate Statistical Analysis*, 5th ed. Prentice Hall Internasional, New Jersey.

Kass, G.V. 1980. *An Exploratory Technique for Investigating Large Quantities of Categorical Data.*Applied Statistic.29:119-127.

Lehman, Thomas and Eheler. *Responder profiling with CHAID and dependency analysis.* Universitt Jena.

Purbayu, Budi Santosa.2005. *Analisis Statistik dengan Microsoft Excel SPSS.* ANDI Yogyakarta.

Ratner, Bruce. 2003. *Statistical Modeling and Analisis for database Marketing.* Chapman Hall, Newyork.

SPSS, Inc. 1998. *Answertree 2.0 user's guide*. United States of America

Tsiatis, Anastasios.A. 2006.*Semiparametric, Theory and Missing Data*. Springer.

Webster, John and Weber, R. 2007. *Introduction to Fungi*. Cambridge University Press.

LAMPIRAN

lampiran 1: Peubah-peubah yang digunakan

NO	P.BEBAS	DESKRIPSI	KATEGORI
1.	**B1**	bentuk tudung	1= lonceng 2= kerucut 3= cembung 4= datar 5= tombol 6= cekung 7= *unknown* (tidak diketahui nilainya)
2.	**B2**	permukaan tudung	1= berserat 2= berlekuk 3= bersisik 4= halus 5= *unknown*
3.	**B3**	warna tudung	1= cokelat 2= abu-abu 3= kekuningan 4= kuning kecokelatan 5= hijau 6= merah muda 7= ungu 8= merah 9= putih 10= kuning 11= *unknown*
4.	**B4**	apakah jamur tergores tidak?	1= Ya 2= Tidak 3= *unknown*
5.	**B5**	aroma jamur	1= almond 2= adas 3= ter kayu

NO	P.BEBAS	DESKRIPSI	KATEGORI
6.	**B6**	pelekatan insang	4= amis 5= busuk 6= apak 7= tajam pedas 8= pedas 9= tidak beraroma 10= *unknown* 1= melekat 2= descending 3= free 4= bertakik 5= *unknown*
7.	**B7**	kerapatan garis-garis insang	1= dekat 2= rapat 3= jauh 4= *unknown*
8.	**B8**	ukuran insang	1= lebar 2=sempit 3= *unknown*
9.	**B9**	warna insang	1= hitam 2= cokelat tua 3= cokelat 4= abu-abu 5= kekuningan 6= hijau 7= jingga 8= merah muda 9= ungu 10= merah 11= putih 12= kuning 13= *unknown*
10.	**B10**	bentuk tangkai	1= membesar Nominal 2= lonjong 3= *unknown*
11.	**B11**	bentuk bagian bawah tangkai	1= umbi 2= club 3= cangkir 4= sama 5= rhizomorph 6= berakar 7= *unknown*

NO.	P.BEBAS	DESKRIPSI	KATEGORI
12.	**B12**	permukaan tangkai di atas cincin	1= berserat 2= bersisik 3= halus 4= sutera 5= *unknown*
13.	**B13**	permukaan tangkai di bawah cincin	1= berserat 2= bersisik 3= halus 4= sutera 5=*unknown*
14.	**B14**	warna tangkai di atas cincin	1= cokelat 2= abu-abu 3= kekuningan 4= kuning kecokelatan 5= jingga 6= merah muda 7= merah 8= putih 9= kuning 10=*unknown*
15.	**B15**	warna tangkai di bawah cincin	1= cokelat 2= abu-abu 3= kekuningan 4= kuning kecokelatan 5= jingga 6= merah muda 7= merah 8= putih 9= kuning 10=*unknown*
16.	**B16**	tipe membran pembungkus	1= parsial 2= universal 3=*unknown*
17.	**B17**	warna membran pembungkus	1= cokelat 2= jingga 3= putih
			4= kuning 5=*unknown*

NO.	PEUBAH	DESKRIPSI	KATEGORI
18.	**B18**	banyaknya cincin	1= tidak ada 2= satu 3= dua 4=*unknown*
19.	**B19**	tipe cincin	1= tidak ada 2= jaring laba-laba 3= evanescent 4= mengembang 5= besar 6= anting 7= pelepah 8= zone 9= *unknown*
20.	**B20**	warna cetakan spora	1= hitam 2= cokelat tua 3= cokelat 4= kekuningan 5= hijau 6= jingga 7= ungu 8= putih 9= kuning 10= *unknown*
21.	**B21**	populasi jamur	1= melimpah 2= bergerombol 3= banyak 4= tersebar 5= beberapa 6= tersendiri 7=*unknown*
22.	**B22**	habitat jamur	1= rumput 2= dedaunan 3= padang rumput 4= jalan setapak 5= perkotaan 6= sampah 7= kayu 8=*unknown*

lampiran 2:Tabel frekuensi peubah yang digunakan

Peubah	**Kategori**	**Jumlah**	**%**	**% Kumulatif**
STATUS	Dapat Dimakan	3916	48,2	48,2
	Beracun	4208	51,8	100
BENTUK TUDUNG	Lonceng	363	4,5	4,5
	Kerucut	4	0,0	4,5
	Cembung	2940	36,2	40.7
	Datar	2500	30,8	71,5
	Tombol	603	7,4	78,9
	cekung	29	0,4	79,3
	Missing	1685	20,7	100
PERMUKAAN TUDUNG	Berserat	1952	24	24
	Berlekuk	3	0,1	24,1
	Bersisik	2785	34,3	58,4
	Halus	2171	26,7	85,1
	Missing	1213	14,9	100
WARNA TUDUNG	Cokelat	1949	24,0	24,0
	Abu-abu	1515	18,6	42,6
	Kekuningan	132	1,6	44,2
	Kuning kecokelatan	34	0,4	44,6
	Hijau	14	0,2	44,8
	Merah muda	118	1,5	46,3
	Ungu	15	0,2	46,5
	Merah	1276	15,7	62,2
	Putih	868	10,7	72,9
	Kuning	898	11	83,9
	Missing	1305	16,1	100
TERGORES	Ya	2711	33,4	33,4
	Tidak	3702	45,5	78,9
	Missing	1711	21,1	100

Peubah	**Kategori**	**Jumlah**	**%**	**% Kumulatif**
AROMA	Almond	357	4,4	4,4
	Adas	359	4,4	8,8
	Ter kayu	158	1,9	10,7
	Amis	496	6,1	16,8
	Busuk	1726	21,2	38,0
	Apak	30	0,4	38,4
	Tajam pedas	221	2,7	41,1
	Pedas	488	6,0	47,1
	Tidak beraroma	2890	35,6	82,8
	Missing	1399	17,2	100
PELEKATAN INSANG	Melekat	177	2,2	2,2
	Free	6774	83,4	85,6
	Missing	1173	14,4	100
KERAPATAN	Dekat	5672	69,8	69,8
Garis-garis	Rapat	1099	13,5	83,3
Insang	*Missing*	1353	16,7	100
UKURAN INSANG	Lebar	4600	56,6	56,6
	Sempit	2116	26,0	82,7
	Missing	1408	17,3	100
WARNA INSANG	Hitam	367	4,5	4,5
	Cokelat tua	660	8,1	12,6
	Cokelat	955	11,8	24,4
	Abu-abu	672	8,3	32,7
	Kekuningan	1526	18,8	51,5
	Hijau	22	0,3	51,8
	Jingga	57	0,7	52,5
	Merah muda	1356	16,7	69,2
	Ungu	437	5,4	74,6
	Merah	90	1,1	75,7
	Putih	1087	13,4	89,1
	Kuning	78	1,0	89,9
	Missing	817	10,1	100
BENTUK TANGKAI	Membesar	2702	33,3	33,3
	Lonjong	3879	47,7	81,0
	Missing	1543	19,0	100
BENTUK BAGIAN	Umbi	3701	45,6	45,6
BAWAH TANGKAI	Club	543	6,7	52,3
	Sama	1093	13,5	65,8
	Berakar	187	2,3	68
	Missing	2600	32	100

Peubah	**Kategori**	**Jumlah**	**%**	**% Kumulatif**
PERMUKAAN TANGKAI DIATAS CINCIN	Berserat	368	4,5	4,5
	Bersisik	21	0,3	4,8
	Halus	1465	18	22,8
	Sutera	5385	66,3	89,1
	Missing	885	10,9	100
PERMUKAAN TANGKAI DIBAWAH CINCIN	Berserat	527	6,5	6,5
	Bersisik	246	3,0	9,5
	Halus	2099	25,8	35,3
	Sutera	4405	54,2	89,6
	Missing	847	10,4	100
WARNA TANGKAI DIATAS CINCIN	Cokelat	386	4,8	4,8
	Abu-abu	523	6,4	11,2
	Kekuningan	384	4,7	15,9
	Kuning kecokelatan	30	0,4	16,3
	Jingga	174	2,1	18,4
	Merah muda	1684	20,7	39,1
	Merah	87	1,1	40,2
	Putih	3911	48,1	88,3
	Kuning	6	0,1	88,4
	Missing	939	11,6	100
WARNA TANGKAI DIBAWAH CINCIN	Cokelat	491	6,0	6,0
	Abu-abu	541	6,7	12,7
	Kekuningan	411	5,1	17,8
	Kuning kecokelatan	29	0,4	18,2
	Jingga	164	2,0	20,2
	Merah muda	1773	21,8	42,0
	Merah	93	1,1	43
	Putih	3976	48,9	91,9
	Kuning	23	0,3	92,3
	Missing	623	7,7	100
TIPE MEMBRAN PEMBUNGKUS	Parsial	6696	82,4	82,4
	Missing	1428	17,6	100
WARNA MEMBRAN PEMBUNGKUS	cokelat	93	1,3	1,3
	Jingga	94	1,2	2,5
	Putih	7217	97,4	99,9
	Kuning	7	0,1	100

Peubah	**Kategori**	**Jumlah**	**%**	**% Kumulatif**
BANYAK CINCIN	Tidak ada	33	0,4	0,4
	Satu	6556	80,7	81,1
	Dua	540	6,6	87,8
	Missing	995	12,2	100
TIPE CINCIN	Tidak ada	30	0,4	0,4
	Evanescent	2354	29,0	29,4
	Mengembang	44	0,5	29,9
	Besar	1148	14,1	44
	Anting	3491	43,0	87
	Missing	1057	13,0	100
WARNA CETAKAN SPORA	Hitam	1687	20,8	20,8
	Cokelat tua	1394	17,2	38
	Cokelat	1778	21,9	59,8
	Kekuningan	40	0,5	60,3
	Hijau	62	0,8	61,1
	Jingga	44	0,5	61, 6
	Ungu	46	0,6	62,2
	Putih	2061	25,4	87,6
	Kuning	45	0,6	88,1
	Missing	967	11,9	100
POPULASI	Melimpah	361	4,4	4,4
	Bergerombol	285	3,5	7,9
	Banyak	349	4,3	12,2
	Tersebar	1109	13,7	25,9
	Beberapa	3367	41,4	67,3
	Tersendiri	1543	19,0	86,3
	Missing	1110	13,7	100
HABITAT	Rumput	1880	23,1	23,1
	Dedaunan	606	7,5	30,6
	Padang rumput	263	3,2	33,8
	Jalan setapak	935	11,5	45,3
	Perkotaan	327	4,0	49, 3
	Sampah	175	2,2	51,5
	Kayu	2743	33,8	85,3
	Missing	1195	14,7	100

lampiran 3: Tabel tabulasi silang antara peubah bebas dengan peubah terikat

Y/B1	1	2	3	4	5	6	Total
0	42	4	1346	1181	433	0	3006
1	321	0	1594	1319	170	29	3433
Total	363	4	2940	2500	603	29	6439

Y/B2	1	2	3	4	Total
0	655	3	1485	1208	3351
1	1297	0	1300	963	3560
Total	1952	3	2785	2171	6911

Y/B3	1	2	3	4	5	6	7	8	9	10	Total
0	878	657	95	10	0	76	0	762	257	557	3292
1	1071	858	37	24	14	42	15	514	611	341	3527
Total	1949	1515	132	32	14	118	15	1276	868	898	6819

Y/B4	1	2	Total
0	512	2495	3007
1	2199	1207	3406
Total	2711	3702	6413

Y/B5	1	2	3	4	5	6	7	8	9	Total
0	2	0	158	493	1721	29	221	482	109	3215
1	355	359	0	3	5	1	0	6	2781	3510
Total	357	359	158	496	1726	30	221	488	2890	6725

Y/B6	1	2	Total
0	16	3334	3350
1	161	3440	3601
Total	177	6774	6951

Y/B7	1	2	Total
0	3093	93	3186
1	2579	1006	3585
Total	5672	1099	6771

Y/B8	1	2	Total
0	1405	1869	3274
1	3195	247	3442
Total	4600	2116	6716

Y/B9	1	2	3	4	5	6	7	8	9	10	11	12	Total
0	55	479	106	451	1507	22	0	581	47	0	229	22	3499
1	312	181	849	221	19	0	57	775	390	90	858	56	3808
Total	367	660	955	672	1526	22	57	1356	437	90	1087	78	7307

Y/B10	1	2	Total
0	1309	1681	2990
1	1393	2198	3591
Total	2702	3879	6581

Y/B11	1	2	3	4	Total
0	1826	43	248	0	2117
1	1875	500	845	187	3407
Total	3701	543	1093	187	5524

Y/B12	1	2	3	4	Total
0	26	7	1292	2191	3516
1	342	14	173	3194	3723
Total	368	21	1465	5385	7239

Y/B13	1	2	3	4	Total
0	132	69	1963	1400	3564
1	395	177	136	3005	3713
Total	527	246	2099	4405	7277

Y/B14	1	2	3	4	5	6	7	8	9	Total
0	373	0	384	30	1	1170	1	1545	6	3510
1	13	523	0	0	173	514	86	2366	0	3675
Total	386	523	384	30	174	1684	87	3911	6	7185

Y/B15	1	2	3	4	5	6	7	8	9	Total
0	425	0	409	29	2	1212	0	1565	23	3665
1	66	541	2	0	162	561	93	2411	0	3836
Total	491	541	411	29	164	1773	93	3976	23	7501

Y/B16	1	Total
0	3167	3167
1	3529	3529
Total	6696	6696

Y/B17	1	2	3	4	Total
0	5	1	3623	7	3636
1	88	93	3594	0	3775
Total	93	94	7217	7	7411

Y/B18	1	2	3	Total
0	32	3435	87	3554
1	1	3122	453	3576
Total	33	6557	540	7130

Y/B19	1	2	3	4	5	Total
0	30	1474	0	1148	732	3384
1	0	880	44	0	2759	3683
Total	30	2354	44	1148	3491	7067

Y/B20	1	2	3	4	5	7	8	9	10	Total
0	214	1546	212	1	70	1	0	1771	2	3817
1	1564	53	1660	47	44	45	45	558	45	4017
Total	1778	1599	1872	48	71	45	45	2329	47	7834

Y/B21	1	2	3	4	5	6	Total
0	0	82	66	399	2090	592	3229
1	361	203	283	710	1277	951	3785
Total	361	285	349	1109	3367	1543	7014

Y/B22	1	2	3	4	5	6	7	Total
0	645	422	30	805	241	18	1007	3168
1	1235	184	233	130	86	157	1736	3761
Total	1880	606	263	935	327	175	2743	6929