

EKSPLORASI STRUKTUR DATA DENGAN METODE CHAID

Aam Alamudi¹, A. H. Wigena¹, Aunuddin¹

RINGKASAN

Metode CHAID adalah metode yang dikembangkan untuk menganalisis keterkaitan struktural dalam data hasil survey. Data tersebut biasanya meliputi satu atau beberapa peubah respon dan peubah-peubah penjelas yang pada umumnya kategorik. Dalam tulisan ini dijelaskan metode CHAID sebagai suatu metode eksploratif dengan contoh kasus penelusuran penciri status keamanan pangan rumah tangga pedesaan menurut peubah-peubah keadaan rumah tangga.

PENDAHULUAN

Struktur keterkaitan antar peubah sering menjadi perhatian dalam berbagai penelitian. Dalam masalah tertentu struktur tersebut mungkin dapat dirumuskan secara jelas sehingga model keterkaitan antar peubah dapat disusun berdasarkan kerangka teoritis tertentu. Dalam hal ini beberapa model diuji dan dipilih model-model yang memuaskan.

Pada kasus lain struktur keterkaitan antar peubah belum difahami secara jelas. Alternatif model sangat banyak sehingga penelusurannya menjadi sulit. Ini terutama apabila penelitian tersebut menyangkut peubah-peubah kategorik yang banyak, yang biasanya terjadi pada penelitian-penelitian dengan data survey. Pada kasus seperti ini diperlukan suatu metode yang efisien untuk menetapkan model yang dapat dipilih.

Metode CHAID adalah salah satu tipe dari metode AID yang diajukan oleh Kass (1980) untuk menyelidiki struktur keterkaitan antar peubah, yaitu antara peubah respon kategorik dengan peubah-peubah penjelas kategorik. Dalam tulisan ini dijelaskan mengenai metode CHAID sebagai suatu metode eksploratif dengan contoh kasus penelusuran penciri status keamanan pangan rumah tangga pedesaan. Metode CHAID di sini digunakan untuk menelusuri peubah-peubah apa saja yang menjadi penciri status keamanan pangan serta peranan peubah-peubah penciri tersebut dalam

menjelaskan status keamanan pangan rumah tangga pedesaan (Alamudi, 1996). Model hubungan antara faktor keadaan rumah tangga dengan status keamanan pangan rumah tangga dengan demikian menjadi lebih mudah disusun.

METODE AID

Metode AID (*Automatic Interaction Detection*) adalah metode yang dikembangkan untuk menganalisis keterkaitan struktural pada data hasil survey (Fielding, 1977, dalam O'Muircheartaigh dan Payne, 1977). Data tersebut biasanya meliputi satu atau beberapa peubah respon dan peubah-peubah penjelas yang pada umumnya kategorik. Ada dua tipe peubah penjelas yang dikenal dalam hal ini, yaitu peubah "monotonik" yang nilai-nilainya bersifat ordinal dan peubah "bebas" yang nilai-nilainya bersifat nominal (Kass, 1980).

Metode AID menganalisis suatu gugus data dengan cara memisahkannya menjadi beberapa kelompok secara bertahap (Fielding, 1977, dalam O'Muircheartaigh dan Payne, 1977). Tahap pertama, seluruh data dibagi menjadi beberapa anak gugus berdasarkan salah satu peubah yang dipilih sedemikian rupa dengan memaksimumkan kriteria tertentu. Masing-masing anak gugus kemudian diperiksa kembali secara terpisah dan dibagi lagi berdasarkan peubah lainnya, demikian selanjutnya sampai tercapai kriteria berhenti tertentu. Dengan cara demikian maka diperoleh kelompok-kelompok

¹ Dosen pada Jurusan Statistika FMIPA-IPB

pengamatan yang mempunyai ciri-ciri respon dan penjas tertentu sehingga keterkaitan diantara peubah-peubah tersebut menjadi jelas.

Hasil analisis dengan metode AID adalah suatu dendogram pemisahan. Dari dendogram ini dapat diperoleh tiga tipe informasi, yaitu (Fielding, 1977, dalam O'Muircheartaigh dan Payne, 1977):

1. **Pengelompokan pengamatan;** pengamatan dikelompokkan kedalam kelompok-kelompok yang relatif homogen dalam kaitannya dengan nilai-nilai peubah penjas dan peubah respon.
2. **Asosiasi antar peubah;** yaitu kecenderungan nilai peubah penjas tertentu berpadanan dengan nilai peubah penjas yang lain.
3. **Interaksi antar peubah penjas;** yaitu peranan silang dua peubah penjas dalam pemisahan pengamatan menurut peubah respon.

Meskipun telah tercatat sejak tahun 1959 dengan nama yang berbeda-beda, versi komputerisasi dari metode ini diperkenalkan pada tahun 1963 oleh J. N. Morgan dan J. A. Sonquist (Anonim, 1982). Kata "*automatic*" dalam AID merujuk kepada penggunaan komputer untuk membuat semua keputusan tentang penjas mana yang digunakan, kapan dan bagaimana digunakannya (Anonim, 1982). Penggunaan komputer pada metode ini sangat penting artinya terutama pada analisis yang melibatkan peubah yang banyak dan kategori yang banyak pula.

Berbagai tipe AID dibedakan menurut peubah respon yang dianalisisnya (Anonim, 1982). Metode AID baku (*standard AID*) digunakan untuk peubah respon interval dengan jumlah kuadrat sebagai kriteria pemilihan peubah. AID Chi-kuadrat (CHAID) dan AID Theta (THAID) digunakan untuk peubah respon nominal atau ordinal. Sebagai perluasan dari AID baku untuk peubah respon berganda dikenal AID peubah ganda (*Multivariate AID, MAID*).

METODE CHAID

Metode CHAID menggunakan kriteria chi-kuadrat dalam pengoperasiannya (Kass, 1980). Dibandingkan dengan THAID, metode CHAID lebih unggul karena (Kass, 1980):

1. Untuk peubah respon dikotom, hasil analisis THAID ternyata berbeda dengan hasil AID.

2. Perilaku teoritik bagi kriteria θ dalam THAID belum banyak diketahui.

Algoritma CHAID adalah sebagai berikut (Kass, 1980):

Tahap 1. Untuk setiap peubah penjas, buat tabulasi silang kategori-kategori peubah penjas dengan kategori-kategori peubah respon.

Tahap 2. Cari pasangan kategori yang mana sub-tabel $2 \times d$ (d adalah banyaknya kategori peubah respon) dari tabel tersebut memiliki angka uji paling kecil. Jika angka ini tidak mencapai nilai kritis, gabungkan kedua kategori ini menjadi satu kategori campuran, dan ulangi tahap ini sehingga angka uji terkecil sub-tabel $2 \times d$ pasangan kategori (kategori campuran) peubah penjas melampaui nilai kritis.

Tahap 3. Untuk setiap kategori campuran yang berisi tiga atau lebih kategori asal, cari pemisahan biner yang memiliki angka uji paling besar. Jika angka uji ini memiliki angka uji di luar nilai kritis, buatlah pemisahan tersebut dan kembali ke **Tahap 2**.

Tahap 4. Hitung tarap nyata untuk masing-masing tabulasi silang yang baru dan perhatikan diantaranya yang memiliki angka uji paling besar, sebut sebagai tabulasi dengan **tarap nyata terbaik**. Jika angka ini lebih besar dari nilai kritis, bagilah data menurut kategori tersebut.

Tahap 5. Jika terjadi pemisahan pada **Tahap 4**, kembali ke **Tahap 1** untuk setiap bagian data hasil pemisahan.

Angka uji dan nilai kritis yang dimaksudkan pada algoritme di atas adalah statistik dan kriteria uji **Chi-kuadrat** apabila tidak terjadi pengurangan tabel kontingensi dari tabel asal. Apabila terjadi pengurangan, dengan c kategori peubah penjas menjadi r kategori ($r < c$), maka nilai kritik tersebut dikalikan dengan suatu pengali (lihat Kass, 1980) sesuai dengan tipe peubahnya. Metode CHAID telah diimplementasikan dalam makro SAS %TCHAD yang ditulis dalam modul SAS/IML pada SAS versi 6.10 (SAS Institute, 1993).

PENELUSURAN PENCIRI STATUS KEAMANAN PANGAN

Data Yang Dianalisis

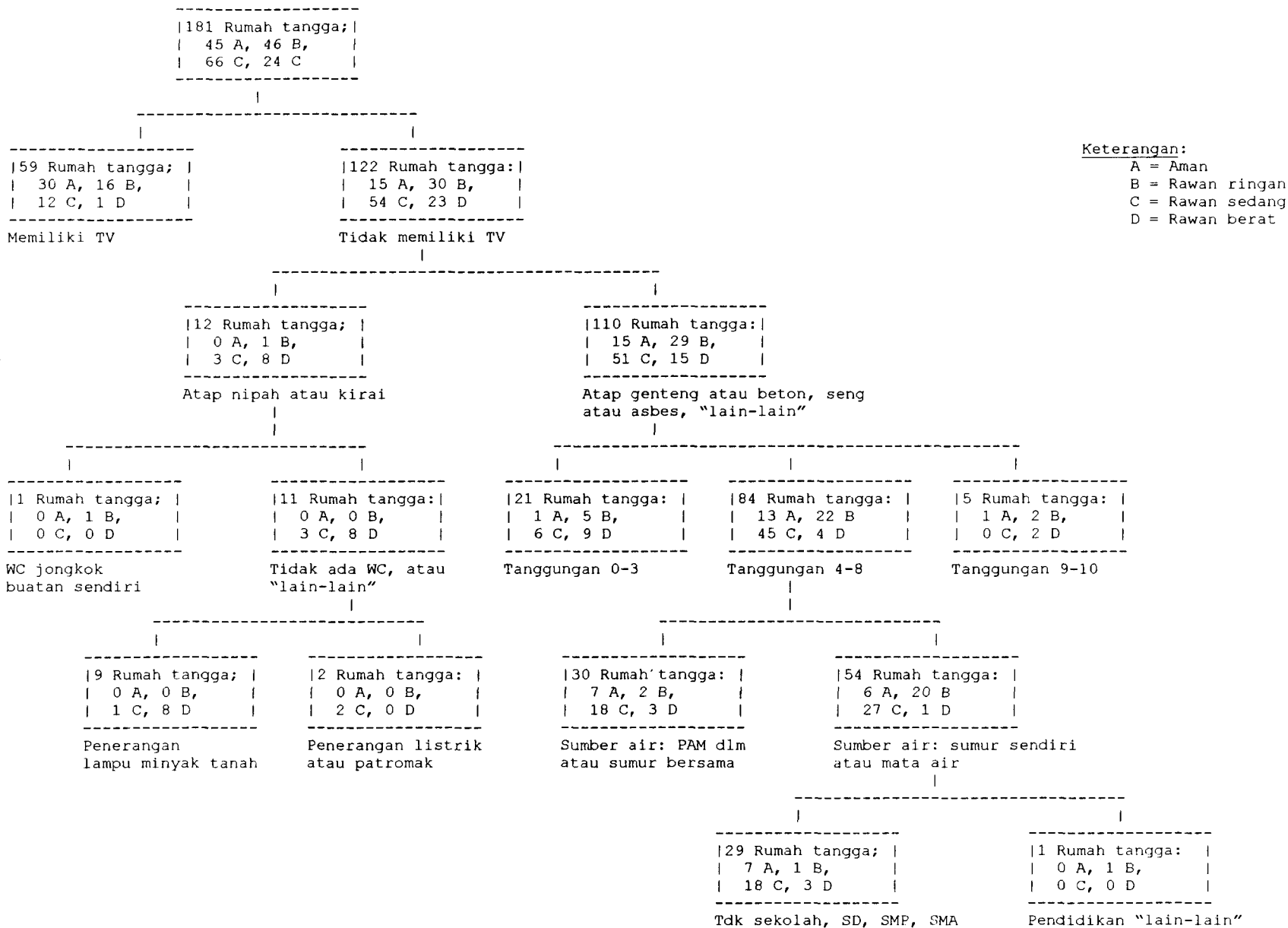
Data yang dianalisis adalah data yang dikumpulkan pada survey "Manajemen Informasi Keamanan Pangan" (Tim Jurusan Statistika, 1993). Data meliputi 181 rumah tangga di Leuwiliang (Kab. Bogor), Parungkuda (Kab. Sukabumi), dan Pameungpeuk (Kab. Garut). Rumah tangga-rumah tangga tersebut dikelaskan kedalam kategori keamanan *status-sekarang* berdasarkan kriteria Taylor (1991), yaitu: 0 (*aman*), 1 (*rawan ringan*), 2 (*rawan sedang*), dan 3 (*rawan berat*). Faktor-faktor keadaan rumah tangga yang digunakan sebagai penjelas adalah sebagai berikut:

1. *Kepemilikan rumah dan pekarangan.*
 - 1 = *Sepenuhnya milik sendiri*
 - 2 = *Milik sendiri masih menyicil*
 - 3 = *Milik keluarga (orang tua/famili)*
 - 4 = *Sewa*
 - 5 = *Pengurus (pamolah) tanah*
 - 6 = *lain-lain*
2. *Lamanya tinggal, termasuk generasi sebelumnya: 1 tahun, 2 tahun,*
3. *Jenis atap yang digunakan.*
 - 1 = *Genteng atau beton*
 - 2 = *Seng atau asbes*
 - 3 = *Nipah atau kirai*
 - 4 = *Lain-lain*
4. *Sumber air untuk minum dan memasak.*
 - 1 = *PAM dalam rumah*
 - 2 = *PAM umum*
 - 3 = *Sumur sendiri*
 - 4 = *Sumur bersama*
 - 5 = *Lambang atau kubangan*
 - 6 = *Sungai atau mata air*
 - 7 = *Lain-lain*
5. *WC yang digunakan.*
 - 1 = *Tidak ada fasilitas khusus*
 - 2 = *WC jongkok buatan sendiri*
 - 3 = *WC jongkok buatan pabrik*
 - 4 = *WC duduk*
 - 5 = *Lain-lain.*
6. *Jenis penerangan yang digunakan.*
 - 1 = *Listrik*
 - 2 = *Petromaks*
 - 3 = *Lampu gas*
 - 4 = *Lampu minyak tanah*
 - 5 = *Lilin*
 - 6 = *Lain-lain*
7. *Kepemilikan radio: 1 = Ya, 2 = Tidak.*
8. *Kepemilikan kendaraan: 1 = Ya, 2 = Tidak.*
9. *Kepemilikan TV: 1 = Ya, 2 = Tidak.*
10. *Banyaknya tanggungan rumah tangga: 0 jiwa, 1 jiwa, 2 jiwa, 3 jiwa, ...*
11. *Pendidikan tertinggi yang dicapai anggota rumah tangga:*
 - 0 = *Tidak sekolah*
 - 1 = *SD*
 - 2 = *Tidak tamat SMP*
 - 3 = *Tamat SMP*
 - 4 = *Tidak tamat SLTA*
 - 5 = *Tamat SLTA*
 - 6 = *Tidak tamat sekolah kejuruan*
 - 7 = *Tamat sekolah kejuruan*
 - 8 = *Tidak tamat S0/S1*
 - 9 = *Tamat S0/S1*
 - 10 = *Tidak tamat S2/S3*
 - 11 = *Tamat S2/S3*
 - 12 = *Lain-lain*
12. *Luas lahan yang diakses: dalam hektar.*
Dalam analisis Lamanya tinggal, Banyaknya tanggungan rumah tangga, dan Luas lahan yang diakses diperlakukan sebagai peubah-peubah monotonik; Kepemilikan rumah dan pekarangan, Jenis atap yang digunakan, Sumber air untuk minum dan memasak, WC yang digunakan, Jenis penerangan yang digunakan, Kepemilikan radio, Kepemilikan kendaraan dan Kepemilikan TV sebagai peubah-peubah bebas; dan Pendidikan tertinggi yang dicapai anggota rumah tangga sebagai peubah float, dengan "lain-lain" sebagai peubah float.

Dendogram Pemisahan CHAID

Dendogram pemisahan CHAID untuk status keamanan pangan pada peubah-peubah keadaan rumah tangga tampak pada Gambar 1. Tarap nyata terbaik pada masing-masing tahap pemisahan tersebut disajikan pada Lampiran 1. Dendogram pemisahan tersebut diperoleh dari analisis dengan nilai kritis yang ditetapkan pada tarap nyata 0.05.

Peubah-peubah yang berperan dalam pengelompokan tampak pada dendogram, yaitu (1) *Kepemilikan TV*, (2) *Jenis Atap*, (3) *Jenis WC*, (4) *Banyaknya tanggungan rumah tangga*, (5) *Jenis penerangan*, (6) *Sumber air untuk minum dan memasak*, dan (7) *Pendidikan tertinggi yang dicapai anggota rumah tangga*. Kepemilikan TV adalah faktor pertama yang



Gambar 1. Dendrogram Pemisahan Rumah Tangga Menurut Status KeAan Pangan dan Peubah-peubah Keadaan Rumah Tangga Pedesaan ($\alpha = 0.05$).

dipilih untuk memisahkan rumah tangga menjadi dua kelompok; yaitu kelompok yang memiliki TV dan kelompok yang tidak memiliki TV. Kelompok pertama meliputi 59 rumah tangga yang terdiri dari 30 *aman*, 16 *rawan ringan*, 12 *rawan sedang*, dan 1 *rawan berat*. Kelompok kedua meliputi 122 rumah tangga yang terdiri dari 15 rumah tangga *aman*, 30 *rawan ringan*, 54 *rawan sedang* dan 23 *rawan berat*.

Kelompok yang tidak memiliki TV berikutnya dibagi menjadi dua kelompok berdasarkan jenis atap yang digunakan. Kelompok pertama adalah yang menggunakan atap nipah atau kirai, meliputi 12 rumah tangga yang terdiri dari 1 *rawan ringan*, 3 *rawan sedang* dan 8 *rawan berat*. Kelompok kedua adalah yang menggunakan atap genteng atau beton, seng atau asbes, atau "lain-lain". Kelompok ini meliputi 110 rumah tangga yang terdiri dari 15 *aman*, 29 *rawan ringan*, 51 *rawan sedang*, dan 15 *rawan berat*. Kelompok yang tidak memiliki TV dan menggunakan atap nipah atau kirai berikutnya dipisahkan berdasarkan jenis WC yang digunakan menjadi kelompok yang menggunakan WC jongkok buatan sendiri dan kelompok yang tidak memiliki fasilitas WC atau menggunakan WC "lain-lain". Kelompok yang menggunakan atap genteng atau beton, seng atau asbes, atau "lain-lain" dipisahkan berdasarkan banyaknya tanggungan rumah tangga menjadi tiga kelompok; yaitu yang tanggungannya 0-3, 4-8, dan kelompok yang tanggungannya 9 atau 10. Terakhir adalah pengelompokan rumah tangga yang tidak memiliki TV, menggunakan atap genteng atau beton, seng atau asbes, atau "lain-lain", memiliki tanggungan 4-8 dan menggunakan PAM dalam rumah atau sumur bersama berdasarkan pendidikan tertinggi yang dicapai anggota rumah tangga.

Pada dendrogram tampak adanya asosiasi antara jenis atap dengan jenis WC yang digunakan, yaitu bahwa rumah tangga yang menggunakan atap nipah atau kirai pada umumnya tidak memiliki WC atau menggunakan WC "lain-lain". Tampak pula adanya asosiasi antara jenis WC dengan jenis penerangan, yaitu bahwa rumah tangga yang tidak mempunyai WC atau menggunakan WC "lain-lain" pada umumnya menggunakan penerangan lampu minyak tanah.

Hasil analisis CHAID menunjukkan bahwa masing-masing peubah penjelas tidak berinteraksi dengan peubah penjelas lain. Pada dendrogram tidak tampak adanya pengaruh silang antar peubah karena peubah-peubah penjelas yang berperan pada masing-masing percabangan berbeda-beda.

Penciri Status Keamanan Pangan

Dua kelompok yang terbentuk pada tahap pertama pemisahan CHAID adalah kelompok yang memiliki TV yang kebanyakan aman dan rawan ringan, dan kelompok yang tidak memiliki TV yang kebanyakan rawan sedang dan rawan berat. Pada kelompok yang memiliki TV, pemisahan lebih lanjut tidak terjadi karena **angka nyata terbaik** pada kelompok ini lebih besar dari 0.05, yaitu 0.6561 (lihat Lampiran 1). Ciri status keamanan pangan pada kelompok ini tidak dapat dirinci lebih lanjut. Pada kelompok yang tidak memiliki TV terjadi pemisahan lebih lanjut dan ciri keamanan pangan tampak lebih jelas.

Rumah tangga yang rawan berat pada kelompok yang tidak memiliki TV dicirikan oleh keadaan yang serba kekurangan, yaitu menggunakan penerangan lampu minyak tanah atau lilin, tidak memiliki WC atau menggunakan WC lain-lain, dan menggunakan atap nipah atau kirai. Pada dendrogram tampak bahwa rumah tangga-rumah tangga yang rawan berat terhimpun pada kelompok yang memiliki ciri-ciri tersebut. Rumah tangga yang aman dicirikan oleh keadaan rumah tangga yang cukup baik, yaitu menggunakan atap genteng, beton, seng atau asbes, atau "lain-lain". Tampak bahwa rumah tangga-rumah tangga yang aman terhimpun pada kelompok yang menggunakan atap genteng atau beton, seng atau asbes, atau "lain-lain".

Penciri rawan sedang dan rawan ringan tidak dapat dipisahkan secara tegas. Banyaknya tanggungan 4-8 menjadi penciri kedua status tersebut. Sumber air menunjukkan status rawan ringan pada kelompok yang tanggungannya 4-8, tetapi tidak menunjukkan status rawan sedang secara jelas. Demikian pula dengan pengelompokan berikutnya berdasarkan tingkat pendidikan anggota rumah tangga. Tidak jelasnya pemisahan ciri status rawan ringan dan rawan sedang ini tampaknya disebabkan oleh tidak jelasnya kedua kriteria status ini (Lihat Taylor dan Phillips, 1992). Ketidakjelasan ini telah dikemukakan oleh Murnihati (1994) dan Sarwono (1994), yang menyatakan baurnya ciri status rawan ringan dan rawan sedang.

Terhimpunnya rumah tangga-rumah tangga yang rawan berat pada kelompok yang tanggungannya 0-3 tidak dapat dijelaskan. Kenyataan ini bertentangan dengan kenyataan umum yang dapat diterima, yaitu bahwa rumah tangga yang tanggungannya kecil semestinya tidak rawan. Keadaan yang bertentangan ini mungkin dapat dijelaskan oleh peubah lain yang belum disertakan dalam analisis.

KESIMPULAN

Metode CHAID merupakan metode yang cukup efisien untuk mengungkapkan keterkaitan struktural dalam data. Metode CHAID dalam kasus ini secara efisien dapat:

1. Menentukan peubah-peubah penjelas yang berperan dalam kaitannya dengan peubah respon; Diperolehnya tujuh peubah yang berperan dalam pemisahan pengamatan, dari dua belas peubah yang dipertimbangkan sebelumnya
2. Memberikan informasi mengenai adanya asosiasi dan interaksi antar peubah penjelas, yang akan sangat berguna dalam penyusunan model.

Dengan diperolehnya dendogram pemisahan CHAID maka analisis selanjutnya menjadi semakin terarah.

DAFTAR PUSTAKA

- Alamudi, A. 1996. Eksplorasi Penciri Fisik Status Keamanan Pangan Rumah Tangga dengan Metode CHAID. Tesis S2. Program Pascasarjana Institut Pertanian Bogor. Bogor.
- Anonim. 1982. Encyclopedia of Statistical Sciences, Vol. I. John Wiley & Sons. New York. Chicester, Brisbane, Toronto, Singapore.
- O'Muirjeartaigh, C. A. dan C. Payne. 1977. The Analysis of Survey Data Vol. I. Exploring Data Structures. John Wilwy & Sons. London, New York, Sydney, Toronto.
- Karmiati, M. 1994. Kajian Perubahan Status Keamanan Pangan Rumah Tangga Dalam Jangka Pendek, Skripsi S1. Jurusan Statistika FMIPA-IPB. Bogor.
- Kass, G. V. 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *App. Statist.* 29, No. 2:119-127.
- Murnihati. 1992. Menelaah Hubungan Antara Status Keamanan Pangan dengan Ciri-ciri Fisik Rumahtangga Pedesaan, Skripsi S1. Jurusan Statistika FMIPA-IPB. Bogor.
- Sarwono, 1994. Pemodelan Hubungan Peubah Penciri Fasilitas Rumahtangga terhadap Status Keamanan Pangan, Skripsi S1. Jurusan Statistika FMIPA-IPB. Bogor.
- SAS Institute. 1993. SAS Sample Library; TCHAIDA, TCHAID Macro Alpha-version. SASWSS.
- Taylor, D. S. 1991. Assessing Household Food Insecurity: A Framework and Questionnaire. University of Guelph.
- Taylor, D. S. dan T. Phillips. 1992. Summary of Food Security Analysis Using the Phase II Questionnaire. SEAMEO-SEARCA, Laguna, Philippines.
- Tim Jurusan Statistika. 1993. Manajemen Informasi dan Sistem Pakar untuk Masalah Keamanan pangan Keluarga di Daerah Pedesaan. Institut Pertanian Bogor. Bogor.