

BAB III

CHI-SQUARED AUTOMATIC INTERACTION DETECTION (CHAID)

3.1 Metode CHAID

Metode CHAID diperkenalkan oleh G. V. Kass pada tahun 1980. Prosedur pada metode CHAID merupakan bagian dari teknik *Automatic Interaction Detection* (AID) dengan menggunakan statistik *chi-square* sebagai alat utamanya (Gallagher *et al.*, 1990). Secara umum, metode CHAID bekerja dengan mempelajari hubungan antara variabel dependen dengan beberapa variabel independen kemudian mengklasifikasikan sampel berdasarkan hubungan tersebut.

Menurut Gallagher *et al.* (1990), CHAID merupakan suatu teknik iteratif yang menguji satu per satu variabel independen yang digunakan dalam klasifikasi dan menyusunnya berdasarkan pada tingkat signifikansi statistik *chi-square* terhadap variabel dependennya. Eherler dan Lehmann (2001) menyatakan bahwa CHAID adalah sebuah metode untuk mengklasifikasikan data kategori dimana tujuan dari prosedurnya adalah untuk membagi rangkaian data menjadi subgrup-subgrup berdasarkan variabel dependennya.

CHAID digunakan untuk membentuk segmentasi yang membagi sebuah sampel menjadi dua atau lebih kelompok yang berbeda berdasarkan sebuah kriteria tertentu. Hal ini kemudian diteruskan dengan membagi kelompok-kelompok tersebut menjadi kelompok yang lebih kecil berdasarkan variabel-variabel independen yang lain. Proses tersebut terus berlanjut sampai tidak ditemukan lagi variabel-variabel independen yang signifikan secara statistik (Kunto dan Hasana, 2006). Hasil dari pengklasifikasian menggunakan metode CHAID akan ditampilkan dalam bentuk diagram pohon (dendogram).

3.2 Variabel dalam Metode CHAID

Variabel dependen dan independen dalam metode CHAID merupakan variabel kategorik. Variabel independen dalam metode CHAID dibedakan menjadi tiga bentuk yang berbeda, yaitu: (Kass, 1980; Gallagher *et al.*, 1990)

Rizka Ismi Nuraini, 2017

***APLIKASI METODE CHI-SQUARED AUTOMATIC INTERACTION DETECTION (CHAID) PADA
KLASIFIKASI MINAT MASYARAKAT UNTUK MENGGUNAKAN INTERNET BANKING*** Universitas
Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

1. Variabel Independen Monotonik (*Monotonic Predictor*)

Kategori-kategori pada variabel ini dapat dikombinasikan atau digabungkan oleh CHAID hanya jika keduanya berdekatan satu sama lain, yaitu variabel-variabel yang kategorinya mengikuti urutan aslinya (data ordinal). Sebagai contoh: usia dan pendapatan.

2. Variabel Independen Bebas (*Free Predictor*)

Kategori-kategori pada variabel ini dapat dikombinasikan atau digabungkan meskipun keduanya berdekatan maupun tidak berdekatan (data nominal). Sebagai contoh: pekerjaan, kelompok etnik, dan area geografis.

3. Variabel Independen Mengambang (*Floating Predictor*)

Kategori-kategori pada variabel ini dapat diperlakukan seperti monotonik kecuali untuk kategori yang *missing value*, yang dapat dikombinasikan dengan kategori manapun.

3.3 Algoritme CHAID

Secara garis besar, langkah-langkah algoritme CHAID dapat dibagi menjadi tiga tahap, yaitu penggabungan (*merging*), pemisahan (*splitting*), dan penghentian (*stopping*) (Kunto dan Hasana, 2006 berdasarkan Magidson dalam Bagozzi (1994)).

3.3.1 Tahap Penggabungan (*Merging*)

Tahap penggabungan merupakan tahap kesatu dalam algoritme CHAID. Pada tahap ini, akan diperiksa signifikansi dari masing-masing kategori variabel independen terhadap variabel dependen. Berikut langkah-langkahnya: (Permana, 2011)

1. Untuk setiap variabel independen, bentuk tabel kontingensi dua arah dengan variabel dependennya.
2. Hitung statistik uji *chi-square* untuk setiap pasang kategori yang dapat dipilih untuk digabungkan menjadi satu, untuk menguji kebebasannya dalam sebuah sub tabel kontingensi $2 \times j$ yang dibentuk oleh sepasang kategori tersebut dengan variabel dependennya yang memiliki sebanyak j kategori.

3. Untuk setiap nilai *chi-square* berpasangan, hitung *p-value* (nilai *p*) berpasangan bersamaan. Diantara pasangan-pasangan yang tidak signifikan, gabungkan sebuah pasangan kategori yang paling mirip (yaitu pasangan yang memiliki nilai *chi-square* berpasangan terkecil dan *p-value* terbesar) menjadi sebuah kategori tunggal dan kemudian lanjutkan ke langkah nomor 4.
4. Periksa kembali kesignifikansian kategori baru setelah digabungkan dengan kategori lainnya dalam variabel independen. Jika masih ada pasangan yang belum signifikan, maka ulangi langkah nomor 3, tetapi jika semua sudah signifikan, maka lanjutkan ke langkah berikutnya.
5. Berdasarkan tabel hasil penggabungan, hitung *p-value* dengan menggunakan koreksi Bonferroni.

3.3.2 Tahap Pemisahan (*Splitting*)

Tahap kedua dari algoritme CHAID adalah pemisahan. Pada tahap ini, akan dipilih variabel independen terbaik sebagai *split node* (simpul pemisah) berdasarkan *p-value* pada tahap penggabungan. Berikut langkah-langkahnya: (Permana, 2011)

1. Pilih variabel independen yang memiliki *p-value* terkecil (paling signifikan), kemudian gunakan variabel independen ini sebagai *split node*.
2. Jika terdapat variabel independen dengan *p-value* kurang dari atau sama dengan tingkat signifikansi alpha ($p\text{-value} \leq \alpha$), maka variabel independen ini digunakan sebagai *split node*. Jika tidak ada variabel independen dengan *p-value* yang signifikan, maka pemisahan tidak dilakukan dan *node* ditentukan sebagai *terminal node* (*node* akhir).

3.3.3 Tahap Penghentian (*Stopping*)

Tahap penghentian dilakukan dengan aturan sebagai berikut: (Permana, 2011)

1. Tidak ada lagi variabel independen yang signifikan menunjukkan perbedaan terhadap variabel dependen.
2. Jika pertumbuhan pohon sekarang sudah mencapai batas nilai maksimum dari spesifikasi pohon, maka proses pertumbuhan akan berhenti. Misalkan

ditetapkan batas kedalaman (*depth*) pertumbuhan pohon klasifikasi adalah tiga, ketika pertumbuhan pohon sudah mencapai kedalaman tiga, pertumbuhan pohon klasifikasi akan dihentikan.

3. Jika ukuran dari *child node* kurang dari ukuran *child node* minimum spesifikasi atau berisi pengamatan-pengamatan dengan banyak yang terlalu sedikit, maka *node* tidak akan di-*split*. Misalkan ditetapkan ukuran minimal *child node* adalah 20, ketika *splitting* menghasilkan ukuran kurang dari 20, maka *node* tersebut tidak akan di-*split*.

3.4 Koreksi Bonferroni

Koreksi Bonferroni adalah suatu proses koreksi yang digunakan ketika beberapa uji statistik untuk kebebasan atau ketidakbebasan dilakukan secara bersamaan (Sharp *et al.* (2002) dalam Kunto dan Hasana, 2006). Koreksi Bonferroni ini biasa digunakan dalam perbandingan berganda.

Ketika terdapat sebanyak M uji perbandingan yang sudah dikatakan bebas satu sama lain, peluang untuk melakukan kesalahan tipe satu (menolak hipotesis yang seharusnya diterima) atau α (dalam satu atau lebih uji-uji tersebut) akan sama dengan 1 dikurangi peluang untuk tidak melakukan kesalahan tipe satu dalam uji-uji tersebut, dimana nilainya akan lebih dari α yang telah ditentukan. Secara matematis, dapat dirumuskan: (Bagozzi (1994), dalam Kunto dan Hasana, 2006)

$$1 - (1 - \alpha)^M > \alpha \quad (3.1)$$

dengan

M : pengali Bonferroni

α : kesalahan tipe I

Andaikan variabel independen memiliki c kategori, setelah dilakukan penggabungan variabel tersebut berkurang menjadi r kategori, sehingga perkalian Bonferroni adalah banyaknya cara yang mungkin dimana c kategori dapat digabungkan menjadi r kategori. Dengan demikian, *p-value* dari uji *chi-square* untuk independensi yang baru merupakan perkaliannya dengan pengali Bonferroni sesuai dengan jenis variabelnya (Gallagher *et al.*, 1990).

Berikut adalah pengali Bonferroni untuk masing-masing jenis variabel independen: (Gallagher *et al.*, 1990)

1. Variabel Independen Monotonik (*Monotonic Predictor*)

$$M = \binom{c-1}{r-1} \quad (3.2)$$

2. Variabel Independen Bebas (*Free Predictor*)

$$M = \sum_{i=0}^{r-1} (-1)^i \frac{(r-1)^c}{i!(r-i)!} \quad (3.3)$$

3. Variabel Independen Mengambang (*Floating Predictor*)

$$M = \binom{c-2}{r-2} + r \binom{c-2}{r-1} \quad (3.4)$$

dengan

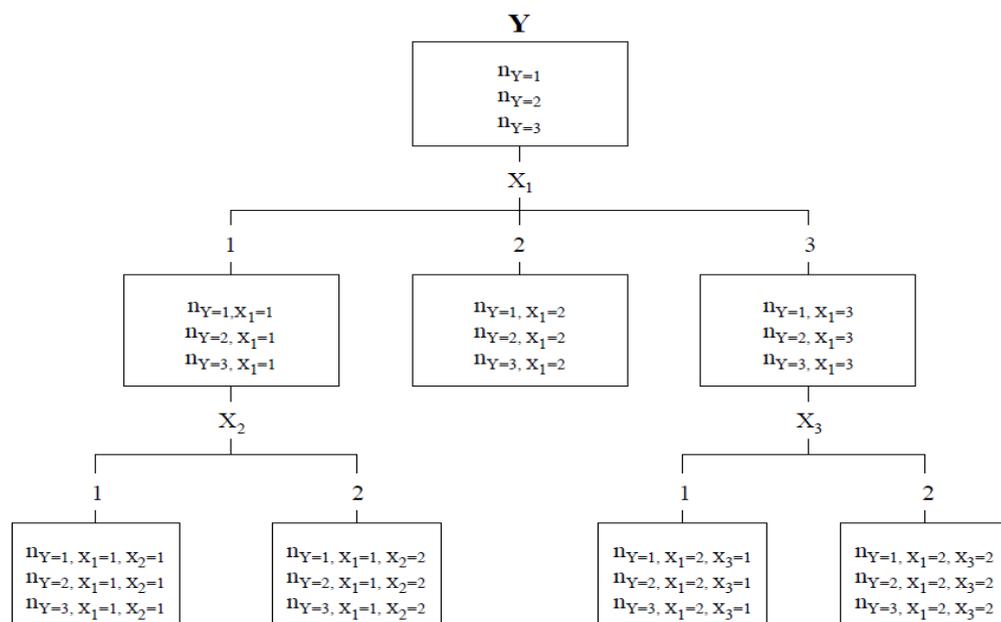
M : pengali Bonferroni

c : banyaknya kategori variabel independen awal

r : banyaknya kategori variabel independen setelah penggabungan

3.5 Diagram Pohon Klasifikasi CHAID

Metode CHAID akan menghasilkan sebuah diagram pohon klasifikasi yang merupakan gambaran dari hasil pembentukan segmen. Berikut adalah gambar diagram pohon CHAID secara umum:



Gambar 3.1 Diagram Pohon dalam Analisis CHAID

Sumber: *Responder Profiling with CHAID and Dependency Analysis*, Eherler and Lehmann, 2001.

Berdasarkan Myers (1996) (Kunto dan Hasana, 2006), diagram pohon CHAID mengikuti aturan “dari atas ke bawah” (*top-down stopping rule*), dimana diagram pohon disusun mulai dari kelompok induk (*parent node*), berlanjut dibawahnya sub kelompok (*child node*) yang berturut-turut dari hasil pembagian kelompok induk berdasarkan kriteria tertentu. *Node* pada ujung pohon yang sudah tidak terdapat percabangan lagi disebut *terminal node*. Masing-masing *node* dari diagram pohon ini menggambarkan sub kelompok dari sampel yang diteliti dan berisi keseluruhan sampel dan frekuensi absolut n_i untuk setiap kategori yang disusun diatasnya (Eherler and Lehmann, 2001).

Terdapat istilah kedalaman (*depth*) pada pohon klasifikasi CHAID yang berarti banyaknya tingkatan *node-node* sub kelompok sampai ke bawah pada *node* sub kelompok yang terakhir. Pada kedalaman kesatu, sampel dibagi oleh X_1 sebagai variabel independen terbaik untuk variabel dependen berdasarkan uji *chi-square*. Setiap *node* berisi informasi tentang frekuensi variabel Y , sebagai variabel dependen, yang merupakan bagian dari sub kelompok yang dihasilkan berdasarkan kategori yang disebutkan (X_1). Pada kedalaman kedua (*node* X_2 dan X_3) merupakan pembagian dari X_1 (untuk *node* ke-1 dan ke-3). Dengan cara yang sama, sampel selanjutnya dibagi oleh variabel independen yang lain, yaitu X_2 dan X_3 , dan selanjutnya menjadi sub kelompok pada *node* ke-4, 5, 6, dan 7 (Eherler and Lehmann, 2001; Kunto dan Hasana, 2006). Pada masing-masing *node* ditampilkan persentase responden untuk setiap kategori dari variabel dependen dan ditunjukkan pula jumlah total responden untuk masing-masing *node* (Myers (1996) dalam Kunto dan Hasana, 2006).