

Klasifikasi Variabel Penentu Kelulusan Mahasiswa FMIPA Unpatti Menggunakan Metode CHAID

FERRY KONDO LEMBANG¹, MEIGA FENDJALANG²

^{1,2}Jurusan Matematika Fakultas MIPA Universitas Pattimura Ambon
Jln. Ir. M. Putuhena Kampus Poka
e-mail: ferrykondolembang@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk mengetahui klasifikasi variabel yang membedakan waktu kelulusan mahasiswa antara tepat waktu dan tidak tepat waktu pada Fakultas MIPA Universitas Pattimura dengan menggunakan Metode CHAID. Penelitian ini dilakukan di Fakultas MIPA Universitas Pattimura, dengan menggunakan sampel penelitian sebanyak 585 orang mahasiswa yang lulus periode Januari 2010 sampai dengan April 2014. Hasil penelitian dengan menggunakan Metode CHAID Menghasilkan pohon klasifikasi yang terdiri dari 4 segmen, dengan variabel yang paling signifikan yaitu jurusan dan IPK. Selanjutnya diperoleh juga ketepatan klasifikasi Metode CHAID yang terbentuk sebesar 85.5 %.

Kata Kunci: Metode CHAID; Ketepatan Klasifikasi; Waktu Kelulusan Mahasiswa

1. PENDAHULUAN

Pendidikan adalah suatu proses pengalaman. Pendidikan yang menjadi persyaratan dasar dalam mencari pekerjaan adalah perguruan tinggi. Pendidikan perguruan tinggi diselenggarakan untuk menghasilkan manusia-manusia yang menguasai ilmu pengetahuan dan teknologi, dan memiliki kepribadian yang unggul. Perguruan tinggi akan mempersiapkan calon-calon sarjana yang handal dan mempunyai ketrampilan di bidangnya. Sebagai lembaga pendidikan, perguruan tinggi tentu menginginkan regenerasi yang baik dimana jumlah mahasiswa yang mencapai gelar sarjana berbanding lurus dengan jumlah mahasiswa baru.

Perguruan tinggi adalah tempat kuliah para mahasiswa yang terdiri dari berbagai macam mahasiswa yang memiliki bakat, minat, kompetensi, dan motivasi yang berbeda-beda. Berhasil tidaknya seorang mahasiswa belajar di perguruan tinggi dapat dilihat dari prestasi belajar yang diperolehnya. Tentunya dalam pencapaian gelar kesarjanaan membutuhkan waktu normal selama empat tahun, tetapi dalam prakteknya mahasiswa tidak selalu dapat menuntaskan studinya selama waktu normal yang telah ditentukan. Faktor-Faktor yang menyebabkan ketidaktepatan waktu kelulusan mahasiswa dapat digolongkan dalam dua faktor : faktor internal dan eksternal. Faktor internal adalah faktor dari dalam diri mahasiswa yang mempengaruhi prestasi belajar, terdiri atas minat, intelegensi, dan motivasi. Sedangkan faktor eksternal adalah faktor dari luar yang mempengaruhi prestasi belajar mahasiswa, terdiri dari peran dosen dalam mendidik mahasiswa, kondisi ekonomi orang tua mahasiswa, dan metode pembelajaran yang digunakan.

Salah satu metode statistika yang dipakai untuk mengklasifikasikan variabel-variabel yang bersifat kategori dengan tujuan membagi rangkaian data menjadi subgrup-subgrup berdasarkan pada variabel respon adalah metode CHAID (Chi-squared Automatic Interaction Detection) (Lehman dan Eherler, 2001). Hasil dari pengklasifikasian dalam CHAID akan ditampilkan dalam sebuah diagram pohon. Makalah ini membahas pengklasifikasian variabel-variabel yang mempengaruhi ketidaktepatan waktu kelulusan mahasiswa F-MIPA Universitas Pattimura Ambon.

Sisa dari makalah ini disusun sebagai berikut. Variabel-variabel dalam analisis CHAID dibahas pada Bagian 2. Langkah-Langkah Analisis CHAID disajikan dalam Bagian 3. Bagian 4 menguraikan Diagram Pohon (Tree Diagram) dari Metode CHAID. Sedangkan bagian terakhir

2 Ferry Kondo Lembang, Meiga Fendjalang

berisikan aplikasi metode CHAID untuk mengklasifikasikan variabel penentu kelulusan mahasiswa FMIPA Unpatti.

2. VARIABEL-VARIABEL DALAM ANALISIS CHAID

Variabel yang digunakan dalam analisis CHAID dibedakan atas variabel terikat dan variabel bebas. Klasifikasi dalam CHAID dilakukan berdasarkan pada hubungan yang ada antara kedua variabel tersebut, oleh karena itu CHAID termasuk dalam metode dependensi dalam menentukan segmentasi. Gallagher (2000) menunjukkan bahwa CHAID akan membedakan variabel – variabel bebasnya menjadi tiga bentuk berbeda, yaitu :

1. Monotonik : Kategori – kategori pada variabel ini dapat dikombinasikan atau digabungkan oleh CHAID jika keduanya berdekatan satu sama lain, yaitu variabel-variabel yang kategorinya mengikuti urutan aslinya (data ordinal). Contohnya usia dan pendapatan.
2. Bebas : Kategori-kategori pada variabel ini dapat dikombinasikan atau digabungkan walaupun keduanya berdekatan atau tidak satu sama lain (data nominal). Contohnya pekerjaan, etnik, dan area geografis.
3. Mengambang (*floating*) : Kategori-kategori pada variabel ini akan diperlakukan seperti monotonik kecuali untuk kategori yang terakhir (yaitu *missing value*), yang dapat berkombinasi dengan kategori manapun..

3. LANGKAH-LANGKAH DALAM ANALISIS CHAID

Menurut Magidson dalam Bagozzi (1994), menerangkan bahwa langkah- langkah analisis CHAID secara garis besar dapat dibagi menjadi tiga tahap, yaitu penggabungan, pemisahan dan penghentian. Tahap – tahap tersebut dapat dijabarkan sebagai berikut:

1. Tahap 1: Penggabungan (*Merging*)

Untuk setiap variabel bebas X_1, X_2, \dots, X_n

- i. Bentuk tabel kontingensi dua arah dengan variabel terikatnya.
- ii. Hitung statistik χ^2 untuk setiap pasang kategori yang dapat dipilih untuk digabung menjadi satu, untuk menguji kebebasannya dalam sebuah sub tabel kontingensi $2 \times j$ yang dibentuk oleh sepasang kategori tersebut dengan variabel terikatnya yang mempunyai sebanyak j kategori.
- iii. Di antara pasangan yang tidak signifikan, gabungkan sebuah pasangan kategori yang paling mirip menjadi sebuah kategori tunggal dan kemudian dilanjutkan ke langkah 4. Apabila semua pasangan kategori yang tersisa signifikan lanjutkan ke langkah 5.
- iv. Untuk suatu kategori gabungan yang terdiri dari tiga kategori atau lebih, uji kategori tersebut untuk melihat apakah suatu kategori variabel bebas seharusnya dipisah dengan menguji kesignifikanan antara kategori tersebut dengan kategori yang lain dalam satu kategori gabungan. Jika didapat nilai χ^2 yang signifikan, pisahkan kategori tersebut dengan yang lain. Jika lebih dari satu kategori yang bisa dipilih untuk dipisah, pisahkan salah satu yang mempunyai nilai χ^2 tertinggi. Kemudian kembali ke langkah 3.
- v. Gabungkan suatu kategori yang mempunyai sedikit pengamatan yang tidak sesuai dengan kategori lain yang paling mirip dengan cara mamilih. Seperti yang diukur oleh nilai χ^2 berpasangan terkecil.
- vi. Hitung *p-value* berdasarkan pada tabel yang telah digabung.

2. Tahap 2: pemisahan (*Splitting*)

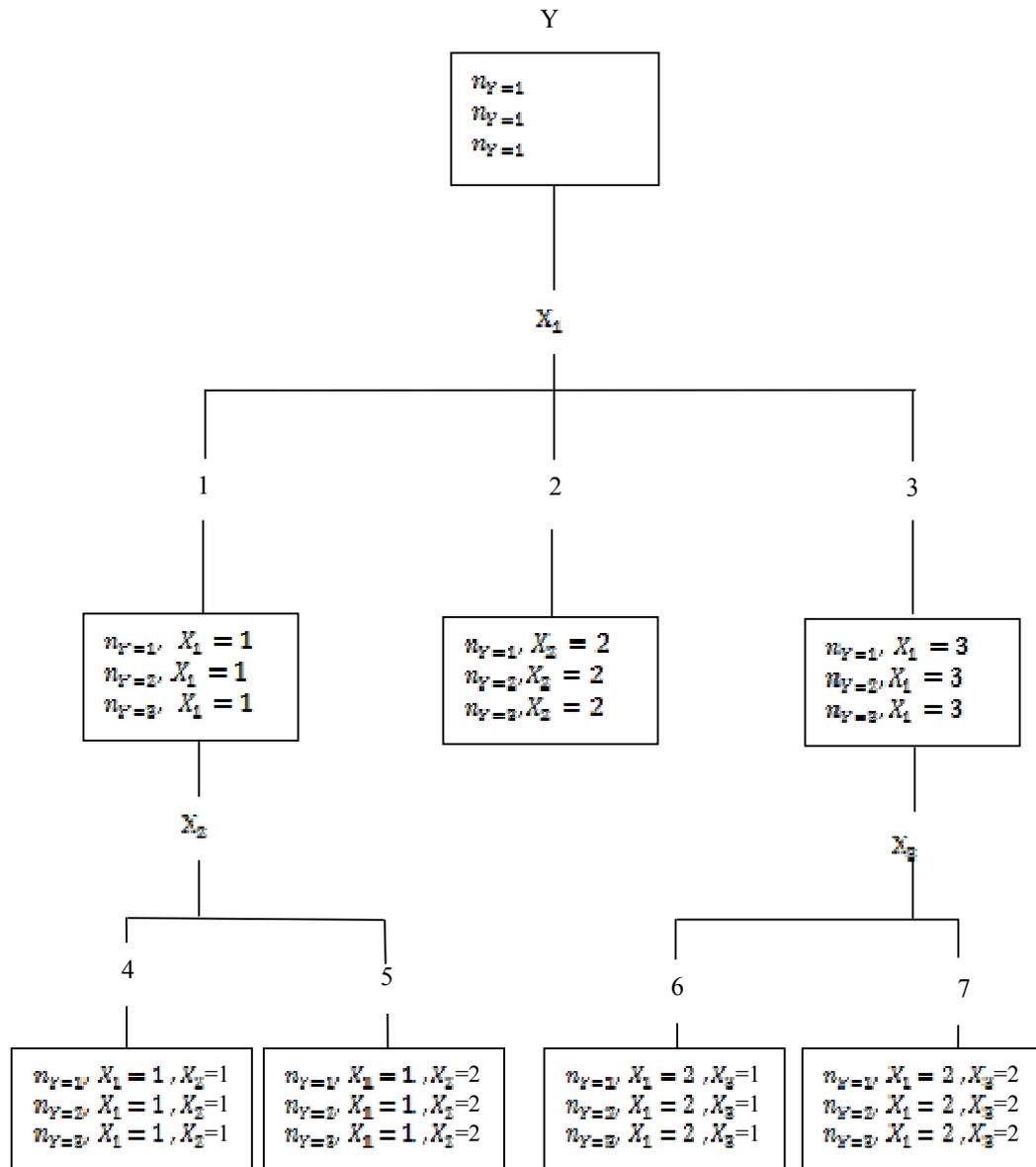
Pilih variabel bebas yang memiliki nilai *p-value* signifikan terbesar. Pembagian kelompok atau pemisahan kelompok dilakukan dengan variabel kelompok tersebut (Gunakan masing – masing kategori variabel bebas tersebut, yang telah digabung secara optimal, untuk menentukan sub pembagian dari kelompok induk menjadi sub kelompok yang baru). Jika tidak ada variabel bebas dengan nilai *p-value* yang signifikan pembagian kelompok atau pemisahan kelompok tidak dapat dilakukan.

3. Tahap 3: Penghentian (*Stopping*)

Kembali ke langkah penggabungan untuk menganalisis sub kelompok berikutnya. Hentikan ketika semua sub kelompok telah dianalisis dan juga telah berisi pengamatan dengan jumlah sedikit.

4. DIAGRAM POHON (TREE DIAGRAM)

Hasil pembentukan segmen dalam CHAID akan ditampilkan dalam sebuah diagram pohon. Secara umum diagram pohon dari CHAID ditunjukkan dalam Gambar 1 (Lehmann dan Eherler, 2001).



Gambar 1 Diagram Pohon Metode CHAID

Diagram pohon CHAID mengikuti aturan “dari atas ke bawah” (*Top-down stopping rule*), di mana diagram pohon disusun mulai dari kelompok induk, berlanjut di bawahnya sub kelompok yang berturut-turut dari hasil pembagian kelompok induk berdasarkan kriteria tertentu (Myers, 1996). Tiap-tiap node dari diagram pohon ini menggambarkan sub kelompok dari sampel yang diteliti. Setiap node akan berisi keseluruhan sampel dan frekuensi absolut ini untuk tiap kategori yang disusun di atasnya. 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 variabel bebas terbaik untuk variabel terikat berdasarkan uji χ^2 . Tiap node berisi

4 Ferry Kondo Lembang, Meiga Fendjalang

informasi tentang frekuensi variabel Y, sebagai variabel terikat, yang merupakan bagian dari sub kelompok yang dihasilkan berdasarkan kategori yang disebutkan (X_1) Pada kedalaman ke-2 (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 penjelas yang lain, yaitu X_2 dan X_3 , dan selanjutnya menjadi sub kelompok pada node ke-4, 5, 6, dan 7 (Lehmann dan Eherler, 2001).

Dari masing-masing node tersebut juga ditampilkan persentase responden untuk tiap-tiap kategori dari variabel terikat, dan juga ditunjukkan jumlah total responden untuk masing-masing node (Myers, 1996).

Secara ringkas, Bagozzi (1994) menyatakan bahwa, diagram pohon, yang merupakan inti dari analisis CHAID, akan berisi:

- Simbol yang menerangkan tentang kategori tertentu (atau kategori-kategori yang telah digabungkan).
- Sebuah ringkasan data dari variabel terikat dalam kelompok tersebut (misalnya persentase respon).
- Ukuran sampel untuk kelompok tersebut, atau biasa dilambangkan dengan "n".

5. APLIKASI METODE CHAID

Sumber data dalam penelitian ini adalah data sekunder yang diperoleh dari kepegawaian F-MIPA Universitas Pattimura. Data yang diambil adalah data kelulusan F-MIPA Universitas Pattimura periode Januari 2010 sampai april 2014. Data tersebut kemudian dipergunakan sebagai variabel dalam metode analisis CHAID.

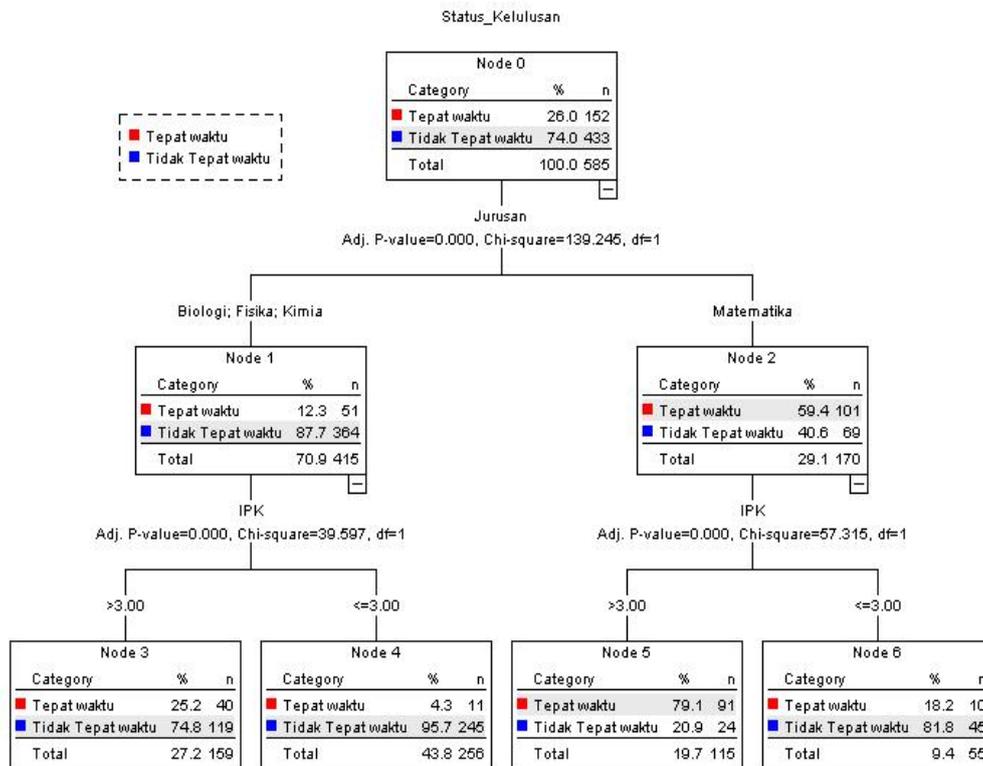
Pada variabel independen, prosedur secara otomatis mengeluarkan variabel independen yang tidak secara signifikan memberikan pengaruh pada model akhir. Dapat dilihat ringkasan pembentukan model pada Tabel 1.

Tabel 1. Pembentukan Model dengan Metode CHAID

Model Summary		
Specifications	Growing Method	CHAID
	Dependent Variable	Status_Kelulusan
	Independent Variables	Jurusan, IPK, Jenis_Kelamin
	Validation	None
	Maximum Tree Depth	3
	Minimum Cases in Parent Node	100
	Minimum Cases in Child Node	50
Results	Independent Variables Included	Jurusan, IPK
	Number of Nodes	7
	Number of Terminal Nodes	4
	Depth	2

Maximum Tree Depth pada bagian *Specification* adalah untuk membatasi pertumbuhan pohon. Pada tabel diatas nilai *Maximum Tree Depth* adalah 3. Sedangkan *minimum cases in parent node* (jumlah minimum kasus pada *parent node*) sebanyak 100 dan *minimum cases in child node* (jumlah minimum kasus pada *child node*) sebanyak 50.

Hasil pembentukan model berupa pohon klasifikasi dengan node sebanyak 7 buah, node akhir (*terminal node*) sebanyak 4 buah dengan level kedalam (*depth*) sebanyak 2. Diketahui variabel yang signifikan dalam membentuk node adalah jurusan dan Indeks Prestasi Kumulatif (IPK). Masing-masing *terminal node* data diartikan sebagai salah satu segmen dari alumni F-MIPA Unpatti segmen-segmen tersebut dapat dilihat pada Gambar 2.



Gambar 2. Diagram Pohon Metode CHAID

Dari Gambar 2, dapat dilihat 4 segmen yang terbentuk berdasarkan klasifikasi, sebagai berikut:

1. Segmen 1 : Alumni jurusan Biologi, Fisika, dan Kimia yang memiliki IPK lebih dari 3.00
2. Segmen 2 : Alumni jurusan Biologi, Fisika dan Kimia yang memiliki IPK kurang dari sama dengan 3.00.
3. Segmen 3 : Alumni jurusan matematika yang memiliki IPK lebih dari 3.00.
4. Segmen 4 : Alumni jurusan matematika yang memiliki IPK kurang dari sama dengan 3.00.

Ketepatan klasifikasi Metode CHAID diartikan sebagai evaluasi dari seberapa bagus Metode CHAID yang terbentuk bekerja. Adapun ketepatan klasifikasi Metode CHAID ditunjukkan pada Tabel 2. Dari tabel ketepatan klasifikasi dari metode CHAID Estimasi risiko yang diperoleh sebesar 0,145 yang menunjukkan bahwa prediksi kategori oleh Metode CHAID (Status kelulusan mahasiswa, tepat waktu atau tidak tepat waktu) klasifikasi yang salah untuk 14.5 % kasus. Sehingga risiko dari klasifikasi yang salah untuk masa studi alumni maksimal sebesar 14.5 % atau Presentase metode CHAID yang terbentuk mengklasifikasikan masa studi alumni secara tepat sebesar 85.5 %.

Tabel 2. Ketepatan Klasifikasi dari metode CHAID

Classification			
Observed	Predicted		
	TEPAT	TIDAK TEPAT	Percent Correct
TEPAT	91	61	59.9%
TIDAK TEPAT	24	409	94.5%
Overall Percentage	19.7%	80.3%	85.5%

Growing Method: CHAID

Dependent Variabel: Status_Kelulusan

DAFTAR PUSTAKA

- Bagozzi, R.P. (1994). *Advances Methods Of Marketing Research*. Blackwell Publisher Ltd. Oxford University.
- Gallagher, C.A. (1999). *An Iterative Approach to Classification Analysis*. www.casact.org/library/ratemaking/90dp237.pdf. Download pada 5 November 2014.
- Lehman, T., dan Eherler, D. (2001). *Responder Profiling with CHAID and Dependency Analysis*.
- Myers, J. H. (1996). *Segmentation and Positioning for Strategic Marketing Decisions*. American Marketing Association: Chicago.