

## ANALISIS CHAID UNTUK IDENTIFIKASI KETEPATAN WAKTU LULUS BERDASARKAN KARAKTERISTIK MAHASISWA

Rindy Anggun Pertiwi\*, Indahwati\*, Farit Mohamad Afendi\*

\*Departemen Statistika Institut Pertanian Bogor

**Ringkasan**—Bogor Agricultural University (IPB) Under-graduated program is one of the best university in Indonesia which also produces the best graduates. From the data of IPB graduated students in 2010-2013 it shows that the percentage of student who graduates on time is 30.94%. It indicates that there are still many students who do not graduate on time. This study aims to look at the characteristics affecting accuracy of graduation on IPB students using CHAID (Chi-squared Automatic Interaction Detection) analysis. The analysis on IPB students CHAID able to classifies students on time with the highest classification probability the characteristics derived from FEM, have a GPA of more than 3.39 and taking a minor during college at IPB, while the lowest classification probability that clasifies IPB students who graduate on time have the characteristics derived from FAPERTA-FAHUTAN, with GPA of less than or equal to 2.87 and from jabodetabek. The analysis on IPB FMIPA CHAID unable to classifies students on time. More over the classification accuracy value for IPB students is 74.6% and for IPB FMIPA students is 80%.

**Keywords**-CHAID analysis; accuracy of graduation; classification accuracy value;

### I. PENDAHULUAN

#### LATAR BELAKANG

Lulus tepat waktu di perguruan tinggi merupakan suatu kewajiban bagi setiap mahasiswa. Selain prestasi akademik yang tinggi, lulus tepat waktu juga menjadi salah satu indikator keberhasilan studi mahasiswa. Banyak faktor yang mempengaruhi ketepatan waktu lulus studi, salah satunya dipengaruhi oleh faktor belajar. Menurut Suryabrata ([1]) faktor belajar terbagi menjadi dua, yaitu faktor yang berasal dari luar diri (eksternal) dan faktor yang berasal dari dalam diri (internal). Mahasiswa program sarjana dikatakan lulus tepat waktu apabila menyelesaikan pendidikan di perguruan tinggi selama empat tahun atau kurang. Lulus tidak tepat waktu akan merugikan negara, perguruan tinggi, keluarga dan terutama diri sendiri.

Institut Pertanian Bogor (IPB) program sarjana merupakan salah satu perguruan tinggi negeri terbaik di Indonesia yang menghasilkan lulusan-lulusan terbaik. Data lulusan mahasiswa IPB tahun 2010-2013 yang lulus tepat waktu sebesar 30.94%. Terlihat masih banyak mahasiswa IPB yang lulus tidak tepat waktu. Salah satu metode statistika yang dapat digunakan untuk menganalisis karakteristik mahasiswa lulus tepat waktu dan tidak tepat waktu adalah metode CHAID (*Chi-squared Automatic Interaction Detection*). Metode

CHAID merupakan salah satu teknik nonparametrik yang dapat digunakan untuk pemilihan peubah dari data berukuran besar dalam menentukan peubah-peubah yang paling berpengaruh. Metode CHAID akan menghasilkan diagram yang mirip dengan diagram pohon keputusan dan menggunakan uji Khi-kuadrat pada pengoperasiannya ([2]). Tujuan dari penelitian ini adalah untuk melihat karakteristik mahasiswa IPB dan FMIPA IPB yang mempengaruhi ketepatan waktu lulus dengan menggunakan metode CHAID

### METODE

Data yang digunakan adalah data sekunder dari Direktorat Administrasi dan Pendidikan (Dit-Ap) IPB. Data tersebut terdiri dari 7024 alumni IPB dan 1317 alumni FMIPA IPB tahun 2010-2013. Data wisuda periode 2010/2011, 2011/2012 terdiri dari 5 tahap wisuda sedangkan periode 2012/2013 terdiri dari 4 tahap wisuda. Data tersebut didapatkan dari mahasiswa yang akan mendaftarkan diri untuk mengikuti wisuda. Penelitian ini menggunakan data alumni secara reguler (mahasiswa yang masuk ke perguruan tinggi setelah lulus dari SMA). Data tersebut terdiri dari satu peubah respon ( $Y$ ) yaitu ketepatan waktu lulus dan delapan peubah penjelas ( $X$ ) yaitu fakultas/departemen ( $X_1$ ), minor ( $X_2$ ), IPK ( $X_3$ ), jenis kelamin ( $X_4$ ), asal daerah ( $X_5$ ), asal sekolah ( $X_6$ ), jalur masuk ( $X_7$ ), dan beasiswa ( $X_8$ ). Peubah jalur masuk yang dipilih hanya tiga yaitu USMI, SNMPTN dan BUD karena jalur masuk yang lain memiliki jumlah yang sedikit sehingga penelitian ini hanya fokus pada tiga jalur masuk. Peubah IPK merupakan IPK ketika mendaftar wisuda, bukan IPK setelah semua nilai keluar, sehingga bisa dikatakan IPK sampai semester 7.

Langkah-langkah analisis yang akan dilakukan pada penelitian ini adalah sebagai berikut :

- 1) Eksplorasi data, jika dari data yang diperoleh ada yang tidak lengkap maka dilakukan pendugaan data hilang untuk melihat dari peubah lain yang berhubungan dengan peubah yang tidak lengkap.
- 2) Melakukan analisis deskriptif terhadap peubah penjelas dan peubah respon.
- 3) Melakukan analisis dengan metode CHAID yang merupakan suatu metode statistika untuk memprediksi keanggotaan objek dalam kelas-kelas peubah tak bebas kategorik. Pohon klasifikasi dibentuk melalui penyekatan data secara rekursif. Menurut Kass ([3]),

algoritma metode CHAID sebagai berikut :

- a) Untuk setiap peubah penjelas, buat tabulasi silang kategori-kategori peubah penjelas dengan kategori peubah respon.
- b) Cari pasangan kategori dari peubah penjelas dengan sub-tabel  $2 \times d$  ( $d$  adalah banyaknya kategori peubah respon) yang mempunyai nilai  $\chi^2$  paling kecil. Kemudian bandingkan  $\chi^2$  yang dihasilkan dengan  $\chi_\alpha^2$  yang telah ditentukan sebelumnya. Jika  $\chi^2 < \chi_\alpha^2$  gabung pasangan ini kedalam satu kategori baru. Nilai  $\chi^2$  didapat dengan rumus :

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \left[ \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right]$$

$$E_{ij} = \frac{n_i \cdot n_j}{n}$$

keterangan :

- $r$  : banyaknya baris
- $c$  : banyaknya kolom
- $i$  : kategori peubah penjelas ke- $i$
- $j$  : kategori peubah respon ke- $j$

- c) Untuk setiap kategori gabungan yang berisi tiga atau lebih kategori asal, cari pemisahan biner yang mempunyai nilai  $\chi^2$  paling besar. Jika  $\chi^2 > \chi_\alpha^2$ , maka buatlah pemisahan baru dan kembali ke tahap 2.
  - d) Hitung nilai  $\chi^{2*}$  dari peubah pejelasan yang telah digabung dengan peubah respon. Cari nilai  $\chi^{2*}$  terbesar dan kemudian bandingkan dengan  $\chi_\alpha^{2*}$ . Jika  $\chi^{2*} \geq \chi_\alpha^{2*}$ , maka bagi data menurut kategori tersebut.
  - e) Jika terjadi pemisahan pada tahap 4, maka kembali ke tahap 1 untuk setiap bagian data hasil pemisahan.
- 4) Apabila terjadi penggabungan pada peubah penjelas, maka dilakukan uji Bonfferoni sebagai pengali terhadap  $\chi_\alpha^2$ . Penggandaan Bonfferoni ada tiga tipe yang dibedakan berdasarkan skala pengukuran peubahnya, yaitu :

- Jika peubah asal berskala nominal digunakan rumus pengganda Bonfferoni tipe bebas, yaitu :

$$B = \sum_{i=0}^{b-1} (-1)^i \frac{(b-i)^r}{i!(b-i)!}$$

keterangan :

- $B$  : pengali Bonfferoni
- $b$  : banyaknya kategori baru
- $r$  : banyaknya kategori asal
- $i$  : kategori baru ke- $i$

Jika peubah asal berskala ordinal digunakan rumus pengganda Bonfferoni tipe monotonik, yaitu

:

$$B = \binom{r-1}{b-1}$$

- Jika peubah asal berskala ordinal, namun terdapat kategori yang belum dapat ditentukan urutannya dalam peubah tersebut digunakan rumus pengganda Bonfferoni tipe mengambang (*float*), yaitu :

$$B = \binom{r-2}{b-2} + r \binom{r-2}{b-1}$$

- 5) Melakukan pemodelan diagram pohon untuk mempermudah interpretasi dan pengambilan kesimpulan.

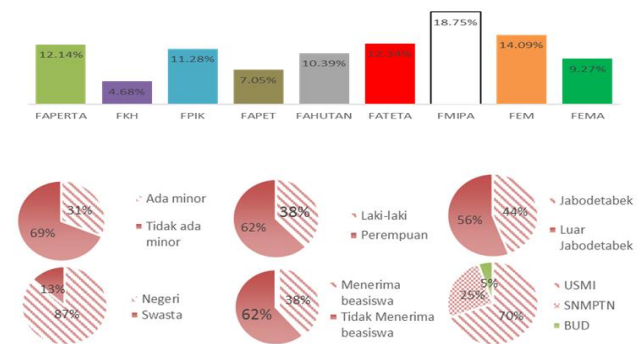
## II. HASIL DAN PEMBAHASAN

### A. Analisis Deskriptif

#### 1) Mahasiswa IPB

Gambar 1 merupakan data wisudawan tahun 2010-2013 dan ternyata wisudawan terbanyak di IPB berasal dari FMIPA dengan persentase sebesar 18.75% sedangkan yang paling sedikit berasal dari FKH dengan persentase sebesar 4.68%. Hal tersebut dikarenakan FMIPA memiliki departemen paling banyak di IPB yaitu delapan departemen sedangkan FKH hanya satu departemen.

Mahasiswa IPB yang mengambil minor selama berkuliah hanya sedikit yaitu 31%. IPB juga didominasi dengan mahasiswa yang berjenis kelamin perempuan, mahasiswa yang berasal dari SMA negeri dan mahasiswa yang masuk dengan jalur USMI.

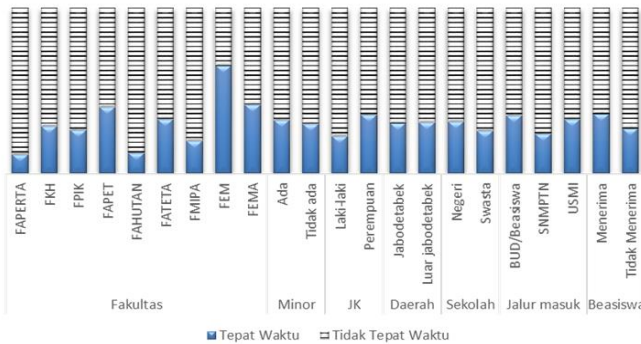


Gambar 1. Sebaran persentase mahasiswa

#### 2) Ketepatan Waktu Lulus Berdasarkan Karakteristik Mahasiswa

Gambar 2 menunjukkan bahwa mahasiswa dari FEM yang paling banyak lulus tepat waktu dengan persentase sebesar 65.45% sedangkan mahasiswa dari

FAPERTA dan FAHUTAN memiliki persentase yang sedikit yaitu 11.84% dan 12.74%. Peubah yang memiliki keragaman yang besar hanya peubah fakultas sedangkan peubah yang lain memiliki keragaman yang kecil.



Gambar 2. Persentase ketepatan waktu lulus mahasiswa IPB berdasarkan karakteristik mahasiswa

**B. Analisis CHAID**

Analisis CHAID mahasiswa IPB menghasilkan 41 klasifikasi. Semua peubah penjelas masuk dalam diagram pohon CHAID dan peubah yang memiliki asosiasi paling kuat (simpul akar) yaitu fakultas. Pohon klasifikasi yang dihasilkan memiliki 68 simpul yang terdiri dari 24 simpul dalam dan 41 simpul akhir. Metode penghentian pohon yang digunakan yaitu minimum amatan pada simpul induk sebanyak 100 amatan dan pada simpul anak sebanyak 50 amatan. Kedalaman pohon tidak dibatasi namun terhenti pada kedalaman empat.

Klasifikasi yang menggolongkan mahasiswa lulus dengan tepat waktu yaitu simpul 57, 54, 56, 44, 67, 42 dan 55 dengan persentase tepat waktu diatas 60%. Klasifikasi yang memiliki peluang paling besar diantaranya yaitu simpul 57 dengan persentase mahasiswa sebanyak 2.9% (n = 206) dari total mahasiswa IPB secara keseluruhan, memiliki ciri-ciri berasal dari FEM, memiliki IPK lebih dari 3.39, dan mengambil minor selama perkuliahan di IPB. Apabila terdapat mahasiswa yang memiliki karakteristik seperti simpul 57, maka peluang mahasiswa tersebut lulus tepat waktu sebesar 82%. Simpul 54 memiliki karakteristik berasal dari FEM memiliki IPK antara 3.08-3.39 dan tidak mengambil minor selama perkuliahan di IPB. Apabila terdapat mahasiswa yang memiliki karakteristik seperti simpul 54 maka peluang mahasiswa tersebut lulus tepat waktu sebesar 73.1%.

Klasifikasi yang menggolongkan mahasiswa lulus dengan tidak tepat waktu yaitu simpul 48, 30, 27, 49 dan 28 dengan persentase tepat waktu dibawah 10%. Klasifikasi

yang memiliki peluang paling kecil diantaranya yaitu simpul 28 dengan persentase mahasiswa sebanyak 3.7% (n = 261) dari total mahasiswa IPB secara keseluruhan. Simpul 28 memiliki ciri-ciri sebagai berikut, berasal dari FAPERTA-FAHUTAN, memiliki IPK kurang dari atau sama dengan 2.87, dan berasal dari Jabodetabek. Apabila terdapat mahasiswa yang memiliki karakteristik seperti simpul 28, maka peluang mahasiswa tersebut lulus tepat waktu sebesar 3.4%. Simpul 49 memiliki ciri-ciri berasal dari FMIPA, memiliki IPK kurang dari atau sama dengan 2.76 dan berasal dari Jabodetabek. Apabila terdapat mahasiswa yang memiliki karakteristik seperti simpul 49 maka peluang mahasiswa tersebut lulus tepat waktu sebesar 3.6%.

Berdasarkan hasil ini, mahasiswa yang banyak tergolong lulus tepat waktu yaitu mahasiswa FEM dan mempunyai IPK lebih dari 3 sedangkan mahasiswa yang banyak tergolong lulus tidak tepat waktu yaitu mahasiswa FAPERTA-FAHUTAN dan FMIPA serta mempunyai IPK kurang dari 3. Hal ini dikarenakan mahasiswa FEM tidak melakukan penelitian di laboratorium untuk data skripsi sehingga mayoritas mahasiswanya menggunakan data sekunder, berbeda dengan FAPERTA-FAHUTAN dan beberapa departemen pada FMIPA yang kebanyakan mahasiswanya melakukan penelitian di laboratorium atau di lapangan. Selain itu, banyak mahasiswa FEM yang sudah mengerjakan skripsi dari awal semester delapan karena tidak banyak lagi mata kuliah yang diambil sehingga bisa fokus mengerjakan skripsi.

Menurut ([4]) ketepatan klasifikasi (*Correct Classification Rate*) yang dihasilkan dengan metode CHAID yaitu sebesar 74.6% yang dapat dilihat pada Tabel I.

Tabel I  
KESESUAIAN KLASIFIKASI CHAID MAHASISWA IPB

Amatan	Dugaan		Total	Ketepatan %
	Tidak tepat waktu	Tepat waktu		
Tidak tepat waktu	4498	353	4851	92.7
Tepat waktu	1430	743	2173	34.2
Total	5928	1096	7024	74.6
Kesalahan (%)	24.12	32.2	25.4	-

Tabel I menunjukkan persentase untuk menduga lulus tidak tepat waktu secara benar (*sensitivity*) sebesar 34.2% dan untuk menduga lulus tepat secara benar (*specificity*) sebesar 92.7%. Model pada penelitian ini lebih bagus untuk menduga mahasiswa yang lulus tidak tepat waktu dibanding manduga mahasiswa yang lulus tepat waktu. Nilai kesalahan

Tabel II  
RINGKASAN ANALISIS CHAID MAHASISWA FMIPA IPB

Klasifikasi	Simpul	Karakteristik ketepatan waktu lulus	Persentase (%)		Banyak mahasiswa (n)
			Tepat waktu	Tidak tepat waktu	
1	9	IPK>3.09; luar jabodetabek; tidak ada minor	46.3	53.7	136
2	8	IPK>3.09; luar jabodetabek;ada minor	28	72	164
3	7	IPK>3.09; jabodetabek	23.2	76.8	224
4	2	2.77<IPK<=3.09	19.2	80.8	401
5	4	IPK<=2.77; luar jabodetabek	9.4	90.6	192
6	5	IPK<=2.77; jabodetabek	3.5	96.5	200

positif yaitu kesalahan ketika observasi mengatakan lulus tepat waktu, namun diprediksi lulus tidak tepat waktu dengan nilai sebesar 32.2%. Nilai kesalahan negatif yaitu kesalahan ketika observasi mengatakan lulus tidak tepat waktu, namun diprediksi lulus tepat waktu dengan nilai sebesar 24.12%.

### C. Analisis CHAID Mahasiswa FMIPA IPB

Analisis ketepatan waktu lulus mahasiswa FMIPA IPB berdasarkan departemen yaitu Statistika 23.08%, GFM 23.85%, Biologi 12.27%, Kimia 15.47%, Matematika 23.13%, Ilmu Komputer 21.70%, Fisika 29.03% dan Biokimia 18.71%. Ringkasan analisis CHAID mahasiswa FMIPA IPB dapat dilihat pada Tabel II. Peubah penjelas yang masuk dalam diagram pohon CHAID untuk mahasiswa FMIPA IPB terdiri dari tiga peubah yaitu peubah IPK, asal daerah dan minor. Analisis CHAID mahasiswa FMIPA IPB menghasilkan 10 simpul dengan 3 simpul dalam dan 6 simpul akhir. Metode penghentian pohon yang digunakan sama dengan analisis mahasiswa IPB dan kedalaman pohon terhenti pada kedalaman tiga.

Peubah yang memiliki pengaruh paling kuat (simpul akar) terhadap ketepatan waktu lulus yaitu IPK dan terlihat pula bahwa IPK berbanding lurus dengan persentase mahasiswa yang lulus tepat waktu. Semakin tinggi IPK maka semakin tinggi persentase mahasiswa yang lulus tepat waktu. Selanjutnya, IPK yang rendah ( $\leq 2.77$ ) dan IPK yang tinggi ( $> 3.09$ ) mempunyai interaksi dengan asal daerah. Persentase mahasiswa yang mempunyai IPK rendah dan berasal dari luar jabodetabek lebih tinggi dibandingkan mahasiswa yang mempunyai IPK rendah dan berasal dari jabodetabek. Persentase mahasiswa yang mempunyai IPK tinggi juga sama, akan tetapi pada IPK tinggi mengalami peningkatan persentase mahasiswa yang lulus tepat waktu lebih curam pada asal daerah luar jabodetabek dibandingkan jabodetabek. Mahasiswa yang mempunyai IPK tinggi dan berasal dari luar jabodetabek juga mempunyai interaksi dengan minor.

Analisis CHAID mahasiswa FMIPA IPB menghasilkan 6 klasifikasi namun tidak ada klasifikasi yang menggolongkan mahasiswa FMIPA IPB lulus dengan tepat waktu karena semua klasifikasi memiliki persentase tepat waktu dibawah

50%. Klasifikasi yang menggolongkan lulus dengan tidak tepat waktu yaitu simpul 2, 4 dan 5. Klasifikasi yang memiliki peluang paling kecil diantaranya yaitu simpul 5 dengan persentase mahasiswa FMIPA IPB sebesar 15.2% ( $n=200$ ) dari total mahasiswa FMIPA IPB secara keseluruhan. Simpul 5 memiliki ciri-ciri sebagai berikut, memiliki IPK kurang dari atau sama dengan 2.77 dan berasal dari jabodetabek. Apabila terdapat mahasiswa yang memiliki karakteristik seperti simpul 5, maka peluang mahasiswa tersebut lulus tepat waktu sebesar 3.5%. Simpul 2 dan simpul 4 bisa dilihat pada Tabel II.

Tabel III  
KESESUAIAN KLASIFIKASI CHAID MAHASISWA IPB

Amatan	Dugaan		Total	Ketepatan %
	Tidak tepat waktu	Tepat waktu		
Tidak tepat waktu	1054	0	1054	100
Tepat waktu	263	0	263	0
Total	1317	0	1317	80
Kesalahan (%)	19.96	0	20	-

Menurut ([4]) ketepatan klasifikasi (*Correct Classification Rate*) yang dihasilkan dengan metode CHAID yaitu sebesar 80% yang dapat dilihat pada Tabel III. Hal ini menunjukkan persentase untuk menduga lulus tidak tepat waktu secara benar (*sensitivity*) sebesar 0% dan untuk menduga lulus tepat secara benar (*specificity*) sebesar 100%. Model pada penelitian ini lebih bagus untuk menduga mahasiswa yang lulus tidak tepat waktu dibanding menduga mahasiswa yang lulus tepat waktu. Nilai kesalahan positif yaitu kesalahan ketika observasi mengatakan lulus tepat waktu, namun diprediksi lulus tidak tepat waktu dengan nilai sebesar 0%. Nilai kesalahan negatif yaitu kesalahan ketika observasi

mengatakan lulus tidak tepat waktu, namun diprediksi lulus tepat waktu dengan nilai sebesar 19.96%.

### III. SIMPULAN

Semua peubah pada mahasiswa IPB masuk dalam diagram pohon dan peubah yang paling berpengaruh dengan ketepatan waktu lulus yaitu fakultas. Peluang klasifikasi tertinggi yang menggolongkan mahasiswa IPB lulus tepat waktu yaitu memiliki ciri-ciri berasal dari FEM, memiliki IPK lebih dari 3.39, dan mengambil minor selama berkuliah di IPB, sedangkan peluang klasifikasi terendah yang menggolongkan mahasiswa IPB lulus tepat waktu yaitu mahasiswa berasal dari FAPERTA-FAHUTAN, memiliki IPK kurang dari atau sama dengan 2.87, dan berasal dari Jabodetabek. Nilai ketepatan klasifikasi mahasiswa IPB sebesar 74.6%.

Peubah pada mahasiswa FMIPA IPB yang masuk dalam diagram pohon CHAID ada tiga yaitu IPK, asal daerah, dan minor. Peubah yang paling berpengaruh dengan ketepatan waktu lulus yaitu IPK. Tidak ada klasifikasi yang menggolongkan mahasiswa FMIPA IPB lulus tepat waktu sedangkan peluang klasifikasi terendah yang menggolongkan mahasiswa FMIPA IPB lulus tepat waktu yaitu memiliki ciri-ciri IPK kurang dari atau sama dengan 2.77 dan berasal dari Jabodetabek. Nilai ketepatan klasifikasi mahasiswa FMIPA IPB sebesar 80%. Model pada penelitian mahasiswa IPB dan FMIPA IPB lebih baik untuk menduga mahasiswa yang lulus tidak tepat waktu daripada mahasiswa yang lulus tepat waktu.

### DAFTAR PUSTAKA

- [1] S. Suryabrata, *Psikologi Pendidikan*, Jakarta: PT Raja Grafindo Persada, 1986.
- [2] Alamudi, A. Wigena, A. H. Aunuddin, *Eksplorasi Struktur Data dengan Metode CHAID*. *Forum Statistika dan Komputasi*. <http://resipatory.ipb.ac.id.pdf> [online: access 2012 Desember 15];.
- [3] Kass, G. V. *An Exploratory Technique for Investigating Large Quantities of Categorical Data*. *Applied Statistics*. <http://links.jstor.org/sici?sici=0035-9254%281980%2929%3A2%3C119%3AAETFIL%3E2.0.CO%3B2-N.pdf> [online: access 2013 Mei 27]; 29(2) : 119-127.
- [4] DW. Hosmer, A. Lemeshow, *Applied Logistic Regression*, Ed ke-2, USA : John Wiley & Sons, Inc, 2000.