

KLASIFIKASI WAKTU KELULUSAN MAHASISWA STIKOM BALI MENGUNAKAN *CHAID REGRESSION – TREES* DAN REGRESI LOGISTIK BINER

¹I Ketut Putu Suniantara, ²Muhammad Rusli

^{1,2}Program Studi Sistem Informasi, STMIK- STIKOM Bali

Email : ¹suniantaraputu@gmail.com

ABSTRAK

Ketidaktepatan waktu kelulusan mahasiswa dipengaruhi oleh banyak faktor. Masalah ketidaktepatan waktu kelulusan mahasiswa perlu mendapatkan perhatian khusus, untuk mengantisipasi masalah tersebut perlu dilakukan analisis klasifikasi faktor-faktor apa yang mempengaruhi ketidaktepatan waktu kelulusan mahasiswa. Salah satu metode klasifikasi adalah analisis regresi logistik dan *CHAID Regression Trees*. Penelitian ini bertujuan untuk mengetahui variabel-variabel yang memengaruhi waktu kelulusan mahasiswa STIKOM BALI dengan menggunakan *CHAID regression trees* dan regresi logistik biner. Adapun variabel-variabel yang digunakan yaitu variabel responnya adalah status kelulusan mahasiswa tepat waktu dan tidak tepat waktu dan variabel penjelasnya adalah Jurusan/prodi, IPK, IPS semester 6, Lama menyusun skripsi, Jenis Kelamin, dan Nilai ujian masuk. Hasil klasifikasi dengan menggunakan metode *CHAID regression trees* menghasilkan 5 klasifikasi yang berdasarkan nilai indek di atas 100. Hasil klasifikasi ini dipengaruhi oleh lama menyusun skripsi, IPK, IP semester 6 dan program studi. Sedangkan hasil klasifikasi dengan menggunakan metode regresi logistik biner dipengaruhi oleh Program studi, IPK dan lama menyusun skripsi. Ketepatan klasifikasi dengan metode *CHAID Regression Tress* lebih baik dibandingkan dengan menggunakan metode regresi logistik biner.

Kata Kunci : *CHAID Regression Trees*, regresi logistik biner, klasifikasi

PENDAHULUAN

Lulus pada tingkat pendidikan sarjana dengan tepat waktu dan langsung mendapatkan kerja adalah harapan setiap mahasiswa. Lulus tepat waktu dipengaruhi oleh banyak faktor. Faktor-faktor tersebut dapat bersumber dari diri mahasiswa (faktor internal) atau dari luar diri mahasiswa (lingkungan/faktor eksternal). Faktor yang bersumber dari diri mahasiswa yaitu intelektual. Namun hal lain yang ada pada diri individu mahasiswa yang juga berpengaruh terhadap kondisi belajar adalah situasi afektif, seperti ketenangan, ketentrangan psikis juga motivasi untuk belajar [1]. Faktor-faktor dari diri mahasiswa yang dimaksudkan di sini

berupa latar belakang mahasiswa, diantaranya aktivitas belajar, umur, jenis kelamin, motivasi, status mahasiswa, fasilitas belajar, tingkat sosial ekonomi, minat jurusan, dan latar belakang yang lain.

Menurut [2] dalam penelitiannya yang berjudul “identifikasi faktor-faktor yang ketepatan waktu kelulusan mahasiswa program pasca sarjana IPB” dengan menggunakan analisis regresi logistik biner bahwa faktor yang mempengaruhi ketepatan kelulusan adalah fakultas, jenis kelamin, jalur masuk, IPK, dan beasiswa. [3] dalam penelitiannya yang berjudul “analisis classification and regression trees (CART) pada lama studi mahasiwa

STIKOM Bali menyatakan bahwa faktor yang memengaruhi lama studi mahasiswa adalah lama penyusunan skripsi, IPK dan jurusan mahasiswa. Sedangkan menurut [4] dalam penelitiannya yang berjudul “Analisis Waktu Kelulusan Mahasiswa Dengan Metode CHAID (Studi Kasus: Fmipa Universitas Udayana)” menyatakan bahwa waktu kelulusan mahasiswa disebabkan oleh jenis kelamin, jurusan, Ipk dan lama skripsi.

Langkah pertama yang harus dilakukan oleh pihak kampus adalah mencari informasi variabel – variabel apa yang mempengaruhi tepat atau tidaknya mahasiswa lulus dengan menggunakan analisis klasifikasi. Analisis yang dapat digunakan untuk klasifikasi metode *CHAID regression trees* dan metode regresi logistik biner. Pada penelitian [5] menyatakan bahwa metode CART memiliki ketepatan prediksi lebih baik dibandingkan dengan metode regresi logistik biner dalam prediksi variabel respon. Penelitian klasifikasi kelulusan dengan menggunakan data mining memang telah banyak dilakukan, akan tetapi klasifikasi dengan perbandingan metode *CHAID regression trees* dan Regresi Logistik Biner pada ketidaktepatan kelulusan mahasiswa masih sedikit. Untuk penelitian ini akan dilakukan analisis klasifikasi ketidaktepatan waktu kelulusan mahasiswa STIKOM BALI dengan menggunakan *CHAID regression trees* dan regresi logistik biner. Penelitian ini bertujuan untuk mengetahui klasifikasi waktu kelulusan mahasiswa STIKOM BALI dengan menggunakan *CHAID regression trees* dan regresi logistik biner dan membandingkan tingkat akurasi analisis *CHAID regression trees* dan analisis regresi logistik biner dalam pengklasifikasian waktu kelulusan mahasiswa STIKOM Bali.

METODE PENELITIAN

Sumber Data dan Variabel Penelitian

Penelitian ini dilakukan di STIKOM – Bali dengan waktu penelitian lima bulan, dimana data yang digunakan dalam penelitian ini data lulusan mahasiswa STIKOM Bali periode 2016 sebanyak 683 mahasiswa. Variabel yang digunakan dalam penelitian ini sebagai berikut:

- Y : Status Kelulusan mahasiswa
- X₁ : jenis kelamin
- X₂ : Program studi
- X₃ : lama skripsi
- X₄ : IPK
- X₅ : IP semester 6
- X₆ : Nilai ujian masuk

Metode Analisis

Adapun langkah–langkah untuk mencapai tujuan penelitian tersebut sebagai berikut:

1. Analisis *CHAID regression trees*, dengan tahapan sebagai berikut:
 - a. Menentukan semua kemungkinan penyekatan pada tiap variabel penjelas.
 - b. Menghitung kehomogenan simpul pada semua variabel penjelas.
 - c. Jika simpul induk telah didapat, maka simpul anak kiri dan kanan dibuat dengan cara yang sama untuk semua variabel penjelas berdasarkan data yang sudah dikelompokkan oleh simpul induk.
 - d. Pembentukan pohon dilakukan sampai dipenuhi suatu aturan penghentian tertentu.
 - e. Pemangkasan pohon dilakukan untuk mendapatkan pohon akhir yang lebih sederhana.
 - f. Setelah diperoleh pengelompokan, kemudian melakukan interpretasi.
2. Analisis regresi logistik biner, dengan tahapan sebagai berikut:
 - a. Menduga parameter
 - b. Melakukan pengujian parameter secara simultan dengan Uji – G.

- c. Melakukan pengujian parameter secara parsial dengan menggunakan Uji Wald.
 - d. Melakukan analisis regresi logistik bertatar.
 - e. Interpretasi koefisien model regresi.
3. Membandingkan hasil analisis CHAID dan regresi logistik biner dengan melihat ketepatan prediksi hasil klasifikasi.

Tahapan-tahapan di atas dilakukan secara otomatis oleh komputer dengan menggunakan software Statistika.

HASIL PENELITIAN

1. Klasifikasi dengan Metode CHAID regression trees

Hasil klasifikasi ketepatan waktu kelulusan mahasiswa STIKOM Bali tahun 2016 menghasilkan 20 simpul dengan 13 simpul terminal dengan 3 kedalaman yaitu jumlah hierarki pada pohon klasifikasi berdasarkan LMS, IPK, IPS dan program studi (Gambar 1).

Ketepatan klasifikasi waktu lulusan dan tidak tepat waktu, yang nantinya diharapkan adalah prediksi kelulusan tepat waktu, maka target prediksinya adalah ketepatan waktu lulusan. Berdasarkan target kategori yaitu ketepatan lulusan (Tepat Waktu) yang dilihat dari indek di atas 100 dari 13 simpul terminal diperoleh 5 simpul terminal sebagai klasifikasi yaitu simpul 18, 16, 1, 17, dan simpul 10 (Tabel 1).

Adapun klasifikasi kelulusan sebagai berikut:

Simpul 18 adalah simpul terminal pertama dengan klasifikasi prediksi ketepatan waktu kelulusan 100% pada lama menyusun skripsi sekitar 1 sampai 2 bulan dengan IPK lebih dari 3,39 dan program studi Sistem Informasi.

Simpul selanjutnya adalah simpul 16 yang di tandai dengan simpul terminal kedua yaitu prediksi waktu ketepatan kelulusan sebesar 97,4% yang dipengaruhi

oleh lama menyusun skripsi sekitar 1 sampai 2 bulan dengan IPK berkisar antara 3,18 sampai dengan 3,39 pada prodi Sistem Informasi.

Simpul 1 sebagai terminal ketiga dengan prediksi waktu ketepatan lulusan 96,2% bila lama menyusun skripsi kurang dari atau sama dengan 1 bulan.

Simpul 17 atau simpul terminal keempat dengan prediksi waktu ketepatan kelulusan sebesar 77,3% yang dipengaruhi oleh lama menyusun skripsi sekitar 1 sampai 2 bulan dengan IPK lebih dari 3,39 pada prodi Sistem Komputer.

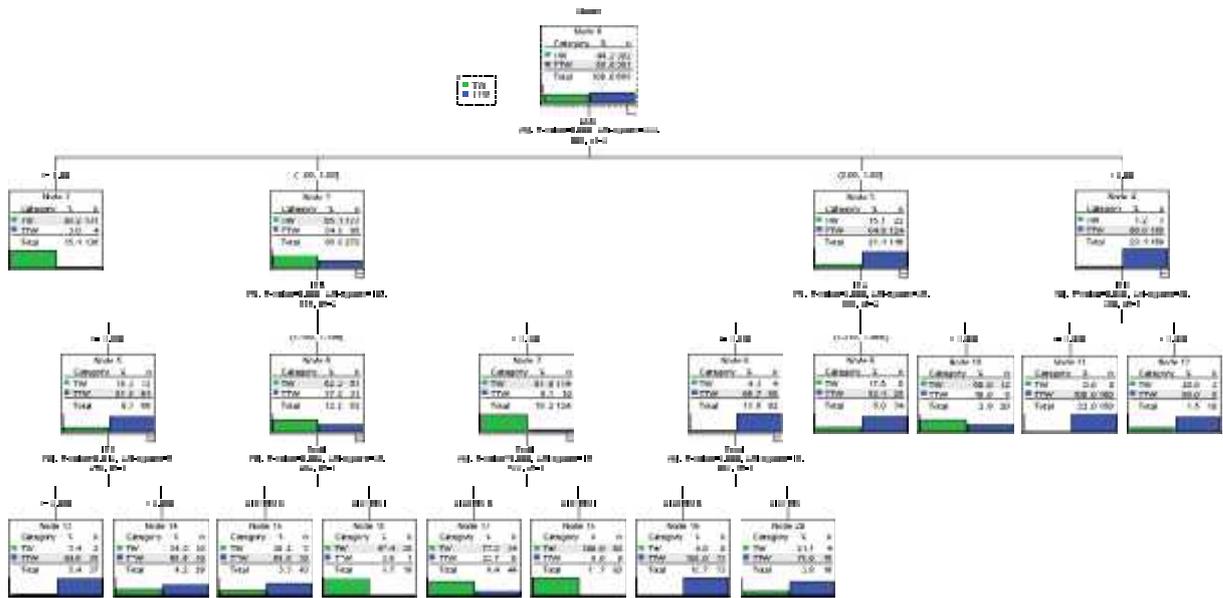
Simpul 10 yang merupakan simpul terminal kelima dengan prediksi waktu ketepatan kelulusan sebesar 60% yang dipengaruhi oleh lama menyusun skripsi sekitar 2 sampai 3 bulan dengan IP semester 6 lebih besar dari 3,48.

Berdasarkan kelima simpul terminal di atas, terlihat simpul 1 memiliki *gain percent* lebih besar yaitu 33,4% yang artinya jika lama menyusun skripsi kurang dari atau sama dengan 1 bulan maka target prediksi ketepatan waktu lulusan pada mahasiswa sekitar 33,4% nilai ini mendekati 42,2%. (Tabel 1).

Tabel 1. Simpul Target Ketepatan Lulusan

Simpul	Gain		Response	Index
	N	Percent		
18	80	26.5%	100.0%	226.2%
16	38	12.6%	97.4%	220.4%
1	101	33.4%	96.2%	217.5%
17	34	11.3%	77.3%	174.8%
10	12	4.0%	60.0%	135.7%
14	10	3.3%	34.5%	78.0%
15	13	4.3%	30.2%	68.4%
20	4	1.3%	21.1%	47.6%
12	2	0.7%	20.0%	45.2%
9	6	2.0%	17.6%	39.9%
13	2	0.7%	5.4%	12.2%
11	0	0.0%	0.0%	0.0%
19	0	0.0%	0.0%	0.0%

Berdasarkan Tabel 2 di atas terlihat ketepatan klasifikasi berada pada tingkat akurasi 91,2% dengan standar error yang sangat kecil yaitu 0,011 dengan resiko kesalahan 8,8%.



Gambar 1. Pohon Klasifikasi Optimal

Tabel 2. Ketepatan Klasifikasi

Aktual	Prediksi		
	Tepat Waktu	Tidak Tepat Waktu	Ketepatan Klasifikasi
Tepat Waktu	265	37	87.7%
Tidak Tepat Waktu	23	358	94.0%
Ket. Klasifikasi Keseluruhan			91.2%

Tabel 3. Uji Model

Chi-square	Df	Signifikan
597,195	6	0,000
597,195	6	0,000
597,195	6	0,000

Adapun model logit yang diperoleh sebagai berikut:

$$g(x) = 13,446 + 2,224Pr\ odi(1) + 4,551IPK + 1,070IPs + 0,041JK(1) + 0,010Nilai\ Ujian\ Masuk + 2,898Lama\ Skripsi$$

2. Klasifikasi dengan Regresi Logistik Biner

Analisis regresi logistik merupakan suatu analisis untuk mencari hubungan atau pengaruh, dimana variabel responnya berbentuk kategori dengan satu atau lebih variabel bebas.

Pada Tabel 3 dijelaskan bahwa ada pangaruh secara bersama antara status kelulusan (Y) dengan variabel bebas (X) yaitu Prodi, IPK, IP Semester 6, Jenis Kelamin, Nilai Ujian Masuk dan Lama Menyusun Skripsi hal ini ditandai oleh nilai G^2 pada model sebesar 597,195 dan nilai probabilitas 0,000 yang lebih kecil dari 0,05.

Tabel 4. Uji Koefisien Model

Variabel	B	Sig.	Exp(B)
Prodi(1)	2.224	0.000	0.108
IPK	4.551	0.000	0.011
IPs	1.070	0.210	0.343
JK(1)	0.041	0.911	0.960
NUM	0.010	0.424	0.990
LMS	2.898	0.000	18.142
Constant	13.446	0.000	691139.578

Pada uji koefisien model terlihat terlihat variabel prodi (1), IPK dan lama menyusun skripsi (LSM) berpengaruh secara parsial terhadap status kelulusan mahasiswa yang ditandai dengan nilai probabilitasnya lebih kecil dari alpha (0,000 < 0,05). Hasil uji koefisien model dapat dilihat pada (Tabel

4) sebagai berikut:

Model regresi di atas memiliki makna berdasarkan nilai *odds ratio* ($exp B$). $Exp(B) = 0,108$ artinya kecenderungan mahasiswa untuk lulus tidak tepat waktu pada program studi SK sebesar 0,108 kali dibandingkan mahasiswa SI. $Exp(B) = 0,011$ artinya kecenderungan mahasiswa untuk lulus tepat waktu sebesar 0,011 kali bila nilai IPK besar. Sedangkan $Exp(B) = 18,142$ artinya kecenderungan mahasiswa lulus tidak tepat waktu bila lama menyusun skripsi meningkat 18,142 kali. Nilai keragaman (R^2) 0,781 yang artinya berarti model yang diperoleh mampu menjelaskan status kelulusan mahasiswa dipengaruhi oleh variabel bebas sebanyak 78,1%.

Tabel 5. Ketepatan Klasifikasi pada Regresi Logistik Biner

Aktual	Prediksi		Ketepatan Klasifikasi
	Tepat Waktu	Tidak Tepat Waktu	
Tepat Waktu	265	36	88,0%
Tidak Tepat Waktu	31	350	91,9%
Ket. Klasifikasi Keseluruhan			90,2%

Berdasarkan Tabel 5, diketahui ketepatan pada regresi logistik biner sudah cukup baik, karena mampu menebak dengan benar 90,2% kondisi yang terjadi.

KESIMPULAN

Berdasarkan hasil dan pembahasan dapat diambil kesimpulan sebagai berikut:

- Hasil klasifikasi dengan menggunakan metode *CHAID regression trees* bahwa status kelulusan mahasiswa dikelompokkan ke dalam 5 klasifikasi yang berdasarkan nilai indek lebih dari 100. Masing-masing klasifikasi dipengaruhi oleh lama menyusun skripsi, IPK, IP semester 6 dan program studi.
- Hasil klasifikasi dengan menggunakan metode Regresi logistik Biner diperoleh bahwa klasifikasi status kelulusan mahasiswa dipengaruhi oleh

program studi, IPK dan lama menyusun skripsi

- Hasil perbandingan ketepatan klasifikasi diperoleh bahwa klasifikasi dengan menggunakan metode *CHAID regression trees* lebih baik yaitu 91,2% sedangkan dengan regresi logistik ketepatan klasifikasi mencapai 90,2%.

Saran yang dapat diberikan dalam penelitian ini adalah memberikan perhatian khusus kepada mahasiswa dalam menyelesaikan/menyusun skripsi supaya dipersingkat, karena berdasarkan kedua metode dinyatakan bahwa lama menyusun skripsi menjadi salah satu alasan klasifikasi dalam status kelulusan mahasiswa tepat waktu atau tidak tepat waktu.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada STIKOM Bali yang telah memberi dukungan financial terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] Suryabrata, S., 2008, Psikologi Pendidikan. Raja Grafindo Persada: Jakarta.
- [2] Nurgenita, Nita, 2015, Identifikasi Faktor-faktor Yang Mempengaruhi Ketepatan Waktu Kelulusan Mahasiswa Program Sarjana IPB. *Skripsi* Departemen Statistika FMIPA IPB Bogor. (Tidak dipublikasikan).
- [3] Suniantara, IKP, 2016, *Pendugaan Lama Studi Mahasiswa STIKOM BALI dengan Menggunakan Regresi Berstruktur Pohon*. Penelitian Internal Tahap I STIKOM BALI 2016.
- [4] Padmini, I.A.S., NLH Suciptawati dan Made Susilawati., 2012, Analisis Waktu Kelulusan Mahasiswa dengan Metode CHAID, (studi kasus: FMIPA Universitas Udayana). *E-jurnal Matematika*, Vol. 1, No. 1, Agustus 2012, 89-93.

- [5] Margasari, A., 2014, Penerapan metode CART (Classification and Regression Trees) dan analisis regresi logistik Biner pada klasifikasi profil mahasiswa FMIPA Universitas Brawijaya. *Studen Journal UB*, vol 2 no 4.
<http://statistik.studentjournal.ub.ac.id/index.php/statistik/article/view/151>
[diakses tanggal 7 Agustus 2015]