

Klasifikasi Ketepatan Masa Studi Mahasiswa FMIPA Unpad Angkatan 2001-2006 dengan Menggunakan Metode *Classification and Regression Trees* (CART)

Tiara Aprilia K, Nurul Gusriani, Kankan Parmikanti

Jurusan Matematika, Fakultas MIPA, Universitas Padjadjaran
Jl. Raya Bandung Sumedang KM 21 Jatinangor Sumedang 45363
Email: gusriani99@gmail.com, parmikanti@yahoo.co.id

ABSTRAK

Salah satu aspek penilaian dalam akreditasi yang dilakukan oleh BAN PT adalah melihat prosentase mahasiswa yang lulus tepat waktu yaitu dengan menempuh masa kuliah dalam delapan semester. Berkaitan dengan hal tersebut makalah ini dibuat berdasarkan penelitian untuk melihat klasifikasi ketepatan masa studi mahasiswa FMIPA Unpad tahun 2001-2006. Metode yang digunakan adalah *Classification and Regression Trees* (CART) yang merupakan salah satu metode pendekatan regresi non parametrik untuk teknik pohon keputusan. Variabel yang digunakan adalah ketepatan masa studi sebagai variabel respon, jenis kelamin, daerah asal, jenis ujian saringan, program studi dan IPK semester II sebagai variabel prediktor. Hasil yang diperoleh menunjukkan bahwa IPK semester II dan program studi yang dipilih mahasiswa menentukan hasil klasifikasi. Variabel IPK semester II menjadi variabel yang paling menentukan dalam klasifikasi dengan tingkat akurasi sebesar 80,87%.

Kata kunci: metode CART, pohon klasifikasi, regresi non parametric

ABSTRACT

One aspect of the assessment in the accreditation process conducted by BAN PT is looking at the percentage of students who graduate on time by taking the college in eight semesters. This paper is based on the research to look at the classification accuracy of the study period 2001-2006 students of the Science Faculty, Universitas Padjadjaran. The method that used is the Classification and Regression Trees (CART) which is one of non-parametric regression approach method to decision tree technique. Variables that used is the accuracy of the study period as the response variable, gender, hometown, type of test sieves, courses and GPA of second semester as a predictor variable. The results show that the second semester GPA and courses are chosen determine the classification results. GPA of second semester became the most decisive variable in the classification with accuracy of 80.87%.

Keywords: CART method, classification trees, non-parametric regression

1. Pendahuluan

Berkembangnya zaman menjadi tantangan negara untuk berkembang lebih baik dari waktu ke waktu. Hal ini menjadikan negara berusaha memberikan pelayanan sebaik mungkin kepada masyarakatnya, salah satunya dalam bidang pendidikan. Seperti yang tertuang dalam Bab I Pasal 1 No.6 Undang-Undang Republik Indonesia No.12 Tahun 2012, perguruan tinggi adalah satuan pendidikan yang menyelenggarakan pendidikan tinggi. Standar akreditasi adalah salah satu cara mengevaluasi perguruan tinggi untuk mengukur dan menetapkan mutu dan kelayakan dalam melaksanakan program-programnya. Menurut Peraturan Menteri Pendidikan Nasional Republik Indonesia No.73 Tahun 2009, salah satu elemen penilaian akreditasi adalah ketepatan masa studi yang ditempuh mahasiswa. Akan tetapi, saat ini banyak permasalahan yang muncul mengenai keterlambatan waktu kelulusan mahasiswa.

Pada program strata satu (S1), rata-rata waktu yang ditempuh untuk menyelesaikan studi adalah 4 tahun atau 8 semester, tetapi banyak juga mahasiswa tingkat akhir yang memiliki masalah keterlambatan waktu kelulusan. Masalah tersebut dapat dipengaruhi beberapa hal, seperti: IPK yang akan mempengaruhi pengambilan banyaknya SKS, masalah finansial, atau masalah-masalah lain yang timbul dari dalam diri mahasiswa itu sendiri maupun dari luar. Oleh karena itu perlu pemetaan mengenai karakteristik mahasiswa yang diprediksi akan

menempuh masa studi ≤ 4 tahun (tepat waktu) atau > 4 tahun (tidak tepat waktu). Pengklasifikasian menjadi alternatif cara untuk melihat karakteristik mahasiswa yang digolongkan lulus tepat waktu dan lulus tidak tepat waktu agar dapat menjadi gambaran untuk pihak jurusan, fakultas, maupun universitas.

Metode yang sering dipakai untuk masalah pengklasifikasian adalah analisis diskriminan dan regresi logistik. Kedua metode ini disebut sebagai metode tradisional karena tidak praktis dalam penggunaannya. Metode tersebut memiliki kelemahan, diantaranya banyaknya variabel prediktor yang menyulitkan untuk menentukan variabel yang paling berpengaruh, asumsi kenormalan yang sulit didapat, interaksi kompleks pada data, dan hasil metode tersebut sulit untuk digunakan (Lewis, 2000).

Menurut Eubank (1999), penggunaan metode regresi nonparametrik menjadi pemecahan masalah untuk mengatasi keterbatasan dalam penggunaan metode parametrik (analisis diskriminan dan regresi logistik) dalam melakukan penelitian.

CART adalah salah satu metode regresi nonparametrik yang digunakan sebagai metode pengklasifikasi. Metode ini dapat digunakan dalam data yang berdimensi tinggi, variabel yang digunakan dapat berupa kontinu maupun kategorik, serta dapat menunjukkan interaksi antar variabel. Pada makalah ini, metode CART diterapkan dalam mengklasifikasi ketepatan masa studi mahasiswa FMIPA Unpad angkatan 2001-2006 yang bertujuan untuk mengetahui variabel yang paling berpengaruh dalam ketepatan masa studi mahasiswa FMIPA Unpad angkatan 2001-2006 dan memberikan tingkat akurasi yang tinggi untuk hasil pengklasifikasian.

2. Tinjauan Pustaka

2.1 *Classification and Regression Trees (CART)*

Classification and Regression Trees atau yang selanjutnya disebut dengan CART merupakan salah satu metode pendekatan regresi nonparametrik yang digunakan untuk teknik pohon keputusan. Model yang dihasilkan merupakan model yang berdasarkan skala variabel respon. Jika variabel respon berbentuk kontinu (skala interval dan rasio) maka model pohon yang akan dihasilkan adalah pohon regresi, sedangkan bila variabel respon berbentuk kategori (skala nominal dan ordinal) maka model pohon yang akan dihasilkan adalah pohon klasifikasi (Breiman, 1984).

Menurut Lewis (2000), ada 4 komponen utama dalam suatu pengklasifikasian, yaitu variabel respon, variabel prediktor, data *learning*, dan data *testing*. Variabel respon adalah suatu karakteristik yang diharapkan dapat diprediksi dengan menggunakan variabel prediktor. Variabel prediktor adalah suatu karakteristik yang secara potensial mempengaruhi variabel yang akan diprediksi. Sehingga secara umum, akan terdapat banyak variabel prediktor yang mempengaruhi variabel respon. Data *learning* adalah himpunan data yang terdiri dari variabel respon dan prediktor yang akan diprediksi hasilnya. Data *testing* adalah himpunan data yang digunakan untuk mengetahui seberapa tepat model yang sudah dibentuk atau klasifikasi yang dihasilkan oleh data *learning*.

Lewis (2000) juga menyebutkan bahwa CART adalah salah satu metode klasifikasi *binary recursive partitioning*. Struktur pohon diperoleh melalui suatu algoritma penyekatan rekursif terhadap variabel prediktor. Penyekatan tersebut dimulai dua simpul anak berdasarkan variabel prediktor yang dianggap paling signifikan menjelaskan variabel responnya. Selanjutnya simpul anak ini disekat lagi masing-masing menjadi dua simpul anak baru. Penyekatan diulang sampai diperoleh kelompok-kelompok pengamatan yang mempunyai ciri-ciri yang relatif homogen berdasarkan variabel respon dan prediktornya.

Proses Pembentukan Pohon Klasifikasi

Pohon klasifikasi dibentuk oleh data *learning* sampel L yang terdiri atas n pengamatan. Menurut Breiman (1984) proses pembentukan pohon klasifikasi terdiri atas 3 tahapan sebagai berikut :

1. Pemilihan Pemilah (*Classifier*)

Pembentukan pohon klasifikasi dengan menggunakan data *learning* sampel L masih bersifat heterogen, sehingga sampel L tersebut perlu dipilah dengan menggunakan aturan pemilahan *goodness-of-split* untuk memilih simpul utama. Hal yang perlu digaris bawahi disini adalah cara untuk menentukan agar himpunan bagian yang dihasilkan dari pemilahan lebih homogen dari pemilahan sebelumnya.

Menurut Breiman (1984) pemilihan pemilah dilakukan dengan cara sebagai berikut:

- i. Mendefinisikan fungsi keheterogenan simpul

Fungsi keheterogenan indeks gini adalah sebagai berikut:

$$i(t) = 1 - \sum_{j=1}^d p^2(j|t) \quad (2.1)$$

dengan :

$i(t)$: indeks gini pada simpul t

j : kelas, dan $j=1,2,\dots,d$

t : simpul pada pohon klasifikasi, dan $t=1,2,3,\dots,m$

$p(j|t)$: peluang kelas j pada simpul t dan $p(j|t) = \frac{n_j(t)}{n(t)}$

$n_j(t)$: banyak pengamatan yang memiliki kelas j pada simpul t

$n(t)$: banyak pengamatan pada simpul t

- ii. Menentukan kriteria *goodness-of-split*

Goodness of split merupakan suatu evaluasi bagi pemilahan yang dilakukan oleh pemilah s pada simpul t . Misal terdapat pemilah s yang akan memilah t menjadi simpul kiri (t_L) dengan proporsi (p_L) dan simpul kanan (t_R) dengan proporsi (p_R), maka *goodness of split* $\Delta i(s,t)$ didefinisikan sebagai penurunan keheterogenan (Breiman, 1984) :

$$\Delta i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad (2.2)$$

Pengembangan pohon dilakukan dengan pencarian semua pemilah yang mungkin pada simpul pertama (t_1). Simpul t_1 akan dipilah menjadi t_2 dan t_3 dengan pemilah s^* yang memberikan nilai penurunan keheterogenan tertinggi, yaitu:

$$\Delta i(s^*,t_1) = \max \Delta i(s,t) \quad (2.3)$$

Cara ini juga dilakukan pada t_2 dan t_3 secara terpisah dan seterusnya.

2. Penentuan Simpul Terminal

Penentuan simpul terminal t akan ditentukan sebagai simpul terminal apabila:

- Tidak terdapat penurunan keheterogenan yang berarti sehingga simpul t tidak akan terpilah lagi.
- Hanya terdapat suatu pengamatan pada tiap simpul anak atau adanya batasan minimum n . Menurut Breiman (1984) pengembangan pohon akan berhenti apabila pada simpul terdapat $n_i < 5$.
- Adanya batasan jumlah tingkat kedalaman pohon maksimal kemudian pohon berhenti.

3. Penandaan Label Kelas

Penandaan label kelas pada simpul terminal ditentukan berdasarkan aturan jumlah terbanyak, yaitu jika,

$$p(j_0|t) = \underset{j}{maks} p(j|t) \tag{2.4}$$

dengan :

$$p(j|t) : \text{peluang kelas } j \text{ pada simpul } t$$

Maka label kelas untuk simpul terminal t adalah j_0 .

Pemilihan Pohon Optimal

Menurut Breiman (1984), pohon klasifikasi yang berukuran besar akan memberikan kesalahan pengklasifikasian paling kecil, sehingga pohon ini cenderung dipilih untuk menduga variabel respon. Tetapi ukuran pohon yang besar akan menyebabkan nilai kompleksitas yang tinggi karena struktur data yang digambarkan cenderung kompleks, sehingga perlu dipilih pohon optimal yang berukuran sederhana tetapi memberikan nilai kesalahan pengklasifikasian cukup kecil. Ada dua jenis penduga pengganti, yaitu : *test sample estimate* dan *cross validation v-fold estimate*. Breiman juga menyarankan penggunaan penduga *cross validation v-fold estimate* untuk menghitung biaya pengganti relatif pada sampel yang jumlahnya kurang dari 3000, sedangkan penggunaan pendekatan *test sample estimate* digunakan untuk jumlah sampel yang lebih besar dari 3000.

Langkah pertama pemilihan pohon optimal adalah menyusun *subtree* (pohon bagian) sebanyak k , yang dimulai dengan menentukan nilai *complexity cost* di simpul t ($g_m(t)$).

Perhitungan *complexity cost* di simpul t ($g_m(t)$) dinyatakan sebagai berikut:

$$g_m(t) = \begin{cases} \frac{R(t)-R(T_k)}{|T_k|-1}, t \notin T_k \\ \infty, t \in T_k \end{cases} \tag{2.5}$$

dengan :

$R(t)$: kesalahan pengklasifikasian pada simpul t

$$\text{dengan } R(t) = \frac{1 - \underset{j}{maks} p(j|t)}{n(t)}$$

T_k : *subtree* ke- k , dan $k = 1, 2, 3, \dots, l$

$R(T_k)$: kesalahan pengklasifikasian pada pohon T_k , dengan $R(T_k) = \sum_m R(t_{m_k})$

$R(t_{m_k})$: kesalahan pengklasifikasian pada simpul t ke m di pohon ke- k

\tilde{T}_k : simpul terminal pada pohon T_k

$|\tilde{T}_k|$: banyaknya simpul terminal pada pohon T_k

Pemangkasan dilakukan secara terus menerus yang dimulai pada cabang terlemah di T_1 atau disebut juga T_{maks} sampai akhirnya membentuk *subtree* terakhir yaitu *root node* (t_1). Cabang yang dipangkas adalah \tilde{t}_m , yaitu cabang yang memiliki nilai $g_m(t)$ terkecil, yaitu :

$$g_m(\tilde{t}_m) = \min_{t \in \tilde{T}_k} g_m(t) \tag{2.6}$$

Setelah *subtree* terbentuk, maka tiap *subtree* akan dibagi ke dalam v bagian. Menurut Breiman (1984), banyak v yang digunakan 10, dengan nilai *cross validation cost* :

$$R^c(T_k) = \frac{1}{n} \sum_{i,j} C(i|j) n_{ij} \tag{2.7}$$

dengan :

n : banyak pengamatan

$C(i|j)$: nilai kesalahan pengklasifikasian,

$$\text{dengan } C(i|j) = \begin{cases} 1, & \text{jika } i \text{ diprediksi sebagai } j \\ 0, & \text{jika } i \text{ diprediksi sebagai } i \end{cases}$$

n_{ij} : banyak pengamatan dengan kelas i yang diprediksi sebagai kelas j

Pohon optimal (T_{k_0}) yang dipilih adalah *subtree* yang meminimumkan $R^c(T_k)$, yaitu :

$$R^c(T_{k_0}) = \min R^c(T_k) \tag{2.8}$$

2.2 Tingkat Akurasi Klasifikasi

Mengevaluasi hasil klasifikasi untuk mengetahui akurasi hasil klasifikasi dilakukan dengan menghitung *total accuracy rate* yang merupakan peluang pengamatan yang diprediksi secara benar oleh fungsi klasifikasi.

Untuk perhitungan *total accuracy rate* dapat melalui Tabel 2.1 berikut :

Tabel 2.1 Hasil Klasifikasi

Aktual	Prediksi		Total
	0	1	
0	n_{00}	n_{01}	$n_{0\cdot}$
1	n_{10}	n_{11}	$n_{1\cdot}$
Total	$n_{\cdot 0}$	$n_{\cdot 1}$	n

Keterangan:

n_{00} : Jumlah pengamatan dari 0 yang tepat diprediksikan sebagai 0

n_{11} : Jumlah pengamatan dari 1 yang tepat diprediksikan sebagai 1

n_{10} : Jumlah pengamatan dari 1 yang tepat diprediksikan sebagai 0

n_{01} : Jumlah pengamatan dari 0 yang tepat diprediksikan sebagai 1

$n_{0\cdot}$: Jumlah pengamatan dari 0

$n_{1\cdot}$: Jumlah pengamatan dari 1

n : Jumlah pengamatan

$$\text{Total accuracy rate} = \frac{\text{Jumlah pengamatan yang benar diprediksi}}{\text{Jumlah pengamatan}}$$

$$= \frac{n_{00} + n_{11}}{n} \tag{2.9}$$

Perhitungan *total accuracy rate* seperti yang dijelaskan pada (2.9), digunakan untuk menentukan tingkat akurasi klasifikasi berdasarkan pohon klasifikasi yang dibentuk.

3. Objek Penelitian

Data yang digunakan adalah data mahasiswa lulusan FMIPA Unpad angkatan 2001-2006 yang diambil dari Sub Bag Pendidikan FMIPA Unpad. Variabel penelitian yang digunakan, yaitu ketepatan masa studi sebagai variabel respon dan jenis kelamin, daerah asal, jalur masuk, program studi, dan IPK semester II sebagai variabel prediktor. Pembagian IPK semester II berdasarkan Panduan Bimbingan Akademik FMIPA Unpad (Program Sarjana, Profesi, dan Diploma III) 2007/2008.

4. Hasil dan Pembahasan

Pada analisis CART, langkah pertama adalah membagi data menjadi data *learning* dan data *testing*. Penggunaan data *learning* dan data *testing* diambil dari data awal dengan membagi data menjadi dua bagian dengan proporsi data *learning* lebih besar dari proporsi data *testing*. Karena tidak adanya aturan khusus mengenai pembagian data, maka proporsi yang digunakan untuk pembagian data *testing* yaitu 10%-49% dari jumlah data. Tingkat akurasi tertinggi dari 40 pembagian data *testing* yang berbeda dihasilkan oleh pembagian data *testing* dan data *learning* 16%:84% dengan tingkat akurasi sebesar 80,87%. Selanjutnya pengklasifikasian ketepatan masa studi dengan metode CART yang dibahas menggunakan pembagian data *testing* dan data *learning* sebesar 16%:84%.

4.1 Pembentukan Pohon Maksimal

Pohon maksimal terdiri dari banyak simpul terminal dari penyekatan rekursif secara biner (*binary recursive partitioning*) pada simpul utama (*root node*) maupun pada simpul dalam (*internal node*) dan dibentuk dari data *learning*. Jumlah data mahasiswa lulusan FMIPA Unpad angkatan 2001-2006 yang digunakan dalam pengklasifikasian ini sebanyak 1883 data yang kemudian dibagi menjadi data *learning* untuk membuat model pohon klasifikasi dan data *testing* untuk perhitungan tingkat akurasi klasifikasi.

Langkah selanjutnya dalam pembentukan pohon maksimal adalah mencari pemilah yang mungkin dari setiap variabel prediktor yang akan menjadi pemilah utama. Variabel prediktor yang menjadi pemilah utama adalah variabel yang paling berpengaruh dalam pembentukan pohon klasifikasi. IPK semester II terpilih sebagai variabel paling berpengaruh dalam pembentukan pohon klasifikasi.

4.2 Pemilihan Pohon Optimal

Proses pemilihan pohon optimal dimulai dengan membentuk *subtree* yang dipangkas secara terus menerus dari pohon maksimal dan berhenti saat terbentuk pohon yang hanya terdiri dari simpul akar (*root node*). Pemangkasan secara terus menerus ini menghasilkan 13 *subtree*. Pohon optimal yang dipilih adalah *subtree* yang memiliki *cross validation cost* (CV cost) paling kecil yang berarti pohon tersebut tingkat kesalahan pengklasifikasiannya paling kecil juga.

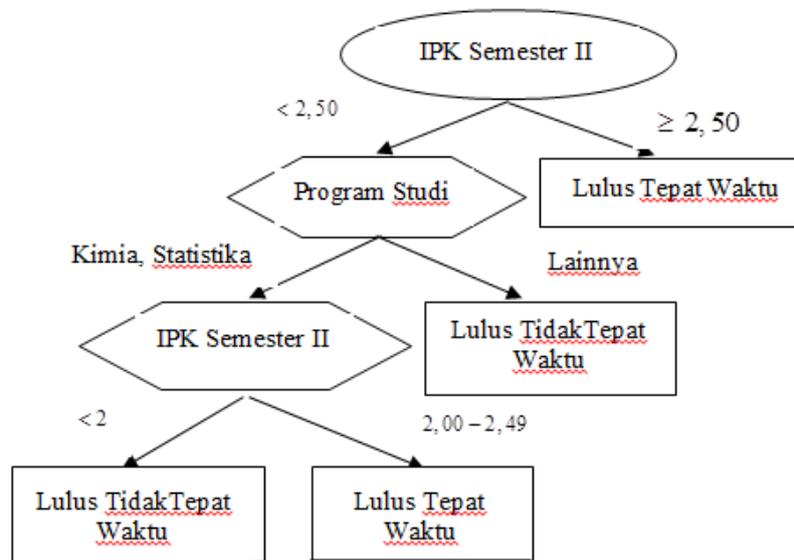
Pohon kesepuluh memiliki CV cost paling kecil yaitu sebesar 0,218191, oleh karena itu pohon ini dipilih sebagai pohon optimal yang dapat mewakili pohon klasifikasi yang struktur pohonnya sederhana dan memiliki nilai kesalahan pengklasifikasian paling kecil.

Di bawah ini adalah pohon optimal yang terpilih dengan 4 simpul terminal.

Penjelasan hasil klasifikasi pohon optimal pada Gambar 4.1 yang dapat dilihat dari simpul terminal, adalah sebagai berikut :

- a) Simpul terminal 3 : mahasiswa lulusan FMIPA Unpad yang memiliki IPK semester II $\geq 2,50$ diprediksi sebagai mahasiswa yang lulus tepat waktu.
- b) Simpul terminal 5 : mahasiswa lulusan FMIPA Unpad yang memiliki IPK semester II $< 2,50$ dan memilih program studi matematika, biologi, atau fisika diprediksi sebagai mahasiswa yang lulus tidak tepat waktu.
- c) Simpul terminal 6 : mahasiswa lulusan FMIPA Unpad yang memiliki IPK semester II $< 2,00$ dan memilih program studi kimia atau statistika diprediksi sebagai mahasiswa yang lulus tidak tepat waktu.

- d) Simpul terminal 7 : mahasiswa lulusan FMIPA Unpad yang memiliki IPK semester II sebesar 2,00–2,49 dan memilih program studi kimia atau statistika diprediksi sebagai mahasiswa yang lulus tepat waktu.



Gambar 4.1 Pohon Optimal

4.3 Hasil Tingkat Akurasi Klasifikasi

Pohon klasifikasi yang dibentuk oleh data *learning* selanjutnya dihitung tingkat akurasi klasifikasinya oleh data *testing*. Tabel 4.1 merupakan hasil program *STATISTICA 8* yang menunjukkan hasil prediksi data *testing* berdasarkan model pohon klasifikasi yang dibentuk oleh data *learning*. Penjelasan mengenai Tabel 4.1 pada dasarnya sama dengan penjelasan Tabel 2.1.

Tabel 4.1 Hasil Klasifikasi Pohon Klasifikasi

Summary Frequency Table (Standard Classification Trees with Deployment(C And RT))				
Marked cells have counts > 10				
(Marginal summaries are not marked)				
Lulus	TreeModelPred Tepat waktu	TreeModelPred Tidak tepat waktu	Row Totals	
Tepat waktu	201	14	215	
Tidak tepat waktu	43	40	83	
All Grps	244	54	298	

Tabel 4.1 menunjukkan bahwa terdapat 201 orang mahasiswa yg lulus tepat waktu juga diprediksi sebagai mahasiswa yang lulus tepat waktu, 14 orang mahasiswa yang lulus tepat waktu diprediksi sebagai mahasiswa yang lulus tidak tepat waktu, 43 orang mahasiswa yang lulus tidak tepat waktu diprediksi sebagai mahasiswa yang lulus tepat waktu, dan 40 orang mahasiswa yang lulus tidak tepat waktu juga diprediksi sebagai mahasiswa yang lulus tidak tepat waktu.

Tingkat akurasi dari hasil klasifikasi pada Tabel 4.1 dapat dihitung menggunakan rumus (2.9), yaitu :

$$Total\ accuracy\ rate = \frac{\text{Jumlah\ pengamatan\ yang\ benar\ diprediksi}}{\text{Jumlah\ pengamatan}} \times 100\%$$

$$= \frac{201 + 40}{298} \times 100\% = 80,87\%$$

Maka tingkat akurasi klasifikasi akurasi masa studi mahasiswa lulusan FMIPA Unpad angkatan 2001-2006 menggunakan metode CART adalah 80,87%.

5. Simpulan

Dari hasil pembahasan dalam penelitian diperoleh simpulan sebagai berikut :

1. Pohon klasifikasi yang terbentuk menggunakan metode CART dengan proporsi data *learning* 84% dan data *testing* 16% menghasilkan tingkat akurasi klasifikasi terbaik dibandingkan dengan proporsi data lainnya. Hasil klasifikasi sebagai berikut :
 - e) Mahasiswa lulusan FMIPA Unpad yang memiliki IPK semester II > 2,49 diprediksi sebagai mahasiswa yang lulus tepat waktu.
 - f) Mahasiswa lulusan FMIPA Unpad yang memiliki IPK semester II sebesar ≤ 2,49 dan memilih program studi matematika, biologi, atau fisika diprediksi sebagai mahasiswa yang lulus tidak tepat waktu.
 - g) Mahasiswa lulusan FMIPA Unpad yang memiliki IPK semester II < 2,00 dan memilih program studi kimia atau statistika diprediksi sebagai mahasiswa yang lulus tidak tepat waktu.
 - h) Mahasiswa lulusan FMIPA Unpad yang memiliki IPK semester II sebesar 2,00 – 2,49 dan memilih program studi kimia atau statistika diprediksi sebagai mahasiswa yang lulus tepat waktu.
2. Variabel yang paling berpengaruh dalam klasifikasi ketepatan masa studi mahasiswa FMIPA Unpad angkatan 2001-2006 adalah IPK semester II.
3. Tingkat akurasi klasifikasi yang didapatkan dari pembagian data *learning* dan data *testing* sebesar 84%:16% adalah 80,87%.

Daftar Pustaka

1. Breiman, L., Friedman, J., Olshen, R. and Stone, C., 1984. *Classification and Regression Trees*. Chapman Hall, New York – London
2. Eubank, Randall L. 1999. *Nonparametric Regression and Spline Smoothing*. Second Edition. Marcel Dekker, Inc. New York.
3. Fakultas Matematika dan Ilmu Pengetahuan Alam. 2007. *Panduan Bimbingan Akademik (Program Sarjana, Profesi, dan Diploma III) 2007/2008*. Bandung: Universitas Padjadjaran
4. Lewis, J.Roger. 2000. *An Introduction to Classification and Regression Tree (CART) Analysis*. Presented at the 2000 Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California.
5. *Peraturan Menteri Pendidikan Nasional Republik Indonesia No 73 Tahun 2009 Tentang Perangkat Akreditasi Program Studi Sarjana (S1)*. 2009. Jakarta : Salinan oleh Kepala Biro Hukum dan Organisasi Departemen Pendidikan Nasional. *Undang-Undang Republik Indonesia No.12 Tahun 2012 Tentang Pendidikan Tinggi*. 2012. Jakarta: Tambahan Lembaran Negara Republik Indonesia No. 5336