

KLASIFIKASI PENERIMA BEASISWA KOPERTIS DENGAN MENGGUNAKAN ALGORITMA C.45

Muhamad Tabrani

Program Studi Teknik Informatika
STMIK Nusa Mandiri Jakarta

Jl. Damai No.8 Warung Jati Barat Marga Satwa Jakarta Selatan
muhammad.mtb@bsi.ac.id

Abstract — Every year college get scholarship coordinator private universities (kopertis), Scholarship devoted to student in a college namely scholarship an increase in academic performance (scholarship PPA) and scholarship assistance learn student (scholarship BBM). The process filing scholarship ppa and bbm through two stages selection selection is the first stage selection in college to determine a candidate scholarship recipients that would be proposed to kopertis, the selection that both the stage in kopertis selection. Many a student who submitted the scholarships as well as surpassing kouta given resulted in the process of select recipients taken longer since select must be in accordance with the criteria so that the recipient scholarship right on target . Based on these problems need a the act of determining scholarship recipients proper. The purpose of this research is to make classifications students scholarship recipients with algorithm C4.5. The results of classifications evaluate and validated with confusion matrix and a curve ROC, the results classifications students scholarship recipients namely algorithm C4.5 with the level of accuracy of 86.88 %, So that it can be applied for the problem the determination of scholarship recipients .

Intisari— Setiap tahunnya perguruan tinggi mendapatkan beasiswa dari Koordinator Perguruan Tinggi Swasta (Kopertis), beasiswa yang ditujukan untuk mahasiswa di perguruan tinggi yaitu Beasiswa Peningkatan Prestasi Akademik (Beasiswa PPA) dan Beasiswa Bantuan Belajar Mahasiswa (Beasiswa BBM). Proses pengajuan beasiswa PPA dan BBM melalui dua tahap seleksi yaitu seleksi pertama merupakan tahap seleksi di perguruan tinggi untuk menentukan calon penerima beasiswa yang akan diusulkan ke kopertis. Tahap seleksi yang kedua yaitu tahap seleksi di kopertis. Banyaknya mahasiswa yang mengajukan beasiswa tersebut serta melebihi kouta yang diberikan mengakibatkan proses penyeleksian penerima memakan waktu yang lama karena penyeleksian harus sesuai dengan kriteria agar penerima beasiswa tepat sasaran. Berdasarkan permasalahan tersebut perlu suatu tindakan

untuk menentukan penerima beasiswa yang tepat. Tujuan Penelitian ini adalah membuat klasifikasi mahasiswa penerima beasiswa dengan algoritma C4.5. Hasil klasifikasi dievaluasi dan divalidasi dengan *confusion matrix* dan kurva ROC, hasilnya klasifikasi mahasiswa penerima beasiswa yaitu algoritma C4.5 dengan tingkat akurasi 86.88%, sehingga dapat diterapkan untuk permasalahan penentuan penerima beasiswa.

Kata Kunci: Algoritma C4.5, Klasifikasi

PENDAHULUAN

Setiap tahunnya masing-masing perguruan tinggi mendapatkan beasiswa dari Koordinator Koordinator Perguruan Tinggi Swasta (Kopertis) beasiswa yang ditujukan untuk mahasiswa di perguruan tinggi swasta yaitu Beasiswa Peningkatan Prestasi Akademik (Beasiswa PPA) dan Beasiswa Bantuan Belajar Mahasiswa (Beasiswa BBM). Beasiswa PPA adalah beasiswa yang diberikan kepada para mahasiswa jenjang Strata Satu dan Diploma Tiga yang mempunyai prestasi akademik baik yaitu minimal Indeks Prestasi Kumulatif (IPK) minimal 3,00. Sedangkan beasiswa BBM adalah beasiswa yang diberikan kepada para mahasiswa jenjang Strata Satu dan Diploma Tiga yang mempunyai prestasi akademik minimal IPK 2,50 dan aktif dalam kegiatan kemahasiswaan serta kondisi orangtuanya kurang mampu. Tujuan dari pemberian beasiswa PPA dan BBM ini untuk membantu mahasiswa yang kurang mampu dalam membayar biaya studi di perguruan tingginya.

Hasil seleksi penerima beasiswa yang dilakukan secara manual dapat berbeda-beda tergantung dari pengambil keputusan yang terlibat dalam proses penentuan penerima beasiswa (Karismariyanti, 2011). Hal ini sering menimbulkan kesalahan dalam penentuan calon penerima beasiswa seperti terpilihnya penerima beasiswa yang kurang tepat. Selain itu proses pengambilan keputusan untuk menentukan calon penerima beasiswa yang dilakukan secara manual sulit dan membutuhkan waktu lama. Untuk membantu dalam pengambilan keputusan

diperlukan suatu klasifikasi mahasiswa penerima beasiswa.

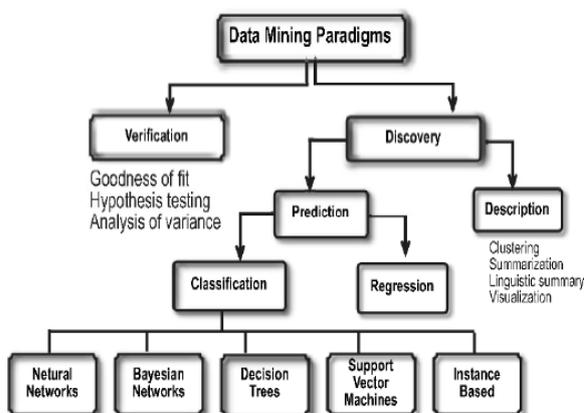
Klasifikasi merupakan salah satu dari teknik data mining yang merupakan proses penempatan objek atau konsep tertentu ke dalam satu set kategori berdasarkan objek yang digunakan. Dari berbagai teknik klasifikasi yang paling populer digunakan adalah *decision tree* (Han & Kamber, 2006). Sedangkan algoritma yang dapat dipakai dalam *decision tree* salah satunya adalah Algoritma C4.5

Algoritma C4.5 dilakukan untuk mencari hasil klasifikasi dengan tingkat akurasi yang tinggi dalam prediksi mahasiswa penerima beasiswa. Rule algoritma dengan tingkat akurasi tertinggi selanjutnya diaplikasikan dalam membuat program seleksi penerima beasiswa PPA dan BBM. Program seleksi penerima beasiswa PPA dan BBM yang digunakan untuk mempermudah dan mempercepat proses seleksi di perguruan tinggi dapat dibangun dengan menggunakan *software* Microsoft Visual Basic 6.0 dalam menerapkan *rule* hasil klasifikasi mahasiswa penerima beasiswa.

BAHAN DAN METODE

Klasifikasi

Klasifikasi atau taksonomi adalah proses menempatkan suatu objek atau konsep kedalam satu set kategori berdasarkan objek atau konsep yang bersangkutan (Gorunescu, 2011). Ada banyak metode data mining yang digunakan untuk tujuan yang berbeda-beda. Metode klasifikasi digunakan untuk membantu dalam memahami pengelompokan data. Klasifikasi sendiri merupakan cabang dari *discovery data mining* seperti yang ditunjukkan pada gambar 1. (Maimon & Rokach, 2010).



Sumber: (Maimon & Rokach, 2010)

Gambar 1. Data mining taksonomi

Algoritma C4.5

Algoritma C4.5 dan pohon keputusan merupakan dua model yang tak terpisahkan, karena untuk membangun sebuah pohon keputusan, dibutuhkan algoritma C4.5. Di akhir tahun 1970 hingga di awal tahun 1980-an, J. Ross Quinlan seorang peneliti di bidang mesin pembelajaran mengembangkan sebuah model pohon keputusan yang dinamakan ID3 (*Iterative Dichotomiser*), walaupun sebenarnya proyek ini telah dibuat sebelumnya oleh E.B. Hunt, J. Marin, dan P.T. Stone. Kemudian Quinlan membuat algoritma dari pengembangan ID3 yang dinamakan C4.5 yang berbasis *supervised learning*.

Menurut (Witten, Frank, & Hall, 2011) serangkaian perbaikan yang dilakukan pada ID3 mencapai puncaknya dengan menghasilkan sebuah sistem praktis dan berpengaruh untuk *decision tree* yaitu C4.5. Perbaikan ini meliputi metode untuk menangani *numeric attributes*, *missing values*, *noisy data*, dan aturan yang menghasilkan *rules* dari *trees*.

Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5 (Larose, 2005) yaitu :

1. Mempersiapkan data *training*. Data *training* biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus :

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2 p_i \dots (1)$$

Keterangan :

S= Himpunan kasus

n = jumlah partisi S

P_i = proporsi S_i terhadap S

Kemudian hitung nilai gain menggunakan rumus :

$$Gain(S,A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots (2)$$

Keterangan :

S = Himpunan Kasus

A = Fitur

n = jumlah partisi atribut A

|S_i| = Proporsi S_i terhadap S

|S| = jumlah kasus dalam S

3. Ulangi langkah ke 2 dan langkah ke 3 hingga semua *record* terpartisi

4. Proses partisi pohon keputusan akan berhenti saat :
 - a. semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut didalam *record* yang dipartisi lagi
 - c. Tidak ada *record* didalam cabang yang kosong

Model Dan Variabel

Data mahasiswa di AMIK "BSI Yogyakarta" tahun angkatan 2008 sampai dengan 2013 terdapat 133 mahasiswa dengan mahasiswa yang mengajukan beasiswa, untuk siswa yang mendapatkan beasiswa kopertis sebanyak 60 mahasiswa baik dalam bentuk beasiswa BBM ataupun beasiswa PPA, Sampai saat ini belum diketahui algoritma yang paling akurat dalam melakukan klasifikasi mahasiswa yang berpotensi diterima dalam beasiswa kopertis (PPA dan BBM).

Atribut dan nilai atribut diperoleh dari data mahasiswa dan data nilai. Adapun atribut yang digunakan dalam penelitian ini antara lain:

1. Semester
Merupakan atribut yang berisi semester aktif mahasiswa yang mengajukan beasiswa
2. IPK
Merupakan atribut yang berisi data nilai Indeks Prestasi Kumulatif (IPK)
Berdasarkan Keputusan Menteri Pendidikan Nasional Republik Indonesia No 232/U/2000 Tentang Pedoman Penyusunan Kurikulum Pendidikan Tinggi dan Penilaian Hasil Belajar mahasiswa Pasal 15 ayat 2 :
IPK sebagai dasar penentuan predikat kelulusan program sarjana dan program diploma adalah:
 - a. IPK 2,00 - 2,75 : memuaskan
 - b. IPK 2,76 - 3.50 : sangat memuaskan
 - c. IPK 3.51 - 4,00 : dengan pujian.
3. Orangtua
Merupakan atribut yang berisi data status keberadaan orangtua masih ada atau sudah tidak ada (yatim).
4. Penghasilan ortu/wali
Merupakan atribut yang berisi penghasilan orangtua/wali perbulannya. penghasilan yang didapat oleh keluarga baik berupa uang ataupun jasa. Berdasarkan penggolongannya, Badan Pusat Statistik (BPS, 2008) membedakan pendapatan menjadi 4 golongan adalah:
 - a. Golongan pendapatan Sangat tinggi, adalah jika pendapatan rata-rata lebih dari Rp. 3.500.000,00 per bulan
 - b. Golongan pendapatan Tinggi, adalah jika pendapatan rata-rata antara Rp.

- 2.500.000,00 s/d Rp. 3.500.000,00 per bulan
- c. Golongan pendapatan Sedang, adalah jika pendapatan rata-rata antara Rp. 1.500.000,00 s/d Rp. 2.500.000,00 per bulan
- d. Golongan pendapatan Rendah, adalah jika pendapatan rata-rata Rp. 1.500.000,00 per bulan
5. Beasiswa
Merupakan atribut yang berisi apakah mahasiswa sebelumnya pernah mendapat beasiswa baik dari kampus atau instansi lainnya
6. Berkas
Merupakan atribut yang berisi berkas yang dikirim lengkap atau tidak lengkap
7. Remark
Merupakan atribut yang berisi DAPAT atau TIDAK pada program Beasiswa kopertis (PPA dan BBM)

Analisa data

Penemuan frase pengetahuan dalam database diciptakan pada *Knowledge Discovery in Databases* (KDD) pertama dalam lokakarya di 1989 untuk menekankan bahwa pengetahuan adalah produk akhir dari penemuan basisdata. KDD adalah analisis eksplorasi secara otomatis dan pemodelan *repository* data yang besar. KDD adalah proses terorganisir untuk mengidentifikasi pola dalam data yang besar dan kompleks dimana pola data tersebut ditemukan yang bersifat sah, baru, dan dapat bermanfaat serta dapat dimengerti (Maimon & Rokach, 2010).

Penemuan pengetahuan dari proses KDD yang terdiri dari sembilan langkah, Proses dimulai dari penentuan KDD dan diakhiri dengan pelaksanaan pengetahuan yang ditemukan.

Berikut ini adalah Sembilan langkah dalam KDD yaitu:

1. *Developing an understanding of the application domain.*
Pada tahap ini dilakukan pengembangan pemahaman domain aplikasi untuk persiapan langkah awal untuk memahami apa yang harus dilakukan dengan banyak keputusan (tentang transformasi, algoritma, representasi, dan lain-lain).
2. *Selecting and creating a data set on which discovery will be performed.*
Memilih dan menciptakan satu set data yang akan digunakan untuk penelitian harus ditentukan, termasuk mencari tahu data apa yang tersedia, memperoleh data tambahan yang diperlukan, dan kemudian mengintegrasikan semua data untuk penemuan *knowledge* ke dalam satu set data

termasuk atribut yang akan dipakai untuk proses tersebut.

Data yang digunakan yaitu data pengajuan beasiswa kopertis di AMIK "BSI Yogyakarta" jurusan Manajemen Informatika tahun 2008 sampai dengan 2013 sejumlah 133 data dimana data tersebut terdapat data mahasiswa yang mendapatkan beasiswa kopertis dan tidak mendapatkan beasiswa.

3. *Preprocessing and cleansing.*

Pada tahap ini kehandalan data ditingkatkan dengan membersihkan data yang tidak lengkap (*missing value*) dan data tidak benar (*noise*).

Data mahasiswa yang ada diseleksi dengan memilih data yang baik dalam hal ini data yang lengkap dan menghilangkan data yang terdapat nilai kosong (*missing value*) dan atau data tidak benar (*noise*). Data yang dipakai untuk klasifikasi sejumlah 133 data.

4. *Data Transformation*

Pada tahap ini disusun dan dikembangkan generasi data yang lebih baik untuk data mining. Tahap ini juga merupakan proses transformasi pada data yang telah dipilih sehingga data tersebut sesuai untuk proses data mining. Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

Data yang sudah disiapkan untuk klasifikasi dibagi menjadi dua untuk data *training* (80%) dan data *testing* (20%). Pembagian data menjadi data *training* dan data *testing* menggunakan teknik sampling random sistematis (*systematic random sampling*). Cara penggunaan teknik sampling random sistematis ini perandoman atau pengundian hanya dilakukan satu kali, yakni ketika menentukan unsur pertama dari sampling yang akan diambil. Penentuan unsur sampling selanjutnya ditempuh dengan cara memanfaatkan interval sampel. Interval sampel adalah angka yang menunjukkan jarak antara nomor-nomor urut yang terdapat dalam kerangka sampling yang akan dijadikan patokan dalam menentukan atau memilih unsur-unsur sampling kedua dan seterusnya hingga unsur ke-n. Interval sampel biasanya dilambangkan dengan huruf k (Sugiana, 2008).

5. *Choosing the Data Mining Algorithm*

Setelah pemilihan jenis data mining yang akan digunakan yaitu klasifikasi, maka selanjutnya menentukan algoritma klasifikasi yang akan digunakan. Pada penelitian ini algoritma yang dipilih adalah C4.5

6. *Employing the Data Mining Algorithm*

Tahap ini dilakukan untuk pengolahan data dengan algoritma yang telah dipilih untuk mendapatkan algoritma terbaik dengan tingkat akurasi yang tinggi dalam klasifikasi mahasiswa penerima beasiswa.

7. *Evaluation*

Dalam tahap ini dilakukan evaluasi dan menafsirkan pola yang didapatkan dari hasil algoritma yang dipakai untuk mengetahui aturan, kehandalan, dan lain-lain. Evaluasi dilakukan dengan menerapkan pola yang didapat dari proses sebelumnya terhadap data *testing* yang disediakan. Evaluasi dilakukan dengan *confusion matrix* dan kurva ROC.

8. *Using the discovered knowledge*

Pada tahap ini menggunakan pengetahuan yang diperoleh dari proses data mining untuk penerapan pada pembuatan program atau aplikasi maupun lainnya. Pengetahuan klasifikasi mahasiswa penerima beasiswa diterapkan pada data baru untuk membuat klasifikasi mahasiswa penerima beasiswa.

HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menentukan akurasi kelayakan pemberian beasiswa yang menggunakan metode algoritma C4.5, Setelah itu membandingkan nilai akurasi dari kedua metode tersebut, dalam menentukan hasil penelitian ini menggunakan data *training* berjumlah 106 data dan data *testing* berjumlah 26.

Algoritma C4.5

Langkah-langkah untuk membuat algoritma C.45 dengan memakai data *training* yang berjumlah 106, yaitu :

- Siapkan data *training* yaitu tabel 3.3 yang berjumlah 106 data.
- Hitung jumlah mahasiswa penerima beasiswa kopertis (PPA dan BBM) yang gagal berdasarkan nilai tiap atribut.
- Hitung nilai *entropy* total dimana diketahui penerima beasiswa kopertis sebanyak 46 mahasiswa dan yang gagal berjumlah 60.

$$Entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i$$

$$= ((-46/106) * \log_2 (46/106)) + ((-60/106) * \log_2 (60/106))$$

$$= 0.98738$$

- Hitung nilai *gain* untuk masing-masing atribut. Kemudian tentukan nilai *gain* tertinggi. Atribut dengan nilai *gain* tertinggi maka atribut tersebut dijadikan sebagai

akar. Sebagai contoh hitung nilai *gain* untuk atribut Semester yaitu:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

$$= 0.98738 - ((36/106 * 0.96408) + (52/106 * 0.96124) + (18/106 * 0))$$

$$= 0.188407$$

Perhitungan nilai *entropy* dan *gain* untuk semua atribut dilakukan untuk mendapatkan nilai *gain* tertinggi yang akan dijadikan sebagai akar. Hasil perhitungannya terlihat di tabel di bawah ini:

Tabel 1. Hasil nilai *entropy* dan *gain* untuk menentukan simpul akar dengan data *training*

Node		Jumlah Kasus (S)	DAPAT	TIDAK	Entropy	Gain
1	Total	106	46	60	0.98738	
	Semester					0.188407
	2	36	14	22	0.96408	
	4	52	32	20	0.96124	
	6	18	0	18	0	
	IPK					0.0956666
	2.00 - 2.75	4	0	4	0	
	2.76 - 3.50	61	20	41	0.91273	
	3.51 - 4.00	41	26	15	0.94744	
	Orang Tua					0.0839632
	ADA	99	39	60	0.96729	
	YATIM	7	7	0	0	
	Penghasilan					0.1159426
	RENDAH	50	32	18	0.94268	
	SEDANG	36	9	27	0.81128	
	TINGGI	10	3	7	0.88129	
	SANGAT TINGGI	10	2	8	0.72193	
	Beasiswa					0.0317983
	TIDAK	102	46	56	0.99306	
	PENERIMA	4	0	4	0	
	Berkas					0.0740404
	LENGKAP	97	46	51	0.99808	
	TIDAK LENGKAP	9	0	9	0	

Sumber : Pengolahan Data (2014)

Berdasarkan hasil perhitungan *gain* pada tabel 1 terlihat atribut semester mempunyai nilai *gain* tertinggi yaitu 0.188407 sehingga atribut semester dijadikan sebagai simpul akar dari pohon keputusan.

Tentukan simpul selanjutnya setelah simpul akar yaitu *node* 1.1 dengan menghitung nilai *entropy* dan *gain* semua atribut berdasarkan atribut Semester dengan nilai 2.

Pembentukan simpul-simpul dengan perhitungan *gain* diperoleh *decision tree* untuk

klasifikasi mahasiswa penerima beasiswa seperti pada gambar di bawah ini:

Tabel 2. Perhitungan *entropy* dan *gain* untuk menentukan *node* 1.1

Node		Jumlah Kasus (S)	DAPAT	TIDAK	Entropy	Gain
1.1	SEMESTER	36	14	22	0.96408	
	2					
	IPK					0.0452619
	2.00 - 2.75	2	0	2	0	
	2.76 - 3.50	21	8	13	0.95871	
	3.51 - 4.00	13	6	7	0.99573	
	Orang Tua					0
	ADA	36	14	22	0.96408	
	YATIM	0	0	0	0	
	Penghasilan					0.1469743
	RENDAH	18	10	8	0.99108	
	SEDANG	13	4	9	0.89049	
	TINGGI	0	0	0	0	
	SANGAT TINGGI	5	0	5	0	
	Beasiswa					0.0409619
	TIDAK	34	14	20	0.97742	
	PENERIMA	2	0	2	0	
	Berkas					0.0852348
	LENGKAP	32	14	18	0.9887	
	TIDAK LENGKAP	4	0	4	0	

Sumber: Pengolahan Data (2014)

Berdasarkan hasil perhitungan *gain* pada tabel 4 terlihat atribut penghasilan mempunyai nilai *gain* tertinggi yaitu 0.1469743 sehingga atribut penghasilan dijadikan sebagai *node* 1.1 dari semester.

Tentukan simpul selanjutnya yaitu *node* 1.2 dengan menghitung nilai *entropy* dan *gain* semua atribut berdasarkan atribut Semester dengan nilai 4.

Tabel 3. Perhitungan nilai *entropy* dan *gain* untuk menentukan *node* 1.2

Node		Jumlah Kasus (S)	DAPAT	TIDAK	Entropy	Gain
1.2	SEMESTER	52	32	20	0.9612	
	4					
	IPK					0.1162921
	2.00 - 2.75	1	0	1	0	
	2.76 - 3.50	26	12	14	0.9957	
	3.51 - 4.00	25	20	5	0.7219	
	Orang Tua					0.1035746
	ADA	45	25	20	0.9911	
	YATIM	7	7	0	0	
	Penghasilan					0.118849
	RENDAH	28	22	6	0.7496	
	SEDANG	14	5	9	0.9403	
	TINGGI	7	3	4	0.9852	
	SANGAT TINGGI	3	2	1	0.9183	
	Beasiswa					0.0548105
	TIDAK	50	32	18	0.9427	
	PENERIMA	2	0	2	0	
	Berkas					0.1446536
	LENGKAP	47	32	15	0.9035	
	TIDAK LENGKAP	5	0	5	0	

Sumber: Pengolahan data (2014)

Berdasarkan hasil perhitungan *gain* pada tabel 3 terlihat atribut penghasilan mempunyai

nilai gain tertinggi yaitu 0.1446536 sehingga atribut berkas dijadikan sebagai *node* 1.2 dari Semester.

Tentukan simpul selanjutnya yaitu *node* 1.2 dengan menghitung nilai *entropy* dan *gain* semua atribut berdasarkan atribut Semester dengan nilai 4.

Tabel 4. Perhitungan nilai *entropy* dan *gain* untuk menentukan *node* 1.3

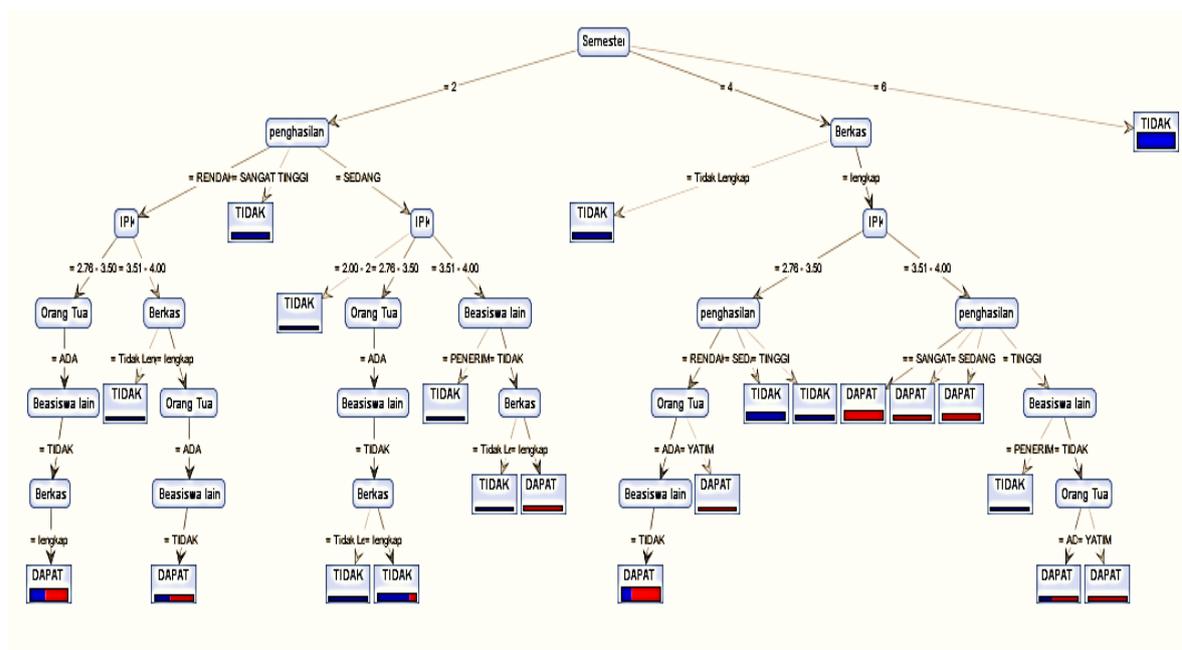
Node		Jumlah Kasus (s)	DAPAT	TIDAK	Entropy	Gain
1.3	SEMESTER	18	0	18	0	
	6					

Sumber: Pengolahan data (2013)

Berdasarkan hasil perhitungan *gain* pada tabel 4 terlihat hasil sudah terklasifikasi.

Tentukan simpul-simpul lainnya dengan mencari *gain* tertinggi dari masing-masing perhitungan *node* berdasarkan *node* yang sudah ditentukan hingga ditemukan hasil klasifikasinya.

Pembentukan simpul-simpul dengan perhitungan *gain* diperoleh *decision tree* untuk klasifikasi mahasiswa penerima beasiswa seperti pada gambar 1



Sumber: Pengolahan data (2014)

Gambar 2. *Decision Tree* Klasifikasi mahasiswa dengan C4.5

Berdasarkan pohon keputusan pada gambar 1, dapat membentuk aturan-aturan, yaitu sebagai berikut:

- R1: If Semester = 2 AND Penghasilan = Rendah AND Ipk = 2.76 - 3.50 AND Orang Tua = Ada AND Beasiswa lain= Tidak AND Berkas = Lengkap THEN DAPAT
- R2: If Semester = 2 AND Penghasilan = Rendah AND Ipk = 3.51 - 4.00 AND Berkas = Tidak Lengkap THEN TIDAK
- R3: If Semester = 2 AND Penghasilan = Rendah AND Ipk = 3.51 - 4.00 AND Berkas = Lengkap AND Orang Tua = Ada AND Beasiswa Lain = Tidak THEN DAPAT
- R4: If Semester = 2 AND Penghasilan = Sangat Tinggi THEN TIDAK
- R5: If Semester = 2 AND Penghasilan = Sedang AND Ipk = 2.00 - 2.75 THEN TIDAK
- R6: If Semester = 2 AND Penghasilan = Sedang AND Ipk = 2.75 - 3.50 AND Orang Tua = Ada AND Beasiswa Lain = Tidak AND Berkas = Tidak Lengkap THEN TIDAK
- R7: If Semester = 2 AND Penghasilan = Sedang AND Ipk = 2.75 - 3.50 AND Orang Tua = Ada AND Beasiswa Lain = Tidak AND Berkas = Lengkap THEN DAPAT
- R8: If Semester = 2 AND Penghasilan = Sedang AND Ipk = 3.51 - 4.00 AND Beasiswa Lain = Penerima THEN TIDAK
- R9: If Semester = 2 AND Penghasilan = Sedang AND Ipk = 3.51 - 4.00 AND Beasiswa Lain = Tidak AND Berkas = Tidak Lengkap THEN TIDAK

- 10. R10: If Semester = 2 AND Penghasilan = Sedang AND Ipk = 3.51 – 4.00 AND Beasiswa Lain = Tidak AND Berkas = Lengkap THEN DAPAT
- 11. R11: If Semester = 4 AND Berkas = Tidak Lengkap THEN TIDAK
- 12. R12: If Semester = 4 AND Berkas = Lengkap AND Ipk = 2.76 – 3.50 Penghasilan = Rendah AND Orang tua = Ada AND Beasiswa Lain = Tidak THEN DAPAT
- 13. R13: If Semester = 4 AND Berkas = Lengkap AND Ipk = 2.76 – 3.50 Penghasilan = Rendah AND Orang tua = Yatim THEN DAPAT
- 14. R14: If Semester = 4 AND Berkas = Lengkap AND Ipk = 2.76 – 3.50 Penghasilan = Sedang THEN TIDAK
- 15. R15: If Semester = 4 AND Berkas = Lengkap AND Ipk = 3.51 – 4.00 Penghasilan = Tinggi THEN TIDAK
- 16. R16: If Semester = 4 AND Berkas = Lengkap AND Ipk = 3.51 – 4.00 Penghasilan = Rendah THEN DAPAT
- 17. R17: If Semester = 4 AND Berkas = Lengkap AND Ipk = 3.51 – 4.00 Penghasilan = Sangat Tinggi THEN DAPAT
- 18. R18: If Semester = 4 AND Berkas = Lengkap AND Ipk = 3.51 – 4.00 Penghasilan = Sedang THEN DAPAT
- 19. R19: If Semester = 4 AND Berkas = Lengkap AND Ipk = 3.51 – 4.00 Penghasilan = Tinggi AND Beasiswa Lain = Penerima THEN TIDAK
- 20. R20: If Semester = 4 AND Berkas = Lengkap AND Ipk = 3.51 – 4.00 Penghasilan = Tinggi AND Beasiswa Lain = Tidak AND Orang Tua = ADA THEN DAPAT
- 21. R21: If Semester = 4 AND Berkas = Lengkap AND Ipk = 3.51 – 4.00 Penghasilan = Tinggi AND Beasiswa Lain = Tidak AND Orang Tua = Yatim THEN DAPAT
- 22. R22: If Semester = 6 THEN TIDAK

Hasil dari pengujian model yang telah dilakukan, dilakukan pengujian tingkat akurasi dengan menggunakan confusion matrix dan kurva ROC/AUC (*Area Under Cover*).

1. *Confusion Matrix*

Tabel 5 adalah perhitungan akurasi data training menggunakan algoritma C4.5. Diketahui dari 106 data training, dengan menggunakan metode algoritma C4.5 didapat 48 data prediksi tidak sesuai dengan tidak, 1 prediksi tidak ternyata Dapat, 12 data prediksi Dapat ternyata tidak, dan 45 data prediksi dapat sesuai dengan Dapat.

Tabel 5. *Confusion Matrix* data training

accuracy: 87.74%			
	true TIDAK	true DAPAT	class precision
pred. TIDAK	48	1	97.96%
pred. DAPAT	12	45	78.95%
class recall	80.00%	97.83%	

Sumber: Pengolahan data (2014)

Tabel 6 adalah perhitungan akurasi data testing menggunakan algoritma C4.5. Diketahui dari 27 data testing, dengan menggunakan metode algoritma C4.5 didapat 14 data prediksi.

Dapat sesuai dengan *Dapat*, 4 prediksi Dapat ternyata Tidak, 1 data prediksi Tidak ternyata dapat, dan 8 data predisi tidak sesuai dengan tidak.

accuracy: 81.48%			
	true DAPAT	true TIDAK	class precision
pred. DAPAT	14	4	77.78%
pred. TIDAK	1	8	88.89%
class recall	93.33%	66.67%	

Sumber: Pengolahan data (2014)

2. Evaluasi dengan Kurva ROC

ROC memiliki tingkat nilai diagnosa yaitu(Gorunescu, 2011):

Akurasi bernilai $0.90 - 1.00 = \text{excellent classification}$

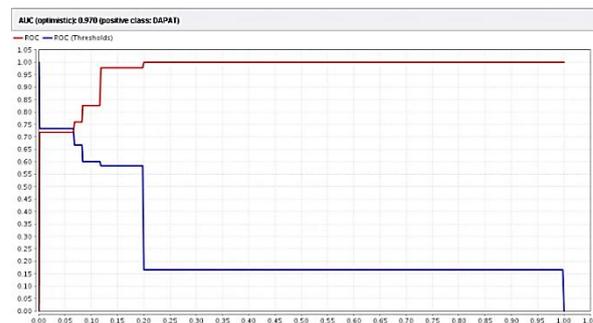
Akurasi bernilai $0.80 - 0.90 = \text{good classification}$

Akurasi bernilai $0.70 - 0.80 = \text{fair classification}$

Akurasi bernilai $0.60 - 0.70 = \text{poor classification}$

Akurasi bernilai $0.50 - 0.60 = \text{failure}$

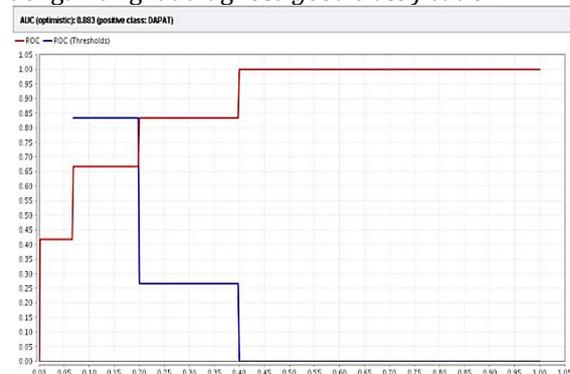
Hasil yang didapat dari pengolahan ROC untuk algoritma C4.5 dengan menggunakan data training sebesar 0.970 dapat dilihat pada gambar 2 dengan tingkat diagnosa *excellent classification*



Sumber: Pengolahan data (2014)

Gambar 2 Grafik ROC dengan data training

Hasil yang didapat dari pengolahan ROC untuk algoritma C4.5 dengan menggunakan data testing sebesar 0.883 dapat dilihat pada gambar 3 dengan tingkat diagnosa *good classification*



Sumber: Pengolahan data(2014)

Gambar 3 Grafik ROC dengan data testing

Berikut aplikasi yang dihasilkan berdasarkan rule yang ada menggunakan aplikasi Visual basic 6.0

Sumber: Pengolahan data (2014)

Gambar 4 Aplikasi klasifikasi Beasiswa

Pada aplikasi untuk klasifikasi mahasiswa pada gambar 4 dihasilkan klasifikasi penerima beasiswa. Input data mahasiswa pada program tersebut sesuai dengan atribut yang dibutuhkan, kemudian klik tombol HASIL KLASIFIKASI, maka secara otomatis tampil hasil klasifikasi mahasiswa mendapatkan beasiswa atau tidak. Untuk menginput kembali data baru klik tombol INPUT DATA LAGI, dan tombol KELUAR digunakan untuk keluar dari aplikasi tersebut.

KESIMPULAN

Dalam pengujian yang telah dilakukan dengan menggunakan atribut diperoleh dari data mahasiswa dan data nilai. Adapun atribut yang digunakan dalam penelitian ini antara lain: Semester, IPK, Orangtua, Penghasilan ortu, Beasiswa, Berkas, Remark. Data di olah dengan menggunakan algoritma C4.5 memiliki tingkat akurasi 87.74%, dengan menggunakan data training ROC 0.970 termasuk dalam kategori *excellent classification* dan pada testing memiliki tingkat ROC yaitu 0.883, termasuk dalam katagori *good classification*. Dan dibuatkan aplikasi klasifikasi beasiswa dengan menggunakan rule yang dihasilkan.

REFERENSI

- Gorunescu, F. 2011. *Data Mining Concepts, Models and Techniques*. Berlin: Springer.
- Han, J., & Kamber, M. 2006. *Data Mining Concepts And Techniques 2nd Edition*. San Fransisco: Elsevier.
- Jones, P. 2001. *Visual Basic: A Complete Course*. London: Continuum.
- Karismariyanti, M. 2011. Simulasi Pendukung Keputusan Penerima Beasiswa

- Menggunakan Metode *Composite Performance Index*. Jurnal Teknologi Informasi Vol.1 , 54-59.
- Kotsiantis, S., & Pintelas. P. 2004. *A Decision Support Prototype Tool for Predicting Student Performance in an ODL Environment. Greece: University of Patras.*
- Larose, D. T. 2005. *Discovering Knowledge in Data An Introduction to Data Mining. New Jersey: John Willey and Sons.*
- Liao, T. W. 2007. *Enterprise Data Mining: A Review and Research Directions. Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications* , 1-109.
- Liberty, J. 2005. *Programming Visual Basic 2005*. USA: O'Reilly Media.
- Maimon, O., & Rokach, L. 2010. *Data Mining and Knowledge Discovery Handbook Second Edition. London: Springer.*
- Powers, D. 2011. *Evaluation: From Precision, Recall and F-Measure To ROC, Informedness, Markedness & Correlation. Journal of Machine Learning Technologies* , 37-63.
- Santiary, P. A. 2012. Sistem Pendukung Keputusan Cerdas dalam Penentuan Beasiswa. Jurnal Logic Vol.12 No.2 , 87-91.
- Suryana, T. 2009. *Visual Basic*. Yogyakarta: Graha Ilmu.
- Vercellis, C. 2009. *Business Intelligence. United Kingdom: John Wiley and Sons.*
- Witten, I. H., Frank, E., & Hall, M. A. 2011. *Data Mining Practical Machine Learning Tools and Techniques Third Edition. Burlington: Elsevier.*
- Wu, X., & Kumar, V. 2009. *The Top Ten Algorithms in Data Mining. USA: CRC Press.*
- www.dikti.go.id/januari-13,2014.
<http://www.dikti.go.id/files/Lemkerma/kepmen232-2000.txt>

BIODATA PENULIS



Muhamad Tabrani, M.Kom
Menyelesaikan S1 Sistem Informasi STMIK Nusa Mandiri Jakarta, S2 Magister Ilmu Komputer di Pascasarjana STMIK Nusa mandiri Jakarta, Saat ini menjadi Dosen tetap di STMIK Nusa Mandiri Jakarta. Bidang ilmu yang ditekuni ilmu komputer khususnya pemograman, beberapa kali mengikuti seminar baik nasional maupun seminar internasional, pemakalah International Seminar on Scientific Issues and Trends (ISSIT) BSI 2014.