

Penggunaan Pohon Keputusan untuk Klasifikasi Tingkat Kualitas Mahasiswa Berdasarkan Jalur Masuk Kuliah

Selly Artaty Zega

Program Studi Teknik Multimedia dan Jaringan, Jurusan Teknik Informatika, Politeknik Negeri Batam
Batam, Indonesia
zega.selly@gmail.com

Abstract—Penelitian ditujukan untuk mengetahui tingkat kualitas mahasiswa program studi Teknik Informatika Politeknik Negeri Batam berdasarkan jalur masuk kuliah dan faktor lainnya sebagai data yang tersimpan dalam basis data mahasiswa. Penelitian dilakukan pada data mahasiswa angkatan 2007-2009, dari data tersebut dilakukan pembagian *data training* dan *data testing*.

Data mining mengekstraksi data untuk menemukan informasi dan *pattern* dalam menentukan tingkat kualitas mahasiswa. Metode *classification* dilakukan pada *data training* hingga menemukan suatu *rule* (proses *learning*) yang dapat diterapkan untuk mengklasifikasikan kategori tingkat kualitas mahasiswa pada data baru. Proses *learning* menggunakan teknik *decision tree C.45* mengubah fakta menjadi pohon keputusan yang merepresentasikan aturan yang lebih mudah dipahami. Selanjutnya validasi terhadap *data testing* dilakukan untuk menemukan model *decision tree* terbaik.

Analisis juga menghasilkan informasi tentang pengaruh jalur masuk kuliah dalam menentukan tingkat kualitas mahasiswa.

Keywords—*data mining; decision tree; classification; rule; jalur masuk; kualitas*

I. PENDAHULUAN

Kualitas mahasiswa menjadi tolak ukur dalam menghasilkan sumber daya manusia yang kompeten. Kualitas tersebut dapat dilihat dari IPK, kemampuan untuk menyelesaikan perkuliahan tepat waktu, Surat Peringatan (SP) dan faktor lainnya. Salah satu faktor yang dapat dikaji adalah jalur masuk kuliah yang memberikan informasi proses seleksi calon mahasiswa. Politeknik Negeri Batam memiliki dua jalur masuk kuliah, yaitu melalui Ujian Masuk Politeknik Negeri Batam (UMPB) dan melalui Penelusuran Minat dan Keterampilan (PMDK).

Jalur masuk kuliah menjadi sarana yang digunakan untuk menjaring mahasiswa masuk perguruan tinggi. Masing-masing jalur masuk kuliah memiliki kualifikasi yang berbeda. Kualifikasi yang diberikan bertujuan untuk melihat kemampuan yang dimiliki calon mahasiswa sebelum menjadi mahasiswa di perguruan tinggi tersebut. Berdasarkan kajian internal yang dilakukan oleh Universitas Indonesia, terlihat bahwa tingkat IPK mahasiswa UI dari lulusan SNMPTN tidak terlalu bagus dibandingkan dengan mahasiswa yang lulus tes Seleksi Masuk (SIMAK UI)

maupun Prestasi dan Pemerataan Kesempatan Belajar (PPKB) [1]. Analisis tersebut memperlihatkan bahwa jalur masuk kuliah memberikan gambaran tingkat kualitas mahasiswa.

Pohon keputusan atau *decision tree* merupakan teknik *data mining* yang digunakan untuk mengeksplorasi data dengan membagi kumpulan data yang besar menjadi himpunan *record* yang lebih kecil dan memperhatikan variabel tujuannya. Teknik ini dapat diterapkan dalam menentukan tingkat kualitas mahasiswa yang melibatkan banyak data. Beberapa penelitian telah banyak dilakukan dengan menggunakan teknik *data mining*, salah satunya penggunaan *Naive Bayes Classifier* dalam evaluasi kinerja akademik mahasiswa [2].

Oleh karena itu, dengan menggunakan teknik pohon keputusan terhadap data yang ada, penulis memandang perlu dilakukan suatu analisis untuk memperoleh informasi tentang parameter yang mempengaruhi tingkat kualitas mahasiswa serta tingkat kualitas mahasiswa berdasarkan jalur masuk kuliah di Politeknik Negeri Batam. Hasilnya dapat digunakan sebagai perbaikan terhadap metode jalur masuk kuliah yang selama ini digunakan.

II. LANDASAN TEORI

A. Kualitas Mahasiswa

Referensi [3] memberikan definisi kualitas mahasiswa adalah tingkat mutu seseorang yang belajar di perguruan tinggi untuk menilai kecakapan dan kepandaiannya. Sesuai dengan sasaran mutu Politeknik Negeri Batam, ada beberapa hal yang menjadi pertimbangan dalam menganalisis kualitas mahasiswa agar dapat diserap dalam dunia kerja. Sesuai dengan data yang tersedia untuk penelitian, maka 3 pertimbangan yang digunakan adalah:

- 1) *Standar seleksi mahasiswa 1 : 3.*
- 2) *85% mahasiswa lulus tepat waktu dengan Indeks Prestasi rata-rata 3,0.*
- 3) *Seluruh lulusan terserap pasar kerja dengan waktu tunggu maksimal tiga bulan.*

B. Data Mining

Data mining adalah proses mengekstraksi dan mengidentifikasi informasi yang bermanfaat, menemukan

hubungan, *pattern* dan kecenderungan dengan memeriksa sekumpulan data dari basis data yang besar[4].

Data mining merupakan tahapan inti (*core*) dalam proses KDD (*Knowledge Discovery in Database*). Tahapan tersebut meliputi pemahaman ruang lingkup, menciptakan data target dan data *selection*, *preprocessing data*, dan menentukan teknik data mining. Teknik tersebut berupa *summarization*, *classification*, *regretion*, *association* dan *clustering*.

C. Metode Klasifikasi

Klasifikasi adalah suatu proses menemukan sebuah model berdasarkan *data training* dan nilai kelas label dari atribut target serta menggunakannya untuk mengklasifikasikan suatu data[5].

Data training adalah data yang siap untuk di-*mining* yang telah melewati *data preprocessing*. Sedangkan *data testing* adalah data yang digunakan untuk menguji *rule* klasifikasi yang diperoleh dari *data training*. Berdasarkan [5]terdapat dua langkah dalam proses *data classification*, yaitu:

1) *Learning*. Proses *learning* membangun algoritma klasifikasi dengan menganalisis atau belajar dari *data training*. Karena kelas label untuk setiap *data training* telah tersedia/diketahui, maka metode klasifikasi tergolong *supervised learning*. Dari analisis *data training* tersebut, terbentuklah *classification rules*.

2) *Classification*. Dari *classification rules*, dilakukan pengujian terhadap *data testing* untuk memperkirakan/mengestimasi akurasi *rule* yang diperoleh. Jika *rule* tepat, maka dapat diaplikasikan pada data yang baru.

D. Teknik Decision Tree Menggunakan Algoritma C4.5 dalam Metode Klasifikasi

Decision tree adalah *flowchart* seperti pohon dimana setiap *node* menunjukkan suatu *test* pada suatu atribut, tiap *branch* merepresentasikan hasil dari *test* tersebut, dan *leaf node* menunjukkan kelas-kelas atau distribusi kelas[5]. Referensi [6] menyatakan bahwa *decision tree* berguna untuk mengeksplorasi data yang sudah melewati tahap *preprocessing* dan menemukan model yang tersembunyi dari data dengan sebuah target variabel sehingga dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan *record* yang lebih kecil dengan memperhatikan variabel tujuannya.

Salah satu algoritma untuk membentuk pohon keputusan adalah C4.5. Algoritma C4.5 adalah pengembangan dari algoritma ID3. Cara kerja algoritma ID3 adalah membuat pohon dengan percabangan awal berupa atribut yang memiliki nilai paling signifikan. Pengembangan algoritma ID3 menjadi C4.5 terlihat dari beberapa perbedaan, yaitu: algoritma C4.5 mampu menangani atribut dengan tipe data diskrit atau kontinu, mampu menangani atribut dengan data kosong serta bisa melakukan pemangkasan cabang. Secara umum langkah untuk membangun algoritma C4.5 adalah[7].

- 1) Memilih atribut yang menjadi *root*.
- 2) Membuat cabang untuk setiap nilai.

- 3) Membagi kasus dalam cabang.
- 4) Mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Persamaan (1) digunakan untuk menghitung *gain* dalam memilih atribut yang menjadi *root*.

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

Keterangan:

- S : himpunan kasus
- A : atribut
- n : jumlah partisi atribut A
- |S_i| : jumlah kasus pada partisi ke-i
- |S| : jumlah kasus dalam S

Persamaan (2) digunakan untuk menghitung nilai *entropy*.

$$Entropy(S) = \sum_{i=1}^n p_i * \log_2 p_i \quad (2)$$

Keterangan:

- S : himpunan kasus
- A : fitur
- n : jumlah partisi S
- p_i : proporsi dari S_i terhadap S

Setelah menghitung *entropy* dan *gain* maka akan menghasilkan *decision tree*. *Decision tree* kemudian diuji untuk mengetahui estimasi keakuratannya. Semakin sedikit *error rate* (kesalahan) yang dihasilkan dari *decision tree* maka semakin akurat *decision tree* yang dihasilkan. Persamaan (3) digunakan untuk menghitung *error rate*.

$$Error\ rate(ER) = \frac{\text{banyaknya data error}}{\text{jumlah data}} * 100\% \quad (3)$$

E. K-Fold Cross Validation

K-Fold Cross Validation merupakan salah satu metode dalam menentukan *data training* dan *data testing* dari keseluruhan data. *K-fold cross validation* mengulang *k*-kali untuk membagi sebuah himpunan data secara acak menjadi *k-subset* yang saling bebas, setiap ulangan disisakan satu *subset* untuk pengujian dan *subset* lainnya untuk pelatihan[9]. Nilai *K* yang disarankan adalah 10 karena lebih akurat dalam ukuran estimasinya[5].

Dalam proses iterasi harus dilakukan sebanyak 10 kali dan dalam bagian S1 sampai S10 harus pernah dijadikan *data testing* sebanyak 1 kali. *K* bernilai 10 sehingga dalam proses iterasi dihasilkan *decision tree* sebanyak 10.

Untuk memilih model *decision tree* yang terbaik dari proses *training* sebaiknya memenuhi syarat-syarat sebagai berikut[8]:

1) Model *decision tree* yang jumlah aturan paling banyak. Semakin banyak jumlah aturan yang diperoleh, penanganan data juga lebih bervariasi. Apabila ditemukan beberapa iterasi yang menghasilkan *rule* yang sama, *rule* tersebut tidak dapat digunakan, hal ini dikarenakan *rule* yang

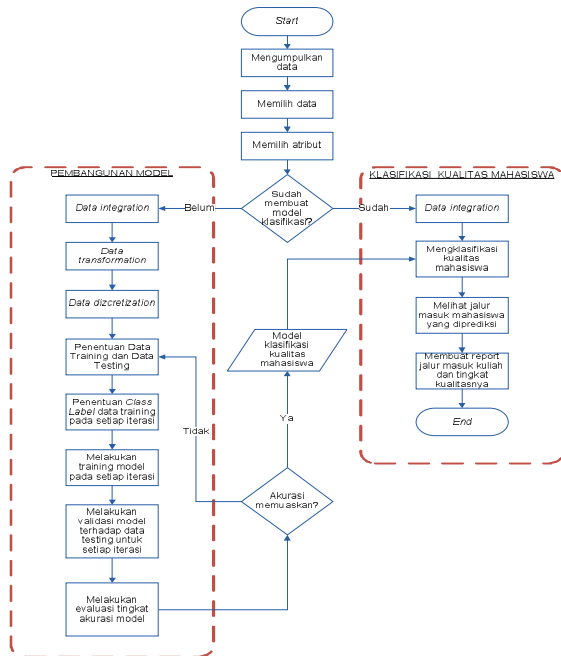
dihasilkan tersebut menangani data yang sama, sehingga tidak variatif.

2) Model *decision tree* yang memiliki tingkat akurasi tertinggi.

3) Model yang mencakup semua kelas target yang mungkin muncul dalam test set.

III. ANALISIS

Proses dalam menganalisis tingkat kualitas mahasiswa ditunjukkan pada Gbr.1.



Gbr 1. Proses Analisis dan Klasifikasi Tingkat Kualitas Mahasiswa

Pada penelitian ini populasi yang diteliti adalah kumpulan mahasiswa Politeknik Negeri Batam Program Studi Teknik Informatika angkatan 2007-2009. Metode pengumpulan data yang digunakan adalah metode sensus. Jumlah data operasional tiga angkatan yang diperoleh sebanyak 331 data. Data yang diambil adalah:

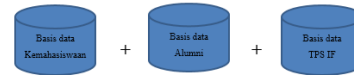
- 1) NIM dan data jalur masuk mahasiswa.
- 2) Data akademis mahasiswa meliputi IP semester 1, IPK, Surat Peringatan (SP), mata kuliah yang mengulang, tidak naik tingkat dan waktu tunggu kerja (2007). Semua data diperhitungkan untuk masa tempuh perkuliahan selama 4 semester untuk angkatan 2008 dan 2 semester untuk angkatan 2009.

A. Data Preprocessing

Data yang akan di-mining harus melewati *data preprocessing*, karena sumber data yang diperoleh dari banyak *database* masih bersifat kotor, tidak lengkap, banyak data yang kosong dan tidak konsisten. Kualitas data mempengaruhi hasil *data mining*.

1) Data Integration

Data integration melakukan penggabungan data dari berbagai sumber data. NIM dan jalur masuk mahasiswa diperoleh dari basis data kemahasiswaan, waktu tunggu kerja mahasiswa yang sudah lulus yang diperoleh dari basis data alumni. IPK, Surat Peringatan, IP semester 1, jumlah mata kuliah yang mengulang, dan tidak naik tingkat diperoleh dari basis data TPS.



Gbr 2. Data Integration

2) Data Transformation

Melakukan perubahan sumber data agar dapat dimengerti ketika di-mining. Label setiap *instance* juga digunakan dalam perhitungan *total score* (total penjumlahan dari perkalian label dengan bobot atribut) yang akan menentukan tingkat kualitas mahasiswa (*class label*).

TABEL I. LABEL SURAT PERINGATAN

Surat Peringatan	Label
Ada	0
Tidak Ada	1

TABEL II. LABEL MENGULANG MATA KULIAH

Mengulang Mata Kuliah	Label
Ada	0
Tidak Ada	1

TABEL III. LABEL TIDAK NAIK TINGKAT

Tidak Naik Tingkat	Label
Ada	0
Tidak Ada	1

3) Data Discretization

Melakukan perubahan data dengan menggolongkan *range* (interval) data tertentu ke level tertentu dan digunakan untuk data numerik. Level yang diperoleh dijadikan *score* untuk perhitungan penentuan tingkat kualitas mahasiswa pada *data training* untuk setiap iterasi. Persamaan (4) digunakan untuk menghitung interval kelas IPK dan IP semester 1. Persamaan (5) untuk mendapatkan jumlah kelas (*k*).

$$c = \frac{x_n - x_1}{k} \quad (4)$$

$$k = 1 + 3,322 \log n \quad (5)$$

Jumlah keseluruhan data yang telah melewati *preprocessing* adalah 331 data. Misalnya untuk menemukan interval IPK pada data iterasi satu (S1) yang berjumlah 298 data.

$n = 298$ (jumlah data training dijelaskan pada tabel IV)

$$k = 1 + 3,322 \log n$$

$$k = 1 + 3,322 \log 298$$

$$k = 9,219346429$$

$k = 9$ (dibulatkan, maka jumlah kelas pada S1 ada 9)

Nilai 9 menjadi jumlah kelas atau label untuk atribut IPK.

Mencari interval (c) IPK pada iterasi satu (S_1)

$$c = \frac{X_n - X_1}{k}$$

X_n = data IPK tertinggi pada *data training* iterasi satu = 3,97

X_1 = data IPK terendah pada *data training* iterasi satu = 0,00

$k=9$

$$c = (3,97 - 0,00)/9 = 0,441111111 \approx 0,44$$

TABEL IV. LABEL IPK DAN IP SEMESTER 1 DATA ITERASI 1

Instance IPK (x)	Label
$0,00 < x <= 0,44$	1
$0,45 < x <= 0,88$	2
$0,89 < x <= 1,32$	3
$1,33 < x <= 1,76$	4
$1,77 < x <= 2,21$	5
$2,22 < x <= 2,65$	6
$2,66 < x <= 3,09$	7
$3,10 < x <= 3,53$	8
$3,54 < x <= 3,97$	9

Pada data iterasi 1 (S_1), nilai IPK tertinggi dan terendah sama dengan nilai IP semester 1, sehingga Tabel IV juga berlaku untuk pemberian label IP semester 1.

B. Pemilihan Data Training dan Data Testing

Metode yang digunakan dalam pemilihan *data training* dan *data testing* adalah dengan menggunakan *K-fold cross validation*. Cara penggunaan 10 *cross validation* dimulai dengan membagi jumlah data yang telah melewati proses *preprocessing* ke dalam 10 bagian (S_1, S_2, \dots, S_{10}). Jumlah data analisis kualitas mahasiswa sebanyak 331 data. Lalu 331 data dibagi sebanyak 10 bagian. Hasil pembagiannya adalah 33,1 dibulatkan menjadi 33 data. Sehingga jumlah data untuk setiap iterasi = 33 data, kecuali pada S_{10} jumlah data = 34 data. Tabel V menyajikan pembagian *data training* dan *data testing*. Tabel VI menyajikan contoh *data training* yang telah melewati *data preprocessing*.

TABEL V. DATA TRAINING DAN DATA TESTING

Iterasi	Data Training		Data Testing	
	Subset	Jumlah Data	Subset	Jumlah Data
1	S2-S10	298	S1	33
2	S1, S3-S10	298	S2	33
3	S1, S2, S4-S10	298	S3	33
4	S1-S3, S5-S10	298	S4	33
5	S1-S4, S6-S10	298	S5	33
6	S1-S5, S7-S10	298	S6	33
7	S1-S6, S8-S10	298	S7	33
8	S1-S7, S9, S10	298	S8	33
9	S1-S8, S10	298	S9	33
10	S1-S9	297	S10	34

TABEL VI. CONTOH DATA TRAINING YANG TELAH MELEWATI DATA PREPROCESSING

N O	ID	IPK	IP SEM 1	SP	MENGU LANG MATA KULIAH	TIDAK NAIK TINGKAT	JALUR MASUK	WAKTU TUNGGU KERJA
1	x7	3,87	3,91	TIDAK ADA	TIDAK ADA	TIDAK ADA	PMDK	BELUM BEKERJA
2	x9	3,64	3,18	TIDAK ADA	TIDAK ADA	TIDAK ADA	PMDK	0 BULAN
3	x10	3,59	3,18	TIDAK ADA	ADA	TIDAK ADA	PMDK	BELUM BEKERJA

N O	ID	IPK	IP SEM 1	SP	MENGU LANG MATA KULIAH	TIDAK NAIK TINGKAT	JALUR MASUK	WAKTU TUNGGU KERJA
4	x12	3,78	3,64	TIDAK ADA	TIDAK ADA	TIDAK ADA	PMDK	BELUM BEKERJA
5	x14	3,55	3,45	TIDAK ADA	TIDAK ADA	TIDAK ADA	PMDK	BELUM BEKERJA

C. Penentuan Tingkat Kualitas Mahasiswa sebagai Class Label berdasarkan Angket

Class label yang digunakan merupakan kategori tingkat kualitas mahasiswa yang diperoleh berdasarkan angket. Angket disebar kepada Bagian Pendidikan, Bagian Alumni dan seluruh Dosen Teknik Informatika.

1) Hasil Angket

Angket ditujukan untuk mengambil sampling pendapat responden mengenai penentuan kualitas mahasiswa. Hasil angket disajikan pada tabel VII:

TABEL VII. PARAMETER KUALITAS MAHASISWA DAN BOBOT AKHIR

Parameter menilai mahasiswa	untuk kualitas	Rata-rata Bobot	Perhitungan	Bobot Akhir
IPK	X_1	32,85	$\frac{32,85}{79,46} \times 100\%$	41,34 %
IP Semester 1	X_2	8,08	$\frac{8,08}{79,46} \times 100\%$	10,16 %
Surat Peringatan	X_3	11,08	$\frac{11,08}{79,46} \times 100\%$	13,94 %
Mengulang mata kuliah	X_4	12,69	$\frac{12,69}{79,46} \times 100\%$	15,97 %
Tidak tingkat	X_5	14,77	$\frac{14,77}{79,46} \times 100\%$	18,59 %
$\sum_{i=1}^5 X_i$		79,46	Total	100 %

Selain itu, angket juga menghasilkan informasi:

a) Responden lebih banyak memilih dua kategori dalam penentuan tingkat kualitas mahasiswa, yaitu "BERKUALITAS" dan "TIDAK BERKUALITAS".

b) Responden lebih banyak memilih jalur masuk kuliah mempengaruhi penentuan kualitas mahasiswa.

c) Responden lebih banyak memilih jalur masuk PMDK yang menghasilkan mahasiswa berkualitas. Hasil ini nantinya akan dibandingkan dengan hasil dari analisis data mining.

2) Penentuan Class Label

Berdasarkan hasil angket, diperoleh bobot parameter penentu kualitas mahasiswa yang disajikan pada tabel VII. Dari bobot tersebut, dilakukan perhitungan *score* yang akan digunakan untuk menentukan interval tingkat kualitas mahasiswa (dalam bentuk kategori). Persamaan (6) dan (7) digunakan untuk menghitung total *score* yang diperoleh dari hasil perkalian label dengan bobot pada setiap parameter penentu kualitas mahasiswa. Perhitungan ini berlaku untuk semua *data training* pada setiap iterasi.

$$\text{Score} = \text{Label} \times \text{Bobot Parameter} \quad (6)$$

$$\text{Total Score} = \sum_{i=1}^5 \text{score} \quad (7)$$

Setelah memperoleh total *score*, selanjutnya nilai tersebut digunakan untuk perhitungan interval tingkat kualitas mahasiswa. Misalnya penentuan kategori tingkat kualitas untuk *data training* pada iterasi 1. Perhitungannya interval kategori tingkat kualitas menggunakan (4).

$$c = \frac{X_n - X_1}{k}$$

X_n = nilai rata-rata *score* tertinggi pada *data training* iterasi satu = 5,12

X_1 = nilai rata-rata *score* terendah pada *data training* iterasi satu = 1,00

$k = 2$ (sesuai hasil angket pertanyaan nomor 3, jumlah kategori yang dipilih adalah 2)

$$c = \frac{4,97 - 0}{2}$$

$$c = 0,26$$

Sehingga, interval *score* yang dipakai dalam menentukan tingkat kualitas mahasiswa pada *data training* S1 disajikan pada tabel VIII.

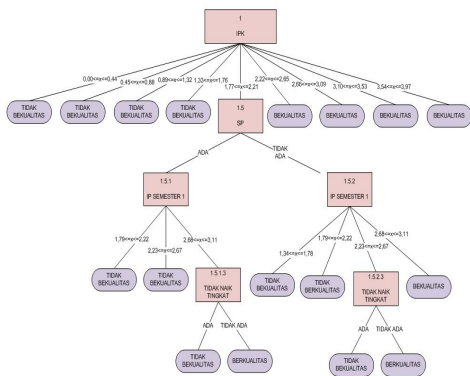
TABEL VIII. CLASS LABEL DATA TRAINING S1

Interval Score (x)	Kategori Tingkat Kualitas (class label)
$1,00 \leq x \leq 3,06$	Tidak Berkualitas
$3,07 \leq x \leq 5,12$	Berkualitas

IV. PENERAPAN METODE KLASIFIKASI

A. Learning

Mengacu pada *k-fold validation* dengan k bernilai 10, maka dilakukan sebanyak 10 kali iterasi dalam pembentukan model *decision tree*. Dari model *decision tree* akan terbentuk *classification rules*. Di antara kesepuluh model yang terbentuk, terdapat iterasi yang memiliki model *decision tree* yang sama, yaitu: pada iterasi **3, 5, 8, 9**, dan **10** yang menghasilkan 9 *rule*.



Gbr 3. Decision Tree Iterasi 3,5,8,9 dan 10

B. Classification

Selanjutnya dilakukan validasi *decision tree* terhadap *data testing* untuk setiap iterasi. Dari hasil validasi tersebut, diperoleh perhitungan *error rate* yang menjadi gambaran *performance* model terhadap *data testing* dan akan menjadi

dasar pengambilan keputusan dalam pemilihan *tree*. Nilai *error rate* setiap iterasi, jumlah *rule* ditampilkan pada tabel IX.

TABEL IX. ERROR RATE MODEL DATA TRAINING

Data testing Iterasi	Error Rate	ID Data Error	Jumlah Rule
S1	0 %		7
S2	0 %		8
S3	0 %		9
S4	0 %		8
S5	0 %		9
S6	0 %		7
S7	3,03 %	y107	11
S8-S10	0 %		9

Validasi model *decision tree* iterasi 7 menemukan satu data yang tidak terklasifikasikan dan satu data *error*, setelah ditelusuri *data testing* iterasi ke-7, data tersebut adalah:

1) Id y92, memiliki IPK 3,97 (tidak masuk range IPK, nilai maksimal IPK iterasi 7 adalah 3,95). Karena jumlah data yang tidak terklasifikasikan pada node ini hanya 1, maka data ini dianggap data yang tidak tertangani, dari pada harus memangkas node tersebut[8].

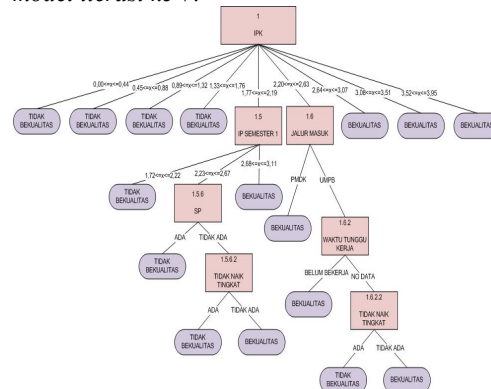
2) Id y107, dengan IPK= 1,91 dan IP semester 1 = 2,80. Sesuai *rule*, seharusnya menghasilkan *class label* = "BERKUALITAS", namun *class label* data ini adalah = "TIDAK BERKUALITAS", sehingga data ini termasuk data *error*.

Ada beberapa syarat yang dipakai dalam pemilihan model *decision tree*[8]:

- Model dengan jumlah aturan (*rule*) paling banyak.
- Model dengan akurasi tinggi. Iterasi 1-6 dan 8-10, nilai *error rate* adalah 0%.
- Model yang mencakup semua kelas target yang mungkin muncul dalam *test set*. Semua *rule* pada iterasi 1-10 sudah mencakup *class label*.

Berdasarkan syarat-syarat pemilihan model *decision tree*, setiap iterasi masing-masing memenuhi dua dari tiga syarat di atas, yaitu syarat 1,3 dan 2,3, sehingga pemilihan *decision tree* berdasarkan:

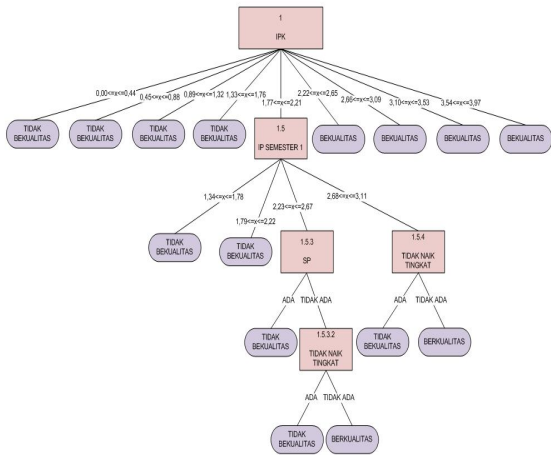
1) Model yang memiliki *rule* terbanyak dan mencakup keseluruhan kelas target *data testing* (syarat 1 dan 3), yaitu model iterasi ke-7.



Gbr 4. Decision Tree Iterasi 7

Gbr. 4 menunjukkan bahwa jalur masuk kuliah memberikan pengaruh dalam menentukan tingkat kualitas mahasiswa, yaitu pada data *range* IPK 2,20 - 2,63.

2) Model yang memiliki akurasi tertinggi dan mencakup keseluruhan kelas target (syarat 2 dan 3). Dalam hal ini, model *decision tree* iterasi 1-6, dan 8-10 memiliki kedua syarat ini, sehingga dapat dipilih salah satu di antaranya. Namun, setelah divisualisasikan menggunakan aplikasi Weka, *decision tree* yang dipilih adalah *decision tree* iterasi ke-4.



Gbr 5. *Decision Tree* Iterasi 4

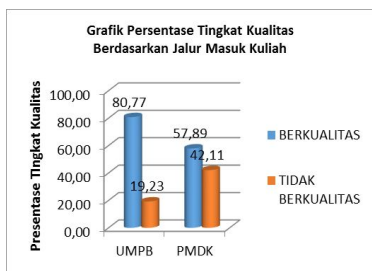
Sehingga, model *decision tree* iterasi 4 dan 7 adalah hasil akhir dari proses klasifikasi tingkat kualitas mahasiswa Politeknik Negeri Batam program studi Teknik Informatika.

C. Penerapan Rule *Decision Tree* Iterasi 4 dan 7 pada Keseluruhan Data

Rule yang dipilih akan diterapkan pada keseluruhan data untuk memperoleh informasi tingkat kualitas mahasiswa setiap jalur masuk kuliah.

1) Penerapan Rule Iterasi 4

Jumlah keseluruhan data yang tertangani sesuai dengan rule iterasi 4 adalah sebanyak 331 data. Grafik persentase mahasiswa yang berkualitas dan tidak berkualitas untuk masing-masing jalur masuk disajikan pada Gbr.3.



Gbr 6. Grafik Tingkat Kualitas Berdasarkan Rule Iterasi 4

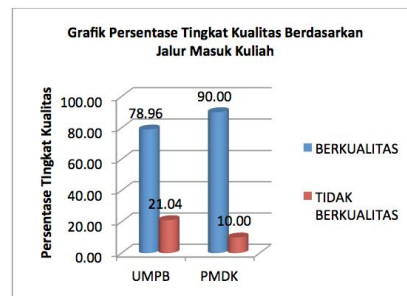
Gbr.6 memberikan informasi bahwa data mahasiswa dengan jalur masuk UMPB memiliki presentase lebih besar dalam menghasilkan mahasiswa berkualitas dibandingkan melalui jalur masuk PMDK.

2) Penerapan Rule Iterasi 7

TABEL X. JUMLAH DATA KLASIFIKASI

JUMLAH DATA	TIDAK TERTANGANI	ERROR	DATA BERSIH UNTUK PERSENTASE
331	1	1	329

Tabel XIII menunjukkan jumlah data pada iterasi 7. Jumlah keseluruhan data yang tertangani sesuai dengan rule iterasi 7 adalah sebanyak 329 data. Grafik persentase mahasiswa yang berkualitas dan tidak berkualitas untuk masing-masing jalur masuk disajikan pada Gbr.7.



Gbr 7. Grafik Tingkat Kualitas Berdasarkan Rule Iterasi 7

Dari grafik di atas diperoleh informasi:

1) Data mahasiswa dengan jalur masuk PMDK memiliki persentase lebih besar dalam menghasilkan mahasiswa yang berkualitas dibandingkan dengan jalur masuk UMPB. Sebaliknya, mahasiswa dengan jalur masuk UMPB memiliki persentase yang lebih lebih besar dalam menghasilkan mahasiswa yang tidak berkualitas dibandingkan jalur masuk PMDK.

2) Persentase mahasiswa yang berkualitas dari data yang diperoleh tergolong besar, hal ini dapat dilihat dari persentase mahasiswa tidak berkualitas yang tidak mencapai setengah dari persentase mahasiswa berkualitas.

V. KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa:

1) Proses *learning* dan *classification* menghasilkan 2 model *decision tree* yang memenuhi persyaratan, yaitu iterasi 4 dan 7.

2) Berdasarkan model *decision tree* iterasi ke 7, PMDK memiliki persentase yang lebih besar dalam menghasilkan mahasiswa yang berkualitas yaitu sebesar 90%, sedangkan melalui UMPB sebesar 78,96%. Hal ini sesuai dengan hasil angket bahwa 7 dari 13 responden memilih lajur masuk PMDK yang menghasilkan mahasiswa berkualitas.

3) Berdasarkan model *decision tree* iterasi ke 4, PMDK memiliki persentase yang lebih kecil dalam menghasilkan mahasiswa yang berkualitas yaitu sebesar 57,89%,

sedangkan melalui UMPB sebesar 80,77%. Hal ini bertolak belakang dengan hasil klasifikasi menggunakan *rule* iterasi 7.

4) Jalur masuk kuliah memiliki pengaruh dalam mengklasifikasikan tingkat kualitas mahasiswa, namun hanya untuk penggunaan *rule* iterasi 7.

5) Berdasarkan *decision tree* iterasi 4 dan 7 terlihat bahwa jalur masuk PMDK memiliki presentasi yang lebih besar dalam menghasilkan mahasiswa yang berkualitas yaitu 90% pada iterasi 7 dibandingkan dengan UMPB pada iterasi 4 sebesar 80,77%.

ACKNOWLEDGE

Terima kasih kepada Mir'atul K. Mufida dan Arsita Agniesari yang telah berpartisipasi pada penelitian ini.

REFERENSI

- [1] Pos Kupang, "Berkualitaskah Lulusan SNMPTN?", (2010, Juni 18). Tersedia: <http://kupang.tribunnews.com/2010/06/18/berkualitaskah-lulusan-snmptn->, diakses pada 6 Mei 2014.

- [2] Ridwan Mujib, Hadi Suyono, dan M. Sarosa. "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Jurnal EECCIS Vol.7*, No. 1(Juni, 2013), hal. 59.
- [3] Departemen Pendidikan Nasional. 2005. *Kamus Besar Bahasa Indonesia*. Jakarta: Balai Pustaka.
- [4] Larose, Daniel T. 2005. *Discovering Knowledge in Data: An Introduction to Data mining*. John Willey & Sons, Inc.
- [5] Han, J. dan Kamber M. 2001. *Data mining Concept and Techniques Chapter 1*. San Fransisco: Morgan Kaufman Publisher.
- [6] Turban, E., dkk. 2005. *Decicion Support System and Intelegant Systems*. Yogyakarta: Andi Offset.
- [7] Kusriani dan Taufiq, Emha. 2009. *Algoritma Data mining*. Yogyakarta: Andi Offset.
- [8] Romansyah Firat. 2009. *Penerapan Teknik Klasifikasi Menggunakan Metode Fuzzy Decision Tree dengan Algoritma Id3 pada Data Diabetes*. *Internetworking Indonesia Journal Vol.1*, No.2 (2009), hal. 45.
- [9] Fu L. 1994. *Neural Network in Computer Science*. Singapura: McGrawHill.
- [10] Basuki Achmad. 2005. *Pemrograman Access File Text dengan Visual Basic*. Surabaya: PENS-ITS.