

ANALISA KLASIFIKASI KUALITAS MAHASISWA LULUSAN BERDASARKAN JALUR PENERIMAAN MENGGUNAKAN ALGORITMA C4.5 (STUDI KASUS: UNIVERSITAS LAMPUNG)

Resalina Oktaria¹, Muhamad Komarudin², Mona Arif Muda³

^{1,2,3} Teknik Informatika, Fakultas Teknik
Universitas Lampung
resalina.oktaria1036@students.unila.ac.id, m.komarudin@eng.unila.ac.id,
mona.batubara@eng.unila.ac.id

ABSTRACT

Universities need a quality evaluation process based on the standards of the National Accreditation Agency for Higher Education (BAN-PT) every year. Therefore, it is necessary for universities knowing the students to evaluate and to maintain the University's Education Efficiency Number (AEE). One of the standards that has been determined by BAN-PT is the quality of students that can be seen from the GPA, the accuracy of completing studies, thesis, path admission, and others. The purpose of this study is to provide information about the quality of students based on SNMPTN and SBMPTN admission with data mining techniques using RapidMiner software in the application of the C4.5 algorithm and using the research method of the Cross-Industry Standard Process for Data Mining (CRIPS-DM). The results of this study were students who has graduated of class I quality was 46% from the SBMPTN path admission, 28% of the SNMPTN path admission, and class II category was 11% from the SBMPTN path admission, and 15% from the SNMPTN path admission. The results of accuracy obtained in decision tree modeling got an accuracy value of 97.46% with an error value of 0.98% and the value of Area Under Curve (AUC) of 0.973 with an error value of 0.014 which is classified into a excellent classification.

Keywords: *Data Mining, C4.5 Algorithm, CRISP-DM, Student Quality, Path Admission*

ABSTRAK

Perguruan tinggi tiap tahunnya memerlukan proses evaluasi mutu berdasarkan standar Badan Akreditasi Nasional-Perguruan Tinggi (BAN-PT). Sehingga perguruan tinggi harus mengetahui kondisi mahasiswanya guna mengevaluasi dan mempertahankan Angka Edukasi Efisiensi (AEE) perguruan tinggi tersebut. Salah satu standar yang telah ditetapkan oleh BAN-PT adalah kualitas mahasiswa lulusan yang dapat dilihat dari IPK, ketepatan menyelesaikan studi, skripsi, jalur penerimaan dan lainnya. Tujuan dari penelitian ini adalah memberikan informasi mengenai klasifikasi kualitas mahasiswa lulusan UNILA berdasarkan jalur penerimaan SNMPTN dan SBMPTN dengan teknik data mining memanfaatkan *software rapidminer* dalam penerapan Algoritma C4.5 dan menggunakan metode penelitian *Cross Industry Standard Process for Data Mining (CRIPS-DM)*. Hasil dari penelitian ini yaitu mahasiswa lulusan dengan kualitas kategori kelas I sebanyak 46% pada jalur penerimaan SBMPTN, 28% jalur penerimaan SNMPTN dan kategori kelas II sebanyak 11% jalur penerimaan SBMPTN dan 15% jalur penerimaan SNMPTN. Hasil pemodelan *decision tree* mendapatkan nilai akurasi sebesar 97,46% dengan nilai *error* sebesar 0,98% serta nilai *Area Under Curve (AUC)* sebesar 0,973 dengan nilai *error* 0,014 yang tergolong ke dalam *excellent classification*.

Kata Kunci: *Data Mining, Algoritma C4.5, CRISP-DM, Kualitas Mahasiswa, Jalur Penerimaan*

DOI: <https://doi.org/10.15408/jti.v12i2.11171>

I. PENDAHULUAN

Perguruan Tinggi tiap tahunnya memerlukan proses evaluasi dan penilaian terhadap mutu perguruan tinggi tersebut. Berdasarkan standar akreditasi yang telah ditetapkan oleh Badan Akreditasi Nasional-Perguruan Tinggi (BAN-PT), salah satu tolak ukur mutu perguruan tinggi yaitu kualitas mahasiswanya, dimana hal tersebut memengaruhi Angka Efisiensi Edukasi (AEE) suatu Perguruan Tinggi. Kualitas mahasiswa dapat dilihat dari Indeks Prestasi Kumulatif (IPK) yang didapat, kemampuan menyelesaikan studi dengan tepat waktu, jalur masuk, masa tenggat kerja, dan faktor lainnya [1].

Jalur masuk merupakan salah satu tolak ukur kualitas mahasiswa yang dijadikan sebagai sarana menjangkau mahasiswa untuk studi di Perguruan Tinggi. Setiap jalur masuk di Perguruan Tinggi memiliki kualifikasi yang berbeda. Kualifikasi bertujuan untuk melihat kemampuan calon mahasiswa yang akan menjadi mahasiswa di Perguruan Tinggi tersebut. Universitas Lampung memiliki beberapa jalur masuk kuliah seperti Seleksi Nasional Masuk Perguruan Tinggi (SNMPTN), Seleksi Bersama Masuk Perguruan Tinggi (SBMPTN), dan jalur lokal seperti Ujian Mandiri, Advokasi, PMPAP. Jalur masuk yang dapat dikaji yaitu SNMPTN dan SBMPTN yang memiliki keterkaitan dalam proses masuk penerimaan mahasiswa baru.

Perkembangan teknologi saat ini menciptakan kondisi akan kaya data namun minim informasi. Universitas Lampung memiliki data yang sangat besar, salah satunya yaitu data mahasiswa. Sehingga perlukan *data scientist* dalam menganalisa data dalam jumlah besar dalam sistem yang ada di Universitas Lampung sehingga dapat ditemukan pola dan Algoritma guna mendapatkan informasi baru dengan memanfaatkan Teknik *data mining* dan menggunakan *tools* dari sistem *data mining* untuk membantu mengolah dan menganalisis data dalam jumlah yang besar. Salah satu Algoritma dalam *data mining* yang dapat membagi data yang besar menjadi *record* yang lebih kecil yaitu Algoritma C4.5. Penelitian terkait yaitu "Penerapan Algoritma C4.5 untuk Klasifikasi Predikat Kelulusan Mahasiswa Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta" oleh Yusuf Sulisty Nugroho pada tahun 2014. Pada penelitian tersebut menggunakan persamaan

Slovin untuk menentukan jumlah data sampel yang digunakan. Data sampel hasil persamaan Slovin sebanyak 341 data mahasiswa. Parameter yang digunakan dalam klasifikasi antara lain asal sekolah, jenis kelamin, jumlah satuan kredit semester, memiliki peran sebagai asisten atau tidak. Hasil dari penelitian yaitu atribut peran sebagai asisten memiliki nilai yang sangat berpengaruh pada proses klasifikasi sebesar 73,91% [2]. Pada penelitian ini bertujuan untuk menerapkan Teknik *data mining* dengan metode *decision tree* Algoritma C4.5 dan memanfaatkan *tools data mining RapidMiner 9.1* terhadap data yang ada untuk memperoleh informasi mengenai klasifikasi kualitas mahasiswa Universitas Lampung berdasarkan jalur masuk SNMPTN, dan SBMPTN sesuai dengan parameter kualitas mahasiswa dari BAN-PT. sehingga dapat digunakan sebagai rekomendasi terhadap peninjauan kembali metode jalur masuk penerimaan mahasiswa baru di Universitas Lampung.

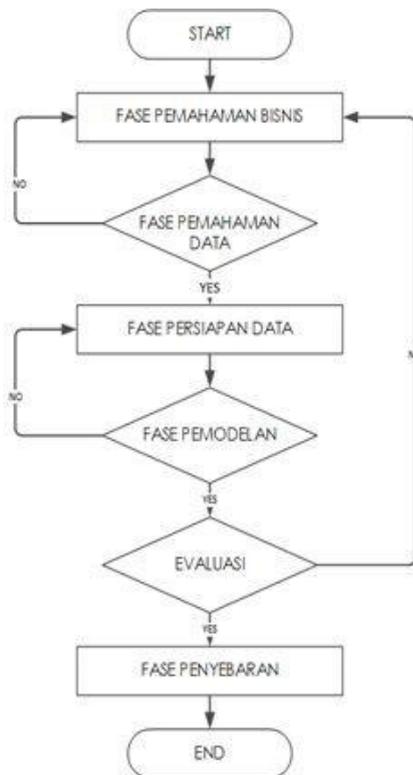
II. METODOLOGI

Alur pada penelitian ini terdapat dua tahapan antara lain studi literatur dan tahapan penelitian menggunakan metode penelitian pada teknik *data mining* yaitu *Cross Industry Standard Process Data Mining (CRISP-DM)* dengan menggunakan teknik *data mining* dan Algoritma C4.5.

Data mining adalah bidang dari beberapa bidang ilmu yang menyatukan teknik dari *mechine learning*, pengenalan pola, statistik, basis data, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari data dalam jumlah besar. *Data mining* merupakan proses mengekstraksi dan mengidentifikasi suatu informasi, menemukan hubungan, menghasilkan *pattern* atau pola dengan data dari basis data yang besar [3].

Decision tree atau pohon keputusan merupakan salah satu metode klasifikasi *data mining*. *Decision tree* dapat diekspresikan dalam bentuk Bahasa *database SQL (Structured Query Language)* untuk mencari *record* pada kategori tertentu. *Data mining* digunakan untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel *input* dan variabel terkait. *Decision tree* atau pohon keputusan adalah salah satu metode dari teknik *data mining* yang digunakan untuk mengeksplorasi data dengan membagi data yang besar menjadi *record* yang lebih kecil dengan

memperhatikan variabelnya [4]. Algoritma dari model *decision tree* antara lain *Iterative Dichotomiser 3* atau *Induction of Decision 3 (ID3)*, *CART*, dan Algoritma *C4.5*.



Gambar 1. Flowchart penelitian

2.1 Fase Pemahaman Bisnis (Business Understanding Phase)

- Menentukan tujuan bisnis

Pada fase ini bertujuan untuk merancang model *data mining* yang dapat digunakan untuk mengklasifikasi tingkat kualitas mahasiswa sesuai parameter BAN-PT (Badan Akreditasi Negeri Perguruan Tinggi) yaitu lama studi, lama skripsi, IPK (Indeks Prestasi Kumulatif), dan jalur masuk (SNMPTN dan SBMPTN). Hasil klasifikasi ini diharapkan dapat dijadikan saran kepada BPPMB dalam membuat keputusan atau kebijakan untuk meningkatkan kualitas mahasiswa melalui jalur penerimaan mahasiswa baru Universitas Lampung.

- Menilai situasi

Salah satu tantangan perguruan tinggi adalah memperbaiki kualitas mahasiswanya agar mudah diserap dalam dunia kerja. Selain itu, kualitas mahasiswa pada suatu perguruan tinggi dapat mempengaruhi Angka Efisiensi Edukasi (AEE). Nilai AEE yang rendah berpengaruh dalam penilaian akreditasi suatu perguruan tinggi.

- Menentukan tujuan *data mining*

Model *data mining* yang dibuat bertujuan untuk mengklasifikasi bagaimana tingkatan kualitas mahasiswa berdasarkan jalur masuk kuliah dan data akademik dari alumni. Pengetahuan ini dapat digunakan oleh manajemen untuk mengambil keputusan atau membuat kebijakan untuk meningkatkan kualitas mahasiswa di Universitas Lampung.

2.2 Fase Pemahaman Data (Data Understanding Phase):

Mengakses Sistem Forum Laporan Pendidikan Tinggi (Forlap DIKTI) guna mendapatkan data alumni Universitas Lampung meliputi Indeks Prestasi Kumulatif (IPK), waktu studi, lama pengerjaan skripsi tahun 2013 hingga 2017 dan data jalur masuk dari basis data Penerimaan Mahasiswa Baru (PMB) Universitas Lampung yang didapatkan dari Badan Pengelola Penerimaan Mahasiswa Baru (BPPMB) Universitas Lampung.

- Mendeskripsikan data

Deskripsi dari atribut data yang akan digunakan dalam membangun model *data mining*.

Tabel 1. Deskripsi data

Atribut	Deskripsi
IPK	Indeks Prestasi Kumulatif adalah ukuran kemampuan mahasiswa sampai pada waktu tertentu yang dapat dihitung berdasarkan jumlah SKS mata kuliah yang diambil sampai pada periode tertentu dikalikan dengan nilai bobot masing masing mata kuliah dibagi dengan jumlah seluruh SKS.
Lama studi	Merupakan lama tembus studi dihitung mulai terdaftar sebagai mahasiswa sampai dinyatakan lulus.
Lama skripsi	Merupakan lama waktu pengerjaan skripsi
Jalur masuk	Merupakan proses masuk perguruan tinggi yang diikuti mahasiswa bersangkutan.

2.3 Fase Persiapan Data (*Data Preparation Phase*)

Pada tahapan persiapan data, ada beberapa hal yang dilakukan antara lain, deskripsi *dataset*, memilih data, membangun data, mengintegrasikan data dan membersihkan data dari 27613 *record* data mahasiswa.

2.4 Fase Pemodelan (*modelling phase*)

Pada penelitian yang menggunakan Algoritma C4.5 menggunakan *software RapidMiner*. Algoritma C4.5 merupakan pengembangan dari Algoritma ID3. Secara umum untuk membangun Algoritma C.45 [2], yaitu:

1. Pilih atribut sebagai akar (*root*).
2. Buat cabang setiap nilai.
3. Membagi kasus dalam cabang

Mengulangi proses untuk setiap cabang hingga semua kasus pada cabang memiliki kelas (*class*) yang sama. Persamaan (1) digunakan untuk menghitung *gain* dalam memilih atribut yang akan menjadi *root*.

$$Gain(S, A) = Entropy(S) - \sum_{i=0}^n p_i \times Entropy(S_i) \quad (1)$$

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*:

$$Entropy(S) = \sum_{i=0}^n p_i \times \log_2 p_i \quad (2)$$

Keterangan:

S : himpunan kasus

A : fitur

n : jumlah partisi S

p_i : proposisi dari S_i terhadap S

Decision tree dihasilkan setelah menghitung nilai *entropy* dan *gain*. Kemudian *decision tree* yang didapatkan diuji untuk mengetahui nilai keakuratan dari *decision tree*.

2.5 Evaluasi (*evaluation*)

Pada tahap ini dilakukan evaluasi terhadap keefektifan dan keakuratan dari model yang terbentuk sebelum digunakan dan menentukan

apakah model dapat mencapai tujuan yang ditetapkan pada fase pemahaman bisnis.

Pada fase evaluasi dilakukan menggunakan metode *Cross validation* merupakan metode untuk memvalidasi keakuratan sebuah model yang dibangun berdasarkan sebuah *dataset* tertentu. Data yang digunakan dalam membangun proses model yaitu *data training* dan *data testing*. Sedangkan data yang digunakan dalam melakukan validasi adalah *dataset*. Bentuk dari *cross validation* adalah *K-Fold cross validation*. Dalam teknik ini dikenal dengan melipat data sebanyak K dan mengulangi atau meniterasi eksperimen sebanyak K juga [3].

2.6 Deployment

Pada tahap ini, pengetahuan atau informasi yang telah diperoleh akan dipresentasikan dalam bentuk laporan sehingga dapat diketahui atau dibaca oleh orang awam. Hasil dari penelitian ini berupa analisa tingkat kualitas mahasiswa di Universitas Lampung yang diharapkan dapat digunakan oleh institusi perguruan tinggi sebagai bahan pertimbangan dalam menentukan langkah guna mengatasi permasalahan penerimaan mahasiswa berdasarkan jalur masuk

III. HASIL DAN PEMBAHASAN

3.1 Persiapan Data (*Preprocessing Data*)

Pada tahap ini dilakukan persiapan data yang akan digunakan dalam penelitian dengan menghilangkan data yang *noise* atau ganjil dari 27613 *record* menjadi 3302 *record*. selanjutnya melakukan transformasi data ke dalam format klasifikasi *data mining* dengan ketentuan standar yang telah ditetapkan oleh BAN-PT. Data yang digunakan adalah IPK, lama studi, lama skripsi, jalur masuk penerimaan (SNMPTN dan SBMPTN).

Tabel 2. *Class data*

Parameter Pengklasasan Data

IPK	> 3.0	Sangat baik
	2.75 – 3.0	Baik
Lama studi	> 4 tahun	Lambat
	≤ 4 tahun	Tepat
Lama skripsi	> 6 bulan	Lambat
	≤ 6 bulan	Tepat

Selanjutnya data dilakukan pengkelasan atau pembuatan label sehingga mempermudah pembacaan hasil dari klasifikasi. Pengkategorian kelas yang dilakukan berdasarkan nilai pedoman dari instrumentasi penilaian BAN-PT [4]. Berikut kategori kelas data yang dilakukan:

Tabel 3. Kategori kelas data

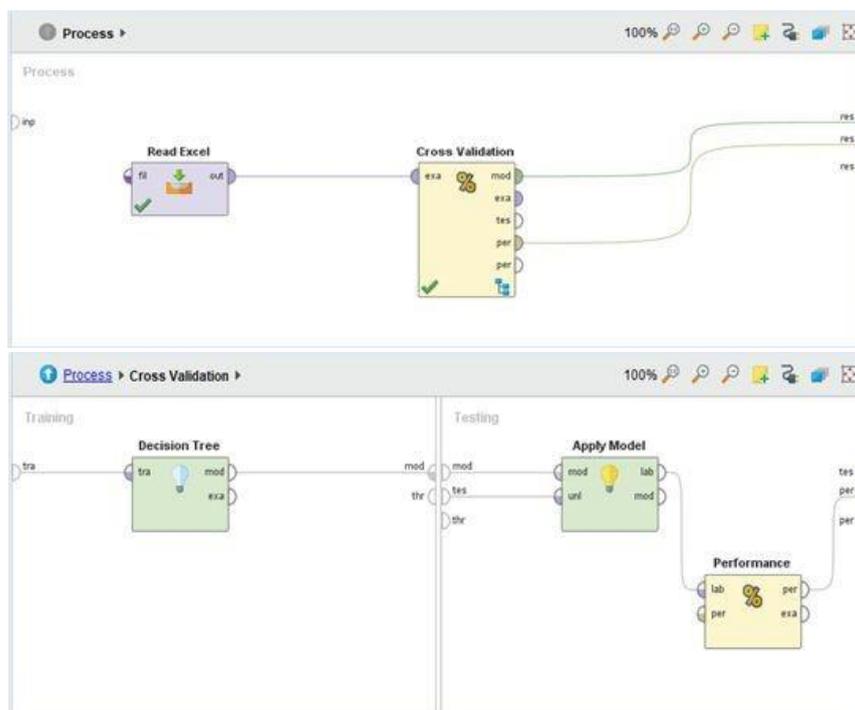
Pengkelasan Data I	
IPK	> 3.0
Lama studi	< 4 tahun
Lama skripsi	< 6 bulan
Pengkelasan Data II	
IPK	< 3.0
Lama studi	> 4 tahun
Lama skripsi	> 6 bulan

Terdapat nilai parameter penilaian yang dilakukan ketika IPK lebih dari 3 maka mendapatkan nilai 4 sedangkan ketika IPK kurang dari 3 maka nilai yang didapatkan adalah 3. Parameter nilai lama menyelesaikan studi ketika studi selesai kurang dari 4 tahun maka mendapatkan nilai 4, namun ketika waktu yang ditempuh lebih dari 4 tahun maka nilai yang didapatkan adalah 3. Kemudian pada

kemampuan menyelesaikan skripsi, ketika waktu yang ditempuh kurang dari 6 bulan maka mendapatkan nilai 4, sedangkan ketika waktu pengerjaan skripsi lebih dari 6 bulan maka nilai yang didapatkan 3. Maka dibuat dua kategori guna melihat kualitas mahasiswa lulusan yang ada di Universitas Lampung yaitu kategori kelas I dan kelas II. Pada kategori kelas I maka mahasiswa lulusan yang termasuk dengan kualifikasi dengan nilai IPK lebih dari 3 sehingga memiliki nilai 4, lama menempuh studi kurang dari 4 tahun dengan nilai 4 dan lama pengerjaan skripsi kurang dari 6 bulan sehingga memiliki nilai 4. Sedangkan mahasiswa yang memiliki IPK, lama studi dan lama skripsi dengan nilai 3 termasuk ke dalam kategori mahasiswa lulusan kelas II.

3.2 Pemodelan (Modelling)

Setelah dilakukan fase persiapan data selanjutnya dilakukan fase pemodelan. Pemodelan klasifikasi pada penelitian ini menggunakan teknik *decision tree* dan Algoritma C4.5 dengan *tools* yang terdapat pada *software RapidMiner 9.1*. *RapidMiner* sendiri digunakan para data *scientist* untuk menganalisa terhadap *data mining* dan *text mining*.

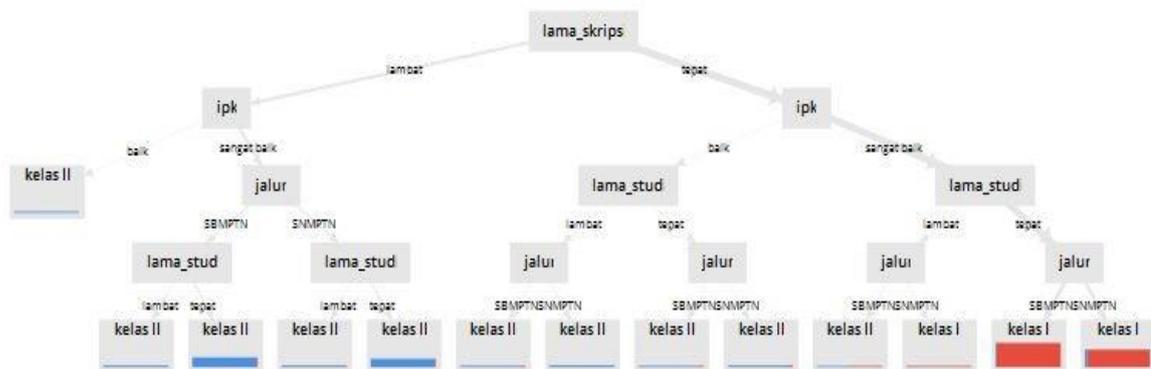


Gambar 2. Rancangan proses klasifikasi

Tabel 4. Perhitungan entropy dan gain

NODE	ATRIBUT	NILAI	SUM	KELAS I	KELAS II	ENTROPHY	GAIN
1	Total		3302	2191	1111	0,921	
	lama_skripsi						0,614
		tepat	2355	2177	178	0,386	
		lambat	947	14	933	0,111	
	lama_studi						0,055
		tepat	3171	2187	984	0,893	
		lambat	131	4	127	0,197	
	ipk						0,054
		sangat baik	3173	2187	986	0,894	
		baik	129	4	125	0,199	
	jalur_masuk						0,005
		SNMPTN	1426	879	547	0,96	
		SBMPTN	1876	1312	564	0,882	

Pemodelan *decision tree* menggunakan RapidMiner menghasilkan *decision tree* seperti di bawah ini:



Gambar 3. Decision Tree C4.5

Pada Gambar 3 merupakan pemodelan klasifikasi terhadap data mahasiswa lulusan pada tahun 2016 dan 2017 berdasarkan jalur penerimaan SNMPTN dan SBMPTN menggunakan *RapidMiner*. Pemodelan yang di dapatkan menunjukkan bahwa data mahasiswa pada kelas I dengan *rule* lama skripsi tepat. IPK sangat baik, lama studi tepat dengan jalur masuk SBMPTN dan SNMPTN memiliki jumlah terbanyak. Setelah didapatkan hasil menggunakan data keseluruhan maka data dibagi lagi menjadi per fakultas dan per program studi yang ada di Universitas Lampung.

```

Tree
lama_skripsi = lambat
| ipk = baik: kelas II (kelas II=11, kelas I=0)
| ipk = sangat baik
| | jalur = SBMPTN
| | | lama_studi = lambat: kelas II (kelas II=28, kelas I=0)
| | | lama_studi = tepat: kelas II (kelas II=478, kelas I=6)
| | | jalur = SNMPTN
| | | | lama_studi = lambat: kelas II (kelas II=26, kelas I=1)
| | | | lama_studi = tepat: kelas II (kelas II=390, kelas I=7)
lama_skripsi = tepat
| ipk = baik
| | lama_studi = lambat
| | | jalur = SBMPTN: kelas II (kelas II=13, kelas I=1)
| | | jalur = SNMPTN: kelas II (kelas II=53, kelas I=0)
| | lama_studi = tepat
| | | jalur = SBMPTN: kelas II (kelas II=16, kelas I=1)
| | | jalur = SNMPTN: kelas II (kelas II=32, kelas I=2)
| ipk = sangat baik
| | lama_studi = lambat
| | | jalur = SBMPTN: kelas II (kelas II=1, kelas I=1)
| | | jalur = SNMPTN: kelas I (kelas II=0, kelas I=1)
| | lama_studi = tepat
| | | jalur = SBMPTN: kelas I (kelas II=22, kelas I=1303)
| | | jalur = SNMPTN: kelas I (kelas II=41, kelas I=868)
    
```

Gambar 4. Rule atau aturan yang diperoleh

Pada Gambar 4 menunjukkan *rule* atau aturan yang diperoleh dari pemodelan menggunakan Algoritma C4.5. Adapun

aturan-aturan yang diperoleh dari *decision tree* untuk menentukan kualitas mahasiswa. Adapun aturan-aturan yang diperoleh dari *decision tree* untuk menentukan kualitas mahasiswa berdasarkan jalur masuk penerimaan, yaitu:

1. Jika lama skripsi = lambat dan IPK = baik maka kelas II
2. Jika lama skripsi = lambat dan IPK = sangat baik dan jalur = SBMPTN dan lama studi = lambat maka kelas II
3. Jika lama skripsi = lambat dan IPK = sangat baik dan jalur = SBMPTN dan lama studi = tepat maka kelas II
4. Jika lama skripsi = lambat dan IPK = sangat baik dan jalur = SNMPTN dan lama studi = lambat maka kelas II
5. Jika lama skripsi = lambat dan IPK = sangat baik dan jalur = SNMPTN dan lama studi = tepat maka kelas II
6. Jika lama skripsi = tepat dan IPK = baik dan jalur = SBMPTN dan lama studi = lambat maka kelas II
7. Jika lama skripsi = tepat dan IPK = baik dan jalur = SNMPTN dan lama studi = tepat maka kelas II

8. Jika lama skripsi = tepat dan IPK = baik dan jalur = SBMPTN dan lama studi = lambat maka kelas II
9. Jika lama skripsi = tepat dan IPK = baik dan jalur = SNMPTN dan lama studi = tepat maka kelas II
10. Jika lama skripsi = tepat dan IPK = sangat baik dan jalur = SBMPTN dan lama studi = lambat maka kelas II
11. Jika lama skripsi = tepat dan IPK = sangat baik dan jalur = SNMPTN dan lama studi = lambat maka kelas I
12. Jika lama skripsi = tepat dan IPK = sangat baik dan jalur = SBMPTN dan lama studi = tepat maka kelas I
13. Jika lama skripsi = tepat dan IPK = sangat baik dan jalur = SNMPTN dan lama studi = tepat maka kelas I

3.3 Fase Evaluasi (*Evolution Phase*)

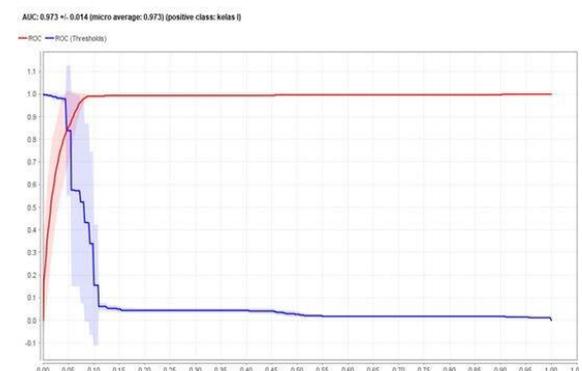
Pada tahap ini dilakukan evaluasi terhadap keefektifan dan kualitas model yang sudah terbentuk sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang ditetapkan pada fase pemahaman bisnis. Kemudian dievaluasi menggunakan metode *k-fold cross validation* dan curva ROC.

accuracy: 97.46% +/- 0.98% (micro average: 97.46%)

	true kelas II	true kelas I	class precision
pred. kelas II	1047	20	98.13%
pred. kelas I	64	2171	97.14%
class recall	94.24%	99.09%	

Gambar 5. Akurasi model *decision tree* algoritma C4.5

Akurasi pemodelan mendapatkan nilai 97,46% dengan nilai *error rate* sebesar 0.98%. Nilai persentase akurasi yang didapatkan dengan metode *cross validation* dengan cara memilih beberapa atribut untuk melakukan validasi silang yang memaksimalkan akurasi dari keseluruhan pohon. Akurasi model yang didapatkan tergolong ke dalam *excellent classification*.



Gambar 6. Grafik ROC (*Receiver Operating Characteristic*)

Pada Gambar 6 menunjukkan *Area Under Curve* (AUC) yang didapatkan pada pemodelan. Nilai AUC yang didapatkan sebesar 0.973 tergolong ke dalam *excellent classification* sehingga dapat digunakan sebagai acuan prediksi yang baik.

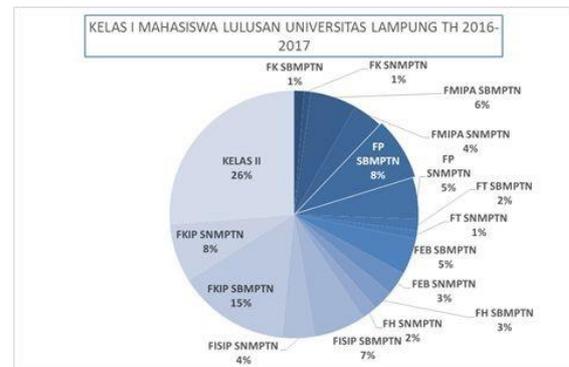
3.4 Deployment

Pada tahap terakhir dalam metode CRISP-DM dilakukan pembuatan laporan hasil pemodelan menggunakan teknik *data mining* dengan Algoritma C4.5 berupa analisa pola *decision tree* yang didapatkan. Data hasil pemodelan dianalisa pola kualitas mahasiswa lulusan Universitas Lampung dari tiap Fakultas dan Program Studi. Setelah mendapat wawasan dari data tersebut dapat dijadikan sebagai saran kepada BPPMB dalam meningkatkan kualitas mahasiswa baru Universitas Lampung berdasarkan jalur penerimaan SNMPTN dan SBMPTN pada tahun penerimaan yang akan datang.



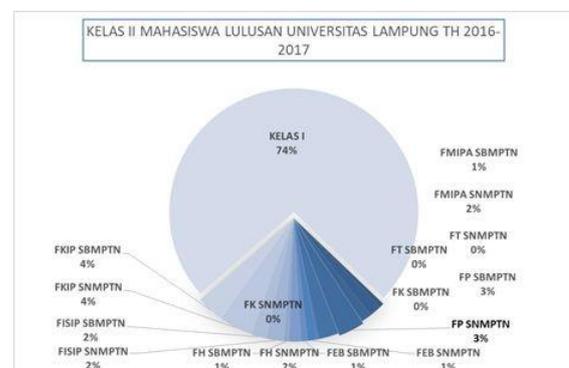
Gambar 7. Persentase mahasiswa lulusan jalur masuk SNMPTN dan SBMPTN Universitas Lampung tahun 2016-2017

Persentase mahasiswa lulusan Universitas Lampung, dengan jalur SBMPTN pada kelas I sebesar 46% sedangkan pada kelas II sebesar 11%. Pada jalur masuk SNMPTN memiliki persentase sebesar 28% pada kelas I dan 15% pada kelas II. Mahasiswa lulusan dengan jalur masuk SBMPTN memiliki jumlah persentase yang lebih besar dibandingkan dengan mahasiswa lulusan dengan jalur masuk SNMPTN.



Gambar 8. Persentase kelas I per Fakultas Universitas Lampung

Pada Gambar 8 di atas menunjukkan perbandingan banyaknya mahasiswa lulusan tahun 2013 dan 2014 kelas I pada jalur masuk SNMPTN dan SBMPTN. Pada kelas I mahasiswa lulusan pada Fakultas Keguruan dan Ilmu Pendidikan dengan jalur SBMPTN sebanyak 15% dan jalur SNMPTN sebanyak 8%.



Gambar 9. Persentase kelas II per Fakultas Universitas Lampung

Pada Gambar 9 di atas menunjukkan perbandingan banyaknya mahasiswa lulusan tahun 2013 dan 2014 kelas I pada jalur masuk SNMPTN dan SBMPTN. Pada kelas II mahasiswa lulusan pada Fakultas Keguruan dan Ilmu Pendidikan dengan jalur SBMPTN sebanyak 4% dan jalur SNMPTN sebanyak 4%.



Gambar 10. Persentase mahasiswa lulusan jalur masuk SNMPTN dan SBMPTN Universitas Lampung tahun 2016

Persentase mahasiswa lulusan Universitas Lampung tahun 2016, dengan jalur SBMPTN pada kelas I sebesar 36%, sedangkan pada kelas II sebesar 20%. Pada jalur masuk SNMPTN memiliki persentase sebesar 28% pada kelas I dan 16% pada kelas II. Data mahasiswa lulusan yang digunakan dalam pemodelan sebanyak 1207 record. Mahasiswa lulusan dengan jalur masuk SBMPTN pada tahun 2016 memiliki jumlah persentase yang lebih besar dibandingkan dengan mahasiswa lulusan dengan jalur masuk SNMPTN.



Gambar 11. Persentase mahasiswa lulusan jalur masuk SNMPTN dan SBMPTN Universitas Lampung tahun 2017

Persentase mahasiswa lulusan Universitas Lampung tahun 2017, dengan jalur SBMPTN pada kelas I sebesar 42% sedangkan pada kelas II sebesar 15%. Pada jalur masuk SNMPTN memiliki persentase sebesar 26% pada kelas I dan 17% pada kelas II. Data mahasiswa lulusan yang digunakan dalam pemodelan sebanyak 2095 records. Mahasiswa lulusan dengan jalur masuk SBMPTN memiliki jumlah persentase yang lebih besar dibandingkan dengan

mahasiswa lulusan dengan jalur masuk SNMPTN.

Hasil yang didapat yaitu kualitas dari mahasiswa lulusan yang melalui jalur penerimaan SBMPTN lebih besar persentasenya dibandingkan dengan jalur SNMPTN, hal tersebut diketahui dari banyaknya mahasiswa SBMPTN yang lulus tepat waktu dan memiliki IPK yang lebih besar serta waktu pengerjaan skripsi yang lebih singkat. Oleh karena itu hasil dari penelitian ini dapat dijadikan sebagai rekomendasi dalam peninjauan kembali metode penerimaan mahasiswa baru yang ada di Universitas Lampung.

IV. PENUTUP

Hasil penelitian dapat disimpulkan bahwa mahasiswa lulusan dengan jalur masuk penerimaan SBMPTN lebih unggul kualitasnya dibandingkan dengan mahasiswa lulusan dengan jalur masuk SNMPTN. Hasil pemodelan keputusan menggunakan Algoritma C4.5 didapatkan akurasi sebesar 97,46 % dengan nilai *error rate* sebesar 0,98%. Pemodelan yang didapatkan memiliki nilai AUC sebesar 0,973 dengan nilai *error rate* sebesar 0,014 yang tergolong ke dalam *excellent classification*. Atribut lama skripsi memiliki nilai *gain* tertinggi sebesar 0,614 sehingga menjadi atribut yang berpengaruh besar pada proses klasifikasi. Hasil pemodelan dengan rule klasifikasi terbesar menghasilkan jumlah persentase mahasiswa lulusan dengan kualitas kategori kelas I sebanyak 36% jalur SBMPTN dan 28% jalur SNMPTN pada tahun 2016, sedangkan pada tahun 2017 kategori kelas I sebanyak 42% jalur SBMPTN dan 26% jalur SNMPTN.

DAFTAR PUSTAKA

- [1] E.T.L. Kusriani, Algoritma Data Mining, 1st ed. Yogyakarta: Andi Offset, 2009.
- [2] Y. S. Nugroho, "Penerapan algoritma C4.5 untuk klasifikasi predikat kelulusan mahasiswa fakultas komunikasi dan informatika universitas
- [3] D. S. S. M.S. Sani Susanto Ph.D. Pengantar Data Mining: Menggali Pengetahuan dari Bongkahan Data. Yogyakarta, 2010.

- [4] F. A. Hermawati, *Data Mining*, 1st ed. Surabaya: Penerbit Andi, 2009.
- [5] BAN-PT, “Buku 1 Naskah Akademik Akreditasi Institusi Perguruan Tinggi,” Badan Akreditasi Nas. Perguru. Tinggi, 2011.
- [6] S. A. Zega, “Penggunaan Pohon Keputusan untuk Klasifikasi Tingkat Kualitas Mahasiswa Berdasarkan Jalur Masuk Kuliah,” *Semin. Nas. Apl. Teknol. Inf. Yogyakarta*, pp. 7–13, 2014.
- [7] D. K. R. Sudrajat¹, I. Irianingsih¹, “Analysis of data mining classification by comparison of C4.5 and ID algorithms,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 755, no. 1, 2016.
- [8] G. Casella, S. Fienberg, and I. Olkin, *An Introduction to Statistical Learning with Applications in R*, vol. 102. New York Heidelberg Dordrecht London: Springer, 2006.
- [9] A. P. Fadillah, “Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ),” *J. Tek. Inform. dan Sist. Inf.*, vol. 1, pp. 260–270, 2015.
- [10] C. Shaerer, *the-Crisp-Dm-Model-the-New-Blueprint-for-Data-Mining* Shearer-Colin, vol. 5, no. 4. 2000.
- [11] L. A. Dennis Aprilla C, Donny Aji BAskoro, *Belajar Data Mining dengan RapidMiner*. Jakarta, 2013.
- [12] A. Asroni, B. Masajeng Respati, and S. Riyadi, “Penerapan Algoritma C4.5 untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta,” *Semesta Tek.*, vol. 21, no. 2, pp. 158–165, 2018.
- [13] T. Seberang, D. Menggunakan, and C. Algoritma, “Data Mining Untuk Klasifikasi Penentuan Peminatan Siswa Sma Negeri 2,” *Semin. Nas. APTIKOM*, pp. 28–29, 2016.
- [14] BAN-PT, *Buku 6-Matriks Penilaian Akreditasi Sarjana*. Badan Akreditasi Nasional Perguruan Tinggi, 2011.
- [15] F. Gorunescu, *Data Mining Concepts, Models and Techniques*. Verlag Berlin