

**ANALISA KLASIFIKASI KUALITAS MAHASISWA LULUSAN
BERDASARKAN JALUR PENERIMAAN MENGGUNAKAN
ALGORITMA C 4.5
(STUDI KASUS : UNIVERSITAS LAMPUNG)**

(Skripsi)

**Oleh :
Resalina Oktaria**



**FAKULTAS TEKNIK
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2019**

ABSTRAK

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

OLEH

RESALINA OKTARIA

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. Seleksi Nasional Masuk Perguruan Tinggi (SNMPTN) dan Seleksi Bersama Masuk Perguruan Tinggi (SBMPTN) merupakan bagian dari jalur penerimaan yang ada di Universitas Lampung (UNILA). 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* (CRISP-DM). Data yang digunakan dalam penelitian ini merupakan data mahasiswa lulusan jalur penerimaan SBMPTN dan SNMPTN pada tahun 2013-2017. 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

ABSTRACT

ANALYSIS OF GRADUATES QUALITY QUALIFICATION BASED ON UNIVERSITY PATH ADMISSION USING C4.5 ALGORITHM (CASE STUDY RESEARCH: UNIVERSITY OF LAMPUNG)

BY

RESALINA OKTARIA

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 National Selection for Higher Education (SNMPTN) and Joint Selection for Higher Education (SBMPTN) are part of the path admission at the University of Lampung (UNILA). 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 data used in this study is the data of students with SBMPTN and SNMPTN path admission in 2013 until 2017. 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.

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

Oleh
Resalina Oktaria

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA TEKNIK

Pada

Program Studi Teknik Informatika
Jurusan Teknik Elektro
Fakultas Teknik Universitas Lampung



**FAKULTAS TEKNIK
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2019**

**Judul Skripsi : ANALISA KLASIFIKASI KUALITAS
MAHASISWA LULUSAN BERDASARKAN
JALUR PENERIMAAN MENGGUNAKAN
ALGORITMA C 4.5
(STUDI KASUS : UNIVERSITAS LAMPUNG)**

Nama Mahasiswa : Resalina Oktaria

Nomor Pokok Mahasiswa : 1415061036

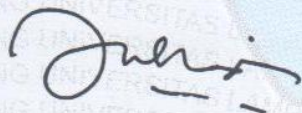
Program Studi : Teknik Informatika

Jurusan : Teknik Elektro

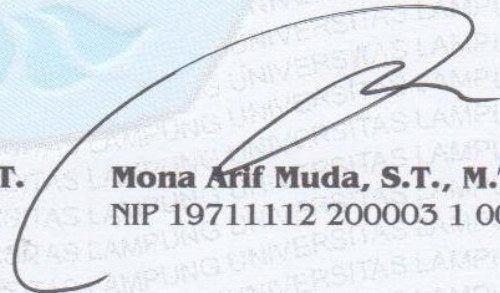
Fakultas : Teknik

MENYETUJUI

1. Komisi Pembimbing



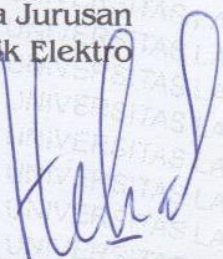
Muhammad Komarudin, S.T., M.T.
NIP 19681207 199703 1 006



Mona Arif Muda, S.T., M.T.
NIP 19711112 200003 1 002

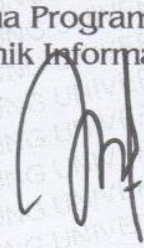
2. Mengetahui

**Ketua Jurusan
Teknik Elektro**



Dr. Herman H. Sinaga, S.T., M.T.
NIP 19711130 199903 1 003

**Ketua Program Studi
Teknik Informatika**



Yessi Mulyani, S.T., M.T.
NIP 19731226 200012 2 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Muhammad Komarudin, S.T., M.T.**

Sekretaris : **Mona Arif Muda, S.T., M.T.**

Penguji

Bukan Pembimbing : **Gigih Forda Nama, S.T., M.T.I.**

2. Dekan Fakultas Teknik



Prof. Suharno, M.Sc., Ph.D.

NIP 19620717 198703 1 002

Tanggal Lulus Ujian Skripsi : **25 Maret 2019**

SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah dilakukan oleh orang lain dan sepanjang sepengetahuan saya tidak terdapat karya atau pendapat yang ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini sebagaimana yang disebutkan dalam daftar pustaka. Selain itu saya menyatakan juga bahwa skripsi ini dibuat oleh saya sendiri.

Apabila pernyataan saya tidak benar, maka saya bersedia dikenai sanksi sesuai dengan hukum yang berlaku

Bandarlampung, 26 Maret 2019

Penulis,



Resalina Oktaria
1415061036

RIWAYAT HIDUP



Penulis dilahirkan di Pringsewu, pada tanggal 25 Oktober 1996, Putri pertama dari dua bersaudara, dari Bapak MHD. Rizali Siregar dan Ibu Yulinawaty S.ST. Pendidikan formal yang pernah ditempuh oleh penulis di saat Sekolah Dasar adalah Sekolah Dasar Fransiskus Pringsewu yang diselesaikan pada tahun 2008. Kemudian meneruskan di Sekolah Menengah Pertama Negeri 1 Pringsewu yang diselesaikan pada tahun 2011. Melanjutkan ke Sekolah Menengah Atas Negeri 1 Gadingrejo dan berhasil tamat di tahun 2014.

Tahun 2014, penulis terdaftar sebagai Mahasiswa Jurusan S1 Program Studi Teknik Informatika, Jurusan Teknik Elektro, Fakultas Teknik, Universitas Lampung. Selama menjadi mahasiswa penulis aktif dalam Organisasi Himpunan Mahasiswa Teknik Elektro (HIMATRO) sebagai Sekretaris Departemen Komunikasi dan Informasi pada tahun 2015-2016 dan Bendahara Umum HIMATRO pada tahun 2016-2017. Penulis melaksanakan Praktek Kerja Lapangan (PKL) di UPT. Teknologi Informasi dan Komunikasi Universitas Lampung pada tahun 2017.

Pada bulan Juli - Agustus 2017 penulis mengaplikasikan ilmu di bidang akademis dengan melaksanakan Kuliah Kerja Nyata (KKN) di Desa Durian, Kecamatan Padang Cermin, Kabupaten Pesawaran, Provinsi Lampung.

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Dengan menyebut nama Allah yang Maha Pengasih lagi Maha Penyayang

SKRIPSI INI KUPERSEMBAHKAN UNTUK

“Ayahanda Alm. MHD. Rizali Siregar dan Ibunda Yulinawaty, terima kasih atas segala doa, pengorbanan, kasih sayang, motivasi dan cinta. Sehingga, dapat mengiringi dalam langkah maupun usaha untuk mencapai keberhasilan.”

“Adikku Kanesia Tahira, yang telah memberikan semangat baik secara motivasi maupun semangat”

“Almamaterku dan Kampus Teknik Universitas Lampung tercinta”

SANWACANA

Bismillahirrahmanirrahim...

Puji syukur ke hadirat Allah SWT, karena atas segala rahmat, hidayah, serta nikmat-Nya, penulis dapat menyelesaikan skripsi ini tepat pada waktunya. Shalawat serta salam tercurah kepada Nabi Muhammad SAW sebagai tauladan umat manusia di dunia.

Skripsi dengan judul **“Analisa Klasifikasi Kualitas Mahasiswa Lulusan Berdasarkan Jalur Masuk Penerimaan Menggunakan Algoritma C4.5 (Studi Kasus : Universitas Lampung)”** disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik pada program Studi Teknik Informatika Jurusan Teknik Elektro Fakultas Teknik Universitas Lampung.

Dalam kesempatan ini penulis mengucapkan terima kasih kepada:

1. Prof. Suharno, M.Sc.,Ph.D. selaku Dekan Fakultas Teknik Universitas Lampung.
2. Dr. Herman H. Sinaga, S.T.,M.T. selaku Ketua Jurusan Teknik Elektro Universitas Lampung.
3. Yessi Mulyani, S.T.,M.T. selaku Kepala Program Studi Teknik Informatika Universitas Lampung.

4. M. Komarudin, S.T.,M.T. selaku Pembimbing Utama, yang telah bersedia meluangkan waktu untuk memberikan pengarahan dan bimbingan dalam mengerjakan penelitian hingga selesai.
5. Mona Arif Muda, S.T.,M.T. selaku Pembimbing Kedua, yang membantu dan memberikan saran dalam mengerjakan penelitian hingga selesai.
6. Gigih Forda Nama, S.T.,M.T.I. selaku Penguji Utama, yang telah membantu sehingga membuat skripsi ini menjadi lebih baik dengan bimbingan dan masukan yang diberikan.
7. Yessi Mulyani, S.T.,M.T. selaku Dosen Pembimbing Akademik yang memberikan bimbingan selama menempuh kuliah di Program Studi Teknik Informatika.
8. Seluruh Dosen Program Studi Teknik Informatika yang telah membagikan ilmunya kepada penulis.
9. Mbak Rika Asliana yang telah membantu penulis dalam hal administrasi di Program Studi Teknik Informatika.
10. Kepada oma dan opa yang tiada henti mendoakan keberhasilan cucunya.
11. Kepada sahabat Mevita Yollanda S.E., Onny Setia Yulianti S.T., Rias Ismawati, Amd.Keb, Nala Maulina, dan May Shinta Amalia S.,Amd.An. untuk semua waktunya yang tidak pernah bosan menemani dan memberikan dukungan kepada penulis selama sembilan tahun terakhir.
12. Kepada para penghuni grup Minion, Restu Pratiwi, Dessy Maya Sary dan Wulan Rahma Izzati, S.T. yang telah memberikan dukungan semangat kepada penulis dan selalu mau direpotkan.

13. Kepada Asep Abdullah Muhammad S.T, Desi Prima S. (ebot), Arif Fauzi, I Nyoman Adi Yudana, Fedra Arya P, Sonny K., Yeni Apriyana S.T, Ginanjar S.T. yang mendukung dan memberikan semangat agar segera menyelesaikan skripsi .
14. Teman seperjuangan seluruh mahasiswa Teknik Informatika dan Teknik Elektro yang telah membantu dalam menyelesaikan masa kuliah. Serta pihak-pihak lain yang tak dapat penulis sebutkan satu per satu pada kesempatan kali ini.

Akhir kata, Penulis meminta maaf karena menyadari masih ada kekurangan pada tugas akhir yang dikerjakan ini, untuk itu kritik dan saran yang membangun masih diperlukan demi kemajuan di masa depan. Semoga Allah membalas kebaikan semua pihak yang telah membantu penulis dalam menyelesaikan Tugas Akhir ini.

Bandarlampung, 07 Maret 2019
Penulis,

Resalina Oktaria

DAFTAR ISI

	Halaman
DAFTAR GAMBAR	xv
DAFTAR TABEL	xviii
DAFTAR SINGKATAN.....	xix
I. PENDAHULUAN	
A. Latar Belakang	1
B. Tujuan Penelitian.....	3
C. Manfaat Penelitian.....	3
D. Rumusan Masalah	3
E. Batasan Masalah.....	4
F. Hipotesis.....	4
G. Sistematika Penelitian	5
II. TINJAUAN PUSTAKA	
A. <i>Data Mining</i>	6
B. Klasifikasi.....	8
C. Teknik <i>Decision Tree</i> Menggunakan Algoritma C4.5 dalam Metode Klasifikasi.....	9
D. <i>K-Fold Cross Validation</i>	11
E. CRISP-DM	12
F. <i>RapidMiner</i>	14
G. Penelitian Terdahulu	15
III. METODOLOGI PENELITIAN	
A. Waktu dan Tempat	17
B. Alat dan Bahan	18
C. Tahapan Penelitian	18
1. Studi Literatur.....	18
2. Tahapan Penelitian dengan Metode CRISP-DM.....	19
IV. HASIL DAN PEMBAHASAN	
A. Fase Persiapan Data	24
B. Fase Pemodelan.....	34
C. Fase Evaluasi.....	44
D. Fase <i>Deployment</i>	46
1. Klasifikasi Mahasiswa Lulusan Universitas Lampung	47

2. Klasifikasi Mahasiswa Lulusan Fakultas Kedokteran.....	53
3. Klasifikasi Mahasiswa Lulusan Fakultas Matematika dan Ilmu Pengetahuan Alam.....	55
4. Klasifikasi Mahasiswa Lulusan Fakultas Pertanian	57
5. Klasifikasi Mahasiswa Lulusan Fakultas Teknik	59
6. Klasifikasi Mahasiswa Lulusan Fakultas Ekonomi dan Bisnis	61
7. Klasifikasi Mahasiswa Lulusan Fakultas Hukum	63
8. Klasifikasi Mahasiswa Lulusan Fakultas Ilmu Sosial dan Ilmu Politik	64
9. Klasifikasi Mahasiswa Lulusan Fakultas Keguruan dan Ilmu Pendidikan	67

V. SIMPULAN DAN SARAN

A. Kesimpulan.....	70
B. Saran.....	71

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR GAMBAR

Gambar	Halaman
2 1. Tahap – tahap pada <i>data mining</i>	7
2 2. Tahapan metode CRISP-DM	13
2 3. <i>Software RapidMiner</i>	15
3.1. Flowchart tahapan penelitian metode CRISP-DM	19
4.1. Data awal Forlap DIKTI	25
4.2. Data awal penerimaan SBMPTN	26
4.3. Data awal penerimaan SNMPTN	26
4.4. <i>Dataset</i> baru jalur masuk dengan NPM mahasiswa	27
4.5. <i>Dataset</i> baru hasil dari proses integrasi data	28
4.6. <i>Dataset</i> baru hasil integrasi data	28
4.7. <i>Dataset</i> yang akan di transformasi	30
4.8. Data yang telah dikelaskan	33
4.9. Rancangan proses klasifikasi	34
4.10. Konfigurasi data	35
4.11. Rancangan proses penerapan model <i>decision tree</i>	36
4.12. <i>Decision tree</i> yang dihasilkan <i>RapidMiner</i>	40
4.13. <i>Rule</i> atau aturan yang diperoleh	41
4.14. Atribut Pemodelan	42
4.15. Hasil klasifikasi	43

4.16. Korelasi paramter pemodelan	43
4.17. Nilai bobot pada <i>rule</i> klasifikasi	44
4.18. Akurasi model <i>decision tree</i> algoritma C4.5	45
4.19. Grafik ROC (<i>Receiver Operating Characteristic</i>).....	46
4.20. Persentase mahasiswa lulusan jalur masuk SNMPTN dan SBMPTN Universitas Lampung tahun 2016-2017.....	47
4.21. Persentase mahasiswa lulusan jalur masuk SNMPTN dan SBMPTN Universitas Lampung tahun 2016.....	48
4.22. Persentase mahasiswa lulusan jalur masuk SNMPTN dan SBMPTN Universitas Lampung tahun 2017.....	49
4.23. Persentase Kelas I per Fakultas Universitas Lampung.....	50
4.24. Persentase Kelas II per Fakultas Universitas Lampung	51
4.25. Grafik kelas mahasiswa lulusan Universitas Lampung per Fakultas	52
4.26. Persentase mahasiswa lulusan jalur masuk SNMPTN dan SBMPTN Fakultas Kedokteran	53
4.27. Grafik kelas mahasiswa lulusan Fakultas Kedokteran	54
4.28. Persentase mahasiswa lulusan jalur masuk SNMPTN dan SBMPTN Fakultas Matematika dan Ilmu Pengetahuan Alam	55
4.29. Grafik kelas mahasiswa lulusan Program Studi Fakultas Matematika dan Ilmu Pengetahuan Alam.....	56
4.30. Persentase mahasiswa lulusan jalur masuk SNMPTN dan SBMPTN Fakultas Pertanian.....	57
4.31. Grafik kelas mahasiswa lulusan Program Studi Fakultas Pertanian.....	58
4.32. Persentase kelas mahasiswa lulusan jalur masuk SNMPTN dan SBMPTN Fakultas Teknik	59
4.33. Garfik kelas Program Studi Fakultas Teknik.....	60
4.34. Persentase kelas mahasiswa lulusan jalur masuk SNMPTN dan SBMPTN Fakultas Ekonomi dan Bisnis	61
4.35. Garfik kelas Program Studi Fakultas Ekonomi dan Bisnis	62
4.36. Persentase mahasiswa lulusan jalur masuk SNMPTN dan SBMPTN Fakultas Hukum.....	63

4.37. Grafik kelas mahasiswa Lulusan Fakultas Hukum.....	64
4.38. Persentase mahasiswa lulusan jalur masuk SNMPTN dan SBMPTN Fakultas Ilmu Sosial dan Ilmu Politik	65
4.39. Garfik kelas mahasiswa lulusan Program Studi Fakultas Ilmu Sosial dan Politik.....	66
4.40. Persentase mahasiswa lulusan jalur masuk SNMPTN dan SBMPTN Fakultas Keguruan Ilmu Pendidikan	67
4.41. Garfik kelas mahasiswa lulusan Program Studi Fakultas Keguruan Ilmu Pendidikan.....	68
4.42. Garfik kelas mahasiswa lulusan Program Studi Fakultas Keguruan Ilmu Pendidikan (lanjutan)	68

DAFTAR TABEL

Tabel	Halaman
3 1. Jadwal penelitian.....	17
3.2. Deskripsi data.....	21
4.1. Paramater <i>class</i> data.....	31
4.2. Kategori Kelas I.....	32
4.3. Kategori Kelas II.....	32
4.4. Hasil tahap persiapan data.....	33
4.5. Jumlah kasus pada <i>dataset</i>	37
4.6. Perhitungan manual nilai <i>gain</i> dan <i>entropy</i>	39

DAFTAR SINGKATAN

1. BAN-PT (Badan Akreditasi Nasional – Perguruan Tinggi)
2. AEE (Angka Efisiensi Edukasi)
3. IPK (Indeks Prestasi Kumulatif)
4. SBMPTN (Seleksi Bersama Masuk Perguruan Tinggi Negeri)
5. SNMPTN (Seleksi Nasional Masuk Perguruan Tinggi Negeri)
6. PMPAP (Penerimaan Mahasiswa Perluasan Akses Pendidikan)
7. CRISP-DM (*Cross Industry Standard Process for Data Mining*)
8. SQL (*Structured Query Language*)
9. ID3 (*Iterative Dechotomiser 3*)
10. CART (*Classification and Regretion Tree*)
11. YALE (*Yet Another Learning Environment*)
12. ETL (*Extraction, Transformation, Loading*)
13. PDDIKTI (Pangkalan Data Pendidikan Tinggi)
14. Forlap DIKTI (Forum Laporan Pendidikan Tinggi)
15. PMB (Penerimaan Mahasiswa Baru)
16. BPPMB (Badan Pengelola Penerimaan Mahasiswa Baru)
17. NPM (Nomor Pokok Mahasiswa)
18. SKS (Satuan Kredit Semester)

- 19. AUC (*Area Under Curve*)
- 20. ROC (*Receiver Operating Characteristic*)
- 21. PC (*Personal Computer*)

I. PENDAHULUAN

A. Latar Belakang

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 menjaring 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. Sedangkan Ujian Mandiri di Universitas Lampung pada lima tahun terakhir mengalami perubahan, pada tahun 2017 Ujian Mandiri berubah menjadi Seleksi Mandiri Masuk Perguruan Tinggi (SMMPTN) Barat yang dilaksanakan oleh 10 daerah yang ada di pulau Sumatera, sedangkan jalur Advokasi dan PMPAP memiliki proses masuk yang berbeda dan tidak dibuka untuk semua jurusan di Universitas Lampung.

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. 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 Universitas Lampung.

B. Tujuan Penelitian

Tujuan dari penelitian ini antara lain:

1. Menerapkan teknik *data mining* pada basis data mahasiswa Universitas Lampung menggunakan metode *decision tree*.
2. Menemukan pola kategori kualitas mahasiswa Universitas Lampung berdasarkan jalur masuk SNMPTN dan SBMPTN.
3. Sebagai masukan dalam peninjauan kembali penggunaan metode dalam jalur seleksi pada Universitas Lampung sehingga menghasilkan mahasiswa yang berkualitas.

C. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah dapat memberikan solusi dalam peninjauan kembali metode jalur masuk mahasiswa baru Universitas Lampung sehingga dapat dijadikan salah satu tolak ukur dalam meningkatkan kualitas mahasiswa dan mutu pendidikan Perguruan Tinggi.

D. Rumusan Masalah

Berdasarkan permasalahan yang ada, perumusan masalah dari penelitian ini antara lain:

1. Bagaimana memperoleh pola dari data yang berjumlah besar agar dapat menghasilkan sebuah informasi?
2. Bagaimana mendapatkan suatu keputusan dengan teknik *decision tree* dan menggunakan Algoritma C4.5 dengan *software RapidMiner*?

3. Bagaimana menampilkan informasi agar dapat digunakan dalam membantu pengambilan keputusan untuk meningkatkan mutu Universitas Lampung?

E. Batasan Masalah

Berdasarkan permasalahan yang ada, perumusan perancangan ini difokuskan pada aspek berikut:

1. Tidak membahas lebih mengenai cara pengoperasian *software RapidMiner*.
2. Informasi yang ditampilkan berupa laporan analisa pola *data mining* tingkat klasifikasi kualitas mahasiswa berdasarkan jalur seleksi masuk SNMPTN dan SBMPTN.
3. Tidak membahas jalur penerimaan lokal dikarenakan kesulitan data.
4. Hanya membahas teknik *data mining* dengan metode *decision tree* dan Algoritma C4.5.
5. Hanya menggunakan data mahasiswa lulusan melalui jalur masuk SNMPTN dan SBMPTN.
6. Pola yang dihasilkan hanya sebagai pendukung keputusan, bukan sebagai faktor utama dalam mengambil keputusan.

F. Hipotesis

Dengan adanya batasan masalah dan berbagai literatur, peneliti memiliki ide untuk melakukan analisis pada basis data mahasiswa Universitas Lampung pada lima tahun terakhir guna mengetahui kualitas mahasiswa yang dimiliki Universitas Lampung menggunakan teknik *data mining* dengan metode *decision tree* dan Algoritma C4.5 yang membagi data sesuai dengan variabel yang ditentukan lalu

mengelompokkan data tersebut dengan kategori penentuan kualitas mahasiswa berdasarkan jalur masuk (SNMPTN dan SBMPTN).

G. Sistematika Penelitian

Sistematika penelitian yang dibuat dalam pembuatan laporan yaitu:

BAB I : PENDAHULUAN

Bab ini membahas tentang latar belakang penelitian, tujuan penelitian, manfaat penelitian, rumusan masalah, batasan masalah, hipotesis dan sistematika laporan.

BAB II : TINJAUAN PUSTAKA

Bab ini membahas tentang penjelasan umum mengenai prinsip, pengetahuan, dan teori penunjang mengenai *data mining*, metode *decision tree*, dan Algoritma C4.5, *k-fold cross validation*, CRISP-DM, serta *software RapidMiner v 9.1*.

BAB III : METODE PENELITIAN

Bab ini berisi waktu dan tempat penelitian, garis besar penelitian yang diusulkan serta metodologi *data mining* yang di usulkan.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini berisi pembahasan dan analisis terhadap hasil yang diperoleh.

BAB V : SIMPULAN DAN SARAN

Bab ini terdiri atas kesimpulan dari pembahasan dan saran mengenai penelitian yang di usulkan untuk pengembangan lebih lanjut.

DAFTAR PUSTAKA

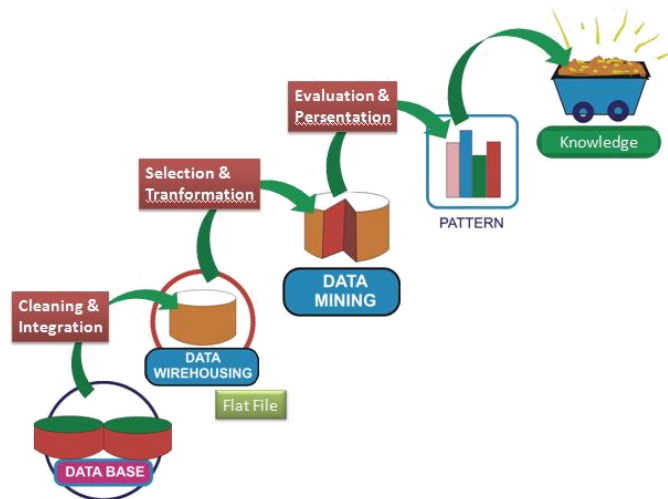
Berisi semua buku acuan yang digunakan untuk membantu penelitian.

II. TINJAUAN PUSTAKA

A. *Data Mining*

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[2].

Secara umum, *data mining* dilakukan untuk memahami lebih jauh perilaku data yang sedang diamati dan untuk memperkirakan atau memprediksi kondisi yang akan terjadi di masa yang akan datang.



Gambar 2.1. Tahap – tahap pada data mining[3]

Pada Gambar 2.1 merupakan tahap yang ada pada teknik *data mining* dengan penjelasan sebagai berikut[3]:

1. Pembersihan Data (*Data cleaning*)

Proses menghilangkan *noise* dan data yang tidak relevan atau konsisten.

2. Integrasi Data (*Data Integration*)

Proses Penggabungan data dari berbagai basis data menjadi basis data baru.

3. Seleksi Data (*Data Selection*)

Proses memilah data yang sesuai pada basis data untuk dianalisis.

4. Transformasi Data (*Data Transformation*)

Proses mengubah data atau menggabung data ke dalam format yang sesuai untuk diproses menggunakan teknik *data mining*.

5. Reduksi Data (*Data Reduction*)

Proses menguraikan data dalam bentuk yang lebih kecil namun tetap menghasilkan analisis yang sama.

6. Proses *Mining*

Proses utama saat menerapkan metode untuk mendapatkan suatu informasi dari data.

7. Evaluasi pola (*Pattern Evaluation*)

Proses mengidentifikasi pola ke dalam *knowledge based* yang didapatkan.

8. Presentasi Pengetahuan (*Knowledge Presentation*)

Proses penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh suatu pengetahuan yang didapatkan pengguna.

Berdasarkan tugas yang dapat dilakukan, *data mining* dibagi menjadi beberapa kelompok[2], yaitu:

1. Fungsi deskripsi (*description*)
2. Fungsi klasifikasi (*classification*)
3. Fungsi pengelompokan (*grouping*)
4. Fungsi asosiasi (*association*), dan
5. Fungsi estimasi (*estimation*)

B. Klasifikasi

Salah satu proses menemukan model yang didapatkan dari *data training* dan nilai kelas label dari atribut target yang akan digunakan untuk mengklasifikasi suatu data[4]. Teknik ini dapat digunakan untuk mengklasifikasi data baru dengan memanipulasi data yang telah diklasifikasi dengan menggunakan hasilnya untuk mendapatkan sejumlah aturan. Klasifikasi bertujuan untuk mengekstraksi pola atau mengklarifikasi humpunan data dari *training set* ke dalam kelas (*class*) tertentu berdasarkan atribut, kemudian model tersebut digunakan untuk mengklarifikasi

atribut yang kelasnya belum diketahui sebelumnya. Algoritma yang digunakan untuk melakukan klasifikasi antara lain *decision tree*, *naïve bayes*, *rule base methods*, *neural networks*[5].

Data training atau data latih adalah data yang sudah ada. Sedangkan *data testing* atau data uji merupakan data yang sudah berkelas atau berlabel yang digunakan untuk menghitung akurasi dari model klasifikasi yang dibentuk. Langkah dalam proses klasifikasi data[4], yaitu:

1. *Learning*, proses membangun Algoritma klasifikasi dengan menganalisis untuk menghasilkan *classification rules*.
2. *Classification*, proses pengujian data testing untuk memperkirakan akurasi *rule* yang diperoleh.

C. Teknik *Decision Tree* Menggunakan Algoritma C4.5 dalam Metode Klasifikasi

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]. Pohon (*tree*) merupakan suatu struktur data yang terdiri atas simpul (*node*) dan rusuk (*edge*). Simpul pada sebuah pohon dibagi menjadi tiga bagian, yaitu simpul akar (*root node*), simpul percabangan atau

internal (*branch/internal node*), dan simpul daun (*leaf node*)[3]. Algoritma dari model *decision tree* antara lain *Iterative Dichotomiser 3* atau *Induction of Decision 3 (ID3)*, *CART*, dan Algoritma C4.5

Algoritma C4.5 merupakan salah satu metode untuk membuat *decision tree* yang digunakan untuk melakukan klasifikasi atau pengelompokan yang bersifat prediktif dengan menggunakan *data training* yang ada. Algoritma C4.5 merupakan pengembangan dari Algoritma *ID3*. Secara umum untuk membangun Algoritma C.45[6], yaitu:

1. Pilih atribut sebagai akar (*root*).
2. Buat cabang setiap nilai.
3. Membagi kasus dalam cabang.
4. 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*.

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

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* :

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

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*. Semakin kecil nilai *error rate* (kesalahan) yang didapatkan, semakin akurat model *decision tree* yang didapatkan. Untuk menghitung nilai *error rate* maka menggunakan persamaan (3) :

$$Error Rate (ER): \frac{\text{banyaknya data error}}{\text{jumlah data}} = 100\% \dots \dots \dots (3)$$

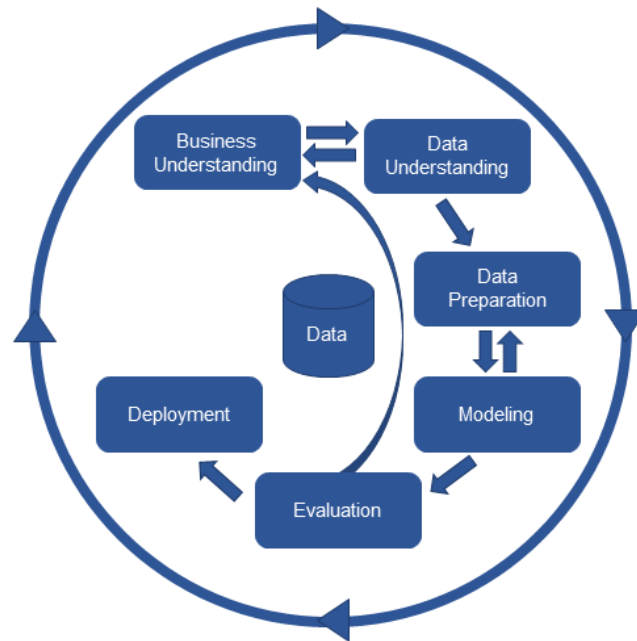
D. K-Fold Cross Validation

Cross validation merupakan metode untuk memvalidasi keakuratan sebuah model yang dibangun berdasarkan sebuah *dataset* tertentu. *Cross validation* digunakan untuk memprediksi atau mengklasifikasi terhadap data baru yang boleh jadi belum ada di dalam *dataset*. Data yang digunakan dalam membangun proses model yaitu *data training* dan *data testing*. Sedangkan data yang digunakan dalam melakukan validasi dalam *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[4]. Lipatan pertama diperlakukan sebagai set validasi, dan metode ini tepat pada lipatan $k - 1$ yang tersisa. Prosedur ini diulang k kali; setiap kali, sekelompok pengamatan yang berbeda diperlakukan sebagai set validasi. Proses ini menghasilkan estimasi k dari kesalahan pengujian, MSE_1 , MSE_2 , ..., MSE_k [7].

E. CRISP-DM

Cross Industry Standard Process for Data mining (CRISP-DM) merupakan standar proses pada *data mining* yang telah dikembangkan pada tahun 1996 yang ditujukan untuk melakukan proses analisis strategi pemecahan masalah secara umum dari bisnis atau untuk penelitian[8]. Pada tahun 1997, CRISP-DM menjadi proyek Uni Eropa di bawah inisiatif pendanaan ESPRIT yang dipimpin oleh lima perusahaan: Integral Solutions Ltd (ISL) , Teradata , Daimler AG , NCR Corporation dan OHRA , sebuah perusahaan asuransi[8]. Pada metode CRISP-DM terdapat enam fase yaitu *Business Understanding Phase* (Fase Pemahaman Bisnis), *Data Understanding Phase* (Fase Pemahaman Data), *Data Preparation* (Persiapan Data), *Modeling Phase* (Fase Pemodelan), *Evaluation Phase* (Fase Evaluasi), dan *Deployment Phase* (Fase Penyebaran). Pada Gambar 2.1 menunjukan tahapan pada metode CRISP-DM.



Gambar 2.2. Tahapan Metode CRISP-DM[9]

Tahapan metode CRISP-DM sebagai berikut :

1. Fase Pemahaman Bisnis (*Business Understanding Phase*)

Memahami tujuan dari perspektif bisnis, kemudian menjadikan pengetahuan tersebut sebagai rencana awal dalam masalah penambangan data.

2. Fase Pemahaman Data (*Data Understanding Phase*)

Melakukan pengumpulan data awal untuk mengidentifikasi kualitas data, menemukan pengetahuan pertama yang ada di dalam data, atau untuk mendeteksi masalah yang menarik untuk membuat hipotesis dengan informasi yang tersembunyi.

3. Persiapan Data (*Data Preparation*)

Melakukan persiapan data yang mencakup semua kegiatan untuk membuat *dataset* final yang berasal dari data mentah

4. Fase Pemodelan (*Modeling Phase*)

Memiliki teknik pemodelan yang dipilih untuk diterapkan.

5. Fase Evaluasi (*Evaluation Phase*)

Menguji model yang sudah dibangun untuk memastikan bahwa data sudah memiliki informasi yang sesuai dengan masalah bisnis.

6. Fase Penyebaran (*Deployment Phase*)

Menggunakan pola yang dihasilkan dari fase pemodelan untuk dipresentasikan ke dalam bentuk seperti laporan mengenai penerapan teknik data mining yang dilakukan atau analisa pola yang di dapatkan.

F. *RapidMiner*

RapidMiner merupakan salah satu *software* yang digunakan dalam pengolahan *data mining*. Pekerjaan yang dilakukan oleh *RapidMiner text mining* adalah berkisar dengan analisis teks, mengekstrak pola-pola dari *dataset* yang besar dan mengombinasikannya dengan metode statistika, kecerdasan buatan, dan *database*[10]. *RapidMiner* sebelumnya bernama YALE (Yet Another Learning Environment), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer di *Artificial Intelligence Unit* dari *University of Dortmund*[10]. *RapidMiner* menyediakan prosedur *data mining* dan *machine learning* di dalamnya termasuk ETL (*extraction, transformation, loading*), *data preprocessing*, visualisasi pemodelan dan evaluasi.



Gambar 2.3. *Software RapidMiner*

G. Penelitian Terdahulu

Berikut ini beberapa penelitian terdahulu yang terkait mengenai teknik *data mining*, metode *decision tree*, Algoritma C4.5 dan metode CRISP-DM yang akan digunakan dalam penelitian antara lain :

1. “Penerapan Algoritma C4.5 untuk Klasifikasi Predikat Kelulusan Mahasiswa Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta” oleh Yusuf Sulistyo 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% [11].
2. “Penerapan Algoritma C4.5 untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta” oleh Asroni, Badrahini Masajeng Respati, dan Slamet Riyadi pada tahun 2018. Pada penelitian ini bertujuan

untuk mengidentifikasi faktor-faktor yang berdampak pada jenis metode klasifikasi pekerjaan berdasarkan Algoritma alumni C 4,5. Parameter yang digunakan dalam pengujian adalah IPK, asal fakultas, dan tahun lulus dengan menggunakan data sebanyak 259 yang berasal dari 3 Fakultas yaitu Fakultas Ekonomi, Fakultas Kedokteran dan Fakultas Teknik. Hasil dari penelitian yaitu lulusan dari Fakultas Ekonomi, Fakultas Kedokteran dan Fakultas Teknik bekerja di sektor swasta[12].

3. “Data Mining untuk Klasifikasi Penentuan Peminatan Siswa SMA Negeri 2 Tenggarong Seberang dengan Menggunakan Algoritma C4.5” oleh Bambang Cahyono, Islamiyah. Pada penelitian ini menggunakan 150 set data yang terdiri dari 70% data pelatihan dan data pengujian adalah 30% dengan menggunakan *software Rapidminer*. Atribut yang digunakan untuk menentukan peminatan siswa terdiri dari 10 atribut antara lain: UN IPA, UN Bahasa Indonesia, UN Bahasa Inggris, UN Matematika, Tes IPA, Tes IPS, Tes Bahasa Indonesia, Tes Bahasa Inggris Hasil dari penelitian ini yaitu tingkat akurasi 84,44% *Confusion Matrix*[13].

III. METODOLOGI PENELITIAN

A. Waktu dan Tempat

Penelitian dan pembuatan tugas akhir ini dilakukan pada :

Waktu : Juli 2018 – Februari 2019

Tempat : Universitas Lampung

Jadwal kegiatan penelitian yang dilakukan dapat dilihat pada Tabel 3.1.

Tabel 3 1. Jadwal Penelitian

No	Aktifitas	Juli	Agus	Sept	...	Jan	Feb	Maret
1	Studi Literatur							
2	Fase Pemahaman Bisnis (<i>Business Understanding Phase</i>)							
3	Fase Pemahaman Data (<i>Data Understanding Phase</i>)							
4	Seminar usul							
5	Persiapan Data (<i>Data Preparation</i>)							
6	Fase Pemodelan (<i>Modeling Phase</i>)							
7	Evaluasi (<i>Evaluation</i>)							
8	Penyebaran (<i>Deployment</i>)							
9	Seminar Hasil							
10	Perbaikan							
11	Ujian Komprehensif							

B. Alat dan Bahan

Beberapa alat dan bahan yang digunakan pada penelitian dan pembuatan Tugas

Akhir ini sebagai berikut :

1. Satu buah *Personal Computer (PC)* ASUS X450J dengan spesifikasi :
 - *Processor* : Intel Core i7-4700HQ
 - *Video Grafik* : VGA Intel HD Graphics 4600 + Nvidia 840M 2GB
 - *RAM* : 4GB DDR3
 - *Sistem Operasi* : Windows 10
2. Perangkat lunak (*software*) pendukung seperti :
 - *RapidMiner ver.7.6* : *software* untuk pengolahan *data mining*
 - *Ms. Excel 2016* : *software* yang digunakan untuk *preprocessing data*

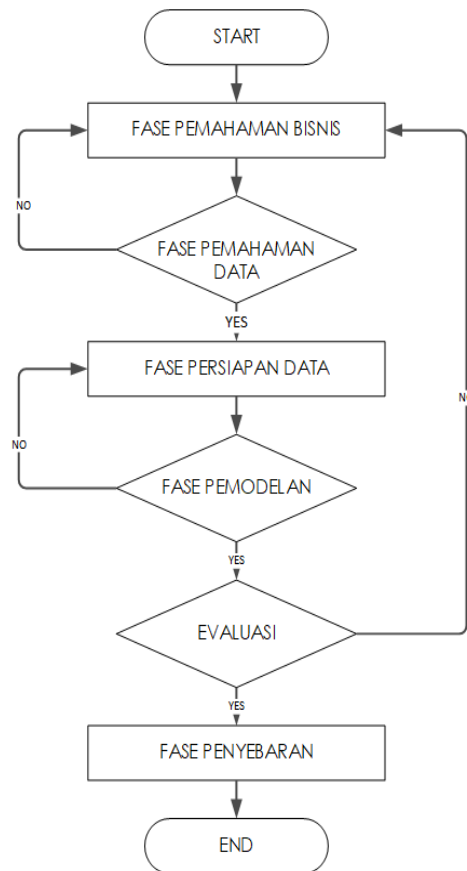
C. Tahapan Penelitian

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)*.

1. Studi Literatur

Studi literatur dilakukan untuk mempelajari ilmu dan penelitian yang pernah dibahas dari berbagai sumber seperti buku, jurnal dan artikel. Informasi yang dipelajari untuk mendukung penelitian ini mengenai *data mining*, klasifikasi, Algoritma C4.5 serta beberapa artikel yang berkaitan dengan penelitian ini.

2. Tahapan Penelitian dengan Metode CRISP-DM



Gambar 3.1. *Flowchart* tahapan penelitian metode CRISP-DM

Pada gambar 3.1 menunjukkan enam tahapan kerja dari metode *Cross Industry Standard Process Data Mining* (CRISP-DM) yang terdiri dari fase pemahaman bisnis (*business understanding phase*), dan fase pemahaman data (*data understanding phase*), fase persiapan data (*data preparation phase*), fase pemodelan (*modelling phase*), fase evaluasi (*evaluation*), dan penyebaran (*deployment*).

1. Fase pemahaman bisnis (*business understanding phase*)

Pada tahapan pemahaman bisnis ada beberapa hal yang dilakukan antara lain, menentukan tujuan bisnis, menilai situasi, dan menentukan tujuan *data mining*.

- 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 di serap 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. Oleh karena itu, diperlukan model klasifikasi untuk melihat kualitas mahasiswa dari proses jalur masuk sebagai penunjang pengambilan keputusan dalam upaya meningkatkan nilai AEE menggunakan teknik *data mining*.

- 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. Fase pemahaman data (*data understanding phase*):

Pada tahapan pemahaman data, ada beberapa hal yang dilakukan antara lain, mengumpulkan data awal, mendeskripsikan data, mengeksplorasi data, dan memverifikasi kualitas data.

- Mengumpulkan data

Data yang digunakan merupakan data yang didapatkan dari Pangkalan Data Pendidikan Tinggi (PDDIKTI) dengan 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* dapat dilihat pada table 3.2.

Tabel 3.2 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.

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 maka akan dilakukan Teknik persiapan data agar kualitas data diperoleh lebih baik dengan cara sebagai berikut :

- Pembersihan data (*data cleaning*) dan integrasi data (*data integration*), pembersihan data bertujuan untuk mengidentifikasi dan menghapus data yang ganjil (*noise*), data yang tidak konsisten dan data yang tidak lengkap (*missing value*). Sedangkan integrasi data merupakan proses menggabungkan data dari beberapa data menjadi data baru.
- Seleksi data (*data selection*) dan transformasi data (*data transformation*), seleksi data merupakan proses memilih data pada *database* yang akan digunakan. Sedangkan transformasi data merupakan proses mengubah data ke format yang sesuai untuk diproses menggunakan teknik *data mining*. Data yang digunakan dalam penelitian ini bernilai kategori untuk model klasifikasi, data ditransformasi ke dalam angka menggunakan *software RapidMiner*.

3. Pemodelan

Pada penelitian yang menggunakan Algoritma C4.5. Untuk melakukan pengukuran dalam akurasi penelitian ini akan menggunakan *framework RapidMiner 9.1*.

4. 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.

5. *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.

V. SIMPULAN DAN SARAN

A. Kesimpulan

1. Hasil pemodelan keputusan menggunakan Algoritma C4.5 didapatkan akurasi sebesar 97,46 % dengan nilai *error rate* sebesar 0,98%.
2. Pemodelan yang didapatkan memiliki nilai AUC sebesar 0,973 dengan nilai *error rate* sebesar 0,014 yang tergolong ke dalam *excellent classification*.
3. Atribut lama skripsi memiliki nilai *gain* tertinggi sebesar 0,614 sehingga menjadi atribut yang berpengaruh besar pada proses klasifikasi.
4. Pemodelan pada penelitian mendapatkan 13 *rule* klasifikasi tingkat kualitas mahasiswa lulusan berdasarkan jalur masuk SNMPTN dan SBMPTN.
5. Berdasarkan penelitian menghasilkan model *decision tree* seperti pada Gambar 4.4 dengan *rule* lama skripsi – IPK - lama studi – jalur (SBMPTN = kelas I), dengan bobot masing - masing, lama skripsi = tepat (1), lama studi = tepat (0,236), IPK = sangat baik (0,233), dan jalur masuk = SBMPTN (0).
6. 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.

B. Saran

Adapun saran dari penelitian ini adalah:

1. Diharapkan pada penelitian selanjutnya menggunakan aplikasi *software* selain *rapid miner* dan menggunakan Algoritma klasifikasi yang lain agar dapat membandingkan dan memberikan hasil yang lebih baik lagi.
2. Pada penelitian selanjutnya diharapkan dapat memasukan semua jalur penerimaan yang ada serta jalur penerimaan dengan mahasiswa Bidikmisi dan menggunakan data mahasiswa lulusan dengan data yang lebih banyak lagi sebagai bahan evaluasi Universitas Lampung.

DAFTAR PUSTAKA

- [1] BAN-PT, “Buku 1 Naskah Akademik Akreditasi Institusi Perguruan Tinggi,” *Badan Akreditasi Nas. Perguru. Tinggi*, 2011.
- [2] E. T. L. Kusrini, *Algoritma Data Mining*, 1st ed. Yogyakarta: Andi Offset, 2009.
- [3] F. A. Hermawati, *Data Mining*, 1st ed. Surabaya: Penerbit Andi, 2009.
- [4] 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.
- [5] D. S. S. . M.S. Sani Susanto Ph.D, *Pengantar Data Mining : Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta, 2010.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] C. Shaerer, *the-Crisp-Dm-Model-the-New-Blueprint-for-Data-Mining-Shearer-Colin*, vol. 5, no. 4. 2000.
- [10] L. A. Dennis Aprilla C, Donny Aji BAskoro, *Belajar Data Mining dengan RapidMiner*. Jakarta, 2013.
- [11] Y. S. Nugroho, “Penerapan algoritma c4.5 untuk klasifikasi predikat kelulusan mahasiswa fakultas komunikasi dan informatika universitas

muhammadiyah surakarta,” *Pros. Semin. Nas. Apl. Sains 1* February, 2014.

- [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