IMPLEMENTASI ALGORITME C4.5 UNTUK KLASIFIKASI KUALITAS MAHASISWA BERDASARKAN JALUR PENERIMAAN (STUDI KASUS: JURUSAN ILMU KOMPUTER FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS LAMPUNG)

(Skripsi)

ALDILA
1717051076

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021

ABSTRACT

IMPLEMENTATION OF ALGORITHM C.45 FOR CLASSIFICATION OF STUDENT QUALITY BASED ON ADMISSION PATH (CASE STUDY: DEPARTMENT OF COMPUTER SCIENCE FACULTY OF MATHEMATICS AND NATURAL SCIENCES UNIVERSITY OF LAMPUNG)

By

ALDILA

The Department of Computer Science is one of the majors in the Faculty of Mathematics and Natural Sciences, University of Lampung. Every 5 years, universities including majors in it require a process to evaluate the quality of students and maintain their accreditation. The quality of students can be seen from GPA, years of study, thesis time, admission, and others. The percentage of student quality can affect the assessment when accrediting. This study aims to analyze the data of graduate students from the Department of Computer Science, University of Lampung using data mining classification techniques and algorithm C4.5 decision tree to produce information about the quality of graduate students based on the SNMPTN and SBMPTN admission. The data mining method used in the classification is the Cross Industry Standard Process for Data Mining (CRISP-DM) and utilizes RapidMiner software in the application of algorithms C4.5 to produce decision trees. This study divided the quality of students into three classes based on the SNMPTN and SBMPTN admission. Based on the results, old classification years of study become the most influential attributes and students from the SBMPTN track are qualified better. The classification process produces an accuracy value of 99.03% with an error rate of 0.97%, which refering to an excellent classification.

Keywords *: Admission Path; Algorithm C4.5; Classification; Student Graduation;*

ABSTRAK

IMPLEMENTASI ALGORITME C.45 UNTUK KLASIFIKASI KUALITAS MAHASISWA BERDASARKAN JALUR PENERIMAAN (STUDI KASUS: JURUSAN ILMU KOMPUTER FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS LAMPUNG)

Oleh

ALDILA

Jurusan S1 Ilmu Komputer adalah salah satu jurusan yang ada di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung. Setiap 5 tahun sekali perguruan tinggi maupun jurusan didalamnya memerlukan proses untuk mengevaluasi kualitas mahasiswa dan mempertahankan akreditasi yang dimiliki. Kualitas mahasiswa dapat dilihat dari IPK, waktu studi, waktu skripsi, jalur penerimaan, dan lainnya. Persentase kualitas mahasiswa dapat mempengaruhi penilaian saat akreditasi. Penelitian ini bertujuan untuk menganalisis data mahasiswa lulusan dari Jurusan Ilmu Komputer Universitas Lampung menggunakan teknik klasifikasi *data mining* dan pohon keputusan algoritme C4.5 untuk menghasilkan informasi mengenai kualitas dari mahasiswa lulusan berdasarkan jalur penerimaan SNMPTN dan SBMPTN. Metode *data mining* yang digunakan dalam klasifikasi yaitu *Cross Industry Standard Process for Data Mining* (CRISP-DM) dan memanfaatkan *software RapidMiner* dalam penerapan algoritme C4.5 untuk menghasilkan pohon keputusan. Penelitian ini membagi kualitas mahasiswa menjadi tiga kelas berdasarkan jalur penerimaan SNMPTN dan SBMPTN. Berdasarkan hasil klasifikasi lama studi menjadi atribut paling berpengaruh dan mahasiswa jalur SBMPTN lebih unggul kualitasnya. Proses klasifikasi menghasilkan nilai akurasi sebesar 99,03% dengan *error rate* sebesar 0,97% yang tergolong dalam *excellent classification.*

Keywords **:** Algoritma C4.5; Kelulusan Mahasiswa; Klasifikasi; Jalur Penerimaan

IMPLEMENTASI ALGORITME C.45 UNTUK KLASIFIKASI KUALITAS MAHASISWA BERDASARKAN JALUR PENERIMAAN (STUDI KASUS: JURUSAN ILMU KOMPUTER FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS LAMPUNG)

Oleh

ALDILA

Skripsi

Sebagai Salah Satu Syarat Untuk Mencapai Gelar
SARJANA KOMPUTER

Pada

Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021

Judul Skripsi : **IMPLEMENTASI ALGORITME C4.5 UNTUK KLASIFIKASI KUALITAS MAHASISWA BERDASARKAN JALUR PENERIMAAN (STUDI KASUS: JURUSAN ILMU KOMPUTER FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS LAMPUNG)**

Nama Mahasiswa : *Aldila*

Nomor Pokok Mahasiswa : 1717051076

Program Studi : Ilmu Komputer

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Dr. Eng. Admi Syarif	**Dewi Asiah Shofiana, S.Komp., M.Kom.**
NIP 19670103 199203 1 003	NIP 19950929 202012 2 030

2. Ketua Jurusan Ilmu Komputer

Didik Kurniawan, S.Si., M.T.
NIP 19800419 200501 1 004

MENGESAHKAN

1. Tim Penguji

Ketua : **Dr. Eng. Admi Syarif** .....

Sekretaris : **Dewi Asiah Shofiana, S.Komp., M.Kom.** .....

Penguji
Bukan Pembimbing : **Didik Kurniawan, S.Si., M.T.** .....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T.
NIP 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : **02 November 2021**

PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : ALDILA

NPM : 1717051076

Jurusan/Prodi : Ilmu Komputer/S1-Ilmu Komputer

Menyatakan bahwa skripsi saya yang berjudul "**Implementasi Algoritme C.45 Untuk Klasifikasi Kualitas Mahasiswa Berdasarkan Jalur Penerimaan (Studi Kasus: Jurusan Ilmu Komputer Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Lampung)**" merupakan karya tulis saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang di skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila dikemudian hari terbukti skripsi saya merupakan hasil penjiplakan atau dibuat orang lain, maka bersedia menerima sanksi berupa pencabutan gelar yang telah saya terima.

Bandar Lampung, 02 November 2021

ALDILA

NPM. 1717051076

RIWAYAT HIDUP

Penulis dilahirkan di Jakarta, 08 September 1997. Penulis merupakan anak pertama dari dua bersaudara, buah hati dari pasangan Bapak Muhamad Arif Arifin dan Ibu Irtawati.

Penulis menyelesaikan Pendidikan Sekolah Dasar (SD) di SDN Sasanawiyata 01 pada tahun 2009, Pendidikan Sekolah Menengah Pertama (SMP) di SMP PGRI 06 Bogor pada tahun 2012, Pendidikan Sekolah Menengah Atas (SMA) di SMAN 08 Bogor pada tahun 2015 dan Pendidikan Diploma III (D3) di Universitas Lampung pada tahun 2018.

Tahun 2019, penulis terdaftar sebagai mahasiswa lanjut studi Jurusan Ilmu Komputer, Program Studi S1 Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung dan menyelesaikan studi sebagai sarjana Ilmu Komputer tahun 2021.

PERSEMBAHAN

Teriring rasa syukur dan cintaku kepada Sang Pencipta yang Maha Kasih,
Allah SWT atas segala berkah-Nya sehingga skripsi ini dapat terselesaikan.

Bismillahirohmannirrohim...

Kupersembahkan karyaku ini kepada

Papa dan Mamaku tercinta
Untuk segala pengorbanannya yang tiada henti memberi
cinta dan kasih sayang yang tak terhingga, kesabaran, perhatian serta iringan doa
yang tak pernah terputus disetiap perjalanan langkah anakmu ini

Adikku Danti dan semua keluarga besar untuk segala doa, motivasi, bantuan serta sumber inspirasiku yang mampu menghantarkanku hingga ke jenjang perguruan tinggi

Sahabat dan orang-orang yang menyayangiku Dendi Irvianto S.T, Rani, Mba Wulan, Mba Dessy, Debora
Terima kasih untuk dukungan dan semangatnya,
Kalian adalah anugerah terindah yang pernah kumiliki

Serta Almamater Tercinta,
Universitas Lampung

MOTTO

"Usaha Tidak Akan Mengkhianati Hasil"

"Karena sesungguhnya sesudah kesulitan itu ada kemudahan"
(QS. Al-Insyirah:7)

SANWACANA

Alhamdulillahirobbil'alamin, segala puji syukur penulis panjatkan kehadirat Allah SWT, karena atas rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi yang berjudul "*Implementasi Algoritme C.45 Untuk Klasifikasi Kualitas Mahasiswa Berdasarkan Jalur Penerimaan Menggunakan Algoritma C4.5 (Studi Kasus: Jurusan Ilmu Komputer Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Lampung)"* sebagai salah satu syarat untuk mendapatkan gelar Sarjana Komputer (S.Kom.) pada Program Studi S1 Ilmu Komputer Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung. Penulis menyadari sepenuhnya bahwa dalam penyusunan skripsi ini tak luput dari kesalahan dan masih jauh dari kesempurnaan.

Skripsi ini tidak akan terselesaikan tanpa adanya bantuan dan motivasi dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis mengucapkan banyak terima kasih kepada:

1. Diriku sendiri atas semangat, usaha, waktu dan pikirannya untuk menyelesaikan skripsi ini;
2. Orang tua tercinta, adik tercinta dan keluarga besar yang selalu memberikan doa, motivasi, cinta dan kasih sayang yang tak terhingga;
3. Bapak Dr. Eng. Admi Syarif sebagai pembimbing utama yang telah membimbing penulis dan memberikan ide, kritik serta saran sehingga penulisan skripsi ini dapat terselesaikan;
4. Ibu Dewi Asiah Shofiana, M.Kom., sebagai pembimbing kedua yang telah membimbing penulis dan memberikan bantuan, kritik serta saran dalam pembuatan skripsi ini;
5. Bapak Didik Kurniawan, S.Si., M.T., selaku pembahas/penguji dan Ketua Jurusan Ilmu Komputer FMIPA Universitas Lampung;
6. Ibu Anie Rose Irawati, M.Cs., selaku pembimbing akademik;

7. Bapak dan Ibu Dosen Jurusan Ilmu Komputer FMIPA Universitas Lampung yang telah memberikan ilmu pengetahuan dan pengalaman yang bermanfaat bagi penulis;
8. Seluruh staf administrasi dan karyawan FMIPA Universitas Lampung khususnya Jurusan Ilmu Komputer;
9. Ibu Astria Hijriani, S.Kom., M.Kom., selaku Sekretaris Jurusan Ilmu Komputer FMIPA Universitas Lampung;
10. Bapak Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T., selaku Dekan FMIPA Universitas Lampung;
11. Dendi Irvianto yang selalu setia menjadi tempat berbagi keluh kesah, memberikan saran, motivasi dan semangat;
12. Teman-temanku Hot Maharani Saragih, Wulan Destyaningsih, Dessy Maya Sari, Debora Ibotona Lumbantoruan, Resalina Oktaria yang telah memberikan saran, motivasi dan semangat;
13. Teman-teman seperjuangan lanjut studi yang telah memberikan saran, motivasi dan semangat;
14. Teman-teman KKN Milda, Yogi, Mia, Abiyu, Dita, yang telah memberikan saran, motivasi dan semangat;
15. Kostan Wisma Aysha yang menjadi tempat tinggalku selama berada di Lampung
16. Almamater tercinta, Universitas Lampung;
17. Serta kepada anda yang membaca skripsi ini, semoga tulisan ini dapat berguna dan bermanfaat bagi anda dan yang lainnya.

Bandar Lampung, 02 November 2021
Penulis,

Aldila

DAFTAR ISI

Halaman

DAFTAR ISI ..... **xii**
DAFTAR TABEL ..... **xiv**
DAFTAR GAMBAR ..... **xv**
I. PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Tujuan Penelitian 5
1.3 Manfaat Penelitian 6
1.4 Rumusan Masalah 6
1.5 Batasan Masalah 6
II. TINJAUAN PUSTAKA **7**
2.1 *Data Mining* 7
2.2 Kecerdasan Buatan 9
2.3 Klasifikasi 10
2.4 Teknik *Decision Tree* Menggunakan Algoritme C4.5 dalam Metode Klasifikasi 11
2.5 Algoritme C4.5 14
2.6 *K-Fold Cross Validation* 16
2.7 CRISP-DM 17
2.8 *RapidMiner* 19
2.9 Seleksi Jalur Masuk 21
2.10 Lama Masa Studi 23
2.11 Kelulusan Mahasiswa 24
III. METODE PENELITIAN **25**
3.1 Waktu dan Tempat 25
3.2 Bahan dan Alat 26

3.3 Tahapan Penelitian 26
3.3.1 Studi Literatur 26
3.3.2 Tahapan Penelitian dengan Metode CRISP-DM 27
IV. HASIL DAN PEMBAHASAN 34
4.1 Fase Persiapan Data 34
4.2 Fase Pemodelan 42
4.3 Fase Evaluasi 52
4.4 Fase *Deployment* 54
V. SIMPULAN DAN SARAN 68
5.1 Simpulan 68
5.2 Saran 68
DAFTAR PUSTAKA 69

DAFTAR TABEL

Tabel Halaman

1.1 Fakultas dan Jurusan di Universitas Lampung 2

3.1 Jadwal Penelitian 25

3.2 Deskripsi Atribut Data 30

4.1 Parameter *Class* Data 39

4.2 Kategori Kelas Satu 40

4.3 Kategori Kelas Dua 40

4.4 Kategori Kelas Tiga 40

4.5 Hasil Tahapan Persiapan Data 41

4.6 Jumlah Atribut Pada *Dataset* 45

4.7 Atribut Pemodelan 51

4.8 Nilai Bobot Pada *Rule* 52

4.9 Perbandingan Kelas Asal Daerah Mahasiswa Lulusan 2017 59

4.10 Perbandingan Kelas Asal Daerah Mahasiswa Lulusan 2018 61

4.11 Perbandingan Kelas Asal Daerah Mahasiswa Lulusan 2019 63

4.12 Perbandingan Kelas Asal Daerah Mahasiswa Lulusan 2020 65

DAFTAR GAMBAR

Gambar Halaman

2.1 *Data Mining* Sebagai Tahapan Dalam Proses *Knowledge Discovery in Database* 8

2.2 Proses Kecerdasan Buatan 10

2.3 Contoh Karakteristik *Decision Tree*..... 13

2.4 Tahapan Metode CRISP-DM..... 18

2.5 Tampilan Utama *RapidMiner* 20

3.1 Tahapan Metode CRISP-DM Dalam Penelitian 28

4.1 Data Awal Penerimaan SNMPTN..... 35

4.2 Data Awal Penerimaan SBMPTN..... 36

4.3 *Dataset* Baru Jalur Masuk Dengan NPM, Asal Sekolah, dan Asal Daerah Mahasiswa 36

4.4 D*ataset* Baru Proses Integrasi Data..... 37

4.5 *Dataset* Hasil Dari Proses Integrasi Data 38

4.6 *Dataset* Proses Transformasi Data 38

4.7 Data Yang Telah Diberi Label Kelas 41

4.8 Rancangan Proses Klasifikasi. 42

4.9 Konfigurasi *Dataset* 43

4.10 Proses Penerapan Model *Decision Tree.*..... 44

4.11 Hasil Pemodelan *Decision Tree* Menggunakan *RapidMiner*..... 49

4.12 Aturan Atau *Rule* Yang Dihasilkan 50

4.13 Hasil *Dataset* Klasifikasi 52

4.14 Akurasi Model *Decision Tree* Dengan Menggunakan Algoritme C4.5. 53

4.15 Klasifikasi *Error Rate* Model *Decision Tree* Dengan Menggunakan Algoritme C4.5 53

4.16 Perbandingan Jumlah Penerimaan Tahun 2013 Jalur SNMPTN dan SBMPTN 54

4.17 Perbandingan Jumlah Penerimaan Tahun 2014 Jalur SNMPTN dan SBMPTN 55

4.18 Perbandingan Jumlah Penerimaan Tahun 2015 Jalur SNMPTN dan SBMPTN 56

4.19 Perbandingan Jumlah Penerimaan Tahun 2016 Jalur SNMPTN dan SBMPTN. 56

4.20 Perbandingan Jumlah Penerimaan Tahun 2017 Jalur SNMPTN dan SBMPTN 57

4.21 Perbandingan Kelas Pada Kelulusan Tahun 2017 Jalur Masuk SNMPTN dan SBMPTN 58

4.22 Perbandingan Kelas Pada Kelulusan Tahun 2017 Jalur Masuk SNMPTN dan SBMPTN 60

4.23 Perbandingan Kelas Pada Kelulusan Tahun 2019 Jalur Masuk SNMPTN dan SBMPTN 62

4.24 Perbandingan Kelas Pada Kelulusan Tahun 2020 Jalur Masuk SNMPTN dan SBMPTN. 64

4.25 Jumlah Keseluruhan Lulusan Setiap Angkatan Jalur SNMPTN. 66

4.26 Jumlah Keseluruhan Lulusan Setiap Angkatan Jalur SBMPTN. 67

I. PENDAHULUAN

1.1 Latar Belakang

Pada zaman sekarang seiring berjalannya waktu, kemajuan teknologi menjadikan banyak masyarakat memerlukan data untuk melakukan suatu proses, sehingga menyebabkan banyak data dapat diperoleh dengan mudah bahkan dengan jumlah yang berlebihan. Data yang terakumulasi jumlahnya semakin banyak, mengakibatkan data yang telah terakumulasi penggunaannya menjadi tidak optimal. Suatu organisasi mengolah data untuk diproses dengan teknik klasifikasi, namun karena memiliki data yang banyak menyebabkan kesulitan dalam pengklasifikasian data untuk kepentingan organisasi tersebut (Cahyono & Islamiyah, 2016).

Manusia yang melakukan kegiatan pengklasifikasian memiliki banyak keterbatasan, terlebih pada kemampuan manusia jika menampung jumlah data yang cukup besar dan ingin diklasifikasikan. Maka dari itu untuk mengurangi tingkat risiko terjadinya kesalahan dalam pengklasifikasian karena jumlah data yang terlalu besar, salah satu solusi untuk mengatasi masalah ini adalah dengan menggunakan teknik klasifikasi di *data mining* (DM).

Data mining adalah sebuah proses untuk menghasilkan sebuah pola atau informasi dari kumpulan data dengan menyaring data yang berukuran besar dan hasilnya sebagai pendukung keputusan. Dalam suatu organisasi *data mining* dapat membantu untuk menghasilkan suatu informasi yang digunakan sebagai pendukung pengambilan keputusan. Tidak hanya organisasi, di dalam dunia pendidikan data yang berlimpah dan berkesinambungan perihal mahasiswa dan alumni terus dihasilkan. Banyaknya jumlah data yang berlimpah menjadikan peluang untuk menerapkan *data mining* dalam pengelolaan data pendidikan yang lebih baik (Cahyono & Islamiyah, 2016).

Perguruan tinggi adalah sarana pendidikan akademik bagi mahasiswa. Perguruan tinggi pada setiap tahunnya melakukan proses evaluasi dan penilaian mutu terhadap perguruan tinggi tersebut. Salah satu tolak ukur dari mutu suatu perguruan tinggi berdasarkan standar akreditasi yang sudah ditetapkan oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) yaitu tingkat kualitas mahasiswanya, karena hal tersebut dapat mempengaruhi Angka Efisiensi Edukasi (AEE). Mutu kualitas mahasiswa dilihat dari nilai akhir mahasiswa yaitu Indeks Prestasi Kumulatif (IPK) mahasiswa, kemampuan dalam menyelesaikan studi tepat waktu, jalur masuk penerimaan, masa tenggat, dan juga faktor lainnya (BAN-PT, 2011).

Tolak ukur kualitas mahasiswa salah satunya adalah dari jalur masuk penerimaan sebagai sarana untuk penerimaan mahasiswa baru. Setiap jalur masuk memiliki standar kualifikasi yang berbeda yang tujuannya untuk menilai kemampuan dari calon mahasiswa yang akan diterima (Oktaria dkk., 2019). Dapat dilihat pada Tabel 1.1 Universitas Lampung memiliki 8 fakultas dan berbagai macam jurusan. Jalur penerimaan yang dimiliki oleh Universitas Lampung yaitu Seleksi Nasional Masuk Perguruan Tinggi (SNMPTN), Jalur Lokal atau Ujian Mandiri, Advokasi, dan PMPAP (Universitas Lampung, 2015). Tabel 1.1 menunjukkan Fakultas dan Jurusan yang ada di Universitas Lampung.

Tabel 1.1 Fakultas dan Jurusan di Universitas Lampung (Universitas Lampung, 2015).

No	**Fakultas**	**Jurusan**
1	FMIPA	a. Ilmu Komputer
		b. Biologi
		c. Kimia
		d. Matematika
		e. Fisika
2	FEB	a. Manajamen
		b. Akuntansi
		c. Ekonomi Pembangunan
		d. Perpajakan

No	Fakultas	Jurusan
		e. Keuangan
		f. Pemasaran
3	HUKUM	a. Ilmu Hukum
4	FKIP	a. Penjas
		b. Bimbingan Konseling
		c. Pendidikan Ekonomi
		d. Pendidikan Sejarah
		e. Pendidikan PKN
		f. Pendidikan Geografi
		g. Pendidikan Matematika
		h. Pendidikan Kimia
		i. Pendidikan Biologi
		j. Pendidikan Fisika
		k. Pendidikan Bahasa Indonesia
		l. Pendidikan Seni
		m. Pendidikan Bahasa Inggris
		n. PGSD
5	Pertanian	a. Agroteknologi
		b. Agribisnis
		c. THP
		d. Peternakan
		e. Kehutanan
		f. Budidaya Perairan
		g. Teknik Pertanian
6	Teknik	a. Teknik Sipil
		b. Teknik Mesin
		c. Teknik Informatika
		d. Teknik Kimia
		e. Teknik Geofisika
7	FISIP	a. Sosilogi
		b. Ilmu Pemerintahan
		c. Adm Bisnis
		d. Adm Negara
		e. Ilmu Komunikasi

No	Fakultas	Jurusan
		f. HI
8	Kedokteran	a. Pendidikan Dokter

Berdasarkan Tabel 1.1 Jurusan S1 Ilmu Komputer Universitas Lampung termasuk salah satu jurusan yang ada di Fakultas Matematika dan Ilmu Pengetahuan Alam. Jurusan Ilmu Komputer memiliki standar program pendidikan tinggi yang memiliki beban studi sebanyak 144 SKS (Satuan Kredit Semester) yang diharapkan dapat ditempuh dalam waktu kurang dari 8 semester. Jalur masuk yang akan digunakan dalam penelitian ini yaitu SNMPTN dan SBMPTN yang memiliki keterkaitan dalam proses masuk penerimaan mahasiswa baru di Jurusan Ilmu Komputer Universitas Lampung.

Data yang akan digunakan adalah data lulusan mahasiswa berdasarkan jalur penerimaan. Untuk menganalisis data tersebut, dibutuhkan suatu teknik klasifikasi yang merupakan bagian dari *data mining*. Terdapat berbagai teknik dalam *data mining*, salah satunya yaitu pohon keputusan *(decision tree). Decision tree* (pohon keputusan) adalah pohon klasifikasi prosedur penalaran untuk mendapatkan jawaban dari masalah yang dikasuskan (Aprilia C dkk., 2013)**.**

Penelitian sebelumnya dilakukan oleh (Zega, 2014) berjudul "penggunaan pohon keputusan untuk klasifikasi tingkat kualitas mahasiswa berdasarkan jalur masuk kuliah". Hasil proses *learning* dan *classification* dalam penelitian ini menghasilkan 2 model *decision tree* yang memenuhi persyaratan, yaitu iterasi 4 dan 7. Berdasarkan iterasi tersebut, dapat diketahui bahwa jalur masuk penerimaan mahasiswa melalui jalur penelusuran minat dan kemampuan (PMDK) memiliki presentasi yang lebih besar dalam menghasilkan mahasiswa yang berkualitas yaitu 90% pada iterasi 7 dibandingkan dengan ujian masuk penerimaan Bersama perguruaan tinggi (UMPB) pada iterasi 4 sebesar 80,77%.

Penelitian lain juga telah dilakukan oleh (Asroni dkk., 2018) dengan judul "Penerapan algoritme C4.5 untuk klasifikasi jenis pekerjaan alumni di

Universitas Muhammadiyah Yogyakarta". Penelitian tersebut bertujuan untuk mengidentifikasi faktor-faktor yang berdampak pada jenis metode klasifikasi pekerjaan berdasarkan algoritme C 4.5. Parameter yang digunakan dalam pengujian adalah IPK, asal fakultas, dan tahun lulus dengan menggunakan sebanyak 259 data yang berasal dari tiga fakultas. Hasil dari penelitian ini menunjukkan bahwa lulusan dari Fakultas Ekonomi, Fakultas Kedokteran, dan Fakultas Teknik sebagian besar bekerja di sektor swasta.

(Risnawati, 2018) juga telah melakukan peneltian dengan judul "Analisis kelulusan mahasiswa menggunakan algoritme C4.5". Penelitian ini bertujuan untuk membandingkan setiap atribut yang mempengaruhi kelulusan mahasiswa. Berdasarkan proses penelitian yang dilakukan terdapat empat hasil analisis mengenai hubungan atribut-atribut yang mempengaruhi kelulusan mahasiswa. Keempat *rule* tersebut dapat menjadi masukan bagi mahasiswa maupun akademik dalam menentukan strategi yang tepat dalam mengatasi permasalahan kelulusan bagi mahasiswa.

Berdasarkan latar belakang permasalahan dan penelitian-penelitian yang pernah dilakukan sebelumnya, maka pada penelitian ini akan dilakukan proses *data mining* dengan menerapkan metode klasifikasi algoritme C4.5 untuk menganalisis data mahasiswa lulusan Jurusan Ilmu Komputer Universitas Lampung. Algoritme C4.5 merupakan algoritme klasifikasi pohon keputusan yang sering digunakan karena kelebihan atau keunggulannya dalam menganalisis dan membagi data yang besar menjadi *record* yang lebih kecil yang hasil akhirnya adalah pohon keputusan dengan tingkat akurasi yang tinggi. Penelitian ini diharapkan dapat menganalisis data dalam jumlah besar hasil, sehingga akhirnya ditemukan pola dari algoritme yang digunakan untuk mendapatkan pengetahuan dan informasi baru mengenai kualitas lulusan pada mahasiswa Jurusan Ilmu Komputer berdasarkan jalur penerimaan SNMPTN dan SBMPTN sesuai parameter kualitas mahasiswa dari BAN-PT.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini yaitu:

1. Mengimplementasikan metode *decision tree* dengan algoritme C4.5 untuk mengklasifikasi kualitas mahasiswa Jurusan Ilmu Komputer Universitas Lampung.
2. Mengevaluasi kinerja dari metode *decision tree* dengan algoritme C4.5.

1.3 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah mendapatkan pengetahuan dan informasi baru mengenai klasifikasi kualitas lulusan yang ada dan dapat memberikan solusi untuk peninjauan kualitas mahasiswa lulusan berdasarkan jalur penerimaannya di Jurusan Ilmu Komputer.

1.4 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah pada penelitian ini adalah bagaimana cara mendapatkan pola dari algoritme C4.5 untuk menghasilkan pengetahuan dan informasi baru mengenai klasifikasi kualitas lulusan pada mahasiswa Jurusan Ilmu Komputer berdasarkan jalur penerimaan SNMPTN dan SBMPTN sesuai parameter kualitas mahasiswa dari BAN-PT.

1.5 Batasan Masalah

Berdasarkan masalah yang ada, pada penelitian ini akan memfokuskan pada aspek berikut:

1. Informasi yang dihasilkan berupa analisis pola dari algoritme C4.5 *data mining*.
2. Data yang digunakan hanya data mahasiswa Jurusan Ilmu Komputer jalur penerimaan SNMPTN dan SBMPTN.
3. Data mahasiswa Jurusan Ilmu Komputer yang digunakan dari tahun 2013-2017.

II. TINJAUAN PUSTAKA

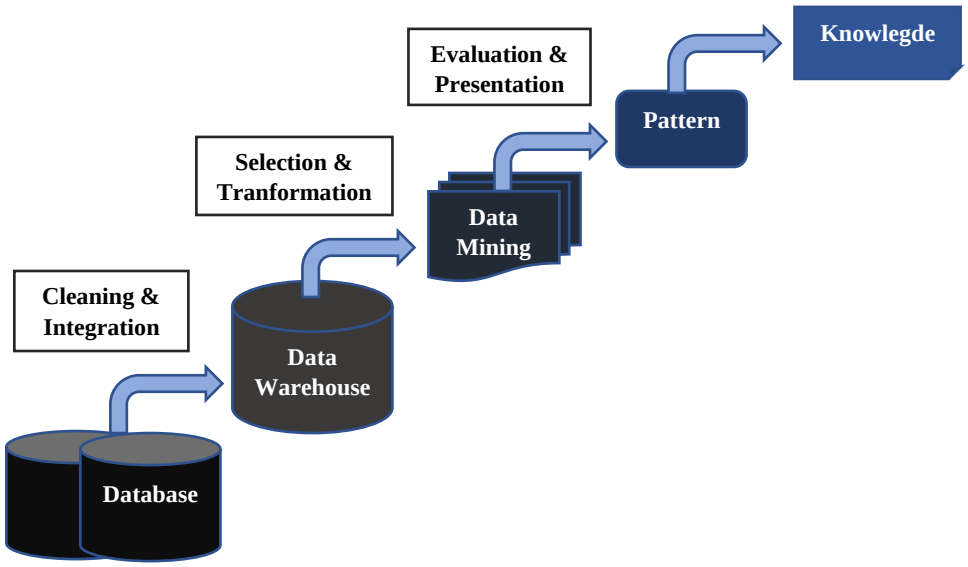
2.1 *Data Mining*

Data mining adalah proses untuk menganalisis data untuk menemukan suatu pola dari kumpulan data. *Data mining* dapat menganalisis data yang berukuran besar menjadi informasi berbentuk pola dan hasilnya memiliki arti sebagai pendukung keputusan (Risnawati, 2018).

Dari definisi yang telah diuraikan simpulannya adalah sebagai berikut:

1. *Data mining* merupakan inti informasi yang ada dalam *database* besar.
2. *Data mining* adalah suatu proses yang meliputi pengumpulan, pemakaian data yang telah tersimpan untuk mencari keteraturan, pola, dan hubungan antar data yang ukurannya besar. Hasil keluaran dari *data mining* dapat dipakai untuk pengambilan keputusan di masa yang akan datang, serta meningkatkan keuntungan dan kecil biaya pengeluarannya.

Data mining juga biasanya disebut dengan *Knowledge Discovery in Database* (KDD). *Data mining* sendiri merupakan bagian dari tahapan yang ada dalam proses KDD. Gambar 2.1 berikut ini yaitu proses *data mining* dalam KDD (Han & Kamber, 2006).



Gambar 2.1. *Data mining* Sebagai Tahapan Dalam Proses *Knowledge Discovery in Database* (Han & Kamber, 2006)

Pada Gambar 2.1 proses di dalam KDD menunjukkan bahwa *data mining* sebagai salah satu tahapan yang ada didalamnya. Berikut ini penjelasan tentang proses yang ada di dalam KDD yaitu (Risnawati, 2018).

1. Pembersihan Data *(Data Cleaning).*

   Proses untuk menghilangkan data yang *noise* dan data yang tidak lengkap.

2. Integrasi Data *(Data Integration).*

   Proses yang dilakukan untuk menggabungkan beberapa data dari berbagai basis data menjadi basis data baru.

3. Seleksi Data *(Data Selection).*

   Pada proses ini pemilihan data yang sesuai untuk dianalisis. Memilih data dari sekumpulan data diperlukan sebelum tahap pencarian informasi dalam KDD dimulai.

4. Transformasi Data *(Data Transformation).*

   Proses mengubah data atau menggabungkan data ke dalam format yang sesuai dan selanjutnya diproses menggunakan teknik *data mining.*

5. Reduksi Data (*Data Reduction*).

   Pada proses ini data diubah ke dalam bentuk yang lebih kecil namun tetap menghasilkan analisis yang sama.

6. Proses *Mining.*

   Proses ini bertujuan untuk mencari pola atau informasi menarik dalam data yang terpilih dengan menggunakan teknik atau metode *data mining* tertentu.

7. Evaluasi pola *(Pattern Evaluation).*

   Proses evaluasi mengenali pola ke dalam *knowledge based* yang didapatkan. Pola pengetahuan yang didapatkan dari proses *data mining* dapat ditampilkan dengan bentuk yang mudah dimengerti oleh berbagai pihak. Pada tahapan proses KDD ini disebut juga dengan *interpretation.*

8. Presentasi Pengetahuan *(Knowledge Presentation).*

   Pada Proses ini penyampaian pengetahuan mengenai metode yang digunakan untuk mendapatkan pengetahuan bagi pengguna.

Pada *data mining* pada umumnya dibagi menjadi beberapa fungsi. Berikut ini fungsi yang ada di dalam *data mining* yaitu (Han & Kamber, 2006):

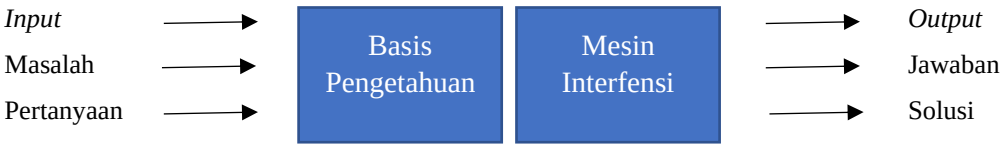
1. Fungsi deskripsi *(description)*
2. Fungsi klasifikasi *(classification)*
3. Fungsi pengelompokan *(grouping)*
4. Fungsi asosiasi *(association)*, dan
5. Fungsi estimasi *(estimation)*

2.2 Kecerdasan Buatan

Sebagai seorang manusia yang memiliki kecerdasan dan memiliki kemampuan untuk menganalisis suatu permasalahan yang ada dengan menggunakan pengetahuan yang dimiliki dalam otaknya dan berdasarkan pengalaman yang pernah dilaluinya. Pengetahuan juga didapati ketika seorang manusia belajar, maka dari itu suatu pembelajaran adalah faktor utama bagi manusia untuk mendapatkan kecerdasan. Namun pengetahuan tidak akan cukup untuk menyelesaikan suatu masalah jika tidak adanya pengalaman, karena pengalaman akan selalu mendatangkan pengetahuan baru. Tetapi hal itu akan menjadi tidak berguna jika seseorang yang sudah memiliki banyak pengalaman tetapi tidak mempunyai akal untuk mengerti semua pengetahuan dan

pengalaman yang telah dimiliki. Penggabungan dari pengetahuan, pengalaman, dan kemampuan berpikir ini yang menjadikan manusia menjadi cerdas dan dapat menyelesaikan berbagai masalah yang akan dihadapi. Berdasarkan pola pikir yang telah dijelaskan kecerdasan buatan akhirnya dibuat supaya mesin mampu bekerja layaknya seorang manusia. Lalu mesin tersebut perlu mempunyai beberapa pengetahuan dan pengalaman dan juga potensi penalaran yang dapat mengubah pengetahuan dan pengalaman sebagai keputusan untuk memecahkan suatu permasalahan (Aprilia C dkk., 2013).

Berdasarkan pernyataan tersebut tentang konsep dibuatnya kecerdasan buatan Kecerdasan buatan adalah salah satu bagian dari cabang ilmu komputer yang mampu membuat sebuah mesin dapat melakukan pekerjaan dan aktivitas seperti dan selayaknya dilakukan oleh manusia. Maka dari itu untuk membentuk sebuah operasi dari kecerdasan buatan ditemukan dua bagian penting yang sangat diperlukan dalam proses kecerdasan buatan (Aprilia C dkk., 2013). Gambar 2.2 menunjukkan proses dari kecerdasan buatan.



Gambar 2.2 Proses Kecerdasan Buatan

2.3 Klasifikasi

Klasifikasi adalah salah satu bagian dari proses untuk mendapatkan model yang dihasilkan dari *data training* dan nilai pada kelas dari atribut tujuan yang akan digunakan untuk mengklasifikasi suatu data (Zega, 2014). Dalam proses ini teknik yang dapat digunakan untuk mengklasifikasikan sebuah data baru dengan memanipulasi data yang sudah diklasifikasikan dengan memanfaatkan hasilnya untuk memperoleh beberapa aturan. Klasifikasi bertujuan untuk menghasilkan pola atau mengelompokkan kumpulan data dari *training set* ke dalam kelas (*class*) tertentu berdasarkan atribut. Selanjutnya model dipakai untuk mengklasifikasikan atribut yang kelasnya belum ditemukan sebelumnya.

Algoritme yang biasanya digunakan untuk klasifikasi yaitu *decision tree, naïve bayes, rule base methods*, dan *neural networks* (Zega, 2014).

Data training atau data latih merupakan data yang tersedia, sedangkan untuk *data testing* atau data uji adalah data yang dikelompokkan atau berlabel yang siap digunakan untuk menghitung akurasi dari model klasifikasi yang telah dibentuk. Berikut ini adalah langkah dalam proses klasifikasi yaitu (Saifudin, 2018):

1. *Learning,* proses untuk membentuk algoritme klasifikasi untuk menganalisis dan menghasilkan *classification rules.*
2. *Classification,* pada proses ini pengujian pada *data testing* dilakukan untuk memperkirakan akurasi *rule* yang akan dihasilkan.

2.4 Teknik *Decision Tree* Menggunakan Algoritme C4.5 dalam Metode Klasifikasi

Teknik *classification decision tree* adalah Teknik klasifikasi yang mudah dimengerti dan banyak digunakan. Pada bagian ini menjelaskan bagaimana *decision tree* beroperasi dan bagaimana *decision tree* dibuat. Ada kalanya untuk mengklasifikasikan suatu objek diperlukan adanya urutan pertanyaan sebelum ditentukan kelompoknya. Jawaban untuk pertanyaan pertama akan mempengaruhi pertanyaan selanjutnya. Di dalam *decision tree*, pertanyaan yang pertama kali diajukan pada simpul akar berada di level 0. Jawaban dari pertanyaan ini ditampilkan di bagian cabang. Jawaban yang ada di dalam cabang akan dilanjutkan dengan pertanyaan selanjutnya melewati simpul yang berikutnya pada level 1 (Darmawan dkk., 2017). Pohon (*tree*) adalah suatu bentuk data yang berisi simpul (*node*) dan rusuk (*edge*). Lalu pada simpul pohon dibagi menjadi tiga bagian yaitu simpul akar (*root node*), simpul percabangan atau internal (*branch/internal node*), dan simpul daun (*leaf node*) (Darmawan dkk., 2017).

Decision tree memiliki karakteristik yang terdiri dari sejumlah elemen yaitu (Kane dkk., 2016) :

1. *Node*

   Bagian yang memiliki variabel, *Node* dapat berupa variabel akar, variabel cabang, dan variabel kelas.

2. *Arm*

   Setiap cabangnya menjelaskan nilai hasil pengujian di *node* bukan di *leaf*.

3. *Node Root*

   Node Root tidak memiliki *input arm* yaitu lengan masukan tetapi memiliki *output arm* yaitu lengan keluar.

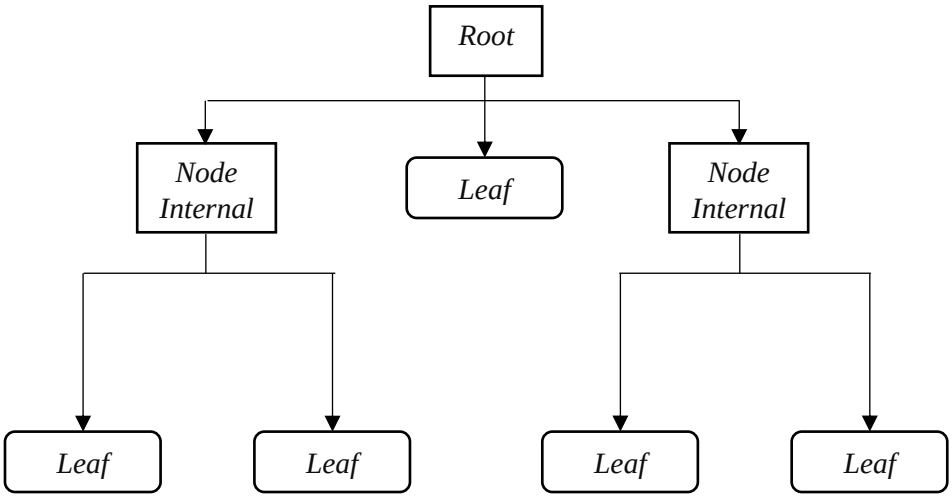
4. *Node internal*

   Pada setiap *node* yang bukan *leaf* (*nonterminal*) memiliki hanya satu *input arm* dan lebih dari satu *output arm*. Pengujian pada *node internal* dinyatakan berdasarkan nilai *fitur*.

5. *Node leaf* (*terminal*)

   Node leaf yaitu *node* yang hanya mempunyai satu *input arm* dan tidak memiliki *output arm*.

Gambaran dari karakteristik yang dimiliki *decision tree* yang telah dijelaskan dapat dilihat pada Gambar 2.3.



Gambar 2.3 Contoh Karakteristik *Decision Tree*

Pada Gambar 2.3 terlihat adanya 4 level pertanyaan. Di setiap level pertanyaan nilai atribut melewati suatu simpul, jawaban dari pertanyaan akan dinyatakan lewat cabang-cabang. Langkah ini akan selesai di suatu simpul jika pada simpul tersebut telah diketahui kelas atau jenis objeknya. Tingkat pertama objek sudah didapati tertera dalam kelas, lalu proses berhenti pada level tersebut. Jika tidak diketahui maka tetap dilanjutkan dengan pertanyaan selanjutnya level berikutnya sampai dengan jelas dan jenis objek dapat ditentukan (Darmawan dkk., 2017).

Sebelum *decision tree* dapat dihasilkan harus menghitung nilai *entropy* dan *gain*-nya terlebih dahulu. Lalu selanjutnya *decision tree* yang diperoleh diuji untuk mengetahui nilai keakuratan dari *decision tree*. Jika nilai kesalahan (*error rate*) semakin kecil maka yang dihasilkan semakin akurat model *decision tree*-nya dan juga sebaliknya jika nilai kesalahan (*error rate*) semakin besar maka semakin tidak akurat hasil model *decision tree*-nya.

Berikut ini adalah Persamaan 1 untuk menghitung nilai *error rate* di dalam *decision tree* (Oktaria dkk., 2019).

$$\textit{Error Rate}\ (\mathrm{ER}): \frac{\text{Jumlah Data Error}}{\text{Jumlah Data}} = 100\% \ \ldots\ldots\ldots\ldots\ (1)$$

Persamaan 1 menunjukkan cara menghitung nilai *error rate* yang ada di *decision tree*, dapat dilihat bahwa jumlah data yang *error* dibagi dengan jumlah data yang ada lalu dikalikan dengan 100%.

2.5 Algoritme C4.5

Algoritme adalah sekumpulan perintah yang tertulis secara sistematis untuk memecahkan masalah logika dari aritmatika. Algoritme C4.5 dan pohon keputusan adalah dua tipe yang tidak terpisahkan, karena untuk membuat suatu pohon keputusan diperlukan algoritme C4.5. Pohon keputusan menyerupai suatu struktur pohon yang terdapat *node internal* untuk menjelaskan atribut-atribut, pada cabang mewakili hasil atribut yang diuji, dan setiap daun

mendeskripsikan kelas. Pohon keputusan bergerak dimulai dari akar paling atas (*root*) (Darmawan dkk., 2017).

Algoritme C4.5 adalah algoritme yang dibutuhkan untuk membuat *decision tree*. Awalnya algoritme C4.5 pada tahun 1970 hingga pada tahun 1980, seseorang bernama J.Ross Quinlann peneliti bidang mesin pembelajaran membuat suatu model pohon keputusan yang diberi nama ID3 (*Iterative Dichotomiser*) meskipun sebelumnya proyek ini telah diciptakan oleh E.B.Hunt, J. Marin, dan P.T.Stone. Lalu oleh Quinlann dibuatlah algoritme yang merupakan pengembangan dari algoritme ID3 dinamakan algoritme C4.5 yang berbasis *supervised learning* (Rismayanti, 2018).

Pada algoritme C4.5 *decision tree* diartikan sebagai salah satu cara untuk memprediksi atau mengklasifikasi suatu objek yang sangat besar. *Decision tree* mampu membagi sekumpulan data yang ukurannya besar menjadi himpunan *record* yang lebih kecil. Algoritme C4.5 merupakan algoritme untuk menghasilkan suatu *decision tree* yang diperlukan untuk melakukan klasifikasi atau pengelompokan objek yang sifatnya predikatif dengan memanfaatkan *data training* pada algoritme C4.5. Umumnya untuk membuat algoritme C4.5 diperlukan tahapan sebagai berikut (Kane dkk., 2016).

1. Pemilihan atribut menjadi variabel akar *(root).*
   Menentukan variabel akar dipilih berdasarkan hasil nilai *gain* tertinggi dari variabel-variabel. Pada persamaan 2 ditunjukkan cara untuk mencari nilai *gain* tertinggi.
2. Membuat cabang untuk setiap nilai.
   Pemilihan cabang yaitu mencari *gain* tertinggi dari variabel-variabel yang ada sama seperti mencari variabel akar.
3. Memecah kasus dalam cabang.
   Membagi kelas di dalam cabang dengan cara yaitu memilih kelas yang terbanyak jika cabang memiliki dua kelas atau lebih maka itu yang akan dipilih.

4. Proses untuk mengulang dilakukan pada setiap cabang diperlukan agar setiap cabang hingga semua kasus mencapai kelas *(class)* masing-masing.

Pada Persamaan 2 berikut ini menunjukkan cara yang digunakan untuk menghitung *gain* dalam memilih atribut yang akan menjadi *root* (Hijriana, Nadiya, 2016).

$$\text{Gain (S,A)} = \textit{Entropy}\ (\text{S}) - \sum_{i=0}^{n} \frac{|\text{Si}|}{|\text{S}|} \times \text{Entropy (Si)} \text{ ..... (2)}$$

Berikut ini keterangan untuk Persamaan 2 yaitu (Hijriana, Nadiya, 2016):

S : Himpunan kasus
A : Atribut
n : Jumlah partisi atribut A.
$|S_i|$: Jumlah kasus pada partisi ke-i
$|S|$: Jumlah kasus keseluruhan

Pada Persamaan 2 sebelum didapatkan nilai *gain*, terlebih dahulu dicari nilai dari *entropy*. *Entropy* yaitu penjelasan tentang proporsi dalam pembagian kelas, untuk nilai *entropy* dimulai dari 0 hingga 1. Diketahui jika nilai *entropy* sama dengan 0, lalu menyatakan jumlah *sample* ada di salah satu kelas. Sementara itu jika nilai *entropy* sama dengan 1 menunjukkan nilai *sample* ada pada tiap-tiap kelas dan berjumlah sama.

Pada Persamaan 3 ditunjukkan perhitungan yang digunakan untuk mencari nilai *entropy* (Hijriana, Nadiya, 2016)*:*

$$Entropy\ (S) = \sum_{i=0}^{n} - \text{Pi} \times log\, 2\ \text{Pi} \text{ .....(3)}$$

keterangan pada Persamaan 3 yaitu (Hijriana, Nadiya, 2016):

S : Himpunan kasus
n : Jumlah partisi dalam S
P_i : Proposisi dari Si terhadap S

2.6 *K-Fold Cross Validation*

K-fold cross validation merupakan salah satu teknik menilai kinerja pada proses mengklasifikasi. Pada *k-fold cross validation* biasanya melakukan proses klasifikasi dengan memerlukan *dataset* yang sama. Kemudian menimbulkan *k* perpecahan dari sekumpulan data yang berubah menjadi *nonoverlapping* dengan proporsi *(k-1)/k* dan pada pengujian yaitu *1/k*.

Untuk mengukur model yang dianjurkan maka dilakukan teknik validasi 10-*fold cross validation* pada *dataset* yang sudah dikumpulkan *dataset* pertama yang akan dibagi (*split*) membentuk data latih (*training data*) dan data uji (*testing data*) dengan menggunakan 10 *fold cross validation*. Untuk data latih dipakai untuk membuat algoritme klasifikasi, lalu data uji akan diperlukan untuk menguji algoritme atau model yang telah siap. Selanjutnya *confusion matrix* dipakai untuk membuat ukuran pada kinerja algoritme yang dianjurkan. Kemudian setelah selesai semua algoritme atau model diuji, kemampuannya dibandingkan untuk melihat model mana yang baik untuk digunakan (Saifudin, 2018).

Pada proses iterasi menggunakan teknik *K-fold cross validation* dapat melakukan sebanyak 10 kali iterasi dan pada tahap S1 hingga S10 akan menjadikan *data testing* sejumlah 1 kali. Lalu untuk *K* yang nilainya 10 maka di dalam proses iterasi menghasilkan *decision tree* berjumlah 10. Sebelum memilih model *decision tree* yang akan digunakan pada proses *training* harus memenuhi syarat-syarat sebagai berikut (Saifudin, 2018):

1. Model *decision tree* yang dihasilkan untuk jumlah aturan (*rule*) yang paling banyak maka akan bertambah banyak jumlah aturan yang didapat dan pada pengerjaan data akan lebih bermacam-macam. Jika didapati sejumlah iterasi yang dihasilkan *rule* yang sama, maka *rule* tidak dapat digunakan. Ini karena hasil *rule* tersebut mengatasi data yang sama. sehingga data tidak *variatif*.
2. Model *decision tree* harus mempunyai tingkat ketepatan akurasi yang tertinggi.

3. Pada setiap model harus meliputi semua kelas target yang memungkinkan muncul dalam *test set.*

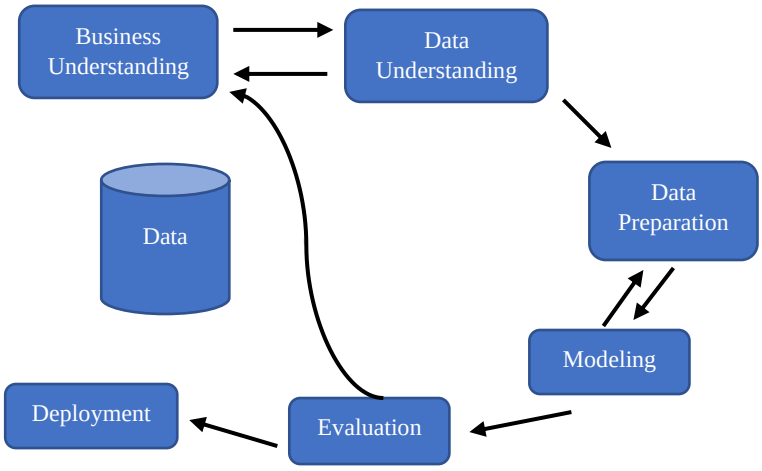
2.7 CRISP-DM

Cross industry standard process for data mining (CRISP-DM) adalah salah satu metode pada *data mining* yang sudah dikembangkan sejak tahun 1966. Biasanya pada metode ini menunjukkan proses analisis strategi untuk memecahkan permasalahan secara umum untuk bisnis dan bisa juga untuk penelitian (Fadillah, 2015).

Di tahun 1997 CRISP-DM dijadikan suatu proyek oleh negara Uni Eropa, untuk pendanaan ESPRIT yang dikepalai atas lima perusahaan: Integral Solutions Ltd (ISL), Teradata, Daimler AG, NCR Corporation dan OHRA (Fadillah, 2015). Gambar 2.4 menerangkan metode yang ada di dalam CRISP-DM terdapat enam tahapan yaitu (Fadillah, 2015):

1. *Business Understanding Phase* (Fase Pemahaman Bisnis)
2. *Data Understanding Phase* (Fase Pemahaman Data)
3. *Data Preparation* (Persiapan Data)
4. *Modeling Phase* (Fase Pemodelan)
5. *Evaluation Phase* (Fase Evaluasi)
6. *Deployment Phase* (Fase Penyebaran).

Berikut ini Gambar 2.4 menunjukkan fase yang terdapat pada metode CRISP- DM.



Gambar 2.4. Tahapan Metode CRISP-DM (Watson, 2000).

Penjelasan Gambar 2.4 tentang fase dalam metode CRISP-DM sebagai berikut (Novandya dkk., 2017):

1. Fase Pemahaman Bisnis (*Business Understanding Phase*)
   Fase ini dilakukan untuk mengerti tujuan dari sudut pandang bisnis, yang selanjutnya membuat pengetahuan untuk persiapan awal dalam permasalahan tersebut.
2. Fase Pemahaman Data (*Data Understanding Phase*)
   Fase ini membuat data awal untuk mendapati kualitas dari data, lalu menghasilkan pengetahuan pertama yang terdapat di dalam data, dan juga untuk menemukan permasalahan menarik agar dapat dibuat hipotesis dengan informasi yang tidak diketahui sebelumnya.
3. Persiapan Data (*Data Preparation*)
   Tahapan melakukan persiapan data yang meliputi semua proses untuk membuat *dataset* akhir yang berawal dari data yang masih belum tidak beraturan.
4. Fase Pemodelan (*Modeling Phase*)
   Fase ini dilakukan teknik pemodelan yang telah dipilih untuk diterapkan.
5. Fase Evaluasi (*Evaluation Phase*)
   Pada fase ini pengujian terhadap model yang sudah dibentuk untuk meyakinkan bahwa data yang memiliki informasi sudah tepat dengan

permasalahan bisnis

6. Fase Penyebaran (*Deployment Phase*)

   Fase penyebaran memerlukan pola yang akan menghasilkan fase dari pemodelan yang disajikan ke dalam bentuk seperti laporan tentang implementasi teknik *data mining* hasil analisis pola yang didapatkan.

2.8 *RapidMiner*

RapidMiner adalah salah satu *tools* atau perangkat lunak yang biasa digunakan dalam menyelesaikan proses analisis bersifat terbuka (*open source*) dan mudah untuk diterapkan. *RapidMiner* merupakan suatu jalan keluar dalam melakukan analisis pada *data mining*, *text mining*, dan analisis prediksi. *RapidMiner* memudahkan pengguna karena menggunakan berbagai teknik di dalamnya seperti teknik deskriptif dan prediksi untuk memberikan pengetahuan bagi penggunanya sehingga dapat menghasilkan keputusan yang diinginkan. Di dalam *RapidMiner* terdapat kurang lebih 500 operator *data mining*, salah satunya operator untuk *input, output, data preprocessing*, dan visualisasi data (Ratniasih, 2019).

RapidMiner pada umumnya digunakan sebagai *tools* untuk analisis data misalnya pada penggunaan *RapidMiner* untuk pengolahan *data mining* dan juga untuk *text mining*. Pada *text mining RapidMiner* digunakan untuk analisis teks, mengekstrak pola-pola dari *dataset* yang berukuran besar, dan menggabungkan dengan metode statistika, kecerdasan buatan, dan *database* (Ratniasih, 2019). Berikut ini Gambar 2.5 menunjukkan tampilan utama dari RapidMiner.

Gambar 2.5. Tampilan Utama *RapidMiner*

Pada Gambar 2.5 yang merupakan tampilan utama dari *RapidMiner* yang merupakan *software data mining*. Sebelum bernama *RapidMiner* pada awalnya bernama YALE (*Yet Another Learning Environment*). Berawal dari versi yang dikembangkan oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer pada tahun 2001 di *Artifical Intelligence Unit* dari *University of Dortmund* (Dennis dkk., 2013). *RapidMiner* menyimpan prosedur dari *data mining* dan *machine learning* di dalamnya termasuk juga ETL (*extraction transformation loading*), *data preprocessing*, visualisasi pemodelan, dan evaluasi (Ratniasih, 2019).

Di dalam *RapidMiner* terdapat GUI (*graphic user interface*) yang dapat digunakan untuk membuat *pipeline analitis*. GUI dapat menghasilkan *file* XML (*extensible markup language*) digunakan untuk proses analisis dan diterapkan menjadi data. Lalu *file* XML dibaca oleh *RapidMiner* kemudian melaksanakan analisis secara otomatis. Adapun beberapa *fitur* yang disediakan oleh RapidMiner yaitu (Ratniasih, 2019):

- Terdapat banyak algoritme yang mendukung proses *data mining* seperti *decision tree*, *self-organization map* dan lainnya.
- *RapidMiner* menyediakan bentuk grafis yang canggih, seperti *histogram*, *tree chart*, dan *3D scatter plots.*
- Terdapat juga bermacam-macam variasi *plugin*, seperti *text plugin* yang

digunakan untuk *analisis text*.

- Adanya prosedur dari *data mining* dan *machine learning* dan termasuk juga ETL (*extraction transformation loading*), *data preprocessing*, visualisasi pemodelan dan evaluasi.
- Pada proses *data mining* terdiri atas operator yang *nestable*, dijelaskan dengan XML, dan terbuat dari GUI.
- Dapat digabungkan dengan proyek *data mining* Weka dan R.

2.9 Seleksi Jalur Masuk

Perguruan tinggi adalah tahapan pendidikan formal yang dilaksanakan setelah menyelesaikan sekolah menengah atas (SMA). Saat lulus para siswa SMA bersaing agar dapat diterima di perguruan tinggi yang diinginkan. Terdapat banyak cara agar para siswa diterima di perguruan tinggi.

Seleksi masuk merupakan salah satu bentuk penerimaan mahasiswa baru untuk diterima di perguruan tinggi negeri. Seleksi masuk diselenggarakan secara serentak di seluruh Indonesia. Seleksi masuk perguruan tinggi negeri (SNMPTN) atau biasa dikenal dengan seleksi jalur undangan. Seleksi jalur undangan adalah seleksi yang berdasarkan pada evaluasi hasil penilaian prestasi akademik siswa yang meliputi nilai rapor, nilai hasil ujian nasional (UN), dan didukung juga dengan prestasi selama menempuh pendidikan sekolah menengah atas. Pada jalur SNMPTN undangan terdapat kuota penerimaan mahasiswa baru yang akan diterima yaitu sekitar 50% dari jumlah total seluruhnya (Nurhasanah dkk, 2015).

Jalur yang selanjutnya yaitu jalur seleksi tertulis atau biasa dikenal dengan seleksi bersama masuk perguruan tinggi negeri (SBMPTN). Sebelumnya SBMPTN bernama SNMPTN-tertulis, dan kemudian diubah menjadi istilah SBMPTN untuk seleksi masuk ke perguruan tinggi dengan tes tertulis. Seleksi pada jalur SBMPTN berdasarkan pada evaluasi hasil tes tertulis yang dilakukan para calon mahasiswa secara langsung.

SBMPTN tes tertulis dilaksanakan secara langsung oleh selaku calon peserta tes para calon mahasiswa menggunakan perlengkapan tes seperti soal yang harus dikerjakan dengan batas waktu yang telah ditentukan. Tes dilaksanakan sebagai bentuk untuk mengukur nilai kompetensi yang didasari sebagai pertimbangan hasil tes tertulis secara langsung, dan menentukan keputusan seorang calon mahasiswa baru untuk diterima melalui jalur SBMPTN. Adapun kuota mahasiswa baru yang akan diterima melalui jalur SBMPTN yaitu sekitar 35% dari total keseluruhannya. Presentase yang tersisa dilaksanakan untuk jalur seleksi tes mandiri yang diadakan oleh masing-masing perguruan tinggi (Pratiwi dkk, 2015).

Menurut (Amirulla, 2013) menjelaskan bahwa adanya alasan utama sebab perguruan tinggi melangsungkan seleksi dalam penerimaan mahasiswa baru, yaitu:

- Pendidikan di perguruan tinggi adalah tempat untuk mempersiapkan calon pemimpin bangsa di waktu yang akan datang, maka dari itu sangat diperlukan suatu cara bahwa calon mahasiswa yang akan menempuh pendidikan di perguruan tinggi mempunyai kualitas yang baik.
- Waktu kesempatan untuk menuntut ilmu di perguruan tinggi adalah kesempatan yang tidak mudah didapatkan, khususnya bagi negara berkembang seperti di Indonesia, menjadikan perguruan tinggi berharap kesempatan yang tidak mudah itu diberikan untuk calon mahasiswa yang memiliki potensial dan berhak mendapatkannya.
- Dengan mengadakan seleksi mengharuskan kandidat mahasiswa memiliki suatu kemampuan yang tinggi.
- Berkesempatan menempuh pendidikan tinggi adalah suatu hal yang sangat berharga, sehingga harus digunakan secara baik efektif dan efisien bagi para calon mahasiswa.

Berdasarkan hal-hal tersebut maka diperlukan suatu sistem yang berkualitas dalam penerimaan calon mahasiswa baru di perguruan tinggi, kemudian para mahasiswa yang mendapatkan kesempatan belajar di perguruan tinggi adalah

mereka yang memiliki potensial akademik dan berhak untuk mendapatkannya dalam hal dalam penerimaan mahasiswa baru dengan menilai dari beberapa aspek yaitu: efektivitas prediksi, efektivitas ekonomi, ekuitas, dan intensitas belajar mengajar (Nurhasanah dkk, 2015).

2.10 Lama Masa Studi

Pendidikan merupakan proses pembelajaran pengetahuan dan keterampilan yang dilakukan oleh masyakarat untuk berkembang. Pendidikan dapat diartikan sebagai pengalaman dalam pembelajaran seseorang di sepanjang hidupnya dan terjadi tidak ada batas usia melainkan terjadi selama hidup (Darmawan dkk, 2017).

Pendidikan yang ada di Indonesia terdiri dari beberapa jenjang. Saat ini perguruan tinggi menjadi syarat dasar dalam pekerjaan, menjadikan perguruan tinggi akan membuat para calon sarjana yang pandai, dan memiliki kemampuan di bidangnya (Darmawan dkk, 2017)

Di dalam kamus besar Bahasa Indonesia (KBBI) istilah sarjana berawal dari bahasa sangkakerta yang artinya adalah sebuah gelar akademik dan diberikan untuk lulusan program pendidikan strata satu (S1). Menurut (Standar Nasional Pendidikan Tinggi Nomor 44 Tahun 2015) pada program pendidikan strata satu untuk menyelesaikan studi telah ditetapkan harus telah menyelesaikan minimal 144 SKS dan setinggi-tingginya 160 SKS dengan waktu lama studi maksimal 6 tahun untuk program sarjana strata satu (S1) dan program diploma empat (D4) atau sarjana terapan. Waktu yang diperlukan untuk mendapatkan gelar sarjana disebut dengan lama masa studi. Lama masa studi terhitung sejak awal seorang mahasiswa menempuh pendidikan hingga dinyatakan lulus, dan apabila lama masa studi mahasiswa terhitung lebih dari 7 tahun mahasiswa tersebut akan mendapati *drop out* (DO) dari perguruan tinggi (Suwena, 2017).

2.11 Kelulusan Mahasiswa

Kelulusan mahasiswa merupakan sebagai tanda telah berakhirnya seorang mahasiswa menyelesaikan pendidikan di perguruan tinggi. Adapun kelulusan mahasiswa merupakan hasil akhir yang penuh pencapaian dan membanggakan dalam menempuh suatu pendidikan di perguruan tinggi pada jenjang tertentu. Bagi mahasiswa strata satu (S1) diharuskan untuk menempuh beban studi yaitu minimal 144 SKS dengan batas waktu kurang dari 8 semester dan paling lama yaitu 14 semester (Risnawati, 2018).

III. METODE PENELITIAN

3.1 Waktu dan Tempat

Kegiatan penelitian dan pembuatan tugas akhir ini waktu dan tempatnya sudah dilaksanakan mulai dari bulan Oktober 2020 hingga Juni 2021 bertempat di Jurusan Ilmu Komputer Universitas Lampung. Berikut ini adalah jadwal kegiatan penelitian yang dilakukan pada penelitian dapat dilihat pada Tabel 3.1.

Tabel 3.1 Jadwal Penelitian

No.	**Aktifitas**	**Okt**	**Nov**	**Des**	**Jan**	**Feb**	**Mar**	**Apr**	**Mei**	**Jun**
1	Studi Literatur	√	√	√	√	√				
2	Fase Pemahaman Bisnis (*Business Understanding Phase*)	√	√	√	√	√				
3	Fase Pemahaman Data (*Data Understanding Phase*)	√	√	√	√	√	√			
4	Seminar usul				√					
5	Persiapan Data (*Data Preparation)*		√	√	√	√				
6	Fase Pemodelan (*Modeling Phase*)		√	√	√	√				

No	Aktivitas	Okt	Nov	Des	Jan	Feb	Mar	Apr	Mei	Jun
7	Evaluasi (*Evaluation*)				√	√	√			
8	Penyebaran (*Deployment*)				√	√	√			
9	Seminar Hasil							√		
10	Perbaikan							√	√	
11	Ujian Komprehensif									√

3.2 Bahan dan Alat

Berikut ini terdapat beberapa peralatan atau perangkat yang akan digunakan selama penelitian ini yaitu:

1. Satu unit *Personal Computer* (PC) ASUS X454Y dengan spesifikasi:
   - *Processor* : AMD Quad Core A8-7410
   - Video Grafik : Radeon Graphics
   - RAM : 4GB
   - Sistem Operasi : Windows 10

2. Perangkat lunak (*Software*) pendukung selama penelitian yaitu:
   - *RapidMiner* ver.7.6 : *software* untuk pengolahan *data mining*
   - *Microsoft Excel* : *software* untuk *preprocessing* data

3.3 Tahapan Penelitian

Pada penelitian ini terdapat dua tahapan utama yaitu studi literatur dan tahapan penelitian menggunakan metode *Cross Industry Standard Process Data Mining* (CRSIP-DM). Berikut ini akan dijelaskan kedua tahapan penelitian yang akan dilakukan.

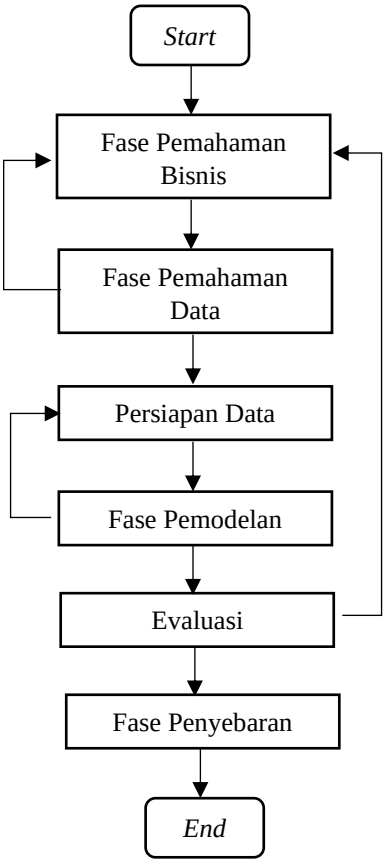
3.3.1 Studi Literatur

Tahap studi literatur adalah tahap pembelajaran konsep tentang pengolahan data menggunakan sistem yang terkomputerisasi. Studi literatur diperoleh dengan cara mengumpulkan jurnal, penelusuran internet, buku-buku yang berkaitan dengan topik penelitian yang sudah pernah dibuat sebelumnya (Asroni dkk, 2018).

3.3.2 Tahapan Penelitian dengan Metode CRISP-DM

Alur pada penelitian ini disesuaikan dengan tahapan pada metode CRISP-DM. Metode CRISP-DM adalah standardisasi *data mining* yang disusun oleh tiga penggagas *data mining market* yaitu Daimler Chrysler (Daimler-Benz), SPSS (ISL), dan NCR. Kemudian metode ini dikembangkan pada berbagai *workshops* antara tahun 1997 sampai dengan 1999. Terdapat lebih dari 300 organisasi berkontribusi dalam proses *modelling* CRISP-DM dan akhirnya CRISP-DM 1.0 dipublikasikan pada tahun 1999 (Fadillah, 2015).

Penggunaan metodologi CRISP-DM pada *data mining* sudah banyak diterapkan dalam penelitian, analisis terhadap kualitas mahasiswa lulusan Jurusan Ilmu Komputer perlu dilakukan agar mengetahui pola yang dihasilkan dari algoritme C4.5. Metode CRISP-DM cocok digunakan untuk tahap penelitian ini. CRISP-DM menyediakan standar proses untuk *data mining* yang diterapkan ke dalam strategi untuk pemecahan masalah bisnis atau pada penelitian. Selain penggunaan metode yang tepat, pemilihan algoritme yang tepat adalah hal yang penting dilakukan dalam *data mining*. Pemilihan algoritme C4.5 sebagai algoritme untuk mempermudah pembentukan pohon keputusan didasari oleh survei yang dipublikasikan Springer "*Top 10 Algorithm in Data Mining*". Algoritme C4.5 ditempatkan pada posisi pertama sebagai algoritme paling banyak digunakan dalam *data mining* (Fadillah, 2015). Pada Gambar 3.1 yang menunjukan tahapan dari metode CRISP-DM yang digunakan dalam penelitian ini.



Gambar 3.1 Tahapan Metode CRISP-DM yang Digunakan Dalam Penelitian

Gambar 3.1 menunjukkan adanya enam tahapan kerja dari metode *Cross Industry Standard Process Data Mining* (CRSIP-DM) yang terdiri dari fase pemahaman bisnis (*business understanding phase*), fase pemahaman data (*data understanding phase*), fase persiapan data (*data preparation phase*), fase pemodelan (*modelling phase*), fase evaluasi (*evaluation*), dan penyebaran (*deployment*) (Watson, 2000). Berikut ini dijelaskan tentang enam tahapan kerja dari metode *Cross Industry Standard Process Data Mining* (CRSIP-DM).

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

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

- Menentukan tujuan bisnis

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

- Menilai situasi

  Salah satu tantangan untuk meningkatkan akreditasi pada suatu jurusan di perguruan tinggi adalah memperbaiki kualitas mahasiswanya agar mudah diterima dalam dunia kerja. Selain itu, kualitas mahasiswa pada setiap jurusan di dalam perguruan tinggi dapat mempengaruhi Angka Efisiensi Edukasi (AEE). Nilai AEE yang rendah berpengaruh dalam penilaian akreditasi. 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 untuk mengambil keputusan atau membuat kebijakan untuk meningkatkan kualitas mahasiswa di Jurusan Ilmu Komputer Universitas Lampung.

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

Pada tahapan pemahaman data ada hal yang harus dilakukan dalam penelitian yaitu, mengumpulkan data awal, mendeskripsikan data, dan mengeksplorasi data.

- Mengumpulkan data

  Data yang digunakan merupakan data yang didapatkan dari Jurusan Ilmu Komputer dengan mengakses Sistem Informasi Akademik UNILA (SIAKADU) untuk mendapatkan data kelulusan meliputi Indeks Prestasi Kumulatif (IPK), waktu studi, lama studi, waktu kelulusan, dan data jalur masuk dari data Penerimaan Mahasiswa Baru (PMB) Jurusan Ilmu Komputer Universitas Lampung.
- Mendeskripsikan data

  Pada Tabel 3.1 berikut ini ditunjukkan deskripsi dari atribut data mahasiswa yang akan digunakan dalam penelitian.

Tabel 3.2 Deskripsi Atribut Data

Atribut	**Deskripsi**
ipk	Indeks Prestasi Kumulatif (IPK) adalah ukuran kemampuan mahasiswa sampai pada waktu tertentu yang dihitung dari jumlah Satuan Kredit Semester (SKS) lalu dikalikan dengan nilai bobot mata kuliah dan dibagi dengan jumlah seluruh SKS.
waktu studi dan lama skripsi	Waktu studi seorang mahasiswa yang dihitung dari mulai terdaftar sebagai mahasiswa sampai dengan lulus.
jalur Masuk	Jalur masuk adalah proses masuk yang diikuti seluruh mahasiswa.

C. Fase Persiapan Data *(Data Preparation Phase)*

Tahapan ini hal yang harus dilakukan adalah mendeskripsikan *dataset*, selanjutnya memilih data, membangun data, mengintregrasikan data, dan membersihkan data mahasiswa yang akan digunakan dalam penelitian ini proses ini dilakukan agar kualitas data yang diperoleh mudah dipahami. Maka dari itu dilakukan dengan cara sebagai berikut:

- Pembersihan data *(data cleaning)*

  Proses ini bertujuan untuk mengidentifikasi dan menghapus data yang tidak terstruktur (*noise*) dan data yang tidak lengkap (*missing value*). Pembersihan data yang dilakukan terdiri dari data mahasiswa lulus dengan jalur masuk SNMPTN, SBMPTN, Ujian Mandiri, PMPAP, dan D3.
- Integrasi data (*data integration*)

  Proses ini dilakukan untuk menggabungkan data dari beberapa data menjadi data baru. Misalnya proses menggabungkan data akademik alumni Jurusan Ilmu Komputer Universitas Lampung dengan data jalur masuk penerimaan mahasiswa baru akan menghasilkan data baru.
- Seleksi data (*data selection*)

  Proses untuk memilah data yang akan digunakan dalam klasifikasi. Pada tahapan ini dilakukan seleksi pada data *database* menggunakan excel. Dilakukannya tahapan ini karena data yang diterima tidak semuanya digunakan, data dipilih sesuai dengan atribut atau variabel yang diperlukan dalam penelitian dengan melakukan seleksi data menjadi *dataset* baru. Contoh dari atribut yang dipilih untuk digunakan pada penelitian ini adalah IPK, lama studi, lama skripsi, dan jalur masuk (SNMPTN dan SBMPTN).
- Data transformasi (*data transformation*)

  Proses untuk mengubah data ke format yang sesuai untuk diproses dengan menggunakan teknik *data mining.* Data yang digunakan dalam penelitian ini bernilai kategori yaitu merupakan data yang variabelnya telah dikelompokkan menjadi beberapa kelompok untuk model klasifikasi menggunakan *software RapidMiner*. Data penelitian yang didapatkan setelah dilakukan pembersihan dan seleksi data terdiri dari empat variabel yaitu data IPK, jalur masuk, waktu studi, dan lama studi.

D. Pemodelan (*Modelling Phase*)

Pada proses pemodelan yang menggunakan algoritme C4.5 untuk melakukan pengukuran akurasi dan juga menggunakan *software*

RapidMiner dan akan menghasilkan *decision tree* dari hasil analisis yang dilakukan. Algoritme C4.5 akan menghasilkan pohon keputusan dengan menentukan atribut mana yang akan menjadi akar, atribut dengan nilai *gain* tertinggi akan dipilih sebagai *node* pertama pada pohon keputusan dan *node* selanjutnya akan diisi oleh atribut-atribut yang bernilai *gain* lebih rendah dan begitu seterusnya. *Node* yang tidak memiliki percabangan akan menunjukkan *output* dari setiap cabang yang biasa disebut dengan *leaf* atau daun. Setelah dilakukan fase persiapan data, pada tahapan ini merupakan pemodelan klasifikasi guna menemukan kumpulan pola atau fungsi yang mendeskripsikan serta memisahkan kelas data yang satu dengan yang lainnya, untuk menyatakan objek tersebut masuk pada kategori tertentu. Pemodelan dilakukan menggunakan *dataset* mahasiswa hasil dari persiapan data. Pada pemodelan klasifikasi penelitian ini menggunakan teknik *decision tree* dan algoritme C4.5 dengan *tools* pada *software RapidMiner*. Di dalam *RapidMiner* terdapat operator *cross validation* yang merupakan operator *data mining* untuk memperkirakan keakuratan dari model *decision tree* yang dibangun dengan melakukan validasi silang pada *dataset*. Operator *cross validation* memiliki subproses percobaan (*training subproses*) dan subproses pengujian (*testing subproses*). Pemodelan *decision tree* dan algoritme C4.5 dibangun dengan cara perhitungan nilai *entropy* dan nilai *gain* guna memilih akar (*root*) dan klasifikasi yang akan diperoleh dengan *decision tree* menghasilkan aturan atau *rule*. Kemudian *decision tree* yang dihasilkan akan diuji untuk mengetahui nilai dari tingkat akurasi.

E. Evaluasi *(evaluation)*

Pada tahap ini dilakukan pengukuran dan evaluasi guna menilai keefektifan dan mengukur keakuratan hasil dari model yang terbentuk sebelum digunakan. Akurasi adalah persentase dari total data uji coba yang dinyatakan benar ketika dilakukan proes identifikasi. Selain itu, pada tahap ini juga diamati apakah hasil sudah sesuai dengan tujuan bisnis yang telah ditetapkan di fase pemahaman bisnis. Pada fase evaluasi akurasi klasifikasi

didapatkan dengan menggunakan metode *k-fold cross validation* untuk memastikan dan memvalidasi tingkat akurasi model yang dibangun. Nilai persentase akurasi yang didapatkan dari metode *k-fold cross validation* ini adalah dengan memilih beberapa atribut dari dataset untuk melakukan validasi silang.

F. *Deployment*

Pada tahap terakhir dalam metode CRISP-DM pengetahuan yang dihasilkan akan ditampilkan dalam bentuk laporan agar dapat dipahami hasil dari pemodelan menggunakan teknik *data mining* dengan algoritme C4.5

Hasil analisis pola *decision tree* yang didapatkan dari data pemodelan kualitas mahasiswa lulusan Jurusan Ilmu Komputer Universitas Lampung diharapkan bermanfaat dan dapat digunakan sebagai saran dalam menentukan langkah untuk mengatasi kelulusan mahasiswa berdasarkan jalur masuk dan meningkatkan kualitas penerimaan mahasiswa baru pada jalur SNMPTN dan SBMPTN tahun penerimaan yang akan datang.

V. SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil klasifikasi kualitas mahasiswa dapat diketahui bahwa lama studi menjadi atribut yang berpengaruh besar pada proses klasifikasi. Mahasiswa jalur masuk SBMPTN lebih unggul kualitasnya dengan asal daerah Kota Bandar Lampung karena memiliki jumlah lulusan tertinggi di setiap tahunnya dibandingkan dengan mahasiswa jalur masuk SNMPTN. Hasil evaluasi kinerja dari *decision tree* dengan algoritme C4.5 menggunakan metode *k-fold cross validation* dengan melakukan validasi silang yang memaksimalkan akurasi dari keseluruhan pohon mendapatkan nilai sebesar 99,03% dengan *error rate* sebesar 0,97%. Akurasi model yang dihasilkan termasuk ke dalam *excellent classification* sehingga dapat digunakan sebagai acuan prediksi yang sangat baik.

5.2 Saran

Berikut ini adalah beberapa saran untuk perkembangan selanjutnya:

1. Diharapkan pada penelitian selanjutnya dapat menggunakan *software data mining* lainnya dan menggunakan algoritme klasifikasi lain untuk menentukan hasil yang lebik baik lagi.
2. Diharapkan pada penelitian selanjutnya dapat memasukan semua jurusan yang ada serta jalur penerimaan lainnya dan menggunakan data mahasiswa yang lebih banyak lagi.
3. Diharapkan adanya penelitian lanjutan mengenai pengklasifikasian jenis pekerjaan dari alumni yang telah lulus.

DAFTAR PUSTAKA

BAN-PT. 2011. *Buku 1 Naskah Akademik Akreditasi Institusi Perguruan Tinggi*. Badan Akreditasi Nas. Perguru. Tinggi. 16 hlm.

Universitas Lampung. 2014. Buku Akademik Universitas Lampung. Universitas Lampung. 91 hlm.

Prasetyo, E. 2012. *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset. 356 hlm.

Han, J. and Kamber M. (2006). *Data Mining, Concepts and Techniques*. Simon Fraser University: Morgan Kaufmann Publisher. 772 hlm.

The Data Warehousing Institute. 2000. *Jurnal Of Data Warehousing*. Vol. 5, No. 4.

Aprilia C, D., Aji Baskoro, D., Ambarwati, L., & Wicaksana, I. W. S. (2013). *Belajar Data Mining Dengan Rapid Minner.*

Asroni, A., Masajeng Respati, B., & Riyadi, S. (2018). *Penerapan Algoritma C4.5 untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta. Semesta Teknika, 21(2), 158–165.*

Cahyono, B., & Islamiyah. (2016). *Data Mining Untuk Klasifikasi Penentuan Peminatan Siswa Sma Negeri 2 Tenggarong Sebrang, dengan Menggunakan Algoritma C4.5.* Seminar Nasional APTIKOM, 28–29.

Darmawan, H. D., Yuniarti, D., & Nasution, Y. N. (2017). *Klasifikasi Lama Masa Studi Mahasiswa Menggunakan Perbandingan Metode Algoritma C.45 dan Regression Tree.* Jurnal Eksponensial, 8(2), 151–160.

Fadillah, A. P. (2015*). Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ).* 1, 260–270.

Hijriana, Nadiya, and R. M. (2016). *Penerapan Metode Decision Tree Algoritma C4. 5 Untuk Klasifikasi Mahasiswa Berprestasi. Al Ulum Jurnal Sains Dan Teknologi,* 2(1), 39–43.

Kane, S. N., Mishra, A., & Dutta, A. K. (2016). *Preface: International Conference on Recent Trends in Physics (ICRTP 2016).* Journal of Physics: Conference Series, 755(1), 0–8.

Novandya, Adhika., Oktria, I. (2017). *Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi*. Jurnal Format, 6(2), 98–106.

Nurhasanah, Purwati, & Ahmad, H. (2015). *Pengaruh Sistem Seleksi Masuk Perguruan Tinggi Terhadap Indeks Prestasi Mahasiswa Jurusan Pendidikan Matematika Universitas Papua (UNIPA).* Prosiding Seminar Nasional, 03, 114–120.

Oktaria, R., Komarudin, M., & Muda, M. A. (2019*). Analisa Klasifikasi Kualitas Mahasiswa Lulusan Berdasarkan Jalur Penerimaan Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Lampung).* Jurnal Teknik Informatika, 12(2), 183–192.

Pratiwi, L. D. B., Wibowo, W., & Zain, I. (2015). *Klasifikasi Nilai Peminat SBMPTN (Seleksi Bersama Masuk Perguruan Tinggi Negeri) ITS dengan Pendekatan Classification and Regression Trees (CART).* Jurnal Sains Dan Seni ITS, 4(2), 2–7.

Ratniasih, N. L. (2019). *Optimasi Data Mining Menggunakan Algoritma Naïve Bayes Dan C4 . 5.* Jurnal Teknologi Informasi Dan Komputer, 5(1), 28–34.

Rismayanti. (2018). *Decision Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika (Studi Kasus : Fakultas Teknik dan Komputer Universitas Harapan Medan).* Jurnal Sistem Informasi, 02(01), 16–24.

Risnawati. (2018). *Analisis Kelulusan Mahasiswa Menggunakan Algoritma C.45.* Jurnal Mantik Penusa, 2(1), 71–76.

Saifudin, A. (2018). *Metode Data Mining Untuk Seleksi Calon Mahasiswa. 10(1), 25–36.*

Suwena, K. R. (2017). *Jalur Penerimaan Mahasiswa Baru Bukan Penentu Prestasi Belajar Mahasiswa. Ekuitas:* Jurnal Pendidikan Ekonomi, 5(2), 1–10.

Zega, S. A. (2014). *Penggunaan Pohon Keputusan untuk Klasifikasi Tingkat Kualitas Mahasiwa Berdasarkan Jalur Masuk Kuliah.* Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta, 7–13.

Ardiansyah, D., Walim. (2018*). Algoritma C4.5 Untuk Klasifikasi Calon Peserta Lomba Cerdas Cermat Siswi SMP Dengan Menggunakan Rapid Miner.* Jurnal Infokar, Vol 1 No.2, 1-8.

Nasrullah, H, A. (2018). Penerapan Algoritma C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop Out. Jurnal Ilmiah, Vol 10 No.2, 1-7.

Harryanto, F, F., Hansun, Seng. Penerapan Algoritma C4.5 Untuk Memprediksi Penerimaan Calon Pegawai Baru di PT. WISE, Jatisi, Vol 3 No,2, 1-9.