

### Klasifikasi Predikat Tingkat Kelulusan Mahasiswa Program Studi Teknik Informatika dengan Menggunakan Algoritma C4.5 (Studi Kasus: STMIK Rosma Karawang)

Karya Suhada<sup>a\*</sup>, Anggi Elanda<sup>b</sup>, Anwar Aziz<sup>c</sup>

<sup>a,b,c</sup>STMIK Rosma, Jl. Kertabumi No.62, Karawang 41311, Indonesia

<sup>a</sup>Corresponding author: karya@rosma.ac.id

---

#### Abstract

*The existence of technology will make it very easy for universities to produce information and facilitate all university activities related to data processing and report generation. The importance of predicting the graduation rate of students in higher education is to minimize the rate of students who do not graduate on time. STMIK Rosma is one of the universities that graduate students every year. The purpose of this study was to analyze and predict student graduation in the Informatics Engineering study program class of 2012 - 2015 with data mining techniques using RapidMiner software in the application of the C4.5 algorithm and using the Cross-Industry Standard Process for Data Mining (CRIPS-DM) research method. The results of this study provide recommendations in the form of strategic information sources that are useful for universities, in this case STMIK Rosma, especially the study program on how to predict student graduation so that it can reduce the percentage of students who do not graduate on time.*

*Keywords : Classification, C4.5 Algorithm, CRISP-DM, Rapid Miner, Student Graduation*

#### Abstrak

Adanya teknologi akan sangat mudah bagi perguruan tinggi untuk menghasilkan informasi dan memudahkan segala aktivitas perguruan tinggi terkait dengan pengolahan data serta pembuatan laporan. Pentingnya memprediksi tingkat kelulusan mahasiswa dalam perguruan tinggi untuk dapat meminimalisir tingkat mahasiswa yang tidak lulus tepat waktu. STMIK Rosma merupakan salah satu perguruan tinggi yang banyak meluluskan mahasiswa setiap tahunnya. Tujuan penelitian ini adalah menganalisis serta memprediksi kelulusan mahasiswa pada program studi Teknik Informatika angkatan 2012 - 2015 dengan teknik data mining menggunakan software RapidMiner dalam penerapan algoritma C4.5 dan menggunakan metode penelitian Cross Industry Standard for Data Mining (CRIPS-DM). Hasil dari penelitian ini memberikan rekomendasi berupa sumber informasi yang strategis berguna bagi perguruan tinggi dalam hal ini STMIK Rosma terutama program studi bagaimana memprediksi kelulusannya sehingga dapat mengurangi presentase mahasiswa yang tidak lulus tepat waktu.

**Kata Kunci :** Algoritma C4.5, CRISP-DM, Klasifikasi, Kelulusan Mahasiswa, Rapid Miner

---

#### 1. Pendahuluan

Perkembangan zaman pada saat ini semakin berkembang pesat terutama teknologi informasi. Dengan adanya teknologi informasi dapat memudahkan pertukaran informasi, sehingga tempat, waktu dan jarak tidak lagi menjadi kendala. Jika dilihat dari beberapa sektor terutama sektor pendidikan, dituntut untuk menyesuaikan dengan perubahan zaman saat ini. Dengan adanya teknologi ini akan sangat mudah bagi perguruan tinggi untuk menghasilkan informasi dan memudahkan segala aktivitas perguruan tinggi terkait dengan pengolahan data serta pembuatan laporan. Laporan bisa berupa informasi yang dapat digunakan untuk kepentingan manajerial untuk menentukan langkah langkah apa

saja yang harus dilakukan dimasa yang akan datang. Laporan bisa didapatkan dan dikumpulkan dalam kurun waktu tertentu sesuai dengan kebutuhan biasanya sebagai perbandingan yaitu setiap lima tahun kebelakang yang nantinya bisa disimpulkan oleh pihak yang berkepentingan dalam hal ini STMIK Rosma.

Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Rosma merupakan perguruan tinggi satu satunya di Karawang yang memiliki 4 program studi yang meluluskan banyak mahasiswa pada setiap tahunnya, salah satunya pada program studi Teknik Informatika. STMIK Rosma ini mengalami kesulitan dalam memprediksi kelulusan mahasiswa/I untuk itu dibutuhkannya suatu sistem yang dapat digunakan untuk memprediksi kelulusan mahasiswa. Jika kelulusan mahasiswa tidak diketahui maka pihak manajemen akan menerapkan suatu kebijakan bagaimana untuk meminimalisir jumlah mahasiswa yang tidak tepat waktu kelulusannya sesuai dengan masa studinya. Dengan adanya data mining ini memiliki tujuan adalah dapat memberikan informasi yang berguna bagi manajemen dan dapat membantu pihak tertentu yang membutuhkannya.

Adapun beberapa penelitian terdahulu yang melatar belakangi penulisan artikel ini adalah penelitian yang dilakukan oleh (Luvia et al., 2016) yang mengklasifikasi predikat keberhasilan mahasiswa di Amik Tunas Bangsa menggunakan Algoritma C4.5, (Sitorus et al., 2019) yang melakukan penerapan Klasifikasi C4.5 dalam meningkatkan sistem pembelajaran mahasiswa, (Fadillah, 2015) memprediksi kelulusan studi mahasiswa menempuh mata kuliah menggunakan metode CRISP-DM untuk (Studi Kasus Universitas XYZ), dan penelitian yang dilakukan oleh (Rahayu et al., 2021) yang mengklasifikasikan ketepatan waktu kelulusan mahasiswa dengan metode naïve bayes.

Berdasarkan uraian diatas maka penulis merasa tertarik untuk membuat sebuah analisis prediksi kelulusan mahasiswa dengan menggunakan Algoritma C4.5 pada akhirnya bisa dimanfaatkan oleh pihak kampus untuk mengatasi permasalahan yang ada sehingga dengan adanya informasi tersebut perguruan tinggi dapat bersaing dengan yang lainnya bahkan semakin maju serta diharapkan dapat menjadi sebuah informasi yang sangat penting sehingga bisa meningkatkan kualitas mahasiswa STMIK Rosma.

## 2. Tinjauan Pustaka

### 2.1. Klasifikasi

Klasifikasi adalah kegiatan dalam mengekstrak dan memprediksi kategori untuk masing-masing data. Klasifikasi melakukan pencarian model yang dapat membedakan kelas data dengan tujuan agar model dapat digunakan untuk memprediksi kelas dari suatu obyek yang kelasnya belum diketahui (Nasrullah, 2018). Sedangkan menurut (Sartika & Sensuse, 2017) klasifikasi merupakan salah satu teknik data mining yang digunakan untuk membangun suatu model dari sampel data yang belum sejenis. Karena menggunakan sekumpulan data untuk dianalisis terlebih dahulu dan pola yang didapatkan dari hasil analisis digunakan untuk pengklasifikasian data uji, teknik klasifikasi ini termasuk dalam *supervised learning*. Proses klasifikasi data terdiri dari pembelajaran dan klasifikasi.

### 2.2. Predikat Kelulusan

Perguruan tinggi memiliki kewajiban dalam mengontrol mahasiswanya untuk menghasilkan lulusan yang berkualitas. Lulusan yang berkualitas dapat dilihat dari indeks prestasi dan lamanya studi yang menghasilkan predikat kelulusan. Adanya beberapa faktor tertentu yang mempengaruhi predikat kelulusan (Imaslihkah et al., 2013). Menurut (Rusdiana & Rosmiati, 2016) penentuan predikan mahasiswa dapat dilakukan dengan mengelompokan data yang didapat berdasarkan jumlah beban studi yang telah ditempuh, IPK yang didapatkan mahasiswa, dan keterangan kelulusan skripsi. Predikat kelulusan dapat diterima oleh mahasiswa jika memenuhi syarat antara lain :

- a. Indeks prestasi kumulatif menunjukn paling sedikit yaitu nilai minimal yang telah ditetapkan oleh kampus.
- b. Jumlah SKS yang telah ditentukan.
- c. Lulus ujian skripsi.

Persyaratan diatas tentunya dapat berbeda disetiap perguruan tinggi. Tentunya mahasiswa dapat dinyatakan lulus apabila memenuhi syarat yang telah ditentukan oleh masing-masing perguruan tinggi tersebut.

### 2.3. Data Mining

Definisi data mining menurut (Permata et al., 2012) menyatakan bahwa data mining berasal dari berbagai ilmu pengetahuan, yang meliputi *machine learning* atau *pattern recognatio*, statistik atau kecerdasan buatan, dan sistem basis data. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Data mining merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Purnia & Warnilah, 2017).

#### 2.4. Algoritma C4.5

Algoritma C4.5 adalah metode pohon keputusan dimana fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Pada algoritma C4.5 pohon keputusan yang dibentuk berdasarkan kriteria-kriteria pembentuk keputusan. Dalam algoritma C4.5 ini terdapat 2 elemen yang harus dipahami yaitu *Entropy* dan *Gain* (Wajhillah & Yulianti, 2017).

Menurut (Honesqi, 2017) menyatakan bahwa algoritma C4.5 merupakan algoritma klasifikasi data bertipe pohon keputusan. Pohon keputusan algoritma C4.5 dibangun dengan beberapa tahapan atau langkah yaitu sebagai berikut :

- a. Pemilihan atribut sebagai akar.
- b. Membuat cabang untuk tiap-tiap nilai.
- c. Membagi kasus dalam cabang.

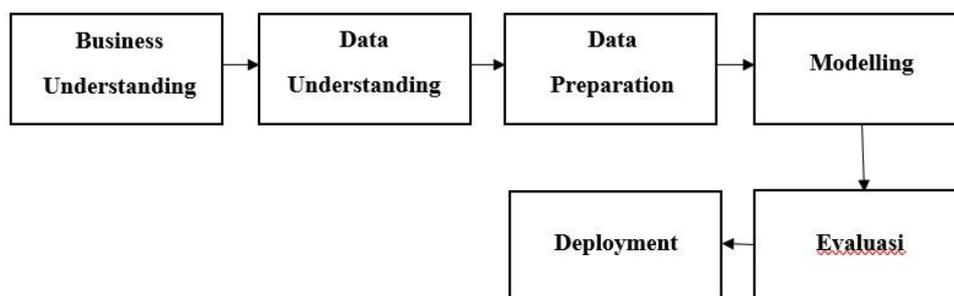
Tahap-tahap ini akan diulangi untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. Dari penyelesaian pohon keputusan maka akan didapatkan beberapa rule suatu kasus. Algoritma C45 memiliki kelebihan yaitu mudah dimengerti, fleksibel, dan menarik karena hasil yang didapatkan dapat divisualisasikan dalam bentuk gambar sebagai pohon keputusan (Honesqi, 2017).

#### 2.5. Rapid Miner

*Rapid Miner* adalah platform perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*), dan analisis prediktif (*predictive analytics*). Aplikasi ini digunakan untuk aplikasi bisnis dan komersial serta untuk penelitian, pendidikan, pelatihan, pembuatan prototype dengan cepat, dan pengembangan aplikasi serta mendukung semua langkah proses pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi dan pengoptimalan. *Rapid Miner* dikembangkan dengan model *open core* (Muhammad et al., 2018).

### 3. Metode

Penelitian ini menyelesaikan masalah dengan menggunakan pendekatan kuantitatif dengan jenis pendekatan eksperimen. Alur pada penelitian ini, disesuaikan dengan tahapan pada metode CRISP-DM. CRISP-DM merupakan metode yang menggunakan model proses pengembangan data yang banyak digunakan para ahli untuk memecahkan masalah. Proses penelitian ini mengacu pada enam tahap CRISP-DM yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi dan penyebaran (Astuti et al., 2019).



Gambar 1. tahap CRISP-DM

#### a. Business Understanding

*Business understanding* adalah awal dari proses CRISP-DM yang merupakan pemahaman terhadap tujuan bisnis, penelitian situasi dan menerjemahkan tujuan bisnis kedalam tujuan data mining. Dalam penelitian ini tentunya membutuhkan pengetahuan untuk mendapatkan data yang mendukung proses klasifikasi predikat kelulusan mahasiswa.

b. *Data Understanding*

*Data understanding* ini dilakukan proses pengumpulan data mahasiswa dan melakukan analisis data dan evaluasi kualitas data untuk mengenali data lebih lanjut dan mencari pengetahuan awal. Sumber data yang digunakan dalam penelitian ini adalah data mahasiswa program studi teknik informatika STMIK Rosma Karawang.

c. *Data Preparation*

*Data preparation* ini dilakukan proses penyusunan dataset final atau dataset yang nantinya akan dijadikan inputan kedalam *modelling data mining*. Tahapan dalam *data preparation* yaitu pemilihan data, transformasi data, dan pembersihan data.

d. *Modeling*

*Modeling* ini dilakukan sesuai teknik pemodelan yang dipilih yang nantinya akan diterapkan pada dataset yang sudah disiapkan untuk mengatasi kebutuhan bisnis yang sesuai. Teknik pemodelan dalam penelitian ini menggunakan teknik klasifikasi, dengan menggunakan algoritma C4.5. *Tools* yang digunakan dalam pemodelan ini yaitu RapidMiner.

e. Evaluasi

Evaluasi ini menguji model yang sudah dibuat dan dilakukan evaluasi keakuratan dan generalitasnya. Tahapan ini dilakukan untuk mengukur sejauh mana model yang telah dipilih memenuhi sasaran bisnis dan bila demikian, apakah perlu lebih banyak model untuk dibuat lagi.

f. *Deployment*

*Deployment* merupakan penyusunan laporan atau presentasi dari *modelling* serta *evaluation* pada proses data mining. Hasil penelitian ini bisa digunakan untuk informasi tingkat kelulusan mahasiswa sebagai bahan kajian bagi perguruan tinggi. Dengan mengetahui informasi tingkat kelulusan dapat memberikan kemudahan bagi staff dalam melakukan pengelolaan mahasiswa. Dan juga dapat memberikan informasi kepada mahasiswa agar dapat lulus dengan tepat waktu.

#### 4. Hasil dan Pembahasan

##### 4.1. *Business Understanding* (Pemahaman Bisnis)

Tahapan pemahaman bisnis ini merupakan pemahaman dari kegiatan penelitian data *mining* yang akan dilakukan. Pemahaman bisnis ini dimulai dari tujuan penelitian dan pemaparan masalah yang ditemukan. Pada tahapan ini terdapat empat fase yaitu *determine business objective*, *assess situation*, *determine data mining goals* dan *produce project plan*.

a. *Determine Business Objective*

Berdasarkan penjelasan yang sudah dijelaskan pada latar belakang sebelumnya, penelitian ini memiliki tujuan yaitu menganalisis prediksi terhadap kelulusan mahasiswa teknik informatika STMIK Rosma Karawang dengan menggunakan metode C4.5. Terdapat dua parameter dalam melakukan prediksi ketepatan kelulusan seorang mahasiswa yaitu lulus tepat waktu dan lulus terlambat. Hal ini didasari dari pengaruh kelulusan mahasiswa tersebut yang dimana dapat dipengaruhi dari jenis kelamin, asal daerah, SKS yang sudah ditempuh, IPK, predikat kelulusan.

b. *Assess Situation*

Seperti informasi yang diketahui bahwa mahasiswa teknik informatika STMIK Rosma Karawang pada angkatan tahun 2012, 2013, 2014, 2015 dan 2016 memiliki jumlah mahasiswa dengan lulus tepat waktu yang cukup kecil. Diketahui bahwa tahun 2012 dari total mahasiswa yang masuk pada tahun tersebut sebanyak 45 mahasiswa dan hanya 40% mahasiswa yang dinyatakan lulus tepat waktu. Pada tahun 2013 juga terdapat total mahasiswa yang masuk pada tahun tersebut sebanyak 63 mahasiswa dan hanya 39% mahasiswa yang dinyatakan lulus tepat waktu. Tahun 2014 terdapat total mahasiswa yang masuk pada tahun tersebut sebanyak 55 mahasiswa dan 55% mahasiswa yang dinyatakan lulus tepat waktu. Tahun 2015 terdapat 95 mahasiswa yang mendaftar dan hanya 23% mahasiswa yang lulus tepat waktu. Dan pada tahun 2016 terdapat mahasiswa yang lulus sebanyak 40% dari total mahasiswa sebanyak 61 mahasiswa.

c. *Determine Data Mining Goals*

Tujuan utama penelitian ini adalah untuk melakukan analisis prediksi kelulusan mahasiswa teknik informatika STMIK Rosma Karawang dengan menggunakan metode C4.5 dan mengevaluasi kinerja dari model. Setelah itu akan diperoleh pengetahuan baru dari data mahasiswa yang sudah diolah. Kemudian hasil dari prediksi yang telah dilakukan menggunakan metode C4.5 dapat dimanfaatkan pihak perguruan tinggi untuk meningkatkan kualitas mahasiswa teknik informatika STMIK Rosma Karawang.

d. *Produce Project Plan*

Rencana penelitian yang akan dilakukan meliputi dari pengumpulan data mahasiswa teknik informatika STMIK Rosma Karawang dari angkatan tahun 2012, 2013, 2014, 2015, hingga 2016 dengan atribut yang digunakan adalah jenis kelamin, asal daerah, SKS yang sudah ditempuh, IPK, predikat kelulusan. Dengan atribut yang sudah ada tersebut maka dilakukan pemodelan dengan menggunakan metode C4.5 dan memanfaatkan *cross validation* untuk mengukur validitas model klasifikasi dengan menggunakan *tools* RapidMiner.

4.2. *Data Understanding* (Pemahaman Data)

Pada tahapan ini dilakukan pengumpulan data akademik mahasiswa pada jurusan teknik informatika STMIK Rosma Karawang berupa nomor ijazah, NIM, nama, Jenis kelamin, asal daerah, tanggal lahir, jumlah SKS, IPK, predikat, dan lama studi. Kemudian data tersebut akan dipelajari dan dipahami untuk mengetahui data apa saja yang dapat dimanfaatkan dari data yang telah diperoleh dalam melakukan analisis prediksi kelulusan mahasiswa nantinya. Pada data *understanding* terdapat empat fase yaitu *collect initial data*, *describe data*, *Explore data* dan *verify quality*.

a. *Collect Initial Data*

Tahapan ini dilakukan pengumpulan data awal di mana data diperoleh dari tata usaha STMIK Rosma Karawang.

Tabel 1. Data Mahasiswa Teknik Informatika STMIK Rosma Karawang

NO	Nomor Seri Ijazah	NIM	Nama	Jenis Kelamin	Asal Daerah	Tanggal Lahir	Jumlah SKS	IPK	Predikat	Lama Studi
1	2014/I	201201	Mahasiswa01	Prempuan	Karawang	27 September 1990	145	2,83	Memuaskan	4
2	20413/I	201202	Mahasiswa02	Laki-Laki	Luar karawang	22 Oktober 1991	145	3,19	Sangat Memuaskan	4
3	20513/I	201203	Mahasiswa03	Laki-Laki	Karawang	13 Juni 1990	145	2,81	Memuaskan	4
4	20513/I	201204	mahasiswa04	Laki-Laki	Karawang	26 Februari 1995	145	2,97	Memuaskan	4
5	20613/I	201205	Mahasiswa05	Laki-Laki	Karawang	31 Desember 1993	145	2,98	Memuaskan	4
6	10013/I	201206	Mahasiswa06	Prempuan	Karawang	03 Maret 1992	145	3,67	Cumlaude	4
7	20013/I	201207	Mahasiswa07	Laki-Laki	Karawang	12 Juni 1993	145	2,80	Memuaskan	4
8	10013/I	201208	Mahasiswa08	Laki-Laki	Karawang	09 September 1984	145	3,57	Cumlaude	4
9	40013/I	201209	Mahasiswa09	Prempuan	Karawang	16 Januari 1996	145	3,03	Sangat Memuaskan	4
10	20013/I	201210	Mahasiswa10	Laki-Laki	Karawang	30 Maret 1996	145	3,63	Sangat Memuaskan	4
11	90013/I	201211	Mahasiswa11	Laki-Laki	Karawang	19 September 1996	145	2,98	Memuaskan	4
12	20013/I	201212	Mahasiswa12	Laki-Laki	Luar karawang	10 Maret 1995	145	3,17	Sangat Memuaskan	4
13	20213/I	201213	Mahasiswa13	Laki-Laki	Luar karawang	21 Juli 1993	145	3,18	Sangat Memuaskan	4
14	20113/I	201214	Mahasiswa14	Laki-Laki	Luar karawang	16 Mei 1996	145	2,88	Memuaskan	4
15	20113/I	201215	Mahasiswa15	Prempuan	Luar karawang	20 Februari 1993	145	2,88	Memuaskan	4
16	20113/I	201216	Mahasiswa16	Laki-Laki	Luar karawang	10 Juli 1991	145	3,01	Sangat Memuaskan	4
17	20113/I	201217	Mahasiswa17	Laki-Laki	Karawang	10 Januari 1998	145	3,01	Sangat Memuaskan	4
18	20113/I	201218	Mahasiswa18	Laki-Laki	Luar karawang	12 Januari 1997	145	3,50	Sangat Memuaskan	4
19	20113/I	201219	Mahasiswa19	Laki-Laki	Luar karawang	13 Januari 1997	145	3,60	Cumlaude	2
20	20113/I	201220	Mahasiswa20	Prempuan	Karawang	14 Januari 1997	145	3,12	Sangat Memuaskan	5

Tabel 1 merupakan data diri mahasiswa teknik informatika STMIK Rosma Karawang. Data yang diperoleh adalah data mahasiswa angkatan 2012, 2013, 2014, 2015, dan juga 2016 yang telah lulus. Dari data tersebut nantinya dapat ditentukan mana saja mahasiswa yang termasuk kedalam kategori mahasiswa dengan lulus tepat waktu dan terlambat. Jumlah data mahasiswa yang lulus diperoleh sebanyak 146 dengan 10 atribut.

a. *Describe Data*

Setelah didapatkan data awal, selanjutnya adalah mendeskripsikan data yang telah didapatkan sebelumnya. Tujuan ini dilakukan untuk mencari sedikit informasi mengenai atribut dari data yang

telah didapat dan memahami data yang akan diolah dalam melakukan prediksi ketepatan kelulusan mahasiswa. Dibawah ini pada tabel 2 merupakan deskripsi atribut semua data yang telah diperoleh sebelumnya.

Tabel 2. Deskripsi Atribut Data

No	Atribut	Jenis	Keterangan
1	Nomor Ijazah	Polynominal	Nomor ijazah mahasiswa
2	NIM	Polynominal	Nomor Induk Mahasiswa
3	Nama	Polynominal	Nama Mahasiswa
4	Jenis Kelamin	Binominal	Jenis Kelamin Mahasiswa
5	Asal Daerah	Binominal	Asal Daerah Mahasiswa
6	Tanggal Lahir	Polynominal	Tanggal Lahir Mahasiswa
7	Jumlah SKS	Integer	Jumlah SKS yang telah ditempuh
8	IPK	Polynominal	Indeks Prestasi Kumulatif (IPK) mahasiswa
9	Predikat	Polynominal	Predikat Mahasiswa
10	Lama Studi	Integer	Lamanya Studi Mahasiswa

b. *Explore Data*

Setelah dilakukannya deskripsi data, fase yang selanjutnya adalah melakukan eksplorasi terhadap data, dan kemudian akan diberikan deskripsi mengenai isi data dari setiap atribut pada data yang diperoleh.

Tabel 3. Eksplorasi Data

No	Atribut	Keterangan	Eksplorasi Data
1	Nomor Ijazah	Nomor ijazah mahasiswa angkatan 2012, 2013, 2014, 2015 dan 2016	Jumlah nomor ijazah mahasiswa: <ul style="list-style-type: none"> <li>• Angkatan Tahun 2012 = 26</li> <li>• Angkatan Tahun 2013 = 31</li> <li>• Angkatan Tahun 2014 = 33</li> <li>• Angkatan Tahun 2015 = 31</li> <li>• Angkatan Tahun 2016 = 25</li> </ul> Total data keseluruhan adalah 146.
2	NIM	Nomor Induk Mahasiswa (NIM) dari tahun angkatan 2012, 2013, 2014, 2015 dan 2016	Jumlah nomor induk mahasiswa: <ul style="list-style-type: none"> <li>• Angkatan Tahun 2012 = 26</li> <li>• Angkatan Tahun 2013 = 31</li> <li>• Angkatan Tahun 2014 = 33</li> <li>• Angkatan Tahun 2015 = 31</li> <li>• Angkatan Tahun 2016 = 25</li> </ul> Total data keseluruhan adalah 146.
3	Nama	Nama mahasiswa prodi Teknik Informatika STMIK Rosma Karawang	Total data keseluruhan adalah 146.
4	Jenis Kelamin	Jenis kelamin mahasiswa berisikan : <ul style="list-style-type: none"> <li>• Laki – Laki</li> <li>• Perempuan</li> </ul>	Jumlah jenis kelamin mahasiswa : <ul style="list-style-type: none"> <li>• Laki – Laki = 108</li> <li>• Perempuan = 38</li> </ul> Total data keseluruhan adalah 146
5	Asal Daerah	Asal daerah mahasiswa berisi : <ul style="list-style-type: none"> <li>• Karawang</li> <li>• Luar Karawang</li> </ul>	Jumlah asal daerah mahasiswa : <ul style="list-style-type: none"> <li>• Karawang = 54</li> <li>• Luar Karawang = 73</li> </ul> Total data keseluruhan adalah 127
6	Tanggal Lahir	Tanggal Lahir Mahasiswa	Total data keseluruhan adalah 127
7	Jumlah SKS	Jumlah Satuan Kredit Semester (SKS) yang telah ditempuh	Jumlah SKS : Min = 64 Max = 145

No	Atribut	Keterangan	Eksplorasi Data
8	IPK	Indeks Prestasi Kumulatif (IPK) mahasiswa pada angkatan 2012 sampai 2016	IPK mahasiswa angkatan 2012-2016 : Min = 2.54 Max = 3.76 Mean = 3.10
9	Predikat	Predikat kelulusan mahasiswa dari angkatan 2012 - 2016	Jumlah predikat kelulusan mahasiswa : • Cumlaude = 8 • Sangat Memuaskan = 91 • Memuaskan = 47 Total data keseluruhan adalah 146.
10	Lama Studi	Lamanya Studi Mahasiswa yang telah lulus dari angkatan 2012 – 2016	Jumlah studi mahasiswa : Min = 2 Tahun Max = 7 Tahun

c. *Verify Data Quality*

Fase verifikasi kualitas data ini akan melakukan pengecekan seperti *missing value*, data yang tidak konsisten, data *noise* atau data *outlier*. Dimana pengecekan ini bertujuan untuk mengetahui seberapa baik kualitas data yang telah didapatkan sebelumnya.

Dapat dilihat pada gambar 4 bahwa data yang telah diperoleh dari tata usaha ternyata masih terdapat *missing value* di beberapa asal daerah dan tanggal lahir. Dapat dilihat data yang mengandung *missing value* ialah baris kosong pada kolom asal daerah dan tanggal lahir.

Tabel 4. *Missing Value Data*

27	Nomor Seri Ijazah	NIM	Nama	Jenis Kela	Asal Daera	Tanggal Lahir	Jumlah SKS	IPK	Predikat	Lama Studi
28	20313/I	201301	Mahasiw	Prempuar	Karawang	27 September 1990	145	2,83	Memuask	4
29	20413/I	201302	Mahasiw	Laki-Laki	Luar karav	22 Oktober 1991	145	3,19	Sangat Me	4
30	20513/I	201304	Mahasisv	Laki-Laki	Karawang	13 Juni 1990	145	2,81	Memuask	4
31	20513/I	201305	mahasisv	Laki-Laki	Karawang	26 Februari 1995	145	2,97	Memuask	4
32	20613/I	201402	Mahasisv	Laki-Laki	Karawang	31 Desember 1993	145	2,98	Memuask	4
33	10013/I	201306	Mahasisv	Prempuar	Karawang	03 Maret 1992	145	2,79	Memuask	3
34	20013/I	201306	Mahasisv	Laki-Laki	Karawang	12 Juni 1993	145	2,80	Memuask	5
35	10013/I	201306	Mahasisv	Laki-Laki	Karawang	09 September 1984	145	3,57	Sangat Me	5
36	40013/I	201306	Mahasisv	Prempuar	Karawang	16 Januari 1996	145	3,03	Sangat Me	5
37	20013/I	201306	Mahasisv	Laki-Laki	Karawang	30 Maret 1996	145	3,63	Sangat Me	5
38	90013/I	201403	Mahasisv	Laki-Laki	Karawang	19 September 1996	145	2,98	Memuask	4
39	20013/I	201404	Mahasisv	Laki-Laki	Luar karav	10 Maret 1995	145	3,17	Sangat Me	4
40	20213/I	201405	Mahasisv	Laki-Laki	Luar karav	21 Juli 1993	145	3,18	Sangat Me	4
41	20113/I	201406	Mahasisv	Laki-Laki	Luar karav	16 Mei 1996	145	2,88	Memuask	4
42	20113/I	201407	Mahasisv	Prempuar	Luar karav	20 Februari 1993	145	2,88	Memuask	4
43	20101/I	201408	Mahasisv	Laki-Laki	Luar karav	10 Juli 1991	145	3,01	Sangat Me	4

4.3. *Data Preparation* (Persiapan Data)

Setelah melalui tahapan pengumpulan data awal dan pemahaman data yang telah diperoleh. Tahapan selanjutnya yaitu melakukan persiapan data sebelum data diproses menggunakan teknik data *mining*. Persiapan data dilakukan untuk mengubah data mentah menjadi data yang siap digunakan sehingga ketika proses prediksi kelulusan mahasiswa dapat berjalan dengan lancar. pada tahapan ini terdapat lima fase yang harus dilakukan yaitu *selet data*, *clean data*, *construct data*, dan *format data*.

a. *Secelect Data*

Dari data awal yang telah dikumpulkan, dipilih beberapa atribut yang nantinya atribut tersebut dianggap sebagai atribut yang berpengaruh terhadap waktu kelulusan mahasiswa. Atribut yang dipilih yaitu NIM, nama, jenis kelamin, asal daerah, tanggal lahir, jumlah sks, IPK, predikat, lama studi, dan stauts kelulusan.

Tabel 5. Atribut Data Terpilih

No	Atribut
1	NIM
2	Nama
3	Jenis Kelamin
4	Asal Daerah
5	Jumlah SKS
6	IPK
7	Predikat
8	Lama Studi
9	Status Kelulusan

Atribut NIM, nama, dan lama studi hanya digunakan sebagai identitas data saja. Sedangkan jenis kelamin, asal daerah, jumlah sks, IPK, predikat, dan lama studi yang akan diproses menggunakan teknik data *mining* dengan metode C4.5. Data yang dibutuhkan dalam melakukan prediksi kelulus ini ialah data mahasiswa angkatan 2012, 2013, 2014, 2015, dan 2016 yang telah lulus. Pada atribut status kelulusan akan berisikan dua buah nilai yaitu “Tepat Waktu” dan “Terlambat”. Dua nilai tersebut diambil dari hasil tranformasi lama studi dan tahun lulus.

b. *Clean Data*

Fase ini dilakukan untuk memastikan bahwa data yang digunakan sudah benar-benar memiliki kualitas yang baik, sehingga ketika melakukan prediksi mendapatkan hasil yang terbaik. Seperti yang telah diketahui data yang telah dikumpulkan memiliki beberapa *missing value* yang dapat dilihat pada gambar 4.2. Oleh karena itu pada tahapan ini data akan diolah untuk menangani data yang terdapat *missing value*. Salah satu cara yang dapat digunakan untuk mengatasi data yang terdapat *missing value* ialah menghapus data yang mengandung *missing value* pada *dataset* yang ada. Dibawah ini adalah gambaran data mahasiswa yang telah diurutkan dari mahasiswa angkatan 2012, 2013, 2014, 2015, dan 2016 yang telah diperoleh dan megandung *missing value*.

Tabel 6. *Missing Value* Data Mahasiswa 2012 – 2016

NIM	Nama	Jenis kelamin	Asal Daerah	Jumlah sks	IPK	Predikat	Lama Studi
201401	Mahasiswa01	Laki-Laki	?	145	3,15	Sangat Memuaskan	4
201402	Mahasiswa02	Laki-Laki	?	145	2,97	Memuaskan	4
201403	Mahasiswa03	Perempuan	?	145	2,97	Memuaskan	3
201502	Mahasiswa04	Laki-Laki	?	145	3,12	Sangat Memuaskan	4
201404	Mahasiswa05	Laki-Laki	?	145	2,92	Memuaskan	4
201405	Mahasiswa06	Perempuan	?	145	2,92	Memuaskan	5
201301	Mahasiswa07	Laki-Laki	?	145	2,94	Memuaskan	4
201406	Mahasiswa08	Laki-Laki	?	145	2,96	Memuaskan	4
201408	Mahasiswa09	Laki-Laki	?	145	3,15	Sangat Memuaskan	4
201409	Mahasiswa10	Laki-Laki	?	145	2,96	Memuaskan	4
201410	Mahasiswa11	Laki-Laki	?	145	3,19	Sangat Memuaskan	4
201411	Mahasiswa12	Laki-Laki	?	145	3,00	Sangat Memuaskan	5
201412	Mahasiswa13	Laki-Laki	?	145	2,92	Memuaskan	4
201313	Mahasiswa14	Laki-Laki	?	145	3,17	Sangat Memuaskan	4

Tabel 6 diatas merupakan data mahasiswa angkatan 2012-2016 yang mengandung *missing value*. *Missing value* dapat dilihat pada baris yang bersikan tanda tanya. Untuk mengatasi hal tersebut, maka akan dilakukan penghapusan data yang mengandung *missing value*. Total keseluruhan data sebanyak 146 data, sebanyak 19 data termasuk *missing value* dihilangkan, sehingga hanya terdapat 127 data yang akan diproses ke tahapan selanjutnya.

Gambar 7. Hasil *Cleaning* Data Mahasiswa 2012 – 2016

NIM	Nama	Jenis Kelamin	Asal Daerah	Jumlah SK	IPK	Predikat	Lama Studi
201201	Mahasiswa01	Laki-Laki	Karawang	145	2,79	Memuaskan	4
201202	Mahasiswa02	Laki-Laki	Luar Karawang	145	3,23	Sangat Memuaskan	4
201205	Mahasiswa03	Perempuan	Luar Karawang	145	3,12	Sangat Memuaskan	4
201206	Mahasiswa04	Laki-Laki	Luar Karawang	145	3,36	Cumlaude	4
201208	Mahasiswa05	Perempuan	Karawang	145	2,94	Memuaskan	4
201209	Mahasiswa06	Perempuan	Luar Karawang	145	3,42	Sangat Memuaskan	4
201209	Mahasiswa07	Perempuan	Luar Karawang	145	3,62	Cumlaude	4
201210	Mahasiswa08	Laki-Laki	Luar Karawang	145	3,29	Sangat Memuaskan	4
201211	Mahasiswa09	Laki-Laki	Luar Karawang	145	3,19	Sangat Memuaskan	4
201212	Mahasiswa10	Laki-Laki	Karawang	145	3	Memuaskan	4
201213	Mahasiswa11	Perempuan	Luar Karawang	145	2,93	Memuaskan	4
201214	Mahasiswa12	Laki-Laki	Luar Karawang	145	3,17	Sangat Memuaskan	4
201215	Mahasiswa13	Laki-Laki	Karawang	145	3,15	Sangat Memuaskan	4
201216	Mahasiswa14	Laki-Laki	Karawang	145	3,17	Sangat Memuaskan	4

Pada tabel 7 diatas merupakan data mahasiswa angkatan 2012, 2013, 2014, 2015, dan 2016 yang sudah dilakukan pembersihan dengan tidak menggunakan baris data yang memiliki *missing value*.

c. *Construct Data*

Fase selanjutnya merupakan langkah untuk mempersiapkan data yang konstruktif yaitu membangun atribut turunan dari atribut yang ada. Pada tahapan ini melakukan penambahan satu buah atribut baru yang dijadikan sebagai *class target*. atribut ini berisikan dua buah nilai yaitu “Tepat Waktu” dan “Terlambat”.

Tabel 8. Atribut Baru

Status Kelulusan
Tepat Waktu
Terlambat

Atribut tersebut didasari pada data lama studi dan tahun kelulusan mahasiswa yang dilihat pada nomor seri ijazah. Di mana mahasiswa yang mendapatkan *class target* atau dikategorikan sebagai mahasiswa yang lulus tepat waktu adalah mahasiswa yang menempuh studi tidak lebih dari 4 tahun. Sehingga data yang telah ditambahkan atribut sebagai berikut.

Tabel 9. Penambahan Atribut Baru Pada Data Mahasiswa STMIK Rosma Karawang

	Nomor Seri Ijazah	NIM	Nama	Jenis Kelamin	Asal Daerah	Tanggal Lahir	SKS	IPK	Predikat	Lama Studi
33	20313/I	200502	Mahasiswa01	Prempuan	Karawang	27 September 1990	145	2,83	Memuaskan	4
34	20413/I	200213	Mahasiswa02	Laki-Laki	Luar karawang	22 Oktober 1991	145	3,19	Sangat Me	4
35	20513/I	200241	Mahasiswa03	Laki-Laki	Karawang	13 Juni 1990	145	2,81	Memuaskan	4
36	20513/I	200310	mahasiswa04	Laki-Laki	Karawang	26 Februari 1995	145	2,97	Memuaskan	4
37	20613/I	200835	Mahasiswa05	Laki-Laki	Karawang	31 Desember 1993	145	2,98	Memuaskan	4
38	10013/I	200401	Mahasiswa06	Prempuan	Karawang	03 Maret 1992	145	2,79	Memuaskan	3
39	20013/I	200501	Mahasiswa07	Laki-Laki	Karawang	12 Juni 1993	145	2,80	Memuaskan	5
40	10013/I	200512	Mahasiswa08	Laki-Laki	Karawang	09 September 1984	145	3,57	Sangat Me	5
41	40013/I	200567	Mahasiswa09	Prempuan	Karawang	16 Januari 1996	145	3,03	Sangat Me	5
42	20013/I	200767	Mahasiswa10	Laki-Laki	Karawang	30 Maret 1996	145	3,63	Sangat Me	5
43	90013/I	200697	Mahasiswa11	Laki-Laki	Karawang	19 September 1996	145	2,98	Memuaskan	4
44	20013/I	200630	Mahasiswa12	Laki-Laki	Luar karawang	10 Maret 1995	145	3,17	Sangat Me	4
45	20213/I	200532	Mahasiswa13	Laki-Laki	Luar karawang	21 Juli 1993	145	3,18	Sangat Me	4
46	20113/I	200758	Mahasiswa14	Laki-Laki	Luar karawang	16 Mei 1996	145	2,88	Memuaskan	4
47	20113/I	200739	Mahasiswa15	Prempuan	Luar karawang	20 Februari 1993	145	2,88	Memuaskan	4
48	20101/I	200588	Mahasiswa16	Laki-Laki	Luar karawang	10 Juli 1991	145	3,01	Sangat Me	4

d. *Transformation*

Fase ini dilakukan untuk mengubah data maka digunakan *transformasi* data sehingga informasi terbaik dapat diambil dengan cara mengubah tipe data standar menjadi kategorikal sehingga dapat digunakan untuk klasifikasi teknik data *minihg* dengan algoritma C4.5.

Pada data yang telah didapatkan terdapat beberapa atribut yang perlu diubah yaitu pada atribut jumlah SKS dan IPK. Jumlah SKS dibagi menjadi dua kategori yaitu jumlah SKS yang ditemui kurang dari atau sama dengan 144 SKS ( $\leq 144$ ) dan jumlah sks yang lebih dari 144 ( $> 144$ ). Pada atribut IPK juga dibagi menjadi dua kategori yaitu IPK yang dimiliki kurang atau sama dengan 3 ( $\leq 3$ ).

Tabel 10. Hasil Transformasi Atribut

No	Jumlah Sks	IPK	Nomor Seri Ijazah	NIM	Nama	Jenis Kel	Asal Daerah	Tanggal Lahir
1	SKS>144	IPK<=3	20011	201201	Mahasiswa01	Pempuan	Karawang	27 September 1990
2	SKS>144	IPK>3	20011	2012012	Mahasiswa02	Laki-Laki	Luar karawang	22 Oktober 1991
3	SKS>144	IPK>3	20011	2012015	Mahasiswa03	Laki-Laki	Karawang	13 Juni 1990
4	SKS>144	IPK>3	20011	2012014	Mahasiswa04	Laki-Laki	Karawang	26 Februari 1995
5	SKS>144	IPK>3	20011	2020154	Mahasiswa05	Laki-Laki	Karawang	31 Desember 1993
6	SKS>144	IPK<=3	20011	2010123	Mahasiswa06	Pempuan	Karawang	03 Maret 1992
7	SKS>144	IPK>3	20011	2010124	Mahasiswa07	Laki-Laki	Karawang	12 Juni 1993
8	SKS>144	IPK>3	20011	2014002	Mahasiswa08	Laki-Laki	Karawang	09 September 1984
9	SKS>144	IPK>3	20011	2014008	Mahasiswa09	Pempuan	Karawang	16 Januari 1996
10	SKS>144	IPK>3	20011	2014009	Mahasiswa10	Laki-Laki	Karawang	30 Maret 1996
11	SKS>144	IPK<=3	20011	20150019	Mahasiswa11	Laki-Laki	Karawang	19 September 1996
12	SKS>144	IPK<=3	20011	20150014	Mahasiswa12	Laki-Laki	Luar karawang	10 Maret 1995
13	SKS>144	IPK>3	20011	20150026	Mahasiswa13	Laki-Laki	Luar karawang	21 Juli 1993
14	SKS>144	IPK>3	20011	20150036	Mahasiswa14	Laki-Laki	Luar karawang	16 Mei 1996

e. *Format Data*

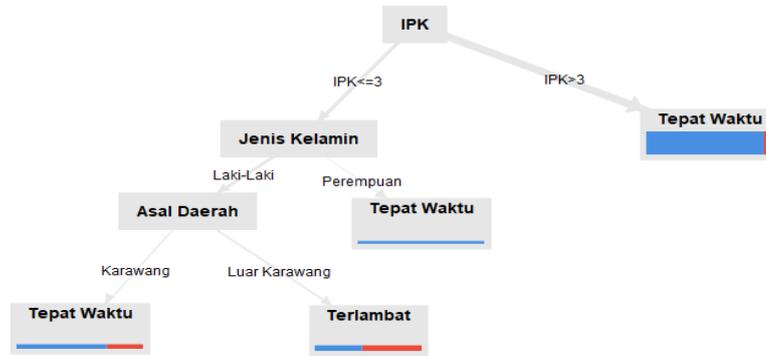
Fase terakhir pada tahapan *preparation* adalah *format data*. Pada fase ini dilakukannya produksi *set data* akhir yang siap digunakan untuk diolah menggunakan *tools* pemodelan data *mining*. *Dataset* yang akan digunakan untuk melakukan prediksi kelulusan yaitu sebagai berikut.

Tabel 11. Atribut Data Terpilih

No	Atribut	Jenis	Keterangan
1	Jenis Kelamin	Binominal	Jenis Kelamin Mahasiswa : Laki-Laki, Perempuan
2	Asal Daerah	Binominal	Asal Daerah Mahasiswa : Karawang, Luar Karawang
3	Jumlah SKS	Integer	Jumlah SKS yang telah ditempuh mahasiswa
4	IPK	Real	Indeks Prestasi Kumulatif (IPK) mahasiswa
5	Predikat	Polynomial	Predikat Mahasiswa : Cumlaude, Sangat Memuaskan, Memuaskan
6	Status Kelulusan	binominal	Status Kelulusan : Tepat Waktu, Terlambat

4.4. *Modeling* (Pemodelan)

Prediksi kelulusan mahasiswa Teknik Informatika STMIK Rosma Karawang ini menggunakan metode *decision tree*. Metode *decision tree* ini dibangun menggunakan *tools* RapidMiner. Data yang digunakan sebanyak 127 data yang terdiri dari dua kelas yaitu tepat waktu dan terlambat untuk mengetahui pola mahasiswa teknik informatika STMIK Rosma Karawang yang dapat menyelesaikan studinya dengan waktu lulus tepat atau terlambat berdasarkan variabel-variabel yang dianjurkan dengan menggunakan metode *decision tree*. Pada penelitian ini untuk mengukur validitas model klasifikasi digunakan *cross validation*. *Cross validation* merupakan salah satu metode untuk menentukan data training dan data testing dari keseluruhan data. *Cross validation* mengulang n-kali untuk membagi himpunan data secara acak. Niali yang disarankan adalah 10 karena memiliki keakuratan yang lebih dalam ukuran estimasinya. Proses iterasi dilakukan sebanyak 10 kali sehingga dalam proses iterasi dihasilkan *decision tree* sebanyak 10.



Gambar 2. Pohon Keputusan Klasifikasi Predikat Kelulusan

Pada gambar 2 terdapat hasil dari klasifikasi predikat kelulusan mahasiswa teknik informatika STMIK Rosma Karawang dengan metode *decision tree*. Dilihat pada pohon keputusan yang sudah terbentuk pola, maka dapat disimpulkan bahwa atribut yang memiliki pengaruh paling tinggi untuk menentukan klasifikasi predikat kelulusan mahasiswa teknik informatika STMIK Rosma Karawang adalah IPK. Hal ini ditunjukkan dengan atribut IPK menempati simpul akar (*root node*).

#### 4.5. Evaluation (Evaluasi)

Implementasi dari metode *decision tree* dengan menggunakan *cross validation* untuk mengukur validitas model dengan sebanyak 127 data mahasiswa teknik informatika STMIK Rosma Karawang menghasilkan keberhasilan sebanyak 109 data. Klasifikasi predikat kelulusan mahasiswa dengan metode *decision tree* ini didapatkan hasil klasifikasi yang disajikan dalam bentuk tabel *confusion matrix*. *Confusion matrix* memberikan informasi mengenai klasifikasi pada konsep *data mining*.

Tabel 12. *Confusion Matrix*

<b>Prediction</b>	<b>Actual</b>	
	Tepat Waktu	Terlambat
Tepat Waktu	99 (TP)	10 (FP)
Terlambat	8 (FN)	10 (TN)

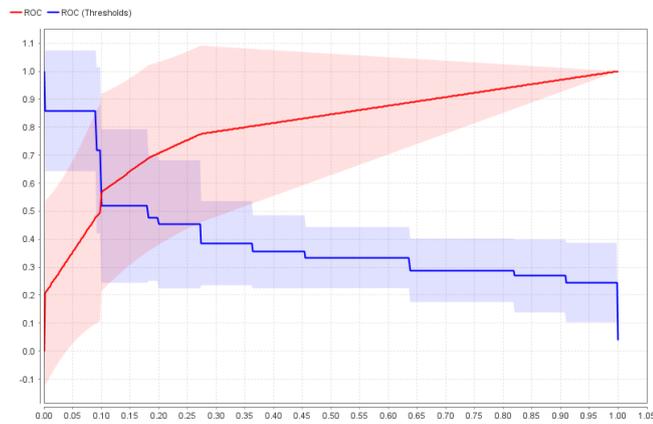
Berdasarkan tabel *confusion matrix* dengan algoritma *decision tree*, diketahui dari data yang berjumlah 127 data didapatkan 99 data tepat waktu benar diprediksi sebagai kelas tepat waktu, sementara 10 data tepat waktu diprediksi terlambat. sedangkan untuk kelas terlambat, 10 data terlambat benar diprediksi terlambat, sementara 8 data terlambat diprediksi tepat waktu.

Setelah diketahui hasil dari tabel *confusion matrix*, maka selanjtnya adalah mencari nilai dari *accuracy*, *recall*, *precision* dan AUC.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% = \frac{99+10}{99+10+8+10} \times 100\% = 85,83\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\% = \frac{99}{99 + 8} \times 100\% = 92,52\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\% = \frac{99}{99 + 10} \times 100\% = 90,83\%$$



Gambar 3. Kurva ROC

Pada gambar 3 menunjukkan bahwa dari model yang telah dibuat dan diuji menghasilkan bentuk kurva dengan nilai AUC sebesar 0,799 yang berarti akurasi dari prediksi adil atau sama.

#### 4.6. Deployment (Penyebaran)

Tahap ini menjelaskan semua alur proses data mining yang telah dilakukan dalam memprediksi ketepatan kelulusan mahasiswa teknik informatika STMIK Rosma Karawang dan mengetahui kinerja dari metode *decision tree* dalam melakukan prediksi. Pada tahap ini terdapat tiga fase yaitu *plan deployment*, *plan monitoring and maintenance*, *produce final report*.

##### a. Plan Deployment

Rencana penyebaran dilakukan dengan membuat laporan kegiatan selama penelitian. Laporan kegiatan berisikan hal-hal yang berkaitan tentang penelitian selama berjalannya penelitian dimulai dari pengumpulan data akademik mahasiswa teknik informatika STMIK Rosma Karawang, pengolahan data, proses data yang terpilih dan pemodelan data menggunakan metode *decision tree* dan mencantumkan hasil klasifikasi yang telah didapat.

##### b. Plan Monitoring and Maintenance

Pemantauan dilakukan untuk memperhatikan setiap langkah pada tahapan proses data *mining* ketika melakukan prediksi. Hal tersebut dilakukan untuk mengetahui hasil evaluasi dari proses data *mining* dalam melakukan prediksi ketepatan kelulusan mahasiswa teknik informatika STMIK Rosma Karawang, sehingga tidak ada satu tahapan yang terlewat. Hasil yang diperoleh harus sesuai berdasarkan pada tahapan pemahaman bisnis yang telah dipaparkan sebelumnya yaitu melakukan prediksi tingkat kelulusan mahasiswa teknik informatika STMIK Rosma Karawang. Hasil prediksi yang dilakukan dengan menggunakan metode *decision tree* berdasarkan nilai *accuracy*, *recall*, *precision*, dan AUC.

##### c. Produce Final Report

Fase ini membuat laporan dari semua kegiatan penelitian mengenai data *mining* dalam melakukan prediksi tingkat kelulusan mahasiswa yang telah dilakukan. Dimulai dengan pemahaman bisnis dan menjelaskan mengenai tujuan dari proses data mining yang dilakukan atau tujuan penelitian yaitu melakukan prediksi tingkat kelulusan mahasiswa teknik informatika STMIK Rosma Karawang. Kemudian melakukan pengumpulan data akademik mahasiswa yang di mana data tersebut selanjutnya digunakan untuk mencari informasi mengenai atribut apa saja yang terkandung di dalam data tersebut. Setelah data diperoleh kemudian melakukan seleksi data untuk menentukan atribut terpilih yang kemudian akan di proses menggunakan teknik dan melakukan pembersihan data untuk menghilangkan *record* yang kosong untuk meningkatkan kualitas data. Selanjutnya data diubah sesuai dengan kebutuhan proses data *mining*. Setelah data siap digunakan, selanjutnya memilih metode *decision tree* sebagai teknik data mining yang akan digunakan. Selanjutnya membangun model dan menguji model yang dibuat. Setelah mengetahui hasilnya, selanjutnya mengevaluasi model yang telah digunakan. Laporan dibuat dalam bentuk presentasi dan proposal skripsi.

## 5. Kesimpulan

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan beberapa hal antara lain:

- a. Penelitian ini melakukan klasifikasi predikat tingkat kelulusan mahasiswa program studi teknik informatika STMIK Rosma Karawang menggunakan metode *desicion tree* dengan menerapkan metodologi CRISP-DM. Penelitian ini diawali dengan pengamatan situasi yang terjadi di program studi teknik informatika STMIK Rosma Karawang. Kemudian melakukan pengumpulan data yang diperlukan, pembersihan data, dan transformasi data sebelum melakukan klasifikasi. Data yang digunakan untuk klasifikasi predikat tingkat kelulusan mahasiswa ini meliputi NIM, nama, jenis kelamin, asal daerah, tanggal lahir, jumlah sks, IPK, predikat, dan lama studi. Setelah data diperoleh sebanyak 146 data, selanjutnya membuat pemodelan dengan menggunakan *cross validation* untuk membagi data training dan data testing. *Dataset* yang digunakan untuk klasifikasi predikat kelulusan yaitu *dataset* mahasiswa angkata 2012, 2013, 2014, 2015, dan 2016 yang sudah lulus. Penggunaan metode *desicion tree* akan menghasilkan *output* berupa pohon keputusan yang menampilkan pola dari klasifikasi predikat kelulusan mahasiswa teknik informatika STMIK Rosma Krawarang. Hasil pengujian berupa tabel *confusion matrix* yang berisikan prediksi-prediksi benar dan salah. Dari prediksi tersebutlah yang akan digunakan untuk mencari nilai *accuracy*, *recall*, *precision*, dan AUC untuk mengevaluasi model prediksi.
- b. Hasil semua pengujian dari pemodelan yang telah dibuat dengan pembagian data testing dan data training dengan menggunakan *cross validation* menghasilkan jumlah nilai *true positif* 99, *true negatif* 10, *false negatif* 8, *false positif* 10 serta nilai *accuracy* sebesar 85,83%, nilai *recall* sebesar 92,52%, nilai *precision* sebesar 90,83%, dan nilai AUC sebesar 0,799.

## References

- Astuti, D., Iskandar, A. R., & Febrianti, A. (2019). Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 1(2), 60–72. <https://doi.org/10.20895/inista.v1i2.71>
- Fadillah, A. P. (2015). Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ). *Jurnal Teknik Informatika Dan Sistem Informasi*, 1(3), 260–270. <https://doi.org/10.28932/jutisi.v1i3.406>
- Honesqi, H. D. (2017). Klasifikasi Data Mining Untuk Menentukan Tingkat Persetujuan Kartu Kredit. *Jurnal Teknoif*, 5(2), 57–62. <https://doi.org/10.21063/jtif.2017.v5.2.57-62>
- Imaslihhah, S., Ratna, M., & Ratnasari, V. (2013). Analisis regresi logistik ordinal terhadap faktor-faktor yang mempengaruhi predikat kelulusan mahasiswa S1 di ITS Surabaya. *Jurnal Sains Dan Seni Pomits*, 2(2), 177–182.
- Luvia, Y. S., Hartama, D., Windarto, A. P., & Solikhun. (2016). Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Amik Tunas Bangsa. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, 1(1). <https://doi.org/10.30645/jurasik.v1i1.12>
- Muhammad, Z., Rahmadhani, R., Rizqifaluthi, H., & Yaqin, M. A. (2018). Process Mining Akademik Sekolah Menggunakan RapidMiner. *Matics: Jurnal Ilmu Komputer Dan Teknologi Informasi*, 10(2), 47–51. <https://doi.org/10.18860/mat.v10i2.5745>
- Nasrullah, A. H. (2018). Penerapan Metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out. *ILKOM Jurnal Ilmiah*, 10(2), 244–250. <https://doi.org/10.33096/ilkom.v10i2.300.244-250>
- Permata, E., Eddy Purnama, I. K., & Hery Purnomo, M. (2012). Klasifikasi Jenis dan Fase Parasit Malaria Plasmodium Falciparum dan Plasmodium Vivax Dalam Sel Darah Merah Menggunakan Support Vector Machine. *Setrum: Sistem Kendali-Tenaga-Elektronika-Telekomunikasi-Komputer*, 1(2), 50. <https://doi.org/10.36055/setrum.v1i2.475>
- Purnia, D. S., & Warnilah, A. I. (2017). Implementasi Data Mining Pada Penjualan kacamata Dengan Menggunakan Algoritma Apriori. *Indonesian Journal on Computer and Information Technology*, 2(2), 31–39.
- Rahayu, T. M., Ningsi, B. A., Isnurani, & Arofah, I. (2021). KLASIFIKASI KETEPATAN WAKTU KELULUSAN MAHASISWA DENGAN METODE NAÏVE BAYES. *Jurnal Media Bina Ilmiah*, 15(10). <http://ejournal.binawakya.or.id/index.php/MBI>
- Rusdiana, & Rosmiati, L. (2016). Aplikasi Berbasis Fuzzy C-Means Dalam Penentuan Predikat Kelulusan Mahasiswa. *Jurnal Ilmiah Ilmu Komputer Fakultas ...*, 2(2), 1–9. <http://www.ejournal.fikom-unasman.ac.id/index.php/jikom/article/view/78>
- Sartika, D., & Sensuse, D. I. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. *Jurnal Teknik Informatika Dan Sistem*

*Informasi*, 1(2), 151–161.

- Sitorus, D. R., Windarto, A. P., Hartama, D., & Damanik, I. S. (2019). Penerapan Klasifikasi C4.5 Dalam Meningkatkan Sistem Pembelajaran Mahasiswa. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 593–597. <https://doi.org/10.30865/komik.v3i1.1665>
- Wajhillah, R., & Yulianti, I. (2017). Penerapan Algoritma C4.5 Untuk Prediksi Penggunaan Jenis Kontrasepsi Berbasis Web. *Klik - Kumpulan Jurnal Ilmu Komputer*, 4(2), 160. <https://doi.org/10.20527/klik.v4i2.98>