

# Implementasi Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa STIMIK ESQ Menggunakan Decision Tree C4.5

Mita Nurul Yatimah  
Program Studi Ilmu Komputer  
Sekolah Tinggi Ilmu Manajemen dan Ilmu Komputer ESQ  
Jl. Tb. Simatupang, Kav. 1 Cilandak Timur, Jakarta 12560  
m.nurul.y@students.esqbs.ac.id

**Abstrak**— Kelulusan mahasiswa merupakan salah satu faktor yang mempengaruhi penilaian akreditasi perguruan tinggi. Semakin banyak mahasiswa lulus tepat waktu (empat tahun) maka akan semakin baik pula penilaian akreditasi perguruan tinggi tersebut. Oleh karena itu penelitian ini bertujuan untuk melakukan analisis faktor yang mempengaruhi kelulusan tepat waktu dan prediksi kelulusan tepat waktu secara dini. Penelitian ini dilakukan di STIMIK ESQ yang merupakan salah satu perguruan tinggi yang ada di Jakarta Selatan. Hasil penelitian ini mampu memberikan manfaat bagi para dosen untuk menganalisis faktor yang mempengaruhi kelulusan tepat waktu dan memprediksi kelulusan mahasiswa secara objektif dan terotomatisasi. Sehingga hasil analisis dan prediksi dapat dijadikan sebagai salah satu acuan untuk memberikan penanganan khusus terhadap kelulusan mahasiswa secara dini. Metode yang digunakan pada penelitian ini yaitu *Decision Tree C4.5* dengan parameter jenis kelamin, prodi, usia, IPS1, SKS1, IPK1, IPS2, SKS2, IPK2, IPS3, SKS3, IPK3, IPS4, SKS4, dan masa studi. Penelitian ini menghasilkan akurasi sebesar 90% 21 rules.

**Kata kunci**— Prediksi, Kelulusan, Decision tree C4.5, Data Mining

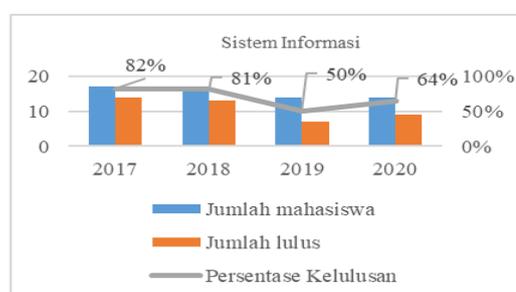
## I. PENDAHULUAN

Perkembangan dunia pendidikan di Indonesia telah memberikan dampak persaingan yang sangat ketat. Hal ini dipicu akibat semakin majunya pendidikan di perguruan tinggi. Salah satu dampak dari persaingan yaitu menghasilkan lulusan yang berkualitas. Adapun kriteria lulusan yang berkualitas diantaranya mampu menyelesaikan masa pembelajaran tepat waktu.

Masa pembelajaran tepat waktu sangat mempengaruhi kualitas dari perguruan tinggi. Kemampuan perguruan tinggi menghasilkan lulusan yang mampu menyelesaikan masa pembelajaran tepat waktu merupakan faktor yang mempengaruhi akreditasi perguruan tinggi. Hal ini sesuai dengan peraturan Badan Akreditasi Nasional Perguruan Tinggi Nomor 3 tahun 2019 tentang Instrumen Akreditasi Perguruan Tinggi yang menyatakan bahwa salah satu indikator penilaian akreditasi adalah persentase lulusan tepat waktu untuk setiap program dari perguruan tinggi [1]. Untuk itu sangat penting

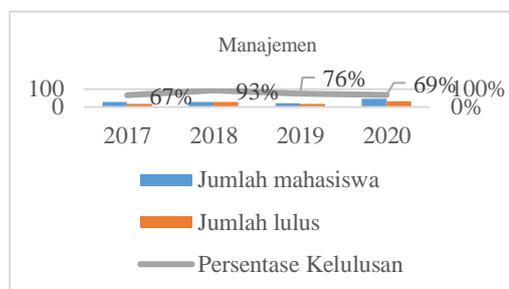
mencari faktor yang mempengaruhi kelulusan tepat waktu di perguruan tinggi.

Studi kasus pada penelitian ini adalah STIMIK (Sekolah Tinggi Ilmu Manajemen Ilmu Komputer) ESQ yang berada di Jakarta Selatan. Adapun yang menjadi objek pada penelitian ini merupakan mahasiswa dari prodi manajemen dan sistem informasi. **Error! Reference source not found.** di bawah merupakan grafik kelulusan mahasiswa program studi informasi tahun 2017 sampai dengan tahun 2020.



Gambar 1. Persentase kelulusan program studi sistem informasi

**Error! Reference source not found.** di bawah merupakan grafik kelulusan mahasiswa program studi manajemen tahun 2017 sampai dengan tahun 2020.



Gambar 2. Persentase kelulusan program studi manajemen

Merujuk pada **Error! Reference source not found.** dan **Error! Reference source not found.** terlihat bahwa adanya pola kelulusan yang mengalami penurunan ataupun kenaikan yang sangat signifikan. Hal ini terlihat pada program studi sistem informasi pada tahun 2018-2019 yang mengalami penurunan drastis sebanyak 31%. Hal ini juga terjadi pada program studi manajemen pada tahun 2017-2018 yang menunjukkan bahwa adanya kenaikan yang signifikan yaitu sebanyak 26%. Namun pada tahun selanjutnya mengalami penurunan terus menerus hingga tahun 2020. Untuk itu maka perlu suatu tindakan untuk mengatasi permasalahan persentase lulusan tepat waktu. Dengan demikian penelitian ini akan melakukan prediksi dan analisis terhadap faktor-faktor yang mempengaruhi kelulusan tepat waktu dengan menerapkan *data mining* untuk proses pengolahan data secara otomatis.

Analisis dan prediksi diharapkan mampu menemukan faktor-faktor yang mempengaruhi dan memperediksi kelulusan tepat waktu. Sehingga dapat dilakukan prediksi kelulusan mahasiswa lebih dini. Manfaat lainnya yaitu untuk menunjang nilai akreditasi serta sistem yang lebih terintegrasi. Data yang digunakan pada penelitian ini merupakan data kelulusan mahasiswa tahun 2017-2020 pada prodi sistem informasi dan manajemen bisnis. Peneliti fokus pada faktor-faktor yang mempengaruhi kelulusan mahasiswa dari semester 1- 4 agar prediksi dapat dilakukan lebih dini.

Penelitian terhadap kelulusan mahasiswa telah banyak dilakukan diantaranya yaitu “Prediksi Kelulusan Mahasiswa Diploma dengan Komparasi Algoritma Klasifikasi”. Penelitian ini mengkomparasi 5 algoritma data mining yaitu *Decision tree* C4.5, *Naive Bayes*, *K-NN*, *rule Induction*, dan *random forest* dengan tujuan untuk menentukan metode yang paling akurat. Data yang digunakan yaitu data mahasiswa AMIK BSI Pontianak angkatan 2013-2017 prodi Informatika sebanyak 394 *record*. Hasil penelitian menunjukkan bahwa algoritma yang paling optimal adalah *Decision tree* C4.5 dengan akurasi sebesar 90,85% [2].

Penelitian selanjutnya dengan judul “Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma *Decision Tree*”. Penelitian ini melakukan komparasi antar algoritma yang dipakai dalam *Decision Tree* yaitu ID3, CHAID, dan C4.5. Tujuan dari penelitian ini yaitu untuk menemukan pola prediksi kelulusan mahasiswa tepat waktu dengan menggunakan teknik *data mining*. Data yang digunakan pada penelitian ini adalah data mahasiswa STIMIK Adhiguna angkatan 2009 dengan atribut usia, jenis kelamin, dan indeks prestasi semester 1-4. Hasil penelitian menunjukkan bahwa *Decision Tree* memiliki kompleksitas yang cukup tinggi karena adanya penelusuran dan pemrosesan nilai pada setiap atributnya dengan tujuan untuk mendapatkan *entropi*. Akurasi tertinggi yang dihasilkan dari penerapan metode *Decision Tree* C4.5, ID3, dan CHAID didapatkan oleh *Decision Tree* C4.5 yaitu sebesar 91,51%. Dapat disimpulkan bahwa *Decision Tree* C4.5 dapat digunakan untuk memprediksi kelulusan mahasiswa [3].

Penelitian selanjutnya yang berjudul “Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus di Universitas Peradaban)”. Penelitian ini menggunakan metode *Decision tree* C4.5, data yang digunakan adalah data kelulusan

mahasiswa universitas Peradaban dengan atribut IPK, SKS, Umur, dan Jenis Kelamin. Tujuan dari penelitian tersebut untuk mengetahui tingkat akurasi algoritma C4.5. Hasil penelitian tersebut menunjukkan bahwa algoritma C4.5 dapat memprediksi kelulusan mahasiswa Universitas Peradaban dengan tingkat akurasi 88,74%, presisi 91,7%, dan recall sebesar 95,34% [4].

Berdasarkan referensi penelitian terdahulu, *Decision tree* memiliki tingkat akurasi yang paling tinggi terutama dengan menggunakan algoritma C4.5 jika dibandingkan dengan metode data mining lainnya. Maka pada penelitian akan digunakan *Decision Tee* C4.5. Dengan menggunakan metode tersebut diharapkan mampu menemukan faktor yang mempengaruhi kelulusan tepat waktu dengan akurasi yang baik.

Pembaharuan pada penelitian ini yaitu terletak pada teknik pemilihan atribut dan teknik pemilihan sampel. Teknik pemilihan atribut menggunakan metode regresi logistik biner, penentuan jumlah anggota sampel menggunakan tabel *krejcie-morgan* dan penentuan anggota sampel menggunakan *stratified random sampling*. Tujuan dari penelitian ini adalah untuk mengetahui faktor yang mempengaruhi kelulusan di STIMIK ESQ dan menerapkan data mining dalam memprediksi kelulusan tepat waktu sedari dini di STIMIK ESQ.

## II. METODE

Langkah untuk melakukan prediksi kelulusan tepat waktu mahasiswa STIMIK ESQ adalah sebagai berikut:

### A. Seleksi Data

Pada umumnya data yang diperoleh dari database memiliki isian yang tidak sesuai atau tidak sempurna seperti *missing value*, data yang tidak valid atau hanya sekedar salah penulisan. Maka pada tahap ini dilakukan pembersihan data dari *missing value* serta data yang tidak relevan. Pembersihan data perlu dilakukan agar data yang diolah memiliki performa yang baik.

Dari total 186 data didapatkan 86 data yang bersih dari *missing value* data yang tidak valid atau hanya sekedar salah penulisan. pembersihan data dilakukan secara manual dengan mengeliminasi data yang memiliki *missing value* dan data yang tidak relevan. Dari 86 data tersebut akan diambil sebagian data sebagai sampel, proses pengambilan sampel dilakukan dengan menggunakan metode *stratified random sampling*.

*Stratified Random Sampling* Merupakan suatu proses pengambilan data yang akan dijadikan sebagai sampel melalui pembagian populasi ke dalam strata tertentu, memilih sampel secara acak dari setiap stratum dan kemudian digabungkan dalam sebuah data yang akan dijadikan sebagai sampel [5]. Komposisi pengambilan sampel dilakukan secara acak dengan jumlah sampel bersifat proporsional sesuai dengan jumlah kelas yang ada untuk kemudian dijadikan sebagai total sampel penelitian [6]. sedangkan penentuan jumlah sampel mengacu pada tabel *krejcie-Morgan*.

Tabel *Krejcie-Morgan* merupakan tabel yang digunakan untuk menentukan jumlah sampel dari  $n$  populasi dengan mengasumsikan tingkat keahwalannya sebesar 95% dan tingkat kesalahannya 5% [7]. Sampel yang diperoleh memiliki tingkat kepercayaan sebesar 95% terhadap keseluruhan populasi.

### B. Preprocessing

*Preprocessing* bertujuan untuk memilih variabel yang mempengaruhi kelulusan tepat waktu. analisis dilakukan dengan menggunakan bantuan aplikasi SPSS dengan teknik Regresi Logistik biner. Regresi Logistik Biner merupakan suatu metode analisis statistik yang berguna untuk menganalisis hubungan antar suatu variabel respon dengan beberapa prediktor [8].

Adapun variabel yang diuji dengan menggunakan metode regresi logistik biner adalah Jenis Kelamin, Usia saat masuk yang diketahui dari pengurangan tahun masuk dengan tahun lahir, Prodi, Indeks Prestasi Semester satu sampai dengan semester 4, Satuan Kredit Semester 1 sampai dengan semester 4, Indeks Prestasi Kumulatif semester 1 sampai dengan semester 4 dan Masa studi sebagai label.

Enam belas variabel tersebut diuji untuk menentukan atribut yang paling berpengaruh terhadap kelulusan tepat waktu, pengujian didasarkan pada nilai signifikansi, atribut dikatakan berpengaruh jika memiliki nilai signifikansi  $\leq 0,05$ . Hasil pengujian menunjukkan bahwa nilai  $\text{sig} \geq 0,05$  yang artinya secara parsial variabel bebas tidak mempengaruhi model. Dari ke-15 variabel tersebut yang mendekati nilai  $\leq 0,05$  adalah variabel SKS4, SKS3, IPK1 dan seterusnya. Hal tersebut menunjukkan bahwa SKS4 memiliki nilai yang paling berpengaruh terhadap kelulusan mahasiswa. Pengujian tersebut juga menunjukkan bahwa model yang dibuat sudah fit.

### C. Transformasi Data

Data dari setiap atribut ditransformasikan ke dalam bentuk kategori agar data dapat diproses dan berada di rentang nilai yang sama, kemudian dinotasikan dengan rentang angka 0-4. Adapun transformasi yang dilakukan adalah sebagai berikut:

- 1) Jenis Kelamin, atribut ini memiliki dua nilai yaitu laki-laki dan perempuan, dimana laki-laki dinotasikan dengan 0 dan perempuan dinotasikan dengan 1.
- 2) Prodi, atribut ini memiliki dua nilai yaitu sistem informasi dan manajemen, dimana manajemen dinotasikan dengan 1 dan sistem informasi dengan 0.
- 3) Usia, atribut usia ditransformasikan seperti pada Tabel 1.

TABEL 1. TRANSFORMASI VARIABEL USIA

Usia	Transformasi
$\leq 15$	3
16-18	2
19-21	1
$\geq 21$	0

4) Indeks Prestasi Semester dan Indeks Prestasi Kumulatif, atribut ini memiliki rentang nilai dari 0,00 sampai dengan 4,00. yang ditransformasikan seperti pada Tabel 2.

5) Satuan Kredit Semester (SKS), atribut ini memiliki rentang nilai antara 0 sampai dengan 24 yang ditransformasi seperti pada Tabel 3

TABEL 2. TRANSFORMASI VARIABEL INDEKS PRESTASI

Indeks Prestasi	Notasi
-----------------	--------

$\geq 3,51$	0
3,01-3,50	1
2,76-3,00	2
2,00-2,75	3
$\geq 1,99$	4

TABEL 3. TRANSFORMASI VARIABEL SKS

Jumlah SKS	Notasi
22-24	0
19-21	1
16-18	2
$\leq 15$	3

6) Masa Studi, atribut ini memiliki dua nilai yaitu “Tepat” dan terlambat, dimana tepat dinotasikan dengan 1 dan terlambat dengan 0.

### D. Proses Data Mining

Berdasarkan data dan atribut yang telah didapat dari proses sebelumnya, proses selanjutnya adalah melakukan pengolahan data menggunakan metode *Decision tree* C4.5. *Decision Tree* merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon yang berisi alternatif-alternatif pemecahan dari suatu permasalahan [9]. Pengertian tersebut sejalan dengan pengertian *Decision Tree* menurut Kamal dan kawan-kawan [10] yang menyatakan bahwa *Decision Tree* merupakan salah satu metode klasifikasi yang direpresentasikan dengan struktur pohon (*Tree*) atribut direpresentasikan oleh node, nilai dari atribut direpresentasikan oleh cabang, kelas direpresentasikan oleh daun.

Node akar (*root*) merupakan level teratas yang biasanya berupa atribut yang memiliki pengaruh paling besar pada suatu kelas tertentu, konsep dari *Decision Tree* adalah mengubah data menjadi suatu model pohon keputusan yang kemudian diubah menjadi sebuah *rule* [11].

## III. HASIL DAN DISKUSI

Populasi yang digunakan pada penelitian ini sebanyak 86 data sedangkan sampel yang digunakan sebanyak 70 data yang terdiri dari 35 data mahasiswa lulus tepat waktu dan 35 data mahasiswa lulus tidak tepat waktu atau terlambat. Penentuan jumlah sampel didasarkan pada tabel *Krejcie Morgan*.

Variabel yang digunakan pada penelitian ini sebanyak 16 data yang terdiri dari variabel bebas yaitu Usia, Jenis Kelamin, Prodi, SKS 1-4, IPS 1-4, dan IPK 1-4. Sedangkan variabel terikatnya adalah variabel masa studi. Untuk menguji apakah variabel bebas secara bersama-sama mempengaruhi model atau tidak, variabel bebas secara sendiri-sendiri (parsial) mempengaruhi model atau tidak dan menentukan model yang dibuat apakah fit atau tidak, dilakukan pengujian dengan menggunakan metode regresi logistik biner. Dari hasil pengujian yang didapatkan sebagai berikut:

### 1. Tabel Omnibus Test

Pada Tabel 4 dapat dilihat bahwa nilai  $\text{sig} < 0,05$  yang artinya secara bersama-sama variabel bebas terbukti mempengaruhi model.

TABEL 4. OMNIBUS TEST

Step	Chi-square	df	Sig.
------	------------	----	------

1	Step	55,209	14	,000
	Block	55,209	14	,000
	Model	55,209	14	,000

## 2. Variabel in the equation

TABEL 5. VARIABLE IN THE EQUATION

Variabel	Sig.
JK(1)	,653
USIA	,369
PRODI(1)	,570
IPS1	,895
IPK1	,033
IPS2	,630
SKS2	,754
IPK2	,898
IPS3	,427
SKS3	,027
IPK3	,818
IPS4	,532
SKS4	,008
IPK4	,269
Constant	,069

Pada Tabel 5 dapat dilihat nilai sig  $\geq 0,05$  yang artinya secara parsial atau sendiri-sendiri variabel bebas tidak mempengaruhi model. Dari ke-15 variabel tersebut yang mendekati nilai  $\leq 0,05$  adalah variabel SKS4, SKS3, IPK1, dan seterusnya. Hal tersebut menunjukkan bahwa SKS4 memiliki nilai yang paling berpengaruh terhadap kelulusan mahasiswa.

## 3. Tabel Hosmer and Lomeshow Test

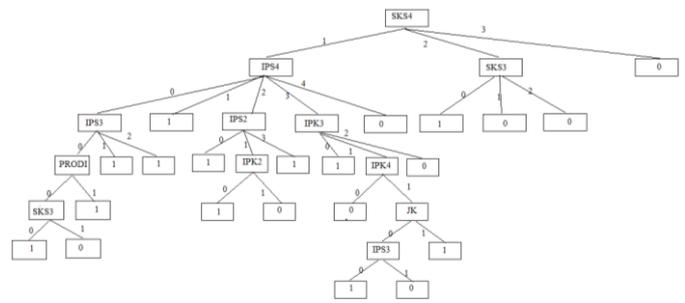
TABEL 6. HOSMER AND LOMESHOW TEST

Step	Chi-square	df	Sig.
1	5,266	8	,729

Pada **Error! Reference source not found.** terlihat bahwa nilai sig  $> 0.05$  yang artinya model yang dibuat terbukti fit.

Berdasarkan data dan atribut yang telah didapat dari proses sebelumnya, proses selanjutnya adalah melakukan pengolahan data dengan metode *Decision tree*. Untuk melakukan prediksi dibutuhkan data latih dan data uji. Data latih berfungsi sebagai data pembelajaran dan data uji berfungsi untuk menguji model yang dihasilkan dari data latih. Pada penelitian ini data latih yang digunakan sebanyak 70% dan data ujinya sebanyak 20% atau setara dengan 21 data.

Pengujian data kelulusan mahasiswa untuk memprediksi kelulusan mahasiswa tepat waktu dengan metode *Decision tree* dilakukan dengan menggunakan bahasa pemrograman *Python* pada *tools Google Colab*. **Error! Reference source not found.** merupakan gambar pohon keputusan prediksi kelulusan tepat waktu.



Gambar 3. Pohon keputusan prediksi kelulusan tepat waktu

Keterangan:

JK (Jenis : 0 (laki-laki) dan 1 (perempuan)

Kelamin)

Usia : 0 ( $\geq 21$ ), 1 (19-21), 2 (16-21), dan 3 ( $\leq 15$ )

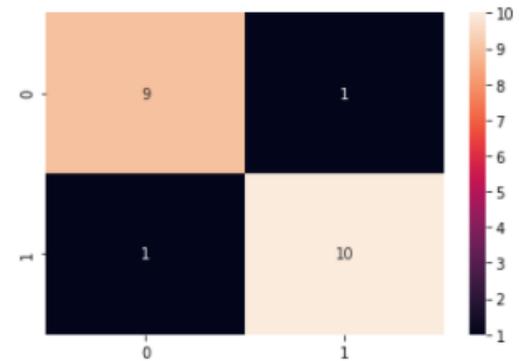
Prodi : 0 (sistem informasi) dan 1 (manajemen)

IPK dan IPS : 0 ( $\geq 3.51$ ), 1 (3.01-3.50), 2 (2.76-3.00), 3 (1-4), dan 4 ( $\leq 1.99$ )

SKS (1-4) : 0 (22-24), 1 (19-21), 2 (16-18), dan 3 ( $\leq 15$ )

Masa Studi : 0 (terlambat) dan 1 (tepat)

Berdasarkan pemodelan yang dilakukan dengan *Decision tree* C4.5 dengan data sebanyak 70 data. Di mana data uji yang digunakan sebanyak 30% atau setara dengan 21 data dan data latih sebanyak 70% atau setara dengan 49 data. **Error! Reference source not found.** merupakan perbandingan antara data aktual dan prediksi.



Gambar 4. Confusion matrix

Keterangan:

0-0 : Jumlah data negatif yang diprediksi negatif sebanyak 9 data

1-0 : Jumlah data positif yang diprediksi negatif sebanyak 1 data

0-1 : Jumlah data negatif yang diprediksi positif sebanyak 1 data

1-1 : Jumlah data positif yang diprediksi positif sebanyak 10 data

Model tersebut menghasilkan akurasi sebesar 90% dengan SKS4 sebagai *root* atau sebagai atribut yang paling berpengaruh. Adapun interpretasi dari model yang telah dibuat disajikan pada **Error! Reference source not found.** berikut:

TABEL 7. INTERPRETASI

Program Studi	Rules
Sistem Informasi	Jika SKS4 sebanyak 19-21 SKS, IPS4 lebih besar atau sama dengan dari 3,51, IPS3 lebih besar atau sama dengan 3,51, Prodi Sistem Informasi, SKS3 sebanyak 22-24 maka diprediksi "lulus tepat waktu"
Manajemen	Jika SKS4 sebanyak 19-21 SKS, IPS4 lebih besar atau sama dengan dari 3,51, IPS3 lebih besar atau sama dengan 3,51, Prodi Sistem Informasi, SKS3 sebanyak 19-21 maka diprediksi "tidak lulus tepat waktu atau terlambat"
Sistem Informasi dan Manajemen	<p>Jika SKS4 sebanyak 19-21 SKS, IPS4 lebih besar atau sama dengan dari 3,51, IPS3 lebih besar atau sama dengan 3,51, Prodi Manajemen maka diprediksi "lulus tepat waktu"</p> <p>Jika SKS4 sebanyak 19-21 SKS, IPS4 lebih besar atau sama dengan dari 3,51, IPS3 berada diantara rentang nilai 3,01-3,50 maka diprediksi "lulus tepat waktu"</p> <p>Jika SKS4 sebanyak 19-21 SKS, IPS4 lebih besar atau sama dengan dari 3,51, IPS3 berada diantara rentang nilai 2,76-3,00 maka diprediksi "lulus tepat waktu"</p> <p>Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 3,01-3,50 maka diprediksi "lulus tepat waktu"</p> <p>Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2,76-3,00, dan IPS2 lebih dari atau sama dengan 3,51 maka diprediksi "lulus tepat waktu"</p> <p>Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2,76-3,00, dan IPS2 lebih dari atau sama dengan 3,51, IPK2 lebih dari 3,51 maka diprediksi "lulus tepat waktu"</p> <p>Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2,76-3,00, dan IPS2 berada pada rentang nilai 2,00-2,75 maka diprediksi "lulus tepat waktu"</p> <p>Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2,00-2,75, IPK3 lebih dari atau sama dengan 3,51 maka diprediksi "lulus tepat waktu"</p> <p>Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2,00-2,75, IPK3 berada pada rentang nilai 3,01-3,50, IPK4 lebih dari atau sama dengan 3,51 maka diprediksi "lulus tepat waktu"</p> <p>Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2,00-2,75, IPK3 berada pada rentang nilai 3,01-3,50, IPK4 berada pada rentang nilai 3,01-3,50, JK (Jenis Kelamin) laki-laki, IPS3 lebih dari 3,51 maka diprediksi "lulus tepat waktu"</p> <p>Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2,00-2,75, IPK3 berada pada rentang nilai 3,01-3,50, IPK4 berada pada rentang nilai 3,01-3,50, JK (Jenis Kelamin) laki-laki, IPS3 berada pada rentang nilai 3,01-3,51 maka diprediksi "tidak lulus tepat waktu atau terlambat"</p> <p>Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2,00-2,75, IPK3 berada pada rentang nilai 3,01-3,50, IPK4 berada pada rentang nilai 3,01-3,50, JK (Jenis Kelamin) laki-laki, IPS3 berada pada rentang nilai 3,01-3,51 maka diprediksi "tidak lulus tepat waktu atau terlambat"</p> <p>Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2,00-2,75, IPK3 berada pada rentang nilai 2,76-3,00 maka diprediksi "lulus tepat waktu"</p> <p>Jika SKS4 sebanyak 19-21 SKS, IPS4 lebih kecil dari</p>

Program Studi	Rules
	<p>atau sama dengan 1,99 maka diprediksi "tidak lulus tepat waktu atau terlambat"</p> <p>Jika SKS4 sebanyak 19-21 SKS, SKS3 berada pada rentang 22-24 SKS maka diprediksi "lulus tepat waktu"</p> <p>Jika SKS4 sebanyak 19-21 SKS, SKS3 berada pada rentang 19-21 "tidak lulus tepat waktu atau terlambat"</p> <p>Jika SKS4 sebanyak 19-21 SKS, SKS3 berada pada rentang 16-18 maka diprediksi "tidak lulus tepat waktu atau terlambat"</p> <p>Jika SKS4 kurang atau sama dengan 15 maka diprediksi "tidak lulus tepat waktu atau terlambat"</p>

#### IV. KESIMPULAN

Berdasarkan seluruh hasil dari tahapan penelitian yang telah dilakukan pada penerapan metode *Decision tree* C4.5 untuk prediksi kelulusan mahasiswa STIMIK ESQ dapat ditarik kesimpulan sebagai berikut:

1. Faktor yang mempengaruhi kelulusan tepat waktu di STIMIK ESQ berdasarkan pohon keputusan *decision tree* C4.5 yaitu SKS4, IPS4, SKS3, IPS3, IPS2, IPK3, Prodi, IPK2, IPK4, dan Jenis Kelamin.

2. Prediksi kelulusan mahasiswa dapat dilakukan dengan menggunakan pendekatan data mining dengan metode *decision tree* C4.5 yang menghasilkan 21 *rules* dengan tingkat akurasi sebesar 90%.

#### DAFTAR PUSTAKA

- [1] Badan Akreditasi Nasional Perguruan Tinggi, "Naskah IAPT 3.0," no. April, pp. 7–9, 2019.
- [2] M. S. Maulana, R. Sabarudin, and W. Nugraha, "Prediksi Ketepatan Kelulusan Mahasiswa Diploma dengan Komparasi Algoritma Klasifikasi," *J. Sist. dan Teknol. Inf.*, vol. 7, no. 3, p. 202, 2019, doi: 10.26418/justin.v7i3.33316.
- [3] H. Romadhona, A., Suprapedi, S. dan Himawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma Decision Tree," *J. Teknol. Inf.*, vol. 13, pp. 69–83, 2017.
- [4] A. F. A. Rahman, Sorikhi, and S. Wartulas, "Prediksi kelulusan mahasiswa menggunakan algoritma c4.5 (studi kasus di universitas peradaban)," vol. 1, no. 2, pp. 70–77, 2020.
- [5] S. F. Ulya, Y. Sukestiyono, and P. Hendikawati, "RANDOM SAMPLING CONFIDENCE INTERVAL," vol. 7, no. 1, pp. 108–119, 2018.
- [6] P. K. Arieska, N. Herdiani, S. Sampling, and E. Relatif, "PEMILIHAN TEKNIK SAMPLING BERDASARKAN," vol. 6, no. 2, 2018.
- [7] W. Wirawan, A. Aghastya, and A. L. Lailya, "No Title," vol. III, pp. 55–61, 2019.
- [8] Y. A. Tampil, H. Komalig, and Y. Langi, "Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif ( IPK ) Mahasiswa FMIPA Universitas Sam Ratulangi Manado Logistic Regression Analysis To Determine Factors Affecting The Grade Point Average ( GPA ) Of FM," 2015.
- [9] M. Hamidah, H. Fitriyah, and I. Arwani, "Implementasi Decision Tree pada Penentuan Kondisi Ruang Berasap Menggunakan Multi-Sensor Berbasis Arduino Uno," vol. 3, no. 4, pp. 3845–3854, 2019.
- [10] I. M. Kamal, T. Hendro, and R. Ilyas, "Prediksi Penjualan Buku Menggunakan Data Mining Di Pt. Niaga Swadaya," *Semin. Nas. Teknol. Inf. Multimed.*, vol. 02, no. February, pp. 49–54, 2017.
- [11] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, "Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5," *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.