

**KLASIFIKASI WAKTU KELULUSAN MAHASISWA
MENGUNAKAN METODE *DECISION TREE***
(Studi Kasus : Fakultas Ilmu Hukum Universitas Ichsan Gorontalo)

Oleh

GEBY ANGGRIANI MUNDOK

T3119071

SKRIPSI

Untuk memenuhi salah satu syarat ujian

Guna memperoleh gelar sarjana



**PROGRAM SARJANA
TEKNIK INFORMATIKA
UNIVERSITAS ICHSAN GORONTALO
GORONTALO
2023**

PERSETUJUAN SKRIPSI

KLASIFIKASI WAKTU KELULUSAN MAHASIWA MENGUNAKAN METODE *DECISION TREE*

(Studi Kasus: Fakultas Ilmu Hukum Universitas Ichsan Gorontalo)

Oleh

GEBY ANGGRIANI MUNDOK

T3119071

SKRIPSI

Untuk memenuhi salah satu syarat ujian
guna memperoleh gelar Sarjana Program
Studi Teknik Informatika, ini telah
disetujui oleh Tim Pembimbing.

Gorontalo, November 2023

Pembimbing I



Amiruddin, M.Kom, MCF
NIDN. 0910097601

Pembimbing II



Zulfrianto Y. Lamasigi, M.Kom
NIDN. 0914089101

PENGESAHAN SKRIPSI

KLASIFIKASI WAKTU KELULUSAN MAHASIWA MENGUNAKAN METODE *DECISION TREE*

Oleh

GEBY ANGGRIANI MUNDOK

T3119071

Diperiksa oleh Panitia Ujian Strata Satu (S1)

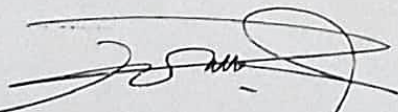
Universitas Ichsan Gorontalo

1. Ketua Penguji
Zohrahayaty, M.Kom
2. Anggota
Rofiq Harun, M.Kom
3. Anggota
Hastuti Dalai, M.Kom
4. Anggota
Amiruddin, M.Kom, MCF
5. Anggota
Zilfianto Y. Lamasigi, M.Kom

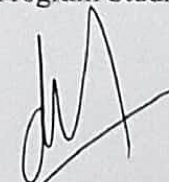
Mengetahui,

Dekan Fakultas Ilmu Komputer

Ketua Program Studi



Irvan Abraham Salihi, M.Kom
NIDN : 0918077302



Sudirman S. Panna, M.Kom
NIDN : 0924038025

PERNYATAAN SKRIPSI

Dengan ini saya menyatakan bahwa:

1. Karya tulis (Skripsi) saya ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) baik di Universitas Ichsan Gorontalo maupun di perguruan tinggi lainnya.
2. Karya tulis (Skripsi) saya ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan Tim Pembimbing.
3. Dalam karya tulis (Skripsi) saya ini tidak terdapat karya atau pendapat yang telah dipublikasikan orang lain, kecuali secara tertulis dicantumkan sebagai acuan/sitasi dalam naskah dan dicantumkan pula dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma- norma yang berlaku di Univeristas Ichsan Gorontalo.

Gorontalo, November 2023

Meterai
100

Geby Anggriani Mundok

ABSTRACT

GEBY ANGGRIANI MUNDOK. T3119071. THE CLASSIFICATION OF STUDENT GRADUATION TIME USING THE DECISION TREE METHOD

A student's graduation time is one indicator of the university's success. Student graduation time is the period required for students to complete their studies. Ideally, students can graduate on time under a specified study period. However, not all students can graduate on time. This study is aimed at classifying whether students graduate on time using the decision tree method. The data employed in this study are the Faculty of Law's graduates of 2015 - 2017, Universitas Ichsan Gorontalo. The attributes used in this classification consist of IPS1, PIS2, IPS3, IPS4, gender, and graduation information. In this study, model optimization performed is by selecting attributes, pruning trees, and measuring inside the tree. The results of this study show that the decision tree method can predict student graduation times with 92% accuracy by producing nine (9) decision rules. It indicates that the decision tree method can be a solution for predicting student graduation times, so it can be a solution to help study programs increase student success in completing their studies.

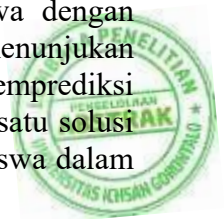


Keywords: classification, graduation time, students, data mining, decision tree, accurate graduation

ABSTRAK

GEBY ANGGRIANI MUNDOK. T3119071. KLASIFIKASI WAKTU KELULUSAN MAHASISWA MENGGUNAKAN METODE *DECISION* TREE

Waktu kelulusan mahasiswa merupakan salah satu indikator keberhasilan perguruan tinggi. Waktu kelulusan mahasiswa merupakan periode waktu yang dibutuhkan mahasiswa untuk menyelesaikan studinya. Idealnya, mahasiswa dapat lulus tepat waktu sesuai dengan masa studi yang ditetapkan. Namun, pada kenyataannya, tidak semua mahasiswa dapat lulus tepat waktu. Peneliti ini bertujuan untuk mengklasifikasi waktu kelulusan mahasiswa apakah lulus tepat waktu atau tidak dengan menggunakan metode *decision tree*. Data yang digunakan dalam penelitian ini adalah data lulusan Fakultas Ilmu Hukum Universitas Ichsan Grontalo dari tahun 2015 hingga 2017 ada. Adapun atribut yang digunakan dalam klasifikasi ini terdiri dari: IPS1, PIS2, IPS3, IPS4, jenis kelamin, keterangan lulus. Pada penelitian ini dilakukan juga optimalisasi model dengan melakukan seleksi atribut, pemangkasan pohon serta pengukuran kedalaman pohon. Hasil penelitian menunjukkan bahwa metode *decision tree* dapat memprediksi waktu kelulusan mahasiswa dengan akurasi 92% dengan menghasilkan sebanyak 9 rule keputusan. Hal ini menunjukkan bahwa metode *decision tree* dapat menjadikan salah satu solusi untuk memprediksi waktu kelulusan mahasiswa, sehingga metode ini dapat menjadi salah satu solusi untuk membantu program studi dalam meningkatkan keberhasilan mahasiswa dalam menyelesaikan studinya.



Kata kunci: klasifikasi, waktu kelulusan, mahasiswa, data mining, *decision tree*, lulus tepat

KATA PENGANTAR

Alhamdulillah, penulis dapat menyelesaikan skripsi ini dengan judul: **“klasifikasi Waktu Kelulusan Mahasiswa Menggunakan Metode *Decision Tree* Studi kasus : Fakultas Ilmu Hukum Unisan Gorontalo”**, untuk memenuhi salah satu syarat Ujian Akhir guna memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.

Penulis Menyadari sepenuhnya bahwa skripsi ini tidak mungkin terwujud tanpa bantuan dan dorongan dari berbagai pihak, baik bantuan moril maupun materil. Untuk itu dengan segala keikhlasan dan kerendahan hati, penulis mengucapkan banyak terima kasih dan penghargaan yang setinggi-tingginya kepada :

1. Ibu Dr. Dra. Juriko Abdussamad, M.Si, selaku ketua Yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo;
2. Bapak Dr. Abdul Gaffar La Tjokke, M.Si, selaku Rektor Universitas Ichsan Gorontalo;
3. Bapak Irvan A. Salihi, S.Kom, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
4. Bapak Sudirman Melangi, M.Kom, selaku Wakil Dekan I Bidang Akademik Fakultas Ilmu Komputer;
5. Ibu Irma Surya Kumala, M.Kom, selaku Wakil Dekan II Bidang Administrasi Umum dan Keuangan Fakultas Ilmu Komputer;
6. Bapak Sudirman S. Panna, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
7. Bapak Amiruddin, M.Kom,MFC selaku Pembimbing I yang telah banyak membimbing dan membantu penulis dalam menyelesaikan skripsi;
8. Bapak Zulfrianto Lamasigi, M.Kom, selaku Pembimbing II yang telah banyak membimbing dan membantu dalam menyelesaikan skripsi;
9. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang telah mendidik dan mengajarkan berbagai disiplin ilmu kepada penulis;

10. Kedua Orang Tua paling berjasa dalam hidup penulis (Bapak Selamat Mundok dan Ibu Nur Ain Dalusa) Terimakasih atas kepercayaan yang telah diberikan kepada saya untuk melanjutkan pendidikan kuliah ,serta cinta do'a motivasi, semangat dan nasihat yang tidak hentinya diberikan kepada anaknya dalam penyusunan skripsi ini;
11. Terimakasih banyak untuk keluarga besar yang selalu memberikan dukungan baik secara moral maupun material ;
12. Rekan-rekan seperjuangan, khususnya Ana Fitriana yang telah banyak memberikan bantuan dan dukungan yang sangat besar kepada penulis selama proses pembuatan penelitian ini;
13. Terakhir, terimakasih untuk diri sendiri, karena telah mampu berusaha keras dan berjuang sejauh ini. Mampu mengendalikan diri dari berbagai tekanan diluar keadaan dan tak pernah memutuskan menyerah sesulit apapun proses penyusunan skripsi ini dengan menyelesaikan sebaik dan semaksimal mungkin, ini merupakan pencapaian yang patut dibanggakan untuk diri sendiri.

Semoga Allah, SWT melimpahkan balasan atas jasa-jasa mereka kepada kami. Penulis menyadari sepenuhnya bahwa apa yang telah dicapai ini masih jauh dari kesempurnaan dan masih banyak terdapat kekurangan, oleh karena itu, penulis sangat mengharapkan adanya kritik dan saran yang konstruktif. Akhirnya penulis berharap semoga hasil yang telah dicapai ini dapat bermanfaat bagi kita semua, Amiin.

Gorontalo, November

Penulis

DAFTAR ISI

PERSETUJUAN SKRIPSI.....	ii
PENGESAHAN SKRIPSI.....	iii
PERNYATAAN SKRIPSI.....	iv
ABSTRAK	v
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	3
1.3 Rumusan Masalah	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.5.1 Manfaat Teoritis.....	4
1.5.2 Manfaat Praktis	4
BAB II LANDASAN TEORI	5
2.1 Tinjauan Studi	5
2.1.1 Data Mining	6
2.1.2 Klasifikasi	7
2.1.3 <i>Decision Tree</i>	9
2.1.4 Algoritma C4.5	9
2.1.5 Penerapan Algoritma <i>Decision Tree C4.5</i>	10
2.1.6 Python	20
2.1.7 <i>Confusion Marix</i>	21

2.2 Kerangka Pikir.....	22
BAB III METODE PENELITIAN	23
3.1 Jenis, Metode, Subjek, Objek, Waktu, dan Lokasi Penelitian.....	23
3.2 Pengumpulan Data	23
3.3 Pemodelan/Abstraksi.....	24
3.4 Pra Pengolahan Data	24
3.5 Validasi.....	24
3.5.1 Pengembangan Model.....	25
BAB IV HASIL PENELITIAN.....	26
4.1 Hasil Pengumpulan Data	26
4.2 Hasil Pemodelan.....	27
4.3 Hasil Proses Pemodelan	37
4.3.1 Tahapan Seleksi	37
4.3.2 Preprocessing	38
4.3.3 Traspormasi	40
4.3.4 Proses Data Mining.....	42
4.3.5 Evaluasi.....	43
4.3.6 Perbaikan Model	47
4.3.7 Klasifikasi Data Testing Baru.....	48
BAB V PEMBAHASAN PENELITIAN	50
5.1 Pembahasan Kinerja Model.....	50
5.2 Pembahasan Hasil Pemodelan	53
5.3 Hasil Pemodelan Setelah Optimalisasi	54
BAB VI KESIMPULAN DAN SARAN.....	57
DAFTAR PUSTAKA	58
LAMPIRAN.....	61

DAFTAR GAMBAR

Gambar 1.1 Grafik Waktu Kelulusan Mahasiswa	2
Gambar 2. 1 Proses Data Mining	6
Gambar 2. 2 Proses Algoritma C4.5	9
Gambar 2. 3 Atribut Data Terpilih	18
Gambar 2. 4 Kurva ROC	20
Gambar 2. 5 Rumus <i>Confussion Matrix</i> 2 Kelas.....	21
Gambar 3. 1 Pengembangan Model Klasifikasi Waktu Kelulusan Mahasiswa	25
Gambar 4. 1 Grafik Jumlah kelas output	30
Gambar 4. 2 Grafik Jumlah Kelas Setelah di Perbaikan	31
Gambar 4. 3 Hasil Perhitungan Confusion Matrix.....	33
Gambar 4. 4 Visualisasi pemodelan <i>Decision tree</i>	33
Gambar 4. 5 Grafik Kepentingan Atribut.....	34
Gambar 4. 6 hasil confusion matrix setelah optimalisasi	36
Gambar 4. 7 Visualisasi model setelah optimalisasi	36
Gambar 4. 8 Gambar Grafik	40
Gambar 4. 9 konversi nilai string numerik	41
Gambar 4. 10 Menghitung Data Statistik.....	41
Gambar 4. 11 Confusion Matrix	43
Gambar 4. 12 Seleksi Atribut	44
Gambar 4. 13 Gambar setelah Perubahan	45
Gambar 4. 14 Gambar mengganti kelas sesuai kondisi	45
Gambar 4. 15 Pengukuran Kedalaman Pohon	47
Gambar 4. 16 Gambar Confusion Matrix	48
Gambar 4. 17 Gambar Prediksi model dengan data dari Excel	49
Gambar 4. 18 Gambar Mengkonversi File Excel.....	49
Gambar 5. 1 <i>Confusion Matrix</i>	50
Gambar 5. 2 Visualisasi Pohon Keputusan Sebelum Optimalisasi.....	53
Gambar 5. 3 Visualisasi Pohon Setelah Optimalisasi.....	54

DAFTAR TABEL

Tabel 2. 1	Penelitian Terkait	5
Tabel 2. 2	Data Mahasiswa STMIK Rosma Karawang.....	11
Tabel 2. 3	Deskripsi Atribut Data	12
Tabel 2. 4	<i>Missing Value</i>	13
Tabel 2. 5	<i>Missing Value Data Mahasiswa 2012 -2016</i>	15
Tabel 2. 6	<i>Hasil Cleaning Data Mahasiswa 2012-2016</i>	15
Tabel 2. 7	<i>Atribut Baru</i>	16
Tabel 2. 8	Atribut Baru Mahasiswa STMIK Rosma Karawang.....	16
Tabel 2. 9	<i>Hasil Transformasi Atribut</i>	17
Tabel 2. 10	Atribut Data Terpilih	17
Tabel 3. 1	Variabel Atribut Data	23
Tabel 3. 2	Data Kelulusan Mahasiswa Fakultas Ilmu Hukum, 2019	24
Tabel 4. 1	Hasil Pengumpulan data	26
Tabel 4. 2	Dataset Hasil Olahan	28
Tabel 4. 3	Data Tidak konsisten	31
Tabel 4. 4	Tabel Dataset Hasil Trasformasi	32
Tabel 4. 5	Hasil Perhitungan Kedalaman Pohon dan Akurasi.....	35
Tabel 4. 6	Data Testing Baru	37
Tabel 4. 7	Hasil Klasifikasi Data Testing Baru	37
Tabel 4. 8	Hasil pembacaan Dataset dari file Excel	38
Tabel 5. 1	Hasil Counfusion Matrix Sesudah Optimalisasi.....	50
Tabel 5. 2	Hasil Counfusion Matrix Sesudah Optimalisasi.....	51
Tabel 5. 3	Rule Keputusan	55

BAB I

PENDAHULUAN

1.1 Latar Belakang

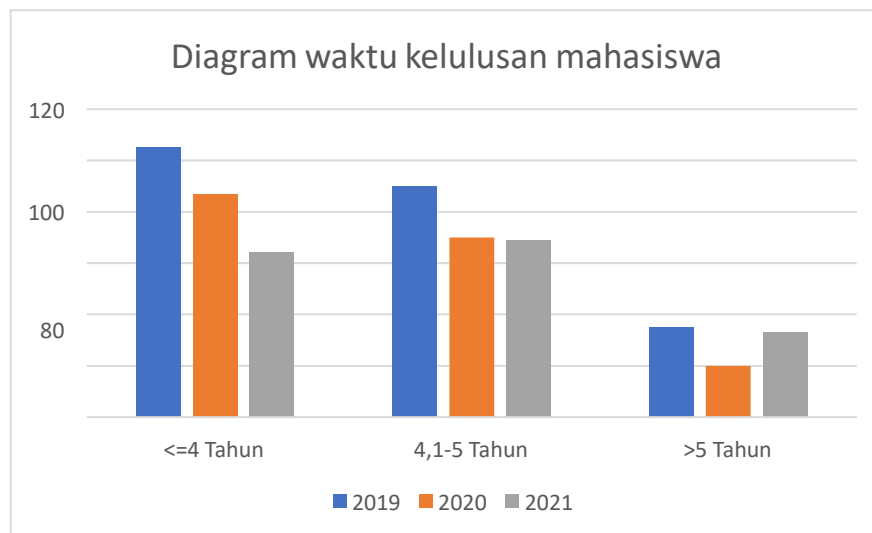
Perkembangan teknologi di zaman pada saat ini semakin berkembang terutama pada Pendidikan. Seiring meningkatnya mutu Pendidikan di tingkat sekolah tinggi atau universitas maka di butuhkan laporan yang dapat mendukung peningkatan prestasi akademik mahasiswa. Laporan ini dapat di peroleh dari data nilai mata kuliah, dan indek prestasi mahasiswa dan masih banyak lagi data-data yang dapat di manfaatkan untuk mendapatkan informasi yang dapat mendukung peningkatan prestasi akademik mahasiwa.

Evaluasi masa studi yang dipengaruhi oleh berbagai keadaan menjadi kriteria dalam penilaian akreditasi program studi. Berdasarkan penelitian ini, pengkategorian kelulusan mahasiswa akan dilakukan berdasarkan I.P. semester I, semester II, semester III, dan semester IV. Jika sebuah sistem dapat mengantisipasi atau memperkirakan kapan mahasiswa akan lulus.

Universitas Ichsan Gorontalo adalah salah satu perguruan tinggi suwasta yang berada di kota Gorontalo. Universitas ichsan Gorontalo memiliki beberapa fakultas salah satunya adalah fakultas hukum yang pada saat ini kesulitan dalam memprediksi waktu kelulusan mahasiswa untuk itu di butuhnya klasifikasi yang dapat di gunakan untuk menentukan waktu kelulusan mahasiswa.

Permasalahan yang di hadapi oleh fakultas hukum adalah jumlah mahasiwa yang masuk dan keluar setiap tahun tidak sebanding dan akan mempengaruhi akreditas prodi tersebut dengan berdasarkan adanya system yang dapat digunakan untuk mengetahui waktu kelulusan mahsiswa, apabila waktu kelulusan dapat di ketahui sejak dini maka pihak akademik dapat menerapkan suatu kebijakan untuk meminimalisir jumlah mahasiwa yang tidak lulusan tepat waktu sesuai dengan masa studinya. Dalam perguruan tinggi mahasiwa merupakan asset yang paling penting

bagi sebuah fakultas oleh karena itu di perlukan tingkat kelulusan mahasiswa tepat waktu.



Sumber : Fakultas Ilmu Hukum UNISAN Gorontalo, 2022

Gambar 1.1 Grafik Waktu Kelulusan

Berdasarkan Gambar 1.1 di atas terlihat bagaimana kelulusan mahasiswa dengan tahun yang berbeda. Jika dilihat dengan seksama mahasiswa yang lulus dari tahun ke tahun rata-rata mengalami peningkatan yang cukup signifikan. Dan juga bisa dilihat terjadi penurunan jumlah wisudawan dari tahun ketahun, sehingga pada kejadian ini dapat di klasifikasikan menggunakan metode *decision tree* untuk mengelompokkan waktu kelulusan mahasiswa di fakultas hukum Universitas Ichsan Gorontalo.

Adapun penelitian yang terkait dengan klasifikasi waktu kelulusan mahasiswa yaitu penelitian dari Reza Maulana, Devi Kumalasari dengan judul “ Analisis Komparasi Agoritma Klasifikasi *Data Mining* Untuk Prediksi Status Kelulusan Mahasiswa Akademik Bina Sarana Informatika”. Penelitian tersebut dilakukan dengan menggunakan metode K-NN, *Decision Tree*, *Naïve Bayes*, *Random Forest* dan *Random Tree* untuk melakukan klasifikasi data mining, untuk memprediksi status kelulusan mahasiswa

Menggunakan algoritma C4.5 dengan pendekatan *data mining* adalah cara

terbaik untuk menangani data-data ini. Nama, mahasiswa, nim, jenis kelamin, jumlah SKS, IPK, dan alamat adalah atribut yang digunakan. Berdasarkan karakteristik tersebut, prosedur pengolahan data dapat membantu institusi pendidikan dalam upaya mendorong mahasiswa untuk meningkatkan nilai IPK dan dapat mendukung akreditasi Fakultas Hukum Universitas Ichsan Gorontalo [2].

Ada beberapa algoritma C4.5 yang digunakan dalam pendekatan klasifikasi data mining yang menggambarkan kategorisasi data sebagai pohon keputusan. Manfaat utama dari metode C4.5 adalah bahwa metode ini dapat menghasilkan pohon keputusan yang cukup akurat dan efisien dalam menangani jenis karakteristik diskrit dan numerik [2].

Algoritma *Decision Tree* (pohon keputusan) adalah suatu algoritma yang mempermudah dalam melakukan pengelompokan karena algoritma ini dapat meningkatkan model prediktif dengan akurasi, kemudahan dalam interpretasi dan stabilitas, serta penilaian antar kriteria dan subkriteria [3].

Pentingnya waktu kelulusan mahasiswa membuat munculnya berbagai penelitian dengan tema ini. Salah satunya adalah penelitian yang berjudul **“Klasifikasi Waktu Kelulusan Mahasiswa Fakultas Hukum Universitas Ichsan Grontalo Dengan Metode *Decision Tree*”**.

1.2 Identifikasi Masalah

Berdasarkan uraian pada latar belakang maka identifikasi masalah dalam penelitian ini adalah melakukan klasifikasi waktu kelulusan mahasiswa di fakultas hukum Universitas Ichsan Gorontalo menggunakan teknik *data mining*.

1.3 Rumusan Masalah

Dari pengertian dan penjelasan masalah di latar belakang maka di peroleh rumusan masalah berupa:

1. Bagaimana penerapan metode Decision Tree Klasifikasi waktu kelulusan mahasiswa Fakultas Hukum ?
2. Bagaimana tingkat akurasi penerapan metode *Decision Tree* dalam klasifikasi waktu kelulusan mahasiswa pada Fakultas Hukum?

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah :

1. Menerapkan metode *Decision Tree* dalam Klasifikasi Waktu Kelulusan Mahasiswa Fakultas Hukum
2. Mengukur tingkat akurasi penerapan Klasifikasi Waktu Kelulusan Mahasiswa menggunakan metode *Decision Tree*

1.5 Manfaat Penelitian

Manfaat penelitian ini menunjukkan apa yang dapat diambil dari hasil penelitian itu.

1.5.1 Manfaat Teoritis

Penelitian ini dapat mengembangkan ilmu pengetahuan, dengan penerapan metode *decision tree*.

1.5.2 Manfaat Praktis

Penelitian ini mendukung proses klasifikasi waktu kelulusan mahasiswa dalam rangka menghasilkan kualitas software yang baik dan memberikan dampak juga pada peningkatan kualitas perguruan tinggi.

BAB II LANDASAN TEORI

2.1 Tinjauan Studi

Merencanakan sistem klasifikasi waktu kelulusan mahasiswa menggunakan Bahasa pemrograman serta menerapkan algoritma *decision tree*.

Tabel 2. 1 Penelitian Terkait

No	PENELITI	JUDUL	TAHUN	METODE	HASIL
1	Reza Maulana ¹ , Devi Kumalasari ² [1].	Analisis Komparasi Agoritma Klasifikasi Data Mining Untuk Prediksi Status Kelulusan Mahasiswa Akademik Bina Sarana Informatika	2019	Menggunakan Metode <i>Decision Tree</i>	Penelitian ini dilakukan dengan menggunakan Metode K-NN, <i>Decision Tree</i> , <i>Naïve Bayes</i> , Random Forest dan Random Tree untuk melakukan klasifikasi data mining untuk memprediksi status kelulusan mahasiswa. Dalam penelitian ini mendapatkan nilai akurasi sebesar 99,97% dari metode <i>Decision Tree</i> yang telah diterapkan.

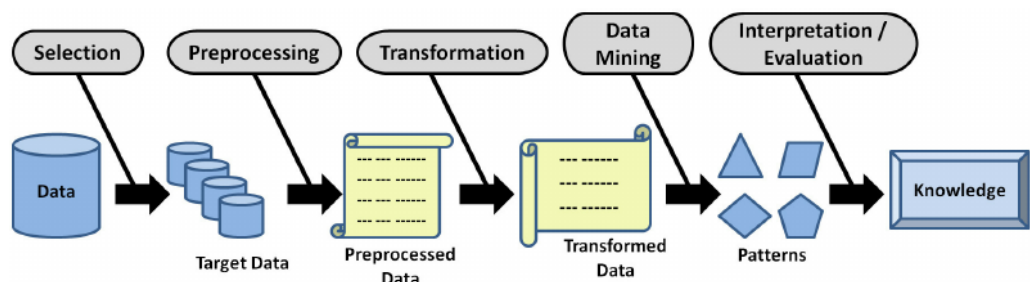
2.1.1 Data Mining

Data mining adalah rangkaian proses yang mengikuti nilai tambah dari satu kumpulan data dalam bentuk pengetahuan. Menurut Ham dan Kamber, “Data mining merupakan proses mengikuti pengetahuan dari sejumlah data besar [6].

Pengorganisasian atau klasifikasi pola yang harus ditemukan dalam data mining adalah tujuan dari data mining. Data mining sering kali dibagi menjadi dua kelompok: deskriptif dan prediktif [6].

Ketentuan data mining biasanya diklasifikasi dengan sistem pendukung keputusan (DSS). Seandainya dalam aplikasi bisnis, informasi yang dihasilkan oleh data mining dapat diklasifikasi dengan alat pengelolaan usaha [6] untuk pengoperasian dan menguji promosi pemesanan yang baik. Kombinasi ini membutuhkan tindakan pasca-pemrosesan untuk menentukan hasil yang menarik dan bermanfaat yang dapat dipersatukan dengan DSS.

Data Mining memiliki Langkah *knowledge discovery in database* (KDD). *Knowledge discovery in database* (KDD) adalah keseluruhan proses menemukan dan pengenalan pola (pattern) dalam data yang sederhana, dan menemukan pola agar baik dan dapat bermanfaat dan mudah dipahami.



Gambar 2. 1 Proses Data Mining

1. Data Selection

Sebelum memulai tahap penambangan data KDD (Knowledge Discovery in Database), pemilihan data dari dataset operasional harus diselesaikan. Hasil seleksi data disimpan dalam file terpisah dari basis data operasional dan digunakan nanti dalam penambangan data [7].

2. *Preprocessing*

Data yang menjadi fokus Knowledge Discovery dalam database harus dibersihkan sebelum melakukan prosedur Knowledge Mining (KDD). Proses pembersihan melibatkan beberapa langkah, seperti menghapus data duplikat, menguji data yang bertentangan, dan memperbaiki celah data seperti kesalahan ketik [7].

3. *Transformation*

Pengkodean data yang dipilih adalah prosedur transformasi yang membuat data dapat digunakan dalam proses penambangan data. Metode pengkodean KDD (Knowledge Discovery in Databases) adalah prosedur inventif yang sangat bergantung pada jenis atau pola data yang diambil dari database [8].

4. *Data Mining*

Data mining adalah metode untuk menemukan pengetahuan yang relevan dan informasi yang dapat digunakan dalam kumpulan data yang sangat besar dengan menggunakan pendekatan matematika, kecerdasan buatan, statistik, dan pembelajaran mesin. [9].

5. *Interpretation / Evaluation*

Sangatlah penting untuk memberikan pola informasi proses data mining dengan cara yang mudah dimengerti. Langkah ini melibatkan penentuan apakah pola atau data yang dihasilkan nantinya bertentangan dengan kebenaran atau asumsi yang telah diketahui. [10].

2.1.2 **Klasifikasi**

Menurut etimologinya, istilah "klasifikasi" berasal dari kata kerja "to classify", yang berarti mengkategorikan atau mengatur segala sesuatu pada posisi yang tepat. Frasa ini menggambarkan teknik untuk mengatur data secara metodis [11].

Menurut *Harrolds Librarians Glossary*, Towa P. Hamakonda, dan J.N.B Tairas (2022) [12]. menjelaskan dalam *Harrolds Librarians Glossary* bahwa kategorisasi adalah pengelompokan secara sistematis dan logis terhadap benda-benda, konsep-konsep, buku-buku, atau hal-hal lain ke dalam kelas-kelas atau kelompok-kelompok tertentu berdasarkan kualitas-kualitas yang sama.

Proses Klasifikasi memiliki beberapa tahapan dalam melakukan *data mining*:

1. Validasi (*Validation*)

Validasi adalah sebuah proses pembagian data awal menjadi dua bagian data, *data training* dan *data testing*. *Data training* adalah data yang akan diolah menggunakan algoritma klasifikasi dan *data testing* adalah data yang akan digunakan dalam proses pengujian menggunakan program komputer. Proses validasi dalam klasifikasi terbagi menjadi dua proses yaitu *cross validation* dan *split validation*. *Cross validation* melakukan pembagian data secara acak sedangkan *split validation* melakukan pembagian data menjadi dua bagian berdasarkan jumlah data.

2. Implementasi Algoritma Klasifikasi

Proses klasifikasi melakukan pengolahan data dengan mengimplementasikan algoritma klasifikasi seperti *support vector machine*, *decision tree*, *k-means* dan *naive baiyes*.

3. *Apply Model*

Apply model digunakan untuk penerapan model klasifikasi yang telah dilatih sebelumnya menggunakan *data training* yang juga akan digunakan dalam pengujian *data testing*.

4. *Performance*

Performance adalah hasil dari pengujian yang dilakukan dalam proses klasifikasi. Hasil yang diukur pada tahap *performance* seperti akurasi, *precision*, dan *recall*.

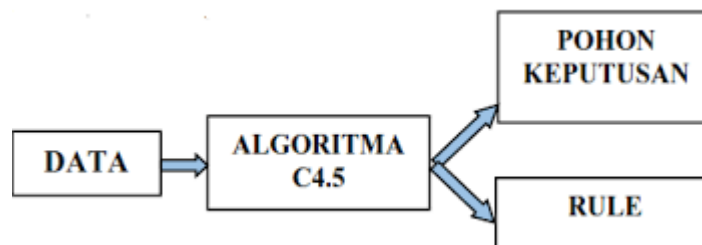
2.1.3 Decision Tree

Struktur data dengan simpul dan sisi disebut pohon. Simpul akar, simpul cabang/internal, dan simpul daun adalah tiga jenis simpul yang membentuk sebuah pohons [13].

Dalam sebuah pohon keputusan, simpul internal dan simpul akar diberi label dengan nama atribut, rusuk diberi label dengan nilai atribut potensial, dan simpul daun diberi label dengan berbagai kelas. Pohon keputusan adalah representasi langsung dari teknik klasifikasi untuk jumlah kelas yang terbatas [13].

2.1.4 Algoritma C4.5

Metode C4.5, yang dapat digunakan untuk menghasilkan pohon keputusan, merupakan algoritme yang paling populer karena mengungguli algoritme lainnya dalam banyak hal. Algoritma ini juga akan menghasilkan rule yang nantinya akan mempermudah dalam proses menggali sebuah data dan informasi [14].



Gambar 2. 2 Proses Algoritma C4.5

Metode pohon keputusan C4.5 melalui beberapa langkah untuk membuat pohon keputusan, termasuk :

1. mengumpulkan data pelatihan. Data awal yang telah diklasifikasikan ke dalam kelompok tertentu dapat digunakan sebagai data pelatihans. Data Training adalah data yang nantinya akan diolah dengan menggunakan algoritma C4.5 dalam pemodelan pohon keputusan dan rule [15].
2. Hitung jumlah akar pohon keputusan. Akar pohon keputusan ditentukan
*Rumus Entropi

berdasarkan atribut data yang dipilih dengan menghitung nilai entropi dan gain dari setiap atribut; akar dari bagian pertama adalah atribut dengan nilai validasi tertinggi. Nilai entropy dari setiap fitur harus dipahami sebelum mencari nilai gain. Perhitungan nilai entropy bisa dihitung dengan menggunakan rumus :

$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2 p_i$$

Keterangan :

S : Himpunan Kasus

i : Jumlah Partisi S

p_i : Proporsi dari S_i terhadap S

Jika nilai entropy pada masing-masing atribut sudah diketahui, kemudian mencari nilai gain dengan rumus :

*Rumus Entropi $Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$ 2.2

Keterangan :

S : Himpunan Kasus

A : Fitur

N : Jumlah Partisi Atribut A

$|S_i|$: Jumlah Kasus Pada Partisi Ke- i

$|S|$: Jumlah Kasus Dalam S

Setelah proses perhitungan nilai entropy dan gain didapatkan, akan menghasilkan model pohon keputusan dan rule yang terbentuk dari proses penggalan data.

2.1.5 Penerapan Algoritma *Decision Tree C4.5*

Adapun penelitian terkait tentang penerapan daripada algoritma *Decision Tree C4.5* yaitu penelitian yang dilakukan oleh Karya Suhada¹, Anggi Elanda², dan Anwar Aziz³ [16].

1. Tahap Collect Initial Data

Pada tahap ini dilakukan pendataan awal dimana informasi diperoleh dari manajemen STMIK Rosma Karawang [16].

Tabel 2. 2 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 2.2 adalah data pribadi mahasiswa IT STMIK Rosma Karawang. Informasi yang diterima menyangkut lulusan rombongan belajar tahun 2012, 2013, 2014, 2015 dan 2016. Berdasarkan informasi tersebut dapat diketahui mahasiswa yang lulus tepat waktu dan yang lulus terlambat. Ada 146 lulusan dengan 10 karakteristik [16].

2. Describe Data

Tahap selanjutnya setelah mengumpulkan data pertama adalah menjelaskan informasi yang telah dikumpulkan sebelumnya. Tujuan dari tujuan ini adalah untuk mempelajari sedikit tentang karakteristik data yang telah dikumpulkan dan memahami data yang akan diproses untuk mengantisipasi kelulusan siswa secara akurat. Ringkasan karakteristik dari semua data yang telah diperoleh sebelumnya disajikan di bawah ini dalam tabel 2.3 [16].

Tabel 2. 3 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

3. 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 [16].

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"> • Angkatan Tahun 2012 = 26 • Angkatan Tahun 2013 = 31 • Angkatan Tahun 2014 = 33 • Angkatan Tahun 2015 = 31 • Angkatan Tahun 2016 = 25 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"> • Angkatan Tahun 2012 = 26 • Angkatan Tahun 2013 = 31 • Angkatan Tahun 2014 = 33 • Angkatan Tahun 2015 = 31 • Angkatan Tahun 2016 = 25 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"> • Laki – Laki • Perempuan 	Jumlah jenis kelamin mahasiswa : <ul style="list-style-type: none"> • Laki – Laki = 108 • Perempuan = 38 Total data keseluruhan adalah 146
5	Asal Daerah	Asal daerah mahasiswa berisi : <ul style="list-style-type: none"> • Karawang • Luar Karawang 	Jumlah asal daerah mahasiswa : <ul style="list-style-type: none"> • Karawang = 54 • Luar Karawang = 73 Total data keseluruhan adalah 127
7	Jumlah SKS	Jumlah Satuan Kredit Semester (SKS) yang telah ditempuh	Jumlah SKS : Min = 64 Max = 145

4. *Verify Data Quality*

Selama langkah verifikasi kualitas data ini, setiap nilai yang hilang, data yang tidak konsisten, data yang berisik, atau data pencilan akan dicari. Sedangkan tujuan dari pemeriksaan ini adalah untuk mengevaluasi kualitas data dari akuisisi data sebelumnya. Gambar 4 mengilustrasikan bagaimana informasi yang dikumpulkan dari administrasi menunjukkan bahwa beberapa asal geografis dan tanggal lahir masih memiliki nilai yang hilang. Seperti yang dapat diamati, kolom untuk asal geografis dan tanggal lahir memiliki baris kosong untuk data yang memiliki nilai yang hilang [16].

Tabel 2. 4 *Missing Value*

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

5. *Select Data*

Beberapa karakteristik dipilih dari data awal untuk diteliti sebagai karakteristik yang mempengaruhi waktu yang dibutuhkan mahasiswa untuk lulus. NIM, nama, jenis kelamin, daerah asal, tanggal lahir, jumlah SKS, IPK, predikat, masa studi, dan status kelulusan adalah karakteristik yang dipilih [16].

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

Karakteristik NIM, nama, dan lama studi hanya digunakan untuk mengidentifikasi data. Dengan menggunakan teknik data mining dan pendekatan C4.5, variabel-variabel berikut ini akan diproses: jenis kelamin, lokasi asal, jumlah SKS, IPK, predikat, dan lama studi. Informasi mengenai mahasiswa yang lulus pada tahun 2012, 2013, 2014, 2015, dan 2016 diperlukan untuk membuat prediksi kelulusan. Dua nilai, "Tepat Waktu" dan "Terlambat", akan dimasukkan ke dalam properti status kelulusan. Dua nilai tersebut berasal dari hasil transformasi durasi studi dan tahun kelulusans [16].

6. *Clean Data*

Untuk mendapatkan hasil terbaik dari pembuatan prediksi, fase ini melibatkan memastikan bahwa data yang digunakan memang berkualitas tinggi. Seperti yang telah diketahui sebelumnya, ada sejumlah nilai yang hilang dalam data yang diperoleh, yang ditampilkan pada tabel 2.3. Oleh karena itu, data tersebut sekarang akan diproses untuk menangani nilai yang hilang. Penghapusan data yang mengandung nilai yang hilang dari kumpulan data saat ini adalah salah satu metode untuk menangani data yang mengandung nilai yang hilang. Data mahasiswa yang telah diurutkan dari mahasiswa angkatan 2012, 2013, 2014, 2015, dan 2016 dan mengandung nilai yang hilang dijelaskan di bawah ini [16].

Tabel 2. 5 *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 2.5 Data siswa yang ditunjukkan di atas, yang memiliki angka yang hilang, berasal dari tahun 2012 hingga 2016. Baris dengan tanda tanya menunjukkan nilai yang hilang. Data dengan nilai yang hilang akan dihapus untuk mengatasi hal ini. Sebanyak 19 data, termasuk data yang memiliki nilai yang hilang, dihapus dari total 146 data, sehingga menyisakan 127 data yang akan diproses ke tahap selanjutnya [16].

Tabel 2. 6 *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 2.6 Data dari mahasiswa angkatan 2012, 2013, 2014, 2015, dan 2016 telah dibersihkan dengan menghindari penggunaan baris data dengan nilai kosong [16].

7. Construct Data

Persiapan data konstruktif menjadi langkah berikutnya, dan ini melibatkan pembuatan karakteristik turunan dari atribut yang sudah ada. Saat ini, kelas target diwakili oleh satu atribut baru yang telah diperkenalkan. Ada dua nilai yang mungkin untuk atribut ini: "Tepat Waktu" dan "Terlambat"s [16].

Tabel 2. 7 Atribut Baru

Tepat waktu
Terlambat

Karakteristik ini didasarkan pada informasi dari nomor seri ijazah mahasiswa, yang menunjukkan lama studi dan tahun kelulusan mereka. Siswa yang belajar tidak lebih dari empat tahun adalah mereka yang menerima nilai tujuan atau yang dianggap lulus tepat waktu. Agar data dapat ditambahkan ke atribut, caranya adalah sebagai berikut [16]:

Tabel 2.8 Penambahan Atribut Baru Pada Data Mahasiswa STMIK Rosma Karawang

No	Nomor Seri Ijazah	Tahun Kelulusan	Nama Mahasiswa	Jenis Kelamin	Alamat	Tanggal Kelahiran	Nilai	Kategori	Target	
34	20313/I	200502	Mahasiwa01	Prempuan	Karawang	27 September 1990	145	2,83	Memuaskan	4
35	20413/I	200213	Mahasiwa02	Laki-Laki	Luar karawang	22 Oktober 1991	145	3,19	Sangat Me	4
36	20513/I	200241	Mahasiwa03	Laki-Laki	Karawang	13 Juni 1990	145	2,81	Memuaskan	4
37	20513/I	200310	mahasiswa04	Laki-Laki	Karawang	26 Febuan 1995	145	2,97	Memuaskan	4
38	20613/I	200835	Mahasiswa05	Laki-Laki	Karawang	31 Desember 1993	145	2,98	Memuaskan	4
39	10013/I	200401	Mahasiswa06	Prempuan	Karawang	03 Maret 1992	145	2,79	Memuaskan	3
40	20013/I	200501	Mahasiswa07	Laki-Laki	Karawang	12 Juni 1993	145	2,80	Memuaskan	5
41	10013/I	200512	Mahasiswa08	Laki-Laki	Karawang	09 September 1984	145	3,57	Sangat Me	5
42	40013/I	200567	Mahasiswa09	Prempuan	Karawang	16 Januari 1996	145	3,03	Sangat Me	5
43	20013/I	200767	Mahasiswa10	Laki-Laki	Karawang	30 Maret 1996	145	3,63	Sangat Me	5
44	90013/I	200697	Mahasiswa11	Laki-Laki	Karawang	19 September 1996	145	2,98	Memuaskan	4
45	20013/I	200630	Mahasiswa12	Laki-Laki	Luar karawang	10 Maret 1995	145	3,17	Sangat Me	4
46	20213/I	200532	Mahasiswa13	Laki-Laki	Luar karawang	21 Juli 1993	145	3,18	Sangat Me	4
47	20113/I	200758	Mahasiswa14	Laki-Laki	Luar karawang	16 Mei 1996	145	2,88	Memuaskan	4
48	20113/I	200739	Mahasiswa15	Prempuan	Luar karawang	20 Februari 1993	145	2,88	Memuaskan	4
49	20101/I	200588	Mahasiswa16	Laki-Laki	Luar karawang	10 Juli 1991	145	3,01	Sangat Me	4

8. Tranformation

Tahap ini melibatkan transformasi data untuk mendapatkan informasi terbaik. Tipe data standar diubah menjadi kategorikal sehingga dapat digunakan untuk kategorisasi algoritme C4.5 untuk pendekatan data mining [16].

Jumlah kredit dan kualitas IPK, khususnya, perlu dimodifikasi dalam data yang telah diperoleh. Jumlah kredit dibagi menjadi dua kelompok, yaitu mereka yang memiliki nilai 144 kredit atau kurang (≤ 144) dan mereka yang memiliki nilai 144 kredit atau lebih (>144). Pembagian kedua dari karakteristik IPK adalah IPK yang dimiliki kurang dari atau sama dengan 3 (tiga)s [16].

Tabel 2. 9 Hasil Transformasi Atribut

No	Jumlah Sks	IPK	Nomor Seri Ijazah	NIM	Nama	Jenis Kel	Asal Daerah	Tanggal Lahir
1	SKS>144	IPK \leq 3	20011	201201	Mahasiswa01	Premputan	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 \leq 3	20011	2010123	Mahasiswa06	Premputan	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	Premputan	Karawang	16 Januari 1996
10	SKS>144	IPK>3	20011	2014009	Mahasiswa10	Laki-Laki	Karawang	30 Maret 1996
11	SKS>144	IPK \leq 3	20011	20150019	Mahasiswa11	Laki-Laki	Karawang	19 September 1996
12	SKS>144	IPK \leq 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

9. Format Data

Pemformatan data adalah langkah terakhir dalam tahap persiapan. Tahap ini melibatkan pembuatan dataset yang sudah jadi, yang dipersiapkan untuk diproses dengan perangkat lunak pemodelan data mining. Dataset berikut ini akan digunakan untuk meramalkan kelulusans [16]:

Tabel 2. 10 Atribut Data Terpilih

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

10. Modeling

menggunakan pohon keputusan untuk memprediksi kelulusan mahasiswa teknologi informasi STMIK Rosma Karawang. Teknologi RapidMiner digunakan untuk menghasilkan pohon keputusan ini. Berdasarkan faktor-faktor yang ditentukan oleh pohon keputusan, data 127 dua kelas tepat waktu dan terlambat digunakan untuk menentukan model mahasiswa teknologi informasi STMIK Rosma Karawang yang dapat menyelesaikan studinya tepat waktu atau terlambat. Dalam penelitian ini, validasi silang digunakan untuk menilai validitas model klasifikasi. Salah satu teknik untuk memisahkan data pelatihan dan pengujian dari keseluruhan kumpulan data adalah validasi silang. Untuk mempartisi set data secara acak, validasi silang diulang sebanyak n kali. Karena 10 memberikan evaluasi ukuran yang lebih baik, maka 10 adalah nilai yang disarankan. Prosedur iterasi diulang sepuluh kali, menghasilkan pembuatan pohon keputusan sepuluh kali selama periode tersebut [16].



Gambar 2. 3 Atribut Data Terpilih

Pada Gambar 2.3 Hasil tersebut diberikan klasifikasi predikat akhir mahasiswa teknologi informasi STMIK Rosma Karawang menggunakan pohon keputusan. Karakteristik yang paling mempengaruhi klasifikasi predikat kelulusan mahasiswa teknologi informasi STMIK Rosma Karawang menurut pohon keputusan yang membentuk template adalah IPK. Fungsi GPA menggunakan simpul akar adalah petunjuk untuk ini [16].

11. Evaluation

Implementasi pendekatan pohon keputusan dengan menggunakan validasi silang untuk mengevaluasi validitas model dengan menggunakan 127 data dari mahasiswa teknik informatika STMIK Rosma Karawang menghasilkan 109 hasil yang berhasil. Ketika predikat kelulusan mahasiswa diklasifikasikan menggunakan pendekatan pohon keputusan, hasil klasifikasi ditampilkan dalam bentuk tabel confusion matrix. Informasi mengenai kategorisasi ide data mining dapat dilihat dari confusion matrix [16].

Tabel 2.11 *Confusion Matrix*

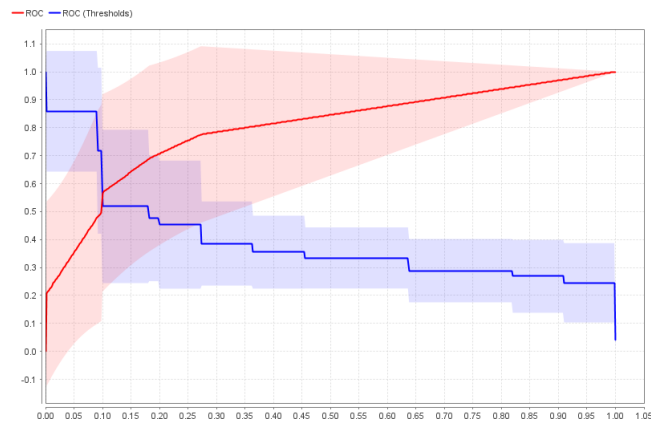
<i>Prediction</i>	<i>Actual</i>	
	Tepat Waktu	Terlambat
Tepat Waktu	99 (TP)	10 (FP)
Terlambat	8 (FN)	10 (TN)

Berdasarkan tabel confusion matrix dan metode decision tree, dari 127 data, 99 data tepat waktu diproyeksikan secara tepat sebagai bagian dari kelas tepat waktu, sedangkan 10 data tepat waktu diantisipasi sebagai data terlambat. Sementara delapan data terlambat diproyeksikan tepat waktu untuk kelas terlambat, sedangkan 10 data terlambat diantisipasi dengan tepat untuk terlambat [16].

Menemukan nilai akurasi, recall, presisi, dan AUC adalah hal berikutnya setelah memahami temuan tabel confusion matrix.

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

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



Gambar 2. 4 Kurva ROC

Pada gambar 2.4 menunjukkan bahwa model yang dikembangkan dan diuji menghasilkan bentuk kurva dengan nilai AUC sebesar 0,799 yang menunjukkan akurasi prediksi cukup baik [16].

12. Hasil

Nilai akurasi sebesar 85.83%, nilai recall sebesar 92.52%, nilai precision sebesar 90.83%, dan nilai AUC sebesar 0.799. Hasil dari semua pengujian pemodelan menggunakan validasi silang dengan jumlah data latih dan uji yang sama memiliki nilai true positive sebesar 99, nilai true negative sebesar 10, nilai false negative sebesar 8, dan nilai false positive sebesar 10 [16].

2.1.6 Python

Python adalah sebuah bahasa pemrograman yang bertujuan umum dan memiliki level tingkat tinggi. Pertama kali dirilis pada tahun 1991 oleh Guido van Rossum selaku pembuat, *python* digunakan untuk pembuatan aplikasi, sebagai alat untuk memerintah suatu komputer, dan melakukan analisis data. *Python* sebagai general-purpose language dapat digunakan untuk membuat sebuah program apa saja dan dapat menyelesaikan berbagai permasalahan yang terjadi [16].

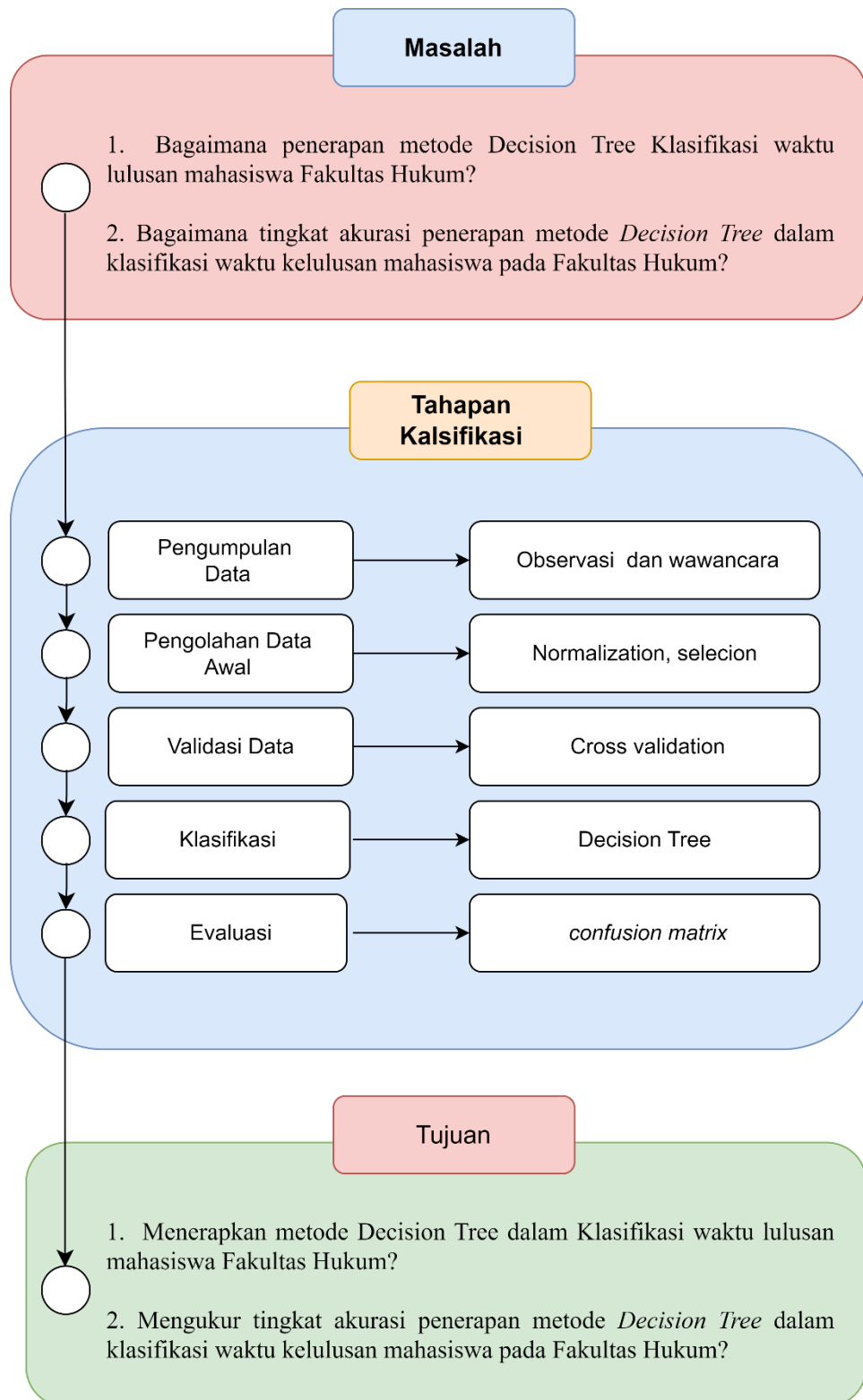
2.1.7 Confusion Marix

Confusion matrix adalah sebuah *tool* untuk mengevaluasi model dari klasifikasi untuk memprediksi objek yang bernilai benar atau salah. *Matrix* dari hasil prediksi yang nantinya akan dibandingkan dengan kelas asli inputan atau informasi nilai aktual dan prediksi pada klasifikasi [16].

	Predicted Positive	Predicted Negative	
Actual Positive	TP <i>True Positive</i>	FN <i>False Negative</i>	Sensitivity $\frac{TP}{(TP + FN)}$
Actual Negative	FP <i>False Positive</i>	TN <i>True Negative</i>	Specificity $\frac{TN}{(TN + FP)}$
	Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

Gambar 2. 5 Rumus *Confussion Matrix* 2 Kelas

2.2 Kerangka Pikir



BAB III

METODE PENELITIAN

2.3 Jenis, Metode, Subjek, Objek, Waktu, dan Lokasi Penelitian

Penelitian ini termasuk dalam kategori penelitian terapan. Menurut jenis informasi yang dianalisis, penelitian ini bersifat kuantitatif. Dan penelitian ini bersifat konfirmatori bila dievaluasi dari perspektif perlakuan data.

Fakultas Hukum Universitas Ichsan Gorontalo adalah tempat studi kasus untuk penelitian ini. Dengan menggunakan metode pohon keputusan, topik penelitian ini adalah prediksi kelulusan mahasiswa Fakultas Hukum Universitas Ichsan Gorontalo. Universitas Ichsan Gorontalo akan menjadi lokasi penelitian ini.

Data primer yang di gunakan pada penelitian ini adalah data mahasiswa fakultas hukum yang di kumpulkan menggunakan Teknik observasi dan wawancara sedangkan data sekunder di kumpulkan menggunakan referensi jurnal dari penelitian sebelumnya dan perpustakaan.

2.4 Pengumpulan Data

Tabel 3. 1 Variabel Atribut Data

No.	Atribut	Jenis	Value	Keterangan
1	Jenis Kelamin	Tuple	Diketahui	Variabel Input
2	Asal Daerah	Varchar	Diketahui	Variabel Input
3	IPS I	Float	Diketahui	Variabel Input
4	IPS II	Float	Diketahui	Variabel Input
5	IPS III	Float	Diketahui	Variabel Input
6	IPS IV	Float	Diketahui	Variabel Input
7	Class Label	Varchar	Tepat Waktu, Tidak Tepat Waktu	Output

Penelitian ini menggunakan pengumpulan data dengan jenis data primer yang dikumpulkan dengan teknik dokumentasi, wawancara, dan observasi yang didapatkan langsung di lapangan.

2.5 Pemodelan/Abstraksi

Tabel 3. 2 Data Kelulusan Mahasiswa Fakultas Ilmu Hukum, 2019

No.	Jenis Kelamin	Asal Daerah	IPS I	IPS II	IPS III	IPS IV	Keterangan Lulus
1	L	Gorontalo	4.00	4.00	3.25	3.26	Tepat
2	L	Luar Gorontalo	3.50	4.00	3.75	3.52	Tepat
3	L	Gorontalo	2.35	3.92	3.83	3.52	Tepat
4	L	Luar Gorontalo	3.60	3.92	3.67	3.43	Tepat
5	L	Luar Gorontalo	3.15	3.50	2.91	0.19	Tidak Tepat
6	L	Luar Gorontalo	2.85	1.96	1.00	2.57	Tidak Tepat
7	L	Luar Gorontalo	4.00	4.00	3.75	3.87	Tepat
8	P	Gorontalo	3.00	3.17	2.78	3.00	Tepat
....
144	L	Gorontalo	2.60	0.65	2.50	3.00	Tidak Tepat

2.6 Pra Pengolahan Data

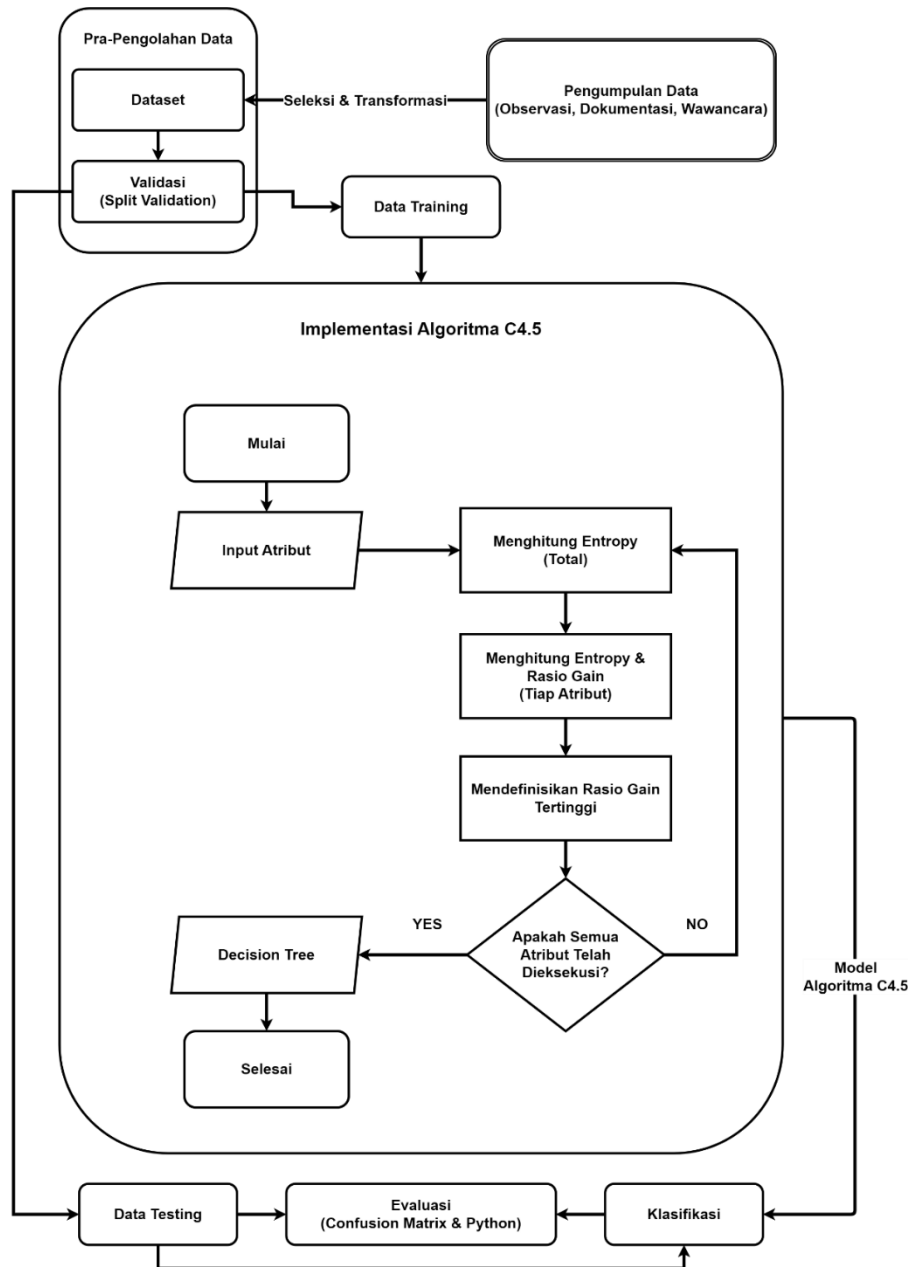
Pemilihan dan transformasi data dilakukan pada data yang akan diolah nantinya. Hal ini dilakukan untuk menetapkan standar atribut yang selanjutnya akan dijadikan sebagai tolok ukur untuk mengukur kontribusi setiap atribut terhadap klasifikasi data.

2.7 Validasi

Tujuan validasi adalah untuk memisahkan data awal menjadi data training dan data testing. Data testing adalah data yang akan digunakan dalam proses pengujian dengan menggunakan program komputer, sedangkan data training adalah data yang akan diproses dengan menggunakan metode klasifikasi. Tergantung pada

jumlah data, pendekatan validasi yang digunakan pada tahap ini adalah split validation, yang membagi data menjadi dua bagian.

2.7.1 Pengembangan Model



Gambar 3. 1 Pengembangan Model Klasifikasi Waktu Kelulusan Mahasiswa

BAB IV

HASIL PENELITIAN

4.1 Hasil Pengumpulan Data

Dataset atau data penelitian yang digunakan pada penelitian ini merupakan data private yang di ambil dari BAAK Universitas Ichsan Gorontalo terkait aktifitas kuliah mahasiswa dari tahun akademik 2017-2017 adapun dataset yang di kumpulkan dapat dilihat pada tabel 4.1.

Tabel 4. 1 Hasil Pengumpulan data

No	NIM	Nama	Tempat Lahir	Kelamin	Tanggal Lahir	Prodi
1	H1115001	MOHAMAD RIZAL MUDA	GORONTALO	L	10/17/1996	S1 ILMU HUKUM
2	H1115003	ANGGRIANI ADRIAN	GORONTALO	P	5/5/1995	S1 ILMU HUKUM
3	H1115012	SUPRAYOGI HIMAWAN	GORONTALO	L	1/3/1997	S1 ILMU HUKUM
4	H1115014	SANTI HANDAYANI HABUGE	GORONTALO	P	7/6/1996	S1 ILMU HUKUM
5	H1115015	ANGCY NELA JUSUF	KABILA	P	11/12/1996	S1 ILMU HUKUM
6	H1115017	APRILIO ALAN PUTRA MOHAMAD	GORONTALO	L	4/7/1998	S1 ILMU HUKUM
7	H1115018	FITRAN ARIFIN	KWANDANG	L	3/14/1996	S1 ILMU HUKUM
8	H1115019	OSTAN TANI	PONELO	L	7/22/1996	S1 ILMU HUKUM
9	H1115020	KURNIAWATI DATAU	GORONTALO	P	2/1/1996	S1 ILMU HUKUM
10	H1115021	MOHAMAD RIYANDI LAGASIUS	BONEPANTAI	L	10/15/1995	S1 ILMU HUKUM
11	H1115022	MOH AGUNG BUMULO	TOMBULILATO	L	12/7/1996	S1 ILMU HUKUM
12	H1115023	MOHAMMAD RIZALDY	GORONTALO	L	8/22/1997	S1 ILMU HUKUM
13	H1115024	VINDRIYANI ABD. AZIS AHMAD	GORONTALO	P	5/26/1997	S1 ILMU HUKUM
14	H1115025	SRI ARLENI ADU	LIMBOTO	P	11/17/1997	S1 ILMU HUKUM
15	H1115026	VIVIAN APRILIANY DUMAKO	GORONTALO	P	4/4/1998	S1 ILMU HUKUM
.....
536	H1115029	MOHAMAD ARIF MASUI	GORONTALO	L	12/11/1996	S1 ILMU HUKUM

No	IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPS7	IPS8	IPS9
1	3.81	3.79	0	3.43	3.67	3.74	3.87	2.54	4
2	3.48	3.26	2.7	3	3	2.06	1.75	3	3.33
3	3.76	3.58	3.74	3.33	3.83	3.48	2.8	4	0
4	3.38	2.95	3.6	3.42	3.22	2.74	2.73	1.13	3
5	3.67	3.11	3.8	3.58	3.39	3.1	3	2.4	0
6	3.67	3.58	3.91	3.42	3.12	3	4		0
7	3.05	3.21	3	3	3.13	3.13	3.17	4	0
8	3.24	3.11	3	2.78	3.05	2.74	2.7	4	0
9	4	3.79	3.78	3.58	3.83	3.48	2.8	0	3.33
10	3.81	3.79	3.17	3.92	3.65	3.76	4	2	0
11	3.9	3.79	3.91	3.75	3.83	3.86	4	2	0
12	3.48	3.58	3.74	3.33	3.39	2.83	2.8	4	0
13	3.67	3.47	3.78	3.67	3.48	3.48	3.8	4	0
14	3.81	4	3.91	3.67	4	3.65	3.8	4	0
15	3.67	4	3.91	3.75	3.74	3.78	3.8	4	0
.....
536	3.9	3.79	4	3.75	3.74	3.86	3.8	2	0

4.2 Hasil Pemodelan

Berdasarkan dataset pada tabel 4.1 diatas dapat di lakukan beberapa pengolahan sebelum di terapkan model .

1. Seleksi Data

Berdasarkan dataset pada tabel 4.1 di atas dapat dilihat bahwa terdapat perbedaan data yang cukup signifikan, sehingga dalam penerapan metode *Decision Tree* data yang dibagi lima yaitu dataset untuk jenis kelamin, IPS1, IPS2, IPS3, dan IPS4 serta tambahan satu atribut yaitu keterangan lulus sebagai keterangan lulus sebagai keterangan keluaran. Hasil dari pemodelan bisa di lihat pada tabel 4.2 di bawah ini:

Tabel 4. 2 Dataset Hasil Olahan

No	Jenis Kelamin	IPS1	IPS2	IPS3	IPS4	keterangan
1	L	3.48	3.26	2.7	3	Tidak Tepat
2	P	3.76	3.58	3.74	3.33	Tepat Waktu
3	L	3.38	2.95	3.6	3.42	Tidak Tepat
4	P	3.67	3.11	3.8	3.58	Tidak Tepat
5	P	3.67	3.58	3.91	3.42	Tepat Waktu
6	L	3.05	3.21	3	3	Tidak Tepat
7	L	3.24	3.11	3	2.78	Tidak Tepat
8	L	4	3.79	3.78	3.58	Tepat Waktu
9	P	3.81	3.79	3.17	3.92	Tepat Waktu
10	L	3.9	3.79	3.91	3.75	Tepat Waktu
11	L	3.48	3.58	3.74	3.33	Tepat Waktu
12	L	3.67	3.47	3.78	3.67	Tepat Waktu
13	P	3.81	4	3.91	3.67	Tepat Waktu
14	P	3.67	4	3.91	3.75	Tepat Waktu
15	P	3	3.68	3.17	3.67	Tepat Waktu
....
536	L	3.9	3.37	3.87	3.58	Tepat Waktu

Berdasarkan tabel 4.2 di atas dilakukan perhitungan jumlah untuk masing-masing kelas hasilnya dapat dilihat pada gambar 4.1 berikut:

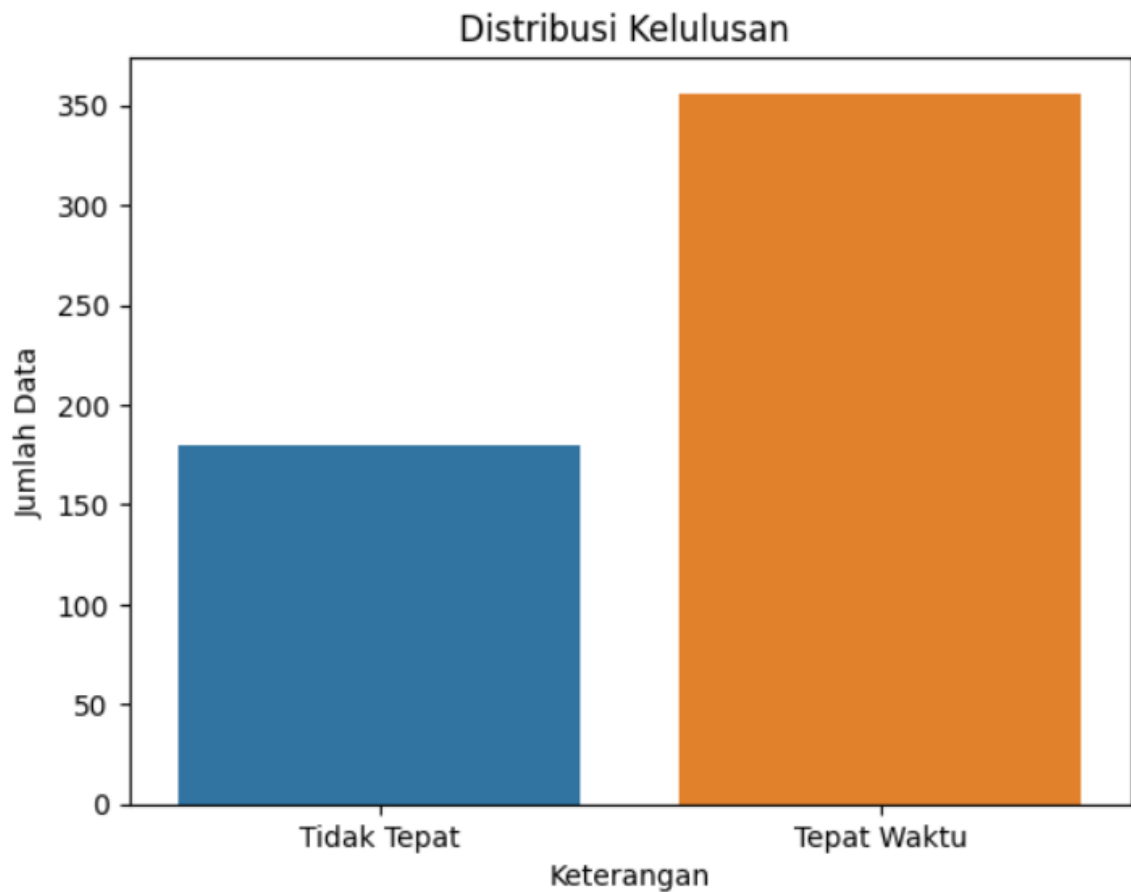
Untuk menentukan root(node) awal dari nilai keputusan ,Langkah pertama dengan menghitung nilai entropy dan gain tertinggi untuk tiap-tiap atributnya. Berikut adalah Langkah-langkah penyelesaian :

Tahap 1 : menghitung nilai entropy total untuk seluruh atribut

Atribut	jumlah	Tidak lulus	Lulus	Entropy
Total	376	123	252	0.914274

Setelah mengetahui nilai entropy total dari semua atribut , masuk ke tahap ke 2 menghitung nilai entropi tiap-tiap atributnya , sebagai berikut :

Atribut	Jumlah	Tidak Lulus	Lulus	Entropy	Gain
IPS 1	Gain				0.512119
	≤ 3.5	375	123	0.91287	
	> 3.5	248	56	0.770629	
	Gain				0.769743
	≤ 3.67	261	59	0.771068	
	> 3.67	208	40	0.706274	
IPS 2	Gain				0.872972
	≤ 3.4	159	86	0.995173	
	> 3.4	216	37	0.660662	
	Gain				0.844376
	≤ 3.58	221	38	0.662136	
	> 3.58	192	30	0.625262	
IPS 3	Gain				0.737688
	≤ 3.5	153	93	0.966177	
	> 3.5	233	35	0.610377	
	Gain				1.343179
	≤ 3.65	193	24	0.541743	
	> 3.65	233	35	0.610377	
IPS 4	Gain				0.894468
	≤ 3.4	151	92	0.965268	
	> 3.4	227	34	0.609289	
	Gain				1.250252
	≤ 3.5	199	24	0.531092	
	> 3.5	317	65	0.731909	



Gambar 4. 1 Grafik Jumlah kelas output

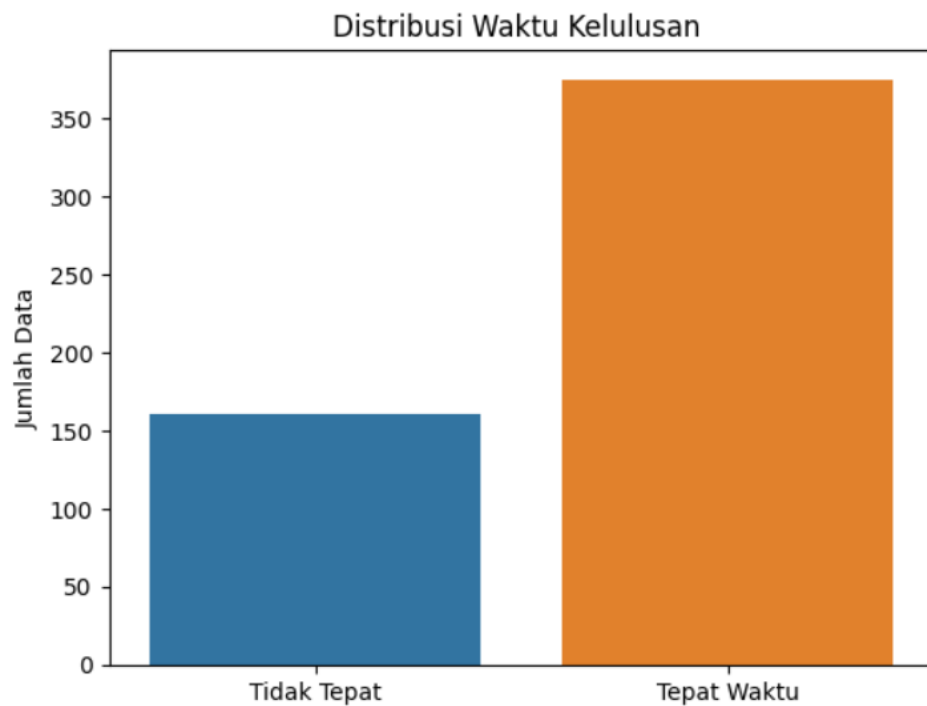
2. Preprocessingg

Pada tahapan ini dilakukan pengecekan data yang kosong atau data yang tidak konsisten terhadap hasil akhir atau kelas output. Berikut contoh sampel data yang tidak konsisten dapat di lihat tabel 4.3 berikut:

Tabel 4. 3 Data Tidak konsisten

No	Jenis Kelamin	IPS 1	IPS 2	IPS 3	IPS 4	Keterangan
1	L	3.81	3.79	3.7	3.83	Tepat waktu,Tidak tepat
2	L	3.9	3.79	3.91	3.75	Tepat waktu,Tidak tepat
3	L	4.0	4.0	3.91	4.0	Tepat waktu,Tidak tepat
4	L	4.0	4.0	4.0	4.0	Tepat waktu,Tidak tepat

Setelah dilakukan perbaikan data kemudian hitung Kembali jumlah data masing-masing kelas dapat dilihat pada gambar 4.2:

**Gambar 4. 2** Grafik Jumlah Kelas Setelah di Perbaikan

3. Transformasi

Pada tahapan ini di lakukan transformasi yaitu data jenis kelamin diubah dalam bentuk nilai 1 dan 2 agar mempermudah proses pemodelan dengan metode *decision tree* sampel hasil transformasi dapat dilihat tabel 4.4 berikut:

Tabel 4. 4 Tabel Dataset Hasil Transformasi

No	Jenis Kelamin	IPS 1	IPS 2	IPS 3	IPS 4	Keterangan
1	1	3.81	3.79	3.7	3.83	Tidak waktu
2	2	3.9	3.79	3.91	3.75	Tepat waktu
3	2	4.0	4.0	3.91	4.0	Tidak waktu
4	1	4.0	4.0	4.0	4.0	Tidak waktu

4. Proses Data Mining

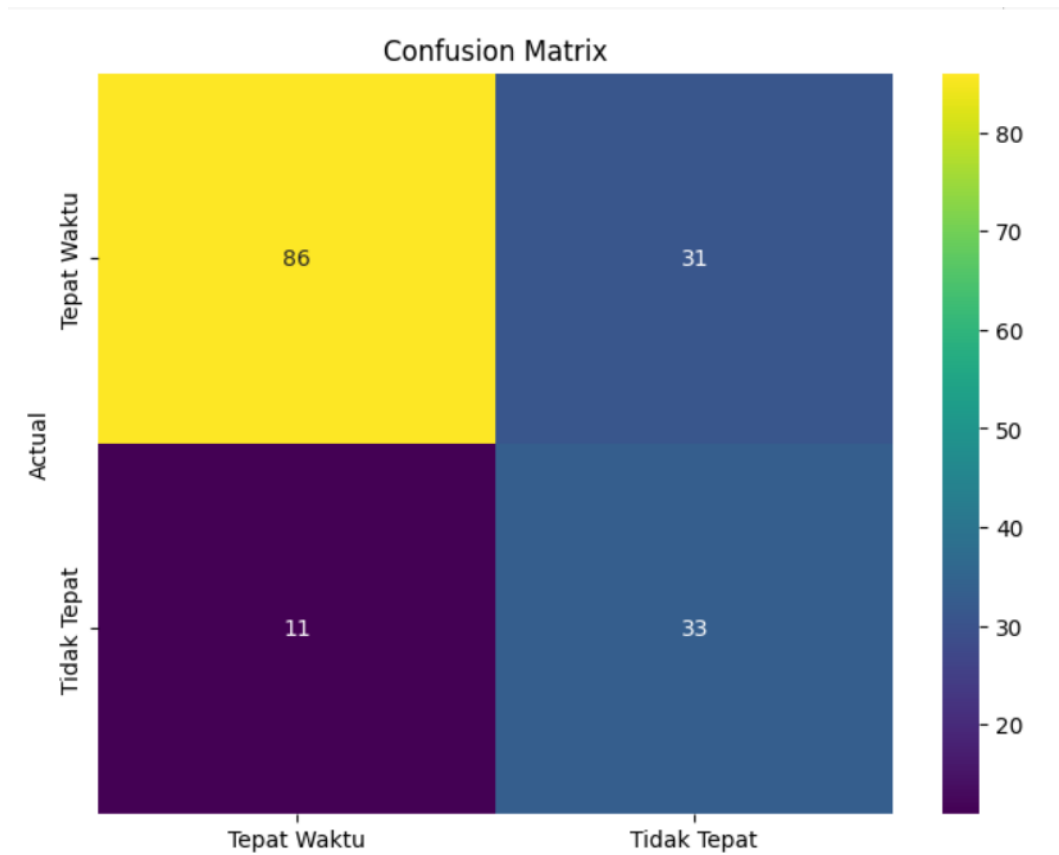
Setelah dilakukan tahapan diatas , maka selanjutnya dilakukan proses data mining terdiri dari penentuan data atribut NO, Jenis Kelamin, IPS1, IPS2, IPS3, IPS4, dan Keterangan , pembagian data trening dan data testing dengan perbandingan 70:30 sehingga didapat Data Trenining sebanyak (375) dan Data Testing sebanyak (161) data . setelah itu dilakukan proses pemodelan dengan metode *decision tree*

5. Evaluasi

Setelah dilakukan proses data mining perlu dilakukan evaluasi terhadap model yang sudah dibuat dengan melihat hasil akurasi dan visualisasi hasil pemodelan

a. evaluasi kinerja model

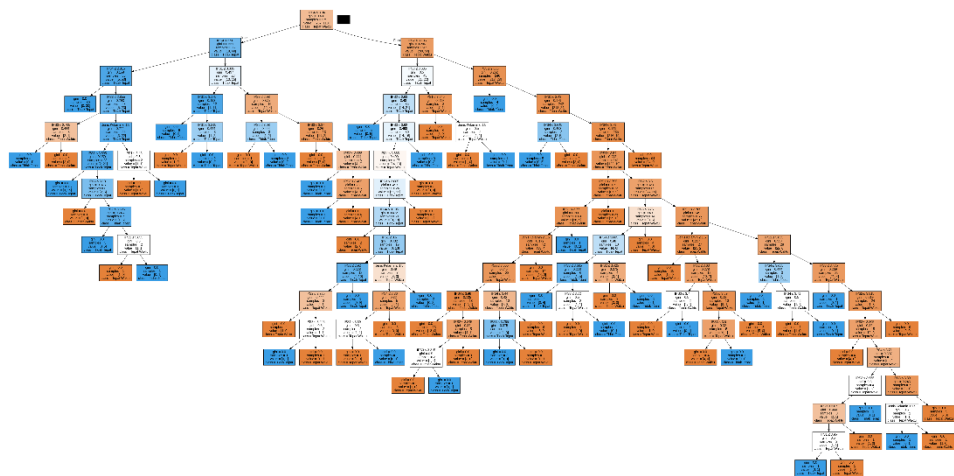
evaluasi kinerja model *decision tree* dengan hasil perhitungan dapat dilihat pada gambar 4.3:



Gambar 4. 3 Hasil Perhitungan Confusion Matrix

b. Visualisasi Hasil Pemodelan

Hasil Pemodelan Metode *decision tree* dapat dilihat visualisasi pada gambar 4.4 berikut:



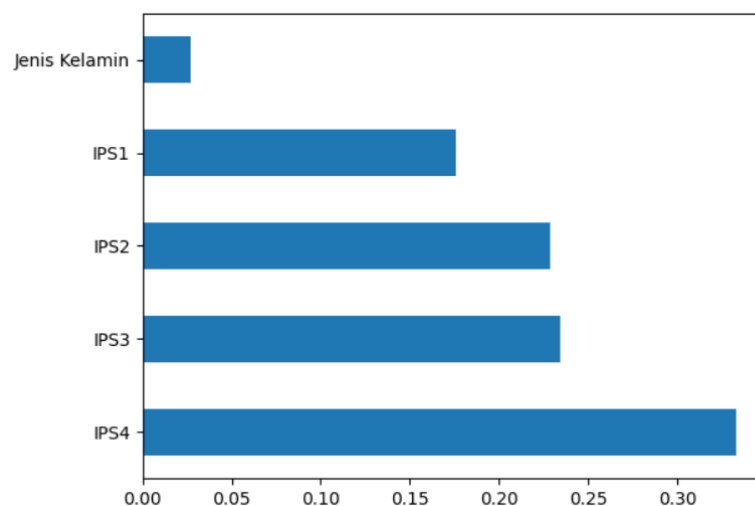
Gambar 4. 4 Visualisasi pemodelan *Decision tree*

6. Perbaikan Permodelan

Berdasarkan hasil evaluasi kinerja model di atas, belum didapatkan hasil yang maksimal sehingga perlu dilakukan perbaikan atau optimalisasi model diantaranya:

a. Optimalisasi Atribut

Beberapa metode bisa digunakan dalam pemilihan atribut yang berpengaruh ke hasil akhir. Pada penelitian digunakan metode *decision tree* dengan hasil perhitungan dapat dilihat pada gambar 4.5 berikut :



Gambar 4. 5 Grafik Kepentingan Atribut

Berdasarkan gambar 4.5 diatas, maka atribut yang digunakan pada pemodelan kedua yaitu : IPS1, IPS2, IPS3, IPS4

b. Pemangkasan pohon

Cara lain bisa gunakan agar model bisa lebih optimal adalah melakukan pemangkasan pohon yang tidak relevan seperti : $IPS4 > 3.75$ dijadikan status Tempat Waktu Semua

c. Mengukur Kedalam Pohon

Salah satu Teknik untuk mendapat pohon keputusan yang tidak terlalu Panjang, maka dapat dilakukan pengukuran kedalaman pohon yang maksimal dengan akurasi yang tinggi. berikut hasil perhitungan kedalaman dan akurasi dapat dilihat pada tabel 4.5 berikut :

Tabel 4. 5 Hasil Perhitungan Kedalaman Pohon dan Akurasi

Kedalaman Pohon	Akurasi
1	87%
2	88%
3	88%
4	90%
5	92%
6	89%
7	91%
8	92%
9	89%
10	89%

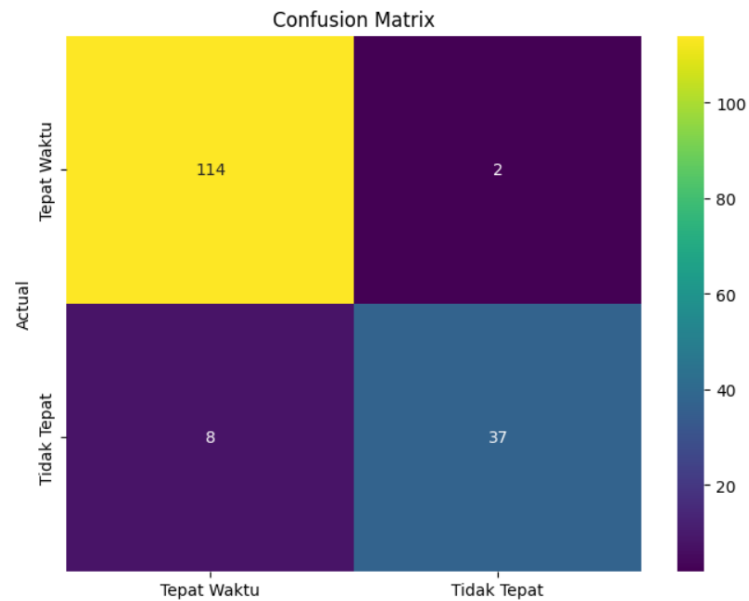
Berdasarkan tabel 4.5 diatas, maka dipilih kedalaman pohon dengan nilai 92%

d. Pemodelan Hasil Optimalisasi

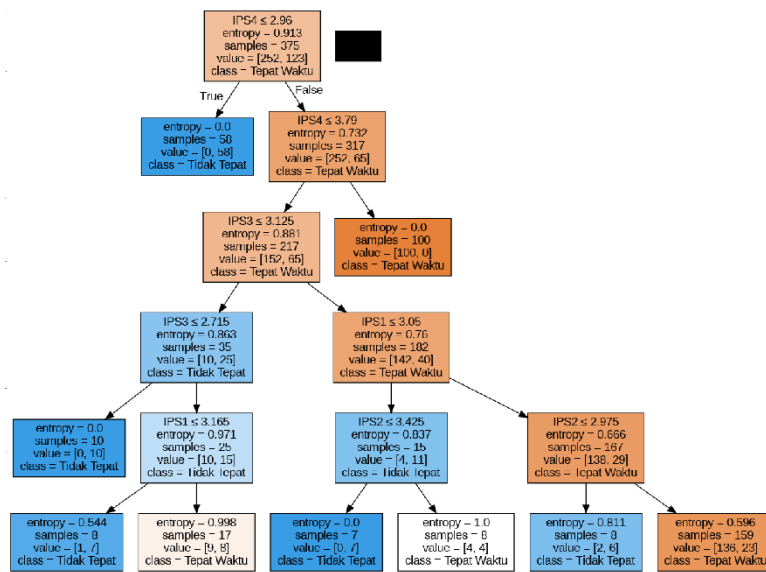
Setelah dilakukan beberapa optimalisasi, selanjutnya diterapkan pada model *disecion tree* kemudian dilakukan evaluasi kinerja Kembali

e. Evaluasi kinerja model setelah optimalisasi

Hasil evaluasi kinerja model setelah dilakukan optimalisasi dapat dilihat pada gambar 4.5 dan gambar 4.6 berikut ini:



Gambar 4. 6 hasil confusion matrix setelah optimalisasi



Gambar 4. 7 Visualisasi model setelah optimalisasi

7. Testing Data Baru

Setelah didapatkan hasil pemodelan yang optimal, maka model tersebut dapat digunakan untuk dilakukan klasifikasi data testing baru dengan memasukan nilai atribut sesuai dengan model yang sudah dilatih. Adapun hasil cobaan klasifikasi data baru dapat dilihat pada tabel 4.6 dan 4.7 berikut:

Tabel 4. 6 Data Testing Baru

No	IPS1	IPS2	IPS3	IPS4	Keterangan
1	3.48	3.26	2.70	3	?
2	3.76	3.58	3.74	3.33	?
3	3.38	2.95	3.60	3.42	?
4	3.67	3.11	3.80	3.58	?
5	3.67	3.58	3.91	3.42	?

Tabel 4. 7 Hasil Klasifikasi Data Testing Baru

No	IPS1	IPS2	IPS3	IPS4	Keterangan
1	3.48	3.26	2.70	3	Tidak Tepat
2	3.76	3.58	3.74	3.33	Tepat Waktu
3	3.38	2.95	3.60	3.42	Tidak Tepat
4	3.67	3.11	3.80	3.58	Tidak Tepat
5	3.67	3.58	3.91	3.42	Tepat Waktu

4.3 Hasil Proses Pemodelan

Pengembangan model *decision tree* menggunakan Bahasa pemrograman Python dengan tools Google Colabotatory dengan tahapan-tahapan berikut:

4.3.1 Tahapan Seleksi

Pada tahapan seleksi data ini dilakukan beberapa proses diantara :

1. Mengaktifkan Penggunaan Google Drive sebagai tempat penyimpanan, Adapun coding dengan python sebagai berikut:

```
#Proses Pengaitan dengan Google Drive
from google.colab import drive
drive.mount('/content/drive')
```

Membaca Dataset file Excel :

```
#Membaca Dataset
df = pd.read_excel("/content/drive/MyDrive/Skripsi/Dataset2.xlsx")
df.head()
```

Hasil outputnya Sebagai berikut :

Tabel 4. 8 Hasil pembacaan Dataset dari file Excel

	No	Jenis Kelamin	IPS1	IPS2	IPS3	IPS4	keterangan
0	1	L	3.48	3.26	2.70	3.00	Tidak Tepat
1	2	P	3.76	3.58	3.74	3.33	Tepat Waktu
2	3	L	3.38	2.95	3.60	3.42	Tidak Tepat
3	4	P	3.67	3.11	3.80	3.58	Tidak Tepat
4	5	P	3.67	3.58	3.91	3.42	Tepat Waktu

menggunakan atribut yaitu: Jenis Kelamin, IPS1, IPS2, IPS3, dan IPS4 sehingga data set yang digunakan dalam proses pemodelan dapat dilihat pada tabel 4.3 berikut:

4.3.2 Preprocessing

Pada tahap preprocessing di lakukan beberapa proses seperti :

1. Melakukan pengecekan data yang kosong atau data yang tidak konsisten terhadap hasil akhir atau kelas output dengan melakukan perintah sebagai berikut

```
# Mengelompokkan data berdasarkan atribut yang ingin diperiksa
grouped_data = df.groupby(['Jenis Kelamin', 'IPS1', 'IPS2', 'IPS3', 'IPS4'])

# Memeriksa kategori kelas untuk setiap kelompok
for key, group in grouped_data:
    unique_classes = group['keterangan'].unique()
    if len(unique_classes) > 1:
        print(f"Kelompok {key}: Kelas yang berbeda = {unique_classes}")
```

Data hasil keluaran dan perintah di atas sebagai berikut :

```
Kelompok ('L', 3.81, 3.79, 3.7, 3.83): Kelas yang berbeda = ['Tepat Waktu' 'Tidak Tepat']
Kelompok ('L', 3.9, 3.79, 3.91, 3.75): Kelas yang berbeda = ['Tepat Waktu' 'Tidak Tepat']
Kelompok ('L', 4.0, 4.0, 3.91, 4.0): Kelas yang berbeda = ['Tepat Waktu' 'Tidak Tepat']
Kelompok ('L', 4.0, 4.0, 4.0, 4.0): Kelas yang berbeda = ['Tepat Waktu' 'Tidak Tepat']
```

2. Melakukan proses validasi data setelah dilakukan perbaikan data dan menghitung Kembali jumlah data masing-masing kelas dengan perintah berikut :

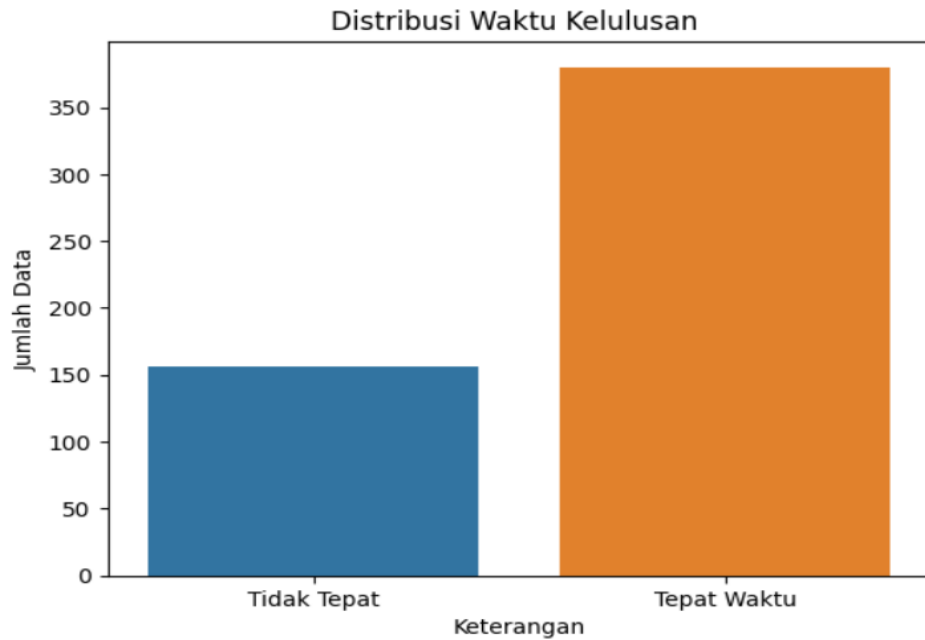
```
# Menghitung jumlah setiap kategori dalam kolom 'Keterangan Lulus'
countplot = sns.countplot(data=df, x='keterangan')

# Menambahkan label pada sumbu x dan y
plt.xlabel('keterangan')
plt.ylabel('Jumlah Data')

# Menambahkan judul
plt.title('Distribusi Waktu Kelulusan ')

# Menampilkan grafik
plt.show()
```

Hasi keluaran dari perintah di atas pada Gambar 4.8



Gambar 4. 8 Gambar Grafik

4.3.3 Traspormasi

Pada proses Traspormasi ini dilakukan beberapa proses berikut :

1. Pada proses mengubah jenis kelamin menjadi angka

```
#Proses Konversi Nilai String ke Numerik
df['Jenis Kelamin'].replace(['L','P'],[1,2], inplace=True)
df.head()
```

Hasil dari keluaran perintah diatas pada gambar 4.9

No	Jenis	Kelamin	IPS1	IPS2	IPS3	IPS4	keterangan
0	1	1	3.48	3.26	2.70	3.00	Tidak Tepat
1	2	2	3.76	3.58	3.74	3.33	Tepat Waktu
2	3	1	3.38	2.95	3.60	3.42	Tidak Tepat
3	4	2	3.67	3.11	3.80	3.58	Tidak Tepat
4	5	2	3.67	3.58	3.91	3.42	Tepat Waktu

Gambar 4. 9 konversi nilai string numerik

2. Pada proses mengubah data statistic

```
#Menghitung data statistik
df.describe()
```

Hasil dari keluaran perintah diatas pada Gambar 4.10

	No	Jenis	Kelamin	IPS1	IPS2	IPS3	IPS4
count	536.000000	536.000000	536.000000	536.000000	536.000000	536.000000	536.000000
mean	268.500000	1.335821	3.539646	3.465187	3.506530	3.404291	
std	154.874142	0.472718	0.681517	0.449919	0.463876	0.509098	
min	1.000000	1.000000	0.000000	0.820000	0.000000	0.780000	
25%	134.750000	1.000000	3.430000	3.197500	3.220000	3.080000	
50%	268.500000	1.000000	3.710000	3.580000	3.610000	3.500000	
75%	402.250000	2.000000	3.900000	3.790000	3.870000	3.830000	
max	536.000000	2.000000	4.000000	4.000000	4.000000	4.000000	

Gambar 4. 10 Menghitung Data Statistik

4.3.4 Proses Data Mining

Pada proses data mining ini di lakukan proses yaitu :

1. Memisahkan atribut label

```
# memisahkan atribut dan label
X = df[['Jenis Kelamin', 'IPS1', 'IPS2', 'IPS3', 'IPS4' ]]
y = df['keterangan']
```

2. Membagi Dataset menjadi data latih dan Data uji

```
# Membagi dataset menjadi data latih & data uji
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
print ('Train set:', X_train.shape, y_train.shape)
print ('Test set:', X_test.shape, y_test.shape)
```

```
Train set: (375, 5) (375,)
Test set: (161, 5) (161,)
```

3. Membuat model *decision tree* dan membuat model dengan data latih

```
# membuat model Decision Tree
tree_model = DecisionTreeClassifier()

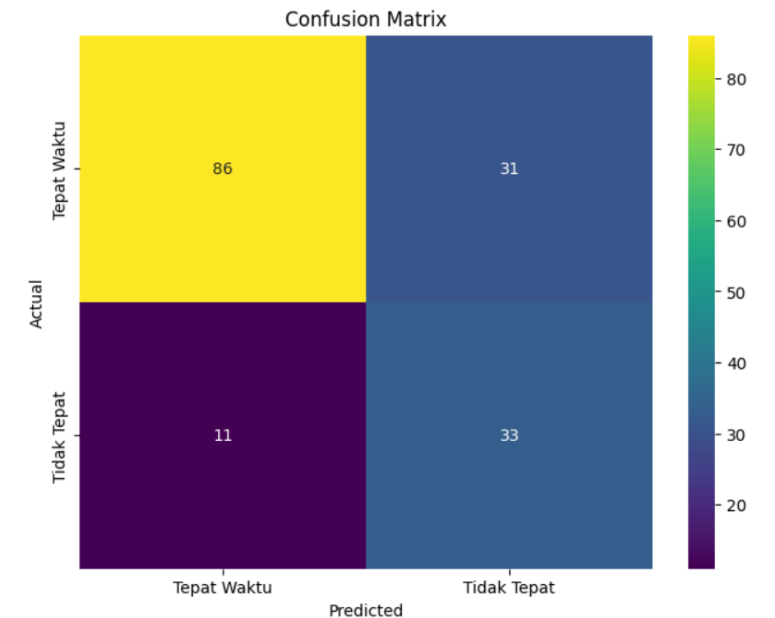
# Melatih model dengan menggunakan data latih
tree_model = tree_model.fit(X_train, y_train)

from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

# 6. Evaluasi Model
from sklearn.metrics import confusion_matrix, classification_report
y_pred = tree_model.predict(X_test)
labels=y
# Membuat confusion matrix
cm = confusion_matrix(y_test, y_pred)

# Mencetak confusion matrix
labels = [ 'Tepat Waktu', 'Tidak Tepat', ]
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', xticklabels=labels, yticklabels=labels, cmap='viridis')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
# Menghitung precision, recall, f1-score, dan support
report = classification_report(y_test, y_pred, target_names=labels)
print(report)
```

Hasil dari keluaran perintah di atas dapat di lihat pada Gambar 4.11



Gambar 4. 11 Confusion Matrix

4.3.5 Evaluasi

Pada proses Evaluasi ini dilakukan beberapa proses berikut :

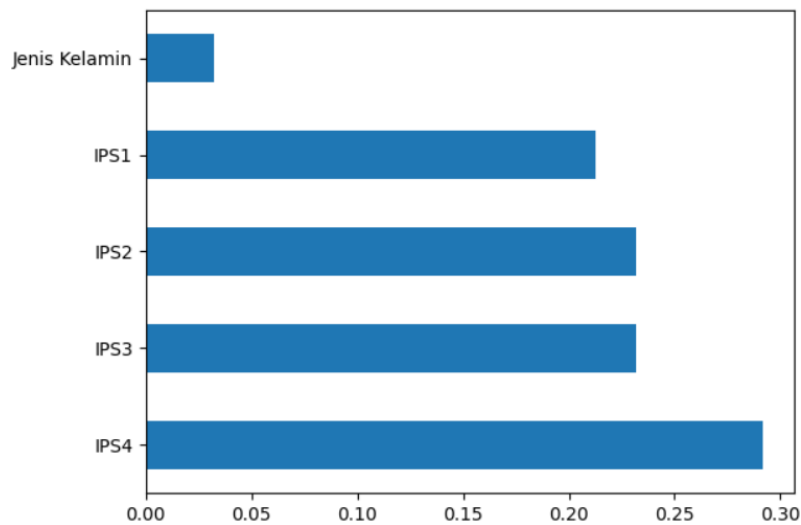
1. Seleksi Atribut

```
from sklearn.ensemble import ExtraTreesClassifier
import matplotlib.pyplot as plt
model = ExtraTreesClassifier()
model.fit(X,y)

print(model.feature_importances_) #use inbuilt class feature_importances of tree based classifiers

#plot graph of feature importances for better visualization
feat_importances = pd.Series(model.feature_importances_, index=X.columns)
feat_importances.nlargest(10).plot(kind='barh')
plt.show()
```

Hasil dari keluaran perintah diatas dapat dilihat pada Gambar 4.12 berikut



Gambar 4. 12 Seleksi Atribut

2. Pemangkasan Pohon

```
# Mengambil atribut yang berpengaruh ke hasil
X = df[['IPS1', 'IPS2', 'IPS3', 'IPS4' ]]
y = df['keterangan']
```

```
# Mengganti kategori kelas sesuai kondisi
df.loc[ (df['IPS4'] > 3.75), 'keterangan'] = 'Tepat Waktu'

# Menampilkan DataFrame setelah perubahan
print(df)
```


Hasil dari keluaran perintah diatas dapat dilihat pada gambar 4.13berikut:

	No	Jenis Kelamin	IPS1	IPS2	IPS3	IPS4	keterangan
0	1	1	3.48	3.26	2.70	3.00	Tidak Tepat
1	2	2	3.76	3.58	3.74	3.33	Tepat Waktu
2	3	1	3.38	2.95	3.60	3.42	Tidak Tepat
3	4	2	3.67	3.11	3.80	3.58	Tidak Tepat
4	5	2	3.67	3.58	3.91	3.42	Tepat Waktu
..
531	532	1	3.14	3.21	3.52	3.42	Tepat Waktu
532	533	1	0.00	3.75	4.00	3.70	Tepat Waktu
533	534	1	3.71	3.58	2.83	3.30	Tepat Waktu
534	535	1	4.00	4.00	4.00	4.00	Tepat Waktu
535	536	2	3.67	3.21	3.00	3.08	Tepat Waktu

[536 rows x 7 columns]

Gambar 4. 13 Gambar setelah Perubahan

3. Menganti kondisi kelas sesuai kondisi

```
# Mengganti kategori kelas sesuai kondisi
df.loc[ (df['IPS1'] <= 2), 'keterangan'] = 'Tidak Tepat'
df.loc[ (df['IPS2'] <= 2), 'keterangan'] = 'Tidak Tepat'
df.loc[ (df['IPS3'] <= 2), 'keterangan'] = 'Tidak Tepat'
df.loc[ (df['IPS4'] <= 2.96), 'keterangan'] = 'Tidak Tepat'

# Menampilkan DataFrame setelah perubahan
print(df)
```

Hasil dari keluaran perintah diatas dapat dilihat pada Gambar 4.14 berikut :

	No	Jenis Kelamin	IPS1	IPS2	IPS3	IPS4	keterangan
0	1	1	3.48	3.26	2.70	3.00	Tidak Tepat
1	2	2	3.76	3.58	3.74	3.33	Tepat Waktu
2	3	1	3.38	2.95	3.60	3.42	Tidak Tepat
3	4	2	3.67	3.11	3.80	3.58	Tidak Tepat
4	5	2	3.67	3.58	3.91	3.42	Tepat Waktu
..
531	532	1	3.14	3.21	3.52	3.42	Tepat Waktu
532	533	1	0.00	3.75	4.00	3.70	Tidak Tepat
533	534	1	3.71	3.58	2.83	3.30	Tepat Waktu
534	535	1	4.00	4.00	4.00	4.00	Tepat Waktu
535	536	2	3.67	3.21	3.00	3.08	Tepat Waktu

[536 rows x 7 columns]

Gambar 4. 14 Gambar mengganti kelas sesuai kondisi

4. Mengambil atribut yang berpengaruh ke hasil dan Membagi dataset menjadi data latih & data uji

```
# Mengambil atribut yang berpengaruh ke hasil
x = df[['IPS1', 'IPS2', 'IPS3', 'IPS4']]
y = df['keterangan']

# Membagi dataset menjadi data latih & data uji
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
print('Train set:', X_train.shape, y_train.shape)
print('Test set:', X_test.shape, y_test.shape)
```

Train set: (375, 4) (375,)
Test set: (161, 4) (161,)

5. Pengukuran Kedalaman pohon

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
# Mencoba berbagai nilai kedalaman pohon
for max_depth in range(1, 11):
    clf = DecisionTreeClassifier(max_depth=max_depth)
    clf.fit(X_train, y_train)
    y_pred = clf.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    print(f"Max Depth = {max_depth}, Accuracy = {accuracy:.2f}")

# Mencoba berbagai nilai minimum sampel untuk percabangan
for min_samples_split in range(2, 11):
    clf = DecisionTreeClassifier(min_samples_split=min_samples_split)
    clf.fit(X_train, y_train)
    y_pred = clf.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    print(f"Min Samples Split = {min_samples_split}, Accuracy = {accuracy:.2f}")
```

Hasil dari keluaran perintah diatas dapat dilihat pada Gambar 4.15 berikut:

```

Max Depth = 1, Accuracy = 0.87
Max Depth = 2, Accuracy = 0.88
Max Depth = 3, Accuracy = 0.88
Max Depth = 4, Accuracy = 0.90
Max Depth = 5, Accuracy = 0.92
Max Depth = 6, Accuracy = 0.89
Max Depth = 7, Accuracy = 0.91
Max Depth = 8, Accuracy = 0.92
Max Depth = 9, Accuracy = 0.89
Max Depth = 10, Accuracy = 0.89
Min Samples Split = 2, Accuracy = 0.84
Min Samples Split = 3, Accuracy = 0.86
Min Samples Split = 4, Accuracy = 0.84
Min Samples Split = 5, Accuracy = 0.84
Min Samples Split = 6, Accuracy = 0.84
Min Samples Split = 7, Accuracy = 0.86
Min Samples Split = 8, Accuracy = 0.86
Min Samples Split = 9, Accuracy = 0.86
Min Samples Split = 10, Accuracy = 0.87

```

Gambar 4. 15 Pengukuran Kedalaman Pohon

4.3.6 Perbaikan Model

Perbaikan model ini adalah Evaluasi Model kedua

```

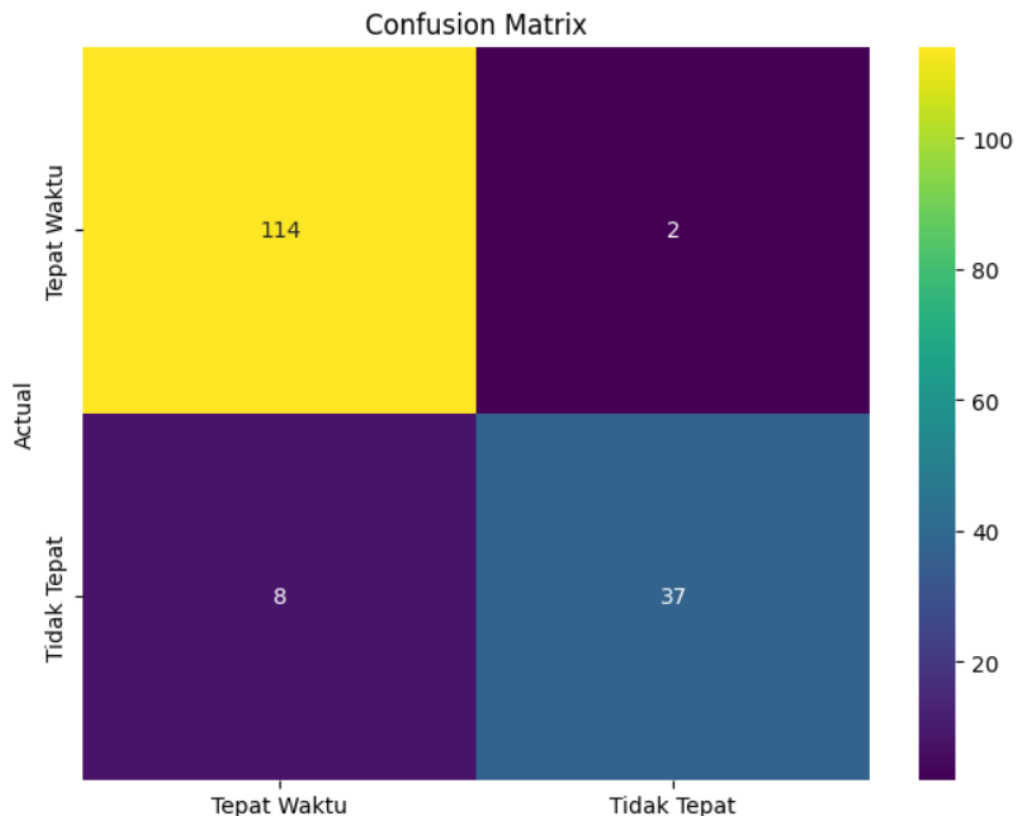
from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

# 6. Evaluasi Model
from sklearn.metrics import confusion_matrix, classification_report
y_pred = tree_model.predict(X_test)
labels=y
# Membuat confusion matrix
cm = confusion_matrix(y_test, y_pred)

# Mencetak confusion matrix
labels = [ 'Tepat Waktu', 'Tidak Tepat', ]
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', xticklabels=labels, yticklabels=labels, cmap='viridis')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
# Menghitung precision, recall, f1-score, dan support
report = classification_report(y_test, y_pred, target_names=labels)
print(report)

```

Hasil dari keluaran perintah diatas dapat dilihat pada Gambar 4.16



Gambar 4. 16 Gambar Confusion Matrix

4.3.7 Klasifikasi Data Testing Baru

Proses klasifikasi data testing baru dilakukan beberapa proses berikut :

1. Prediksi model dengan data baru dengan data dari Excel

```
#prediksi model dengan data baru dengan data dari Excel
data_tes = pd.read_excel("/content/drive/MyDrive/Skripsi/Data baru 5.xlsx")
data_tes.head()
```

Hasil dari keluaran perintah diatas dapat di lihat pada Gambar 4.17

	No	IPS1	IPS2	IPS3	IPS4	keterangan
0	1	3.48	3.26	2.70	3.00	Tidak Tepat
1	2	3.76	3.58	3.74	3.33	Tepat Waktu
2	3	3.38	2.95	3.60	3.42	Tidak Tepat
3	4	3.67	3.11	3.80	3.58	Tidak Tepat
4	5	3.67	3.58	3.91	3.42	Tepat Waktu

Gambar 4. 17 Gambar Prediksi model dengan data baru dengan data dari Excel

2. Mengkonversi file Excel

```
#Mengkonversi File Excel .xlsx ke .csv
data_tes.to_csv("/content/drive/MyDrive/Skripsi/Data baru 5.csv", index=False)
data_tes.info()
```

Hasil dari keluaran perintah diatas dapat dilihat pada Gambar 4.18

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0   No           5 non-null      int64
1   IPS1         5 non-null      float64
2   IPS2         5 non-null      float64
3   IPS3         5 non-null      float64
4   IPS4         5 non-null      float64
5   keterangan   5 non-null      object
dtypes: float64(4), int64(1), object(1)
memory usage: 368.0+ bytes
```

Gambar 4. 18 Gambar Mengkonversi File Excel

BAB V

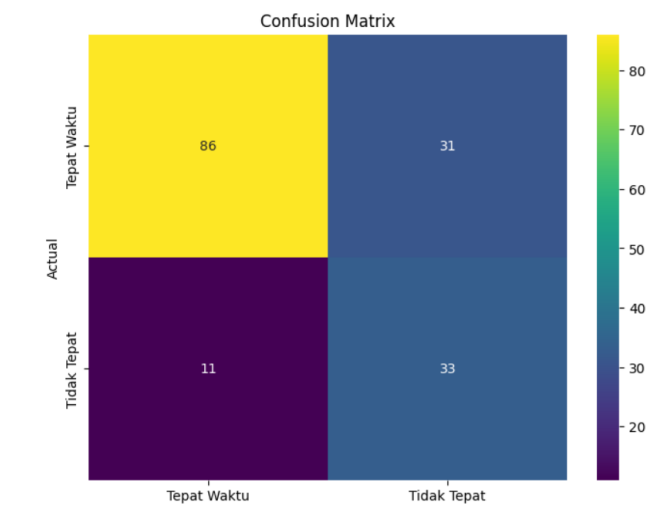
PEMBAHASAN PENELITIAN

4.1 Pembahasan Kinerja Model

Pemodelan Disecion tree dalam klasifikasi waktu kelulusan mahasiswa dilakukan uji coba pemodelan di lakukan sebanyak dukali pengujiannya ,setelah dilakukan obtimalisasi.

1. kinerja model sebelum obtimalisasi

setelah dilakukan pemodelan decision tree selanjutnya dilakukan pengujian model dengan menggunakan data testing sebanyak 161 adapun hasil pengujian akurasi model menggunakan confusion matrix sebagai mana yang di tampilkan pada Gambar 5.1 beikut.



Gambar 5. 1 *Confusion Matrix*

Berdasarkan tabel 5.1 di atas dapat dilakukan perhitungan lebih lanjut terkait akurasi presisi dan *Recall* sebagai berikut :

Keterangan	Precision	Recall	<i>F1-score</i>	Jumlah Data
Tidak Tepat	0.89	0.74	0.80	117
Tepat waktu	0.52	0.74	0.61	44
<i>Accuracy</i>	0.74			

2. kinerja model setelah optimalisasi

setelah dilakukan pemodelan decision tree selanjutnya dilakukan pengujian model kedua dengan menggunakan data testing sebanyak 161 adapun hasil pengujian akurasi model menggunakan confusion matrix sebagai mana yang di tampilkan pada Gambar 5.2 berikut.

Tabel 5. 1 Hasil Counfusion Matrix Sesudah Optimalisasi

Keterangan	Precision	Recall	<i>F1-score</i>	Jumlah Data
Tidak Tepat	0.93	0.98	0.96	116
Tepat waktu	0.95	0.82	0.88	45
<i>Accuracy</i>	0.94			

Pada gambar 5.1 ddiatas, terdapat hasil nilai akurasi dari proses prediksi mahasiswa menggunakan Confusion Matrix dengan nilai akurasi sebesar 90% yang diperoleh dari data testing yang berjumlah 162 data, dengan jumlah mahasiswa tepat waktu sebanyak 114 orang, jumlah mahasiswa tidak lulus tepat waktu sebanyak 37 orang, dan jumlah mahasiswa yang seharusnya tidak lulus tepat waktu tetapi diprediksi tepat waktu sebanyak 8 orang serta jumlah mahasiswa yang seharusnya lulus tepat waktu tetapi diprediksi sebagai tidak tepat waktu sebanyak 2 orang. Untuk mengukur performance metrics dari model evaluasi dengan Confusion Matix, dapat menggunakan accuracy, precission, dan recall. perhitungan manualnya menggunakan persamaan berikut :

1. Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FB+FN} \times 100 \%$$

$$Accuracy = \frac{114+37}{114+37+8+2} = \frac{151}{161} \times 100 \%$$

$$= 0,93 \times 100 \% = 93\%$$

2. Precision

$$Precision = \frac{TP}{TP+FP} \times 100 \%$$

$$Precision = \frac{114}{114+8} = \frac{114}{122} \times 100 \%$$

$$= 0,93 \times 100 \% = 93 \%$$

3. Recall

$$Recall = \frac{TP}{TP+FN} \times 100 \%$$

$$Recall = \frac{114}{114+2} = \frac{114}{116} \times 100 \%$$

$$= 0,98 \times 100 \% = 98\%$$

Berdasarkan perhitungan manual diatas, bisa disimpulkan bahwa evaluasi dengan menggunakan Confussion Matrix sudah sesuai dengan memperoleh nilai accuracy 93% , precission 93% dan recall 98%.

- Precision adalah salah satu metrik yang digunakan dalam evaluasi kinerja model klasifikasi, precision mengukur sejauh mana model klasifikasi mampu mengindetifikasi data positif dengan benar dari semua data yang di prediksi sebagai positing.
- Recall adalah salah satu metrik evaluasi yang digunakan dalam berbagai bidang seperti klasifikasi data, mengukur sejauh mana model atau system mampu mengidentifikasi seluruh instance dari kelas yang sebenarnya positif dalam sebuah dataset.
- *F1-score* adalah salah satu metrik evaluasi yang umum digunakan dalam

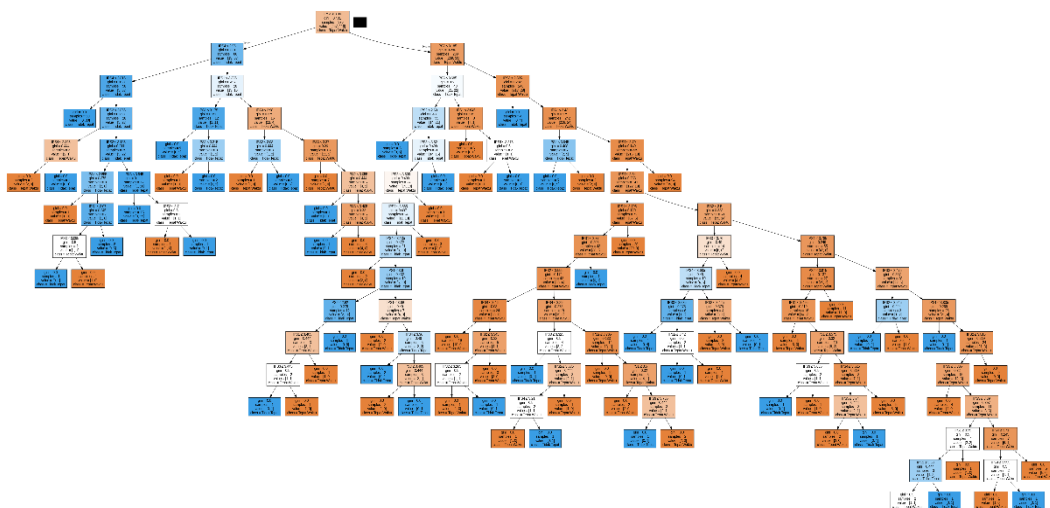
evaluasi kinerja model klasifikasi.

Berdasarkan pengujian kinerja model *Decision Tree* yang dilakukan dengan 2 kali uji coba yaitu sebelum optimalisasi model didapatkan hasil akurasi sebesar 74% sebagai mana dapat dilihat pada tabel 5.1 di atas dan setelah dilakukan optimalisasi model didapatkan hasil akurasi sebesar 94% seperti pada 5.4 di atas, demikian dapat dilihat bahwa terjadi peningkatan akurasi sebesar 20% setelah optimalisasi model dilakukan.

4.2 Pembahasan Hasil Pemodelan

Hasil pemodelan *Decision Tree* pada Klasifikasi Waktu Kelulusan Mahasiswa dilakukan 2 kali tahapan ujicoba yaitu pemodelan yang dilakukan seperti seleksi atribut, mengukur kedalaman pohon dan pemangkasan pohon. Adapun hasil masing-masing ujicoba tersebut diuraikan pada pembahasan berikut

Hasil pemodelan *Decision tree* sebelum melakukan optimalisasi didapatkan hasil visualisasi pohon sebagai di tampilkan pada Gambar 5.3 berikut ini :



Gambar 5. 2 Visualisasi Pohon Keputusan Sebelum Optimalisasi

Berdasarkan Gambar 5.2 diatas dapat dilihat bahwa pohon keputusan yang terbentuk belum maksimal dan masi terlalu banyak jalur keputusan yang tidak signifikan yaitu terdapat lebih dari 15 jalur keputusan , hal ini terjadi karena masi terdapat data yang tidak konsisten terhadap kelas output, singga untuk mengatasi masalah tersebut perlu dilakukan proses optimalisasi model.

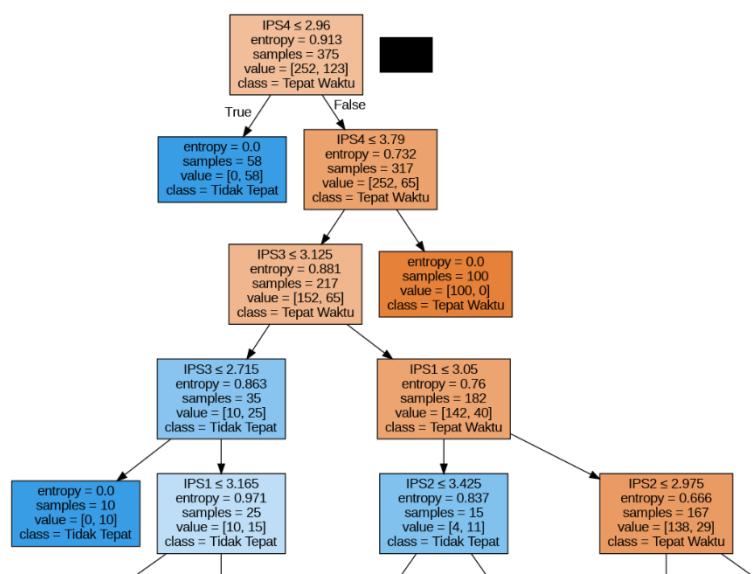
4.3 Hasil Pemodelan Setelah Optimalisasi

Hasil pemodelan *Decision Tree* setelah dilakukan optimalisasi model terkait seleksi atribut , pemangkasan pohon, dan mengukur kedalam pohon .

Pada optimalisasi seleksi data atribut terpilih atribut IPS1, IPS2, IPS3, IPS4 sesuai hasil pengukuran menggunakan metode *Disicion Tree*

Pada optimalisasi pemangkasan pohon dilakukan pemangkasan $IPS > 3.72$ dijadikan kelasnya Menjadi Tepat Waktu semua , untuk $IPS < 2.96$ dijadikan kelasnya menjadi Tidak Tepat Waktu.

Pada optimalisasi pengukuran kedalaman didapatkan hasil kedalaman pohon 5 yang memiliki tingkat akurasi tertinggi, sehingga nilai kedalaman ini yang digunakan dalam pemodelan *Decision Tree* Berdasarkan hasil optimalisasi tersebut diatas



Gambar 5. 3 Visualisasi Pohon Setelah Optimalisasi

dijadikan sebagai pemodelan *Decision Tree* yang digunakan dalam penelitian ini dengan hasil visualisasi pada pohon keputusan dapat dilihat pada Gambar 5.3 berikut ini:

Berdasarkan Gambar 5.4 di atas dapat dilihat bahwa pohon keputusan yang terbentuk sudah lebih sederhana dan jumlah jalur atau rule keputusan yang terbentuk sisa 9 rule. Adapun jalur atau rule keputusan secara detail diuraikan pada Tabel 5.3 berikut ini :

Tabel 5. 2 Rule Keputusan

No	Rule Keputusan
1	Jika $IPS4 < 2.96$ maka Tidak Tepat
2	Jika $IPS4 > 2.96$ dan $IPS4 < 3.79$ dan $IPS3 < 3.125$ dan $IPS3 < 2.715$ maka Tidak Tepat
3	Jika $IPS4 > 2.96$ dan $IPS4 < 3.79$ dan $IPS3 < 3.125$ dan $IPS3 < 2.715$ dan $IPS1 < 3.165$ maka Tidak Tepat
4	Jika $IPS4 > 2.96$ dan $IPS4 < 3.79$ dan $IPS3 < 3.125$ dan $IPS3 < 2.715$ dan $IPS1 > 3.165$ maka Tepat Waktu
5	Jika $IPS4 > 2.96$ dan $IPS4 > 3.79$ maka Tepat Waktu
6	Jika $IPS4 > 2.96$ dan $IPS4 > 3.79$ dan $IPS3 < 3.125$ dan $IPS1 > 3.05$ dan $IPS2 > 2.975$ maka Tidak Tepat
7	Jika $IPS4 > 2.96$ dan $IPS4 > 3.79$ dan $IPS3 < 3.125$ dan $IPS1 > 3.05$ dan $IPS2 < 2.975$ maka Tepat Waktu
8	Jika $IPS4 > 2.96$ dan $IPS4 > 3.79$ dan $IPS3 < 3.125$ dan $IPS1 > 3.05$ dan $IPS2 > 3.425$ maka Tepat Waktu
9	Jika $IPS4 > 2.96$ dan $IPS4 > 3.79$ dan $IPS3 < 3.125$ dan $IPS1 > 3.05$ dan $IPS2 < 3.425$ maka Tidak Tepat

Berdasarkan rule keputusan di atas, Klasifikasi Waktu Kelulusan Mahasiswa dapat ditentukan dengan cukup melihat pohon atau rule keputusan tersebut. Rule keputusan ini juga bisa diterapkan pada sistem informasi akademik di Fakultas kedepannya agar dengan mudah bisa dilakukan monitoring atau evaluasi Waktu Kelulusan Mahasiswa.

BAB VI

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang sudah dilakukan terkait klasifikasi waktu kelulusan mahasiswa dapat disimpulkan bahwa :

1. Penerapan metode *Decision tree* didapatkan rule keputusan sebanyak 5 rule, dengan atribut yang terpilih yaitu: IPS1, IPS2, IPS3, IPS4 hasil didapatkan berdasarkan hasil optimalisasi model yaitu seleksi fitur, pemangkasan pohon dan pengukuran kedalaman pohon.
2. Penerapan metode *Decision Tree* pada Klasifikasi Waktu kelulusan Mahasiswa di dapatkan hasil akurasi sebesar 94% dengan demikian metode klasifikasi *decision tree* dapat digunakan untuk mengklasifikasi waktu kelulusan mahasiswa, dengan menggunakan metode ini , perguruan tinggi dapat mengidentifikasi mahasiswa yang beresiko tidak lulus tepat waktu dan memberikan intervensi yang di perlukan untuk membantu mereka meyelesaikan studinya

Berdasarkan kesimpulan laporan tersebut diatas, peneliti dapat memberikan saran untuk peneliti selanjutnya, yaitu:

1. Penelitian selanjutnya dapat dilakukan dengan menggunakan data yang lebih besar dan lebih beragam. Hal ini akan dapat meningkatkan akurasi dari metode *decision tree*.
2. Penelitian dapat juga di lakukan untuk menguji efektivitas metode *decision tree* dalam pemprediksi waktu kelulusan mahasiswa di berbagai perguruan tinggi. Hal ini akan dapat memberikan gambaran yang lebih luas tentang efektivitas metode ini.

DAFTAR PUSTAKA

- [1] R. Maulana and D. Kumalasari, “Analisis Komparasi Agoritma Klasifikasi Data Mining Untuk Prediksi Status Kelulusan Mahasiswa Akademik Bina Sarana Informatika,” 2019.
- [2] A. Kadir Jaelani, “Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Akademi Farmasi Jember.”
- [3] S. P. Y. Pramana, “Proyek Machine Learning ,” 2021.
- [4] D. Dalbergio, M. N. Hayati, and Y. N. Nasution, “Klasifikasi Lama Studi Mahasiswa Menggunakan Metode Algoritma C5.0 Pada Stui Kasus Data Kelulusan Mahasiswa Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Mulawarman Tahun 2017,” 2019.
- [5] F. Rezkika *et al.*, “Klasifikasi Masa Tunggu Alumni Untuk Mendapatkan Pekerjaan Menggunakan Algoritma C4.5 (Studi Kasus : Fasilkom Unsika),” 2021.
- [6] M. I. A. Akaseh, “Aplikasi Data Mining Untuk Pemilihan Jurusan IPA dan IPS untuk Siswa Kelas X Pada Sekolah Menengah Umum Negeri 1 Telaga Kabupaten Gorontalo,” 2020.
- [7] “Pengelompokkan Kasus Covid-19 Di Indonesia Menggunakan Metode Density-Based Spatial Clustering Of Applicatios With Noise (Dbscan).” [Online]. Available: <http://repository.unimus.ac.id>
- [8] H. Handayani, “Pengelompokan Indsutri Mikro dan Kecil Di Indonesia Menggunakan Kohonen Self Organizing Maps (SOM),” 2017.
- [9] E. Triyanto, H. Sismoro, and A. D. Laksito, “Implementasi Algoritma Regresi Linear Berganda Untuk Meprediksi Produksi Padi di Kabupaten Bantul,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 4, no. 2, pp. 66–75, Jul. 2019, doi: 10.36341/rabit.v4i2.666.

- [1] R. Maulana and D. Kumalasari, “Analisis Komparasi Agoritma Klasifikasi Data Mining Untuk Prediksi Status Kelulusan Mahasiswa Akademik Bina Sarana Informatika,” 2019.
- [2] A. Kadir Jaelani, “Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Akademi Farmasi Jember.”
- [3] S. P. Y. Pramana, “Proyek Machine Learning ,” 2021.
- [4] D. Dalbergio, M. N. Hayati, and Y. N. Nasution, “Klasifikasi Lama Studi Mahasiswa Menggunakan Metode Algoritma C5.0 Pada Stui Kasus Data Kelulusan Mahasiswa Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Mulawarman Tahun 2017,” 2019.
- [5] F. Rezkika *et al.*, “Klasifikasi Masa Tunggu Alumni Untuk Mendapatkan Pekerjaan Menggunakan Algoritma C4.5 (Studi Kasus : Fasilkom Unsika),” 2021.
- [6] M. I. A. Akaseh, “Aplikasi Data Mining Untuk Pemilihan Jurusan IPA dan IPS untuk Siswa Kelas X Pada Sekolah Menengah Umum Negeri 1 Telaga Kabupaten Gorontalo,” 2020.
- [7] “Pengelompokkan Kasus Covid-19 Di Indonesia Menggunakan Metode Density-Based Spatial Clustering Of Applicatios With Noise (Dbscan).” [Online]. Available: <http://repository.unimus.ac.id>
- [8] H. Handayani, “Pengelompokan Indsutri Mikro dan Kecil Di Indonesia Menggunakan Kohonen Self Organizing Maps (SOM),” 2017.
- [9] E. Triyanto, H. Sismoro, and A. D. Laksito, “Implementasi Algoritma Regresi Linear Berganda Untuk Meprediksi Produksi Padi di Kabupaten Bantul,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 4, no. 2, pp. 66–75, Jul. 2019, doi: 10.36341/rabit.v4i2.666.
- [10] S. Septiawanti, “Penerapan Metode Case-Based Reasoning Untuk Prediksi Data Nasabah yang Berpotensi Membuka Simpanan Deposito,” 2018.

- [11] R. B. Mahaputra, “Rancang Bangun Aplikasi Ujian Akhir Semester Online Untuk Mengukur Pencapaian Kompetensi Siswa”.
- [12] D. A. Novilla, R. Goejantoro, F. Deny, and T. Amijaya, “Classification of Insurance Data Customers Using Naive Bayes Method (Case Study : PT. Prudential Life MT. Haryono Street Samarinda),” *Jurnal EKSPONENSIAL*, vol. 10, no. 2, 2019.
- [13] A. H. Nasrullah, “Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris,” vol. 7, no. 2, 2021, [Online]. Available: <http://ejournal.fikom-unasman.ac.id>
- [14] G. Taufik and D. Jatmika, “Penerapan Algoritma C4.5 Untuk Klasifikasi Keberhasilan Pengiriman Barang,” vol. 6, no. 1, 2021.
- [15] Z. Gustiana, “Penerapan Algoritma C 4.5 Dalam Sistem Pendukung Keputusan Evaluasi Kinerja Fasilitator Pamsimas (Studi Kasus di Kabupaten Kampar),” 2020.
- [16] K. Suhada, A. Elanda, and A. Aziz, “Klasifikasi Predikat Tingkat Kelulusan Mahasiswa Program Studi Teknik Informatika dengan Menggunakan Algoritma C4.5 (Studi Kasus: STMIK Rosma Karawang),” 2021.

LAMPIRAN

```
#Mengimport Library
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.tree import DecisionTreeClassifier

%matplotlib inline
#Proses Pengaitan dengan Google Drive
from google.colab import drive
drive.mount('/content/drive')

#Membaca Dataset
df =
pd.read_excel("/content/drive/MyDrive/Skripsi/Dataset2.xlsx")
df.head()

#Mengkonversi File Excel .xlsx ke .csv
df.to_csv("/content/drive/MyDrive/Skripsi/Dataset2.csv",
index=False)
df.info()

# Menghitung jumlah setiap kategori dalam kolom 'Keterangan
Lulus'
df['keterangan'].value_counts()

# Menghitung jumlah setiap kategori dalam kolom 'Keterangan
Lulus'
countplot = sns.countplot(data=df, x='keterangan')

# Menambahkan label pada sumbu x dan y
plt.xlabel('Keterangan')
plt.ylabel('Jumlah Data')

# Menambahkan judul
plt.title('Distribusi Kelulusan ')

# Menampilkan grafik
plt.show()
```

```

# Mengelompokkan data berdasarkan atribut yang ingin diperiksa
grouped_data = df.groupby(['Jenis Kelamin', 'IPS1', 'IPS2',
'IPS3', 'IPS4'])

# Memeriksa kategori kelas untuk setiap kelompok
for key, group in grouped_data:
    unique_classes = group['keterangan'].unique()
    if len(unique_classes) > 1:
        print(f"Kelompok {key}: Kelas yang berbeda =
{unique_classes}")

# Mengganti kategori kelas sesuai kondisi
df.loc[(df['Jenis Kelamin'] == 'L') & (df['IPS1'] == 3.81) &
(df['IPS2'] == 3.79) & (df['IPS3'] == 3.7) & (df['IPS4'] ==
3.83), 'keterangan'] = 'Tepat Waktu'
df.loc[(df['Jenis Kelamin'] == 'L') & (df['IPS1'] == 3.9) &
(df['IPS2'] == 3.79) & (df['IPS3'] == 3.91) & (df['IPS4'] ==
3.75), 'keterangan'] = 'Tepat Waktu'
df.loc[(df['Jenis Kelamin'] == 'L') & (df['IPS1'] == 4.0) &
(df['IPS2'] == 4.0) & (df['IPS3'] == 3.91) & (df['IPS4'] == 4.0),
'keterangan'] = 'Tepat Waktu'
df.loc[(df['Jenis Kelamin'] == 'L') & (df['IPS1'] == 4.0) &
(df['IPS2'] == 4.0) & (df['IPS3'] == 4.0) & (df['IPS4'] == 4.0),
'keterangan'] = 'Tepat Waktu'
# Menampilkan DataFrame setelah perubahan
print(df)

# Menghitung jumlah setiap kategori dalam kolom 'Keterangan
Lulus'
df['keterangan'].value_counts()

# Menghitung jumlah setiap kategori dalam kolom 'Keterangan
Lulus'
countplot = sns.countplot(data=df, x='keterangan')

# Menambahkan label pada sumbu x dan y
plt.xlabel('Keterangan')
plt.ylabel('Jumlah Data')

# Menambahkan judul
plt.title('Distribusi Waktu Kelulusan ')

# Menampilkan grafik
plt.show()

```

```

#Proses Konversi Nilai String ke Numerik
df['Jenis Kelamin'].replace(['L','P'],[1,2], inplace=True)
df.head()

#Menghitung data statistik
df.describe()

# memisahkan atribut dan label
X = df[['Jenis Kelamin','IPS1','IPS2', 'IPS3','IPS4' ]]
y = df['keterangan']

# Membagi dataset menjadi data latih & data uji
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=0)
print ('Train set:', X_train.shape,  y_train.shape)
print ('Test set:', X_test.shape,  y_test.shape)

# membuat model Decision Tree
tree_model = DecisionTreeClassifier()

# Melatih model dengan menggunakan data latih
tree_model = tree_model.fit(X_train, y_train)

from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

# 6. Evaluasi Model
from sklearn.metrics import confusion_matrix,
classification_report
y_pred = tree_model.predict(X_test)
labels=y
# Membuat confusion matrix
cm = confusion_matrix(y_test, y_pred)

```

```

from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

# 6. Evaluasi Model
from sklearn.metrics import confusion_matrix,
classification_report
y_pred = tree_model.predict(X_test)
labels=y
# Membuat confusion matrix
cm = confusion_matrix(y_test, y_pred)

# Mencetak confusion matrix
labels = [ 'Tepat Waktu', 'Tidak Tepat', ]
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', xticklabels=labels ,
yticklabels=labels, cmap='viridis')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
# Menghitung precision, recall, f1-score, dan support
report = classification_report(y_test, y_pred,
target_names=labels)
print(report)

dot_data = StringIO()
filename = "/content/drive/MyDrive/Skripsi/Pohon1.png"
featureNames = ['Jenis Kelamin', 'IPS1', 'IPS2', 'IPS3', 'IPS4']
targetNames = ['Tepat Waktu', 'Tidak Tepat' ]
out = tree.export_graphviz(tree_model,
feature_names=featureNames, out_file=dot_data, class_names=
np.unique(y_train), filled=True, special_characters=True,
rotate=False)
graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png(filename)
img = mpimg.imread(filename)
plt.figure(figsize=(200, 100))
plt.imshow(img, interpolation='nearest')

```

```

from six import StringIO
import pydotplus
import matplotlib.image as mpimg
from sklearn import tree

%matplotlib inline

from sklearn.ensemble import ExtraTreesClassifier
import matplotlib.pyplot as plt
model = ExtraTreesClassifier()
model.fit(X,y)

print(model.feature_importances_) #use inbuilt class
feature_importances of tree based classifiers

#plot graph of feature importances for better visualization
feat_importances = pd.Series(model.feature_importances_,
index=X.columns)
feat_importances.nlargest(10).plot(kind='barh')
plt.show()

# Mengambil atribut yang berpengaruh ke hasil
X = df[['IPS1', 'IPS2', 'IPS3', 'IPS4' ]]
y = df['keterangan']

# Mengganti kategori kelas sesuai kondisi
df.loc[ (df['IPS4'] > 3.75), 'keterangan'] = 'Tepat Waktu'

# Menampilkan DataFrame setelah perubahan
print(df)

# Mengganti kategori kelas sesuai kondisi
df.loc[ (df['IPS1'] <= 2), 'keterangan'] = 'Tidak Tepat'
df.loc[ (df['IPS2'] <= 2), 'keterangan'] = 'Tidak Tepat'
df.loc[ (df['IPS3'] <= 2), 'keterangan'] = 'Tidak Tepat'
df.loc[ (df['IPS4'] <= 2.96), 'keterangan'] = 'Tidak Tepat'

# Menampilkan DataFrame setelah perubahan
print(df)

```

```

# Mengganti kategori kelas sesuai kondisi
df.loc[ (df['IPS4'] > 3.75), 'keterangan'] = 'Tepat Waktu'

# Menampilkan DataFrame setelah perubahan
print(df)

# Mengambil atribut yang berpengaruh ke hasil
X = df[['IPS1', 'IPS2', 'IPS3', 'IPS4' ]]
y = df['keterangan']

# Membagi dataset menjadi data latih & data uji
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=0)
print ('Train set:', X_train.shape, y_train.shape)
print ('Test set:', X_test.shape, y_test.shape)

from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
# Mencoba berbagai nilai kedalaman pohon
for max_depth in range(1, 11):
    clf = DecisionTreeClassifier(max_depth=max_depth)
    clf.fit(X_train, y_train)
    y_pred = clf.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    print(f"Max Depth = {max_depth}, Accuracy = {accuracy:.2f}")

# Mencoba berbagai nilai minimum sampel untuk percabangan
for min_samples_split in range(2, 11):
    clf =
DecisionTreeClassifier(min_samples_split=min_samples_split)
    clf.fit(X_train, y_train)
    y_pred = clf.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    print(f"Min Samples Split = {min_samples_split}, Accuracy =
{accuracy:.2f}")

```

```

# membuat model Decision Tree
#tree_model = DecisionTreeClassifier()
clf = DecisionTreeClassifier(max_depth=5, min_samples_split=3,
min_samples_leaf=4, criterion="entropy")
#clf = DecisionTreeClassifier(max_depth=5, min_samples_split=5)
# Melatih model dengan menggunakan data latih
tree_model = clf.fit(X_train, y_train)

from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

# 6. Evaluasi Model
from sklearn.metrics import confusion_matrix,
classification_report
y_pred = tree_model.predict(X_test)
labels=y
# Membuat confusion matrix
cm = confusion_matrix(y_test, y_pred)

# Mencetak confusion matrix
labels = [ 'Tepat Waktu', 'Tidak Tepat', ]
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', xticklabels=labels ,
yticklabels=labels, cmap='viridis')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
# Menghitung precision, recall, f1-score, dan support
report = classification_report(y_test, y_pred,
target_names=labels)
print(report)

```

```

import os

# Where to save the figures
#content/drive/MyDrive/Skripsi
PROJECT_ROOT_DIR = "."
CHAPTER_ID = "decision_trees"
IMAGES_PATH = os.path.join(PROJECT_ROOT_DIR, "Skripsi",
CHAPTER_ID)
os.makedirs(IMAGES_PATH, exist_ok=True)

def save_fig(tree_model, tight_layout=True, fig_extension="png",
resolution=300):
    path = os.path.join(IMAGES_PATH, tree_model + "." +
fig_extension)
    print("Saving figure", tree_model)
    if tight_layout:
        plt.tight_layout()
    plt.savefig(path, format=fig_extension, dpi=resolution)

from graphviz import Source
from sklearn.tree import export_graphviz
export_graphviz(
    tree_model,
    out_file=os.path.join(IMAGES_PATH, "lulus_tree7.dot"),
    feature_names = ['IPS1', 'IPS2', 'IPS3', 'IPS4'],
    class_names = ['Tepat Waktu', 'Tidak Tepat' ],
    rounded= True,
    filled =True)
Source.from_file(os.path.join(IMAGES_PATH, "lulus_tree7.dot"))

dot_data = StringIO()
filename = "/content/drive/MyDrive/Skripsi/Pohon2.png"
featureNames = ['IPS1', 'IPS2', 'IPS3', 'IPS4']
targetNames = ['Tepat Waktu', 'Tidak Tepat' ]
out = tree.export_graphviz(tree_model,
feature_names=featureNames, out_file=dot_data, class_names=
np.unique(y_train), filled=True, special_characters=True,
rotate=False)
graph = pydotplus.graph_from_dot_data(dot_data.getvalue())

```



```

graph.write_png(filename)
img = mpimg.imread(filename)
plt.figure(figsize=(200, 100))
plt.imshow(img, interpolation='nearest')

# prediksi model dengan data baru 1(['IPS1', 'IPS2', 'IPS3',
'IPS4'])
hasil=tree_model.predict([[2.3,2.3,3.6,2]])
print(hasil)

#prediksi model dengan data baru dengan data dari Excel
data_tes = pd.read_excel("/content/drive/MyDrive/Skripsi/Data
baru 5.xlsx")
data_tes.head()

#Mengkonversi File Excel .xlsx ke .csv
data_tes.to_csv("/content/drive/MyDrive/Skripsi/Data baru 5.csv",
index=False)
data_tes.info()

dataX = data_tes[['IPS1','IPS2', 'IPS3','IPS4' ]]
print(dataX)

# prediksi model dengan data baru (IPS1, IPS2,IPS3,IPS4)
hasil=tree_model.predict(dataX)
print(hasil)

```



**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,
RISET DAN TEKNOLOGI**
UNIVERSITAS ICHSAN
(UNISAN) GORONTALO

SURAT KEPUTUSAN MENDIKNAS RI NO. 84/D/O/2001

JL. Achmad Nadjamuddin No. 17 Telp. (0435) 829975 Fax (0435) 829976 Gorontalo

SURAT KETERANGAN
NO : 166/UNISAN-G/SKT/XI/2023

Yang bertanda tangan dibawah ini :

Nama : Amiruddin, S.Kom., M.Kom, MCF
NIDN : 0910097601
Jabatan : Wakil Rektor Bidang Akademik, Tata Kelola & Sistem Informasi

Menyatakan,

Nama : Geby Anggriani Mundok
Nim : T3119071
Prodi : Teknik Informatika
Fakultas : Ilmu Komputer

Bahwa yang bersangkutan diberikan izin penelitian di Universitas Ichsan Gorontalo dengan judul penelitian **“Klasifikasi Waktu Kelulusan Mahasiswa Fakultas Hukum Menggunakan Metode Decision Tree”** Kepada calon peneliti diharapkan:

1. Mematuhi segala ketentuan dan peraturan yang berlaku di Universitas Ichsan Gorontalo
2. Data penelitian bersifat rahasia dan tidak di izinkan menyampaikan informasi kepada pihak-pihak lain yang tidak berkepentingan.

Demikian surat keterangan ini dibuat untuk dapat dipergunakan sebagaimana mestinya,

Gorontalo, 21 November 2023

An. Rektor
Wakil Rektor Bidang Akademik, Tata Kelola &
Sistem Informasi



Amiruddin, S.Kom., M.Kom, MCF
NIDN. 0910097601

Tembusan:

1. Rektor Universitas Ichsan Gorontalo
2. Wakil Rektor I Bidang Akademik, Tata Kelola & Sistem Informasi
3. Mahasiswa yang bersangkutan
4. Arsip



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO
FAKULTAS ILMU KOMPUTER

SURAT KEPUTUSAN MENDIKNAS RI NOMOR 84/D/O/2001
Jl. Achmad Najamuddin No. 17 Telp. (0435) 829975 Fax (0435) 829976 Gorontalo

SURAT REKOMENDASI BEBAS PLAGIASI
No. 483/FIKOM-UIG/R/XI/2023

Yang bertanda tangan di bawah ini :

Nama : Irvan Abraham Salihi, M.Kom
NIDN : 0928028101
Jabatan : Dekan Fakultas Ilmu Komputer

Dengan ini menerangkan bahwa :

Nama Mahasiswa : Geby Anggriani Mundok
NIM : T3119071
Program Studi : Teknik Informatika (S1)
Fakultas : Fakultas Ilmu Komputer
Judul Skripsi : Klasifikasi Waktu Kelulusan Mahasiswa Menggunakan Metode Decision Tree

Sesuai hasil pengecekan tingkat kemiripan skripsi melalui aplikasi **Turnitin** untuk judul skripsi di atas diperoleh hasil *Similarity* sebesar **9%**, berdasarkan Peraturan Rektor No. 32 Tahun 2019 tentang Pendeteksian Plagiat pada Setiap Karya Ilmiah di Lingkungan Universitas Ichsan Gorontalo dan persyaratan pemberian surat rekomendasi verifikasi calon wisudawan dari LLDIKTI Wil. XVI, bahwa batas kemiripan skripsi maksimal 30%, untuk itu skripsi tersebut di atas dinyatakan **BEBAS PLAGIASI** dan layak untuk diujikan.

Demikian surat rekomendasi ini dibuat untuk digunakan sebagaimana mestinya.

Mengetahui
Dekan,



Irvan Abraham Salihi, M.Kom
NIDN. 0928028101

Gorontalo, 20 November 2023
Tim Verifikasi,



Zulfrianto Y. Lamasigi, M.Kom
NIDN. 0914089101

Terlampir :
Hasil Pengecekan Turnitin



Similarity Report ID: oid:25211:47009579

PAPER NAME

SKRIPSI_T3119071_GEBY ANGGRIANI M
UNDOK.pdf

AUTHOR

Geby Anggriani Mundok gebyanggriani@
gmail.com

WORD COUNT

8967 Words

CHARACTER COUNT

50692 Characters

PAGE COUNT

69 Pages

FILE SIZE

3.0MB

SUBMISSION DATE

Nov 20, 2023 11:21 AM GMT+8

REPORT DATE

Nov 20, 2023 11:22 AM GMT+8

● 9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 9% Internet database
- 3% Publications database
- Crossref database
- Crossref Posted Content database
- 0% Submitted Works database

● Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less than 10 words)



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO
FAKULTAS ILMU KOMPUTER
UPT. PERPUSTAKAAN FAKULTAS
SK. MENDIKNAS RI NO. 84/D/0/2001

Jl. Achmad Nadjamuddin No.17 Telp(0435) 829975 Fax. (0435) 829976 Gorontalo

SURAT KETERANGAN BEBAS PUSTAKA

No : 014/Perpustakaan-Fikom/XI/2023

Perpustakaan Fakultas Ilmu Komputer (FIKOM) Universitas Ichsan Gorontalo dengan ini menerangkan bahwa :

Nama Anggota : Geby Aggriani Mundok

No. Induk : T3119071

No. Anggota : M202376

Terhitung mulai hari, tanggal : Jumat, 10 November 2023, dinyatakan telah bebas pinjam buku dan koleksi perpustakaan lainnya.

Demikian keterangan ini di buat untuk di pergunakan sebagaimana mestinya.



Gorontalo, 10 November 2023

**Mengetahui,
Kepala Perpustakaan**

Apriyanto Alhamad, M.Kom

NIDN : 0924048601

RIWAYAT HIDUP PENELITI



Nama : Geby Anggriani Mundok
 NIM : T3119071
 Tempat, Tgl Lahir : Toluaya, 08 Juli 2001
 Pekerjaan : Mahasiswa
 Agama : Islam
 Jenis Kelamin : Perempuan
 Fakultas/Jurusan : Ilmu Komputer/Teknik
 Informatika
 Konsentrasi : Software Engineering
 Alamat : Jl, Makasar
 Email : gebyanggriani@gmail.com

Riwayat Pendidikan:

1. SDN Soguo (2007 – 2013)
2. SMP Negeri Molibagu (2013 – 2016)
3. SMA Negeri 1 Bolaang Uki (2016 – 2019)
4. Universitas Ichsan Gorontalo (2019 – Sekarang)

