

**PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU
MENGGUNAKAN METODE *DECISION TREE***

(Studi Kasus : Program Studi Teknik Informatika Fakultas Ilmu Komputer Unisan Gorontalo)

Oleh

FRENGKI DAUD

T3119131

SKRIPSI

**Untuk memenuhi salah satu syarat ujian
Guna memperoleh gelar sarjana**



**PROGRAM SARJANA
TEKNIK INFORMATIKA
UNIVERSITAS ICHSAN GORONTALO
GORONTALO
2022**

HALAMAN PERSETUJUAN SKRIPSI

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN METODE *DECISION TREE*

(Studi Kasus : Program Studi Teknik Informatika Fakultas Ilmu Komputer Unisan Gorontalo)

Oleh

FRENGKI DAUD

T3119131

SKRIPSI

Untuk memenuhi salah satu syarat ujian
Guna untuk memperoleh gelar sarjana
Program Studi Teknik Pemimping

Telah disetujui dan siap diseminarkan

Gorontalo, Maret 2023

Pembimbing I


Rezqiwati Ishak, M.kom
NIDN : 0903087901

Pembimbing II


Warid Yunus, M.kom
NIDN : 0914059001

PENGESAHAN SKRIPSI

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN METODE **DECISION TREE**

(Studi Kasus : Program Studi Teknik Informatika Fakultas Ilmu Komputer Unisan Gorontalo)

Oleh

FRENGKI DAUD

T3119131

SKRIPSI

Diperiksa oleh Panitia Ujian Strata Satu (S1)

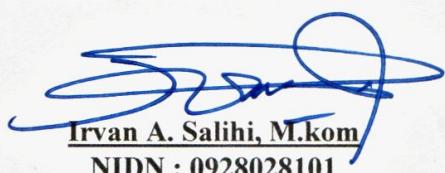
Universitas Ichsan Gorontalo

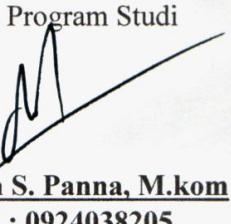
1. Ketua Pengujii
Zohrahayati, M.Kom
2. Anggota
Muis Nanja, M.Kom
3. Anggota
Maryam Hasan, M.Kom
4. Anggota
Rezqiwati Ishak, M.Kom
5. Anggota
Warid Yunus, M.Kom

Mengetahui,

Dekan Fakultas Ilmu Komputer

Ketua Program Studi


Irvan A. Salihi, M.kom
NIDN : 0928028101


Sudirman S. Panna, M.kom
NION : 0924038205

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 dari tim pembimbing.
3. Dalam karya tulis (Skripsi) saya ini tidak terdapat karya atau pendapat yang telah di publikasikan 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 Universitas Ichsan Gorontalo.

Gorontalo, Maret 2023

Yang Membuat Pernyataan

Frengki Daud

ABSTRACT

FRENGKI DAUD. T3119131. THE PREDICTION OF ON-TIME STUDENT GRADUATION USING THE DECISION-TREE METHOD

A student graduation prediction system can provide information related to the number of students who graduate on time or not on time. This system is significantly important to be used by the campus to facilitate obtaining student graduation data and providing accurate information. The student graduation prediction system must continue to be developed. The method used in this research is the Decision Tree method with a descriptive research type. The subjects of this research are informatics engineering students at Universitas Ichsan Gorontalo using several attributes including gender, IPS1, IPS2, IPS3, and IPS4. The sampling of this research employs literature and document study through observation of the research site. The results of this research data are presented in tabular form. The results are in the form of 178 students who graduated on time. Those who graduated not on time are 186 students, obtained from the test results using training data, as well as the accuracy value of the prediction results obtained, namely 90%. The value of Precision 93% and Recall 88% is gained from the prediction of test data using the Confusion Matrix model evaluation. Based on the results of the application and testing of this decision-tree method, future researchers are expected to get a better and more accurate accuracy value. They can also develop an application program that can be used later.

Keywords: *decision tree, prediction, student, accuracy*

ABSTRAK

FRENGKI DAUD. T3119131. PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN METODE DECISION TREE

Sistem prediksi kelulusan mahasiswa dapat memberikan informasi terkait jumlah mahasiswa yang lulus tepat waktu atau tidak tepat waktu. Sistem ini sangatlah penting digunakan oleh pihak kampus agar dapat memudahkan dalam mendapatkan data kelulusan mahasiswa serta dapat memberikan informasi yang akurat. Karena itu, sistem prediksi kelulusan mahasiswa ini harus terus dikembangkan. Metode penelitian yang digunakan dalam penelitian ini adalah dengan menggunakan metode *Decision Tree* dengan jenis penelitian deskriptif. Subjek penelitian ini adalah mahasiswa Teknik informatika universitas ichsan Gorontalo dengan menggunakan beberapa atribut diantaranya : jenis kelamin, IPS1, IPS2, IPS3 dan IPS4. Pengambilan sampel penelitian ini dengan cara studi kepustakaan dan studi dokumen melalui observasi ke lokasi penelitian. Adapun hasil dari data penelitian ini disajikan dalam bentuk tabel. Adapun hasilnya adalah berupa 178 mahasiswa yang lulus tepat waktu sedangkan untuk mahasiswa yang lulus tidak tepat waktu sebanyak 186 yang didapatkan dari hasil pengujian dengan menggunakan data training, serta nilai akurasi dari hasil prediksi yang didapatkan yakni sebesar 90%, nilai *Precision* 93% dan *Recall* 88% yang didapatkan dari hasil prediksi data uji dengan menggunakan evaluasi model *Confussion Matrix*. Dari hasil implementasi serta pengujian dari metode *decision tree* ini, untuk para peneliti selanjutnya diharapkan agar lebih mendapatkan nilai akurasi yang lebih baik dan lebih akurat serta dapat mengembangkan sebuah program aplikasi yang bisa digunakan nantinya.

Kata kunci: *decision tree*, prediksi, mahasiswa, akurasi

KATA PENGANTAR

Alhamdulillah, penulis dapat menyelesaikan skripsi ini dengan judul **“Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree (Studi Kasus : Program Studi Teknik Informatika Fakultas Ilmu Komputer Unisan Gorontalo)”,** untuk memenuhi salah satu syarat ujian guna untuk memperoleh gelar sarjana 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. Ibu Rezqiwati Ishak, M.Kom, selaku Pembimbing I yang telah banyak membimbing penulis selama ini;
8. Bapak Warid Yunus, M.Kom, selaku Pembimbing II yang telah banyak membimbing penulis selama ini;
9. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang telah mendidik dan mengajarkan berbagai disiplin ilmu kepada penulis;

10. Kepada Orang tua khususnya Ibu sebagai penyemangat dan yang telah memberikan doa, dukungan yang sangat-sangat besar kepada penulis;
11. Rekan-rekan seperjuangan yang telah banyak memberikan bantuan dan dukungan moril yang sangat besar kepada penulis;
12. Sahabat-sahabat yakni : Dival, Dzikri, Febri, Muin, Ferdi dan Rahmat, yang telah memberikan saya dukungan dan motivasi kepada penulis;
13. Kepada Musisi Tate Mcrae, Taylor Swift dan Coldplay, Neptune, Matthew Zeitler, Boy In Space sebagai inspirasi dan penyemangat penulis dalam Menyusun skripsi ini.;
14. Kepada Fearless Motivation, Really Slow Motion, Kim Dong Hyeok, Peder B. Helland, Ivan Torrent, dan Lee Nyum selaku *Music Score Composer* yang menjadi *moodbooster*/penyemangat penulis dalam menyusun skripsi ini;
15. Kepada semua pihak yang ikut membantu dan mendukung penulis dalam penyelesaian skripsi ini yang tak sempat penulis sebutkan satu persatu.

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, Maret 2023

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN SKRIPSI	ii
PENGESAHAN SKRIPSI.....	iii
PERNYATAAN SKRIPSI.....	iv
ABSTRACT	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
2.2 Tinjauan Pustaka	7
2.2.1 Mahasiswa.....	7
2.2.2 Data Mining	7
2.2.3 <i>Decision Tree</i>	9
2.2.4 <i>Python</i>	10
2.2.5 <i>Confussion Matrix</i>	10
2.2.6 Penerapan Metode <i>Decision Tree</i>	11
2.2.7 Perangkat Lunak.....	15
2.2.8 Kerangka Pikir	16

BAB III METODE PENELITIAN	17
3.1 Jenis, Metode, Subjek, Objek, Waktu, dan Lokasi Penelitian.....	17
3.2 Pengumpulan Data	17
3.3 Pemodelan	19
3.3.1 Pra Pengolahan Data	19
3.3.2 Validasi Data.....	19
3.3.3 Pengembangan Model.....	20
3.3.4 Evaluasi Model.....	20
BAB IV HASIL PENELITIAN.....	21
4.1 Hasil Pengumpulan Data	21
4.2 Hasil Pemodelan.....	22
4.3 Hasil Proses Pemodelan	23
4.3.1 Tahapan Seleksi	24
4.3.2 Tahapan Preprocessing.....	25
4.3.3 Tahap Transformation.....	28
4.3.4 Tahap Data Mining	29
4.3.5 Tahap Interpretation / evaluation	34
BAB V PEMBAHASAN PENELITIAN	42
5.1 Pembahasan Kinerja Model.....	42
5.2 Pembahasan Model.....	44
5.2.1 Proses Pemodelan Pohon Keputusan dari Hasil Prediksi	44
5.2.2 Proses Prediksi dengan Model Pohon Keputusan.....	44
5.2.3 Proses Pembuatan Path Storage File Hasil Pohon Keputusan	44
5.2.4 Proses Hasil Keluaran Pohon Keputusan	45
BAB VI KESIMPULAN DAN SARAN.....	47
6.1 Kesimpulan.....	47
6.2 Saran	47
DAFTAR PUSTAKA	48
LAMPIRAN	

DAFTAR GAMBAR

Gambar 2.1 Diagram hubungan data mining.....	8
Gambar 2.2 Proses <i>Knowledge Discoveryin Database</i>	8
Gambar 2.3 Node Keputusan Akhir	13
Gambar 2.4 Rule berdasarkan role akhir	14
Gambar 2.5 Kerangka Pikir	16
Gambar 4.1 Grafik Waktu Kelulusan Mahasiswa	28
Gambar 4.2 Visualisasi Pohon Keputusan	33
Gambar 4.3 Grafik Persetase Atribut	34
Gambar 4.4 Hasil Prediksi Dengan Model <i>Confusion Matrix</i>	35
Gambar 4.5 Grafik Waktu Kelulusan Mahasiswa	37
Gambar 4.6 Visualisasi Pohon Keputusan	40
Gambar 4.7 Grafik Persentase Atribut	41
Gambar 5.1 Evaluasi dengan <i>Confusion Matrix</i>	42
Gambar 5.2 Pohon Keputusan	47

DAFTAR TABEL

Tabel 1.1 Data Mahasiswa Baru Prodi Teknik Informatika.....	2
Tabel 2.1 Penelitian Terkait dengan <i>Decision Tree</i>	5
Tabel 2.2 Kelompok Dataset	11
Tabel 2.3 Atribut klasifikasi mahasiswa drop out	12
Tabel 2.4 Hasil perhitungan pada Dataset.....	12
Tabel 2.5 Perangkat Lunak Pendukung.....	15
Tabel 3.1 Variabel Data.....	18
Tabel 3.2 Data Mahasiswa Tahun 2018	18
Tabel 4.1 Hasil Pengumpulan Data Mahasiswa Prodi Teknik Informatika	21
Tabel 4.2 Data Training Final	23
Tabel 4.3 Dataset Mahasiswa Hasil Olahan	24
Tabel 4.4 Dataset Training	30
Tabel 4.5 Dataset Testing	31
Tabel 4.6 Entropi untuk tiap-tiap atribut	32
Tabel 5.1 Entropi Total	46
Tabel 5.2 Role Akhir	47
Tabel Jadwal Penelitian.....	50

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perguruan tinggi adalah ukuran sejauh mana pendidikan yang harus ditempuh setiap orang. untuk memperoleh gelar sarjana sesuai dengan bidang ilmu yang diminati. Perguruan Tinggi yang baik adalah perguruan tinggi penerbit sekolah pendidikan untuk mahasiswa. Mahasiswa sering disebut sebagai organisasi jaringan kecerdasan lebih luas dari dalam organisasi [1]. Data kelulusan mahasiswa dapat memberikan informasi yang bermanfaat bagi universitas jika data tersebut digunakan dengan maksimal.

Universitas Ichsan Gorontalo merupakan salah satu kampus swasta di Gorontalo yang banyak diminati oleh banyak orang. Kampus ini memiliki enam fakultas diantaranya : Fakultas Ilmu Komputer, Fakultas Ekonomi, Fakultas Hukum, Fakultas Pertanian, Fakultas Ilmu Sosial dan Ilmu Politik dan Fakultas Teknik.

Fakultas Ilmu Komputer merupakan salah satu fakultas yang ada di Universitas Ichsan Gorontalo. Fakultas Ilmu Komputer Ini menawarkan tiga program studi, yaitu sebagai berikut: Program Studi Teknik Informatika, Desain Komunikasi Visual (DKV) dan Sistem Informasi. Setiap tahunnya fakultas ini mendapatkan mahasiswa baru cukup banyak, salah satu penerima mahasiswa paling banyak adalah program studi Teknik Informatika. Akan tetapi, jumlah mahasiswa yang lulus tepat waktu tidak sebanding dengan jumlah mahasiswa yang masuk.

Dari jumlah mahasiswa baru yang masuk yang dapat menyelesaikan pendidikannya dalam waktu yang tepat dalam setiap kurang dari 20% di setiap masa kelulusan. Hal ini tentunya berdampak pada evaluasi akreditasi, oleh karena itu diperlukan upaya dini untuk mengetahui faktor-faktor apa saja yang mempengaruhi kemampuan seorang mahasiswa untuk menyelesaikan studinya tepat waktu.

Tabel 1.1 Data Mahasiswa Baru Prodi Teknik Informatika

Tahun	Jumlah Mahasiswa yang Masuk	Jumlah Mahasiswa yang Lulus Tepat Waktu
2012	297	55
2013	280	59
2014	307	56
2015	227	51
2016	194	28
2017	194	12
2018	167	5

Berdasarkan pada tabel diatas bisa dilihat bahwa adanya penurunan jumlah mahasiswa yang lulus tepat waktu, dimana setiap tahunnya jumlah mahasiswa yang lulus tepat waktu tidak sampai setengah dari jumlah mahasiswa yang masuk. Beberapa alasan berkontribusi untuk ini, termasuk: faktor ekonomi, kuliah sambil kerja, dan cuti yang diajukan oleh mahasiswa itu sendiri. Selain itu adanya kurang kesadaran atau keseriusan dari mahasiswa dalam mengikuti proses perkuliahan yang mengakibatkan mahasiswa tersebut menjadi malas, sering bolos, tidak mengerjakan tugas sehingga mahasiswa tersebut harus mengulang mata kuliah di semester berikutnya. Diperlukan strategi untuk mengatasi hal tersebut yang memungkinkan pihak program studi teknik informatika untuk memprediksi jumlah kelulusan tepat waktu mahasiswa sehingga pihak program studi teknik informatika dapat mengetahui seberapa banyak jumlah mahasiswa yang akan lulus tepat waktu.

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Untuk mengukur nilai akurasi dapat digambarkan dengan menghitung nilai galat (*error*) dan persen galat. Semakin rendah nilai galat dan persen galat yang diperoleh maka akurasi semakin membaik [2].

Prediksi kelulusan mahasiswa ini dilakukan karena cara ini merupakan salah satu cara yang bisa dilakukan oleh pihak jurusan tentang kiat-kiat yang diambil guna untuk menjadikan lulusan terhadap mahasiswa bisa tepat waktu. Kiat-kiat tersebut bisa berupa memberikan bimbingan khusus kepada mahasiswa tingkat akhir atau kegiatan khusus lainnya yang bisa dilakukan oleh pihak jurusan agar mahasiswa dapat menyelesaikan studi tepat waktu. Yang menjadi alasan peneliti mengangkat topik penelitian tentang prediksi karena, peneliti melihat masih banyaknya jumlah mahasiswa yang tidak dapat menyelesaikan studinya dengan tepat waktu serta tidak sebanding dengan jumlah mahasiswa yang masuk

Berdasarkan penelitian yang dilakukan oleh Ade Yuliana¹, Duwi Bayu Pratomo² tentang “Algoritma *Decision Tree* (C4.5) Untuk Memprediksi Kepuasan Mahasiswa Terhadap Kinerja Dosen Politeknik TEDC Bandung” Berdasarkan validasi silang, uji-t, dan evaluasi f-measure, akurasinya 94,62% dan f-measure 96,99% [3].

Berdasarkan hasil dari penelitian terkait di atas, maka penggunaan metode *Decision Tree* ini sangat cocok digunakan pada penelitian ini dikarenakan algoritma ini bisa memberikan nilai akurasi yang baik untuk melakukan proses prediksi kelulusan tepat waktu mahasiswa.

Adanya permasalahan diatas, maka peneliti tertarik untuk melakukan penelitian yang diberi judul **“Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Decision Tree* (Studi Kasus : Program Studi Teknik Informatika Fakultas Ilmu Komputer Unisan Gorontalo)”**.

1.2 Identifikasi Masalah

Latar belakang masalah informasi adalah Program Studi Teknik Informatika Universitas Ichsan Gorontalo, dimana lebih sedikit mahasiswa yang selesai tepat waktu daripada yang mendaftar.

1.3 Rumusan Masalah

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

1. Bagaimana hasil penerapan Metode *Decision Tree* untuk memprediksi kelulusan tepat waktu mahasiswa ?.
2. Seberapa tinggi tingkat akurasi pada hasil prediksi kelulusan tepat waktu mahasiswa menggunakan metode *Decision Tree* ?.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah :

1. Mengetahui apakah penerapan daripada Metode *Decision Tree* ini berhasil untuk memprediksi kelulusan terhadap mahasiswa
2. Mengetahui apakah terjadi peningkatan terhadap lulusan tepat waktu mahasiswa daripada penelitian sebelumnya.

1.5 Manfaat Penelitian

Adapun manfaat yang didapatkan dari penelitian ini antara lain :

1. Dapat memberikan informasi tentang hasil prediksi kelulusan mahasiswa tepat waktu maupun yang tidak tepat waktu.
2. Agar bisa membantu sistem kerja yang ada pada prodi Teknik Informatika untuk mengevaluasi kelulusan tepat waktu mahasiswa.

1.5.1 Manfaat Teoritis

Kontribusi terhadap terobosan ilmiah dan teknologi, khususnya dalam ilmu komputer, membantu memperbarui kemampuan *Decision Tree* dalam memperkirakan kelulusan tepat waktu mahasiswa.

1.5.2 Manfaat Praktis

Kontribusi ide, upaya, perhatian, atau solusi bagi software developer guna mendukung proses prediksi dalam rangka di hasilkanya kualitas software yang baik dan memberikan dampak juga pada peningkatan kualitas perguruan tinggi.

BAB II

LANDASAN TEORI

2.1 Tinjauan Studi

Berikut merupakan daftar penelitian terdahulu yang berkaitan dengan judul dan metode yang digunakan :

Tabel 2.1 Penelitian Terkait dengan Decision Tree

No	PENELITI	JUDUL	TAHUN	METODE	HASIL
1	Rahmi Ramadhani ¹ , Heka Hendriyani ² [4].	Prediksi Prestasi Siswa Berbasis Data Mining Menggunakan Algoritma <i>Decision Tree</i> (Studi Kasus : SMKN 2 Padang)	2021	Menggunakan Metode <i>Decision Tree</i>	Penelitian ini menggunakan Algoritma <i>Decision Tree</i> untuk memprediksi data prestasi siswa dengan menggunakan metode <i>Decision Tree</i> , dalam penelitian ini telah didapatkan nilai presisi sebesar 90,31%. Yang didapatkan dari hasil analisis menggunakan metode pohon keputusan J48 serta Aplikasi WEKA.

No	PENELITI	JUDUL	TAHUN	METODE	HASIL
2	Hermanto Wahano, Dwiza Riana [5].	Prediksi Calon Pendonor Darah Potensial Dengan Algoritma <i>Naïve Bayes</i> , <i>K-Nearest Neighbors</i> dan <i>Decision Tree C4.5</i>	2020	Menggunakan Metode <i>Decision Tree</i> <i>Naïve Bayes</i> dan <i>K-Nearest Neighbors</i>	Pada penelitian ini, penggunaan algoritma <i>Decision Tree</i> <i>Naïve Bayes</i> dan <i>K-Nearest Neighbors</i> menghasilkan nilai akurasi tertinggi yaitu algoritma <i>Decision Tree C4.5</i> dengan nilai akurasi 93,83% lebih baik dari <i>Naïve Bayes</i> dengan nilai akurasi 85,15% dan <i>K-Nearest Neighbors</i> dengan nilai akurasi 84,10%.
3	Nurdiana Handayani, Herry Wahyono, Joko Trianto, Dwi Sidik Permana[6].	Prediksi Tingkat Resiko Kredit dengan Data Mining Menggunakan Algoritma <i>Decision Tree C.45</i>	2021	Menggunakan Metode <i>Decision Tree C.45</i>	Dari hasil evaluasi menggunakan <i>Confusion Matrix</i> didapatkan akurasi yang dihasilkan untuk 1.153 data training dengan 91 data testing dan enam atribut yang digunakan menghasilkan akurasi sebesar 79%. yang dimana hasil tersebut dapat dipengaruhi oleh beberapa faktor diantaranya : jumlah data training, data testing dan atribut yang digunakan.

2.2 Tinjauan Pustaka

2.2.1 Mahasiswa

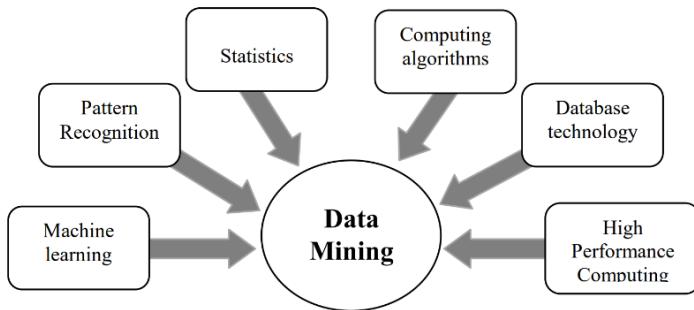
Mahasiswa adalah peserta didik yang belajar di tingkat perguruan tinggi, menurut Kamus Besar Bahasa Indonesia (KBBI) (Depdiknas, 2012). Mahasiswa memegang peranan penting dalam pencapaian tujuan pembangunan nasional, sedangkan perguruan tinggi adalah lembaga pendidikan yang secara eksplisit dibebankan tugas dan kewajiban untuk membina mahasiswa sesuai dengan tujuan pendidikan tinggi. Perguruan tinggi dapat mencapai tujuannya apabila Tridharma Perguruan Tinggi diterapkan, yaitu kemampuan menyelenggarakan pendidikan, melakukan penelitian, dan pengabdian kepada masyarakat (Undang-Undang Republik Indonesia Nomor 12 Tahun 2012) [7].

Mahasiswa adalah orang yang senantiasa belajar dan mengikuti mata kuliahnya, dimana dalam mengikuti serangkaian perkuliahan sangat dipengaruhi oleh bakat mahasiswa, karena sebagian mahasiswa telah bekerja atau aktif dalam kegiatan organisasi kemahasiswaan. (Ganda, 2004) [8]. Syarat-syarat mahasiswa agar lulus tepat waktu antara lain : 1) IPK minimal 3.00 dan $IPK > 3.75$ dengan predikat cumlaude, 2) SKS yang diperoleh harus sebanyak 144 SKS, 3) Tidak mendapatkan nilai error untuk mata kuliah serta, 4) Mengikuti seluruh aktifitas dan kegiatan perkuliahan.

2.2.2 Data Mining

Data mining telah ada sejak tahun 1990-an, ketika pekerjaan berbasis data menjadi lebih lazim di bidang-bidang seperti pemasaran dan bisnis, penelitian dan teknik, serta seni dan hiburan. Data mining, menurut beberapa ahli, adalah suatu tahapan dalam pemeriksaan pengetahuan dalam database, juga dikenal sebagai *Knowledge Discovery in Database* (KDD). Teknik mengumpulkan data yang relevan dan memahami pola dari kumpulan data yang sangat besar dikenal sebagai data mining. Sumber data termasuk database, gudang data, web, repositori, dan input data ke dalam

sistem dinamis (Han, 2006). Data mining juga terkait dengan statistik, pembelajaran mesin, pengenalan pola, algoritma komputasi, teknologi basis data, dan komputasi kinerja tinggi [9].



Gambar 2.1 Diagram hubungan data mining

Secara sistematis, langkah utama untuk melakukan data mining terdiri dari tiga tahap, yaitu sebagai berikut (Gonunescu, 2011);

1. Eksplorasi atau pemrosesan awal data

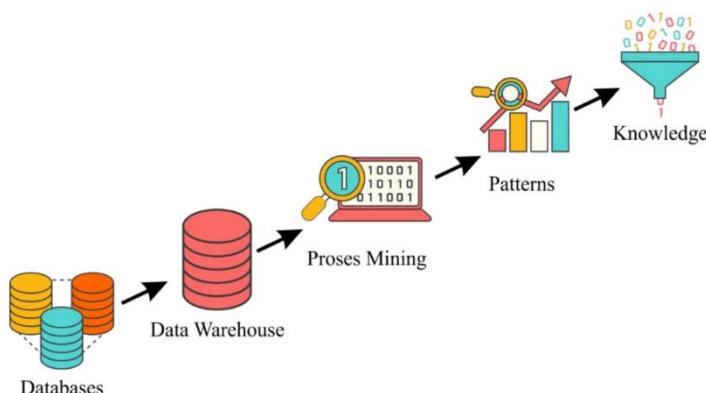
Eksplorasi data atau pemrosesan awal meliputi pembersihan data, normalisasi, transformasi, manajemen nilai yang hilang, pengurangan dimensi, pemilihan subset fitur, dan sebagainya.

2. Membangun model dan validasi

Konstruksi model dan validasi memerlukan analisis banyak model dan memilih model yang menawarkan hasil terbaik. Metode seperti klasifikasi, regresi, analisis cluster, dan asosiasi digunakan dalam pembuatan model.

3. Penerapan

Model yang dipilih kemudian diterapkan pada data baru untuk memberikan kinerja yang baik pada masalah yang diteliti.



Gambar 2.2 Proses *Knowledge Discovery in Database* (KDD)

2.2.3 Decision Tree

Algoritma C4.5 dan pohon keputusan adalah model yang tidak dapat dipisahkan karena metode C4.5 diperlukan untuk menghasilkan pohon keputusan, Pada akhir 1970-an dan awal 1980-an, J. Ross Quinlan, seorang peneliti pembelajaran mesin, membuat model pohon keputusan bernama ID3 (Iterative Dichotomiser), meskipun proyek ini sebelumnya telah dibangun oleh E.B Hunt, J. Marin, dan P.T Stone, Quinlan membangun sebuah metode berdasarkan pembelajaran terawasi yang disebut C4.5 dari pengembangan ID3 [10].

Serangkaian peningkatan ke ID3 memuncak dalam pembuatan C4.5, sistem pohon keputusan yang praktis dan berpengaruh. Metode untuk menangani properti numerik, nilai yang hilang, data yang bising, dan aturan yang menghasilkan aturan dari pohon adalah beberapa peningkatan [10].

Pendekatan pohon keputusan (C4.5) memiliki tahapan sebagai berikut:

- 1.) Mengumpulkan data pelatihan.
- 2.) Menentukan akar pohon.
- 3.) Menentukan gain. Untuk menjadikan suatu atribut sebagai akar berdasarkan atribut yang ada dengan nilai Gain tertinggi, Persamaan 1 dapat digunakan untuk menghitung Gain [10] :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|Si|}{|S|} * Entropy(Si) \dots\dots\dots(1)$$

Keterangan :

1. A = Variabel
 2. v = Nilai yang mungkin untuk variable A
 3. $|S_i|$ = Jumlah sampel untuk nilai i
 4. $|S|$ = Jumlah sampel untuk seluruh sampel data
 5. $Entropy(S_i)$ = Entropy untuk sampel yang memiliki nilai i

Ulangi langkah kedua sampai setiap cabang telah selesai. Sementara itu, kita dapat menggunakan persamaan 2 untuk menghitung nilai Entropi sebagai berikut:

$$\text{Entropy}(S) = \sum_{i=1}^n -pi * \log_2 pi \dots \dots \dots (2)$$

Keterangan :

1. S = Himpunan kasus
2. n = Jumlah partisi S
3. pi = Proposisi S_i terhadap S

2.2.4 Python

Python adalah bahasa pemrograman populer yang sangat cocok untuk pengembangan Linux. Python bukan hanya bahasa pemrograman yang kompeten; Python juga ada dalam terobosan teknis lain yang harus kita pelajari. Python adalah bahasa pemrograman yang banyak digunakan di banyak sistem operasi. Guido van Rossum dari Amsterdam, Belanda, mendesain Python, sebuah bahasa pemrograman populer yang mendukung paradigma pemrograman berorientasi objek atau scripting [11].

2.2.5 Confusion Matrix

Confusion matrix menurut Han dan Kamber (2011) dapat diartikan sebagai suatu alat yang memiliki fungsi untuk melakukan analisis apakah classifier tersebut baik dalam mengenali tuple dari kelas yang berbeda. Nilai dari *True-Positive* dan *True-Negative* memberikan informasi ketika *classifier* dalam melakukan klasifikasi data bernilai benar, sedangkan *False-Positive* dan *False-Negative* memberikan informasi ketika classifier salah dalam melakukan klasifikasi data [12].

		Predicted class		Total
		yes	no	
Actual class	yes	TP	FN	P
	no	FP	TN	N
	Total	P'	N'	P + N

2.2.6 Penerapan Metode *Decision Tree*

Berikut merupakan penerapan metode *Decision Tree* dalam penyelesaian berdasarkan rumus yang digunakan untuk melakukan proses Prediksi Kelulusan Tepat Waktu Mahasiswa berdasarkan studi yang dilakukan oleh (Amaul Husna Nasrullah : 2018) [13].

Tabel 2.2 Kelompok Dataset

Kelas Target	Data yang digunakan			
	Data Training 50%	Data Validasi 25%	Data Testing 25%	Jumlah Data
Non DO	120	50	25	220
DO	100	50	50	200
Jumlah	220	100	100	420

Data juga dihasilkan dengan memilih dan memodifikasi pengumpulan data (lampiran a). Tujuan dari pemilihan data adalah untuk menemukan variabel yang relevan untuk penyelidikan ini. Dan transformasi data yang digunakan untuk merubah kumpulan data agar konten informasi terbaik dapat diekstraksi dan dimasukkan ke dalam alat penambangan dengan cara yang benar [13].

Dengan menggunakan cara kerja pendekatan C4.5 mulai dari data training sampai dengan dihasilkannya entropi dan mengembangkan aturan-aturan yang dimanfaatkan sebagai pola untuk memilih siswa yang memiliki kecenderungan putus sekolah merupakan tahap selanjutnya dari penelitian ini. Atribut adalah komponen data yang mencerminkan sifat atau kualitas objek data. Dalam literatur, dimensi, fitur, atribut dan variabel sering bergantian untuk digunakan. Karakteristik yang dipertimbangkan pada penelitian ini didasarkan pada sumber penelitian terdahulu [13].

Tabel 2.3 Klasifikasi Atribut drop out mahasiswa [13].

No.	Atribut	Nilai Atribut
1	Jenis Kelamin	Laki-laki Perempuan
2	Umur	<20 Tahun 20 s/d 25 Tahun >25 Tahun
3	Agama	Muslim Non Muslim
4	Asal Daerah	Dalam Kota Gorontalo Luar Kota Gorontalo Luar Provinsi
5	Kelas	Kelas Reguler Kelas Karyawan
6	Nilai IPS Semester I	<2,75 2,75 - 3,00 >3,00
7	Nilai IPS Semester II	<2,75 2,75 - 3,00 >3,00
8	Nilai IPS Semester III	<2,75 2,75 - 3,00 >3,00
9	Nilai IPS Semester IV	<2,75 2,75 - 3,00 >3,00

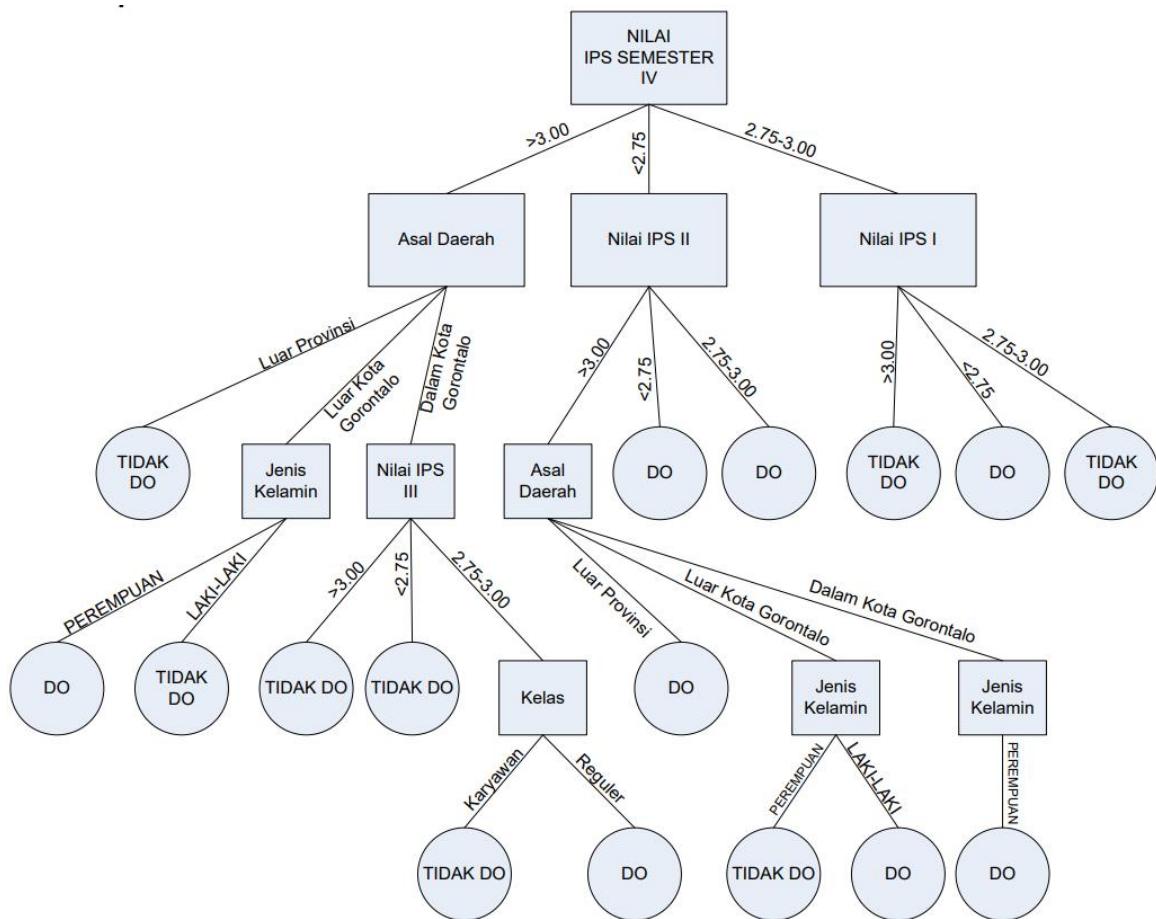
Atribut-atribut yang dipilih, serta nilai gain terbesar dari setiap atribut yang akan menjadi root awal, akan digunakan untuk menentukan Root (Node). Hitung nilai entropy terlebih dahulu, kemudian nilai gain atribut. Nilai entropy yang diperoleh dari persamaan 1 adalah sebagai berikut [13] :

$$Entropi(S) = \left(-\left(\frac{41}{98}\right)x \log_2 \left(\frac{41}{98}\right) \right) + \left(-\left(\frac{57}{98}\right)x \log_2 \left(\frac{57}{98}\right) \right) = 0.9806857$$

Tabel 2.4 Perhitungan hasil pada Dataset [13].

Total Kasus	Sum(DO)	Sum(TIDAK DO)	Entropi Total
98	41	57	0.9806857

Setelah menghitung nilai entropi untuk setiap contoh, setiap atribut dan nilainya diperiksa. Nilai gain setiap atribut dihitung menggunakan persamaan (2). Setelah itu, pohon keputusan dibangun berdasarkan node yang dipilih, menghasilkan node keputusan akhir seperti yang ditunjukkan di bawah ini ;



Gambar 2.3 Node keputusan akhir [13].

Keputusan Simpul akhir, dapat diamati bahwa atribut yang merujuk pada agama sudah tidak digunakan lagi sebagai simpul karena nilai gain yang didapatkan untuk atribut agama dalam perhitungan nilai gain simpul kedua adalah -4.53. 200 dari dataset dan 98 data yang telah di cleaning.

Selanjutnya, seperti yang ditunjukkan pada tabel di bawah ini, hasil pohon keputusan diekstraksi menjadi aturan-aturan pada tabel berikut :

Id	Rule
1	JIKA Nilai IPS IV = 2.75 - 3.00 DAN Nilai IPS I = 2.75 - 3.00 MAKA Klasifikasi =TIDAK DO
2	JIKA Nilai IPS IV = 2.75 - 3.00 DAN Nilai IPS I = <2.75 MAKA Klasifikasi =DO
3	JIKA Nilai IPS IV = 2.75 - 3.00 DAN Nilai IPS I = >3.00 MAKA Klasifikasi =TIDAK DO
4	JIKA Nilai IPS IV = <2.75 DAN Nilai IPS II = 2.75 - 3.00 MAKA Klasifikasi =DO
5	JIKA Nilai IPS IV = <2.75 DAN Nilai IPS II = <2.75 MAKA Klasifikasi =DO
6	JIKA Nilai IPS IV = >3.00 DAN Asal Daerah = Luar Provinsi MAKA Klasifikasi =TIDAK DO
7	JIKA Nilai IPS IV = <2.75 DAN Nilai IPS II = >3.00 DAN Asal Daerah = Luar Provinsi MAKA Klasifikasi = DO
8	JIKA Nilai IPS IV = >3.00 DAN Asal Daerah = Dalam Kota Gorontalo DAN Nilai IPS III = <2.75 MAKA Klasifikasi = TIDAK DO
9	JIKA Nilai IPS IV = >3.00 DAN Asal Daerah = Dalam Kota Gorontalo DAN Nilai IPS III = >3.00 MAKA Klasifikasi = TIDAK DO
10	JIKA Nilai IPS IV = >3.00 DAN Asal Daerah = Luar Kota Gorontalo DAN Jenis Kelamin = Laki-Laki MAKA Klasifikasi = TIDAK DO
11	JIKA Nilai IPS IV = >3.00 DAN Asal Daerah = Luar Kota Gorontalo DAN Jenis Kelamin = Perempuan MAKA Klasifikasi = DO
12	JIKA Nilai IPS IV = <2.75 DAN Nilai IPS II = >3.00 DAN Asal Daerah = Dalam Kota Gorontalo DAN Jenis Kelamin = Laki-Laki MAKA Klasifikasi = TIDAK DO
13	JIKA Nilai IPS IV = <2.75 DAN Nilai IPS II = >3.00 DAN Asal Daerah = Dalam Kota Gorontalo DAN Jenis Kelamin = Perempuan MAKA Klasifikasi = DO
14	JIKA Nilai IPS IV = <2.75 DAN Nilai IPS II = >3.00 DAN Asal Daerah = Luar Kota Gorontalo DAN Jenis Kelamin = Laki-Laki MAKA Klasifikasi = DO
15	JIKA Nilai IPS IV = <2.75 DAN Nilai IPS II = >3.00 DAN Asal Daerah = Luar Kota Gorontalo DAN Jenis Kelamin = Perempuan MAKA Klasifikasi = TIDAK DO
16	JIKA Nilai IPS IV = >3.00 DAN Asal Daerah = Dalam Kota Gorontalo DAN Nilai IPS III = 2.75 - 3.00 DAN Kelas = Karyawan MAKA Klasifikasi = TIDAK DO
17	JIKA Nilai IPS IV = >3.00 DAN Asal Daerah = Dalam Kota Gorontalo DAN Nilai IPS III = 2.75 - 3.00 DAN Kelas = Reguler MAKA Klasifikasi = DO

Gambar 2.4 Rule berdasarkan role akhir [13].

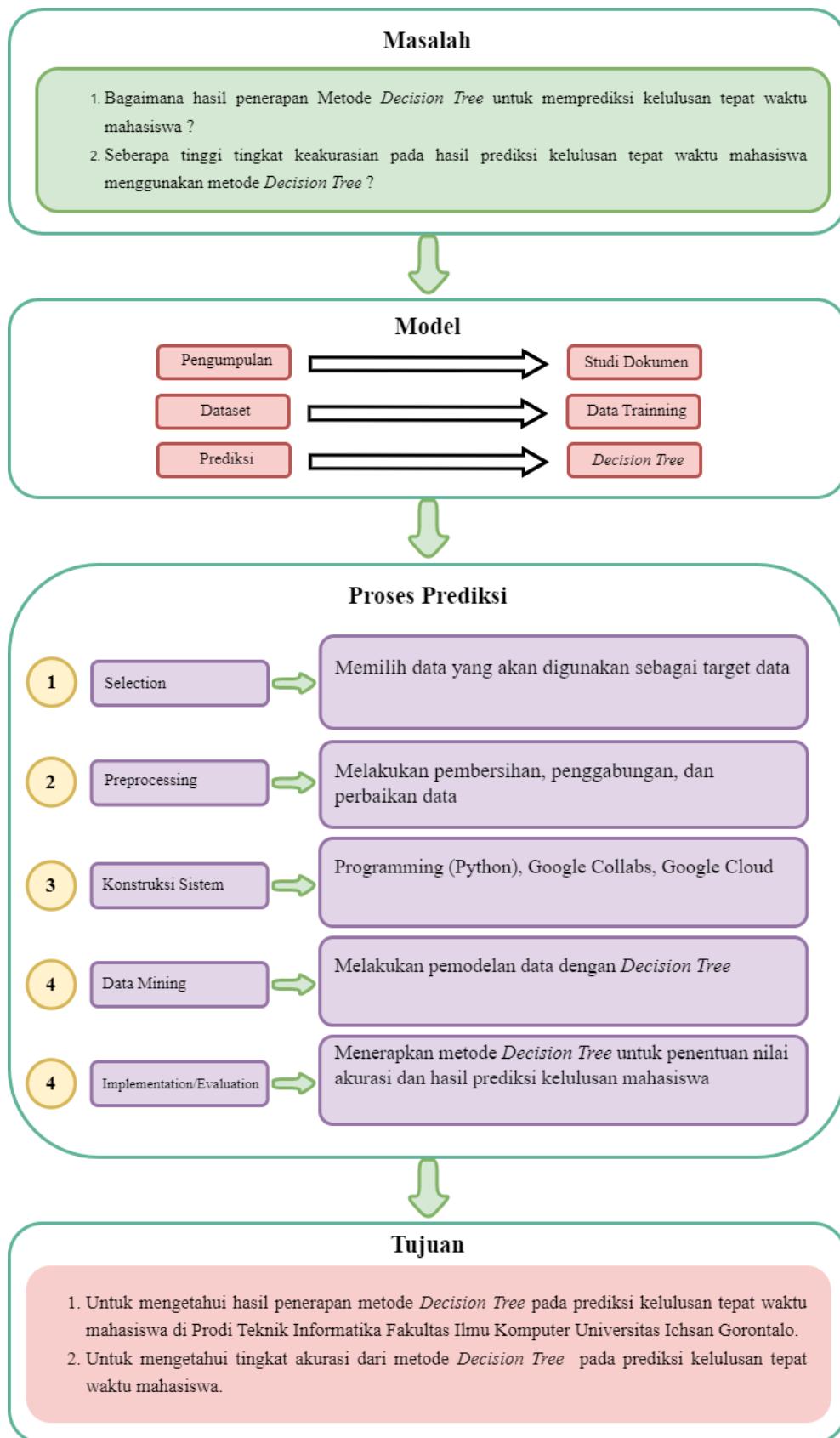
Tujuh belas (17) pedoman dikembangkan dalam mengkategorikan siswa yang berisiko putus sekolah. Selanjutnya 17 kelas (aturan) tersebut dapat digunakan sebagai kelas/kelompok untuk mengidentifikasi siswa yang mungkin putus sekolah sebelum terdaftar sebagai mahasiswa di FEKON UNISAN Gorontalo. Karena hanya mempertimbangkan siswa yang di Drop Out dan juga yang tidak Drop Out, maka akurasi beserta tingkat kesalahan tidak dihitung.

2.2.7 Perangkat Lunak

Tabel 2.5 Perangkat Lunak Pendukung.

No.	Perangkat Lunak Pendukung	Berfungsi
1	Python	Bahasa pemrograman tingkat tinggi, penulisan kode/sintaks lebih sederhana, bersifat <i>open-source</i> , <i>cross-platform</i> dan cocok digunakan untuk <i>Data Scientist</i> serta didukung banyak <i>library</i> seperti : <i>Numpy</i> , <i>SciPy</i> , <i>Pandas</i> , <i>Scikit-Learn</i> , dan <i>Matplotlib</i>
2	Google Colaboratory (Google Colab)	<i>Integrated Development Environments (IDE)</i> berbasis Web yang akan digunakan untuk menuliskan koding Bahasa Python

2.2.8 Kerangka Pikir



Gambar 2.5¹⁶Kerangka Pikir

BAB III

METODE PENELITIAN

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

Penelitian ini merupakan penelitian terapan berdasarkan tingkat penerapannya. Sedangkan penelitian kuantitatif digunakan untuk mengolah jenis informasi dalam penelitian ini. Seperti yang dapat diamati dari penanganan data, penelitian ini bersifat konfirmatori.

Penelitian ini memiliki studi kasus di Program Studi Teknik Informatika Fakultas Ilmu Komputer Unisan Gorontalo. Sedangkan topik penelitian ini adalah prediksi kelulusan mahasiswa tepat waktu pada Program Studi Teknik Informatika Fakultas Ilmu Komputer Unisan Gorontalo dengan menggunakan pendekatan Decision Tree. Penelitian ini akan berlangsung di Universitas Ichsan Gorontalo yang beralamatkan di jl. Drs. Achmad Nadjamuddin, Limba U Dua, Kota Selatan Kota Gorontalo dari bulan September sampai Februari 2022.

3.2 Pengumpulan Data

Untuk melakukan pengumpulan data, data yang digunakan adalah :

1. Penelitian Data Primer (Lapangan)

Untuk mendapatkan data primer yang diperlukan, yaitu data langsung dari objek penelitian, yang berlokasi di Program Studi Teknik Informatika Fakultas Ilmu Komputer Unisan Gorontalo. Sebagai hasilnya digunakan pendekatan :

- a. Observasi : Pendekatan ini memungkinkan analisis sistem secara langsung. Studi ini dilakukan dengan mengumpulkan data tentang jumlah siswa yang masuk dan lulus tepat waktu antara tahun 2012 dan 2018. Tabel 3.1 menunjukkan variabel dan jenis data terkait.

Tabel 3.1 Variabel Data

No.	Name	Type	Keterangan
1	Jenis Kelamin	Tuple	Variabel Input
2	IPS_1	Float	Variabel Input
3	IPS_2	Float	Variabel Input
4	IPS_3	Float	Variabel Input
5	IPS_4	Float	Variabel Input
6	Keterangan Lulus	Varchar	Output

Tabel 3.2 Data Mahasiswa Tahun 2018

Jumlah Mhs Masuk	Tahun	Jenis Kelamin	IPS I	IPS II	IPS III	IPS IV	Keterangan Lulus Tepat Waktu
1	2018	L	2.14	2.09	0.17	3	Tidak
2	2018	L	3.23	3.57	3.54	3.22	Ya
3	2018	P	3.87	3.08	3.87	3.63	Ya
4	2018	P	3.55	3.87	2.88	3.85	Ya
5	2018	L	4	3.3	3.75	4	Ya
6	2018	P	3.45	3.61	3.71	3.87	Ya
7	2018	P	3.36	3.09	3.05	3.29	Tidak
...
167	2018	L	0	0	0	1.75	Tidak
Jumlah Mhs Lulus Tepat Waktu							5

2. Penelitian Data Sekunder (Kepustakaan)

Teknik kepustakaan diperlukan untuk mengumpulkan data sekunder untuk melengkapi data primer. Data sekunder ini berasal dari penelitian kepustakaan yang memiliki dasar-dasar teori. Teknik kepustakaan digunakan untuk menganalisis sistem dengan menggunakan contoh-contoh dari bahan penelitian. Selanjutnya, analisis sistem mencari faktor-faktor berupa catatan, buku, majalah, dan sebagainya yang berhubungan dengan penelitian.

3.3 Pemodelan

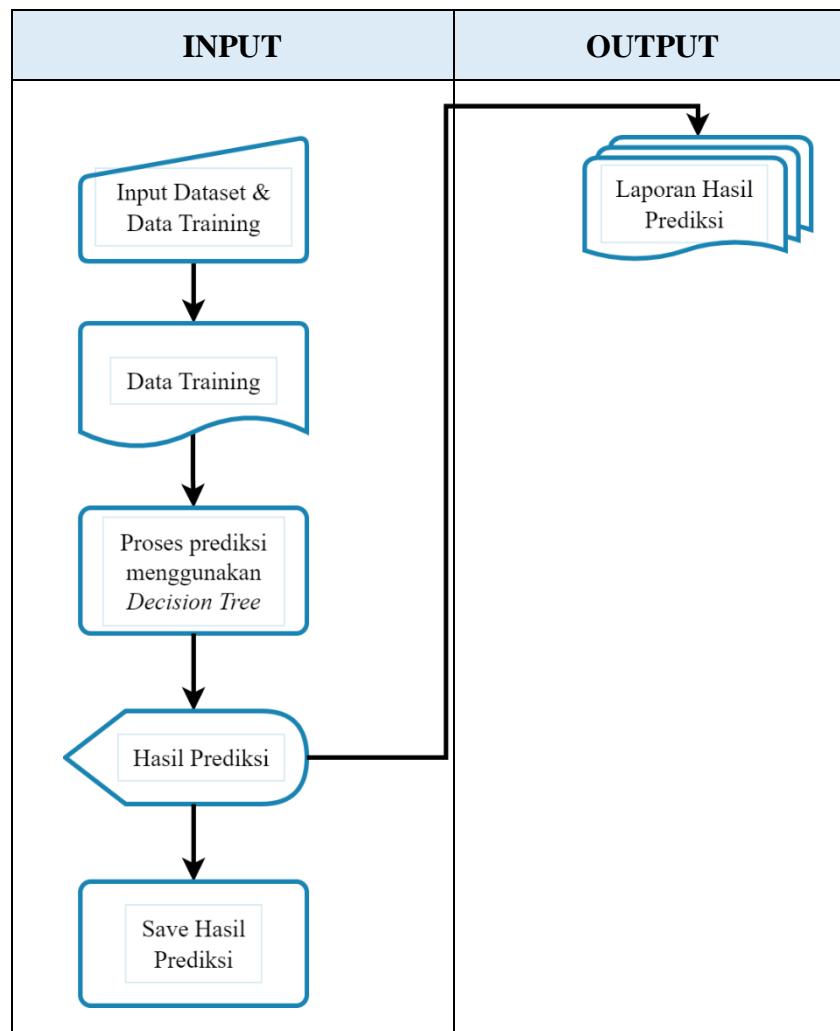
3.3.1 Pra Pengolahan Data

Seleksi dan transformasi data dilakukan pada data yang akan diolah nantinya. Hal ini dilakukan untuk menentukan kriteria atribut, yang nantinya akan menjadi nilai ukur dari setiap atribut untuk proses kategorisasi data.

3.3.2 Validasi Data

Validasi membagi data awal menjadi dua bagian: data training dan data testing. Data training adalah informasi yang akan diproses oleh algoritma klasifikasi, sedangkan data testing akan digunakan dalam tahap pengujian oleh program komputer. Pada level ini, pendekatan validasi yang digunakan adalah split validation, yaitu membagi data menjadi dua bagian berdasarkan jumlah data.

3.3.3 Pengembangan Model



Gambar 3.1 Sistem yang diusulkan

3.3.4 Evaluasi Model

Akurasi, presisi, dan recall dari model yang dihasilkan kemudian ditentukan menggunakan Python dan Confusion Matrix.

BAB IV

HASIL PENELITIAN

4.1 Hasil Pengumpulan Data

Pada tahap ini, peneliti mengumpulkan data yang akan digunakan pada penelitian ini. Untuk mendapatkan data, peneliti sebelumnya melakukan observasi dengan cara datang langsung ke lokasi penelitian yakni di prodi Teknik Informatika Unisan Gorontalo dan melakukan studi dokumen untuk mendapatkan data tersebut. Setelah melakukan observasi, peneliti akhirnya mendapatkan data mahasiswa prodi Teknik Informatika dari tahun 2012 s/d 2020 yang di sajikan dalam Tabel 4.1 Hasil Pengumpulan Data Mahasiswa Prodi Teknik Informatika dibawah ini :

Tabel 4.1 Hasil Pengumpulan Data Mahasiswa Prodi Teknik Informatika

No.	NIM	Nama	Tempat Lahir	Jenis Kelamin	Tanggal Lahir	Prodi
1	T3112012	Ferdy Yanto Ibrahim	Gorontalo	L	8/20/1992	S1 Teknik Informatika
2	T3112015	Adelia Kasim	Limboto	P	8/3/1992	S1 Teknik Informatika
3	T3112027	Adryan Alamri	Gorontalo	L	12/1/1989	S1 Teknik Informatika
4	T3112029	Irwin Hasan	Gorontalo	L	12/6/1988	S1 Teknik Informatika
5	T3112032	Alfianus Manoppo	Manado	L	10/2/1992	S1 Teknik Informatika
6	T3112041	Asrin Bumulo	Gorontalo	P	8/13/1993	S1 Teknik Informatika
7	T3112049	Onang Karim	Gorontalo	L	8/20/1995	S1 Teknik Informatika
8	T3112050	Hasrawati Husain	Kayubulan	P	11/9/1993	S1 Teknik Informatika
9	T3112053	Adrian Danial	Gorontalo	L	10/12/1991	S1 Teknik Informatika
10	T3112055	Sugianto	Gorontalo	L	10/7/1992	S1 Teknik Informatika
11	T3112057	Abdul Muhaimin Halid	Gorontalo	L	8/7/1985	S1 Teknik Informatika
12	T3112058	Widyawati Abdul Gani	Gorontalo	P	9/27/1993	S1 Teknik Informatika
13	T3112059	Ando Novriyanto Bodjo	Gorontalo	L	11/7/1994	S1 Teknik Informatika
14	T3112063	Havid Karim Mau	Gorontalo	L	10/20/1992	S1 Teknik Informatika
15	T3112072	Ismet Abdulllah	Labanu	L	10/27/1991	S1 Teknik Informatika
16	T3112073	Novprijan Hasan	Kabila	L	6/3/1994	S1 Teknik Informatika
17	T3112077	Djamaludin Maseke	Gorontalo	L	11/7/1993	S1 Teknik Informatika
18	T3112085	Moh Akbar Mile	Gorontalo	L	5/17/1992	S1 Teknik Informatika
19	T3112088	Rahim Bulotio	Wonggarasi Timur	L	10/20/1989	S1 Teknik Informatika
20	T3112093	Hermanto Maniku	Momalia	L	6/12/1990	S1 Teknik Informatika
21	T3112095	Riski Rahman	Sopi	L	7/16/1990	S1 Teknik Informatika
22	T3112101	Ahmad Ayuba	Telaga	L	8/18/1993	S1 Teknik Informatika
23	T3112112	Eko Susanto Saputra	Toli Toli	L	11/10/1993	S1 Teknik Informatika
24	T3112120	Supriyanto Bahu	Gorontalo	L	12/29/1991	S1 Teknik Informatika
25	T3112121	Moh Syarif Djupandang	Buol	L	12/22/1992	S1 Teknik Informatika
....
520	T3120131	Niko Arianto Ointu	Biak	L	10/10/2000	S1 Teknik Informatika

No.	Status Lulus	Shift	Kelas	IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPS7	IPS8	IPS9	IPS10
1	Lulus	N	7	3.50	3.75	3.64	3.70	3.44	2.00	3.07	2.50	1.86	1.33
2	Lulus	N	10	3.80	3.75	3.87	3.65	3.19	2.35	1.83	0.00	2.06	3.00
3	Drop-Out	N	6	3.77	3.78	3.64	3.77	0.00	0.00	0.00	0.00	0.00	0.00
4	Drop-Out	N	8	3.69	3.88	3.98	3.89	3.26	0.00	0.00	0.00	0.00	0.00
5	Drop-Out	N	8	3.66	3.87	3.76	3.88	3.65	2.95	3.05	3.00	0.00	0.00
6	Drop-Out	N	6	3.97	3.67	3.30	3.55	0.00	0.00	0.00	0.00	0.00	0.00
7	Drop-Out	N	10	3.88	3.64	3.50	3.58	0.00	0.00	0.00	0.00	0.00	0.00
8	Lulus	N	9	3.67	3.59	3.24	3.50	3.40	3.48	3.00	3.40	0.00	0.00
9	Drop-Out	N	6	3.59	3.98	3.65	3.80	0.00	0.00	0.00	0.00	0.00	0.00
10	Drop-Out	N	8	3.98	3.87	3.77	3.77	0.00	0.00	0.00	0.00	0.00	0.00
11	Lulus	N	7	3.87	3.68	3.89	3.69	3.40	2.50	2.45	2.81	3.50	0.00
12	Drop-Out	N	8	3.68	3.95	3.88	3.66	0.00	0.00	0.00	0.00	0.00	0.00
13	Lulus	N	9	3.95	3.88	3.55	3.97	3.46	3.22	2.44	3.00	0.00	0.00
14	Lulus	N	7	3.57	3.79	3.58	3.76	2.80	3.40	2.83	2.05	3.25	0.00
15	Drop-Out	N	6	4.00	3.76	3.68	3.98	2.95	2.74	2.26	0.40	0.00	0.00
16	Drop-Out	N	7	3.86	3.58	3.95	4.00	0.00	0.00	0.00	0.00	0.00	0.00
17	Lulus	N	9	3.85	3.67	3.88	3.76	3.17	1.58	0.00	0.77	1.60	1.80
18	Drop-Out	N	7	3.65	3.87	3.76	3.88	3.29	2.50	2.63	1.82	3.57	0.57
19	Drop-Out	N	6	3.88	3.64	3.78	3.97	0.00	0.00	0.00	0.00	0.00	0.00
20	Drop-Out	R	2	3.79	3.75	3.88	3.65	0.00	0.00	0.00	0.00	0.00	0.00
21	Drop-Out	R	1	3.76	3.78	3.87	3.79	0.00	0.00	0.00	0.00	0.00	0.00
22	Drop-Out	N	2	3.58	3.88	3.67	3.78	2.57	2.75	3.10	2.38	0.25	0.00
23	Drop-Out	R	4	3.70	3.83	4.00	3.87	0.00	0.00	0.00	0.00	0.00	0.00
24	Lulus	N	8	3.57	3.76	3.87	3.65	3.35	2.80	2.53	2.72	2.45	3.33
25	Drop-Out	R	4	3.87	3.76	3.88	3.65	2.50	2.37	2.44	0.00	0.00	0.00
....
520	Aktif	N	20	2.90	3.50	3.33	3.26	3.38	3.63	2.67	3.00	0.00	0.00

4.2 Hasil Pemodelan

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

Perhitungan data training yg digunakan dalam perhitungan manual adalah data training akhir (final) yang sudah di pilih atribut yang berpengaruh dan hasil

pemisahan antara data testing dan data training, Adapun data training nya sebagai berikut:

Tabel 4.2 Data Training Final

No	IPS2	IPS3	IPS4	Ket_Lulus
0	3.75	3.64	3.70	Tepat Waktu
1	3.75	3.87	3.65	Tepat Waktu
2	3.78	3.64	3.77	Tepat Waktu
3	3.88	3.98	3.89	Tepat Waktu
4	3.87	3.76	3.88	Tepat Waktu
5	3.67	3.30	3.55	Tidak Tepat Waktu
6	3.64	3.50	3.58	Tidak Tepat Waktu
7	3.59	3.24	3.50	Tidak Tepat Waktu
8	3.98	3.65	3.80	Tepat Waktu
9	3.87	3.77	3.77	Tepat Waktu
10	3.68	3.89	3.69	Tepat Waktu
11	3.95	3.88	3.66	Tepat Waktu
12	3.88	3.55	3.97	Tepat Waktu
13	3.79	3.58	3.76	Tepat Waktu
14	3.76	3.68	3.98	Tepat Waktu
15	3.58	3.95	4.00	Tepat Waktu
16	3.67	3.88	3.76	Tepat Waktu
17	3.87	3.76	3.88	Tepat Waktu
18	3.64	3.78	3.97	Tepat Waktu
19	3.75	3.88	3.65	Tepat Waktu
....
519	3.50	3.33	3.26	Tidak Tepat Waktu

4.3 Hasil Proses Pemodelan

Berdasarkan tabel 4.2 diatas merupakan hasil pemodelan dari dataset mahasiswa untuk judul skripsi Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Decision Tree* untuk mendapatkan hasil tersebut maka dilakukan tahapan sebagai berikut :

4.3.1 Tahapan Seleksi

Pada tahapan ini dilakukan seleksi atribut yang akan digunakan, berdasarkan pengumpulan data pada tabel 4.1 diatas, dilakukan seleksi atribut sehingga hanya menggunakan atribut yaitu : Jenis Kelamin, IPS1, IPS2, IPS3 dan IPS4 sehingga dataset yang akan digunakan dalam proses pemodelan dapat dilihat pada tabel 4.3 berikut :

Tabel 4.3 Dataset Mahasiswa Hasil Olahan

No.	Tahun	JK	IPS1	IPS2	IPS3	IPS4
1	2012	L	3.50	3.75	3.64	3.70
2	2012	P	3.80	3.75	3.87	3.65
3	2012	P	3.77	3.78	3.64	3.77
4	2012	P	3.69	3.88	3.98	3.89
5	2012	L	3.66	3.87	3.76	3.88
6	2012	L	3.97	3.67	3.65	3.55
7	2012	L	3.88	3.64	3.50	3.58
11	2012	P	3.67	3.59	3.70	3.50
12	2012	L	3.59	3.98	3.65	3.80
13	2012	L	3.98	3.87	3.77	3.77
14	2012	L	3.87	3.68	3.89	3.69
15	2012	L	3.68	3.95	3.88	3.66
16	2012	L	3.95	3.88	3.55	3.97
17	2012	L	3.57	3.79	3.58	3.76
18	2012	L	4.00	3.76	3.68	3.98
19	2012	P	3.86	3.58	3.95	4.00
....
520	2018	L	3.09	3.54	3.63	3.87

Proses diatas masih dilakukan secara manual menggunakan Microsoft Excel yang nantinya akan diolah menggunakan Bahasa pemrograman python menggunakan tools Google Colaboratory (Colab Notebooks).

4.3.2 Tahapan Preprocessing

Pada tahap ini dilakukan serangkaian proses diantaranya :

4.3.2.1 Mengumpulkan Data

1. Import Library Python

Langkah pertama yaitu, mengimport library-library yang akan digunakan :

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
```

2. Proses pengaitan Google Drive

Pada proses ini, dataset yang akan digunakan akan disimpan di google drive untuk itu perlu dikaitkan Google Drive Google Colaboratory dengan memasukan perintah berikut :

```
from google.colab import drive
drive.mount('/content/drive')
```

3. Proses Membaca Dataset

Selanjutnya adalah proses membaca dataset yang telah disimpan dari google drive, dengan memasukan perintah berikut :

```
df = pd.read_excel('/content/drive/MyDrive/Skripsi/dataset6.xlsx')
df.head()
```

Sedangkan hasil keluaran dari syntax diatas adalah sebagai berikut :

No.	Tahun	JK	IPS1	IPS2	IPS3	IPS4	Ket_Lulus
0	1	2012	L	3.50	3.75	3.64	3.70 Tepat Waktu
1	2	2012	P	3.80	3.75	3.87	3.65 Tepat Waktu
2	3	2012	P	3.77	3.78	3.64	3.77 Tepat Waktu
3	4	2012	P	3.69	3.88	3.98	3.89 Tepat Waktu
4	5	2012	L	3.66	3.87	3.76	3.88 Tepat Waktu

4. Mengkonversi dataset Excel ke CSV

Pada proses ini dilakukan konversi file excel ke format csv agar dataset yang digunakan lebih sederhan. Dengan menuliskan perintah berikut :

```
df.to_csv("/content/drive/MyDrive/Skripsi/dataset.csv",
          index=False)
df.info()
```

4.3.2.2 Menelaah Data

1. Info Dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 520 entries, 0 to 519
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   No          520 non-null    int64  
 1   Tahun        520 non-null    int64  
 2   JK           520 non-null    object  
 3   IPS1         520 non-null    float64 
 4   IPS2         520 non-null    float64 
 5   IPS3         520 non-null    float64 
 6   IPS4         520 non-null    float64 
 7   Ket_Lulus    520 non-null    object  
dtypes: float64(4), int64(2), object(2)
memory usage: 32.6+ KB
```

Berdasarkan info diatas, terdapat data yang memuat informasi terkait dataset yang digunakan, informasi tersebut menampilkan berupa : Jumlah total data yang digunakan, jumlah kolom dalam data, jenis type data yang digunakan dalam dataset 5 (tipe data: integer,object,float) dan jumlah penggunaan memori yang digunakan

2. Data Statistik

	JK	IPS1	IPS2	IPS3	IPS4
count	520.000000	520.000000	520.000000	520.000000	520.000000
mean	1.384615	3.530596	3.614962	3.410135	3.400192
std	0.486973	0.448693	0.575888	0.633752	0.622290
min	1.000000	0.800000	0.100000	0.080000	0.100000
25%	1.000000	3.430000	3.600000	3.250000	3.260000
50%	1.000000	3.650000	3.750000	3.550000	3.570000
75%	2.000000	3.780000	3.880000	3.790000	3.780000
max	2.000000	4.000000	4.000000	4.000000	4.000000

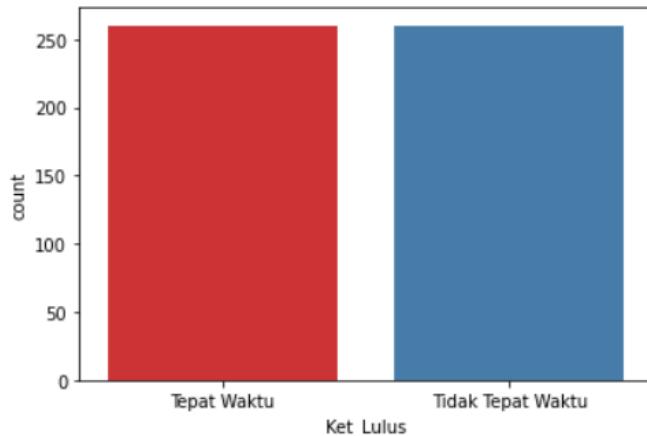
Dalam data statistik ini, menampilkan sejumlah informasi berupa jumlah data yang digunakan, nilai rata-rata, standar deviasi, nilai minimum, nilai dalam range 25%, 50% dan 75% serta nilai maximum dari setiap atribut yang digunakan.

4.3.2.3 Memvalidasi Data

Pada tahap ini akan menampilkan hasil dari balance data untuk status kelulusan, dengan menuliskan perintah berikut :

```
sns.countplot(df['Ket_Lulus'], palette='Set1')
```

Setelah perintah diatas dijalankan maka akan muncul grafik balance data untuk status kelulusan mahasiswa seperti pada gambar dibawah ini :



Gambar 4.1 Grafik Waktu Kelulusan Mahasiswa

Berdasarkan gambar diatas, bisa dilihat bahwa jumlah mahasiswa yang lulus tepat waktu dengan yang tidak tepat waktu itu seimbang.

4.3.3 Tahap Transformation

1. Memilih dan memilah data

Pada tahap ini, akan dilakukan proses pemilihan dan pemilahan terhadap data yang akan digunakan. Beberapa atribut yang tidak digunakan akan dikeluarkan sehingga hasil dari pemilihan dan pemilahan data adalah sebagai berikut :

```
df.drop(['No.', 'Tahun'], axis=1, inplace=True)
df.head()
```

Setelah perintah diatas dijalankan, hasil keluarannya sebagai berikut :

	JK	IPS1	IPS2	IPS3	IPS4	Ket_Lulus
0	1	3.50	3.75	3.64	3.70	Tepat Waktu
1	2	3.80	3.75	3.87	3.65	Tepat Waktu
2	2	3.77	3.78	3.64	3.77	Tepat Waktu
3	2	3.69	3.88	3.98	3.89	Tepat Waktu
4	1	3.66	3.87	3.76	3.88	Tepat Waktu

2. Membersihkan Data

Pada tahap ini akan dilakukan proses pembersihan data, dimana akan dilakukan pengecekan terhadap data yang kosong dan duplikat

3. Merekonstruksi Data

Pada tahap ini, akan dilakukan proses pengubahan atau konversi terhadap atribut yang bernilai string menjadi numerik. Salah satu atribut string pada dataset adalah atribut jenis kelamin. Untuk mengkonversi atribut jenis kelamin tersebut diperlukan perintah sebagai berikut :

```
df['JK'].replace(['L', 'P'], [1, 2], inplace=True)
df.head()
```

Maka hasil dari perintah diatas adalah sebagai berikut :

No.	Tahun	JK	IPS1	IPS2	IPS3	IPS4	Ket_Lulus
0	1	2012	1	3.50	3.75	3.64	3.70 Tepat Waktu
1	2	2012	2	3.80	3.75	3.87	3.65 Tepat Waktu
2	3	2012	2	3.77	3.78	3.64	3.77 Tepat Waktu
3	4	2012	2	3.69	3.88	3.98	3.89 Tepat Waktu
4	5	2012	1	3.66	3.87	3.76	3.88 Tepat Waktu

4.3.4 Tahap Data Mining

1. Membagi Data Training dan Data Testing

Pada tahap ini, akan dilakukan pembagian terhadap dataset menjadi dua bagian yaitu data training dan data testing. Untuk proses pembagian datanya dengan membagi data training sebesar 70% dan data testing sebesar 30%. Dimana untuk data training berjumlah 364 data dan data testing berjumlah 156 data. Untuk proses pembagiannya dapat dilihat sebagai berikut :

```
# 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=123)
print ('Train set:', X_train.shape, y_train.shape)
print ('Test set:', X_test.shape, y_test.shape)
```

Setelah perintah diatas dijalankan, akan muncul hasil keluaran berikut :

⇒ Train set: (364, 5) (364,)
Test set: (156, 5) (156,)

Adapun juga data training yang bisa dilihat pada tabel 4.4 berikut :

Tabel 4.4 Dataset Training

No.	IPS2	IPS3	IPS4	Ket_Lulus
0	3,75	3,64	3,70	Tepat Waktu
2	3,78	3,64	3,77	Tepat Waktu
3	3,88	3,98	3,89	Tepat Waktu
5	3,67	3,30	3,55	Tidak Tepat Waktu
9	3,87	3,77	3,77	Tepat Waktu
11	3,95	3,88	3,66	Tepat Waktu
13	3,79	3,58	3,76	Tepat Waktu
14	3,76	3,68	3,98	Tepat Waktu
16	3,67	3,88	3,76	Tepat Waktu
17	3,87	3,76	3,88	Tepat Waktu
18	3,64	3,78	3,97	Tepat Waktu
19	3,75	3,88	3,65	Tepat Waktu
20	3,78	3,87	3,79	Tepat Waktu
22	3,83	4,00	3,87	Tepat Waktu
23	3,76	3,87	3,65	Tepat Waktu
24	3,76	3,88	3,65	Tepat Waktu
25	3,67	3,00	2,83	Tidak Tepat Waktu
27	3,83	2,36	2,75	Tidak Tepat Waktu
28	3,78	3,87	3,61	Tepat Waktu
29	3,65	3,64	3,80	Tepat Waktu
30	3,56	2,41	2,27	Tidak Tepat Waktu
31	3,74	3,86	3,80	Tepat Waktu
32	3,61	3,00	3,57	Tidak Tepat Waktu
33	3,50	3,77	3,48	Tidak Tepat Waktu
35	4,00	3,52	3,61	Tepat Waktu
.....
519	3,50	3,33	3,26	Tepat Waktu

Berikut adalah data testing yang akan dituangkan dalam tabel 4.5 berikut :

Tabel 4.5 Dataset Testing

No.	IPS2	IPS3	IPS4	Ket_Lulus
1	3,75	3,87	3,65	Tepat Waktu
4	3,87	3,76	3,88	Tepat Waktu
6	3,64	3,50	3,58	Tidak Tepat Waktu
7	3,59	3,24	3,50	Tidak Tepat Waktu
8	3,98	3,65	3,80	Tepat Waktu
10	3,68	3,89	3,69	Tepat Waktu
12	3,88	3,55	3,97	Tepat Waktu
15	3,58	3,95	4,00	Tepat Waktu
21	3,88	3,67	3,78	Tepat Waktu
26	3,71	2,89	2,80	Tidak Tepat Waktu
34	3,83	3,73	3,48	Tidak Tepat Waktu
37	3,52	3,20	3,61	Tepat Waktu
45	3,65	3,24	3,50	Tidak Tepat Waktu
46	3,57	3,57	3,61	Tepat Waktu
49	3,52	3,76	3,71	Tepat Waktu
52	3,70	3,28	3,57	Tidak Tepat Waktu
54	3,55	4,00	3,74	Tepat Waktu
55	3,52	3,77	3,61	Tepat Waktu
56	3,91	2,80	3,57	Tidak Tepat Waktu
59	3,57	3,74	3,67	Tepat Waktu
60	3,52	3,62	3,74	Tepat Waktu
65	4,00	3,64	3,61	Tepat Waktu
71	3,50	3,75	3,24	Tidak Tepat Waktu
74	3,92	3,79	3,75	Tepat Waktu
75	3,50	3,67	2,83	Tidak Tepat Waktu
76	3,58	3,58	3,33	Tidak Tepat Waktu
....
517	3,50	3,50	3,78	Tepat Waktu

Untuk menentukan root (node) awal dari pohon keputusan, Langkah pertama yaitu dengan menghitung dan mencari nilai entrophy dan gain tertinggi untuk tiap-tiap atributnya. Berikut adalah Langkah Langkah penyelesaiannya:

Tahap 1 : Menghitung nilai entropi total untuk seluruh atribut

Atribut	Jumlah	Tidak Lulus	Lulus	Entropy
Total	364	186	178	0.99965154

$$Entropi(S) = \left(-\frac{186}{364} \times \log_2\left(\frac{186}{364}\right)\right) + \left(-\frac{178}{364} \times \log_2\left(\frac{178}{364}\right)\right) = 0.99965154$$

Setelah menghitung nilai entropi total dari semua atribut, Langkah selanjutnya adalah tahap 2 menghitung nilai entropi tiap-tiap atribut nya, sebagai berikut:

Tabel 4.6 Entropi untuk setiap atribut

Atribut		Jumlah	Tidak Lulus	Lulus	Entropy	Gain
IPS2	Gain				0.15081777	
	<=3.58	87	78	9	0.47983202	
	>3.58	277	108	169	0.96472962	
	Gain				0.163279368	
	<=3.75	191	140	51	0.83714969	
IPS3	>3.75	173	46	127	0.83551375	
	Gain				0.337433351	
	<=3.40	124	118	6	0.27950555	
	>3.40	240	68	172	0.85995305	
	Gain				0.299262426	
IPS4	<=3.55	185	151	34	0.68828496	
	>3.55	179	35	144	0.71289899	
	Gain				0.322319226	
	<=3.38	124	117	7	0.31319505	
	>3.38	240	69	171	0.8654699	
Gain					0.628848364	
	<=3.55	184	172	12	0.34781691	
	>3.55	180	14	166	0.39430024	

Pada table 4.6 diatas, setelah di hitung seluruh atribut untuk mencari nilai entropi selanjutnya menghitung dan mencari nilai gain tertinggi dari

semuanya. Dan di dapatkan nilai gain tertinggi yaitu 0.628848364 sehingga yang menjadi root (node) awal dari pohon keputusan adalah IPS4.

2. Pemodelan dengan *Decision Tree*

Pada tahap ini dilakukan proses pemodelan dengan menggunakan metode *decision tree*, dimana pemodelan ini yang nantinya akan menciptakan sebuah pohon keputusan akhir. Untuk koding pemodelan ini bisa di buat dengan menjalankan perintah berikut :

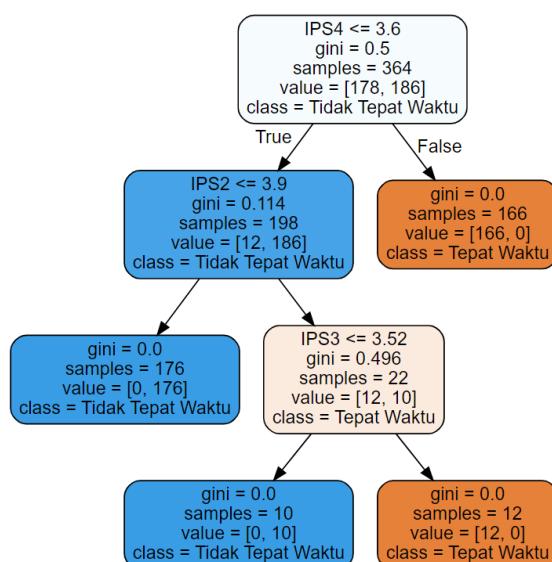
```
from sklearn.tree import DecisionTreeClassifier

# membuat model Decision Tree
tree_model = DecisionTreeClassifier()

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

3. Visualisasi Model

Setelah membuat model untuk *decision tree*, maka pada tahap ini akan menampilkan pohon keputusan dari hasil pemodelan yang telah dibuat sebelumnya yang memuat beberapa node (rule). Hasil visualisasi model *Decision Tree* dapat dilihat pada gambar berikut :



Gambar 4.2 Visualisasi Pohon Keputusan

4.3.5 Tahap Interpretation / evaluation

1. Evaluasi Model

Pada tahap ini akan dilakukan proses evaluasi model *Decision Tree*, dimana pada proses ini akan memprediksi nilai akurasi dari dataset. Dataset yang diprediksi menggunakan data uji yang telah dibagi sebelumnya. dengan menuliskan perintah berikut :

```
# Evaluasi Model
from sklearn.metrics import accuracy_score

y_pred = tree_model.predict(X_test)

acc_secore = round(accuracy_score(y_pred, y_test), 3)

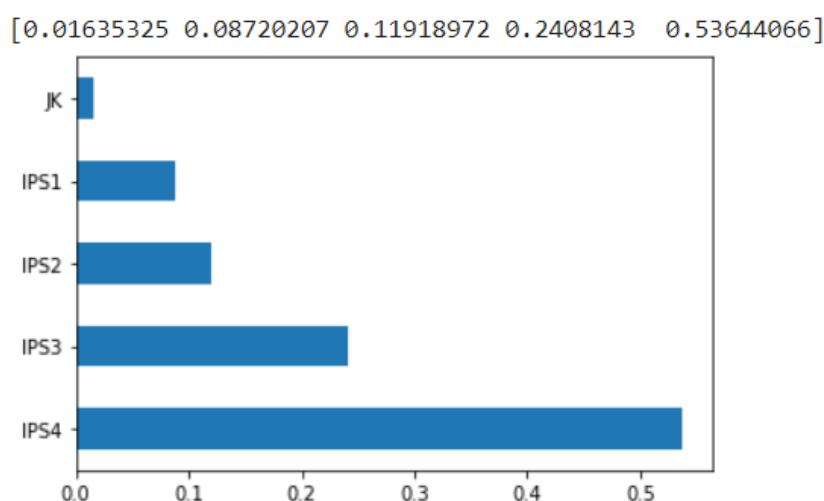
print('Accuracy: ', acc_secore)
```

Setelah perintah diatas dijalankan maka akan keluar nilai akurasi sebagai berikut :

↳ Accuracy: 0.904

2. Evaluasi Proses

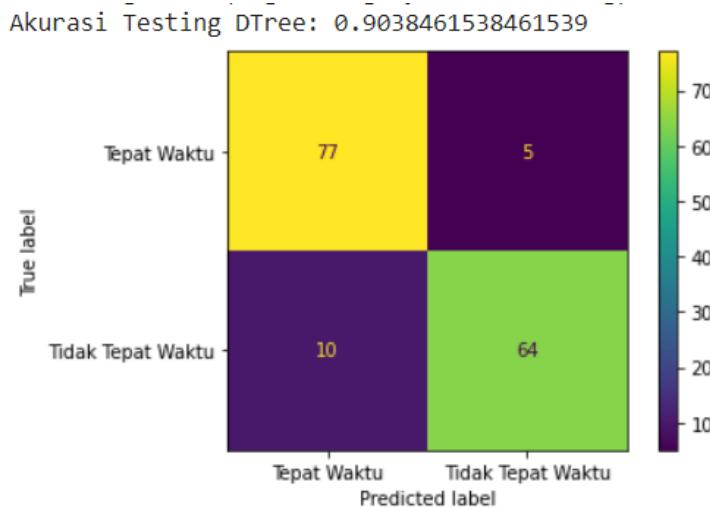
Pada tahap ini, dilakukan proses evaluasi terhadap variabel atau atribut-atribut yang tidak berpengaruh, yang di tampilkan pada grafik berikut ini :



Gambar 4.3 Grafik Persentase Atribut

Berdasarkan grafik diatas, terdapat atribut jenis kelamin dan IPS1 yang tidak berpengaruh dalam proses prediksi dan pada hasil pohon keputusan, sehingga perlu adanya penghapusan atribut yang tidak berpengaruh tersebut.

3. Perbaikan Model



Gambar 4.4 Hasil Prediksi Dengan Model *Confusion Matrix*

Berdasarkan pada gambar diatas bahwa, hasil prediksi dengan menggunakan model *confusion matrix*, mendapatkan nilai akurasi sebesar 90%, yang diperoleh dari data uji. Akan tetapi, perlu di rebuild kembali karena adanya penghapusan atribut yang tidak berpengaruh. Sehingga adapun susunannya sebagai berikut :

1. Import Library Python

Langkah pertama yaitu, mengimport library-library yang akan digunakan :

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
```

2. Proses pengaitan Google Drive

Pada proses ini, dataset yang akan digunakan akan disimpan di google drive untuk itu perlu dikaitkan Google Drive Google Colaboratory dengan memasukan perintah berikut :

```
from google.colab import drive
drive.mount('/content/drive')
```

3. Proses Membaca Dataset

Selanjutnya adalah proses membaca dataset yang telah disimpan dari google drive, dengan memasukan perintah berikut :

```
df = pd.read_excel('/content/drive/MyDrive/Skripsi/dataset6.xlsx')
df.head()
```

Sedangkan hasil keluaran dari syntax diatas adalah sebagai berikut :

No.	Tahun	JK	IPS1	IPS2	IPS3	IPS4	Ket_Lulus
0	1	2012	L	3.50	3.75	3.64	3.70 Tepat Waktu
1	2	2012	P	3.80	3.75	3.87	3.65 Tepat Waktu
2	3	2012	P	3.77	3.78	3.64	3.77 Tepat Waktu
3	4	2012	P	3.69	3.88	3.98	3.89 Tepat Waktu
4	5	2012	L	3.66	3.87	3.76	3.88 Tepat Waktu

4. Mengkonversi dataset Excel ke CSV

Pada proses ini dilakukan konversi file excel ke format csv agar dataset yang digunakan lebih sederhan. Dengan menuliskan perintah berikut :

```
df.to_csv("/content/drive/MyDrive/Skripsi/dataset.csv",
index=False)
df.info()
```

5. Data Statistik

	JK	IPS1	IPS2	IPS3	IPS4
count	520.000000	520.000000	520.000000	520.000000	520.000000
mean	1.384615	3.530596	3.614962	3.410135	3.400192
std	0.486973	0.448693	0.575888	0.633752	0.622290
min	1.000000	0.800000	0.100000	0.080000	0.100000
25%	1.000000	3.430000	3.600000	3.250000	3.260000
50%	1.000000	3.650000	3.750000	3.550000	3.570000
75%	2.000000	3.780000	3.880000	3.790000	3.780000
max	2.000000	4.000000	4.000000	4.000000	4.000000

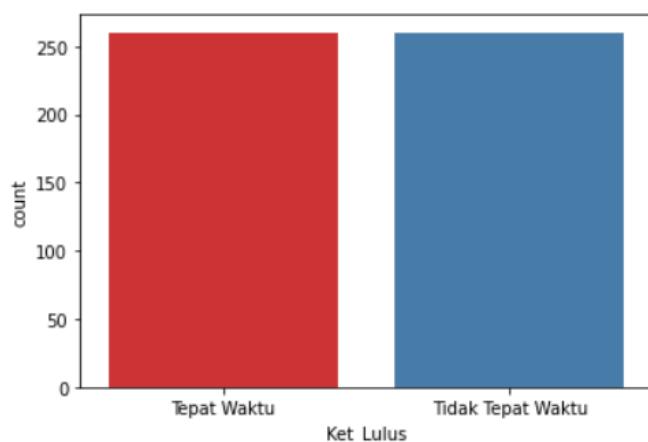
Dalam data statistik ini, menampilkan sejumlah informasi berupa jumlah data yang digunakan, nilai rata-rata, standar deviasi, nilai minimum, nilai dalam range 25%, 50% dan 75% serta nilai maximum dari setiap atribut yang digunakan.

6. Validasi Data

Pada tahap ini akan menampilkan hasil dari balance data untuk status kelulusan, dengan menuliskan perintah berikut :

```
sns.countplot(df['Ket_Lulus'], palette='Set1')
```

Setelah perintah diatas dijalankan maka akan muncul grafik balance data untuk status kelulusan mahasiswa seperti pada gambar dibawah ini :



Gambar 4.5 Grafik Waktu Kelulusan Mahasiswa

Berdasarkan gambar diatas, bisa dilihat bahwa jumlah mahasiswa yang lulus tepat waktu dengan yang tidak tepat waktu itu seimbang.

7. Memilih dan Memilah Data

Pada tahap ini, akan dilakukan proses pemilihan dan pemilahan terhadap data yang akan digunakan. Beberapa atribut yang tidak digunakan akan dikeluarkan sehingga hasil dari pemilihan dan pemilahan data adalah sebagai berikut :

```
df.drop(['No', 'Tahun', 'JK', 'IPS1'], axis=1, inplace=True)
df.head()
```

Setelah perintah diatas dijalankan, hasil keluarannya sebagai berikut :

	<u>IPS2</u>	<u>IPS3</u>	<u>IPS4</u>	<u>Ket_Lulus</u>
0	3.75	3.64	3.70	Tepat Waktu
1	3.75	3.87	3.65	Tepat Waktu
2	3.78	3.64	3.77	Tepat Waktu
3	3.88	3.98	3.89	Tepat Waktu
4	3.87	3.76	3.88	Tepat Waktu

8. Membersihkan Data

Pada tahap ini akan dilakukan proses pembersihan data, dimana akan dilakukan pengecekan terhadap data yang kosong dan duplikat

9. Merekonstruksi Data

Pada tahap ini, akan dilakukan proses pengubahan atau konversi terhadap atribut yang bernilai string menjadi numerik. Salah satu atribut string pada dataset adalah atribut jenis kelamin. Untuk mengkonversi atribut jenis kelamin tersebut diperlukan perintah sebagai berikut :

```
df['JK'].replace(['L', 'P'], [1, 2], inplace=True)
df.head()
```

Maka hasil dari perintah diatas adalah sebagai berikut :

No.	Tahun	JK	IPS1	IPS2	IPS3	IPS4	Ket_Lulus
0	1	2012	1	3.50	3.75	3.64	3.70 Tepat Waktu
1	2	2012	2	3.80	3.75	3.87	3.65 Tepat Waktu
2	3	2012	2	3.77	3.78	3.64	3.77 Tepat Waktu
3	4	2012	2	3.69	3.88	3.98	3.89 Tepat Waktu
4	5	2012	1	3.66	3.87	3.76	3.88 Tepat Waktu

10. Pemodelan Dengan *Decision Tree*

Pada tahap ini dilakukan proses pemodelan dengan menggunakan metode *decision tree*, dimana pemodelan ini yang nantinya akan menciptakan sebuah pohon keputusan akhir. Untuk coding pemodelan ini bisa di buat dengan menjalankan perintah berikut :

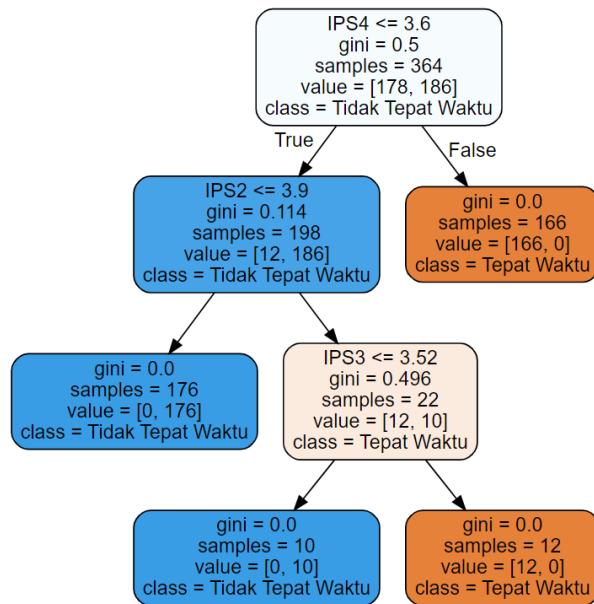
```
from sklearn.tree import DecisionTreeClassifier

# membuat model Decision Tree
tree_model = DecisionTreeClassifier()

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

11. Visualisasi Pohon Keputusan

Setelah membuat model untuk *decision tree*, maka pada tahap ini akan menampilkan pohon keputusan dari hasil pemodelan yang telah dibuat sebelumnya yang memuat beberapa node (rule). Hasil visualisasi model *Decision Tree* dapat dilihat pada gambar berikut :

**Gambar 4.6** Visualisasi Pohon Keputusan

12. Evaluasi Model

Pada tahap ini akan dilakukan proses evaluasi model *Decision Tree*, dimana pada proses ini akan memprediksi nilai akurasi dari dataset. Dataset yang diprediksi menggunakan data uji yang telah dibagi sebelumnya. dengan menuliskan perintah berikut :

```

# Evaluasi Model
from sklearn.metrics import accuracy_score

y_pred = tree_model.predict(X_test)

acc_secore = round(accuracy_score(y_pred, y_test), 3)

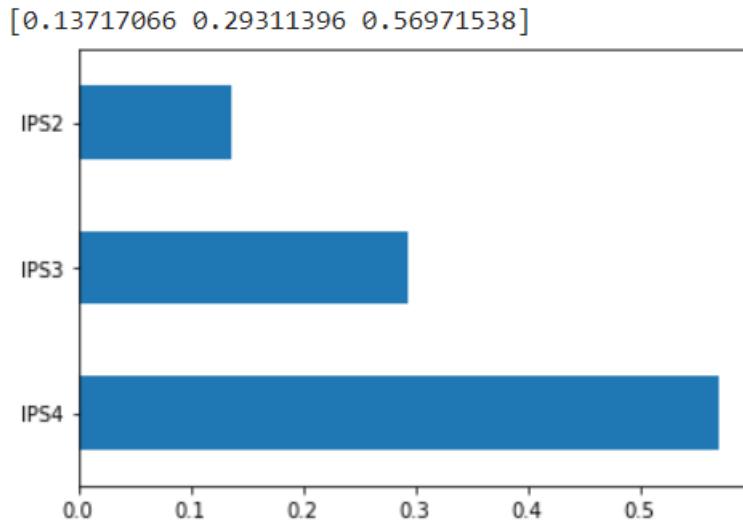
print('Accuracy: ', acc_secore)
  
```

Setelah perintah diatas dijalankan maka akan keluar nilai akurasi sebagai berikut :

→ Accuracy: 0.904

13. Evaluasi Proses

Pada tahap ini, dilakukan proses evaluasi terhadap variabel atau atribut-atribut yang tidak berpengaruh, yang di tampilkan pada grafik berikut ini :



Gambar 4.7 Grafik Persentase Atribut

Berdasarkan grafik diatas, adalah atribut-atribut final yang berpengaruh terhadap hasil dari pohon keputusan yaitu IPS2, IPS3 dan IPS4.

BAB V

PEMBAHASAN PENELITIAN

5.1 Pembahasan Kinerja Model

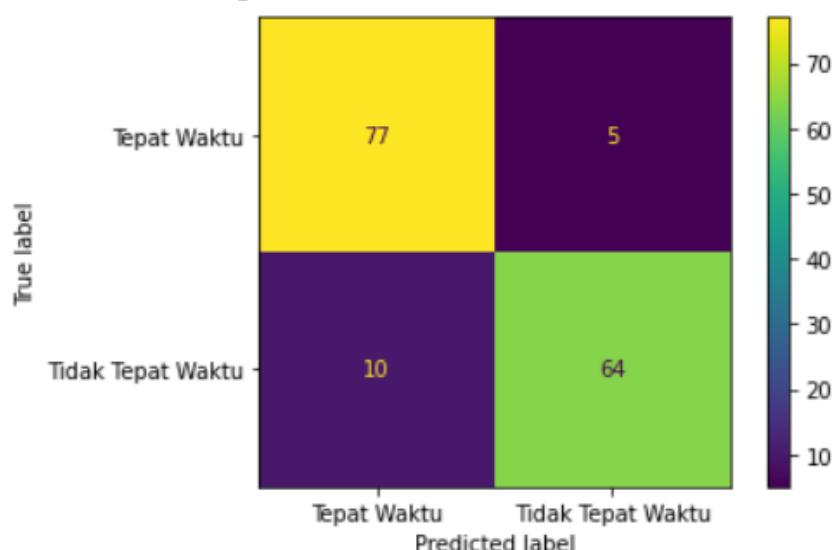
Kinerja metode *Decision Tree* dalam pengklasifikasian data tentu sangat di dasarkan dari banyaknya data yang diperoleh, agar tidak memakan waktu yang lama dalam melakukan uji coba beberapa jumlah prediksi yang tepat, maka diperlukan suatu teknik untuk menentukan jumlah prediksi yang optimal yaitu dengan menggunakan teknik *Confusion Matix*. Berikut adalah sintaks penggunaan teknik *Confusion Matrix* :

```
# Evaluasi Model dengan confusion_matrix
from sklearn.metrics import accuracy_score, plot_confusion_matrix

prediksi = tree_model.predict(X_test)
plot_confusion_matrix(tree_model, X_test, y_test)
accuracy = accuracy_score(y_test, prediksi)
print('Akurasi Testing DTee:', accuracy)
```

Setelah perintah tersebut dijalankan, maka ditampilkan hasil keluaran seperti berikut :

Akurasi Testing DTee: 0.9038461538461539



Gambar 5.1 Evaluasi dengan *Confusion Matrix*

Pada gambar 4.1 diatas, terdapat hasil nilai akurasi dari proses prediksi mahasiswa menggunakan *Confusion Matrix* dengan nilai akurasi sebesar 90% yang diperoleh dari data testing yang berjumlah 156 data, dengan jumlah mahasiswa tepat waktu sebanyak 77 orang, jumlah mahasiswa tidak lulus tepat waktu sebanyak 64 orang, dan jumlah mahasiswa yang seharusnya tidak lulus tepat waktu tetapi diprediksi tepat waktu sebanyak 10 orang serta jumlah mahasiswa yang seharusnya lulus tepat waktu tetapi diprediksi sebagai tidak tepat waktu sebanyak 5 orang. Untuk mengukur performance metrics dari model evaluasi dengan *Confusion Matix*, dapat menggunakan *accuracy*, *precision*, dan *recall*. perhitungan manualnya menggunakan persamaan berikut :

1. Accuracy

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (\text{Persamaan 1})$$

$$\begin{aligned} \text{Accuracy} &= \frac{77+64}{77+64+5+10} = \frac{141}{156} = \frac{47}{52} \times 100\% \\ &= 0.90 * 100\% = 90\% \end{aligned}$$

2. Precision

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \quad (\text{Persamaan 2})$$

$$\begin{aligned} \text{Precision} &= \frac{77}{77+5} = \frac{77}{82} \times 100\% \\ &= 0.93 * 100\% = 93\% \end{aligned}$$

3. Recall

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (\text{Persamaan 3})$$

$$\begin{aligned} \text{Recall} &= \frac{77}{77+10} = \frac{77}{87} \times 100\% \\ &= 0.88 * 100\% = 88\% \end{aligned}$$

Berdasarkan perhitungan manual diatas, bisa disimpulkan bahwa evaluasi dengan menggunakan *Confusion Matrix* sudah sesuai dengan memperoleh nilai *accuracy* 90%, *precision* 93% dan *recall* 88%.

5.2 Pembahasan Model

5.2.1 Proses Pemodelan Pohon Keputusan dari Hasil Prediksi

setelah didapatkan jumlah prediksi yang optimum, maka selanjutnya adalah proses penerapan model *decision tree* dan pohon keputusan serta melatih model menggunakan data training, dengan menuliskan perintah berikut :

```
from sklearn.tree import DecisionTreeClassifier

# membuat model Decision Tree
tree_model = DecisionTreeClassifier()

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

5.2.2 Proses Prediksi dengan Model Pohon Keputusan

Selanjutnya adalah proses prediksi model dengan pohon keputusan, dengan menuliskan perintah berikut :

```
# prediksi model dengan tree_model.predict([[ 'IPS II',
'IPS III', 'IPS IV']])
hasil=tree_model.predict([[3.8, 3.6, 4]])
print(hasil)
```

keluaran sebagai berikut :

```
['Tepat Waktu']
/usr/local/lib/python3.8/
warnings.warn(
```

5.2.3 Proses Pembuatan Path Storage File Hasil Pohon Keputusan

Langkah selanjutnya adalah proses pembuatan path atau jalur penyimpanan hasil pohon keputusan, dengan menuliskan perintah berikut :

```

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

```

5.2.4 Proses Hasil Keluaran Pohon Keputusan

Langkah terakhir adalah, proses pemnbuatan pohon keputusan.

Berikut adalah hasil dari pohon keputusan yang terbentuk, dengan menuliskan perintah berikut :

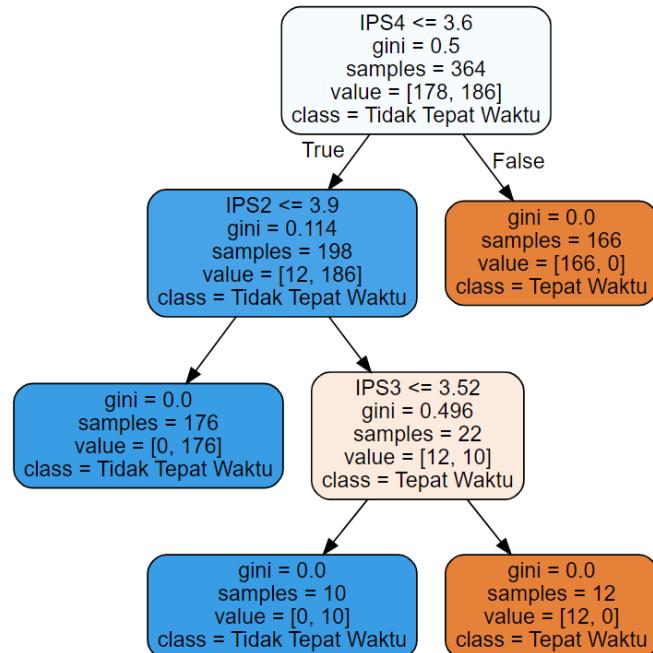
Tabel 5.1 Entropi Total

```

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 = ['IPS2', 'IPS3','IPS4'],
    class_names = ['Tepat Waktu', 'Tidak Tepat Waktu']
,
    rounded= True,
    filled =True)
Source.from_file(os.path.join(IMAGES_PATH, "lulus_tree7
.dot"))

```

Setelah perintah diatas dijalankan, maka akan keluar hasil keluaran dari pohon keputusan yang telah terbentuk seperti berikut ini :



Gambar 5.2 Pohon Keputusan

Selanjutnya, seperti yang ditunjukan pada tabel dibawah ini adalah hasil ekstraksi dari pohon keputusan diatas yang membentuk aturan sebagai berikut :

Tabel 5.2 Role Akhir

No.	Rule
1	Jika IPS4 > dari 3.6 Maka Tepat Waktu
2	Jika IPS4 <= 3.6 dan IPS2 <= 3.9 Maka Tidak Tepat Waktu
3	Jika IPS4 <= 3.6 dan IPS2 <= 3.9 dan IPS3 <= 3.52 Maka Tidak Tepat Waktu
4	Jika IPS4 <= 3.6 dan IPS2 <= 3.9 dan IPS3 > 3.52 Maka Tepat Waktu

Terdapat 4 aturan yang didapatkan untuk memprediksi status kelulusan mahasiswa. Selanjutnya 4 aturan tersebut dapat digunakan sebagai kelas/kelompok untuk memprediksi mahasiswa yang lulus tepat waktu ataupun tidak tepat waktu di Prodi Teknik Informatika Unisan Gorontalo.

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Dari hasil penelitian dengan sistem untuk memprediksi kelulusan mahasiswa tepat waktu menggunakan metode *Decision Tree* pada Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo, maka pada akhir laporan penelitian ini penulis menyimpulkan bahwa :

1. Peneliti dapat mengetahui cara prediksi kelulusan mahasiswa tepat waktu menggunakan metode *Decision Tree* pada Prodi Teknik Informatika Fikom Unisan Gorontalo.
2. Peneliti juga dapat mengetahui hasil penerapan daripada metode *Decision Tree* dalam memprediksi kelulusan mahasiswa tepat waktu dengan mendapatkan nilai akurasi sebesar 90%.
3. Dari dataset awal yang digunakan, ditemukan beberapa atribut yang tidak berpengaruh dalam hasil pohon keputusan yakni : atribut JK dan IPS1 sehingga diperlukan untuk menghapus atribut tersebut.

6.2 Saran

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

1. Penelitian selanjutnya dapat mengoptimalkan metode *Decision Tree* dengan menambahkan jumlah data agar dapat mendapatkan hasil yang jauh lebih baik lagi.
2. Dapat dikembangkan dengan membuat program aplikasi untuk memprediksi kelulusan mahasiswa tepat waktu menggunakan metode *Decision Tree*.
3. Peneliti selanjutnya dapat menambahkan variabel berpengaruh seperti lama studi sehingga proses prediksi dan akurasi bisa lebih optimal.

DAFTAR PUSTAKA

- [1] A. Rohman, “Model Algoritma K-Nearest Neighbor (K-NN) Untuk Prediksi Kelulusan Mahasiswa,” 2015.
- [2] M. Kafil, “Penerapan Metode K-Nearest Neighbors Untuk Prediksi Penjualan Berbasis Web Pada Boutiq Dealove Bondowoso,” 2019.
- [3] A. Yuliana and D. B. Pratomo, “Seminar Nasional Inovasi Teknologi Algoritma Decision Tree (C4.5) Untuk Memprediksi Kepuasan Mahasiswa Terhadap Kinerja Dosen Politeknik TEDC Bandung,” 2017.
- [4] Ramadhani Rahmi and Hendriyani Yeka, “Prediksi Prestasi Siswa Berbasis Data Mining Menggunakan Algoritma Decision Tree (Studi Kasus : SMKN 2 Padang),” 2021.
- [5] H. Wahono and D. Riana, “Prediksi Calon Pendonor Darah Potensial Dengan Algoritma Naïve Bayes, K-Nearest Neighbors dan Decision Tree C4.5,” *JURIKOM (Jurnal Riset Komputer)*, vol. 7, no. 1, p. 7, Feb. 2020.
- [6] N. Handayani, H. Wahyono, J. Trianto, and D. S. Permana, “Prediksi Tingkat Risiko Kredit dengan Data Mining Menggunakan Algoritma Decision Tree C.45,” *(Jurnal Riset Komputer)*, vol. 8, no. 6, pp. 2407–389, 2021.
- [7] M. P. Dr. Ir. Ch. Wariyah *et al.*, “Prokrastinasi Akademik Dalam Penyelesaian Skripsi,” 2014.
- [8] Ganda and Yahya, “Petunjuk Praktis Cara Mahasiswa Belajar di Perguruan Tinggi,” 2004.
- [9] Muslim Much Aziz *et al.*, “Data Mining Algoritma C4.5,” 2019.

- [10] D. Yunita, “Perbandingan Algoritma K-Nearest Neighbor Dan Decision Tree Untuk Penentuan Risiko Kredit Kepemilikan Mobil,” vol. 103, no. 2, 2017.
- [11] Suhesti Tian, “Bahasa Pemrograman Python,” 2014.
- [12] M. Fluorida Fibrianda and A. Bhawiyuga, “Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naïve Bayes Dan Support Vector Machine (SVM),” 2018.
- [13] A. H. Nasrullah, “Penerapan Metode C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop Out,” 2018.