

LAPORAN_PENELITIAN_LPPM_- _2020.pdf *by*

Submission date: 19-Jun-2023 11:11PM (UTC+0700)

Submission ID: 2119136166

File name: LAPORAN_PENELITIAN_LPPM_-_2020.pdf (2.01M)

Word count: 10653

Character count: 63439

LAPORAN PENELITIAN

Klaster Penelitian Pengembangan Pendidikan Tinggi

**Penerapan Konsep *Data Science* dan *Machine Learning* untuk
Kelangsungan dan Keberhasilan Studi Mahasiswa UIN SUSKA
Riau**



PENELITI

Ketua : Dr. Elin Haerani, S.T., M.Kom

NIDN : 2023058101

Anggota : Fadhilah Syafria, S.T., M.Kom

NIDN : 2007108502

**LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU**

2021

ABSTRAK

Abstrak – Salah satu penentu kualitas Perguruan Tinggi adalah tingginya tingkat keberhasilan mahasiswa. Tingkat keberhasilan mahasiswa dapat dilihat pada kelulusan mahasiswa yang tepat waktu. Universitas Islam Negeri Sultan Syarif Kasim Riau adalah salah satu perguruan tinggi negeri di Riau yang memiliki total mahasiswa mencapai 30.000 mahasiswa. Dari seluruh mahasiswa yang aktif terdapat mahasiswa yang tidak aktif. Keberadaan mahasiswa yang tidak aktif mempengaruhi mahasiswa sehingga lulus tidak tepat waktu. Tidak adanya informasi yang terkait dengan faktor penyebab mahasiswa yang tidak aktif sehingga sampai terancam *drop out* dan informasi serta data apa yang mempengaruhi mahasiswa dapat lulus tepat waktu, maka universitas selalu mengevaluasi kinerja mahasiswanya. Hasil evaluasi disimpan dalam *database* akademik sehingga data dapat digunakan sebagai pendukung pengambilan keputusan oleh universitas. Pada penelitian ini menggunakan konsep *data science* untuk mengeksplorasi dan mengekstraksi sekumpulan data dari *database* sehingga dari kumpulan data tersebut ditemukan model atau pola, serta wawasan baru yang dapat digunakan sebagai alat untuk pengambilan keputusan. Setelah data dieksplorasi, maka digunakan konsep *machine learning* untuk mengenali dan mengklasifikasi data. Metode yang digunakan adalah Metode *Decision Tree*. Hasil penelitian menunjukkan bahwa kedua konsep ini dapat memberikan hasil sesuai dengan yang diharapkan. Berdasarkan hasil pengujian bahwa, Atribut yang berpengaruh terhadap keberhasilan studi mahasiswa adalah Indeks Prestasi Kumulatif (IPK) dan akurasi tingkat pengenalan maksimal adalah 88,19%.

Kata Kunci – *Data science, Decision Tree, Lulus Tepat Waktu, Machine Learning*

DAFTAR ISI

<i>ABSTRAK</i>	<i>i</i>
<i>DAFTAR ISI</i>	<i>ii</i>
<i>DAFTAR TABEL</i>	<i>iv</i>
<i>DAFTAR GAMBAR</i>	<i>v</i>
<i>BAB I PENDAHULUAN</i>	<i>1</i>
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Batasan masalah	6
1.4 Tujuan penelitian.....	7
1.5 Manfaat Penelitian	7
1.6 Sistematika Penulisan.....	7
<i>BAB II TINJAUAN PUSTAKA</i>	<i>9</i>
2.1 Data Mining.....	9
2.1.1 Tahap – Tahap <i>Data Mining</i>	9
2.1.2 Teknik <i>data mining</i>	11
2.2 Decision Tree.....	12
2.3 Decision Tree C4.5.....	15
2.4 Penelitian yang Relevan	16
<i>BAB III METODOLOGI PENELITIAN</i>	<i>21</i>
3.1 Identifikasi Masalah	22
3.2 Perumusan Masalah	22
3.3 Pengumpulan Data	22
3.4 Analisa	23
3.5 Hasil dan Pembahasan	23
3.6 Pengujian	23
3.7 Kesimpulan dan Saran.....	24

BAB IV ANALISA DAN PEMBAHASAN.....	25
4.1 Pengumpulan Data	25
4.2 Praproses	27
4.2.1 Data Cleaning	27
4.2.2 Data Integrasi.....	27
4.2.3 Data Selection.....	27
4.2.4 Data Transformation.....	28
4.3 Data Mining (Algoritma Decision Tree C4.5).....	30
4.4 Interpretasi / Evaluation	40
4.5 Pengujian	42
4.5.1 Pengujian untuk mendapatkan Atribut yang berpengaruh	42
4.5.2 Pengujian dengan menggunakan perbandingan Data Latih dan Data Uji	44
4.5.3 Kesimpulan Pengujian.....	51
4.5.4 Analisis Hasil Pengujian.....	52
BAB V KESIMPULAN DAN SARAN.....	53
5.1 Kesimpulan.....	53
5.2 Saran	54
DAFTAR PUSTAKA.....	vi

DAFTAR TABEL

Tabel 1. Penelitian yang Relevan	22
Tabel 2. Data Kelulusan Mahasiswa	32
Tabel 3. Data hasil Seleksi	34
Tabel 4. Kategori IPK	35
Tabel 5. Kategori Lama Studi	35
Tabel 6. Daya hasil transformasi	35
Tabel 7. Hasil Perhitungan Entrophy dan Gain	37
Tabel 8. Hasil Perhitungan Entrophy dan Gain untuk Node IPK Sedang	41
Tabel 9. Hasil Perhitungan Entrophy dan Gain untuk Node IPK Tinggi	43
Tabel 10. Kesimpulan Pengujian	57

DAFTAR GAMBAR

Gambar 1. Tahap-tahap Data mining	16
Gambar 2. Konsep Decision Tree	18
Gambar 3. Contoh Penggunaan Metode Decision Tree untuk Menentukan Jenis Buah	20
Gambar 4. Flowchart Metodologi Penelitian	27
Gambar 5. Pohon Keputusan	45
Gambar 6. Grafik Perbandingan Akurasi	58

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perguruan Tinggi atau dikenal juga dengan nama Universitas adalah sebagai lembaga ilmiah yang menyelenggarakan program pendidikan akademik atau profesional dalam sejumlah disiplin ilmu pengetahuan. Pengertian pendidikan menurut (UU Sisdiknas No.20 tahun 2003) adalah suatu usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya dan masyarakat.

Perguruan Tinggi atau Universitas merupakan salah satu penyelenggara pendidikan bagi mahasiswa. Perguruan Tinggi memiliki peranan yang besar dalam mengorganisir kualitas Pendidikan agar dapat menghasilkan sumber daya manusia yang kompeten dan berpengetahuan luas serta kreatif. Salah satu faktor yang menentukan kualitas suatu Perguruan Tinggi adalah tingginya tingkat keberhasilan mahasiswa dan rendah nya tingkat kegagalan mahasiswa. Tingkat keberhasilan mahasiswa dapat dilihat pada kelulusan mahasiswa yang tepat waktu, sedangkan kegagalan mahasiswa dilihat pada status mahasiswa yang *drop out* (DO). Semakin tinggi tingkat keberhasilan mahasiswa maka menjadi cerminan bahwa semakin bagus kualitas Perguruan Tinggi tersebut begitu juga sebaliknya. Pernyataan ini dikuatkan juga dalam evaluasi instrument penilaian akreditasi program studi dari BAN PT menyatakan bahwa persentase mahasiswa yang lulus tepat waktu adalah salah satu elemen penilaian akreditasi universitas [1]. Selain kelulusan tepat waktu, masalah lain yang mempengaruhi kualitas perguruan tinggi adalah masalah mahasiswa yang tidak aktif (alpha studi). Mahasiswa tidak aktif adalah mahasiswa yang tidak mendaftar pada awal semester atau tidak menghadiri kuliah selama setidaknya satu semester (alpa studi).

Universitas Islam Negeri Sultan Syarif Kasim Riau (UIN Suska Riau) adalah salah satu perguruan tinggi negeri di Riau yang memiliki total mahasiswa mencapai 30.000 mahasiswa. Dari seluruh total mahasiswa yang aktif terdapat mahasiswa yang tidak aktif. Keberadaan mahasiswa yang tidak aktif tentu mempengaruhi mahasiswa sehingga menjadi lulus tidak tepat waktu, semakin banyak siswa yang tidak aktif, semakin banyak siswa yang lulus tidak tepat waktu. Dalam Buku Pedoman Akademik UIN Suska Riau [11] dinyatakan bahwa pada akhir semester 2 dilakukan evaluasi tahap pertama untuk kelanjutan studi mahasiswa, dan pada akhir semester 8 dilakukan evaluasi tahap kedua yang dilaksanakan oleh Rektor. Mahasiswa program S1 yang memiliki IPK kurang dari 2,0 setelah dievaluasi dinyatakan gagal dan/atau *drop-out* (DO). Mahasiswa yang tidak melakukan pendaftaran dianggap tidak aktif (alpha studi) dua semester berturut-turut atau tidak, dikenakan sanksi akademik berupa pemutusan studi (*drop-out*).

Berdasarkan uraian di atas, siswa tidak aktif atau alpha studi dapat berpotensi menimbulkan masalah kelulusan yang tidak tepat waktu dan berpotensi *drop out* yang dapat menurunkan kualitas pendidikan dan penilaian akreditasi Perguruan Tinggi. Tidak adanya informasi yang terkait dengan faktor-faktor yang penyebab mahasiswa yang tidak aktif sehingga sampai terancam *drop out* dan informasi serta data apa yang mempengaruhi mahasiswa dapat lulus tepat waktu. Untuk alasan ini, setiap universitas selalu mengevaluasi kinerja mahasiswanya. Hasil evaluasi disimpan dalam database akademik sehingga data dapat digunakan sebagai pendukung pengambilan keputusan oleh universitas. Sejumlah besar data ini membuka peluang untuk menghasilkan informasi yang bermanfaat bagi universitas.

Data science adalah ilmu atau teknik untuk mengeksplorasi dan mengekstraksi sekumpulan data atau database sehingga dari kumpulan data tersebut ditemukan model, bentuk atau pola serta wawasan baru yang dapat digunakan sebagai alat untuk pengambilan keputusan. Peran *data science* sangat penting di era *Big Data* saat ini. Di mana berbagai jenis data baik terstruktur dan tidak terstruktur dapat dianalisis dengan benar. *Data science* adalah kombinasi dari

inferensi data, pengembangan algoritmik, dan teknologi untuk memecahkan masalah analitik yang kompleks. Konsep *data science* lebih kepada bagaimana mengekstraksi (penggalian) atau analisis prediksi data yang akan disaring dan menemukan data yang benar untuk menghasilkan data yang akurat sesuai dengan data aktual.

Dalam hal mengekstraksi atau menggali informasi mahasiswa pada ukuran data yang besar tidak dapat dilakukan dengan mudah. Teknologi *data mining* adalah bidang penelitian interdisipliner yang intinya adalah persimpangan antara *machine learning*, statistik, dan basis data. Saat ini *data mining* sedang dikembangkan untuk digunakan untuk menyelesaikan masalah yang melibatkan pendidikan. Istilah *Machine Learning* dapat didefinisikan sebagai hasil dari penggunaan algoritma untuk menggunakan data, mempelajarinya dan kemudian memprediksinya. *Machine Learning* terdiri dari analisis statistik dan analisis prediktif yang digunakan untuk menemukan pola dan menangkap wawasan tersembunyi berdasarkan data yang dirasakan.

Beberapa penelitian yang berkaitan dengan ³ *data mining* telah digunakan sebagai upaya untuk meningkatkan kualitas Perguruan Tinggi [5] [6] [7]. *Data mining* adalah istilah yang digunakan untuk menggambarkan penemuan pengetahuan dalam suatu basis data. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang berguna dan pengetahuan terkait dari berbagai database besar [8]. Fungsi dari *data mining* adalah

a. Fungsi prediksi

Merupakan proses menemukan pola dari data dengan menggunakan beberapa variabel untuk memprediksi variabel lain dari tipe atau nilai yang tidak diketahui.

b. Fungsi deskripsi.

Merupakan suatu proses untuk menemukan karakteristik data yang penting dalam suatu basis data.

c. Fungsi klasifikasi.

Klasifikasi adalah proses untuk menemukan model atau fungsi untuk menggambarkan kelas atau konsep suatu data. Proses yang digunakan untuk menggambarkan data penting dan dapat memprediksi tren data masa depan.

d. Fungsi asosiasi.

Proses ini digunakan untuk menemukan hubungan yang terkandung dalam nilai atribut dari set data.

Salah satu Teknik *data mining* adalah teknik klasifikasi. Teknik klasifikasi adalah teknik pembelajaran untuk memprediksi nilai dari variabel kategori. Variabel yang digunakan dalam penelitian ini adalah data-data akademik mahasiswa yang tersimpan dalam database universitas seperti Nama, Nim, IP, IPK, nilai mata kuliah, status kelulusan yang terdapat dalam aplikasi I-Raise (*Integrated Academic Information System*), yaitu System Informasi Akademik UIN Suska Riau . Dalam penelitian ini untuk mendukung upaya universitas dalam meningkatkan kualitas mahasiswa, maka diperlukan suatu alat bantu atau *tools* yang dapat memberikan informasi kepada universitas. Untuk memberikan informasi kepada universitas atau pengambil keputusan berdasarkan begitu banyak data mahasiswa tentu tidak mudah karena harus mengelola data sebesar itu yang memiliki banyak *field* dan jumlah *record*. Teknologi *data mining* adalah salah satu alat untuk mengekstraksi data dalam database besar dan dengan spesifikasi kerumitan yang tinggi. Penelitian ini menggunakan teknik *data mining* yang akan menghasilkan informasi faktor apa saja yang mempengaruhi mahasiswa lulus tepat waktu dan mahasiswa lulus tidak tepat waktu yang bisa berujung kepada *drop out*. Setelah faktor- faktor yang paling mempengaruhi diketahui dengan cara menghitung korelasi antar variabel, maka faktor-faktor yang paling berpengaruh akan diambil menjadi variabel yang akan menjadi masukan untuk mempelajari suatu pola data agar dapat menghasilkan informasi berupa hasil klasifikasi apakah mahasiswa lulus tepat waktu atau tidak tepat waktu.

³ Penelitian ini akan menggunakan teknik klasifikasi *data mining*. Algoritma klasifikasi cukup populer adalah pohon keputusan (*decision tree*), naif bayes, dan jaringan saraf tiruan [10]. Beberapa penelitian menggunakan algoritma

klasifikasi *data mining* yang digunakan untuk memprediksi, antara lain *decision tree* (pohon keputusan), *neural network*, *naïve bayes*, *K-Nearest Neighbor* dan jaringan saraf tiruan [2] [3] [4]. Penelitian yang dilakukan [2] dengan melakukan uji klasifikasi algoritma perbandingan yaitu antara *logistik regresion*, *naive bayes*, jaringan saraf tiruan, dan pohon keputusan (*decision tree*) menunjukkan bahwa pohon keputusan (*decision tree*) telah kinerja tertinggi. Pohon keputusan terbukti memiliki kinerja yang baik dibandingkan dengan algoritma klasifikasi lain. Algoritma *decision tree* bekerja dengan membentuk pohon keputusan yang dapat disimpulkan aturan klasifikasi tertentu, salah satunya adalah algoritma C4.5. Penelitian terkait mengenai algoritma C4.5 pada kasus lain [17][18][19]. Algoritma *decision tree* ini memiliki beberapa keunggulannya adalah [12] :

- a. Mudah diinterpretasikan
- b. Tidak memerlukan biaya mahal saat membangun algoritma ini
- c. Dapat menggunakan data numerik dan kategorikal
- d. Proses pengambilan keputusan dapat dipahami dengan mudah
- e. Stabil dan cepat saat digunakan dalam set databesar.

Berdasarkan latar belakang diatas maka dalam penelitian ini akan dilakukan analisis data akademik untuk mengklasifikasi kelangsungan dan keberhasilan studi mahasiswa UIN Suska Riau memanfaatkan konsep *data mining* dengan menggunakan metode klasifikasi *decision tree* C4.5. Dengan adanya penelitian ini diharapkan dapat menghasilkan model yang dapat mendukung aplikasi akademik universitas yang sudah ada. Dengan adanya penelitian ini bagi Program Studi Teknik Informatika merupakan salah satu upaya meningkatkan kualitas lulusan mahasiswa, hasil prediksi dapat menjadi bahan untuk melakukan evaluasi terhadap hasil studi mahasiswa. Evaluasi dilakukan secara berkala pada setiap tahun ajaran, diberikan oleh Prodi kepada Dosen PA untuk memantau kualitas akademik mahasiwa, sebagai upaya untuk meningkatkan kualitas universitas melalui kualitas lulusan mahasiswa

Hasil penelitian ini diharapkan dapat menjadi potensi besar bagi suatu universitas untuk menentukan kebijakan strategis bagi UIN Suska Riau dalam mengantisipasi mahasiswa yang terprediksi tamat tidak tepat waktu atau

kemungkinan *drop out* dan memberikan perhatian khusus kepada mahasiswa yang terprediksi lulus tepat waktu. Hasil penelitian juga dapat memperlihatkan faktor yang paling mempengaruhi tingkat keberhasilan studi mahasiswa. Penelitian ini juga memberikan informasi penting terkait rencana strategis bagi universitas berdasarkan interpretasi dari hasil penelitian. Dengan adanya tindakan ini maka akan dapat membantu visi dan misi dari UIN Suska Riau yaitu sebagai penyelenggara pendidikan dan pengajaran untuk melahirkan sumber daya manusia yang berkualitas secara akademik dan profesional serta memiliki integritas pribadi sebagai sarjana muslim.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang penelitian ini, perumusan masalah yang diambil sebagai berikut: "Bagaimana menerapkan konsep *Data Science* dan *Machine Learning* melalui proses *Data Mining* dan *algoritma decision tree* untuk Klasifikasi kelangsungan dan keberhasilan studi mahasiswa dan melihat faktor yang paling mempengaruhi tingkat keberhasilan studi mahasiswa di UIN Suska Riau sebagai upaya untuk meningkatkan kualitas universitas melalui kualitas lulusan mahasiswa".

1.3 Batasan masalah

Keterbatasan masalah untuk penelitian ini meliputi:

- a. Penelitian ini mengambil sample data akademik dari program studi S1 Teknik Informatika Fakultas Sains dan Teknologi UIN Suska Riau.
- b. Data yang digunakan merupakan data kelulusan mahasiswa Teknik informatika dari tahun 2011 sampai tahun 2015
- c. Variabel masukan yang digunakan adalah Nim, IP tiap semester, SKS tempuh tiap semester, IPK dan status kelulusan.
- d. Keluaran yang akan dihasilkan adalah model / *rule* yang bisa digunakan untuk memprediksi status kelulusan mahasiswa dan mengetahui atribut yang paling berpengaruh terhadap kelulusan mahasiswa.

1.4 Tujuan penelitian

Berdasarkan rumusan masalah, tujuan dari penelitian ini adalah menerapkan konsep *Data Science* dan *Machine Learning* melalui proses *Data Mining* dan *algoritma decision tree* untuk Klasifikasi kelangsungan dan keberhasilan studi mahasiswa dan melihat faktor yang paling mempengaruhi tingkat keberhasilan studi mahasiswa di UIN Suska Riau sebagai upaya untuk meningkatkan kualitas universitas melalui kualitas lulusan mahasiswa.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

- a. Bagi penulis, penelitian ini berguna untuk menambah wawasan tentang *data science* dan *machine learning* dengan teknik *data mining* dan algoritma *decision tree*.
- b. Untuk UIN Suska Riau, sebagai upaya untuk meningkatkan kualitas universitas melalui kualitas lulusan mahasiswa.
- c. Hasil penelitian ini juga dapat digunakan sebagai referensi untuk penelitian lebih lanjut.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam laporan penelitian ini dibagi menjadi lima bab, yaitu:

- a. BAB I PENDAHULUAN
Bab ini berisi latar belakang, perumusan masalah, batasan masalah, manfaat penelitian, tujuan penelitian dan sistematika penulisan.
- b. BAB II TINJAUAN LITERATUR
Bab ini menjelaskan studi sebelumnya (tinjauan pustaka) dan teori yang mendukung penelitian (landasan teori).
- c. BAB III METODOLOGI PENELITIAN
Bab ini berisi metode dan langkah yang akan digunakan dalam penelitian.
- d. BAB IV PEMBAHASAN

Bab ini berisi hasil dan diskusi dari penelitian yang dilakukan. Diskusi ini berisi analisis data, langkah penyelesaian, desain dan analisis proses dan hasil.

e. **BAB V KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan dan saran untuk pengembangan penelitian lebih lanjut.

BAB II

TINJAUAN PUSTAKA

2.1 Data Mining

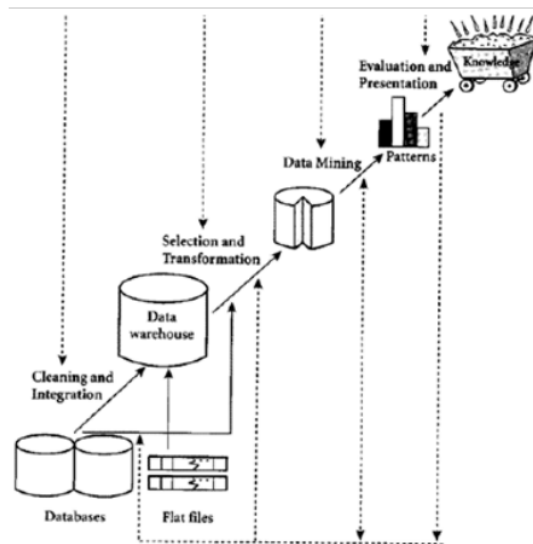
Data mining adalah proses pencarian pola-pola yang tersembunyi (*hidden pattern*) berupa pengetahuan (*knowledge*) yang tidak diketahui sebelumnya dari suatu sekumpulan data yang mana data tersebut dapat berada dalam *database*, dan *dataware*, atau media penyimpanan informasi yang lain. *Data mining* menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstrak dan mengidentifikasi informasi yang berguna dan pengetahuan terkait dari berbagai database besar [13].

Data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari data yang sangat besar [14]. Menurut Pramudiono (2007) *data mining* juga dapat disebut sebagai rangkaian proses untuk mengeksplorasi nilai tambah dalam bentuk pengetahuan yang belum diketahui secara manual dari kumpulan data. Menurut Davies (2004), *data mining* memiliki beberapa karakteristik, yaitu:

- a. *Data mining* berkaitan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang belum diketahui sebelumnya.
- b. *Data mining* biasanya menggunakan data yang sangat besar
- c. *Data mining* berguna untuk membuat keputusan penting, terutama dalam pembuatan strategi dan kebijakan.

2.1.1 Tahap – Tahap Data Mining

Istilah *data mining* dan *knowledge discovery in database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Walaupun sebenarnya *data mining* sendiri adalah bagian dari tahapan proses dalam KDD. Proses KDD secara garis besar dapat dilihat pada Gambar 1 berikut : [15]



Gambar 1. Tahap-tahap *Data mining*

- a. *Data cleaning*, untuk membersihkan data dari noise data dan data yang tidak konsisten. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data
- b. *Data integration*, mengkombinasikan atau mengintegrasikan beberapa sumber data.
- c. *Data selection*, mengambil data-data yang relevan dari database untuk dianalisis.
- d. *Data transformation*, mentransformasikan data *summary* ataupun operasi agregasi.
- e. *Data mining*, merupakan proses yang esensial dimana metode digunakan untuk mengekstrak pola data yang tersembunyi dengan menggunakan Teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
- f. *Interpretation / Evaluasi*, pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak

yang berkepentingan. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.1.2 Teknik data mining

Ada lima jenis teknik analisis yang dapat diklasifikasikan dalam *data mining*, yaitu : [14]

1. Asosiasi

Teknik asosiasi adalah teknik penambangan untuk menemukan aturan asosiatif antara kombinasi atribut. Contoh aturan asosiatif dari analisis pembelian di supermarket diketahui seberapa besar kemungkinan bagi pelanggan untuk membeli roti bersama dengan susu. Dengan pengetahuan ini, pemilik supermarket dapat mengatur penempatan barang-barang mereka atau merancang strategi pemasaran menggunakan kupon diskon untuk kombinasi barang tertentu.

2. Klasifikasi.

Klasifikasi adalah proses menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri dapat menjadi aturan "if-then", dalam bentuk pohon keputusan, rumus matematika atau jaringan saraf. Proses klasifikasi biasanya dibagi menjadi dua fase: pembelajaran dan pengujian. Dalam fase pembelajaran, beberapa data yang telah dikenal untuk kelas data diumpankan untuk membentuk model perkiraan. Kemudian pada tahap uji coba model yang telah terbentuk diuji dengan beberapa data lain untuk menentukan keakuratan model. Jika akurasi cukup, model ini dapat digunakan untuk prediksi kelas data yang tidak diketahui.

3. Pengelompokan (*Clustering*).

Tidak seperti association rule mining dan klasifikasi di mana kelas data telah ditentukan sebelumnya, pengelompokan melakukan pengelompokan data

tanpa berdasarkan pada kelas data spesifik. Bahkan pengelompokan dapat digunakan untuk memberi label pada kelas data tidak dikenal. Oleh karena itu pengelompokan sering diklasifikasikan sebagai metode pembelajaran yang tidak diawasi. Prinsip pengelompokan adalah untuk memaksimalkan kesamaan antara anggota satu kelas dan meminimalkan kesamaan antar kelompok. Clustering dapat dilakukan pada data yang memiliki beberapa atribut yang dipetakan sebagai ruang multidimensional.

4. Estimasi.

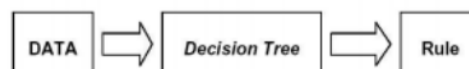
Estimasi hampir sama dengan klasifikasi, kecuali variable target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun dengan record lengkap menyediakan nilai dari variable target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variable target dibuat berdasarkan nilai variable prediksi

5. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan datang di masa mendatang.

2.2 Decision Tree

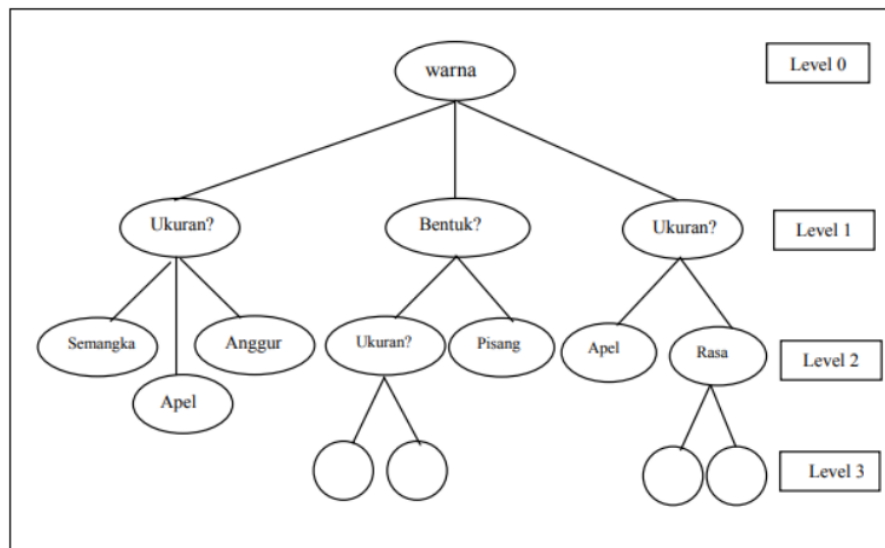
Decision Tree adalah salah satu metode klasifikasi yang kuat dan terkenal. Decision Tree mengubah fakta yang besar menjadi pohon keputusan yang merepresentasikan aturan, aturan tersebut dapat dengan mudah untuk diinterpretasikan oleh manusia. Decision Tree juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah variable input dengan sebuah variable target [16]. Konsep dari decision tree adalah mengubah data menjadi pohon keputusan yang akan menghasilkan aturan-aturan keputusan (*rule*)



Gambar 2. Konsep Decision Tree

Proses pada pohon keputusan adalah mengubah bentuk data yang biasanya berbentuk table dengan atribut dan record menjadi model pohon, lalu pohon akan dirubah menjadi *rule* , lalu menyederhanakan *rule* . Decision Tree memiliki kemampuan untuk membreak down proses pengambilan keputusan yang kompleks menjadi lebih simple sehingga memudahkan dalam memahami pengetahuan (*knowledge*).

2 Seringkali untuk mengklasifikasikan obyek, kita ajukan urutan pertanyaan sebelum bisa kita tentukan kelompoknya. Jawaban pertanyaan pertama akan mempengaruhi pertanyaan berikutnya dan seterusnya. Dalam decision tree, pertanyaan pertama akan kita tanyakan pada simpul akar pada level 0. Jawaban dari pertanyaan ini dikemukakan dalam cabang-cabang. Jawaban dalam cabang akan disusul dengan pertanyaan kedua lewat simpul yang berikutnya pada level 1. Dengan memperhatikan decision tree dalam Gambar 3 akan nampak ada 4 level pertanyaan. Dalam setiap level ditanyakan nilai atribut melalui sebuah simpul. Jawaban dari pertanyaan itu dikemukakan lewat cabang-cabang. Langkah ini akan berakhir di suatu simpul jika pada simpul tersebut sudah ditemukan kelas atau jenis obyeknya. Kalau dalam satu tingkat suatu obyek sudah diketahui termasuk warna Ukuran? Bentuk? Ukuran? Semangka Anggur Apel Ukuran? Pisang Apel Rasa Level 0 Level 1 Level 2 Level 3 dalam kelas tertentu, maka kita berhenti di level tersebut. Jika tidak, maka dilanjutkan dengan pertanyaan di level berikutnya hingga jelas ciri-cirinya dan jenis obyek dapat ditentukan [20].



Gambar 3. Contoh Penggunaan Metode Decision Tree untuk Menentukan Jenis Buah

Pada umumnya beberapa ciri kasus yang cocok untuk diterapkan metode *decision tree* adalah [20] :

1. Data dinyatakan dengan pasangan atribut dan nilainya. Misalnya atribut suatu data adalah temperatur dan nilainya adalah dingin. Dalam contoh atribut warna buah ada beberapa nilai yang mungkin yaitu hijau, kuning, merah.
2. Label/output data biasanya bernilai diskrit. Output ini bisa bernilai ya atau tidak, sakit atau tidak sakit, diterima atau ditolak. Dalam beberapa kasus mungkin saja outputnya tidak hanya dua kelas, tetapi penerapan *decision tree* lebih banyak untuk kasus *binary*.
3. Data mempunyai *missing value*. Misalkan untuk beberapa data, nilai dari suatu atributnya tidak diketahui. Dalam keadaan seperti ini *decision tree* masih mampu memberi solusi yang baik

Metode Decision Tree memiliki beberapa keunggulan dibandingkan metode lainnya untuk *database* yang besar, yaitu :

- a. Memiliki kecepatan yang relati cepat
- b. Dapat diubah menjadi rule klasifikasi dengan mudah dan sederhana

- c. Dapat menggunakan *query* SQL untuk mengakses *database*

2.3 Decision Tree C4.5

Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 [14]. Berikut adalah langkah-langkah algoritma C4.5 :

1. Pilih atribut sebagai akar

Pemilihan atribut sebagai akar didasarkan pada nilai *Gain* tertinggi dari seluruh atribut yang ada. Untuk menghitung nilai *Gain* digunakan Persamaan 2.1 berikut :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i) \quad (2.1)$$

Keterangan :

S = Himpunan kasus

A = Atribut

n = Jumlah partisi atribut A

|S_i| = Jumlah kasus pada partisi ke-i

|S| = Jumlah kasus dalam S

Sebelum mendapatkan nilai *Gain* , ada satu hal harus dilakukan, yaitu mencari nilai *Entropy*. *Entropy* digunakan untuk menentukan seberapa informatif sebuah input atribut untuk menghasilkan output atribut. Persamaan 2.2 menunjukkan persamaan untuk menghasilkan nilai *Entropy* :

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (2.2)$$

Keterangan :

S = Himpunan kasus

n = Jumlah partisi S

p_i = Proporsi dari S_i terhadap S

2. Buat cabang untuk tiap-tiap nilai
3. Bagi kasus dalam cabang

4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

2.4 Penelitian yang Relevan

Berikut ini adalah beberapa studi yang relevan dengan penelitian yang dibangun, yaitu:

Tabel 1. Penelitian yang Relevan

No	Nama (Tahun)	Judul	Deskripsi	Hasil
1	Ahmad, Sulistiyani & Saputra (2018)	Using Fuzzy K-Nearest Neighbor for Predicting University Students Graduation In Teknokrat	Variabel yang digunakan IP-1, IP-2, IP-3, IP-4, SPMB, Asal Sekolah, Jenis Kelamin dan Masa studi. Metode yang digunakan adalah Fuzzy-KNN	Akurasi prediksi sebesar 77,35%
2	Nugroho dan Sulistyono (2014)	Penerapan Algoritma C4.5 untuk Klasifikasi Predikat Kelulusan Mahasiswa fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta	Variabel yang digunakan Jurusan sekolah, Jenis kelamin, Asal sekolah, Rerata jumlah SKS per semester, rerata jumlah SKS, pernah menjadi asisten. Metode yang digunakan adalah C4.5	Variabel yang paling tinggi pengaruhnya terhadap predikat kelulusan adalah partisipasi mahasiswa menjadi asisten
3	Hastuti (2012)	Analisa Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif	Variabel yang digunakan adalah Jenis Kelamin, Usia saat mendaftar, Kota asal, Status domisili, Agama, Marital, Asal sekolah, Status kerja, Asal biaya, Pekerjaan orang	Hasil komparasi menunjukkan bahwa metode <i>decision tree</i> adalah metode yang paling akurat dibandingkan dengan

			tua, Penghasilan orang tua, IPS 1 – IPS 4, SKS 1 – SKS 4, dan Status skripsi. Metode data mining yang digunakan adalah <i>logistic regression, decision tree, naïve bayes, dan neural network.</i>	ketiga metode lainnya, yaitu dengan akurasi 95,29%
4	Al Amin, Indahwati & Anggraini	Analisis Ketepatan Waktu Lulus Berdasarkan Karakteristik Mahasiswa FEM dan FAPERTA Menggunakan Metode Chart	Variabel yang digunakan adalah Departemen, Minor, IPK, Jenis Kelamin, Asal Daerah, Asal Sekolah, Jalur Masuk, dan Beasiswa. Metode yang digunakan adalah Metode Chart	Model pohon klasifikasi FEM lebih baik dalam menduga mahasiswa lulus tepat waktu sedangkan model pohon klasifikasi FAPERTA lebih baik dalam menduga mahasiswa lulus tidak tepat waktu
5	Vera, Romero & Ventura (2011)	Predicting School Failure Using Data Mining	Vaiabel yang digunakan adalah 15 variabel yang berasal dari 77 variabel. Metode yang digunakan JRip, NNge, Prism, Rido, J48, SimpleCart, ADTree, Random Tree dan REPTree	Metode yang menghasilkan akurasi terbaik adalah Metode ADTree
6	Suhartinah & Erna (2010)	Graduation Prediction Of	Vaiabel yang digunakan adalah	Akurasi C.4.5 adalah

		Gunadarma University Students Using Naïve Bayes Algorithm And C4.5 Algorithm	data NEM, IP DNS semester 1, IP DNS semester 2, IPK DNU semester 1-2, gaji orang tua, pekerjaan orang tua. Metode yang digunakan adalah Metode <i>Naïve Bayes</i> dan Metode C4.5	85,7% sedangkan Naïve Bayes adalah 80,85
7	Ernawati (2008)	Prediksi Status Keaktifan Studi Mahasiswa dengan Algoritma C5.0 dan K-Nearest Neighbor	Variabel yang digunakan adalah data identitas mahasiswa dan data IPK mahasiswa pada tiga program studi tahun angkatan 2000 – 2005. Metode yang digunakan C5.0 dan K-Nearest Neighbor	Rata-rata keberhasilan algoritma C5.0 dan KNN mencapai akurasi di atas 90%. Namun algoritma C5.0 tetap dianggap sebagai algoitma yang sangat membantu dalam melakukan klasifikasi karena karakteristik data yang diklasifikasi dapat diperoleh dengan jelas baik dalam bentuk struktur pohon keputusan maupun aturan if-then
8	Kusrini &	Implemetatio	Membahas	Aplikasi

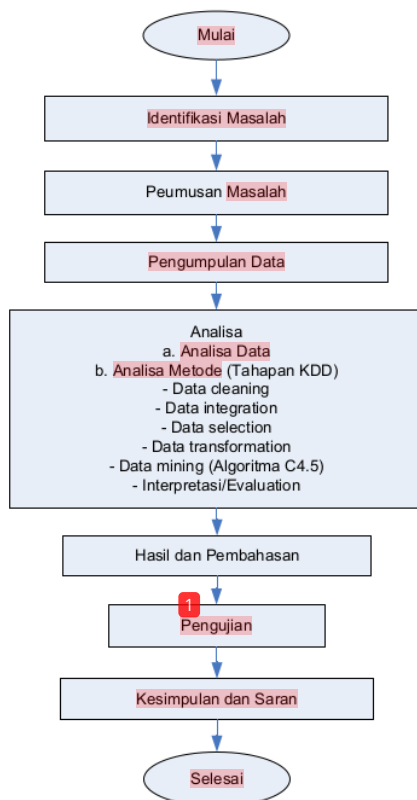
	Hartati (2007)	n Of C4.5 Algorithm To Evaluate The Cancellation Possibility O New Student Applicants At STEMIK Amikom Yogyakarta	tentang pembatalan calon mahasiswa baru. Calon mahasiswa yang telah berhasil lulus tes masuk membatalkan keputusan untuk masuk universitas dengan mengabaikan tahap berikutnya dari poses penerimaan mahasiswa baru yaitu pendaftaran ulang. Jika pendaftaran ulang dapat dideteksi dini, maka manajer eksekutif dapat membuat setiap upaya untuk menjaga calon mahasiswa baru melalui proses penerimaan dan selanjutnya mengurangi tingkat pembatalan masuk. Metode yang digunakan adalah Metode C4.5	yang dibangun dapat menghasilkan pohon keputusan yang sesuai dengan variabel dan data kasus yang dibeikan oleh user. Tingkat akuisi sangat tergantung pada variable terpilih yang akan menjadi dasar untuk membuat pohon keputusan
--	----------------	---	--	--

Beberapa penelitian di atas menjadi acuan dalam penelitian ini dalam menentukan variable masukan yang akan menjadi atribut pada metode yang digunakan yaitu metode *decision tree* C4.5. Metode *decision tree* C4.5 diterapkan karena berdasarkan beberapa penelitian terkait menunjukkan hasil pengenalan

yang sangat baik, bahkan lebih baik dibandingkan dengan beberapa metode klasifikasi lainnya.

BAB III METODOLOGI PENELITIAN

Pada bab ini akan dijelaskan tahap – tahap yang dilakukan dalam melakukan penelitian. Tahapan penelitian berguna agar pelaksanaan penelitian dapat berjalan dengan baik dan sistematis serta memenuhi tujuan yang diinginkan. Susunan tahapan penelitian dijelaskan pada gambar 4 *flowchart* dibawah ini :



Gambar 4. *Flowchart* Metodologi Penelitian

Dari *flowchart* Metode Penelitian pada gambar 4 terdapat beberapa proses penelitian diantaranya adalah :

3.1 Identifikasi Masalah

Pada tahapan ini dilakukan identifikasi terhadap masalah yang ditemukan pada saat penelusuran dalam kasus menerapkan konsep *data mining* dan algoritma *decision tree* untuk klasifikasi kelangsungan dan keberhasilan studi mahasiswa di UIN Suska Riau, sebagai upaya untuk meningkatkan kualitas universitas melalui kualitas lulusan mahasiswa. Adapun masalah yang teridentifikasi yaitu :

- a) Data akademik mahasiswa yang menumpuk
- b) Banyaknya mahasiswa Jurusan Teknik Informatika yang tidak lulus tepat waktu sehingga dapat menyebabkan kegagalan dalam menyelesaikan studi yang berujung kepada *drop out* (DO).
- c) Ketidaktahuan faktor-faktor apa saja yang mempengaruhi keberhasilan studi.
- d) Banyaknya mahasiswa yang gagal akan mempengaruhi akreditasi jurusan bahkan universitas.

3.2 Perumusan Masalah

Dari identifikasi masalah yang dilakukan maka ditemukan rumusan masalah Bagaimana menerapkan konsep *data mining* dan algoritma *decision tree*, untuk klasifikasi kelangsungan dan keberhasilan studi mahasiswa di UIN Suska Riau, sebagai upaya untuk meningkatkan kualitas universitas melalui kualitas lulusan mahasiswa.

3.3 Pengumpulan Data

Pengumpulan data merupakan tahap ketiga dalam metodologi penelitian. Pada tahap pengumpulan data ditentukan jenis data apa yang digunakan untuk penelitian, bahan-bahan yang akan dibutuhkan selama penelitian. Tahap pengumpulan data merupakan tahap yang penting dari suatu penelitian. Data yang digunakan untuk penelitian harus benar-benar akurat dan jelas sumbernya. Data yang digunakan merupakan Data Wisudawan Jurusan Teknik Informatika dari tahun 2011 sampai tahun 2015.

3.4 **Analisa**

Pada tahapan ini gambaran tentang penelitian yang akan dilakukan telah diketahui. Terdapat 2 proses analisa yang dilakukan, yaitu analisa terhadap data dan analisa terhadap tahapan KDD. Analisa data dilakukan terhadap data akademik mahasiswa Jurusan Teknik Informatika, dimana data mahasiswa yang digunakan adalah mahasiswa yang sudah lulus. Data wisudawan yang diambil mulai dari data lulusan periode pertama pada tahun 2004 sampai periode terakhir tahun 2019. Data tersebut akan melalui tahapan KDD yaitu *Data cleaning*, *Data Integration*, *Data Selection*, *Data Transformation*. Tahapan tersebut dilakukan bertujuan untuk mendapatkan data yang valid untuk dijadikan dan digunakan dalam penelitian ini. Pada tahapan KDD terjadi perubahan data sehingga menghasilkan atribut *dataset* yang baik untuk diteliti dengan tujuan untuk pemahaman terhadap isi *record* pada data akademis. Setelah tahapan tersebut dilakukan, selanjutnya akan masuk pada tahap Data Mining. Pada tahap ini akan menggunakan metode *Decision Tree C4.5*.

Untuk melihat faktor yang paling mempengaruhi keberhasilan studi pada penelitian ini digunakan korelasi antar variable menggunakan correlation matriks dengan algoritma association rules negative.

3.5 **Hasil dan Pembahasan**

Pada tahap ini akan diperoleh hasil dari penelitian. Klasifikasi dengan *decision tree* memberikan hasil berupa aturan-aturan klasifikasi dalam bentuk pohon keputusan, *if-then*, serta jumlah ketepatan (akurasi) dari data yang diklasifikasi. Hasil ini kemudian akan dilakukan pembahasan agar memberikan pengetahuan baru pada kasus yang diangkat.

3.6 **Pengujian**

Setelah dilakukan tahapan hasil dan pembahasan maka akan dilakukan pengujian dan menampilkan hasil dari penelitian yang telah dibuat. Tahap pengujian diperlukan sebagai ukuran bahwa penelitian dapat dijalankan sesuai

dengan tujuan. Pada penelitian ini akan dilakukan pengujian tingkat akurasi, yaitu mengukur tingkat kebenaran pengenalan dan performa metode Decision Tree C4.5. Pengujian dilakukan dengan mengubah-ubah perbandingan jumlah data latih dan data uji. Adapun split data yang akan diuji adalah 90:10, 80:20, 70:30, 60:40 dan 50:50 .

3.7 **Kesimpulan dan Saran**

Tahap ini berisikan tentang kesimpulan penelitian ini dan hasil yang didapatkan. Tahap ini juga berisikan hal yang disimpulkan dan disarankan penulis bagi pembaca untuk melakukan pengembangan terhadap penelitian ini kedepannya.

BAB IV

ANALISA DAN PEMBAHASAN

Secara garis besar model yang dirancang pada penelitian ini terdiri dari 2 bagian utama, yaitu tahapan Praproses data dan tahapan data mining (Decision Tree C4.5). Langkah pertama yang harus dilakukan sebelum pembuatan model adalah mengumpulkan data penelitian. Data yang telah dikumpulkan akan dilakukan tahapan praproses. Data hasil praproses kemudian akan diolah dengan menggunakan metode Decision Tree C4.5 sehingga menghasilkan model berupa Pohon / *Tree*.

4.1 Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data kelulusan mahasiswa Teknik informatika dari tahun 2011 sampai tahun 2015. Data yang dikumpulkan berjumlah 632 record data. Data ini diambil dari Pusat Teknologi dan Pangkalan Data (PTIPD) Universitas Islam Negeri Sultan Syarif Kasim (UIN SUSKA) Riau. Data memiliki atribut NIM, IP Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, IP Semester 5, IP Semester 6, IP Semester 7, IP Semester 8, SKS Tempuh Semester 1, SKS Tempuh Semester 2, SKS Tempuh Semester 3, SKS Tempuh Semester 4, SKS Tempuh Semester 5, SKS Tempuh Semester 6, SKS Tempuh Semester 7, SKS Tempuh Semester 8, IPK, serta Kelulusan. Atribut - atribut tersebut akan dikelola dan digunakan sebagai masukan dan kelas keluaran pada metode Decision Tree. Berikut adalah data kelulusan mahasiswa yang diperoleh :

Tabel 2. Data Kelulusan Mahasiswa

NIM	IP SMTR 1	IP SMTR 2	IP SMTR 3	IP SMTR 4	IP SMTR 5	IP SMTR 6	IP SMTR 7	IP SMTR 8	SKS TEMP UH SMTR 1	SKS TEMP UH SMTR 2	SKS TEMP UH SMTR 3	SKS TEMP UH SMTR 4	SKS TEMP UH SMTR 5	SKS TEMP UH SMTR 6	SKS TEMP UH SMTR 7	SKS TEMP UH SMTR 8	IPK	KELULUSAN
11151100238	3,15	2,87	3,43	2,83	3,06	2,82	3,42	1,75	17	21	20	20	20	24	18	16	3,07	TEPAT WAKTU
11151100013	2,48	2,88	2,96	2,37	2,95	2,82	3,35	2,45	17	17	20	20	18	21	15	20	2,93	TEPAT WAKTU
11151100212	3,41	3,65	3,46	3,78	3,63	3,25	3,72	4	17	19	22	20	20	24	13	4	3,57	TIDAK TEPAT WAKTU
11151100109	3,23	3,23	3,33	2,9	3,42	2,97	3,22	1,81	17	17	22	21	20	24	13	16	3,18	TIDAK TEPAT WAKTU
11151100110	3,09	3,16	3,07	2,94	2,97	2,91	3,2	3,9	17	21	22	20	20	20	16	12	2,97	TEPAT WAKTU
11151100123	2,72	3,05	3,12	2,66	2,97	2,7	3,36	2,8	17	19	20	22	20	19	16	10	3,08	TIDAK TEPAT WAKTU
11151100132	3,01	3,31	3,13	2,78	2,77	2,69	3,41	3,87	17	21	20	20	20	19	11	7	3,12	TIDAK TEPAT WAKTU
11151100264	2,19	2,89	2,77	2,31	2,5	2,66	2,45	3,07	17	17	20	20	18	21	17	17	3	TIDAK TEPAT WAKTU
11151100268	2,96	3,28	3,32	2,78	3,36	3,05	3,29	3,79	17	19	23	22	20	24	13	10	3,07	TEPAT WAKTU
11151100352	2,29	2,55	2,48	2,29	2,78	2,65	2,93	3,06	17	17	20	17	17	18	13	20	3	TIDAK TEPAT WAKTU
11151101770	2,54	3,14	3,18	3,02	3,23	3,3	3,19	3,6	17	17	20	22	22	24	17	15	3,06	TEPAT WAKTU
11151100481	2,62	2,81	2,1	2,07	2,78	2,46	3,55	2,08	17	19	20	18	17	20	13	14	2,77	TIDAK TEPAT WAKTU
11151101805	2,89	3,14	3,27	2,52	3,01	2,05	2,48	1,58	17	19	20	22	20	21	17	12	2,97	TIDAK TEPAT WAKTU
11151100532	2,82	2,74	2,87	2,54	2,85	0	2,61	2,93	17	19	24	19	18	0	14	13	2,92	TIDAK TEPAT WAKTU
...
11151101738	1,79	2,41	2,62	2,01	3,28	2,7	1,84	2,26	17	17	18	20	17	24	17	14	2,93	TIDAK TEPAT WAKTU

Data yang sudah dikumpulkan selanjutnya akan dilakukan tahap praproses dan tahap data mining yang ada di dalam tahapan KDD.

4.2 Praproses

4.2.1 Data Cleaning

Pada umumnya, data yang diperoleh, baik dari database suatu perusahaan maupun eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau hanya sekedar salah ketik. Data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya. Pembersihan data juga akan mempengaruhi performansi dari system data mining karena data yang akan ditangani akan berkurang jumlah dan kompleksitasnya. Berdasarkan pengecekan pada data kelulusan, tidak ditemukan data yang hilang dan tidak valid, sehingga tahapan cleaning data tidak dilakukan.

4.2.2 Data Integrasi

Tahap ini adalah mengkombinasikan atau mengintegrasikan beberapa sumber data. Integrasi data dilakukan pada atribut-attribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, nomor pegawai, tempat lahir, agama dan lain sebagainya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil data yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Data kelulusan tidak bersumber dari banyak sumber data, sehingga tidak perlu dilakukan proses integrasi data.

4.2.3 Data Selection

Proses pada tahap ini adalah menseleksi atribut data yang akan digunakan untuk proses data mining (klasifikasi). Hanya atribut data yang dibutuhkan dan

relevan saja yang akan digunakan sebagai atribut pada proses data mining (klasifikasi). Berdasarkan analisis maka total atribut yang digunakan dalam penelitian ini adalah 11 atribut yaitu Nim, Indeks Prestasi (IP) Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, Satuan Kredit Semester (SKS) Tempuh Semester 1, Sks Tempuh Semester 2, Sks Tempuh Semester 3, Sks Tempuh Semester 4, Indeks Prestasi Kumulatif (IPK) serta Kelulusan. Data selection dapat dilihat pada Tabel 3 berikut:

Tabel 3. Data hasil Seleksi

NIM	IP SMTR 1	IP SMTR 2	IP SMTR 3	IP SMTR 4	SKS TEMP UH SMTR 1	SKS TEMP UH SMTR 2	SKS TEMP UH SMTR 3	SKS TEMP UH SMTR 4	IPK	KELULUSAN
11151100238	3,15	2,87	3,43	2,83	17	21	20	20	3,07	TEPAT WAKTU
11151100013	2,48	2,88	2,96	2,37	17	17	20	20	2,93	TEPAT WAKTU
11151100212	3,41	3,65	3,46	3,78	17	19	22	20	3,57	TIDAK TEPAT WAKTU
11151100109	3,23	3,23	3,33	2,9	17	17	22	21	3,18	TIDAK TEPAT WAKTU
11151100110	3,09	3,16	3,07	2,94	17	21	22	20	2,97	TEPAT WAKTU
11151100123	2,72	3,05	3,12	2,66	17	19	20	22	3,08	TIDAK TEPAT WAKTU
11151100132	3,01	3,31	3,13	2,78	17	21	20	20	3,12	TIDAK TEPAT WAKTU
11151100264	2,19	2,89	2,77	2,31	17	17	20	20	3	TIDAK TEPAT WAKTU
11151100268	2,96	3,28	3,32	2,78	17	19	23	22	3,07	TEPAT WAKTU
11151100352	2,29	2,55	2,48	2,29	17	17	20	17	3	TIDAK TEPAT WAKTU
11151101770	2,54	3,14	3,18	3,02	17	17	20	22	3,06	TEPAT WAKTU
11151100481	2,62	2,81	2,1	2,07	17	19	20	18	2,77	TIDAK TEPAT WAKTU
11151101805	2,89	3,14	3,27	2,52	17	19	20	22	2,97	TIDAK TEPAT WAKTU
11151100532	2,82	2,74	2,87	2,54	17	19	24	19	2,92	TIDAK TEPAT WAKTU
...
11151101738	1,79	2,41	2,62	2,01	17	17	18	20	2,93	TIDAK TEPAT WAKTU

4.2.4 Data Transformation

1 Tahap selanjutnya adalah mentransformasikan data yang akan digunakan dalam tahap data mining. Tahap ini berfungsi untuk mengubah data ke dalam bentuk yang sesuai dengan kebutuhan pada tahap data mining, sehingga lebih mudah dipahami. Pengubahan data terjadi pada SKS tempuh dan IPK yang akan

dikategorikan ke dalam bentuk range pada Tabel 4 dan Tabel 5 berikut :

Tabel 4. Kategori IPK

No	IPK	Hasil Transformasi
1	> 3.50	Tinggi
2	3.00 - 3.50	Sedang
3	< 3.00	Rendah

Tabel 5. Kategori Lama Studi

No	SKS Tempuh	Hasil Transformasi
1	> 17	Rendah
2	17-19	Sedang
3	20-21	Tinggi
4	22-24	Sangat Tinggi

1 Setelah melalui tahap transformasi dengan cara mengkategorikan atribut

SKS tempuh dan atribut IPK, maka hasil perolehan data dapat dilihat pada Tabel 6 berikut :

Tabel 6. Daya hasil transformasi

NIM	IP SMTR 1	IP SMTR 2	IP SMTR 3	IP SMTR 4	SKS TEMP UH SMTR 1	SKS TEMP UH MTR 2	SKS TEMP UH SMTR 3	SKS TEMP UH SMTR 4	IPK	KELULUSAN
11151100238	Sedang	Rendah	Sedang	Rendah	Sedang	Tinggi	Tinggi	Tinggi	Sedang	TEPAT WAKTU
11151100013	Rendah	Rendah	Rendah	Rendah	Sedang	Sedang	Tinggi	Tinggi	Rendah	TEPAT WAKTU
11151100212	Sedang	Tinggi	Sedang	Tinggi	Sedang	Sedang	Sangat Tinggi	Tinggi	Tinggi	TIDAK TEPAT WAKTU
11151100109	Sedang	Sedang	Sedang	Rendah	Sedang	Sedang	Sangat Tinggi	Tinggi	Sedang	TIDAK TEPAT WAKTU
11151100110	Sedang	Sedang	Sedang	Rendah	Sedang	Tinggi	Sangat Tinggi	Tinggi	Rendah	TEPAT WAKTU
11151100123	Rendah	Sedang	Sedang	Rendah	Sedang	Sedang	Tinggi	Sangat Tinggi	Sedang	TIDAK TEPAT WAKTU
11151100132	Sedang	Sedang	Sedang	Rendah	Sedang	Tinggi	Tinggi	Tinggi	Sedang	TIDAK TEPAT WAKTU
11151100264	Rendah	Rendah	Rendah	Rendah	Sedang	Sedang	Tinggi	Tinggi	Sedang	TIDAK TEPAT WAKTU
11151100268	Rendah	Sedang	Sedang	Rendah	Sedang	Sedang	Sangat Tinggi	Sangat Tinggi	Sedang	TEPAT WAKTU
11151100352	Rendah	Rendah	Rendah	Rendah	Sedang	Sedang	Tinggi	Sedang	Sedang	TIDAK TEPAT WAKTU
11151101770	Rendah	Sedang	Sedang	Sedang	Sedang	Sedang	Tinggi	Sangat Tinggi	Sedang	TEPAT WAKTU
11151100481	Rendah	Rendah	Rendah	Rendah	Sedang	Sedang	Tinggi	Rendah	Rendah	TIDAK TEPAT WAKTU

11151101805	Rendah	Sedang	Sedang	Rendah	Sedang	Sedang	Tinggi	Sangat Tinggi	Rendah	TIDAK TEPAT WAKTU
11151100532	Rendah	Rendah	Rendah	Rendah	Sedang	Sedang	Sangat Tinggi	Rendah	Rendah	TIDAK TEPAT WAKTU
...
11151101738	Rendah	Rendah	Rendah	Rendah	Sedang	Sedang	Sedang	Tinggi	Rendah	TIDAK TEPAT WAKTU

4.3 Data Mining (Algoritma Decision Tree C4.5)

Data mining adalah proses mengeksplorasi dan menganalisa data dalam jumlah yang besar yang bertujuan untuk menemukan suatu pola atau informasi yang menarik dari data yang tersimpan dalam jumlah yang besar dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Tahap ini merupakan inti dari tahapan KDD yang dilakukan untuk menganalisis data yang telah dibersihkan. Metode yang digunakan dalam proses data mining adalah Decision Tree C4.5.

Konsep dari decision tree adalah mengubah data menjadi pohon keputusan yang akan menghasilkan aturan-aturan keputusan (*rule*). Berikut adalah langkah-langkah pada Algoritma C4.5 :

1. Pilih atribut sebagai akar

Atribut yang dipilih menjadi akar adalah atribut yang memiliki Gain tertinggi. Nilai Entropy dan Gain dihitung berdasarkan Persamaan 2.1 dan Persamaan 2.2 berikut :

$$\begin{aligned} \text{Entropy Total} &= ((-56/568) \cdot \log_2(-56/568) + (-512/568) \cdot \log_2(-512/568)) \\ &= 0,45786 \end{aligned}$$

Entropy Atribut **IP SMTR 1** :

$$\begin{aligned} \text{Rendah} &= ((-12/256) \cdot \log_2(-12/256) + (-244/256) \cdot \log_2(-244/256)) \\ &= 0,27297086 \end{aligned}$$

$$\begin{aligned} \text{Sedang} &= ((-18/203) \cdot \log_2(-18/203) + (-185/203) \cdot \log_2(-185/203)) \\ &= 0,43201464 \end{aligned}$$

$$\text{Tinggi} = ((-26/109) \cdot \log_2(-26/109) + (-83/109) \cdot \log_2(-83/109))$$

$$= 0,7925907$$

$$\begin{aligned} \text{Gain Atribut IP SMTR} &= 0,45786 - ((256/568)*0,27297086 + \\ &\quad (203/568)*0,43201464 + \\ &\quad (109/568)*0,43201464) \\ &= 0,03498696 \end{aligned}$$

Untuk nilai Entrophy dan Gain Atribut yang lain, dapat dilihat pada Tabel 7 berikut ini :

Tabel 7. Hasil Perhitungan Entrophy dan Gain

	Atribut	Jumlah	TEPAT WAKTU	TIDAK TEPAT WAKTU	Entrophy	Gain
Total		568	56	512	0,46451494	
IP SMTR 1	Rendah	256	12	244	0,27297086	0,03498696
	Sedang	203	18	185	0,43201464	
	Tinggi	109	26	83	0,7925907	
IP SMTR 2	Rendah	216	2	214	0,07584151	0,06738195
	Sedang	280	33	247	0,52316976	
	Tinggi	72	21	51	0,87086447	
IP SMTR 3	Rendah	277	11	266	0,24096644	0,02944914
	Sedang	202	29	173	0,59349796	
	Tinggi	89	16	73	0,6795852	
IP SMTR 4	Rendah	347	21	326	0,32950066	0,02607878
	Sedang	176	22	154	0,54356444	
	Tinggi	45	13	32	0,86728162	
SKS TEMPUH SMTR 1	Rendah	4	0	4	0	0,00105849
	Sedang	564	56	508	0,46674338	
	Tinggi	0	0	0	0	
	Sangat Tinggi	0	0	0	0	
SKS TEMPUH SMTR 2	Rendah	31	2	29	0,34511731	0,01552903
	Sedang	198	9	189	0,26676499	
	Tinggi	339	45	294	0,56491415	

	Sangat Tinggi	0	0	0	0	
SKS TEMPUH SMTR 3	Rendah	7	0	7	0	0,01798613
	Sedang	60	2	58	0,2108423	
	Tinggi	200	12	188	0,32744492	
	Sangat Tinggi	301	42	259	0,58301942	
SKS TEMPUH SMTR 4	Rendah	27	2	25	0,38094659	0,00653543
	Sedang	93	4	89	0,25593004	
	Tinggi	223	26	197	0,51948447	
	Sangat Tinggi	225	24	201	0,48977901	
IPK	Rendah	147	0	147	0	0,06924171
	Sedang	360	36	324	0,46899559	
	Tinggi	61	20	41	0,91273416	
GAIN MAX						0,06924171

Berdasarkan Tabel 7, diperoleh nilai Gain tertinggi adalah nilai Gain pada atribut **IPK** dengan nilai **0,06924171**, sehingga atribut ini menjadi akar pada pohon yang terbentuk.

IPK

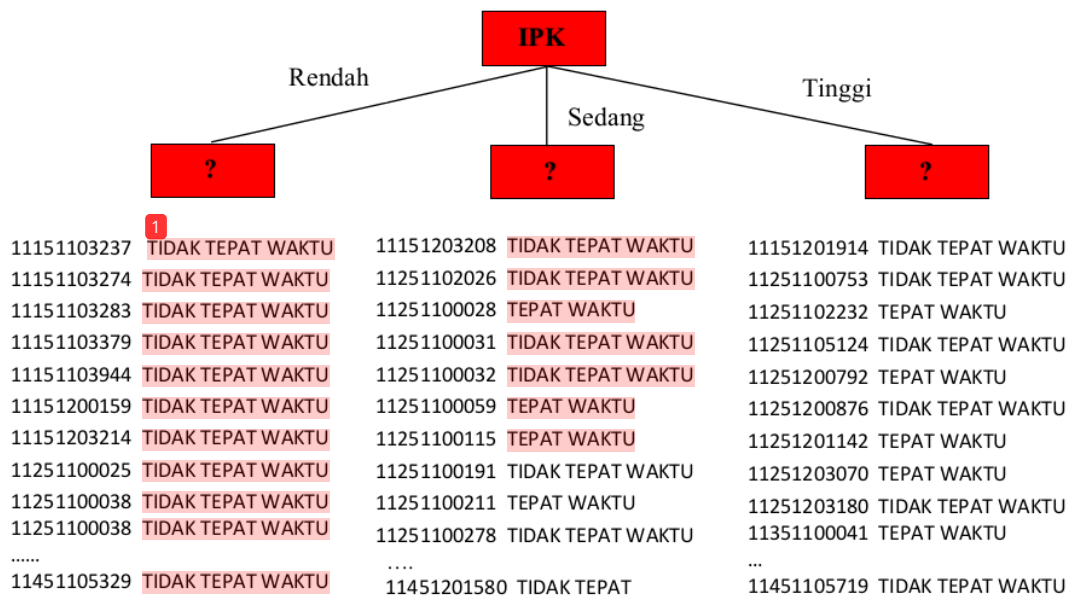
2. Buat cabang untuk tiap-tiap nilai

Setelah akar pohon dipilih, Langkah selanjutnya adalah menentukan cabang-cabang. Cabang-cabang merupakan instance dari atribut yang terpilih sebagai akar pada proses sebelumnya. Karena atribut yang terpilih sebagai akar adalah **IPK**, maka cabang-cabangnya adalah **Rendah**, **Sedang** dan **Tinggi**.

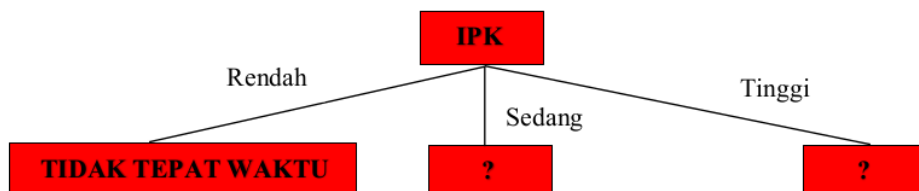


3. Bagi kasus dalam cabang

Pada tahap ini, kasus / data akan dibagi ke dalam cabang sesuai dengan atribut dan *instance* yang sesuai . Daftarkan semua kasus pada data dengan IPK Rendah, IPK Sedang dan IPK Tinggi.



Berdasarkan hasil pembagian kasus di atas, Mahasiswa dengan IPK Rendah semuanya lulus Tidak Tepat Waktu (semua data memiliki kelas yang sama), sehingga *node* pada cabang ini tidak dikembangkan lagi. Sedangkan untuk IPK Sedang dan IPK Tinggi, terdapat Mahasiswa dengan kelulusan **Tepat Waktu dan Tidak Tepat Waktu** (data memiliki kelas yang berbeda), sehingga *node* pada cabang ini perlu dikembangkan lagi.



4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Selanjutnya akan dicari node untuk IPK Sedang dan IPK Tinggi. Proses penentuan *node* dilakukan dengan dengan cara yang sama, yaitu dengan mencari Entrophy dan Gain. Atribut yang tersisa dari tahap sebelumnya akan dicari kembali Entrophy dan Gain nya. Atribut dengan Gain tertinggi akan terpilih sebagai cabang selanjutnya. Untuk mencari Entrophy dan Gain dilakukan dengan cara yang sama seperti perhitungan di atas.

Menentukan node untuk IPK Sedang :

$$\begin{aligned} \text{Entropy Total} &= ((-36/357) \cdot \log_2(-36/357) + (-321/357) \cdot \log_2(-321/357)) \\ &= 0,47165375 \end{aligned}$$

Entropy Atribut IP SMTR 1 :

$$\begin{aligned} \text{Rendah} &= ((-12/136) \cdot \log_2(-12/136) + (-124/136) \cdot \log_2(-124/136)) \\ &= 0,43055187 \end{aligned}$$

$$\begin{aligned} \text{Sedang} &= ((-15/163) \cdot \log_2(-15/163) + (-148/163) \cdot \log_2(-148/163)) \\ &= 0,44319161 \end{aligned}$$

$$\begin{aligned} \text{Tinggi} &= ((-9/58) \cdot \log_2(-9/58) + (-49/58) \cdot \log_2(-49/58)) \\ &= 0,62263432 \end{aligned}$$

$$\begin{aligned} \text{Gain Atribut IP SMTR 1} &= 0,47165375 - ((136/357) \cdot 0,43055187 + \\ &\quad (163/357) \cdot 0,44319161 + \\ &\quad (58/357) \cdot 0,62263432) \\ &= 0,00412412 \end{aligned}$$

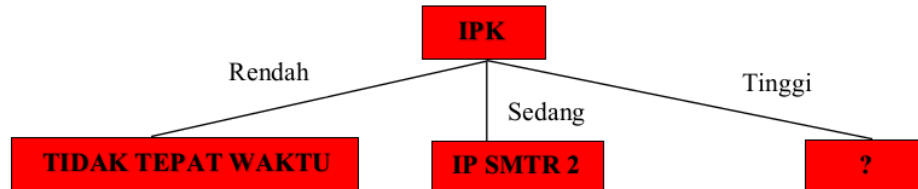
Untuk nilai Entrophy dan Gain Atribut yang lain, dapat dilihat pada Tabel 8 berikut ini :

Tabel 8. Hasil Perhitungan Entrophy dan Gain untuk Node IPK Sedang

	Atribut	Jumlah	TEPAT WAKTU	TIDAK TEPAT WAKTU	Entrophy	Gain
Total		357	36	321	0,47165375	
IP SMTR 1	Rendah	136	12	124	0,43055187	0,00412412
	Sedang	163	15	148	0,44319161	
	Tinggi	58	9	49	0,62263432	
IP SMTR 2	Rendah	95	2	93	0,14731067	0,03064007
	Sedang	234	27	207	0,51594693	
	Tinggi	28	7	21	0,81127812	
IP SMTR 3	Rendah	142	11	131	0,39318261	0,01116699
	Sedang	167	23	144	0,57824845	
	Tinggi	48	2	46	0,24988229	
IP SMTR 4	Rendah	201	21	180	0,48303068	0,00050425
	Sedang	141	14	127	0,46674338	
	Tinggi	15	1	14	0,35335934	
SKS TEMPUH SMTR 1	Rendah	1	0	1	0	0,00043019
	Sedang	356	36	320	0,47254722	
	Tinggi	0	0	0	0	
	Sangat Tinggi	0	0	0	0	
SKS TEMPUH SMTR 2	Rendah	17	0	17	0	0,01628188
	Sedang	117	7	110	0,32676598	
	Tinggi	223	29	194	0,55756115	
	Sangat Tinggi	0	0	0	0	
SKS TEMPUH SMTR 3	Rendah	3	0	3	0	0,01991415
	Sedang	35	1	34	0,18717626	
	Tinggi	123	7	116	0,31505695	
	Sangat Tinggi	196	28	168	0,59167278	

SKS TEMPUH SMTR 4	Rendah	14	2	12	0,59167278	0,0071288
	Sedang	53	2	51	0,23181305	
	Tinggi	142	16	126	0,50793536	
	Sangat Tinggi	148	16	132	0,49418293	
GAIN MAX						0,03064007

Berdasarkan Tabel 8, diperoleh nilai Gain tertinggi adalah nilai Gain pada atribut **IP SMTR 2** dengan nilai **0,03064007** , sehingga atribut ini menjadi node pada IPK Sedang.



Menentukan node untuk IPK Tinggi :

$$\begin{aligned} \text{Entropy Total} &= ((-19/58) \cdot \log_2(-19/58) + (-39/58) \cdot \log_2(-39/58)) \\ &= 0,91244118 \end{aligned}$$

Entropy Atribut **IP SMTR 1** :

$$\begin{aligned} \text{Rendah} &= ((-0/0) \cdot \log_2(-0/0) + (-0/0) \cdot \log_2(-0/0)) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Sedang} &= ((-3/9) \cdot \log_2(-3/9) + (-6/9) \cdot \log_2(-6/9)) \\ &= 0,91829583 \end{aligned}$$

$$\begin{aligned} \text{Tinggi} &= ((-16/49) \cdot \log_2(-16/49) + (-33/49) \cdot \log_2(-33/49)) \\ &= 0,91134238 \end{aligned}$$

$$\begin{aligned} \text{Gain Atribut IP SMTR 1} &= 0,91244118 - ((0/58) \cdot 0 + (9/58) \cdot 0,91829583 + \\ &\quad (49/58) \cdot 0,91134238) \\ &= 1,9819E-05 \end{aligned}$$

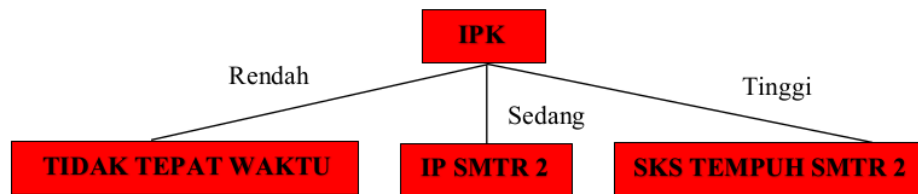
Untuk nilai Entrophy dan Gain Atribut yang lain, dapat dilihat pada Tabel 9 berikut ini :

Tabel 9. Hasil Perhitungan Entrophy dan Gain untuk Node IPK Tinggi

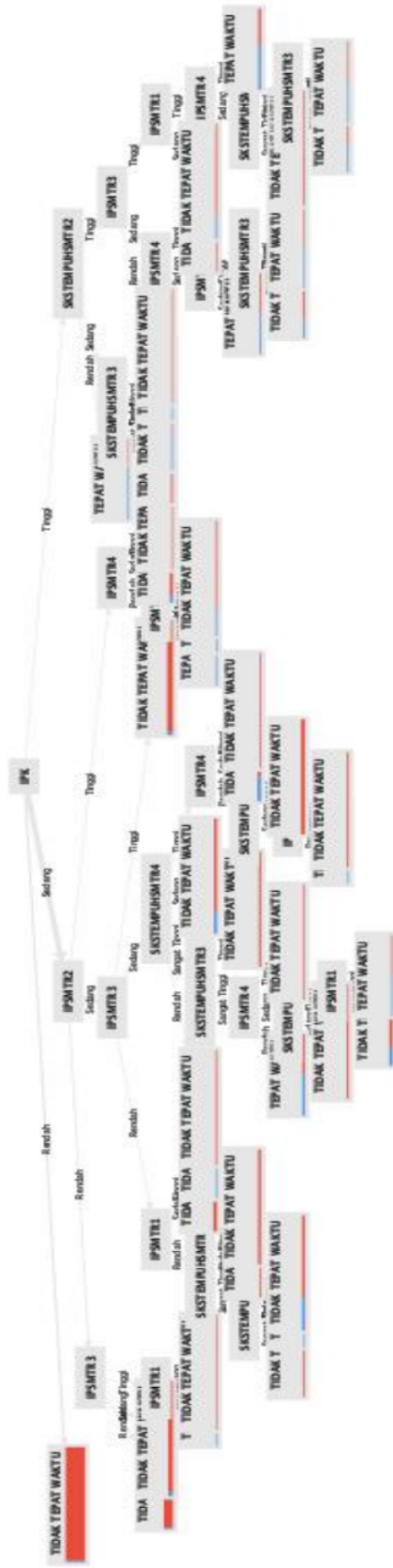
	Atribut	Jumlah	TEPAT WAKTU	TIDAK TEPAT WAKTU	Entrophy	Gain
Total		58	19	39	0,91244118	
IP SMTR 1	Rendah	0	0	0	0	1,9819E-05
	Sedang	9	3	6	0,91829583	
	Tinggi	49	16	33	0,91134238	
IP SMTR 2	Rendah	0	0	0	0	0,00276686
	Sedang	16	6	10	0,954434	
	Tinggi	42	13	29	0,89262301	
IP SMTR 3	Rendah	2	0	2	0	0,02022653
	Sedang	18	6	12	0,91829583	
	Tinggi	38	13	25	0,92681906	
IP SMTR 4	Rendah	1	0	1	0	0,01698554
	Sedang	28	8	20	0,86312057	
	Tinggi	29	11	18	0,95755348	
SKS TEMPUH SMTR 1	Rendah	1	0	1	0	0,00997803
	Sedang	57	19	38	0,91829583	
	Tinggi	0	0	0	0	
	Sangat Tinggi	0	0	0	0	
SKS TEMPUH SMTR 2	Rendah	3	2	1	0,91829583	0,02800185
	Sedang	10	2	8	0,72192809	
	Tinggi	45	15	30	0,91829583	
	Sangat Tinggi	0	0	0	0	
SKS TEMPUH SMTR 3	Rendah	0	0	0	0	0,01212378
	Sedang	3	1	2	0,91829583	
	Tinggi	11	5	6	0,99403021	
	Sangat Tinggi	44	13	31	0,87566339	

SKS TEMPUH SMTR 4	Rendah	1	0	1	0	0,02191478
	Sedang	2	2	0	0	
	Tinggi	23	9	14	0,96563613	
	Sangat Tinggi	32	8	24	0,81127812	
GAIN MAX						0,02800185

Berdasarkan Tabel 10, diperoleh nilai Gain tertinggi adalah nilai Gain pada atribut **SKS TEMPUH SMTR 2** dengan nilai **0,02800185** , sehingga atribut ini menjadi node pada IPK Tinggi.



Langkah selanjutnya adalah menentukan node untuk level selanjutnya. Proses dilakukan dengan cara yang sama. Setelah pengulangan proses dilakukan, maka dihasilkan Pohon / Tree yang seperti pada gambar di bawah ini :



Gambar 5. Pohon Keputusan

4.4 Interpretasi / Evaluation

Dalam tahap ini, merupakan hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi yang dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Hasil yang terbentuk dari Metode C4.5 adalah berupa Pohon keputusan. Gambar 5 di atas adalah model interpretasi yang terbentuk. Berdasarkan Pohon yang terbentuk, diperoleh *rule* sebagai berikut:

```
IPK = Rendah: TIDAK TEPAT WAKTU {TEPAT WAKTU=2, TIDAK TEPAT
WAKTU=149}
IPK = Sedang
|   IPSMTR2 = Rendah
|   |   IPSMTR3 = Rendah: TIDAK TEPAT WAKTU {TEPAT WAKTU=1, TIDAK
TEPAT WAKTU=66}
|   |   IPSMTR3 = Sedang: TIDAK TEPAT WAKTU {TEPAT WAKTU=1, TIDAK
TEPAT WAKTU=30}
|   |   IPSMTR3 = Tinggi
|   |   |   IPSMTR1 = Rendah: TEPAT WAKTU {TEPAT WAKTU=1, TIDAK
TEPAT WAKTU=1}
|   |   |   IPSMTR1 = Sedang: TIDAK TEPAT WAKTU {TEPAT WAKTU=0,
TIDAK TEPAT WAKTU=2}
|   |   IPSMTR2 = Sedang
|   |   |   IPSMTR3 = Rendah
|   |   |   |   IPSMTR1 = Rendah
|   |   |   |   |   SKSTEMPUHSMTR3 = Sangat Tinggi
|   |   |   |   |   SKSTEMPUHSMTR4 = Sangat Tinggi: TIDAK TEPAT
WAKTU {TEPAT WAKTU=0, TIDAK TEPAT WAKTU=5}
|   |   |   |   |   SKSTEMPUHSMTR4 = Sedang: TEPAT WAKTU {TEPAT
WAKTU=1, TIDAK TEPAT WAKTU=1}
|   |   |   |   |   SKSTEMPUHSMTR4 = Tinggi: TIDAK TEPAT WAKTU
{TEPAT WAKTU=4, TIDAK TEPAT WAKTU=10}
|   |   |   |   |   SKSTEMPUHSMTR3 = Sedang: TIDAK TEPAT WAKTU {TEPAT
WAKTU=0, TIDAK TEPAT WAKTU=4}
|   |   |   |   |   SKSTEMPUHSMTR3 = Tinggi: TIDAK TEPAT WAKTU {TEPAT
WAKTU=0, TIDAK TEPAT WAKTU=14}
|   |   |   |   IPSMTR1 = Sedang: TIDAK TEPAT WAKTU {TEPAT WAKTU=0,
TIDAK TEPAT WAKTU=22}
|   |   |   |   IPSMTR1 = Tinggi: TIDAK TEPAT WAKTU {TEPAT WAKTU=1,
TIDAK TEPAT WAKTU=2}
|   |   |   IPSMTR3 = Sedang
|   |   |   |   SKSTEMPUHSMTR4 = Rendah: TIDAK TEPAT WAKTU {TEPAT
WAKTU=0, TIDAK TEPAT WAKTU=4}
|   |   |   |   SKSTEMPUHSMTR4 = Sangat Tinggi
|   |   |   |   |   SKSTEMPUHSMTR3 = Sangat Tinggi
|   |   |   |   |   IPSMTR4 = Rendah: TEPAT WAKTU {TEPAT WAKTU=6,
TIDAK TEPAT WAKTU=6}
|   |   |   |   |   IPSMTR4 = Sedang
```

```

| | | | | | | SKSTEMPUHSMTR2 = Sedang: TIDAK TEPAT WAKTU
{TEPAT WAKTU=0, TIDAK TEPAT WAKTU=7}
| | | | | | | SKSTEMPUHSMTR2 = Tinggi 1
| | | | | | | IPSMTR1 = Sedang: TIDAK TEPAT WAKTU
{TEPAT WAKTU=3, TIDAK TEPAT WAKTU=16}
| | | | | | | IPSMTR1 = Tinggi: TEPAT WAKTU {TEPAT
WAKTU=2, TIDAK TEPAT WAKTU=2}
| | | | | | | IPSMTR4 = Tinggi: TIDAK TEPAT WAKTU {TEPAT
WAKTU=0, TIDAK TEPAT WAKTU=3}
| | | | | | | SKSTEMPUHSMTR3 = Tinggi: TIDAK TEPAT WAKTU {TEPAT
WAKTU=0, TIDAK TEPAT WAKTU=9}
| | | | | | | SKSTEMPUHSMTR4 = Sedang: TIDAK TEPAT WAKTU {TEPAT
WAKTU=3, TIDAK TEPAT WAKTU=13}
| | | | | | | SKSTEMPUHSMTR4 = Tinggi
| | | | | | | IPSMTR4 = Rendah
| | | | | | | SKSTEMPUHSMTR2 = Sedang 1
| | | | | | | IPSMTR1 = Rendah: TEPAT WAKTU {TEPAT
WAKTU=1, TIDAK TEPAT WAKTU=1}
| | | | | | | IPSMTR1 = Sedang: TIDAK TEPAT WAKTU {TEPAT
WAKTU=0, TIDAK TEPAT WAKTU=6}
| | | | | | | SKSTEMPUHSMTR2 = Tinggi: TIDAK TEPAT WAKTU
{TEPAT WAKTU=0, TIDAK TEPAT WAKTU=22}
| | | | | | | IPSMTR4 = Sedang: TIDAK TEPAT WAKTU {TEPAT
WAKTU=4, TIDAK TEPAT WAKTU=15}
| | | | | | | IPSMTR4 = Tinggi: TIDAK TEPAT WAKTU {TEPAT
WAKTU=0, TIDAK TEPAT WAKTU=3}
| | | | | | | IPSMTR3 = Tinggi: TIDAK TEPAT WAKTU {TEPAT WAKTU=1, TIDAK
TEPAT WAKTU=38}
| | | | | | | IPSMTR2 = Tinggi
| | | | | | | IPSMTR4 = Rendah 1
| | | | | | | IPSMTR1 = Rendah: TEPAT WAKTU {TEPAT WAKTU=2, TIDAK
TEPAT WAKTU=1}
| | | | | | | IPSMTR1 = Sedang: TEPAT WAKTU {TEPAT WAKTU=2, TIDAK
TEPAT WAKTU=0}
| | | | | | | IPSMTR1 = Tinggi: TIDAK TEPAT WAKTU {TEPAT WAKTU=1,
TIDAK TEPAT WAKTU=3}
| | | | | | | IPSMTR4 = Sedang: TIDAK TEPAT WAKTU {TEPAT WAKTU=1, TIDAK
TEPAT WAKTU=17}
| | | | | | | IPSMTR4 = Tinggi: TIDAK TEPAT WAKTU {TEPAT WAKTU=0, TIDAK
TEPAT WAKTU=2}
IPK = Tinggi
| | | | | | | SKSTEMPUHSMTR2 = Rendah: TEPAT WAKTU {TEPAT WAKTU=2, TIDAK
TEPAT WAKTU=1}
| | | | | | | SKSTEMPUHSMTR2 = Sedang 1
| | | | | | | SKSTEMPUHSMTR3 = Sangat Tinggi: TIDAK TEPAT WAKTU {TEPAT
WAKTU=0, TIDAK TEPAT WAKTU=6}
| | | | | | | SKSTEMPUHSMTR3 = Sedang: TIDAK TEPAT WAKTU {TEPAT WAKTU=1,
TIDAK TEPAT WAKTU=2}
| | | | | | | SKSTEMPUHSMTR3 = Tinggi: TEPAT WAKTU {TEPAT WAKTU=1, TIDAK
TEPAT WAKTU=1}
| | | | | | | SKSTEMPUHSMTR2 = Tinggi
| | | | | | | IPSMTR3 = Rendah: TIDAK TEPAT WAKTU {TEPAT WAKTU=0, TIDAK
TEPAT WAKTU=2}
| | | | | | | IPSMTR3 = Sedang
| | | | | | | IPSMTR4 = Sedang

```



```

| | | | IPSMTR2 = Sedang: TEPAT WAKTU {TEPAT WAKTU=2,
TIDAK TEPAT WAKTU=2}
| | | | IPSMTR2 = Tinggi
| | | | SKSTEMPUHSMTR3 = Sangat Tinggi: TIDAK TEPAT
WAKTU {TEPAT WAKTU=1, TIDAK TEPAT WAKTU=5}
| | | | SKSTEMPUHSMTR3 = Tinggi: TEPAT WAKTU {TEPAT
WAKTU=1, TIDAK TEPAT WAKTU=1}
| | | | IPSMTR4 = Tinggi: TIDAK TEPAT WAKTU {TEPAT WAKTU=0,
TIDAK TEPAT WAKTU=2}
| | | | IPSMTR3 = Tinggi
| | | | IPSMTR1 = Sedang: TIDAK TEPAT WAKTU {TEPAT WAKTU=1,
TIDAK TEPAT WAKTU=4}
| | | | IPSMTR1 = Tinggi
| | | | IPSMTR4 = Sedang
| | | | SKSTEMPUHSMTR4 = Sangat Tinggi: TIDAK TEPAT
WAKTU {TEPAT WAKTU=0, TIDAK TEPAT WAKTU=4}
| | | | SKSTEMPUHSMTR4 = Tinggi
| | | | SKSTEMPUHSMTR3 = Sangat Tinggi: TIDAK
TEPAT WAKTU {TEPAT WAKTU=1, TIDAK TEPAT WAKTU=3}
| | | | SKSTEMPUHSMTR3 = Tinggi: TEPAT WAKTU
{TEPAT WAKTU=1, TIDAK TEPAT WAKTU=1}
| | | | IPSMTR4 = Tinggi: TEPAT WAKTU {TEPAT WAKTU=7,
TIDAK TEPAT WAKTU=6}

```

4.5 Pengujian

Berdasarkan model berupa Pohon dan Rule yang terbentuk, maka bisa dilakukan proses pengujian.

4.5.1 Pengujian untuk mendapatkan Atribut yang berpengaruh

Berikut adalah fungsi pada python untuk mendapatkan atribut yang berpengaruh :

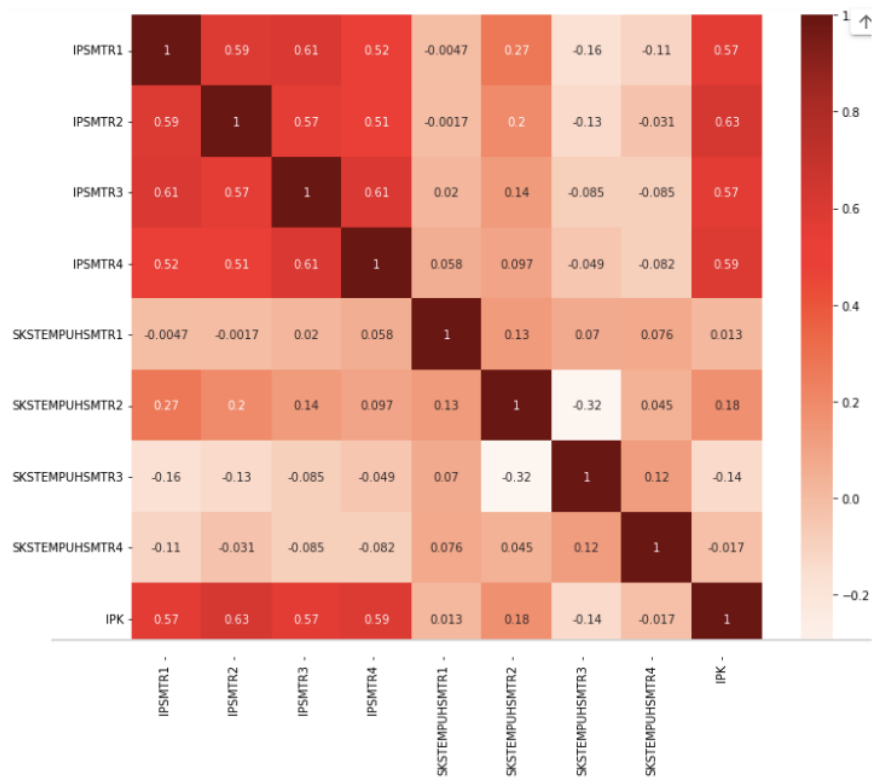
```

# MELIHAT KORELASI ANTAR VARIABEL
import matplotlib.pyplot as plt
import seaborn as sns

#Using Pearson Correlation
plt.figure(figsize=(12,10))
cor = X.corr()
sns.heatmap(cor, annot=True, cmap=plt.cm.Red)
plt.show()

```

Berdasarkan fungsi di atas, maka didapatkan hasil sebagai berikut :



Berdasarkan gambar di atas, maka Atribut yang berpengaruh pada kelangsungan dan keberhasilan studi mahasiswa prodi Teknik Informatika UIN SUSKA Riau adalah :

```
# LIHAT VARIABEL YANG BERPENGARUH
cor_target = abs(cor("IPK"))
relevant_features = cor_target[cor_target>0.1] #Variable dengan trashold korelasi > 0.1
relevant_features

IPSMTR1      0.570518
IPSMTR2      0.627394
IPSMTR3      0.569127
IPSMTR4      0.587603
SKSTEMPUHSMTR2 0.175936
SKSTEMPUHSMTR3 0.137077
IPK          1.000000
Name: IPK, dtype: float64
```

Berdasarkan gambar di atas, maka Atribut yang paling berpengaruh pada kelangsungan dan keberhasilan studi mahasiswa prodi Teknik Informatika UIN SUSKA Riau adalah Atribut IPK (Indeks Prestasi Kumulatif), diikuti dengan IP

Semester 2, IP Semester 4, IP Semester 1, IP Semester 3, SKS Tempuh Semester 2 dan SKS Tempuh Semester 3.

4.5.2 Pengujian dengan menggunakan perbandingan Data Latih dan Data Uji

Pada scenario ini, pengujian dilakukan dengan merubah-rubah jumlah data latih dan data ujinya, yaitu 50% Data Latih : 50% Data Uji (50:50), 60:40 , 70:30 , 80:20 , dan 90:10. Pengujian ini bertujuan untuk melihat apakah jumlah data latih berpengaruh terhadap kinerja Metode Decision Tree. Berikut adalah hasil pengujiannya :

1. Proporsi 50% Data Latih : 50% Data Uji

Berikut adalah fungsi untuk membagi data menjadi 50% Data Latih dan 50% Data Uji :

```
# PEMBAGIAN DATA
from sklearn.model_selection import train_test_split

# proporsi : Data Latih (50%) , Data Uji (50%)
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size = 0.5, random_state =10)
#X_train, X_test, y_train, y_test = train_test_split(X_new, y, train_size = 0.7, random_state =10)
print("Banyak data latih: ", len(X_train))
print("Banyak data uji : ", len(X_test))
```

Banyak data latih: 316
Banyak data uji : 316

Dengan proporsi ini, maka jumlah data latih dan data ujinya menjadi seimbang, yaitu 316 Data Latih dan 316 Data Uji. Berikut adalah hasil pelatihan dan akurasi dengan proporsi 50:50

```
# PELATIHAN MENGGUNAKAN DT
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics

dt = DecisionTreeClassifier(
    random_state=10,
    min_samples_split = 2,

    criterion='gini'
)

dt.fit(X_train, y_train)
y_pred = dt.predict(X_test)
score = metrics.accuracy_score(y_test, y_pred)
print("Akurasi dengan menggunakan Decision Tree: ", score)
```

Akurasi dengan menggunakan Decision Tree: 0.8291139240506329

Berdasarkan fungsi di atas, maka diperoleh akurasi 82,91%

Berikut adalah hasil confusion matrix nya :

```
# EVALUASI
conf_mat = metrics.confusion_matrix(y_test, y_pred)

print('Confusion Matrix:')
print(conf_mat)
print()

score = metrics.classification_report(y_test, y_pred)
print('Evaluasi Model Terbaik (Accuracy, Precision, Recall, F1-score):\n')
print(score)
```

Confusion Matrix:

```
[[ 6 26]
 [28 256]]
```

Evaluasi Model Terbaik (Accuracy, Precision, Recall, F1-score):

	precision	recall	f1-score	support
TEPAT WAKTU	0.18	0.19	0.18	32
TIDAK TEPAT WAKTU	0.91	0.90	0.90	284
accuracy			0.83	316
macro avg	0.54	0.54	0.54	316
weighted avg	0.83	0.83	0.83	316

Berdasarkan fungsi di atas maka diperoleh Confusion Matriks nya :

	TEPAT WAKTU	TIDAK TEPAT WAKTU
TEPAT WAKTU	6	26
TIDAK TEPAT WAKTU	28	256

$$Akurasi = \frac{6 + 256}{6 + 26 + 28 + 256} \times 100\% = 82,91\%$$

- Proporsi 60% Data Latih : 40% Data Uji

Berikut adalah fungsi untuk membagi data menjadi 60% Data Latih dan 40%

Data Uji :

```
# PEMBAGIAN DATA
from sklearn.model_selection import train_test_split

# proporsi : Data Latih (60%) , Data Uji (40%)
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size = 0.6, random_state =10)
#X_train, X_test, y_train, y_test = train_test_split(X_new, y, train_size = 0.6, random_state =10)
print("Banyak data latih: ", len(X_train))
print("Banyak data uji : ", len(X_test))
```

Banyak data latih: 379
Banyak data uji : 253

Dengan proporsi ini, maka jumlah data latih dan data ujinya menjadi 379 Data Latih dan 253 Data Uji. Berikut adalah hasil pelatihan dan akurasi dengan proporsi 60:40

```
# PELATIHAN MENGGUNAKAN DT
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics

dt = DecisionTreeClassifier(
    random_state=10,
    min_samples_split = 2,

    criterion='gini'
)

dt.fit(X_train, y_train)
y_pred = dt.predict(X_test)
score = metrics.accuracy_score(y_test, y_pred)
print("Akurasi dengan menggunakan Decision Tree: ", score)
```

Akurasi dengan menggunakan Decision Tree: 0.8695652173913043

Berdasarkan fungsi di atas, maka diperoleh akurasi 86,96%

Berikut adalah hasil confusion matriks nya :

```
# EVALUASI
conf_mat = metrics.confusion_matrix(y_test, y_pred)

print('Confusion Matrix:')
print(conf_mat)
print()

score = metrics.classification_report(y_test, y_pred)
print('Evaluasi Model Terbaik (Accuracy, Precision, Recall, F1-score):\n')
print(score)
```

Confusion Matrix:
[[5 21]
 [12 215]]

Evaluasi Model Terbaik (Accuracy, Precision, Recall, F1-score):

	precision	recall	f1-score	support
TEPAT WAKTU	0.29	0.19	0.23	26
TIDAK TEPAT WAKTU	0.91	0.95	0.93	227
accuracy			0.87	253
macro avg	0.60	0.57	0.58	253
weighted avg	0.85	0.87	0.86	253

Berdasarkan fungsi di atas maka diperoleh Confusion Matriks nya :

	TEPAT WAKTU	TIDAK TEPAT WAKTU
TEPAT WAKTU	5	21
TIDAK TEPAT WAKTU	12	215

$$\text{Akurasi} = \frac{5 + 215}{5 + 21 + 12 + 215} \times 100\% = 86,96\%$$

3. Proporsi 70% Data Latih : 30% Data Uji

Berikut adalah fungsi untuk membagi data menjadi 70% Data Latih dan 30%

Data Uji :

```
# PEMBAGIAN DATA
from sklearn.model_selection import train_test_split

# proporsi : Data Latih (70%) , Data Uji (30%)
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size = 0.7, random_state =10)
#X_train, X_test, y_train, y_test = train_test_split(X_new, y, train_size = 0.7, random_state =10)
print("Banyak data latih: ", len(X_train))
print("Banyak data uji : ", len(X_test))
```

Banyak data latih: 442
Banyak data uji : 190

Dengan proporsi ini, maka jumlah data latih dan data ujinya menjadi 442 Data Latih dan 190 Data Uji. Berikut adalah hasil pelatihan dan akurasi dengan proporsi 70:30

```
# PELATIHAN MENGGUNAKAN DT
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics

dt = DecisionTreeClassifier(
    random_state=10,
    min_samples_split = 2,

    criterion='gini'
)

dt.fit(X_train, y_train)
y_pred = dt.predict(X_test)
score = metrics.accuracy_score(y_test, y_pred)
print("Akurasi dengan menggunakan Decision Tree: ", score)
```

↳ Akurasi dengan menggunakan Decision Tree: 0.8789473684210526

Berdasarkan fungsi di atas, maka diperoleh akurasi 87,89%

Berikut adalah hasil confusion matriks nya :

```
# EVALUASI
conf_mat = metrics.confusion_matrix(y_test, y_pred)

print('Confussion Matrix:')
print(conf_mat)
print()

score = metrics.classification_report(y_test, y_pred)
print('Evaluasi Model Terbaik (Accuracy, Precision, Recall, F1-score):\n')
print(score)
```

Confussion Matrix:
[[1 20]
[3 166]]

Evaluasi Model Terbaik (Accuracy, Precision, Recall, F1-score):

	precision	recall	f1-score	support
TEPAT WAKTU	0.25	0.05	0.08	21
TIDAK TEPAT WAKTU	0.89	0.98	0.94	169
accuracy			0.88	190
macro avg	0.57	0.51	0.51	190
weighted avg	0.82	0.88	0.84	190

Berdasarkan fungsi di atas maka diperoleh Confusion Matriks nya :

	TEPAT WAKTU	TIDAK TEPAT WAKTU
TEPAT WAKTU	1	20
TIDAK TEPAT WAKTU	3	166

$$Akurasi = \frac{1 + 166}{1 + 20 + 3 + 166} \times 100\% = 87,89\%$$

4. Proporsi 80% Data Latih : 20% Data Uji

Berikut adalah fungsi untuk membagi data menjadi 70% Data Latih dan 30%

Data Uji :

```
# PEMBAGIAN DATA
from sklearn.model_selection import train_test_split

# proporsi : Data Latih (80%) , Data Uji (20%)
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size = 0.8, random_state = 10)
#X_train, X_test, y_train, y_test = train_test_split(X_new, y, train_size = 0.8, random_state = 10)
print("Banyak data latih: ", len(X_train))
print("Banyak data uji : ", len(X_test))
```

Banyak data latih: 505
Banyak data uji : 127

Dengan proporsi ini, maka jumlah data latih dan data ujinya menjadi 505 Data Latih dan 127 Data Uji. Berikut adalah hasil pelatihan dan akurasi dengan proporsi 80:20

```

▶ # PELATIHAN MENGGUNAKAN DT
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics

dt = DecisionTreeClassifier(
    random_state=10,
    min_samples_split = 2,

    criterion='gini'
)

dt.fit(X_train, y_train)
y_pred = dt.predict(X_test)
score = metrics.accuracy_score(y_test, y_pred)
print("Akurasi dengan menggunakan Decision Tree: ", score)

```

↳ Akurasi dengan menggunakan Decision Tree: 0.8818897637795275

Berdasarkan fungsi di atas, maka diperoleh akurasi 88,19%

Berikut adalah hasil confusion matriks nya :

```

▶ # EVALUASI
conf_mat = metrics.confusion_matrix(y_test, y_pred)

print('Confussion Matrix:')
print(conf_mat)
print()

score = metrics.classification_report(y_test, y_pred)
print('Evaluasi Model Terbaik (Accuracy, Precision, Recall, F1-score):\n')
print(score)

```

↳ Confussion Matrix:
[[0 12]
[3 112]]

Evaluasi Model Terbaik (Accuracy, Precision, Recall, F1-score):

	precision	recall	f1-score	support
TEPAT WAKTU	0.00	0.00	0.00	12
TIDAK TEPAT WAKTU	0.90	0.97	0.94	115
accuracy			0.88	127
macro avg	0.45	0.49	0.47	127
weighted avg	0.82	0.88	0.85	127

Berdasarkan fungsi di atas maka diperoleh Confusion Matriks nya :

	TEPAT WAKTU	TIDAK TEPAT WAKTU
TEPAT WAKTU	0	12
TIDAK TEPAT WAKTU	3	112

$$Akurasi = \frac{0 + 112}{0 + 12 + 3 + 112} \times 100\% = 88,19\%$$

5. Proporsi 90% Data Latih : 10% Data Uji

Berikut adalah fungsi untuk membagi data menjadi 70% Data Latih dan 30%

Data Uji :

Berikut adalah fungsi untuk membagi data menjadi 70% Data Latih dan 30%

Data Uji :

```
# PEMBAGIAN DATA
from sklearn.model_selection import train_test_split

# proporsi : Data Latih (90%) , Data Uji (10%)
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size = 0.9, random_state =10)
#X_train, X_test, y_train, y_test = train_test_split(X_new, y, train_size = 0.9, random_state =10)
print("Banyak data latih: ", len(X_train))
print("Banyak data uji : ", len(X_test))
```

Banyak data latih: 568
Banyak data uji : 64

Dengan proporsi ini, maka jumlah data latih dan data ujinya menjadi 568

Data Latih dan 64 Data Uji. Berikut adalah hasil pelatihan dan akurasi dengan

proporsi 90:10

```
# PELATIHAN MENGGUNAKAN DT
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics

dt = DecisionTreeClassifier(
    random_state=10,
    min_samples_split = 2,

    criterion='gini'
)

dt.fit(X_train, y_train)
y_pred = dt.predict(X_test)
score = metrics.accuracy_score(y_test, y_pred)
print("Akurasi dengan menggunakan Decision Tree: ", score)
```

Akurasi dengan menggunakan Decision Tree: 0.875

Berdasarkan fungsi di atas, maka diperoleh akurasi 87,5%

Berikut adalah hasil confusion matriks nya :

```
# EVALUASI
conf_mat = metrics.confusion_matrix(y_test, y_pred)

print('Confussion Matrix:')
print(conf_mat)
print()

score = metrics.classification_report(y_test, y_pred)
print('Evaluasi Model Terbaik (Accuracy, Precision, Recall, F1-score):\n')
print(score)
```

Confussion Matrix:
[[0 5]
 [3 56]]

Evaluasi Model Terbaik (Accuracy, Precision, Recall, F1-score):

	precision	recall	f1-score	support
TEPAT WAKTU	0.00	0.00	0.00	5
TIDAK TEPAT WAKTU	0.92	0.95	0.93	59
accuracy			0.88	64
macro avg	0.46	0.47	0.47	64
weighted avg	0.85	0.88	0.86	64

Berdasarkan fungsi di atas maka diperoleh Confussion Matriks nya :

	TEPAT WAKTU	TIDAK TEPAT WAKTU
TEPAT WAKTU	0	5
TIDAK TEPAT WAKTU	3	56

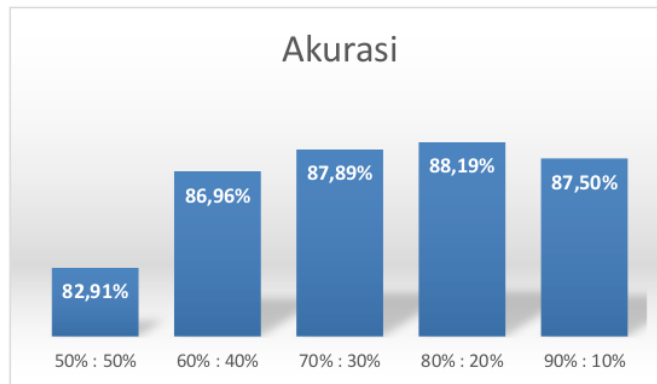
$$Akurasi = \frac{0 + 56}{0 + 5 + 3 + 56} \times 100\% = 87,5\%$$

4.5.3 Kesimpulan Pengujian

Berdasarkan scenario pengujian yang telah dilakukan, maka dapat diperoleh Tabel dan Grafik dibawah ini :

Tabel 10. Kesimpulan Pengujian

Perbandingan Data	Akurasi
50 : 50	82,91 %
60 : 40	86,96 %
70 : 30	87,89 %
80 : 20	88,19 %
90 : 10	87,5 %



Gambar 6. Grafik Perbandingan Akurasi

4.5.4 Analisis Hasil Pengujian

Beberapa hal yang dapat disimpulkan dari hasil pengujian yang telah dilakukan berdasarkan hasil analisis adalah sebagai berikut :

1. Berdasarkan pengujian, diperoleh atribut yang paling berpengaruh terhadap kelulusan mahasiswa prodi Teknik informatika adalah Atribut Indeks Prestasi Kumulatif (IPK).
2. Berdasarkan pengujian, diperoleh atribut yang tidak berpengaruh terhadap kelulusan mahasiswa prodi Teknik informatika adalah SKS Tempuh Semester 1 dan SKS Tempuh Semester 2. Hal ini disebabkan karena pada semester 1 dan 2, mahasiswa hanya bisa mengambil SKS sesuai dengan jumlah paket di kurikulum, artinya mahasiswa belum bisa mengambil SKS ke atas atau ke bawah.
3. Berdasarkan hasil pengujian, akurasi tertinggi berada pada perbandingan data latih dan data uji 80% : 20% yaitu 88,19%
4. Berdasarkan hasil pengujian, akurasi terendah berada pada perbandingan data latih dan data uji 50% : 50% yaitu 82,91%

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Bedasarkan hasil penelitian dan pembahasan yang mengacu pada tujuan penelitian ini, maka dapat disimpulkan bahwa:

1. Berhasil menerapkan konsep *Data Science* dan *Machine Learning* melalui proses *Data Mining* dan *algoritma decision tree* untuk Klasifikasi waktu studi mahasiswa dan melihat atribut yang paling mempengaruhi tingkat keberhasilan studi mahasiswa.
2. Berdasarkan pengujian, diperoleh atribut yang paling berpengaruh terhadap kelulusan mahasiswa prodi Teknik informatika adalah Atribut Indeks Prestasi Kumulatif (IPK).
3. Berdasarkan pengujian, diperoleh atribut yang tidak berpengaruh terhadap kelulusan mahasiswa prodi Teknik informatika adalah SKS Tempuh Semester 1 dan SKS Tempuh Semester 2. Hal ini disebabkan karena pada semester 1 dan 2, mahasiswa hanya bisa mengambil SKS sesuai dengan jumlah paket di kurikulum, artinya mahasiswa belum bisa mengambil SKS ke atas atau ke bawah.
4. Akurasi pengenalan tertinggi adalah 88,19%
5. Berdasarkan pengujian, perbandingan data latih dan data uji berpengaruh terhadap tingkat akurasi pengenalan Metode Decision Tree, semakin besar penggunaan data latih, semakin baik tingkat akurasi pengenalan.
6. Output penelitian dapat menjadi rekomendasi kepada pengelola Program Studi untuk mengevaluasi kinerja mahasiswanya sebagai upaya untuk meningkatkan kualitas lulusan mahasiswa.

5.2 Saran

Bedasarkan hasil penelitian dan pembahasan yang mengacu pada batasan penelitian ini, maka dapat disarankan bahwa:

1. Bagi peneliti perlunya kajian secara lebih mendalam agar model yang dikembangkan dapat memenuhi kebutuhan mahasiswa dan kebutuhan Jurusan dan penekanannya bukan pada produk semata, tetapi dapat dilihat dalam proses yang berlangsung
2. Hasil penelitian yang telah di lakukan dapat menjadi rujukan, kerangka kerja dan pedoman dalam melaksanakan pengembangan model klasifikasi lain nya.

DAFTAR PUSTAKA

- [1] BANPT, Buku I Naskah Akademik Akreditasi Instistusi Perguruan Tinggi. 2007.
- [2] K. Hastuti, “Analisis Komparasi Algoritma Klasifikasi *Data mining* Untuk Prediksi Mahasiswa Non-Aktif,” no. Semantik, 2012.
- [3] Ernawati, “Prediksi Status Keaktifan Studi Mahasiswa Dengan Algoritma C5.0 dan KNearest Neighbor,” 2008.
- [4] C. M. Vera, C. Romero, and S. Ventura, “Predicting School Failure Using *Data mining*,” J. Educ. Data Min., 2011.
- [5] Hui Yu ; Zhong-qiu Zhang, “The Application of *Data mining* Technology in Employment Analysis of University Graduates”, 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)
- [6] Anoopkumar M ; A. M. J. Md. Zubair Rahman, “A Review on *Data mining* techniques and factors used in Educational *Data mining* to predict student amelioration”, 2016 International Conference on *Data mining* and Advanced Computing (SAPIENCE), IEEE
- [7] Tismy Devasia ; Vinushree T P ; Vinayak Hegde “Prediction of students performance using Educational *Data mining*”, 2016 International Conference on *Data mining* and Advanced Computing (SAPIENCE), IEEE
- [8] Turban, E, 2005, *Decision Support Systems and Intelligent Systems Edisi Bahasa Indonesia Jilid 1*. Andi: Yogyakarta.
- [9] Halil Ibrahim Bulbul ; Özkan Unsal, “Comparison of Classification Techniques used in Machine Learning as Applied on Vocational Guidance Data”, 2017 10th International Conference on Machine Learning and Applications and Workshops
- [10] Ranjeet Devarakonda ; Michael Giansiracusa ; Jitendra Kumar, “Machine Learning and Social Media to Mine and Disseminate Big

Scientific Data”, 2018 IEEE International Conference on Big Data (Big Data)

- [11] Buku Panduan Akademik UIN Suska Riau, 2018
- [12] Gorunescu, *Data mining* Concept, Models and Techniques. Verlag Berlin Heidelberg: Springer, 2011.
- [13] Davies, and Paul Beynon, 2004, “Database System Third Edition”, Palgrave macmillan, New York.
- [14] Larose, D. T., 2006, “*Data mining* Methods And Models”, John Wiley & Sons, Inc., Hoboken, New Jersey.
- [15] Han, J. And Kamber, M, 2006, “*Data mining* Concept and Techniques Second Edition”, Morgan Kauffma, San Fransisco
- [16] Berry, Michael J.A dan Linoff, Gordon S.. “*Data mining* Techniques For Marketing, Sales, Customer Relationship Management Second Editon”. 2004United States of America: Wiley Publishing, Inc.
- [17] Nugroho, Yusuf Sulisty. “Penerapan Algoritma C4. 5 Untuk Klasifikasi Predikat Kelulusan Mahasiswa Fakultas Komunikasi Dan Informatika Universitas Muhammadiyah Surakarta”. 2014In , A1–6.
- [18] Suhartinah & Erna, “Graduation Prediction Of Gunadarma University Students Using Naïve Bayes Algorithm And C4.5 Algorithm”, 2010, Universitas Gunadharma, Jakarta
- [19] Kusrini & Hartati, “Implemetation Of C4.5 Algorithm To Evaluate The Cancellation Possibility O New Student Applicants At STEMIK Amikom Yogyakarta” 2009, Jurnal Dasi Issn: 1411- 3201, Vol. 10 No. 1, Maret 2009
- [20] Santosa B. 2007. *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis, Teori dan Aplikasi*. Graha Ilmu Yogyakarta.

LAPORAN_PENELITIAN_LPPM_-_2020.pdf

ORIGINALITY REPORT

10%

SIMILARITY INDEX

12%

INTERNET SOURCES

5%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1

repository.uin-suska.ac.id

Internet Source

6%

2

123dok.com

Internet Source

2%

3

docplayer.info

Internet Source

2%

Exclude quotes On

Exclude matches < 2%

Exclude bibliography On