

**PENERAPAN DATA MINING DALAM MENETUKAN MASA STUDI
MAHASISWA MENGGUNAKAN ALGORITMA C4.5**

SKRIPSI

WIRDANA RAUDA NINGSIH

71154061



**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUMATERA UTARA**

MEDAN

2020

**PENERAPAN DATA MINING DALAM MENETUKAN MASA STUDI
MAHASISWA MENGGUNAKAN ALGORITMA C4.5**

SKRIPSI

Diajukan untuk Memenuhi syarat Mencapai Gelar Sarjana Komputer

WIRDANA RAUDA NINGSIH

71154061



**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUMATERA UTARA**

MEDAN

2020

PERSETUJUAN SKRIPSI

Hal : Surat Persetujuan Skripsi

Lamp :-

Kepada Yth,

Dekan Fakultas Sains dan Teknologi

Universitas Islam Negeri Sumatera Utara Medan

Assalamu'alaikum Wr. Wb.

Setelah membaca, meneliti, memberikan petunjuk, dan mengoreksi serta mengadakan perbaikan, maka kami selaku pembimbing berpendapat bahwa skripsisaudara,

Nama : Wirdana Rauda Ningsih
Nomor Induk Mahasiswa : 71154061
Program Studi : Ilmu Komputer
Judul : Penerapan Data Mining Dalam Menentukan
Masa Studi Mahasiswa Menggunakan
Algoritma C4.5

dapat disetujui untuk segera *dimunaqasyahkan*. Atas perhatiannya kami ucapkan terimakasih.

Medan, 26 Oktober 2020 M
09 Rabiul Awal 1441H

Komisi Pembimbing,

Pembimbing I,



Dr. Mhd. Furqan, S.Si, M.Comp.Sc.
NIP. 198008062006041003

Pembimbing II,



Muhammad Ikhsan, S.T., M.Kom
NIP. 19830415201101 1 008

KEPAJANTARAN AGAMA REPUBLIK INDONESIA
UNIVERSITAS ISLAM NEGERI SUMATERA UTARA
MEDAN
MEDAN KEMAHSISWAAN DAN TEKNOLOGI
Jl. Lain No. 1 Medan 20131
Telp. (061) 4111111 / 4111112 / 4111113 / 4111114 / 4111115 / 4111116 / 4111117 / 4111118 / 4111119 / 4111120 / 4111121 / 4111122 / 4111123 / 4111124 / 4111125 / 4111126 / 4111127 / 4111128 / 4111129 / 4111130 / 4111131 / 4111132 / 4111133 / 4111134 / 4111135 / 4111136 / 4111137 / 4111138 / 4111139 / 4111140 / 4111141 / 4111142 / 4111143 / 4111144 / 4111145 / 4111146 / 4111147 / 4111148 / 4111149 / 4111150 / 4111151 / 4111152 / 4111153 / 4111154 / 4111155 / 4111156 / 4111157 / 4111158 / 4111159 / 4111160 / 4111161 / 4111162 / 4111163 / 4111164 / 4111165 / 4111166 / 4111167 / 4111168 / 4111169 / 4111170 / 4111171 / 4111172 / 4111173 / 4111174 / 4111175 / 4111176 / 4111177 / 4111178 / 4111179 / 4111180 / 4111181 / 4111182 / 4111183 / 4111184 / 4111185 / 4111186 / 4111187 / 4111188 / 4111189 / 4111190 / 4111191 / 4111192 / 4111193 / 4111194 / 4111195 / 4111196 / 4111197 / 4111198 / 4111199 / 4111200

SURAT PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan dibawah ini:

Nama : Wirdana Rauda Ningsih

Nim : 71154061

Program Studi : Ilmu Komputer

Judul : Penerapan Data Mining dalam Menentukan masa
Studi Mahasiswa Menggunakan Algoritma C4.5

Menyatakan dengan sebenarnya bahwa skripsi yang saya serahkan ini benar-benar merupakan hasil karya saya sendiri, kecuali kutipan-kutipan dari ringkasan- ringkasan yang semuanya saya jelaskan sumbernya.

Apabila kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil ciplakan, maka gelar dan ijazah yang diberikan oleh institut batal saya terima.

Medan, 26 Oktober 2020



Wirdana Rauda Ningsi
NIM. 71154061



KEMENTERIAN AGAMA REPUBLIK INDONESIA
UNIVERSITAS ISLAM NEGERI SUMATERA UTARA MEDAN
FAKULTAS SAINS DAN TEKNOLOGI
Jl. IAIN No. 1 Medan 20235
Telp. (061) 6615683-6622925, Fax. (061) 6615683
Url: <https://saintek.uinsu.ac.id>, E-mail: saintek@uinsu.ac.id

PENGESAHAN SKRIPSI

Nomor: 115/ST/ST.V.2/PP.01.1/01/2021

Judul : Penerapan Data Mining Dalam Menentukan
Masa Studi Mahasiswa Menggunakan
Algoritma C4.5
Nama : Wirdana Rauda Ningsih
NomorIndukMahasiswa : 71154061
ProgramStudi : Ilmu Komputer
Fakultas : Sains danTknologi

Telah dipertahankan di hadapan Dewan Penguji Skripsi Program Studi Ilmu
Komputer Fakultas Sains dan Teknologi UIN Sumatera Utara Medan dan
dinyatakan LULUS.

Padahari/ Tanggal : Senin 26 Oktober2020
Tempat/Media : Via Zoom Meeting

Tim Ujian Munaqasyah,

Ketua
Furqan

Dr. Mhd. Furqan, S.Si, M.Comp.Sc.
NIP. 198008062006041003

Dewan Penguji,

Penguji I,

Furqan

Dr. Mhd. Furqan, S.Si, M.Comp.Sc.
NIP. 198008062006041003

Penguji II,

Muhammad Iqbal

Muhammad Iqbal, S.I, M.Kom
NIP.198304152011011008

Penguji III,

Rakhmat Kurniawan

Rakhmat Kurniawan, ST., M.Kom
NIP. 198503162015031003

Penguji IV,

Yusuf Ramadhani Nasution

Yusuf Ramadhani Nasution, M.Kom
NIB. 1100000075

Mengesahkan,
Dekan Fakultas Sains dan Teknologi
UIN Sumatera Utara Medan,

Dr. H.M Jamil, M.A.
NIP. 196609101999031002

ABSTRAK

Institusi Perguruan tinggi dalam menentukan masa studi mahasiswa menggunakan algoritma data mining C4.5, agar tidak menimbulkan penimbunan data yang terjadi karena banyaknya jumlah data mahasiswa setiap tahunnya. Analisis dengan metode statistik sederhana sulit diterapkan sehingga perlu diterapkan metode klasifikasi teknik data mining. Data mining adalah proses yang menggunakan satu atau lebih teknik pembelajaran komputer (pembelajaran mesin) untuk menganalisis dan mengekstrak pengetahuan secara otomatis. Salah satu caranya adalah dengan menggunakan Algoritma C4.5 untuk suatu keputusan atau penyelesaian. Hal tersebut menjadi dasar bagi penulis dalam menentukan masa studi mahasiswa Program Studi Teknik Informatika berdasarkan data mahasiswa menggunakan data mining pada tugas akhir ini.

Tujuan dari penelitian ini adalah dengan menggunakan pohon keputusan berbasis algoritma C4.5 dan diimplementasikan ke suatu aplikasi yaitu RapidMiner diharapkan dapat meningkatkan keakuratan analisa masa studi mahasiswa. Pada penelitian dibahas dengan memantau hasil belajar di STMIK Budi Darma berupa nilai IPK dan Jumlah SKS yang belum akurat untuk menentukan seorang mahasiswa lulus tepat waktu atau tidak. Pada Penelitian ini untuk mengklasifikasikan kelulusan mahasiswa digunakan teknik data mining dengan algoritma C4.5 dan diimplementasikan ke Rapid Miner, hal tersebut bertujuan untuk melihat hasil perkembangan mahasiswa apakah dapat lulus tepat waktu atau tidak. Dari hasil penelitian terbukti bahwa algoritma C4.5 lebih akurat dibandingkan analisa yang dilakukan oleh analis mahasiswa. Hal ini dibuktikan dengan hasil evaluasi penelitian bahwa algoritma C4.5 mampu menganalisa tingkat ketepatan waktu mahasiswa menyelesaikan masa studinya.

Kata Kunci: Masa Studi, Mahasiswa, Data mining, Algoritma C4.5,

ABSTRACT

Higher education institutions in determining student study periods use the C4.5 data mining algorithm, so as not to cause data hoarding that occurs because of the large number of student data each year. Analysis with simple statistical methods is difficult to apply so it is necessary to apply classification methods of data mining techniques . Data mining is a process that employs one or more computer learning techniques (machine learning) to analyze and extract knowledge automatically. One of the methods is to use the C4.5 Algorithm for a decision or settlement. This becomes the basis for the author in determining the study period of the students of the Informatics Engineering Study Program based on student data using data mining in this thesis.

The purpose of this research is to use a decision tree based on the C4.5 algorithm and implement it into an application, namely Rapid Miner, which is expected to improve the accuracy of the analysis of the student's study period. In the study, it was discussed by monitoring the learning outcomes at STMIK Budi Darma in the form of inaccurate GPA and number of credits to determine whether a student graduated on time or not. In this study, to classify students' completeness, data mining techniques were used with the C4.5 algorithm and implemented to the Rapid Miner, which aims to see the results of student development whether they graduate on time or not. From the research results, it is proven that the C4.5 algorithm is more accurate than the analysis performed by student analysts. This is evidenced by the results of the research evaluation that the C4.5 algorithm is able to analyze the level of timeliness of students completing their studies.

Keywords: Study Period, Student, Data mining, C4.5 Algorithm,

KATA PENGANTAR

Puji dan syukur penulis ucapkan kepada ALLAH SWT yang telah memberikan nikmat berupa kesehatan, kesempatan kepada penulis sehingga mampu menyelesaikan Proposal skripsi ini.

Maka sehubungan dengan Proposal skripsi ini, segala bentuk hasil dituliskan kedalam bentuk Proposal skripsi ini sebagai persyaratan dalam menyelesaikan mata kuliah Proposal skripsi di Program Studi Ilmu Komputer. Tujuan utama dari penulisan Proposal skripsi ini adalah untuk memantapkan teori dan praktek yang telah dipelajari di kampus dan dapat diselesaikan dengan serta diaplikasikan.

1. Prof. Dr. Saidurrahman, M.Ag. Selaku Rektor Universitas Islam Negeri Sumatera Utara Medan yang telah memberikan kesempatan kepada penulis untuk pelaksanaan kegiatan penyusunan tugas akhir khusus tahun akademik 2019/2020.
2. Bapak Dr. H. M. Jamil, M.A selaku Dekan Fakultas Sains dan Teknologi.
3. Bapak Dr. Mhd. Furqan, S.Si.,M.Comp.Sc selaku Ketua Program Studi Ilmu Komputer sekaligus sebagai Dosen Pembimbing I penulis.
4. Bapak Muhammad Ikhsan, ST.,M.Kom selaku Dosen Pembimbing II penulis yang telah berkontribusi membantu dalam memberikan ide, saran, keritik, dan bimbingannya kepada penulis selama mengerjakan proposal skripsi ini.
5. Kedua orang tua yang sangat penulis sayangi dan cintai yang rela mengorbankan apapun demi kelancaran penulis dalam menyelesaikan pendidikan
6. Kepada saudara-saudari kandung yang penulis sayangi yang bayak memberikan semangat dan kepedulian yang luar biasa.
7. Kepada teman-teman sekelas Ilmu Komputer II yang telah memberikan semangat.
8. Dan semua pihak yang telah membantu penulis namun tidak dapat disebutkan satu persatu.

Penulis menyadari bahwa Proposal skripsi ini masih jauh dari kata sempurna. Untuk itu penulis mengharapkan saran perbaikan dari para penguji, pembimbing dan pembaca demi kesempurnaan dari Proposal skripsi ini. Demikian penyusunan Proposal

skripsi ini di tulis. Semoga Proposal skripsi ini dapat bermanfaat bagi pembacadan dapat digunakan dengan semestinya.

Medan, 26 Oktober 2020

Penulis,

WIRDANA RAUDA NINGSIH

NIM. 71154061

DAFTAR ISI

KATA PENGANTAR.....	iii
DAFTAR ISI.....	v
DAFTAR GAMBAR.....	viii
DAFTAR TABEL	vii
DAFTAR LAMPIRAN.....	viix
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian.....	6
1.5. Manfaat Penelitian.....	6
BAB II TINJAUAN PUSTAKA	7
2.1 Data Mining.....	7
2.1.1. Posisi Data Mining Dalam Berbagai Disiplin Ilmu	8
2.1.2 Operasi Data Mining	8
2.1.3 Pengolahan Data Mining	10
2.1.4 Penerapan Data Mining.....	11
2.2 Klsifikasi	13
2.3 <i>Decision Tree</i>	17
2.4 Algoritma C4.5.....	18
2.4.1 Prinsip Kerja Algoritma C4.5	16
2.5 Identifikasi.....	20
2.6 Rapid Miner.....	21

2.6	Penelitian Terdahulu	21
BAB III METODOLOGI PENELITIAN		21
3.1.	Tempat Dan Waktu	21
3.1.1	Tempat Penelitian	21
3.1.2	Waktu Dan Jadwal Pelaksanaan Penelitian	21
3.2.	Bahan dan Alat Penelitian	22
3.2.1	<i>Hardware</i> (Perangkat Keras).....	22
3.2.2	<i>Software</i> (Perangkat Lunak).....	22
3.3.	Teknik Pengumpulan Data	24
3.4	Penerapan Algoritma	24
3.5.	Analisa Kebutuhan	24
3.6.	Implementasi	25
3.7.	Pengujian.....	25
BAB IV HASIL DAN PEMBAHASAN		26
4.1.	Pembahasan	26
4.2.	Analisa Data	26
4.3.	Representasi Data.....	27
4.4.	Hasil Analisis Data.....	39
4.5.	Hasil Pengujian	47
4.6.	Pengujian.....	47
4.7.	Penerapan	44
BAB V KESIMPULAN DAN SARAN.....		45
5.1.	Kesimpulan.....	45
5.2.	Saran.....	45
DAFTAR PUSTAKA.....		Error! Bookmark not defined.

DAFTAR GAMBAR

Gambar	Judul Gambar	Halaman
Gambar 2.1	Posisi data mining diantara beberapa bidang ilmu	8
Gambar 2.2	Knowledge Discovery (KDD)	10
Gambar 2.3	Tampilan Welcome Perspective	18
Gambar 3.1	Algoritma C4.5	23
Gambar 4.1	Decision Tree node 1 penentuan masa studi mahasiswa.....	36
Gambar 4.2	Decision Tree node 1.1 penentuan masa studi mahasiswa.....	37
Gambar 4.3	Decision Tree node 1.2 penentuan masa studi mahasiswa.....	39
Gambar 4.4	Operator rapidminer	40
Gambar 4.5	Read Excel pada main process	41
Gambar 4.6	Import Configuration Wizard	41
Gambar 4.7	Set Role pada Main Process	42
Gambar 4.8	Decision tree algoritma C4.5 pada main process	42
Gambar 4.9	Graph Decision tree C4.5	43
Gambar 4.10	Hasil Role Decision Tree Algoritma C4.5	43

DAFTAR TABEL

Tabel	Judul Tabel	Halaman
Tabel 3.1	Waktu dan Jadwal Penelitian	21
Tabel 3.2	Perangkat Keras	22
Tabel 3.3	Perangkat Lunak.....	22
Tabel 4.1	Keterangan Data Alumni.....	27
Tabel 4.2	Data Alumni Tahun 2019	28
Tabel 4.3	Klasifikasi Atribut Input	31
Tabel 4.4	Atribut Target.....	32
Tabel 4.5	Data Transformasi	32
Tabel 4.6	Hasil Perhitungan Gain dan Entropy Pada tahap1	36
Tabel 4.7	Hasil Perhitungan Gain dan Entropy Pada tahap 1.1	37
Tabel 4.8	Hasil Perhitungan Gain dan Entropy Pada tahap 1.2	38

DAFTAR LAMPIRAN

Lampiran	Judul Lampiran
1.	Surat Keterangan Penelitian
2.	Daftar Riwayat Hidup
3.	Kartu Bimbingan Skripsi

BAB I PENDAHULUAN

1.1. Latar Belakang Masalah

Dalam Tasmara (2004), telah dijelaskan bahwasanya islam adalah agama yang sedang bberkmabang dan maju dalam bidang-bidang pendidikan maupun ilmu pengetahuan , yang kita ketahuai bahwasanya islam tidak hanya sebatas shalat dan ibadah saja tapi islam juga sebagai kaffa yang artinya (menyeluruh), dan agama selain itu juga menjadi komprehensif ,sumber inspiratif ataupun inpirasi bagi manusia dalam setiap kehidupan yang dijalannya. Hal ini dapat di lihat dari klasifikasi al qur'an yang di terapkan pada QS, Al-Muijaddilah ayat 11 yaitu :

Dalam QS. Al-Mujaadilah ayat 11, Allah SWT berfirman:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِذَا قِيلَ لَكُمْ تَفَسَّحُوا فِي الْمَجَالِسِ فَافْسَحُوا
يَفْسَحِ اللَّهُ لَكُمْ وَإِذَا قِيلَ انشُرُوا فَانشُرُوا يَرْفَعِ اللَّهُ الَّذِينَ
آمَنُوا مِنْكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ

Artinya: “Hai orang -orang yang beriman apabila dikatakan kepadamu : lapang-lapanglah dalam majlis “. maka niscaya allah akan memberikan kelapangan padamu, dan apabila di katakan : “ Berdirilah” maka engkau berdirilah , niscaya allah akan meninggikan orang yang beriman di antara mu dan orang akan memberi ilmu pengetahuan dan beberapa derajat dan allah selalu mengetahui apa yang sedang kamu kerjakan.”

Islam mengatakan , dalam mencari ilmu pengetahuan oleh seseorang ialah bukan karena sesuatu hal yang tidak diyakini melainkan menjadi kewajiban dan menjadi suatu hal yang wajar, sehingga semua muslim yang harus memiliki tanggung jawab, ini berbeda dengan para sifat -sifat yunani dan shopis yang dianggap bahwa pengetahuan adalah imajinasi atau hanya hayalan.

Pada bahasa arab, telah digambarkan sebuah pengetahuan yang beristilah Al- ilm, Al- Ma'rifah dan Al syu'ur. Maka pandangan terhadap dunia islam, yang pertama dan yang terpenting adalah salah satu sifat Allah SWT. Al ilm berasal dari akar kata ilm dan diambil pula dari kata alamah yang artinya tanda, simbol, atau lambang, yang dapat dikenal, tetapi alamah juga mempunyai arti pengetahuan, rencana, karakteristik, dan sebagai petunjuk. Karena ma;lam (jamak ma'alim) yang mempunyai arti petunjuk jalan atau yang menunjukkan sebuah diri seseorang. Hal ini sama dengan terhadap kata alam yang artinya sebagai petunjuk, setelah itu tanpa tujuan Al quran mempunyai ayat yang memiliki arti yang baik dan memiliki wahyu terhadap masalah ialah QS Al- Hasyr ayat 18

QS. Al- Hasyr Ayat 18

يَا أَيُّهَا الَّذِينَ آمَنُوا اتَّقُوا اللَّهَ وَانْتَظِرُوا نَفْسَ مَا قَدَّمْتُمْ لِغَدٍ
وَانْتَظِرُوا اللَّهَ إِنَّ اللَّهَ خَبِيرٌ بِمَا تَعْمَلُونَ

Artinya : Wahai orang-orang yang beriman! Bertaqwalah engkau kepada Allah SWT dan hendklah memperhatikan setiap orang atas perbuatannya di hari esok (akhirat), dan maka bertaqwalah engkau kepada allah sungguh Allah maha melihat dan teliti atas apa yang telah dikerjakan oleh umatnya

Mahasiswa adalah suatu aspek yang sangat diperlukan untuk menjadi patokan dalam keberhasilannya suatu penyelenggara dalam program bidang study, berdasarkan penilaian akreditasi pada program study berdasarkan matriks , Badan Akreditasi Nasional Perguruan Tinggi (2008) telah mengatakan bahwa suatu elemen penilaian Akreditasi Universitas berada pada persentase pada mahasiswa yang kelulusannya tepat pada waktunya.

Peraturan yang telah di tetapkan Permendikbud No. 49 tahun (2004) mengenai Standar Nasional Pendidikan Tertinggi (SN-PT) menetapkan bahwa, beban dalam belajar minimal mahasiswa strata S1/D1V adalah 144 SKS (Satuan

Kredit Semester) untuk menyelesaikan semua SKS mahasiswa yang telah diberikan batas waktu 4-5 tahun (8-10 semester), tetapi nyatanya tidak semua mahasiswa dapat menyelesaikan SKS tepat waktu sesuai jangka waktu yang telah ditetapkan, tidak sedikit pula mahasiswa yang menempuh studi S1 lebih dari batas waktu yang telah ditetapkan, atau bahkan sampai telah terancam Drop Out, adapun permasalahan keterlambatannya kelulusan mahasiswa saat ini adalah Program Study karena terbatasnya pihak Program Study untuk dapat melakukan prediksi masa study dari awal, dengan adanya prediksi dari dini tersebut pihak Program Study dapat melakukan, perencanaan, pengawalan study dan bimbingan yang lebih intensif pada mahasiswa yang telah terindikasi lulus tidak tepat waktu dan terancam Drop Out. (Eko Prasetyo,2012).

Prediksi ialah Sebuah proses yang diperkirakan akan terjadi di masa depan berdasarkan dari informasi masa lalu dan masa kini yang telah dimiliki. Masa study mahasiswa yang telah berkaitan dengan kelulusannya adalah suatu kejadian di masa depan yang biasanya di prediksi melalui data masa lalu dan riwayat akademik mahasiswa. (Andri,2013).

Berdasarkan masalah yang telah didapat, maka kita perlu membangun sebuah model untuk memprediksi masa studi mahasiswa, model tersebut sangat bisa membantu program study untuk membantu memprediksi masa study nya mahasiswa, proses itu dilakukan saat pertengahan masa study yang biasanya ketepatan dengan evaluasi di 2 tahun pertama yaitu di empat semester. di buku pedoman pendidikan strata s1 ataupun sarjana syarat untuk menyelesaikan masa kuliah adalah kelulusan yang mempunyai syarat kriteria penilaian seperti IPK minimal 2.00. Peraturan evaluasi data yang menggunakan atribut penelitian yaitu data akademik dari 1-4, biasa atribut yang digunakan adalah Indeks Prestasi Kumulatif (IPK), SKS tempuh, SKS Lulus, dan hasil Kelulusan, hasil prediksi telah diklarifikasikan menjadi 4 klarifikasi ialah Cepat, Tepat, Terlambat dan DO. Dengan ditetapkannya model prediksi masa study mahasiswa maka dapat membantu pihak Program Study dan para Dosen Pembimbing Akademik

sehingga penanganan dapat dilakukan lebih awal dan dapat meminimalisir kemungkinan mahasiswa akan Drop Out.

Pada model Matematika terdapat pendukung yang diprediksi pada masa study mahasiswa ialah algoritma *Decission Tree C4.5*. Algoritma ini termasuk sebuah data untuk mengembangkan algoritma ID3, *Decision Tree C4.5*, Algoritma tersebut adalah pengembangan dari algoritma ID3, dimana algoritma tersebut sebagai klasifikasi dan prediksi. Metode *Decission Tree* telah bekerja mengubah data fakta yang cukup besar sehingga muncullah sebuah keputusan yang akan di persentasikan, dengan aturan-aturan tertentu yang dapat di pahami dan mudah untuk di mengerti karena metode memakai bahasa alami, dan beberapa penelitian menggunakan algoritma C4.5. untuk mengklarifikasi dan memprediksi IPK dan kelulusan pada setiap mahasiswa. (Arif dan Hakim.2015)

Algoritma C4.5 algoritma yang telah di pakai membentuk pohon keputusan(*DecissionTree*).Pohon keputusan adalah sebuah metode mengklarifikasi sebuah keputusan dan memprediksi yang sudah terkenal. Pohon keputusan ini sangat berguna untuk mengesplor beberapa data, kemudian data juga dapat ditemukan data yang telah sembunyi diantara banyaknya calon variabel yang sudah di input sesuai dengan target. Algoritma yang biasa di pakai agar terbentuk pohon keputusan adalah : ID3, CART, dan C4.5, Algoritma C4.5 adalah pengembangan dari algoritma ID3, adapun proses pada pohon keputusan tersebut adalah mengubah bentuk data (tabel) menjadi model pohon, kemudian mengubah model Tree jadi rule, lalu kemudian disederhanakan rule. (Sofi Defianty.2016)

Decission Tree merupakan *flowchart* yang memiliki *Tree* (pohon), yang setiap simpulnya memakai atribut terhadap cabang , cara untuk memprentasekannya adalah, simpul bersifat seperti daun itu termasuk atribut kelas. *Decission Tree* memiliki jalan cerita dari akar nya sampai daunnya yang sudah menggapai prediksi kelas tersebut.

Data mining termasuk sebuah proses yang telah mempekerjakan satu atau lebih teknik sebagai pembelajaran komputer (*Machine Learning*) agar dapat menganalisis serta mengekstasi pengetahuan (*knowledge*) dengan otomatis. Antara lain adanya pembelajaran berbasis induksi (*Induction Based Learning*), yaitu pembentukan definisi secara umum yang biasanya dilakukan dengan cara observasi contohnya spesifik konsep yang mau dipelajari, adapun salah satunya ialah Algoritma C4.5 untuk pengambilan suatu keputusan dan penyelesaiannya. Hal tersebut menjadi dasar bagi penulis dalam menentukan masa studi Mahasiswa program studistara 1 Teknik Informatika berdasarkan data-data Mahasiswa dengan menggunakan data mining pada skripsi ini. (Heni Sulastri, 2017:299-305)

Pada Tujuan dari penelitian ini adalah mengetahui masa studi mahasiswa. Berdasarkan permasalahan di atas, maka penulis akan melakukan penelitian tentang **“Penerapan Data Mining dalam menentukan masa studi Mahasiswa Menggunakan Algoritma C4.5”**

1.2. Rumusan Masalah

Dengan mengacu pada uraian latar belakang masalah di atas, maka permasalahan yang dibahas dan diteliti adalah :

1. Bagaimana mengimplemantasikan algoritma C4.5 dalam menentukan masa studi mahasiswa?

1.3. Batasan Masalah

Berdasarkan permasalahan yang dikemukakan di atas maka batasan masalah pada penelitian ini adalah:

1. Penelitian dilakukan berdasarkan data mahasiswa yang sudah lulus dan data yang terdapat di program studi.
2. Pengambilan data pada Program Sarjana (S1) tahun 2018 sampai dengan tahun 2019 yang didapatkan dari pusat data dan informasi Universitas Budi Darma.

3. Mahasiswa cuti kuliah tidak diperhitungkan dalam penentuan estimasi masa studi mahasiswa
4. Aplikasi yang digunakan untuk pengujian algoritma C4.5 yaitu Rapid Miner.

1.4. Tujuan Penelitian

Dalam pembuatan laporan ini terdapat beberapa maksud dan tujuan yang hendak dicapai dan sesuai dengan yang diharapkan. Adapun Penelitian ini bertujuan untuk menerapkan:

1. Untuk meng implementasi Algoritma C4.5 agar dapat menentukan Study Mahasiswanya.

1.5 Manfaat Penelitian

Berdasarkan masalah yang telah di dapatkan adapun manfaat yang dicapai oleh peneliti ialah :

1. Dapat mengetahui kelemahan dan kelebihan pengambil keputusan untuk prediksi waktu masa studi mahasiswa
2. Dapat memprediksi mahasiswa sehingga berguna bagi prodi untuk meningkatkan akreditasi.
3. Diminimalisir oleh peneliti mahasiswa yang lulus tidak tepat waktu

BAB II TINJAUAN PUSTAKA

2.1 Data Mining

Data mining adalah tempat mengolah suatu pekerjaan satu atau lebih dalam teknik pembelajaran komputer. (*Machine Learning*) juga bermanfaat untuk menganalisis pengetahuan (*Knowledge*) secara otomatis. Pengertian lain adalah sebagai pembelajaran berbasis induksi (*induction-based learning*) yaitu sebagai pembentuk yang dilakukan dengan cara observasi seperti contohnya konsep konsep yang akan kita pelajari. (*Knowledge Discovery in Databases (KDD)*) ialah sebagai penerapan pada metode saintifik yang ada pada data mining, konteks, *data mining* adalah salah satu langkah dari proses KDD.

Data Mining adalah proses yang dapat menghasilkan sebuah informasi yang berguna sebagai basis data, *Data Mining* juga berfungsi sebagai penguat informasi yang dihasilkan dari data data besar yang akan membantu mengambil sebuah keputusan.

Data Mining adalah bidang yang dapat di dapat dari bidang ke ilmuwan agar dapat disatukan sebagai teknik pembelajaran mesin, serta mengenal pola, Statistik, Database, dan memvisualisasi untuk menangani masalah dalam mengambil sebuah informasi dari database yang besar. (Larose,2005)

Data Mining dapat dikelompokkan menjadi 2 yaitu (Tan et al, 2005)

1. *Descriptive Mining*, suatu proses menemukan sebuah karakter yang penting di dalam basis data, teknik ini termasuk *Descriptive Mining* seperti *Clustering Association, dan Sequential Mining*
2. *Predictive*, suatu proses menemukan sebuah pola dari data sehingga dapat digunakan untuk variabel lain pada masa depan, salah satu tekniknya dapat di dalam *Predictive Mining* adalah klasifikasi, Karena itulah fungsi satu ini dikatakan sebagai fungsi prediksi sama halnya dengan melakukan *predictive analisis*. Fungsi ini juga bisa digunakan untuk memprediksi sebuah variabel tertentu yang tidak ada dalam suatu data. Sehingga fungsi ini memudahkan dan

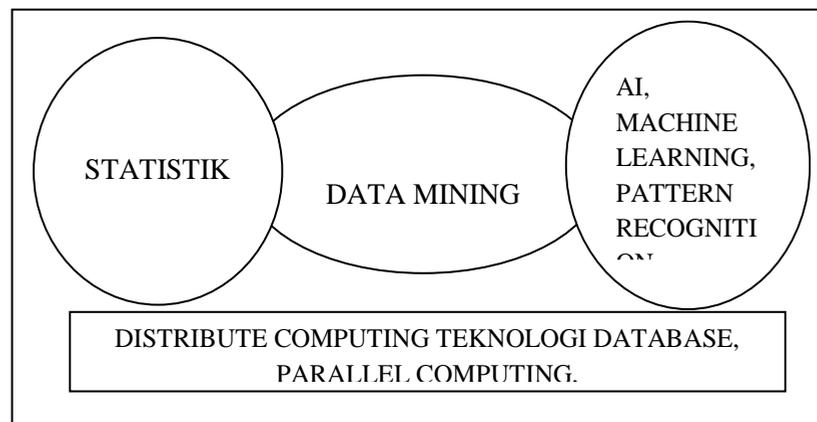
menguntungkan bagi siapapun yang memerlukan prediksi yang akurat untuk membuat hal penting tersebut menjadi lebih baik. Fungsi Data mining yang lainnya yaitu : characterization, discrimination, association, classification, clustering, outlier and trend analysis, dll.

- **Multidimensional concept description** Karakterisasi dan diskriminasi, Atau berfungsi untuk Menggeneralisasikan, meringkas, dan membedakan karakteristik data, dll.
- **Frequent patterns, association, correlation**
- **Classification and prediction**, Membangun model (fungsi) yang menggambarkan dan membedakan kelas atau konsep untuk prediksi masa depan. Misalnya, Mengklasifikasikan negara berdasarkan (iklim), atau mengklasifikasikan mobil berdasarkan (jarak tempuh gas)
- **Cluster analysis**, Membuat data grup untuk membentuk kelas baru. Misalnya, Memaksimalkan kesamaan intra-kelas & meminimalkan kesamaan antar kelas
- **Outlier analysis**, Objek data yang tidak sesuai dengan perilaku umum dari data, Berguna dalam deteksi penipuan, analisis peristiwa langka.
- **Trend and evolution analysis**, Trend dan penyimpangan: misalnya Analisis regresi atau Mining Penambangan pola berurutan: misalnya, Kamera digital, atau Analisis periodisitas dan Analisis berbasis kesamaan.
- **Other pattern-directed or statistical analyses.**

Teknik yang telah dibuat Data Mining bermaksud untuk mengetahui cara menelusur data yang ada dan dapat membangun model, dan kemudian modelnya di jadikan sebuah pola yang bisa dikenali pada pola lain dan tidak sedang berada di basis data yang tersimpan, Data mining juga dapat mengelompokkan sebuah data yang bertujuan pengguna tahu mengenai pola universal data yang sudah ada. Anomali data sangat perlu di deteksi supaya dapat diketahui tindak lanjut yang akan di kerjakan selanjunya, ini bertujuan sebagai pendukung kegiatan operasi terhadap perusahaan hingga tujuan akhir pada perusahaan tercapai.

2.1.1. Posisi Data Mining Dalam Berbagai Disiplin Ilmu

Para ahli telah berusaha untuk menentukan posisi- posisi di bidang *Data Mining* yang di antaranya sebagai bahasan dalam *Data mining* dengan bidang lainnya, walaupun tidak sampai seratus persen sama dengan bidang yang lain, tetapi mempunyai karakteristik yang sama dalam beberapa hal. Kesamaan yang ada di dalam metode *Data Mining* seperti dibidang Statistic adalah sampel estimasi dan pengujian hipotesis. Persamaan dan kecerdasan yang dibuat (*Artificial Intelligence*), mengenal pola (*Pattern Recognition*), dan pembelajaran mesin (*Machine Learning*) termasuk Algoritma pencarian, Teknik pemodelan, dan teori pembelajaran, seperti yang terlihat pada gambar.



Gambar 2.1 : Posisi data mining di antara beberapa bidang ilmu

Sumber: Hermawati, FajarAstuti. 2013

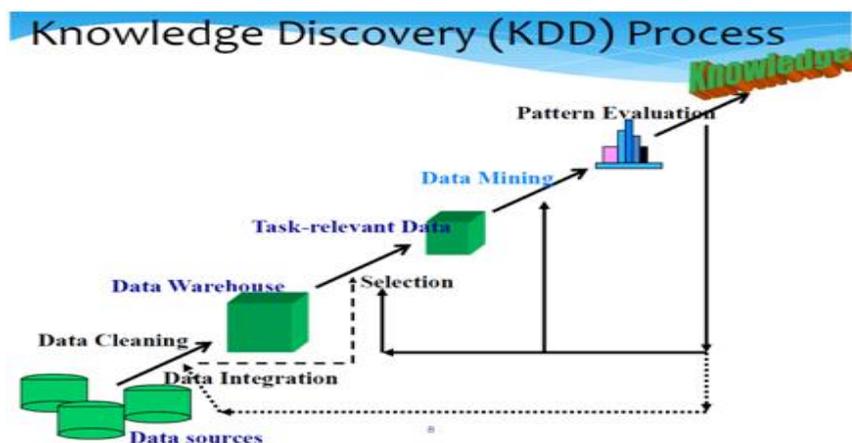
2.1.2 Operasi Data Mining

Operasi *Data Mining* terbagi menjadi dua sifat yaitu, Sifat prediksi (*Prediction Driven*) berfungsi untuk menyelesaikan pertanyaan dan sesuatu yang bersifat transparan. Operasi yang di prediksi mempalidasi sebuah hipotesis, *Querying idan pelapor* (Misalnya: *Spreadsheet* dan *Pivot Tabel*), analisis Multi Dimensi (*Dimensional Summary*), OLAP (*Online Analitic Processing*), dan analisis statistic. Penemuan (*discovery Driven*) ini mempunyai sifat yang terbuka atau Transparan yang biasa digunakan untuk menjawab sebuah pertanyaan

“Mengapa?”. Operasi yang digunakan dalam penemuan ini untuk menganalisis data serta mengeksplorasi, pemodelan prediktif, segmentasi database, analisis keterkaitan (*link Analysis*) serta mendeteksi deviasi. (Hermawati,Fajar Astuti.2013)

Adapun tahap-tahap dalam proses menggunakan *Data Mining* yang merupakan sebagai proses *Knowledge Discovery in Database* (KDD) ialah:

1. Mengetahui domain sebuah aplikasi agar yang menggunakan tahu bagaimana cara menggali pengetahuan dari sasaran pengguna
2. Membuat sasaran data utama menjadi pilihan data yang fokus pada sub- set pada data.
3. Membersihkan serta mentrasfer data melalui eliminasi *outliers*, *missing*, *value*, dan memilih reduksi dimensi.
4. Yang menggunakan Algoritma *data mining* termasuk dari *assosiasi*, *sekuensial*, *klasifikasi*, *klasterisasi* dan lain-lain.
5. Menginterpretasikan, mengevaluasi, serta mengvisualisasi suatu pola agar terlihat jika sesuatu yang baru dan menarik dapat dilakukan jika itu memang di perlukan.



Gambar 2.2 : Knowledge Discovery (KDD)

Sumber: Hermawati, FajarAstuti. 2013

2.1.3 Pengolahan Data Mining

Pengelolaan *Data Mining* memiliki beberapa metode pengolahan yaitu :

(Hermawati, FajarAstuti. 2013)

1. Prediksi (*Predictive*).

Prediksi suatu tempat pengolahan, Data Mining untuk melakukan prediksi, tujuan dari metode ini membangun sebuah model prediksi untuk suatu nilai yang dimiliki ciri- ciri , contohnya seperti Algoritma *linier Regression*, *Neural Network*, *Support Vector Machine*, dan lainnya.

2. Asosiasi (*Association*)

Asosiasi teknik yang ada pada Data Mining ,mempelajari tentang hubungan antar data Contoh dari penggunaannya sebagai tempat menganalisis perilaku mahasiswa yang datang tidak tepat waktu. Misalnya jika ada beberapa mahasiswa mempunyai jadwal dengan para dosen A dan Dosen B, maka mahasiswa yang datang terlambat dapat menjadi seperti contoh Algoritma *FP-Growth A priori*, dan lainnya.

3. Klastering (*Clustering*)

Clustearing atau pengelompokan yang merupakan suatu teknik untuk mengelompokkan data kedalam suatu kelompok tertentu. Contoh algoritmanya *K-Means*, *K-Medoids*, *Self-Organisation Map (SOM)*, *Fuzzy C-Means*, dan lain-lain. Contoh untuk *clustering*: Terdapat lima pulau di Indonesia: Sumatera, Kalimantan, Jawa, Sulawesi dan Papua. Ada lima pulau yang telah jadi klaster berdasarkan waktu Indonesia Barat (Sumatera, kalimantan, dan Jawa) Waktu Indonesia Tengah (Sulawesi) dan Waktu Indonesia Timur (Papua)

4. Klasifikasi (*Classification*)

Klasifikasi adalah sebuah teknik yang mengklasifikasikan atau mengklompokkan data, berbeda dengan *Clustering* yang terletak pada sebuah data, karena *Clustering* tidak memiliki variabel yang dependen, sedangkan *classification* harus mempunya variabel-variabel tertentu yang dependen

contohnya seperti Algoritma yang menggunakan metode ID3 dan *K Nearest Neighbors*.

2.1.4 Penerapan Data Mining

Penerapan *Data Mining* terbagi dalam beberapa bidang :

1. Analisa Pasar dan Manajemen

Masalah yang dapat di selesaikan dengan *Data mining* adalah dapat membentuk sebuah target pasar, melihat pola pembelian dari waktu ke waktu, Cross Market Analysis, Profil Customer, Nilai Loyalitas Customer, dan Informasi Summary.

2. Analisa Perusahaan dan Manajemen Resiko

Solusi yang biasanya di selesaikan oleh Data mining adalah Perencanaan Keuangan dan Evaluasi Aset , serta Perencanaan Sumber daya (*resource planning*), Persaingan (*Competition*).

3. Telekomunikasi

Perusahaan telekomunikasi juga menggunakan data mining bertujuan sebagai mencari tahu seberapa banyak transaksi yang telah masuk . dan transaksi yang harus dikerjakan secara manual.

4. Keuangan

Financial Crimes Enforcement Network ,terkenal di amerika serikat yang belakangan ini juga memakai data mining sebagai tempat menambang teriliun an seperti property rekening bank dan transaksi keuangan lainnya untuk mendeteksi transaksi-transaksi keuangan yang mencurigkan (seperti *money laundry*).

5. Asuransi

Data mining dimanfaatkan sebagai mengidentifikasi suatu layanan kesehatan yang biasanya tidak perlu tapi tetap dilakukan oleh pasien yang sering berasuransi.

6. Olahraga

Data mining biasanya digunakan sebagai analisis statistik pada permainan tim olahraga dalam rangka mencapai keunggulan bersaing (*competitive advantage*).

7. Astronomi

Data mining berguna dalam bidang astronomi di ilmu ruang angkasa, seperti telah ditemukannya 22 quasar dengan bantuan *data mining*.

8. Internet Web surf-aid

IBM Surf-aid juga menggunakan data minim untuk mendata terhadap akses pada halaman web yang berhubungan dengan pemasaran yang berguna untuk melihat perilaku dan minat customer serta melihat keefektifan pemasaran melalui Web.

Teknik Data Mining terdapat beberapa teknik yang biasa disebut literatur, Ada 3 teknik Data Mining yang terpopuler (Santo 1999) ialah :

1) *Association Rule Mining*.

Association Rule Mining merupakan teknik data yang menemukan asosiasi yang menjadi atribut contohnya seperti aturan asosiasi yang menjadi pasar swalayan agar penempatan dan perancangan strategi pemasaran agar saat adanya diskon atau kombinasi barang mempunyai kupon ataupun struk.

2) *Klastering (Clustering)*

Clustering dapat di ketahui berbeda dengan *Association Rule Mining* dimana data telah di tentukan kebenarannya, *Clustering* biasa dipakai dalam label pada kelas yang belum di ketahui maka dari itu *Clustering* dikelompokkan

sebagai metode unsur *pervised learning*. kelas data yang belum diketahui, setelah itu *clustering* sering digolongkan sebagai metode unsur *prevised learning*. Prinsip *clustering* adalah memaksimalkan kesamaan antar *cluster*. *Clustering* dapat dilakukan kepada data yang telah dimiliki oleh beberapa atribut yang dipetakan sebagai ruang multidimensi.

3) Klasifikasi,

Klasifikasi memiliki tujuan misalnya, golongan pada hasil yang dapat dipisahkan dengan tiga karakter seperti Pendapatan tertinggi, Pendapatan sedang, hingga pendapatan terendah

2.1.5 Metode Data Mining

Dalam melakukan pengumpulan informasi tentunya terdapat metode, Metode tersebut akan membantu dalam proses menemukan data. Data mining akan menyediakan perencanaan dari ide hingga implementasi akhir.

1. Proses pengambilan Data

Bagaimana proses pengambilan data? Tadi di atas sudah dijelaskan tentang KDD atau Knowledge discovery (mining) in databases. Dengan KDD tersebut itulah kamu dapat melakukan proses pengambilan data. Proses atau tahapan-tahapan tersebut dimulai dari data mentah dan berakhir dengan pengetahuan atau informasi yang telah diolah. Nah proses tersebut sebagai berikut :

- a. **Data Cleansing**, Proses dimana data-data yang tidak lengkap, mengandung error dan tidak konsisten dibuang dari koleksi data. Ketahui juga data lifecycle management untuk mengetahui tentang pengolahan data.
- b. **Data Integration**, Proses integrasi data dimana yang berulang akan dikombinasikan.

- c. **Selection**, Proses seleksi atau pemilihan data yang relevan terhadap analisis untuk diterima dari koleksi data yang ada.
- d. **Data Transformation**, Proses transformasi data yang sudah dipilih ke dalam bentuk mining procedure melalui cara dan agresi data.
- e. **Data Mining**, Proses yang paling penting dimana akan dilakukan berbagai teknik yang diaplikasikan untuk mengekstrak berbagai pola-pola potensial untuk mendapatkan data yang berguna.
- f. **Pattern Evolution**, Sebuah proses dimana pola-pola menarik yang sebelumnya sudah ditemukan dengan identifikasi berdasarkan measure yang telah diberikan
- g. **Knowledge Presentation**, Merupakan proses tahap terakhir, Dalam hal ini digunakan teknik visualisasi yang bertujuan membantu user dalam mengerti dan menginterpretasikan hasil dari penambangan data.

2.1.6 Permasalahan dalam Data Mining

Bukan soal yang mudah dalam mengumpulkan informasi dan melakukan penambangan data yang nantinya data tersebut berguna kedepannya, Banyak sekalian permasalahan yang bisa ditemui saat melakukan penambangan data. Apa saja permasalahan dalam data mining tersebut?

1. Metodologi Mining

- a. Menambang berbagai jenis pengetahuan dari berbagai tipe data
- b. Kinerja: efisiensi, efektivitas, dan skalabilitas
- c. Evaluasi pola: masalah ketertarikan
- d. Memasukkan pengetahuan latar belakang
- e. Menangani kebisingan dan data yang tidak lengkap
- f. Metode penambangan paralel, terdistribusi dan tambahan
- g. Integrasi pengetahuan yang ditemukan dengan yang ada: fusi pengetahuan

2. User interaction

- a. Bahasa kueri penambangan data dan penambangan ad-hoc
- b. Ekspresi dan visualisasi hasil penambangan data
- c. Penambangan pengetahuan interaktif di berbagai tingkatan abstraksi

3. Applications and social impacts

- a. Penambangan data khusus domain & penambangan data tak terlihat
- b. Perlindungan keamanan data, integritas, dan privasi

2.1.7 Contoh Penerapan Data Mining

Penambangan data bisa digunakan di berbagai sektor, Mulai dari sektor bisnis, manajemen, keuangan dan lain sebagainya. Berikut Contoh penerapan Data mining di beberapa sektor :

1. Market Analysis dan Management

Dalam sektor pemasaran biasanya data mining digunakan untuk Pemasaran target, manajemen hubungan pelanggan (CRM), analisis pasar, cross selling, segmentasi pasar.

a. Target Pemasaran,

Misalnya menemukan kelompok pelanggan “model” yang memiliki karakteristik yang sama: minat,tingkat pendapatan, kebiasaan belanja, dll. atau menentukan pola pembelian pelanggan dari waktu ke waktu.

b. Analisis lalu lintas pasar, Menemukan hubungan / hubungan antar produk penjualan, & prediksi berdasarkan asosiasi tersebut.

c. Profiling pelanggan, Jenis pelanggan apa yang membeli produk apa (pengelompokan atau klasifikasi)

d. Analisis kebutuhan pelanggan, Misalnya identifikasi produk terbaik untuk berbagai kelompok pelanggan, Memprediksi faktor apa yang akan menarik pelanggan baru, Penyediaan informasi

ringkasan, Laporan ringkasan multidimensi, Informasi ringkasan statistik (kecenderungan dan variasi pusat data)

3. Corporate Analysis & Risk Management

Penerapan Data mining dalam sektor perusahaan biasanya digunakan untuk prediksi, retensi pelanggan, underwriting yang lebih baik, kontrol kualitas, analisis kompetitif.

- **Perencanaan keuangan dan evaluasi aset**, Misalnya analisis dan prediksi arus kas, analisis klaim kontinjensi untuk mengevaluasi aset, analisis cross-sectional dan time series (rasio keuangan, tren analisis, dll.)
- **Planning Perencanaan sumber daya**, Misalnya merangkum dan membandingkan sumber daya dan pengeluaran
- **Persaingan**, Misalnya memantau pesaing dan arah pasar, mengelompokkan pelanggan ke dalam kelas dan penetapan harga berbasis kelas prosedur, dan mengatur strategi penetapan harga di pasar yang sangat kompetitif.

4. Fraud Detection & Mining Unusual Patterns

Data mining juga berfungsi untuk mencari dan mendeteksi fraud pada sebuah sistem. Dengan menggunakan data mini maka akan bisa melihat dari jutaan transaksi yang masuk.

- **Pendekatan**: Clustering & konstruksi model untuk penipuan, analisis outlier
- **Aplikasi**: Layanan kesehatan, ritel, layanan kartu kredit, telecomm. Misalnya Asuransi otomatis, Pencucian uang, Asuransi kesehatan, Telekomunikasi, Analisis pola yang menyimpang dari norma yang diharapkan, Industri retail, Dll.

2.2 Klasifikasi

Classification adalah proses yang dapat menemukan fungsi dan model penjelasan yang biasa dibedakan menjadi konsep atau data kelas, yang mempunyai tujuan agar dapat memperkirakan kelas sebagai objek yang tidak diketahui (Han dan Kamber, 2006). Klasifikasi juga merupakan fungsi pembelajaran yang dipetakan (mengklasifikasi) sebuah unsur data ke dalam salah satu dari beberapa kelas, dengan memiliki tujuan dapat memperkirakan kelas dari suatu objek yang labelnya diketahui. Data input di klasifikasikan adalah koleksi dari *record*.

Record sudah dikenal sebagai himpunan yang sudah ditentukan menjadi tuple (X, Y) , X yang dimaksud sebuah himpunan atribut, dan Y sebagai atribut tertentu yang sudah ditentukan menjadi label kelas. Teknik ini dapat digunakan oleh *Decision Tree*, *Classifier*, *rule-based classifier*, *neural-network*, *support vector machine*, dan *naive bayes classifier*. Teknik algoritme sebuah pembelajaran yang telah diidentifikasi model yang memberi hubungan sesuai atribut label kelas dari data yang sudah diinput.

2.3 Decision Tree

Decision tree (pohon keputusan) termasuk diagram alir di sebuah struktur pohon pada setiap internal *Node* di notasikan pada atribut yang dapat diuji dan cabang nya dapat dipersentasikan hasilnya atribut tes, dan *Leaf Node* di persentasikan atribut kelas (Han & Kamber, 2001)

Ada 3 jenis *Node* pada *Decision Tree* adalah :

1. *Root Node* adalah *Node* yang terletak paling atas, *node* ini tidak dapat menginput dan tidak punya *output* yang lebih dari satu.
2. *Internal Node*, adalah *node* yang bercabang, *node* ini mempunyai satu *input* yang memiliki *output* minimal dua.

3. *Leaf Node* atau *Terminal Node* adalah *Node* terakhir, *Note* ini cuma memiliki satu penginput dan tidak memiliki *Output*.

Menggunakan *Decission Tree* harus memperhatikan atribut apa yang telah dipilih sebagai pemisah objek, dapat di urutkan struktur mana yang lebih dulu *Tree*, *Criteria*, pemberhentian dan *prunning*.

Decission Tree adalah metode prediksi yang sangat baik dan sudah terkenal. Metode ini telah dapat mengubah suatu fakta menjadi sebuah keputusan yang di presentasikan memiliki aturan, aturan dapat dipahami dengan memakai bahasa alamia dan mereka juga bisa di eksperikan dalam bentuk bahasa yang berbasis data misalnya *Structured Query Language* untuk mencari *record* pada kategori tertentu. *Decission Tree* sangat berguna sebagai pengeksplorasi data, mendapatkan hubungan yang hilang dengan sejumlah variabel yang memiliki target. (Kusrini, 2009).

2.4 Algoritma C4.5

Algoritma Data Mining C4.5 adalah suatu Algoritma yang dapat digunakan sebagai klasifikasi dan segmentasi pengelompokan yang bersifat prediktif. Klasifikasi ini juga sebagai Data mining yang mempunya tujuan untuk menemukan sebuah pola yang berharga dara data yang ukurannya relative besar hingga sampai yang sangat besar. Algoritma C4.5 sebagai pengembangan dari Algoritma ID3. *Decisson Tree* mempunya kemiripan denngan struktuk sebuah pohon yang terdapat *Node* internal (Bukan Daun) yang di deskripsikan sebagai atribut, cabang-cabangnya menggambarkan hasil atribut untuk diuji, pada daun menggambarkan sebbuah kelas.

Decission Tree mulai bekerja dari akar paling atas , sehingga mendapatkan jumlah data yang diuji, contohnya X yang menyatakan kelas , dan data X belum diketahui, *Decission Tree* akan mencari tahu dari akar sampai *node* terakhir, pada nilai setia atribut sesuai dengan data X di uji agar mengetahui sesuai atau tidak

dengan aturan dan keputusan *Decision Tree*, setelah itu *Decision Tree* bertugas memprediksi kelas dari *Tupel X*.

Decision Tree mencocokkan antara Eksplor dan model *Decision Tree* yang sangat baik sehingga pada tahap awal di proses pemodelan dijadikan model akhir dari bermacam macam teknik. *Decision Tree* memiliki struktur yang biasa di pakai untuk membagi data besar menjadi sebuah himpunan *record* yang lebih kecil, dan menerapkan rangkaian aturan dari sebuah keputusan. (Berry dan Linoff. 2004)

Decision Tree merupakan kumpulan- kumpulan dari aturan yang jumlah populasinya heterogen yang lebih kecil sehingga homogen memperhatikan variabel serta tujuannya. *Decision Tree* dibangun dengan baik dengan cara manual dan dapat tumbuh dengan cara otomatis, salah satu algoritma pohon keputusan agar membuat model data yang belum di klasifikasi variabel ini dikelompok kan secara pasti dan model pohon keputusan melebihi dari 21 arah yang terhitung *probabiliti* tiap record terdapat kategori untuk mengklasifikasi record untuk dikelompokkan kedalam satu kelas. Pohon keputusan juga berguna untuk mengestimasi nilai dari serbuah variabel *continue* adapun beberapa teknik sesuai dengan kasus ini argoritma yang biasa di pakai agar membentuk pohon keputusan ialah, ID3, CART, dan C4.5 (Iarose.2006)

Data yang terdapat di dalam sebuah pohon keputusan biasa mempunyai bentuk tabel dengan atribut dan *record*. Atribut pada parameter digunakan sebagai kriteria dalam membentuk pohon contoh nya, menentukan main tenis hal yang harus diperhatikan seperti, cuaca, angin, dan temperatur. Arribut ini menyatakan bahawa dfata solusi perdata yang di sebut targer atribut, yang dimiliki nilai *instance*. Misalnya atribut *instance* seperti cuaca yang cerah, berawan, dan hujan (Basuki dan Syarif, 2003).

2.4.1 Prinsip Kerja Algoritma C4.5

Pada pembelajaran Algoritma C4.5 mempunyai 2 cara kerja ialah (Selvia.et.2014) :

1. Pembuatan pohon keputusan tujuan adanya algoritma penginduksi pohon keputusan merupakan konstruksi struktur data pohon yang digunakan sebagai prediksi kelas atau *record* yang tidak memiliki kelas. C4.5 konstruksi pohon keputusan *Divide and Conquer*. Algoritma sebagai pemecah kasus yang sangat baik dalam menghitung dan membandingkan *gain ratio*, lalu Node-node telah berbentuk level *Algoritma Divide and Conquer* sampai berbentuk daun
2. Pembuatan Aturan-aturan (*Rule Set*). Aturan dari pohon keputusan akan membentuk sebuah *ifthen*. aturan yang menelusuri pohon keputusan dari akar sampai daun. Pada node juga terbentuk sesuatu *IF* sedangkan untuk nilai terdapat daun yang terbentuk hasil atau *THEN*.

2.5 Identifikasi

Identifikasi adalah kata yang berasal dari *identify* artinya adalah meneliti, menelaah, identifikasi juga kegiatan untuk mencari, menemukan, mengumpulkan, meneliti, mencatat data, dan informasi dari kebutuhan. Secara kebutuhan terdapat 2 kategori yaitu kebutuhan yang bersifat terdesak dan kebutuhan yang bersifat terduga amaksudnya tidak terdesak.

Adapun fungsi identifikasi dan tujuan dari identifikasi terhadap program ini adalah sebagai memberitahu masalah serta kebutuhan program yang sangat diinginkan masyarakat. Agar mengetahui banyak sumber yang dapat di manfaatkan program pendukung ini dapat mempermudah menyusun rencana program yang akan dilakukan. Fungsi sebuah program di kembangkan menjadi kebutuhan masyarakat, data yang telah di kumpulkan akan digunakan sebagai penyusun program yang telah di pengaruhi oleh pengelola program.

2.6 Rapid Miner

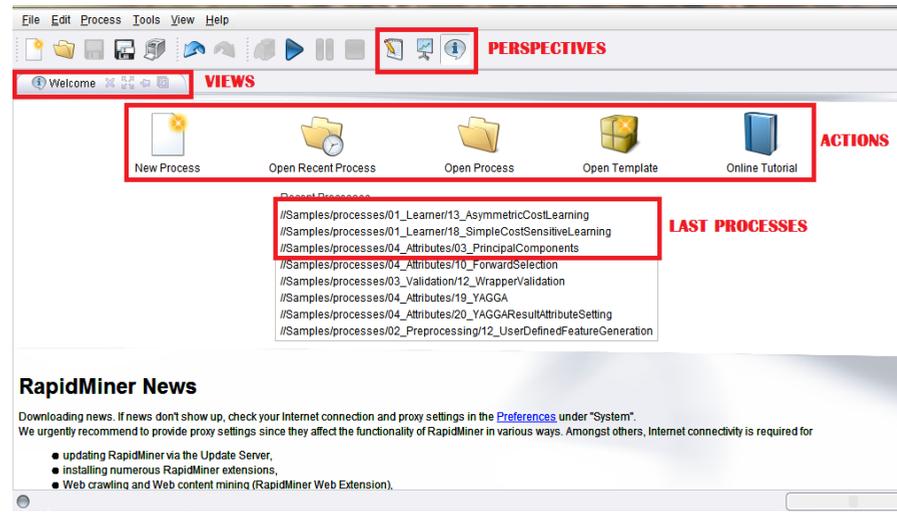
Rapid Miner sebuah perangkat lunak yang mempunyai sifat yang selalu terbuka atau transparan, (*Open Source*). *Rapid Miner* juga dapat memberikan solusi untuk menganalisis sebuah data, Data Mining, Text Mining, memiliki 500 operator data mining. *Rapid Miner* berguna sebagai teknik Deskripsi dan prediksi untuk memberi wawasan pada pengguna hingga menjadi keputusan yang terbaik. *Rapid Miner* mempunyai kurang lebih 500 operator *Data Mining*, termasuk operator sebagai *Input dan Output data Processing* dan *Visualisasi*.

Rapid Miner juga menjadi software yang dapat berdiri dengan sendirinya agar analisis data pada mesin Data Mining dapat diintegritkan pada produknya. *Rapid Miner* telah menyediakan GUI (*Graphic User Interface*) sbagai perancang sebuah *pipeline analisis*. GUI (*Graphic User Interface*) sebagai perancang *pipeline Analisis*. GUI menghasilkan file XML (*Extensible Markup Language*) yang di definisikan sebagai proses analisis sebagai keinginan pengguna di terapkan ke suatu data, lalu file tersebut dibaca oleh *Rapid Miner* agar dapat menjalankan analisis dengan cara otomatis.

Rapid Miner telah menyelesaikan tampilan yang *user Friendly* agar memudahkan pengguna saat aplikasi telah dijalankan . Tampilan *Rapid Miner* telah dikenal dengan sebuah *Perspective*. *Rapid Miner* mempunya 2 *Perspective* ialah:

1. *Welcome Perspective*

Saat membuka aplikasi yang akan di sambut oleh tampilan yang dinamakan *Welcome Perspective*, ditunjukkan pada gambar 2.4



Gambar 2.3 Tampilan Welcome Perspective

Sumber : Denis Aprilla, 2013

Komputer yang telah tersambung dengan internet di bagian bawah *Welcome Perspective* dapat menampilkan berita baru mengenai *Rapid Miner* ini dinamakan *News*. yang terlihat pada daftar *Last Processing (Recent Processing)*, bagian yang di tampilkan pada daftar analisis yang baru saja di lakukan. Ini akan mempermudah kita jika ingin melanjutkan proses yang sebelumnya yang telah ditutup, dengan cara mengklik dua kali dalam proses daftar.

RapidMiner sebelumnya bernama **YALE (Yet Another Learning Environment)**, dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund. RapidMiner didistribusikan di bawah lisensi AGPL (GNU Affero General Public License) versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan menggunakan RapidMiner di lebih dari 40 negara. RapidMiner sebagai software open source untuk data mining tidak perlu diragukan lagi karena software ini sudah terkemuka di dunia. RapidMiner menempati

peringkat pertama sebagai Software data mining pada polling oleh KDnuggets, sebuah portal data-mining pada 2010-2011.

RapidMiner menyediakan GUI (Graphic User Interface) untuk merancang sebuah pipeline analitis. GUI ini akan menghasilkan file XML (Extensible Markup Language) yang mendefinisikan proses analitis keinginan pengguna untuk diterapkan ke data. File ini kemudian dibaca oleh RapidMiner untuk menjalankan analisis secara otomatis.

RapidMiner memiliki beberapa **sifat** sebagai berikut:

- Ditulis dengan bahasa pemrograman Java sehingga dapat dijalankan di berbagai sistem operasi.
- Proses penemuan pengetahuan dimodelkan sebagai operator trees
- Representasi XML internal untuk memastikan format standar pertukaran data.
- Bahasa scripting memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen.
- Konsep multi-layer untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
- Memiliki GUI, command line mode, dan Java API yang dapat dipanggil dari program lain.

Beberapa **Fitur** dari RapidMiner, antara lain:

- Banyaknya algoritma data mining, seperti decision tree dan self-organization map.
- Bentuk grafis yang canggih, seperti tumpang tindih diagram histogram, tree chart dan 3D Scatter plots.

- Banyaknya variasi plugin, seperti text plugin untuk melakukan analisis teks.
- Menyediakan prosedur data mining dan machine learning termasuk: ETL (extraction, transformation, loading), data preprocessing, visualisasi, modelling dan evaluasi
- Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI
- Mengintegrasikan proyek data mining Weka dan statistika R

2.7 Penelitian Terdahulu

Penelitian terdahulu terkait dengan algoritma C4.5, ditinjau dari bidang study penelitian yaitu :

Penelitian oleh Siti Nurhayani, Kusriani, Emha Taufiq Lutfi, 2016 dengan judul Prediksi Mahasiswa Drop Out menggunakan Metode Support Vector Machine, hasil penelitiannya adalah tingginya tingkat keberhasilan mahasiswa dan rendah pula tingkat kegagalan mahasiswa di sini dapat dicerminkan oleh kualitas dari suatu perguruan tinggi. Ada pula faktor kegagalan mahasiswa dan berujung dengan Drop Out, adapun salah satu mengatasi permasalahan tersebut adalah dengan dilakukannya prediksi dengan metode Support Vector Machine , yang mana metode ini mencari hyperplane yang paling optimal dimana dua pola kelas di pisahkan dengan semaksimal mungkin , adapun parameter nya digunakan sebagai Support Vector Machine hanyalah parameter kernel dalam satu parameter C yang diberikan finalty pada titik data yang di klarifikasikan secara acak.

Dalam Support Vector Machine bobot (w) dan biasa (b) merupakan solusi global optimum dari quadratic programming agar cukup sekali running yang dapat dihasilkan solusi yang selalu sama untuk pilihan sekali running solusi yang selalu sama untuk pilihan kernel running yang menghasilkan solusi yang sama untuk

pilihan kernel dan parameter yang sama, dari Penerapan Support Vector Machine diharapkan untuk memperoleh hasil terbaik dalam memprediksi mahasiswa Drop Out.

Penelitian oleh Khafizh Hastuti dengan judul Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif, dengan hasil penelitian Mahasiswa non aktif yang mana mahasiswa yang telah terhenti studi dan tidak melakukan registrasi administratif, mahasiswa yang mempunyai status non aktif dan cenderung Drop Out. Tinggi persentase mahasiswa dengan status non aktif mempengaruhi nilai akreditasi yang tinggi di universitas . perlu diketahui adanya faktor penyebab mahasiswa memiliki status non aktif adalah dengan mengklarifikasi atau memprediksi menggunakan data mining maka perlu dilakukan komparasi di tingkat akurasi dari masing- masing algoritma , Algoritma yang biasa digunakan adalah logistic regression, decision tree, naive, bayes, dan neural network data yang diambil banyaknya 3861 mahasiswa program studi Teknik Informatika, Sistem Informasi dan Desain Komunikasi Visual Universitas Dian Nuswantoro. Hasil ,dari proses ini dapat dievaluasikan dengan *cross validation* , *confusion matrix*, *ROC Curve* dan T-Test untuk mengetahui algoritma Klasifikasi data mining yang paling tepat dan akurat untuk memprediksi mahasiswa yang Non Aktif

Penelitian yang dilakukan oleh Dwi Utari, 2015, dengan judul Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Metode *Decision Tree* C4.5, dengan hasil penelitian keberhasilan dan kegagalan pada study mahasiswa merupakan suatu cerminan kualitas pada setiap perguruan tinggi. Mahasiswa yang tidak aktif yang merupakan siswa yang tidak melakukan registrasi pada tahap awal di semester atau mahasiswa tersebut tidak ikut dalam maha perkuliahan berlangsung setidaknya satu semester, dan selain itu banyak juga mahasiswa yang lulus tidak tepat waktu dan setelah itu mahasiswa meningkatkan jumlah mereka itu berpotensi untuk Drop Out.

Lulus dengan tepat waktu ialah elemen penilaian yang ada di akreditasi perguruan tinggi, dengan ini mengatasi masalah nya dengan cara menerapkan metode data mining dengan metode, c4.5, data mining yang berpotensi non aktif selain itu definisi menggunakan CRSP-DM memakai data mahasiswa S1 fakultas ilmu Komputer . proses untuk memvalidasi siswa nya menggunakan Confusion Matrix, yang nilai kosong vdan IPS semester nilai terbaik 97.60% adapun rasio data 90% percobaan ini biasa nya digunakan pada mahasiswa smester 3 dan ips semester 2.60, sks pada semester 3 sudah di ambil dan kosong, ips smester 4 suatu potensilebihh memilih menjadi mahasiswa tidak aktif dan non aktif.

BAB III

METODOLOGI PENELITIAN

3.1. Tempat Dan Waktu

3.1.1 Tempat Penelitian

Tempat Penelitian adalah tempat di mana penulis melakukan penelitian yang berguna untuk mengambil data untuk mengembangkan penelitian, peneliti menentukan lokasi di STMIK Budi Darma Medan yang lokasinya di JL. Sisingamangaraja No. 338, Siti Rejo I, Kec. Medan Kota, Kota Medan, Sumatera Utara 20219

3.1.2 Waktu Dan Jadwal Pelaksanaan Penelitian

Waktu dan jadwal yang digunakan untuk penelitian menentukan masa studi Mahasiswa Menggunakan Algoritma C4.5 ini ditunjukkan pada Tabel 3.1

Tabel 3.1 Waktu dan Jadwal Penelitian

No.	Waktu	Jadwal Penelitian			
		November	Desember	Januari	Februari
1.	Perencanaan				
2.	Pengumpulan Data				
3.	Analisis Data dan Perancangan Sistem				
4.	Pengujian Sistem				

3.2. Bahan dan Alat Penelitian

Pembuatan Skripsi ini di butuhkan alat dan bahan , agar dapat melancarkan proses pengerjaan penelitian skripsi seperti

3.2.1 *Hardware* (Perangkat Keras)

Hardware adalah sebuah perangkat keras yang di manfaatkan sebagai penelitian ini sbgai berikut

Tabel 3.2 Perangkat Keras

No.	Nama Perangkat Keras
1.	<i>Processor core i3 Core 2,4 Ghz</i>
2.	Kapasitas <i>Harddisk 500GB</i>
3.	Besar Memory RAM 2GB

3.2.2 *Software* (Perangkat Lunak)

Software adalah sebuah perangkat lunak yang biasa digunakan sebagai bahan penelitian untuk menyelesaikan skripsi ini seperti :

Tabel 3.3 Perangkat Lunak

No.	Nama Perangkat Lunak
1.	Windows 64 Bit
2.	Microsof Office Excel
3.	Rapid Miner

3.3. Teknik Pengumpulan Data

Pada tahap ini pengumpulan data ini, akan ada dua langkah yang dilakukan yaitu:

a. Studi Observasi

Observasi adalah salah satu proses yang dilakukan secara mengamati langsung ke lokasi dan pencatatan secara langsung dan sistematis mengenai masalah-masalah yang akan di teliti. Observasi adalah salah satu sebagai ltempat

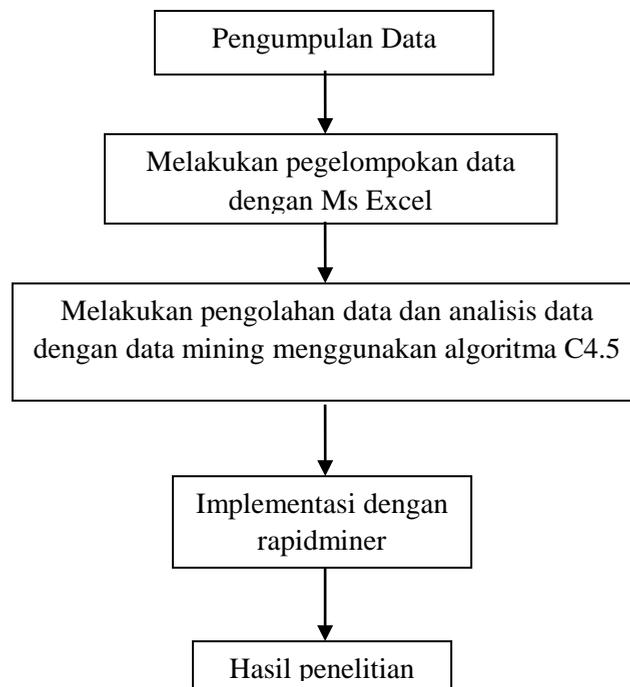
pengumpulan data yang sesuai dengan tujuan si peneliti yang direncanakan dan di catat dengan sistematis, dan di kontrol dengan keadaan (*reliabilitas*) dan kebenarannya (*validation*) dan kali ini penulis selaku observasi telah melakukan observasi pada STMIK Budi Darma Medan

b. Studi Literatur

Studi literature yaitu studi yang dilakukan penelitian dalam mencari teori-teori yang bersangkutan dengan penelitian yang dilakukan. Seperti teori-teori data mining, teori metode-metode, ataupun yang bersangkutan lainnya yang mana bias berupa jurnal, majalah atau buku.

3.4. Penerapan Algoritma

Pada penelitian ini akan dibahas bagaimana pencarian *frequent itemset* menggunakan algoritma C4.5 untuk menentukan masa studi Mahasiswa Menggunakan Algoritma C4.5 berbasis dapat digambarkan dalam blok diagram sebagai berikut:



Gambar 3.1 Blok diagram algoritma C4.5

- Pengumpulan data yang dilakukan dengan memprediksi masa studi mahasiswa, Prediksi masa studi dilakukan pada pertengahan studi mahasiswa ataubertepatan dengan evaluasi studi dua tahun pertama
- Melakukan pengelompokan data dengan Microsoft Excel untuk memudahkan proses analisa maka data terlebih dahulu dikelompokkan berdasarkan tahun angkatan yang akan menjadi atribut pada Microsoft excel
- Melakukan proses pengolahan data dan proses analisis data dengan menggunakan algoritma C4.5 setelah data diolah pada excel selanjutnya data alumni mahasiswa diolah dan diklasifikasikan menggunakan algoritma C4.5 untuk menghasilkan pohon keputusan (*decision tree*)
- Implementasi dengan *rapidminer*, data yang sudah diolah akan diuji dengan aplikasi *rapidminer* untuk menentukan output penelitian
- Hasil penelitian yang di peroleh yaitu penentuan masa studi mahasiswa dengan aplikasi *rapidminer* dalam bentuk pohon keputusan dan aturan (*rule*)

3.5. Analisa Kebutuhan

Analisa kebutuhan system membahas secara garis besar kebutuhan system Tujuannya adalah memahami dengan sesungguhnya kebutuhan dari system yang akan dibuat. Sistem ini dianalisis dengan sistem yang berisi tentang informasi yang segala sesuatunya berkaitan dengan *Data Mining*, dalam proses analisis kebutuhan system akan mencakup dua hal pokok yaitu analisis kebutuhan fungsional dan analisis kebutuhan non fungsional.

- a. Analisis kebutuhan fungsional adalah Analisa yang memiliki tujuan untuk mendapatkan informasi terkait dengan kebutuhan yang berisikan proses-proses yang nantinya akan dilakukan dalam Penerapan data mining dalam menentukan masa studi Mahasiswa menggunakan algoritma C4.5

- b. Analisis kebutuhan non fungsional adalah kebutuhan perangkat yang dibutuhkan untuk mengembangkan sebuah sistem, kebutuhan sistem non fungsional adalah sebagai penelitian adalah:

Spesifikasi Hardware :

1. Processor core i3 Core 2,4 Ghz
2. Kapasitas Harddisk 500GB
3. Besar Memory RAM 2GB

Spesifikasi software :

4. Sistem Operasi Windows 64 Bit
5. Microsof Office Excel
6. Aplikasi Rapid Miner

3.6. Implementasi

Implementasi pada sistem, dilakukan pada tahap pengujian yang selesai dilakukan. Sistem atau aplikasi yang telah dirancang selesai dibangun dan berjalan sesuai yang direncanakan dan sesuai dengan tujuan dari penelitian ini adalah agar menerapkan sistem Data Mining yang dapat menentukan hasil pada masa studi Mahasiswa yang menggunakan Algoritma C4.5

3.7. Pengujian

Di tahap ini adalah tahap pengujian yang dilakukan dengan uji coba terhadap Algoritma yang sudah ditetapkan oleh sistem untuk menentukan apakah algoritma tersebut sudah sesuai dengan tahap ananalisa yang telah dilakukan sebelumnya. Apabila pada tahap ini algoritma belum sesuai dengan tahap ananalisa maka akan kembali lagi ketahap sebelumnya untuk melakukan perbaikan penerapan algoritma tersebut.

BAB IV

HASIL DAN PEMBAHASAN

4.1. Pembahasan

Setelah melalui tahap perancangan pada bab sebelumnya maka setelah itu dilanjut ke tahap berikutnya yaitu, tahap pembahasan mengenai isi skripsi dan hasil yang sudah ditemukan dari penerapan data mining yang sedang dibahas oleh penulis yaitu Menentukan Masa Study Mahasiswa Menggunakan Algoritma C4.5. dimana hasil yang di peroleh secara manual dan melalui komputerisasi dan mengetahui hasil ada perbedaan atau tidak sama sekali.

4.2. Analisa Data

Pada penelitian ini data yang digunakan untuk menerapkan data mining untuk menentukan masa Study terhadap mahasiswa dengan metode algoritma C4.5 yang data nya di dapat dari STIK Budi Darma Program Study Manajemen Informatika (D3). Setelah proses observasi dilakukan untuk mengetahui dan mengamati secara langsung objek yang diteliti dan juga melakukan proses pengumpulan data di STMIK Budi Darma Program Studi Manajemen Informatika (D3).

Data yang sudah di analisa adalah NPM ialah Nama , Jenis kelamin, Asal Daerah , Permintaan Konsentrasi , Asal Sekolah, IPK serta tahun lulus. Pada atribut targer mahasiswa seperti Lulus “ Tepat Waktu” dan Lulus “ Tidak tepat Waktu” Analisa yang digunakan adalah Algoritma C4.5 yang memiliki sampel 44 yang di ambil dengan secara acak terhadap data mahasiswa, hasil penelitian di tunjukkan pada model pohon keputusan dengan aturan mahasiswa yang lulus “Tepat Waktu” adalah siswa yang memiliki IPK tinggi dan siswa yang memiliki IPK sedang dengan focus Palliative. Adapun table keterangan alumni adalah sebagia berikut:

Tabel 4.1 Keterangan Data Alumni

No.	Input	Keterangan
1.	NPM	Nomor Pokok Mahasiswa
2.	Nama Lengkap	Nama Mahasiswa
3.	Jenis Kelamin	Laki-laki/perempuan
4.	Asal Daerah	Asal domisili daerah mahasiswa (Medan/Luar Medan)
5.	Peminatan Konsentrasi	Merupakan kelas peminatan pada Program Studi Manajemen Informatika (D3) STMIK Budi Darma yaitu Manajemen Informatika dan Akutansi Komputer
6.	Asal Sekolah	Asal sekolah mahasiswa yaitu SMK dan SMK
7.	IPK	Merupakan indeks prestasi kumulatif mahasiswa
8.	Tahun Lulus	Tahun pada saat mahasiswa lulus dan ditentukan mahasiswa lulus tepat waktu atau tidak tepat waktu.

4.3. Representasi Data

Pada representasi data ini yaitu untuk membuat decision Tree maka data yang dianalisa seperti NPM yang terdiri dari Nama Siswa, Jenis kelamin, Asal Daerahnya, Asal sekolah, IPK serta tahun lulusnya dan 2 pilihan target yaitu Lulus “Tepat Waktu” atau “Tidak Tepat Waktu”, data yang sudah dianalisis menggunakan Algoritma C4.5 tersebut dan menggunakan sampel sebanyak 44 hasil penelitian menunjukkan model pohon keputusan dengan aturan mahasiswa yang lulus pada tepat waktu adalah mahasiswa- mahasiswa lulus “tepat waktu” adalah mahasiswa yang mendapatkan nilai IPK yang tinggi dan mahasiswa yang memiliki nilai IPK standart, tidak tinggi dan tidak rendah harus lebih *Focus Palliative*.

Decission Tree memiliki alur yang telah di telusuri dan dapat disimpulkan dari simpul akar ke simpul daun yang biasa memegang p[rediksi kelas. *Decission Tree* merupakan metode yang biasa digunakan sebagai peng klasifikasian dan

memprediksi data agar memudahkan dalam interpretasi hasil yang didapat, agar dapat membangun *Decision Tree* di butuhkan Algoritma C4.5 yang merupakan dua model yang tidak berpisah. Pengembangan C4.5 adalah tempat mengatasi sebuah *Missing Value, Continue Data, and Prunning*, yang diinput oleh Algoritma C4.5 yang merupakan *Training Samples dan Training samples*. *Training sample* ini digunakan sebagai sebuah tree yang kebenarannya telah diuji, dan *field-field* data akan digunakan sebagai parameter di dalam klasifikasi data adalah sampel.

Adapun data alumni di buat untuk mempermudah proses data mining untuk menentukan masa bidang study dengan dilakukannya pemilihan atribut, sebagian data atribut akan ada yang di transformasikan. Sebanyak 44 sampel data alumni yang lulus tahun 2019 diambil secara acak dari data yang diberikan STMIK Budi Darma Program Studi Manajemen Informatika (D3). Sistem Informasi yang digunakan untuk menentukan masa studi (dijelaskan pada analisis pengumpulan data), ada pada table 4.2 berikut ini.

Tabel 4.2. Data Alumni Tahun 2019

NPM	Nama	JK	Asal Daerah	Konsentrasi	Asal Sekolah	IPK	TAHUN LULUS
16020 50	Ade Pinta Uli	P	Deli Tua	Managemen Informatika	SMA	3.13	2019
15021 27	Adriani Lasmalita Sipayung	P	Pulu Raja	Managemen Informatika	SMK	3.62	2019
16020 56	Agita Carina Sembiring	P	Bandar Betsy	Managemen Informatika	SMK	3.76	2019
16020 54	Aprilini Franciska Samosir	P	Batam	Komputer Akuntansi	SMK	3.08	2019

NPM	Nama	JK	Asal Daerah	Konsentrasi	Asal Sekolah	IPK	TAHUN LULUS
1602018	Aston Rivandy Dalimunthe	L	Galang	Managemen Informatika	SMK	3.21	2019
1502105	Ayudina Pasha	P	Medan	Managemen Informatika	SMA	3.38	2019
1602064	Christina Mariana Tampubolon	P	Simpang gambir	Managemen Informatika	SMK	3.64	2019
1602096	Desi Rezeki Br Purba	P	Simpang Empat	Managemen Informatika	SMK	3.75	2019
1602046	Dita Ashary Putri	P	Tanjung Saraelan g	Managemen Informatika	SMA	3.50	2019
1802076	Elsa Manurung	P	Pematan g Siantar	Managemen Informatika	SMA	3.53	2019
1602041	Elsa Triana Saragih	P	Pinangso ri	Managemen Informatika	SMA	3.54	2019
1602053	Emira Waruwu	P	Sidikalan g	Managemen Informatika	SMA	3.77	2019
1602071	Jenny Ester Panjaitan	P	Tuktuk Bariangi	Managemen Informatika	SMA	3.44	2019
1602063	Johansyah	L	Bireun	Komputer Akuntansi	SMA	3.26	2019
1602084	Joi Fanni Siregar	P	Medan	Komputer Akuntansi	SMK	3.07	2019

NPM	Nama	JK	Asal Daerah	Konsentrasi	Asal Sekolah	IPK	TAHUN LULUS
16020 27	Juliana Ribka Simanjuntak	L	Medan	Managemen Informatika	SMK	3.39	2019
16020 23	Juliem Charisman Zalukhu	L	Tebing Tinggi	Komputer Akuntansi	SMA	3.18	2019
16020 25	Junimasdayanti Lase	L	Balige	Komputer Akuntansi	SMK	3.29	2019
18020 75	Khairun Nisa	P	Siborong -borong	Komputer Akuntansi	SMA	3.00	2019
16020 78	Labora Christina Silalahi	P	Sukaram ai	Komputer Akuntansi	SMA	3.26	2019
16020 22	Larisma Decsanta Br Sitohang	P	Simantin	Komputer Akuntansi	SMA	3.14	2019
16020 61	Loisa Revina Manalu	P	Parsaora n	Komputer Akuntansi	SMA	3.42	2019
16020 04	M. Agung Priari	L	Lbn Panjaitan	Komputer Akuntansi	SMK	3.29	2019
16020 40	Mayang Sari	P	Resdes	Komputer Akuntansi	SMA	3.39	2019
16020 92	Muhammad Rizki Alfianda	L	Medan	Komputer Akuntansi	SMK	3.36	2019
16020 38	Nadira Sabila	P	Medan	Komputer Akuntansi	SMA	3.17	2019

NPM	Nama	JK	Asal Daerah	Konsentrasi	Asal Sekolah	IPK	TAHUN LULUS
1602017	Nur Arfah Eliza	P	Hapoltahan Nauli	Komputer Akuntansi	SMA	3.69	2019
1602080	Octaviany Eka Syah Putri Sebayang	P	Medan	Komputer Akuntansi	SMA	3.52	2019
1602011	Paulina Tanjung	P	Ujung Pandang	Managemen Informatika	SMA	3.48	2019
1602068	Putra Natalis Ziliwu	L	Nagatonga	Komputer Akuntansi	SMA	3.49	2019
1602098	Rajasaid Natid Harahap	L	Huta Samosir	Managemen Informatika	SMK	3.59	2019
1602094	Rany Bunga Afriyanti Manurung	P	Medan	Komputer Akuntansi	SMK	3.15	2019
1602014	Rosdiana Simamora	P	Sampuar a	Komputer Akuntansi	SMA	3.10	2019
1602083	Sarah Aulia	P	Indrapur a	Managemen Informatika	SMK	3.15	2019
1702008	Srilestari Agustina Marpaung	P	Hutagalung	Komputer Akuntansi	SMK	3.36	2019
1502011	Suparno	L	Santar Jehe	Komputer Akuntansi	SMA	3.21	2019
1602048	Syahril Ginting	L	Tanjung Tiram	Komputer Akuntansi	SMK	3.30	2019

NPM	Nama	JK	Asal Daerah	Konsentrasi	Asal Sekolah	IPK	TAHUN LULUS
1602049	Vebni Desmita Natalia Siringoringo	P	Sumbul	Komputer Akuntansi	SMA	3.47	2019
1602003	Vetrisna Elisabeth Gultom	P	Cinta Damai	Komputer Akuntansi	SMK	3.15	2019
1602072	Vivi Nove Rianti Hia	P	Medan	Managemen Informatika	SMA	3.31	2019
1602042	Wilda Yati	P	Paruloha n	Managemen Informatika	SMA	3.00	2019
1602008	Yeni Syahfitri	P	Tigaling ga	Komputer Akuntansi	SMA	3.15	2019
1602006	Yuli Kartika	P	Lumban Simbolo n	Managemen Informatika	SMA	3.00	2019
1602065	Yuni Kartika	P	Laepinan g	Komputer Akuntansi	SMA	3.50	2019
1602110	Syuhra Wirdi	P	Silalahi	Komuter Akutansi	SMA	3.21	2018
1602029	Ricky Zulkarnain Putra	P	Parparea n	Managemen Informatika	SMK	3.86	2018
1602039	Azrian Finanda Lubis	P	Medan	Komuter Akutansi	SMK	3.21	2019
1602030	Muhammad Firdaus	P	Suhimah asar	Komuter Akutansi	SMK	3.43	2019

NPM	Nama	JK	Asal Daerah	Konsentrasi	Asal Sekolah	IPK	TAHUN LULUS
16020 67	Suci Ayu Lestari Br. Regar. D	L	Bandar Lama	Managemen Informatika	SMK	3.21	2018
16021 03	Hotma Fitriani Sianturi	P	Pasir Jae	Managemen Informatika	SMK	3.36	2018
16021 19	Anil Syahputra	P	Medan	Managemen Informatika	SMA	3.68	2018
16020 04	Ailul Imran	P	PT.ABM . T.Panji	Komputer Akutansi	SMA	3.20	2019
16020 53	Sheila Regina Manalu	P	Duri	Managemen Informatika	SMA	3.15	2018
16021 24	Anju Mikhael Pratomo	P	Janggir leto	Managemen Informatika	SMA	3.72	2018
16021 36	Sylvi Fadhlyani Nasution	P	Medan	Managemen Informatika	SMK	3.89	2018
16022 08	Hernika Situmorang	L	Medan	Komputer Akutansi	SMK	3.02	2019

Setelah menentukan basis analisa data, maka tahap selanjutnya melakukan proses perhitungan dengan metode C4.5 untuk menentukan masa study mahasiswa sebagai berikut :

1. *Cleaning Data*

Data adalah salah satu dokumen para alumni program study Manajemen Informatika D3, berdasarkan proses *Cleaning Data* dapat digunakan sebagai model atribut Decision Tree yang terdiri dari data Jenis kelamin, Asal Daerah, peminatan, konsentrasi, asal sekolah, dan IPK mahasiswa. Kelima atribut

tersebut di gunakan sebagai data *input* sedang *Output* yang ditentukan berdasarkan NIM dan tahun lulus alumni

2. Transformasi Data

Berdasarkan data ulumni yang sudah melalui proses *cleaning* makan dilakukan transformasi data, data yang dapat ditransformasikan adalah pada jenis kelamin, Asal daerah di mana dia tinggal, Asal sekolah sebelum masuk universitas, serta IPK nilai yang di dapat oleh para mahasiswa. Data yang diinput merancang model *Decision Tree*. Transformasi telah dilakukan dengan membuat klasifikasi terhadap masing-masing atribut yang ditentukan, lihat tabel 4.3

Tabel 4.3 Klasifikasi pada Atribut *Input*

No.	Atribut	Class
1.	Ipk	<3,00 (Rendah)
		3,00<=IPK<3,50 (Sedang)
		>=3,50(Tinggi)
2.	Jenis Kelamin	Laki-Laki
		Perempuan
3.	Asal Daerah	Medan
		Luar Medan
4.	Asal Sekolah	SMA
		SMK
5.	Peminatan	Manajemen Informatika
		Komputer Akutansi

Pada table 4.3 telah dijelaskan bahwasanya 5 atribut yang biasa digunakan untuk menginput data, memp[unyai klasifikasi masing-masing agar mempermudah proses *Data Mining*. Ke 5 atribut tersebut akan dianalisa menggunakan Algoritma C4.5 untuk dapat di tentukan secara tepat waktu , atau tidak tepat waktu kelulusan

mahasiswa selama menyelesaikan masa study nya sesuai dengan atribut atau target yang telah di tentukan Lihat tabel 4.4

Tabel 4.4 Atribut Taget

Atribut Target	Keterangan
Tepat Waktu	Masa Studi ≤ 1.5 Tahun
Tidak Tepat Waktu	Masa Studi > 1.5 Tahun

Dari table 4.4 telah di jelaskan bahwasanya mempunyai 2 target yang peneliti inginkan adalah Lulus secara tepat waktu dan lulus secara tidak tepat waktu hal ini agar dapat menentukan sebagai metode masa study mahasiswa. Analisis yang di tetapkan dengan 4 sampel data alumni ditransformasikan sehingga menghasilkan keputusan terlihat pada tabel 4.5 sebagi berikut:

Tabel 4.5. Data Tranformasi

NPM	Nama	JK	Asal Daerah	Konsentrasi	Asal Sekolah	IPK	TAHUN LULUS
1602050	Ade Pinta Uli	P	Luar Medan	Managemen Informatika	SMA	Sedang	Tepat Waktu
1502127	Adriani Lasmalita Sipayung	P	Luar Medan	Managemen Informatika	SMK	Tinggi	Tepat Waktu
1602056	Agita Carina Sembiring	P	Luar Medan	Managemen Informatika	SMK	Tinggi	Tepat Waktu
1602054	Aprilini Franciska Samosir	P	Luar Medan	Komputer Akuntansi	SMK	Sedang	Tidak Tepat Waktu

NPM	Nama	JK	Asal Daerah	Konsentrasi	Asal Sekolah	IPK	TAHUN LULUS
1602018	Aston Rivandy Dalimunthe	L	Luar Medan	Managemen Informatika	SMK	Sedang	Tepat Waktu
1502105	Ayudina Pasha	P	Medan	Managemen Informatika	SMA	Sedang	Tepat Waktu
1602064	Christina Mariana Tampubolon	P	Luar Medan	Managemen Informatika	SMK	Tinggi	Tepat Waktu
1602096	Desi Rezeki Br Purba	P	Luar Medan	Managemen Informatika	SMK	Tinggi	Tepat Waktu
1602046	Dita Ashary Putri	P	Luar Medan	Managemen Informatika	SMA	Tinggi	Tepat Waktu
1802076	Elsa Manurung	P	Luar Medan	Managemen Informatika	SMA	Tinggi	Tepat Waktu
1602041	Elsa Triana Saragih	P	Luar Medan	Managemen Informatika	SMA	Tinggi	Tepat Waktu
1602053	Emira Waruwu	P	Luar Medan	Managemen Informatika	SMA	Tinggi	Tepat Waktu
1602071	Jenny Ester Panjaitan	P	Luar Medan	Managemen Informatika	SMA	Sedang	Tepat Waktu
1602063	Johansyah	L	Luar Medan	Komputer Akuntansi	SMA	Sedang	Tidak Tepat Waktu
1602084	Joi Fanni Siregar	P	Medan	Komputer Akuntansi	SMK	Sedang	Tidak Tepat Waktu

NPM	Nama	JK	Asal Daerah	Konsentrasi	Asal Sekolah	IPK	TAHUN LULUS
1602027	Juliana Ribka Simanjuntak	L	Medan	Managemen Informatika	SMK	Sedang	Tepat Waktu
1602023	Juliem Charisman Zalukhu	L	Luar Medan	Komputer Akuntansi	SMA	Sedang	Tepat Waktu
1602025	Junimasdayanti Lase	L	Luar Medan	Komputer Akuntansi	SMK	Sedang	Tidak Tepat Waktu
1802075	Khairun Nisa	P	Luar Medan	Komputer Akuntansi	SMA	Sedang	Tepat Waktu
1602078	Labora Christina Silalahi	P	Luar Medan	Komputer Akuntansi	SMA	Sedang	Tidak Tepat Waktu
1602022	Larisma Decsanta Br Sitohang	P	Luar Medan	Komputer Akuntansi	SMA	Sedang	Tidak Tepat Waktu
1602061	Loisa Revina Manalu	P	Luar Medan	Komputer Akuntansi	SMA	Sedang	Tidak Tepat Waktu
1602004	M. Agung Priari	L	Luar Medan	Komputer Akuntansi	SMK	Sedang	Tidak Tepat Waktu
1602040	Mayang Sari	P	Luar Medan	Komputer Akuntansi	SMA	Sedang	Tepat Waktu

NPM	Nama	JK	Asal Daerah	Konsentrasi	Asal Sekolah	IPK	TAHUN LULUS
1602092	Muhammad Rizki Alfianda	L	Medan	Komputer Akuntansi	SMK	Sedang	Tidak Tepat Waktu
1602038	Nadira Sabila	P	Medan	Komputer Akuntansi	SMA	Sedang	Tepat Waktu
1602017	Nur Arfah Eliza	P	Luar Medan	Komputer Akuntansi	SMA	Tinggi	Tepat Waktu
1602080	Octaviany Eka Syah Putri Sebayang	P	Medan	Komputer Akuntansi	SMA	Tinggi	Tepat Waktu
1602011	Paulina Tanjung	P	Luar Medan	Managemen Informatika	SMA	Sedang	Tepat Waktu
1602068	Putra Natalis Ziliwu	L	Luar Medan	Komputer Akuntansi	SMA	Sedang	Tepat Waktu
1602098	Rajasaid Natid Harahap	L	Luar Medan	Managemen Informatika	SMK	Tinggi	Tepat Waktu
1602094	Rany Bunga Afriyanti Manurung	P	Medan	Komputer Akuntansi	SMK	Sedang	Tidak Tepat Waktu
1602014	Rosdiana Simamora	P	Luar Medan	Komputer Akuntansi	SMA	Sedang	Tepat Waktu
1602083	Sarah Aulia	P	Luar Medan	Managemen Informatika	SMK	Sedang	Tepat Waktu
1702008	Srilestari Agustina Marpaung	P	Luar Medan	Komputer Akuntansi	SMK	Sedang	Tidak Tepat Waktu

NPM	Nama	JK	Asal Daerah	Konsentrasi	Asal Sekolah	IPK	TAHUN LULUS
1502011	Suparno	L	Luar Medan	Komputer Akuntansi	SMA	Sedang	Tepat Waktu
1602048	Syahril Ginting	L	Luar Medan	Komputer Akuntansi	SMK	Sedang	Tidak Tepat Waktu
1602049	Vebni Desmita Natalia Siringoringo	P	Luar Medan	Komputer Akuntansi	SMA	Sedang	Tepat Waktu
1602003	Vetrisna Elisabeth Gultom	P	Luar Medan	Komputer Akuntansi	SMK	Sedang	Tidak Tepat Waktu
1602072	Vivi Nove Rianti Hia	P	Medan	Managemen Informatika	SMA	Sedang	Tepat Waktu
1602042	Wilda Yati	P	Luar Medan	Managemen Informatika	SMA	Sedang	Tepat Waktu
1602008	Yeni Syahfitri	P	Luar Medan	Komputer Akuntansi	SMA	Sedang	Tepat Waktu
1602006	Yuli Kartika	P	Luar Medan	Managemen Informatika	SMA	Sedang	Tepat Waktu
1602065	Yuni Kartika	P	Luar Medan	Komputer Akuntansi	SMA	Tinggi	Tepat Waktu
1602110	Syuhra Wirdi	P	Luar Medan	Komputer Akuntansi	SMA	Sedang	Tepat Waktu
1602029	Ricky Zulkarnain Putra	P	Luar Medan	Managemen Informatika	SMK	Tinggi	Tepat Waktu

NPM	Nama	JK	Asal Daerah	Konsentrasi	Asal Sekolah	IPK	TAHUN LULUS
1602039	Azrian Finanda Lubis	P	Medan	Komputer Akutansi	SMK	Sedang	Tidak Tepat Waktu
1602030	Muhammad Firdaus	P	Luar Medan	Komputer Akutansi	SMK	Sedang	Tidak Tepat Waktu
1602067	Suci Ayu Lestari Br. Regar. D	L	Luar Medan	Managemen Informatika	SMK	Sedang	Tepat Waktu
1602103	Hotma Fitriani Sianturi	P	Luar Medan	Managemen Informatika	SMK	Sedang	Tepat Waktu
1602119	Anil Syahputra	P	Medan	Managemen Informatika	SMA	Tinggi	Tepat Waktu
1602004	Ailul Imran	P	Luar Medan	Komputer Akutansi	SMA	Sedang	Tidak Tepat Waktu
1602053	Sheila Regina Manalu	P	Luar Medan	Managemen Informatika	SMA	Sedang	Tepat Waktu
1602124	Anju Mikhael Pratomo	P	Luar Medan	Managemen Informatika	SMA	Tinggi	Tepat Waktu
1602136	Sylvi Fadhlyani Nasution	P	Medan	Managemen Informatika	SMK	Tinggi	Tepat Waktu
1602208	Hernika Situmorang	L	Medan	Komputer Akutansi	SMK	Sedang	Tidak Tepat Waktu

3. Merancang *Decision Tree* dengan Algoritma C4.5

Dari penjelasan tabel 4.5, untuk menentukan *root* (akar) dari *Decision tree* dalam menentukan masa studi mahasiswa maka dilakukan perhitungan *entropy* (1) dan *gain* (2). Berikut hasil hitung atau jumlah *entropy* dan *gain* pada tahap pertama, lihat pada tabel 4.6.

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

S : Himpunan Kasus

A : Atribut

N : Jumlah Partisi Atribut A

|S_i| : Jumlah Partisi Kasus pada Partisi Ke-i

|S| : Jumlah Kasus dalam M

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=0}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i)$$

S : Himpunan Kasus

N : Jumlah Partisi dalam S

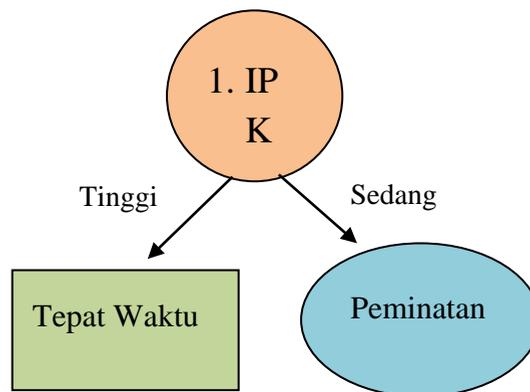
P_i : Proporssi dari S_i Terhadap S

Tabel 4.6. Hasil perhitungan *Gain* dan *Entropy* pada tahap 1.

Th p	Atribut	Kelas	Jumlah Kasus	Tepat Waktu	Tidak Tepat Waktu	Entropy (D1)	Entro phy (D)	Gain
1	TOTAL		56	39	17	0.89		
	IPK	Rendah	0	0	0	0.00	0.70	<u>0.18</u>
		Sedang	40	23	17	0.98		
		Tinggi	16	16	0	0.00		

	Jenis	Laki-Laki	9	6	3	0.92	0.89	0.00
	Kelamin	Perempuan	47	33	14	0.88		
	Asal	Medan	12	7	5	0.98	0.87	0.01
	Daerah	Luar Medan	44	32	12	0.85		
	Asal	SMA	32	27	5	0.63	0.79	0.10
	Sekolah	SMK	24	12	12	1.00		
	Peminatan	Manajemen Informatika	26	26	0	0.00	0.51	0.18
		Komputer Akuntansi	30	13	17	1.72		

Lihat hasil tabel 4.6. telah diketahui bahwa *gain* adalah yang tertinggi ada pada Peminatan dan IPK yaitu 0.18, akan tetapi pada IPK terdapat satu yang dapat digunakan lebih mungkin sebagai node akar. Dari hasil perhitungan pada IPK terdapat IPK-Sedang yang akan di gambarkan oleh *decision tree* untuk itu untuk hasil sementara seperti pada gambar 1.



Gambar 4.1. *Decision Tree* Node 1 Penentuan Masa Studi Mahasiswa

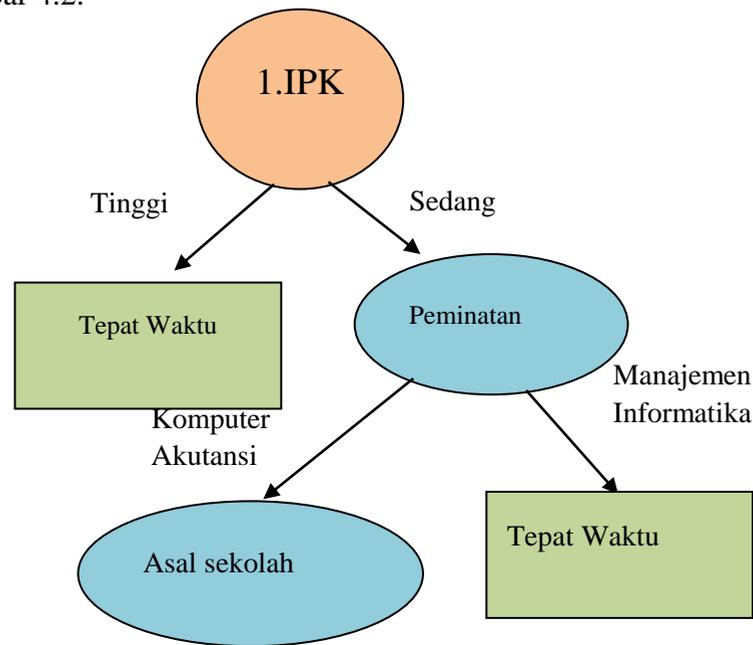
Ada berapa aspek nilai dari IPK ialah dari yang Rendah, Sedang, dan Tinggi. Adapun dari 44 sampel data yang telah di analisa ada 2 yang dapat di klasifikasikan IPK yang rendah dan yang Tinggi , dikarenakan IPK yang terendah tidak termasuk dari klasifikasi 44 sampel yang di analisis. Hasil penilaian yang tinggi dapat di klasifikasikan menjadi 1 keputusan dan hasil nilai yang rendah juga menjadi 1 keputusan dan dilakukan proses perhitungan kembali *entropy* dan *gain*, lihat pada tabel 4.7

Tabel 4. 7. Hasil Perhitungan *Gain* dan *Entropy* pada Node 1.1

Node	Atribut	Kelas	Jumlah Kasus	Tepat Waktu	Tidak Tepat Waktu	Entropy (D1)	Entropy (D)	Gain
1.1	IPK-Sedang		40	23	17	0.98		
	Jenis Kelamin	Laki-Laki	6	3	3	1.00	0.98	0.00
		Perempuan	34	20	14	0.98		
	Asal Daerah	Medan	9	4	5	0.99	0.97	0.01
		Luar Medan	31	19	12	0.96		
	Asal Sekolah	SMA	23	18	5	0.76	0.81	0.18
		SMK	17	5	12	0.87		
	Peminatan	Manajemen Informatika	13	13	0	0.00	0.57	<u>0.42</u>

		Komputer Akuntansi	27	10	17	0.99		
--	--	-----------------------	----	----	----	------	--	--

Dari tabel 4.7 diketahui bahwa atribut dengan nilai yang tertinggi adalah peminatan ialah 0.42, hingga peminatan menjadi Node dari cabang IPK yang mempunyai nilai sedang maka dapat digambarkan *decision tree* sementara seperti pada gambar 4.2.



Gambar 4.2. *Decision Tree* Node 1.1 Penentuan Masa Studi Mahasiswa

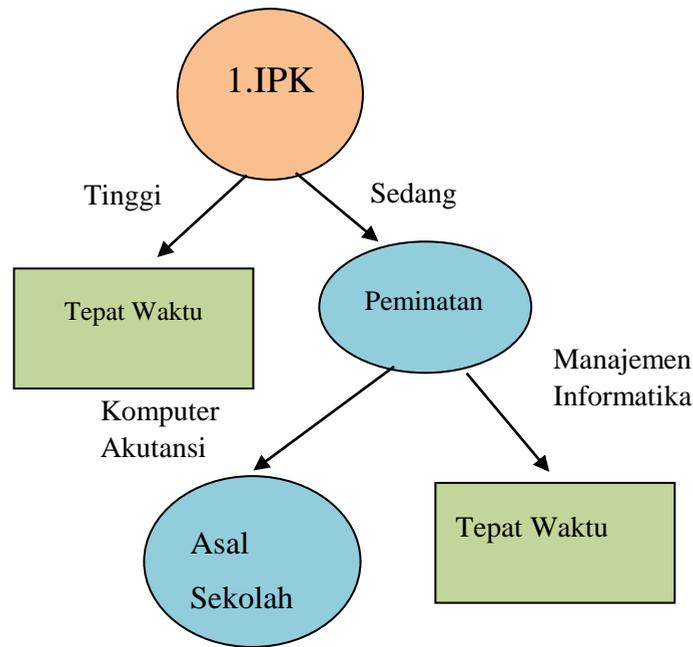
Untuk atribut Peminatan Konsentrasi jurusan memiliki 2 nilai atribut yaitu Manajemen Informatika, dan Komputer Akuntansi. Nilai atribut “ Mnagemen informatika “ dapat di klasifikasikan sebagai kasus menjadi 1 keputusan ialah lulus “ tepat Waktu”, sedangkan nilai dari atribut “ Komputer Akuntansi” menjadi 1 keputusan hingga di lakukan perhitungan *entrophy dan gain* yang lanjut.

Tabel 4.8 Pehitungan *Gain* dan *Entropy* pada Node 1.2

Node	Atribut	Kelas	Jumlah Kasus	Tepat Waktu	Tidak Tepat Waktu	Entropy (D1)	Entropy (D)	Gain
1.2	Peminatan + IPK	Komputer Akutansi + Sedang	15	9	6	0.97		
	Jenis Kelamin	Laki-Laki	1	1	0	0.00	0.9	0.1
		Perempuan	14	8	6	0.99		
	Asal Daerah	Medan	2	0	2	0.00	0.8	0.2
		Luar Medan	13	9	4	0.89		
	Asal Sekolah	SMA	9	9	0	0.00	0.0	1.0
		SMK	6	0	6	0.00		

Dari tabel 4.8 diketahui pada atribut dengan nilai tertinggi adalah atribut “Asal Sekolah” yaitu dapat menjadi Node cabang dari IPK yang bernilai “Sedang” dan Perminatan bernilai “Teknik Informatika”, Atribut asal sekolah memiliki 2 nilai yaitu SMA dan SMK yang telah menjadi kasus yang sudah di klasifikasikan menjadi 1 keputusan ialah Lulus secara Tidak Tepat Waktu dan Nilai Atribut SMA yang menjadi sudah di klasifikasikan juga menjadi 1 keputusan yang lulus “Tepat Waktu” pada nilai atribut “Asal Sekolah” tersebut sudah di klasifikasikan menjadi kasus yang telah dihitung dari nilai *entropy* dan *gain* Node 1.2 yang digambarkan *Decission Tree* byang menentukan masa study mahasiswa STMIK

Budi Darma Program Studi Manajemen Informatika (D3) seperti terlihat pada gambar 4.3 dibawah ini.



Gambar 4.3. Decision Tree Node 1.2 Penentuan Masa Studi Mahasiswa

4.4. Hasil Analisa Data

Adapun hasil analisa data yang telah di uji, maka kita dapat menyimpulkan bahwa hasil dari analisa dari Penentuan Masa Studi Mahasiswa Menggunakan Metode Algoritma C4.5 adalah

1. Jika Mahasiswa memiliki IPK “Tinggi” maka pada masa study Mahasiswa dapat lulus dengan “ Tepat Waktu”
2. Jika Mahasiswa memiliki IPK “ Sedang” dan peminatan adalah Teknik Informatika maka pada masa study Mahasiswa dapat lulus dengan “ Tidak Tepat Waktu”
3. Jika Mahasiswa memiliki IPK “ Rendah” dan peminatan adalah Teknik Informatika maka pada masa study Mahasiswa dapat lulus dengan “ Perbaikan“

4. Jika mahasiswa memiliki IPK “Sedang” dan peminatan adalah “Komputer Akutansi” dan asal sekolah adalah SMA maka study Mahasiswa Lulus “ Tepat Waktu”

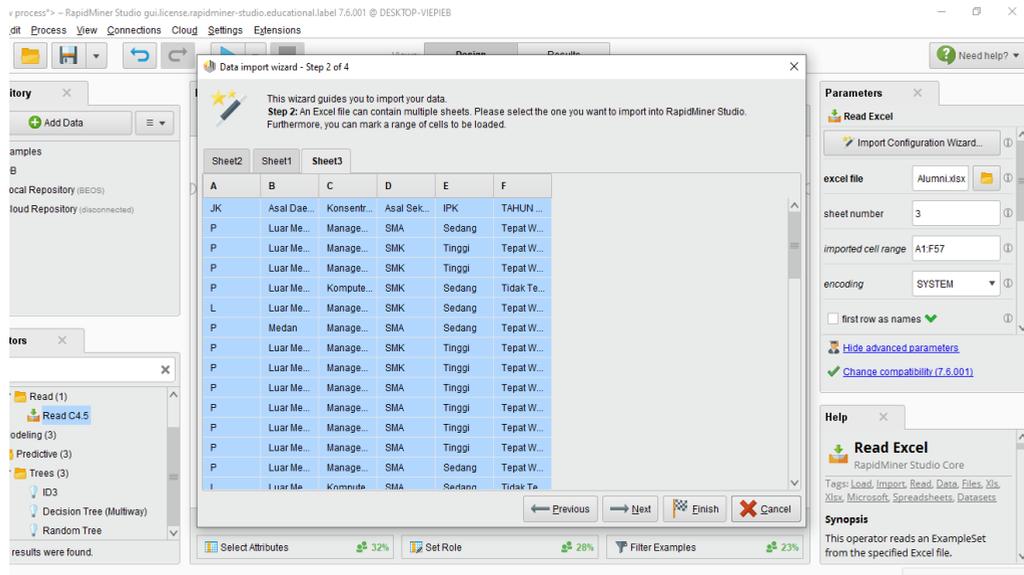
4.5. Hasil Pengujian

Implementasi hasil pengujian penerapan data mining dalam menentukan masa studi mahasiswa menggunakan algoritma C4.5 dengan melakukan pengujian menggunakan tools *Rapidminer*. Dimana dengan adanya pengujian sistem ini akan diketahui apakah hasil yang diperoleh secara manual sama dengan hasil yang diperoleh secara komputerisasi

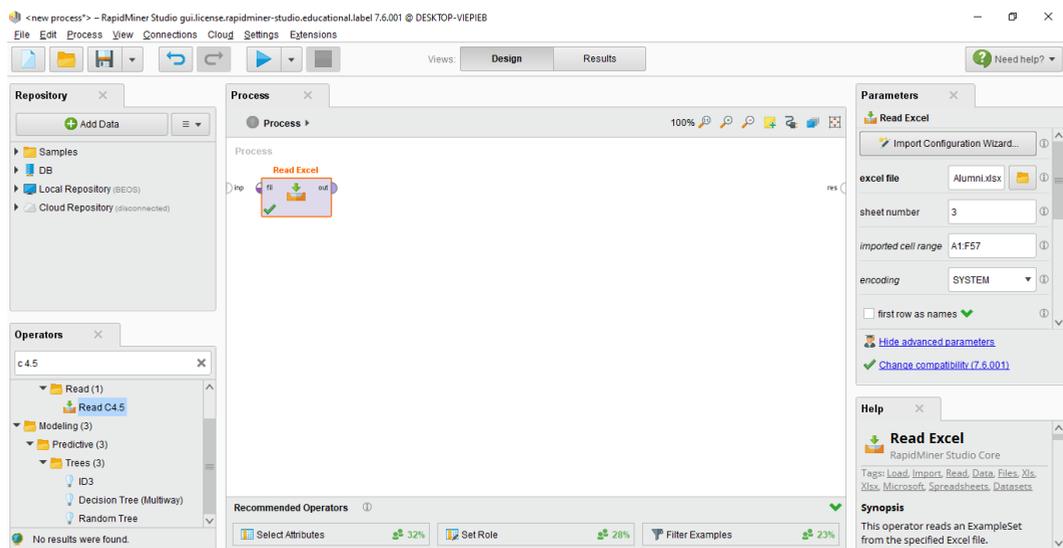
4.6. Pengujian

Pada pengujian ini menggunakan aplikasi *Rapidminer*, Adapun tahapan pengujian penerapan data mining dalam menentukan hasil masa study Mahasiswa dengan Algoritma C4.5 dengan aplikasi *Rapidminer* adalah:

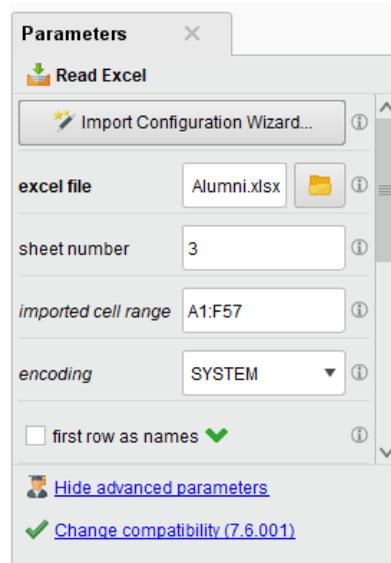
Pada penelitian ini jenis data yang diuji adalah data berupa *Microsoft Excel*, Yang file datanya disimpan dalam *File Extension.xlsx* maka menu yang dipilih adalah menu operator dan pilih submenu *Data Access* → *Read Excel* yang dapat dilihat pada gambar 4.4:

Gambar 4.4 Operator *Rapid miner*

Kemudian menu *read excel* di *double* klik atau di klik dan di *drag* ke halaman *main process* seperti pada tampilan gambar 4.5 berikut ini :

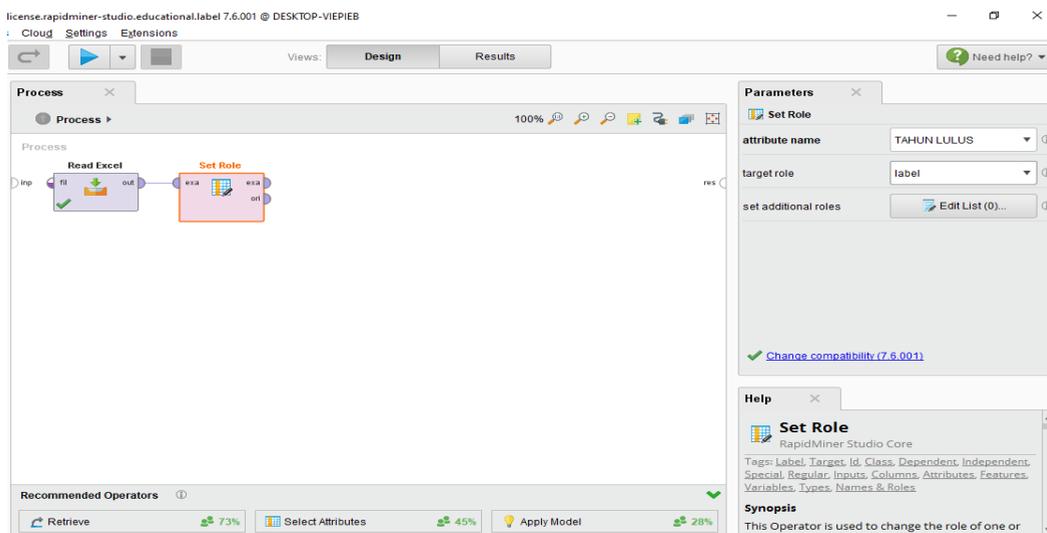
Gambar 4.5 Read Excel Pada *Main Process*

Setelah Sub Menu *Read Excel* telah berada pada halaman *main process* maka selanjutnya dilakukan *import data* dengan pemilihan menu *import data* seperti tampilan pada gambar 4.6 berikut ini



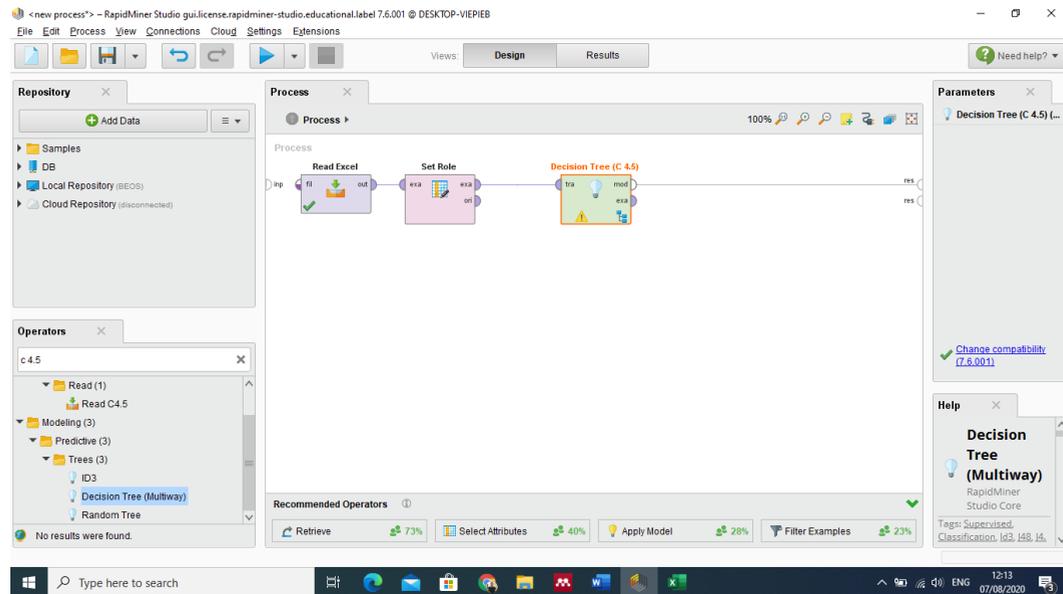
Gambar 4.6 *Import Configuration Wizard.*

Kemudian menu *Set Role* di *double* klik atau di klik dan di *drag* ke halaman *main process* seperti pada tampilan gambar 4.7 berikut ini:



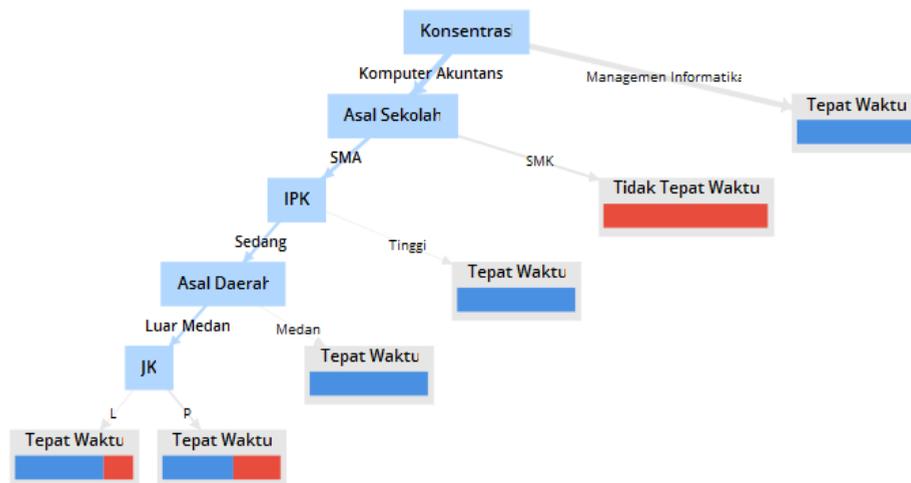
Gambar 4.7 *Set Role Pada Main Process*

Kemudian menu Decision Tree algoritma C 4.5 di *double* klik atau di klik dan di *drag* ke halaman *main process* seperti pada tampilan gambar 4.8 berikut ini:

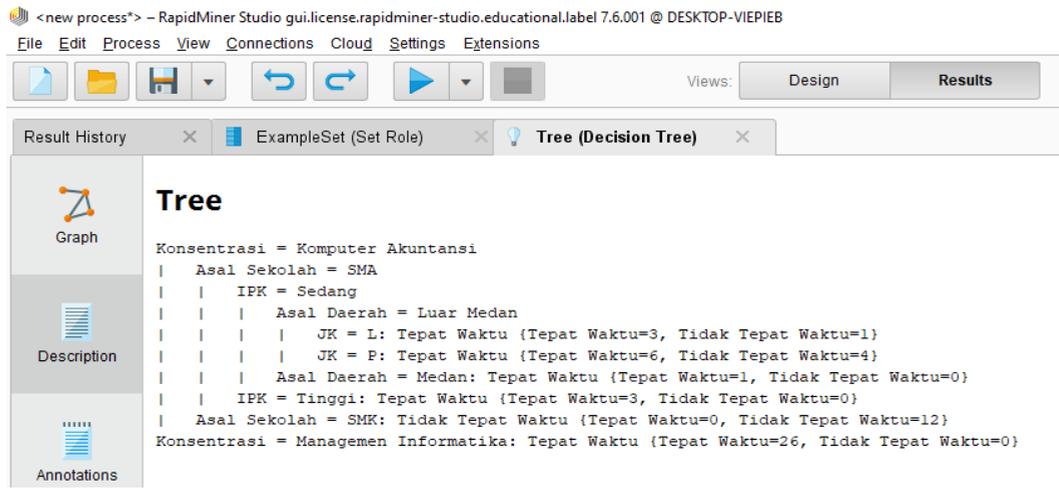


Gambar 4.8 Decision Tree algoritma C4.5 Pada *Main Process*

Adapun hasil yang diperoleh dari pengujian bentuk graph dan description rule decision tree algoritma C4.5 seperti gambar di bawah ini:



Gambar 4.9 Graph Decision Tree algoritma C 4.5



Gambar 4.10 Hasil Rule Decision Tree algoritma C4.5

4.7. Penerapan

Penerapan data mining dalam menentukan masa studi mahasiswa menggunakan algoritma C4.5 di STMIK Budi Darma Program Studi Manajemen Informatika (D3). Dengan adanya penelitian dalam menentukan masa studi mahasiswa STMIK Budi Darma Program Studi Manajemen Informatika (D3), jika masa studi pada mahasiswa diketahui lebih sedikit maka pada pihak jurusan Program Study Mahasiswa Managemen Informatika (D3) dilakukan pada tindakan agar mahasiswa dapat lulus dengan waktu kurang dari lima tahun, dan tepat pada lima tahun sesuai dengan Peraturan Menteri Pendidikan.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan penelitian yang dilakukan selama membuat sistem ini, maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Pada penerapan *Data Mining* dengan metode yang digunakan Algoritma C4.5 yang menghasilkan sebuah informasi untuk prediksi masa studi secara tepat waktu di universitas STMIK Budi Darma pada Program Studi Manajemen Informatika (D3) hingga tingkat akurasi yang dihasilkan 755 dengan jumlah 56 data tes
2. Penerapan data mining dalam menentukan masa studi mahasiswa menggunakan algoritma C4.5 di STMIK Budi Darma Program Studi Manajemen Informatika (D3) sehingga mempermudah pihak kampus dalam mengatur strategi manajemen untuk kedepannya agar mahasiswa bisa lulus tepat waktu.
3. Penerapan dari hasil data mining yang ditentukan dari masa Study Mahasiswa sehingga dapat kita ketahui mahasiswa “Lulus secara Tepat Waktu” untuk mahasiswa yang mempunyai nilai IPK tinggi dan “Lulus Secara Tidak Tepat Waktu” mahasiswa memiliki nilai IPK sedang dengan konsentrasi jurusan komputer akutansi dan lulusan SMK berdasarkan Algoritma C4.5 yang dilakukan oleh peneliti.

5.2. Saran

Adapun saran untuk pengembangan aplikasi mendatang adalah :

1. Penerapan pada *Data Mining* terhadap penelitian ini yang menggunakan metode Algoritma C4.5 yang berfungsi sbagai perbandingan dengan Algoritma lainnya dengan kasus yang sama. Hingga dapat diketahui

Algoritma mana yang lebih baik digunakan agar dapat menentukan masa studi mahasiswa

2. Analisa yang dihasilkan pada penelitian ini masih merupakan analisa mendasar dan belum sempurna sehingga perlu dilakukan pengembangan yang lebih spesifik oleh penelitian selanjutnya
3. Sebaiknya untuk pengujian pada algoritma C4.5 ataupun algoritma data mining lainnya dapat digunakan tools pengujian yang lain.

DAFTAR PUSTAKA

- A. I. G. Heni Sulastri, "Jurnal Teknologi dan Sistem Informasi PENERAPAN DATA MINING DALAM PENGELOMPOKAN PENDERITA THALASSAEMIA," *Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 299–305, 2017.
- Andri, Y. N. Kunang, and S. Murniati, "Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Pada Universitas Bina Darma Palembang," *Semin. Nas. Inform. 2013 (semnasIF 2013)*, vol. 2013, no. A, pp. 56–63, 2013.
- Defiyanti, Sofi; Pardede, D.L.Crispina;,(n.d.). "PERBANDINGAN KINERJA ALGORITMA ID3 DAN C4.5".1-5.
- Darmawan, Hadi Dwi, Desi Yuniarti, and Yuki Novia Nasution. "Klasifikasi Lama Masa Studi Mahasiswa Menggunakan Perbandingan Metode Algoritma C.45 dan Algoritma Classification and Regression Tree." *Jurnal Eksponensial* Vol.8, No. 2, 2017, pp.151-160.
- Dosen pendidikan. (2015). 161 Pengertian Bahasa Menurut Para Ahli. (<http://www.dosenpendidikan.com/50-pengertian-bahasa-menurut-para-ahli/>), tanggal akses: 27 Maret 2016. Eka Listiana, Much Aziz Muslim, "Penerapan *Adaboost* Untuk *Klasifikasi Support Vector Machine* Guna Meningkatkan Akurasi Pada Diagnosa *Chronic Kidney Disease*," "Fakultas Teknik – Universitas Muria Kudus 875," no. 2015, pp. 875–881, 2017.
- Eko Prasetyo, "*Data Mining Konsep dan Aplikasi menggunakan Matlab*". Yogyakarta, 2012.
- Elsa Paskalis Krisda Orpa, Eva Faja Ripanti, Tursina "Model Prediksi Awal Masa Studi Mahasiswa Menggunakan Algoritma Decision tree c4.5" *Jurnal Sistem dan Teknologi Informasi* Vol. 7, No. 4
- J. Arief and R. Hakim, "APLIKASI DATA MINING UNTUK MENGHASILKAN POLA KELULUSAN SISWA Budanis Dwi Meilani , Nofi Susanti," pp. 1–7, 2015
- Kamagi, David Hartanto., & Hansun, Seng. (2014). Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. *ULTMATICS*, Vol: VI, No: 1, ISSN: 2085-4552. Universitas Multimedia Nusantara Tangerang.

- Listiana, Mila., Sudjalwa, Gunawan, Dedi. (2015). Perbandingan Algoritma Decision Tree (C4.5) Dan Naïve Bayes Pada Data Mining Untuk Identifikasi Tumbuh Kembang Anak Balita (Studi Kasus: Puskesmas Kartasura). Naskah Publikasi. Universitas Muhammadiyah Surakarta. Juli 2015.
- L. Dini, U. Sekolah, T. M. Informatika, D. Komputer, N. Mandiri, and R. S. Wahono, "Integrasi Metode Information Gain Untuk Seleksi Fitur dan Adaboost Untuk Mengurangi Bias Pada Analisis Sentimen Review Restoran Menggunakan Algoritma Naïve Bayes," *J. Intell. Syst.*, vol. 1, no. 2, pp. 120–126, 2015.
- R. Agus, Suprapedi, H. Himawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma Decision Tree", *Jurnal Teknologi Informasi*, Vol.13 No.1, Januari 2017.
- RABIT (Jurnal Teknologi dan Sistem Informasi Univrab) ISSN CETAK : 2477-2062 VOL. 2 No. 2, Juli 2017 ISSN ONLINE : 2502-891X233
- Sunjana, 2010, Aplikasi Data Mining Data Mahasiswa dengan Metode Klasifikasi Decision Tree, SNATI 2010, Yogyakarta, 19 Juni 2010.
- Swastina, Liliana, 2013, Penerapan Algoritma C4.5 untuk Penentuan Prodi Mahasiswa, *Jurnal GEMA AKTUALITA*, No.1, Vol.2, 93-98.
- Y. Elmande dan P.P. Widodo, "Pemilihan Criteria Splitting Dalam Algoritma Iterative Dichotomiser 3 (ID3) Untuk Penentuan Kualitas Beras : Studi Kasus Pada Peru Bulog Divre Lampung," *Jurnal TELEMATIKA MKOM*, vol. 4, no. 1, ISSN :2085-725X, Maret 2012.

LAMPIRAN 1

SURAT KETERANGAN PENELITIAN

Medan, 11 Maret 2020

Nomor : 076/STMIK-BD/EXT/III.2020
Lamp : -
Hal : Izin Riset

Kepada Yth,
Dekan Fak. Sains dan Teknologi
Univ. Islam Negeri Sumatera Utara Medan
Di
Tempat

Dengan Hormat,
Berdasarkan surat dari Dekan Fak.Sains dan Teknologi UINSU Medan, tentang permohonan Izin dari Fak Sains dan Teknologi UINSU Medan di STMIK Budi Darma atas nama:

No	NIM	Nama Mahasiswa	Asal Perguruan Tinggi
1	71154061	Wirdana Rauda Ningsih	Univ. Islam Negeri Sumatera Utara Medan

Maka, memberikan izin kepada mahasiswa tersebut di atas untuk melaksanakan riset di STMIK Budi Darma Medan dengan tujuan mendapatkan data-data berupa sampel sesuai dengan topik penelitian yang dibahas, dengan syarat mahasiswa tersebut mengikuti aturan dan mekanisme yang berlaku di lingkungan STMIK Budi Darma Medan.

Demikian Surat keterangan ini, dibuat dengan sebenarnya agar dapat dipergunakan sebagaimana mestinya.


STMIK BUDI DARMA
Ketua
Mesran S.Kom, M.Kom
NIDN : 0124087801

LAMPIRAN 2
DAFTAR RIWAYAT HIDUP



I. Data Pribadi

Nama : Wirdana Rauda Ningsih
Tempat/ Tanggal Lahir : Medan, 15 Juni 1997
Alamat : Dusun III Pematang Guntung
Agama : Islam
Jenis Kelamin : Perempuan
Status : Belum Menikah
No. Hp : 0852-7513-3102
Email : wirdanarauda176@gmail.com

II. Pendidikan formal

1. 2002 – 2003 RA AR-Ridho – Sialang Buah
2. 2003 – 2009 SD Negeri 105406 Pematang Guntung
3. 2009 – 2012 Pondok Pesantren Darul Mukhlisin – Sei Rampah
4. 2012 – 2015 SMK Negeri 1 Pantai Cermin
5. 2015 – 2020 Universitas Islam Negeri Sumatera Utara, Fakultas Sains dan Teknologi, Program Studi IlmuKomputer S1

LAMPIRAN 3
KARTU BIMBINGAN SKRIPSI

KARTU BIMBINGAN SKRIPSI
Semester Gasal/Genap Tahun Akademik /

Nama : <u>WIJAPANA RAUDA N</u>	Pembimbing I : <u>Dr. MAD Furqon S.Si. M.com.P.Sc</u>
NIM : <u>71154061</u>	Pembimbing II : <u>Mhd Ikhsan M.Wom</u>
Prog. Studi : <u>UMU Komputer</u>	SK Pembimbing :
Judul Skripsi : <u>Penerapan Data Mining dalam Menentukan Masa Depan Mahasiswa menggunakan Algoritma C4.5</u>	

P E R T	PEMBIMBING I			PEMBIMBING II		
	Tgl.	Materi Bimbingan	Tanda Tangan	Tgl.	Materi Bimbingan	Tanda Tangan
I	<u>03/02/2020</u>	<u>Revisi BAB I sampai III</u>		<u>02/10/2019</u>	<u>- Revisi Judul proposal - Revisi Penulisan latar belakang, rumusan masalah, tujuan dan manfaat penelitian</u>	
II	<u>11/02/2020</u>	<u>DISKUSI BAB I sampai BAB III</u>		<u>23/10/2019</u>	<u>- ACC BAB I Lanjutan BAB II</u>	
III	<u>26/02/2020</u>	<u>ACC seminar Proposal</u>		<u>18/11/2019</u>	<u>- Perbaikan penulisan pada BAB II - Penyesuaian tahapan metode sesuai teori pemodelan sistem dan teori yang digunakan</u>	
IV	<u>23/10/2020</u>	<u>DISKUSI dan REVISI BAB IV</u>		<u>02/12/2019</u>	<u>- ACC BAB II - Lanjut ke BAB III</u>	
V	<u>07/10/2020</u>	<u>Penambahan Daftar Pustaka</u>		<u>16/12/2019</u>	<u>- Menjelaskan diagram Kasa Penelitian pada BAB III</u>	

VI	24/11/2020	acc keseluruhan skripsi akhir	01/01/2020	-Perbaiki daftar Pustaka, kata Pengantar, dan daftar isi	
VII			21/01/2020	Acc Keseluruhan proposal skripsi	
VIII			21/09/2020	Revisi BAB IV	
IX			01/10/2020	DISKUSI BAB V	
X			21/10/2020	acc keseluruhan skripsi akhir	

Medan, 24 oktober 2020.
 An. Dekan
 Ketua Jurusan/Program Studi

.....

 NIP.

Catatan: Pada saat bimbingan, kartu ini harus diisi dan ditandatangani oleh pembimbing