

**PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN  
METODE *DECISION TREE* DENGAN PENERAPAN  
ALGORITMA C4.5**

**SKRIPSI**

oleh :

**SUSI MASHLAHAH**

**NIM. 07650150**



**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2013**

**PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN  
METODE *DECISION TREE* DENGAN PENERAPAN  
ALGORITMA C4.5**

**SKRIPSI**

**Diajukan Kepada :  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk Memenuhi Salah Satu Persyaratan Dalam  
Memperoleh Gelar Sarjana Komputer (S.Kom)**

**Oleh :  
SUSI MASHLAHAH  
NIM. 07650150**

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2013**

**HALAMAN PERSETUJUAN**

**PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN  
METODE *DECISION TREE* DENGAN PENERAPAN  
ALGORITMA C4.5**

**SKRIPSI**

Oleh:  
**SUSI MASHLAHAH**  
NIM: 07650150

Telah Disetujui,  
Malang 16 Januari 2013

Dosen Pembimbing I

Dosen Pembimbing II

**M. AINUL YAQIN, M.Kom**

NIP. 197610132006041004

**MUHAMMAD FAISAL, M. T**

NIP. 19740502005011007

Mengetahui:

Ketua Jurusan Teknik Informatika

**RIRIEN KUSUMAWATI, M. Kom**

NIP. 197203092005012002

**HALAMAN PENGESAHAN****PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN METODE  
DECISION TREE DENGAN PENERAPAN ALGORITMA C4.5****SKRIPSI**

Oleh:

**Susi Mashlahah  
NIM. 07650150**

Telah Dipertahankan Di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Sarjana Teknik Informatika (S.Kom)

Tanggal, 15 Januari 2013

<b>Susunan Dewan Penguji:</b>	<b>Tanda Tangan</b>
1. <b>Penguji Utama</b> : Fatchurrochman, M.Kom NIP. 197007312005011002	( )
2. <b>Ketua Penguji</b> : Syahiduz Zaman, M.Kom NIP. 197005022005011005	( )
3. <b>Sekretaris Penguji</b> : M. Ainul Yaqin, M.Kom NIP. 197610132006041004	( )
4. <b>Anggota Penguji</b> : Muhammad Faisal, M.T NIP. 197405102005011007	( )

Mengetahui dan Mengesahkan  
Ketua Jurusan Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang

**Ririen Kusumawati, M.Kom**  
NIP. 197203092005012002

## SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini saya:

Nama : Susi Mashlahah

NIM : 07650150

Fakultas/Jurusan : Sains dan Teknologi/Teknik Informatika

Judul Penelitian : Prediksi Kelulusan Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritma C4.5

Menyatakan dengan sebenar-benarnya bahwa skripsi yang saya buat tidak terdapat unsur-unsur penjiplakan karya penelitian atau karya ilmiah yang pernah dilakukan atau dibuat oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata hasil penelitian ini terbukti terdapat unsur-unsur jiplakan, maka saya bersedia untuk mempertanggungjawabkan, serta diproses sesuai peraturan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya dan tanpa paksaan dari siapapun.

Malang, 16 Januari 2013

Yang Membuat Pernyataan,

Susi Mashlahah

NIM: 07650150

## MOTTO

لَا تُؤَخِّرْ عَمَلَكَ إِلَى الْغَدِ مَا تَقْدِرُ أَنْ تَعْمَلَهُ الْيَوْمَ

*“Janganlah mengakhirkan pekerjaanmu hingga esok hari, yang kamu dapat  
mengejakannya hari ini”*



## HALAMAN PERSEMBAHAN

*Karya ini saya persembahkan kepada :*

*Kedua orang tua yang telah mempercayaku, menyayangiku dan mendo'akanku,  
ibunda Hj. Sulistiani dan ayahanda H. Sukirman*

*Kedua kakakku Hajat Aris Sulistiawan dan Taufiqurrohman,  
Yangtelah memberikan tauladan yang baik serta memotivasi diriku untuk  
senantiasa mencapai kesuksesan dan kebaikan.*

*Teman, rekan dan sahabatku di UIN Maliki Malang,  
terutama teman-teman jurusan Teknik Informatika,  
sukses selalu untuk semuanya.*

*Kepada setiap orang yang membantuku menjalani kehidupan ini  
hingga terhindar dari kesia-siaan*

## KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Alhamdulillah, segala puji dan syukur dengan tulus kami persembahkan ke hadirat Allah SWT, karena hanya dengan petunjuk dan hidayah-Nya peneliti mampu menyelesaikan tugas akhir yang berjudul *Prediksi Kelulusan Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritma C4.5*

Shalawat serta salam peneliti haturkan pada junjungan Nabi Muhammad SAW yang memberikan motivasi bagi umat Islam, khususnya bagi peneliti untuk selalu berproses menuju insan yang memiliki intelektualitas tinggi dan berakhlak mulia.

Penyelesaian skripsi ini merupakan suatu pekerjaan sangat berat bagi peneliti yang fakir ilmu, namun berkat ma'unnah Allah SWT dan bantuan dari berbagai pihak baik berupa materiil maupun moril, akhirnya tugas akhir ini dapat terselesaikan dengan baik. Oleh karena itu peneliti menyampaikan rasa hormat, ungkapkan terima kasih serta penghargaan setinggi-tingginya kepada:

1. Prof. Dr. H. Imam Suprayogo, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sutiman Bambang Sumitro SU., DSc. selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Ririen Kusumawati, M.Kom. selaku Ketua Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

4. Ainul Yaqin, M.Kom, dosen pembimbing I yang dengan sabar memberikan arahan, saran dan motivasi pada peneliti sehingga skripsi ini dapat terselesaikan dengan baik.
5. Mohammad Faisal, M.T selaku Pembimbing II yang telah membimbing penulis dalam menyelesaikan skripsi ini.
6. Ibu Sulistiani dan Bapak Sukirman yang setiap waktu bersujud dan berdo'a demi kelancaran dan kesuksesan penulisan skripsi ini hingga tercapai cita-cita anaknya.
7. Seluruh Dosen yang telah mengajarkan banyak hal dan selalu memberikan semangat untuk terus berproses hingga akhir perkuliahan peneliti.

Skripsi ini tentunya tiada terbebas dari kekurangan dan kelemahan. Oleh karena itu peneliti mengharapkan kritik dan saran penyempurna untuk perbaikan di masa mendatang. Penulis berharap semoga skripsi ini bisa dibaca oleh banyak orang, terutama civitas akademika Universitas Islam Negeri Maulana Malik Ibrahim Malang. Selain itu peneliti berharap semoga skripsi ini dapat memberikan nilai guna baik bagi peneliti maupun bagi pembaca. Amin Ya Robbal'Alamin.

Malang, 16 Januari 2013  
Peneliti,

Susi Mashlahah

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>ii</b>
<b>HALAMAN PERSETUJUAN .....</b>	<b>iii</b>
<b>HALAMAN PENGESAHAN .....</b>	<b>iv</b>
<b>HALMAN PERNYATAAN .....</b>	<b>v</b>
<b>MOTTO .....</b>	<b>vi</b>
<b>HALAMAN PERSEMBAHAN .....</b>	<b>vii</b>
<b>KATA PENGANTAR .....</b>	<b>viii</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR GAMBAR.....</b>	<b>xii</b>
<b>DAFTAR TABEL .....</b>	<b>xiv</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>xv</b>
<b>ABSTRAK .....</b>	<b>xvi</b>
<b>BAB I PENDAHULUAN</b>	
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	6
1.3 Batasan Masalah.....	6
1.4 Tujuan Penelitian.....	7
1.5 Manfaat Penelitian .....	7
1.6 Metode Penelitian .....	7
1.7 Sistematika Penulisan .....	9
<b>BAB II TINJAUAN PUSTAKA</b>	
2.1 Decision Tree.....	10
2.2 Algoritma C4.5 .....	13
2.3 Penelitian Terkait .....	21
<b>BAB III ANALISIS DAN PERANCANGAN</b>	
3.1 Gambaran Umum Sistem .....	25
3.2 Analisa Data .....	26
3.3 Alat dan Bahan yang diperlukan .....	27

3.4 Deskripsi Sistem .....	28
3.5 Perancangan Sistem .....	34
3.5.1 Ekstraksi Dokumen.....	34
3.5.2 Analisa .....	38
3.5.3 Konteks Diagram .....	39
3.5.4 Data Flow Diagram.....	39
3.5.5 Entity Relational Diagram .....	41
3.5.6 Desain Database.....	42
3.6 Perancangan Antar Muka .....	44
3.7 Perancangan Uji Coba .....	48
3.7.1 Tujuan Uji Coba.....	48
3.7.2 Skenario Uji Coba.....	49
<b>BAB IV HASIL DAN PEMBAHASAN</b>	
4.1 Deskripsi Program .....	50
4.2 Impelementasi Sistem .....	60
4.2.1 Proses Perhitungan Data Training.....	61
4.2.2 Membuat Total Entropy .....	62
4.2.3 Membuat Gain .....	63
4.2.4 Menentukan Gain maksimal .....	63
4.2.5 Menentukan Node Baru .....	64
4.2.6 Perulangan Proses Perhitungan.....	65
4.2.7 Proses Rule.....	67
4.2.8 Perhitungan Target.....	68
4.3 Hasil Uji Coba .....	69
4.4 Hubungan Penelitian Dengan Kaidah Islam .....	73
<b>BAB V PENUTUP</b>	
5.1 Kesimpulan .....	77
5.2 Saran .....	77
<b>DAFTAR PUSTAKA.....</b>	<b>79</b>
<b>LAMPIRAN-LAMPIRAN .....</b>	<b>81</b>

## DAFTAR GAMBAR

Gambar 2.1 Konsep Decision Tree .....	11
Gambar 2.2 Pohon Keputusan Hasil Perhitungan Node 1 .....	17
Gambar 2.3 Pohon Keputusan Hasil Perhitungan Node 1.1 .....	19
Gambar 2.4 Pohon Keputusan Hasil Perhitungan Node 1.1.2 .....	20
Gambar 3.1 Blok Diagram Sistem .....	29
Gambar 3.2 Flowchart Penentuan Root .....	31
Gambar 3.3 Flowchart Penentuan Cabang .....	32
Gambar 3.4 Flowchart Penentuan Node .....	33
Gambar 3.5 Sitemap Aplikasi .....	34
Gambar 3.6 Konteks Diagram .....	39
Gambar 3.7 DFD Level 1 .....	39
Gambar 3.8 DFD Level 2 Proses Admin .....	40
Gambar 3.9 DFD Level 2 Proses Operator .....	40
Gambar 3.10 ERD Conceptual .....	41
Gambar 3.11 ERD Physical .....	42
Gambar 3.12 Halaman Login.....	44
Gambar 3.13 Halaman Utama.....	44
Gambar 3.14 Halaman Training.....	45
Gambar 3.15 Halaman Entropy dan Gain.....	45
Gambar 3.16 Halaman Rule.....	46
Gambar 3.17 Halaman Target.....	47
Gambar 3.18 Halaman Laporan.....	48
Gambar 4.1 Halaman Login.....	50
Gambar 4.2 Halaman Utama .....	51
Gambar 4.3 Halaman Training .....	52
Gambar 4.4 Halaman Tampil Data Training .....	53
Gambar 4.5 Halaman Import Data Training .....	53
Gambar 4.6 Halaman Rule .....	54
Gambar 4.7 Halaman Operator .....	56
Gambar 4.8 Halaman Target .....	57

Gambar 4.9 Halaman Hasil Prediksi Tidak Lulus Tepat Waktu.....	58
Gambar 4.10 Halaman Hasil Prediksi Lulus Tepat Waktu.....	58
Gambar 4.11 Halaman Multi Target .....	59
Gambar 4.12 Halaman Laporan .....	60
Gambar 4.13 Proses Prediksi .....	69



## DAFTAR TABEL

Tabel 2.1	Contoh Data Sampel.....	14
Tabel 2.2	Perhitungan Node 1 .....	16
Tabel 2.3	Perhitungan Node 1.1 .....	18
Tabel 2.4	Perhitungan Node 1.1.2 .....	19
Tabel 3.1	Predikat Kelulusan.....	37
Tabel 3.2	Tabel Desain Input .....	40
Tabel 3.3	Tabel Desain Output .....	41
Tabel 3.2	Tabel Desain Proses .....	41
Tabel 3.5	Tabel Training .....	42
Tabel 3.6	Tabel Hasil Target .....	42
Tabel 3.7	Tabel Multi .....	43
Tabel 3.8	Tabel Tes .....	43
Tabel 3.9	Tabel Rule .....	43
Tabel 3.10	Tabel User .....	43
Tabel 4.1	Tabel Aturan (Rule) Hasil Uji Coba.....	70
Tabel 4.2	Tabel Hasil Pengujian Tingkat Kebenaran .....	71

## DAFTAR LAMPIRAN

Lampiran 1 : Perhitungan Manual



## ABSTRAK

Mashlahah, Susi. 2013. **Prediksi Kelulusan Mahasiswa Menggunakan Metode *Decision Tree* Dengan Penerapan Algoritma C4.5**. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) M. Ainul Yaqin, M.Kom. (II) Mohammad Faisal, M.T.

**Kata kunci :** Klasifikasi, *Decision Tree*, Algoritma C4.5

Universitas Islam Negeri Maulana Malik Ibrahim Malang merupakan Perguruan Tinggi Negeri di Indonesia yang setiap tahunnya kuota mahasiswa yang diterima makin bertambah, namun tidak semua mahasiswa dapat lulus tepat waktu sesuai dengan masa studi yang ditempuh sehingga mengakibatkan penumpukan jumlah mahasiswa yang tidak lulus sesuai dengan masa periode kelulusannya.

Berdasarkan latar belakang tersebut penelitian dilakukan untuk membuat sebuah sistem menggunakan teknik klasifikasi yang dapat mengolah data dalam jumlah besar untuk menemukan pola yang terjadi pada data mahasiswa. Pengolahan data tersebut digunakan untuk memprediksi kelas yang belum diketahui yaitu prediksi kelulusan mahasiswa. Teknik klasifikasi yang digunakan yaitu *decision tree* dengan penerapan algoritma C4.5.

Inputan yang digunakan yaitu berupa atribut dari data mahasiswa meliputi asal daerah, jenis sekolah, jalur masuk, pengalaman pesantren, indeks prestasi kumulatif (IPK), dan indeks prestasi per semester mulai dari semester 1 sampai 5. Data mahasiswa tersebut merupakan data sampel training yang digunakan dalam penyusunan *decision tree*. Berdasarkan pada pengujian menggunakan data mahasiswa yang sudah lulus dari tahun 2005 sampai 2008, akurasi kecocokan pada sistem ini mencapai 82,79% sehingga dapat digunakan untuk memprediksi kelulusan mahasiswa yang belum diketahui.

## ABSTRACT

Mashlahah, Susi. 2013. **Prediction of Students Graduation Using *Decision Tree* Method with the Implementation of Algorithm C4.5**. Thesis. Major Study of Informatics Engineering, Faculty of Science and Technology, Islamic State University of Maulana Malik Ibrahim of Malang. Supervisors: (I) M. Ainul Yaqin, M.Kom. (II) Mohammad Faisal, M.T.

**Keywords:** Classification, *Decision Tree*, Algorithm C4.5

Islamic State University of Maulana Malik Ibrahim of Malang is a public college in Indonesia that receives more increased quota of students year to year, but all of students are not able to graduate just in time in accordance with study period pursued so it results in accumulation of the number of students who are not graduated in accordance with the graduation periods.

Based on that background, the research is conducted to create a system using technique of classification which can manage data in big number to meet with pattern in students data. The data processing is used to predict unknown-yet class, that is the prediction of students graduation. Technique of classification used is *decision tree* with the implementation of algorithm C4.5.

Input used is an attribute of students data including the origin of region, types of school, way of university entrance, experience of pesantren, accumulative grade point average (IPK), and grade point average of each semester, from first semester up to fifth semester. That students data is training sample data used in arranging *decision tree*. Based on the testing that use graduated students data from 2005 up to 2008, accuracy of compatibility in this system reaches 82,79%, so it can be used to predict unknown-yet students graduation.

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Jurusan Teknik Informatika Fakultas Sains Dan Teknologi di UIN Maulana Malik Ibrahim Malang pada tahun 2004 memiliki jumlah mahasiswa sebanyak 89, sedangkan yang dapat lulus tepat waktu yaitu sebanyak 22 mahasiswa, sehingga diperoleh prosentase kelulusan sebesar 24% pada periodenya. Pada tahun 2005 memiliki jumlah mahasiswa sebanyak 137, sedangkan yang dapat lulus tepat waktu yaitu sebanyak 17 mahasiswa, sehingga diperoleh prosentase kelulusan sebesar 12% pada periodenya. Pada tahun 2006 memiliki jumlah mahasiswa sebanyak 129, sedangkan yang dapat lulus tepat waktu yaitu sebanyak 9 mahasiswa, sehingga diperoleh prosentase kelulusan sebesar 6% pada periodenya. Pada tahun 2007 memiliki jumlah mahasiswa sebanyak 156, sedangkan yang dapat lulus tepat waktu yaitu sebanyak 9 mahasiswa, sehingga diperoleh prosentase kelulusan sebesar 5% pada periodenya. Pada tahun 2008 memiliki jumlah mahasiswa sebanyak 158, sedangkan yang dapat lulus tepat waktu yaitu sebanyak 20 mahasiswa, sehingga diperoleh prosentase kelulusan sebesar 12% pada periodenya. Berdasarkan prosentase diatas dapat diketahui nilai rata-rata dari prosentase kelulusan selama 5 tahun yaitu sebesar 12%.

Dari keterangan tersebut diketahui bahwa tiap memasuki tahun ajaran baru kuota mahasiswa yang diterima makin bertambah, namun tidak semua mahasiswa dapat lulus tepat waktu sesuai dengan masa studi yang ditempuh sehingga mengakibatkan jumlah mahasiswa semakin bertambah banyak. Hal tersebut berdampak pula pada peningkatan jumlah data

pribadi dan data akademis dari semua mahasiswa yang masih terdaftar. Data dalam jumlah besar tersebut apabila diolah menggunakan beberapa metode khusus akan memberikan informasi baru yang dapat dimanfaatkan dalam hal lain.

Jumlah mahasiswa yang lulus dan mahasiswa baru yang masuk setiap tahunnya tidak sebanding, untuk itu perlu adanya suatu sistem yang dapat digunakan untuk memprediksi kelulusan mahasiswa. Sistem prediksi kelulusan mahasiswa ini memerlukan informasi yang ada untuk mengetahui apakah seorang mahasiswa dapat lulus tepat waktu atau tidak. Apabila kelulusan mahasiswa dapat diketahui sejak dini maka pihak akademik dapat menerapkan suatu kebijakan untuk meminimalisir jumlah mahasiswa yang tidak lulus tepat waktu sesuai dengan masa studinya.

Sistem prediksi kelulusan ini menggunakan metode data mining dalam pencarian informasi dari data berskala besar. Data mining bertujuan untuk mendapatkan informasi yang berguna dari sekumpulan dokumen dan dapat membantu pihak-pihak tertentu yang membutuhkan informasi tersebut.

Dengan memanfaatkan informasi tersebut suatu instansi ataupun perguruan tinggi dapat bersaing dan menjadi lebih maju. Suatu sistem informasi merupakan salah satu sumberdaya yang sangat penting dalam meningkatkan keunggulan persaingan karena sistem informasi adalah gabungan dari manusia, perangkat keras, perangkat lunak, jaringan komunikasi, sumber data, kebijakan dan prosedur yang terorganisir yang mampu menyimpan, mengambil dan mendistribusikan informasi di dalam sebuah organisasi. System informasi merupakan amanah yang harus dijaga dan disampaikan secara benar. Dalam Firman Allah dalam QS. An-Nisaa' : 58 dijelaskan bahwa :

﴿ إِنَّ اللَّهَ يَأْمُرُكُمْ أَنْ تُؤَدُّوا الْأَمَانَاتِ إِلَىٰ أَهْلِهَا وَإِذَا حَكَمْتُمْ بَيْنَ النَّاسِ أَنْ تَحْكُمُوا بِالْعَدْلِ إِنَّ اللَّهَ نِعِمَّا يَعِظُكُمْ بِهِ إِنَّ اللَّهَ كَانَ سَمِيعًا بَصِيرًا ﴾

“*Sesungguhnya Allah menyuruh kamu menyampaikan amanat kepada yang berhak menerimanya, dan (menyuruh kamu) apabila menetapkan hukum di antara manusia supaya kamu menetapkan dengan adil. Sesungguhnya Allah memberi pengajaran yang sebaik-baiknya kepadamu. Sesungguhnya Allah adalah Maha mendengar lagi Maha Melihat.*” (QS. An-Nisaa’ : 58)

Ayat ini turun ketika Ali mengambil kunci Ka’bah dari Utsman bin Thalhah Al-Hajabi, pejaga Ka’bah, secara paksa ketika Nabi SAW. datang ke Makkah pada waktu penaklukan kota Makkah, tetapi dia menolaknya dan berkata: “*Seandainya aku tahu bahwa dia adalah Rasulullah SAW. aku tidak akan menolaknya.*” Kemudian Rasulullah SAW. memerintahkan agar kunci itu dikembalikan padanya dan bersabda : “*Ambillah untuk selamanya dan seterusnya.*” Lalu dia merasa takjub akan hal itu. Lantas Ali membacakan ayat tersebut kepadanya dan dia langsung masuk Islam. Dan ketika meninggal dunia, dia memberikan kunci itu kepada adiknya, Syaibah, kemudian terus dipegang oleh anaknya. Kendati ayat ini turun karena ada sebab yang khusus, tetapi keumumannya bisa dijadikan sebagai pegangan. Sebaik-baik sesuatu ialah menyampaikan amanat dan menetapkan hukum dengan adil.

Ada pula hadits yang menjelaskan tentang penyampaian amanat, salah satunya yaitu dari Abdullah bin Amr *radhiyallahu ta’ala ‘anhu*, bahwa Nabi *shallallahu ‘alaihi wa sallam* bersabda,

بَلِّغُوا عَنِّي وَلَوْ آيَةً

“*Sampaikanlah dariku walau hanya satu ayat*” (HR Al-Bukhari : 3274)

Ada beberapa poin kandungan dalam hadits diatas, pertama yaitu Nabi SAW. memerintahkan untuk menyampaikan perkara agama dari beliau, karena Allah SWT telah

menjadikan agama ini sebagai satu-satunya agama bagi manusia dan jin. Kedua menyampaikan dalil dari Al Qur'an atau sebagiannya dan dari As Sunnah, baik sunnah yang berupa perkataan (*qauliyah*), perbuatan (*amaliyah*), maupun persetujuan (*taqririyah*), dan segala hal yang terkait dengan sifat dan akhlak mulia Nabi *shallallahu 'alaihi wa sallam*. Menyampaikan dapat juga secara makna dan pemahaman terhadap nash-nash yang ada. Ketiga, yaitu penyampaian ilmu, ilmu yang disampaikan harus benar dan hendaknya disertai penguasaan yang baik agar bebas dari kerancuan atau kesalah fahaman terhadapnya.

Pada kesempatan kali ini peneliti membuat aplikasi data mining menggunakan metode *Decision Tree* dengan penerapan algoritma C4.5 yang akan digunakan dalam memprediksi kelulusan mahasiswa. Aplikasi ini akan diimplementasikan pada mahasiswa Jurusan Teknik Informatika Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang. Dengan adanya aplikasi ini diharapkan dapat menjadi sebuah informasi penting untuk bisa meningkatkan kualitas mahasiswa di UIN Maulana Malik Ibrahim Malang sehingga nantinya siap menjadi profesional-profesional muda yang tidak hanya intelek namun memiliki keagungan akhlak yang luhur sesuai dengan visi UIN Maulana Malik Ibrahim Malang yang berpijak pada (QS. Al Imran : 190-191)

إِنَّ فِي خَلْقِ السَّمَوَاتِ وَالْأَرْضِ وَأَخْتِلَافِ اللَّيْلِ وَالنَّهَارِ لَآيَاتٍ لِأُولِي الْأَلْبَابِ ﴿١٩٠﴾ الَّذِينَ يَذْكُرُونَ اللَّهَ قِيَمًا وَقُعُودًا وَعَلَىٰ جُنُوبِهِمْ وَيَتَفَكَّرُونَ فِي خَلْقِ السَّمَوَاتِ وَالْأَرْضِ رَبَّنَا مَا خَلَقْتَ هَذَا بَطْلًا سُبْحَانَكَ فَقِنَا عَذَابَ النَّارِ ﴿١٩١﴾

Artinya : “*Sesungguhnya dalam penciptaan langit dan bumi, dan silih bergantinya malam dan siang terdapat tanda-tanda bagi orang-orang yang berakal. (yaitu) orang-orang yang mengingat Allah sambil berdiri atau duduk atau dalam keadan berbaring dan mereka memikirkan tentang penciptaan langit dan bumi (seraya berkata): "Ya Tuhan kami, tiadalah Engkau menciptakan ini dengan sia-sia, Maha Suci Engkau, maka peliharalah kami dari siksa neraka.*”(QS. Ali Imron:190-191)

“*Sesungguhnya pada penciptaan langit dan bumi*” dan keajaiban-keajaiban yang terdapat pada keduanya “*serta pergantian malam dan siang*” dengan datang dan pergi serta bertambah dan berkurang “*menjadi tanda-tanda*” atau bukti-bukti atas kekuasaan Allah swt. “*bagi orang-orang yang berakal*” artinya yang mempergunakan pikiran mereka. “*Yakni orang-orang yang*” menjadi 'na`at' atau badal bagi yang sebelumnya “*mengingat Allah di waktu berdiri dan duduk dan ketika berbaring*” artinya dalam keadaan bagaimana pun juga sedang menurut Ibnu Abbas mengerjakan salat dalam keadaan tersebut sesuai dengan kemampuan “*dan mereka memikirkan tentang kejadian langit dan bumi*” untuk menyimpulkan dalil melalui keduanya akan kekuasaan Allah, kata mereka: “*Wahai Tuhan kami! Tidaklah Engkau ciptakan ini*” maksudnya makhluk yang kami saksikan ini “*dengan sia-sia*” menjadi hal sebaliknya semua ini menjadi bukti atas kesempurnaan kekuasaan-Mu “*Maha Suci Engkau*” artinya tidak mungkin Engkau akan berbuat sia-sia “*maka lindungilah kami dari siksa neraka*”.

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang di atas maka diperoleh rumusan masalah sebagai berikut:

Bagaimana memprediksi kelulusan mahasiswa menggunakan metode *Decision Tree* dengan penerapan algoritma C4.5?

## **1.3. Batasan Masalah**

Batasan masalah pada penelitian ini adalah :

1. Ruang lingkup penelitian dibatasi pada data mahasiswa Jurusan Teknik Informatika Fakultas Sains Dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

2. Data sampel maupun data uji menggunakan cara import yang berisi beberapa field di antaranya yaitu nim, asal daerah, asal sekolah, jalur masuk, IPK, IP semester 1-5, pesantren dan keterangan kelulusan.
3. Kriteria kelulusan tepat waktu yaitu lama studi 4 tahun dihitung dari tahun lulus dikurangi tahun masuk mahasiswa. Bila lebih dari 4 tahun maka termasuk pada klasifikasi kasus tidak lulus tepat waktu.

#### **1.4 Tujuan Penelitian**

Tujuan penelitian ini yaitu membuat aplikasi prediksi kelulusan mahasiswa jurusan Teknik Informatika Fakultas Sains dan Teknologi di Universitas Islam Negeri Maulana Malik Ibrahim Malang dengan menggunakan metode *Decision Tree* dengan menggunakan algoritma C4.5.

#### **1.5 Manfaat Penelitian**

Adapun beberapa manfaat yang didapat antara lain :

1. Dapat memberikan informasi prediksi kelulusan mahasiswa yang lulus tepat waktu maupun yang tidak tepat waktu.
2. Dapat membantu mengevaluasi kelulusan mahasiswa pada sistem kerja ataupun sistem akademik yang berjalan di jurusan.

#### **1.6 Metode Penelitian**

##### **a. Study Literatur**

Materi studi literatur :

1. Salah satu teknik data mining yaitu klasifikasi dan metode *Decision Tree*.
2. Metode *Decision Tree* dan teknik pemrograman menggunakan PHP.
3. Algoritma C4.5 dan parameter yang digunakan.
4. Penelitian-penelitian yang berkaitan dengan penelitian penulis.
5. Kajian Islam tentang prediksi kelulusan mahasiswa

**b. Pengumpulan Data**

Data yang diperoleh berasal dari Jurusan Teknik Informatika UIN Maulana Malik Ibrahim Malang sebagai sumber utama tentang desain database dan data mahasiswa yang akan diteliti.

**c. Analisis dan Desain**

Pada proses ini menjelaskan bagaimana aplikasi prediksi kelulusan mahasiswa akan dibangun. Pada tahap ini penulis melakukan perancangan desain *interface*, desain alur sistem, *flowchart* aplikasi dan DFD aplikasi.

**d. Pembuatan Aplikasi**

Hasil dari perancangan dan desain sistem diimplementasikan ke dalam sebuah *source code* dengan bahasa pemrograman PHP. Penulisan *source code* PHP dengan menggunakan editor Macromedia Dreamweaver 8, sedangkan untuk penyimpanan data menggunakan database MySQL.

**e. Pengujian Aplikasi**

Pengujian aplikasi dilakukan untuk memastikan bahwasanya implementasi yang dilakukan sudah sesuai dengan perancangan dan desain sistem, serta dapat memastikan bahwa tidak ada kekeliruan dalam penerapan algoritma ke dalam aplikasi.

## f. Penyusunan Laporan

Setelah semua proses dilakukan, hal terakhir yang dilakukan penulis adalah menyusun laporan dari kegiatan penelitian. Penyusunan laporan bertujuan untuk mendokumentasikan semua hasil penelitian yang dilakukan penulis.

### 1.7 Sistematika Penulisan

Pembuatan skripsi ini dilakukan dengan pembagian bab sebagai berikut:

#### BAB I : PENDAHULUAN

Pada bab ini membahas mengenai latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, metode penelitian, dan sistematika penulisan.

#### BAB II : TINJAUAN PUSTAKA

Bab ini menjelaskan tentang *Decision Tree* dan algoritma C4.5. Adapun literatur yang digunakan meliputi buku referensi dan dokumentasi *internet*.

#### BAB III : ANALISIS DAN PERANCANGAN

Pada bab ini dijelaskan tinjauan organisasi, tahap perencanaan, desain dan perancangan sistem informasi tentang prediksi kelulusan mahasiswa Jurusan Teknik Informatika Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang menggunakan metode *Decision Tree* dengan penerapan algoritma C4.5.

**BAB IV : HASIL DAN PEMBAHASAN**

Bab ini menjelaskan tahapan implementasi dan uji coba dari perancangan sistem serta analisis hasil.

**BAB V : KESIMPULAN DAN SARAN**

Pada bab ini berisi kesimpulan dari pembahasan dan saran yang bermanfaat untuk pengembangan skripsi ini.



## BAB II

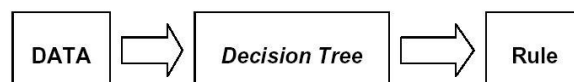
### TINJAUAN PUSTAKA

#### 2.1 Decision Tree

Pohon keputusan merupakan salah satu metode klasifikasi yang kuat dan terkenal. Metode pohon keputusan mengubah fakta yang besar menjadi pohon keputusan yang merepresentasikan aturan, aturan tersebut dapat dengan mudah untuk diinterpretasi oleh manusia. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah variabel input dengan sebuah variabel target (Berry & Linoff, 2004).

Model pohon keputusan terdiri dari sekumpulan aturan untuk membagi sejumlah populasi yang heterogen menjadi lebih kecil (homogen) dengan memperhatikan variabel tujuannya. Variabel tujuan biasanya dikelompokkan dengan pasti dan model pohon keputusan lebih mengarah pada perhitungan probabilitas dari tiap-tiap record terhadap kategori tersebut atau untuk mengklasifikasi record dengan mengelompokkannya dalam satu kelas. Sebuah pohon keputusan dapat dibangun dengan menerapkan salah satu algoritma pohon keputusan untuk memodelkan himpunan data yang belum terklasifikasi kelasnya (Kusrini, 2009).

Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan.



Gambar 2.1 Konsep *Decision Tree*

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan record. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan *Tree*. Misalkan untuk menentukan main tenis, kriteria yang diperhatikan adalah cuaca, angin dan temperatur. Salah satu atribut merupakan atribut yang menyatakan data solusi per-item data yang disebut dengan target atribut. Atribut memiliki nilai-nilai yang dinamakan dengan instance.

Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule. Langkah pertama yang dilakukan dalam pembuatan pohon keputusan yaitu menghitung nilai *Entropy* total dari jumlah data yang dijadikan sampel, selanjutnya yaitu mengelompokkan variabel dan menghitung nilai *Gain* pada tiap atribut. Setelah dihitung menggunakan rumus algoritma C4.5 maka atribut yang mempunyai nilai *Gain* tertinggi akan menjadi akar dan atribut lainnya menjadi cabang, kemudian dari cabang akan dihitung kembali atribut apa lagi yang mempunyai nilai *Gain* tertinggi. Tahapan perhitungan akan berulang secara continue sampai beberapa tingkatan hingga mencapai nilai akhir yaitu keputusan “Yes” dan “No” dalam kasus ini yaitu “Lulus tepat waktu” dan “Tidak lulus tepat waktu”.

Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk membreak down proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Pohon Keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Pohon keputusan memadukan antara eksplorasi data dan pemodelan, sehingga sangat bagus

sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain.

*Decision Tree* juga disebut sebagai diagram alir yang berbentuk seperti struktur pohon yang mana setiap internal node menyatakan pengujian terhadap suatu atribut, setiap cabang menyatakan output dari pengujian tersebut dan node daun (leaf node) menyatakan distribusi kelas. Node yang paling atas disebut sebagai node akar (root node). *Decision Tree* digunakan untuk mengklasifikasikan suatu sampel data yang belum diketahui kelasnya ke dalam kelas-kelas yang sudah ada. Jalur pengujian data adalah pertama semua data harus melalui root node dan terakhir adalah melalui leaf node yang akan menyimpulkan prediksi kelas bagi data tersebut. Atribut data harus berupa data kategorik, bila kontinu maka atribut harus didiskretisasi terlebih dahulu.

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

- a. Memiliki kecepatan yang relatif lebih cepat
- b. Dapat diubah menjadi rule klasifikasi dengan mudah dan sederhana
- c. Dapat menggunakan query SQL untuk mengakses database
- d. Dapat dibandingkan tingkat akurasi dengan metode lainnya

## 2.2 Algoritma C4.5

Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C.45. Algoritma C.45 merupakan pengembangan dari algoritma ID3 (Larose, 2005). Secara umum algoritma C4.5 digunakan untuk membangun pohon keputusan adalah sebagai berikut (Kusrini, 2009):

- a. Pilih atribut sebagai akar.
- b. Buat cabang untuk tiap-tiap nilai.
- c. Bagi kasus dalam cabang.
- d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *Gain* digunakan rumus seperti tertera dalam persamaan berikut :

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

Keterangan :

- S : himpunan kasus  
 A : atribut  
 n : jumlah partisi atribut A  
 |S<sub>i</sub>| : jumlah kasus pada partisi ke-i  
 |S| : jumlah kasus dalam S

Setelah mendapatkan nilai *Gain*, ada satu hal lagi yang perlu dilakukan perhitungan yaitu mencari nilai *Entropy*. *Entropy* digunakan untuk menentukan seberapa informatif sebuah input atribut untuk menghasilkan output atribut.

Rumus dasar dari *Entropy* tersebut adalah sebagai berikut :

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

Keterangan :

- S : himpunan kasus  
 n : jumlah partisi S

$p_i$  : proporsi dari  $S_i$  terhadap  $S$

Untuk memudahkan penjelasan mengenai algoritma C4.5 berikut ini disertakan contoh kasus yang dituangkan dalam Tabel 2.1.

Tabel 2.1 Contoh Data Sampel

NO	CUACA	SUHU	KELEMBABAN	ANGIN	MAIN
1	Cerah	Panas	Tinggi	Kecil	Tidak
2	Cerah	Panas	Tinggi	Besar	Tidak
3	Berawan	Panas	Tinggi	Kecil	Ya
4	Hujan	Sejuk	Tinggi	Kecil	Ya
5	Hujan	Dingin	Normal	Kecil	Ya
6	Hujan	Dingin	Normal	Besar	Ya
7	Berawan	Dingin	Normal	Besar	Ya
8	Cerah	Sejuk	Tinggi	Kecil	Tidak
9	Cerah	Dingin	Normal	Kecil	Ya
10	Hujan	Sejuk	Normal	Kecil	Ya
11	Cerah	Sejuk	Normal	Besar	Ya
12	Berawan	Sejuk	Tinggi	Besar	Ya
13	Berawan	Panas	Normal	Kecil	Ya
14	Hujan	Sejuk	Tinggi	Besar	Tidak

Dalam kasus yang tertera pada Tabel 2.1, akan dibuat pohon keputusan untuk menentukan main tenis atau tidak dengan melihat keadaan cuaca, temperatur, kelembaban dan keadaan angin. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar
2. Buat cabang untuk masing-masing nilai

3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Hasil perhitungan nilai *Entropy* digunakan untuk menghitung nilai *Gain*. Untuk menghitung *Gain* digunakan rumus seperti tertera diatas.

Berikut ini adalah penjelasan lebih rinci mengenai masing-masing langkah dalam pembentukan pohon keputusan dengan menggunakan algoritma C4.5 untuk menyelesaikan permasalahan pada Tabel 2.1.

Menghitung jumlah kasus, jumlah kasus untuk keputusan Ya, jumlah kasus untuk keputusan Tidak, dan *Entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut Cuaca, Suhu, Kelembaban, dan Aangin. Setelah itu lakukan penghitungan *Gain* untuk masing-masing atribut. Hasil perhitungan ditunjukkan oleh Tabel 2.2.

Tabel 2.2 Perhitungan Node 1

Node		Jml Kasus (S)	Tidak (S <sub>1</sub> )	Ya (S <sub>2</sub> )	<i>Entropy</i>	<i>Gain</i>
1	TOTAL	14	4	10	0.863120569	
	CUACA					0.258521037
	BERAWAN	4	0	4	0	
	HUJAN	5	1	4	0.721928095	
	CERAH	5	3	2	0.970950594	

	SUHU						0.183850925
		DINGIN	4	0	4	0	
		PANAS	4	2	2	1	
		SEJUK	6	2	4	0.918295834	
	KELEMBABAN						0.370506501
		TINGGI	7	4	3	0.985228136	
		NORMAL	7	0	7	0	
	ANGIN						0.005977711
		KECIL	8	2	6	0.811278124	
		BESAR	6	4	2	0.918295834	

Baris Total kolom *Entropy* pada Tabel 3.2 dihitung dengan rumus sebagai berikut:

$$Entropy(Total) = \left( -\frac{4}{14} * \log_2 \left( \frac{4}{14} \right) \right) + \left( -\frac{10}{14} * \log_2 \left( \frac{10}{14} \right) \right)$$

$$Entropy(Total) = 0.863120569$$

Sementara itu nilai *Gain* pada baris Cuaca dihitung dengan menggunakan rumus *Gain*, sebagai berikut :

$$Gain(Total, Cuaca) = Entropy(Total) - \sum_{i=1}^n \frac{Cuaca}{|Total|} * Entropy(Cuaca)$$

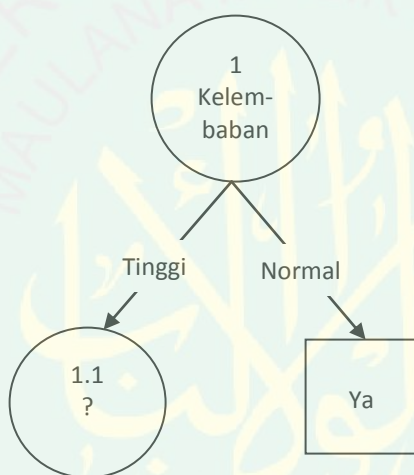
$$Gain(Total, Cuaca) = 0.863120569 - \left( \left( \frac{4}{14} * 0 \right) + \left( \frac{5}{14} * 0.723 \right) + \left( \frac{5}{14} * 0.97 \right) \right)$$

$$Gain(Total, Cuaca) = 0.23$$

Dari hasil pada Tabel 2.2 dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah Kelembaban yaitu sebesar 0.37. Dengan demikian Kelembaban dapat menjadi node akar. Ada 2 nilai atribut dari Kelembaban yaitu Tinggi dan Normal. Dari kedua nilai atribut

tersebut, nilai atribut Normal sudah mengklasifikasikan kasus menjadi 1 yaitu keputusannya Ya, sehingga tidak perlu dilakukan perhitungan lebih lanjut. Atribut normal yang sudah terklasifikasi kasusnya dengan keputusan Ya menjadikan 1 aturan atau rule yang terbentuk, tetapi untuk nilai atribut Tinggi masih perlu dilakukan perhitungan lagi karena belum terklasifikasikan kasusnya.

Langkah pertama dalam membangun pohon keputusan yaitu memilih atribut sebagai akar. Dari hasil perhitungan diatas dapat digambarkan pohon keputusan sementara tampak seperti Gambar 2.3



Gambar 2.2 Pohon Keputusan Hasil Perhitungan Node 1

Menghitung jumlah kasus, jumlah kasus untuk keputusan Ya, jumlah kasus untuk keputusan Tidak, dan *Entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut Cuaca, Suhu dan Angin yang dapat menjadi node akar dari nilai atribut Kelembaban yaitu Tinggi. Setelah itu lakukan penghitungan *Gain* untuk masing-masing atribut. Hasil perhitungan ditunjukkan oleh Tabel 2.3.

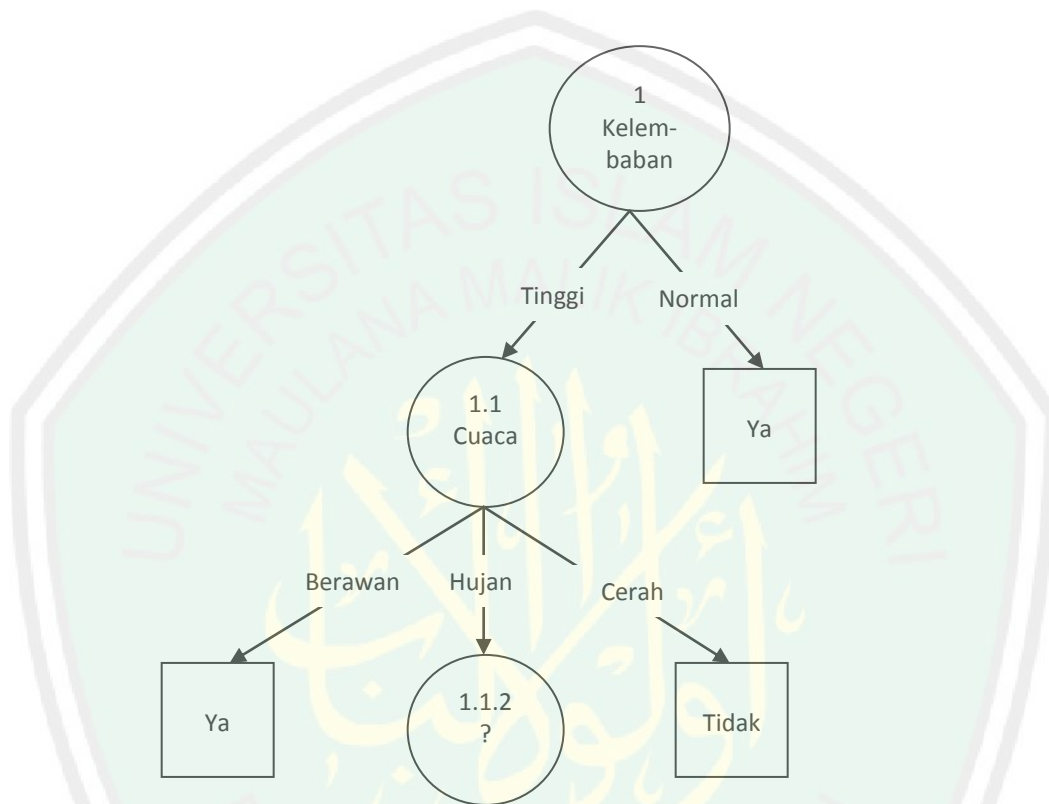
Tabel 2.3 Perhitungan Node 1.1

Node		Jml Kasus (S)	Tidak (S <sub>1</sub> )	Ya (S <sub>2</sub> )	Entropy	Gain
1.1	KELEMBABAN—TINGGI	7	4	3	0.985228136	
	CUACA					0.69951385
	BERAWAN	2	0	2	0	
	HUJAN	2	1	1	1	
	CERAH	3	3	0	0	
	SUHU					0.020244207
	DINGIN	0	0	0	0	
	PANAS	3	2	1	0.918295834	
	SEJUK	4	2	2	1	
	ANGIN					0.020244207
	KECIL	4	2	2	1	
	BESAR	3	2	1	0.918295834	

Dari hasil pada Tabel 2.3 dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah Cuaca yaitu sebesar 0.67. Dengan demikian Cuaca dapat menjadi node cabang dari nilai atribut Tinggi. Ada 3 nilai atribut dari Cuaca yaitu Berawan, Hujan dan Cerah. Dari ketiga nilai atribut tersebut, nilai atribut Berawan sudah mengklasifikasikan kasus menjadi 1 yaitu keputusannya Ya dan nilai atribut Cerah sudah mengklasifikasikan kasus menjadi satu

dengan keputusan Tidak, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut Hujan masih perlu dilakukan perhitungan lagi.

Pohon keputusan yang terbentuk sampai tahap ini ditunjukkan pada gambar 2.3 berikut:



Gambar 2.3 Pohon Keputusan Hasil Perhitungan Node 1.1

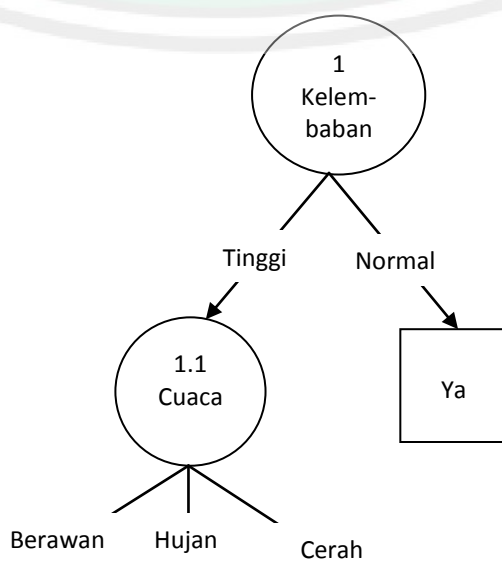
Menghitung jumlah kasus, jumlah kasus untuk keputusan Ya, jumlah kasus untuk keputusan Tidak, dan *Entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut Suhu dan Angin yang dapat menjadi node cabang dari nilai atribut Hujan. Setelah itu lakukan penghitungan *Gain* untuk masing-masing atribut. Hasil perhitungan ditunjukkan oleh Tabel 2.4.

Tabel 2.4 Perhitungan Node 1.1.2

Node	Jml Kasus (S)	Tidak (S <sub>1</sub> )	Ya (S <sub>2</sub> )	<i>Entropy</i>	<i>Gain</i>

1.1.2	KELEMBABAN—TINGGI dan CUACA—HUJAN		2	1	1	1	
	SUHU						0
		DINGIN	0	0	0	0	
		PANAS	0	0	0	0	
		SEJUK	2	1	1	1	
	ANGIN						1
		KECIL	1	0	1	0	
		BESAR	1	1	0	0	

Dari hasil pada tabel 2.4 dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah Angin yaitu sebesar 1. Dengan demikian Angin dapat menjadi node cabang dari nilai atribut Hujan. Ada 2 nilai atribut dari Angin yaitu Kecil dan Besar. Dari kedua nilai atribut tersebut, nilai atribut Kecil sudah mengklasifikasikan kasus menjadi 1 yaitu keputusannya Ya dan nilai atribut Besar sudah mengklasifikasikan kasus menjadi satu dengan keputusan Tidak, sehingga tidak perlu dilakukan perhitungan lebih lanjut untuk nilai atribut ini. Pohon keputusan yang terbentuk sampai tahap ini ditunjukkan pada Gambar 2.5.



#### Gambar 2.4 Pohon Keputusan Hasil Perhitungan Node 1.1.2

Dengan memperhatikan pohon keputusan pada Gambar 2.5, diketahui bahwa semua kasus sudah masuk dalam kelas. Dengan demikian, pohon keputusan pada Gambar 2.5 merupakan pohon keputusan terakhir yang terbentuk.

Dari hasil pohon keputusan yang terbentuk hingga terakhir menghasilkan 5 aturan atau rule keputusan dari target yang ingin dicapai yaitu bermain tenis atau tidak. Rule tersebut diantaranya yaitu :

1. Jika kelembaban normal maka bermain.
2. Jika kelembaban tinggi dan cuaca cerah maka tidak bermain.
3. Jika kelembaban tinggi dan cuaca berawan maka bermain.
4. Jika kelembaban tinggi dan cuaca hujan dan angin kecil maka bermain.
5. Jika kelembaban tinggi dan cuaca hujan dan angin besar maka tidak bermain.

Dari aturan-aturan yang telah dihasilkan nantinya akan digunakan untuk mencocokkan kasus pada variabel target yang diinputkan. Tiap record akan dikoreksi statusnya dengan aturan yang ada sehingga dapat diklasifikasikan dalam satu kelas berdasarkan target yang ingin diketahui.

## 2.3 Penelitian Terkait

Penelitian yang terkait dalam pembuatan aplikasi ini antara lain :

### 1. Graduation Prediction Of Gunadarma University Students Using Naïve Bayes Algorithm And C4.5 Algorithm

Penelitian ini dilaksanakan oleh Marselina Silvia Suhartinah dan Ernastuti pada tahun 2010.

Penelitian dilakukan untuk mencari dan menemukan pola yang terdapat pada data mahasiswa berdasarkan data NEM, IP DNS semester 1, IP DNS semester 2, IPK DNU semester 1-2, gaji orang tua dan pekerjaan orang tua, untuk memprediksi mahasiswa yang lulus atau tidak lulus sesuai dengan waktu studi dengan menggunakan algoritma naive bayes dan C4.5, kemudian membandingkan hasil dan akurasi kedua algoritma tersebut.

Setelah dilakukan uji coba dapat disimpulkan beberapa hal sebagai berikut :

1. Akurasi kebenaran dari 22 set data mahasiswa pada proses uji coba yaitu 85,7% untuk algoritma C4.5 dan 80,85% untuk algoritma Naïve Bayes.
2. Dengan menggunakan algoritma C4.5 kesalahan yang dihasilkan dalam proses prediksi lebih sedikit karena C4.5 melakukan klasifikasi record-record ke dalam kelas tujuan yang ada.
3. Algoritma *Decision Tree* memiliki kompleksitas yang lebih besar. Karena pada algoritma C4.5 setiap nilai dalam suatu atribut ditelusuri dan diproses untuk mendapatkan entropi masing-masing nilai yang akan digunakan untuk mencari ukuran purity masing-masing atribut yang dinyatakan dengan *information Gain*. Proses penelusuran ini akan membentuk sebuah pola berupa pohon keputusan.

4. Algoritma Naïve Bayes bila diimplementasikan menggunakan data yang digunakan dalam proses training akan menghasilkan nilai kesalahan yang lebih besar karena pada naive bayes nilai suatu atribut adalah independent terhadap nilai lainnya dalam satu atribut yang sama. Namun memiliki akurasi akurasi yang lebih tinggi bila dimplementasikan ke data yang berbeda dari data training dan kedalam data yang jumlahnya lebih besar.

Perbedaan dengan penelitian ini yaitu penulis menerapkan algoritma naïve bayes dan C4.5 dalam prediksi kelulusan mahasiswa yang dapat lulus sesuai dengan waktu studi menggunakan Java Netbeans sedangkan pada penelitian ini menggunakan PHP.

## **2. Implementation Of C4.5 Algorithm To Evaluate The Cancellation Possibility Of New Student Applicants At STMIK Amikom Yogyakarta**

Penelitian ini dilakukan oleh Kusriani dan Sri Hartati pada tahun 2007.

Pada penelitian ini penulis membahas tentang kasus pembatalan calon mahasiswa baru. Calon mahasiswa yang telah berhasil lulus tes masuk membatalkan keputusan untuk masuk universitas dengan mengabaikan tahap berikutnya dari proses penerimaan mahasiswa baru yaitu pendaftaran ulang. Jika pembatalan pendaftaran dapat dideteksi dini, maka manajer eksekutif dapat membuat setiap upaya untuk menjaga calon mahasiswa baru melalui proses penerimaan dan selanjutnya mengurangi tingkat pembatalan masuk.

Tahapan yang dilalui penulis pada paper ini yaitu :

1. Menentukan atribut yang dipilih
2. Menentukan *range* value dari atribut
3. Tranfer dari data kandidat ke *case* data

#### 4. Membentuk *Decision Tree*

Kesimpulan yang didapat dari paper ini yaitu, aplikasi yang dibangun dapat menghasilkan pohon keputusan yang sesuai dengan variabel dan data kasus yang diberikan oleh user. Tingkat akurasi dari data prediksi dari aplikasi ini sangat tergantung kepada variabel terpilih yang akan menjadi dasar untuk membuat pohon keputusan.

Perbedaan dengan penelitian ini yaitu penulis membuat aplikasi data mining berbasis desktop dengan bahasa pemrograman Visual BASIC, sedangkan pada penelitian ini berbasis web dengan menggunakan bahasa pemrograman PHP. Dalam paper ini obyek yang diteliti berbeda sehingga parameter yang digunakan sebagai atribut dalam perhitungan juga berbeda, namun dalam perhitungannya memiliki konsep yang sama yaitu mengkonversi rumus C4.5 dengan semua atribut yang digunakan.

## BAB III

### ANALISIS DAN PERANCANGAN

#### 3.1 Gambaran Umum Sistem

Input dari sistem yang akan dibangun yaitu berupa data mahasiswa dalam bentuk tabel yang berisi beberapa atribut meliputi jalur masuk mahasiswa, asal sekolah, asal daerah, indeks prestasi kumulatif, indeks prestasi semester 1 sampai semester 5 dan keterangan pesantren. Data mahasiswa tersebut akan diklasifikasikan berdasarkan target yang ditentukan dan dihitung menggunakan metode *Decision Tree* yaitu algoritma C4.5 untuk mencari nilai *Entropy* dan informasi *Gain*. Setelah proses perhitungan selesai maka akan menghasilkan rule atau kondisi yang digunakan dalam penentuan keputusan pada proses prediksi. Output dari sistem ini yaitu keterangan lulus tepat waktu atau tidaknya tiap mahasiswa yang diprediksi.

Sistem yang dibangun memiliki kemampuan sebagai berikut:

1. Melakukan perhitungan terhadap data yang diperoleh dengan mengubah formatnya menjadi bentuk tabel data sampel training
2. Melakukan perhitungan data untuk menentukan nilai *Gain* dan *Entropy*
3. Melakukan update dengan mentraining kembali sampel data yang baru
4. Menyarankan atau memberikan pengetahuan baru dalam penentuan target yang akan diprediksi. Dalam kasus ini yaitu tepat atau tidaknya waktu kelulusan mahasiswa dengan masa studi yang ditempuh.

### 3.2 Analisis Data

Pada langkah awal dalam analisis data ini akan ditentukan beberapa atribut yang digunakan sebagai parameter dalam pengklasifikasian data sampel. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon. Pada ilmu statistik khususnya pada bab sampling terdapat penjelasan mengenai beberapa metode pengambilan sampel. Sampel yang baik yaitu yang dapat menggambarkan (mewakili) populasinya. Untuk memperoleh sampel yang baik diperlukan metode yang baik dalam pemilihan anggota sampel. Sampel nonrandom merupakan salah satu metode pengambilan sampel dimana pemilihan sampel dengan cara ini menggunakan pengetahuan dan opini dari peneliti terhadap obyek yang akan diteliti. Sedangkan dalam sebuah model pohon keputusan berisi aturan yang membagi sejumlah populasi yang homogen menjadi lebih kecil (heterogen).

Dalam pengklasifikasian data sampel, parameter diambil dari data induk mahasiswa dan data kelulusan mahasiswa. Dari hubungan tingkat kelulusan dengan data induk mahasiswa dan data akademik mahasiswa, tidak semuanya dijadikan parameter, namun hanya diambil beberapa atribut saja yang dirasa nilainya bisa mempengaruhi terhadap kelulusan seorang mahasiswa. Adapun beberapa hal yang akan diproses yaitu :

1. Hubungan tingkat kelulusan dengan jalur masuk mahasiswa
2. Hubungan tingkat kelulusan dengan jenis asal sekolah
3. Hubungan tingkat kelulusan dengan asal daerah

4. Hubungan tingkat kelulusan dengan pengalaman belajar di pesantren
5. Hubungan tingkat kelulusan dengan IPK
6. Hubungan tingkat kelulusan dengan Indeks Prestasi Semester 1-5

Dalam kasus ini pemilihan atribut diatas berdasarkan pertimbangan bahwa jumlah nilai variabelnya tidak banyak sehingga diharapkan mahasiswa yang masuk dalam satu klasifikasi nilai variabel tersebut cukup banyak. Misalnya atribut pernah belajar di pesantren atau tidak. Jumlah nilai variabelnya hanya dua yaitu pernah dan tidak pernah, sehingga mahasiswa yang pernah belajar dipesantren dan yang tidak pernah dapat diketahui cukup banyak. Sedangkan seperti nama dan alamat tidak dipilih karena jumlah nilai variabelnya cukup besar.

### 3.3 Alat dan Bahan yang diperlukan

Sistem informasi yang dibangun ini membutuhkan perangkat lunak sebagai penunjang agar dapat berjalan dengan sesuai dengan fungsinya. Kebutuhan tersebut diantaranya:

#### a. Kebutuhan Hardware

Sebuah computer PC/laptop untuk melakukan perancangan dan pembangunan sistem dengan spesifikasi sebagai berikut:

1. Prosesor Core 2 Duo (atau di atasnya)
2. Memory minimal 1 GB
3. Free Hardisk Minimal 20 GB

### **b. Kebutuhan Software**

Selain kebutuhan hardware penulis juga membutuhkan kebutuhan software untuk melakukan perancangan dan pembuatan sistem. Adapun software tersebut adalah:

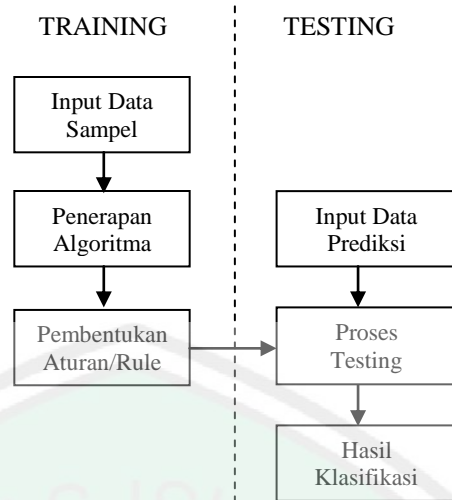
1. Web browser seperti Internet Explorer 6 atau di atasnya
2. Power Designer untuk melakukan perancangan sistem
3. Appserv untuk web server
4. My SQL untuk melakukan penyimpanan data
5. Adobe Photoshop untuk mendesain tampilan aplikasi
6. Microsoft Office untuk membuat dokumentasi dan laporan hasil penelitian

### **c. Kebutuhan Data**

Data yang dibutuhkan yaitu data mahasiswa meliputi beberapa atribut yaitu data jalur masuk mahasiswa, asal sekolah, asal daerah, indeks prestasi kumulatif, indeks prestasi semester 1 sampai 5, keterangan pesantren, tahun masuk dan tahun lulus mahasiswa.

## **3.4 Deskripsi Sistem**

Secara garis besar alur sistem yang akan dibangun akan dijelaskan dengan blok diagram sebagai berikut :



Gambar 3.1 Diagram Blok Sistem

Pada gambar diagram blok 3.1 dapat dijelaskan bahwasanya sistem dimulai dengan dua langkah :

1. Proses Training.

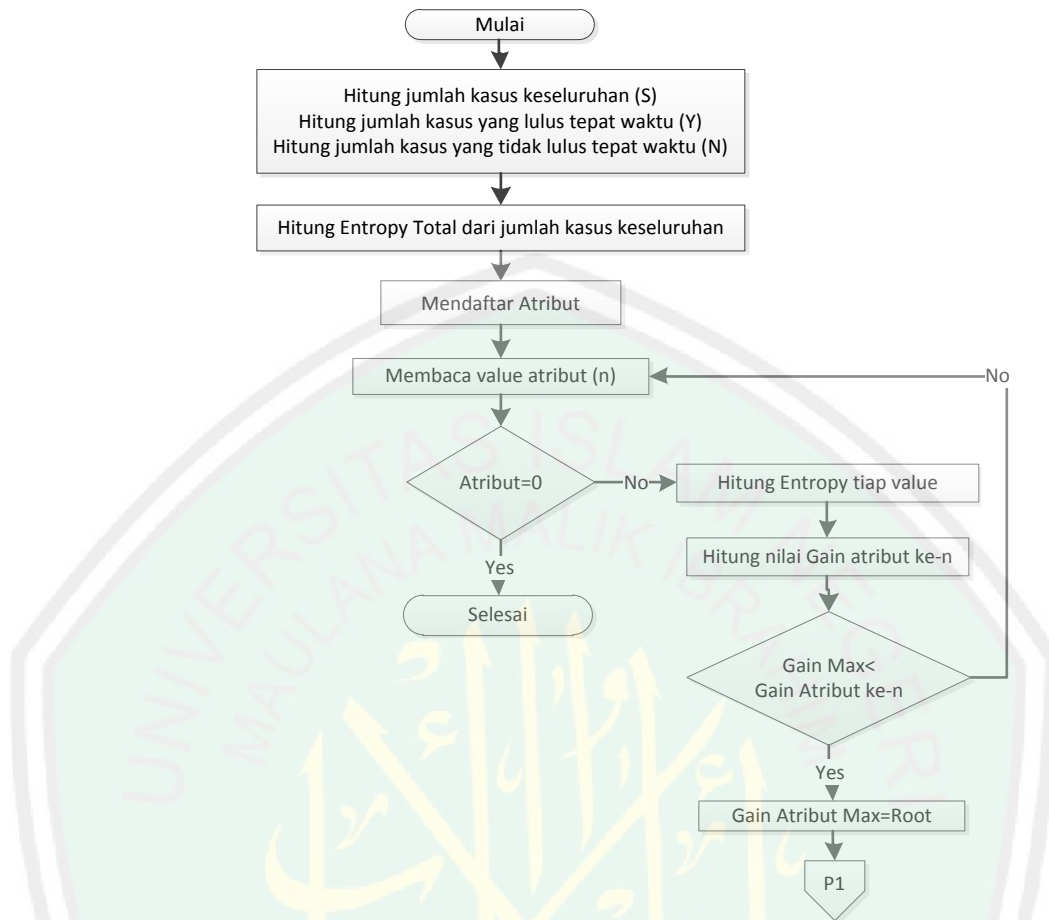
Pada proses training yaitu memasukkan data sampel kedalam tabel yang disiapkan untuk proses perhitungan. Tabel tersebut meliputi atribut, jumlah data keseluruhan, jumlah data yang sudah terklasifikasi berdasarkan target yang ditentukan, dalam kasus ini yaitu lulus tepat waktu atau tidak, serta kolom nilai *Entropy* dan *Gain*. Tahapan selanjutnya yaitu penerapan algoritma C4.5 yaitu menghitung nilai *Entropy* dan *Gain* pada tiap-tiap atribut untuk dijadikan bentuk *Tree*. *Tree* merupakan bentuk aturan klasifikasi yang akan diterapkan pada proses testing.

## 2. Proses Testing

Pada proses testing ini langkah yang dilakukan yaitu memasukkan data uji atau data prediksi. Atribut yang digunakan pada proses testing ini harus sesuai dengan atribut pada proses training. Setiap data atribut akan dibandingkan dengan aturan yang sudah terbentuk pada perhitungan data training sebelumnya. Selanjutnya data tersebut akan diklasifikasikan berdasarkan target yang ingin diketahui yaitu data mahasiswa dengan kondisi atribut ini dapat lulus tepat waktu atau tidak.

Di bawah ini merupakan flowchart pembentuk pohon keputusan yang berjalan pada sistem :

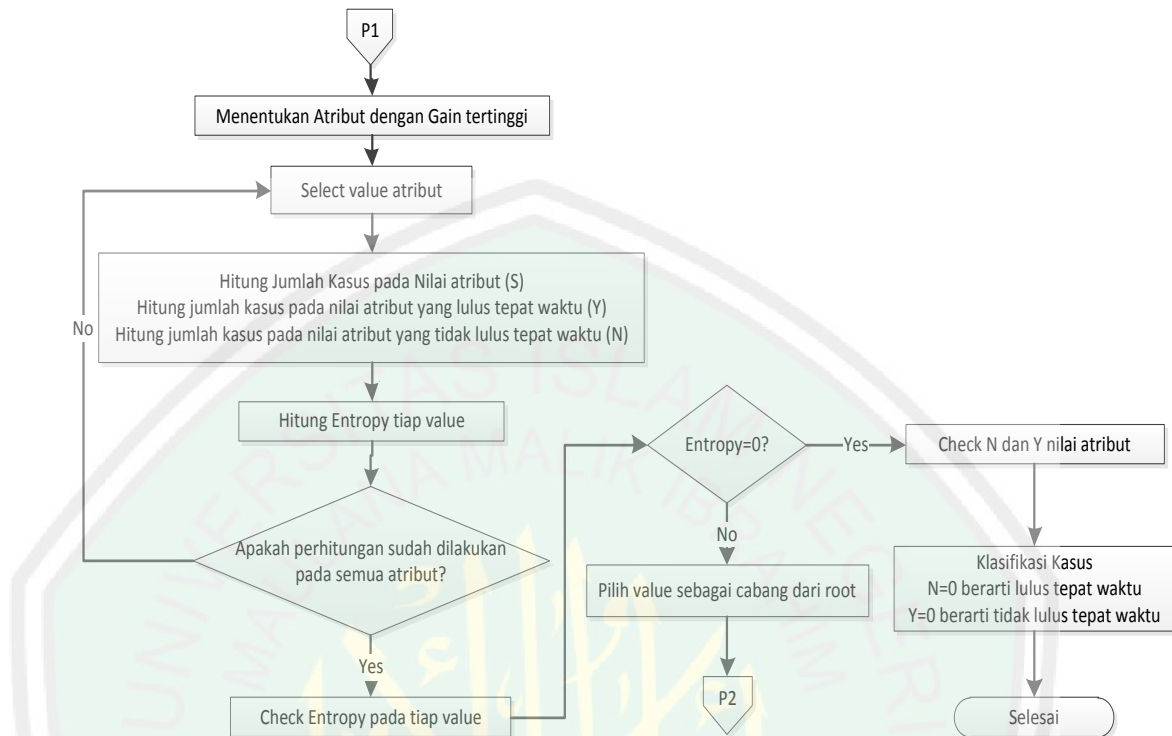




Gambar 3.2 Flowchart Penentuan Root

Berikut sistem yang berjalan pada saat penentuan akar(root):

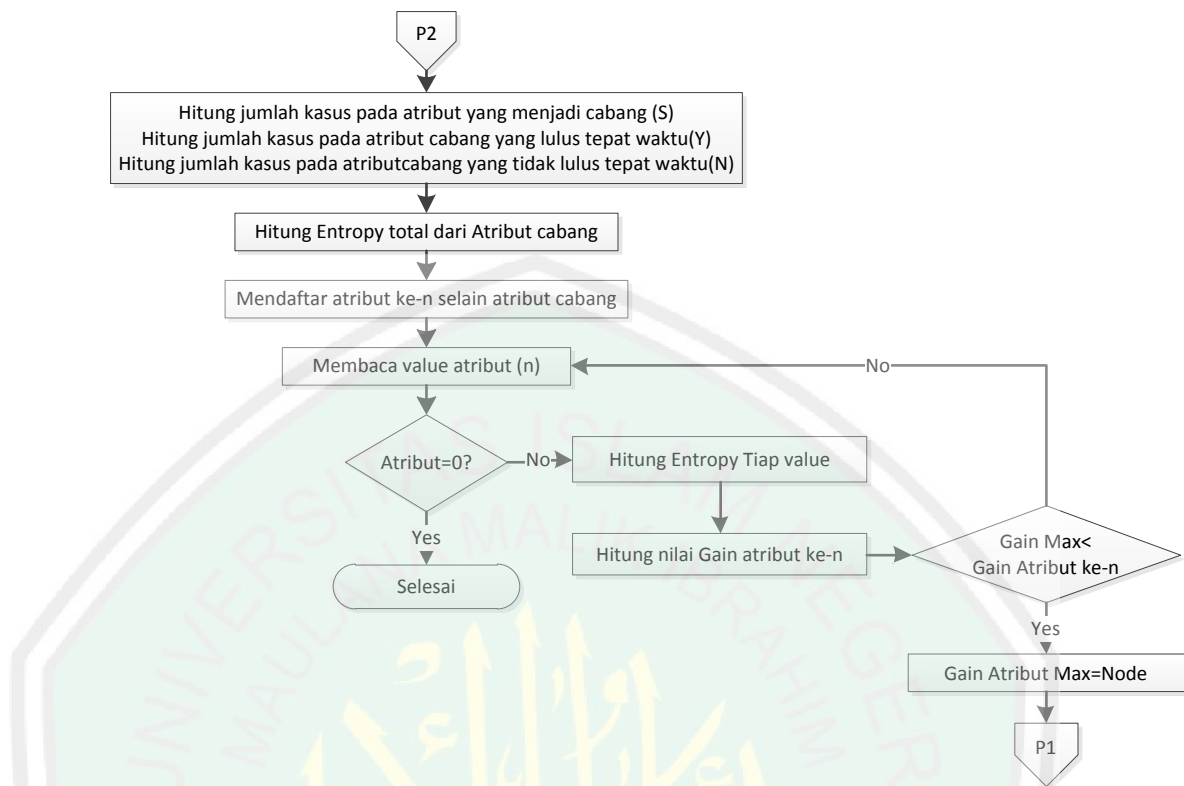
- Sistem akan menghitung total nilai informasi dari data trainingnya
- Sistem akan mendaftarkan atribut A
- Tiap-tiap atribut akan dihitung nilai *Entropy* dan *Gain*nya
- Sistem akan membandingkan nilai *Gain* terbesar dari tiap-tiap atribut
- Setelah nilai *Gain* terbesar didapat, maka sistem akan memilih atribut dengan *Gain* terbesar sebagai atribut terbaik untuk dijadikan akar
- Sistem akan melakukan proses ini sampai semua atribut dalam daftar habis dihitung



Gambar 3.3 Flowchart Penentuan Cabang

Berikut sistem yang berjalan pada saat penentuan cabang pada *Tree* yang terbetuk:

- Sistem akan memilih atribut dengan *Gain* tertinggi
- Nilai yang ada pada atribut tertinggi akan diklasifikasikan berdasar target yang ingin dicapai yaitu lulus tepat waktu(Y) atau tidak lulus tepat waktu(N)
- Tiap nilai atribut akan dihitung *Entropy* masing-masing hingga semua atribut habis.
- Nilai *Entropy* yang nol(0) akan dikoreksi untuk penentuan klasifikasi kasus
- Bila nilai *Entropy* lebih dari nol(0) maka nilai tersebut akan dijadikan cabang pada node selanjutnya.



Gambar 3.4 Flowchart Penentuan Node

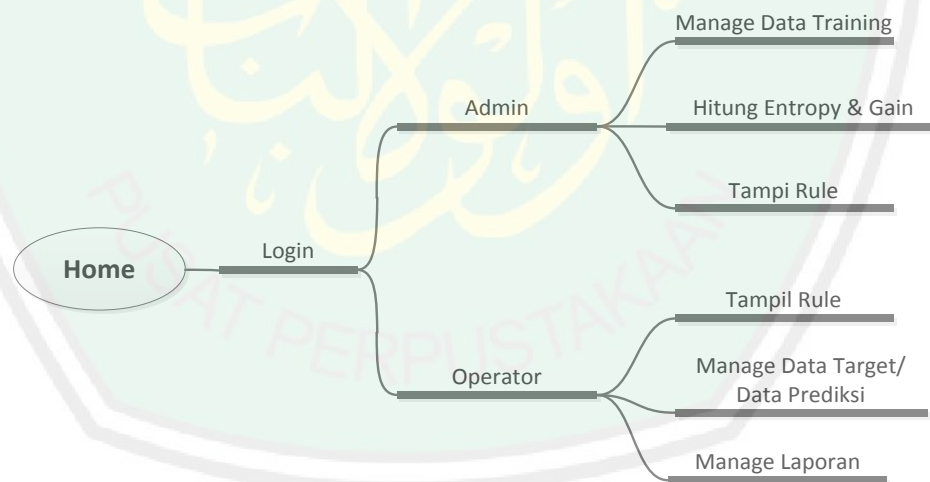
Berikut sistem yang berjaan saat penentuan node pada pembentukan *Tree*:

- a. Sistem akan menghitung jumlah kasus pada atribut dengan *Gain* tertinggi
- b. Menghitung nilai *Entropy* total dari atribut dengan *Gain* tertinggi
- c. Mendaftar atribut A selain atribut dari *Gain* tertinggi
- d. Menghitung *Entropy* dari masing-masing nilai atribut ke-n hingga habis terhitung semua
- e. Menghitung nilai *Gain* dari masing-masing atribut
- f. Menentukan *Gain* maksimal sebagai penentuan node selanjutnya

- g. Setelah node selanjutnya terpilih maka proses perhitungan akan berulang lagi mulai dari penentuan cabang dan penentuan node hingga semua atribut habis dieksekusi dan mencapai end of *Tree*.

## 1. Sitemap Aplikasi

Perancangan aplikasi akan dibuat secara sederhana dan friendly user agar lebih mudah dalam pengoperasiannya. Aplikasi dibuat dengan dua level user yaitu admin dan Operator. Admin memiliki hak akses pada management data training atau data sampel. Sedangkan operator memiliki hak akses management data target atau data prediksi dan akses keterangan kelulusan mahasiswa. Berikut ilustrasi sistem yang berjalan pada aplikasi :



Gambar 3.5 Sitemap Aplikasi

## 3.5 Perancangan Sistem

### 3.5.1 Ekstraksi Dokumen

## 1. Pembersihan data

Dalam hal pengumpulan data, kriteria atribut yang dipilih meliputi hal-hal yang bersifat akademis dan erat hubungannya dengan kelulusan mahasiswa. Data tersebut nantinya akan dicari nilai *Entropy* dan *Gain*nya untuk dapat dijadikan pola dalam pembentukan *Decision Tree* sehingga akan terlihat atribut apa saja yang paling besar pengaruhnya terhadap kelulusan mahasiswa. Data pengujian yang digunakan dalam penelitian ini adalah data mahasiswa jurusan teknik informatika UIN Maulana Malik Ibrahim Malang, data tersebut meliputi:

### a. Data Induk Mahasiswa

Dalam data induk mahasiswa terdapat banyak atribut, namun dalam hal ini dipilih beberapa atribut yang memiliki pengaruh terhadap kelulusan diantaranya yaitu :

1. Proses masuk digunakan untuk mengetahui hubungan antara tingkat kelulusan dengan jalur masuk yang digunakan mahasiswa.
2. Asal sekolah, digunakan untuk mengetahui tingkat kelulusan mahasiswa dengan jenis asal sekolah pada waktu tingkat SLTA.
3. Asal daerah, digunakan untuk mengetahui tingkat kelulusan mahasiswa dengan asal daerah karena lingkungan juga mempengaruhi pembentukan pribadi seseorang.
4. Pernah belajar di pesantren atau tidak, digunakan sebagai perbandingan antara mahasiswa yang pernah belajar di pesantren dan yang tidak terhadap tingkatan kelulusan mahasiswa mengingat mata kuliah di

Universitas Islam Negeri ini sangat bervariasi seperti PKPBA dan PKPBI.

b. Data Kelulusan Mahasiswa

Data kelulusan mahasiswa yang digunakan yaitu :

1. IP semester 1 sampai 5 digunakan untuk mengetahui perkembangan nilai akademis mahasiswa per semester yang mempengaruhi kelulusan mahasiswa.
2. IPK, digunakan untuk mengetahui tingkat prestasi akademis terhadap kelulusan mahasiswa.

2. Integrasi data

Dalam proses integrasi data ini menggabungkan beberapa atribut dari data induk mahasiswa dan data kelulusan mahasiswa menjadi satu tabel.

3. Transformasi data

Transformasi data merupakan proses perubahan atau penggabungan data ke dalam format yang sesuai untuk diproses dalam data mining. Seringkali data yang akan digunakan dalam proses data mining mempunyai format yang belum langsung bisa digunakan, oleh karena itu perlu dirubah formatnya. Dalam hal ini data yang ada diubah menjadi bentuk tabel yang sudah dikelompokkan berdasarkan target yang akan diprediksi yaitu ketepatan waktu kelulusan

seorang mahasiswa. Untuk atribut seperti nilai IPK dan IP per Semester akan dikelompokkan berdasarkan parameter sebagai berikut :

Tabel 3.1 Predikat Kelulusan

No.	Nilai Indeks Prestasi	Nilai Huruf	Predikat
1	3,51 – 4,00	A	Dengan Pujian (Cumlaude)
2	3,00 – 3,50	B	Sangat Memuaskan
3	2,51 – 2,99	C	Memuaskan
4	2,00 – 2,50	D	Cukup

#### 4. Penggunaan Algoritma C4.5

Data mining merupakan metode pencarian informasi (pengetahuan) baru yang terkandung dalam data yang berjumlah besar. *Decision Tree* merupakan salah satu metode klasifikasi dan prediksi yang kuat dan terkenal dalam penerapan data mining. Algoritma C4.5 merupakan algoritma yang digunakan dalam pembuatan pola *Decision Tree* berdasarkan perhitungan dari data yang diperoleh.

#### 5. Evaluasi pola yang ditemukan

Menentukan rule atau pola yang bisa digunakan dalam perhitungan prediksi kelulusan berdasarkan data training yang telah diperoleh hingga mendapatkan hasil akhir yaitu informasi prediksi kelulusan mahasiswa, hasil itu nantinya akan dicocokkan dengan data sesungguhnya sehingga dapat diketahui tingkat kesalahan dari pola yang didapat. Apabila pola yang didapat mempunyai tingkat kesalahan yang rendah maka pola itu dapat berlaku dan digunakan

dalam prediksi data testing untuk mengetahui informasi dari prediksi kelulusan mahasiswa.

#### 6. Presentasi pengetahuan

Memperlihatkan data hasil prediksi berdasarkan target yang ingin dicapai yaitu prediksi tentang informasi lulus atau tidaknya seorang mahasiswa berdasarkan data-data dari atribut yang digunakan.

### 3.5.2 Analisa

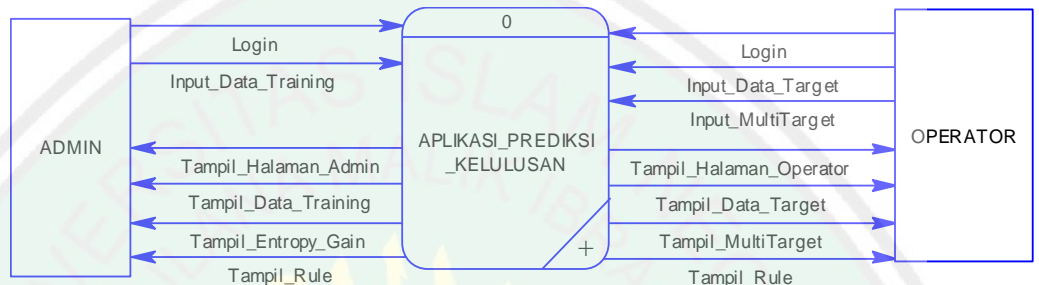
Secara umum, fase algoritma C4.5 dalam membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai root
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Langkah awal yang dilakukan yaitu mengelompokkan data training berdasarkan target yang ingin dicapai yaitu lulus atau tidaknya seorang mahasiswa, lalu langkah selanjutnya yaitu menghitung nilai *Entropy* dan *Gain* berdasarkan pengelompokan data training yang ada. Nilai tertinggi dari beberapa atribut akan dijadikan akar atau root pada pola *Decision Tree* dan sisa atribut yang ada akan dihitung kembali dengan rule yang sama yaitu dihitung dengan mencari nilai *Gain* tertinggi untuk dijadikan cabang dari akar pertama. Rule perhitungan berjalan

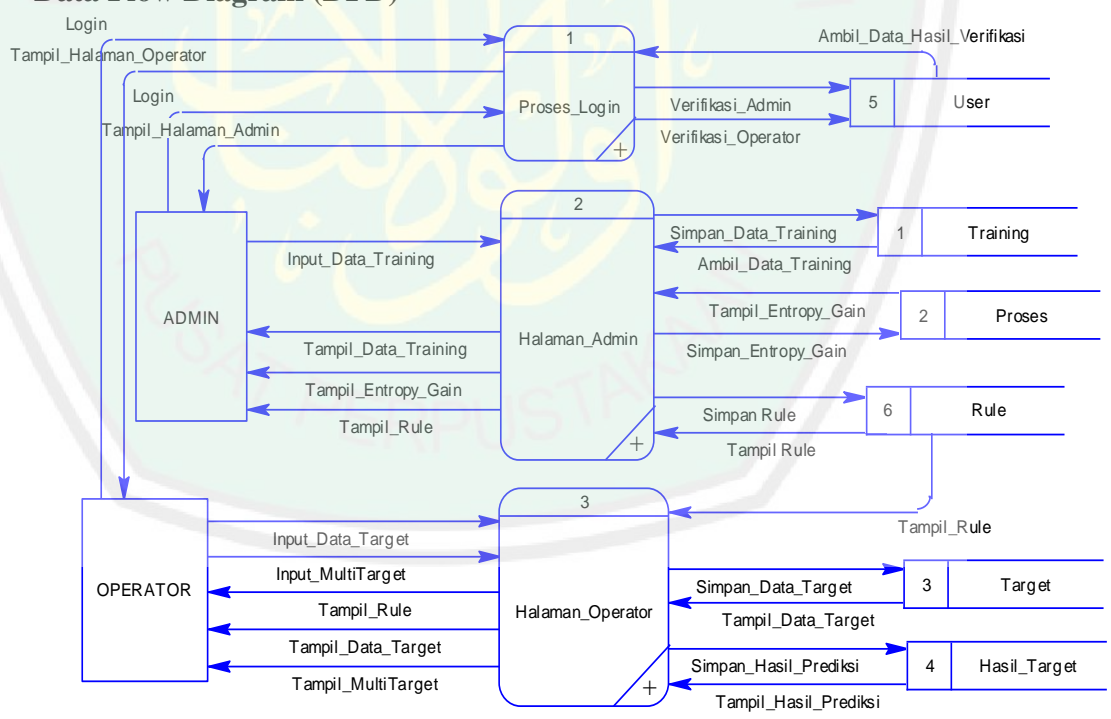
continue hingga semua atribut dapat diketahui hasil akhirnya pada pola *Tree* yang terbentuk yaitu “Y” berarti Lulus Tepat Waktu dan “N” berarti Tidak Lulus Tepat Waktu.

### 3.5.3 Konteks Diagram

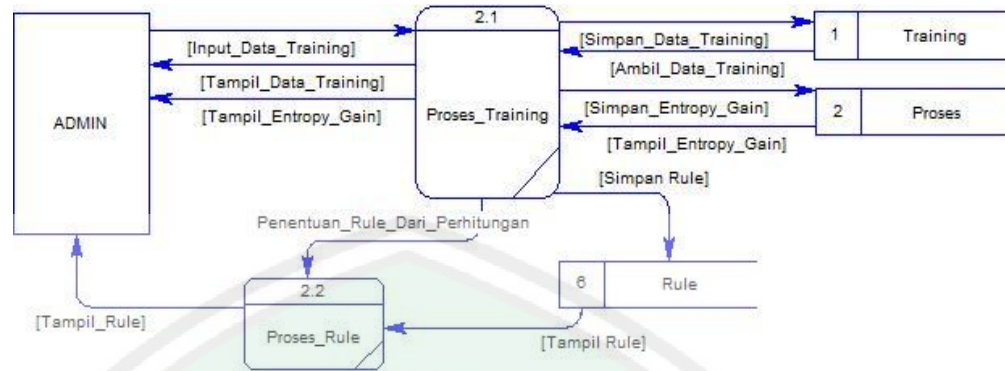


Gambar 3.6 Konteks Diagram

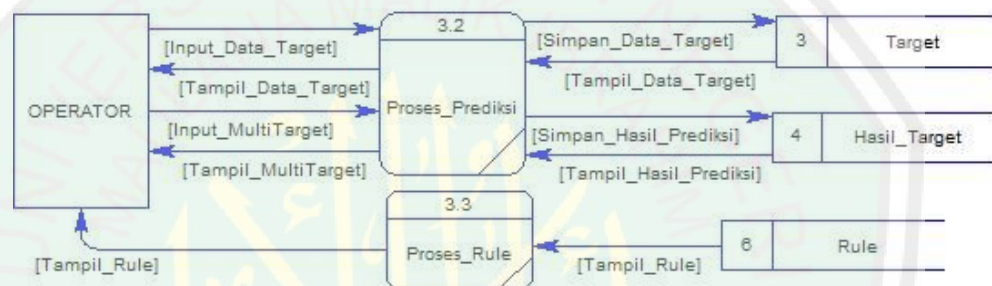
### 3.5.4 Data Flow Diagram (DFD) Level 1



Gambar 3.7 Data Flow Diagram (DFD) Level 1



Gambar 3.8 Data Flow Diagram (DFD) Level 2 Proses Admin



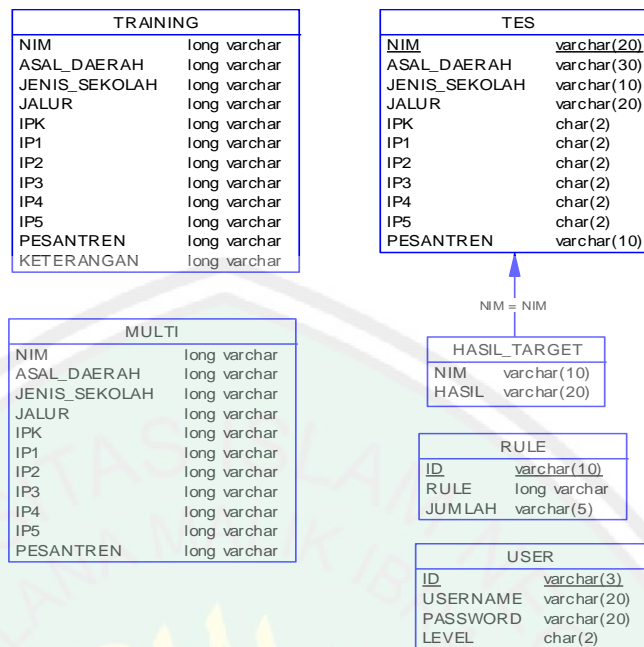
Gambar 3.9 Data Flow Diagram (DFD) Level 2 Proses Operator

## 1. Desain Input

Tabel 3.2 Tabel Desain Input

Nama input	Alat untuk mengentry data	Bentuk input	Yang menyediakan data	Yang mengentry data	Periode input	Deskripsi input	Data/ informasi yang dientry-kan
<b>Input Data Training</b>	Keyboard dan mouse	Data	Jurusan	Admin	Flexible	Berisikan data atribut yang digunakan dalam proses perhitungan	Nim, asal_daerah, asal_sekolah, jalur_masuk, ipk,ip1,ip2,ip3,ip4, ip5,pesanrten, keterangan_lulus
<b>Input Data Target</b>	Keyboard, mouse	Data	Jurusan	Operator	Flexible	Berisikan data atribut yang akan diprediksi kelulusannya	Nim, asal_daerah, asal_sekolah, jalur_masuk, ipk,ip1,ip2,ip3,ip4, ip5, pesanrten





Gambar 3.11 Entity Relational Diagram Physical

### 3.5.6 Design Database

Database Yang Digunakan :

#### 1. Tabel Training

Tabel 3.5 Tabel Training

No	Nama field	Type data
1	nim	Text
2	asal_daerah	Text
3	jenis_sekolah	Text
4	jalur	Text
5	ipk	Text
6	ip1	Text
7	ip2	Text
8	ip3	Text
9	ip4	Text
10	ip5	Text
11	Pesantren	Text
12	Keterangan	Text

#### 2. Tabel Hasil\_Target

Tabel 3.6 Tabel Hasil\_Target

No	Nama field	Type data	Fungsi
1	Nim	Varchar (20)	Foreign key
2	Hasil	Varchar (20)	

## 3. Tabel Multi

Tabel 3.7 Tabel Multi

No	Nama field	Type data
1	nim	Text
2	asal_daerah	Text
3	jenis_sekolah	Text
4	jalur	Text
5	ipk	Text
6	ip1	Text
7	ip2	Text
8	ip3	Text
9	ip4	Text
10	ip5	Text
11	Pesantren	Text

## 4. Tabel Tes

Tabel 3.8 Tabel Tes

No	Nama field	Type data	Fungsi
1	nim	Varchar (20)	Primary key
2	asal_daerah	Varchar (30)	
3	jenis_sekolah	Varchar (10)	
4	jalur	Varchar (20)	
5	ipk	Char (2)	
6	ip1	Char (2)	
7	ip2	Char (2)	
8	ip3	Char (2)	
9	ip4	Char (2)	
10	ip5	Char (2)	
11	Pesantren	Varchar (20)	

## 5. Tabel Rule

Tabel 3.9 Tabel Rule

No	Nama field	Type data	Fungsi
1	Id	Int (11)	Foreign key
2	Rule	Text (100)	
3	Jumlah	Varchar (5)	

## 6. Tabel User

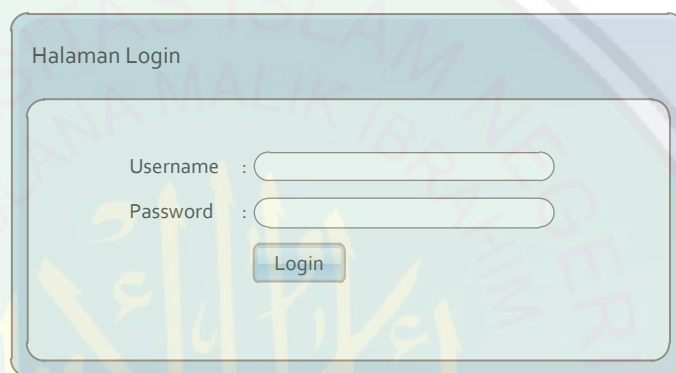
Tabel 3.10 Tabel User

No	Nama field	Type data	Fungsi
1	Id	Varchar (3)	Primary key
2	Username	Varchar (20)	
3	Password	Varchar (20)	
4	Level	Char(2)	

### 3.5.7 Perancangan Antar Muka

#### a. Halaman Login

Pada aplikasi ini mempunyai dua user yaitu admin dan operator. Admin mempunyai hak akses management data training dan perhitungan nilai *Entropy* dan *Gain* untuk mendapatkan rule prediksi. Sedangkan operator mempunyai hak akses management data target, tampil rule dan tampil laporan.



The image shows a login form titled "Halaman Login". It contains two input fields: "Username" and "Password", each followed by a colon and a text input box. Below these fields is a blue "Login" button.

Gambar 3.12 Halaman Login

#### 1. Halaman Utama

Halaman utama terdiri dari 3 bagian yaitu header, menu dan content.

Berikut rancangan interface pada aplikasi :



The image shows a wireframe of the main page layout. At the top, it says "APLIKASI PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA C4.5". Below this, there are three distinct sections: "HEADER", "MENU", and "CONTENT", each represented by a rectangular box.

Gambar 3.13 Halaman Utama

## 2. Halaman Training

Halaman ini digunakan untuk memasukkan data training atau data sampel baru dari data mahasiswa yang ada lalu memproses perhitungan nilai *Entropy* dan *Gain* untuk menghasilkan rule prediksi.

Gambar 3.14 Halaman Training

## 3. Halaman *Entropy* dan *Gain*

Halaman ini berisi nilai *Entropy* dan *Gain* dari hasil perhitungan tiap atribut pada data training menggunakan algoritma C4.5.

Gambar 3.15 Halaman *Entropy* dan *Gain*

#### 4. Halaman Rule

Halaman rule ini berisi aturan prediksi yang telah terklasifikasi menjadi dua yaitu aturan dengan keputusan lulus tepat waktu dan aturan dengan keputusan tidak lulus tepat waktu yang dihasilkan dari perhitungan data training sebelumnya.



Gambar 3.16 Halaman Rule

#### 5. Halaman Target

Halaman target berisi management data target atau data mahasiswa yang akan diprediksi kelulusannya. Halaman target ini terdapat dua macam cara prediksi, yang pertama yaitu memprediksi kelulusan mahasiswa secara individu atau satu per satu, dan yang kedua memprediksi kelulusan mahasiswa secara bersamaan dengan cara import file dari excel.

APLIKASI PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA C<sub>4.5</sub>

HEADER

TRAINING ENTROPY & GAIN RULE TARGET LAPORAN

IMPLEMENTASI TARGET

MASUKKAN INISIAL (NIM)

PILIH ASAL DAERAH

PILIH ASAL SEKOLAH

PILIH JALUR MASUK

MASUKKAN INDEKS PRESTASI KUMULATIF (IPK)

MASUKKAN INDEKS PRESTASI SEMESTER 1

MASUKKAN INDEKS PRESTASI SEMESTER 2

MASUKKAN INDEKS PRESTASI SEMESTER 3

MASUKKAN INDEKS PRESTASI SEMESTER 4

MASUKKAN INDEKS PRESTASI SEMESTER 5

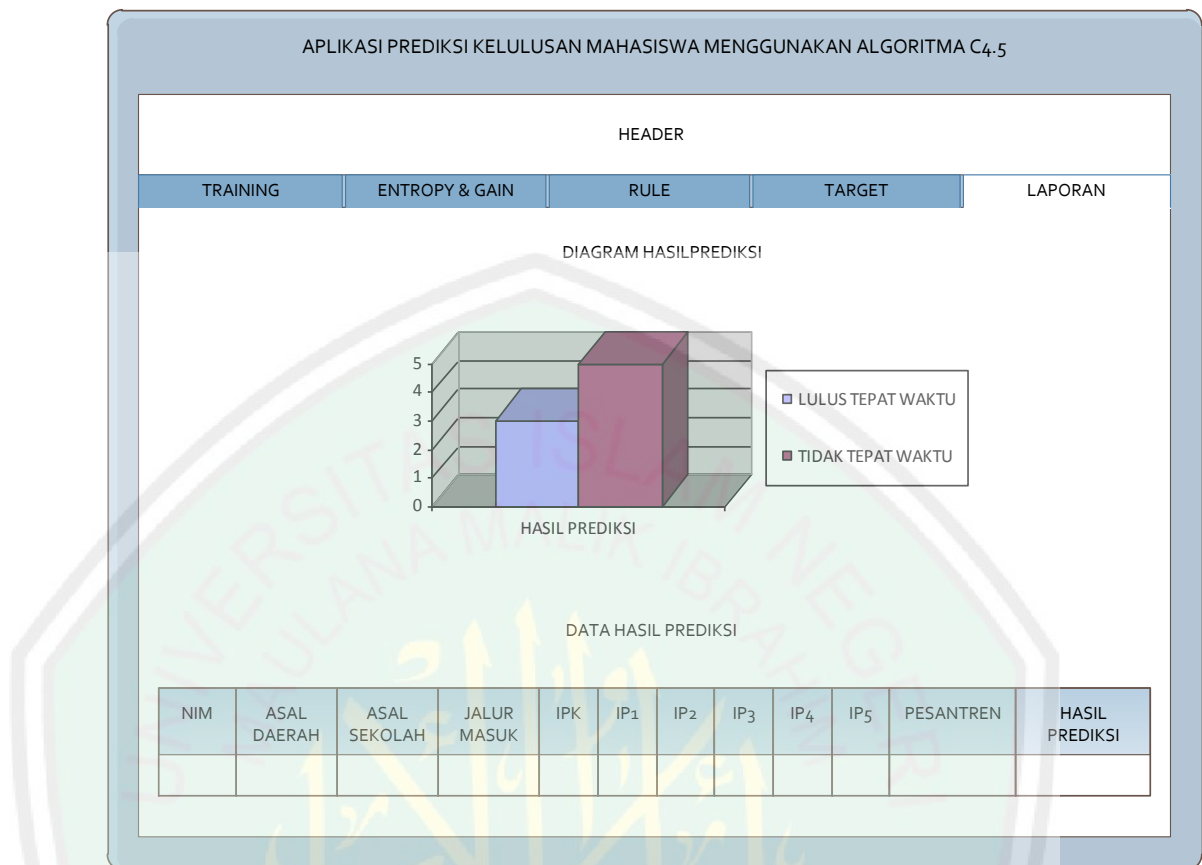
KETERANGAN PESANTREN

CEK KETEPATAN

Gambar 3.17 Halaman Target

## 6. Halaman Laporan

Halaman ini berisi keterangan dari hasil prediksi kelulusan meliputi jumlah mahasiswa yang diprediksi lulus tepat waktu dan tidak beserta data hasil prediksi kelulusan mahasiswa.



Gambar 3.18 Halaman Laporan

### 3.6 Perancangan Uji Coba

Pada subbab ini dilakukan perancangan uji coba terhadap algoritma C<sub>4.5</sub> yang telah dibangun. Pada uji coba ini akan membandingkan akurasi yang diperoleh dari hasil perhitungan dengan data sesungguhnya.

#### 3.6.1 Tujuan Uji Coba

Beberapa hal yang menjadi tujuan pengujian sistem ini adalah :

- a. Memeriksa apakah algoritma C<sub>4.5</sub> dapat berjalan dengan baik.

- b. Pada fase test model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tersebut. Bila akurasinya mencukupi model ini dapat dipakai untuk prediksi kelas data yang belum diketahui.
- c. Mengidentifikasi pola-pola dalam knowledge based yang ditemukan untuk menilai apakah hipotesa yang ada telah tercapai.

### 3.6.2 Skenario Uji Coba

Dalam proses prediksi kelulusan mahasiswa ini disiapkan data mahasiswa yang telah lulus dari berbagai angkatan, data tersebut terdiri dari atribut-atribut pilihan yang diperlukan dalam perhitungan. Uji coba dilakukan sebanyak tiga kali dengan jumlah data training yang berbeda-beda. Pertama yaitu 60 data, kedua 79 data dan ketiga 90 data mahasiswa akan digunakan sebagai sampel pada proses training dan 93 data akan digunakan pada proses testing. Data tersebut akan diproses menggunakan metode *Decision Tree* untuk menemukan klasifikasi kasus yang belum diketahui, sehingga data hasil prediksi dengan data asli dapat dibandingkan tingkat kecocokan atau akurasinya.

Dalam kasus ini target yang ingin dicapai yaitu prediksi ketepatan waktu kelulusan mahasiswa sesuai dengan masa studi mahasiswa 4 tahun atau lebih. Bila masa studi kurang dari atau sama dengan 4 tahun berarti diklasifikasikan menjadi kasus yang lulus tepat waktu, bila lebih dari 4 tahun maka diklasifikasikan menjadi tidak lulus tepat waktu.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Deskripsi Program

Perancangan *desain interface* pada bab III diimplementasikan dengan menggunakan bahasa pemrograman web yaitu PHP pada editor dreamweaver 8 dan ditampilkan pada web browser sehingga menghasilkan *interface* aplikasi sebagai berikut:

##### 1. Halaman Login

Pada aplikasi ini terdapat dua user yaitu admin dan operator. Admin memiliki hak akses manajemen data training dan rule, sedangkan untuk operator memiliki hak akses manajemen data target serta laporan hasil prediksi berupa diagram. Berikut tampilan login pada program :



Gambar 4.1 Halaman Login

## 2. Halaman Admin

Halaman admin memiliki hak akses manajemen data training serta perhitungan *Entropy* dan *Gain* sesuai rumus algoritma C4.5 dan membentuk suatu rule dengan menerapkan metode *Decision Tree*. Rule tersebut nantinya akan menjadi klasifikasi keputusan dari data mahasiswa yang akan diprediksi. Inputan data training ini berupa import file excel dalam jumlah banyak. Berikut tampilan pada program :



Gambar 4.2 Halaman Utama

### a. Halaman Data Training

Halaman ini merupakan form inputan data sampel training yang meliputi beberapa atribut yaitu nim, asal daerah, asal sekolah, jalur masuk,

indeks prestasi kumulatif (IPK), indeks prestasi semester 1 sampai 5, keterangan pesantren dan keterangan kelulusan.

Pada halaman training terdapat dua menu yaitu tampil data training, dan import data training dari file excel. Tampilan awal merupakan hasil perhitungan dari data training yang diinputkan berupa nilai *Entropy* dan *Gain*. Pada form ini atribut dari masing-masing mahasiswa telah dikelompokkan dalam bentuk tabel, lengkap dengan keterangan jumlah mahasiswa yang lulus tepat waktu dan tidak tepat waktu beserta nilai *Entropy* dan *Gain* dari masing-masing atribut sesuai dengan hasil perhitungan algoritma C4.5, yang tampak seperti gambar dibawah ini:

Atribut	Nilai atribut.	S	N	Y	ENTROPY	GAIN
		301	102	199	0.923733682177	
					0	0
asal_daerah	ACEH	2	1	1	1	0

Gambar 4.3 Halaman Training

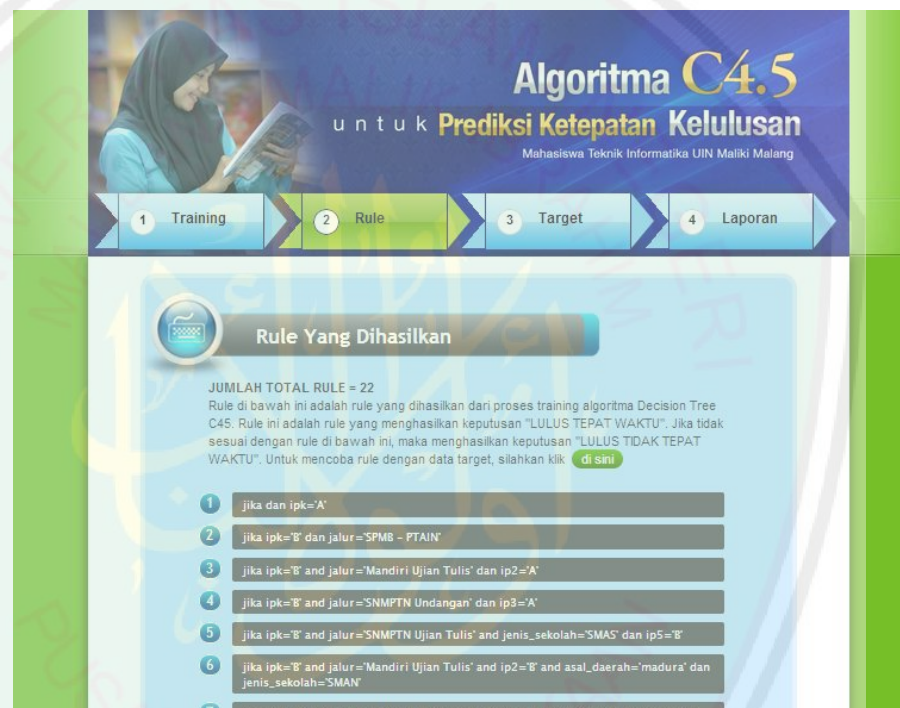
NO	NIM	Asal Daerah	Sekolah	Jalur	IPK	IP1	IP2	IP3	IP4	IP5	Pesantren	Keterangan	Delete
1	4550006	JAWA TIMUR	MAN	Mandiri Ujian Tulis	B	B	C	B	B	B	tidak	Tidak tepat	
2	4550009	JAWA TIMUR	SMAN	Mandiri Ujian Tulis	B	A	B	B	B	B	ya	tepat waktu	
3	4550010	JAWA TIMUR	MAN	Mandiri Prestasi	B	B	B	B	B	B	tidak	tepat waktu	

Gambar 4.4 Halaman Tampil Data Training

Gambar 4.5 Halaman Import Data Training

## b. Halaman Rule

Halaman rule ini merupakan halaman yang berisi kondisi dari data atribut yang dijadikan aturan dalam penentuan keputusan dalam prediksi kelulusan. Aturan tersebut diperoleh berdasarkan perhitungan *Entropy* dan *Gain* dari data sampel training yang diperoleh. Berikut tampilannya.



Gambar 4.6 Halaman Rule

Pada proses rule terdapat beberapa perulangan pada perhitungan berdasarkan perolehan data training atau data latih. Pada perulangan pertama terdapat beberapa *function* yaitu *function* awal yang menghasilkan gain tertinggi dan *function* getNode yang digunakan untuk mengklasifikasikan kasus menjadi aturan kasus lulus tepat waktu (Y), tidak lulus tepat waktu (N), atau Node baru. Perulangan kedua terdapat beberapa *function* yaitu

ambilLanjut dan getNodeLanjut. Perbedaannya yaitu atribut dengan gain tertinggi yang dihasilkan dari perhitungan sebelumnya tidak dimasukkan dalam array perhitungan selanjutnya.

```

while ($row2 = mysql_fetch_array($result))
{
    $y=(int)$row2['Y'];
    $n=(int)$row2['N'];

    if ($y>0 && $n>0)
    {
        $jmlCari+=1;
        $nodeBuat['cari'].$jmlCari=$row2[$nodeBaru];
    }

    else if ($y>0 || $n>0)
    {
        if($y>0) {
            $nodeBuat[$row2[$nodeBaru]]='yes';
            $aYes[]=$row2[$nodeBaru];
        }
        else
        {
            $nodeBuat[$row2[$nodeBaru]]='no';
            $no[]=$nodeBaru;
        }
    }
}

$hasilNode=array('atribut'=>$nodeBaru,'jenis'=>$nodeBuat,
'jml'=>$jmlRow,'level'=>$level, 'jmlCari'=>$jmlCari,
'where'=>$where);
//print_r($hasilNode);
return $hasilNode;

```

Source code dibawah ini digunakan untuk menampilkan rule yang dihasilkan dari perhitungan data training.

```

//CETAK RULE
//mengambil data array dari database
<?
$tesRule=unserialize(base64_decode($Hasil["rule"]))
;
$j=1;
foreach ($tesRule as $aturan)

```

```

        {
            list($nYes, $atribut, $where) = $aturan;
            echo '<li> <span>'. $j. '</span><a href="">'. ' jika ' .
            $where. ' dan ' . $atribut. ' '. $nYes. '</a></li>';
            $j++;
        } ?>

```

### 3. Halaman Operator

Halaman operator memiliki hak akses manajemen data target yang meliputi input data prediksi dan laporan hasil prediksi kelulusan mahasiswa berupa diagram. Input data prediksi dapat dilakukan secara satu persatu atau secara bersamaan dalam jumlah banyak dengan tehnik import data dari file excel. Berikut tampilan pada program :



Gambar 4.7 Halaman Operator

#### a. Halaman Target

Halaman target ini merupakan halaman untuk memprediksi kelulusan mahasiswa. Proses prediksi dapat dilakukan dengan dua cara yaitu menginputkan data per individu dan menginputkan data menggunakan import file excel sehingga dapat memprediksi banyak mahasiswa sekaligus. Atribut

yang diinputkan hampir sama seperti form training namun disini belum diketahui keterangan kelulusannya, setelah semua atribut diisi dan diproses maka akan muncul keterangan kelulusan dari tiap mahasiswa. Atribut yang diperlukan yaitu NIM, Asal Daerah, Asal Sekolah, Jalur Masuk, IPK, IP semester 1 sampai dengan 5, dan Pengalaman Pesantren. Berikut tampilan inputan data target per individu :

Gambar 4.8 Halaman Target

Setelah semua data mahasiswa berhasil diinputkan maka proses selanjutnya yaitu mengklasifikasikannya dengan membandingkan atribut data

mahasiswa dengan rule yang ada, sehingga hasil prediksi dapat menunjukkan keterangan lulus tepat waktu atau tidaknya mahasiswa tersebut, berikut tampilannya :



Gambar 4.9 Hasil Prediksi Tidak Lulus Tepat Waktu



Gambar 4.10 Hasil Prediksi Lulus Tepat Waktu

### b. Halaman Multi Target

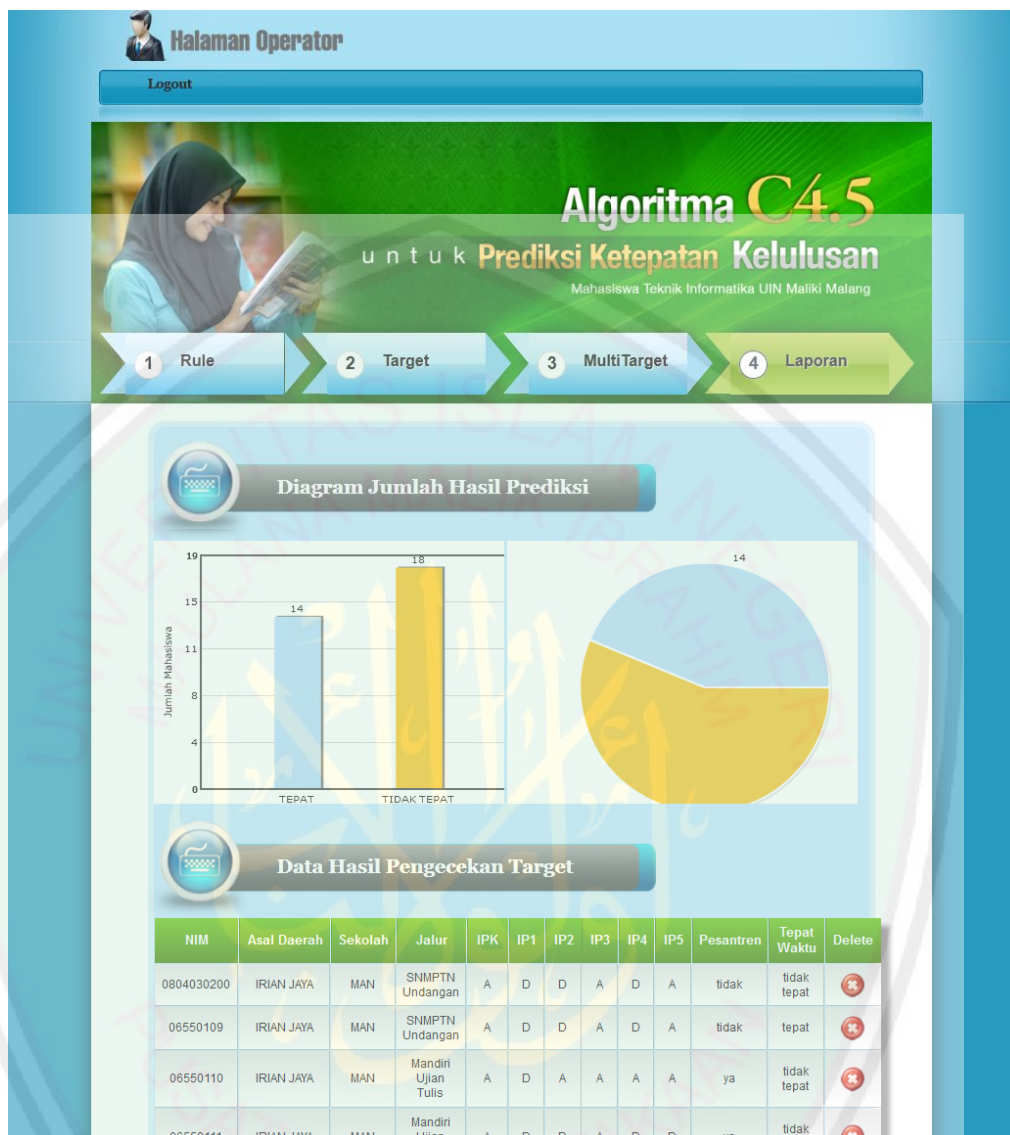
Halaman ini merupakan form inputan untuk data target yang dilakukan secara bersamaan yaitu memprediksi beberapa mahasiswa sekaligus dengan cara import file excel. Berikut tampilan pada program:



Gambar 4.11 Halaman Multi Target

### c. Halaman Laporan

Halaman laporan ini merupakan halaman yang berisi informasi jumlah mahasiswa yang lulus dan tidak lulus berdasarkan perhitungan data target yang telah eksekusi. Data tersebut divisualkan dalam bentuk diagram. Berikut tampilannya.



Gambar 4.12 Halaman Laporan

## 4.2 Implementasi Sistem

Pada tahap implementasi ini akan membahas alur algoritma dan sistem yang telah dirancang pada bab 3 diatas. Secara umum, fase algoritma C4.5 dalam membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai root
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Dibawah ini merupakan langkah-langkah perhitungan algoritma C4.5 yang diaplikasikan didalam program, mulai dari proses pembentukan *Tree*, penentuan rule hingga proses prediksi dan laporan mengenai jumlah kelulusan mahasiswa.

#### 4.2.1 Proses Perhitungan Data Training

Sebelum proses perhitungan dimulai, semua data diklasifikasikan berdasarkan target tujuan yaitu mengelompokkan data mahasiswa yang lulus tepat waktu dan yang tidak tepat waktu berdasarkan jumlahnya. Setelah itu baru dihitung nilai *Entropy* dan *Gain* untuk menentukan atribut terbaik yang ambil sebagai akar (root). Proses ini berdasarkan flowchart pada Gambar 3.2

```
//MENGHITUNG JUMLAH KASUS LULUS TEPAT WAKTU DAN TIDAK

function awal() {
    echo '<br><h1>di function awal </h1>';
    $id='1';
    $level='1';
    $atribut=array
('asal_daerah','jenis_sekolah','jalur','ipk','ip1','ip2','i
p3','ip4','ip5','pesantren');

    $queryTotal = "SELECT sum(S) as S,
                    sum(Y) as y,
                    sum(N) as n
```

```

from
  (SELECT jenis_sekolah,
  sum(jumlah) as S,
  SUM(IF(keterangan = 'Tidak tepat', jumlah, 0)) AS Y,
  SUM(IF(keterangan = 'tepat waktu', jumlah, 0)) AS N

FROM
  (SELECT Count(jenis_sekolah) as jumlah,jenis_sekolah,
  IF(keterangan = 'tepat waktu', 'tepat waktu', 'Tidak
tepat') AS keterangan
FROM
  training_baru
GROUP BY
  jenis_sekolah,
  keterangan
)
as temp_table
GROUP BY
  jenis_sekolah
)
as total";
$aGain=array();
$resultTot = mysql_query($queryTotal) or
die(mysql_error());
    $rowTot = mysql_fetch_array($resultTot);
    $S = $rowTot['S'];
    $y = $rowTot['y'];
    $n = $rowTot['n'];

```

#### 4.2.2 Membuat Total Entropy

Setelah semua data berhasil diidentifikasi jumlahnya berdasarkan pengelompokkan kelulusan maka langkah selanjutnya yaitu menghitung *Entropy* total dari semua jumlah data keseluruhan dan menghitung *Entropy* dari masing-masing nilai atribut yang ada. Berikut tampilan kode program yang dikonversi berdasarkan rumus *Entropy* pada halaman 14.

```

//MEMBUAT TOTAL ENTROPI

$EntropyTot=@(((-$rowTot['y']/$rowTot['S']))
*log($rowTot['y']/$rowTot['S'],2))
+((-$rowTot['n']/$rowTot['S'])*
log($rowTot['n']/$rowTot['S'],2));

```

### 4.2.3 Membuat *Gain*

Apabila nilai *Entropy* sudah diketahui maka langkah selanjutnya yaitu menghitung nilai *Gain* dari tiap-tiap atribut. Setelah semua atribut diketahui nilai *Gain*nya maka atribut dengan nilai *Gain* terbesar akan dijadikan akar (*root*) dalam pembentukan *Decision Tree* pada node 1. Selanjutnya sistem akan menunjuk atribut dengan *Gain* terbesar sebagai acuan dalam pembentukan pohon seterusnya. Nilai dalam atribut yang dijadikan root akan menjadi cabang dalam penentuan node 2 dan seterusnya. Proses akan berulang hingga semua atribut telah habis terhitung. Berikut tampilan kode program yang dikonversi berdasarkan rumus *Gain* pada halaman 13.

```
//MENJUMLAHKAN NILAI ENTROPY DARI SEMUA NILAI ATRIBUT
    $minGain+= ($row['S']/$S*$Entropy);

    if ($number==0)
    {
        $Gain=$EntropyTot-$minGain;

//MEMASUKKAN NILAI GAIN YANG DIHASILKAN DAN INDEX ARRAYNYA
ADALAH ATRIBUT
        $aGain[$atribut[$i]]=$Gain;
    }

    if ($number==0)
    {
        $Gain=$EntropyTot-$minGain;
        $aGain[$atribut[$i]]=$Gain;
    }
```

### 4.2.4 Menentukan *Gain* Maksimal

Nilai *Gain* yang ada akan dibandingkan dan dicari nilai *Gain* dari atribut mana yang paling tinggi. Apabila ditemukan maka atribut dengan *Gain* maksimal akan dipilih menjadi akar.

```
//MENENTUKAN GAIN MAKSIMAL

$nMax=(max($aGain));

    $nNode=array_search( $nMax,$aGain);

    $indexAtribut =array_search( $nNode,$atribut);

    $nilaiNode=
array('level'=>$level,'node'=>$nNode,'Gain'=>$nMax,'iAtribut'=>$i
ndexAtribut);

    return $nilaiNode;
```

#### 4.2.5 Menentukan Node Baru

Proses pembuatan pohon selanjutnya setelah diketahui akarnya yaitu menentukan node selanjutnya. Tiap nilai atribut akan dijadikan cabang pada node selanjutnya. Namun tidak semua cabang menjadi node, ada sebagian cabang yang terhenti atau mencapai *end of Tree* ketika sudah dapat terklasifikasi dalam keputusan lulus tepat waktu “Y” atau tidak lulus tepat waktu “N”. Cabang yang dijadikan node selanjutnya yaitu ketika nilai atributnya belum terklasifikasi keputusannya.

```
//MENENTUKAN NODE BARU

function getNode($arrayNode,$level)
{
    echo '<br><h1>di function getNode </h1>';
    $nodeBaru= $arrayNode['node'];

    $queryGetNode = "
    SELECT    $nodeBaru,
              sum(jumlah) as S,
              SUM(IF(keterangan = 'tepat waktu', jumlah, 0)) AS Y,
              SUM(IF(keterangan = 'Tidak tepat', jumlah, 0)) AS N
    FROM
    (SELECT
    Count( $nodeBaru ) as jumlah,  $nodeBaru,
    IF(keterangan = 'tepat waktu', 'tepat waktu', 'tidak tepat') AS
    keterangan
```

```

FROM training_baru GROUP BY $nodeBaru,keterangan )
as temp_table GROUP BY $nodeBaru ";

$result = mysql_query($queryGetNode) or die(mysql_error());
$jmlRow=mysql_num_rows($result);
$nodeBuat=array();
$jmlCari=0;
$i=0;
while ($row2 = mysql_fetch_array($result))
{ $y=(int)$row2['Y'];
  $n=(int)$row2['N'];
  if ($y>0 && $n>0)
  { $jmlCari+=1;
    $nodeBuat['cari'].$jmlCari]=$row2[$nodeBaru];
  }
  else if ($y>0 || $n>0)
  { if($y>0)
    { $nodeBuat[$row2[$nodeBaru]]='yes';}
    else
    { $nodeBuat[$row2[$nodeBaru]]='no';}
  } }
$hasilNode=array('atribut'=>$nodeBaru,'jenis'=>$nodeBuat,
                 'jml'=>$jmlRow,'level'=>$level,
                 'jmlCari'=>$jmlCari, 'where'=>'');
return $hasilNode;
}

```

#### 4.2.6 Perulangan Proses Perhitungan

Proses pembentukan pohon mulai dari menghitung jumlah kasus, menghitung *Entropy*, *Gain* dan penentuan node akan terus berulang sampai semua atribut yang ada berhasil dieksekusi secara keseluruhan.

```

//PERULANGAN PROSEES PERHITUNGAN

$ambilAwal=awal();
  $nodeBaru[$lvl][$m-1]= getNode($ambilAwal,$n);
  for($a=0;$a<(int)$nodeBaru[$lvl][$m-1]['jmlCari'];$a++)
    $atribut2[$lvl][0][$a]= array
('asal_daerah','jenis_sekolah','jalur','ipk','ip1','ip2','ip3','i
p4','ip5','pesantren');

  $ok=true;

  do
  {

    $jmlNode=0;
    $as=0;
    for($j=0;$j<$m;$j++){

```

```

        for ($i=0;$i<$nodeBaru[$lvl][$j]['jmlCari'];$i++)
        {
$vAtribut[$lvl][$j][$i]=$nodeBaru[$lvl][$j]['atribut'];
        $key =
array_search($vAtribut[$lvl][$j][$i], $atribut2[$lvl][$j][$i]);
        unset($atribut2[$lvl][$j][$i][$key]);
        $atribut[$lvl][$j][$i] =
array_values($atribut2[$lvl][$j][$i]);
        $where = $nodeBaru[$lvl][$j]['where'];

$nilaiAtribut=$nodeBaru[$lvl][$j]['jenis']['cari'.($i+1)];

        if ($n>0)
        {
            $where.=' and ';
        }

        $where .=
$vAtribut[$lvl][$j][$i].'\''.$nilaiAtribut.'\'';
        //echo '<br> WHERE perulangan =
        '.$where.'<br>';

$hasilAmbilLanjut=ambilLanjut($vAtribut[$lvl][$j][$i],$nilaiAtribut,
$atribut[$lvl][$j][$i],$where,$level.$i,$lvl);
        if($hasilAmbilLanjut['node']=="end"){
            $ok=false;
            break;
        }

$getNode=getNodeLanjut($hasilAmbilLanjut,$where,$lvl.'.'.($i+1));
        $nodeBaru[$lvl+1][$as]=$getNode;

        for ($k=0;$k<(int)$getNode['jmlCari'];$k++){

            $atribut2[$lvl+1][$as][$k]=$atribut2[$lvl][$j][$i];
            }
            $as++;
            $ms++;
        }
        }
        $n++;
        $m=$ms;
        $ms=0;
        $lvl++;
    }
} while (cek($nodeBaru[$lvl]) && $ok);

```

#### 4.2.7 Proses Rule

Pada proses *rule*, perhitungan yang dilakukan sesuai dengan prosedur metode *decision tree* yaitu mencari atribut dengan *gain* tertinggi untuk dijadikan akar, selanjutnya menentukan cabang berdasarkan nilai atribut didalamnya. Ada tiga kondisi klasifikasi kasus yang digunakan dalam penentuan rule yaitu :

1. Jika nilai  $N=0$  dan  $Y>0$  maka menghasilkan satu rule dengan keputusan lulus tepat waktu.
2. Jika nilai  $N>0$  dan  $Y=0$  maka menghasilkan satu rule dengan keputusan tidak lulus tepat waktu.
3. Jika  $N\geq 0$  dan  $Y\geq 0$  atau keduanya sama-sama mempunyai nilai, maka cabang itu membutuhkan perhitungan lagi sebagai node selanjutnya, sampai terdapat salah satu nilai diantara Y dan N untuk diklasifikasikan aturannya.

```

/// MENGAMBIL NILAI RULE YANG YES DAN DISIMPAN DALAM
DATABASE
    $p=0;
    $hasilrule=array();
    $jmlRule=0;
    foreach ($nodeBaru as $nb)
    {
        foreach ($nodeBaru[$p] as $nb1)
        {
            $hitung=0;
            foreach ($nb1["jenis"] as $key =>$nb2)
            {
                if ($nb2=='yes' || $nb2=='no')
                {
                    $jmlRule++;
                    if ($nb2=='yes')
                    {
                        $hasilrule[]=array ('0'=>'\'.$key.\'',
                        '1'=>$nb1["atribut"], '2'=>$nb1["where"]);
                    }
                }
            }
            $p++;
        }
    }
}

```

#### 4.2.8 Perhitungan Target

Proses perhitungan data target ini diinputkan manual dengan mengoreksi atribut yang diinputkan dengan rule hasil perhitungan dari data sampel training yang ada. Apakah kondisi atribut dari data mahasiswa yang diinputkan mempunyai keputusan lulus tepat waktu atau tidak.

```
//PROSES PERHITUNGAN TARGET

function cekHasil($nim,$rule)
{
    $hasilCek='tidak tepat';

    foreach ($rule as $aturan)
    {
        list($nYes, $atribut, $where) = $aturan;

        if ($where=='')
        {
            $whereCari = $atribut.$nYes;
        }
        else
        {
            $whereCari = $where.' and '.$atribut.$nYes;
        }

        $queryCari = 'SELECT * from tes where
        '.$whereCari.' and nim='.$nim;

        $queryhHasil=mysql_query($queryCari);
        $rowDataHasil = mysql_fetch_assoc($queryhHasil);

        if ( $rowDataHasil['nim']==$nim) {
            $hasilCek='tepat';
        }
    }
}
```

### 4.3 Hasil Uji Coba

Pada proses pengujian pada dasarnya membandingkan hasil prediksi dari aplikasi dengan data sesungguhnya. Prediksi kelulusan pada aplikasi ini dapat dilakukan dengan dua cara, yang pertama yaitu memprediksi kelulusan mahasiswa per individu dengan menginputkan data mahasiswa yang akan diprediksi satu per satu lalu dicek ketepatannya. Yang kedua yaitu menginputkan data testing dengan cara import file excel 97-2003 Workbook, cara kedua ini dapat memprediksi kelulusan beberapa mahasiswa sekaligus.

Proses prediksi pada aplikasi dilakukan melalui tahapan sebagai berikut :



Gambar 4.13 Proses Prediksi

1. Data diinputkan dengan field asal daerah, asal sekolah, jalur masuk, ipk, ip semester 1-5, dan pesantren. Atribut yang digunakan hampir sama seperti pada inputan data training, perbedaanya ada pada keterangan kelulusan. Pada data training keterangan kelulusan sudah diketahui agar klasifikasi kasus dapat dihitung menggunakan metode *Decision Tree* untuk mendapatkan aturan prediksi. Sedangkan pada data testing, atribut keterangan kelulusan belum diketahui dan akan diprediksi hasil kelulusannya.
2. Setelah berhasil diinputkan, tiap record dari atribut data testing akan dicocokkan dengan rule atau aturan yang terbentuk saat proses perhitungan data training.
3. Bila data testing yang diinputkan mempunyai record yang sama dengan rule diatas maka akan terklasifikasi menjadi kasus 'Y' yaitu lulus tepat waktu.

Sedangkan bila record dari data atribut yang dimasukkan tidak sama maka akan terklasifikasi menjadi keputusan ‘N’ yaitu tidak tepat waktu.

Uji coba dilakukan sebanyak tiga kali menggunakan jumlah data sampel training yang berbeda-beda. Yang pertama yaitu menggunakan 60 data mahasiswa, kedua 79 data dan ketiga 90 data. Sedangkan untuk data testing atau data uji menggunakan 93 data set mahasiswa Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pengumpulan data uji diambil dari beberapa data mahasiswa angkatan 2005 sampai 2008 yang sudah lulus. Hasil uji coba tersebut akan dicocokkan dengan hasil data sebenarnya apakah mahasiswa bersangkutan lulus tepat waktu atau tidak.

Aturan klasifikasi kasus lulus tepat waktu yang terbentuk pada proses training yaitu :

Tabel 4.1 Aturan (Rule) Hasil Uji Coba

Jumlah Data	Aturan (Rule) yang Didapat
60 Data	<ol style="list-style-type: none"> <li>1. Jika JALUR MASUK SPMB-PTAIN maka lulus tepat waktu</li> <li>2. Jika JALUR MASUK SNMPTN UNDANGAN dan JENIS SEKOLAH SMAN maka lulus tepat waktu</li> <li>3. Jika JALUR MASUK MANDIRI UJIAN TULIS dan ASAL DAERAH MADURA dan JENIS SEKOLAH SMAN maka lulus tepat waktu</li> <li>4. Jika JALUR MASUK SNMPTN UJIAN TULIS dan IP5 B dan JENIS SEKOLAH SMAS maka lulus tepat waktu</li> <li>5. Jika JALUR MASUK SNMPTN UNDANGAN dan JENIS SEKOLAH MAS dan ASAL DAERAH JAWA TIMUR maka lulus tepat waktu</li> </ol>

79 Data	<ol style="list-style-type: none"> <li>1. jika JALUR MASUK SPMB-PTAIN maka lulus tepat waktu</li> <li>2. jika JALUR MASUK SNMPTN UNDANGAN dan JENIS SEKOLAH MAS maka lulus tepat waktu.</li> <li>3. jika JALUR MASUK SNMPTN UNDANGAN dan JENIS SEKOLAH SMAN maka lulus tepat waktu.</li> <li>4. jika JALUR MASUK MANDIRI UJIAN TULIS dan ASAL DAERAH MADURA dan JENIS SEKOLAH SMAN maka lulus tepat waktu.</li> <li>5. jika Jalur Masuk SNMPTN UJIAN TULIS dan IP5 B dan JENIS SEKOLAH SMAS maka lulus tepat waktu.</li> </ol>
90 Data	<ol style="list-style-type: none"> <li>1. jika IPK A maka lulus tepat waktu</li> <li>2. jika IPK B dan JALUR MASUK SPMB-PTAIN maka lulus tepat waktu</li> <li>3. jika IPK B dan JALUR MASUK MANDIRI UJIAN TULIS dan IP2 A maka lulus tepat waktu</li> <li>4. jika IPK B dan JALUR MASUK SNMPTN UNDANGAN dan IP3 A maka lulus tepat waktu</li> <li>5. jika IPK B dan JALUR MASUK SNMPTN UJIAN TULIS dan JENIS SEKOLAH SMAS dan IP5 B maka lulus tepat waktu</li> <li>6. jika IPK B dan JALUR MASUK MANDIRI UJIAN TULIS dan IP2 B dan ASAL DAERAH MADURA dan JENIS SEKOLAH SMAN maka lulus tepat waktu</li> <li>7. jika IPK B dan JALUR MASUK SNMPTN UNDANGAN dan IP3 B dan JENIS SEKOLAH MAS dan ASAL DAERAH JAWA TIMUR maka lulus tepat waktu</li> </ol>

Dari pengujian tersebut dapat diketahui tingkat kebenaran pada tabel 4.2 dibawah ini :

Tabel 4.2 Hasil Pengujian Tingkat Kebenaran

No	Jumlah Data Training	Jumlah Aturan (Rule)	Jumlah Data Testing	Jumlah Benar	Akurasi (%)
1	60	17	93	60	64.51
2	79	19	93	66	70.96
3	90	22	93	77	82.79

Penentuan tingkat akurasi dapat dihitung menggunakan rumus sebagai berikut :

$$\frac{\text{Jumlah data yang benar}}{\text{Jumlah data uji (testing)}} \times 100\%$$

1. Prosentase kebenaran dari 60 data yaitu  $= \frac{60}{93} \times 100\%$   
 $= 0,6451 \times 100\%$   
 $= 64,51\%$
2. Prosentase kebenaran dari 79 data yaitu  $= \frac{66}{93} \times 100\%$   
 $= 0,7096 \times 100\%$   
 $= 70,96\%$
3. Prosentase kebenaran dari 90 data yaitu  $= \frac{77}{93} \times 100\%$   
 $= 0,8279 \times 100\%$   
 $= 82,79\%$

Dari hasil uji coba diatas menunjukkan bahwa penggunaan algoritma C4.5 sebagai pembentuk pohon keputusan dalam sistem prediksi kelulusan mahasiswa ini dapat digunakan karena telah terbukti keakuratannya. Aturan yang dihasilkan dari pohon keputusan menunjukkan penentuan klasifikasi kasus berdasarkan target yang ingin dicapai yaitu keputusan lulus tepat waktu atau tidak. Semakin banyak data sampel training yang dipakai, maka semakin besar pula tingkat akurasi kebenaran prediksinya.

#### 4.4 Hubungan Penelitian Dengan Kaidah Islam

Hasil dari penelitian yang telah dilakukan, memperlihatkan bahwa dengan adanya aplikasi prediksi kelulusan mahasiswa ini nantinya akan memberikan sebuah perubahan yang konstruktif utamanya dalam peningkatan mutu pendidikan. Berkaitan dengan hal tersebut maka diperlukan sebuah keberanian dan usaha yang nyata untuk mewujudkan sebuah perubahan yang lebih baik dalam dunia pendidikan khususnya di perguruan tinggi. Senada dengan hal tersebut, Allah SWT telah berfirman dalam Al-Qur'an Surat Ar-Ra'd 13 Ayat 11 yang berbunyi :

لَهُ مُعَقِّبَاتٌ مِّنْ بَيْنِ يَدَيْهِ وَمِنْ خَلْفِهِ يَحْفَظُونَهُ، مِنْ أَمْرِ اللَّهِ ۗ إِنَّ اللَّهَ لَا يُغَيِّرُ مَا بِقَوْمٍ حَتَّىٰ يُغَيِّرُوا مَا بِأَنْفُسِهِمْ ۗ وَإِذَا أَرَادَ اللَّهُ بِقَوْمٍ سُوءًا فَلَا مَرَدَّ لَهُ ۗ وَمَا لَهُمْ مِنْ دُونِهِ مِنْ وَالٍ ﴿١١﴾

Artinya : “Bagi manusia ada malaikat-malaikat yang selalu mengikutinya bergiliran, di muka dan di belakangnya, mereka menjaganya atas perintah Allah. Sesungguhnya Allah tidak merubah keadaan sesuatu kaum sehingga mereka merubah keadaan yang ada pada diri mereka sendiri. dan apabila Allah menghendaki keburukan terhadap sesuatu kaum, Maka tak ada yang dapat menolaknya; dan sekali-kali tak ada pelindung bagi mereka selain Dia.” (QS. Ar-Ra'd : 11)

(Baginya) manusia (ada malaikat-malaikat yang selalu mengikutinya bergiliran) para malaikat yang bertugas mengawasinya (di muka) di hadapannya (dan di belakangnya) dari belakangnya (mereka menjaganya atas perintah Allah) berdasarkan perintah Allah, dari gangguan jin dan makhluk-makhluk yang lainnya. (Sesungguhnya Allah tidak mengubah keadaan sesuatu kaum) artinya Dia tidak mencabut dari mereka nikmat-Nya (sehingga mereka mengubah keadaan yang ada pada diri mereka sendiri) dari keadaan yang baik dengan melakukan perbuatan durhaka. (Dan apabila Allah menghendaki keburukan terhadap suatu kaum) yakni

menimpakan azab (maka tak ada yang dapat menolaknya) dari siksaan-siksaan tersebut dan pula dari hal-hal lainnya yang telah dipastikan-Nya (dan sekali-kali tak ada bagi mereka) bagi orang-orang yang telah dikehendaki keburukan oleh Allah (selain Dia) selain Allah sendiri (seorang penolong pun) yang dapat mencegah datangnya azab Allah terhadap mereka.

Dalam hal perubahan ini tidak semata-mata berlaku pada tataran iman dan taqwa saja melainkan juga dalam dunia pendidikan yang menggunakan berbagai inovasi sebagai sarana peningkatan dari jenjang yang satu ke jenjang yang lebih tinggi secara bertahap. Dengan bantuan aplikasi ini diharapkan dapat dijadikan acuan untuk menyeleksi calon-calon mahasiswa sehingga nantinya akan diperoleh bibit-bibit unggul generasi penerus bangsa yang lebih baik.

Dalam proses pengembangan mutu pendidikan ini memerlukan suatu kerjasama yang baik antara mahasiswa dan pihak perguruan tinggi. Proses belajar mengajar, pelatihan, bimbingan dan motivasi sangat berperan dalam peningkatan mutu seorang individu. Dalam hal ini telah dijelaskan dalam Al-Qur'an surat Al-Maidah ayat 2 yang berbunyi,

يٰۤاَيُّهَا الَّذِيْنَ ءَامَنُوْا لَا تَحْلُوْا شَعْبِرَ اللّٰهِ وَلَا الشَّهْرَ الْحَرَامَ وَلَا اَهْدٰى وَلَا الْفَلَيْدَ وَلَا ءَامِيْنَ الْبَيْتِ  
الْحَرَامِ يَبْتَغُوْنَ فَضْلًا مِّنْ رَبِّهِمْ وَرِضْوَانًا ۚ وَاِذَا حَلَلْتُمْ فَاصْطَادُوْا ۗ وَلَا تَجْرِمْنٰكُمْ سَنَآئِنُ قَوْمٍ اَنْ  
صَدُّوْكُمْ عَنِ الْمَسْجِدِ الْحَرَامِ اَنْ تَعْتَدُوْا ۗ وَتَعَاوَنُوْا عَلٰى الْبِرِّ وَالتَّقْوٰى ۗ وَلَا تَعَاوَنُوْا عَلٰى الْاِثْمِ  
وَالْعُدُوْنِ ۗ وَاتَّقُوا اللّٰهَ ۗ اِنَّ اللّٰهَ شَدِيْدُ الْعِقَابِ ﴿٢﴾

*Artinya : “Hai orang-orang yang beriman, janganlah kamu melanggar syi'ar-syi'ar Allah, dan jangan melanggar kehormatan bulan-bulan haram, jangan (mengganggu) binatang-binatang had-ya, dan binatang-binatang qalaa-id, dan jangan (pula) mengganggu orang-orang yang mengunjungi Baitullah sedang mereka mencari*

*kurnia dan keredhaan dari Tuhannya dan apabila kamu Telah menyelesaikan ibadah haji, Maka bolehlah berburu. dan janganlah sekali-kali kebencian(mu) kepada sesuatu kaum Karena mereka menghalang-halangi kamu dari Masjidilharam, mendorongmu berbuat aniaya (kepada mereka). dan tolong-menolonglah kamu dalam (mengerjakan) kebajikan dan takwa, dan jangan tolong-menolong dalam berbuat dosa dan pelanggaran. dan bertakwalah kamu kepada Allah, Sesungguhnya Allah amat berat siksa-Nya. ”(QS. Al-Maidah : 2)*

(Hai orang-orang yang beriman, janganlah kamu melanggar syiar-syiar Allah) jamak sya`iraturun; artinya upacara-upacara agama-Nya. Melanggar yaitu dengan berburu di waktu ihram (dan jangan pula melanggar bulan haram) dengan melakukan peperangan padanya (dan jangan mengganggu binatang-binatang hadya) yakni hewan yang dihadiahkan buat tanah suci (serta binatang-binatang berkalung) jamak dari qilaadatun; artinya binatang yang diberi kalung dengan kayu-kayuan yang terdapat di tanah suci sebagai tanda agar ia aman, maka janganlah ada yang mengganggu baik hewan-hewan itu sendiri maupun para pemiliknya (jangan pula) kamu halalkan atau kamu ganggu (orang-orang yang berkunjung) atau menuju (Baitulharam) dengan memerangi mereka (sedangkan mereka mencari karunia) artinya rezeki (dari Tuhan mereka) dengan berniaga (dan keridaan) daripada-Nya di samping berkunjung ke Baitullah tidak seperti pengertian mereka yang salah itu. Ayat ini dimansukh oleh ayat Bara`ah. (Dan apabila kamu telah selesai) dari ihram (maka perintahkan berburu) perintah di sini berarti ibahah atau memperbolehkan (dan sekali-kali janganlah kamu terdorong oleh kebencian) dibaca syana-aanu atau syan-aanu berarti kebencian atau kemarahan (kepada suatu kaum disebabkan mereka telah menghalangi kamu dari Masjidilharam untuk berbuat aniaya) kepada mereka dengan pembunuhan dan sebaGainya. (Bertolong-tolonglah kamu dalam kebaikan) dalam mengerjakan yang

dititahkan (dan ketakwaan) dengan meninggalkan apa-apa yang dilarang (dan janganlah kamu bertolong-tolongan) pada ta`aawanu dibuang salah satu di antara dua ta pada asalnya (dalam berbuat dosa) atau maksiat (dan pelanggaran) artinya melampaui batas-batas ajaran Allah. (Dan bertakwalah kamu kepada Allah) takutlah kamu kepada azab siksa-Nya dengan menaati-Nya (sesungguhnya Allah amat berat siksa-Nya) bagi orang yang menentang-Nya.

Pada surat Al-Maidah ayat 2 di atas dijelaskan untuk saling tolong menolong dalam hal kemanusiaan dengan cara berqurban. Pada kenyataanya tolong menolong juga dapat dilakukan melalui berbagai macam cara, salah satunya yaitu saling membantu antar sesama umat dalam hal kebaikan seperti mendorong pada perbaikan dan pengembangan diri.

Dengan merenungkan ayat-ayat di atas diharapkan setiap orang dapat mengetahui kesadaran untuk menambah mutu pendidikan dalam rangka menimba ilmu di jalan Allah SWT. sehingga di masa depan akan tercipta generasi penerus bangsa yang baik dan bermutu.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Kesimpulan yang diperoleh dari hasil pengerjaan skripsi tentang sistem prediksi kelulusan mahasiswa menggunakan algoritma C4.5 yaitu :

1. Dari hasil uji coba menggunakan 60 data sampel, pola yang dibentuk mempunyai akurasi kecocokan sebesar 64,51%, sedangkan dari 79 data sampel menghasilkan 70,96%, dan dari 90 data sampel menghasilkan 82,79%.
2. Semakin banyak data sampel yang digunakan maka semakin besar pula kebenaran prediksinya, sehingga algoritma yang dibuat dapat digunakan untuk memprediksi kelas yang belum diketahui yaitu memprediksi kelulusan mahasiswa dapat tepat waktu atau tidak.
3. Sistem untuk memprediksi kelulusan mahasiswa ini telah berhasil dibuat menggunakan metode *Decision Tree* yaitu algoritma C4.5.

#### 5.2 Saran

Beberapa saran dari penulis untuk pengembangan penelitian lebih lanjut yaitu :

1. Membuat inputan secara dinamis untuk menambah atau mengubah atribut - atribut beserta *value* didalamnya yang digunakan dalam proses perhitungan algoritma C4.5.

2. Membuat bentuk *Tree* secara visual mengenai atribut yang bernilai *Gain* paling tinggi hingga proses perhitungan selesai sehingga dapat diketahui apa saja atribut yang menempati posisi *root*, *branch* dan *leaf*.



## DAFTAR PUSTAKA

- Alqur'an dan Terjemahnya*. 1971. Jakarta: Yayasan Penyelenggara Penterjemah/Pentafsir Al-Qur'an
- Al-Imam Jalaluddin Muhammad Al-Mahalli dan Al-Imam Jalaluddin Abdurrahman As-Suyuth. 2011. *Tafsir Jalalain*. Surabaya: Elba Fitrah Mandiri Sejahtera.
- Bahtiar, Agus. 2008. *PHP Script Most Wanted*. Yogyakarta: Andi Offset.
- Berry, Michael J.A. dan Gordon S. Linoff. 2004. *Data Mining Techniques For Marketing, Sales, Customer, Relationship Management*. Second Edition. Wiley Publishing, Inc.
- Faisal, Muhammad. 2008. *Sistem Informasi Manajemen Jaringan*. Malang: UIN-Malang Press.
- Harini, Sri dan Ririen Kusumawati. 2007. *Metode Statistika*. Jakarta: Prestasi Pustakaraya.
- Imarah, Musthafa M. 1933. *Terjemah Jawahirul Bukhari*. Penj. M.Zuhri. Darul Ihya' Indonesia.
- Kadir, Abdul. 2008. *Tuntunan Praktis Belajar Database Menggunakan MySQL*. Yogyakarta: Andi Offset.
- Kusrini, dan Emha Taufiq Luthfi. 2009. *Alogaritma Data Mining*. Yogyakarta: Andi Publisher.
- Kusrini. 2007. Design And Implementation Of Building Decision Tree Using C4.5 Algorithm.  
<http://elearning.amikom.ac.id/index.php/download/karya/586/a733b5873027ad0abaac6682499a3914> (diunduh pada tanggal 16 April 2012).
- Kusrini dan Sri Hartati. Implementation Of C4.5 Algorithm To Evaluate The Cancellation Possibility Of New Student Applicants At Stmik Amikom Yogyakarta.  
<http://repository.gunadarma.ac.id/bitstream/123456789/737/1/B-71.pdf> (diunduh pada tanggal 30 Mei 2012).
- Larose, Daniel .T. 2005. *Discovering Knowledge in Data*. New Jersey: John Willey & Sons.
- Pudjiantoro, dkk. 2011. Penerapan Data Mining Untuk Menganalisa Kemungkinan Pengunduran Diri Calon Mahasiswa Baru.  
<http://yudiagusta.files.wordpress.com/2008/09/051-060-knsi2011-009-penerapan-data-mining-untuk-menganalisa-kemungkinan-pengunduran-diri-calon-mahasiswa-baru.pdf> (diunduh pada tanggal 21 Mei 2012).

- Pramudiono, Iko. 2003. Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data.  
<http://www.scribd.com/doc/39630436/Data-Mining> (diunduh pada tanggal 31 Mei 2012).
- Santosa, Budi. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu: Yogyakarta.
- Suhartinah dan Marselina Silvia dan Ernastuti. 2010. Graduation Prediction Of Gunadarma University Students Using Algorithm And Naive Bayes C4.5 Algorithm.  
<http://www.papers.gunadarma.ac.id/index.php/industry/article/viewFile/816/787> (diunduh pada tanggal 30 Mei 2012).
- Sunjana. 2010. Aplikasi Mining Data Mahasiswa Dengan Metode Klasifikasi Decision Tree.  
<http://journal.uui.ac.id/index.php/Snati/article/viewFile/1857/1633> (diunduh pada tanggal 23 Agustus 2012).
- Suyanto. 2011. *Artificial Intelligence*. Informatika. Bandung.
- Widiarto dan Muchamad Piko Henry. 2011. Pengambilan Pola Kelulusan Tepat Waktu Pada Mahasiswa Stmik Amikom Yogyakarta Menggunakan Data Mining Algoritma C4.5. Yogyakarta: Sekolah Tinggi Manajemen Informatika Dan Komputer Amikom.  
[http://repository.amikom.ac.id/index.php/add\\_downloader/Publikasi\\_04.22.0400.pdf/1201](http://repository.amikom.ac.id/index.php/add_downloader/Publikasi_04.22.0400.pdf/1201) (diunduh pada tanggal 30 Mei 2012).
- <http://lecturer.eepis-its.edu/~tessy/lecturenotes/db2/bab10.pdf>  
(diunduh pada tanggal 23 Agustus 2012).

## LAMPIRAN 1 PERHITUNGAN MANUAL

Pada perhitungan manual disini menggunakan 79 data mahasiswa yang telah lulus dari berbagai angkatan. Semua data yang ada dimasukkan dalam bentuk tabel agar lebih mudah proses perhitungannya.:

ASAL DAERAH	JENIS SEKOLAH	JALUR MASUK	IPK	IP1	IP2	IP3	IP4	IP5	PESANTREN	KETERANGAN
jawa timur	MAN	Mandiri Ujian Tulis	C	B	B	B	B	B	Tidak	Tidak tepat
jawa timur	MAS	SNMPTN Ujian Tulis	B	C	C	B	B	C	Tidak	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	B	B	C	C	B	Tidak	Tidak tepat
madura	SMAN	SNMPTN Ujian Tulis	B	A	B	B	B	B	Tidak	Tidak tepat
jawa timur	MAN	Mandiri Ujian Tulis	B	C	B	C	B	B	Ya	Tidak tepat
jawa barat	SMKS	SNMPTN Ujian Tulis	B	B	B	B	B	B	Tidak	Tidak tepat
jawa timur	MAS	Mandiri Ujian Tulis	B	C	B	C	B	B	Ya	Tidak tepat
jawa timur	MAN	Mandiri Ujian Tulis	B	C	B	C	C	B	Tidak	Tidak tepat
jawa timur	MAS	Mandiri Ujian Tulis	B	B	C	B	B	C	Ya	Tidak tepat
jawa timur	MAS	Mandiri Ujian Tulis	B	B	B	B	B	B	Ya	Tidak tepat
jawa timur	MAN	SNMPTN Undangan	B	B	B	B	B	B	Tidak	Tidak tepat
jawa timur	SMAN	Mandiri Prestasi	B	B	B	B	B	B	Ya	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	B	B	B	B	B	Tidak	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	C	B	C	B	B	C	Tidak	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	B	C	B	B	B	Tidak	Tidak tepat
sumatra	MAN	Mandiri Ujian Tulis	B	B	B	C	B	C	Ya	Tidak tepat
sumatra	MAS	Mandiri Ujian Tulis	B	C	C	B	B	B	Ya	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	B	B	C	C	B	Tidak	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	B	C	B	B	C	Tidak	Tidak tepat
sumatra	SMAN	Mandiri Ujian Tulis	B	B	B	C	C	B	Tidak	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	C	C	B	B	B	Tidak	Tidak tepat
jawa timur	SMAS	Mandiri Ujian Tulis	B	D	C	B	B	C	Tidak	Tidak tepat
jawa timur	MAN	SNMPTN Ujian Tulis	B	B	B	B	B	C	Ya	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	D	C	B	B	B	Tidak	Tidak tepat
jawa timur	SMAN	Mandiri Prestasi	B	B	B	B	B	A	Tidak	Tidak tepat
jawa timur	MAN	Mandiri Ujian Tulis	C	A	B	B	B	B	Ya	Tidak tepat
jawa timur	SMAN	SNMPTN Ujian Tulis	B	B	B	C	B	B	Tidak	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	C	B	B	B	C	Ya	Tidak tepat
sumatra	MAS	Mandiri Ujian Tulis	B	B	B	C	C	B	Ya	Tidak tepat
sumatra	MAN	Mandiri Ujian Tulis	B	B	C	C	C	B	Tidak	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	B	C	C	B	B	Tidak	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	B	B	B	C	B	Tidak	Tidak tepat
jawa timur	MAN	Mandiri Ujian Tulis	B	C	C	B	B	B	Tidak	Tidak tepat

jawa timur	SMAN	Mandiri Prestasi	B	B	B	C	B	B	Tidak	Tidak tepat
jawa timur	SMAN	SNMPTN Undangan	B	C	B	B	B	B	Tidak	Tidak tepat
jawa timur	MAN	SNMPTN Ujian Tulis	B	C	B	C	B	B	Tidak	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	C	B	B	C	B	C	Ya	Tidak tepat
jawa timur	MAS	Mandiri Ujian Tulis	C	B	B	C	C	C	Ya	Tidak tepat
jawa timur	MAS	Mandiri Ujian Tulis	B	C	C	C	B	B	Ya	Tidak tepat
sumatra	MAS	SNMPTN Undangan	B	B	B	B	A	A	Ya	Tidak tepat
jawa timur	MAS	Mandiri Ujian Tulis	C	C	C	B	B	B	Ya	Tidak tepat
jawa timur	SMAS	Mandiri Ujian Tulis	B	B	B	B	C	B	Ya	Tidak tepat
kalimantan	MAS	Mandiri Ujian Tulis	C	B	B	B	A	B	Ya	Tidak tepat
jawa timur	MAN	SNMPTN Ujian Tulis	B	B	B	B	B	B	Tidak	Tidak tepat
jawa timur	MAN	SNMPTN Ujian Tulis	B	B	B	B	B	B	Tidak	Tidak tepat
jawa timur	MAN	Mandiri Ujian Tulis	C	C	B	B	B	B	Ya	Tidak tepat
madura	SMAS	Mandiri Ujian Tulis	B	B	B	C	C	B	Ya	Tidak tepat
jawa timur	MAS	Mandiri Prestasi	C	B	B	B	A	A	Tidak	Tidak tepat
jawa timur	SMAS	Mandiri Ujian Tulis	C	B	B	B	C	B	Tidak	Tidak tepat
jawa timur	SMKS	Mandiri Ujian Tulis	C	C	C	B	B	B	Ya	Tidak tepat
jawa timur	MAS	Mandiri Ujian Tulis	C	B	B	B	C	B	Tidak	Tidak tepat
jawa timur	SMKN	Mandiri Ujian Tulis	B	C	C	B	B	B	Ya	Tidak tepat
jawa timur	MAN	Mandiri Ujian Tulis	B	C	B	C	B	B	Tidak	Tidak tepat
jawa timur	SMAS	Mandiri Ujian Tulis	B	C	C	B	C	B	Tidak	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	C	B	B	B	B	Tidak	Tidak tepat
jawa timur	MAN	SNMPTN Ujian Tulis	B	A	B	B	B	B	Tidak	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	C	C	B	C	B	Tidak	Tidak tepat
jawa timur	MAN	SNMPTN Ujian Tulis	B	B	B	B	B	A	Ya	Tidak tepat
jawa timur	MAN	SNMPTN Undangan	B	B	B	B	B	B	Tidak	Tidak tepat
jawa timur	MAN	SNMPTN Undangan	B	B	B	B	B	B	Tidak	Tidak tepat
jawa timur	MAN	Mandiri Ujian Tulis	B	B	C	C	B	B	Tidak	Tidak tepat
jawa timur	SMAS	Mandiri Ujian Tulis	B	C	B	B	B	B	Ya	Tidak tepat
jawa timur	SMAN	SNMPTN Ujian Tulis	B	B	B	C	C	B	Tidak	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	B	B	B	B	C	Tidak	Tidak tepat
madura	SMAN	Mandiri Ujian Tulis	A	A	B	B	B	B	Tidak	tepat waktu
jawa timur	MAS	SNMPTN Undangan	B	B	B	B	B	B	Tidak	tepat waktu
jawa timur	SMAN	SNMPTN Undangan	A	B	B	B	B	A	Tidak	tepat waktu
madura	SMAN	Mandiri Ujian Tulis	B	B	B	B	B	B	Tidak	tepat waktu
jawa timur	SMAN	SNMPTN Ujian Tulis	B	B	B	B	B	B	Tidak	tepat waktu
jawa timur	SMAN	SNMPTN Undangan	B	B	B	A	B	B	Tidak	tepat waktu
jawa timur	MAN	SNMPTN Ujian Tulis	B	B	B	B	B	B	Tidak	tepat waktu
jawa timur	SMKS	SPMB - PTAIN	B	A	B	B	B	B	Ya	tepat waktu
jawa timur	SMAS	SNMPTN Ujian Tulis	B	B	B	B	B	B	Tidak	tepat waktu

jawa timur	SMAN	SNMPTN Ujian Tulis	B	B	B	B	B	B	Tidak	Tidak tepat
jawa timur	SMAS	SNMPTN Ujian Tulis	B	B	B	B	B	A	Ya	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	C	C	B	B	C	Ya	Tidak tepat
jawa timur	SMAN	Mandiri Ujian Tulis	B	B	C	C	B	B	Ya	Tidak tepat
jawa timur	SMAS	Mandiri Prestasi	B	B	B	B	B	A	Tidak	Tidak tepat
madura	SMAS	Mandiri Ujian Tulis	B	C	C	B	B	B	Ya	Tidak tepat

Berikut langkah-langkah perhitungannya :

1. Data dikelompokkan berdasarkan atribut beserta nilai didalamnya.
2. Menghitung jumlah data pada tiap nilai atribut yang ada
3. Mengklasifikasi data yang sudah dihitung menjadi dua kelompok berdasarkan target tujuan yaitu lulus tepat waktu (Y) dan tidak lulus tepat waktu (N).
4. Menghitung Entropy total dari 79 data mahasiswa
5. Menghitung Entropy dari masing-masing nilai atribut
6. Menghitung Gain dari tiap atribut
7. Mencari atribut dengan gain tertinggi untuk dijadikan root
8. Menentukan nilai atribut yang akan dijadikan cabang
9. Menentukan node selanjutnya dari atribut yang terpilih berdasarkan nilai gain tertinggi

Berikut rumus algoritma C4.5 yang digunakan pada pembentukan decision tree :

- a. Rumus mencari nilai Entropy :

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

Keterangan : S : himpunan Kasus

n : jumlah partisi S

pi : proporsi dari Si terhadap S

- b. Rumus mencari nilai Gain :

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

Keterangan : S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

|Si| : jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

### A. PENENTUAN ROOT

NODE	ATRIBUT	NILAI ATRIBUT	S	Y	N	ENTROPY	GAIN
0	<b>TOTAL</b>		<b>79</b>	<b>9</b>	<b>70</b>	<b>0.511639784</b>	
	ASAL DAERAH						0.042559524
		JAWA BARAT	1	0	1	0	
		JAWA TIMUR	66	7	59	0.487917993	
		KALIMANTAN	1	0	1	0	
		MADURA	5	2	3	0.970950594	
		SUMATRA	6	0	6	0	
	JENIS SEKOLAH						0.063865605
		MAN	20	1	19	0.286396957	
		MAS	14	1	13	0.371232327	
		SMAN	31	5	26	0.637387499	
		SMAS	10	1	9	0.468995594	
		SMKN	1	0	1	0	
		SMKS	3	1	2	0.918295834	
	<b>JALUR MASUK</b>						<b>0.121386942</b>
		MANDIRI PRESTASI	5	0	5	0	
		<b>MANDIRI UJIAN TULIS</b>	<b>49</b>	<b>2</b>	<b>47</b>	<b>0.246022578</b>	
		<b>SNMPTN UJIAN TULIS</b>	<b>16</b>	<b>3</b>	<b>13</b>	<b>0.69621226</b>	
		<b>SNMPTN UNDANGAN</b>	<b>8</b>	<b>3</b>	<b>5</b>	<b>0.954434003</b>	
		SPMB PTAIN	1	1	0	0	
	IPK						0.106076167
		A	2	2	0	0	
		B	65	7	58	0.49291578	
		C	12	0	12	0	
	IP1						0.080415815
		A	5	2	3	0.970950594	
		B	50	7	43	0.584238812	
		C	22	0	22	0	
		D	2	0	2	0	
	IP2						0.057624521
		B	57	9	48	0.629249224	
		C	22	0	22	0	
	IP3						0.089437905
		A	1	1	0	0	
		B	57	8	49	0.58515699	
		C	21	0	21	0	
	IP4						0.045528553
		A	3	0	3	0	
		B	61	9	52	0.60365225	

	C	15	0	15	0	
IP5						0.028952697
	A	7	1	6	0.591672779	
	B	60	8	52	0.566509507	
	C	12	0	12	0	
PESANTREN						0.030741192
	TIDAK	50	8	42	0.634309555	
	YA	29	1	28	0.216396932	

Data dikelompokkan berdasarkan atribut dan nilai atributnya lalu dihitung jumlah keseluruhan, jumlah mahasiswa yang lulus tepat waktu dan tidak, lalu dihitung nilai entropy dan gain masing-masing atributnya.

Baris TOTAL kolom ENTROPY pada Tabel diatas dihitung dengan rumus, sebagai berikut :

$$Entropy (Total) = \left( -\frac{9}{79} * \log_2 \left( \frac{9}{79} \right) \right) + \left( -\frac{70}{79} * \log_2 \left( \frac{70}{79} \right) \right)$$

$$Entropy(Total) = 0.511639784$$

Perhitungan entropy pada tiap nilai atribut dihitung dengan cara yang sama dengan entropy total :

$$Entropy (Asal Daerah, Jawa Timur) = \left( -\frac{7}{66} * \log_2 \left( \frac{7}{66} \right) \right) + \left( -\frac{59}{66} * \log_2 \left( \frac{59}{66} \right) \right)$$

$$Entropy(Asal Daerah, Jawa Timur) = 0.487917993$$

Sementara itu nilai GAIN pada baris ASAL DAERAH dihitung dengan menggunakan rumus Gain, sebagai berikut

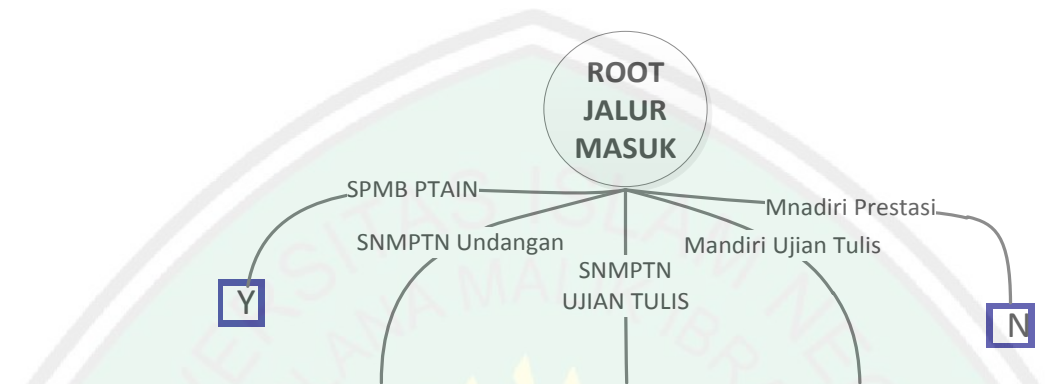
$$Gain (Total, Asal Daerah) = Entropy(Total) - \sum_{i=1}^n \frac{|Asal Daerah|}{|Total|} * Entropy(Asal Daerah)$$

$$Gain (Total, Asal Daerah) = 0.511 - \left( \left( \frac{1}{79} * 0 \right) + \left( \frac{66}{79} * 0.487 \right) + \left( \frac{1}{79} * 0 \right) + \left( \frac{5}{79} * 0.970 \right) + \left( \frac{6}{79} * 0 \right) \right)$$

$$Gain(Total, Asal Daerah) = 0.0425$$

Dari hasil pada table diatas dapat diketahui bahwa atribut dengan Gain tertinggi adalah JALUR MASUK yaitu sebesar 0.1213. Dengan demikian JALUR

MASUK menjadi node akar. Ada 5 nilai atribut dari JALUR MASUK yaitu MANDIRI PRESTASI, MANDIRI UJIAN TULIS, SNMPTN UJIN TULIS, SNMPTN UNDANGAN dan SPMB-PTAIN. Berikut Tree yang terbentuk :



Dari kelima nilai atribut tersebut diklasifikasikan berdasarkan nilai Y dan N dari atribut dengan gain terbesar yaitu sebagai berikut :

ATRIBUT	NILAI ATRIBUT	S	Y	N	ENTROPY	GAIN
JALUR MASUK						0.121386942
	MANDIRI PRESTASI	5	0	5	0	
	MANDIRI UJIAN TULIS	49	2	47	0.246022578	
	SNMPTN UJIAN TULIS	16	3	13	0.69621226	
	SNMPTN UNDANGAN	8	3	5	0.954434003	
	SPMB PTAIN	1	1	0	0	

Apabila nilai  $Y = 0$  dan  $N$  mempunyai nilai maka atribut tersebut menghasilkan 1 Rule yaitu Tidak Lulus Tepat Waktu, sedangkan bila  $N = 0$  dan  $Y$  mempunyai nilai maka akan menghasilkan 1 Rule Lulus Tepat waktu. Berdasarkan table diatas nilai atribut MANDIRI PRESTASI diketahui nilai  $Y = 0$  dan  $N = 5$  sehingga menghasilkan 1 Rule yaitu JIKA JALUR MASUK MANDIRI PRESTASI, MAKA TIDAK LULUS TEPAT WAKTU. Nilai atribut SPMB PTAIN mempunyai nilai  $Y=1$  dan  $N = 0$  sehingga menghasilkan 1 Rule lagi yaitu JIKA JALUR MASUK SPMB PTAIN, MAKA LULUS TEPAT WAKTU.

Sedangkan untuk MANDIRI UJIAN TULIS, SNMPTN UJIAN TULIS DAN SNMPTN UNDANGAN sama-sama mempunyai nilai  $Y$  dan  $N$  sehingga perlu dihitung lagi untuk mencari Node selanjutnya.

## B. LEVEL 0

Metode perhitungan untuk node selanjutnya sama dengan perhitungan Root diatas yaitu mengelompokkan data berdasarkan atribut dan dihitung jumlah mahasiswa yang lulus tepat waktu dan tidak lalu dicari nilai Gain tertinggi untuk dijadikan Node selanjutnya. Perbedaan pada perhitungan node selanjutnya yaitu atribut dengan gain tertinggi sebelumnya (root) tidak ikut dihitung lagi. Atribut total diganti dengan nilai atribut yang akan dihitung :

NODE	ATRIBUT		S	Y	N	ENTROPY	GAIN
<b>0</b>	<b>JALUR MASUK</b>	<b>MANDIRI UJIAN TULIS</b>	<b>49</b>	<b>2</b>	<b>47</b>	<b>0.246022578</b>	
	<b>ASAL DAERAH</b>						<b>0.164389925</b>
		JAWA TIMUR	39	0	39	0	
		KALIMANTAN	1	0	1	0	
		<b>MADURA</b>	<b>4</b>	<b>2</b>	<b>2</b>		<b>1</b>
		SUMATRA	5	0	5	0	
	<b>JENIS SEKOLAH</b>						<b>0.054595805</b>
		MAN	10	0	10	0	
		MAS	10	0	10	0	
		SMAN	20	2	18	0.468995594	
		SMAS	7	0	7	0	
		SMKN	1	0	1	0	
		SMKS	1	0	1	0	
	<b>IPK</b>						<b>0.110665956</b>
		A	1	1	0	0	
		B	37	1	36	0.179256067	
		C	11	0	11	0	
	<b>IP1</b>						<b>0.12122609</b>
		A	2	1	1	0	
		B	26	1	25	0.235193382	
		C	19	0	19	0	
		D	2	0	2	0	
	<b>IP2</b>						<b>0.03388982</b>
		B	28	2	26	0.371232327	
		C	21	0	21	0	
	<b>IP3</b>						<b>0.025751514</b>
		B	32	2	30	0.337290067	
		C	17	0	17	0	
	<b>IP4</b>						<b>0.02206877</b>
		A	1	0	1	0	

		B	34	2	32	0.322756959	
		C	14	0	14	0	
	IP5						0.013759068
		B	39	2	37	0.291818257	
		C	10	0	10	0	
	PESANTREN						0.038423953
		TIDAK	26	2	24	0.391243564	
		YA	23	0	23	0	

$$\text{Entropy (Jalur Masuk, Mandiri Ujian Tulis)} = \left(-\frac{2}{49} * \log_2 \left(\frac{2}{49}\right)\right) + \left(-\frac{47}{49} * \log_2 \left(\frac{47}{49}\right)\right)$$

$$\text{Entropy (Jalur Masuk, Mandiri Ujian Tulis)} = 0.246$$

$$\text{Entropy (Asal Daerah, Madura)} = \left(-\frac{2}{4} * \log_2 \left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} * \log_2 \left(\frac{2}{4}\right)\right)$$

$$\text{Entropy (Asal Daerah, Madura)} = 1$$

Gain tertinggi dari JALUR MASUK MANDIRI UJIAN TULIS yaitu ASAL DAERAH, berikut perhitungannya :

$$\text{Gain (Jalur Masuk Mandiri Ujian Tulis, Asal Daerah)} =$$

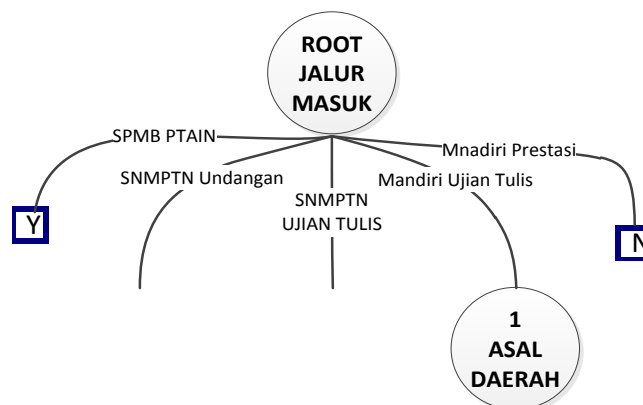
$$\text{Entropy (Jalur Masuk Mandiri Ujian Tulis)} - \sum_{i=1}^n \frac{|Asal Daerah|}{|Jalur Masuk Mandiri Ujian Tulis|} * \text{Entropy (Asal Daerah)}$$

$$\text{Entropy (Asal Daerah)}$$

$$\text{Gain (Jalur Masuk Mandiri Ujian Tulis, Asal Daerah)} = 0.246 - \left(\left(\frac{39}{49} * 0\right) + \left(\frac{1}{49} * 0\right) + \left(\frac{4}{49} * 1\right) + \left(\frac{5}{49} * 0\right)\right)$$

$$\text{Gain (Jalur Masuk Mandiri Ujian Tulis, Asal Daerah)} = 0.1643$$

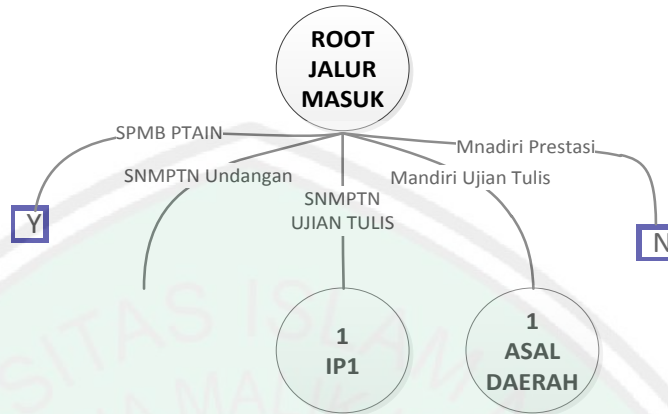
Sehingga Tree yang terbentuk sebagai berikut :



Selanjutnya yaitu menghitung nilai atribut SNMPTN UJIAN TULIS untuk mendapatkan Node selanjutnya, berikut tabel perhitungannya :

NODE	ATRIBUT		S	Y	N	ENTROPY	GAIN
0	JALUR MASUK	SNMPTN UJIAN TULIS	16	3	13	0.69621226	
	ASAL DAERAH						0.04031641
		JAWA BARAT	1	0	1	0	
		JAWA TIMUR	14	3	11	0.749595257	
		MADURA	1	0	1	0	
	JENIS SEKOLAH						0.08675289
		MAN	7	1	6	0.591672779	
		MAS	1	0	1	0	
		SMAN	5	1	4	0.721928095	
		SMAS	2	1	1	1	
		SMKS	1	0	1	0	
	IPK						0
		B	16	3	13	0.69621226	
	IP1						0.087753667
		A	2	0	2	0	
		B	12	3	9	0.811278124	
		C	2	0	2	0	
	IP2						0.019404671
		B	15	3	12	0.721928095	
		C	1	0	1	0	
	IP3						0.062990517
		B	13	3	10	0.779349837	
		C	3	0	3	0	
	IP4						0.019404671
		B	15	3	12	0.721928095	
		C	1	0	1	0	
	IP5						0.087753667
		A	2	0	2	0	
		B	12	3	9	0.811278124	
		C	2	0	2	0	
	PESANTREN						0.062990517
		TIDAK	13	3	10	0.779349837	
		YA	3	0	3	0	

Dari perhitungan tabel diatas menghasilkan tree sebagai berikut :

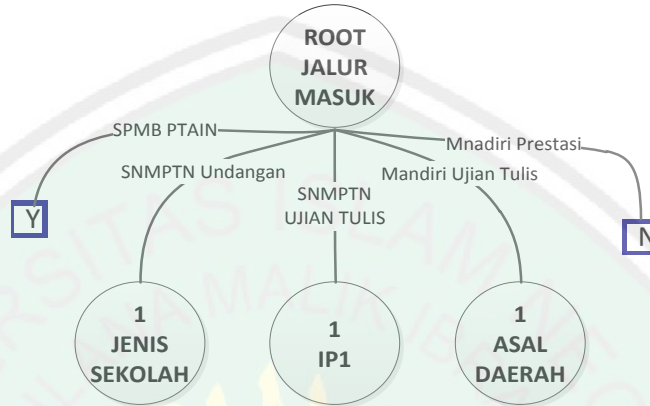


Selanjutnya yaitu menghitung nilai atribut SNMPTN UNDANGAN untuk mendapatkan Node selanjutnya, berikut tabel perhitungannya :

NODE	ATRIBUT		S	Y	N	ENTROPY	GAIN
<b>0</b>	<b>JALUR MASUK</b>	<b>SNMPTN UNDANGAN</b>	<b>8</b>	<b>3</b>	<b>5</b>	<b>0.954434003</b>	
	ASAL DAERAH						0.092359384
		JAWA TIMUR	7	3	4	0.985228136	
		SUMATRA	1	0	1	0	
	<b>JENIS SEKOLAH</b>						<b>0.360073065</b>
		MAN	3	0	3	0	
		MAS	2	1	1	1	
		SMAN	3	2	1	0.918295834	
	IPK						0.199203505
		A	1	1	0	0	
		B	7	2	5	0.863120569	
	IP1						0.092359384
		B	7	3	4	0.985228136	
		C	1	0	1	0	
	IP2						0
		B	8	3	5	0.954434003	
	IP3						0.199203505
		A	1	1	0	0	
		B	7	2	5	0.863120569	
	IP4						0.092359384
		A	1	0	1	0	
		B	7	3	4	0.985228136	
	IP5						0.015712127
		A	2	1	1	1	
		B	6	2	4	0.918295834	
	PESANTREN						0.092359384

		TIDAK	7	3	4	0.985228136	
		YA	1	0	1	0	

Dari perhitungan tabel diatas menghasilkan tree sebagai berikut :



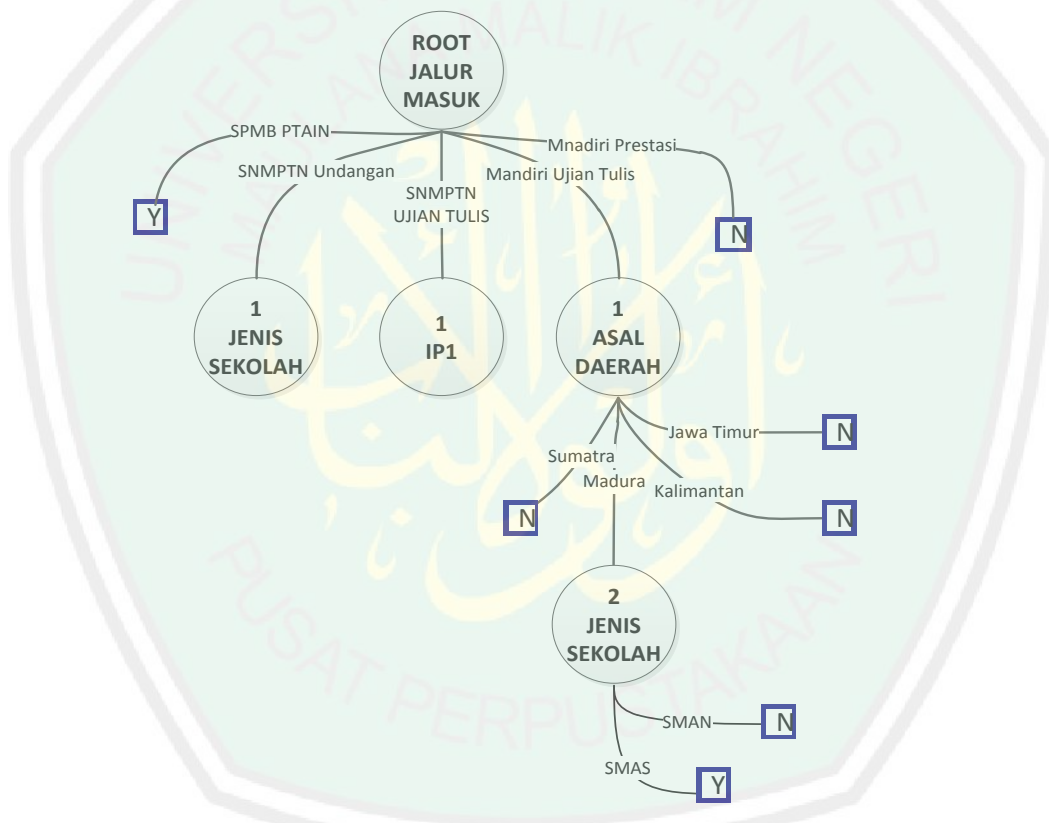
**C. LEVEL 1**

Langkah selanjutnya yaitu menghitung nilai entropy dari atribut JENIS SEKOLAH, IP1, DAN ASAL DAERAH. Sisa atribut yang akan dihitung yaitu JENIS SEKOLAH, IPK, IP1, IP2, IP3, IP4, IP5, dan PESANTREN. Berikut tabel perhitungannya :

NODE	ATRIBUT		S	Y	N	ENTROPY	GAIN
1	JALUR MASUK	MANDIRI UJIAN TULIS					
	ASAL DAERAH	MADURA	4	2	2		1
	JENIS SEKOLAH						1
		SMAN	2	2	0		0
		SMAS	2	0	2		0
	IPK						0.311278124
		A	1	1	0		0
		B	3	1	2	0.918295834	
	IP1						0.5
		A	1	1	0		0
		B	2	1	1		1
		C	1	0	1		0
	IP2						0.311278124
		B	3	2	1	0.918295834	
		C	1	0	1		0
	IP3						0.311278124
		B	3	2	1	0.918295834	
		C	1	0	1		0

	IP4						0.311278124
		B	3	2	1	0.918295834	
		C	1	0	1	0	
	IP5						0
		B	4	2	2	1	
	PESANTREN						1
		TIDAK	2	2	0	0	
		YA	2	0	2	0	

Dari perhitungan tabel diatas menghasilkan tree sebagai berikut :

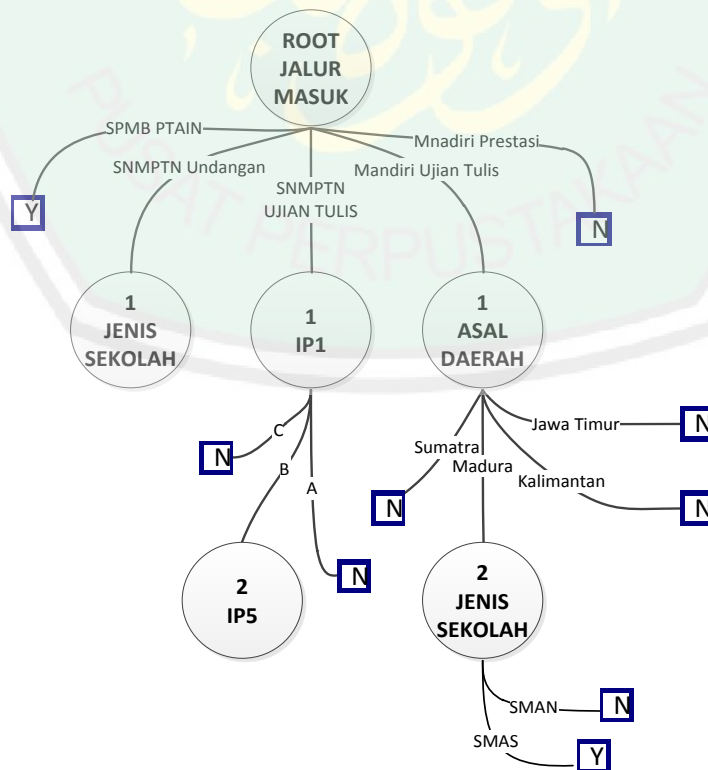


Selanjutnya yaitu menghitung nilai atribut SNMPTN UJIAN TULIS dan IP1 B untuk mendapatkan Node selanjutnya, berikut tabel perhitungannya :

NODE	ATRIBUT		S	Y	N	ENTROPY	GAIN
<b>1</b>	<b>JALUR MASUK</b>	<b>SNMPTN UJIAN TULIS</b>					
	IP1	B	12	3	9	0.811278124	
	ASAL DAERAH						0.036373099
		JAWA BARAT	1	0	1	0	
		JAWA TIMUR	11	3	8	0.845350937	
	JENIS SEKOLAH						0.073382043
		MAN	5	1	4	0.721928095	
		SMAN	4	1	3	0.811278124	

		SMAS	2	1	1	1
		SMKS	1	0	1	0
IPK						0
		B	12	3	9	0.811278124
IP2						0
		B	12	3	9	0.811278124
IP3						0.076869042
		B	10	3	7	0.881290899
		C	2	0	2	0
IP4						0.036373099
		B	11	3	8	0.845350937
		C	1	0	1	0
IP5						0.122556249
		A	2	0	2	0
		B	9	3	6	0.918295834
		C	1	0	1	0
PESANTREN						0.122556249
		TIDAK	9	3	6	0.918295834
		YA	3	0	3	0

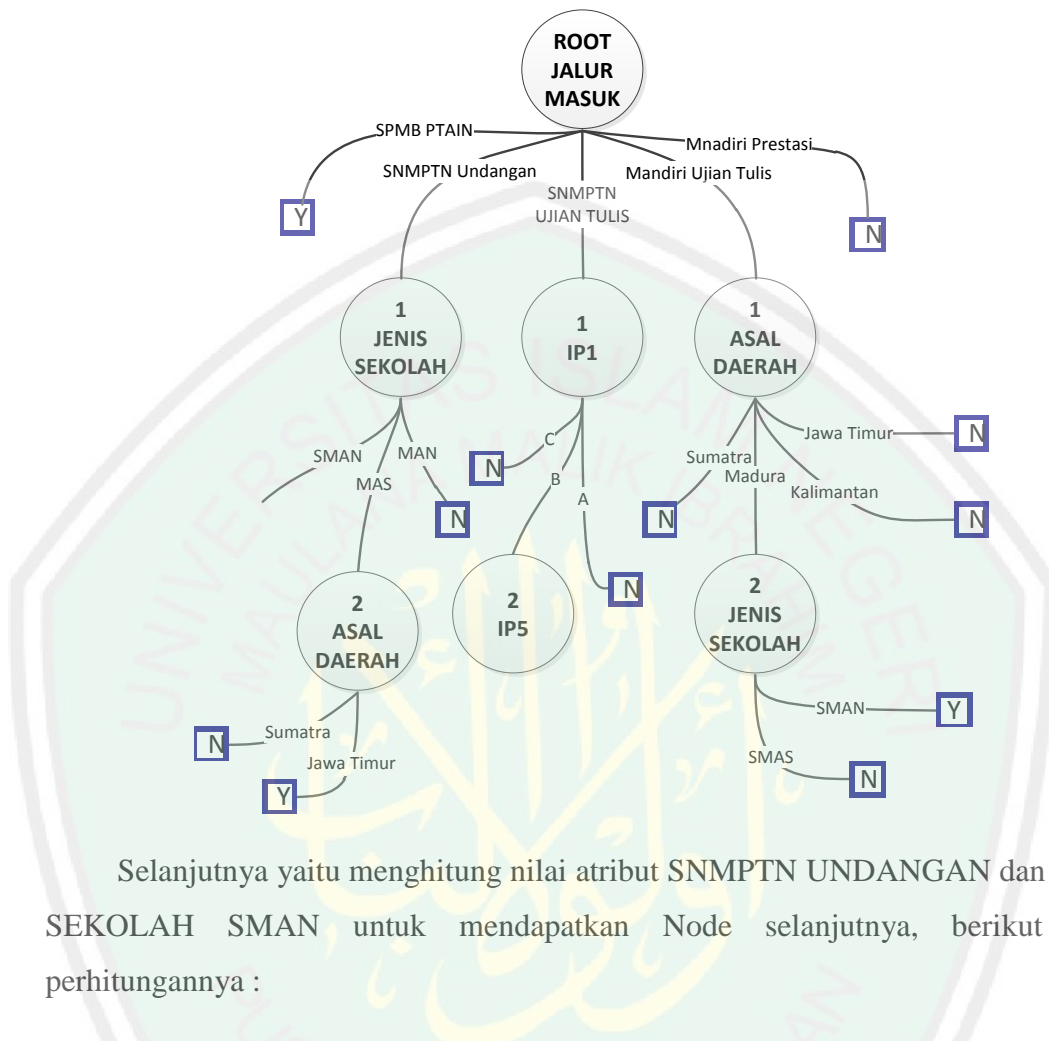
Dari perhitungan tabel diatas menghasilkan tree sebagai berikut :



Selanjutnya yaitu menghitung nilai atribut SNMPTN UNDANGAN dan JENIS SEKOLAH MAS untuk mendapatkan Node selanjutnya, berikut tabel perhitungannya :

NODE	ATRIBUT		S	Y	N	ENTROPY	GAIN
<b>1</b>	<b>JALUR MASUK</b>	<b>SNMPTN UNDANGAN</b>					
	<b>JENIS SEKOLAH</b>	<b>MAS</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>1</b>	
	<b>ASAL DAERAH</b>						<b>1</b>
		JAWA TIMUR	1	1	0	0	
		SUMATRA	1	0	1	0	
	IPK						0
		B	2	1	1	1	
	IP1						0
		B	2	1	1	1	
	IP2						0
		B	2	1	1	1	
	IP3						0
		B	2	1	1	1	
	IP4						1
		A	1	0	1	0	
		B	1	1	0	0	
	IP5						0.5
		A	1	0	1	1	
		B	1	1	0	0	
	PESANTREN						1
		TIDAK	1	1	0	0	
		YA	1	0	1	0	

Dari perhitungan tabel diatas menghasilkan tree sebagai berikut :

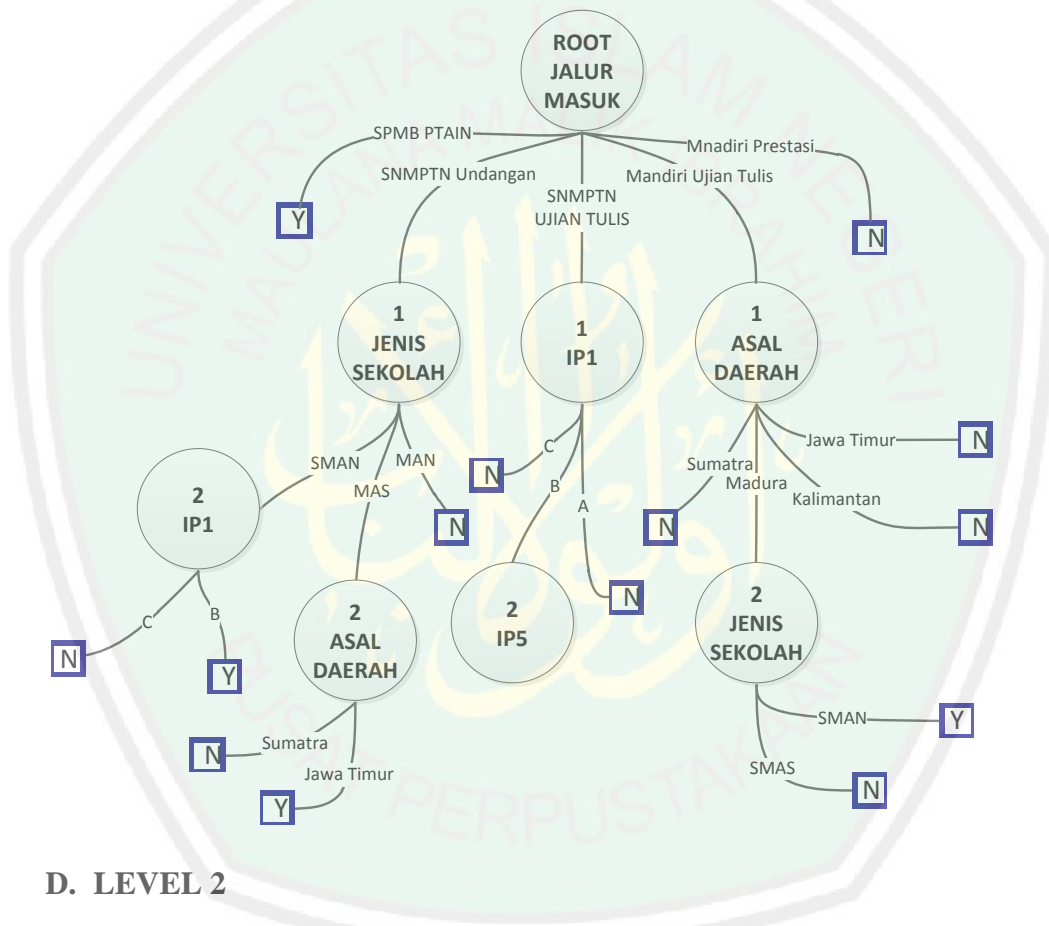


Selanjutnya yaitu menghitung nilai atribut SNMPTN UNDANGAN dan JENIS SEKOLAH SMAN untuk mendapatkan Node selanjutnya, berikut tabel perhitungannya :

NODE	ATRIBUT		S	Y	N	ENTROPY	GAIN
1	JALUR MASUK	SNMPTN UNDANGAN					
	JENIS SEKOLAH	SMAN	3	2	1	0.918295834	
	ASAL DAERAH						0
		JAWA TIMUR	3	2	1	0.918295834	
		IPK					0.251629167
		A	1	1	0	0	
		B	2	1	1	1	
	IP1						0.918295834
		B	2	2	0	0	
		C	1	0	1	0	
	IP2						0
		B	3	2	1	0.918295834	
	IP3						0.251629167
		A	1	1	0	0	
		B	2	1	1	1	
	IP4						0

		B	3	2	1	0.918295834	
	IP5						0.251629167
		A	1	1	0	0	
		B	2	1	1	1	
	PESANTREN						0
		TIDAK	3	2	1	0.918295834	

Dari perhitungan tabel diatas menghasilkan tree sebagai berikut :



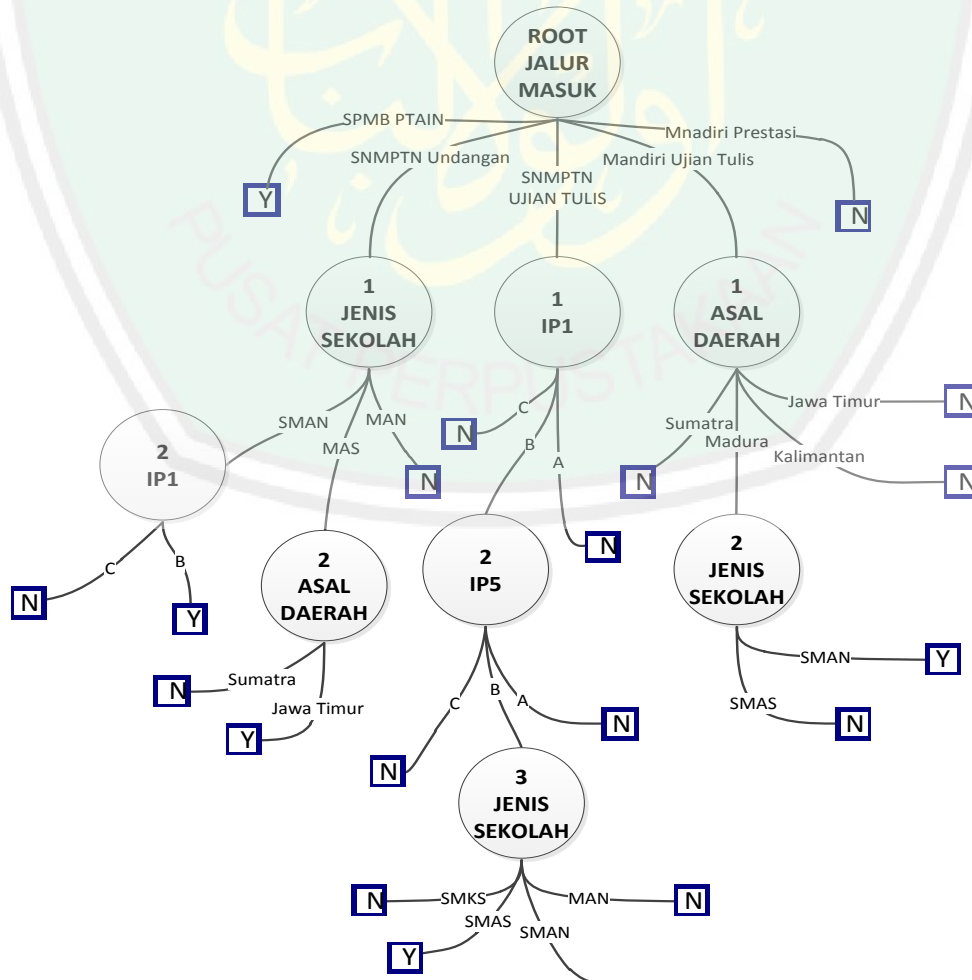
**D. LEVEL 2**

Langkah selanjutnya yaitu menghitung nilai entropy dari atribut IP5. Sisa atribut yang akan dihitung yaitu ASAL DAERAH, JENIS SEKOLAH, IPK, IP2, IP3, IP4, dan PESANTREN. Berikut tabel perhitungannya :

NODE	ATRIBUT		S	Y	N	ENTROPY	GAIN
2	JALUR MASUK	SNMPTN UJIAN TULIS					
	IP1	B					
	IP5	B	9	3	6	0.918295834	
	ASAL DAERAH						0.069910054
		JAWA BARAT	1	0	1	0	
		JAWA TIMUR	8	3	5	0.954434003	

	JENIS SEKOLAH						0.251629167
	MAN	3	1	2	0.918295834		
	SMAN	4	1	3	0.811278124		
	SMAS	1	1	0	0		
	SMKS	1	0	1	0		
IPK							0
	B	9	3	6	0.918295834		
IP2							0
	B	9	3	6	0.918295834		
IP3							0.152007284
	B	7	3	4	0.985228136		
	C	2	0	2	0		
IP4							0.069910054
	B	8	3	5	0.954434003		
	C	1	0	1	0		
PESANTREN							0
	TIDAK	9	3	6	0.918295834		

Dari perhitungan tabel diatas menghasilkan tree sebagai berikut :

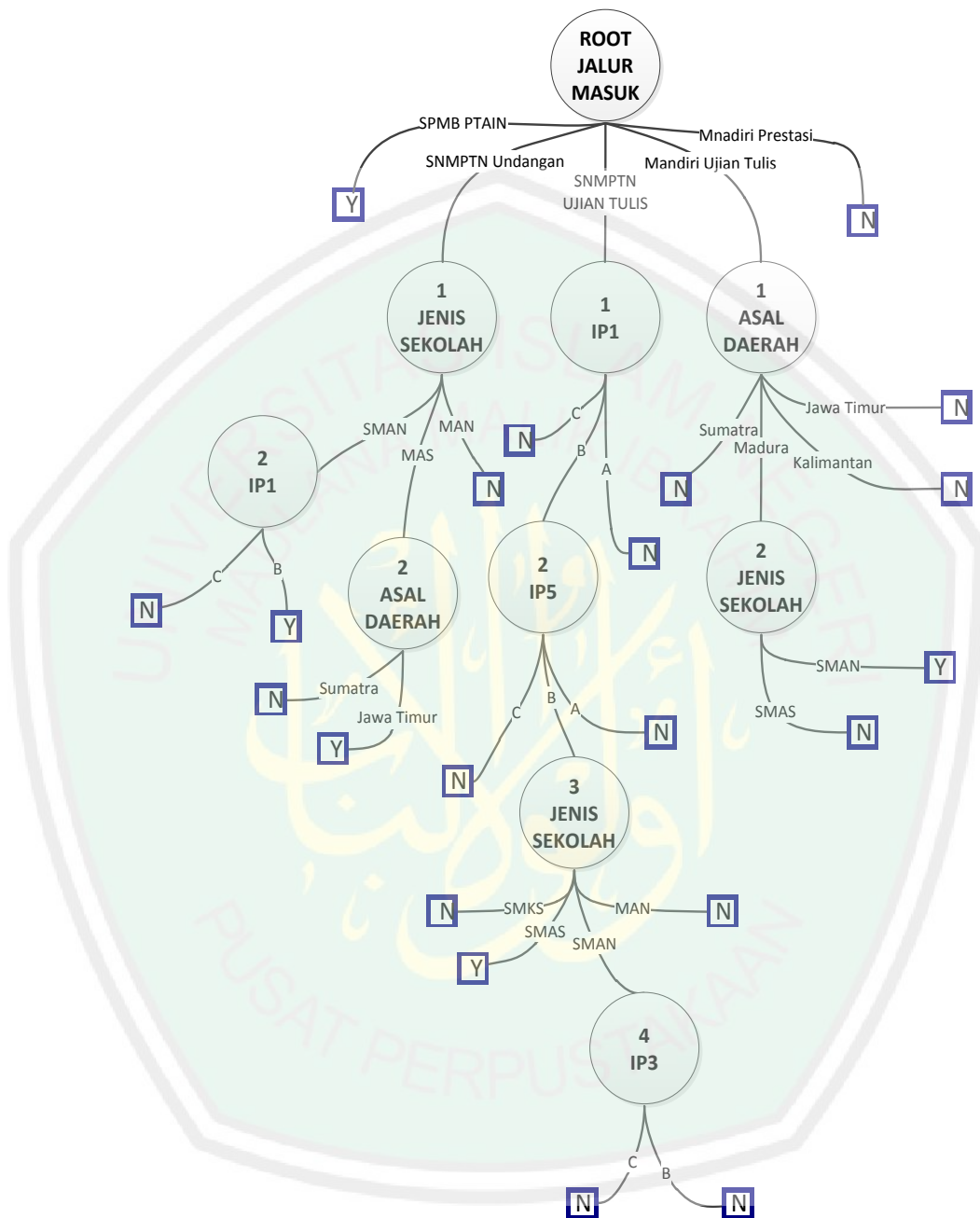


### E. LEVEL 3

Langkah selanjutnya yaitu menghitung nilai entropy dari atribut JENIS SEKOLAH. Sisa atribut yang akan dihitung yaitu ASAL DAERAH, IPK, IP2, IP3, IP4, dan PESANTREN. Berikut tabel perhitungannya :

NODE	ATRIBUT		S	Y	N	ENTROPY	GAIN
<b>3</b>	<b>JALUR MASUK</b>	<b>SNMPTN UJIAN TULIS</b>					
	IP1	B					
	IP5	B					
	JENIS SEKOLAH	SMAN	4	1	3	0.811278124	
	ASAL DAERAH						0
		JAWA TIMUR	4	1	3	0.811278124	
	IPK						0
		B	4	1	3	0.811278124	
	IP2						0
		B	4	1	3	0.811278124	
	IP3						0.311278124
		B	2	1	1		1
		C	2	0	2		0
	IP4						0.122556249
		B	3	1	2	0.918295834	
		C	1	0	1		0
	PESANTREN						0
		TIDAK	4	1	3	0.811278124	

Perhitungan pada level ini merupakan akhir dari proses pembentukan tree, semua atribut telah berhasil diklasifikasikan. Dari perhitungan tabel diatas menghasilkan tree sebagai berikut :



Rule atau aturan dari perhitungan 79 data sampel training menghasilkan 5 rule dengan keputusan lulus tepat waktu dan 16 rule dengan keputusan tidak lulus tepat waktu. Aturan ini digunakan pada proses testing yaitu memprediksi kelulusan mahasiswa yang belum diketahui. Tiap field dari data yang akan diprediksi dicocokkan dengan rule atau aturan yang terbentuk saat proses training. Berikut keterangannya:

**Prediksi Lulus Tepat Waktu (Y)**

1. Jika Jalur Masuk SPMB PTAIN maka Lulus Tepat Waktu
2. Jika Jalur Mandiri Ujian Tulis Dan Asal Daerah Madura dan Asal Sekolah SMAN maka Lulus Tepat Waktu
3. Jika Jalur Masuk SNMPTN Undangan dan Asal Sekolah MAS dan Asal Daerah Jawa Timur maka Lulus Tepat Waktu
4. Jika Jalur Masuk SNMPTN Undangan dan Asal Sekolah SMAN dan IP1 B maka Lulus Tepat Waktu
5. Jika Jalur Masuk SNMPTN Ujian Tulis dan IP1 B Dan IP5 B dan Jenis Sekolah SMAS maka Lulus Tepat Waktu

**Prediksi Tidak Lulus Tepat Waktu (N)**

1. Jika Jalur Masuk Mandiri Prestasi maka Tidak Lulus Tepat Waktu
2. Jika Jalur Masuk Mandiri Ujian Tulis dan Asal Daerah Jawa Timur maka Tidak Lulus Tepat Waktu
3. Jika Jalur Masuk Mandiri Ujian Tulis dan Asal Daerah Kalimantan maka Tidak Lulus Tepat Waktu
4. Jika Jalur Masuk Mandiri Ujian Tulis dan Asal Daerah Sumatra maka Tidak Lulus Tepat Waktu
5. Jika Jalur Masuk SNMPTN Ujian Tulis dan IP1 A maka Tidak Lulus Tepat Waktu
6. Jika Jalur Masuk SNMPTN Ujian Tulis dan IP1 C maka Tidak Lulus Tepat Waktu
7. Jika Jalur Masuk SNMPTN Undangan dan Asal Sekolah MAN maka Tidak Lulus Tepat Waktu
8. Jika Jalur Masuk Mandiri Ujian Tulis dan Asal Daerah Madura dan Asal Sekolah SMAS maka Tidak Lulus Tepat Waktu
9. Jika Jalur Masuk SNMPTN Ujian Tulis dan IP1 B dan IP5 A maka Tidak Lulus Tepat Waktu
10. Jika Jalur Masuk SNMPTN Ujian Tulis dan IP1 B dan IP5 C maka Tidak Lulus Tepat Waktu

11. Jika Jalur Masuk SNMPTN Undangan dan Jenis Sekolah MAS dan Asal Daerah Sumatra maka Tidak Lulus Tepat Waktu
12. Jika Jalur Masuk SNMPTN Undangan dan Jenis Sekolah SMAN dan IP1 C maka Tidak Lulus Tepat Waktu
13. Jika Jalur Masuk SNMPTN Ujian Tulis dan IP1 B dan IP5 B dan Jenis Sekolah MAN maka Tidak Lulus Tepat Waktu
14. Jika Jalur Masuk SNMPTN Ujian Tulis dan IP1 B dan IP5 B dan Jenis Sekolah SMKS maka Tidak Lulus Tepat Waktu
15. Jika Jalur Masuk SNMPTN Ujian Tulis dan IP1 B dan IP5 B dan Jenis Sekolah SMAN dan IP3 B maka Tidak Lulus Tepat Waktu
16. Jika Jalur Masuk SNMPTN Ujian Tulis dan IP1 B dan IP5 B dan Jenis Sekolah SMAN dan IP3 B maka Tidak Lulus Tepat Waktu

