



LAPORAN TUGAS AKHIR
IMPLEMENTASI DATA MINING ALGORITMA C4.5 DALAM
MEMPREDIKSI CALON MAHASISWA PADA PROGRAM
STUDI SISTEM INFORMASI UNIVERSITAS DIAN
NUSWANTORO

Disusun Oleh:

Nama : Alexander Devanda
Wisnu P
NIM : A12.2012.04604
Program Studi : Sistem Informasi - S1

FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO
SEMARANG
2016



LAPORAN TUGAS AKHIR
IMPLEMENTASI DATA MINING ALGORITMA C4.5 DALAM
MEMPREDIKSI CALON MAHASISWA PADA PROGRAM
STUDI SISTEM INFORMASI UNIVERSITAS DIAN
NUSWANTORO

Laporan ini disusun guna memenuhi salah satu syarat untuk menyelesaikan
program studi Sistem Informasi S-1 pada Fakultas Ilmu Komputer
Universitas Dian Nuswantoro

Disusun Oleh:

Nama : Alexander Devanda
Wisnu P
NIM : A12.2012.04604
Program Studi : Sistem Informasi - S1

FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO
SEMARANG

2016

PERSETUJUAN PROPOSAL TUGAS AKHIR

NB: Halaman ini tidak perlu dihapus dan dicetak

Silakan download halaman ini di siadin setelah ujian

PERSETUJUAN LAPORAN TUGAS AKHIR

NB: Halaman ini tidak perlu dihapus dan dicetak

Silakan download halaman ini di siadin setelah ujian

PENGESAHAN DEWAN PENGUJI

NB: Halaman ini tidak perlu dihapus dan dicetak
Silakan download halaman ini di siadin setelah ujian

PERNYATAAN KEASLIAN TUGAS AKHIR

NB: Halaman ini tidak perlu dihapus dan dicetak
Silakan download halaman ini di siadin setelah ujian

**PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH
UNTUK KEPENTINGAN AKADEMIS**

NB: Halaman ini tidak perlu dihapus dan dicetak
Silakan download halaman ini di siadin setelah ujian

UCAPAN TERIMAKASIH

Segala rasa syukur penulis panjatkan kehadiran Yesus Kristus atas segala karunia-Nya, sehingga laporan tugas akhir dengan judul “Implementasi Data Mining Algoritma C4.5 dalam Memprediksi Calon Mahasiswa Pada Program Studi Sistem Informasi Universitas Dian Nuswantoro”.

Skripsi disusun sebagai syarat untuk mencapai gelar sarjana komputer di Universitas Dian Nuswantoro Semarang. Skripsi ini dapat terselesaikan dengan baik tidak lepas dari dukungan dan bantuan dari berbagai pihak. Oleh karena itu penulis menyampaikan terima kasih kepada :

1. Dr. Ir. Edi Noersasongko, M.Kom, selaku Rektor Universitas Dian Nuswantoro Semarang.
2. Dr. Abdul Syukur, selaku Dekan Fasilkom.
3. Affandy, M.Kom, Ph.D, selaku Ketua Program Studi Sistem Informasi.
4. Acun Kardianawati, M.Kom, selaku pembimbing tugas akhir yang telah memberikan ide, saran, serta masukan yang berkaitan dengan penelitian yang dilakukan penulis.
5. Seluruh dosen di Fakultas Ilmu Komputer Sistem Informasi Universitas Dian Nuswantoro Semarang yang telah memberikan ilmu dan pengalamannya sehingga penulis dapat mengimplementasikan ilmu yang telah disampaikan.
6. Seluruh staf PSI Universitas Dian Nuswantoro yang telah memberikan data-data untuk keperluan penyusunan tugas akhir.
7. Ayah, Ibu dan Adik yang selalu memberikan kasih sayang, cinta, doa, dan semangat.
8. Untuk Kekasih, Dela Restania A yang telah memberikan waktu, doa, nasihat, ide dan motivasi kepada penulis untuk menyelesaikan Skripsi
9. Para sahabat seperjuangan M. Hadiyan, Risky Adrianto, Irya Adi S, Gufran, Heryono yang telah memberikan dukungan, semangat dan ide.

10. Keluarga pers mahasiswa “wartadinus” dan Keluarga BEM-KM Universitas Dian Nuswantoro kabinet “metamorfosa” yang selalu memberikan dukungan dalam menyelesaikan skripsi ini
11. Dan semua teman-teman FIK Universitas Dian Nuswantoro, khususnya untuk teman-teman pada program studi Sistem Informasi angkatan 2012 yang tidak bisa disebutkan satu persatu.

Semoga Tuhan Yang Maha Esa memberikan balasan yang lebih besar kepada beliau-beliau, dan pada akhirnya penulis berharap bahwa penulisan laporan tugas akhir ini dapat bermanfaat dan berguna sebagaimana fungsinya.

Semarang, 28 juli 2016

Penulis

ABSTRAK/RINGKASAN

Klasifikasi prediksi calon mahasiswa yang mendaftar pada program studi Sistem Informasi untuk memprediksi apakah calon mahasiswa mendaftar pada program studi Sistem Informasi atau tidak mendaftar pada program studi Sistem Informasi di Universitas Dian Nuswantoro. Klasifikasi prediksi menggunakan data mining dengan algoritma C4.5. Data yang digunakan pada penelitian yaitu data calon mahasiswa 2013 hingga 2015. Atribut yang digunakan dalam proses data mining algoritma C4.5 yaitu atribut tahun masuk, jenis kelamin, asal sekolah, kota asal, provinsi, jurusan, gelombang pendaftaran, tahun ijazah, dan jumlah nilai Ujian Nasional. Proses data mining dilakukan pada data training yaitu calon mahasiswa yang mendaftar pada program studi Sistem Informasi periode 2013 dan periode 2014 yang akan menghasilkan pohon keputusan atau rules. Metode pada tahap evaluasi dilakukan pada penelitian ini menggunakan confusion matrix dan nilai akurasi. Pengujian dilakukan 3 kali dengan prosentase 70% data training dan 30% data testing, data training 80% dan data testing 20%, pengujian terakhir dengan 90% data training dan 10% data testing. Didapatkan nilai akurasi tertinggi yaitu 80% pada pengujian ketiga dengan prosentase 90% data training dan 10% data testing. Hal ini membuktikan bahwa algoritma C4.5 akurat dalam memprediksi calon mahasiswa pada program studi Sistem Informasi Universitas Dian Nuswantoro.

Kata kunci : *Prediksi, Data Mining, Klasifikasi, C4.5, Sistem Informasi*

xv + 85 halaman; 31 gambar; 10 tabel; 2 lampiran

Daftar Acuan: 10 (2011 – 2015)

ABSTRACT

Predictions classification of prospective students who enroll in the information systems programs to predict whether the prospective student enrolls in information systems program or do not enroll in the information systems at the Universitas Dian Nuswantoro. Predictions classification used mining data by algorithm C4.5. Data used in this research is data of prospective students in 2013 and 2015. Attributes that are used in the process of data mining algorithms that attribute C4.5 year in, gender, home school, home town, province, department, a wave of registration, the diploma and the score of the National Examination. The data mining process is done on the training data that prospective students who enroll on the courses and information systems 2013 period 2014 period will result in a decision tree or rules. Method in the evaluation phase conducted in this study using the confusion matrix and accuracy values. Tests carried out three times with a percentage of 70% of the training data and testing data is 30%, 80% of training data and testing data is 20%, the final test with 90% of the training data and 10% of data testing. The highest value obtained accuracy of 80% in the third test with a percentage of 90% and 10% training data the data testing. This proves that the algorithm C4.5 accurate in predicting the students who enroll in Information System Program in Universitas Dian Nuswantoro.

Keywords : *Prediction, Data Mining, Classification, C4.5, Information System*
xvii + 85 pages; 31 images; 10 tables; 2 appendixes
References: 10 (2011 – 2015)

DAFTAR ISI

Halaman Judul	ii
Persetujuan Proposal Tugas Akhir	iii
Persetujuan Laporan Tugas Akhir	iv
Pengesahan Dewan Penguji	v
Pernyataan Keaslian Tugas Akhir	vi
Pernyataan Persetujuan Publikasi Karya Ilmiah Untuk Kepentingan Akademis ..	vii
Ucapan Terimakasih	viii
Abstrak/Ringkasan	x
Abstract	xi
Daftar Isi	xii
Daftar Gambar.....	xv
Daftar Tabel	xvii
Daftar Lampiran	xviii
Bab 1 Pendahuluan	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
Bab 2 TINJAUAN PUSTAKA	5
2.1 Tinjauan Pustaka	5
2.2 Data Mining.....	7
2.2.1 Pengertian Data Mining	7

2.2.2	Pengelompokan <i>Data Mining</i>	7
2.2.3	Process Knowledge Discovery in Database.....	8
2.3	<i>Classification Rules Mining</i>	10
2.3.1	<i>Decision tree</i>	11
2.3.2	Struktur Pohon Keputusan (<i>Decision Tree</i>).....	12
2.4	Algoritma C4.5.....	13
2.4.1	Langkah membangun pohon keputusan menggunakan C4.5.....	15
2.4.2	Entropy.....	15
2.4.3	Gain.....	17
2.4.4	Contoh Kasus C4.5	18
2.5	Confusion Matrix	19
2.6	Program Studi Sistem Informasi	20
2.6.1	Jalur Pendaftaran Calon Mahasiswa	21
2.7	Hubungan Data Mining dengan Sistem Informasi dan Universitas	22
2.7.1	Data mining dengan sistem informasi.....	22
2.7.2	Data Mining dengan Universitas.....	23
2.7.3	Penerapan Data Mining.....	25
2.8	WEKA 3.8.....	27
Bab 3	METODE PENELITIAN	29
3.1	Metode Pengumpulan Data	29
3.1.1	Survey	29
3.1.2	Studi Pustaka.....	29
3.2	Sumber Data.....	29
3.2.1	Data Primer	30
3.2.2	Data Sekunder	30

3.3	Metode yang Diusulkan	30
3.4	Langkah Implementasi pada proses KDD.....	30
Bab 4 HASIL DAN PEMBAHASAN		32
4.1	Pengumpulan Data	32
4.2	Penyeleksian Data	32
4.3	<i>Preprocessing/ Cleaning</i>	33
4.4	<i>Transformation data</i>	35
4.5	<i>Data Mining</i>	36
4.6	Implementasi menggunakan Aplikasi WEKA 3.8	71
4.7	<i>Intepretation/evaluation</i>	76
Bab 5 PENUTUP		81
5.1	Kesimpulan.....	81
5.2	Saran.....	81
DAFTAR PUSTAKA		82
LAMPIRAN.....		85

DAFTAR GAMBAR

Gambar 2.1 ProcessKnowledge Discovery in Database	9
Gambar 2.2 Struktur Pohon Keputusan	13
Gambar 2.3 Hasil Akhir Pohon Keputusan.....	19
Gambar 2.4 Cakupan Ilmu pada Data Mining	23
Gambar 2.5 Tugas Data Mining pada Sistem Institusi atau perusahaan.....	26
Gambar 2.6 Tampilan Proses Program WEKA 3.8	28
Gambar 4.1 Data Sample dengan 10 Atribute	33
Gambar 4.2 Data yang Telah Dilakukan Proses Cleaning.....	34
Gambar 4.3 Data yang Telah Dilakukan Proses Transformasi.....	35
Gambar 4.4 Data yang Telah Dilakukan Proses Transformasi.....	36
Gambar 4.5 Data Training Sample dari Nomor 1 - 34.....	37
Gambar 4.6 Data Training Sample Nomor 35-68.....	38
Gambar 4.7 Data Training Sample dari Nomor 69-100.....	39
Gambar 4.8 Pohon Keputusan Node 1	48
Gambar 4.9 Pohon Keputusan Node 2.....	56
Gambar 4.10 Pohon Keputusan Node 3.....	62
Gambar 4.11 Akhir Pohon Keputusan	67
Gambar 4.12 Pohon Keputusan Menggunakan WEKA	68
Gambar 4.13 Hasil Pengujian Sampel Data Training dan Data Testing.....	71
Gambar 4.14 Tampilan Awal WEKA.....	72
Gambar 4.15 Langkah Processing	72
Gambar 4.16 Proses Mencari Data	73

Gambar 4.17 Data Dibaca Oleh WEKA	73
Gambar 4.18 Proses Data Mining Algoritma C4.5 di WEKA.....	74
Gambar 4.19 Memasukkan Data Testing.....	74
Gambar 4.20 Memilih Class/ Atribbute	75
Gambar 4.21 Memulai Data Mining	75
Gambar 4.22 Hasil Data Mining Algoritma C4.5	76
Gambar 4.23 Hasil Pengujian 1	77
Gambar 4.24 Hasil Pengujian 2	78
Gambar 4.25 Hasil Pengujian 3	79

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait	5
Tabel 2.2 contoh kasus algoritma C4.5	18
Tabel 2.3 Confusion Matrix	20
Tabel 4.1 Tabel untuk Membentuk Node 1	47
Tabel 4.2 Tabel untuk Membentuk Node 2	54
Tabel 4.3 Tabel untuk Membentuk Node 3	61
Tabel 4.4 Tabel untuk Membentuk Node 4	66
Tabel 4.5 Pengujian 1	77
Tabel 4.6 Pengujian 2	78
Tabel 4.7 Pengujian 3	79

DAFTAR LAMPIRAN

Lampiran 1. Data Calon Mahasiswa yang mendaftar tahun 2013-2015.....	85
Lampiran 2. Jurnal Algoritma C4.5	125

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Dunia pendidikan erat kaitannya dengan siswa - siswi yang merupakan salah satu substansi yang perlu diperhatikan dan dibimbing. Siswa-siswi juga sebagai penerjemah terhadap suatu ilmu pengetahuan dan memiliki tugas untuk mendalami ilmu yang telah ada. Lulusan Sekolah Menengah Atas yang tiap tahunnya meningkat menyebabkan diperlukan tempat untuk melanjutkan cita-cita. Banyak yang memilih untuk meneruskan ke jenjang yang lebih tinggi menurut minat mereka, dan ada juga yang langsung turun ke dunia pekerjaan.

Semakin majunya era teknologi, semakin beragam pula ilmu pengetahuan ditawarkan dan minat akan seseorang dalam mendalami ilmu tersebut. Ilmu computer merupakan salah satu bidang ilmu yang banyak dibutuhkan didalam dunia pekerjaan. Kebutuhan yang dituntut untuk memahami komputasi, memicu minat lulusan Sekolah Menengah Atas dan sederajat yang ingin melanjutkan ke bangku perkuliahan untuk mendalami ilmu komputer. Banyak perguruan tinggi yang menawarkan studi ilmu komputer, dimana ilmu computer terbagi banyak jurusan-jurusan ilmu salah satunya yaitu jurusan Teknologi Informasi (TI), jurusan Sistem Informasi (SI) dan jurusan Desain Komunikasi Visual (DKV).

Universitas Dian Nuswantoro yang terletak di kota Semarang Jawa Tengah, merupakan universitas yang menawarkan studi ilmu komputer yaitu program studi Teknologi Informasi (TI) , Sistem Informasi (SI) dan Desain Komunikasi Visual (DKV). Namun dengan seiring perkembangan komunikasi dan penambahan program studi baru pada Universitas Dian Nuswantoro, Sistem Informasi telah mengalami penurunan calon mahasiswa yang mendaftar pada program studi tersebut. Penurunan terjadi pada tahun 2013 hingga tahun 2015. Jika program

studi Sistem Informasi terus mengalami penurunan maka akan berdampak pada akreditasi program studi tersebut.

Melihat kondisi tersebut dibutuhkan penelitian untuk menganalisis data yang dimiliki oleh program studi Sistem Informasi. Data yang dimanfaatkan untuk memprediksi calon mahasiswa adalah data mahasiswa tahun 2015 (Sebagai data testing) dan data calon mahasiswa tahun 2013-2014 pada program studi Sistem Informasi (Sebagai data training). Prediksi ini menggunakan algoritma C4.5 hasil akhirnya akan membentuk pohon keputusan dalam menentukan strategi yang tepat dan menghindari penurunan jumlah mahasiswa pada program Sistem Informasi.

Data mining merupakan kegiatan proses pengolahan data dengan melakukan perhitungan yang tepat untuk menghasilkan ilmu yang berguna. Dalam Penggunaan algoritma C4.5, dapat digunakan untuk menentukan dan memprediksi suatu objek. Dengan melakukan perhitungan data mining metode klasifikasi algoritma C4.5 kita dapat membuat pohon keputusan yang akan membantu program studi Sistem Informasi dalam memprediksi calon mahasiswa terhadap program studi Sistem Informasi di Universitas Dian Nuswantoro.

Berdasarkan permasalahan yang dibahas tersebut, maka penulis tertarik untuk mengambil judul dengan tema “IMPLEMENTASI DATA MINING ALGORITMA C4.5 DALAM MEMREDIKSI CALON MAHASISWA PADA PROGRAM STUDI SISTEM INFORMASI UNIVERSITAS DIAN NUSWANTORO”

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah dibahas diawal, penulis dapat merumuskan masalahnya yaitu bagaimana mengimplementasikan data mining algoritma C4.5 dalam memprediksi calon mahasiswa pada program studi Sistem Informasi di Universitas Dian Nuswantoro?

1.3 Batasan Masalah

Karena ruang lingkup masalah yang sangat luas di dalam penyusunan laporan tugas akhir ini maka penulis akan membatasi masalahnya yaitu:

1. Dataset yang digunakan data calon mahasiswa yang mendaftar pada program studi Sistem Informasi tahun 2013 - 2015
2. Pengolahan data hanya digunakan untuk menganalisa program studi Sistem Informasi dan hubungannya dengan minat calon mahasiswa pada program studi Sistem Informasi.
3. Menggunakan program WEKA 3.8 sebagai alat untuk pengujian

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dikemukakan diatas, maka tujuan dari penelitian ini adalah dengan mengimplementasikan data mining metode klasifikasi algoritma C.4.5 dapat memprediksi calon mahasiswa pada program studi Sistem Informasi Fakultas Ilmu Komputer Universitas Dian Nuswantoro dengan melihat dari pohon keputusan yang terbentuk dari hasil pengujian. Hasil dari pengujian algoritma C4.5 akan berguna untuk meningkatkan calon mahasiswa yang akan mendaftar ke program studi Sistem Informasi Fakultas Ilmu Komputer Universitas Dian Nuswantoro.

1.5 Manfaat Penelitian

Beberapa manfaat yang didapatkan dari penelitian ini adalah:

1. Bagi Program Studi Sistem Informasi

Membantu pihak program studi Sistem Informasi dalam meningkatkan jumlah mahasiswa dengan meningkatkan kualitas program studi Sistem Informasi dalam penentuan pohon keputusan program studi Sistem Informasi dan memberikan pengetahuan baru bagi pihak program studi Sistem Informasi dan sebagai strategi program studi Sistem Informasi dalam membantu meningkatkan calon mahasiswa terhadap program studi tersebut.

2. Bagi Universitas

Sebagai bahan informasi dan referensi kaitannya terhadap penelitian tentang minat calon mahasiswa terhadap program studi Sistem Informasi dengan menggunakan metode algoritma C4.5.

3. Bagi Penulis

- a. Sebagai sarana dalam menerapkan ilmu yang telah dipelajari selama masa perkuliahan di jurusan Sistem Informasi S-1.
- b. Sebagai salah satu syarat agar dapat menyelesaikan pendidikan di jurusan Sistem Informasi S-1 pada Fakultas Ilmu Komputer UDINUS.

BAB 2

TINJAUAN PUSTAKA

Penelitian tugas akhir yang dituliskan ini dibuat dengan terlebih dahulu melakukan studi keperpustakaan yang bersumber dari beberapa jurnal dan beberapa penelitian yang telah adadengan mendukung dan mengangkat topik yang serupa. Beberapa referensi yang diambil sebagai landasan teori diantaranya yaitu jurnal dan buku yang mengenai algoritma dan metode yang sesuai dengan topik.

2.1 Tinjauan Pustaka

Beberapa jurnal dan artikel yang terkaitdengan metode data mining teknik klasifikasi dengan algoritma yang sama dengan topik penelitian ini yaitu:

1. Implementasi Teknik Data Mining untuk Memprediksi Tingkat Kelulusan pada Universitas Bina Dharma Palembang (Andri,Yesi Novaria Kunang ,Sri Murniati , 2013) [1]
2. Analisa dan Penerapan Algoritma C4.5 Untuk Prediksi Studi Mahasiswa Berdasarkan Data Nilai Akademik (Selvi Lorena br Ginting, Wendi Zarman,Ida Hamidah, 2014) [2]
3. Penerapan Decision Tree Untuk Penentuan Pola Data Penerimaan Mahasiswa Baru (Aradea, Satriyo A, Ariyan Z,Yuliana A, 2011) [3]

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
1.	Andri,Yesi Novaria Kunang ,Sri Murniati ,	data kelulusan pada gudang data saat ini belum dimanfaatkan dengan	Data Mining metode <i>Classification algoritma</i>	Penerapan Algoritma J48 pada teknik Data Mining sangat efisien dalam mengklasifikasikan data mencapai keakuratan diatas 90%.

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
	2013	maksimal, sehingga kelulusan dari mahasiswa belum dikaji dan dilihat secara mudah.	<i>J48</i>	Hal ini menunjukkan bahwa algoritma <i>J48</i> tersebut memiliki performa yang sangat baik dalam melakukan klasifikasi pola
2.	Selvi Lorena br Ginting, Wendi Zarman, Ida Hamidah, 2014	Banyak terjadi mahasiswa – mahasiswi fakultas ilmu computer jurusan teknik informasi S1 yang masa studi tidak tepat waktu yang seharusnya 8 tahun masa studi tetapi telat hingga lebih dari 8 tahun	Data Mining metode <i>Classification algoritma C4.5</i>	Dapat memprediksikan masa studi dari mahasiswa-mahasiswi fakultas ilmu computer jurusan teknik informasi S1 di Universitas UNIKOM
3	Aradea, Satriyo A, Ariyan Z, Yuliana A,	Kurangnya penentuan pola data dalam	Metode decision tree algoritma	Untuk membantu klasifikasi Penerimaan Mahasiswa Baru dengan 3 atribut skor nilai UN,

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
	2011	penerimaan mahasiswa baru	ID3	prioritas progdi, dan jurusan SMA

2.2 Data Mining

2.2.1 Pengertian Data Mining

Data mining merupakan sebuah proses yang dapat mempekerjakan 1(satu) atau lebih dari teknik-teknik pembelajaran komputer (*machine learning*) untuk dapat menganalisa dan mengekstrasikan berbagai macam pengetahuan (*knowledge*) secara otomatis [1](Luthfi, 2009).beberapa ahli berpendapat bahwa Secara sederhana data mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies, 2004) sedangkan Pramudiono (2007) mengemukakan bahwa Data mining juga sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data. Beberapa teknik yang terdapat pada data mining adalah sebagai berikut:

1. *Clasification*
2. *Clustering*
3. *Association rule Discovery*
4. *Regression*
5. *Deviation detection*

2.2.2 Pengelompokan Data Mining

Pada data mining terdapat beberapa pengelompokan, pengelompokan tersebut berdasarkan tugas yang dikerjakan[4]:

1. Deskripsi

Mencari cara dalam menentukan sebuah pola dan kecenderungan yang ada didalam data. Deskripsi dari pola-pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali Variabel target untuk estimasi lebih mengarah *numeric* daripada variable kategori. Model dibangun menggunakan *record* lengkap yang akan memberikan nilai dari variable target sebagai nilai prediksi. Selanjutnya pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variable prediksi.

3. Prediksi

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

4. Klasifikasi

Dalam klasifikasi variabel, tujuan bersifat kategorik. Misalnya, kita akan mengklasifikasikan suhu dalam tiga kelas yaitu suhu panas, suhu sejuk, suhu dingin.

5. Pengclusteran

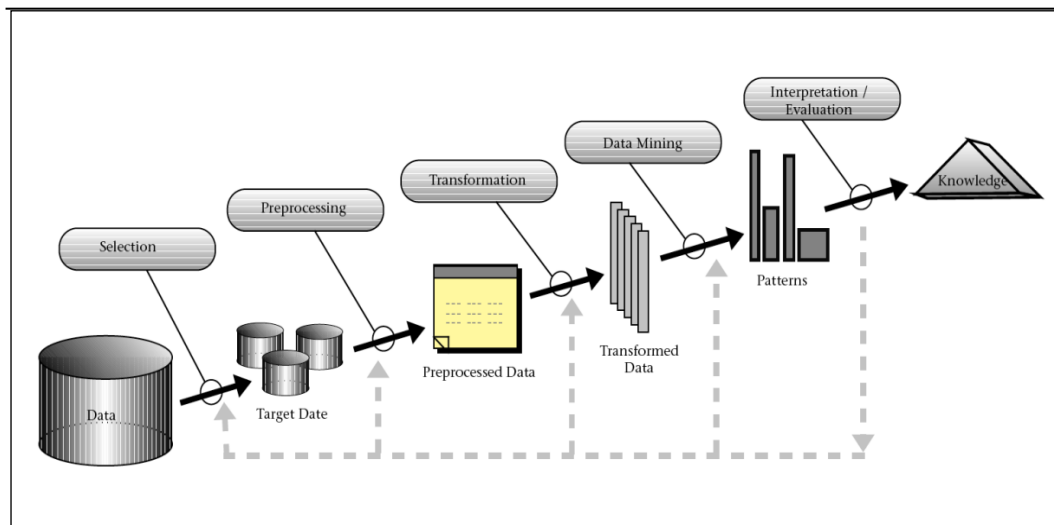
Merupakan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. kumpulan *record* yang memiliki kemiripan satu dengan yang lainya dan tidak memiliki kemiripan dengan *record-record* dalam cluster lain.

6. Asosiasi

Asosiasi merupakan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum.

2.2.3 Process Knowledge Discovery in Database

Istilah dalam data mining dan *knowledge discovery in database* (KDD) cukup sering digunakan dalam menjelaskan suatu proses untuk pencarian informasi tersembunyi pada suatu basis data yang besar[4]



Gambar 2.1 Process Knowledge Discovery in Database

Proses-proses yang terdapat pada KDD dibagi menjadi 5 , yaitu:[4]

1. Data Selection

Tahapan yang dilakukan dengan melakukan sebuah seleksi atau melakukan sebuah pemilihan data-data dari suatu kumpulan data operasional sebelum melakukan tahapan selanjutnya yaitu penggalian informasi dalam proses KDD. Data dari hasil penyeleksian tersebut dapat disimpan di dalam suatu berkas yang terpisah dari data operasional yang nantinya data seleksi akan digunakan untuk proses data mining.

2. Pre-Processing/Cleaning

Proses selanjutnya pada proses data mining yaitu melakukan pembersihan data atau cleaning data. Tahapan ini melakukan aktivitas pembuangan data yang dianggap tidak membantu , membuang data yang terdapat unsur duplikasi data, pemeriksaan data inkonsisten, dan juga melakukan perbaikan kesalahan-kesalahan yang terdapat pada data seperti contoh *tipografi*. Pada tahapan ini juga dilakukan proses untuk enrichment yang merupakan sebuah proses yang berguna untuk menambahkan atau memperbanyak data yang telah ada dengan informasi lain seperti data atau informasi eksternal.

3. *Transformation*

Proses *transformation* disebut juga dengan proses *coding* merupakan tahapanmentransformasikan pada data yang telah dipilih, sehingga data yang telah ada dapat sesuai dan dapat digunakan pada proses data mining. Jenis ataupun pola dari sebuah informasi yang di cari di dalam basis data bergantung pada proses ini.

4. *Data Mining*

Proses data mining merupakan suatu tahapan dalam mencari sebuah pola ataupun sebuah informasi menarik di dalam data terpilih dengan menggunakan teknik ataupun metode tertentu. Dengan beragamnya metode dan algoritma yang terdapat pada data mining, dibutuhkan sebuah pemilihan metode dan algoritma yang tepat dan akurat agar keseluruhan proses KDD dan tujuan yang diharapkan dapat tercapai dengan baik.

5. *Interpretation/Evaluation*

Hasil dari proses data mining yang berasal dari sebuah pola informasi perlu ditampilkan didalam bentuk yang dapat mudah dipahami oleh pihak perusahaan. Tahapan ini merupakan rangkaian dari proses KDD yang disebut dengan intepretation. Pada tahapan ini yaitu pemeriksaan apakah pola atau informasi yang telah didapatkan akan bertentangan dengan fakta ataupun hipotesis yang telah ada sebelumnya ataukah tidak.

2.3 *Classfication Rules Mining*

Klasifikasi merupakan proses yang terdapat pada data mining yang memiliki tujuan untuk menemukan sebuah pola yang berharga dari data yang memiliki ukuran relatif besar hingga data yang memiliki ukuran sangat besar. Data tersebut untuk saat ini dikelola menggunakan *Database Management System* (DBMS) baik database atau data warehouse.

Sampel-sampel yang akan disimpan secara keseluruhan dengan waktu secara bersamaan di memory. Algoritma C4.5 merupakan sebuah hal algoritma data

mining metode klasifikasi yang terkenal dan populer pada kelompok algoritma pohon keputusan. Algoritma C4.5 mengkonstruksi dan membangun sebuah pohon keputusan. Pada tahap selanjutnya yaitu tahap klasifikasi, pohon keputusan dipergunakan untuk memprediksi kelas-kelas dari sebuah permasalahan yang kelas-kelas tersebut belum diketahui. Prinsip kerja algoritma C4.5 ini dalam pemrosesannya adalah membaca keseluruhan permasalahan dari *storage* dan kemudian di masukkan ke dalam sebuah memori, kemudian melakukan kegiatan komputasi untuk membaca sampel-sampel di memori dengan mengkonstruksi dan membangun sebuah pohon keputusan.

Dengan menggunakan pendekatan algoritma C4.5 memiliki kelemahan yang termasuk dalam kategori skalabilitas adalah algoritma tersebut hanya dapat digunakan untuk menangani.

2.3.1 Decision tree

Pohon keputusan atau yang disebut dengan decision tree adalah pohon analisis pemecahan masalah yang pengambilan keputusan mengenai pemetaan dengan membuat suatu alternatif-alternatif yang nantinya digunakan untuk pemecahan masalah yang dapat diambil dari permasalahan tersebut. Decision tree atau pohon keputusan merupakan alat yang baik dalam membantu memilih beberapa tujuan dari beberapa tindakan[5]. Pohon keputusan ini juga menunjukkan banyak faktor kemungkinan-kemungkinan yang dapat mempengaruhi beberapa alternatif dari sebuah keputusan, yang kemudian disertai dengan sebuah estimasi dari hasil akhir yang akan diperoleh bila kita mengambil alternatif keputusan tersebut.

Decision tree mempergunakan struktur-struktur hierarki untuk pembelajaran *supervised*. Alur proses dari decision tree atau pohon keputusan diawali dari *root node* hingga menuju ke *leaf node* yang dilakukan secara rekursif. Dimana percabangan-percabangan yang ada digunakan untuk menyatakan kondisi yang harus dipenuhi dan pada setiap ujung-ujung pohon keputusan digunakan untuk menyatakan kelas dari suatu data.

Proses yang terdapat pada pohon keputusan adalah melakukan proses pengubahan bentuk data (tabel) menjadi bentuk model pohon (*tree*) yang kemudian mantransisi ke bentuk model pohon menjadi sebuah aturan (*rule*). Metode tersebut dipergunakan untuk melakukan perkiraan nilai diskret dari sebuah fungsi target yang dimana fungsi tersebut merupakan pembelajaran yang direpresentasikan oleh pohon keputusan (*decision tree*). Pohon keputusan atau *decision tree* terdiri dari banyak himpunan *IF...THEN*. Setiap *path* yang ada didalam tree kemudian disatukan dengan aturan-aturan, yang dimana premis tersebut terdiri atas beberapa kumpulan node-node dan kemudian disimpulkan dari beberapa aturan atas kelas-kelas yang dihubungkan ke *leaf node* dari path tersebut.[4]

2.3.2 Struktur Pohon Keputusan (*Decision Tree*)

Struktur pohon keputusan merupakan struktur yang berbentuk pohon, umumnya bentuk dari sebuah pohon terdapat 3 bagian yaitu akar (*root*), cabang dan daun (*leaf*). Pada struktur pohon keputusan juga terdiri dari tiga bagian yaitu :

1. *Root node*

Root node merupakan node yang letaknya berada pada bagian paling teratas dari suatu pohon.

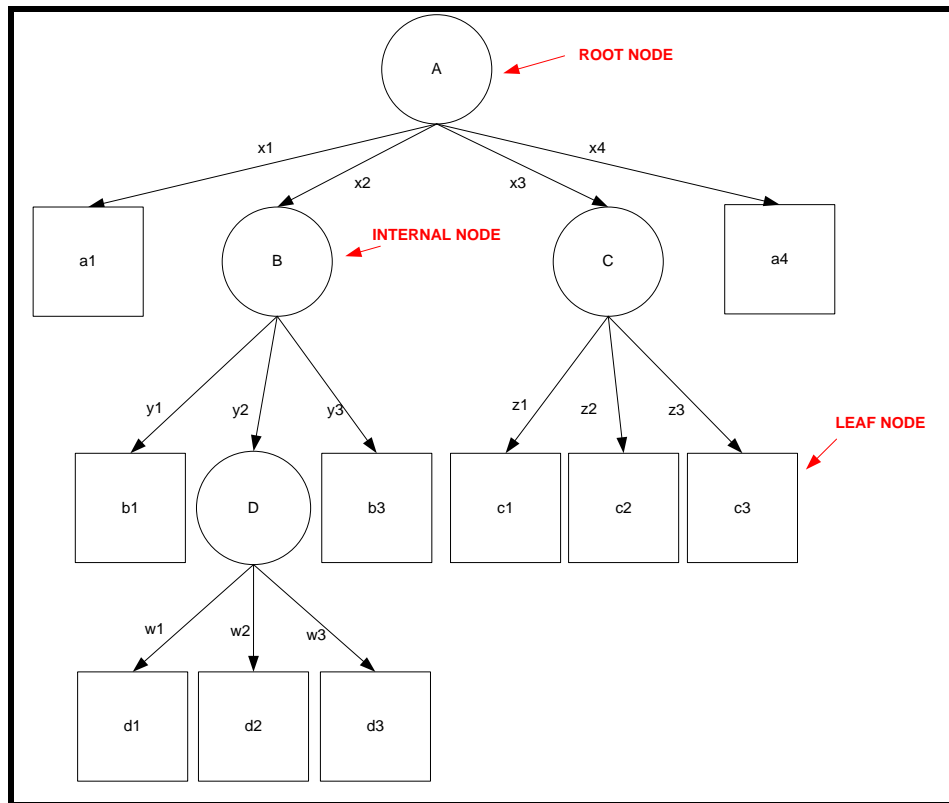
2. *Internal node*

Internal Node merupakan *node* pada percabangan, Pada bagian node ini biasanya terdapat satu bagian input dan memiliki minimal 2 output\

3. *Leaf node*

Node merupakan *node* pada bagian akhir, pada bagian ini hanya memiliki satu input saja dan tidak memiliki output. Pada bagian leaf node pada pohon keputusan setiap *leaf node* dapat menandai label kelas.

Pada pohon keputusan pada seluruh bagian percabangan dapat menyatakan sebuah kondisi yang harus dipenuhi dan setiap ujung-ujung pohon dapat menjadi alat untuk menyatakan nilai kelas data. Gambar dibawah adalah bentuk struktur pohon keputusan.



Gambar 2.2 Struktur Pohon Keputusan

Bentuk bulat pada sebuah pohon keputusan dapat mengartikan sebagai node akar (*root node*) dan juga merupakan node cabang (*internal node*). Namun pada node akar akan selalu berada pada bagian paling atas tanpa memiliki input, sedangkan node cabang memiliki input. Bentuk kotak pada pohon keputusan mengartikan node daun (*leaf node*). Setiap bagian node daun berisi nilai-nilai atribut dari node-node cabang atau node-node akar.[4]

2.4 Algoritma C4.5

Algoritma C4.5 adalah algoritma yang termasuk dalam data mining tipe klasifikasi data dengan menggunakan teknik pohon keputusan sebagai alat untuk pengambilan keputusan. Pohon keputusan merupakan teknik yang terkenal dan disukai karena memiliki kelebihan-kelebihan khusus. Kelebihan tersebut seperti contoh dapat melakukan pengolahan data yang berbentuk numerik (kontinyu) dan diskret, dapat menjadi solusi jika terdapat nilai atribut yang hilang, aturan-aturan yang dihasilkan oleh pohon keputusan dapat diinterpretasikan dan merupakan

algoritma tercepat di antara algoritma-algoritma di data mining yang menggunakan memori utama di komputer.

Algoritma C4.5 membangun pohon keputusan dari data-data pelatihan yang telah di buat, yang berupa kasus-kasus atau *record* dalam basis data. Setiap atribut-atribut yang terdapat pada pohon keputusan dapat berisi data-data yang berupa diskret atau kontinyu (numerik). C4.5 juga dapat menyelesaikan permasalahan yang permasalahan tersebut tidak memiliki unsur nilai untuk sebuah atau lebih atribut. Namun, atribut kelas tersebut hanya dapat bertipe diskret dan tidak boleh kosong. Terdapat tiga prinsip kerja dari algoritma C4.5 yaitu sebagai berikut :

1. Pembuatan Pohon Keputusan

Obyek dari algoritma yang menghasilkan sebuah pohon keputusan adalah membangun struktur-struktur data pohon (dinamakan pohon keputusan) yang akan digunakan untuk melakukan prediksi kelas dari sebuah permasalahan atau *record-record* baru yang belum memiliki kelas. Algoritma ini memilih pemecahan permasalahan yang terbaik dengan melakukan penghitungan dan melakukan perbandingan *gain ratio*, dan *node-node* akan terbentuk pada level berikutnya, dilakukan terus-menerus hingga membentuk daun-daun.

2. Pemangkasan Pohon Keputusan dan melakukan Evaluasi (Opsional)

Dikarenakan pohon keputusan yang terbentuk dapat berukuran besar dan sulit untuk dibaca, Algoritma C4.5 dapat melakukan penyederhanaan pohon dengan melakukan pemangkasan-pemangkasandengan berdasarkan nilai tingkat kepercayaan (*confidence level*). Selain untuk melakukan pengurangan ukuran pohon, pemangkasan berfungsi juga untuk mengurangi kesalahan prediksi pada record baru.

3. Pembuatan Aturan-Aturan dari Pohon Keputusan (Opsional)

Aturan-aturan yang terdapat pada pohon keputusan yang berbentuk *if-then* merupakan aturan yang diturunkan dari pohon keputusan dengan melakukan penelusuran dari akar hingga ke daun. Setiap node yang terbentuk dan persyaratandari pencabangannya akan diberikan pada *if*, sedangkan untuk nilai pada daun pohon keputusan akan dituliskan dengan *then*.

Setelah semua aturan-aturan telah dibuat, maka akan dilakukan penyederhanaan aturan (digabung atau diperumum)[6].

2.4.1 Langkah membangun pohon keputusan menggunakan C4.5

Adapun langkah-langkah dalam membangun dan membentuk pohon keputusan dengan menggunakan data mining algoritma C4.5 adalah[6]:

1. Pohon dimulai dengan sebuah simpul yang merepresentasikan sampel data pelatihan yaitu dengan membuat simpul akar.
2. Jika semua sampel berada dalam kelas yang sama, maka simpul ini menjadi daun dan dilabeli menjadi kelas. Jika tidak, *gain ratio* akan digunakan untuk memilih atribut split, yaitu atribut yang terbaik dalam memisahkan data sampel menjadi kelas-kelas individu.
3. Cabang akan dibuat untuk setiap nilai pada atribut dan data sampel akan dipartisi lagi.
4. Algoritma ini menggunakan proses rekursif untuk membentuk pohon keputusan pada setiap data partisi. Jika sebuah atribut sudah digunakan di sebuah simpul, maka atribut ini tidak akan digunakan lagi di simpul anak-anaknya.
5. Proses ini berhenti jika dicapai kondisi seperti berikut :
 - a. Semua sampel pada simpul berada di dalam satu kelas.
 - b. Tidak ada atribut lainnya yang dapat digunakan untuk mempartisi sampel lebih lanjut. Dalam hal ini akan diterapkan suara terbanyak. Ini berarti mengubah sebuah simpul menjadi daun dan melabelinya dengan kelas pada suara terbanyak.

2.4.2 Entropy

Entropi dan informasi memiliki hubungan untuk menyediakan perilaku-perilaku jangka panjang dari pemrosesan secara acak dan sangat berguna dalam menganalisis data. Perilaku yang terdapat dalam pemrosesan secara acak juga merupakan faktor utama dalam mengembangkan sebuah pengkodean untuk memberikan sebuah teori informasi. Entropi merupakan bagian pengukuran

ketidakpastian rata-rata yang melihat dari kumpulan data ketika tidak mengetahui hasil dari sumber informasi. Hal tersebut berarti bahwa seberapa besar pengukuran –pengukuran informasi yang kita tidak dimiliki. Hal tersebut juga dapat menunjukkan jumlah dari rata-rata informasi yang akan diterima dari hasil pengolahan sumber informasi. Dalam memperoleh nilai *gain ratio* untuk pembentukan pohon keputusan, diperlukan menghitung terlebih dahulu nilai-nilai informasi dalam satuan bits dari suatu kumpulan-kumpulan objek Rumus untuk mendapatkan entropi dengan menggunakan rumus sebagai berikut :

$$Entropy(X) = \sum_{j=1}^k p_j * \log_2 \frac{1}{p_j} = - \sum_{j=1}^k p_j * \log_2 p_j \quad (2.1)$$

dimana,

X : Himpunan Kasus

k : jumlah partisi X

p_j : Proporsi X_j terhadap X

Entropi split yang akan membagi X dengan n record menjadi himpunan-himpunan X_1 dengan n_1 baris dan X_2 dengan n_2 baris adalah :

$$EE(X_1, X_2) = \frac{n_1}{n} E(X_1) + \frac{n_2}{n} E(X_2) \quad (2.2)$$

Besar nilai *Entropy(X)* dapat untuk menunjukkan bahwa X adalah atribut yang lebih acak. atribut yang lebih kecil dari nilai *Entropy(X)* menggambarkan atribut pada bagian ini merupakan sedikit lebih acak yang signifikan untuk data mining. Nilai entropi mencapai nilai minimum 0, ketika semua p_j lain = 0 atau pada kelas

yang sama. Nilainya mencapai maksimum $\log_2 k$, ketika semua nilai p_j adalah sama dengan $1/k$ [2]

2.4.3 Gain

Pada pembentukan pohon keputusan algoritma C4.5, Gain didasarkan pada penurunan entropi setelah dataset yang kemudian dibagi pada setiap atribut[7], atribut dengan nilai *gain ratio* tertinggi akan dipilih sebagai atribut split untuk menjadi simpul. Rumus dari gain ratio adalah sebagai berikut [6]

$$gainratio(a) = \frac{gain(a)}{split(a)} \quad (2.3)$$

Dimana $gain(a)$ adalah *information gain* dari atribut a untuk himpunan sampel X dan $split\ info(a)$ menyatakan entropi atau informasi potensial yang didapat pada pembagian X menjadi n sub himpunan berdasarkan telaahan pada atribut a . Sedangkan $gain(a)$ didefinisikan sebagai berikut :

$$gain(a) = info(X) - info_a(X) \quad (2.4)$$

Untuk rumus $split\ info(a)$ adalah sebagai berikut :

$$gain(a) = - \sum_{j=1}^k \frac{|X_j|}{|X|} * \log_2 \left(\frac{|X_j|}{|X|} \right) \quad (2.5)$$

Dimana X_i menyatakan sub himpunan ke- i pada sampel X .

Dengan kata lain rumus untuk menghitung nilai gain ratio untuk dipilih sebagai atribut dari simpul yang ada sebagai berikut ini :

$$Gainratio(a) = Entropy(X) - \sum_{j=1}^k \frac{|X_j|}{|X|} * Entropy(X_j) \quad (2.6)$$

Keterangan

a :Atribut

k :Jumlah partisi atribut a

X :Jumlah kasus

X_j :Himpunan ke- j pada sampel X

2.4.4 Contoh Kasus C4.5

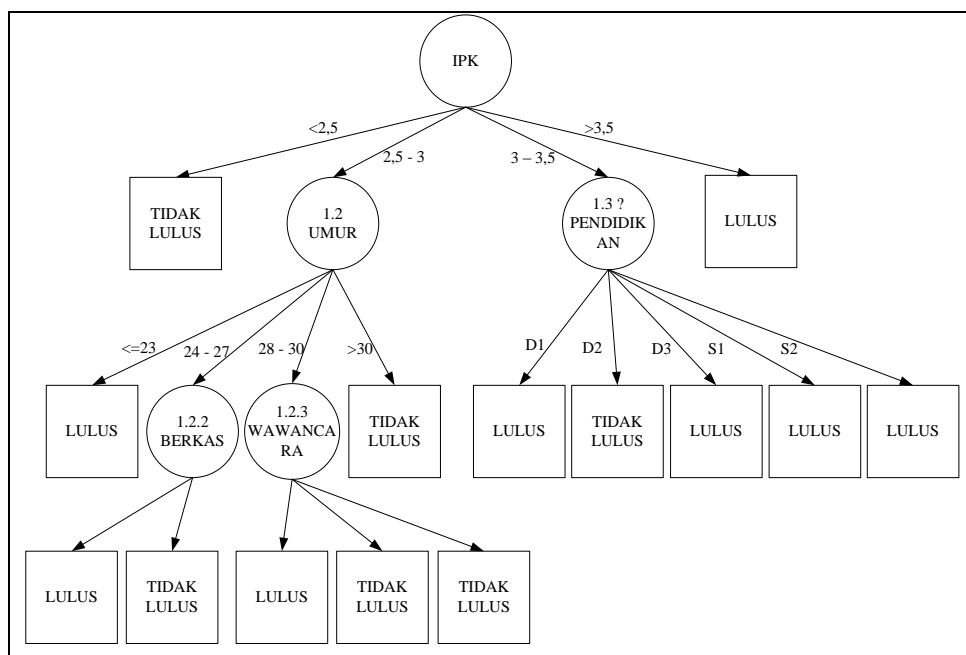
Contoh permasalahan yang dapat diselesaikan dengan menggunakan data mining algoritma C4.5 sebagai berikut:

Tabel 2.2 Contoh Kasus Algoritma C4.5

no	Pendidikan	Jurusan	Wawancara	Kelengkapan	Umur	Hasil
1	S1	AKUNTANSI	BAIK	LENGKAP	27	LULUS
2	S1	AKUNTANSI	CUKUP	TDK LENGKAP	28	TIDAK LULUS
3	S1	AKUNTANSI	CUKUP	LENGKAP	32	TIDAK LULUS
4	S2	INFORMATIKA	BAIK	LENGKAP	29	LULUS
5	S1	INFORMATIKA	BAIK	LENGKAP	26	LULUS
6	S2	INFORMATIKA	BAIK	TDK LENGKAP	28	LULUS
7	D3	INFORMATIKA	CUKUP	TDK LENGKAP	27	TIDAK LULUS
8	D3	AKUNTANSI	CUKUP	TDK LENGKAP	22	LULUS
9	S2	MANAJEMEN	CUKUP	LENGKAP	29	LULUS
10	S2	MANAJEMEN	BURUK	LENGKAP	29	TIDAK LULUS
11	S1	AKUNTANSI	BAIK	LENGKAP	27	LULUS
12	D2	MANAJEMEN	BAIK	LENGKAP	33	TIDAK LULUS
13	D2	MANAJEMEN	BURUK	TDK LENGKAP	22	LULUS
14	S1	AKUNTANSI	BAIK	TDK LENGKAP	31	TIDAK

						LULUS
15	S1	AKUNTANSI	BURUK	TDK LENGKAP	21	TIDAK LULUS
16	S1	AKUNTANSI	BAIK	LENGKAP	25	LULUS
17	S2	MATEMATIKA	CUKUP	LENGKAP	25	LULUS

Kemudian dari permasalahan diatas akan membentuk pohon keputusan (Decision tree) sebagai berikut :



Gambar 2.3 Hasil Akhir Pohon Keputusan

2.5 Confusion Matrix

Confusion matrix merupakan metode pengujian yang digunakan untuk mengukur kinerja suatu model klasifikasi. Pengujian confusion matrix dilakukan dengan menggunakan data testing yang dicocokkan dengan hasil dari klasifikasi data training maka akan mendapatkan akurasi, recall, dan Precision. *Confusion matrix* menggunakan table yang berisi benar atau tidak benar dari hasil data testing yang telah dicocokkan dengan klasifikasi data training. Contoh dari tabel *confusion matrix* dapat dilihat dari tabel 1 *confusion matrix*.

Tabel 2.3 Confusion Matrix

		PREDICTED CLASS	
		CLASS = 1	CLASS = 0
ACTUAL CLASS	CLASS = 1	F11	F10
	CLASS = 0	F01	F00

Dimana penjelasannya sebagai berikut :

F11 = Jika hasil prediksi positif dan data sebenarnya positif

F10 = Jika hasil prediksi negatif dan sedangkan nilai sebenarnya positif

F01 = Jika hasil prediksi positif dan sedangkan nilai sebenarnya negatif

F00 = Jika hasil prediksi negatif dan data sebenarnya negatif

$$Akurasi = \frac{F11 + F10}{F11 + F10 + F01 + F00} \quad (2.7)$$

Akurasi merupakan presentase dari hasil data yang benar. Sedangkan precision merupakan ukuran dari akurasi suatu kelas tertentu yang telah diprediksi. Adapun rumus untuk mencari precision adalah sebagai berikut:

$$Precision = \frac{F11}{F01 + F11} \quad (2.8)$$

Sedangkan recall merupakan persentase dari data dengan nilai positif yang nilai prediksinya juga positif. Adapun rumus untuk mencari recall adalah sebagai berikut:

$$Recall = \frac{F11}{F10 + F11} \quad (2.9)$$

2.6 Program Studi Sistem Informasi

Sistem Informasi merupakan program studi yang mengkombinasikan antara teknologi informasi dan kegiatan menggunakan teknologi tersebut dalam

mendukung kegiatan operation dan manajemen. Sistem Informasi digunakan untuk rujukan kepada interaksi antara orang, proses algoritma, data dan juga teknologi. Pada program studi Sistem Informasi kumpulan informasi akan disimpan pada basis data menggunakan media teknologi dan digunakan dalam pengambilan keputusan disebuah institusi atau organisasi.

Pada sistem informasi terdapat tiga komponen utama yaitu :

1. Input: Proses memasukkan data yang nantinya akan diolah dan dijadikan sebagai informasi yang nantinya akan disajikan untuk masyarakat.
2. Proses: Suatu kegiatan pengolahan atau pemorsesan data yang telah dimasukkan untuk dapat menghasilkan sebuah informasi yang berguna
3. Output: Informasi-informasi yang telah diproses akan berbentuk data-data yang nantinya akan dimanfaatkan oleh masyarakat luas.

2.6.1 Jalur Pendaftaran Calon Mahasiswa

Terdapat 3 jalur pendaftaran yang dapat dilakukan oleh calon mahasiswa untuk mendaftar pada Universitas Dian Nuswantoro yaitu: [8]

1. Jalur PMDK : Jalur untuk penerimaan mahasiswa baru dengan tanpa tes untuk mendaftar pada seluruh program studi yang berstrata satu (S-1) dan seluruh Program Diploma Empat (D-IV). Penyeleksian dapat melihat dari prestasi akademik yang semasa Sekolah Menengah Atas dan non akademik siswa dapat memperlihatkan dengan nilai hasil studi semasa sekolah menengah atas dan piagam penghargaan kejuaraan. Terdapat 2 cara untuk mendaftarn jalur PMDK yaitu secara manual dan online [8]
2. Jalur Reguler: jalur untuk penerimaan mahasiswa baru untuk seluruh jenjang strata satu (S-1), Program Vokasi (D3), dan Diploma 4 (D4) dengan melalui serangkaian tes akademik yang dilakukan dengan mandiri atau kolektif. Pendaftaran untuk Jalur Reguler dapat dilakukan secara langsung pada kesekretariatan Pendaftaran mahasiswa baru Universitas Dian Nuswantoro atau dapat juga dilakukan secara online [8]

3. jalur Penyetaraan/UMBPTS: jalur untuk melanjutkan kejenjang yang lebih tinggi untuk lulusan program studi vokasi atau pindah studi baik dari universitas Dian Nuswantoro atau luar Universitas Dian Nuswantoro.

2.7 Hubungan Data Mining dengan Sistem Informasi dan Universitas

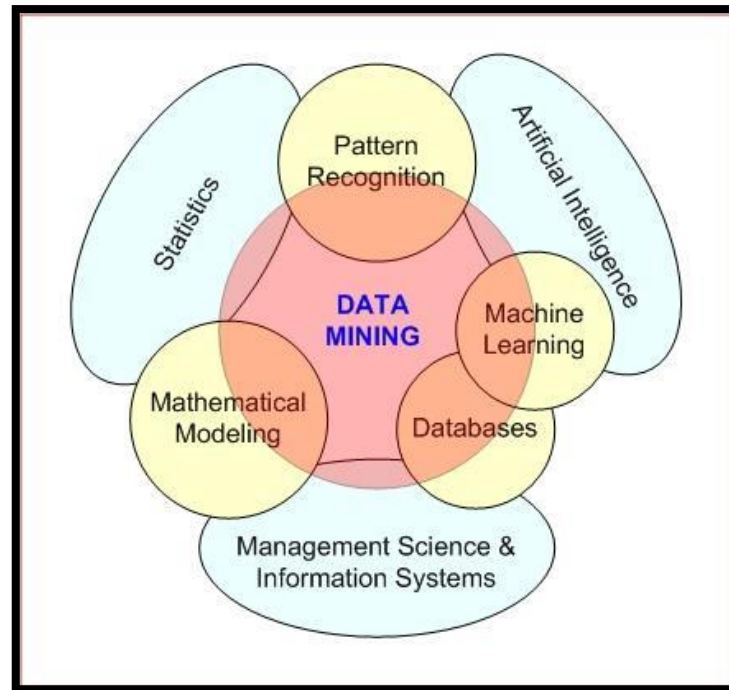
2.7.1 Data mining dengan sistem informasi

Hubungan antara data mining dengan program studi Sistem Informasi data mining digunakan sebagai alat untuk membantu dan mempermudah dalam mengelola data dan menggali atau mengekstrak data dengan berbagai metode yang telah dikomputasikan pada sebuah institusi atau perusahaan, nantinya data-data pada perusahaan atau institusi tersebut akan diolah dalam data mining dengan menghasilkan sebuah pola-pola yang berhubungan untuk menghasilkan informasi yang dibutuhkan.

Sistem informasi digunakan untuk rujukan kepada interaksi antara orang, proses algoritma, teknologi. pada program studi sistem informasi kumpulan informasi akan disimpan pada basis data menggunakan media teknologi dan digunakan dalam pengambilan keputusan disebuah institusi atau organisasi. Nantinya menghasilkan sebuah data, dari data tersebut akan menghasilkan sebuah informasi yang berguna. Dalam pengertian sistem informasi tersebut peran data mining dapat membantu dalam mempercepat proses pengolahan data-data yang terdapat pada sistem informasi perusahaan. Set data yang terdapat pada sistem informasi perusahaan memiliki jumlah yang besar dan luas dapat dilakukan pengolahan data menggunakan data mining agar data-data tersebut dapat menghasilkan informasi-informasi yang berguna dalam pengambilan keputusan institusi maupun perusahaan.

Data mining merupakan salah satu mata kuliah yang mempelajari mengenai pencarian informasi dari kumpulan data dalam jumlah besar dengan menggunakan algoritma tertentu serta telah lama diterapkan pada program studi teknologi informasi Universitas Dian Nuswantoro. Beberapa tahun belakangan program studi sistem informasi di Universitas Dian Nuswantoro mulai menerapkan data mining.

Data mining mencakup beberapa ilmu yang dapat mempergunakan data mining pada proses pembelajarannya. Secara garis besar cakupan beberapa ilmu tersebut yaitu:



Gambar 2.4 Cakupan ilmu pada data mining

Pada gambar 2.4 dijelaskan bahwa sistem informasi masuk dalam cakupan data mining disimpulkan bahwa data mining dapat bekerja dalam ilmu tersebut. Perbedaan antara data mining pada teknologi informasi dan data mining pada sistem informasi sebenarnya memiliki tujuan yang sama, Tetapi yang membedakan pada hasil dan penerapannya. Pada program studi sistem informasi data mining mengarah pada analisa hasil yang diperoleh dan dimanfaatkan sebagai alat atau metode pengambilan keputusan. Sedangkan pada program studi teknologi informasi lebih mengarah dan mengedepankan pada pengembangan atau pembuatan sistem perhitungan data mining.

2.7.2 Data Mining dengan Universitas

Sedangkan hubungan dengan universitas yaitu pada sebuah universitas seperti contoh Universitas Dian Nuswantoro di dalamnya terdapat set data dalam

jumlah yang besar. Kemudian dengan proses data mining, data yang berjumlah besar tersebut seperti contoh data calon mahasiswa, data KRS dan sebagainya kemudian akan diolah dan dimanfaatkan untuk menghasilkan sebuah informasi serta dapat juga dipergunakan dalam proses pengambilan keputusan dalam meningkatkan startegi pemasaran.

Pengambilan keputusan pemilihan beberapa tindakan alternatif yang ada untuk mencapai satu atau beberapa tujuan yang telah diterapkan (Turban, 2005). Pada dasarnya pengambilan keputusan merupakan suatu bentuk pemilihan dari berbagai alternatif tindakan yang mungkin dipilih, yang prosesnya melalui mekanisme tertentu dengan harapan akan menghasilkan suatu keputusan yang terbaik[9].

Dalam proses pengambilan keputusan terdapat 4 proses tahapan yaitu intelligence, design , choice dan implementation [9]. Data mining dapat diergunakan dalam pengambilan keputusan pada tahap ke tiga yaitu choice Tahap ini adalah proses menemukan, mengembangkan, dan menganalisis alternatif tindakan yang bisa dilakukan. Tahap ini meliputi proses untuk mengertimasalah, menurunkan solusi, dan menguji kelayakan solusi. Hasil dari penerapan data mining dapat diterapkan pada tahapan ke empat yaitu implementation.

Pada penerapannya data mining diterapkan pada institusi yang berbasis sistem informasi. Indonesia khususnya belum banyak menerapkan data mining sebagai alat untuk menggali sebuah informasi dari kumpulan data yang terdapat pada sebuah institusi maupun sebuah universitas. Data mining belum banyak diterapkan pada universitas tetapi, data mining banyak dipergunakan sebagai bahan pembuatan tugas akhir atau skripsi dan pembelajaran pada mata kuliah.

Hasil yang perlu dipertimbangkan dan dikembangkan dalam penerapan data mining ditingkat universitas yaitu perlu mempertimbangkan dan mengembangkan dari segi penerapan hasil data mining adalah prediksi hal inierat kaitannya dengan sistem pengambilan keputusan. Prediksi perlu dilakukan dan diperhitungkan agar dapat melakukan tindakan pencegahan atau tindakan preventif atau tindakan yang belum dilakukan dengan kata lain yang akan muncul. Sehingga dari prediksi tersebut dapat dikembangkan untuk hasil yang telah dicapai sebelumnya.

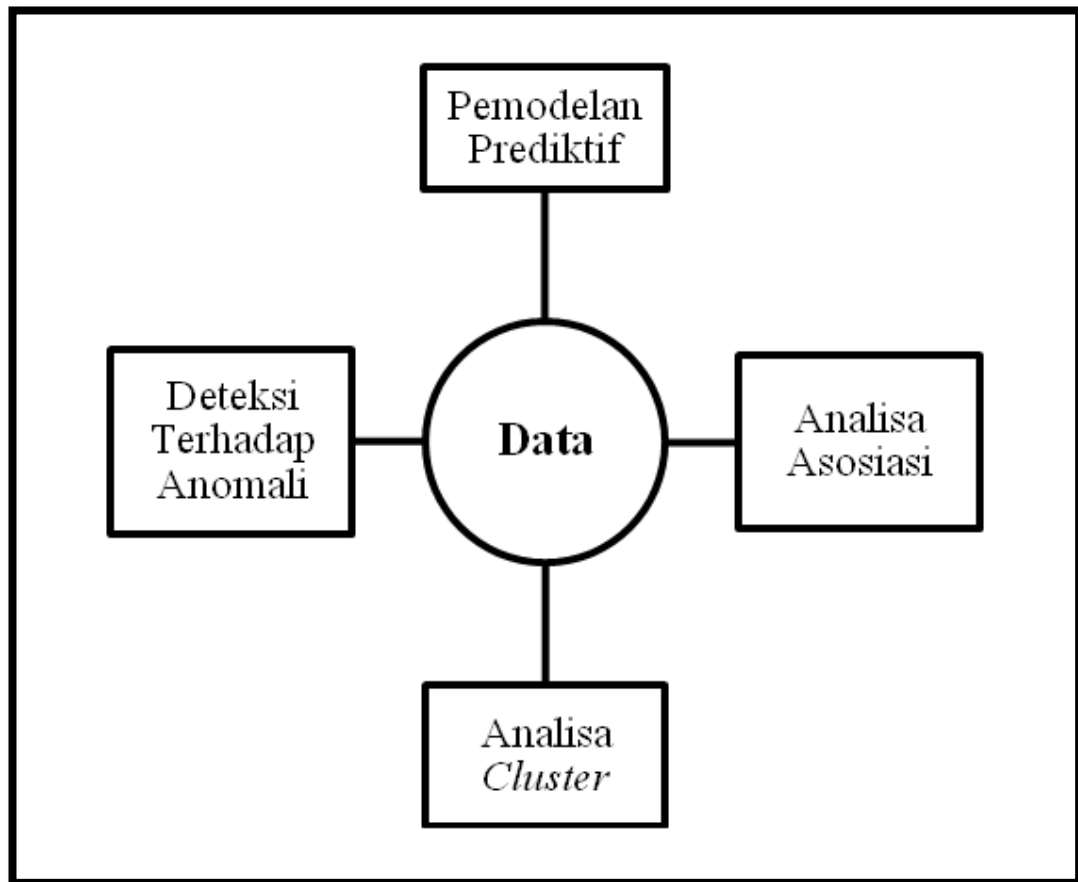
Penerapan tersebut harus diterapkan dengan benar dan menggunakan metode yang tepat dan dijalankan secara penuh akan menjadi efektivitas pada program studi Sistem Informasi Universitas Dian Nuswantoro yaitu dalam mengurangi pengeluaran sumber daya karena promosi akan difokuskan pada salah satu daerah tertentu dan meningkatkan jumlah mahasiswa dengan cara meningkatkan promosi. Sedangkan efisiensi penggunaan data mining pada program studi Sistem Informasi Universitas Dian Nuswantoro yaitu mempercepat dalam proses pengambilan keputusan.

2.7.3 Penerapan Data Mining

Perusahaan atau institusi di luar Indonesia telah banyak menerapkan data mining pada proses perkembangan bisnisnya. Salah satu perusahaan asing yang telah menerapkannya yaitu perusahaan IBM dan VISA.

Pada perusahaan IBM menggunakan algoritma data mining untuk mendata akses halaman website khususnya yang berkaitan dengan pemasaran yang berguna untuk melihat perilaku dan minat customer serta melihat keefektifan pemasaran melalui website. Sedangkan pada VISA menggunakan data mining untuk melihat kesuksesan dari program – program customer loyalty yang terdapat pada perusahaan tersebut. Data mining tidak hanya cocok digunakan oleh industri-industri yang mengelola informasi secara intensif seperti keuangan dan perbankan, tetapi juga perusahaan apa saja yang ingin memanfaatkan data untuk manajemen pelanggan dengan lebih baik[10].

Secara garis besar pengolahan data mining dibagi menjadi dua kategori utama, yaitu tugas deskriptif dan tugas prediktif [9]. Pada intinya tugas data mining yaitu melakukan analisa asosiasi, pemodelan prediktif, analisa cluster dan prediksi anomali. Berikut gambar tugas data mining :



Gambar 2.5 Tugas Data mining pada sistem institusi atau perusahaan

Pemodelan prediktif mengacu pada proses membentuk model untuk menentukan variabel target sebagai fungsi dari variabel penjelas. Ada dua tipe dari pemodelan prediktif, yaitu klasifikasi (*classification*) yang digunakan untuk variabel target yang diskret, dan regresi (*regression*) yang digunakan untuk variabel target yang kontinyu. Analisa asosiasi digunakan untuk menemukan pola yang mendeskripsikan fitur-fitur data yang saling berhubungan. Pola-pola ini biasanya digambarkan dalam bentuk aturan implikasi. Analisa *cluster* merupakan proses untuk mencari kelompok-kelompok data, sedemikian sehingga data yang berada dalam satu kelompok memiliki kemiripan dibandingkan data yang terletak pada kelompok lain. Deteksi anomaly merupakan proses identifikasi data yang

memiliki perbedaan karakteristik yang signifikan dengan data yang lain atau yang dikenal dengan istilah *outlier* [9].

2.8 WEKA 3.8

WEKA merupakan salah satu aplikasi data mining yang berbasis java dan open source atau dapat digunakan secara gratis. WEKA merupakan paket learning machine yang praktis. Nama WEKA berasal dari Waikato Environment Knowledge Analysis. aplikasi berbasis java ini dapat dengan mudah menyelesaikan permasalahan data mining secara cepat khususnya klasifikasi yang merupakan dasar-dasar pendekatan data mining[11]. WEKA memiliki *tools* yaitu *pre processing data*, klasifikasi, regresi, lustering, asosiasi dan visualisasi. Pada WEKA terdapat 2 format yang dapat digunakan yaitu arff dan cmv.C4.5 pada aplikasi ini dinamakan J48. C4.5 dan J48 merupakan satu algoritma yang sama tetapi hanya berbeda nama. Terdapat komponen-komponen pada klasifikasi WEKA yaitu: [12]

1. *Use training set*

Merupakan fitur sebagai alat untuk pengetesan data mining klasifikasi dengan menggunakan data training itu sendiri.

2. *Supplied test set*

Merupakan fitur yang digunakan untuk pengetesan dengan menggunakan data lain . fitur inilah yang digunakan untuk melakukan prediksi terhadap data set.

3. *Cross Validation*

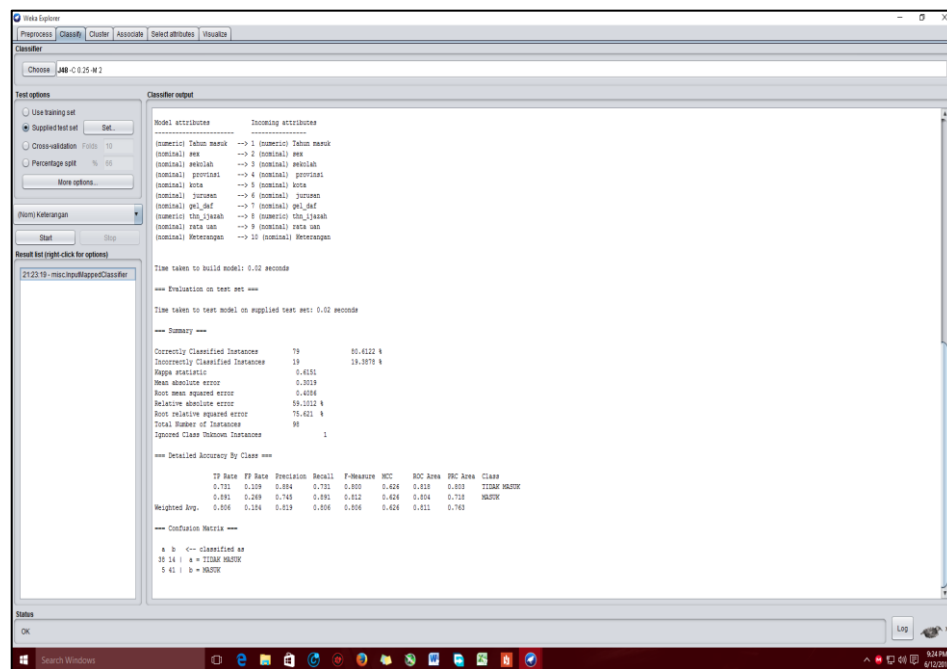
Fitur cross validation membagi berapa *fold* yang akan digunakan. Terdapat nilai default yaitu 10.

4. Percentage Split

Fitur yang digunakan untuk pengujian data dengan menggunakan k% dari data set.k merupakan masukan dari user itu sendiri.

5. Start

Merupakan tombol yang digunakan untuk memulai proses pengujian setelah memilih fitur yang ada.



Gambar 2.6 Tampilan Proses program WEKA 3.8

BAB 3

METODE PENELITIAN

3.1 Metode Pengumpulan Data

Di dalam sebuah kegiatan penelitian dibutuhkan pengumpulan data-data dan metode pengumpulan data yang memiliki perananan penting untuk memberikan keakuratan dan kualitas dari data yang akan digunakan didalam proses penelitian. Terdapat metode – metode yang dapat digunakan untuk mendukung peneletian adalah:

3.1.1 Survey

Survey merupakan sebuah kegiatan dengan melakukan pengamatan secara langsung pada sebuah perusahaan atau universitas untuk memperoleh data-data yang dapat digunakan didalam penyusunan laporan tugas akhir, seperti data calon mahasiswa baru pada program studi sistem informasi Universitas Dian Nuswantoro.

3.1.2 Studi Pustaka

Metode yang dapat digunakan oleh peneliti didalam mencari sebuah informasi yang terkait dengan tema penelitian dan metode dari jurnal peneitian yang memiliki kemiripan tema dengan melakukan studi keperpustakaan dan mencari jurnal penelitian pada media digital yang berhubungan dengan algoritma data mining. Dapat juga pada buku-buku yang berhubungan langsung pada algoritma yang digunakan oleh peneliti.

3.2 Sumber Data

Sumber-sumber data penelitian yang digunakan pada penelitian dapat yang diperoleh dan dikelompokan menjadi 2, yaitu:

3.2.1 Data Primer

Data primer merupakan data yang didapatkan dengan melakukan sebuah pengamatan secara langsung terkait dengan penelitian seperti melakukan survey dan wawancara, dimana hasil dari kegiatan survey dan wawancara tersebut dapat digunakan untuk bahan penelitian didalam penerapan *Classification Rules* yang berguna untuk memprediksi calon mahasiswa yang dapat dilihat pada pembuatan pohon keputusan.

3.2.2 Data Sekunder

Data sekunder merupakan data yang didapatkan dalam bentuk sebuah informasi seperti pada jurnal, literatur, buku dan juga media digital; seperti internet yang berkaitan langsung dengan metode dan tema penelitian yang disusun oleh penulis.

3.3 Metode yang Diusulkan

Metode yang diusulkan dalam penelitian ini yaitu menggunakan metode data mining algoritma C.45 dalam memprediksi mahasiswa pada program studi Sistem Informasi dengan menggunakan pohon keputusan pada data calon mahasiswa baru pada program studi Sistem Informasi Universitas Dian Nuswantoro.

3.4 Langkah Implementasi pada proses KDD

Langkah untuk menerapkan proses Knowledge Discovery in Databases (KDD) sebagai berikut:

1. Pengumpulan Data

Data pada penelitian tugas akhir ini bersumber pada data calon mahasiswa pada program studi Sistem Informasi tahun 2013-2015 Universitas Dian Nuswantoro.

2. Seleksi Data

Diambil beberapa field data dari sekumpulan field yang terdapat pada pendaftaran mahasiswa baru sehingga penyeleksian data akan mengambil data yang dibutuhkan.

3. *Preprocessing Data/Cleaning Data*

Preprocessing data atau cleaning data melakukan proses pembersihan atau cleaning dengan memilih beberapa atribut pada data calon mahasiswa yang mendaftar pada Fakultas Ilmu Komputer yang menjadi fokus penelitian berupa atribut jurusan ketika SMA, alamat sekolah, nilai ujian nasional, pilihan program studi dan atribut lain seperti agama, nomor induk mahasiswa yang tidak digunakan akan dihapus.

4. Transformasi Data

Tahap pada transformasi data dalam pemrosesan data mining pada algoritma C4.5 sangat diperlukan karena dalam kegiatan proses data mining yang akan dikomputerisasikan akan memerlukan bentuk – bentuk data yang dapat untuk diintegrasikan dengan aplikasi/tools yang dapat digunakan seperti WEKA atau SPMF.

5. Data Mining

Pada tahapan data mining ini memusatkan agar mendapatkan banyak pola dari hasil ekstraksi data calon mahasiswa yang mendaftar yang telah ditransformasikan dengan menggunakan algoritma C4.5. Saat proses data mining dengan membentuk gain dan entropy yang selanjutnya akan membentuk pohon keputusan.

6. Interpretasi dan Evaluasi

Banyak pola yang telah dibentuk pada pohon keputusan yang telah pohon tersebut akan diartikan dan disajikan dengan bentuk yang lebih sederhana dan mudah untuk dipahami dalam membantu proses pengambilan keputusan.

BAB 4

HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Dalam tugas akhir ini sumber data yang digunakan berasal dari data calon mahasiswa baru pada rentang 3 tahun terakhir dimulai pada tahun 2013 hingga tahun 2015 pada program studi Sistem informasi Universitas Dian Nuswantoro. Data yang diperoleh dari PSI Universitas Dian Nuswantoro yang kemudian akan diolah dengan menggunakan proses KDD. Untuk mendapatkan pohon keputusan yang nantinya berguna untuk program studi Sistem Informasi.

4.2 Penyeleksian Data

Proses pertama dalam proses data mining dengan melakukan penyeleksian data dari data calon mahasiswa pada tahun 2013-2015 jurusan sistem informasi. Pada keseluruhan attribute yang diperoleh berjumlah 14 attribute yaitu nomor pendaftaran, tahun masuk, NIM, jenis kelamin(Sex), nama sekolah, nama provinsi, nama kota, jurusan SMA, gelombang pendaftaran, tahun ijazah, jumlah pelajaran uan, jumlah nilai uan, rata-rata ujian nasional dan keterangan. Kemudian variable atau attribute yang dipakai dalam proses data mining akan diambil 10 attribute dari 14 attribute yang diperoleh yaitu tahun masuk, jenis kelamin (Sex), nama sekolah, nama provinsi, nama kota, jurusan SMA (jurusan), gelombang pendaftaran (gel_daf), tahun ijasah (thn_ijazah) ,jumlah nilai ujian nasional dan keterangan . Yang terlihat pada table dibawah ini:

Tahun masuk	sex	nama sekolah	nama prov	nama kota	jurusan	gel_daf	thn_ijazah	jumlah nilai uan	Keterangan
2015	LAKI-LAKI	SMA Kristen YSKI	JAWA TENGAH	SEMARANG	IPS	PMDKI	2015	45	MASUK
2015	LAKI-LAKI	SMK NEGERI 7 SEMARANG	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2014	40.3	TIDAK MASUK
2015	LAKI-LAKI	SMK TELEKOMUNIKASI TUNAS HARAPAN	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2015	47.2	TIDAK MASUK
2015	LAKI-LAKI	SMA Kebon Dalem	JAWA TENGAH	SEMARANG	IPA	PMDKI	2015	37.4	TIDAK MASUK
2015	PEREMPUAN	SMA Islam Hidayatullah	JAWA TENGAH	SEMARANG	IPA	PMDKI	2015	45.3	TIDAK MASUK
2015	PEREMPUAN	SMA NEGERI 6	JAWA TENGAH	SEMARANG	IPA	PMDKI	2015	38.5	TIDAK MASUK
2015	PEREMPUAN	SMA Negeri 1 Jeparo	JAWA TENGAH	JEPARA	IPS	PMDKI	2015	35.75	MASUK
2015	LAKI-LAKI	MADRASAH ALIYAH	JAWA TENGAH	SEMARANG	IPA	PMDKI	2015	42	TIDAK MASUK
2015	LAKI-LAKI	SMA NEGERI 3 PATI	JAWA TENGAH	PATI	IPA	PMDKI	2015	44.6	TIDAK MASUK
2015	LAKI-LAKI	SMA Kristen Tri Tunggal	JAWA TENGAH	SEMARANG	IPS	PMDKI	2015	36.4	MASUK
2015	LAKI-LAKI	SMA NEGERI 1 KENDAL	JAWA TENGAH	KENDAL	IPS	PMDKI	2015	48	MASUK
2015	PEREMPUAN	SMA NASIMA SEMARANG	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2015	44.95	TIDAK MASUK
2015	PEREMPUAN	SMA NEGERI 8	JAWA TENGAH	SEMARANG	IPA	PMDKI	2015	44.4	TIDAK MASUK
2015	LAKI-LAKI	SMK TELEKOMUNIKASI TUNAS HARAPAN	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2015	50	TIDAK MASUK
2015	LAKI-LAKI	SMK TELEKOMUNIKASI TUNAS HARAPAN	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2015	43.9	TIDAK MASUK
2015	LAKI-LAKI	SMK TELEKOMUNIKASI TUNAS HARAPAN	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2015	32.85	TIDAK MASUK
2015	LAKI-LAKI	/N	JAWA TENGAH	SEMARANG	IPS	PMDKII	2015	0:00	TIDAK MASUK
2015	PEREMPUAN	SMA Islam Sultan Agung 1	JAWA TENGAH	SEMARANG	IPA	PMDKII	2015	0:00	TIDAK MASUK
2015	LAKI-LAKI	SMA Theresiana 1	/N	SEMARANG	IPA	PMDKII	2015	0:00	MASUK
2015	PEREMPUAN	SMK PERDANA	/N	SEMARANG	Multimedia	PMDKII	2015	0:00	TIDAK MASUK
2015	LAKI-LAKI	SMK YAYASAN PHARMASI	/N	SEMARANG	Farmasi	PMDKI	2015	31.5	TIDAK MASUK
2015	LAKI-LAKI	SMA PGRI 2	/N	SEMARANG	-	PMDKII	2015	0:00	MASUK
2015	LAKI-LAKI	SMK NEGERI 8 SEMARANG	/N	SEMARANG	Teknik Komputer & Jaringan	PMDKII	2015	31.5	MASUK
2015	PEREMPUAN	SMA NEGERI 2 KUDUS	/N	KUDUS	IPA	PMDKI	2015	49.67	TIDAK MASUK
2015	LAKI-LAKI	SMA NEGERI 1 GUBUG	/N	SEMARANG	IPA	PMDKII	2015	0:00	MASUK
2015	LAKI-LAKI	/N	/N	SEMARANG	IPA	PMDKII	2015	0:00	MASUK
2015	PEREMPUAN	/N	/N	UNGARAN	-	PMDKII	2015	31.5	MASUK
2015	PEREMPUAN	/N	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2015	47.5	TIDAK MASUK
2015	LAKI-LAKI	/N	JAWA TENGAH	UNGARAN	IPA	PMDKII	2015	0:00	MASUK
2015	LAKI-LAKI	SMA NEGERI 1 UNGARAN	JAWA TENGAH	UNGARAN	IPA	PMDKII	2015	0:00	MASUK

Gambar 4.1 Data Sample Dengan 10 Atribute

4.3 Preprocessing/ Cleaning

Data-data yang telah dikumpulkan dari data calon mahasiswa program studi Sistem Informasi dari tahun 2013-2015 dilakukan pembersihan data dimana data yang kosong pada salah satu attribute atau beberapa attribute akan dibuang dan tidak digunakan. Data yang digunakan adalah data yang pada 10 attribute dari attribute dari 14 attribute yang diperoleh yaitu tahun masuk, jenis kelamin (Sex), nama sekolah, nama provinsi, nama kota, jurusan SMA (jurusan), gelombang pendaftaran (gel_daf), tahun ijazah (thn_ijazah), jumlah nilai ujian nasional dan variable output yaitu keterangan . Keseluruhan attribute harus terisi dan lengkap akan digunakan untuk proses data mining algoritma C4.5.

Setelah melakukan pembersihan data dengan melakukan pembuangan data dimana data yang tidak penuh pada setiap attribute atau sebagian atribut akan dibuang dan tidak dipakai dalam proses data mining. Data yang digunakan adalah data yang penuh atau terisi pada setiap atributnya seperti gambar dibawah ini setelah melakukan pembersihan data :

Tahun masuk	sex	nama sekolah	nama prov	nama kota	jurusan	gel. daf.	thn. ijaz.	jumlah nilai	uan	Keterangan
2015	LAKI-LAKI	SMA Kristen YSKI	JAWA TENGAH	SEMARANG	IPS	PMDKI	2015	45	MASUK	
2015	LAKI-LAKI	SMK NEGERI 7 SEMARANG	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2014	40.3	TIDAK MASUK	
2015	LAKI-LAKI	SMK TELEKOMUNIKASI TUNAS HARAPAN	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2015	47.2	TIDAK MASUK	
2015	LAKI-LAKI	SMA Kebon Dalem	JAWA TENGAH	SEMARANG	IPA	PMDKI	2015	37.4	TIDAK MASUK	
2015	PEREMPUAN	SMA Islam Hidayatullah	JAWA TENGAH	SEMARANG	IPA	PMDKI	2015	45.3	TIDAK MASUK	
2015	PEREMPUAN	SMA NEGERI 6	JAWA TENGAH	SEMARANG	IPA	PMDKI	2015	38.5	TIDAK MASUK	
2015	PEREMPUAN	SMA Negeri 1 Jepara	JAWA TENGAH	JEPARA	IPS	PMDKI	2015	35.75	MASUK	
2015	LAKI-LAKI	MADRASAH ALYAH	JAWA TENGAH	SEMARANG	IPA	PMDKI	2015	42	TIDAK MASUK	
2015	LAKI-LAKI	SMA NEGERI 3 PATI	JAWA TENGAH	PATI	IPA	PMDKI	2015	44.6	TIDAK MASUK	
2015	LAKI-LAKI	SMA Kristen Tri Tunggal	JAWA TENGAH	SEMARANG	IPS	PMDKI	2015	36.4	MASUK	
2015	LAKI-LAKI	SMA NEGERI 1 KENDAL	JAWA TENGAH	KENDAL	IPS	PMDKI	2015	48	MASUK	
2015	PEREMPUAN	SMA NASIMA SEMARANG	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2015	44.95	TIDAK MASUK	
2015	PEREMPUAN	SMK NUSA PUTERA	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2015	47.5	TIDAK MASUK	
2015	PEREMPUAN	SMA NEGERI 2 KUDUS	JAWA TENGAH	KUDUS	IPA	PMDKI	2015	49.67	TIDAK MASUK	
2015	PEREMPUAN	SMA NEGERI 8	JAWA TENGAH	SEMARANG	IPA	PMDKI	2015	44.4	TIDAK MASUK	
2015	LAKI-LAKI	SMK TELEKOMUNIKASI TUNAS HARAPAN	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2015	50	TIDAK MASUK	
2015	LAKI-LAKI	SMK TELEKOMUNIKASI TUNAS HARAPAN	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2015	43.9	TIDAK MASUK	
2015	LAKI-LAKI	SMK TELEKOMUNIKASI TUNAS HARAPAN	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2015	32.85	TIDAK MASUK	
2015	LAKI-LAKI	SMK YAYASAN PHARMASI	JAWA TENGAH	SEMARANG	Farmasi	PMDKI	2015	31.5	TIDAK MASUK	
2015	PEREMPUAN	SMK NUSA PUTERA	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKI	2015	49.78	TIDAK MASUK	
2015	PEREMPUAN	SMK NEGERI 8 SEMARANG	JAWA TENGAH	SEMARANG	Rekayasa Perangkat Lunak	PMDKI	2015	37	TIDAK MASUK	
2015	LAKI-LAKI	SMA NEGERI 9	JAWA TENGAH	SEMARANG	IPA	PMDKI	2015	41.5	TIDAK MASUK	
2015	LAKI-LAKI	SMA Theresiana 1	JAWA TENGAH	SEMARANG	IPS	PMDKII	2015	51.2	MASUK	
2015	LAKI-LAKI	SMA Theresiana 1	JAWA TENGAH	SEMARANG	IPS	PMDKII	2015	36.8	MASUK	
2015	PEREMPUAN	SMA Theresiana 1	JAWA TENGAH	SEMARANG	IPS	PMDKII	2015	36.86	MASUK	
2015	PEREMPUAN	SMA Theresiana 1	JAWA TENGAH	SEMARANG	IPA	PMDKII	2015	39.6	TIDAK MASUK	
2015	PEREMPUAN	SMA Kesatrian 2	JAWA TENGAH	SEMARANG	IPS	PMDKII	2015	42.7	MASUK	
2015	PEREMPUAN	SMA Islam Sultan Agung 1	JAWA TENGAH	SEMARANG	IPA	PMDKII	2015	48.8	TIDAK MASUK	
2015	LAKI-LAKI	SMA Theresiana 1	JAWA TENGAH	SEMARANG	IPA	PMDKII	2015	38.75	TIDAK MASUK	
2015	PEREMPUAN	SMK PERDANA	JAWA TENGAH	SEMARANG	Multimedia	PMDKII	2015	48.62	TIDAK MASUK	
2015	LAKI-LAKI	SMA Kesatrian 1	JAWA TENGAH	SEMARANG	IPA	PMDKII	2015	31	TIDAK MASUK	
2015	LAKI-LAKI	SMA PGRI 2	JAWA TENGAH	SEMARANG	IPS	PMDKII	2015	44.25	MASUK	
2015	LAKI-LAKI	SMK NEGERI 8 SEMARANG	JAWA TENGAH	SEMARANG	Teknik Komputer & Jaringan	PMDKII	2015	32.8	TIDAK MASUK	
2015	PEREMPUAN	SMA NEGERI 1 Toroh	JAWA TENGAH	SEMARANG	IPA	PMDKII	2015	37.65	TIDAK MASUK	
2015	LAKI-LAKI	SMA NEGERI 1 GUBUG	JAWA TENGAH	SEMARANG	IPA	PMDKII	2015	37.8	TIDAK MASUK	
2015	LAKI-LAKI	SMA Islam Hidayatullah	JAWA TENGAH	SEMARANG	IPA	PMDKII	2015	40.5	TIDAK MASUK	
2015	LAKI-LAKI	SMA Islam Hidayatullah	JAWA TENGAH	UNGARAN	IPA	PMDKII	2015	49.97	TIDAK MASUK	
2015	LAKI-LAKI	SMA NEGERI 1 UNGARAN	JAWA TENGAH	UNGARAN	IPA	PMDKII	2015	45.75	TIDAK MASUK	
2015	LAKI-LAKI	SMA NEGERI 1 UNGARAN	JAWA TENGAH	UNGARAN	IPA	PMDKII	2015	44.78	TIDAK MASUK	
2015	LAKI-LAKI	SMA NEGERI 1 UNGARAN	JAWA TENGAH	UNGARAN	IPA	PMDKII	2015	44.35	TIDAK MASUK	

Gambar 4.2 Data yang Telah dilakukan Proses Cleaning

4.4 Transformation data

Tahap transformasi data merupakan tahap ketiga dari proses KDD dimana proses tersebut melakukan klasifikasi data. Seluruh data yang telah diambil kemudian dilakukan transformasi yang disesuaikan dengan attribute yang ada. Pada attribute sekolah diklasifikasikan menjadi 2 yaitu swasta dan negeri, provinsi diklasifikasikan menjadi 2 yaitu Jawa dan luar Jawa, kota dibedakan sesuai jarak kota asal dengan Universitas Dian Nuswantoro yaitu menjadi jauh dan dekat, jurusan diklasifikasikan menjadi IPA, IPS dan kejuruan, kemudian pada attribute gelombang di klasifikasikan menjadi 3 yaitu PMDK, REGULER DAN UMBPTS dan untuk attribute nilai dibedakan nilai >40 dinyatakan kecil sedangkan nilai <40 dinyatakan besar. Proses transformasi data dapat dilihat pada table dibawah :

Tahun masuk	sex	nama_sekolah	nama_sekolah	provinsi	nama_prov	kota	nama_kota
2015	LAKI-LAKI	swasta	SMA Kristen YSKI	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	negeri	SMK NEGERI 7 SEMARANG	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	swasta	SMK TELEKOMUNIKASI TUNAS HARAPAN	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	swasta	SMA Kebon Dalem	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	PEREMPUAN	swasta	SMA Islam Hidayatullah	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	PEREMPUAN	negeri	SMA NEGERI 6	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	PEREMPUAN	negeri	SMA Negeri 1 Jepara	Jawa	JAWA TENGAH	jauh	JEPARA
2015	LAKI-LAKI	swasta	MADRASAH ALIYAH	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	negeri	SMA NEGERI 3 PATI	Jawa	JAWA TENGAH	jauh	PATI
2015	LAKI-LAKI	swasta	SMA Kristen Tri Tunggal	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	negeri	SMA NEGERI 1 KENDAL	Jawa	JAWA TENGAH	jauh	KENDAL
2015	PEREMPUAN	swasta	SMA NASIMA SEMARANG	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	PEREMPUAN	swasta	SMK NUSA PUTERA	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	PEREMPUAN	negeri	SMA NEGERI 2 KUDUS	Jawa	JAWA TENGAH	jauh	KUDUS
2015	PEREMPUAN	negeri	SMA NEGERI 8	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	swasta	SMK TELEKOMUNIKASI TUNAS HARAPAN	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	swasta	SMK TELEKOMUNIKASI TUNAS HARAPAN	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	swasta	SMK TELEKOMUNIKASI TUNAS HARAPAN	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	swasta	SMK YAYASAN PHARMASI	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	PEREMPUAN	swasta	SMK NUSA PUTERA	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	PEREMPUAN	negeri	SMK NEGERI 8 SEMARANG	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	negeri	SMA NEGERI 9	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	swasta	SMA Theresiana 1	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	swasta	SMA Theresiana 1	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	PEREMPUAN	swasta	SMA Theresiana 1	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	PEREMPUAN	swasta	SMA Theresiana 1	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	PEREMPUAN	swasta	SMA Kesatrian 2	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	PEREMPUAN	swasta	SMA Islam Sultan Agung 1	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	swasta	SMA Theresiana 1	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	PEREMPUAN	swasta	SMK PERDANA	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	swasta	SMA Kesatrian 1	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	negeri	SMA PGRI 2	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	negeri	SMK NEGERI 8 SEMARANG	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	PEREMPUAN	negeri	SMA NEGERI 1 Toroh	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	negeri	SMA NEGERI 1 GUBUG	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	swasta	SMA Islam Hidayatullah	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	swasta	SMA Islam Hidayatullah	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	negeri	SMA NEGERI 1 UNGARAN	Jawa	JAWA TENGAH	dekat	UNGARAN
2015	LAKI-LAKI	negeri	SMA NEGERI 1 UNGARAN	Jawa	JAWA TENGAH	dekat	UNGARAN
2015	LAKI-LAKI	negeri	SMA NEGERI 1 UNGARAN	Jawa	JAWA TENGAH	dekat	UNGARAN
2015	LAKI-LAKI	negeri	SMK NEGERI 2 Rembang	Jawa	JAWA TENGAH	dekat	REMBANG
2015	LAKI-LAKI	negeri	SMA NEGERI 1 KENDAL	Jawa	JAWA TENGAH	jauh	KENDAL
2015	LAKI-LAKI	negeri	SMA NEGERI 1 BATANG	Jawa	JAWA TENGAH	jauh	BATANG
2015	PEREMPUAN	swasta	SMK TELEKOMUNIKASI TUNAS HARAPAN	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	negeri	SMA Negeri 14	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	negeri	SMA PGRI 2	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	negeri	SMA Negeri 10	Jawa	JAWA TENGAH	dekat	SEMARANG
2015	LAKI-LAKI	negeri	SMA Negeri 10	Jawa	JAWA TENGAH	dekat	SEMARANG

Gambar 4.3 Data yang Telah dilakukan Proses Transformasi

jurusan	jurusan	gel_daf	thn_ijazah	jumlah_nilai	jumlah_nilai_us	Keterangan
IPS	IPS	PMDKI	2015	BESAR	45	MASUK
KEJURUAN	Teknik Komputer & Jaringan	PMDKI	2014	BESAR	40.3	TIDAK MASUK
KEJURUAN	Teknik Komputer & Jaringan	PMDKI	2015	BESAR	47.2	TIDAK MASUK
IPA	IPA	PMDKI	2015	BESAR	37.4	TIDAK MASUK
IPA	IPA	PMDKI	2015	BESAR	45.3	TIDAK MASUK
IPA	IPA	PMDKI	2015	BESAR	38.5	TIDAK MASUK
IPS	IPS	PMDKI	2015	KECIL	35.75	MASUK
IPA	IPA	PMDKI	2015	BESAR	42	TIDAK MASUK
IPA	IPA	PMDKI	2015	BESAR	44.6	TIDAK MASUK
IPS	IPS	PMDKI	2015	KECIL	36.4	MASUK
IPS	IPS	PMDKI	2015	BESAR	48	MASUK
KEJURUAN	Teknik Komputer & Jaringan	PMDKI	2015	BESAR	44.95	TIDAK MASUK
KEJURUAN	Teknik Komputer & Jaringan	PMDKI	2015	BESAR	47.5	TIDAK MASUK
IPA	IPA	PMDKI	2015	BESAR	49.67	TIDAK MASUK
IPA	IPA	PMDKI	2015	BESAR	44.4	TIDAK MASUK
KEJURUAN	Teknik Komputer & Jaringan	PMDKI	2015	BESAR	50	TIDAK MASUK
KEJURUAN	Teknik Komputer & Jaringan	PMDKI	2015	BESAR	43.9	TIDAK MASUK
KEJURUAN	Teknik Komputer & Jaringan	PMDKI	2015	KECIL	32.85	TIDAK MASUK
KEJURUAN	Farmasi	PMDKI	2015	KECIL	31.5	TIDAK MASUK
KEJURUAN	Teknik Komputer & Jaringan	PMDKI	2015	BESAR	49.78	TIDAK MASUK
KEJURUAN	Rekayasa Perangkat Lunak	PMDKI	2015	KECIL	37	TIDAK MASUK
IPA	IPA	PMDKI	2015	BESAR	41.5	TIDAK MASUK
IPS	IPS	PMDKII	2015	BESAR	51.2	MASUK
IPS	IPS	PMDKII	2015	KECIL	36.8	MASUK
IPS	IPS	PMDKII	2015	BESAR	36.86	MASUK
IPA	IPA	PMDKII	2015	KECIL	39.6	TIDAK MASUK
IPS	IPS	PMDKII	2015	BESAR	42.7	MASUK
IPA	IPA	PMDKII	2015	BESAR	48.8	TIDAK MASUK
IPA	IPA	PMDKII	2015	KECIL	38.75	TIDAK MASUK
KEJURUAN	Multimedia	PMDKII	2015	BESAR	48.62	TIDAK MASUK
IPA	IPA	PMDKII	2015	KECIL	31	TIDAK MASUK
IPS	IPS	PMDKII	2015	BESAR	44.25	MASUK
KEJURUAN	Teknik Komputer & Jaringan	PMDKII	2015	KECIL	32.8	TIDAK MASUK
IPA	IPA	PMDKII	2015	KECIL	37.65	TIDAK MASUK
IPA	IPA	PMDKII	2015	KECIL	37.8	TIDAK MASUK
IPA	IPA	PMDKII	2015	BESAR	40.5	TIDAK MASUK
IPA	IPA	PMDKII	2015	BESAR	49.97	TIDAK MASUK
IPA	IPA	PMDKII	2015	BESAR	45.75	TIDAK MASUK
IPA	IPA	PMDKII	2015	BESAR	44.78	TIDAK MASUK
IPA	IPA	PMDKII	2015	BESAR	44.35	TIDAK MASUK
KEJURUAN	Perikanan	PMDKII	2015	BESAR	43.7	TIDAK MASUK
IPA	IPA	PMDKII	2015	BESAR	43.05	TIDAK MASUK
IPS	IPS	PMDKII	2015	BESAR	48	MASUK
KEJURUAN	Teknik Komputer & Jaringan	PMDKII	2015	BESAR	40.2	MASUK
IPA	IPA	PMDKII	2015	KECIL	36.8	TIDAK MASUK
IPA	IPA	PMDKII	2015	BESAR	45.25	TIDAK MASUK
IPS	IPS	PMDKII	2015	BESAR	49.95	MASUK
IPS	IPS	PMDKII	2015	BESAR	45.35	MASUK

Gambar 4.4 Data yang Telah dilakukan Proses Transformasi

4.5 Data Mining

Data Mining dengan menggunakan algoritma C4.5 akan menghasilkan *Decision Tree* atau pohon keputusan. Pada penghitungan data mining algoritma C4.5 dipecah menjadi data training dan data testing. Kemudian akan mengambil contoh atau jumlah sample data untuk pembentukan *Decision Tree*. Sample data training akan diambil 100 sample data training, dari data calon mahasiswa pada tahun 2013 dan tahun 2014 secara acak dan data telah dilakukan proses cleaning dan

transformasi menjadi data training. Data calon mahasiswa pada tahun 2015 akan dijadikan data testing dan diambil 100 data dengan attribute yang sama. 100 data dapat dilihat pada table berikut ini:

Tahun masuk nim	sex	nama_sekolah	keterangan	kota	Jurusan	gel_daf	tho_ijazah rata uan	Keterangan
2014 A11.2014.08402	LAKI-LAKI	NEGERI	JAWA	DEKAT	IPA	REGULER	2014 BESAR	TIDAK MASUK
2014 A11.2014.08354	LAKI-LAKI	NEGERI	JAWA	DEKAT	IPS	REGULER	2014 BESAR	TIDAK MASUK
2014 A12.2014.05194	LAKI-LAKI	NEGERI	JAWA	DEKAT	IPA	REGULER	2014 BESAR	MASUK
2014 A12.2014.05171	LAKI-LAKI	NEGERI	JAWA	DEKAT	IPA	REGULER	2014 BESAR	MASUK
2014 A12.2014.05148	PEREMPUAN	SWASTA	JAWA	DEKAT	IPA	REGULER	2014 BESAR	MASUK
2014 A12.2014.05227	LAKI-LAKI	NEGERI	JAWA	JAUH	IPS	REGULER	2014 KECIL	MASUK
2014 A11.2014.08683	LAKI-LAKI	NEGERI	JAWA	JAUH	IPA	REGULER	2014 BESAR	TIDAK MASUK
2014 A12.2014.05161	LAKI-LAKI	NEGERI	JAWA	JAUH	KEJURUAN	REGULER	2014 BESAR	MASUK
2014 A11.2014.08014	PEREMPUAN	NEGERI	JAWA	DEKAT	KEJURUAN	PMDK	2014 BESAR	TIDAK MASUK
2014 A11.2014.08032	PEREMPUAN	NEGERI	JAWA	JAUH	IPA	PMDK	2014 BESAR	TIDAK MASUK
2014 A11.2014.08337	PEREMPUAN	SWASTA	JAWA	DEKAT	KEJURUAN	PMDK	2014 BESAR	TIDAK MASUK
2014 A11.2014.08024	PEREMPUAN	NEGERI	JAWA	JAUH	IPA	PMDK	2014 BESAR	TIDAK MASUK
2014 A11.2014.08026	LAKI-LAKI	NEGERI	JAWA	JAUH	IPA	PMDK	2014 BESAR	TIDAK MASUK
2014 A12.2014.05064	PEREMPUAN	NEGERI	JAWA	JAUH	IPA	PMDK	2014 BESAR	TIDAK MASUK
2014 A11.2014.08012	PEREMPUAN	NEGERI	JAWA	DEKAT	IPA	PMDK	2014 BESAR	TIDAK MASUK
2014 A11.2014.08627	LAKI-LAKI	NEGERI	JAWA	DEKAT	IPS	PMDK	2014 BESAR	TIDAK MASUK
2014 A11.2014.08019	LAKI-LAKI	SWASTA	JAWA	DEKAT	IPA	PMDK	2014 BESAR	TIDAK MASUK
2014 A14.2014.01822	PEREMPUAN	NEGERI	JAWA	DEKAT	IPA	PMDK	2014 BESAR	TIDAK MASUK
2014 A12.2014.05076	PEREMPUAN	NEGERI	JAWA	JAUH	IPS	PMDK	2014 BESAR	MASUK
2014 A12.2014.05066	PEREMPUAN	NEGERI	JAWA	JAUH	IPS	PMDK	2014 BESAR	MASUK
2014 A11.2014.08082	LAKI-LAKI	SWASTA	JAWA	DEKAT	IPA	PMDK	2014 BESAR	TIDAK MASUK
2014 A12.2014.05164	LAKI-LAKI	SWASTA	JAWA	DEKAT	IPS	PMDK	2014 KECIL	MASUK
2014 A12.2014.05084	LAKI-LAKI	NEGERI	JAWA	DEKAT	IPA	PMDK	2014 BESAR	MASUK
2014 A12.2014.05083	PEREMPUAN	NEGERI	JAWA	DEKAT	IPS	PMDK	2014 BESAR	MASUK
2014 A11.2014.08133	LAKI-LAKI	SWASTA	JAWA	DEKAT	KEJURUAN	PMDK	2014 BESAR	TIDAK MASUK
2014 A11.2014.08045	PEREMPUAN	SWASTA	JAWA	JAUH	KEJURUAN	PMDK	2014 BESAR	TIDAK MASUK
2014 A11.2014.08044	LAKI-LAKI	SWASTA	JAWA	JAUH	KEJURUAN	PMDK	2014 BESAR	TIDAK MASUK
2014 A12.2014.05088	LAKI-LAKI	SWASTA	JAWA	JAUH	KEJURUAN	PMDK	2014 BESAR	MASUK
2014 A12.2014.05089	LAKI-LAKI	NEGERI	JAWA	DEKAT	KEJURUAN	PMDK	2014 BESAR	MASUK
2014 A12.2014.05090	PEREMPUAN	NEGERI	JAWA	DEKAT	IPS	PMDK	2014 BESAR	MASUK
2014 A11.2014.08121	LAKI-LAKI	SWASTA	JAWA	DEKAT	KEJURUAN	PMDK	2014 BESAR	TIDAK MASUK
2014 A11.2014.08084	PEREMPUAN	SWASTA	JAWA	DEKAT	IPA	PMDK	2014 KECIL	TIDAK MASUK
2014 A12.2014.05101	LAKI-LAKI	NEGERI	JAWA	DEKAT	IPA	PMDK	2014 BESAR	MASUK

Gambar 4.5 Data Training Sample dari Nomor 1 - 34

2014 LAKI-LAKI	SWASTA	JAWA	DEKAT	IPS	PMDK	2014 BESAR	MASUK
2014 LAKI-LAKI	NEGERI	JAWA	JAUH	IPS	PMDK	2014 BESAR	MASUK
2014 LAKI-LAKI	NEGERI	JAWA	JAUH	IPS	PMDK	2014 BESAR	MASUK
2014 LAKI-LAKI	SWASTA	JAWA	DEKAT	IPS	PMDK	2014 BESAR	MASUK
2014 PEREMPUAN	SWASTA	JAWA	DEKAT	IPS	PMDK	2014 BESAR	MASUK
2014 LAKI-LAKI	NEGERI	JAWA	JAUH	IPA	PMDK	2014 BESAR	TIDAK MASUK
2014 LAKI-LAKI	SWASTA	JAWA	DEKAT	KEJURUAN	PMDK	2014 BESAR	TIDAK MASUK
2014 LAKI-LAKI	SWASTA	JAWA	JAUH	IPS	PMDK	2014 BESAR	MASUK
2014 LAKI-LAKI	NEGERI	JAWA	JAUH	IPS	PMDK	2014 KECIL	MASUK
2014 LAKI-LAKI	NEGERI	JAWA	JAUH	KEJURUAN	PMDK	2014 BESAR	MASUK
2014 LAKI-LAKI	NEGERI	JAWA	JAUH	IPS	PMDK	2014 KECIL	MASUK
2014 LAKI-LAKI	NEGERI	JAWA	JAUH	IPA	PMDK	2014 BESAR	MASUK
2014 LAKI-LAKI	SWASTA	JAWA	DEKAT	IPS	PMDK	2014 BESAR	MASUK
2014 LAKI-LAKI	SWASTA	JAWA	DEKAT	IPA	PMDK	2014 BESAR	MASUK
2014 PEREMPUAN	SWASTA	JAWA	DEKAT	IPS	PMDK	2014 BESAR	MASUK
2014 LAKI-LAKI	SWASTA	JAWA	JAUH	KEJURUAN	REGULER	2014 BESAR	MASUK
2014 LAKI-LAKI	SWASTA	JAWA	JAUH	KEJURUAN	REGULER	2014 BESAR	TIDAK MASUK
2014 LAKI-LAKI	NEGERI	JAWA	JAUH	IPS	REGULER	2014 BESAR	MASUK
2014 LAKI-LAKI	NEGERI	JAWA	JAUH	IPA	REGULER	2014 BESAR	MASUK
2014 PEREMPUAN	NEGERI	JAWA	JAUH	IPA	REGULER	2014 BESAR	MASUK
2014 LAKI-LAKI	NEGERI	JAWA	JAUH	IPS	REGULER	2014 KECIL	MASUK
2014 LAKI-LAKI	NEGERI	JAWA	JAUH	KEJURUAN	REGULER	2014 BESAR	MASUK
2014 PEREMPUAN	NEGERI	LUAR JAWA	JAUH	IPA	REGULER	2014 BESAR	MASUK
2014 LAKI-LAKI	NEGERI	JAWA	JAUH	IPA	REGULER	2014 BESAR	MASUK
2014 PEREMPUAN	NEGERI	JAWA	JAUH	IPS	REGULER	2014 KECIL	MASUK
2014 LAKI-LAKI	NEGERI	JAWA	JAUH	IPA	REGULER	2014 BESAR	MASUK
2014 LAKI-LAKI	SWASTA	JAWA	DEKAT	IPS	REGULER	2014 BESAR	MASUK
2014 LAKI-LAKI	NEGERI	JAWA	JAUH	IPS	REGULER	2014 BESAR	MASUK
2014 LAKI-LAKI	SWASTA	JAWA	DEKAT	IPA	REGULER	2014 BESAR	MASUK
2014 LAKI-LAKI	SWASTA	LUAR JAWA	JAUH	Bahasa	REGULER	2014 KECIL	MASUK
2014 LAKI-LAKI	SWASTA	JAWA	DEKAT	IPA	REGULER	2014 KECIL	MASUK
2014 PEREMPUAN	NEGERI	JAWA	DEKAT	KEJURUAN	REGULER	2014 BESAR	MASUK
2014 LAKI-LAKI	NEGERI	JAWA	DEKAT	KEJURUAN	REGULER	2014 KECIL	MASUK
2014 LAKI-LAKI	SWASTA	JAWA	DEKAT	IPS	REGULER	2014 BESAR	MASUK

Gambar 4.6 Data Training Sample Nomor 35-68

2014 LAKI-LAKI	SWASTA	JAWA	DEKAT	IPS	REGULER	2014 BESAR	MASUK
2014 PEREMPUAN	NEGERI	JAWA	DEKAT	IPA	REGULER	2014 BESAR	MASUK
2013 PEREMPUAN	NEGERI	JAWA	JAUH	IPA	PMDK	2013 BESAR	TIDAK MASUK
2013 LAKI-LAKI	NEGERI	JAWA	JAUH	IPA	PMDK	2013 KECIL	TIDAK MASUK
2013 LAKI-LAKI	SWASTA	JAWA	DEKAT	IPA	PMDK	2013 BESAR	TIDAK MASUK
2013 PEREMPUAN	NEGERI	JAWA	JAUH	IPA	PMDK	2013 BESAR	MASUK
2013 LAKI-LAKI	NEGERI	JAWA	DEKAT	IPS	PMDK	2012 BESAR	TIDAK MASUK
2013 PEREMPUAN	NEGERI	JAWA	JAUH	KEJURUAN	PMDK	2013 BESAR	MASUK
2013 LAKI-LAKI	NEGERI	JAWA	JAUH	IPS	PMDK	2013 BESAR	MASUK
2013 LAKI-LAKI	NEGERI	JAWA	DEKAT	KEJURUAN	PMDK	2013 KECIL	TIDAK MASUK
2013 LAKI-LAKI	SWASTA	JAWA	DEKAT	IPS	PMDK	2013 KECIL	MASUK
2013 LAKI-LAKI	NEGERI	JAWA	JAUH	IPA	PMDK	2013 KECIL	TIDAK MASUK
2013 LAKI-LAKI	NEGERI	JAWA	JAUH	IPS	REGULER	2013 BESAR	MASUK
2013 LAKI-LAKI	NEGERI	JAWA	JAUH	IPS	REGULER	2013 BESAR	MASUK
2013 LAKI-LAKI	NEGERI	JAWA	JAUH	KEJURUAN	REGULER	2013 BESAR	TIDAK MASUK
2013 LAKI-LAKI	NEGERI	JAWA	JAUH	IPS	REGULER	2013 BESAR	MASUK
2013 LAKI-LAKI	NEGERI	JAWA	DEKAT	IPS	REGULER	2013 KECIL	MASUK
2013 LAKI-LAKI	SWASTA	JAWA	DEKAT	KEJURUAN	REGULER	2013 BESAR	MASUK
2013 LAKI-LAKI	NEGERI	JAWA	DEKAT	IPS	REGULER	2013 KECIL	TIDAK MASUK
2013 LAKI-LAKI	SWASTA	JAWA	DEKAT	IPS	REGULER	2013 BESAR	MASUK
2013 LAKI-LAKI	SWASTA	JAWA	DEKAT	IPS	REGULER	2013 BESAR	MASUK
2013 LAKI-LAKI	NEGERI	JAWA	JAUH	IPS	REGULER	2013 BESAR	TIDAK MASUK
2013 PEREMPUAN	NEGERI	JAWA	JAUH	IPS	REGULER	2013 KECIL	MASUK
2013 PEREMPUAN	NEGERI	LUAR JAWA	JAUH	IPS	REGULER	2013 KECIL	MASUK
2013 PEREMPUAN	SWASTA	JAWA	JAUH	IPS	REGULER	2013 BESAR	MASUK
2013 LAKI-LAKI	NEGERI	JAWA	JAUH	IPS	REGULER	2013 BESAR	MASUK
2013 LAKI-LAKI	SWASTA	JAWA	JAUH	KEJURUAN	REGULER	2013 KECIL	MASUK
2013 PEREMPUAN	SWASTA	LUAR JAWA	JAUH	IPA	REGULER	2013 KECIL	MASUK
2013 LAKI-LAKI	SWASTA	JAWA	JAUH	KEJURUAN	REGULER	2013 BESAR	MASUK
2013 LAKI-LAKI	NEGERI	JAWA	JAUH	IPA	REGULER	2013 BESAR	MASUK
2013 PEREMPUAN	SWASTA	JAWA	JAUH	KEJURUAN	REGULER	2013 BESAR	MASUK
2013 LAKI-LAKI	SWASTA	JAWA	DEKAT	IPS	REGULER	2013 BESAR	MASUK
2013 PEREMPUAN	NEGERI	JAWA	JAUH	IPA	PMDK	2013 BESAR	TIDAK MASUK
2013 PEREMPUAN	NEGERI	JAWA	DEKAT	IPA	UMBPTS	2013 BESAR	MASUK

Gambar 4.7 Data Training Sample dari Nomor 69-100

Setelah sample data training telah dilakukan proses cleaning dan transformasi. Kemudian dilakukan proses penghitungan entropy dan gain untuk menentukan *decision tree*. Langkah pertama dengan mencari jumlah keseluruhan data sample dan melakukan penghitungan *entropy* dan *gain*.

Jumlah data : 100 data

Jumlah data dengan keterangan masuk : 68 data

Jumlah data dengan keterangan tidak masuk : 32 data

1. Menghitung nilai entropy dari total data sample:

$$\begin{aligned}
 \text{Entropy data} &= \left(\left(-\frac{68}{100} \right) * \log_2 \left(\frac{68}{100} \right) \right) + \left(\left(-\frac{32}{100} \right) * \log_2 \left(\frac{32}{100} \right) \right) \\
 &= (-0,68 * -0,556) + (-0,32 * -1,6438) \\
 &= 0,378 + 0,5260 \\
 &= 0,9040
 \end{aligned}$$

2. Melakukan penghitungan pada seluruh entropy pada setiap attribute:

$$\begin{aligned}
 \text{Entropy tahun 2014} &= \left(\left(-\frac{46}{68} \right) * \log_2 \left(\frac{46}{68} \right) \right) + \left(\left(-\frac{22}{68} \right) * \log_2 \left(\frac{22}{68} \right) \right) \\
 &= (-0,6764 * -0,5639) + (-0,3235 * -1,6428) \\
 &= 0,4691 + 0,5266 \\
 &= 0,9040
 \end{aligned}$$

$$\begin{aligned}
 \text{Entropy tahun 2013} &= \left(\left(-\frac{22}{32} \right) * \log_2 \left(\frac{22}{32} \right) \right) + \left(\left(-\frac{10}{32} \right) * \log_2 \left(\frac{10}{32} \right) \right) \\
 &= (-0,6875 * -0,5406) + (-0,3125 * -1,6780) \\
 &= 0,37166 + 0,5244 \\
 &= 0,8961
 \end{aligned}$$

$$\begin{aligned}
 \text{Entropy sex laki-laki} &= \left(\left(-\frac{48}{69} \right) * \log_2 \left(\frac{48}{69} \right) \right) + \left(\left(-\frac{21}{69} \right) * \log_2 \left(\frac{21}{69} \right) \right) \\
 &= (-0,6956 * -0,5325) + (-0,3043 * -1,7162)
 \end{aligned}$$

$$\begin{aligned}
&= 0,3641 + 0,5222 \\
&= 0,8863 \\
\text{Entropy sex perempuan} &= \left(\left(-\frac{20}{31} \right) * \log_2 \left(\frac{20}{31} \right) \right) + \left(\left(-\frac{11}{31} \right) * \log_2 \left(\frac{11}{31} \right) \right) \\
&= (-0,6451 * -0,6322) + (-0,3548 * -1,4947) \\
&= 0,4078 + 0,5303 \\
&= 0,9381 \\
\text{Entropy sekolah Negeri} &= \left(\left(-\frac{42}{63} \right) * \log_2 \left(\frac{42}{63} \right) \right) + \left(\left(-\frac{21}{63} \right) * \log_2 \left(\frac{21}{63} \right) \right) \\
&= (-0,6667 * -0,5849) + (-0,333 * -1,584) \\
&= 0,3899 + 0,5277 \\
&= 0,9176 \\
\text{Entropy sekolah Swasta} &= \left(\left(-\frac{31}{37} \right) * \log_2 \left(\frac{31}{37} \right) \right) + \left(\left(-\frac{11}{37} \right) * \log_2 \left(\frac{11}{37} \right) \right) \\
&= (-0,8378 * -0,2552) + (-0,2972 * -1,7500) \\
&= 0,2138 + 0,5201 \\
&= 0,7339 \\
\text{Entropy provinsi Jawa} &= \left(\left(-\frac{65}{96} \right) * \log_2 \left(\frac{65}{96} \right) \right) + \left(\left(-\frac{31}{96} \right) * \log_2 \left(\frac{31}{96} \right) \right) \\
&= (-0,6770 * -0,5625) + (-0,3229 * -1,6307) \\
&= 0,3808 + 0,5265 \\
&= 0,9073 \\
\text{Entropy provinsi luar Jawa} &= \left(\left(-\frac{4}{4} \right) * \log_2 \left(\frac{4}{4} \right) \right) + \left(\left(-\frac{0}{4} \right) * \log_2 \left(\frac{0}{4} \right) \right)
\end{aligned}$$

$$= (-1 * -0) + (-0 * -0)$$

$$= 0 + 0$$

$$= 0$$

$$\text{Entropy kota dekat} = \left(\left(-\frac{29}{46} \right) * \log_2 \left(\frac{29}{46} \right) \right) + \left(\left(-\frac{17}{46} \right) * \log_2 \left(\frac{17}{46} \right) \right)$$

$$= (-0,6304 * -0,6656) + (-0,3966 * -1,4361)$$

$$= 0,4196 + 0,5696$$

$$= 0,9891$$

$$\text{Entropy kota jauh} = \left(\left(-\frac{39}{54} \right) * \log_2 \left(\frac{39}{54} \right) \right) + \left(\left(-\frac{15}{54} \right) * \log_2 \left(\frac{15}{54} \right) \right)$$

$$= (-0,7222 * -0,4694) + (-0,2778 * -1,847)$$

$$= 0,3990 + 0,5133$$

$$\text{Entropy jurusan IPS} = \left(\left(-\frac{36}{40} \right) * \log_2 \left(\frac{36}{40} \right) \right) + \left(\left(-\frac{4}{40} \right) * \log_2 \left(\frac{4}{40} \right) \right)$$

$$= (-0,9 * -0,1520) + (-0,1 * -3,3219)$$

$$= 0,1368 + 0,3321$$

$$= 0,4689$$

$$\text{Entropy jurusan IPA} = \left(\left(-\frac{19}{36} \right) * \log_2 \left(\frac{19}{36} \right) \right) + \left(\left(-\frac{17}{36} \right) * \log_2 \left(\frac{17}{36} \right) \right)$$

$$= (-0,5277 * -0,9220) + (-0,4722 * -1,0824)$$

$$= 0,4866 + 0,5112$$

$$= 0,9978$$

$$\text{Entropy jurusan}$$

$$\text{Kejuruan} = \left(\left(-\frac{13}{23} \right) * \log_2 \left(\frac{13}{23} \right) \right) + \left(\left(-\frac{10}{23} \right) * \log_2 \left(\frac{10}{23} \right) \right)$$

$$= (-0,5652 * -0,8231) + (-0,4347 * -1,2016)$$

$$\begin{aligned}
&= 0,4652 + 0,5223 \\
&= 0,9875 \\
\text{Entropy jurusan} &= \left(\left(-\frac{1}{1} \right) * \log_2 \left(\frac{1}{1} \right) \right) + \left(\left(-\frac{0}{1} \right) * \log_2 \left(\frac{0}{1} \right) \right) \\
\text{Bahasa} &= (-1 * -0) + (- * -0) \\
&= 0 + 0 \\
&= 0 \\
\text{Entropy gelombang} &= \left(\left(-\frac{26}{51} \right) * \log_2 \left(\frac{26}{51} \right) \right) + \left(\left(-\frac{25}{51} \right) * \log_2 \left(\frac{25}{51} \right) \right) \\
\text{PMDK} &= (-0,5098 * -0,9719) + (-0,4902 * -1,0286) \\
&= 0,4955 + 0,5042 \\
&= 0,9909 \\
\text{Entropy gelombang} &= \left(\left(-\frac{43}{48} \right) * \log_2 \left(\frac{43}{48} \right) \right) + \left(\left(-\frac{5}{48} \right) * \log_2 \left(\frac{5}{48} \right) \right) \\
\text{Regular} &= (-0,8958 * -0,1586) + (-0,1042 * -3,2630) \\
&= 0,1425 + 0,3400 \\
&= 0,9909 \\
\text{Entropy gelombang} &= \left(\left(-\frac{1}{1} \right) * \log_2 \left(\frac{1}{1} \right) \right) + \left(\left(-\frac{0}{1} \right) * \log_2 \left(\frac{0}{1} \right) \right) \\
\text{UMBPTS} &= (-1 * -0) + (-0 * -0) \\
&= 0 + 0 \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\text{Entropy ijazah 2014} &= \left(\left(-\frac{46}{68} \right) * \log_2 \left(\frac{46}{68} \right) \right) + \left(\left(-\frac{22}{68} \right) * \log_2 \left(\frac{22}{68} \right) \right) \\
&= (-0,6764 * -0,5639) + (-0,3235 * -1,6280) \\
&= 0,3814 + 0,5266 \\
&= 0,9080
\end{aligned}$$

$$\begin{aligned}
\text{Entropy ijazah 2013} &= \left(\left(-\frac{22}{31} \right) * \log_2 \left(\frac{22}{31} \right) \right) + \left(\left(-\frac{9}{31} \right) * \log_2 \left(\frac{9}{31} \right) \right) \\
&= (-0,7096 * -0,4947) + (-0,2903 * -1,7842) \\
&= 0,3510 + 0,5179 \\
&= 0,8689
\end{aligned}$$

$$\begin{aligned}
\text{Entropy ijazah 2012} &= \left(\left(-\frac{0}{1} \right) * \log_2 \left(\frac{0}{1} \right) \right) + \left(\left(-\frac{1}{1} \right) * \log_2 \left(\frac{1}{1} \right) \right) \\
&= (-0 * -0) + (-1 * -0) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\text{Entropy nilai besar} &= \left(\left(-\frac{53}{80} \right) * \log_2 \left(\frac{53}{80} \right) \right) + \left(\left(-\frac{27}{80} \right) * \log_2 \left(\frac{27}{80} \right) \right) \\
&= (-0,6625 * -0,5940) + (-0,3375 * -1,5670) \\
&= 0,3935 + 0,5288 \\
&= 0,9223
\end{aligned}$$

$$\begin{aligned}
\text{Entropy nilai kecil} &= \left(\left(-\frac{15}{20} \right) * \log_2 \left(\frac{15}{20} \right) \right) + \left(\left(-\frac{5}{20} \right) * \log_2 \left(\frac{5}{20} \right) \right) \\
&= (-0,75 * -0,4150) + (-0,25 * -2) \\
&= 0,3112 + 0,5 \\
&= 0,8112
\end{aligned}$$

3. Setelah dilakukan perhitungan *entropy* dari seluruh *attribute* yang ada, kemudian melakukan pencarian nilai *gain* dari setiap *attribute*. Dengan rumus seperti :

$$Gain\ ratio(a) = Entropy(X) - \sum_{j=1}^k \frac{|X_i|}{|X|} * Entropy(X_i) \quad (4.1)$$

$$\begin{aligned} \text{Gain} &= 0,9040 - \left(\left(\frac{68}{100} * 0,9957 \right) + \left(\frac{32}{100} * 0,8961 \right) \right) \\ \text{(tahun)} &= 0,9040 - (0,6770 + 0,2867) \\ &= 0,0597 \end{aligned}$$

$$\begin{aligned} \text{Gain} &= 0,9040 - \left(\left(\frac{69}{100} * 0,8863 \right) + \left(\frac{31}{100} * 0,9381 \right) \right) \\ \text{(sex)} &= 0,9040 - (0,6115 + 0,2908) \\ &= 0,0017 \end{aligned}$$

$$\begin{aligned} \text{Gain} &= 0,9040 - \left(\left(\frac{63}{100} * 0,9176 \right) + \left(\frac{37}{100} * 0,7339 \right) \right) \\ \text{(sekolah)} &= 0,9040 - (0,6115 + 0,2908) \\ &= 0,0545 \end{aligned}$$

$$\begin{aligned} \text{Gain} &= 0,9040 - \left(\left(\frac{96}{100} * 0,9073 \right) + \left(\frac{4}{100} * 0 \right) \right) \\ \text{(provinsi)} &= 0,9040 - (0,6115 + 0,2908) \\ &= 0,033 \end{aligned}$$

$$\begin{aligned} \text{Gain} &= 0,9040 - \left(\left(\frac{46}{100} * 0,9891 \right) + \left(\frac{54}{100} * 0,7381 \right) \right) \\ \text{(kota)} &= 0,9040 - (0,4549 + 0,3985) \end{aligned}$$

$$= 0,0506$$

$$\begin{aligned} \text{Gain} &= \\ \text{(jurusan)} &= 0,9040 - \left(\left(\frac{40}{100} * 0,4689 \right) + \left(\frac{36}{100} * 0,9978 \right) + \right. \\ &\quad \left. \left(\frac{23}{100} * 0,9875 \right) + \left(\frac{1}{100} * 0 \right) \right) \end{aligned}$$

$$= 0,9040 - (0,1875 + 0,3592 + 0,2271 + 0)$$

$$= 0,9040 - 0,7738$$

$$= 0,1302$$

$$\begin{aligned} \text{Gain} &= \\ \text{(gelombang)} &= 0,9040 - \left(\left(\frac{51}{100} * 0,9909 \right) + \left(\frac{48}{100} * 0,4825 \right) + \left(\frac{1}{100} * 0 \right) \right) \end{aligned}$$

$$= 0,9040 - (0,5053 + 0,2316 + 0)$$

$$= 0,1671$$

$$\begin{aligned} \text{Gain} &= \\ \text{(ijazah)} &= 0,9040 - \left(\left(\frac{68}{100} * 0,9080 \right) + \left(\frac{31}{100} * 0,8689 \right) + \left(\frac{1}{100} * 0 \right) \right) \end{aligned}$$

$$= 0,9040 - (0,6174 + 0,2693 + 0)$$

$$= 0,017$$

$$\begin{aligned} \text{Gain (nilai)} &= \\ &= 0,9040 - \left(\left(\frac{80}{100} * 0,9223 \right) + \left(\frac{20}{100} * 0,8112 \right) \right) \end{aligned}$$

$$= 0,9040 - (0,7378 + 0,1622 + 0)$$

$$= 0,0040$$

Hasil perhitungan *entropy* dan *gain* dari masing masing *atribute* menghasilkan tabel seperti berikut:

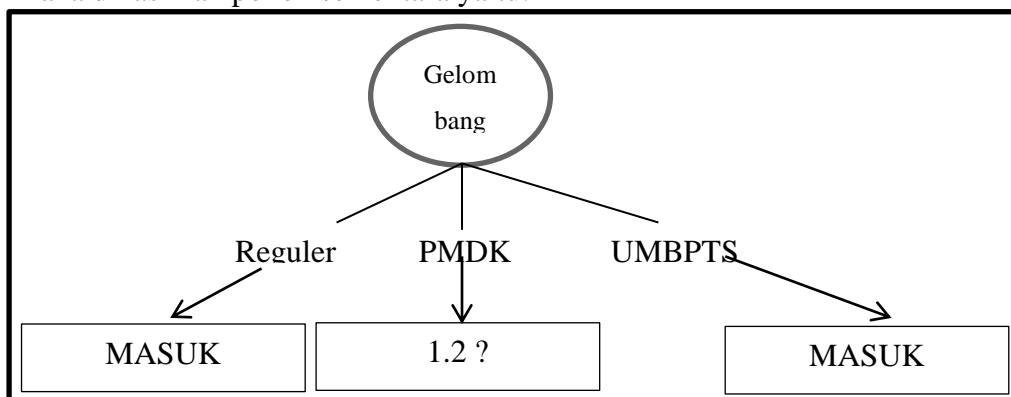
Tabel 4.1 Tabel untuk Membentuk Node 1

Atribut		Banyak Kasus	Masuk	Tidak masuk	Entropy	Gain
Total kasus		100	68	32	0,9040	
Tahun						0,0597
	2014	68	46	22	0,9957	
	2013	32	22	10	0,8961	
Sex						0,0017
	Laki-laki	69	47	21	0,8863	
	Perempuan	31	20	11	0,9381	
Sekolah						0,0544
	Negeri	63	42	21	0,9176	
	Swasta	37	31	11	0,7339	
Provinsi						0,033
	Jawa	96	65	31	0,9073	
	Luar jawa	4	4	0	0	
Kota						0,0506
	Dekat	46	29	17	0,9891	
	Jauh	54	39	15	0,7381	
Jurusan						0,1302
	IPS	40	36	4	0,4689	
	IPA	36	19	17	0,9978	
	Kejuruan	23	13	10	0,9875	
	Bahasa	1	1	0	0	
Gelombang						0,1671
	PMDK	51	26	25	0,9909	
	REGULER	48	43	5	0,4825	
	UMBPTS	1	1	0	0	

Atribut		Banyak Kasus	Masuk	Tidak masuk	Entropy	Gain
Ijazah						0,017
	2014	68	46	22	0,9080	
	2013	31	22	9	0,2693	
	2012	1	0	1	0	
Nilai UN						0,0040
	Besar	80	53	27	0,9223	
	Kecil	20	15	5	0,8112	
					MAX =	0,1671

Setelah dilakukan penghitungan *entropy* dan *gain* didapatkan nilai tertinggi adalah *attribute* gelombang yaitu sebesar 0,1671. Hasil itulah *attribute* gelombang menjadi node akar. Pada *attribute* gelombang terdapat 3 nilai *attribute* yaitu PMDK, Rreguler dan UMBPTS. Pada nilai *attribute* UMBPTS dengan hasil 0 menghasilkan nilai masuk dan regular dengan nilai 0,48725 menghasilkan nilai masuk hasil tertinggi pada *attribute* gelombang yaitu PMDK dengan nilai 0,9909 belum menghasilkan suatu keputusan maka akan menjadi node akar.

4. Maka dihasilkan pohon sementara yaitu:



Gambar 4.8 Pohon Keputusan dengan Node 1

Perhitungan dilakukan dengan mencari nilai dari *attribute* selain yang menjadi *node* akar (**Gelombang**), yaitu dengan mencari jumlah kasus untuk hasil Masuk, Tidak masuk, dan nilai *Entropy* dari semua kasus saat **gelombang = PMDK**. Dan kemudian melakukan perhitungan untuk mendapatkan nilai *Gain*, dan *attribute* dengan nilai *Gain* terbesar, maka akan menjadi *node* cabang dari nilai *attribute* **PMDK**. Kemudian dilakukan penghitungan

Node 1.2:

Jumlah data dengan Gelombang = PMDK : 51

Jumlah data dengan Gelombang = PMDK yang hasil Masuk : 26

Jumlah data dengan Gelombang = PMDK yang hasil Tidak Masuk : 25

5. Hitung entropy total data node 1.2 :
6. Melakukan penghitungan nilai entropy pada setiap *attribute* kecuali pada *attribute* gelombang :

$$\begin{aligned}
 \text{Entropy gelombang PMDK} &= \left(\left(-\frac{26}{51} \right) * \log_2 \left(\frac{26}{51} \right) \right) + \left(\left(-\frac{25}{51} \right) * \log_2 \left(\frac{25}{51} \right) \right) \\
 &= (-0,5098 * -0,9719) + (-0,4902 * -1,0286) \\
 &= 0,4955 + 0,5042 \\
 &= 0,9909
 \end{aligned}$$

$$\begin{aligned}
 \text{Entropy tahun 2014} &= \left(\left(-\frac{21}{40} \right) * \log_2 \left(\frac{21}{40} \right) \right) + \left(\left(-\frac{19}{40} \right) * \log_2 \left(\frac{19}{40} \right) \right) \\
 &= (-0,525 * -0,9296) + (-0,475 * -1,0740) \\
 &= 0,4880 + 0,5101 \\
 &= 0,9981
 \end{aligned}$$

$$\begin{aligned}
 \text{Entropy tahun 2013} &= \left(\left(-\frac{4}{11} \right) * \log_2 \left(\frac{4}{11} \right) \right) + \left(\left(-\frac{7}{11} \right) * \log_2 \left(\frac{7}{11} \right) \right) \\
 &= (-0,3636 * -1,4594) + (-0,6363 * -0,6520)
 \end{aligned}$$

$$= 0,5306 + 0,4149$$

$$= 0,9455$$

$$\text{Entropy sex laki-laki} = \left(\left(-\frac{15}{27} \right) * \log_2 \left(\frac{15}{27} \right) \right) + \left(\left(-\frac{12}{27} \right) * \log_2 \left(\frac{12}{27} \right) \right)$$

$$= (-0,5555 * -0,8479) + (-0,4444 * -1,1699)$$

$$= 0,4710 + 0,5194$$

$$= 0,9904$$

$$\text{Entropy sex perempuan} = \left(\left(-\frac{8}{19} \right) * \log_2 \left(\frac{8}{19} \right) \right) + \left(\left(-\frac{11}{19} \right) * \log_2 \left(\frac{11}{19} \right) \right)$$

$$= (-0,4210 * -1,2479) + (-0,5789 * -0,7884)$$

$$= 0,5253 + 0,4564$$

$$= 0,9817$$

$$\text{Entropy sekolah negeri} = \left(\left(-\frac{16}{31} \right) * \log_2 \left(\frac{16}{31} \right) \right) + \left(\left(-\frac{15}{31} \right) * \log_2 \left(\frac{15}{31} \right) \right)$$

$$= (-0,5161 * -0,9541) + (-0,4838 * -1,0473)$$

$$= 0,4924 + 0,5066$$

$$= 0,9990$$

$$\text{Entropy sekolah swasta} = \left(\left(-\frac{10}{20} \right) * \log_2 \left(\frac{10}{20} \right) \right) + \left(\left(-\frac{10}{20} \right) * \log_2 \left(\frac{10}{20} \right) \right)$$

$$= (-0,5 * -1) + (-0,5 * -1)$$

$$= 0,5 + 0,5$$

$$= 1$$

$$\text{Entropy provinsi jawa} = \left(\left(-\frac{25}{51} \right) * \log_2 \left(\frac{25}{51} \right) \right) + \left(\left(-\frac{26}{51} \right) * \log_2 \left(\frac{26}{51} \right) \right)$$

$$= (-0,4901 * -1,0285) + (-0,5098 * -0,9719)$$

$$\begin{aligned}
&= 0,5041 + 0,4955 \\
&= 0,9996 \\
\text{Entropy provinsi luar jawa} &= \left(\left(-\frac{0}{0} \right) * \log_2 \left(\frac{0}{0} \right) \right) + \left(\left(-\frac{0}{0} \right) * \log_2 \left(\frac{0}{0} \right) \right) \\
&= (-0 * -0 + (-0 * -0)) \\
&= 0 \\
\text{Entropy kota dekat} &= \left(\left(-\frac{13}{27} \right) * \log_2 \left(\frac{13}{27} \right) \right) + \left(\left(-\frac{14}{27} \right) * \log_2 \left(\frac{14}{27} \right) \right) \\
&= (-0,4814 * -1,0544) + (-0,5185 * -0,9475) \\
&= 0,5076 + 0,4912 \\
&= 0,9988 \\
\text{Entropy kota jauh} &= \left(\left(-\frac{13}{24} \right) * \log_2 \left(\frac{13}{24} \right) \right) + \left(\left(-\frac{11}{24} \right) * \log_2 \left(\frac{11}{24} \right) \right) \\
&= (-0,5416 * -0,8845) + (-0,4583 * -1,1255) \\
&= 0,4790 + 0,5158 \\
&= 0,9948 \\
\text{Entropy jurusan IPA} &= \left(\left(-\frac{5}{20} \right) * \log_2 \left(\frac{5}{20} \right) \right) + \left(\left(-\frac{15}{20} \right) * \log_2 \left(\frac{15}{20} \right) \right) \\
&= (-0,25 * -2) + (-0,75 * -0,4150) \\
&= 0,5 + 0,3112 \\
&= 0,8112 \\
\text{Entropy jurusan IPS} &= \left(\left(-\frac{17}{19} \right) * \log_2 \left(\frac{17}{19} \right) \right) + \left(\left(-\frac{2}{19} \right) * \log_2 \left(\frac{2}{19} \right) \right) \\
&= (-0,8947 * -0,1604) + (-0,1052 * -3,2479) \\
&= 0,1435 + 0,3416 \\
&= 0,4851
\end{aligned}$$

$$\begin{aligned}
\text{Entropy jurusan} &= \left(\left(-\frac{4}{12} \right) * \log_2 \left(\frac{4}{12} \right) \right) + \left(\left(-\frac{8}{12} \right) * \log_2 \left(\frac{8}{12} \right) \right) \\
\text{kejuruan} &= (-0,3333 * -1,5849) + (-0,6666 * -0,5849) \\
&= 0,5282 + 0,3899 \\
&= 0,9181
\end{aligned}$$

$$\begin{aligned}
\text{Entropy ijazah 2014} &= \left(\left(-\frac{18}{40} \right) * \log_2 \left(\frac{18}{40} \right) \right) + \left(\left(-\frac{22}{40} \right) * \log_2 \left(\frac{22}{40} \right) \right) \\
&= (-0,45 * -1,520) + (-0,55 * -0,8624) \\
&= 0,5184 + 0,4743 \\
&= 0,9927
\end{aligned}$$

$$\begin{aligned}
\text{Entropy ijazah 2013} &= \left(\left(-\frac{18}{40} \right) * \log_2 \left(\frac{18}{40} \right) \right) + \left(\left(-\frac{22}{40} \right) * \log_2 \left(\frac{22}{40} \right) \right) \\
&= (-0,45 * -1,520) + (-0,55 * -0,8624) \\
&= 0,5184 + 0,4743 \\
&= 0,9927
\end{aligned}$$

$$\begin{aligned}
\text{Entropy nilai besar} &= \left(\left(-\frac{21}{42} \right) * \log_2 \left(\frac{21}{42} \right) \right) + \left(\left(-\frac{21}{42} \right) * \log_2 \left(\frac{21}{42} \right) \right) \\
&= (-0,5 * -1) + (-0,5 * -1) \\
&= 0,5 + 0,5 \\
&= 1
\end{aligned}$$

$$\begin{aligned}
\text{Entropy nilai kecil} &= \left(\left(-\frac{4}{8} \right) * \log_2 \left(\frac{4}{8} \right) \right) + \left(\left(-\frac{4}{8} \right) * \log_2 \left(\frac{4}{8} \right) \right) \\
&= (-0,5 * -1) + (-0,5 * -1) \\
&= 1
\end{aligned}$$

7. Nilai gain dari setiap attribute:

$$\begin{aligned}
 \text{Gain (tahun)} &= 0,9009 - \left(\left(\frac{40}{51} * 0,9981 \right) + \left(\frac{11}{51} * 0,9455 \right) \right) \\
 &= 0,9009 - (0,7828 + 0,2039) \\
 &= 0,0042
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (sex)} &= 0,9009 - \left(\left(\frac{27}{51} * 0,9904 \right) + \left(\frac{19}{51} * 0,9817 \right) \right) \\
 &= 0,9009 - (0,5245 + 0,3657) \\
 &= 0,1005
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (sekolah)} &= 0,9009 - \left(\left(\frac{31}{51} * 0,9990 \right) + \left(\frac{20}{51} * 1 \right) \right) \\
 &= 0,9009 - (0,6072 + 0,3921) \\
 &= 0,0026
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (kota)} &= 0,9009 - \left(\left(\frac{27}{51} * 0,9988 \right) + \left(\frac{24}{51} * 0,9948 \right) \right) \\
 &= 0,9009 - (0,5287 + 0,4681) \\
 &= 0,9968
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (jurusan)} &= 0,9009 - \left(\left(\frac{20}{51} * 0,8112 \right) + \left(\frac{19}{51} * 0,4851 \right) + \left(\frac{12}{51} * 0,9181 \right) \right) \\
 &= 0,9009 - (0,3181 + 0,1807 + 0,2160) \\
 &= 0,2761
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (ijazah)} &= 0,9009 - \left(\left(\frac{40}{51} * 0,8112 \right) + \left(\frac{10}{51} * 0,4851 \right) + \left(\frac{0}{51} * 0 \right) \right) \\
 &= 0,9009 - (0,7785 + 0,1903 + 0) \\
 &= 0,0221
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (nilai)} &= 0,9009 - \left(\left(\frac{42}{51} * 1 \right) + \left(\frac{8}{51} * 1 \right) \right) \\
 &= 0,9009 - (0,8235 + 0,1568) \\
 &= 0,0106
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain (provinsi)} &= 0,9009 - \left(\left(\frac{51}{51} * 0,9996 \right) + \left(\frac{0}{51} * 0 \right) \right) \\
 &= 0,9009 - (0,9996 + 0) \\
 &= 0,0087
 \end{aligned}$$

Hasil perhitungan entropy dan gain dari masing masing atribut menghasilkan table seperti berikut :

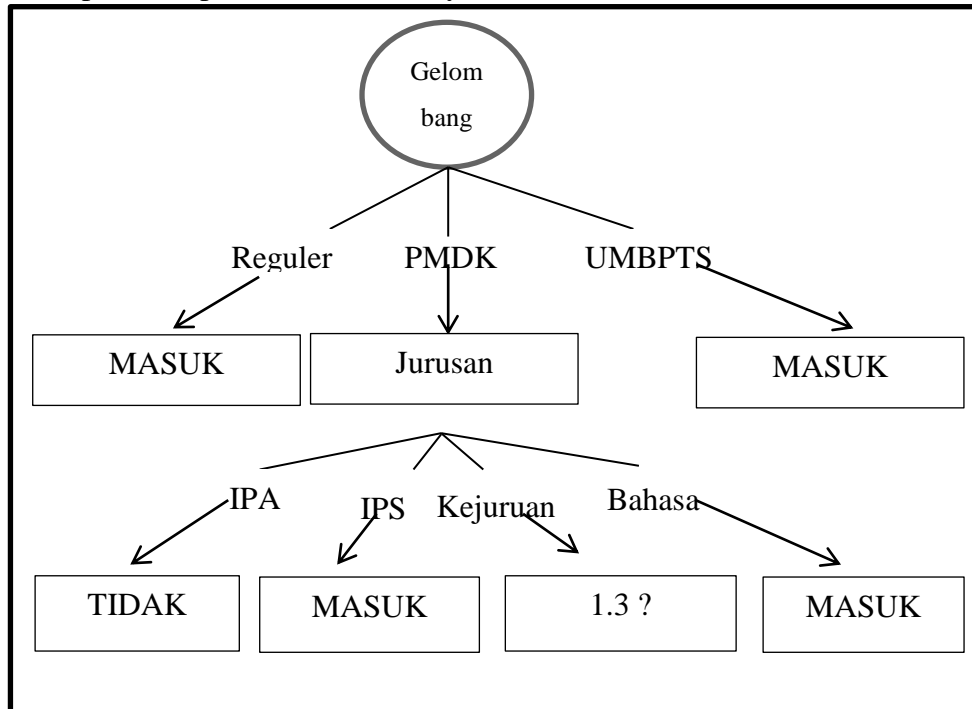
Tabel 4.2 Tabel untuk Membentuk Node 2

Atribut		Banyak Kasus	Masuk	Tidak masuk	Entropy	Gain
PMDK		51	26	25	0,9009	
TAHUN						0,0042
	2014	40	21	19	0,9981	
	2013	11	4	7	0,9455	
Sex						0,1005
	Laki-laki	27	15	12	0,9904	
	Perempuan	19	8	11	0,9817	
Sekolah						0,0026
	Negeri	31	16	15	0,9990	
	Swasta	20	10	10	1	
Provinsi						0,0087
	Jawa	51	25	26	0,9996	
	Luar jawa	0	0	0	0	
Kota						0,0059
	Dekat	27	13	14	0,9988	
	Jauh	24	13	11	0,9948	

Atribut		Banyak Kasus	Masuk	Tidak masuk	Entropy	Gain
Jurusan						0,2761
	IPS	20	5	15	0,8112	
	IPA	19	17	2	0,4851	
	Kejuruan	12	4	8	0,9181	
	Bahasa	0	0	0	0	
Ijazah						0,0221
	2014	40	18	22	0,9927	
	2013	10	4	6	0,9708	
	2012	1	0	1	0	
Nilai UN						0,0106
	Besar	42	21	21	1	
	Kecil	8	4	4	1	
					MAX =	0,2761

Dari hasil perhitungan pada tabel diatas, diketahui bahwa nilai Gain terbesar yaitu sebesar 0,2761, Sehingga atribut **jurusan** menjadi **node cabang** dari gelombang = PMDK. Pada *attribute* jurusan = kejuruan dengan nilai *entropy* terbesar = 0,9181 maka dilakukan perhitungan untuk mencari nilai *Gain*, dan *attribute* dengan nilai *Gain* terbesar, maka akan menjadi *node cabang* dari nilai *attribute* jurusan = **kejuruan**

8. Maka pohon keputusan sementara yaitu:



Gambar 4.9 Pohon Keputusan Node 2

Perhitungan dilakukan dengan mencari nilai dari *attribute* selain yang menjadi *node* akar (**Jurusan**), yaitu dengan mencari jumlah kasus untuk hasil Masuk, Tidak masuk, dan nilai *Entropy* dari semua kasus saat **gelombang = PMDK dan Jurusan = Kejuruan**. Dan kemudian melakukan perhitungan untuk mendapatkan nilai *Gain*, dan *attribute* dengan nilai *Gain* terbesar, maka akan menjadi *node* cabang dari nilai *attribute* **Kejuruan**.

Node 1.3:

Jumlah data dengan Gelombang = PMDK dan jurusan = kejuruan : 12

Jumlah data dengan Gelombang = PMDK dan jurusan = kejuruan

Hasil yang masuk : 4

Jumlah data dengan Gelombang = PMDK dan jurusan = kejuruan

Hasil yang tidak masuk : 8

$$\begin{aligned}
 \text{Entropy gelombang} &= \left(\left(-\frac{4}{12} \right) * \log_2 \left(\frac{4}{12} \right) \right) + \left(\left(-\frac{8}{12} \right) * \log_2 \left(\frac{8}{12} \right) \right) \\
 \text{PMDK dan kejuruan} &= (-0,3333 * -1,584) + (-0,666 * -0,5849) \\
 &= 0,5282 + 0,3899 \\
 &= 0,9181
 \end{aligned}$$

9. Melakukan penghitungan nilai entropy pada setiap attribute kecuali pada attribute gelombang dan jurusan:

$$\begin{aligned}
 \text{Entropy tahun 2014} &= \left(\left(-\frac{3}{10} \right) * \log_2 \left(\frac{3}{10} \right) \right) + \left(\left(-\frac{7}{10} \right) * \log_2 \left(\frac{7}{10} \right) \right) \\
 &= (-0,3 * -1,7369) + (-0,7 * -0,5145) \\
 &= 0,5210 + 0,3606 \\
 &= 0,8812
 \end{aligned}$$

$$\begin{aligned}
 \text{Entropy tahun 2013} &= \left(\left(-\frac{1}{2} \right) * \log_2 \left(\frac{1}{2} \right) \right) + \left(\left(-\frac{1}{2} \right) * \log_2 \left(\frac{1}{2} \right) \right) \\
 &= (-0,5 * -1) + (-0,5 * -1) \\
 &= 0,5 + 0,5 \\
 &= 1
 \end{aligned}$$

$$\begin{aligned}
\text{Entropy sex laki-laki} &= \left(\left(-\frac{3}{8} \right) * \log_2 \left(\frac{3}{8} \right) \right) + \left(\left(-\frac{5}{8} \right) * \log_2 \left(\frac{5}{8} \right) \right) \\
&= (-0,375 * -1,4150) + (-0,625 * -0,6780) \\
&= 0,5306 + 0,4237 \\
&= 0,9543
\end{aligned}$$

$$\begin{aligned}
\text{Entropy sex} &= \left(\left(-\frac{1}{4} \right) * \log_2 \left(\frac{1}{4} \right) \right) + \left(\left(-\frac{3}{4} \right) * \log_2 \left(\frac{3}{4} \right) \right) \\
\text{Perempuan} &= (-0,25 * -2) + (-0,75 * -0,4150) \\
&= 0,5 + 0,3112 \\
&= 0,8112
\end{aligned}$$

$$\begin{aligned}
\text{Entropy sekolah} &= \left(\left(-\frac{1}{7} \right) * \log_2 \left(\frac{1}{7} \right) \right) + \left(\left(-\frac{6}{7} \right) * \log_2 \left(\frac{6}{7} \right) \right) \\
\text{swasta} &= (-0,1428 * -2,803) + (-0,8571 * -0,2223) \\
&= 0,4008 + 0,1906 \\
&= 0,5914
\end{aligned}$$

$$\begin{aligned}
\text{Entropy sekolah} &= \left(\left(-\frac{3}{5} \right) * \log_2 \left(\frac{3}{5} \right) \right) + \left(\left(-\frac{2}{5} \right) * \log_2 \left(\frac{2}{5} \right) \right) \\
\text{negeri} &= (-0,6 * -0,7369) + (-0,4 * -1,3219) \\
&= 0,4421 + 0,5287 \\
&= 0,9708
\end{aligned}$$

$$\begin{aligned}
\text{Entropy provinsi} &= \left(\left(-\frac{4}{12} \right) * \log_2 \left(\frac{4}{12} \right) \right) + \left(\left(-\frac{8}{12} \right) * \log_2 \left(\frac{8}{12} \right) \right) \\
\text{jawa} &= (-0,3333 * -1,5849) + (-0,6666 * -0,5849) \\
&= 0,5282 + 0,3899
\end{aligned}$$

$$\begin{aligned}
&= 0,9181 \\
\text{Entropy kota jauh} &= \left(\left(-\frac{3}{5} \right) * \log_2 \left(\frac{3}{5} \right) \right) + \left(\left(-\frac{2}{5} \right) * \log_2 \left(\frac{2}{5} \right) \right) \\
&= (-0,6 * -0,7369) + (-0,4 * -1,3219) \\
&= 0,4421 + 0,5287 \\
&= 0,9708 \\
\text{Entropy kota dekat} &= \left(\left(-\frac{1}{7} \right) * \log_2 \left(\frac{1}{7} \right) \right) + \left(\left(-\frac{6}{7} \right) * \log_2 \left(\frac{6}{7} \right) \right) \\
&= (-0,1428 * -2,8073) + (-0,8571 * -0,2223) \\
&= 0,4008 + 0,1906 \\
&= 0,5914 \\
\text{Entropy ijazah 2014} &= \left(\left(-\frac{3}{10} \right) * \log_2 \left(\frac{3}{10} \right) \right) + \left(\left(-\frac{7}{10} \right) * \log_2 \left(\frac{7}{10} \right) \right) \\
&= (-0,3 * -1,7369) + (-0,7 * -0,5147) \\
&= 0,5210 + 0,5602 \\
&= 0,8812 \\
\text{Entropy nilai besar} &= \left(\left(-\frac{4}{11} \right) * \log_2 \left(\frac{4}{11} \right) \right) + \left(\left(-\frac{7}{11} \right) * \log_2 \left(\frac{7}{11} \right) \right) \\
&= (-0,3636 * -1,4594) + (-0,6363 * -0,6520) \\
&= 0,5306 + 0,4149 \\
&= 0,9455
\end{aligned}$$

10. Nilai gain dari setiap attribute:

$$\begin{aligned}
\text{Gain (tahun)} &= 0,9181 - \left(\left(\frac{10}{12} * 0,8812 \right) + \left(\frac{2}{12} * 1 \right) \right) \\
&= 0,9181 - (0,7343 + 0,1666) \\
&= 0,0171
\end{aligned}$$

$$\begin{aligned}
\text{Gain (sex)} &= 0,9181 - \left(\left(\frac{8}{12} * 0,9543 \right) + \left(\frac{4}{12} * 0,8112 \right) \right) \\
&= 0,9181 - (0,6361 + 0,2703) \\
&= 0,0116
\end{aligned}$$

$$\begin{aligned}
\text{Gain (sekolah)} &= 0,9181 - \left(\left(\frac{7}{12} * 0,5914 \right) + \left(\frac{5}{12} * 0,9708 \right) \right) \\
&= 0,9181 - (0,3449 + 0,4045) \\
&= 0,1686
\end{aligned}$$

$$\begin{aligned}
\text{Gain (provinsi)} &= 0,9181 - \left(\left(\frac{12}{12} * 0,9181 \right) + \left(\frac{0}{12} * 0 \right) \right) \\
&= 0,9181 - (0,9181 + 0) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\text{Gain (kota)} &= 0,9181 - \left(\left(\frac{5}{12} * 0,9708 \right) + \left(\frac{7}{12} * 0,5914 \right) \right) \\
&= 0,9181 - (0,4045 + 0,3449) \\
&= 0,1686
\end{aligned}$$

$$\begin{aligned}
\text{Gain (ijazah)} &= 0,9181 - \left(\left(\frac{10}{12} * 0,8812 \right) + \left(\frac{2}{12} * 1 \right) \right) \\
&= 0,9181 - (0,7343 + 0,2) \\
&= 0,0162
\end{aligned}$$

$$\begin{aligned}
\text{Gain (nilai)} &= 0,9181 - \left(\left(\frac{11}{12} * 0,9455 \right) + \left(\frac{1}{12} * 0 \right) \right) \\
&= 0,9181 - (0,8667 + 0) \\
&= 0,0513
\end{aligned}$$

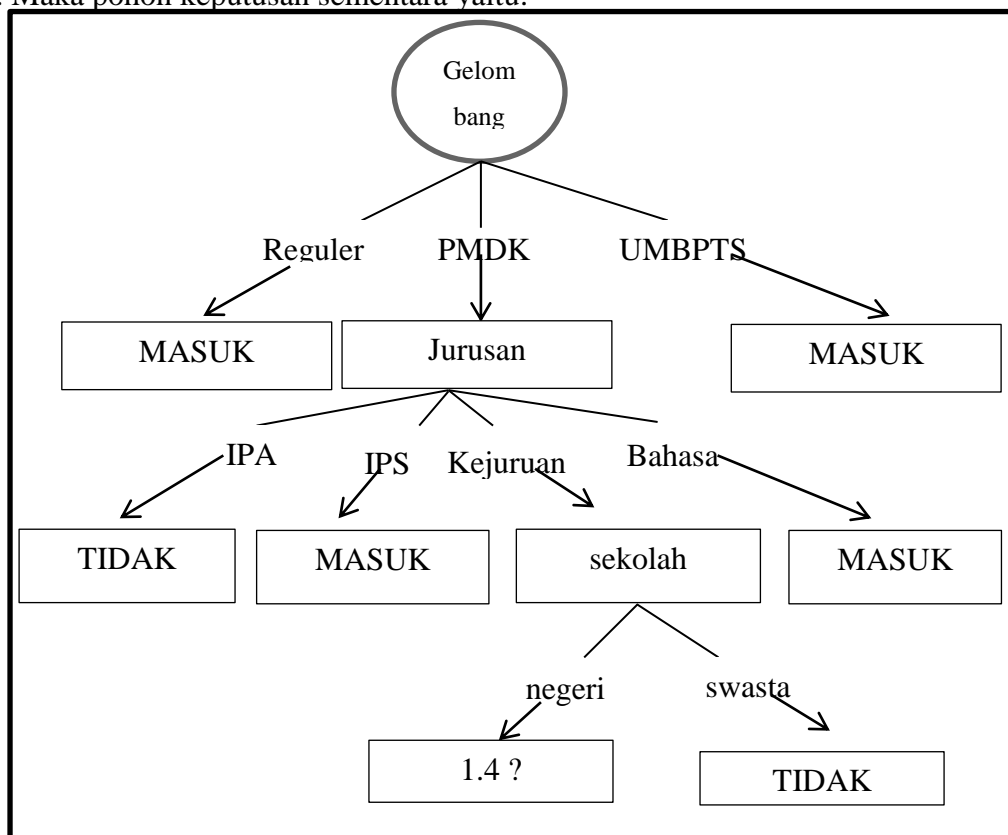
Hasil perhitungan *entropy* dan *gain* dari masing masing *attribute* menghasilkan tabel seperti berikut :

Tabel 4.3 Tabel untuk Membentuk Node 3

Atribut		Banyak kasus	Masuk	Tidak masuk	Entropy	Gain
PMDK dan kejuruan		12	4	8	0,9181	
Tahun						0,0171
	2014	10	3	7	0,8812	
	2013	2	1	1	1	
Sex						0,0116
	Laki-laki	8	3	5	0,9543	
	Perempuan	4	1	3	0,8812	
Sekolah						0,1686
	Negeri	5	3	2	0,9708	
	Swasta	7	1	6	0,5914	
Provinsi						0
	Jawa	12	4	8	0,9181	
	Luar jawa	0	0	0	0	
Kota						0,1686
	Dekat	7	1	6	0,5914	
	Jauh	5	3	2	0,9708	
Ijazah						0,0162
	2014	10	3	7	0,8812	
	2013	2	1	1	1	
	2012	0	0	0	0	
Nilai UN						0,0513
	Besar	11	4	7	0,9455	
	Kecil	1	0	1	0	
					MAX =	0,1686

Dari hasil perhitungan pada tabel diatas, diketahui bahwa terdapat nilai *Gain* yang memiliki nilai *gain* yang sama dan merupakan *gain* terbesar yaitu *attribute* kota dan sekolah dengan *gain* 0,1686 , Sehingga *attribute* kota dan sekolah dapat dipilih salah satu sebagai cabang. Dengan itu sekolah dipilih sebagai cabang selanjutnya. Pada *attribute* sekolah = negeri dengan nilai *entropy* terbesar = 0,9708 maka dilakukan perhitungan kembali yang akan menjadi *node* cabang dari nilai atribut sekolah = **negeri**.

11. Maka pohon keputusan sementara yaitu:



Gambar 4.10 Pohon Keputusan Node 3

Perhitungan dilakukan dengan mencari nilai dari *attribute* selain yang menjadi *node* akar (**Sekolah**), yaitu dengan mencari jumlah kasus untuk hasil Masuk, Tidak masuk, dan nilai *Entropy* dari semua kasus saat **gelombang = PMDK dan Jurusan = Kejuruan dan Sekolah =Negeri**. Dan kemudian melakukan perhitungan untuk mendapatkan nilai *Gain*, dan *attribute* dengan nilai *Gain* terbesar, maka akan menjadi *node* cabang dari nilai *attribute* **Negeri**.

Node 1.4:

Jumlah data dengan

Gelombang = PMDK, jurusan=kejuruan, sekolah = negeri : 5

Jumlah data Gelombang = PMDK, jurusan=kejuruan,
sekolah = negeri yang hasil Masuk : 3

Jumlah data dengan Gelombang = PMDK, jurusan=kejuruan,
sekolah = negeri hasil Tidak Masuk : 2

Entropy gelombang = $\left(\left(-\frac{3}{5} \right) * \log_2 \left(\frac{3}{5} \right) \right) + \left(\left(-\frac{2}{5} \right) * \log_2 \left(\frac{2}{5} \right) \right)$
PMDK kejuruan dan negeri

$$\begin{aligned} &= (-0,6 * -0,7369) + (-0,4 * -1,321) \\ &= 0,4421 + 0,5287 \\ &= 0,9708 \end{aligned}$$

12. Melakukan penghitungan nilai entropy pada setiap attribute kecuali pada attribute gelombang dan jurusan:

$$\begin{aligned} \text{Entropy tahun 2014} &= \left(\left(-\frac{2}{3} \right) * \log_2 \left(\frac{2}{3} \right) \right) + \left(\left(-\frac{1}{3} \right) * \log_2 \left(\frac{1}{3} \right) \right) \\ &= (-0,6666 * -0,5849) + (-0,3333 * -1,5849) \\ &= 0,3899 + 0,5282 \\ &= 0,9181 \end{aligned}$$

$$\begin{aligned} \text{Entropy tahun 2013} &= \left(\left(-\frac{1}{2} \right) * \log_2 \left(\frac{1}{2} \right) \right) + \left(\left(-\frac{1}{2} \right) * \log_2 \left(\frac{1}{2} \right) \right) \\ &= (-0,5 * -1) + (-0,5 * -1) \\ &= 1 \end{aligned}$$

$$\begin{aligned} \text{Entropy sex laki-laki} &= \left(\left(-\frac{2}{3} \right) * \log_2 \left(\frac{2}{3} \right) \right) + \left(\left(-\frac{1}{3} \right) * \log_2 \left(\frac{1}{3} \right) \right) \\ &= (-0,6666 * -0,5849) + (-0,3333 * -1,5849) \\ &= 0,3899 + 0,5282 \\ &= 0,9181 \end{aligned}$$

$$\begin{aligned} \text{Entropy sex Perempuan} &= \left(\left(-\frac{1}{2} \right) * \log_2 \left(\frac{1}{2} \right) \right) + \left(\left(-\frac{1}{2} \right) * \log_2 \left(\frac{1}{2} \right) \right) \\ &= (-0,5 * -1) + (-0,5 * -1) \end{aligned}$$

$$\begin{aligned}
&= 0,5+0,5 \\
&= 1 \\
\text{Entropy kota jauh} &= \left(\left(-\frac{2}{2} \right) * \log_2 \left(\frac{2}{2} \right) \right) + \left(\left(-\frac{0}{2} \right) * \log_2 \left(\frac{0}{2} \right) \right) \\
&= (-1 * -0) + (-0 * -0) \\
&= 0 \\
\text{Entropy provinsi jawa} &= \left(\left(-\frac{3}{5} \right) * \log_2 \left(\frac{3}{5} \right) \right) + \left(\left(-\frac{2}{5} \right) * \log_2 \left(\frac{2}{5} \right) \right) \\
&= (-0,6 * -0,7369) + (-0,4 * -1,321) \\
&= 0,4421 + 0,5287 \\
&= 0,9708 \\
\text{Entropy provinsi luar} &= \left(\left(-\frac{0}{0} \right) * \log_2 \left(\frac{0}{0} \right) \right) + \left(\left(-\frac{0}{0} \right) * \log_2 \left(\frac{0}{0} \right) \right) \\
\text{jawa} &= 0 \\
\text{Entropy ijazah 2014} &= \left(\left(-\frac{2}{3} \right) * \log_2 \left(\frac{2}{3} \right) \right) + \left(\left(-\frac{1}{3} \right) * \log_2 \left(\frac{1}{3} \right) \right) \\
&= (-0,6666 * -0,5849) + (-0,3333 * -1,5849) \\
&= 0,3899 + 0,5282 \\
&= 0,9181 \\
\text{Entropy ijazah 2013} &= \left(\left(-\frac{1}{2} \right) * \log_2 \left(\frac{1}{2} \right) \right) + \left(\left(-\frac{1}{2} \right) * \log_2 \left(\frac{1}{2} \right) \right) \\
&= (-0,6666 * -0,5849) + (-0,3333 * -1,5849) \\
&= 0,5 + 0,5 \\
&= 1 \\
\text{Entropy nilai besar} &= \left(\left(-\frac{3}{4} \right) * \log_2 \left(\frac{3}{4} \right) \right) + \left(\left(-\frac{1}{4} \right) * \log_2 \left(\frac{1}{4} \right) \right) \\
&= (-0,75 * -0,4150) + (-0,25 * -2) \\
&= 0,3112 + 0,5 \\
&= 0,8112 \\
\text{Entropy nilai nilai} &= \left(\left(-\frac{0}{1} \right) * \log_2 \left(\frac{0}{1} \right) \right) + \left(\left(-\frac{1}{1} \right) * \log_2 \left(\frac{1}{1} \right) \right) \\
&= (-0 * -0) + (-1 * -0)
\end{aligned}$$

$$= 0$$

13. Nilai gain dari setiap attribute:

$$\text{Gain (tahun)} = 0,9708 - \left(\left(\frac{3}{5} * 0,9180 \right) + \left(\frac{2}{5} * 1 \right) \right)$$

$$= 0,9708 - (0,5508 + 0,4)$$

$$= 0,02$$

$$\text{Gain (ijazah)} = 0,9708 - \left(\left(\frac{3}{5} * 0,9180 \right) + \left(\frac{2}{5} * 1 \right) \right)$$

$$= 0,9708 - (0,5508 + 0,4)$$

$$= 0,02$$

$$\text{Gain (sex)} = 0,9708 - \left(\left(\frac{3}{5} * 0,9180 \right) + \left(\frac{2}{5} * 1 \right) \right)$$

$$= 0,9708 - (0,5508 + 0,4)$$

$$= 0,02$$

$$\text{Gain (nilai)} = 0,9708 - \left(\left(\frac{4}{5} * 0,8112 \right) + \left(\frac{1}{5} * 0 \right) \right)$$

$$= 0,9708 - (0,6489 + 0)$$

$$= 0,3219$$

$$\text{Gain (kota)} = 0,9708 - \left(\left(\frac{3}{5} * 0,9180 \right) + \left(\frac{2}{5} * 0 \right) \right)$$

$$= 0,9708 - (0,5508 + 0)$$

$$= 0,42$$

$$\text{Gain (provinsi)} = 0,9708 - \left(\left(\frac{5}{5} * 0,9708 \right) + \left(\frac{0}{5} * 0 \right) \right)$$

$$= 0,9708 - (0,9708 + 0)$$

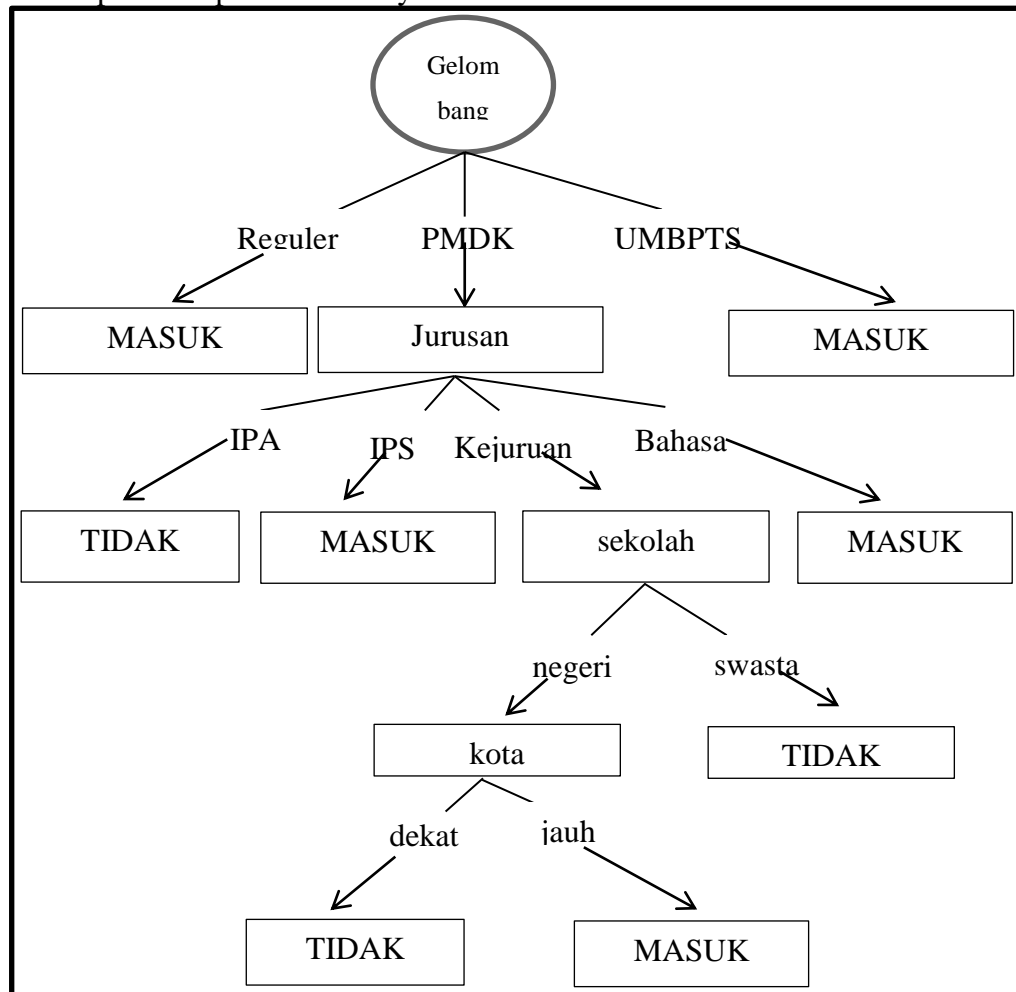
$$= 0$$

Hasil perhitungan entropy dan gain dari masing masing atribut menghasilkan table seperti berikut :

Tabel 4.4 Tabel Untuk Membentuk Node 4

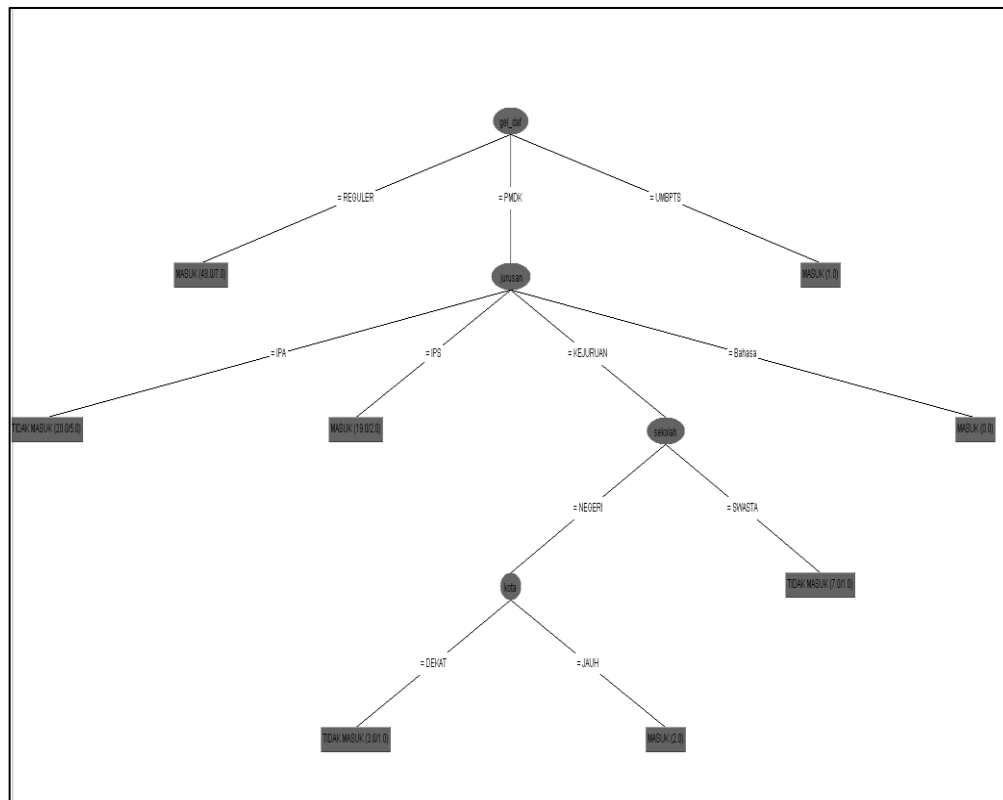
Atribut		Banyak kasus	Masuk	Tidak masuk	Entropy	Gain
PMDK , kejuruan dan Negeri		5	3	2	0,9708	
Tahun						0,02
	2014	3	2	1	0,9181	
	2013	2	1	1	1	
Sex						0,02
	Laki-laki	3	2	1	0,9181	
	Perempuan	2	1	1	1	
Provinsi						0
	Jawa	5	3	2	0,9708	
	Luar jawa	0	0	0	0	
Kota						0,4200
	Dekat	3	1	2	0,9181	
	Jauh	2	2	0	0	
Ijazah						0,02
	2014	3	2	1	0,9181	
	2013	2	1	1	1	
	2012	0	0	0	0	
Nilai UN						0,3219
	Besar	4	3	1	0,8112	
	Kecil	1	0	1	0	
					MAX =	0,4200

14. Maka pohon keputusan akhir yaitu:



Gambar 4.11 Gambar Akhir Pohon Keputusan

15. Hasil dari pengujian menggunakan aplikasi WEKA dengan jumlah 100 sample data calon mahasiswa yang memilih program studi system informasi tahun 2013-2014 yang dijadikan data training dan 100 sample data calon mahasiswa yang mendaftar pada program studi system informasi tahun 2015 menjadi data testing. Membentuk pohon keputusan seperti gambar dibawah ini :



Gambar 4.12 Pohon Keputusan Menggunakan WEKA

16. Pola pohon keputusan
- Jika gelombang = regular maka masuk sistem informasi
 - Jika gelombang = UMBPTS maka Masuk sistem informasi
 - Jika gelombang = PMDK dan jurusan = IPS maka masuk sistem informasi
 - Jika gelombang = PMDK dan jurusan = Bahasa maka masuk sistem informasi
 - Jika gelombang = PMDK dan jurusan = kejuruan dan sekolah = negeri dan kota = jauh maka masuk sistem informasi

- f. Jika gelombang = PMDK dan jurusan = IPA maka tidak masuk sistem informasi
 - g. Jika gelombang = PMDK dan jurusan = kejuruan dan sekolah = swasta maka tidak masuk sistem informasi
 - h. Jika gelombang = PMDK dan jurusan = kejuruan dan sekolah = negeri dan kota = dekat maka tidak masuk sistem informasi
17. Dari uji coba 100 sample data training yang kemudian dilakukan uji coba dengan sample 100 sample data testing yang dilakukan menggunakan aplikasi WEKA 3.8 memiliki keakuratan 80,61% dengan perhitungan menggunakan confusion matrix yaitu:

$$Akurasi = \frac{F11 + F10}{F11 + F10 + F01 + F00} \quad (4.2)$$

$$Precision = \frac{F11}{F01 + F11} \quad (4.3)$$

$$Recall = \frac{F11}{F10 + F11} \quad (4.4)$$

$$\begin{aligned}
 \text{Akurasi} &= \frac{38 + 41}{38 + 41 + 14 + 5} \\
 &= 78/98 * 100\% \\
 &= 80,61\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Precision} &= \frac{38}{38 + 5} \\
 \text{(tidak masuk)} &= 38/43 \\
 &= 0,884
 \end{aligned}$$

$$\begin{aligned}
 \text{Precision} &= \frac{41}{41 + 14} \\
 \text{(masuk)} &= 41/55 \\
 &= 0,745
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{41}{5 + 41} \\
 \text{(masuk)} &= 41/46 \\
 &= 0,891
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall (tidak masuk)} &= \frac{38}{14 + 38} \\
 &= 38/52 \\
 &= 0,731
 \end{aligned}$$

18. Hasil penghitungan keakuratan , precision dan recall secara manual akurat dengan penghitungan pada aplikasi WEKA 3.8 Terlihat pada gambar dibawah :

```

=== Summary ===

Correctly Classified Instances      79           80.6122 %
Incorrectly Classified Instances    19           19.3878 %
Kappa statistic                    0.6151
Mean absolute error                 0.3019
Root mean squared error             0.4086
Relative absolute error             59.1012 %
Root relative squared error         75.621 %
Total Number of Instances          98
Ignored Class Unknown Instances     1

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      0.731    0.109    0.884     0.731    0.800      0.626    0.818     0.803    TIDAK MASUK
      0.891    0.269    0.745     0.891    0.812      0.626    0.804     0.718    MASUK
Weighted Avg.   0.806    0.184    0.819     0.806    0.806      0.626    0.811     0.763

=== Confusion Matrix ===

  a  b  <-- classified as
38 14 | a = TIDAK MASUK
 5 41 | b = MASUK

```

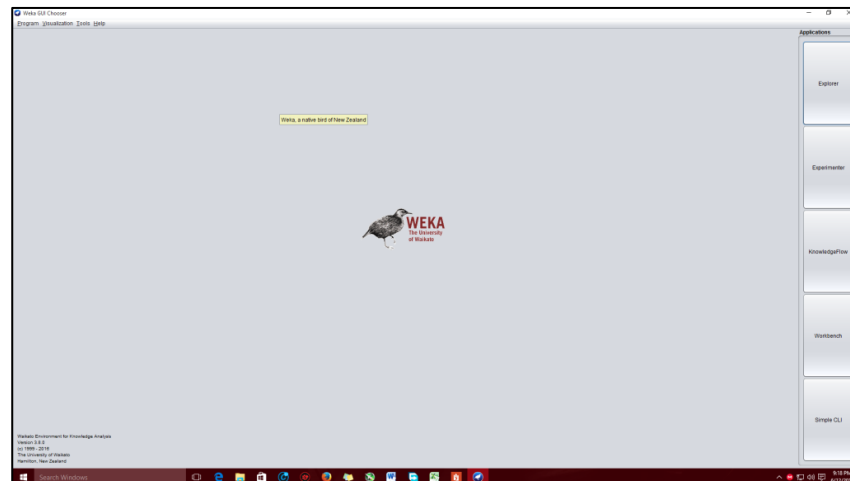
Gambar 4.13 Hasil Pengujian *Sample Data Training* dengan *Data Testing*

4.6 Implementasi menggunakan Aplikasi WEKA 3.8

Pada tahapan implementasi data yang digunakan merupakan seluruh data training dari data calon mahasiswa yang mendaftar pada program studi Sistem Informasi tahun 2013-2014 dan data testing merupakan data calon mahasiswa yang pada program studi Sistem Informasi tahun 2015. Data tersebut telah melewati proses dari tahapan KDD yaitu *preprocessing* dan proses *cleaning* yang kemudian akan digunakan untuk membangun pohon keputusan / decision tree pada WEKA 3.8

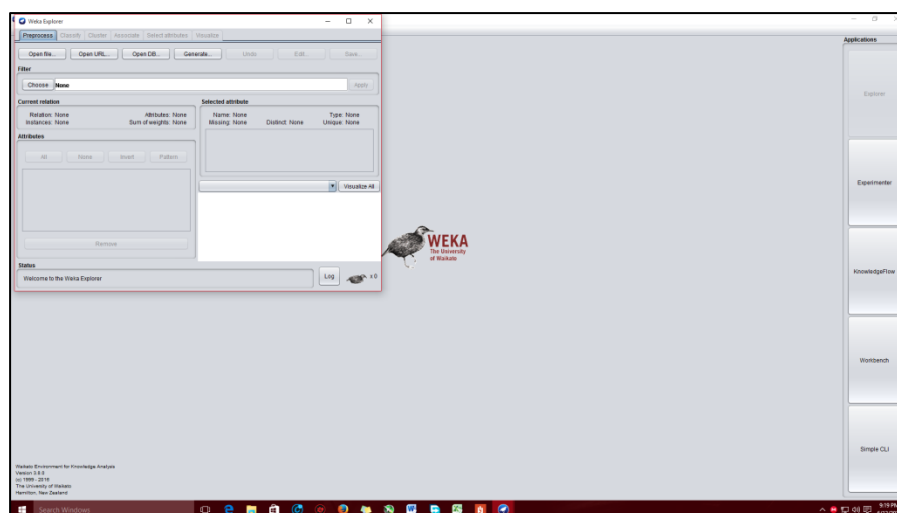
Langkah-langkah implementasi seperti berikut:

1. Buka Aplikasi WEKA 3.8 yang akan berjalan pada WEKA Java



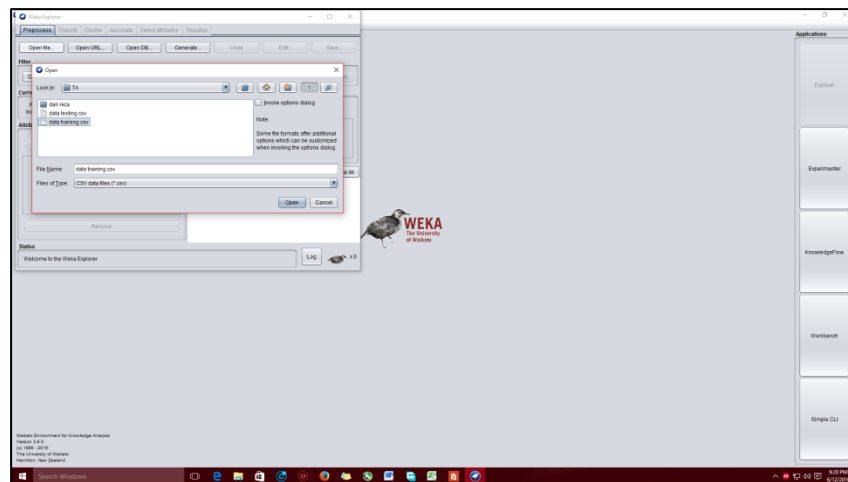
Gambar 4.14 Tampilan Awal WEKA

2. Klik *explore* kemudian jendela baru akan terbuka terdapat bar yang berisi preproses, klasifikasi, cluster dan lainnya. Pilih *preprocessing* terlebih dahulu kemudian klik *open file*.



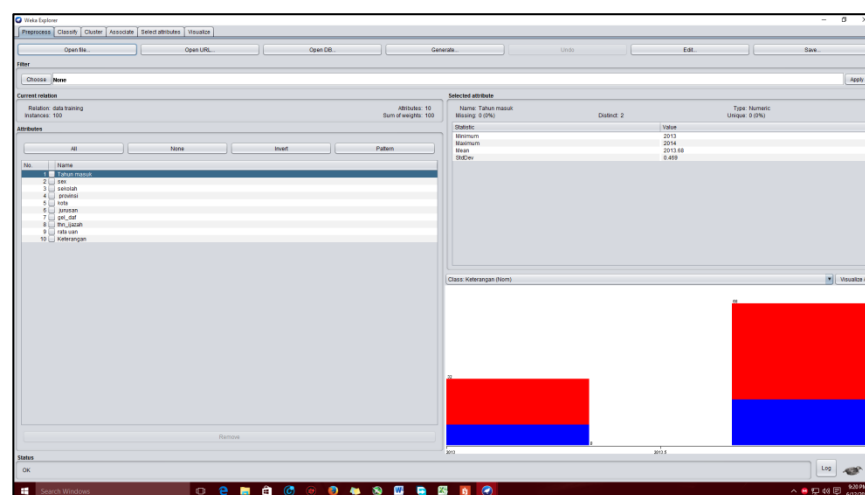
Gambar 4.15 Langkah *Preprocessing*

- Setelah *open file* diklik maka akan muncul jendela baru cari file data training dengan format cmv. Data tersebut telah dilakukan proses *cleaning* dan transformasi dengan isi keseluruhan data *training* berjumlah 505 data dari 2013-2014. Kemudian tekan *open*.



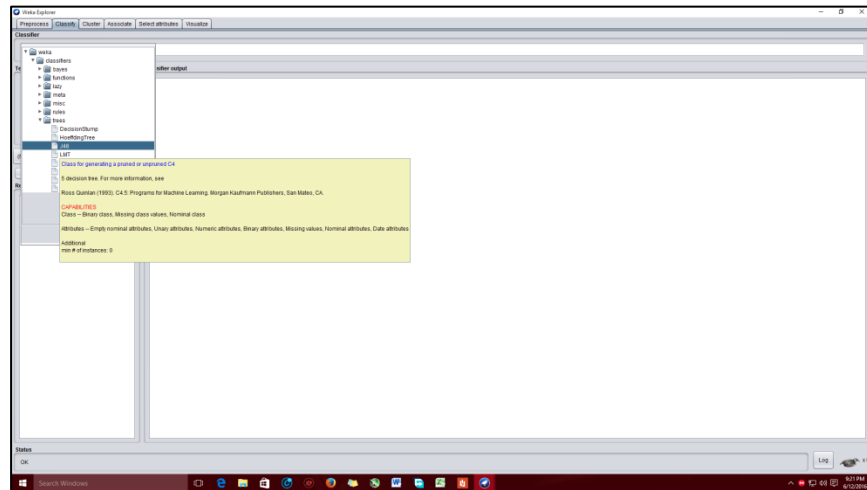
Gambar 4.16 Proses Mencari Data

- Kemudian data pada *file* akan dibaca oleh aplikasi WEKA seperti gambar dibawah ini:



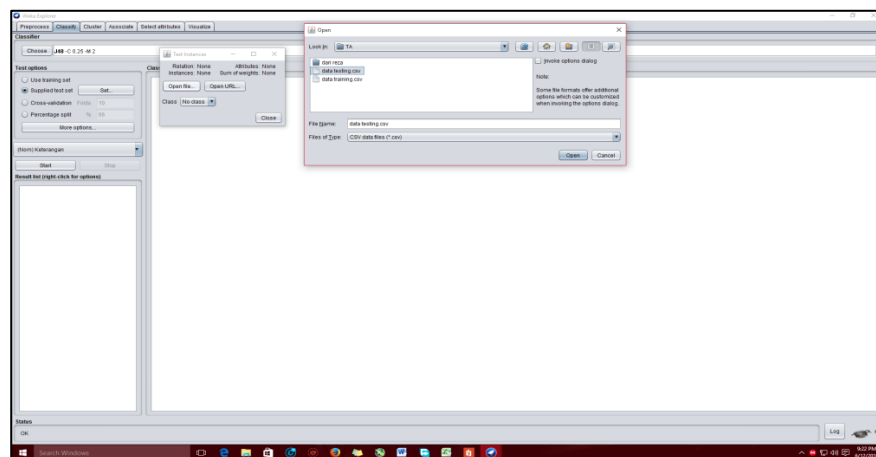
Gambar 4.17 Data dibaca Oleh WEKA

5. Pilih *classify* pada bar diatas maka akan masuk pada jendela *classify* yang terdapat bar algoritma, *test option*, *result*. Pilih *classifier* klik choose cari algoritma C.45 pada sub bagian *tree* (pada WEKA C45 yaitu J48).



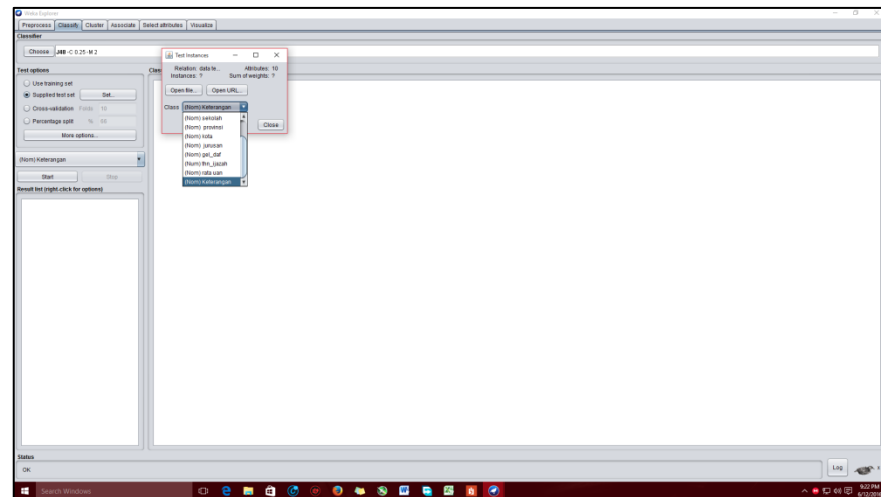
Gambar 4.18 Proses Data Mining Algoritma C4.5 di WEKA

6. Setelah *classifier* terisi dengan J48 pada bagian *test option* pilih sub supplied *test set* klik *set*. Kemudian akan muncul jendela baru untuk memilih data yang digunakan untuk testing. Open file cari data dengan nama “data testing.csv” *format attribute* dari data *testing* harus sma seperti data *training*. Data *testing* berisi keseluruhan data calon mahasiswa tahun 2015.



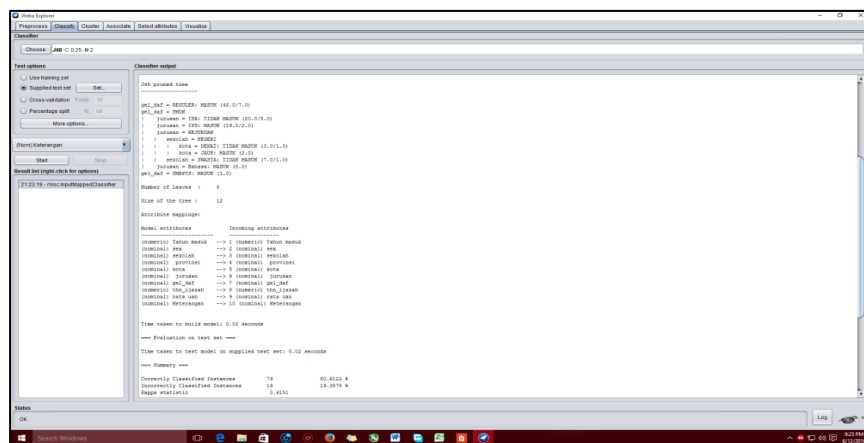
Gambar 4.19 Memasukan Data Testing

7. Pilih *class* yang akan menjadi variabel target dari seluruh *attribute* yang ada .
pilih *class / attribute* keterangan lalu *close*



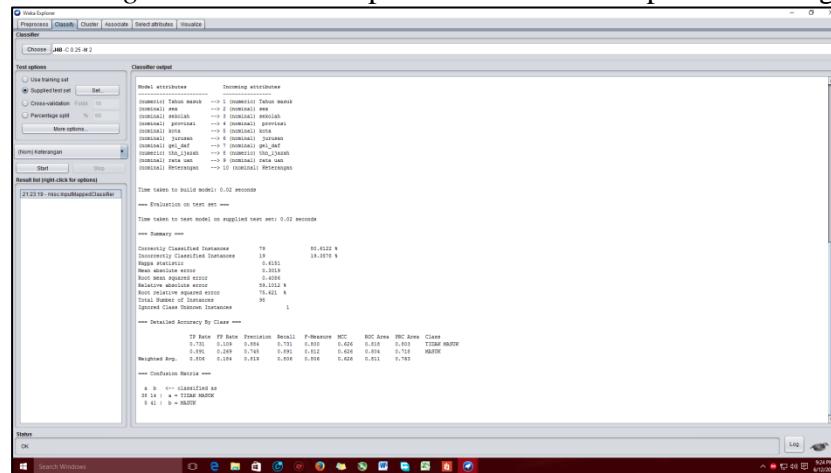
Gambar 4.20 Memilih Class/Atribut

8. Setelah semuanya siap dari *data training.cmv* dan *data testing.cmv* tekan start yang berada diatas result.



Gambar 4.21 Memulai Proses Data Mining

9. Maka akan menghasilkan hasil seperti gambar dibawah ini. Jika data *training* dan data *testing* cocok maka setiap *attribute* tidak terdapat data hilang.



Gambar 4.22 Hasil Data Mining Algoritma C4.5

4.7 Interpretation/evaluation

Hasil keseluruhan dari data set calon mahasiswa tahun 2013 hingga tahun 2105 dengan melakukan pengujian terhadap validitas yang memiliki tujuan untuk mengetahui apakah solusi dari pembentukan pohon keputusan tersebut telah valid ataukah tidak.

Data set tersebut dibagi menjadi 2 yaitu *data training* berisi calon mahasiswa yang mendaftar pada program studi Sistem Informasi pada tahun 2013 dan 2014 dan *data testing* berisi calon mahasiswa yang mendaftar pada program studi sistem informasi pada tahun 2015. *Data training* dan *data testing* akan dilakukan pengujian 3 kali dengan menggunakan aplikasi WEKA 3.8 dan memiliki prosentase sebesar :

- 1 *Data training* 70% dan *data testing* 30%
- 2 *Data training* 80% dan *data testing* 20%
- 3 *Data training* 90% dan *data testing* 10%

Pengujian pertama dengan *Data training* 70% dan *data testing* 30%

Tabel 4.5 Pengujian Pertama

<i>Data training</i>	<i>Data testing</i>	Akurasi	Precision	Recall
70%	30%	68,371%	0,765	0,684

```

Attribute mappings:

Model attributes          Incoming attributes
-----
(numeric) Tahun_masuk    --> 1 (numeric) Tahun_masuk
(nominal) sex             --> 2 (nominal) sex
(nominal) nama_sekolah    --> 3 (nominal) nama_sekolah
(nominal) provinsi       --> 4 (nominal) provinsi
(nominal) kota            --> 5 (nominal) kota
(nominal) jurusan        --> 6 (nominal) jurusan
(nominal) gel_daf         --> 7 (nominal) gel_daf
(numeric) thn_ijazah      --> 8 (numeric) thn_ijazah
(nominal) jumlah_nilai    --> 9 (nominal) jumlah_nilai
(nominal) Keterangan      --> 10 (nominal) Keterangan

Time taken to build model: 0.03 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

=== Summary ===

Correctly Classified Instances      80          68.3761 %
Incorrectly Classified Instances    37          31.6239 %
Kappa statistic                    0.3841
Mean absolute error                 0.3943
Root mean squared error             0.4458
Relative absolute error             68.4712 %
Root relative squared error         73.0315 %
Total Number of Instances          117

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          0.608   0.158   0.889    0.608   0.722    0.422   0.818    0.875   TIDAK MASUK
          0.842   0.392   0.508    0.842   0.634    0.422   0.818    0.628   MASUK
Weighted Avg.   0.684   0.234   0.765    0.684   0.693    0.422   0.818    0.795

=== Confusion Matrix ===

 a  b  <-- classified as
48 31 | a = TIDAK MASUK
 6 32 | b = MASUK

```

Gambar 4.23 Hasil Pengujian 1

Pengujian kedua dengan *Data training* 80% dan *data testing* 20%

Tabel 4.6 Pengujian Kedua

<i>Data training</i>	<i>Data testing</i>	Akurasi	Precision	Recall
80%	20%	70,0855%	0,793	0,701

```

Attribute mappings:

Model attributes              Incoming attributes
-----
(numeric) Tahun_masuk        --> 1 (numeric) Tahun_masuk
(nominal) sex                 --> 2 (nominal) sex
(nominal) nama_sekolah        --> 3 (nominal) nama_sekolah
(nominal) provinsi            --> 4 (nominal) provinsi
(nominal) kota                --> 5 (nominal) kota
(nominal) jurusan             --> 6 (nominal) jurusan
(nominal) gel_daf              --> 7 (nominal) gel_daf
(numeric) thn_ijazah           --> 8 (numeric) thn_ijazah
(nominal) jumlah_nilai         --> 9 (nominal) jumlah_nilai
(nominal) Keterangan           --> 10 (nominal) Keterangan

Time taken to build model: 0.01 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

=== Summary ===

Correctly Classified Instances      82          70.0855 %
Incorrectly Classified Instances    35          29.9145 %
Kappa statistic                    0.4241
Mean absolute error                 0.417
Root mean squared error             0.4936
Relative absolute error             73.9608 %
Root relative squared error         83.7793 %
Total Number of Instances          117

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
              0.895   0.392   0.523     0.895   0.660     0.473   0.713    0.483    MASUK
              0.608   0.105   0.923     0.608   0.733     0.473   0.713    0.852    TIDAK MASUK
Weighted Avg.   0.701   0.199   0.793     0.701   0.709     0.473   0.713    0.732

=== Confusion Matrix ===

  a  b  <-- classified as
34  4  | a = MASUK
21 48  | b = TIDAK MASUK

```

Gambar 4.24 Hasil Pengujian 2

Pengujian ketiga dengan *Data training* 90% dan *data testing* 10%

Tabel 4.7 Pengujian Ketiga

<i>Data training</i>	<i>Data testing</i>	Akurasi	Precision	Recall
90%	10%	80,3419%	0,802	0,803

```

Attribute mappings:

Model attributes          Incoming attributes
-----
(numeric) Tahun_masuk    --> 1 (numeric) Tahun_masuk
(nominal) sex            --> 2 (nominal) sex
(nominal) nama_sekolah   --> 3 (nominal) nama_sekolah
(nominal) provinsi       --> 4 (nominal) provinsi
(nominal) kota           --> 5 (nominal) kota
(nominal) jurusan        --> 6 (nominal) jurusan
(nominal) gel_daf         --> 7 (nominal) gel_daf
(numeric) thn_ijazah     --> 8 (numeric) thn_ijazah
(nominal) jumlah_nilai   --> 9 (nominal) jumlah_nilai
(nominal) Keterangan     --> 10 (nominal) Keterangan

Time taken to build model: 0.01 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

=== Summary ===

Correctly Classified Instances      94          80.3419 %
Incorrectly Classified Instances    23          19.6581 %
Kappa statistic                    0.5487
Mean absolute error                 0.3728
Root mean squared error             0.4106
Relative absolute error             67.5683 %
Root relative squared error        72.1742 %
Total Number of Instances          117

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          0.861   0.316   0.850     0.861   0.855     0.549   0.812    0.864    TIDAK MASUK
          0.684   0.139   0.703     0.684   0.693     0.549   0.812    0.622    MASUK
Weighted Avg.   0.803   0.258   0.802     0.803   0.803     0.549   0.812    0.786

=== Confusion Matrix ===

  a  b  <-- classified as
68 11 | a = TIDAK MASUK

```

Gambar 4.25 Hasil Pengujian 3

Pada pengujian yang dilakukan selama 3 kali berturut-turut dengan prosentase berbeda dengan menggunakan algoritma C4.5 dan menggunakan aplikasi WEKA 3.8. Menghasilkan akurasi yang berbeda-beda , pada pengujian pertama dengan prosentase *data training* 70% dan *data testing* 30% menghasilkan akurasi 68,371%, Selanjutnya pada pengujian kedua dengan prosentase *data training* 80% dan *data testing* 20% menghasilkan prosentase 70,0855%, dan pengujian terakhir dengan *data training* 90% dan *data testing* 10% menghasilkan akurasi 80,3419%. Dengan hasil ini disimpulkan bahwa semakin banyak data training yang diproses maka akan lebih akurat hasil yang akan didapatkan untuk melakukan sebuah prediksi.

BAB 5

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil perhitungan algoritma C4.5 diambil kesimpulan sebagai berikut:

1. factor dominan calon mahasiswa yang mendaftar pada program studi sistem informasi yaitu berdasarkan gelombang pendaftaran. Dari 739 data calon mahasiswa dengan melakukan 3 kali percobaan pengujian algoritma C4.5 dengan *data training* 90% dan 10% didapatkan akurasi sebesar 80,3419%.
2. Algoritma C4.5 merupakan metode algoritma yang cocok untuk diterapkan dalam mencari prediksi dari data calon mahasiswa tahun 2013-2015. Hasil prediksi tersebut dapat digunakan oleh program studi Sistem Informasi untuk memperoleh informasi yang dapat dipergunakan dalam menyusun startegi menarik bagi calon-calon mahasiswa baru.

5.2 Saran

1. Jumlah dataset yang digunakan untuk penelitian selanjutnya hendaknya ditambah dan diperbanyak agar menghasilkan prediksi yang lebih akurat.
2. Jumlah *attribute* yang digunakan pada penelitian ini masih melihat 10 *attribute*. Untuk pengembangannya perlu dilihat dari faktor lain seperti alasan calon mahasiswa memilih program studi dan bakat calon mahasiswa sehingga prediksi akan lebih tepat.

DAFTAR PUSTAKA

- [1] Yesi Novaria Kunang , Sri Murniati Andri, "IMPLEMENTASI TEKNIK DATA MINING UNTUK MEMPREDIKSI," *IMPLEMENTASI TEKNIK DATA MINING UNTUK MEMPREDIKSI*, no. ISSN: 1979-2328, May 2013.
- [2] Wendi Zarman, Ida Hamidah Selvia Lorena Br Ginting, "ANALISIS DAN PENERAPAN ALGORITMA C4.5 DALAM DATA MINING UNTUK," *ANALISIS DAN PENERAPAN ALGORITMA C4.5 DALAM DATA MINING UNTUK*, no. ISSN: 1979-911X, Nov. 2014.
- [3] satriyo A, Ariyan Z, Yuliana A Aradea, "Penerapan Decision Tree untuk menentukan pola data penerimaan mahasiswa baru," *Penerapan Decision Tree untuk menentukan pola data penerimaan mahasiswa baru*, vol. 7, no. 1693-9670, 2011.
- [4] Kusrini & Emha Taufiq Luthfi, *Algoritma Data Mining*, Theresia Ari Prabawati, Ed.: C.V ANDI OFFSET, 2009.
- [5] Mind Tools Ltd. (1996-2016) Mindtools. [Online]. <https://www.mindtools.com/dectree.html>
- [6] triisant. (2015, juni) dokumen.tips. [Online]. <http://dokumen.tips/documents/algoritma-c45.html>
- [7] Dr. Saed Sayad. (2010-2016) An Introduction to Data Mining. [Online]. http://www.saedsayad.com/decision_tree.htm
- [8] Udinus. (2012-2016) admisi universitas Dian Nuswantoro. [Online]. <http://admisi.dinus.ac.id/pmbonline/menu/jalurpendaftaran>
- [9] Tyas. (2016) Informatika. [Online]. <http://informatika.web.id/proses-pengambilan-keputusan.htm>

- [10] (2012) Infomoneter. [Online]. <http://infomoneter.com/data-mining-sebagai-solusi-bisnis/>
- [11] erdi susanto. (2011) erdisusanto.com. [Online]. <http://www.erdisusanto.com/2012/06/data-mining-menggunakan-weka.html>
- [12] University of Waikato. Weka. [Online]. <http://www.cs.waikato.ac.nz/ml/weka/>
- [13] Adi Sucipto, "PREDIKSI KREDIT MACET MELALUI PERILAKU NASABAH PADA KOPERASI SINPAM PINJAM DENGAN MENGGUNAKAN ALGORITMA KLASIFIKASI C4.5," *DISPROTEK*, vol. Volume 6 no. 1, Januari 2015.
- [14] Liliana Swastina, "Penerapan Algoritma C4.5 Untuk Penentuan jurusan mahasiswa," *GEMA AKTUALITA*, vol. Vol. 2 No. 1, juni 2013.
- [15] Yusuf Sulistyo Nugroho, "PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI PREDIKAT KELULUSAN MAHASISWA FAKULTAS KOMUNIKASI DAN INFORMATIKA UNIVERSITAS MUHAMMADIYAH SURAKARTA," *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014*, no. ISSN: 1979-911X, 15 November 2014.
- [16] Hilda Amalia Evicienna, "ALGORITMA C4.5 UNTUK PREDIKSI HASIL PEMILIHAN LEGISLATIF DPRD DKI JAKARTA," *Techno Nusa Mandiri*, vol. Vol. IX No.1, Maret 2013.
- [17] Seng Hansun David Hartanto Kamagi, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *ULTIMATICS*, vol. Vol. VI, No. 1, no. ISSN 2085-4552, Juni 2014.
- [18] Rika Yunitarini, Mochammad Kautsar Sophan Windy Julianto, "ALGORITMA C4.5 UNTUK PENILAIAN KINERJA KARYAWAN," *Universitas Trunojoyo Madura*, vol. VOL. IX NOMOR 2, no. ISSN : 1978-

0087, JUNI 2014.

- [19] G.H. Golub, Wolter R. Mears, and Fajrian Nur Adnan, *Matrix Computations*. 2nd ed. Baltimore: The Johns Hopkins University Press., 1989.
- [20] Bob Kuttner, "The Declining Middle," *Atlantic Monthly*, vol. IX, no. 2, pp. 60-70, July 1983.
- [21] Kharis Widiyatmoko, "Rekayasa Perangkat Lunak Translator English-Indonesia dan Indonesia-English," Universitas Dian Nuswantoro, Semarang, Tesis Magister Komputer 2003.
- [22] E. A. Lindgren, "Screen Room Air Inlet and Wave Guard.," Patent 2, 925,457, July 12, 1960.
- [23] Rico Andrea. (2005, July) <http://www.cbl.ncsu.edu/CBLDocs/Bench.html>. [Online]. [12](#)
- [24] Satriyo A,Ariyan Z,Yuliana A Aradea, "Penerapan Decision Tree untuk Penentuan Pola," *Jurnal Penelitian Sitrotika*, vol. Volume 7 Nomor 1, no. ISSN: 1693-9670, July 2011.
- [25] Sukma Putri Utari, "IMPLEMENTASI METODE C4.5 UNTUK MENENTUKAN GURU," *Pelita Informatika Budi Dharma*, vol. Volume : IX, Nomor: 3, no. ISSN : 2301-9425, April 2015.

LAMPIRAN

Lampiran 1. Data Calon Mahasiswa mendaftar pada program studi SI tahun 2013-2015