

# PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI PREDIKAT KEBERHASILAN MAHASISWA DI AKADEMI FARMASI JEMBER

ABDUL KADIR JAELANI. (13 1065 2012)

## ABSTRAK

Predikat keberhasilan mahasiswa di perguruan tinggi merupakan hal utama yang menjadi acuan bagi seorang mahasiswa selama menjalani pendidikan. Data mahasiswa yang terus meningkat dari tahun ke tahun adalah bukti untuk mengembangkan *Data Mining*. Banyaknya jumlah mahasiswa akan diproses untuk mengetahui berdasarkan kriteria apa saja Mahasiswa layak mendapatkan predikat keberhasilannya dengan beberapa atribut seperti, jenis kelamin, prosentase kehadiran, waktu perkuliahan, nilai UN dan asal rumah. Dalam hal ini, peneliti melakukan analisis di Akademi Farmasi Jember. Analisis yang digunakan adalah menggunakan *Data Mining* dengan metode C4.5 dan diproses menggunakan *software Rapidminer* untuk membuat pohon keputusan. Proses validasi yang digunakan adalah *split validation*, hasil menunjukkan nilai akurasi terbaik pada percobaan *split ratio* 0,9 dengan akurasi mencapai 92,59%. Berdasarkan percobaan yang dilakukan, mahasiswa yang berasal dari luar jember dan memiliki nilai UN >80 berpotensi menjadi mahasiswa berpredikat sangat baik.

**Kata Kunci** : Predikat mahasiswa; metode C4.5; *rapidminer*; *decision tree*.

## 1. PENDAHULUAN

### 1.1. Latar belakang

Perguruan tinggi sebagai institusi pendidikan telah memiliki data akademik dan administrasi dalam jumlah yang sangat besar, namun hanya sebagian kecil data tersebut dimanfaatkan (khususnya dalam penyusunan evaluasi diri). Data akademik mahasiswa merupakan data yang dihimpun dari hasil kegiatan proses belajar mengajar selama mengikuti studi di suatu perguruan tinggi. Data tersebut antara lain: data pribadi mahasiswa, data rencana studi, dan data hasil studi (nilai dan indeks prestasi). Jumlah data yang terakumulasi dari tahun ke tahun perlu dilakukan analisis untuk dapat membuka peluang dihasilkannya informasi yang berguna dalam pembuatan alternatif keputusan bagi manajemen perguruan tinggi (Suprawoto, 2016).

Seperti halnya pada proses penerimaan mahasiswa baru dalam sebuah perguruan

tinggi akan menghasilkan data-data baru berupa profil dari mahasiswa baru tersebut. Tahap selanjutnya adalah mahasiswa akan melakukan kegiatan pembelajaran di setiap semester, sehingga dapat diketahui data indeks prestasi mahasiswa pada akhir semester. Hal ini akan terjadi secara berulang pada sebuah perguruan tinggi. Apabila dilakukan pengolahan data pada kedua sumber data tersebut maka dapat diketahui berbagai informasi yang bermanfaat dalam membantu menentukan strategi promosi penerimaan mahasiswa baru tahun berikutnya. Pengolahan tersebut dapat dilakukan menggunakan metode Data Mining dengan algoritma Decision Tree C4.5. Hasil dari pengolahan tersebut akan sangat membantu pihak perguruan tinggi dengan kondisi prodi atau fakultas tersebut masih belum lama berdiri, karena membutuhkan calon mahasiswa yang mempunyai potensi dalam bidang akademik dan pengaruhnya terhadap perguruan tinggi

adalah dapat meningkatkan jumlah mahasiswa dengan nilai akademik yang cukup tinggi.

Kondisi tersebut sama dengan kondisi yang dialami Akademi Farmasi Jember. Data mahasiswa tersebut perlu dilakukan pengolahan data dengan hasil yang menjelaskan tentang berbagai informasi – informasi mengenai persebaran wilayah mahasiswa sehingga informasi dari hasil pengolahan tersebut dapat membantu bagian promosi Akademi Farmasi Jember untuk menentukan strategi promosi.

Berdasarkan data yang diperoleh dari Pangkalan Data Pendidikan Tinggi (PDDIKTI), jumlah mahasiswa baru Akademi Farmasi Jember tahun 2015 berjumlah 92 mahasiswa, pada tahun 2016 berjumlah 106 mahasiswa sedangkan pada tahun 2017 berjumlah 135 mahasiswa. Pada 2 tahun terakhir ini terjadi peningkatan pendaftaran mahasiswa baru, pentingnya sebuah strategi promosi untuk lebih menarik minat calon mahasiswa maka dilakukan analisis dari historis data mahasiswa dengan cara melakukan pengolahan data tersebut.

Metode pengolahan yang cocok untuk mengolah data tersebut adalah menggunakan metode *Data Mining* dengan algoritma C4.5. Atribut yang dibutuhkan dalam pengolahan adalah asal mahasiswa, jenis kelamin, waktu kuliah, nilai UN, dan IPK mahasiswa. Berdasarkan atribut tersebut selanjutnya proses pengolahan data-data dapat membantu lembaga pendidikan dalam upaya memotivasi mahasiswa untuk meningkatkan nilai IPK, serta dapat meningkatkan akreditasi di Akademi Farmasi Jember

Banyak teknik klasifikasi di *Data Mining* salah satunya adalah Algoritma C4.5 yang menyajikan klasifikasi data kedalam bentuk pohon keputusan. Kelebihan utama Algoritma C4.5 dapat membuat pohon keputusan (*decision tree*) yang efisien menangani atribut tipe diskrit dan tipe *diskrit- numerik*, mudah untuk

diinterpretasikan dan memiliki tingkat akurasi yang dapat diterima (Han 2001). Kelemahan Algoritma C4.5 salah satunya terdapat di skalabilitas yaitu data training hanya dapat digunakan dan disimpan secara keseluruhan pada waktu yang bersamaan di memori (Veronica, 2007).

Maka dari itu, pada penelitian kali ini akan mengimplementasikan “Penerapan Algoritma C4.5 untuk Klasifikasi Predikat Keberhasilan Mahasiswa di Akademi Farmasi Jember” yang akan digunakan untuk membantu bagian akademik Akademi Farmasi Jember sebagai bahan identifikasi mengolah data mahasiswa Akademi Farmasi Jember agar dapat meningkatkan jumlah mahasiswa yang lebih berkualitas.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, dapat disimpulkan permasalahan pada penelitian ini yaitu bagaimana menerapkan Algoritma C4.5 untuk klasifikasi mahasiswa berpredikat baik dan sangat baik beserta nilai akurasi, presisi dan recall .

## 1.3. Batasan Masalah

Terdapat beberapa batasan masalah yang diangkat sebagai parameter pengerjaan tugas akhir ini diantaranya sebagai berikut :

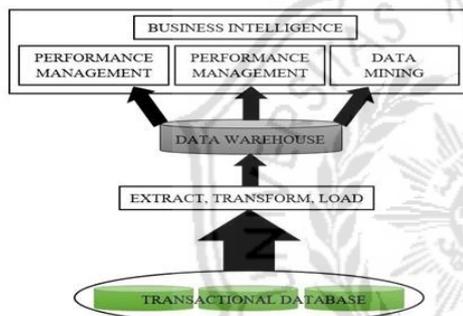
1. Data yang digunakan adalah data mahasiswa Akademi Farmasi Jember.
2. Algoritma yang digunakan adalah Algoritma Decision Tree C4.5 .
3. Atribut yang digunakan adalah asal mahasiswa, jenis kelamin, waktu kuliah, prosentase kehadiran, nilai UN, dan IPK
4. Mahasiswa dengan IPK lebih dari 3,00.

## 2. DASAR TEORI

### 2.1. Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

Data Mining merupakan suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Kegunaan data mining adalah untuk menspesifikasikan pola yang harus ditemukan dalam tugas data mining. Salah satu teknik dalam data mining yaitu untuk membangun sebuah model dalam penelusuran data.



**Gambar 2.1** Posisi data mining dalam business intelligence

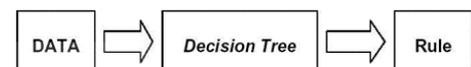
### 2.2 Decision Tree

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

Model pohon keputusan terdiri dari sekumpulan aturan untuk membagi sejumlah

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

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



**Gambar 2.2** Konsep Decision Tree

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

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

- Memiliki kecepatan yang relatif lebih cepat
- Dapat diubah menjadi rule prediksi/klasifikasi dengan mudah dan sederhana
- Dapat dibandingkan tingkat akurasi dengan metode lainnya

### 2.3 Algoritma C4.5

Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C.45. Algoritma C.45 merupakan pengembangan dari algoritma ID3 (Larose, 2005). Secara umum algoritma C4.5

digunakan untuk membangun pohon keputusan adalah sebagai berikut (Kusrini, 2009) :

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

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

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

Keterangan :

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

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

Rumus dasar dari Entropy tersebut adalah sebagai berikut :

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

Keterangan :

- S : himpunan kasus
- n : jumlah partisi S
- p<sub>i</sub> : proporsi dari S<sub>i</sub> terhadap S

## 2.4 Akademi Farmasi Jember

Akademi Farmasi Jember merupakan salah satu lembaga pendidikan farmasi diploma III dan dalam binaan Kopertis Wilayah VII Jawa Timur yang didirikan untuk menjawab kebutuhan masyarakat Jawa Timur khususnya dan Indonesia pada umumnya terhadap tenaga kesehatan yang profesional. Atas tantangan kebutuhan tersebut, dirintislah upaya pendirian Akademi Farmasi (AKFAR) pada tanggal 25 September 2009, Akademi Farmasi (AKFAR) mendapatkan Ijin Penyelenggaraan Institusi dan Pendirian

Akademi Farmasi (AKFAR) berdasarkan SK Mendiknas Nomor: 147/D/O/2009 untuk Jenjang Diploma III Farmasi, dan saat ini AKFAR sudah meluluskan 6 angkatan.

## 2.5 Rapid Miner

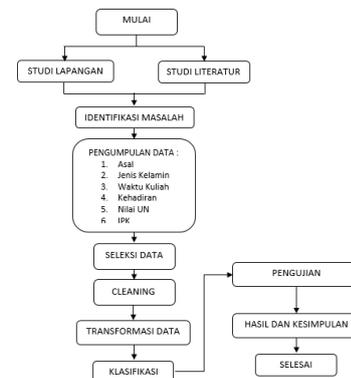
Rapid Miner merupakan perangkat lunak yang bersifat terbuka (open source). Rapid Miner adalah sebuah solusi untuk melakukan analisis prediksi. Rapid Miner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. Rapid Miner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. Rapid Miner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. Rapid Miner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

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

## 3. METODE PENELITIAN

### 3.1 Tahap-tahap penelitian

Bab ini menjelaskan tentang prosedur yang digunakan untuk mengumpulkan, mengelola dan menganalisa data guna menunjang penelitian. Tahapan yang akan dilakukan dalam penelitian ini sebagai berikut :



## 3.2. Penjelasan flowchart penelitian

### 3.2.1. Studi Lapangan dan Studi Literatur

Studi lapangan yaitu melakukan observasi secara langsung di Akademi Farmasi Jember, sedangkan studi literature yaitu melakukan studi untuk mencari materi-materi dan metode yang sesuai dengan penelitian yang akan dilakukan dan dijadikan landasan untuk mendukung pembuatan laporan.

### 3.2.2. Identifikasi Masalah

Mengidentifikasi masalah dari penelitian akan dilakukan. Berikut adalah masalah yang teridentifikasi, bagaimana menerapkan Algoritma C4.5 untuk klasifikasi mahasiswa predikat keberhasilan mahasiswa berdasarkan atribut yang dipakai, yang kemudian akan disampaikan kepada bagian akademik untuk menjadi bahan pertimbangan.

### 3.2.3. Pengumpulan Data

Tahapan ini dilakukan untuk mengumpulkan data-data yang digunakan dalam tugas akhir ini, data yang digunakan adalah :

Asal : asal tempat tinggal mahasiswa  
Jenis Kelamin : jenis kelamin mahasiswa  
Waktu Kuliah : waktu perkuliahan yang dilaksanakan mahasiswa  
Kehadiran : prosentase jumlah kehadiran mahasiswa  
Nilai UN : nilai ujian nasional  
IPK : indeks prestasi kumulatif

### 3.2.4. Seleksi Data

Tugas akhir ini dilakukan di Akademi Farmasi Jember. Data yang digunakan adalah data mahasiswa angkatan 2015 sampai 2017 dengan IPK lebih dari atau sama dengan 3,0.

### 3.2.5. Cleaning Data

Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk data mining, seperti data atau informasi eksternal.

### 3.2.6. Transformasi Data

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam data mining merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

### 3.2.7. Klasifikasi (Menggunakan Algoritma C4.5)

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

Terdapat beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5, yaitu :

- Menyiapkan *data training*. *Data training* diambil dari data riwayat yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
- Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung gain dari atribut, hitung dahulu nilai *entropy*.
- Langkah selanjutnya setelah melakukan perhitungan *entropy*, hitung nilai *Gain*.
- Ulangi langkah ke-2 sehingga semua kasus terpartisi.

### 3.2.8. Pengujian

Pengujian yang dilakukan dalam tahapan ini adalah menguji dan memeriksa apakah sistem sudah sesuai atau tidak. Maka dari itu, peneliti menggunakan metode Accuracy sebagai parameter uji untuk mengukur kinerja dari algoritma.

### 3.2.8. Hasil dan Kesimpulan

Membuat kesimpulan dari hasil penelitian yang berdasarkan dari tujuan penelitian yang sudah ditentukan.

#### 4. Implementasi Sistem

Implementasi sistem adalah tahap penerapan sistem yang telah dibuat pada tahap perancangan. Implementasi bertujuan untuk menerjemahkan perancangan berdasarkan hasil analisis dalam bahasa yang dapat dimengerti oleh mesin.

##### 4.1.1 Perangkat Keras Yang Digunakan

Spesifikasi perangkat keras (hardware) yang digunakan dalam pembangunan sistem ini adalah sebagai berikut:

1. Processor Dual Core berkecepatan 3.2 GHz
2. Kapasitas RAM 2 GB
3. Kapasitas Harddisk 500 GB
4. Monitor LCD 14" beresolusi 1366x768 piksel
5. Mouse USB
6. Keyboard USB

##### 4.1.2 Perangkat Lunak Yang Digunakan

Spesifikasi perangkat lunak (software) yang digunakan dalam pembangunan sistem ini adalah sebagai berikut:

1. Sistem Operasi: Microsoft Windows 7.
2. Aplikasi Rapid Miner versi 5.3

#### 4.2 Analisis Metode Klasifikasi Decision Tree C4.5

##### 4.2.1. Data alternatif mahasiswa

Pada Tugas akhir ini akan dibahas cara mengimplementasikan Algoritma C4.5 pada data mahasiswa angkatan tahun 2015 sampai 2017 guna untuk klasifikasi predikat mahasiswa.

Data mahasiswa angkatan tahun 2015 sampai 2017 sebanyak 330, sedangkan setelah dilakukan cleaning data berjumlah 270 mahasiswa, dikarenakan nilai IPK kurang dari 3,00.

Data alternatif mahasiswa didapat dari bagian akademik yang tercatat dalam format excel yang akan diolah untuk mendapatkan informasi. Adapun tabel data mahasiswa dapat dilihat pada tabel berikut :

| No  | NIM       | Asal                     | Jenis Kelamin | Waktu Kuliah | Kehadiran | UN  | IPK  |
|-----|-----------|--------------------------|---------------|--------------|-----------|-----|------|
| 1   | 151251391 | Kaliwates, jember        | Perempuan     | Pagi         | 100%      | 85  | 3,60 |
| 2   | 151251392 | Tempurejo, jember        | Laki - laki   | Pagi         | 80%       | 68  | 3,20 |
| 3   | 151251394 | Galingan, lumajang       | Laki - laki   | Pagi         | 80%       | 77  | 3,30 |
| 4   | 151251395 | Patrang, jember          | Perempuan     | Pagi         | 75%       | 76  | 3,19 |
| 5   | 151251396 | Bungatan, situbondo      | Perempuan     | Pagi         | 80%       | 62  | 3,30 |
| 6   | 151251397 | Tegalsari, banyuwangi    | Perempuan     | Pagi         | 100%      | 83  | 3,72 |
| 7   | 151251398 | Klakah, lumajang         | Perempuan     | Pagi         | 75%       | 68  | 3,19 |
| 8   | 151251399 | Kaliwates, jember        | Perempuan     | Pagi         | 100%      | 85  | 3,61 |
| 9   | 151251401 | Umbulsari, jember        | Perempuan     | Pagi         | 100%      | 76  | 3,62 |
| 10  | 151251402 | Umbulsari, jember        | Perempuan     | Pagi         | 75%       | 69  | 3,18 |
| 11  | 151251403 | Semboro, jember          | Perempuan     | Pagi         | 80%       | 68  | 3,44 |
| 12  | 151251404 | Sumbersuko, lumajang     | Perempuan     | Pagi         | 100%      | 88  | 3,75 |
| 13  | 151251405 | Patrang, jember          | Laki - laki   | Pagi         | 75%       | 60  | 3,14 |
| 14  | 151251406 | Wuluhan, jember          | Perempuan     | Pagi         | 80%       | 71  | 3,49 |
| 15  | 151251407 | Tempel, lumajang         | Perempuan     | Pagi         | 75%       | 69  | 3,17 |
| 16  | 151251408 | Rambipuji, jember        | Perempuan     | Pagi         | 75%       | 68  | 3,08 |
| 17  | 151251409 | Sumbersari, jember       | Perempuan     | Pagi         | 75%       | 61  | 3,01 |
| 18  | 151251410 | Sumbersari, jember       | Perempuan     | Pagi         | 75%       | 65  | 3,13 |
| 19  | 151251412 | Patrang, jember          | Laki - laki   | Pagi         | 75%       | 62  | 3,03 |
| 20  | 151251413 | Kalisat, jember          | Perempuan     | Pagi         | 80%       | 68  | 3,26 |
| ... | ...       | ...                      | ...           | ...          | ...       | ... | ...  |
| 267 | 171251718 | Ajung, jember            | Perempuan     | pagi         | 80%       | 76  | 3,40 |
| 268 | 171251719 | Ajung, jember            | Perempuan     | sore         | 80%       | 81  | 3,35 |
| 269 | 171251720 | Kebonsari, jember        | Laki - laki   | sore         | 80%       | 71  | 3,45 |
| 270 | 171251722 | Kaliboto Kidul, lumajang | Perempuan     | sore         | 100%      | 87  | 3,60 |

Untuk melakukan proses perhitungan algoritma C4.5 yang perlu dimasukkan pertama kali yaitu melakukan transformasi data. Untuk atribut asal mahasiswa dibagi dalam dua kelas, yaitu jember dan luar jember, untuk atribut kehadiran dan nilai Ujian Nasional dibagi dalam dua kelas variabel yaitu kurang dari 80% dan lebih dari 80%, sedangkan untuk atribut nilai IPK dibagi dalam dua tipe label, jika nilai IPK 3.00 s/d 3.50 adalah baik dan jika > 3.50 adalah sangat baik. Adapun data yang telah melalui transformasi adalah sebagai berikut :

| No  | Asal        | jenis kelamin | waktu kuliah | kehadiran | UN  | IPK         |
|-----|-------------|---------------|--------------|-----------|-----|-------------|
| 1   | jember      | Perempuan     | Pagi         | >80       | >80 | Sangat Baik |
| 2   | jember      | Laki - laki   | Pagi         | >80       | <80 | Baik        |
| 3   | luar jember | Laki - laki   | Pagi         | >80       | <80 | Baik        |
| 4   | jember      | Perempuan     | Pagi         | <80       | <80 | Baik        |
| 5   | luar jember | Perempuan     | Pagi         | >80       | <80 | Baik        |
| 6   | luar jember | Perempuan     | Pagi         | >80       | >80 | Sangat Baik |
| 7   | luar jember | Perempuan     | Pagi         | <80       | <80 | Baik        |
| 8   | jember      | Perempuan     | Pagi         | >80       | >80 | Sangat Baik |
| 9   | jember      | Perempuan     | Pagi         | >80       | <80 | Sangat Baik |
| 10  | jember      | Perempuan     | Pagi         | <80       | <80 | Baik        |
| 11  | jember      | Perempuan     | Pagi         | >80       | <80 | Baik        |
| 12  | luar jember | Perempuan     | Pagi         | >80       | >80 | Sangat Baik |
| 13  | jember      | Laki - laki   | Pagi         | <80       | <80 | Baik        |
| 14  | jember      | Perempuan     | Pagi         | >80       | <80 | Baik        |
| 15  | luar jember | Perempuan     | Pagi         | <80       | <80 | Baik        |
| 16  | jember      | Perempuan     | Pagi         | <80       | <80 | Baik        |
| 17  | jember      | Perempuan     | Pagi         | <80       | <80 | Baik        |
| 18  | jember      | Perempuan     | Pagi         | <80       | <80 | Baik        |
| 19  | jember      | Laki - laki   | Pagi         | <80       | <80 | Baik        |
| 20  | jember      | Perempuan     | Pagi         | >80       | <80 | Baik        |
| ... | ...         | ...           | ...          | ...       | ... | ...         |
| 267 | Jember      | Perempuan     | pagi         | >80       | <80 | Baik        |
| 268 | Jember      | Perempuan     | sore         | >80       | >80 | Baik        |
| 269 | Jember      | Laki - laki   | sore         | >80       | <80 | Baik        |
| 270 | luar jember | Perempuan     | sore         | >80       | >80 | Sangat Baik |

##### 4.2.2. Pemilihan Atribut Sebagai Akar

Untuk memilih atribut sebagai akar, didasarkan pada nilai Gain tertinggi dari

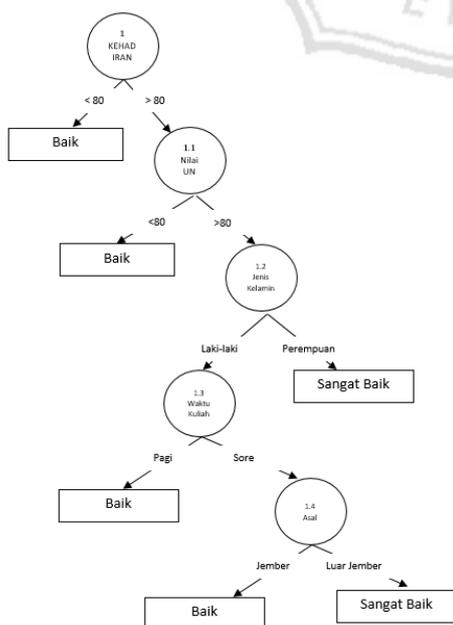
atribut- atribut yang ada. Hasil perhitungan nilai Entropy digunakan untuk menghitung nilai Gain. Berikut adalah hasil perhitungan dengan rumus entropy dan nilai Gain :

|           |             | JML KASUS | BAIK | SANGAT BAIK | ENTROPY     | GAIN       |
|-----------|-------------|-----------|------|-------------|-------------|------------|
| TOTAL     |             | 270       | 176  | 94          | 0,932403312 |            |
| ASAL      |             |           |      |             |             | 0,00038467 |
|           | JEMBER      | 145       | 96   | 49          | 0,922825073 |            |
|           | LUAR JEMBER | 125       | 80   | 45          | 0,942683189 |            |
| JK        |             |           |      |             |             | 0,03122547 |
|           | LAKI-LAKI   | 31        | 28   | 3           | 0,458685816 |            |
|           | PEREMPUAN   | 239       | 148  | 91          | 0,958572204 |            |
| WK        |             |           |      |             |             | 0,00103811 |
|           | PAGI        | 142       | 95   | 47          | 0,915930508 |            |
|           | SORE        | 128       | 81   | 47          | 0,948488076 |            |
| KEHADIRAN |             |           |      |             |             | 0,19518976 |
|           | >80         | 215       | 121  | 94          | 0,988593762 |            |
|           | <80         | 55        | 55   | 0           | 0           |            |
| UN        |             |           |      |             |             | 0,18215179 |
|           | >80         | 131       | 54   | 77          | 0,977648278 |            |
|           | <80         | 139       | 122  | 17          | 0,535942343 |            |

Dari hasil perhitungan tabel dapat diketahui bahwa atribut dengan gain tertinggi adalah kehadiran yaitu sebesar 0.19518976. dengan demikian atribut kehadiran dapat menjadi node akar. Ada 2 nilai atribut dari kehadiran yaitu >80 dan <80. Nilai atribut <80 sudah mengklasifikasikan kasus menjadi satu yaitu “baik” sehingga tidak perlu dilakukan perhitungan lebih lanjut. Tetapi untuk nilai atribut kehadiran >80 masih perlu dilakukan perhitungan lanjutan.

#### 4.2.3. Membuat Cabang untuk masing-masing nilai.

Dari hasil perhitungan diatas dapat digambarkan pohon keputusan akhir seperti gambar berikut :



Aturan atau rule untuk mahasiswa berpredikat baik :

1. Jika nilai kehadiran <80 maka berpredikat baik
2. Jika nilai kehadiran >80 dan nilai UN <80 maka berpredikat baik
3. Jika nilai kehadiran >80, nilai UN >80, berjenis kelamin laki-laki dan waktu kuliah pagi maka berpredikat baik
4. Jika nilai kehadiran >80, nilai UN >80, berjenis kelamin laki-laki, waktu kuliah sore dan berasal dari jember maka berpredikat baik

Aturan atau rule untuk mahasiswa berpredikat sangat baik :

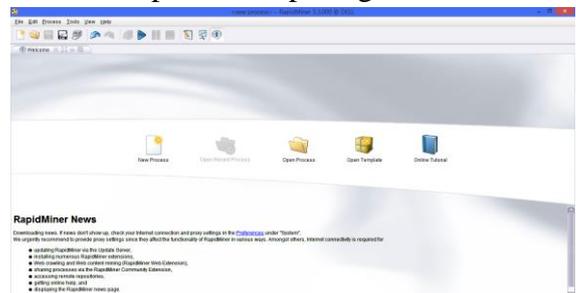
1. Jika nilai kehadiran >80, nilai UN >80, berjenis kelamin laki-laki, waktu kuliah sore dan berasal dari luar jember maka berpredikat sangat baik
2. Jika nilai kehadiran >80, nilai UN >80, berjenis perempuan maka berpredikat sangat baik

#### 4.3 Implementasi Aplikasi *Rapid Miner*

Implementasi pada aplikasi *rapid miner* dilakukan untuk mempermudah proses algoritma C4.5.

##### 4.3.1 Halaman Beranda

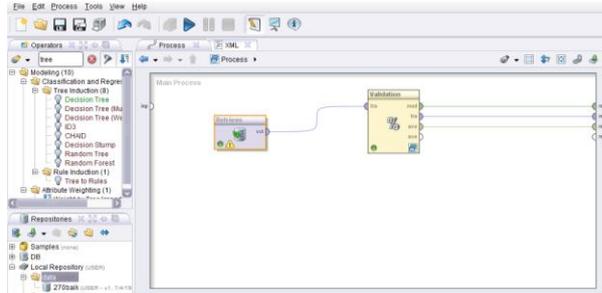
Halaman ini tampilan pertama sekali saat sistem dijalankan, untuk bisa mengakses beberapa tampilan menu yang ada pada halaman beranda. Halaman beranda dapat dilihat pada gambar berikut :



##### 4.3.2 Halaman New Proses

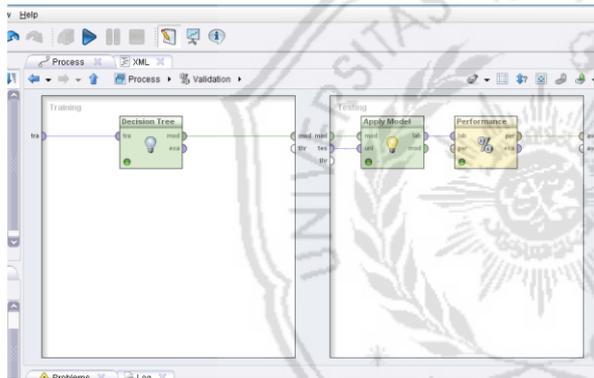
Halaman *new process* merupakan halaman untuk menampilkan semua menu dan proses, bagian proses ini adalah untuk menambahkan proses-proses yang digunakan

untuk mengolah data. Tampilan halaman proses dapat dilihat pada gambar berikut :



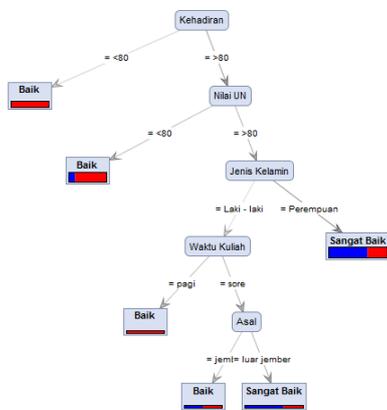
#### 4.3.3 Operator Decision Tree

Untuk membuat *decision tree* dengan menggunakan *Rapid Miner*, kita membutuhkan operator *Decision tree*, operator ini terdapat pada *View Operators*. Untuk menggunakannya pilih *Modelling* pada *View Operator*, lalu pilih *Classification and Regression*, lalu pilih *Tree Induction* dan pilih *Decision Tree*, masukkan pada operator Validasi. Seperti dilihat pada gambar berikut :



#### 4.3.4 Halaman Result View

Halaman *result view* dapat dilihat pada gambar berikut :



Selain menampilkan hasil *decision tree* berupa *graph* atau tampilan pohon keputusan,

*Rapid Miner* juga menyediakan tool untuk menampilkan hasil berupa *text view* dengan mengklik button *Text View* seperti yang tampak pada gambar berikut :

```

Tree
Kehadiran = <80: Baik {Sangat Baik=0, Baik=55}
Kehadiran = >80
| Nilai UN = <80: Baik {Sangat Baik=17, Baik=78}
| Nilai UN = >80
| | Jenis Kelamin = Laki - laki
| | | Waktu Kuliah = pagi: Baik {Sangat Baik=0, Baik=4}
| | | Waktu Kuliah = sore
| | | | Asal = jember: Baik {Sangat Baik=1, Baik=1}
| | | | Asal = luar jember: Sangat Baik {Sangat Baik=2, Baik=1}
| | | Jenis Kelamin = Perempuan: Sangat Baik {Sangat Baik=74, Baik=37}
  
```

Jika dilihat berdasarkan hasil pohon keputusan klasifikasi predikat keberhasilan mahasiswa, bahwa atribut yang mempunyai pengaruh utama untuk mendapatkan predikat keberhasilan adalah kehadiran yang menempati sebagai simpul akar.

#### 4.4 Pengujian Accuracy dengan split validation

*Split Validation* adalah teknik validasi yang membagi data menjadi dua bagian secara acak, sebagian sebagai data *training* dan sebagian lainnya sebagai *data testing*. Dengan menggunakan *Split Validation* akan dilakukan percobaan *training* berdasarkan *split ratio* yang telah ditentukan sebelumnya, untuk kemudian sisa dari *split ratio data training* akan dianggap sebagai data *testing*. *Data training* adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan *data testing* adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian kebenaran atau keakuratan hasil pembelajaran.

Berdasarkan nilai *True Negative (TN)*, *False Positive (FP)*, *False Negative (FN)*, dan *True Positive (TP)* dapat diperoleh nilai akurasi, presisi dan recall. Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan rumus berikut :

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. Presisi dapat diperoleh dengan rumus berikut :

$$Presisi = \frac{TP}{FP + TP} * 100\%$$

Sementara itu, recall menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem. Nilai recall diperoleh dengan rumus berikut :

$$Recall = \frac{TP}{FN + TP} * 100\%$$

Dengan Jumlah data sebanyak 270 mahasiswa, dilakukan dalam 9 percobaan dengan tingkat akurasi yang berbeda, hasil pengujian dapat dilihat pada tabel berikut :

| <b>Split Ratio 0.9</b> | True Sangat Baik | True Baik | Class Precision |
|------------------------|------------------|-----------|-----------------|
| Prediksi Sangat Baik   | 8                | 2         | 80%             |
| Prediksi Baik          | 0                | 17        | 100%            |
| Class Recall           | 100%             | 89.47%    |                 |
| Akurasi                | 92.59%           |           |                 |
| <b>Split Ratio 0.8</b> | True Sangat Baik | True Baik | Class Precision |
| Prediksi Sangat Baik   | 18               | 5         | 78.26%          |
| Prediksi Baik          | 1                | 30        | 96.77%          |
| Class Recall           | 94.74%           | 85.71%    |                 |
| Akurasi                | 88.89%           |           |                 |
| <b>Split Ratio 0.7</b> | True Sangat Baik | True Baik | Class Precision |
| Prediksi Sangat Baik   | 25               | 8         | 75.76%          |
| Prediksi Baik          | 1                | 47        | 97.92%          |
| Class Recall           | 96.15%           | 85.45%    |                 |
| Akurasi                | 88.89%           |           |                 |
| <b>Split Ratio 0.6</b> | True Sangat Baik | True Baik | Class Precision |
| Prediksi Sangat Baik   | 39               | 13        | 75.00%          |
| Prediksi Baik          | 2                | 54        | 96.43%          |
| Class Recall           | 95.12%           | 80.60%    |                 |
| Akurasi                | 86.11%           |           |                 |

| <b>Split Ratio 0.5</b> | True Sangat Baik | True Baik | Class Precision |
|------------------------|------------------|-----------|-----------------|
| Prediksi Sangat Baik   | 25               | 4         | 86.21%          |
| Prediksi Baik          | 30               | 76        | 71.70%          |
| Class Recall           | 45.45%           | 95.00%    |                 |
| Akurasi                | 74.81%           |           |                 |
| <b>Split Ratio 0.4</b> | True Sangat Baik | True Baik | Class Precision |
| Prediksi Sangat Baik   | 43               | 19        | 69.35%          |
| Prediksi Baik          | 19               | 81        | 81.00%          |
| Class Recall           | 69.35%           | 81.00%    |                 |
| Akurasi                | 76.54%           |           |                 |
| <b>Split Ratio 0.3</b> | True Sangat Baik | True Baik | Class Precision |
| Prediksi Sangat Baik   | 34               | 13        | 72.34%          |
| Prediksi Baik          | 37               | 105       | 73.94%          |
| Class Recall           | 47.89%           | 88.98%    |                 |
| Akurasi                | 73.54%           |           |                 |
| <b>Split Ratio 0.2</b> | True Sangat Baik | True Baik | Class Precision |
| Prediksi Sangat Baik   | 69               | 36        | 65.71%          |
| Prediksi Baik          | 10               | 101       | 90.99%          |
| Class Recall           | 87.34%           | 73.72%    |                 |
| Akurasi                | 78.70%           |           |                 |
| <b>Split Ratio 0.1</b> | True Sangat Baik | True Baik | Class Precision |
| Prediksi Sangat Baik   | 75               | 84        | 47.17%          |
| Prediksi Baik          | 11               | 73        | 86.90%          |
| Class Recall           | 87.21%           | 46.50%    |                 |
| Akurasi                | 60.91%           |           |                 |

Tabel diatas menunjukkan bahwa akurasi paling tinggi didapat dari *Split Ratio* 0,9 dengan akurasi 92,59%.

Hasil pengujian *split validation* dari aplikasi *Rapid Miner* dapat dilihat dari gambar berikut :

|                  | true Sangat Baik | true Baik | class precision |
|------------------|------------------|-----------|-----------------|
| pred Sangat Baik | 8                | 2         | 80.00%          |
| pred Baik        | 0                | 17        | 100.00%         |
| class recall     | 100.00%          | 89.47%    |                 |

**PerformanceVector**

PerformanceVector:  
accuracy: 92.59%  
ConfusionMatrix:  
True: Sangat Baik Baik  
Sangat Baik: 8 2  
Baik: 0 17

## 5. Kesimpulan dan Saran

### 5.1 Kesimpulan

Kesimpulan dari penyusunan laporan penelitian ini adalah sebagai berikut.

1. Metode klasifikasi C4.5 dapat diterapkan pada klasifikasi mahasiswa berprestasi di Akademi Farmasi Jember, sehingga metode ini sangat membantu pihak akademik.
2. Dari hasil penerapan metode klasifikasi C4.5 menggunakan 5 atribut tersebut dapat diketahui mahasiswa mana saja yang masuk dalam kedua kelas klasifikasi tersebut.
3. Dalam percobaan pertama dengan split ratio 0,9 didapat hasil akurasi sebesar 92,59%, percobaan kedua dengan split ratio 0,8 didapat hasil akurasi sebesar 88,89%, percobaan ketiga dengan split ratio 0,7 didapat hasil akurasi sebesar 88,89%, setelah dilakukan 9 kali percobaan rata-rata akurasi sebesar 80,1% sehingga bisa dikatakan algoritma C4.5 layak digunakan untuk klasifikasi mahasiswa berprestasi.
4. Berdasarkan percobaan yang dilakukan, mahasiswa yang berasal dari luar jember dan memiliki nilai UN >80 berpotensi menjadi mahasiswa berpredikat sangat baik

### 5.2 Saran

Pada penelitian ini masih sangat jauh dari kata sempurna, disarankan bagi penelitian selanjutnya agar :

1. Penelitian selanjutnya diharapkan dapat mengembangkan metode klasifikasi C4.5 dengan lebih banyak atribut.
2. Dalam klasifikasi mahasiswa berprestasi dapat menggunakan metode yang lain seperti algoritma extreme learning machine (ELM), K-NN dan Naïve.

## DAFTAR PUSTAKA

- Akademi Farmasi Jember, 2017. Buku Panduan Akademik Mahasiswa Tahun Akademik 2017, Jember.
- Aprillia Dennis, Ambarwati Lia, 2013. Belajar Data Mining dengan *RapidMiner*, Jakarta.
- Haryati Siska, 2015. Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu), *Jurnal Media Infotama Vol 11 No. 2*, pp 130-138.
- Kusrini , Luthfi. E T, 2007, *Algoritma Data Mining*, Penerbit Andi
- Kamagi, David Hartanto dan Seng Hansun. 2014. Implementasi Data mining dengan Algoritma C 4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. *Jurnal Seminar Nasional Teknologi Informasi dan Multimedia 2014*. ISSN: 2085-4552.
- Syahputra Guntur, 2014. Penerapan Algoritma C4.5 Dalam Analisa Kelayakan Penerima Bonus Tahunan Pegawai (Studi Kasus : PT. Multi Pratama Nauli Medan), *Jurnal Mantik Penusa Vol 16 No. 2*, pp 1-10