

## ANALISA POLA KELULUSAN MAHASISWA PADA SEKOLAH TINGGI MANAJEMEN INFORMATIKA & KOMPUTER ASIA MALANG DENGAN MENGGUNAKAN ALGORITMA ITERATIVE DICHOTOMISER 3 (ID3)

Dwi Anggih Yosepta <sup>1)</sup>, Tria Aprilianto <sup>2)</sup>

<sup>1</sup> STMIK Asia Malang,

e-mail: [dwianggiyosepta@yahoo.co.id](mailto:dwianggiyosepta@yahoo.co.id) <sup>1</sup>

<sup>2</sup> STMIK Asia Malang,

e-mail: [raptorapril@gmail.com](mailto:raptorapril@gmail.com) <sup>2</sup>

### Abstrak

Sekolah Tinggi Manajemen Informatika & Komputer ASIA Malang merupakan salah satu sekolah tinggi yang berada dalam lingkup pembinaan Kopertis Wilayah VII Jawa Timur yang dituntut untuk dapat menghasilkan lulusan dengan mutu yang baik. Tetapi kenyataannya tingkat kelulusan pada setiap angkatan masih terbilang cukup rendah bahkan tidak sedikit mahasiswa yang harus di dropout. Berdasarkan data mahasiswa lulus angkatan 2008-2011, rata-rata kelulusan mahasiswa sebesar 47%. Penelitian ini bertujuan untuk menganalisa pola kelulusan mahasiswa berdasarkan data induk mahasiswa serta riwayat perkuliahannya.

Algoritma ID3 atau Iterative Dichotomiser 3 merupakan algoritma yang dapat digunakan untuk menganalisa pola kelulusan mahasiswa. Analisa dilakukan dengan cara menghitung entropy dan information gain pada setiap atribut yang selanjutnya membentuk pohon keputusan.

Hasil penelitian ini diperoleh pola kelulusan mahasiswa yang kemudian dapat di generate menjadi pohon keputusan. Atribut yang mempengaruhi kelulusan mahasiswa adalah Indeks Prestasi Semester. Dari 1089 data yang digunakan untuk penelitian dihasilkan 27 pola/rule dan tingkat akurasi yang dihasilkan oleh perangkat lunak cukup baik yaitu 99,39% dengan pengujian yang dilakukan terhadap 163 dataset yang diambil secara acak. Sedangkan pengujian dengan data baru dihasilkan akurasi sebesar 100% dengan menguji 69 data.

**Kata kunci:** Analisa Pola Kelulusan, Kelulusan Mahasiswa STMIK ASIA Malang, ID3

### Abstract

*Sekolah Tinggi Manajemen Informatika & Komputer ASIA Malang is one of the university that is within the scope of the coordination of Kopertis VII Jawa Timur that is claimed to be able to produce graduations with a good quality. But in fact the rate of graduation in any generation is still low even not a few students who should be dropped. Based on the data of students graduated generation 2008-2011, the average graduation students amounted to 47%. This research aims to analyze the patterns of student graduation based on student data as well as the lecture history.*

*ID3 algorithm or Iterative Dichotomiser 3 is an algorithm that can be used to analyze patterns of passing students. The analysis was performed by calculating the entropy and information gained on any attribute which further form the decision tree.*

*The result of this research obtained student graduation pattern which then can generate into a decision tree. Attributes that affect the graduation is student achievement index. From 1089 of the data used for the research resulted 27 patterns and the accuracy that is resulted by the software is good enough i.e 99.39% with testing of 163 datasets taken randomly. While testing with the new data resulted 100% accuracy by testing 69 datum.*

**Keywords:** Graduation pattern analysis, Students graduation in STMIK ASIA Malang, ID3

## 1. PENDAHULUAN

Dewasa ini setiap lembaga penyelenggara pendidikan dituntut untuk dapat menghasilkan

lulusan yang memiliki daya saing tinggi khususnya pada tingkatan perguruan tinggi yang dapat digunakan untuk bersaing pada era

globalisasi. Lulusan yang berdaya saing tinggi pastinya mempunyai mutu yang baik dalam hal prestasi serta ketepatan waktu kelulusan. Oleh sebab itu perguruan tinggi harus mampu melakukan monitoring dan evaluasi terhadap mahasiswa sehingga faktor-faktor yang menjadi penghalang dapat diminimalkan.

Sekolah Tinggi Manajemen Informatika & Komputer ASIA Malang merupakan salah satu sekolah tinggi yang berada dalam lingkup pembinaan Kopertis Wilayah VII Jawa Timur yang juga dituntut untuk dapat menghasilkan lulusan yang memiliki mutu yang baik. Tetapi kenyataannya tingkat kelulusan pada setiap angkatan masih terbilang cukup rendah bahkan tidak sedikit mahasiswa yang harus di dropout. Berdasarkan data mahasiswa lulus yang diperoleh dari Biro Administrasi Akademik (BAA) Sekolah Tinggi Manajemen Informatika & Komputer ASIA Malang, rata-rata tingkat kelulusan mahasiswa angkatan 2008-2011 sebesar 47%. Hal tersebut dipengaruhi oleh beberapa faktor antara lain : kurangnya strategi mahasiswa dalam belajar, sistem pengajaran yang kurang baik, adanya perbedaan kemampuan setiap mahasiswa, serta latar belakang mahasiswa itu sendiri. Adanya kenyataan bahwa dalam sebuah proses perkuliahan selain akan ada mahasiswa yang lulus juga ada mahasiswa yang harus di dropout, hal ini akan menjadi tantangan besar bagi Sekolah Tinggi Manajemen Informatika & Komputer ASIA Malang untuk mencari langkah-langkah yang perlukan agar mahasiswa dapat menyelesaikan studinya sehingga tidak sampai terjadi dropout.

Dengan Memanfaatkan data induk mahasiswa, data status kemahasiswaan, dan data riwayat perkuliahan mahasiswa dapat diketahui pola kelulusan mahasiswa dengan teknik data mining menggunakan algoritma ID3. Oleh karena itu penulis mencoba merumuskan masalahnya yaitu bagaimana menerapkan algoritma *Iterative Dichotomiser 3* (ID3) untuk melakukan analisa pola kelulusan mahasiswa pada Sekolah Tinggi Manajemen Informatika & Komputer ASIA Malang?.

## 2. LANDASAN TEORI

### Data Mining

Secara sederhana data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies, 2004). Data mining juga disebut sebagai serangkaian

proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pramudiono, 2007). Data mining, sering juga disebut sebagai *knowledge discovery in database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso, 2007).

Data mining juga didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semi otomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan dan data yang dibutuhkan dalam jumlah besar (Witten, 2005). Data mining berkaitan dengan bidang ilmu-ilmu lain, seperti database system, data warehousing, statistik, machine learning, information retrieval, dan komputasi tingkat tinggi. Selain itu, data mining didukung oleh ilmu lain seperti neural network, pengenalan pola, spatial data analysis, image database, signal processing (Han, 2006).

Data mining memiliki beberapa karakteristik diantaranya sebagai berikut :

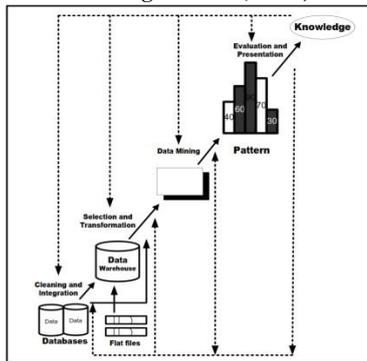
1. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
3. Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi (Davies, 2004).

Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa data mining adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu kumpulan data (database) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. Kata mining sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu data mining sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (artificial intelligence), machine learning, statistik dan database. Beberapa metode yang sering disebutkan dalam literatur data mining antara lain clustering, classification, association rules mining, neural network, genetic algorithm dan lain-lain (Pramudiono, 2007).

### Tahap-tahap Data Mining

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap. Tahapan

tersebut terdiri dari 7 tahap dan bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledge base* (Han, 2006).



**Gambar Tahapan Data Mining**

Tahapan-tahapan pada data mining adalah sebagai berikut :

1. Pembersihan Data (Data Cleaning)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi Data (Data Integration)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

3. Seleksi Data (Data Selection)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil

dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analysis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

4. Transformasi Data (Data Transformation)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5. Proses Mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi Pola (Pattern Evaluation)

Untuk mengidentifikasi pola-pola menarik ke dalam knowledge base yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

7. Presentasi Pengetahuan (Knowledge Presentation)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining.

**Algoritma Iterative Dichotomiser 3 (ID3)**

*Iterative Dichotomiser 3* (ID3) adalah algoritma decision tree learning (algoritma

pembelajaran pohon keputusan) yang paling dasar yang dikembangkan oleh J. Ross Quinlan sejak tahun 1986. Algoritma ini melakukan pencarian secara menyeluruh (greedy) pada semua kemungkinan pohon keputusan (Han, 2006). Algoritma ID3 membentuk pohon keputusan dengan cara pembagian dan menguasai sampel secara rekursif dari atas ke bawah. Algoritma ID3 dimulai dengan semua data yang ada sebagai akar dari pohon keputusan, sebuah atribut yang dipilih akan menjadi pembagi dari sampel tersebut. Untuk setiap atribut dari cabang yang telah dibentuk, semua sampel yang memiliki nilai yang sama dengan atribut cabang akan masuk dalam anggotanya dan dinamakan anak cabang (Nugroho, 2007). Input dari algoritma ini adalah sebuah database dengan beberapa variabel yang juga dikenal dengan atribut. Setiap masukan dalam database menyajikan sebuah objek dari domain yang disebut dengan variabel bebas (independent variable). Sebuah atribut didesain untuk mengklasifikasikan objek yang disebut dengan variabel tidak bebas (dependent variable). Algoritma ID3 dapat diimplementasikan menggunakan fungsi rekursif (fungsi yang memanggil dirinya sendiri). Algoritma ID3 berusaha membangun decision tree (pohon keputusan) secara top-down (dari atas kebawah), mulai dengan pertanyaan : "atribut mana yang pertama kali harus dicek dan diletakkan pada root?". Pertanyaan ini dijawab dengan mengevaluasi semua atribut yang ada dengan menggunakan suatu ukuran statistik (yang banyak digunakan adalah information gain) untuk mengukur efektivitas suatu atribut dalam mengklasifikasikan kumpulan sampel data. Berikut adalah cara kerja dari algoritma ID3 :

1. Pemilihan atribut dengan menggunakan Information Gain.
2. Pilih atribut dimana nilai gainnya paling besar.
3. Buat simpul yang berisi atribut tersebut.
4. Proses perhitungan Information Gain akan terus dilaksanakan sampai semua data masuk ke dalam kelas yang sama, sedangkan atribut yang telah dipilih tidak diikutkan lagi dalam perhitungan nilai information gain.

Secara ringkas, langkah kerja algoritma ID3 dapat digambarkan sebagai berikut (Defiyanti & Pardede, 2010) :

1. Hitung entropy dan information gain dari setiap atribut dengan menggunakan rumus:  
 $Entropy(S) = (- p+ \log_2 p+) + (- p- \log_2 p-)$

Keterangan :

S = Himpunan kasus atau ruang (data) sampel yang digunakan untuk training.

p+ = Jumlah yang bersolusi positif (mendukung) pada data sampel untuk kriteria tertentu.

p- = Jumlah yang bersolusi negatif (tidak mendukung) pada data sampel untuk kriteria tertentu.

$$Gain(S,A) = Entropy(S) - \sum \frac{|S_v|}{|S|} Entropy(S_v)$$

Keterangan :

A = atribut

V = suatu nilai yang mungkin untuk atribut A

|SV| = jumlah sampel untuk nilai v

|S| = jumlah seluruh sampel data

Entropy(Sv) = entropy untuk sampel-sampel yang memiliki nilai v

Entropy(S) = 0 jika semua contoh pada S berada dalam kelas yang sama.

Entropy(S) = 1 jika semua jumlah contoh positif dan jumlah negatif dalam S adalah sama.

Entropy(S) = < 1 jika jumlah contoh positif dan negatif dalam S tidak sama.

Tujuan dari pengukuran nilai information gain adalah untuk memilih atribut yang akan dijadikan cabang pada pembentukan pohon keputusan. Dalam hal ini, dipilih atribut yang memiliki nilai information gain terbesar.

2. Bentuk simpul yang berisi atribut tersebut.
3. Ulangi proses perhitungan information gain yang akan terus dilaksanakan sampai semua data telah termasuk dalam kelas yang sama. Atribut yang telah dipilih tidak diikutkan lagi dalam perhitungan nilai information gain.

ID3 berhenti jika atribut sempurna mengklasifikasikan training sets atau secara rekursif mengoperasikan nilai n, dimana n adalah banyaknya nilai kemungkinan dari suatu untuk mendapatkan atribut terbaik. Adapun sampel data yang digunakan oleh ID3 memiliki beberapa syarat, yaitu :

1. Deskripsi atribut nilai, atribut yang sama harus mendeskripsikan tiap contoh dan memiliki jumlah nilai yang sudah ditentukan.
2. Kelas yang sudah didefinisikan sebelumnya, suatu atribut contoh harus sudah didefinisikan, karena tidak dipelajari oleh ID3.
3. Kelas-kelas yang diskrit, kelas harus digambarkan dengan jelas. Kelas yang kontinu dipecah-pecah menjadi kategori-kategori yang relatif.
4. Jumlah contoh (example) yang cukup, karena pembangkitan induktif digunakan, maka dibutuhkan test case yang cukup untuk membedakan pola yang valid dari peluang suatu kejadian.

5. Pemilihan atribut pada ID3 dilakukan dengan properti statistik, yang disebut dengan information gain.

### 3. HASIL DAN PEMBAHASAN

#### Analisa Masalah

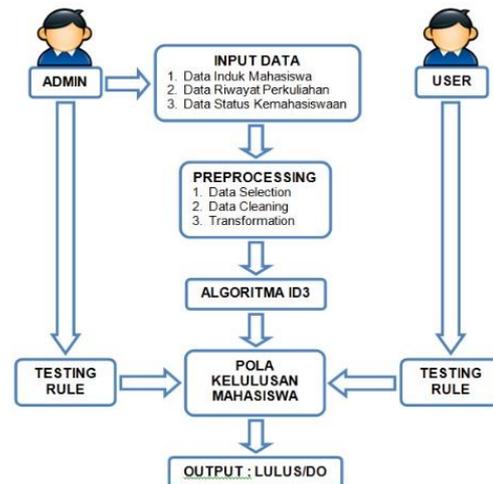
Sekolah Tinggi Manajemen Informatika & Komputer ASIA Malang merupakan salah satu sekolah tinggi di Jawa Timur yang memiliki tingkat kelulusan masih terbilang rendah. Tingkat kelulusan yang rendah disebabkan oleh beberapa faktor diantaranya : kurangnya strategi mahasiswa dalam belajar, sistem pengajaran yang kurang efektif dan adanya perbedaan kemampuan pada setiap mahasiswa. Dari data kelulusan yang diperoleh dari Biro Administrasi Akademik (BAA) Sekolah Tinggi Manajemen Informatika & Komputer ASIA Malang, jumlah mahasiswa lulus angkatan 2008-2011 sampai yudisium gelombang IV semester ganjil tahun akademik 2016/2017 sebanyak 1118 dengan rata-rata kelulusan sebesar 47%. Jumlah tersebut tergolong relatif kecil jika dibandingkan dengan total awal mahasiswa angkatan 2008-2011 yang masuk ke Sekolah Tinggi Manajemen Informatika & Komputer ASIA Malang yaitu sebanyak 2368. Dibalik tingkat kelulusan yang rendah pada Sekolah Tinggi Manajemen Informatika & Komputer ASIA Malang, adanya tingkat dropout mahasiswa yang terbilang cukup tinggi yang dipengaruhi oleh beberapa faktor, salah satunya adalah tingkat keberhasilan studi di tiap semester yang kurang memuaskan.

Melihat selisih jumlah antara mahasiswa lulus dan dropout pada setiap angkatan yang tidak terpaut jauh, hal ini menjadi permasalahan untuk dievaluasi guna dicari penyebabnya. Dengan memanfaatkan sistem data mining dibuat sebuah perancangan aplikasi untuk menganalisa pola kelulusan mahasiswa. Dengan adanya aplikasi ini diharapkan dapat dimanfaatkan oleh para pemangku pendidikan di Sekolah Tinggi Manajemen Informatika & Komputer ASIA Malang untuk menganalisa tingkat kelulusan mahasiswa dalam masa studinya serta kemungkinan dropout yang akan terjadi.

#### Analisa Sistem

Sistem ini merupakan sistem yang dikembangkan untuk membuat analisis terhadap gabungan dari data induk mahasiswa, data riwayat perkuliahan, dan data status kemahasiswaan mahasiswa guna mengetahui pola kelulusan mahasiswa pada Sekolah Tinggi

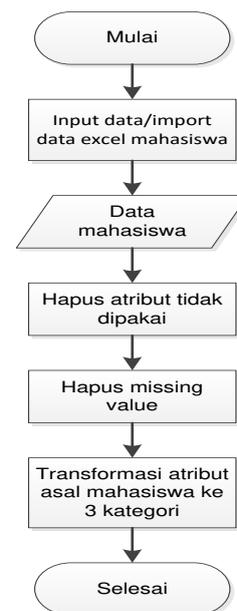
Manajemen Informatika & Komputer ASIA Malang. Tidak semua atribut dari data induk mahasiswa, data riwayat perkuliahan, dan data status kemahasiswaan akan dianalisis tetapi hanya beberapa atribut yang berguna karena data yang terlalu kompleks akan membuat proses mining memakan waktu lama dan tingkat hubungannya pun rendah sehingga sangat mempengaruhi pola/rule yang akan dihasilkan.



Gambar Blok Sistem

#### Preprocessing

Preprocessing data didalam sistem analisa pola kelulusan mahasiswa pada Sekolah Tinggi Manajemen Informatika & Komputer ASIA Malang berfungsi untuk mengolah data training agar data training tersebut memiliki value yang valid sebelum dilakukan proses mining. Alur preprocessing data dapat dilihat pada gambar di bawah ini :



Gambar Alur Preprocessing

**Pemodelan Dataset**

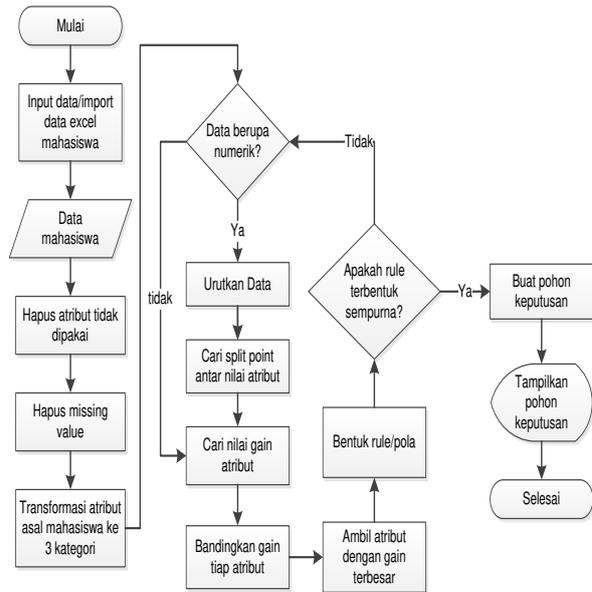
Pemodelan dataset dilakukan untuk memetakan tipe, status, dan pengelompokan dari dataset yang akan digunakan untuk proses mining agar dalam proses mining tidak terjadi kesalahan data pengolahan data untuk masing-masing tipe variabel. Berikut adalah pemodelan dataset analisa pola kelulusan mahasiswa STMIK ASIA Malang.

**Tabel Pemodelan Dataset**

No	Variabel	Tipe	Status	Pengelompokan
1	Jenis Kelamin	Nominal	Input	-
2	Usia	Numerik	Input	-
3	Inputan	Nominal	Input	-
4	Kota Asal Mahasiswa	Nominal	Input	- Malang - Luar Malang - Luar Jatim
5	Program Studi	Nominal	Input	-
6	SKS smt 1	Numerik	Input	-
7	IP smt 1	Numerik	Input	-
8	SKS smt 2	Numerik	Input	-
9	IP smt 2	Numerik	Input	-
10	SKS smt 3	Numerik	Input	-
11	IP smt 3	Numerik	Input	-
12	SKS smt 4	Numerik	Input	-
13	IP smt 4	Numerik	Input	-
14	SKS smt 5	Numerik	Input	-
15	IP smt 5	Numerik	Input	-
16	SKS smt 6	Numerik	Input	-
17	IP smt 6	Numerik	Input	-
18	Jml. Cuti	Numerik	Input	-
19	Jml. Nonaktif	Numerik	Input	-
20	Status	Nominal	Traget	-

**Proses Mining**

Proses mining merupakan proses utama dalam pencarian pola kelulusan mahasiswa pada Sekolah Tinggi Manajemen Informatika dan Komputer ASIA Malang. Dengan menggunakan dataset yang ada, dilakukan perhitungan Entropy dan Information Gain yang berfungsi untuk menentukan atribut yang akan menjadi Root Node, Node Level 1, maupun Node level-level selanjutnya yang pada akhirnya node-node tersebut akan membentuk sebuah pohon keputusan sehingga pola/rule dari kelulusan mahasiswa di Sekolah Tinggi Manajemen Informatika dan Komputer ASIA Malang akan diketahui. Gambar di bawah ini adalah alur proses mining analisa pola kelulusan mahasiswa.



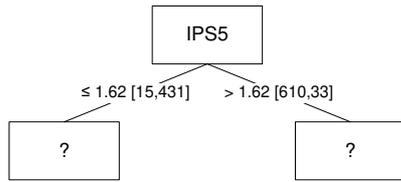
**Gambar Alur Proses Mining**

Atribut-atribut yang ada dilakukan perhitungan entropy dan dicari information gainnya. Tabel di bawah ini adalah daftar nilai information gain dari penentuan *root node*.

**Tabel Information Gain Root Node**

Atribut	Information Gain
Jenis Kelamin (JK)	0.0070
Usia (Usia)	0.0090
Inputan (Inputan)	0.0071
Kota Asal Mahasiswa (KtMhs)	0.0009
Program Studi (PS)	0.0032
SKS Semester 1 (SKS1)	0.0101
IP Semester 1 (IPS1)	0.2995
SKS Semester 2 (SKS2)	0.3121
IP Semester 2 (IPS2)	0.4503
SKS Semester 3 (SKS3)	0.5016
IP semester 3 (IPS3)	0.6041
SKS Semester 4 (SKS4)	0.6417
IP Semester 4 (IPS4)	0.6972
SKS Semester 5 (SKS5)	0.6278
<b>IP Semester 5 (IPS5)</b>	<b>0.7248</b>
SKS Semester 6 (SKS6)	0.6599
IP Semester 6 (IPS6)	0.6923
Jumlah Cuti (Cuti)	0.0075
Jumlah Nonaktif (Nonaktif)	0.1362

Berdasarkan tabel di atas karena nilai information gain terbesar adalah IP Semester 5 (IPS5), maka IP Semester 5 dijadikan root awal.



Gambar Root Node

Implementasi Dan Pengujian

Implementasi antar muka menggambarkan bentuk tampilan grafis dari sistem, berikut ini adalah tampilan dari sistem yang telah dibuat.

1. Form Login

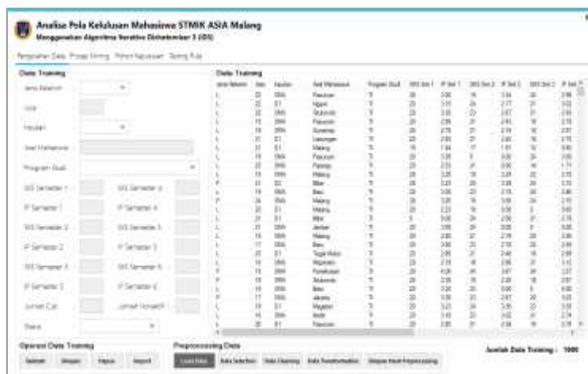
Form login adalah form yang tampil pertama kali saat aplikasi dijalankan. Pada aplikasi ini login dibedakan menjadi 2 yaitu Admin dan User biasa.



Gambar Form Login

2. Form Pengolahan Data

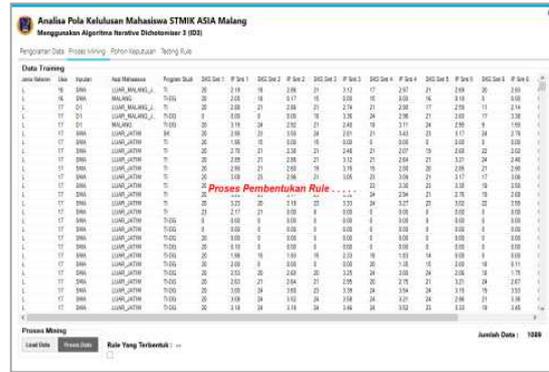
Form pengolahan data digunakan untuk mengolah data training/data mentah menjadi dataset yang sesuai untuk proses mining melalui preprocessing.



Gambar Form Pengolahan Data

3. Form Proses Mining

Form proses mining digunakan untuk mengolah dataset untuk membentuk pola/rule dengan cara menghitung entropy dan mencari information gain dari hasil perhitungan tersebut.



Gambar Form Proses Mining

4. Form Pohon keputusan

Form pohon keputusan berfungsi untuk menampilkan pohon keputusan hasil dari pembentukan rule pada saat proses mining. Rule yang sudah terbentuk juga dapat diekspor ke dalam bentuk excel.



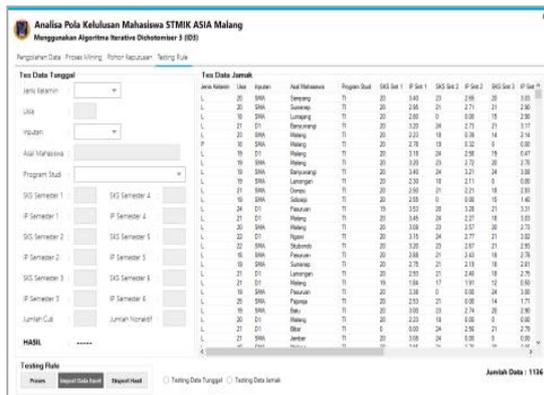
Gambar Form Pohon Keputusan



Gambar Form Eksport Rule

5. Form Testing Rule

Form testing rule digunakan untuk melakukan testing perangkat lunak untuk menguji keakuratan dan juga digunakan untuk melakukan analisa terhadap data baru untuk mengetahui status dari mahasiswa tersebut apakah lulus atau dropout.



Gambar Form Testing Rule

Proses pengujian dilakukan dengan dua model inputan data yaitu yang pertama inputan dengan menggunakan data yang diambil secara acak dari dataset dan yang kedua adalah inputan data baru. Pada pengujian pertama, data yang digunakan di ambil sebanyak 15% dari dataset yaitu sejumlah 163 data dari total 1089 data. Setelah melakukan pengujian dengan memasukkan 163 data pada program, diketahui terdapat 162 data yang sesuai dengan output, dan 1 data yang tidak sesuai sehingga tingkat akurasi program adalah sebagai berikut :

$$\begin{aligned} \text{Akurasi} &= (\text{Jml data benar} / \text{Jml data}) * 100\% \\ &= (162/163) * 100\% \\ &= 99,39\% \end{aligned}$$

Pada pengujian yang kedua dengan menggunakan data baru sebanyak 69 data, diketahui terdapat 69 data yang sesuai dengan output, dan tidak ditemukan data yang tidak sesuai sehingga tingkat akurasi program adalah 100%.

#### 4. PENUTUP

##### Kesimpulan

Berdasarkan penjelasan serta uraian yang telah dibahas, maka dapat diperoleh beberapa kesimpulan sebagai berikut :

1. Faktor yang paling berpengaruh terhadap kelulusan mahasiswa adalah Indeks Prestasi Semester (IPS). Pada penelitian ini adalah Indeks Prestasi Semester 5.
2. Penggunaan algoritma Iterative Dichotomiser 3 untuk proses mining mampu membentuk pohon keputusan sampai lebih dari empat level dan menghasilkan 27 pola/rule.
3. Proses testing dengan 163 data (15% dari dataset) yang diambil secara acak menghasilkan akurasi sebesar 99.39%. Proses tersebut menghasilkan 162 data yang sesuai dengan data sebenarnya.
4. Proses testing dengan 69 data baru menghasilkan akurasi sebesar 100%. Proses

tersebut menghasilkan 69 data yang sesuai dengan data sebenarnya.

#### Saran

Mengingat keterbatasan penulis, berikut adalah beberapa saran yang perlu disampaikan untuk pengembangan data mining ini antara lain :

1. Untuk penelitian selanjutnya dapat menggunakan data induk mahasiswa dengan jumlah data training, range angkatan dan atribut-atribut yang lebih banyak dan juga data riwayat perkuliahan berupa Indeks Prestasi yang lebih banyak pula sehingga rule yang terbentuk dapat lebih kompleks dan akurat.
2. Untuk menjaga keakuratan rule maka perlu dilakukan penambahan data dan dilakukan proses mining ulang secara berkala.
3. Penelitian tentang pencarian pola kelulusan mahasiswa menggunakan algoritma ID3 memiliki tingkat keakuratan hasil yang cukup baik namun tetap diharapkan adanya penelitian terkait kelulusan mahasiswa dengan menggunakan algoritma yang lain.

#### 5. REFERENSI

- Al-Bahra. *Analisa dan Desain Sistem Informasi I*. Jakarta. Elek Media Komutindo. 2005
- Bramer, Max. *Principles of Data Mining*. London. Springer. 2007
- Davies, and Paul Beynon. *Database Systems Third Edition*. New York. Palgrave Macmillan. 2004
- Defiyanti, Sofi. dan Crispina Pardede, D.L. *Perbandingan Kinerja Algoritma ID3 dan C4.5 Dalam Klasifikasi Spam-Mail*. Jurnal Informatika, Universitas Gunadharma. 2010
- Gorunescu, Florin. *Data Mining Concepts, Models, and Techniques*. Verlag Berlin Heidelberg. Springer. 2011
- Han, J. and Kamber, M. *Data Mining Concepts and Techniques Second Edition*. San Francisco. Morgan Kauffman Publishers. 2006
- Han, J, Kamber, M and Pei, J. *Data Mining Concepts and Technique. Third Edition*. Massachusetts. Elsevier, Inc. 2011
- Hartono, Jogyianto. *Analisis dan Desain Sistem Informasi. Pendekatan Terstruktur Teori dan Praktek Aplikasi Bisnis*. Yogyakarta. Andi Publishing. 1999
- Herawati, Fajar Astuti. *Data Mining*. Yogyakarta. Andi Publishing. 2013

- Hoffer, J. A., Ramesh, V., & Topi, H. *Modern Database Management, (11th Edition)*. New Jersey. Prentice Hall. 2012
- Jogiyanto. *Analisis & Desain Sistem Informasi Pendekatan Terstruktur Teori dan Praktik Aplikasi Bisnis*. Yogyakarta. Andi Publishing. 2005
- Kristanto, Andi. *Perancangan Sistem Informasi dan Aplikasinya*. Yogyakarta. Gava Media. 2003
- Kusrini, dan Emha Taufik Luthfi. *Algoritma Data Mining*. Yogyakarta. Andi Publishing. 2009
- Nugroho, Fanuel., Kristanto, Herianto., & Oslan, Yetli. *Validasi Suatu Alamat Menggunakan Decision Tree Dengan Algoritma ID3*. Jurnal Informatika, Universitas Kristen Duta Wacana. 2007
- Pramudiono, I. *Pengantar Data Mining : Menambang Permata Pengetahuan di Gunung Data*. 2007. <http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining.zip> (diakses pada tanggal 10 September 2016)
- Prasetyo, Eko. *Data Mining-Konsep Dan Aplikasi Menggunakan Matlab*. Yogyakarta. Andi Publishing. 2012
- Santosa, Budi. Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta. Graha Ilmu. 2007
- Sudirman. *Implementasi Teknik Data Mining Classification Dengan Metode Decission Tree Untuk Menentukan Tingkat Kelulusan Mahasiswa*. Surabaya. 2012: Prosiding Seminar Nasional Manajemen Teknologi XVI
- Turban, Efraim, Jay E. Aronson & Ting Peng Liang. Decision Support Systems and Intelligent Systems (Sistem Pendukung Keputusan dan Sistem Cerdas). Yogyakarta. Andi Publishing. 2007
- Widodo, Prabowo Pudjo, dkk. *Penerapan Data Mining Dengan Matlab*. Bandung. Rekayasa Sains. 2013
- Witten, Ian H., & Eibe Frank. *Data Mining Practical Machine Learning Tools And Techniques With Java Implementations*. San Diego. Morgan Kaumann Publishers. 2000
- Witten, I. H and Frank, E. *Data Mining : Practical Machine Learning Tools and Techniques Second Edition*. San Francisco. Morgan Kauffman Publishers. 2005