

**IMPLEMENTASI ALGORITMA C4.5 DALAM APLIKASI UNTUK
MEMPREDIKSI JUMLAH MAHASISWA YANG MENGULANG
MATA KULIAH DI STMIK AMIKOM YOGYAKARTA**

NASKAH PUBLIKASI



diajukan oleh

Jefri

09.11.3387

kepada
**SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
AMIKOM YOGYAKARTA
YOGYAKARTA
2013**

NASKAH PUBLIKASI

**IMPLEMENTASI ALGORITMA C4.5 DALAM APLIKASI UNTUK
MEMPREDIKSI JUMLAH MAHASISWA YANG MENGULANG
MATA KULIAH DI STMIK AMIKOM YOGYAKARTA**

Disusun oleh

Jefri

09.11.3387

Dosen Pembimbing

Dr. Kusrini, M.Kom

NIK.190302106

Tanggal 03 Juni 2013

**Ketua Jurusan
S1 Teknik Informatika**

Sudarmawan, MT.

NIK. 190302035

**IMPLEMENTATION OF ALGORITHM C4.5 IN APPLICATION TO PREDICT THE
NUMBER OF STUDENTS THAT REPEAT COURSES IN
STMIK AMIKOM YOGYAKARTA**

**IMPLEMENTASI ALGORITMA C4.5 DALAM APLIKASI UNTUK MEMPREDIKSI
JUMLAH MAHASISWA YANG MENGULANG MATA KULIAH
DI STMIK AMIKOM YOGYAKARTA**

Jefri
Kusrini
Jurusan Teknik Informatika
STMIK AMIKOM YOGYAKARTA

ABSTRACT

Every semester there are always problems that appear when the new semester will enter the problem of students who repeat courses, while college semesters have been completed then the management will be overwhelmed appropriate class schedule of courses that will be held.

Therefore this application will help the academic in order to predict how many students will repeat courses in each period or each semester, so the academic side can overcome the problem that will arise as early as possible eg if the number of students who repeat a lot of the academic authorities could add new classes or if the amount can be included with a few existing classes. In this application implementing the algorithm C4.5.

Algorithm C4.5 is an algorithm that is used to form a decision tree, the decision tree method we can change the fact that very large into a decision tree that represents the rules. The decision trees are also useful to explore the data, find the hidden relationship between the number of candidate input variables to a target variable.

Keywords: Algorithm C4.5, Decision Trees, Course

1. Pendahuluan

Perkembangan dunia pendidikan saat ini sangat pesat, itu dikarenakan kesadaran masyarakat akan manfaat dari pendidikan sudah tinggi sehingga peningkatannya setiap tahun selalu mengalami kenaikan. Dengan perkembangan yang pesat itu dapat menimbulkan masalah bagi setiap perguruan tinggi dalam manajemen mahasiswa dengan baik.

STMIK AMIKOM Yogyakarta merupakan salah satu perguruan tinggi yang sukses menarik banyak mahasiswa disetiap tahunnya, terbukti dengan meningkatnya jumlah pendaftar calon mahasiswa baru disetiap periode. Maka dengan peningkatan ini pihak akademik harus meningkatkan juga kualitas dan kenyamanan bagi mahasiswa dalam menjalani studinya.

Oleh karena itu dibutuhkan sebuah sistem yang bisa membantu dalam manajemen mata kuliah, namun pihak akademik selalu mendapat kesulitan dalam memprediksi jumlah mahasiswa yang akan mengulang mata kuliah pada setiap semesternya atau setiap periode. Sehingga dalam mengatur jumlah kelas yang akan diadakan bagi mahasiswa yang ingin mengulang mata kuliah tertentu masih dilakukan dengan manual.

2. Landasan Teori

2.1 Data Mining

Data Mining didefinisikan sebagai sebuah proses untuk menemukan hubungan, pola dan trend baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2005).

Data mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahawa data mining mewarisis banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. Berawal dari beberapa disiplin ilmu, data mining bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani :

1. Jumlah data yang sangat besar
2. Dimensi data yang tinggi
3. Data yang heterogen dan berbede sifat

Menurut para ahli, *data mining* merupakan sebuah analisa dari observasi data dalam jumlah besar untuk menemukan hubungan yang tidak diketahui sebelumnya dan

metode baru untuk meringkas data agar mudah dipahami serta kegunaannya untuk pemilih data (David Hand,2001).

2.2 Pengelompokan *Data Mining*

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Larose, 2005):

2.2.1 Prediksi

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

Contoh prediksi dalam penelitian:

- a. Prediksi harga gula dalam tiga dekade yang akan datang
- b. Prediksi keadaan cuaca pada suatu tempat apakah akan terang, mendung, hujan dan sebagainya.

Beberapa metode atau teknik yang digunakan dalam klasifikasi dan estimasi dapat juga digunakan (untuk keadaan yang tepat) untuk memprediksi.

2.2.2 Klasifikasi

Didalam pengklasifikasian terdapat target variabel katagori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga katagori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

Contoh lain klasifikasi dalam penelitian adalah:

- a. Menenttukan apakah suatu traksaksi kartu kredit merupakan transaksi yang curang atau tidak
- b. Mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk katagori penyakit apa.

2.3 Pohon Keputusan (*Decision Tree*)

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu.

Pohon Keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Karena pohon keputusan memadukan antara eksplorasi data dan pemodelan, dia sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain (Kusrini, 2009).

2.4 Sistem Pendukung Keputusan

Manusia merupakan bagian dari alam karena hidupnya yang tidak terlepas dari alam. Proses kehidupan manusia merupakan unsur yang semakin lama semakin mendominasi unsur-unsur lainnya di alam. Hal ini dikarenakan manusia diberi kemampuan-kemampuan untuk berkembang dan selalu menginginkan hal baru. Sehingga proses yang terjadi disekelilingnya dan dalam dirinya dirasakan serta diamatinya dengan semua indra yang dimilikinya, lalu ia berbuat dan bertindak.

Dalam menghadapi segala proses yang terjadi disekelilingnya dan didalam dirinya, hampir setiap saat manusia membuat atau mengambil keputusan dan melaksanakannya. Hal ini dilandasi dengan asumsi bahwa segala tindakan dilakukan secara sadar merupakan pencerminan hasil proses pengambilan keputusan dalam pikirannya sehingga sebenarnya manusia sudah sangat terbiasa dalam membuat keputusan. Menurut Mangkusubroto dan Tresnadi, jika keputusan yang diambil tersebut perlu dipertanggung jawabkan kepada orang lain atau prosesnya memerlukan pengertian pihak lain, maka perlu untuk diungkapkan sasaran yang akan dicapai (Suryadi dan Ramdhani, 1998).

2.5 Algoritma C4.5

Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

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

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti yang tertera 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_i| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Sebelum mendapatkan nilai Gain adalah dengan mencari nilai Entropi. Entropi digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut. Rumus dasar dari Entropi adalah sebagai berikut:

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

Keterangan:

S : Himpunan Kasus

A : Fitur

n : Jumlah partisi S

p_i : Proporsi dari S_i terhadap S

2.6 JAVA

Tahun 1991 sekelompok insinyur Sun dipimpin oleh Petrick Naughton dan James Gosling ingin merancang bahasa komputer untuk perangkat konsumen seperti *cable TV Box*. Karena perangkat tersebut tidak memiliki banyak memory, jadi harus ada bahasa yang berukuran kecil dan mengandung kode yang liat. Juga karena manufaktur-manufaktur yang berbeda memilih *processor* yang berbeda pula, maka bahasa harus bebas dari manufaktur manapun. Proyek yang dikerjakan mereka dikasi nam kode "*Green*".

2.7 Gambaran Umum MySQL

2.7.1 Sekilas Tentang MySQL

MySQL (My Structure Query language) adalah salah satu *DataBase Management System (DBMS)* dari sekian banyak DBMS seperti Oracle, MS SQL, Postagre SQL, dan lainnya. MySQL berfungsi untuk mengolah database menggunakan bahasa SQL. MySQL bersifat open source sehingga kita bisa menggunakan secara gratis. Pemrograman PHP juga sangat mendukung/*support* dengan database MySQL (Anhar, 2010).

MySQL merupakan software yang terdistribusikan secara gratis walaupun ada versi untuk komersial. Barulah sejak versi 3.23.19, MySQL dikategorikan software berlisensi GPL, yaitu dapat dipakai tanpa biaya untuk kebutuhan apapun. Awalnya, MySQL dapat diberbagai sistem operasi, termasuk Windows. MySQL menjadi database server *open source* yang sangat populer dan merupakan *Database Relation (RDMS)*, yang mempunyai kemampuan cepat untuk menjalankan perintah SQL (*Structure Query Language*) dengan multi-thread dan multi-user. Dengan melihat kemampuan itu, maka MySQL dijadikan database server handal tambahan feature terus dikembangkan agar lebih optimal kerjanya untuk mendapatkan source dan dokumentasi lengkap dapat diakses melalui situs www.mysql.com.

3. Analisis dan Perancangan Sistem

3.1 Analisis Sistem

Analisis merupakan suatu tahapan mendiskripsikan data-data yang akan digunakan sebagai dasar dari suatu produk (perangkat lunak), fungsi dan kinerja, menunjukkan *Interface* perangkat lunak, membangun batasan yang harus dipenuhi oleh suatu perangkat lunak. Sedangkan sistem merupakan kumpulan elemen-elemen yang saling berkaitan dan bertanggung jawab memproses masukan sehingga menghasilkan suatu keluaran. Pengertian ini merupakan pendekatan sistem yang lebih menekankan pada elemen dan komponennya. Analisis sistem yang dipaparkan dalam pembahasan ini merupakan gambaran secara keseluruhan kendala-kendala yang ada dalam aplikasi *data mining* yang berbasis algoritma C4.5 pada sistem STMIK AMIKOM Yogyakarta.

3.1.1 Analisis Data

Data dari sistem adalah berupa atribut yang dimiliki oleh mahasiswa, nilai atribut, dan nilai kemungkinannya. Data yang dimaksud adalah data yang mempunyai minimal dua kolom atribut. Satu kolom sebagai kolom atribut masukan dan satu kolom sebagai

kolom atribut target. Dari setiap kolom terdapat nilai-nilai yang akan dipergunakan untuk kalkulasi, dan nilai dari setiap atribut harus bersifat diskret.

Ketentuan lain yang harus dipenuhi agar masukan dapat diproses dengan lancar adalah peletakan kolom target harus berada pada posisi terakhir dari kolom pada tabel masukan. Sistem akan membaca masukan dengan atribut target berada pada kolom terakhir dari tabel, maka dari itu selain kolom terakhir sistem akan mengenalinya sebagai atribut masukan dari sistem. Beberapa komponen variable yang digunakan yaitu :

1. **Nilai.** Variabel nem berisi seluruh kemungkinan nilai setiap mata kuliah yang dimiliki oleh mahasiswa untuk diisi pada proses input program. Nilai yang sudah ditentukan pada program ini antara lain A, B, C, D, dan E.
2. **Jenis Kelamin.** Variabel ini berisi data jenis kelamin dari setiap mahasiswa yang akan digunakan sebagai pemebentuk keputusan. Pengelompokan yang ada berdasarkan ketentuan yang dibuat program adalah LAKI-LAKI dan PEREMPUAN
3. **Tahun Angkatan.** Variabel angkatan berisi seluruh kemungkinan angkatan dari setiap mahasis. variabel yang sudah ditentukan pada program berdasarkan hasil pengelompokan survei antara lain 2010, 2011, dan 2012.
4. **Jurusan Kuliah.** Variabel jurusan kuliah berisi seluruh kemungkinan jurusan yang akan diambil oleh mahasiswa pada perguruan tinggi STMIK AMIKOM Yogyakarta. Pengelompokan jurusan kuliah dibagi menjadi 4 berdasarkan ketentuan perguruan tinggi yaitu S1 Teknik Informatika, S1 Sistem Informasi, D3 Teknik Informatika, D3 Manajemen Informatika.
5. **Agama.** Variabel agama adalah variabel yang berisi tentang agama dari setiap mahasiswa. Pengelompokan yang terdapat pada program ini antara lain dibedakan menjadi lima bagian yaitu ISLAM, KRISTEN PROTESTAN, KRISTEN KATOLIK, BUDDHA dan HINDU.
6. **Keputusan.** Variabel keputusan merupakan data yang berfungsi untuk menentukan hasil keputusan. Dalam pengelompokan data sudah ditentukan secara tetap agar tidak terjadi kesalahan dalam perhitungan proses program. Data keputusan hanya memiliki dua buah nilai yaitu "YA" dan "TIDAK".

3.1.2 Analisis Model

Kebutuhan masukan dari sistem adalah berupa atribut yang dimiliki oleh sebuah data nilai atribut dan nilai kemungkinannya yang dibuat kedalam sebuah data tabel. Data tabel yang dimaksud adalah data yang mempunyai minimal dua kolom atribut. Satu kolom sebagai kolom atribut masukan dan satu kolom sebagai kolom atribut target.

Ketentuan lain yang harus dipenuhi agar masukan dapat diproses dengan lancar adalah peletakan kolom target harus berada pada posisi terakhir dari kolom pada tabel

masukan. Sistem akan membaca masukan dengan atribut target berada pada kolom terakhir dari tabel, maka dari itu selain kolom terakhir sistem akan mengenalinya sebagai atribut masukan dari sistem. Berikut adalah data untuk menentukan apakah sebuah mahasiswa tersebut akan mengulang mata kuliah atau tidak dengan nilai-nilai kemungkinannya.

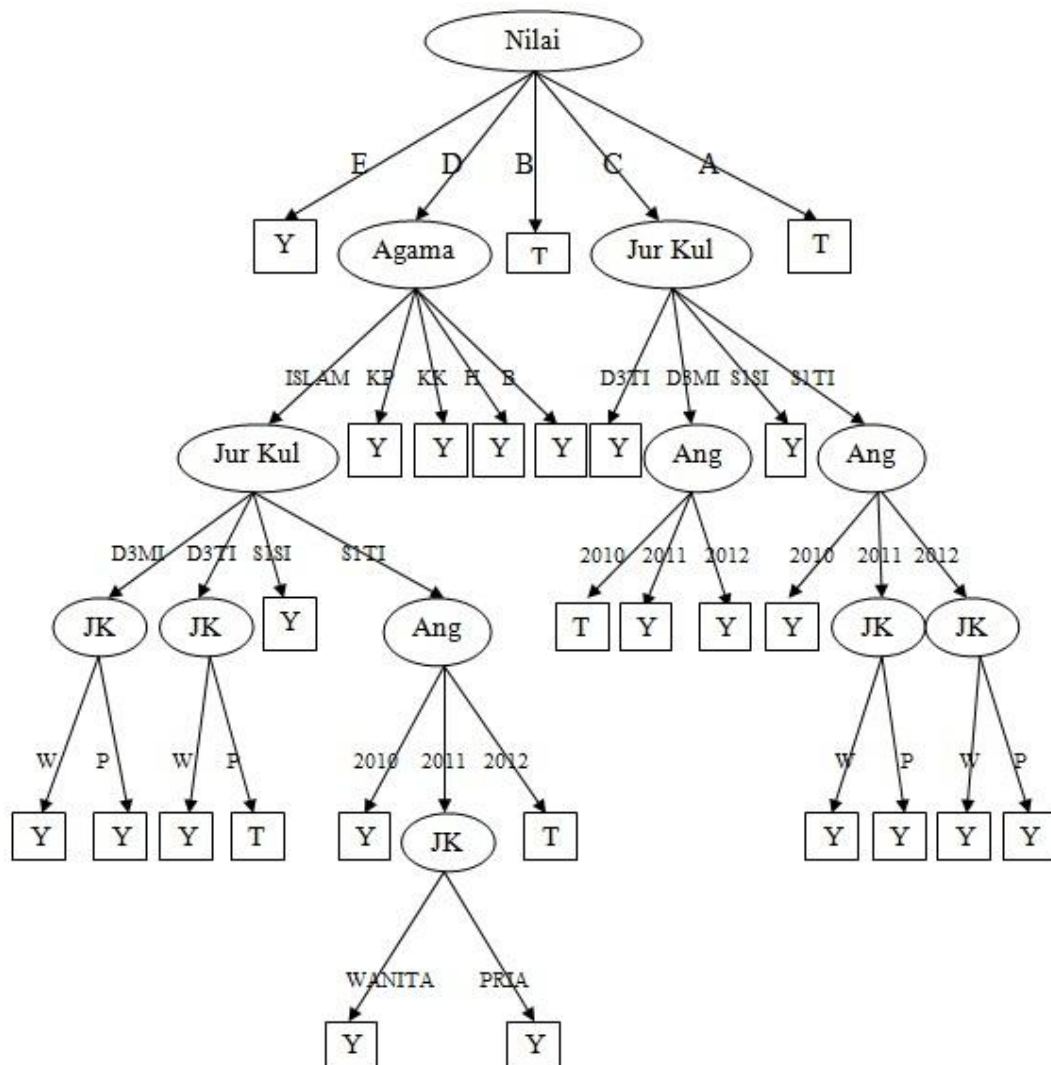
Berikut ini adalah penjelasan lebih terperinci mengenai tiap-tiap langkah dalam pembentukan pohon keputusan dengan menggunakan algoritma C4.5 untuk menyelesaikan permasalahan dalam tugas akhir ini. Salah satu proses kalkulasi dari *entropy* adalah proses *kalkulasi* nilai *entropy TOTAL* yaitu dengan jumlah sampel 100 data.

Tabel 3.1 Perhitungan Node 1

Note			Jlh (s)	Tidak (S1)	Ya (S2)	Entropy	Gain
1	TOTAL		100	34	66	0,92481870	
	THN						0,01780631
		2010	20	4	16	0,72192809	
		2011	50	18	32	0,94268318	
		2012	30	12	18	0,97095059	
	JK						0,00217246
		PRIA	58	21	37	0,94438719	
		WANITA	42	13	29	0,89262301	
	JUR KUL						0,01072363
		S1 TI	36	10	26	0,85240517	
		S1 SI	24	10	14	0,97986875	
		D3 TI	22	7	15	0,90239328	
		D3 MI	18	7	11	0,96407876	
	AGAMA						0,01591289
		ISLAM	75	26	49	0,93105582	
		PROTES	7	2	5	0,86312056	
		KATOLIK	12	4	8	0,91829583	
		BUDDHA	4	2	2	1	
		HINDU	2	0	2	0	
	NILAI						0,21197155
		A	3	3	0	0	
		B	7	7	0	0	
		C	58	19	39	0,91244118	
		D	26	5	21	0,706274089	
		E	6	0	6	0	

Dari tabel hasil kalkulasi di atas dapat kita lihat hasil *entropy* dan *gain* yang diperoleh oleh masing-masing atribut. Dan dapat dilihat juga hasil pohon keputusan berdasarkan kolom *node* yang dihasilkan. Pada *node* pertama dihasilkan variabel Nilai sebagai pohon pertama. Pohon pertama ditentukan dari melihat hasil *gain* terbesar.

Setelah itu untuk menentukan hasil keputusan akhir dapat dilihat dari hasil *entropy* yang hasilnya nol. Apabila telah di dapat hasil *entropy* nol maka selanjutnya kita lihat hasil *variabel* keputusan yang paling banyak nilainya.



Gambar 3.1 Hasil Akhir Pohon Keputusan

4. Hasil Penelitian dan Pembahasan

4.1 Implementasi Sistem

Tahap implementasi sistem (*system implementation*) merupakan tahap meletakkan sistem supaya siap untuk dioperasikan. Tahap ini termasuk juga kegiatan menulis kode program jika tidak digunakan paket perangkat lunak aplikasi (Hartono, 1999). Implementasi juga merupakan penerapan dari elemen-elemen yang telah didesain dalam bentuk pemrograman untuk menghasilkan suatu tujuan yang dibuat berdasarkan kebutuhan.

4.2 Implementasi Interface

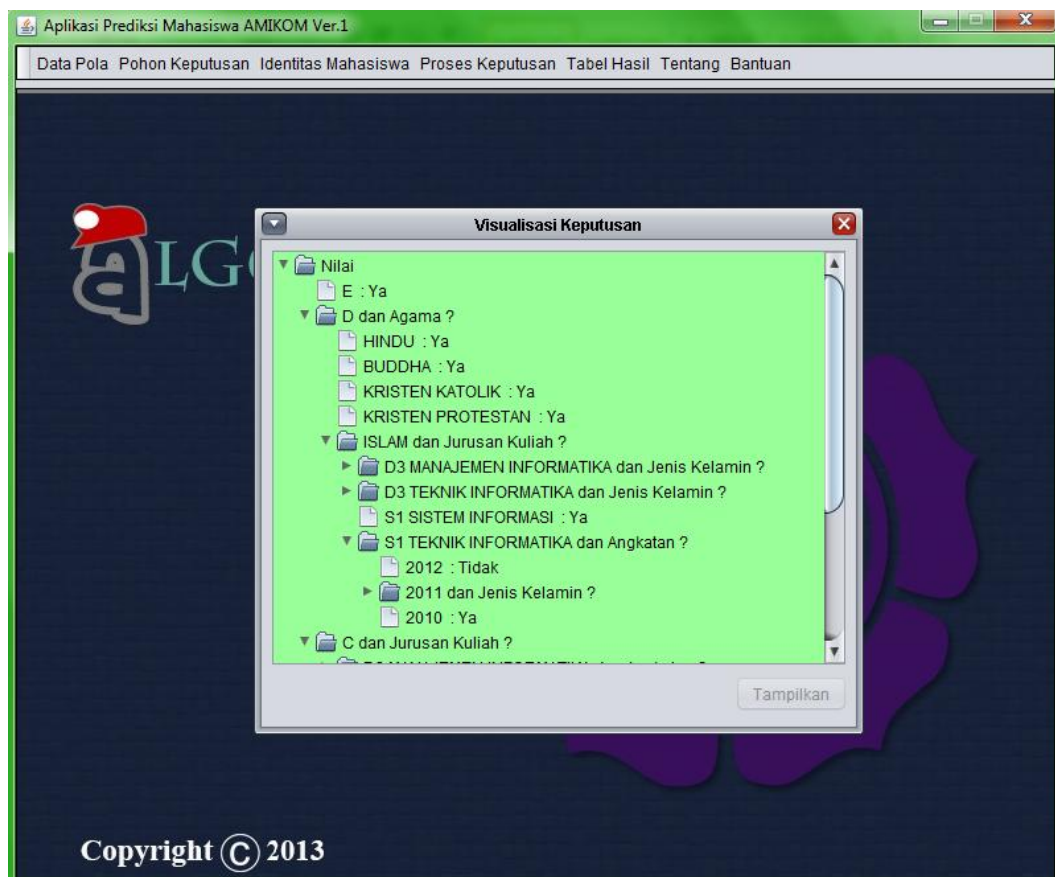
4.2.1 Halaman Utama



Gambar 4.1 Halaman Utama

Form utama dan merupakan halaman yang berada paling depan ada saat program dijalankan. Form utama biasanya digunakan untuk mengontrol semua modul atau bagian program yang ada dalam sistem yang dibuat. Didalam form utama, kita dapat mengontrol program-program lain dengan cara mengkaitkannya dengan link. Link tersebut dapat dibuat dengan menggunakan button.

4.2.2 Halaman Pohon Keputusan



Gambar 4.2 Halaman Pohon Keputusan

Pada form visualisasi keputusan membantu user agar mudah memahami pohon keputusan yang telah dibangun. Visualisasi keputusan didapatkan dari proses pembentukan pola.

4.3 Halaman Proses Keputusan

Aplikasi Prediksi Mahasiswa AMIKOM Ver.1

Data Pola Pohon Keputusan Identitas Mahasiswa Proses Keputusan Tabel Hasil Tentang Bantuan

NIM Cari

MATA KULIAH

NILAI ☐

AGAMA ☐

JURUSAN KULIAH ☐

TAHUN ANGKATAN ☐

JENIS KELAMIN ☐

PREDIKSI

Mulai Proses Simpan Ke Database

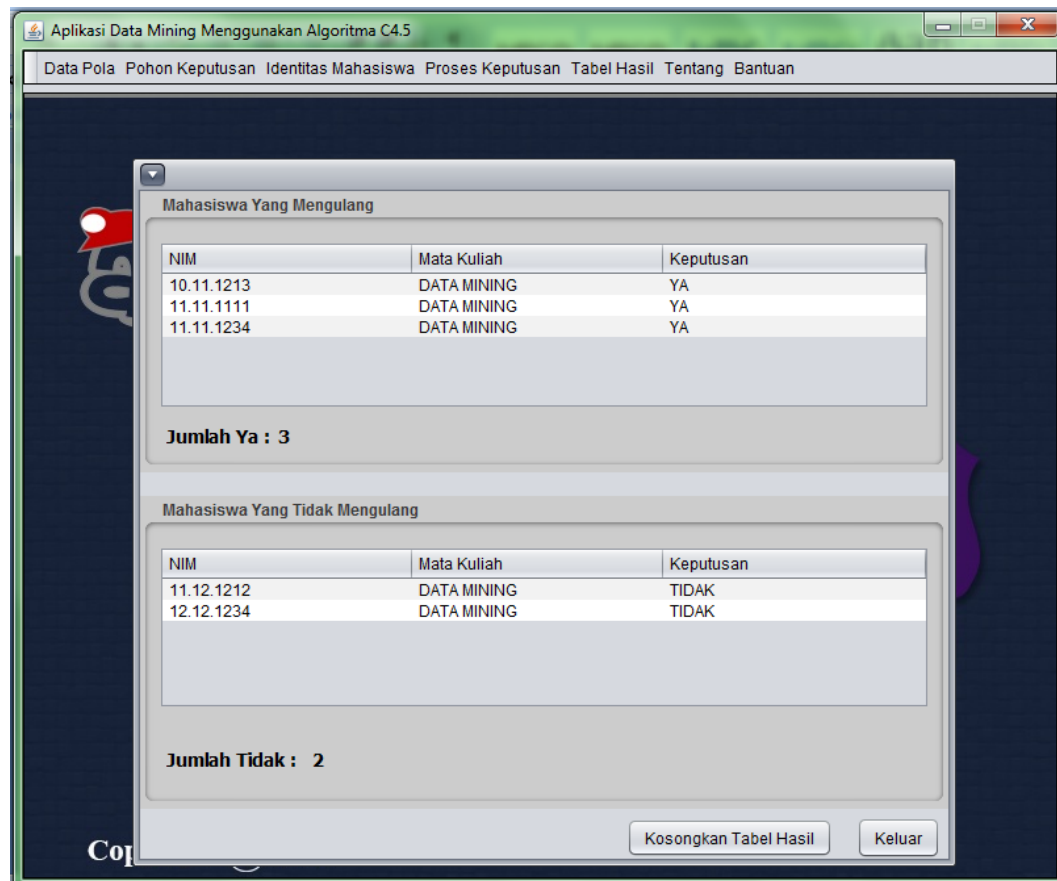
Tutup

Copyright © 2013

Gambar 4.3 Halaman Proses Keputusan

Halaman ini berfungsi untuk menentukan hasil keputusan yang disesuaikan dengan hasil pohon keputusan yang telah dibuat sebelumnya dengan data Mahasiswa yang diambil Tabel Identitas Mahasiswa. Halaman ini menentukan data mahasiswa itu akan mengulang mata kuliah atau tidak.

4.4 Halaman Tabel Hasil



Gambar 4.4 Halaman Tabel Hasil

Halaman tabel hasil akan muncul untuk menampilkan hasil dari semua prediksi yang nantinya bisa dilihat jumlah mahasiswa yang mengulang mata kuliah dan yang tidak mengulang mata kuliah.

5. Penutup

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan dapat diambil beberapa kesimpulan, antara lain:

1. *Data mining* dapat digunakan untuk membantu manajemen STMIK AMIKOM Yogyakarta dalam memprediksi jumlah mahasiswa yang akan mengulang mata kuliah sesuai pola yang terbentuk.

2. Hasil penelitian ini sebagai gambaran bagi mahasiswa STMIK AMIKOM Yogyakarta melihat implementasi dari teknik *data mining* di lapangan.
3. Perancangan pohon keputusan memudahkan dalam proses penalaran penentuan pola keputusan yang terbentuk.
4. Sistem ini telah memberi kemudahan bagi user untuk menentukan hasil keputusan yang mudah dimengerti dalam bentuk visualisasi pohon keputusan.
5. Data yang bisa diambil hanya file yang berekstensi *.csv (Comma Delimited).
6. Dari hasil pengujian sistem ini terdapat kelemahan yaitu pada proses prediksi, hanya dapat dilakukan satu per satu.

5.2 Saran

Mengingat keterbatasan yang dimiliki oleh penulis, baik pengetahuan, waktu, maupun pemikiran, maka penulis dapat memberikan beberapa gambaran sebagai saran yang dapat dipakai sebagai acuan dalam pengembangan aplikasi ini di masa yang akan datang, antara lain:

1. Selama melakukan penelitian ada variabel yang dibutuhkan namun tidak dimiliki oleh pihak akademik. Oleh karena itu saran kepada STMIK AMIKOM Yogyakarta agar menambah variabel pada form pengisian identitas mahasiswa berupa *range* status ekonomi keluarga atau orangtua mahasiswa. (perlu penyesuaian lagi)
2. Untuk membuat hasil pola *data mining* ini bisa optimal dan *real* di lapangan seperti apa yang ada dalam penelitian ini, hendaknya STMIK AMIKOM Yogyakarta membuat kategori mengulang mata kuliah pada saat Remedial atau ikut ulang kelas Reguler.
3. Menyempurnakan segala kekurangan program yang belum diketahui oleh penulis. Seperti hal menambah variabel-variabel data dari segi sosial dan psikologi mahasiswa yang analisisnya dapat dilakukan pada jenjang lebih lanjut.
4. Memberikan sistem yang baik dalam pembentukan pola keputusan basisdata yang ada

DAFTAR PUSTAKA

- Turban, E, dkk. 2005. *Decicion Support Systems and Intelligent Systems*. Yogyakarta: Andi Offset.
- Kusrini, Luthfy, E.T. 2009. *Algoritma Data Mining*. Yogyakarta: Andi.
- Widiarta, Wisnu. 2007. *Dari Pascal ke Java*. Bandung: Informatika.
- Basuki, A dan Syarif, I . 2003. *Decision Tree*. <http://www2.eepis-its.edu/~basuki/lecture/DecisionTree.pdf>. diakses pada tanggal 12 Juni 2012.
- Said, F.L. 2009. *Data mining – Konsep Pohon Keputusan*. <http://fairuzelsaid.wordpress.com/2009/11/24/data-mining-konsep-pohon-keputusan/>. diakases pada tanggal 12 juni 2012.