

# **SISTEM APLIKASI DATA MINING UNTUK MENAMPILKAN INFORMASITINGKAT KELULUSAN MAHASISWA**

**Dwi Retnosari**

Departement Teknik Informatika Universitas Islam Kalimantan, Banjarmasin

E-mail: dwiretnosarisari@yahoo.co.id

## **ABSTRAK**

Kegiatan yang paling menonjol dalam suatu pendidikan tinggi adalah pengolahan informasi yang tersedia yang berkaitan dengan kegiatan yang ada di dalam pendidikan tersebut, mulai dari pengelolaan penerimaan mahasiswa sampai kelulusan mahasiswa. Peningkatan jumlah mahasiswa pada perguruan tinggi Universitas Islam Kalimantan (UNISKA) menuntut adanya suatu kerja sistem yang baik. Pada saat ini belum adanya sistem untuk menentukan kelulusan mahasiswa pada perguruan tinggi Universitas Islam Kalimantan (UNISKA). Dengan adanya sistem ini diharapkan akan meningkatkan kualitas mahasiswa Universitas Islam Kalimantan (UNISKA) sebelum terjun langsung di dalam dunia kerja. *Data mining* merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut. Dengan memanfaatkan data induk mahasiswa dan data kelulusan mahasiswa, diharapkan dapat menghasilkan informasi tentang tingkat kelulusan dengan data induk mahasiswa melalui teknik *data mining*. Kategori tingkat kelulusan diukur dari lama studi dan IPK. Algoritma yang digunakan adalah algoritma *apriori*, informasi yang ditampilkan berupa nilai *support* dan *confidence* dari masing-masing kategori tingkat kelulusan.

Kata kunci : data mining, *algoritma apriori*, *tingkat kelulusan*, *data induk mahasiswa*

## **ABSTRACT**

The most prominent activities in a higher education is processing the available information relating to existing activities in the education, ranging from the management of student admissions to graduation. An increasing number of students in colleges Islamic University of Kalimantan (UNISKA) demands a good working system. At this time the lack of a system to determine college graduation at the Islamic University of Kalimantan (UNISKA). With the existence of this system is expected to improve the quality of student Islamic University of Kalimantan (UNISKA) before plunging directly into the work world. Data mining is mining or the discovery of new information by looking for certain patterns or rules of a number of large amounts of data that are expected to treat the condition. By leveraging master data and data graduation student, is expected to yield information about the graduation rates of students with master data through data mining techniques. Category in measuring graduation rates of long study and GPA. The algorithm used is a priori algorithm, the information displayed in the form of support and confidence values of each each category graduation rates.

Key words: *data mining*, *apriori algorithm*, *graduation rates*, *student parent data*

## 1. PENDAHULUAN

Perguruan tinggi saat ini dituntut untuk memiliki keunggulan bersaing dengan memanfaatkan semua sumber daya yang dimiliki. Selain sumber daya sarana, prasarana, dan manusia, sistem informasi adalah salah satu sumber daya yang dapat digunakan untuk meningkatkan keunggulan bersaing. Sistem informasi dapat digunakan untuk mendapatkan, mengolah dan menyebarkan informasi untuk menunjang kegiatan operasional sehari-hari sekaligus menunjang kegiatan pengambilan keputusan strategis.

Dewasa ini kita sering juga memperbincangkan tentang bagaimana indeks prestasi dalam perguruan tinggi. Beberapa aspek yang mencakup tentang sistem informasi kelulusan juga banyak dibahas di berbagai perguruan tinggi sehingga banyak hal yang harus di ketahui tentang Peningkatan jumlah mahasiswa pada perguruan tinggi dengan menggunakan suatu system kerja yang baik. Pada saat ini belum adanya sistem untuk menentukan kelulusan mahasiswa pada perguruan tinggi Universitas Islam Kalimantan (UNISKA) . Dengan adanya sistim ini diharapkan akan meningkatkan kualitas mahasiswa Universitas Islam Kalimantan (UNISKA) sebelum terjun langsung di dalam dunia kerja.

Kemajuan teknologi informasi dewasa ini, kebutuhan akan informasi yang akurat sangat dibutuhkan dalam kehidupan sehari-hari, sehingga informasi akan menjadi suatu elemen penting dalam perkembangan masyarakat saat ini dan waktu mendatang. Namun kebutuhan informasi yang tinggi kadang tidak diimbangi dengan penyajian informasi yang memadai, sering kali informasi tersebut masih harus di gali ulang dari data yang jumlahnya sangat besar. Metode tradisional untuk menganalisis data yang ada, tidak dapat menangani data dalam jumlah besar.

Pemanfaatan data yang ada di dalam sistem informasi untuk menunjang kegiatan pengambilan keputusan, tidak cukup hanya mengandalkan data operasional saja, diperlukan suatu analisis data untuk menggali potensi-potensi informasi yang ada. Para pengambil keputusan berusaha untuk memanfaatkan gudang data yang sudah dimiliki untuk menggali informasi yang berguna membantu mengambil keputusan, hal ini mendorong munculnya cabang ilmu baru untuk mengatasi masalah penggalian informasi atau pola yang penting atau menarik dari data dalam jumlah besar, yang disebut dengan *data mining*. Penggunaan teknik *data mining* diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya tersembunyi di dalam gudang data sehingga menjadi informasi yang berharga.

Di dalam peraturan akademik Universitas Islam Kalimantan bidang pendidikan tahun 2009 pada BAB I pasal 1 ayat 2 di sebutkan bahwa "Program Sarjana (S1) reguler adalah program pendidikan akademik setelah pendidikan menengah, yang memiliki beban studi sekurang-kurangnya 144 (seratus empat puluh empat) sks (satuan kredit semester) dan sebanyak-banyaknya 160 (seratus enam puluh ) sks yang dijadwalkan untuk 8 (delapan) semester dan dapat ditempuh dalam waktu kurang dari 8 (delapan) semester dan paling lama 14 (empat belas) semester" (Peraturan Akademik, 2009). Dengan memanfaatkan data induk mahasiswa dan data kelulusan mahasiswa, dapat diketahui informasi tingkat kelulusan mahasiswa melalui teknik *data mining*.

## 2. TINJAUAN PUSTAKA

### 2.1 Pengertian *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* adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *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* didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar (Witten, 2005).

Karakteristik data mining sebagai berikut

- Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
- 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 koleksi 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 intelligent*), *machine learning*, statistik dan *database*. Beberapa metode yang sering disebut-sebut dalam literatur *data mining* antara lain *clustering*, *classification*, *association rules mining*, *neural network*, *genetic algorithm* dan lain-lain (Pramudiono, 2007).

### 2.2 Pengenalan Pola, *Data Mining*, dan *Machine Learning*

Pengenalan pola adalah suatu disiplin ilmu yang mempelajari cara-cara mengklasifikasikan obyek ke beberapa kelas atau kategori dan mengenali kecenderungan data. Tergantung pada aplikasinya, obyek-obyek ini bisa berupa pasien, mahasiswa, pemohon kredit, *image* atau *signal* atau pengukuran lain yang perlu diklasifikasikan atau dicari fungsi regresinya (Santoso, 2007).

*Data mining*, sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Sehingga istilah *pattern recognition* jarang digunakan karena termasuk bagian dari *data mining* (Santoso, 2007).

*Machine Learning* adalah suatu area dalam *artificial intelligence* atau kecerdasan buatan yang berhubungan dengan pengembangan teknik-teknik yang bisa diprogramkan dan belajar dari data masa lalu. Pengenalan pola, *data mining* dan *machine learning* sering dipakai untuk menyebut sesuatu yang sama. Bidang ini bersinggungan dengan ilmu probabilitas dan statistik kadang juga optimasi. *Machine learning* menjadi alat analisis dalam *data mining*.

### **2.3 Tahap-Tahap Data mining**

Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledge base*.

Tahap-tahap *data mining* ada 6 yaitu :

#### **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 menyatkan pengambilan aksi nantinya.

### 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.

### 6. Evaluasi pola (*pattern evaluation*),

Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* 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* (Han, 2006).

## 3. METODOLOGI PENELITIAN

Dalam Aplikasi *Data Mining* ini terdapat dua buah form. Form pertama merupakan halaman awal yang berisi perintah pengambilan data pemilihan atribut data induk mahasiswa, input *threshold*, perintah proses *mining* dan tombol keluar aplikasi. Form kedua merupakan halaman report *data mining* yang berisi hasil proses *data mining* yaitu tabel nilai *support* dan *confidence*. Perintah-perintah dalam form pertama berupa tombol ambil data untuk melakukan proses pengambilan data, inputan teks *threshold* untuk memasukkan nilai *threshold*, inputan *combo box* untuk memilih jurusan, inputan *combo box* untuk memilih atribut yang akan diproses *mining*, tombol proses untuk

perintah proses *mining* dan tombol keluar untuk perintah keluar aplikasi. Selain tombol tersebut terdapat tabel data hasil dari perintah yang diberikan

#### 4. PEMBAHASAN

##### 4.1. Rancangan Pengujian

Dalam pengujian Aplikasi *Data Mining* ini digunakan teknik pengujian *Black Box*. Teknik yang digunakan dalam pengujian *Black Box* antara lain :

1. Digunakan untuk menguji fungsi-fungsi khusus dari perangkat lunak yang dirancang.
2. Kebenaran perangkat lunak yang diuji hanya dilihat berdasarkan keluaran yang dihasilkan dari data atau kondisi masukan yang diberikan untuk fungsi yang ada tanpa melihat bagaimana proses untuk mendapatkan keluaran tersebut dan bagaimana hasil dari proses *mining*.
3. Dari keluaran yang dihasilkan, kemampuan program dalam memenuhi kebutuhan pemakai dapat diukur sekaligus dapat diketahui kesalahan-kesalahannya.

Identifikasi dan pelaksanaan pengujian dapat dilihat pada tabel 4.2.

Tabel 4.2 Identifikasi dan pelaksanaan pengujian

No	Kelas Uji	Butir Uji	Pengujian	Jenis Pengujian
1.	Fungsi pengambilan data	Menekan tombol ambil data	Pengujian Sistem	
2.	Fungsi bersihkan Data	Menekan tombol bersihkan data	Pengujian Sistem	
3.	Fungsi Proses	Menekan tombol	Pengujian	<i>Black Box</i>

Tabel 4.3 Hasil Uji Aplikasi Data Mining

Deskripsi	Pengujian	Yang Diharapkan	Evaluasi Hasil	Didapat
Fungsi pengambilan data	Menekan tombol ambil data	Tampilan data gabungan pada Tabel	Data gabungan muncul pada <i>gridview</i> gabungan	Data gabungan muncul pada <i>gridview</i> gabungan

Fungsi bersihkan data	Menekan tombol bersihkan Data	Tampilan data gabungan yang sudah bersih dari data kotor dan atribut tidak dipakai	Data gabungan Yang Sudah Bersih muncul pada <i>gridview</i> gabungan bersih	Data gabungan yang sudah bersih muncul pada <i>gridview</i> gabungan bersih
Fungsi Proses <i>Mining</i>	Menekan tombol Proses <i>mining</i>	Tampilkan form report <i>mining</i> berupa tabel hasil proses mining dengan nilai support dan confidence dan Menampilkan masing-masing Kategori dengan nilai <i>Confidence</i> paling tinggi	Tampil form report dengan hasil report proses <i>mining</i>	Form report dengan tabel hasil proses <i>mining</i> dan tampilan masing- masing kategori dengan nilai <i>confidence</i> paling tinggi

#### 4.2. Analisis Hasil Uji

Berdasarkan pengujian di atas, dapat disimpulkan bahwa Aplikasi *Data Mining* dapat digunakan untuk mengetahui hubungan data kelulusan dengan data induk mahasiswa.

### 5. KESIMPULAN

#### 5.1. Kesimpulan

- a. Kesimpulan yang dapat diambil dalam pengerjaan tugas akhir ini adalah Aplikasi *Data Mining* ini dapat digunakan untuk menampilkan informasi tingkat kelulusan. Informasi yang ditampilkan berupa nilai *support* dan *confidence* hubungan antara tingkat kelulusan dengan data induk mahasiswa. Semakin tinggi nilai *confidence* dan *support* maka semakin kuat nilai hubungan antar atribut. Data induk mahasiswa yang diproses *mining* meliputi data proses masuk, data asal sekolah, data kota mahasiswa, dan data program studi.
- b. Hasil dari proses *data mining* ini dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang faktor yang mempengaruhi tingkat kelulusan khususnya faktor dalam data induk mahasiswa.

## 5.2. Saran

Untuk pengembangan Aplikasi *Data Mining* lebih lanjut, dapat menggunakan algoritma lain, misal algoritma *FP-Growth*. Perbedaannya adalah algoritma *apriori* harus melakukan *scan database* setiap kali iterasi, sedangkan algoritma *FP-Growth* hanya melakukan satu kali *scan database* diawal.

### DAFTAR PUSTAKA

- Anonim, 2009 "*Peraturan Akademik Universitas Diponegoro Bidang Pendidikan*", Semarang.
- Bramer, Max, 2007, "*Principles of Data Mining*", Springer, London.
- Chintakayala, Padmini. 2005. "*Beginners Guide for Software Testing : Symbiosys Technologies*".
- Davies, and Paul Beynon, 2004, "*Database Systems Third Edition*", Palgrave Macmillan, New York.
- Elmasri, Ramez and Shamkant B. Navathe, 2000, "*Fundamentals of Database Systems. Third Edition*", Addison Wesley Publishing Company, New York.
- Han, J. and Kamber, M, 2006, "*Data Mining Concepts and Techniques Second Edition*". Morgan Kauffman, San Francisco.
- Kadir, Abdul, 1999, "*Konsep dan Tuntunan Praktis Basis Data*", Penerbit Andi, Yogyakarta.
- Kusrini, dan Emha Taufik Luthfi, 2009, "*Algoritma Data Mining*", Penerbit Andi, Yogyakarta.
- Pramudiono, I. 2007. *Pengantar Data Mining : Menambang Permata Pengetahuan di Gunung Data*. <http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining.zip> Diakses pada tanggal 15 Maret 2009
- Rainardi, Vincent, 2008, "*Building a Data Warehouse with Examples in SQL Server*", Springer, New York.
- Pressman, Roger S, 1997, "*Software Engineering:A Practitioner's Approach*."The McGraw-Hill Companies, Inc., New York.
- Santosa, Budi, 2007, "*Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*", Graha Ilmu, Yogyakarta.
- Sommerville, Ian, 2003, "*Software Engineering (Rekayasa Perangkat Lunak)/ Edisi 6/Jilid 1*" Erlangga, Jakarta.