

# IMPLEMENTASI DATA WAREHOUSE DAN DATA MINING: STUDI KASUS ANALISIS PEMINATAN STUDI SISWA

**Eka Miranda**

Fakultas Ilmu Komputer, Jurusan Sistem Infromasi, Universitas Bina Nusantara  
Jln. KH Syahdan No.9, Palmerah, Jakarta Barat 11480  
ekamiranda@yahoo.com

## ABSTRACT

*This paper discusses the implementation of data mining and their role in helping decision-making related to students' specialization program selection. Currently, the university uses a database to store records of transactions which can not directly be used to assist analysis and decision making. Based on these issues then made the data warehouse design used to store large amounts of data and also has the potential to gain new data distribution perspectives and allows to answer the ad hoc question as well as to perform data analysis. The method used consists of: record analysis related to students' academic achievement, designing data warehouse and data mining. The paper's results are in a form of data warehouse and data mining design and its implementation with the classification techniques and association rules. From these results can be seen the students' tendency and pattern background in choosing the specialization, to help them make decisions.*

**Keywords:** *specialization studies, data warehousing, data mining*

## ABSTRAK

*Tulisan ini membahas implementasi data mining dan peranannya dalam membantu pengambilan keputusan yang terkait dengan pemilihan program peminatan studi siswa. Saat ini universitas menggunakan database untuk menyimpan record transaksi. Sementara, database tidak dapat digunakan secara langsung untuk membantu analisis dan pengambilan keputusan. Berdasarkan permasalahan ini, dibuat desain data warehouse yang dapat menyimpan data dalam jumlah besar, berpotensi mendapatkan sudut pandang baru dari sebaran data, dan memungkinkan untuk memberikan pelaporan dan jawaban ad hoc, termasuk melakukan analisis data. Metode yang digunakan: analisis record yang terkait dengan prestasi akademik siswa, mendesain data warehouse, dan data mining. Hasil dari tulisan ini berupa desain data warehouse dan data mining serta implementasinya dengan teknik classification and association rules. Dari hasil ini dapat dilihat kecenderungan dan pola latar belakang siswa dalam memilih peminatan, sehingga siswa terbantu dalam mengambil keputusan.*

**Kata kunci:** *peminatan studi, data warehouse, data mining*

# PENDAHULUAN

## Latar Belakang

Universitas perlu memanfaatkan secara optimal sumber daya yang dimilikinya seperti infrastruktur, fasilitas, orang, data serta pengetahuan untuk dapat bersaing. Data dapat diproses menjadi informasi dan kemudian dapat didistribusikan untuk mendukung kegiatan operasional sehari-hari yang nantinya dapat juga diolah lebih lanjut dan digunakan untuk kegiatan pengambilan keputusan. Hambatan yang sering dialami adalah tidak semua data terintegrasi sehingga terkadang informasi yang dihasilkan tidak lengkap dan akurat.

*Data warehouse* bukan hanya suatu arsip atau fasilitas penyimpanan data. Pengembangan *data warehouse* ditujukan untuk mengintegrasikan, menggali dan membuat intisari informasi yang penting dari data yang tersebar. *Data warehouse* dikembangkan untuk mempertemukan suatu permintaan bagi Manajemen Informasi dan Analisis yang tidak bisa dijumpai pada basis data. Aplikasi basis data tidak mampu mempertemukan kebutuhan ini untuk beberapa pertimbangan: pengolahan pelaporan data yang banyak/besar dapat mengurangi waktu tanggap aplikasi basis data; perancangan aplikasi basis data tidaklah dioptimalkan untuk analisis informasi dan pelaporan; kebanyakan organisasi mempunyai lebih dari satu aplikasi basis data, sehingga pelaporan secara keseluruhan organisasi tidak bisa didukung dari sistem tunggal; pengembangan laporan pada aplikasi basis data sering memerlukan program komputer spesifik yang mahal.

Aktifitas lebih lanjut kemudian adalah analisis informasi dan data yang ada di *data warehouse*, yang memungkinkan untuk menampilkan dan menganalisis informasi dari berbagai sudut pandang (dimensi), teknik-teknik *data mining* kemudian digunakan untuk menggali data dan mendapatkan pemahaman yang tersembunyi dari data serta informasi yang ada untuk membantu dalam kegiatan pengambilan keputusan, dalam hal ini membantu siswa dalam menentukan program peminatan studi yang akan diambilnya (Azimah, Suchyo, 2007).

## Perumusan Masalah

Masalah yang diformulasikan dari kasus analisis peminatan studi siswa terdiri atas: Bagaimana menghasilkan informasi yang dibutuhkan pihak universitas dan siswa yang terkait dengan kegiatan akademik siswa? Bagaimana memproses data menjadi informasi yang dapat digunakan oleh universitas dan siswa dalam mengambil keputusan yaitu keputusan yang terkait dengan pemilihan program peminatan studi siswa? Bagaimana membantu pihak universitas dalam mengambil keputusan dengan cepat dan akurat?

Dari ketiga isu di atas terlihat bahwa tidak hanya dibutuhkan tempat untuk menyimpan data tapi juga menghasilkan informasi serta mampu melakukan analisis data dari berbagai dimensi dan menggali nilai tersembunyi dari data yang ada untuk membantu pengambilan keputusan.

## Tujuan

Tujuan dari penulisan tulisan ini adalah untuk menjelaskan bagaimana *data warehouse* dan *data mining* digunakan dalam kegiatan akademik yaitu terkait dengan analisis terhadap data akademik siswa yang nantinya dapat menghasilkan informasi yang dapat dianalisis dari berbagai sudut pandang, kemudian dengan teknik *data mining* yang tepat dapat digunakan untuk membantu pihak universitas dan siswa dalam menentukan program peminatan studi yang akan diambil.

## Studi Pustaka

### Data warehouse

Membuat *data warehouse* tidak sekedar memindahkan data operasional ke dalam *data warehouse*, memindahkan data dari beberapa sumber berpotensi menimbulkan masalah besar terutama terkait dengan isu tidak konsistennya data. *Data warehouse* tidak hanya digunakan dalam melakukan *loading*, integrasi, dan menyimpan data dalam jumlah besar, namun juga berpotensi untuk mendapatkan sudut pandang baru dari sebaran data, dan memungkinkan untuk memberikan pelaporan dan jawaban dari pertanyaan pengguna yang bersifat *ad hoc* secara cepat dan lebih baik.

### Tahapan untuk membangun data warehouse terdiri atas:

#### Identifikasi sumber data

Langkah pertama sebelum mulai mengembangkan *data warehouse* yaitu identifikasi sumber data. Perlu mencari tahu apa data yang dibutuhkan untuk diletakkan ke dalam *data warehouse*.

#### Membangun ETL (Extraction, Transformation, Loading) tool

Setiap *data warehouse* memiliki data yang berbeda persyaratan, oleh karena itu, *ETL tool* yang sesuai adalah solusi yang lebih baik untuk memenuhi persyaratan.

#### Ekstraksi

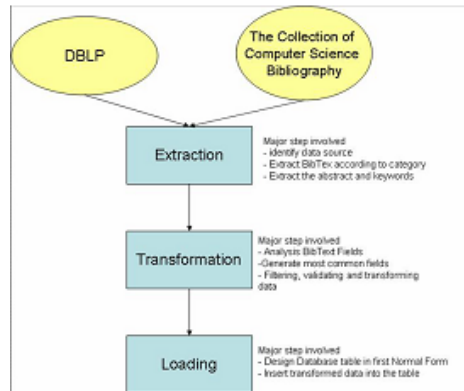
Perlu menentukan sistem *database* yang akan digunakan dan juga mencari tahu data apa yang diperlukan sebelum mendapatkannya. Penurunan biaya perangkat keras dan penyimpanan telah mengatasi masalah pada menghindari duplikasi data dan juga kekhawatiran pada kurangnya ruang penyimpanan sebagai tempat penyimpanan data yang berlebihan atau yang tidak diperlukan. Namun, mungkin tidak ada alasan untuk menyimpan data yang tidak diperlukan dan telah diidentifikasi tidak berguna dalam proses pengambilan keputusan. Oleh karena itu, diperlukan untuk mendapatkan hanya ekstrak data yang relevan sebelum membawa ke *data warehouse* (Mallach, 2000).

#### Transformasi

Setelah penggalian data dari berbagai sumber, transformasi dibutuhkan untuk menjamin konsistensi data. Agar mengubah data ke dalam *data warehouse* dengan benar, perlu mengetahui cara pemetaan bidang sumber data eksternal ke *data warehouse*. Transformasi dapat dilakukan selama ekstraksi data atau saat memuat data ke dalam *data warehouse*. Integrasi ini bisa menjadi masalah yang kompleks ketika jumlah sumber data menjadi lebih besar.

#### Loading

Setelah proses penggalian, mengubah dan pembersihan telah dilakukan, data diambil ke dalam *data warehouse*. Pemuatan data dapat dikategorikan ke dalam dua jenis; pemuatan data yang saat ini ada dalam *database* operasional dan pemuatan pembaruan *data warehouse* dari perubahan yang telah terjadi dalam *database* operasional. Untuk menjamin kesegaran data, *data warehouse* perlu untuk terus memperbaharui datanya. Banyak persoalan yang perlu untuk dipertimbangkan terutama saat memuat data yang sudah diperbaharui ke *data warehouse*. Sementara melakukan pemutakhiran *data warehouse*, perlu memastikan bahwa tidak ada data yang dilepaskan dan juga untuk memastikan *overhead* minimum selama proses *scanning* file yang ada.

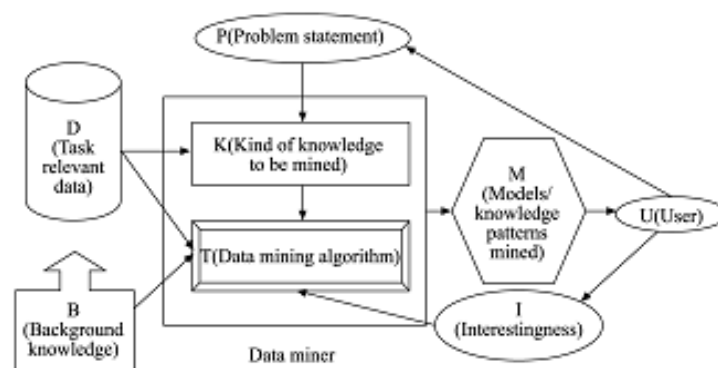


Gambar 1. Tahap-tahap membangun data warehouse.

Sumber: Wah, Teh Ying, Peng, Ng Hooi, Hok, Ching Sue, Building Data Warehouse, *Proceedings of the 24th South East Asia Regional Computer Conference*, , 2007, Bangkok, Thailand

## Data Mining

*Data mining* (DM) adalah proses mencari dan menggali hubungan antar data yang berguna bagi penggunanya. *Data mining* (DM) juga merupakan proses mencari data untuk menemukan yang sebelumnya tidak diketahui dari hubungan antara data yang menarik untuk penggunanya (Han dan Kamber, 2001). Tahap terpenting dalam mendapatkan pengetahuan dari dalam *database* adalah *data mining*. DM telah menjadi bidang ilmu tersendiri (Fayyad, 1996; Chen dan Liu, 2005). *Data mining* adalah subjek yang terus berkembang (Chen dan Liu, 2005) di mana digunakan untuk mendapatkan pengetahuan (*knowledge discovery*) dari dalam *database* yang mengacu pada keseluruhan proses mulai dari *low-level* data sampai ke *high-level* data. *Data mining* adalah proses untuk menemukan suatu pola dan *trend* dari data. Tujuan dari proses ini adalah menyusun data yang berjumlah besar untuk mendapatkan informasi baru. *Data mining* juga dapat berarti eksplorasi dan analisis data secara otomatis atau semi otomatis dari data yang jumlahnya besar untuk mendapatkan pola dan aturan yang bermakna bagi penggunanya (Fayyad, U., Shapiro, G.P. dan Smyth, P. (1996). Chang (2006) dan Yu dan Chen (2001) menjelaskan bahwa proses data mining terdiri atas enam proses kerja utama yang dapat dilihat pada Gambar 2.



Gambar 2. Proses kerja data mining.

Sumber: Chang, Chan-Chine and Chen, Ruey-Shun, Using data mining technology to solve classification problems A case study of campus digital library, *The Electronic Library* Vol. 24 No. 3, 2006 pp. 307-321, Emerald Group Publishing Limited

*Data mining* adalah proses mencari *trend* dan pola dalam data. Tujuan dari proses ini adalah untuk menyortir besar jumlah data dan menemukan informasi baru. Manfaat dari *data mining* adalah pengetahuan baru ditemukan dan hasil ditindaklanjuti.

Tujuan dari *data mining* adalah untuk mengekstrak informasi yang berharga dan baru dari data yang ada. Secara umum, data mining meliputi fungsi utama sebagai berikut: klasifikasi, clustering, estimasi, prediksi, dan deskripsi.

Teknologi *data mining* dapat dibagi antara teknologi tradisional dan halus. Statistik analisis merupakan perwakilan dari teknologi tradisional. Adapun teknologi *data mining* halus adalah semua jenis kecerdasan buatan yang digunakan, yang lebih sering digunakan adalah jenis pohon keputusan, jaringan syaraf, algoritma genetika, logika *fuzzy* dan aturan induksi. Penggunaan berbagai jenis aplikasi yang berbeda sering dapat menyebabkan hasil yang berbeda.

Han dan Kamber (2001), dan Buja dan Lee (2001), model *data mining* yang umum digunakan adalah:

### ***Classification***

Mengklasifikasikan item data ke dalam salah satu dari beberapa standar kategoris kelas.

### ***Regression***

Memetakan sebuah item data ke sebuah variabel nilai prediksi nyata.

### ***Clustering***

Memetakan sebuah item data ke dalam salah satu dari beberapa *cluster*, dimana *cluster* mengelompokkan item data berdasarkan matrik kesamaan atau model kepadatan probabilitas.

### ***Rule generation***

Ekstrak aturan klasifikasi dari data.

### ***Discovering association rule***

Asosiasi menggambarkan hubungan antara berbagai atribut.

### ***Summarization***

Memberikan gambaran solid untuk subset data.

### ***Dependency modelling***

Menggambarkan dependensi yang signifikan antara variabel.

### **Star Schema**

*Star schema* merupakan paradigma modeling yang paling banyak digunakan, di mana di dalamnya mengandung antara lain sebuah tabel pusat yang besar tanpa adanya *data redundancy* di dalamnya, yang biasa disebut dengan *fact table* (tabel fakta). Selain itu, di dalam *Star schema* juga mengandung satu set tabel yang lebih kecil, yang biasa disebut dengan tabel dimensi. Berikut adalah beberapa karakteristik utama dari *Star schema* antara lain: Pusat dari *Star schema* adalah tabel fakta,

Tabel fakta berisi indikator-indikator kinerja pokok, Obyek-obyek informasi dan waktu adalah kunci utama tabel fakta, Tabel-tabel yang ada di sekeliling tabel fakta adalah tabel dimensi, Tabel dimensi berisi data mengenai objek-objek informasi atau waktu, Tabel fakta dan tabel dimensi direlasikan dengan *key* yang ada, *Star scheme* diimplementasikan menggunakan teknologi *relational database*.

## METODE

Tahapan yang dilakukan untuk desain dan implementasi *data warehouse* dan *data mining* pada kasus Analisis Peminatan Studi Siswa terdiri atas: (1) Analisis data yang terkait dengan kegiatan akademik siswa; (2) Mendesain dan membangun *data warehouse*; (3) Implementasi teknik *data mining*.

## HASIL DAN PEMBAHASAN

### Pengumpulan dan Analisis Data

#### Mendeskripsikan data yang digunakan:

Data yang digunakan berasal dari proses registrasi siswa, perhitungan nilai siswa dan *record* akademik siswa.

#### Analisis *attribute*:

Melakukan analisis terhadap *record* akademik siswa untuk menentukan *attribute* yang akan digunakan dalam membangun *data warehouse*.

#### Data *Pre-processing*:

Mendeskripsikan proses yang terkait dengan pengumpulan data yang akan digunakan untuk membangun *data warehouse*, penggunaan teknik *data mining*, proses tersebut juga meliputi *data cleaning*, *data integration*, *data transformation* dan *data reduction*.

#### Mendesain dan Membangun *Data Warehouse*

Pada tahap desain *data warehouse* ini juga termasuk di dalamnya mendesain arsitektur *data warehouse*, merencanakan sumber data dan *dimensional data modeling* dalam bentuk *Star schema*.

#### Merencanakan Sumber Data

Data yang digunakan berasal dari proses registrasi siswa, perhitungan nilai siswa dan *record* akademik siswa.



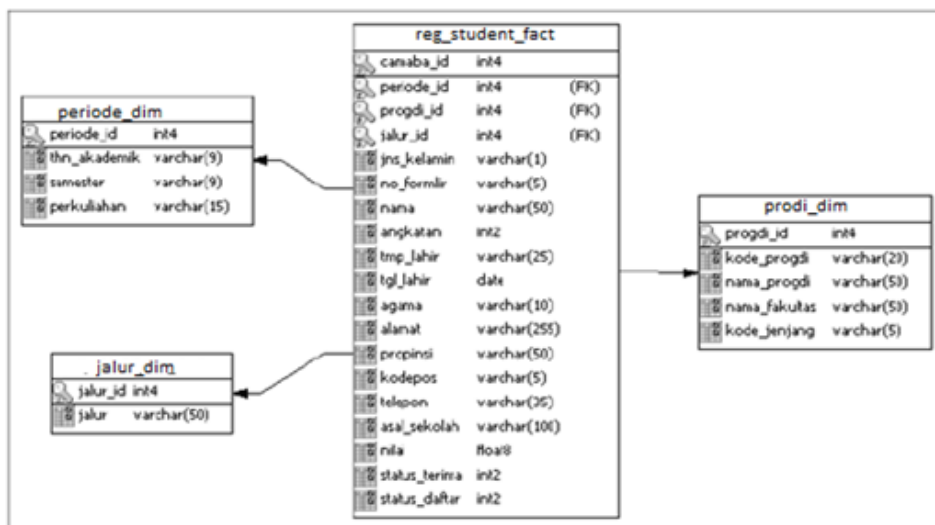
Gambar 3. Arsitektur Logikal Data Warehouse.

### Dimensional Data Modeling

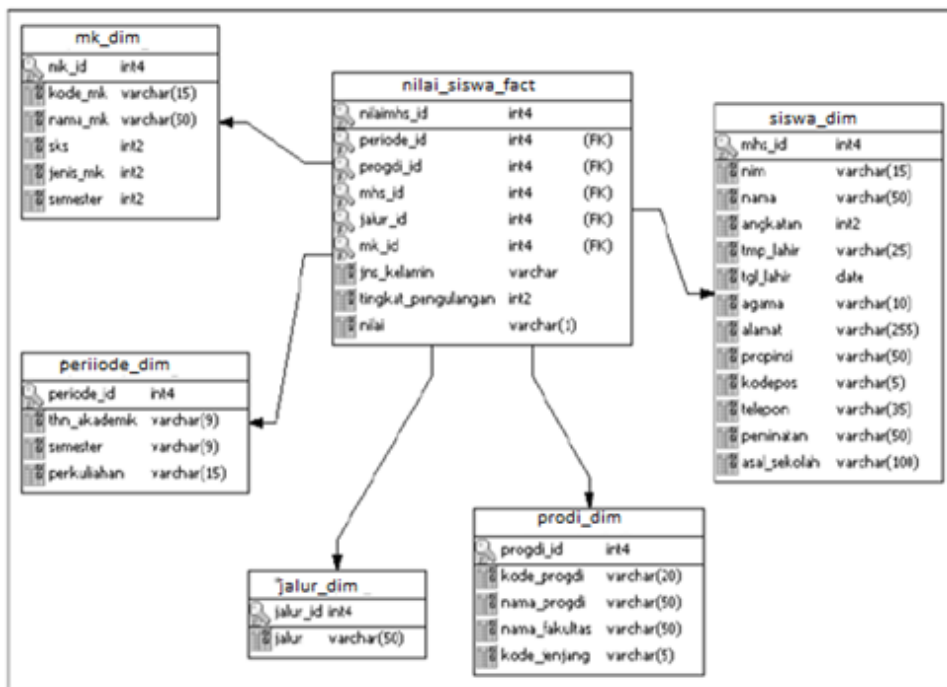
Skema yang digunakan untuk membuat data model adalah *Star schema*, di mana terdapat satu *fact table* dan beberapa *dimension table*. Menggunakan *Star schema* memungkinkan menganalisis data dari berbagai dimensi. *Fact table* merupakan tabel yang digunakan untuk menghubungkan data yang terkait digunakan dalam aktifitas evaluasi akademik siswa. *Dimension table* adalah tabel yang berisi data yang menampilkan pandangan dari berbagai sudut pandang. Dengan adanya *dimension table* memungkinkan untuk membuat laporan yang bersifat dinamis dari berbagai sudut pandang.

### Mining Data

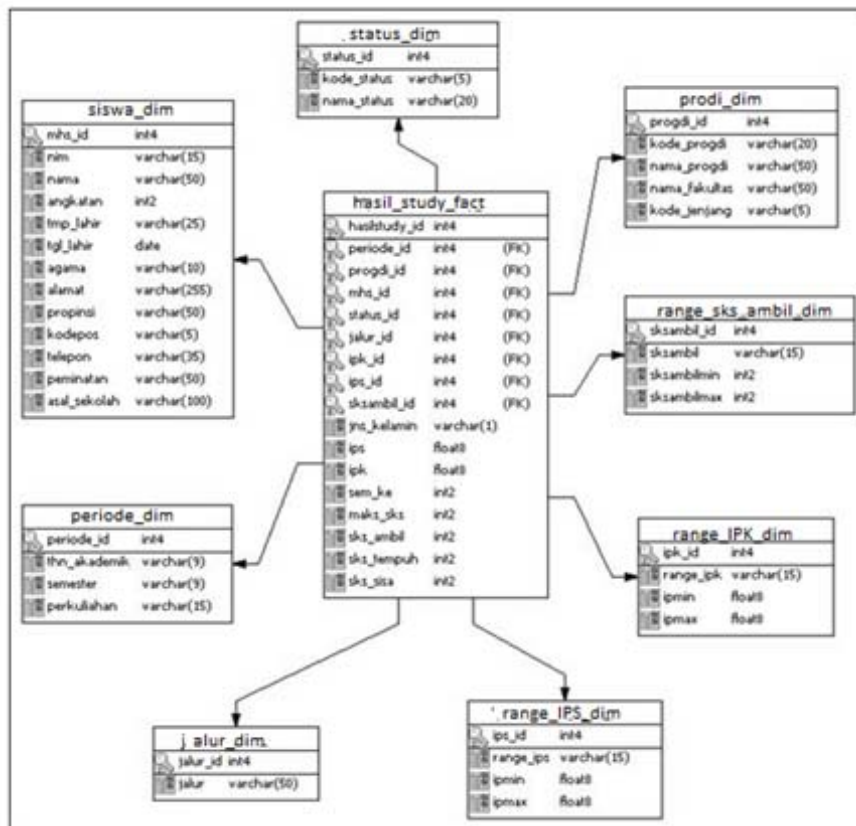
Teknik *data mining* yang digunakan untuk menjawab kasus di tulisan ini adalah *classification and association rules*. *Classification* digunakan untuk memprediksikan suatu *categorical class labels (discrete atau nominal)* dengan mengklasifikasikan data berdasarkan *training set* dan *values (class labels)* pada *attribute* yang digunakan untuk mengklasifikasikan data yang baru. Menggunakan teknik *classification and association rules* dapat dilihat kelompok siswa dan pola yang ada terkait dengan pemilihan program peminatan studi siswa.



Gambar 4. Star schema registrasi siswa.



Gambar 5. Star schema perhitungan nilai siswa.



Gambar 6. Star schema akademik siswa.



## Modeling Data Mining

Membuat model *data mining* yang terkait dengan informasi yang dibutuhkan dan karakteristik data yang ada pada *data warehouse*. Program peminatan yang digunakan pada tulisan ini adalah program peminatan yang ada di jurusan Sistem Informasi terdiri atas: Corporate IS (CIS), E-Business (EBIS), Business Intelligence (BI) dan Applied Database (AD). Data yang digunakan adalah data akademik siswa yang memiliki IPK lebih besar dari atau sama dengan 2.75 agar diperoleh pola siswa yang berhasil di dalam studinya terkait dengan program peminatan yang dipilihnya. Karakteristik data yang digunakan pada proses *data mining* dapat dilihat pada Table 1. Dari data tersebut kemudian dicari relasi antara nilai yang dicapai dari tiap mata kuliah yang merupakan persyaratan tiap program peminatan. Program peminatan yang dipilih berfungsi sebagai *output* dan nilai tiap mata kuliah persyaratan sebagai *input*.

Tabel 1  
*Data Mining Model*

No	Description	Type	Remark
1	Concentration	Discrete	Output
2	English I	Discrete	Input
3	Management	Discrete	Input
4	Introduction to Business	Discrete	Input
5	Information System Concepts	Discrete	Input
6	Introduction to Information Technology	Discrete	Input
7	Algorithm and Object Oriented Programming Methods	Discrete	Input
8	Accounting	Discrete	Input
9	Information and Business Process	Discrete	Input
10	Information Systems Development	Discrete	Input
11	Human Computer Interaction	Discrete	Input
12	Programming I	Discrete	Input
13	Programming I Lab	Discrete	Input
14	Statistics	Discrete	Input
15	Marketing Management	Discrete	Input
16	Information System in Management	Discrete	Input
17	Information System Analysis and Design	Discrete	Input
18	Programming II	Discrete	Input
19	Programming II Lab	Discrete	Input
20	English II	Discrete	Input
21	Business Mathematics	Discrete	Input
22	Introduction to Database System	Discrete	Input
23	Computer Network	Discrete	Input
24	Advanced Information System Analysis and Design	Discrete	Input
25	Web-based Programming	Discrete	Input
26	Web-based Programming Lab	Discrete	Input

## Evaluasi Hasil Data Mining

Pada tulisan ini *minimum support* yang digunakan adalah 10%. *Support* adalah probabilitas suatu transaksi berisi  $X \rightarrow Y$ . Total data yang digunakan adalah 100 data. Hasil yang diperoleh adalah: Karakteristik pola siswa yang memilih program peminatan CIS:

Management = B dan Introduction to Information Technology = B

Information Systems Concepts = B dan Object Oriented Programming Algorithm and Methods = C

Information Systems Development = B dan Programming I = C

Information Systems in Management = A dan Programming II = C  
 Information Systems Analysis and Design = C and Introduction to Database Systems = C  
 Business Mathematics = C dan Advanced Information Systems Analysis and Design = C

Karakteristik pola siswa yang memilih program peminatan EBIS:

Management = B dan Introduction to Information Technology = B  
 Introduction to Business = B dan Object Oriented Programming Algorithm and Methods = C  
 Information and Business Process = B dan Programming I = C  
 Information Systems Analysis and Design = A dan Programming II = C  
 Business Mathematics = B dan Advanced Information Systems Analysis and Design = B  
 Introduction to Database Systems = B dan Web-based Programming = C

Karakteristik pola siswa yang memilih program peminatan AD:

English I = B dan Object Oriented Programming Algorithm and Methods = A  
 Introduction to Business = B dan Introduction to Information Technology = A  
 Human Computer Interaction = A dan Web-based Programming = B  
 Information Systems Analysis and Design = A dan Programming II = A  
 Introduction to Database Systems = A dan Programming I = A  
 Advanced Information Systems Analysis and Design = B dan Computer Network = B

Berdasarkan pola yang ada di atas terlihat karakteristik bahwa jika nilai yang didapat untuk mata kuliah Introduction to Database Systems = A dan Programming = A maka siswa cenderung untuk memilih program peminatan Applied Databases, sedangkan siswa yang memiliki nilai mata kuliah Introduction to Databases Systems = C dan Programming = C cenderung memilih program peminatan CIS atau EBIS. Hal ini didukung juga bahwa mata kuliah Introduction to Database dan Programming terkait erat dengan program peminatan Applied Databases. Siswa yang memiliki nilai mata kuliah Information Systems Analysis and Design = A dan Programming II = C lebih condong memilih program peminatan EBIS, hal ini didukung juga bahwa mata kuliah Information Systems Analysis and Design terkait erat dengan program peminatan EBIS. Siswa yang memiliki nilai mata kuliah Information Systems in Management = A dan Programming II = C lebih condong memilih program peminatan CIS. Pada tulisan ini tidak melibatkan program peminatan BI.

Tabel 2

Hasil tes menggunakan *classification based association rules*

Rule	Cover %	Conf %	Cover Count	Sup Count	Sup %
Introduction to Database Systems = A dan Programming I = A → class AD	72%	39.9%	72	28	28%
Information Systems Analysis and Design = A dan Programming II = A → class AD	83%	60.24%	83	50	50%
English I = B dan Object Oriented Programming Algorithm and Methods = A → class AD	86%	65.11%	86	56	56%
Information Systems Analysis and Design = A dan Programming II = C → class EBIS	76%	51.32%	76	39	39%
Information and Business Process = B dan Programming I = C → class EBIS	73%	47.95	73	35	35%
Business Mathematics = B dan Advanced Information Systems Analysis	77%	57.14%	77	44	44%

and Design = B → class EBIS					
Information Systems in Management = A dan Programming II = C	72%	51.39%	72	37	37%
→ class CIS					
Information Systems Development = B dan Programming I = C	79%	64.55%	79	51	51%
→ class CIS					
Information Systems Analysis and Design = C dan Introduction to Database Systems = C	87%	83.90%	87	73	73%
→ class CIS					

*Support Count* adalah frekuensi munculnya suatu *itemset* dan jumlah transaksi yang mengandung *itemset* tersebut.

## PENUTUP

*Data Warehouse* dapat digunakan untuk menyimpan data dalam jumlah besar dan juga berpotensi untuk mendapatkan sudut pandang baru dari sebaran data, dan memungkinkan untuk memberikan pelaporan dan jawaban dari pertanyaan pengguna yang bersifat ad hoc termasuk juga untuk melakukan analisis terhadap *record* akademik siswa untuk mendukung pengambilan keputusan pemilihan program peminatan siswa. Proses transformasi dari *database* yang menyimpan data transaksi akademik siswa mampu memindahkan data penting yang berguna bagi universitas untuk melakukan analisis. *Star schema* yang telah dirancang dapat memenuhi fungsinya sebagai suatu tempat penampungan data yang memungkinkan para pengambil keputusan untuk pergi hanya ke satu tempat untuk mengakses seluruh data yang ada tentang suatu masalah

Implementasi *data warehouse* dan teknik *data mining* pada universitas dapat membantu mengidentifikasi karakteristik pola siswa dalam mengambil program peminatan yang ada di jurusan Sistem Informasi. Hasilnya adalah diperoleh karakteristik pola siswa yang memilih program peminatan CIS, EBIS dan Applied Databases dilihat dari prestasi akademiknya, yang pada akhirnya dapat membantu siswa lain dalam mengambil keputusan.

## DAFTAR PUSTAKA

- Azimah, A., Suyahco, D., Giri, Y. (2007). Implementasi Data Warehouse untuk Menunjang Kegiatan Akademik. *Seminar Nasional Sistem dan Informatika*.
- Buja, A., Lee, Y. (2001). Data mining criteria for tree-based regression and classification. *International Conference on Knowledge Discovery and Data Mining*, 27-36.
- Chang, C., Chen, R. (2006). Using data mining technology to solve classification problems: A case study of campus digital library. *The Electronic Library*, 03, 307-321.
- Chen, S.Y., Liu, X. (2005). Data mining from 1994 to 2004: an application-oriented review. *International Journal of Business Intelligence and Data Mining*, 01 (01), 4-11.

- Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996), The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39 (11), 07-34.
- Han, J., Kamber, M. (2001), *Data Mining: Concepts and Techniques*. San Mateo, CA: Morgan Kaufmann.
- Mallach, E. G. (2000). *Decision Support and Data Warehouse Systems*. United States: McGraw-Hill
- Wah, T. Y., Peng, N. H., Hok, C. S. (2007). Building Data Warehouse. *Proceedings of the 24th South East Asia Regional Computer Conference*.
- Yu, S.C., Chen, R.S. (2001). Developing an XML Framework for an Electronic Document Delivery System. *The Electronic Library*, 19 (20), 102-110.