

Penggalian Informasi Potensial dari Basis Data di Perguruan Tinggi

Sri Andayani
Jurusan Pendidikan Matematika FMIPA UNY

Abstrak

Basis data di perguruan tinggi menyimpan banyak informasi yang dapat lebih dioptimalkan pengkajiannya guna memprediksikan potensi perguruan tinggi. Teknik penggalian data, *data mining*, yang merupakan teknik penggalian potensi informasi yang tersembunyi dari sebuah basis data yang besar, dapat dimanfaatkan untuk kepentingan di atas. Teknik tersebut awalnya banyak dipakai dalam dunia bisnis, untuk memprediksikan keuntungan bisnis. Dalam perkembangannya, *data mining* banyak dipakai dalam berbagai bidang, akan tetapi masih jarang yang menggunakannya dalam bidang pendidikan.

Penerapan *Data mining* di perguruan tinggi dapat memberikan banyak manfaat sebagai dasar pengambilan langkah strategis peningkatan kualitas dan mutu input, proses, manajemen dan output perguruan tinggi.

Kata kunci: Basis data, data mining, perguruan tinggi.

Pendahuluan

Banyak komponen yang terkait dunia pendidikan, khususnya di perguruan tinggi, yang mempunyai data-data yang hanya diakumulasi tanpa kemudian dievaluasi atau ditindaklanjuti. Berbagai data akademik dan non akademik, baik yang terkait dengan data input, proses, maupun output dari perguruan tinggi yang tidak diolah lebih lanjut, ataupun dikaji secara mendalam untuk mengetahui apakah ada keterkaitan atau pola-pola tertentu dari data-data tersebut.

Sebuah perguruan tinggi dalam menjalankan fungsinya, selayaknya telah mengumpulkan ribuan bahkan jutaan data yang terkait dengan ketiga hal tersebut. Karenanya, basis data di sebuah perguruan tinggi akan mempunyai skala yang besar dan luas. Semisal data input, perguruan tinggi mempunyai mekanisme untuk menjarang mahasiswa baru, sedemikian sehingga mahasiswa yang diterimanya adalah yang paling berkualitas. Penjarangan calon mahasiswa baru tersebut tentu saja menghasilkan skor tes masuk, maupun data kuantitatif lain. Namun demikian sangat jarang dilakukan pengkajian yang lebih komprehensif terhadap data calon mahasiswa baru tersebut, dan melihat keterkaitan dengan unsur-unsur yang lain.

Salah satu tantangan terbesar dalam perguruan tinggi adalah bagaimana memprediksikan potensi perguruan tinggi di masa depan, termasuk kualitas input, proses dan outputnya. Mengantisipasi derasnya arus informasi, maka data-data yang ada di perguruan tinggi seharusnya lebih dioptimalkan pemanfaatannya, yang dapat menjawab prediksi yang dibutuhkan. Prediksi tersebut akan sangat berguna jika disertakan dalam pengambilan kebijakan yang menyangkut input, proses (termasuk di dalamnya manajemen) maupun output perguruan tinggi. .

Salah satu cara untuk mengatasi masalah tantangan yang berkaitan dengan memprediksikan kualitas perguruan tinggi adalah dengan menganalisis data-data yang ada menggunakan teknik *data mining*. *Data mining* telah banyak menunjukkan kesuksesannya dalam memprediksikan keuntungan di dunia bisnis, akan tetapi masih jarang digunakan untuk memprediksikan keuntungan dalam dunia pendidikan .

Data Mining dan keunggulannya

Beberapa definisi tentang *data mining* mengarahkan *data mining* sebagai sebuah teknik mengeksplorasi data yang sebelumnya tidak penting menjadi sebuah informasi yang sangat berharga. Dalam http://en.wikipedia.org/wiki/Data_mining, *data mining* didefinisikan sebagai proses ekstraksi sebuah informasi yang implisit, sebelumnya belum diketahui, dan sangat potensial manfaatnya dari sebuah data. Definisi kedua menyebutkan *data mining* sebagai ilmu pengekstrakan informasi yang sangat berguna dari sejumlah besar data atau basis data. Merceron dan Yacef berpendapat *Data mining* berhubungan dengan pencarian pola baru dalam sejumlah besar data (http://www.it.usyd.edu.au/~kalina/publis/merceron_yacef_aied05.pdf).

Beberapa manfaat yang dapat diberikan oleh *data mining*, yang dilakukan terhadap sebuah basis data yang besar dan berkualitas, adalah sebagai berikut:

- (1). *Data mining* dapat memprediksi kecenderungan data dan tingkah lakunya. *Data mining* dirancang dapat menemukan informasi prediktif dari basis data yang besar secara otomatis.
- (2). Menemukan pola yang sebelumnya tidak diketahui. *Data mining* menjelajahi kandungan basis data dan secara otomatis akan menemukan pola-pola data yang sebelumnya tersembunyi.

Data mining adalah teknologi baru yang handal dan berpotensi besar guna membantu sebuah organisasi/perusahaan/lembaga menjadi lebih memperhatikan informasi yang paling penting dari basis data yang dimiliki. *Data mining* memprediksikan kecenderungan yang akan terjadi di masa depan, sehingga memungkinkan para pengambil kebijakan lebih proaktif dalam mengambil kebijakan dengan didasari informasi yang akurat. Thearling (<http://www.thearling.com/text/dmwhite/dmwhite.htm>.) menyebutkan bahwa *data mining* telah

memberikan kinerja yang melampaui teknik pengambilan keputusan yang ada sebelumnya.

Lebih lanjut, Thearling menyatakan *data mining* dapat menjawab pertanyaan-pertanyaan penting seputar dunia bisnis yang akan memakan waktu lama jika dikerjakan dengan cara tradisional. Teknik tersebut membongkar kandungan informasi tersembunyi dari sebuah basis data, menemukan prediksi informasi yang mungkin lepas dari pengamatan para ahli sekalipun.

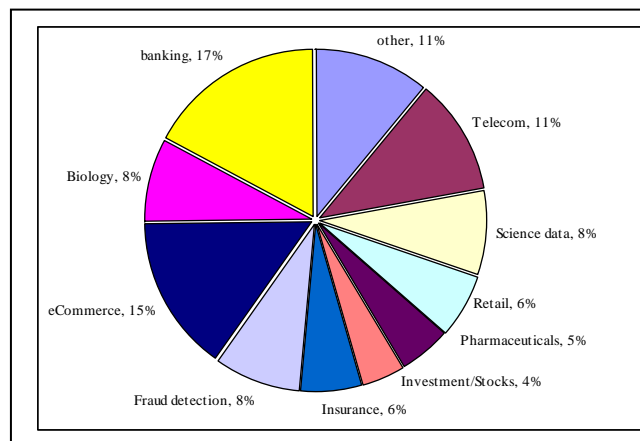
Data mining sangat sesuai diterapkan dalam komunitas bisnis dikarenakan adanya dukungan dari 3 (tiga) teknologi yang dewasa ini telah matang. Ketiga teknologi yang dimaksud adalah (1). Teknologi basis data yang mampu melakukan pengelolaan data dalam jumlah besar, (2). Teknologi komputer multiprocessor yang handal, dan (3). Algoritma *data mining*.

Beberapa alasan munculnya pemanfaatan *data mining* dalam berbagai bidang adalah (Thomas, http://airpo.binghamton.edu/conference/jan2004/Thomas_data_mining.pdf): (1). Sejumlah besar data disimpan dalam basis data, (2). Basis data tersebut terlalu besar untuk dianalisis dengan menggunakan teknik statistika secara tradisional, (3). Pengidentifikasian pola data dalam basis data akan sangat membantu menentukan hal-hal yang akan menjadi keuntungan dan ketidakuntungan organisasi, serta pelanggan yang memberikan kontribusi terbesar.

Aplikasi Data Mining

Data mining banyak diaplikasikan dalam berbagai bidang, diantaranya dalam bidang bisnis, dan riset-riset di laboratorium ilmu pengetahuan. Salah satu penerapan *data mining* dalam bidang bisnis adalah untuk mengidentifikasi target pemasaran yang dilakukan melalui promosi lewat surat. Cara tersebut ditempuh guna memaksimalkan pengembalian investasi dalam pengiriman surat promosi bisnis di periode berikutnya.

Diagram dalam gambar 1 berikut menunjukkan persentase penerapan *data mining* dalam beberapa bidang, yang dihasilkan melalui poling di www.kdnuggets.com (Pal & Mitra, 2004:13).



Gambar 1. Bidang-bidang aplikasi *data mining*

Perspektif dalam *data mining*

Data mining merupakan perpaduan beberapa konsentrasi riset yang berbeda, yakni *machine learning*, *pattern recognition*, statistik, basis data, kecerdasan buatan, penalaran, sistem pakar, visualisasi data dan komputasi performa tinggi (Pal & Mitra, 2004:14). Meskipun demikian, pada dasarnya *data mining* dapat dipandang dari 3 (tiga) perspektif yang berbeda, yaitu perspektif basis data, statistik dan *pattern recognition*.

Kumpulan data yang dikelola dalam sistem manajemen basis data adalah obyek yang akan ditelaah dengan teknik *data mining*. Penggalan data yang tersembunyi dalam basis data dengan menggunakan *data mining* kadang dianggap sama dengan memanfaatkan *query* dalam bentuk yang lebih sempurna.

Istilah *data mining* telah muncul dalam ilmu statistik lama sebelum dikenal di dalam ilmu komputer. Perspektif statistik memandang *data mining* sebagai eksplorasi analisis data dari sejumlah besar data yang kompleks, yang dilakukan secara otomatis dengan bantuan teknologi komputer. Salah satu piranti statistik yang digunakan dalam *data mining* adalah pemodelan data (*clustering*, *regresi*).

Perspektif ketiga adalah *pattern recognition* (pengenalan pola). Aspek yang membuat *pattern recognition* sangat tepat digunakan dalam *data mining* adalah kemampuannya dalam belajar atau melakukan induksi. *Pattern recognition* secara otomatis menganalisis data dan mengidentifikasi hubungan antara atribut dan entitas dalam basis data untuk membentuk model yang memudahkan pengkajian keterkaitan antara atribut dengan kelasnya.

Berangkat dari perkembangannya yang merupakan perpaduan dari berbagai disiplin ilmu, *data mining* dapat dilakukan dengan menggunakan teknik yang sangat bervariasi. Beberapa teknik yang sering digunakan adalah: (1). Jaringan syaraf tiruan, (2). Pohon keputusan (*decision tree*) yang melambangkan serangkaian keputusan, (3). Algoritma genetika, (4). Metode tetangga terdekat, (5). Induksi aturan (*rule induction*)

Penerapan *data mining* di Perguruan Tinggi

Data mining adalah metodologi analisis data yang penting yang telah sukses dalam banyak bidang. Biasanya *data mining* digunakan dalam dunia bisnis, dan masih sangat jarang digunakan dalam dunia pendidikan. Namun demikian menjadi sebuah tantangan untuk menggunakan *data mining* dalam bidang pendidikan. Perkembangan *world wide web* yang menghasilkan banyak data dengan banyak topik menjadikan *data mining* berkembang dengan pesat, dikarenakan teknik tersebut sangat relevan digunakan untuk mengeksplorasi informasi dari dunia maya tersebut.

Beberapa penerapan *data mining* dalam pendidikan antara lain adalah untuk memprediksi tipologi *learning outcome*, dan memprediksi alumni yang memberikan komitmen paling menjanjikan.

Penerapan lainnya adalah untuk menganalisa aktivitas dalam dunia pendidikan misalnya untuk mengevaluasi aktivitas, sistem pembelajaran, dan membentuk model interaksi antara mahasiswa dengan sistem. *Data mining* juga telah digunakan sebagai metode untuk mengekstraksi penilaian (*assessment*) jangka pendek sehingga dapat diketahui rata-rata waktu mengerjakan tugas (Tanimoto, <http://www.educationaldatamining.org/UM2007/Tanimoto.pdf>).

Data mining telah digunakan untuk mengeksplorasi data alumni, menemukan alumni manakah yang berpotensi untuk memberikan donasi yang besar bagi almamater. Selain itu Merceron dan Yacef juga mengevaluasi model pembelajaran yang menarik bagi mahasiswa dan dosen dalam perspektif pedagogi (http://www.it.usyd.edu.au/~kalina/publis/merceron_yacef_aied05.pdf).

Tabel 1 menunjukkan beberapa pertanyaan kritis dalam dunia perguruan tinggi yang ekuivalen dengan pertanyaan dalam lingkup bisnis pribadi, yang dapat dijawab oleh data mining (Luan, J. <http://www.pse.pt/Documentos/Data%20mining%20in%20higher%20education.pdf>.)

Tabel 1. Beberapa permasalahan dalam lingkup pribadi yang ekuivalen dengan masalah dalam perguruan tinggi

| Lingkup pribadi | Perguruan tinggi |
|---|--|
| Siapakah pelanggan yang paling memberikan keuntungan? | Siapakah mahasiswa yang paling banyak mengambil sks? |
| Siapakah pengunjung web site yang paling sering berkunjung? | Mahasiswa manakah yang paling sering mengulang matakuliah? |
| Siapakah pelanggan yang paling setia? | Siapakah mahasiswa yang paling gigih di universitas? |
| Pelanggan manakah yang cenderung menaikkan pesannya? | Siapakah alumni yang memberikan donasi paling besar? |

Luan, J berpendapat bahwa perguruan tinggi akan lebih menemukan aplikasi yang cukup besar dan luas bagi *data mining* daripada penerapannya dalam dunia bisnis (www.cabrillo.edu/services/pro/oir_reports/UCSFpaper.pdf). Hal tersebut dikarenakan perguruan tinggi mengemban 3 tugas utama yang secara intensif dapat menjadi wahana yang tepat bagi penerapan *data mining*. Ketiga hal tersebut adalah penelitian yang berhubungan dengan penemuan ilmu pengetahuan – ilmu pengetahuan baru, pembelajaran yang berkaitan dengan proses transfer pengetahuan, dan penelitian institusi dalam kaitannya dengan penerapan pengetahuan untuk pengambilan keputusan.

Banyak hal yang dapat diprediksikan informasi masa depannya dengan menggunakan *data mining*. Beberapa hal berikut merupakan garapan yang menarik dikaji dengan menggunakan *data mining*.

- a. Pelacakan alumni. Beberapa penelitian menunjukkan bahwa *data mining* memberikan hasil yang memuaskan dari data pelacakan alumni. Hasil tersebut akan memberikan kontribusi positif bagi pengembangan institusi di masa mendatang.
- b. Memprediksi kebutuhan *stakeholder*. Lulusan sebuah perguruan tinggi akan cepat diserap oleh pasar kerja jika kemampuannya sesuai dengan kebutuhan *stakeholder*. *Data mining* dapat menjawab tantangan prediksi kebutuhan *stakeholder* berdasarkan basis data yang dimiliki perguruan tinggi.
- c. Memprediksi tingkat kualitas calon mahasiswa baru. Proses penjarangan mahasiswa baru dari tahun ke tahun akan meninggalkan sejumlah data-data

- calon mahasiswa, yang dapat digunakan untuk melihat seperti apakah kualitas calon mahasiswa baru di sebuah perguruan tinggi di masa mendatang.
- d. Memprediksi tingkat kualitas lulusan. Proses pembelajaran yang terjadi di perguruan tinggi membuat terkumpulnya data-data akademik dari mahasiswa, yang jika dikaji lebih mendalam dapat dimanfaatkan untuk mengetahui pola kualitas lulusan perguruan tinggi.
 - e. Tingkat serapan pasar kerja. Data alumni perguruan tinggi yang telah bekerja dari tahun ke tahun, selayaknya menjadi perhatian bagi pengambil kebijakan di perguruan tinggi sehingga dapat ditentukan tingkat serapan pasar kerja terhadap lulusan perguruan tinggi tersebut.

Memperhatikan luasnya cakupan yang diemban oleh perguruan tinggi, tentu masih banyak informasi potensial yang dapat digali data-data yang ada di sebuah perguruan tinggi. Sudah saatnya perguruan tinggi memanfaatkan teknik pengambilan keputusan yang lebih akurat guna menjawab tantangan derasnya arus informasi di abad ini.

Penutup

Basis data yang dimiliki oleh suatu perguruan tinggi perlu digali lebih dalam dengan *data mining* untuk mendapatkan informasi berharga yang sangat bermanfaat bagi pengembangan perguruan tinggi di masa mendatang. Pola-pola data yang terdapat dalam basis data kemudian diubah menjadi sebuah model dan digunakan untuk memprediksi kecenderungan data yang ada dengan keakuratan yang tinggi. Sebagai dampaknya, diharapkan institusi lebih mudah mengalokasikan sumber daya yang dimilikinya secara bijaksana.

DAFTAR PUSTAKA

- Luan, Jing. *Data Mining Applications in Higher Education*.
<http://www.pse.pt/Documentos/Data%20mining%20in%20higher%20education.pdf>. Diakses tgl 20 Juli 2007
- Luan, Jing. *Mining as Driven by Knowledge Management in Higher Education, Persistence Clustering and Prediction*.
www.cabrillo.edu/services/pro/oir_reports/UCSFpaper.pdf. Diakses tgl 1 Agustus 2007

Merceron, A & Yacef, K. *Educational Data Mining: a Case Study*.
http://www.it.usyd.edu.au/~kalina/publis/merceron_yacef_aied05.pdf. diakses tgl
Juli 2007.

Pal, Shankar K & Mitra, Pabitra. 2004. *Pattern Recognition algorithms for data mining*.
CRC Press.

Tanimoto, Steven L. *Improving the Prospects for Educational Data Mining*.
<http://www.educationaldatamining.org/UM2007/Tanimoto.pdf>. Diakses tgl 20
Juli 2007.

Thearling, K. *An Introduction to Data Mining*.
<http://www.thearling.com/text/dmwhite/dmwhite.htm>. Diakses tgl 21 Juli 2007

Thomas, E. *Data mining: definition and decision tree examples*.
http://airpo.binghamton.edu/conference/jan2004/Thomas_data_mining.pdf.
Diakses tgl 20 Juli 2007

http://en.wikipedia.org/wiki/Data_mining