

EDUCATIONAL DATA MINING (KONSEP DAN PENERAPAN)

Fitri Marisa

Program Studi Teknik Informatika
Universitas Widyagama Malang
Jl. Borobudur No. 35 Malang (0341)492282
e-mail: fitrimarisa@widyagama.ac.id

ABSTRACT

Electronic Learning (E-Learning), especially Web-based learning or intelligent tutorial system has yielded information about the interaction between students and the system. This information is stored in a database in the form of a web log. Within a certain period it will be a large amount of information. From the collection of large amounts of data that can be explored using data mining methods to generate new patterns that can be useful for improving the effectiveness of computer-based learning process. This paper will discuss how data mining can be utilized to improve the effectiveness of computer-based learning process. In its application the data will be processed in three stages: the collection, transformation, and analysis. Then the techniques used in the analysis algorithm is Association rules, classification, and clustering.

Keywords: data mining, Educational data mining, E-Learning

PENDAHULUAN.

Di era teknologi informasi seperti saat ini tentunya ketersediaan data di segala bidang sangatlah melimpah. Dengan fenomena tersebut akan sangat berguna jika dipikirkan bagaimana membangun pola yang menjadikan data yang berlimpah tersebut menghasilkan data-data yang memiliki banyak manfaat bagi pemilik data itu sendiri maupun pihak eksternal yang membutuhkannya. Maka dikembangkanlah ilmu *Data Mining*. Dalam definisinya *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). Sependapat dengan Witten (2005) yang mengatakan bahwa “*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”.

Salah satu bidang yang akan diangkat dalam pembahasan makalah ini adalah

penerapan data mining dalam dunia pendidikan khususnya dunia pendidikan. Di dunia pendidikan tentunya terdapat banyak sumber yang menghasilkan data melimpah antara lain data kinerja pengajar, kinerja administrasi dan pelayanan, data rekam jejak alumni, data akademik, maupun data proses pembelajaran.

Mungkin selain yang disebutkan diatas masih banyak sumber data yang bisa digali. Namun pembahasan ini akan menfokuskan pada satu sumber yaitu proses pembelajaran dengan studi kasus pembelajaran berbantuan komputer (*computer aided learning system*). Sistem pembelajaran berbantuan komputer dapat diimplementasikan sebagai sistem tutorial berbasis web (*web-based tutoring tool*) [Merceron, 2005]. Dalam mekanismenya bahwa dalam tutorial berbasis, terdapat interaksi antara pebelajar dengan sistem yang disimpan dalam database baik berupa basis data SQL maupun *web log*. Dengan rentang waktu yang panjang maka data yang dihasilkan akan semakin berlimpah, maka dari berlimpahnya data tersebut berpotensi untuk digali sehingga menghasilkan pola-pola baru dari data yang bermanfaat untuk efektifitas proses pembelajaran. Pola penggalian data tersebut menggunakan ilmu Data Mining.

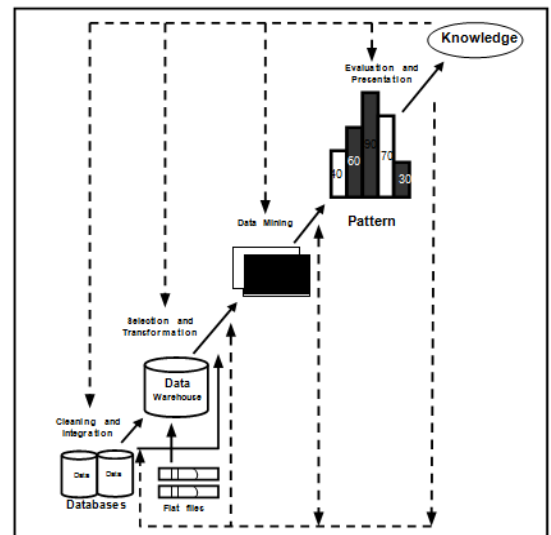
LANDASAN TEORI

1. Penerapan Educational Data Mining.

Sebenarnya *data mining* merupakan suatu langkah dalam *knowledge discovery in databases* (KDD). *Knowledge discovery*

sebagai suatu proses terdiri atas pembersihan data (*data cleaning*), integrasi data (*data integration*), pemilihan data (*data selection*), transformasi data (*data transformation*), *data mining*, evaluasi pola (*pattern evaluation*) dan penyajian pengetahuan (*knowledge presentation*). (Ayub, 2007: 22).

Pendapat lain mengatakan bahwa data mining terdiri dari 6 tahap yang terpola dalam gambar 1 sebagai berikut :



Gambar 1: Tahapan data mining Han: 2006)

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 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*),

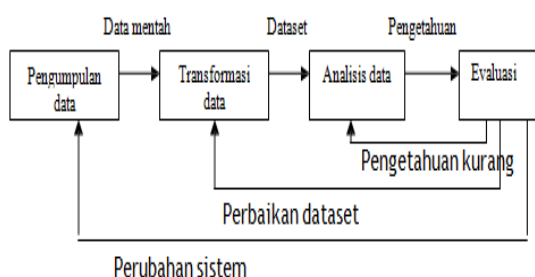
Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa 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)

Sedangkan menurut Nilakant (2004) “Kerangka proses *data mining* yang akan dibahas tersusun atas tiga tahapan, yaitu pengumpulan data (*data collection*), transformasi data (*data transformation*), dan analisis data (*data analysis*). Dari proses tersebut dapat digambarkan sebagai berikut:



Gambar 2: Kerangka proses Data mining
(adaptasi dari Ayub, 2007)

Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in.

(www.educationaldatamining.org)

diakses 7 Pebruari 2014 jam 13.40.

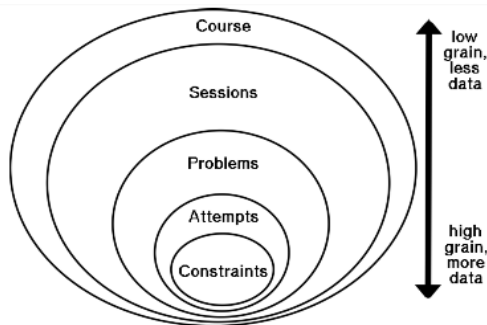
Dari definisi tersebut dikatakan bahwa Educational data mining adalah disiplin ilmu yang mendalami tentang pembangunan pola-pola yang unik yang dihasilkan dari proses pembelajaran yang berguna untuk memahami pebelajar secara lebih baik sehingga dapat diupayakan bagaimana merancang pola yang cocok.

Dari kerangka proses data mining yang digabungkan dengan pola educational data mining maka dapat dijabarkan tahapan proses data mining sebagai berikut:

A. Pengumpulan data.

Dalam kerangka ini proses dimulai dari pengumpulan data yang menghasilkan data mentah, kemudian ditransformasikan sehingga menghasilkan file yang dapat dibaca oleh tools data mining. Hasil dari transformasi tersebut akan dilakukan analisis, yang biasanya menggunakan analisis statistik ataupun visualisasi informasi. Hasil evaluasi pengetahuan yang dihasilkan *data mining* inilah dapat digunakan sebagai acuan kebutuhan pengetahuan yang lebih lengkap, perbaikan kumpulan data (dataset) atau perubahan pada sistem.

Berikut ini adalah menggambarkan interaksi pembelajaran dalam berbagai layer.



Gambar 3: Model interaksi pembelajaran.

Dari Gambar 3 digambarkan layer yang dengan lapisan mulai dari lapisan dalam yaitu Constrains (Aturan) – Attempts (Usaha)- Problems (soal)- Session (sesi)- Course (Materi). Suatu contoh pebelajar dalam melakukan tes yang pada saat pengerjaan sebelum pebelajar memutuskan menjawab soal yang dianggap benar tentunya ada kegiatan yang dilakukan seperti misal menjawab jawaban yang salah, kemudian direvisi kembali, berapa lama waktu yang dibutuhkan untuk menghasilkan jawaban yang benar, apakah siswa melakukan pelanggaran, bagaimana pola pelanggaran yang dilakukan, apakah siswa menjawab dengan salah, seperti apa pola kesalahan pebelajar. Hal-hal tersebut sedetail mungkin harus bisa direkam oleh sistem yang akhirnya dapat digunakan untuk menyusun representasi kognitif siswa atau disebut model siswa.

Hasil pengelompokan data tersebut dapat digambarkan taksonomi variable tentang usaha pebelajar dalam memecahkan pertanyaan, seperti yang dicontohkan pada gambar 4.

Kegiatan	Variabel	Keterangan
Persiapan tes	• umpan balik yang tersedia • banyaknya soal • banyaknya usaha • tingkat kesulitan soal • konteks soal	Informasi yang digunakan pebelajar sebelum mencoba menjawab soal
Waktu pelaksanaan	waktu yang diperlukan untuk menjawab soal	informasi tentang usaha menjawab
Evaluasi hasil	• aturan (relevan, dipenuhi, dilanggar) • tingkat umpan balik yang diminta dan solusi • permintaan melihat	informasi yang berhubungan dengan hasil dari suatu usaha.

Gambar 4: contoh Taksonomi variabel dari usaha menjawab soal (adaptasi dari Ayub, 2007)

Data mentah yang dihasilkan dari pengumpulan data, biasanya tersimpan dalam bentuk beberapa tabel basis data. Karena analisis data umumnya dilakukan terhadap suatu tabel tunggal, maka perlu dilakukan penggabungan (join) beberapa tabel yang relevan. Hasilnya adalah suatu struktur yang disebut dengan dataset, [Ayub, 2007]. Kemudian dataset ini digolongkan menjadi 2 yaitu kumpulan atribut (vertikal) dan kumpulan instan (horisontal).

Setiap atribut mempunyai tipe data, yang dapat berupa string, angka atau yang lainnya. Untuk nilai atribut berhingga, disebut atribut nominal. Instans adalah data yang dihasilkan dari suatu kejadian riil, yang dicatat dalam beberapa atribut.

	atribut-1	atribut-2	. . .	atribut-n
instans-1	$x_{1,1}$	$x_{1,2}$	. . .	$x_{1,n}$
instans-2	$x_{2,1}$	$x_{2,2}$	. . .	$x_{2,n}$
. . .	. . .	. . .	. . .	. . .
instans-m	$x_{m,1}$	$x_{m,2}$	. . .	$x_{m,n}$

Gambar 5: Format dataset (Adaptasi dari Nikalant: 2004)

B. Transformasi.

Dalam proses transformasi ini dapat dilakukan dengan beberapa cara yaitu filtrasi dan konversi atribut. Filtrasi yaitu mensoting atribut-atribut yang tidak relevan sehingga menghasilkan data yang lebih rekomended. Sedangkan konversi atribut adalah mengubah atribut yang bernilai kontinyu menjadi nominal, karena ada beberapa perangkat data mining yang tidak bisa menghitung nilai kontinyu.

Dalam konversi atribut ini terdapat dua cara yaitu teknik scanning yaitu melakukan penelusuran terhadap semua nilai kontinyu menjadi domain atribut nominal. Cara berikutnya adalah teknik binning yaitu mendefinisikan kumpulan kelas nominal untuk setiap atribut, kemudian menetapkan setiap nilai atribut ke dalam salah satu kelas. Misalnya, jika domain atribut numerik mempunyai nilai dari 0 sampai dengan 100, domain tersebut dapat dibagi menjadi empat bin (0..24, 25..49, 50..74, 75..100). Setiap nilai atribut akan dikonversi menjadi atribut nominal yang berkorespondensi dengan salah satu bin.

C. Analisis Data.

Proses analisis data dengan menerapkan teknik *data mining* dapat dilakukan melalui analisis statistik atau dengan pendekatan *machine learning*. Analisis data pembelajaran dengan pendekatan *machine learning* akan menggunakan tiga teknik, yaitu *association rules*, *clustering*, dan *classification* [Nilakant,

2004]

Menurut Huda (2010: 14) *Association rules* (aturan asosiasi) atau *affinity analysis* (analisis afinitas) berkenaan dengan studi tentang “apa bersama apa”. Sebagai contoh dapat berupa berupa studi transaksi di supermarket, misalnya seseorang yang membeli susu bayi juga membeli sabun mandi. Pada kasus ini berarti susu bayi bersama dengan sabun mandi. Karena awalnya berasal dari studi tentang *database* transaksi pelanggan untuk menentukan kebiasaan suatu produk dibeli bersama produk apa, maka aturan asosiasi juga sering dinamakan *market basket analysis*

Aturan asosiasi ingin memberikan informasi tersebut dalam bentuk hubungan “if-then” atau “jika-maka”. Aturan ini dihitung dari data yang sifatnya probabilistik (Santoso, 2007). Bentuk umum aturan dalam *association rule* adalah :

$$(X = x_i) \rightarrow (Y = y_i) \text{ [sup,conf]}$$

dengan :

$$X = \{ x_1, x_2, \dots, x_n \},$$

$$Y = \{ y_1, y_2, \dots, y_m \},$$

sup = probabilitas bahwa suatu instans dalam dataset mengandung $X \cup Y$, conf = probabilitas kondisional bahwa instans yang mengandung X juga mengandung Y.

Association rule merupakan salah satu metode *data mining* yang menjadi dasar dari berbagai metode *data mining* lainnya. tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (*frequent pattern mining*) menarik perhatian banyak peneliti

untuk menghasilkan algoritma yang efisien. Penting tidaknya suatu aturan assosiatif dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu prosentase kombinasi item tersebut. dalam *database* dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan assosiatif. Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan assosiatif yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat minimum untuk *confidence* (*minimum confidence*) (Pramudiono, 2007).

Ada beberapa algoritma yang sudah dikembangkan mengenai aturan asosiasi, namun ada satu algoritma klasik yang sering dipakai yaitu algoritma *apriori*. Ide dasar dari algoritma ini adalah dengan mengembangkan *frequent itemset*. Dengan menggunakan satu item dan secara rekursif mengembangkan *frequent itemset* dengan dua item, tiga item dan seterusnya hingga *frequent itemset* dengan semua ukuran.

Untuk mengembangkan *frequent set* dengan dua item, dapat menggunakan *frequent set item*. Alasannya adalah bila set satu item tidak melebihi *support minimum*, maka sembarang ukuran itemset yang lebih besar tidak akan melebihi *support minimum* tersebut. Secara umum, mengembangkan set dengan *fc-item* menggunakan frequent set dengan $k - 1$ item yang dikembangkan dalam langkah sebelumnya. Setiap langkah memerlukan sekali pemeriksaan ke seluruh isi *database*.

Dalam asosiasi terdapat istilah *antecedent* dan *consequent*, *antecedent* untuk

mewakili bagian “jika” dan *consequent* untuk mewakili bagian “maka”. Dalam analisis ini, *antecedent* dan *consequent* adalah sekelompok item yang tidak punya hubungan secara bersama (Santoso, 2007).

Assosiation Rules dapat dimanfaatkan untuk menemukan kesalahan yang sering terjadi pada saat pebelajar mengerjakan latihan soal. Suatu contoh jika pebelajar melakukan kesalahan X dan Z, maka mereka juga melakukan kesalahan Y, misalnya dengan *support* 40% dan *confidence* 55% akan ditulis sebagai

$$X \text{ and } Z \rightarrow Y [40\%, 55\%]$$

Aturan tersebut dapat dibaca sebagai berikut : dari 40% siswa yang melakukan kesalahan X dan kesalahan Z maka 55% diantaranya melakukan kesalahan Y.

Algoritma Assosiation Rules juga dapat menyatakan hubungan antara beberapa atribut yang berbeda, misalnya kesalahan X pada konsep A menimbulkan kesalahan Z pada konsep C, yang ditulis sebagai

$$X \text{ and } A \rightarrow Z \text{ and } C$$

Teknik *classification* bekerja dengan mengelompokkan data berdasarkan data training dan nilai atribut klasifikasi. Aturan pengelompokan tersebut akan digunakan untuk klasifikasi data baru ke dalam kelompok yang ada.

Classification dapat direpresentasikan dalam bentuk pohon keputusan (*decision tree*). Setiap node dalam pohon keputusan menyatakan suatu tes terhadap atribut dataset,

sedangkan setiap cabang menyatakan hasil dari tes tersebut. Pohon keputusan yang terbentuk dapat diterjemahkan menjadi sekumpulan aturan dalam bentuk IF condition THEN outcome.

KESIMPULAN

Education data mining adalah suatu teknik penerapan data mining dalam sistem pembelajaran berbantuan komputer. Pada prosesnya adalah diawali dengan pengumpulan data, yang dilanjutkan dengan transformasi data, dan diakhiri dengan analisis data. Dalam pengumpulan data, yang mesti didefinisikan adalah suatu model interaksi antara pebelajar sistem yang bertujuan untuk menghasilkan ketetapan data yang harus terekam dari suatu proses pembelajaran yang diamati. Kemudian model interaksi pebelajar-sistem tersebut dapat tersusun atas beberapa lapisan untuk memungkinkan penangkapan data yang lebih kaya. Transformasi data mengubah data mentah menjadi dataset yang siap dilakukan penghitungan selanjutnya (dianalisis). Transformasi dapat dilakukan pada instans dataset maupun pada atribut dari dataset melalui filtrasi ataupun konversi. Sementara itu analisis data hasil pembelajaran dapat dilakukan dengan menerapkan teknik algoritma *association rules*, *classification*, dan *clustering* untuk menghasilkan pengetahuan yang dapat membantu untuk proses pembelajaran selanjutnya.

DAFTAR PUSTAKA

- Ayub, Mewati. 2007. “Proses data mining dalam pembelajaran berbantuan komputer”. Vol 2. No. 1. Jurnal Sistem Informasi. Universitas Kristen Maranatha.
- Davies, and Paul Beynon, 2004, “*Database Systems Third Edition*”, Palgrave Macmillan, New York.
- Han, J. and Kamber, M, 2006, “*Data Mining Concepts and Techniques Second Edition*”. Morgan Kauffman, San Francisco.
- Merceron, A., Yacef, K. (2005). Educational Data Mining : a Case Study, http://www.it.usyd.edu.au/~kalina/publis/merceron_yacef_aied0
- Nilakant, K. (2004). Application of Data Mining in Constraint Based Intelligent Tutoring System, www.cosc.canterbury.ac.nz/research/reports/HonsReps/2004/hons_04
- Pramudiono, I., 2007, *Algoritma Apriori*, <http://datamining.japati.net/cgi-bin/indodm.cgi?bacaarsip&1172210143> Di
- Pramudiono, I. 2007. *Pengantar Data Mining : Menambang Permata Pengetahuan di Gunung Data*. <http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining.zip>.
- Santosa, Budi, 2007, “*Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*”, Graha Ilmu, Yogyakarta.
- Witten, I. H and Frank, E. 2005. *Data Mining : Practical Machine Learning Tools and Techniques Second Edition*. Morgan Kauffman : San Francisco