DATA MINING AS A TECHNIQUE TO ANALYZE THE LEARNING STYLES OF STUDENTS IN USING THE LEARNING MANAGEMENT SYSTEM

Eka Miranda
Department of Information System, Faculty of Computer, Binus University
Jl. KH Syahdan no.9, Palmerah Jakarta Barat 11480
ekamiranda@yahoo.com

ABSTRACT
Higher education is becoming a big business, with huge investments in IT technology supporting Learning Management System (LMS). LMS is a software application which is used by student in learning process. The purpose of this paper is to analyze the use of LMS based on activities and results of student learning and seeks to address the interface between individual learning that use LMS data to support decision-making and course design. A model of data warehouse is built to evaluate by means of a case study the usefulness validity of analyses performed. Data mining is considered the non-trivial extraction of implicit, previously unknown, and potentially useful information from data. This paper analyzes students' activities records using LMS then building a data warehouse, and data mining. The paper finds that after mining data, students can be classified into different groups according to the activities using LMS. The data mining results shows that students can be categorized into four clusters; each cluster has its own characteristics. Integration of data warehouse, data mining and LMS produce a framework that uses transactional data to be transformed into knowledge that can be used to maximize the teaching-learning process of lecturer and students.

Keywords: Learning Management System, Data-mining, Data-warehouse

1. INTRODUCTION

Development of Learning Management System (LMS) is intended to allow students helped in the learning process. It has become a common phenomenon; various applications of information systems developed in the business world, but not put to good use. LMS applications, which have been built can be used to evaluate students activities. One measure that can be used in evaluating the utilization of this application is an activity done by the students and lecturers use the application and the results obtained by users.

In using the LMS application, students are faced with the decision making process. The decision making process requires the provision of information to a user that can be used to facilitate decision-making. Users need to be facilitated by knowledge. Knowledge of learning management systems can be provided using mining data and design data warehouses. Research data mining and data warehouse design associated with the LMS has been done by Liezl van Dyk and Pieter Conradie.

The purpose of this paper is to analyze the use of LMS based on activities and results of student learning and seeks to address the interface between individual learning that use LMS data to support decision-making and course design.

This research used the following procedures using data mining and online analytical processing technology:

a. confirmation of the goals of data mining – determine the problems to be solved by data mining
b. data selection – select the data from library massive databases
c. data processing – data cleaning, error removal, and data format consistency
d. data transformation – format adjustment, the joining or division of data fields
e. data storage – storing the data in an appropriate data repository
f. data dredging – classifying, sorting, and aggregating data to discover patterns and rules in order to assist decision-makers in making vital decisions
g. User-relevant feedback – apply data mining mechanism and deliver query results to users: users then respond to the results.

2. LITERATURE REVIEW

2.1 Learning Management System (LMS)

Learning Management System (LMS) developed to facilitate independent student learning. LMS is software that is used to display the lecture material, manage the learning activities and web-based results. LMS supports the administration, delivery of content, assessment, tracking and monitoring, collaboration and communication. LMS users can be classified as lecturers, students, and administrators. Through the LMS lecturer can create a syllabus,

materials, manage files, create exam questions, quizzes, assignments, monitor student activity, giving the value, process value, interact with students through forums, and sending electronic mail and see the progress of students. Through the LMS students can access the syllabus, materials, tasks, send tasks, view the results of the assessment, do the problems, quizzes, managing files, interacting with other students and lecturers through the forums and sending electronic mail and group work. Through the LMS administrator can register the course, determine the faculty and courses, record and manage data backup websites.

2.2 Data Warehouse

Usually companies use databases in their daily operations such as recording transactions, administration of delivery of goods, inventory, payroll and other things that are commonly referred to as OLTP (online transaction processing). With the growing need for data analysis and to maintain excellence in competition, many companies are also building its own database which is specifically used to support decision-making process or also commonly referred to as OLAP (online analytical processing). Unlike OLTP queries that only use simple operations and repeatedly, for OLAP queries are usually more complicated, are ad hoc, and do not involve update data. OLAP also did not use data of daily operations for granted, but using data that has been summarized by the data model, called the data cube. Data cube is a multidimensional data presentation such as the type of goods, time, and location. The dimensions of the data cube can be made high rise; for example, the location dimension can be divided into cities, provinces and countries. With this user can easily get summary information from a wider dimension levels or general operation called roll-up contrast with drill-down operation, the user can draw information from a more detailed level dimension. The data cube is available in the data warehouse allows users to analyze data daily operation of the various points of view, and very useful for evaluating a business assumptions. The data warehouse itself periodically filled data from the OLTP after undergoing cleaning and data integration. Step build a data warehouse consists of:

1. Defining dimension tables
2. Defining fact tables
3. Creating a data warehouse schema

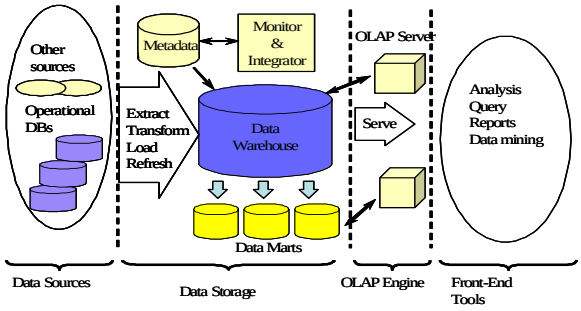

Figure 1. Data Warehouse: A Multi-Tiered Architecture

2.3 Data mining

Knowledge discovery in databases refers to the overall process of turning low-level data into high-level knowledge. An important step in the knowledge discovery in databases process is data mining. Data mining is the process of finding trends and patterns in data (Groth, R. (2000)). The objective of this process is to sort large quantities of data and discover new information. The benefit of data mining is to turn this newfound knowledge into actionable results. Data mining is the exploration and analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns and rules (Berry, M.J.A. and Linoff, G.S. (1997), (1999)). The work process of data mining is composed of eight primary tasks (Yu, S.-C. and Chen, R.-S. (2001)).

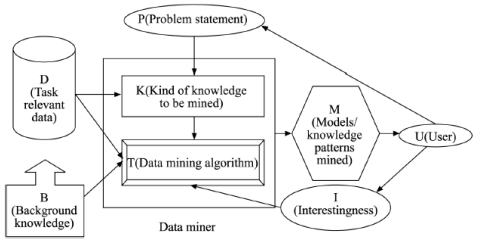

Figure 2. The work process of data mining

The goal of data mining is to extract valuable and new information from existing data. Data mining technology can be divided between traditional and refined technologies. Statistical analysis is representative of traditional technology.

A particular data-mining algorithm is usually an instantiation of the model search preference components. The most common model functions in current data mining practice to be the following:

a. Classification – classifies a data item into one of several predefined categorical classes
b. Regression – maps a data item into a real valued prediction variable

c. Clustering – maps a data item into one of several clusters, where clusters are natural groupings of data items based on similarity metrics or probability density models
d. Rule generation – extracts classification rules from the data
e. Discovering association rules – describes association relationships among different attributes
f. Summarization – provides a compact description for a subset of data
g. Dependency modeling – describes significant dependencies among variables
h. Sequence analysis – models sequential patterns, like time-series analysis.

3. RESEARCH METHODOLOGY

This paper analyzes students' activities records using LMS with the following techniques:
a. Data analysis
b. Building a data warehouse
c. Data mining

3.1 Data collection and analysis

Data collected and analyzed from Student Information System (SIS), which the system is used to registration process and allocation of the curriculum of courses and students. Quite a lot of data types, relationship between data are drawn by relational model. Therefore constructed relational relationship among the data (Van Dyk, Liezl, Conradie, Pieter (2007)), as can be seen in Figure 3. Figure 3. shows that the data warehouse will consist of tables Mata_Kuliah, Siswa, Pengajar, Mata_Kuliah, and Jurusan. Each table consists of various attributes in which each of these attributes must be defined with the appropriate field name, type, and size (Inmon, WH (2002)). Other tables are: TRMata_Kuliah and TRSiswa

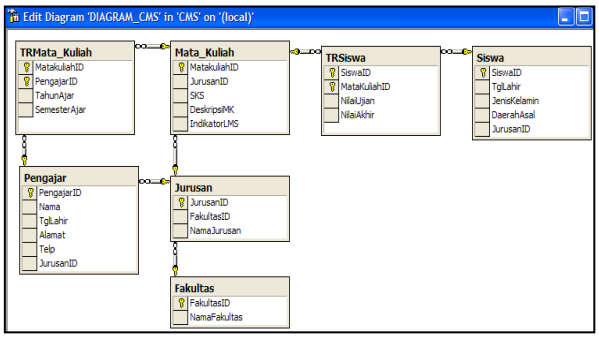

Figure 3. Relational database for SIS

3.2 Building a data warehouse

Development of data warehouses using star schema. Star scheme was chosen because this modeling paradigm most widely used, where it contains a large central table with no data redundancy in it, commonly called the fact table. In addition, in the star schema also contains a set of tables that are smaller, so-called dimension tables. Here are some of the main characteristics of star schemes include: The center of the star schema is the fact tables. Fact tables contain key; objects and time information is key to the fact tables. Tables that surround the fact table is dimension table; dimension table contains data about objects or time information. The fact table and dimension tables are related with the existing key. Star schema is implemented using relational database technology (Inmon, WH (2002)). SQL Server 2005 can be used to implement "click stream" and "aggregate click stream" schema. Figure 5. shows the star schema "click stream" and Figure 6. shows the "aggregate click stream." Figure 4. Shows the dimensions of which can be produced from dimension tables used in the star schema, dimension table are: Student, Mata_kuliah and results. The combination of dimensions resulting from the star schema is described in the form of dimensional data cube.

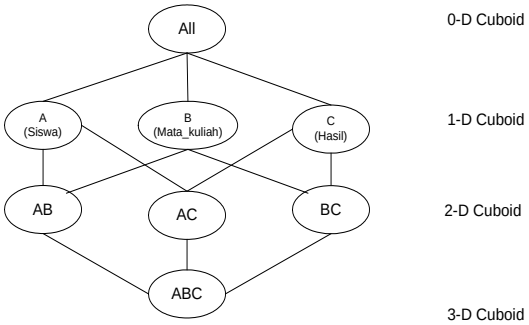

Figure 4. 3-Dimensional data cube for star schema

"Click stream" schema is a composite body, containing the action a user through a web browser (Van Dyk, Liezl, Conradie, Pieter (2007)). "Click stream" is actually in the form of web server logs, where each web server log records associated with a one-page incident. Each time a student or lecturer log into the LMS, participate in discussions, complete the electronic quizzes or reading electronic documents, then a transaction occurs (Van Dyk, Liezl, Conradie, Pieter (2007)). Data transactions are recorded according to the type and timing of activities.

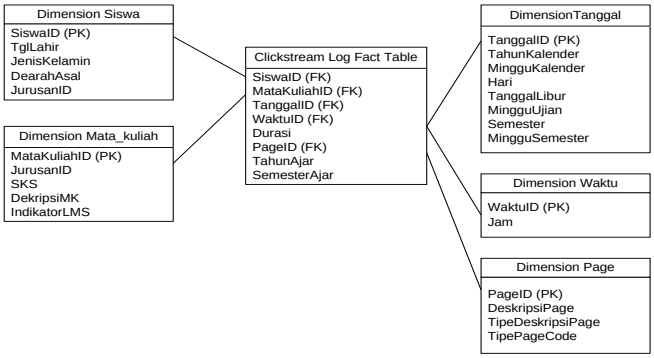

Figure 5. Star schema : Click stream

"Click stream" table is built using the feature "dimension." Thus the data warehouse with "click stream" have "dimension" to the table of Siswa, Mata_Kuliah, Waktu, Tanggal, and Pages. "Aggregate click stream" schema is the physical lines that exist in the database. This line is usually made from the sum of other records. By "Aggregate click stream" allow to make a list and an index of student learning style. "Aggregate click stream" contains Siswa as a dimension, associated with average student key and the total number of students accesses the LMS, also used Hasil and Mata_Kuliah as a dimension.

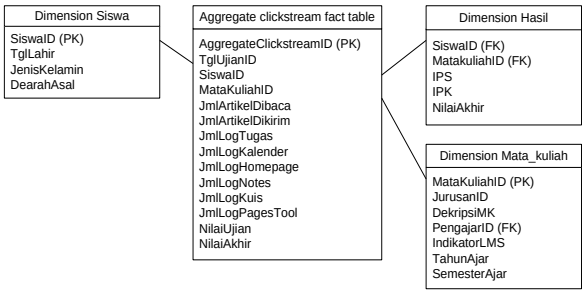

Figure 6. Star schema : Aggregate click stream

3.3 Data mining process

The seven steps in the data mining process for student data are:

a. Step 1. Establish mining goals. Deciding what the desired results are.
b. Step 2. Select data. Deciding which data are useful, which attributes are worth considering, and how big the sample size should be.
c. Step 3. Pre-process data. Filter out noisy, erroneous, or irrelevant data, and handle missing data.
d. Step 4. Transform data. Where possible, reduce the number of data attributes or extract new ones from existing data attributes. Combine data tables and project the data onto working spaces – tables that represent the optimal abstraction level for the problem of interest.
e. Step 5. Store data. Integrate and store data at a single site under a unified scheme.
f. Step 6. Mine data. Perform appropriate data mining functions and algorithms according to mining goals. Typically, analysts first construct data cubes to provide multi-dimensional views of the data. Then they perform online analytical mining using the multi-dimensional data cube structure for knowledge discovery.
g. Step 7. Evaluate mining results. Perform various operations such as knowledge filtering from the output, analyzing the usefulness of extracted knowledge, and presenting the results to the user for feedback. The feedback from this step can prompt changes to earlier steps.

The development objective in the LMS data mining is to gain knowledge related to learning activities of students in using the LMS. Attributes that are used in exploring the data and its correlation with the data warehouse star schema can be seen in Table 1.

Table 1. Attribute to mining data

Attribute Name	*Table*
SiswaID (key)	Dimensi Siswa
DaerahAsal	Dimensi Siswa
MataKuliahID (key)	Dimensi Mata_Kuliah
JurusanID	Dimensi Mata_Kuliah
PengajarID	Dimensi Mata_Kuliah
TahunAjar	Clickstream Log Fact Table
SemesterAjar	Clickstream Log Fact Table
TanggalID (key)	Dimensi Tanggal
TahunKalender	Dimensi Tanggal
MingguKalender	Dimensi Tanggal
WaktuID (key)	Dimensi Waktu
PageID (key)	Dimensi Halaman
AggregateClickstreamID (key)	Aggregate clickstream fact
TglUjianID	Aggregate clickstream fact

Table dimensions used were Siswa and Mata_Kuliah because learning activities at the LMS involving students and courses. Date and Time dimension table is selected to be able to describe a fact from the standpoint of time. While the dimension table page is used to obtain the facts from the student activity dimension in using the LMS.

Coffield (2004) emphasized the importance to determine the status of the instruments used to study related to students' learning styles, so there are some questions to be asked: Does the instrument consistently? Is the instrument valid? How should teachers respond to students' learning styles? This paper focuses only on the third question and uses data mining approach to the analysis, results are used to recommend lecturer in making decisions and respond to students' learning styles. This paper using Felder index of learning styles (Felder and Silverman, 2005) to measure students' learning styles, where there are four dimensions of learning styles: Cluster 1, Active-reflective, usually make observations and think before doing and understanding the information in a way to discuss it. Cluster 2, Sensory-intuitive, more likely to learn on the basis of a fact, solve the problem using a method to find a possibility and connection. Cluster 3, Visual-verbal, usually remembering something based on the views and get more information than words, both written and oral explanations. Cluster 4, Sequentially-globally. These groups usually understand a thing globally first and then learn the details in sequence.

4. EVALUATION OF MINING RESULTS

"Homepage" should be seen as a gateway, and not a tool. To access all other tools students always have to go to "Homepage". Correlation was found between using the "Homepage" and learning style or performance, which may be caused by correlation with another tool. Students with learning styles sensory and reflective more use of LMS. Reflective students who prefer to observe and think before doing. Since the LMS is used as a learning tool by students with reflective and sensory learning styles, teachers should strive to meet the learning needs of students. There is a similar correlation between the "Homepage" and "Content" on the UTS and exercise. Correlation in this case is the effect of intrinsic motivation, not the cause of increasing students' learning process. It seems that the LMS administration element ("E-mail", "Calendar" and "MyGrades) does contribute to the UTS results, but the academic achievements of students can be seen in the UAS results. Instrument rating (Assignments, Discussions write) indicated in many cases there is a significant correlation with the UTS results, which asserts that the way the course material is displayed in the LMS effect on student performance. In this paper an approach to data mining can provide to advice teachers with information to support decision making and responding to their students' learning styles. Table 2 shows only the aggregate results, but the data warehouse design have tracing facility that allows the lecturer to see the steps to deal with subjects that are taught.

Table 2. can be seen that there is a link between the use of LMS tools with students' learning styles. Cluster 1, active-reflective, usually conduct observation and think before doing and understanding the information in a way to discuss it. They are more frequently access the homepage, read the discussion topics, see the academic calendar, information value and access the assignment of student teachers. Cluster 2, sensory-intuitive, more likely to learn on the basis of a fact, solve the problem using a method to find a possibility and connection, rarely access the LMS compared to cluster 1. Sensory-intuitive students often access lecture material, reading the discussions and access the homepage.

Table 2. Number of modules that show a significant correlation between LMS tool frequency and learning style

Tools LMS	Home Page	Discussion (membaca)	Discussion (menulis)	E-mail	Calendar	Content	MyGrade	Assignment	Quiz
Number of modules under investigation for which this tool is used	47	21	20	10	15	32	39	34	9
Reflective	6	5	0	1	5	1	4	3	1
Active	0	0	0	0	0	0	0	0	0
Intuitive	3	4	0	0	1	5	1	1	1
Sensing	0	0	2	0	0	0	0	0	0
Visual	1	1	2	0	1	1	1	1	1
Verbal	1	1	2	0	1	1	1	1	1
Sequential	5	0	1	0	1	0	0	0	1
Global	0	2	0	0	0	2	0	0	0
Results : UTS	13	4	2	3	6	6	9	6	6
Results assessments	9	2	1	2	2	4	3	4	4
Results : UAS	12	0	2	0	0	8	2	0	0

Cluster 3, visual-verbal, most often access the LMS compared to clusters 1 and 2. This group usually remembers things based on the views and get more information than words, both written and oral explanations. Students with visual-verbal learning style more use of LMS to write the topic and content

of discussion. Cluster 4, sequential-global, more frequent use of LMS than clusters 2 and 3. These groups usually understand a thing globally first and then learn the details in sequence. Students with sequential-global learning style most often access the homepage. Sometimes they also access the contents of lectures and reading material and content of discussion topics.

Of the 200 respondents who use LMS in their learning activities related to the percentage of students with four dimensions of learning styles described by Felder can be seen in Table 3.

Table 3. Percentage of students associated with the four dimensions of learning styles

Active (%)	Reflective (%)	Sensing (%)	Intuitive (%)	Visual (%)	Verbal (%)	Sequential (%)	Global (%)
46	34	52	29	77	3	42	38

5. CONCLUSION

Advances in technology have dramatically enhanced the level of productivity and efficiency of Higher Education Institutions. Many lecturers draw manually the same type of queries from LMS data for purposes of action research. On the other hand, more and more technology providers create data warehouses to support many kinds of decision-making. It is possible and valuable to follow a data mining approach to facilitate the queries drawn by individual action from Learning Management Systems (LMS). Data warehouse tables are created consisting of the dimensions: Siswa, Mata_Kuliah and Waktu. Table facts: "Click Stream" which is manifest in the form of web server logs, where each web server log records associated with a one-page student access LMS events. Exploring data cluster analysis performed premises; the objects are clustered or grouped based on the principle of maximizing the intra-class similarity and minimizing the interclass similarity. That is, clusters of objects are formed so that objects within a cluster have high similarity in comparison with the one another, but are very dissimilar to objects in other clusters. Each cluster can be view as a class of objects, from which rules can be derived. Felder index of learning styles used in this paper to measure students' learning styles, where there are four dimensions of learning styles: Active-reflective; Sensory-intuitive, Visual-verbal; and Sequentially-globally. With a focus on how should lecturer respond learning styles and use of data mining approaches the analysis results can ultimately be used to recommend lecturer in making decisions and respond to students' learning styles.

REFERENCES :

Berry, M.J.A. and Linoff, G.S. (1997), Data Mining Techniques for Marketing, Sales, and Customer Support, Wiley, New York, NY.

Berry, M.J.A. and Linoff, G.S. (1999), Mastering Data Mining: The Art and Science of Customer Relationship Management, Wiley, New York, NY.

Coffield, F., Mosely, D., Hall, E. and Ecclestone, K. (2004), *Should We be Using Learning Styles? What Research Has to Say About Practice*, technical report, Learning and Skills Research Centre, Learning and Skills Development Agency, Regent, London.

Felder, R, Spurlin, J., 2005, "*Applications, reliability and validity of the index of learning styles*", International Journal of Engineering Education, Vol. 21 No. 1, pp. 103-12

Groth, R. (2000), Data Mining: Building Competitive Advantage, Prentice-Hall, Englewood Cliffs, NJ

Inmon, W.H. (2002). *Building The Data Warehouse*. Third edition. John Wiley & Sons. New York.

Van Dyk, Liezl, Conradie, Pieter (2007), "*Creating business intelligence from course management systems*", Campus-Wide Information Systems Vol. 24 No. 2, pp. 120-133, Emerald Group Publishing Limited

Yu, S.-C. and Chen, R.-S. (2001), "Developing an XML framework for an electronic document