

Educational Data Mining (EDM) as a Model for Students' Evaluation in Learning Environment

Nurul Hidayat

Department of Computer Science and
Electronics Gadjah Mada University
Yogyakarta Indonesia
Universitas Jenderal Sudirman
Purwokerto - Indonesia
nurul@ugm.ac.id

Retantyo Wardoyo

Department of Computer Science and
Electronics Gadjah Mada University
Yogyakarta Indonesia
rw@ugm.ac.id

Azhari SN

Department of Computer Science and
Electronics Gadjah Mada University
Yogyakarta Indonesia
arism@ugm.ac.id

Abstract— *The theory of Educational Data Mining (EDM) is a branch of knowledge domain that presence to have a narrow relationship to mine data from the educational institution. The EDM principally uses computational approaches to analyze educational data to dig up some information related to the context of teaching and learning activities. EDM is able to manage learning materials and the environment through a Learning Management System. The purpose of this study is to identify a knowledge to conduct an analysis of student motivation on e-Learning system based on data mining techniques. We use a dataset of the learning system activities in the Program of Distance Learning hosted by the second batch of APTIKOM Consortium. In this study, the association rules and classification techniques are used to identify patterns of knowledge and reorganize the virtual course based on patterns found. The expected result of this research can contribute to a model of Data Pre-Preparation Process and its steps from Moodle log data as a reference for the researchers. For the future research, we recommend using different datasets from techniques to comprehend divers of result expectations.*

Keywords— Educational Data Mining, Learning Management System, Model.

I. INTRODUCTION

The web-based education system has been applied as an important medium to support the student activities and faculty in the educational institution level. The main objective is to open communication channels to become the broadest system of collaboration between some entities. Furthermore, in the implementation of e-Learning, learners can take the web-based classes to improve their knowledge at any time, any place to enable to create their online classroom and monitor student's performance.

Moodle is one of the Learning Management System (LMS) platform that supports educators to create an effective online learning system. However, behind many conveniences gained by the lecturers, at the same time, this system did not provide the functionality to access the behavior and motivation of the learning environment. Conceptually, such as that delivered by Mostow that record or log student activity in Moodle system can show student interactions includes reading, writing, taking exams, and perform various assignments [1].

In addition, the magnitude of the accumulated per day of a course on e-Learning system is very difficult to analyze by manually. Although there are some tools that able to help to make report useful, which in turn is very valuable to analyze the behavior patterns of students, according to Garcia tools do not offer special features for faculty to be able to track

and evaluate the activities of all students during the class [2].

Data Mining Technique is the methodology used to extract valuable information in the extracting data on a web-based system. In the case of this research is the study of data mining in a system LMS. Moreover, data mining can analyze the results of the relevant information and produce a different perspective to understand better about the activities so that students can tailor courses to what kind of system can be applied to a student based on the motivation and the knowledge base of their own activity.

In a study, Baker stated that the Educational Data Mining (EDM) related to developing methods to explore a unique type of data in the educational setting. Thus, this method can make a better understanding of the students and the setting in which they should be learning their environment at the same time [3]. On the one hand, the increase in both software, as well as the role of education as a repository or database of student information situation, has created large repositories of data reflecting how the students learn and interact [4]. The same goes for Castro, he argued that the use of the Internet in education institution has created a new context, known as e-learning or web-based education. This terminology in which a large amount of information about the interaction of teaching and learning are endlessly generated and ubiquitously available on the system in the activity log or logs activity [5]. Still, according to Mastowm that all information stored in the activity log provides a gold mine for education data [6].

Further, data mining comes to solve problems, not only because of the intrinsic complexity but also due to it has largely rooted in the ever-changing world of business. In detailed, Chapman revealed that it can comprehend not just as a collection of data analysis methods, but as the process of data analysis include an instrument from understanding the data, pre-processing and evaluation process modeling and how to implement this mechanism [7].

Consequently, it is still unusual for preferential attention to its own methods of Data Mining. In other words, it generally bridging the fields of traditional statistics, pattern recognition, and machine learning. The aim is to provide analytical solutions to problems in various fields, such as biomedical, engineering, and business, to name just a few. An aspect that may make the Data Mining unique is giving a special attention to the compatibility of modeling techniques with new Information Technologies (IT) perspective and

database technology by typically focused on large databases, heterogeneous and complex.

For this historical background, data mining can be used to extract knowledge from the e-learning system through the analysis of the information. It is provided in the form of data generated by their users. In this case, the main goal became the finding patterns of use of the system by educators and students to find a certain pattern of student learning behavior.

In a long study conducted by Romero, he produced a survey that could be recognized that this is not the first time a similar business has been done: a collection of papers that cover most of the important topics in the field was simultaneously presented in [8]. The findings of a survey commissioned from a different angle to different interests of potential readers: research surveyed can be seen as displayed along two axes: the problem of data mining and methods and e-learning applications. This organization-surveyed content should allow the reader to access information in a way that is more compact and self-contained from it.

II. RELATED WORKS

In some studies have shown a close relationship between the field of Artificial Intelligence (AI) and Machine Learning Data Mining as a source techniques and methods in performing a processing of a terminology database [9], [10], [11]. Meanwhile, Baker [9] established an opportunity in AI research and education on the basis of three models of the educational process. The model as a scientific tool is used as a means to understand and predict some aspects of the educational situation. The models as components, according to some characteristics of the teaching or learning process are used as a component of educative artifacts. In addition, the model as the basis for the design of artifacts education. It is helping design a computer tool for education by providing design methodology and system components, or even by limiting the range of tools that may be available to students.

Some interesting research results presented by Margo [12] and Tang [13] presents how data mining techniques can be successfully incorporated into the e-learning environment and how they can improve learning tasks done. Still, Tang argued [13] that clustering of the data suggested as a means to promote group-based collaborative learning and to provide additional student diagnosis. In-depth review, the possibility of applying the model of web mining and web usage clustering techniques to meet some of the current challenges in distance education is presented in the form of specific data representation [11].

The proposed approach can improve the effectiveness and efficiency of distance education in two ways: (1) on the one hand, the discovery of the aggregate and individual pathways for students could help in the development of customized education effectively by giving an indication of how to organize the best courseware educator organizations, and (2) on the other hand, the structure of the virtual

knowledge can be identified through the method of web mining. The discovery of association rules can allow for web-based distance tutor to identify patterns of knowledge and reorganize the virtual course based on patterns found.

An analysis of how the techniques Machine Learning, as a general source for Data Mining techniques have been used to automate the construction and induction of student models, as well as background knowledge required for modeling the students [14]. In this paper, we defined the difficulties, suitability, and potential of applying machine learning techniques for modeling student was facilitated.

A literature search shows that most of the related researches have been deployed some data mining techniques to analyze the behavior of the students' learning, as follows:

- a) Romero et. al [8], this study demonstrates the usefulness of data mining techniques in the course management system and the rules can help to classify the students as well as to detect the source of anomalous values received from the student activities.
- b) Beck [15], Biswas et. al [16], and Brusilovsky [17] presents an approach to classify students to predict their final grades based on the features extracted from the log data mining techniques of web-based data system of education such as association rule mining is applied.
- c) Charciolo [18] provides that assessment to extract patterns and evaluating the activities on the field line and the classification and association rule mining algorithms.

III. PROPOSED METHOD

From the functionality of the model, it can be defined that Data Mining is an approach to data analysis from different perspectives to summarize the results of useful information. The data mining process such as that delivered by Fayyad [19] and Frawle [20] uses a number of rules as machine learning, statistical, and visualization techniques to locate and knowledge present in a form easily comprehend.

LMS Moodle system is a well-known open source software for learning management system in particular that offers teachers to create and manage an online class effectively. The students should not be required to attend courses in the room, to do the exercises and projects, as well as to take the exam in class. Also, in class e-Learning, students need to take a pre and post quiz online, to review and use materials from e-Learning systems and to participate in the assignment or discussion.

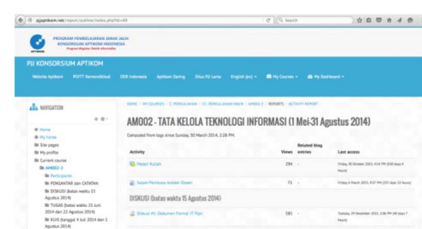


Fig 1. Moodle Log Report Screen

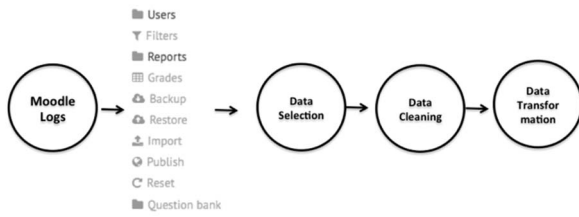


Fig 2. Data Pre-Preparation Process in e-Learning

Figure 2 illustrates on Data Pre-Preparation Process in e-Learning. In the initial section, researchers get Moodle Logs that contains several important attributes include User Logs, Filter, Reports, Grades, Backup, Restore, Import, Publish, Reset, and are also Question Bank. At this point it should be ensured that all the attributes that have activity logs can be lifted with a good and perfect. Stage 2 is a Data selection, where the system will ensure that data will be processed first through the selection process so that not all the attributes to be processed in parallel. Then, in Phase 3 Data Cleaning is done where the system will do the cleaning or filter the data logs that have interests in accordance with the objectives and key researchers when conducting mining. Lastly, is Stage 4 is Data transformer, where the system has completed the process by doing a full transformation of the mining activities undertaken.

IV. RESULTS AND DISCUSSION

In the application of data mining in e-learning system as communicated by Romero et. al [21] is a repeating cycle in which knowledge is mined must enter the circle of the system and guide, facilitate and enhance the overall learning, not just turn data into knowledge, but also filter the knowledge mined for decision-making.

Meanwhile, the data mining process e-learning consists of four steps in the same general data mining process, as seen in Figure 3:

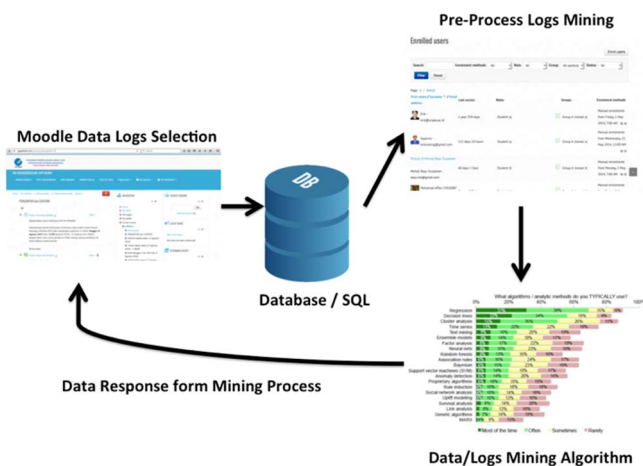


Fig 3. Mining Steps From Moodle Logs Data

Figure 3 shows the mining steps from logs data in Moodle which has four major activities as may be described in detail as follows:

1. Moodle Data Selection, the logs are filled or used by students and the use and interaction of information stored in the database. Data stored is all interactive activity between the three actors are students, faculty, and administrators.
2. Pre-Process Logs Mining, the data is cleaned up and converted into a suitable format to be mined. In order to preprocess the data Moodle. This condition is necessary in order to use the system database administrator or some specific preprocessing tool, so that the mining process can be done more easily.
3. Data / Logs Mining Algorithm, data mining algorithms applied to establish and implement a model that finds and summarizes knowledge of interest to the user (faculty, students and administrators). To do so, either general or specific data mining tools, or data mining tools can be used commercially or for free.
4. Response Data Mining Process form, or models obtained results are interpreted and used by the instructor for further action. Instructors can use the information found to make decisions about students and Moodle course activities to improve student-learning motivation.

Additionally, Moodle in particular can be delivered that the system cannot save the log as a text file. Instead, keep a log in a relational database, for instance MySQL, SQL Server and PostgreSQL are best supported, but can also be used with Oracle, Access, Interbase, and others). Databases more robust, flexible and yet vulnerable. Text log file that is specific to collect detailed access and high-level usage information from all the services available in the CMS. Based on the results of the investigation in practice, the system Moodle database has about 145 tables that are interrelated. But we do not need all this information and it is also necessary to convert it into the required format used by the data mining algorithms.

Based on these rationales, we have to perform the previous steps to pre-process data Moodle. Data preprocessing allows the original data to be converted into a form suitable for use by a particular data mining algorithms or frameworks. It is important to notice that the preprocessing tasks usually performed by an administrator, not by the instructors themselves. In this step, the session that have been identified since the user might also have to logging and eliminates the need for a regular user and session identification tasks. Thus, the data collected by Learning Management System may require less cleaning and preprocessing of data collected by other web-based systems.

V. CONCLUSIONS

The purpose of this study was to provide a model of Data Pre-Preparation Process in e-Learning where in this framework there are four steps in the pre-preparation is Moodle Logs, Data Selection, Data Cleaning and Data Transformation. Then, in the development of this research also produced a model of Mining Steps From Moodle Logs Data that has 4 activities. The limitation of this research is in the data pre-preparation step, where we need to refine the dataset very carefully to avoid the mistranslation of the data. In the future works, we recommend using different datasets, case study, and methods to get better results of data analysis.

REFERENCES

- [1] Mostow, J., Beck, J., Cen, H., Cuneo, A., Gouvea, E., Heiner, C. (2005). An educational data mining tool to browse tutor-student interactions: Time will tell! In *Proceedings of the Workshop on Educational Data Mining*, 15–22.
- [2] Garcia, E., Romero, C., Ventura, S., Castro, C. (2009). An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, 19, 99–132.
- [3] Baker, R., Yacef, K. (2010). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 3-17.
- [4] Koedinger, K., Cunningham, K., Skogsholm A., Leber, B. (2008). An open repository and analysis tools for fine-grained, longitudinal learner data. In *1st International Conference on Educational Data Mining*, Montreal, 157-166.
- [5] Castro, F., Vellido, A., Nebot, A. Mugica, F. (2007). Applying Data Mining Techniques to e- Learning Problems. In: Jain, L.C., Tedman, R. and Tedman, D. (eds.) *Evolution of Teaching and Learning Paradigms in Intelligent Environment*. Studies in Computational Intelligence, 62, Springer-Verlag, 183-221.
- [6] Mostow, J., Beck, J. (2006). Some useful tactics to modify, map and mine data from intelligent tutors. *Journal Natural Language Engineering*, 12 ,2, 195-208.
- [7] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: *CRIPS-DM 1.0 Step by Step Data Mining Guide*. CRISP-DM Consortium (2000).
- [8] Romero, C., Ventura, S., De Bra, P.: Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware. *User Modeling and User-Adapted Interaction* 14(5) (2004) 425-464.
- [9] Baker, M.: The Roles of Models in Artificial Intelligence and Education Research: A Prospective View. *International Journal of Artificial Intelligence in Education* 11 (2000) 122-143.
- [10] Fasuga, R., Sarmanova, J.: Usage of Artificial Intelligence in Education Process. In: *International Conference for Engineering Education & Research, ICEER2005*. Tainan, Taiwan (2005).
- [11] Ha, S.H., Bae, S.M., Park, S.C.: Web Mining for Distance Education. In: *IEEE International Conference on Management of Innovation and Technology, ICMIT'00*. (2000) 715-719.
- [12] Margo, H.: Data Mining in the e-Learning Domain. *Computers & Education* 42(3) (2004) 267-287.
- [13] Tang, C., Lau, R.W., Li, Q., Yin, H., Li, T., Kilis, D.: Personalized Courseware Construction Based on Web Data Mining. In: *The First international Conference on Web information Systems Engineering, WISE'00*. IEEE Computer Society. June 19 - 20, Washington, USA (2000) 204-211.
- [14] Sison, R., Shimura, M.: Student Modelling and Machine Learning. *International Journal of Artificial Intelligence in Education* 9 (1998) 128-158.
- [15] Beck, J.E., Woolf, B.P.: High-Level Student Modeling with Machine Learning. In: Gauthier, G., et al. (eds.): *Intelligent Tutoring Systems, ITS 2000*. Lecture Notes in Computer Science, Vol. 1839. Springer-Verlag, Berlin Heidelberg New York (2000) 584-593.
- [16] Biswas, G., Leelawong, K., Belyne, K., Viswanath, K.: Developing Learning by Teaching Environments that Support Self-Regulated Learning. In: *The 7th International Conference on Intelligent Tutoring Systems*. Maceió, Brazil (2004).
- [17] Brusilovsky, P.: Adaptive Hypermedia. *User Modelling and User Adapted Interaction, Ten Year Anniversary Issue* 11(1-2) (2001) 87-110.
- [18] Carchiolo, V., Longheu, A., Malgeri, M., Mangioni, G.: Courses Personalization in an e- Learning Environment. In: *The 3rd IEEE International Conference on Advanced Learning Technologies, ICALT'03*. July 9-11 (2003) 252-253.
- [19] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurasamy, "Advances in Knowledge Discovery and Data Mining", AAAI/MIT Press, 1996.
- [20] W. Frawley, G. Piatetsky-Shapiro, and C. Matheus, "Knowledge Discovery in Databases: An Overview". *AI Magazine*, Fall 1992, pp. 213-228.
- [21] Romero, C., & Ventura, S. (2006). *Data mining in e-learning*. Southampton, UK: Wit Press.