Data Analysis in E-Learning System of Gunadarma University
by Using Knime

Dian Kusuma Ningtyas
tyaz_tyaz_tyaz@student.gunadarma.ac.id

Farah Virnawati
virtha_1408@student.gunadarma.ac.id

Prasetyo
prasetyo@student.gunadarma.ac.id

Tirta Paramitta
eatha_020688@student.gunadarma.ac.id

I. Wayan Simri W.
iwayan@staff.gunadarma.ac.id

Gunadarma University
Jl. Margonda Raya 100, Depok, Indonesia, 16424

Abstract

Nowadays, E-Learning has become a preferred choice of learning method because it can reduce operational cost in establishing educational activities. It also offers time efficiency, flexibility in the way of attending courses, and ease of access on teaching material. The main problem is how to make user more interested in accessing and fulfilling user requirements to provide suitable E-Learning system. One way to meet that challenge is by analyzing user behaviors. The purpose is to give a recommendation and evaluation in developing E-Learning system.

In this paper, we will use log data from E-Learning system of Gunadarma University. Using Knime, we have done some analysis to explore the pattern of user’s behaviors. Later, the pattern will be used as an input to decide a way to enhance E-Learning system.

Keywords: access, data analysis, E-Learning, Knime, pattern, user behavior

1. Introduction

E-Learning is one of learning methods that appears as a personalization demand of human resource development supported with Information Technology sector. E-Learning enables user to access course material by using internet, intranet, or other media. Now E-Learning apparently becomes a preferred choice because it offers some benefits, such as reduced operational cost, ease of access, broader participant, and minimal use of equipment and printed material. E-Learning can also be intended to those who have several limitations in attending physical class.

With the great number of benefits offered, now many institutions implement E-Learning as a learning solution to their staffs, especially in educational sector, like Gunadarma University. E-Learning has also been applied to their staffs and students. Moreover, E-Learning becomes a local issue released by Ministry of National Education in National Education Meeting on February 4th - 6th 2008. It makes E-Learning considered as a way to enrich the material of learning activities which are still performed classically[1]. Based on that fact, E-Learning has been considered as the most prospective way to establish learning environment especially in Indonesia.

In the new established E-Learning system, it is important to continuously evaluate and monitor effects of its implementation. The main problem is how to provide E-Learning system that can attract and really suitable with user demand. One way to explore user demands is by analyzing their behavior when accessing E-Learning system. Primary result expected from behavioral analysis is to give a recommendation in enhancing E-Learning system. This recommendation result is very suitable for an institution, like Gunadarma University, that recently establishes E-Learning system.

This analysis is intended to two parties, developer and E-Learning user. The developer of E-Learning system can observe some effects and tendencies that really happen in E-Learning implementation. The effects and tendencies can become input for evaluation to improve or decide the next step in E-Learning system development. Users can also take the benefits of this analysis. With a better E-Learning system, they will become more interested in accessing and also increasing their learning interests.

This paper is organized as follows. In section 2, we will explain recent approach used for optimizing E-Learning. In
section 3, we will give a brief overview to Knime as tool that used in this paper. Section 4 presents our case study that can be solved by using Knime. Section 5 shows the result and discussion to our case study. Section 6 concludes the paper.

2. Recent Approach For E-Learning Optimization

In fact, there are a lot of methods which can be used to optimize the effectiveness of learning process through E-Learning. It is proven with some papers discuss about methodologies used to solve problem in E-Learning. In paper "Discovering Student Preferences in E-Learning", Cristina Carmina, Gladys Castillo and Eva Millan propose to use adaptive machine learning algorithms to learn about the student's preferences over time. First, they use all the background knowledge available about a particular student to build an initial decision model based on learning styles. This model can then be fine-tuned with the data generated by the student's interactions with the system in order to reflect more accurately his/her current preferences[4].

Then Felix Castro, Alfredo Vellido, Angela Nebot, and Francisco Mugica explain more detail about several techniques of data mining which deal with Artificial Intelligent. Also, they provide a taxonomy of e-learning problems to which Data Mining techniques have been applied, including, for instance: Students classification based on their learning performance; detection of irregular learning behaviors; e-learning system navigation and interaction optimization; clustering according to similar E-learning system usage; and systems adaptability to students requirements and capacities in paper "Applying of Data of Mining Techniques to E-Learning Problems"[5].

There is also Wengang Liu trying to introduce an ecological method which reveals latent user models that can be used for various pedagogical purposes in educational applications. This method consists of six layers which are the raw data layer, the factor data layer, the data mining layer, the measurement layer, the metric layer and the application layer. Liu also applies clustering tools to dynamically analyze and group learners to match their behavior and performance, and create new metrics and measurements to represent user models[6].

Besides the studies about method used in Data Mining, there are also many writers have tried to analyze the internal issue of E-Learning. one of them is David Monk, by paper entitled "Using of Data of Mining for e-Learning Decision Making", he tried to examine the path learners followed when offered the course in a custom virtual learning environment (VLE), which is applied at University of Glamorgan, structured by tasks, course materials and learning resources. A better understanding of how learners accessed the electronic course materials was needed to evaluate the effectiveness of developing and delivering courses in this way. By combining the data content of activity with the user's profiles were possible to examine alternate information perspectives and reveal patterns in large volume data sets. Mining data in this way provides ways to learn about learners in order to make effective decisions regarding teaching methods, delivery models and infrastructure investment[7].

There is also Azizul Azhar bin Ramli who is trying to implement the high level process of Web Usage Mining using basic Association Rules algorithm called Apriori Algorithm, in order to produce the university E Learning (UUM Educare) portal usage patterns and user behaviors, using commercial data Web mining tools (WebLog Expert Lite 3.5 and Sawmill 7) and ARunner 1.0 (prototype of GUI Christian Borgelt Apriori tool by Shamrie Sainin, FTM, UUM), it has identified several Web access patterns. This analysis includes descriptive statistic and Association Rules for the portal including support and confidence to represent the Web usage and user behavior, which is explained in his paper entitled "Web Usage Mining Apriori Algorithm: UUM Learning Care Portal Case"[3]. What we have done in this paper is relatively a simple approach. By using Knime tool, we have discovered some results of our behavior analysis. Mostly, we use statistical methodology combined with query manner to retrieve some facts.

3. Knime

KNIME is a modular data exploration platform that enables the user to visually create data flows (often referred to as pipelines), selectively executes some or all analysis steps, and later investigates the results through interactive views on data and models[Knime].

KNIME was developed (and will continue to be expanded) by the Chair for Bioinformatics and Information Mining at the University of Konstanz, Germany. The group headed by Michael Berthold is also using KNIME for teaching and research at the University. Quite a number of new data analysis methods developed at the chair are integrated in KNIME. KNIME is released under the Aladdin Free Public License, and can be run on Windows and Linux.

KNIME based version has already incorporated over 100 processing nodes for data I/O, preprocessing and cleansing, model, analysis and data mining as well as various interactive views, such as scatter plots, parallel coordinates and others. It includes all analysis modules of the well-known Weka data mining environment (http://www.cs.waikato.ac.nz/ml/weka/) and additional plugins allow R-scripts (www.r-project.org) to be run, offering access to a vast library of statistical routines. Screenshot of Knime can be seen in figure 1 below.
Figure 1. Screenshot of Knime

Knime can be obtained as a single package of binary file or bundled in eclipse plugin. In this paper, we use the single package that can compile directly in Java environment. The eclipse bundled package is for development purposes, which offer code-level access.

4. Log Data Analysis using Knime

One approach that can be used to optimize E-Learning system is by analyzing user behavior. In this paper we have done log data analysis in the E-Learning system to find some facts related to E-Learning implementation. Those facts are in form of user behavior pattern and statistical result.

Log data contain NPM (student ID), access time (in UNIX time format), IP address, course code, course full-name, user action, and URL address that captured in the system, shown in figure 2.

<table>
<thead>
<tr>
<th>username</th>
<th>time</th>
<th>ip</th>
<th>course</th>
<th>full-name</th>
<th>action</th>
<th>url</th>
</tr>
</thead>
<tbody>
<tr>
<td>12345678</td>
<td>112406656</td>
<td>120</td>
<td>105.211 152</td>
<td>5</td>
<td>Kejuruanagaan (CA) view</td>
<td>view.php?id=1</td>
</tr>
<tr>
<td>12543876</td>
<td>114650989</td>
<td>120</td>
<td>106.211 152</td>
<td>5</td>
<td>Kejuruanagaan (CA) view</td>
<td>view.php?id=2</td>
</tr>
<tr>
<td>12543886</td>
<td>114650989</td>
<td>120</td>
<td>106.211 152</td>
<td>5</td>
<td>Kejuruanagaan (CA) view</td>
<td>view.php?id=3</td>
</tr>
<tr>
<td>12543897</td>
<td>114650989</td>
<td>120</td>
<td>106.211 152</td>
<td>5</td>
<td>Kejuruanagaan (CA) view</td>
<td>view.php?id=4</td>
</tr>
<tr>
<td>12543898</td>
<td>114650989</td>
<td>120</td>
<td>106.211 152</td>
<td>5</td>
<td>Kejuruanagaan (CA) view</td>
<td>view.php?id=5</td>
</tr>
</tbody>
</table>

Figure 2. Entry at Used E-Learning System Log

Data analysis is implemented to explore E-Learning user’s behavior in E-Learning system of Gunadarma University. Problem of analysis may occur in the level of understanding and cleaning from log data of E-Learning system. At the beginning, we have more than 600,000 record along 2006-2008 period with cleaning level needed almost 11 percent. Cleaning process done to the entry which does not have username, IP address, and access time. Besides the data consistency, data cleaning also done for action that not so significant. The insignificant action is such action with occurrence level under the defined threshold value. After cleaning process, we have 546,449 records for log data along March 16th 2006 until June 3rd 2008.

The next step is to find user access pattern. Our primary concern is user time tendency in E-Learning system when accessing and course spreading toward student department, faculty, and generation year. Beside that, we will show several statistic results such as the most accessed course, the most accessed action, average of access duration, the number of IP inside and outside campus, and activity rate per department.

5. Result and Discussion

5.1 Experimental Result

After cleaning the log data, we import it to Knime and do some data analysis process. Knime supports several data types, such as any text file, ARFF, database, and it also provides artificial data generator. In this paper, we use text file (.txt) for the data.

The first analysis is the most favorite course based on the number of access on each course. In Knime, we do some steps to get the expected result, shown in figure 3.

In figure 3, there are 4 operators performed in Knime. The first operator is File Reader, which is needed to read the source data log. The second operator is GroupBy, which is needed to group records in the data log based on a field (for this case we group the records based on course names). The next operator is Sorter, used to sort the records based on a field (for this case we sort the data log based on the number of NPM (student ID). The last operator is Interactive Table, to view the result in a table. After executing all these operators, we get that the most favorite lecture is Ilmu Budaya Dasar (KA).

The second analysis is the action mostly occurs in the E-Learning system. The step is the same as the first analysis. The differences are the chosen fields in GroupBy operator, which is based on actions, and the chosen field in Sorter operator, which is based on the number of URL.
After executing all those operators, we get that the most action is view.

Third analysis is about average of access duration. Based on the following equation:

\[
\text{Accessing Time} = \frac{(1,212,477,975 - 1,142,499,905)}{546,449} \text{access} \\
= \frac{69,978,070}{546,449} \text{access} \\
= 128.06 \text{s/access} \\
= 2 \text{minutes/access}
\]

where:
1,212,477,975 = last accessing time
1,142,499,905 = the first accessing time
546,449 = the number of entry

From the equation above, we can see that the average of access duration is about 2 minutes per access.

The fourth analysis is the number of IP inside campus environment (local ip). The steps in Knime are shown in figure 4.

In figure 4, after File Reader operator, we can see that there are four stages to get the number of local ip. Each of the stages has sequence of operators which are not different from the previous analysis. First stage is needed to count the number of local ip in block A. Second stage is needed to count the number of local ip in block B. Third stage is needed to count the number of local ip in block C. The difference between those stages is only in regular expression used in the Row Filter operator. And then fourth stage is needed to count all of ip (local or public) accessed the system. After execute all these operators, the result is the number of local IP accessed is less than public IP.

The fifth analysis is most active department. In Knime, we do some steps to get the expected result, shown in figure 5.

In figure 5, we see that there are 5 operators performed in Knime. The first operator is File Reader, same as previous analysis, which is used to read the log data. The second operator is GroupBy, for this result we group the records based on NPM. There are two third operators after GroupBy operator, which are Sorter, for this result we sort the data log based on the number of actions, and Histogram (interactive). The last operator is Interactive Table, which is needed to view the result in a table. After execute all these operators, we get that the most active department accessed E-Learning system is Sistem Informasi.

From the fifth analysis, we can have another analysis. We can analyze which generation year in Sistem Informasi department is the most active. To know that, we must perform some steps in Knime, shown in figure 6.

In figure 6, we see that there are 5 operators performed in Knime. The first operator is File Reader, same as previous analysis, which is used to read the log data. The second operator is Row Filter, which is needed to filter the records based on a field (for this result we filter the data based on NPM in Sistem Informasi department. The next operator is GroupBy, for this result we group the records based on generation year. The last two operators are Interactive Table and Pie Chart. Pie Chart operator is used to make a pie chart based on the result shown in Interactive Table. After execute all these operators, we got the most active generation year in Sistem Informasi department is 2007.

We can also analyze profoundly the most accessed
6. Conclusion

Data analysis using Knime in E-Learning system of Gunadarma University is able to give the expected result. Released under the Aladdin Free Public License, Knime is capable to handle large data set, perform filtering, clustering, sorting, and give the result in table or chart. Also, Knime can be executed in Windows or Linux operating system.

By implementing data analysis, we can see user behavior and access pattern in E-Learning system of Gunadarma University. The results are addressed to give a recommendation to improve the E-Learning system, for example: socialization to gain more access to E-Learning system, enhance course material quality, and infrastructure improvement. The result of this analysis can also give an evaluation for implementation of E-Learning system in Gunadarma University in order to make it better.

The number of access from outside of the campus environment indicates two things. First, students are not satisfied with the facilities in campus to access V-Class. Because of that, it is important to improve infrastructure of the campus so that the students can feel comfortable to access V-Class in campus environment. Second, access from outside of the campus environment needs a better bandwidth supply and management to be able to handle number of access especially when peak access occurs.

In advanced, Knime's ability can be used as a background process to capture data and analyze it automatically. This is possible since Knime developer version can be accessed freely and give us opportunity to embed our self-defined operator. With this ability, Knime can be well-collaborated with E-Learning system to capture long-term data.

5.2 Discussion

As a benchmark, we also compare those results with manual approach. Our manual approach is developed by PHP and mySQL and the results are exactly as the same as what we have got using Knime. For our purpose, Knime is satisfying enough and can meet our desired result. Moreover, Knime is easier than using our manual approach since it allows us to process the file directly. GUI environment that Knime possessed also makes file processing simpler without getting stuck in syntax and command.

Figure 6. Steps to get the most active generation in Sistem Informasi

Figure 7. Steps to get the most accessed lecture in generation year 2007

In figure 7, we see that there are 5 operators performed in Knime. The first operator is File Reader, same as previous analysis. The second operator is Row Filter, which is needed to filter the records based on a field (for this result we filter the data based on student ID). The third operator is GroupBy, for this result we group the records based on course name. The next operator is Sorter, for this result we sort the data log based on student ID. The last operator is Interactive Table. After execute all these operators, we get that the most accessed lecture for generation year 2007 is Ilmu Budaya Dasar (KA).

References