



INTRODUCTION OF EDUCATIONAL DATA MINING BY USING A VARIETY OF TECHNIQUES IN ORDER TO ACHIEVE THE GOAL FROM THE MOODLE LMS

Ali Akhtar¹ⁱ,

Mohammad Serajuddin²,

Zafrul Hasan³

¹Researcher,

College of Pharmacy,

King Saud University,

Riyadh, Kingdom of Saudi Arabia

²Researcher,

Dental University Hospital,

King Saud University,

Riyadh, Kingdom of Saudi Arabia

³Researcher, College of Nursing,

King Saud University,

Riyadh, Kingdom of Saudi Arabia

Abstract:

Different works relating to this specialty have been done in recent years and several data extraction approaches have been used to solve numerous educational problems. This analysis compares the Felder-Silverman Learning Style Model component of student activity in Moodle class with three data mining algorithms for the identification of knowledge presentation dimension (visual/verbal) learning style. This study analyzes Moodle LMS student log data using data mining strategies to identify their learning styles that rely on one aspect of the learning style of Feld-Silverman: visual/verbal. The WEKA compares various classification algorithms as classified J48 Decision Tree, Naive Bayes and Portion. The selected classifiers were evaluated using a 10-fold cross validation. The tests revealed that at 71.18 percent the Naive Bays achieve the strongest score.

Keywords: data mining, educational data mining, techniques, Moodle LMS, Weka, Felder- Silverman learning style model

ⁱ Correspondence: email aakhtar@ksu.edu.sa

1. Introduction

Because of the great contributions that technology provides today to all disciplines, it can be clarified that the way we interact and the walls that have fallen thanks to the Internet, and to the volume of data collected and processed in various information systems, are evolving across existence. Therefore, the universe progresses adaptively would not make a mistake. Decisions are not supposed to be made in this transition, particularly if these decisions are decisive for the future of organizations¹.

One of the core goals of all educational settings is to ensure that learning mechanisms make it possible for students to comprehend and develop. This is where data mining Educational Data Mining (EDM) has essential importance for schools and for all organizations that promote diverse mechanisms of learning practices for their acronym in English.

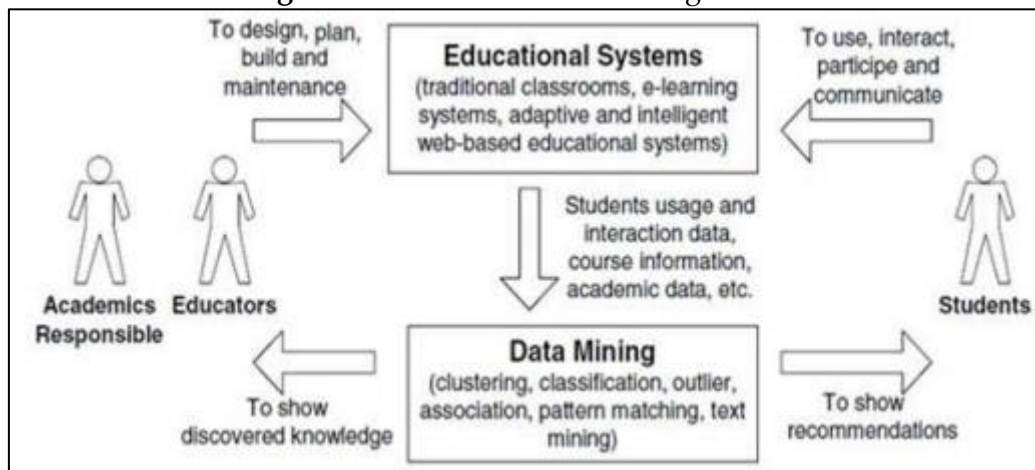
Data mining offers many data collection methods. The vast volume of data available in student libraries are greater than the human capacity to interpret and retrieve the most valuable knowledge without automatic processing. Discovery of knowledge (DK) is the nontrivial retrieval from a vast archive of implied, undefined, and possibly useful details. In KD, data mining was used to find trends for users' needs. The pattern description represents a subset of data in language. A great deal of data is rendered possible by the usage of technology in school systems. EDM contains a large quantity of knowledge that gives a better understanding of students and their learning processes. DM methods are used to interpret instructional data and to address problems of curriculum. EDM collects interesting, interpretable, valuable and new knowledge from educational data in a way similar to other extraction processes in DM techniques. EDM is, however, primarily designed to create approaches using particular data forms in educational systems. These approaches are used to improve understanding of instructional phenomena, pupils and environments. Developing analytical methods combining data and theory can contribute to improving T&L efficiency²⁻⁷. The objective of this study is to introduce education data mining using a range of techniques to accomplish the objective of the Moodle LMS.

1.1 Educational data mining (EDM)

The Information Mining Protocols were now used in various and informal circumstances. The usage of information mining is known as data mining in education (EDM). • Education Data Mining is a new order which creates technologies that investigate one type of information from an educational system and use those techniques to more efficiently obtain undergraduates and research conditions. Educational Data Mining is an educational database mining website that aims at:

The key components of EDM are partners of education, MD approaches, tools and strategies, education, tasks and results that meet the educational objectives. The information mining methodology in education is an iterative knowledge exploration approach involving definition, checking and refining hypotheses.

Figure 1: Educational Data Mining Process



1.2 Educational tasks and data mining techniques

Data analyze and visualize Statistics and knowledge on simulation are the two key tools used the most for this mission. Feedback for instructor strengthening Several DM techniques were used, but association rule mining was the most boring and shows fascinating connections between variables.

1.3 Educational data mining users and tools

Including:

- 1) Students 1. The students want to learn the concerns and methods of the students to strengthen their comprehension and implementation.
- 2) Teachers. Educators strive to see what method of instruction they should utilize and how they can develop their educational practices. Teachers should utilize EDM extensibility in order to decide if the learning program should be organized to organized and choose the correct strategies and technologies to target students for the greatest results. In particular, the filtration of knowledge used with the technique on human judgment gives educators an open door to take advantage of EDM on the grounds that this leads teachers to quickly identify designs that may complement their showing techniques during their residency or develop future classes. Teachers should use symbols that demonstrate obviously content pleasure and dedication and advances in screen learning.
- 3) Analysis providers. Researchers rely on the advancement and evaluation of adequacy knowledge excavation programs.
- 4) Managers and Administrators are responsible for the distribution of funds to organizations.

1.4 Student performance evaluation methods

The aim of the study is to explain the use of the recurrence of the repetitions, the arrangement of neural systems, the various neurosystems, e.g. back proliferation and feed forwards, the bayesian system, rules based frameworks, the bayesian system,

relationship investigation and relapse. Different types of bunching, affiliation rule mining, human trust and Markov Chain Investigation, as well como a section of this later stage are also referenced as undergraduate behavior.

2. Review of literature

Singh, and Yadav (2020) affirmed that Educational data mining (EDM) is an open research interdisciplinary area that improves information research strategies in academic fields. EDM computer methodologies should examine educational information so that instructive enquiries can be considered. It therefore provides innate information on the learning process and the successful arrangement of education. This paper provides an exhaustive report on current and important research in this field. The research focuses on strategies to examine educational information in order to develop models to improve the exhibitions and to improve institutional appropriateness. This paper adds and consigns written texts, recognizes and intercedes in the weighty work of professors and professional bodies. We acknowledge inquiries that provide an encouragement to correct the illumination and strengthen the more impotent portion of the foundation. The consequences of these examinations provide understanding of academic process improvement strategies, the provision of the under-study, the accuracy of data mining calculations and the development of open-source updates⁸.

Manjarres, et al. (2018) decared that Educational Data Mining is an emerging field that aims to develop methods to explore large quantities of information from educational environments, in order to understand the behavior, interests and results of students better. Different works relating to this specialty have been done in recent years and multiple data extraction techniques have been used to address various educational problems. The purpose of this paper is to review the work that has been used by data mining techniques to solve specific educational problems and to classify them in different scenarios⁹.

Blagojevic, and Micic (2013) presents a method of PDCA (Plan, Do, Check, Act), which was developed and implemented in the Cacak Faculty of Technology, University of Kragujevac, to improve intelligent e-Learning system Web-based reports. The focus of the workshop is to improve LMS or e-learning systems by predicting student behavior patterns and adapting the structure of such electronic courses. A current learning management framework is being improved by utilizing data processing methods and improving the efficiency of courses using personalized modules. This study presents the system design, implementation and assessment. For further work, the continuous improvement of the system created for PDCA and additional modules should be considered and the present and future results should be comparatively assessed¹⁰.

Lile, A. (2011) stated that Educational data mining has recently become a new field of research to extract knowledge from e-learning systems and to determine patterns. The Albanian education system is currently struggling to classify and customize instruction and forecast the quality of student experiences, including the identification of student

needs. The mining of educational data provides a variety of methods for solving these challenges through the education system. In order to accomplish the task of discovering the uses it aims to implement educational data mining by explaining step by step methods such as weighting attributes, relève, hi-squared, insecurity, grouping, association mining (Apriori). Analyzing mining results makes it possible for educational institutions to allocate resources more effectively and to organize learning to improve student learning experience and improve profits. The results of the experiments showed that the model of data mining presented in this research could obtain comprehensive and logical feedback from the LMS data describing the learning patterns of the students. For this work, the data mining tools Rapid Miner (v5.0) and Weka (v3.6.2) were used to mine Moodle's data, which were taken from the computer technology students of University Epoka, in the "C Programming – CEN112" course, during Spring Semester 2009-2010¹¹.

3. Methodology

3.1 Research design

The results of the ILS (Index of Learning Styles) questionnaire were used in the paper with 63 students of two courses who took part in the same "user interface" subject and data collected from Moodle during the course development. 44 questions (11 questions each), based on the theory of Felder Silverman, were answered from ILS (index of styles of study) questionnaire (Felder and Solomon, 2001). Their learning style can be calculated for each dimension based on the results of this questionnaire. During the course, moodle and data were collected concerning their interaction with this course. Moodle log data provides student activity information such as: what materials, access time, etc. they visited. Based on the results, three grading algorithms were implemented to detect the most precise technology in data mining that can predict student learning styles from their interaction with Moodle.

3.2 Data collection

Data from Moodle's questionnaire and log data are collected for each student. They are classified into one of five classes for each dimension, based on the questionnaire (information process, information perceived, information presentation and learning process) (table 1-4).

Table 1: Information Process Dimension

Classes	Count
Moderate Activist	14
Strong Activist	4
Balanced	41
Moderate Reflector	4
Strong Reflector	0

Table 2: Information Perceived Dimension

Classes	Count
Moderate Sensing	31
Strong Sensing	18
Balanced	13
Moderate Intuitive	1
Strong Intuitive	0

Table 3: Information Presentation Dimension

Classes	Count
Moderate Visual	29
Strong Visual	8
Balanced	22
Moderate Verbal	4
Strong Verbal	0

Table 4: Learning Process Dimension

Classes	Count
Moderate Sequential	3
Strong Sequential	0
Balanced	54
Moderate Global	5
Strong Global	1

3.3 Arrangement of data

Middle records of the students who have completed the questionnaire are filtered from the log data. These documents contain the activities performed by each student during the course using Moodle. The relationship history of the students with the learning objects is used to model their learning types.

3.4 Tools to be used

J48 Decision Tree, Naive Bayes and Part were used to classify students' learning styles using the Moodle Data Log. To use Weka tool, the data has been transformed to Weka file format in order to determine the best classification algorithm for learning style prediction (arff).

4. Data analysis

The information relating to the information presentation dimension is recorded for each student. The lessons that have been best used in visual education are graphics, tables, images, presentations with slides and videos, while text-based lessons with aims and content have been the teaching materials for verbal learners (Dung and Florea, 2012).

The course has instructional materials that are marked visually or verbally, depending on the above categorization. For every student who has recorded the number of visual material and verbal material he/she has accessed. The average of access for each

category for all students is calculated in order to compare the number of accesses of each student with each other.

Let's mention the number of visual material accesses to a student as n_{visual} and the average number of visits to all students in this category. Two K_{Low} and K_{High} parameters were used to categorize the number of access to each pseudo-code for the following categories {LOW, MEDIUM and HIGH}:

```
if  $n_{\text{visual}}$  less than  $a_{\text{visual}} * K_{\text{Low}}$  then  
VISUAL_CATEGORY = LOW  
else if  $n_{\text{visual}}$  less than  $a_{\text{visual}} * K_{\text{High}}$   
VISUAL_CATEGORY = MEDIUM  
else  
VISUAL_CATEGORY = HIGH
```

The same methodology is used for verbal data.

Three pairs of values for the parameters K_{Low} and K_{High} (0.33,0.66); (0.4, 0.7) and (0.5,0.8) have been used.

5. Results

Since the student's learning process was improved, the data mining techniques used for classification were: J48, Parti and Naive Bayes to provide information about the course grade and styles which the questionnaires identified. The data mining techniques were applied to students. No significant results were found. When the analyzed information was found to be different for each dimension of students who failed and did not follow any pattern.

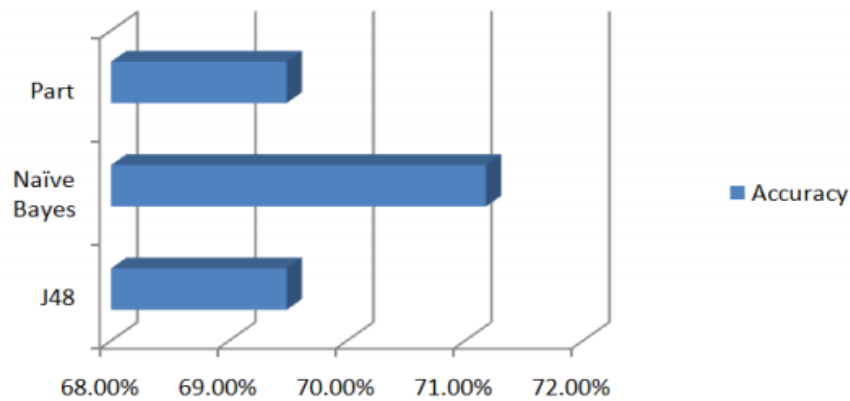
The data from the previous portion was used to perform the following experiments. Another attribute was added, which is the student's level of results based on the average of his grades. Three labels {LOW, MEDIUM, HIGH} defined the level of performance of the students.

The information presentation data for the dimension are imbalanced in Table 4. To prevent the overfitting problem the Weka Resample Filter produces a random data set sub-sample using either a replacement sample or without a replacement. Increased the number of instances in the Verbal class, which has the fewest, and the number of instances in the Visual class, which has the maximum.

The classification algorithms used were J48, Naive Bayes, and PART. These algorithms are applied to similar data in order to find the students' learning style in the information presentation dimension. Several experiments were carried out in order to combine attributes to define the best result attributes. For the three algorithms the best result is reached by using these coefficients: {0.5,0.8} and the level of performance for the students, which defines the number of visits for each material category.

Figure 2 shows the results of the three algorithms.

Figure 2: The accuracy of the three chosen algorithms
Accuracy



The best accuracy is achieved by the 71.18 percent algorithm of Naive Bayes as can be seen from this graph. The detailed accuracy by class in Figure 3 shows that the average area under the curve of the ROC Area is improved, showing that it is the best grade.

Figure 3: Detailed accuracy by class of Naive Bayes algorithm

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.808	0.303	0.677	0.808	0.737	0.502	0.731	0.679	Visual
	0.600	0.088	0.833	0.600	0.698	0.549	0.726	0.643	Balanced
	0.750	0.078	0.600	0.750	0.667	0.613	0.757	0.782	Verbal
Weighted Avg.	0.712	0.182	0.733	0.712	0.711	0.537	0.732	0.678	

6. Conclusions

We are currently developing a special Moodle data mining tool for online instructors that would avoid the need for CMS managers to help these instructors prepare or apply mining techniques. It has an intuitive and useful interface to data collection and preprocesses data from Moodle automatically to make data collection easier, thanks to its parameter-free data mining algorithms, to configure and execute data mining technology. Data from the ILS questionnaire and Moodle's log data were collected from 63 students from the User Interface course and log data from Moodle. The Weka algorithms were: J48, Naive Bayes and PART, respectively. It is thought that visual and verbal materials are visited with a coefficient to determine how much access the student has and how much performance the student has in order to predict the student style of learning on the visual/verbal aspect. The results showed that Naive Bayes algorithm achieves best accuracy at a precision of 71.18 percent. There was no relationship between style of learning and grades taken in the course. This results show that there is no preferred style of study, but the learning material for each style should be used to improve the student's performance.

Conflict of Interest Statement

The authors declare no conflicts of interests.

About the Authors

Mr Ali Akhtar holds Master in Computer Application from Jamia Hamdard University, New Delhi, India, Currently working as Researcher in department of IT and Quality Unit, College of Pharmacy, King Saud University, Riyadh-KSA.

Mr. Mohammad Serajuddin holds M.Sc in Computer Science from Jamia Hamdard University, New Delhi, India, Currently working as Researcher in IT department Dental University hospital, King Saud University, Riyadh-KSA.

Mr. Zafrul Hasan holds M.Sc in Computer Science from Jamia Hamdard University, New Delhi, India, Currently working as Researcher in College of Nursing, King Saud University, Riyadh-KSA.

References

1. C. Romero, S. Ventura, P. G. Espejo, and C. Hervás, "Data mining algorithms to classify students", in EDM, pp. 8–17, 2008.
2. S.-T. Wu, "Knowledge discovery using pattern taxonomy model in text mining," 2007.
3. J. Mostow and J. Beck, "Some useful tactics to modify, map and mine data from intelligent tutors," *Natural Language Engineering*, vol. 12, no. 02, pp. 195–208, 2006.
4. S. K. Mohamad and Z. Tasir, "Educational data mining: A review," *Procedia-Social and Behavioral Sciences*, vol. 97, pp. 320–324, 2013.
5. R. Baker et al., "Data mining for education," *International encyclopedia of education*, vol. 7, pp. 112–118, 2010.
6. C. Romero, S. Ventura, M. Pechenizkiy, and R. S. Baker, *Handbook of educational data mining*. CRC Press, 2010.
7. R. S. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *JEDM-Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.
8. Singh, M. R. K., & Yadav, M. S. Review on Data Mining Techniques and Variables Used in Educational Data Mining.
9. Manjarres, A. V., Sandoval, L. G. M., & Suárez, M. S. (2018). Data mining techniques applied in educational environments: Literature review. *Digital Education Review*, (33), 235-266.
10. Blagojević, M., & Micić, Ž. (2013). A web-based intelligent report e-learning system using data mining techniques. *Computers & Electrical Engineering*, 39(2), 465-474.
11. Lile, A. (2011). Analyzing e-learning systems using educational data mining techniques. *Mediterranean Journal of Social Sciences*, 2(3), 403-403.

Creative Commons licensing terms

Authors will retain the copyright of their published articles agreeing that a Creative Commons Attribution 4.0 International License (CC BY 4.0) terms will be applied to their work. Under the terms of this license, no permission is required from the author(s) or publisher for members of the community to copy, distribute, transmit or adapt the article content, providing a proper, prominent and unambiguous attribution to the authors in a manner that makes clear that the materials are being reused under permission of a Creative Commons License. Views, opinions and conclusions expressed in this research article are views, opinions and conclusions of the author(s). Open Access Publishing Group and European Journal of Open Education and E-learning Studies shall not be responsible or answerable for any loss, damage or liability caused in relation to/arising out of conflict of interests, copyright violations and inappropriate or inaccurate use of any kind content related or integrated on the research work. All the published works are meeting the Open Access Publishing requirements and can be freely accessed, shared, modified, distributed and used in educational, commercial and non-commercial purposes under a [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).