



## RESEARCH ARTICLE

# Educational Data Mining to Improve the Academic Performance in Higher Education

Alyaa A. Mahdi\*

Department of Computer Science, Cihan University-Erbil, Erbil, Kurdistan Region - F.R. Iraq

## ABSTRACT

Globalization and innovation are unlimited interest for the public sector and private business in the world, especially in the higher education institutions. Educational data mining (EDM) is mainly one of the business processes nowadays that attempt to bring the global innovation through improving and enhancing their processes and procedures to fulfill all the requirements and needs of the students as well as the institutions. The EDM considered mostly concern with any research concerning the applications of the DM and developing innovative techniques for DM in the educational sector. This study mainly combined the use of the powerful online E-learning management system (Moodle) with DM tools to improve the performance and effectiveness of the learning and teaching manners using the innovative daily data that collected from the educational institutions. Learning analytical enhance rubric (LAe-R) is the main tool that can retrieve the data from the activities of the students on the Moodle. The use of the LAe-R can lead for many benefits for both of the educators and students, for example, support the struggling students, a better way for students' advising, enhance the administration processes and enhance the performance of the students to learn from each other.

**Keywords:** Data mining, educational data mining, Moodle, learning analytical enhance rubric

## INTRODUCTION

Data mining (DM) playing a paramount role in many sectors such as industry, agriculture, retail, banks, health care, and education. Due to the globalization and the innovation in the past few decades, the educational institutions faced great challenges and to improve the administrative processes as well as to enhance and develop the academic activities. The International Educational DM Society defined the Educational DM (EDM): "An emerging discipline, concerned with developing methods for exploring the unique types of data that come from the educational settings. As well as using those methods to better understand students, and the settings which they learn in." EDM is principally concerned with analyzing the data that mainly generated from the process educational setup using different systems. Therefore, EDM's aim is to develop models that enhance and improve the teaching and learning process as well as the effectiveness of the institutions.<sup>[1]</sup>

DM is playing an important role in higher education, which is mainly, improves the level of academics processes. In every society, higher education institutions are mainly considered the center of knowledge and innovations; it is very important to use the huge data that available in these institutions to improve the level of their operations to prepare the new generations of students that equipped with the new skills to run in different sectors of the society. Hence, it is

necessary to develop the systems of DM that able to use the available educational data in the higher education institutions that will lead to offer high-quality educational processes to students as well as to enhance the management and research processes that will lead to the innovations and discoveries.<sup>[2]</sup>

Collaborative learning (EDM and LAe) is usually used to deal with the issues that are related to providing instructional strategies that support the collaboration of the students that are working together by groups. The main key indicator was the interaction to measure the effectiveness of the collaboration of the users' activity logs from their learning platform, which was used as the main key tool for the activities of the learners inferring to fit the demanded behaviors.<sup>[3]</sup>

EDM is considered as an emerging field in the area of DM. However, the educational setting is mainly used DM tools to

### Corresponding Author:

Alyaa A. Mahdi, Department of Computer Science,  
Cihan University-Erbil, Erbil, Kurdistan Region - F.R. Iraq.  
E-mail: alyaa.asaad@cihanuniversity.edu.iq

**Received:** Aug 24, 2020

**Accepted:** Sep 12, 2020

**Published:** Dec 20, 2020

**DOI:** 10.24086/cuesj.v4n2y2020.pp13-18

Copyright © 2020 Alyaa A. Mahdi. This is an open-access article distributed under the Creative Commons Attribution License

discover and analyze students' performance, to predict their results, and to avoid dropout also emphasis on the good and poor academic performers then provide feedback for the instructors and faculties as well as visualization of data is to improve the assessment process. The modern educational institutions mainly need DM techniques to set strategies and plans for the future. Mainly, the top-level educational authorities of the institutions may apply the outcomes of the EDM results to realize the main key trends and behaviors of the students' performance, which may lead to design the innovative pedagogical strategies.<sup>[4]</sup>

## BACKGROUND

### EDM

Mainly the EDM is considered as a developed research field to collect big amounts of students' data then analyze their activities to understand their behavior, which is the base of creating intelligent EDM. The analysis of learning and knowledge of the students' performance is the platform used in a learning environment to measure, scrutinize, as well as report the patterns of learning behavior to improve teaching and learning processes.<sup>[5]</sup> EDM is concerning developing unique methods as well as discovering the educational environment for a better understanding of the students' behavior. In another word, EDM uses a large repository of students' data from the system such as the Moodle with a view to give information regarding how a student is learning process. As well as, the internet and client/service interface utilization that provides a very big repository of the information about students learning process and of learning materials availability and teaching methods.<sup>[6]</sup>

Recently, EDM garnered the significance for improving and personalizing students' learning process. Many researchers focused on predicting the performance of the learners mainly such a task is not simple to achieve, especially during the course delivery time.<sup>[7]</sup>

Learning analytics and EDM have been efficient and effective really within a short time. However, the estimates of (student-at-risk) systems are mainly employed at the scale of higher education as well as the adaption of new learning techniques and systems that are used widely in higher education. As well as there is a steady stream of models and discoveries in a very wide range of once challenging areas to the study: The collaborative learning classroom participation, as well as online connections, students' engagement, and motivation, the meta-cognition as well as self-regulated learning, and many other areas.<sup>[8]</sup>

EDM provides appropriate opportunities to evaluate and assess students in an enhanced technological learning system that is considered an easy and efficient platform if we compare it to the traditional systems. To evaluate the students, many factors have to be taken into consideration like their login and logout on the learning system, the time that they spent on each page, participation in the classwork, and the discussion form.<sup>[9]</sup> EDM tools are embedded as a plugin tool in the Moodle and the use of these tools is easy even for the non-experts.<sup>[10]</sup> EDM adoption requires a structure and steps for an efficient design that can be upgradable to the system. The methodological

structure of this study is mainly based on using data called the cross-industry standard process for data mining (CRISP-DM). Figure 1 shows the steps and the processes of implementing CRISP-DM.

CRISP-DM technique is mainly recorded as one of the successful guidelines or ways of implementing a DM system. According to Sahay and Mehta,<sup>[11]</sup> the foundation of the CRISP-DM method was by the European Community Research Fund then the statistical companies such as SPSS and NCR developed it. The following guidelines discussed briefly the implementation of the DM in the educational sector:<sup>[2]</sup>

1. Understanding the business: The main aim is to understand the objectives and structure of the educational institutions to fulfill the requirements.
2. Understanding the data: The main aim is how to deal with and represent the collected data.
3. Data preparation: To transfer and clean the collected data to finalize the data of Moodle.
4. Modeling: The main aim is to apply the technique to represent the CRISP algorithm.
5. Evaluation: To measure if the model fulfills the requirements of the educational institutions.
6. Deployment: Implementing and starting the work in the project.

Educational Data Mining has significant importance in the higher educational sector. Various studies showed that the application of EDM can enhance the teaching and learning processes in the universities they studied.

For example, at Ajman University, different colleges of science and technology in the United Arab Emirates significantly after applying the EDM mentioned that the performance of the students is not very dependent on the academic efforts, but some other factors that may have equal to greater influences. EDM motivating and helping the universities to perform EDM on the data of their students' to find out important results

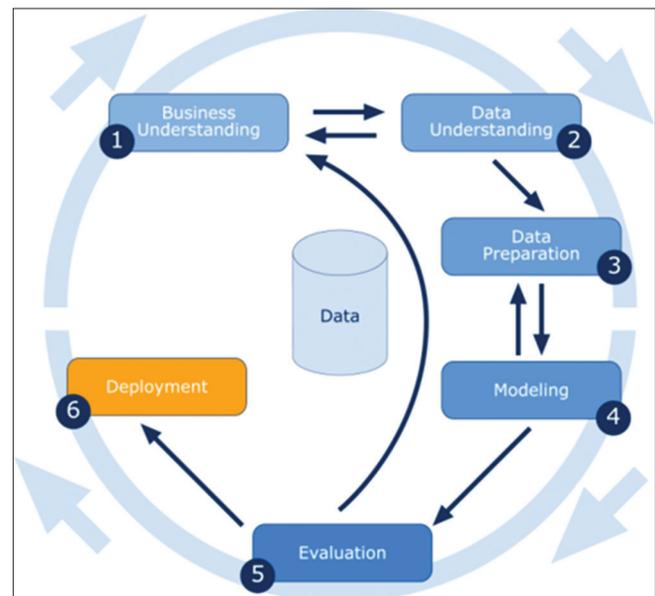


Figure 1: Steps of data mining

that can help both the university, educators, and students in several ways.<sup>[12]</sup> As well as another EDM application, the case was in from VBS Purvanchal University, Jaunpur (Uttar Pradesh), the Department of Computer Applications courses, Master of Computer Applications to predict their students' performance by the end of each semester. Significantly, the use of EDM supports of both the educators and students to enhance the division their student to identify those students that mainly needed special attention to reducing the ration of fail as well as to take suitable action for the examination of the next semester.<sup>[13]</sup>

### Higher Education Institutes in Iraq

The Iraqis have a wide intellectual history from the ancient of the Mesopotamian civilizations the remarkable accomplishments period throughout the Arab-Islamic in the 9<sup>th</sup> century. The higher education in Iraq dates back to the last century in the year 1908 that Law College in Baghdad was established, then Academies College was established in the 1950s.

A number of the other colleges as well as the academies later all these colleges combined as Baghdad University in the 1960s. The flourishing of the higher education sectors was between the 1960s and the 1970s. The Iraqi college of sciences was the nexus for the research and the literature. Medicine and science colleges attracted students from many countries to Baghdad University.

The Iraqi universities according to surveys of the UNESCO in the year 2004 were about 200 colleges also about 800 departments and 28 specialized institutes, in addition to the American style private universities in the Sulaimaniya and Koya. However, only Missan and Muthanna governorates were without universities. Then, in 2006, the public universities included in Iraq Sulaimaniya, Dahuk, Koya, Tikrit, Kirkuk, Mosul, Anbar, Baghdad, Irbil, Islamic Studies, al-Nahrain, al-Mustansiriyya, Diyala, Technology, Qadisiyya, Karbala,

Wasit, Kufa, Thi Qar, Babil, and Basra. In addition, the number of private universities was (7) in Baghdad and (2) in Kurdistan Region.<sup>[14]</sup>

According to the UNESCO survey in the year 2003, the number of students in the non-profit colleges was about 400,000 students, as shown in details of Figure 2:

- The ratios of the students were 58% of male students and 42% of female students.
- Approximately 50% of the students attended Baghdad universities (Baghdad, al-Nahrain, al-Mustansiriyya, Technology, and Islamic studies).
- Near to 32% of the students studied the education, 28% were in engineering and sciences colleges, 15% in the law and social sciences colleges, 13% in the medical college, and 12% were in humanities college. The total number of the academic staff was 19,000 member, 36% from them had doctorate, 58% of them had master degree, and also only 6% of them had bachelor degree.

Most of the Iraqi universities did not have the system of data collection, but mainly in the last years, some of the universities, especially private universities, were started to utilize student information management systems with a view to collect and manage information of the students. Some universities mainly used systems like Moodle for the academic purposes. However, these systems have produced data and information regarding the students in the universities. The generated data were used to produce some reports to improve the operations of learning, teaching, and research at these universities.

### DATA COLLECTION SYSTEM

In this study, the data collection was mainly suggested to be used by the Iraqi universities and the Kurdistan region. This system will mainly contain the modules for uploading and downloading lectures and assignments with the sign-in from



Figure 2: The main actions in the Moodle learning system

the students and the lecturers. In addition, it includes the modules for scheduling academic activities. Mainly that will provide the use time and resources efficiently and effectively for the university. Moodle is a good example of such a system. This system if it's deployed by the university then used by the lecturers and students will mainly provide an efficient tool to enhance the academic environment.

### Moodle as an Academic Tool

Moodle is basically a free online learning management system that provides for administrators, educators, and students by a robust and dynamic platform for all the academic processes. Moodle system aims are to help the instructors to present the knowledge better and to support in acquiring skills, knowledge, and experiences.<sup>[15]</sup> This online system supports the academic staff to enhance their performance and for better assessment methods for their students. Furthermore, this system can be used by many other business processes such as the schools, the university, and the firms. The Moodle is a free download system, configure, and also an open-source system that can serve various business aspects.<sup>[16]</sup> Moodle system is robust with some training that can be modified and they are available to the user for any higher education institute. The most important activities in Moodle are shown in Figure 2.

### DM Tool Based on Moodle

Learning analytical enhance rubric (LAE-R) is considered as one of the main tools that can retrieve the data intelligently from the activities of the students on the Moodle to predict the behaviors of the learners to produce clear patterns or trends. Mainly this tool works on collecting data from the online activities of the students such as their login and logout from the system, their participation in the discussions, and

many other activities for a better understanding of their behaviors that will lead to a better assessment for the students. Rubric is an assessment tool developed installed and imbedded as a plug-in on computers of the users that can extracts the interaction students within the Moodle system. Figure 3 shows the mentioned learning system environment.

### Analysis and Visualization

An efficient visualization tool called Gismo that shows the end result of the LAE-R for the educators using graphs of any other descriptive tool. For instance, Figure 4 shows the students' representation login activities in Moodle; using this analysis, the teachers can easily assess their dormant students.

The analysis includes the pedagogical coherent system assessment aspects, which analyze how the students are accessing the lectures for example as a PDF and/or word formats on the Moodle, also studying time per each page of the lecture. The Excel pivot tables' provide the statistics of learning to show the data on spreadsheets report. From a big volume of data available on the Moodle using, the mentioned tools will provide the ability for the educators to extract the valuable information.<sup>[17]</sup> Furthermore, this tool helps the students to access their assessment and to get a comparative analysis and efficient recommendation.

The platforms of online technological aim to facilitate scalable peer-to-peer training and instructor instructors' learning dominant channels.<sup>[18]</sup> Combining Moodle (online learning management system) with the powerful DM tools like LAE-R provides an efficient and effective system for exploring a huge amount of educational data that available to the educational institute also this will provide many improvements to the educational process.

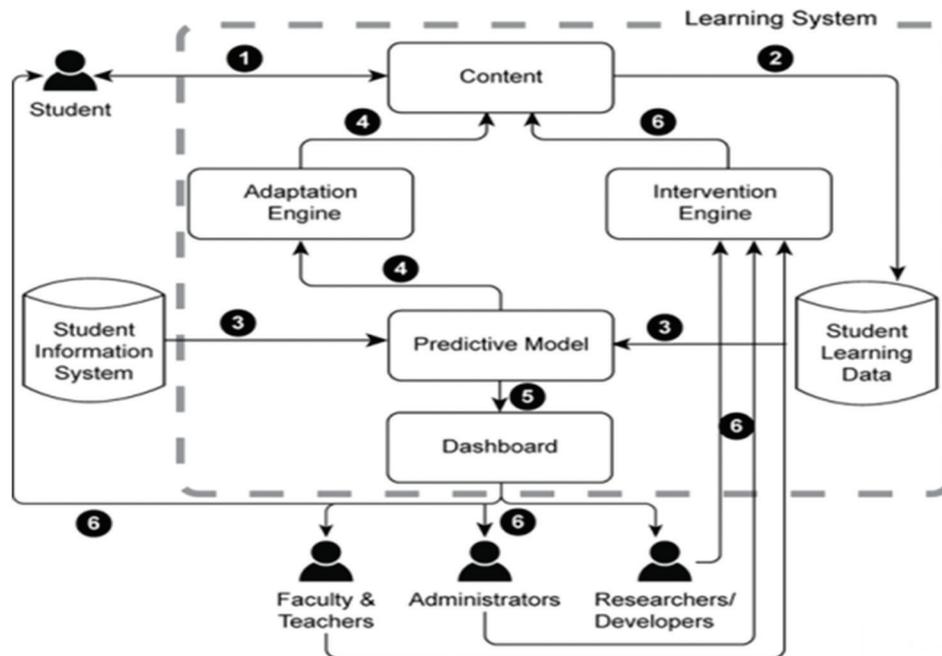
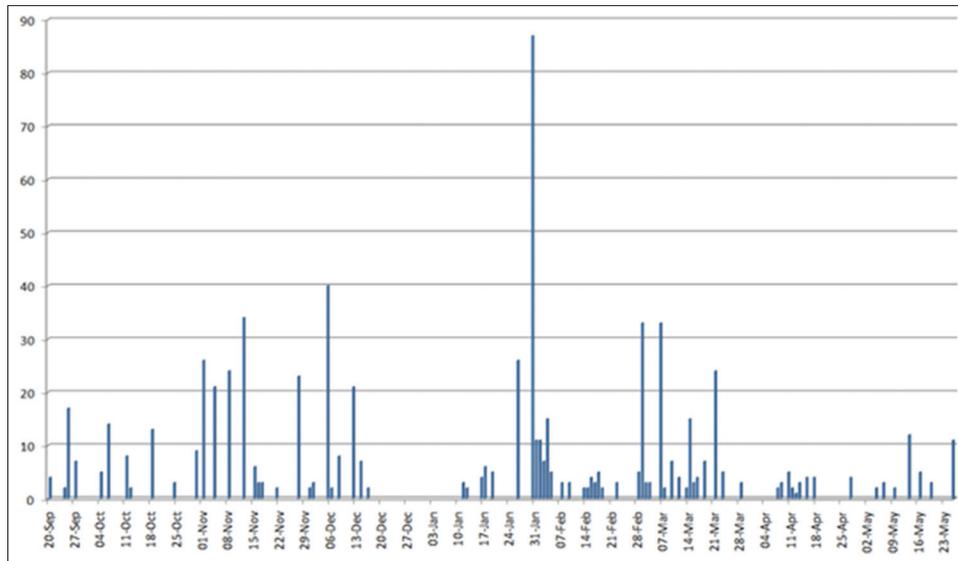


Figure 3: Framework of learning analytical rubric



**Figure 4:** Information about students login in full semester

## CONCLUSION AND FUTURE WORK

The study illustrates how the combination of combining between the online learning management system with the DM applications provides an effective and efficient tool for the higher education institutions using the tools that have been mentioned in this study that enhances the evaluation for the students' performance from their online activities on the Moodle system.

From the educator's perspective, the professor, or educators are able to support the struggling students, especially by providing a better way to advise students. Also, it is profitable for the marketing of the universities and supports and enhance the administration processes. For that reason, it will provide the capability to improve and enhance efficiency in the educational process.

On the other hand, from the student's perspective, it encourage and help them to enhance their performance and to learn from each other, leading them to be creative.

## ACKNOWLEDGMENT

Gratitude and appreciations to my family for giving me the support from various aspects and encouraging me to develop and complete my work. Special thanks for Ministry of Higher Education and Scientific Research in North Iraq that provides me by the adequate information.

## REFERENCES

1. A. Dutt, M. A. Ismail and T. Herawan. A Systematic Review on Educational Data Mining. United States: IEEE Access, 2017.
2. M. Chalaris, S. Gritzalis, M. Maragoudakis, C. Sgouroupolou and A. Tsolakidis. Improving quality of educational processes providing new knowledge using data mining techniques. *Procedia-Social and Behavioral Sciences*, vol. 147, pp. 390-397, 2014.
3. H. Aldowah, H. Al-Samarraie and W. M. Fauzy. Educational data mining and learning analytics for 21<sup>st</sup> century higher education: A review and synthesis. *Telematics and Informatics*, vol. 37, pp.

- 13-49, 2019.
4. S. Hussain, R. Najoua and N. A. Dahan. Educational Data Mining and Analysis of Students' Academic Performance Using WEKA. Berlin, Germany: ResearchGate, 2018.
5. Z. Papamitsiou, A. Economides and I. N. Sneddon. Learning Analytics with Excel Pivot Tables. Berlin, Germany: Moodle Research Conference Beuth University of Applied Sciences, pp. 115-121, 2012.
6. C. Romrto, S. Ventura, and E. Garcia. Data mining in course management systems: Moodle case study and tutorial. *Computers and Education*, vol. 51, no. 1, pp. 368-384, 2008.
7. M. N. Injadat, A. Moubayed, A. B. Nassif and A. Shami. Systematic Ensemble Model Selection Approach for Educational Data Mining. Netherlands: Elsevier, 2020.
8. R. S. Baker, Challenges for the Future of Educational Data Mining: The Baker Learning Analytics Prizes. *Journal of Educational Data Mining*, vol. 11, no. 1, pp. 1-17, 2019.
9. M. Petropoulou, K. Kasimatis, I. Dimopoulos and S. Retalis. LAe-R: A new learning analytics tool in moodle for assessing students' performance. *Bulletin of the Technical Committee on Learning Technology*, vol. 16, no. 1, pp. 2-5, 2014.
10. J. Hung, K. Rice, and A. Saba. An educational data mining model for online teaching and learning. *Journal of Educational Technology Development and Exchange*, vol. 5, no. 2, pp. 77-94, 2012.
11. A. Sahay, and K. Mehta. Assisting Higher Education in Assessing, Predicting, and Managing Issues Related to Student Success: A Web-based Software using Data Mining and Quality Function Deployment. Las Vegas: Academic and Business Research Institute Conference, 2010.
12. A. Saa. Educational data mining and students' performance prediction. *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 5, p. 070531, 2016.
13. B. K. Baradwaj and S. Pal. Mining Educational Data to Analyze Students Performance. *International Journal of Advanced Computer Science and Applications*, vol. 2, no. 6, pp. 63-69, 2011.
14. I. Harb. Higher Education and the Future of Iraq. United States: Institute of Peace, 2008.
15. H. F. Hasan, M. Nat and V. Z. Vanduhe. Gamified Collaborative Environment in Moodle. United States: IEEE Access, 2019.
16. B. Ngom, I. Niang, and H. Guillermet. Enhancing Moodle for Offline Learning in a Degraded Connectivity Environment.

- Tangier, Morocco: IEEE, 2012.
17. H. Dierenfeld, and A. Merceron. Learning Analytics with Excel Pivot Tables. Berlin, Germany: Moodle Research Conference, Beuth University of Applied Sciences, pp. 115-121, 2012.
  18. V. Z. Vanduhe, M. Nat and H. F Hasan. Continuance Intentions to Use Gamification for Training in Higher Education: Integrating the Technology Acceptance Model (TAM), Social Motivation, and Task Technology Fit (TTF). United States: IEEE, 2020.