THE EFFECT OF APPLYING DATA MINING TECHNIQUES (FUZZY-C MODEL) TO INVESTIGATE THE PERFORMANCE OF LIBYAN STUDENTS DURING USING E-LEARNING PLATFORM

Salem Msaoud Adrugi
Computer Department, Faculty of Education
Elmergib University, Libya
salem.adrugi@elmergib.edu.ly

Abstract—**This research study presents a methodology to investigate the performance of learners in Libyan universities using e-learning platform based on data-mining algorithms and extract the challenges which act as barriers to improve performance of learner through e-learning platforms. Investigating these issues, data-mining technical is introduced. Clustering based fuzzy algorithm is applied to two data groups related to five groups of learners (two groups are neglecting) the three groups of data consist of (committed, non- committed, active) students, The number of learner visits to the platforms is measured, it was 4247, and after clean processing the samples it is reduced to 1288. The clean stage removed the hits which are formed from using the search-engine. Results showed that the non-committed learners had the smallest size this due to many factors such as the difficulty to use platforms for some learners, also there is the poor network connection. We could see that data-mining tools are effectively ways to investigating the problems in e-learning improvement.**

Keywords: E-Learning, Libyan Universities, Fuzzy, Education, Barriers, Data-Mining

I. INTRODUCTION

Growing of technical applications made it is necessary to use this technology in all the fields especially in the educational sector, E-learning become spread on wide range especially during Corona pandemic. Which led to real thinking about improving the e-learning means. Several platforms become widespread in Libyan universities. This method has more advantages compared with traditional learning. It support self-learning which increase the level of ability of self-thinking and open the rang for creativity. Nowadays, the advanced technologies of educational systems become spreading on wide range [1]. These systems collect a major amount of information that is very importance in evaluating behavior of learner and supporting teachers in discovering of potential errors. However, the great amount of data getting from this systems which could produce daily creates difficulty in data management traditionally, so that it is necessary to propose new techniques which help in this process.

To achieve this goal, Data Mining is using [1], data-miming includes four stages, the first stage is the problem defining and the aim and the targets are formulating, the second is preparation of study data and collecting the information, this stage may takes the majority of study time [2.] It should identifying the source of data, cleaning and formatting in preselected. Third stage is Modeling and estimating the evaluation and the parameters should optimized. Finally deployment stage, which presents the process of organization data-mining results, [3]. In the past years, searches had been concentrated to implement diverse approaches of data mining. An overview could be observed [4], the exploration of modern knowledge related to learners is done by these tools [5].

II. PROBLEM STATEMENT

The quality improvement of e-learning systems is one of the major challenges. To achieve that, performance of learners should be evaluated [6]. Traditional (statistical) ways used to evaluate the learner online do not meeting the requirements due to the immensity of data which extracted from network [7], new concepts appeared with e-learning platforms such as the number of visits to the site, the multiple downloads, the traditional ways not capable to

dealing with this data so data-mining technologies are used in this case.

III. The Main Goal

The main objective is presented one of the data-mining tools (Fuzzy-C) to discover the students' performance and discuss the issues that facing e-learning quality development. Also this model aims to remove the challenges related to data of web.

IV. Research Questions

1. What is proportion of interactive of learners on platforms?
2. what the effects of fineness of quality on the results in the case of data-mining model based on fuzzy-C application?
3. What is proportion of frequently visits on platforms ?
4. Does the proposed strategy of data-mining offer good index about the e-learning affectivity ?

V. Related Work

A. E-Learning and Data-Mining

E-learning represents untraditional process to get the knowledge by using Information and communication technology. According to this concept, the educational content is obtainable online and provide feedback to the activities of learning for students. This concept implemented in many countries. There are many types of mass media witch facilitate the online learning by transmitting text, voice, video, additionally the CD-ROM, learning based on computer, smart tablets, extranet, web grounded education. This is a convenient way to flexible education. This process requires specialists in education have good expertise in ICT and professional to utilize systems of e-learning [8]. There are several benefits of e-learning for students, e-learning increases the learner's skills in utilizing computers and other technical application, e-learning include advantages for student such as:

1. The student could get access a several resources all over the word.
2. Effective tool that could ease education that can improve the student imagination and provide good visuals.
3. Support students and instructors' communication. E-learning also useful for instructors such as:
   - Simplify their works with regard to student presence, grades by using a computer.
   - Simplify keeping of record and rendering of diagnostic data.
   - The ability to communicate and participate with their counterparts and sharing educational units through computer.

Despite this features e-learning has some drawback which related to:

1. The absence of meeting face to face with teacher.
2. Loss of necessary personal interactivities.
3. Increasing the eventuality of illegal behaviors like plagiarism or deception.
4. For low self-motivation students, rates of success had decreased.
5. The low self-regulation students tend to not specify enough time to complete tasks.

E-learning has two basic forms to offer the information and learning contents to the learner synchronous form**,** in this case, both teachers and their students connect online directly at the same time, that means there is need to provide sharing application, sessions of whiteboard and rooms for chat. Video-conferencing is very important to support this kind of e-learning.

This process provides the ability to share questions and getting answers easily. The other is Asynchronous, for this kind, the communication online is not occurring at the same time, the students log in their platforms without specific time, they download their files and send to their teachers [9].

Recently, Libya government founded a powerful infrastructure. It supports research to ensure offering a convenient materials and media to distance learning. Presenting a modern technology to any society create challenges due to the different opinions and attitudes towards those new technique, it is suitable to take the cultural aspect of the various users into consideration. Another important point is the language.

Libya is an Arabian Islamic country, and the formal language is Arabic, e-libraries, e-learning contents, webs, are in English, this may create some difficulty. So that, it should be concerning in English language skills for all ages from learner and teachers. To make e-learning more powerful, achieve the advantages and follow the advanced countries it should supporting research to define the barriers to enhance this system and finding solutions to remove these negative factors.

[10], Because that e-learning is an electronic system, a great amount of data is produced, this fact offered reasons to enhancement the skills of decision for teachers and their students as like. Data mining is a technique which used to classify the selected data to make the accuracy of prediction at the maximum level [11], figure (1) showed the main types of data-mining algorithm. Data-mining has a significant role in developing the quality of e-learning. it is a technique that allows to process a great amount of data to discover and observe the deep information in educational fields.

However, it has some drawbacks related to the security as well as privacy. Due to the huge data processing, the implementation of this technique is very cost and complex. Additionally, it demands ICT experts to design the models which used in processing to meet the required results. Also, it is not exactly accurate.

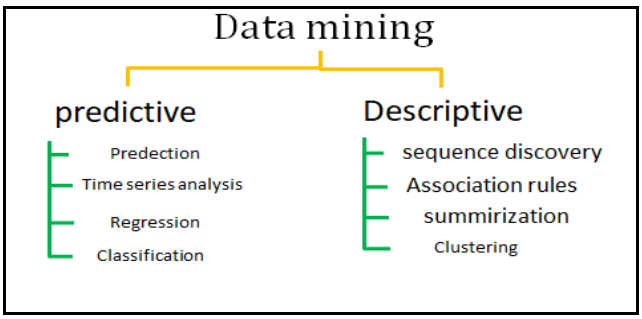

Figure 1. Classification of Data Mining

There are many approaches belong to data-mining, some of these approaches which used to apply in learning based online system are:

1. Fuzzy-Logic:

It is widely used for educational evaluation process, It is based on simulating human thinking. Figure (2) presented the stages to implementation this approach(Almohammadi,2017). Fuzzy-logic could measure the reacting between learns and transform it to linguistic expressions. This system could be used to predict the behavior of learners and it could combined with other algorithms to achieve more useful characteristics for processing more great and complex data.

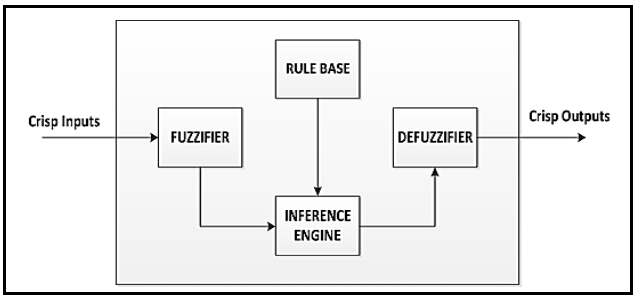

Figure 2. Fuzzy Logic Structure

2. Artificial Intelligence (Neural-Networks):

This kind works to simulate behavior of human and his actions. The structure of this strategy resembles the nerves of human. By training the proposed network, process the studied data processed then the results are obtained. There are many basic approaches are introduced in data-mining fields. Also many improvement and collaboration algorithm is designed based on these basic methods.

B. E-Learning Systems and Data Mining Techniques:

Learners and teachers had low level of knowledge related to use the computer [12]. Libyan students had the eagerness to engage with using technical application for different learning issues [13]. Recently, institutions of high educational had started to use ICT in learning process. In this case, there was a need to present the concept of virtual universities, this is a platform which offers all the services of university online.

[14] implemented qualitative to explore e-learning in the Saudi Arabian universities to develop knowledge on the development of Saudi Arabia's remoteness educational system. The collection of data made up of 28 in-depth, individual interviews with teachers and educational content creator to catch insider point of views, then objective analysis of essence backing documents of the universities was done. Three important objective fields were insulated and analyzed, (a) the rising pains of remoteness learning, (b) challenges related to learning integration, (c) instructional and technical aligning.

Viable System Model (SVM) presented explanation about the cause of Saudi Arabia's remoteness education system stayed applicable despite passing through periods of considerable variation. Complex interaction of the factors that affecting delivered learning related to the studied educational system [15].

[16] studied the factors affecting negatively on e-learning in Libya to find the solutions. Data was collected by interview additionally questionnaire, the sample study included learners and instructor of Al Asmarya Islamic university. It found that the major difficulties for e-learning presented by the insufficient knowledge related to advanced learning technologies, also financial backup is weak.

[17] presented study to discover the relation between learners opinions and their intellectual trends related to learning online and the demographic features and the ability to use the modern instruments of technology, also evaluated their skills in this field. [18] Designed questionnaire for both learner and teacher about the benefits and drawback of e-learning during quarantine period due to COVID-19, the results presented that the workload is reduced on the university and increased on learners, also the weakness of internet during this period is one of the most considerable challenges during this period.

[19] concentrated on providing an review on techniques of data-mining to expect the performance of learners and the specified an algorithm to introduce the features of the studies data. [20]. Presented a theoretical comparison to explain that the collaboration between e-learning and data-mining is more powerful. It was shown that data-mining has an important role in e-learning and it could predicted the performance. [21].

Presented data-mining algorithm depending on clustering, students were divided into collections and founded depending on great database. [22] discovered some techniques to support the instructor to make a right decision by create groups of learners. [23] presented that techniques for big volume of data in the education system, data-miming could be mange this data for adoptive systems.

[24] Used data-mining to expected number of registered learners to discussion the challenges that affect the learners commitment, this had implemented by decision-tree. [25] Identified the unobserved relation between the courses which learners did not success in them to introduce the patterns of learner failure based on data-mining.

VI. METHODOLOGY

Clustering based on fuzzy technique is a method applied on a large scale to get models based fuzzy from collected data. It achieves effective performance in different fields. The traditional concept of Clustering depend on exactly one of clusters, but Fuzzy-C offers possibility to represent data related to multi- clusters synchronously through the memberships. In this algorithm, a cluster represents an initial sample which made up of cluster represents the center, and information about form and size of this cluster. The parameters related to form and size specify the cluster extension in various direction of a particular area.

The Fuzzy-C algorithm which is introduced by (Bezdek, [26, 27] seeks to define partitioning based fuzzy of a specified training group. The following equation represents the objection-function minimizing to achieve the optimum solution:

$$f(u, C_1, C_2, \ldots\ldots, C_c) = \sum_{i=1}^{c}\sum_{j=1}^{m} uij^m \left\|x_j - c_i\right\|^2 \quad (1)$$

u_{ij} have a range (0-1), c_i : represent the center of Fuzzy set i.

m is weighting and it represents the exponent of memberships. In this algorithm, the matrix of membership u is permitted to get only 0 & 1

$$\sum_{i=1}^{c} u_{ij} = 1\,, \qquad \forall_j = 1,2,\ldots,n \quad (2)$$

partitioning based fuzzy is achievement through a repeated optimization process of the objective-function which is mentioned above and through updating u_{ij} and c_i :

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^m \, x_j}{\sum_{j=1}^{n} u_{ij}^m} \quad (3)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\left\|x_j - c_j\right\|}{\left\|x_j - c_k\right\|} \right)^{-\frac{2}{m-1}}} \quad (4$$

Fuzzy-C algorithm is implemented by these point as following:

1. Define the initial set of the matrix of membership with values (0 or 1) according to Eq. (2)
2. Compute the cluster-centers c_i ; $(i = 1,2,\ldots,c)$.
3. Calculate the objective-function as stated by Eq. (1), Stop whether it is lower a specific tolerance or its enhancement over prior iteration is under a specific threshold.
4. Calculate novel matrix for membership according to Eq. (4)
5. Return the number 2

When the variation between partition based fuzzy matrix for two repetition consecutive is below than ε, It can be stop the interactions.

The data registered in the server represents the access of platform by multiple learners. The basic data resources of platform integrated with miming are the data from server-side additionally client-side in our study (student).

A. Specifications of Data:

In this study, the data groups are selected from platform access-logs for investigating the performance of students in e-learning system. This data return to learners of Libyan universities. The behavior of students in online educational distance from their attending on this platforms can be classified at these categories:

1. Committed student: downloads a small amount of notes and follow this notes continually.
2. Uncommitted student: he downloads a big amount of notes, this implies to he did not log into platform for long time.
3. Non-interactive student: this student had not shared and reacted with the online courses also the did not download learning contents.
4. Active student: this kind is working on the tasks and assignments.
5. Non-existence student**:** this kind does not visit the platform and following his e-learning program.

B. Preparation Process of Studied Data:

One of the most important issues related to data-mining technology is the quality of selected data because the bad quality gives results have low quality. So that, this step is very important to success the data-mining algorithm. In this paper the penetration process comprises of two stages: cleaning process, abstraction process, also normalization. Cleaning stage includes hits removing from search-engine. Also it should be clean the data which is formed due to the visits of students who did not download the teachers notes, this could be implemented by eliminating this visits. The number of visits on the platform was 4248 which is decreased to 1287 after cleaning process.

C. Remove The Non-Interactive and Non-Existence from The Studied Data:

Those students did not participate and they are not interactive with the platform also they do not downloading so that cleaning is unnecessary for this type of students.

D. Keeping The Security of Students and Privacy of Studied Data:

This is a necessary step for identification of platform visitors. Particular areas of educational site should protected and the only learner could access by their passwords and IDs. The privacy protect includes also access to learners marks, assignments and task submission, password change.

E. Abstraction (for studied data):

This related to the type of students as above mentioned, some learners continue viewing notes regularly, however another students log in at particular times such as when the final exams become close.

To represent and simulate the learner behavior, it is presented the following assumptions:

- On/off platform access (the values are binary 0&1)
- Morning/evening access (assume 9 a.m. to 9 p.m.)
- Lab/non lab access
- Number of visits (Hits, represented as decimal (0-10).
- Number of downloads (represented as decimal (0-10).

To represent each learner visiting the platform.

VII. Analysis Study and Results:

By using Fuzzy-C technique, the learners classified into clusters. The five types of students (committed,

uncommitted, Non-interactive, active, Non-existence) which represented by the clusters. The data related to web-mining is formed from the data related to server-side and student-side. access log of Web-server also is a source of the data due to recording the behavior of browsing of platform visitors. This is represented by table(1).

The obtained data from logs to platforms for discussing 2 courses related to (computer-science& programming).

Table 1. Data Before Processing and After Processing

Set of Data	Hits number	Hits number after clean	Visits Number	Visits number after clean
1st Course	361508	342010	24734	7662
2nd Course	40053	37015	4247	1288

Table 2. Results of Fuzzy-C For 1st Data Group

Types of students	Behavior of each type of students					size
	Camp	Time	Lab	Hit	Req	
committed	0.001	0.75	0.35	0.47	0.70	1904
active student	0.97	0.93	0.67	0.99	1.19	2550
Non committed	0.68	0.821	0.46	3.19	5.99	396

Table 3. Results of Fuzzy-C For 2st Data Group

Types of students	Behavior of each type of students					size
	Camp	Time	Lab	Hit	Req	
committed	0.49	0.75	0.31	2.07	3.98	161
active student	0.54	0.71	0.43	2.43	2.77	1000
Non committed	0.58	0.77	0.46	2.26	4.85	25

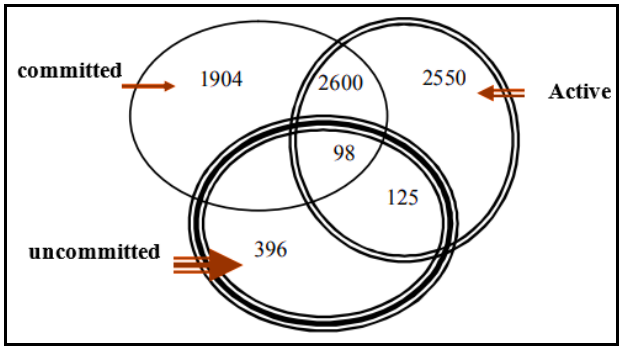

Figure 3. Fuzzy-C For 1st Data Group

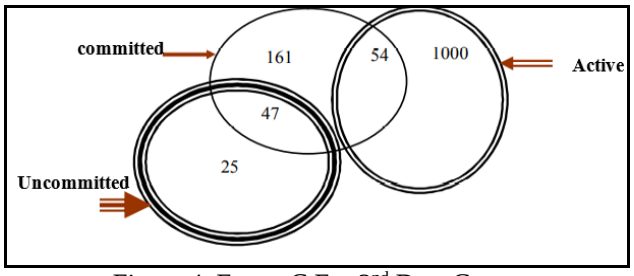

Figure 4. Fuzzy-C For 2nd Data Group

By neglecting the non-interactive and non-existence students and take three types into consideration, the result had shown the student behavior on the e-learning course, from table 1&2 it can be observed that :

1. Uncommitted students number of visits was lower than the number of both the active and committed students. Also, we can notice the raising number of hits related to non-committed students because the big amount of downloads process.
2. The active students size was the biggest.
3. The committed students size was smaller than active and the bigger than non-committed students. This is explained by the continuous downloading files and following the platform regularly.

Table 4. Clusters Size /Real Value Ratio

Data	(Clusters size /real value) %		
	Committed/real	Active/real	non committed/real
1st	81%	88%	90%
2nd	88%	90%	96%

Table 5. Questionnaire for students views

Students number%	Students' views
22.03%	Agree that e-learning is only an interactive way
16.99%	Refuse platforms
31.1%	Hybrid of traditional and e-learning
22.99%	Didn't use platforms
7.01%	Didn't use platforms for specific causes

According to above mentioned tables, some learners do not view e-learning as a suitable way because there are some materials require real presence and face-to-face way, some learners proposed that the educational system represented by e-learning integrated with traditional way for particular martials.

From the above mentioned results, It is observed that learners who has the smallest value of size were the non-committed group, this implies that there are factors effect on the performance. The reasons of this results could be relate to the learner such as the complexity of using the technical

application, reasons related to financial problems, the low quality of network in some places also play major role in non-download their learning files regularly and reduce their visits to the platforms. These challenges result in reduce the efficiency of e-learning system. Data-mining techniques is a good choose to process the data in electronic systems and it offers more features than the statistical methods.

VIII. Conclusion

In this study, it concentrated on applying Fuzzy-C model. Fuzzy conception is one of the data-mining tools. This model is used to discover the factors affecting in e-learning platforms in Libyan universities. The data is divided to five collections. Two of them are removed due to their low performance on the platforms. The random visitors and downloads are measured which form data-set then reprocessing is done for data. From the results it can be shown that the active learners have the biggest size, this implies that these learners continue their courses and follow the e-learning process regularly also this gives information about the number and time of visits of these learners.

It should be observed that the performance of active learner is better than the other groups. To increase affectivity of e-learning, it should be work on increase the interactive between learner and check the reasons which acts like barriers to develop the online education in Libyan universities. It can be seen that the data- mining is given good evolution of the data related to the learners' performance online, this could be implying how we could develop these platforms to raise the level of the high-education in Libyan universities.

IX. Future work

It should improve the e-learning tools and improvement the methods for specific materials which face a serious difficulty and need to use charts and symbols in the exams. Also, should providing courses for teachers about the importance of developing e-learning and how to apply these improvements, this increases the interactive. Improve the quality of network is an important point to make the communication easily, working on synchronous e-learning more than Asynchronous because the first type ensures interactive with students.

With regard to the administrators: it should get parameters about the ability to improve the platform efficiency and make it suitable to the learners' behavior, improvement the e-learning programs. Based on data-mining algorithms, so that, new approaches should be proposed and examined for data processing, to achieve more accurate in investigating the learners performance.

References

[1] Al Badree, A., (2007), Problems of Libyan higher education, Symposium of HE and Development in the jamahiriya (libya).

[2] Rhema, Amal, Miliszewska, Iwona, Sztendur, Ewa M. (2013). Attitudes towards e-Learning and Satisfaction with Technology among Engineering Students and Instructors in Libya, Proceedings of Informing Science & IT Education Conference.

[3] Elkaseh, A., (2015), An Investigation of the Factors for Adopting E-learning in Libyan Higher Education for Learning and Teaching. Phd of Murdoch University.

[4] Walabe & Luppicini, R., (2020), Exploring E-Learning Delivery in Saudi Arabian Universities, International Journal of E-learning & Distance Education, Volume 35.

[5] Al-Salman, S., Haider, A, (2021), Jordanian University Students' Views on Emergency Online Learning during COVID-19.

[6] Romero, et al., Herrera Evolutionary algorithms for subgroup discovery in e-learning: A practical application using Moodle data Expert Systems with Applications, 36, pp.1632-1644, (2009).

[7] Al-Azawei, A., Parslow, P., Lundqvist, K., (2016), Barriers and Opportunities of E-Learning Implementation in Iraq: A Case of Public Universities, International Review of Research in Open and Distributed Learning Volume 17, Number 5.

[8] Hammad, A., Zohry, M., (2020), Obstacles Hindering the Implementation of E-learning in the Faculties of Tourism and Hotels in Egyptian Public Universities, Journal of Association of Arab Universities for Tourism and Hospitality. Vol.18, No.2.

[9] Al Ghawali, E., Yahia, S., Alrshah, M., Challenges of Applying E-Learning in the Libyan Higher Education System, Al Asmarya Islamic University, Libya.

[10] Rhema, A., Miliszewska, I., (2014), Analysis of Student Attitudes towards E-learning: The Case of Engineering Students in Libya, Volume 11.169-190.

[11] Maatuk, A., Elberkawi, K., Aljawarneh, S., Rashaideh, H., Alharbi, H., (2021), The COVID 19 pandemic and E learning: challenges and opportunities from the perspective of students and instructors, Journal of Computing in Higher Education.

[12] Dihoum, B., Salih, H., Al-Qiblawi, S., (2022), Activating E-Learning in Schools in Libya Case Study: Bashir Solah High School, Journal of Electronics and Communication Engineering Research. 24-30.

[13] Kenan, T., Pislaru, C., (2012), Challenges Related to the Implementation of E-learning in Higher Education Institutions in Libya. Computing and Engineering Researchers' Conference, University of Huddersfield. 116-122.

[14] Alshref, H.,Abas, H., Bakar, N.,(2021), The Adoption of ICT in Libyan Higher Education Institutions (LHEIs): Theoretical Models and Challenges, Journal of Physics: Conference Series.

[15] Rhema, A., Miliszewska, I., (2010), Towards E-Learning in Higher Education in Libya, Issues in Informing Science and Information Technology, Volume 7, 423-437.

[16] Othman, Aisha, (2010), Investigating an On-line Teaching and Learning Environment for the University of Omar AlMukhtar, Libya, University of Omar Al-Mukhtar, Libya. Master thesis.

[17] Benghet, M., Helfert, M., (2014), Factors Influencing the Acceptance of F-learning Adoption in Libya's Higher Education Institutions. International Conference e-Learning, 405-408.

[18] Adrah, M., Elmarash, G., (2020), The Virtual University: Trends and Challenges in Libya Elmergib University as Case Study, Third Conference for Engineering Sciences and Technology, 1-11.

[19] Almihoub, A., Zaid, E., Bianchi, B., (2022), An Application of Web-Based E-Learning in Online Learning for Students at Future High School Libya, Bulletin of Science Education,52-57.

[20] Kamal, J., Dave, M., (2022), Analysing the Adoption of E-Learning Experience in School Education System of India: Challenges and Issues, Jagannath University Research Journal (JURJ), 15-22.

[21] Almohammadi, K., Hagras, H., Alghazzawi, D., Aldabbagh,G., (2017), A Survey of Artificial Intelligence Techniques Employed For Adaptive Educational Systems Within E-Learning Platforms., JAISCR, , Vol. 7, 47-64.

[22] Saleh, M., Palaniappan, S., Abdalla, N., (2021), Predicting Student Performance using Data Mining Techniques in Libyan High Schools, Edukasi, Volume 15, 91-100.

[23] Cui, Y., Chen, F., Shiri, A.,Fan, Y. Predictive analytic models of student success in higher education: A review of methodology. Inf. Learn. Sci. 2019, 120, 208-227

[24] Bezdek J., Pattern Recognition with Fuzzy Objective Function, Plenum, New York, (1981).

[25] Bezdek J. Some, non-standard clustering algorithms, In: Legendre, P. & Legendre, L. Developments in Numerical Ecology. NATO ASI Series, Vol. G14. Springer-Verlag, (1987).

[26] Shahiri, A., Husain, W., Rashid, N., (2015), The Third Information Systems International Conference A Review on Predicting Student's Performance using Data Mining Techniques,Procedia Computer Science 72, 414-422.

[27] Prakash, R., Hanumanthappa,M., Kavitha, V., (2014) Big data in educational data mining and learning analytic, Comput. Commun. Eng., vol. 2, 7515-7520.