

Empirical Study of Data and Web Mining in Education

Yehia Helmy

Faculty of Commerce & Business
Administration
Helwan University, Cairo, Egypt
ymhelmy@yahoo.com

Osama Emam

Faculty of Computers and
Artificial Intelligent
Helwan University, Cairo, Egypt
emam_o_e@yahoo.com

Ayman Khedr

Faculty of Computers and
Artificial Intelligent,
Helwan University, Cairo, Egypt
aymankhedr@gmail.com

Mahmoud Bahloul

Faculty of Commerce & Business
Administration, Helwan
University, Cairo, Egypt
mahmoud_bhloul@yahoo.com

Abstract — Growth of a nation is strongly determined by the quality of its education system. Over the past few decades, educational sector has expanded tremendously, the growing work has caused constant change in the syllabus and course structure of various systems. The biggest issues that educational institutions face are the rapid growth of educational data and the use of that data to improve decision quality. So there is always a need to build an Information system that can learn from the data smoothly which achieved successfully by using various data and web mining approaches. Data mining techniques are commonly used for numerous studies in education. Also, look at the vast volume of data displayed on the web pages, there is a great emphasis on collecting this information from various organizations and businesses to use it to their best advantage. This is achieved by web mining that uses the data mining techniques and algorithms to extract information and knowledge directly from the Web. We presents in this paper empirical studies in the Data Mining and Web Mining fields related to Education Sectors to assess the actual state of the universities and to identify the main guidance for future action.

Keywords: Data mining (DM), Web mining (WM), Education Institutions, Education Data Mining (EDM).

1. INTRODUCTION

Today Universities not only generate graduates but also vast amounts of data from their systems. Educational institutions are essential part of our society, and play a significant role in the growth and development of the nation. Therefore, the question that arises is how a higher educational institution can meet the demands for its strategic use of this instructional data. Across the globe, the education sector has undergone a major change in its performance, which can be defined as an industry, and faces challenges like every other industry.

Currently, a massive amount of data in various data sources outweighs the human capabilities to analyze and retrieve the most valuable information without the assistance of automated analytical techniques. Discovery of knowledge is the process of the non-trivial extraction from a broad database of tacit, unknown and potentially useful information [1].

Data Mining (DM) delivers powerful techniques for different fields including education. Data mining techniques are being applied in many specialized universities today to analyze available data and to gather information and insights in support of decision making. Work in data mining in education field has been recorded since the mid-1990s and interest in this field of research has grown since then, and since 2008 Educational Data Mining (EDM) has been organized annually [2,4].

EDM was concerned with developing methods for evaluating different types of data from educational settings and using those methods to better understand students and their learning environments. The educational settings hold information related to the educational process such as student attendance, performance, and profile information [1].

The growing of internet usage has contributed to vast quantities of unprocessed data being generated as a result of technological development so automated tools or intelligent methods are needed find and evaluate the needed information. Web has become the main education and e-commerce

platforms, therefore it is important to track and evaluate the users' access habits. So, Web Mining is called extracting information from the web using data mining techniques. Web mining is categorized into three major sections: Web Structure Mining (WSM), Web Content Mining (WCM) and Web Usage Mining (WUM) [3].

The scope of this research paper presents literature on the use of data mining and web mining in Education. The remainder of this paper is organized as follows. Section 2 discusses Overview on Data Mining. In section 3 discusses Overview on Web Mining. Section 4 presents Literature Review for Data Mining. Section 5 presents Literature Review for Web Mining and conclusions are discussed in section 6.

2. OVERVIEW ON DATA MINING

Data mining is a process of analyzing and deriving knowledge from large amounts of data stored in the form of Data Warehouse. Data Mining or the discovery of knowledge has become the field of increasing interest as it helps to interpret data from various viewpoints and translate it into useful information [2]. The use of data mining in education is increasingly of research interest. DM has taken on a variety of fields including bioinformatics, ecommerce, fraud detection and recently, education as well, and in the field of educational research is known as Educational Data Mining (EDM).

The Educational Data Mining is field concerned with developing methods of discovering knowledge from educational environment data. The Educational data mining have many goals as optimum instructional sequences are categorized using advanced domain models which enhance the learning content and also can obtain accurate details from the students by predicting their patterns of learning behavior. The discovered knowledge can be used predict student enrolment in a particular program, to organize the syllabus, detect abnormal values in student outcome sheets, detect unfair means used in online examination, etc. [5].

The tasks in data mining can either be descriptive or predictive. Descriptive data mining uses association rule

mining, clustering, etc. techniques to identify patterns hidden in broad data collection, and to assist in smart decision making. Predictive data mining constructs models for predicting the class of a new data set using rule set, decision tree, neural nets, and support vectors etc. and the common key data mining education technologies are generally broken down into association rule, regression, classification prediction, clustering and diagnosis [6,7].

- Classification is one of the types of prediction which classifies data based on the training set to create a pattern and uses this pattern to classify new data (test set). The test data are used in classification system to determine the efficacy of the classification rules. If the usefulness of the rules is appropriate then rules should be used for new data sets.
- Clustering is the process of grouping records in similar classes, and different from records in other classes. Methods of clustering are attempted to find the approximate solution to a problem. It is an iterative process of discovering the knowledge which involves the methods of trial and failure. Changing the parameters of the model and the data pre-processing will be absolutely necessary to achieve the desired results.
- Association technique is used to search through a huge data set for the most frequently available data item. The main objective of the association technique is to search for the most impressive association and interrelationship between a vast set of data. Apriori for example is a common algorithm for the rule of association.

EDM is the use of data mining techniques to collect data from an educational context. Extract data from educational sector has broad spread applications that can be used for different purposes such as forecasting students' needs, evaluating their progress and helping the institutions. The data in educational institutions can be personal or academic data that can be used to understand the behavior of students by simulate a student models, help instructors, improve teaching and improve e-learning systems and other benefits [8].

2.1 EDUCATIONAL DATA MINING TASKS

A- Organization of Syllabus:

One use of data mining is to classify similar subjects in a broad educational institution's syllabi of educational programs to optimize the use of resources for educational institutes. Currently, syllabic organization is affected by various factors such as associated, competing or partnering university programs, lecturer availability and expert judgments.

B- Predicting the Registration of Students in an Educational Program:

Getting and advancing over other educational organizations involves deep and adequate expertise for better assessment, evaluation, preparation, and decision-making. Data mining helps organizations recognize the secret trends in databases, patters extracted are used to construct data mining models and can thus be predicted with high accuracy in performance and behavior.

C- Predicting Student Performance:

Data mining provides several tasks that could be used to test success of the students through extract knowledge that describes students' performance in end semester examination. Also identifying the dropout rates and students who need particular attention and providing adequate advice to the teacher.

D- Detecting Cheating in Online Examination:

Now online tests are conducted remotely over the Internet for a few days and one of the basic problems is scam occurs. Data mining techniques can propose models that can help organizations detect online assessments and prevent cheats.

E- Identifying Abnormal / Erroneous Values:

The data stored in a database may reflect outlier, exceptional cases or data objects that are incomplete. As a result, the exactness of the patterns discovered may be poor. One of Outlier Analysis applications can be detecting the abnormal values in the students' result sheet [9, 10].

2.2 APPLICATIONS OF DATA MINING

Data mining in various ways is commonly used in many sectors of the world today. Many organizations have now commenced to use data mining to compete with the existing data analysis environment. Various mining methods and techniques are used to get fast and simple analyses of the trends and patterns of the prevalent industry and to create a quick and efficient market trend analysis.

There are many Data Mining applications such as financial data analysis, telecommunication industry, retail industry, science & engineering, web mining and etc. Web Services applications are rapidly growing through creates a huge amount of Web-based data that has its own features. This makes exploration more demanding within the Web Data Mining field [11].

3. OVERVIEW ON WEB MINING

Web mining is a sub-part of Data Mining in which different mining techniques are applied to the produced data and also working on the web to find out interesting and useful patterns previously unknown. Web is a phenomenon which is very large and stems popularity from the fact that it provides a vast wealth of information about nearly every imaginable subject.

Web mining is classified into three types of mining according to the different analyzed data sources. Web content mining focuses on extract information of web page content. The extract of knowledge from user navigation data when visiting a website that means web usage mining. Finally, web structure mining that extract knowledge from hyperlinks of the web as shown in Figure 1[13].

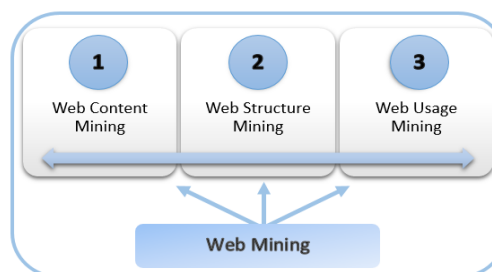


FIGURE 1: CATEGORIES OF WEB MINING

A. WEB CONTENT MINING

Web use mining is a very useful area to analyzing users and their web content behavior. Web content may

include a very wide array of data such as large volumes of multimedia data as images, video and animations, in addition to the typical hypertext or document data that exists in different formats (i.e. pdf). • There is multimedia web mining focuses on mining methods applied to multimedia data and also opinion mining and sentiment analysis that Identify and acknowledge the opinions and emotions of user posts and comments in web forums , blogs and other social media sites.

Web content mining in education includes methods of extracting useful educational content from the Internet to meet unique consumer needs and valuable information that may aid educators in the learning process and in decision making. The online tools providing educational content reflect the general pattern of the online where knowledge is evolving constantly. Text mining and opinion mining are web content mining techniques that can be used to gain knowledge from the textual data.

B. WEB USAGE MINING

Web usage mining focuses on knowledge discovery from user activity while browsing websites and web applications. This data can be found in different formats and in a variety of locations, such as web server access logs, user logs, cookies, and proxy server logs that depending on whether server, client or proxy level takes place. The browsing behavior that the user records will then be stored in remote databases.

In educational environments, web usage mining can be used to consider student learning patterns and the learning process context in which they occur. Web usage mining can be used in educational environments to answer about the most that navigational routes the learners have followed and which of these routes contribute to successful learning or determine the pages/topics that students skip and the time spend with a page etc.

C. WEB STRUCTURE MINING

Web structure mining means knowledge extraction based on the structure of the Web. The Web contains a large variety of non-unifying objects that can be connected to one another either by hyperlinks or other types of social connections. Web structure mining is also gathering knowledge from a web graph which is created either by extracting links between web pages or by extracting other types of links resulting from interaction and communication between social web users.

Web structure mining also has applications in education that focus on social software applications and services as a result of the rapid evolution of Web tools such as blogs, Facebook, Twitter and so on. The exploration of social relationships may be used for drawing important conclusions about the learner behavior and to explore communities that are established in these environments [12, 14, 15].

4. LITERATURE REVIEW FOR DATA MINING

In [16], focused on the identification of attributes which influenced the performance of the third semester students. The effect of emotional quotient parameters was developed on the placement. J48 and Random Tree are two decision tree algorithms were used to construct the model that finding the impact of different attributes on the performance of the students. Random tree gave higher accuracy of prediction than J 48.

The research [17] have considered students who are pursuing Master of Computer Application (MCA) degree from Pune University through some of attributes that will

predict student's performance in the university examination. Neural network is a technique used to select the attributes from a set of attributes, and important attributes are identified and rules are generated based on the accuracy of correctly classified data. The data collection of 60 MCA students was collected from Pimpri Chinchwad College of Engineering, Pune University's department of M.C.A by using WEKA tool implemented in java language. The result indicates that if a student scores poorly in graduation and performs poorly in attendance and assignment then chances are he / she will perform low in unit testing. This will result in poor University result results.

In [18], the standard of higher education institutions implies offering the programs that most likely meet the needs of the education system's students, academic staff and other participants. This paper introduced a new Data Mining concept to be implemented in institutions of higher education. The proposed model "DATA MINING ADMISSION MODEL (DMAM)" assists in decision-making processes at the strategic levels of higher institutions and also controls the student admission disciplines. This paper introduced eight alternatives to admission using the same technique with different applicability rules in order to demonstrate many advantages and drawbacks of each alternative used, as well as to identify the best form of admission among these. The DMAM Admission Approach proposed was the best alternative and recommended for use in Saudi universities.

In [19], the proposed Higher Educational Data Mining System (HEDMS) framework deals with the creation of methods that discover valuable knowledge from data derived from the educational system. The collected data types historical and usage data reside in educational institution databases. The system proposed helps to obtain adequate results that consist of several steps in their case study starting with data collection, pre-processing, implementation of data mining techniques and results visualization. They obtained data from Moodle database of students. The proposed "HEDMS" would help the teachers at the right time in their decision making. The purpose of this study is to examine the decision-making process for instructors to help their students in the delivery of any course linked to every research department prior to any exams that could be categorized and clustered on the basis of Moodle use as in their case study. This paper used K-mean algorithm cluster that divided into 3 clusters each one of them had some features as follow:

Group 1 shows non-active students' cluster and it had 12% of all data.

Group 2 shows very active students' cluster and it had 42% of all data.

Group 3 shows active students' cluster and it had 46% of all data

In [20], the main Data Mining techniques are classification and prediction and are commonly used in various fields. Local classifier methods can be used to generalize global model rules. Applying decision tree algorithm Random tree provides greater accuracy in single data sets compared to other decision tree algorithm. Predicting at-risk students and providing them better preparation to enhance their success would definitely support their individual results and also the profile of academic institutions. Classification rules are generated in each node from training data sets to create Global model at central side. Using weka runtime environment and applied Random decision tree algorithm on "Parul institute students" data sets of each having 1000 records.

In paper [21], 150 students obtained as a data source from the Knowledge Institute of Computer Science and Technology Professional, data mining analysis, preprocessing, culminating in an effective mining analysis of the student achievement table data source. This paper makes full use of the performance data of the students, draws the relationship between curriculum and curriculum, and analyzes the curriculum order, also makes full use of clustering method and the association's mining algorithm. Based on the Apriori algorithm, the results of 60 points below the score set to "0" are looking for k-frequent sets. The relationship between mining courses exceeding 85 courses, the minimum support = 0.05, the score of 85 points or more records was set at "0" the association rules produced 2- set.

In [22], attempted to analyze the instructors' data associated with the student assessment to enhance the quality of the education and show the factors that influence the performance of the students. The student performance prediction is primarily tied to the quality of the teaching process. Several data classification algorithms are applied to the Turkey Student Assessment dataset to predict student achievement, investigate teacher performance, and find the best high accuracy classification algorithm. The data set is tested and analyzed using four J48 DT, MLP, NB, and SMO data classifiers. It is found that to predict the teacher results, using the attribute evaluation method on the dataset is helpful. The results show that the accuracy of the performance increased when the worst ranked attributes are removed compared to the case when the algorithms run on the dataset with all the attributes.

In [23], presented that the proposed quality monitoring system for teaching can effectively monitor the quality of the teaching. Data items that have a high correlation with teaching quality are evaluated using data mining techniques, and the higher education quality management model is built on the basis of the core teaching quality. The results can be used to modify existing indicators of teaching quality assessment, and to improve and enrich the five-in-one national quality assurance system for higher education. In [24], made full use of the system for universities can achieve more accurate and efficient management, including intelligent learning and teaching.

5- LITERATURE REVIEW FOR WEB MINING

In [25], the web accessibility plays very important role whilst anyone of students accessing educational websites. This study used k-means clustering algorithm on the educational website data and categorized the educational websites into fast, moderate, and slow by the clustering algorithm. Educational website data is classified by many algorithms such as Naïve Bayes, RBFNetwork, 48, and SMO algorithms. The educational dataset is clustered using simple k means clustering algorithm and it was loaded into Weka. The result classified using four different classification algorithms available in Weka to determine the web accessibility data. The experimental results show that the accuracy of J48 and Naïve Bayes algorithms is high, and it takes less time to build model.

In [26], introduced a various methods to investigate behavioral patterns of users within e-learning systems. The authors reviewed, quantitative methods which focus on specific segments of data that represent points of interest on and provide an overall perspective, also qualitative methods that applied to provide a more detailed perspective. Evaluation occurred using web analytics and advances to methods of data analysis, such as Formal Concept Analysis, towards the state of the art concept. This paper presented a

quantitative statistical analysis performed by the on-site PHP web statistics tool WATEC, through a quantitative-qualitative representation using conceptualization, knowledge clustering (FCA) and force directed graphs to a fully qualitative analysis.

In [27], paper focused on web structure retrieval, so that Web Pages as nodes and hyperlinks as edges to classify the web page as a common web page or related webpage. Concentrating on the Web structure mining category to define the URL structure content analysis defined for its domain target achievement. The final result is a web crawls structure oriented to an educational university based on its ac. in characteristics.

In [28], carried out various analyzes of web log data such as Visitor Hit Analysis, Page View Analysis, Time Analysis, Source Website Visitors, Pages Accessed Website and Document Number Downloaded. The research is undergoing a time analysis and page view analysis, Web Log Expert lit study. Web log data about user access has been obtained from the website of the educational institution. The time analysis examines the various time of day, weekdays and month days that most users got on the website.

In [29], presented a review techniques, trend, tools, and some of application areas from the Web mining. This paper proposes a cross-analytical approach between the skills acquired during university training and the skills sought by employers in Morocco.

In [30], paper explained a web data mining based modeling method is proposed for the quality of information-based education in colleges and universities. The quality of information-based education can be effectively improved by developing and reviewing the teaching content, collecting data characteristics according to the assessment results, and adapting the teaching methods and modes in a targeted way. The results show that the efficacy of the test results is between 10 percent and 60 percent compared to that of the information-based quality assessment of education under the traditional model, and its effectiveness shows a significant downward trend with the increase in the number of test data.

TABLE 1. SUMMARY DESCRIPTION OF DATA AND WEB MINING IN EDUCATION SECTOR

Ref	Year	Objective(s)	Advantages	Disadvantages
[16]	2014	The goal of this paper is to predict the output of MCA students during the third semester.	Early prediction of poor performance students helps managers take timely action to improve their performance through additional coaching and counseling.	Need to develop a decision support system to help the authorities identify the weak students and take timely action.
[17]	2014	The aim of this paper evaluate students' performance and select certain attributes which generate rules through association rule mining.	Predict the student's university performance by identifying students who perform poorly in unit Test, Attendance, Assignment and Graduation unit and providing additional guidance.	Number of attributes are little and not characterized about specific student. The results not accurate about specific values.
[18]	2015	The main purpose of this paper is to suggest a DM model for filtering out students who meet all eligibility criteria that match the admission methods. The suggested model could also determine the best admission approach a student should apply to at Saudi Arabia's King Abdul-Aziz University.	This research article suggested a new model to promote the proper vision of choosing the best form of enrolment to increase the efficacy of the educational cycle within Saudi universities and at the same time boost the performance of the graduates.	Need to check the method with data from other institutions in order to learn the accuracy and convergence of this algorithm so as to produce better results. Also need to obtain a relationship between the Admission methodologies of different Saudi universities. There is no KPIs suitable for all students separately.
[19]	2015	This research focuses on how data mining methods are applied to higher education systems through using the most common methods in the education sector on the most common framework called the Moodle system.	Distinct and descriptive defined for how to use different data mining techniques to enhance the learning of students by supporting teacher decision.	This results related to some data mining techniques not all of them so evaluation for another techniques needed and also need to apply with more datasets
[20]	2015	The aim of this paper prediction of student performance in distributed environment through classification techniques	Distributed data set Environment give higher accuracy for test data set.	Working with distributed data set environment always need a dedicated effort.
[21]	2016	This paper combines the Apriori algorithm and clustering technology with student achievement data through data mining techniques and performs curriculum correlation analysis.	The method of mining association rule is more reasonable and scientific than the traditional method of analysis of dependence between courses that provides a scientific foundation for management and decision-making at universities.	There are several variables that are not taken into account in the collection of attributes field data set The content of information is not comprehensive.
[22]	2016	The paper's aim to predict the performance of the instructor	The student predictive performance accuracy is	Need for more classification algorithms to apply more datasets in

		and investigate the factors affecting the achievements of the students in improving the quality of the education system.	increased by the attribute evaluation method on the data set.	order to see impacts of the various performance of algorithms on various datasets.
[23]	2018	The objective of paper is improving the quality assurance system and the quality of teaching in universities and colleges. Also it provides a reference for educational authorities to make scientific decisions.	A perfectly system of monitoring the quality of higher education teaching is implemented.	There is lack of information hinders the reproduction of past results for expanding and developing new studies which focus only on teaching and there is no specific recommendations.
[24]	2018	In this paper making full use of the system for universities can achieve more accurate and efficient management, including intelligent learning and teaching. Furthermore it will improve the standard of higher education and teaching. Results:	To provide accurate service and support scientific decision-making, researchers can conduct horizontal correlation comparison and longitudinal historical analyzes.	This paper lacks practical application to his model and focuses without providing any indicators on a particular field.
[25]	2015	The present study identifies the rapid, moderate, or slow accessibility of educational websites. Weka is used as a tool for data clustering and classification by this research.	SMO algorithm efficiency is too poor because accuracy is lower, and time is higher. Kappa statistics are lower, and a high error rate compared to other algorithms.	There is no comparison between other techniques Unable to detect a greater diversity of intrusions
[26]	2015	This paper present how an investigation can benefit from using both quantitative and qualitative methods through conducted on a locally developed e-learning platform called PULSE.	This research detecting behavioral patterns through proved the effectiveness of a mix of quantitative vs. qualitative approaches and improving the e-learning environment and its content.	Need to find more factors for determine the appropriate learning style such as types of ideal behaviors to which users adhere over time by using the e-learning system.
[27]	2016	This paper conducts a thorough analysis of the web structure retrieval scheme against the variant effect of periodic web pages in the field of educational that can be implemented with optimal performance strategies planned.	Easy method for storage and retrieval of periodic Site data rates	implement real time of object representation in the motive of educational Domains to restoration web structure and need more factors to analysis data efficiently
[28]	2017	This paper based Visit study of samples conducted using web log data from academic educational institution where the focus is on the data collection in web servers of academic educational institution and implement	This paper will boost the sector of trade, manufacturing, and education.	This paper focuses only one perception and there is no analysis for recommendations.

		through using Web Log Expert lite 9.3 tools and analyze.		
[29]	2018	This research apply the latest trends in web mining and data mining in the field of education and jobs in Morocco, through cross-analysis between university-acquired skills and the skills employers are searching for.	Conduct a quantitative and qualitative study of the competencies taught during the university period and the competences required by the job market	This research lacks practical applicable of the prototype of study and also there are no results.
[30]	2019	This work suggests a quality model of information-based education in Web-based data mining colleges and universities to effectively address some of the issues.	Compared to a traditional model, the quality model of information-based education in colleges and universities is better based on web data mining performance.	Need to using new data from various sources for foresight more perceptions. This research lacks practical applicable to his model and focus on specific area without gives any indicators.

6. CONCLUSION

The usage of technology in education generates a huge amount of data each day quickly which has become a primary goal for all researchers all over the world. Mining indicates things valuable to extract. Hence, it can be said that the field of data mining in the education sector is growing rapidly and has the advantage of incorporating new algorithms and techniques established in the various data mining areas. Education has also been affected by improvements in web mining, as a massive amount of education institutions services are available on the internet. In this study we discussed the various techniques of data mining and web mining that can support the education institutions through generating important information for decision makers. A structure will be presented in the immediate future as an improvement on the shortcomings indicated in the study's analyzed techniques.

REFERENCES

- [1] Martínez-Álvarez, Francisco, et al. "A survey on data mining techniques applied to electricity-related time series forecasting." *Energies* 8.11 (2015): 13162-13193.
- [2] Romero, Cristobal, and Sebastian Ventura. "Educational data mining: A survey from 1995 to 2005." *Expert systems with applications* 33.1 (2007): 135-146.
- [3] Bharti, Pooja M., and Tushar J. Raval. "Improving Web Page Access Prediction using Web Usage Mining and Web Content Mining." 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 2019.
- [4] Nachmias, Rafi. "Web mining and higher education: Introduction to the special issue." *The Internet and Higher Education* 2.14 (2011): 65-66.
- [5] Patil, Shubhangi D., Ratnadeep R. Deshmukh, and D. K. Kirange. "Adaptive Apriori Algorithm for frequent itemset mining." 2016 International Conference System Modeling & Advancement in Research Trends (SMART). IEEE, 2016.
- [6] Ahmad, Fadhilah, Nur Hafieza Ismail, and Azwa Abdul Aziz. "The prediction of students' academic performance using classification data mining techniques." *Applied Mathematical Sciences* 9.129 (2015): 6415-6426.
- [7] Priyadarshini, Nandita, and Mitrabinda Ray. "A Review: Data Mining Techniques in Education Academia." DIAS/EDUDM@ ISEC. 2017.
- [8] Rambola, Radha Krishna, Mrunmayee Inamke, and Shloka Harne. "Literature Review-Techniques and Algorithms Used for Various Applications of Educational Data Mining (EDM)." 2018 4th International Conference on Computing Communication and Automation (ICCCA). IEEE, 2018.
- [9] Sarala, V., and J. Krishnaiah. "Empirical study of data mining techniques in education system." *International Journal of Advances in Computer Science and Technology (IJACST)* 4.1 (2015): 15-21.
- [10] Hussain, Sadiq, et al. "Educational data mining on performance of under graduate students of dibugarh university using r." *International Journal of Computer Applications* 114.11 (2015): 10-16.
- [11] Sethi, Sidhant, Dheeraj Malhotra, and Neha Verma. "Data mining: current applications & trends." *International*

- Journal of Innovations in Engineering and Technology 6.4 (2016): 586-589.
- [12] Kleftodimos, Alexandros, and Georgios Evangelidis. "An overview of web mining in education." Proceedings of the 17th Panhellenic Conference on Informatics. 2013.
- [13] Nagarathna, Ms B., and Dr M. Moorthi. "A study on web content mining and web structure mining." International Journal of Modern Trends in Engineering and Research (IJMTER) 2.05 (2015).
- [14] Chopra, Preeti, and Md Ataulah. "A survey on improving the efficiency of different web structure mining algorithms." International Journal of Engineering and Advanced Technology (IJEAT) 2.2249 (2013): 8958.
- [15] Johnson, Faustina, and Santosh Kumar Gupta. "Web content mining techniques: a survey." International Journal of Computer Applications 47.11 (2012).
- [16] Mishra, Tripti, Dharminder Kumar, and Sangeeta Gupta. "Mining students' data for performance prediction." Proceedings of international conference on advanced computing & communication technologies. 2014.
- [17] Borkar, Suchita, and Kannan Rajeswari. "Attributes selection for predicting students' academic performance using education data mining and artificial neural network." International Journal of Computer Applications 86.10 (2014).
- [18] Al-Twijri, Mohammed I., and Amin Y. Noaman. "A new data mining model adopted for higher institutions." Procedia Computer Science 65 (2015): 836-844.
- [19] Khedr, Ayman E., and Ahmed I. El Seddawy. "A proposed data mining framework for higher education system." International Journal of Computer Applications 113.7 (2015).
- [20] Parmar, Krina, Dineshkumar Vaghela, and Priyanka Sharma. "Performance prediction of students using distributed data mining." 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). IEEE, 2015.
- [21] Feng, Shaorong. "Application of Data Mining in University Teaching and Management." 2016 6th International Conference on Management, Education, Information and Control (MEICI 2016). Atlantis Press, 2016.
- [22] Ahmed, Ahmed Mohamed, Ahmet Rizaner, and Ali Hakan Ulusoy. "Using data mining to predict instructor performance." Procedia Computer Science 102 (2016): 137-142.
- [23] JIANG, Xiaobing. Teaching Quality Monitoring of Higher Education Based on Data Mining. In: 2018 International Conference on Sports, Arts, Education and Management Engineering (SAEME 2018). Atlantis Press, 2018.
- [24] Dong, Jing. "Analysis and Designing of Educational Big Data Mining System for Higher Education." 2018 8th International Conference on Social science and Education Research (SSER 2018). Atlantis Press, 2018.
- [25] Kumbhar, V. S., and K. S. Oza. "Educational web mining using weka." International Journal of Engineering Trends and Technology (IJETT) 26.3 (2015).
- [26] Dragoş, Sanda, Diana Haliţă, and Christian Săcărea. "Behavioral pattern mining in web based educational systems." 2015 23rd International Conference on Software, Telecommunications and Computer Networks (SoftCOM). IEEE, 2015.
- [27] Victor, S. P., and Mr M. Xavier Rex. "Analytical implementation of web structure mining using data analysis in educational domain." International Journal of Applied Engineering Research 11.4 (2016): 2552-2556.
- [28] Singh, Satya Prakash. "Analysis of web site using web log expert tool based on web data mining." 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). IEEE, 2017.
- [29] Sellamy, K., et al. "Web mining techniques and applications: Literature review and a proposal approach to improve performance of employment for young graduate in Morocco." 2018 International Conference on Intelligent Systems and Computer Vision (ISCV). IEEE, 2018.
- [30] Tang, Ruiming, et al. "Research on Quality Modeling of Information Education in Colleges and Universities Based on Web Data Mining." 2019 International Conference on Smart Grid and Electrical Automation (ICSGEA). IEEE, 2019.