Grenze International Journal of Engineering and Technology, June Issue

Comparative Study on Educational Data Mining and Learning Analytics for Improving the Teaching-Learning Environment

Rushikesh Shete[1], Kavita Kathane[2] and Palash Gourshettiwar[3]
[1-2]YCCE, Department of Computer Science and Engineering, Nagpur, Maharashtra, India
Email: sheterushikesh7@gmail.com, kolarkarkavita@gmail.com
[3]FEAT, DMIHER(DU),Dept of Computer Science and Medical Engg, Sawangi (M), Wardha, Maharashtra, India
Email: palashg.feat@dmiher.edu.in

***Abstract*— Education is evolving a fast. After Covid-19 many different ways are introduced to students to learn and grow. Digital Platforms are booming in this era of self-learning. Educational data mining (EDM) gives the knowledge about the growth and understanding of students. It is important to know the outcome of the learning, irrespective of the platform. This paper gives a general review of what is known about educational data mining and how it might be used in teaching and learning. In this work we review different studies about EDM and Learning Analytics for betterment of learning environments and to enhance teaching techniques.**

***Index Terms*— Education, Data Mining, Clustering, Big Data.**

I. INTRODUCTION

Educational Data Mining (EDM) can have the great impact on education, for now work in this area is boomed[1]. Researches knows the dimensions of how organizations can use data mining (DM) for escalating student learning or related educational techniques like clustering and predictions have had considerable advantages to education and the use of Big Data (BD)[2].

EDM and learning analytics focuses on facts and in training, learning and proposing needs and understanding the investigation on student's results and efforts to make new strategies to apply on teaching and learning[3]. The question arises why data mining? Because, DM goes to granule of data and gives fine results. Which can be helpful to understand student's behavior, knowledge, their grasping? Various data mining tools can be used for extracting meaningful information from the data. And this data can be further used for analytics. Learning analytics helps organization to look after different parameters for improving system[4]. EDM and Learning analytics will help teachers to reconize the problems students are facing in understanding the subject and will help them further to improve teaching methodologies, to introduce novel approaches in teaching.

In the field of education teaching and getting results from the exam is not the only thing now. In recent time many more changes are observed[5]. Now it is not only about the chalk and board way, it's more than that. Traditional way is effective but it is not the only way for teaching, now it is also observed how students are learning whatever they are being taught and what outcome they are getting. It is now important that what student is gaining out of the subject and how it is going to help students in their future assignments. EDM will help institution to study the

Grenze ID: 01.GIJET.10.2.117

results, to study the teaching ways, to understand the impact of teaching methods on student's learning[6]. EDM and learning analytics will help to study the whether the methods used are effective or not, what kind of changes can be done in teaching method, and how it will benefit students.

A) How to Use Data?

The enormous size of statistics, facts, and figures etc. are being used by the educational data mining class to validate study discovery. All together more information, forecasts about student knowledge, dropout rates, and motivational mood become considerably accurate[7]. Large data sets allowed us to better understand particular student groupings, which improves adaptively and personalization for each person.

B) What Type of Data Should be Collected?

More and more information on education is being available. More and more of it is being created as a natural byproduct of the usage of educational technology. Some of it is produced as a result of instructors' efforts to record grades, while other parts come from school administrative systems. [7]

C) Data Mining

Data mining is a process of extracting information from the large data set. Extracted information can be used for discovering new patterns, to understand the data. There are two types of Data Mining Predictive and Descriptive Data Mining. The information gathered after data mining is used for the purpose of analysis. This information can be used by officers, managers, administrators or any person for making business decisions[3]. Data mining gives the processed information in particular format. This processed data is used for performing analysis.

D) Learning Analytics

In order to understand and enhance learning outcomes, learning analytics generally refers to the gathering and analysis of data on students and their settings.
Governments, educational institutions, testing agencies, and companies offering massive open online courses are gathering information about students and their learning processes[3]. However, until the relatively recent discovery of the techniques and instruments to do so, most of that data remained untouched.
Learning analytics can lead to the emergence of fresh approaches aimed at enhancing teaching quality and learning outcomes. Additionally, it can equip students with valuable insights to guide their educational choices effectively.

II. Literature Survey

Increasing volume of data describes Big Data. In every teaching organization or in institution or in any program student's data is growing and growing. This data can be used to understand the student's progress and their falls to. Authors tried to enhance the process of EDM using clustering[8]. During research authors collected data of 84 under-graduate students as they need to work on improving student performance and unambiguousness of obtained model. Authors grouped student's data as per their final marks they secured in their final exams by using a data mining technique, clustering.
Authors have tried to improve the process of EDM. They utilized all data logs to exhibit a student's conduct model. This method first used the clustering technique, to get the end result goal to collecting understudies with comparative mark or qualities[9]. Afterwards they applied data mining to discover more accurate outcome of understudy conduct. After getting results authors got clarity of specific model is producing better and unambiguous results than the general model. Authors used two divers clustering approaches, first manual, in which they used student's marks and second is Automatic, by calculating student's collaboration and the course to gather students utilizing a clustering. The data sets used were collected from Moodle 2.0. 84 undergraduate perusing Psychology degrees were given a task for investigating their performance. Students were informed to take an interest in an e-learning program about "Figuring out how to learn". Program was divided into 11 units, which were assigned to students every week; they had to work on each unit within 15 days span in educating hours. Every unit has three different levels; declarative learning level, Procedural learning level and Conditional level[10].
To collect data regarding courses diverse variables were defined. While performing experiments on data, authors decided to use complete data of all 84 students also they divided data in two data sets using EM clustering algorithm[11]. Further they performed different experiments on data sets considering divers variables. Clusters were formed on the basis on number of students who passed and failed. Their work lead them to some outcomes about teaching learning analytics like proper schedule of classes, improvement in teaching skills, enhancement in learning process, some improvements in infrastructure at institution level. In this work they put forward to use

clustering to escalate instructive process mining and streamline both execution and size of the model got. The transversal of essential learning gives clarity to the model.

Authors have provided an overview about EDM and its application in teaching and learning[12]. They focus on important aspects of Big Data, including technologies and the recent applications of Big Data in education. At last, they examined and described the difficulties and problems experienced in EDM and give a converse of potential upcoming paths for the research and education circle. EDM offers both scholars and practitioners a lot of exciting opportunities. This field helps higher education institutions find practical solutions for enhancing institutional efficiency and student learning.

According to the authors learning analytics tools can aid consortium intensify decision making and ameliorate the caliber and merit of the students' learning experience. It can assist supervisor in allocating resources with the most impact when faced with obstacles like budget constraints and international competitiveness. As authors talk through this paper, current implementations of Big Data technologies in education have been valuable. By identifying hidden trends in huge databases, EDM enables educational institutions to monitor student outcomes and better utilize resources. This would help address issues like attrition, retention and recruitment to organizations for building predictive models [13].

In this article by Arianne and Cleyton [3], extend experience records of an intervention project which addresses the use of Data Mining and Learning Analytics for teaching and learning processes in a public school giving basic education. The project intents to pick out, among eight mathematical points, in which students have more difficulties. Authors developed digital games referred to the subjects, which raised relevant data for creating a data set, in which authors applied Learning Analysis techniques, EDM and Unsupervised Learning strategies together. Authors have used softwares like Jclic and Orange. Jclic used for development and applications of the games used in this paper, it has innate strategy of creating teaching material it is because of this quality this software was chosen. This software helps user to create multimedia pursuits in diverse areas of information. Jclic is a program which bestow to the teaching and learning processes[14].

Orange Data Mining Canvas is used for the Knowledge Discovery in Databases methodology. They used K-means algorithm to find clusters. K-means is mostly used unsupervised algorithm for clustering. The results showed it was viable to identify the contents that the students had trouble through the set of applied methods, reaching the research goal. The results support the potential for gathering, utilizing, and analyzing data in order to create a more individualized and immersive learning experience. They also aim to oversee and handover teachers and administrators of school with assertive feedback in order to help them make wise decisions regarding the educational environment [15].

As there is great advancement in technology and the need for inculcating digital area in education new ways have been magnified by using computers, mobile, tablets, and many more devices which emphasis on the use of learning games in education. Along with the traditional approach using games in teaching method have helped students to understand the subject in a better way also it brings out the interest and motivation in student.

Methodological path used by Arianne and Cleyton was divided in four phases and and different tools were used in each phase for project execution. First phase was planning, in this phase they used Project Feasibility analysis. In the first phase eight different mathematical contents were used to work on. In the subsequent stage, known as Games development, following the formulation of game development strategies and information processing, the Jclic Author tool was employed for the creation of games. Third phase is application, to implement this phase they used tools as JClic Player, JClic reports and Excel. In this phase the developed games were executed and applied with students in computer lab. For this they used JClic Player Software. To collect the data for individual reports JClic Reports were used. Lastly, the information collected using softwares were converted in research dataset using Excel (manually). And in final phase Application the KDD they used JClic Reports and Orange Canvas for data mining and Machine Learning application. From this they obtained visual programming structures and mechanisms for exploratory data analysis. After getting dataset k-means clustering was used to classify and group the data. To check the relationship between the data used for analysis Scatter plot and silhouette plot viewers were used together. Final outcome was satisfying as the eight key points worked very well in the games. This feedback allowed teachers to be more assertive and helped them to build new strategies so that they can overcome the problems encountered in teaching learning[16].

In their work cited as [4], authors have introduced a structured approach to semi-supervised learning. This method involves training a concept-based classifier alongside an explicit semantic analysis (ESA) classifier. The aim is to extract a common set of prerequisite rules from a diverse range of online educational resources. With some excellent achievements attained in particular engineering topics, a working prototype is developed to show the viability of the idea. Our idea also illuminates a variety of promising lines of inquiry for the future. A functioning prototype is developed with some impressive achievements attained in particular technical topics to show its

viability. More importantly, our idea illuminates a variety of promising lines of inquiry for the future. Domain structure discovery is one out of many demanding study areas in EDM that aims to identify the organization of knowledge in an educational area, like forming the precondition conditions amid distinct knowledge components using online educational resources. No matter if it is done manually or semi-automatically, the process of such formulation is quite difficult and time-consuming due to the enormous number of knowledge components in particular areas.

In this article the authors have introduced a systematic and effective framework for employing a semi-supervised learning method[17]. This approach involves co-training a concept-based classifier utilizing the bag-of-words (BOW) technique with an explicit semantic analysis (ESA) classifier. The objective is to generate a shared set of precedence rules by analyzing a diverse range of online educational materials. One of the most difficult researches in EDM, domain structure discovery, is to understand the organization of knowledge in the educational domain, for example, formulating the precondition conditions in the midst of various knowledge components through online educational resources. The huge quantity of knowledge components in particular areas, however, make the process of such formulation frequently exceedingly difficult and time-consuming, whether done manually or partially mechanically. More importantly, our suggestion for a semi-supervised learning approach sheds light on a variety of potential directions, such as the incorporation of more advanced optimizers like evolutionary algorithms to create more optimized learning ways for individualized learning, as well as the academic and other effects of our suggestions on sophisticated e-Learning systems for further investigation[18].

Sagardeep Roy and Dr. Shailendra Narayan Singh, in Reference [5] examine the most recent Big Data techniques and technologies like Like MongoDB, Hadoop, Orange, Weka in the context of educational data mining and learning analytics. In conclusion they dictate that the amount of data in education grows daily, big data technologies and methodologies are more in demand than ever. The use of Big Data technologies and methods has become more prevalent. A total of 20 publications were pulled from Google Scholar and examined for "Educational Data Mining" and "Learning Analytics." It is evident from both themes that the majority of authors are primarily interested in student performance prediction. Predicting student performance has emerged as a brand-new area of study.

Li Chin and Kyong Jin Shim constructed a system utilizing SAS® Enterprise Miner (EM) and SAS® JMP Pro[19]. Their study concentrated on seven modules across four sessionThey examined students' online interactions, including their assessment outcomes and other important learning behaviors such as how often they accessed learning materials, and the eventual outcome of each module. In this study Authors mainly used SAS® Enterprise Miner (EM) and SAS® JMP in the data mining tasks and predictive modeling. To extract, transform and perform the exploratory data analysis task SAS® JMP is used. To build predictive models SAS® Enterprise Miner (EM) is used.

To cross check the predictive analytics results with the results obtained from SAS® Enterprise Miner (EM), they developed predictive model using Python[20]. Two modules were identified to associate with at risk students, first CCC has the highest rate of dropout among all pupils second FFF. The dataset is split into training (80%) and testing (20%) sets for each of the regions. The testing set was used to evaluate and validate the models' prediction power while the training set was used to estimate model parameters. According to findings of the investigation, students' evaluation scores and their online use of learning materials are reliable indicators of how well a module would turn out for them. It was difficult to predict the "Fail" result. In order to accurately differentiate between "fail" and "withdrawn" cases, we intend to examine students' learning footprints and incorporate them into prediction. This research will be expanded to actual academic environments. In order to identify pupils who are at danger, we intend to use data from both online and offline assessments. Important indicators for passing, failing, and withdrawing instances found in this experiment offer educators valuable information for creating effective intervention strategies to keep students on track.

III. Experiment, Result & Discussion

To understand the students learning ability we gave students s task related to their course subject. The task was to Creating and configuring Networks and apply routing algorithm using CISCO Packet Tracer. Total 151 students were assigned this task. They were taught the concept of Network topology, IP addressing, Devices used in Network. Also they were taught the tool CISCO Packet Tracer.

Objective of the practice was students should be able learn how to create types of network topology, Static Routing and connecting networks together using Routers using CISCO packet tracer. Expected Outcome was students should be able to create networks using CISCO packet, and will be able to apply static routing algorithms to the network, and will understand router configurations. In static routing, we have to manually assign the IP address.

Also, we provide network addresses to the router so that a packet can travel to different networks. This all information is feed to the topology manually. Benefit of entering all the details is that network administrator will have knowledge of whole network he/she is handling. Following Figure shows the topology created. Also, it will show different addresses assigned to the end devices and router's interfaces.

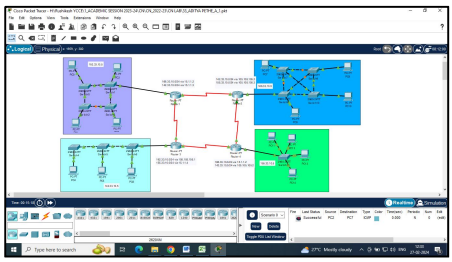

Figure 1. Topology showing four different networks connected using routers and Successful Transmission of Packet from one network to other.

In this method student was provided with the video link to understand the creation of topology of network and how to use it. Students were given a task. Impact of the task was more than 100 students were able to solve the task. Around 20-25 students were not able to perform router configuration. 15-20 students were not able to solve the task at first but after guiding and helping them to solve simple assignments they could perform the task.
After executing the experiment, we conducted the feedback for the activity we got following results. From this feedback we tried to know whether the activity was challenging or not. 86% students found that the activity was challenging. From this activity we could know that 95% students understand the concept of network topology, as they had to create four different topologies. 97% students agreed that they understood the concept of routing configuration using CISCO Packet Tracer and overall 95% of students were able to solve the activity and understand the objective of the activity or task.

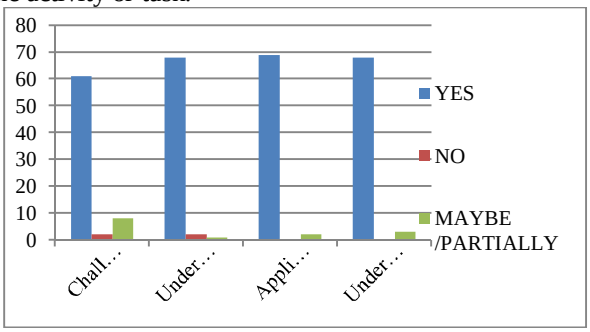

Figure 3. Results of Feedback taken from the students

IV. Conclusions

This paper presents various research work carried out on the use and implementation of Educational Data Mining for getting better results by learning analytics by various research scholars. In this we have reviewed different methods implemented by authors. Clustering, a data mining technique is mostly used on data set obtained from educational data. Datasets with various parameters can be formed to perform educational data mining. Motivation for this review was to understand EDM and Learning Analytics so that it will help us to build a better learning environment for students, also to recognize the need of new tools and techniques for teaching and learning to meet students need and to get good results.

References

[1] S. MacNeill, L. M. Campbell, and M. Hawksey, "Analytics for education," in Reusing Open Resources: Routledge, 2014, pp. 154-165.
[2] C. Romero, S. J. W. I. R. D. M. Ventura, and K. Discovery, "Educational data science in massive open online courses," vol. 7, no. 1, p. e1187, 2017.
[3] M. Bienkowski, M. Feng, and B. J. O. o. E. T. Means, US Department of Education, "Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief," 2012.

[4] Y. B. Rajabalee, M. I. Santally, and F. Rennie, "The use of learning analytics to improve online learning outcomes: A systematic literature review," 2019.
[5] A. N. Applebee and J. A. J. E. J. Langer, "EJ Extra: A Snapshot of Writing Instruction in Middle Schools and High Schools [FREE ACCESS]," vol. 100, no. 6, pp. 14-27, 2011.
[6] H. Aldowah, H. Al-Samarraie, W. M. J. T. Fauzy, and Informatics, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," vol. 37, pp. 13-49, 2019.
[7] Y. Qian, C. X. Li, X. G. Zou, X. B. Feng, M. H. Xiao, and Y. Q. J. C. A. i. E. E. Ding, "Research on predicting learning achievement in a flipped classroom based on MOOCs by big data analysis," vol. 30, no. 1, pp. 222-234, 2022.
[8] C. Romero, S. J. W. I. R. D. m. Ventura, and k. discovery, "Data mining in education," vol. 3, no. 1, pp. 12-27, 2013.
[9] M. Bibi et al., "A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for twitter sentiment analysis," vol. 158, pp. 80-86, 2022.
[10] K. Morgan-Short, M. Faretta-Stutenberg, K. A. Brill-Schuetz, H. Carpenter, P. C. J. B. L. Wong, and cognition, "Declarative and procedural memory as individual differences in second language acquisition," vol. 17, no. 1, pp. 56-72, 2014.
[11] A. Dutt, M. A. Ismail, and T. J. I. A. Herawan, "A systematic review on educational data mining," vol. 5, pp. 15991-16005, 2017.
[12] M. W. Rodrigues, S. Isotani, L. E. J. T. Zarate, and Informatics, "Educational Data Mining: A review of evaluation process in the e-learning," vol. 35, no. 6, pp. 1701-1717, 2018.
[13] K. Alshehhi, S. B. Zawbaa, A. A. Abonamah, and M. U. J. A. o. E. J. Tariq, "Employee retention prediction in corporate organizations using machine learning methods," vol. 27, pp. 1-23, 2021.
[14] A. S. Torcate and C. M. de Oliveira Rodrigues, "Educational Data Mining with Learning Analytics and Unsupervised Algorithms: Analysis and Diagnosis in Basic Education," in 2021 XVI Latin American Conference on Learning Technologies (LACLO), 2021, pp. 67-74: IEEE.
[15] G. Casey, T. J. I. R. o. R. i. O. Evans, and D. Learning, "Designing for learning: Online social networks as a classroom environment," vol. 12, no. 7, pp. 1-26, 2011.
[16] J. J. I. J. o. e. Fu and D. u. ICT, "Complexity of ICT in education: A critical literature review and its implications," vol. 9, no. 1, pp. 112-125, 2013.
[17] Y. Ouali, C. Hudelot, and M. J. a. p. a. Tami, "An overview of deep semi-supervised learning," 2020.
[18] K. Colchester, H. Hagras, D. Alghazzawi, G. J. J. o. A. I. Aldabbagh, and S. C. Research, "A survey of artificial intelligence techniques employed for adaptive educational systems within e-learning platforms," vol. 7, no. 1, pp. 47-64, 2017.
[19] L. C. Ho and K. J. Shim, "Data mining approach to the identification of at-risk students," in 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 5333-5335: IEEE.
[20] A. Fortino, "Data mining and predictive analytics for business decisions: a case study approach," 2023.