

# AN EFFICIENT SECURE KEY TECHNOLOGIES AND APPLICATIONS IN MASSIVE OPEN ONLINE COURSE LEARNING PLATFORM FOR EDUCATIONAL BIG DATA ANALYTICS

<sup>1\*</sup>M SANDHYA VANI, P H SWARNA REKHA<sup>2</sup>

<sup>12</sup>ASSISTANT PROFESSOR, SRI INDU INSTITUTE OF ENGINEERING AND TECHNOLOGY, SHERIGUDA,

TELANGANA, INDIA

**Abstract**— with the quick advancement of the Internet and correspondence innovation, online education has drawn increasingly consideration, online learning platforms, then again, store massive student social data and educational data. Step by step instructions to adequately analyze and use the data to enhance the nature of online education has turned into a key issue critically should have been explained in the field of big data in education (BDE), educational data mining (EDM) is actually a viable and viable technique and means of applying BDE. Therefore, EDM is an important scholastic research hotspot in the field of EDM. Right off the bat, the paper introduces the essential ideas of BDE, EDM and online learning platform, and then explains on the procedure of how educational data mining transforms crude data into information. Finally, the key technologies of data mining are characterized according to their utilizations, and give its application in online education scene. The paper can give some guidance to the examination and use of educational data mining dependent on online education.

**Keywords-** component; educational data mining; online education; big data in education; online learning platforms

## I. INTRODUCTION

With the dangerous advancement of big data, big data has begun a noteworthy transformation of the occasions, and extraordinary changes, even substantial changes and improvements in financial aspects, governmental issues, human science and many science classes have occurred [1]. Big data likewise profoundly affects the education industry. Online education involves a critical position in the field of education, and produces massive online and detectable data consistently [2]. Therefore, the combination of big data and online education has turned into an inevitable necessity for the advancement of the occasions.

At present, the most troublesome piece of the customary education mode lies in the accumulation and analysis of data. The online education platform stores huge measures of data that are constant, screen capable and quantifiable, including student conduct data and the educational data of the platform [3]. Be that as it may, the data itself is just data, did not intuitively demonstrate any profitable learning, how to utilize these educational data to transform these data into information and learning and give administrations to educational basic leadership, teaching streamlining and scholarly enhancement is the estimation of BDE. The challenges of processing innovation of big data mainly center around the unpredictability of data, and

EDM is an important specialized means to understand the utilization of BDE [4]. Through the analysis and mining of the relevant data in the online learning platforms, the connection between students' practices and their inner intelligent connections are found, and the teaching plan is better formulated dependent on it.

## II. BDE, EDM AND ONLINE LEARNING PLATFORMS

### A. The Concept and Basic Connotation of BDE

Big data refers to a large scale of data which is rich in information in different forms of media, network and other media [5]. Big data on education (BDE) is a subset of big data, which refers to data in education. In fact, big data is a nebulous concept that has not yet formed an accepted definition. Even so, there is a difference between big data and past data, and its basic connotation can be summed up in 4V [6] Volume Variety Value and Velocity of which specific meanings are shown in table I.

Information Technology is a key enabler of many of today's technological innovations that have led to improvements that benefit society in many areas including education, business, security and health care. Technological advancements in digital technologies have led to an increase in the number of people accessing and using digital devices worldwide. [1] States that digital connectivity in Africa is driven by greater ownership of mobile devices, and is the driving force for the technological revolution on the African continent. Technological innovations, increased affordability and use of digital devices have led to a situation where extremely large amounts of data are being generated through use of these devices worldwide. Big Data Analytics can therefore help institutions of higher learning make use of data that is generated in their ecosystems. A report by [2], indicates that smart phone ownership rates in emerging and developing nations are rising at an extraordinary rate, climbing from a median of 21% in 2013 to 37% in 2015; and that overwhelming majorities in many countries surveyed reported owning some form of mobile device, even if they are not considered smart phones. A good example is Kenya, where smart phones have become increasingly common, while food is unfortunately scarce [3]. This trend is expected to continue, with data being rapidly generated and collected at a tremendous rate and will therefore contribute to towards big data. This pool of data when harnessed has the potential of helping to solve the statistical tragedy that has characterized many developing countries, and crises that are related to unreliable data, particularly in Africa. Big Data refers to data sets that are large enough to pose significant challenges when using commonly available or traditional tools and infrastructures to collect, manage, process and analyze the data within a tolerable amount of time. [4] Points out that Big Data are concerned with emergent suite of technologies that can process mass volumes of data of various types at faster speeds than ever before. [5, 6] observe that the attributes of big data are best described as high-volume, high-velocity and high-variety information assets which by their nature demand cost-effective, innovative forms of information processing that can enable enhanced insight and decision making. Big Data therefore changes the way we approach data analysis, inspiring entirely new families of information services and necessitating new processing models and knowledge representations. Big Data consists of extensive datasets; primarily in the characteristics of volume, variety, velocity, and/or variability that require a scalable architecture for efficient storage, manipulation, and analysis [7]. [8] Defines big data as distribution of data systems across horizontally-coupled independent resources to achieve the scalability needed for the efficient processing of extensive datasets. [9] Describes Big Data as data that is too big in terms of volume, and moves too fast in terms of its acquisition and analysis speed that normally exceeds the processing capacity of conventional database management systems. Big Data is identified by a number of fundamental characteristics which include the following 1) Volume: which refers to colossal amounts of data collected that present a challenge to conventional information technology structures to store, process analyze and present. 2) Velocity: refers to the rapid pace at which information is generated and flows within an organization. The velocity dimension is also concerned with the capability of an organization understanding and responding to events as they occur [10]. 3) Variety: which refers to data in diverse formats both structured and unstructured. This variety of data formats come from interaction between people and the Information Technology systems such as sensors and web applications. 4) Veracity: which refers to the quality and authenticity of the data, and looks at how data that is being stored is meaningfully mined to address the problem being analyzed 5) Value: refers to the relevance of the data to Big Data Analytics and extent to which the data is utilized to generate value of the insights, benefits and business processes within an organization [4, 8, 11, 12, 13]. Big data comprises large and complex data sets which are difficult to process using traditional data processing applications and common database management tools.

However, big data is in a nascent stage, for educators, the concept of big data and its applications are vague [5]. First, the paper elaborates the related concepts of BDE, EDM and online education platform, so that educators have a deeper understanding of the relationship and development of those concepts. Based on this, this paper mainly includes two aspects: educational data mining process and key mining technology classification: the educational data mining process mainly explains how to extract raw data from the database to provide value for education decision-makers and learners key mining technology classification is to analyze the technologies used in educational data mining process, the paper not only categorizes these technologies but also attempts to explain the differences and connections between them, namely, when different technologies should be applied to a scenario. Finally, we put forward some prospects of EDM in online education

Type	Interpretation	Example
Volume	Huge Amounts of data	The data volume is generally at the TB level, or even PB level
Variety	Many types of data	Data sources can be data such as documents, web logs, video and audio, can also be generated from the Internet, Shared cloud, Internet of things, etc.
Value	Low density	The availability of useful data on decisions may be one thousandth or less of a huge Amount of data.
Velocity	Quick input and output	Feature data is growing fast, processing speed is fast, and the time efficiency is high which is the distinguishing feature of big Data difference from traditional data mining.

### ***B. The Concept and Purpose of EDM***

Data mining is a procedure of extracting the information and learning from extensive, incomplete, fluffy and random data in which individuals are not known in advance, but rather conceivably valuable information and learning [7]. As of late, with the advancement of educational informationization and the ascent of new online education display spoken to by MOOC, an ever increasing number of analysts are worried about the EDM [8]. Educational data mining

EDM is really the utilization of data mining innovation in education and the question of analysis and processing is to BDE, its motivation is to find out and take care of the issues in education by using the key technologies of data mining to mine big data in the online learning platform, to obtain the important and significant information in the teaching procedure and to enhance the nature of teaching and learning.

### ***C. Online Learning Platforms and Types***

Online education is a web - based learning conduct. Through relevant online learning platforms, students accomplish auspicious, asset sharing and zero-distance learning knowledge and approach [9]. As of now, the education platforms can be partitioned into three classifications. The first is the customary online education platform; it is mainly appropriated in the form of video and sound through TV or the Internet, and is generally utilized in remote regions and zones with poor educational quality. The second is another sort of online education platform spoken to by massive open online course MOOC [10], the esteem introduction of MOOC's initiative is that "anyone can learn high caliber educational assets whenever and anywhere". In view of EDM, MOOC has broadened learning apparatuses and rich educational modules assets, integrates mechanical research and application innovation, and utilizations efficient accuracy tests and widely associated interpersonal organizations to give students learning knowledge anytime, anywhere. The third is little private open course (SPOC), which compares to MOOC. SPOC isn't an option to moocs, yet rather the inheritance, flawlessness and transcendence of moocs [11]. It assimilates and inherits advanced thoughts of MOOC, for example, top notch assets, precise testing, opportune criticism and customized proposal, and so on. Because of the high dropout rate of MOOC and the absence of vivid learning knowledge, SPOC has changed the conditions for the use of the course and increased the cooperation rate of the students and instructors. "Private" in SPOC empowers students to have a feeling of possession and a feeling of criticalness. "Small" in SPOC enables the student's cooperation in learning to enhance extraordinarily.

### III. DATA MINING ON EDUCATION AND ITS PROCESS

Figure 1 indicates where EDM transforms data into learning during data analysis and processing. From the point of view of summed up data mining, the processing stream includes 3 phases: data readiness, data mining and analysis and assessment. From an educational point of view, this is a cycle of discovering information from data created by the educational condition and reusing it to enhance the educational condition. From the point of perspective of data positioning, the data of each stage may transform the data of another phase of different targets.

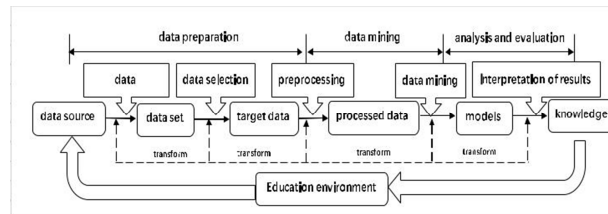


Figure 1.Process of EDM.

#### A. Preparation of Educational Data

The establishment of EDM is that data, mass, extravagance and nature of educational data are vital to the result of EDM. Educational data arrangement for the most part includes data integration, data choice and preprocessing. Online education data exist in the form of organized data, (for example, understudy database, and so on.), unstructured data, (for example, teaching recordings, English sound and understudy pictures, and so forth.) and semi-organized data, (for example, warning mail, Java works and teaching asset store, and so forth.). Be that as it may, not all data sources can complete data mining, yet selecting the corresponding data dependent on data mining purposes as the data wellspring of this educational data mining, which is the procedure of data determination. The gathered data frequently contain some important information missing, incorrect or containing commotion, inconsistency and different issues, data preprocessing innovation can enhance data quality to meet the necessities of educational data mining, normal preprocessing activities can utilize data cleansing (filling in missing qualities, removing exceptions, and so on.), data decrease (reducing high-dimensional data to low - dimensional, and so forth.) and data transformations (transforming data into another form of articulation, and so on.). Data arrangement is the reason for follow-up work and specifically determines its quality and effectiveness.

#### B. Educational Data Mining

EDM utilizes a portion of the key technologies of data mining to enhance the nature of online learning by modeling BDE and finding out the connection between's student's scholarly performance and learning conduct, teaching reason and teaching procedure [12]. In request to accomplish the objective, the basic key technologies of education data mining are commonly isolated into ASSOCIATION principle, relapse and forecast, order, clustering and analysis. For instance, Apriori is a commonplace calculation for ASSOCIATION rule. Using Apriori calculation, it is discovered that there is a connection between the performance of the understudy and their involvement in doing task, internal evaluation tests, attendance and so forth., which would recognize the normal and underneath normal understudies and to enhance their performance to give great results[13]. Coordinations expectation technique is a typical calculation for anticipate mining, educators can build up accomplishment forecast display by run of the mill student conduct indexes, through the expectation show instructors can understand the learning circumstance of the students in advance, give students exact guidance and intervention, let students come back to the ordinary learning track[14]. The over two precedents demonstrate that distinctive data mining calculations ought to be utilized in various teaching situations. Therefore, according to the sorts of online education data and the teaching issues to be illuminated the reasonable data mining calculations are chosen. And deciding how to utilize the calculation on these data increases the proficiency of research as well as enhance the estimation of learning.

### C.Expression and Interpretation of the Results

The articulation and interpretation of results allude to the analysis and synopsis of the aftereffects of the uncovered data, and set forward significant standards and reestablish them to the meaning individuals can understand. The articulation and interpretation of results is the most generally utilized representation innovation. In spite of the fact that there is some understanding of the data subsequent to mining and insights of BDE, it not intuitive. The outcomes pictured by the form will demonstrate the qualities of data, with the goal that instructors can plainly understand the consequences of digging out and settle on an exact teaching choice, for example, students' precision rate, time of study, learning bend and so on. The outcomes obtained at this stage are important information, for students, the outcomes can be set up learning information model to furnish students with point by point learning criticism and exhortation; for the education, the relationship can be summed up online teaching procedures and learning results, and then enhance their teaching strategies.

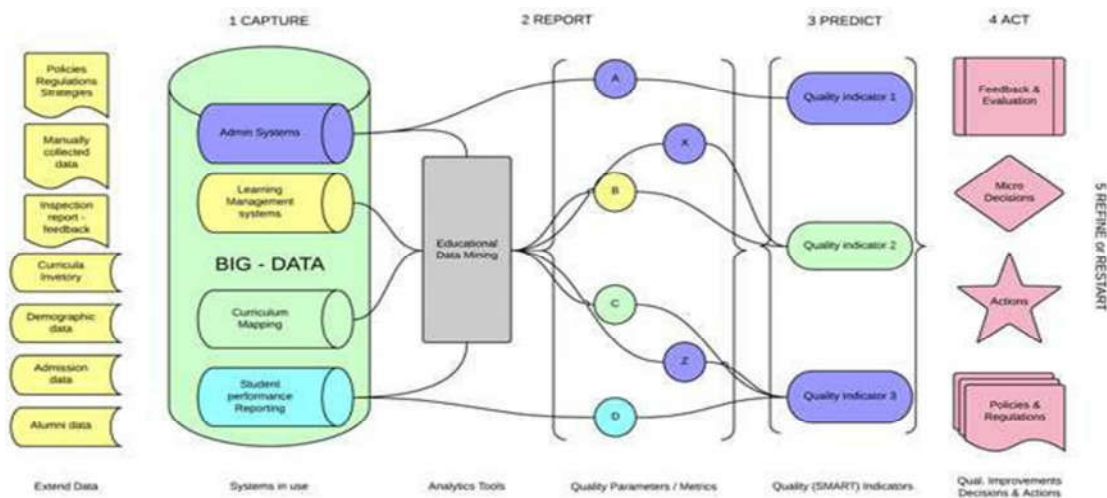
Big Data presents to Institutions a decent structure for efficiently utilizing the huge swath of data in shaping the eventual fate of advanced education [4]. [4] Citing [15] states that Big Data application in advanced education is ascribed to mechanical innovations and improvements; which have catalyzed the development of analytics in advanced education. According to [16], Data Warehouses and Cloud Computing combined with more prominent responsibility for gadgets by end clients in the educational biological system are making it conceivable to gather, manage and maintain massive measures of data. These Information Technologies are important assets that when abused by approach producers are helpful in driving institutional procedure and arrangement making for what's to come. Information Technology makes accessible modern platforms that give computing power important to plowing through massive measures of data, and turning these masses into meaningful information. Data mining technologies when connected utilize engaging and inferential measurements to get designs from the massive measures of gathered data for noteworthy information[17].An representation of Big Data Environment is exhibited by figure 2 beneath. Big Data Analytics is relevant in addressing a significant number of pressing issues for education frameworks [14] , key among them are 1) increasing teacher adequacy; 2) harnessing insights from learning encounters; 3) delivering education for all that may likewise be custom fitted for individual students needs; and 4) equipping understudies with relevant aptitudes for their future

Professions. Institutions of higher learning are looked with strain to enhance nature of learning results while cutting expenses and to manufacture new esteem associations with understudies. This is feasible when these institutions have the ability of leveraging the data gathered during the enlistment stages, finance sourcing and instructional procedure to help more noteworthy learning results. Big Data technologies have been altered by innovations in database technologies, PC equipment, particularly memory and capacity limits, and increased accessibility of system bandwidth. According to [19], The Hadoop biological community which includes Pig, Hive, Mahout, and R-Hadoop is a genuine case of the Big Data technologies that can be sent in educational frameworks. Hadoop and NoSQL databases are fit for enabling analysis of vast heterogeneous datasets at exceptional paces.

Explain that Hadoop is an open-source structure for conveyed computing that empowers processing of extensive datasets through level versatility. The Apache Hadoop venture for instance creates open-source programming for solid, versatile and dispersed computing. Actually, organizations, for example, Yahoo, Google, and Face book have invested in the Apache Hadoop Project with a perspective of addressing their Big Data needs. The Apache Hadoop programming library system permits disseminated processing of expansive data sets crosswise over bunches of PCs through basic programming models, and is intended to scale up from single machines to different machines each offering neighborhood calculation and capacity [13, 19].The Education division is these days becoming more innovation arranged. The advanced insurgency in developing nations is leading to improvement of new technologies, for example, pervasive computing gadgets and the Massive Open Online Courses which are all drastically transforming the mode and availability to teaching and learning [4]. These Massive open online courses (MOOC) are generating colossal measures of data that are relevant for Big Data Analytics [21]. [22] Underscores this by stating that the period of cloud and versatile computing is opening up many open doors for revolutionizing education. Massive open online courses (MOOC) are another wonder in advanced education in which free online courses, open to a boundless number of understudies are offered by Institutions of higher learning. MOOC portrays courses offered completely online where students can obtain endless supply of the courses they enlisted in, and offers a change in outlook in the conveyance of learning. The virtual classroom is another important apparatus on which distance learning programs depend on, and on which instructors can share learning material through the webinar. These instruments, together with other learning management frameworks produce a great deal of data which can be put away and analyzed for use in forecasting



[21]. In such cases Big Data Analytics is helpful in enabling institutions using such technologies to get an incentive from these wellsprings of data. Enhancements in Internet openness mean that big data analytics can be connected to help diminish the expenses of education and enhance understudy performances by examining and offering altered and self-managed learning answers for students [23]. [24] Argue that advanced education institutions that attention on better instructional quality and the procurement and maintenance of understudies and workforce generally have utilized data stockrooms and business intelligence apparatuses to give an account of and analyze data. Big Data analytics can possibly emphatically affect all the real zones that are of importance for an institution of learning; in regions, for example, understudy enlistment and maintenance, integrated information management and reporting, operational cost management, administrative compliance and research. According to report by [18], through analytics learning institutions can perform exhaustive analyses of understudy and learning data to settle on informed choices on future course offerings in request to provide food for the requirements of potential and existing understudies



**Figure 2:** Analytics Driven Conceptual Framework: Adapted from

#### IV. THE KEY TECHNOLOGIES OF DME

DME can assemble models through analyzing and mining education-related data. These models can be understudy performance forecast models, versatile educational modules suggestion models, and so on. In the field of education, data mining technologies are commonly isolated into five classes according to the innovation data types and issues to be unraveled, to be specific, ASSOCIATION guideline, relapse and expectation, grouping, clustering and analysis. This article will at that point quickly introduce the 5 technologies and give their application situations in online education.

##### A. ASSOCIATION Rule

The objective of ASSOCIATION rules is to find the ASSOCIATION or connection between's the data set, and the most acclaimed of them are Apriori calculation and the FP-Growth calculation. ASSOCIATION rule mining is generally connected in online platform. It utilizes ASSOCIATION principles to uncover the internal standards of teaching, understudy advancement mode and discipline connection rules, to all the more likely organize teaching exercises.

##### B. Regression and Prediction

The relapse strategy is a numerical technique to find the relationship between's reliant factors and independent factors. Relapse analysis is partitioned into linear relapse, numerous relapses and nonlinear relapse. In linear relapse, the data is displayed using a straight line, multivariate relapse is an augmentation of the linear relapse,

involving numerous factors, nonlinear relapse is a nonlinear model with adding polynomial form in essential linear model. Calculated relapse calculation is a generally utilized relapse and expectation calculation. For instance, while analyzing the connection between student's practices indexed and learning finishing rate on the MOOC platform, the relationship coefficient of the indexes can be determined by using the calculated relapse calculation to obtain the learning expectation show.

### **C.Classification**

Order is an administered learning process. Its motivation is to develop a grouping capacity or arrangement demonstrate dependent on the attributes of the data set, which can delineate sorts of tests into a given classification. There are many grouping techniques, the most widely recognized are KNN, neural systems, bolster vector machines and choice trees. For instance, the online learning platforms have an information suggestion framework. At the point when students look for a certain course, the framework will consequently suggest comparable courses with the goal that students have more customized decisions. The labeling of these information points is a run of the mill use of characterization.

### **D.Clustering**

The way toward dividing an accumulation of physical or theoretical articles into bunches with comparative items is called clustering. Each question is exceptionally like a protest in a similar bunch and less like protests in different groups. The contrast among clustering and order is that clustering isn't assembled according to explicit criteria yet inherent rationale between data. Nobody knows whether the data will be isolated into a few gatherings or what kind of gathering. The normal clustering strategies include K-means, Hierarchical Clustering, and Gaussian blended clustering and so on. For instance, there are contrasts in the dimension of many students in online learning process. Instructors utilize clustering techniques to bunch understudies who have comparable learning foundations, providing a more dependable quantitative reason for understudies' general dimension analysis and focused on guidance.

### **E.Diagnosis**

In data mining, the objective of the symptomatic technique is to find few data questions that are irregular in the data set, known as exception. Exceptions are generally viewed as unusual or boisterous data, however some of the time these data objects contain critical information. Symptomatic strategies can be founded on factual techniques, strategies dependent on distance or vicinity, and strategies dependent on thickness. In SPOC, after the determination analysis technique, instructors find that a student is partitioned into low class, however its scientific index is greatly high, or one's performance has been steady for a certain timeframe, yet at a point of time falls strongly these anomalies may stow away important information, educators can propelled customized teaching exercises according to the explicit circumstance of individual students.

### **CONCLUSION**

Big data is of progressive significance to online education. With such colossal and confused data on the online learning platforms, it is especially important to consider and apply these BDE. From the viewpoint of scholarly research and innovation application, the paper expounds the essential ideas of BDE, EDM and online education, the point by point procedure of educational data mining, the grouping of educational data mining technologies, and the need to focus during the time spent mining. For instance, extraordinary calculations are chosen according to various application situations, and instances of explicit situations utilized by some key technologies are given. The reason for this paper is to empower different specialists or instructors to gain a superior understanding of the manners by which educational data mining is utilized in the education of big data and better use educational data mining systems to enhance the nature of online education. The improvement of big data has brought many open doors for online education; the rise of new mining technologies will certainly give better techniques and means to the utilization of EDM in the field of online education. Using these new technologies will likewise prompt more logical analysis and higher-esteem information, and better administration for online education. In any case, care must be taken to address the managerial, moral and

specialized difficulties and constraints of educational data mining, all of which need to additionally reinforce the exploration on the educational data mining of online education.

## REFERENCES

1. Zhang and S. Qin, "A brief analysis of the key technologies and applications of educational data mining on online learning platform," 2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA), Shanghai, 2018, pp. 83-86. Doi: 10.1109/ICBDA.2018.8367655.
2. X. Jin, B. W. Wah, X. Cheng, and Y. Wang, "Significance and Challenges of Big Data Research," *Big Data Research*, vol. 2, no. 2, 59-64, 2015/06/01/ 2015.
3. D. Zhenhua, "The Application of MOOC and Big Data Technology in the University Computer Foundation Teaching," *Computer & Telecommunication*, vol. 6, p. 037, 2014.
4. H. F. O'Neil and R. S. Perez, *Web-based learning: Theory, research, and practice*. Routledge, 2013.
5. S. K. Mohamad and Z. Tasir, "Educational data mining: A review," *Procedia-Social and Behavioral Sciences*, vol. 97, pp. 320-324, 2013.
6. S. John Walker, "Big data: A revolution that will transform how we live, work, and think," ed: Taylor & Francis, 2014.
7. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of "big data" on cloud computing: Review and open research issues," *Information Systems*, vol. 47, pp. 98-115, 2015.
8. R. Baker, "Data mining for education," *International encyclopedia of education*, vol. 7, no. 3, pp. 112-118, 2010.
9. Nawrot and A. Doucet, "Building engagement for MOOC students: introducing support for time management on online learning platforms," in *Proceedings of the 23rd International Conference on World Wide Web*, 2014, pp. 1077-1082: ACM.
10. J. Emanuel, "Online education: MOOCs taken by educated few," *Nature*, vol. 503, no. 7476, pp. 342-342, 2013.
11. Christensen, A. Steinmetz, B. Alcorn, A. Bennett, D. Woods, and E. J. Emanuel, "The MOOC phenomenon: Who takes massive open online courses and why?" 2013.
12. P. Guo, "MOOC and SPOC, Which One is Better?" *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 8, 5961-5967, 2017.
13. A. Dutt, M. A. Ismail, and T. Herawan, "A Systematic Review on Educational Data Mining," *IEEE Access*, 2017.
14. D. M. D. Angeline, "Association rule generation for student performance analysis using apriori algorithm," *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)*, vol. 1, no. 1, pp. 12-16, 2013.
15. W. Zhang, X. Huang, S. Wang, J. Shu, H. Liu, and H. Chen, "Student Performance Prediction via Online Learning Behavior Analytics," in *2017 International Symposium on Educational Technology (ISET)*, 2017, pp. 153-157.