International Journal of Computer Sciences and Engineering
Vol.**11**, Issue **3**, pp.**28-33**, March **2023**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

Research Paper

Revolutionizing Online Education: Integrating Machine Learning and Data Analysis into LMS

Parveen Singh[1], Meenakshi Handa[2], Anwal Ul Haq[3*]

[1,2]Department of Computer Sciences Govt SPMR College of Commerce, Jammu
[3]Department of Management, Govt SPMR College of Commerce, Jammu

Received: 15/Feb/2023; **Accepted**: 20/Mar/2023; **Published**: 31/Mar/2023. **DOI**: https://doi.org/10.26438/ijcse/v11i3.2833

***Abstract*:** In 2020, the events that transpired revealed the fragility of society and its vulnerability to abrupt shifts in governing paradigms. The outbreak of COVID-19 pandemic globally altered the manner in which people engage in activities such as communication, work, study, and interaction. This resulted in a significant change in the way society operates, including education. To accommodate the new reality, education embraced the use of technology, specifically information and communication technologies. One such example is the increased reliance on learning management systems as a platform for resource management and educational activities. This proposal seeks to enhance the learning experience by incorporating artificial intelligence and data analysis into learning management systems. The aim is to establish robust educational models in the new normal, where students have access to virtual assistants for guidance during online learning.

***Keywords*:** analysis of data; artificial intelligence; machine learning; online education

1. Introduction

The COVID-19 pandemic has brought about significant changes in society, affecting various sectors, including health, education, and industry. With the vulnerabilities and weaknesses exposed by the pandemic, universities and research departments have an essential role to play in addressing these challenges and developing new models based on the lessons learned. The integration of Information and Communication Technologies (ICT) has been critical in combating the pandemic and maintaining essential sectors' functionality. The higher education sector has already integrated these technologies into its operations, allowing education to persist amidst the challenges posed by the pandemic. However, to ensure that students can continue learning under any circumstances, it is essential to implement adaptive education models that incorporate advanced technologies and revisit concepts and tools that were previously overlooked [4,5].

Learning Management Systems (LMSs) were once viewed as basic storage platforms, but the current pandemic situation has made them essential for students to stay connected with their schools. The shift to virtual/online education has resulted in decreased motivation and difficulty adapting, affecting student performance. To overcome these challenges, students need a suitable learning environment that offers access to resources and monitoring systems to track their progress and personalized activities based on their individual needs. This study aims to enhance online education by integrating Artificial Intelligence (AI), data analysis, and an LMS. The study is based on an existing online education model from a university in Ecuador, using the LMS as a tool to provide students with access to resources and activities. Integrating AI, such as machine learning, into this model enhances the management of student performance by providing a sense of security and support during their academic pursuits [11].

The AI system collects data from student interactions and data analysis to identify patterns, classify students, and tailor the model to meet their individual needs. The virtual assistant greets students when they access the LMS, providing them with personalized feedback to enhance their learning experience. The system generates alerts that are reviewed and discussed with the teacher to improve the activity or modify the course methodologies, ultimately enhancing the learning experience through the integration of technology and student interaction in an online education model. The study consists of two crucial components: the university infrastructure and the integration of new technologies. The university infrastructure serves as the foundation for data processing and management, providing a functional architecture for integrating technologies such as big data and AI. The integration of new technologies, such as data analysis and AI, must work in harmony and be constantly evaluated to enable the AI system to learn and make more successful decisions with each instance generated in the analysis. The study chooses two online education courses related to ICT as the sample, as students have a sufficient level of familiarity with ICT.

2. Related Works

Several studies have explored the integration of AI and educational data analysis tools in LMS [8], but they do not prioritize enhancing the learning experience. Some papers suggest using data mining algorithms to uncover student weaknesses in a particular course [6], which is then used to address these weaknesses by the relevant individuals in charge of learning quality. Other studies propose more complex models that integrate business intelligence (BI) frameworks [7], enabling the identification of factors contributing to student attrition in virtual or online educational models by combining multiple data sources to provide a more detailed analysis.

The focus of several studies is to enhance the teacher's ability to create better learning models and methodologies using AI in LMS. Specialized AI techniques have been shown to improve user interaction and continuously learn from interactions [8]. This proposed work is distinct from previous studies because it integrates both AI and data analysis in a centralized system, creating a virtual assistant that manages student information and performs personalized monitoring. Unlike other studies that only analyze data within the LMS, integrating multiple data sources through big data technology provides the assistant with a comprehensive understanding of each student's needs and expectations, resulting in more adaptable decision-making by AI [1]. This integration enables swift and efficient evaluations of student performance, ultimately enhancing the learning experience.

Preliminary Concepts

The primary goal of data analysis is to extract insights and knowledge from a dataset by applying statistical and computational methods. This process involves several steps, including data cleaning, processing, modeling, and interpretation. Data analysis has broad applications across industries, including finance, healthcare, and government, among others. It helps organizations to make data-driven decisions, optimize their operations, and gain a competitive advantage. In healthcare, data analysis is used to identify patterns and trends in patient data, improve diagnoses, and develop personalized treatment plans. In finance, it is used to assess risk, detect fraud, and forecast market trends. Similarly, in the education sector, data analysis is used to measure student progress, identify areas of improvement, and evaluate the effectiveness of educational programs.

Artificial Intelligence in Education

The field of artificial intelligence (AI) involves the development of systems that can simulate human intelligence. AI systems are capable of learning from experience, solving problems, evaluating information, and performing logical operations. They can also analyze large amounts of data, detect patterns, and make accurate predictions automatically. The emergence of Information and Communication Technology (ICT) has revolutionized education, providing access to quality learning for everyone regardless of location or schedule. Online education, or distance education, leverages technological tools to create a virtual classroom accessible over the internet. It offers the flexibility of asynchronous learning, allowing students to interact with instructors and peers at designated times. Online education was created to accommodate individuals with work, family, or geographic constraints who cannot attend traditional classrooms. Interactive models that facilitate student engagement with content, instructors, and peers, easy accessibility from any location with internet access, both synchronous and asynchronous options, and access to online resources at any time are all key features of online education.

3. Method

The study conducted involved a university in Ecuador that offers two modes of study - traditional face-to-face approach and online education [30]. The traditional mode is teacher-centered, where the teacher determines the learning content and evaluation criteria, while the student is passive and must adhere to pre-set schedules [9]. This approach may not be ideal for a comprehensive learning model, as it can be subjective. On the other hand, the online education mode has evolved over ten years, with the integration of Information Technology (IT) and establishment of a technological foundation [30]. This mode caters to individuals who cannot attend in-person classes due to their schedules and obligations, and operates on a Learning Management System (LMS) platform, where courses are designed for each degree program. The student is required to complete three mandatory activities - completing tasks, evaluations, and participating in forums, which can be accessed through a module with access to all relevant resources [30].

The university's technology infrastructure is an advantage for the study, as it can support a high volume of transactions and services, with information security being maintained by the IT department. Figure 1 shows the technology architecture, including a layer for data analysis, which allows for the integration of AI, data analysis, and the learning management system without compromising data [30]. However, the online education mode faces challenges such as a high dropout rate and low academic effectiveness, which may be due to students lacking proper study methods and discipline to adhere to their own schedules. To address these challenges, the AI system is refined to enhance the online education experience, with the data analysis model being established based on the variables and inquiries to be addressed [30].

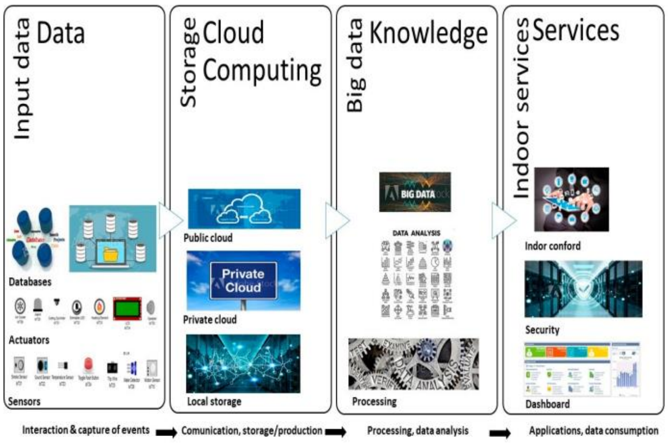

Figure 1. Technological architecture of a university for the management of an online education modality [10].

The layer in charge of data entry is accountable for gathering information from all systems and tools utilized by students [10]. This layer also takes into account data generated by students on social networks related to their student experience.The knowledge layer is dedicated to data analysis and employs big data architecture. It drives the online education model by processing data from various sources, analyzing it with data mining algorithms, and passing the information to the AI system. This AI system then generates insights and interacts with students and learning authorities.

The LMS service layer serves as the integration point for all the systems and layers discussed earlier, consolidating and presenting the information to users of the online education mode. The way the information is displayed can be customized and presented through various systems that are related to the educational model.

Analysis of Data

The importance of data analysis in this study cannot be overstated due to the vast volume and variety of data that needs to be processed. Big data technology is well-suited to meet these requirements, as it is designed to analyze data from various sources. Student-generated data from interactions with the Learning Management System (LMS) is stored in a structured manner in a separate database. However, relying solely on this data would limit the granularity of analysis and only provide segmented scores for each activity. To gain a more complete understanding of students' learning processes, additional information, such as socio-economic background and past academic performance, should also be integrated into the analysis. This includes information typically stored by universities and information obtained from students through sources such as social networks, both structured and unstructured. The objective is to collect information from all available sources to gain deeper insights into students' learning behaviors.

Hadoop Operation

The MapReduce computational process is executed within a cluster. This process involves assigning tasks to different servers within the cluster. Hadoop, which is a popular tool for MapReduce processing, manages the distribution of data between nodes in the cluster to reduce network traffic.

The MapReduce process consists of two main phases, which are the Map phase and the Reduce phase. During the Map phase, the initial data processing occurs, and data is mapped into key-value pairs. This phase is responsible for filtering and sorting the data, as well as grouping it by key.

The Reduce phase, on the other hand, further processes the data and is divided into two sub-phases: shuffling and reduction. The shuffling phase involves transferring data between nodes based on their key values. During the reduction phase, the data is aggregated and processed to produce the final output.

Once the MapReduce process is complete, the user receives the result generated by the cluster. This result could be in the form of a report, a graph, or any other output that was specified by the user. The MapReduce process is widely used in big data processing, machine learning, and data analytics applications.

Phases for the Implementation of Machine Learning

Before considering the technical approach, it is essential to identify the business objective that the machine learning tool aims to achieve. The goals can range from boosting conversions, lowering customer churn, to enhancing user satisfaction [6]. The key is to have a clear understanding of the aspect to be improved, to concentrate efforts and resources in that direction, avoiding implementation of a solution that exceeds the original goal [12].

The illustration in Figure 3 depicts the various stages of the machine learning process and how
they are interconnected.

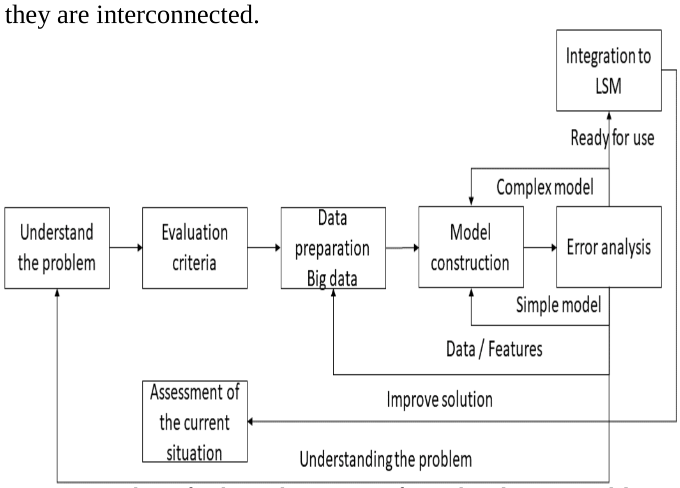

Figure 2. Phases for the implementation of a machine learning model.

To effectively build a machine learning model, it is crucial to establish a clear evaluation criterion. This involves choosing an appropriate error measure, depending on the type of problem being solved. For regression problems, root-mean-square error is commonly used, while cross-entropy is typically used for classification problems. In binary classification, accuracy and completeness measures are commonly utilized. Choosing the appropriate evaluation criterion is crucial for accurately assessing the performance of the model.

Before building the model, it is also important to perform an exploratory analysis of the data. This involves using descriptive statistics, correlations, and graphs to gain a better understanding of the data and the story it is trying to tell. This analysis can help determine if the data is sufficient and relevant enough to build a model.

In some cases, the problem may originate from a field with limited knowledge. In such cases, establishing collaborative relationships with individuals who have a deep understanding of the problem can be crucial in gaining insight into the data. Overall, the process of building a machine learning model involves multiple steps, including data preparation, feature engineering, model selection, and evaluation, among others.

Evaluating the current solution: Determine if there is an existing solution and measure its performance. Compare it to the machine learning model to determine feasibility. If there is no existing solution, create a simple solution and compare its performance to the machine learning model. Stick with the simple solution if it's comparable to the machine learning solution.

Preparing the data: Data preparation involves dealing with incomplete data. Actions such as deletion, imputation with a reasonable value, imputation using a machine learning model, or using a machine learning technique that handles incomplete data can be used. Harmonize data from multiple sources, calculate relevant features, and express data in intuitive ways to enhance performance.

Developing the Model: Select the machine learning technique and algorithm, and the algorithm will automatically learn to produce correct outcomes using prepared historical data.

Error Analysis: Identify areas of improvement for the machine learning results. Verify the model's ability to generalize and achieve accurate results when presented with new data. Repeating the previous phases multiple times can lead to exceptional results.

Model Integration into System: Integrate the model into the system, automate the data preparation process, and monitor errors automatically. Create interfaces for the data so that the model can obtain data automatically and the system can use its predictions automatically.

Integration of Big Data, Machine Learning, and LMS

A model, similar to the one depicted in Figure 4, is utilized for the integration of systems and new technology. The LMS (Learning Management System) has a vast amount of data regarding all student activities and interactions with the platform. Although the interaction is indirect, the LMS database often records the length of time each student spends actively using the platform, as well as the regular schedule in which each student logs in [10]. Additionally, data from administrative and other academic databases are also added to this information. This leads to a comprehensive analysis that involves a greater number of variables, processed by the big data architecture.

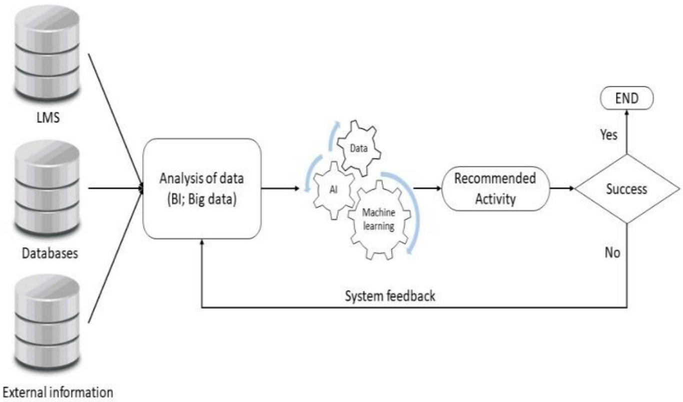

Figure 3. Big data integration model—Machine learning and LMS.

The big data architecture in its initial stage focuses on gathering data from various sources, which can be both structured and unstructured [9]. The collected data is then processed and made useful for the AI to use in machine learning. The machine learning algorithm is tasked with identifying patterns in the data, which it then uses to classify individuals based on certain characteristics. The goal is for the system to understand the needs of each group, allowing it to suggest strategies to improve the way activities are presented [5].

In addition to providing students with personalized learning experiences, learning management systems can also recommend activities based on individual student needs. After activities have been recommended, the system enters an analysis phase in which it evaluates the effectiveness of the activities by analyzing the grades that students receive. If the results indicate improvement, the process is repeated. However, if the system detects that the results fall below the university's average mark, feedback is integrated into the analysis phase to further improve the learning process, with the process being repeated until satisfactory results are achieved.

4. Discussion and Results

The integration of technology in education has revolutionized the way student performance is monitored. The traditional method of relying on teacher evaluation has been replaced by a new model that utilizes continuous analysis by the system and machine learning. The implementation of this model enables the system to identify students at risk of poor academic performance and provide early warnings, thus preventing further decline in academic performance. Unlike the conventional method, the new model uses the student's previous history to generate projections, which enables the system to detect potential difficulties in subjects related to a previously failed one, and also possible course repetition.
The machine learning model used in this study recommends further learning activities based on the student's performance in each activity. The model uses the best results a student has achieved to make decisions on which activities would be most beneficial. It also identifies groups of students that certain activities, such as rapid evaluations using true or false questions, do not meet the needs of, and suggests alternative activities that promote active learning. This approach prioritizes the development of active learning, offering students a range of activities tailored to their individual needs and abilities.

The proposed online education model was evaluated through exercises involving two parallel classes from an administrative career, each consisting of 24 students, monitored for a period of 16 weeks, equivalent to one academic term. The sample comprised students from the fourth level, as this level was chosen because students have taken all computer science subjects, making them better equipped to adapt to a technology-integrated model. The university's online education model adheres to a standardized 16-week format, divided into two seven-week segments and a final evaluation. The courses, created and registered within the LMS (Moodle in this case), were divided into modules corresponding to each week of the term and comprised a main module providing information on the study type, subject matter, and assigned tutor, as well as a syllabus and study guide outlining the topics and activities. Each weekly module was further divided into sections containing resources, activities, and information for asynchronous meetings with the tutor.

In the "resources section," the tutor is responsible for uploading all the material related to the week's topic. This material must be aligned with the subject's learning outcomes, and the tutor usually provides their own material such as presentations, exercise solutions, and readings. The material may also include supporting resources such as videos, articles, etc.

The "activities section" contains tasks for the students to complete each week. This includes an opinion forum where students critically discuss a topic raised by the tutor. Additionally, students must complete tasks based on Bloom's taxonomy which aims to ensure students acquire new skills and knowledge. The levels of Bloom's taxonomy are knowledge, understanding, application, analysis, evaluation, and creation. Lastly, students must take a questionnaire-style evaluation to encourage them to read the resources.

The final section retains information pertaining to the asynchronous meeting with the instructor. The purpose of this meeting is for students to have the opportunity to ask the tutor any questions they may have or receive feedback on topics discussed. Each meeting is 60 minutes long, and participation is not mandatory. If desired, students can review the recorded meeting as many times as they see fit.

In order to determine the factors contributing to student dropout, variables are established once the integrated model is defined. These variables include: the university degree equivalent to the average score a student received during their secondary studies, the number of successful course completions, the number of times enrolled within a specified time frame, the specific courses taken (with a range of 1 to 20, coded based on average course load), student gender, and age (between 19 and 30 years old). The objective is to identify the causes of university dropout, as previous studies have considered abandonment as a failure to continue enrollment in consecutive semesters.

For the first exercise, big data access to logs of teacher and student activities is required. These logs are typically stored in MySQL. The data collected from various sources undergoes processing and transformation to obtain clean data, which is then analyzed using Hadoop to uncover patterns in student behavior.

Most students in the forums demonstrate a high level of learning, but some receive low grades due to either not registering their participation or submitting non-objective contributions. While the tasks based on Bloom's taxonomy generally result in mean values that meet the activity requirements, the lowest performing activity is the questionnaire-type evaluations. These evaluations consist of 10 questions to be completed in 20 minutes, with an expected response time of two minutes per question. Unfortunately, these evaluations yield extremely low results that do not contribute to the learning process.

To address this issue, the big data result is fed into an AI for machine learning and decision-making purposes. The integrated AI model includes an analysis of data from the LMS, which records the time students spend reading the teacher's materials, as well as data from a survey of students that covers the amount of time they took to answer each question. The data from this analysis is subjected to a naive Bayes data mining algorithm, which conducts a detailed analysis on 51 instances to determine the reason for the below-expected scores in evaluations. Out of the 51 instances, 48 were found to be accurate with an accuracy rate of 94.1176%, which was deemed as a reliable result to form the analysis decision.

The analysis revealed that the allotted time of 2 minutes per question hinders the success of the evaluation. Findings were compared to the number of evaluations that were closed due to time expiration, with 18 instances identified as affected and one instance mistakenly detected as a difficulty with the evaluation. Additionally, 15 instances were observed where students did not dedicate enough time to reading the teacher's materials. The evaluation difficulty was also considered, with 15 instances recorded as being a cause and two instances recorded as being due to the time assigned per question. Based on these results, the machine learning model recommended that the tutor increase the response time per question to 2 minutes and 30 seconds, resulting in a total evaluation time of 25 minutes.

The model incorporates additional variables into the analysis and makes decisions based on the outcomes. Data was gathered from the LMS and a student survey, and the model enables weight adjustments to ensure effective decision-making, which becomes crucial when results need to be adjusted. The model's quick action allows prompt corrections to be made before a situation escalates into a problem, particularly when evaluations are conducted concurrently.

5. Conclusions

The education sector has undergone significant changes with the advent of virtual, online, or hybrid learning models, which are now at the forefront of learning and research. These models have integrated technology as a crucial aspect to enhance education, with the adoption of student-centric educational models holding the potential to improve learning outcomes and address issues such as high dropout rates and low academic performance.

The integration of technology in education has allowed for the development of systems that serve as valuable aids to both students and teachers. These systems facilitate students' calendar management, generating reminders, notifications, and events to keep them informed of their required activities. The system also provides continuous support to help students improve their performance, while enabling teachers to monitor each student's learning progress through a thorough analysis of the system's data.

This work discussed here serves as an evaluation tool for online education, generating indicators based on different learning methodologies. This helps improve the resources and activities in online education, making it a more efficient and effective model, reducing student dropout rates, and conserving economic resources. The model also plays a crucial role in monitoring students and ensuring educational quality, freeing up universities to reallocate human resources. The model is based on a granular analysis of large amounts of student data, and in the future, it is proposed to incorporate blockchain and internet of things technologies for increased security and data collection. These technologies could further enhance the evaluation model and provide additional layers of security and data analysis.

References

[1]. Li, H., Liu, S.M., Yu, X.H., Tang, S.L. & Tang, C.K. (2020). Coronavirus disease 2019 (COVID-19): Current status and future perspectives. International Journal of Antimicrobial Agents, 55, 105951, 2020. doi: 10.1016/j.ijantimicag.2020.105951

[2]. Riofrio, G., Encalada, E., Guaman, D. & Aguilar, J. (2015, October). Business intelligence applied to learning analytics in student-centered learning processes. In Proceedings of the Latin American Computing Conference. Arequipa, Peru. pp.1-10, 2015.

[3]. Beldarrain, Y. (2006). Distance education trends: Integrating new technologies to foster student interaction and collaboration. Distance Education, 27, 139-153, 2006. doi: 10.1080/01587910600789498

[4]. Hssina, B., Bouikhalene, B. & Merbouha, A. (2017). Europe and MENA Cooperation Advances in Information and Communication Technologies (Vol. 520). In A. Rocha, S. Mohammed & C. Felgueiras (Eds.), Springer International Publishing. 2017. doi: 10.1007/978-3-319-46568-5_43

[5]. Villegas-Ch, W., Lujan-Mora, S. & Buenano-Fernandez, D. (2017, November). Application of a Data Mining Method in to LMS for the Improvement of Engineering Courses in Networks. In Proceedings of the 10th International Conference of Education, Research and Innovation. Seville, Spain. pp. 6374-6381, 2017.

[6]. Comendador, B.E.V., Rabago, L.W. & Tanguilig, B.T. (2016, March). An educational model based on Knowledge Discovery in Databases (KDD) to predict learner's behavior using classification techniques. In Proceedings of the IEEE International Conference on Signal Processing, Communications and Computing, Shanghai, China. pp. 1-6, 2016.

[7]. Kim, T. & Lim, J. (2019). Designing an Efficient Cloud Management Architecture for Sustainable Online Lifelong Education. Sustainability, 11, 1523, 2019. doi: 10.3390/su11061523

[8]. Ferguson, R. (2013). Learning analytics: Drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4, 304-317, 2013. doi: 10.1504/IJTEL.2013.057405

[9]. Lee, S.J., Lee, H. & Kim, T.T. (2018). A study on the instructor role in dealing with mixed contents: How it affects learner satisfaction and retention in e-learning. Sustainability, 10, 850, 2018. doi: 10.3390/su10030850

[10]. Lee, J., Song, H.D. & Hong, A.J. (2019). Exploring factors, and indicators for measuring students' sustainable engagement in e-learning. Sustainability, 11, 985, 2019. doi: 10.3390/su11040985

[11]. Villegas-Ch, W., Palacios-Pacheco, X., Buenaño-Fernandez, D. & Luján-Mora, S. (2019). Comprehensive learning system based on the analysis of data and the recommendation of activities in a distance education environment. International Journal of Engineering Education, 35, 1316-1325, 2019.

[12]. Darcy, A.M., Louie, A.K. & Roberts, L.W. (2016). Machine learning and the profession of medicine. JAMA, 315, 2016.

