Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 220 (2023) 348–355

Procedia Computer Science

www.elsevier.com/locate/procedia

The 14th International Conference on Ambient Systems, Networks and Technologies (ANT)
March 15-17, 2023, Leuven, Belgium

Towards IoT-Big Data architecture for future education

Khadija Ahaidous[a,b*], Mohamed Tabaa[a], Hanaa Hachimi[b]

[a]*LPRI Lab, EMSI Casablanca, Morocco*
[b]*LGS Lab, USMS Beni-Mellal, Morocco*

Abstract

New technology, through the IoT, Big Data and AI ecosystem, is the foundation on which tomorrow's solutions are created. Improved processes with better decision making are the actors that drive the evolution of tomorrow's services. This new technology has enabled innovations in several sectors including medical, manufacturing, automation, energy, education, and others. It should be noted that in recent years, Industry 4.0 and the COVID-19 sanitary crisis have forced some remarkable transformations and contributions to people's quality of life, safety, health and education. Currently, the education sector in all its levels (primary, secondary and university) re-quires the use of this technology and is moving more specifically towards a system known as "education 4.0" to meet the different demands of this sector. In this paper, we present a literature review of the use of Big Data analytics in the education sector in particular. We discuss the different technologies, architectures, and applications of Big Data analytics in this growing sector, and we propose a new architecture based on the new technology applied in education.

Keywords: Big Data Analytics, IoT, Education 4.0, New technology, Architecture, Applications, Higher Education, new architecture.

1. Introduction

The number of internet users in 2021 is estimated by 4.9 billion [1], this huge number of users means increasing the amount of data generated daily, and the challenge behind, is to know how to benefit from it. Higher education is like any other field, which follows digital transformation, thanks to the positive impact of the last pandemic (COVID'19) that was able to push and accelerate this transformation [2] [3].

* Corresponding author. Tel.: +2-126-3473-0903.
E-mail address: k.ahaidous@usms.ma

10.1016/j.procs.2023.03.045

According to the Scopus database, the researcher community start to be interested in this particular area of education from 2016. The number of papers that are published in the last 10 years was 352, among which 98 papers published only in 2021 (>27%).

These results were found by looking on Scopus database, for research papers that contain at least “Big Data” and “Higher Education” in their keywords. All these factors give to authors of this paper, the motivation to discover the potential of Big Data in this important field.

The main objective of this paper is to focus on Big Data applications in Higher Education, in the era of Education 4.0. For this reason, we conduct this review based on research papers that are published, in the last 5 years (Between 2018 and 2022), from Google Scholar and Scopus. The searching keywords mainly used are: “Big Data”, “Higher Education” and “Education 4.0”.

This paper is organized into 5 sections: the first section is an introduction to Education 4.0 and the history behind, the second section presents a detailed definition of Big Data, its characteristics, the architecture used today, and the most popular tools used in Higher Education. The third section presents some Big Data applications combined with latest technologies like Artificial Intelligence. The fourth part discusses many Big Data applications related to IoT, machine learning and virtual reality. The final part, we propose our architecture followed by a conclusion and perspectives.

2. Education 4.0

The appearance of the fourth industrial revolution impacts a lot of industries, including education sector. From Education 1.0 (late 18^{th} century) that was based on traditional environment and methods, to Education 4.0 that is essentially based on new technologies and innovation, the actors of education systems seek to equip students and teachers with the skills necessary for 21st century [4].

To deep dive into the steps of this transformation, we start with the first generation -Education 1.0-, this generation was characterized by systems mechanization like paper-making-machine and typewriter, so the teachers was seen as a knowledgeable and smart person, but in the other side, students almost has no potential in this system. After the second industrial revolution, this period was an electricity period and mass production. At that time, the main source of information was libraries, teacher’s role changed to be only an information source, but student’s role still passive. At the end of 20th century, the focus was on the automation and control. The learning process moved to a new era, by using online tools and multimedia course materials. So, at this point, teachers and students started to collaborate to build their knowledge. Actually, we survive in the fourth industrial revolution, which the innovative ideas are developing each day, and the self-learning centred on students is the main approach followed.

According to [2], Education 4.0 is guided by the most innovative technologies, including artificial intelligence, data management, cloud computing, robots and sustainable technologies, however, to get a successful teaching, teachers, educators and parents must have a certain level of maturity in these fields.

There are many countries that started to develop Education 4.0 architectures in their universities: Iraq, at university of Kufa [5], China in medical education [6], many universities and colleges in United States (US) [7] and United Kingdom (UK) that is considered as the most developed model of Higher Education in the world [8]. The section 4 contains some examples of Big Data use cases combined with other technologies and applied in Higher Education.

3. Big Data

3.1. Definition

According to the Cambridge dictionary [9], Big Data refers to very large sets of data that are produced by people using the internet, and that can only be stored, understood, and used with the help of special tools and methods. This is in accordance with Oxford dictionary [10] that refers to it as sets of information that are too large or too complex to handle, analyze or use with standard methods.

Examples of Big Data type vary between what is posted on social media, like tweets and videos and what is stored from sensors, cameras and so on.

3.2. Characteristics

Big Data is measured in terabytes, petabytes... opposed to traditional data, and it is recognized thanks to a set of characteristics. The way data is stored, processed, and analyzed is essentially impacted by these characteristics. There is not only one way to describe the characteristics of Big Data, in their paper [11], authors represent the evolution of those characteristics, going from 3 V's : volume, velocity and variety [7] [12] [13], to 5 V's: volume, velocity, variety, veracity and value [14] to 7 V's: 5V's plus visualization and variability to the previous set [15].

Over and above that, the researchers add three more characteristics: volatility, viscosity, and viability (See Fig. 2).

According to the above, 10 V's are used today. They are listed and described below:

Volume	It refers to a large amount of Data, generated by different sources. Nowadays, the size of Data is not challenging, due to the fast evolution of Data storage technologies (e.g., Cloud technology), and the decrease of the price of these solutions [16].
Value	The finality of Big Data is getting the results that help in decision making and business growth. So, having useful data is one of the most important aspects in Big Data.
Velocity	The third V in this list refers to the speed of Data Generation [16] or Data Production [17] . In Big Data Analysis, Data should move quickly to be available at the right times.
Variety	Represents the different forms of Data. It can be a simple text, images, videos, audios etc.
Veracity	Represents reliability and accuracy of data. The credibility of data is what gives it value and ensures its quality making it dependable.
Variability	is one of the most challenging characteristics in Big Data. It represents the consistent change of data, in term of the flow and speed of data loaded.
Visualization	refers to the method you should follow to explore data in a comprehensive way and easy to understand.
Volatility	describes data lifetime i.e., how much time do we need before we say that data is no longer useful or relevant.
Viability	not all data is useful, the challenge is how to keep only the suitable data. For example, by calculating the correlation between two variables, we can decide if this data is significant and viable or not.
Validity	Validity and veracity may have almost the same meaning, but not the same concept. Validity refers to the understandability and the intelligibility of Data.

3.3. Big Data Architecture

Big Data architecture refers to the structure/layout of the systems relied on to treat the high volumes of data. Components of this architecture vary according to the age of data going from data sources to the result displayed to the end user.

Generally, Big Data architecture can be split into three layers: Storage layer, Processing Layer and Access layer.

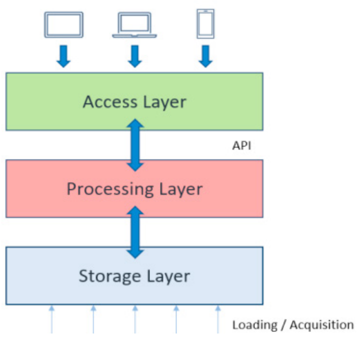

Figure 1- Big Data Architecture

Storage Layer: There are so many sources of data that should be stored and one of the main obstacles to how implementing this data in Big Data Analytics [18]. Data can be received from Learning Management Systems (LMS), Students Information System (SIS), log files, sensors, IoT etc. [19]. Obviously, this Data does not follow a unique format, it can be classified into three categories: structured data (e.g., students' information stored in SQL Databases), semi-structured data (e.g., log files stored in CSV, JSON or XML files) or unstructured data (e.g., Sensor's Data...).
All these data types must be stored and archived in a trusted way to be used in demand. Without a doubt, Hadoop was and is still associated with Big Data as a solution of storing and processing data for later. Hadoop might be used only as a storage system through HDFS or combined with Business Intelligence tools (Data warehouses…).

Processing Layer: The second layer represents the core of Big Data Analysis Systems (BDAS). There are two types of processing: Batch processing and Stream Processing:

- Batch processing: is where the processing happens to data that is stored over a period, e.g., estimating a student skill, based on historical data, using machine learning algorithm and tools like Spark.
- Stream processing: is where the processing happens in real time, e.g., fraud detection.

Access Layer: With IoT devices, results of analysis can be viewed using any connected object to internet, from laptops to smart watches.

4. Big Data Application in Higher Education

Higher Education ecosystem can use Big Data in different departments: Finance, Marketing, Pedagogy, Human Resources and Research [20] and it has a direct connection to six stakeholders [21]: Learners, educators, educational researchers, course developers, learning institutions, and education administrators. Using Big Data, learners can interact with course material instantly, and give and receive feedback. Making use of this data, they can also be provided with useful information about their gaps and weaknesses. As for educators, from now on, they no longer need to wait for assessments to evaluate students, they can analyze the performance of a class, or specifically the performance of a student, on a microscopic level, periodically, to know what is wrong and help students enhance their knowledge.

Educational researchers, in their turn, can use the data collected to suggest and test new learning models and practices. Based on, students/educators' feedback and students' level, the course developers can adapt existing courses, suggest new materials, and design new courses to meet the needs. The learning institutions also can benefit from this data, by targeting and reaching potential students. Lastly, education administrators can easily get an idea about the effectiveness of the programs launched and the education strategy followed.

4.1. Big Data & Internet of Things

Internet of Things (IoT) is a concept that was invented in 1999 by Kevin Ashton [22], it is one of buzzwords in digital world that transforms any objects into an intelligent object and give to human beings the power to manipulate those objects. The combination of Big Data and IoT plays a key role in E-learning platforms. It can be of massive help in the transformation of education. The fourth Big Data characteristic "Variety" is discernible in the IoT field: all types of data can be captured by the means of devices installed in learning environments and made more engaging.

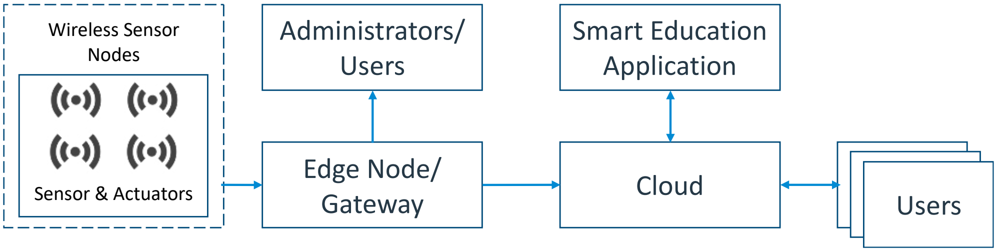

Figure 2 A Basic IoT Architecture [23]

The author of the article [23] believes in the continuity of digital transformation and the revolutionary concept of IoT, which is necessary for better engagement among students and teachers. For this reason, the author describes the IoT requirements for smart education, and he suggests a generic IoT architecture which leads to Big Data.

In order to design an architecture, the researcher needs to specify a list of requirements. These requirements are divided into two categories: functional and non-functional. The non-functional requirements are related to the implementation and operation of IoT itself, like availability and scalability. However, the functional requirements refer to IoT actors like application support, device management etc.

The figure 3 shows the suggested architecture, it is made up of five layers:

a. Learning Data Collection Layer (LDCL): represents physical components: sensors, gateways, mobile phones etc.
b. Learning Data Integration Layer (LDIL): provides an abstract level to homogenize data access and services provided by LDCL.
c. Learning Information Access Layer (LIAL): is a canal level, that provides access to information, using a cloud service, which will be used by the superior layer.
d. Learning Knowledge Processing Layer (LKPL): is a processing layer, that deal with LIAL.
e. Learning Application Layer (LAL): provides a last layer which end-users will interact with, using dashboards, APIs.

To evaluate this architecture, the researcher follows weighted sum model, to validate 24 criteria.

Another concrete example can be found in [24]. The authors choose smartwatches as an IoT device to take the advantage of a multitude of embedded sensors, like Pedometer, Heart rate and Motion. The experience was about answering two hypotheses: “How students will accept the use of wearable device from the initial pre-test to the post experimental test?” and “Are the participants notice changes of their subjective opinion about wearable devices?”

By collecting and analyzing data from participant’s smartwatches, for one month, the authors found that students were more committed to accept the use of smartwatches, they notice changes of their subjective opinion about wearable devices, also, they can use these data to encourage and motivate students to have a better healthy life.

4.2. Big Data & Machine Learning

Big Data as an architecture, is not enough, it should be combined with other tools and techniques, to be more beneficial. In this section, we discuss the use case of machine learning algorithms applied to Ecuadorian university data [25]. To test the approach, the authors started working on a small dataset, with the aim of scaling it on a Big Data architecture using Hadoop and MangoDB.

Based on a historical academic data (2016 - 2018) of 335 students in computer engineering, the authors propose a methodology to monitor and predict grades in education, in order to obtain the best results in term of prediction.

The first step in this process is collecting and cleaning data from the Academic Management System (AMS). The second step consists of analyzing the training dataset using classification algorithms to identify similar groups of students. Thereafter, using decision tree algorithm, the researchers test different attributes: Partial Grades (PGs) and Final Grade (FG) to find the best correlation, i.e., the best prediction to know if a student will pass a certain subject or not. Finally, the last step is about visualization of results in an easy way to be interpreted.

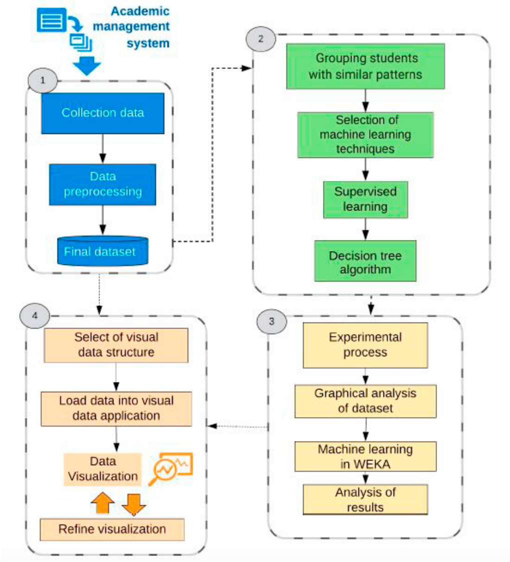

Figure 3 Proposed Methodology [25]

As mentioned above, to be approved, this approach requires an application on a large scale and more details about the classification algorithm used.

Other use cases from US universities [7], use predictive analysis to:

- Identify at-risk students and increase retention rate by targeting students and advising them. (Concordia University Wisconsin, Delaware State University, Temple University)
- Predict students' success and failure and help them. (Washburn University, University of Maryland, Arizona State University)
- Identify potential donors. (Michigan State University, University of California - Santa Barbara)
- Develop adaptative learning to enhance and accelerate learning [18]. (Colorado Technical University)
- Improve student's academic experience and boost graduation rate. (Arizona State University)

All the examples, cited above and other, have been proved effective:

- Arizona State University boost graduation by 20%
- Concordia University Wisconsin increased student retention rate to 82%.
- Delaware State University increased student retention rate to 70%.
- Open University increased student retention rate by 2.1% compared to a previous year and generated an estimated £1.8 million. [21]

4.3. Big Data & Virtual Reality

Virtual Reality (VR) refers to a set of images and sounds, produced by a computer, that seem to represent a place or a situation [26], in other words, VR is a "simulation system that integrates a full range of computer-generated computer graphics, computer simulation, artificial intelligence, sensing, display, network parallel processing and other technologies." [27]

The first model of headset was appeared in 1960 called "Telesphere Mask". So far, the development still in progress and it takes several forms: Augmented Reality (AR), Mixed Reality (MR), Extended Reality (XR) [27]. Recently, we can notice that the use of VR is exposed and affects many areas including Education. In their research paper [28], the authors conducted a review of the effect of VR in Education and summarize the outcomes of some other papers.

5. Our proposed architecture

In the literature, there is no single or universal Big Data architecture on which all research works agree. According to some t, this architecture can be divided into three layers: the storage layer, the processing layer and the access layer.

Due to the limitations of this type of architecture and in order to meet the IoT requirements for a smart campus, an improved version has been suggested, based on six layers: Perception Layer, Network Layer, Storage Layer, Processing Layer and Application Layer.

- Perception Layer: or sensor layer. Its major responsibility is to recognise objects/things and collect data from them.
- Network Layer: It works like a bridge between Perception Layer and Storage Layer. Depending on the sensor's type, a protocol network will be used: LoRa for long-range transmissions, WiFi for short-range transmissions with a greater volume of data.
- Storage Layer: It combines all types of Databases needed in a smart campus: SQL for relational databases, NoSQL for no documents, objects etc.
- Processing Layer: Processing is required for every piece of data that was previously saved. Based on the type of data and the final application, the system provides several algorithms that will be used to address the ultimate requirement.
- Application Layer: It defines all the applications that will be used in a smart campus by administrators, teachers, students…
- Security Layer: It is an essential layer that will be provided in many aspects of the architecture, namely in Application Layer and Storage Layer.

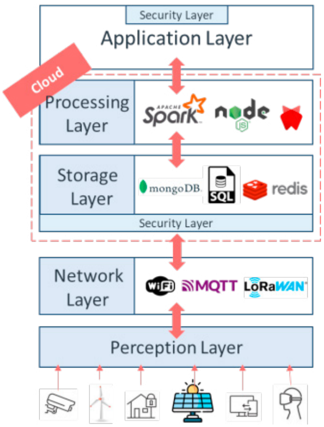

Figure 4 Our proposed architecture

6. Conclusion

The education sector is moving more towards the use of new technology to manage and create adequate training content for a new generation that is increasingly connected. For their part, schools and universities need to record all possible academic data from all available sources: Learning Management Systems (LMS), student information, courses, etc. to get the maximum benefit. Big Data and advanced technological tools like AI, virtual reality, digital simulation tools, IoT must collaborate with each other for the redesign of tomorrow's education models. In this article, we have presented a literature review regarding the use of Big Data as a must-have technology for the development of applications for education. In the paper, we, also, present the tools and examples of Big Data in this rapidly evolving sector.

Future work will focus more on the development and the deployment of a Big Data architecture, proposed above, in an engineering school in Morocco. The idea is to develop an architecture capable of responding to the massive need to use several technological tools and simulations for a better management of training for engineers.

References

[1] 'Number of internet users worldwide 2021', *Statista*. https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/ (accessed May 26, 2022).

[2] L. I. González-Pérez and M. S. Ramírez-Montoya, 'Components of Education 4.0 in 21st Century Skills Frameworks: Systematic Review', *Sustainability*, vol. 14, no. 3, p. 1493, Jan. 2022, doi: 10.3390/su14031493.

[3] D. Gašević, 'Learning analytics in higher education – Stakeholders, strategy and scale', *The Internet and Higher Education*, p. 5, 2022.

[4] J. Miranda *et al.*, 'The core components of education 4.0 in higher education: Three case studies in engineering education', *Computers & Electrical Engineering*, vol. 93, p. 107278, Jul. 2021, doi: 10.1016/j.compeleceng.2021.107278.

[5] A. Y. Mjhool, A. H. Alhilali, and S. Al-augby, 'A proposed architecture of big educational data using hadoop at the University of Kufa', *IJECE*, vol. 9, no. 6, p. 4970, Dec. 2019, doi: 10.11591/ijece.v9i6.pp4970-4978.

[6] K.-J. Liu, Y.-D. Cao, Y. Hu, and L.-J. Wei, 'Application status and development of big data in medical education in China', p. 8.

[7] M. Attaran, J. Stark, and D. Stotler, 'Opportunities and challenges for big data analytics in US higher education: A conceptual model for implementation', *Industry and Higher Education*, vol. 32, no. 3, pp. 169–182, Jun. 2018, doi: 10.1177/0950422218770937.

[8] B. Williamson, 'The hidden architecture of higher education: building a big data infrastructure for the "smarter university"', *Int J Educ Technol High Educ*, vol. 15, no. 1, p. 12, Dec. 2018, doi: 10.1186/s41239-018-0094-1.

[9] 'BIG DATA | meaning in the Cambridge English Dictionary'. https://dictionary.cambridge.org/dictionary/english/big-data (accessed May 19, 2022).

[10] 'big-data noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com'. https://www.oxfordlearnersdictionaries.com/definition/english/big-data (accessed Apr. 29, 2022).

[11] A. Alkhalil, M. A. E. Abdallah, A. Alogali, and A. Aljaloud, 'Applying Big Data Analytics in Higher Education: A Systematic Mapping Study', *International Journal of Information and Communication Technology Education*, vol. 17, no. 3, pp. 29–51, Jul. 2021, doi: 10.4018/IJICTE.20210701.oa3.

[12] M. Beerkens, 'An evolution of performance data in higher education governance: a path towards a "big data" era?', *Quality in Higher Education*, vol. 28, no. 1, pp. 29–49, Jan. 2022, doi: 10.1080/13538322.2021.1951451.

[13] M. Anil, 'An Efficient Way of Applying Big Data Analytics in Higher Education Sector for Performance Evaluation', *IJCA*, vol. 180, no. 23, pp. 25–32, Feb. 2018, doi: 10.5120/ijca2018916434.

[14] S. Pratsri and P. Nilsook, 'Design on Big data Platform-based in Higher Education Institute', *HES*, vol. 10, no. 4, p. 36, Oct. 2020, doi: 10.5539/hes.v10n4p36.

[15] S. S. Ali, 'A PROPOSED MODEL BASED ON CLOUD COMPUTING TECHNOLOGY TO IMPROVE HIGHER EDUCATION INSTITUTIONS PERFORMANCE', p. 17.

[16] N. Khan, M. Alsaqer, H. Shah, G. Badsha, A. A. Abbasi, and S. Salehian, 'The 10 Vs, Issues and Challenges of Big Data', in *Proceedings of the 2018 International Conference on Big Data and Education*, Honolulu HI USA, Mar. 2018, pp. 52–56. doi: 10.1145/3206157.3206166.

[17] A. A. Hadwer, D. Gillis, and D. Rezania, 'Big Data Analytics for Higher Education in The Cloud Era', in *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*, Suzhou, China, Mar. 2019, pp. 203–207. doi: 10.1109/ICBDA.2019.8713257.

[18] S. S. Chaurasia, D. Kodwani, H. Lachhwani, and M. A. Ketkar, 'Big data academic and learning analytics: connecting the dots for academic excellence in higher education', p. 27.

[19] M. Y. Amare and S. Simonova, 'Learning analytics for higher education: proposal of big data ingestion architecture', *SHS Web of Conf.*, vol. 92, p. 02002, 2021, doi: 10.1051/shsconf/20219202002.

[20] M. Jha, S. Jha, and L. O'Brien, 'Re-engineering Higher Education Learning and Teaching Business Processes for Big Data Analytics', in *Business Information Systems*, vol. 354, W. Abramowicz and R. Corchuelo, Eds. Cham: Springer International Publishing, 2019, pp. 233–244. doi: 10.1007/978-3-030-20482-2_19.

[21] S. Ray and M. Saeed, 'Applications of Educational Data Mining and Learning Analytics Tools in Handling Big Data in Higher Education', in *Applications of Big Data Analytics*, M. M. Alani, H. Tawfik, M. Saeed, and O. Anya, Eds. Cham: Springer International Publishing, 2018, pp. 135–160. doi: 10.1007/978-3-319-76472-6_7.

[22] M. Abdel-Basset, G. Manogaran, M. Mohamed, and E. Rushdy, 'Internet of things in smart education environment: Supportive framework in the decision-making process', p. 12.

[23] K. Palanivel, 'SMART EDUCATION ARCHITECTURE USING THE INTERNET OF THINGS (IOT) TECHNOLOGY', vol. 9, no. 4, p. 28, 2019.

[24] M. I. Ciolacu, L. Binder, P. Svasta, I. Tache, and D. Stoichescu, 'Education 4.0 – Jump to Innovation with IoT in Higher Education', in *2019 IEEE 25th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, Cluj-Napoca, Romania, Oct. 2019, pp. 135–141. doi: 10.1109/SIITME47687.2019.8990825.

[25] D. Buenaño-Fernández, D. Gil, and S. Luján-Mora, 'Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study', *Sustainability*, vol. 11, no. 10, p. 2833, May 2019, doi: 10.3390/su11102833.

[26] 'VR'. https://dictionary.cambridge.org/dictionary/english/vr (accessed Jun. 03, 2022).

[27] S. Xiao, 'VR Open Computer Network Virtual Laboratory Based on Big Data Technology', *J. Phys.: Conf. Ser.*, vol. 1648, no. 4, p. 042105, Oct. 2020, doi: 10.1088/1742-6596/1648/4/042105.

[28] Z. Alrababah and S. Shorman, 'Review of Effects of the Virtual Reality in Education', in *The Big Data-Driven Digital Economy: Artificial and Computational Intelligence*, vol. 974, A. M. A. Musleh Al-Sartawi, Ed. Cham: Springer International Publishing, 2021, pp. 445–459. doi: 10.1007/978-3-030-73057-4_33.