Application of Machine Learning on Educational Big Data: A Review

Syyada Shumaila Khurshid[1], Ravi Kumar Singh Pippal[2]

[1]*MTech Scholar,* [2]*Professor*

[1]*Department of Computer Science Engineering, Vedica Institute of Technology, Bhopal, India*

[2]*Department of Computer Science Engineering, Vedica Institute of Technology, Bhopal, India*

shumailakhurshid74@gmail.com[1] ravesingh@gmail.com[2]

Abstract — ***Education is being transformed by machine learning and big data technologies, which is shifting it away from the standard classrooms and toward a mixed learning paradigm that includes both physical and digital elements. Data and artificial intelligence (AI) are being used to improve the quality of education while also making it more flexible. With this, decision-making skills can be enhanced. As a result, this paper examines the role of machine learning and big data in today's educational system.***

Keywords — ***Big Data, Educational Big Data, Machine Learning.***

I. Introduction

More people can afford scientific computing now that Internet of Things and Big Data are in such high demand. Large amounts of data must be evaluated and transformed on a regular basis to keep up with changing market trends that create a wide range of needs. Big data can be useful in higher education because of this. Education delivery has changed dramatically as a result of technological advancements [1]. Some examples of technology used in education include mobile devices (such as smartphones and tablets), educational platforms and services, conference calling and remote access systems (such as laptops and desktop computers) and others with which assessment specialists (such as test developers), researchers, university faculty members and educational decision makers interact. If you use just one access log file to track institutional activity, you'll create a ton of data. But educational systems aren't equipped to deal with it or harness its potential for long-term, systemic quality improvement. More than ever, we must deal with the problems in our educational system. Because of its potential value in facilitating knowledge discovery, big data and analytics were also taken into consideration.

II. BIG DATA AND EDUCATION

It is usual today to refer to massive data sets that have only just been developed or are already in existence as “big data”. It’s used across a wide range of industries. Huge amounts of data from a variety of sources and formats are not uncommon in government, business, and the non-profit sector. Large amounts of data are known as “big data” [1]. They’re generated on a regular basis and, in some cases, are being analyzed right now. It’s all about speed, volume and variety when it comes to data: what’s the source, what’s the format, what’s the type.

A system or domain has high volume, diversity, and speed all at the same time, even if these 3 characteristics are viewed as “small” in another domain. This is referred to as big data. The only thing that needs to be done in this scenario is to raise awareness about the constraints of information filtering and assessment. Megabytes to petabytes in data size are possible, relying on the scope. Thus, big data is domain-specific and can refer to a wide range of dimensions and types based on the domain. Big data In all of these domains, understanding data through sophisticated analytical processing and enabling the development of new procedures and strategies based on that understanding is a challenge [3].

Business intelligence and analytics [4] have shown to be a very effective way of their future influence on industries in terms of their contribution to data value in numerous settings and in the field. Those It’s been established that learning, studying, and analysing all have educational benefits [5]. There are a multitude of people to talk to about data-driven educational quality activities, assessment specialists, researchers, computer scientists, academics, and including policy makers of all types and levels.

III. EDUCATIONAL BIG DATA (EBD)

Data volume, diversity, and speed all coexist together in higher education. Every day, a vast amount of educational data is collected and generated from many sources and in various formats in the university environment. Learning objectives, teaching materials, activities, exam results, and other types of data related to education and quality improvement processes and procedures comprise educational data, which ranges from what results from students’ use.

Due to the weak support for big data in learning and the amount and type of this data in higher education, special methods need to be used to find valuable additional clues that are currently hidden in the information [6]. These strategies can be adapted and successfully used to manipulate big data in education with techniques from other fields dominated by big data. These strategies could be implemented to help generate data "related to student learning and teaching skills" and to highlight regions in huge educational data that can be modified at low cost, such as: actual student progress according to the proposed course [7]. Big data and analytics have shown great promise lately to support a wide variety of initiatives in higher education. Administrative decision-making and the allocation of organizational resources issue such measures, which aim to keep students from struggling by detecting them early, improve teaching methods and reshape the conservative view of a scheme into a system of large datasets. The curriculum and its content are a key component

of educational big data since they may be used effectively for studying and enhancing higher education [9].

A. Objectives of EBD

1. Predicting and identifying the learning habits of students.
2. Increasing the quality of the educational experience.
3. Recognizing students' development and faculty members' teaching styles.
4. To create a course structure and analyze the results of that structure on student learning.
5. A learning model is developed using data mining or machine learning techniques.
6. Improves administrative structure and resource organization with the help of this approach Aids in enhancing the efficiency of educational programs.

IV. RELATED WORK

Yu et al. [15] focuses on the extraction of payloads from large online educational data using learning transfer with Hadoop to create the Online Educational Data Classification Framework (OEDCF) and design a Tr_MAdaBoost algorithm. Liet al. [16] Data used such as information on student performance in a vocational school and information on campus one-card consumption. First, the original data is preprocessed so that the processed data can meet the basic requirements of data analysis. After preprocessing, the data is then discretized using data discretization technology so that the data can meet the a priori algorithm requirements for the assignment rules. Finally, the Apriori algorithm is used to find the relationship between student performance and consumption data from an on-site card. Santos et al. [17] proposed a computational approach that uses data mining in education and various supervised learning techniques to evaluate and assess the behavior of various predictive models identify. Profile of Higher Education Students at Risk in the Context of Brazilian Higher Education. Ketu et al. [18] focused on classification models for use in educational data mining. Classification models are used to identify the appropriate subject for science students. The experiment is designed to improve student performance by comparing the performance of five assessment models and then predicting the corresponding academic performance in each specialization.

Shao et al. [19] proposed an optimized mining algorithm to analyze students' learning status using dynamic data.

Researchers used social network analysis' traditional centrality metric, as well as statistical performance metrics, to evaluate the study's results. As a result, many efforts have been made to establish an effective educational data big data analytical approach. Figure 3 depicts the most common issues and limits with the methodology and analytical approach that need to be addressed.

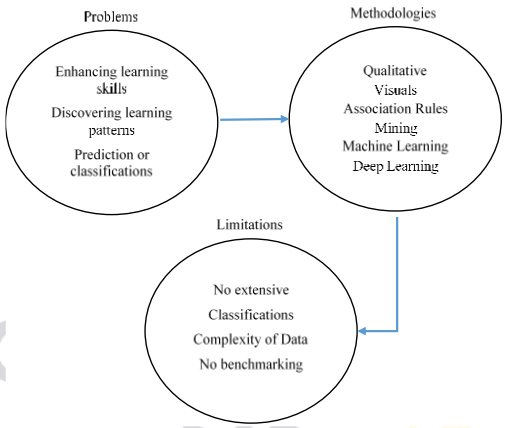

Figure1: Observations from Existing Educational Big Data

The following are the key focuses of educational machine learning and data mining applications:

1. Predicting Future Behavior of Student: Simulations is being used to promote student Profiles depends on the quality of learners and may use adequate data such as their motivation and learning behavior to do so.
2. Identify or Improving Domain Models: Using data mining techniques, it is feasible to discover new models or improve on existing ones.
3. Study of Educational Support: The educational support system can be studied using the learning system.
4. Scientific Knowledge of Learners: Educational Data Mining research and technology tools, as well as Machine Learning, can be utilized to develop student models.
5. Techniques of Machine Learning and Their Use in EBD: Supervised and unsupervised learning are the two basic Machine Learning approaches, and they are both employed in educational databases.

V. IMPACT OF MACHINE LEARNING ON EDUCATION SYSTEM

Machine Learning's growing importance has resulted in rapid growth. Artificial Intelligence is the next step in the tech boom since it brings together machines and humans. It's complicated, but it affects nearly every aspect of our daily lives. Currently, ML shows their advancement in our day-to-day activities. Education isn't exempt from the advances made possible by machine learning. The following are some of the most common educational applications of machine learning:

1. Understand Individual Level of Student: By employing Machine Learning in a classroom study, students with different personalities can learn about each other's strength and weakness.
2. To Provide Education to Special Ability Students: The resources available to educate students with special needs

are still quite restricted in the modern world. Though Machine Learning could be used to teach it more readily.

3. Grouping Students based on Different Indicators: Students are sorted into groups depending on a variety of factors, such as personal qualities and test scores. These tasks make use of Data Mining and Machine Learning. Classification and clustering are just two forms of data mining. Model-based clustering and hierarchical agglomerative clustering are both used. To locate pupils with similar learning styles, we use a clustering technique.
4. Grading Students: Calculating Student Grades Students' answer scripts can be evaluated with the aid of machine learning. To analyze the answer scripts objectively, we may need to use AI and ML.
5. Learning as Per Your Capability: Instead than forcing children to learn anything in a set amount of time, learning at one's own ease and pace is the most scientific technique. Every kid possesses a certain amount of learning and comprehension potential. If a teacher had the option of teaching each student individually (using Machine Learning), it will be much easier.
6. Communication: With the help of Machine Learning, communication can be established between teacher and student anytime and anywhere. Students are automatically paired with their friends, which aids in learning and expanding networks.
7. Differentiation: Students and teachers will be able to connect with the resources they need with the availability of Machine Learning in the classroom. The Internet is the most popular resource these days.

VI. BIG DATA PLATFORMS

Different platforms are capable of handling large amounts of data. Platforms such as Hadoop and Spark are frequently utilized in big data applications. Large data sets and real-time analytics are two common uses for Apache Spark. Table 1 lists various big data platforms, along with the areas in which they can be used.

Table 1: Description of different big data platform

Platform	Description
MapReduce	Working with large datasets in parallel. Utilization of low-cost groupings. Data support for both structured and unstructured sources.
HDFS	A filesystem that is distributed, scalable, and portable. Large files are stored using a master-slave architecture.
Hive	Analyses and summarizes data and results.
NoSQL	Data preservation and retrieval
Spark	Flow data and perform iterative procedures Analyzer for Query Graph analyzer with a large processing capacity.
Samza	Use the same computer for storage and processing, and don't add any extra RAM.

VII. CONCLUSION

By utilizing technology in education, we're creating big data, which includes student, teacher, and school data. This article explores the ways in which machine learning can improve the current educational system. Educational databases can use machine learning applications to anticipate future outcomes in data. These applications can also be used in current systems. In order to strengthen the overall education system, educational institutions will benefit from this research by better allocating people and material resources.

REFERENCES

[1] X. Bai et al., "Educational Big Data: Predictions, Applications and Challenges," Big Data Res., vol. 26, p. 100270, Nov. 2021.

[2] J. T. Wassan, "Discovering Big Data Modelling for Educational World," Procedia, 176, 2015, pp. 642–649.

[3] D. J. Lemay, C. Baek, and T. Doleck, "Comparison of learning analytics and educational data mining: A topic modeling approach," Comput. Educ. Artif. Intell. 2021.

[4] M. Alam and M. R. Devjani, "Analyzing energy consumption patterns of an educational building through data mining," J. Build. Eng., vol. 44, 2021.

[5] M. Lnenicka, H. Kopackova, R. Machova, and J. Komarkova, "Big and open linked data analytics: a study on changing roles and skills in the higher educational process," Int. J. Educ. Technol. High. Educ. 2020 171, vol. 17, no. 1, 2020, pp. 1–30.

[6] M. I. Baig, L. Shuib, and E. Yadegaridehkordi, "Big data in education: a state of the art, limitations, and future research directions," Int. J. Educ. Technol. High. Educ. 17(1), 2020, pp. 1–23.

[7] Y. Wang and Y. Yang, "Research on Higher Vocational Teaching Quality Improvement Based on Educational Big Data," IEEE, 2020, pp. 227-230.

[8] X. Yu and S. Wu, "Typical Applications of Big Data in Education," 2015 IEEE, 2015, pp. 103-106.

[9] S. Roy and S. N. Singh, "Emerging trends in applications of big data in educational data mining and learning analytics," IEEE, 2017, pp. 193-198.

[10] K. Abe, "Data Mining and Machine Learning Applications for Educational Big Data in the University," IEEE, 2019, pp. 350-355.

[11] K. L. Ang, F. L. Ge and K. P. Seng, "Big Educational Data & Analytics: Survey, Architecture and Challenges," in IEEE Access, vol. 8, pp. 116392-116414, 2020, doi: 10.1109/ACCESS.2020.2994561.

[12] N. Zhang, B. Geng, W. Hu and R. Wen, "The Applications of Big Data Analysis in Student Management Education," IEEE, 2021, pp. 55-58.

[13] H. Hu and H. Yan, "A Study on Discovery Method of Hot Topics Based on Smart Campus Big Data Platform," IEEE, 2016, pp. 176-179, doi: 10.1109/ICITBS.2016.10.

[14] A. Z. Bhat and I. Ahmed, "Big data for institutional planning, decision support and academic excellence," IEEE, 2016, pp. 1-5, doi: 10.1109/ICBDSC.2016.7460353.

[15] L. Yu, X. Wu and Y. Yang, "An Online Education Data Classification Model Based on Tr_MAdaBoost Algorithm," in Chinese Journal of Electronics, vol. 28, no. 1, pp. 21-28, 1 2019.
[16] Z. Li, "New Employee Student Repast Big Data Analysis Research Application," IEEE, 2020, pp. 583-586.
[17] K. J. de O. Santos, A. G. Menezes, A. B. de Carvalho and C. A. E. Montesco, "Supervised Learning in the Context of Educational Data Mining to Avoid University Students Dropout," 2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT), Macei??, Brazil, 2019, pp. 207-208.
[18] N. Ketui, W. Wisomka and K. Homjun, "Using Classification Data Mining Techniques for Students Performance Prediction," IEEE, 2019, pp. 359-363.
[19] Z. Shao, H. Sun, X. Wang and Z. Sun, "An Optimized Mining Algorithm for Analyzing Students' Learning Degree Based on Dynamic Data," in IEEE Access, vol. 8, pp. 113543-113556, 2020.