Survey on Big Data Analytics for the Educational Sector : Abiding Challenges and Contributions

Shrikant Bhandalkar, Aniruddha Jathar

Department of Computer Engineering, Zeal College of Engineering and Research, Pune, Maharashtra, India

ABSTRACT

In today's world, the pace of development in technology is at full tilt, which leads to the creation of enormous amount of data. This data which is created at very high-speed has variety and exists in a very huge size is known as big data. Big data analytics can be used in the educational sector for the betterment and to tackle the issues which can be faced while working with the educational data. Even though big data analytics is an excellent way to deal with most of the issues, but in some cases some of the problems get neglected or get don't recognized in their pre stage. Which then reside for a long period of time and then lead to some critical issues. For that, this paper focuses on the different types of challenges that can be faced while utilizing big data analytics for the betterment of the educational sector. This paper also covers, various contributions of big data analytics that have been done in the educational sector, in this way making us able to focus on the abiding issues/problems.

Keywords: Big data; Big data for Education; Big Data analytics; Big Data Challenges; Higher education; Education.

I. INTRODUCTION

As the world around us is developing with an extreme rate of development there is a rapid increase of technology in every single sector like health and medical, business, social media, educational sector, etc. Also, today's world is revolutionized by the Internet by providing an extremely huge network and platform for every person, industries and organizations. As this technology gets utilized and also develops at the same time there is a generation of enormous amount of data through various sources and platforms. As the data is generated from various sources the data varies in the type having the heterogenous behaviour in the nature. The data which is generated might be in various forms like image, audio, text data or might be sensor generated data, logs, electronic mails, etc. This data is always generated with a high velocity in the real time world which is the major factor as compared to the size of the data. Also, the volume of this kind of data is very high which can be typically scaled into exabytes (EB). This type of data which has Variety, high Velocity of generation and exists in a huge Volume is referred as Big Data. According to McKinsey the term Big Data refers to *datasets whose size is beyond the ability of*

typical database software tools to capture, store, manage, and analyse. Big data has the characteristics which are known as 3Vs of big data. These three Vs are nothing but Volume, Velocity and Variety. The figure.1[4] below shows the 3Vs of Big Data.[1][2][3]
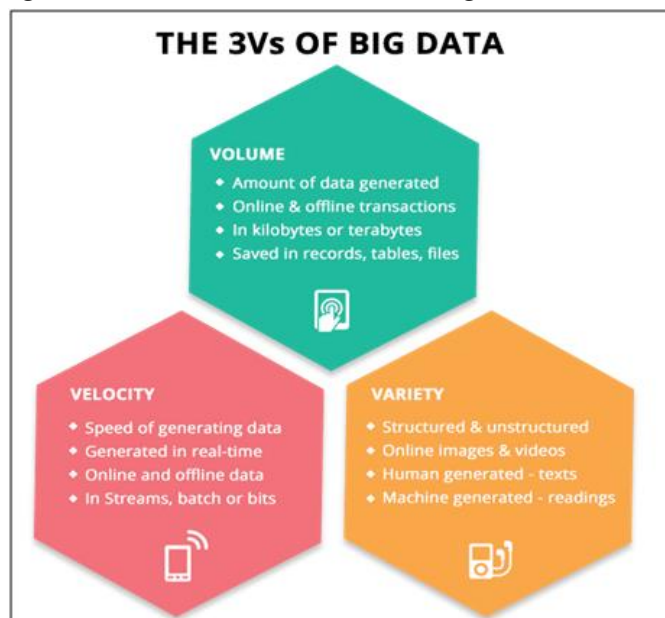

Figure.1- The 3Vs of Big Data

As the Big data has its own characteristics the challenges/issues emerging into it are also different and need different treatment to deal with. The challenges like computing and sharing, accessing the data, data integration, etc are major problems faced while dealing with big data [5]. As the educational data also matches with the characteristics of the big data, so big data analytics can be used for the purpose of extraction of educational analytics from the educational data [7]. Some of the applications of Big Data Analytics in the educational sector which are previously implemented and are being used for the educational betterment proposes are like enhancing the results of students, customized programs, better grading system, etc [6]. But while performing analytics also some of the challenges can be confronted. The challenges like storing data, proper handling of data, appropriate disposal of data, maintaining the security and confidentiality of the data, etc are some of the issues which may arise. some of the challenges can be easily tackled but some

problems reside for a long amount of time. These problems can be said as abiding problems, which stay for a long period of time because of not getting recognized in the previous stages or getting neglected or remain unfocused. As the part of abiding problems in the big data analytics have not been focused much, so this paper discloses some of the most important issues which are unrecognized while performing analytics and can be a major challenge later.

The remainder of the paper is structured as follows. The section 1 of the paper gives us the basic idea about the big data, big data analytics and challenges into the introduction. The section 2 of the paper discusses about the section I and section II of the Literature survey. The section I of literature survey presents the information about the general/basic and the long-term challenges/problems for the big data analytics which can be faced while performing analysis on the educational, student, organization data of the educational sector. The section II in the literature survey discusses about the various kinds of contributions by utilizing the methods of big data analytics, giving us an idea about the lack of attention in the different area. And the conclusions have been drawn in the final section which is section 3 of paper.

II. LITERATURE SURVEY

I. Challenges for big data analytics in the educational sector

There are different types of challenges while performing big data analytics for the intended purpose, in this case for the educational big data. As the data size of the big data is enormous and which varies in the type whether it can be student data, social media data, etc. The one of the most important factors is Handling of the educational data in an effective manner [7]. It should be ensured that the data is being stored and disposed in a proper manner. If the data has been not handled properly it can result into some critical issues. The issues of data alteration, data loss or also some issues like access of data by

unauthorized users may arise and which might have a serious impact on the environment. "Data handling requires adequate planning, development of procedures, and training and supervision of research staff to ensure that data is stored, archived or disposed off in a safe and secure manner that preserves the integrity of research data as well as simplifies data management." [11].

As the data handling is an important factor for big data analytics there is always the need of experts and the skilled persons. Which takes us to our next challenge, which is the lack of talent and skilled peoples. The people having the in-depth knowledge of big data analytics and are able perform some critical and deep analytics whose work will be loyal to the value which can be extracted from the data. There is currently a huge demand of data analysts and data scientists in the field, which gives us a full view about the enormous size of data being generated and needed to be processed currently. Also, there exist more issues like, the amount of money required to hire these highly intellectual persons.

The lack of analytics skills and unavailability of internal resources which are required to take the advantages of the pulled-out insights [8]. Considering the case that, even if there is an availability of skilled and talented peoples, there remains a barrier of unable to pull out the expected value if there is not an availability of required equipment and infrastructure to work with. Such issues can easily result into poor quality of data extraction. In that case there will not be an effective use of the given resource (in this case educational data).

As big data is an advanced technology there is always a need of experienced staff and a skilled person [9]. As being a new and advanced technology in the industry, growing with high speed, big data analytics requires a deep study of the field to understand the complex things in big data, but only deep study is not sufficient in the field while working on important and sensitive data it requires skills and a lot of experience to be able to understand the working of

actual process, so there is always a need of experienced person in the field, who in actual knows how to wield the huge amount of data and which actions to be perform on the right time when required. Other challenges were also identified like, way of dealing with the new types of data, adaptation of new and more advanced analytics techniques [8]. As big data analytics is a new and daily emerging technology there is always a room to improve and due to that, there is always development going into the sector. And also, various kinds of problems also start to emerge daily. As per the rate of development it difficult to learn everything in such a short amount of time. So, this kind of problems can be considered the long-term problems which can exist until some type of the technology reaches its peak point.

The traditional data management done with the help of RDBMS is not able to handle the enormous amount of heterogenous data as the storing with this method requires more expensive hardware [10]. Similarly considering the amount of money required to hire field specialists to handle data, the hardware required to store the huge amount of data into a distributed as well as in a well-organized and easily accessible format also requires the expensive hardware. If we consider, in this case where the data which will be handled or stored is the Educational Data, which will be mainly handled by the organisations and the universities, due to this it can have a huge effect on the all-over infrastructure of the respected organisation/universities in the financial manner. There are several reasons to for the failure of RDBMS to handle the Big Data. Reasons like, the process of 'Scaling' is very hard to achieve because vertical scaling is limited and expensive, also the complexity of the query cannot be determined which is required for the achievement of the expected results. Combining of multiple queries is done in the process of pulling out the insights, as pulling out the insights requires the merging of data which is highly normalized which is present in database and as the normalization process requires complex queries and

combining these queries can slow down the overall performance of the machine and the slowness will result in to more requirement of time to calculate the expected result [12].

Some of the challenges in data storage are Data representation, data life cycle management, redundant data and also data confidentiality [10]. As, in this case the data we are dealing with is the valuable student data and the educational organisation's data, so the issues regarding confidentiality may result in leakage of confidential data which might affect the reputation of the organisations if the data will be misused. Also, if the data life cycle is not managed in a proper manner, then the problems like over collection of data, poor management of data and hoarding deprecated data may rise [13]. This type of problems may arise more frequently if there is a lack of well-designed architecture and a flexible platform which is also one of the major problems. Lack of comprehensive platform tailored for the educational institutions [13]. Planning, developing and implementing such kind of diverse platform itself is a time consuming and complex process. As, it includes all of the major factors like data gathering, processing, storing the data according to its sensitiveness which also requires planning and choosing the suitable architecture, and the last and most important factor is to provide insights and gain value by performing analytics on the educational data. As this process is interminable so it also takes a lot of time to optimize which makes this an abiding problem.

II. Contributions of Big data Analytics in the Educational sector

The paper discusses about how the big data analytics is able to benefit the higher educational sector and the business contributors in the sector just by using the fundamental capabilities of the big data analytics. The paper also states that the better absorption of knowledge and better performance of students can be expected, by developing the pre alerting system which will be based on the big data analytics and is needed to be deployed in an appropriate manner. The Research paper also includes an experimental study report, which gives us an idea about how the big data analytics can be used for betterment of higher studies, by studying the data of 309 postgraduate students which was collected through the LMS and EASI. The research issues the need of performing the analytics on the behavioural student data which is in the unstructured form. Also, by the use of Moodle Online Survey System for the collection of data more deeper and various types of analytics were performed. This research provided a well-planned architecture for the betterment and evolution of big data analytics in the educational sector, which will be beneficial to the students and the educational organisations [14].

As there was an absence of a flexible, well planned and well-designed platform, the authors of this paper contributed by developing a full-fledged platform for the educational sector which can be utilized for various purposes, which will benefit the educational organisations to drive and manipulate the educational data. The paper covers to major parts which consist of first designing the platform and the another one is focused on the implementation. As per the authors the platform is especially designed to tailor the educational data in real time, also having capability to monitor the effects caused due to decision making with the help of feedback loop, and also providing the platform which can be effortlessly utilized by the non-data scientist and the educational organization peoples. The platform was implemented with the help of cloud computing platform and using various architectures leading us to tackle number of problems related to vacancy of platform [7]. As the problem of absence of platform can be an abiding problem depending on certain conditions but as per the previously discussed paper that is a huge contribution in the field providing a solution to the one of the critical issues.

In today's world mainly the use of big data analytics in the educational sector is focused on, how the big

data analytics can be applied and can be utilized in the sector to earn maximum value possible. So, the applications of big data in the educational sector for the enhancement of teaching and learning process is discussed by the authors of this paper [15]. The paper is aimed for the provision of a teaching and learning model and also providing a clear view of benefits by the usage of big data analytics for the educational purpose also providing a big data model for educational sector. The paper also focused on the topic of two-way feedback for both students and teachers by measuring the whole process, resulting into betterment in the students learning and teachers teaching performance. According to the authors the use of survey and modelling approach has been done to make the big data analytics applications in educational sector more noticeable making us able to recognize the applications and potential of big data analytics.

As discussed int the previous paragraph, applications of big data analytics can be used for various kind of purposes to extract and utilize the resource(data) to gain the required value. The applications like Prediction of student's performance, Course Recommendation, Retention analytics and attrition risk and etc., have been discussed in the paper by the authors for better recognition of big data analytics applications. The paper is mainly focused on the usage of analytical tools which can contribute to Learning Analytics and Educational data mining [16]. By studying the paper, the problem of big data analytical applications getting unnoticed can be a key take away from this paper, which put forth another abiding problem which is under looked.

As pulling insights from the provided data itself is a difficult task and also one of the long-lasting problems, the stages like decision making which are performed after the insight pulling process are affected due to the previous steps if not done properly, so some of the factors which can be used to get better decision making are discussed in this paper by the authors. The paper discusses about the factors like sources from

where the data needs to be collected for better decision-making purposes and improving universities performance and factors also like choosing the better tools like HIVE and Apache Hadoop and techniques like descriptive analytics, diagnostic analytics, predictive analytics and prescriptive analytics for better decision making [17].

When the data is large in size it needs to be secured and handled very carefully and properly as it may contain sensitive data. In today's era even though data is being stored on clouds which helps deployment of big data but also brings some of the challenges related to security in big data. The paper discusses about many types of security issues which can be problematic in future. The paper gives and deep description about the education under the network environment also including issues like security of storage and access control, it also includes the strategies related providing safety and also provision of safety framework using the standard and also the supervision methods to ensure the security of the environment in which the big data exists [18].

III. CONCLUSION

In the educational sector Big Data Analytics holds and important role as it serves the sector in the major ways by usage of its various techniques and the number of its applications. These techniques and applications hold a lot of potential and can benefit the sector in more effective manner if used properly. The challenges which are responsible for the unachievable potential were discussed and noted. The problems which were residing into the sector form a long-time were put forth for the better recognition and to be more focused. Also, the problems which have their pre-existing solutions were also discussed for the optimization of the previously given solution according to the newly developed techniques and tools. The major abiding problems like lack of skilled persons, lack of experienced person, unavailability of tools, absence of an optimized and flexible platform,

etc., were majorly focused. The contributions were also included for the better understanding of the previously discussed platforms, problems, architectures and the issues faced while implementation and while performing big data analytics. Observing through the contrast between the highly overlooked and under looked challenges/topics the more focus was done on the under looked topics making us able to deal with the problems/challenges which were left out and resulting into taking another step towards the betterment into the educational sector with the help of Big Data Analytics.

IV. REFERENCES

[1]. Mohd, Ayesha. (2020). EVOLUTION OF BIG DATA AND TOOLS FOR BIG DATA ANALYTICS. Journal of interdisciplinary cycle research. 12. 309-316.

[2]. Sun, Zhaohao & Strang, Kenneth David & Li, Rongping. (2018). Big Data with Ten Big Characteristics. 56-61. 10.1145/3291801.3291822.

[3]. Roberta Pastorino, Corrado De Vito, Giuseppe Migliara, Katrin Glocker, Ilona Binenbaum, Walter Ricciardi, Stefania Boccia, Benefits and challenges of Big Data in healthcare: an overview of the European initiatives, European Journal of Public Health, Volume 29, Issue Supplement_3, October 2019, Pages 23–27, https://doi.org/10.1093/eurpub/ckz168

[4]. https://www.coforge.com/salesforce/blog/data-analytics/understanding-the-3-vs-of-big-data-volume-velocity-and-variety/

[5]. Mardis, Elaine. (2016). The challenges of big data. Disease Models & Mechanisms. 9. 483-485. 10.1242/dmm.025585.

[6]. Applications of big data analytics. Referred from-- Top 5 Interesting Big Data Applications in Education [2022] | upGrad blog

[7]. A. A. Munshi and A. Alhindi, "Big Data Platform for Educational Analytics," in IEEE Access, vol. 9, pp. 52883-52890, 2021, doi: 10.1109/ACCESS.2021.3070737.

[8]. Attaran, Mohsen & Stark, John & Stotler, Derek. (2018). Opportunities and Challenges for Big Data Analytics in American Higher Education- A Conceptual Model for Implementation. Industry and Higher Education. 32. 10.1177/0950422218770937.

[9]. Bamiah, Mervat & Brohi, Sarfraz & Bashari Rad, Babak. (2018). Big data technology in education: Advantages, implementations, and challenges. Journal of Engineering Science and Technology.

[10]. Big Data Storage and Challenges" M.H. Padgavankar et al "by S R Gupta - IJCSIT) International Journal of Computer Science and Information Technologies, Vol.5(2)2014http://www.ijcsit.com/docs/Volume%205/vol5issue02/ijcsit20140502284.pdf

[11]. Article on Data handling. Referred form- Data Handling (hhs.gov)

[12]. article related to challenges faced while using RDBMS. Referred from--Can RDBMS handle Big Data?. Introduction: This article points out… | by Harshit Dawar | Analytics Vidhya | Medium

[13]. Article on data life cycle. Reference-- Data Lifecycle Management (Definition and Framework) | Talend

[14]. S. Jha, M. Jha and L. O'Brien, "A Step towards Big Data Architecture for Higher Education Analytics," 2018 5th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), 2018, pp. 178-183, doi: 10.1109/APWConCSE.2018.00036.

[15]. A. Mohammed, S. Kumar, S. P. Singh and R. Prakash Sharma, "Enhancing Teaching and Learning in Educational Institutes Using the Concept of Big Data Technology," 2018 International Conference on Computing, Power and Communication Technologies (GUCON),

2018, pp. 1038-1041, doi: 10.1109/GUCON.2018.8674982.

[16]. S. Roy and S. N. Singh, "Emerging trends in applications of big data in educational data mining and learning analytics," 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, 2017, pp. 193-198, doi: 10.1109/CONFLUENCE.2017.7943148.

[17]. M. A. Segooa and B. M. Kalema, "Improve Decision Making Towards Universities Performance Through Big Data Analytics," 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD), 2018, pp. 1-5, doi: 10.1109/ICABCD.2018.8465132.

[18]. T. Fupeng, "Research on Education Big Data Security Strategy under Network Environment," 2021 IEEE 6th International Conference on Big Data Analytics (ICBDA), 2021, pp. 19-22, doi: 10.1109/ICBDA51983.2021.9403030.