Application of Big Data Analytics for the development of eLearning

Deepshikha Aggarwal
Professor, Department of Information Technology, Jagan Institute of Management Studies, Delhi, India

Abstract: In today's world technology is dominating every field. The education sector is taking complete benefit of technology by utilizing it in classrooms and outside. The success of eLearning is one of the best examples of usage of technology for education. One of the most amazing aspects of eLearning is that it provides the facility of anywhere and anytime learning and is constantly evolving to accommodate the learning needs of the learners. Due to the emerging technologies and improved learning management systems, the possibilities are endless for the future of eLearning. A concept that has emerged as a highly useful tool in the eLearning community is "big data". In this paper we have explained what exactly is big data and how does it benefit the eLearning industry. The impact of Big Data is seen everywhere including the businesses and education and it has become indispensable in the 21st century world. The applications are used in almost every sector and field of human endeavor. Big Data has been an integral part of eLearning for a number of years. Big Data gives the opportunity to examine learner preferences and thus improve the overall eLearning strategy. It can help in tracking timing, device usage, and determining the strengths and weaknesses of an eLearning program. Big data analysis gives the ability to customize the eLearning experience and improve the elements that may be falling short of expectations. Big data analytics has direct application for the development of an effective eLearning programme.

Keywords: Big Data, eLearning, big data analytics, lifelong learning, technology enhanced learning.

I. INTRODUCTION

Big Data refers to the large amount of data collected from multiple sources every second. It is not just a large volume of data but also includes different new types of data. It is the data that is too large, complex and dynamic. Therefore, it is difficult for any conventional data tools to capture, store, manage and analyse the big data. But due to advancements in technology, this data can now be analysed and used by various sectors such as healthcare, government, retail, manufacturing and e-Learning.

The term Big Data has actually emerged from the open source community as they are the ones trying to find faster and more scalable solutions to store and process the large amounts of data, instead of using traditional data warehousing approaches. Big Data analytics is the process of analysing the large volume of data that is the big data. The business organisations use data not only for day to day transaction processing but to make strategic decisions as well. Business organisations have been storing and analysing large volumes of data since the advent of data warehouse systems in the early 1990s. But the kind of data is changing, and the changes are leading to complexity in managing and analysing of data. Most businesses today run on structured data (numbers and categories). However, this does not reflect the complexity on the nature of available corporate data and their untapped hidden business value. The actual data generated by organisations is unstructured with a variety of formats such as text, video, audio, images, etc. This unstructured data goes into the corporate data warehouses. The term 'Data warehouse' refers to a central repository of data or a centralised database that can be utilised for better decision making. Big Data describes data that is too big and moves too fast, thus exceeding the processing capacity of conventional database systems [14]. It also covers innovative techniques and technologies to capture, store, distribute, manage and analyse larger sized data sets with diverse structures.

Big Data refers to emergent suit of technologies that can process mass volumes of data of various types at faster speeds than ever before [15]. This conceptualisation of Big Data was echoed by Forrester defining Big Data as "technologies and techniques that make capturing value from data at an extreme scale economical." The term economical suggests that if the costs of extracting, processing and making use of data exceed the advantages to be collected, then it is not worth indulging in the process.

Big Data is identified by a number of fundamental characteristics [10]. Key among them are:

• Volume—large amount of information to store, process, and transfer, analyze and present.

• Velocity—relating to increasing rate at which information flows within an organization

• Veracity refers to the biases, noise and abnormality in data. It also looks at how data that is being stored, and meaningfully mined to the problem being analyzed.

• Variety—referring to data in diverse format both structured and unstructured.

• Verification—refers to data verification and security.

• Value—most importantly, has the data been utilized to generate value of the insights, benefits and business processes, etc. within an organization

Douglas [10] in the Gartner's report proposed the three of the most common properties of Big Data. The report made

three fundamental observations: the increasing size of data, the growing rate at which it is produced and the cumulative range of formats and representations employed, proposed threefold definition encompassing the "three Vs" (Volume, Velocity and Variety).

There are also other properties of Big Data such as data validity, which refers to accuracy of data, and volatility, a concept associated with the longevity of data and their relevance to analysis outcomes, as well as the length required to store data in a useful form for appropriate value-added analysis.

II. ROLE OF BIG DATA IN eLEARNING

ELearning is the process of learning using any kind of device such as a computer, laptop, tablet, smartphone, etc. in this kind of learning, the learning content is made available to the users and the users or learners can use this content for learning as per their convenience [11]. ELearning is also referred to as anytime and anywhere learning. The number of learners in eLearning is huge as there is no restriction of availability of physical classrooms and other infrastructure. Since there are very large number of people involved, the data to be handled is also large. In the process of eLearning when the learners interact with the content in their course, they produce data which is very large due to the large number of users and therefore comes under the category of Big Data. We can collect and track this data through learning management systems (LMSs), social networks and other media. This data is about how the learners interact with different aspects of the eLearning course that they are taking. The Big Data collected through the different sources is processed and analysed to understand the behaviour of the learners and design the eLearning content according to the demands of the eLearning industry.

III. BENEFITS OF BIG DATA FOR e-LEARNING

Big Data can play an important role in understanding the learning patterns of the learners more effectively because it aims to track a learner's experience in an eLearning course and this data is then analysed to study the behaviour of the learners and thus leads to the development of appropriate eLearning environments. By examining the digital footprints of the users, we can track the learner's journey throughout the entire learning experience.

By analysing the big data collected from a LMS (Learning Management System), we can see which portions of the course were too easy and which were difficult. This data can be collected by giving the facility of a small assessment after each module and depending on how the learners perform in these assessments; we can judge how well the learner understood the content [6]. If majority of learners are able to score well, it indicates that the content is easy to understand. Whereas, if more learners are scoring less, the content needs to be modified.

There are other aspects of the course content that need to be tracked and analysed. We have to track the pages revisited often, sections recommended to peers, learning styles preferred and the time of day most learning activities take place.

The most important factor for success of eLearning is the flexibility that the learners have related to the time and place of learning. Big Data may change the way in which the eLearning programmes are designed by enabling developers to personalize courses to accommodate the individual needs of the learners. This is possible because the learning patterns of the users enable the designers to collect the data and analyse it so that the eLearning content is provided as per the user requirements. This process of data collection and analysis allows e-Learning professionals to continuously raise the standards for effective and exceptional eLearning courses. Various learning management system like Lectora Express – The Easy LMS or CourseMill LMS help to easily track, record and analyse the Big Data from the learners.

IV. BENEFITS OF BIG DATA TO eLEARNING PROFESSIONALS

There are a variety of benefits that big data offers to eLearning professionals.

It allows eLearning professionals to understand how the learners are utilising the information and which learning needs are being mostly addressed. Since the concept of big data is based upon collection of data, therefore in case of eLearning also the data is collected to indicate the usage pattern of users for the eLearning resources [2]. This data can be analysed to find patterns such as to determine how effective is a real time scenario than a text-based problem solving activity in teaching a concept. This will facilitate the future design of eLearning content as per the need of the user.

Big data enables the eLearning professionals to identify the areas that need to be modified within an eLearning course or module. The data can be collected on the basis of time spent by the learners on each module. This data is analysed to check if multiple learners are taking an excessive amount of time to finish a particular module. If this happens it may indicate that the module needs to be improved in order to make it more easily understandable for the learners. But this decision in not just based on the time spent as there may be different reasons such as leaving the system on and not actually reading by the user can lead to spending more time on a particular module. But when we use big data we analyse huge number of values so there is a better probability of obtaining accurate results as compared to conducting a study on few users.

The big data analysis provides information on which eLearning modules are visited the most and lets the eLearning course designers create more courses which are preferred by the learners. The data collection is enhanced through social learning where we try to analyse which eLearning modules or links are shared with other learners via social media like Facebook [4].

With the latest learning management systems, data is received almost immediately, rather than having to wait for

long periods of time to receive assessments. This means that eLearning professionals can begin implementing changes or utilizing the data to fine tune their eLearning strategy right away.

Based upon the analysis of the eLearning patterns of the users, eLearning professionals can predict which factors enable the learners to excel and what makes them struggle through the course. This way, they can develop their eLearning courses in such a way that the learners can achieve the best possible results.

V. IMPLEMENTATION PROCESS OF BIG DATA IN eLEARNING

The process of implementation of big data analytics in eLearning is depicted in Fig 1. Big data analytics of the eLearning data not only leads to the development of an effective eLearning system but also makes way for continuous improvements. The implementation of big data in eLearning is done in the following stages:

A) Data Collection
The first stage in the process is collection of data. To have a good analysis, the data collection plays a major role. The sources of data have to be identified and then the data is collected. In the world of technology there are numerous resources of data on the internet and the right sources have to be identified as per the requirements of the analysis. The sources can be the learning management systems, web logs, social networking data and so on. In data analytics, what goes into the analysis decides the effectiveness of results of analysis. Therefore it is very important that the right kind of data is collected and sent for analysis.

B) Data Compilation
The second stage is compilation of the collected data as per the goals and objectives of the analysis. In this stage the data is organised and cleaned. When the data is collected there are many fields in the data which may not be relevant to the concerned analysis. Therefore it is required that the data is organized as per the results required. Data cleaning is another important stage in big data analytics as the dirty data may lead to inaccurate results.

C) Data Analysis
The third and main stage is actual data analysis to find trends and patterns in the eLearning resource usage of the learners. These trends and patterns thus analysed, help the eLearning designers and content developers to modify the eLearning content and create a platform for effective learning. The main advantage of eLearning is the tailored learning feature where the learning can be done as per the needs and pace of the learner. The analysis of data leads to the results that indicate the learning patterns and usage statistics of the eLearning platforms and

enable the developers to design the eLearning system tailored to the needs and demands of the learners.

D) Feedback
The work of the eLearning providers does not end after the development of the eLearning platform. There are continuous updations and modifications that have to be carried out on a regular basis. The developers of the eLearning platform have to look for continuous feedback from the users and identify the problem areas. As the learners are using an eLearning platform, they come across different issues such as some may find the pace of the course faster or slower, others may find the language difficult and so on. The eLearning course design is in the continuous process of improvement and therefore is able to provide a learning environment as per the needs of the individual learners. This is what makes eLearning an effective tool as it is accommodating the changing needs of the learning community. To be a successful eLearning platform, the continuous modifications are very important.

Fig 1: Implementation of Big Data Analytics in eLearning

VI. BIG DATA ANALYTICS IN MOOCs

An emerging face of online learning is MOOCs. A massive open online course (MOOC) is a model for delivering learning content online to any person who wants to take a course, with no limit on attendance.

MOOCs are based on video lectures, multiple-choice quizzes, discussion forums and document sharing [16]. Lessons are delivered on a regularly on a fixed timeline, and students complete their lessons in the given timeline only. The main difference between eLearning and MOOCS is that MOOCs are based on a fixed timeline. The students in a MOOCs course study the content of the course and take up assignments and have the option to share and discuss their comments and views in forums or social media networks. Also the teachers make the learning process interactive by posting questions, assignments, comments and communicating with students on a regular basis [17].

The reports of various analyses conducted on MOOCs data complains about the problems such as low completion rate of the courses, lack of interaction, falling motivation levels of the learners, etc. [18]. To have an effective system of learning this kind of data is collected for analysis. The collection and analysis of data about student actions on online learning environments makes the job easier for educationists and researchers to understand the learner behaviors and handle the various concerns.

Online learning environments such as MOOCs provide a rich source of knowledge mining opportunity. By logging mouse clicks, forums activity, quiz performance, login frequency, time spent on tasks and tracking videos interactivity, data analytics researchers can build an enormous amount of data logs. This database of information, if interpreted appropriately, can help the developers of MOOCs courses to understand the learning patterns of the students and design the course content as per the learning needs. Thus data analytics in MOOCs is highly useful.

As compared to the other fields of application of big data analytics, many of the goals of an online learning scenario cannot be directly measured. But if the correct data is collected and analyzed, the results can help in improvement, control and development of a successful online learning platform.

VII. CHALLENGES FOR IMPLEMENTATION OF BIG DATA

The widespread popularity of big data has brought with itself a set of challenges too. The education sector is a comparatively new user of big data and therefore is still in the process of learning to handle these challenges. The major challenges faced by the eLearning community are:

A) Security and Privacy

With usage of big data analytics in the eLearning scenario, there is a continuous data gathering from the learning activities of the students. There is a continuous research on the subject with concerns from parents over the privacy of their children's data.

While this is a serious concern, the positives in using big data override other issues. Using big data can help improve student performance because it leads to a targeted approach to teaching.

It's important to take the security and privacy issues seriously in the educational organizations, especially in deciding who would access and process the data gathered. Hacking and other cyber threats are not the only ways in which data thefts take place. In some cases the internal employees may be stealing private data. The educational institutes that are promoting eLearning have to design various strategies to maintain trust among the staff, students, and their parents.

B) Lack of organised data

Data analysis is done for eLearning with the aim of improving the overall experience of the learners and to improve the output of the programme. An effective analysis requires an effective collection of data. The procedures adopted in collecting data should therefore follow a proper protocol. There have to be guidelines on what kind of data should be collected and how should it be organised for an effective analysis. There should also be a proper backup so that critical data is not lost a cyber-attack or power outage.

Also, we have to decide from whom we choose to collect the data from and the there has to be a data cleaning mechanism to assure data quality.

C) Correct Usage of Data

Big data analytics is not just collecting and analysing data but it has to be done in a way that the results are directly beneficial for the field of application. A strategy has to be rolled out to map how the big data analytics will be used for helping the learners and improving the process of eLearning.

The results of analysis can help improve the educational procedures of the school or college. It can also help parents of the students to keep a track of the learning of their children, including things they can do at home to improve the learning ability.

D) Data Protection and Sharing

The educational institutes going for the implementation of big data analytics for the development of their eLearning programmes need to invest in efficient security tools to protect their data. While automation systems can provide data backup, but we also need ways to prevent serious cyber issues like hacking, denial of service attacks, or ransomware.

The eLearning department has to work in collaboration with the IT department, to include processes such as encryption and multi-step authentication. This assures the data isn't easily accessible by unauthorised persons even if the learners and teachers are logging into the eLearning environment from separate computers.

There has to be a proper control mechanism to guide with who all the data is shared with in the organization. Sharing

it with parents of students can educate them on how to best approach their children's learning needs, and sharing it with school staff and teachers helps in refinement of the curriculum.

VIII. CONCLUSION

Big data plays an important role in the education sector. It offers extraordinary prospects for educators to reach out and instruct learners in new ways. Big data analysis of the learners' data gives the course designers a deeper understanding of students' learning experience, and thereby helps them to evaluate the strengths and weaknesses of the eLearning system. In this paper, we have summarized different ways in which big data can impact the eLearning scenario.

With big data in the education sector, it is possible to monitor student actions and when we are collecting data on a very large scale as in big data, the analysis is more accurate. There are different metrics on which the learners' are analysed such as how long they take to answer a question, which sources they use for exam preparation, which questions do they skip, etc. these type of questions can help the eLearning course designers to modify and update the content of the eLearning courses. The data can be used to customize the eLearning modules for individual student as per the requirement and learning speed.

The use of big data for education can also be extended to what is called 'blended learning' which is a combination of online and offline learning. In this case the students are attending classes in college or school and they can also access the learning content online through different eLearning modules. This gives them the opportunity to follow classroom teaching and also the ability to learn at their own pace through eLearning. They have the possibility for offline guidance by professors as in the case of MOOCs that are developed and delivered around the world now and have become extremely popular in the education sector.

Big data has the potential to revolutionize the learning industry in the coming years. Smarter students will have a positive impact on organizations and society. Therefore, it's time we embrace big data in the education sector.

IX. REFERENCES

[1] https://elearningindustry.com/big-data-in-elearning-future-of-elearning-industry.

[2] https://elearningindustry.com/8-steps-create-develop-effective-mobile-training-strategy.

[3] https:// elearningindustry.com/directory/ software-categories/learning-management-systems.

[4] C. Pappas, "Big Data in eLearning: The Future of eLearning Industry," (2014).

[5] https://elearningindustry.com/8-steps-create-develop-effective-mobile-training-strategy.

[6] Daniel, B., Big Data and analytics in higher education: Opportunities and challenges. British Journal of Educational Technology, 2015. 46(5): p. 904-920.

[7] Ramesh, V., P. Parkavi, and K. Ramar, Predicting student performance: a statistical and data mining approach. International Journal of Computer Applications, 2013. 63(8): p. 35-39.

[8] Baepler, P. & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. International Journal for the Scholarship of Teaching and Learning, http://digitalcommons.georgiasouthern.edu/ ij-sotl/vol4/iss2/17

[9] Baer, L. & Campbell, J. (2011). Game changers. EDUCAUSE. http://net.educause.edu/ir/library/pdf/pub72034.pdf

[10] Douglas, L. (2001). 3D data management: controlling data volume, velocity and variety. Gartner Report. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

[11] Norris, D. M. (2011). Seven Things You Should Know About First Generation Learning Analytics. EDUCAUSE Learning Initiative Briefing.

[12] Yan, J. (2013). Big Data, Bigger Opportunities. www.meritalk.com/pdfs/bdx/bdx-whitepaper-090413.pdf.

[13] Basak, A. (2013). A Lap Around Big Data With Microsoft HD Insight. http://architects.dzone.com/articles/lap-around-bigdata-microsoft

[14] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. et al (2011). Big data: the next frontier for innovation, competition, and productivity. Retrieved October 7, October 30, 2014, from http:// www.mckinsey.com/ Insights/MGI/ Research/Technology_and_Innovation/Big_data_The_next_frontier for_innovation

[15] Yan, J. (2013). Big Data, Bigger Opportunities. Retrieved from: www.meritalk.com/pdfs/bdx/bdx-whitepaper-090413.pdf [27, July 2014].

[16] Khalil, H., & Ebner, M. (2013). Interaction Possibilities in MOOCs–How Do They Actually Happen. In International Conference on Higher Education Development (pp. 1-24).

[17] Khalil, M., & Ebner, M. (2015). Learning Analytics: Principles and Constraints. In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications (pp. 1326-1336).

[18] Khalil, M. & Ebner, M. (2016). What Massive Open Online Course (MOOC) Stakeholders Can Learn from Learning Analytics. Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy. Springer.