

# Big Data Analytics Model for the Education Sector



Bhavin D. Shah, D. B. Choksi

**Abstract:** *With an array of innovative technologies, the 21st century has struck at our door. Many sectors are embracing the implementation of a revolutionary technology called Big Data Analytics. The education sector has lately jumped on the bandwagon. With the new evolving education technology capable of generating Big Data, stakeholders have firmly started adopting the use of Big Data Analytics in the education sector. This paper proposes a model for Big Data Analytics depicting various required components, for the education sector. Various sources of education data provide the necessary input to the model and model produces useful results in the form of the output using various tools and technologies of Big Data ecosystem. The outputs are going to be consumed by various stakeholders of the education sector. The paper also describes the concept of Big Data, new technologies influencing the education sector and explains why Big Data Analytics is going to be core technology in analyzing the data and taking the world of education to new heights. We have also listed beneficiaries and benefits.*

**Keywords:** *Big Data Analytics, Educational Data Mining, Higher Education, Learning Analytics*

## I. INTRODUCTION

The 21st century has arrived with a number of disruptive innovations. Internets of Things (IoT), Hyperloop, 3D printing, Blockchain are some prominent arrivals. Data are becoming essential vitamins for every human being in routine life. People have started realizing that the impact of data is hidden within data itself and that belief has started a quest for the new technologies for extracting useful nutrients from the commodity named data. Big Data and related technologies have emerged as a disruptive force in the data analysis world. Big Data Analytics (BDA) has already built its rock-solid foundation in some sectors like retail, health care, finance, media, and entertainment. It has started its presence felt in other sectors also. The education sector is not an exception. Stakeholders in the education domain have turned their face optimistically towards this promising technology for the unprecedented analytics input to enable them to remain competitive in this fast-changing academic world. Education is being transformed by the use of digital technology.

Education has occupied a prominent place in the UN 2015 Millennium Development Goals. All countries' development plans also include education as a top priority. The reason is that education is a compelling indicator and a powerful Instrument of human progress [1]. The U.S. Dept. of Education observes different ways to use data of an online learning system for improving instruction in its National Education Technology Plan [2]. Nowadays education is no longer confined within the school or college boundary. Learning continues through lifelong education, vocational courses, online education and training on the job also. Through education, any human being connects (generates/consumes) with an abundance of data whether raw or processed in early years of his life as he undergoes many hours of classwork and homework 5-6 days/week for years. This causes educational data sets to be broad and deep compared to other sectors such as e-commerce where customers spend a few minutes each day or users busy with social media for a few hours daily. This data is now an area of focus and we are trying to capture it due to the presence of technology called Big Data Analytics.

Though Big Data has been more of a priority in scientific, industrial and public sectors than the education sector, it has the potential to transform the educational landscape radically. By applying Big Data in the education sector, we can get analytics results that we could have never imagined. It is going to be instrumental in opening new frontiers in this sector. We can bring a drastic change in the learning cycle having an infant child and matured worker as two extremes. Researchers use a lot of various tools to understand student quality and how to develop the course structure. Instructors use Social Networks Adapting Pedagogical Practice (SNAPP) to understand various students' blogs. This software can decide about students' interest in the course based on the interaction of the students. Learning Management System is another tool that is very popular now a day to handle the education system [3]. Virtual classroom is another important tool which is gaining popularity. All the distance programs are based on it. All these tools generate a lot of data that can be stored and analyzed. This paper is about the use of Big Data Analytics in the education sector. Section 2 discusses the basics of Big Data. Section 3 enlightens evolving education technology. Section 4 elaborates on the relation of Big Data with the education sector. In Section 5, we propose a model for Big Data Analytics in the education sector. We have also listed stakeholders and expected benefits of analytics they can reap.

**Revised Manuscript Received on October 30, 2019.**

\* Correspondence Author

**Bhavin D. Shah\***, CSE Dept. Nirma Institute of Technology, Ahmedabad, Gujarat, India. Email: shahbhavind@gmail.com

**D. B. Choksi**, PG Dept. of Comp.Sci., Sardar Patel University, Vallabh Vidyanagar, Gujarat, India. Email: dbchoksi@yahoo.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

## II. BIG DATA REVOLUTION

Big Data is a big term in the context of definition as there is no consensus on the standard definition of this term. Everyone has his own view. To put in simple terms, Big Data represents that set of data that is outside the scope of management of traditional relational database technology. Though relational is not the only database technology used so far, it's almost four decades of dominance has made it a synonym of database technology. In Big Data related technology, the way data is collected, stored and analyzed is significantly different from the traditional database technology. Basically the term "Big Data" has surfaced due to volume, velocity, and variety (3 V's of Big Data) of the data. Volume refers to the quantity of data, velocity refers to the speed of data generation and variety refers to the type of data. These 3 V's play a vital role to decide whether data fall in the Big Data category or not. Massive data gathering caused by the proliferation of a plethora of different digital devices has provoked data analysts for unprecedented analysis of data to help the stakeholders in various sectors worldwide. The speed with which this data is generated compels processing and analyzing of the data to be faster than the data generation rate. Data processed so far was mainly of structured nature. But nowadays data generated is structured, unstructured or hybrid. Traditional database technologies are exhibiting signs of frustration to process these Big Data. This has necessitated the search for alternative data processing technology. It is worth to note that the scope of the term Big Data is not restricted to the type of data, it also covers volume and velocity of data generation also.

## III. EVOLVING EDUCATION TECHNOLOGY

Major revamp in teaching methodologies and subsequently the entire education process across the private and public educational institutes worldwide is attributed to the emergence of disruptive digital education technology. Centuries long teaching is now transforming from analog (blackboard) to digital (smartboard). These new sets of technologies have proved to be an enabler for collecting a vast amount of digital data. Here, we elaborate on some of these path-breaking innovations.

*E-learning*: Though its impact on students is still an immense debatable issue, both developing and developed nations have embraced this new type of learning technology. With the proliferation of personal digital assistants coupled with ubiquitous access to a fast and reliable internet connection, this new form of learning has started to be used globally. Education is no longer bound to local and within campuses. Distance learning is now a reality by exploiting e-learning. Lectures and exams are conducted in virtual classrooms.

*Learning Management System (LMS)*: LMS is a software application for the administration, documentation, tracking, reporting, and delivery of educational courses or training programs and development programs [4]. The content format can be text, audio, video, etc. Accessibility, durability, reusability, interoperability, adaptability, etc. are advantages of LMS [5]. Modular Object-Oriented Dynamic Learning Environment (Moodle) is a widely accepted open-source example of LMS.

*Smartboard*: It is a modern technology-enabled, the interactive black (white) board. It helps teachers to be more effective and productive and enhances the learning experience of the students. Numbers of modern schools have adopted this innovative apparatus. Its maintenance is low and it is environment-friendly too as it can save tons of papers.

*Virtual Reality (VR)*: This technology combines the best of personal and online education. This is a very realistic and cost-effective way to visit unthinkable scenarios [6]. Microsoft's HoloLens allows students to visualize the functioning of the human body in 3-D. Though the penetration of VR is low, it is going to be a great game changer.

*Augmented Reality (AR)*: Using AR, one can add a layer of information to a live view. Google glass is an example of AR. It is having a profound value addition in the education domain.

In addition to the above technologies, bar code scanners, Optical Mark Recognition (OMR), eye trackers, IoT, AI are worth to just mention that are going to shape the future of education.

## IV. BIG DATA IN EDUCATION

The key question of the discussion is - what is the relation between education and Big Data? Is the education sector a suitable candidate that can leverage the power of Big Data Analytics? The answer is positive. The use of advanced technologies discussed in the previous section in the higher education domain generates data that fall in the Big Data category. Confined to data analytics with the help of emerging data analysis tools and techniques, these data are proving to be a gold mine. The use of Big Data in education can be classified into two areas namely Educational Data Mining (EDM) and Learning Analytics (LA). Generally, EDM looks for new patterns in data and develops new algorithms and/or new models, while learning analytics applies known predictive models in instructional systems [2]. Educational Data Mining is a discipline that is emerging. It is related to developing methods that will explore the educational data and with the help of these methods, we can gain insights in academic settings and can understand the students better [7].

EDM can help in following [8]:

- Find out the student activities that can result better learning and help to achieve good marks
- Finding out effective topic sequence for a particular student
- Identify features of the online learning environment that aid in better planning
- Find out the actions indicating satisfaction and engagement

LA can provide guidance for:

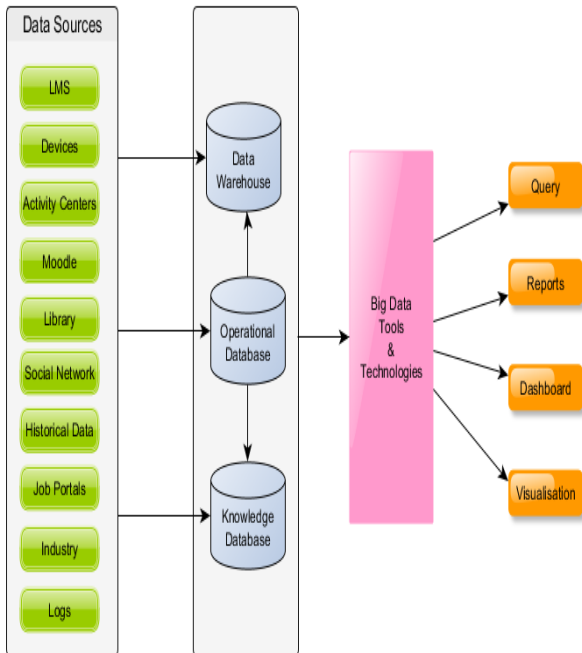
- Deciding time when students are ready to move to the next topic
- Figuring out the grade that a student can bring without anyone's intervention
- Work out risk conditions when a student will not complete a course

- Deciding intervention of counselor for a student

## V. BIG DATA ANALYTICS

### A. Model

Fig. 1 shows the input-processing-storage-output model for Big Data Analytics in the education sector.



**Fig. 1 Big Data Analytics Model**

The processing part comprises of Big Data tools and technologies that can be of different levels of complexities depending on desired outcomes. In an educational setup, a variety of data sources generate Big Data - the data having attributes like volume, velocity, and variety. These sources can be Learning Management System, devices like sensors, cameras, RFID, eye trackers, etc. Different activity centers (mostly on campus) can also input vital data of the students' activities – both curricular and extra-curricular. A platform like Moodle can also be an important data source. University social network can cater to data like posts, messages, comments to the analytics system. Library data can give useful insight into the students' learning process. Timing of the issue of books and other study materials possess the immense potential and provide vital clues during analytics. Historical data sources represent students' past educational data provided sharing is agreed upon by the organizations that own that data. Job portals provide important data about current industry demands and expected level of knowledge to be possessed by the job-seeking candidates. The industry can provide data pertaining to the professional progress of university students after completing the course successfully. There can be more addition to the list of data sources. The whole data can be classified as long term data, transactional data and value-added data that are destined for a data warehouse, operational database, and knowledge database respectively. Some data stores can be part of the Big Data ecosystem of databases. Data may be directly stored from source to database in its pure form or it can undergo some transformation process before it getting stored in the database.

Big Data landscape [9] is quite complex having a plethora of technologies for different operations to be performed. Depending on the system's need, data analysts will select a set of technologies to carry out the important analysis task. Finally, outcomes can be in terms of query outputs, reports, dashboard or sophisticated charts/diagrams produced by an array of rich visualization applications.

### B. Stakeholders

Fig 2 presents stakeholders associated with Big Data in the education sector.



**Fig. 2 Stakeholders in the education domain**

Here is a comprehensive list of stakeholders (beneficiaries) and few expected benefits they can gain with the use of BDA.

- 1) Students
  - Personalized learning experience
  - Exclusive personalized recommendations of study material like video/text/references to gain a better understanding of the topic/subject
  - Improve learning methodology
  - Better Career decisions
- 2) Faculty Members
  - Improve teaching skills
  - Factors analysis for student's poor performance
  - Predicting the future performance of the student
  - Assess the effectiveness of different learning tools and techniques used in teaching
  - Focus on research/self upgrading
  - Analysis of the gap between student's learning and performance
- 3) Parents/Guardians
  - Insights of student's performance and behavior
  - Early warning indicators
  - Micro-level assessment of class activities

## Big Data Analytics Model for the Education Sector

- 4) Universities/Colleges/Schools/Administrators
  - Increase student retention rate
  - Reduce dropout rates
  - Build new innovative scalable education models
  - Improved national and international collaboration
  - Launch new long term and short term niche courses
  - Design dynamic curriculums that reduce academic-industry gap
  - Better assessment of affiliated institutes/ colleges
  - Promotion of research
  - Track student career chain
  - Better Accreditation Ranking
- 5) Employers/Industry
  - Procuring skilled employees
  - Reducing training time and budget
  - Establish a strong academic pipeline
- 6) Government
  - Help in Education reforms
  - Increase in higher education literacy rate
  - New course/academic setup decisions
  - Improvement in rural areas
  - Reduce child labor
  - Narrowing education gap between developing and developed nations
- 7) Publishers/Authors
  - Measure the efficacy of learning material through content analytics
  - Anticipate need of students and faculties
  - Offer customized and more relevant rich material
- 8) Online Education Providers
  - Deliver personalized, world-class education on a global scale
  - Increase the reach
  - Use concept-level learning analytics to improve learning experiences and outcome

### C. Barriers

Irrespective of how big and promising any technology is, if barriers to its implementations are not addressed then it will not be able to deliver its promised best. The following are some of the barriers that can potentially slow down the realization of the power of Big Data Analytics in education.

General:

- Scarcity of skilled workforce
- Issues related to privacy and security of data
- Poor data quality
- Selecting the right data for analysis
- Seamless connectivity

Education Domain-specific:

- Major education policy reform challenge
- Resistance to digital transition
- Incremental nature of innovation
- Creating a data-sharing network[10] – Data system's interoperability
- Inadequate technology infrastructure access

- Difficulty in establishing success benchmarks for performance measurement
- Uniform software platform

## VI. CONCLUSION

Big Data Analytics is no longer a hype. It has become a reality and widely accepted software technology worldwide. Successful implementation of this trend-setting innovative technology has given phenomenal and remarkable results in various sectors across all the levels of management. Compared to other sectors, the education sector has remained less explored by this modern age inevitable tool called “Big Data Analytics”. Increased focus on education by developing countries coupled with the revolution in the educational approaches, powered by Big Data tools and technologies have sparked a phenomenal momentum in this domain. Big Data tools and technologies are attaining a level of maturity and stability. A new model of Big Data Analytics depicting input, processing, storage and output elements is proposed in this paper. Application of the proposed model on the repository of educational data has a tremendous potential of delivering unprecedented useful insights to the key decision-makers and other stakeholders. Though there are certain barriers, if they are properly handled and their impacts are diluted, this new age analytics is going to change the landscape of education in the world. The ultimate beneficiary will be the student community.

## REFERENCES

1. Education and Skills 2.0: New Targets and Innovative Approaches World Economic Forum eBook [http://www3.weforum.org/docs/GAC/2014/WEF\\_GAC\\_EducationSkills\\_TargetsInnovativeApproaches\\_Book\\_2014.pdf](http://www3.weforum.org/docs/GAC/2014/WEF_GAC_EducationSkills_TargetsInnovativeApproaches_Book_2014.pdf).
2. Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief: UD Department of Education Office of Educational Technology <http://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf>
3. Big Data in Education - Saptarshi Ray Lecturer Great Lakes Institute of Management [http://www.researchgate.net/publication/259443597\\_BIG\\_DATA\\_IN\\_EDUCATION](http://www.researchgate.net/publication/259443597_BIG_DATA_IN_EDUCATION)
4. [https://en.wikipedia.org/wiki/Learning\\_management\\_system#cite\\_note-ellis-1](https://en.wikipedia.org/wiki/Learning_management_system#cite_note-ellis-1)
5. Long, Phillip D. (2004). Encyclopedia of Distributed Learning. Thousand Oaks: SAGE Publications, Inc. pp. 291–293. doi:10.4135/9781412950596.n99
6. <https://www.u-planner.com/blog/disruptive-education-technologies-poi-sed-to-change-higher-learning>
7. International Educational Data Mining Society <http://www.educationdatamining.org/>
8. <https://edtechmagazine.com/higher/article/2013/02/personalized-learning-sweeping-college-campuses-courtesy-big-data-infographic>
9. Big Data and AI Landscape 2018 [http://mattturck.com/wp-content/uploads/2018/07/Matt\\_Turck\\_FirstMark\\_Big\\_Data\\_Landscape\\_2018\\_Final.png](http://mattturck.com/wp-content/uploads/2018/07/Matt_Turck_FirstMark_Big_Data_Landscape_2018_Final.png)
10. Big Data for Education: Data Mining, Data Analytics, and Web Dashboards <http://www.brookings.edu/research/papers/2012/09/04-education-technology-west>

## AUTHORS PROFILE



**Bhavin Shah**, completed his Bachelor of Engineering in Computer Engineering from Gujarat University and Master of Computer Application from Indira Gandhi National Open University. He is currently pursuing a Ph.D. from Sardar Patel University. His interest areas are Big Data Analytics, Graph Database. He has more than 22 years of industry and academic experience. He is currently serving as Assistant Professor at Nirma Institute of

Technology, Ahmedabad, India.



**D. B. Choksi**, received a Ph.D. in Computer Science from Sardar Patel University in 1996. He has been serving the Post-Graduate Department of Computer Science of the Sardar Patel University since 1989. He is a recipient of the prestigious Young Scientist Award from the Indian Science Congress in the year 1992. He has guided several Ph.D. scholars. He has held various key positions in University and academic institutions. He has contributed research in several areas including

system software, computer networks, and distributed computing.