# Emergence of Unstructured Data and Scope of Big Data in Indian Education

Dr. S S Kolhatkar
Marathwada Mitramandal College of Commerce
Pune, India

Mr. S P Kolhatkar
Tech Mahindra
Pune, India

Mrs. M Y Patil
Marathwada Mitramandal College of Commerce
Pune, India

Mrs. M S Paranjape
Université Paris-Dauphine
London, United Kindom

*Abstract*—The Indian Education sector has grown exponentially in the last few decades as per various official reports[22]. Large amount of information pertaining to education sector is generated every year. This has led to the requirement for managing and analyzing the structured and unstructured information related to various stakeholders. At the same time there is a need to adapt to the dynamic global world by channelizing young talent in appropriate domains by cognizing and deriving the knowledge about individual student preferences hidden within the vast amount of education data. The derived knowledge is about getting finer information related to courses, facility and quality of institutes, etc and also analyzing unstructured educational learning resources present in the form of multimedia data. Also, the desire to cater to stakeholders for decision making related to courses, admissions, career planning etc has accentuated big data analytics.

Various MIS or ERP systems handling structured information for educational applications exist in order to aid in the administration and managerial process. These systems are useful in customizing software application as per institutes or courses, generating various customizable reports and aiding the decision making process related to institutes. The need for storing, maintaining and analyzing unstructured information related to multimedia content online has generated a need for big data and data analytics. This paper is about the emergence of unstructured data, comparison of features provided by relational databases and big data and to identify the scope of big data in the Indian education sector.

*Keywords—Big Data; Indian Education; Unstructured data; Big Data Analytics; Comparative of Big data and Relational Database; Scope of Big Data based Applications*

## I. INTRODUCTION

The Indian education sector is at a threshold of cognizing that instead of being reactive, it needs to  be proactive by leveraging technology and analytics. The development and deployment of various software systems is aimed at providing availability, reliability and transparency of information to various stakeholders like trusts, faculty members, students etc. of the education sector. At the same time, the use of right technology enables businesses to integrate information across all departments, enabling all stakeholders to have access to one version of the truth. The existence of Apereo[28] and Open Hatch[27] is a success for education sector based software applications.

The education system generates, maintains and analyzes large amount of data generated through various sources. This data is related to academic, non-academic, research, learning resources, examination, admission, training & placement etc. The nature of such data as shown in Fig. 1 is varied in nature as given below.
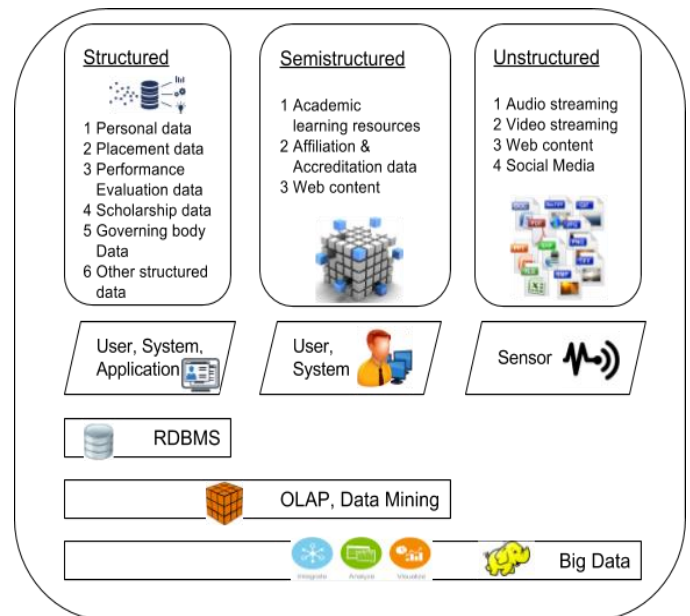


Fig. 1.   Type of data and its application area

### A. Structured Data

It refers to kinds of data with a high level of organization, such as information in a relational database. Transactional software systems work on structured data mostly for the purpose of querying and maintaining. In education system, the student academic information, scholarship (& benefits) information, placement data, examination data, administrative information etc; is identified as structured data which is maintained and queried for customized reports.

## B. Semi-structured Data

It is a form of structured data that does not conform with the formal structure of data models associated with relational databases or other forms of data tables, but nonetheless contains tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. Software systems for education sector that use comma separated values (CSV), eXtensible Markup Language (XML) and JavaScript Object Notation (JSON) technologies have semi structured data.

## C. Unstructured Data

It often includes text and multimedia content. Examples include e-mail messages, word processing documents, videos, photos, audio files, presentations, webpages and many other kinds of business documents. While these sorts of files may have an internal structure, they are still considered unstructured because the data they contain does not fit neatly in a relational database. The learning resources and social networking sites are examples of unstructured data. The video and audio streaming of classroom is also unstructured data. Our paper focuses on learning resources and multimedia data for unstructured data.

Such developing disparity in education data points to the fact that software systems pertaining to education sector also need to accommodate and reflect the change in their functioning.

## II. COMPARISON OF RELATIONAL DATABASE AND BIG DATA

The software systems that have been or are being developed, use data storage as a basic need. But due to the change in nature of data (structured or unstructured), the software systems have also evolved from traditional systems to more complex systems. The increasing demand for various software systems enable the investment of resources to facilitate for appropriate outcomes. Technology is a powerful tool that businesses invest in to move with greater speed and certainty for competitive edge. The underlying data storage in education sector basically uses the relational databases or the big data. Let us consider features of data storage in education sector. This is discussed below.

## A. Relational Databases

A relational database management system (RDBMS) is a database management system (DBMS) that is based on the relational model as invented by E. F. Codd. Many of the databases in widespread use are based on the relational database model. RDBMSs have been a common choice for the storage of information in new databases used for financial records, logistical information, academic data, personnel data, and other applications. A few of RDBMS are Oracle Database, Microsoft SQL Server, MySQL, IBM DB2, IBM Informix, etc. which are used to store the structured information by defining the inter relationships between data. The relational database features along with their interpretation for education sector is as given below.

- Use Provides data to be stored in tables - Structured data related to the education sector can be normalized

and stored. The structured data is related to student information, course / programme information, staff (teaching and non-teaching) member information, examination information, academic monitoring information etc.

- Persists data in the form of rows and columns - The structured data can be stored as atomic values with the cell as the identification.

- Provides facility primary key, to uniquely identify the rows - Structured data is inter related across various entities, departments etc; for which the concept of primary key and foreign key is used successfully to correlate as well as differentiate for identification.

- Creates indexes for quicker data retrieval - The facility of index creation as provided by the relational databases is used for faster information access or retrieval which is beneficial given the large amount of academic and supporting records.

- Provides multi user accessibility that can be controlled by individual users - Various users having individual role and responsibility for education software systems need to be given authenticity and access rights accordingly so as to ensure data management at different management and security levels.

As seen in the above 5 features, the nature of data is structured. The unstructured data like social networking, learning resources, multimedia data etc which is generated on a large scale by the stakeholders, is beyond the scope or management of the RDBMS. Relational database management systems and desktop statistics and visualization packages often have difficulty handling big data. The work instead requires "massively parallel software running on tens, hundreds, or even thousands of servers". Let us consider the nature and type of data used in education in current times with respect to big data.

## B. Big Data

Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate. In Indian education sector the data that can be qualified as big data is related to learning resources, images and social networking, along with structured and semi structured data. It can be described by the following characteristics:

- Volume - It represents the quantity of generated and stored data. The size of the data determines the value and potential insight and whether it can actually be considered big data or not. The semi structured and unstructured data is large in case of Indian Education sector as the use of ICT has been well assimilated in the functioning of the sector.

- Variety - It means the type and nature of the data. The multimedia data represents a variety in the type and nature of data. It includes text, presentations, spreadsheets, images, audio, video, tweets, posts, blogs etc. which forms a large part of the learning resources in the education sector. The data stored in database for

software systems implemented for this sector can also be considered as an important type of data.

- Velocity - In this context, the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development represents the velocity. The demand for online storage and functioning of the education sector, has resulted in large amount of data generation (for example Google Apps) and data processing (for example online tests or feedback) to an equally large audience.

- Variability - This characteristic represents the inconsistency of the data set can hamper processes to handle and manage it. A class that represents an entity in this sector may have its own distinct data set but would need to be stored as a set nevertheless which is not efficient in case of relational databases. The inefficiency is due to memory wastage or processing bottleneck.

- Veracity - The quality of captured data can vary greatly, affecting accurate analysis. Educational data and metadata has limited veracity as the method or means of data capture are standard and reliable.

The emergence and popularity of big data exists due to various factors but among them the expectations for accountability to stakeholders through reports, analysis or results; is the greatest. The other reasons identified for its importance are demands for evidence to guide and support decision making, finding metrics that matter to institutions and individuals and the need for a technology[3][6][7][10][14] platform that provides a means to the end. Big Data relates to extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour[2][16][17][19][25][32] and interactions. In general, there are four main sources of big data in Indian education as given below :-

- Users : The education sector has student as the main users of various software applications. They are the main sources for generating a large amount of data related to admissions, attendance, examinations, placement etc. The students access and generate e learning resources. They also participate and engage themselves in social networking where large data exists in the form of tweets, posts etc. The teachers are important users of software systems and contribute in generating data related to curricular, co-curricular and extracurricular activities. The decision makers or the managing body is also an important user who actively frame and design policies or strategies based on the different MIS reports. Apart from these three users, there are other administrative staff members, who ensure the smooth functioning of education institutes by maintaining and generating data about the social and legal aspects of the various courses along with trustees and board members.

- Application : The need to enforce a control and reward mechanism over various universities, colleges, institutes etc; there are governing bodies which maintain the data that has been generated by users (as discussed above) using their customized applications. The data may be maintained on cloud or company server but definitely it provides a basis for informative and analytical reports that are crucial to policy and decision making. There are a variety of applications used in education, such as Massive Open Online Course - MOOC (provides interactive user forums to support community interactions among students, professors, and teaching assistants - http://mooc.org), Moodle (Moodle is a learning platform designed to provide educators, administrators and learners with a single robust, secure and integrated system to create personalised learning environments - https://moodle.org), Google Apps ( is a brand of cloud computing, productivity and collaboration tools, software and products developed by Google - https://www.google.com/edu/products/productivity-tools/) etc are used to enhance the teaching learning experience.

- System: The need for customized information at various managerial and administrative levels in education is different. Customized applications have been developed for individual purposes such as DOMO[29] (Domo bills itself as a business intelligence tool, and the goal of the product is to bring a business together with its data, and to be able to measure multiple data sources against one another. It automatically pulls in data, in real time, from spreadsheets, social media, on-premise storage, databases, cloud-based apps, and data warehouses. Data points are presented on a customizable dashboard so employees can view the information in a way that is easy to digest and relevant.), ERP-Microsoft Dynamics® AX 2012 (It is Microsoft's enterprise resource planning software products), MIS - EMIS, online learning - Lynda.com, etc.

- Sensors: Sensors in buildings enable tracking of students and the time that they spend in the classroom, dormitory, cafeteria, or in the library. The effectiveness of their instructor can be partly determined by analysis of student sentiment. Sensors are increasingly providing critical information gathered on devices. Data critical to research might be gathered directly from sensors in semi-structured form.

The above four data sources are mostly captured from devices such as mobiles, microphone, reader/scanner, science facilities, program / software, camera and social media. In order to implement software system for big data, there is big data analytics and integration platform as shown in Fig. 2 consisting of data integration, data management and data analytics.
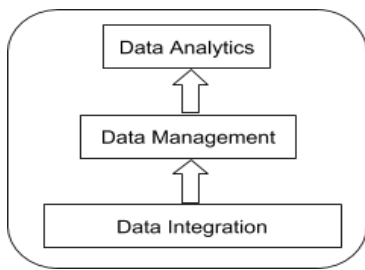
Fig. 2.   Big Data Analytics and Integration Platform

*1) Data Integration*

Data integration is the combination of technical and business processes used to combine data from disparate sources into meaningful and valuable information. The education sector generates and maintains data which is flat file data, data in database etc. A complete data integration solution gives trusted data from a variety of sources. This process follows a defined series of steps to manage the integration of data as shown in Fig. 3.
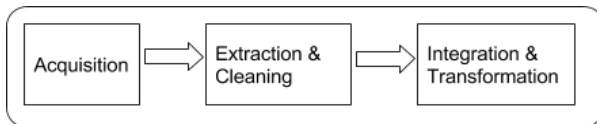


Fig. 3.   Data Integration

*a)* Capturing Data in numerous forms (structured, semi-structured and unstructured) from various sources requires a standardized approach to be maintained across various tools including security, metadata and look and feel.

*b)* Data Cleansing and Quality Management requires identifying and repairing inaccurate, incomplete and redundant data to maintain consistency on quality of the availability of the data. It also includes Profiling, Parsing and Standardization, Generalized Cleansing, Matching, Monitoring and Enrichment.

*c)* Transformation of data depending on the situation. Based on the nature of the source of the data the extracted information is transformed regardless of the format, complexity or file size.

*2) Data Management*

Big data management is the organization, administration and governance of large volumes of both structured and unstructured data. The goal of big data management is to ensure a high level of data quality and accessibility for business intelligence and big data analytics applications. There are a few end user tools for such data management. They are Big Data Portal, Reporting, MS Office Integration, Mobile BI, Ad hoc query, Dashboards etc.

*3) Data Analytics*

Big data analytics is the process of examining large data sets to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information through modeling, analysis and interpretation as

shown in Fig 4. It allows the forecast of data models and accelerate decision making through realization of Big Data solutions. There are various analytical tools as mentioned below.

*a)* Online Analytical Processing (OLAP)[30]: Creating a multidimensional data store from a vast pool of detailed data. The data offered are often summarized data using fast query and calculation performance delivered in real time. For example educational statistical data.

*b)* Predictive Modeling and Data Mining[31]: Identifying variables that allow enterprises to forecast scenarios with aid of mathematical models. These models are often integrated into the reporting, dashboarding or OLAP phases. For example classification or clustering patterns in education.

*c)* Sentiment analysis[32]: It uses natural language processing techniques to read and interpret the meaning of the textual information. Era of customer-focused business organizations has led to the proliferation of sentiment analysis. For example polarity of student's sentiments about courses, organization, individuals and events.

*d)* Advanced visualization and visual discovery[33]: ADV has emerged as a significant technique to extract knowledge from data. It enables data exploration with interactive visualization tools like FLOT, ProfitBricks etc. For example interactive charts, panning and zooming for popularity of courses and performance rates
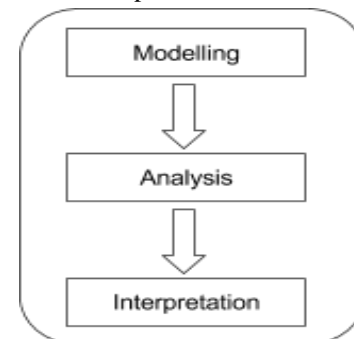


Fig. 4.   Data Analytics

The characteristics and architecture of big data is difficult to implement. There are challenges like communication between educationists and software system companies, differences in educational system and practices etc. These may lead to challenges in analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying, updating and information privacy. Accuracy in big data may lead to more confident decision making, and better decisions can result in greater operational efficiency, cost reduction and reduced risk for the stakeholders. At the same time, the use of relational databases for online transaction processing (OLTP) and big data for online analytical processing (OLAP) need to coexist for the education sector. Table 1 shows comparison of relational and big data on various factors which are both essential but for their individual purpose.

TABLE I.     COMPARISON OF RELATIONAL DATABASES AND BIG DATA

| Factors | Relational Databases | Big Data |
|---|---|---|
| Data Types and Formats | Structured | Structured, semi structured, unstructured |
| Data Integrity | High | Low |
| Data Storage | Stored in rows and columns Rows contain all the information about one specific entry/ entity and the columns contain varied data points | Encompasses varied host of databases Each database has different data storage model The important ones are: document, graph. Key-Value and columnar. |
| Storage Volume | Gigabytes to Terabytes | Terabytes, Petabytes and beyond |
| Scalability | Scale up with more powerful HW | Scale out with additional Servers |
| Data Processing Distribution | Limited or None | Distributed Across Cluster |
| Economics | Expensive Hardware & Software | Commodity Hardware & Open Source software |
| ACID Compliancy (Atomicity, Consistency, Isolation, Durability) | The vast majority of relational databases are ACID compliant. | Varies between technologies but many NoSQL solutions sacrifice ACID compliancy for performance and scalability. |
| Scalability & Cost | Vertical scaling More data requires bigger servers which increases the cost Difficult and time consuming to scale the database | Horizontal scaling Multiple servers for easy manageability and cost efficiency |
| Schemas and Flexibility | Schema is rigid Each column must be decided before the data entry Addition of columns to existing database needs to be handled | Dynamic in nature Information can be added anytime. The rows and columns can be empty |
| Example | On Line Transaction Processing Applications | Big data analytics |

The Table 1 attempts to make comparison of relational database and big data based on the type and volume of data and the software applications that handle this storage. Both types of storages are required as they serve different purpose viz OLTP and OLAP. Both ie relational database and big data handle large volume of data though big data takes an upper hand. But the important difference is the type of data that they handle as explained in the next section.

## III. USE / TYPE OF DATA STORAGE IN EDUCATION SOFTWARE SYSTEMS

The demand for fast and available information for various education system reports, requires the development and deployment of MIS / ERP software systems designed for education sector. Such systems require the use of relational databases for the capture, storage and maintenance of data that accumulates over a period of time. The advent of cloud computing[34] along with the development and progress of the underlying IT infrastructure has greatly aided in the development and implementation of such advanced and flexible software systems.

At the same time, the need for adapting to a dynamic world with the aid of information / data; has tremendously increased. The demand for fast information instantaneously is rising and it has to be fulfilled for competitive advantage. This demand is facilitated by deploying ERP / MIS systems specifically for education sector where data is structured and stored using relational databases.

### A. RDBMS in Education software system

The education data is structured as well as unstructured. The structured data is managed by well-defined software systems that rely on the principles of RDBMS. Here the focus is on data capture for clarity and transparency; data storage for long term purpose of information authentication; and lastly data presentation for the purpose of reporting. The reports are of various types depending on the level of organizational or management structure or hierarchy. The bottom management generates transactional reports that focus on the regular activity. The middle management generates MIS reports that are more directed towards planning and organizing of resources. The top management software system is designed so as to take decisions about resources, results and ramifications.

The legacy systems are well suited for structured data, but there needs to be cognizance of increase in amount of unstructured data. Such unstructured large amount of data may be considered as big data.

### B. Big Data in Indian Education

The unstructured data is associated with details related to social networking, learning resources, multimedia data etc. The generation and management of big data is an important scenario in most of the sectors. There are many companies and software systems that harness the benefits of big data in architecting the course of their businesses. For example, Knewton's pioneering approach to adaptive learning draws on each student's own history, how other students like them learn, and decades of research into how people learn to improve future learning experiences.

Currently, the underlying data storage is used for reporting and analytical purposes. There are numerous statistical reports with various aspects; few of which are as discussed below.

*a)* Course / College wise Reports to represent the educational institutes distribution based on region or facilities along with its intake capacity

*b)* Student Admission Reports to statistically know about course, gender, category, region, etc of admitted students

*c)* Teaching Staff Reports that mention the qualification and contribution of teachers

*d)* Infrastructure Reports to represent the existing and developing campus details

*e) Examination Reports to know the performance of students and counsel for career development*

There are several other reports of education sector which together help various governing bodies to identify and channelize young talent for human resource development. Along with this well-defined and structured data, the Indian Education is witnessing an undefined epoch in social networking and student resources which include learning resources, images, posts etc. At the same time, there is a need for contemplating on the aptitude, inclination and perseverance from the students' viewpoint to cope up with the global challenges of a dynamic world. Such a cross environment integration and analysis lays the foundation for data analytics and the initiation and implementation of big data in education.

## IV. Software Applications in Indian Education

The education sector practices and promotes the use of various enterprise systems for the management and development of academics; thereby facilitating the stakeholders with accessible and available information. The concept of cloud computing / storage has tremendously aided in developing, implementing and deploying such software systems. As discussed above, most of the software systems handle structured data whereas few provide support for semi structured or unstructured data. This makes the need to develop software systems to work on big data, very important. Big Data that has been created by taking data from traditional systems will have structured data.

In an article by Benjamin Herold [34], there is a mention of analysing various multimedia inputs like classroom video streaming or performance scores with respect to other activities. Large amount of information is captured and stored. There are few software applications that are currently involved in the analysis of the data (mostly structured as it is generated from a software application implementing relational database) for decision making. Furthermore, the generation of unstructured data and its subsequent analysis, demands the development and deployment of new software applications. This unstructured information is what can be studied and analysed to predict or confirm about various aspects based on the structured and unstructured data. These aspects are as given below.

- Human Behaviour can be analysed based on students' gestures, reactions or responses obtained through classroom video streaming to know about individual aptitude / abhorrence for particular subjects / areas, interaction with others etc. Such analysis guides the

parents in understanding their wards and realizing their interests, better. Generally, this analysis acts as a guide to parents during the child's formative years.

- Other way round, study of a student's attitude can be done in relation to student's marks to show how poor performance causes bad behaviour. Teachers, students and parents together discuss corrective steps for improvement. Here, the analysis is done on examination data which is structured.

- The identification of subject expert / popular teacher can be done by analysing teacher performance of teacher in terms of students' behaviour in class or reciprocation of students for a particular teacher. This analysing requires classroom streaming ie unstructured data.

- Inter college / university / education boards analysis can be done on structured examination data to know about its correlation with education medium, region, syllabus, inclination of population etc which is unstructured data.

- The aptitude and examination data ie structured data can be used to assist the students while choosing various further courses along with understanding current and future industry trade requirements which needs to be weighed and quantified. Such analysis will channelize the students to be practical professionals in terms of skill sets and salary expectations thereby aiding them to choose right course and right direction. It will also enlighten the industry and colleges in terms of finding industry institution gap. Likewise, well directed efforts can be made so as to propel the student knowledge as each campus requirement and outputs will be stored as structured data.

The above given aspects mention the use of structured as well as unstructured data for analysis purpose that can be used as conclusion or as corroborative evidence. The implementation and use of big data, therefore becomes advantageous in different situations and its scope is given next.

## V. Scope of Big Data

The software applications in Indian education are based on the current need and their architecture is based on the type of data storage needed to support it. There is large amount of academic data, performance evaluation data etc which is generated and maintained by current software applications; whereas personal data like posts, photographs etc, as well as formal data like notes, ebooks, online resources etc. need to be maintained and analysed for benefit of various stakeholders. Big data provides the infrastructure to maintain such unstructured data for analysing and finally visualizing, which can be classified as student engagement analysis, predictive analysis and sentiment analysis; as given below.

### A. Improving Student Engagement

A projected application of big data is in the area of adaptation of business intelligence for improving student

engagement. Student engagement is a vital antecedent to student achievement and organizational success. Big data analytics is used to identify the potential risk of disengagement based on the data inputs from online and offline resources. For example, Students could be asked to scan their id cards before joining classes, seminars and other events in the institution. Frequency of accessing virtual learning environment, library attendance, visits to the information desk could be other sources of data acquisition. Timely monitoring and interpretation of student engagement behaviours leads to potential actions to eliminate the impact of disengagement.

*B. Predictive Analysis*

Coupled with the predictive analytics, Big data anticipates to contribute to education by finding solutions to deal efficiently with bottleneck subjects, create means to advise students and colleagues accurately about their career inclinations and for forming a customized student-specific learning package that consistently influence students. Semi structured data from sensors is ingested by NoSQL databases and analyzed using predictive analytics at lower cost and more effectively. The next generation of students has smart phones and other access to devices connected to distant institutions. Following are the four applications of Big Data's role in education:

- Big Data driven Policies - Policymakers with live information on the quality and impact of education policies will be faced with the power to make information-loaded decisions. Such decisions will pertain to varying dimensions of higher education including finance, enrollment, choice of college, and career inclination of students. Big Data will allow policymakers to form a system of policies that would sync objectives of the educational institutions with available resources to produce intended results.

- Safety of Big Data - Gradually, student data will also shift to cloud services, to allow unfussy sharing and coordination of admission and transfers. As more information enters the cloud, there will be a need to set up walls that determine what kind of information and how much of it can be accessible to various stakeholders.

- Big Data will expand through Collaboration - Since Big Data maintains and manages unstructured data, institutions would have to move into the cloud age of collaboration thereby catering to needs of different types of students, recognizing the educational resources available collectively.

- Injecting Meaning into Big Data - There needs to be an inclusion of Big-Data positions like 'Predictive Analyst' within staffing and a subsequent rise in innovation-based job titles. Issues like economic affordability, dropout rates, retention and broadcast of content through new means (like Online Open Courses) have taken the front seat in the Big Data mission.

*C. Sentiment Analysis*

Sentiment analysis may also be used on education data such as online video streaming data to know about the impact of infrastructure, lectures, teachers etc so as to design customized learning courses. This type of analysis exists for social media content where the input data is in the form of tweets, posts, images etc.

Analytics can help in correlate attendance with scores to identify the target scores and related minimum number of classes required for schools to track data on the performance of students and teachers. However in the absence of dedicated and structured process, the data is not stored for a substantial duration to generate meaningful insights. Social media analytics is another emerging area of application. Important aspects of students learning styles, behaviour and preferences can now be gauged from formal groups that the educational institutes may have on social media platforms.

## VI. SUMMARY / CONCLUSION

The success of any software application or system is as a result of feedback from users and its subsequent improvisation based on current and future trends. The advent of unstructured data in the human life and its subsequent analysis for better understanding of the situation, makes the implementation of big data for any sector as desirable and beneficial. Our contribution in this paper is the discussion related to differences in various types of data being generated from different sources with reference to the Indian Education system along with the discussion of differences as observed in big data and relational data. We have identified the need for big data in Indian education sector along with its scope. We suggest that further research on the performance of predictive and sentiment analysis will greatly benefit organizations that are planning for implementing big data based software systems.

REFERENCES

[1] J.Fan, F. Han, H. Liu "Challenges of big data analysis", National Science Review, 1 (2) (2014), pp. 293–314

[2] R. Feldman, "Techniques and applications for sentiment analysis", Communications of the ACM, 56 (4) (2013), pp. 82–89

[3] Yanqing Duan & et al, "Big data in higher education: an action research on managing student engagement with business intelligence", University of Bedfordshire Repository, 2013.

[4] Xinguo Yu & Shuang Wu, "Typical Applications of Big data in Education", International Conference of Educational Innovation through Technology, 2015

[5] Peter MIchalik, Jan stofa & Iveta Zolotova, "Concept Definition for Big Data Architecture in Education System", IEEE 12th international Symposium Applied Machine Intelligence and Informatics, SAMI January 2014, Slovakia.

[6] Ling Cen, Dymitr Ruta, and Jason Ng, "Big Education: Opportunities for Big Data Analytics", IEEE International Conference on Digital Signal Processing (DSP), July 2015.

[7] Said Rabah Azzam, Ylber Ramadani, "Reforming Education Sector through Big Data", IEEE International Conference on Cloud Computing and Big Data Analysis, 2016.

[8] Robin G. Qiu & et al, "A Big Data Approach to Assessing the US Higher Education Service", 12th International Conference on Service Systems and ServiceManagement (ICSSSM), 2015.

[9] Mang Chen & et al, "The Positioning and Construction of Education Ecosystem Base on Big Data", 2nd International Conference on Information Management (ICIM), 2016

[10] M.M.M.A. Riffai & et al, "The Potential for Big Data to Enhance the Higher Education Sector in Oman", 3rd MEC International Conference on Big Data and Smart City (ICBDSC), 2016.

[11] Yuri Demchenko & et al, "Instructional Model for Building effective Big Data Curricula for Online and Campus Education", IEEE 6th International Conference on Cloud Computing Technology and Science, 2014.

[12] Shaoying Li and Jun Ni, "Evolution of Big-Data-enhanced Higher Education Systems", Eighth International Conference on Internet Computing for Science and Engineering, 2015, 978-1-5090-0454-6/15 $31.00 © 2015 IEEE DOI 10.1109/ICICSE.2015.53

[13] "Improving Higher Education Performance with Big Data Architect's Guide and Reference Architecture Introduction", Oracle Enterprise Architecture White Paper, April 2015.

[14] B.Tulasi, "Significance of Big Data and Analytics in Higher Education" International Journal of Computer Applications, (0975– 8887) Volume 68 – No.14, April 2013.

[15] Anurag Bhatt & Manish joshi, "Statistical Analysis of Impact of Social Networking Sites on Present Technical Educational Environment", Volume 4, Issue4, April 2014, ISSN: 2277 128X.

[16] Enock Kanyesigye et al "Sentiment Analysis of Reviews Using Hadoop" Vol-2 Issue-2 2016 International Journal of Advance Research and Innovative Ideas in Education

[17] Manisha Shinde-Pawar "Formation of Smart Sentiment Analysis Technique for Big Data" International Journal of Innovative Research in Computer and Communication Engineering

[18] Jon Tupitza, "Introduction to Windows Azure HDInsight", 2014 JTIT Consulting, LLC

[19] Ellen Wagner & Joel Hartman, "Welcome to the Era of Big Data and Predictive Analytics in Higher Education", SHEEO, Colorado

[20] Saga Briggs,"Big Data in Education: Big Potential or Big Mistake?", Innovation Excellence

[21] Mohd Ujaley, "Data Analytics in education sector to see high growth", Express Computer, November 6, 2015

[22] Government of India, Ministry of Human Resource and Development

[23] Educational software From Wikipedia

[24] Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments Ben Williamson Journal of Education Policy Published online: 29 Apr 2015

[25] "Big Data and Visualization: Methods, Challenges and Technology Progress" Lidong Wang et al in Digital Technologies, 2015, Vol. 1, No. 1, 33-38, Science and Education Publishing

[26] "An Odd Couple: Pairing Big Data and Behavioral Science?" Sahana Rajan April 2016

[27] "Technical Guide to Development and Documentation" by OpenHatch

[28] Apereo Documentation at https://www.apereo.org

[29] DOMO https://www.domo.com/industries/education

[30] Shirin Mirabedini , Seyedeh Fatemeh Nourani,"The Research on OLAP for educational Data Analysis" , International Research Journal of Applied and Basic Science, 2014, ISSN 2251-838X / Vol, 8 (2): 224-230

[31] Amirah Mohamed Shahiria, Wahidah Husaina , Nur'aini Abdul Rashida , "A Review on Predicting Student's Performance using Data Mining Techniques", The Third Information Systems International Conference, Volume 72, 2015, Pages 414-422

[32] Sunghwan Mac Kim,Rafael A. Calvo, "Sentiment Analysis in Student Experiences of Learning", Educational Data Mining 2010, The 3rd International Conference on Educational Data Mining, Pittsburgh

[33] Data Visualization and Discovery for Better Business Decisions by David Stodder (http://www.pentaho.com/sites/default/files/uploads/resources/data_visualization_and_discovery_for_better_business_decisions.pdf)

[34] Cloud Computing for Standard ERP Systems : Reference ERP Framework and Research Agenda. Schubert, Petra; Adisa, Femi. Koblenz: University Koblenz-Landau, 2011.

[35] http://www.edweek.org/ew/articles/2016/01/13/the-future-of-big-data-and-analytics.html