



## Big Data Mining in Education Sector: Critical Issues and Current Status in Indian Context

**Bhavin D Shah\***

Som Lalit Inst.of Computer Applications  
Gujarat University, Ahmedabad  
Gujarat, India

**D. B. Choksi**

G H Patel P.G. Dept. of Comp. Sci.  
Sardar Patel University, Vallabh Vidyanagar  
Gujarat, India

**Abstract**— *Information and communications technology is revolutionizing education sector. Data about students' online activities are collected in large volumes and analysed to produce and recommend actions that improve outcomes. Being a largest democracy, India is also trying to join this e-learning era. In terms of the number of Internet users, India stands third in the world. India has immense potential to use big data and mining in education sector. However, there are some critical issues that are blocking fast adoption of this technology. In this paper, we discuss critical issues that are not allowing India to leverage the result of educational data analytics. We also discuss initiatives taken by government and corporate, current status and stakeholders in current digital learning revolution age.*

**Keywords**— *Big Data, Data Mining, Online Education, e-learning, Data Analytics*

### I. INTRODUCTION

The impact of ICT in all sectors including industry has been profound in India. Education sector is also under high influence of ICT now. Pedagogical methodologies today are being transformed drastically with the use of technology. Big data is playing a key role in it. Big data in education sector is associated with collecting vast amounts of data from the digitized activities of students, parents, faculty and staff, transforming it into information, and producing or recommending actions aimed at improving institution outcomes[11]. Indian education system - especially higher education - must leverage innovation in technology to help all stakeholders. We should explore new opportunities that are knocking the doors. Big data and analytics have an immense potential to make education instrumental in economic growth of India. Gartner has highlighted Big data as one of the top 10 strategic technologies impacting education in 2015[11]. Applying data mining techniques to educational data for knowledge discovery is significant to educational organizations as well as students. Knowledge driven data supports educational decision support system[19]. Big Data mining can provide personalized learning experience to students. It can help faculty in assessing students' progress. Universities, institutes, publishers, authors, parents and government are also other stakeholders who can be greatly benefited with this technology. But there are some critical issues that need to be addressed. This paper attempts to discuss critical issues of Big Data mining in Indian education sector. It also discusses initiatives taken by government in the direction of embracing new technology to transform the way education is imparted in the country.

### II. CRITICAL ISSUES IN INDIAN CONTEXT

To benefit from Big Data mining in education sector, we need to collect huge volumes of digital data from digitized (online) or non-digitized (offline) activities of students, parents, faculty and staff, institutions, universities. Switching to digital learning in India, needs radical change in the whole education process. There are number of serious barriers that need to be addressed. Some of them are discussed here.

#### ***Absence of rich educational data set***

Educational data mining requires rich data warehouse to be ready. In India, there is a very little focus in this direction. India's higher education system is the world's third largest in terms of students, next to China and United States[4]. But educational institutes are not still having clear cut data warehousing strategy for storage of educational data. The data which is required for analysis may not be available on storage devices and if it is available then may not be in enough volume or structured form. This will make data mining ineffective. Online education is in very primitive stage at universities level. As a result we have non-availability of data of students' online activities. Some private players are offering online courses, but enrollment level of these courses is very less which restricts generic research.

#### ***Internet penetration and infrastructure***

The success of online education and subsequent data mining depends on use of Internet. In India, Internet connectivity and related infrastructure in rural areas is very poor. Though the growth of Internet users in India is phenomenal (number wise third in the world with approximate 243 million users in 2014), it is approximately only 19% of total population. So it is long to go to make Internet ubiquitous in India. India's share of world population is 17.5% while share of world

Internet users is only 8.33% [5]. Among the share of activities social networking is with 25% share followed by services 23% and entertainment 11% [14]. To increase the penetration of Big data in education sector, we need high penetration of Internet in the country and prepare infrastructure suitable for e-learning.

Spectrum is huge problem on infrastructure side in India that needs to be addressed rapidly to meet the demands.

Other important factor is the price of broadband. In India, price for Internet connectivity one has to pay is very high compared to Europe and other parts of Asia [6]. This barrier needs to be addressed fast to increase the penetration of Internet in India.

#### ***Inadequate autonomy to universities and Affiliation***

Universities are student data banks. They are sources of massive data required for data mining. India has both public and private universities that impart higher education. Public universities are supported by central and state government while private universities are supported by societies or bodies. India has totally 693 universities as on Nov 2014[16]. Out of these 45 are central universities, 325 are state universities, 128 are deemed universities and 195 are private universities.

In India, a higher education institution is subjected to several restrictions and regulations at various levels. UGC is main regulatory body controlling public as well as private universities. It is also responsible for awarding autonomous stature to institutions. UGC autonomy has resulted in various tussles with universities. University can excel only if autonomy is provided to it. In our country universities have been deprived of autonomy they need to function in interest of students. Governor of the state is the by default chancellor of the state university who is appointed by the President of India as a union representative of the state.

Financial aid has become most powerful instrument in the hands of state government to curtail the autonomy of the university and colleges [3]. Different commissions and committees have examined the Indian higher education system post independence, have identified the maladies that affect it and have also suggested remedies. However we have failed to follow up on the recommendations. The maladies identified by the Kothari Commission over three decades ago still exist. The Kothari commission emphasized that the proper sphere of institution autonomy is in three fields; (1) Selection of students, (2) Appointment and promotion of teachers, (3) Determination of courses of study, methods of teaching and selection of areas and problems of research [3].

In India, most of the colleges are affiliated to some university and there is no bar to the number of colleges that can be affiliated to a university. In some cases some universities are having hundred plus colleges affiliated to it. For instance at present, Osmania University has 901 colleges and Pune University has 811 colleges affiliated to it [17]. The aim of affiliation was to standardize education quality but analysis of results is disappointing. In top 100 universities less than 50 years age, India's IIT – Guwahati ranks 87[1].

The system of affiliation prevents good institutes to prosper as they are dragged back along with other colleges. The system of affiliation in India has been modeled on the University of London. Ironically the same model was abandoned in University of London in 1858[17].

Giving autonomy to universities in higher education is the need of hour to bring level of higher education at par with other foreign universities. Universities should be given power to design innovative courses, change/upgrade the syllabi as per industry need. They should be given freedom to implement innovative but effective methods of teaching.

#### ***Literacy rate and Enrollment ratio***

The big concern over success of online education in India is that India has largest illiterate population in the world. Compared to other countries of the world, India is far behind in literacy rate. India has literacy rate of 74% well below the average literacy rate of 84% of the world [10]. Gross Enrollment Ratio (GER) of higher education is 17.9% which is far below the world average of 27% as well as other emerging countries such as China(26%) and Brazil(36%)[20]. As far as computer literacy is concerned, in 2013, there were 185 mn computer literates out of 381 mn urban population and 125 mn computer literates out of 889 mn rural population [13]. English is the language that is preferred to be used by most of the online education providers. Though India is second largest as per total English speakers, it accounts only 10% of the Indian population [12]. To bring more students to online education, providers will have to prefer option of regional languages.

#### ***Mindset***

In spite of many advantages of online courses, people are still reluctant in adopting online courses. Key reason behind this is the mindset of the students, parents and employers. We feel that we can get job only if we are possessing degree from valid university. To add to fuel, our society also accepts degree holders and shows great respect to them. That is the reason today enrollment in colleges is still fat though many of them believe it to be unproductive. Students are ready to spend on handsome fees but they want degree from university only. It is degree, not skill that allows or disallows you to enter the interview room though we know degree is not a real measure of the candidate's potential. Unless this mindset is changed, quality online courses will not be able to attract quality students and if we are not able to collect digital data of student activities, it is difficult for us to eat fruits of analytics.

#### ***Power crisis***

Data analysis is done on machines that are driven by power. Power cuts are common throughout India. According to Central Electricity Authority Jan 2015 report, power supply was less than the demand in all the regions of India[7]. Statistically deficit was 6.8% in northern region, 0.5% in western region, 2.3% in southern region, 1.6% in eastern region

and 10.4% in north eastern region. Deficit for peak demand in the same regions were 38%, 1.6%, 4.2%, 1% and 10.3% respectively. In terms of particular states the peak shortage deficit was in Delhi, Uttar Pradesh, Maharashtra, Karnataka, Tamilnadu, Andhra Pradesh, Bihar, Odisha and West Bengal. Some of the states in the list are strategically important states with respect to development in IT sector. Surprisingly 4 metro cities Mumbai, Delhi, Chennai and Kolkata and IT hubs like Hyderabad, Bangalore and Pune belong to these states.

Daily consumption of energy per capita wise India's rank is 23<sup>rd</sup> with 2.02 KWh while Canada and US are 1<sup>st</sup> and 2<sup>nd</sup> with 51.5 KWh and 39.25 KWh respectively [8]. It simply means that with such less consumption compared to other countries if India is facing deficit what if per capita consumption increases to a level near US or Canada? Other important statistics that is worth to mention is that out of total energy produced, renewable power plants constituted only 28.43% while remaining 71.57% was from non-renewable power plants [9]. India has to increase the power production to meet future requirements and focus should be on maximum generation from renewable sources.

### **Lack of research**

Data mining is an interesting area of research. Research should be interwoven with education. Our education system is lacking thrust of research. As a result we are not able to get quality researchers in form of graduates. In comparisons of other countries, investment in R&D is less in India. In China, research scholars accounts for 0.08% of total population compared to India's 0.02% [23]. The quality of research is also not best in comparison with other countries.

### **Accreditation**

Accreditation addresses quality of institutions and education. Education of database and related technologies requires high quality. Our higher education is not at par of global quality standards. In India National Assessment and Accreditation Council (NAAC) assesses and accredits institutions of higher education. At present accreditation is voluntary for higher education institutions. Out of 612 universities in the country only 172 have been accredited. NAAC follows the process of grade accreditation only and does not take threshold accreditation. Grade is only relative value assigned to a university and is not an absolute attribute of quality. So it is high time now to do the compulsory quality assessments of universities in India and enforce international standards to improve the quality. A legislation namely National Accreditation Regulatory Authority for Higher Educational Institutions Bill, 2010 has been introduced in parliament which proposes to make accreditation mandatory for all higher education institutions[18].

### **Faculty shortage, Quality norms, Training and Lack of motivation**

Data mining is highly technical and versatile subject. To teach data mining and related subjects to students, technically sound and skilled teachers are required. There is an acute shortage of good teachers in higher education. The institutes of higher learning under HRD ministry face a faculty shortage to an extent of 35%. Data in following table shows current status of vacancy in various prestigious institutes of higher education [20].

TABLE – I

Type	Name	Faculty Strength (Sanctioned)	Faculty Strength (Current)	Vacant	Vacancy %
IITs	Indian Institute of Technology	6653	4079	2574	39%
IIITs	Indian Institute of Information Technology	253	162	91	36%
IIMs	Indian Institute of Management	766	577	189	25%
IISERs	Indian Institute of Science Education and Research	861	746	115	13%
SPAs	School of Planning and Architecture	672	618	54	8%
NITs	National Institute of Technology	6056	4292	1764	29%
CUs	Central Universities	16328	10058	6270	38%
Total	31589	20532	11057	35%	

Student / Faculty ratio is 29:1 in IIITs, 17:1 in NITs and 16:1 in IITs. Some institutes run show by appointing ad-hoc faculties at salaries close to minimum wages and for years at a stretch. This is demoralizing and results in deterioration in quality [22].

UGC has introduced Academic Performance Indicator – API to assess teacher's performance and recruitment and promotions of faculties are done based on this indicator. But introduction of this indicator has led to proliferation of poor quality journals, conferences [22]. It has resulted in competition of points rather than producing quality teachers.

Training programs of faculties are also few in numbers and those are conducted are formal, lack of enthusiasm and quality.

### **Industry academia bridge**

Industry embraces new technology first followed by academia. It is necessary that industry should communicate this new adoption to academic institutions as fast as possible so academic institutions can bring that technology in their curricula and subsequently industry can benefit with skilled workforce. In India there is wide gap between industry and academia. Industry requires skilled workforce but the necessary structured model of collaboration between industry and academia is hard to find. It has resulted in shortage of skilled manpower across various industries in India. To address this issue partnership between industry and academia needs to be strengthened and there must be seamless exchange of vital data between both of them. According to NASSCOM, each year over 3 mn graduates and postgraduates are added to workforce, however of these only 25% of technical and 10-15% of other graduates are considered employable by the rapidly growing IT and ITES segments [21]. Universities and institutions are not agile enough to upgrade syllabi to meet industry need. As a result graduates passing out are not industry ready.

### **III. INITIATIVES AND CURRENT STATUS**

In spite of above mentioned critical issues, there has been progress in education reforms and embracing new technology in education sector in India. Some initiatives and current state are discussed here.

Government of India has initiated process to bring in new education policy. National Mission on Education through ICT (NMEICT) is a landmark initiative of the Ministry of Human Resource Development (MHRD) to address all the education related needs of students, teachers and lifelong learners.

The mission aims to provide just-in-time quality educational resources and teachers 24 X 7 to learners [15].

Objectives of the mission are:

- Empowering and enabling students by ensuring equity and access to education through the use of ICT
- Connecting over 400 universities and 20000 colleges all over India through high-speed networks
- Improving faculty quality by using a unique synchronous training methodology
- Ensuring equity by providing access to expensive equipment to students even in remote corners through innovative use of ICT
- Making available e-content and educational videos created by the best teachers

Three cardinal principles of education policy – access, equity and quality are to be served through three major components:

- 1) Broadband connectivity to all colleges and universities – 60% of the budget is earmarked for connectivity. BSNL and MTNL has been assigned task to provide broadband connectivity.
- 2) Low cost access and computing devices for students and teachers – Ultra low cost computing device AKASH tablet has been developed to enable students to access the e-content.
- 3) High quality e-content generation – National Program on Technology Enhanced Learning (NPTEL) is joint initiative of IITs and IISc to provide e-learning through online web and video based courses in engineering, science and humanities streams. Virtual labs have been established for students to carry out experiments.

Private players like coursera, edx, edureka are also offering courses and certification online. Some courses are also provided free by them. The courses offered include big data related courses like Hadoop and related technologies also.

Government has taken initiatives to promote research in the country. It includes successive increase in plan allocations for scientific departments, setting up new research institutes, introducing attractive scholarships in research, encouraging public-private partnerships etc[23]. Between 2011 and 2013, more than 100 companies including Facebook and LinkedIn, have opened their R&D centers in India [2].

In mid of 2013, IBM signed an agreement with leading universities across the country to use IBM's smarter computing technology. The Indian Institute of Technology (IIT) Madras, The Indian Institute of Science Education and Research (IISER), Bhopal, Karunya University – Coimbatore, and the Shiv Nadar University (SNU) have selected IBM High Performance Computing (HPC) solutions, IBM System x iDataPlex servers, IBM PureSystems, IBM Power servers and storage solutions, virtualization equipment and software to aid high end research, refine processes and elevate the IT infrastructure in the universities. IBM also collaborates with business schools across India to help and enable business managers of tomorrow, address the need for skills in Big Data and Analytics. Over 10 institutions in India, like Sharda University and Galgotias University, have setup an IBM Business Analytics Lab, where participants are offered a two year full time MBA in Business Analytics in association with IBM. IBM India also runs 'Project Pravin', aimed at bridging the gap between what is taught as part of curriculum in campuses and the real life situations business organizations encounter [24].

To reduce industry academia gap, many IT companies are partnering with engineering colleges and universities. Infosys has started program called 'Campus Connect', Wipro has started Wipro Academy of Software Excellence in association with BITS Pilani, Tech Mahindra is planning to start engineering college[21]. Surely these are some of the welcome steps in the direction of improving relationship between industry and academia and creating strong bond.

UGC has made accreditation must by issuing a notification. The mandatory rule will cover every general-stream institute that has either completed six years or has provided education to two batches [25].

### **IV. CONCLUSION**

The technology is changing the world at a rapid pace. Every industry in the world is greatly impacted by this digital revolution. Not only developed nations, developing nations are aligning their strategies to this change. India is giant

democracy in the world. It has enormous rich youth potential who can transform the economy. What keeps away them from doing this is lack of proper higher education and development of vocational skills. If education process can be transitioned to digital era fully, we can use results of analytics in improving knowledge, skills and power of our youth greatly. Initiatives and efforts discussed above have to be accelerated at great speed to achieve this and put India in the front list of nations having rich and matured education model.

#### REFERENCES

- [1] Times Higher Education Website. [Online]. Available: <http://www.timeshighereducation.co.uk/world-university-rankings/>
- [2] EY – Higher Education in India : Vision 2030 [www.ey.com/...India.../EY-Higher-education-in-India-Vision-2030.pdf](http://www.ey.com/...India.../EY-Higher-education-in-India-Vision-2030.pdf)
- [3] Dr M M Gandhi Principal Professor and Head PG Dept of Commerce and Management, Jaysingpur college of Arts, Commerce, Science and Computer Science. Autonomy and Accountability in Higher Education An Indian perspective Available: <http://www.iosrjournals.org/iosr-jrme/papers/Vol-3%20Issue-5/F0353337.pdf?id=7370>
- [4] Dr J D Singh, Higher Education in India – Issues, Challenges and Suggestions <http://www.gvctesangaria.org/websiteimg/publications/jdarticle.pdf>
- [5] Internet live stats Website. [Online]. Available: <http://www.internetlivestats.com/internet-users/india/>
- [6] [http://en.wikipedia.org/wiki/National\\_broadband\\_plan#India](http://en.wikipedia.org/wiki/National_broadband_plan#India)
- [7] Government Of India, Ministry of power, Power Sector Executive Summary [http://www.cea.nic.in/reports/monthly/executive\\_rep/jan15.pdf](http://www.cea.nic.in/reports/monthly/executive_rep/jan15.pdf)
- [8] [http://en.wikipedia.org/wiki/Electric\\_energy\\_consumption](http://en.wikipedia.org/wiki/Electric_energy_consumption)
- [9] [http://en.wikipedia.org/wiki/Electricity\\_sector\\_in\\_India](http://en.wikipedia.org/wiki/Electricity_sector_in_India)
- [10] [http://en.wikipedia.org/wiki/Literacy\\_in\\_India](http://en.wikipedia.org/wiki/Literacy_in_India)
- [11] Gartner highlights the top 10 strategic technologies impacting education in 2015 <http://www.gartner.com/newsroom/id/2994417>
- [12] [http://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_English-speaking\\_population](http://en.wikipedia.org/wiki/List_of_countries_by_English-speaking_population)
- [13] Internet in India 2013 [http://www.imrbint.com/downloads/Report-BB55685%20IAMAI%20ICUBE\\_2013-Urban+Rural-C1.pdf](http://www.imrbint.com/downloads/Report-BB55685%20IAMAI%20ICUBE_2013-Urban+Rural-C1.pdf)
- [14] Digital India 2014 - [http://thedmti.com/images/Digital\\_India\\_2014.pdf](http://thedmti.com/images/Digital_India_2014.pdf)
- [15] NMEICT – National Mission on Education through ICT Website. [Online]. Available: <http://www.nmeict.ac.in>
- [16] <http://www.ugc.ac.in/oldpdf/alluniversity.pdf>
- [17] Is loss of autonomy trapping higher education? <http://indiai.org/blog/is-loss-of-autonomy-trapping-higher-education/>
- [18] Dr Shashi Tharoor (MoSHRD) in Rajya Sabha - Standard of Universities <http://pib.nic.in/newsite/erelease.aspx?relid=89764>
- [19] Pratiyush Guleria and Manu Sood, “Data mining in education: A review on the knowledge discovery perspective”, International Journal of Data Mining and Knowledge Management Process(IJDKP) Vol 4, No 5 September 2014 <http://aircse.org/journal/ijdkp/papers/4514ijdkp04.pdf>
- [20] Rakesh Dubbudu, 35% faculty positions vacant in Government Higher education institutes <https://factly.in/35-faculty-positions-in-higher-education-india-institutes/>
- [21] Lokesh Mehra, Regional Manager, Corporate Responsibility, South Asia, CISCO Systems Bridging the skills gap with industry: Academia partnerships [http://www.cisco.com/web/IN/about/network/academia\\_partnerships.html](http://www.cisco.com/web/IN/about/network/academia_partnerships.html)
- [22] Arun Kumar – The Hindu, When accountability is not institutional <http://www.thehindu.com/opinion/lead/lead-article-when-accountability-is-not-institutional/article6764600.ece?css=print>
- [23] Smriti Irani, HRD Minister in Lok Sabha <http://timesofindia.indiatimes.com/home/education/news/Lack-of-funds-affects-research-in-India-Smriti-Irani-says/articleshow/39782734.cms>
- [24] IBM News room - <http://www-03.ibm.com/press/in/en/pressrelease/42015.wss>
- [25] UGC makes accreditation must [http://www.telegraphindia.com/1130305/jsp/nation/story\\_16634387.jsp#.VSP2qtyUeNA](http://www.telegraphindia.com/1130305/jsp/nation/story_16634387.jsp#.VSP2qtyUeNA)