Big Data Technologies in Scholarship Management of Technical Education System

Harinder Pal Singh, *Harpreet Singh

Desh Bhagat University, Mandi Gobindgarh & Department of Technical Education Govt of Punjab, India

*Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI, USA

Abstract: *The Applications of Big Data management systems in technical education have been increasing in the last few years. Due to the easy availability of the mobile phones and other intelligent devices along with internet facilities students have started using mobile phones, Laptops to access online content and instructions, study materials and submission of necessary academic inputs to the institutions/universities etc. Student's online activities generate enormous amount of data, where used and unused data requires Big Data technologies and tools to process and make use of the large amounts of data involved. This study looks into the recent applications of Big Data technologies in scholarship management of technical education system and presents techniques on handling Educational Data Mining and Learning Analytics.*

Key words: Technical Educational, Scholarship Data, Data Mining, Big Data, Learning Analytics etc.

1.1 Introduction

Research in the technical education system has resulted in several new pedagogical improvements. Today the global education system has become a competition within the institutions for the attention of prospective students in order to increase the number of successful admissions. Due to advances in the computer technology and faster internet facilities, it is much easier for schools and colleges to reach out to more and more students and to encourage them for admissions and other related activities. The online admission and offline computer based admission process in technical education system has increased the volume of the data related to the academic process. The institutions now adopting the advanced models to analyze access, and manage vast volumes of this data. Such data in technical education system can be further classified in different ways such as data related to the admission, evaluation and scholarships etc. Based upon the above data institutions/ universities are working hard to identify relevant talent pools in order to appeal to suitable candidates. It is also very hard to eliminate the students with fake data and duplicate data coming from various sources. In order to identify these tools, big data analysis comes in handy. Apart from identifying a particular tool, it also proving opportunity to improve the decision-making process for technical educational institutions, while examining

previous student data. Whereas world education systems are already applying such tools and using such practices for business intelligence, financial analytics, predictive analytics and strategy management etc. The prime focus of this study is on development of a data-based architecture that enables the addition of a variety of sources, while managing the scholarships. These data sources can include – students' personal information, their past educational institutions, results, certificates, parental income and dropout rates etc.

1.2 Big Data

The term "Big Data" is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data that are too large or complex to be dealt with by traditional data-processing application software [1]. Examples of Big Data include the amount of data shared in the internet every day, Industrial data, educational institute data, medical data, server of search engines, YouTube videos, Facebook and twitter feeds and mobile phone location data etc. [2]

2.1 Challenges in handling of Big Data

There are several challenges that need to address while handling Big Data. Those challenges include

2.1.1 Storage:

Hard disks nowadays are of high storage capacity in the range of terabytes, the amounts of data generated through internet everyday are in the order of Exabyte. Though the data generated in technical education is not as large as all the data generated through internet. The traditional RDBMS tools are sufficient to store or process such Big Data. If the data exceeds the limit compression technology can be used to compress the data at rest and in memory.

2.1.2 Data Analysis:

As data depends upon the application to application as data of online learning sites is different in structure and the size as compared to data of vision sensor based surveillance system. Similarly data related to banking is different in size as compare to educational institution data. In such a case analysis of the data may consume a lot of time and resources. To overcome this, scaled out architectures are used to process the data in a distributed manner. Data can be split into smaller pieces and can be processed in a vast number of computers.

3.1 Techniques for handling Big Data

The challenges faced in processing Big Data technologies are overcome by using various techniques [3 and 4]. The most popular techniques used in educational data mining are listed below.

- *Regression* – Regression is used in predicting values of a dependant variable by estimating the relationship among variables using statistical analysis
- *Averaging*-averaging used to average the independent variable
- *Nearest Neighbour*– In this technique the values are predicted based on data information.
- *Clustering*– Clustering involves grouping of data of an n-dimensional space where n is the number of variables.
- *Classification*– Classification is the identification of the category/class to which a value belongs.

3.1.1 Open source tools and techniques

Several Open source tools are available to handle big data such as **MongoDB.** MongoDB is a cross platform document oriented database management system used to manage data in the form of table. **Hadoop** is another architecture that deals with the data of clusters of the networked computers. **MapReduce** is a modelling tool that enables processing of huge amount of data in parallel on large clusters of compute nodes. There are some python based tools like **Orange** used to processing of big data [4, 6 and 7].

4.1 Applications of Big Data Techniques in scholarship management in technical education system

There is a wide range of applications in Big Data techniques that can be used in a variety of ways in learning analytics in scholarship management in technical education system [8]:

4.1.1 Personal data of students enrolled under various institutions: MongoDB is a cross platform document oriented database management system used to manage data in the form of table to store the personal data of the students like their national identity number, bank accounts number, and family income etc.

4.1.2 Analysis of data and estimation: In technical education system analysis of student data helps in providing the analysis of variables such as how many students belonging to particular sections of the society became graduates in various degrees and diplomas, who otherwise could not have been the part of educated population etc. Analysis provided the economic condition of the people and the state as a whole. After completion of the course the placement of such students contribute to analyse the GDP growth of the state and the country. Analysis shows that   every year about 80,000 students of weaker section are enrolled for getting scholarship in a year in the Punjab state of India under Technical Education Department. From year 2014-18 around state

government spent more than Indian rupees (INR) 1200 Crores in the form of scholarships for technical education only. Analysis also classified the area of specialization of enrolled students in different technical education streams.

4.1.3 Risk Detection:

Data security and information integrity is a big challenge in institutional data as the personal data and information of applicants can be stolen online. For example if national identity number of the student or bank accounts is stolen by hackers it can lead to financial loss to the applicant students. Leakage of such personal and classified data can lead to various scams. So risk detection and analysis and using various security techniques like modern encryption algorithms are proposed to be inbuilt in the data mining system.

4.1.4 Performance Prediction:

The performance prediction of students whether he/she is continuing in his studies after availing the benefit of scholarship need to be ascertained before granting the scholarship application for the next semester/year. His Board/University scores need to be linked using various data tools to the Data Base Management system. If he do not appear or pass any of the subjects his application is liable to be rejected till he passes the requisite no of subjects and re apply for scholarship of next semester/year. In the proposed study data alert has been implemented. Drop out rates can be ascertained while analysing the data so finally the decision making can be improved for further award of scholarships.

4.1.5 Data Visualization: Technical educational data become more and more complex as it grow in size. Data can be visualized using data visualization techniques to easily identify the trends and relations in the data just by looking on the visual reports.

4.1.6 Intelligent feedback: Learning systems can provide intelligent and immediate feedback to students in response to their inputs which will improve student interaction and performance. It is proposed to implement Mobile app that can be developed by linking application submission transaction for scholarship applications till the approval happens.

4.1.7 Social network analysis: A social network is defined as a social structure of individuals, who are related (directly or indirectly to each other) based on a common relation of interest, e.g. friendship, trust, etc. Social network analysis is the study of social networks to understand their structure and behaviour. Social network analysis has gained prominence due to its use in different applications-from product marketing (e.g. viral marketing) to search engines and organizational dynamics (e.g. management). Recently there has been a rapid increase in interest regarding social network analysis in the data mining community. The basic motivation is the demand to exploit

knowledge from copious amounts of data collected, pertaining to social behaviour of users in online environments. A prime example of this are the research efforts dedicated towards the Enron email dataset. Data mining based techniques are proving to be useful for analysis of social network data, especially for large datasets that cannot be handled by traditional methods.

4.1.8 Planning and scheduling: Benefits derived from innovative scheduling techniques range across the entire spectrum of the educational experience, frequently providing for substantial change in the roles of students, teachers, and administrators. Under such a scheduling system, students can choose instruction to meet their needs and their college authorities can assume greater participation. In an online scholarship data system, scheduling is provided as student requires adding identification number. Once it gets verified it gives access to the student to submit application online. Application goes to the institute head that verifies the student credentials, and forwards it to the Sanctioning authority. The concerned sanctioning authority sanctions the scholarship claim, and forwards it to the Line Department which verify it and approved the online payment to the student into his Bank account. Handing thousands of such students applying and getting scholarship is objective study of planning and scheduling.

Conclusions

Big data management and handling of scholarship in technical education system is presented. It is concluded that for handling of big data various tools and techniques are required, we used MongoDB open source tool which is a cross platform document oriented database management system to manage data in the form of table. Tools used to manage big data are given in section 4.1 it is found that with use of above the memory saving, threat assessment and analysis of data has significantly been improved.

References

[1] Feng Xia ; Wei Wang ; Teshome Megersa Bekele ; Huan Liu , Big Scholarly Data: A Survey, IEEE Transactions on Big Data (Volume: 3 , Issue: 1 , 18 - 35 March 1 2017)

[2] Jiaying Liu, Tao Tang, Wei Wang, Bo Xu, Xiangjie Kong, Feng Xia, "A Survey of Scholarly Data Visualization", Access IEEE, vol. 6, pp. 19205-19221, 2018

[3] Jefry Dean and Sanjay Ghemwat, .MapReduce: Simplified data processing on large clusters, Communications of the ACM, Volume 51 pp. 107-113, 2008

[4] Carlos Ordonez, Algorithms and Optimizations for Big Data Analytics: Cubes, Tech Talks, University of Houston, USA 2017.

[5] Lamyae Hbibi, Hafid Barka, "Big Data: Framework and issues", Electrical and Information Technologies (ICEIT) 2016 International Conference on, pp. 485-490, 2016.

[6] Leo Willyanto Santoso, Yulia, "Data Warehouse with Big Data Technology for Higher Education", Procedia Computer Science, vol. 124, pp. 93, 2017.

[7] R. Kimball, M. Ross, The Data Warehouse toolkit, Indianapolis US:John Wiley & Sons, Inc., 2013.

[8] A. S. Drigas, P. Leliopoulos, "The Use of Big Data in Education", *International Journal of Computer Science Issues (IJCSI)*, vol. 11, no. 5, September 2014,