BIG DATA ANALYSIS IN EDUCATION SYSTEM

[1]**Arvind Kumar Pandey,** [2] **Dr. Asha Ambhaikar**
[1] Research Scholar, Kalinga University, Raipur
[2] Professor, Kalinga University, Raipur

Email: - arvonline@yahoo.com

ABSTRACT

Big Data is the term used to refer to a huge volume of data with diverse variety and complex data sets which may be heterogeneous in nature with complex structure. The capacity to analyze and examine huge volume of complex data which may be heterogeneous in nature has been an objective of various educational institutions for quite some time to get some meaning information. The traditional tools have been failing to process such complex data and to reach to a tangible decision. Educational institutions have been trying to find a way to analyze data and improve student performance. In education system, Big Data provides an opportunity to use the analysis tools to examine large volume of data strategically to improve educational quality and help students' determination and outcome. It also helps to build a dynamic and better education system which will help every student to achieve maximum benefit by improving the quality of education. Students, teachers and management leave a continual set of data during the time of work at the institution. These datasets can be used to gain insight into the performance at individual student level as well as at the institution level. This paper is a study on the use of Big Data Analysis tools in Education. It aims to discuss the analysis tools used in big data to improve education by analyzing and benefiting from the unused data sets.

Keywords: *Big Data Analysis, Complex datasets, Decision Making, Educational system, Student performance*

1. INTRODUCTION

Today the organizations generate huge amount of data on a daily basis. This massive data generated daily is known as Big Data. Gartner's definition of Big data: Big data is data that contains greater variety arriving in increasing volumes and with ever-higher velocity. This is known as the three Vs of Big Data [1].

Big data is data in huge volume with diverse variety and with ever-higher velocity. Few more characteristics have been added to it like Veracity and Value. In simple words, we can define big data as larger, more complex data sets, especially from new data sources. These huge volumes of data can be used to deal with problems that have been difficult for you or organizations to handle earlier.

Big data is not only about how much data is being collected but it is about how it is being utilized. Every organization utilizes data according to its own need and convenience. The more efficiently an organization uses its data, the more capability it has grown. The organization can collect data from several sources and evaluate it to find solutions which will have tremendous impact in the organization [2]

The Different Vs of Big Data
1)Volume:
Volume refers to the amount of data available which need to be evaluated as per requirement for a desired result.[3] It can be in the form of text, videos, music and large images, etc. As the volume of the data grows the techniques, applications used to evaluate them also need to be reevaluated.[4]

2)Velocity:
Velocity is referred to the rate at which the new data is being generated, stored and acted upon. For example websites like Facebook store almost 1 billion of images every day. 500 million of tweets are posted on twitter, 0.4 million hours of videos are uploaded on YouTube and 3.5 billion searches performed in Google every day. Big data has techniques to manage this huge in flow of data and at the same time processes it faster to avoid hold ups.

3)Variety:
Variety in Big data refers to all the unstructured, semi-structured and structured data that has been generated. The structured data are texts, picture, videos, tweets. The semi-structured and unstructured data are emails, hand written text, voice mails, ECG readings, etc. [3] They come under this category as they require further processing to derive meaning data.

There are few more Vs defined now-a-days.

4)Veracity: Veracity refers to the quality and legitimacy of the data . It verifies the validity of the data which can be used to cater the different needs of an organization.

5)Value: Value refers to the significance of the data to Big Data Analysis and extent to which the data is utilized to generate information which is of usefulness for an organization

Big Data Analysis

Big Data Analysis is the method of examining large amount of data to explore hidden patterns, relationships and other insights. With all the latest technologies, it is possible to analyze data and get the results immediately. Big data Analysis help institutions to process data and the outcome of it is used for decision making to reach to a desired goal.

Big Data Analytics provides organization with complete details of customers, which permits a lot of customized client information. It helps the organization with the ability to add a variety of data from various sources in real-time.

Big Data Analysis reduces data redundancy so that the organizations can acquire unique view of the clients that helps in the compiling of a comprehensive record of the behavior of each client. Big data can be intimidating, but with the use of Big Data Analysis any institution can use it efficiently to address its needs. As the amount of data continues to grow, the organization also starts growing in all the terms by using the stored raw data to generate relevant information.

Big Data Analysis in Education

Organizations are collecting and storing huge amount of data from varied sources may be it digital or so. They are also interpreting these huge data namely big data as per their business needs with the latest computing and machine learning technologies. The big data analysis is now not only limited to businesses but now it is efficiently used in education.[5]

With the advent of new technologies and growing number of educational institutions be it schools, colleges, universities, - they are collecting data from different sources. Internet and open government information are now-a-days very popular source of data which are efficiently collected by different organizations and are used after being processed in a customized way.

Educational institutions like schools, colleges, universities and other educational bodies collect and compile lot of data about faculties and students. This data can be analyzed to get information that can be used to improve the functioning/performance of the institutions. Collection and analysis of student data includes – attendance, grades, test scores, and disciplinary issues.

The analyzed report can be used in the varying educational requirements. The big data analysis provides the schools and institutions with information about the student conduct and performance. It also helps the institutions to plan and redesign its working methods where students can learn in a more interesting and a better way.

i. Helps in Better Decision-Making

One of the important features of big data analysis is to know how to make use of it. Institutions will be in a better position if they make use of suitable techniques to analysis big data which they keep on collecting, storing and examining. Utilizing this data coming from various sources, institutions can use various techniques to improve teaching methods which are accepted by the students.

ii. Students' Results

A student's scholastic performance can be calculated on the basis of his/her examinations and the prevailing results.

Every student produces an extra ordinary result in his or her lifetime which can be examined for a better-quality understanding of a student's behavior to make the most superlative learning condition.

Big Data analysis monitors students' advancement like classroom performance, desired subjects, curriculum activity interests, the time they take to finish an examinations and numerous different things in a student's educational condition.

A progress report can be developed that will exhibit the interest as well as concern areas of a student.

iii. Career Forecast

After scrutinizing the performance report of the students, the school/college/university officials can understand the strengths and weaknesses of the students and help them in recommending the area of interest they have. This will also help them identifying the field in which they can seek out a profession.

In this case, the big data analysis can help the student who is passionate about a specific subject/course. The big data analyzer can advise the students to follow what is right for them to follow.

iv. Assessment

Big Data Analysis helps a teacher to assess, observe, and

react, in real time to a student's understanding of the subject. Analysis can help teachers to follow a particular teaching style and deal with student needs before their final examination results. It is an important development for teachers as it strengthens their skills to deal with any biases that may be towards the performance of the students.[6]

v. Improvise the Learning Experience

Big data analysis helps teachers to make courses interesting for students with different level of knowledge from introduction level to expert level. Using big data analysis teachers can understand where the students are facing problems or are performing well. Based on this analysis, teachers can prepare different starting study material for each student in the same subject. This will always help improve student interest in the subject and point out when a particular learning material need to be given to the students to start.

vi. Design new Courses

Major challenges in professional institutions or universities are to efficiently understand the requirement of businesses and design a market oriented program to meet their demands. Big data analysis can be used efficiently to understand the demand and recent trends. It also helps to create basic courses and learning materials around emerging concepts.

2. TOOLS AND TECHNIQUES

In Education system, data are in more hierarchical form. Data is kept in an unorganized form at various levels in an Educational institution starting from the student level, the class level, the teacher level and the school level which are nested inside one another at various levels as shown in **Figure 1**. [7]
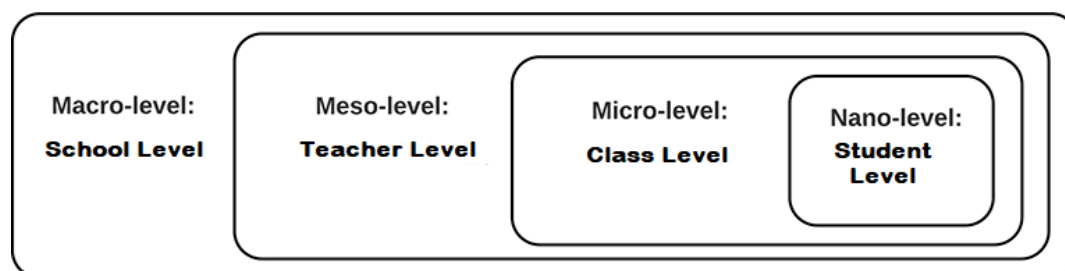
Researchers have identified a set of objectives for big data analysis in educational system [8]:

1. By creating a student model which will help to predict a student's future learning behavior by incorporating detailed information like student's knowledge, motivation, metacognition, and attitudes

2. By discovering and improving domain models that characterize the content to be learned and optimal instructional sequences

3. By studying the impact of various kinds of pedagogical support which can be provided by learning software; and

4. Through the help of computational models which incorporates model of the student, the program, and the software pedagogy, researchers can develop the understanding level about learners and learning.

A wide variety of techniques have been developed and adapted for big data analysis. These techniques can be adapted from various fields including computer science, mathematics, and statistics. This means that an organization must adopt a flexible and multidisciplinary approach in order to get desired results.

Researchers are continuously working on it and the story of big data is still being written. New methods, techniques and technologies continue to be developed to solve new problems.

2.1 TECHNIQUES FOR ANALYZING BIG DATA

Researchers have developed a wide range of techniques and technologies associated with big data analysis which has been adapted to collect, manipulate, analyze and visualize on diverse datasets ranging from small to too big which are drawn from various fields such as statistics, applied mathematics, economics, and computer science. The list of techniques and tools is ever growing as the researchers continue to improve the existing techniques and develop new

Figure 1

ones. Some of these techniques can be used for smaller datasets while some others are used to handle large datasets.

We are going to discuss the popular techniques which are used by almost every institution to work on huge datasets.

A/B testing, Data mining, Association rule mining, Data fusion and data integration, Machine learning, Neural Networks, and Visualization.

There are various other data analysis techniques which include Association rule learning, Ensemble learning, Genetic algorithms, Natural language processing, Network analysis, Optimization, Pattern recognition, Predictive modeling, Regression, Sentiment analysis, Signal processing, Spatial analysis, Statistics, Supervised learning, Simulation, Time series analysis, Unsupervised learning and many are being added by researchers on regular basis.

A/B testing: A/B testing is also known as split testing or bucket testing. It is a very using technique for web developers and professionals engaged in digital marketing. This technique is used to compare a control group with a vast variety of test groups in order to determine what changes can be recommended to improve a given objective variable.

Data mining: In Data mining, various techniques and methods are combined to extract knowledge in the form of patterns from large datasets from database management systems. The extracted knowledge is very important and significant for the decision makers of any organization. These techniques include association rule mining, classification, cluster analysis, and regression.

Association rule mining: Association rule mining technique is used to establish relationship of attributes and tuples, by discovering frequently occurring item sets in database. It is used to find frequent patterns, associations or correlations among sets of items or elements in data repositories. Finally, the result consists of a confidential value and support value which is again used to identify the pattern and solution.

Data fusion and data integration: A set of techniques are combined to analyse and integrate data from multiple data sources to get accurate, reliable and useful information. In this technique, various techniques are combined to resolve conflicts from a collection of different sources and to find reliable information which can be more accurate than the information generated from a single source of data.

Machine Learning: Machine learning is a part of Artificial Intelligence (AI) which is used for big data analysis. It includes software that can be used to analysis and learn from data to extract useful information. It focuses on enabling computational systems to learn from data to automate a given task to recognize complex patterns and make intelligent decisions. It is generally used in the area of applications like forecasting, decision making. [9]

Neural Networks: Neural network method has been established to be very useful technique for statistical analysis of a big data set. It is the most preferred techniques when information is to be sought out of non-linear patterns. Neural network techniques can be used for pattern recognition and optimization. Some neural network applications involve supervised learning while others use unsupervised learning.

Visualization: Visualization makes the organizations to look beyond the individual records and easily identify dependencies and correlations hidden inside large data sets. Different visualization techniques can be used to create images, diagrams and animations to analyze, understand, communicate, and improve the results of big data analysis

2.2 TOOLS FOR ANALYZING BIG DATA

Big Data Analysis software is widely used in providing meaningful analysis of a large set of data. This software helps in finding current market trends, customer preferences, and other information.

Technologies in big data analysis can be categorized in two groups – Operational Big Data Technologies and Analytical Big Data Technologies. Operational Big Data Technologies indicate the amount of data generated on a daily basis which includes data from online transactions, social media. Generally, this type of data is generated from online trading and purchasing from Flipkart, Amazon, online ticket bookings for flight, railways, movies and many more.

Data generated from Operational Big Data Technologies acts as raw data to be feed in the process of Analytical Big Data Technologies. Analytical Big Data Technologies refers to advance adaptation of Big Data Technologies which is technically complicated compared to Operational Big Data Technologies. In this category, huge data generated are considered from the areas namely, weather forecasting, time series analysis, stock marketing, and medical records.[10]

The number of technologies keeps on growing on regular basis to support the development in the area of big data analysis techniques. Some of them are listed below:

Artificial Intelligence: Artificial Intelligence can be used to change data analytics by studying results generated from past analysis tasks so that future results can be generated faster with more accuracy. Future predictions can be made accurately based on the current events and by studying the enormous amount of data, along with analytic results from past analytical queries. One of such example of artificial intelligence model is weather predictions. The two subsets of Artificial Intelligence – machine learning and deep learning has played a major role in the advancement of AI technologies. Machine learning is mostly associated with predictive analytics which enables systems to look at historical data, recognize patterns, build models, and predict future outcomes. As contrast to machine learning, deep learning is a subset of machine learning which can be used to simulate the working of a human brain creating artificial neural networks that use multiple layers of algorithm to analyse data. It can be used in big data technologies, which allows analytical tools to recognize content in images and videos and then process this accordingly.

Business intelligence (BI): Business Intelligence is the technology involved in process analysis of data to help the business executives and the corporate heads to draw business decisions. Business Intelligence technologies can be used in Big Data analysis to find the solutions of complex problems after reading past data from a data warehouse or data mart. Business Intelligence technologies can be used to generate periodic reports and display information on dashboards of real-time management.

NoSQL Databases: NoSQL includes a wide range of distinct database technologies that are used developing to design modern applications. It describes a non-relational database technology that delivers a method for storing and retrieval of data without SQL. They are generally deployed in real-time web applications and big data analytics. One of the major advantage of using NoSQL databases is that it can store unstructured data and delivers faster performance, and tenders flexibility while dealing with varieties of data types at a huge scale. It uses data structures that are different from those accounted by default in relational databases, it makes computations quicker in NoSQL. There are more 250 NOSQL databases available today and the list is ever growing. The commonly used includes MongoDB, Redis, and Cassandra. The major companies using NoSQL databases include Facebook, Google and Twitter which store terabytes of user data every single day.

R Programming: R is an open-source programming language used and leadingly used for data science, consisting of powerful functions to tackle all problems related to Big Data processing. It is a free software majorly used for statistical computing and graphical visualization. It has become a de facto standard among statisticians for developing statistical software and is majorly used for statistical software development and data analysis.

Big Data Analysis using Excel: Excel is a very powerful data analysis tool and almost all big and small institutions are using Excel in their day to day functioning starting from compilation of attendance to fees collection to marks compilation. Excel uses the ability to analyze data using powerful skills that helps in making better decisions. Excel is one of the top tools for data analysis and it has built-in features like pivot tables which is the most popular analytic tool.

3. CONCLUSION

This paper presents the different tools and techniques used in Big Data Analysis in Education System. In education system also, large volume of data is generated in different forms. It is very difficult to handle these large amount of data and creating relationships among them to analyze and reach to a certain result. This paper also discusses about the popular tools and techniques used for analysis of big data in education system.

REFERENCES

[1] https://www.oracle.com/in/big-data/guide/what-is-big-data.html

[2] https://www.newgenapps.com/blog/importance-benefits-competitive-advantage-big-data

[3] https://www.whishworks.com/blog/big-data/understanding-the-3-vs-of-big-data-volume-velocity-and-variety

[4] https://blog.sqlauthority.com/2013/10/02/big-data-what-is-big-data-3-vs-of-big-data-volume-velocity-and-variety-day-2-of-21/

[5] https://www.newgenapps.com/blog/5-dramatic-impacts-of-big-data-on-education

[6] https://master.edhec.edu/news/three-ways-educators-are-using-big-data-analytics-improve-learning-process

[7] B R Prakash, Dr.M. Hanumanthappa and VasanthaKavitha. "Big Data in Educational Data Mining and Learning Analytics", International Journal of Innovative Research in Computer and Communication Engineering, 2014

[8] Jeong, H., and G. Biswas. 2008. "Mining Student Behavior Models in Learning-by-Teaching Environments." In Proceedings of the 1st International Conference on Educational Data Mining, Montréal, Québec, Canada, 127–136.

[9] Big data for development: applications and techniques

[10] Introduction to Big Data in Education and Its Contribution to the
Quality Improvement Processes,
https://www.intechopen.com/books/big-data-on-real-world-
applications/introduction-to-big-data-in-education-and-its-
contribution-to-the-quality-improvement-processes

IJSER