Edu Data Mining

SHOBHANA AVINASH GAIKWAD[1]

1 Professor, Computer Technology, Bharati Vidyapeeth Institute of Technology, Maharashtra, India

ABSTRACT

Education institute are the nursery for the future minds of the nation. Education institute have large amount of student data like basic personal information, attendance, marks, achievements etc. Online learning platform provide an opportunity to capture fine gain data about student online activities will generate large amount of structured and unstructured data. But it is found that educational system are notoriously poor in managing the data and taking advantage of this generated data There are two research area for Big Data mining in education called educational data mining and learning analytics. Educational data mining is suit for the computational and psychological methods and research approach for the understanding how student learn, predict student future learning behavior. Learning analytics is becoming defined as an area of research and application and is related to academic analytics actions and predication analytics. As recently there are lot of research in education data mining and some researcher started treating data in education system as big data problem, we have done survey of various researches in education.

Keyword: - *Education, ERP, Data Mining, Big Data, Hadoop, Association Rules*

---

1. INTRODUCTION

Recent years we have witnessed rapid growth of the Internet, which has become an important medium to deliver digital content to Web users instantaneously. The recent report shows that online courses are spreading and developing around the world [1]. An educational institute and universities started providing e-learning facilities for providing contents and using ERP for better management of student data have opened up opportunities to collect and analyze student data, to discover patterns and trends in those data and to make new discoveries and test hypothesis about how students learn [6].

An educational institute have vast amount of data like student personal information like name, student registration number, birth date, and address etc. and curriculum information like code, name, semester achievements, assignments status unit tests and semester marks etc. When students are learning online, there are multiple opportunities to exploit the power of the technology for formation of for formative assessment [1].

Now many universities and college and have started online forum where student share their day today encounter share their knowledge seek help for problem also provide opportunity to capture fine grain data about student can provide deep insight view [4]. Big Data as a term has been among the biggest trends of the last three years, leading to an upsurge of research, as well as industry and government applications. The process of handling big data encompasses collection, storage, transportation and exploitation. “Big data” has come to be defined by the four V’s— Volume, Velocity, Veracity, and Variety [2]. Recently there is massive research in field of education data mining. We have done the survey of various works in education data mining.

2. EDUCATION DATA MINING

Educational data mining is emerging as a research area with a suit of computational and psychological methods and research approaches for understanding how students learn. Education data mining have targeted various issues like Understanding real learning behaviour of a students, the detection of students who show signs of frustration, students' academic achievements (successes and failures), students' drop out, and students' financial behaviour. We have done the survey of various research in education domain.

The technology of data mining could discover useful knowledge from mass data, so it has been broadly applied in various fields. There are various research to analyse the exam grade of college students. There is attempt to find the weighted association rules to analyse grades of college- wide examination course [6].

However, a review of engineering education literature reveals scant research on the use of online discussion forums for engineering learning beyond the classroom [4]. In case of online discussion there is an various ongoing research to find what student's sought help for and for what purpose, which words are frequently used by students who seek homework help?, what are the major themes of discussion?. To answer these questions there are various attempt to use keyword analysis, topic modeling, linguistic inquiry and word count analysis. There is an also attempt to mining social media data for understanding students' learning experiences. Students are more active on social media sites like Facebook, twitter, LinkedIn where they share there day today encounters, experiences, emotions, stress, achievements [3].

Analysis of these informal discussions on social media can give vast amount of information about student learning experience, opinions, feeling, and concerns. Analysis of such data will provide key information for education decision makers. There is an attempt to mine twitter data, have collected about 28500 tweets using the hash tag #engineering Problems over a period of 14 months, and a second data set of 39,095 tweets using the geo-code (longitude and latitude) of Purdue University, West Lafayette. This research proposes a workflow to bridge and integrate a qualitative research methodology and large-scale data mining techniques. Research applies the algorithm to another large-scale and unexplored data set, so that the manual method is augmented. Research can keep refining the model based on further human feedback like the cycle [3].

In a successful engineering education, continuous improvement should be fulfilled through all phases: curriculum, speech or talk listening, practical training, industrial visiting, contest attending and undergraduate research [1]. Once a student is predicted to get bad grade, its teacher can strengthen the student's learning difficulty at the beginning of a course. So research focusing on developing data mining tool which can forecast students' performance to the instructor. In order to solve this issue, research proposes a decision support of integrating data mining, fuzzy logic and Bayesian networks to qualitatively predict students' learning. The research have proposed four steps: 1) using fuzzy theory to identify the influence factors on learning outcomes through related work analysis and questionnaires; 2) Using data mining to analyses the relationships among the factors and then to construct influence diagram; (3) using machine learning to establish the fuzzy inference relations through questionnaires and parameter learning; (4) using the model to predict the exam scores at the beginning of course and thereby to help students enhance their scores according to their weakness [1].

3. BIG DATA

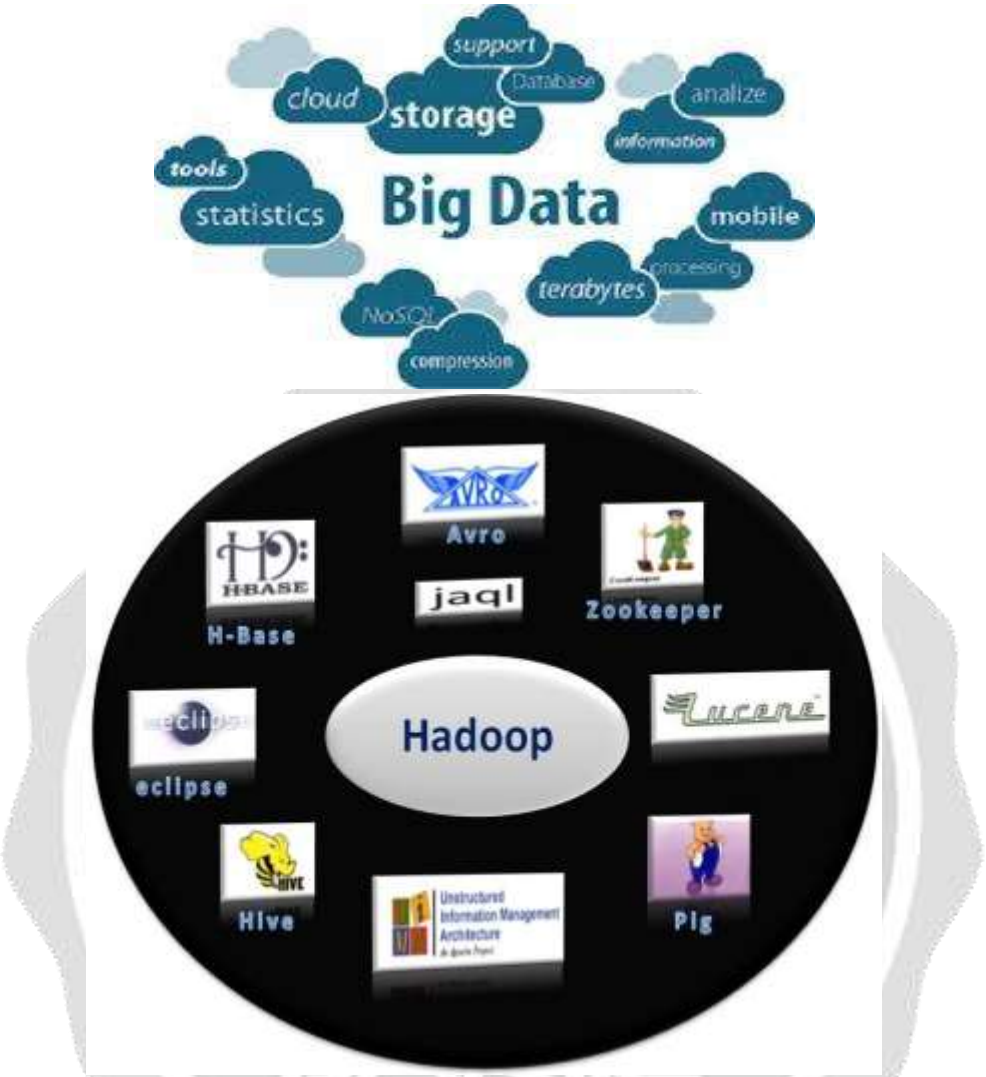


Fig- 1: Big data and Hadoop

Big data is one of the "hottest" phrases being used today. Big Data is actually the data that is big in size and various companies or social media platforms generate it. If you consider a social media like Facebook, just storing the number of active users currently would itself be a big data as there would be millions of people accessing the Facebook at a time around the globe. Big Data has become a catchphrase to describe data so large that it is not amenable to processing or analysis using traditional database and software techniques; such Big Data is noted for its volume, varieties of data types, and rapid accumulation. IBM estimates that 2.5 quintillion bytes of data are created daily, and that 90% of the data being used in the world today was generated in the past couple of years [2]. The advent of Big Data poses considerable enterprise challenges: what portion of this Big Data get stored; how is this storage managed; how quickly can this Big Data be analyzed to enable enterprises to take quick action on enhance productivity, meet or change directions, reduce risk, and more; how can this data be stored securely; what data privacy issues are involved; and so on [2].

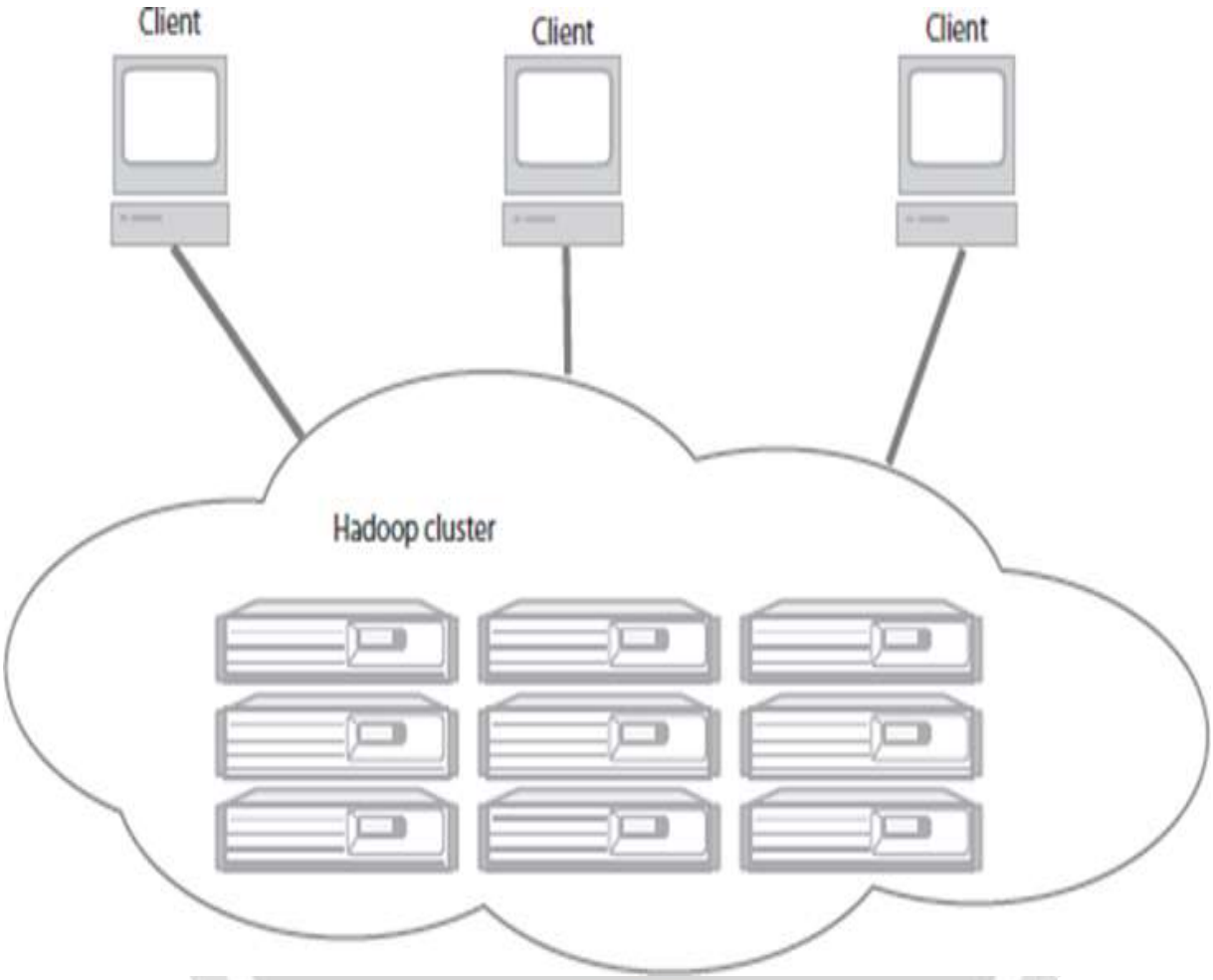
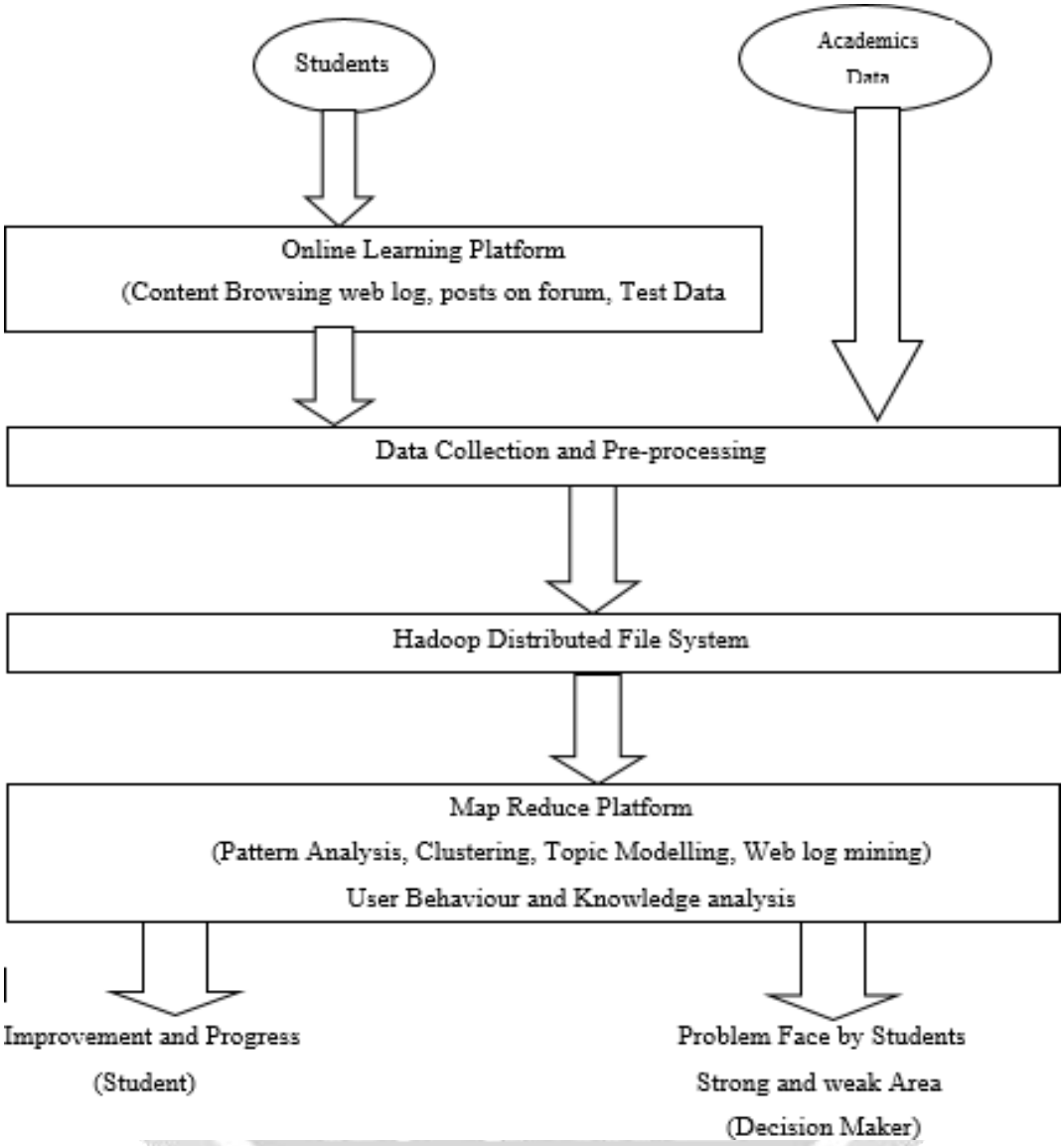


Fig- 2: Hadoop Cluster

The solution to deals with big data problem is Hadoop distributed processing platform. At the center of the Hadoop is Map Reduce framework and Hadoop distributed file system. MapReduce is a data processing model. Its greatest advantage is the easy scaling of data processing over multiple computing nodes. Here the data processing primitives are called mappers and reducers. Decomposing a data processing application into mappers and reducers is sometimes nontrivial. But, once you write an application in the MapReduce form, scaling the application to run over hundreds, thousands, or even tens of thousands of machines in a cluster is merely a configuration change [2].

Now if we consider number student pursuing the bachelor degree in universities like Mumbai, Pune, Solapur, and Shivaji in Maharashtra, there is a huge amount of student academic data in universities [7]. And now days various universities have started online course so these can produce huge amount of data and mostly that data are in various forms like relation data in tabular format, web site logs and forum data are semi structure format. So various researches like research center of Microvity private ltd. have started treating education data as big data problem to improve the education quality. Below is the system architecture of Microvity education analysis platform.



.

Fig- 3: Microvity Education Platform

4. CONCLUSIONS

1) Thus we can do the survey of various research in big data in education data mining. Our survey system will help the student to find solution to their problems.

2) We have survey systems that will improve the students and our suggestion will help to keep student interested in learning and explaining new topic.

3) We have survey systems that will also help the education decision maker by providing key information about student so they can take appropriate measure to improve education quality.

5. REFERENCES

[1] Alana M. de Morais, Joseana M.F.R. Araujio and Evandro B. Costa, "Monitoring Student Performance Using Data Clustering and Predictive Modeling", IEEE 2014. J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[2] Zhi-Hua Zhou, Nitesh V. Chawla, Yaochu Jin, and Graham J. Williams, "Big Data Opportunity And Challenges:

[3] Discussion From Data Analysis Prespective", IEEE Compitational intelligence magazine, Nov. 2014

[4] Xin Chen, Mihaela Vorvoreanu and Krishna Madhavan, "Mining Social Media for Understanding Students Learning Experience", IEEE Transaction on Learning Technologies, Vol. 7, No. 3, Jul.-Sep., 2014.

[5] "Towards on Understanding of ECE Students Online Home Help Forum", IEEE, 2014.

[6] Hadautho Roberto Barros da Silva and Paulo Jorge Leitao Adeodato, "A Data Mining Approach for Undergraduate Students Retention", IEEE WCCI, Jun. 2012.

[7] Xin-hua Zhu, Ya-qiong Deng, Qing-ling Zeng, "Analysis on Course Grade of College-wide Examination Based on Mixed Weighted Association Rules Mining Algorithm", IEEE International Conference on Computer Application and System Modeling,2010.

[8] Asma A. Al-shargabi and Ali N. Nusari, "Discovering Vital Patterns from UST Students Data by Applying Data Mining Techniques", IEEE 2010.

[9] Ai Yubing and Yubing I, Zhang Jianping, "Application of Data Mining in Distance Learning Evaluation", IEEE International Forum on Information Technology and Application, 2010.

[10] Rajnish Dass, "Application Rule Mining for Behavioral Analysis of School Students: A Case from India", 2nd Hawaii International Conference on System Science, 2009.