ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Biology, Engineering, Science and Technology (IJARBEST)
Vol.2, Issue.2, February 2016

Exploring Educational Dataset using Data Mining Technique

Dr. M. Thangamani[1], T.Selvakumar[2]
[1]Assistant Professor, Kongu Engineering College, Tamilnadu, India
[2]ECE Final Year, Kongu Engineering College, Tamilnadu, India

***Abstract*—**Educational Data Mining is a research field related with the application of data mining, statistics and machine learning to information generated from educational settings from universities and colleges. This paper is a exploration of the various concepts employed in educational data mining whose outcomes are highly advantageous to the students for their performance to be upgraded in a systematized manner. Many diverse approaches such as clustering, decision tree algorithm and association rule mining have been discussed which highly focuses on the furtherance of the student's academic betterment.

Index Terms— Data mining, Clustering, Decision Tree Algorithm, MOOC

I. Introduction

Data mining is a term which refers to the processing of huge amount of database called as big data which cannot be processed very easily. Nowadays, since the amount of data available in any institutions is so high, it is very difficult to process them. Due to this, the challenges faced are analyzing the data, capturing, visualizing the data, making secure environment for the processed data. Analysis of data sets is so tedious that it needs a secure and efficient tool like hadoop and cloudera. Educational big data are the data collected from universities which are taken for mining so that the prediction of the future learning behavior of the students can be done very easily.

II. Related Works

Jose del Campo-Avila et.al [1] presented student models from the data acquired by a specific Web-based tool for adaptive testing using data mining. In this case, many comparisons have been carried out and relations have been found between the continual assessment during the semester and the final evaluation for the student's improvement. Machine learning method is mostly used in big data analytical systems nowadays. This has resulted in the penetration of Big Data technologies and tools into education to process the huge amount of data involved [2]. The field of data mining technique can enhance the educational resource organization mode, because the students use it easily and rapidly. In this case [3], data mining helps in gathering Multi-level knowledge, and also gives out an excellent way in Internet educational resource sharing.

Syed Tanveer Jishan states [4] that the exactness of the prediction model upgrades on the application of discretization and oversampling methods. The author concludes that the combined use of oversampling and optimum equal width binning will increase the precision of the student's grade prediction in the final semester. Educational data mining helps to overcome the problem of predicting student's performance. The various classification methods and clustering methods have been used with the help of WEKA software [5] for the improvement of the learning skills of the students. The main advantage is that the timing needed for generating clusters is very less. Nowadays, a large set of data is collected from the medical database. There may be some relations between each data. A new algorithm has been developed [6] which depends on C4.5 to perform the process of mining data for medicine applications and the proposed algorithm is checked with two datasets. The result proves to be an efficient one. However, the disadvantage is that it takes a large amount of time for the computational process.

Educational data mining is performed at different levels for school students, University students, etc. An ideal career has been introduced [7] for ideal students and then it is compared with the generic student using different techniques. Educational data mining plays a vital role in improving the learning ability of the students [8]. In this case, many non-statistical methods have been used to explain the gathered information and the emerged value of Learning Analytics/Educational Data Mining has been taken into consideration. The main advantage of this method is that both the teaching and technical improvements can be achieved using this method. Data mining not only helps the engineering college students [9] but also the polytechnic students. In this paper, the polytechnic students have been considered and they have been analyzed with their family backgrounds such as father's occupation, mother's occupation, etc. In addition to this, their academic background, i.e., the percentage of marks in SSC and their mode of admission are taken into consideration. With all these parameters, the status of any student can be found out easily. If a particular student is found to be at risk, he will be trained to the maximum level and eventually, his grade will be increased. Educational data mining is the most powerful tool for enhancing the skill set of the students [10]. In this case, there are two techniques applied. The first focuses on the record of the EDM developments; the second focuses on analyzing the content which depends on the result from the mining approach. i.e., strength, weakness, opportunities and threats (SWOT) analysis has been made. Thus, analyzing all these things helps in improving the student's performance.

Normally, educational institutions need to use different methods to improvise the learning skills of the students [11]. In this paper, four major areas have been focused. They are: User modeling, User grouping or profiling, Domain modeling and Trend analysis. The benefits of this approach are that the decision makers can suggest a proper solution for a particular problem of the student if the same needs to be improved because each student will have different kinds of problems. Nowadays, e-learning is increased at leaps and bounds compared to book learning. In this paper, the similarities and differences between Educational Data Mining and Learning Analytics have been evaluated and the Massive Open Online Course (MOOC) has been suggested [12] as an innovative method because in MOOCS, data are gathered from hundreds of students and they are used for reframing the course materials which will help for the students who will come to learn in the future.

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Biology, Engineering, Science and Technology (IJARBEST)
Vol.2, Issue.2, February 2016

This paper investigates the application of data mining in the field of education and helps the students to choose the competitive course and pursue the business [13]. Here, neural network has been used to classify the large complex data. The application of data mining technique finds its path in Students' Enrolment Prediction, Predicting Students' Profiling, Curriculum Development, etc. Online instructors and e-learning administrators will be benefited through this technique. The most important data mining techniques [14] used are statistics, visualization, classification, clustering and association rule mining.

In the college examination, GPA is the most important factor that determines the learning skill of any student. In this paper, Mashael A. Al-Barrak and Muna Al-Razgan proposed a model by taking student's data including their final GPA and J48 decision tree algorithm has been applied to find out the classification rules. As a result, better improvement has been achieved in predicting the student's final GPA [15]. In the field of data mining techniques, many algorithms such as clustering algorithm, classification algorithm, prediction algorithm are being applied [16]. But in the case of educational dataset, the usage of these algorithms is completely low. So in this paper, all these algorithms have been combined. Algorithms such as k-means clustering algorithm and C-means clustering algorithm have been analyzed in detail. These algorithms have been used to predict the final GPA of a student. Educational data mining is achieved with the help of many kinds of algorithms such as clustering and decision algorithms [17]. In this paper, both the techniques have been applied combined. Here, the marks of the students from the university examination have been taken and then applied both the algorithms. Finally, the result shows that the decision tree which was constructed using J48 algorithm provides better results than the clustering k-means algorithm. Data mining comes out as a big tool for analytics.

It is applied in many areas. This paper focuses on how to use this data mining technique efficiently in the educational side. More and more comparisons have been carried out and the performance is varying from one method to another.

III. Conclusion

Educational data mining is the best method for improving the performance of the students. The big data analytics applied for educational dataset proves to be of immense use for the benefit of the students. Various techniques applied in the data mining technique such as decision tree are highly efficient in analyzing the educational big data.

References

[1] José del Campo-Ávila, Ricardo Conejo, Francisco Triguero, Rafael Morales-Bueno, Mining Web-based Educational Systems to Predict Student Learning Achievements, International Journal of Artificial Intelligence and Interactive Multimedia, Vol. 3, No 2. Pp.49-54, 2015.

[2] Katrina Sin and Loganathan Muthu,Application of Big Data in Education Data Mining and Learning Analytics – A Literature Review, ICTACT Journal on Soft Computing, Special Issue on Soft Computing Models For Big Data, Vol. 5, No.4, Pp.1035-1049, 2015.

[3] Rui Chen, Yang Zhao**,** The Network Educational Resources Construction on Meteorology Based on Data Mining, Open Journal of Social Sciences, Vol. 3, Pp.1-5, 2015.

[4] Syed Tanveer Jishan, Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique, Decision Analytics, a Springer Open Journal, Vol. , Pp. 1-25, 2015.

[5] Brahamjit Pannu1, Mr. Puneett Sharma, A Comparative data mining technique on education data, International Journal For Technological Research In Engineering, Vol. 2, Pp. 2124 -2127, Issue 9, May-2015.

[6] Kasra Madadipouya1, A new decision tree method for data mining in medicine, Advanced Computational Intelligence: An International Journal Vol.2, No.3, Pp.31-37, 2015

[7] Renza Campagni, Donatella Merlini , Renzo Sprugnoli, Maria Cecilia Verri, Data mining models for student careers, Expert Systems with Applications, Pp. 5508–5521, 2015

[8] Zacharoula Papamitsiou and Anastasios A. Economides, Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence.

[9] Jyoti Bansode Shah And Anchor Kutchhi, Mining Educational Data to Predict Student's Academic Performance, International Journal on Recent and Innovation Trends in Computing and Communication , Vol. 4, No.1, 2016.

[10] Alejandro Pena-Ayala, Educational data mining: A survey and a data mining-based analysis of recent works, Expert Systems with Applications, Pp.1432-1462, 2014

[11] S. Lakshmi Prabha, Dr.A.R.Mohamed Shanavas, Educational data mining applications, Operations Research and Applications: An International Journal , Vol. 1, No. 1, Pp. 23-29, 2014.

[12] Laura Calvet Liñán1 and Ángel Alejandro Juan Pérez2, Educational Data Mining and Learning Analytics: differences, similarities, and time evolution, Universities and Knowledge Society Journal, Pp.98-112, 2015.

[13] Dr. Mohd Maqsood Ali, Role of data mining in education sector, International Journal of Computer Science and Mobile Computing, Vol. 2, No. 4, Pp.374– 383, 2013.

[14] Cristóbal Romero, Sebastián Ventura, Enrique García, Data mining in course management systems: Moodle case study and tutorial, Computers & Education, Elsevier, 2007.

[15] Mashael A. Al-Barrak and Muna Al-Razgan, Predicting Students Final GPA Using Decision Trees: A Case Study, International Journal of Information and Education Technology, Vol. 6, No. 7, Pp.528-533, 2016.

[16] Ashish Dutt, Saeed Aghabozrgi, Maizatul Akmal Binti Ismail, and Hamidreza Mahroeian, Clustering Algorithms Applied in Educational Data Mining, International Journal of Information and Electronics Engineering, Vol. 5, No. 2, Pp. 112-116, 2015.

[17] Ritika Saxena, Educational Data Mining: Performance Evaluation of Decision Tree and Clustering Techniques using WEKA Platform, HCTL

[18] Open International Journal of Technology Innovations and Research, Vol.14, Pp.1-10, 2015.

[19] Pooja Thakar, Anil Mehta, Ph.D, Manisha, Ph.D , Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue, International Journal of Computer Applications, Vol. 110, No. 15, Pp. 60-68, 2015.

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Biology, Engineering, Science and Technology (IJARBEST)
Vol.2, Issue.2, February 2016

Authors Biography

Dr. M. Thangamani possesses nearly 20 years of experience in research, teaching, consulting and practical application development to solve real-world business problems using analytics. Her research expertise covers Medical data mining, machine learning, cloud computing, big data, fuzzy, soft computing, ontology development, web services and open source software. She has published nearly 70 articles in refereed and indexed journals, books and book chapters and presented over 67 papers in national and international conferences in above field. She has delivered more than 60 Guest Lectures in reputed engineering colleges on various topics. She has got best paper awards from various education related social activities in India and Abroad. She has organized many self-supporting and government sponsored national conference and Workshop in the field of data mining, big data and cloud computing. She continues to actively serve the academic and research communities. She is on the editorial board and reviewing committee of leading research journals, which includes her nomination as the Associate Editor to International Journal of Entrepreneurship and Small & Medium Enterprises at Nepal and on the program committee of top international data mining and soft computing conferences in various countries. She is also seasonal reviewer in IEEE Transaction on Fuzzy System, international journal of advances in Fuzzy System and Applied mathematics and information journals. She has organizing chair and keynote speaker in international conferences in India and countries like Malaysia, Thailand and China. She has Life Membership in ISTE, Member in CSI, International Association of Engineers and Computer Scientists in China, IAENG, IRES, Athens Institute for Education and Research and Life member in Analytical Society of India. She is currently working as Assistant Professor at Kongu Engineering College at Perundurai, Erode District.

Mr. T. Selvakumar is currently pursuing B.E final year in Electronics and Communication Engineering at Kongu Engineering College, Perundurai. He achieved District 2nd in HSc Examination in the year 2012. He has been awarded for the same. He is an active member in IEEE and IETE. He has been placed in Accenture Management Consulting Services Company. He has attended workshops conducted by IEEE. He holds the post of Additional Secretary in ECE Association at Kongu Engineering College. He has attended Teacher Training Programme in Yoga conducted by Temple of Consciousness, Aliyar. He has finished the certified course namely “Yoga for Youth Empowerment”.