Second International Conference on Nexgen Technologies

Sengunthar Engineering College, Tiruchengode, Namakkal Dist. Tamilnadu (India)

8ᵗʰ - 9ᵗʰ March 2019   www.conferenceworld.in   ISBN : 978-93-87793-75-0

EDUCATIONAL DATA MINING

Henny Yashmi.R.S[1], Asfaq Ahamed.A[2], Arun Vadivu.S[3],
Savitha.V[4], Dr.Kalaikumaran[5]

[1,2,3]*Department of Computer Science and Engineering,*

SNS College of Technology, Coimbatore (India)

[4]*Assistant Professor,* [5]*Professor and Head,*

Department of Computer Science and Engineering, SNSCT, Coimbatore (India)

ABSTRACT

Presently educational institutions compile and store huge volumes of data such as student enrolment and attendance records, as well as their examination results. Mining such data yields stimulating information that serves its handlers well. Rapid growth in educational data points to the fact that distilling massive amounts of data requires a more sophisticated set of algorithms. This issue led to the emergence of the field of Educational Data Mining (EDM). Traditional data mining algorithms cannot be directly applied to educational problems, as they may have a specific objective and function. This implies that a preprocessing algorithm has to be enforced first and only then some specific data mining methods can be applied to the problems. One such preprocessing algorithm in EDM is Clustering. Many studies on EDM have focused on the application of various data mining algorithms to educational attributes. Therefore, our project provides over three decades long (1983-2016) systematic literature review on clustering algorithm and its applicability and usability in the context of EDM.Ourproject classifies the student's record based on achievement. Their result is predicted through data mining with the help of the internal marks scored. Future insights are outlined based on the literature reviewed, and avenues for further research are identified.

Keywords: Data Mining,Extraction,Fuzzy Classification, Fuzzy Clustering

I. INTRODUCTION

Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. The biggest challenge is to analyse the data to extract meaningful information that can be used to solve a problem or for the growth of the business. There are powerful tools and techniques available to mine data and find insights from it.The data mining process breaks down into five steps. First, organizationscollect data and load it into their data warehouses. Next, they store and manage the data, either on in-house servers or the cloud. Business analysts, management teams and information technology professionals access the data and determine how they want to organize it. Then, application software sorts the data based on the user's results, and finally, the end user presents the data in an easy-to-share format,

Second International Conference on Nexgen Technologies

Sengunthar Engineering College, Tiruchengode, Namakkal Dist. Tamilnadu (India)

8th - 9th March 2019   www.conferenceworld.in   ISBN : 978-93-87793-75-0

such as a graph or table.Data mining serves the primary purpose of discovering patterns among large volumes of data and transforming data into more refined/actionable information

II.BACKGROUND

In[1], *Pauziah Moud Arshad, Usamah Binmat, Norlida Buniyamin "Educational Data Mining for prediction and classification of student's achievement"*

This paper highlights the importance of using student data to drive improvement in education planning. It then presents techniques of how to obtain knowledge from databases such as large arrays of student data from academic Institution databases. Further, it describes the development of a tool that will enable faculty members to identify, predict and classify students based on academic performance measured using Cumulative Grade point average (CGPA) grades.

In [2], *Maizatul Akmar Ismail, Tutut Herawanand Ashish Dutt "A Systematic Review on Educational Data Mining"*

This paper has also outlined several future insights on educational data clustering based on the existing literatures reviewed, and further avenues for further research are identified.The key advantage of the application of clustering algorithm to data analysis is that it provides relatively an unambiguous schema of learning style of students given a number of variables like time spent on completing learning tasks, learning in groups, learner behavior in class, classroom decoration and student motivation towards learning. Clustering can provide pertinent insights to variables that are relevant in separating the clusters.

In[3],*Vandhana Dahiya's work survey on educational datamining*, the proposed solution explains the objective of educational data mining into 3 categories educational, administrative, commercial.

In[4]*, Christobal Romero and Sebastian Ventura's data mining in education*, the proposed system introduces and reviews key milestones and the current state of affairs in the field of EDM, together with specific applications, tools, and future insights.

In[5]**,***Rajni Jindal and Malaya Dutta Borah's survey on educational data mining and research trends*,this survey work focuses on components, research trends (1998-2012) of Educational Data Mining highlightingits related tools, techniques and outcomes.

III. PREDEFINED SYSTEM

In Existing system, the Clustering and Classification Algorithms in Education Data Mining Process are used. In Clustering process c-means , K-means and classification process SVM classification or naive bayes Algorithms are used for clustering and classification process. After the Clustering process is completed the clustered input is passed to the Classifier process. It is an unsupervised approach for analyzing data in statistics, machine learning, pattern recognition, DM, and bio informatics. It refers to collecting similar objects together to form a group or cluster. Each cluster contains objects that are similar to each other but dissimilar to the objects of other groups. This approach when applied to analyze the dataset derived from educational system is termed as Educational Data Clustering (EDC). An educational institution environment broadly involves three types of actors namely Teacher, student and the environment. Interaction between these three actors generates voluminous data that can systematically be clustered to mine invaluable

Second International Conference on Nexgen Technologies

Sengunthar Engineering College, Tiruchengode, Namakkal Dist. Tamilnadu (India)

8th - 9th March 2019   www.conferenceworld.in   ISBN : 978-93-87793-75-0

information. Data clustering enables academicians to predict student performance, associate learning styles of different learner types and their behaviours and collectively improve upon Institutional performance. Researchers, in the past have conducted studies on educational datasets and have been able to cluster students based on academic performance in examination

IV. PROPOSED SYSTEM

In proposed system, the Fuzzy Classification and Clustering Algorithm for Educational Data Mining Process is used. Fuzzy clustering (also referred to as soft clustering) is a form of clustering in which each data point can belong to more than one cluster.

In Fuzzy Classification process it based on Student all semester marks details, total counting and College overall result. It gives the Classification result. This Classification result based Fuzzy Graph generate process it based on Education Department Based

4.1   ADVANTAGES

- ✓ Mining process speed is high
- ✓ Educational datas can be easily accessed
- ✓ The preprocessing speed is high
- ✓ This process fully concentrates on educational purpose

V. OVERVIEW OF THE SYSTEM

Input design is the process of connecting the user-originated inputs into a computer to used formats. The goal of the input design is to make data entry Logical and free from errors. Errors in the input database controlled by input design this application is being developed in a user-friendly manner. The forms are being designed in such a way that during the processing the cursor is placed in the position where the data must be entered. An option of selecting an appropriate input from the values of validation is made for each of the data entered. Concerning clients comfort the project is designed with perfect validation on each field and to display error messages with appropriate suggestions. Help managers are also provided whenever user entry to a new field he/she can understand what is to be entered. Whenever user enter a error data error manager displayed user can move to next field only after entering a correct data

Computer output is the most important and direct source of information to the user. Efficient intelligible output design should improve the system's relationship with the user and admin in decision-making. Output design generally refers to the results generated by the system. For many end users on the basis of the output the evaluate the usefulness of the application. Efficient software must be able to produce and efficient effective reports

VI. MODULES

- ➤ *ADMIN LOGIN*
- ➤ *LOAD DATA/ PREPROCESSING*

Second International Conference on Nexgen Technologies

Sengunthar Engineering College, Tiruchengode, Namakkal Dist. Tamilnadu (India)

8ᵗʰ - 9ᵗʰ March 2019   www.conferenceworld.in   ISBN : 978-93-87793-75-0

> *FUZZY CLASSIFICATION AND*
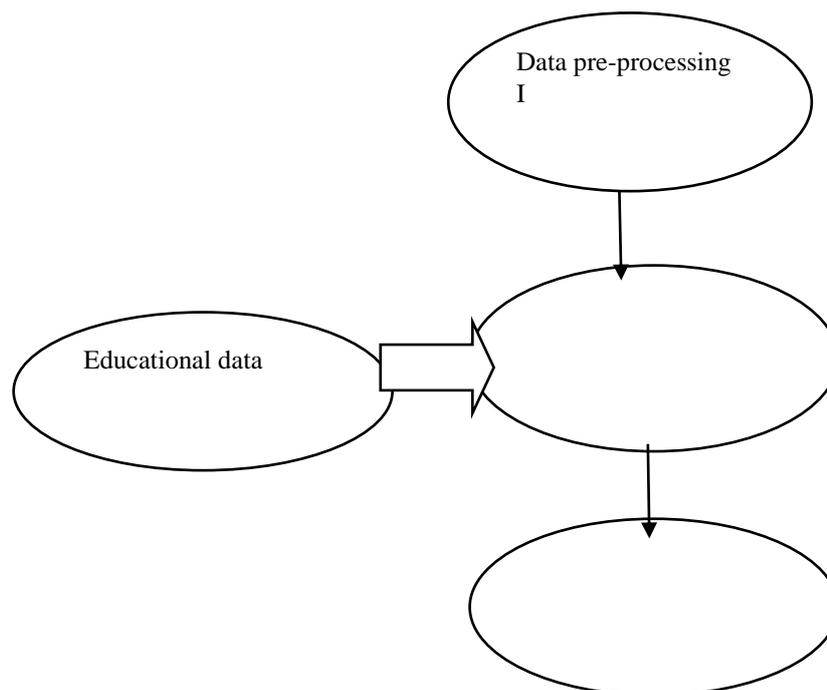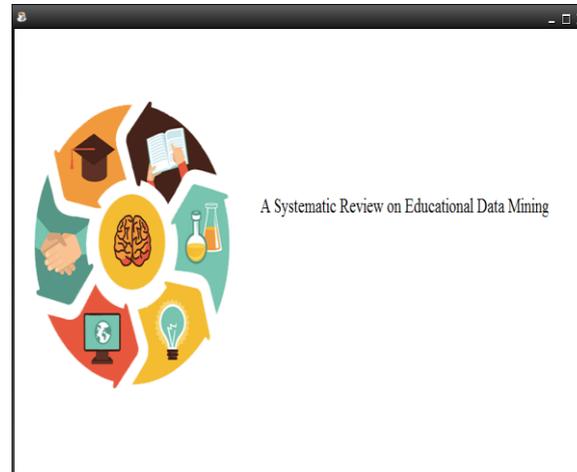> *CLUSTERING*
> *CLASSIFICATION REPORT*
> *GRAPH REPORT*

6.1 MODULE DESCRIPTION

6.1.1 Admin Module

In this admin login page provide the Process security the admin is handle in overall process so we implement the admin login security it avoid the un authentication person.

If the correct user was the login it send the successful message else it show the alert message.

6.1.2 Load Data/ Preprocessing

In this module, the admin may select the any Education Dataset (2013,2014,2015) in this dataset many type of attributes are there for Ex(Student department, Semester result ,Register/Roll

Fig 6.1 preprocessing

6.1.3   Fuzzy Classification and Clustering

Fuzzy clustering (also referred to as soft clustering) is a form of clustering in which each data point can belong to more than one cluster.Clustering or cluster analysis involves assigning data points to clusters such that items in the same cluster are as similar as possible, while items belonging to different clusters are as dissimilar as possible. Clusters are identified via similarity measures these similarity measures include distance, connectivity, and intensity Different similarity measures may be chosen based on the that

Second International Conference on Nexgen Technologies

Sengunthar Engineering College, Tiruchengode, Namakkal Dist. Tamilnadu (India)

8ᵗʰ - 9ᵗʰ March 2019   www.conferenceworld.in   ISBN : 978-93-87793-75-0

Fuzzy classification is the process of grouping elements into a fuzzy set whose membership function is defined by the truth value of a fuzzy propositional function. After the Classification process is completed the data was stored in department, class ,total ,average wise.This data's are stored in databases we calculate the over all class count for each department and total year college student education level report.a or the application.
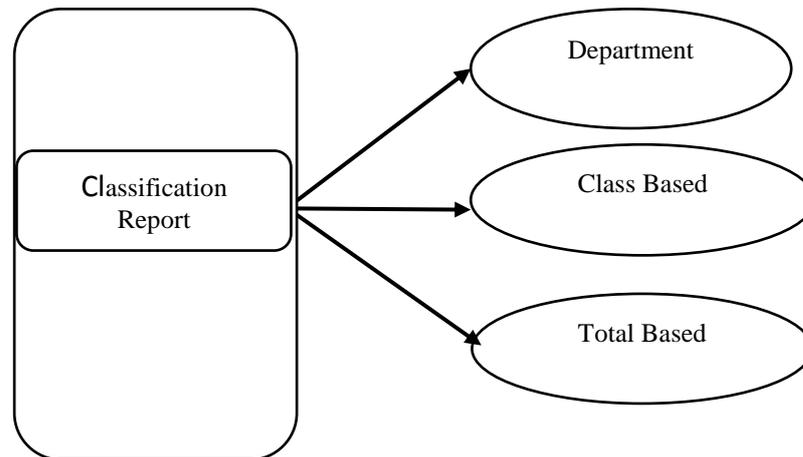
Fig 6.2 fuzzy classification

6.1.4   GRAPH REPORT
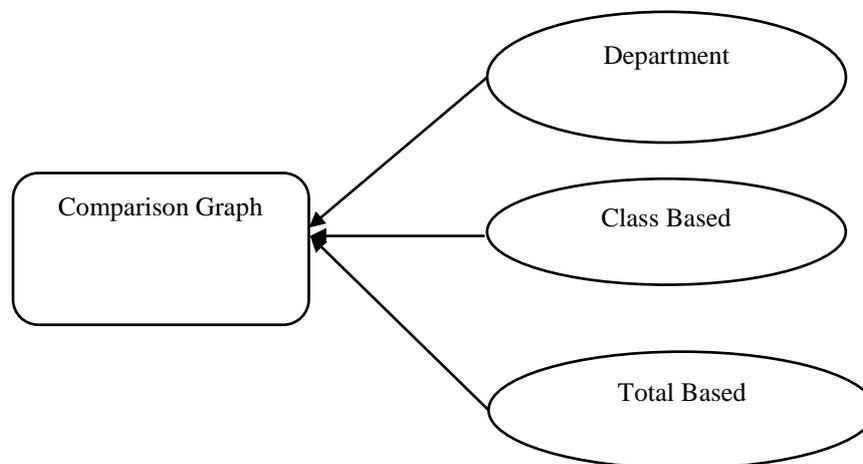
Fig 6.3 Graph Report

Finally the classification report is completed it generate the Graph for based on Student department, class, total average, semester result

Second International Conference on Nexgen Technologies

Sengunthar Engineering College, Tiruchengode, Namakkal Dist. Tamilnadu (India)

8th - 9th March 2019   www.conferenceworld.in   ISBN : 978-93-87793-75-0

VII. UPSHOTS

Fig : 7.1 Home page

Fig : 7.2 Login module

Fig 7.3 Department wise classification

Second International Conference on Nexgen Technologies

Sengunthar Engineering College, Tiruchengode, Namakkal Dist. Tamilnadu (India)

8th - 9th March 2019   www.conferenceworld.in   ISBN : 978-93-87793-75-0

Fig 7.4 Mark wise Classification

Fig 7.5 Result

Second International Conference on Nexgen Technologies

Sengunthar Engineering College, Tiruchengode, Namakkal Dist. Tamilnadu (India)

8th - 9th March 2019   www.conferenceworld.in   ISBN : 978-93-87793-75-0
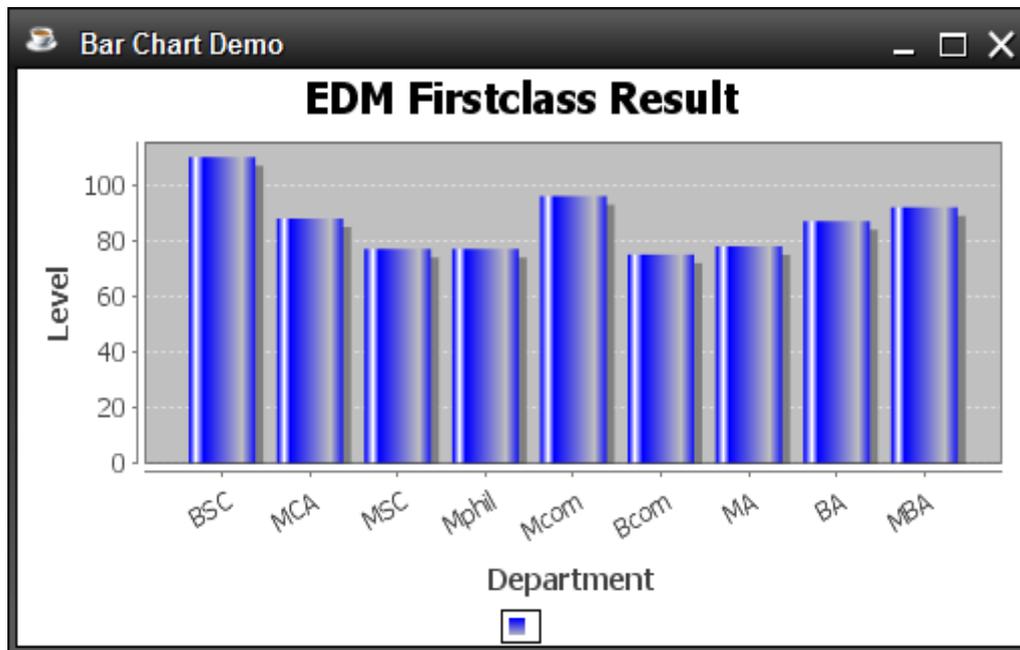
Fig 7.6 First Class Graph

VIII. CONCLUSION

In our Existing system following the Clustering and Classification Algorithms in Education Data Mining Process In Clustering process c-means , K-means and classification process SVM classification or naive bayes Algorithms are used for clustering and classification Process After the Clustering process is completed the clustered input passing the Classifier process It is an unsupervised approach for analyzing data in statistics, machine learning, pattern recognition, DM, and bio informatics. It refers to collecting similar objects together to form a group or cluster. Each cluster contains objects that are similar to each other but dissimilar to the objects of other groups. This approach when applied to analyze the dataset derived from educational system is termed as Educational Data Clustering (EDC). An educational institution environment broadly involves three types of actors namely Teacher, student and the environment. Interaction between these three actors generates voluminous data that can systematically be clustered to mine invaluable information. Data clustering enables academicians to predict student performance, associate learning styles of different learner types and their behaviors and collectively improve upon Institutional performance. Researchers, in the past have conducted studies on educational datasets and have been able to cluster students based on academic performance in examinations.So it was one of the drawback of to accessing educational data's.To accessing educational data's the mining process may slow.The basic concepts and basic algorithms may only using the educational mining process

REFERENCES

[1]   Amershi, S., Conati, C. and Maclaren, H., (2006) "Using Feature Selection and Unsupervised Clustering to Identify Affective Expressions in Educational Games", in Proc Of The Intelligent Tutoring Systems Workshop on Motivational and Affective Issues. pp. 25-30.

[2]   Amershi, S., and Conati, C., (2009) "Combining unsupervised and supervised classification to build user models for exploratory learning environments" Journal of Educational Data Mining.Vol.1, No.1,pp. 18-71.

[3]   Mercheron, A., and Yacef, K. (2005), "Educational Data Mining: a case study" in Proc. Conf. on Artificial Intelligence in Education Supporting Learning through Intelligent and Socially Informed Technology. IOS Press, Amsterdam, The Netherlands, pp. 467-474.

[4]   Al-shargabi, A.A. and Nusari, A. N (2010), "Discovering Vital Patterns From UST Students Data by Applying Data Mining Techniques", in Proc. Int. Conf. On Computer and Automation Engineering, China: IEEE, 20102,547-551.DOI:10.1109/ICCAE.2010.5451653.

[5]   Badri, M., and El Mourad, T (2011) "Measuring job satisfaction among teachers in Abu Dhabi:
design and testing differences" in Proc.Int. Conf. on NIE, 4th Redesigning Pedagogy.Singapore.

[6]   Baker, R.S, Corbett, A.T., Koedinger, K.R (2004) , "Detecting Student Misuse of Intelligent Tutoring Systems" in Proc. Lecture Notes in Computer ScienceVol.3220,531-540.

[7]   Baker, R.S.J.D.,and Yacef, K.(2009), "The state of Educational Data Mining in 2009:A review and future vision" Journal of Educational Data Mining, Vol.1,No. 1,pp.3-17.

[8]   Nelson,B., Nugent, R., Rupp, A. A.(2012), "On Instructional Utility, Statistical Methodology, and the Added Value of ECD: Lessons Learned from the Special Issue", Journal of Educational Data Mining.Vol.4, No.1,pp.224-230.

[9]   Baradwaj,B.K., and Pal,S.(2011), "Mining Student Data to Analyze Students' Performance".International Journal of advanced Computer Science and applications.2,6International Journal of Database Management Systems (IJDMS) Vol.5, No.3, June 2013

[10]   Aggrawal,C.C, and Yu, P.S.(2009), "A Survey of Uncertain Data Algorithm and Applications" IEEE Transactions on Knowledge and Data Engineering, Vol.21,No. 5,pp.609-623.

[11]   Lee, C.H., Lee, G., Leu, Y.(2009), "Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning" .Expert Syst. Appl. J.Vol.36, pp.1675-1684.

[12]   Chrysostomu K. el al.(2009), "Investigation of users' preference in interactive multimedia learning systems: a data mining approach",Taylor and Francis online journal Interactive learning environments. Vol. 17,No. 2.

[13]   Conati,C., Muldner,K., and Carenini G.,(2006)., ".From Example Studying to Problem Solving via Tailored Computer-Based Meta-Cognitive Scaffolding: Hypotheses and Design", Technology, Instruction, Cognition, and Learning - Special Issue on Problem Solving Support in Intelligent Tutoring System.Vol.4, No.1-54.

Second International Conference on Nexgen Technologies

Sengunthar Engineering College, Tiruchengode, Namakkal Dist. Tamilnadu (India)

8th - 9th March 2019   www.conferenceworld.in   ISBN : 978-93-87793-75-0

[14]   Alberg, D., Last, M., and Kandel, A (2012), "Knowledge discovery in data streams with regression tree methods" John Wiley & Sons, Inc,2,pp.69-78,DOI:10.1002/widm.51

[15]   Dey,D., and Sarkar,S.(2002) "Generalized Normal Forms for Probabilistic Relational Data" IEEE Transactions on Knowledge and Data Engineering, Vol.14,No.3,pp. 485-497.DoI:10.1109/TKDE.2002. 1000338.

[16]   Perera,D. et al.(2009), "Clustering and sequential pattern mining of online collaborative data", IEEE Transactions on Knowledge and Data Engineering,Vol.21, No.6,pp.759-772.