ISSN:0975-9646

Study of Data Mining in Higher Education-A Review

Khushbu N.Shah[1], Misika R. Patel[2], Neha V. Trivedi[3], Priyanka N. Gadariya[4], Richa H. Shah[5], Ms. Nehal Adhvaryu[6]

Department of Computer Science and Technology,
Uka Tarsadia University Bardoli, Gujarat, India.

***Abstract*-Education sector widely used conceptual tools of "Data Mining" in prediction of student's performance. In this paper the challenges that are faced by higher education sector are discused and from that we will try to predict student's performance using tools. In this paper we also organize data using weka data mining tool which helps to recognize the data.**

***Keywords*-Classification, Clustering, Data Mining, Decision Tree, Education Data Mining (EDM), Higher Education.**

I. INTRODUCTION

Data mining is a method of extracting hidden predictive information from large databases. It is a new technology with good potential to help Universities or institutions. It is used to focus on the most vital information in their data warehouses. Figure 1 shows how we can create knowledge from data warehouse using data mining.

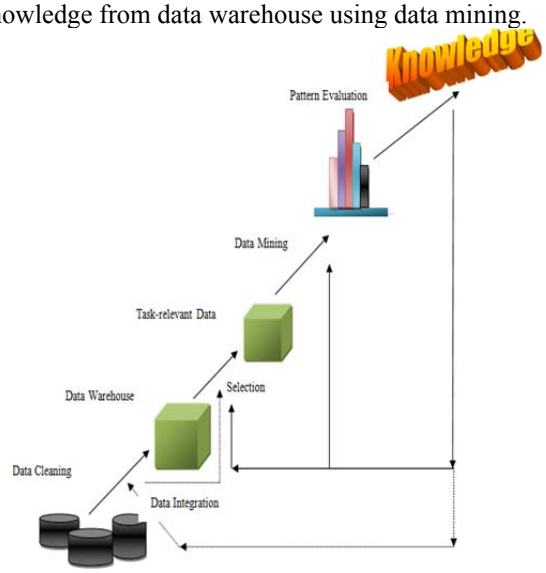

Fig. 1 Steps of creating knowledge using Data Mining

Using data mining techniques, university can predict the result of students. The university can use this information to focus on poor students, who have more risk of failure by giving intellectual features. The actual task of data mining is to examine huge quantities of data in order to extract previously unknown patterns such as cluster analysis, anomaly detection and association rule mining.

These patterns are used in additional analysis. Data mining tasks can be categorized as

- **Anomaly detection:** Identification of unusual data or data errors which require for further investigation.
- **Association rule learning:** Searches for link between variables.
- **Clustering:** Task of discovering groups in the data that have similar characteristic in some way with each other.
- **Classification:** Task of generalizing identified structure to apply for new data.
- **Regression:** It attempts to find a method which builds the data with very few error.
- **Summarization:** It provides more solid representation of the data set. It includes task like visualization as well as report generation of data.

II. RELATED WORK

A. *Literature Review*

In this report Author investigates the perceptions of Knowledge Management within Higher Education , and presents the nature of academics and universities. It focuses on two aspects of the case study – the characteristics of universities and academics that support the implementation of KM, and the perceptions of Knowledge Management and its challenges for implementation within the higher education sector [4].

In many domains Data Mining techniques are important and from that Data Mining techniques are used to improve the ability of higher education institution. If Data Mining Techniques like clustering, decision tree and association are applied to higher education process, it would help to improve in students life cycle management, their performance, as well as in selection of courses etc. In this report author had given brief introduction on data mining techniques like cluster analysis, decision tree, factor analysis, regression analysis [10].

In this report Author had used Classification technique to evaluate student performance. There are many approaches that are used for classification and from that author had used decision tree method. Information's like student's attendance in classroom, seminar, class test, and assignment mark of student were collected from the student's management system, to predict the student performance at the end of the semester [12].

Author had used J48 Algorithm for predict student's performance and for this task classification technique is used in this report. J48 algorithm is used to classifies the data in the form of decision tree and using this decision tree we can easily identify the weak student [15].

In this report Author discussed about which data mining techniques can be applied in the field of education and to identify which data mining technique is suited for

what kind of application in the form of a conceptual model. For Example Classification technique is better to predict student's performance [18].

In this report author survey positive relationship student's performance and the nature of the university. Author had select student using cluster sampling technique. This technique is used for grouping the students [7].

In this paper author had overcome problem in ID3 Algorithm and developed weighted ID3 Algorithm. ID3 algorithm is one of the famous algorithm to generate decision trees. ID3 algorithm has a fault that it disposed attributes with many values. So, Author overcomes this fault by using gain ratio. In this process author gives weight to each attribute at every decision making point [19].

Author conducted study on the student by using classification technique on category, language and background qualification and it was found that whether new comer students will performer or not. Author conducted a comparative study to find out private tutoring percentage in different countries. Author used decision tree model to predict the final grade of students [20].

Higher education teachers are often interested in prediction of student's result before their exams. During classes they try to predict performance of students. For this problem author described that some data mining techniques like neural networks may not be able to accomplish the learning task as small datasets and it can't provide enough data to fill the gaps between too small samples [21].

In this report author discussed on quality education. It does not mean production of high level knowledge but meaning of quality education is that education is produced to students in well-organized manner so they can learn without any kind of problem [23].

III. METHODS AND ALGORITHAMS

A. *Methods*

1) *Cluster Analysis:* Clustering is the process of assigning a set of objects, so that the objects in the same cluster are more similar to each other. Application of clustering in educational sector can help institutes group individual student into class of similar performance.

2) *Classification:* Classifcation is a process of mining patterns that can classify future data into known classes.

3) Factor Analysis: It is one kind of statistical method used to describe inconsistency among observed, correlated variables. It has two types, Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA). EFA is used to uncover the underlying structure of variables. In CFA indicator selection of variables are done on the basis of previous theory.

*4) Regression Analysis:*It includes techniques for modeling and analyzing several variables. Regression analysis estimates the conditional probability of the dependent variable with respect to independent variables. It is mostly used for prediction. It explores relationship between independent variables and dependent variables. There are two methods widely used are linear regression and ordinary least squares regression.

B. *Algorithms*

*1) J48 Algorithm:*It is used for classifying and prediction. For the classification technique J48 algorithm was chosen (based on the C4.5 algorithm from the machine learning), for being one of the most used Weka tool that offers a more stability between precision, speed and interpretability of results. It classifies data in the form of decision tree and using this decision tree we can easily identify the weak students. The classification learning was also used to predict the student's result.

2) ID3 Algorithm: ID3 stands for **Iterative Dichotomiser 3.** It is one kind of Mathematical Algorithm used to generate a decision tree. It will generate the tree from the top down, with not any backtracking.

Advantage of ID3

- It is used to build tree very quickly.
- It will generate a short tree.
- Only need to test attributes until all data is classified.
- It will reduce number of tests.

TABLE 1 Cluster analysis methods

Name	**Description**
Partitioning	Suppose we create one database with n objects, now this method construct k which indicates partition of data. Every partition will present a cluster and k≤n. It will classify the data into groups .
Hierarchical	This method is used to create the hierarchical decomposition of data objects and we can classify hierarchical method on the basis of how the hierarchical decomposition.
Density-based	This method is based on the view of density. The basic idea for this method is to carry on cluster analysis as long as the density in the neighborhood object exceeds some threshold.
Grid-Based	In this method the objects will form a grid. The space of object is quantized into limited number of cells which form a grid structure.
Constraint-based	In this method the clustering is performed on the bases of incorporation of user or all application oriented constraints.

Decision Tree

Decision Tree is one kind of decision support tool that uses a tree like graph of decisions including its utility. Decision Trees are commonly used in research, specifically in decision analysis, for recognize the strategy as well as to reach a purpose. Another use of decision trees is for calculating conditional probabilities. Figure 2 shows a simple decision tree to understand result even with small data.

Advantages:

- It does not need domain knowledge.
- It is easy to understand by human.
- Classification steps of decision tree are very simple and quick.
- Provide a clear mention of which fields are most important for prediction.

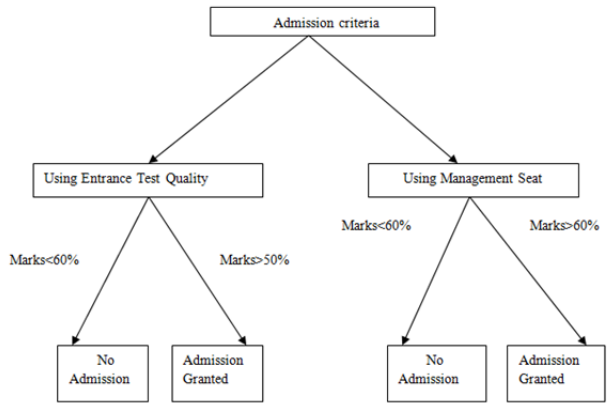

Fig. 2 Decision Tree based upon Admission Criteria

IV. Application Of Data Mining In Higher Education

Data Mining is used in different types of application area like Banking, Security Applications, Educational Data Mining, Agriculture etc. The main concerned are is in educational systems.

Educational Data Mining (EDM) is concerned with developing methods for exploring the distinctive type of data that come from educational setting, and we can use those methods for better understanding in educational system. There are two kind of key area in EDM. First is Mining Student's performance and another one is mining enrollment data. Techniques of data mining like classification, clustering, categorization are used in prediction of student's performance, in study and visualization of data, in development of educational sector etc. Study and Visualization of Data is used to show up useful information and its support in decision making. For this task two techniques are used, Statistics and Visualization. Statistics include collection, analysis, interpretation and presentation of data. Visualization uses graphic techniques for better understanding and to analyze data in educational system.

V. Conclusion And Future Work

The current educational system does not involve any kind of prediction about student's result based in their performance. The system does not deal with failure student. In current situation it doesn't analysis the student. Another common problem in universities some students may realize lost in the crowd. Whether they are struggling to find help with coursework or having difficulty in choosing the appropriate courses they need. The proposed model identifies the students and then it helps teacher to act before a student drops for recourse allocation with confidence gained from knowing how many students are likely passed or fail.

It is observed that current DM tools are too complex for educators to use. One possible solution is the development of tools that use a default algorithm for each task and secondly, DM tool has to b integrated into the e-learning environment so that results generated with DM techniques could be easily applied.

References

[1] http://www.knowledge-management-tools.net/knowledge-information-data.html Date:9/10/2014

[2] University press(India) Private Limited (2009)."Data Mining Techniques." Universities press, ISBN 978 81 7371 3804, Chapter 1, 2, 3,p. Date:11/12/2014

[3] Morgan Kaufmann Publishers An Imprint of Elsevier Science "Data Mining Concepts and Techniques" ISBN 81-7867-023-2 ,Chapter 7,p.279 Date:11/12/2014

[4] Dr. R .Shukla, "KNOWLEDGE MANAGEMENT IN HIGHER EDUCATION", International Journal of Reviews, Surveys and Research (IJRSR)".

[5] J.Ranjan and S.Khalil , "Application of Knowledge Management in Management Education : A Conceptual Framework" Journal of Theoretical and Applied Information Technology., pp.15-25,2007.

[6] A.Laoufi, S.Mouhim, E.Megder, C.Cherkaoui and D.Mammass , "Using Knowledge Management in Higher Education: Research Challenges and Opportunities" Journal of Theoretical and Applied Information Technology.,vol.31 no.2, pp.100-108,Sep-2011

[7] M.Ramaswami,R.Bhaskaran , "A CHAID Based Performance Prediction Model in Educational Data Mining " , Journal of Computer Science Issues, vol.7 , Issue 1 , No.1 , January 2010

[8] A.Laoufi, S.Mouhim, E.Megder, C.Cherkaoui and D.Mammass, "Using Knowledge Management in Higher Education: Research Challenges and Opportunities" Journal of Theoretical and Applied Information Technology.,vol.31 no.2, pp.100-108,Sep-2011

[9] P.Sinha, M. Arora, N. Mishra, "Framework for a Knowledge Management Platform in Higher Education Institutions", ISSN: 2231-2307, Volume-2, Issue-4, September 2012

[10] M.Goyal, R.Vohra, "Applications of Data Mining in Higher Education", International Journal of Computer Science Issues, Vol.9, Issue 2, No1, March 2012.

[11] S.Kumar, Dr. S. Gupta," Role of Knowledge Management Systems (KMS) in Multinational Organization: An Overview", Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 2, Issue 10, Oct 2012

[12] B.Baradwaj, S.pal, "Mining Educational Data to Analyze Students' Performance" (IJACSA) International Journal Of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011.

[13] C.Janicot, S.Mignon," Knowledge codification in audit and consulting firms: a conceptual and empirical approach", Journal of Knowledge Management Research & Practice 10,pp. 4–15, 2012

[14] L.Abdillah, " MANAGING INFORMATION AND KNOWLEDGE SHARING CULTURES IN HIGHER EDUCATION INSTITUTIONS",Mar-2011

[15] M.Bhullar, A.Kaur, "Use of Data Mining in Education Sector" , Proceedings of the World Congress on Engineering and Computer Science 2012 , Vol I, Oct 2012

[16] J.Denford, Y.Chan, "Knowledge strategy typologies: defining dimensions and relationships", Journal of Knowledge Management Research & Practice ,102–119,2011
[17] V.Kumar, A.Chadha , "An Empirical Study of the Applications of Data Mining Techniques in Higher Education" , Vol. 2, No.3, Mar-2011
[18] S.Bhanap, R. Kulkarni, "Student - Teacher Model for Higher Education System", ISSN : 2279-0535 Volume : Volume –II Issue : Issue-III,2013
[19] S.Dhanda, S.D, R.L, "Predicting Students' Performance using Modified ID3 Algorithm", International Journal of Engineering and Technology, ISSN:0975-4024 , Vol 5 No 3 Jun-Jul 2013
[20] U.Pandey , S.Pal , "A Data Mining View on Class Room Teaching Language" , ISSN:1694-0814, International Journal of Computer Science Issues, Vol. 8, Issue.2, Mar-2011
[21] S.Natek, M.Zwilling, "Data Mining For Small Student Data Set - Knowledge Management System For Higher Education Teachers", Active citizenship by Knowledge Management & Innovation, Management Knowledge and Learning International Conference , pp.19-21 ,June 2013.
[22] Dr. Y. Cader, "Knowledge Management and Knowledge-based Marketing "Journal of Business Chemistry., vol.4 , Issue 2 ,pp.46-58 , May-2007
[23] U.Pandey, B. Bhardwaj and S.pal , "Data Mining as a Torch Bearer in Education Sector", Technical Journal of LBSIMDS.