

A Survey on Analysis the Students Mind in Different Area

Krishna Parmar¹, Huma Khan²

¹PG Scholar (M. Tech, CSE), Department of CSE, Rungta College of Engineering and Technology, Raipur, Chhattisgarh 492009, India

²Assistant Professor (CSE), Department of CSE, Rungta College of Engineering and Technology, Raipur, Chhattisgarh 492009, India

Abstract: Over the past decade there has been a rapid growth in higher education system. A lot of new institutions have come up both from public and private sector offering variety of courses for under graduating and post graduating students. The rates of enrolments for higher education has also increased but not as much as the number of higher institutions are increasing. It is a concern for today's education system and this gap has to be identified and properly addressed to the learning community. Hence it has become important to understand the requirement of students and their academic progression. Educational Data Mining helps in a big way to answer the issues of predictions and profiling of not only students but other stake holders of education sectors. This paper discusses the application of various Data Mining tools and techniques that can be effectively used in answering the issues of predictions of student's performance and their profiling. This study explores multiple factors theoretically assumed to affect students' performance in higher education, and finds a qualitative model which best classifies and predicts the students' performance based on related personal and social factors. Educational Data Mining refers to techniques, tools, and research designed for automatically extracting meaning from large repositories of data generated by or related to people's learning activities in educational settings.

Keywords: Data Mining, Educational Data Mining, Prediction, Profiling Data Mining, Clustering, Relationship Mining, Education, Students, Performance, Patterns, Classification

1. Introduction

It is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings. It is important to study and analyse educational data especially students' performance. Educational Data Mining (EDM) is the field of study concerned with mining educational data to find out interesting patterns and knowledge in educational organizations.

The main objective of higher education institutions is to provide quality education to its students. One way to achieve highest level of quality in higher education system is by discovering knowledge for prediction regarding enrolment of students in a particular course, isolation of traditional classroom teaching model, detection of unfair means used in online examination, detection of abnormal values in the result sheets of the students, prediction about students' performance and so on. By this task we extract knowledge that describes students' performance in examination. It helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate counseling and take remedial classes. The biggest challenge for today's education sector is effectively managing all the data of students staff, faculties as well as various facility which they are using in their regular academics by effective and efficient way. This raw data is very important, but managing all this data is very laborious job if we have to handle it manually. Data mining is one of the effective techniques of extracting mining data from large database. Educational data mining is one of the sub branch of data mining where all the data related to result, performance, behavioral analysis of students is extracted.

While the analysis of educational data is not itself a new practice, recent advances in educational technology,

including the increase in computing power and the ability to log fine grained data about students' use of a computer-based learning environment, have led to an increased interest in developing techniques for analyzing the large amounts of data generated in educational settings. Analysing students' data and information to classify students, or to create decision trees or association rules, to make better decisions or to enhance student's performance is an interesting field of research, which mainly focuses on analysing and understanding students' educational data that indicates their educational performance, and generates specific rules, classifications, and predictions to help students in their future educational performance.

The objective of this study is to identify relations between students' personal and social factors, and their academic performance.

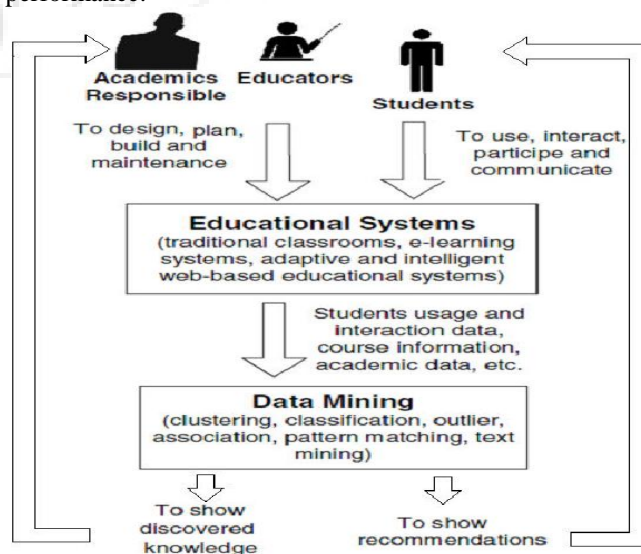


Figure 1: A cycle of Applying Data Mining in Educational System

Data mining is a vast technique which searches retrieves and analyze data from a data repository say as warehouse which is required for business purpose. It is the process of studying data from different perceptions and briefing it into useful information. The information obtained from data mining can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of investigative tools used for searching and testing data. It allows users to analyze data from many different scopes or angles, classify it, and summarize the relationships identified among the retrieved data. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

Now a day, the educational institutes have to require financial assistance for their research and other operational activities as the government funding has been limited to aided institutes. Employing a decision support system is a positive way to use data to manage, operate, and evaluate educational institute in a better way.

Depending on the quality and availability of the underlying data, such a system could address a wide range of problems by cleansing data from any combination of education records maintenance system. Mining data from data warehouse can be a ready and effective system for the decision makers. Data warehouse is a subject oriented, integrated, time-variant, and non-volatile collection of data. This data helps experts to take knowledgeable decisions in an educational institution.

Features of Data Warehouses

- 1) It is a database, which is kept separate from the organization's operational database.
- 2) No regular updating done in a data warehouse.
- 3) It possesses consolidated ancient data, through which organization can analyze its business.
- 4) It helps executives to organize, understand, and use their data to take strategic decisions.
- 5) It helps in the integration of diversity of application systems.

Despite significant progress over the last ten years, Indian higher education is faced with following challenges:

Challenge 1 (A Gap between Supply and Demand): India has a low rate of enrolment in higher education, at only 12% (post graduation level), compared with 26% in China. There is enormous unmet demand for higher education. By 2020, the Indian government aims to achieve 30% gross enrolment, which will mean providing 40 million university places, an increase of 14 million in six years.

Challenge 2 (Poor Quality of teaching and learning): The system is beset by issues of quality in many of its institutions: a chronic shortage of faculty, poor quality teaching, outdated and rigid curricula and pedagogy, lack of

accountability and quality assurance and separation of research and teaching.

Challenge 3 (Limited research Capacity): With a very low level of PhD enrolment, India does not have enough high quality researchers; there are few opportunities for interdisciplinary and multidisciplinary working, lack of early stage research experience; a weak ecosystem for innovation, and low levels of industry engagement.

Reasons behind lower rate of enrolments in higher education, poor quality of teaching, learning and research needs to be discovered in order to achieve desired growth of any nation.

Nowadays many tools and techniques are available to us that can change or improve the education system. Growing digitization of educational data has helped researcher to easily capture these available data and extract meaningful information to take corrective decisions.

There are great openings in the educational data through which our knowledge towards education is tremendously increased as compared in the past decade. Educational data is increasing rapidly as more and more education system is going online. It has opened new areas like new computer supported interactive learning methods and tools-intelligent tutoring system, simulation games which has 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 through online classes. The data collected from online learning systems can be aggregated over large numbers of students and can contain many variables that data mining algorithms can explore for model building.

In today's era educational systems try to offer a customized learning method, by building a model of the individual's goals, attitude and knowledge. Educational Data Mining can be seen as an iterative cycle of hypothesis formation, testing and refinement.

2. Data Mining Definition and Techniques

Data mining, also popularly known as Knowledge Discovery in Database, refers to extracting or "mining" knowledge from large amounts of data. Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. While data mining and knowledge discovery in database are frequently treated as synonyms, data mining is actually part of the knowledge discovery process. The sequences of steps identified in extracting knowledge from data are shown in Figure 2.

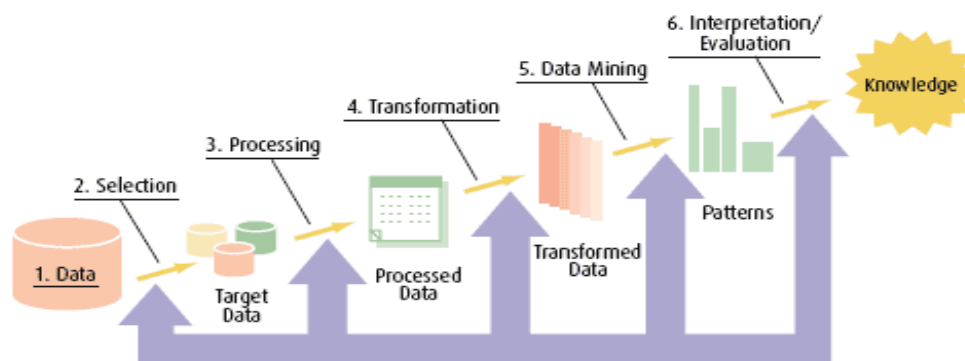


Figure 2: The steps identified in extracting knowledge from data

There are several major data mining techniques have been developed and used in data mining projects recently including **association, classification, clustering, prediction and sequential patterns**.

1) Association

Association is one of the best known data mining technique. In association, a pattern is discovered based on a relationship of a particular item on other items in the same transaction.

Types of association rules:

- Types of values handled
- Boolean association rules
- Quantitative association rules
- Levels of abstraction involved
- Single-level association rules
- Multilevel association rules
- Dimensions of data involved
- Single-dimensional association rules
- Multidimensional association rules

2) Classification

Classification is a classic data mining technique based on machine learning. Basically classification is used to classify each item in a set of data into one of predefined set of classes or groups. Classification is referred to as supervised learning as the classes are determined before examining the data and unsupervised classification when the objects or classes are no known in advance. Classification is the oldest data mining techniques. Thus it is the most familiar and most popular data mining technique. The Bayesian classification can be used on the student's grade. If the student has 'A' grade then the student's chance of selection in a multinational company can be calculated using conditional Bayesian probability. Similarly the lower grade student's chances of selection can be calculated using Bayesians classification.

Classification Techniques are:

- Regression
- Distance
- Decision Trees
- Rules
- Neural Networks

3) Clustering

Clustering is the process of organizing objects into groups whose members are similar in some way. The cluster is therefore a collection of objects which are "similar" between them and are 'dissimilar' to the objects belonging to other clusters. Clustering is referred as unsupervised learning, it is

similar to classification except that the groups are not predefined, but rather defined by data alone. Clustering is the process of partitioning a collection of objects into groups, called clusters such that 'similar' objects fall into the same groups. Clustering methods are broadly classified into two categories based on the cluster structure which produce hierarchical cluster and partitioning cluster. K – Mean clustering can be used to classify students with similar characteristics in same cluster or group while removing others showing dissimilarities into some other clusters thus it can partition the similar groups from dissimilar groups by continuously measuring the Euclidean distance from cluster mean of similar groups.

4) Prediction

The prediction as it name implied is one of a data mining techniques that discovers relationship between independent variables and relationship between dependent and independent variables. Most commonly used prediction technique is regression analysis. It consists of one or more than one predictor variables. Regression can be used for continuous as well as attribute variables. Prediction is based on the relationship between a thing that is known and a thing need to be predicted that is if certain attributes like domain knowledge and communication level of a student is known than his/her placement possibility can be predicted using multiple regression. Here placement possibility is dependent variable generally denoted by y and domain knowledge and communication level are independent variable generally denoted by x.

5) Sequential Patterns

Sequential patterns analysis in one of data mining technique that seeks to discover similar patterns in data transaction over a business period. The uncover patterns are used for further business analysis to recognize relationships among data. A sequence is an ordered list of events, denoted $\langle e_1 e_2 \dots e_L \rangle$.

3. Goals for Analyzing the Students Mind in Different Area

- 1) Analysis and visualization of data** - It is used to focus meaningful information and support decision making.eg. Course administrators and educators will analyze the usage information and students activities during course to get a brief idea of a student's learning. Visualization information and statics are the two main methods that have been used for this task. The detailed

mathematical analysis of educational data can give us information like where students enter and exit, the most important pages students browse, how many number of downloads of e-learning resources, quantity of pages browsed and total amount of time for browsing of these different pages. It also provides information about reports on monthly and weekly usage of tendencies, usage summaries, material studied by students, the series in which they study topics, patterns of studying activity, timing and sequencing of activities.

- 2) **Predicting student's future learning behavior** - With the use of student modeling, this goal can be achieved by creating student models that incorporate the learner's characteristics, including detailed information such as their knowledge, behaviors and motivation to learn.
- 3) **Discovering or improving domain models** - Through the various methods and applications of EDM, discovery of new and improvements to existing model is possible.
- 4) **Studying the effects of educational support** - It can be achieved through learning systems. It contains what a learner knows, what the user experience is like, what a learner's behavior and Motivation is, and how satisfied users are with online learning. EDM can be applied in modeling, user knowledge, user behavior and user experience
- 5) **Advancing scientific knowledge about learning and learners** - By building and incorporating student models, the field of EDM research and the technology and software used.
- 6) **Predicting Students' Admission in Higher Education** - As more and more institutes are established from private sectors, the student expectation from these institutes is also increasing. They are taking admission in any new course only after screening various factors that are considered important for their overall growth.
- 7) **Predicting Students' Profiling** - EDM can also be used as an effective tool in profiling students based on both hard as well as soft skills. The hard factors include academic background, grades and achievements while soft factor includes communication, behavior, attitude, hobbies etc. Different Data Mining Techniques and algorithms have been used for this task.
- 8) **Predicting Students' Performance** - Data Mining is most popularly used to predict performance of students. Various prediction tools are available like regression and correlation analysis, decision tree, Bayesian networks, neural network etc.
- 9) **Teachers' teaching performance** - There can be various measures to judge teacher's teaching performance. Student feedback is a popular measure but often it gives distorted results. It is because there is high correlation found between marks of the student and feedback of the teacher.
- 10) **Curriculum Development** - In order to capture what is required by the student and how it can be delivered, various quality tools are used. Quality Function Deployment (QFD) is one of them but it is time consuming process to reach inferences.
- 11) **Students' Targeting** - Segmentation, Targeting and positioning are important steps in marketing a product or services. Right kind of student for right kind of course is very important to achieve student satisfaction.
- 12) **Predicting Students Survival in a course** - The problem of student drop out is frequent in professional courses. Often student find themselves misfit for the course and thus withdraw from it.
- 13) **Predicting Students' Placement opportunities** - Another big challenge in higher education is providing placement to students. Most of the institutions are struggling in this domain. With students becoming more and more demanding, quality placement of students is not only crucial but also very important in creating brand for institutes.
- 14) **Planning and scheduling** - Planning and scheduling is used to enhance the traditional educational process by planning future courses, course scheduling, planning resource allocation which helps in the admission and counseling processes, developing curriculum, etc. Different data mining techniques used for this task are classification, categorization, estimation, and visualization. Decision trees, link analysis and decision forests have been used in course planning to analyze enrollee's course preferences and course completion rates in extension education courses. Educational training courses have been planned through the use of cluster analysis, decision trees, and back propagation neural networks in order to find the correlation between the course classifications of educational training. Decision trees and Bayesian models have been proposed to help management institutes to explore the probable effects of changes in recruitments, admissions and courses.
- 15) **User Modeling** User modeling encompasses what a learner knows, what the user experience is like, what a learner's behavior and motivation are, and how satisfied users are with online learning. EDM can be applied in modeling user knowledge, user behavior and user experience.
- 16) **Organization of Syllabus** Presently, organization of syllabi is influenced by many factors such as affiliated, competing or collaborating programs of universities, availability of lecturers, expert judgments and experience. This method of organization may not necessarily facilitate students' learning capacity optimally. Exploration of subjects and their relationships can directly assist in better organization of syllabi and provide insights to existing curricula of educational programs. One of the applications of data mining is to identify related subjects in syllabi of educational programs in a large educational institute.
- 17) **Detecting Cheating in Online Examination** - Now a day's exams are conducted online remotely through the Internet and if a fraud occurs then one of the basic problems to solve is to know: who is there? Cheating is not only done by students but the recent scandals in business and journalism show that it has become a common practice. Data mining techniques can propose models which can help organizations to detect and to prevent cheats in online assessments. The models generated use data comprising of different student's personalities, stress situations generated by online assessments, and common practices used by students to cheat to obtain a better grade on these exams.

4. Applications of Algorithms in Education Mining

Number of universities and students is increasing day by day; we think that data mining technology can help improving the education standard and consequently causing high ratio of successful candidate, low ratio of students' drop-out and maximizing education system efficiency. Following is a detail of the algorithms used in education mining.

- 1) **C 4.5** - A classifier system takes input from the cases described by values and attributes and output a classifier that can accurately predict classes of new cases. C 4.5 is a descendant of CLS(Common Language Specification) and IDE(Interactive Development Environment), creates classifier and generated decision tree. It can also make classifier in most comprehensive rule-set forms.
- 2) **Support Vector Machine (SVM)** - Support Vector Machine (S.V.M) is considered an efficient tool to train data. It offers accurate methods among algorithms. SVM is the most worked upon algorithm for training purposes and a lot of research is still going on. SVM can find classification function in two- class learning tasks. The metric for "best" can be realized geometrically. It is good because of its generalization ability.
- 3) **Apriori** - A popular way to find frequent data item sets is by comparing explosion. After data items are obtained then we can easily generate association rules. There are different steps for that.
 - a) Generate C_{k+1} for it
 - b) Generate C_{k+1} for item sets of k+1
 - c) Calculate support
 - d) Put items for minimum support for f_{k+1}
- 4) **The Expectation Maximization Algorithm**- It provides a flexible mathematical approach to modeling and clustering of data on randomly observed basis. This can be used to cluster continuous data. Expectation Maximization algorithm is used to model distribution of random phenomenal data.

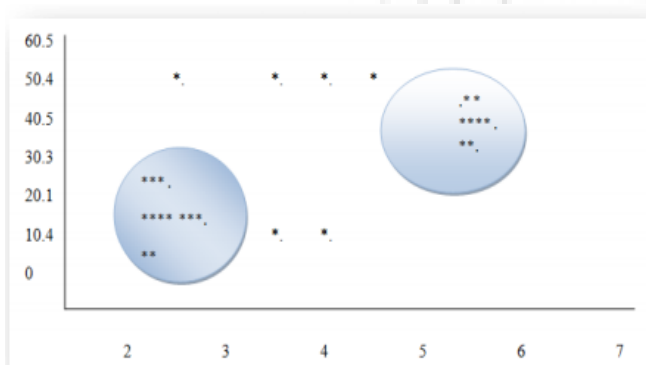


Figure 3: EM clustering of Old Faithful Data

E and M parts EM algorithms give (ML) estimation of normal components. On (k+1) turn of EM, log of the data is taken.

- 5) **Page Ranker**- Page Ranker was given forth by Brin and Larry Page in 1998. On this algorithm's basis they built Google, which has an excellent success ratio. It produces a static ranking of different web pages in sense

that pager value is determined offline and does not depend on the online queries. Pager Ranker formula:

- a) A hyperlink point the value of the page is an implicit conveyance for authority. Thus more links a page receives more prestige it has.

- b) Pages that go to I is also considered good

Page Rank algorithm given by Lawrence Page and Sergey Brin is in a lot of publications.

It is as under...

$$PR(A) = (1-d) + d (pr(t_i) / C(t_i) + \dots + PR(T_n)/C(t_n))$$

Where, PR(A) is a Page Rank of Page A

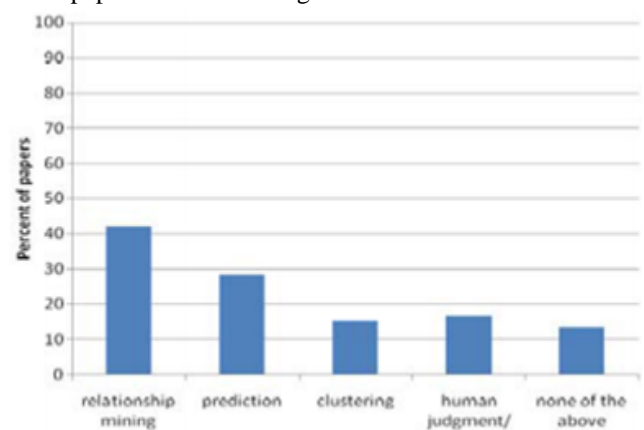
PR(T_n) is Page Rank.

- 6) **Clustering Algorithm**- Clustering is a grouping of data division of data into groups of similar objects. It is very crucial in data mining applications such as information retrieval and text mining, scientific data exploration, web analysis, spatial database applications, medical diagnostics, marketing etc. Data Clustering is unsupervised and statistical data analysis technique. It is used to divide similar type of data into homogeneous groups. Following algorithms are used in education mining in Clustering.

- a) K-Mean Clustering Algorithm is one of the best clustering algorithms in data mining.
- b) K-Means is a non-hierarchical clustering - method that seeks to partition the data into the form of one or more clusters. This method partitions the data into clusters so that the data having the same characteristics are grouped into one cluster and the data that have different.

5. Trends in Educational Data Mining Methods

Romero and Ventura's survey of Educational Data Mining research from 1995 to 2005, 60 papers were stated that developed EDM methods to answer research questions of applied interest. Relationship mining methods of various types were the most prominent type of EDM research between 1995 and 2005. 43% of papers in those years involved relationship mining methods. Prediction was the second most prominent research area, with 28% of papers in those years involving prediction methods of various types. Human judgment data analysis and clustering followed with 17% and 15% of papers. The full distribution of methods across papers is shown in Figure 4.



Whereas relationship mining was leading between 1995 and 2005, in 2008- 2009 it slipped to fifth place, with only 9% of papers involving relationship mining. Prediction, which was in second place between 1995 and 2005, moved to the leading position in 2008-2009, representing 42% of EDM 2008 papers. Human judgment/exploratory data analysis and clustering remain in approximately the same position in 2008-2009 as 1995-2005, with (respectively) 12% and 15% of papers.

Motivation

The need of building data warehouse for the educational institute is generated due to following two main reasons:

- a) Internal sources
- b) External Sources

Internal Sources includes incapability of current operational systems to provide required information for parameter driven analysis external sources include competitive factors.

Generally following problems are faced by all educational Institutes:

- The data is stored in different sources in distributed locations.
- Users find difficulty in locating the reports needed by them.
- The user interface for the current operational system is not satisfactory and is confusing and hard to use for decision makers.
- When the consolidated report from two or more different subject area is required, it is almost impossible.
- There is no easy way to get assistance.

Thus by using a decision support system is a positive way to use data to manage, operate, and evaluate educational institute in a better way. Depending on the quality and availability of the underlying data, such a system could address a wide range of problems by cleansing data from any combination of education records maintenance system.

6. Conclusion

In this paper, the classification task is used on student database to predict the students division on the basis of previous database. As there are many approaches that are used for data classification, the decision tree method is used here. Information's like Attendance, Class test, Seminar and Assignment marks were collected from the student's previous database, to predict the performance at the end of the examination. This study will help to the students and the teachers to improve the division of the student. This study will also work to identify those students which needed special attention to reduce fail ratio and taking appropriate action for the next examination. Data mining is a tremendously vast area that includes employing different techniques and algorithms for pattern finding. The algorithms discussed in this paper are the ones used in education mining. These algorithms have shown a remarkable improvement in strategies like course outline formation, teacher student understanding and high output and turn out ratio. An appreciable research is still being done on various algorithms. I hope this review paper appreciates the current algorithm researchers and inspires the new ones to explore further. A lot of interest has been seen in

Educational Data Mining these days because a large number of students are enrolling for higher education. Through Educational Data Mining institutions' researchers and stakeholders can bring more and more satisfaction amongst student's fraternity. Educational data mining finds its application not only in descriptive and predictive analytics but also in prescriptive analytics where suitable actions can also be prescribed. Understanding students, appropriate profiling and accurate predictions will not only increase the quality of education but also increase good learning experience to the students' fraternity. Due to more and more usage of internet by students today, huge data is available about them. Through data mining, we can extract useful information that can help the education system to formulate appropriate strategies for our youths. Thus this paper gives a brief overview on Educational Data mining various techniques and applications used in Educational sector for improving quality of higher Education.

References

- [1] Mrs. Bharati M. Ramageri, "Data Mining Techniques and Applications," Indian Journal of Computer Science and Engineering, Vol. 1 No. 4, pp. 301-305
- [2] Hemlata Sahu, Shalini Shirma and Seema Gondhalakar, "A Brief Overview on Data Mining Survey," International Journal of Computer Technology and Electronics Engineering (IJCTEE), Vol.1, Issue 3, pp.114-121
- [3] Kalyani M Raval, "Data Mining Techniques," International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2 Issue 10, pp.439-442
- [4] Sangeeta Goele, Nisha Chanana, "Data Mining Trend In Past, Current And Future," International Journal of Computing & Business Research, in Proc. I-Society 2012, 2012.
- [5] Acharya S., N Madhu, (2012) "Discovery of students academic patterns using data mining techniques" International Journal of Computer Science and Engineering, Vol 4, no. 06, J
- [6] Baker, R.S.J.d. (2010) Data Mining for Education. In Mc Gaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition), vol. 7, pp. 112-118. Oxford, UK: Elsevier.
- [7] Berry, J.A. Michael and Linoff S. Gordon., (2004) "Data Mining Techniques for Marketing, Sales, and Customer Relationship Management." Second Edition, Wiley Publishing.
- [8] "EducationalDataMining.org". 2013. Retrieved 2013-07-15. "<http://educationaldatamining.org/EDM2008/>" Retrieved 2013-09-04.
- [9] Baker, Ryan. "Data Mining for Education" (PDF). Oxford, UK: Elsevier. Retrieved 9 February 2014.
- [10] Romero, C. and Ventura, S. (2007) "Educational data mining: A Survey from 1995 to 2005", Expert Systems with Applications (33).
- [11] Monika Goyal and Rajan Vohra, "Applications of Data Mining in Higher Education", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012.
- [12] <http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm>

- [13] Manjunath T. N., Ravindra S. Hegadi, Umesh I. M., and Ravikumar G. K., "Realistic Analysis of Data Warehousing and Data Mining Application in Education Domain " Cristobal Romero* and Sebastian Ventura, "Data mining in education"
- [14] B.Namratha and Anurag Group of Institutions," "Educational Data Mining –Applications and Techniques"
https://www.tutorialspoint.com/dwh/dwh_overview.htm
- [15] Heikki, Mannila, Data mining: machine learning, statistics, and databases, IEEE, 1996.
- [16] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2000.
- [17] Alaa el-Halees, "Mining students data to analyze e-Learning behavior: A Case Study", 2009.
- [18] Baradwaj, B.K. and Pal, S., 2011. Mining Educational Data to Analyze Students' Performance. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011.
- [19] Ahmed, A.B.E.D. and Elaraby, I.S., 2014. Data Mining: A prediction for Student's Performance Using Classification Method. World Journal of Computer Application and Technology, 2(2), pp.43-47.
- [20] Bhardwaj, B.K. and Pal, S., 2012. Data Mining: A prediction for performance improvement using classification. (IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 4, April 2011.

