

Data Mining in Education Sector: A Review

Jaya Srivastava

Ph.D. Scholar, Jaipur National University, Jaipur

Dr Abhay Kumar Srivastava

Assistant Professor, Jaipuria Institute of Management, Lucknow

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.

Keywords: Data Mining, Educational Data Mining, Prediction, Profiling

Introduction

The development of any country depends upon the educational background of its citizens. Most of the developed countries have a high literacy rate. Developing countries are also witnessing a growing demand of higher education. Since this growing demand could not be met only by government Universities and colleges hence large numbers of private universities and institutions have been established in the last decade and the trend will continue in the future also.

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

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.

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system is beset by issues of quality in many of its institutions: a chronic shortage of faculty, poor quality teaching, outdated and rigid curricula and pedagogy, lack of accountability and quality assurance and separation of research and teaching.

Limited research Capacity: With a very low level of PhD enrolment (only 1% in year 2012-13), 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 on line 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.

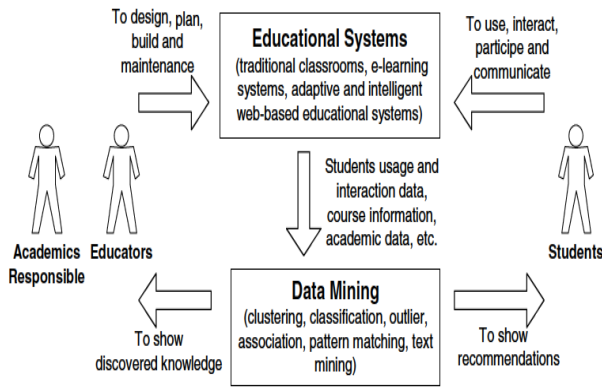


Fig 1. Data Mining in Education System

Educators are responsible for designing, planning, building and maintaining educational systems, while students use and interact with them.

The application of data mining is different for educators and students. For students, the goal is to discover activities, resources and learning tasks that improve their learning, based on their attitude and likings while for educators, the goal is to have more feedback from students for evaluating the structure of the course content and its effectiveness on the learning process; to classify students based on their needs; to discover information to improve the adaptation and customization of the course, etc.

Educational Data Mining (EDM) is upcoming field in Knowledge discovery. Due to widespread growth of higher education, predictions related to student’s performance can be accurately done through EDM. Not only predictions, classification, associations and grouping can also be done with perfection using statistical and software tools. The Education system can be equipped with more information relating to future drop out of students and their success in enrolled courses. Not only students but other stake holders could be benefitted by EDM. Nowadays interactive e-learning methods and tools have opened an opportunity to collect and scrutinize student data.

In the educational field, data mining techniques can generate useful patterns that can be used both by educators and learners. Not only may EDM assist educators to improve the instructional materials and to establish a decision process that will modify the learning environment or teaching approach, but it may also provide recommendations to learners to improve their learning and to create individual learning environments.

The paper focuses on the role of data mining in the education sector.

Related Work

There are two main taxonomies for categorizing data mining techniques in educational systems:

Romero-Ventura (2007) and Baker-Yassef (2009). These are depicted in Figure 2 in which some categories are shown as related.

Romero-Ventura’s taxonomy encompasses four categories and focus more on applications of educational data mining to web data. Although "Statistics and visualization" cannot formally be considered data mining, the authors include it as is often the starting point of any study.

The second and third categories are usual tasks found in data mining projects while the fourth category can be viewed as an extension of data mining to text data and it is very related to web content mining (2007)

Baker-Yassef’s (2009) taxonomy first aligns with Romero-Ventura’s "Statistical and visualization" category (2007) and the fifth is building models from knowledge discovery. Baker (2010) identified four areas of application of data mining in education: improve students’ model, discovering models of the knowledge structure, studying the pedagogical support provided by learning software and scientific discovery about learning and learners

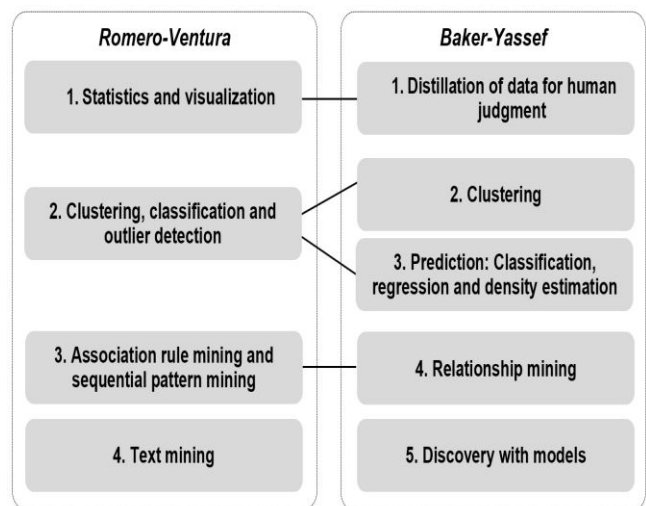


Fig. 2. Taxonomies in Educational Data Mining

Siraj F. (2011) et al found that data mining can be used to understand correlations and patterns in existing enrolment data by using Predictive modeling techniques. The results of this study can help for those associated with the registration and education process of students.

Verma et al(2012) suggested the use of data mining techniques such as clustering, decision tree and association to improve student’s performance in higher education process. The output of the process can be used by the faculty, consultant to present the most appropriate courses to students. It can also be used by educational course planner to have more advance strategies on student course planning

He & Wu (2013) examines students’ participation and learning behavior by using LVS (Live Video Streaming) of data mining and text mining techniques. They found that students use chat messages to communicate positive emotions, negative emotions, and expressions of social support. In addition, there is no positive correlation

between the number of chat messages and final grades. Maqsood (2013) states that data mining can be used to report and analyze the data that help in preparing marketing strategies for targeted students.

Samira et al. (2012) discusses the interface between data mining research in higher education and legal implications of data. They stated that there is ‘hardly any research that focuses on applications for improving higher education using data in a holistic way. They suggest that teachers can use data mining to find innovative way to help improving and teaching as well as develop assessment procedures. Cios et al. (2007) indicates that data mining allows for automated discovery of patterns and trends in the data.

Tsai et al. (2011) examined computer proficiency test using data mining technique (cluster analysis, decision tree) to assess students’ computer literacy before admitting into higher education. They concluded that data mining techniques assist universities to identify a number of groups who need reinforcement training and promote their computer proficiency more efficiently.

Cristobal Romero and Sebastian Ventura (2008) reviewed different data mining methods and techniques for classifying students based on their Moodle usage data. They also used real data from seven moodle courses with Cordaba University students to find out better classifier models.

Pena et al. (2009) studied application of data mining in the web based education field. Data mining applied in student modeling, tutoring, content, and assessment areas. They conclude that education data mining help in understanding learning outcomes, identify students; behavior, and characterize groups of students.

Tools Used in Educational Data Mining:-

The most commercial data mining tools which provide multiple data mining functions and multiple knowledge discovery techniques and the environment on which they run are listed below:-

Name of Tool and Developer	Source (Commercial)	Function/ Features	Techniques/ Tools	Environments
MSSQL Server 2005 (Microsoft) An	Commercial	Provides DM functions both in relational db system and Data Warehouse (DWH) system environment.	Integrates the algorithms developed by third party vendors and application users.	Windows, Linux
SPSS Clementine (IBM)	Commercial	Provides an integrated data mining development environment for end users and developers.	Association Mining, Clustering, Classification, Prediction and visualization tools	Windows, Solaris, Linux
Enterprise Miner (SAS Institute)	Commercial	Provides variety of statistical analysis tools	Association Mining, Classification, Regression, Time series analysis, Statistical analysis, Clustering	Windows, Solaris, Linux Insightful Miner
Oracle Data Mining (Oracle Corporation)	Commercial	Provides an embedded DWH infrastructure for multidimensional data Analysis	Association Mining, Classification, Prediction, Regression, Clustering, Sequence similarity search and analysis.	Windows, Mac, Linux
DBMiner (DBMiner technology Inc)	Commercial	Provides multiple mining algorithms .Data-cube-based on-line analytical mining ,frequent-pattern mining functions and integrated visual classification methods	Discovery-driven OLAP analysis, association, classification, and clustering	Windows, Linux

Name of Tool and Developer	Source (Commercial)	Function/ Features	Techniques/ Tools	Environments
Intelligent Miner (IBM)	Commercial	Provides tight integration with IBB’s DB2 relational database system, Scalability of Mining Algorithm	Association Mining, Classification, Regression, Predictive modelling, Deviation detection, Clustering, Sequential Pattern	Windows, Solaris, Linux

Education data mining Techniques focusing students’ performance:

In data mining many methodologies are been used according to the related problems in which one should have the in-depth and complete knowledge of the problem according to that one could decide the technique and algorithm to be applied on the dataset. The methodology should be used in such a manner that it should be able to yield an accurate prediction of the performance which should be beneficial for the students and for the institutions. More than one hundred and fifty papers, journals, surveys

help to establish proper methodology for identifying the best proper methodological pattern for the problem.

The most important methodologies used are Classification, Prediction, and Clustering. These techniques are generally used in almost all of the applications and are described below:

Classification

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 not known in advance. Classification is the oldest data mining technique. 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 then his 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 Bayesian classification.

Prediction

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 .

Clustering

Clustering is referred to 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.

Applications of Data Mining in Education Sector

Data mining can be applied in the following functional areas of an education sector.

1. 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.

Yadav et al (2012) noted that the important attribute in predicting students enrolment is found to be Graduation Stream (GS). The study shows that the student past academic performance can be used to create a decision tree model using ID3 algorithm that is used for predicting student's enrollment in MCA course. The Study found that students with background of mathematics or computers performed better in MCA course than students with other backgrounds.

Classification and regression (CART) are the commonly used Data Mining techniques for doing such predictions. Kovacic, J. Zlatko (2010) used CART for predicting students' admission.

Aksenova et al. (2006) build predictive model for fresher, existing and returned students at both graduate and undergraduate levels. On the basis of pre-enrolment data they recommended students who are classified in the category of "High Risk" should be mentored in order to avoid any drop outs.

2. 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. Naïve Bayes, Bayes Net, Support Vector Machines, Logistic regression and decision trees are mainly used for student profiling. Consequently, cluster analysis can be done to students' profiling and separate marketing strategies can be prepared to target segmented students. Cluster analysis is also called data segmentation (Sinha et al., 2010).

3. 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. Kumar and Uma (2009) studied students' performance in the course using data mining techniques, particularly classification techniques such as Naïve Bayes and Decision tree based on students ID and marks scored in course. Furthermore, they suggest that data mining process can be done to the teachers for classifying performance which helps in improving higher education system. Data mining methods helps students and teachers to improve students' performance.

Tair M. M. Abu (2012) et al showed the usage of Data Mining in higher education particularly to discover knowledge about student's performance. The paper proposes various techniques to improve the performance of graduate students. It proposes to use two classification methods namely Rule Induction and Naïve Bayesian

classifier to predict the grade of the graduate student. Student's clustering was done by using K-means clustering algorithm. Distance-based Approach and Density-Based Approach outlier approach was used to detect all outliers in the data.

4. Teachers' teaching performance:

There can be various measures to judge teacher's teaching performance. Student feedback is a popular measure but often it gives skewed results. It is because there is high correlation found between marks of the student and feedback of the teacher. Several DM techniques have been used in this task, although association rule mining has been the most common. Mardikyan and Badur (2012) identified factors that affect instructors' teaching performance in university by using stepwise regression and decision tree of data mining techniques. Instructor attitude, Teacher status, student attendance, and students' feedback affect teaching performance.

5. 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 very time consuming process to reach inferences. Colored Petri Nets is a mining tool to meet generally used for curriculum development. Curriculum revision has been done by association rule mining in order to identify and understand whether curriculum revisions can affect students in a university. Hsia et al. (2008) study course preferences, completion rates and profession of enrollees by using data mining algorithm such as decision tree, link analysis and decision forest. They found the correlation between course category and enrollee professions. They lay emphasis on importance of data mining in building curriculum and marketing in the field of higher education.

6. 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. Woo et al. (2005) defines customer targeting as "a process of building strategy towards specific customers." They indicate that customer map is the visualization method for customer targeting. Customer map helps in building customer-oriented strategy. It is a "novel technique to find right target customers who are homogenous with characteristics, values and needs. It is organized with three dimensions of customer targeting: Customer Value (usage and behaviour), customer characteristics (demographic and Psychographic), and customers' needs (complaints and satisfaction). Target customers are detected and targeting strategies can be derived from a customer map.

7. 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. Massa and Puliafito (1999)

studied the problem of university dropout by using Markov Chains, a new data mining technique. They conclude that behavior of homogenous group can be studied through Markov chain. It can be used to define clusters of students associated with different dropout risk degree.

8. Predicting Students' course selection:

Selection of course by a student depends on various factors such as student's interest, grades, further demand etc. Kardan et al. (2013) determine various factors influencing student course selection using neural networks such as characteristics, students' workload, course grades, course type, course duration, and number of time conflicts, final examination time and students' demand. These factors are used as input of neural network modeling. Furthermore, Guo (2010) analyse and predict student course satisfaction using neural networks. He found that number of students enrolled to a course and high distinction rate in final grading are the two most influential factors to student course satisfaction

9. 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.

Shreenath Acharya & Madhu N (2012) proposed a model for predicting student's placement based on historical information of database which can be utilized by educational institutions to reveal some interesting patterns that could be analyzed to plan their future activities. It has been found to be genuinely useful for the higher authorities like Director and Head of the departments.

Conclusion :

A lot of interest has been seen in EDM these days because a large number of students are enrolling for higher education. Through EDM 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.

References

1. 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*

2. Affendey, L.S., I.H.M. Paris, N. Mustapha, M.N. Sulaiman and Z. Muda, (2010). Ranking of influencing factors in predicting students academic performance. *Information. Technologies. Journal.*, 9: 832-837.
3. Aksenova, S. Sretlana., Zhang, Du and Lu, Meilin. (2006) "Enrolment prediction through Data Mining" *IEEE Conference Publications*: 510-515
4. Baker, R.S.J.d., Yacef, K. (2009) *The State of Educational Data Mining in 2009: A Review and Future Visions*. *Journal of Educational Data Mining*, 1 (1), 3-17
5. He, Wu. "Examining students' online interactions in alive video streaming environment using data mining and text mining" *Computers in Human Behaviour*, 29 (2013): 90-102.
6. Baker, R.S.J.d. (2010) *Data Mining for Education*. In McGaw, 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. Berry, J.A. Michael and Linoff S. Gordon.,(2004) "Data Mining Techniques for Marketing, Sales, and Customer Relationship Management." Second Edition, Wiley Publishing,.
9. B.K. Bharadwaj and S. Pal. "Data Mining (2011): Mining Educational Data to Analyze students performance", *International Journal of Advanced Computer Science and Applications*, (IJACSA), Vol. 2, No. 6,.
10. C. Romero and S. Ventura,(2007) "Educational data mining: A survey from 1995 to 2005," *Expert Systems with Applications*, vol. 33, pp. 135–146.
11. Chen, Mu-Chen, Chiu, Ai-Lun and Chang, Hsu-Hwa.(2005) "Mining changes in customer behaviour in retail marketing." *Expert System with Applications*, 28: 773-781
12. Chen, Chi-Kuang., Shie, An-Jin., and Yu, Chang-Hsi.(2012) "A Customer-oriented Organisational diagnosis model based on data mining of customer complaints databases" *Expert System with Applications*, 39: 786-792
13. Cios, K.J., Swiniarski, W.R., Pedrycz, W. and Kurgan, A. L., *Data Mining(2007): A Knowledge Discovery Approach*, Springer, New York
14. Guo, W. William (2010), "Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction." *Expert System with Applications*, 37 3358-3365.
15. Giudici, Paolo.(2013) "Applied Data Mining: Statistical Methods for Business and Industry", John Wiley and Sons Ltd.,.
16. Hsia, Tai-Chang., Shie, An-Jin and Chen, Li-Chen., (2008) "Course Planning of extension education to meet market demand by using data mining technique: an example of Chinkuo Technical University in Taiwan." *Expert System with Applications*, 34: 596-602.
17. Kardan, A. Ahmad., Sadegh, Hamid, Ghidary, Shiry. Saeed., and Sani, Reza Fani., Mohammad. (2013) "Prediction of student course selection in online higher education institutions" 65: 1-11
18. Kovacic, J. Zlatko, (2010) "Early prediction of student success: Mining students' enrolment data, in proceedings of informing science and IT education Conferences (InSITE), 647-665.
19. Kumar, A.N.V. and Uma, G.V. (2009) "Improving academic performance of students by applying data mining techniques." *European Journal of Scientific Research*, No. 4: 526-534
20. M. Bienkowski, M. Feng, and B. Means(2012), "Enhancing teaching and learning through educational data mining and learning analytics: An issue brief," tech. rep.,
21. Maqsood, A. Mohammed, (April 2013) "Customer Relationship Management in B-Schools: An Overview." *International Journal of Computer Sciences and Management Research*, Vol. 2 issue 4
22. Mardikyan, Sona and Badur, Bertan. (2011) "Analysing teaching performance of instructors using data mining techniques." *informatics in Education*, Vol.10, No.2,: 245-257.
23. Massa, S and Puliafito, P.P. (1999) "An application of data mining to the problem of the university students' dropout using Markov chains." in Zytknow, J.M and Rauch, J (Eds.), *Principles of Data mining and Knowledge Discovery*, third European conference, PKDD 99, Prague, Czech Republic, September 15-18, proceedings, Springer, vol. 1784, 51-60
24. Nguyen Thai Nghe, Paul Janecek and Peter Haddawy (2007), "A Comparative Analysis of Techniques for Predicting Academic Performance", 37th ASEE/IEEE *Frontiers in Education Conference*,.
25. N. R. Draper and H. Smith (1998), *Applied Regression Analysis*. New York: Wiley,.
26. Olson L. David and Delen, Dursun.(2008) "Advanced Data Mining Technique", Springer.
27. Pena, Alejandro. Dominguez, R., and Medd, J (2009), "Education Data Mining: A sample of review and study case." *Academic World and Education and Research Center*, 2:118-139
28. P. Cortez and A. Silva (2008), "Using Data Mining To Predict Secondary School Student Performance", In *EUROSIS*, A. Brito and J. Teixeira (Eds.), pp.5-12.
29. P. Espejo, S. Ventura, and F. Herrera (2010), "A survey on the application of genetic programming to classification," *IEEE Trans. Syst.,Man, Cybern. C, Appl. Rev.*, vol. 40, no. 2, pp. 121–144,.
30. Q. A. AI-Radaideh, E. M. AI-Shawakfa, and M. I. AI-Najjar(2006), "Mining Student Data using Decision Trees", *International Arab Conference on Information Technology(ACIT'2006)*, Yarmouk University, Jordan,.

30. Ramageri, M. Bharti "Data Mining Techniques and Applications (Dec-Jan 2010-2011)." *Indian Journal of Computer sciences and Engineering*, Vol.1 No.4.: 301-305.
31. Romdhae, L.B., Fadhel, N. and AyeB, B. (2010) "An Efficient approach for building customer profiles from business data." *Expert System with Applications*, 37: 1573-1585
32. Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (2010) Introduction. In Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (Eds.) *Handbook of Educational Data Mining*. Boca Raton, FL: CRC Press, pp. 1-8.
33. Romero, C. and Ventura, S. (2008). "Data Mining Algorithms to Classify Students" *EDM Proceedings*
34. Ramesh V., Thenmozhi P., Ramar. K. (2012), "Study Of Influencing Factors Of Academic Performance of Students: A Data Mining Approach" *International Journal Of Scientific And Engineering Research*, Vol.No.3 Issue no.7.
35. Samira, AlEtia, Ipperciel, Donald and Hammad, Ahmed (2012). "Implications and challenges to using data mining in educational research in the Canadian context." *Canadian Journal of Education*, 35, 2: 101-119.
36. Sembiring S., Zarlis. M, Hartama D (2011). "Prediction of students academic performance by the application of data mining technique" *IACSIT*
37. S. Kotsiantis, C. Pierrakeas, and P. Pintelas (2004), "Prediction of Student's Performance in Distance Learning Using Machine Learning Techniques", *Applied Artificial Intelligence*, Vol. 18, No. 5, pp. 411-426.
38. Sinha, Nidhi., Vandana, Ahuja., and Medury, Y. (2010) "Effective Brand Management through customer profiling using clustering, *IUP Journal of Brand Management*, 21-33
39. Siraj F., Abdoulha A. M. "Mining enrolment data using predictive and descriptive approaches" *Applied Sciences, College of Art and Sciences*, University Utara Malaysia, 21 Jan 2011
40. Sunita B Aher and LOBO L.M.R.J (2011)., "Data Mining in Educational System using WEKA" *International Conference on Emerging Technology Trends (ICETT)*.
41. Sukanya M., Biruntha S., Karthik S. (2012), "Data Mining: Performance Improvement in Education Sector using Classification and Clustering Algorithm" *International Conference on Computing and Control Engineering (ICCCE 2012)*, 12 & 13 April.
42. Tiwari, Mahendra. Randhir Singh and Neeraj Vimal., (Feb 2013) "An Empirical Study of Application of Data Mining Techniques for Predicting Student Performance in Higher Education." *International Journal of Computer Sciences and mobile Computing*, Vol.2 Issue 2: 53-57.
43. Tuffery, Stephane. (2011) "Data Mining and Statistics for Decision making". John Wiley and Sons Ltd.
44. Tair Abu. M. M., Elhalees E. L. M.A. (2012), "Mining educational data to improve students performance: A case study", *International Journal of information and communication technology research*, vol. 2 No. 2.
45. Tsai, C.F., Tsai, C.T., Hung, C.S., Hwang, P.S. (2011)., "Data Mining Techniques for identifying students at risk of failing a computer proficiency test requires for graduation." *Australian Journal of Educational Technology*, 27(3): 481-498
46. Verma S., Thakur R.S. and Jalori S. (2012), "A study of the applications of Data Mining techniques in Higher education", *International Journal of Computer and Communication Technology*, Vol.3, Issue-3,
47. Woo Ji-young., Bae, Sung Min and Park, Sang Chan. (2005) "Visualisation method for customer targeting using customer map." *Expert System with Applications*, 28: 763-772
48. Witten, H. Ian and Frank, Eibe (2005). *Data Mining: Practical Machine Learning Tools and Techniques*, Second Edition, Morgan Kaufmann.
49. Yadav S. K., Pal S. (2012) "Data mining application in enrolment management : A case study", *International Journal of Computer Application* Vol. 41 No. 5, March
50. Zafra A., Romero C., Ventura S. (2011)., "Multiple instance learning for classifying students in learning management systems" *Journal of Expert Systems with Applications, Elsevier*
51. Z. J. Kovacic (2010), "Early prediction of student success: Mining student enrollment data", *Proceedings of Informing Science & IT Education Conference*.
52. Z. N. Khan (2005), "Scholastic Achievement of Higher Secondary Students in Science Stream", *Journal of Social Sciences*, Vol. 1, No. 2, pp. 84-87.