A Comprehensive Survey of Students Performance Using Various Data Mining Techniques

Dr. B. Umadevi[1], R. Dhanalakshmi[2]

[1]Assistant Professor & Head, P.G. & Research Department of Computer Science,
Raja Doraisingam Govt. Arts College, Sivaganga, TamilNadu, India.

[2]Research Scholar, P.G. & Research Department of Computer Science, Raja Doraisingam Govt. Arts College, Sivaganga, TamilNadu, India.

Abstract: ***The educational institutions are key resources for producing the good students to provide better services for the society.It is mandate for every educational institute to understand the competency level of every students, inorder to study and know the performance. The key factors for identifying the performance is being not only controlled with limited parameter but also with clear data . So it is inevitable to include some standards and calibration measures to make the study of students' performance. In accompanying with this factor,we have so many tools and techniques are available to predict the results. But today the modern method, Datamining is evolving with so many techniques. Among them EDM(Educational Data Mining)is much popular and useful for making such a research. In our survey paper we would like to focus and analyse the various Data Mining Techniques to brought the clarity in students' results and faculties contribution to make this one as success.***

Keywords: EDM, IQ, Classification, Clustering, Prediction

1. Introduction

The main aim of higher education institutes is to give excellence education to its students and to improve the quality of managerial decisions. One way to meet the top level of excellence in higher education system is by discovering knowledge from educational observations to study the main attributes that may affect the students' performance. The discovered knowledge can be used to stretch forth a helpful and beneficial recommendations to the academic planners in higher education institutes to develop their decision making process, to develop students academic performance and trim down failure rate, to better understand students' behaviour, to help instructors, to improve teaching and many other benefits.

Data mining is concerned with the analysis of data and they use different software techniques to find the unknown and unpredicted patterns and their relationships in the data set. The techniques of data mining are categorized into two groups: they are supervised learning and unsupervised learning. [1] Data mining techniques have been pertained to predict the academic performance of the students based on their socioeconomic condition and earlier academic Performances. Classification is one of the data mining methods of predictive types that classifies data (Constructs a pattern) based on the training set and use the pattern to classify a new date (testing set).[2]The prediction of the student's performance has become one of the most significant needs to develop the quality of performance. There is a need of data mining in an educational organization for the students as well as academics responsible.

Educational data mining is a rising regulation that endorses the new techniques to extract the new data that come from educational settings and by using those techniques, a greater prediction can be done for students' behavior, academic performance, subject interest etc. [3]

The following Figure1 which elucidate that educators can use the applications of EDM to determine how to design, plan, build and maintain the educational system and design best methods to deliver the course information and tools to use to engage their learners for optimal learning outcomes.Students can also benefit from the discovered knowledge by using the EDM tools to suggest activities and resources that they can use based on the insights from the past or similar learners.

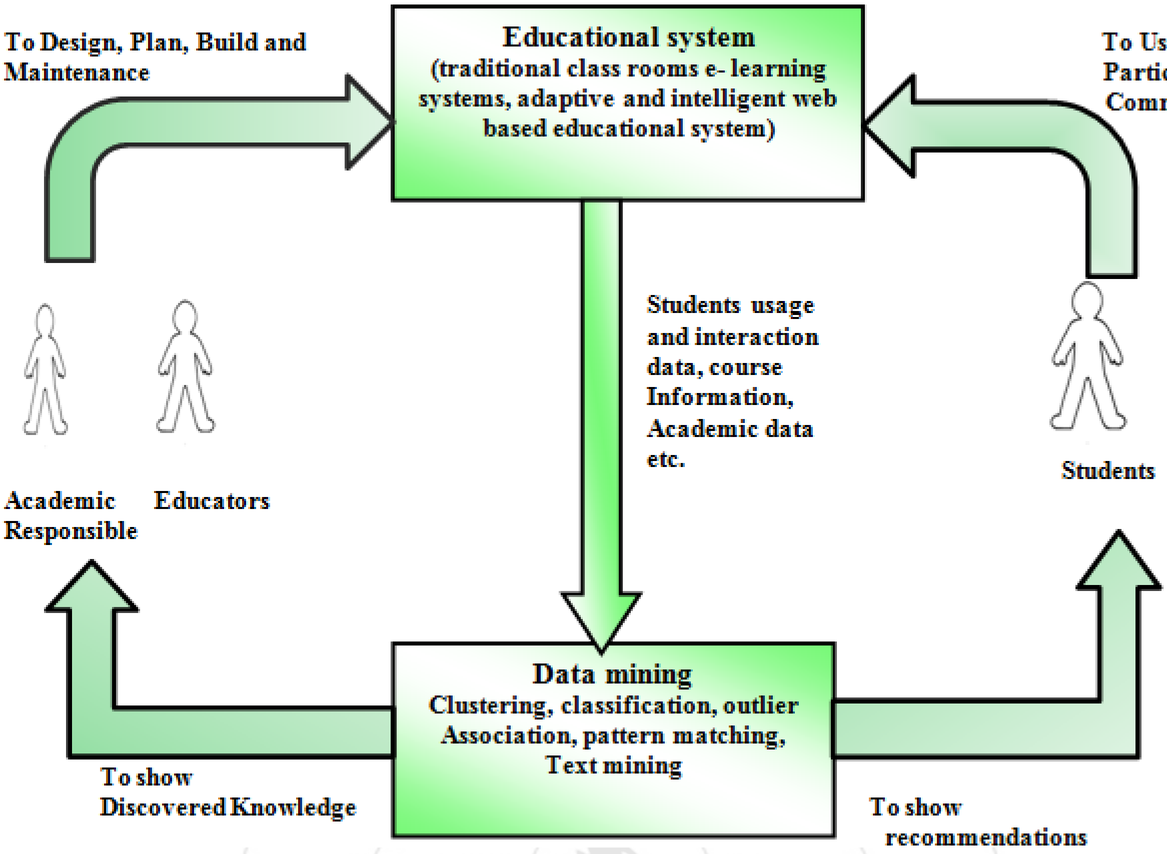
To Design, Plan, Build and Maintenance
Educational system
(traditional class rooms e- learning systems, adaptive and intelligent web based educational system)
To Use, Interact Participate and Communicate
Students usage and interaction data, course Information, Academic data etc.
Academic Responsible
Educators
Students
Data mining
Clustering, classification, outlier Association, pattern matching, Text mining
To show Discovered Knowledge
To show recommendations

Figure 1: Data mining system by Academic Educators cum students

2. Related Works

Yadav and Pal [4] obtained VBS University student's data like Discipline, Category, Student grade in high school, Admission type, medium and family size from the previous student database to predict the students who are likely to fail with the help of ID3, C4.5 and CART Algorithm. They observed that C4.5 is the best algorithm for predicting student result.

Hijazi et al [5] conducted a study along the student carrying out by taking a sample of 300 pupils (225 males, 75 females) from a group of colleges affiliated to Punjab university of Pakistan. He stated as Student's attitude towards attendance in class, hours spent in study on a daily basis after college, students' family income, students' mother's age and mother's education are much related with student performance" was formed.

Romero and Ventura [6], have a survey of educational data mining between 1995 and 2005. They inferred that educational data mining is a growing area of research and it has a special need not presented in other domains. Their work should be oriented towards the educational domain of data mining.

R.Shanmuga Priya [7] presented a paper on improving the student's performance using Educational Data Mining based by selecting 50 students from Hindustan College of Arts and Science, Coimbatore, India. By using the decision tree classifier on 8 attribute, it was identified that the class assessment, seminar, attendance, lab practical is used to predict the student functioning. This prediction will aid to the teacher to pay specific heed of students and improve student confidence in their studies.

C. Marquez, et al [8] had done research on diagonising the components that impress the low performance of pupils at different educational degrees. They obtained 670 middle school students' data from Zacatecas, Mexico. They used classification algorithms on various chosen components and found sociological, economic or educational characteristics that may be more relevant in the prediction of low academic performance in school students.

Pandey and Pal [9] had done research on the student performance by selecting 600 students from various colleges of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayes Classification of group, accent and background adequacy, it was deduced that whether newcomer students will perform or not.

Ryan S.J.D. Baker[10] and Kalinayacef reviewed the history and developments in the field of educational data mining (EDM) 2009. They focused on the increased importance on prediction, the development of work using existing models to make scientific discoveries.

Abeer and Elaraby [11] conducted a similar research that mainly targets on generating classification rules and predicting students' performance in a preferred course program based on previously filed students' behavior and activities. They processed and examined previously enrolled students' data in a specific course program across 6 years (2005–10), with multiple attributes taken from the university database. As a result, this study was able to predict, to a assured extent, the students' final grades in the preferred course program, as well as, "help the student's to enhance the student's performance, to identify those students who needed special attention to diminish failing ration and taking appropriate action at right time"[11].

Sudheep Elayidom , Sumam Mary Idikkula & Joseph Alexander [x4] proved that the technology named data mining can be very effectively applied to the domain called employment prediction, which helps the students to choose a good branch that may fetch them placement. A generalized framework for similar problems has been projected.[12]

3. EDM Phases

EDM generally consists of four phases [13]:

The first stage of educational data mining is to find the relationships between data of educationalenvironment. The aim of implementing these relationships is to utilize these relationships invarious data mining techniques like classification, clustering, regression etc.

The second phase of educational data mining is validation of discovered similarities betweendata so that over fitting can be avoided.

The third phase is to make predictions for future on the basis of ratified relationships in learning environment.

The fourth phase is supporting decision making progress with the help of predictions.

4. Applications of EDM

Educational data mining research examines the different ways that course management systems (CMS) data can be mined to provide novel kinds of pupil behavior.Solutions can help staff and staff with improving learning and sustaining educational processes, which in turn improve institutional effectiveness.

4.1 Student Retention and Attrition

Research has deduced that educational data mining can be applied to detect at risk pupils and help institutions become much more proactive in identifying and responding to those students (Luan, 2002). He applied data mining as a approach to expect what types of students would drop out of school, and then return to school later on. This research is important because it demonstrated the successful application of data mining tools to supportstudent retention endeavors.

Lin (2012) was able to create predictive models based on incoming students' data. The models were able to give short-term accuracy for predicting which categories of students would gain from student retention programs on campus.

4.2 Personal Learning Environments and Recommender Systems

Personal learning environments and personal recommendation systems also directly relate to educational data mining. Personalized learning environments focus on yielding the various tools, services, and artifacts so that the system can accomodate to student's learning needs on the fly (Mödritscher, 2010).

Recommender systems must be adapted when they are used in educational contexts because the recommendations should coincide with educational objectives. The reason is that it is not probable to assign existing recommender systems plainly to educational data because they are highly domain dependent (Santos & Boticario, 2010).

4.3 EDM And Course Management Systems

A large number of researchers within EDM focus directly on course management systems and how they can be improved to support student learning outcomes and student success. One research group built up a rearranged data mining toolbox that works inside the course administration framework and permits no – expert users to get data mining information for their courses (García, Romero, Ventura, & de Castro, 2011).

In an online educational environment, learner commitment is an important aspect of student success. Students' engagement with the course substance can be investigated utilizing information mining strategies to decide whether there are disengaged learners (Cocea & Weibelzahl, 2009).

5. Goals for educational data mining in educational field

Baker and Yacef [14] describe the following four goals of EDM:

5.1 Predicting student's future learning behaviour

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.

5.2 Findinging or improving domain models

Through the different strategies and uses of EDM, revelation of new and changes to existing models are conceivable.It characterizes the content to be learned and optimal instructional sequences.

5.3 Studying the effects of educational support

It can be achieved through learning systems. There are various techniques available in datamining to study the effects of educational support .

5.4 Advancing scientific knowledge about learning and learners

By building and incorporating student models, the field of EDM research and the technology and software used. Evolving scientific knowledge about learning and learners through building computational models that combine models of the student, the domain,and the software's pedagogy.

6. Data Mining Approaches in Prediction of Students Performance

Educational data mining (EDM) is a new stream in the data mining research field. It uses many approaches such as decision tree, rule induction, neural networks, k-nearest neighbour, naïve Bayesian . By applying these methods, many kinds of knowledge can be set up such as association rules, classifications and clustering.

6.1 Classification

Classification is generally assigned data mining strategy, which utilizes an arrangement of pre classified attributes to build up a model that can group the population of records at large. This approach regularly employs decision tree or neural network based classification algorithms. The data classification process involves learning and classification. In learning training data are analysed by algorithm. In classification test data are used to estimate the accuracy of the rules .

Classification is the most natural and best data mining procedure used to classify and anticipate values.Educational Data Mining (EDM) is no exemption to this fact, hence it can be used to analyze collected students' information through a survey, and provide classifications based on the collected data to predict and classify the students' performance in their upcoming semester.

6.1.1 Decision trees

Decision tree techniques are easy to understand and implement, It allows the addition of new possible scenarios, It helps to find worst, best and expected values for different scenarios, It can be combined with other decision tree techniques to generate rules easily[15]. This technique has many disadvantages as the number of training data increases like over fitting. It does not handle numeric data and pruning may become cumbersome. Decision trees can be worn to figure the understudy's lead in an instructive situation, his enthusiasm towards a subject or his result in the examination.

6.1.2 Bayesian Classifier

It includes Naive Bayes algorithm and its variations**.** This technique is simple and easy to understand, requires a small amount of training data to estimate the parameters, Fast Space efficient, Insensitive to irrelevant features and handles both real and discrete data well[16]. Patterns that are discovered by Bayesian Classifier from educational data can be used to enhance decision making in terms of finding students at risk, decreasing student dropout rate, increasing students' success and increasing students learning outcome.

6.1.3 Neural Networks

Neural network is another preferred technique used in educational data mining. The advantage of neural network is that it has the capability to spot all possible interactions between predictor variables. It includes algorithm like Multilayer Perception.[17] This technique is a generalized method, works well with noise. But it does not scale well from the small research system to large real-time system. It is possible to model an Artificial neural network that can be used to predict a candidate's performance based on some given pre admission data for a given student.

6.1.4 Support Vector Machine

A powerful Support Vector Machine (SVM) which was first proposed by Vapnik and it has a great potency of interest in the machine learning research community. Several past studies have described that the SVM generally has a proficient in delivering the high accuracy in classification when compared to other data classification algorithms. There are several advantages of SVM such as it uses greatest marginal hyper plane for classifying linearly separable data. In Educational Data Mining SVM Classifier can provide valuable information to departmental faculty members in making decisions.

6.1.5 C4.5 Tree

The most commonly, and nowadays probably the most widely used decision tree algorithm is C4.5. Professor Ross Quinlan created a decision tree algorithm named as C4.5 in 1993; it represents the result of research that traces back to the ID3 algorithm (which is also proposed by Ross Quinlan in 1986). C4.5 has additional features such as handling missing values, categorization of continuous attributes, pruning of decision trees, rule derivation, and others. Basic construction of C4.5 algorithms uses a method known as divide and conquer to construct a suitable tree from a training set. It can be used in educational data mining to predict academic performance of learners.

6.2 Clustering

In clustering, the goal is to find data points that naturally group together, splitting the full data set into a set of clusters. Clustering is especially benefical in cases where the most common categories within the data set are not known in advance.[18] Clusters can be created at several different possible grain-sizes: such as, schools could be clustered together (to examine similarities and differences between schools), students could be clustered together (to examine sameness and differences between students), or student actions could be clustered together (to investigate patterns of behaviour) Clustering algorithms can either start with no earlier hypotheses about clusters in the data (such as the k-means algorithm with randomized restart), or start from a specific hypothesis, possibly generated in earlier research with a distinct data set (using the Expectation Maximization algorithm to iterate towards a cluster hypothesis for the new data set).

The quality of a set of clusters is typicallyassess with reference to how well the set of clusters fits the data, relative to how much fit might be expected solely by chance given the number of clusters, using statistical metrics such as the Bayesian Information Criterion.[19]

6.3 Prediction

In prediction, the goal is to build up a model which can gather a single aspect of the data (predicted variable) from some blend of different aspects of the information (predictor factors).

This is a common advent in programs of research that attempt to predict student educational outcomes (Romero et al, 2008) without predicting intermediary or mediating components first. In a second type of handling, prediction approaches are applied in order to expect what the output value would be in contexts where it is not desirable to directly obtain a label for that construct.

Baker et al (2008) developed a prediction model by using observational methods to label a small data set, building up an expectation show utilizing consequently gathered information from cooperations amongst understudies and the product for indicator factors, and afterward approving the model's exactness when summed up to extra understudies and settings.. They were then able to learn their research question in the context of the full data set. Commonly, there are three types of prediction: classification, regression, and density estimation. In classification, the anticipated variable is a binary or categorical variable. In regression, the anticipated variable is a consistent variable. Various chosen regression proceduresin educational data mining comprised linear regression, neural networks, and support vector machine regression.[19]

7. Conclusion

The study was made with the help of research papers published by various authors research work. From their works they focused various approaches were used for the prediction. The different parameters were used by the researcher to classify the student's quality assessment according to their capability and IQ(Internal Quality) factors. Their research includes various educational data for prediction. It concludes so many data mining methods are available for performance analysis. It is an eye opening for conducive research in the field of educational data mining.

References

[1] Amirah Mohamed Shahiria, Wahidah Husaina,Nur'aini Abdul Rashida, "A Review on Predicting Student's Performance using Data Mining Techniques", The Third Information Systems International Conference, Procedia Computer Science 72 pp 414 – 422, 2015.

[2] B. Umadevi, D.Sundar, Dr.P.Alli,"A Study on Stock Market Analysis for Stock Selection – Naïve Investors' Perspective using Data Mining Technique", International Journal of Computer Applications (0975 – 8887),Vol 34– No.3, 2011.

[3] NityaUpadhyay, VinodiniKatiyar, "A Survey of the Classification Techniques In Educational Data Mining", International Journal of Computer Applications Technology and Research, Vol.3,Issue 11, pp 725 – 728, 2014.

[4] S. K. Yadav and S. Pal, "Data Mining: A prediction for performance improvement of Engineering students using classification", World of Computer Science and Information Technology Journal (WCSIT), Vol. 2, No. 2, pp51-56, 2012.

[5] Hijazi,S.T., and Naqvi, R.S.M.M., " Factors Affecting Student's Performance: A Case of Private Colleges", Bangladesh e-Journal of Sociology, 2006.

[6] Romero, C. And Ventura, S., "Educational data mining: A Survey from 1995 to 2005", Expert Systems with Applications (33), pp. 135-146, 2007.

[7] ShanmugaPriya,"Improving the student's performance using Educational data mining", International Journal of Advanced Networking and Application, Vol.4, pp1680-1685, 2013

[8] C. Marquez-Vera,C.Romeroand S.Ventura, "Predicting School Failure Using Data Mining",2011.

[9] U.K. Pandey, and S. Pal, "Data Mining: A prediction of performer or underperformer using classification", (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), pp.686-690.

[10] BakerRSJd, Yacef K, "The state of educational datamining in 2009: A review and future visions", 2009.

[11] Ahmed, A.B.E.D. and Elaraby, I.S., " Data Mining: A prediction for Student's Performance Using Classification Method", World Journal of Computer Application and Technology, Vol 2, pp.43-47, 2014.

[12] Sudheep Elayidom, Sumam Mary Idikkula& Joseph Alexander,"A Generalized Data mining Framework for Placement Chance Prediction Problems" , International Journal of Computer Applications, Volume 31– No.3, 2011.

[13] Tripti,DwivediDiwakar Singh, "Analyzing Educational Data through EDM Process: A Survey", International Journal of Computer Applications, Vol 136 ,No.5, 2016

[14] Baker RS,Yacef K, " The state of educational data mining in 2009: A review and Future visions". JEDM-Journal of Educational Data Mining, 2009.

[15] B. Umadevi, ,D.Sundar, Dr.P.Alli, "An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50",International Journal of Data Mining & Knowledge Management Process (IJDKP),Vol.3, No.1, 2013.

[16] B. UmadeviD.Sundar, Dr.P.Alli,"An Optimized Approach to Predict the Stock Market Behavior and Investment Decision Making using Benchmark Algorithms for Naive Investors", Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on (IEEE Xplore Digital Library),pg1 -5.,2013

[17] K.R.Kavyashree,LakshmiDurga,"A Review on Mining Students' Data for Performance Prediction", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, 2016

[18] B. Umadevi,,D.Sundar, Dr.P.Alli, "Novel Framework For The Portfolio Determination Using PSO Adopted Clustering Technique", Journal of Theoretical and Applied Information Technology", Vol. 64 No.1, 2014

[19] Ryan S.J.d. Baker, “ Data Mining for Education”, International Encyclopedia of Education (3rd edition). Oxford, UK: Elsevier, 2009.

Author Profile

Dr. B. Umadevi has received her Doctoral degree in Computer Science from Manonmaniam Sundaranar University, Tirunelveli, India. Currently working as Assistant Professor & Head- P.G and Research Department of Computer Science, Raja Doraisingam Government Arts College, Sivagangai-Tamilnadu, India. She has over 22 years of Teaching Experience and published her research papers in various International, National Journals and Conferences. Her research interests include DataMining, Soft Computing and Evolutionary Computing. She got the Best Paper Award for her publication in the IEEE International Conference on Computational Intelligence and Computing Research held on 27th Dec 2013 at VICKRAM College of Engineering and Technology.

R. Dhanalakshmi is a M.Phil Research Scholar in PG & Research Department Of Computer Science, Raja Doraisingam Government Arts College, Sivaganga, Tamilnadu, India. Her research interest includes in Data mining and its applications.