Classification Techniques Applied to Educational Data

Manjula Sanjay Koti[a] and Samyukta D. Kumta[a]

[a]*Master of Computer Applications, Dayanada Sagar Academy of Technology and Management, Bangalore, Karnataka, India*
E-mail: manjula.dsce@gmail.com, samyukta.dsatm@gmail.com

Abstract: The recent advancements in educational technology has led to increase in computing power and the student's use of computer based learning environment wherein sophisticated techniques can be applied to analyse huge amount of data which is generated from educational settings. The analysis of educational data is not a new practice. It provides us with an enhanced understanding of students' knowledge & effective valuations of their progress. Education system faces a lot of issues and Data mining provides a set of techniques, which can favour the educational system to overcome these issues and improve the quality of education. This paper explores about educational data mining, its phases and goals. We have also explained the various classification techniques that could be applied on educational data.

Keyword: Educational Data Mining, Universities, Educators, Classification

1. INTRODUCTION

Data mining aims at discovering insightful, interesting and novel patters, as well as descriptive, understandable, and predictive models from large-scale data[4].It is concerned with finding new patterns in large amount of data. Data mining finds applications in various domains and one such area is in educational environment.

The important information stored in data warehouse could be analysed using data mining technology, where it extracts the hidden predictive information which helps in effective decision making and hence it is termed as knowledge discovery in database (KDD) [2]. Data mining discovers hidden trends and patterns through explicit knowledge base, sophisticate analytics skills and domain knowledge [1, 3].

This paper is organised as follows: Section I deals with Introduction to data mining, Section II explores about Educational Data mining, goals and phases. Section III enlists the Literature Survey. Section IV deals with various approaches to classification techniques. Section V gives the Conclusion.

2. EDUCATIONAL DATA MINING

Educational Data Mining(EDM) has come up with lot of innovations in terms of improvisation of educational software and student learning[5]. One of the promising and emerging area of research is Education data mining is that it tries to extract useful previously unknown patterns from educational database for better understanding, improved educational performance and assessment of the student learning process [3].

EDM aims at developing methods to find the unique kind of data that emerges from education standards. These methods can be utilized to understand the environment in which the students learning behaviour happens. The methods that are utilized by EDM are quite different from the standard Data mining methods. EDM is an emerging discipline in the research field by data mining [6,7] that focus on evolving methods to explore the unique kinds of data which arise from educational settings. These methods intern helps to understand students and educational settings which they learn through[10].

A prediction model can be develop to identify the slow learners and distinguish students with low academic achievement where this model can be utilised by teachers, parents and educational planners to analyse the student's patterns in terms of current behaviour that could be associated with positive and negative outcomes of the past and also discussions could be made to rectify the problems associated with that [4,8].

2.1. The Major goals of EDM are

1. One can predict the students learning behaviour by developing the student models and this can be achieved through the parameters viz., student's knowledge, behaviour, academic information etc[12].

2. Upgradation of the domain models that distinguishes between the relevant content to be learnt and perfect pedagogical sequences

3. Analyse the effects of certain types of instructional support provided through learning.

4. Upgrading scientific knowledge in learners [17].

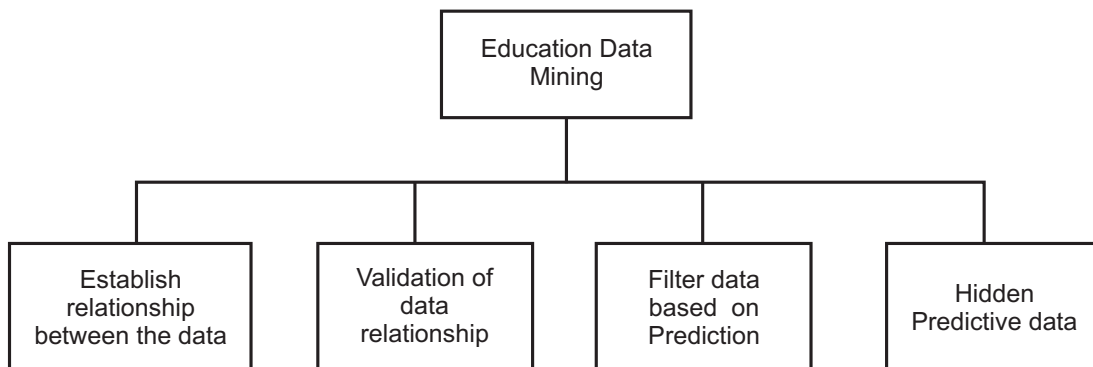## 2.2 EDM can be categorised into four different phases.

Figure 1: Phases of EDM

Phase 1: The relationship between the data is established by exploring the training set of data. This process can be achieved through the various Data mining algorithm such as classifier, regression etc...

Phase 2: Validation of data relationships is done which can be used for prediction in learning environment. Accordingly, decisions can be made on the basis of prediction and this can be utilised to make various policies for the institute.

Phase 3 and 4: Filtering data takes place when missing data, redundant and entries are eliminated [14,17].

2.3. Different actors of EDM

The various users of EDM are Students, Course developers, Administrator, Educators and Universities. Fig 2 shows the various users of EDM.
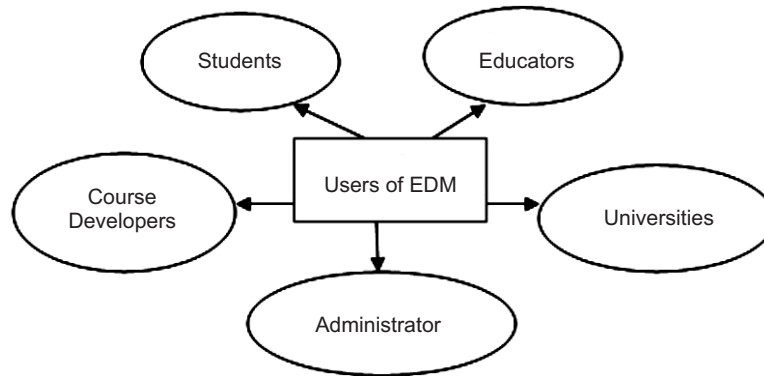
Figure 2: Actors of EDM

Students are interested in their learning behaviour, their needs and methods to improve their experience and performance. This can be achieved by activities, resources, learning tasks, courses, relevant discussions and books [11,14].

Educators use EDM to (*i*) Acquire objective feedback about instruction (*ii*) Analysing students learning and behaviour (*iii*) Identify which students require support (*iv*) Predicting student performance (*v*) Categorising learners into groups (*vi*) Discover the patterns for regular and irregular learners (*vii*) Determine the most effective activities of the students (*viii*) Advance the adaption and customization of courses, etc.

Course Developers use EDM to evaluate and Manage the courseware for (*i*) Improvisation of student learning (*ii*) Evaluating the structure of course content and its effectiveness in the learning process (*iii*) Atomised construction of student models and tutor model[13] (*iv*) Comparing data mining methods to recommend the useful one for each task (*v*) Developing specific data mining tools for education purposes etc[15].

Universities uses EDM to (*i*) Enrich the decision processes in higher learning institutions and restructure the efficiency in the decision making process (*ii*) Accomplish the specific objectives to recommend certain courses that might be valuable for each class of learners (*iii*) Find the most cost-effective way of improving retention and grades (*iv*) Identify the most qualified learners for graduation (*v*) Help to admit students who will do well in university etc.

Administrators Use EDM to (*i*) Develop the best way to provide institutional resources to students (human and material) and educational offers (*ii*) Utilize the available resources more effectively (*iii*) Improvisation of academic educational program and evaluate the effectiveness of the distance learning program (*iv*) Upgrading teacher's knowledge by providing educational trainings (*v*) Setting up the parameters for improving web-site efficiency and adapting it to users (optimal server size, network traffic distribution, etc.)[14].

3. LITERATURE SURVEY

In [1], the author has suggested that Data Mining is an emerging methodology that is used in educational field to enhance the understanding of learning process. In [9], the authors have discussed about Student performance in an university courses, which is of great concern to the higher education, where several factors may affect the performance and provide information on predictive modelling which is a combination of mathematics, computer science and domain expertise. In [10], the author has conducted a study on students learning behaviour, the students evaluation factors like class quizzes, mid and final exam assignment etc. He has also recommended that all these correlated information should be conveyed to the class teacher before the conduction of final exam. This study will help the teachers to reduce the drop out ratio to a significant level and improve the performance of students.

In [5], the authors has made a study on education system of "Portugal", the results showed that a good predictive accuracy can be achieved, provided that the first and/or second school period grades are available. In [11], the author has discussed about the positive role of video conferencing technologies in online education. This study particularly concentrates on the creation of brainstorming style-discussions and small group meeting which are fundamental to many of modern educational techniques. In [12], the authors used genetic algorithms and a combination of multiple classifiers to predict students' final grades. In [13] the authors has presented how educational data mining prediction methods can be used to develop student models. They have used a variety of variables to predict whether a student will make a correct answer. This work has inspired a great deal of later educational data mining work – student modelling is a key theme in modern educational data mining, and the paradigm of testing EDM models' ability to predict future correctness – advocated strongly by Beck & Woolf – has become very common.

In [8] the author states that "Educational data mining" (EDM) is a new growing research area and the essence of data mining concepts are used in the educational field for the purpose of extracting useful information on the behaviour of students in the learning process. He has developed a predictive data mining model for students' performance so as to identify the slow learners and study the influence of the dominant factors on their academic performance. In [15], the author has suggested an classification technique for predicting the student's performance in examination. In [16], the author has presented a new approach of classification to predict the placement of students. This approach provides the relation between academic records and placement of students. In this analysis, various classification algorithms are employed by using data mining tools like WEKA for study of student's academic records.

In [17][18], the authors discuss that in the field of education data mining it is found that analysing different literature for predicting the performance, placement prediction, scholarship prediction for recruitment process different factor plays a vast role to measures the accurate result which include: student grade in secondary exam, mother's education, living location, medium of teaching, family income, student other habits, student family status, student learning activities, learning behaviour, language known to him, decision making capability, grades in last year exam and many more criteria related to student habits are co related to analyse the prediction of various fields.

4. CLASSIFICATION TECHNIQUES

Classification is an unsupervised technique that can be used to find the unknown patterns which in turn are used to predict the classes or instances. Classification technique is divided into two phases: the training and testing phase. Training phase deals with data records with known classes which are used to build models. In the testing phase, the model which is constructed in training phase is used to predict the unknown classes.

One of the technique in data mining is classification, which is used to predict the group membership for data instances[16]. Some of the traditional classification methods are decision tree induction, k-nearest neighbour classifiers, Bayesian networks, support vector machines, rule based classification, case-based classification, case-based reasoning, fuzzy logic techniques, genetic algorithm, rough set approach and so on[19].

classification, based on quantitative information the discrete items are positioned into various groups. This technique generalises known structure, which can be related to new data.

Classification tasks can be utilised to predict academic success, the course outcomes, Succeeding in the next task, Metacognitive skills, habits, and motivation [18].

4.1. Classification approaches

Classification algorithms are divided into Lazy learners and eager learners. Algorithms such as decision tree classifier, Bayesian classifier, support vector classifier, linear regression are eager learners. Nearest neighbour classifiers are lazy learners, since they wait until a test instance is arrived for classification to perform generalization[20].

Decision trees are the best known classification paradigm. A decision tree represents a set of classification rules in a tree form [9]. It is one of the analytical modelling approaches used in statistics, data mining and machine learning. Decision tree methods are used to evaluate students' performance using datasets such as, students' final grades, marks obtained in particular courses etc …[12,15]

Bayesian Classification technique used to determine whether the given tuple belongs to a particular class or not. It exhibit high accuracy on, when applied to large data sets. The two fundamentals methods of Bayes' Algorithm are Naïve Bayes and Bayesian Network. Bayesian models are fast to train, evaluate and have a high accuracy in many domains [12].

Artificial neural network is another popular technique selected as one of best prediction method in educational data mining[13].The attributes analysed by neural network are admission data, students attitude towards self-regulated learning and academic performance.

K-nearest neighbour classifiers represent a totally different approach to classification. This algorithm gives the best performance with the good accuracy. K-nearest algorithm will help to identify the students' performance for slow learners, average learners, good learners and excellent learner.

Support vector machines (SVMs) are an ideal method used to predict students' performance. It is suited well in small datasets. The studies demonstrated that support vector machine method has acquired the highest prediction accuracy in identifying students at risk of failing.

Linear regression is actually not a classification method, but it works well, when all attributes are numeric. For example, passing a course depends on the student's points, and these points can be predicted by linear regression.

Classification has many applications in both traditional education and modern educational technology. The best results are achieved, when classifiers can be learnt from real data, but in educational domain the data sets are often too small for accurate learning [19,20].

5. CONCLUSION

Education System faces a lot of issues and data mining techniques can favour the educational system to overcome these issues and improve the quality of education. These Data mining techniques can be implemented in education system to predict the academic performance of student. Based on these predictions, the academic performance of student can be enhanced. In order to improve the performance quality, prediction of student performance has become the prominent focus. We have discussed about the goals and phases of Educational Data Mining along with the various classification algorithms that could be applied on the education data. Also a thorough literature survey has been done to explore about educational data mining and various techniques of Data mining that could be applied on such kind of data.

6. ACKNOWLEDGEMENT

REFERENCES

[1]   G.Paul Suthan 1 and Lt.Dr. Santhosh Baboo "Hybrid CHAID a key for MUSTAS Framework in Educational Data Mining" ISSN (Online): 1694-0814 IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011

[2]   K.Shanmuga Priya,A.V.Senthil Kumar "Improving the Student's Performance Using Educational Data Mining" Int. J. Advanced Networking and Applications ISSN : 0975-0290 Volume: 04 Issue:04 Pages:1680-1685 (2013).

[3]   Pooja Gulati ,Dr. Archana Sharma Research Scholar, "Educational Data Mining for Improving Educational" Quality International Journal of Computer Science and Information Technology &amp, IIS University Computer Science Department IRACST Security (IJCSITS), ISSN: 2249-9555 Vol. 2, No.3, June 2012

[4] T ryan s.j.d. baker, A Review and Future Visions Department of Social Science and Policy Studies Worcester Polytechnic Institute Worcester, ma usa and kalina yacef School of Information Technologies University of Sydney Sydney, NSW Australia.2009

[5] Jaseena K.U. 1 and Julie M. David "ISSUES , CHALLENGES , AND SOLUTIONS : BIG DATA MINING" Department of Computer Applications, M.E.S College, Marampally,Aluva, Cochin, India.2014

[6] Mrs. M.S. Mythili , Dr. A.R.Mohamed Shanavas "An Analysis of students' performance using classification algorithms Bishop Heber College, Tiruchirappalli 620 017,TamilNadu, India.

[7] E. Baker, "International Encyclopedia of Education" (3 rd edition), Oxford, UK: Elsevier-2010.

[8] M. Ramaswami and R. Bhaskaran "A Study on Feature Selection Techniques in Educational Data Mining". Journal of Computing, Vol. 1, December 2009.

[9] Al-Radaideh, A. Qasem, E. M. Al-Shawakfa, and M.I. Al-Najjar, "Mining Student Data Using Decision Trees", International Arab Conference on Information Technology, 2006.

[10] Ayesha, Shaeela, T. Mustafa, A.R. Sattar, M.Inayat Khan, "Data Mining Model for Higher Education System", European Journal of Scientific Research,Vol.43, pp.24-29,2010.

[11] Sam Redferm and Neil Naughton, "Collaborative Virtual Environments to Support Communication and Community in Internet-Based Distance Education," Journal of Information Technology Education, vol. 1, no. 3, pp. 201-211, 2002

[12] B. Minaei-Bidgoli and B. Punch, "Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System," Genetic and Evolutionary Computation, vol. 2, pp. 2252–2263, 2003

[13] Beck, J., & Woolf, B High-level student modeling with machine learning. Proceedings of the 5th International Conference on Intelligent Tutoring Systems, 584–593, 2000.

[14] Kumar S. Anupama and N. vijaylaxmi ,Efficiency of Decision trees in predicting Student's Acedemic performance. M.2011

[15] Pal A. kumar and Pal S. Classification Model of Prediction for Placement of Students. . 2013

[16] A Survey on the Classification Techniques In Educational Data Mining Nitya Upadhyay RITM Lucknow, India Vinodini Katiyar Shri   Ramswaroop Memorial University Lucknow, India.2014

[17] Pal, B. K. quot;Mining Educational Data to Analyze Students Performance".IJACSA. 2011

[18] S.Pal, U. a. "Data Mining :Aprediction of performer or underperformer using classification"(IJCSIT). IJCSIT, 2011

[19] Espejo, P., Ventura, S., Herrera, F. A Survey on the Application of Genetic Programming to Classification. IEEE Transactions on Systems, Man, and Cybernetics-Part C.2010.

[20] W. H ̈am ̈al ̈ainen and M. Vinni Classifiers for educational data mining Hand Book of Educational Data Mining.2010