

Survey on Prediction the Performance of Students Using Educational Data

^AKinjal Jakhariya ^BShantanu Santoki

^AStudent, M.E.(Computer), Gardi Vidyapith, Gujarat, India

^BAssistant Professor, CSE/IT Department, Gardi Vidyapith, Gujarat, India

Abstract—Education is very wide and important subject. Today education database storage is growing rapidly. From That database we can predict very important the performance of students and improve them. We have many data mining techniques for the prediction of student performance. We can compare all the method and analysis the obtain result.

Keywords- Prediction, Student , Performance , Education

I. INTRODUCTION

Data mining is also called Knowledge Discovery in databases (KDD), is the field of determine useful information from huge amounts of data. Data mining has been applied numerous arenas including bank, medical, census data, airline passenger records and super market, bioinformatics, business etc. Now a day data mining is also used in Education. Applying data mining technique in Educational setting is called as educational data mining (EDM). Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students [5]. Although data mining has been successfully implemented in the business world for some time now, its use in higher education is still relatively new, i.e. its use is intended for identification and extraction of new and potentially valuable knowledge from the data.[13] Educational Data Mining has various techniques like regression, prediction, clustering, Decision Tree, classification, Association Rules etc.

Today the important challenge that higher education faces, is reaching a stage to facilitate the universities in having more efficient, effective and accurate educational processes[14]. Predicting students’ performance proved to be helpful for identifying poor performers and it can enable tutors to take remedial measures at an earlier stage, even from the very beginning of an academic year using only students’ demographic data, in order to provide additional help to the groups at risk [2]. We can also predict Academic trends and pattern also. We can obtain data from any educational background. We can construct class work, schedule, method, materials, extra classes, plan using educational data mining. Prediction models that include all personal, social, psychological and other environmental variables are necessitated for the effective prediction of the of the students performance.

We can show Data Mining methodology in figure 1.[4]

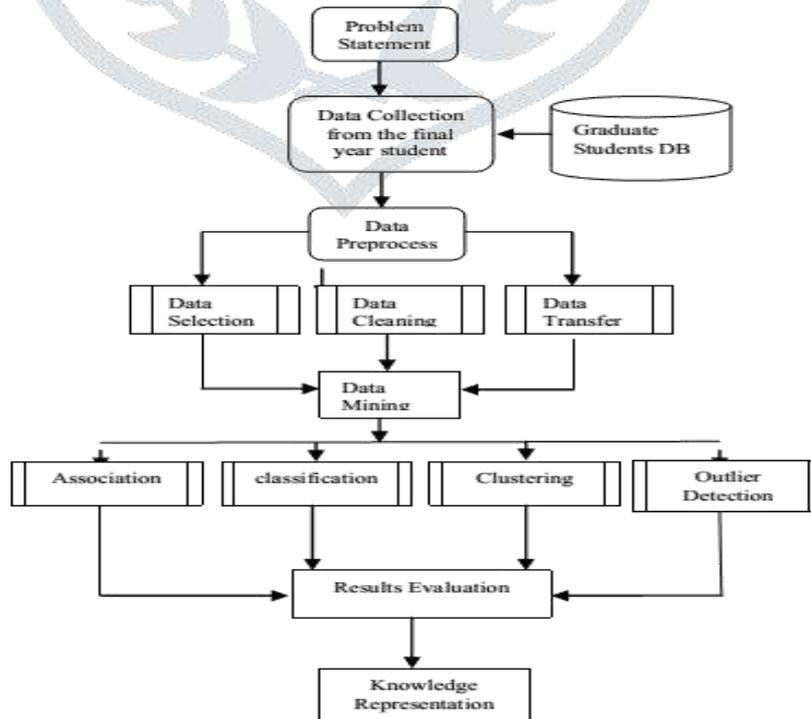


Fig 1. Data Mining Methodology

Educational Data Mining cycle is given below fig 2.[11].

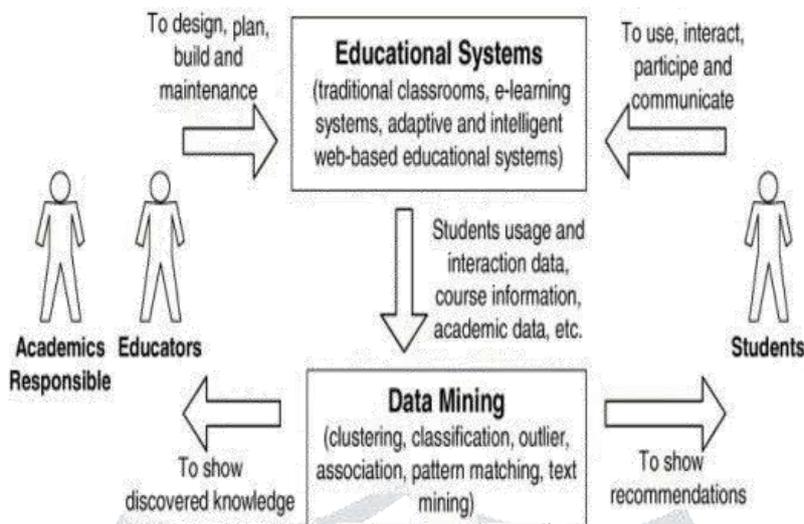


Fig 2.The cycle of applying data mining in educational Systems[11].

II. RELATED WORK

There are a lot of techniques and applications available to predict the student’s performance based on Grade.

We can predict the performance using backpropagation and counterpropagation. Backpropagation and counterpropagation networks can be trained to give reasonable predictions of student performance in Calculus.[7] Backpropagation networks with a sufficient number of hidden units (25 or 30) achieve very high levels of accuracy in predicting outcomes, for either the training data or testing data that is reasonably similar to the training data. Although backpropagation nets take significantly longer to train than counterpropagation networks, the use of the Nguyen-Widrow initialization of the weights provides an extremely easy and effective method of improving the speed of training.[7]

Association Rule mining is very popular and interesting method for find relationship between two variable from dataset. Associations rules are satisfied same time minimum support and minimum confidence which is user specified. We can show work methodology of association rule mining in figure 2.[3]

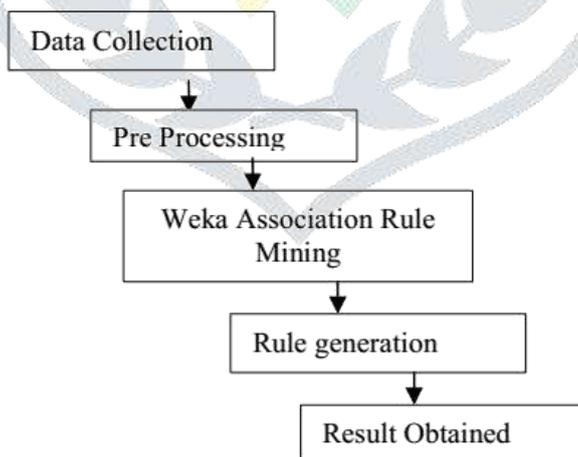


Fig 3 Association rule mining methodology

The association rule mining can be viewed as a two-step process:

- 1) Find all frequent item sets: Each of these item sets will occur at least as frequently as a predetermined minimum support count.
- 2) Generate strong association rules from the frequent item sets: The rules must satisfy minimum support and confidence. These rules are called strong rules.[4]

Sum of Absolute Difference(SAD) and Sum of Squared Difference(SSD) is another method for predict the performance of students. This method is originally comes from the Image Processing. To find Correlation between image blocks, Sum of Absolute Difference is widely used[10]. So, it can be considered as a Correlation method. It can be applied on images by taking absolute differences between every pixel from original block to corresponding pixel in the other image block. After that all absolute differences are summed together. The least value of sum is considered as the best matching block for that image.[1] We have six equation of different method which we can use for prediction student performance. It is shown in following table 1 [1].

In SAD algorithm we have to make reference mark sheet of the student and then compare the actual marks from the reference marksheet and find the difference between them and final summed all together. Following is some steps of SAD algorithm[6]:

1. Computation of difference between corresponding elements
2. Determine the absolute value of each difference
3. Add all absolute value

From this algorithm we can predict the future result from the internal marks of the students. For the testing our prediction result we can make software. For SAD and SSD algorithm we can make software dot net platform[1].

We can use the distance measurement equation for prediction the student performance. We can use for this students previous examination marks. Distance measurement equation is Euclidean Distance, Manhattan Distance, Squared Euclidean Distance, Bray Curtis Distance, Canberra Distance, Chessboard Distance, Cosine Distance. Distance measure between KNOWN(p, q, r) and UNKNOWN(xj, yj, zj)[8]

Euclidean Distance	$\sqrt{\sum_j [Abs(p-x_j)^2 + Abs(q-y_j)^2 + Abs(r-z_j)^2]}$
Manhattan Distance	$\sum_j [Abs(p-x_j) + Abs(q-y_j) + Abs(r-z_j)]$
Squared Euclidean Distance	$\sum_j [Abs(p-x_j)^2 + Abs(q-y_j)^2 + Abs(r-z_j)^2]$
Bray Curtis Distance	$\frac{\sum_j [Abs(p-x_j) + Abs(q-y_j) + Abs(r-z_j)]}{\sum_j [Abs(p+x_j) + Abs(q+y_j) + Abs(r+z_j)]}$
Canberra Distance	$\sum_j \frac{Abs(p-x_j)}{p+x_j} + \frac{Abs(q-y_j)}{q+y_j} + \frac{Abs(r-z_j)}{r+z_j}$
Chessboard Distance	$MAX[Abs(p-x_j), Abs(q-y_j), Abs(r-z_j)]$
Cosine Distance	$1 - \frac{px_j + qy_j + rz_j}{\sqrt{p^2 + q^2 + r^2} \sqrt{x_j^2 + y_j^2 + z_j^2}}$

Table1 Similarity measures for numerical data[8]

In paper 8 the author used the Canberra Distance algorithm and take the data of VIT university for the prediction result. They use the c# language and use dot net software for the testing. They take the three test cat1 , cat2 and quiz for the prediction of marks and give grade from the reference marks. Then from the marks they make the output cluster.

We can also predict the student performance using web educational data. classifying students in order to predict their final grade based on features extracted from logged data in an education web-based system. To improve the classification performance we can used classifier. As more instructors develop educational materials for their courses to use with the LONCAPA system, the content of this database grows. [9]. They used Genetic Algorithm for improve accuracy. To used GA algorithm they improve accuracy atleast 10%. It has considerably gained in importance and thousands of web courses have been deployed in the past few years. But many of the current web-based courses are based on static learning materials, which do not take into account the diversity of students.[15]

Now a days No of Graduate students are increase, so to find a job is very difficult. We can make a decision tree using ID 3 algorithm. The decision tree obtained through data mining reveals the relationship of several key factors affecting the graduates' employment. At present, many units require graduates to obtain the appropriate degree, without which would make job seeking very difficult; in addition, despite the fact that one does obtain a degree, with poor professional accomplishments, communication skills, and foreign language capabilities, it is still very difficult to secure a job [12]. This decision tree is used for prediction of non graduate employment is possible, and we can guide this students to improve themselves in their weakest point. We can take a factor for employment or non employment is based on communication skill, professional achievements, English capabilities.

They apply Apriori algorithm to the academic records of a group of students and obtain the association rules which will help in student profiling. And also use K-means clustering to group the students categorically. They used 20% minimum support for association rule mining. They use students academic record file and various parameter like Exam Marks, Term Work Grades, attendance, practical marks. They used weka to get best association rule and that based we can profile students performance as good, satisfactory and poor. Here minimum support is take 20%. Since the data used to demonstrate this is not very large, we have made use of a low minimum support value. The higher the minimum support we provide, the stronger the association rules we obtain. Based on our input file, the best obtained rule states that if the exam marks are 80-100 and if the attendance is High, then the student can conveniently be profiled as Good. There are totally 101 instances. These 101 instances have been partitioned into 2 clusters with 57 and 44 instances. Clustering is mainly done to find the centers of the natural clusters in the data. If we go manually huge data set to identify students trends and behavior, it is very difficult. Instead is we can use of data mining technique on huge data set of academic record and easily group them, identify hidden patterns of their learning style, behaviour and performance. [12]

III. CONCLUSION AND FUTURE WORK

This paper surveys that prediction of students performance current trend in data mining. Student can predict about success and failure about examination. From the prediction of students performance we can warn the students that they must improve their performance by hardworking. The main objective of educational institute is to provide best education for their students and improve their performance. So this prediction model is the best for that educational institute. Predict the performance of students and help them to improve their result by their institute.

In above table 1 has six equation. There are two algorithm SAD and SSD is used for the prediction. Another four algorithm we can used for the prediction of student performance

ACKNOWLEDGEMENT

I am deeply indebted & would like to express mgratitude to my thesis guide Prof. Santanu Santoki and Prof. Prashant Mehta , B. H. Gardi College of Engineering & Technology for his great efforts and instructive comments in the dissertation work.I would also like to extend my gratitude to Prof.Hemal Rajyaguru, Head of the Computer Science & Engineering Department, B. H. Gardi College of Engineering & Technology for his continuous encouragement and motivation.I would also like to extend my gratitude to Prof. Vaseem Ghada, PG Coordinator, B. H. Gardi College of Engineering & Technology for his continuous support and cooperation.I should express my thanks to my dear friends & my classmates for their help in this research; for their company during the research, for their help in developing the simulation environment.I would like to express my special thanks to my family for their endless love and support throughout my life. Without them, life would not be that easy and beautiful.

REFERENCES

- [1] Mining on Student's Records to Predict the Behavior of Students , Kartik N. Shah, Shantanu Santoki, Himanshu Ghetia, L.Ramanathan, IEEE - International Conference on Research and Development Prospects on Engineering and Technology(ICRDPET 2013) March 29,30 - 2013 Vol.4
- [2] Predicting Students' Marks in Hellenic Open University, Sotiris B. Kotsiantis & Panayiotis E. Pintelas, Proceedings of the Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05).
- [3] Predicting Students Academic Performance Using Education Data Mining, Suchita Borkar, K. Rajeswar, IJCSMC, Vol. 2, Issue. 7, July 2013, pg.273– 279
- [4] A Study on Student Data Analysis Using Data Mining Techniques, Umamaheswari. K S. Niraimathi, International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 8, August 2013.
- [5] A Survey on Predicting Student Performance, A.Dinesh Kumar ,Dr.V.Radhika Sri Krishna Arts and Science College Coimbatore, India, International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014, 6147-6149.
- [6] Minimum Sum of Absolute Differences implementation in a single FPGA device, Joaquin Olivares, Javier Hormigo,

Julio Villalba, and Ignacio Benavides

- [7] Predicting Performance from test scores using backpropagation and counterpropagation I. v. fauset and w. elwasif,iee.
- [8] Clustering students' based on previous academic performance, Kartik N. Shah, Srinivasulu Kothuru and S. Vairamuthu, International Journal of Engineering Research and Applications (IJERA), Vol. 3, Issue 3, May-Jun 2013, pp.935-939.
- [9] Predicting Student Performance: An Application Of Data Mining Methods With The Educational Web-Based System Lon-Capa, Behrouz Minaei-Bidgoli Deborah A. Kashy Gerd Kortemeyer, William F. Punch , 33rdASEE/IEEE Frontiers in Education Conference, November 5-8, 2003, Boulder, CO.
- [10] Sum of Absolute Differences Algorithm in Stereo Correspondence Problem for Stereo Matching in Computer Vision Application, Rostam Affendi Hamzah osman Abd Rahim Zarina Mohd Noh ,Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on, (Volume:1), 652 – 657.
- [11] Survey Of Data Mining In Socio-Academic Perspective, Prof. K. Vijayalakshmi, International Journal Of Scientific & Technology Research Volume 2, Issue 9, September 2013 .
- [12] Application of data mining in employment instruction of undergraduate students, ZangWeihua, Guo rui, Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on (Volume:3), 772 – 775.
- [13] Application of Data Mining in Educational Databases for Predicting Academic Trends and Patterns, Suhem Parack, Zain Zahid, Fatima Merchant ,Technology Enhanced Education (ICTEE), 2012 IEEE International Conference on, 3-5 Jan. 2012,1 – 4.
- [14] Data Mining Approach For Predicting Student Performance ,Edin Osmanbegović , Mirza Suljić, Economic Review – Journal of Economics and Business, Vol. X, Issue 1, May 2012 .
- [15] Mining Association Rules in Student's Assessment Data, Dr. Varun Kumar, Anupama Chadha, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012.
- [16] A Study on Educational Data Mining, Smitha Harikumar, International Journal of Computer Trends and Technology (IJCTT) – volume 8 number 2– Feb 2014.

JETIR

AUTHORS BIOGRAPHY



Mrs. Kinjal Jakhariya completed B.E in information technology from Government Engineering College Modasa and M.E in CE (pursuing) from B .H. Gardi college of engineering and technology, Rajkot Gujarat.



Mr. Shantanu Santoki completed B.E. in CE from Government Engineering College Chandkheda. And M.Tech. completed in CE from VIT university, Chennai. Currently working as assistant professor in B.H. Gardi College of Engineering and Technology.