



ISSN: 0975-766X  
CODEN: IJPTFI  
Research Article

Available Online through  
[www.ijptonline.com](http://www.ijptonline.com)

## PREDICTION OF STUDENT PERFORMANCE IN EDUCATIONAL DATA MINING - A SURVEY

T.Archana\* and Usha Devi Gandhi  
SCOPE and SITE, VIT University, Vellore.  
Email: [archana.t@vit.ac.in](mailto:archana.t@vit.ac.in)

Received on 10-07-2016

Accepted on 29-08-2016

### Abstract:

Adaption of global culture in education system has paved an imbalance in intake and pass out ratio of educational institutions. It opens up an enormous research work to try balancing this imbalance. This paper surveys on the research works focusing on students' performance prediction, analysis to improve performance, increase student retention published since 2004 to till date.

**Keywords:** Student performance prediction, Education Data Mining, Dropout, Retention, Traditional classroom, online tutoring system.

### Introduction

In the past few decades, many educational institutions have sprung up. The intake of students to this institution is decently high. But the number of student passes out has a high degree of deviation to its intake. This is due to the increase in low performance of the student and high dropouts from institution. Educational institutes has enormous data on socio, economic, academic, skill etc. of a student, but this data is kept as a junk. It is rarely used to refer the alumni record. Educational Data Mining addresses this problem. Education Data Mining refers to techniques, tools and research designed for automatically extracting meaning from large repositories of data generated from learners activities in an educational environment.[26]. Prediction is a Data mining method used to discover the unknown value of a variable. Student data can be made used to predict the behavioural pattern of performer and non-performer. This may in turn help the tutor to take extra remedial care on the student and institution can frame policies and a measure to handle the non-performers on a timely basis highly increases the performers. In developing county like India, dropouts from educational institution are a common scene. Mostly students with high potential at the time of joining an institute may fall prey to this category. There are several factors like social, economic, emotional, academic may be the cause of the dropouts. Early detection of such factors may help the student to retain the studies.

## Comprehensive Review of Literature

Educational system can be classified into two types:[32]

- a. Traditional classroom environment which provides a real face-to-face contact between the students and teacher.
- b. Online tutoring system which provides a virtual classroom environment with a remote student –teacher communication.[31]

The prediction and analysis of students' performance can be applicable to online and offline environments. Various research works has been discussed in this paper since 2000 to till date.

- i. Prediction of student's performance in a Traditional Classroom environment.
- ii. Early Prediction of student Drop out and Retention in a Traditional Classroom environment.
- iii. Prediction of student's performance in Online tutoring system.

### Prediction of student's performance in a Traditional Classroom environment.

Yannick Meier Xu et al [1], proposed an algorithm that predicts the final grade of each student in the course. It issues a timely and personalized prediction result to the instructor, followed by the proper intervention of the instructor to improve individual performance of every student.

Raheela Asif et al [2], Investigated the evolution of academic performance of a student during the lifetime of the course using x-means clustering algorithm.

Parneet Kaura et al [4], studied the prediction of slow learners among the students in a high school using a predictive data mining model using classification based algorithm.

Tripti Mishra et al [9], attempted in including emotional data about a student as a key factor of predicts the student performance along with the social and academic data is concerned, using J48 and Random tree algorithm they developed a prediction model.

Pallavi Kulkarni et al [10] , applied incremented learning technique in building the predicting model and compared four different classifiers: Naïve bayes , KSTAR, IBK and nearest neighbour.

Suchita Borkar et al [11], build a prediction model based on the association rule data mining technique to predict the students' performance in final exam.

Mohammed M. Abu Tair et al [15], proposed a method to improve graduate students performance and ways to overcome the problem of low grade by using various data mining techniques .They employed association rule and naïve Bayesian to predict the grades of the students, K-means clustering to group the students, density based approach and distant based approach to perform outlier detection.

Dorina Kabakchieva [16], build a forecasting model to predict the students' performance based on university score and number of failures using classification algorithm.

Md. Hedayetul Islam Shovon et al [17], included Internal scores obtained by the student in Quiz and assignment and attendance to predict the final grade of the students by applying decision tree and k-means clustering algorithm.

Zlatko J. Kovačić et al[19], introduced socio-demographic data collected during enrolment of a student as a key factor to predict the students success using Classification and Regression Technique (CART)

Nguyen Thai-Nghe et al [21], build a recommender system for predicting student performance. They proposed a mapping of educational data to user in recommendation system using matrix factorization method.

V. O. Oladokun et al[23], applied a multi-layer perceptron topology on student enrolment data to predict the performance of a student who is currently enrolled for a course in an university.

C. Chen et al[25] , proposed a prediction method build from key formative assessment rule to predict the performance of the learner based on the compilation of learning portfolios.

J. Frey berger et al [30], employed logistic regression and robust ridge regression algorithm to predict when a student will get right answer to a question and to find the students success rate.

### **Early Prediction of student Drop out and Retention in a Traditional Classroom environment.**

Erman Yukselturk et al [8], examined the prediction of students dropout from a course through various classifiers such as K-nearest neighbour, Decision tree, Naïve bayes, Neural network. Feature selection algorithm is used to find the factor which influences the student dropout.

C.Marquez-vera et al[12], predicted the high school students failure and dropout from school by white box classification method and decision tree.

Saurabh Pal[13],build a predictive model to forecast the student dropout in an engineering college using Machine learning algorithm such as ID3, C4.5 and CART.

Jaroslav Bayer et al[14],forecasted the dropouts and failures in a high school based on the social behaviour data of a student using machine learning methods.

Ying Zhang et al[22], proposed a method to retain students by early prediction of student at risk and to tailor the risk factors by using naïve Bayes algorithm.

Camilo Ernesto Lopez Guarin et al [3], proposed a model to predict the attrition in the academic performance at enrolment of a certain course using naïve bayes and decision tree classifiers based on the academic and non-academic data about the student.

### **Prediction of student's performance in online tutoring system.**

Nick.Z.Zacharis[5], developed a model to predict the students outcome in a MOODLE LMS – Supported blended learning course and found the significant correlation between the different online activities and grade of a particular course using stepwise multivariate regression.

Xing Wanli et al[6], projected a methodology which combines the perspectives from learning analytics, Educational Data Mining theory to predict students' performance in a Computer Supported Collaborative Learning (CSCL) environment using Genetic Programming algorithm.

Suhang Jiang et al [7], proposed a predictive model to forecast the students final performance in a course conducted in a MOOC, based on the assignment performance of the students in the first week of the course along with the social interactions made by the student within MOOC using a logistic regression.

S.Kotsiantis [20], prophesied the students' performance in distance education based on the continuous data generated in the log file using a combinational incremental ensemble technique.

C. Romero et al [24], conducted a comparative study on different data mining algorithm to predict the performance of a student as a final grade on a MOODLE Course.

W.H'am'al'ainen et al [27], carried a comparative study on various machine learning methods to predict the success rate of an online tutoring system.

E. Ayers et al [28], build a prediction methodology to determine the score of a student in the final exam based on the data obtained from the previous performance of the same student in an online tutoring system using Bayesian network.

Etchellset al [29], applied decision tree algorithm to predict the student performance and to provide lessons on time based on the prediction in an online tutoring system.

**Conclusion:** This paper is a survey of prediction and analysis of student performance; it surveys the most relevant work in this area till date. Many research works has been carried out in this area, still this area has its charm to attract so many researchers as it addresses the key issue of the quality of student performance and retention. Online tutoring system is gaining importance many unexplored peaks in the student performance and retention can be explored in the

future. The real world implementation of EDM to the educational institutions will open up a way to increase powerful young performance with high potential to develop the nation.

## **References:**

1. Yannick Meier Xu, Jie Xu, OnurAtan, Mihaela van der schaar,"Predicting grades", IEEE Transactions on signal processing, vol 64, no4, February 15,2016.
2. Raheela Asif, Agathe Merceron, Mahmood khan Pathan, "Investigating Performance of students: a longitudinal study", ACM, March 2015.
3. Camilo Ernesto Lopez Guarin, Elizabeth leonguzman, Fabio.A.Gonzalez, "A model to predict low academic performance at a specific enrolment using data mining", IEEE Revista, Vol.10.No.3, August 2015.
4. ParneetKaura, ManpreetSinghb, Gurpreet Singh Josanc, "Classification and prediction based data mining algorithms to predict slow learners in education sector "3rd International Conference on Recent Trends in Computing 2015(ICRTC-2015)
5. Nick.Z.Zacharis, "A Multivariate approach to predicting student outcomes in web-enabled blended learning courses, Internet and higher education 27(2015),44-53.
6. Xing Wanli, GuoRui, Petakovic Eva, Goggins Sean, "Participation-based student final performance prediction model through interpretable genetic programming: Integrating analytics, educational data mining and theory", Computers in Human behaviour 47(2015),168-181.
7. Suhang Jiang, Mark Warschauer School of Adrienne E.Williams, Diane O'Dowd, Katerina Schenke, "Predicting MOOC Performance with Week 1 Behavior", Proceedings of the 7<sup>th</sup> international conference on educational data mining,2014.
8. Erman Yukselturk, Serhat Ozekes," Predicting dropout student: an application of data mining methods in an online education program", European Journal of Open, Distance and e-Learning Vol. 17 / No. 1 – 2014.
9. Tripti Mishra, Dr.Dharminder Kumar, Dr. Sangeeta Gupta, "Mining Students Data for Performance Prediction", Fourth International conference on advanced computing and communication technologies,2014.
10. Pallavi Kulkarni , Roshani Ade ,"Prediction of based on Incremental Learning ", International Journal of Computer Applications (0975 8887) Volume 99 No.14, August 2014
11. SuchitaBorkar, K. Rajeswari, "Predicting Students Academic Performance Using Education Data Mining", International Journal of Computer Science and Mobile Computing, Vol. 2, Issue. 7, July 2013, pg.273 – 279.

12. C.Marquez-vera, C.Romero , S.Ventura, “Predicting School Failure Using Data Mining”,IEEE Journal of Latin-American Learning Technologies, Vol.8,No.1, February 2013.
13. Dr. Saurabh Pal, ”Mining Educational Data Using Classification to Decrease Dropout Rate of Students”, International Journal Of Multidisciplinary Sciences And Engineering, VOL. 3, NO. 5, MAY 2012,[ISSN: 2045-7057]
14. Jaroslav Bayer, Hana Bydžovská, Jan Gěryk, Tomáš Obšiváček , “Predicting drop-out from social behaviour of students”, Proceedings of the 5th International Conference on Educational Data Mining,
15. Mohammed M. Abu Tair, Alaa M. El-Halees, “Mining Educational Data to Improve Students’ Performance: A Case Study “, International Journal of Information and Communication Technology Research, Volume 2 No. 2, February 2012.
16. Dorina Kabakchieva, “ Student Performance Prediction by using Data Mining Classification Algorithms”, International Journal of computer Science and Management Research, vol 1, Issue 4, November 2012
17. Md. Hedayetul Islam Shovon , MahfuzaHaque, “An approach of improving students’ academic Performance by using K-means clustering algorithm and Decision tree”, International Journal of Advanced Computer Science and Applications, Vol.3, No. 8, 2012
18. Umesh Kumar Pandey S. Pal,” Data Mining : A prediction of performer or underperformer using classification“, International Journal of Computer Science and Information Technologies, Vol. 2 (2) , 2011, 686-690
19. Zlatko J. Kovačić, “Early Prediction of Student Success: Mining Students Enrolment Data” , Proceedings of Informing Science & IT Education Conference (InSITE) 2010
20. S.Kotsiantis, K. Patriarcheas, M.Xenos, “A combinational incremental ensemble of classifiers as a technique for predicting students performance in distance education”, Knowledge-Based systems 23 (2010),529-535
21. Nguyen Thai-Nghe\*, Lucas Drumond\*, ArtusKrohn-Grimberghe\*, Lars Schmidt-Thieme, “Recommender System for Predicting Student Performance”, Procedia Computer Science 1 (2010)
22. Ying Zhang, SamiaOussena, Tony Clark, Hyeonsook Kim, “USE DATA MINING TO IMPROVE STUDENT RETENTION IN HIGHER EDUCATION – A CASE STUDY “,2008.
23. V. O. Oladokun, A. T. Adebajo, and O. E. Charles-Owaba, “Predicting student’s academic performance using artificial neural network: A case study of an engineering course,” Proc. Pacific J. Sci. Technol., vol. 9, no. 1, pp. 72–79, 2008.

24. C. Romero, S. ventura, C. herv'as, and P. Gonzales, "Data mining algorithms to classify students," in Proc. Int. Conf. Educ. Data Mining, Montreal, Canada, 2008, pp. 8–17.
25. C. Chen, M. Chen, and Y. Li, "Mining key formative assessment rules based on learner portfolios for web-based learning systems," in Proc. IEEE Int. Conf. Adv. Learn. Technol., Niigata, Japan, 2007, pp. 1–5.
26. C. Romero, S. Ventura, "Educational data mining: A survey from 1995 to 2005", *Expert system with applications* 33 (2007), 135-146.
27. W.H'am'al'ainen and M.Vinni, "Comparison of machine learning methods for intelligent tutoring systems," in Proc. Int. Conf. Intell. Tutoring Syst., Taiwan, 2006, pp. 525–534.
28. E. Ayers and B. W. Junker, "Do skills combine additively to predict task difficulty in eighth grade mathematics?" in Proc. AAAI Workshop Educ. Data Mining, Menlo Park, CA: AAAI Press, 2006, pp. 14–20.
29. T. A. Etchells, A. Nebot, A. Vellido, P. J. G. Lisboa, and F. Mugica, "Learning what is important: Feature selection and rule extraction in a virtual course," in Proc. Eur. Symp. Artif. Neural Netw., Brussels, Belgium, 2006, pp. 401–406.
30. J. Freyberger, N. T. Heffernan, and C. Ruiz, "Using association rules to guide a search for best fitting transfer models of student learning," in Proc. Workshop Analyzing Student-Tutor Interaction Logs Improve Educ. Outcomes, Alagoas, Brazil, 2004, pp. 1–10.
31. Sheard, J. Ceddia, J. Hurst, J., & Tuovinen, J., "Inferring student behaviour from website interaction: a usage analysis", *Journal of education and information technologies*, 8(3), 2003, 245-266.
32. Johnson, S., Arago, S., Shaik, N., & Palma-Rivas, N., "Comparative analysis of learner satisfaction and learning outcome in online and face-to-face learning environment. *Journal of interactive learning research*", 2000, 29-49.

**Corresponding Author:**

**T.Archana\***,

**Email:** [archana.t@vit.ac.in](mailto:archana.t@vit.ac.in)