

Survey Report on: EDM for Prediction of Academic Trends & Patterns

Trupti S. Diwan¹, Bharati Dixit²

ME Student, MIT College of Engineering, Pune, Maharashtra, India

Faculty, MIT College of Engineering, Pune, Maharashtra, India

Abstract: *Predicting student failure at school has changed into a difficult challenge due to both the large number of factors that can affect the reduced performance of students and the imbalanced nature of these kinds of datasets. This paper surveys the two elements needed to make prediction on Students' Academic Performances which are parameters and methods. This paper also proposes a framework for predicting the performance of engineering students. Genetic programming can be used to predict student failure/success. Ranking algorithm is used to rank students according to their credit points. The framework can be used as a basis for the system implementation & prediction of Students' Academic Performance in Higher Learning Institute.*

Keywords: Classification, Educational data mining, Student failure, Grammar-based genetic programming

1. Introduction

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, and the settings which they learn in. EDM has contributed to theories of learning investigated by researchers in educational psychology and the learning sciences.

Predicting student failure at school is really a key problem and it is becoming essential for educational experts to better realize why therefore several youth crash to complete their school studies. But, this is a difficult problem to eliminate as a result of wide range of risk factors or students characteristic that could influence school failure, such as age, ethnic, cultural, household, or instructional history, socioeconomic status position, emotional profile, and academic development [1]. Actually, this problem is generally known as "one thousand factors problem" [3].

Before this [2], study has been done on pinpointing the main factors that affect the lower efficiency of students such as failure and dropping out at various educational levels like primary, secondary, and higher etc. Tinto model is just like this dilemma [8]. That model suggest that the student's cultural and academic integration to the instructional institution may be the key determinant of completion and identifies some essential influences on integration like the student's history, personal information of student like family background, previous schooling details, previous academic efficiency, and interactions between student and school staff.

Lately, there has been an agreement that identification and counteractive action of students disappointment at school and early intervention bode well than remediation [6]. In this line, a successful method to discover student failure at school is the use of Data Mining (DM) practices which have been properly put on other study places such as ecommerce, where their use has become very popular. DM is area of the process of Knowledge Discovery and Data Mining (KDD) and is recognized to function as non-trivial removal of previously not known and probably helpful, legitimate, and

comprehensible data from a sizable level of data [4]. DM is named Educational Data Mining (EDM) when applied to an educational situation [7] and is worried with establishing practices to examine distinctive kinds of information in educational settings and using these practices to better understand students and the options in that they learn. Good instances of how to use EDM methods can be found to produce types that predict dropping out and student failure specifically [5].

In this paper, white box algorithms are used because the types obtained have the form of IF-THEN classification or prediction principles that report the causes for classification and crucial dependence relations between data. This kind of concept includes two parts. The rule antecedent (the IF part) contains a conjunction of m situation on values of interpreter properties, although the concept consequent (the THEN part) contains a prediction about the value of a goal attribute or class.

Within any situation, the final objective is always to manage to identify the reasons that cause school failure and to supply identification of students who display those characteristics (i.e. the facets which have the maximum influence on school failure) in order to offer them proper and customized assist in an effort to reduce failure in school. A new algorithm is proposed for predicting school failure as a classification task. Many studies have been done to evaluate proposed algorithm with other classification algorithms. Different DM methods have been applied to take to improve the precision of the classification model and to handle the problems of large dimensionality and imbalanced data.

2. Literature Review

This report [9] is overview of their state of the art with respect to EDM and surveys the absolute most relevant perform of this type to date. Each study has been categorized, not only by the kind of data and DM methods applied, but also and more importantly, by the type of instructional job that they resolve.

EDM has been presented as an upcoming study area linked to several well-established areas of study including, intelligent tutoring techniques, adaptive hypermedia, e-learning, data mining, web mining etc. We've observed how fast EDM is growing as reflected in the raising quantity of contributions printed annually in Global Conventions and Journals, and the quantity of specific instruments especially created for applying data mining algorithms in instructional data/environments. So, it could be stated that EDM is currently approaching their adolescence, that's, it is no further in their early days but is not even a mature area.

Actually, some intriguing potential lines have been identified but also for it to become a more mature area, it can be required for scientists to produce more unified and collaborative studies instead of the current array of multiple personal proposals and lines. Hence, the full integration of data mining in the instructional environment will become a reality, and completely operative implementations (both industrial and free) could be produced accessible not only for scientists and designers but also for additional users.

From [10], based on the outcomes of Apriori algorithm and E means clustering in Weka, on the academic history file, it can be deduce that the crucial responsibility that data mining can play in the subject of training and teaching. It would be very difficult to physically proceed through the huge group of academic documents to recognize the scholar students and the pattern in which they learn. Instead, If we take advantage of data mining techniques on the wide range of academic history, we are able to easily categorize the students, recognize hidden patterns about their learning variations, discover undesirable behavior of student and accomplish student profiling. In That manner, data mining can truly be a significant tool and part of highly sophisticated educational techniques.

Predicting student placement class manually by teachers is an arduous task. To resolve this problem, we are able to use data mining to help predict the classification [11]. Before apply, we need to know the best algorithm that suitable with this data. Since of that, Pratiwi, O.N. did this experiment. This report proposes six formulas that will use to categorize the student's data. From the first test, that uses all characteristics, show that the very best classification methods are J48, and with the percentage accuracy of equally test are 79.60%. This test shows that the data balancing can also be very useful for improving accuracy. Thus, it is essential process of locating the very best characteristics in classification process. Additionally the algorithm will be in to the device so that it could be used to help teacher predicting the placement class of students. The machine can also support students to give early information about their major.

Nowadays academic success of students of any skilled Institution has become the significant concern for the management. An early prediction of students at risk of poor performance assists the management to take reasonable action to improve their performance through extra instruction and counseling. This report [12] mainly focused on various qualities that inspired student semester performance. Effect of mental quotient parameters on

placement has been established. Random tree gave larger reliability of prediction than J48.

The performance of equally the algorithm is acceptable; nevertheless, larger over all reliability (94.418%) was obtained by Random Tree implementation as compared to J48 with 88.372% accuracy. Also the True Good Rate, Accuracy and Recall measures of Random tree are higher than J48 and consistent with the corresponding accuracy.

In [13] classification as a technique to predict an appropriate career for a student is suggested by Elakia and Gayathri. A student must be disciplined and shouldn't be vulnerable to abuse which can influence their future (career). Hence a prediction on the opportunity of the student getting violent in future is performed which can help the organization to provide counseling to the appropriate student to control the violence in the initial phases itself. Different Classifiers have been considered because of their precision and efficiency and acceptable classifier was used.

Educational Data mining tools are typically designed more for power and adaptability than for simplicity. The majority of the existing data mining tools are too complex to use for teachers and their features go well past the extent of what a teacher may need to do. Current tools for mining data from a particular course may be helpful only to its developers. There are no universal tools or re-utilizing tools or methods that can be applied to any educational system. Conventional data mining algorithms are not tuned to consider the educational context.

3. Propose Work

Figure1 shows Architectural view of EDM for prediction of academic trends and patterns. In this, first we will gather the dataset from the educational system. After that we will do preprocessing step on the dataset. In preprocessing step there are three sub steps such as data cleaning, selection of attributes, and data balancing.

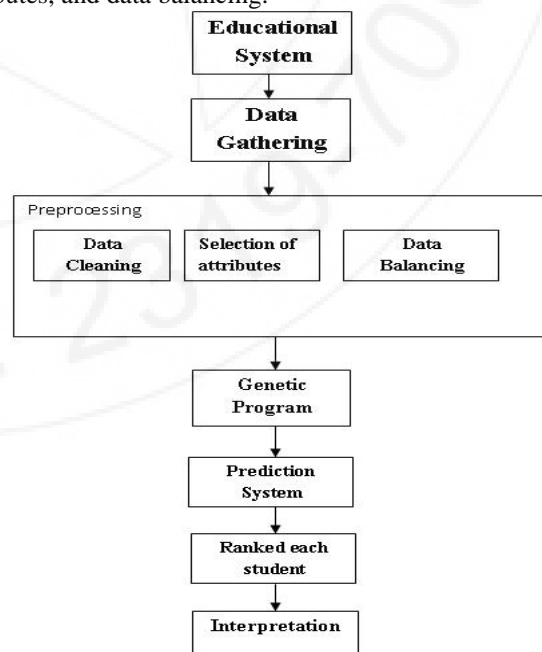


Figure 1: Architecture Diagram

After that the processing of mining will be done. In the mining process we will use genetic programming; the algorithm is used for classification/ prediction based on grammar. We will also try to predict the student failure as soon as possible by using latest algorithm C4, C5. Because the earlier the better, in order to detect students at risk in time before it is too late.

4. Conclusion

As observed, this paper surveys the prediction in data mining. The main objective of Educational Institution is to provide the best quality education to its student and to improve their behavior. Prediction of student academic performance is helps to teachers to predict about student success and failure in examination.

Predicting student failure at school can be quite a hard task not just because it is really a multifactor issue (where there are a lot of particular, household, cultural, and economic facets that may be influential) but in addition as the accessible data are generally imbalanced (most of the students go to another course). To resolve these issues, we demonstrate the usage of various DM calculations and techniques for predicting student failure. We've carried out many tests using actual data about 1000 students from some Indian university. We've used various classification techniques and genetic programming for predicting the academic position or ultimate student performance by the end of the course.

References

- [1] Aloise-Young PA, Chavez EL (2002) Not all school dropouts are the same: Ethnic differences in the relation between Reason for leaving school and adolescent substance use. *Psychol Sch* 39(5):539-547
- [2] Araque F, Roldan C, Salguero A (2009) Factors influencing university dropout rates. *Comput Educ* 53:563-574
- [3] Hernández MM (2002) Causas del Fracaso Escolar. In: XIII Congreso de la Sociedad Española de Medicina
- [4] Klösgen W, Zytkow JM (2002) Handbook of data mining and knowledge discovery. Oxford University Press, London
- [5] Kotsiantis S, Patriarcheas K, Xenos M(2010) A combinational incremental ensemble of classifiers as a technique for predictin students performance in distance education. *Knowl-Based Syst* 23(6):529-535
- [6] Slavin RE, Karweit NL, Wasik BA(1994) Preventing early school failure. Allyn & Bacon, Needham Heights
- [7] Romero C, Ventura S (2010) Educational data mining: a review of the state of the art. *IEEE Trans Sys Man Cybern* 6:601-618
- [8] Tinto V (1987) Leaving college: rethinking the causes and curse of students attrition. University of Chicago Press, Chicago
- [9] Cristóbal Romero, Member, IEEE, Sebastián Ventura, Senior Member, IEEE, "Educational Data Mining: A Review of the State-of-the-Art", *IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews*, Vol. Xx, No. X, 200x.

- [10] Suhem Parack, Zain Zahid, Fatima Merchant , "Application of Data Mining in Educational Databases for Predicting Academic Trends and Patterns", *Technology Enhanced Education* , 2012 IEEE International Conference on 3-5 Jan. 2012
- [11] Oktariani Nurul Pratiwi, "Predicting Student Placement Class using Data Mining", *Teaching, Assessment and Learning for Engineering (TALE)*, 2013 IEEE International Conference on 26-29 Aug. 2013
- [12] Mishra, T., Kumar, D. gupta,S.," mining Students' Data for Prediction Performance", *Advanced Computing & Communication Technologies (ACCT)*, 2014 Fourth International Conference on 8-9 Feb. 2014
- [13] Elakia, Gayathri, Aarthi, Naren J," Application of Data Mining in Educational Database for Predicting Behavioural Patterns of the Students ", *Elakia et al, / (IJCSIT) International Journal of Computer Science and Information Technologies*, Vol. 5 (3) , 2014, 4649-4652