International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

Student's Performance Prediction Using FP-Tree Data Mining Techniques

Eliakim Ombati Akama

Student of MSc (CS), Jrnrvu University, Udaipur, India

Abstract: ***Student performance is an essential part in higher learning institutions. But it has remained a challenge in predicting students' performance in India due to the huge volume of student data in educational databases. There are three reasons as to why it still remains a challenge. First, lack of investigations on the factors affecting students' achievements in their respective courses. Secondly, the study on the existing prediction methods is not sufficient to identify the most suitable methods for the performance of students in India. Data mining techniques can be used to find hidden patterns from these huge educational databases, which can be further used to predict the student performances. After getting the predicted result, the performance of the student can be improved by engaging with desirable assistance.***

Keywords: Data mining, Data Mining Techniques, FP-Tree

1. Introduction

Student performance is an integral part in higher learning institutions. This is because one of the criteria for high quality university is based on its excellent record of academic achievement. In a corporate world, recruitment, employments measures are based on the grading system. In academics, grading is the criteria for measuring the potentiality of students. Institutions use the final year grades to evaluate the performance of a student.

Most of the high learning institutions in India evaluate their performance of their students by their final grades. Final grades are based on course structure, assessment mark, final exam score and also extra-curricular activities. This evaluation is very important to maintain student performance and the effectiveness of learning process.

Although most teachers, trainers, and even professors, they don't want to be held responsible for learner motivation [1], it's important to identifying and analyzing the area of need for each student so that their performance can be improved by engaging with desirable assistance. Educational data mining is one of the criteria that can be used to analyze the performance of students and with data mining techniques, better solutions can be achieved to improve the overall performance not only students but also the institutional excellence.

2. Motivations

Why predict student performance?
To identify weak students early and intervene:
To enhance learning:especially in online assessment to complement study and learning experiences.
To reduce drop-out rates: Dropout rates are negative to our economy on income
Student recruitment: this enables the institution to do replacement of students upon completion of their course study. It helps parents to understand the best field for their children.

3. Data Mining

Data mining is the process of discovering interesting knowledge from a large amount of data stored in databases, data warehouse or other information repositories. It is a process that analyses huge amount of data for useful information to reveal trends and pattern which otherwise may be undetected. It has attracted a great deal of attention in the information industry and society as a whole in recent years due to the wide availability of huge imminent need of turning such data into useful information and knowledge. Data mining can also be termed as knowledge discovery database (KDD).The main aim of data mining is to extract information from dataset and convert it to a structure that is easy to understand.

Data mining is an iterative process consisting of the following steps:

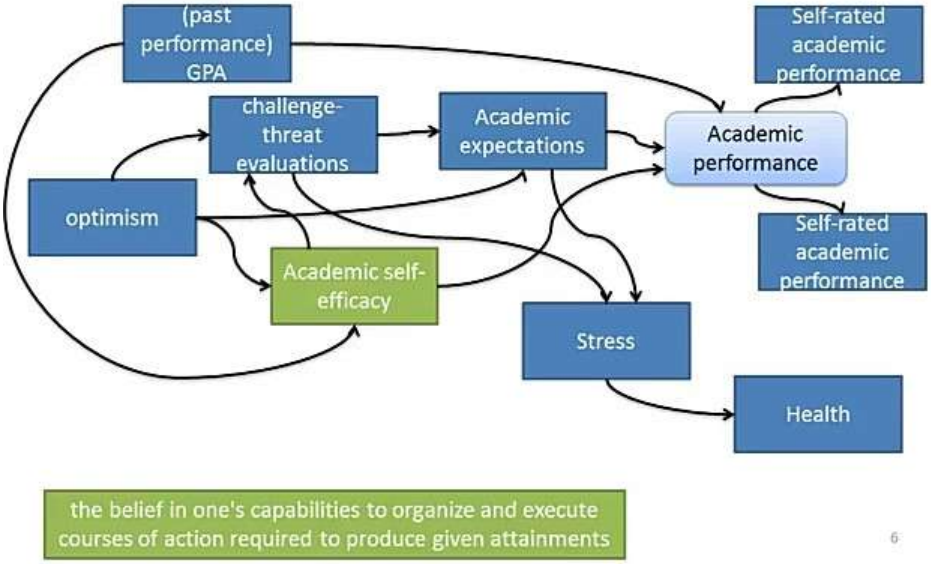
- Data cleaning: removing of noise and inconsistence data
- Data integration: Data from multiple data sources is combined
- Data selection: Task essential data is retrieved
- Data transformation: respective data is transformed and summarized
- Data mining: An essential process where intelligent methods are applied to extract data patterns
- Pattern evaluation: Identifying interesting patterns
- Knowledge presentation: Visualization and knowledge representation techniques are used to present the mined knowledge to the user

4. Literature

There are lots of definitions on students' performance based on the previous literature. Usamah et al. (2013) stated that students' performance can be obtained by measuring the learning assessment and co-curriculum [2].In 1960s they used structured equation model, they were basically getting a pair of correlation variables and then the structured equation model could tell you what are the variable that are related to each other, they used these methods because they didn't have the they kind of data or platform that we have nowadays where we can correct lots of data.

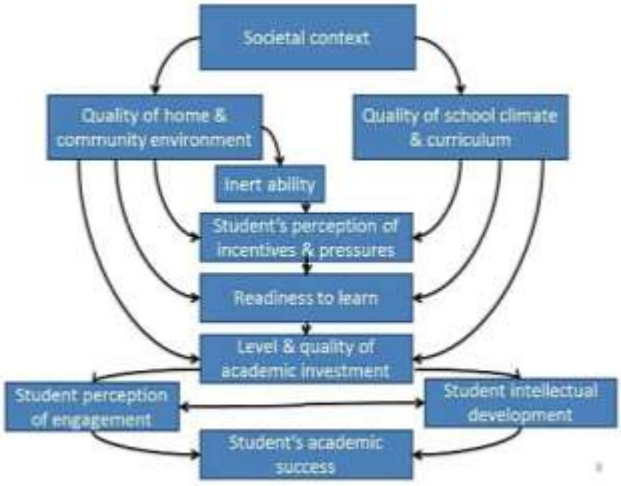
4.1 Chemer, Hu, and Garcia's model

When Chemer, Hu, and Garcia's tried to understand the problem they interviewed the students and then they did some hypothesis from that survey like academic self-efficacy and attitude to learning, that is a belief in one's capability to organize and execute course of action required to produce a given attainment or academic performance of achievement .



4.2 Rossi and Montgomery's model

This model focus much more on the environmental condition for student's home and the environmental condition within their school.



4.3 Findings

Most poor educated young people do not become lifelong welfare recipients or career criminals. Too many of them labor hours at dead end jobs for wages that fail to raise their families out of poverty [9]: they enroll in store-front vocational colleges that immerse them in depth and fail to prepare them for promised career opportunities; they struggle to read the employment application or the letter from their child's teacher that demands more literacy skills than they possess: they die at earlier ages from illnesses and diseases that related to poverty [3].
Although it is difficult to prescribe a "one size fits all" approach to motivate students, researchers suggested that some general patterns do appear to hold true for a wide range of students [8] [1].

5. Methodology

There exist different data mining techniques like prediction, sequential patterns, association, classification, and clustering .The objective of the system is to predict a student's status as weaker or non-weaker student and placement status. To predict this status input academic and placement related details of previous year students. By applying data mining techniques on previous year data, a structure predicting the status will be obtained. This structure will be applicable to predict status of the students in current year before they appear for the actual examination. If students are identified as weaker, extra academic inputs can be provided by the college. This will help in improving overall performance of college or university.

Based on parameters like aggregate, 12th marks, and certification, another module for placement prediction will be designed. This module will evaluate placement status of current year considering previous year placements. Performance of student and company criteria will be two important factors of placement prediction.

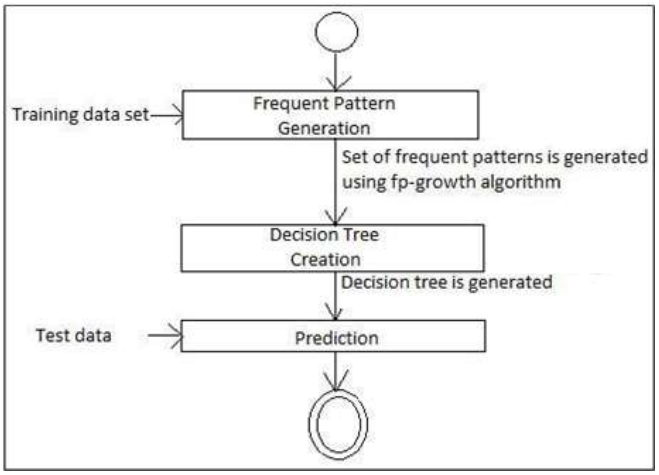


Figure 1.1: Methodology Flow

The details required are collected from the ERP database and placement office of respective college. This input includes personal details like name, date of birth, address etc. and academic details like mid-semester marks, end-semester marks, attendance and placement information etc. The values obtained for marks and attendances are numerical. In order to ease processing these numerical values are converted into categorical data as grades A, B, C, D, and E.

The proposed model uses techniques of association and decision tree. In association, a pattern is discovered based on a relationship between items in the same transaction. A decision tree is a tool that support decisions based on a tree-like graph or model of decisions. FP-growth algorithm which is an association technique is used. It gives set of frequent patterns from the input dataset

The available dataset is pre-processed for each subject according to its prerequisite subjects. FP growth algorithm is applied on these datasets separately. Frequent patterns leading a student to pass in a particular subject are obtained when algorithm is applied on dataset. A decision tree is generated by frequent pattern obtained from dataset. The leaf nodes of the decision tree are performance status of student as weaker or non-weaker. This tree is finally traversed to identify the performance of current year student. Student's placement status will be determined by using clustering algorithm.

6. Conclusion

Predicting student performance is mostly useful to help the educator and leaners in improving their learning and teaching process. Predicting student academic performance, predicting educational dropout student in near future, predicting institute placement and admission in a new academic year is most useful for educators, management and even education policy maker. This paper has reviewed previous studies on predicting students' performance with various analytical methods more so we have emphasized on FP Tree and placement prediction. It will help the education system to monitor the students' performance in a systematic way and help reduce the number of suicides caused by parent's pressure to students due to poor performance.

7. Recommendation

From the research done, Data mining measures should be applied in every institution so that they can foster the education performance. It should not be an option but a formality for all institutions to stimulate or lay out a data mining technique that predicts and sets roadmap measures that can be taken to ensure quality learning.

References

[1] J.M.Keller. Motivational Design for Learning and Performance 2010, pp. 1 DIO 10.1007/978-1-4419-1250-3_1.
[2] U. bin Mat, N. Buniyamin, P. M. Arsad, R. Kassim, An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention, in: Engineering Education (ICEED), 2013 IEEE 5th Conference on, IEEE, 2013, pp. 126–130.
[3] R Rossi; A Montgomery, Educational Reforms and Students at Risk; A Review of the current State of the Art 1994 URL:https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=152206
[4] John M Keller, Motivational Design for Learning and Performance: The ARCS model Approach
[5] Deimann, Keller, A review of challenges to maintain persistence in online learning 2006
[6] M. Mayilvaganan, D. Kalpanadevi, Comparison of classification techniques for predicting the performance of students' academic environment, in: Communication and Network Technologies (ICCNT), 2014 International Conference on, IEEE, 2014, pp. 113–118.
[7] M. M. A. Tair, A. M. El-Halees, Mining educational data to improve students' performance: a case study, International Journal of Information 2 (2).
[8] Anderman et al 1999 E.M.Anderman, M.L. Maehr, C.Midgley, Declining motivation after the transition to middle school: Schools can make a difference A journal of research and development in education, 32(1999),pp. 131-147
[9] Montgomery, Alesia; andothers, Educational Reforms and students at risk: A review of the current State of the Art Oct 1993