Education Data Mining – A Prediction Model and Its Affecting Parameters for Prediction in Higher Education

Jaimin N. Undavia,
Assistant Professor,
CMPICA, CHARUSAT,
Changa, Gujarat.

Atul Patel,
In charge Principal,
CMPICA, CHARUSAT,
Changa, Gujarat
.
Nikhil P. Shah,
Assistant Professor,
Dharmsinh Desai University,
Nadiad, Gujarat.

Abstract

The data is the most vital part of any business or organization. Apart from its importance, its amount is increasing very rapidly. So, these huge amounts of data can be used very effectively with some mining techniques for predicting the performance and best suitable higher education as well. The higher education is tending to prepare professionals with high degree of knowledge, moral value and expertise in their particular discipline. We may achieve the highest level of quality in higher education is by discovering knowledge of prediction regarding the academic history of particular student. The academic performance of the student may be influenced by many factors, so it becomes necessary to develop predictive data mining model for students. In this paper we have analyzed students of MCA 2012 batch of Charutar University of Science & Technology. At the end we will conclude the result against our parameters.

Keywords: Data Mining, Education Data Mining, Prediction Model.

1. Introduction

The most prominent feature of Data Mining is the ability to discover hidden patterns and relationships helpful in decision making. It has been successfully used in different areas including the educational environment. Educational data mining is an interesting research area which extracts useful, previously unknown patterns from educational database for better understanding, improved educational performance and assessment of the student learning process [4]. It is concerned with developing methods for exploring the unique types of data that come from educational environment which include students' results repository and student's academic performance so far.

The term Data Mining comes with some handy techniques for prediction, classification, etc.

Basically the mining process can be categorized into 2 groups depending its capabilities.

   A. Automated predictions of trends and behavior.

B. Automated prediction of previously unknown patterns.

This research falls under the second category in which we have tried to discover new pattern from historical data.

In this paper, we have analyzed some pass out students' records and then we can predict the new record and the student will be assisted to choose best discipline or course for higher study. As the student performance is influenced by many parameters, here first we will identify parameters.

2. Background

Most of the higher education institutes are working for their quality educations. However they are facing many problems in achieving their quality objectives.

The main reason for that is the lacking in knowledge. This lacking may be in counseling, planning, registration, and evaluation, marketing and proper guidance in selection of higher study.

The main objective of EDM is to discover the hidden patterns, association and relations to discover the hidden knowledge through different data mining techniques. The knowledge discovered by data mining techniques would enable the higher learning institutions in making better decisions, having more advanced planning in directing students, predicting individual behaviors with higher accuracy, and enabling the institution to allocate resources and staff more effectively. It results in improving the effectiveness and efficiency of the educational processes.

3. Related Work

Data Mining is also referred as knowledge discovery. In this research we are trying to discover the knowledge from the education field to discover better understanding of the affecting parameters in higher education of student [6]. The mining used in educational environments is called education data mining.

Han and Kamber [7] describes data mining software that allow the users to analyze data from different dimensions, categorize it and summarize the relationships which are identified during the mining process.

Pandey and Pal [8] conducted study on the student performance based by selecting 600 students from different colleges of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayes Classification on category, language and background qualification, it was found that whether new comer student will performer or not.

Pandey and Pal [9] conducted study on the student performance based by selecting 60 students from a degree college of Dr. R. M. L. Awadh University, Faizabad, India. They have used association rule mining to find out interestingness of student for class teaching language.

Tissera et al. [11] presented a real-world experiment conducted in an ICT educational institute in Sri Lanka, by analyzing students' performance. They applied a series of data mining task to find relationships between subjects in the undergraduate syllabi. They used

association rules to identify possible related two subjects' combination in the syllabi, and apply correlation coefficient to determine the strength of the relationships of subject combinations identified by association rules. As a result, the knowledge discovered can be used for improving the quality of the educational programs.

Ramaswami and Bhaskaran [12] have constructed a predictive model called CHAID with 7-class response variable by using highly influencing predictive variables obtained through feature selection so as to evaluate the academic achievement of students at higher secondary schools in India.

4. Data Collection

Here I have collected the data from the Charusat University, Changa MCA 2012 students. These data are analyzed using classification method to predict the student's performance

5. Parameters for Classification

In this research paper, I have used the students who have completed their MCA degree. Our key steps of research at this point are as under.
- Deciding the parameter list.
- value of each parameter for each students.
- Analyze final result of each student.
- Check against the proposed classification model.
- Verify the actual result and the result generated by the proposed model.

Affecting parameters are extracted and list out here for reference.

Parameter Name	Description	Value
SM	SSC Percentage	A - <90% B- 70% - 89% C - 60% - 69% D – 50% - 59% E – 40% - 49% F < 40%
ST	No of attempts in SSC	1 or 2
HM	HSC Percentage	A - <90% B- 70% - 89% C - 60% - 69% D – 50% - 59% E – 40% - 49% F < 40%
HT	No of attempts in HSC	1 or More
HS	Stream in HSC	SC – Science CM – Commerce AT - Arts
GR	Graduation	BA, BCom, BSc, BBA, BCA, BE
GP	Graduation Percentage	D – Distinction F – First Class S – Second Class T – Third Class
IM	Medium of instruction up to school	Guj, Eng.
GM	Medium of instruction in Graduation	Guj, Eng.
GR	Rank in GCET	Best of 2 GCET

SM - Students grade in SSC of state board education. Students who are in state board appear for five subjects each carry 100 marks. Grades are assigned to all students using following mapping.
A - <90%, B- 70% - 89%, C - 60% - 69%, D – 50% - 59%, E – 40% - 49%, F - < 40%.

ST – Total attempts in SSC examination. More than 2 attempted is not considered.

HM - Students grade in HSC of state board education. Grades are calculated as per marks in main subjects depending on stream. Grades are assigned to all students using following mapping.
A - <90%, B- 70% - 89%, C - 60% - 69%, D – 50% - 59%, E – 40% - 49%, F - < 40%.

HT – Total attempts in HSC examination. More than 1 attempted is not considered.

HS – As per GHSEB, student can opt any of three available streams. Streams are assigned to all students using following mapping.
SC – Science, CM – Commerce, AT – Arts.

GR – This variable will be assigned the value according to the graduation degree of student. Degree will be assigned to all students using following code. BA, BCom, BSc, BBA, BCA, BE.

GP - Grade Obtained in Graduation. Marks/Grade obtained in graduation. It is also split into four class values. Class is assigned to all students using following mapping.
D – Distinction, F – First Class, S – Second Class, T – Third Class

IM – This parameter indicates the medium of instruction during school. It will be assigned to each student using following mapping.
Gujarati - Guj, English – Eng.

GM – This parameter indicates the medium of instruction of graduation. It will be assigned to each student using following mapping.
Gujarati - Guj, English – Eng.

GR- This parameter indicates the marks obtained in common entrance test.

6. Evaluation of Parameters

For the development of model, each parameter has to be evaluated and required to find its impact on final result.

7. Future Work & Conclusion

As a part of conclusion, here we have gathered the most affecting parameters for the student's further study. As a space for the proposed work, in our next paper we will test each parameter against the model.

Reference:
1. Mining Education Data to predict student's Retention: A Comparative Study. IJCSIS, Vol.10, No. 2, 2012. Surjit kumar yadav.

2. Data Mining: A prediction for performance improvement using classification, IJCSIS, Vol.9, No. 4, Brijeshumar.

3. Educational Data Mining: a Case Study - Agathe MERCERON and Kalina YACEF.

4. B. Dogan, A. Y. Camurcu. "Association Rule Mining from an Intelligent Tutor" Journal of Educational Technology Systems Volume 36, Number 4 / 2007-2008, pp 433 – 447, 2008.

5. Study of Trends in Higher Education - R. Lakshman Naik, S.S.V.N. Sarma, B. Manjula, D. Ramesh - International Journal of Computer Trends and Technology- Sep to Oct Issue 2011.

6. Alaa el-Halees, "Mining Students Data to Analyze e-Learning Behavior: A Case Study", 2009.

7. Han,J. and Kamber, M., "Data Mining: Concepts and Techniques", 2nd edition. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor, 2006.

8. Pandey, U. K. and Pal, S., "Data Mining: A prediction of performer or underperformer using classification", (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), 2011, 686-690, ISSN:0975-9646.

9. Pandey, U. K. and Pal, S., "A Data Mining View on Class Room Teaching Language", (IJCSI) International Journal of Computer Science Issue, Vol. 8, Issue 2, March -2011, 277-282, ISSN:1694-0814.

10. a classification model for predicting the suitable study track for school students qasem a. al-radaideh, ahmad al ananbeh, and emad m. al-shawakfa.

11. Tissera R., Athauda I., and Fernando C., Discovery of Strongly Related Subjects in the Undergraduate Syllabi using Data Mining, ICIA, IEEE, 2006.

12. Ramaswami M., and Bhaskaran R., CHAID Based Performance Prediction Model in Educational Data Mining, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 1, No. 1, 2010.