Ascertaining the Factors Influencing Students' Performance for Engineering Pedagogy

Sandeep Singh Rawat[1], S. Sreenatha Reddy[2], Devi Prasad Mishra[3] and Salma Sultana[4]
Guru Nanak Institutions, Hyderabad, India
hodcse.gnit@gniindia.org[1], principal.gnit@gniindia.org[2], deviprasadmishra@gniindia.org[3] and salma4534@yahoo.com[4]

***Abstract*—The education domain offers ground for many interesting and challenging data mining applications like astronomy, chemistry, engineering, climate studies, geology, oceanography, ecology, physics, biology, health sciences and computer science. We study the application of data mining to educational data collected from Guru Nanak Institutions, Hyderabad, India.**

We applied very distinctive techniques like association rule and classification algorithms. This work presents an approach for classifying students in order to predict their final grade based on features extracted from student's data in an educational system. This application can help both educators and students, and improve their quality of work. Finally, we analyze the distribution of information across students, and identify factors that predict the number of successful (pass) students.

Keywords*— *Data Mining, Educational System, Engineering Pedagogy and Prediction.

I. INTRODUCTION

Society expects alot from education system, especially from engineering. They expect an engineer to have a strong scientific background, competent technical skills, a good awareness of the social concerns linked with their profession. The role of the engineer is to devise, design and provide machines, devices, vehicles, systems, structures or facilities which improve, sustain or defend the material for better living standard and to make available the technical expertise, services necessary to operate and maintain them.

Present technical education systems are carried out in a very complex and highly competitive environment. The key challenge for any technical education system is to analyze students' performance and to identify their behaviors for their further development. An educational institution needs to have an approximate prior knowledge of enrolled students to predict their performance in future academics. This helps them to identify promising students and also provides them an opportunity to pay attention to, and improve those who would probably get lower grades. In the existing system we do not have any method to predict the students' performance. In some cases institutes consider the performance of student based on previous test and then take measures like special classes, seminars and guest lectures. The shortcoming with existing system are:

- Manual determination of results: Papers are evaluated and then marks are given and finally performance is analyzed.
- Done after the examination: The procedure starts after the examination is completed.
- No scope of improving the students' performance: Since the results are determined, there is no scope for improvement.
- Doesn't consider additional factors: Additional factors such as students' previous performance record is not considered in the existing system.
- Chance of occurrence of human errors: Since this is a manual procedure, the probability of human errors is more.
- Directly the final result is declared: The result is declared and then performance is analyzed.

The rest of the paper is organized as follows. Section two discusses the related work, section three formulates the proposed approach, sections four explains the experimental results and presents simulations that demonstrate the performance of students and give some intuition about the observed results. We conclude in section five with mention of the future likely enhancements of the system.

II. RELATED WORK

Education is viewed as a critical factor in contributing to the long-term economic well-being of the country. An education institution often has many diverse and varied sources of information, which helps it to identify how best to achieve good student performance. The education domain offers many interesting and challenging applications for data mining [1].

It is important to note that there are some factors that are more closely associated with higher performance than others that have a direct effect on the policies and structure of education systems. There is a rich body of research in educational effectiveness that has highlighted factors at different levels of the education system that appear to be more closely associated with higher performance. The aspects and processes that influence the development of education system can be categorized into key concepts such as ability, social background, segregation between schools and neighborhood, teaching conditions in schools, public and hidden differentiation, the relationship between the highest level of education and the labor market and the relationship between education and other social sectors [4].

The quality of the national curriculum is based on the extent to which it meets individual traits, the requirements of financial state of the nation, the needs of society and the future challenges

and aspirations of the nation. It solicits to promote individual and national achievement through the recognition of the different caliber and needs of learners. The current curriculum requires strengthening in terms of developing values and catering for needs and aspirations for self-reliance and entrepreneurship, in order to produce a productive and self-sustaining citizen. Depending on the degree of centralization or decentralization within the education system, decision makers are benefitted at various levels of education systems. Identifying what works at the school level can provide valuable information on potential levers for educational reform, such as initial teacher training, training programs for school managers, in-service teacher training, student enrolment policies, accountability or even policies to restructure the entire school system [5]. We aim to shed light on essential issues in pedagogic strategy, such as:

- The role schools play in quality and equity of student performance;
- The relative impact of school climate and school resources on student performance;
- The degree to which school systems seem to foster selectivity in education; and
- The impact of decentralization and public versus private schooling.

Educational data mining differs from knowledge discovery in other domains in several ways. Higher education will find larger and wider applications for data mining than its counterpart in the business sector, because higher education institutions carry three duties that are data mining intensive:

- Scientific research that relates to the creation of knowledge,
- Teaching that concerns with the transmission of knowledge, and
- Institutional research that pertains to the use of knowledge for decision-making.

Frequently educational institutions find that the students' performance is being affected by several factors and nobody is really sure what the result will be. The system we are considering now will solve all these problems. This allows the administration to keep track of the students' performance in an effective manner. The main objective of this work is to validate and maximize the potential of each and every individual student. This allows the instructors to identify students at risk and provide advice ahead of the exams. There are instances where one wants to analyze the data as early as first year. The instructor may use these findings to reflect on his/her teaching and re-design the course material accordingly [2, 3, 4].

In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. In data mining, a decision tree describes data but not decisions; rather the resulting classification tree can be an input for decision making. Clustering and cluster visualization can be used to identify a particular behavior among failing students. Table 2.1 shows the different classification algorithms with accuracy rate and processing time.

Table 2.1: Different classification algorithms

Classification algorithm	Definition	Accuracy	Processing time (minute)
ID3	This algorithm recursively partitions the training dataset till the record sets belong to the class label using depth first greedy technique [6].	91	<1
C4.5	C4.5 can deal with datasets that have patterns with unknown attribute values. C4.5 can also deal with the case of attributes with continuous domains by discretization [7].	91	<1
CART	CART is unique from other Hunt's based algorithm as it is also use for regression analysis with the help of regression trees [8].	89	<3
SLIQ	It partitions a training data set recursively using breadth-first greedy strategy that is integrated with pre-sorting technique during the tree building phase [9].	79	<5
SPRINT	This algorithm associates the class label along with the record identifier for each value in the attribute lists [10].	87	<2
CHAID	In growth phase of the tree construction, this algorithm uses information gain, an entropy based measure, to select the best splitting attribute, and the attribute with the highest information gain is selected as the splitting attribute [11].	88	<1

III. PROPOSED APPROACH

This work provides a better way to predict students' performance and foresee the result of the student before he/she writes the final exams. In the present scenario, the examination answer sheets are corrected and then based on the marks secured, the performance of the student is analyzed. The proposed work predicts the students' performance well before the results are declared. The instructors may use these findings to reflect on his/her teaching and re-design the course material. The inclusion of certain additional factors such as students' previous performance track record i.e. the marks secured in 12th/Intermediate makes it even more effective in analyzing the capabilities of the student. The results are automated, accurate and thoroughly help the college administration in further enhancing the students' performance as they are predicted before the final examinations and hence there is always a scope of improvement in the student's performance.

The Figure 1 shows the different classes used to represent the system. Basically there are only five classes. The classes are associated with each other.

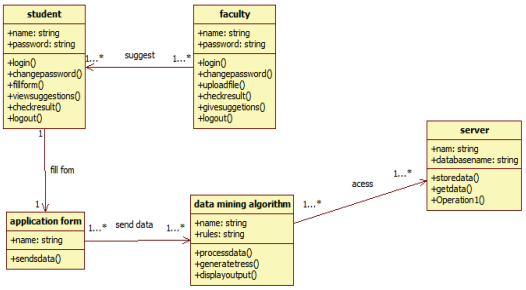
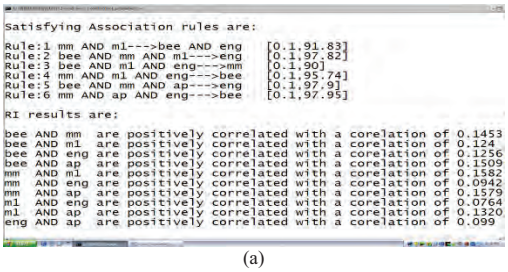


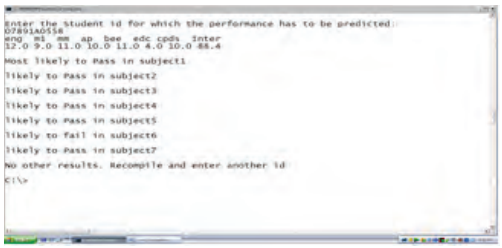
Figure 1: Class diagram of proposed system

IV. EXPERIMENT RESULTS

We have developed simulated examples to determine the performance of the students. In our experimental testing we have used java and sql server in windows environment at Guru Nanak Institutions. The different attributes present in the table are the student ID, Name and marks in each individual subject along with his/her 12th/ intermediate marks. Two separate programs have been developed to test the students' performance.



(a)



(b)

Figure 2 Association rules showing relations between the different attributes for 1st year students

First, programs mine the association rules among different attributes. Based on minimum support, confidence and rule interest values, all the frequent itemsets and then association rules are generated. Figure 2 (a) shows the relation among different attributes for determining the factors, which influence the students' performance. Figure 2 (b) shows the students' performance for a given student id.

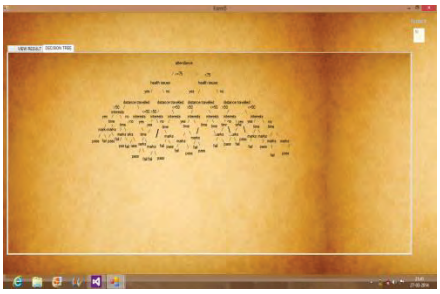


Figure 3: Decision tree generation

In the Figure 3 decision tree is displayed with all the attributes as features. The tree contains the attributes like attendance, distance travelled, interests, marks, time spend on social networking sites.

V. CONCLUSION AND FUTURE SCOPE

This paper acts as a guideline or roadmap for instructor to identify which part of their educational processes can be enhanced and how they can improve their traditional processes by getting advantages of it. The model is also presented as a guideline for technical educational system to improve their decision-making processes.

Many directions for future enhancements are open. Among them, we can mention:

- this work primarily used the support, confidence, rule interest measures and c4.5 to extract the pattern but future research will be done using others interestingness measures for finding hidden pattern and knowledge in making better education system.
- the testing of the proposed work on a real distributed platform and
- we can use our model as a process model to develop an appropriate system for any other technical institute.

REFERENCES

[1] Yiming Ma, Bing Lu, Ching Kian Wong, Philip S.Yu, and Shuik Ming Lee, "Targeting the Right Students Using Data Mining", Conference on Knowledge Discovery and Data Mining, pp 457-464, 2000.

[2] Agathe Merceron and Kalina Yacef, "Interestingness Measures for Association Rules in Educational Data", First International Conference on Educational Data Mining, pp 57-66, 2008.

[3] Sandeep Singh Rawat, H. S. Saini, S. Sreenath Reddy and Ranjan Dandpat, "Tracking Lab Activity in Technical Education System: A Case-Study at the Guru Nanak Institutions (India)", Procedding of International conference on Transformations in Engineering Education (springer), Hubli, 16-18 January 2014.

[4] Sandeep Singh Rawat and Srilakshmi, "Determination of Factors Influencing the Managing Student Enrollment in Higher Education System", Proceeding of International Conference on Information & Communication Technology, pp 279 – 284, DIT, Dehradun, July 2007.

[5] Jaap Dronkers, "Features of Educational Systems as Factors in the Creation of Unequal Educational Outcomes", Book titled Quality and Inequality of Education, ISBN 978-90-481-3992-7, pp 299 – 327, Springer Netherlands, 2010.

[6] John Ross Quinlan, "Induction of Decision trees", Book titled Machine Learning, pp 81-206, March 1986.

[7] J.R. Quinlan, “Improved use of continuous attributes in c4.5”, Journal of Artificial Intelligence Research, pp 77-90, 1996.

[8] Chipman, Hugh A., Edward I. George, and Robert E. McCulloch. "Bayesian CART model search", Journal of the American Statistical Association, pp 935-948, 1998.

[9] Manish Mehta, “SLIQ: A Fast Scalable Classifier for Data Mining”, Decision Trees from large databases, pp 41-77, 1996.

[10] John Shafer,Rakesh Agrawal, Manish Mehta, “ SPRINT: A Scalable Parallel Classifier for Data Mining”, Proceedings of 22nd VLDB Conference, Mumbai, pp1-12, 1996.

[11] Biggs D, B. de Ville, and E Suen, “A method of choosing multiway partitions for classification and decision trees”, Journal of Applied Statistics, Vol. 18, pp 49-62, 1991.