Decision Support Systems Using Data Mining Techniques for Higher Education: A Case Study of College

Sachin Bhoite¹ and Ajay Kumar²

¹ Department of Computer Science, Indira College, affiliated to university of Pune, Maharashtra, India.
² Department of Management, JSPM College, affiliated to university of Pune, Maharashtra, India.
sac78in@gmail.com

ABSTRACT
According to National Knowledge Commission, Higher education in India refers to education beyond school (class 12). The medium term macro objective with regard to higher education would be to increase the gross enrolment ratio to at least 15% by 2015. In today’s competitive world improving & maintaining quality of education in society is challenging task for organizations. In modern times, the knowledge is a key ingredient to the development of the society. As the core competitive ability, the knowledge has become the most important factor to evaluate the quality of education. In many situations it will help to decision makers to better evaluate & enhance the higher educational organizations. To evaluation of such knowledge using mining techniques some interlinking between education domain & Data Mining is required. In this interlinking generally Data mining & Knowledge discovery both terms can be used as interchangeably because Data Mining is the key to KDD process. The effective interlinking between DM & higher education is very important to the sustainable development of the education. In this paper one case study is carried out for analyzing performance of first year (Bachelor of Computer Science) student & its dependency on their last year (Higher Secondary Education) performance. This analysis is done by statistical Data mining (DM) techniques. Using Hypothesis testing (DM) technique we will get patterns of students academic performance for predicate the profile of the students. This study is helpful to prediction of students profile for admitting them in the course. It is helpful to Decision Support Systems (DSS) are well suited technologies to provide decision support in the higher education environments.

KEYWORDS
Data Mining, KDD, DSS, Hypothesis testing, Interlinking, prediction, National Knowledge Commission.

1. INTRODUCTION
In the higher education the huge amount of growth in data in administration of students, courses, teaching, non-teaching staff & managerial systems has been result of current powerful & flexible Information Technology. To utilize such huge data decision makers have to take fast & correct decision to compete in the society. One of the necessary components to manage this difficult task is new techniques and tools for processing the large amount of generated data for extracting some useful knowledge and patterns. In the other words a new technology is required to mine valuable knowledge from the great deal of data. This powerful technology is data mining. Although there are many standard techniques of Data Mining the selection, implementation & if required modification of the technique according to our domain requirements is most cost & time effective task Data Mining & Knowledge Discovery are current areas of research that are influenced by many disciplines including databases, information retrieval, statistics, AI, Visualization and machine learning. Data mining has broad range of applications [1, 2]. It can improve customer service, better target marketing campaigns, identify high-risk clients, and improve production processes. In short, because it can help you or your company make or save money [3].

Data mining is the core part of the Knowledge Discovery in Database (KDD) process. The KDD process may consist of the following steps: data selection, data cleaning, data transformation, pattern searching (data mining), and finding presentation, finding interpretation and finding evaluation. Data mining and KDD are often used interchangeably because data mining is the key to the KDD process [4, 5].

An important Knowledge Discovery and Data Mining goal is to “turn data into knowledge”. For education domain knowledge acquired through such methods are useful for developing educational system such as maximizing educational system efficiency, decreasing student's drop-out rate, increasing student's promotion rate, increasing student's retention rate, increasing student's transition rate, increasing educational improvement ratio, increasing student's success, increasing student's learning outcome, and reducing the cost of system processes. In order to achieve the above quality improvement, we need a data mining system that can provide the needed knowledge and insights for the decision makers in the higher educational system [8].

The information technology's development would provide a better platform in the management of education. The purpose of this paper is to analyze performance of F. Y. students & its dependency on their HSC performance. For this analysis one case study is carried out by collecting two years data of the (F.Y.) students. Cleaning & preprocessing it in useful manner & finally applying Hypothesis Testing Mining (Chi-Square)
technique on it. This study is useful as prediction to decide students profile to admit in the course. This paper is structured as follows Section 2, explains Review of Related Works section 3, briefly explains the data mining & its technology in higher educational system. Section 4, demonstrates and explains the case study in education domain. Section 5, explains results obtained are discussion Section 6, briefly concludes the whole paper and identifies the further works. In India National Knowledge Commission would imply more than doubling the scale of higher education within the next few years. Further the system needs to be expanded without diluting quality and in fact by raising the standard of education imparted and making higher education more relevant to the needs and opportunities of a knowledge society. So in support of this, the educators have to better evaluate & enhance the higher educational organizations [7].

2. REVIEW OF RELATED WORKS

In this section, some related researches about data mining in higher education using mining techniques Decision tree, association rules are briefly reviewed.

The main idea is that data mining technology can discover the hidden patterns, associations, and anomalies from educational data. This knowledge can improve the decision making processes in higher educational systems [1, 2, 8, 9, 10, 12, and 13]

1. The three basic functional indicators in higher education are a. Educational System Input b. Process c. Output & all the detailed activities under each functional indicator.
2. The main components the data Mining for higher educational processes.
3. To understand the objectives of higher education
4. To understand educational data.
5. Preparing data & ends up with modeling & evaluation
6. Finally converting the result to understandable by humanity/end user.

3. BRIEF REVIEW OF DATA MINING

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified.

There are many techniques or algorithms available for performing the different tasks [1, 2]. For example Market Basket Analysis and Apriori algorithm are techniques that are used for association rule mining [2].

3.1. EXAMPLES OF INFORMATION EXTRACTED USING QUERY LANGUAGE & DATA MINING

- **Query language is used for**
  - List customers who use credit card to purchase more than Rs 1000 worth groceries.

- **Data mining is used for**
  - List patients who had at least one heart attack
  - Develop a general profile of credit card customers
  - Determine patients whose lifestyle is prone to getting a heart attack in near future.

The tasks of data mining are very diverse and distinct because many patterns exist in a large database. Different methods and techniques are needed to find different kinds of patterns. Based on the patterns we are looking for, tasks in data mining can be classified into Summarization, classification, clustering, association and trend analysis. Finding the patterns and regularities in the data evolutions along the dimension of time can be fascinating.

Data mining adopted its techniques from many research areas including statistics, machine learning, association Rules, neural networks, and so on. Decision tree induction: the best known ML framework was found to be able to handle large-scale problems due to its computational efficiency, to provide interpretable results, and, in particular, able to identify the most representative attributes for a given task [6].

3.2. HIGHER EDUCATIONAL SYSTEM

Nowadays the higher educational system has great impact of information technology. It involves Internet, Intranet, Extranet, the technology of advance database management system, the data collection and gain technology, different platform independent, reliable less expensive application software. But which are not sufficient to tackle the problems encountered in higher education. It prevents them to achieve their quality objectives. For example, many educational systems do not have accessed to necessary knowledge for accurate prediction of student's academic outcome so that proactive academic intervention can not be taken on time and they would not be able to forestall a student from failing before he or she even knows it. The result of such problems affect in increasing student's drop-out rate, decreasing student's transition, promotion, and retention rate which is more or less different in various colleges & universities.[9] Also we don’t have prompt system to decide profile of the student to admit them for a particular course. We believe that data mining technology can help us to interlink these two areas. The hidden patterns, knowledge that are discovered by data mining techniques from educational data can improve decision making processes in higher educational systems. This improvement can bring advantages such as maximizing educational system efficiency, decreasing student's drop-out rate, increasing student's promotion rate, increasing student's retention rate, increasing student's transition rate, increasing educational improvement ratio, increasing student's success, increasing student's learning outcome, and reducing the cost of system processes [9, 10, 12].

4. CASE STUDY IN EDUCATION DOMAIN
The Higher Secondary School (HSC) exam is stepping stone for students. Specifically for science students many carrier development options are available. With the engineering and medical field, study in computer science is equivalent option now a day. The computer science is one of the attractive and challenging fields. Though the meritious and quality students have obtain the computer science for graduation. The degree’s syllabus and examination pattern is different form higher secondary level. There are serious controversies found in first year degrees’ result. The student got good marks in HSC not succeed in first year result as well as those students pass with lower class in HSC achieve a first class and distinction in degrees’ first year.

All the students are not having equal qualities and marks. Some educational thinkers and analyst well noted that, marks and achievement in HSC reflects the student’s further performance and quality. The HSC mark and class reveals the future study and achievements. Today’s assessment process is purely based on the marks of the students but there are very less concentration on the students past performance and record. On contrarily to this failing of students in examination the blame goes to educational institute and teaching faculties. Therefore for present study researchers want to know the relationship of first year degree level marks with HSC marks. Such type of the study also important for finding linkages and issues of student’s performance in T.Y. exam and course offered educational institute.

4.1. COMPUTER SCIENCE COURSES AT INDIRA COLLEGE OF COMMERCE AND SCIENCE, PUNE (ICCS)

ICCS is carrying BCA, B.Sc., and M.Sc. in computer sciences. The first year degree computer science courses have 160 intake capacity affiliated to University of Pune, Maharashtra, India except year 2005-06. In 2005-06 the capacity was 80. Admissions are done according to university norms. The F.Y. B.Sc. admission is done on the basis of H.S.C. merit and as per reservation norms.

The faculty in ICCS is having in-depth theoretical as well as practical knowledge. Guest lectures are organized for needy subjects by subject experts. It helps to broaden their view.

4.2. OBJECTIVES

The objectives of this study are demarcated below:

1. To study the available Data Mining & Knowledge Discovery technologies.

2. To investigate the use of Data Mining & Knowledge Discovery technology for the available data of higher education institutions.

3. To study available data, suggest initiatives for capturing necessary data and identify methods to discover useful information from it. (Knowledge or Unknown patterns)

4. To investigate ways of facilitating storage of such useful information/knowledge and make it available to Knowledge workers or end users.

5. To decide profile of students on the basis of performance analysis & provide necessary support to solve their issues.

Apart from these

i. To analyze the H.S.C. & T.Y.B.Sc (Comp Sci.) Students marks in year 2007 to 2010.

ii. To evaluate the performance of T.Y.B.Sc (Comp Sci.) students with their past examination.

iii. To find dependency of T.Y.B.Sc (Comp Sci.) class with H.S.C class.

4.3. MATERIALS AND METHODS

1) Data Source- Present study is based on the exam class of the student’s in H.S.C. & T.Y.B.Sc (Comp Sci.). The exam marks of the student’s are collected form ICCS. The all selected student’s exam performance is belonging to one institute though the sample is homogeneous.

2) Study area and Limitation of study- For present study Indira College of Commerce and Science, Pune-33 selected. Though the result of this study is limited for Indira College of Commerce and Science, Pune-33 in year 2007-08, 2008-09 and 2009-10.

3) Statistical operations- Various statistical tools are used in present study; these are total, average, percentile, chi-square test etc.

<table>
<thead>
<tr>
<th>H. S. C.</th>
<th>T.Y.B.Sc (Comp Sci.) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
</tr>
<tr>
<td>F</td>
<td>3</td>
</tr>
<tr>
<td>S</td>
<td>1(3)</td>
</tr>
<tr>
<td>P</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
</tr>
</tbody>
</table>

4) Hypothesis- For drawing the dependency for two variable the following hypothesis is defined-

“Attributes i.e. Student’s class in H.S.C. exam & T.Y.B.Sc (Comp Sci.) exam is independent.”

Due to space problem all the life cycle of mining is not mentioned over here. All the collected data is made available by in required form for chi-square test for all the three years.

5. RESULT AND DISCUSSION

Table no. 1. Students performance in T.Y.B.Sc (Comp Sci.) exam and its’ dependency on H.S.C. performance in year 2007-08

Copy Right © INDIACom-2011 ISSN 0973-7529 ISBN 978-93-80544-00-7

In above first 3 tables students performance in T.Y.B.Sc.(Comp. Sci.) exam and its’ dependency on H.S.C. performance is calculated & displayed from collected data for Chi-square test of independence of two attributes for three different years. From Table No. 4 observations all three years value of $X^2 cal > X^2 tab$ So it is noted that hypothesis is rejected in all study years. As the chi-square’s calculated value is significant than the table value (df=12) in all respective years. Therefore it is clearly stated that first year B.Sc.(Comp. Sci.) exam’s performance (class) is dependent on the performance in H.S.C. exam.

CONCLUSION

Data Mining is very important tool for discovering meaningful patterns or knowledge, understanding relationship among the available data. In studied case from the available bulk amount of student’s data essential data is selected, cleaned & integrated for these work. By applying hypothesis testing (statistical) Data Mining techniques on these data the dependency of two years marks is tested. So from the above result student profile for admitting the studied course is predicted. Such type of analysis is useful for the Decision Support Systems (DSS) to provide decision support in the higher education environments.

FUTURE SCOPE

This type of analysis will help in making standard of the institute in the society. So further this type of study is also carried out for [3]

1) Evaluation of students assessment, teachers assessment, Course assessment, Industrial training assessment
2) Student course registration
3) Student performance, Lecturer performance
4) Student examination
5) Student behavioral consulting, Program selection counseling

REFERENCES

[3] Delavari N, Dr. Mohammad Reza Beikzadeh, Dr. Somnuk P.” Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System”, Dominican Republic ITHEF 6th Annual International Conference Juan Dolio
Conference on Knowledge Discovery in Databases and Data Mining, Newport Beach, California, 1997.


