Data Mining in Higher Education System and the Quality of Faculty Affecting Students Academic Performance: A Systematic Review

Hema Malini B H¹, L Suresh²
¹ Associate Professor, BMSITM, Bengaluru
² Principal, Cambridge Institute of Technology, Bengaluru

Abstract. Higher education plays a very important role in the global scenario in the present day situation. Basic education has become a preliminary requirement. Since the education rate is increasing day by day and competition for placement is increasing in the campus, the quality of the potential faculty also plays a very important role in carving the student’s future. Hence, the present study considers the various factors of faculty which affects the quality of the students produced to the global market. In the present paper the personal and professional credentials of five potential faculties are considered. The result analysis for the subjects they have handled for four consecutive years is collected. An effort is made to map the faculty credentials with the outcome produced. The inferences are drawn based on the performance of students.

Keywords: Educational Data mining, Teaching-learning process, Academic performance.

1 Introduction
Teaching has become a challenging task in the present day scenario wherein all information is obtained by just a click of the mouse on the internet. The availability of online courses, study material on the web has provided more choice for the students. The various schemes like CBSE, ICSE provide more exposure to the present day students. So, the time the students enter the professional courses, already have an exposure to the technology. The expectation of students is more. The demand in the industry is also more. Hence, the faculties have to remain updated at all times to handle the present day students.

The increased number of schools and colleges to cater to the needs of students has provided more competition in the field. The students’ academic performance mainly depends on the quality of teaching.

The present study deals with collecting faculty information like the date of birth, institution studied, percentage secured in UG and PG examinations, human attitude and tries to relate the faculty caliber with the student performance by relating it to the subjects the faculties have handled in past 4 years. The result analysis of all eight semesters in four consecutive years is considered and study is done. An effort is made to check whether the faculty background, the attitude of the faculty has any impact on the student performance.

1.1 Data Mining Techniques
There are various techniques utilized by Educational Data Mining such as classification which includes k-nearest neighbor, Naïve Bayesian, Decision tree and Rule based induction in association etc. For Educational Data Mining, the data mining techniques will be used. The key techniques in Data Mining are:

Association: A simple correlation between two or more items is made, frequently of the same type to identify patterns.

Classification: Classification can be used to build up an idea of the type of customer, item, or object by describing multiple attributes to identify a particular class.

Clustering: Clustering allows to use common attributes in different classifications to identify clusters.

Prediction: Prediction is an extensive topic and runs from predicting the failure of components or machinery, to identifying fraud and even the prediction of company profits.
Sequential patterns: Sequential patterns are a useful method for identifying trends, or regular occurrences of similar events.

Decision trees: The decision tree can be used as a part of the selection criteria, or to support the use and selection of explicit data within the overall structure.

Combinations: In practice, it's very rare that you would use one of these exclusively. A combination of any two techniques might be used.

2. Literature Survey
The work done in from 2012 to 2017 is collected in the area of Educational Data mining. The same is discussed below in detail.

Leena et al. [1], discuss that the Quality Education in any of the country depends on the student’s characteristics and parameters like institution name, current market trend etc. Defining the relevant parameters which affect the performance of students and teaching learning process is a leading research area in educational data mining.

Rohit Ahlawat et al. [2] conclude that K-means clustering technique was performed on the crime rate, enrolment, and companies’ dataset. The data for analysis performed on the clusters formed showed that enrolments for higher education is very less in north east states while enrolments were high for southern states like Kerala, Karnataka etc. This trend of enrolments may be due to crime rate of state, status of proper infrastructure, existence of good companies or population.

Deka Ganesh Chandra et al. [3] explore the different phases of e-learning development from an organizational perspective and explore the different phases where data mining can be applied. The possibility of applying EDM in Learning Management Systems (LMS) with the help of SCORM (Sharable Content Object Reference Model) standard is also discovered. SCORM is a widely used e-learning standard maintained by more than 90% of the LMS vendors in the market. The gaps of SCORM and the next generation LMS technology is also discussed in this paper.

Suhem Parack et. al [4] discuss the application of data mining in education for student grouping and profiling. They use an Apriori algorithm for student profiling i.e. discovering co-relations among set of items. K-means clustering is used for grouping students that assigns a set of observations into subsets.

It would be very tough to physically go through the huge set of academic records to identify the student trends and performance and the pattern in which they study.

By using data mining techniques on the large amount of academic record, it is easy to categorize the students, identify the hidden patterns about their learning pattern, find the undesirable student behaviour and perform student profiling.

A. Banumathi; et. al [5] focus on upgrading the Indian education system by using clustering in Data mining. It is very difficult to predict the behaviour of students from large data base. Cluster analysis sorts the data given into some significant groups. In this paper, the authors attempt to analyse student’s data in different approach beyond the indicated patterns (good, average, poor) through newly proposed UCAM (Unique clustering with Affinity Measures) clustering algorithm.

This paper published by Fábio de Paula Santos et. al [6] is part of a doctoral thesis which aims to propose an evaluation model, for future application, using Educational Data Mining techniques to analyse the responses of pupils obtained during an Institutional Teaching Evaluation. The proposed model applies the Sentiment Analysis to identify the teaching practices are positive or negative from the outlook of students from a Higher Education Institution.

The significance of this research exist in listening to students and making sure their involvement in the discussions on the future of Higher Education supporting changes in the teachers’ pedagogical practices.

Diego Buenaño Fernández et. al [7] compare the technical features of three open source tools (RapidMiner, Knime and Weka) as they are used in educational data mining. A practical case study of the academic records is carried out on three engineering programs in an Ecuadorian university.
3. Faculty Data Collection

The faculty personal information like the institution studied, year of passing the degrees, total experience is collected and same is tabulated in Table 1. To start with the faculties of department of Computer Science and Engineering are considered. On an average, the faculties considered have more than a decade of teaching experience. So, over the years, the college they have studied may not make much difference. They are equally competent.

<table>
<thead>
<tr>
<th>Faculty</th>
<th>Year of Passing</th>
<th>UG % Secured</th>
<th>Year of Passing</th>
<th>PG % Secured</th>
<th>Experience in Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faculty1</td>
<td>1995</td>
<td>56%</td>
<td>2005</td>
<td>62%</td>
<td>20</td>
</tr>
<tr>
<td>Faculty2</td>
<td>2001</td>
<td>61%</td>
<td>2007</td>
<td>78%</td>
<td>13</td>
</tr>
<tr>
<td>Faculty3</td>
<td>1998</td>
<td>58%</td>
<td>2009</td>
<td>7.9 CGPA</td>
<td>13</td>
</tr>
<tr>
<td>Faculty4</td>
<td>2004</td>
<td>68%</td>
<td>2011</td>
<td>69%</td>
<td>7.5</td>
</tr>
<tr>
<td>Faculty5</td>
<td>2005</td>
<td>67%</td>
<td>2012</td>
<td>FC</td>
<td>12</td>
</tr>
</tbody>
</table>

4. Result Analysis: Pass Percentage

In Engineering, the examinations are conducted in semester schemes. The result analysis of five potential faculties is collected for consecutive 4 years. Usually a faculty handles one/two theory subjects in addition to laboratory and other workloads. The results of those subjects is considered. The student data viz., the number of students appeared, the number of students passed and the pass percentage is collected. The result analysis is done. The overall pass percentage is tabulated. The same is tabulated in Table 2.

<table>
<thead>
<tr>
<th>Faculty</th>
<th>2016-17</th>
<th>2015-16</th>
<th>2014-15</th>
<th>2013-14</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odd</td>
<td>Even</td>
<td>Odd</td>
<td>Even</td>
</tr>
<tr>
<td>Faculty1</td>
<td>96.5</td>
<td>96.5</td>
<td>96.4</td>
<td>96.4</td>
</tr>
<tr>
<td></td>
<td>91.2</td>
<td>91.2</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>Faculty2</td>
<td>96.5</td>
<td>98.4</td>
<td>100</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>98.2</td>
<td>100</td>
<td>96.9</td>
</tr>
<tr>
<td>Faculty3</td>
<td>96.5</td>
<td>96.5</td>
<td>87.8</td>
<td>87.8</td>
</tr>
<tr>
<td></td>
<td>86.2</td>
<td>86.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Faculty4</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>98.1</td>
<td>98.1</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Faculty5</td>
<td>93.4</td>
<td>93.4</td>
<td>87.5</td>
<td>87.5</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
5 Inferences Drawn:
The result of analysis of five potential faculties is collected for 4 consecutive years. The following inferences can be drawn.

The average results are consistently above 90%. It shows that the faculties are focusing on OBE (Outcome Based Education). The institution where the faculty has studied, does not affect the students they teach since over years, they gain experience. The students’ performance is consistently monitored and they are professionally trained to face the university examinations. The results in the higher semesters / final year are considerably high, since the students are already aware of preparing and facing the university examinations. The results in the lower semesters are considerably low since the students come from various backgrounds, have to accustom themselves for professional courses, understand the university requirements and prepare for semester end examinations. The focus should be provided to improve the performance of lower semester examinations.

6 Conclusions
The data collected from faculties clearly show that the performance of students is considerably very good. The faculty contribution is also very high. The continuous monitoring of students, identifying slow learners, conducting remedial classes etc. are regularly done. This can happen only by experienced, dedicated and highly committed staff. So, the work clearly indicates that quality staffs play a significant role in promoting higher education. Work has to be carried out on improving the performance of students in lower semesters. Role of faculty in identifying such students and providing them support has to be researched.

7 References


6. Fábio de Paula Santos; Carla Pineda Lechugo; Ismar Frango Silveira-Mackenzie, "Speak well" or "complain" about your teacher: a contribution of Education Data Mining in the evaluation of teaching practices, 2016 International Symposium on Computers in Education (SIEE), Pages: 1 – 4, IEEE Conference Publications.
