Applications of Data Mining in the Indian Higher Education Sector: A Review

Harsh Vardhan Pant

Ph. D. Scholar, Graphic Era Hill University, Bhimtal campus

Prof. Manoj Chandra Lohani, Graphic Era Hill University, Bhimtal campus

Dr. Jeetendra Pande,Assistant Professor, School of Computer Sciences & Information Technology, Computer Science.

ABSTRACT

Among the biggest challenges that the Indian higher education faces today is predicting the education-related aspirations of students. Educational institutions need information such as which students will enrol in specific courses, which students are likely to need more support in one or more subjects, students' probable academic attainment, the success of new courses offered etc. The answer to all these problems lies in the rapidly evolving field of *Data Mining*. Data mining can help educational institutions take decisions more accurately and efficiently. This paper discusses the broad areas to which data mining is applicable, the scope of educational data mining and discusses various data mining tools and techniques that can be effectively used in assisting educational institutions.

Keywords: Educational Data Mining, Data Mining Techniques, Applications of Data Mining

Introduction

No country can really develop unless its citizens are educated - Nelson Mandela

Generally, developed countries have high literacy rates. Thus, the literacy rate of Pakistan in the year 2015 was 58.7 % while that of the USA was 92.2 %, as per the 'World Fact Book' for 2015 [1]. Since this huge gap cannot be filled only by universities and colleges set up by the government, a large number of private universities, educational institutions and online courses have come up in India during the last two decades and this trend will continue in the foreseeable future.

Despite having made significant progress in the recent past, Indian higher education is faced with following broad challenges:

• Gap between Supply and Demand

Year	No. of Institutes	No. of Sanctioned Seats	No. of students enrolled	Vacancies
2019-2020	537	2,00,000	-	-
2018-2019	501	2,49,505	1,12,032	1,37,473
2017-2018	518	2,66,666	1,46,087	1,20,579
2016-2017	523	2,76,842	1,44,587	1,32,255

Table-1 From the newspaper 'The Hindu', 19 May Chennai"

In 2018-19, only 45% of the seats in engineering colleges in the states were filled. Deemed universities did slightly better with 59% of its seats being filled. Although data for previous years are not available for deemed universities, it is obvious that they too have been facing the same set of problems as Anna University's affiliated colleges. [2]

• Poor quality of teaching and learning

According to the 2018 Parliamentary Standing Committee Report on Human Resource Development, entitled 'Demand for Grants 2018-19 of the Department of Higher Education', approximately 35 per cent of teaching positions at various Central Universities under the purview of UGC are lying vacant (1,323 Professors; 2,217 Associate Professors, 2,457 Assistant Professors). [3]

A review of literature suggests that the Indian higher education system is suffering from several issues, i.e. quality of institutions; a chronic shortage of faculty; poor quality teaching; outdated and rigid curricula and pedagogy; lack of accountability and quality assurance.

• Deficient quality of research

The number of students pursuing Ph. D. has registered an increase from 0.1 million in 2013-14 to nearly 0.2 million in 2017-18 (AISHE, 2017), which is a positive development from the viewpoint of the supply of qualified teaching staff [23]. However, it is much less than that in most other significant countries.

In terms of the number of researchers engaged in research activities, India has merely 119 researchers per million prople, whereas Japan has 5,287 and the US has 4,484. Even in absolute terms, the number of researchers in India is much smaller than that in the US, China, Japan, Russia and Germany. However, the number of technicians engaged in research activities in India is not as small, suggesting that India has more technicians per researcher, as compared to most other countries. The numbers of doctoral degrees awarded in science and engineering in India is a little over 6,000, compared to 9,000 in China and 25,000 in the US. In China, this figure has increased rapidly from a little over 1,000 in 1990 to over 9,000 now. In comparison, there has been a modest increase in India. The National Science Foundation (NSF) Science and Engineering

Indicators – 2002 shows that in the US about 4% of the science and engineering graduates finish their doctorates. This figure is about 7% for Europe and merely 0.4% for India. **[4]**

Thus, various studies indicate that, in India, not only the number of researchers is deficient but enough high-quality research is not being conducted. There are few opportunities for interdisciplinary and multidisciplinary work; a glaring lack of early-stage research experience; a weak ecosystem for innovation and low level of industry engagement.

Data mining can help meet these challenges, which can improve the prevailing education system. The growing digitization of educational data has helped researchers easily extract meaningful information, thus helping in taking corrective measures. Data mining, which involves the extraction of hidden predictive information from large databases, is a powerful technology with great potential to help the education sector focus on the most important information in their data warehouses.

Data Mining Techniques

Various data mining techniques are as follows:

Tracking patterns: This technique helps recognize patterns in datasets. For example, we might see that the sale of an umbrella product spikes just before the winter or summer weather.

Classification: Classification allows categorization of various attributes together into discernable categories, which can be used to draw conclusions or serve other purposes. For example, if we're evaluating data on students' grades and performance histories, we might be able to classify them as 'Good', 'Better' and 'Best'. We could then use these classifications to learn more about those students.

Association: Association is related to tracking patterns but is more specific to dependently linked variables. We may notice that when the grade of a student is best then the student's academic participation will also be good.

Outlier detection: In many cases, simply recognizing the overarching pattern cannot provide a clear understanding of the dataset. For example, if the best performers are almost exclusively female but during the July month of one particular year there's a huge spike in the performance of male students, we might want to investigate the spike and see what drove it so that it can be better understood.

Clustering: Clustering is very similar to classification but involves grouping chunks of data together, based on their similarities. For example, we might choose to cluster different demographics of students into different locations, based on their performance.

Regression: Regression, used primarily as a form of planning and modelling, is used to identify the likelihood of the existence of a certain variable, given the presence of other variables. The main focus of regression is to help uncover the exact relationship between two or more variables in a given dataset.

Prediction: Prediction is one of the most valuable data mining techniques since it's used to project the type of data likely to be present in the future, e.g. prediction of student performance in semester examinations.

Applications of Data mining in the education sector

Data mining can be applied to the education sector in various ways, the more notable ones being:

Social-emotional skills

Using social-emotional skills, a person acquires the capacity to understand, analyze, express and manage emotions. He also learns how to develop a relationship with others. Thus, this is a non-academic skill that plays a major role in defining the learning capabilities of students.

Using data mining techniques it is possible to-gather large quantities of information and, incorporating it with other tools, produce dynamic and informatics result. When applying various predictive analytical techniques in the educational database, data mining helps teachers better understand the motivation of their students.

Measuring the instructor's performance

The performance of students depends substantially on teachers. Linda Darling-Hammond **[5],** in the report entitled 'How Teacher Performance Assessments Can Measure and Improve Teaching' dwell on the connection between teacher performance and student outcomes.

Traditional techniques for evaluating the performance of teachers have mostly been subjective and manual. Thus, student reviews of teachers' performance have been the standard tool for quantifying teaching methodologies. However, these techniques are quite inefficient and time-consuming. Furthermore, reading student reviews and creating an analogy is a tiresome task.

Data mining techniques are applicable not only to recorded data but also real-time data. As a result, real-time data collection and teacher monitoring are possible, along with analysis of said data being collected. Furthermore, using Natural Language Processing, it is possible to analyze the sentiments and end-reviews of students for providing a comprehensive analysis of teacher performance.

Innovating the curriculum

In the prevailing competitive scenario, if educational institutions do not keep updated with the demands of the industry and do not provide access to appropriate courses for students, their existence would be in jeopardy. Thus, it is a challenge for them to keep up with changes in industries. In order to accommodate this, many educational institutions have started using data mining for analyzing emerging trends in the market.

Aher et al. **[6]** have presented a paper in which they have explained how the combination of clustering algorithm – Simple K-means Algorithm & association rule algorithm- Apriori Association Rule is useful in Course Recommender system.

So, in combination with various statistical and monitoring techniques, data mining can be helpful to course creators in keeping pace with industrial evolution.

Monitoring student's requirement

For teaching to lead to effective learning, high-quality assessment is imperative. Assessment, therefore, needs to be integrated into the teaching-learning process. It needs to be designed in such a manner that it becomes a powerful means of influencing the quality of what teachers teach and what learners learn **[23]**.

Predicting students' course selection

Selection of course by students depends on various factors. Using neural networks, Kardan et al. **[7]** have determined various factors that influence student course selection, such as, students' workload, course grades, course type, course duration, time conflicts, final examination time and students' demand. These factors are used as inputs for neural network modelling. Furthermore, Guo **[8]** analyses and predicts students' satisfaction with courses, using neural networks. He has found that the number of students enrolled in a course and a high distinction rate in the final grading are the two most influential factors affecting student course satisfaction.

Predicting students' placement opportunities

A big challenge in higher education is student placement, in ensuring which most institutions have been struggling. With students becoming more and more demanding, ensuring quality placement of students is not only crucial for educational institutions' survival but also critical to the creation and maintenance of their brand equity.

Shreenath Acharya & Madhu N **[9]** proposed a model for predicting student's placement based on historical information of database which can be utilized by educational institutions to reveal some interesting patterns that could be analyzed to plan their future activities. It has been found to be genuinely useful for the higher authorities like the Director and the Head of the departments.

Conclusion

Using educational data mining techniques, researchers and stakeholders can enhance student fraternity's satisfaction. Application of data mining in the educational sector can help us find students' requirements, not only in a descriptive and predictive analytical manner but also in a prescriptive one so that suitable actions can also be prescribed. Understanding students, appropriate profiling and accurate predictions will not only increase the quality of education but also help the education system in formulating appropriate strategies for youth. Due to rapidly increasing usage of the internet by students, an increasing number of students is taking admission in online and other MOOC courses. A huge data will be available about them, so there will have need the solutions and know their requirement of student's real-time. Our next objective will be to know the requirements and behaviours of MOOC students.

References

[1] Mr. Abbas Ali, "Difference between developed and developing countries", "https://www.slideshare.net/malikabbas982/difference-between-developed-and-developing-countries"

[2] News Paper, The Hindu, 19 May Chennai "Supply-demand gap in engineering seats to continue, "https://www.thehindu.com/news/cities/chennai/supply-demand-gap-in-engineering-seats-to-continue/article27175363.ece"

[3] Web Link, "Low student-teacher ratio vs quality of higher education", https://www.dnaindia.com/analysis/column-low-student-teacher-ratio-vs-quality-of-higher-education-2705415

[4] MALLIKARJUN I MINCH, ASSISTANT PROFESSOR, DEPARTMENT OF POLITICAL SCIENCE, COLLEGE, SINDAGI, -586128 DIST- BIJAPUR (KARNATAKA), "HIGHER EDUCATION AND RESEARCH IN INDIA", International Journal of Social Science & Interdisciplinary Research, ISSN 2277 3630 IJSSIR, Vol. 2 (3), MARCH (2013)

[5] Linda Darling-Hammond October 2010, "Evaluating Teacher Effectiveness", "Article in Phi Delta Kappan • March 2012 DOI: 10.2307/41497541"

[6] Sunita B. Aher, Lobo L.M.R.J, Prediction of Course Selection by Student using Combination of Data Mining Algorithms in E-learning, International Journal of Computer Applications (0975 – 8887)

[7] Ahmad A.Kardan, Hamid Sadeghi, Saeed Shiry Ghidary, Mohammad Reza FaniSania "Prediction of student course selection in online higher education institutes using neural network", Elsevier computer and education volume 65, July 2013

[8] Sonja ISLJAMOVIC, Milija SUKNOVIC, Faculty of Organizational Sciences, University of Belgrade, PREDICTING STUDENTS' ACADEMIC PERFORMANCE USING ARTIFICIAL NEURAL NETWORK : A CASE STUDY FROM FACULTY OF ORGANIZATIONAL SCIENCES, ICEMST 2014: International Conference on Education in Mathematics, Science & Technology

[9] Mr. Shreenath Acharya , Ms. Madhu N, Information Science & Engg department, St. Joseph Engineering College, Discovery of students' "academic patterns using data mining techniques" International Journal on Computer Science and Engineering (IJCSE)