



INSPECTION OF DATA MINING TECHNIQUES IN HIGHER EDUCATION SYSTEM

Sunil Kumar Tiwari

MATS University

Archi Dubey

MATS University

ABSTRACT Higher education is essential for the growth and development of a country. Higher education has been under incomparable pressure to provide better access to our institutions. One of the key points of the higher education is the increase in education data explosion. This information is fast growing without the management and organization facilities. We believe that managing this huge amount of information is a daunting task. However, with new strategies and tools of data mining, we can easily process large quantities of information generated in business processes and find useful knowledge and information. Data Mining is a technique for extracting prediction information hidden from large databases. It is a powerful technology with significant potential that helps universities or higher education institutions to focus on key information in their data warehouses. In this paper we focus on the various data mining techniques that are useful for solving various problems happening in higher education of India.

KEYWORDS : Higher Education, Data mining, Clustering, Decision Tree, prediction.

INTRODUCTION

Higher education has been under unprecedented pressures to provide greater access to our institutions, ensure student learning and success, respond to rapidly changing demographics, increase accountability, and develop greater efficiencies as we face significantly reduced resources, especially in the public sector. Strategic enrolment planning is a critical tool in responding to these pressures and overcoming them. Our ability to recruit, enrol, retain, and graduate a diverse, high-quality student body is significantly enhanced by a dynamic and comprehensive plan that includes both a short term and long-term focus, and that is seen as part of an ongoing and adaptive process. Now a day's educational data are collected and stored in educational institution but not studied properly to understand the hidden information from it. Earlier statistical methods were used for this purpose but advent of computer added an advantage of complex study, algorithmic approach, speedy process etc. feature which make it more useful in place of only statistical process [1].

The financial system of the country is direct depends on the education of the students who have a effects in the industry. Excellence in education Institutions are taken into account by the student success rate and the capabilities of institutions are measured by withheld rate of vulnerable students. The different aspects as an individual, social and psychological will be helpful measuring the academic performance of the students. That can lead to discover the students who are at risk and this help management act quickly. The academic achievements of the students are measured socio-economic and university level power. This process is performed by using data mining techniques for educational purposes. The class determination is carried out in advance examine the data for it to be called attended lessons [2]. On the basis of previous scientific achievements and Socio-economic situations of student performance were measured using data mining techniques. The classification divides the data into predefined sets or groups Course groups. We often talk about supervised learn because classes are stubborn before verification of the data. The patterns discovered by data mining methods from educational data can be used to improve the decision-making process identify vulnerable students, reduce school dropout rate, increases student success and increases Student learning outcome [3].

The discovery of educational data is becoming a research area with many calculus and psychologists research methods and approaches to understanding as students learning. Data mining education in Data mining recovers hidden knowledge through the application of various data mining techniques like clustering, rules mining, web mining, test mining, neuronal networks, Bayesian Networks and many others which gives us a final result and if it needs a request then to change the raw data again when filtered. If needed Data from educational institute can be collected in large numbers. Students and many variables can hold these data exploration may seek algorithms for various data modelling. Data mining is also known as knowledge discovery. Database is a technique used in various disciplines look for meaningful

relationships between variables. In a large database, it removes hidden knowledge and finds the database and the necessary information according to the requirement. It is done by implementing various data mining algorithms in learning datasets. This recording passes through several stages of Data mining and finally filtered data coming from user as per requirement. It plays one important role in educational institution.

Measuring student achievement helps interns for their future. Let the students completely decide their wearer according to the understanding of society, they have waiting for the skills of the students as a skill like communication, percentage, positive and good attitude and performance. When the education planner leads education plans ahead of time, it can protect those risky students who have also improved earlier with total productivity [4]. Learning needs are different for different groups of students; they can see using data mining techniques in education [5].

I. DATAMINING CONCEPT

Data mining is a way to extract hidden estimate information from large databases. This is a very new and huge technology which has great potential to help the university and Institutions. It is used to focus on the most important information in the data warehouse. Data mining is a powerful, new and emerging technology with great potential for information systems. This can be defined as the automated process of extracting relevant knowledge and information, including models, associations, trees, changes, trends, anomalies, and meaningful structures from large data sets. Our main idea is that hidden patterns, associations, classifications and anomalies discovered through data mining techniques can help fill this gap in higher education [6].

The term "data mining" is often used for the two distinct processes as "knowledge discovery" and "prediction" processes. Knowledge Discovery provides explicit information that is readable and understandable by a person at the end of the user. Forecasts or predictive models provide forecasts of future events that may contribute to improvement, may be transparent and legible in some approaches, and opaque in others, such as neural networks. Data mining is based on the use of real-world data. These data are extremely vulnerable to co linearity because real-world data can have unknown relationships with each other. Data mining is the entire process of applying computer methods, including new knowledge discovery technologies and technologies. Data mining tools predict future trends and behaviours, enabling institutions to make proactive, appropriate and knowledge-based decisions. The automated prospective analytics offered by data mining technology goes beyond the analysis of past events provided by retrospective tools, typically decision support systems. Data mining tools can answer questions from institutions that have traditionally taken too long to resolve. Higher education institutions may use the classification technique to perform a complete student characteristics analysis or use estimation and prediction technique to predict the likelihood of different outcomes such as

portability and choice of option, the right career choice, the level and success of the courses [7].

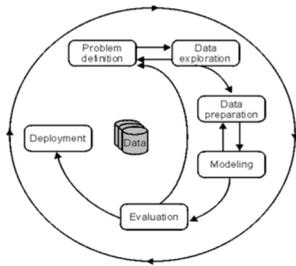


Figure: Data Mining Process

I. DATA MINING TECHNIQUES IN EDUCATION SYSTEM

The exploration of educational data is becoming a research area with a range of computer and psychological methods and research approaches to understanding how students learn. Education Data mining in data mining brings hidden knowledge through the use of various data mining techniques. Some of the data mining techniques are decision tree, clustering, Bayesian classifier, neural network etc.

i. Decision tree:

Decision tree is a kind of decision support tool that uses a tree as a decision graph, including its usefulness. Decision trees are often used in research, especially in decision analysis, to identify strategies and achieve a goal. Another use of decision trees is the calculation of conditional probabilities [6]. Many algorithms used in decision trees are ID3, CART, C4.5, Random tree and CHIAID.

Some of the advantages of a decision tree is as follows [6]:

- Domain knowledge is not required.
- Human understandable concept and user friendly.
- Classification steps of decision tree are very simple and quick.
- Provide a clear mention of which fields are most important for prediction.

ii. Clustering:

Clustering is the attribution process of a lot of objects, making objects in the same cluster which are more similar to each other. Application of Clustering in the education system which can help the institutes of single group classroom student having similar authority. It's a process of assembling a group of abstract objects into classes of parallel objects. All data in groups is partitioned first and then clustered. Then the labels are mentioned in all groups and here the data objects are treated as a separate set. This technique is flexible and allows changes. It also allows the distinction of groups of positive characteristics. The various areas such as data analysis, pattern recognition and image processing focus mainly on grouping techniques [2].

iii. Neural Network:

Recognizing trends and extracting risk patterns by deriving meanings from complex data is a considerable asset for neural networks. One of the main advantages of a neural network is finding a task from given data for preliminary training or experience. Neural networks can identify all possible interactions of variables between different predictors. This is the main reason for using neural networks in the exploration of educational data [4].

iv. Bayesian Classifier:

This is a very simple technique that requires less data preparation to calculate the parameters. Unrelated functions are insensitive and the classifier is well organized to handle real and different data. The conditional class independence between the subsets of the variables of the Bayesian classifier is unique. A graphical model of causality provides a learning process of execution [2].

II. LITERATURE WORK KEEN ON DATA MINING TECHNIQUES IN HIGHER EDUCATION SYSTEM

In [3], K.R.Kavyashree, LakshmiDurga has proposed that a country's growth is strongly measured by the quality of its education system. Education sector has witnessed sea change in its functioning. Today it is recognized as an industry and as an industry it is facing challenges. The challenges of higher education being decrease in students' success rate and their leaving the course without completion. An early prediction of students' failure may help the management provide

timely counselling as well as coaching to increase the success rate and student retention. Data mining are widely used in educational field to find new hidden patterns from student's data which are used to understand the problem. Classification is one of the prediction type classifiers that classifies data based on the training set and uses the pattern to classify a new data. Aim of the project is to develop an internetworking application that uses data mining technique to predict the students' performance based on their behaviour. This paper explored the link between emotional skills of the students along with the socio economic and previous academic performance parameters using Naive Bayes Classifier technique.

In [8] TriptiDwivedi, Diwakar Singh has proposed that the survey of student background history which helps academic planners in institute to give right direction to student. If the class of student is predicted in midsession of institute in final year then it will be easy for the academic planner to plan some important workshop for the enhancement of performance of student which helps it in placement at the end of academic session. In educational institute data mining techniques plays an important role in each activities of institute whether it is academic, cultural, examination and training and placement etc. in which Educational Data Mining which is a field of data mining helps a lot to find the actual filtered data in various field of department in institute. Hidden knowledge through data mining techniques is extracted from large database which helps to predict the pattern in such activities. It plays a great role in predictions of student data for placement and performance.

In [9], Abdulmohsen Algainhas proposed that the data mining techniques are used to extract useful knowledge from raw data. The extracted knowledge is valuable and significantly affects the decision maker. Educational data mining (EDM) is a method for extracting useful information that could potentially affect an organization. The increase of technology use in educational systems has led to the storage of large amounts of student data, which makes it important to use EDM to improve teaching and learning processes. EDM is useful in many different areas including identifying at-risk students, identifying priority learning needs for different groups of students, increasing graduation rates, effectively Assessing institutional performance, maximizing campus resources, and optimizing subject curriculum renewal. This paper surveys the relevant studies in the EDM field and includes the data and methodologies used in those studies.

In [5], Amirah Mohamed Shahiria, Wahidah Husaina and Nuraini Abdul Rashidah has proposed that the Predicting student's performance becomes more Challenging due to the large volume of data in educational databases. Currently in Malaysia, the lack of existing system to analyse and monitor the student progress and performance is not being addressed. There are two main reasons of why this is happening. First, the study on existing prediction methods is still insufficient to identify the most suitable methods for predicting the performance of students in Malaysian institutions. Second is due to the lack of investigations on the factors affecting student's achievements in particular courses within Malaysian context. Therefore, a systematically literature review on predicting student performance by using data mining techniques is proposed to improve students achievements. The main objective of this paper is to provide an overview on the data mining techniques that have been used to predict student's performance. This paper also focuses on how the prediction algorithm can be used to identify the most important attributes in a student's data. We could actually improve student's achievement and success more effectively in an efficient way using educational data mining techniques. It could bring the benefits and impacts to students, educators and academic institutions.

In [10], Heba Mohammed Nagy, Walid Mohamed Aly, Osama Fathy Hegazy has proposed that the educational data mining is a specific data mining field applied to data originating from educational environments, it relies on different approaches to discover hidden knowledge from the available data. Among these approaches are machine learning techniques which are used to build a system that acquires learning from previous data. Machine learning can be applied to solve different regression, classification, clustering and optimization problems. In their research "Student Advisory Framework" that utilizes classification and clustering to build an intelligent system. This system can be used to provide pieces of consultations to a first year university student to pursue a certain education track where he/she will likely succeed in, aiming to decrease the high rate of academic failure among these students. A real case

study in Cairo Higher Institute for Engineering, Computer Science and Management is presented using real dataset collected from 2000–2012. The dataset has two main components: pre-higher education dataset and first year courses results dataset. Results have proved the efficiency of the suggested framework.

In [11], Monika Goyal and Rajan Vohra has proposed that the data analysis plays an important role for decision support irrespective of type of industry like any manufacturing unit and education system. There are many domains in which data mining techniques play an important role. This paper proposes the use of data mining techniques to improve the efficiency of higher education institution. If data mining techniques such as clustering, decision tree and association are applied to higher education processes, it would help to improve students' performance, their life cycle management, selection of courses, to measure their retention rate and the grant fund management of an institution. This is an approach to examine the effect of using data mining techniques in higher education.

In [12], U.K. Pandey and S. Pal has proposed that from ancient period in India, educational institution embarked to use class room teaching. Where a teacher explains the material and students understand and learn the lesson. There is no absolute scale for measuring knowledge but examination score is one scale which shows the performance indicator of students. So it is important that appropriate material is taught but it is vital that while teaching which language is chosen, class notes must be prepared and attendance. This study analyses shows the impact of language on the presence of students in class room. The main idea is to find out the support, confidence and interestingness level for appropriate language and attendance in the classroom. For this purpose association rule is used.

In [13], Dr. Mohd Maqsood Ali has proposed that the Universities either public or private and its colleges enrol thousands of students into various courses or programs every year. They collect information from students at the time of admissions and store the same in computers. Understanding the benefits of data is essential from business point of view. Data can be used for classifying and predicting the students' behaviour, performance, dropouts as well as teachers' performance. Therefore, this paper "Role of data mining in education sector" examines the role of data.

III. IMPLEMENTATION OF DATA MINING TECHNIQUES IN HIGHER EDUCATION

Data Mining has various applications, ranging from marketing and promotion of goods, services or products, to research in the field of artificial intelligence, life sciences and criminal investigations, to high-level government intelligence. Because of its widespread use and the complexity of creating data mining applications, a large number of data mining tools have been developed over the decades. Each tool has its advantages and disadvantages.

Educational Data Mining (EDM) is about developing methods to explore the different types of data coming from the educational environment. We can use these methods to better understand the education system. There are two types of key areas in the EDM. The first is the performance of Mining Student and the second is the mining record data. Data mining techniques such as classification, grouping, and categorization are used to predict student performance, study and visualize data, develop the education sector, and more. The study and visualization of data are used to provide useful information and support in decision making. Two techniques are used for this task: Statistics and Visualization. Statistics include the collection, analysis, interpretation and presentation of data. Visualization uses graphical techniques to better understand and analyse data in the education system [6].

IV. CONCLUSION

The current education system does not include any prediction of student achievement because of its performance. The system does not process abandonment. In the current situation, the student is not analysed. Another common problem in universities is that some students feel lost in the mass. They may have trouble finding help for classes, or have trouble choosing the right courses they need.

The current education system does not predict whether a percentage of students has failed or succeeded because of its performance. The system does not process abandonment. Nowadays, students' academic success is the main topic of management in all professional institutes.

Improving student performance with additional guidance and guidance from the beginning will help management take timely action to reduce the percentage of poor people Appearance of students. Classification and grouping operations can predict more accurate results to improve the success rate of Students in universities. In future we can enhance the data mining techniques and tools for finding out the passing and failing percentage rate and effective gross enrolment value of universities.

REFERENCES

1. Sunil Kumar Tiwari, Archi Dubey, Concise Discussion on Growth of Higher Education in India: PARIPEX, Volume-8 | Issue-5 | May-2019 | PRINT ISSN No. 2250 - 1991
2. T.Thilagaraj, Dr.N Sengottaiyan, A Review of Educational Data Mining in Higher Education System : Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering pp. 349–358 DOI: 10.15439/2017R87 ACSIS, Vol. 10 ISSN 2300-5963
3. K. R. Kavyashree, Lakshmi Durga – “A Review on Mining Students Data for Performance Prediction” - International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 4, April 2016.
4. G. Gray, C. McGuinness, P. Owende – “An Application of Classification models to predict learner progression in tertiary education” - Advance Computing Conference (IACC), 2014 IEEE International, IEEE, 2014, pp. 549–554.
5. Amira Mohamed Shahria, Wahidah Husaina and Nuraini Abdul Rashida – “A Review on Predicting Students performance using Data Mining Techniques” - The Third Information Systems International Conference, Procedia Computer Science 72 (2015) 414 – 422.
6. Khushbu N. Shah, Misika R. Patel, Neha V. Trivedi, Priyanka N. Gadariya, Richa H. Shah, Ms. Nehal Adhvaryu, Study of Data Mining in Higher Education - A Review, International Journal of Computer Science and Information Technologies, Vol. 6 (1) , 2015, 455-458.
7. Ajinkya Kunjir, Poonam Pardeshi, Shrinik Doshi, Karan Naik, Recommendation of Data Mining Technique in Higher Education, IICER, ISSN (e): 2250 – 3005 Volume, 05 Issue, 03 March – 2015 .
8. Tripti Dwivedi, Diwakar Singh – “Analyzing Educational Data through EDM Process: A Survey” - International Journal of Computer Applications (0975 – 8887) Volume 136 – No. 5, February 2016.
9. Abdulmohsen Algarni - “Data Mining in Education” - (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 6, 2016
10. Heba Mohammed Nagy, Walid Mohamed Aly, Osama Fathy Hegazy – “An Educational Data Mining System for Advising Higher Education Students” - International journal of Computer, Electrical, Automation, Control and Information Engineering Vol:7, No:10, 2013.
11. Monika Goyal and Rajan Vohra – “Applications of Data Mining in Higher Education” - IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012.
12. U. K. Pandey and S. Pal – “A Data mining view on class room teaching language” International Journal of Computer Science Issue, Vol. 8, Issue 2, pp. 277-282, ISSN:1694-0814, 2011.
13. Dr. Mohd Maqsood Ali - “Role of data mining in education sector” - International Journal of Computer Science and Mobile Computing Vol. 2, Issue. 4, April 2013.
14. Morgan Kaufmann Publishers An Imprint of Elsevier Science “Data Mining Concepts and Techniques” ISBN 81-7867-023-2, Chapter 7, p.279 Date:11/12/2014