COVID-19: A survey on change in learning method from classroom to virtual using Educational Data Mining

Sulakshana Kishor Vispute[1], Dr. Jitendra Janardan Ahirrao[2], Dr. Shubhangi M. Potdar[3]

[1]*Assistant Professor, MCA department, DES's NMITD, University of Mumbai, India*
[2]*Associate Professor, Department of Commerce, Smt. Dankunwar Mahila Mahavidyalay Jalana*
[3]*Assistant Professor, MCA department, DVVPF's Institute of Business Management & Rural Development, Ahmednagar.*

Abstract

The process of extracting hidden predictive information from a large data set is called data mining and it is also referred to as Knowledge Discovery from Database i.e. KDD. Time series analysis, prediction, regression, clustering, association rule mining, classification are few data mining techniques used by many sectors for extracting hidden patterns from their huge data warehouses. Educational data mining is an emerging field where data mining techniques are applied to educational data. Corona virus is declared as a pandemic and most of the countries have declared lockdown as a preventive measure against it. During this lockdown, most of the higher education institutes organized online lectures for their students. This paper surveys to understand what kind of teaching students preferred in the future either online or classroom for which data mining classification technique was used with the J48 algorithm. And the result shows that students preferred classroom teaching instead of online teaching.

Keywords: Corona virus, COVID-19, Data mining, Educational data mining, Classification, Decision tree, J48 etc.

Introduction

Corona virus (COVID-19) is declared as a pandemic and the whole world is fighting against it. COVID-19 is a highly infectious disease which can easily spread from human to human. In India, the first case was reported in January 2020 and till the date; more than 1 Lakh cases were reported. Government of India declared the first lockdown on 24 March 2020, before that all schools and colleges were temporarily shutdowns from 16th of March 2020 as a preventive measure against the COVID-19 pandemic. This pandemic significantly affects the education sector ranging from nursery school to higher education. This is a very crucial time for the education sector because most of the exams like 10th, 12th board examinations, the entrance test for various universities, PG courses and competitive exams are held during this period. It also affects pre-primary, primary, and secondary school exams and admission process. Most of the higher education institutes start online teaching during this lockdown period for the completion of their pending syllabus. This lockdown period changed the way of teaching methods from the classroom to online or virtual teaching. Students are also experiencing this new way of teaching i.e. online teaching.

Data mining is an emerging field among researchers because of its characteristics like it can handle complex data very efficiently as well as it can extract hidden predictive information from a huge database and many more. Many sectors like banking, telecommunication, E-commerce, etc. are using data mining for extracting desired knowledge which would assist them in decision making and business forecasting. The educational sector also uses Data Mining for different applications like students' profiling, students' retention, performance prediction, course management, enrollment management, recommender system, etc. When data mining applied to educational data it is referred to be Educational Data mining i.e. EDM. The main objective of this paper is to understand what is the students' feedback or opinion about online teaching and classroom teaching by applying a data mining classification technique using a decision tree algorithm. A questionnaire has been designed for this survey to understand which form of teaching students' would prefer in future virtual or online teaching or classroom teaching.

Data Mining: The mining of data means extracting novel but the valid discovery of patterns or desired information from a huge data set. It is also known as KDD ie. Knowledge discovery from the database. Decision-makers can use this useful information in designing plans for their organization. Clustering, association rule mining, regression, time series analysis, classification, prediction are few data mining techniques.

Clustering: Clustering is an unsupervised technique use to divide the collection of objects into groups or clusters such that one group or cluster must have a collection of objects having the same features.

Association Rule Mining: Association rule mining is used to finding out significant dependencies or relationships among different variables.

Regression: Regression is a mathematical based data mining technique that helps in identifying correlation among different variables.

Time series analysis: It is used to determine similarities among values, classify them based on their behavior, and predict future values based on past values eg. Stock exchange.

Classification: Classification is a supervised technique is used to classify different objects in predefined classes. One class can have a collection of objects which share the same characteristics.

Decision tree: It builds the classification model in the form of a tree structure. The decision tree contains a root, intermediate node, and leaf nodes. Root and intermediate nodes represent attributes while leaf nodes are class labels. The arcs from an internal node to its sub node are labeled with distinct outcomes of the test. It breaks down a training dataset into smaller subsets while at the same time an associated decision tree is incrementally developed.

ID3: ID3 is a core algorithm used to build a decision tree. It follows the top-down approach and uses Entropy or information gain for building a decision tree.

Information Gain formulae:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Entropy formulae:

$$Entropy = \sum_{j} - p_j \log_2 p_j$$

Jai Ruby & David (2014) discusses REPTree, simple CART, ID3, BFTree, J48; MLP algorithms used to analyzed and predict the academic performance of students. The author surveyed 165 students' of Arts and Science colleges and on that data classification algorithms were applied and in results, it shows that MLP gives more accurate results than other algorithms and ID3 is the second-highest in the accuracy. Mital and Zaman (2016) discussed the basics of Data Mining, it's architecture, techniques, processes, tasks, and challenges faced by it. along with this the author also discusses applications of data mining in different sectors like E-commerce, telecommunication, finance, medical, education, etc. It is also helpful in customer relationship management. Sulakshana (2018) briefs about data mining techniques like classification, time series analysis, prediction, clustering, etc, the process of EDM, different EDM modules, data mining, and educational data mining applications in different sectors, tools used by EDM, types of data on which EDM works, etc. Sulakshana & Ahirrao (2020) conducted a survey to understand the pattern of the second year MCA students while selecting elective subjects. For this decision tree algorithm was applied on collected data using the WEKA tool and the result shows the ID3 algorithm gives good results than other algorithms in terms of accuracy.

Problem Statement

The COVID-19 pandemic is the first and foremost health crisis due to which most of the countries announced the lockdown as a preventive measure against this pandemic. From mid of March 2020, the Indian government also closes schools, colleges, and universities for precaution. The structure of learning, teaching, and assessment methodologies were the first to be affected by these closures. This period is a very crucial time for the higher education sector because half of the syllabus is yet to be completed. As a solution, most of the institutes adopt a new way of teaching ie. Online/virtual teaching through which teachers can complete their syllabus. Due to this pandemic situation students' way of learning is suddenly changing from chalk duster to online and every institute needs to know students' opinion about this change so during this lockdown period and post COVID-19 they can make necessary changes in an existing system and in the future, they can give best services to their student.

Data Preparations

The data set used for this study was obtained from the MCA department of Navinchandra Mehta Institute of Technology (Mumbai). The initial size of data is 80 for session 2019-20. It is very important to improve the performance of online teaching so the institute would attract more and better students in the future.

The questionnaire and its expected answers are listed below

Table1: Questionnaire and Expected Answers

Sr. No	Question	Possible values/ answers
1.	Do your college scheduled online lectures for all subjects?	Yes / No
2	Do you attend all scheduled online lectures?	Yes / No
3	Online lectures are good to understand Theoretical subjects?	Yes / No
4	Online lectures are good to understand Technical/Practical subjects?	Yes / No
5	Do online lectures increase your interest in subjects?	Yes / No
6	In the future which forms of teaching you preferred?	Online teaching / Classroom teaching

Model Construction:

There are many open source tools are available for data mining like Rapid miner, Orange, KNIME, WEKA, etc. Out of which WEKA explorer is a very popular tool because of its user-friendly GUI. It covers major data mining packages like preprocessing, classification, clustering, association, etc. Primarily it was designed to analyze data and prediction for the agricultural sector but nowadays it is used for a variety of applications areas ranging from medical, E-commerce, telecommunication to the education sector. This software is JAVA based software so that it is free and portable to use. It can efficiently handle different data mining tasks like preprocessing, clustering, classification, regression, and many more. For the above questionnaire, 60 samples are collected in the CSV file named feedback_onlineteaching.csv. This file is loaded in WEKA explorer's classify tab, where more than 15 decision tree algorithms are available like J48, ID3, CART, etc. out of which the present study uses the J48 classification algorithm.

Results Obtained

Table2: Collected Responses

Sr. No	Questions	Responses	
1	Do your college scheduled online lectures for all subjects?	Yes(100%)	No(0%)
2	Do you attend all scheduled online lectures?	Yes(80%)	No(20%)
3	Online lectures are good to understand Theoretical subjects?	Yes(63%)	No(37%)
4	Online lectures are good to understand Technical/Practical subjects?	Yes(44%)	No(56%)
5	Do online lectures increase your interest in subjects?	Yes(10%)	No(90%)
6	In the future which forms of teaching you preferred?	Online Teaching (4%)	Classroom teaching (96%)

Findings

The result shows online or virtual teaching cannot increase students' interest in their subjects. For the future, only 4 per cent of students select virtual teaching and 96 per cent are comfortable with classroom teaching. For finding out the reasons behind this after taking interview of a few students came across with few challenges faced by students while attending online lectures,

1. A proper internet connection is not available.

2. Sometimes teachers are not audible.

3. Some students do not have a laptop.

4. They are not able to connect on time.

5. Sometimes they are disconnected due to internet issues.

6. Virtual learning of a theoretical subject is boring.

7. It is difficult to understand practical concepts through online teaching.

8. They could not pay attention due to disturbing ambiance and mental stress.

Conclusion

This study shows that even though few students found online or virtual teaching suitable for theoretical subjects, few of them found it suitable for technical or practical subjects, it also helps increase the interest of students in few subjects still students feel that in future classroom teaching is always preferred than virtual teaching. Unavailability of proper gadgets, internet speed, and lack of attention are some major reasons responsible for this. When the J48 algorithm is applied to skewed data it will not give the desired result. Instead of using different online apps, every institute should put in place its own LMS so that E-data will be more secured. Online teaching should continue as a supplement to classroom teaching to reinforce learning.

References:

Sulakshana Vispute, D. J. (2020). Use of Data Mining Techniques on Educational Data for Prediction of Elective Selections. NCMAT.

vispute, S. (2018). A study of Educational Data Mining(EDM) and NCMAT.

Zaman, S. M. (2016). A review of Data Mining Literature. International Journal of Computer Science and Information ecurity Vol. 14, No. 11.

https://en.wikipedia.org/wiki/COVID-19_pandemic_in_India

https://government.economictimes.indiatimes.com/news/education/covid-19-pandemic-impact-and-strategies-for-education-sector-in-india/75173099

https://voxeu.org/article/impact-covid-19-education

https://www.indiatoday.in/education-today/featurephilia/story/effect-of-covid-19-on-campus-steps-taken-by-colleges-1668156-2020-04-17

https://www.jagranjosh.com/general-knowledge/impact-of-coronavirus-on-education-in-india-1587642880-1

https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses