STM JOURNALS

Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 9, Issue 2, 2022
DOI Journal: 10.37591/JoAIRA

http://computers.stmjournals.com/index.php?journal=JoAIRA&page=index

Review JoAIRA

Implementation of Data Mining Approach to Find the Adaptability of Students in Online Education During COVID-19 Pandemic

Trailokya Raj Ojha*

Abstract

The global COVID-19 pandemic has severely affected every aspect of human life, including education. The virus' stunning spread created havoc in the educational system, causing educational institutions to close. As an effect, students must quickly adopt to the change to synchronous online learning. This study identified the different aspects affecting the adaptability level of students in online class. It also identifies the student's adaptability level in different circumstances in online classes. A sample of 1205 school, college and university students from Bangladesh has been examined using data mining technique to find their adaptability level. The result shows that college and university students have more adaptability than the school students.

Keywords: Online learning, adaptability, COVID-19, K-means algorithm, Kaggle repository, educational data mining

INTRODUCTION

The pandemic spread of Novel Corona virus, also known COVID-19, has significantly disrupted our daily life, including education as well. As an effect, all the educational institutions were closed, and online classes were started as an alternative teaching approach. For most of the educational institutions, resource persons and students, online education was new thing. It was totally a new thing, so it was difficult to adopt the system as traditional face to face teaching learning approach. The student's potential for using online tools, their technological capacity for accessing online courses, and the method in which the instructors handle learning activities all influence how well they will adapt to online education and how they will respond to it.

Although schools, colleges, and universities were closed indefinitely, both academic institutions and students experimented with various study methods to finish their coursework within the deadlines established by the academic calendar. These measures undoubtedly caused some inconvenience, but they also gave rise to fresh instances of educational innovation that utilized digital interventions. However, COVID-19 has inspired educational institutions all across the world to put inventive ways to practice quickly [1]. But it had many challenges including adaption of the system from both students and resource person's side. In this study, the adaptability level of students in online classes in different circumstances has been analyzed.

Different predictive models such as association, classification and clustering have been used to

***Author for Correspondence**
Trailokya Raj Ojha
E-mail: trailokyaro@nec.edu.np

Assistant Professor, Department of Computer Science Engineering, Nepal Engineering College, Bhaktapur, Nepal

Received Date: July 26, 2022
Accepted Date: August 09, 2022
Published Date: August 16, 2022

Citation: Trailokya Raj Ojha. Implementation of Data Mining Approach to Find the Adaptability of Students in Online Education During COVID-19 Pandemic. Journal of Artificial Intelligence Research & Advances. 2022; 9(2): 1–8p.

predict the adaptability of students in online class. This study's accomplishment will assist in the development of various methods for estimating students' capacity for online learning. In this study, Apriori, J48 and K-Means algorithm were implemented in WEKA to predict the result. Data was obtained from the Kaggle source. The accuracy measure that represents the percentage of correctly classified instances is used for judging the performance of the attributes. The results of this study can help in deciding the environment of learning in online platform to encourage effective learning. A brief overview of the literature is followed by an explanation of the methodology and the data in the section that follows. Then, the obtained result has been analyzed and discussed followed by concluding remarks.

LITERATURE REVIEW

Data mining is also known as knowledge discovery in database which is the area of uncovering new knowledge from the large database [2]. For the mining process, data mining uses statistics, machine learning, and artificial intelligence [3]. Educational data mining or data mining in education is the application area of data mining research used for gathering and analyzing data stored in e-learning system to find the students useful information and knowledge [4]. The newly growing field of "educational data mining" is concerned with creating tools for examining student data.

The COVID-19 pandemic has resulted in the temporary suspension of physical classroom instruction. Students were put in an unusual predicament as a result, making it difficult for them to see the future clearly.

During isolation due to COVID-19, all the educational institutes conducted online classes; so, various devices like laptop (35.83%)/smartphone (57.98%)/tablet (4.89%)/desktop (0.65%) and internet access were used for e-learning. The mobile data pack was the source of internet for 82% of students [1, 5]. Mostly students (62%) prefer WhatsApp for communication [5].

The rapid and forced switch from traditional to online education has had a negative effect on high school and university students' readiness in the context of the COVID-19 pandemic [6]. These consequences are caused by problems with technology access and internet networks, as well as by reduced instruction quality.

During the isolation phase in Romania, parents were responsible for ensuring their children's access to and participation in the online learning opportunities set up by the country's educational institutions. To participate in online classes, you will need to have a computer, a phone, a tablet, and internet access. The families or educational institutions of the students may be the owners of this equipment. The whole responsibility for any damage to borrowed technology and communication equipment is with the parent or legal guardian. Platforms, online instructional materials, and virtual libraries can all be accessed for free [7].

According to Kebritchi *et al*., prior technical skill training is required for effective use of computers and the internet to support students in online education [8]. The students' impression of and attitude toward the internet, their proficiency with the English language, and their time management skills are additional crucial considerations. The Hung *et al*. indicators of the success of online education include factors like student control, self-directed learning, computer and internet quality, and the effectiveness of online communication [9].

Regarding student age, some research revealed no relationship between age and online learning contentment or performance [10], while others found that older students are more likely than younger ones to finish online courses [11, 12]. For instance, in one research of online learning [11], successful students had an average age of 28, compared to unsuccessful students who had an average age of 25. Senior students may perform better in online learning because they have higher levels of rehearsal,

Journal of Artificial Intelligence Research & Advances
Volume 9, Issue 2
ISSN: 2395-6720

elaboration, critical thinking, and metacognitive self-regulation as they get older, according to Colorado and Eberle [13]. Each of these factors may help older students succeed in online courses.

METHODOLOGY

To determine the efficacy of online education during the COVID-19 epidemic, this study uses the machine learning method to build a prediction model. The data set obtained from Kaggle repository were examined to find the possible and best relationship between them. The subsequent stage involved preprocessing and data modification using data mining techniques. Following the implementation of several data mining techniques including association, clustering, and classification on those data, the knowledge representation of the outcome was used. The working methodology of the research is shown in Figure 1.

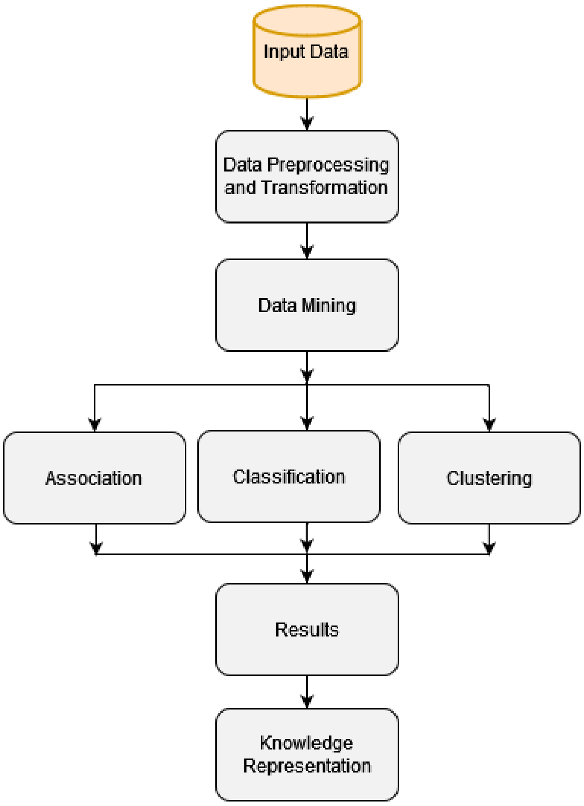

Figure 1. Working methodology.

Data Description

This section is divided into three subsections that include data descriptions, algorithms, and implementation techniques.

The data set used in this study was obtained from Kaggle repository. It contains the record of 1205 students from different universities, colleges, and schools of Bangladesh from December 10, 2020, to February 5, 2021. In this study we are using only 10 attributes among 14 attributes of the data set. One of them is the response variable, and the remaining nine are predictor variables. The specific details of these variables are given below:

- *Gender:* This attribute depicts the gender of the students. Boy and girl are the two possible values for the attribute.
- *Age*: This attribute depicts the age of the students. The possible values for the attribute are 1–5, 6–10, 11–15, 16–20, 21–25 and 26–30 years.
- *Education level:* It depicts the educational level of students. The possible values for the attribute are school, college and university.

- *IT Student:* This attribute shows that whether the student is IT student or not. Yes and no are the possible values for the attribute.
- *Load-shedding:* It depicts the load-shedding condition during online class. Low and high are the possible values for the attribute.
- *Internet type:* It represents what type of internet is being used by students during online class. Wi-Fi and mobile data are two possibilities for the attribute's possible values.
- *Class duration:* This attribute indicates the daily class duration. It has three values that are 0, 1–3 and 3–6.
- *Self LMS:* The attributes depict whether the institution has its own LMS or not. The attribute has two possible values either true or false.
- *Device*: This attribute shows what kind of device the students are using for online class. Mobile, computer and tab are the possible values for the attribute.
- *Adaptability level:* This is the data set's primary response variable. It depicts how much adaptive the students are to the online class. There are three characteristics: low, moderate, and high. The attributes of datasets with possible values are depicted in Table 1.

Table 1. Description of attributes.

Attribute	**Variable type**	**Possible value**
Gender	Predictor Variable	Boy Girl
Age (years)	Predictor Variable	1 to 5, 6 to 10, 11 to 15, 16 to 20, 21 to 25 and 26 to 30
Education Level	Predictor Variable	School, College, University
IT Student	Predictor Variable	Yes No
Load-shedding	Predictor Variable	Low High
Internet Type	Predictor Variable	Wi-Fi Mobile Data
Class Duration (hr)	Predictor Variable	0, 1 to 3, 3 to 6
Self LMS	Predictor Variable	Yes No
Device	Predictor Variable	Mobile Tablet Computer
Adaptability Level	Response Variable	Low Moderate high

Algorithm Description

The algorithms used in this study are Apriori, J48 and K-Means algorithm which are described in this section.

Apriori Algorithm

The apriori algorithm is a set of procedures to be used in order to identify the most frequent item set in a given database. The data mining method repeatedly performs the join and prune processes until the most frequent item set is obtained. The apriori algorithm's primary objective is to establish an association rule between different things. The correlation between two or more elements are explained by the association rule.

Item set X directly makes up the first candidate item set Y1 in the initial iteration. Assume that if X={x1, x2,….., xn}, then Y1={{x1},{x2},...,{xn}}. First of all, the candidate item set Y_k for the Kth iteration arises in accordance with the frequent item set L_{k-1} discovered in the previous iteration. Then distribute a counter with an initial value of zero to each item in the set, and scan database Z in the correct chronological order. As long as every activity is included in each item set, the counter for those item sets will rise. In accordance with the actual value of |Z| and the minimal support level of a specific Y_k of the common item set, the support level can be determined after all affairs have been scanned. Continue until no more mew items appear [14].

J48 Algorithm

A decision tree has several advantages for data mining and makes it simple for users to understand and execute. With incomplete or incorrect datasets, it can nevertheless produce improved predictions. J48 creates the decision tree using the idea of knowledge gain. The J48 algorithm divides the dataset into smaller subsets by using each attribute's choice as a function of information gain. The splitting process comes to an end when all instances in the subsets belong to the same class, which is determined by the attribute with the biggest information gain. On the other hand, J48 can manage both continuous and discrete attributes when features are not given any information gain [15].

K-means Algorithm

K-means algorithm is used for clustering which divides the data into groups of similar objects. Clustering helps to identify the weight of different values of attributes. By using a straightforward matching dissimilarity measure for categorical objects, modes rather than cluster means, and a frequency-based mechanism to update modes during the clustering process, the k-means algorithm is able to minimize the clustering cost function. This can be defined as follows: given a set X={X1, .. , Xn} of n numerical data objects, a natural number $k \leq n$, and a distance measured, the k-means algorithm aims at finding a partition Y of X into k non-empty disjoint clusters Y1, . . . , Yk with $Yi \cap Yj = \emptyset$ and $\bigcup k$ i=1 Yi=X such that the overall sum of the squared distances between data objects and their cluster centers is minimized [16].

Implementation Techniques

Figure 1 depicts the graphical representation of the stepwise methods for forecasting the result of the research. The details of the procedure are explained below:

1. *Input Data*: After collecting data from Kaggle repository, the data set has been used as an input for the system.
2. *Data Preprocessing and Transformation*: It is the technique of transforming raw data into understandable format. Real data is frequently erroneous, inconsistent, and frequently incomplete. Preparing raw data to be acceptable for a machine learning model is known as data preprocessing. It uses a variety of transformation techniques, including data cleaning, integration, reduction, transformation, and discretization.
3. *Data Mining*: It is a computer process that uses techniques from the junction of artificial intelligence, machine learning, statistics, and database systems to find patterns in massive data sets. We used data mining techniques such as association, clustering, and classification in this study. The Apriori algorithm is used to find the association rules, the J48 classifier algorithm for classification, and the K-Means algorithm for clustering.
4. *Result*: This section displayed the findings from the data mining approaches.
5. *Knowledge Representation*: This section showed the information and knowledge that had been extracted along with an explanation.

RESULTS

The experiment was conducted using data set from Kaggle repository which contains the record of 1205 students from different universities, colleges and schools of Bangladesh. To understand the student's adaptability level and factors affecting online education system, educational data mining technique was used. Such technique includes association, classification and clustering. Educational data mining can be used for multiple purposes such as students' academic performance prediction, their adoption level in online classes, result analysis and classification of student's performance [17].

Association Rule

In transaction databases or other data repositories, association rule mining seeks to extract intriguing correlations, recurrent patterns, relationships, or casual structures among groupings of items. Apriori algorithm was used in WEKA to create the association rule model. Figure 2 depicts the process's result.

```
Best rules found:

1. IT Student=No Device=Mobile 834 ==> Self Lms=No 777    <conf:(0.93)> lift:(1.13) lev:(0.07) [88] conv:(2.51)
2. IT Student=No Self Lms=No 834 ==> Device=Mobile 777    <conf:(0.93)> lift:(1.11) lev:(0.06) [75] conv:(2.29)
3. IT Student=No 901 ==> Self Lms=No 834    <conf:(0.93)> lift:(1.12) lev:(0.07) [90] conv:(2.31)
4. IT Student=No 901 ==> Device=Mobile 834    <conf:(0.93)> lift:(1.1) lev:(0.06) [76] conv:(2.11)
5. IT Student=No Load-shedding=Low 787 ==> Self Lms=No 727    <conf:(0.92)> lift:(1.12) lev:(0.06) [77] conv:(2.25)
```

Figure 2. Result of the association rules.

After applying apriori algorithm in WEKA, five best rules were found which are explained as:

- Rule 1: If the student is non-IT student and is using mobile device as a tool for online class, it is likely not to have self LMS in his/her school.
- Rule 2: If the student is non-IT student and does not have self LMS, they are likely to use mobile device for online classes.
- Rule 3: If the student is non-IT student, it is likely to have no self LMS.
- Rule 4: Non-IT students seem to use mobile as online study device.
- Rule 5: If the student is not an IT student and has low load shedding, they do not have self LMS.

Classification

Building a model that can categorize a class of objects in order to predict the classification or missing attribute value of future objects is known as classification. The classification result of 1205 students obtained using J48 classifier in WEKA is shown in Figure 3.

```
=== Summary ===

Correctly Classified Instances         947               78.5892 %
Incorrectly Classified Instances       258               21.4108 %
Kappa statistic                          0.61
Mean absolute error                      0.1946
Root mean squared error                  0.322
Relative absolute error                 51.5957 %
Root relative squared error             74.166  %
Total Number of Instances             1205

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.862    0.255    0.785      0.862    0.822      0.613    0.874     0.883     Moderate
                 0.767    0.123    0.805      0.767    0.785      0.650    0.894     0.846     Low
                 0.400    0.019    0.656      0.400    0.497      0.479    0.883     0.543     High
Weighted Avg.    0.786    0.183    0.782      0.786    0.780      0.616    0.883     0.840

=== Confusion Matrix ===

   a   b   c   <-- classified as
 539  77   9 |   a = Moderate
 100 368  12 |   b = Low
  48  12  40 |   c = High
```

Figure 3. Classification result.

The confusion matrix from the outcome provides a solution to the classification problem as 539 students were correctly classified as, 86 students were incorrectly classified as moderate adaptability level. 368 students were correctly classified and 112 students were incorrectly classified as low adaptability level. And 40 students were correctly classified and 60 students were incorrectly classified as high adaptability level.

Journal of Artificial Intelligence Research & Advances
Volume 9, Issue 2
ISSN: 2395-6720

Clustering

Finding groupings of objects that are similar to one another and distinct from other groups of objects is the process of clustering. The K-Means algorithm is used for clustering in this study. It helped to find the adaptability level of students in online classes. The result obtained from using K-Means algorithm in WEKA is depicted in Figure 4. The result shows that 558 boys, which is 46% of total students, have moderate adaptability, whereas 647 girls, that is 54% of total students, have low adaptability in online class. It shows that school girls of the age group 11–15 years have low adaptability level in online classes. The cluster visualization is shown in Figure 5.

```
Clusterer output
Cluster 0: Boy,26-30,University,Yes,Low,'Mobile Data',1-3,No,Mobile,Moderate
Cluster 1: Boy,11-15,School,No,Low,'Mobile Data',1-3,No,Mobile,Low

Missing values globally replaced with mean/mode

Final cluster centroids:
                                  Cluster#
Attribute            Full Data           0           1
                      (1205.0)     (558.0)     (647.0)
======================================================
Gender                     Boy         Boy        Girl
Age                      21-25       21-25       11-15
Education Level         School  University      School
IT Student                  No         Yes          No
Load-shedding              Low         Low         Low
Internet Type      Mobile Data        Wifi Mobile Data
Class Duration             1-3         1-3         1-3
Self Lms                    No          No          No
Device                  Mobile      Mobile      Mobile
Adaptivity Level      Moderate    Moderate         Low




Time taken to build model (full training data) : 0.07 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       558 ( 46%)
1       647 ( 54%)
```

Figure 4. Clustering result.

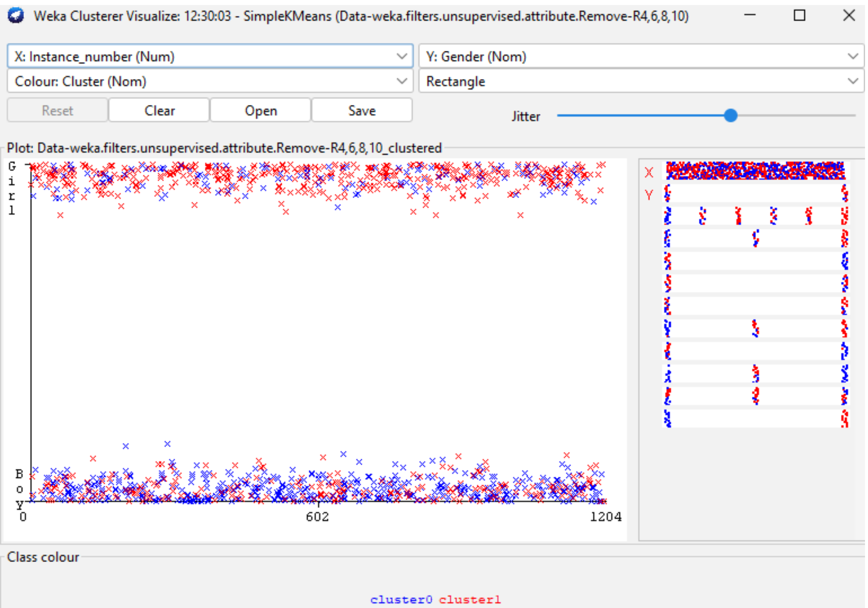

Figure 5. Cluster visualization.

CONCLUSION

The study has been conducted using nine different classifiers to predict the student's adaptability level in online classes during global COVID-19 pandemic. Different data mining techniques were used to achieve the result. To understand the student's adaptability level and factors affecting online education system, educational data mining technique was used. Such technique includes association, classification, and clustering. From the observation, it can be confirmed that the male students of the age group 21–25 years have moderate adaptability in online class compared to the other age group students. And it is also concluded that the female students of age group 11–15 years have the low adaptability in online class. The result of the study also shows that the adaptability level of students in online class is affected by different factors such as load shedding, device used to take online classes, availability and type of internet and age group.

REFERENCES

1. Butnaru GI, Niță V, Anichiti A, Brînză G. The effectiveness of online education during COVID 19 pandemic: a comparative analysis between the perceptions of academic students and high school students from Romania. Sustainability. 2021 May 10; 13(9): 5311.
2. Baker RS, Yacef K. The state of educational data mining in 2009: A review and future visions. Journal of Educational Data Mining (JEDM). 2009 Oct 1; 1(1): 3–17.
3. Abd Elaal SA. E-learning using data mining. Chin-Egypt Res J. 2013.
4. Baepler P, Murdoch CJ. Academic analytics and data mining in higher education. International Journal for the Scholarship of Teaching & Learning (IJ-SoTL). 2010 Jul 1; 4(2): Article 17.
5. Muthu prasad T, Aiswarya S, Aditya KS, Jha GK. Students' perception and preference for online education in India during COVID-19 pandemic. Social Sciences & Humanities Open (SSHO). 2021 Jan 1; 3(1): 100101.
6. Dorn E, Hancock B, Sarakatsannis J, Viruleg E. COVID-19 and student learning in the United States: The hurt could last a lifetime. McKinsey & Company. 2020 Jun 1; 1–9.
7. UNESCO. Distance Learning Solutions, More on UNESCO's COVID-19 Education Response [Online]. Available from https://en.unesco.org/covid19/educationresponse/solutions
8. Kebritchi M, Lipschuetz A, Santiague L. Issues and challenges for teaching successful online courses in higher education: A literature review. J Educ Technol Syst. 2017 Sep; 46(1): 4–29.
9. Hung ML, Chou C, Chen CH, Own ZY. Learner readiness for online learning: Scale development and student perceptions. Comput Educ. 2010 Nov 1; 55(3): 1080–90.
10. Biner PM, Summers M, Dean RS, Bink ML, Anderson JL, Gelder BC. Student satisfaction with interactive telecourses as a function of demographic variables and prior telecourse experience1. Distance Educ. 1996 Jan 1; 17(1): 33–43.
11. Dille B, Mezack M. Identifying predictors of high risk among community college telecourse student. Am J Distance Educ. 1991 Jan 1; 5(1): 24–35.
12. Wojciechowski A, Palmer LB. Individual student characteristics: Can any be predictors of success in online classes. Online Journal of Distance Learning Administration (OJDLA). 2005 Nov; 8(2): 13.
13. Colorado JT, Eberle J. Student demographics and success in online learning environments. Emporia State Research Studies (ESRS). 2010; 46(1): 4–10.
14. Yabing J. Research of an improved apriori algorithm in data mining association rules. International Journal of Computer and Communication Engineering (IJCCE). 2013 Jan 1; 2(1): 25–27.
15. Onik AR, Haq NF, Alam L, Mamun TI. An analytical comparison on filter feature extraction method in data mining using J48 classifier. Int J Comput Appl. 2015; 124(13): 1–8.
16. San OM, Huynh VN, Nakamori Y. An alternative extension of the k-means algorithm for clustering categorical data. Int J Appl Math Comput Sci. 2004; 14(2): 241–7.
17. Prabha SL, Shanavas ARM. Application of Educational Data mining techniques in e-Learning-A Case Study. Int J Comput Sci Inf Technol. 2015; 6(5): 4440–4443. [Online]. Available: www.ijcsit.com

