A Study on the Effectiveness of Online Learning During Covid 19 Pandemic

Sebastian Jose[1], Aleena Thomas[2], Asst.Prof. Ambily Merlin Kuruvilla[3]
[1]BCA Final Year Student, SAINTGITS College of Applied Sciences, Kottayam, Kerala, India
[2]BCA Final Year Student, SAINTGITS College of Applied Sciences, Kottayam, Kerala, India
[3]Asst. Professor Department of Computer Applications, SAINTGITS College of Applied Sciences, Kottayam, Kerala, India

Abstract : *Over the past decade there is a rapid growth in most of the sectors. Many of the industries has been come up from both private and public sector offering graduate and postgraduates with certifications. One of the main motive of education institutions is to improve individual performance thereby increasing their overall performance of the institution. Educational Data Mining has been very successful in achieving and predicting those conclusions in order to improve their performance. But most of the educational institution fails to give quality education due to the lack of education information and poor ways to create efficient learning. They never utilize the new technology and techniques to improve learning and only relay on traditional education.*

*As Covid-19 has made people to adapt into online learning in the place of traditional learning, it has been a dilemma for the educators to predict the efficiency of online learning. Learning Management System (**LMS**) creates same environment as the offline learning. It has some advantages and disadvantages too. Through this paper we try to predict the effectiveness of online learning through the LMS. These LMS are implemented to conduct classes on regular basis. Data Mining uses classification techniques such as J48, Random Forest, MLP, Bagging etc to conduct the analysis.*

Keywords : **Data mining, Educational data mining (EDM), Classification, J48, Random Forest, bagging, multilayer perceptron (MLP), Learning Management System (LMS).**

1. Introduction

Education is a process of providing learning or the acquisition of imparting knowledge to produce moral belief and skills in an individual. The development of a country depends on the education acquired by the citizen. India is a developing country and most of the states in India have high literacy rate.

The world is fighting an invisible threat called Covid-19 [1] which has made all the sector from industrial to education sector in a fragile way. So it was been a challenging mission for the education experts to overcome this condition. The learning management system (**LMS**) has been promoted for enhancement of learning which made into a system called e-learning or online learning. There are cons as well as pros for virtual learning as compared to traditional(offline) mode of learning. One of the cons regarding the virtual learning platform is that the difficulty in learning practical sessions and the comprehension of effectiveness of learning process in students. The pros include covering of the syllabus without offline mode of learning. It has helped in managing time. We became more aware about the new virtual tools for learning such as Skype, Teams, Google Meet, Zoom, WebEx, Google Drive, Dropbox, Basecamp etc which has been more popularized during the span of Covid-19 pandemic. It has added flexibility and self-motivation in learning. Due to the online learning students will have an opportunity to create network around the globe. This helped them to broaden their culture and views.

Data mining is a very promising area for making these types of decisions called Educational Data Mining or **EDM.** It is important for us to use the medium to improve the quality of education to improve the efficiency of learning. We use many classification techniques and algorithms such J48, Random Forest, MLP, Bagging etc for finding the efficiency, which is one of the advantage of EDM.

2. Proposed Methodology

As we can use data mining to predict the performance of students, a survey was taken from a group of students to predict the efficiency of online learning during Covid-19 pandemic . Questions were prepared based on the criteria about their resources, availability of internet connection, activeness in the curriculum learning etc. By using Google forms we collected 120 responses from a group of college students in Kottayam District of Kerala . These responses were collected and evaluated using the Weka tool. The results are visualized using the classification technique such as J48, MLP, Random Forest , Bagging etc.

4. Data collection

The following are the attributes used in the data collection.

ATTRIBUTE	DESCRIPTION
E-mail Address	E-mail ID of the student
Effectiveness	Effectiveness of online learning
Health Issues	Health issues of online learning
Availability of resources	Availability df resources which affect learning
Internet Connection	Stable internet connection at home
Familiarity of software	Familiarity of software tools used for online learning
Interest of continuation	Interest of on-going online learning

Table 1. Attributes and Description

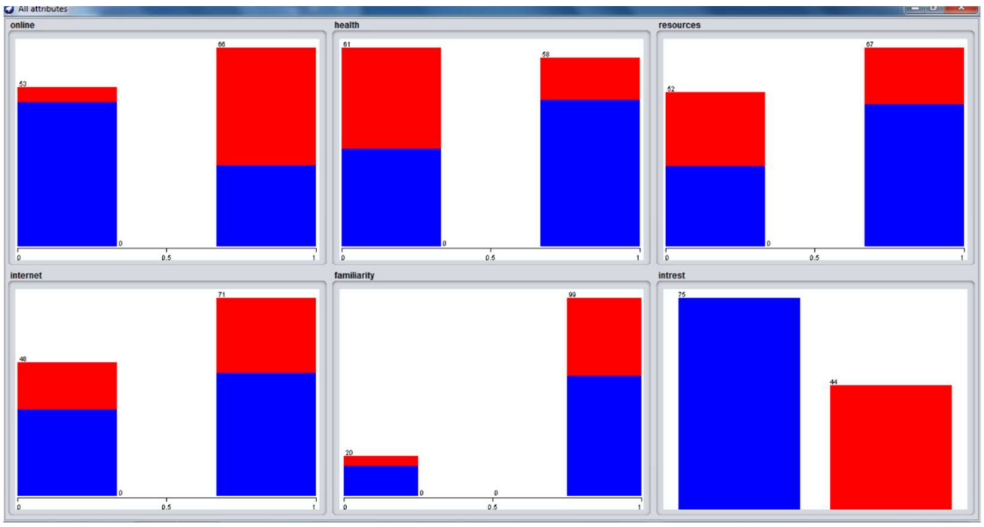

Figure 1. Graphical representation of attributes and values

5. Existing Methodology

The application of data mining would be different for different participants in the educational sector. This may be divided into the students and the educators. For students, data mining can be applicable for discovering their activities but for educators it could be for predicting the student's performance, effectiveness of learning process etc. Without data mining, we could only draw this conclusion by plotting graphs manually, questionnaires etc. This would make the analysis difficult and prone to error (human error).

5.1. Data mining

Data mining is commonly used technique to extract beneficial information from a set of raw unused informatory data. Data mining is also referred as Knowledge Discovery in Databases (**KDD)** which has many application in various fields such as science, genetics, sales, marketing etc . Data mining include many steps such as extraction , transformation , loading data into warehouses, storage and management of data etc. These results are then manipulated into different visual representation such as decision trees, visualisation graph etc for ease of analysis. Most of the data used for data mining is retrieved from database, data warehouses and other information repositories. All the useful data is retrieved through data mining algorithm and the main unsupervised [7] algorithm used are **K Mean Algorithm and** supervised [7] algorithms **Navie Bayes Algorithm**, **J48**, **Random Forest**, **Appropri Algorithm** etc.

5.2. Educational data mining

Educational data mining is an upcoming field in knowledge Discovery due to the high educational predictions. Rather than prediction, classification and grouping can be done accurately through EDM. It gives an answer to the problems behind the profiling ,efficiency, students performance etc [1].

5.3. Classification

Classification is one of the key techniques used to predict instances of various data sets to receive a valid decision classified data. It should be discrete and accurate in order to classify them according to the different classification techniques.

5.4 Applying Machine Learning Techniques

5.4.1 Random forest

Random forest is flexible and supervised learning algorithm [4] which is used for classification techniques as well as in regression. As forest is made into trees, random forest algorithm creates a set of decision making trees which creates a predicted form of each decision tree and finally gets a prominent solution by means of process called voting. This method is used for classification problems. It is better than single decision tree because it reduces each method by averaging the results.

5.4.2 Multilayer Perceptron (MLP)

MLP falls in artificial neural network that evaluates the received data sets into valuable outputs. MLP is one of the main method used for the planning. Graphs regarding the MLP contains multiple layers in a directed graph [2] . It can be connected to one node to another . It can be connected to a node other than input node. MLP uses the back propagation, which is an important tool for increasing the accuracy in predicting the data sets [2] . MLP is also one of the individual functions that follows Cybenko Theorem. It can be used to create mathematical models by regression analysis.

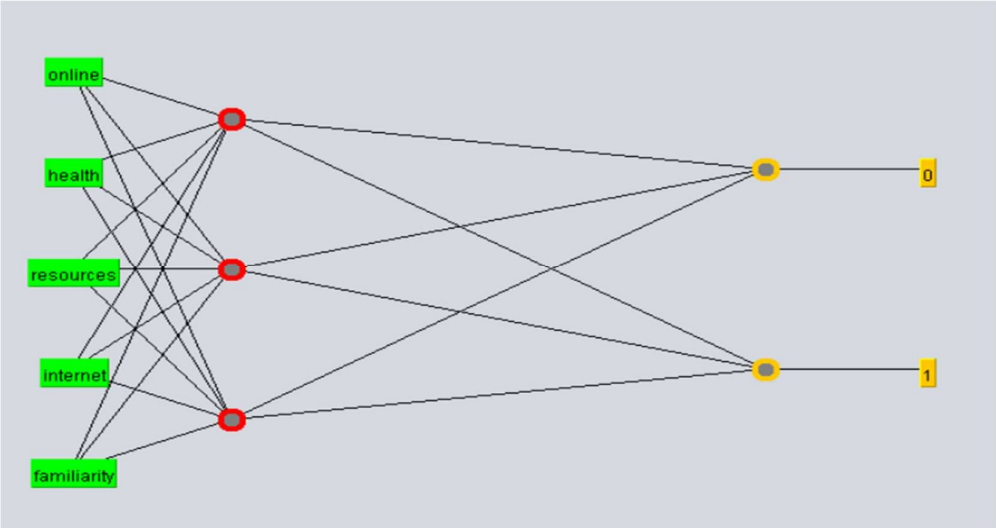

Figure 2. Hidden layers generated by MLP corresponding to the proposed data set

5.4.3 J48

J48 is an Supervised deep learning algorithm used to make decisions from decision tree [4]. It is developed by Ross Quinan. J48 is an extension of ID 3 algorithm.

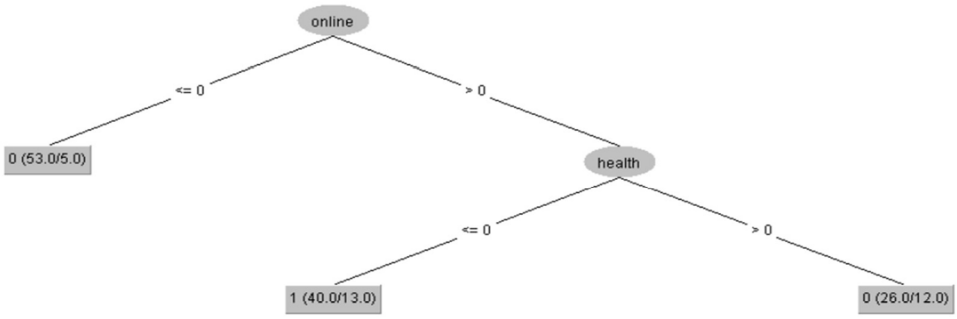

Figure 3. Trees generated by J48 for the proposed data set

5.4.4 Bagging

Bagging is a machine learning meta-algorithm which is used to improve stability and accuracy of statistical classification and regression. Even though it is commonly used for decision-making, it can be also used for model averaging approach. It makes prediction by generating additional data set from the predefined data set with repetitions to produce multiple data sets of original data.

5.5. Learning Management System (LMS)

Learning Management System is mainly engaged with remote learning. Learning Management System is a software application which is used for administration, educational training programs and other functions. Learning Management System was first approved in the school or educational sector but now it is found more in the corporate marketing sector. In data mining Learning Management System is mainly engaged in educational and analytical margin. The data received can be converted into useful information to provide quality predictions in education.

6. Implementation

Implementation of the data set is done through three stages. First stage involves the data collection from various group of students, 2nd Stage mainly involves removing the duplicated data and extracting the relevant data from the database and the last step involved the evaluation classification techniques to form a decision to obtain the decision tree.

6.1. Data collection Process

Data mining used in educational sector is gaining popularity due to the effectiveness in gaining accurate and flexible data. This dataset is collected from a group of 120 students of different institutions in Kottayam District of Kerala. This dataset was used to find the efficiency of the learning through online platforms during the pandemic span of time. This paper mainly focus on two phase:

- Firstly they are classified into their attributes as yes or no .
- Secondly this paper deals with the use of classification techniques such as J48 , Random Forest, MLP etc by constantly checking the size of the data set and using the pre-processing techniques such as data cleaning, data transformation, data integration and data reduction to produce useful data.

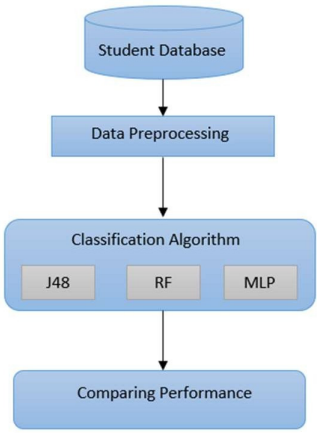

Figure 4. Process of implementation

Table 2. Student Related Variables

ATTRIBUTES	POSSIBLE VALUES
E-mail	Email ID of the students
Effectiveness	**{ Yes, No }**
Health Issues	**{ Yes, No }**
Availability Of Resources	**{ Yes, No }**
Stable Internet Connection	**{ Yes, No }**
Familiarity Of Software Tools	**{ Yes, No }**
Interest In Continuation	**{ Yes, No }**

7. Experimentation and result analysis

This survey and classification is done by using Weka software tool which is open source machine language . It is graphical user interface based tool . It includes tools such as pre-

processing, clustering, regression etc. From the collected dataset 120 samples were taken through Google forms and stored in in MS Excel and converted as .arff files.

The following is the results obtained from these classification techniques:

Table 3

Algorithm	Accuracy	TP Rate	FP Rate	Precision	Recall	F-Measure
J-48	74.8	74.8	30.8	74.4	74.8	74.5
Random Forest	77.3	77.3	22.7	78.4	77.3	77.6
MLP	75.6	75.6	24.6	76.8	75.6	75.9
Bagging	76.5	76.5	24.1	77.4	76.5	76.7

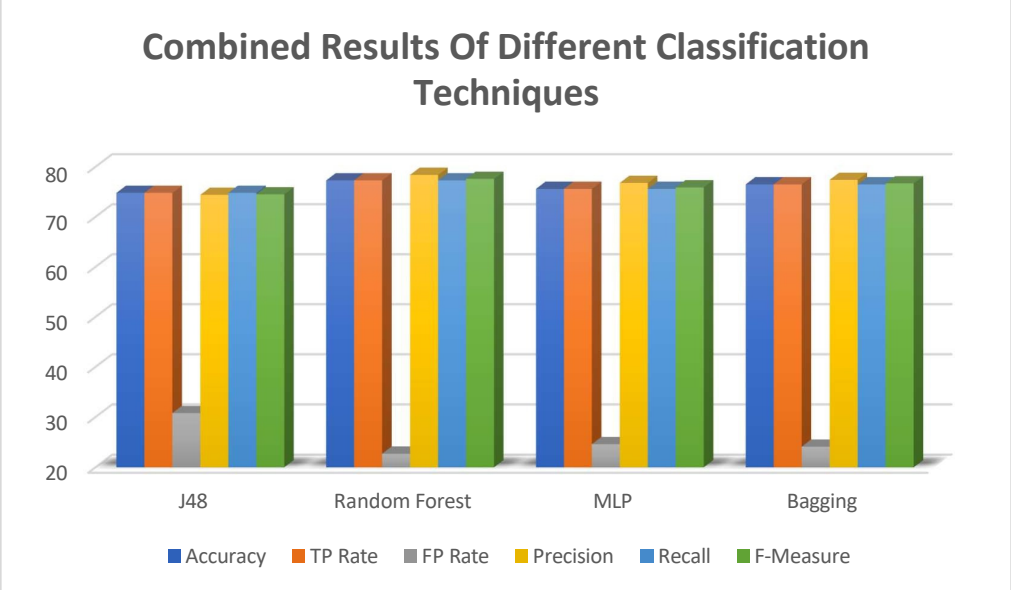

Figure 5. Combined Results Of Different Classification Techniques

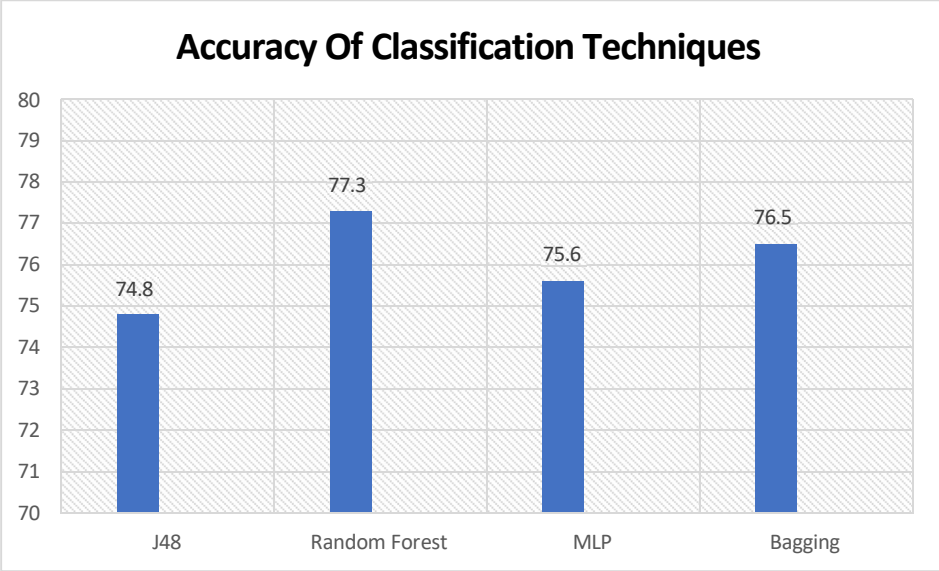

Figure 6. Accuracy of different classification techniques

For calculating the comparison between different classification algorithms Accuracy sensitivity specificity Matrixes are observed here the Kappa statistics is also used for managing the performance.

Accuracy, Sensitivity and Specificity are calculated by using the following formulae [2] :

$$\text{Accuracy} = \frac{(\text{True positive} + \text{True negative})}{(\text{True Positive} + \text{True Negative} + \text{True Positive} + \text{False negative})} \quad (1)$$

$$\text{Sensitivity} = \frac{\text{True positive}}{(\text{True positive} + \text{False Negative})} \quad (2)$$

$$\text{Specificity} = \frac{\text{True Negative}}{(\text{True Negative} + \text{False Positive})} \quad (3)$$

Table 4. Results of classification performance analysis

Algorithm	CCI	ICCI	RMASE
J-48	89	30	0.4075
Random Forest	92	27	0.3807
MLP	80	29	0.3921
Bagging	91	28	0.4017

CCI : Correctly Classified Instances

ICCI : Incorrectly Classified Instances

RMASE : Root Mean Absolute Square Error

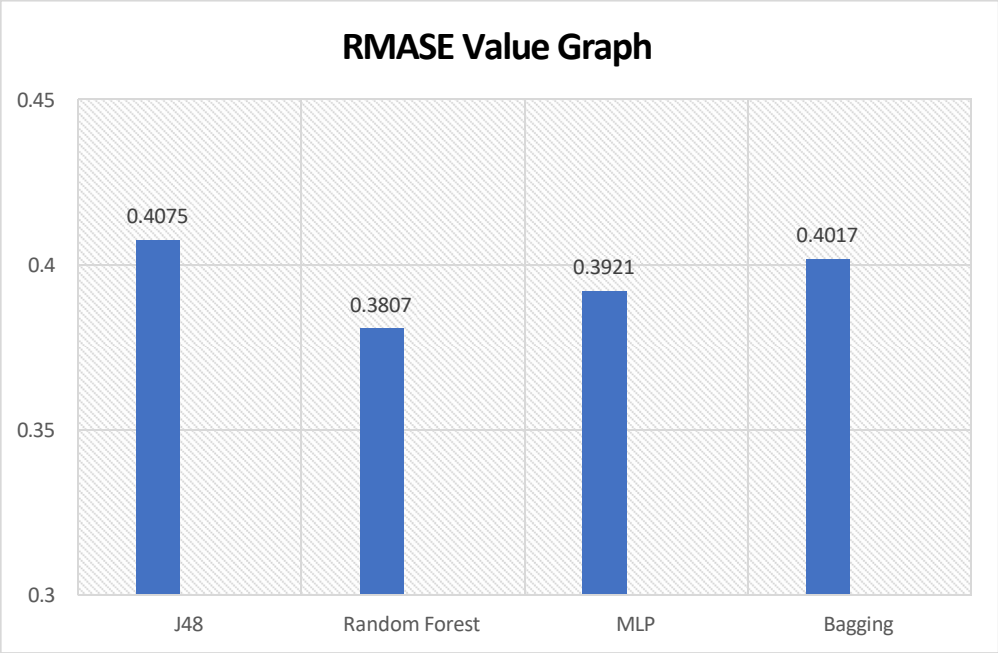

Figure 7. Performance of classifiers corresponding to RMASE

Table 5. Results of classification performance analysis

Algorithm	Kappa statistics	Accuracy
J-48	0.4487	74.8
Random Forest	0.5287	77.3
MLP	0.4938	75.6
Bagging	0.509	76.5

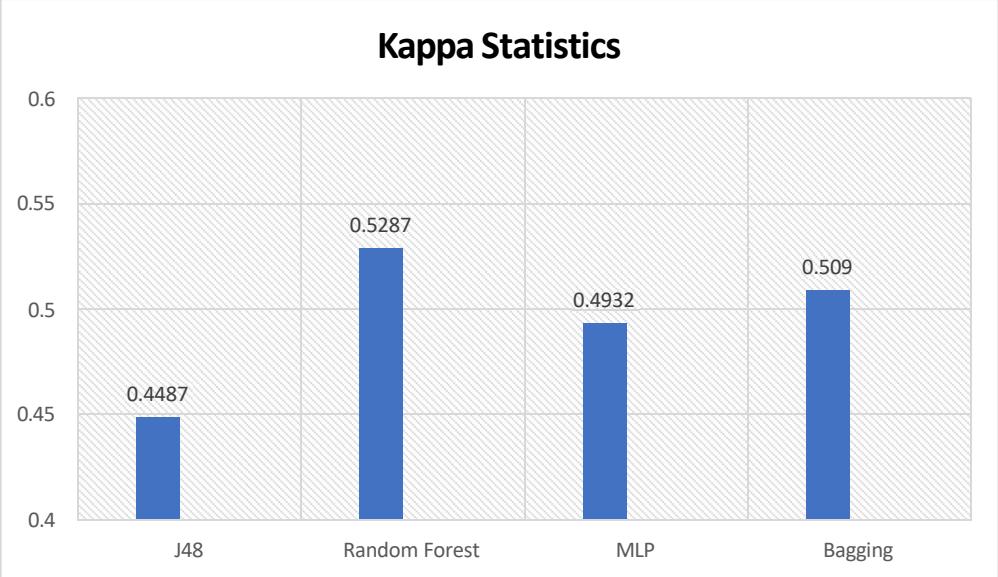

Figure 8. Evaluation of kappa statistics for analyzing the performance of the various classification techniques

Through these observations we conclude that the accuracy of these classification techniques is more in Random Forest than other techniques. The kappa statistics is used to find the precise performance of various classification techniques. In these graphs, Random Forest contains high statistical performance than other techniques. **RMASE** or **Root Mean Absolute Squared Error** is used to measure the average squares of the errors. It should always be positive because of the randomness of the accurate estimate. As the RMASE

value increases, absolute accuracy of the possible error decreases. Random forest have the smallest RMASE value. From the above observations, random forest is the most appropriate technique for predicting efficiency in learning during the span of Covid-19 pandemic.

8. Conclusion

In this paper we predict the efficiency of the online learning through various classification technique. This study helps us to find which aspect affects the online learning. Studies show that over half of the students are facing health issues do to the hectic class hours and more than half of the students disagree in for the online studies. Results from the Google forms were not enough to finalize the conclusion. With the Weka tool, we used many classification techniques such as J48, Random Forest, Bagging and MLP. The results which decide the accuracy of the efficiency in learning are on the basis of Precision values, RMASE, Kappa statistics etc. On the terms of accuracy, the graph shows that the classification techniques which is more accurate is the Random Forest. RMASE value is mainly used in forecasting as well as in regression analysis to verify experimental results. The Kappa statistics is taken as the account chance agreement. When two measurements settle at only a single chance level, the value of Kappa is said to be 0. When two measurements settle perfectly, the value of Kappa is said to be 1.0 [9]. Here the Kappa statistics is greater in random forest, which is 0.5287. This means that they agree at only one chance level. From this paper the best classification technique used for predicting the efficiency of online learning during the span of Covid-19 pandemic is the random forest.

References

[1] *Rositsa Doneva Plovdiv University "Paisii Hilendarski" Silvia Gaftandzhieva Plovdiv University "Paisii Hilendarski" Md. Akhtarul Islam Khulna University,"COVID-19 LOCKDOWN, ONLINE LEARNING AND STUDENTS' PERFORMANCE, November 2020 DOI: 10.21125/iceri.2020.0784*

[2] *Dr. N. V Balaji, Ambily Merlin Kuruvilla, "Heart disease prediction system using Correlation Based Feature Selection with Multilayer Perceptron Approach", IOP Conference Series: Material Science and engineering.*

[3] *Jaya Srivastava Ph.D. Scholar, Jaipur National University, Jaipur Dr Abhay Kumar Srivastava Assistant Professor, Jaipuria Institute of Management, Lucknow, "DATA MINING IN EDUCATION SECTOR: A REVIEW ",Special Conference Issue: National Conference on Cloud Computing & Big Data*

[4] *Ambily Merlin Kuruvilla, Dr.N.V.Balaji ,"PREDICTING DIABETES MELLITUS USING FEATURE SELECTION AND CLASSIFICATION TECHNIQUES IN MACHINE LEARNING ALGORITHMS",Karpagam JCS Vol.13 Sep. - Oct. 2019*

[5] *Amjad Abu Saa Information Technology Department Ajman University of Science and Technology Ajman, United Arab Emirates,"EDUCATIONAL DATA MINING & STUDENTS' PERFORMANCE PREDICTION",(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 5, 2016*

[6] ***6**. K. Prasada Rao Research Scholar Dept.of CSE AITAM, Tekkali M.V.P. Chandra Sekhara Rao, PhD Professor, Dept.of.CSE R.V.R.&J.C College of Engg., B. Ramesh Assistant Professor Dept.of CSE AITAM, Tekkali,,"PREDICTING*

LEARNING BEHAVIOR OF STUDENTS USING CLASSIFICATION TECHNIQUES", April 2016International Journal of Computer Applications 139(7):15-19 DOI: 10.5120/ijca2016909188

[7] *Ambily Merlin Kuruvilla, Dr.N.V.Balaji , "A REVIEW AND ANALYSIS ON DATA MINING ALGORITHMS TO PREDICT DIABETES", Journal of Information and Computational Science.*

[8] *Jui-long Hung Boise State University, Ke Zhang Wayne State University, "REVEALING ONLINE LEARNING BEHAVIOURS AND ACTIVITY PATTERNS AND MAKING PREDICTIONS WITH DATA MINING TECHNIQUES IN ONLINE TEACHING", MERLOT Journal of Online Learning and Teaching Vol. 4, No. 4, December 2008*

Web Resources.

[9] https://www.sciencedirect.com/science/article/pii/B9780123705440500276