***Type: xxx***

Effectiveness of Online Classes DuringCovid-19 Pandemic Using Machine Learning Approach

Saheeb Tareque1, AsrafiAkter1, NazibaNasir1, TahiaTazin1 and Mohammad Monirujjaman Khan1\*

1Department of Electrical and Computer Engineering, North South University, Bashundhara, Dhaka-1229, Bangladesh

\*Corresponding Author: Mohammad Monirujjaman Khan. Email: monirujjman.khan@northsouth.edu

Received: XX Month 202X; Accepted: XX Month 202X

**Abstract:** The deadly COVID-19 started its journey from the December 2019 from the country china. For most of the patients it’s like a mild fever and non-specific gastrointestinal symptoms to a lesser extent. Aged people with previous illness like diabetics, heart problem, and blood pressure etcetera suffer the most. Corona virus 169,094,726 confirmed cases of COVID-19, including 3,512,510 deaths, were reported to the World Health Organization till May 26, 2021.Since the covid-19 lifestyle has changed for almost everyone. That’s when E-learning had a bigger impact for all the students around the world. It is no different for the students of Bangladesh too. In order to keep the study of the students, the government of Bangladesh took a decision and advised the educational institutions to take classes through online. While attending online classes students are facing many opportunities and obstacles like: disruption in class, health issues, financial issues, save time from traffic. However, being captive, they have utilized their time a lot better and this has reflected on their performances too. Machine Learning is a sub-region of man-made reasoning, which predicts outcomes depending on the features on a given dataset. In this paper with the help of machine learning approach effectiveness of online classes during Covid-19 pandemic was formed. By creating our own dataset, model was trained and then prediction was done depending on the different features of the dataset. The collecting data that helped to create a model and which gave the highest accuracy over student’s performance. For classification, multiple techniques were used, including Linear Discriminant Analysis (LDA), Logistic Regression, and K-Nearest Neighbor (KNN). Prediction score Accuracy of 81.05 % by LDA, 86.3% by Logistic Regression and 84.2% by KNN was achieved. The highest prediction score was achieved by Logistic Regression (86.3%). The accuracy shows the effectiveness of online classes during pandemic through Machine Learning approach. It shows how the online classes are effective for a particular student when the input fields are filled. Overall, after performing the three algorithms (K-Nearest Neighbor, Logistic Regression, Linear Discriminant Analysis) with satisfactory results and successfully predicting with high level of accuracy while maintaining keeping up with its objective of being ease and easy to apprehend.

**Keywords:** E-learning; student performance; pandemic; machine learning

**1 Introduction**

Most of the people pre-covid-19 never thought about attending online classes for once. But ever since the outbreak of covid-19, everything started getting out of their regular life-style and people started social distancing with one another and with that on campus classes were shutoff and the new era of online classes began that never witnessed so much people before. It was no different for the students of Bangladesh too. Having no options left, University Grant Commission of Bangladesh (UGC) instructed all universities to start taking online classes. The COVID-19 pandemic emphasizes the need of increasing remote learning quality for advanced education. According to projections as of April 2, 2020, 3,278 higher education institutions and 22.3 million students are affected around the world [1]. So far, 10,200 faculty members have taught a total of 203,200 classes to more than 9.2 million students.[2].

The world has become so advanced in technology. One such technological advancement is machine learning. The phrase ‘machine learning’ was first heard from Arthur Samuel in 1952. Machine Learning is a sub-discipline of artificial reasoning in which the phrase refers to the ability of IT frameworks to discover solution to problems on their own by observing patterns in data sets. Machine learning could be applied to a variety of areas, including student progress, image recognition, traffic prediction, product suggestions, self-driving cars, spam filtering, and malware detection. By utilizing machines, the instructor can monitor each student on an individual level and evaluate their learning progress, individually. Machines can also give extra learning examples of the students, which help instructors to determine the most ideal methods of teaching the students.

In this COVID-19 crisis situation educational institutions are currently based on online studies. As a result, with the undeniable rise of e-learning, whereby education is embraced remotely and on computerized stages [3], training has changed dramatically. With this dramatic shift away from the study hall in many parts of the world, some are wondering whether internet learning would continue to thrive post-pandemic, and what such a shift would mean for general education. Bangladesh being a developing country, can’t provide electricity and stable internet 24x7. Hence students face disruption while attending online classes. Since students are doing online classes from their home and have to look at the screen for a longer period of time, they are facing vision problems, back pain. A lot of students used to bear their university tuition fees by earning money through tutoring even that stopped post covid-19. As a result, they are getting depressed that leads to mental health issues. And for those who do have access to the right technology, learning online can be more effective to them in a number of ways. Hence in this paper the input fields for a particular student that has been focused on are electricity, internet, device, hardware courses, mental health issues, depression, dependency on private, class participation, economic condition, cheating in exam, grading policy etc. The different algorithms used in this paper takes the input fields mentioned above to produce the prediction output.

In [4] the author has highlighted how dissect inputs assembled from e-Campus framework by utilizing Methods and the outcome would be positive, negative and unbiased. Classification algorithms used for research model. Decision tree-classifier, MLP classifier, XGB, and SVC were utilized in the study model to estimate the likelihood of a result with only two values. The Logistic regression approach yielded a 77.5 percent correctness ratio. The author of [5] focuses on student commitment predictions in an E-learning system. In a social science course at the Open University (OU), several machine learning (ML) algorithms were employed to discriminate low-commitment students in order to analyze the influence of participation on student achievement. To create predictive (learning analytic) models that predict student involvement, researchers used six types of machine learning classifiers (decision trees, JRIP, J48, gradient-boosted trees (GBT), classification and regression tree (CART), and a Naive Bayes classifier (NBC). Following an analysis of the data, it was determined that the J48, DT, JRIP, and GBT classifiers were suitable for predicting low-engagement pupils. In [6] the author has talked about the challenges in e-learning that personalized classification in grade prediction by machine learning approach. The first part of personalization is predicting the grades, and the second is writing the recommendation. Therefore, algorithms that have been performed in this project are Linear Regression, Polynomial Regression and the accuracy are 90%. Logistic regression, Support Vector Machine (SVM), Artificial Neural Network (ANN), and Decision Trees are used for classification, and their accuracy is 78 percent. In [7], the author underlined how work recommends integrating technology like artificial intelligence and data analysis with learning management systems to improve learning. The AI model included the analysis, data from the LMS in connection to the time students devote to reading the teacher's resource, and data from a student survey in which the amount of time they had to answer each question was discussed. The naive Bayes data mining approach was used to process the results of this study. The author of [8] presented a method for assessing the impact of elements on the e-learning system. From there, it's a matter of explaining the relevance of each component and prioritizing construction investment based on the K-means clustering algorithm and data from students who have participated in the system. To discover the model fit, the data classification approach is used with the methods Multilayer Perceptron (MP), Random Forest (RF), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Nave Bayes (NB).The forecast findings have an accuracy of up to 81.52%. The author has addressed in the field where e-learning is evaluated in terms of definitions and characteristics in [9]. In addition, the numerous challenges that the many participants in this process face are discussed. Similarly, some of the works proposed in the literature to address these issues. After that, there's a quick rundown of some of the most common machine learning and DA techniques. Finally, some of the current research opportunities that apply such methodologies are presented to provide insights into the locations that warrant more exploration and analysis.[10] presents a decision tree algorithm-based model that proposes the optimum algorithm based on performance. This model included three built classifiers (J48, Random Tree, and RepTree), as well as polls filled out by students. Finally, the J48 method was determined to be the best method when compared to the Random Tree and RepTree algorithms. The TP rate characteristic receives the highest estimate of 63.4%, while Precision (which refers to a positive predictive value) similarly receives the highest estimate of 62.9%. All the above systems use different machine learning algorithms to find out better accuracy using Logistic regression, Decision trees, K-Nearest Neighbors, Random Forest, Support vector machine etc. The highest accuracy that was achieved was 90%. However Linear Discriminant Analysis (LDA) algorithm was not used in the above systems.

This paper presents effectiveness of online classes during pandemic based on machine learning algorithms. Three algorithms (Logistic Regression, K-Nearest Neighbor and Linear Discriminant Analysis) have been used. There have been many journals which relates that topic but there hasn’t been any journal published yet that has worked with all three algorithms especially LDA mentioned above. The dataset created in this paper was prepared solely by each of the members involved in completing this paper. No similar work has been done like the outcome of this paper.

Introduction is presented in section one. Section two discusses methods and methodology and section three provide the results and analysis. Finally, in section four conclusion of the presented work is provided.

**2 Method and Methodology**

This section focuses on the approaches and methodologies used to achieve the objectives. Initially to create the dataset a survey of 30 questions was done and the responses (500) was collected in a excel csv file. After cleaning the dataset, features of the project were selected. The process of applying feature scaling or standardization to independent variables or data characteristics. It essentially aids in the normalization of data within a specific range. Feature scaling is done so that the data becomes standard. This step is done by StandardScaler() method. Choosing the data fields was a challenging task. A dilemma was faced in choosing what and how many data fields should be there in the dataset. The dataset was created in such a way to make sure that there are no outliers so that to get the correct predictions and the model was not over-fitting. Cleaning the dataset was another challenging task. The Machine Learning algorithms that have been used in this project are (Logistic Regression, K-Nearest Neighbor, Linear Discriminant Analysis) each algorithm giving accuracy above 80 percentage. The first subsection presents a simple block diagram of the system; following that, the design of the algorithms utilized in this project is highlighted. Finally, equations and mathematical expressions are shown that are required in algorithms mentioned above.

***2.1 Tools***

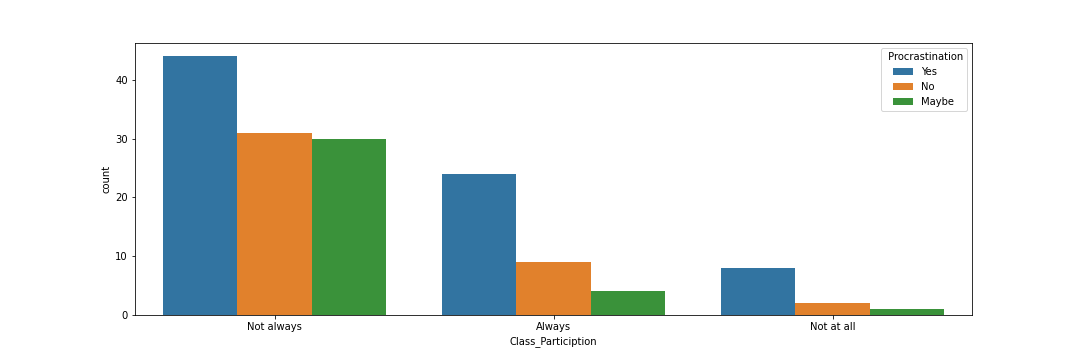
Jupyter Notebooks are powerful, shareable and open-source web based interactive environment that provide the ability to perform data visualization in the same environment. It is mainly used for python because it is used with Machine learning, Artificial Intelligence (AI), as well as Deep learning. The advantages of using python over other programming languages is that it has many libraries that doesn’t makes the code lengthy and the application of python is everywhere starting from professionally building web apps to performing data analysis, web scraping, build games, machine learning application and powerful visualizations. NumPy is a Numeric and NumArray extension. It's an open-source numerical Python library with multi-dimensional arrays and matrix data structures. It may be used to execute a variety of mathematical operations on arrays, including trigonometric functions, statistical procedures, and random number generators. As a result, there are a lot of mathematical, algebraic, and transformation functions in the library. Pandas is a popular open-source Python tool for machine learning, data science, and data analysis. It is built on top of NumPy, a Python module that supports multi-dimensional arrays. Many of the time-consuming and repetitive activities associated with working with data, such as data cleaning, data fill, data normalization, and data visualization, are made simple using Pandas. Matplotlib is a fantastic Python visualization tool for 2D array charts. Matplotlib is a multi-platform data visualization package based on NumPy arrays and meant to integrate with the larger SciPy stack. One of the most significant advantages of visualization is that it provides visual access to enormous amounts of data. Matplotlib has a variety of plots such as line, bar, scatter, histogram, and so on. Seaborn, a Python data visualization library based on Matplotlib, is another option. It has a high-quality interface for creating visually appealing and useful statistical graphs.

As a result, Seaborn is a library that can graphically display data contained in an array, list, or any other data structure.

***2.2 Data Visualization***

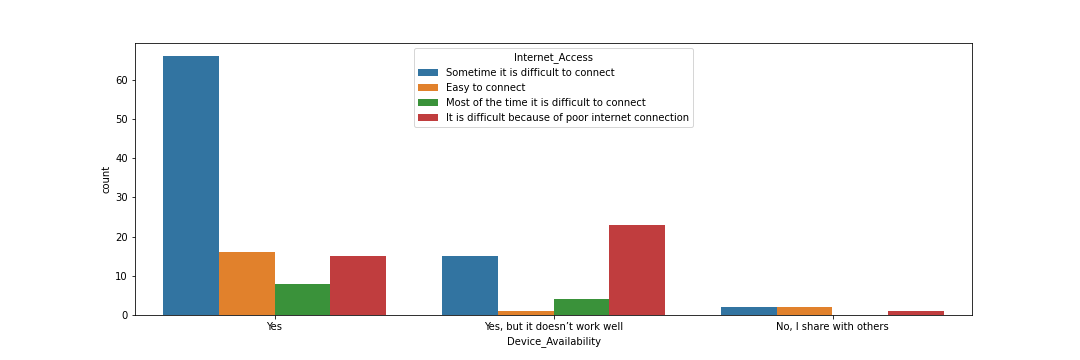
The answers of the respondent were stored in a csv file. For every question (field) the answers were stored in a specific column and to see the relationship between different fields we formed different bar graphs for two different fields and how one affects the other.

Fig.1 shows the relationship between the two fields Class Participation vs Procrastination. The figure illustrates that the least number of students are those who do not participate at class completely and are confused whether they are attentive or not. Maximum number of students is those who don’t class participate and are non-attentive and that number is above 40. And the students who excel in class participation and are also attentive are less than 10 in number in total.



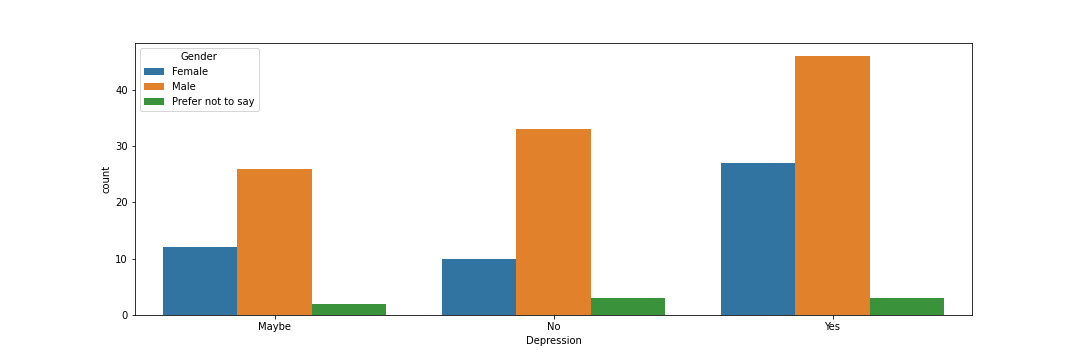
**Figure 1:** Bar chart of Class Participation vs Procrastination

Fig. 2 shows the relationship between the two fields Device availability vs Internet access. The figure illustrates that the least number of students (below 5) do not share their device have very poor internet connection. The maximum number of people (above 60) has devices to access into the class but more often the face problem with the internet connection.

****

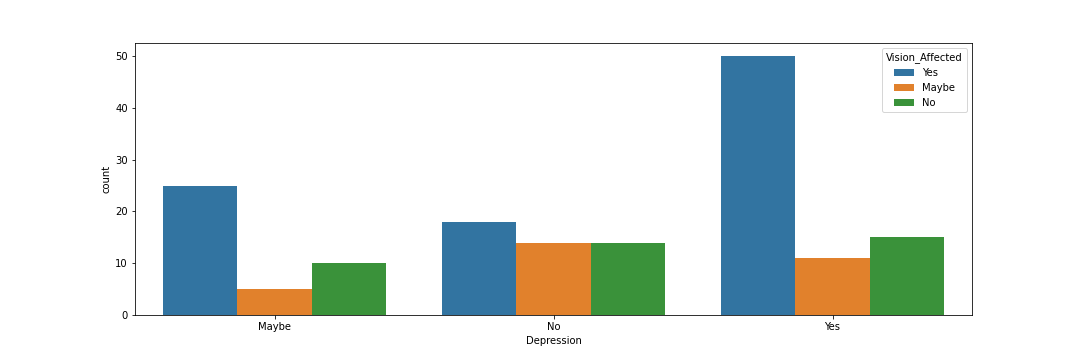
**Figure 2:** Bar chart of Device availability vs Internet access

Fig. 3 shows the relationship between the two fields Depression vs Gender. The figure illustrates that the least number of students which is less than 5, do not prefer to say their gender and are confused whether they are depressed or not. The maximum numbers (above 45) of people who are depressed are male.



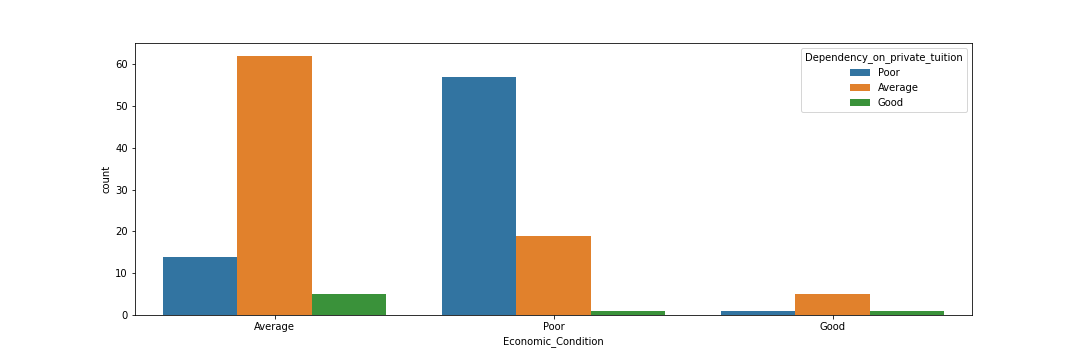
**Figure 3:** Bar chart of Depression vs Gender

Fig. 4 shows the relationship between the two fields Depression vs Vision Affected. The figure illustrates that the least number of students (less than 5) are confused for both depression and Vision affect during this pandemic. The maximum number of people (50) is male who thinks their vision is getting affected.

****

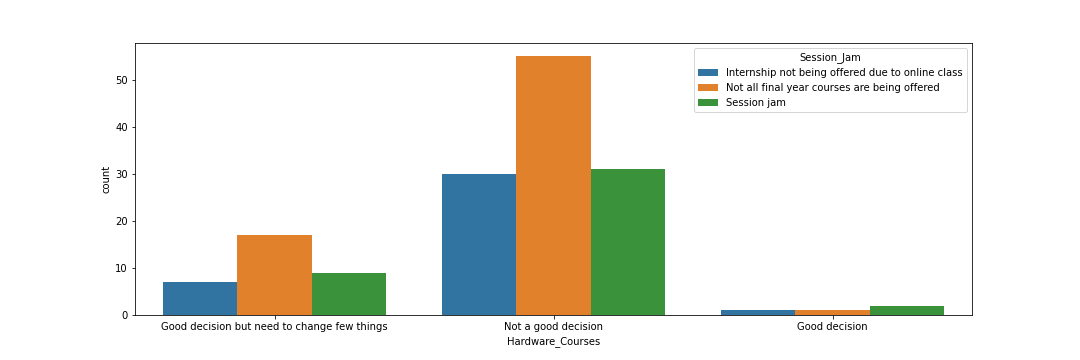
**Figure 4:** Bar chart of Depression vs Vision Affected

Fig. 5 shows the relationship between the two fields Economic condition vs dependency on private tuition. The figure illustrates that very few students (below 3) needed to do private tuition in order to keep their family going during this pandemic. The maximum number of people (above 60) has average economic condition and don’t rely on private tuition too much.

****

**Figure 5:** Bar chart of Economic condition vs dependency on private tuition

Fig. 6 shows the relationship between the two fields Hardware courses vs session jam. The figure illustrates that the most students (above 54) feel taking hardware courses is not a good decision as not all final year courses are not being offered. The least number of students (below 3) who thinks taking a hardware course is a good decision and they will experience session jam since not all final year courses are not being offered.

****

**Figure 6:** Bar chart of Hardware courses vs session jam

Fig. 7 shows the relationship between the two fields Internet connection vs paper submission issues. The figure illustrates that the most students (above 30) can’t complete their paper on time due to sudden power failure. And the least number of people (below 8) face difficulty to submit exam script due to poor internet connection.

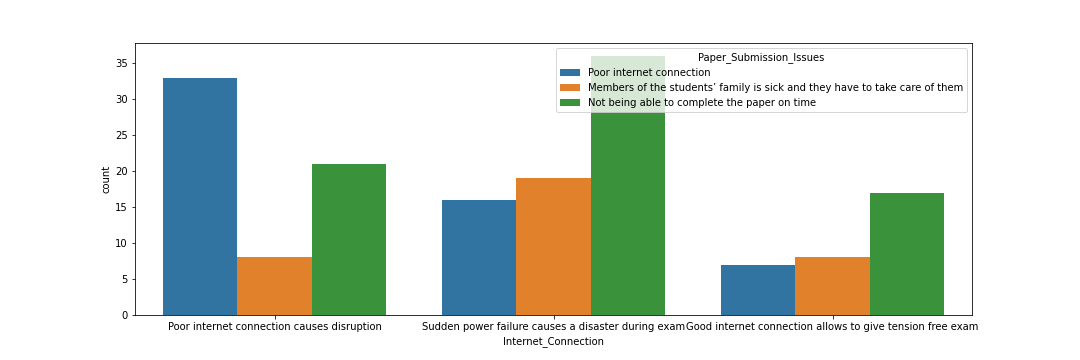
**Figure 7:** Bar chart of Internet connection vs paper submission issues

Fig. 8 shows the relationship between the two fields Teacher’s co-operation vs grading policy. The figure illustrates that the most students (above 27) go through more mental stress during online exam when their teacher is slightly helpful during office hour. However, as the graph suggests there is no correlation between grading policy and teacher’s co-operation.

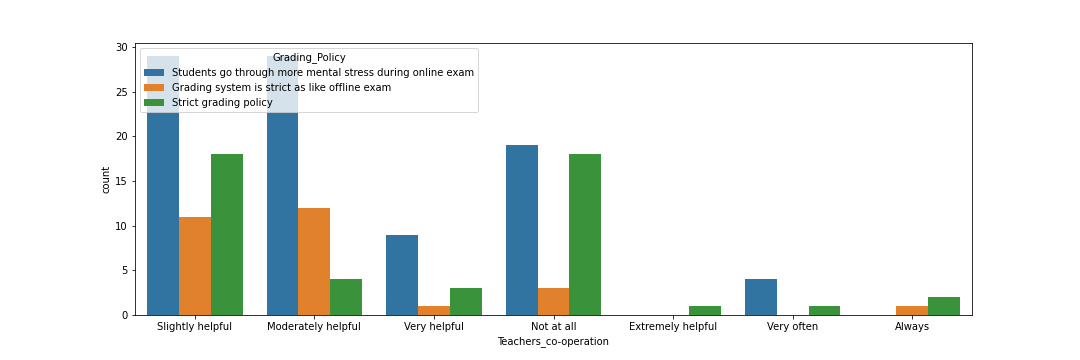
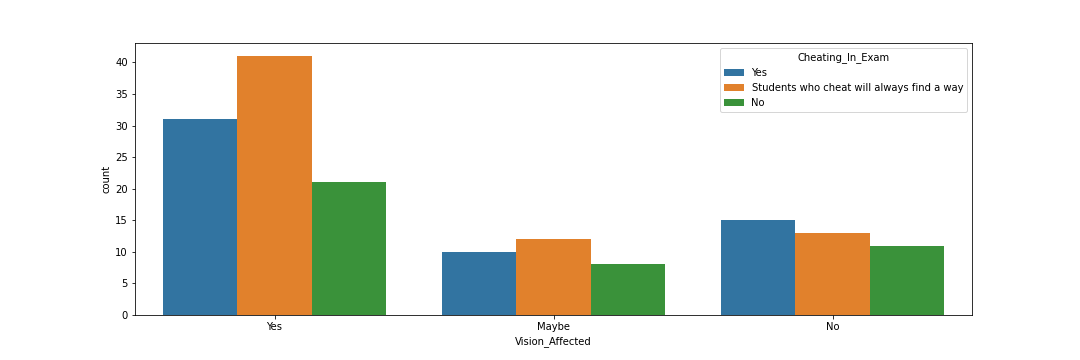
**Figure 8:** Bar chart of Teachers co-operation vs grading policy

Fig. 9 shows the relationship between the two fields Vision affected vs Cheating in exam. The figure illustrates that the least number of students (below 8) do not cheat in exam are confused about their vision getting affect. The maximum number of student (above 40) who thinks their vision is getting affected are the ones who cheat in exam no matter what.

**Figure 9:** Bar chart of Vision affected vs Cheating in exam

***2.3Outline of Full System***

Fig. 10 shows the entire system is depicted as a block diagram. It depicts the actions that were taken to complete this work. Feature scaling is a strategy for standardizing the data's independent features. It is used to handle drastically changing values during data pre-processing. If feature scaling is not performed, the machine learning algorithm will treat higher values as lower values, regardless of the unit used. Data cleaning is the process of identifying the parts of data that are wrong, incomplete, erroneous, irrelevant, or missing, and then changing, replacing, or deleting them as needed. At first the raw dataset goes through data cleaning and then features are selected with the help of feature engineering after that on the given dataset it is spitted into two parts: training and testing (70:30). The accuracy that gets the main attention is when new data is fed to the system to predict its outcome known as the test accuracy. However, training accuracy is also required which is found after applying algorithm on the training model so that it can be understood whether the model is over-fitting or under-fitting. If it’s neither over-fitting nor under-fitting that it can be said that the model is working perfectly to predict the test accuracy.

**Figure 10:** Full system block diagram showing the process.

**Train Model**

**Evaluate Model**

**Score Model**

**Learning Algorithm**

**Training Data**

**Data Cleaning**

**Feature Engineering**

**New Data**

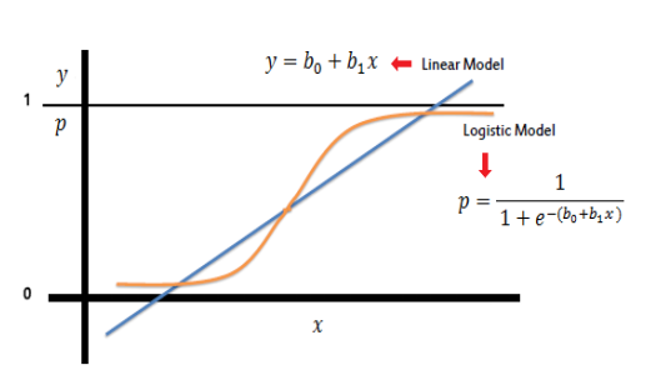
**Dataset**

***2.4 Applied Algorithms***

*2.4.1 Logistic Regression*

Logistic regression is a machine learning approach that is similar to linear regression except that it predicts whether something is true or false rather than something continuous like weight. Simply said, logistic regression produces a binary result (0 or 1). In the case of input, however, logistic regression can be used with both categorical and continuous data. Rather than fitting a straight line to the data, logistic regression uses an S-shaped ‘Sigmoid function,' which converts inputs into a 0–1 range.

Fig. 11 shows the similarity between logistic regression and linear regression is clearly visible. The sigmoid function is applied on top of the linear model to produce the logistic model graph limited to values between 0 and 1. The hypothesis function, also known as the predictive function P(X), calculates the probability of an output equal to 1 given a particular input. P(X) = P(Y=1|X), in other words. The output is defined as 1 if the probability is larger than 0.5. Otherwise, the output is set to 0 (zero). A cost function must be utilized to determine the coefficient vector. However, unlike linear regression, the squared error function will not work here since the sigmoid function used to construct the hypothesis function will cause the output to be wavy. Instead, a cost function based on the log function that offers highest likelihood is utilized. The logistic regression approach can also be used to classify several classes. Instead of a single hypothesis function, numerous functions would be used, each assessing the probability of the output being of a specific class.



**Figure 11:** Showsthe data is fitted in both linear and logistic model [4].

Fig. 12 shows a diagram of the steps involved in logistic system. The responses in CSV file are in categorical forms. Hence at first by the help of get dummies function which is found inside the pandas’ library, the categorical values are converted into numerical values. Once the p value is calculated, if we find it is lower than 0.05 then we reject it [12]. Because chances are very low that it would be of some use in our model because the change in predictor’s value is related to change in responsible variable.

**Predictors with a p value greater than 0.05**

**Predictors with p value > 0.1**

**Multi**

**Variable LR**

**Predictors with p value <=0.05**

**Logistic Regression**

**Figure 12:** The steps involved in logistic regression.

*2.4.2 K-Nearest Neighbor*

KNN which stands for K-Nearest Neighbor is a very simple algorithm used to solve classification problems. It is instant base learning which means it does not learn anything in training period. It simply saves the training dataset and uses it to make real-world predictions. As a result, this makes it faster than other algorithms.

Fig. 13 shows the K-Nearest Neighbor (KNN) algorithm procedure. At first, its need to load data and then the value of K has to be initialized and for each sample in the training data it’s obvious to the distance between the query point and the current point the distance will be graded from little too large after that. Get the labels of the first k entries in a sorted collection by picking the first k entries.

**Define K**

**Compute Euclidean distance**

**Distance is sorted**

**Take K nearest Neighbors**

**Apply simple Majority**

**Figure 13:** The flowchart of K-Nearest Neighbor classifier procedure.

K is the number of neighbors in K-Nearest Neighbor (KNN), for reliable results, choosing the precise value of K, also known as parameter tuning, is necessary. Sqrt(n), where n is the total number of data points, is used to find a value for K. However, extreme caution must be taken to ensure that an odd value of k is chosen at all times to avoid confusion between two classes. Because K-Nearest Neighbor is a lazy learner, it can be utilized when data is labeled, data is noise-free, and the dataset is small. From the training set, it does not learn a discriminative function. The nearest neighbor is usually calculated using Euclidean distance. If there are two points (x, y) and (a, b), the Euclidean distance formula [13] will be applied. We can also employ Manhattan and Minkowski distances, among other things. The unknown data point's Euclidean distance will be determined using all of the points in the dataset. The majority neighbors will be taken for KNN. In our project the converted numerical features will be calculated for a person to know the effectiveness of online class’s base on his or her given parameters and our model will show its accuracy rate. All Scikit-learn libraries need to import to complete this paper. Import Pandas, NumPy, import train\_test\_split, StandardScaler, K-neighbors Classifier and Confusion matrix. KNN will predict the accuracy base on training dataset. Feature Scaling is so important during training period. Confusion matrix will evaluate the model. It's a N x N matrix for evaluating a classification model's performance, where N is the number of target classes. The matrix compares the actual goal values to the machine learning model's predictions.

*2.4.3 Linear Discriminant Analysis*

Linear Discriminant Analysis (LDA) is a machine learning approach for linear classification. This approach improves the performance of a probabilistic model based on the distribution of observations for each input variable. The conditional likelihood of an example's impacts on each class is then calculated, and the class with the highest probability is chosen. It's a straightforward categorization model, and the method we're applying for the probabilistic model generates accurate predictions even when these assumptions are upset. LDA, or Linear Discrimination Analysis, computes summary statistics for input features based on class labels, such as mean and standard deviation. All of these statistics are used to help the model adapt to new data and learn from previous data. LDA makes predictions by calculating the likelihood that a new example belongs to each class label based on the values of each input feature. The class with the highest probability is then allowed to participate in the example. LDA is a straightforward application of BAYES theorem for classification. The LDA classifiers are based on the assumption that each observation inside a class comes from a normal distribution with a common variance and a class-specific mean vector.

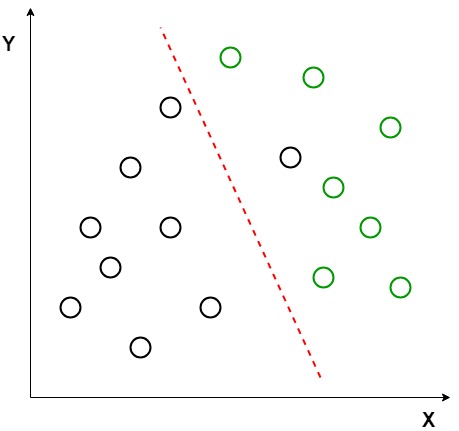
If there are two classes in Fig. 14, for example, they must be accurately separated. Classes can have a variety of characteristics. Using only one feature to categorize them may result in some overlap.

As a result, the number of features required for proper classification will continue to grow.



**Figure 14:** Overlapping used in LDA to separate two classes [21]

Assume there are two sets of data points to be classified, each of which corresponds to a different class. When the data points are plotted on the 2D plane, as shown in the Figure 15 2D graph, there is no straight line that can finally isolate the two classes of data points. Thus, LDA (Linear Discriminant Analysis) is utilized in this case, which reduces the 2D graph to a 1D graph to increase the separability between the two classes.



**Figure 15:** 2D Graph that illustrates two classes of the data points [21]

Linear Discriminant Analysis (LDA) uses both axes (X and Y) to create a new axis and projects data onto a new axis in order to increase the separation between the two categories and reduce the 2D graph to a 1D graph.

***2.******5Equations and Mathematical Expressions***

Logistic Regression equation:

Euclidean Distance Equation:

Manhattan Distance Equation:

**3 Results Analysis**

In this section discussion was done in finding out the effectiveness of online classes during pandemic through machine learning approach. Firstly, it started with a survey where the responses were converted into a csv file. After that from raw data, feature selection was done once data cleaning was accomplished. And from this collecting data a model was created that gave the highest accuracy in predicting student’s performance. How actually they did in offline class in contrast to online class and also figure out the variables that are responsible in altering the output the model is build, then focus on improving the accuracy of prediction. After analyzing the student’s performance during pandemic whether the online class is really effective or not.

**The number of correctly identified samples (True positive + True Negative) divided by the total number of samples (True Positive + True Negative + False Positive + False Negative) is the model's accuracy rate. The error rate is the percentage of times the model predicts one class but actually predicts another. In other words, it's the ratio of the number of samples that were incorrectly classified (False Positive + False Negative) to the total number of samples (True Positive + True Negative + False Positive + False Negative).**

**TP + TN**

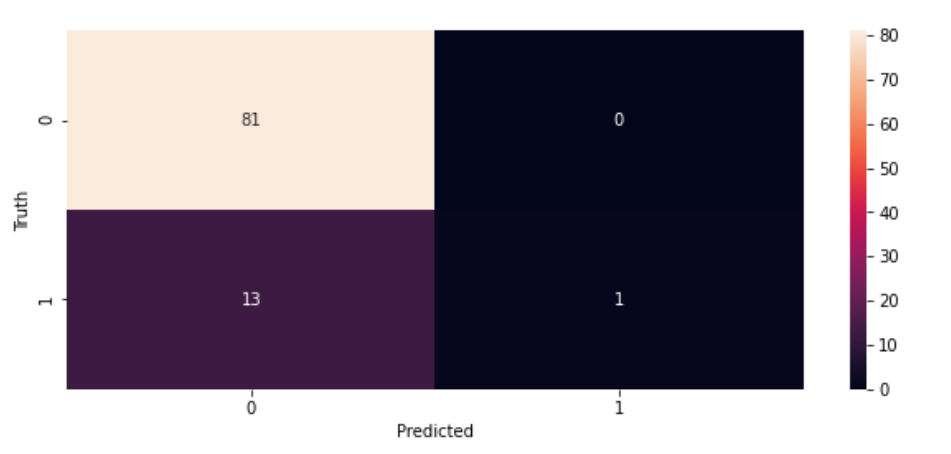
**Accuracy =**

**TP + TN + FP + FN**

Table 2 shows the confusion matrix of Logistic Regression. As the table suggests the model predicts the right class correctly (Effective) 81 times and wrong class (Not Effective) correctly 1 time. The model prediction is wrong (FP+FN) in predicting a definite class out of two classes for 13+0 =13 times. As a result, the model's real accuracy is (81+1)/ (81+1+13+0) = 0.863 = 86.3 percent.

**Table 2:** Confusion Matrix of Logistic Regression

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TP** | **TN** | **FP** | | **FN** |
| 81 | 01 | | 00 | 13 |

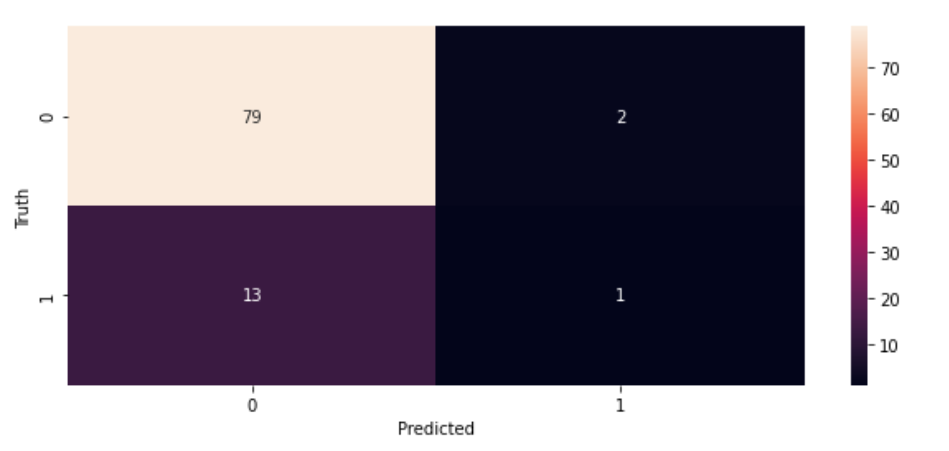


**Figure 16:** Confusion Matrix of Logistic Regression

Table 3 shows the confusion matrix of K-Nearest Neighbor. As the table suggests the model predicts the right class correctly (Effective) 79 times and wrong class (Not Effective) correctly 1 time. The model prediction is wrong (FP+FN) in predicting a definite class out of two classes for 13+2 =15 times. As a result, the model's true accuracy is (79+1) / (79+1+13+2) = 0.842 = 84.2 percent.

**Table 3:** Confusion Matrix of K-Nearest Neighbor

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TP** | **TN** | **FP** | | **FN** |
| 79 | 01 | | 02 | 13 |

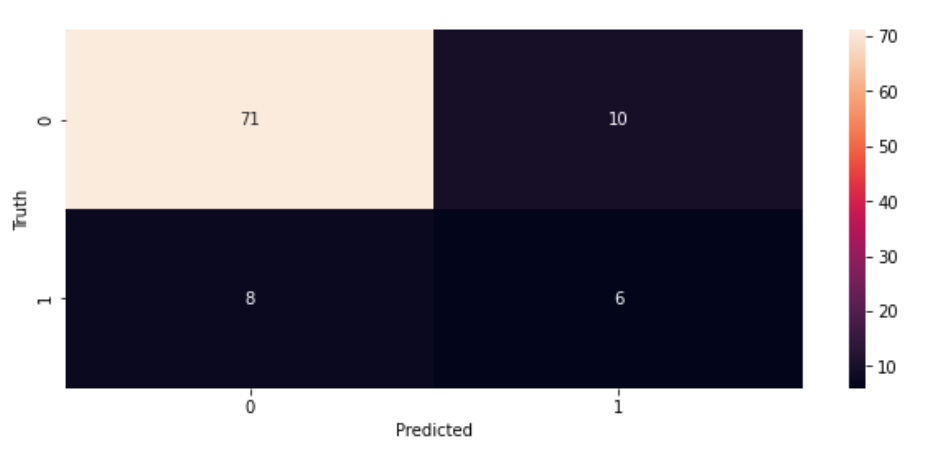
****

**Figure 17:** Confusion Matrix of K-Nearest Neighbor

Table 4 shows the confusion matrix of Linear Discriminant Analysis. As the table suggests the model predicts the right class correctly (Effective) 71 times and wrong class (Not Effective) correctly 6 times. The model prediction is wrong (FP+FN) in predicting a definite class out of two classes for 10+8 =18 times. As a result, the model's true accuracy is (71+6) / (71+6+10+8) = 0.8105 = 81.05 percent.

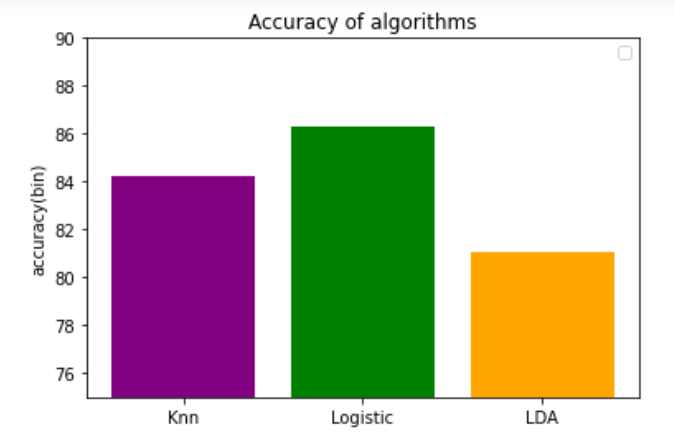
**Table 4:** Confusion Matrix of Linear Discriminant Analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TP** | **TN** | **FP** | | **FN** |
| 71 | 06 | | 10 | 08 |

****

**Figure 18:** Confusion Matrix of Linear Discriminant Analysis

Fig. 19 shows the accuracy that is achieved by the three different algorithms. The score accuracy gives the right percentage in predicting the effectiveness when different fields such as class participation, procrastination, depression, internet access, device availability, teacher’s cooperation, economic condition, hardware courses, vision affected, dependency on private tuitions, university offering resources are provided as inputs.



**Figure 19:** Prediction Accuracy

Table 5 shows the accuracy of different algorithms used in this paper. In our proposed algorithms we have achieved 86.3% accuracy using Logistic Regression (LR), 84.2% using K-Nearest Neighbor (KNN) and 81.5% using Linear Discriminant Analysis (LDA). The accuracy shows the effectiveness of online classes during pandemic through Machine Learning approach. It shows how the online classes are effective for a particular student when the input fields are filled. There have been several articles published on that topic, but none have worked with all three methods, particularly the LDA stated above. Each of the members participating in the completion of this study produced the dataset used in this article. This paper's results are unlike anything else that has been done before.

**Table 5:** Accuracy comparison with different Algorithms

|  |  |
| --- | --- |
| Algorithms Name | Accuracy |
| Logistic Regression | **86.3%** |
| K-Nearest Neighbors | 84.2% |
| Linear Discriminant Analysis | 81.5% |

Table 6 shows the comparison of existing journals with proposed journal. All the existing journals have some similarities with our proposed journals. In [4] the author has displayed how dissect inputs assembled from e-Campus framework by utilizing methods and the result would be positive, negative and unbiased. Classification algorithms used for research model. Decision tree-classifier, MLP classifier, XGB, and SVC were utilized in the study model to estimate the likelihood of a result with only two values. In [4] score of 77.5%accuracy was achieved via the Logistic regression algorithm. In [6] it was mainly discussed about the challenges in e-learning that personalized classification in grade prediction by machine learning approach. The first part of personalization is predicting the grades, and the second is writing the recommendation. The accuracy score was90% for Linear Regression, Polynomial Regression. In this paper, accuracy was attained 78% using Classification Logistic Regression, Support Vector Machine (SVM), Artificial Neural Network (ANN), and Decision Trees techniques. The author of [8] presented a method for assessing the impact of elements on the e-learning system. From there, it's a matter of explaining the relevance of each component and prioritizing construction investment based on the K-means clustering algorithm and data from students who have participated in the system. To discover the model fit, the Multilayer Perceptron (MP), Random Forest (RF), K-nearest neighbor (KNN), Support Vector Machine (SVM), and Nave Bayes (NB) methods are used. The accuracy of the prediction results is up to 81.52%. In this paper prediction score Accuracy of 81.05 % by LDA, 86.3% by Logistic Regression and 84.2% by KNN was achieved. The highest prediction score was achieved by Logistic Regression (86.3%).

**Table 6:** Accuracy comparison with other papers

|  |  |  |
| --- | --- | --- |
| Name | Algorithm | Accuracy |
| This paper | LR | **86.3%** |
| This paper | KNN | 84.2% |
| This paper | LDA | 81.5% |
| Ref [4] | LR | 77.5% |
| Ref [6] | LR | 78% |
| Ref [8] | KNN | 81.52% |

**4 Conclusion**

Using a Machine Learning technique, this research examines the effectiveness of online lessons during the Covid-19 pandemic. Three algorithms were used for the prediction of effectiveness of online classes. Prediction score Accuracy of 81.05 % was achieved by LDA, 86.3% by Logistic Regression and 84.2% by KNN. The highest prediction score was achieved by Logistic Regression (86.3%). The prediction score suggests how the online classes are effective for a particular student when the input fields are filled. Our collecting data helped us to create a model to predict student’s performance. How actually they did in offline class in contrast to online classes. After Analyzing the student’s performance during pandemic, it can be understood whether the online class is really effective or not. Afterwards having a look at the results, researchers will focus on different fields that hamper student’s performance and work on it so that the problem gets resolved before the start of next semester. They will think of a way that can get the students attention more in ongoing online class. A possible solution could be trying different ways to make the student class participation and creating discussion forums so that whenever they have a question on their mind, they can get help. Chances are high that the dataset that has been created by us will be of use for the developing countries only because some of the fields on our dataset i.e., problem of electricity, availability of smart devices etc. Will have mixed answers (Yes or no) whereas in developed countries availability of smart devices and electricity is not of a concern for anyone. However, if there are researchers from developed countries that are keen to work on the online class’s effect in developing countries then the dataset of ours could come of great help for them. Even though, we will be working on our paper during this pandemic but it can also be used later on post pandemic. Hopefully this covid-19 gets over very soon and if there is a similar situation like this in near future then our project could be of great help. One of the benefits of E-Learning for students who work while studying is that, because the online class is recorded, they can watch the video of that class whenever they want after work. Other methods can be used to improve the system in the future, and the dataset we collected can be used to find better accuracy.

**References**

1. S. Jung,A. Akhmetzhanov, K. Hayashi, N. Linton, Y. Yang, B. Yuan, et al., "Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: Inference Using Exported Cases", *Journal of Clinical Medicine,* vol. 9, no. 2, p. 523, 2020.
2. Q. Pham, D. Nguyen, T. Huynh-The, W. Hwang, P. Pathirana, "Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic: asurvey on the state-of-the-arts", *IEEE Access,* vol. 8, pp. 130820-130839, 2020.
3. S. Ryu, H. Gao, J. Wong, E. Shiu, J Xiao, M. Fong, B. Cowling, et al., "Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings—international travel-related measures", *Emerging Infectious Diseases*, vol. 26, no. 5, pp. 961-966, 2020.
4. U. Osmanoğlu, O. Atak, K. Çağlar, H. Kayhan, T. Can, "Sentiment analysis for distance education course materials: a machine learning approach", *Journal of Educational Technology and Online Learning*, vol. 3, no. 1, pp. 31-48, 2020.
5. M. Hussain, W. Zhu, W. Zhang, S. Abidi, "Student engagement predictions in an e-learning system and their impact on student course assessment scores", *Computational Intelligence and Neuroscience,* vol. 2018, pp. 1-21, 2018.
6. P. Fung and D. Roth, "Guest editors introduction: machine learning in speech and language technologies", *Machine Learning*, vol. 60, no. 1-3, pp. 5-9, 2005.
7. W. Villegas-Ch, M. Román-Cañizares and X. Palacios-Pacheco, "Improvement of an online education model with the integration of machine learning and data analysis in an LMS", *Applied Sciences*, vol. 10, no. 15, p. 5371, 2020.
8. D. Lu, H. Le and T. Vu, "The factors affecting acceptance of e-learning: a machine learning algorithm approach", *Education Sciences*, vol. 10, no. 10, p. 270, 2020.
9. A. Moubayed, M. Injadat, A. Nassif, H. Lutfiyya, A. Shami, "E-learning: challenges and research opportunities using machine learning & data analytics", *IEEE Access*, vol. 6, pp. 39117-39138, 2018.
10. A. Hamoud, A. Hashim and W. Awadh, "Predicting student performance in higher education institutions using decision tree analysis", *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 2, p. 26, 2018.
11. P. Larrañaga et al., "Machine learning in bioinformatics", *Briefings in Bioinformatics*, vol. 7, no. 1, pp. 86-112, 2006.
12. S. Kotsiantis, I. Zaharakis and P. Pintelas, "Machine learning: a review of classification and combining techniques", *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159-190, 2006.
13. T.Wuest, D. Weimer, C. Irgens, K. Thoben, "Machine learning in manufacturing: advantages, challenges, and applications", *Production & Manufacturing Research*, vol. 4, no. 1, pp. 23-45, 2016.
14. D. Lieber, M.Stolpe, B. Konrad, J. Deuse, K. Morik, "Quality prediction iniInterlinked manufacturing processes based on supervised & unsupervised machine learning", *Procedia CIRP*, vol. 7, pp. 193-198, 2013.
15. D. Aha, D. Kibler and M. Albert, "Instance-based learning algorithms", *Machine Learning*, vol. 6, no. 1, pp. 37-66, 1991.
16. A. Samuel, "Some studies in machine learning using the game of checkers", *IBM Journal of Research and Development,* vol. 3, no. 3, pp. 210-229, 1959.
17. D. Arora, "A review of machine learning techniques over big data case studies", *International Journal of Innovative Research in Computer Science & Technology,* vol. 8, no. 3, 2020.
18. A. Tleubaev, "Application of machine learning methods for subject classification of the internet domains", *Machine Learning and Data Analysis,* vol. 4, no. 3, pp. 192-214, 2018.
19. C. CDC Weekly, "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) — China, 2020", *China CDC Weekly*, vol. 2, no. 8, pp. 113-122, 2020.
20. D. Zhang, M. Orgun and K. Zhang, "Special issue on information visualization in machine learning and applications", *Journal of Visual Languages & Computing,* vol. 33, pp. 1-2, 2016.
21. "ML | Linear Discriminant Analysis - GeeksforGeeks", GeeksforGeeks, 2021. [Online]. Available: <https://www.geeksforgeeks.org/ml-linear-discriminant-analysis/>.