Students' Adaptability Level Prediction in Online Education using Machine Learning Approaches

Md. Mahmudul Hasan Suzan
Department of CIS
Daffodil International University
Dhaka, Bangladesh
mahmudul16-410@diu.edu.bd

Nishat Ahmed Samrin
Department of CIS
Daffodil International University
Dhaka, Bangladesh
nishat16-380@diu.edu.bd

Al Amin Biswas
Department of CSE
Daffodil International University
Dhaka, Bangladesh
alaminbiswas.cse@gmail.com

Md. Aktaruzzaman Pramanik
Department of CSE
Daffodil International University
Dhaka, Bangladesh
a.pramanikk@gmail.com

***Abstract*—Online Education has become a buzzword since the COVID-19 hit the World. Most of the educational institutions went online to continue educational activities while developing countries like Bangladesh took a significant period of time to ensure online education at every education level. Students of several levels also faced many difficulties when they got introduced to online education. It is important for the decision-makers of educational institutions to be informed about the effectiveness of online education so that they can take further steps to make it more beneficial for the students. Our main motivation is to contribute to this matter by analyzing the relevant factors associated with online education. In this work, we have collected students' information of all three different levels(School, College, and University) by conducting both online and physical surveys. The surveys form consists of an individual's socio-demographic factors. To get an idea about the effectiveness of online education we have applied several machine learning algorithms named Decision Tree (DT), Random Forest (RF), Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and also Artificial Neural Network(ANN) on our dataset to predict the adaptability level of the students to online education. Among used algorithms, the Random Forest classifier achieved the best accuracy of 89.63% and outperformed other algorithms.**

***Index Terms*—Online Education, Machine Learning, Prediction, Random Forest, Classification**

I. Introduction

Modernism is rising with time, we are progressing in each field but how far have we reached in online education? Have we been skilled enough to adapt ourselves to online education? The purpose of this paper is to give some focus on the current online education scenario in Bangladesh. We have tried to predict a student's adaptability level to the online education system. Traditional education has vastly changed within the last couple of years. Mostly in the Covid-19 situation, online classes are the only solution to pursue education. In this research, we have tried to explore the gap between online and physical classroom-based education in the perspective of Bangladesh. We analyzed the potential moderating effects of gender, age, level, location, class-time, internet connection quality, govt./non-govt.institution, institution's own LMS availability, and adaptability level of students. We found 8.298% students only adapt to online education highly. The lack of smooth internet connection, lack of digital skills, electricity problem, poor network connection, lack of proper instruction, resistance to convert, etc. are the major obstacles for online education during the Covid-19 pandemic situation [1]. Before covid-19 students were not much aware of online education. But the sudden hit of the pandemic students are still struggling to adapt themselves to online education. A significant difference regarding students performances and satisfaction confirmed many benefits of online education[2]. It is a major object to work out on online education.

In Bangladesh, students are comfortable with traditional physical classes but an unexpected closure in March 2020 has changed everything. The circumstances reveal that online education is the solution to pursue further education. To perceive student experience, we conducted a survey nation-wide to collect information from different institutions. A descriptive survey was used to find student's experience in online education. About 51.87% of the students adapt online classes properly and most of the 39.83% hardly adapt to online education. The first challenge for students is adaptability to online learning. In Bangladesh, internet connection is limited in rural areas which is also a major issue. It is a big challenge for the Govt. to expand the online education system to every level. To properly execute online education, students must have the motivation to take it out appropriately [3]. The opportunities of online education are a lot, it demands suitable execution.

Therefore, the major contributions we have proposed in this work are:

- Socio-demographic factors can play a significant role to ensure the smooth execution of the online class. So, predicting the student's adaptability level based on these factors helps to understand the decision-makers to take

2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) | 978-1-7281-8595-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCNT51525.2021.9579741

necessary steps to mitigate the issues.
- To understand the predictive capability of the classifier, a comparison of the model's performance is performed.
- We used KNN, Decision Tree, Random Forest, Naive Bayes, SVM, and ANN to predict the student adaptability and achieved the highest prediction accuracy of 89.63% using Random Forest classifier.

Several parts of this paper are organized in below ways: Literature reviews of the existing works have been presented in section II. Section III describes the methodology and details about data collection, data prepossessing, the algorithms used, and implementation. Section IV includes an analysis of the performance of algorithms and an evaluation of models. And finally, the conclusion has been described including our future work in Section V.

II. Literature Review

The contemporary situation excessively exhibits the importance of the online education system. Technological growth permits us to find a way to create online education systems. Technology provides a solution to virtual or remote learning. In current circumstances, aspects of education are going to be held in digitization. For these changes, students have to take up the challenge of adaptation to online education. In the discussion to follow, we give an outline of the learnings from the analysis of related works regarding online education.
In [2] and [4] the researchers have studied improvement of the online education model. Rojan et al. [2] showed a remarkable difference regarding students' performances, satisfaction and firmed many benefits of online education for students. It exhibits similar performances and it shows similar offers and student satisfaction both off-campus and on-campus. William et al. [4] focused on formative assessment for better learning and the exploratory results show 85% of students replied they learn more in online education. Mainly the researchers tried to improve the assessment system for students and teachers, also self and peer assessment for students and teachers.

Rolim et al. [5] in their paper they used a Supervised ML algorithm to recognize the existence of satisfactory practices, they focused on feedback that they collected from the LMS courses. William et al. [4] studied the betterment of the Online Education Model with the integration of Machine Learning and Data Analysis in a Learning Management System(LMS). Researchers of [6] focused on students' academic performance by applying Machine Learning. This paper presents the results and evolution of the project which aims to prepare and measure the performance of some Machine Learning algorithms for analysis purposes and prediction of the student's academic performance in the course. Monica et al. [7] studied that Education is moving on online and course content available on digital platforms. So they analyze based on Neural Networks, Support Vector Machine(SVM), Decision Tree, and Cluster Analysis. Their prediction accuracy of Blended Learning was not satisfactory. The exploratory results show Online education is better than blended education. Kuck et al. [8] studied how much machine learning is effective in education. This paper's main aim was to assess the potential of applying machine learning in the education sector. They defined the main four categories, First, one is Grading students. Machine learning is able to grade students by detaching human biases. They tried to improve the assessment of problem-solving in education. The second one is improving student retention. Then predicting student performance and the last one is testing students.

Xiaofeng et al. [9] studied select student characteristics to predict student pass rates in online education. This paper tried to predict student pass rates and tried to discover the most effective machine learning algorithm to find out more important student features affecting learning. They used three algorithms DT, SVM, DNN to construct a feature model. Mingjie et al. [10] studied that the dropout rate is a serious problem in online education or E-learning courses. They tried to predict a workable solution to stop dropout students.

In [1] and [11], the researcher has studied that COVID-19 is a concern to global education systems. For Corona disease, more than 100 countries closed schools. Their study shows that the effects of coronavirus on education are very horrible and they find numerous barriers that hamper students and instructors' interaction in online education in order to continue learning during COVID-19 lockdown. Also, they found rural area is not updated, not had digital skills, technological barriers, individual barriers, domestic barriers, institutional barrier, communication barrier, poor electricity, have network issues, lack of proper training, absence of finance, resistance to change, etc. are the extensive barriers for online education during the COVID-19 pandemic. So, in our work, we tried to find out the student's adaptability to online education during this pandemic situation.

III. Methodology

This section is designed to present the methodology of our work. It is divided into three parts: A. Data Collection, B. Data Preprocessing, C. Description of Models. Details of data collection and data preprocessing are described in subsections A and B respectively. And subsection C includes the description of the models used for prediction and analysis.

A. Data Collection

The origin of the dataset used in this research work is the online and offline survey. We collected student data on the different levels (university, schools, and colleges). Across all, we collected 1205 data from our survey(from December 10, 2020, to February 5, 2021) and the 14 attributes are age, gender, level, govt./non-govt. institution, location, IT student or not, educational background, load shedding level, internet quality, class-time, the economic condition of the family, device type used while attending classes, and institution's own LMS availability. A brief description of the data set attributes has been explained in table I.

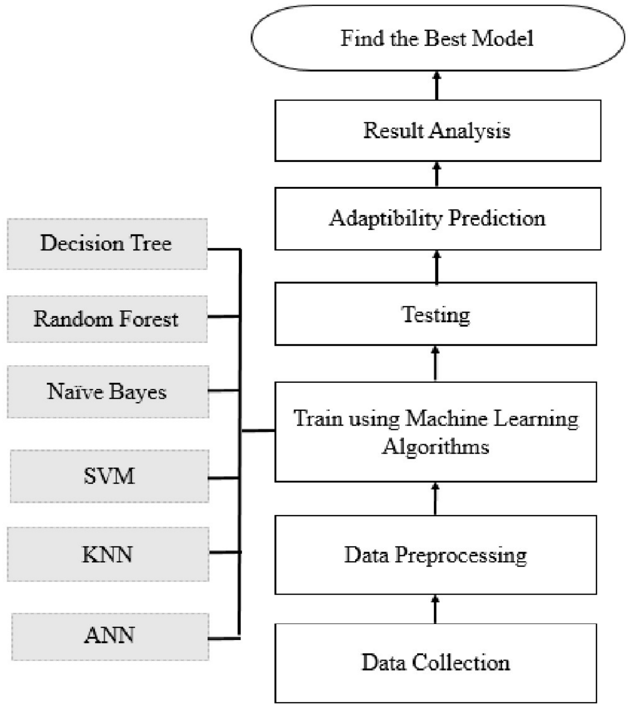

Fig. 1: Methodology for predicting the adaptability level in the online education

B. Data Prepossessing

In the data collection phase, we have used the Bengali medium for both online and offline surveys. As the raw dataset was in Bengali, first of all, we have manually converted the data into the English language for easy further processing. After that, we have converted the string attribute values into numeric values to make it more understandable and convenient for the models. We have used the digits 0, 1, 2 to represent the low, moderate, and high adaptability level respectively. Table I represents details of all 14 attributes we have used and their assigned class values in data preprocessing phase.

TABLE I: Attribute details with their probable values

Variable Name	Full-form	Variable Type	Probable Value
GT	Gender type	Independent	Girl(0), Boy (1)
ARTS	Age range of the student	Independent	Around 1 to 5 (0), 6 to 10 (1), 11 to 15 (2),16 to 20 (3), 21 to 25 (4), 26 to 30 (5), 30+(6)
EIL	Education institution level	Independent	School (0), College (1), University (2)
EIT	Education institution type	Independent	Non Government Ins (0), Government Ins (1)
SITS	Studying as IT student	Independent	No (0), Yes (1)
ISLT	Is student location in town	Independent	No (0), Yes (1)
LLS	Level of load shedding	Independent	Low (0), High (1)
FCF	Financial condition of family	Independent	Poor (0), Mid (1), Rich (2)
ITUMD	Internet type used mostly in device	Independent	2G (0), 3G (1), 4G (2)
DUMC	Device used mostly in class	Independent	Tab (0), Mobile (1), Computer (2)
NCT	Network connectivity type	Independent	Mobile Data (0), Wifi (1)
DCD	Daily class duration	Independent	0 (0), 1 to 3 Hours (1), 3 to 6 Hours (2)
IOLA	Institution's own LMS availability	Independent	No (0), Yes (1)
ALTS	Adaptability level of the student	Dependent	Low (0), Moderate(1), High (2)

The bar chart (Fig. 2) is a visual representation of our dataset which indicates the adaptability level of students of several age groups. While, the pie chart (Fig. 3) illustrates the scenario of adaptability in online education of several education levels(school, college, and university).

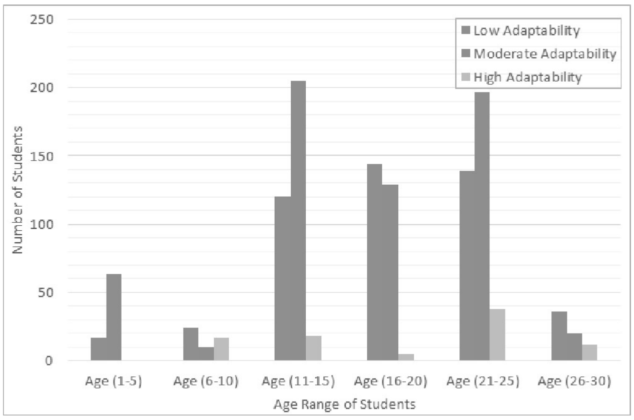

Fig. 2: Bar chart of adaptability of various age levels

C. Description of Models

There are lots of ML algorithms to predict the possible outcomes of student adaptability levels. We have trained and tested our dataset with various ML algorithms. The algorithms that were used for prediction and analysis are K-Nearest Neighbor, Decision Tree, Random Forest, Naive Bayes, Support Vector Machine, and Artificial Neural Network.

- Decision Tree: We used the Decision Tree classifier as one of our predictive models. A DT is built like a flow-chart tree structure in which each internal node represents a test of a feature and the leaf nodes represent the final corresponding output. It is the most commonly used algorithm because it provides a rapid and effective method of categorizing datasets that is easily understandable and implemented compared to other classification algorithms. The DT structure begins with a root node and involves splitting data into smaller and smaller subsets containing instances

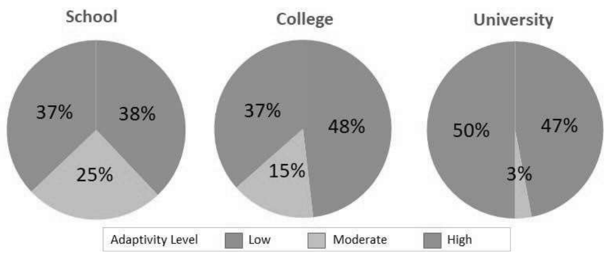

Fig. 3: Pie chart of adaptability of several education levels

of similar values. Therefore the entropy was often used in the DT algorithm to calculate the weight of the student homogeneous characteristics. Each decision tree in the ensemble is built based on the principle of recursive partitioning [12], which eventually showed a tree structure which is understandable by a human. The entropy[13] in set s can be calculated as follows:

$$H(S) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{1}$$

where p(i) defined the probability of an event occurring. Information gain was used to determine which feature will be split during the construction of the tree nodes to determine the best feature in the data set [14]. It can be expressed as follows:

$$Gain(S, A) = H(s) \sum \frac{|Sv|}{|s|} H(Sv) \tag{2}$$

- Random Forest: Random forest is a supervised learning algorithm capable of classification and regression. Just as its name suggests it is a set of decision trees that works as an ensemble. The main advantages of RF are that each tree protects each other from their own individual errors. The working principle of the random forest model is shown in Fig. 4 [15].
In this paper, we have used the ensemble based technique to combine the predictions from multiple machine learning algorithms together to make more accurate predictions than any individual model. The margin function that is used for RF is:

$$\bar{r}(X, D_n) = \mathbb{E}_{\Theta}[r_n(X, \Theta, D_n], \tag{3}$$

where Θ_1 and Θ_2 are the outputs of Θ and Θ and $\mathbb{E}_{\Theta}$, respectively, of a fortuitous variable.
- Naive Bayes: Naive Bayes is also a supervised learning algorithm that employs the Bayes theorem and is well-suited to very high-dimensional datasets. Naive Bayes works with the assumption of feature independence. Even in any case if these features are interdependent or upon the existence of the other features in the dataset, these are still considered independently in NB. This independence assumption simplifies the computation process and makes the algorithm considered naive. In our case, we believe that particular student adaptability for online education may have weak to strong dependencies. For example, a student with a good mobile network will have good adaptability in online learning. In this paper, we have used the Bayes Theorem as follow:

$$P(S_k|F) = \frac{P(F|S_k)P(F_k)}{P(F)} \tag{4}$$

Here, a feature vector is $F = (f_1, f_2, ..., f_n)$ and a class variable is denoted as S_k. Where,
- $P(F_k|X)$ the posterior probability.
- $P(F|S_k)$ the likelihood.
- $P(S_k)$ the prior probability of class.
- P(F) the prior probability of predictor.
Given the input, the prior probability of predictor P(X) is constant, hence we can get:

$$P(S_k|X) \propto P(S_k) \prod_{i=1}^{n} P(x_i|S_k) \tag{5}$$

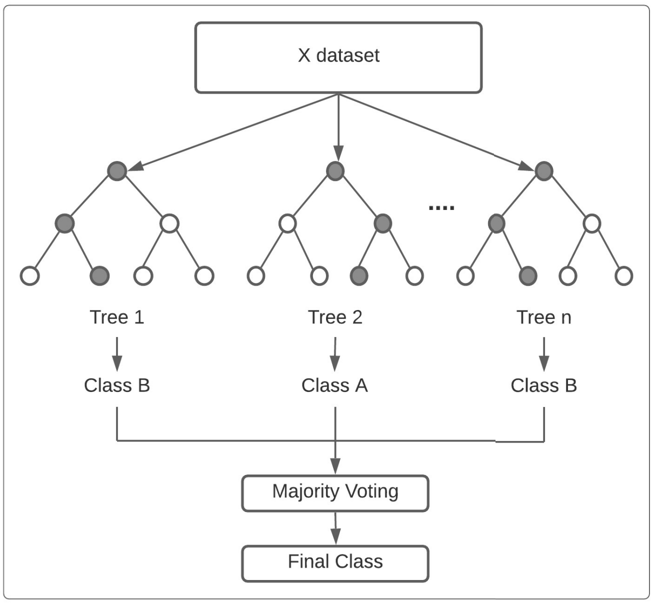

Fig. 4: Working Procedure of Random Forest Classifier

- Support Vector Machine (SVM): In this study, we have moreover utilized SVM for the prediction of the adaptability level of the students in spite of the fact that it does not back multi-classification natively. So for our datasets, we've implemented it with the aid of the LinearSVC one-vs-one approach. This OVO SVM works by dividing the primary dataset into a single binary classification for only the points of each pair of classes, ignoring the points of the third class, and then testing each point with all of the C(C-1)/2 models. In our case, the main purpose was to find the best maximum margin hyperplane for separating classes for an effective prediction.

- K-Nearest Neighbour(KNN): Among the many ML algorithms, KNN is one of the simplest algorithms. The reason for the KNN popularity can be attributed to its easy quick interpretation and low time for calculation [16]. KNN selects the number k of neighbors, calculates the distance function of k number of neighbors, processes the distance function to assign the class most frequently between its k nearest neighbors. The formula used to calculate the distance vector is:

$$d(q,p) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \tag{6}$$

where q_n denotes one observation's attribute values and p_n denotes another observation's attribute values.
- Artificial Neural Network (ANN): The artificial neural network (ANN) is a more advanced type of machine learning algorithm based on the idea of simulating the human brain. [17]. Artificial neural networks have nodes, which are similar to neurons of the human brain, and are coupled by many layers, effectively representing any complex input in a layered abstraction and producing a better output prediction. [18]. Therefore, in online education, the ANN algorithm can be used to predict student adaptation. Inside the neural network, neurons receive data from external sources and relay it to deeper levels. Based on the network's function, the hidden layers process the data and deliver the final output data to the last output layer. The simplified view of a neural network is shown in Fig. 5.

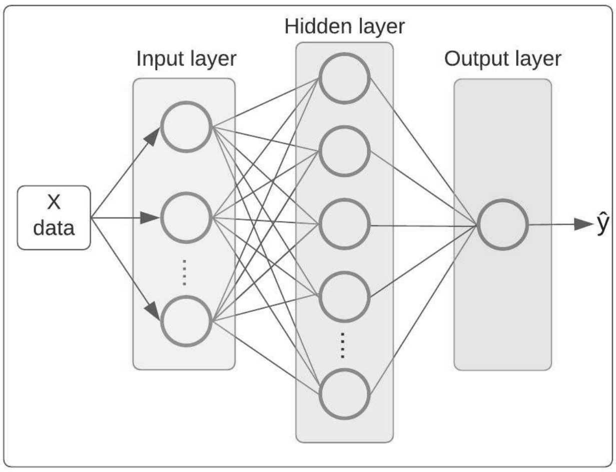

Fig. 5: Simplified view of ANN

In this study the activation function called Rectifier Linear Unit (ReLu) was used with the formula shown as follows.

$$f(x) = max(0,x) \tag{7}$$

The intent of this ReLu function is to elect whether the neuron should be triggered or not. The function generates the best non-linear decision boundary for a neuron's output. Because most real-life data is now non-linear, we need the neurons to learn these illustrations as well. [16]. Only 1000 iterations were used to train the data in this experiment.

IV. Result and Analysis

In this section, the obtained outcome of each classifier is described. This part has been divided into two sub-parts namely Performance Evaluation and Performance Analysis of the Applied Models. The in-detailed descriptions and analysis are given below.

A. Performance Evaluation

A number of measures can be used to assess the performance of ML models. Precision, Recall, F1 score, and Accuracy are the most important characteristics used to assess a model's performance. The value of the confusion matrix which is generated during the testing of the model is considered to calculate the score of the precision, recall, F1-Score, and accuracy. The formulas [19] that are used in these computations are given in equations 8, 9, 10, 11.

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{10}$$

$$F1-Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{11}$$

Here, True positives are represented by TP, true negatives by TN, false positives by FP, and false negatives by FN.

B. Performance Analysis of Applied Models

Six distinct machine-learning methods, including the K-Nearest Neighbor, Decision Tree, Random Forest, Naive Bayes, Support Vector Machine, and Artificial Neural Network, were deployed to forecast the various outcomes of the student's adaptability level in this study. The objective was to get a more reliable predictive model by comparing the performance of these models. We used 80% of the data as a training set and the remaining 20% as a test set. As we have worked with multi-class (three) problems, each classifier has generated a 3*3 matrix[20]. Generated confusion matrix by the applied models based on the test data is presented in TABLE II.

To assess the model's performance, we employed four sorts of evaluation pointers: Precision, Recall, Accuracy, and F1-Score. TABLE III shows the accuracy, precision, recall, and F1-score produced by the various models. It is observed that the Random Forest classifier has the top accuracy rate (89.63%) of the six machine learning models we employed for prediction, while the Support Vector Machine has the lowest accuracy rate (66.80%).

It is also observed that the Random Forest model achieved the highest class-wise precision that is 0.89, 0.90, 0.88 for the

TABLE II: Generated Confusion Matrix by Applied Models

	Model	Class Name	Predicted		
			Low	Moderate	High
Actual	DT	Low adaptability	93	8	2
		Moderate adaptability	11	103	1
		High adaptability	3	5	15
	RF	Low adaptability	97	4	2
		Moderate adaptability	11	104	0
		High adaptability	0	8	15
	NB	Low adaptability	72	30	1
		Moderate adaptability	23	88	4
		High adaptability	1	11	11
	SVM	Low adaptability	56	47	0
		Moderate adaptability	10	105	0
		High adaptability	0	23	0
	KNN	Low adaptability	76	23	4
		Moderate adaptability	17	97	1
		High adaptability	0	12	11
	ANN	Low adaptability	67	15	7
		Moderate adaptability	7	112	2
		High adaptability	1	9	21

class low adaptability, moderate adaptability, and high adaptability respectively. Here, the class-wise recall also has been calculated. For the class "Low adaptability", Random forest achieved the highest score that is 0.94. For the class "Moderate adaptability" and "High adaptability", ANN achieved the highest score that is 0.93 and 0.68 respectively. In the case of most of the real scenarios where imbalanced class distribution is the normal fact, F1-score is a better metric to evaluate the model. The more the F1 values indicate the better performance of the models.

TABLE III: Accuracy, Precision, Recall and F1-Score of the Applied Models

Model	Class Name	Accuracy	Precision	Recall	F1 Score
DT	Low adaptability	87.56%	0.87	0.90	0.89
	Moderate adaptability		0.89	0.90	0.89
	High adaptability		0.83	0.65	0.73
RF	Low adaptability	89.63%	0.89	0.94	0.92
	Moderate adaptability		0.90	0.90	0.90
	High adaptability		0.88	0.65	0.75
NB	Low adaptability	70.95%	0.75	0.70	0.72
	Moderate adaptability		0.68	0.77	0.72
	High adaptability		0.69	0.48	0.56
SVM	Low adaptability	66.80%	0.85	0.54	0.66
	Moderate adaptability		0.60	0.91	0.72
	High adaptability		0.00	0.00	0.00
KNN	Low adaptability	76.348%	0.82	0.74	0.78
	Moderate adaptability		0.73	0.84	0.79
	High adaptability		0.69	0.48	0.56
ANN	Low adaptability	82.99%	0.89	0.75	0.82
	Moderate adaptability		0.82	0.93	0.87
	High adaptability		0.70	0.68	0.69

From TABLE III, Random Forest achieved the highest class-wise F1-score that is 0.92, 0.90, 0.75 for the class of low adaptability, moderate adaptability, and high adaptability respectively.

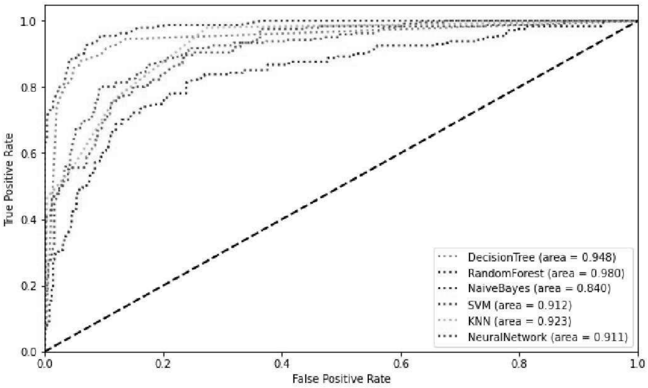

Fig. 6: Micro-Average ROC Curve

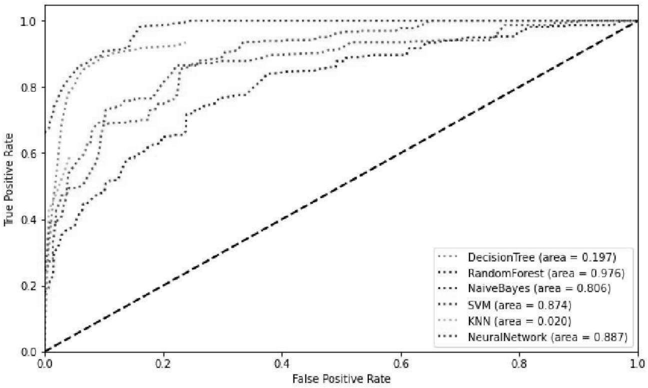

Fig. 7: Macro-Average ROC Curve

We have plotted (Fig. 6, Fig 7) the micro and macro average ROC curve to understand the quality of the predicted output by the models. Fig. 6 and Fig. 7 shows that the Random Forest classifier's Area Under the Curve (AUC) (0.980 for micro-average ROC curve and 0.976 for macro-average ROC curve) is larger enough compared to others model which indicates the better performance of the Random Forest over the other models.

From the above aforementioned analysis of the result, it is being perceived that the Random Forest model beat the other models in terms of performance indicators. From the micro-average and macro-average ROC curves, it is observed also that the Random Forest model generates the quality output. So, the Random Forest model can be a good choice for the students' adaptability prediction in online education.

V. Conclusions

In this paper, using Machine Learning, we have done forecasting the degree of adaptation among Bangladeshi students to online education during the COVID-19 epidemic. We have collected data from both online and offline surveys of several education levels. We have applied machine learning models for the prediction. We used K-Nearest Neighbor, Decision Tree, Support Vector Machine, Random Forest, Naive Bayes, and also Artificial Neural Network. Machine Learning models have shown satisfactory performance for the prediction. Among all the models used, Random Forest Classifier achieved the best prediction accuracy of 89.63%. Our study has revealed the overall perceptions about online education at the school, college, and university levels of students in Bangladesh perspective. This work could be beneficial for the decision-makers of the education sector to get a proper idea about the current online education system and the level of adaptability of the students to online education. In the future, we will try to explore how these socio-demographic factors affect the mental health of the students in the case of online education.

References

[1] E. M. Onyema, N. C. Eucheria, F. A. Obafemi, S. Sen, F. G. Atonye, A. Sharma, and A. O. Alsayed, "Impact of coronavirus pandemic on education," *Journal of Education and Practice*, vol. 11, no. 13, pp. 108–121, 2020.

[2] R. Afrouz and B. R. Crisp, "Online education in social work, effectiveness, benefits, and challenges: A scoping review," *Australian Social Work*, vol. 74, no. 1, pp. 55–67, 2021.

[3] M. Tabassum, S. E. Mannan, M. I. Parvej, and F. Ahmed, "Online education during covid-19 in bangladesh: University teachers' perspective," *Aquademia*, vol. 5, no. 1, p. ep21005, 2021.

[4] D. Wiliam, "Assessment in education: Principles, policy & practice," *Assessment in Education: Principles, Policy and Practice*, vol. 15, no. 3, pp. 253–257, 2008.

[5] A. P. Cavalcanti, R. F. L. de Mello, V. Rolim, M. André, F. Freitas, and D. Gaševic, "An analysis of the use of good feedback practices in online learning courses," in *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, vol. 2161. IEEE, 2019, pp. 153–157.

[6] V. L. Uskov, J. P. Bakken, A. Byerly, and A. Shah, "Machine learning-based predictive analytics of student academic performance in stem education," in *2019 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 2019, pp. 1370–1376.

[7] M. Ciolacu, A. F. Tehrani, R. Beer, and H. Popp, "Education 4.0—fostering student's performance with machine learning methods," in *2017 IEEE 23rd International Symposium for Design and Technology in Electronic Packaging (SIITME)*. IEEE, 2017, pp. 438–443.

[13] M. S. Zulfiker, N. Kabir, A. A. Biswas, P. Chakraborty, and M. M. Rahman, "Predicting students' performance of the private universities of bangladesh using machine learning approaches," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, pp. 672–679, 2020.

[8] D. Kučak, V. Juričić, and G. ambić, "Machine learning in education-a survey of current research trends." *Annals of DAAAM & Proceedings*, vol. 29, 2018.

[9] X. Ma, Y. Yang, and Z. Zhou, "Using machine learning algorithm to predict student pass rates in online education," in *Proceedings of the 3rd International Conference on Multimedia Systems and Signal Processing*, 2018, pp. 156–161.

[10] M. Tan and P. Shao, "Prediction of student dropout in e-learning program through the use of machine learning method." *International Journal of Emerging Technologies in Learning*, vol. 10, no. 1, 2015.

[11] R. E. Baticulon, J. J. Sy, N. R. I. Alberto, M. B. C. Baron, R. E. C. Mabulay, L. G. T. Rizada, C. J. S. Tiu, C. A. Clarion, and J. C. B. Reyes, "Barriers to online learning in the time of covid-19: A national survey of medical students in the philippines," *Medical science educator*, pp. 1–12, 2021.

[12] C. Strobl, J. Malley, and G. Tutz, "An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests." *Psychological methods*, vol. 14, no. 4, p. 323, 2009.

[14] "Machine learning 101: Decision tree algorithm for classification," *Available Online: https://www.analyticsvidhya.com/blog/2021/02/machine-learning-101-decision-tree-algorithm-for-classification/*, Access Date: 10/05/2021.

[15] A. A. Biswas, A. Majumder, M. J. Mia, I. Nowrin, and N. A. Ritu, "Predicting the enrollment and dropout of students in the post-graduation degree using machine learning classifier," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 11, pp. 3083–3088, 2019.

[16] "Understanding math behind knn (with codes in python)," *Available Online: http://www.datascribble.com/blog/machine-learning/understanding-math-behind-knn-codes-python/*, Access Date: 22/05/2021.

[17] W. S. Alaloul and A. H. Qureshi, "Data processing using artificial neural networks," in *Dynamic Data Assimilation-Beating the Uncertainties*. IntechOpen, 2020.

[18] J. Zhan and B. Dahal, "Using deep learning for short text understanding," *Journal of Big Data*, vol. 4, no. 1, pp. 1–15, 2017.

[19] A. A. Biswas, A. Majumder, M. J. Mia, R. Basri, and M. S. Zulfiker, "Career prediction with analysis of influential factors using data mining in the context of bangladesh," in *Proceedings of International Conference on Trends in Computational and Cognitive Engineering*. Springer, 2021, pp. 441–451.

[20] M. J. Mia, A. A. Biswas, A. Sattar, and M. Habib, "Registration status prediction of students using machine learning in the context of private university of bangladesh," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 1, pp. 2594–2600, 2019.