# Students adaptability level in online education

**Report for Project number-18**

**Submitted by Group-7**

**Group Member:**

1. Darshan Prajapati (B23CH1032)
2. Malav Parekh (B23MT1029)
3. Zalak Kanzaria (B23MT1023)
4. Prateek Agrahari (B23BB1033)

## Course instructor : Avinash sharma

## T. A name : Om Patel

**Project Title:** *Since as a beginner in machine learning it would be a great opportunity to try some techniques to predict the outcome of Students' Adaptability Level Prediction in Online Education using Machine Learning Approaches.*

**Abstract**: E-learning is a way of teaching where lessons are shared using technology, without the limits of time or place. It became very popular during the Covid-19 pandemic and has been used in all levels of education. Since many schools were not ready for the sudden switch to e-learning during the pandemic, it caused several technical problems. It's important for school leaders to understand how effective e-learning is so they can improve it for students. In this study, data from Kaggle was used to analyze how well students adapt to e-learning. Various machine learning and deep learning techniques were used to predict students' adaptability to e-learning.

**Introduction:** E-learning has not only gained prominence due to the pandemic but has also opened the doors to new educational possibilities. It allows students to learn at their own pace and from any location, making education more flexible and accessible. However, the rapid shift highlighted the unpreparedness of many educational institutions. Technical infrastructure issues such as inadequate software, lack of digital literacy among instructors, and challenges in managing large online classrooms posed significant hurdles. Many students faced difficulties due to limited access to devices or stable internet connections, creating disparities in learning experiences.

To address these issues, the study employed various machine learning algorithms, such as Random Forest, Logistic Regression, and Decision Tree Classifiers, among others. These algorithms analyzed student data to predict adaptability to the e-learning environment. This adaptability refers to how well students cope with the challenges of online education, including engagement with the platform, time management, and technical navigation. By accurately predicting adaptability, institutions can take proactive measures to support students who may struggle in an online learning environment. This helps in personalizing education, ensuring no student is left behind due to technological challenges.

The deep learning model incorporated in the study further enhanced the predictive capabilities. Unlike traditional machine learning algorithms, deep learning works by mimicking the human brain's neural networks, allowing it to process more complex data and make more accurate predictions. This level of analysis is crucial, as it can capture intricate patterns in how students interact with e-learning platforms, such as their performance on assignments, participation in virtual discussions, or engagement during live lectures.

Looking forward, e-learning is expected to continue playing a significant role in the future of education. The pandemic has demonstrated that hybrid learning models, which combine online and face-to-face instruction, may become a permanent fixture in education systems worldwide. Such models offer the flexibility of e-learning while retaining the benefits of in-person interaction. However, for this model to succeed, educational institutions must focus on improving digital infrastructure, providing training for educators, and ensuring equitable access for all students. Addressing these factors will help create a more resilient and adaptable educational system that can withstand future crises, whether they are caused by pandemics, natural disasters, or other disruptions.

In conclusion, while the pandemic accelerated the adoption of e-learning, it also revealed gaps that need to be filled for it to be effective and inclusive. By using machine learning and deep learning techniques to predict student adaptability, this study offers a data-driven approach to improving online education. The insights gained can guide decision-makers in implementing strategies that support both students and instructors, ultimately enhancing the quality and accessibility of education in a post-pandemic world.

---

**Literature review:**

The literature review explores e-learning as an educational model that, especially during the COVID-19 pandemic, enabled students from diverse locations to continue learning through online platforms. However, e-learning comes with challenges, such as technical limitations, lack of real-time interaction, assessment issues, and limited access to practical learning, which have impacted its effectiveness in technical subjects.

In conclusion, this study highlights the effectiveness of machine learning models in predicting the adaptability level of students to online education, particularly during the COVID-19 pandemic in Bangladesh. By analyzing data collected through extensive surveys across various educational levels, this research identified Random Forest as the most reliable model with an accuracy of 89.63%, outperforming other algorithms like KNN, Decision Tree, and ANN. These findings provide valuable insights for educational policymakers and institutions to understand students' adaptability challenges and make informed decisions to enhance the online learning experience. Future studies may delve deeper into how socio-demographic factors influence students' mental health in online education.[2]

In conclusion, the analysis indicates that students generally struggle to adapt to online education, particularly high school students facing significant academic pressures. The use of mobile devices with small screens for learning further exacerbates this challenge, reducing learning efficiency and effectiveness. Additionally, the absence of direct teacher supervision in online settings contributes to decreased student engagement and self-discipline, leading to procrastination and subpar performance. To enhance the quality and accessibility of online education, it is recommended that schools consider hybrid learning models and provide adequate technological resources, particularly for students from disadvantaged backgrounds.[3]

In conclusion, this research highlights the effectiveness of using a modified ensemble machine learning model to predict students' adaptability in online entrepreneurship education. The study demonstrates that combining multiple algorithms—such as Random Forest, Neural Networks, and CART—yields higher prediction accuracy and robustness, outperforming individual models. The proposed model's ability to handle complex data relationships and assess feature importance makes it an essential tool for educators to identify students requiring support, tailor interventions, and enhance online learning experiences. This work not only contributes to improved educational outcomes but also emphasizes the growing role of technology in fostering entrepreneurial skills and adaptability among students. Future research could explore the integration of such predictive models into user-friendly platforms for real-time adaptability assessments and personalized educational pathways.[4]

In conclusion, this study employed a range of machine learning and deep learning techniques to predict students' adaptability to e-learning during the COVID-19 pandemic. The dataset, sourced from Kaggle, was analyzed using various models, with the Decision Tree Classifier achieving 92% accuracy and a proposed deep learning model reaching 94.67% accuracy. These findings underline the potential of advanced predictive methods to guide educational institutions in understanding student adaptability and enhancing the e-learning experience. This research can aid decision-makers in addressing challenges in virtual learning and ensuring a supportive environment for diverse student needs.[5]

Research has focused on several aspects:

**1. Feature Selection in Predicting Student Success:** A study explored using decision trees (DT), support vector machines (SVM), and deep neural networks (DNN) to predict student pass rates, emphasizing which student features best indicate learning outcomes.

**2. Addressing Dropout Rates**: High dropout rates in e-learning are a concern. Researchers have aimed to develop methods to identify and retain at-risk students.

**3. COVID-19's Impact on Education:** The pandemic led to widespread school closures, exposing issues like limited internet access, inadequate training, and financial barriers, especially in rural areas, hindering student and teacher interaction.

**4. Enhancements in E-learning Models:** Studies have highlighted improvements in student satisfaction, similar performance levels on and off-campus, and the value of formative assessments to enhance learning experiences.

**5. Machine Learning in E-learning**: Machine learning has shown potential for grading, enhancing student retention, predicting academic performance, and conducting assessments. For example, algorithms like neural networks and SVMs have been used, but results indicate that e-learning is more effective than blended learning.

This review underscores efforts to optimize e-learning by addressing technical barriers, understanding factors affecting student adaptability, and leveraging machine learning to improve educational outcomes.

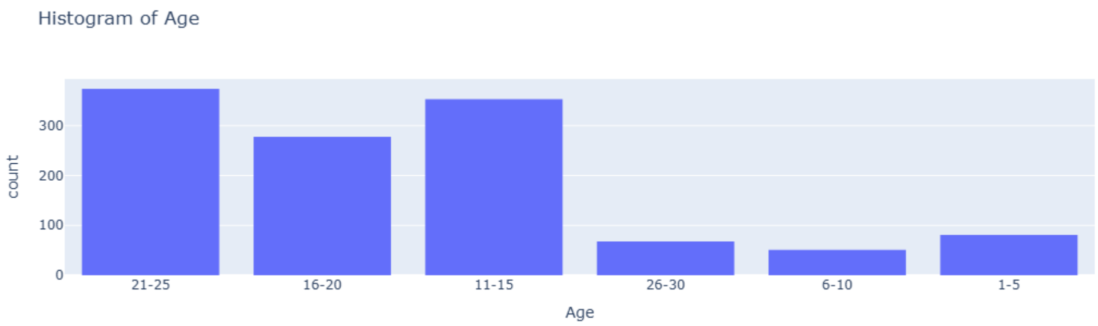
**METHODOLOGY:** Methodology is the systematic approach and set of principles guiding how research or analysis is conducted to achieve reliable results.

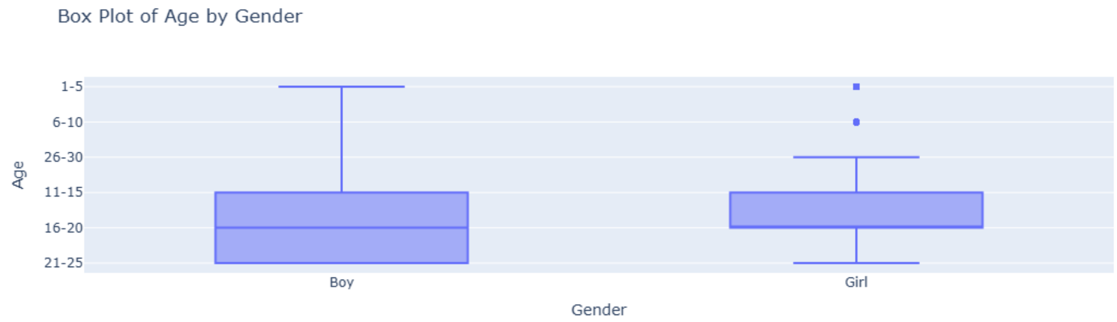
### A. Data Collection

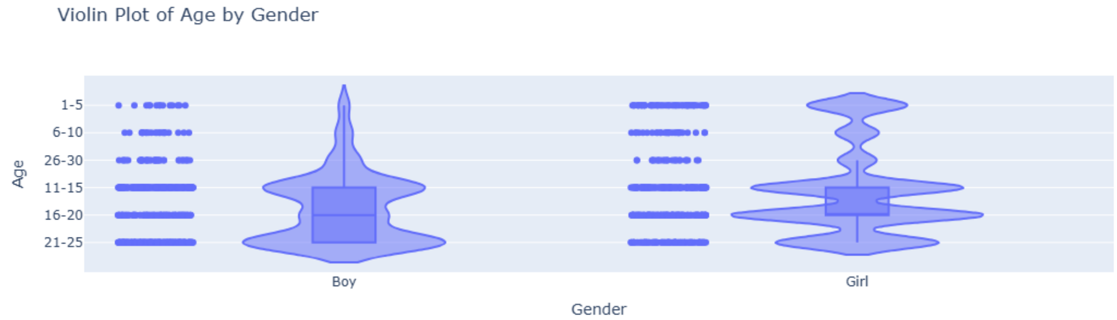
The data for this study was sourced from Kaggle, specifically from the dataset titled ***"Students Adaptability Level in Online Education"*** by Md Mahmudul Hasan Suzan. This dataset provides insights into various factors influencing student adaptability in e-learning environments.

| Features | Feature Type | Possible Values |
| --- | --- | --- |
| Gender | Input | Girl, Boy |
| Age range | Input | 21-25 , 11-15, 16-20 , 1-5 , 26-30, 6-10 |
| Education institution level | Input | School , University , College |
| Education institution type | Input | Non Government, Government |
| Studying as IT student or not | Input | No, Yes |
| Is student location in town | Input | Yes, No |
| Level of load shedding | Input | Low, High |
| Financial condition of family | Input | Mid, Poor , Rich |
| Internet type | Input | Mobile Data, Wifi |
| Network connectivity type | Input | 4G, 3G, 2G |
| Daily class duration | Input | 1-3, 3-6, 0 |
| Institution's own LMS availability | Input | No<br>Yes |
| Device | Input | Mobile, Computer, Tab |
| Adaptability level of the student | Output | Low<br>High |

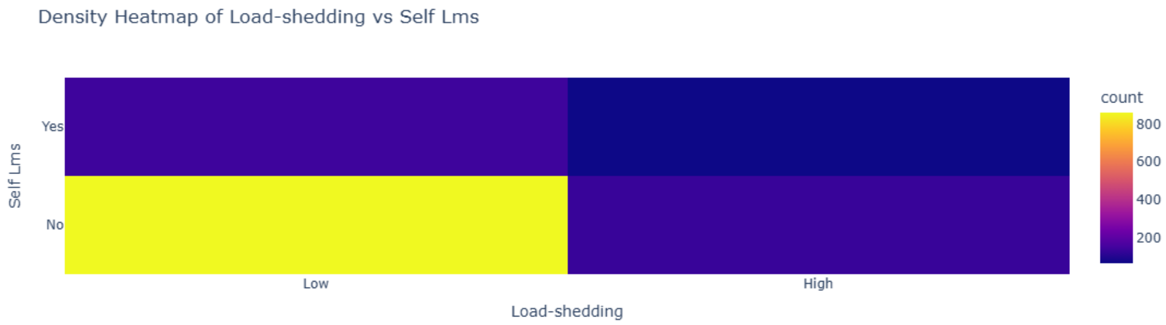
### B. Data Analysis

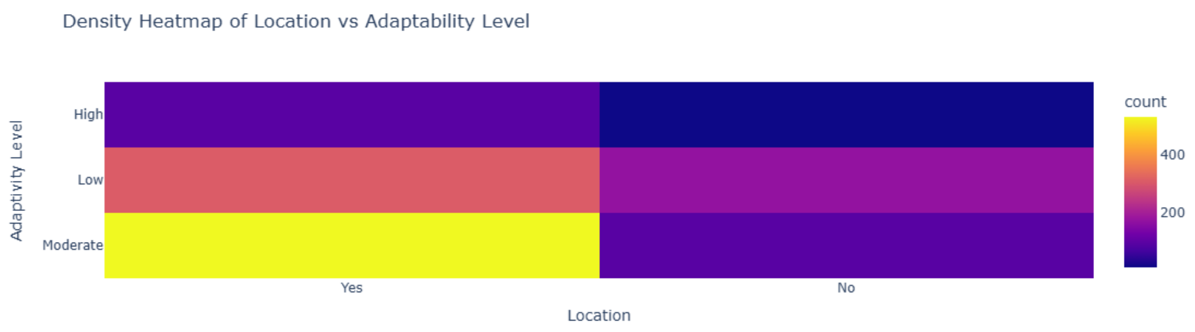


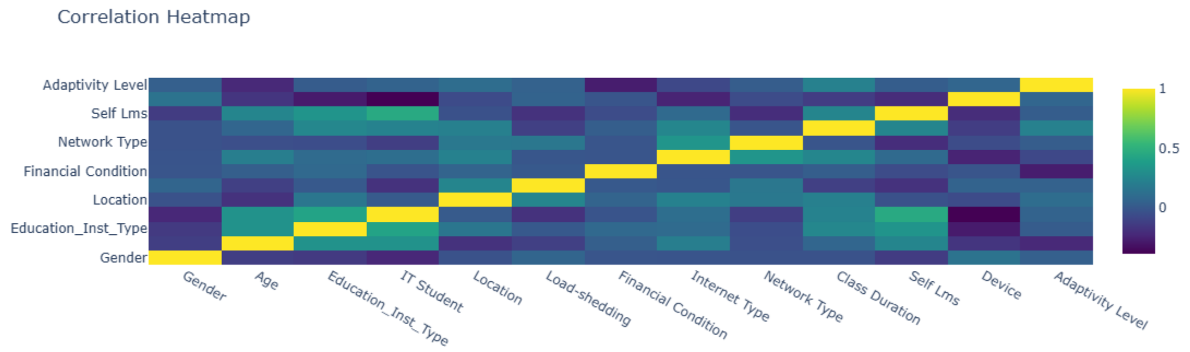



A violin plot of age by gender shows the distribution of age within each gender group. It highlights the density and spread of age data, allowing for easy comparison between genders. From this plot, you can identify differences in the central tendency, variation, and any potential skewness or outliers between male and female participants in the dataset.






## Data Preparation

All of the 14 features of the dataset are of categorical type: Gender, Education Level, Institution Type, IT Student, Location, Loadshedding, Financial Condition, Internet Type, Network Type, Class Duration, Self LMS, Device, Adaptability Level, and Age. Therefore, we labeled encoded categorical values

First merge 2 features Education Level and Institution Type and create new feature Education type

1. **Data Cleaning:**

   **Handling Missing Values:**Identify and address any missing values in the dataset. This can be done by removing records with missing values, imputing values using mean/median/mode, or using more advanced methods like K-nearest neighbors (KNN) imputation.

   **Removing Duplicates:**Check for and remove duplicate entries to ensure each customer is represented only once.

2. **Data Transformation:**

   **Normalization/Standardization:** Scale numerical features, such as Age, Tenure, and Total Spend, to ensure they contribute equally to the model. Normalization (scaling to a range) or standardization (scaling to a mean of 0 and standard deviation of 1) are common techniques.

**Encoding Categorical Variables:** Convert categorical features like Gender, location, education level , IT student and other as well into numerical format using techniques such as one-hot encoding or label encoding. This allows machine learning algorithms to process the categorical data.

3. **Feature Engineering:**

**Selecting Relevant Features:** Identify and retain only the most relevant features for the model. Techniques like Recursive Feature Elimination or feature importance analysis using tree-based models can help in this process.

4. **Outliers detection and treatment:**

**Identifying Outliers:** Use statistical methods or visualization tools (e.g., box plots) to detect outliers in numerical features that may skew the results.

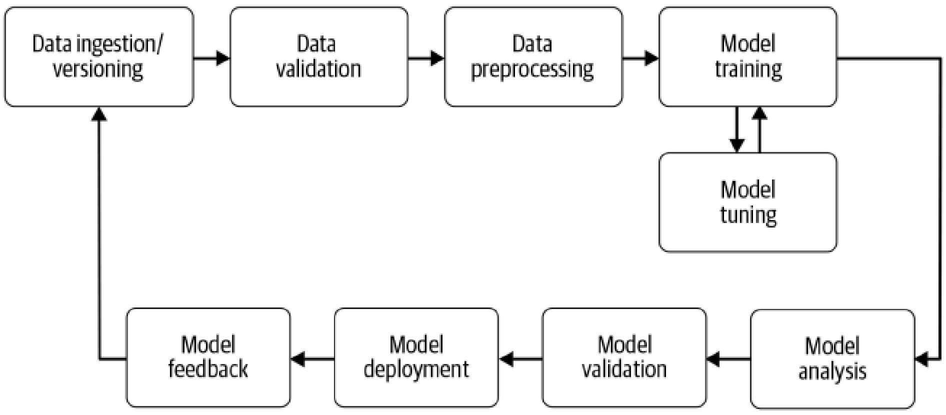
**Handling Outliers:** Decide whether to remove, cap, or transform outliers based on their impact on the dataset.

5. **Splitting the dataset:**

**Train-Test Split:** Divide the dataset into training and testing sets (commonly 75% for training and 25% for testing) to evaluate the model's performance on unseen Data.

**C. Description of Models Used in the Study**

We used several machine learning algorithms to predict student adaptability levels in e-learning education. Our dataset was trained and tested using five key algorithms: Support Vector Machine (SVM), XGBoost, Random Forest, Decision Tree, and Deep Learning models. These algorithms help us analyze and understand how well students can adapt to online learning environments.



## 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

To evaluate the performance of machine learning models, several measures can be used, with Precision, Recall, F1 Score, and Accuracy being the most important. These metrics are calculated using the confusion matrix, which is created during the model testing phase. The confusion matrix helps us understand how well the model is making predictions.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

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

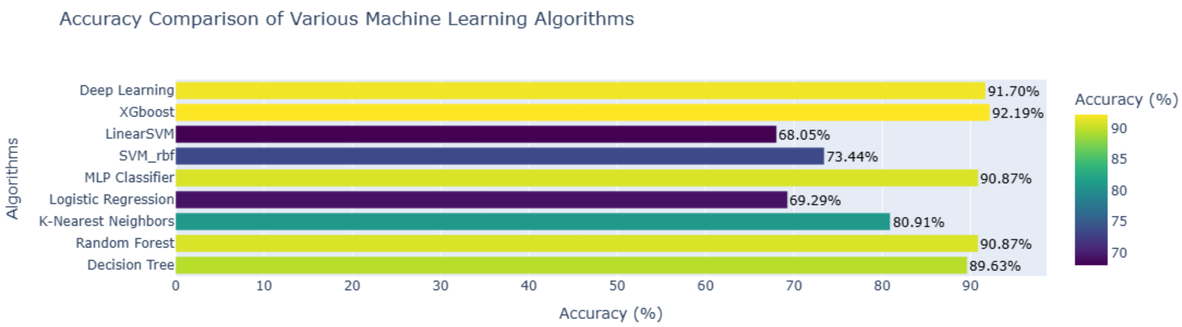
$$Precision = \frac{TP}{TP + FP}$$

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

In this study, we used several machine learning algorithms to predict student adaptability levels in e-learning. Specifically, we focused on five key algorithms: Support Vector Machine (SVM), XGBoost, Random Forest, Decision Tree, and a Deep Learning model. These algorithms were chosen for their effectiveness in analyzing how well students can adapt to online learning environments.

**Table 2**: Performance of the Machine and Deep Learning Algorithms

| Model Type | Model Name | Accuracy | Precision | Recall | F1_score | Time in Sec |
|---|---|---|---|---|---|---|
| Machine Learning | DecisionTreeClassifier | 89.63% | 0.90 | 0.8962 | 0.89 | 0.01 |
| | RandomForestClassifier | 90.87% | 0.91 | 0.9087 | 0.91 | 0.16 |
| | KNeighborsClassifier | 80.91% | 0.81 | 0.8091 | 0.81 | 0.00 |
| | LogisticRegression | 69.29% | 0.81 | 0.8091 | 0.81 | 0.03 |
| | LinearSVC | 68.05% | 0.71 | 0.6804 | 0.67 | 0.01 |
| | SVC | 73.44% | 0.77 | 0.7344 | 0.73 | 0.03 |
| | MLP Classifier | 90.87% | 0.91 | 0.9087 | 0.91 | 3.29 |
| | XGboost | 92.19% | 0.92 | 0.92 | 0.92 | 3.16 |
| | | | | | | |
| Deep Learning | Proposed Deep Learning Model | 91.70% | 0.91 | 0.91 | 0.91 | 3.14 |




## CONCLUSIONS

In this study, we explored the effectiveness of various machine learning and deep learning models in predicting students' adaptability to e-learning during the COVID-19 era. Using data from Kaggle, we evaluated five main models: Support Vector Machine (SVM), XGBoost, Random Forest, Decision Tree, and a Deep Learning model. Among these, the XGBoost Classifier achieved an Accuracy of 92.194 %, with corresponding Precision, Recall, and F1-Score at 92.00%. In comparison, our proposed Deep Learning model performed better, achieving an Accuracy of 91.29%, Precision of 91%, Recall of 91%, and F1-Score of 91%.

- **XGBoost** has the highest accuracy at **92.19%**, followed by **Deep Learning** at **91.7%**.
- **Random Forest** and **MLP Classifier** both perform well with **90.87%** accuracy.
- **LinearSVM** and **Logistic Regression** have lower accuracy, around **68-69%**.

This study provides valuable insights for educational decision-makers into students' adaptability to e-learning, which can help improve online learning strategies and tailor support to enhance student engagement and success in digital learning environments.

**References:**


1. www.kaggle.com
2. Md. Mahmudul Hasan Suzan, Nishat Ahmed Samrin, Al Amin Biswas, Md. Aktaruzzaman Pramanik, July 6-8, 2021 - IIT - Kharagpur
3. Advances in Education, Humanities and Social Science Research, “Analyzing students’ adaptability for online education”, Harbin University of Commerce, Harbin, 150028, China, Volume-5-(2023)
4. Amit Malik , Edeh Michael Onyema, Surjeet Dalal, Umesh Kumar Lilhore, Darpan Anand , Ashish Sharma , Sarita Simaiya, Array 19 (2023) 100303
5. Zakaria K. D. AlKayyali , Bassem S. Abu-Nasser , Ashraf M. Taha , Qasem M. M. Zarandah , Samy S. Abu-Naser, University Malaysia of Computer Science & Engineering (UNIMY), Cyberjaya, Malaysia