Student’s Adaptability Level in Online Learning Classifier

Aditya Jain Harshit Jain Vaibhav Wali Vasu Khanna

1. Abstract

Since Covid-19 online education has become very crucial. It has been due to the fact that it has enabled the people to continue their education even when the whole world had been hit by the pandemic. People had pursued their studies, sitting in front of electronic devices in their homes all over the world. The increasing importance of online education led to the emergence of the need to determine the student’s adaptability level in online education. The student’s adaptability level in online education depends on various factors which includes the characteristics of the student itself, student’s location, access to technology and internet etc. Different students with varying factors faced different difficulties in online education. So, it becomes extremely important for educational institutions to predict the adaptability level in online education for the student with given constraints w.r.t their location, internet and technology access, age etc. Using ML techniques, predicting the adaptability level beforehand is very beneficial for the students as it helps improve adaptability level further to get an optimal level for that scenario.

2. Introduction

There is a sudden substantial increase in the importance of online education with the advent of Covid-19 as it was the only medium left to pursue education for an individual. But the major drawback of online education is that there is not an effective one-on-one interaction between teachers and students, so it becomes extremely important to predict a student's adaptability level.

Different students have different adaptability levels depending on various factors related to a student viz. Gender, Age, Education level, Institution type, location, internet access, etc. Prediction of adaptability level is done using different Machine Learning techniques (logistic regression, Naive Bayes, SVM, Decision Trees, Random Forests, KNN and ANN) and their performance is compared to find the best classifier.

3. Literature Survey

Technology allows for virtual or remote learning. Thanks to technological advancements, we can now create online education systems. Aspects of education will be held in digitization under current circumstances. Students must accept the challenge of adapting to online education to make these changes. In the following discussion, we will provide an overview of the findings from the analysis of related works on online education.

In **[1]** and **[2],** the researchers have studied the improvement of the online education model.

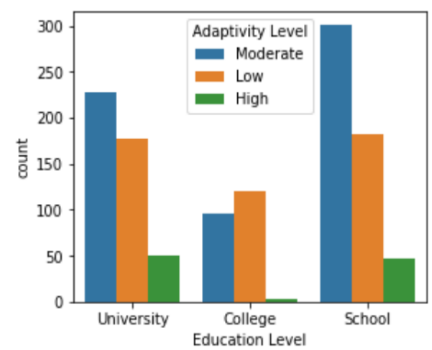
Online education decisions should be based on evidence of effectiveness rather than the assumption that face-to-face interaction is superior. A demonstration by Rojan et al. **[1]** helped us observe a significant difference in student performance and satisfaction and made us realize many benefits of online education for students. It has comparable off-campus and on-campus performances, offers, and student satisfaction, but communication has been difficult.

An investigation was done by William et al. **[2]** on how to improve the Online Education Model by using Machine Learning and Data Analysis in a Learning Management System (LMS), He also focused on formative assessment for better learning and the exploratory results show that 85% of students responded that they learn more in online education. The researchers primarily attempted to improve the assessment system for students and teachers. Also included are self and peer evaluations for students and teachers.

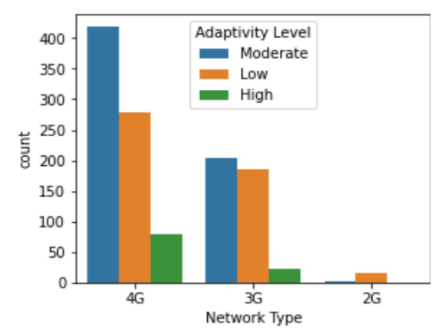
The researchers have examined how the ongoing pandemic is a worry for international education systems in **[3]** and **[4].** During the pandemic, a vast majority of the countries shut down their schools. The research studies in **[3]** and **[4]** demonstrates the terrible impacts of the coronavirus on education and identifies a number of obstacles that prevent interactions between students and instructors in online learning during the pandemic. They further took their research into depth and found out that as the rural areas didn't have adequate digital skills and had a lot of barriers including the technological barriers, domestic barriers, financial inadequacy, poor electricity, hence, they faced even more challenges as compared to the people in urban areas.

4A. **Data Description**

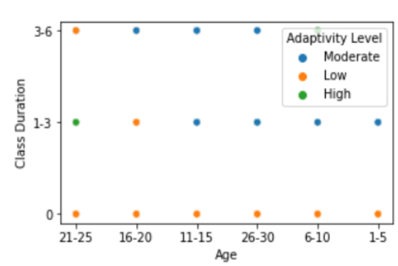
The dataset has 13 independent and 1 dependent features. There are 1205 samples in the dataset. Some of the features have binary values like Yes/No and Low/High while others have multiple values. Since the data is categorical, no outlier is present. A brief description of the dataset attributes with their possible values has been provided in the below f.



The above countplot depicts the distribution of different adaptivity levels as per the different types of education level.



The above countplot depicts the distribution of different adaptivity levels as per the different types of network.

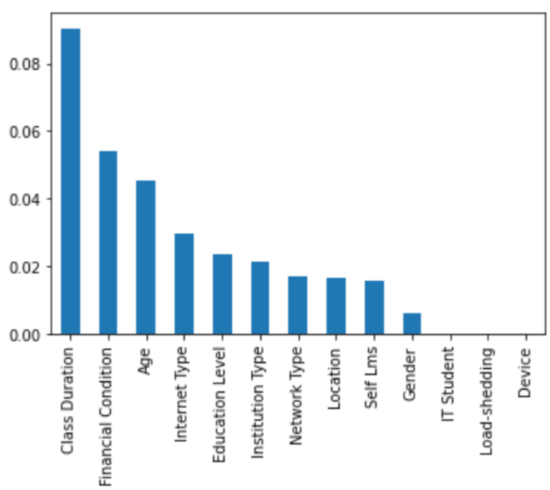


Scatter plot of class duration vs age groups

4B. **Data Preprocessing**

In the given dataset :

1. There are no **null values** present.
2. Feature Transformation: The data present is categorical, so the string values have been scaled to Integer for model prediction.
3. Feature selection: On the basis of Information Gain, some of the attributes like IT Student, Education level and Internet type which provide low information gain for Adaptivity level(dependent variable) are dropped.



The above plot shows the information gain by different independent variables on ‘Adaptivity Level’. It is clear that the features like IT Student, Load shedding and Device have nearly zero information gain and hence are dropped while selecting features to improve model performance.

Table1: Attribute details

Variable	**Variable Type**	**Possible values**
Gender	Independent	Girl(1), Boy(0)
Age	Independent	Around 1-5 (5), 6-10 (4), 11-15 (1), 16-20 (2), 21-25 (3), 26-30 (0)
Education Level	Independent	School(1), College(0), University(2)
Institution Type	Independent	Non Govt(1), Govt(0)
IT Student	Independent	No(0), Yes(1)
Location (Is Town)	Independent	No(0), Yes(1)
Load-shed ding	Independent	Low(1), High(0)
Financial Condition	Independent	Poor(1), Mid(0), Rich(2)
Internet Type	Independent	2G(0), 3G(1), 4G(2)
Network Type	Independent	Mobile Data(0), Wifi(1)
Class Duration	Independent	0 Hours(0), 1-3 Hours(1), 3-6 Hours(2)
Self lms	Independent	No(0), Yes(1)
Device	Independent	Tab(2), Mobile(1), Computer(0)
Adaptivity Level	Dependent	Low(1), Moderate(2), High(0)

5. **Methodology**

After preprocessing the data, we have used several Machine Learning models to predict the Adaptivity Level based on the features in the sample.
We have used the following classifiers:
1. Logistic Regression: It is a type of regression used in case of classification problems. It learns a linear relationship from the given dataset and then introduces a non-linearity in the form of the Sigmoid function.

2. **Gaussian Naive Bayes:** It is a type of classification model which uses Bayes algorithm. It is easy and fast in multiclass classification as it needs less training data. It is used to determine the benchmark performance of the models.

3. Random Forests Classifier: It is ensemble learning of Decision Trees(which provides interpretability and is non-parametric in nature) where some weak classifiers are combined and the prediction is done by majority voting for classification problems.

4. Decision Tree Classifier: A decision tree is a non-parametric supervised learning algorithm which provides interpretability while doing classification. At each level, a feature is chosen as per its information gain or entropy for classifying data and final classification is obtained at the leaf level.
5. SVM : A support vector machine (SVM) is a supervised learning algorithm to classify or predict data groups. The goal of the SVM is to determine the unique decision boundary known as Optimum Separating Hyperplane (OSH) that can segregate n-dimensional space into the required number of regions for classification.

6.KNN: The k-nearest neighbors' algorithm, also referred to as KNN or k-NN, is a supervised learning classifier that uses proximity to make classifications or predictions about the grouping of a single data point. It assumes the similarity between the new case/data and available cases and puts the new case into the category/cluster that is most similar to the available categories and then stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a good suite category by using K-NN algorithm.

7. ANN : Artificial Neural Network comprises several different layers viz. input layer, one or more hidden layer, and output layer. Each node connects to the node in the previous and next layer having a certain weight associated with it. The node can be activated or deactivated depending upon the output it generates.
6. **Results and Analysis**

Model	**Class Name**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
LR	Low	69.71%	0.88	0.39	0.54
	Moderate		0.74	0.56	0.64
	High		0.66	0.87	0.75

NB	Low	70.12%	0.58	0.61	0.59
	Moderate		0.72	0.61	0.66
	High		0.70	0.80	0.75
RF	Low	86.72%	1.00	0.67	0.80
	Moderate		0.88	0.85	0.86
	High		0.84	0.92	0.88

From the above observations, we can say that Random Forests give best performance amongst all the classifiers used for prediction as it has more accuracy, precision, recall, F1-score than other classifiers.

7. **Conclusion**

In this project, we have tried to forecast the student's adaptability level in online education using various Machine Learning models. We have used Logistic regression, Naive Bayes, Random Forest and Random Forest gives the best accuracy because it does ensemble learning on decision trees and decision trees works well on categorical data. Till now, we have just used the above three models and rest of the models viz. KNN, SVM, and ANN will be used further. This work would be beneficial for the educational decision makers to help them improve the quality of education for students.

Contribution

Dataset description: Aditya, Harshit
Model training: Aditya, Harshit
Analysis: Aditya, Harshit, Vaibhav, Vasu
Report: Aditya, Harshit, Vaibhav, Vasu
Literature Review & Slides: Vasu, Vaibhav

REFERENCES:

Dataset from Kaggle

https://www.kaggle.com/datasets/mdmahmudulhasansuzan/students-adaptability-level-in-online-education

Research Papers

[1] 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.

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

[3] 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.

[4] 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. Clariational survey of medical students in the philippines," Medical scieon, and J. C. B. Reyes, "Barriers to online learning in the time of covid-19: A nnce educator, pp. 1–12, 2021.