Chinthaka Abesekara

Dialog axiata PLC

Capstone Project

Effectiveness of online education

Table of Contents

[Introduction 2](#_Toc103425922)

[Data 2](#_Toc103425923)

[Methodology 3](#_Toc103425924)

[Results 4](#_Toc103425925)

[Conclusion 4](#_Toc103425926)

# Introduction

With the breakout of the COVID-19 pandemic, the whole world’s education structure collapsed. Therefore, most of the institutions started online education. Adapting to this digital education mode is affected by age limits, social background, etc. From this experiment we try to identify patterns in which students are adapting to Online education and create a Machine learning model to predict how best a student is adapting to online education.

# Data

For this I used the data set available in <https://www.kaggle.com/datasets/mdmahmudulhasansuzan/students-adaptability-level-in-online-education> which is based on the research paper ([Students' Adaptability Level Prediction in Online Education using Machine Learning Approaches](https://www.researchgate.net/publication/355891881_Students'_Adaptability_Level_Prediction_in_Online_Education_using_Machine_Learning_Approaches)) that is done based on an analysis in Bangladesh.

This dataset contains 14 columns as follows,

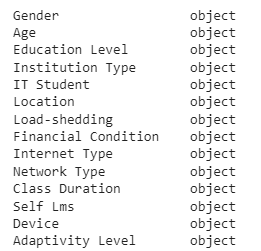
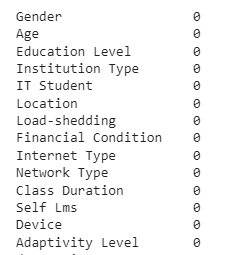
|  |  |
| --- | --- |
| Column | Description |
| Gender | Gender of the student |
| Age | Age range that student belongs to |
| Education Level | Whether student is in School, College or a University student |
| Institution Type | Whether student studies in Government or private sector |
| IT Student | Whether the student is an IT student or not |
| Location | Whether the student lives in an urban area or not |
| Financial Condition | Financial status of the student |
| Internet Type | Internet connection type the student uses to connect (WIFI or mobile data) |
| Network Type | Whether student is in 2G, 3G or 4G network on his/her classes |
| Class Duration | Class duration |
| Device | Whether student uses mobile phone, tab or laptop to connect |
| Adaptivity Level | How good the student has adapted to Online education |
| Load-shedding | Unknown |
| Self LMS | Unknown |

This data set consisted of 1205 rows and all columns had values (no missing values)

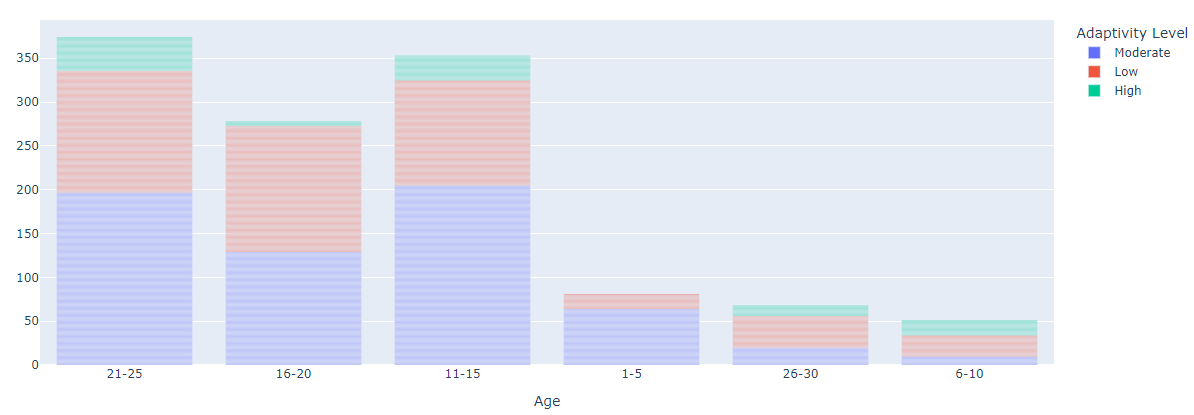
# Methodology

For this I used Google collab as the Notebook IDE.

In data inspection, I identified that all these features are in String categorical types and no missing data for any columns in the data.

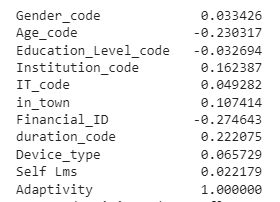
 

In addition to that when inspecting the sampling, it is noticed that majority of data was for students of age > 11. Hence, the sample is not balanced.



Next, to train a Machine Learning model we should have a numeric representation of data. So, I converted all fields to numeric using LabelEncoding and binary encoding.

When inspecting the correlation matrix of the data, it is noticed that age group, financial condition, class durations and location has significant contribution to the student adaptation to online education. Therefore, take these as feature variables (X).



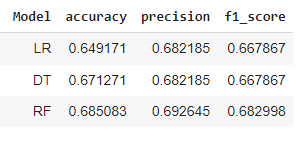
Subsequently, I split the data set to training data and testing data.

Next, I created a Logistic Regression model and trained it using a train data set. From this trained model I got prediction and prediction probabilities using test data. In this model I recorded accuracy, precision score and f1 score values.

Then similarly to create and train regression model, I created Decision Tree Classifier model and Random Forest Classifier models and recorded their accuracy, precision score and f1 score values separately.

After that, I created a data frame with above recorded accuracy, precision score and f1 score values separately against each model type. From this, we can identify which is the best model to be selected.

# Results



# Conclusion

With this result it can be concluded that Random Forest Classifier gives the best scores in all three measurements. So, it was selected as the model to be used.

# 