kaggle

KaggleX- Showcase

Cohort 3

Oct 2023

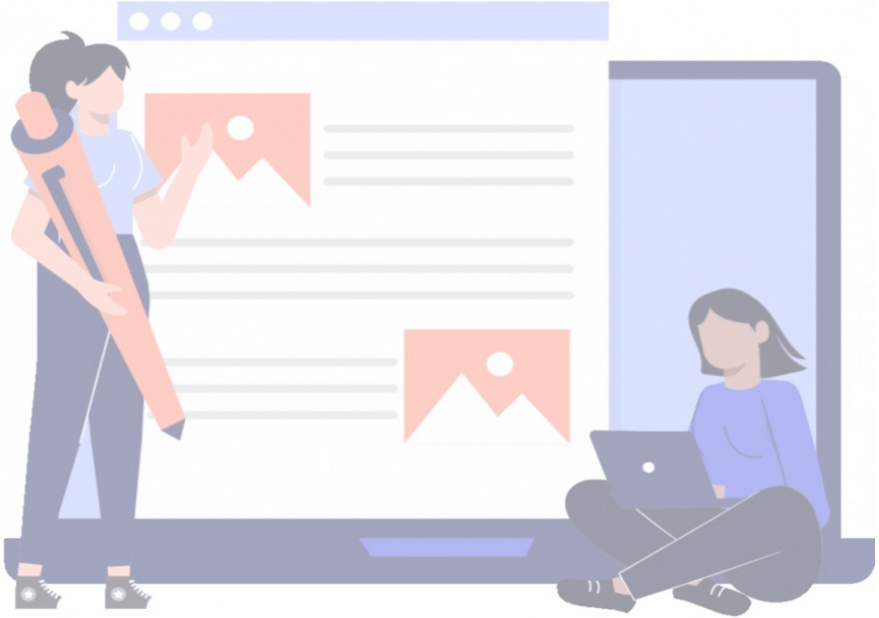
Predicting Student Adaptivity Level in Online Education using Multiple Machine Learning Models with Exploratory Data Analysis (EDA)

Garima Sharma

Mentor : Moushmi Dasgupta

AGENDA

1. Background
2. What I have Learned?
3. Understanding the Problem Statement
4. Data Collection and Understanding
5. Data Pre-Processing
6. Exploratory data analysis
7. Model Training
8. Choose best model

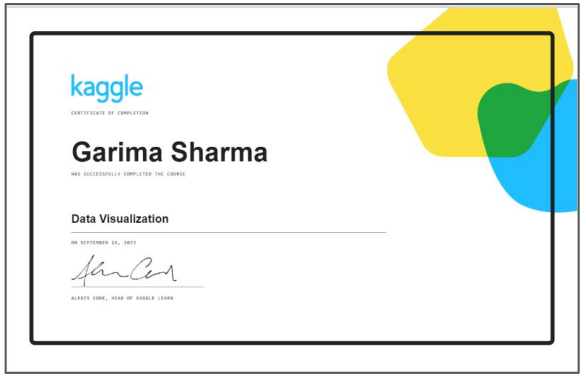


Background

- Bachelor’s in Artificial Intelligence and Robotics from Dayalbagh Educational Institute, Agra, Uttar Pradesh, India completed in May, 2023.
- I really enjoy Data Analytics and I'm eager to learn more. I'm planning to pursue a master's degree in this field to delve even deeper into it.
- I completed certifications from Kaggle Learn, where I delved into various architectures and put them into practice using Kaggle Compiler.
- And, I've also begun my journey on the Machine Learning Engineer Track through Google Cloud Skills Platform. This will further expand my knowledge of machine learning and provide valuable hands-on experience.
- Presently working as Data Analyst Intern in Pixel Truth Brand Safety in Agra, India.
- Special thanks to my mentor, Mrs. Moushmi Dasgupta, whose guidance and support have been invaluable throughout this journey.

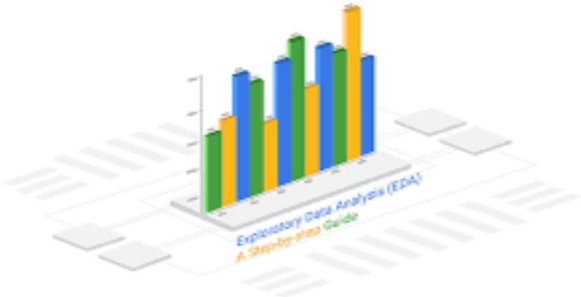
Certifications

I am pleased to share my accomplished certificates from Kaggle Learn.



What I have Learned?

- Different Exploratory Analysis Methods.
- Different basic Machine Learning Algorithms and its practical use.
- Different advanced Machine Learning Algorithms and its practical use.



Understanding the Problem Statement

- Since the emergence of the Covid-19 virus pandemic, the world of education seems to have slow down for many underprivileged student.
- Online education is a solution for implementing learning during the pandemic or other calamities such as war etc.
- It is required to find the adaptivity of students pursuing online education.
- We need to analyze the dataset that measures the extent to which students rely on their adaptive skills in the context of online education.

Data Collection and Understanding

Data Understanding

Column Group	Column Name	Data Type
User Profile	Gender, Age. Financial Condition, IT Student, Location, Load Shedding	Categorical Data
Education Profile	Education Level, Institution Type, Self LMS, Class Duration	
User Service	Internet Type, Network Type, Device	
Target Column	Adaptivity Level	

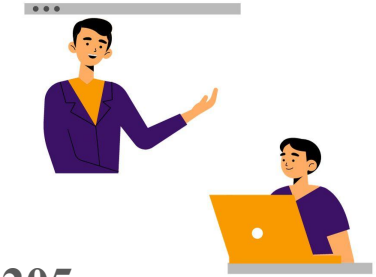
Dataset

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

Data Pre-Processing

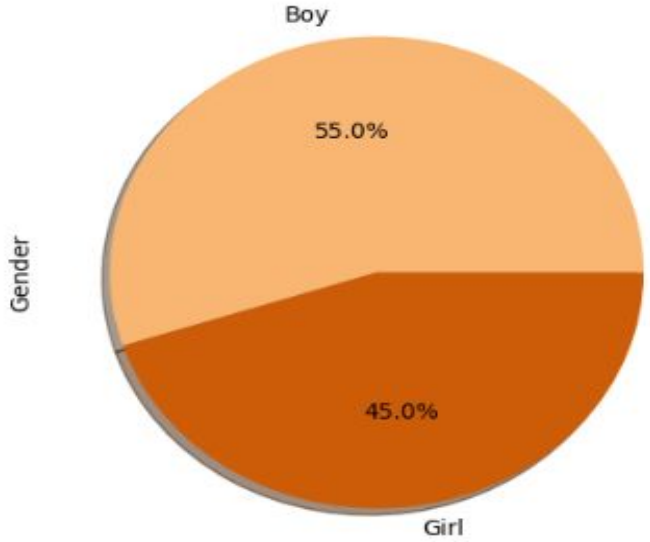
```
1 df.info()
```

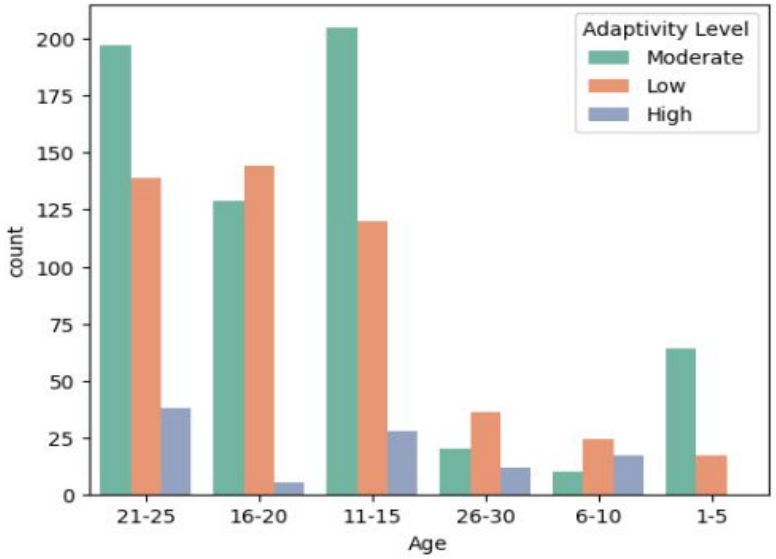
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1205 entries, 0 to 1204
Data columns (total 14 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Gender               1205 non-null   object
 1   Age                  1205 non-null   object
 2   Education Level      1205 non-null   object
 3   Institution Type     1205 non-null   object
 4   IT Student           1205 non-null   object
 5   Location             1205 non-null   object
 6   Load-shedding        1205 non-null   object
 7   Financial Condition  1205 non-null   object
 8   Internet Type        1205 non-null   object
 9   Network Type         1205 non-null   object
 10  Class Duration       1205 non-null   object
 11  Self Lms             1205 non-null   object
 12  Device               1205 non-null   object
 13  Adaptivity Level     1205 non-null   object
dtypes: object(14)
memory usage: 131.9+ KB
```



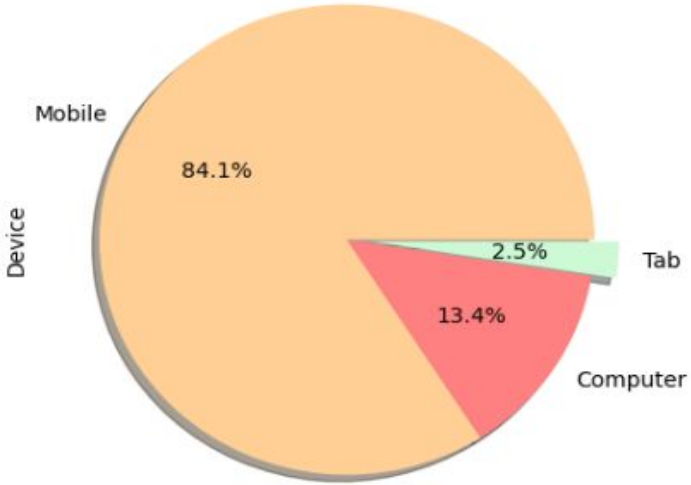
- The total data recorded is 1205
- There are 13 category features and one target feature.
- The data set has no missing values.
- All data type is “object”.

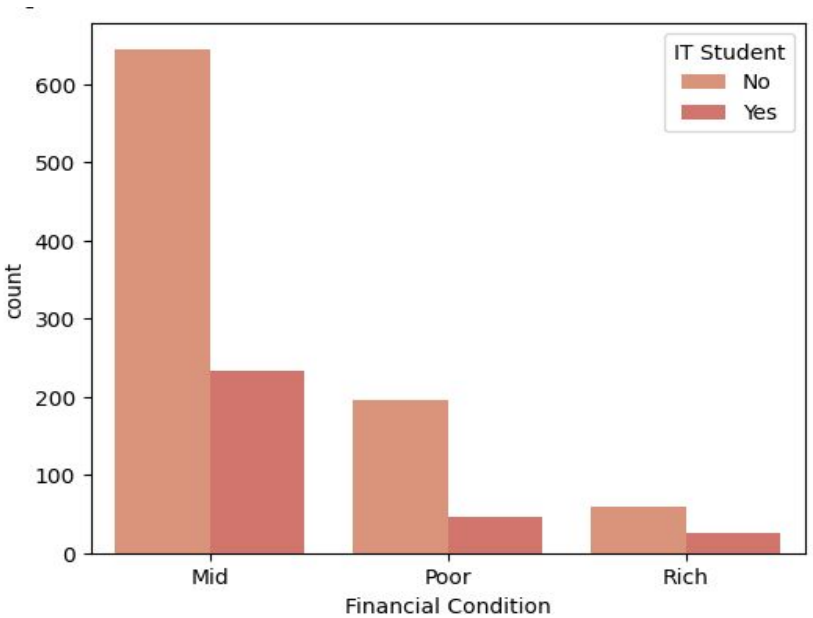
Exploratory data analysis

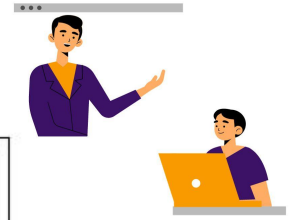




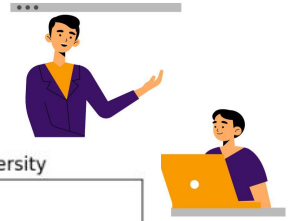
Exploratory data analysis

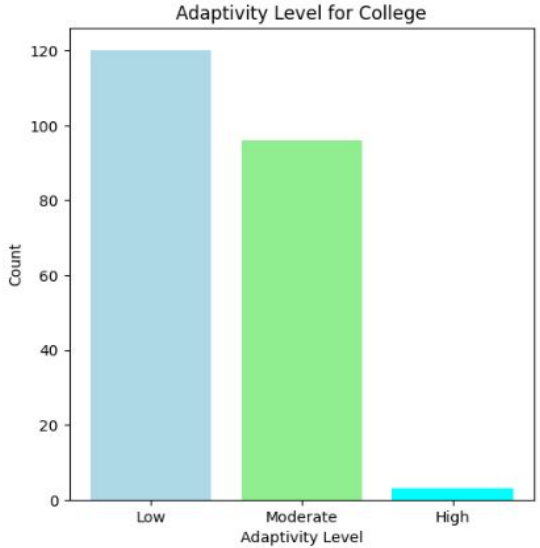


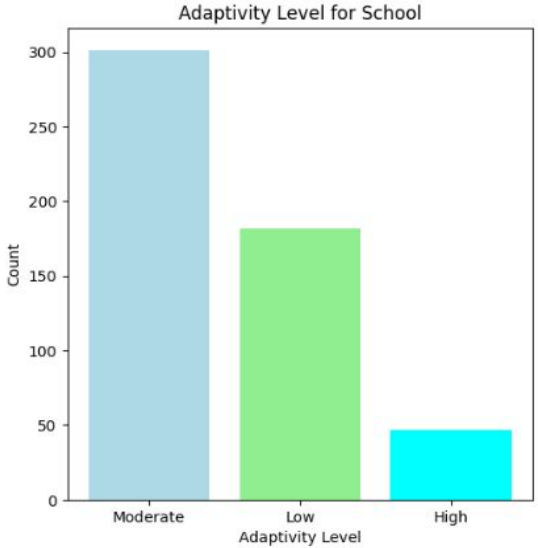


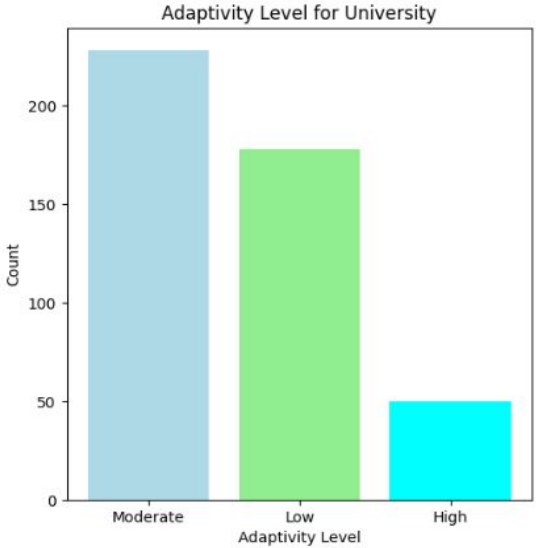


Exploratory data analysis









Classification of Categorical Dataset Using Machine Learning-Based Algorithm

Basic Classification Algorithm is used :

1. SVC
2. Gaussian Naives Bayes
3. Logistic Regression
4. Decision tree
5. Random forest
6. KNN

Advanced Classification Algorithm is used :

1. XGboost
2. Ensemble-Voting Classifier
3. Gaussian Process Classifier
4. Neural Network
5. Adaboost
6. Cat boost

Comparative Accuracy Analysis of Each Model

Basic Classification Algorithm Accuracies

S. No.	Model	Test Accuracy
1.	Logistic Regression	0.68
2.	Gaussian Naive Bayes	0.68
3.	Support Vector Classification	0.71
4.	K-Nearest Neighbors (KNN)	0.78
5.	Decision Tree Classifier	0.89
6.	Random Forest Classification	0.91

Comparative Accuracy Analysis of Each Model

Advanced Classification Algorithm Accuracies

S. No.	Model	Test Accuracy
1.	XGBoost	0.76
2.	Neural Network	0.87
3.	Ensemble-Voting Classifier	0.89
4.	Gaussian Process Classifier	0.89
5.	CatBoost	0.90
6.	AdaBoost	0.92

Choose Best Model

Top 5 Highest Model Accuracies:

1. AdaBoost with 92%.
2. Random Forest with 91%.
3. CatBoost with 90%.
4. Gaussian Process Classifier with 89%.
5. Decision Tree Classifier and Ensemble - Voting Classifier with same 88%.

Project Links

- **Google Colab Notebook – https://colab.research.google.com/drive/1X52hazlrJRDl7vkXYX4GY9N9pnyV6mDQ?usp=sharing**

- **Kaggle Notebook – https://www.kaggle.com/code/garimasharma05/analysing-student-performance-based-onlineeducat**

- **Youtube Link – https://youtu.be/tkrBluoVPE4?si=E8EHWtB-QdA6sm4c**

I would like to thank kaggle for giving me this opportunity to learn more and expand my horizons. The lessons I have learnt and the community I have gain has truly helped me grow and love the field of Data Analytics even more.