"Predicting Student Adaptivity at Online Education System"

A

Report

on

Project based learning

For the partial fulfilment of

Master of Computer Application

Semester – III (2024-25)

By

Ankur Singh Chauhan
(MCAN1CA23039)

Under the Supervision of

Dr. Sanjay Jain

Professor, Department of CSE, SOET,

ITM University

Title: "Predicting Student Adaptivity at Online Education System"

Abstract:

The objective of this study is to solve a classification problem by utilizing a heterogeneous dataset with 1,205 entries spanning 14 features. The dataset contains information on age, gender, education level, kind of institution, and internet connectivity, among other demographic, educational, and technological characteristics. The main goal is to create and assess classification models that may be used to forecast adaptivity level.

The study's findings will help comprehend the variables affecting the classification results and offer useful insights. Potential Adaptivity Level ramifications arise from the findings. In order to improve the model's performance and applicability, future work will investigate possible enhancements and additional refinements.

Dataset Source:

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

Problem Type: Classification

Dataset Information

Columns:

- **Gender**: Gender of the student.
- **Age**: Age ranges
- **Education Level**: Different education levels represented.
- **Institution Type**: Types of institutions.
- **IT Student**: Whether the student is an IT student or not.
- **Location**: Is student location in town.
- **Load-shedding**: Impact or presence of load-shedding.
- **Financial Condition**: Financial status of student's family.
- **Internet Type**: Types of internet connections.
- **Network Type**: Types of network connections.
- **Class Duration**: Duration of classes.
- **Self LMS**: Institution's own LMS availability.
- **Device**: Types of devices used.
- **Adaptivity Level**: Level of adaptivity reported.

Target Feature: *Adaptivity Level*

Snapshot of Dataset:

```
data.head()
```

	Gender	Age	Education Level	Institution Type	IT Student	Location	Load-shedding	Financial Condition	Internet Type	Network Type	Class Duration	Self Lms	Device	Adaptivity Level
0	Boy	21-25	University	Non Government	No	Yes	Low	Mid	Wifi	4G	3-6	No	Tab	Moderate
1	Girl	21-25	University	Non Government	No	Yes	High	Mid	Mobile Data	4G	1-3	Yes	Mobile	Moderate
2	Girl	16-20	College	Government	No	Yes	Low	Mid	Wifi	4G	1-3	No	Mobile	Moderate
3	Girl	11-15	School	Non Government	No	Yes	Low	Mid	Mobile Data	4G	1-3	No	Mobile	Moderate
4	Girl	16-20	School	Non Government	No	Yes	Low	Poor	Mobile Data	3G	0	No	Mobile	Low

Columns Info & Dtype:

```
<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
```

Shape:

```
data.shape
```

```
(1205, 14)
```

Literature Survey:

1. Qusay AL-Btoush: Accuracy = 91.4% with XGBClassifier
   Used these models: LogisticRegression, DecisionTreeClassifier, SVC
   RandomForestClassifier, KNeighborsClassifier, MLPClassifier, XGBClassifier
   Notebook: https://www.kaggle.com/code/qusaybtoush1990/students-adaptability-accuracy-91-4/notebook

2. Wonduk: Accuracy = 89.9% with tuned XGBClassisfier
   Used these models: LogisticRegression, KNearest, RandomForest, XGBClassifier, CatBoostClassifier
   Notebook: https://www.kaggle.com/code/wonduk/predict-eda-on-adaptivity-in-online-education

3. Georgy Zubkov: Accuracy = 92% with XGBClassifierr
   Used SMOTE for oversampling the data, and RandomForestClassifier, KNeighborsClassifier, SVC, LogisticRegression ,XGBClassifier
   Notebook: https://www.kaggle.com/code/georgyzubkov/students-adaptability-eda-and-mini-ml/notebook

4. NoNameDataScientist: Accuracy = 86% using KNN
   Used undersampling and oversampling.
   Used thse models: LogisticRegression, KNearest, RandomForest, AdaBoostClassifier, VotingClassifier, MultinomialNB
   Notebook: https://www.kaggle.com/code/noname666666/adaptivity-level-prediction/notebook

5. Vishnu U: Accuracy = 93.33% using DecisionTreeClassifier
   Used SMOTE oversampling
   Used these models: DecisionTreeClassifier
   Notebook: https://www.kaggle.com/code/vishnu0399/adaptability-analysis-of-online-education-system/notebook

Exploratory Data Analysis

1. Distribution of Features
2. Features Values Count with respect to Adaptivity Level
3. Correlation Heatmap
4. Relationship Between Features
5. Class Balance Check
6. Boxplots to analyse the Distribution of Numerical Features Across Categories

Objectives of the Proposed Work

1. Creation of model which helps in determining whether the student can adapt to online education or not.
2. Getting to know about the feature importance
3. Learn new techniques for preprocessing data
4. Reduce problems related to online education

Implementation Plan

1. **Preprocessing**
    - Data Acquisition
    - Label Encoding
    - Standard Scaling

2. **EDA**
3. **Handing Class Imbalance**
    - Use undersampling or oversampling to balance the dataset
    - Use SMOTE or other related technique

4. **Model Building**
    - Split into dependent and independent variables
    - Train test split
    - Train the various models

5. **Performance Evaluation**
    - Accuracy Score
    - Classification Report
    - Confusion Matrix

6. **Model Optimization**
    - Hyper Parameter Tuning
    - Apply PCA if required
    - Feature selection
    - Ensemble or Hybrid Models

Proposed ML Models

1. XGBClassifier
2. CatBoostClassifier
3. AdaBoostClassifier
4. RandomForestClassifier
5. DecisionTreeClassifier
6. VotingClassifier
7. StackClassifier

Tools Used

Anaconda Framework, Jupyter NoteBook, Pandas, ScikitLearn, Imbalance, Matplotlib, Seaborn

Expected Outcomes

1. **Accuracy**
   To get better accuracy than the previous works done. If not achieved, then at least get the max available.
   To achieve accuracy better than 90%

2. **Feature Importance**
   Understanding Key Predictors, which can help institutions to know which areas to improve.

3. **Class Imbalance Management**
   After applying balancing techniques, the model will handle minority classes better, and will be unbiased.

4. **Unbiased and Fair Model**
   We expect to create a model which will be unbiased and fair in its overall predictions.

References

1. Suzan, M.M.H.; Samrin, N.A.; Biswas, A.A.; Pramanik, A. Students' Adaptability Level Prediction in Online Education Using Machine Learning Approaches. In Proceedings of the 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 6–8 July 2021;

2. Loderer, K.; Rinas, R.; Daumiller, M. Student adaptability, emotions, and achievement: Navigating new academic terrains in a global crisis. Learn. Individ. Differ. 2021, 90, 102046

3. Peng, H.; Ma, S.; Spector, J.M. Personalized Adaptive Learning: An emerging pedagogical approach enabled by a smart learning environment. In Foundations and Trends in Smart Learning; Lecture Notes in Educational Technology; Springer: Singapore, 2019;

4. Haleem, A.; Javaid, M.; Qadri, M.A.; Suman, R. Understanding the role of digital technologies in education: A review. Sustain. Oper. Comput. 2022,

5. Jang, Y.; Choi, S.; Jung, H.; Kim, H. Practical early prediction of students' performance using machine learning and eXplainable AI. Educ. Inf. Technol. 2022,

6. Adadi, A.; Berrada, M. Peeking Inside the Black-Box: A survey on Explainable Artificial Intelligence (XAI). IEEE Access 2018.