IML PROJECT

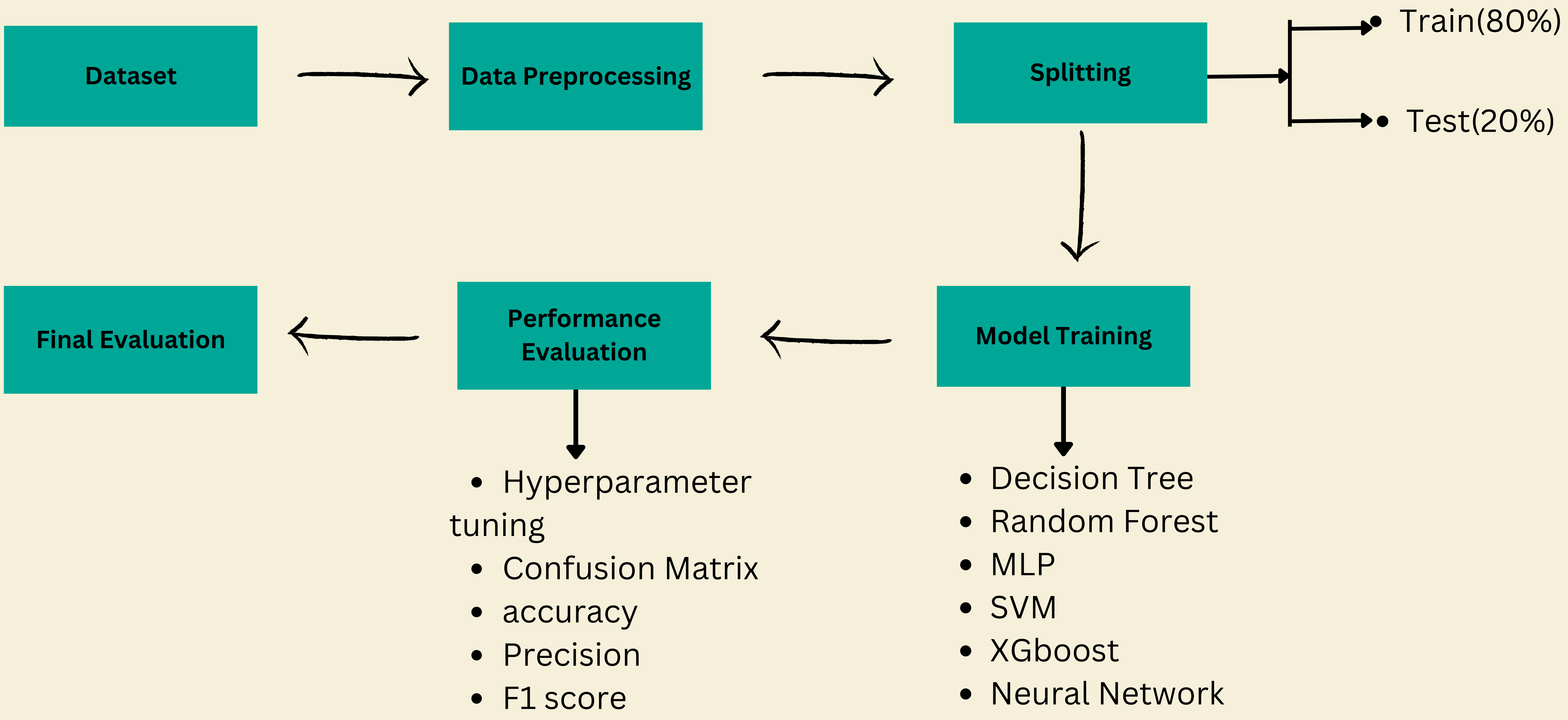
STUDENTS ADAPTABILITY LEVEL IN ONLINE EDUCATION

PROJECT OVERVIEW

This project focuses on predicting students' adaptability to e-learning, especially important in the shift towards online education. By understanding adaptability, educational institutions can provide tailored support to enhance learning outcomes.

Our approach combines data analysis and machine learning techniques to identify key factors influencing adaptability. We conduct Exploratory Data Analysis (EDA), apply various predictive models, and evaluate their performance to determine the best model for accurate predictions.

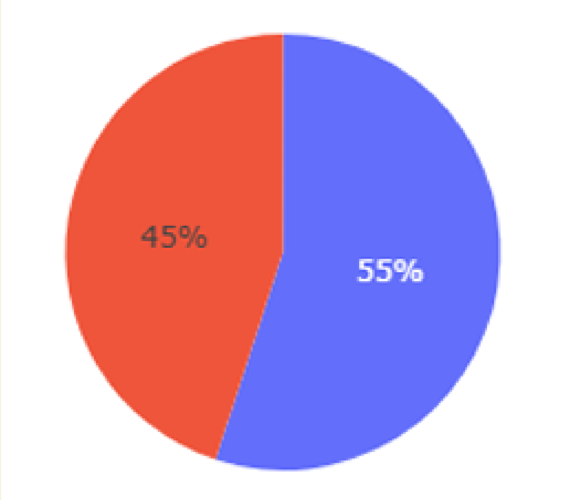
Flowchart for ML Model



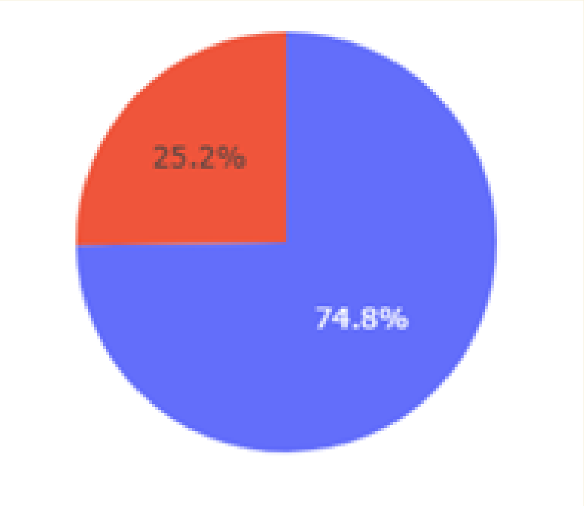
EXPLORATORY DATA ANALYSIS (EDA) AND GRAPH ANALYSIS - PART 1

Initial dataset analysis shows key features like gender, institution type, and internet access, which impact adaptability. This helps in understanding the distribution and trends within the dataset.

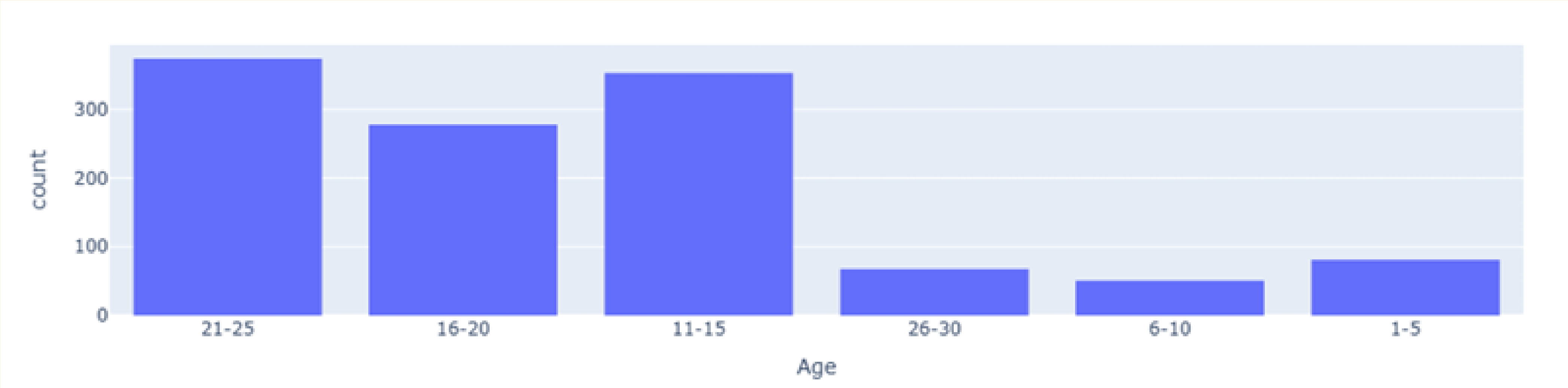
01 Distribution of Gender



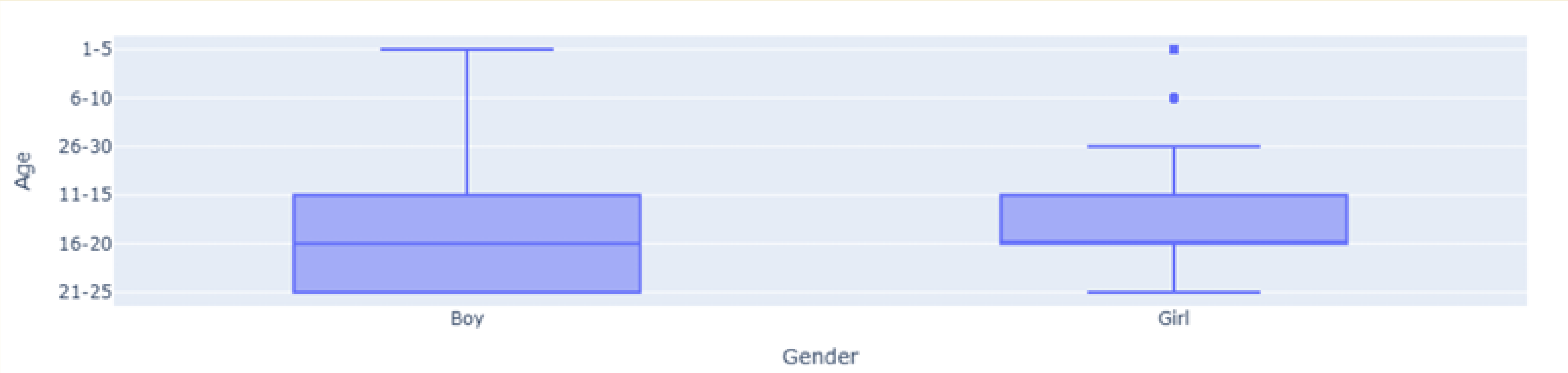
02 Distribution of IT Student



03 Histogram of Age



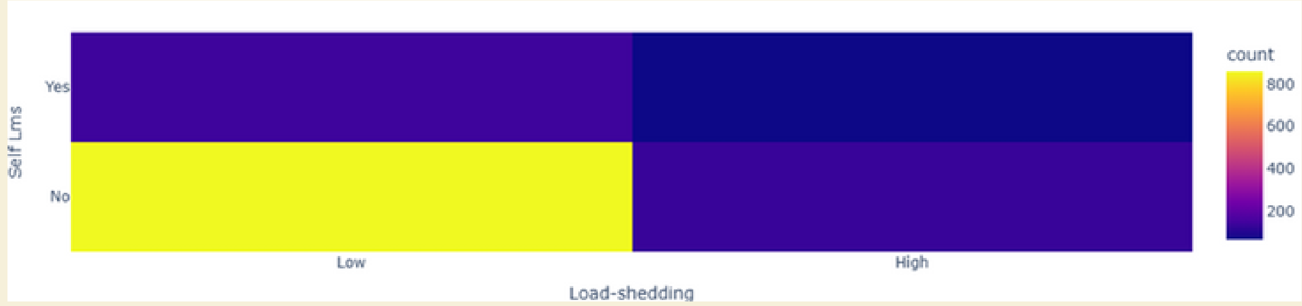
04 Box Plot of Age by Gender



EXPLORATORY DATA ANALYSIS (EDA) AND GRAPH ANALYSIS - PART 2

01

Density Heatmap of Load-shedding vs Self Lms

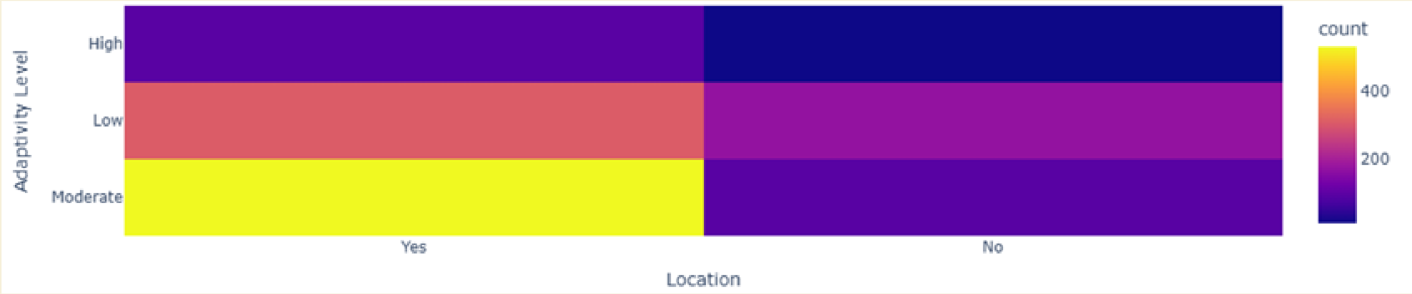


02

Sunburst Plot of Load-shedding and Self Lms

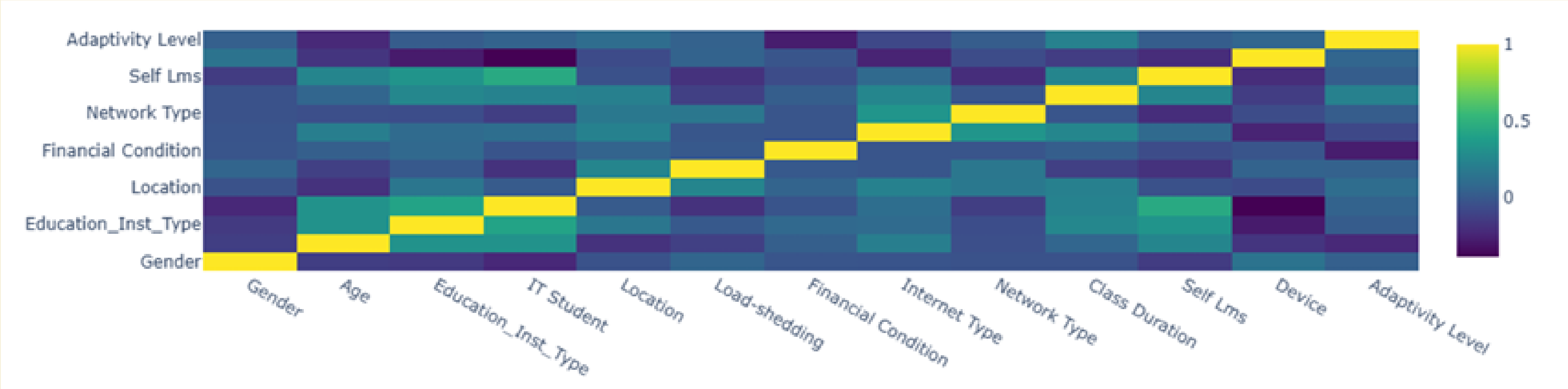
03

Density Heatmap of Location vs Adaptability Level



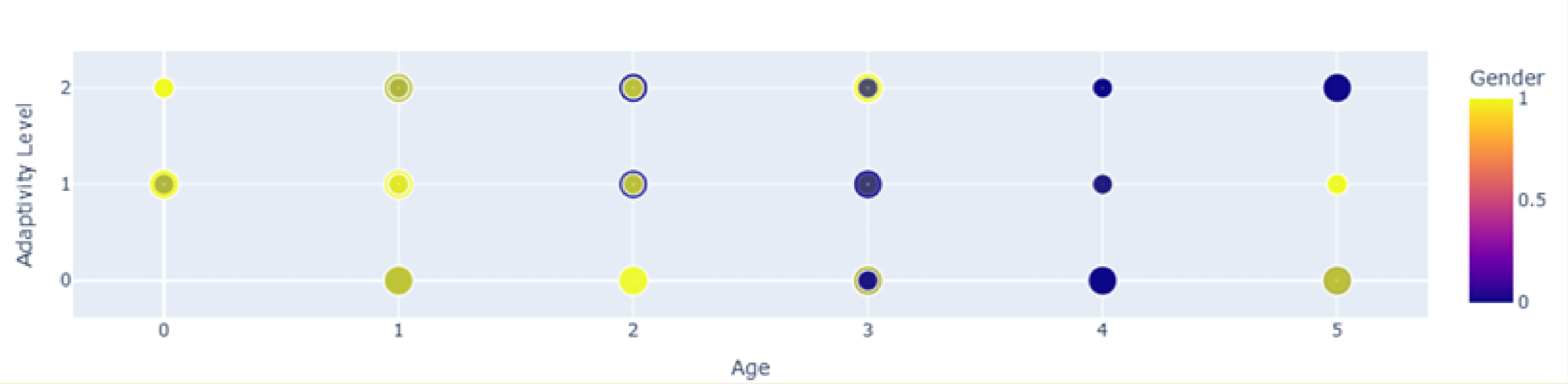
04

Correlation Heatmap

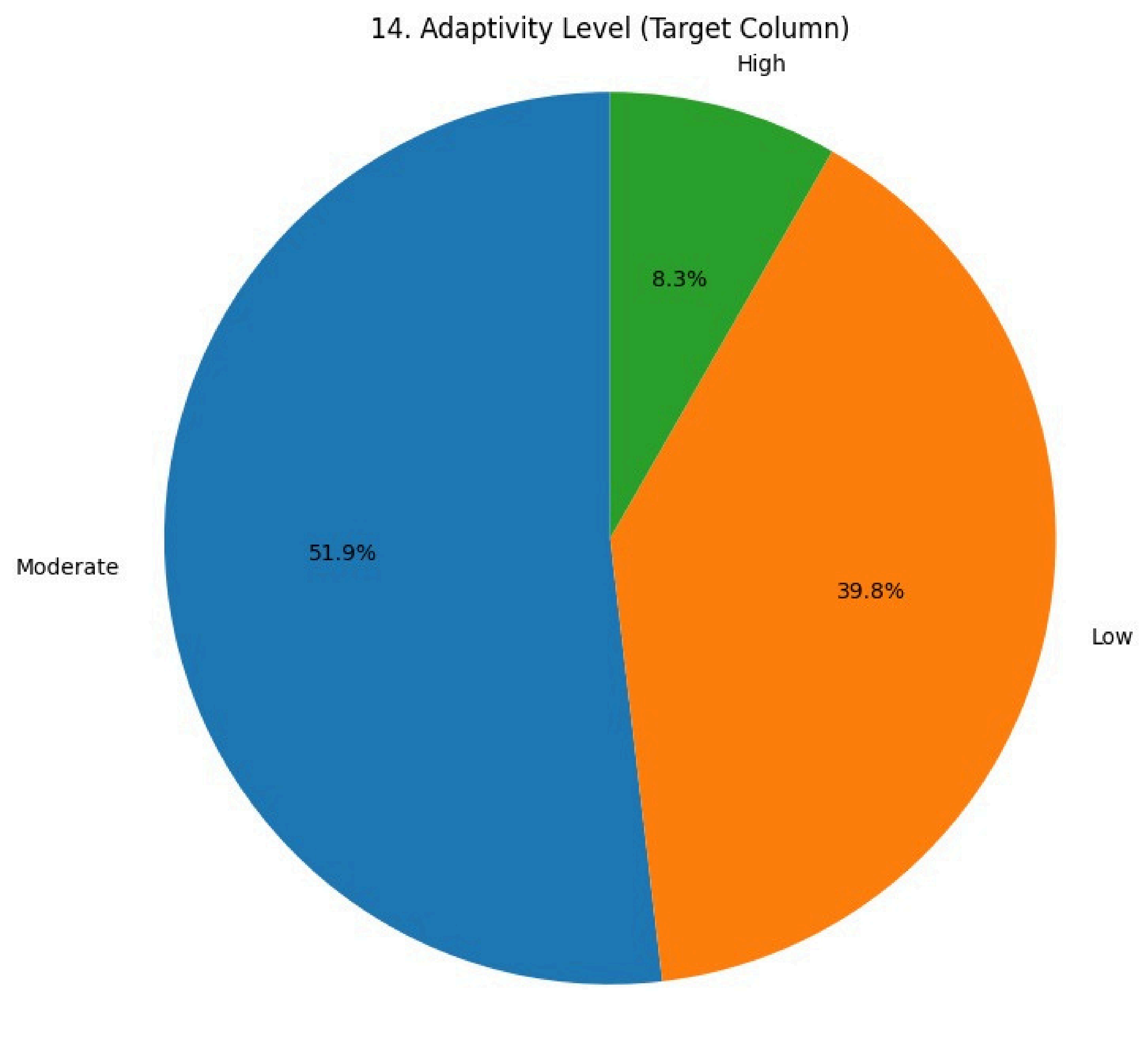


Bubble Chart of Age vs Adaptability Level, Size by Satisfaction

05



PRE-PROCESSING AND EXPLORATORY DATA ANALYSIS (EDA)



Data preprocessing is essential for enhancing machine learning accuracy, and it encompasses several critical steps:

1. **Data Cleaning**: This involves addressing missing values, eliminating duplicates, and correcting inconsistencies through methods like imputation or deletion.

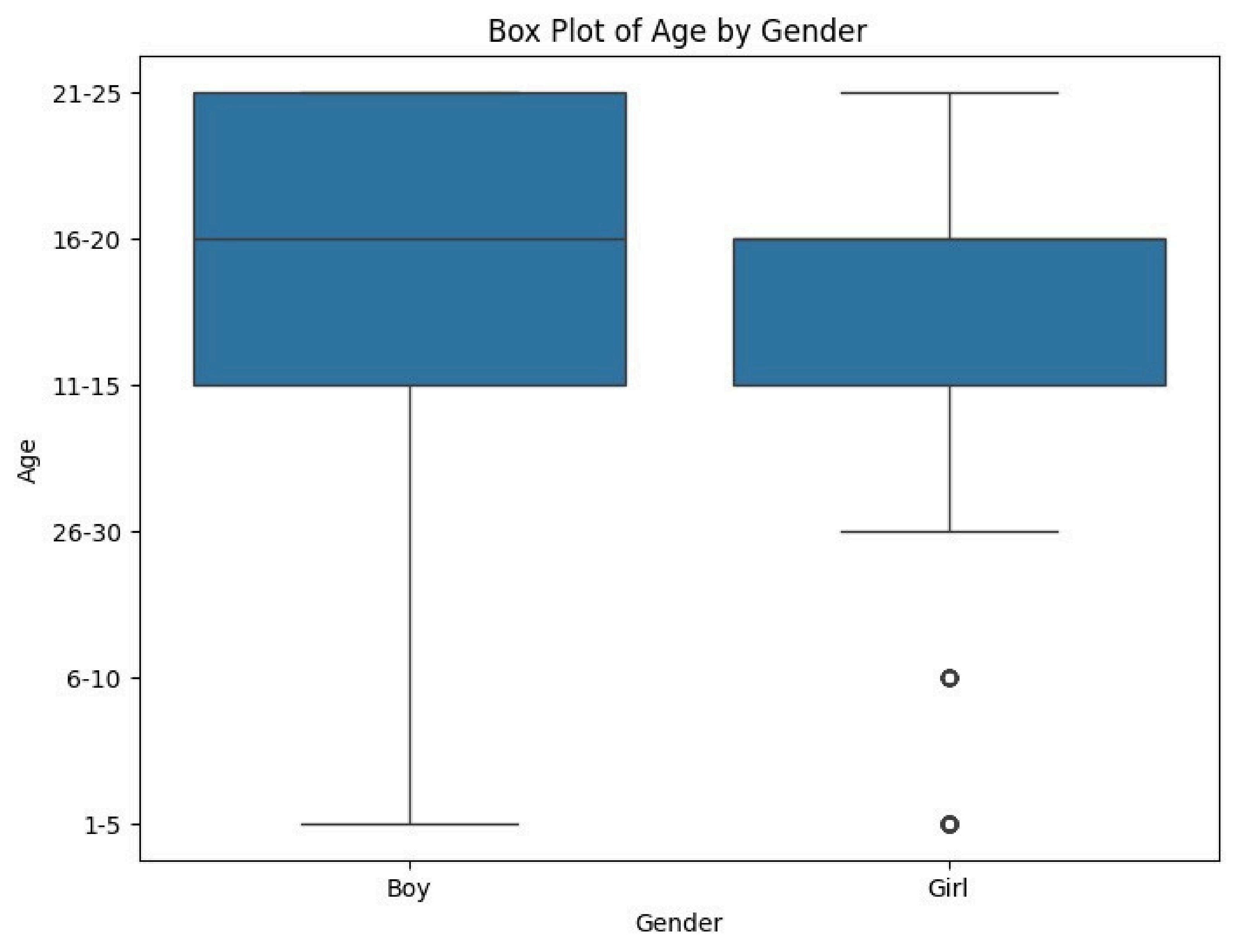
2. **Feature Scaling**: It is important to normalize numerical features using techniques such as Min-Max Scaling or Standardization to ensure optimal algorithm performance.

3. **Encoding Categorical Variables**: Categories need to be converted into numerical formats, which can be achieved through One-Hot Encoding or Label Encoding.

4. **Feature Selection**: Identifying and selecting relevant features is vital for improving model performance and reducing the risk of overfitting.

Once preprocessing is complete, it is beneficial to visualize data distributions using a pie chart to assess class balance, which aids in better model training and evaluation.

FURTHER



This box plot shows the age distribution by gender, which is relevant for predicting students' adaptability to e-learning:

- **Boys** have a wider age range, with the median age around 16–20 and several younger outliers.
- **Girls** show a narrower range with a similar median age but fewer outliers.

These age differences can help in identifying which groups may need more support adapting to online education, informing a tailored approach to enhance learning outcomes.

ALGORITHMS USED IN THE PROJECT

To predict adaptability, we used:

- Decision Tree - It is a tree-like structure where each internal node tests on attribute, each branch corresponds to attribute value and each leaf node represents the final decision or prediction. The decision tree algorithm falls under the category of Supervised learning .
- Support Vector Machine (SVM) - SVM focus on finding the maximum separating hyperplane between the different classes in the target feature, making them robust for both binary and multiclass classification.
- Random Forest - It uses feature randomness which ensures low correlation among decision trees .
- XGBoost - (Check next slides...)

ABOUT XGBOOST

XGBoost was chosen for its high accuracy and efficiency with large datasets. This algorithm uses boosting to improve predictive accuracy, making it well-suited for complex classification tasks.

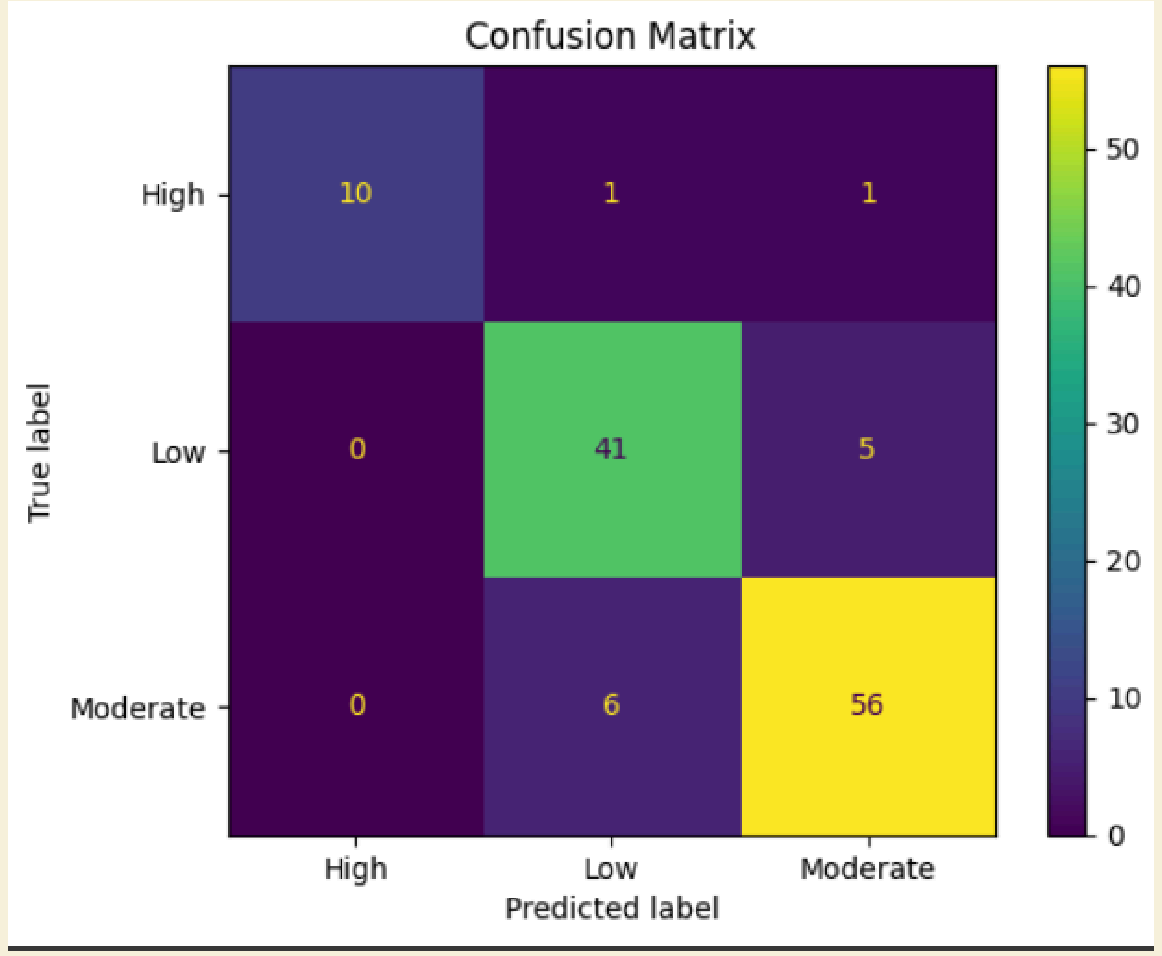
Features of XGBoost -

XGBoost offers regularization, which allows you to control overfitting by introducing L1/L2 penalties on the weights and biases of each tree.

It has ability to handle sparse data sets using the weighted quantile sketch algorithm. This algorithm allows us to deal with non-zero entries in the feature matrix

XGBOOST PROCESS VISUALIZATION

The XGBoost process involves sequentially improving model performance by correcting errors from previous models, resulting in higher accuracy for adaptability predictions.



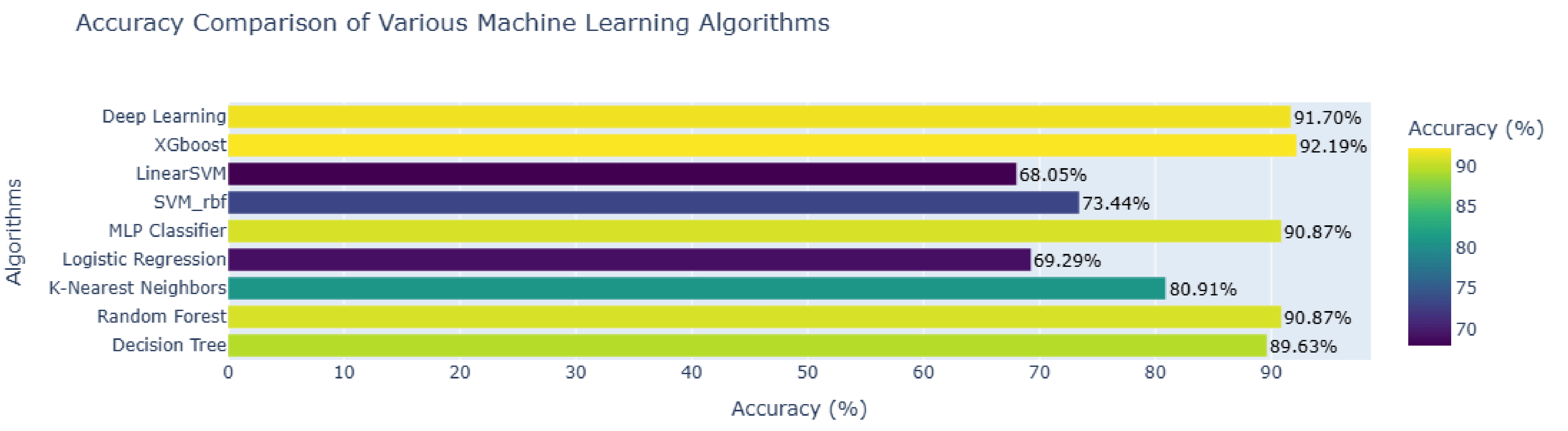
The XGBoost algorithm can also be divided into two types based on the target values:

1.Classification boosting is used to classify samples into distinct classes, and in xgboost, this is implemented using XGBClassifier.

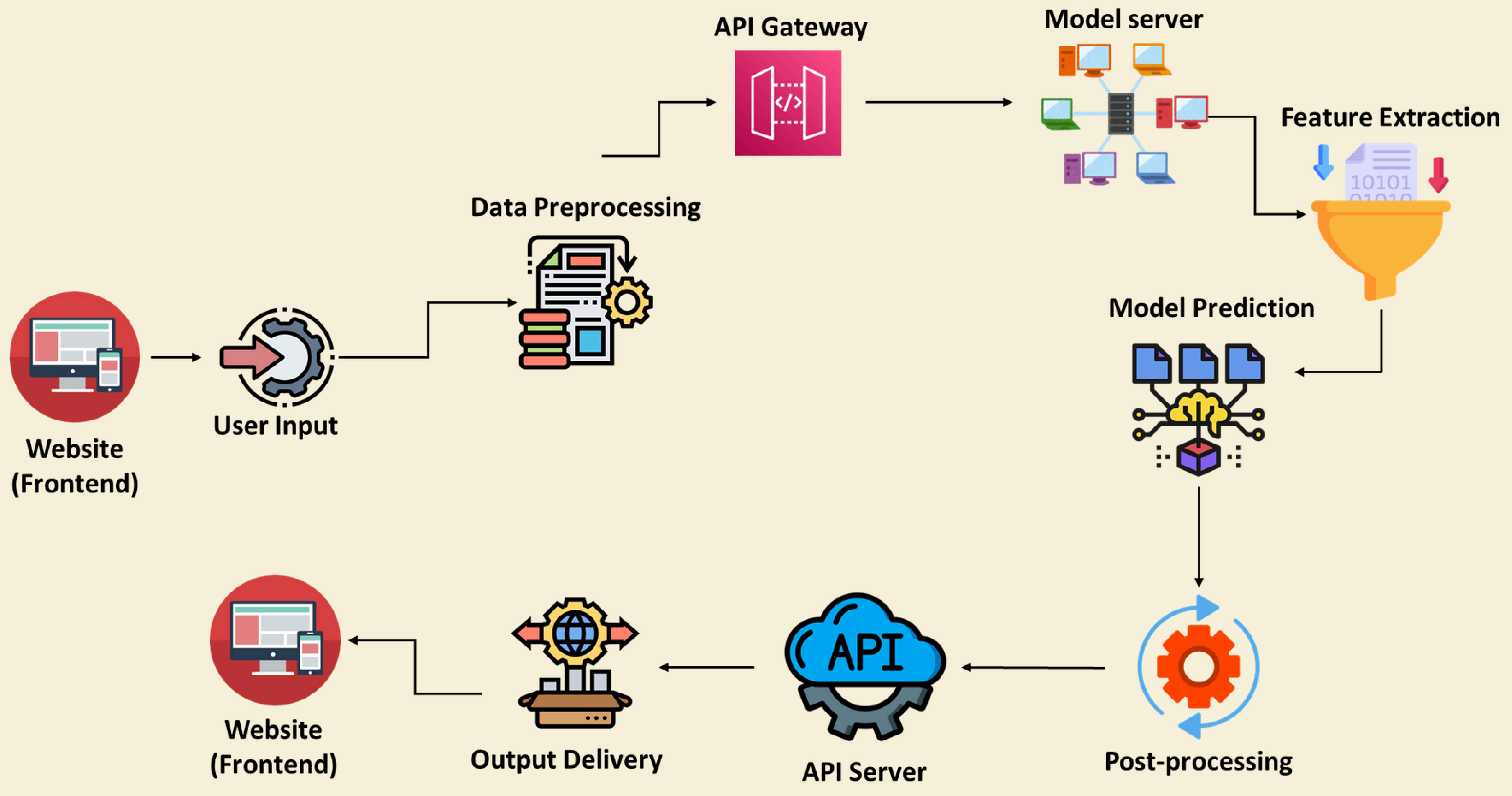
2. Regression boosting is used to predict continuous numerical values, and in xgboost, this is implemented using XGBRegressor.

ACCURACY TABLES

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



Summary:
Best Fold Accuracy: 95.87%
Average Accuracy: 92.11%


API Gateway
Model server
Feature Extraction
10101
Data Preprocessing
Model Prediction
Website
(Frontend)
User Input
API
Website
(Frontend)
Output Delivery
API Server
Post-processing

USE CASE

1. Early Intervention for At-Risk Students: Identifies students struggling with online learning, allowing institutions to provide targeted support like tutoring or counseling, improving their chances of success.

2. Personalized Learning and Resource Allocation: Assesses adaptability to tailor learning paths and allocate resources efficiently, ensuring each student receives the support suited to their needs.

3. Course Design Improvement: Provides insights to enhance online courses with features like interactive content and flexible deadlines, making programs more engaging and accessible for all adaptability levels.

CONCLUSION

Our analysis shows that XGBoost is the most effective model for predicting adaptability levels. This project provides insights that can help educational institutions create adaptive e-learning environments tailored to student needs.

```
print("Thank you")
```

THANK YOU

GROUP 7

Darshan Prajapati (B23CH1032)
Malav Parekh (B23ME1029)
Prateek Agrahari (B23BB1033)
Zalak Kanzaria (B23MT1023)