 

Boot Camp on Artificial Intelligence

**Practical Assignment - 1**

**Group Members:**  **PRN:**

Mihir Patil 122B1B213

Siddhesh Patil 122B1B218

Mihir Pande 122B1B200

Rohan Mengal 122B1B175

Shrawani Pande 122B1B201

Yash Mulay 122B1B178

Mubin Tamboli 22TCOMP2071

Shailesh Galande 22TCOMP2024

Dhanashree Patil 22TCOMP2093

Shivani Pandey 22TCOMP2109

**Date of Submission: Maximum Marks:**

**Lab Assignment 1: Data Cleaning, Encoding, and Scaling**

**Problem Statement:**

You are provided with a dataset containing information about students in a school. The dataset includes columns such as StudentID, Name, Gender, Age, MathScore, EnglishScore, and TotalScore. The dataset has the following issues:

1. Missing values in the MathScore and EnglishScore columns.

2. The Gender column contains categorical data.

3. The TotalScore column is incorrect (it should be the sum of MathScore and EnglishScore).

**Tasks:**

Your task is to:

1. Handle the missing data by replacing missing MathScore and EnglishScore with the mean of the respective columns.

2. Encode the Gender column using label encoding (0 for female, 1 for male). 3. Correct the TotalScore column.

4. Standardize the MathScore, EnglishScore, and TotalScore columns using IQR Method. **Dataset:**

Use the CSV file named students.csv

**Lab Assignment 2: Feature Engineering and Outlier Detection**

**Problem Statement:**

You are provided with a dataset containing sales data for a retail store. The dataset includes columns such as ProductID, ProductCategory, Price, QuantitySold, and Revenue. However, the dataset has some inconsistencies:

1. The Revenue column is incorrect and needs to be calculated as Price \* QuantitySold. 2. Some Price values seem too high or too low and may be outliers.

**Tasks:**

Your task is to:

1. Create a new column Revenue by multiplying Price and QuantitySold.

2. Detect and remove outliers in the Price column using the IQR Method.

3. Normalize the Price and Revenue columns using Min-Max scaling. **Dataset:**

Use the CSV file named sales.csv