**Name : Ali Azhar**

**Roll Number : SU92-BSAIM-S24-026**

**Section : 4A**

**“Documentation of Task 2”**

**“Gradient Boosting Model for Transport Prediction”**

**Objective:**

This program predicts whether a passenger is Transported or not using a **Gradient Boosting Classifier**. It is based on data from the **train.csv** and **test.csv** files.

**Description:**

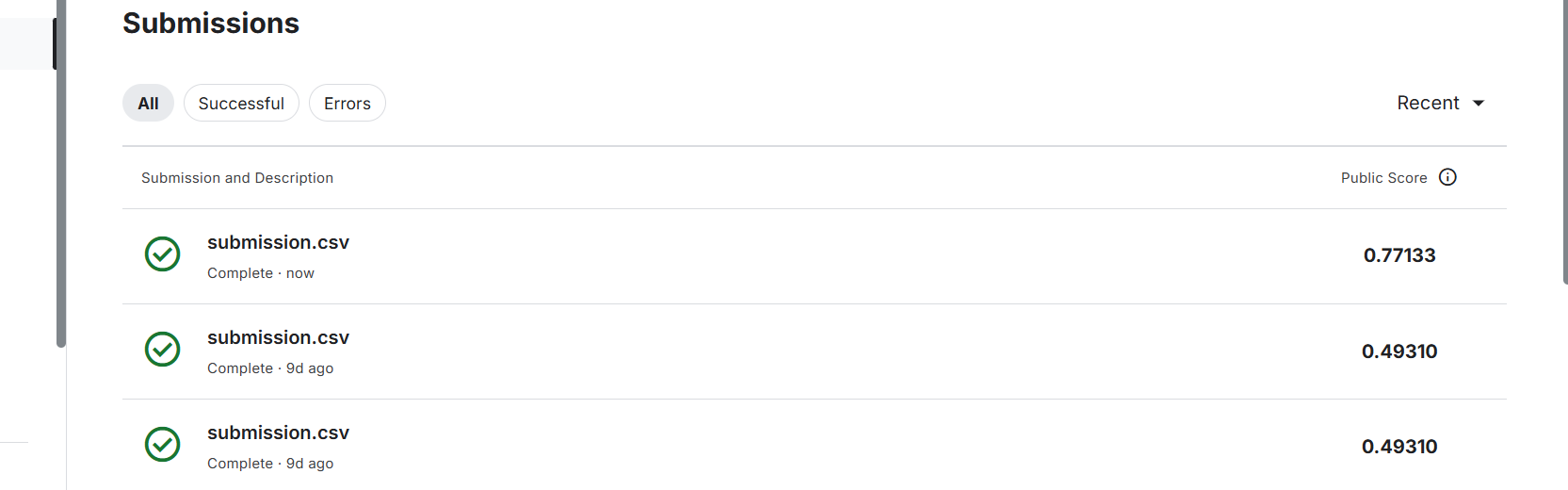
The program performs **data preprocessing**, trains a **machine learning model**, evaluates its **accuracy**, and generates a **submission file** with predictions.

**Steps Involved:**

1. **Import Libraries:**  
   Uses pandas, numpy, and scikit-learn for data handling, preprocessing, and modeling.
2. **Data Loading:**
3. Reads train.csv and test.csv files.
4. **Data Cleaning:**
   * Fills missing values (with “Missing” or median).
   * Converts categorical and boolean data using **Label Encoding**.
   * Scales numerical features with **StandardScaler**.
5. **Model Training and Testing:**
   * Splits the data into training and validation sets.
   * Trains a **Gradient Boosting Classifier**.
   * Calculates and prints the model’s accuracy.
6. **Final Model and Submission:**
   * Retrains the model on full training data.
   * Predicts results for the test dataset.
   * Saves predictions to submission.csv.

**Output:**

* Displays model accuracy in the console.
* Generates a file named **submission.csv**

****