# Documentation for train_test.ipynb

This notebook focuses on **Train-Test Splitting, Model Application, and Evaluation** and is designated as **Lab 11 Task**.

## Modeling Steps

1. **Data Preparation:**
   - The preprocessed dataset is loaded from preprocessed.csv.
   - The data is split into **Features (X)** (all columns except the last one) and the **Target variable (Y)** (the last column, which is the encoded Continent label).
   - Any remaining categorical columns in X are converted to integer representation using pd.factorize().
2. **Train-Test Splitting:**
   - The dataset is split into training and testing sets using train_test_split.
   - The **test size** is set to **30%** (test_size=.3), and the data is **not shuffled** (shuffle=False).
3. **Model Application and Evaluation:**
   - Multiple classification algorithms were tested, including **Decision Tree**, **Random Forest**, **Gaussian Naive Bayes**, **K-Nearest Neighbors**, **Logistic Regression**, and **Support Vector Classifier (SVC)**.
   - Key metrics—**Accuracy**, **Precision**, **Recall**, and **F1-score**—were calculated and printed for each model.
   - A **bar chart** was generated to visually compare the performance (accuracy) of the different models.
4. **Final Results (Decision Tree Classifier):**
   - The final report focuses on the Decision Tree Classifier, showing a detailed performance report.
   - **Accuracy Score: 0.9333**.
   - A full **Classification Report** is displayed, showing precision, recall, and f1-score for each predicted class.