

The Superior University

|  |  |  |
| --- | --- | --- |
| Name: Muhammad Bilal Sajid | Roll No: 018 | Course: PAI Lab |
| Semester: 4th | Section: BSAI 4A | Department: |
| Submitted To: | Total Marks: | Date: |

**Lab 2**

**Task: Kaggle Competition: Spaceship Titanic**

**Spaceship Titanic Classification Using Machine Learning**

**Introduction**

This project involves predicting passenger survival on the "Spaceship Titanic" dataset using a Random Forest Classifier. The dataset contains various features related to passengers, and the goal is to classify whether they were transported to another dimension.

**Code Explanation**

**1. Data Loading and Exploration**

* The dataset is loaded using pandas.read\_csv().
* Basic exploratory data analysis (EDA) is performed using .head(), .info(), and .describe() to understand the structure and characteristics of the dataset.

**2. Data Preprocessing**

* Unnecessary columns (Name, Cabin, PassengerId) are dropped to reduce noise in the dataset.
* Missing values in numerical columns are filled with their median values to maintain data consistency.
* Categorical variables are transformed using **One-Hot Encoding**, making them suitable for machine learning models.

**3. Feature Engineering and Scaling**

* Numeric features are standardized using StandardScaler() to normalize the dataset.
* The dataset is prepared for training by encoding categorical variables.

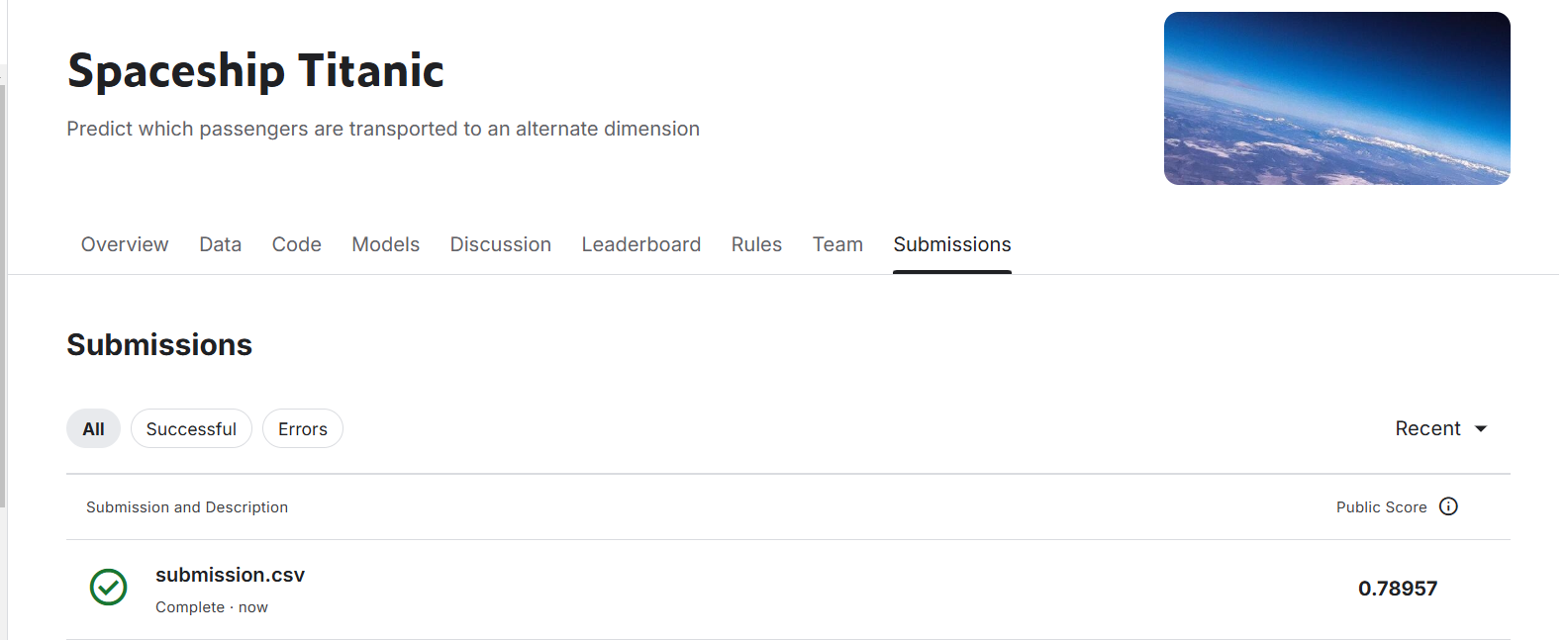
**4. Model Training and Evaluation**

* The dataset is split into training and validation sets using train\_test\_split() with an 80-20 split.
* A **Random Forest Classifier** is trained on the processed data.
* Model accuracy is evaluated using accuracy\_score() to measure the performance.

**Model Performance**

* **Random Forest Accuracy:** Displayed in the console.
* The model performance can be further improved using hyperparameter tuning or additional feature engineering.

**Output Screenshot**

****