Task 02

Explanation of code:

Spaceship Titanic:

This is a project on predicting whether a spaceship passenger was taken to another dimension. The data has information such as age, home planet, VIP status, expenditure in different facilities, and whether the passenger was in cryosleep.

Step-by-Step Breakdown:

Loading the Data:

The data is loaded using pandas.read_csv(), and some initial exploration (head(), tail(), info(), describe(), shape) is to understand an idea of the data.

Handling Missing Values:

- Missing numerical values (e.g., Age, RoomService, FoodCourt, etc.) are replaced with the mean of the column.
- Missing categorical values (HomePlanet, Destination) are replaced with the most frequent (mode) value.
- Certain columns (VIP, CryoSleep) are encoded to 0 and 1 (Boolean values).

Feature Encoding:

- Non-numeric columns (HomePlanet, Destination) are encoded to numbers with LabelEncoder().

Dropping Unnecessary Columns:

- Columns such as PassengerId, Name, and Cabin are dropped because they don't help in prediction.

Preparing Data for Machine Learning:

- The dataset is separated into features (X) and target (y, which is Transported).
- The data are separated into training and validation sets by train_test_split().

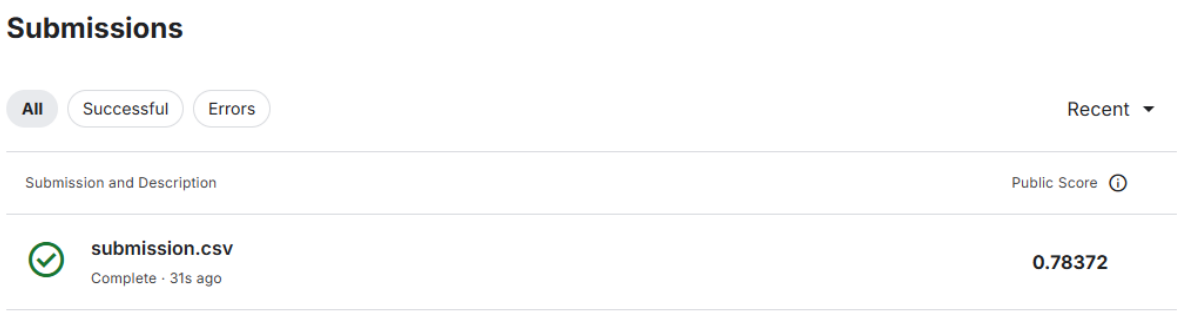
Creating the Model:

- A RandomForestClassifier (machine learning model) is trained on the dataset.
- The model is evaluated on the validation set, and accuracy is printed.

Making Predictions on Test Data:

- Same preprocessing steps are carried out for the test dataset.
- Predictions are made, and a submission file (submission.csv) is generated.

Accuracy:


Submissions
All
Successful
Errors
Recent
Submission and Description
Public Score
submission.csv
Complete · 31s ago
0.78372