Name: M Shan Iftikhar

Roll No: BSAI-4A-017

Lab task 2

Kaggle Competition: Spaceship Titanic

The objective is to forecast whether a passenger was carried to another universe upon the collision of the spaceship.

Loading the Data

Importing the training and test data (train.csv and test.csv) with pandas. The data includes details about passengers, including their cabin, expenditure, and whether they were shipped.

Handling Missing Values  
Missing values are handled differently depending on the data type  
For categorical columns (object type), missing values are filled using the mode, which is the most frequent value.  
For numerical columns, missing values are filled using the mean value.

**Feature Engineering**  
The Cabin column is divided into CabinDeck and CabinSide in order to pull deck and side columns.  
New column,TotalSpending,is formed by adding all columns related to spending (RoomService, FoodCourt, ShoppingMall, Spa, and VRDeck).  
The PassengerId column is divided into group and number in order to pull group and individual identifiers.  
Columns such as Cabin, PassengerId, and Name, which are not of use, are dropped to diminish noise in the data.  
**Encoding Categorical Columns**  
Categorical columns are converted into numerical format using one-hot encoding (pd.get\_dummies).   
The training and test datasets are aligned to ensure they have the same columns, and any missing columns in the test set are filled with zeros.  
**Separating Target Variable**  
The target variable, Transported, is separated from the training data (y\_train).  
**Feature Scaling**  
The numerical features are scaled by StandardScaler in order to standardize the data. This will ensure that each feature makes the same contribution to the model performance.  
**Model Training**  
RandomForestClassifier is utilized for training the model. The model is set up with 100 estimators and a deterministic random seed (random\_state=42) to achieve reproducibility.  
**Training the Model**  
The model is trained on the preprocessed training data (X\_train and y\_train).  
**Predictions**  
The trained model is used to predict the target variable (Transported) for the test dataset (X\_test).

