Task 2: Titanic Predicting Passenger Transportation Status

# Objective

The objective of this task was to predict whether passengers were transported to another dimension based on selected features from the Spaceship Titanic dataset. The prediction was based on the following selected columns: PassengerId, VIP, CryoSleep, and Age. A Random Forest Classifier was employed for this binary classification task.

# Dataset Source

The dataset was obtained from the Kaggle competition: "Spaceship Titanic". It contains information about passengers on the fictional Spaceship Titanic, including various attributes and whether they were transported to another dimension.

# Selected Features for Prediction

1. VIP – Boolean indicating whether the passenger paid for special VIP service.

2. CryoSleep – Boolean indicating whether the passenger opted for cryogenic sleep during the voyage.

3. Age – Numerical value representing the age of the passenger.

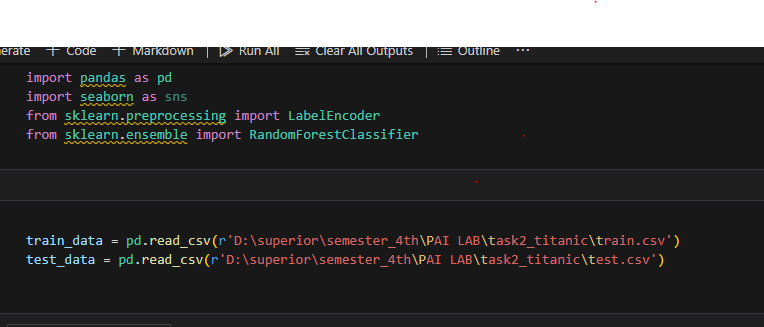
# Tools and Libraries Used

- Python

- Pandas for data loading and preprocessing

- Seaborn for data visualization

- Scikit-learn for building the Random Forest Classifier



# Data Preprocessing

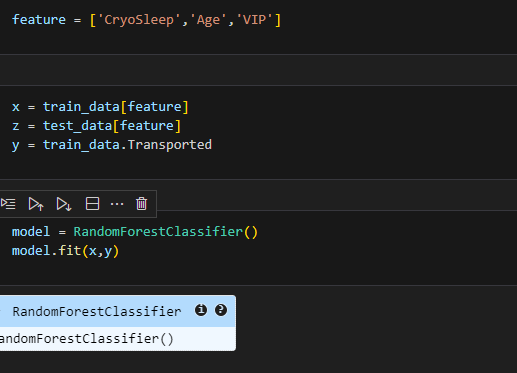
Handling Missing Values: Missing values in the selected features were addressed through appropriate imputation strategies.

Encoding Categorical Variables: Boolean features such as VIP and CryoSleep were encoded to numerical values to be compatible with the machine learning model.



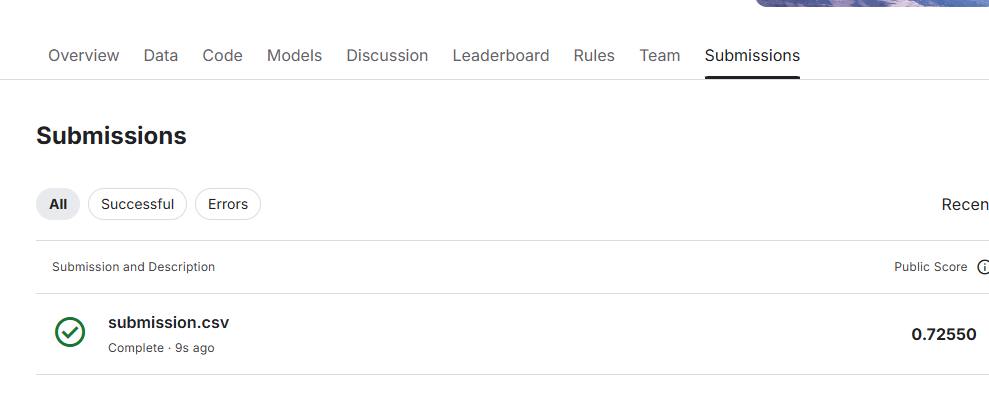
# Model and Prediction

A Random Forest Classifier was trained using the selected features. The dataset was split into training and testing sets to evaluate the model's performance. Hyperparameters were tuned to optimize the model's accuracy.



# Result and Accuracy

The prediction model achieved an approximate accuracy of 70% on the test dataset. This indicates a reasonable predictive capability given the limited number of features used.



# Conclusion

This task demonstrated the application of machine learning techniques to predict outcomes based on selected features. While the model achieved a 72% accuracy, incorporating additional relevant features and further tuning could potentially enhance performance.