

## 

***Spaceship Titanic :***

*Result File*

**LAB 02: Kaggle Competition Spaceship Titanic**

**Introduction:**

This project is basically a Kaggle competition task named as Spaceship Titanic. This project involves **data processing** and **model training** using **machine learning** techniques. The dataset includes various attributes, and the objective is to preprocess the data, perform exploratory data analysis (EDA), and train machine learning models for predictive analysis.

**Code Explanation**

1. **Data Loading and Preprocessing**

The dataset is loaded using **pandas.read\_csv()**.

**Both:**

train\_df = pd.read\_csv("train.csv")

test\_df  = pd.read\_csv("test.csv")

The initial structure of the dataset is explored using .shape, .columns, .dtypes, and .describe().

Missing values are handled using in this to make data ready to use for model.

**2. Feature Engineering and Scaling**

The dataset undergoes preprocessing by dropping unnecessary columns

(**Name, Cabin, and PassengerId**).

Missing values in numerical columns are replaced with the **median**.

Categorical columns are transformed using **One-Hot Encoding** with OneHotEncoder().

The transformed dataset is concatenated back for further processing.

The dataset is divided into two parts:

* + **Training dataset** for model learning.
  + **Test dataset** for evaluating results.

Missing values in categorical columns are replaced with the **most frequent values**.

One-hot encoding is applied using pd.get\_dummies() to convert categorical variables into numerical form.

Features are standardized using StandardScaler() to normalize the dataset.

**4. Model Training and Evaluation**

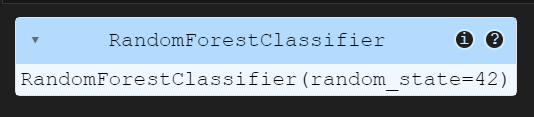
The dataset is split into training and validation sets (train\_test\_split() with an 80-20 split).

models are trained:

**RandomForestClassifier**

Models are evaluated using Checking accuracy.

The results of both models are displayed.



**5. Final Predictions and Submission**

The test dataset undergoes the same preprocessing steps as the training dataset.

Missing columns in the test dataset are aligned with the training dataset.

The trained model is used to predict outcomes.

Predictions are saved to a CSV file for submission.

**Conclusion**

This notebook focuses on data preprocessing, exploratory analysis, and model training. The trained models provide predictions based on structured input data. And submitted on Kaggle the output is given below:

