







 mikabenson /
Space_ship-project



 Code  Issues  Pull requests  Actions  Projects  Wiki  Security  Insights  Setting



★ 0 stars  0 forks  1 watching  Branches  Activity
 Tags

 Public repository

  1 Branch  0 Tags  




















mikabenson Final update for the project

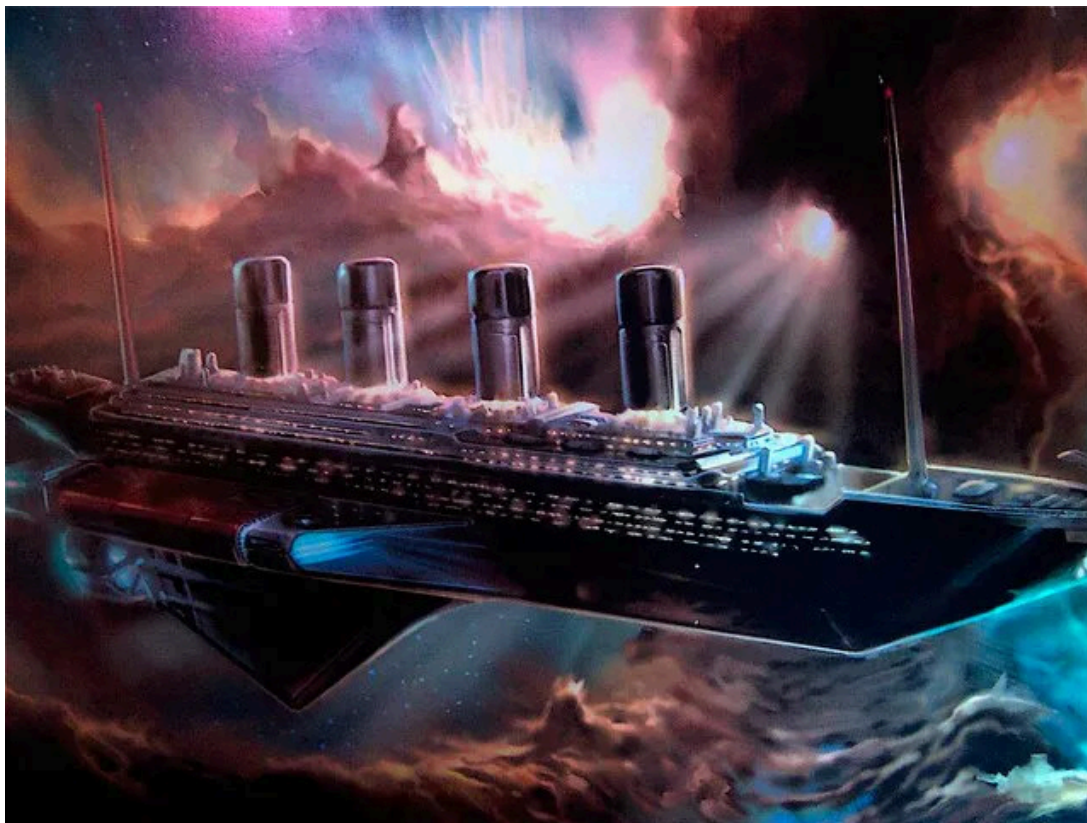
501dd48 · 5 minutes ago 

 README.md	tuning my models	3 hours ago
 data_2.csv	saving to pdf	1 hour ago
 decision_tree	first commit	last week
 sample_submission.csv	saving to pdf	1 hour ago
 spaceship.pdf	Final update for the project	5 minutes ago
 student.ipynb	saving to pdf	1 hour ago
 student.ipynb.pdf	saving to pdf	1 hour ago
 test.csv	saving to pdf	1 hour ago
 train.csv	saving to pdf	1 hour ago

 README



Spaceship Titanic: Predicting Alternate Dimension Transportation for the Passengers caused by the anomaly 🚀



Introduction 🌟

Here is my Data Science and Machine Learning project. I created this repository which houses my project from the Kaggle Spaceship Titanic competition. I'm working on trying to understand the mystery behind the passengers who were transported to an dimension during the Spaceship Titanic collision with the anomaly

[Link to my Submission Notebook](#)

Key Insights & Results:

- One of the most critical factors in predicting passenger transportation was the "CryoSleep" feature, showing a strong correlation of over 0.4. Passengers who opted for cryosleep were more likely to be transported to an alternate dimension.
- Through rigorous analysis and modeling, I achieved a commendable score of 0.80, securing a top 28% ranking among 2062 teams. My model's predictions played a vital role in the rescue mission's success.

About the Challenge 🌐

The Spaceship Titanic competition tasks us with solving a cosmic mystery. By leveraging data science skills, we aim to predict the fate of passengers who encountered a spacetime anomaly during their voyage.

[Challenge Link](#)

Challenge Details 📄

- **Goal:** Classify passengers as transported or not transported to an alternate dimension.

- **Datasets:** The competition provides a training dataset with personal records and a test dataset for predictions.
- **Evaluation:** Submissions are evaluated based on classification accuracy.

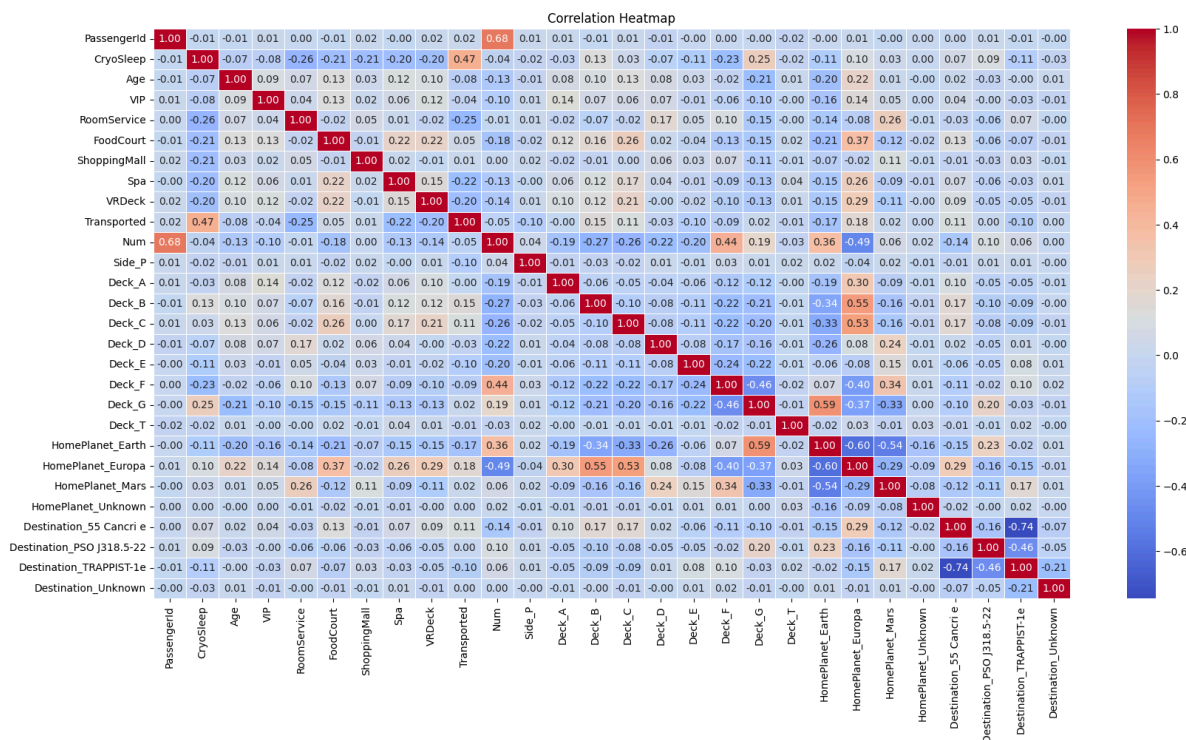
Project Files

Key files related to this project:

- [train.csv](#) - Training dataset with personal records and ground truth labels.
- [test.csv](#) - Test dataset for predictions.
- [sample_submission.csv](#) - Sample submission file with the required format.
- [My Notebook](#) or [My Kaggle Notebook](#) - My notebook with code and analysis.

My Approach

1. **Reading Datasets:** I began by loading the provided datasets, both the training and test data.
2. **Handling Missing Values:** I addressed missing data in the dataset, ensuring that no valuable information was lost.
3. **Checking Class Distribution:** To understand the balance between transported and non-transported passengers, I examined the distribution of classes in the training dataset.
4. **Data Visualization :** I perform some visualization on the train csv to check for the features that affected the rate of transportation on the passengers



5. **Feature Engineering:** To improve model performance, I engineered new features. For example, I extracted additional information from the "numeric" column that had expenditure, summing them to the spending cost

6. **One-Hot Encoding:** I prepared the categorical data for modeling by performing one-hot encoding, a necessary step for many machine learning models.
7. **Scaling:** I prepared the numeric data for modeling by performing scaling which is a necessary step for data Normalization.
8. **Extracting (X, y):** I separated the feature matrix (X) and the target variable (y) from the training dataset, ensuring that the data was ready for model training.
9. **Creating the Models:** I created my baseline logistic regression model which was my reference model, then evaluated the accuracy score, plotted the ROC curve. I also improved my baseline model by applying the following model:
 - DecisionTreeClassifier
 - RandomForestClassifier
 - I performed tuning on these models to optimize their performance, checked its accuracy score and finally created the ROC curve
10. **Making Predictions:** I got the predictions from the three models.

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

● Jupyter Notebook 100.0%