

Cosmic Passenger Dimension Prediction Project Report

Introduction

In this project, the task is to predict whether a passenger aboard the Spaceship Titanic was transported to an alternate dimension during its collision with a spacetime anomaly. This cosmic mystery requires analyzing personal records recovered from the spaceship's damaged computer system to classify passengers accurately.

Dataset Overview

The dataset comprises personal records for approximately two-thirds (~8700) of the passengers, with features including PassengerId, HomePlanet, CryoSleep status, Cabin number, Destination, Age, VIP status, and billing amounts at luxury amenities. The target variable, Transported, indicates whether a passenger was transported to another dimension.

Data Preprocessing

Loaded the dataset and performed exploratory data analysis to understand feature distributions and the target variable.

Encoded categorical features using LabelEncoder for modeling purposes.

Split the dataset into training and testing sets for model training and evaluation.

Model Development

Utilized the following classification algorithms:

Logistic Regression: Simple and effective for binary classification tasks.

Support Vector Machine (SVM): Effective for classification tasks in high-dimensional spaces.

Trained each model using the training data and evaluated their performance on the testing data.

Model Evaluation

Evaluated the performance of each model based on:

Accuracy Score: Percentage of correctly classified instances.

Classification Report: Analysis of precision, recall, and F1-score for each class.

Conclusion

This project demonstrates the application of multiple machine learning models to predict whether passengers aboard the Spaceship Titanic were transported to an alternate dimension. By preprocessing the data, training logistic regression, linear regression, and SVM models, accurate predictions can be made to assist in rescuing the lost passengers and changing history.

Next Steps

Experiment with additional feature engineering techniques to potentially improve model performance.

Fine-tune hyperparameters of SVM and logistic regression models for better accuracy.

Consider ensemble methods such as random forests or gradient boosting for further experimentation.