

# **Predicting Survival Outcomes from Titanic Disaster**

A Machine Learning Approach

Sai Kiran Chikatimalla

# Introduction:

- ▶ The Titanic disaster is one of the most analyzed incidents in history.
- ▶ This project aims to predict whether a passenger survived using various machine learning algorithms.
- ▶ The dataset includes diverse features such as age, gender, and class, which are crucial for making predictions about survival chances.

## Dataset Overview:

- ▶ Dataset Overview Dataset Size: 418 passenger entries  
Each entry indicates whether a passenger survived (1) or not (0).
- ▶ The dataset is likely used in conjunction with other Titanic datasets for building and evaluating predictive models.

# Methodology:

## Data Preprocessing:

- ▶ Handling missing values
- ▶ Scaling features
- ▶ Feature Engineering:

Creating new features such as family size and gender indicator8.

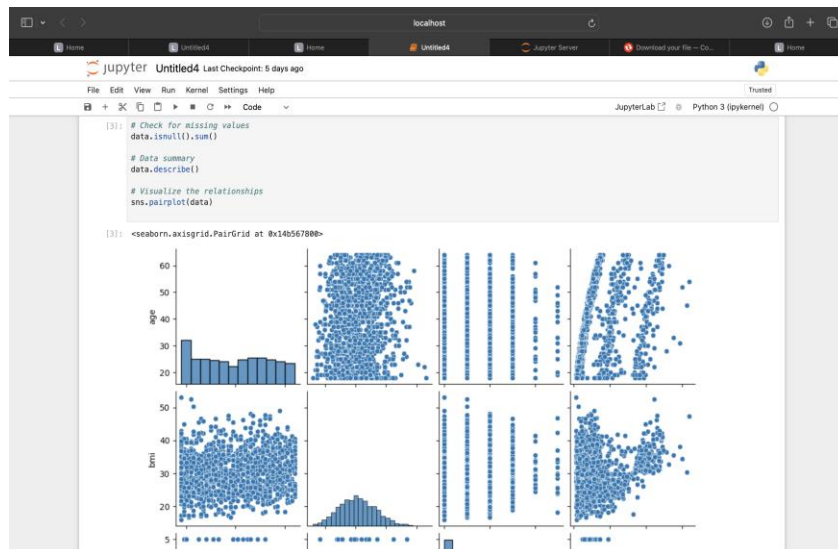
- ▶ Algorithms Tested
- ▶ Logistic Regression
- ▶ Random Forest Classifier
- ▶ Support Vector Machine (SVM)
- ▶ Neural Network

# Model Evaluation:

- ▶ Models will be evaluated using:
  - ▶ Accuracy
  - ▶ Precision
  - ▶ Recall
  - ▶ F1-score
  - ▶ ROC-AUC.
- ▶ Cross-validation will be applied to avoid overfitting and perform hyperparameter tuning.

# Expected Results:

- ▶ Anticipate better performance from Random Forest and Neural Network models compared to simpler algorithms like Logistic Regression.
- ▶ These models can capture complex relationships in the data.  
age sex bmi children smoker region charges
- ▶ 0 19 female 27.900 0 yes southwest 16884.92400
- ▶ 1 18 male 33.770 1 no southeast 1725.55230
- ▶ 2 28 male 33.000 3 no southeast 4449.46200
- ▶ 3 33 male 22.705 0 no northwest 21984.47061
- ▶ 4 32 male 28.880 0 no northwest 3866.85520



## Discussion:

- ▶ Some models, like Random Forest, excel in handling nonlinear relationships. Neural Networks, while accurate, have drawbacks such as longer training times and lower interpretability<sup>10</sup>.
- ▶ Limitations include small dataset size and missing data, affecting model generalizability.

## Conclusion:

- ▶ The project demonstrates that machine learning models, particularly Random Forests and Neural Networks, can significantly improve survival prediction accuracy.
- ▶ Future enhancements could involve deep learning models, better feature engineering, and larger datasets to improve model performance.