

# Titanic Survival Prediction

Your Name

March 20, 2025

## Overview

This project predicts whether a passenger survived the Titanic disaster using machine learning. The dataset contains information such as passenger age, gender, fare, class, and embarkation point. By applying data preprocessing, exploratory data analysis (EDA), and machine learning models, we build a classifier to predict survival outcomes.

## Dataset

- **Source:** Titanic dataset from Kaggle or open-source repositories.
- **Features:** Passenger attributes like age, gender, class, fare, and embarkation point.
- **Target Variable:** Survival (0 = No, 1 = Yes).

## Project Workflow

### 1. Data Loading and Exploration

- Read the dataset using Pandas.
- Display dataset summary, missing values, and distributions.

### 2. Data Preprocessing

- Handle missing values:
  - Fill missing **Age** values with median.
  - Fill missing **Embarked** values with mode.
  - Drop **Cabin** due to excessive missing values.
- Convert categorical features (**Sex**, **Embarked**) into numerical values.

### 3. Exploratory Data Analysis (EDA)

- Visualize survival rates based on:
  - Sex
  - Passenger class (Pclass)
  - Age, Fare, and Embarked location
- Check feature correlations with survival.

### 4. Feature Engineering

- Convert categorical features using encoding.
- Create new relevant features for better prediction.

### 5. Model Training and Evaluation

- Train different ML models:
  - Logistic Regression
  - Decision Tree
  - Random Forest
  - Support Vector Machine (SVM)
- Evaluate models using:
  - Accuracy
  - Precision, Recall, F1-score
  - Confusion Matrix
- Select the best model based on performance.

## Results

### Insights

- Certain factors (e.g., **gender**, **class**, **fare**) significantly influenced survival.
- The best-performing model achieved **high accuracy** and provided valuable insights into survival patterns.

## Installation

## Usage

### Prerequisites

Ensure you have the following installed:

- Python 3.x
- Pandas, NumPy
- Matplotlib, Seaborn
- Scikit-learn

### Running the Project

```
# Clone the repository
git clone https://github.com/your-username/Titanic_Survival_Prediction.git
cd Titanic_Survival_Prediction
```

```
# Install dependencies
pip install -r requirements.txt
```

```
# Run the Jupyter Notebook
jupyter notebook Titanic_Survival_Prediction.ipynb
```

### Future Improvements

- Use more advanced feature engineering techniques.
- Experiment with deep learning models (e.g., Neural Networks).
- Tune hyperparameters for better model performance.

### Contributing

Contributions are welcome! Feel free to fork the repo and submit a pull request.

### License

This project is licensed under the **MIT License**.