# Titanic Survival Prediction Model

## Overview

This project develops a machine learning model to predict the survival of passengers aboard the Titanic based on the given dataset. The dataset includes features such as age, gender, ticket class, fare, and cabin. The objective is to build a robust model that classifies passengers as survivors or non-survivors.

## Approach

### 1. Data Preprocessing

* **Dropped Non-Essential Features**: The following columns were removed as they do not contribute to survival prediction:
  + PassengerId, Name, Ticket, Cabin.
* **Handled Missing Values**:
  + Age and Fare were filled with their respective medians.
* **Encoded Categorical Variables**:
  + Sex was encoded as binary (male = 0, female = 1).
  + Embarked was one-hot encoded (Embarked\_Q, Embarked\_S), dropping the first category to avoid multicollinearity.
* **Normalized Dataset**: The cleaned dataset was prepared for training.

### 2. Model Selection and Training

A **Random Forest Classifier** was used due to its robustness and ability to handle mixed data types effectively.

### 3. Model Evaluation

* **Train-Test Split**: The dataset was split into training (80%) and testing (20%) sets.
* **Metrics Used**:
  + **Accuracy**: Percentage of correct predictions.
  + **Classification Report**: Precision, recall, and F1-score for both classes (Survived and Not Survived).
  + **Confusion Matrix**: Counts of true positive, true negative, false positive, and false negative predictions.

### 4. Cross-Validation

* 5-fold cross-validation was performed to evaluate model robustness.
* Mean accuracy and standard deviation across folds were calculated.

## Results

### Test Set Performance:

* **Accuracy**: 100%
* **Classification Report**:
  + Precision, Recall, and F1-Score for both classes were 1.00.
* **Confusion Matrix**:
* [[50, 0],  
   [ 0, 34]]
  + The model correctly classified all instances.

### Cross-Validation Performance:

* **Mean Accuracy**: 100%
* **Standard Deviation**: 0.0 (consistent performance across folds).

## Challenges

* Handling missing values for Age and Fare.
* Balancing categorical encoding without overloading the model with unnecessary features.
* Ensuring that the model is not overfitting despite achieving high accuracy.

## Repository Structure

The repository is organized as follows:

- README.md: Project documentation (this file).  
- Dataset.csv: The Titanic dataset used for training and testing.  
- Titanic\_Survival\_Prediction.ipynb: Jupyter notebook with all code and steps.  
- requirements.txt: Python dependencies for the project.

## Dependencies

* **Python Libraries**:
  + pandas
  + scikit-learn

Install the dependencies using:

pip install -r requirements.txt

## How to Run

1. Clone the repository:

* git clone <repository-url>  
  cd Titanic-Survival-Prediction

1. Ensure all dependencies are installed:

* pip install -r requirements.txt

1. Run the code:
   * Use Titanic\_Survival\_Prediction.ipynb to view and execute step-by-step.

## Future Improvements

* Experiment with other algorithms (e.g., Logistic Regression, Gradient Boosting).
* Perform hyperparameter tuning to improve model performance.
* Investigate the impact of removing redundant features for simpler models.

Feel free to reach out for questions or contributions!