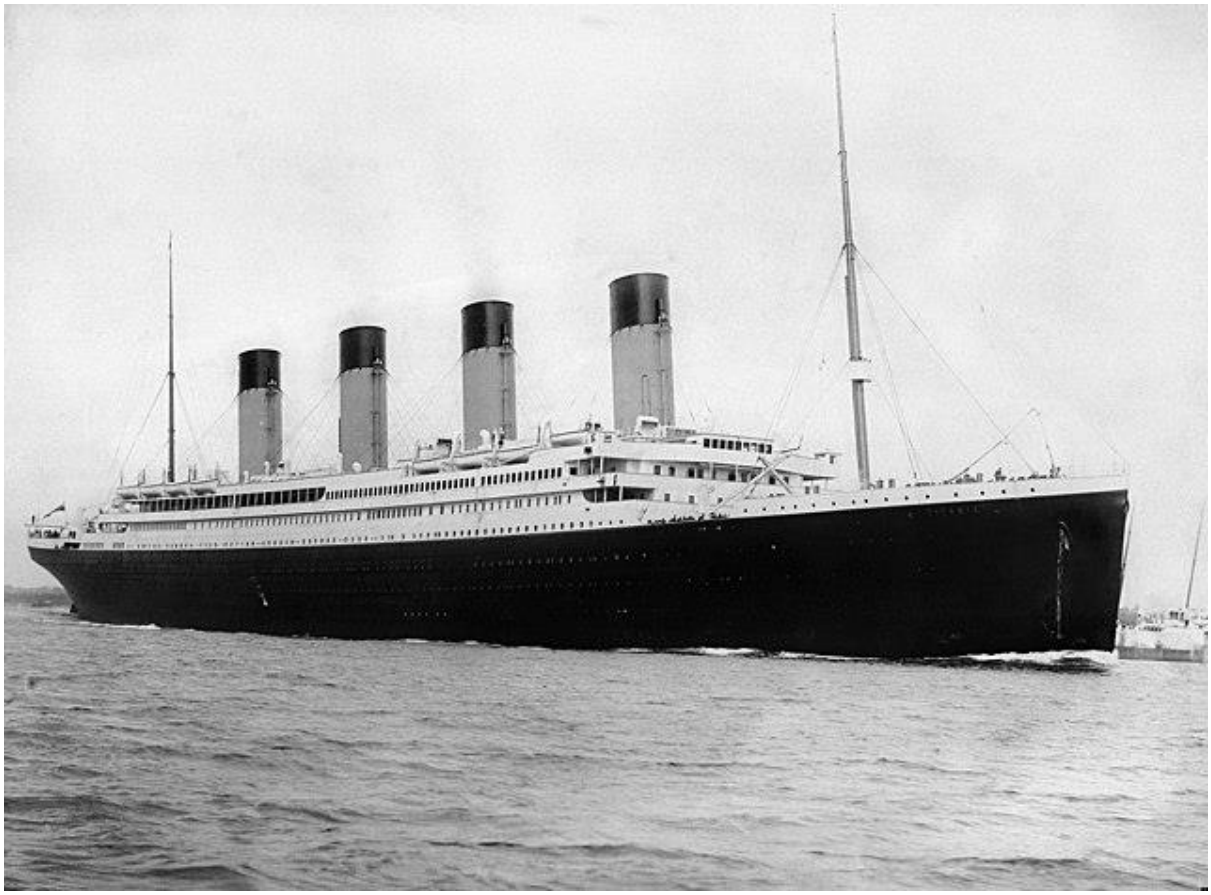
Titanic Survival Prediction



Overview

This project is a Python-based data analysis and machine learning project that predicts whether a person would have survived the sinking of the Titanic based on various features such as age, gender, passenger class, and more. It uses the Titanic dataset, which contains information about passengers on board the Titanic.

Table of Contents

Requirements

Before running the project, you need to have the following requirements installed:

- Python 3.x
- Required Python libraries (scikit-learn, pandas, matplotlib, seaborn)

You can install the required libraries using pip:

bashCopy code

pip install scikit-learn pandas matplotlib seaborn

Usage

To use this project:

1. Clone this repository:

bashCopy code

git clone https://github.com/yourusername/titanic-survival-prediction.git

2. Navigate to the project directory:

bashCopy code

cd titanic-survival-prediction

3. Run the **titanic_survival_prediction.py** script:

bashCopy code

python titanic_survival_prediction.py

4. Follow the on-screen instructions to input passenger information for survival prediction.

Data

The Titanic dataset used in this project contains the following columns:

- PassengerId
- Survived
- Name
- Age
- SibSp (Number of Siblings/Spouses Aboard)
- Parch (Number of Parents/Children Aboard)
- Ticket
- Fare
- Cabin
- Gender (Male or Female)
- Pclass (Passenger Class: 1st, 2nd, or 3rd)

- Embarked (Port of Embarkation: C, Q, or S)

Data Preprocessing

- Missing values in the "Age" and "Embarked" columns are filled with appropriate values.
- Categorical variables ("Gender," "Pclass," "Embarked") are one-hot encoded.

Model Training

A Random Forest Classifier is trained on the preprocessed data to predict survival outcomes.

Data Visualization

- Survival rate by gender is visualized using a count plot.
- Survival rate by passenger class is visualized using a count plot.
- The age distribution of passengers is visualized using a histogram.

Predicting Survival

You can predict whether a person would have survived the Titanic sinking by providing their information in the **predict_df** DataFrame within the script. Replace the values in **predict_df** with the person's information, and the script will predict their survival outcome.