

Title: Titanic Survival Prediction ML Solution Document

Aim

This project aims to predict whether a passenger survived the Titanic disaster using machine learning. We used the classic Titanic dataset, which includes variables like age, sex, fare, and passenger class to build and evaluate predictive models.

Dataset Overview

The dataset contains information on 891 passengers, including:

Passenger class (Pclass)

Sex

Age

Number of siblings/spouses aboard (SibSp)

Number of parents/children aboard (Parch)

Fare paid

Port of embarkation (Embarked)

Survival (0 = Did not survive, 1 = Survived)

Some values (e.g., Age and Embarked) had missing entries and were handled during preprocessing.

Data Cleaning & Preprocessing

Filled missing values in the 'Age' column with the median.

Filled missing values in 'Embarked' with the most frequent value (mode).

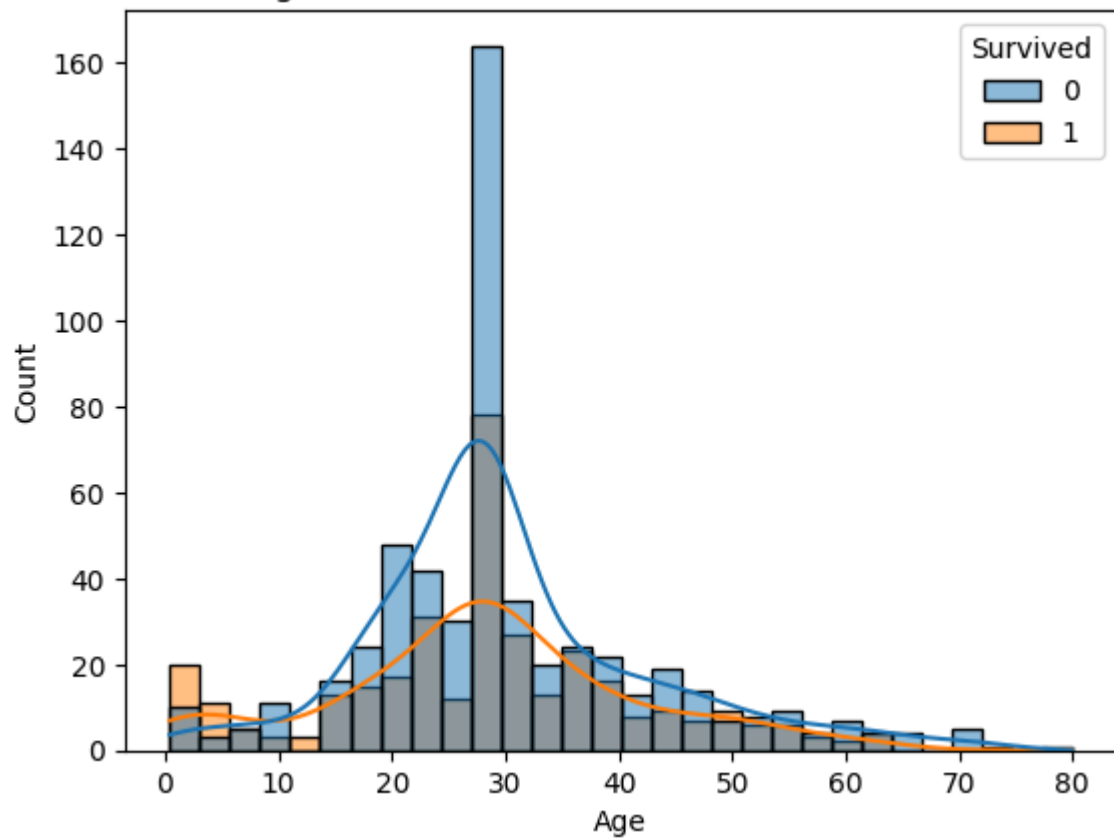
Encoded categorical features:

'Sex': male \rightarrow 0, female \rightarrow 1

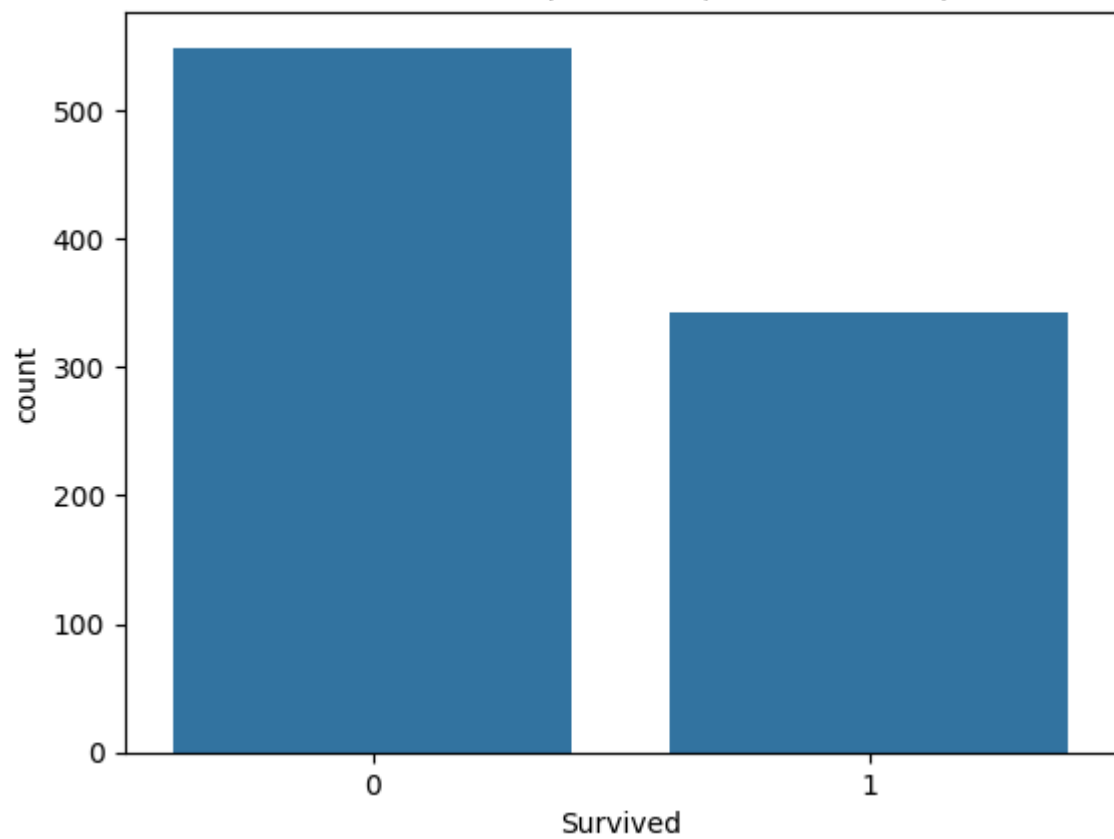
'Embarked': S \rightarrow 0, C \rightarrow 1, Q \rightarrow 2

Selected relevant features: ['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare', 'Embarked']

Age Distribution: Survivors vs Non-Survivors

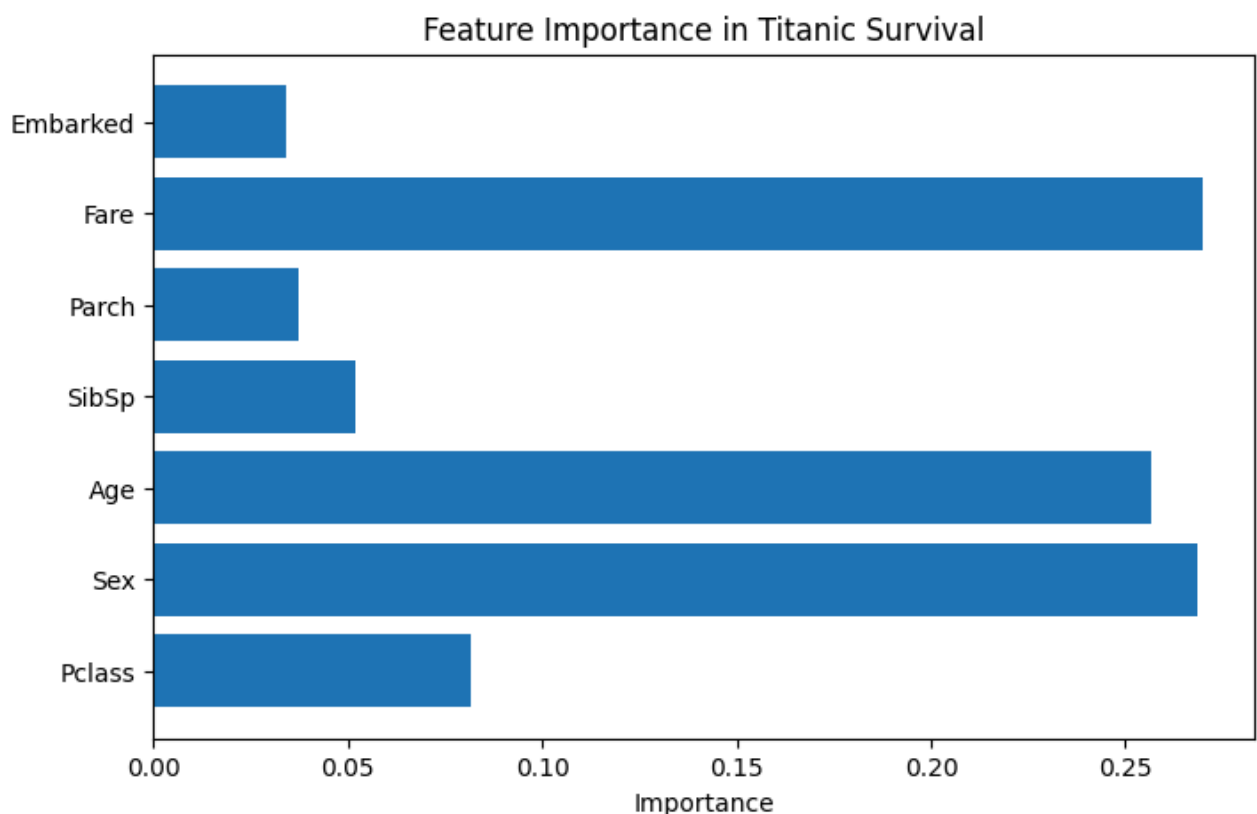


Survival Count (0 = Died, 1 = Survived)



Exploratory Data Analysis (with Plots)

- Survival by Gender: Female passengers had a much higher survival rate.
- Survival by Pclass: Passengers in first class were more likely to survive.
- Age Distribution: Younger passengers had higher chances of survival.
- Fare vs Survival: Those who paid higher fares were more likely to survive.



Modeling

- Model Used: Logistic Regression, Random Forest Classifier
- TrainTest Split: 80% training, 20% testing
- Evaluation Metrics:
 - Accuracy
 - Precision, Recall, F1score
 - Confusion Matrix

```
♦ Logistic Regression Results
Accuracy: 0.7988826815642458
Confusion Matrix:
[[89 16]
 [20 54]]
Classification Report:
              precision    recall  f1-score   support

     0       0.82         0.85         0.83         105
     1       0.77         0.73         0.75          74

 accuracy          0.80         0.80         0.80         179
 macro avg         0.79         0.79         0.79         179
 weighted avg         0.80         0.80         0.80         179
```

```
♦ Random Forest Results
Accuracy: 0.8268156424581006
Confusion Matrix:
[[92 13]
 [18 56]]
Classification Report:
              precision    recall  f1-score   support

     0       0.84         0.88         0.86         105
     1       0.81         0.76         0.78          74

 accuracy          0.83         0.83         0.83         179
 macro avg         0.82         0.82         0.82         179
 weighted avg         0.83         0.83         0.83         179
```

Conclusion

Using Random Forest, we were able to predict the survival of Titanic passengers with an accuracy of over 80%. The most significant factors were gender, ticket class, and fare. The model can be further enhanced by engineering new features and fine-tuning parameters.

