Titanic Survival Prediction Report

1. Introduction

This project involves building a classification model to predict the survival of Titanic passengers based on features like age, sex, class, and fare. Logistic Regression was used for the prediction task.

2. Data Summary and Preprocessing

Dataset Used:

The dataset used was the Titanic dataset from Kaggle, consisting of 891 entries and 12 columns, including both categorical and numerical features.

Data Preprocessing:

Missing values in 'Age' were filled with the median. Categorical variables like 'Sex' and 'Embarked' were encoded appropriately. The 'Cabin' feature was dropped due to excessive missing values.

3. Model and Evaluation

Model:

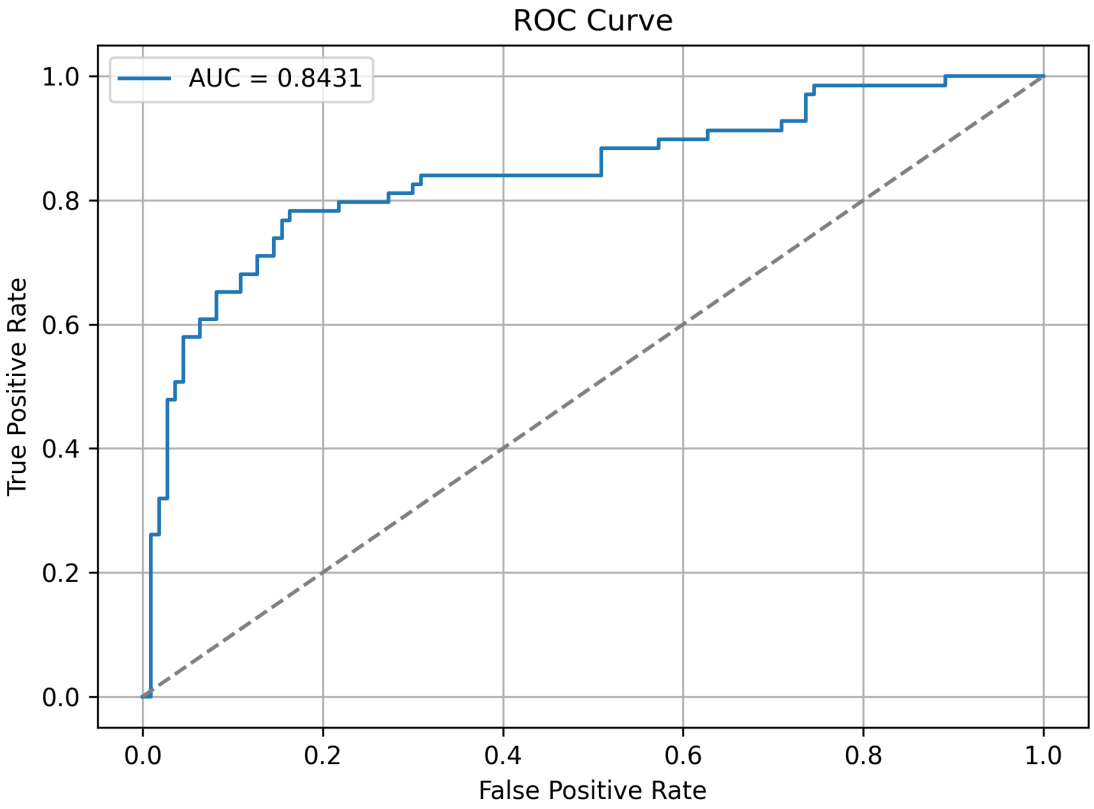
Logistic Regression

Evaluation Metrics:

- Accuracy: ~81%
- Precision, Recall, F1-score, and AUC were also used to assess performance.

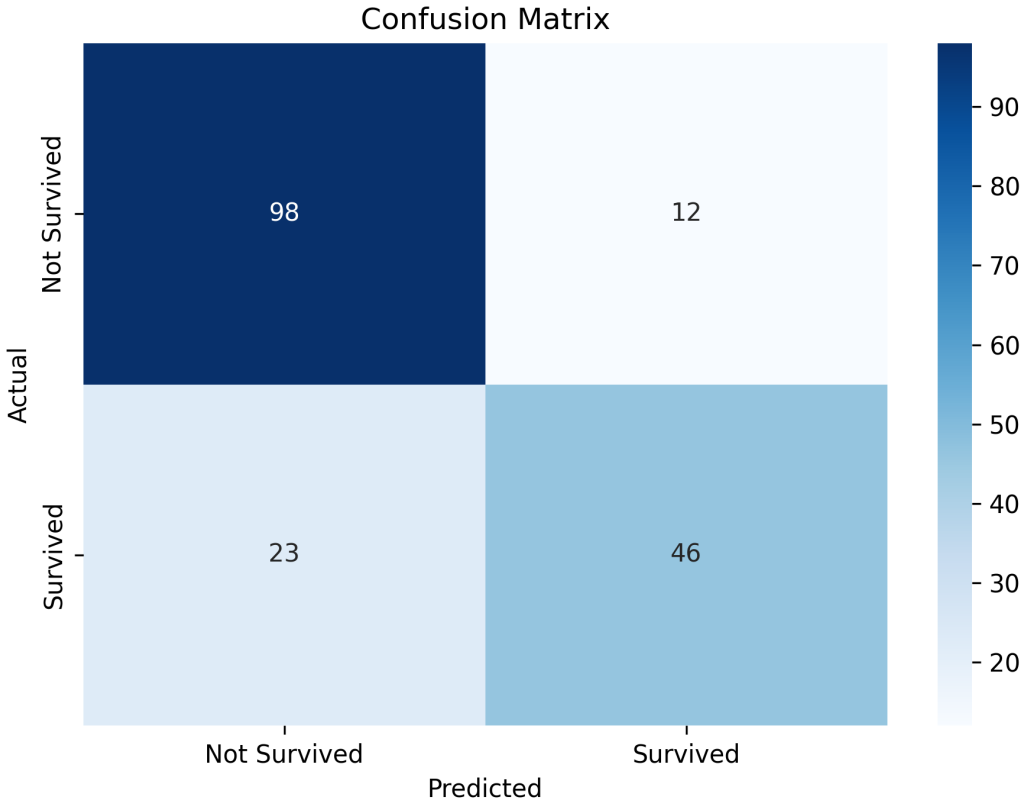
ROC Curve (AUC = 0.8431)

Titanic Survival Prediction Report



Confusion Matrix

Titanic Survival Prediction Report



4. Conclusion

Conclusion:

The model demonstrated good predictive ability with an AUC of 0.84. The ROC curve shows strong discrimination between classes. The confusion matrix highlights a balanced performance across survival classes.