Team : Bhavesh Naidu Kulluru, Haarkarandeep  
  
Titanic Dataset Binary Classification Report

# 1. Dataset Overview

The Titanic dataset is a famous dataset used in binary classification tasks. It contains information about the passengers aboard the Titanic, including whether they survived or not. The key features used for prediction include:  
- Pclass: Passenger class (1st, 2nd, 3rd)  
- Sex: Gender of the passenger  
- Age: Age of the passenger  
- Fare: Fare paid by the passenger  
- SibSp: Number of siblings/spouses aboard  
- Parch: Number of parents/children aboard  
- Embarked: Port of embarkation (C, Q, S)  
- Survived: Target variable indicating if the passenger survived (1) or not (0)  
  
We have picked dataset from Kaggle.com : <https://www.kaggle.com/code/sandragracenelson/titanic-dataset-prediction/input>

# 2. Preprocessing

Handling Missing Data:  
- Age: Missing values were filled using the median age.  
- Embarked: Missing values were filled with the most common embarkation point (mode).

Encoding:  
- Categorical Variables: The categorical variables like Sex and Embarked were converted to numerical values using label encoding.  
  
Splitting the Data:  
- The dataset was split into a training set (80%) and a test set (20%) using train\_test\_split().

# 3. Model Choice

The Perceptron model was chosen for this task. The Perceptron is a simple linear binary classifier that updates its weights iteratively to minimize classification errors. It was selected because:  
- It's easy to implement.  
- It's well-suited for linearly separable data, which was assumed after initial analysis of the features.  
- It’s computationally efficient for this binary classification task.

# 4. Model Performance

The following metrics were used to evaluate the Perceptron model:  
Accuracy: 0.6433566433566433

Precision: 0.6

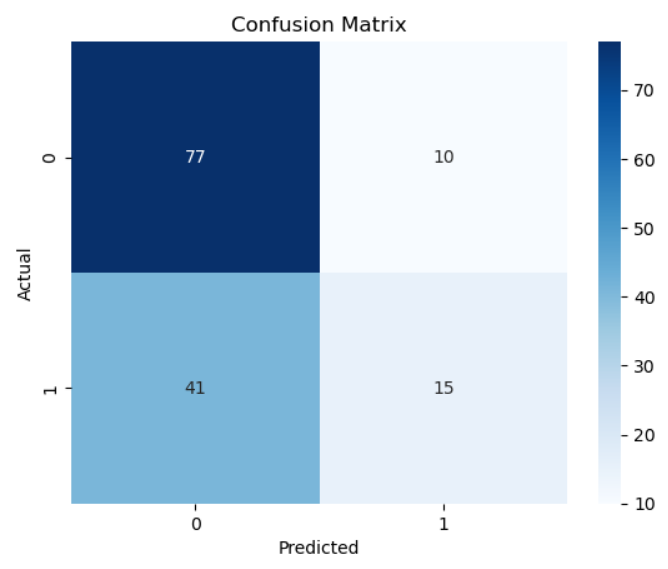
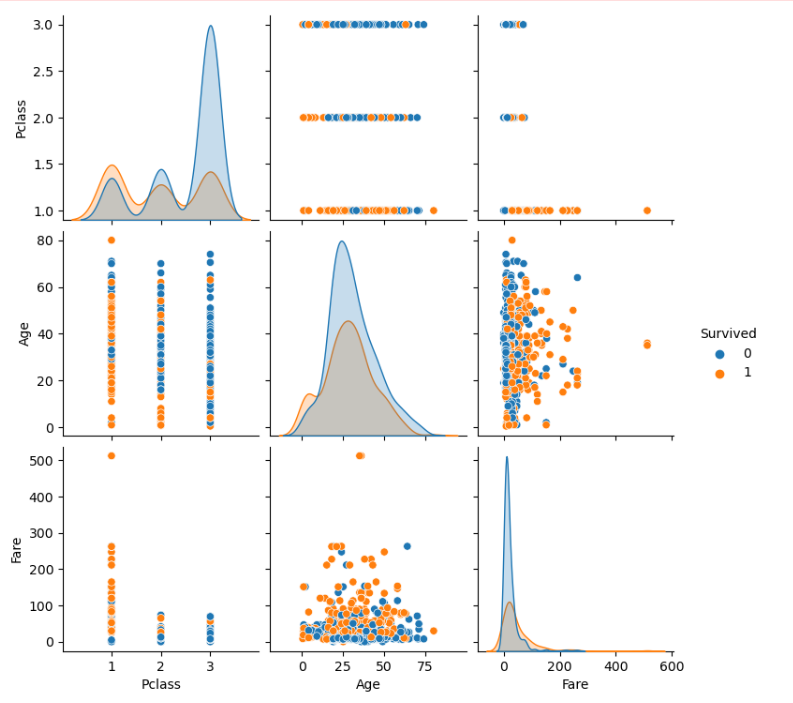
Recall: 0.26785714285714285

F1 Score: 0.37037037037037035

While the accuracy was moderate, the precision and recall values indicate that the model struggled to generalize well on the test data, particularly in identifying the positive class (Survived).

# 5. Visualizations

Confusion Matrix:  
A confusion matrix was generated to visualize the performance of the model:

   
SNS Pairplot  
  


# 6. Insights and Recommendations

Insights:  
- The model performed reasonably well on non-survivor predictions but had difficulty predicting survivors, as shown by the low recall and F1-score.  
- Features like Sex and Pclass seemed to have the most influence on the survival prediction.  
  
Recommendations for Improvement:  
- Feature Engineering: Further feature engineering could improve performance. For example, combining SibSp and Parch into a single feature representing family size could capture more nuanced relationships.  
- Advanced Models: Trying more advanced models like Support Vector Machines, Random Forests, or Gradient Boosting could lead to better predictive performance.  
- Hyperparameter Tuning: Tuning the learning rate and iterations of the Perceptron model may help it converge better and provide better results.