# Titanic Dataset Binary Classification Report

## Team Members:

Bhavesh Naidu Kulluru, Haarkarandeep, Sathwik Chowdary

## 1. Dataset Overview

The Titanic dataset is a widely used dataset for binary classification tasks. It contains information about passengers aboard the Titanic, including features such as: - 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) The target variable 'Survived' indicates if a passenger survived (1) or not (0).

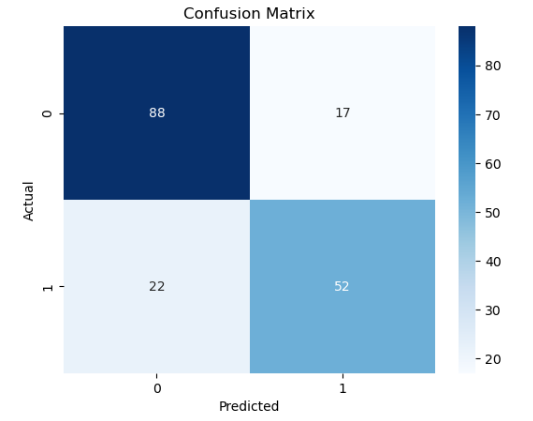
## 2. Data Preprocessing

To prepare the dataset for modeling, the following preprocessing steps were performed:  
- Missing values in the 'Age' column were filled using the median age.  
- Missing values in the 'Embarked' column were filled with the most frequent value (mode).  
- Categorical variables, such as 'Sex' and 'Embarked,' were encoded numerically using label encoding.  
- The dataset was split into training (80%) and testing (20%) sets using train\_test\_split.  
- Feature scaling was applied using StandardScaler to normalize the features.

## 3. Model Choice

The Support Vector Machine (SVM) model with a linear kernel was chosen for this task. SVM was selected due to its effectiveness in binary classification tasks and its ability to find the optimal hyperplane that separates the two classes.

## Model Performance



The SVM model's performance was evaluated using the following metrics:  
- Accuracy: 78.21%  
- Precision: 75.36%  
- Recall: 70.27%  
- F1 Score: 72.73%

The confusion matrix provided insights into the classification results:  
Confusion Matrix:  
[[88, 17],  
 [22, 52]]

## 5. Insights and Recommendations

Insights:  
- The model achieved reasonable accuracy and performed well in predicting both survived and non-survived classes.  
- Features such as 'Sex' and 'Pclass' were critical in determining survival.

Recommendations for Improvement:  
- Experimenting with different kernels (e.g., RBF) and hyperparameters in the SVM model.  
- Performing advanced feature engineering, such as combining 'SibSp' and 'Parch' into a single family size feature.  
- Using ensemble methods like Random Forests or Gradient Boosting for potentially better performance.  
  
