Understanding the problem:

Predict which passengers survived the Titanic shipwreck using passenger data.

Task = Binary Classification

Predict: Survived (0 or 1)

Metric: Accuracy

Input Files: train.csv, test.csv, gener\_submission

Dataset Features:

PassengerId – unique ID

Pclass – passenger class (1st, 2nd, 3rd)

Name – passenger name (can extract title: Mr, Mrs, Miss, etc.)

Sex – gender

Age – age in years (some missing values)

SibSp – # of siblings/spouses aboard

Parch – # of parents/children aboard

Ticket – ticket number

Fare – ticket price

Cabin – cabin number (many missing)

Embarked – port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton)

Survived – target variable (only in train.csv)