Bài toán đặt ra dựa trên một thảm họa có thật trong lịch sử, vụ chìm tàu Titanic. Đầu vào là 2 file csv, 1 file train.csv chứa dữ liệu huấn luyện và 1 file test.csv chứa dữ liệu kiểm tra.   
File train.csv có chứa các thuộc tính pclass, survived, name, sex, age, sibsp, parch, ticket, fare, cabin, embarked, boat , body, home.dest

File test.csv cũng chứa các thông tin như trên, và cần xác định thông tin cho survived

Yêu cầu bài toán là xác định survived từ các thuộc tính còn lại trong tập kiểm tra.

Có thể kiểm tra kết quả bằng file gender\_submission.csv