TEBONG ROLAND

DSC680-T301 Applied Data Science (2243-1)

1.1 Project 1: Proposal and Data Selection/Milestone 1

Titanic Survival Prediction

Topic:

"A Machine Learning Approach to Analyzing Titanic Survival Rates"

Context:

The sinking of the Titanic is one of the most infamous shipwrecks in history. On April 15, 1912, during her maiden voyage, the widely considered “unsinkable” RMS Titanic sank after colliding. with an iceberg. Unfortunately, there weren’t enough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew. While there was some element of luck involved in surviving, it seems some groups of people were more likely to survive than others.

Business Problem:

This project aims to understand and predict survival rates during maritime disasters, using the Titanic tragedy as a case study. By analyzing passenger data, we seek to identify key factors that influenced survival chances. This could provide insights for modern maritime safety protocols and evacuation strategies, enhancing passenger safety in current and future maritime travel. The objective of this problem is to build a predictive model that answers the question: “what sorts of people were more likely to survive?” using the passenger data (i.e., name, age, gender, socioeconomic class and more.).

Dataset:

The primary dataset for this project will be sourced from the "Titanic: Machine Learning from Disaster" competition on Kaggle.

- survival - If passenger will survive or not (Yes - 1, No - 0).

- pclass - Ticket class (1st, 2nd, and 3rd).
- sex - Sex of the passenger (M / F).
- Age - Age in years.
- sibsp - The number of siblings / spouses aboard the Titanic.
- parch - The number of parents / children aboard the Titanic.
- ticket - Ticket number.
- fare - Passenger fare.
- cabin - Cabin number.
- embarked - Port of embarkation.