Introduction

Takuto Yoshida

2022-12-14

Introduction

Titanic was the name of a British luxury liner with a gross tonnage of 46,328 tons. On her maiden voyage to New York in 1912, she collided with an iceberg off the coast of Newfoundland on April 14 and sank. It was the world's largest maritime accident, claiming the lives of 1514 of the 22008 people on board. This was the first time that the internationally established distress signal "SOS" was sent out. Later, this maritime accident was featured in Hollywood, where Leonardo DiCaprio and Kate Winslet gave glamorous and tragic performances. The film won 11 Academy Awards in 1998, including Best Picture, and its total U.S. box office gross of $659,363,944 ranks seventh on the U.S. all-time list. Predicting the Titanic's survivors is the gateway problem for Kaggle, the world's largest platform for data scientists. Using this data set, I created multiple models and evaluated their performance to determine which model performed the best.

Objective

What I have done in this project is the following. * To make the data set tidy so that it can be analyzed. * Evaluate each variable and create variables to be used for forecasting. * Create multiple models and select the model with the best predictive performance.

Methods

For each variable, I evaluated its association with mortality to determine if it was useful for prediction. I created variables that I considered useful for forecasting. If there were missing values, I substituted values in multiple ways to ensure that there were no missing values in the data. Finally, multiple machine learning models were created using these variables, and their predictive ability was evaluated.