Machine Learning From Disaster

Calculating the Survival Rate for Passengers in the Titanic

Miguel Melo Ochoa, Alex Hayet, Francisco Gomez, Maeki Kashana

CS549 - Machine Learning
Professor Xin Zhang
May 9, 2025
San Diego State University

I. Introduction and Research Problem

All details for this project will be provided by a competition titled "Titanic - Machine Learning from Disaster" introduced by kaggle.com [1].

This following section will provide a brief background and a statement of the problem this project is addressing.

A. Introduction

This project will build a predictive model that aims to find the type of people who are going to survive the sinking of the Titanic, a well renowned ship that was believed to be "unsinkable" until it hit an iceberg in 1912, resulting in the death of 1502 passengers/crew.

This project will utilize passenger information, such as name, age, gender, socio-economic class, etc., to determine what sorts of groups were more likely to survive than others during the sinking of the Titanic.

This data aforementioned above will be provided by the Kaggle competition and the goal of this project is to engineer code that uses the data to calculate the survival rate of various groups of people.

B. Research Problem

This project will be creating a predictive model that will dictate the type of individuals that were the most likely to survive the sinking of the Titanic.

To give more detail, this project will be provided with a lot of testing data in order to make predictions on who will survive. In consequence of this data, this team will be able to create a predictive machine learning model that utilizes the data provided for this project to make predictions using various models previously covered in this course. The specific details regarding any related work that relates to this project will be discussed in depth in later sections.

To get more into specifics, this project will utilize these following .csv files to find a solution for this problem: test.csv and train.csv.

For the train.csv file, it will contain information on who survived the sinking of the Titanic in order to train the model that was developed.

For the test.csv file, it is used to see how well the model that was developed performs on unseen data. This file will be used to make predictions on who survived based on the model that was trained using train.csv.

II. Related Work

III. Methodology and Technical Details

A. Methodology

B. Technical Details

IV. Experimental Results and Analysis

A. Experimental Results

B. Analysis

V. Conclusion and Future Work

A. Conclusion

B. Future Work

VI. Contributions

A. Maeki Kashana

1) Significant Contributions to the Final Report
   a) Introduced the Research Problem
   b) Discussed, at length, any past work in the course that relates to this project
   c) Summarized findings and suggested possible future directions
2) Created the README file

B. Miguel Melo Ochoa

1)

C. Alex Hayet

1)

D. Francisco Gomez

1) Methodology and Technical Details
2) Provided ideas on how to implement the code

References

[1] Cubierski. W., “Titanic - Machine Learning From Disaster,” Kaggle.com, 2012.