# Project Proposal - Survival in the Titanic

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

San Diego State University

CS549 - Machine Learning

Professor Xin Zhang

April 4, 2025

# Project Proposal - Survival in the Titanic

## Contents

## Problem Statement

All details for this project will be provided by a competition titled "Titanic - Machine Learning from Disaster" introduced by kaggle.com (Cukierski, 2012).

### Project Description and Statement

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.

### Project Motivations

What motivated this group to take on the challenge of creating a predictive model that will dictate the type of individuals that were the most likely to survive the sinking of the Titanic is that this project provided an adequate challenge from a machine learning perspective.

To give more detail, this project will provide 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 on who will survive. Moreover, this project will enable the team to practice the machine learning concepts previously learned in order to find a solution for this problem.

In addition, another motivation that this team had for undertaking this specific project is that this project requires the primary use of the "numpy" and "pandas" python packages in order to find a solution. As a result, this project will enable this team to become more skillful in using these packages to manipulate data and apply machine learning concepts.

## Dataset Description

## Planned Methodology

a

## Evaluation Metrics & Expected Outcomes

When looking at the evaluation metrics of the Titanic problem, the problem uses binary classification in order to show if a passenger survives the tragedy. In this instance, if a passenger is marked with a 1 this means they survived and a 0 would mean that they did not survive. When wanting to make a prediction on whether or not a passenger will survive there are a few metrics that we could use in order to find out the most likely outcome for the passenger. The first metric to look at would be the gender of the passengers. When looking at the data it seems that female passengers had a much larger survival rate than men passengers. Only about 20% of men survived aboard the Titanic while around 75% of women survived. This is largely due to the fact that women and children were the number one priority to get on the lifeboats first. Passenger class is another metric that affects the survival rate of passengers. First class passengers were more likely to survive in the event of sinking due to being higher up in the ship than second or third class passengers. Age is another factor that helped in the survival rate for passengers on the Titanic. As stated earlier, women and children were the highest priority passengers to get onto the lifeboats. This had a large impact on the survival rate for children aboard the Titanic and gave them a better chance at surviving. The ports in which passengers had boarded the Titanic from do show a correlation to survival. This is a socio-economic factor that does link back to the class of passengers as we can see a connection between the ports passengers boarded from to a class level on the boat. For example, Cherbourg has a lot of passengers that boarded into the first class.which means that they had a higher chance of surviving than passengers that boarded from Queenstown which had more third class passengers than Cherbourg. These are just some of the metrics that can be used to find expected outcomes in the event of the Titanic.

## Work Distribution

**Maeki Kashana**

1. Describe the problem statement and motivation
2. Download data required for project
3. Set up coding environment for the project
4. Help with finding a solution for the problem with other team members and work on sections of the code to implement the solution.

**Miguel Melo Ochoa**

**Alex Hayet**

**Francisco Gomez**

## References

Cukierski, W. (2012). Titanic - machine learning from disaster [Kaggle].