**Introduction**

The sinking of the RMS titanic is one of the most infamous shipwrecks in history. On April 15, 1912 the largest passenger liner ever made during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This sensational tragedy shocked the entire world and the international community and led to better safety regulations for ships. One of the reasons that the shipwreck resulted in such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others. The purpose of this analysis is to explore factors that influenced a person’s likelihood to survive. While there could hardly be a more chaotic event than frightened people scrambling to escape a sinking ship, the disaster is famous for saving “women and children first”.

The project is all about which sort of people traveling mostly survived and were likely to survive and we’ll try to predict who they were. Considering age, class, gender, family and some other important features of the passengers, the analysis is done. Most of the project is data exploration and basic model building. With this project, we will try to analyze which sort of people were given most importance in the process of evacuation after the wreck.

**Research Question:**

Who were the people most likely to survive in such an accident, where the survival was just a matter of luck and what sorts of people were likely to survive?

**Machine learning model**

I will be using Random Forest model as a decision analysis model and to create a model predicting survival on the Titanic. The data set I am using inherently support decision trees. The number of features in the set are less, so randomforest approach wouldn’t take many computational resources. It gives us more accuracy when compared to decision tree.

I will use R programming language to build the code and to visualize my data set I would probably use bar graphs, plots and try to understand the relationship between variables. The data I am using is already divided in to test and training datasets to build the prediction model, so test data can be used to test the performance of learned model.

My focus would be considering some of the important features that could impact the data analysis and predict which passengers survived the tragedy. The practical use of the results of the analysis would be that - Rather than rely on assumption, we will know exactly who were the people most likely to survive? I am sure how deepen this analysis goes we will be benefited practically from the results.