# Titanic: Machine Learning from Disaster

**Predict survival on the Titanic using Excel, Python, R & Random Forests**

If you're new to data science and machine learning, or looking for a simple intro to the Kaggle competitions platform, this is the best place to start. Continue reading below the competition description to discover a number of tutorials, benchmark models, and more.

# Competition Description

Source: <https://www.kaggle.com/c/titanic>

The sinking of the RMS Titanic is one of the most infamous shipwrecks in history. On April 15, 1912, 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 international community and led to better safety regulations for ships.

One of the reasons that the shipwreck led to 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, such as women, children, and the upper-class.

In this challenge, we ask you to complete the analysis of what sorts of people were likely to survive. In particular, we ask you to apply the tools of machine learning to predict which passengers survived the tragedy.

# Evaluation

Source: <https://www.kaggle.com/c/titanic/details/evaluation>

The historical data has been split into two groups, a 'training set' and a 'test set'.  For the training set, we provide the outcome ( 'ground truth' ) for each passenger.  You will use this set to build your model to generate predictions for the test set.

For each passenger in the test set, you must predict whether or not they survived the sinking ( 0 for deceased, 1 for survived ).  Your score is the percentage of passengers you correctly predict.

The Kaggle leaderboard has a public and private component.  50% of your predictions for the test set have been randomly assigned to the public leaderboard ( the same 50% for all users ).  Your score on this public portion is what will appear on the leaderboard.  At the end of the contest, we will reveal your score on the private 50% of the data, which will determine the final winner.  This method prevents users from 'overfitting' to the leaderboard.

# Prizes

Source: <https://www.kaggle.com/c/titanic/details/prizes>

The confidence to go forward and compete for some serious $$$$$$

# Other References

* Youtube Tutorial by David Langer - <https://www.youtube.com/playlist?list=PLTJTBoU5HOCRrTs3cJK-PbHM39cwCU0PF> (github source for this tutorial <https://github.com/EasyD/IntroToDataScience>)

# Data

Source: <https://www.kaggle.com/c/titanic/data>

* test.csv
* train.csv

VARIABLE DESCRIPTIONS:

survival Survival

(0 = No; 1 = Yes)

pclass Passenger Class

(1 = 1st; 2 = 2nd; 3 = 3rd)

name Name

sex Sex

age Age

sibsp Number of Siblings/Spouses Aboard

parch Number of Parents/Children Aboard

ticket Ticket Number

fare Passenger Fare

cabin Cabin

embarked Port of Embarkation

(C = Cherbourg; Q = Queenstown; S = Southampton)

SPECIAL NOTES:

Pclass is a proxy for socio-economic status (SES)

1st ~ Upper; 2nd ~ Middle; 3rd ~ Lower

Age is in Years; Fractional if Age less than One (1)

If the Age is Estimated, it is in the form xx.5

With respect to the family relation variables (i.e. sibsp and parch)

some relations were ignored. The following are the definitions used

for sibsp and parch.

Sibling: Brother, Sister, Stepbrother, or Stepsister of Passenger Aboard Titanic

Spouse: Husband or Wife of Passenger Aboard Titanic (Mistresses and Fiances Ignored)

Parent: Mother or Father of Passenger Aboard Titanic

Child: Son, Daughter, Stepson, or Stepdaughter of Passenger Aboard Titanic

Other family relatives excluded from this study include cousins,

nephews/nieces, aunts/uncles, and in-laws. Some children travelled

only with a nanny, therefore parch=0 for them. As well, some

travelled with very close friends or neighbors in a village, however,

the definitions do not support such relations.