Data-Set Source - <https://www.kaggle.com/c/titanic/data?train.csv>

Number of attributes - 11

Number of instances - 891

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)

Survival survival (0=No, 1= Yes)

Speical 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.

Removed attributes passenger no.(unique attribute), name(unique attribute), cabin(only few values are available and are almost unique), ticket no.(duplicates to Pclass attribute) & ticket fare (duplicates to Pclass attribute)

Embarked

4 – S

2 – C

3 – Q

Age

NA -> Replaced by average

Categorized into 3 different range of values

1 -> 18 and below

2 -> 18 above and below 40

3 -> above 40

Sex

Male -> 1

Female -> 2