1. For the scoring mechanism implement linear regression and do ROC/AUC analysis.
2. Implement logistic regression to compare the ROC/AUC w.r.t linear regression.

WRT Data Cleaning:-

1. Removed “Name” column and “Ticket description” column from the training data set.
2. Removed “Cabin number” column for two reasons, 1. Not too many entries are given, 2.Cannot form any relational analysis whether the passenger survived or not.
3. Of the 891 entries in the training set, 177 entries have no age entered. To keep the mean for ages of the entire dataset constant, mean age of (891-177=714) entries is filled for all unfilled entries.
4. Filled the age as 30 for all the 177 missing entries
5. For ease of data understanding, male=0, female=1
6. For ease of data understanding,