

Group 1 did the Titanic challenge to predict which passengers survived or not, using logistic regression (forward+backward), decision tree, and random forest.

The report is concise, graphs are easy to understand and conclusions are relevant. Use of unique Random Forest model to fill in missing age values was quite creative. The report also experimented with different models for prediction and accurately summarized the most important features to their chosen model (random forest.)

Result of 0.82 is very good. The level of feature engineering used is a key strength of this report, as Cai and Song extrapolated many new and potentially relevant features (isAlone, fare per person). Adding interaction terms to regression model was unique but more discussion could be given on which features require said interaction terms (perhaps some domain knowledge).

Slides seem well prepared but data overview part far too long - maybe no need to explain each variable step by step, just understanding which variables are categorical vs which are quantitative enough. As a result, first part of presentation vastly overshadowed 2nd and more important part (model selection and results).

Quality of writing: 4

Presentation: 2.5

Creativity: 4

Confidence in assessment: 3

=====