Features used: Age, Pclass, SibSp, Parch, Fare and Embarked

Train Data Processing Decisions

(1) Data over-sampled to balance distribution of survived.

(2) Age and Fare standardized to have mean 0 and standard deviation 1.

(3) 70% of training data used to train and 30% of data used to validate

(4) Observations with missing age values removed from the data set.

Test Data Prediction Method

(Step 1) Random Forests was used to predict all observations, except observations missing age values.

(Step 2) Decision Tree was used to specifically predict the observations with missing age values.

(Step 3) Decision Tree, Random Forests, K-Nearest Neighbours and Logistic Regression were all fitted and combined to confidently classify approx 50% of the test data.