Presentation

The sinking of the Titanic is one of the most famous shipwrecks in history. 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. And this is the focus of the feature engineering.

Here is the dictionary of the the data set. The dataset have these features….

The training set has 891 samples and test set has 418 samples, the train set have all of this 12 features the test set dont include the survival feature since it is the target. and both sets have some.missing values.

Parameters

max\_depth (int) – Maximum tree depth for base learners.

learning\_rate (float) – Boosting learning rate (xgb’s “eta”)

n\_estimators (int) – Number of trees to fit.

verbosity (int) – The degree of verbosity. Valid values are 0 (silent) - 3 (debug).

objective (string or callable) – Specify the learning task and the corresponding learning objective or a custom objective function to be used (see note below).

booster (string) – Specify which booster to use: gbtree, gblinear or dart.

tree\_method (string) – Specify which tree method to use. Default to auto. If this parameter is set to default, XGBoost will choose the most conservative option available. It’s recommended to study this option from parameters document.

n\_jobs (int) – Number of parallel threads used to run xgboost.

gamma (float) – Minimum loss reduction required to make a further partition on a leaf node of the tree.

min\_child\_weight (int) – Minimum sum of instance weight(hessian) needed in a child.

max\_delta\_step (int) – Maximum delta step we allow each tree’s weight estimation to be.

subsample (float) – Subsample ratio of the training instance.

colsample\_bytree (float) – Subsample ratio of columns when constructing each tree.

colsample\_bylevel (float) – Subsample ratio of columns for each level.

colsample\_bynode (float) – Subsample ratio of columns for each split.

reg\_alpha (float (xgb's alpha)) – L1 regularization term on weights

reg\_lambda (float (xgb's lambda)) – L2 regularization term on weights

scale\_pos\_weight (float) – Balancing of positive and negative weights.

base\_score – The initial prediction score of all instances, global bias.

random\_state (int) –