Grid Search
applicable params for each classifier
CV = 5
data : master_X_train
master_y_train

Model Setup for full validation
'best' model from grid search results
data : X_train, y_train
verify results consistent with Grid Search results : X_test, y_test (1x)

10-fold Cross Validation
data : master_X_train
master_y_train

Best Overall 10-fold cross-validated model
report results from golden data set
golden_X_test
golden_y_test

20% of full data
sacred test set
golden_X_test
golden_y_test

80% of full data set
master_X_train
master_y_train

multinomial logistic regression
params = p1, p2, p3 ...

decision tree
params = p1, p2, p3 ...

random forest
params = p1, p2, p3 ...

naïve bayes
params = p1, p2, p3 ...

X_train
y_train
(80% of master_train)

X_test
y_test
(20% of master_train)

multinomial logistic regression
params = p1, p2, p3 ...

decision tree
params = p1, p2, p3 ...

random forest
params = p1, p2, p3 ...

naïve bayes
params = p1, p2, p3 ...

multinomial logistic regression
params = p1, p2, p3 ...

decision tree
params = p1, p2, p3 ...

random forest
params = p1, p2, p3 ...

naïve bayes
params = p1, p2, p3 ...

best ?

??? model
params = p1, p2, p3 ...