Random forest classifier

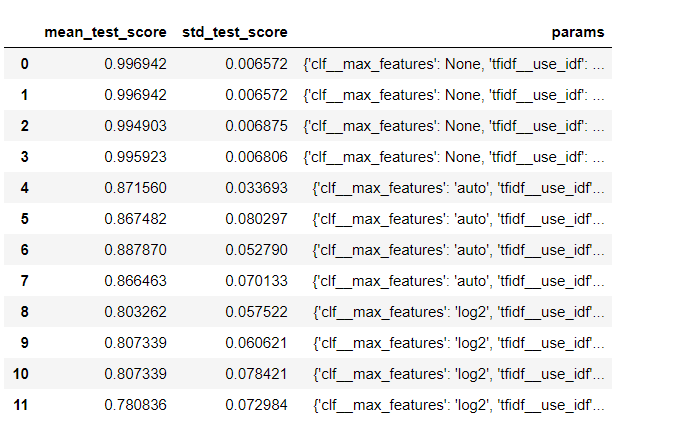
Parameters

* max\_features-maximum no of features Random forest is allowed to try in individual tree

more features improves the performance of the model.

parameter values-

1. None-n\_features
2. sqrt/auto-sqrt(n\_features)
3. log2-log2(n\_features)



Best result-> None

0.9969418960244648

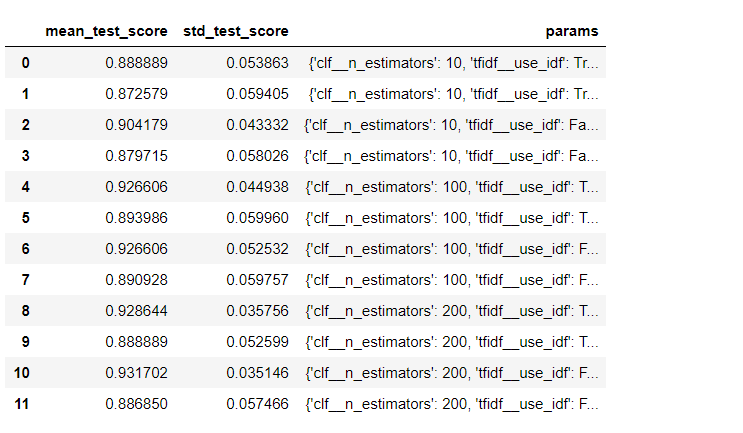
{'clf\_\_max\_features': None, 'tfidf\_\_use\_idf': True, 'vect\_\_ngram\_range': (1, 1)}

* n\_estimators- No of trees want to build

High no of trees. Better performance

Parameter values-

1. 10,100,200



Best result-

0.9317023445463812

{'clf\_\_n\_estimators': 200, 'tfidf\_\_use\_idf': False, 'vect\_\_ngram\_range': (1, 1)}

* min\_sample\_leaf- Minimum no of samples required to be at a leaf.

Leaf-end node of the decision tree

Values- 1, 10, 50, 100

Smaller leaf makes more prone to capture noise in train data.

Best result-

0.9011213047910296

{'clf\_\_min\_samples\_leaf': 1, 'tfidf\_\_use\_idf': False, 'vect\_\_ngram\_range': (1, 1)}

