

# INFO

---

## Libraries Used:

The external libraries used are scikit-learn and nltk. Comments are provide to describe what each method does.

The following are some of the methods that have been used.

**1) DecisionTreeClassifier():** Implements the decision tree classifier. The default criterion used is "GINI" and it automatically splits taking the best attribute based on GINI index.

Attributes: a) Max\_depth: The maximum depth of the tree. If not specified the nodes are expanded until all the leaves are pure.

**2) KNeighborsClassifier():** Implements K nearest neighbours.

Attributes:

a) n\_neighbors: Number of neighbours to use .

b) weights : Weight functions used for our prediction. We used Euclidean distance metric for both the algorithms.

**3) AdaboostClassifier():** Implements Adaboost classifier. This classifier is a meta estimator which begins fitting a classifier on the original dataset and then fits additional copies of the classifier on the same dataset, but the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on difficult areas.

Attributes:

a) Base\_estimator : Specifies the base estimator from which the boosted ensemble is built. We used DecisionTreeClassifier() as base estimator.

b) n\_estimator: Specifies the maximum number of estimators at which the boosting is terminated.

c) learning\_rate : Learning rate shrinks the contribution of each of the classifier. There is a tradeoff between learning\_rate and n\_estimators.

The following are some of the other important methods used:

**1) Predict():** This is used for testing the classifier.

**2) metrics.accuracy\_score(predicted, groundtruth):** This gives the accuracy of prediction made on the testing.