Support Vector Machine

SVM is a supervised machine learning model. The main objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N being the number of features) provided the hyper-plane distinctly classify the data points.

Class: SVC(kernel = "rbf", random_state = 0)

kernel{‘linear’, ‘poly’, ‘rbf’, ‘sigmoid’, ‘precomputed’}, default=’rbf’

Kernel rbf radial basis function. The RBF kernel function for two points X_1 and X_2 computes the similarity or how close they are to each other

https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

Decision Tree

It is a supervised machine learning algorithm wherein data is frequently split according to a certain variable. The decision tree asks a series of questions about the attributes of the record. Each time it receives an answer, it further asks up a question till it reaches about a conclusion on the label of the class record.

Decision tree uses the tree representation to solve the problem in which each leaf node corresponds to a class label and attributes are represented on the internal node of the tree.

Class : DecisionTreeClassifier from sklearn.tree

criterion{“gini”, “entropy”}, default=”gini”

Gini Index is a metric to measure how often a randomly chosen element would be incorrectly identified.

It means an attribute with lower Gini index should be preferred.

Sklearn supports “Gini” criteria for Gini Index and by default, it takes “gini” value.

The Gini Index is calculated by subtracting the sum of the squared probabilities of each class from one. It favors larger partitions.

Iterative Dichotomiser 3 (ID3): This algorithm uses Information Gain

C4.5: Information gain or gain ratio

Classification and Regression Tree(CART)

https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html

Random Forest Machine

Random forest classifier creates a set of decision trees from randomly selected subset of training set. It then aggregates the votes from different decision trees to decide the final class of the test object. The

hyperparameters n estimators, max features, min sample leaf are used for increasing the predictive power and model's speed.

Class : RandomForestClassifier

For our model we have set n estimator value to 100 i.e 100 decision trees are created, and their average value is taken. Similarly, criterion is set to ”gini”(calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly), min samples split is set to 2, max features is set to ”auto”, random state is set to 0(to produce the same result every time)

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

Every class first fit method for training and then predict method for testing