a classifier is an algorithm that maps the input data to a specific category. 5 classification algorithms applied to the dataset

Logistic Regression: Predictive Learning Model. To determine output in this classifier, we use a statistical method to analyse the dataset. These data sets can have one or more than one independent values. The output is calculated with data in which there could be two outputs. The aim of this classification algorithm is to find the relationship between the dichotomous category and predictor variables

KNN A cluster of labeled points are used to understand how the other points should be labelled. For labelling a new point it checks the already labelled points which could be closest to the point to be labelled, i.e closest to the neighbour. In this way depending on the votes of the neighbour the new point is labelled the same label which most neighbours have.

Decision Trees: This classification algorithm builds the regression models. These models are builded in form of structure which is similar to tree – a tree like structure is created by this classifier. It keeps on dividing the data set into subsets and smaller subsets which develops an associated tree, incrementally. The decision tree is finally created which has decision nodes and leaf nodes. In this tree the leaf node will have details about the classification or the decision taken for classification whereas the decision will have branches. The highest decision node which will be at the top of the tree will correspond to the root node. This will be the best predictor.

Random Forest: The regression and other tasks work by building a group of decision trees at training data level and during the output of the class, which could be the mode of classification or prediction regression for individual trees. This classifier accuracy for decision trees practice of overfitting the training data set

Naive Bayes Classifier: The classification here is based on Bayes Theorem, it assumes independent predictors. In simple words, this classifier will assume that the existence of specific features in a class is not related to the existence of any other feature. If there is dependency among the features of each other or on the presence of other features, all of these will be considered as an independent contribution to the probability of the output. This classification algorithm is very much useful to large datasets and is very easy to use

SVM: In this algorithm each point which is a data item is plotted in a dimensional space, this space is also known as n dimensional plane, where the n represents the number of features of the data. The classification is done based on the differentiation in the classes, these classes are data set points present in different planes.