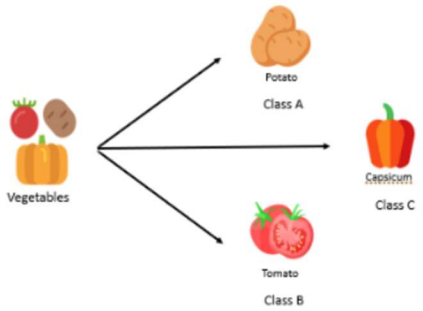
Classification

It is used to identify the category of new observations on basis of training data. A program learns from the given dataset or observation and then classifies new observation into a number of classes or groups such as yes or no, spam or not spam. Classes can be called as targets, labels or categories. **Binary classifier** if classification problem has only two possible outcomes. **Multi-class classifier** if classification problem has more than two outcomes. You can classify something based on its traits if you consider it a group of items, such as a collection of vegetables. For example, you could classify the potatoes, tomatoes, and peppers into A, B, and C categories.



Random Forest

The term "Random Forest" refers to a collection of decision trees. Each tree is assigned a class, and the tree "votes" for that class to categorize a new item based on its characteristics. The classification with the highest votes is selected by the forest (over all the trees in the forest). Researchers use Random Forest at work in various enterprises, including banking, stock trading, medicine, and internet commerce. It forecasts factors like customer behavior, patient demographics, and safety that help these firms run efficiently.

It is an ensemble learning method primarily used for classification and regression tasks. It operates by constructing multiple decision trees during training and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The main idea is to combine the predictions of several base estimators to improve generalization and robustness over a single estimator.

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.