# Supervised Machine Learning Algorithms

In machine learning, classification is a supervised learning approach in which the computer program learns from the data input given to it and then uses this learning to classify new observation.
This data set may simply be bi-class (like identifying whether the person is male or female or that the mail is spam or non-spam) or it may be multi-class too.
Some examples of classification problems are: speech recognition, handwriting recognition, bio metric identification, document classification etc.

Here we have the types of classification algorithms in Machine Learning:

- Logistic Regression
- Naive Bayes Classifier
- Stochastic Gradient
- K Nearest Neighbour
- Decision Trees
- Random Forest
- Support Vector Machines

**Structured Data Classification**

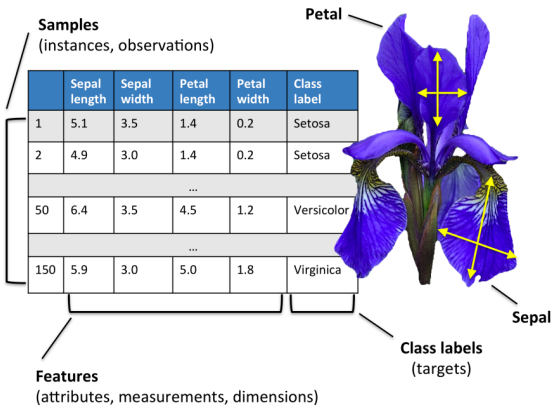
Classification can be performed on structured or unstructured data.
Classification is a technique where we categorise data into a given number of classes. The main goal of a classification problem is to identify the category/class to which a new data will fall under.

# Few of the terminologies encountered in machine learning – classification:

**Classifier:**
An algorithm that maps the input data to a specific category.

**Classification model:**
A classification model tries to draw some conclusion from the input values given for training.
It will predict the class labels/categories for the new data.

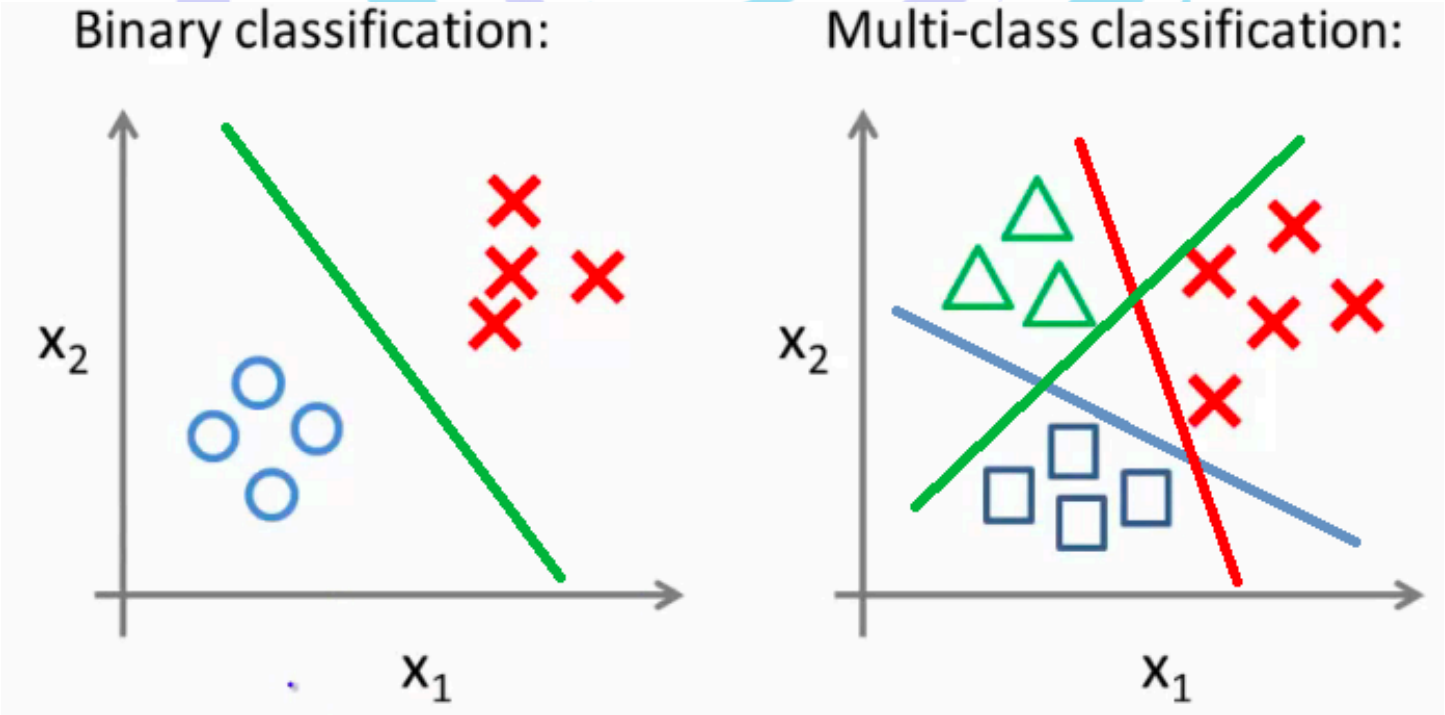
**Feature:**
A feature is an individual measurable property of a phenomenon being observed.




| | Sepal length | Sepal width | Petal length | Petal width | Class label |
|---|---|---|---|---|---|
| 1 | 5.1 | 3.5 | 1.4 | 0.2 | Setosa |
| 2 | 4.9 | 3.0 | 1.4 | 0.2 | Setosa |
| ... | | | | | |
| 50 | 6.4 | 3.5 | 4.5 | 1.2 | Versicolor |
| ... | | | | | |
| 150 | 5.9 | 3.0 | 5.0 | 1.8 | Virginica |

**Binary Classification:**
Classification task with two possible outcomes. Eg: Gender classification (Male / Female)

**Multi class classification:**
Classification with more than two classes. In multi class classification each sample is assigned to one and only one target label. Eg: An animal can be cat or dog but not both at the same time




**Multi label classification:**
Classification task where each sample is mapped to a set of target labels (more than one class). Eg: A news article can be about sports, a person, and location at the same time.

# The following are the steps involved in building a classification model

**Step 1:**
Initialise the classifier to be used.

**Step 2:**
Train the classifier:
All classifiers in scikit-learn uses a **fit**(X, y) method to fit the model(training) for the given train data X and train label y.

**Step 3:**
Predict the target:
Given an unlabeled observation X, the **predict**(X) returns the predicted label y.

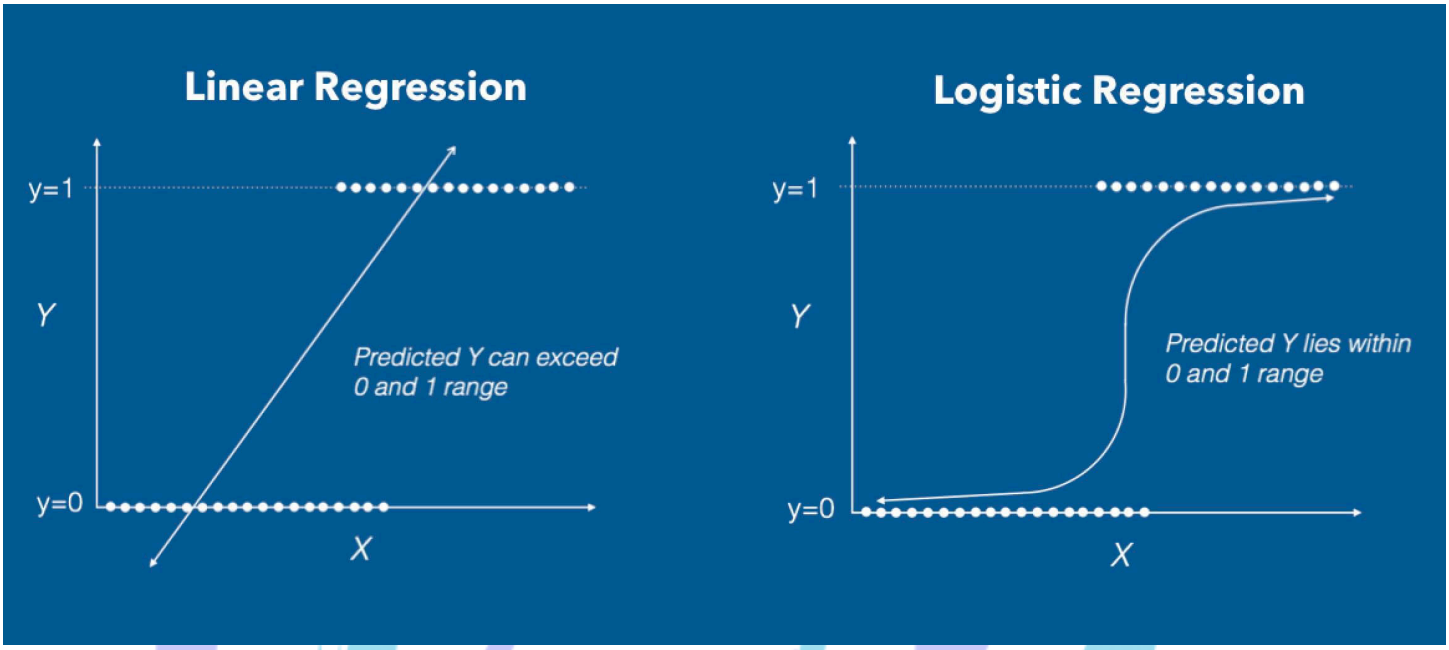
**Step 4:**
Evaluate the classifier model

Example :

```
from sklearn import tree

# Data Set
BalllsFeatures = [[35,1],[47,1],[90,0],[48,1],[90,0],[35,1],[92,0],[35,1],[35,1],[35,1],
[96,0],[43,1],[110,0],[35,1],[95,0]]

# Features
Names = [1,1,2,1,2,1,2,1,1,1,2,1,2,1,2]

# Step 1 Initialise the classifier
clf = tree.DecisionTreeClassifier()

# Step 2 Train the classifier
clf = clf.fit(BalllsFeatures,Names)

# Step 3 Predict the target
print(clf.predict([[44,1]]))
```

# Classification Algorithms

## Logistic Regression Algorithm (Predictive Learning Model)




**Definition**:

Logistic regression is a machine learning algorithm for classification. In this algorithm, the probabilities describing the possible outcomes of a single trial are modelled using a logistic function.
It is a statistical method for analysing a data set in which there are one or more independent variables that determine an outcome.
The outcome is measured with a dichotomous variable (in which there are only two possible outcomes).
The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables.

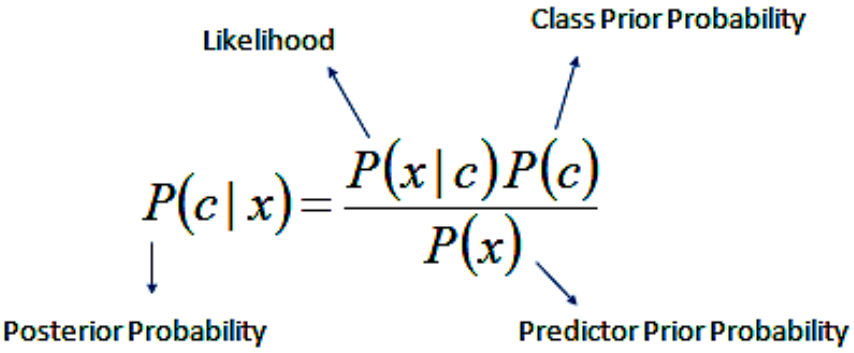
**Advantages**:

Logistic regression is designed for this purpose (classification), and is most useful for understanding the influence of several independent variables on a single outcome variable.

**Disadvantages**:

Works only when the predicted variable is binary, assumes all predictors are independent of each other, and assumes data is free of missing values.

# Naïve Bayes Algorithm
# (Generative Learning Model)




$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

**Definition:**

Naive Bayes algorithm based on Bayes' theorem with the assumption of independence between every pair of features.
Naive Bayes classifiers work well in many real-world situations such as document classification and spam filtering.
It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.
Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability.
Naive Bayes model is easy to build and particularly useful for very large data sets.
Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

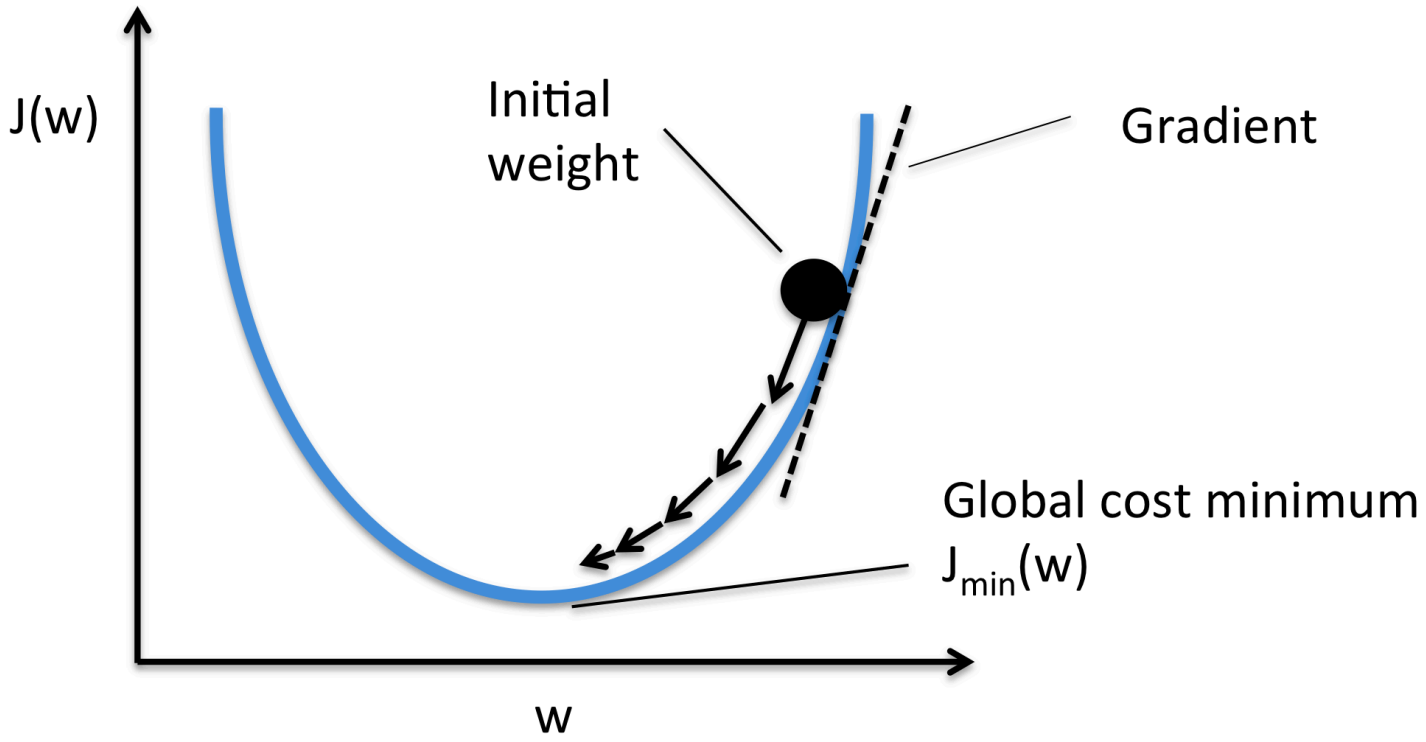
**Advantages**:

This algorithm requires a small amount of training data to estimate the necessary parameters. Naive Bayes classifiers are extremely fast compared to more sophisticated methods.

**Disadvantages:**

Naive Bayes is is known to be a bad estimator.

# Stochastic Gradient Descent Algorithm




**Definition:**

Stochastic gradient descent is a simple and very efficient approach to fit linear models.
It is particularly useful when the number of samples is very large.
It supports different loss functions and penalties for classification.

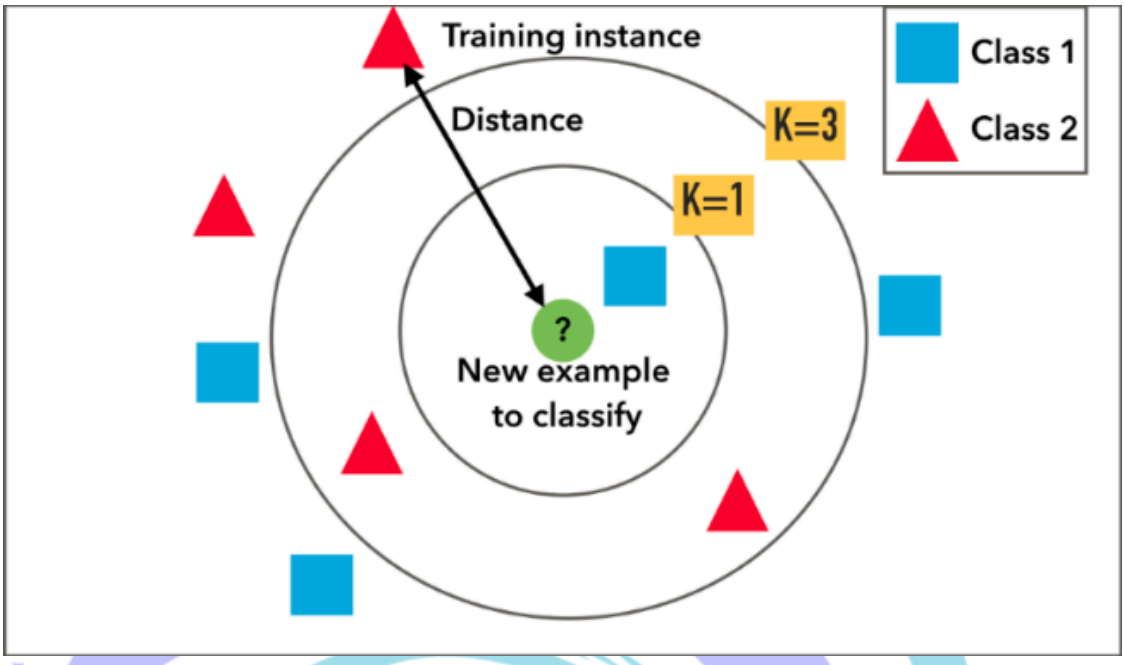
**Advantages:**

Efficiency and ease of implementation.

**Disadvantages:**

Requires a number of hyper-parameters and it is sensitive to feature scaling.

# K-Nearest Neighbours Algorithm




**Definition:**

Neighbours based classification is a type of lazy learning as it does not attempt to construct a general internal model, but simply stores instances of the training data. Classification is computed from a simple majority vote of the k nearest neighbours of each point.
The k-nearest-neighbors algorithm is a classification algorithm, and it is supervised: it takes a bunch of labelled points and uses them to learn how to label other points.
To label a new point, it looks at the labelled points closest to that new point (those are its nearest neighbors), and has those neighbors vote, so whichever label the most of the neighbors have is the label for the new point (the "k" is the number of neighbors it checks).

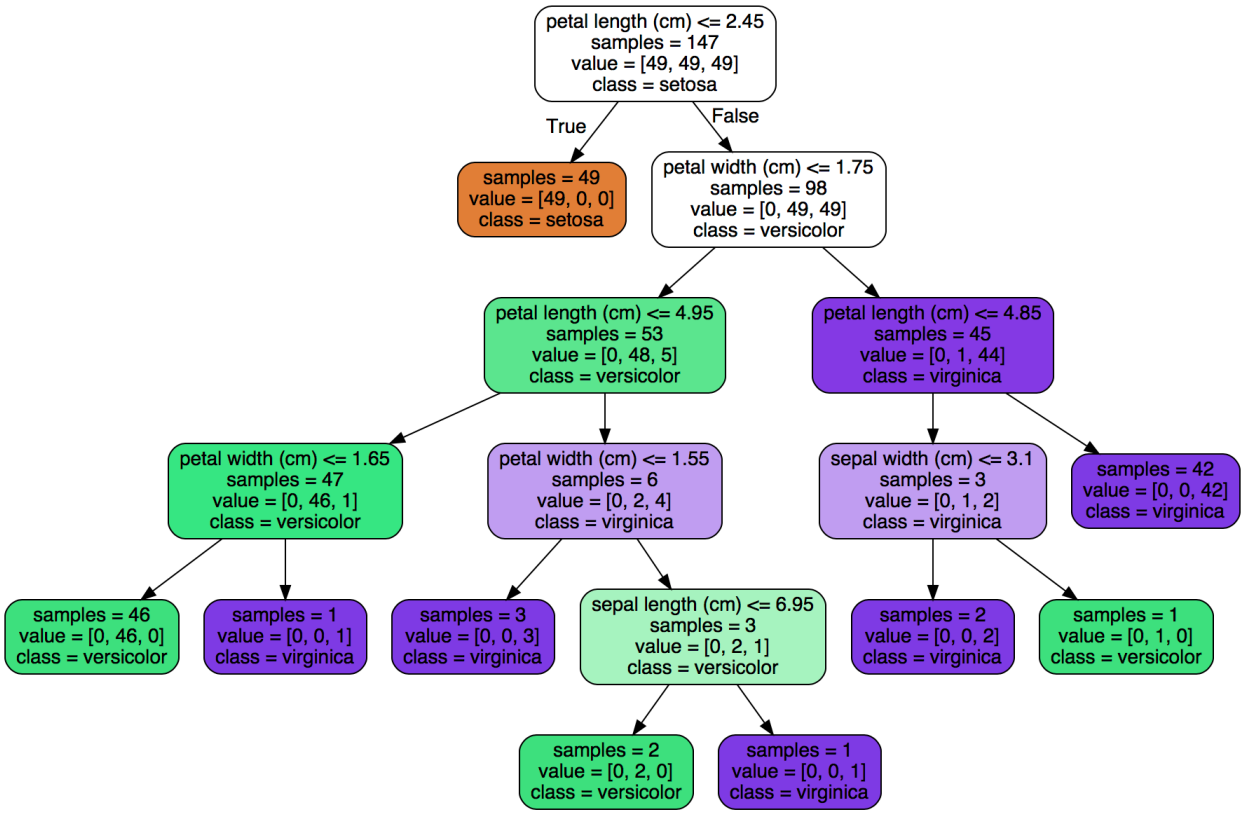
**Advantages:**

This algorithm is simple to implement, robust to noisy training data, and effective if training data is large.

**Disadvantages:**

Need to determine the value of K and the computation cost is high as it needs to computer the distance of each instance to all the training samples.

# Decision Tree Algorithm




**Definition:**

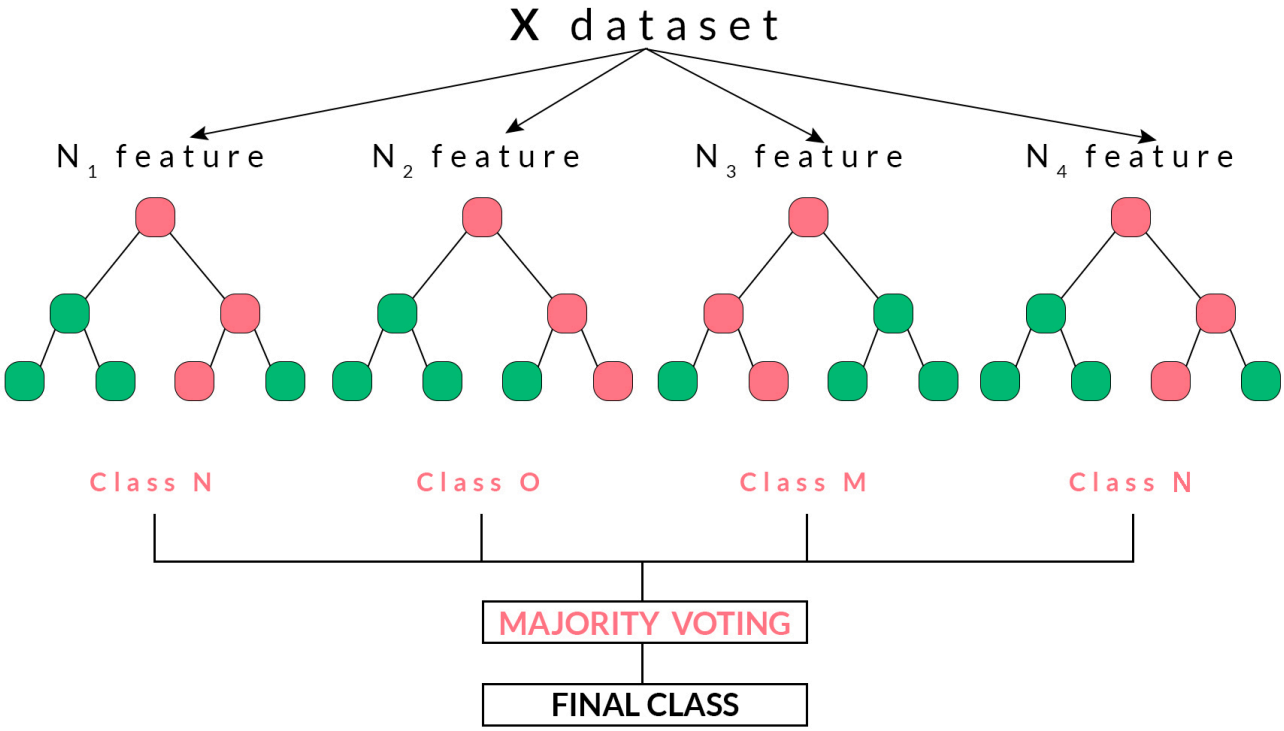
Given a data of attributes together with its classes, a decision tree produces a sequence of rules that can be used to classify the data.
Decision tree builds classification or regression models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed.
The final result is a tree with decision nodes and leaf nodes.
A decision node has two or more branches and a leaf node represents a classification or decision.
The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

**Advantages:**

Decision Tree is simple to understand and visualise, requires little data preparation, and can handle both numerical and categorical data.

**Disadvantages:**
Decision tree can create complex trees that do not generalise well, and decision trees can be unstable because small variations in the data might result in a completely different tree being generated.

# Random Forest Algorithm




**Definition:**

Random forest classifier is a meta-estimator that fits a number of decision trees on various sub-samples of datasets and uses average to improve the predictive accuracy of the model and controls over-fitting.
The sub-sample size is always the same as the original input sample size but the samples are drawn with replacement.
Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.
Random decision forests correct for decision trees' habit of over fitting to their training set.

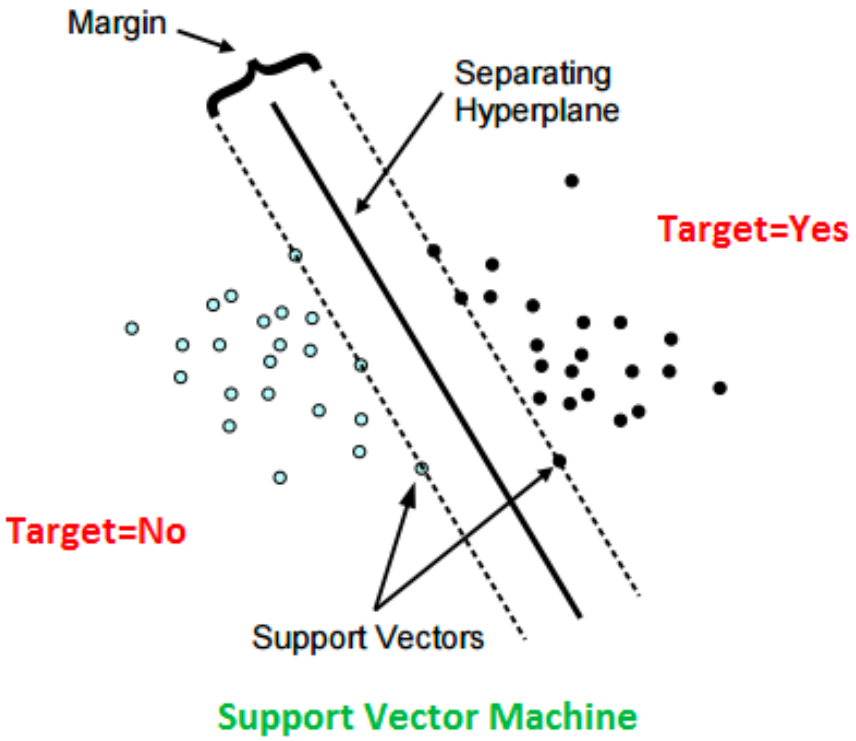
**Advantages:**

Reduction in over-fitting and random forest classifier is more accurate than decision trees in most cases.

**Disadvantages:**

Slow real time prediction, difficult to implement, and complex algorithm.

# Support Vector Machine Algorithm




**Definition**:

Support vector machine is a representation of the training data as points in space separated into categories by a clear gap that is as wide as possible.
New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

**Advantages**:

Effective in high dimensional spaces and uses a subset of training points in the decision function so it is also memory efficient.

**Disadvantages**:

The algorithm does not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation.

**We can compare each algorithm by considering its Accuracy and F1-Score**

**Accuracy:**

(True Positive + True Negative) / Total Population

• Accuracy is a ratio of correctly predicted observation to the total observations. Accuracy is the most intuitive performance measure.
• True Positive: The number of correct predictions that the occurrence is positive
• True Negative: The number of correct predictions that the occurrence is negative

**F1-Score:**

(2 x Precision x Recall) / (Precision + Recall)

• F1-Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. F1-Score is usually more useful than accuracy, especially if you have an uneven class distribution.
• Precision: When a positive value is predicted, how often is the prediction correct?
• Recall: When the actual value is positive, how often is the prediction correct?