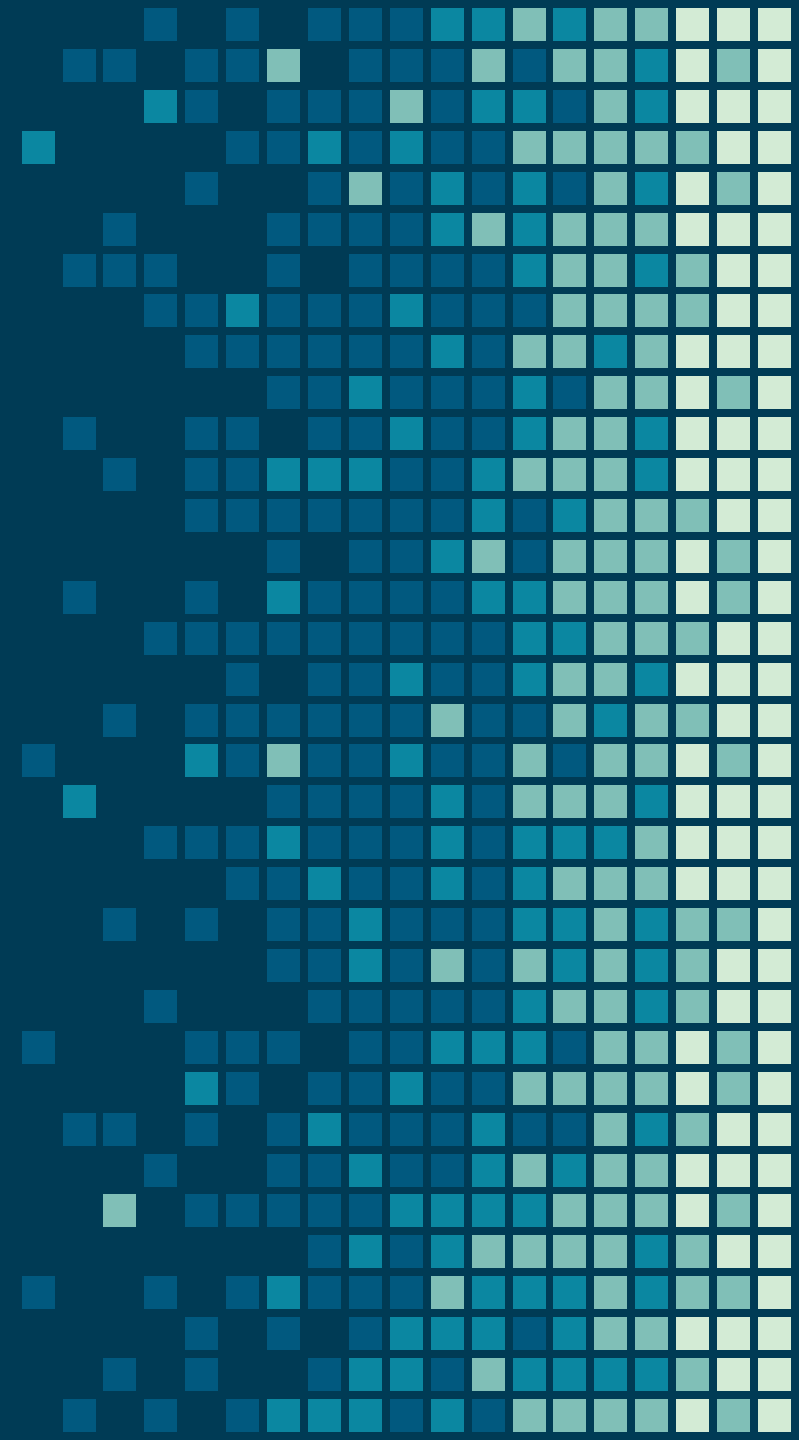


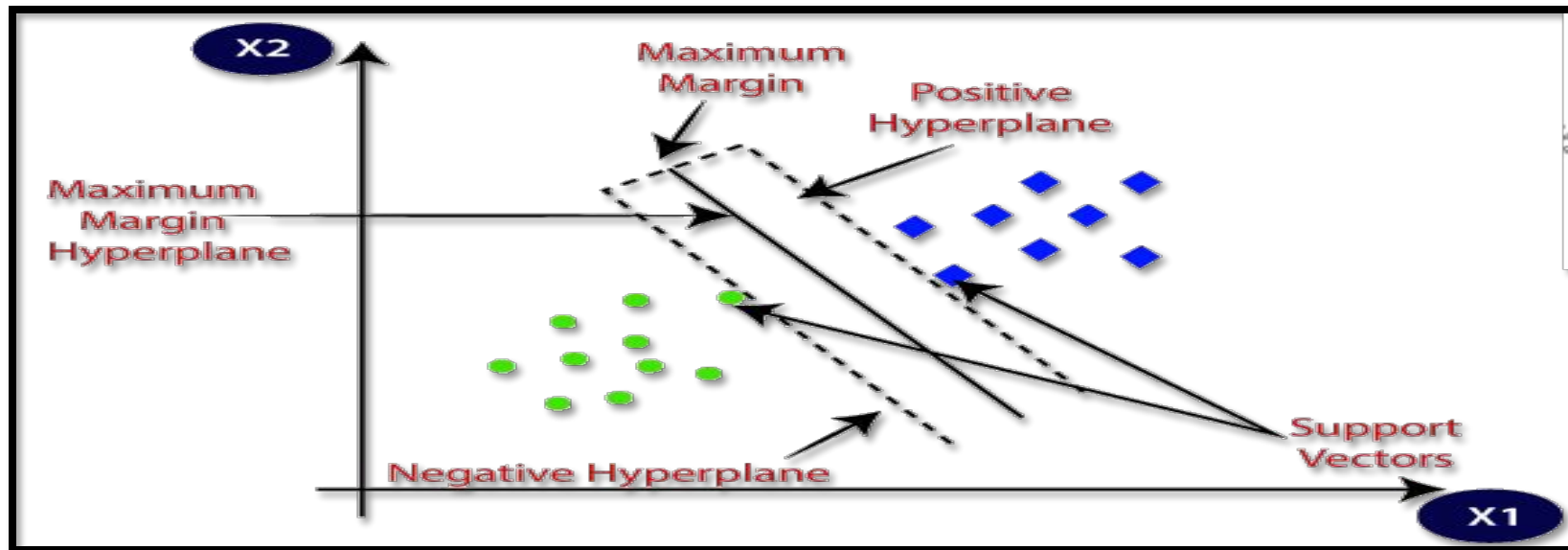
Data Science

UNIT III



Support Vector Machine

- Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms.
- Which is used for Classification as well as Regression problems.



- The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future.
- This best decision boundary is called a hyperplane.
- SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

Example

- SVM can be understood with the example that we have used in the KNN classifier. Suppose we see a strange cat that also has some features of dogs, so if we want a model that can accurately identify whether it is a cat or dog, so such a model can be created by using the SVM algorithm. We will first train our model with lots of images of cats and dogs so that it can learn about different features of cats and dogs, and then we test it with this strange creature. So as support vector creates a decision boundary between these two data (cat and dog) and choose extreme cases (support vectors), it will see the extreme case of cat and dog. On the basis of the support vectors, it will classify it as a cat.



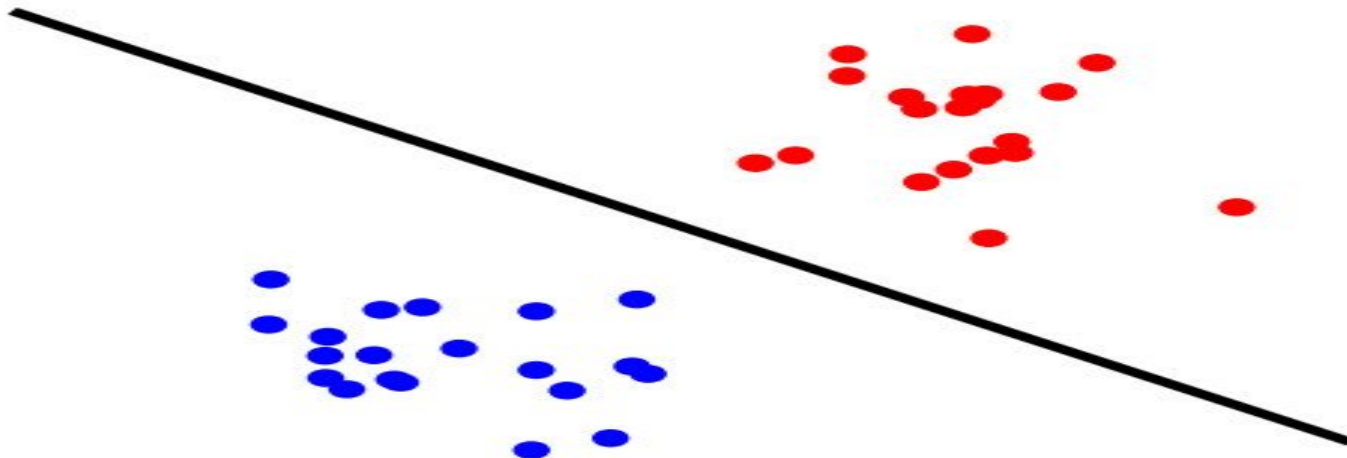
SVM algorithm can be used

- Face detection
- image classification
- text categorization

Types of SVM

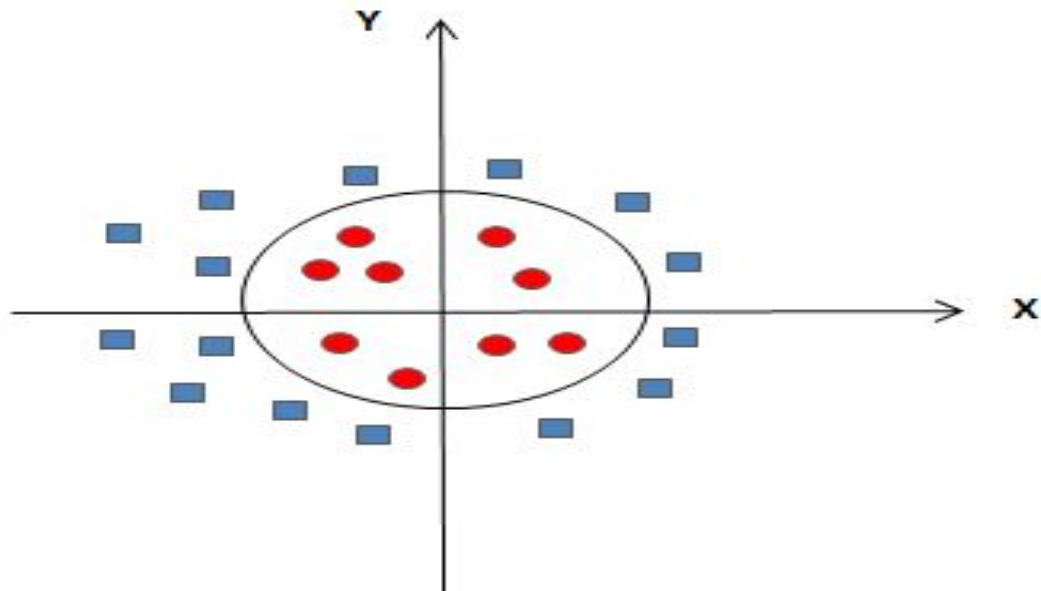
- Linear SVM:

Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.



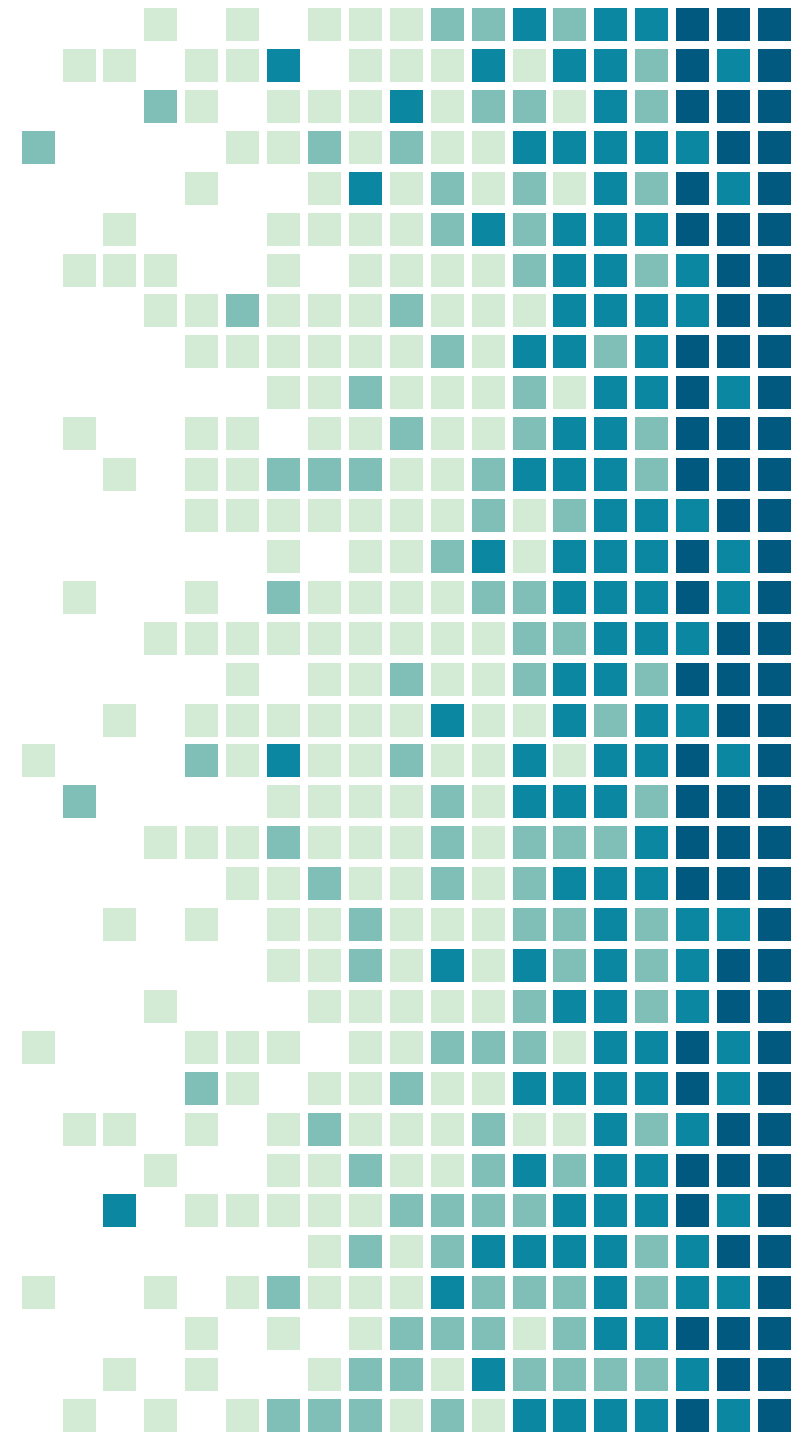
Non-linear SVM:

- Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.



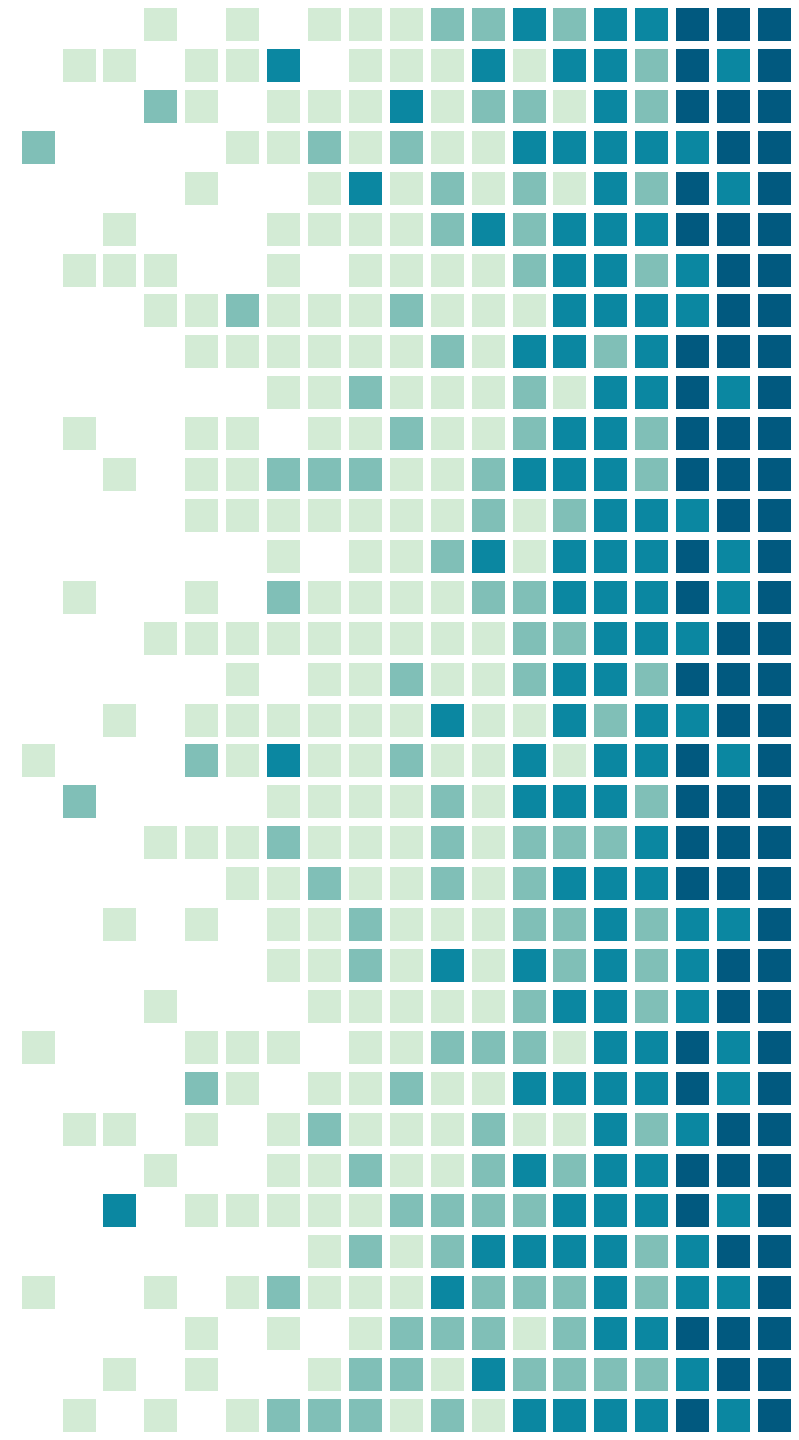
k-means clustering

- k-means clustering is one of the simplest algorithms which uses unsupervised learning method to solve known clustering issues.
- It groups the unlabeled dataset into clusters.
- k-means clustering require following two inputs.
 1. k = number of clusters
 2. Training set(m) = $\{x_1, x_2, x_3, \dots, x_m\}$



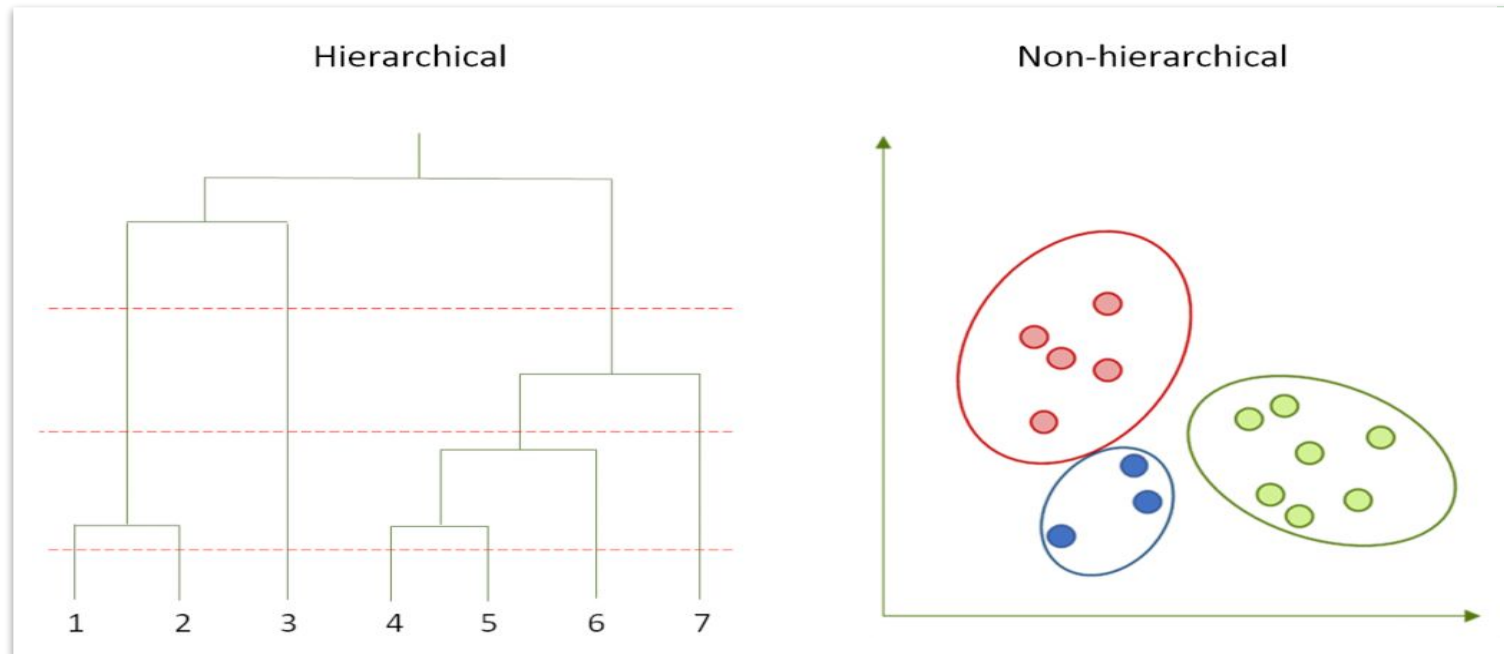
K-means clustering Application

- Market segmentation
- Document Clustering
- Image segmentation
- Image compression
- Customer segmentation



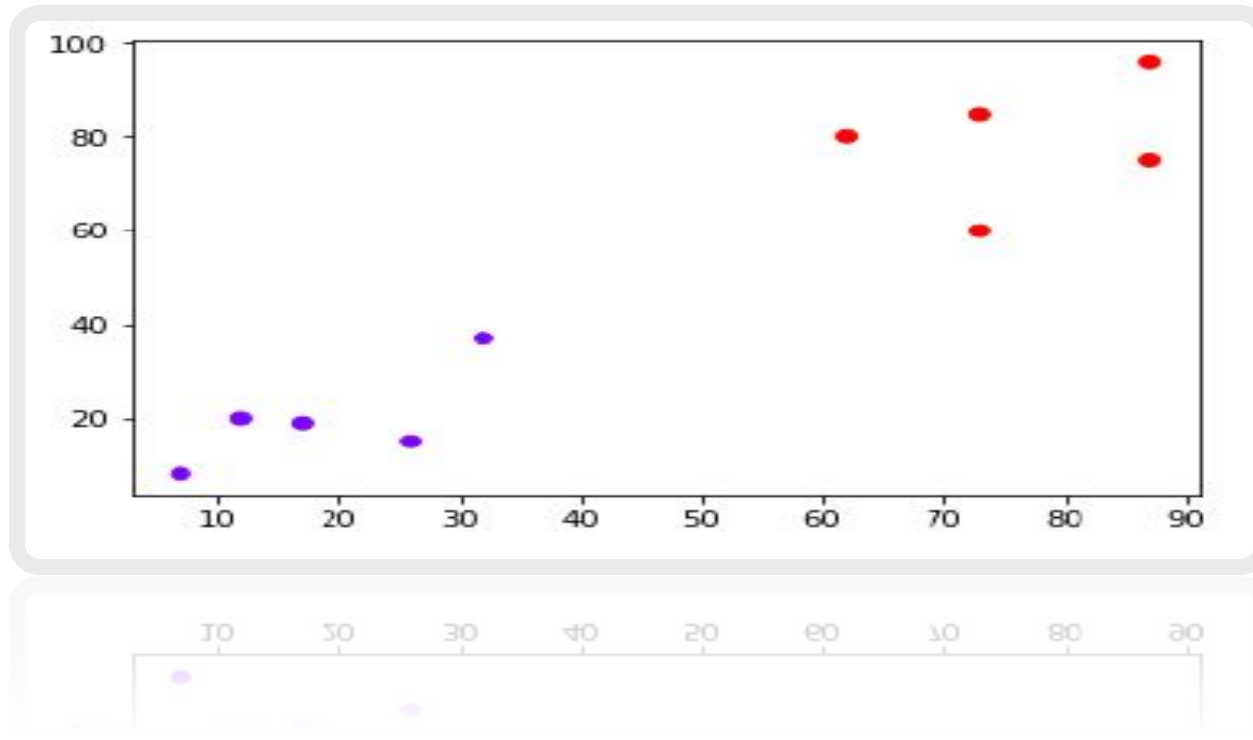
Hierarchical clustering

- In hierarchical clustering, the clusters are not formed in a single step rather it follows series of partitions to come up with final clusters. It looks like a tree as visible in the image. this tree-shaped structure is known as the **dendrogram**.



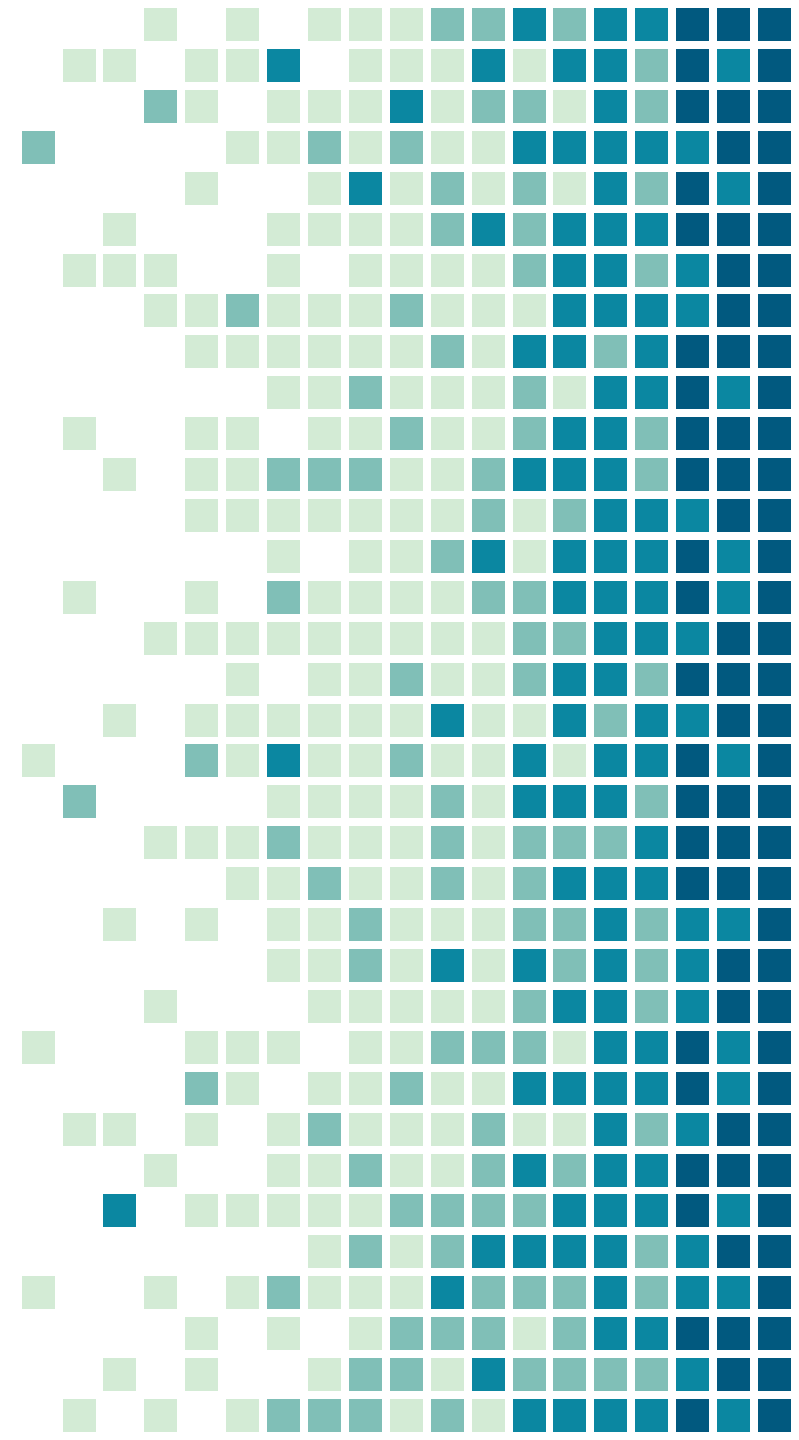
K-Nearest Neighbour

- K-nearest neighbors (KNN) algorithm is a type of supervised ML algorithm which can be used for classification.

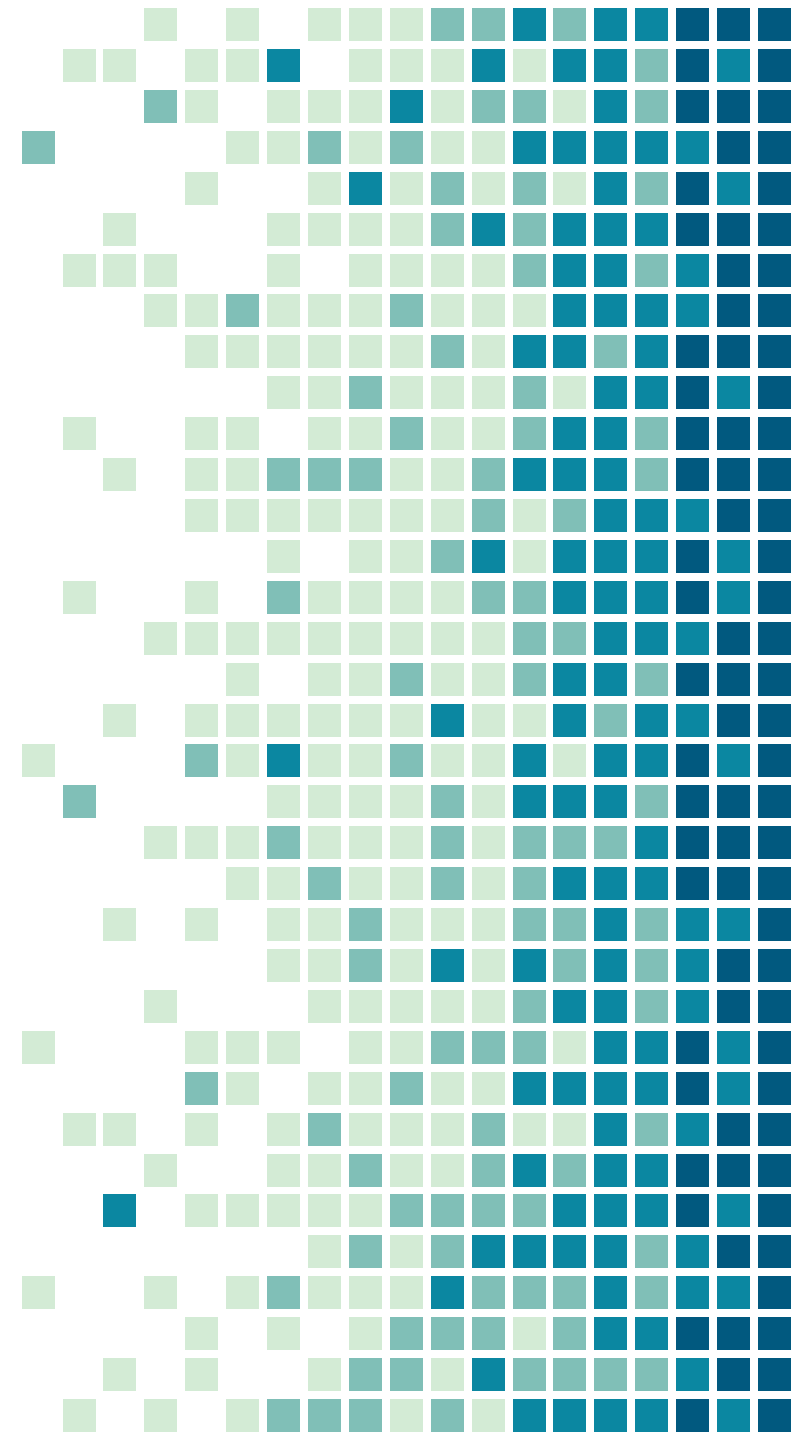
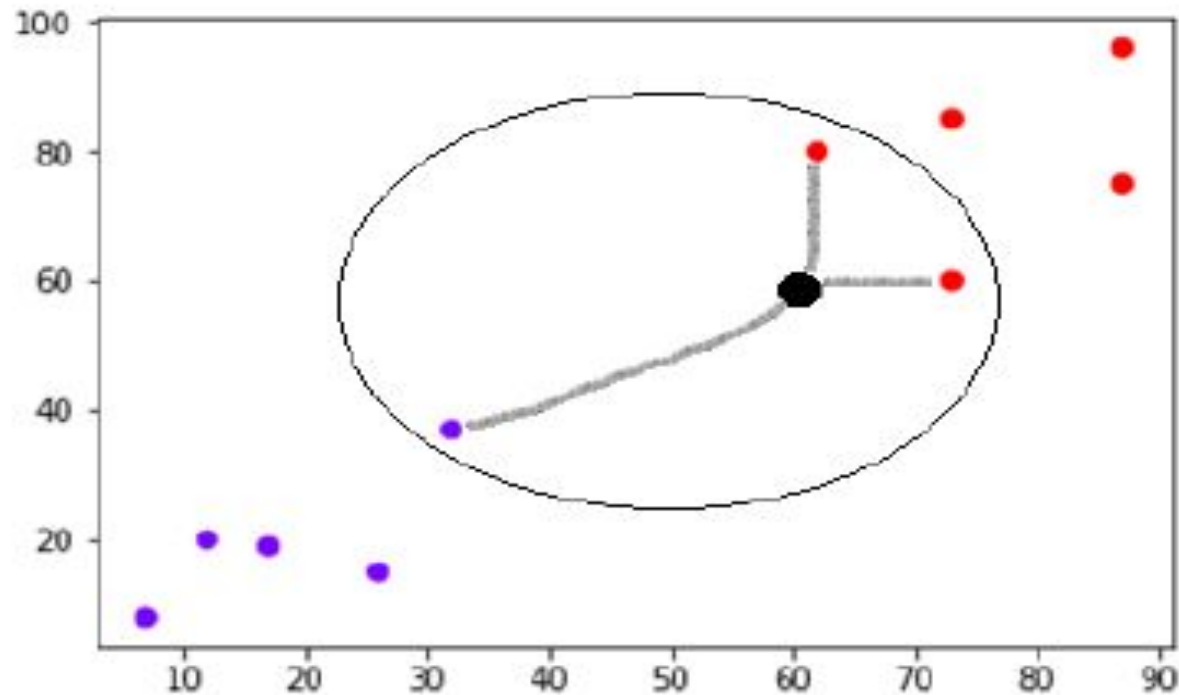


Working of KNN Algorithm

- K-nearest neighbors (KNN) algorithm uses ‘feature similarity’ to predict the values of new datapoints which further means that the new data point will be assigned a value based on how closely it matches the points in the training set.



- We can see in the above diagram the three nearest neighbors of the data point with black dot. Among those three, two of them lies in Red class hence the black dot will also be assigned in red class.



Applications of KNN

- Banking System

KNN can be used in banking system to predict whether an individual is fit for loan approval? Does that individual have the characteristics similar to the defaulters one?

- Calculating Credit Ratings

KNN algorithms can be used to find an individual's credit rating by comparing with the persons having similar traits.

