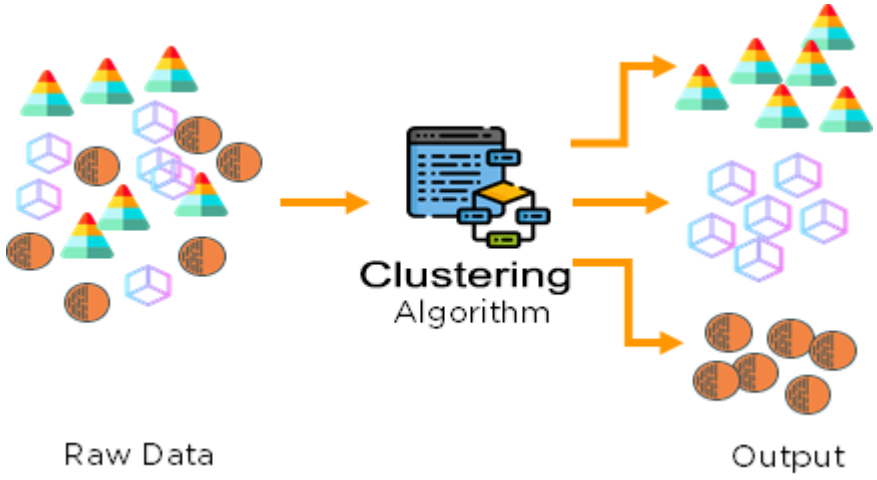
Clustering

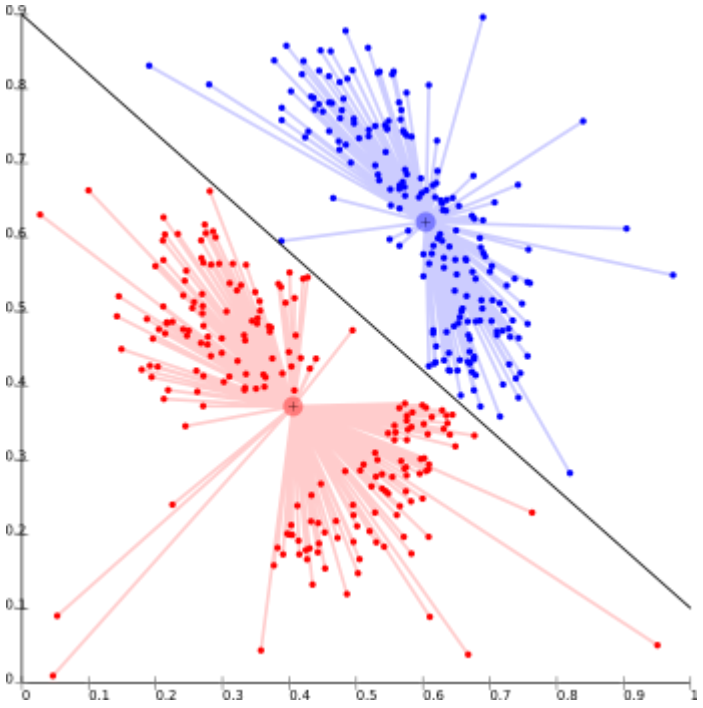
- It is basically a type of unsupervised learning method . An unsupervised learning method is a method in which we draw references from datasets consisting of input data without labelled responses.
- Generally, it is used as a process to find meaningful structure, explanatory underlying processes, generative features, and groupings inherent in a set of examples.
- Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups.



Types Of Clustering :

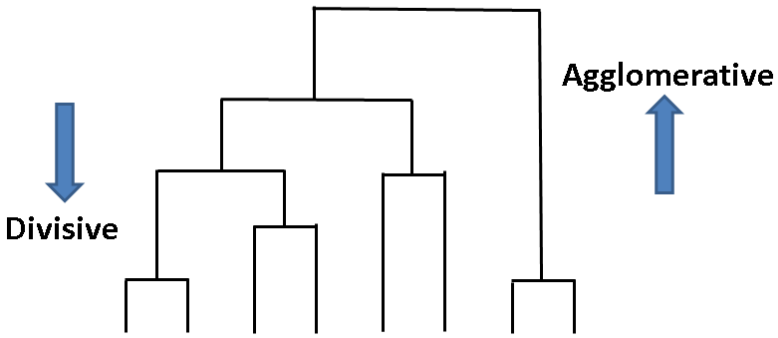
1. Centroid-based clustering:

- This is basically one of iterative clustering algorithm in which the clusters are formed by the closeness of data points to the centroid of clusters.
- Here , the cluster center i.e. centroid is formed such that the distance of data points is minimum with the center.
- Data is grouped into centroids based on how close they are to the center of the centroids.
- For Ex- K – means algorithm is one of popular example of this algorithm.



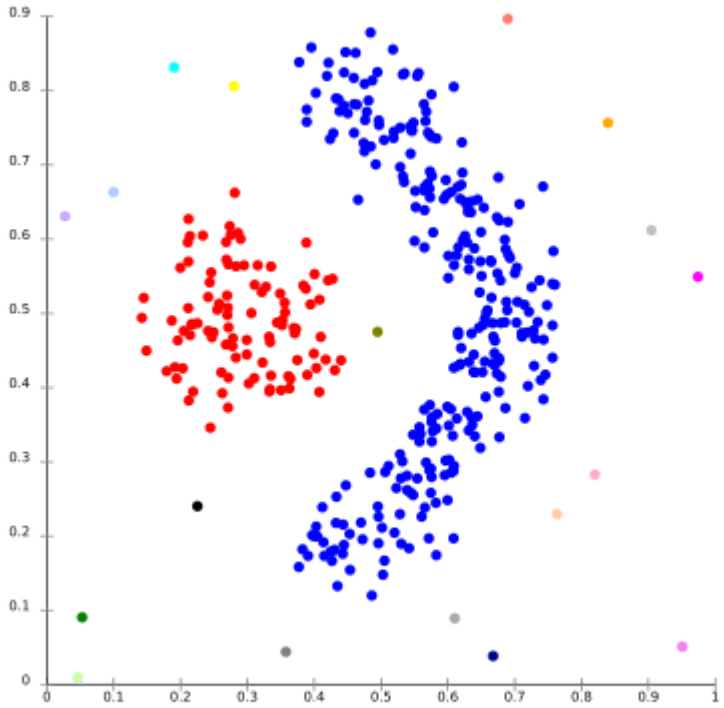
2. Hierarchical clustering:

- When you want machine to find the right number of clusters. Each data item is considered as a cluster and then data items are grouped together based on their distance recursively until optimum clusters of data are calculated.
- Algorithms include Agglomerative clustering/ hierarchical clustering.



3. Density Models:

- In this clustering model there will be a searching of data space for areas of varied density of data points in the data space .
- For Ex- DBSCAN and OPTICS.



- There are also other clustering algorithm types such as distribution based clustering which uses underlying probability distribution to group data into clusters. refer this: https://www.geeksforgeeks.org/different-types-clustering-algorithm/?ref=rp (https://www.geeksforgeeks.org/different-types-clustering-algorithm/?ref=rp)

Applications of Clustering in different fields :

1. Identifying Fake News
2. Spam filter
3. Marketing and Sales
4. Classifying network traffic
5. Identifying fraudulent or criminal activity
6. Earthquake studies etc