**Clustering**

**https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68**

The clustering algorithm is an unsupervised learning algorithm. There is no label for model training. **We look for patterns present in the dataset to group them into different clusters.**

Solving a clustering problem is different from solving a supervised learning problem. The steps involved are,

**1. Data exploration**

Like any supervised learning problem, we start with data exploration. The difference is we focus more on the **visual analysis.** The goal here is to understand if there are any clear clusters.

**2. Choosing the right algorithm**

The next step is to identify the right set of algorithms. The choice of the algorithm would depend on the problem itself. In most cases, we would create clusters where one item belongs to only one cluster. Sometimes we need to allow items to belong to more than one cluster. In the first case, we use algorithms such as K-Mean and Hierarchical Clustering. For the latter, we use Fuzzy Clustering. Customer profiling is a typical example of fuzzy clustering. It would be normal for customers to exhibit many characteristics. Here it is fair to allow them to be part of multiple clusters with a probability.

**3. The ideal number of clusters**

It is critical to identify the ideal number of clusters. There are techniques and methods to identify the best number of clusters. The technique depends on the algorithm.

**4. Characteristics of the clusters**

A clustering problem is not considered solved after identifying the clusters. It is not enough to show the business the different clusters. **It requires further analysis to prove the clusters are correct. Visual analysis plays a key role here. It is easy to pick insights from visual data.**

<https://www.kaggle.com/kushal1996/customer-segmentation-k-means-analysis/notebook>

https://www.kaggle.com/ibtesama/getting-started-with-a-movie-recommendation-system

https://www.youtube.com/playlist?list=PLeo1K3hjS3uvCeTYTeyfe0-rN5r8zn9rw