This module deals with recommendation of movies and health issues. To illustrate this, we consider the company “Netflix”. We shall study how the Netflix recommend movies for a user. Netflix is an online movies streaming service based on subscription. Netflix database consists of movies with user rankings. **Collaborative Filtering** is a technique which uses other user preferences to predict our recommendations. We can also use movie information to predict recommendations to a user based on his/her ranking to the movies. For example, if a user gives a good rank to thriller movies, then we could recommend movies of thriller genre to that user. This technique is called as **Content Filtering.**

Netflix uses **Hybrid recommendation system** which is a combination of both collaborative and content filtering. Content filtering uses a method called **clustering.** This method uses unsupervised learning method – a method to group the given cluster based on a condition and building the predictive models for producing the outcome rather than predicting outcome directly. There are different algorithms for clustering like Hierarchical, K clustering and many more.

The first step in clustering is to define distance between two data points. For this we use Euclidean distance formula - √ (x2−x1)2+ (y2−y1)2+ (z2−z1)2.

Hierarchical clustering starts with each data point in its own cluster. Hierarchical clustering is of two types – **Bottom up approach and Top down approach**. Top down approach is also called as **k-mean cluster** approach. In Bottom up approach, every variable or data point is considered as an individual cluster. All these individual clusters/data points are grouped into a single cluster and then centroid of the cluster is obtained. The distance between this centroid and other clusters/data points are obtained. The data point with minimum distance to cluster is again grouped into another cluster. Then again the centroid is calculated and the process goes on till all the clusters/data points are fused. Dendrogram is a graphical representation of Hierarchical cluster.

Next, prediction of occurrence heart attack. Based on diagnostic data of a person, we study how to predict if there is any occurrence of heart attack. Random forest algorithm is used for this task. Two other algorithms that can be used for analysis are – spectral algorithm, k – mean algorithm.

Clustering methods are also used to group digital images. They partition image to clusters based on differences in pixel, colours, intensity or texture. Edge detection, region – growing methods are used to cluster digital images. These methods are applied in face recognition, finger print recognition and many more.