Cluster analysis is a technique used in machine learning that attempts to find clusters of observations within a dataset. This is unsupervised Learning.

The goal of cluster analysis is to find clusters such that the observations within each cluster are quite similar to each other, while observations in different clusters are quite different from each other.

An Example.

Below the Model/Algorithm Clusters Fruits into Similar Groups.
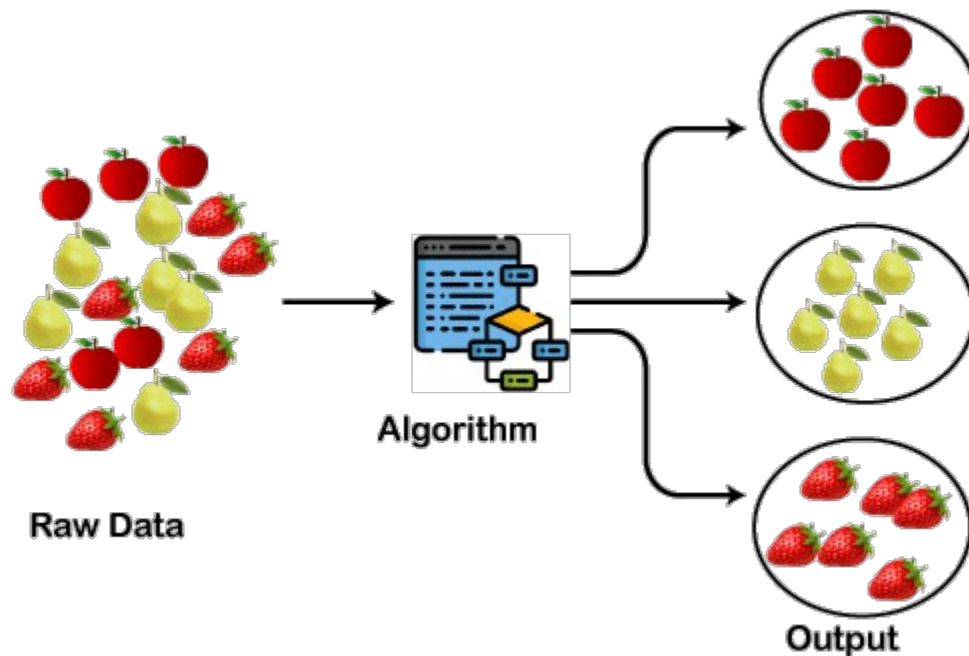
The following examples show how cluster analysis is used in various real-life situations.

Example 1: Retail Marketing

Retail companies often use clustering to identify groups of households that are similar to each other.

For example, a retail company may collect the following information on households:

1. Household income
2. Household size
3. Head of household Occupation
4. Distance from nearest urban area

They can then feed these variables into a clustering algorithm to perhaps identify the following clusters:

1. Cluster 1: Small family, high spenders
2. Cluster 2: Larger family, high spenders
3. Cluster 3: Small family, low spenders
4. Cluster 4: Large family, low spenders

The company can then send personalized advertisements or sales letters to each household based on how likely they are to respond to specific types of advertisements.

Example 2: Streaming Services

Streaming services often use clustering analysis to identify viewers who have similar behavior.

For example, a streaming service may collect the following data about individuals:

1. Minutes watched per day
2. Total viewing sessions per week
3. Number of unique shows viewed per month

Using these metrics, a streaming service can perform cluster analysis to identify high usage and low usage users so that they can know who they should spend most of their advertising dollars on.

Example 3: Sports Science

Data scientists for sports teams often use clustering to identify players that are similar to each other.

For example, professional basketball teams may collect the following information about players:

1. Points per game
2. Rebounds per game
3. Assists per game
4. Steals per game

They can then feed these variables into a clustering algorithm to identify players that are similar to each other so that they can have these players practice with each other and perform specific drills based on their strengths and weaknesses.

Example 4: Email Marketing

Many businesses use cluster analysis to identify consumers who are similar to each other so they can tailor their emails sent to consumers in such a way that maximizes their revenue.

For example, a business may collect the following information about consumers:

1. Percentage of emails opened
2. Number of clicks per email
3. Time spent viewing email

Using these metrics, a business can perform cluster analysis to identify consumers who use email in similar ways and tailor the types of emails and frequency of emails they send to different clusters of customers.

Example 5: Health Insurance

Actuaries at health insurance companies often used cluster analysis to identify "clusters" of consumers that use their health insurance in specific ways.

For example, a Data Scientist may collect the following information about households:

1. Total number of doctor visits per year
2. Total household size
3. Total number of chronic conditions per household
4. Average age of household members

A Data Scientist can then feed these variables into a clustering algorithm to identify households that are similar. The health insurance company can then set monthly premiums based on how often they expect households in specific clusters to use their insurance.

References

https://en.wikipedia.org/wiki/Cluster_analysis

https://developers.google.com/machine-learning/clustering/overview