Unit-4:

Data Science Applications:

★ Prediction and Elections:

- "Prediction" refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when forcasting the likelihood of a particular outcome, such as whether or not a Customer will churn in 30 days. The algorithm will generate probable values for an unknown variable for each record in the new data, allowing the Model builder to identify what that value will most likely be.
- "Prediction" — In some Case prediction means you are Predicting the future outcome, such as when you are using Machine learning to determine the next best action in a marketing Campaign.

And

In other Case it Can be like an educated guess, Examples to check whether or not a transaction that already occured was fradulent. In this Case the transaction has already happened, but you are making an educated guess about whether or not it was legal, allowing you to take the

appropriate action.

→ Why it is Important :

- It allows business to make highly accurate guesses as to likely outcomes of a question based on historical data, which can be about all kinds of things - Customer churn likelihood, possible fraudulent activity, and more.
- They provide the business with insights that result in precise business value.

→ Elections :

- Specific Algorithms are ~~used~~ designed to analyze collected data on voters to predict.
  - election outcomes.
  - advertising tactics to promote voter support and participation (online, door-to-door, phone, mail).
- Taking Raw Data and making it sensible.

A Recommendations and business analytics :

- When you visit an application or a website for the very first time, it will probably show you its most popular content among its various users, but once you keep visiting it for a while, it starts recommending the items you should read, buy, watch, listen, or spend your time for.

- this means that recommedation systems are based on three important factors :
  - (1.) Users.
  - (2.) Content.
  - (3.) Ratings.
- The Combination of users, Content and Ratings creates two different approaches for creating recommendation systems, which are :
  - (1) Content based.
  - (2) Collaborative filtering.

(1) Conten based.

- it is based on user data or Content (Content means music, videos, products to buy etc).
- User data or Content is used to target a new user that falls under the same Category of users.

(2) Collaborative filtering :

- it is very sophisticated as Compared to Content-based recommedation systems.
- They are based on the ratings or comments given by the users and their purpose is to predict the ratings for each Content and each user.

- It is more accurate than Content-based recommendation system, because it works on large databases with more Computational power.

✱ Business Analytics :

- it is the statistical study of business data to gain insights.
- uses mostly structured data.
- Does not involve much coding: it is more statistics oriented.
- whole analysis is based on statistical Concepts.
- it studies trends and patterns specific to business.
- Top Industries where business analytics is used : finance, healthCare, marketing, retail, supply chain, telecommunications.

→ Why Business Analytics :

- it bridges the gap b/w information technology and business by using analytics to provide data-driven recommendations.
- the business part requires deep business understanding, while the analytics part requires an understanding of data, statistics and Computer Science.

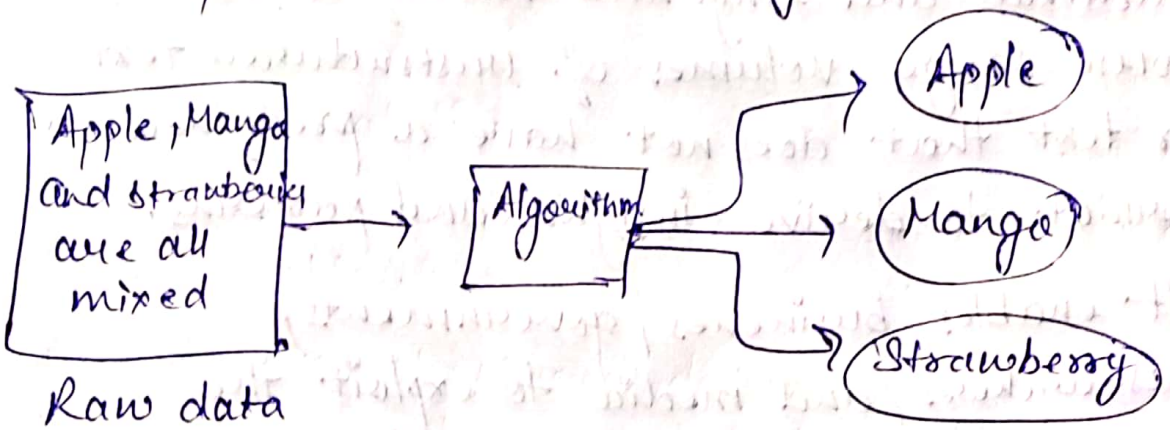
☆ Clustering and text analytics :-

- Data scientists and others use clustering to gain important insights from data by observing what groups (or clusters) the data points fall into when they apply a clustering algorithm to the data.
- It is used to identify groups of similar objects in datasets with two or more variable quantities.
  this data may be collected from marketing, biomedical, or geospatial databases.
- there are many clustering algorithms, simply because there are many notions of what a cluster should be or how it should be defined.
- It is unsupervised Machine learning.
- After applying clustering techniques, each cluster or group is provided with a cluster-ID.

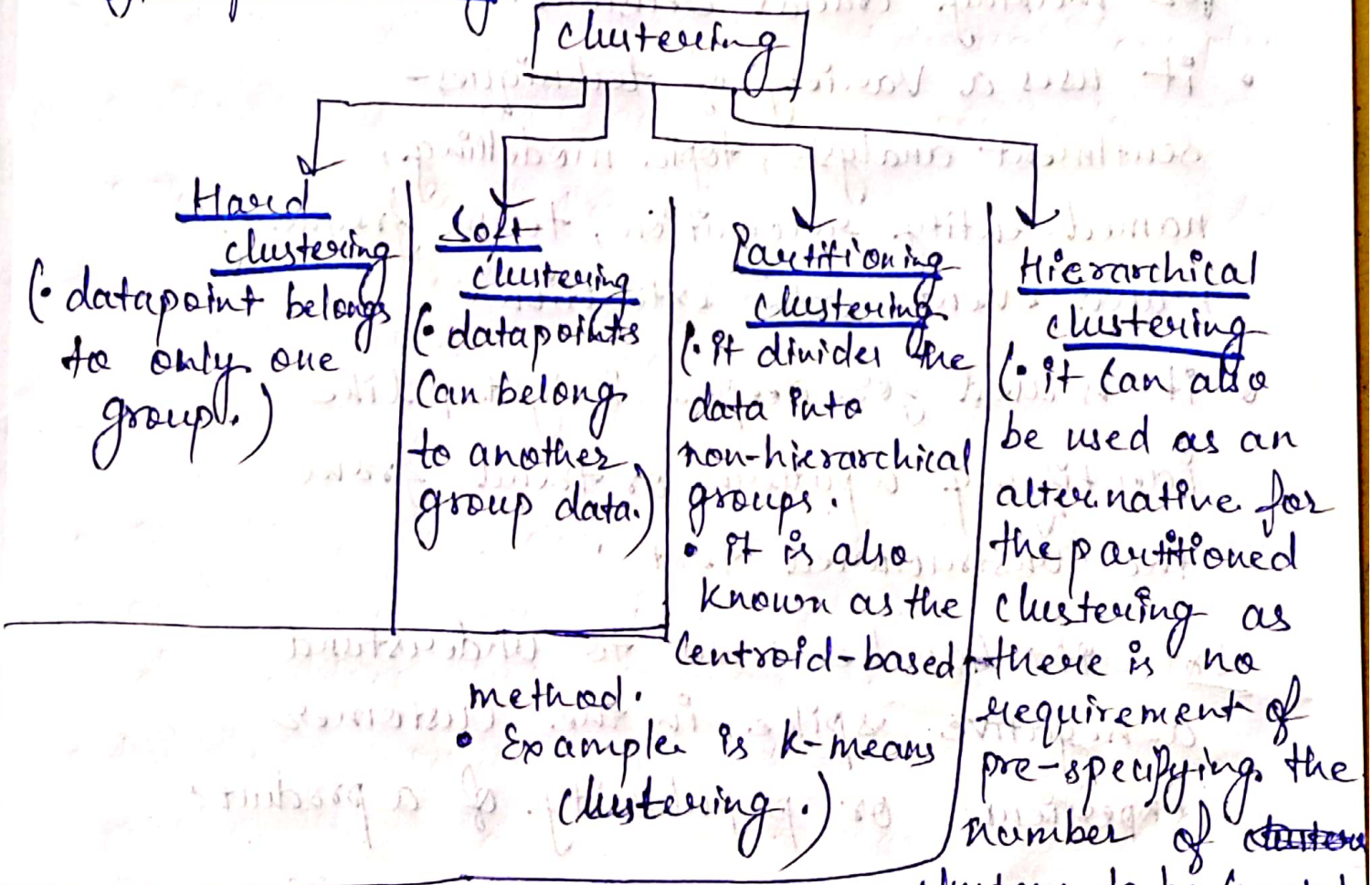
Ex:- a In any shopping mall - t-shirts are grouped in one section, and trousers are at other sections.

- Amazon also uses clustering in the recommendation system.

- Netflix also uses this technique to recommend the movies and web-services to its users, as per the watch history.



→ Types of clustering :-



→ Text Analytics :

- it combines a set of machine learning, statistical and linguistic techniques to process large volumes of unstructured text or text that does not have a predefined format, to derive insights and patterns.
  - it enables businesses, governments, researchers, and media to exploit the enormous content at their disposal for making crucial decisions.
  - it uses a variety of techniques - Sentiment analysis, topic modelling, named entity recognition, term frequency, and event ~~extar~~ extraction.
  - it is used for deeper insights, like identifying a pattern or trend from the unstructured text.

Ex: It can be used to understand a negative spike in the customer experience or popularity of a product.