Unit-V
Unsupervised Machine Learning Models

Computer Department
AVPTI Rajkot

Unit Outcomes

- Define Unsupervised Learning
- List types of Unsupervised Learning
- Advantage and disadvantage of Unsupervised machine learning
- Differentiate Supervised and Unsupervised Learning

Topics

- Introduction of Unsupervised Learning
- Types of Unsupervised Learning :Clustering and Association
- Advantage and disadvantage of Unsupervised machine learning
- Differentiate Supervised and Unsupervised Learning

Unsupervised Learning

- Unsupervised learning is a machine learning technique in which models are not supervised using training dataset.
- Instead, models itself find the hidden patterns and insights from the given data.
- It can be compared to learning which takes place in the human brain while learning new things.
- Unsupervised learning is a type of machine learning in which models are trained using unlabeled dataset and are allowed to act on that data without any supervision.

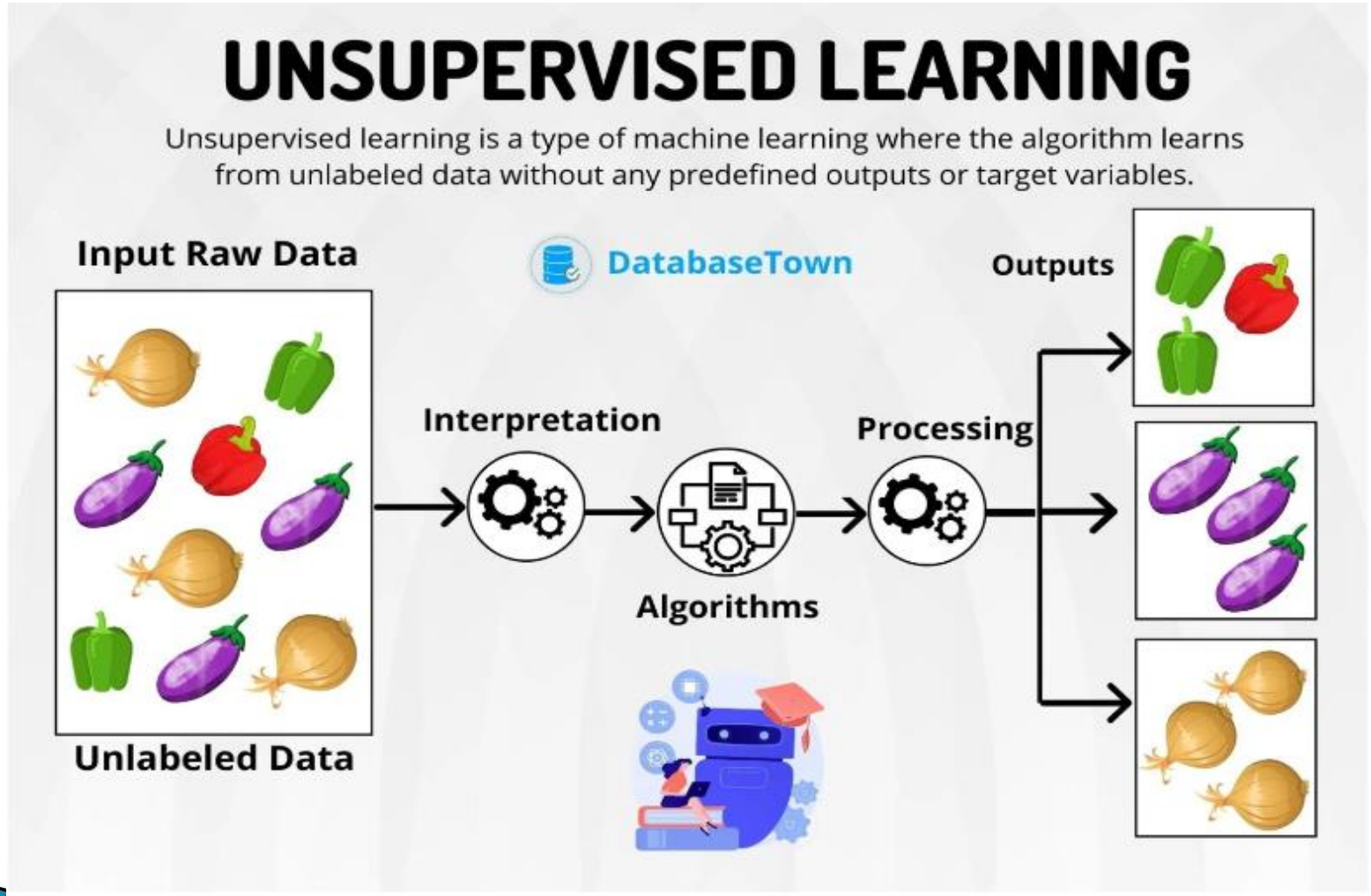
Continue…

- Unsupervised learning cannot be directly applied to a regression or classification problem because unlike supervised learning, we have the input data but no corresponding output data.
- The goal of unsupervised learning is to **find the underlying structure of dataset, group that data according to similarities, and represent that dataset in a compressed format**.

Why use Unsupervised Learning?

- Unsupervised learning is helpful for finding useful insights from the data.
- Unsupervised learning is much similar as a human learns to think by their own experiences, which makes it closer to the real AI.
- Unsupervised learning works on unlabeled and uncategorized data which make unsupervised learning more important.
- In real-world, we do not always have input data with the corresponding output so to solve such cases, we need unsupervised learning.

Working of Unsupervised Learning



Continue…

- Here, we have taken an unlabeled input data, which means it is not categorized and corresponding outputs are also not given.
- Now, this unlabeled input data is fed to the machine learning model in order to train it.
- Firstly, it will interpret the raw data to find the hidden patterns from the data and then will apply suitable algorithms such as k-means clustering, Decision tree, etc.
- Once it applies the suitable algorithm, the algorithm divides the data objects into groups according to the similarities and difference between the objects.

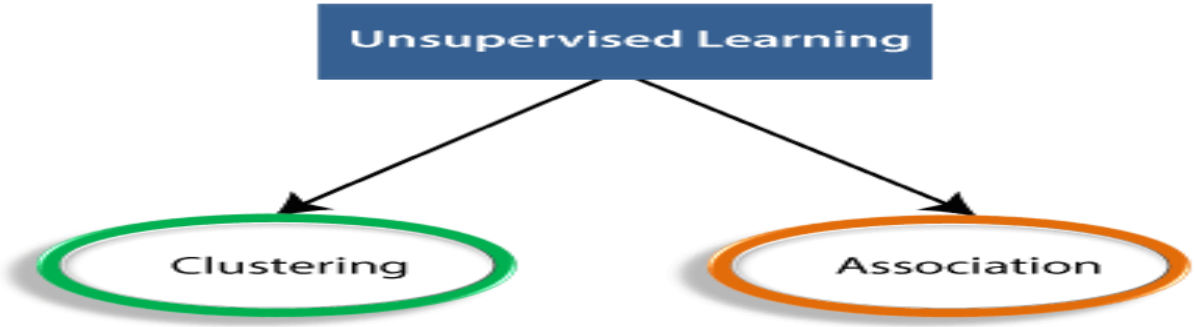
Real world examples of unsupervised Learning

- Data exploration,
- Customer segmentation,
- Recommender systems,
- Target marketing campaigns, and
- Data preparation and visualization
- Computer vision
- Medical imaging
- Anomaly detection

List unsupervised learning algorithms

- **Clustering Algorithm**
  - Hierarchical clustering
  - K-means clustering
  - K-NN (k nearest neighbors)
  - Principal Component Analysis
  - Singular Value Decomposition
  - Independent Component Analysis
- **Association Algorithm**
  - Apriori Algorithm
  - Eclat Algorithm **(Equivalence Class Transformation)**
  - F-P Growth Algorithm (Frequent Pattern)

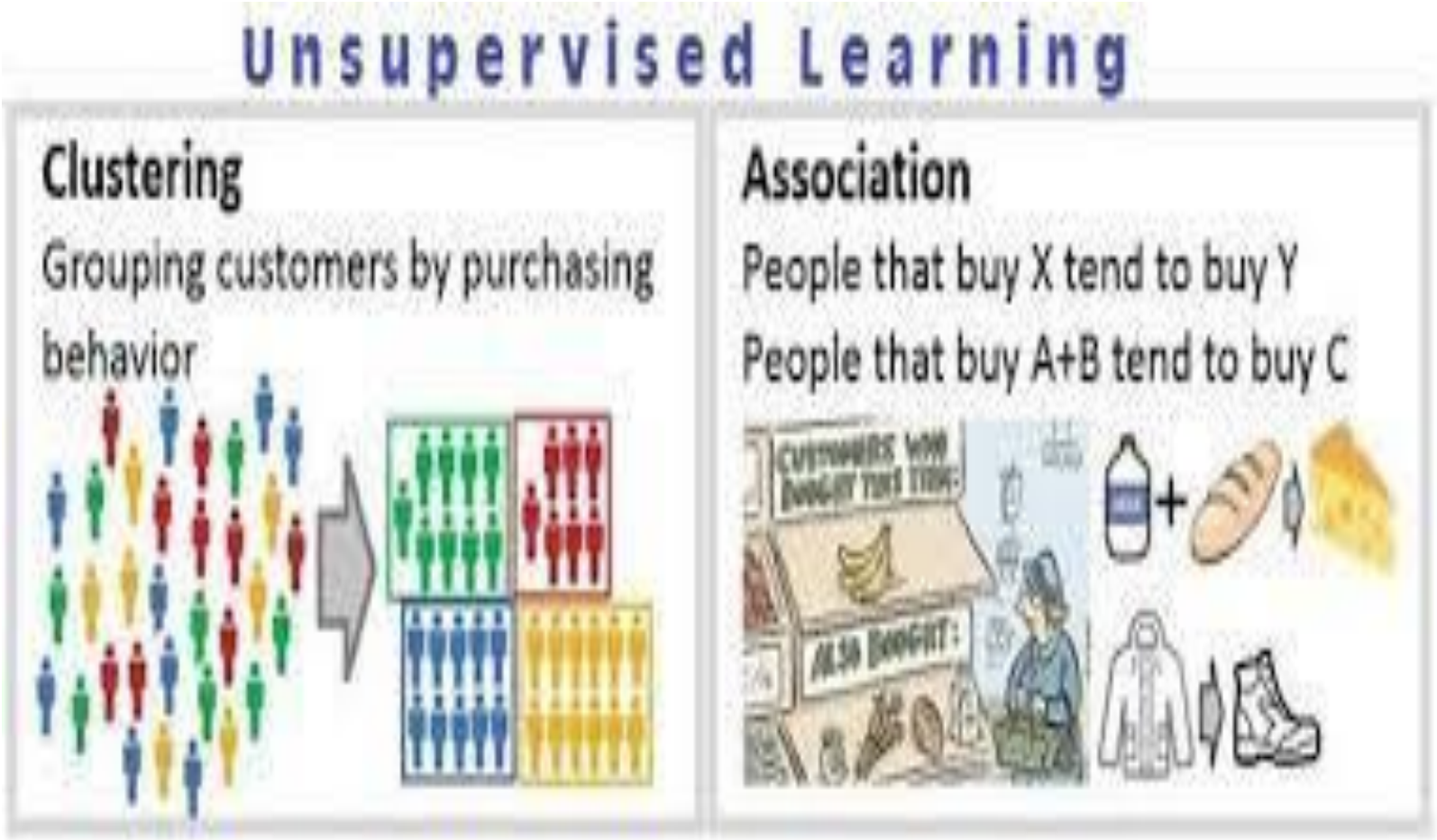
Types of Unsupervised Learning Algorithm



•**Clustering**: Clustering is a method of grouping the objects into clusters such that objects with most similarities remains into a group and has less or no similarities with the objects of another group. Cluster analysis finds the commonalities between the data objects and categorizes them as per the presence and absence of those commonalities.

•**Association**: An association rule is an unsupervised learning method which is used for finding the relationships between variables in the large database. It determines the set of items that occurs together in the dataset. Association rule makes marketing strategy more effective. Such as people who buy X item (suppose a bread) are also tend to purchase Y (Butter/Jam) item. A typical example of Association rule is Market Basket Analysis.

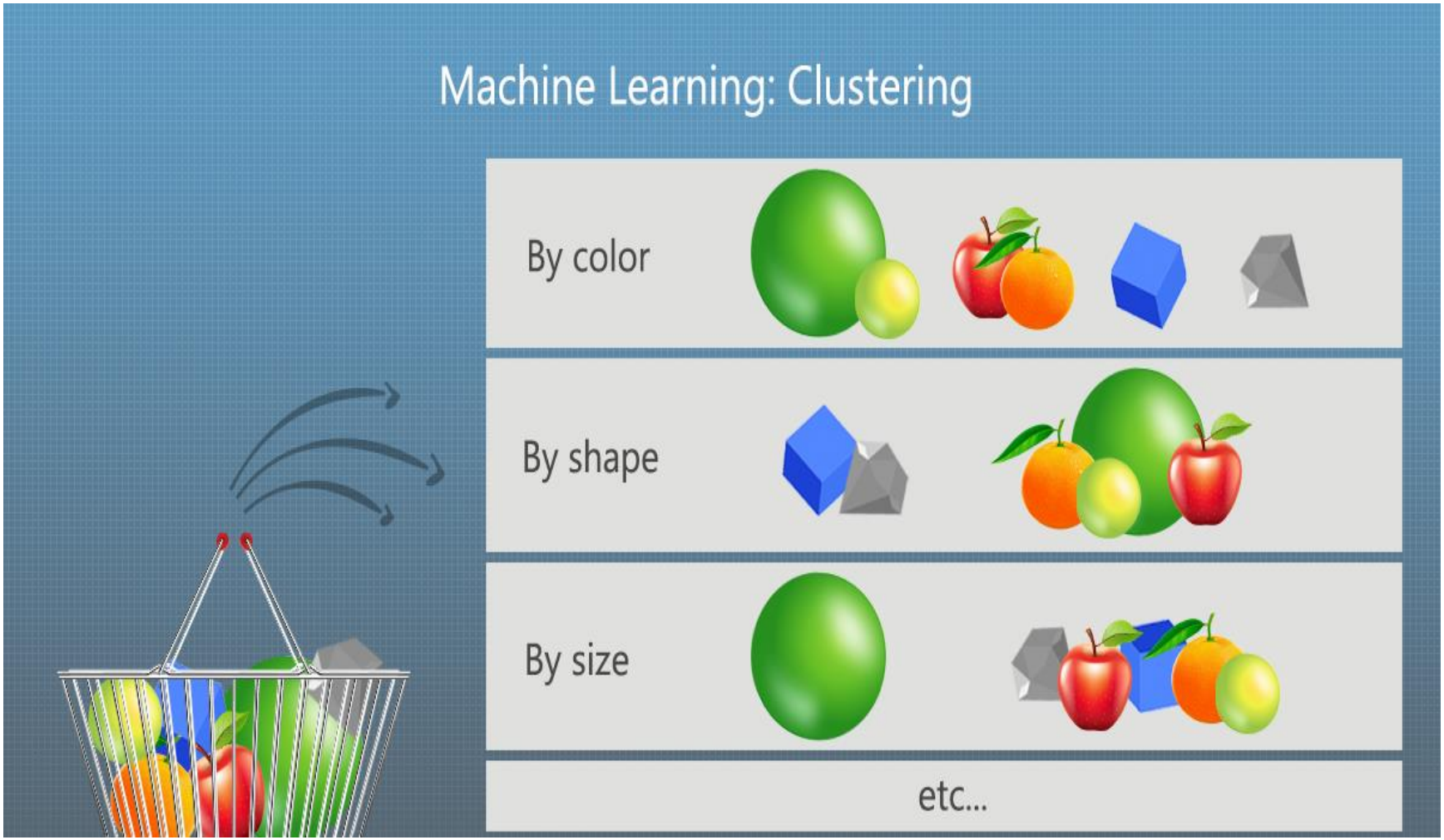
Continue…


Unsupervised Learning
Clustering
Grouping customers by purchasing behavior
Association
People that buy X tend to buy Y
People that buy A+B tend to buy C

Clustering

- **Clustering** is the task of dividing the population or unlabeled 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.
- It is basically a collection of objects on the basis of similarity and dissimilarity between them.

Continue…



Types of Clustering

- **Hard Clustering:** In this, each input data point either belongs to a cluster completely or not.
- **Soft Clustering**: In this, instead of putting each input data point into a separate cluster, a probability or likelihood of that data point being in those clusters is assigned.

Clustering Methods

- Connectivity-based Clustering (Hierarchical clustering)
- Centroids-based Clustering (Partitioning methods)
- Distribution-based Clustering
- Density-based Clustering (Model-based methods)
- Fuzzy Clustering
- Constraint-based

Real world applications/examples

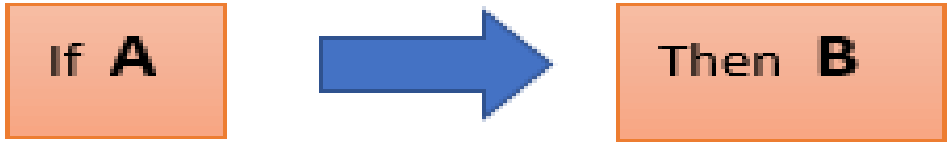
- Recommendation engines
- Market and Customer segmentation
- Social Network Analysis
- Biological Data Analysis,
- Medical Imaging Analysis
- Identification of Cancer Cells
- Identifying Fake News
- Identifying fraudulent activity
- Document Analysis

Association

- Association rule learning is a type of unsupervised learning technique that checks for the dependency of one data item on another data item.
- It tries to find some interesting relations or associations among the variables of dataset.
- It is based on different rules to discover the interesting relations between variables in the database.

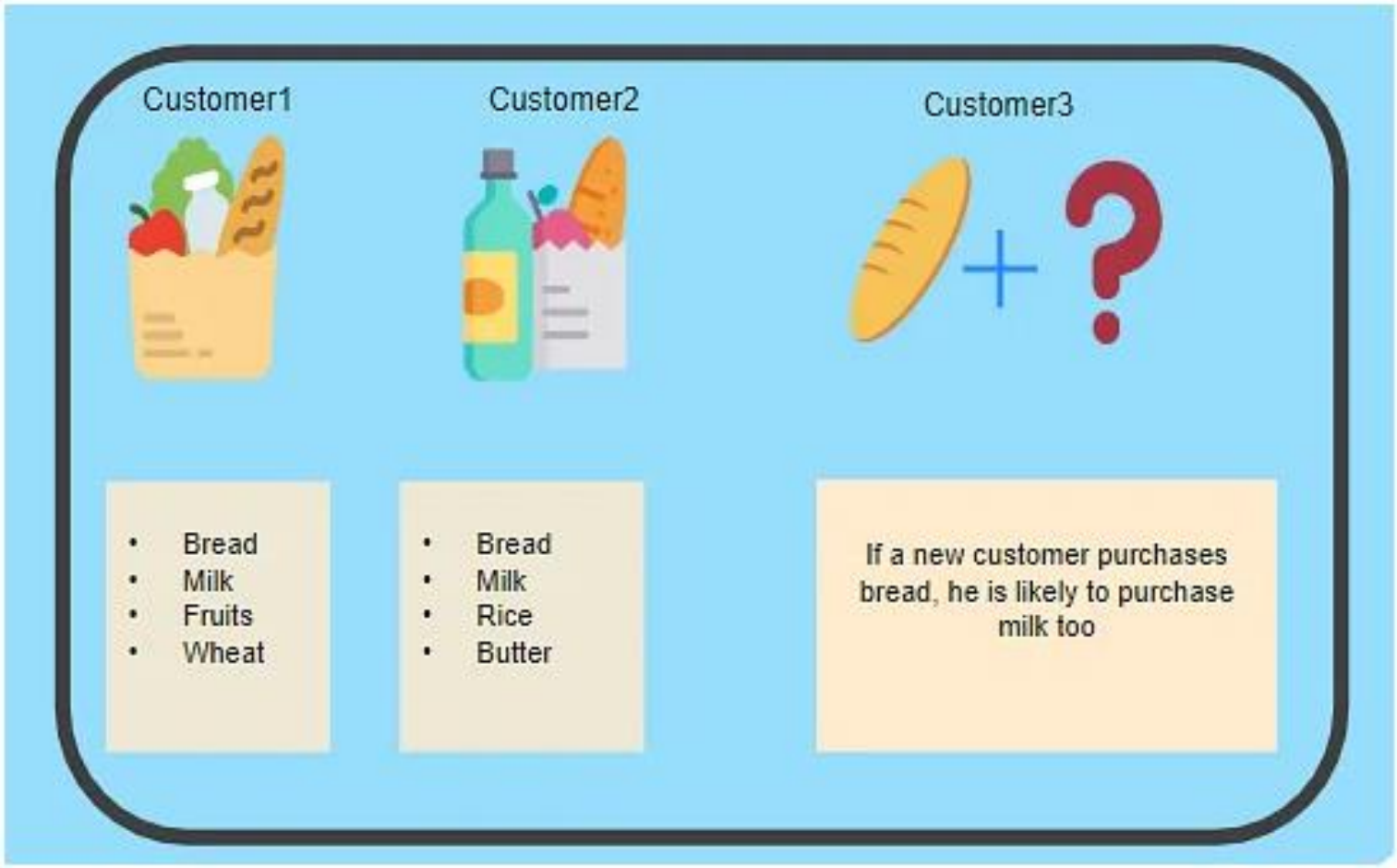
How does Association Rule Learning work?

- Association rule learning works on the concept of If and Else Statement, such as if A then B.



- Here the If element is called antecedent, and then statement is called as Consequent.
- These types of relationships where we can find out some association or relation between two items is known as single cardinality.
- It is all about creating rules, and if the number of items increases, then cardinality also increases accordingly.

Continue...


Customer1
Customer2
Customer3
• Bread
• Milk
• Fruits
• Wheat
• Bread
• Milk
• Rice
• Butter
If a new customer purchases bread, he is likely to purchase milk too

Real world applications/examples

- Market Basket Analysis
- Medical Diagnosis
- Banking services
- Web usage mining and intrusion detection
- Bioinformatics

Advantages and Disadvantages of Unsupervised Learning

- **Advantages**:
- Unsupervised learning is used for more complex tasks as compared to supervised learning because, in unsupervised learning, we don't have labeled input data.
- Unsupervised learning is preferable as it is easy to get unlabeled data in comparison to labeled data.
- **Disadvantages**:
- Unsupervised learning is intrinsically more difficult than supervised learning as it does not have corresponding output.
- The result of the unsupervised learning algorithm might be less accurate as input data is not labeled, and algorithms do not know the exact output in advance.

Applications of Unsupervised Learning

- Network Analysis: Unsupervised learning is used for identifying plagiarism and copyright in document network analysis of text data for scholarly articles.
- Recommendation Systems: Recommendation systems widely use unsupervised learning techniques for building recommendation applications for different web applications and e-commerce websites.
- Anomaly Detection: Anomaly detection is a popular application of unsupervised learning, which can identify unusual data points within the dataset. It is used to discover fraudulent trans.
- Customer segmentation: Interesting buyer persona profiles can be created using unsupervised learning. This helps businesses to understand their customers' common traits and purchasing habits, thus, enabling them to align their products more accordingly.

Differentiate Supervised and Unsupervised Learning

Supervised Learning	Unsupervised Learning
Supervised learning algorithms are trained using labeled data.	Unsupervised learning algorithms are trained using unlabeled data.
Supervised learning model takes direct feedback to check if it is predicting correct output or not.	Unsupervised learning model does not take any feedback.
The goal of supervised learning is to train the model so that it can predict the output when it is given new data.	The goal of unsupervised learning is to find the hidden patterns and useful insights from the unknown dataset.
In supervised learning, input data is provided to the model along with the output.	In unsupervised learning, only input data is provided to the model.

Continue…

Supervised Learning	Unsupervised Learning
Supervised learning needs supervision to train the model.	Unsupervised learning does not need any supervision to train the model.
Less Computational Complexity	More Computational Complex
Supervised learning can be used for those cases where we know the input as well as corresponding outputs.	Unsupervised learning can be used for those cases where we have only input data and no corresponding output data.
Supervised learning model produces an accurate result.	Unsupervised learning model may give less accurate result as compared to supervised learning.

Continue…

Supervised Learning	Unsupervised Learning
Supervised learning is not close to true Artificial intelligence as in this, we first train the model for each data, and then only it can predict the correct output.	Unsupervised learning is more close to the true Artificial Intelligence as it learns similarly as a child learns daily routine things by his experiences.
Supervised learning can be categorized in **Classification** and **Regression** problems.	Unsupervised Learning can be classified in **Clustering** and **Associations** problems.
It includes various algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, Multi-class Classification, Decision tree,etc.	It includes various algorithms such as Clustering, KNN, and Apriori algorithm.

Happy Learning