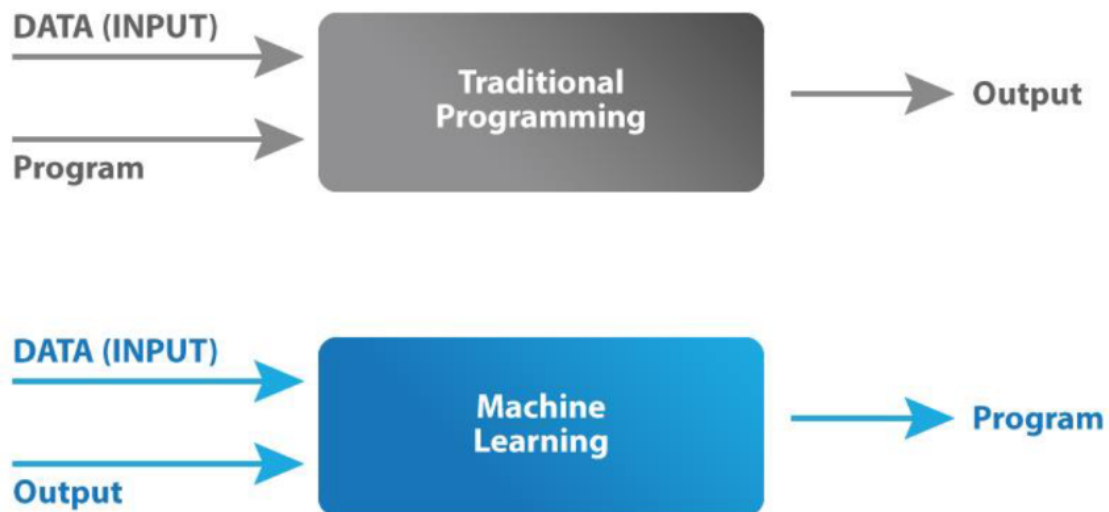


ML-1

Q) What is machine learning?

Ans. Machine learning is that field of computer science in which we train the computer without doing any programming.

Evolved from `pattern recognition` and `computational learning theory` from AI.



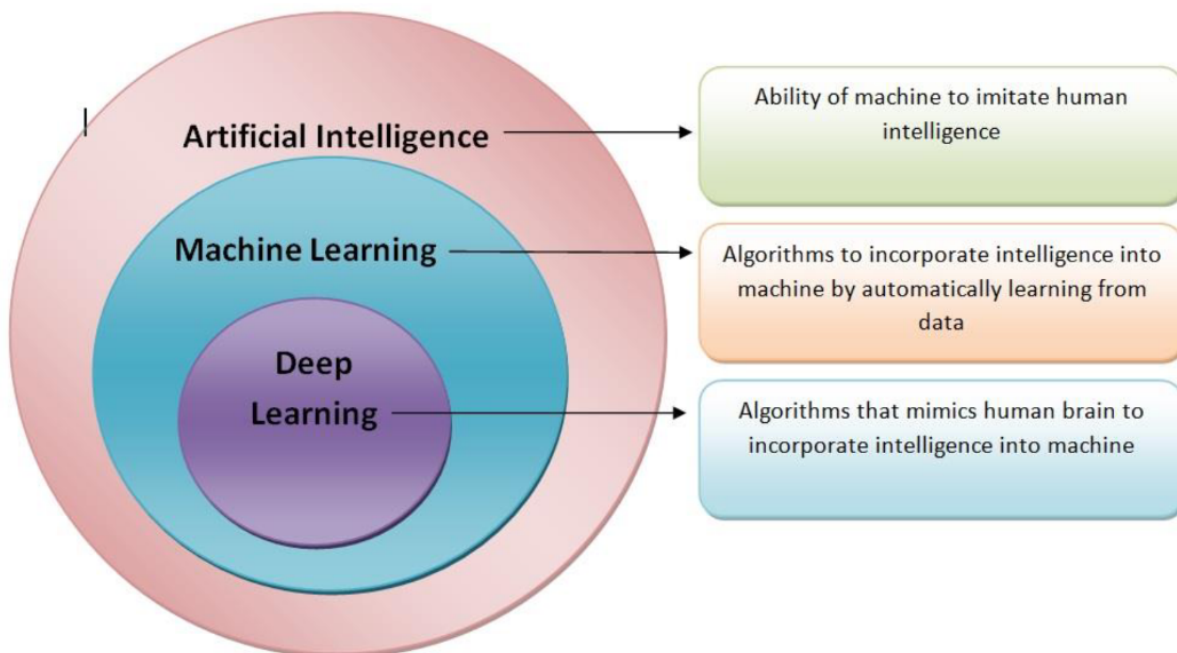
History of ML

- In 1764, `Thomas bayes` started this new technology by giving his `foundatitions on statistics`.
- In 1842, `Ada Lovelace` gave his `foundations on algorithm`.
- In 1847, `George Boole` gave `Boolean Logic`.

- In 1936, **Alan Turing** proposed a machine that can learn.
- In 1952, **Arthur Samuel** created the first computer programme at **IBM**.
- In 1959, **MADALINE** - the first **artificial neural network** was invented.
- In 2006, **Geoffrey Hinton** invents the term **Deep Learning**.
- In 2016, **AlphaGo** was the first machine which defeated the human player in the famous game **GO**.

AI vs DL vs ML

DL(Deep Learning) is just a subset of ML which employs artificial neural networks for complex tasks.



Data Science vs Data Analytics vs AI/ML/DL

Data Science is a term which contains data analytics , data mining , machine learning and little bit of deep learning.

Data Science actually predicts the future based on their past while Data Analytics takes the meaningful insights of the data sources.

Data Scientist

Data scientists apply statistics, machine learning, and analytic approaches to answer critical business questions. Data scientists interpret and deliver the results of their findings by using visualization techniques, building data science apps, or narrating exciting stories about the solutions to their data (business) problems. They work with data sets of different sizes and run algorithms on large data sets. Data scientists must be current with the latest automation and machine learning technologies. The requirements to perform these roles include statistical and analytical skills, programming knowledge (Python, R, Java), and familiarity with Hadoop, a collection of open-source software utilities that facilitates working with massive amounts of data. Data scientists are data wranglers who organize and deliver value from data.

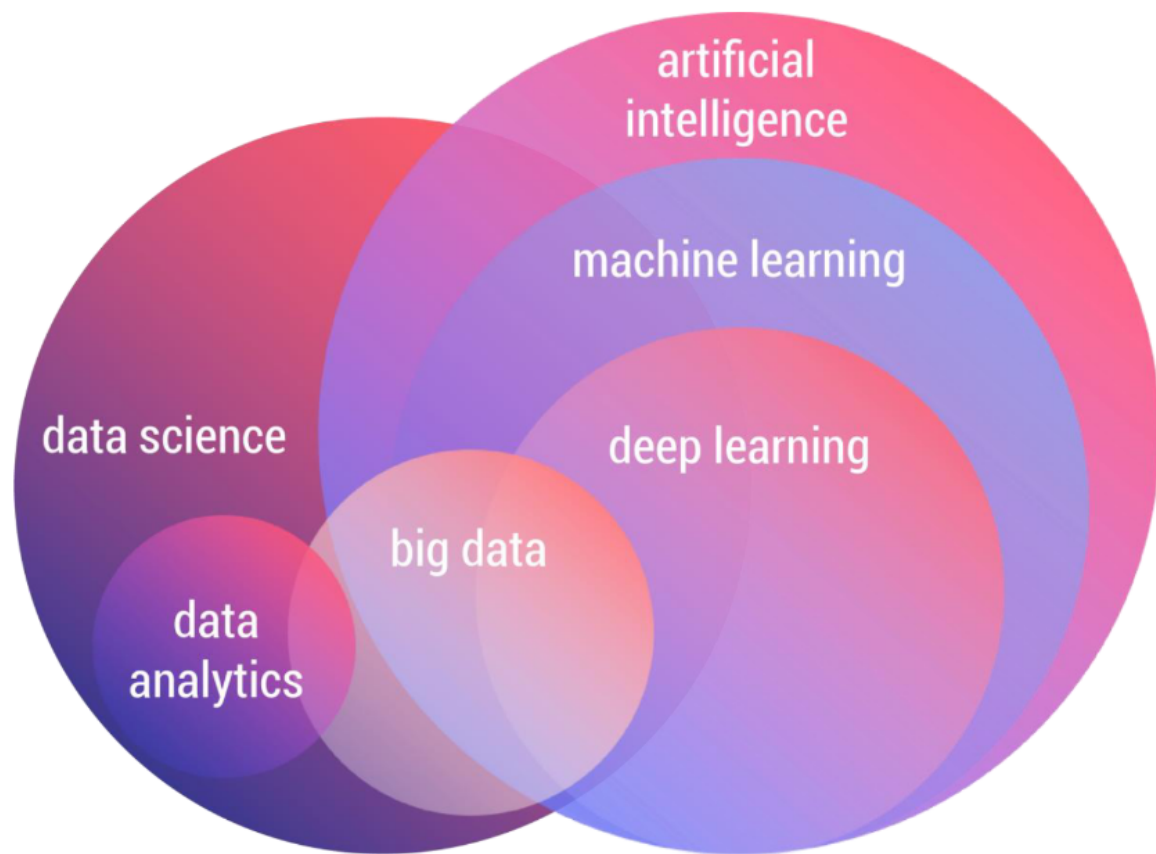
Data Analyst

Data analysts query, process, provide reports, and summarize and visualize data. They leverage existing tools and methods to solve a problem. They help people, such as business analysts, to understand specific queries with ad-hoc reports and charts. Data analysts must understand basic statistical principles, cleaning different data types, visualization, and exploratory data analysis. In short, data analysts analyze data to help businesses and other organizations make informed decisions.

Data Engineer

Data engineers are responsible for building and operationalizing data pipelines to collect and organize data. They ensure the accessibility and availability of quality data for data scientists and data analysts by integrating data from disparate sources and performing data cleaning and transformation. Skills needed for data

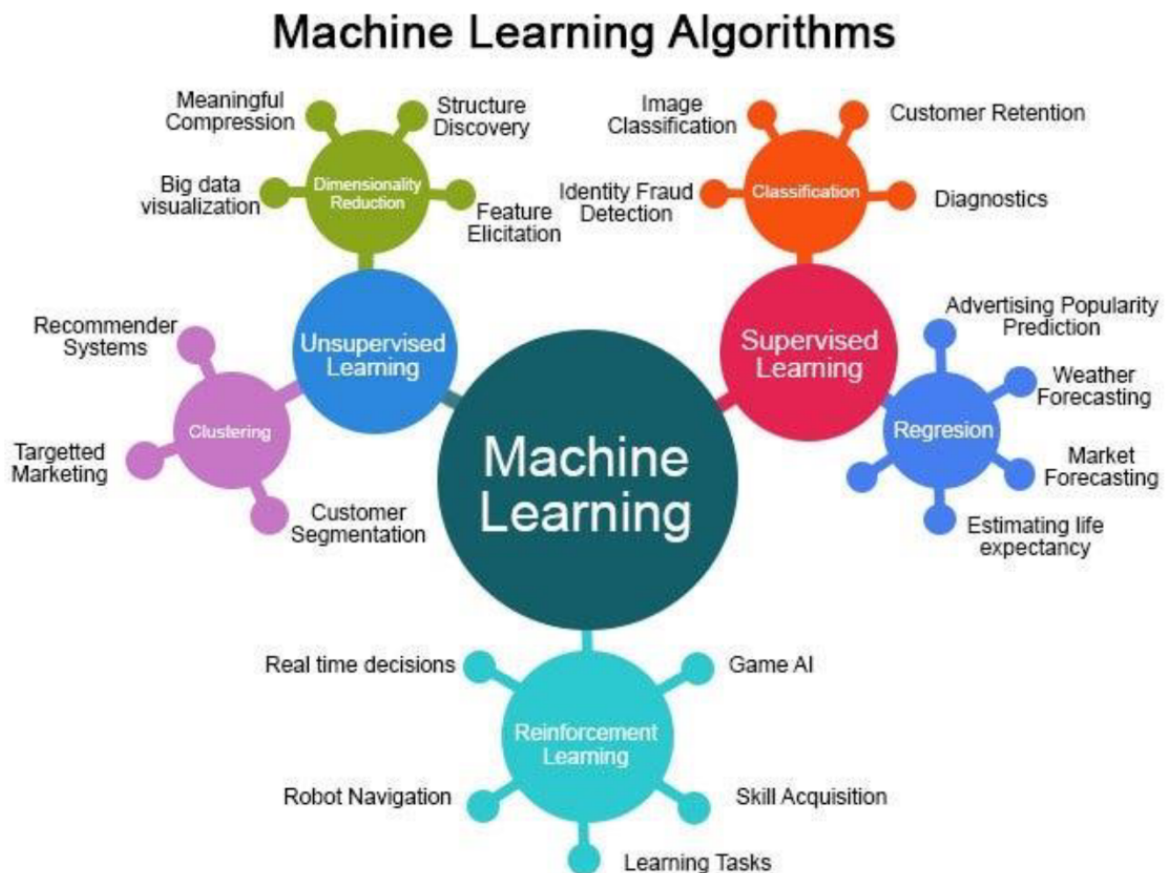
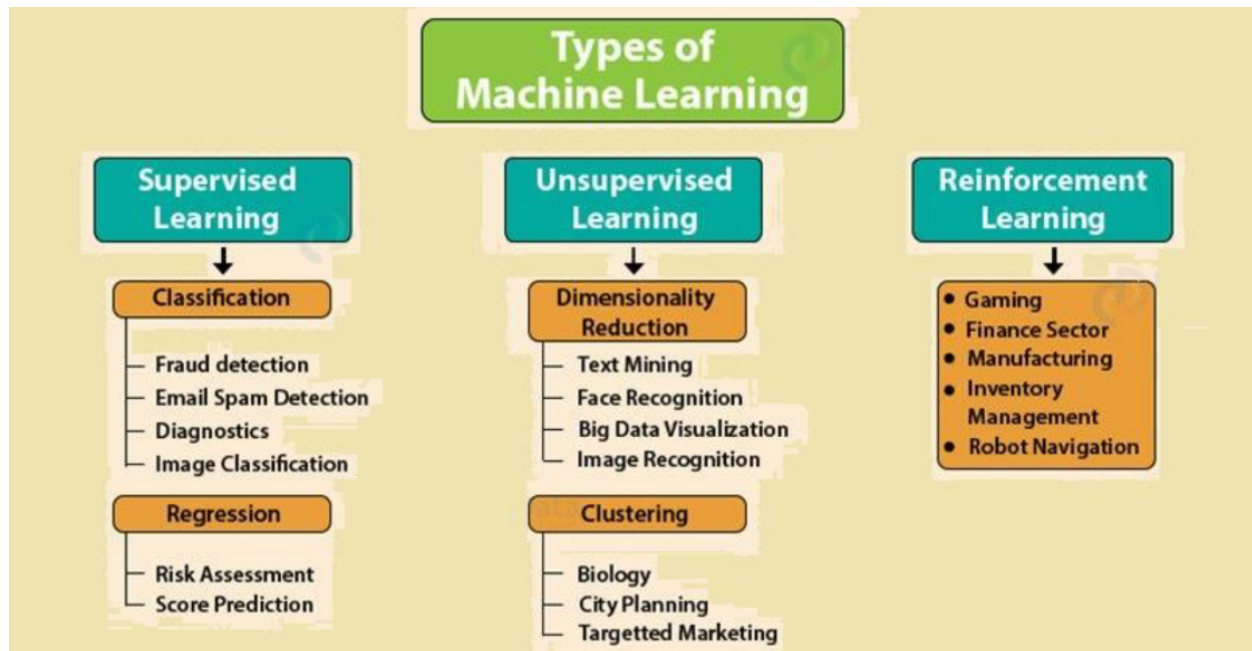
engineering roles include understanding the architecture, tools, and methods of data ingestion, transformation and storage; and proficiency with multiple programming languages (including Python and Scala). In summary, data engineers build and operate the data infrastructure needed to prepare data for further analysis by data analysts and scientists.



Types of machine learning

There are 3 types of machine learning

- Supervised machine learning
- Unsupervised machine learning
- Reinforcement machine learning



1)Supervised Machine Learning

It is a model which gets trained on the labelled dataset. labelled dataset is the dataset which has both inputs and outputs. It points the correct output and marks its range resp.

Example: If we have 10,000 Students Data who have placed or some un placed and we want to find the new predictive result output.

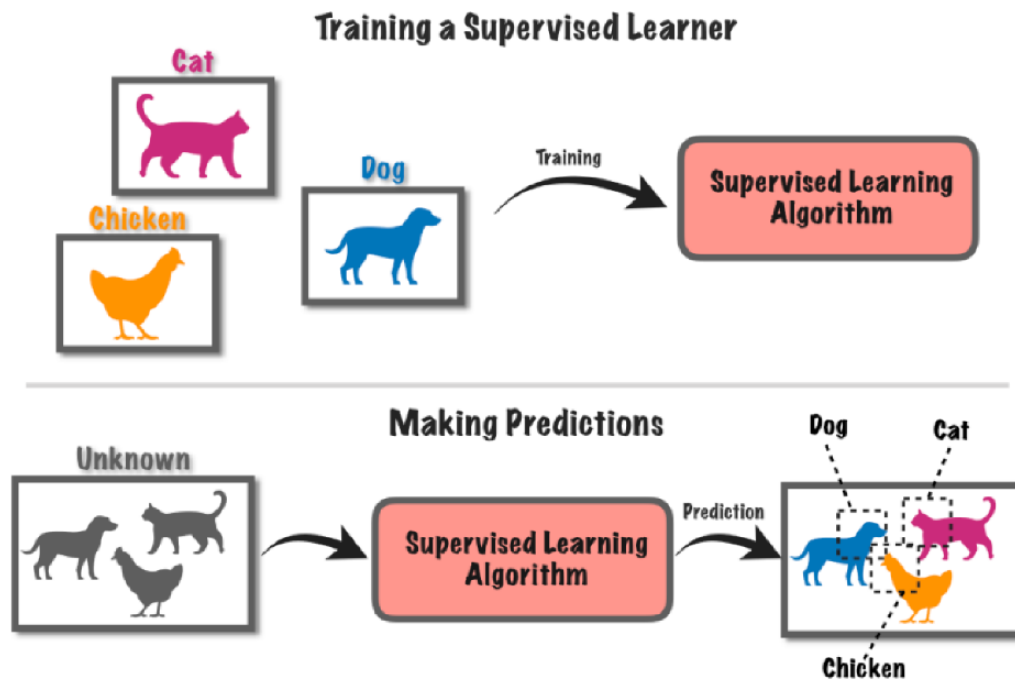
					Output
S.No	Student_Name	St_Percentage	St_CGPA	St_Placement	Package
1	Nishant	87%	7.8	Yes	8 LPA
2	Swati	75%	7.2	No	0 LPA
3	Ragini	91%	8.4	Yes	10 LPA
4	Salman	68%	6.5	No	0 LPA
.					
.					
10K	XYZ	00%	0.0	No	0

Conclusion: Draw Mathematical Relationship for generate a new result or indicate the prediction.

Input (87%, 7.8) + Output (Learning ML) = (Program) New Result Output

There are 2 categories of supervised machine learning

1. Regression: it gives the numerical output.
2. Classification: It gives the categorical output.



Advantages of supervised machine learning

- Can have high accuracy as they are trained on the labelled dataset.
- it can be used in pre-trained models which saves time and resources when making new models.

Disadvantages of supervised machine learning

- it may struggle in the data which is not present in the training data.
- it can be costly.

Applications of supervised learning

- Image Classification

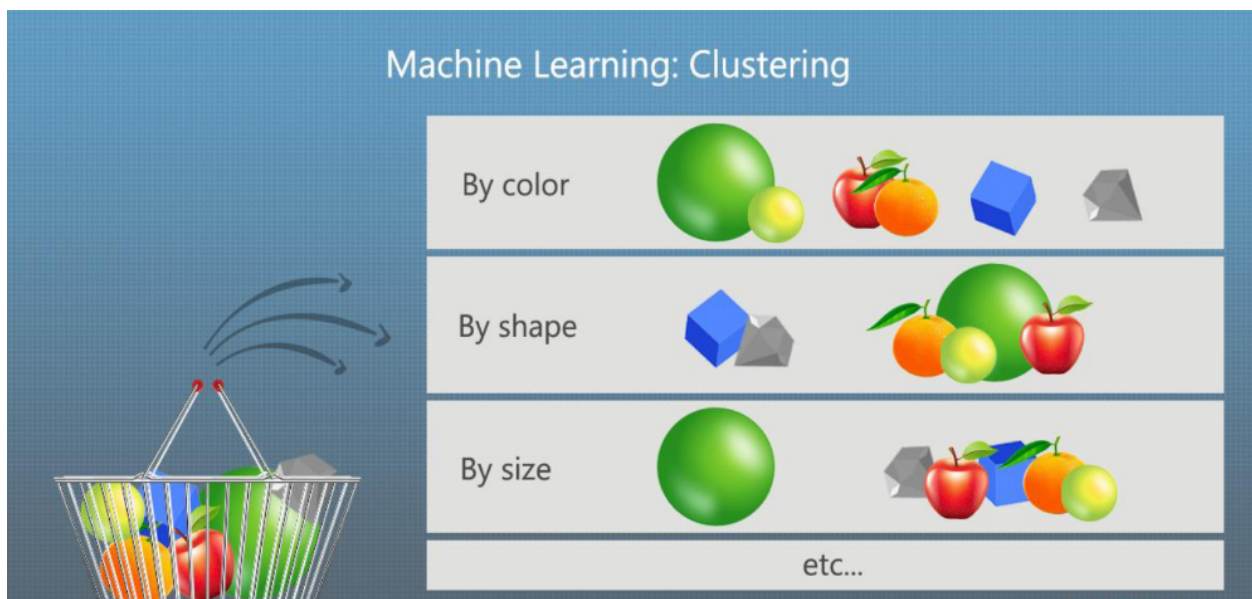
- Speech Recognition
- Fraud Detection
- Email spam Detection
- Weather Forecasting

2) Unsupervised machine learning

In this type we don't actually give the output to the machine but just give the raw data and let the machine make its own algorithm and group the data accordingly and make its own patterns and relationships

Types of unsupervised machine learning

1. **Clustering**: collection of objects on the basis of similarities and dissimilarities.

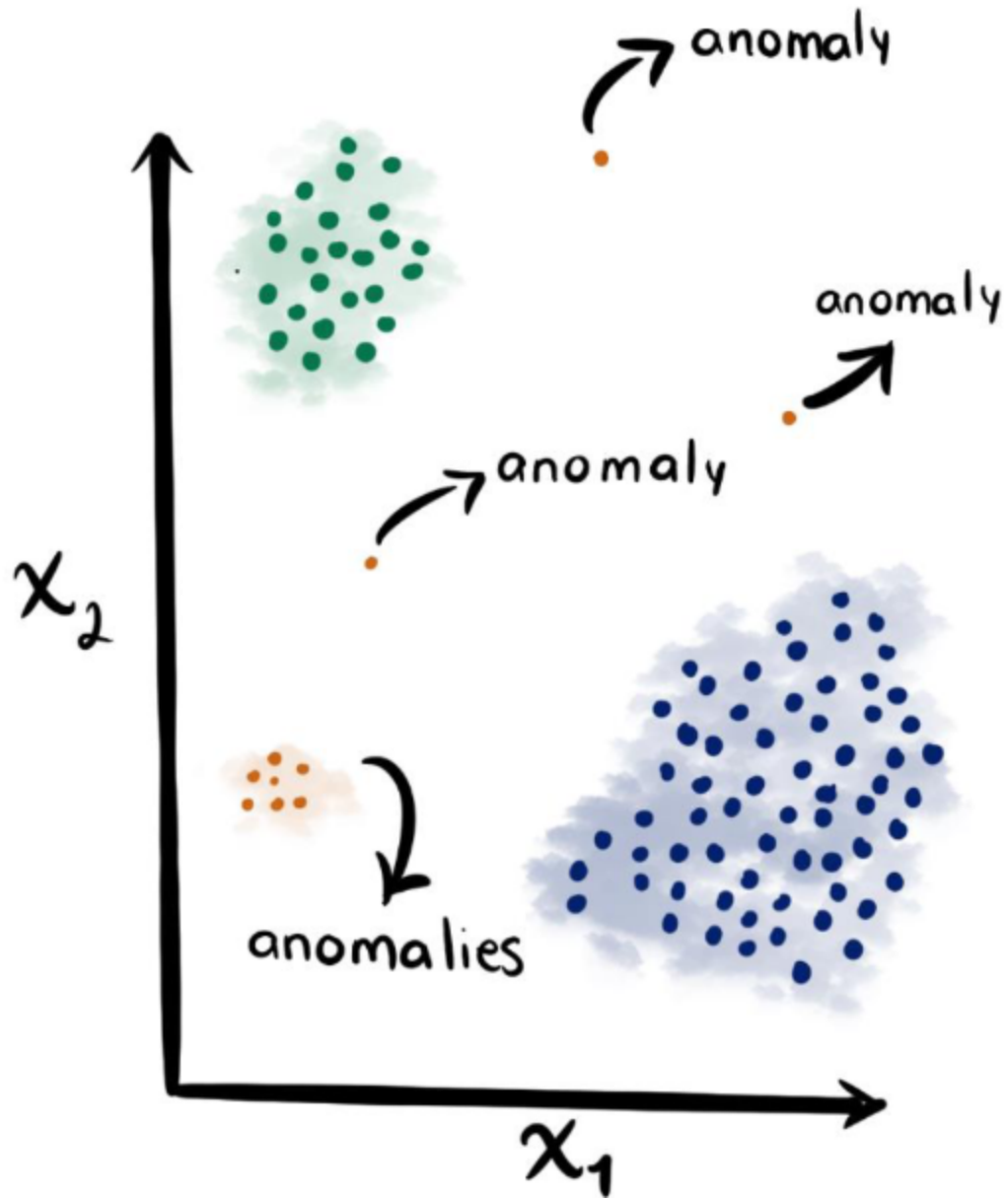


2. **Dimensionality reduction**: Retaining the number of features (dimensions) in the dataset.

Type of Rooms	No Of Rooms	Room Size
Bedroom	3	10*15 10*10 10*12
Washroom	2	6*6 6*8
Kitchen	1	10*8

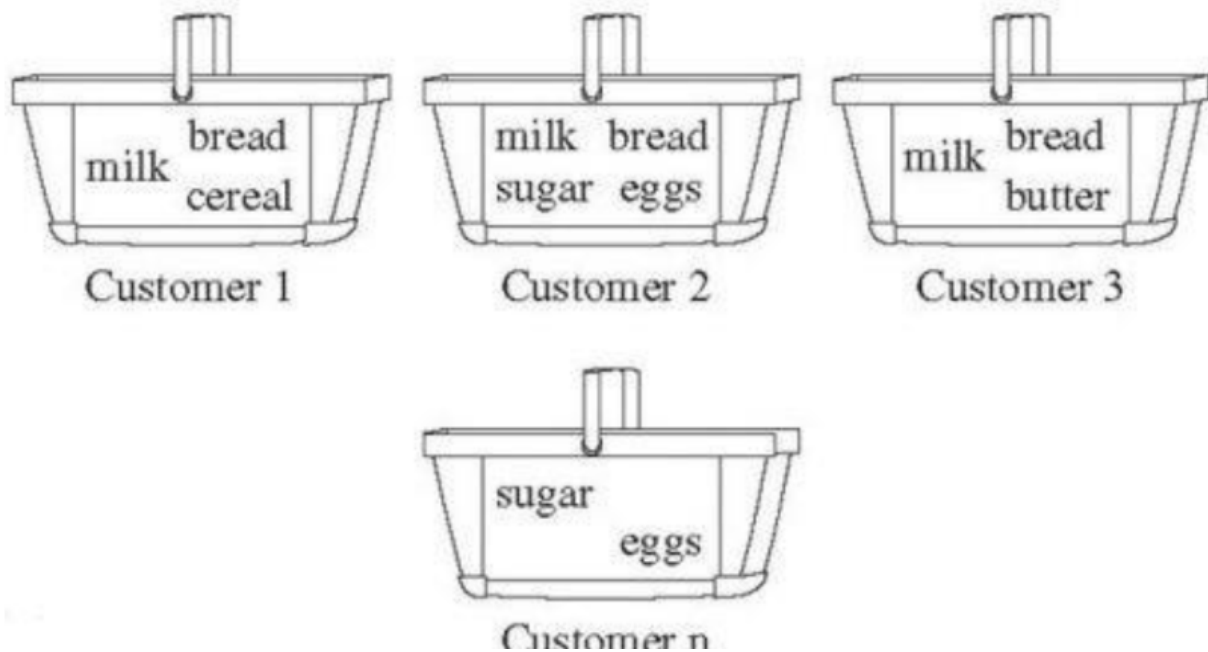
Area
370 Sqft
84 Sqft
80 Sqft

3. **Anomaly detection** : it is used to detect the anomaly , important concepts in ml and data mining as it helps to get the usual behavior's or some errors in the specific data.



4. **Association Rule learning**: it is used to uncover different hidden relationships in the dataset .

Example: Study many purchase product in shopping mart.



Advantages of Unsupervised learning

- helps to discover hidden patterns and various relationships.
- Reduces the effort for data labelling.

Disadvantages of unsupervised learning

- without labels it is difficult to get the accuracy of the model.
- cluster interpretability may not be clear and the interpretations may not be meaningful.

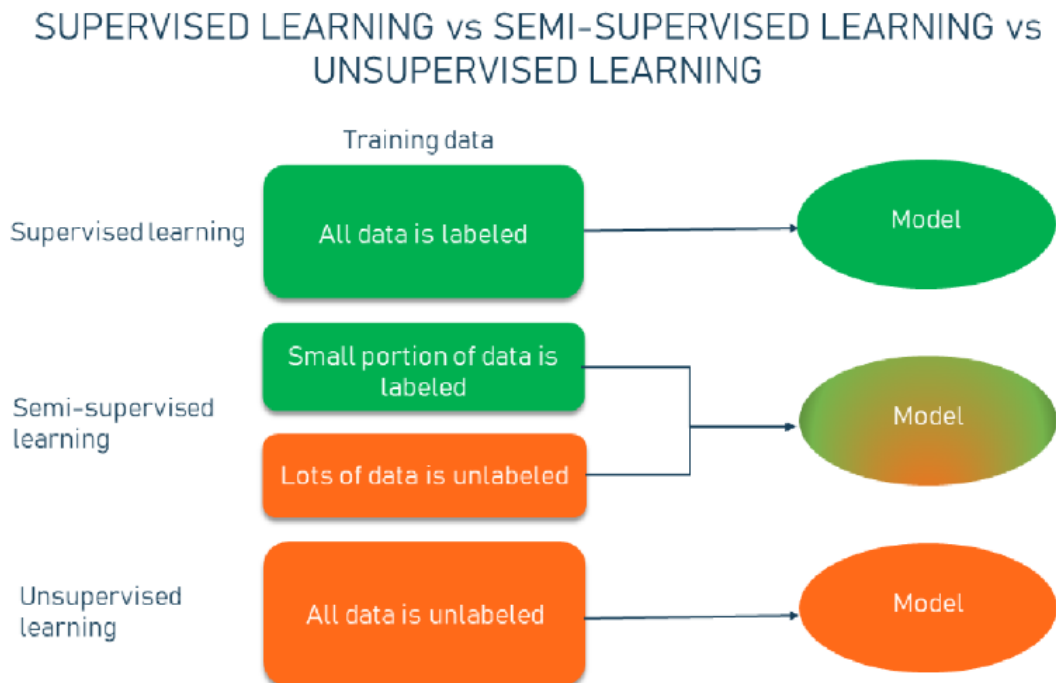
Applications of Unsupervised learning

- Recommendation Systems
- Image segmentation
- Genomic data science
- Customer behavior analysis

3) Semi supervised machine learning

we need to ignore the labelling as its very costly that is the reason of there creation.

Example: Google photos



4) Reinforcement learning

it is a method in which we award when the machine is correct and we punish if the machine is not correct .

In basic , it interprets the environment and take the decisions accordingly and learn through there trial period and errors.

