Data Science and ML

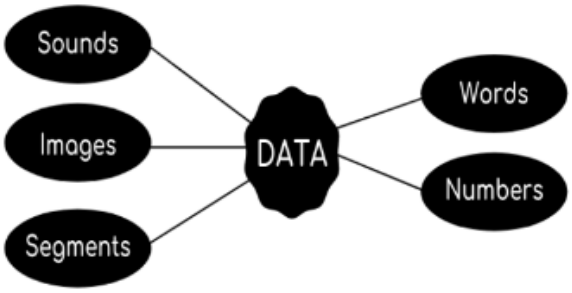
ML ALGORITHMS

Yeasin Bhuiyan

What is Data?

Data is defined as a collection of organized or unorganized facts, concepts, or instructions in a formalized manner suitable for communication, interpretation, processing by humans, or some automatic means such as computers, ATMs.

The main examples of data are phone numbers, weights, prices, costs, number of items sold, product names, addresses, registration marks, etc.



What is Information?

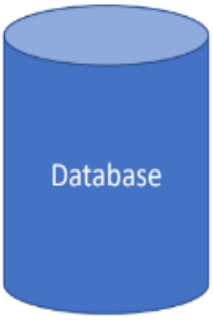
Information is defined as the **processed data** which helps us in making **decisions**. It is an intelligent form of data. such as bills, profits, reports, receipts, comparison of sales figures, merit list, printed documents, etc.

For example, the students' marks and their roll numbers form the data, and their report card/sheet is the information that helps us decide which student stood at which rank in the class.

What is Database?

A database is an organized collection of related data or information stored and accessed electronically within a computer system.

For example, MySQL, SQL Server, MongoDB, Oracle Database, PostgreSQL, Informix, Sybase, etc. are all examples of different databases. These modern databases are managed by DBMS. Structured Query Language, or SQL as it is more widely known, is used to operate on the data in a database.



What is Data Science?

- Computer Science
- Mathematics
- Statistics

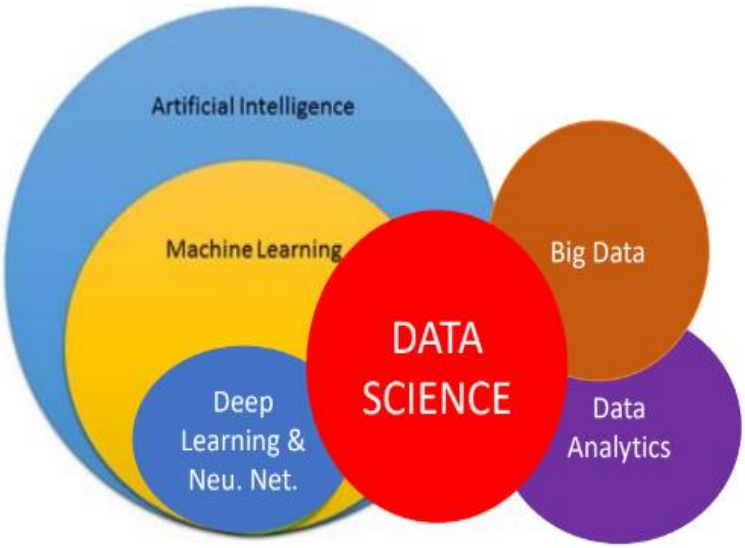
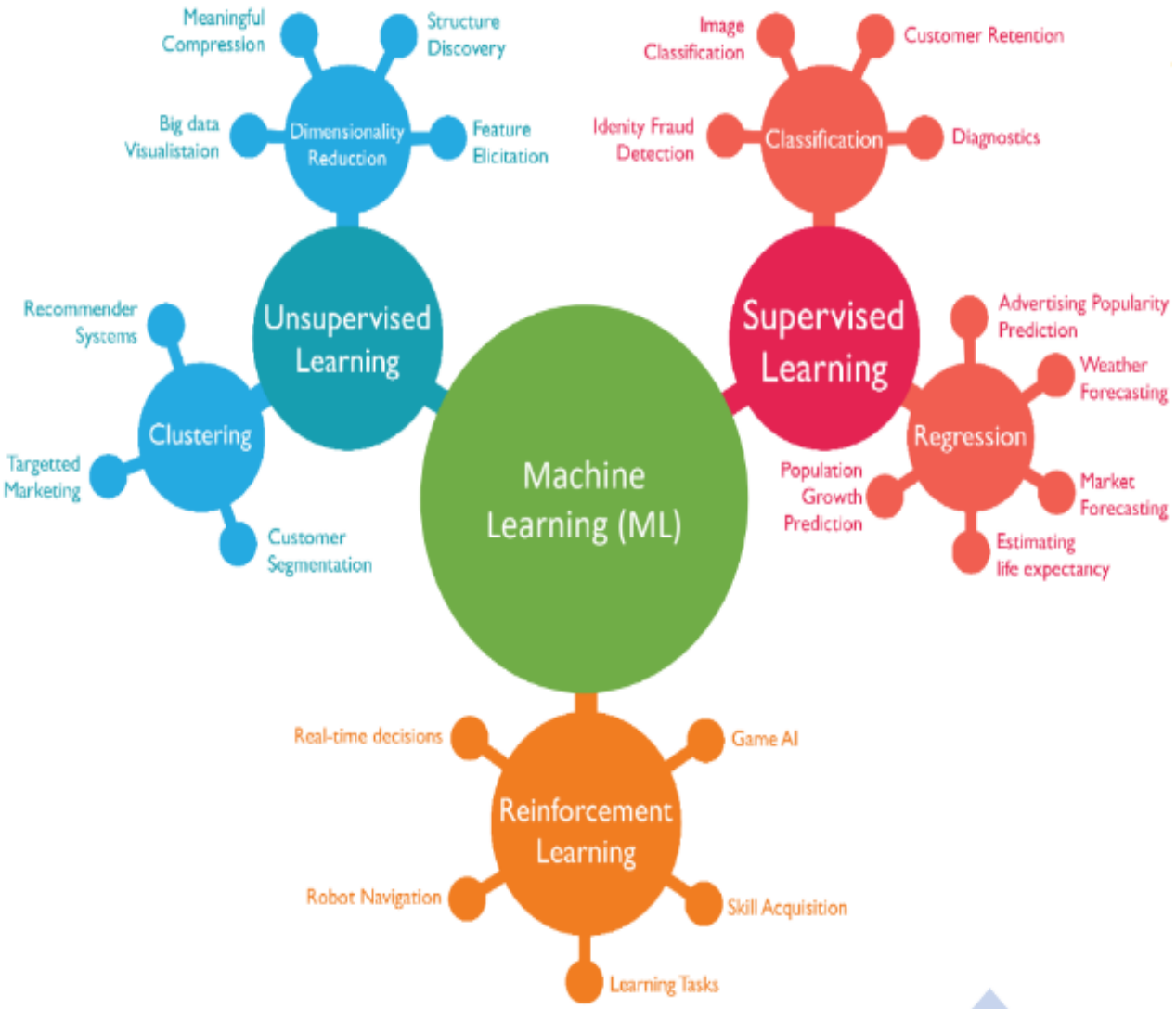
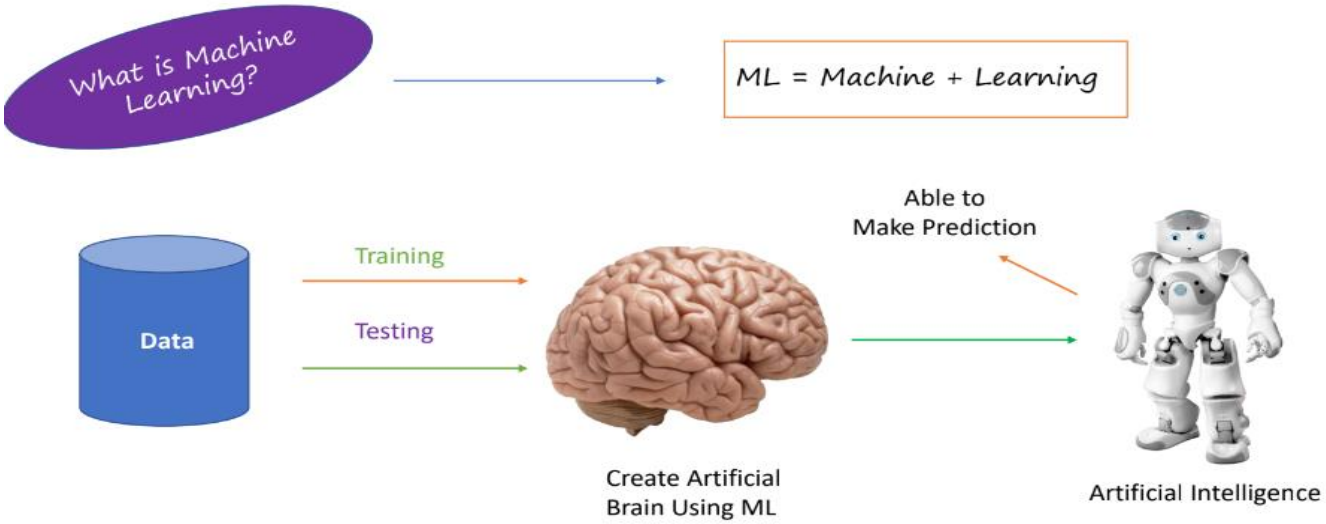
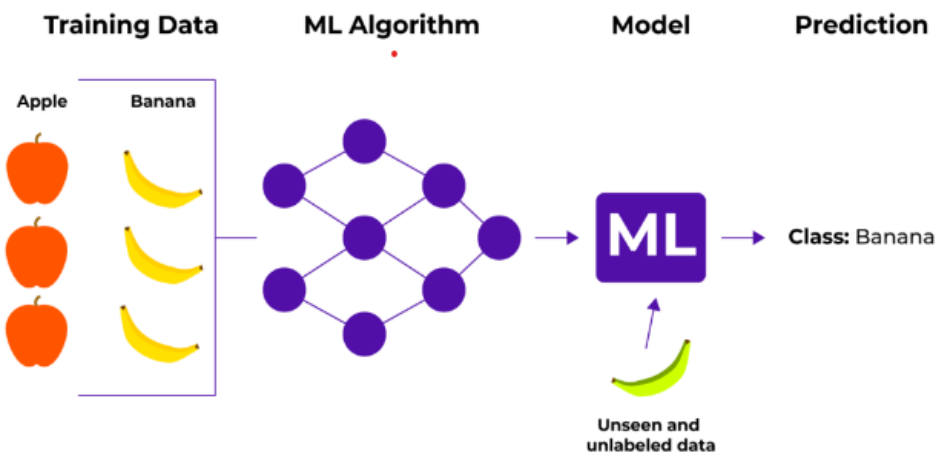


Fig: Venn Diagram


Machine Learning (ML)
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks


What is Machine Learning?
ML = Machine + Learning
Data
Training
Testing
Create Artificial Brain Using ML
Able to Make Prediction
Artificial Intelligence

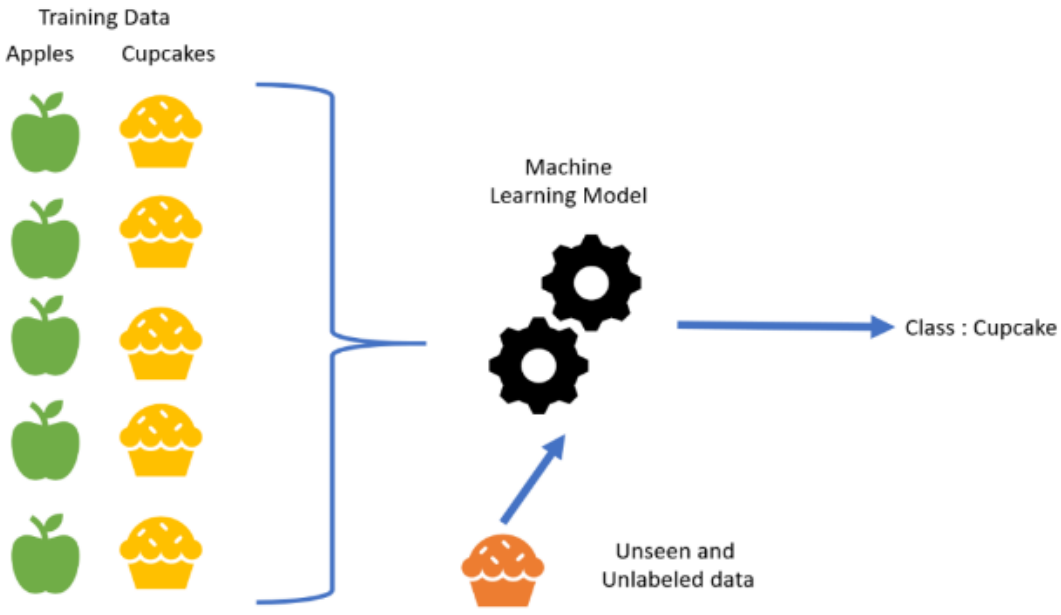
Supervised Learning:



Mind it

- Labeled
- At least two classes

Let's have an Example -



Regression

Let's have another Example-

Output Column

No	X1 transactior	X2 house age	X3 distanc	X4 number	X5 latitude	X6 longitude	Area
1	2012.917	32	84.87882	10	24.98298	121.54024	37.9
2	2012.917	19.5	306.5947	9	24.98034	121.53951	42.2
3	2013.583	13.3	561.9845	5	24.98746	121.54391	47.3
4	2013.5	13.3	561.9845	5	24.98746	121.54391	54.8
5	2012.833	5	390.5684	5	24.97937	121.54245	43.1
6	2012.667	7.1	2175.03	3	24.96305	121.51254	32.1
7	2012.667	34.5	623.4731	7	24.97933	121.53642	40.3
8	2013.417	20.3	287.6025	6	24.98042	121.54228	46.7
9	2013.5	31.7	5512.038	1	24.95095	121.48458	18.8
10	2013.417	17.9	1783.18	3	24.96731	121.51486	22.1
11	2013.083	34.8	405.2134	1	24.97349	121.53372	41.4
12	2013.333	6.3	90.45606	9	24.97433	121.5431	58.1
13	2012.917	13	492.2313	5	24.96515	121.53737	39.3
14	2012.667	20.4	2469.645	4	24.96108	121.51046	23.8
15	2013.5	13.2	1164.838	4	24.99156	121.53406	34.3
16	2013.583	35.7	579.2083	2	24.9824	121.54619	50.5
17	2013.25	0	292.9978	6	24.97744	121.54458	70.1
18	2012.75	17.7	350.8515	1	24.97544	121.53119	37.4
19	2013.417	16.9	368.1363	8	24.9675	121.54451	42.3
20	2012.667	1.5	23.38284	7	24.96772	121.54102	47.7

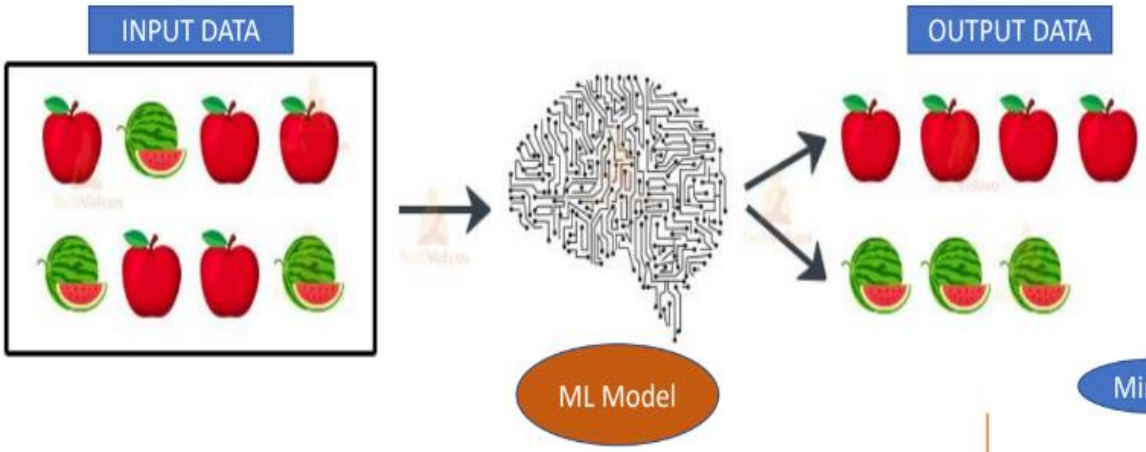
Features:

- X1
- X2
- X3
- X4
- X5
- X6

Mind it

- Labeled
- Continuous Data

Unsupervised Learning



Mind it

- Un-Labeled
- At least two classes

Regression Algorithm:

- Linear Regression
- XGBoost (Regressor & Classifier)
- AdaBoost (Regressor & Classifier)

Classification Algorithms: (Also Have Regressor)

- Logistic Regression
- Decision Tree
- Naïve Bayes
- Random Forest
- SVM
- KNN
- CatBoost

Unsupervised Algorithm:

- Clustering (K-means, O-Cluster)
- PCA
- Feature Selection