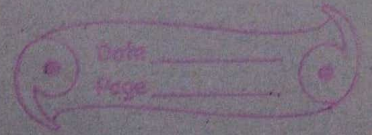
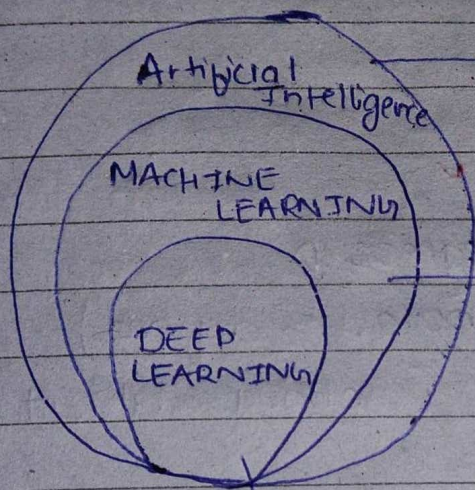


# MACHINE LEARNING



Machine Learning is a subset of artificial intelligence. It focuses mainly on the designing of systems, thereby allowing them to learn and make predictions based on some experience which is data in case of machines.



Artificial Intelligence

AI - A Technique which enables machines to mimic human behavior

ML - subset of AI techniques which use statistical methods to enable machines to improve with experience

Deep learning

Subset of ML which make the computation of multi-layer neural network feasible

## Machine Learning Types

- Supervised
- Unsupervised
- Reinforcement

Supervised ex- Cortana, Weather app, Biometric attendance

~~Unsupervised~~



Supervised learning - Supervised learning is when the model is getting trained on a labelled dataset. Labelled dataset is one which have both input and output

## Supervised learning

Classification  
(defined labels)  
[Dataset is Discrete]

↓

userID	Gender	age	salary
15624510	M	19	19000
15810944	M	35	20000
15668575	F	26	43000

Regression  
(no labels defined)  
[Continuous dataset]

→ values  
can only be  
0 or 1

Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
10.62261758	986.88209	54.1937313	195.723456	3.275823
13.59184184	987.8729248	189.281202	189.2951202	2.901876
17.70494885	986.1119835	192.9278384	192.927838	1.234867

can be anything b/w  
1-5



# Unsupervised Learning

It's a type of learning where we don't give target to our model while training i.e. training model has only input parameter values. The model by itself has to find which way it can learn.

Ex-

Customer ID	Gender	Age	Annual Income	Spending score (1-100)
1	M	19	15	39
2	F	21	15	81
3	F	20	16	6
4	F	23	16	77
5	F	31	17	40

Don't know what to predict?

- **Clustering** - Broadly this technique is applied to group data based on different patterns, our machine model finds. For example in above figure we are not given output parameters value, so this technique will be used to group clients based on the input parameters provided by our data.
- **Association** - This technique is a rule based ML techniques which finds out some very useful relations b/w parameters of a large dataset. For eg shopping stores use algorithms based on this technique to find out relationship b/w sale of one product wrt to others.

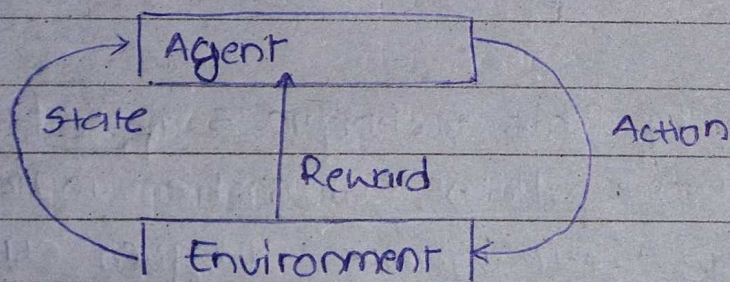


Date \_\_\_\_\_  
Page \_\_\_\_\_

sale based on customer behavior. Once trained well such models can be used to increase their sales by planning different offers

- **Semi Supervised learning** - we use this techniques when we are dealing with a data which is a little bit labelled and rest large portion of it is unlabelled. we can use unsupervised techniques to predict labels and then feed these labels to supervised techniques. This technique is mostly applicable in case of image data-sets where usually all images are not available.

**Reinforcement Learning -**



In this technique, model keeps on increasing its performance using a Reward Feedback to learn the behavior or pattern. These algorithms are specific to a particular problem eg. Google Self Driving Car, AlphaGo where a bot competes with human kind even itself to getting better and better performance of Go Game.



Date \_\_\_\_\_  
Page \_\_\_\_\_

Each time we feed in data, they learn and add the data to its knowledge that is training data. So, more it learns the better it get trained and hence experienced.

Agents observe input.

Agent performs an action by making some decisions. After its performance, agent receives reward and accordingly reinforce and the model stores in state-action pair of information.

### Data Vs Information Vs Knowledge

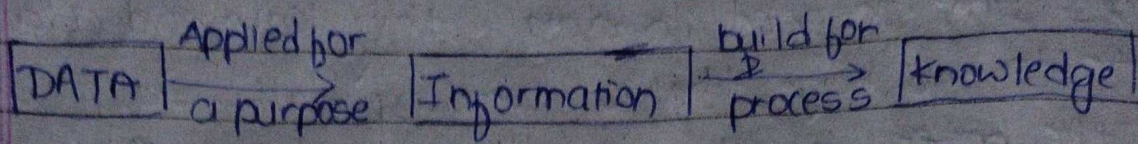
**Data** - Unprocessed facts and figures without any added interpretation or analysis.

"The price of the oranges is \$80 per kg."

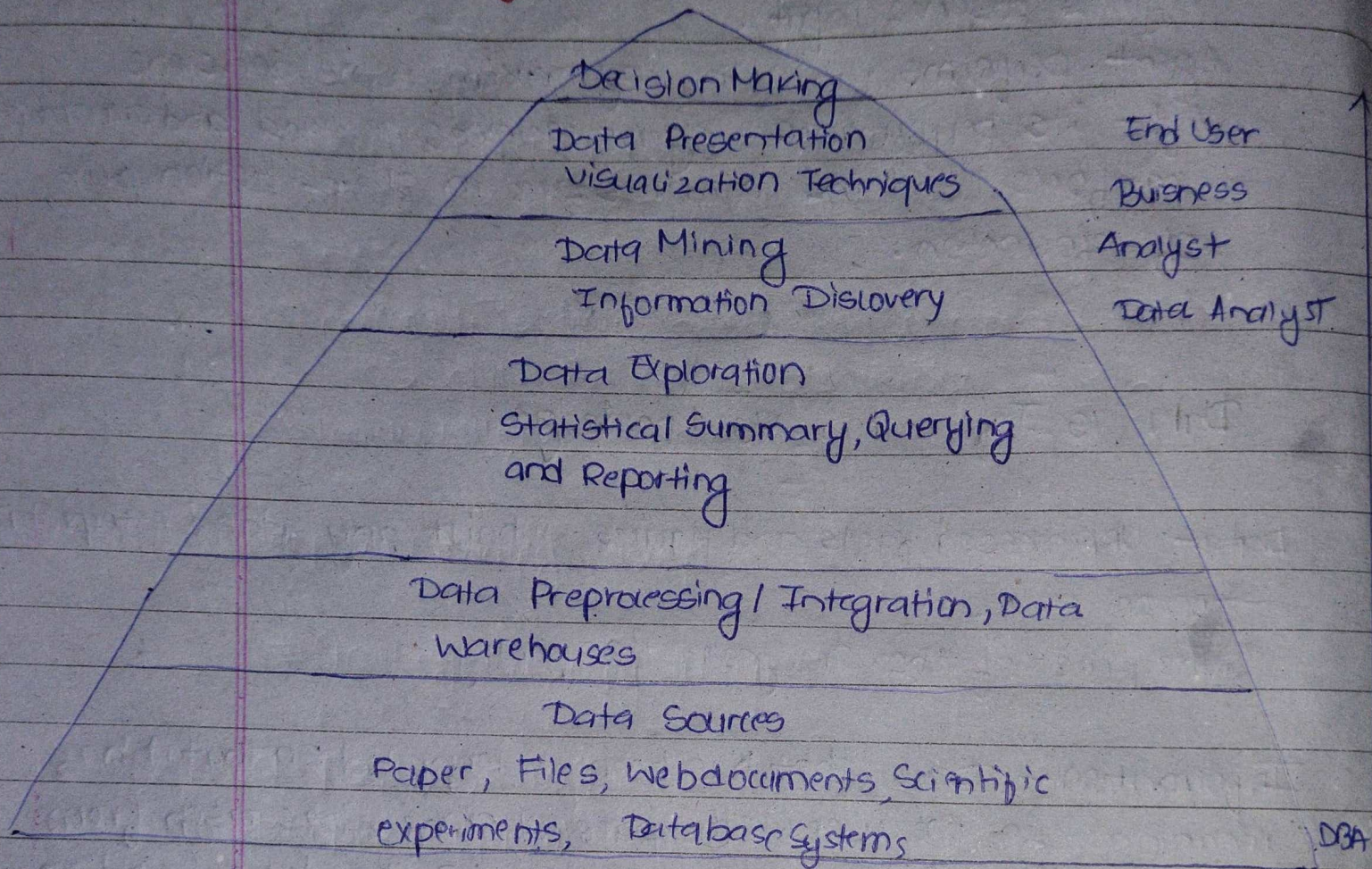
**Information** - Data that has been interpreted so that it has meaning for the user. "The price of oranges has risen from \$70 to \$80 per kg" gives meaning to the data and so is said to be information to someone who tracks orange prices.

**Knowledge** - Combination of information, experience and insight that may benefit the individual or the organisation. "When orange prices go up by \$10 per kg, it's likely that market growth will also be affected in some ways" is knowledge.





## DATA Mining and Business Intelligence



### Nominal Data (Categorical)

What is your gender?

- M - Male
- F - Female

What is your hair colour?

- Brown
- Black
- Blonde
- Gray
- Other



Nominal Scales are used for labelling variables, without any quantitative value. "Nominal" scales could simply be called "labels".

- In simple words it cannot be compared i.e. male < Female etc

### Ordinal Data:

How do you feel today?

- 1 - Very unhappy
- 2 - Unhappy
- 3 - OK
- 4 - Happy
- 5 - Very happy

It refers to quantities that have a neutral ordering. With ordinal data you cannot state with certainty whether the intervals b/w each value are equal.

### Interval Data:

Interval Data is like ordinal except we can say the intervals b/w each value are equally split. The most common example is temperature in degrees Fahrenheit. The difference b/w 29 and 30 degrees is the same magnitude as the difference b/w 78 & 79.

### Binary Data:

Data which can only be in two forms. Let's say yes or no, 0 or 1. It can be considered as a special case of nominal, ordinal or interval data. It is very



commonly used in case of classification tasks

Count:

A — 500

B — 1500

Useful for making bar graph

Time:

Time data is cyclic data which repeats continuously

This data can be in the daily, weekly, monthly and annually.

```
import numpy as np
```