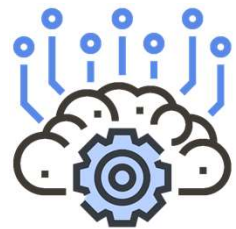# Introduction to AI and ML
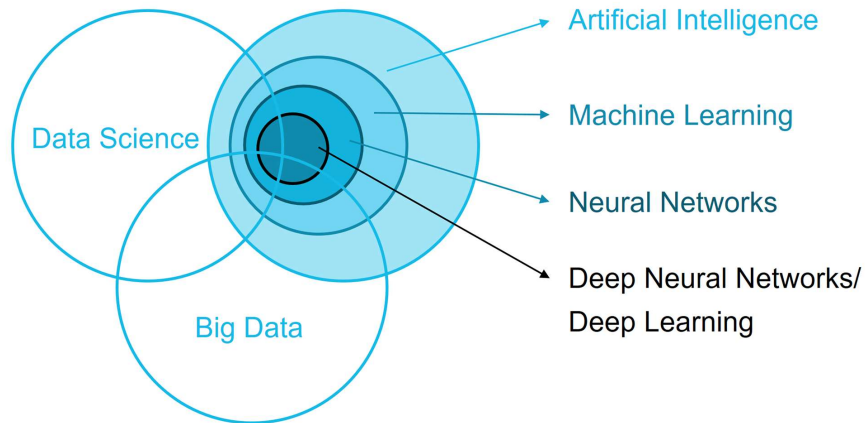
Dr. Shailesh Sivan

1

OVERVIEW

➢Confused?

➢What is machine learning?

➢Programming vs Learning

➢Machine Learning Pipeline

➢When to use Machine Learning?
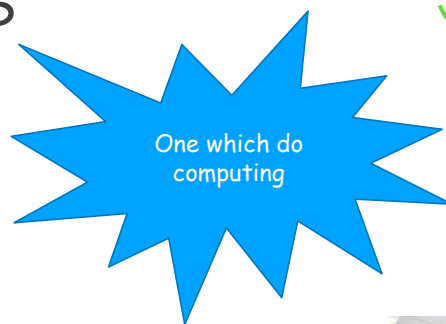
➢ML Application

➢Types of Learning

2

CONFUSED ?

Artificial Intelligence

Machine Learning

Neural Networks

Deep Neural Networks/
Deep Learning

Data Science

Big Data

3

3

COMPUTER SYSTEMS

- An electronics device

- Capable of storing and processing data

- Can be used for controlling other devices

One which do computing

4

ROBOTS AND COMPUTER

Are Computers Intelligent or Dumb ?

COMPUTERS ARE DUMB ☹ !

➢ Computer is not a magical device.

➢ It performs only those works which man can does

➢ But with very high speed and reliable accuracy.

➢ It has no any intelligence quality or thinking power

5

ARTIFICIAL INTELLIGENCE

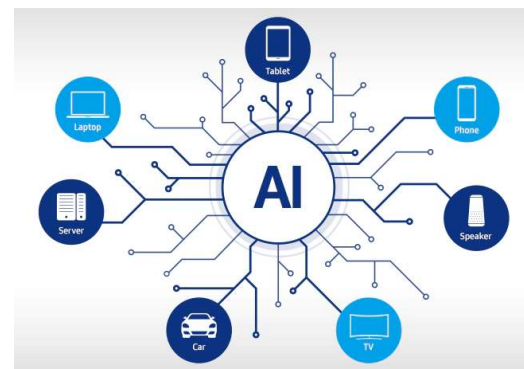**Can We make Computers Intelligent ?**
Yeah Of course !

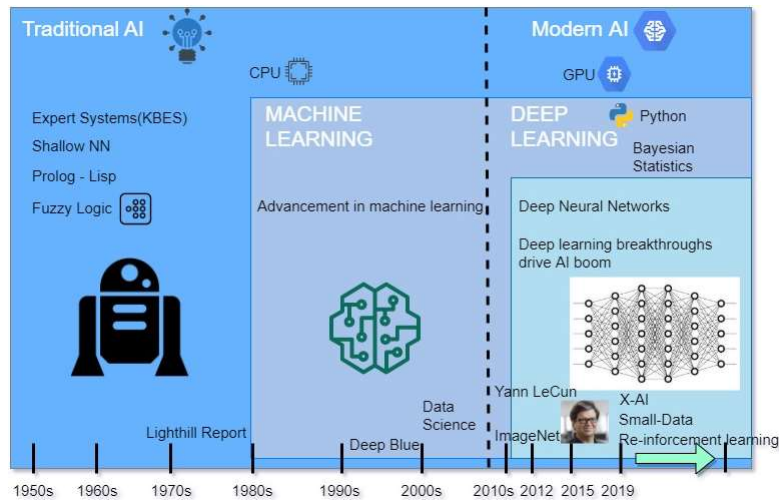\+ Knowledge = **Artificial Intelligence**

Artificial Intelligence is a scientific domain
that's deals with making intelligence to system

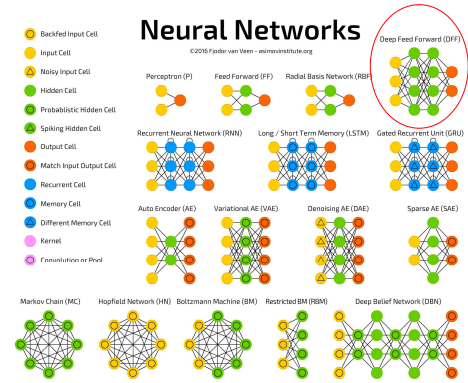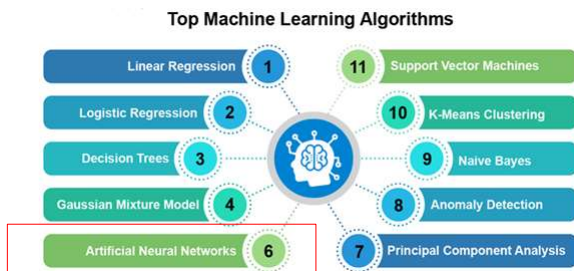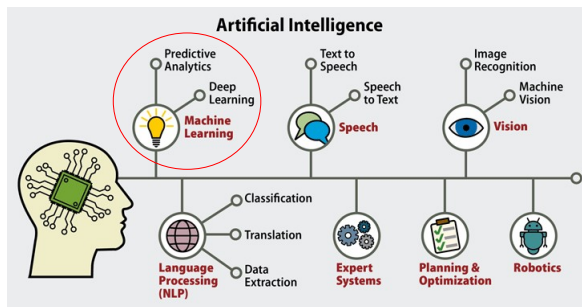To make a system "**think like a human**"

6

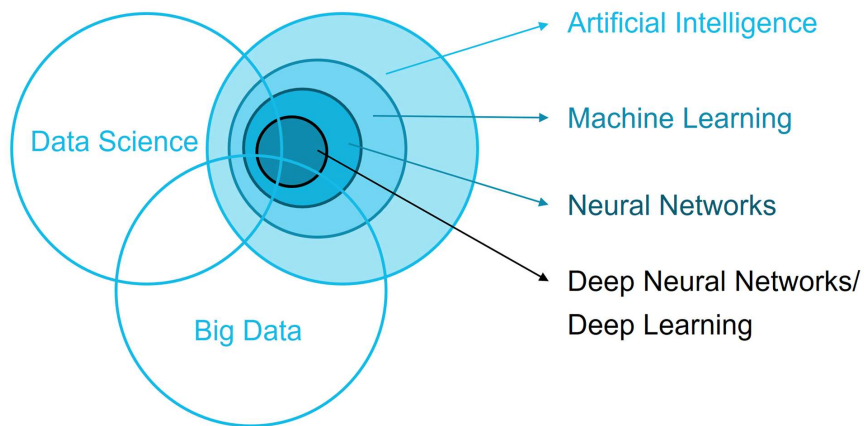# ARTIFICIAL INTELLIGENCE

7

ARTIFICIAL INTELLIGENCE

8

9

CONFUSED ?

- Artificial Intelligence
- Machine Learning
- Neural Networks
- Deep Neural Networks/ Deep Learning
- Data Science
- Big Data
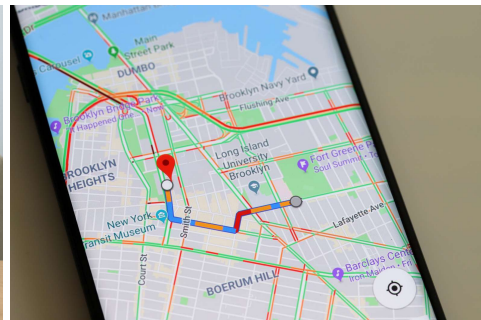
10

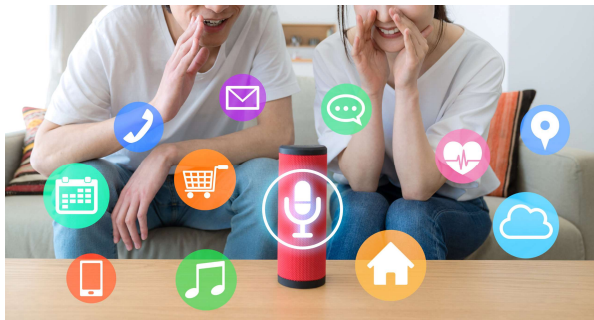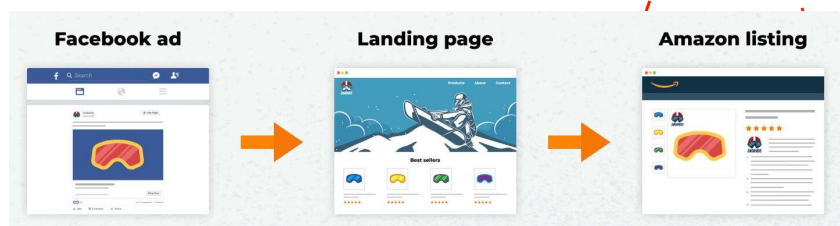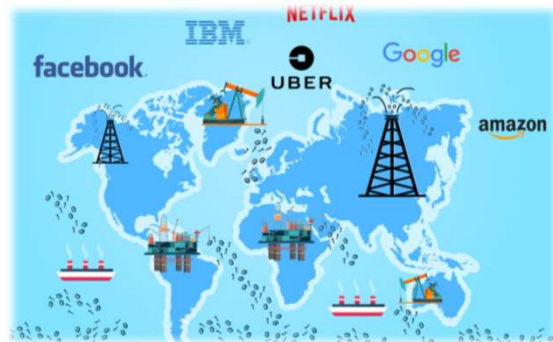10

DATA SCIENCE
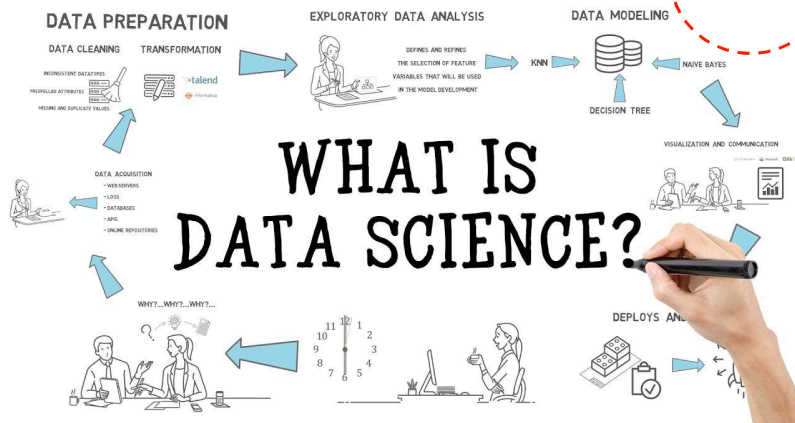
IS DATA IMPOTANT?

WHY DATA IS IMPORTANT ?

- Make Informed Decisions
- Get The Results You Want
- Find Solutions To Problems
- Back Up Your Arguments
- Be Strategic In Your Approaches
- Keep Track Of It All
- Know What You Are Doing Well
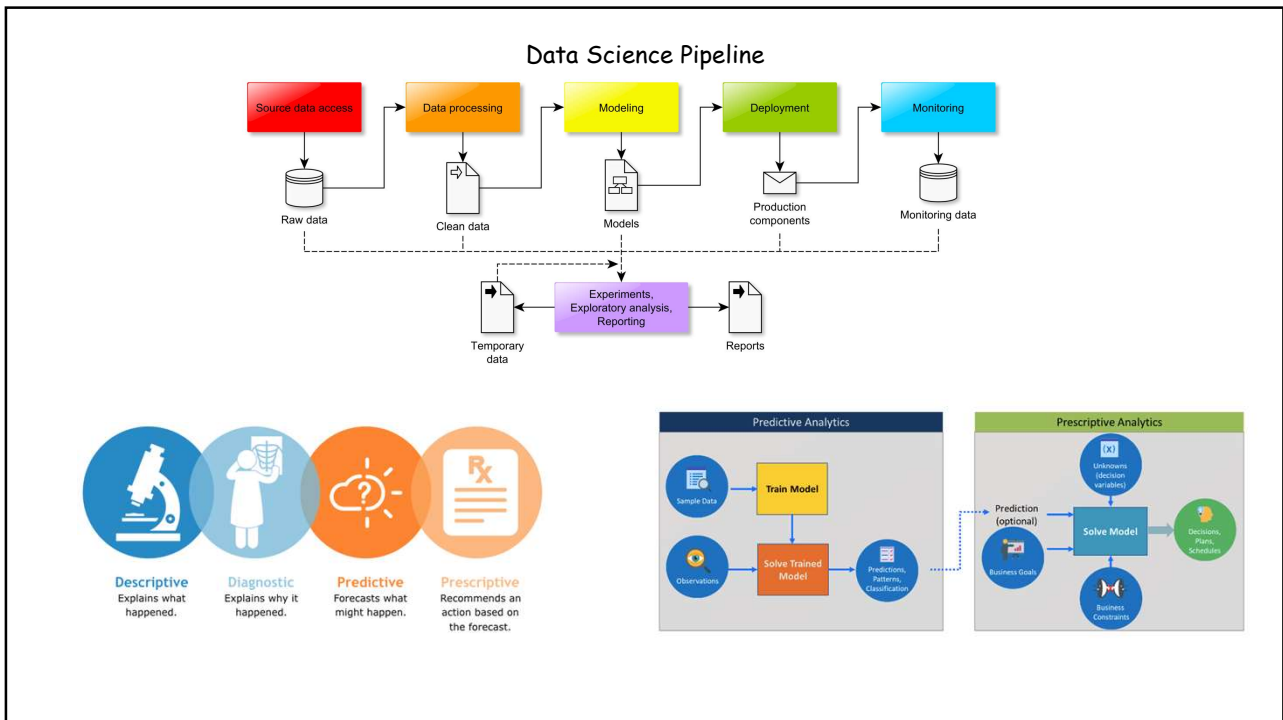
DATA IS NEW OIL

13

REALITY ABOUT DATA

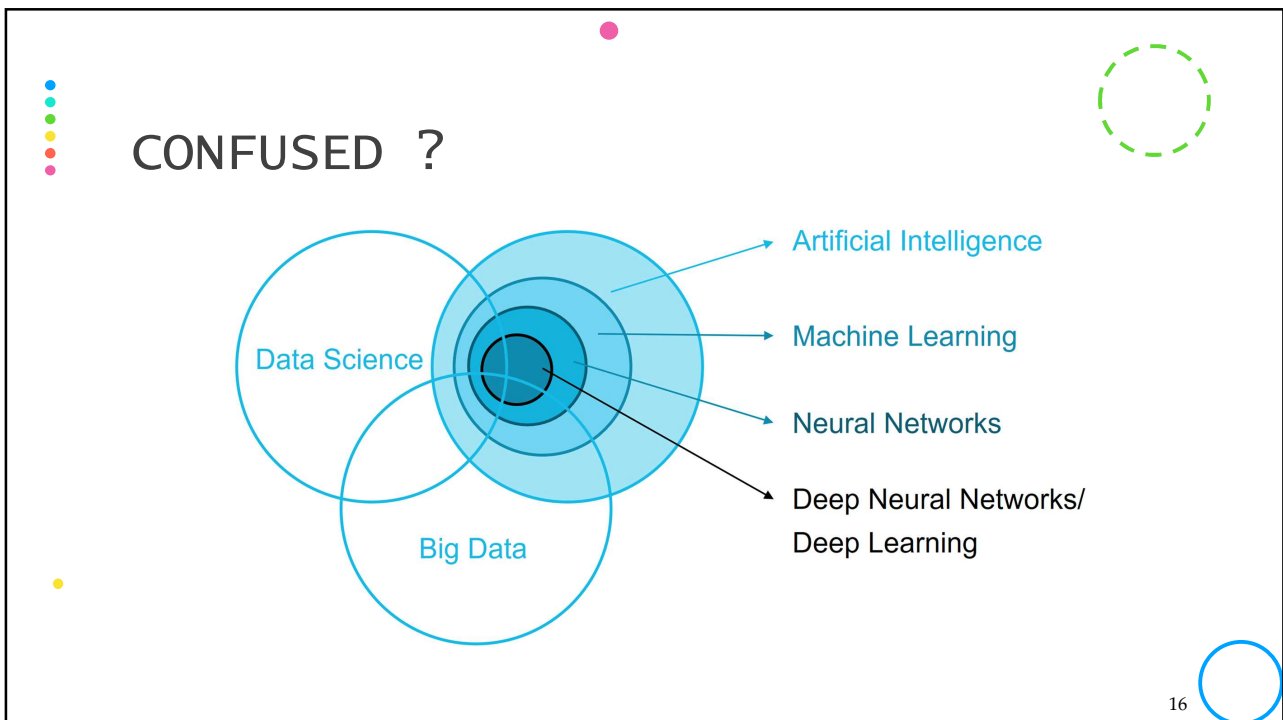**SOLUTION ?**

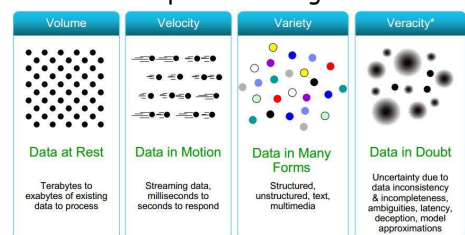We are now **drowing** in data ! but **starving** for insights ☹
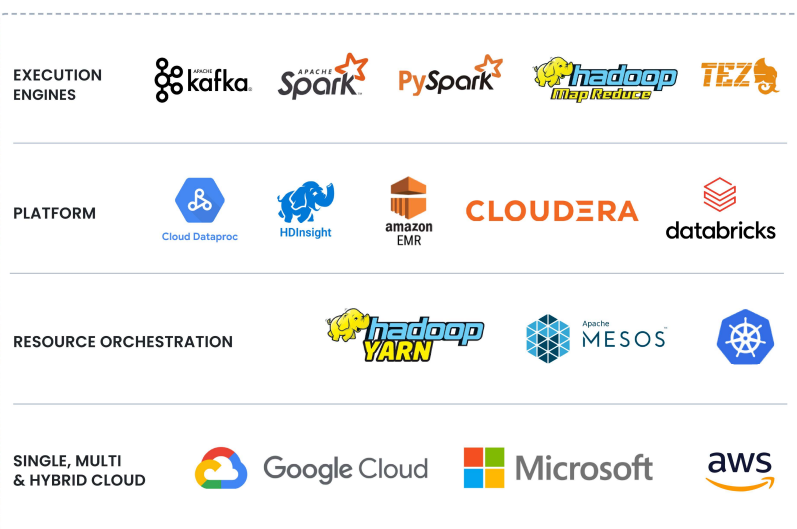
14

15

16

Big Data

Is Big Data = Large Data ?

Properties of Bigdata

17

Big Data

WHAT IS MACHINE LEARNING?

"Learning is any process by which a system
improves performance from experience."
- Herbert Simon

Definition by Tom Mitchell (1998):

Machine Learning is the study of algorithms that
- improve their performance P
- at some task T
- with experience E.
A well-defined learning task is given by <P,T,E>

19

19

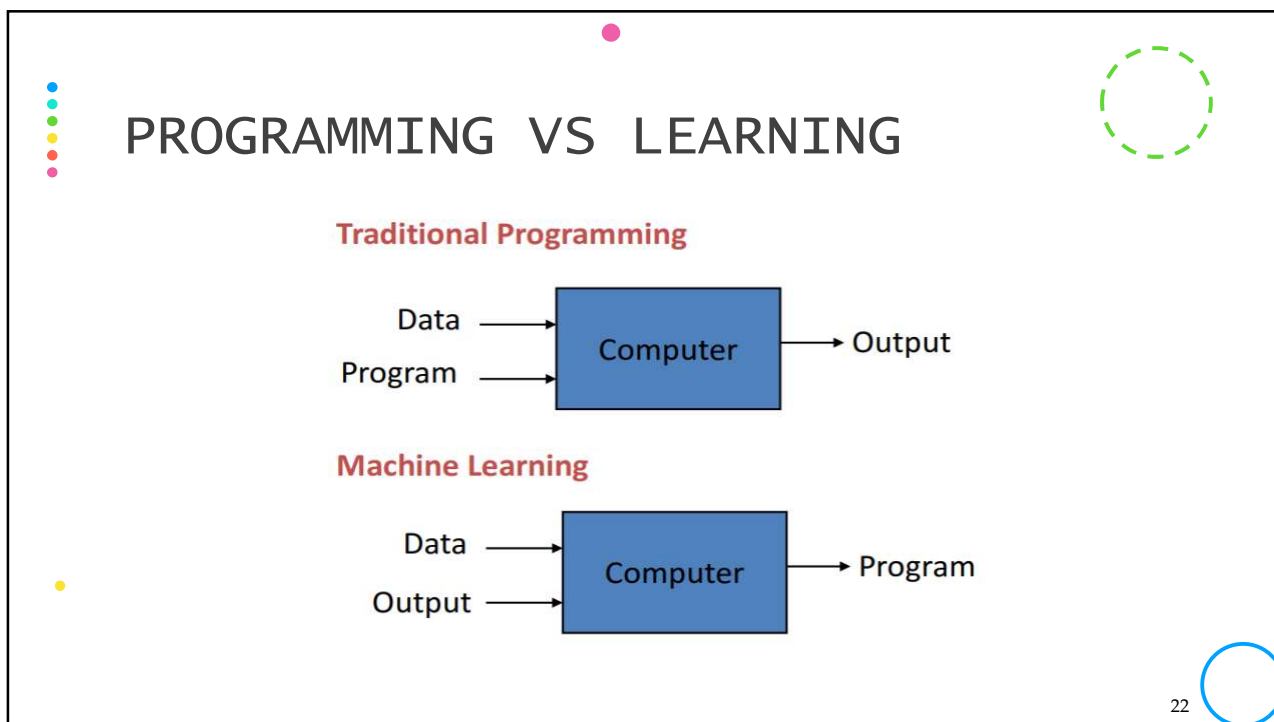This is a shirt we used to wear.

Color: Green
Size : Large
Type : Formal

Color: Green, red
Size : Large, small
Type : Formal, casual

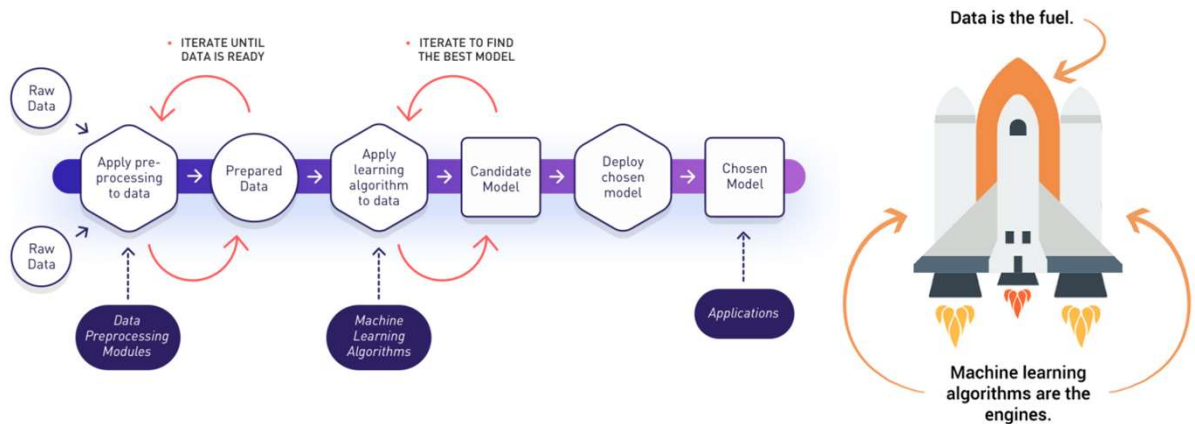Is this a shirt ?
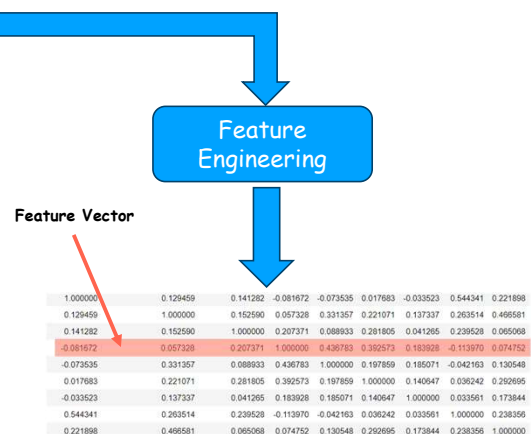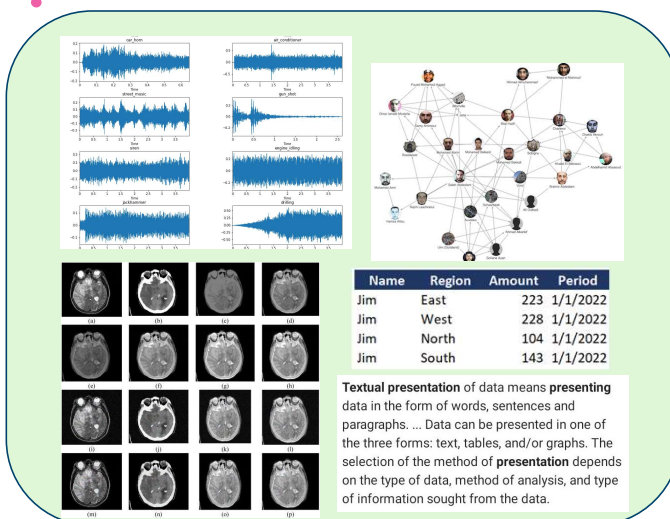
Ok

This is also a shirt

20

PROGRAMMING VS LEARNING

MACHINE LEARNING PIPELINE

DATA AS FEATURE VECTORS

WHEN TO MACHINE LEARNING?

ML is used when:

- Human expertise does not exist (navigating on Mars)

- Humans can't explain their expertise (speech recognition)

- Models must be customized (personalized medicine)

- Models are based on huge amounts of data (genomics)

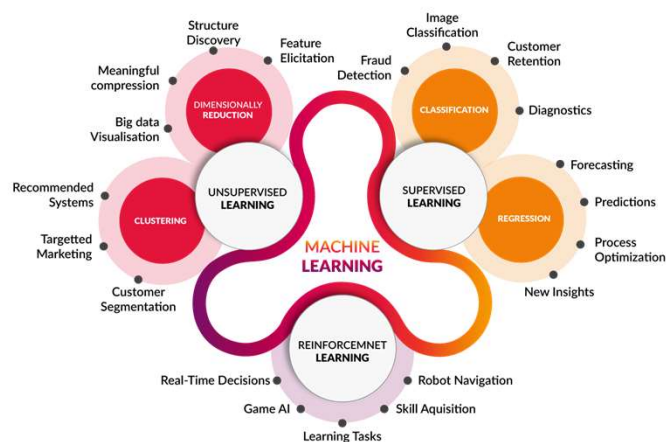Learning isn't always useful - There is no need to "learn" to calculate payroll

25

TYPES OF LEARNING
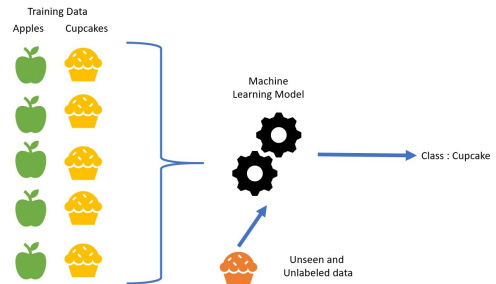

26

SUPERVISED LEARNING

Input : Labeled Data

$X\ (features)$	$Y\ (labels)$
$x_{11}, x_{12}, x_{13}, \ldots \ldots \ldots \ldots x_{1n}$	y_1
.	.
.	.
.	.
$x_{k1}, x_{k2}, x_{k3} \ldots \ldots \ldots \ldots x_{kn}$	y_k

Training Data
Apples   Cupcakes

Machine
Learning Model

Class : Cupcake

Unseen and
Unlabeled data

Goal : Construct a predictor $f : X \rightarrow Y$

to minimize the error between $\widehat{Y}, Y$
where, $\widehat{Y} = f(X)$

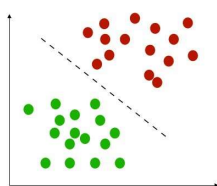Use : using predictor to predict $\widehat{y} = f(\widehat{x})$

For the unknown input $\widehat{x}$

27

27

SUPERVISED LEARNING

Regression
What is the temperature going to be tomorrow?

PREDICTION
84°

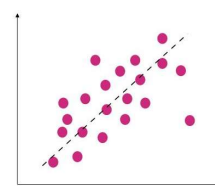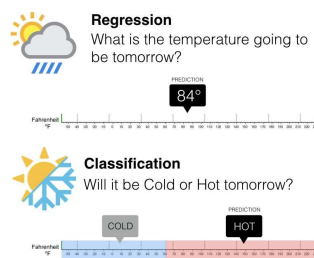**Classification**
Will it be Cold or Hot tomorrow?

COLD   PREDICTION HOT

Classification

- features and discrete labels

- maps an input to discrete label(class)

- Eg: spam or not, type of cancer

Regression

- features and continues real values

- Predict a real value for an input

- Eg: gold price, temperature
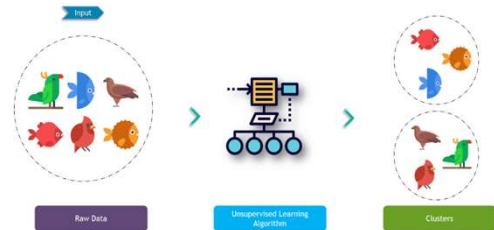
28

28

UNSUPERVISED LEARNING

Input : Unlabeled Data

OUTLIERS

X $(features)$
$x_{11}, x_{12}, x_{13}, \ldots \ldots \ldots \ldots x_{1n}$
.
.
.
$x_{k1}, x_{k2}, x_{k3} \ldots \ldots \ldots \ldots x_{kn}$

hidden

Goal : Construct a analyzer to find the relationship between inputs

$$x_1, x_2, x_3, \ldots, x_k$$

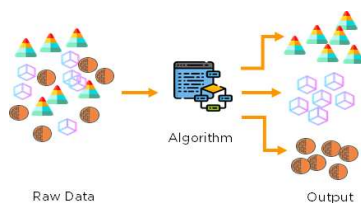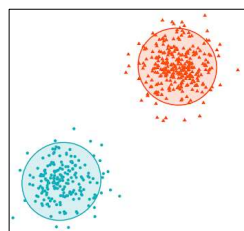Use : Group or associate inputs according to their similarity
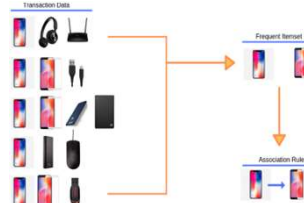
29

UNSUPERVISED LEARNING

Clustering

Algorithm

Raw Data

Output

ID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}
...	...

market basket transactions

{Diapers, Beer}   Example of a frequent itemset
{Diapers} → {Beer}   Example of an association rule

Transaction Data

Frequent Itemset
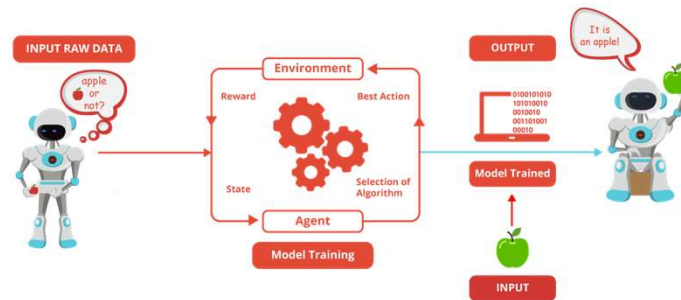
Association Rule

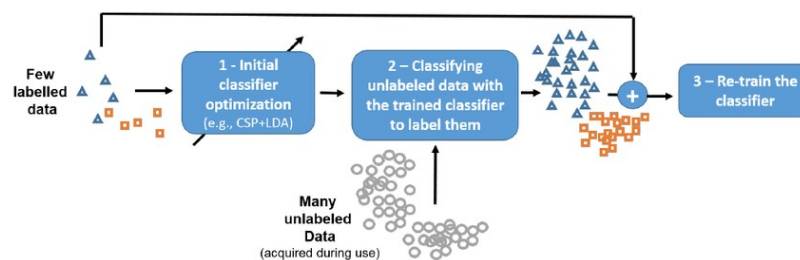**Itemset and rule mining**

30

REINFORCEMENT LEARNING

- Use software agents
- Based on rewards
- Objective is to maximize rewards for better learning

31

SEMI SUPERVISED LEARNING

In a nutshell, semi-supervised learning (SSL) is a machine learning technique that uses a small portion of labeled data and lots of unlabeled data to train a predictive model

32

Dr. Shailesh Sivan

+91 8907230664

shaileshsivan@cusat.ac.in

https://shaileshsivan.info

33

34