

# Machine Learning - Overview

**DR DV Ramana**  
**Academic Advisor**  
**And Program Manager**

# Installation steps

To install Jupyter using Anaconda, just go through the following instructions

Download Anaconda. We recommend downloading Anaconda's latest Python 3 version (currently Python 3.7).

<https://www.anaconda.com/products/individual>

## Anaconda Installers



Python 3.7

64-Bit Graphical Installer (466 MB)

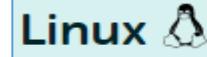
32-Bit Graphical Installer (423 MB)



Python 3.7

64-Bit Graphical Installer (442 MB)

64-Bit Command Line Installer (430 MB)



Python 3.7

64-Bit (x86) Installer (522 MB)

64-Bit (Power8 and Power9) Installer (276 MB)

Install the version of Anaconda which you downloaded, following the instructions on the download page.

Steps:

Visit [Anaconda.com/downloads](https://www.anaconda.com/downloads)

Select Windows

Download the .exe installer

Open and run the .exe installer

# Agenda - Theory

What is machine learning?

How Machine Learning Works

Algorithm types of Machine learning

Supervised and Unsupervised Learning

Uses of Machine learning

Evaluating ML techniques

# Agenda -LAB

1. Explore Datasets Online (can refer Kaggle, UCI ML, etc.)

2. Load dataset in google colab

3. Explore Datasets Online (can refer Kaggle, UCI ML, etc.)

4. Print first five values and last five values in dataset

5. Check Correlation Between fields present in dataset

## What is Data

Data is any set of characters that is gathered and translated for some purpose, usually analysis

Data can be defined as a representation of facts, concepts, or instructions in a formalized manner, which should be suitable for communication, interpretation, or processing by human or electronic machine

Facts and statistics collected together for reference or analysis

The quantities, characters, or symbols on which operations are performed by a computer, being stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media

Things known or assumed as facts, making the basis of reasoning or calculation.

Data is a collection of raw, unorganized facts and details like text, observations, figures, symbols and description of things etc.

Data does not carry any specific purpose and has no significance by itself

Data is measured in terms of bits and bytes – which are basic units of information in the context of computer storage and processing

# What is Science

Science is the pursuit and application of knowledge and understanding of the natural and social world following a systematic methodology based on evidence

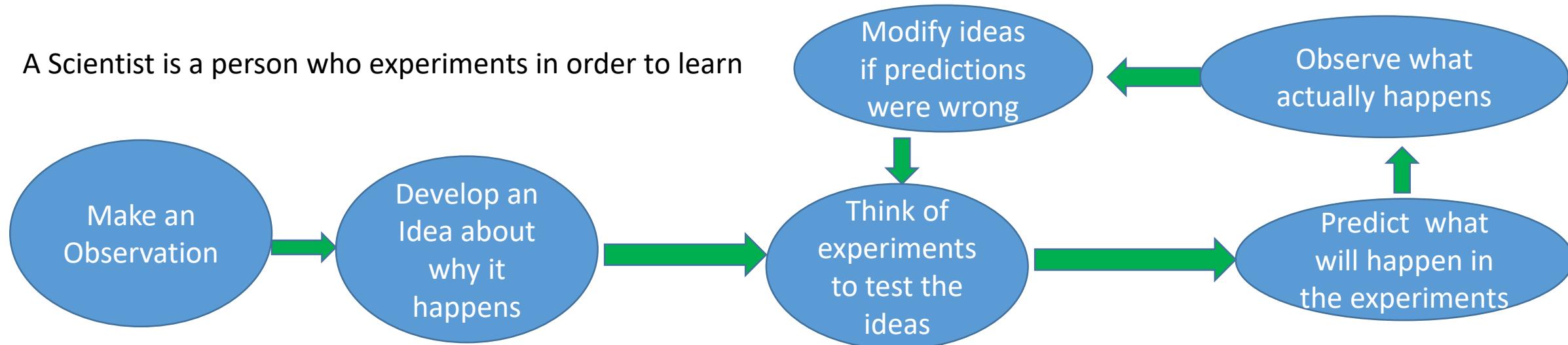
Science can be thought of as both a body of knowledge (the things we have already discovered), and the process of acquiring new knowledge (through observation and experimentation—testing and hypothesizing).

Science is a method (or tool) enabling the discovery of new observations, by the collation of data, development of models, and subsequent use of those models for prediction

Models themselves undergo change over time

Science is the knowledge gained from using observations and experiments to describe and explain the world around us

A Scientist is a person who experiments in order to learn

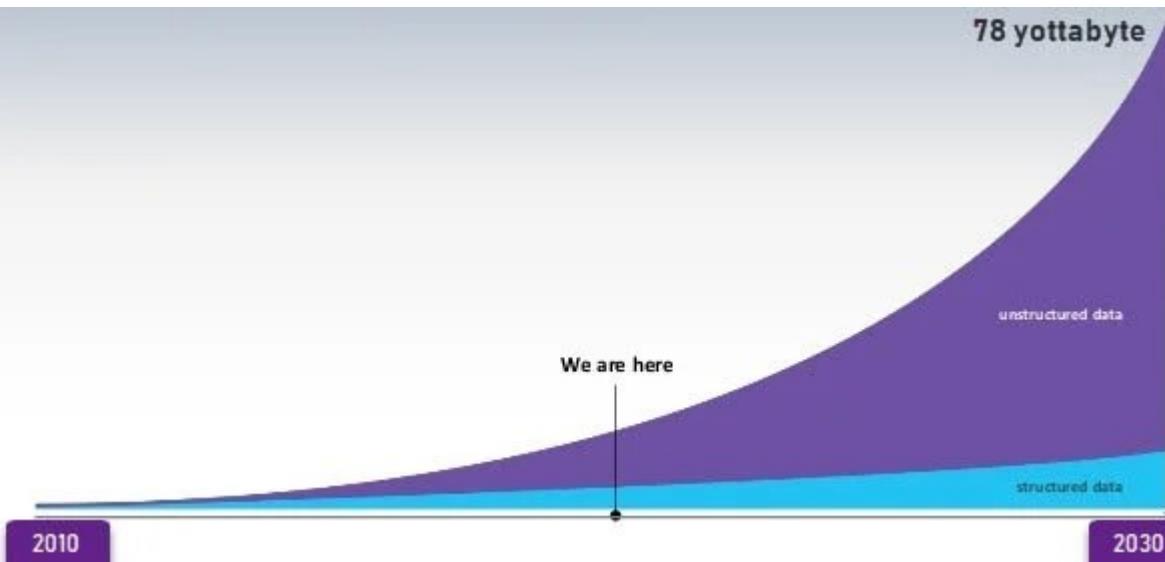


## Big Data

Big Data is a collection of data that is huge in volume, yet growing exponentially with time

Big Data is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently

Big data is also a data but with huge size



Big Data is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently

Big Data is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently

Big Data is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently

## DARK Data

Dark data is data which is acquired through various computer network operations but not used in any manner to derive insights or for decision making

The ability of an organization to collect data can exceed the throughput at which it can analyze the data

Dark data makes up most of the information companies collect

Unstructured data

videos

Images

E-mails and other points that can't be inserted into a spreadsheet – falls into this category

but businesses also collect a lot of information that for one reason or another never gets analyzed

Dark data is data which is acquired through various computer network operations but not used in any manner to derive insights or for decision making

The ability of an organization to collect data can exceed the throughput at which it can analyze the data

## DARK Data

All the information that a person can dig out on Google or Bing search constitutes only 5% of the total data

The rest of the 95 % of the data, the unindexed data exists in the deep web

Financial records

Medical and Legal documents

Government and Organization specific data repositories are some examples

Data like pictures and videos add to this data, it becomes a complex amalgamation of unstructured data

The problem goes beyond data visibility as more and more companies are into putting their data behind the firewall

Unstructured data presents a variety of challenges pertaining to

Data management

SLAs around data discovery and classification, and

the lack of tools and infrastructure to process the data

## DARK Data

There is a certain inertia in the business management towards exploring dark data as

It can jeopardize the normal business process putting the team at unease which is used to working with traditional data.

With companies harboring their data on public cloud platforms like

Amazon

Google, and Microsoft, and

service offerings such

as PaaS and SaaS, the journey towards dark data analytics is going to get smooth

## DARK Data

Dark data is data which is acquired through various computer network operations but not used in any manner to derive insights or for decision making

Dark Data refers to

"Any information assets that organizations collect, process, and store during regular business activities but generally fail to use for other purposes"

Dark Data is all the unknown and untapped data across an organization, generated by systems, devices and interactions

Example

we might have the number of sales we are making. Also, we may have the number of visits the customer has made to a store. Or the number of times he/she has visited our online store to make a particular purchase. But this information about the number of visits will be gone unprocessed.

## DARK Data

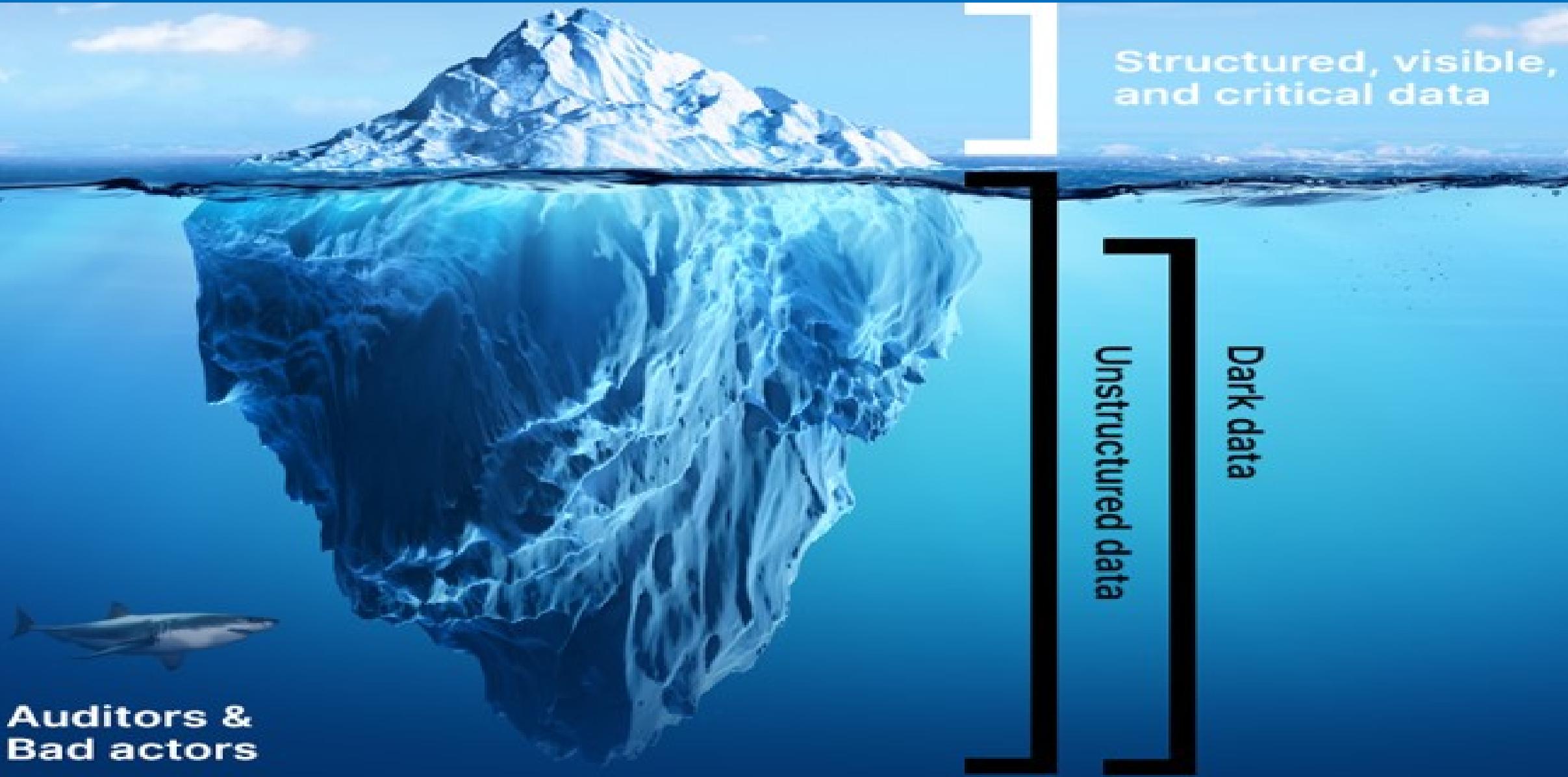
Dark data is data which is acquired through various computer network operations but not used in any manner to derive insights or for decision making

Dark Data refers to

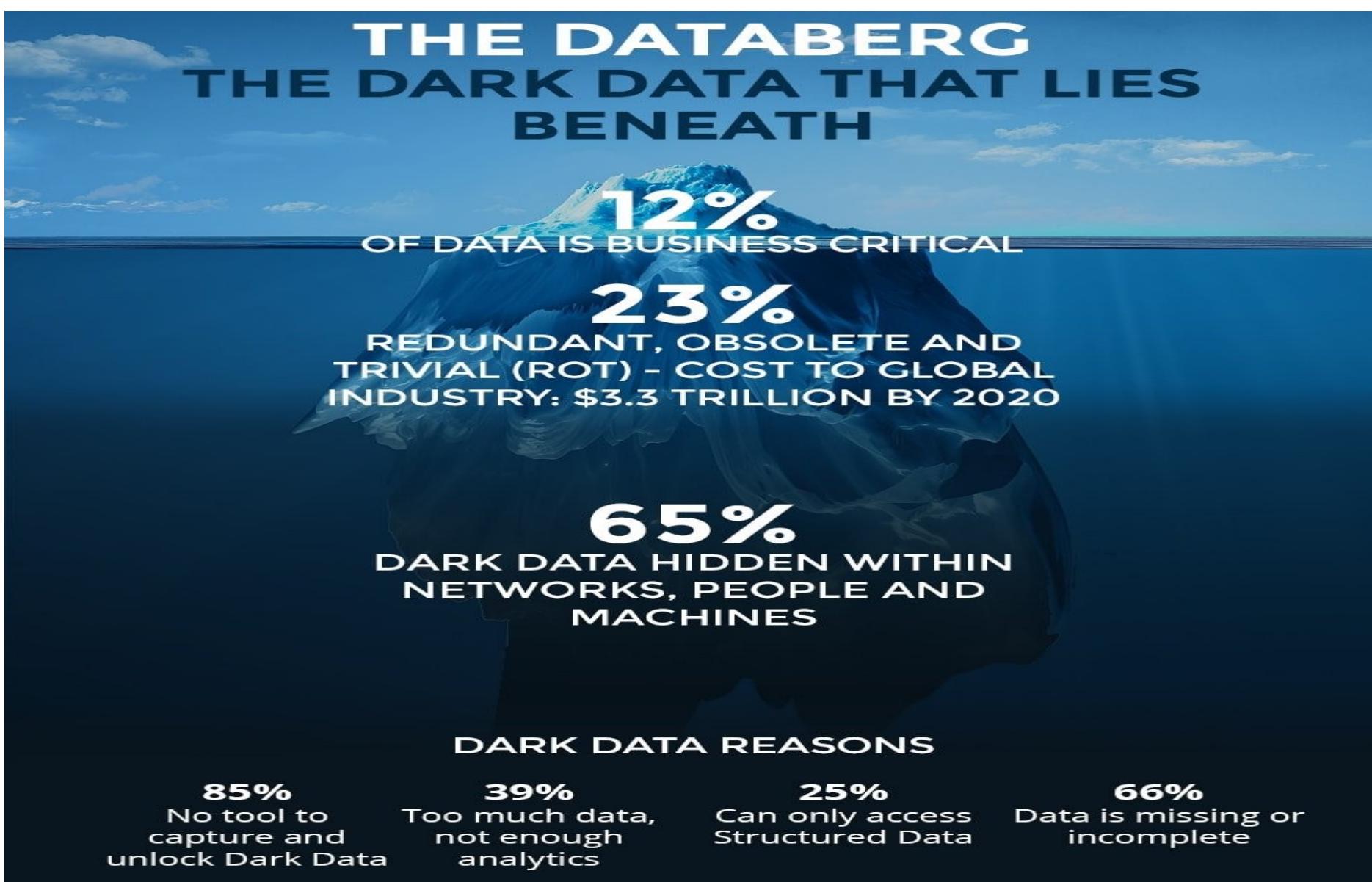
"Any information assets that organizations collect, process, and store during regular business activities but generally fail to use for other purposes"

Dark Data is all the unknown and untapped data across an organization, generated by systems, devices and interactions

# DARK Data



## DARK Data



## Dark Data Analytics

Dark Data analytics is a software that helps the organizations to better locate, tap and identify their Dark Data

Dark Data analyzed here can be used for critical business making.

Dark Data is all the unknown and untapped data across an organization, generated by systems, devices and interactions

Organized Dark Data analytics overcome the limitations and risks caused by unorganized Dark Data

Dark Data enables the less technical employees to understand the needs and requirements of their organization

Dark Data analytics solution helps in providing a

Comprehensive, Insightful and Accurate understanding of users' data to the organizations

Dark Data analytics helps in creating a big picture of their environment.

## Dark Data Analytics

To fuel AI powered solutions.

To increase data wealth stored

To produce deeper and more accurate insights

To create more productive and innovative business strategies

Helps the organization in building management and leadership processes

To improve quality assurance

To detect and correct the process errors

To secure vulnerabilities

For potential compliance violations

To create new data management skills

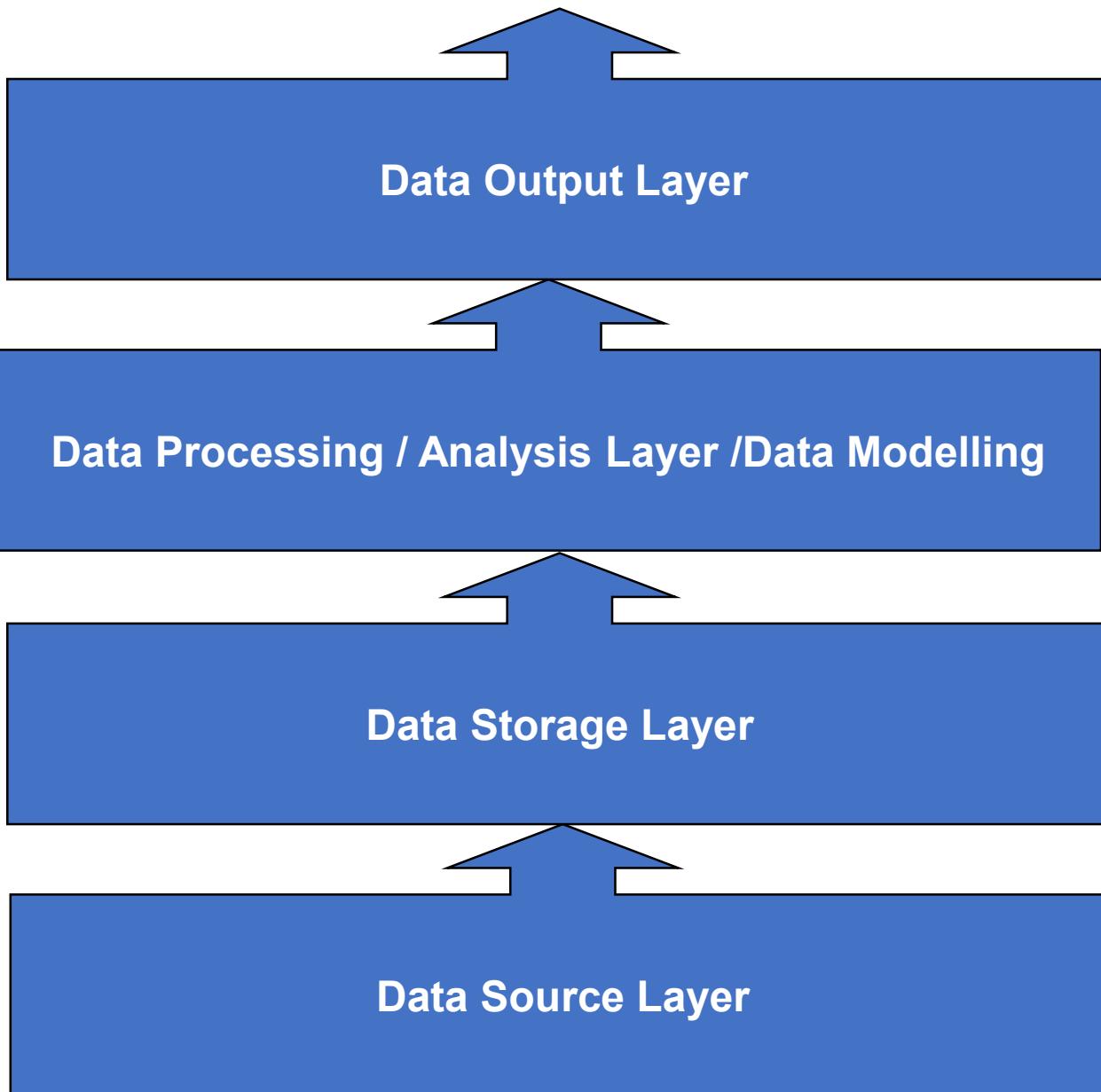
For growing rapid technologies, such as IoT

For long- and short-term analysis

To access quantifiable results

To improve the healthcare facilities

# Simplest and most accurate breakdown



# Data Sources Layer

Data arrives at your organization

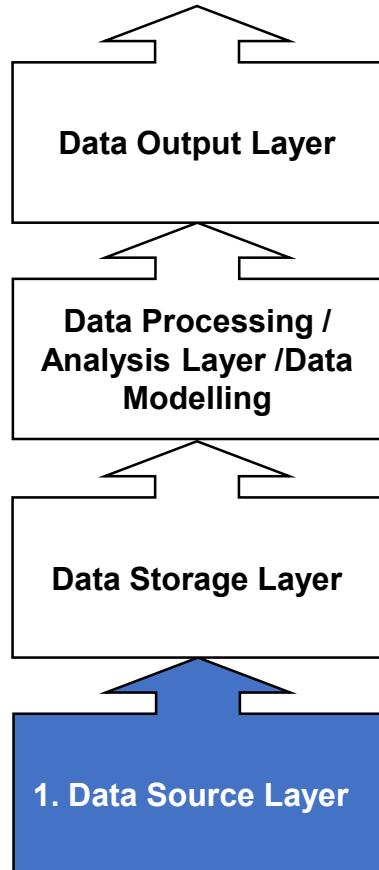
It includes everything from  
Sales records,  
Customer database,  
Feedback,  
Social media channels,  
Marketing list

Email archives and any data gleaned from monitoring or measuring aspects of your operations.

One of the first steps in setting up a data strategy is

- Assessing what you have here, and
- Measuring it against what you need to answer the critical questions you want help

You might have everything you need already, or you might need to establish new sources.



# Data Storage Layer

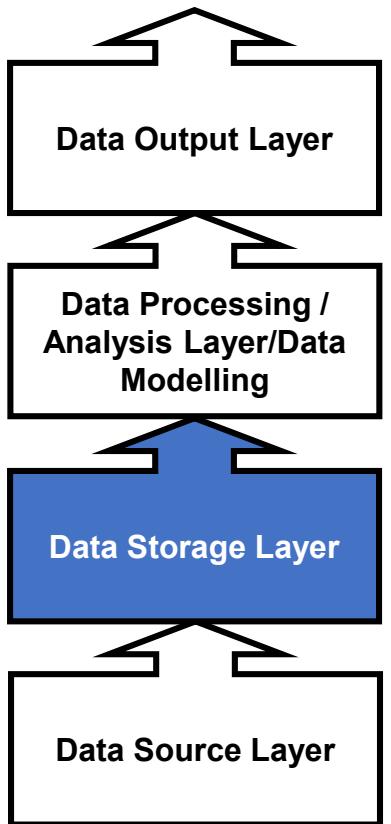
Big Data lives, once it is gathered from your sources.

As the volume of data generated and stored by companies has started to explode, sophisticated but accessible systems and tools have been developed – such as

Apache Hadoop DFS (distributed file system), or Google File System, to help with this task.

Storing data that of the computer system will understand (the file system) and need a system for organizing and categorizing it in a way that people will understand – the database.

Hadoop has its own, known as HBase, but others including Amazon's DynamoDB, MongoDB and Cassandra (used by Facebook), all based on the NoSQL architecture, are popular too.



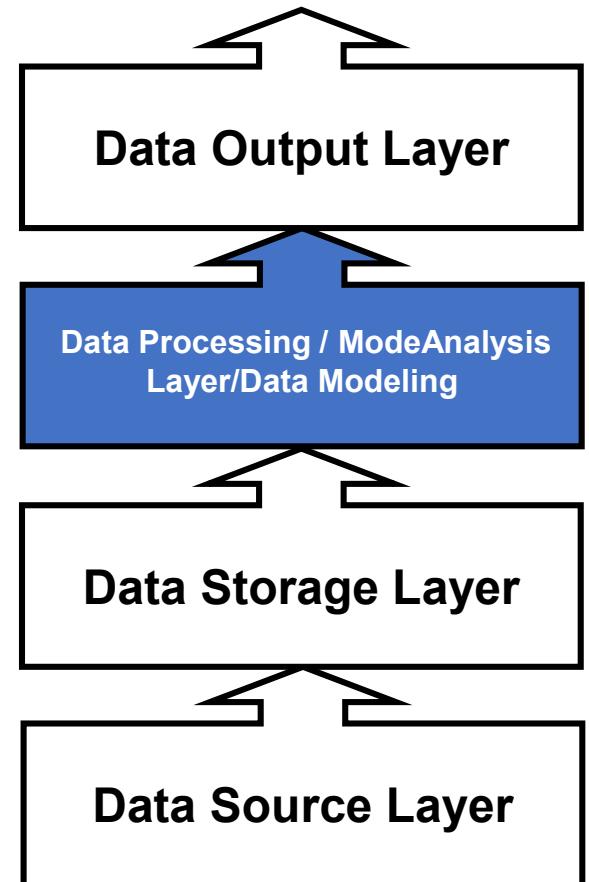
# Data processing/ analysis layer

When you want to use the data you have stored to find out something useful, you will need to process and analyse it.

A common method is by using a MapReduce tool. Essentially, this is used to select the elements of the data that you want to analyze, and putting it into a format from which insights can be cleaned.

If you are a large organization which has invested in its own data analytics team, they will form a part of this layer, too.

Tools such as Apache PIG or HIVE to query the data, and might use automated pattern recognition tools to determine trends, as well as drawing their conclusions from manual analysis.

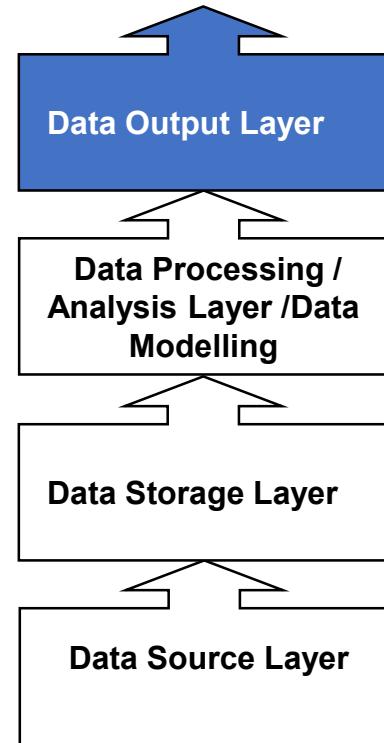


# Data output layer

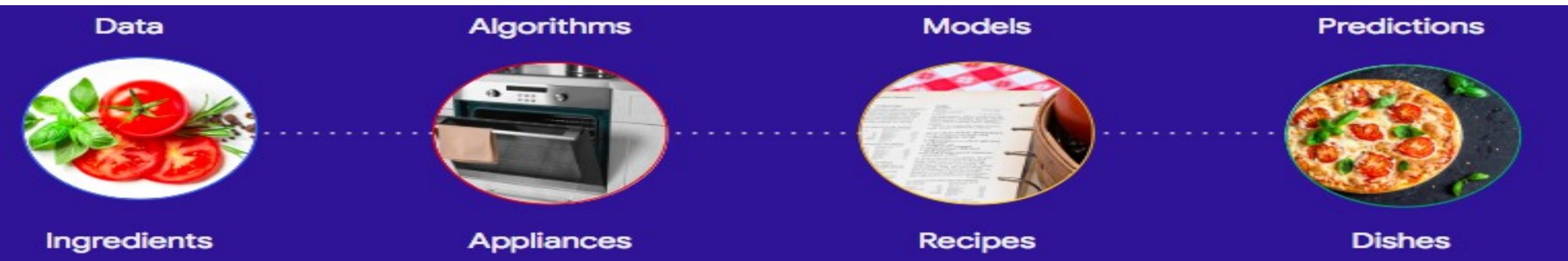
Insights gleaned through the analysis is passed on to the people who can take action to benefit from them.

Clear and concise communication (particularly if your decision-makers don't have a background in statistics) is essential, and this output can take the form of reports, charts, figures and key recommendations.

Big Data system's main task is to show, at this stage of the process, how measurable improvement in at least one KPI that can be achieved by taking action based on the analysis you have carried out



# Data science Skills, Scenario



Open an industrial-scale pizzeria that makes innovative pizzas, you need the big team or you need to partner with providers/consultants.

If you want to make a unique pizza or two this weekend—caramelized anchovy surprise, anyone?—then you still need to think about all the components we mentioned.

You're going to decide what to make (role 1)

Which ingredients to use (roles 2 and 3)

Where to get ingredients (role 0)

In charge of documentation and Training (role 6)

How to customize the recipe (role 5)

How to give it a taste test (role 4) before serving someone you want to impress, but for the casual version with less at stake, you can do it all on your own.

if your goal is just to make standard traditional pizza, you don't even need all that: get hold of someone else's tried and tested [recipe](#) (no need to reinvent your own) along with [ingredients](#) and start cooking!

# Machine Learning

Data science is the field of study that combines

Domain expertise

Programming skills, and

Knowledge of mathematics and statistics to extract meaningful insights from data

Data scientist is a professional responsible for

Data Collecting

Data Analyzing, and

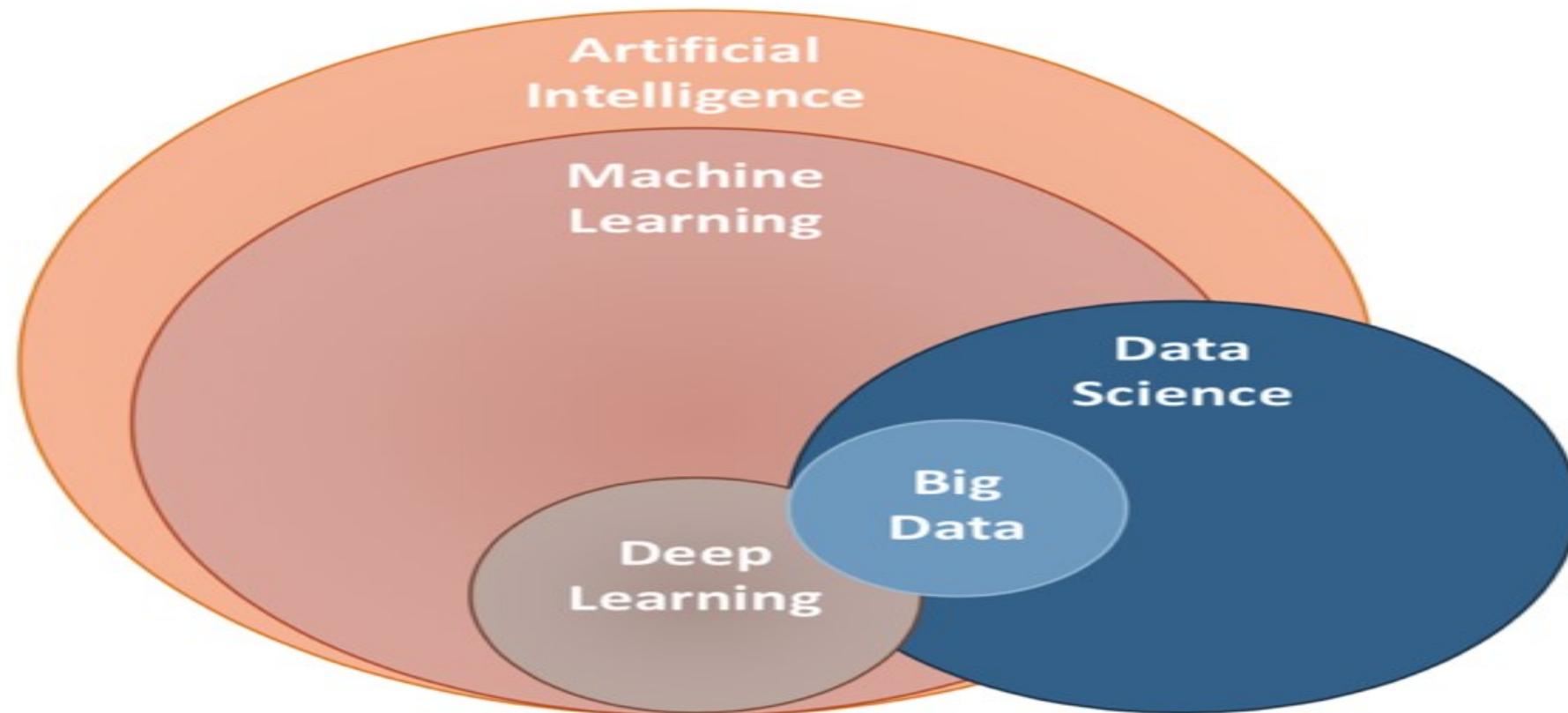
Interpreting extremely large amounts of data.

Big Data is larger, more complex data sets, especially from new data sources

# Machine Learning

Machine learning is a subfield of artificial intelligence, which is broadly defined as the capability of a machine to imitate intelligent human behavior

Machine Learning (ML) is a sub-field of artificial intelligence (AI) that focuses on building applications that can automatically and periodically learn and improve from experience or through gathered data and help in solving practical problems.



# Machine Learning

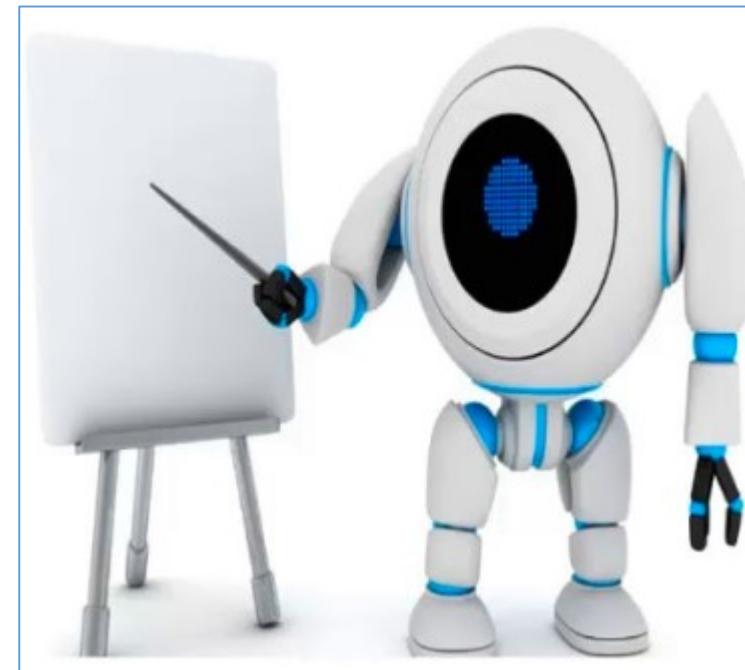
Machine learning is the most dynamically developing field of data science today due to a number of recent theoretical and technological breakthroughs.

Machine learning is the most dynamically developing field of data science today due to a number of recent theoretical and technological breakthroughs.

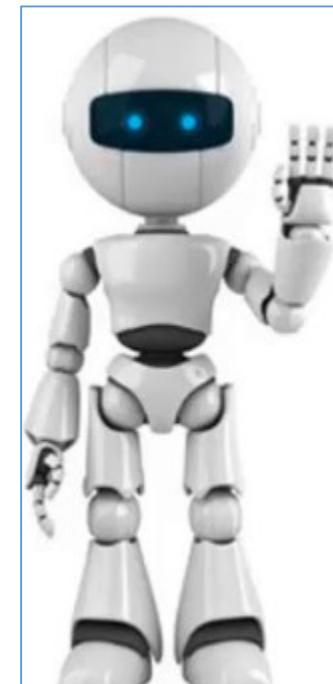
Learn from Experience



Learn from Experience Data



Follow Instructions



# How Machine Learning Works

Machine Learning process starts with gathering the data

The data is a collection of pairs (input, output).

Inputs could be anything,

Example

Email messages

Images, or

Sensor readings.

Outputs are usually real numbers, or labels

Example

"spam", "not\_spam",

"true", "false",

"eyes", "nose" etc .

# How Machine Learning Works



# Machine Learning

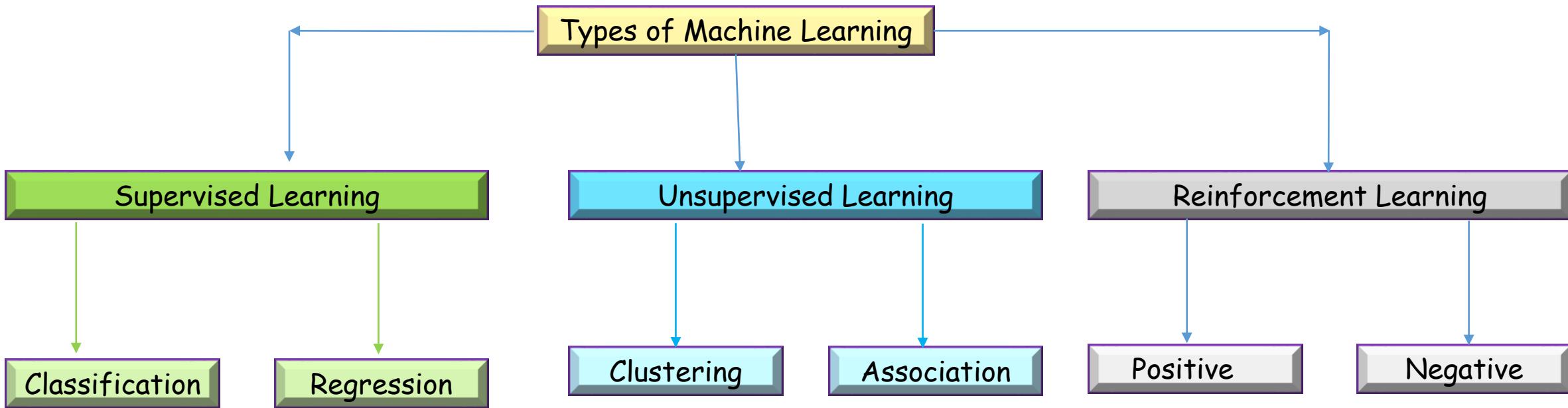
Class of Tasks

Performance of Measure (P)

Well Defined Experience

Inductive Learning

# Machine Learning



# How Machine Learning Works

## Supervised Learning

Trained data is given to the machine having input and output variables, which is based on the characteristics and data sets

### Example

E-mails converted into feature vectors and with each feature vector (used to represent characteristics) having label spam or not spam is given to the machine as training data in order to prepare the machine to make predictions with help of provided data or new input data.

# Machine Learning

## Supervised Learning

### Classification

Classification problem is when the output variable is a category

#### Example

"Spam" or "Not\_Spam" (in case of email) or "green" or "blue" (in case of balls)

### Regression

Regression problem is when the output variable is a real value or numeric outcome

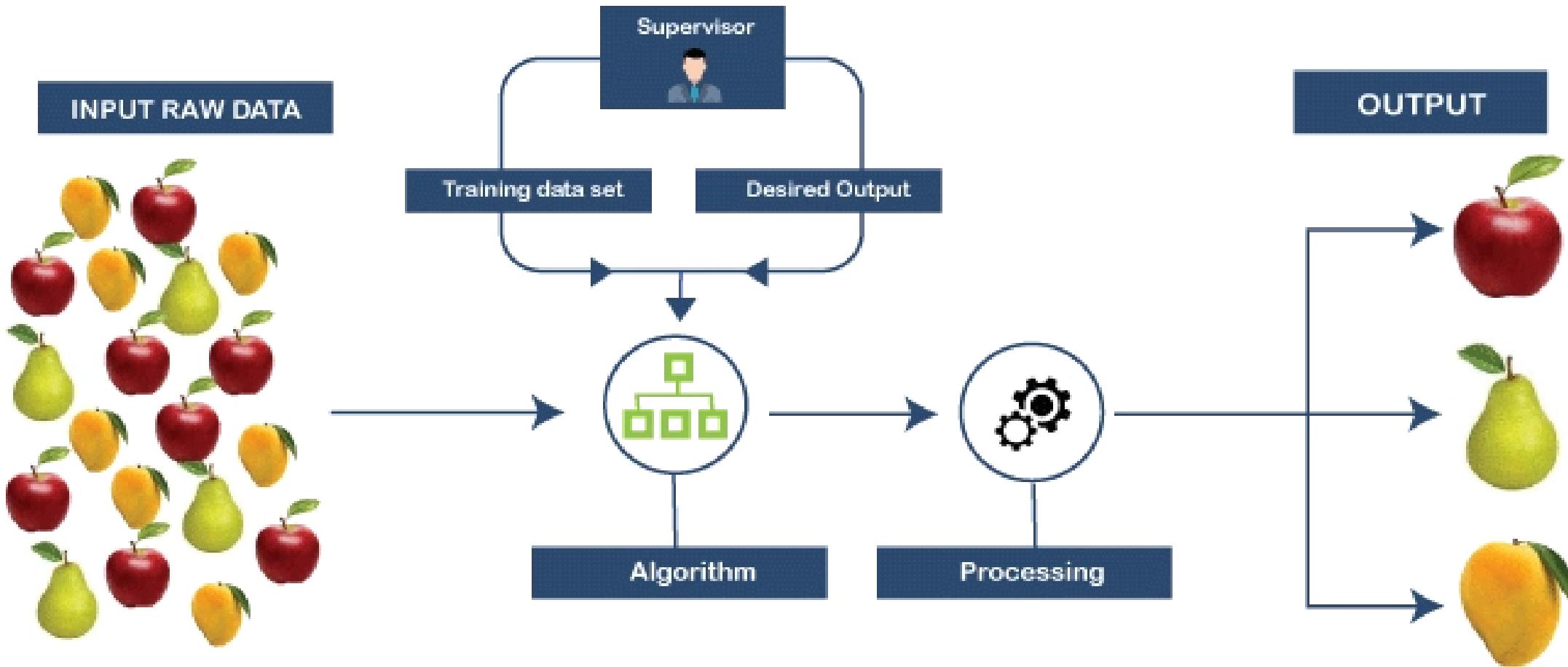
#### Example

House price prediction for a given feature like price, size, and location.

# Machine Learning

## Supervised Learning

### SUPERVISED LEARNING



# Machine Learning

## How Supervised Learning Works

### Supervised Learning



Suppose you had a basket and it is filled with some different kinds of fruits, your task is to arrange them as groups.

For understanding let me clear the names of the fruits in our basket

We have four types of fruits. They are: apple, banana, grape and cherry.

### Supervised Learning

You already learn from your previous work about the physical characters of fruits

So arranging the same type of fruits at one place is easy now.

Your previous work is called as **training data** in data mining.

Already learn the things from your train data, this is because of **response variable**.

Response variable mean just a **decision variable**.

You can observe response variable below (**FRUIT NAME**) .

# Machine Learning

## How Supervised Learning Works

### Supervised Learning



No.	Size	Color	Shape	Fruit Name
1	Big	Red	Rounded shape with a depression at the Top	Apple
2	Small	Red	Heart-shaped to nearly globular	Cherry
3	Big	Green	Long curviing cylinder	Banana
4	Small	Green	Round to oval. Bunch shape cylindrical	Grape

### Supervised Learning

Suppose you have taken an new fruit from the basket then you will see the size , color and shape of that particular fruit. If size is Big , color is Red , shape is rounded shape with a depression at the top, you will conform the fruit name as apple and you will put in apple group.

Likewise for other fruits also.

Job of groping fruits was done and happy ending.

You can observe in the table that a column was labeled as “**FRUIT NAME**” this is called as response variable.

If you learn the thing before from training data and then applying that knowledge to the test data(for new fruit), This type of learning is called as **Supervised Learning**

**Classification** come under Supervised learning.

### Steps Involved in Supervised Learning

First Determine the type of training dataset

Collect/Gather the labelled training data.

Split the training dataset into training dataset, test dataset, and validation dataset.

Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc

Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.

Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

## Unsupervised Learning

No trained data is given to the machine or we can say unlabeled or uncategorized data is given to the machine

Machine with the help of certain algorithms collects and categorized all the data on the basis of assumption and figure out the unknown pattern of data and group them in manner

### Example

Clustering DNA patterns to analyze evolutionary biology

# Machine Learning

## Unsupervised Learning

### Clustering

Machine tries to search inherit groups in a given data set such as grouping customers by purchasing behavior

### Example

Retail company may collect the following information on households:  
Household income. Household size

### Association

Collection of zero or more items is termed an itemset

Association learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y

### Example

{Beer, Diapers, Milk} is an example of a 3-itemset. The null (or empty) set is an itemset that does not contain any items.

# Machine Learning

## How Unsupervised Learning Works

### Unsupervised Learning



No.	Size	Color	Shape	Fruit Name
1	Big	Red	Rounded shape with a depression at the Top	Apple
2	Small	Red	Heart-shaped to nearly globular	Cherry
3	Big	Green	Long curviing cylinder	Banana
4	Small	Green	Round to oval. Bunch shape cylindrical	Grape

### Unsupervised Learning

Suppose you had a basket and it is filled with some different types fruits, your task is to arrange them as groups.

This time you don't know any thing about that fruits, honestly saying this is the first time you have seen them.

so how will you arrange them.

What will you do first??

You will take a fruit and you will arrange them by considering physical character of that particular fruit. suppose you have considered color.

Then you will arrange them on considering base condition as **color**.

Then the groups will be some thing like this.

# Machine Learning

## How Unsupervised Learning Works

### Unsupervised Learning



No.	Size	Color	Shape	Fruit Name
1	Big	Red	Rounded shape with a depression at the Top	Apple
2	Small	Red	Heart-shaped to nearly globular	Cherry
3	Big	Green	Long curviing cylinder	Banana
4	Small	Green	Round to oval. Bunch shape cylindrical	Grape

### Unsupervised Learning

RED COLOR GROUP: apples & cherry fruits.

GREEN COLOR GROUP: bananas & grapes.

So now you will take another physical character such as **size** .

RED COLOR AND BIG SIZE: apple.

RED COLOR AND SMALL SIZE: cherry fruits.

GREEN COLOR AND BIG SIZE: bananas.

GREEN COLOR AND SMALL SIZE: grapes.

job done happy ending.

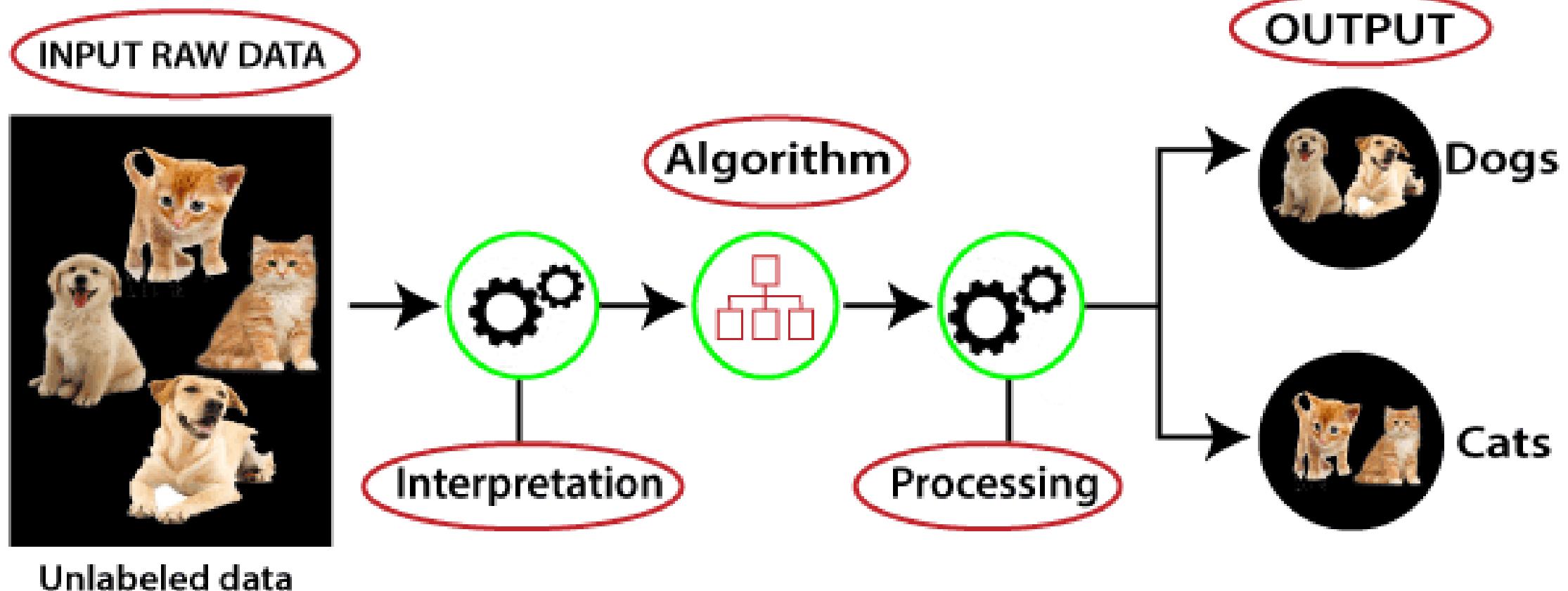
Here you didn't know learn any thing before  
,means no train data and no response variable

This type of learning is known unsupervised learning.

clustering comes under unsupervised learning.

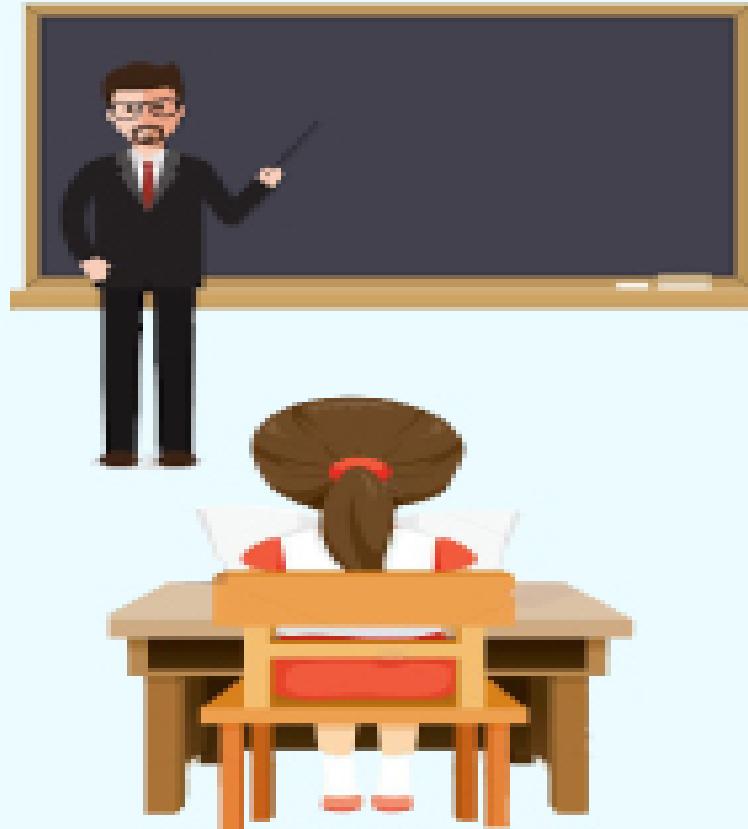
# Machine Learning

## Unsupervised Learning

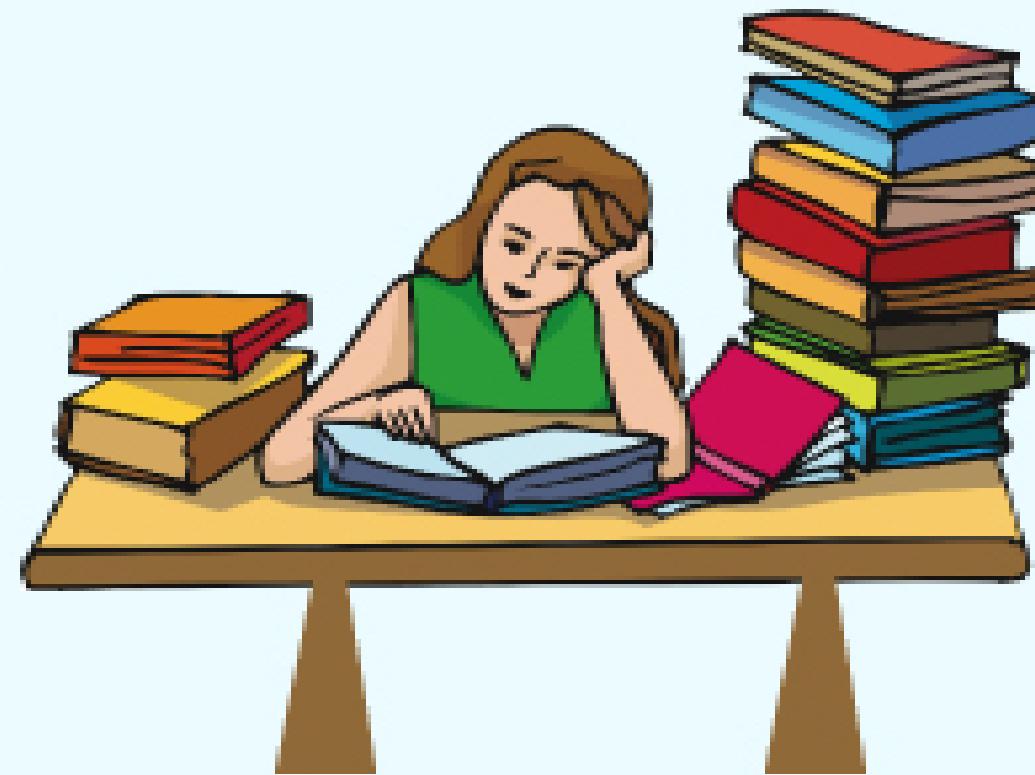


# Major Techniques of Machine Learning

## Supervised Learning



## Unsupervised Learning



# Machine Learning

## Reinforcement Learning

Machine is able to learn from the consequence of its actions by past experience or new explorations like when we play a new game, we take time to learn and experience over it, we try certain things to accumulate the right possibilities same as machine do.

### Example

Your music player or YouTube, showing you the similar playlist on the basis of your play or search history and its keep changes as your preference do

# Machine Learning

## Reinforcement Learning

### Positive

Positive reinforcement is the most common type of reinforcement learning that helps models to maximize the performance on a given task

### Example

A student calls out during class to get attention.

When the teacher responds, i.e. pays attention to the disruptive student, this response acts as positive reinforcement – therefore, the probability that the student will call out again increases

# Machine Learning

## Reinforcement Learning

### Negative

Negative reinforcement is used for maintaining a minimum performance standard instead of maximizing the performance

### Example

Doing away with homework for the weekend if students behave well in class.

Freedom to escape extra classes if students perform well in exams.

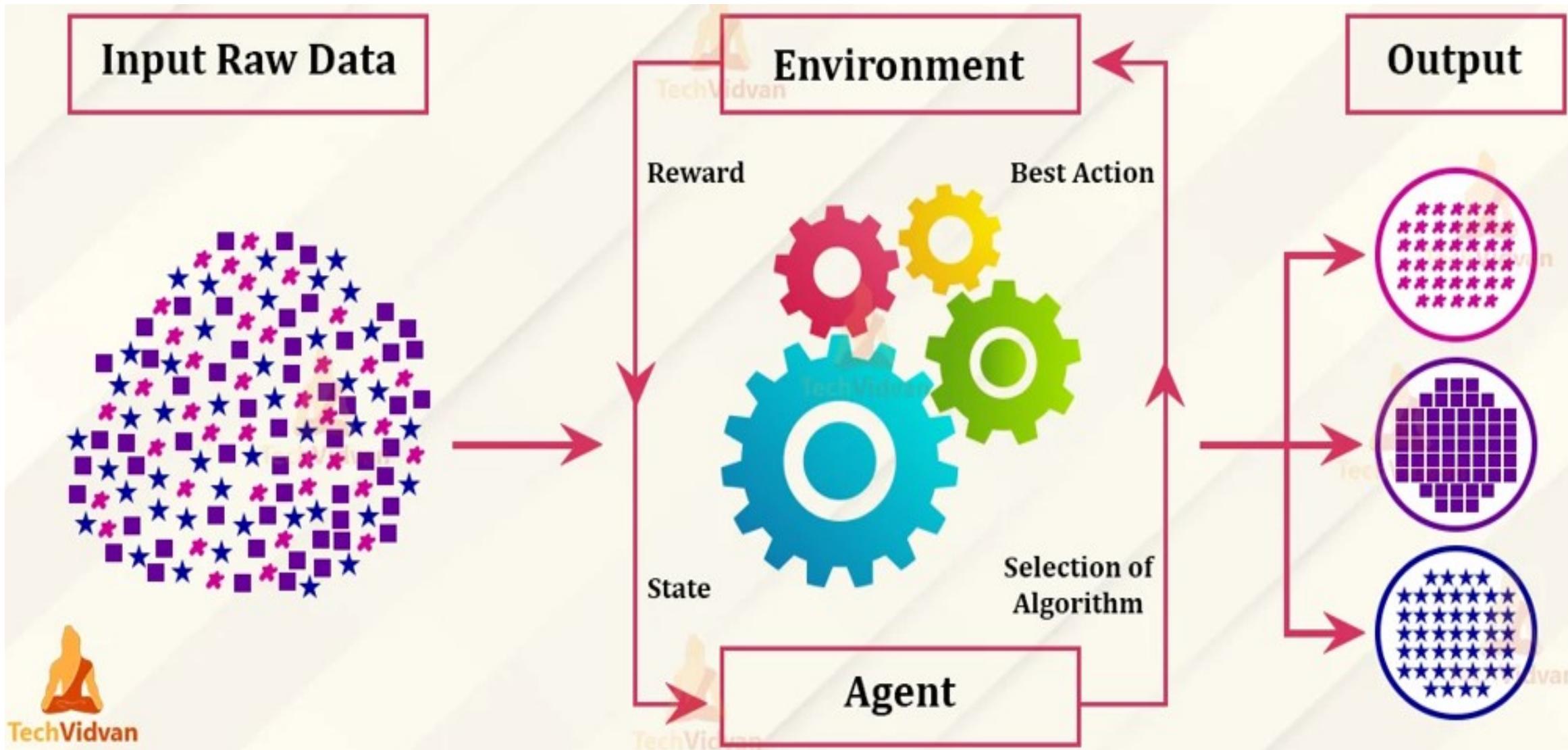
Removing the code of silence in the classroom once everyone completes their tasks.

Leaving early from the house to avoid traffic jams

Giving a pass to the car behind to avoid its honking

# Machine Learning

## Reinforcement Learning



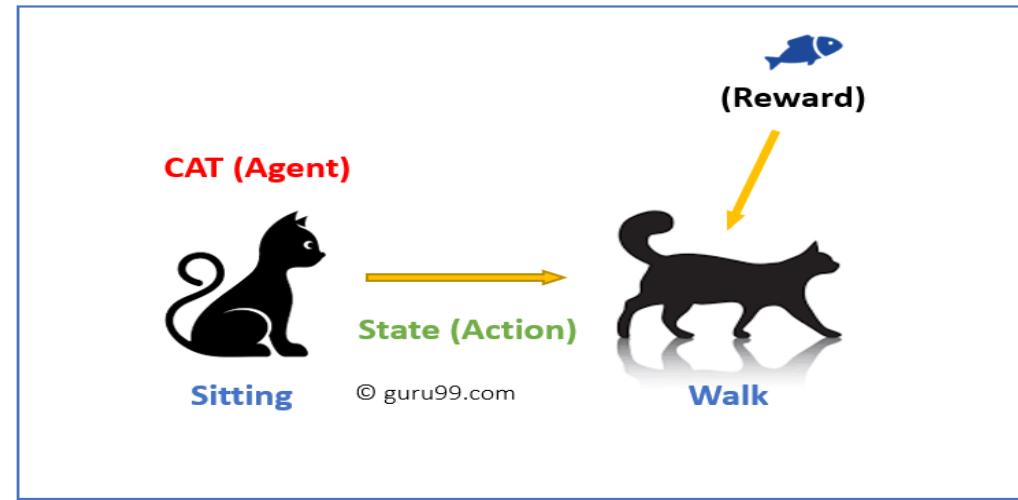
# ML Models Reinforcement Learning

## How Reinforcement Learning works?

Consider the scenario of teaching new tricks to your cat

As cat doesn't understand English or any other human language, we can't tell her directly what to do. Instead, we follow a different strategy.

We emulate a situation, and the cat tries to respond in many different ways. If the cat's response is the desired way, we will give her fish.



Now whenever the cat is exposed to the same situation, the cat executes a similar action with even more enthusiasm in expectation of getting more reward(food).

That's like learning that cat gets from "what to do" from positive experiences.

At the same time, the cat also learns what not do when faced with negative experiences.

# Machine Learning

## Semi-Supervised Machine Learning

Semi-supervised machine learning is a combination of supervised and unsupervised machine learning methods

Semi-supervised learning can be divided into Inductive Learning and Transductive Learning

Inductive learning refers to learning from specific (training) examples and trying to generalize the condition for the entire input space

Transduction or transductive learning is used in the field of statistical learning theory to refer to predicting specific examples given specific examples from a domain

Deductive learning is a more instructor-centered approach to education.

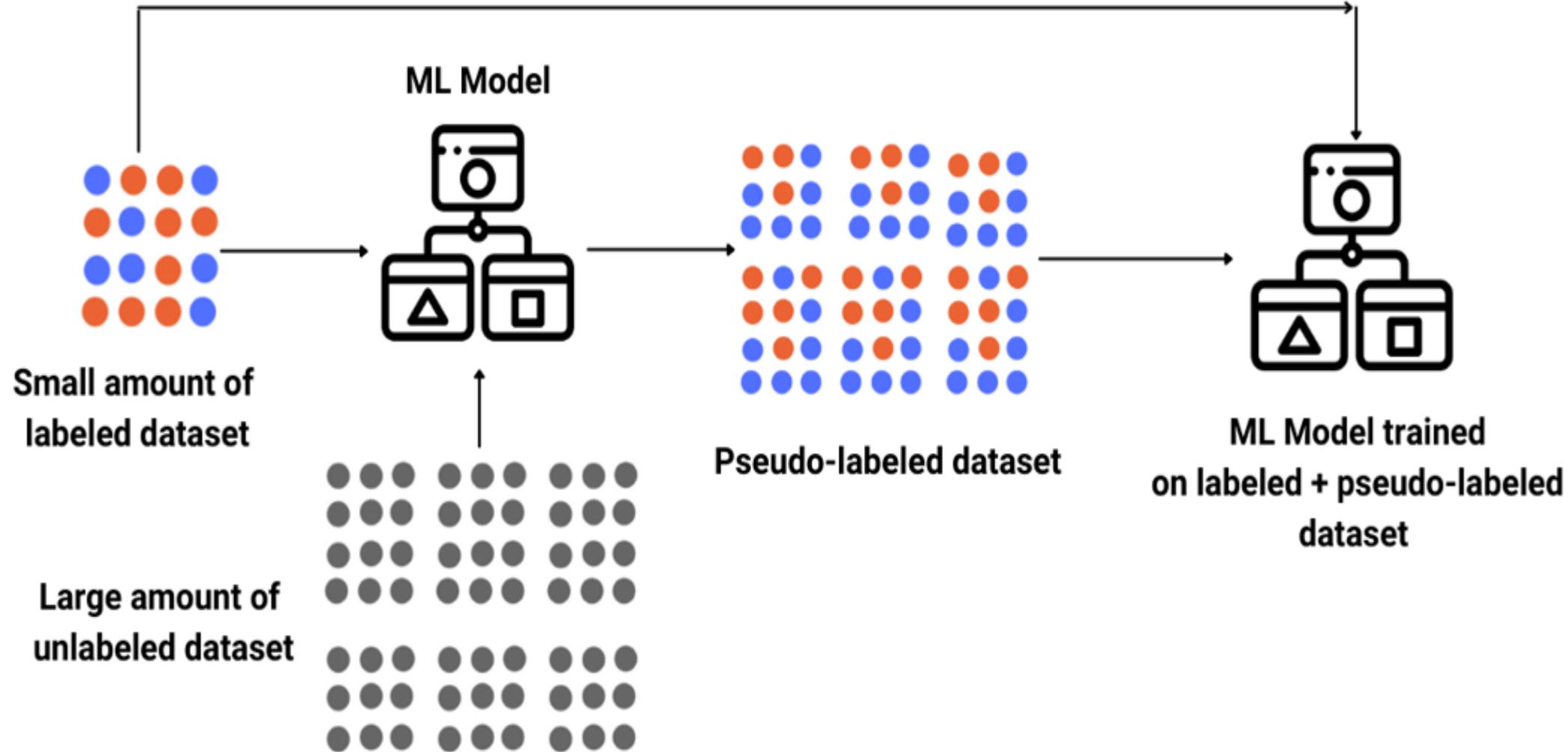
### Example

Labeling audio is a very resource- and time-intensive task, so semi-supervised learning can be used to overcome the challenges and provide better performance.

When the teacher responds, i.e. pays attention to the disruptive student, this response acts as positive reinforcement – therefore, the probability that the student will call out again increases

# Machine Learning

## Semi-Supervised Machine Learning



# How does Machine Learning work?

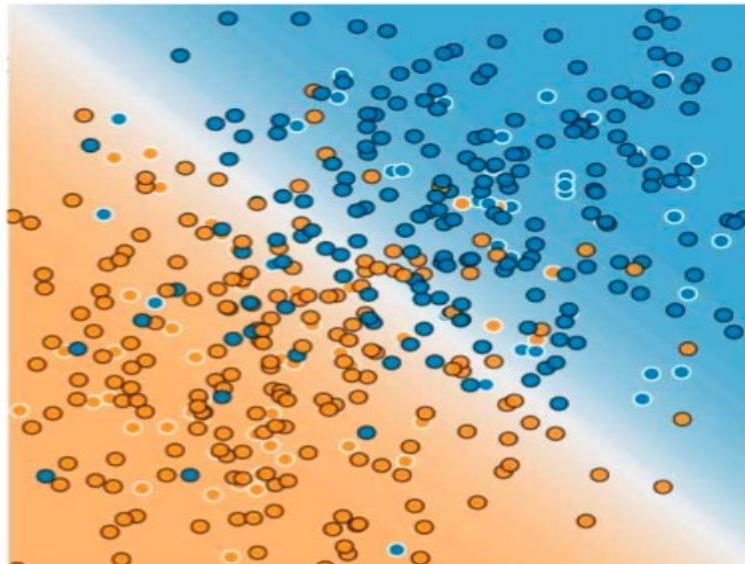
## Workflow follows these simple steps:

**Training set** is the one on which we train and fit our model basically to fit the parameters whereas **test data** is used only to assess performance of model

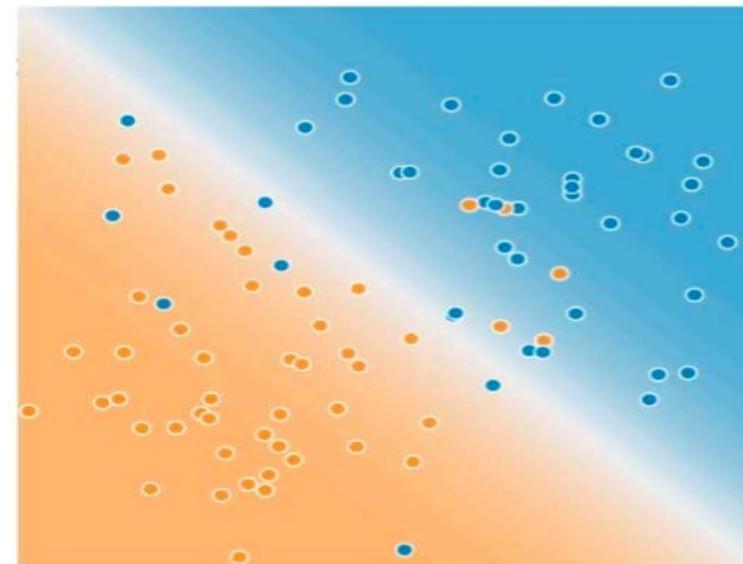
**Training data's** output is available to model whereas **testing data** is the unseen **data** for which predictions have to be made.

**Training set**—subset to train a model.

**Test set**—a subset to test the trained model.



Training Data

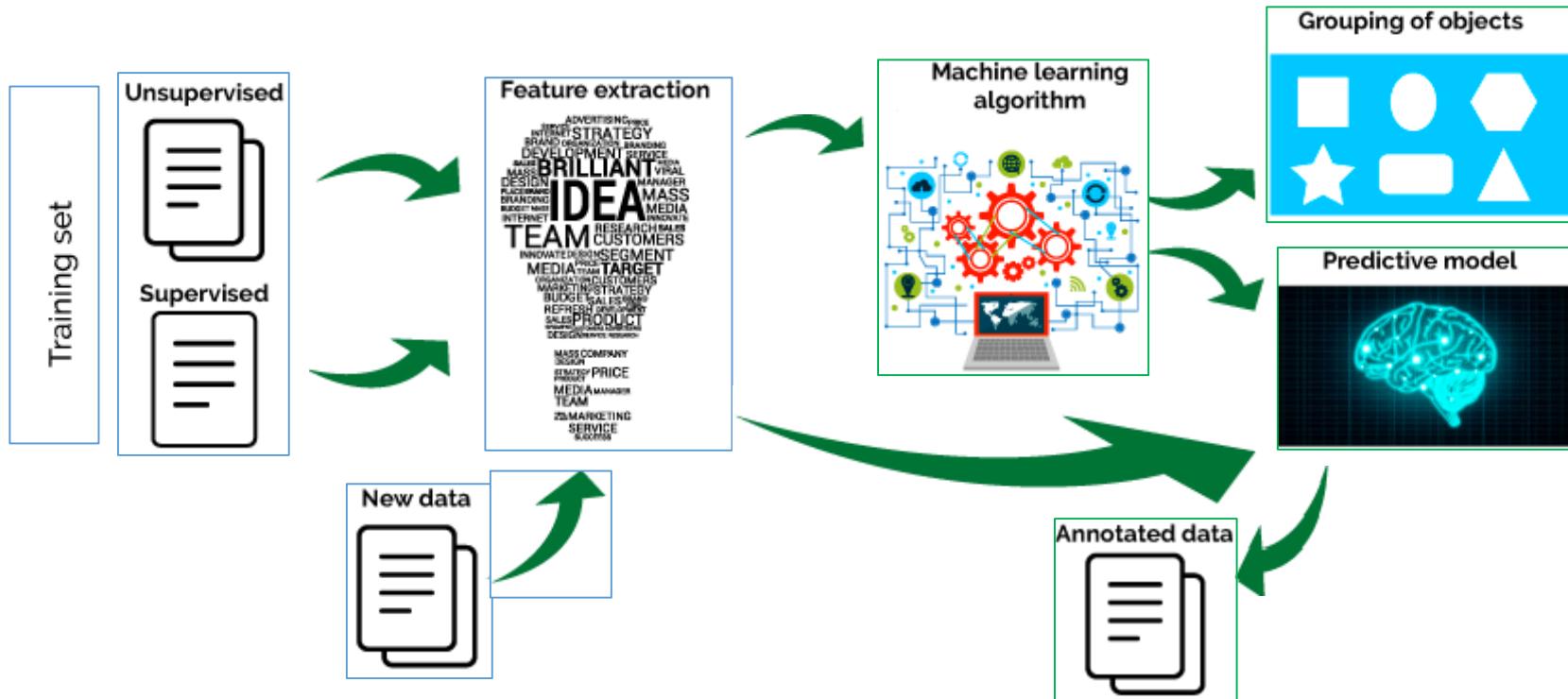


Test Data

# Machine Learning

Machine learning is the ability of computers to learn without being explicitly programmed.

Machine learning is the most dynamically developing field of data science today due to a number of recent theoretical and technological breakthroughs.



Machine Learning impacts society in a very influential way

Amazon uses Machine Learning to give you suggestion, what you can further buy.

Banks also use Machine Learning to approve Loans.

Telcos use customers data to segment them.

# How does Machine Learning work?

## Workflow follows these simple steps:

### Collect data.

Use your digital infrastructure and other sources to gather as many useful records as possible and unite them into a dataset.

### Prepare data

Prepare your data to be processed in the best possible way.

Data preprocessing and cleaning procedures can be quite sophisticated, but usually, they aim at filling the missing values and correcting other flaws in data, like different representations of the same values in a column (e.g. *December 14, 2016* and *12.14.2016* won't be treated the same by the algorithm).

### Split data

Separate subsets of data to train a model and further evaluate how it performs against new data.

### Train a model

Use a subset of historic data to let the algorithm recognize the patterns in it.

### Test and validate a model

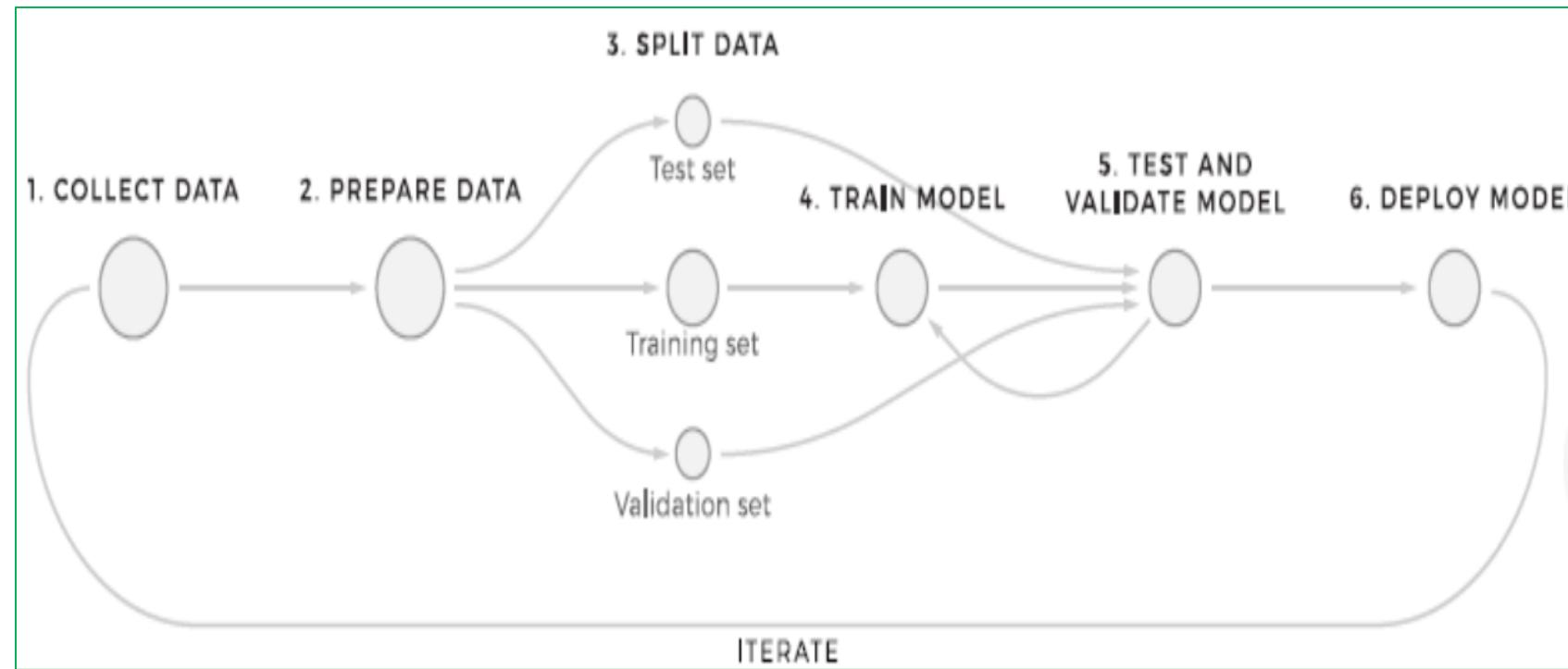
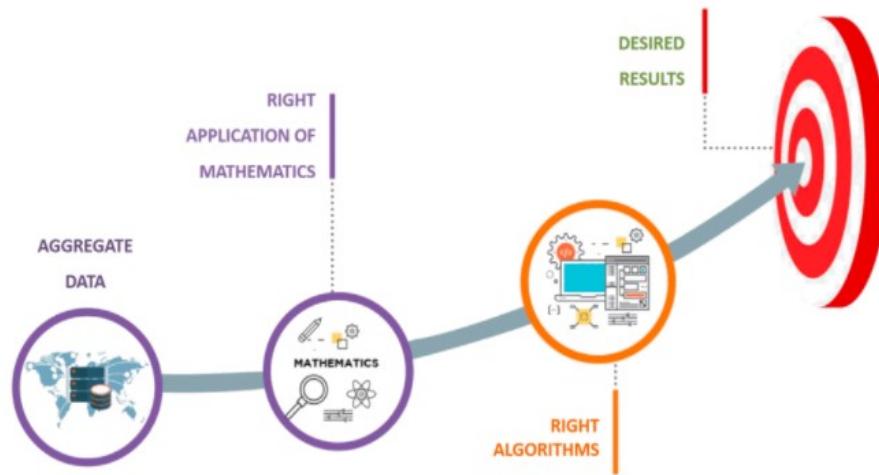
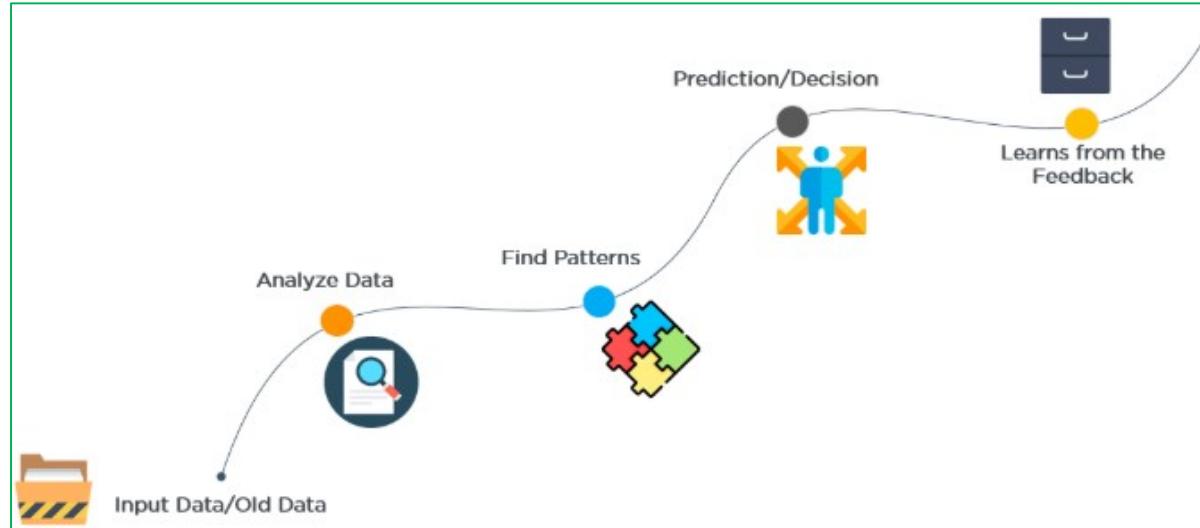
Evaluate the performance of a model using testing and validation subsets of historic data and understand how accurate the prediction is.

### Deploy a model

### Iterate

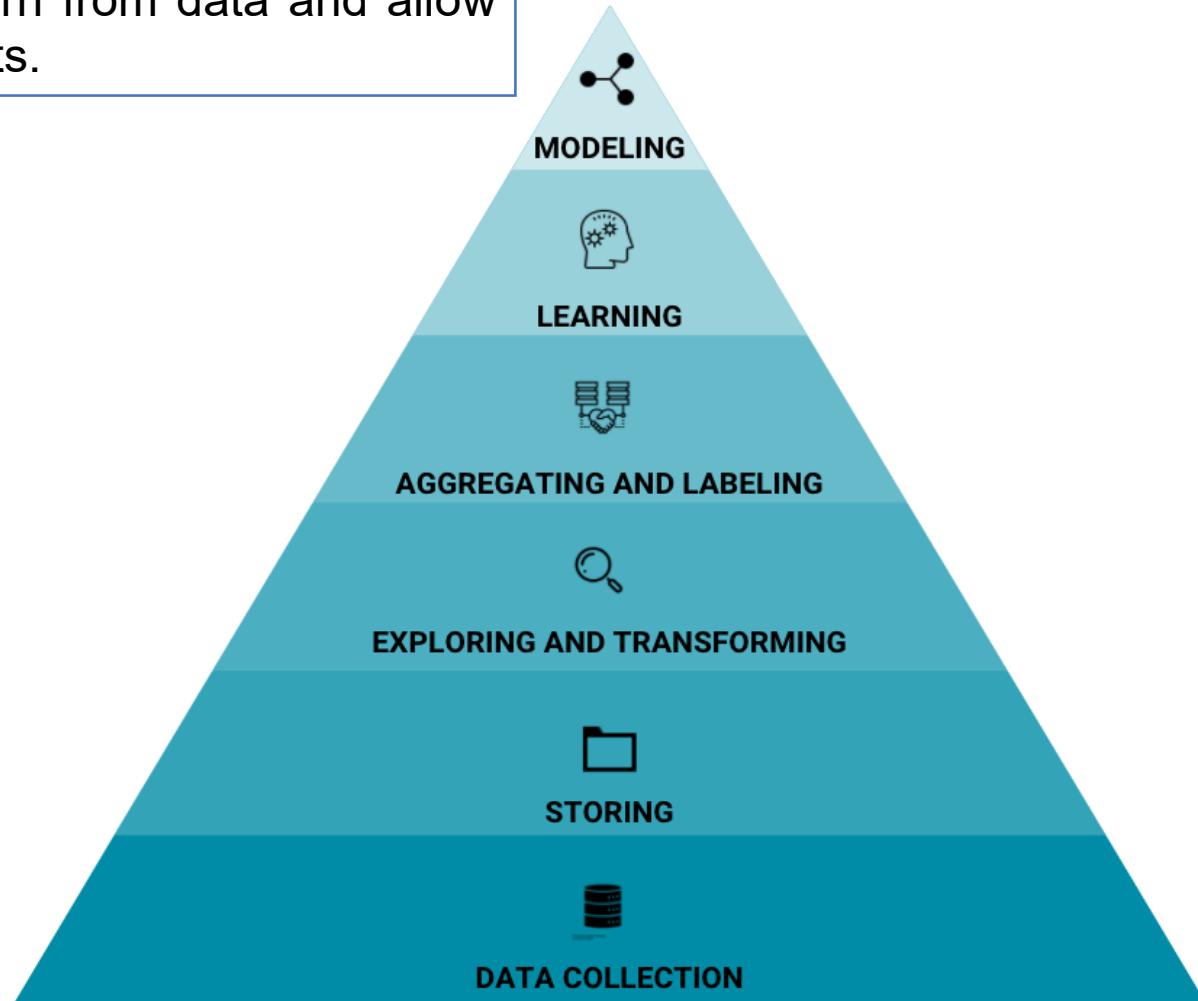
Embed the tested model into your decision-making framework as a part of an analytics solution or let users leverage its capabilities (e.g. better target your product recommendations).

# How does Machine Learning work?



# Machine Learning

Machine learning algorithms, inspired by the human learning process, iteratively learn from data and allow computers to find hidden insights.



# Evaluating ML Techniques

Evaluation helps us to know which algorithm best suits the given dataset for solving a particular problem

Machine Learning it is called as "Best Fit"

It evaluates the performance of different Machine Learning models, based on the same input dataset

Model evaluation aims to estimate the generalization accuracy of a model on future (unseen/out-of-sample) data

Methods for evaluating a model's performance are divided into 2 categories

Holdout and

Cross-validation.

Both methods use a test set (i.e data not seen by the model) to evaluate model performance

# Evaluating ML Techniques

How well is my model doing? Is it a useful model?

Will training my model on more data improve its performance?

Do I need to include more features?

Model evaluation aims to estimate the generalization accuracy of a model on future (unseen/out-of-sample) data

Methods for evaluating a model's performance are divided into 2 categories

Holdout and

Cross-validation.

Both methods use a test set (i.e data not seen by the model) to evaluate model performance

# Evaluating ML Techniques

Various ways to evaluate a machine learning model's performance

Confusion matrix

Accuracy

Precision

Recall.

Specificity

F1 score

Precision-Recall or PR curve

ROC (Receiver Operating Characteristics) curve

Iterate

DVR

# Evaluating ML Techniques

Various ways to evaluate a machine learning model's performance

## Confusion matrix or Classification metrics

When performing classification predictions, there's four types of outcomes that could occur

True Positive

True Negative

False Positive

False Negative

# Evaluating ML Techniques

Various ways to evaluate a machine learning model's performance

## Confusion matrix or Classification metrics

### True Positive

True positives are when you predict an observation belongs to a class and it actually does belong to that class

The model has predicted yes, and the actual value was also true

# Evaluating ML Techniques

Various ways to evaluate a machine learning model's performance

## Confusion matrix or Classification metrics

### True Negative

True negatives are when you predict an observation does not belong to a class and it actually does not belong to that class

Model has given prediction No, and the real or actual value was also No

# Evaluating ML Techniques

Various ways to evaluate a machine learning model's performance

## Confusion matrix or Classification metrics

### False Positive

False positives occur when you predict an observation belongs to a class when in reality it does not

The model has predicted Yes, but the actual value was No. It is also called a Type-I error.

False positives is also called a Type-I error.

# Evaluating ML Techniques

Various ways to evaluate a machine learning model's performance

## Confusion matrix or Classification metrics

### False Negative

False negatives occur when you predict an observation does not belong to a class when in fact it does

The model has predicted no, but the actual value was Yes, it is also called as Type-II error

False negative is also called as Type-II error

# Evaluating ML Techniques

Various ways to evaluate a machine learning model's performance

## Calculations using Confusion Matrix:

Confusion Matrix in Machine learning evaluates the performance of the classification models, when they make predictions on test data, and tells how good our classification model is

Confusion Matrix in Machine learning not only tells the error made by the classifiers but also the type of errors such as it is either type-I or type-II error

With the help of the Confusion Matrix in Machine learning, we can calculate the different parameters for the model, such as accuracy, precision, etc

# Evaluating ML Techniques

Various ways to evaluate a machine learning model's performance

## Calculations using Confusion Matrix:

### Classification Accuracy

Classification Accuracy is one of the important parameters to determine the accuracy of the classification problems

Classification Accuracy defines how often the model predicts the correct output

Classification Accuracy can be calculated as the ratio of the number of correct predictions made by the classifier to all number of predictions made by the classifiers

The formula is given below:

Classification Accuracy can be calculated as the ratio of the number of correct predictions made by the classifier to all number of predictions made by the classifiers

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

# Evaluating ML Techniques

Various ways to evaluate a machine learning model's performance

## Calculations using Confusion Matrix:

### Misclassification rate

Misclassification rate is also termed as Error rate, and it defines how often the model gives the wrong predictions

The value of error rate can be calculated as the number of incorrect predictions to all number of the predictions made by the classifier

The formula is given below:

$$\text{Error rate} = \frac{FP + FN}{TP + FP + FN + TN}$$

# Evaluating ML Techniques

Various ways to evaluate a machine learning model's performance

## Calculations using Confusion Matrix:

### Precision

Precision can be defined as the number of correct outputs provided by the model or out of all positive classes that have predicted correctly by the model, how many of them were actually true

Precision can be calculated using the below formula

$$\text{Precision} = \frac{TP}{TP+FP}$$

# Evaluating ML Techniques

Various ways to evaluate a machine learning model's performance

## Calculations using Confusion Matrix:

### Recall

Recall is defined as the out of total positive classes, how our model predicted correctly

The recall must be as high as possible

$$\text{Recall} = \frac{TP}{TP+FN}$$

# Evaluating ML Techniques

Various ways to evaluate a machine learning model's performance

## Calculations using Confusion Matrix:

### F-measure

If two models have low precision and high recall or vice versa, it is difficult to compare these models

for this purpose, we can use F-score

This score helps us to evaluate the recall and precision at the same time

The F-score is maximum if the recall is equal to the precision

F-measure can be calculated using the below formula

$$\text{F-measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

# Evaluating ML Techniques

Various ways to evaluate a machine learning model's performance

## Calculations using Confusion Matrix:

### Null Error rate

Null Error rate defines how often our model would be incorrect if it always predicted the majority class

As per the accuracy paradox, it is said that "the best classifier has a higher error rate than the null error rate."

# Evaluating ML Techniques

Various ways to evaluate a machine learning model's performance

## Calculations using Confusion Matrix:

### ROC Curve

The ROC is a graph displaying a classifier's performance for all possible thresholds

The graph is plotted between the true positive rate (on the Y-axis) and the false Positive rate (on the x-axis)

# Uses of Machine learning

Machine learning is used in

Internet search engines,

Email filters to sort out spam

Websites to make personalized recommendations

Banking software to detect unusual transactions, and

Lots of apps on our phones such as voice recognition

Help with the diagnosis of diseases

# Uses of Machine learning

## Image Recognition

Image recognition referred to as a digital image and for these images, the measurement describes the output of every pixel in an image.

Face recognition is also one of the great features that have been developed by machine learning only

Image recognition helps to recognize the face and send the notifications related to that to people

## Voice Recognition

Machine learning (ML) also helps in developing the application for voice recognition. It also referred to as virtual personal assistants (VPA).

Machine learning (ML) will help you to find the information when asked over the voice

one mobile app called Google Allo and smartphones are Samsung S8 and Bixby

# Uses of Machine learning

## Predictions

Machine Learning helps in building the applications that predict the price of cab or travel for a particular duration and congestion of traffic where can be found

While booking the cab and the app estimates the approximate price of the trip that is done by the uses of machine learning only

## Videos Surveillance

It helps to detect the crime or any miss happening that is going to happen before it happens

It helps in tracking the unusual behavior of people like napping on benches and standing still from a long time, stumbling etc. and it will create an automatic alert to the guards or people who all are posted there and they can help to avoid any issues or problems.

# Uses of Machine learning

## Predictions

Machine Learning helps in building the applications that predict the price of cab or travel for a particular duration and congestion of traffic where can be found

While booking the cab and the app estimates the approximate price of the trip that is done by the uses of machine learning only

## Social Media Platform

Social Media is being used for providing better news feed and advertisement as per the user's interest is mainly done through the uses of machine learning only

There are many examples like friend suggestions, page suggestions for Facebook, songs, and videos suggestion on YouTube.

Social Media also provides the technique to extract useful information from images and videos

# Uses of Machine learning

## Spam and Malware

Email clients use a number of spam filtering and these spam filters are continuously getting updated and these are mainly done by the uses of machine learning

Rule-based, multi-layer and tree induction are some of the techniques that are provided by machine learning

Rule-based, multi-layer and tree induction are some of the techniques that are provided by machine learning

Number of malware are detected and these are detected mainly by the system security programs that are mainly helped by machine learning only

## Customer Support

Most of the reputed companies or many websites provide the option to chat with a customer support representative

Answers are given by the chatbot which extracts the information from the website and provides the answer to customers

Better and understand the queries quickly and faster and also provides a good result by giving appropriate result and it is done by the uses of machine learning only.

Many physicians use chatbots with speech recognition capabilities to discern patterns in symptoms

# Uses of Machine learning

## Search Engines

Search engines available while searching to provide the best results to customers

Many machine learning algorithms created for searching the particular user query like for google

Whatever the page is being opened by the users for a particular topic frequently that will remain at the top of the page for a long time

## Fraud and Preference.

Companies to keep track of money laundering like Paypal

Paypal uses the set of tools to help them to check or compare the millions of transactions and make secure transactions

# Transforming towards the Data Science Career

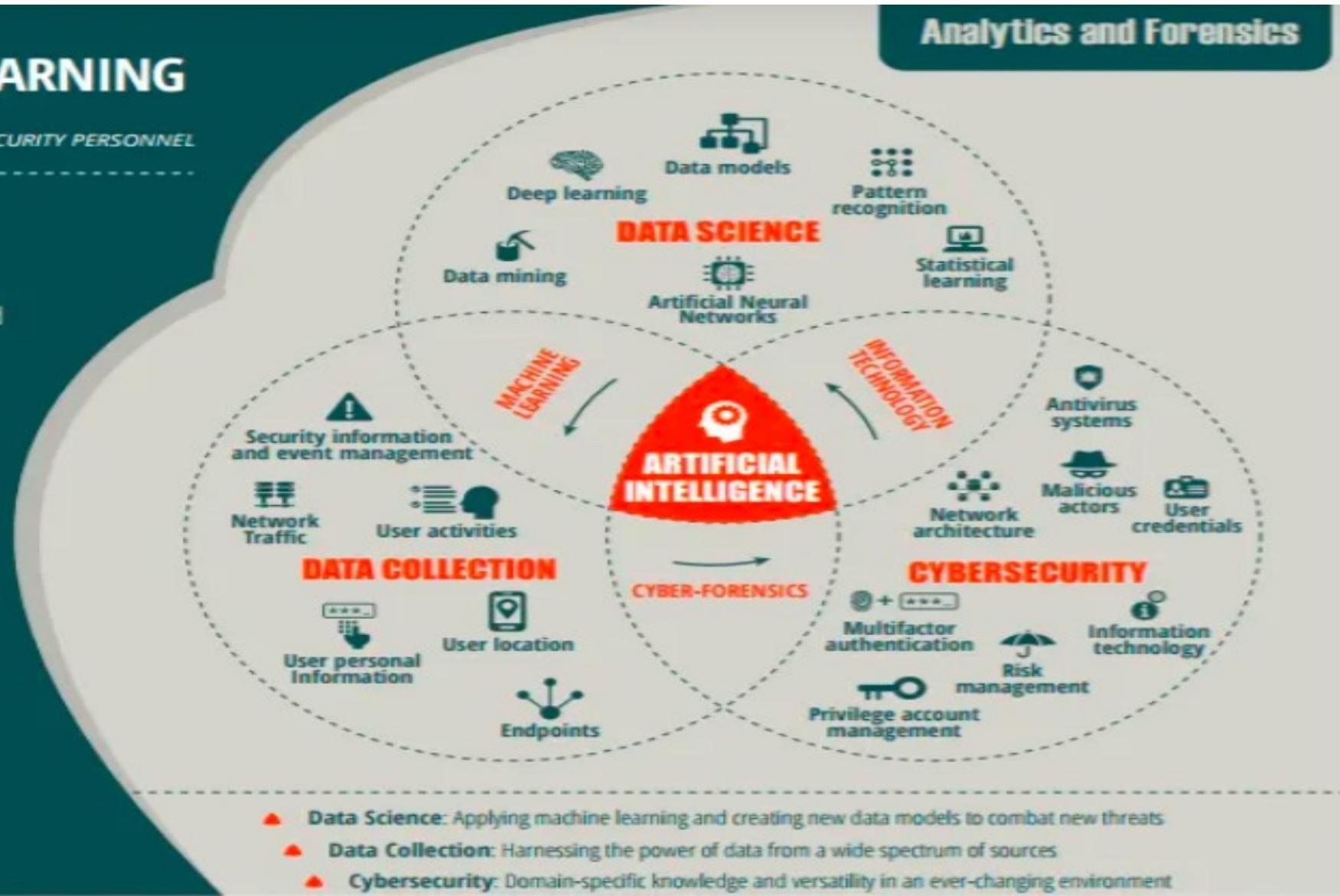
## Data Science used in Cyber Security

Data Science through effective Analytics Systems help enhance the cyber security industry

### WHAT CAN MACHINE LEARNING DO FOR CYBERSECURITY?

A POTENT NEW ARSENAL FOR IT AND CYBERSECURITY PERSONNEL

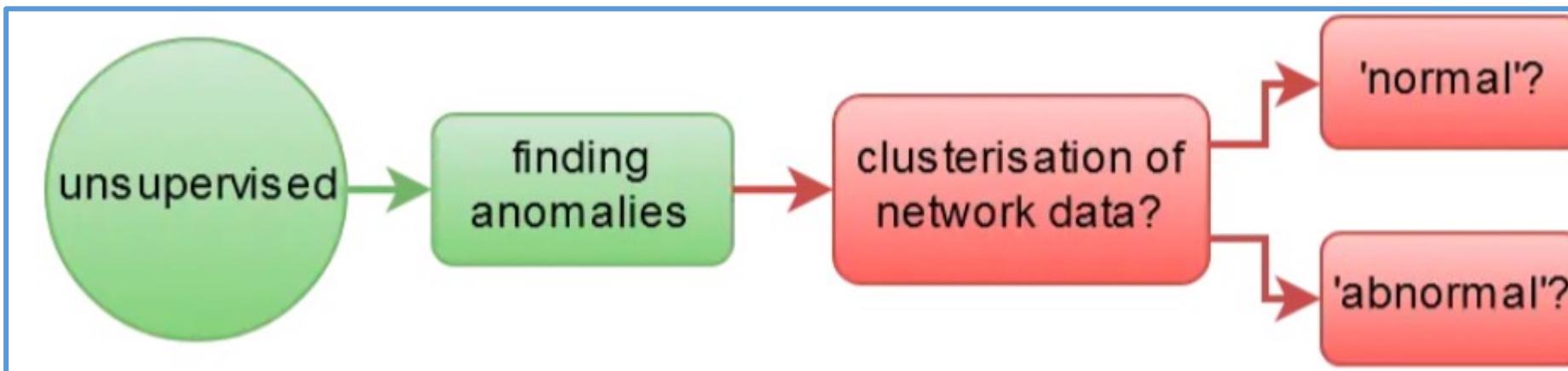
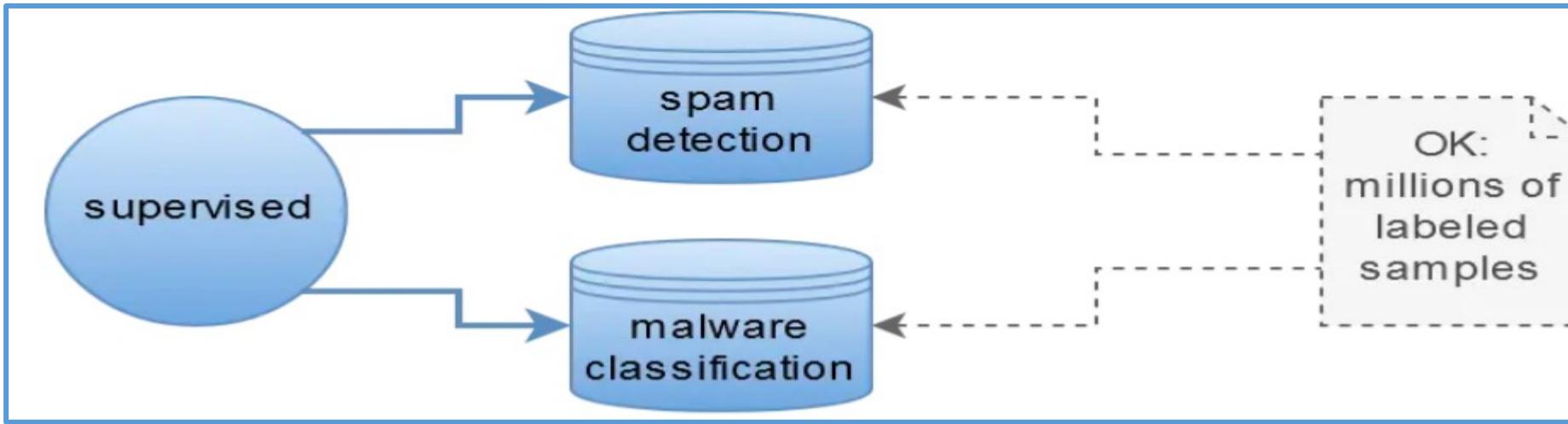
- User entity behavioral analytics, deep learning, automation
- Assist IT professionals and defend against new cyberthreats
- Better predictive models, lower FPR, distill new metrics
- Fraud and anomaly detection
- Defend against new cyberthreats
- Better use of internal data and global repositories
- Tackle device influx and enhanced data loss prevention (DLP) solutions



# Transforming towards the Data Science Career

## Data Science used in Cyber Security

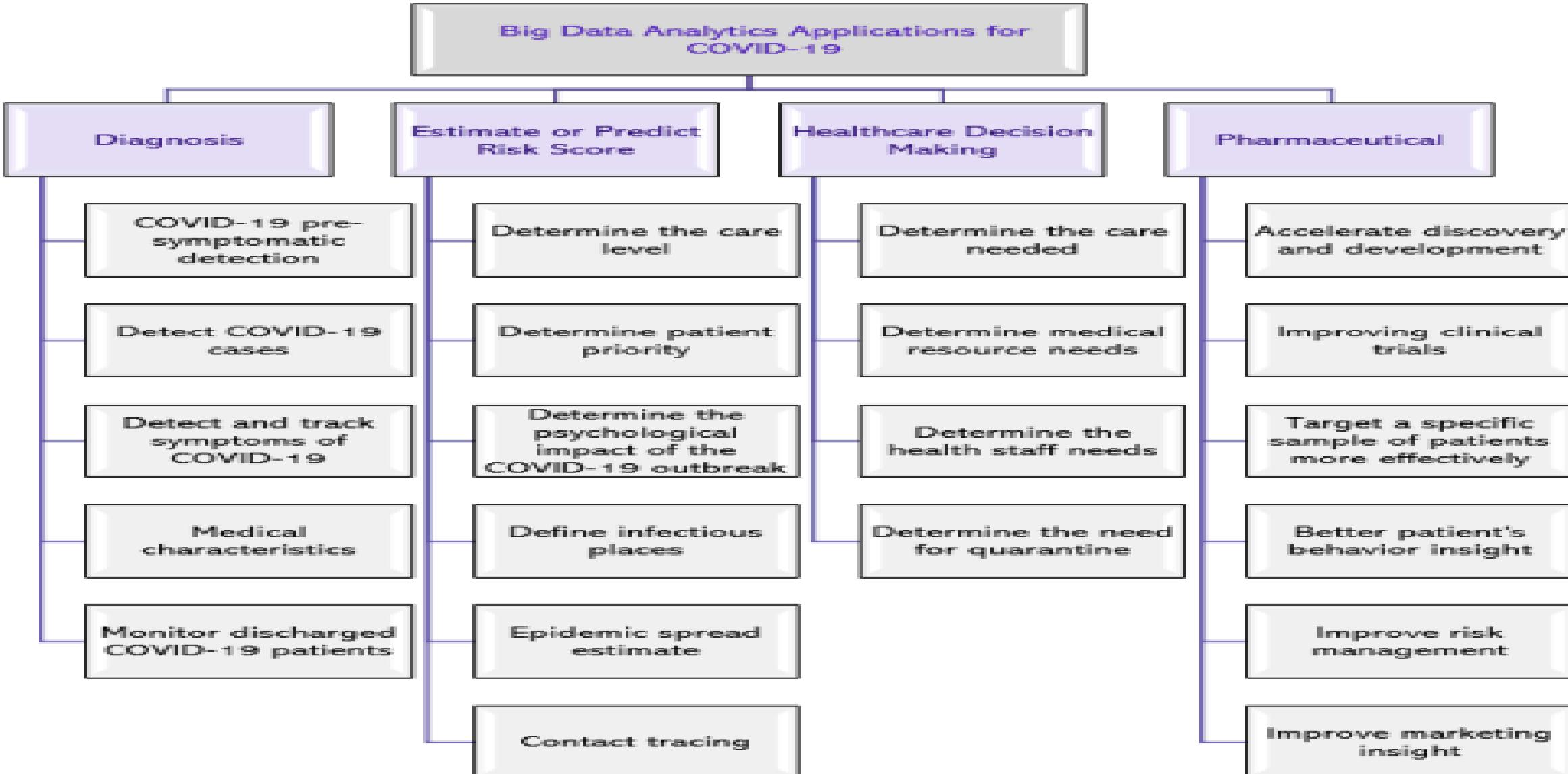
How Machine Learning cyber security can safeguard organizations:



# Transforming towards the Data Science Career

## Data Science used in Cyber Security

### Potential Application areas of Big Data Analytics for COVID-19



### Potential Application areas of Big Data Analytics for COVID-19

#### Type and source of medical data

##### Social Media Data

i.e., Virous spread news, medical advice, ect.

##### ED and ICU Records

i.e., Admission data, capacity, medical resources, abd medical staff, ect.

##### Passenger and Travel Data

i.e., Flight number, data, source, destination, departure airport, ect.



##### Mobile and Sensor Data

i.e., Activity, location, physiological, ect.

##### Patient Medical Characteristics

i.e., Vital signs, symptoms, ect.

##### Hospital Resource Records

i.e., Capacity, medical resources and medical staff, ect.

## Data Science used in Cyber Security

How has analytics proved to be beneficial in the success story of BharatMatrimony?

### **Data and analytics as a key enabler in all decision making processes.**

For acquisition marketing, we track the performance of marketing channels and campaigns and measure the return on investment which, in turn, is used to optimize the allocation of spends in the future

Identify customer needs, based on their profile and activities to identify their product propensity

Segment and profile customers based on the value delivered and identify gaps in service delivery

Use scoring models to identify and stall suspicious user activity

Artificial intelligence (AI) algorithm is helping the users on matrimony sites to find a match not only based on their preferences but also by observing and understanding user's behavior and suggesting real-time appropriate profiles

AI algorithm knows more about what you are looking for than you might yourself! This helps us find the best matches for our users - not just by searching according to preferences but by observing user behavior and suggesting similar matches

The logo consists of the letters "LAB" in a stylized, black, hand-drawn font. The letter "L" has a small dot above it. The letter "A" has a small circle at its top center. The letter "B" has a small circle at its bottom right corner.

Explore Datasets Online (can refer Kaggle, UCI ML, etc.)

a) Load dataset in google colab

b) Print first five values and last five values in dataset.

c) check correlation between fields present in dataset



Any Questions

Dr DV Ramana  
Academic Advisor  
And Program Manager

Mail Address: [pythonpmg@gmail.com](mailto:pythonpmg@gmail.com)

To contact: +91 9959423084