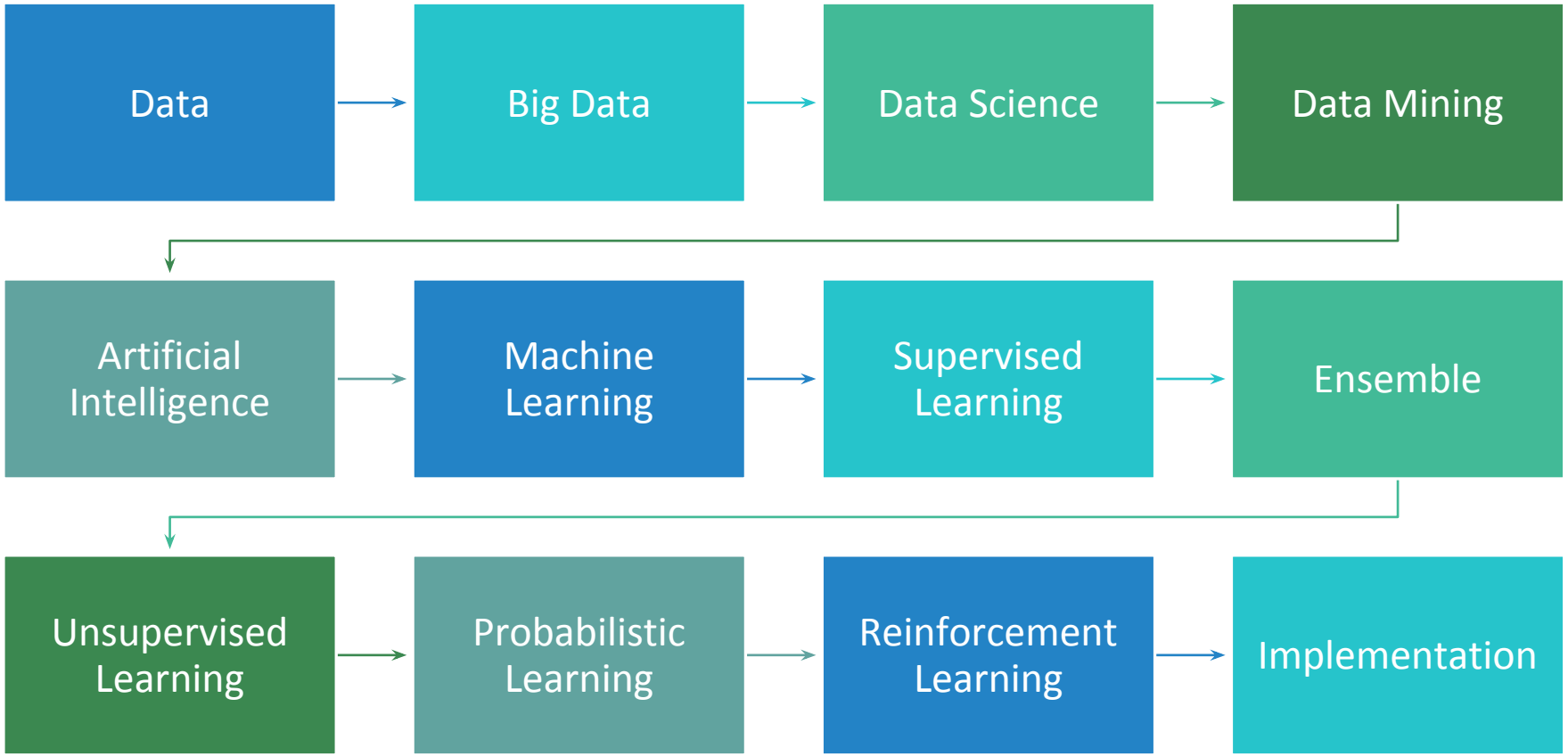
Data Science

INDEED INSPIRING INFOTECH

WWW.INDEEDINSPIRING.COM

INFO@INDEEDINSPIRING.COM

Agenda



DATA

Raw Facts & Figures

DATA

Processed Data = Information

VIDEO FILES | AUDIO FILES

DATA FORMATS

TEXT FILES | SENSOR DATA

BIG DATA

The 6 Vs of Big Data

VOLUME

VARIETY

VELOCITY

VERACITY

VALUE

VARIABILITY

The six Vs of big data

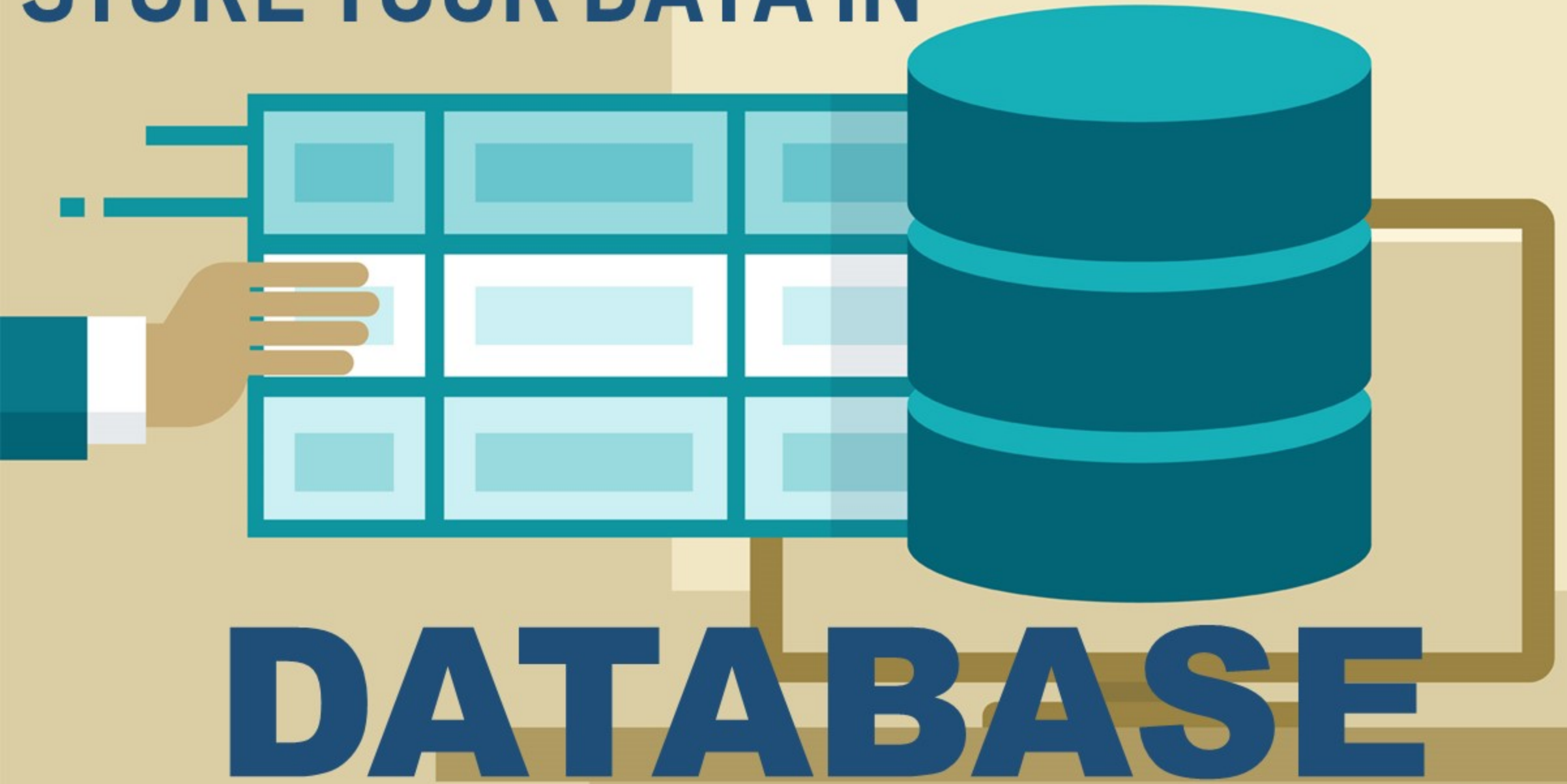
Big data is a collection of data from various sources, often characterized by what's become known as the 3Vs: *volume*, *variety and velocity*. Over time, other Vs have been added to descriptions of big data:

VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources.	The types of data: structured, semi-structured, unstructured.	The speed at which big data is generated.	The degree to which big data can be trusted.	The business value of the data collected.	The ways in which the big data can be used and formatted.






STORE YOUR DATA IN



DATABASE

How do you store DATA?

Lockers

Almirah

Cupboard

DATABASE

A database is **an organized collection of data**, generally stored and accessed electronically from a computer system.

Example: Database can contain 1 or more tables.

ID	Name	Contact	Email
1	Kushal Sharma	9762203269	Kushal@indeedinspiring.com
2	Mahadev Bhumbar	9850603269	info@indeedinspiring.com
3	Nitish Kumar	8149102080	info@prushal.com

OLTP

Online Transaction Processing

OLAP

Online Analytical Processing

Data Science

Data Science

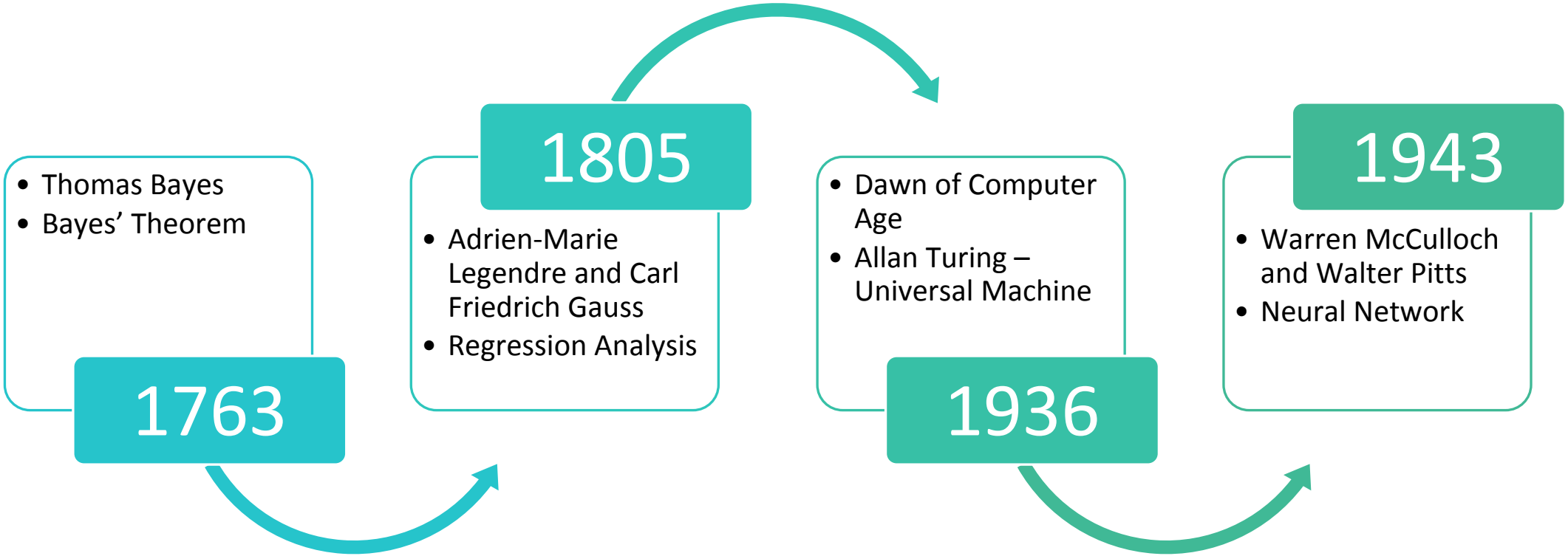
Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data.

Datarobot.com says

Data science is the field of study that combines domain expertise, programming skills, and knowledge of math and statistics to extract meaningful insights from data.

In turn, these systems generate insights that analysts and business users translate into tangible business value.

History Says…

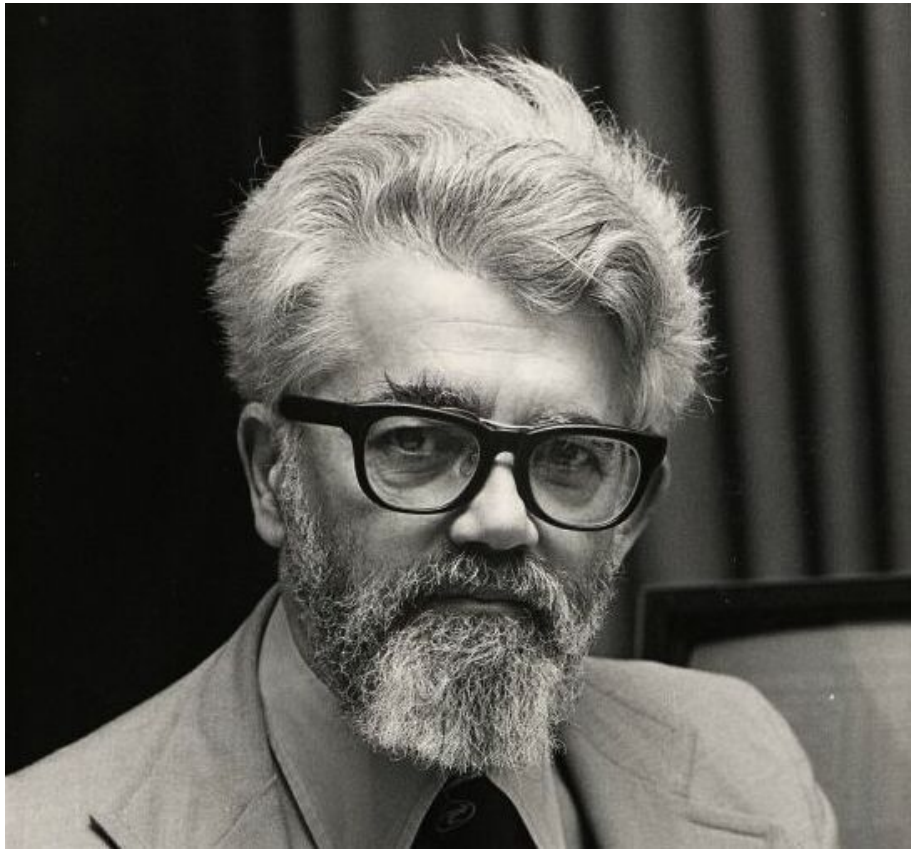


Data Mining

Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.

Artificial Intelligence

John McCarthy first coined the term artificial intelligence in **1956** when he invited a group of researchers from a variety of disciplines including language simulation, neuron nets, complexity theory and more to a summer workshop called the **Dartmouth Summer Research Project** on Artificial Intelligence to discuss what would ultimately become the field of AI.



What researchers have been doing so far?

Thinking	Machines
Artificial	Intelligence
Information	Processing

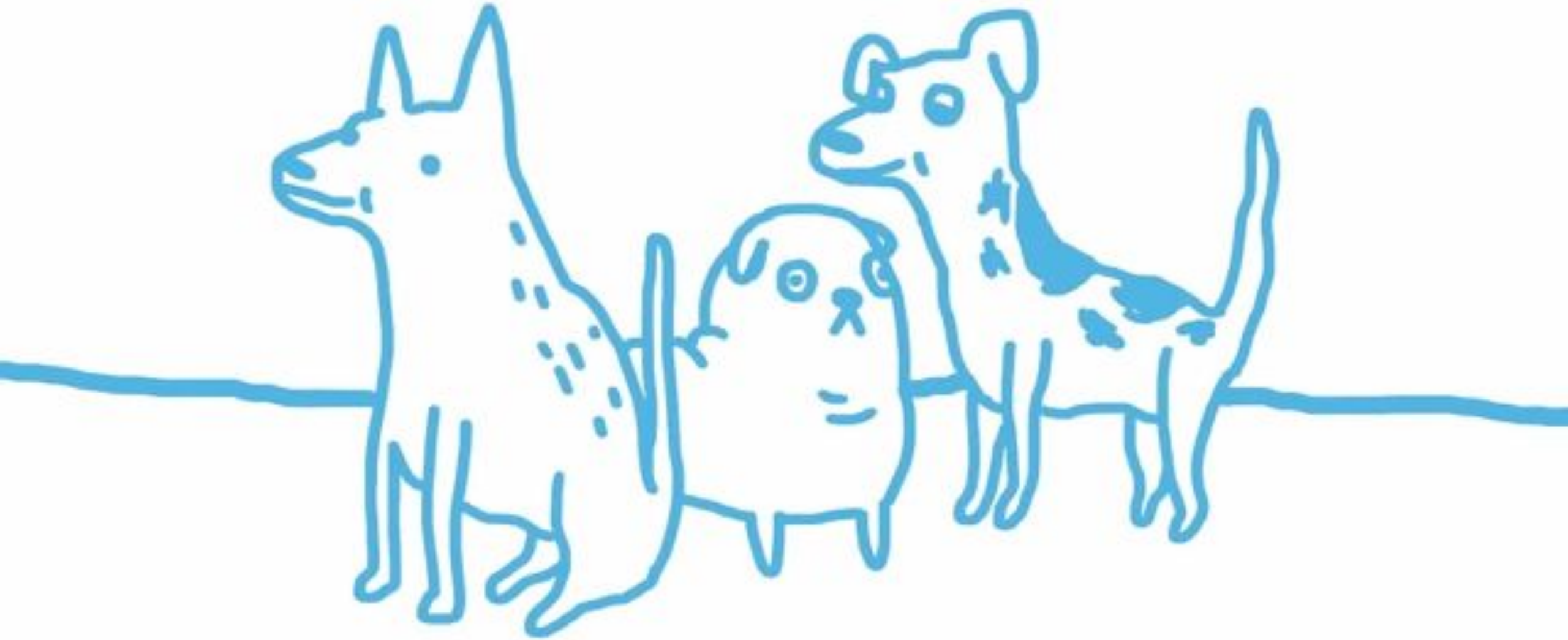
AI – The English Oxford Living Dictionary

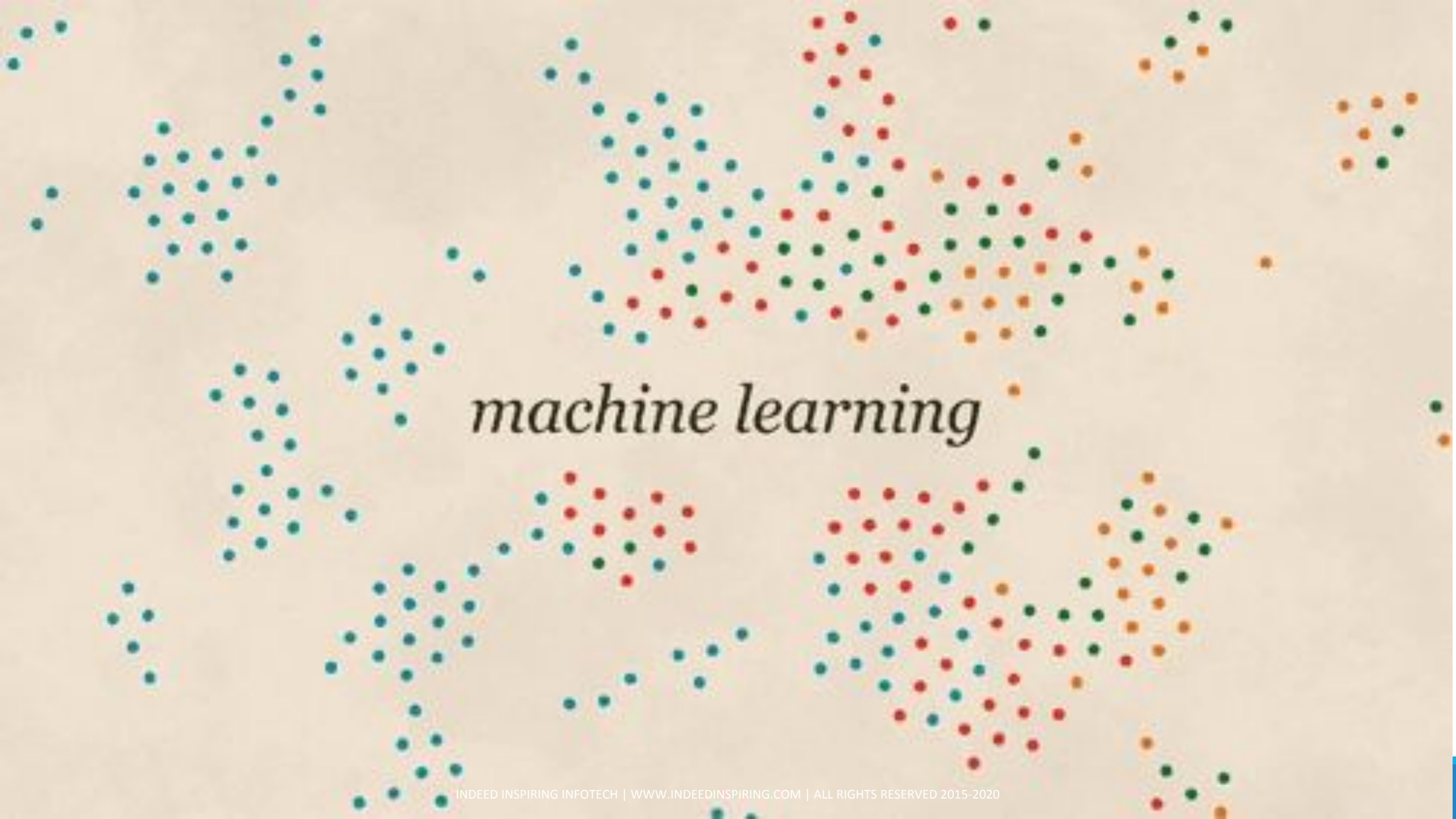
The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.

Merriam-Webster defines AI

A branch of computer science dealing with the simulation of intelligent behavior in computers.

The capability of a machine to imitate intelligent human behavior.





Machine Learning

Arthur Samuel (1901-1990), an American pioneer in the field of computer gaming and artificial intelligence, coined the term “machine learning” in **1959**.

He defined it as a “field of study that gives computers the ability to learn without being explicitly programmed”.

Machine Learning

Machine learning is an application of **artificial intelligence (AI)** that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

Deep Learning

Deep learning is a subset of machine learning in artificial intelligence (AI) that has networks capable of learning unsupervised from data that is unstructured or unlabeled.

Also known as **deep neural learning** or **deep neural network**.

Learning Paradigms

Supervised Learning

Ensemble Learning

Unsupervised Learning

Probabilistic Learning

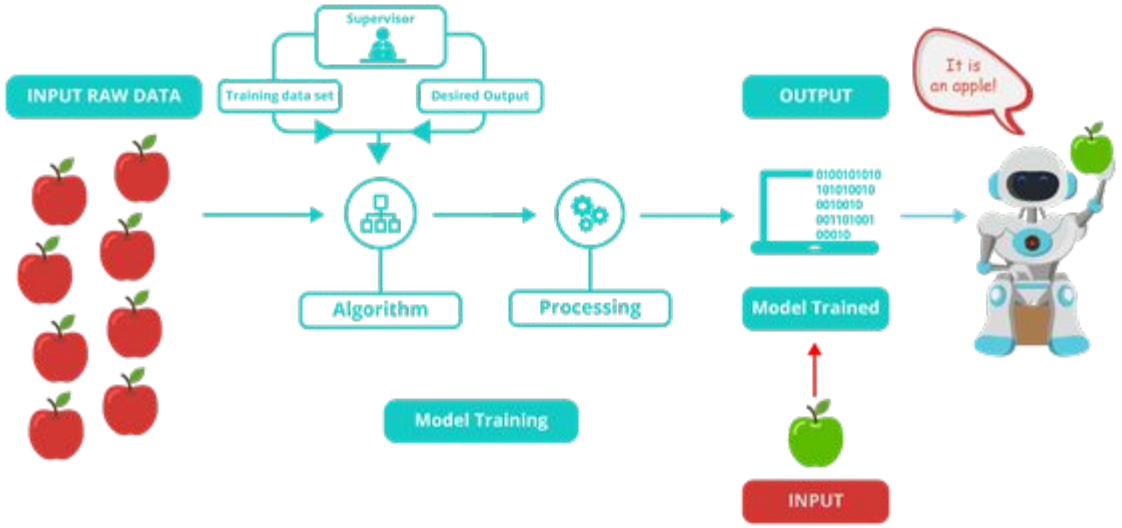
Reinforcement Learning

Supervised Learning – Introduction

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs.

It infers a function from labeled training data consisting of a set of training examples.

Supervised Learning – Example



Supervised Learning – i/o Variables

Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.

$Y = f(X)$

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

Learn Supervised Learning in 5 minutes: https://youtu.be/WKqshCFxX-E

Supervised Learning – Approach

Supervised learning problems can be further grouped into regression and classification problems.

Classification:

- A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”.
- Learn Decision Tree Algorithm in 7 minutes: https://youtu.be/LymTZR-aeQg

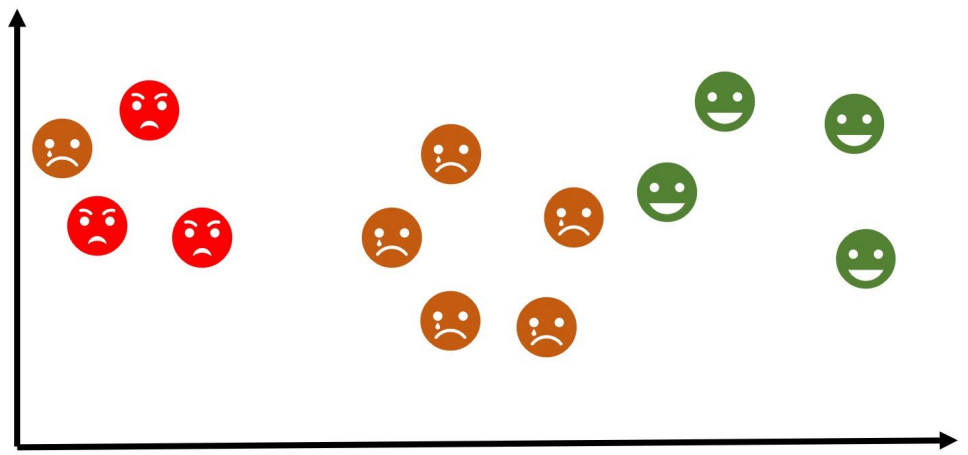
Regression:

- A regression problem is when the output variable is a real value, such as “dollars” or “weight”.
- Learn Regression Analytics in 10 minutes: https://youtu.be/lfsO05GJhpE

Unsupervised Learning – Definition

Unsupervised learning is a type of machine learning that looks for previously undetected patterns in a data set with no pre-existing labels and with a minimum of human supervision.

In unsupervised learning, an AI system may group unsorted information according to similarities and differences even though there are no categories provided.



Learn Unsupervised Learning in 8 minutes: https://youtu.be/TQvzUvzG9as

Unsupervised Learning: Apriori

Apriori algorithm, a classic algorithm, is useful in mining frequent itemsets and relevant association rules.

Usually, you operate this algorithm on a database containing a large number of transactions. One such example is the items customers buy at a supermarket.

It has got this odd name because it uses ‘prior’ knowledge of frequent itemset properties.

Learn Apriori Algorithm in 8 minutes: https://youtu.be/hoyBSnFM-Bo

Apriori: Story of Vegetable Seller

Kamala – the vegetable seller, always bundles onions and potatoes together.

She even offers a discount to people who buy these bundles.

Why does she do so?

Apriori: Story of Vegetable Seller

She realizes that people who buy potatoes also buy onions.

Therefore, by bunching them together, he makes it easy for the customers.

At the same time, she also increases her sales performance.

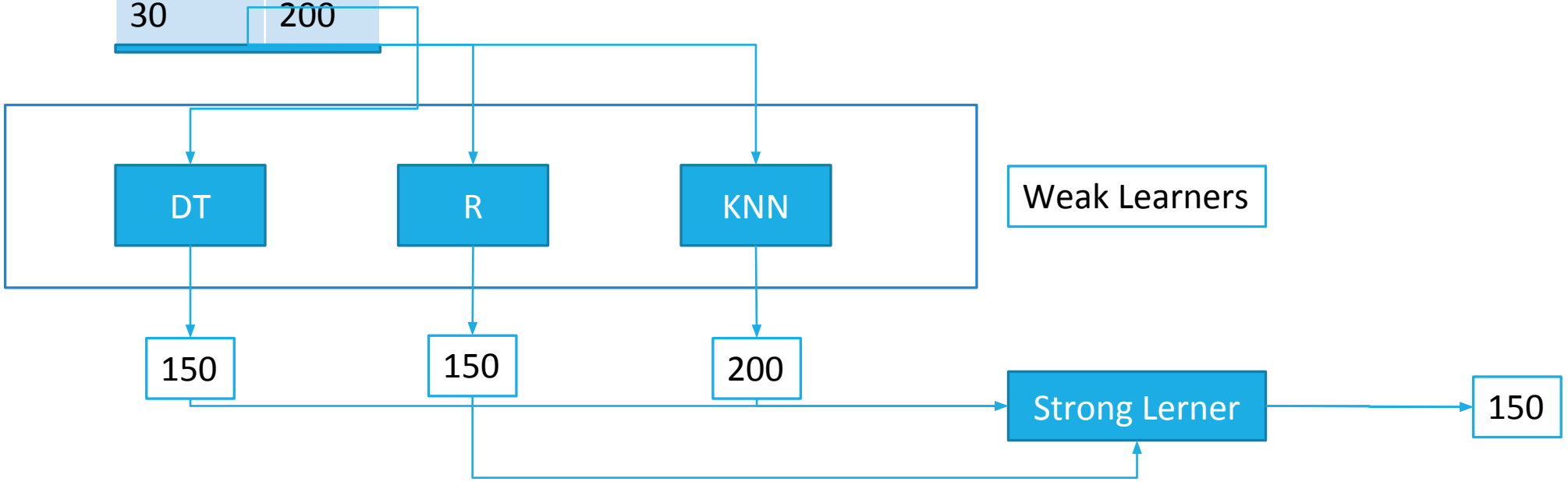
Ensemble Learning

Ensemble learning is the process by which multiple models, such as classifiers or experts, are strategically generated and combined to solve a particular computational intelligence problem.

Ensemble Learning: Max Voting

Weight	Price
10	100
20	150
30	200

Weight = 28



Probability

Tossing a Coin

H | T

Total number of events: 2

Probabilities:

H = ½

T = ½

Probability is the branch of mathematics concerning numerical descriptions of how likely an event is to occur or how likely it is that a proposition is true. Probability is a number between 0 and 1, where, roughly speaking, 0 indicates impossibility and 1 indicates certainty

Conditional Probability

Conditional probability is the probability of one event occurring with some relationship to one or more other events. For example:

Event A is that it is raining outside, and it has a 0.3 (30%) chance of raining today.

Event B is that you will need to go outside, and that has a probability of 0.5 (50%).

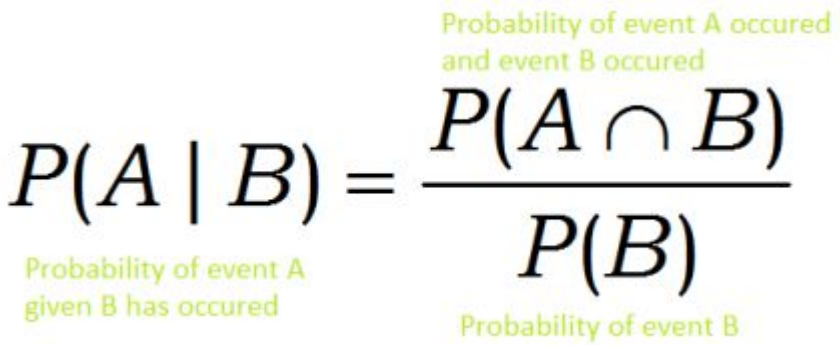
A conditional probability would look at these two events in relationship with one another, such as the probability that it is both raining and you will need to go outside.

The formula for conditional probability is:

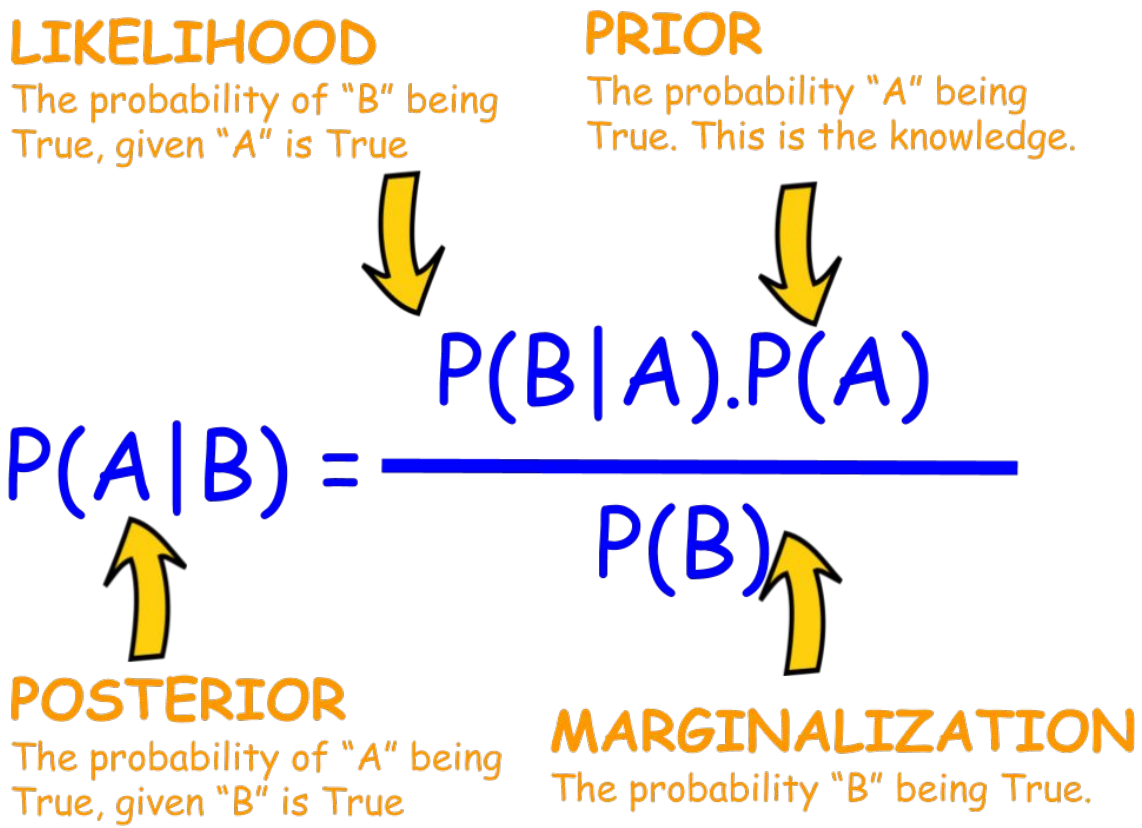
P(B|A) = P(A and B) / P(A)

which you can also rewrite as:

P(B|A) = P(A∩B) / P(A)



Bayes Theorem



Reinforcement Learning Analogy

Consider the scenario of teaching a dog new tricks.

The dog doesn't understand our language, so we can't tell him what to do.

Instead, we follow a different strategy.

We emulate a situation.

How Reinforcement Learning Works?

Your dog is an "agent" that is exposed to the **environment**. The environment could be your house, with you.

The situations they encounter are analogous to a **state**. An example of a state could be your dog standing and you use a specific word in a certain tone in your living room

Our agents react by performing an **action** to transition from one "state" to another "state," your dog goes from standing to sitting, for example.

After the transition, they may receive a **reward** or **penalty** in return. You give them a treat! Or a "No" as a penalty.

The **policy** is the strategy of choosing an action given a state in expectation of better outcomes.

The Reinforcement Learning Process

Some Useful Links

Topic	Link
Supervised Learning	https://youtu.be/WKqshCFxX-E
Decision Tree Algorithm	https://youtu.be/LymTZR-aeQg
Regression Analytics	https://youtu.be/lfsO05GJhpE
Unsupervised Learning	https://youtu.be/TQvzUvzG9as
Apriori Algorithm	https://youtu.be/hoyBSnFM-Bo
Game Theory Tutorial	Part1: https://youtu.be/CYE-NtTP5vw Part 2: https://youtu.be/_rde-0VwMkY Part 3: https://youtu.be/fWPNJUmvJXY
Webinar Recording	https://youtu.be/kTPyo7KKM7Q

Thank You!

EXAMPLE: SELF DRIVING CAB

KUSHAL@INDEEDINSPIRING.COM