

# What is Machine Learning (ML)?

Applied Machine Learning with R

[www.therbootcamp.com](http://www.therbootcamp.com)

@therbootcamp

January 2019

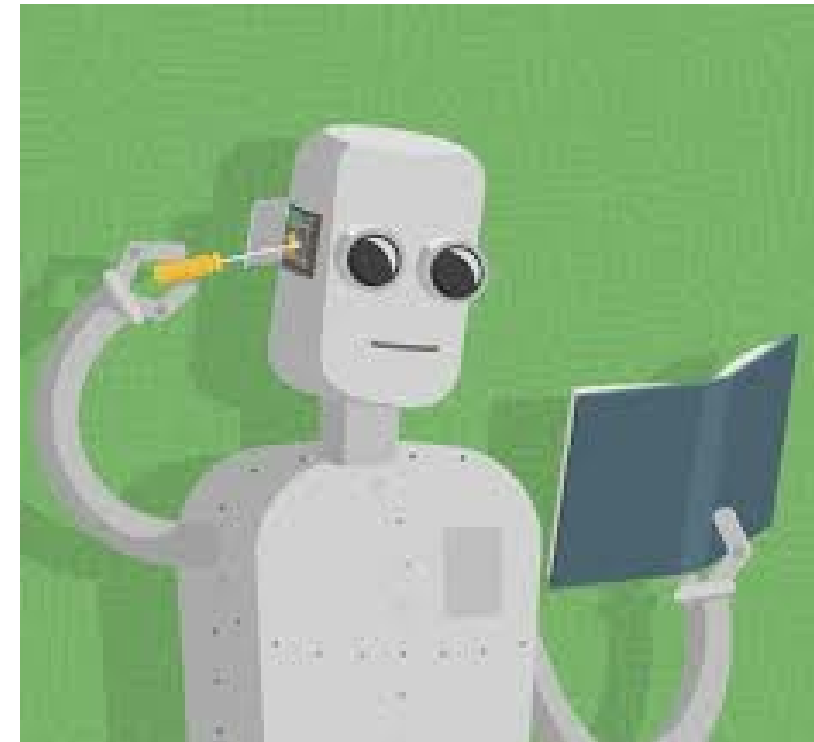
# What do you think?

No Googling :)

# What is Machine Learning?

Machine learning (ML) is a **field of artificial intelligence** that uses **statistical techniques** to give computer systems the ability to **"learn"** (e.g., progressively improve **performance** on a specific task) **from data**, without being explicitly programmed.

[Wikipedia](#)



[www.towardsdatascience.com](http://www.towardsdatascience.com)

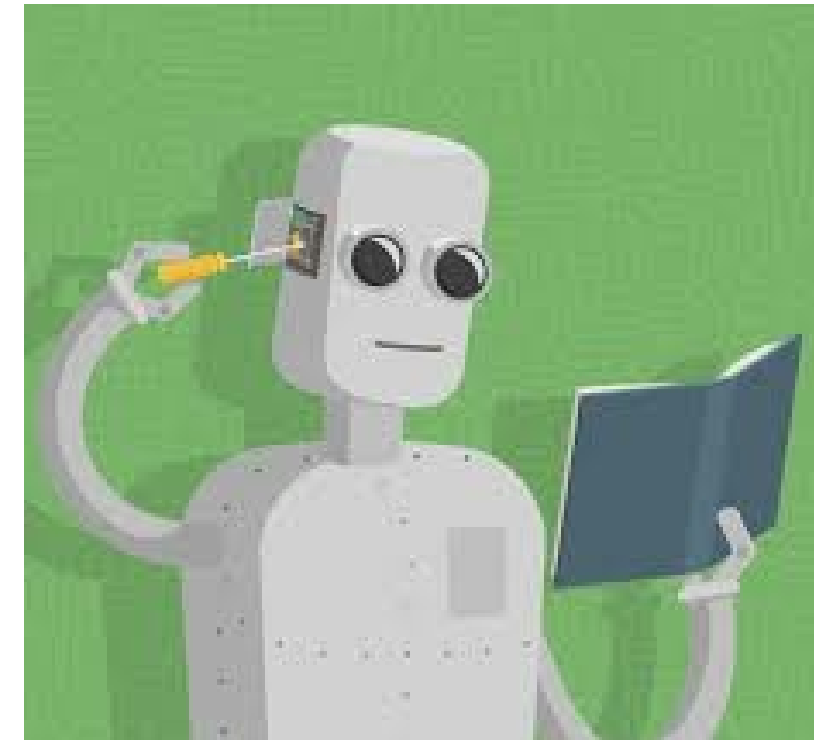
# What is Machine Learning?

Machine learning (ML) is a **field of artificial intelligence** that uses **statistical techniques** to give computer systems the ability to **"learn"** (e.g., progressively improve **performance** on a specific task) **from data**, without being explicitly programmed.

[Wikipedia](#)

## What does this mean?

- Machine Learning is about making decisions.
- To do machine learning, you need (clean, relevant) data!
- ML algorithms try to automatically figure out which information is important and which is not.
- ML algorithms are always guided by a



[www.towardsdatascience.com](http://www.towardsdatascience.com)

# Examples

## Predicting Heart Attacks

You are an intake nurse at an emergency room.

A patient comes in complaining of chest pain and thinks they are having a heart attack



[ucsf.edu](https://www.ucsf.edu)

## Predicting sales

You are an analyst at a retail corporation.

The executive team is considering whether or not to open a new retail location in Basel.



[location.ch](https://www.location.ch)

# Solution 1: Intuition

Many decisions are based on an individual's intuitive 'gut' judgment.

"I have a feeling this patient is having a heart attack"

"Sales will be at least 100,000 CHF per month, trust me on this."

## What are potential problems with intuition?



[ucsf.edu](https://www.ucsf.edu)



[location.ch](https://www.location.ch)

# Solution 1: Intuition

## What are potential problems with intuition?

- Might not tell you anything about      it makes a prediction.
- Could be based on reasons other than accuracy (e.g.; self protection)
- Without rigorous testing, impossible to know if **critical information is being ignored**.
- If the decision isn't programmed, it might be **impossible to replicate** (and improve) in the future.
- Intuition is rarely consistently tracked and evaluated. When wrong, intuition can always be defended 'in hindsight' (ML too!).



ucsf.edu




location.ch

One very influential world leader really trusts his intuition...








## PRESIDENT TRUMP ON HIS INTUITION

"...I have a gut, and my gut tells me more sometimes than anybody else's brain can ever tell me."

November 27, 2018 | The Washington Post

**DON'S TAKE**

**PRESIDENT TRUMP IS GOVERNING BY NOT GOVERNING**



10:09 PM ET

**DON LEMON**

# Solution 2: Data-Driven Machine Learning

Here's how a data-driven, ML approach would look:

	diagnosis	age	sex	cp	trestbps	chol	fbs	restecg
1	FALSE	63	1	ta	145	233	1	hypertrophy
2	TRUE	67	1	a	160	286	0	hypertrophy
3	TRUE	67	1	a	120	229	0	hypertrophy
4	FALSE	37	1	np	130	250	0	normal
5	FALSE	41	0	aa	130	204	0	hypertrophy
6	FALSE	56	1	aa	120	236	0	normal
7	TRUE	62	0	a	140	268	0	hypertrophy
8	FALSE	57	0	a	120	354	0	normal
9	TRUE	63	1	a	130	254	0	hypertrophy
10	TRUE	53	1	a	140	203	1	hypertrophy
11	FALSE	57	1	a	140	192	0	normal

Based on **historical data** from past patients **at this hospital**, a **regression model**, using the patient's **age, cholesterol level, and ecg**, **predicts** the probability that this patient is having a heart attack is only **20%**.



gstatic.com



WK Digital

# Solution 2: Data-Driven Machine Learning

## What are the benefits of ML?

- Algorithms fit to your past data can tell you which variables are important and which are not.
- Make explicit, quantitative predictions of variables of interest.
- Many can give you probability estimates, and estimated errors, rather than single decisions or point estimates.
- Can reveal novel insights about your data!
- Can be programmed and automated.

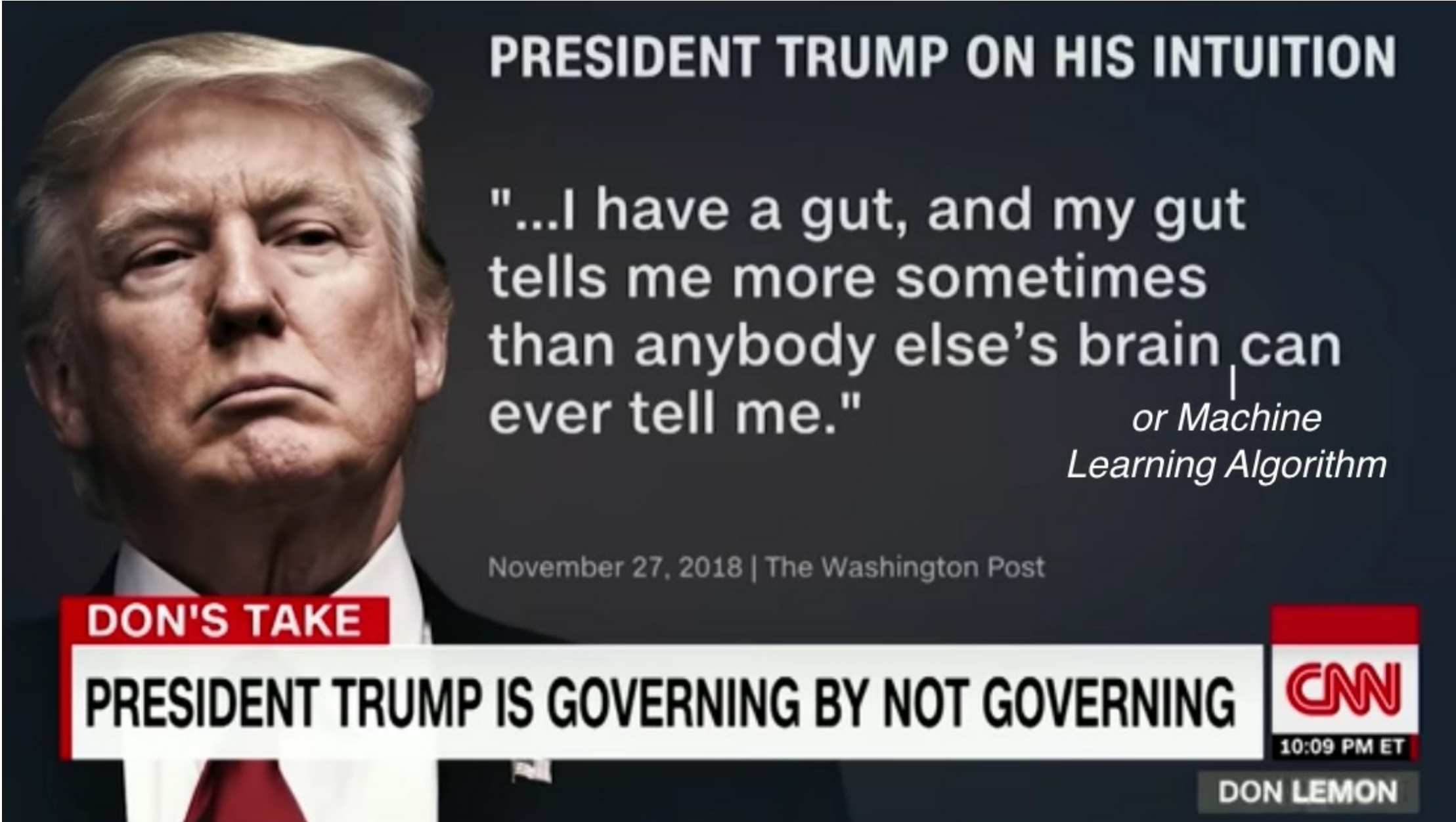


gstatic.com



WK Digital





**PRESIDENT TRUMP ON HIS INTUITION**

"...I have a gut, and my gut tells me more sometimes than anybody else's brain can ever tell me."

*or Machine Learning Algorithm*

November 27, 2018 | The Washington Post

**DON'S TAKE**

**PRESIDENT TRUMP IS GOVERNING BY NOT GOVERNING**

**CNN**  
10:09 PM ET

**DON LEMON**

# Data Terminology

Term	Definition	Example
Data (Tidy)	Data represented in a rectangular format with rows and columns	Excel spreadsheet, .CSV
Case	A specific observation of data	A patient, a site
Feature	An individual, measurable property of cases	Age, temperature, country
Criterion	Something you want to predict	Total sales, success (yes or no)

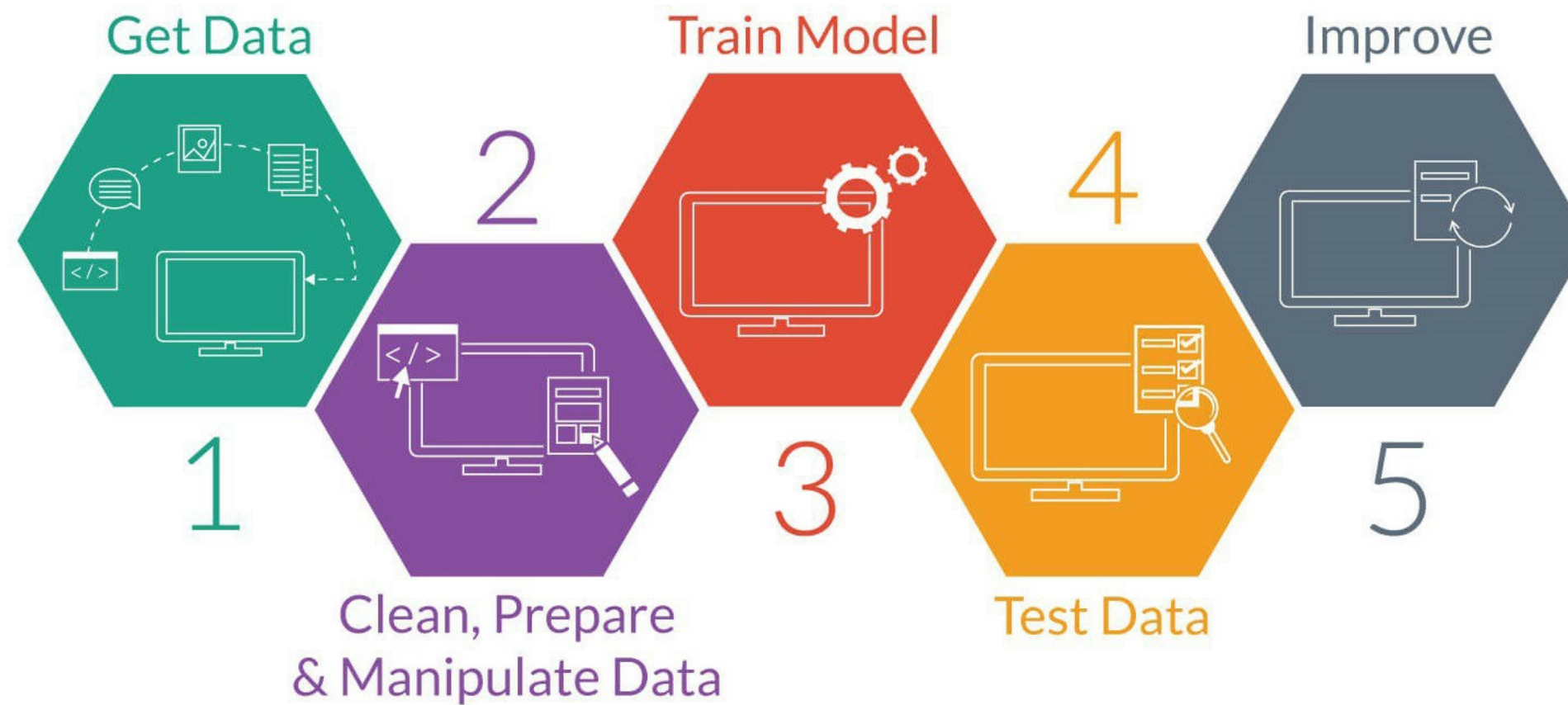
**Criterion**      **Features**

↓      ↙      ↓      ↘

**Cases** →

	diagnosis	age	sex	cp	trestbps	chol	fbs
1	FALSE	63	1	ta	145	233	1
2	TRUE	67	1	a	160	286	0
3	TRUE	67	1	a	120	229	0
4	FALSE	37	1	np	130	250	0
5	FALSE	41	0	aa	130	204	0
6	FALSE	56	1	aa	120	236	0
7	TRUE	62	0	a	140	268	0
8	FALSE	57	0	a	120	354	0
9	TRUE	63	1	a	130	254	0
10	TRUE	53	1	a	140	203	1
11	FALSE	57	1	a	140	192	0

# What are the steps of machine learning?



Medium.com

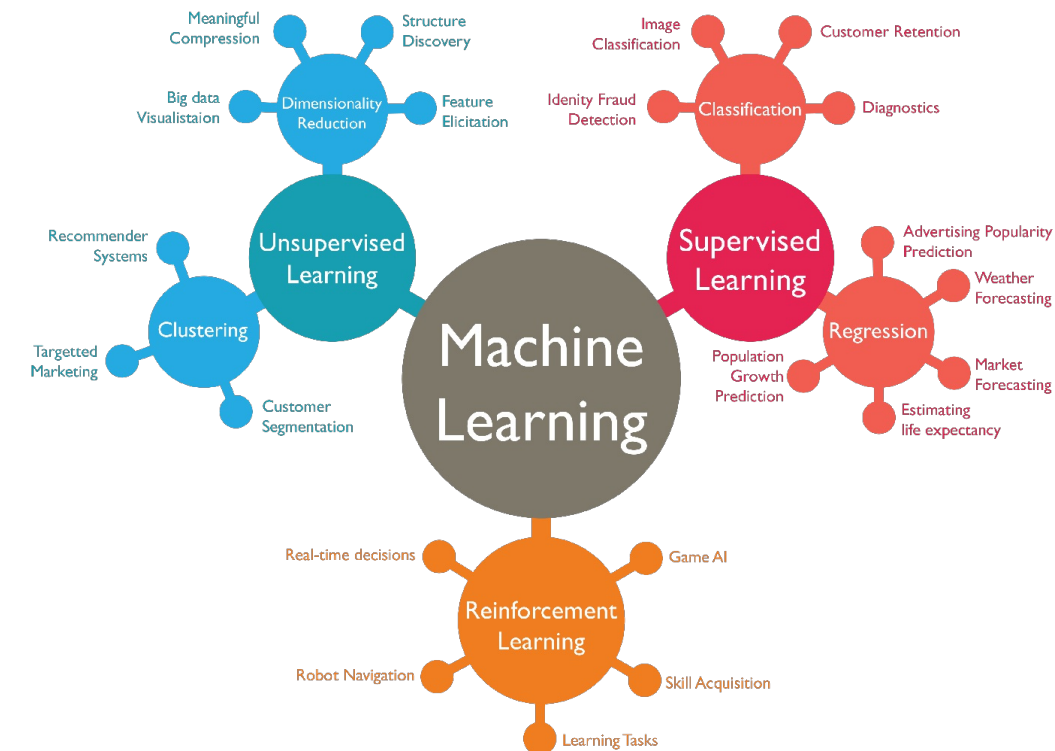
# Types of machine learning tasks

There are many types of machine learning tasks, each of which call for different models.

We will focus on **Supervised** tasks.

Three general categories

Type	Example
Reinforcement Learning	Robot navigation, game playing AI
Unsupervised	Customer segmentation
Supervised	Classification, regression



Wordstream.com



# Types of machine learning tasks

There are many types of machine learning tasks, each of which call for different models.

We will focus on **Supervised** tasks.

Three general categories

Type	Example
Reinforcement Learning	Robot navigation, game playing AI
Unsupervised	Customer segmentation
Supervised	Classification, regression

## Reinforcement Learning



[towardsdatascience.com](https://towardsdatascience.com)



MIT Technology Review

# Types of machine learning tasks

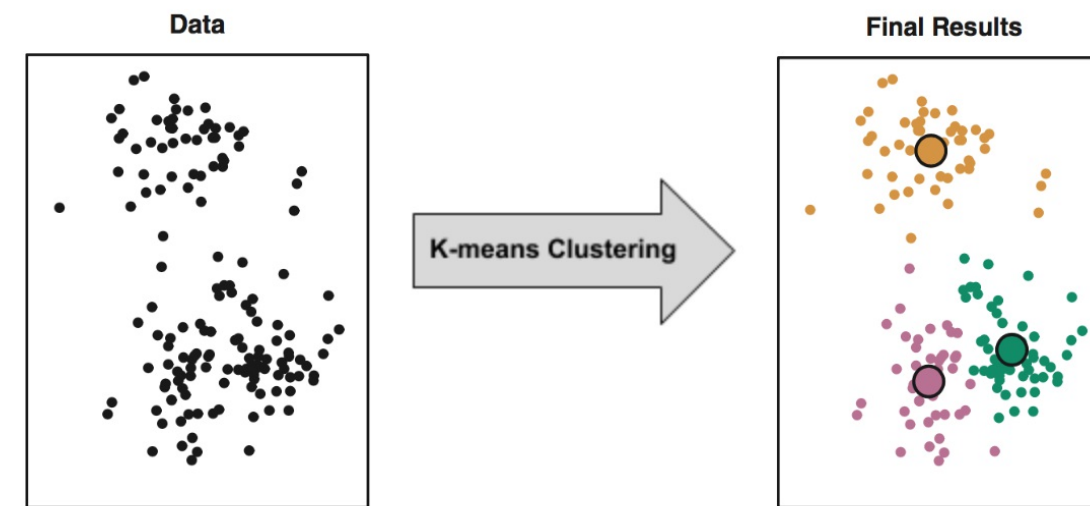
There are many types of machine learning tasks, each of which call for different models.

We will focus on **Supervised** tasks.

Three general categories

Type	Example
Reinforcement Learning	Robot navigation, game playing AI
Unsupervised	Customer segmentation
Supervised	Classification, regression

## Unsupervised learning



iotforall.com

# Types of machine learning tasks

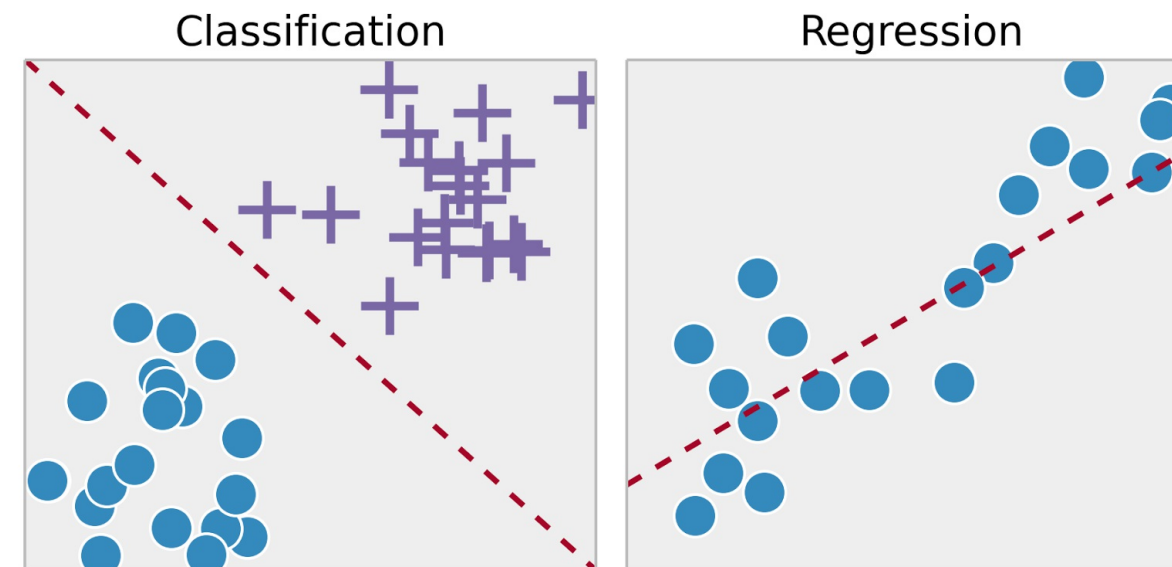
There are many types of machine learning tasks, each of which call for different models.

We will focus on **Supervised** tasks.

Three general categories

Type	Example
Reinforcement Learning	Robot navigation, game playing AI
Unsupervised	Customer segmentation
Supervised	Classification, regression

## Supervised learning



# Types of machine learning tasks

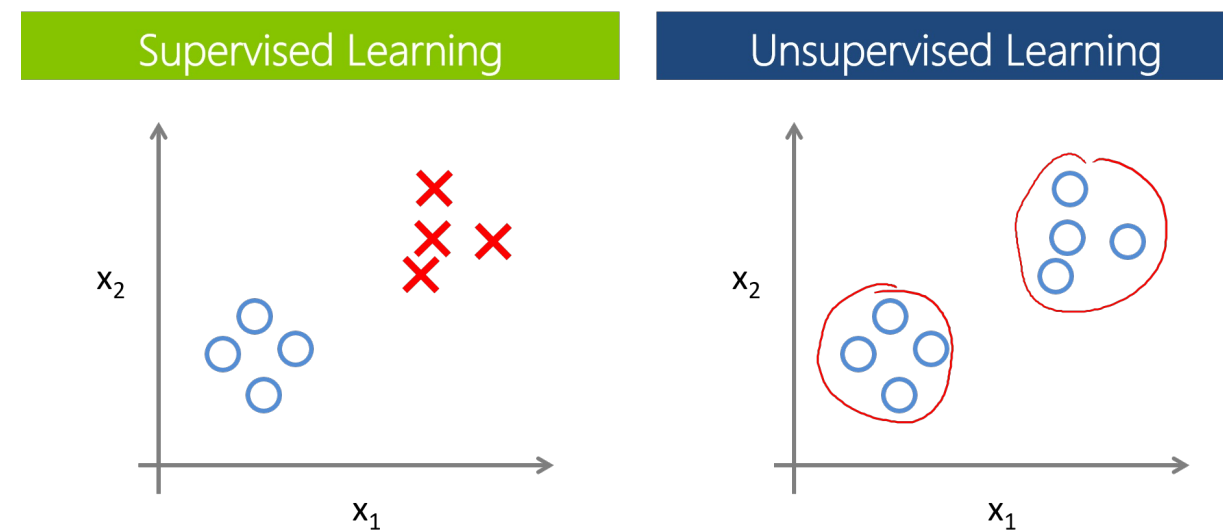
There are many types of machine learning tasks, each of which call for different models.

We will focus on **Supervised** tasks.

Three general categories

Type	Example
Reinforcement Learning	Robot navigation, game playing AI
Unsupervised	Customer segmentation
Supervised	Classification, regression

## Unsupervised vs. Supervised learning



# What is a model?

# What is a model?

It's not magic.

# What is a model?

It's not magic.

A model is a **mathematical** (computational) representation of data.

# What is a model?

It's not magic.

A model is a **mathematical** (computational) representation of data.

That can be **programmed**, and used to understand, and **predict relationships** between data



# Three of the models we will work with

## Regression

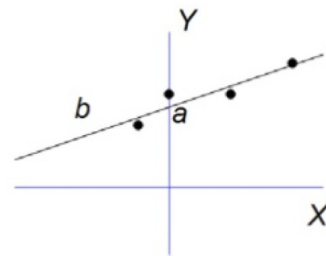
Linear regression equation  
(without error)

$$\hat{Y} = bX + a$$

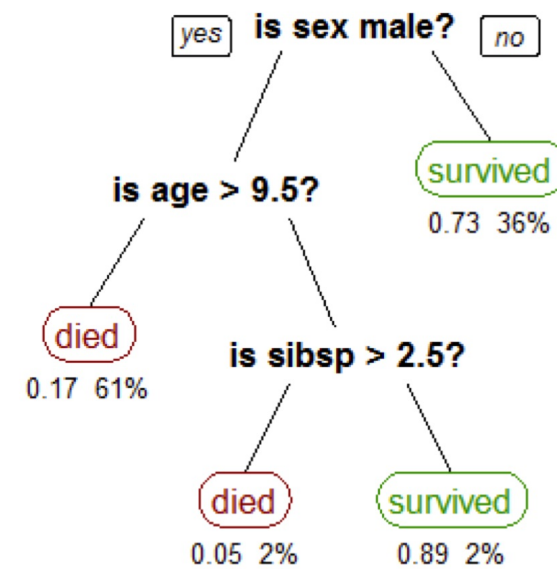
predicted values of Y

$b$  = slope = rate of predicted  $\uparrow/\downarrow$  for Y scores for each unit increase in X

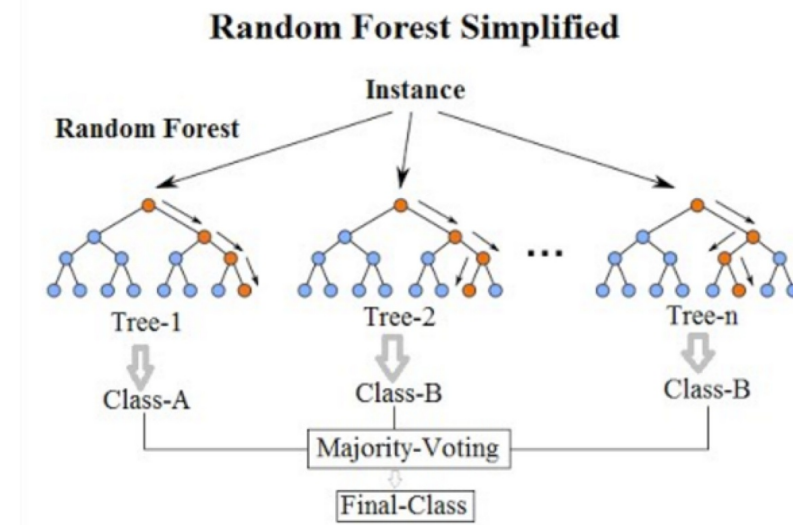
Y-intercept = level of Y when X is 0



## Decision Tree



## Random Forest



# Why do we need a model at all?

## Theoretical

Which features are most **important** in predicting the criterion and which can be safely ignored?

What is the **relationship** between features and the criterion?

- Strong? Weak? Positive? Negative?

Make **predictions** for new data.

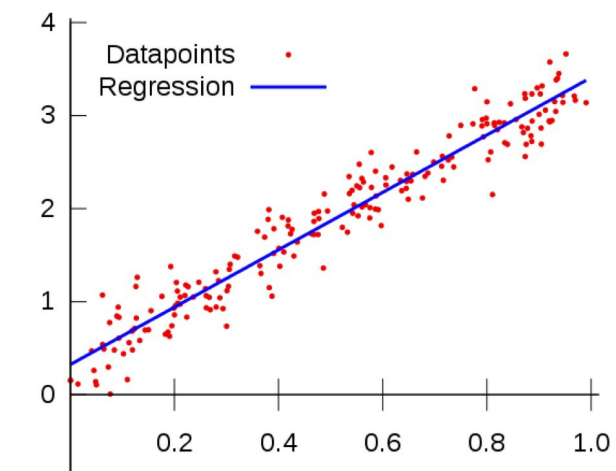
- Is this patient having a heart attack?
- How much will this new site sell?

Improve predictions over time with new data.

## Data

	diagnosis	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang
1	FALSE	63	1	ta	145	233	1	hypertrophy	150	0
2	TRUE	67	1	a	160	286	0	hypertrophy	108	1
3	TRUE	67	1	a	120	229	0	hypertrophy	129	1
4	FALSE	37	1	np	130	250	0	normal	187	0
5	FALSE	41	0	aa	130	204	0	hypertrophy	172	0
6	FALSE	56	1	aa	120	236	0	normal	178	0
7	TRUE	62	0	a	140	268	0	hypertrophy	160	0
8	FALSE	57	0	a	120	354	0	normal	163	1
9	TRUE	63	1	a	130	254	0	hypertrophy	147	0
10	TRUE	53	1	a	140	203	1	hypertrophy	155	1
11	FALSE	57	1	a	140	192	0	normal	148	0
12	FALSE	56	0	aa	140	294	0	hypertrophy	153	0

## Model



# Why do we need a model at all?

## Practical

Need a computational / mathematical representation of the data that can be programmed and **productionised**.

A model is like a **step-by-step recipe**, telling you exactly what data you need, and how to combine it.

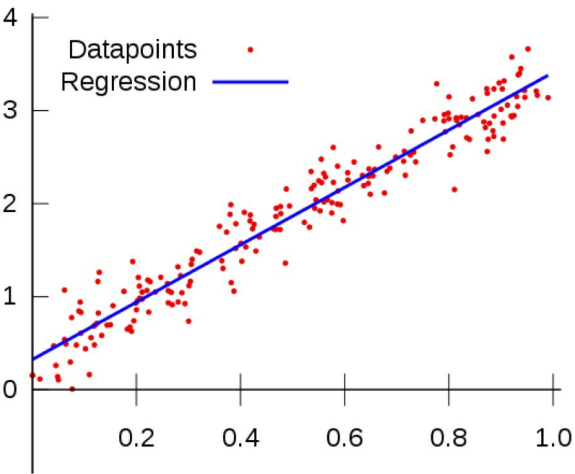


amazon.com

## Data

	diagnosis	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang
1	FALSE	63	1	ta	145	233	1	hypertrophy	150	0
2	TRUE	67	1	a	160	286	0	hypertrophy	108	1
3	TRUE	67	1	a	120	229	0	hypertrophy	129	1
4	FALSE	37	1	np	130	250	0	normal	187	0
5	FALSE	41	0	aa	130	204	0	hypertrophy	172	0
6	FALSE	56	1	aa	120	236	0	normal	178	0
7	TRUE	62	0	a	140	268	0	hypertrophy	160	0
8	FALSE	57	0	a	120	354	0	normal	163	1
9	TRUE	63	1	a	130	254	0	hypertrophy	147	0
10	TRUE	53	1	a	140	203	1	hypertrophy	155	1
11	FALSE	57	1	a	140	192	0	normal	148	0
12	FAI SF	56	0	aa	140	294	0	hypertrophv	153	0

↓  
**Model**  
↓



Questions?

Schedule