

# **DATA HACK 2023 MAKER WORKSHOP: JAVASCRIPT PROGRAMMING, DATA VISUALISING, AND MACHINE LEARNING - PART II**

Bernard Suen  
Center for Entrepreneurship  
Chinese University of Hong Kong

**Today's agenda.**

1. Tying up previous loose ends: **Github** and **WordPress**.
2. **Data Science** approach to **AI**, **Machine Learning**, and **Deep Learning** using the **CRISP-DM** data cycle.
3. Use of the **methodology map** for determining the appropriate approach to tackle a data problem.
4. The machine learning **pipeline**.
5. Introduction to the **Teachable Machine** for getting started with supervised learning and deep learning.
6. Use of Teachable Machine trained models in **ML5.js**/  
**P5.js** for ML application and deployment.

**Use [Github](#) to host your  
web application.**



Search or jump to...

Pull requests Issues Codespaces Marketplace Explore

Notifications + Profile

suentze2020 / datahack Public

Pin

Unwatch 1

Fork 1

Star 0

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

main

1 branch

0 tags

Go to file

Add file

<> Code

suentze2020 Update README.md

● 48aed66 2 minutes ago 28 commits

html\_css\_workshop

DataHack Workshop Work Files

yesterday

js\_workshop\_i

Datahack Workshop work files

yesterday

js\_workshop\_ii

DataHack Workshop Work Files

yesterday

sql\_workshop

DataHack Workshop Work Files

yesterday

web\_scraping\_workshop

DataHack Workshop Work Files

yesterday

wordpress\_workshop

DataHack Workshop Work Files

yesterday

.DS\_Store

Datahack Workshop work files

yesterday

.gitattributes

Initial commit

last week

README.md

Update README.md

2 minutes ago

README.md



## datahack

<https://github.com/suentze2020/datahack>

About



No description, website, or topics provided.

Readme

0 stars

1 watching

1 fork

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Environments 1

github-pages In progress



Search or jump to...

Pull requests Issues Codespaces Marketplace Explore



suentze2020 / datahack Public

Pin

Unwatch 1

Fork 1

Star 0

<> Code Issues Pull requests Actions Projects Wiki Security Insights Settings

## General

Access

Collaborators

Moderation options

Code and automation

Branches

Tags

Actions

Webhooks

Environments

Codespaces

Pages

Security

Code security and analysis

Deploy keys

## General

Repository name

datahack

Rename

**Template repository**

Template repositories let users generate new repositories with the same directory structure and files. [Learn more](#).

**Require contributors to sign off on web-based commits**

Enabling this setting will require contributors to sign off on commits made through GitHub's web interface. Signing off is a way for contributors to affirm that their commit complies with the repository's terms, commonly the [Developer Certificate of Origin \(DCO\)](#). [Learn more about signing off on commits](#).

## Social Preview

Upload an image to customize your repository's social media preview.

Images should be at least 640x320px (1280x640px for best display).

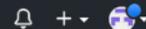
[Download template](#)





Search or jump to...

Pull requests Issues Codespaces Marketplace Explore



suentze2020 / datahack Public

Pin Unwatch 1 Fork 1 Star 0

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

## General

### Access

### Collaborators

### Moderation options

### Code and automation

#### Branches

#### Tags

#### Actions

#### Webhooks

#### Environments

#### Codespaces

#### Pages

### Security

#### Code security and analysis

#### Deploy keys

## GitHub Pages

GitHub Pages is designed to host your personal, organization, or project pages from a GitHub repository.

Your site is live at <https://suentze2020.github.io/datahack/>

Last deployed by suentze2020 yesterday

Visit site

...

## Build and deployment

### Source

Deploy from a branch ▾

### Branch

Your GitHub Pages site is currently being built from the main branch. [Learn more](#).

main ▾

/ (root) ▾

Save

Learn how to [add a Jekyll theme](#) to your site.

Your site was last deployed to the [github-pages](#) environment by the [pages build and deployment](#) workflow.

[Learn more about deploying to GitHub Pages using custom workflows](#)

**Embed [Github](#) website inside a [WordPress](#) website using [iframe](#).**

[Tutorials ▾](#)[References ▾](#)[Exercises ▾](#)[Sign Up](#)[Pro](#)[Get Certified](#)[HTML](#)[CSS](#)[JAVASCRIPT](#)[SQL](#)[PYTHON](#)[JAVA](#)[PHP](#)[BOOTSTRAP](#)[HOW TO](#)[W3.CSS](#)[C](#)[C++](#)[C#](#)[REACT](#)[R](#)[JQUERY](#)[DJANGO](#)

## HTML Reference

[HTML by Alphabet](#)[HTML by Category](#)[HTML Browser Support](#)[HTML Attributes](#)[HTML Global Attributes](#)[HTML Events](#)[HTML Colors](#)[HTML Canvas](#)[HTML Audio/Video](#)[HTML Character Sets](#)[HTML Doctypes](#)[HTML URL Encode](#)[HTML Language Codes](#)[HTML Country Codes](#)[HTTP Messages](#)[HTTP Methods](#)[PX to EM Converter](#)[Keyboard Shortcuts](#)

## HTML Tags

[`<!-->`](#)[`<!DOCTYPE>`](#)

# HTML <iframe> Tag

[◀ Previous](#)[Complete HTML Reference](#)[Next ▶](#)

## Example

An inline frame is marked up as follows:

```
<iframe src="https://www.w3schools.com" title="W3Schools Free Online Web Tutorials"></iframe>
```

[Try it Yourself »](#)

More "Try it Yourself" examples below.

[https://www.w3schools.com/tags/tag\\_iframe.ASP](https://www.w3schools.com/tags/tag_iframe.ASP)

W Datahack 2023 0 + New View Page Howdy, admin

Dashboard Posts Media Post Grid Combo Pages All Pages Add New

Comments Restaurants Appearance Plugins Users Tools Settings Pods Admin All Export WP Data Access

Ultimate CSV Importer Free FakerPress Collapse menu

## Edit Page [Add New](#)

Page updated. [View page](#)

### Data Visualization Demonstration

Permalink: <https://dev-hackathon-2023.pantheonsite.io/embedded-website/> [Edit](#)

[Add Media](#) [Pods Shortcode](#) Visual Text

```
<iframe style="width: 100vw !important; height: 100vh;" title="Data Visualization Demonstration" src="https://suentze2020.github.io/datahack/js_workshop_i/data_viz/dataviz04.html" frameborder="0" scrolling="no"></iframe>
```

Word count: 0 Last edited by admin on February 9, 2023 at 9:15 am

**Publish** [Preview Changes](#)

- Status: Published [Edit](#)
- Visibility: Public [Edit](#)
- Rewrites: 33 [Browse](#)
- Published on: Feb 8, 2023 at 07:35 [Edit](#)

[Move to Trash](#) [Update](#)

**Page Attributes**

Parent: (no parent) Order: 0

Need help? Use the Help tab above the screen title.

**Editor** [Switch to block editor](#)

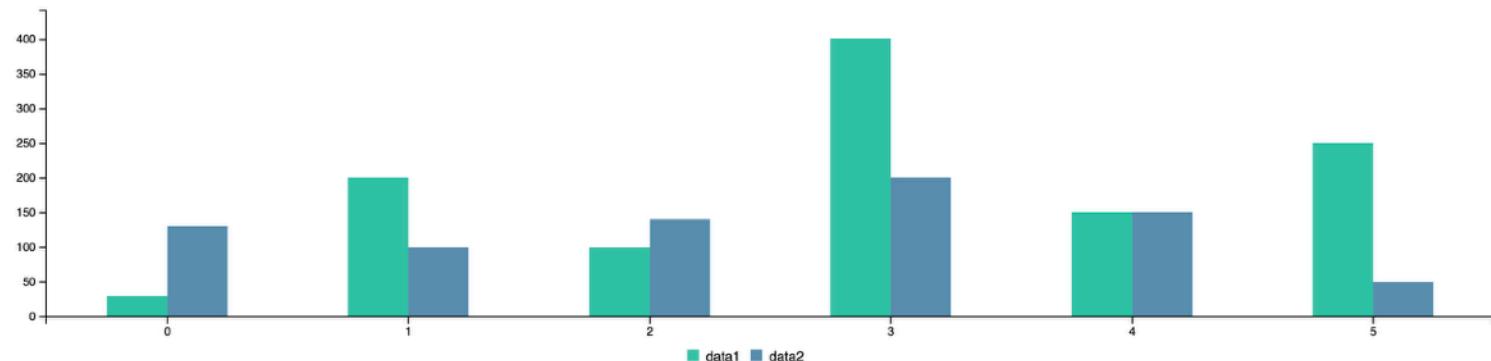
**Featured image**

Datahack 2023 

Home Privacy Policy Sample Page Post Listing Restaurant List Data Visualization Demonstration

## Data Visualization Demonstration

Top navbar Demo 1 Demo 2 Demo 3 Demo 4



**Free resources for improving  
your web publishing.**

[Photos](#)[Illustrations](#)[Vectors](#)[Videos](#)[Music](#)

...

[Explore](#) ▾[Sign up](#)

# Stunning free images & royalty free stock

Over 1.7 million+ high quality stock images and videos shared by our talented community.

[Images](#) ▾

Popular images: easter, wallpaper, coronavirus, nature, corona, love, business, food, spring, money, covid, computer

Free image by mac231

<https://pixabay.com/>

Pick Color From Image      Color Picker

...pick your color online

November 23 - 25, 2021 

OPEN

Ad



PALETTE:



Show more

Use Your Image

We think data protection is important! [No data is sent.](#) The magic happens in your browser.

<https://imagecolorpicker.com/>

**! NEW FEATURE:** You can now create a gradient out of 3 colors!



Beta

ColorSpace

NEW

PALETTES

GRADIENT

3-COLOR-GRADIENT

CONTACT INFO



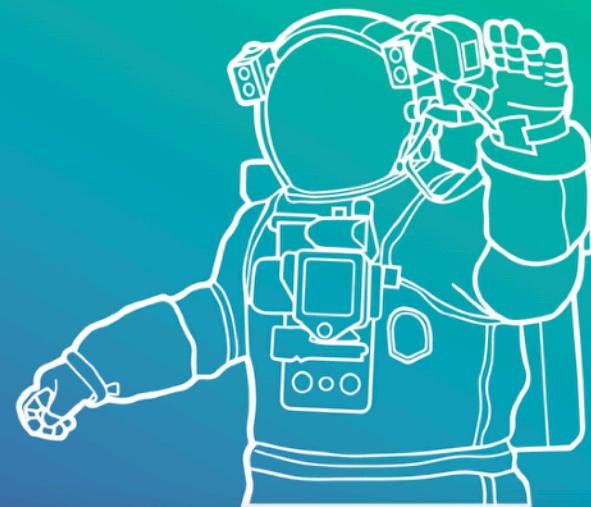
# Never waste Hours on finding the perfect Color Palette again!

## JUST ENTER A COLOR!

And Generate nice Color Palettes

#59AB6E

GENERATE



<https://mycolor.space/>



chrome web store



suentze2020@gmail.com ▾

[Home](#) > [Extensions](#) > [ColorZilla](#)



# ColorZilla

[colorzilla.com](http://colorzilla.com)

Featured

2,729 [i](#)

| [Developer Tools](#) | 3,000,000+ users

[Remove from Chrome](#)

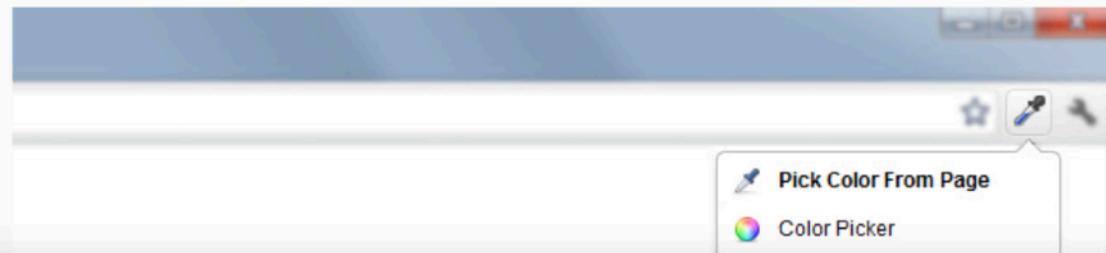
[Overview](#)

[Privacy practices](#)

[Reviews](#)

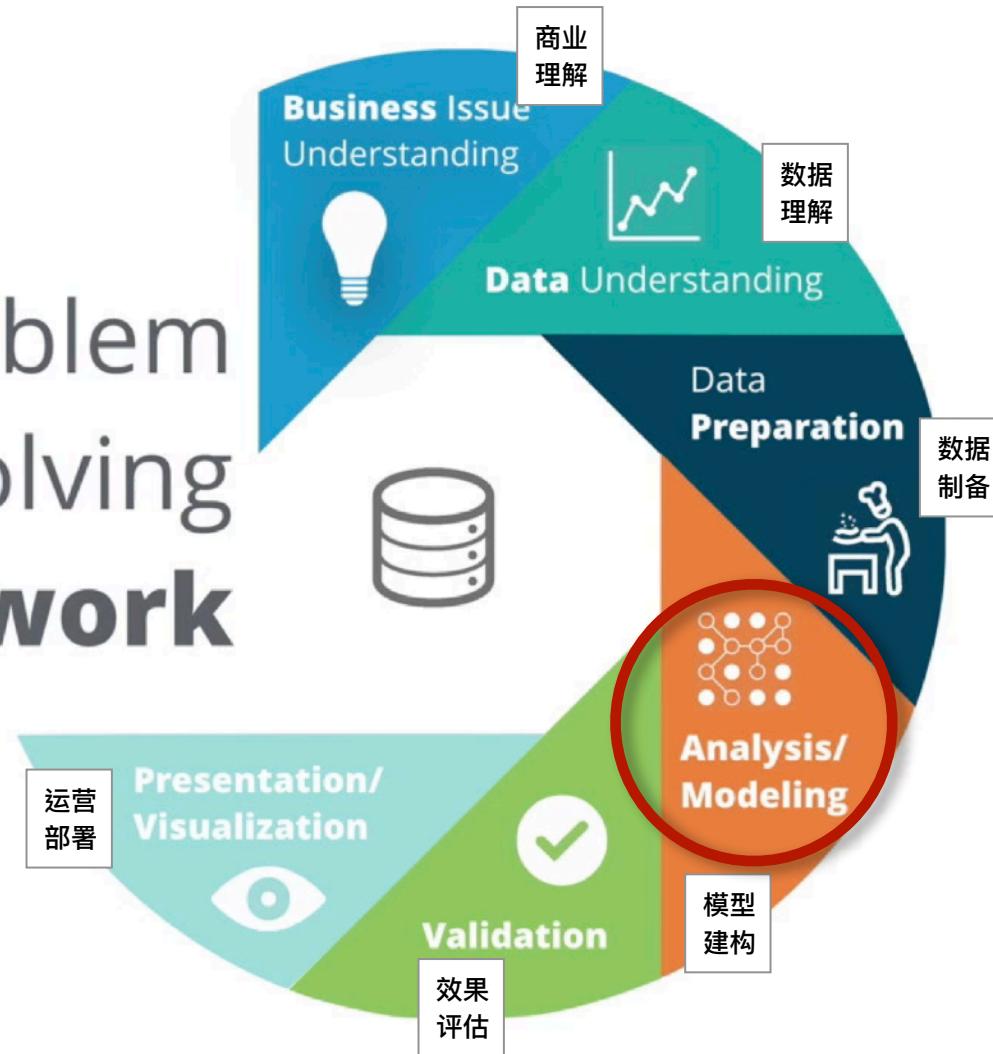
[Support](#)

[Related](#)



# **Data Science Approach to AI, Machine Learning and Deep Learning.**

# Problem Solving Framework



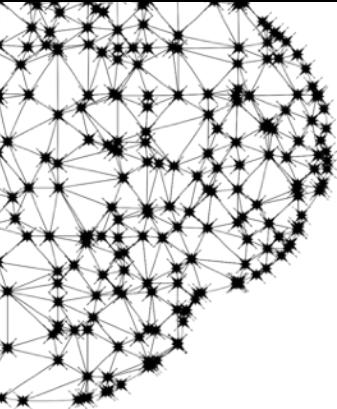
**Which analytics model to use? And why?**



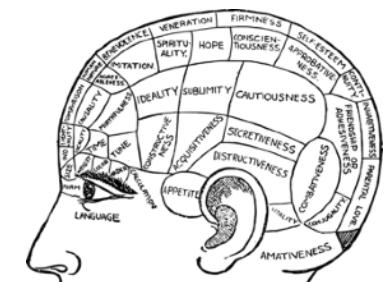
# **The Methodology Map**

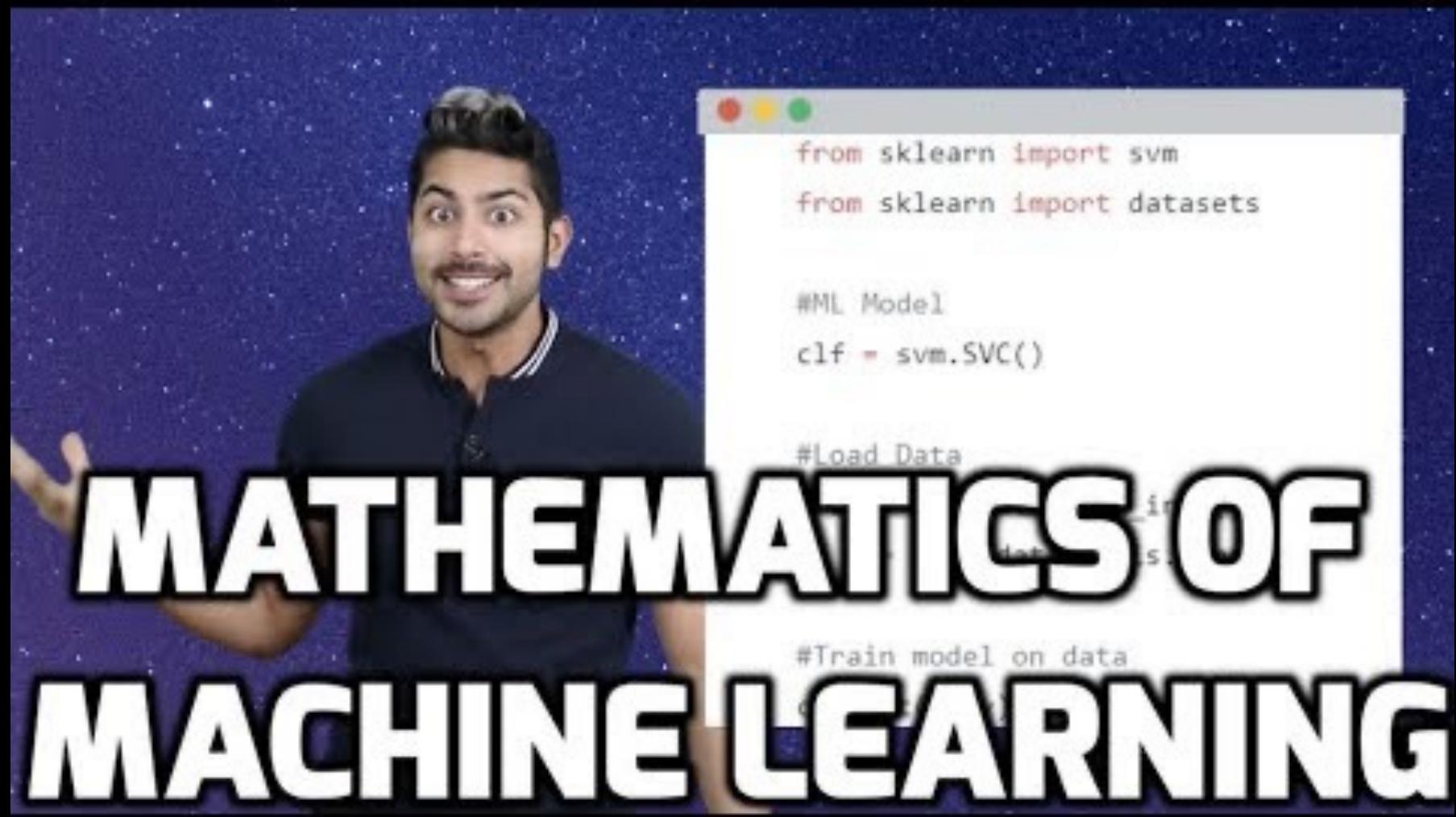
聚焦未来					聚焦过去和现在
Predict Outcome					Data Analysis
Data Rich			Data Poor	Geospatial	
Numeric		Classification		A/B Test	Segmentation
Continuous	Count	Binary	Non Binary	<ul style="list-style-type: none"> <li>Inferential Statistics (t-test, Chi square, etc.)</li> </ul>	Aggregation
<ul style="list-style-type: none"> <li>Linear Regression,</li> <li>Multiple Regression</li> </ul>	<ul style="list-style-type: none"> <li>Count Regression</li> </ul>	<ul style="list-style-type: none"> <li>Logistic Regression</li> <li>Decision Tree</li> </ul>	<ul style="list-style-type: none"> <li>Random Forest</li> <li>K-Nearest Neighbour</li> </ul>		Descriptive
<b>Popular tools:</b> Scikit-learn, Tensorflow, PyTorch, Keras, ML5.js, Brain.js, Knime, Orange, BigQuery ML etc.				e.g. SPSS, Jamovi	e.g. SQL, SPSS, Jamovi

Source: Udacity Model Selection Methodology Map



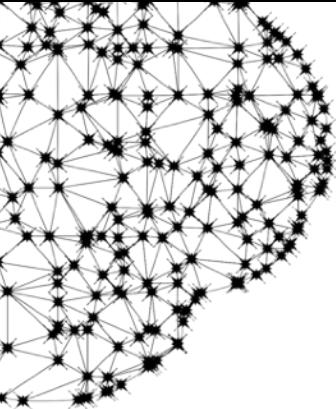
# **Mathematical Foundation of Machine Learning**



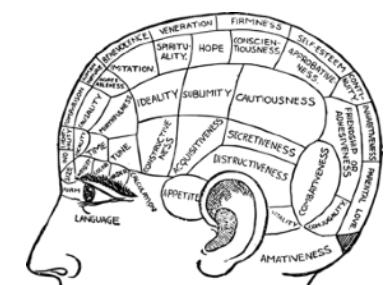


Source: <https://www.youtube.com/watch?v=8onB7rPG4Pk&t=1s>

**Does that mean you need to know all those maths to implement codes in Python to perform the machine learning functions (e.g. Least Square - linear shaped or Sigmoid function - S shaped line fitting)?**



# Well, how does the machine learn in the first place?



# Gradient descent



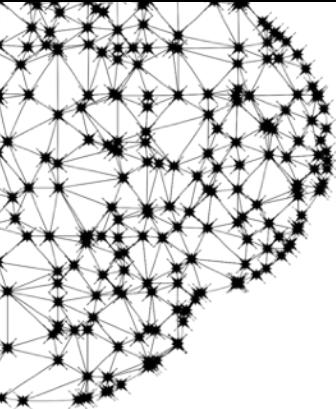
Mount Errorest



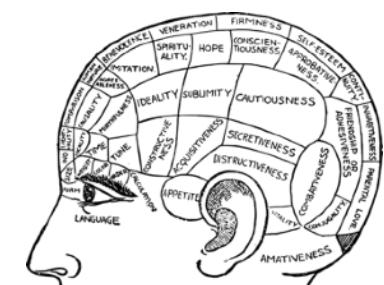
<https://www.youtube.com/watch?v=IpGxLWOIZy4>

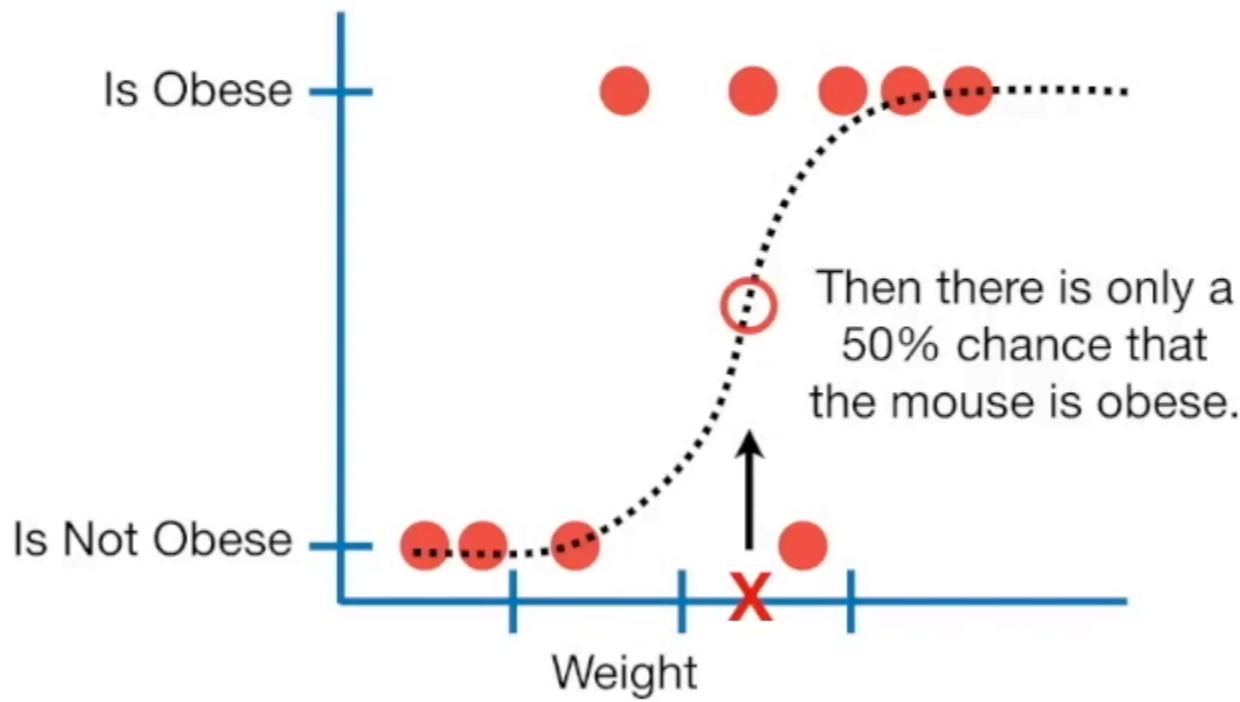
聚焦未来				聚焦过去和现在	
Predict Outcome				Data Analysis	
Data Rich		Data Poor		Geospatial	
Numeric		Classification		A/B Test	Segmentation
Continuous	Count	Binary	Non Binary	<ul style="list-style-type: none"> <li>Inferential Statistics (t-test, Chi square, etc.)</li> </ul>	Aggregation
<ul style="list-style-type: none"> <li>Linear Regression,</li> <li>Multiple Regression</li> </ul>	<ul style="list-style-type: none"> <li>Count Regression</li> </ul>	<ul style="list-style-type: none"> <li>Logistic Regression</li> <li>Decision Tree</li> </ul>	<ul style="list-style-type: none"> <li>Random Forest</li> <li>K-Nearest Neighbour</li> </ul>		Descriptive
<b>Popular tools:</b> Scikit-learn, Tensorflow, PyTorch, Keras, ML5.js, Brain.js, Knime, Orange, BigQuery ML etc.				e.g. SPSS, Jamovi	e.g. SQL, SPSS, Jamovi

Source: Udacity Model Selection Methodology Map



# From Linear Regression to Logistic Regression





**How is Machine Learning related to  
Predictive Analytics and Deep Learning?  
What are their relationships to Data Science?**

## (Supervised Learning)

### Predictive Analytics

预测分析/机器学习/深度学习

- Linear Regression
- Decision Tree
- Forest Model
- Boosted Model

- Count Regression
- Logistic Regression
- Decision Tree

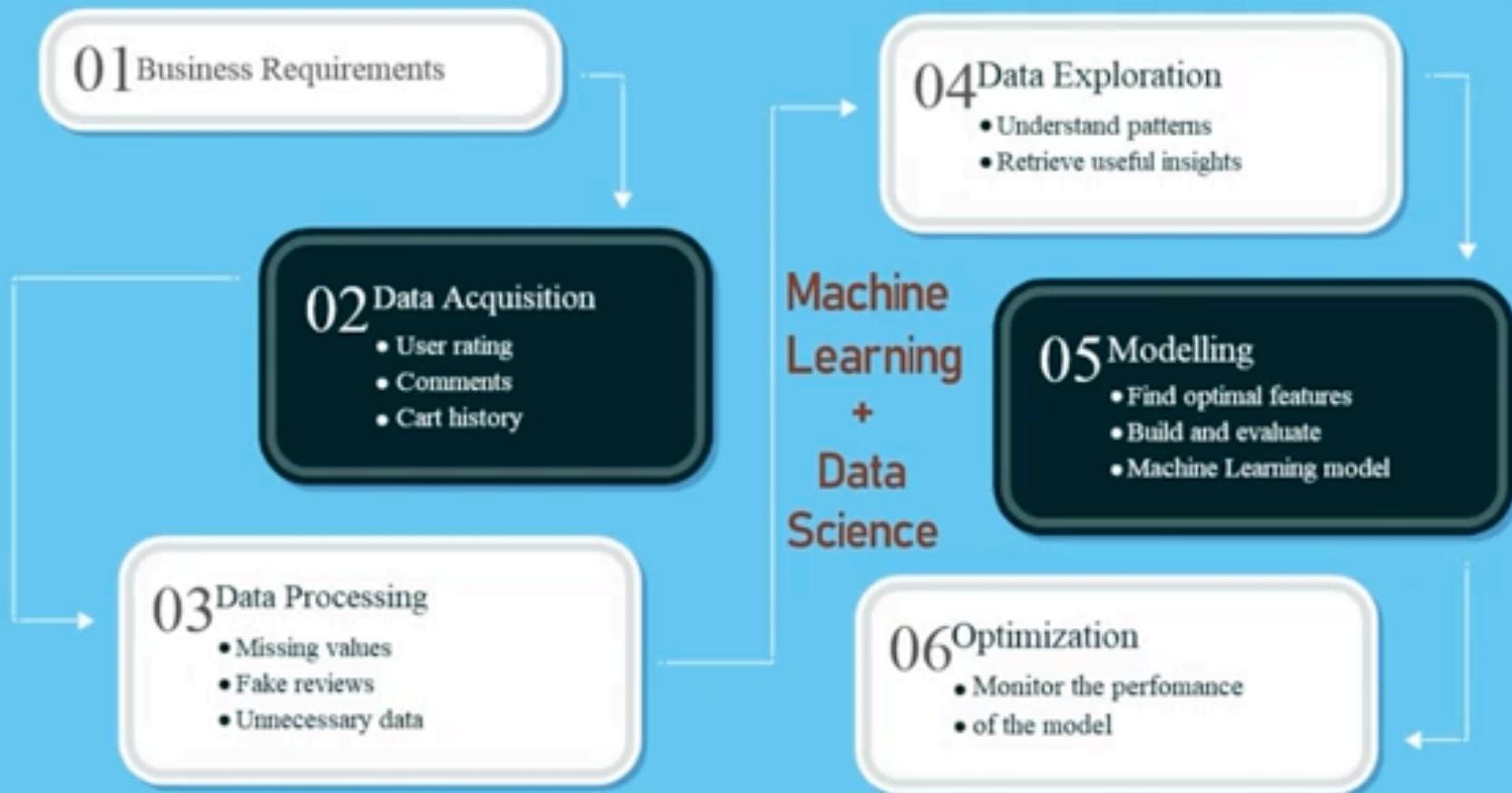
- Forest Model
- Booted Model

## (Unsupervised Learning)

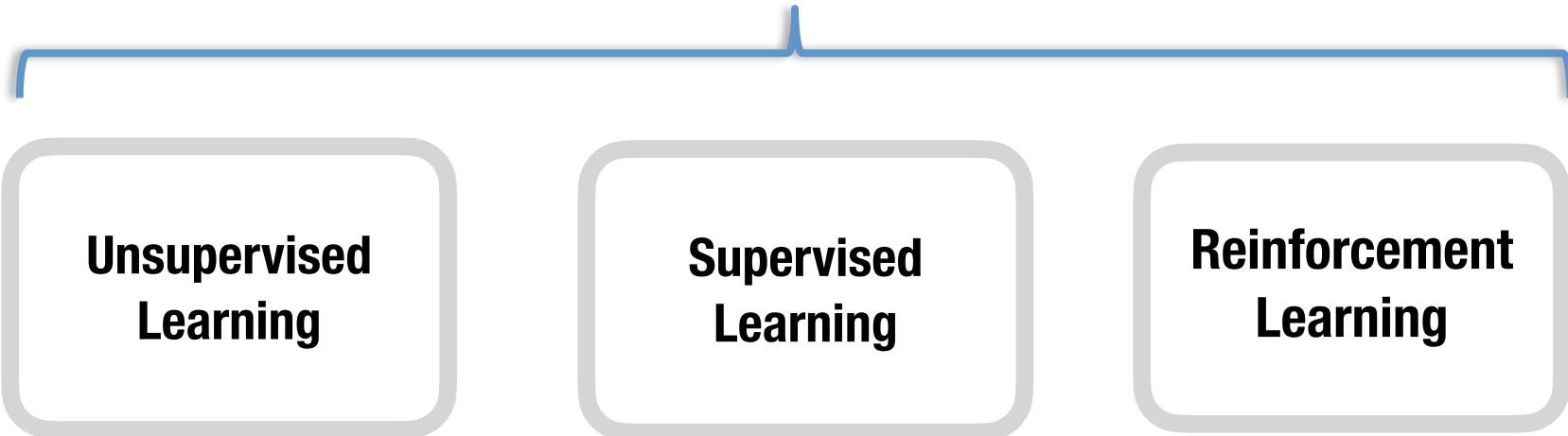
Inferential Statistics & Hypothesis Testing  
推论统计与假设检验

Explorative Data Analysis & Descriptive Statistics  
探索性数据分析与描述性统计

Source: Udacity Model Selection Methodology Map



# **Machine Learning**



```
graph TD; A[Machine Learning] --> B[Unsupervised Learning]; A --> C[Supervised Learning]; A --> D[Reinforcement Learning]
```

**Unsupervised  
Learning**

**Supervised  
Learning**

**Reinforcement  
Learning**

**“Machine Learning is a sub-field of Artificial Intelligence used in programming the computers to learn on its own from data fed to it. The data can be labelled, unlabelled and environmentally triggered through reinforced interactions.”**

## **Supervised Learning**

- Used for prediction of categorical and numerical outcome.
- Data has to be labelled and separated into training set and testing set before model building.
- Apply different algorithms and evaluate which one has best fit.

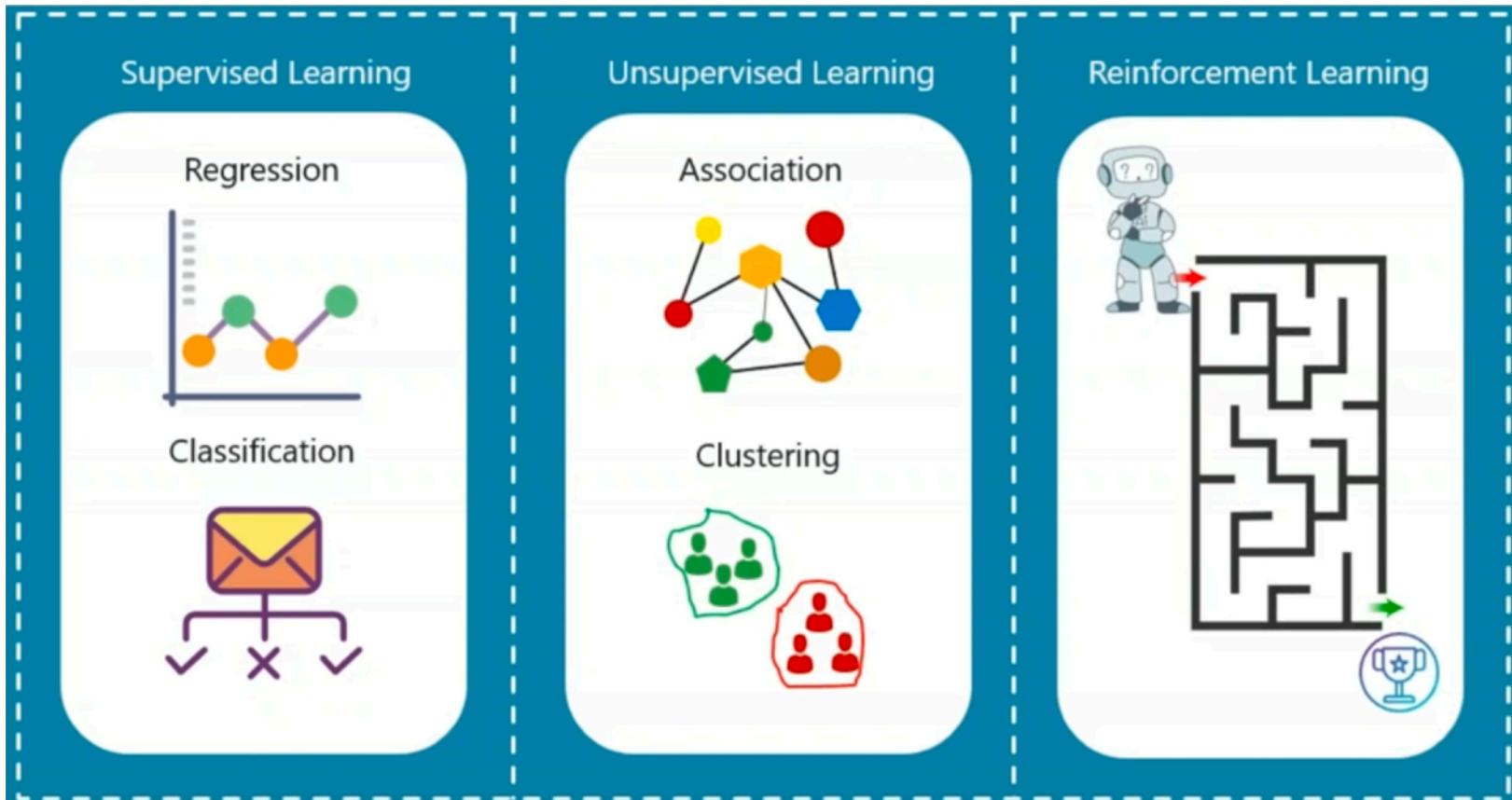
## **Unsupervised Learning**

- Used in the exploratory stage of data preparation to find out patterns (clusters).
- Data is not labelled.
- Used in performing dimension reduction to help extract the essential features for preparing datasets used in model building.

## **Reinforcement Learning**

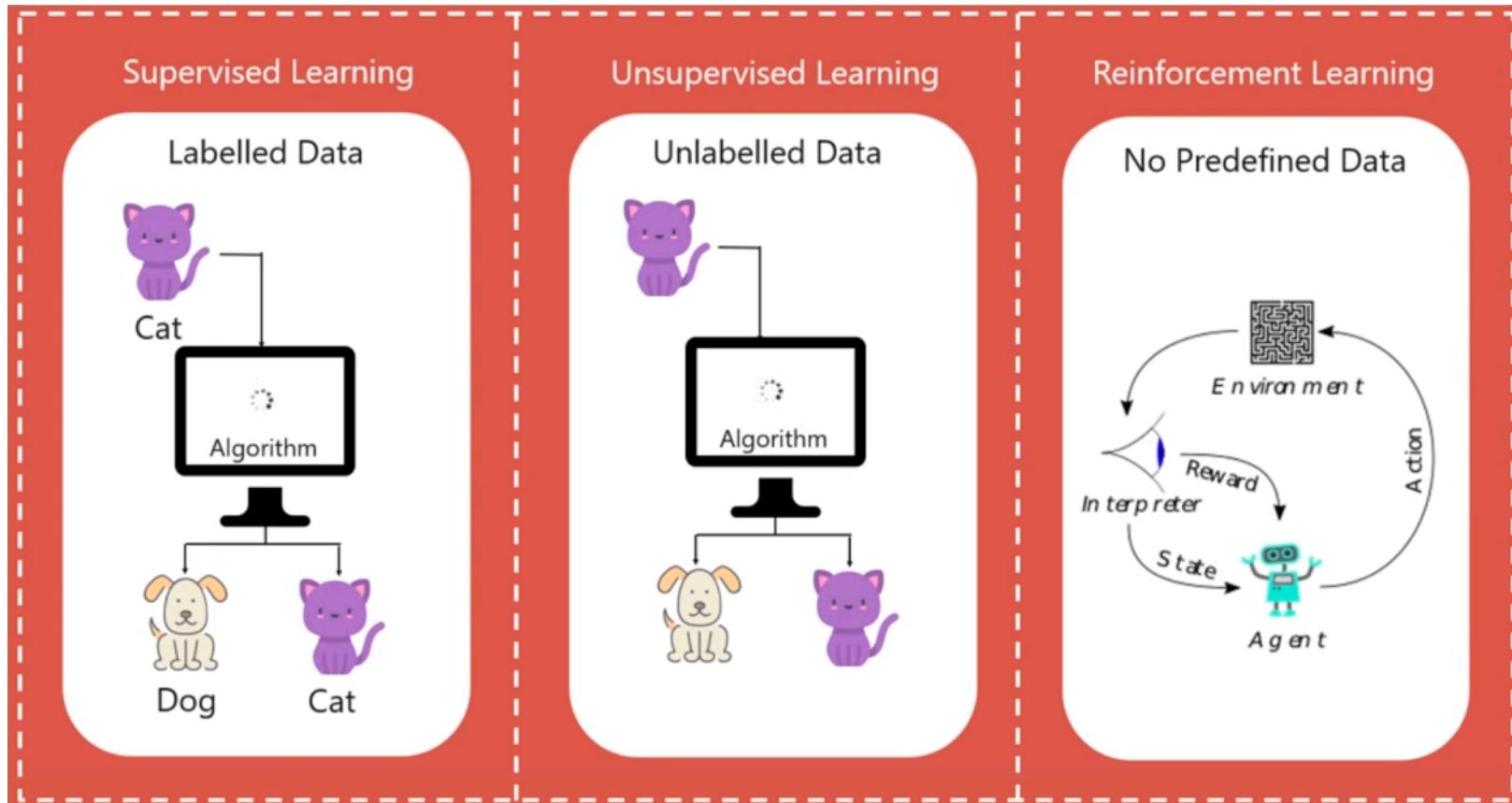
- The ML program is turned into a software agent, navigating through a problem space to reach a goal by trial and error.
- Throughout the course of interaction with the environment, feedbacks will be given to steer the agent toward the goal.

## Type of Problems



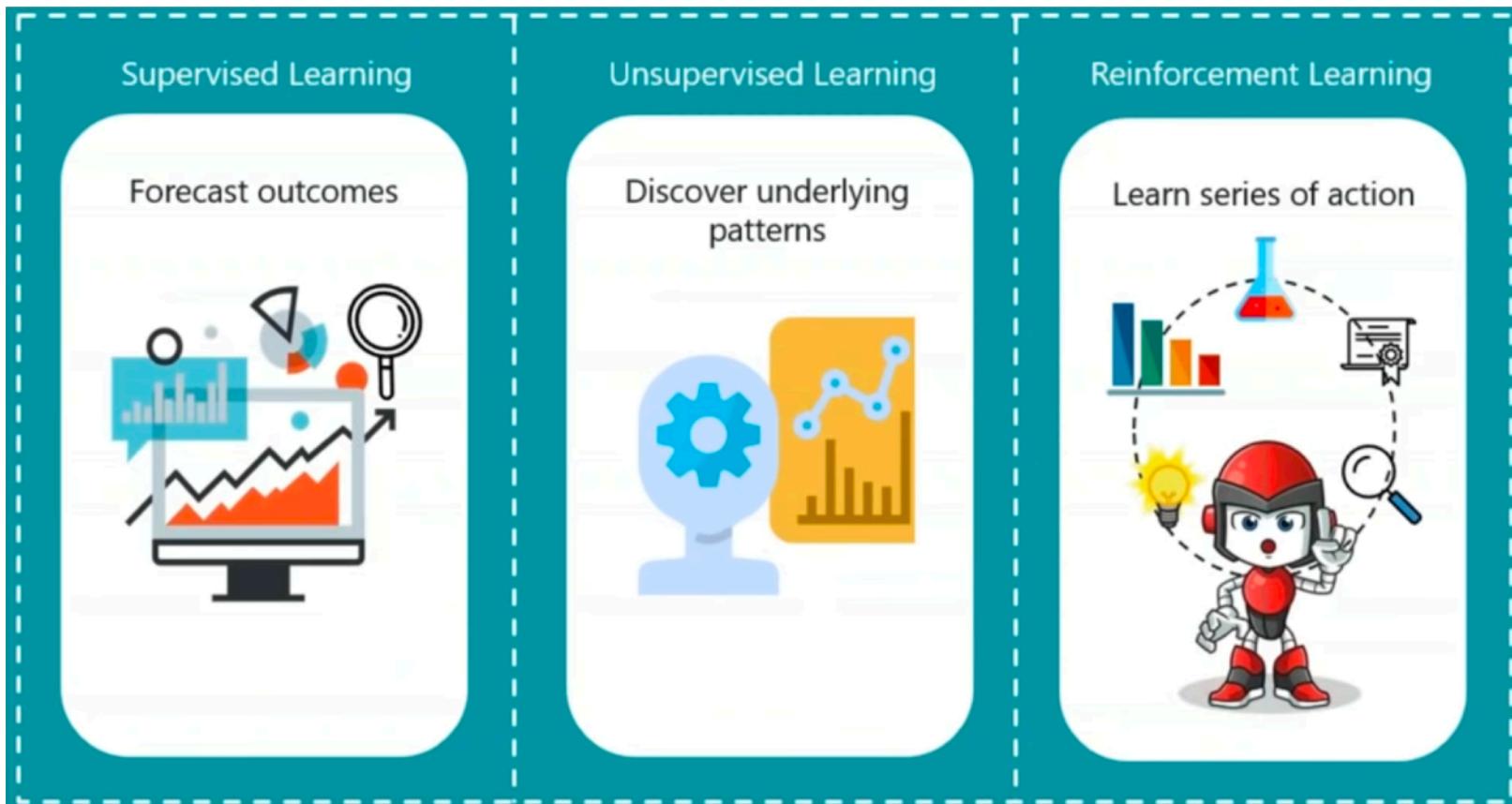
Source: Edureka!

## Type of Data



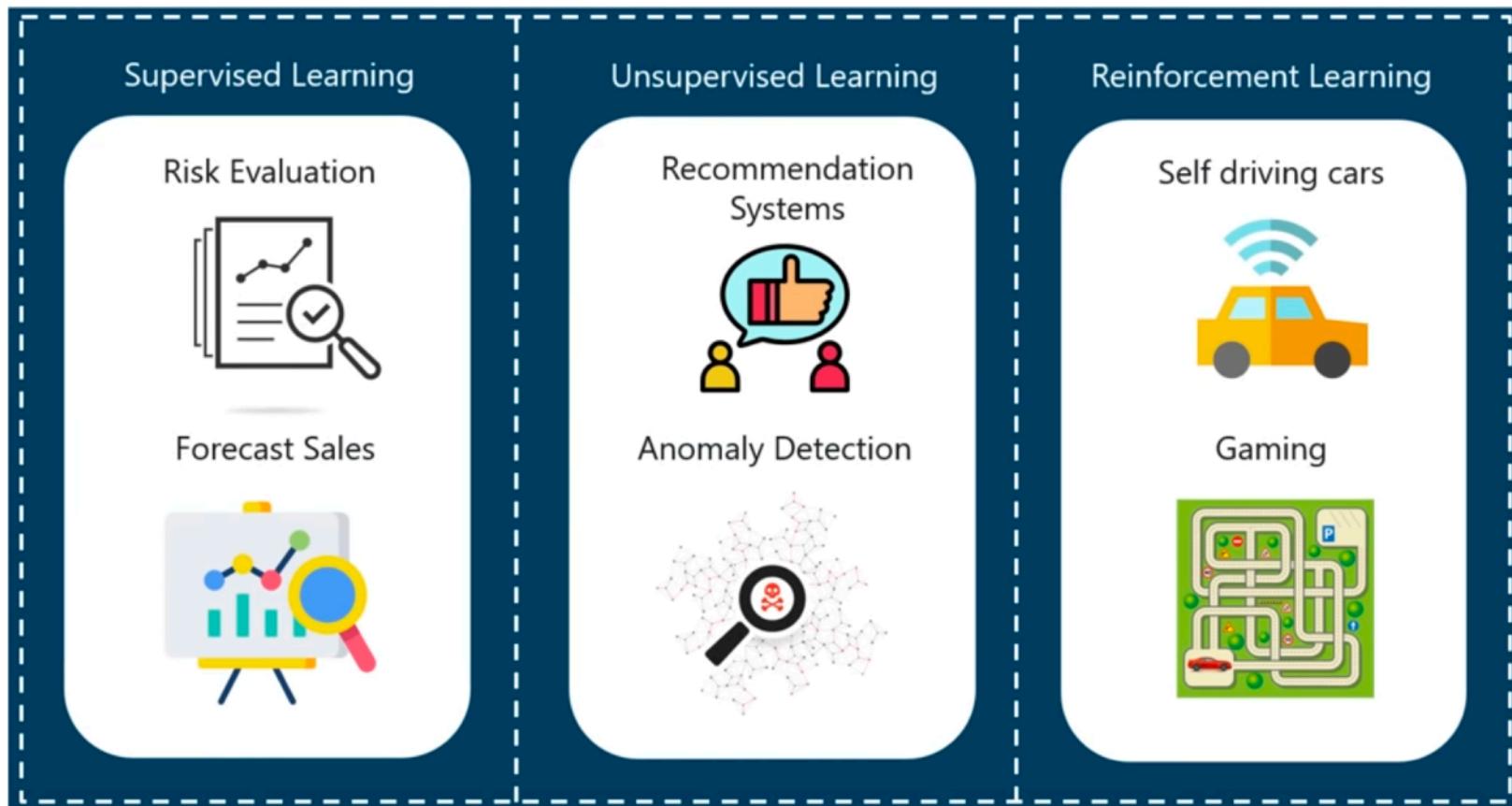
Source: Edureka!

## Aim



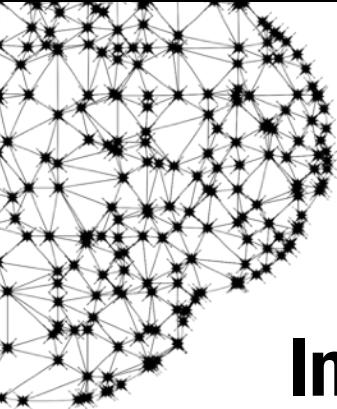
Source: Edureka!

# Applications



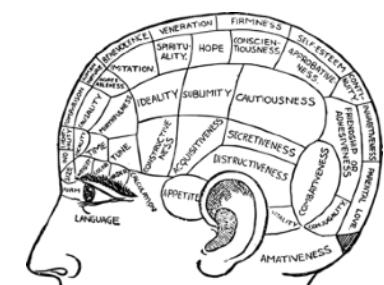
Source: Edureka!

**How does the machine perform its “magic”?**



**In traditional programming, **human** is the  
main programmer?**

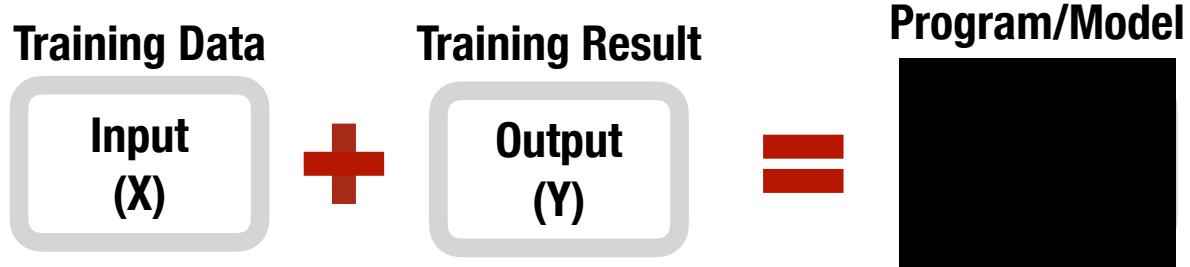
**In machine learning, the **computer** is the  
main programmer?**



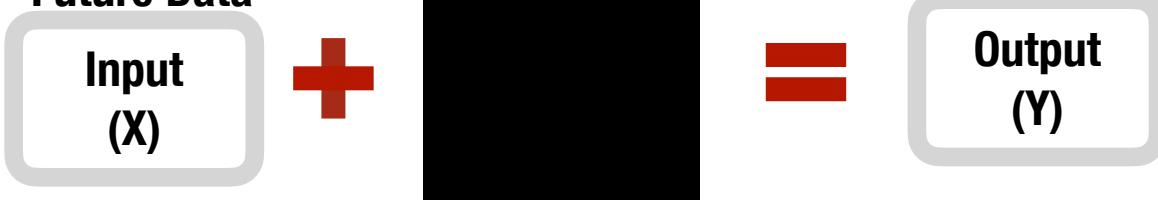
**Traditional  
Programming**



**Machine  
Learning**



**Testing/  
Future Data**



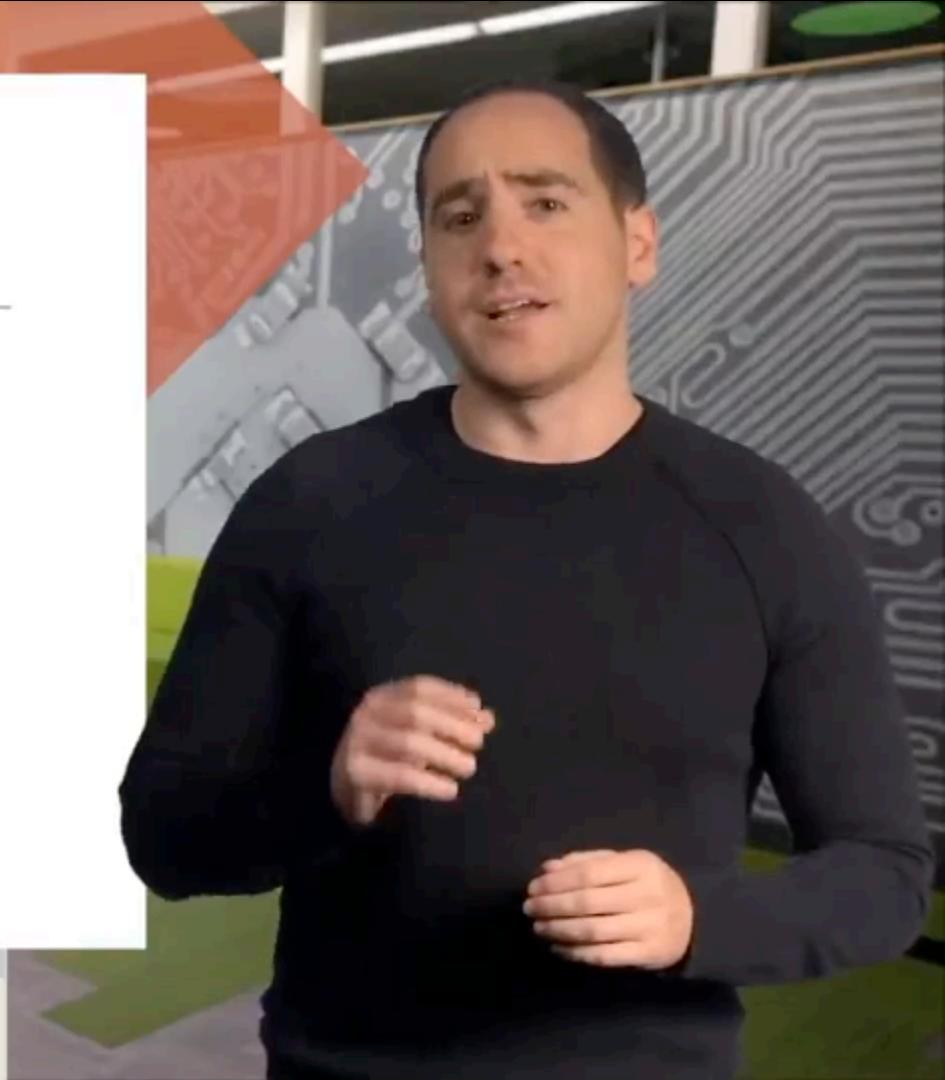
# {ML} Recipes

---

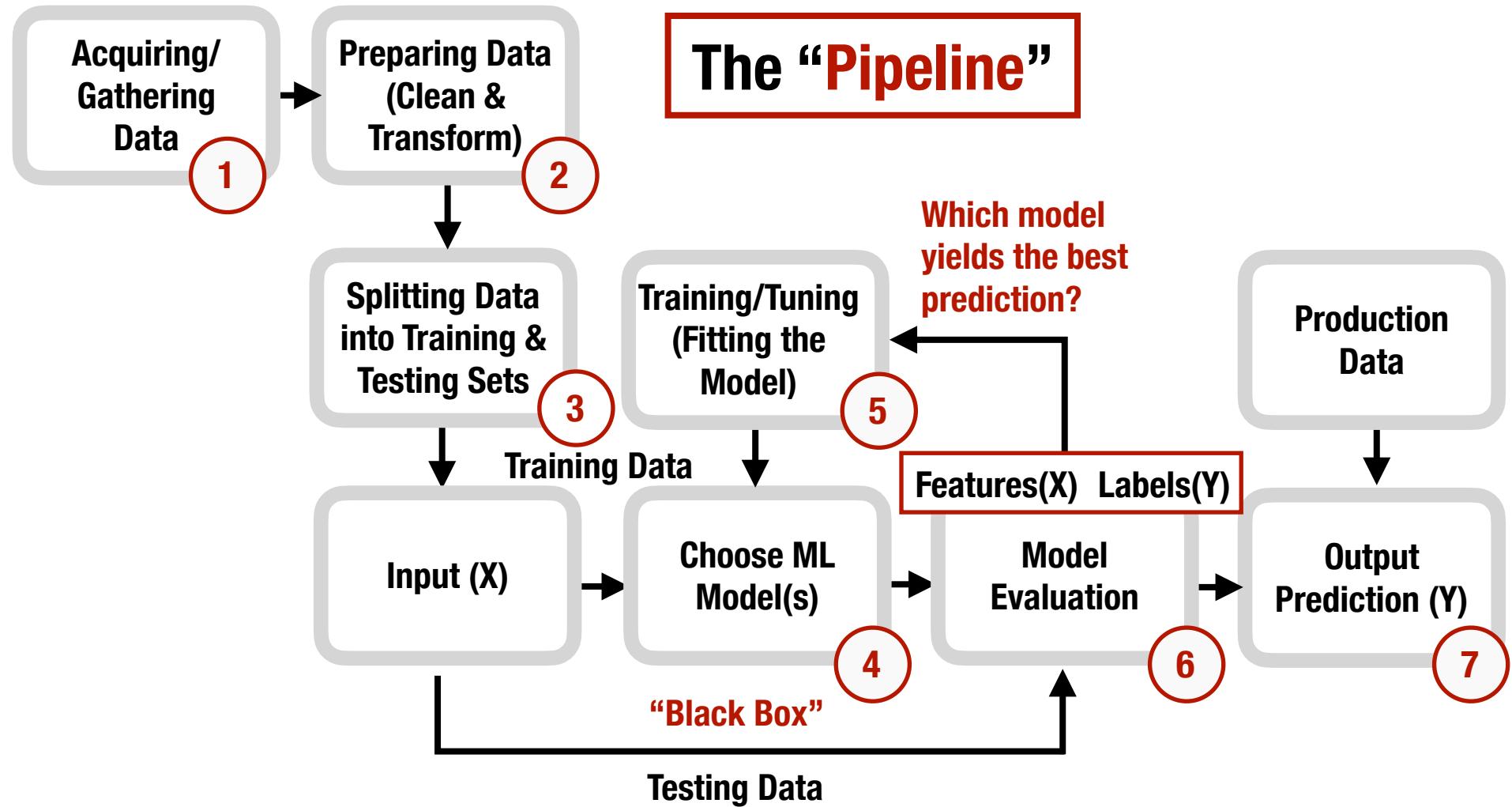
Josh Gordon presents:

Let's Write a Pipeline

 Google Developers



## The “Pipeline”



**“A machine learning pipeline** is used to help **automate machine learning workflows**. They operate by enabling a sequence of data to be transformed and correlated together in a model that can be tested and evaluated to achieve an outcome, whether positive or negative.”

Source: What is a Pipeline in Machine Learning? How to create one? by Shashanka M



ANALYTICS

# What Data Scientists Really Do, According to 35 Data Scientists

by Hugo Bowne-Anderson

August 15, 2018

[Summary](#) [Save](#) [Share](#) [Comment 8](#) [Print](#) **\$8.95** Buy Copies



[https://hbr.org/2018/08/what-data-scientists-really-do-according-to-35-data-scientists?referral=03758&cm\\_vc=rr\\_item\\_page.top\\_right](https://hbr.org/2018/08/what-data-scientists-really-do-according-to-35-data-scientists?referral=03758&cm_vc=rr_item_page.top_right)

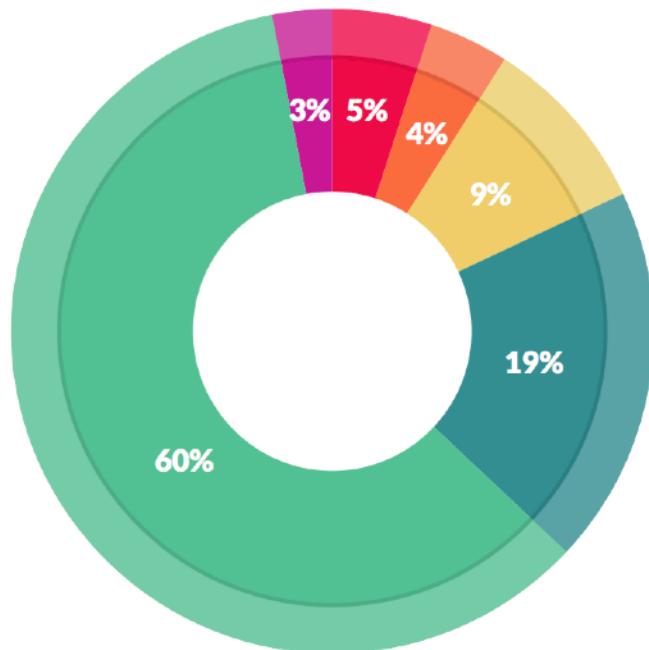


**What data scientists do.** We now know how data science works, at least in the tech industry. First, data scientists lay a solid data foundation in order to perform robust analytics. Then they use online experiments, among other methods, to achieve sustainable growth. Finally, they build machine learning pipelines and personalized data products to better understand their business and customers and to make better decisions. In other words, in tech, data science is about infrastructure, testing, machine learning for decision making, and data products.

Source: By Hugo Bowen-Anderson  
August 15, 2018  
Harvard Business Review

## How a Data Scientist Spends Their Day

Here's where the popular view of data scientists diverges pretty significantly from reality. Generally, we think of data scientists building algorithms, exploring data, and doing predictive analysis. That's actually not what they spend most of their time doing, however.



What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets; 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

**Modelling  
and  
Evaluation**

Source: Data Science 2016 Report by CrowdFlower

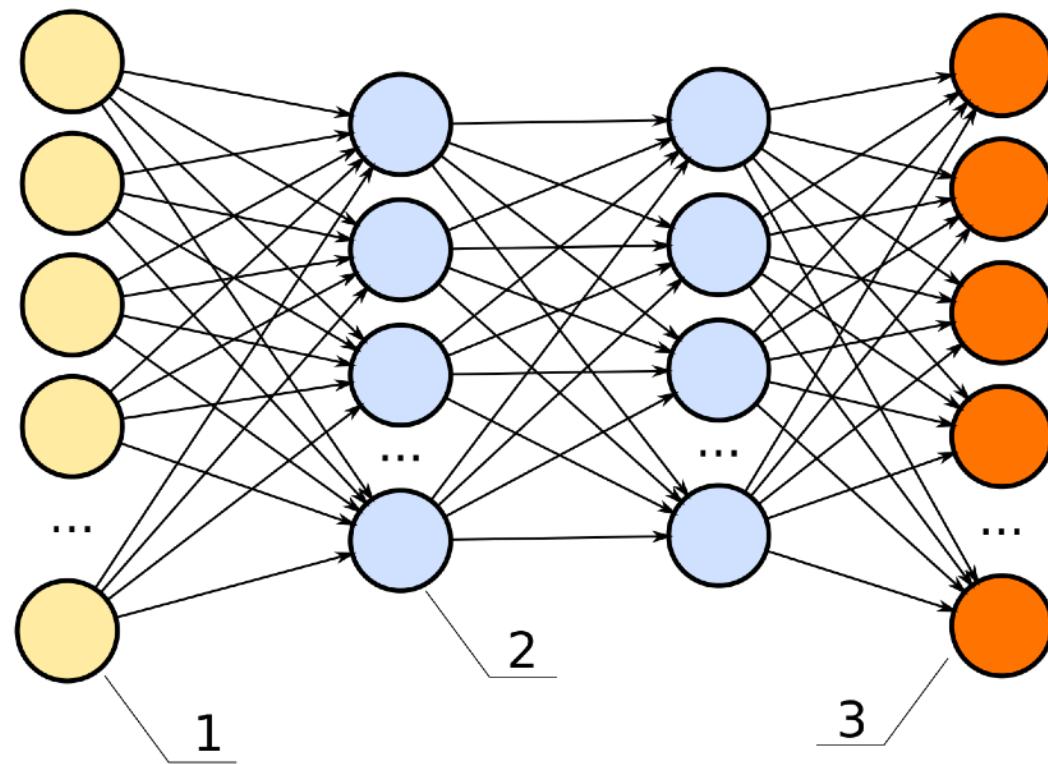
# **Introduction to “Deep Learning”**

**“Deep Learning is a sub-field of Machine Learning which utilises the artificial neural network model to learn on its own from data. The reason that it is called “deep” is because the network is formed with many layers of connection. Output from one layer will become input to another layer, leading to the solution of the problem.”**

Input Layer

Hidden Layers

Output Layer

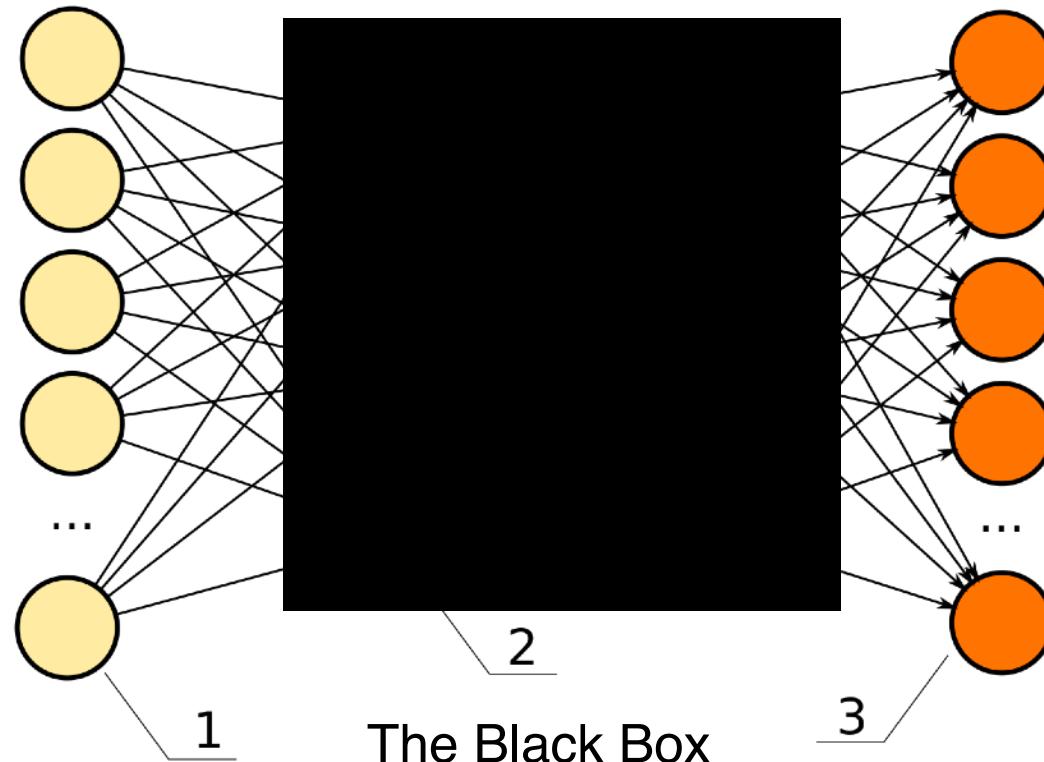


Source: Wikipedia

Input Layer

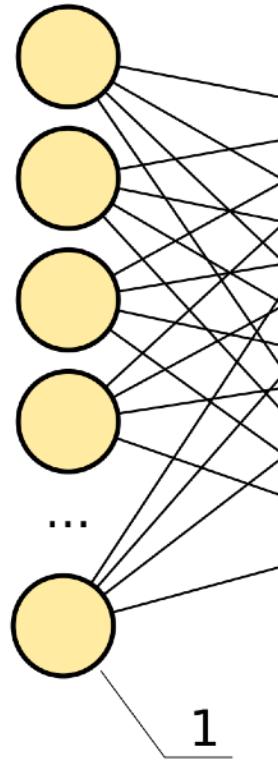
Hidden Layers

Output Layer



Source: Wikipedia

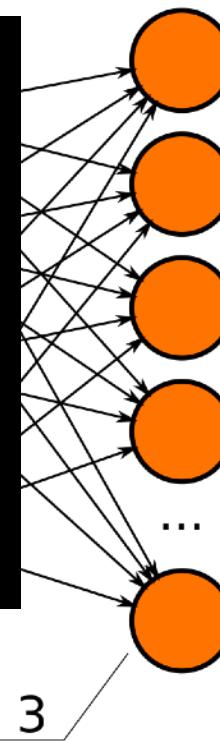
Input Layer



Hidden Layers

- How many hidden layers?
- How many nodes in each layer?
- Activation function
- The effect of learning rate and momentum (how fast and accurate) on the weight and bias.
- Iteration and desired error level (optimisation)

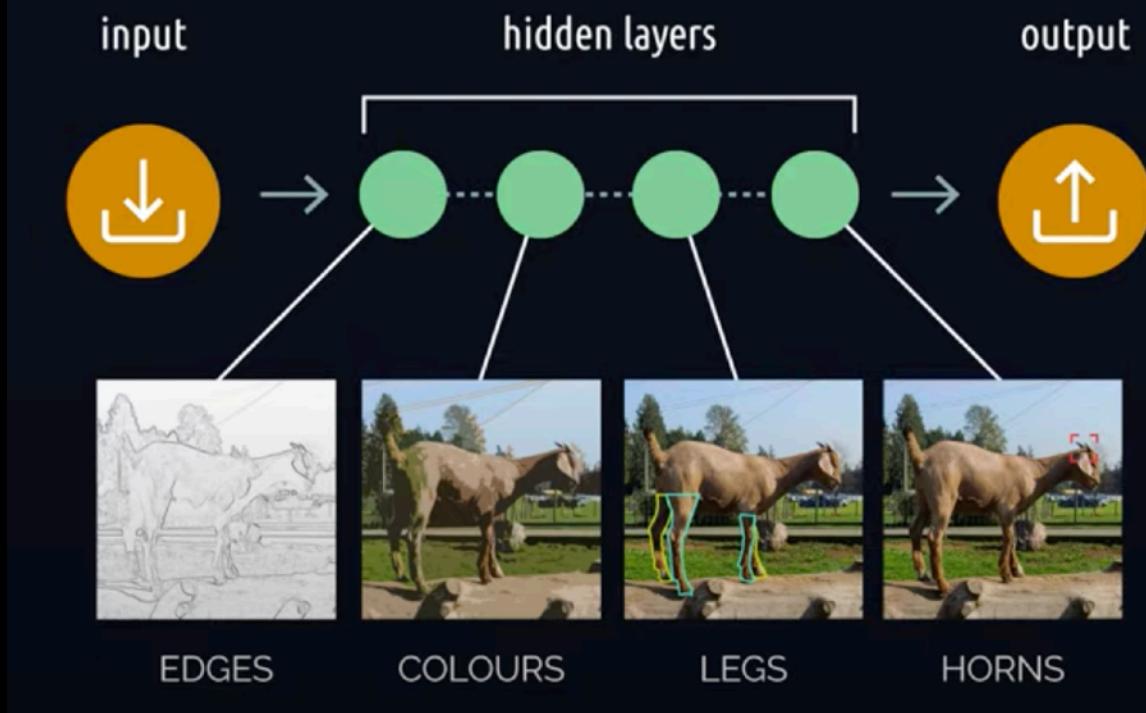
Output Layer



The Black Box

Source: Wikipedia

# *Hidden Layers*



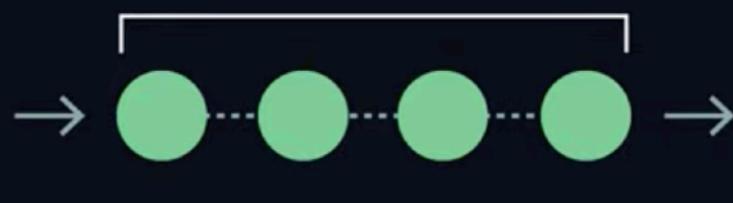
Source: [Python Simplified](#)

# *Training*

input



hidden layers



output



goat



goat

prediction

Source: [Python Simplified](#)

# *Weights*

connect between the nodes of our layers  
and help us determine how much impact  
each node has on the input.



Source: [Python Simplified](#)

# *Weights*

Optimization

we keep **adjusting** the weights until  
most of our examples are **correctly**  
**classified** / predicted.



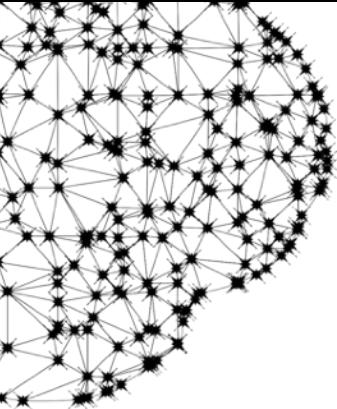
Source: [Python Simplified](#)

# **The “Black Box” *vs.* Explainable & Interpretable AI**



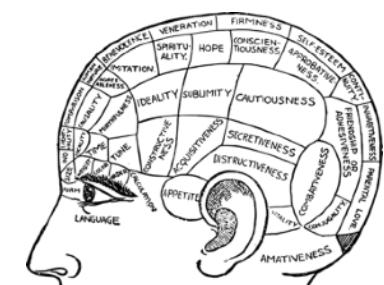
# Comparison Between PA, ML & DL

	Predictive Analytics	Machine Learning	Deep Learning
Supervised Learning	X	X	X
Unsupervised Learning		X	X
Reinforcement Learning		X	X
Data Intensive			X
GPU Processing Intensive			X



# Deep Learning for Everyone

## (Neural Network Model)



[About](#)   [FAQ](#)   [Get Started](#)

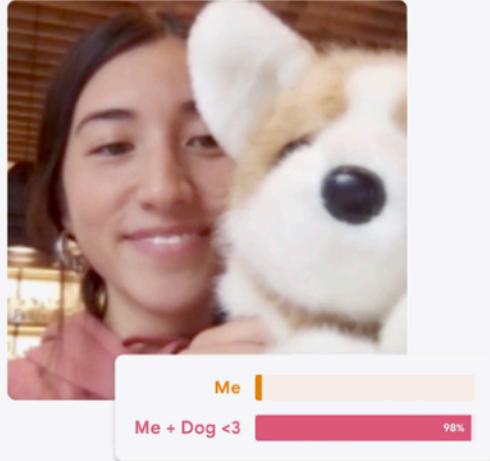
# Teachable Machine

**Train a computer to recognize your own images, sounds, & poses.**

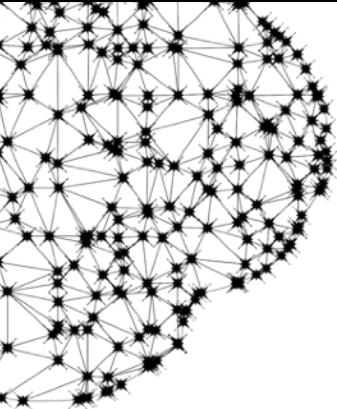
A fast, easy way to create machine learning models for your sites, apps, and more – no expertise or coding required.

[Get Started](#)

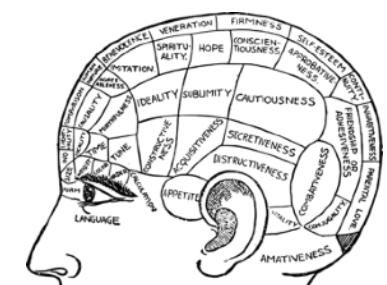
TensorFlow.js   p5.js   Coral   Node.js   GPU



Source: <https://teachablemachine.withgoogle.com/train>



# How does the machine perform its “**magic**”?





### Output

Too  
Early



Ripe



Too Late



No  
Banana



# **Introduction to**

# **ML5.js**

Tracks &gt; A Beginner's Guide to Machine Learning in JavaScript

# A BEGINNER'S GUIDE TO MACHINE LEARNING IN JAVASCRIPT

Welcome to "A Beginner's Guide to Machine Learning in JavaScript"! In this series, I'll teach the concepts behind machine learning using the ml5.js library.

**Main track**

10 chapters, 23 videos

**Languages**p5.js, ml5.js, [Show more](#)**Topics**machine learning (ML), ml5.js, [Show more](#)

## A Beginner's Guide to Machine Learning with ml5.js

**Languages** p5.js, ml5.js, [Show more](#)**Topics** machine learning (ML), ml5.js, [Show more](#)

## TRACK STOPS

## TIMESTAMPS

## INTRODUCTION

A Beginner's Guide to Machine Learning with ml5.js

## CLASSIFICATION

○ ml5.js: Image Classification with MobileNet

○ ml5.js: Webcam Image Classification

○ Object Detection with COCO-SSD

Source: <https://thecodingtrain.com/tracks/ml5js-beginners-guide>



# Friendly Machine Learning for the Web

A neighborly approach to creating and exploring artificial  
intelligence in the browser.

Source: <https://learn.ml5js.org/#/>



Type to search

ml5

## Welcome

[Getting Started](#)[FAQ](#)

## Tutorials

[Introduction to ml5.js](#)[Running a local web server](#)[Promises and Callbacks in ml5](#)

## Reference

[Overview](#)

## Helpers

[NeuralNetwork](#)[FeatureExtractor](#)[KNNClassifier](#)[kmeans](#)

## Image

[ImageClassifier](#)[- Description](#)[- Quickstart](#)

# ImageClassifier



Label: robin, American robin, *Turdus migratorius*  
Confidence: 0.99

## Description

You can use neural networks to recognize the content of images. `ml5.imageClassifier()` is a method to create an object that classifies an image using a pre-trained model.

It should be noted that the pre-trained model provided by the example below was trained on a database of approximately 15 million images (ImageNet). The ml5 library accesses this model from the cloud. What the algorithm labels an image is entirely dependent on that training data -- what is included, excluded, and how those images are labeled (or mislabeled).

**Train your own image classification model with Teachable Machine:** If you'd like to train your own custom image classification model, try [Google's Teachable Machine](#).

## Quickstart

Source: <https://learn.ml5js.org/#/reference/image-classifier>



14,197,122 images, 21841 synsets indexed

[Home](#) [Download](#) [Challenges](#) [About](#)

Not logged in. [Login](#) | [Signup](#)

**ImageNet** is an image database organized according to the **WordNet** hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. The project has been **instrumental** in advancing computer vision and deep learning research. The data is available for free to researchers for non-commercial use.

**Mar 11 2021. ImageNet website update.**

© 2020 Stanford Vision Lab, Stanford University, Princeton University [imagenet.help.desk@gmail.com](mailto:imagenet.help.desk@gmail.com) Copyright infringement

Source: <https://www.image-net.org/index.php>



## DEMO

- Tutorials
- Model and Data Provenance
- Acknowledgements
- Source Code

## Sound

[SoundClassification](#)[PitchDetection](#)

## Text

[CharRNN](#)[Sentiment](#)[Word2Vec](#)

## Utils

[utils](#)

## Contributing

[Contributor Notes](#)[Maintenance Notes](#)[Style Guide](#)[Reference guidelines](#)[Development guidelines](#)

## Design guidelines



## Description

Real-time object detection system using either [YOLO](#) or [CocoSsd](#) model.

## Quickstart

```
const video = document.getElementById('video');

// Create a ObjectDetector method
const objectDetector = ml5.objectDetector('cocossd', {}, modelLoaded);

// When the model is loaded
function modelLoaded() {
```

js

Source: <https://learn.ml5js.org/#/reference/object-detector>

## News

- We are pleased to announce the [LVIS 2021 Challenge and Workshop](#) to be held at ICCV.
- Please note that there will not be a COCO 2021 Challenge, instead, we encourage people to participate in the LVIS 2021 Challenge.
- We have partnered with the team behind the open-source tool [FiftyOne](#) to make it easier to download, visualize, and evaluate COCO
- [FiftyOne](#) is an open-source tool facilitating visualization and access to COCO data resources and serves as an evaluation tool for model analysis on COCO.

## What is COCO?



COCO is a large-scale object detection, segmentation, and captioning dataset. COCO has several features:

- ✓ Object segmentation
- ✓ Recognition in context
- ✓ Superpixel stuff segmentation
- ✓ 330K images (>200K labeled)
- ✓ 1.5 million object instances
- ✓ 80 object categories
- ✓ 91 stuff categories
- ✓ 5 captions per image
- ✓ 250,000 people with keypoints

## Collaborators

Tsung-Yi Lin Google Brain

Genevieve Patterson MSR, Trash TV

Matteo R. Ronchi Caltech

Yin Cui Google

Michael Maire TTI-Chicago

Serge Belongie Cornell Tech

Lubomir Bourdev WaveOne, Inc.

Ross Girshick FAIR

James Hays Georgia Tech

Pietro Perona Caltech

Deva Ramanan CMU

Larry Zitnick FAIR

Piotr Dollár FAIR

## Sponsors



CVDF



Microsoft

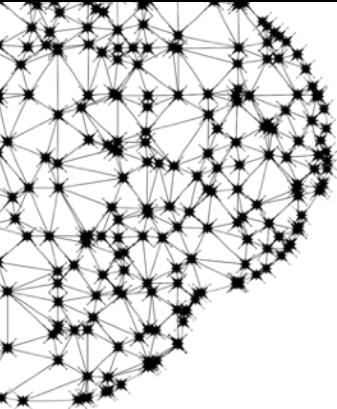


facebook

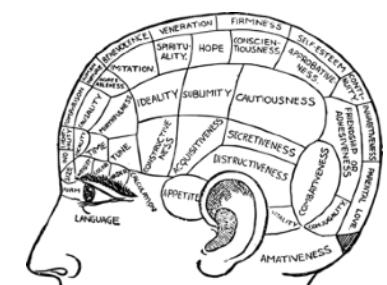


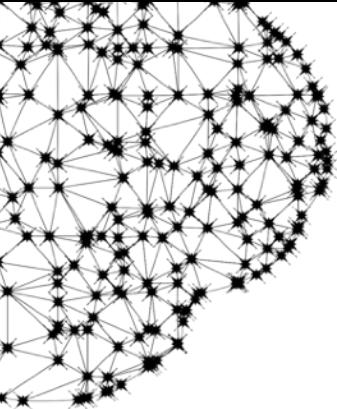
Mighty Ai

Source: <https://cocodataset.org/#home>

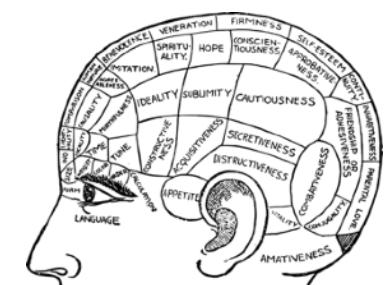


To work with in data science and machine learning, **data** is the **key**.

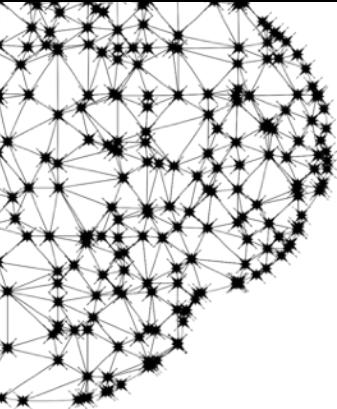




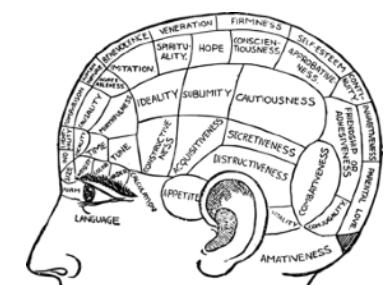
**Yes, data is important. But **what** kind of data?**







**How to **avoid** bias? How should the data  
be **collected**? These are important issues  
for us to think about beyond the hype!**



**Thank you for your time!**