

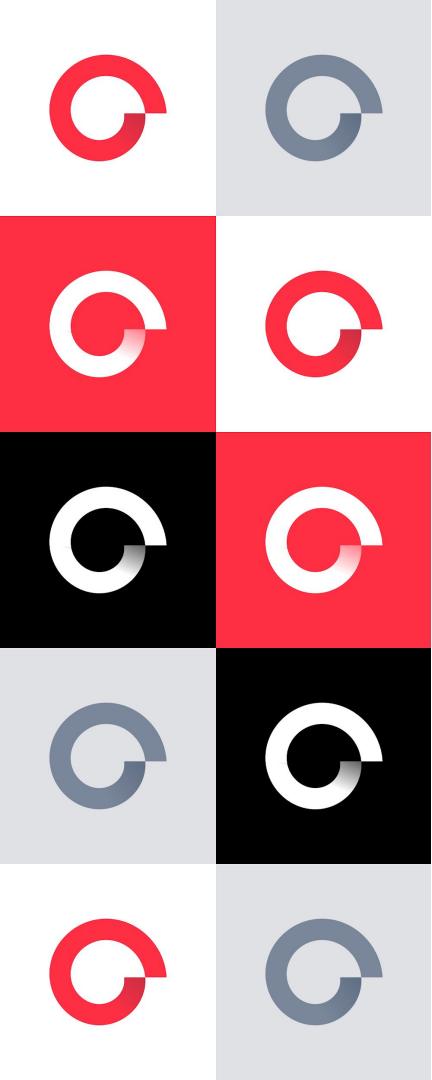
# BTG Academy

Analytics Track

October 16th, 2019

## WTF is Machine Learning?

1. Artificial Intelligence for Dummies
2. Machine Learning Procedure
3. Machine Learning Intuition
4. ~~Jupyter Notebook Hands On~~



# BTG Academy

Analytics Track

October 16th, 2019

## WTF is Machine Learning?

1. Artificial Intelligence for Dummies
2. Machine Learning Procedure
3. Machine Learning Intuition
4. ~~Jupyter Notebook Hands On~~

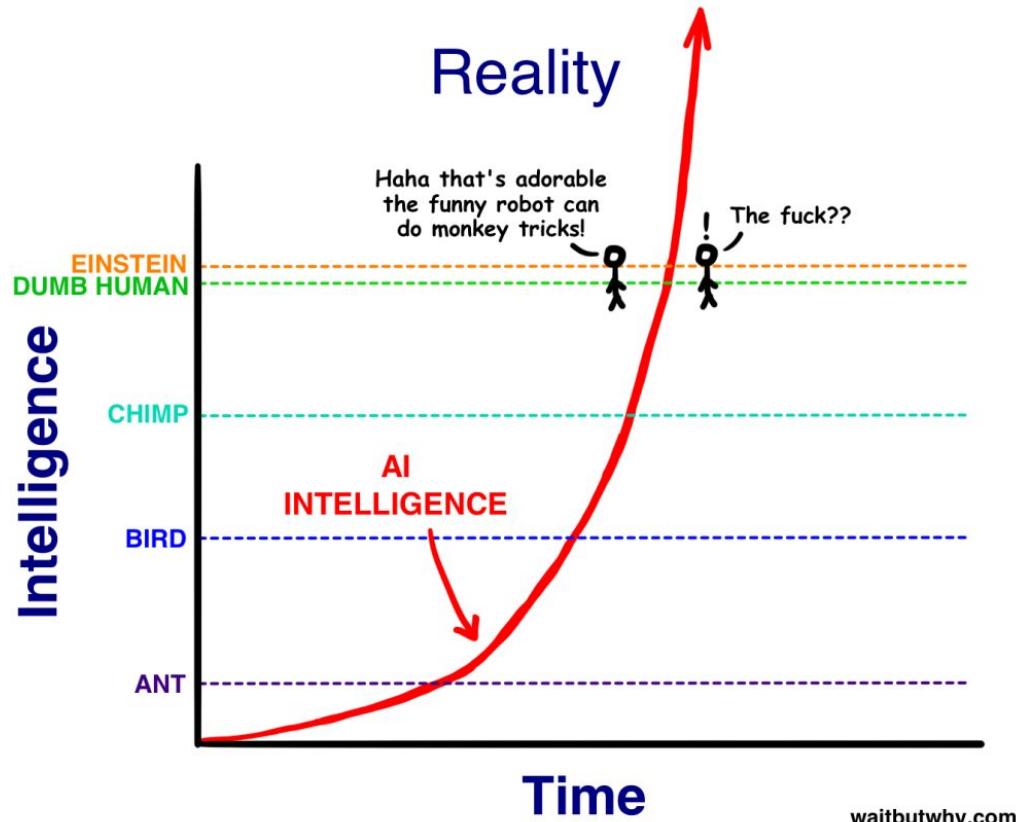


# Artificial Intelligence for Dummies

**AI Caliber 1) Artificial Narrow Intelligence (ANI):** Sometimes referred to as *Weak AI*, Artificial Narrow Intelligence is AI that specializes in *one* area. There's AI that can beat the world chess champion in chess, but that's the only thing it does. Ask it to figure out a better way to store data on a hard drive, and it'll look at you blankly.

**AI Caliber 2) Artificial General Intelligence (AGI):** Sometimes referred to as *Strong AI*, or *Human-Level AI*, Artificial General Intelligence refers to a computer that is as smart as a human *across the board*—a machine that can perform any intellectual task that a human being can. Creating AGI is a *much* harder task than creating ANI, and we're yet to do it. Professor Linda Gottfredson describes intelligence as “a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly, and learn from experience.” AGI would be able to do all of those things as easily as you can.

**AI Caliber 3) Artificial Superintelligence (ASI):** Oxford philosopher and leading AI thinker Nick Bostrom [defines](#) superintelligence as “an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills.” Artificial Superintelligence ranges from a computer that’s just a little smarter than a human to one that’s trillions of times smarter—across the board. ASI is the reason the topic of AI is such a spicy meatball and why the words “immortality” and “extinction” will both appear in these posts multiple times.



## Dictionary

Search for a word



# ar·ti·fi·cial in·tel·li·gence

/ ərdə'fiSHəl in'teləjəns/

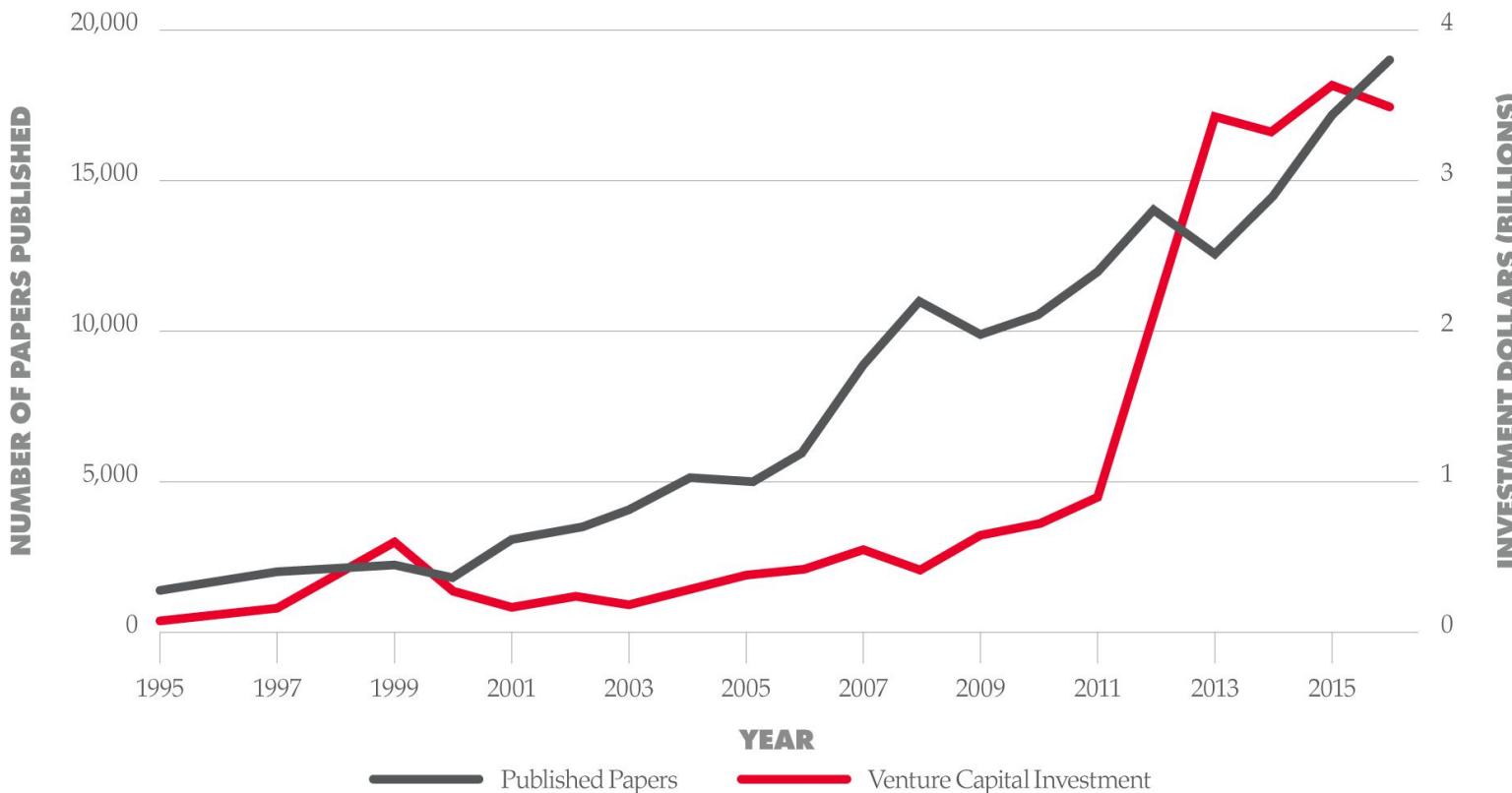
*noun*

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



Translations, word origin, and more definitions

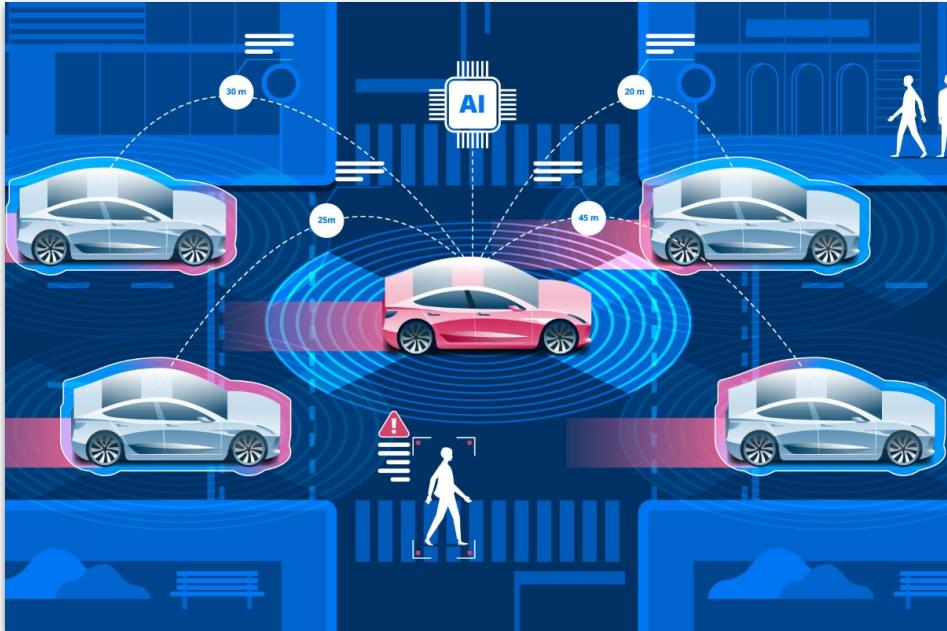
## GROWTH OF AI OVER TIME





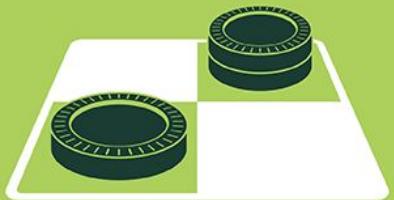






## ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



1950's

1960's

1970's

1980's

## MACHINE LEARNING

Machine learning begins to flourish.



1990's

2000's

2010's

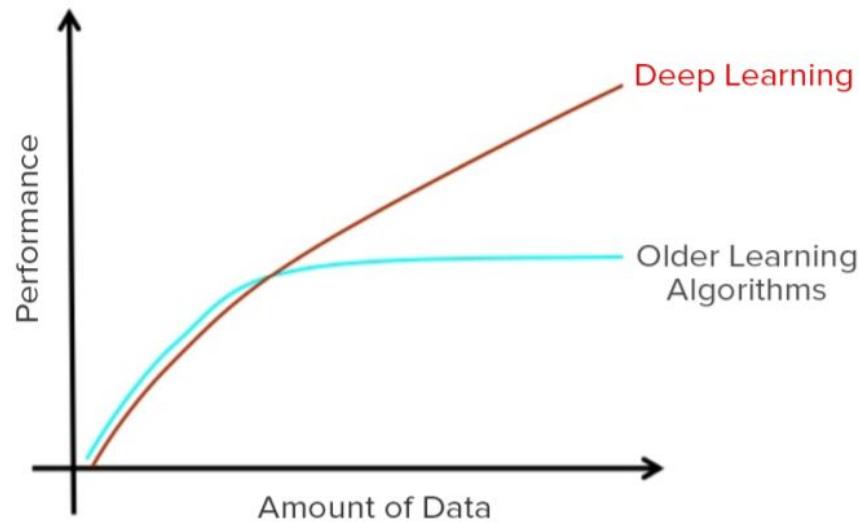
## DEEP LEARNING

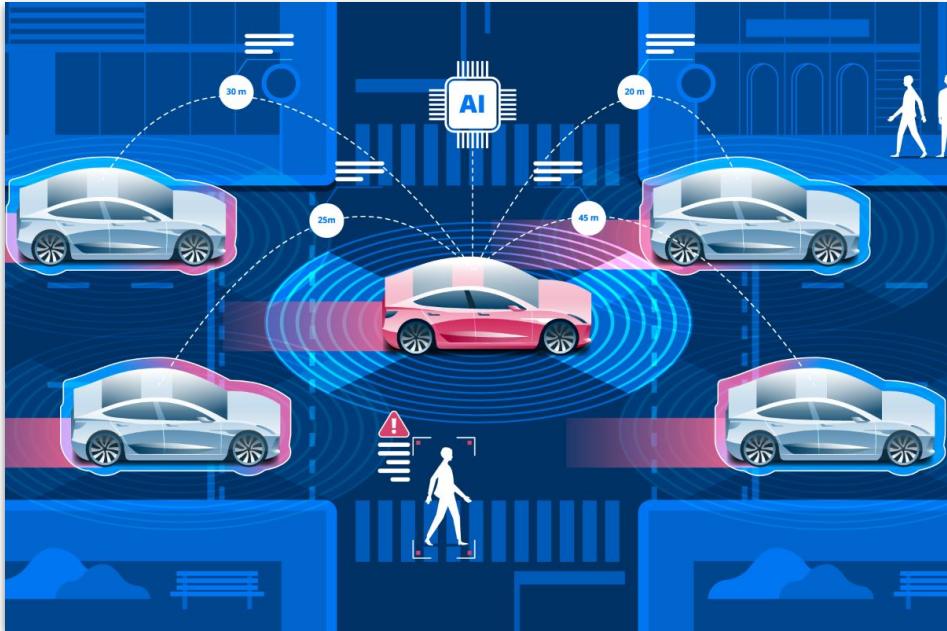
Deep learning breakthroughs drive AI boom.

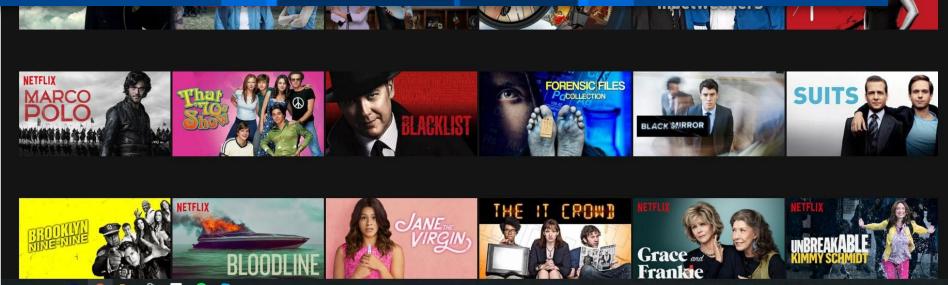
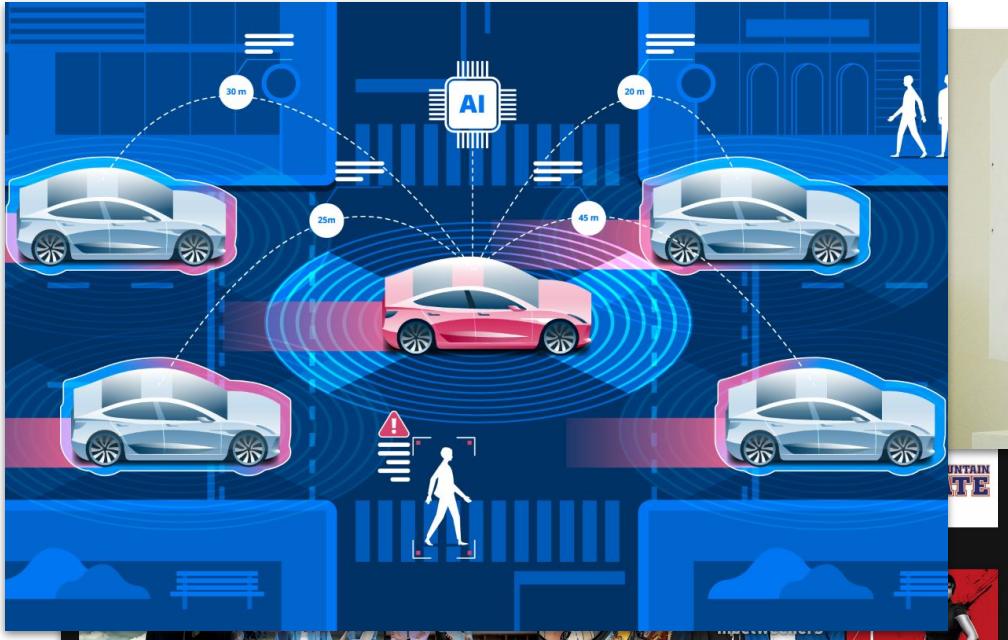


Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

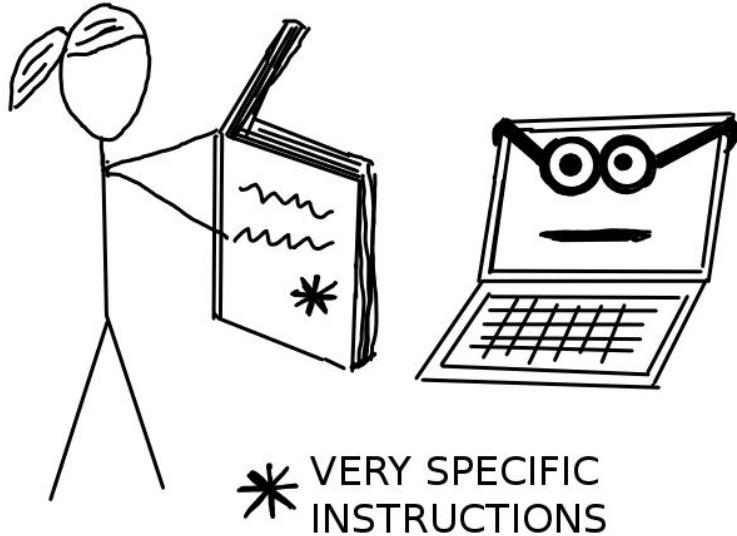
Big Data  
+  
Computing  
Power



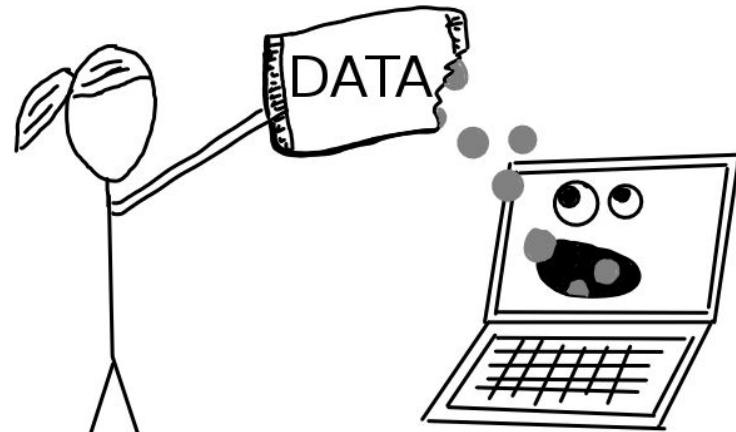




## Without Machine Learning



## With Machine Learning



Supervised  
Learning

Unsupervised  
Learning

# Machine Learning

Reinforcement  
Learning

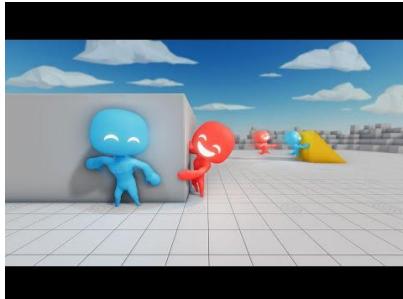
Supervised  
Learning

With Labels

Unsupervised  
Learning

Without Labels

# Machine Learning

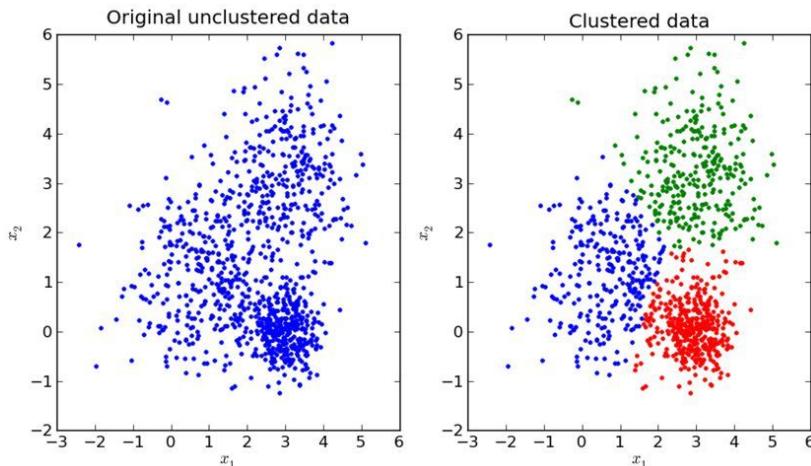


Reinforcement  
Learning



# Unsupervised Learning

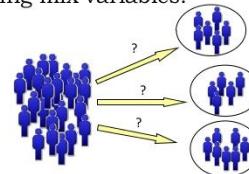
Without Labels



## Application

### Example: Market Segmentation

- Process of dividing a total market into groups of consumers who have similar needs and who respond similarly to marketing mix variables.



## Supervised Learning

With Labels

### Classification



“Dog”



“Cat”

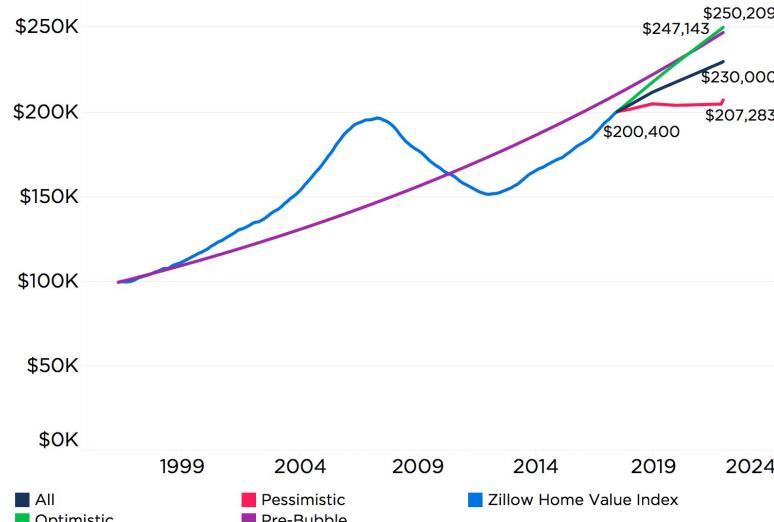


# Supervised Learning

With Labels

## Regression

U.S. Home Prices Expectations for the Next Five Years



Powered by pulseonomics

Source: Zillow Home Price Expectations Survey, 2017 Q3

# Interpolation

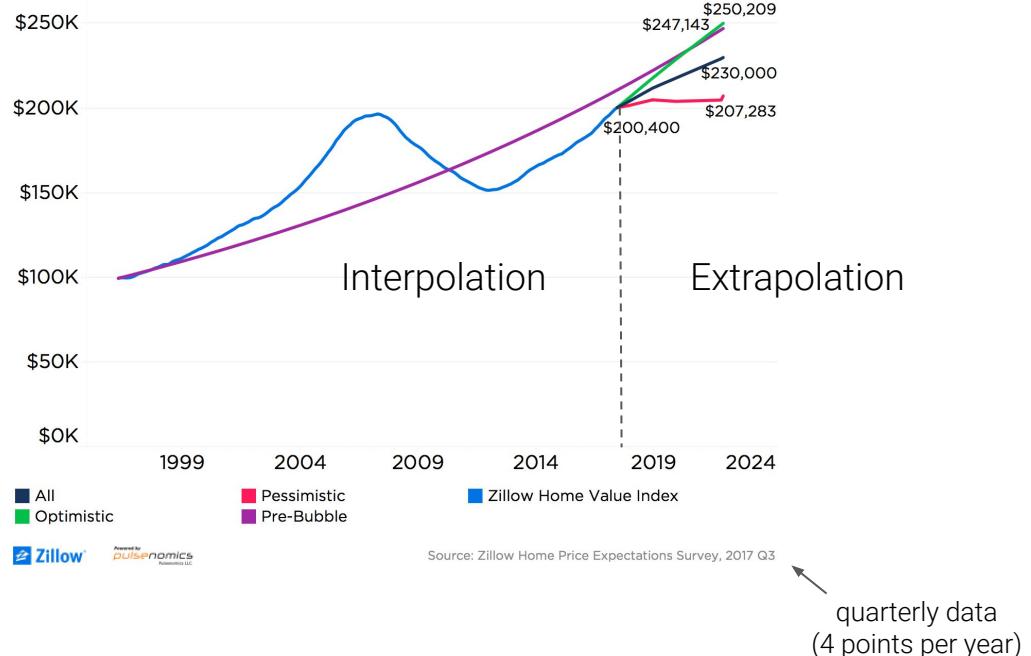
“What was the home price expectation in **December 2010?**”

# Extrapolation

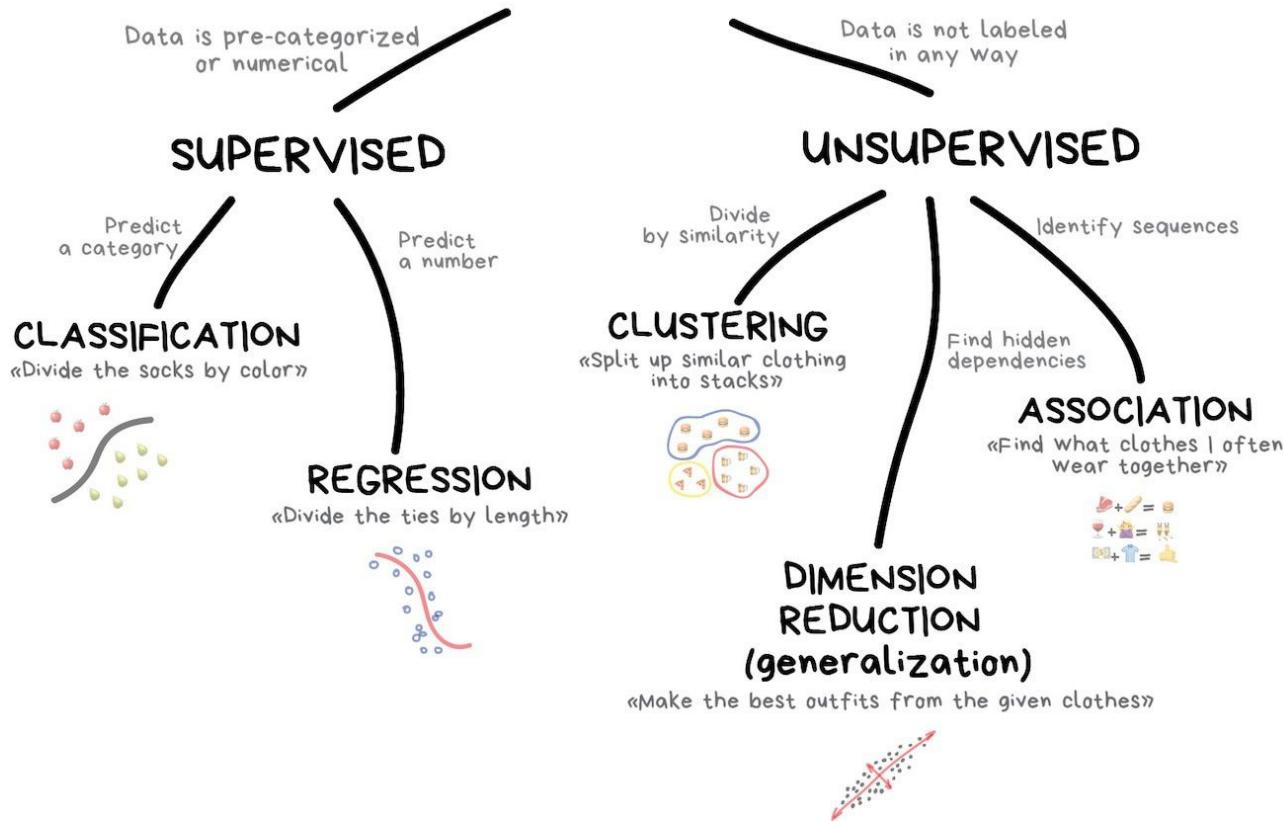
“What is the home price expectation in **2024?**”

# Regression

U.S. Home Prices Expectations for the Next Five Years



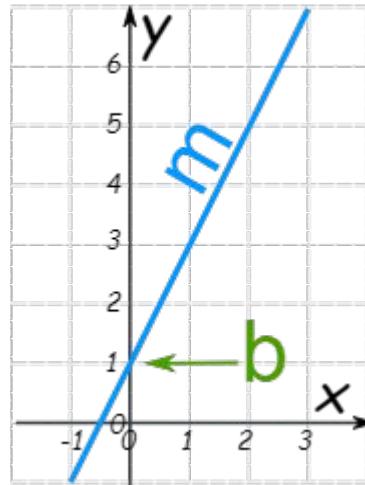
# CLASSICAL MACHINE LEARNING



# Machine Learning Procedure

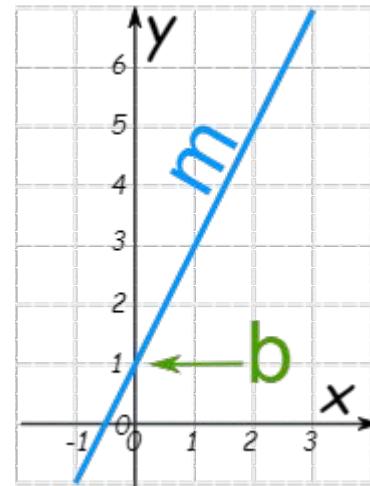
# Line of Best Fit

$$y = mx + b$$



# Machine Learning

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

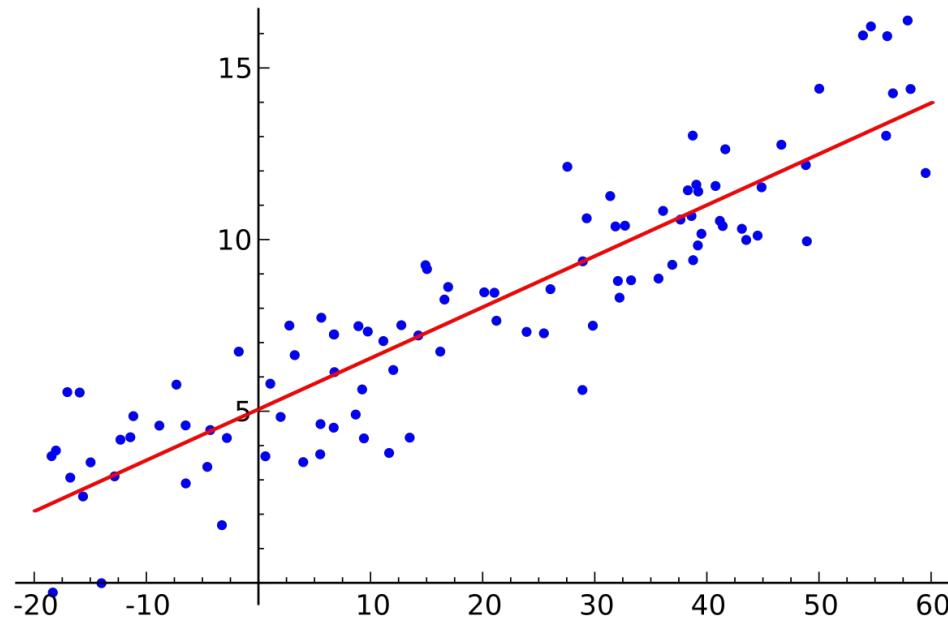


# Loss Function

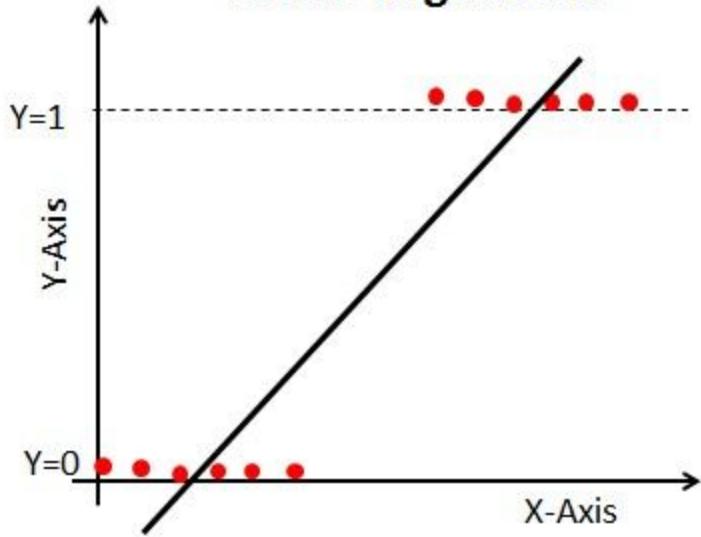
$$Q_{\text{LS}}(\beta_0, \beta_1) = \sum_{i=1}^n (Y_i - (\beta_0 + \beta_1 X_i))^2$$

↑  
Least Squares Error

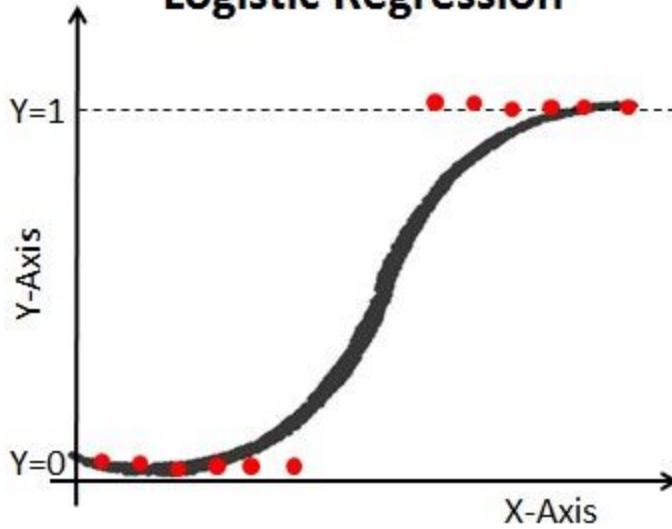
# Linear Regression



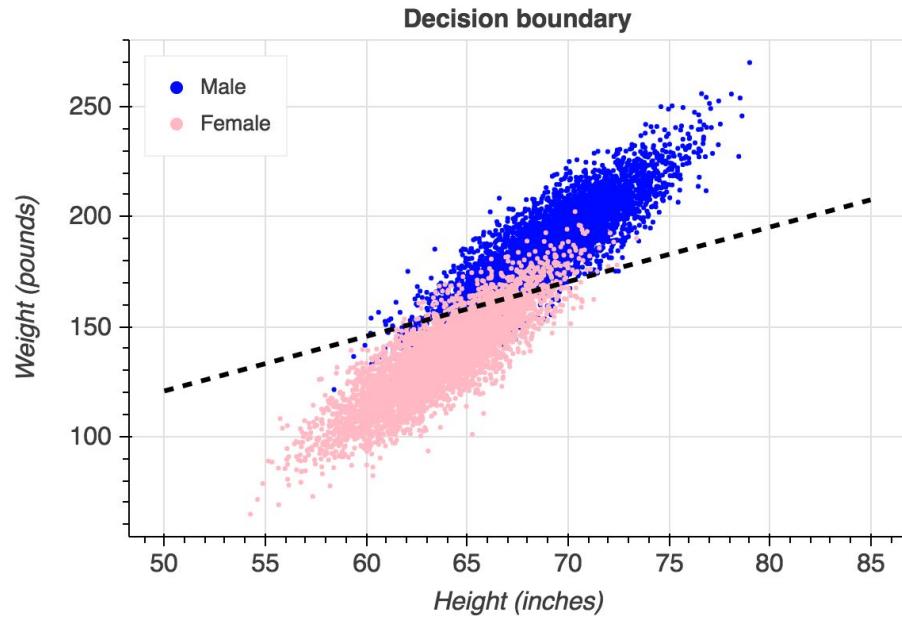
### Linear Regression



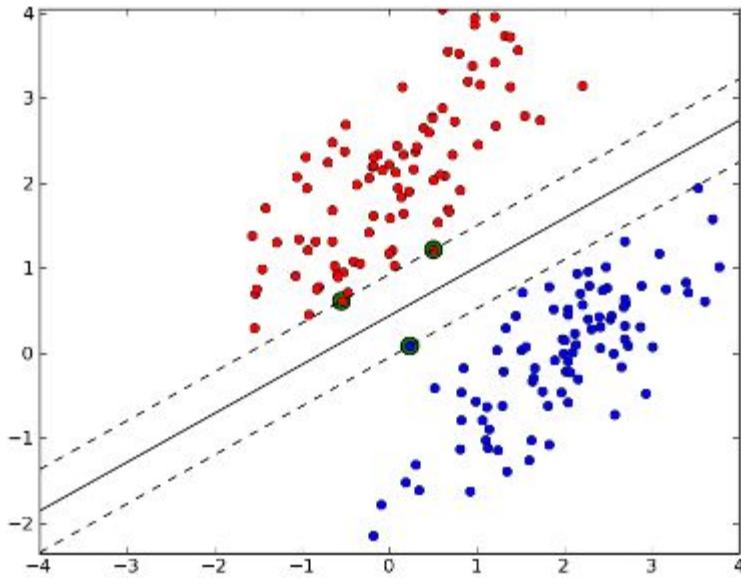
### Logistic Regression

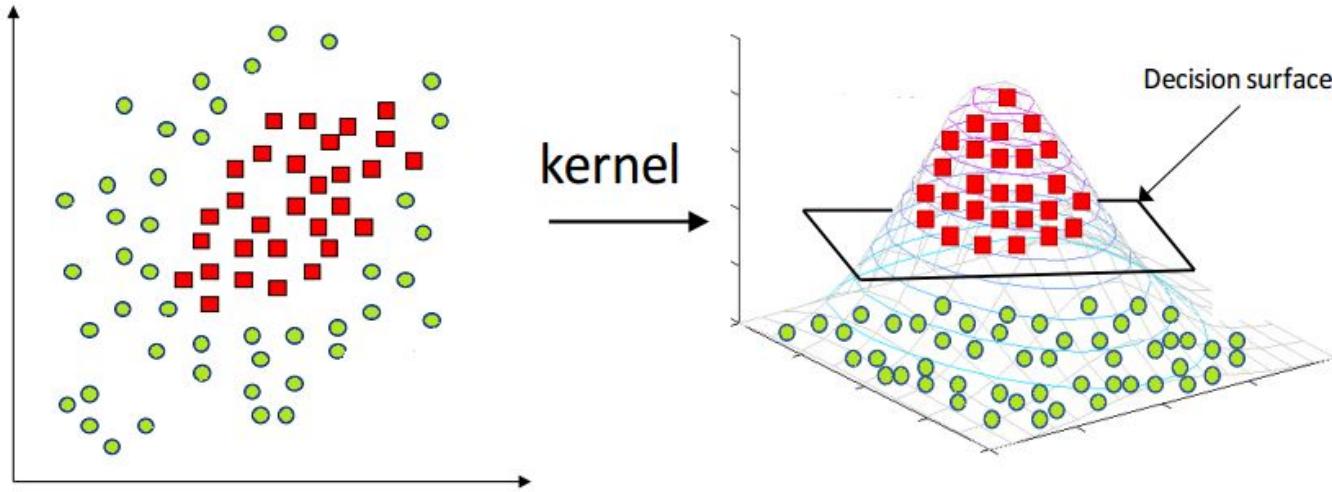


# Logistic Regression

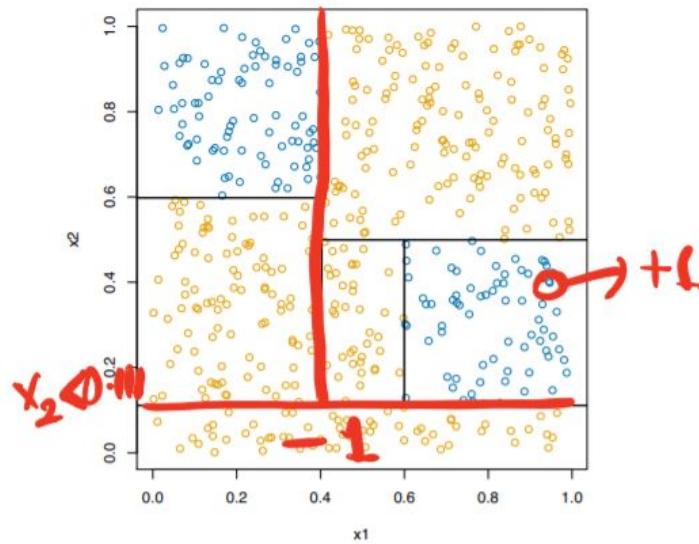
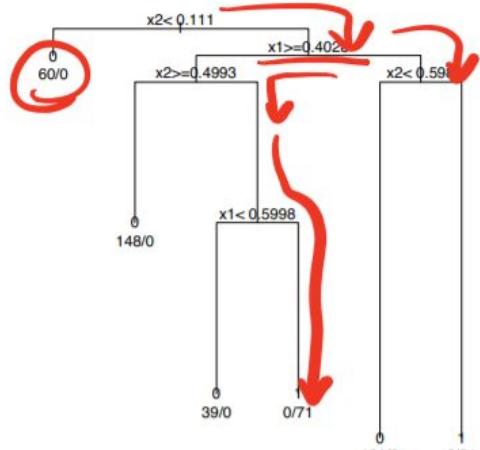


# Support Vector Machines

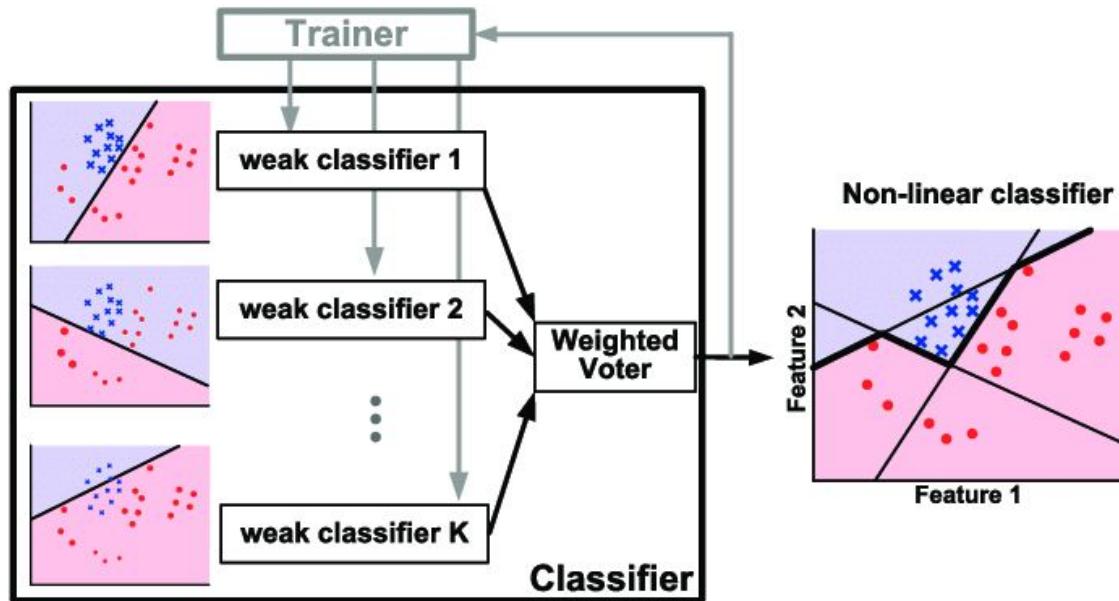




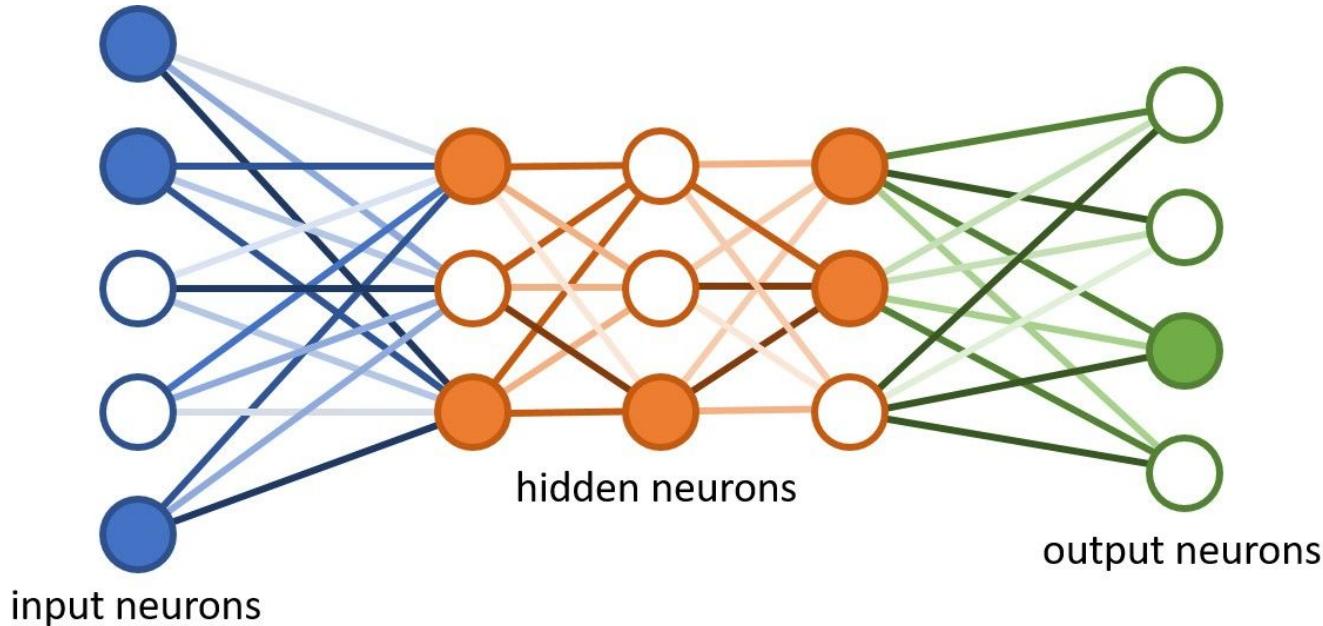
# Decision Trees



# Boosting

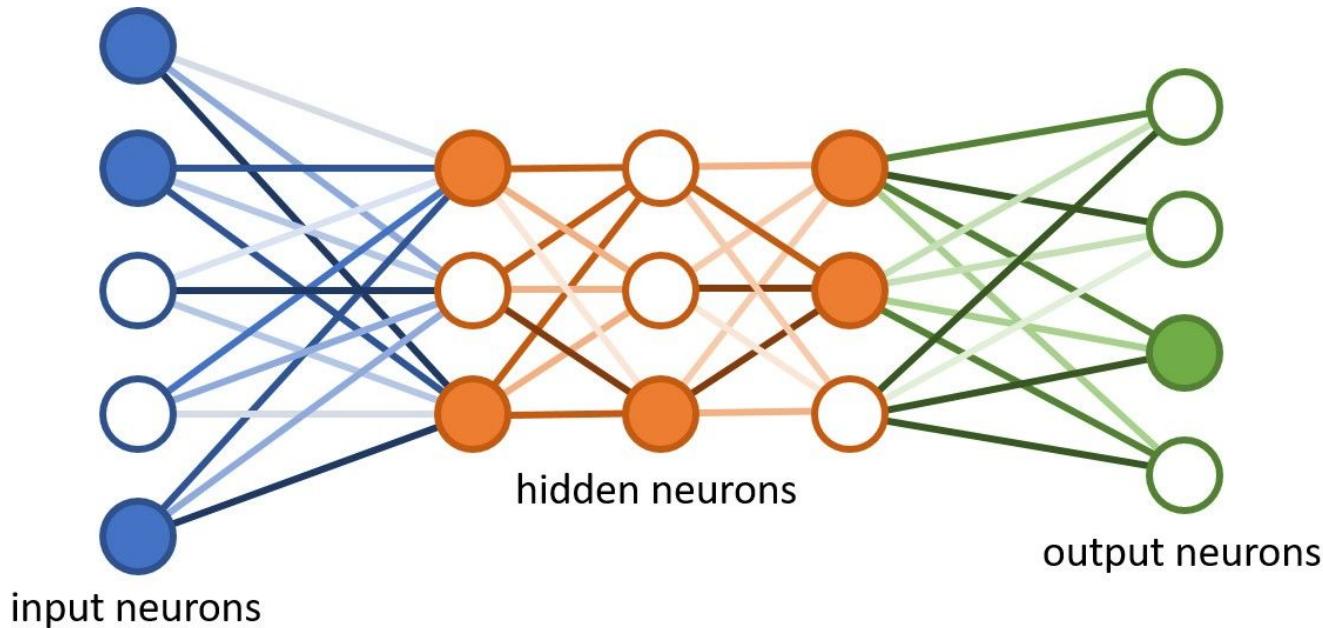


# Neural Networks

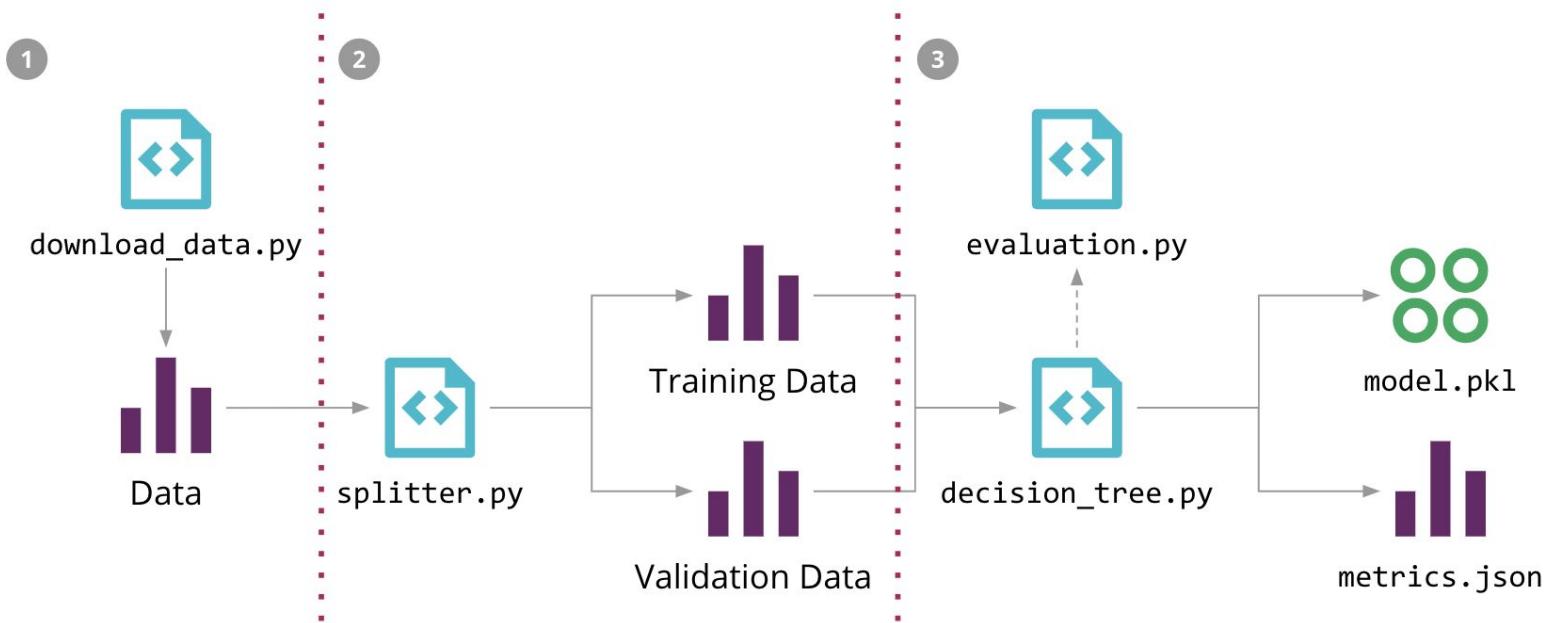


Deep  
Learning!

# Neural Networks



# ML Pipeline



# Feature Engineering

Size	Bedrooms	Bathrooms	Address	Price
3,012 sqft	5	2.5	5022 Forbes Avenue	\$500,000
1,402 sqft	2	1	241 Walnut Street	\$94,000
2,331 sqft	3	2	302 Craig Street	\$156,000

# Feature Engineering

Size	Bedrooms	Bathrooms	Distance to University	Price
3,012 sqft	5	2.5	0.8 miles	\$500,000
1,402 sqft	2	1	2.2 miles	\$94,000
2,331 sqft	3	2	1.3 miles	\$156,000

# Feature Engineering

**IN**

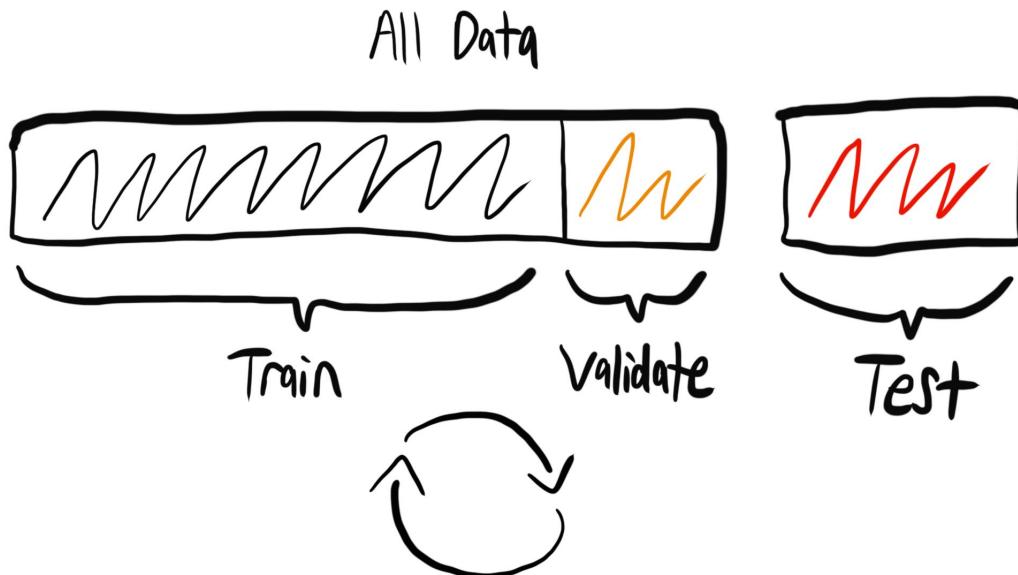


**OUT**

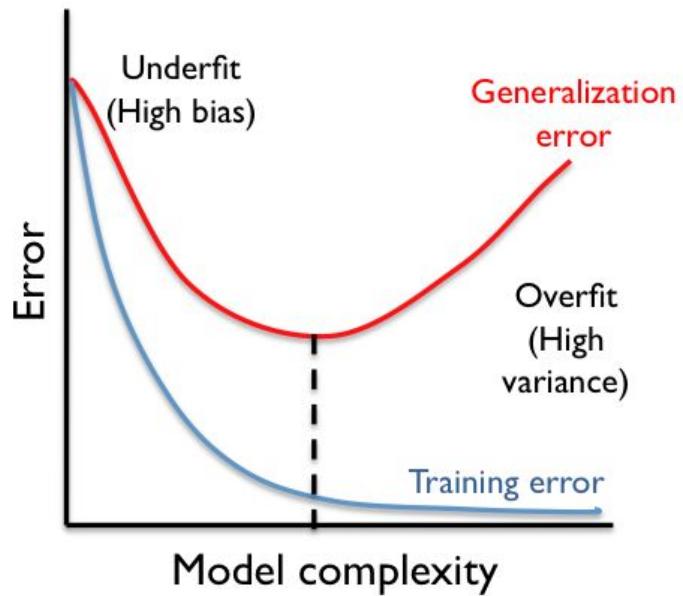


=

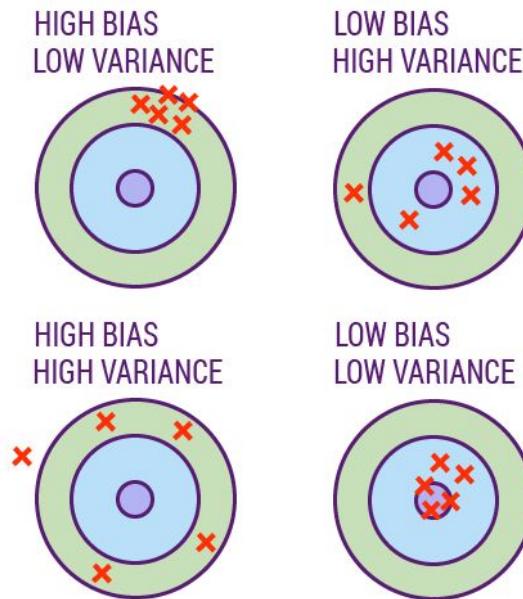
# Model Evaluation



# Model Evaluation



# Model Evaluation



# Recommended Classes

10-301 Introduction to Machine Learning

36-462 Data Mining

15-281 Artificial Intelligence

15-388 Practical Data Science

Coursera Machine Learning

# Machine Learning Intuition

# RESUME

Street Name: 1  
70000 City Name  
Tel: 0000 5555555  
E-Mail: emailname@server.com

## SUMMARY

- Experience in commercial engines development
- Expert knowledge in programming
- Strong experience in software design and architecture, animation, network programming, performance optimization
- 10 years of development experience. Worked on projects in various industries.
- Management of a small team of engineers

## WORK EXPERIENCE

07/2007 - Present Company Name Ltd. (United States)  
Lead Position Name

Working on new innovative project

- Sed sed ipsum et tortor ornare ullamcorper nec quis orci.
- Suspendisse nec urna sit amet arcu volutpat imperdiet vitae et velit.
- Donec et ipsum interdum, vulputate augue eu, aliquam ipsum.
- Integer sed turpis tempus sem laoreet pellentesque vitae tincidunt diam.
- Maecenas mattis mauris non neque fermentum, vel gravida turpis dignissim.
- Aliquam rhoncus quam eu eros ullamcorper laculis.

01/2005 – 07/2007 Company GmbH  
Position Name

Worked mostly on engine development:

- Nulla non metus id neque tempor suscipit.
- Suspendisse bibendum elit et nulla euismod, vitae aliquet i...
- accumsan.

# Resumes

School	Major	Python?	Performance
MIT	Computer Science	Yes	4.12
CMU	Information Systems	Yes	4.75
Community College	Computer Science	No	2.35

# Resumes

Top Tech School?	Major	Python?	Performance
Yes	Computer Science	Yes	4.12
Yes	Information Systems	Yes	4.75
No	Computer Science	No	2.35

# Resumes

Top Tech School?	Major	Python?	Performance
Yes	Computer Science	Yes	4.12
Yes	Information Systems	Yes	4.75
No	Computer Science	No	2.35

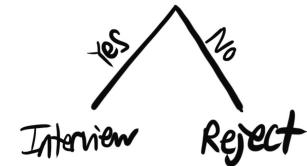
# Resumes

Top Tech School?	Major	Python?	Performance
Yes	Computer Science	Yes	4.12
Yes	Information Systems	Yes	4.75
No	Computer Science	No	2.35
No	Information Systems	Yes	

# Resumes

Top Tech School?	Major	Python?	Performance
Yes	Computer Science	Yes	4.12
Yes	Information Systems	Yes	4.75
No	Computer Science	No	2.35

Predicted Performance > 4.0 ?





# Payments

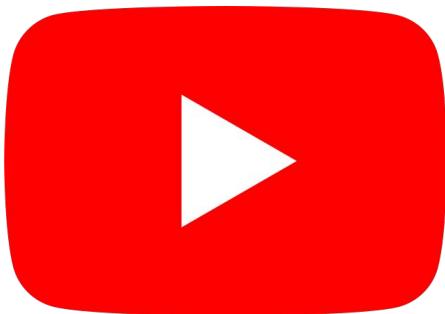
Time	Location	Merchant	Fraud?	Spend
4:50 PM	Pittsburgh	Gas	No	\$5.00
6:30 PM	Pittsburgh	Food	No	\$23.43
7:11 PM	New York	Jewelry	Yes	\$120.67

# Payments

Time	Location	Merchant	Fraud?	Spend
4:50 PM	Pittsburgh	Gas	No	\$5.00
6:30 PM	Pittsburgh	Food	No	\$23.43
7:11 PM	New York	Jewelry	Yes	\$120.67

# Payments

Time	Location	Merchant	Fraud?	Spend
4:50 PM	Pittsburgh	Gas	No	\$5.00
6:30 PM	Pittsburgh	Food	No	\$23.43
7:11 PM	New York	Jewelry	Yes	\$120.67
10:11 PM	Pittsburgh	Movie		\$21.02



YouTube

# Uploads

Views	Likes	Title	Length	Time
2,132,434	43,945	...	05:35	11:00 AM
534,938	32,583	...	17:34	02:30 PM
321,453	4,032	...	14:32	04:23 AM

# Uploads

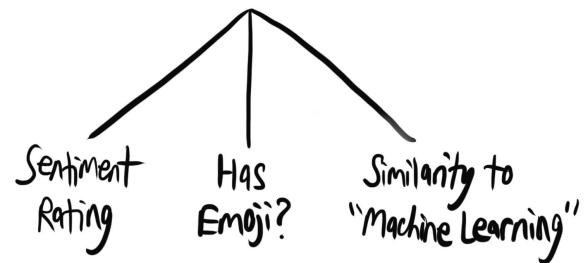
Views	Likes	Title	Length	Time of Day
2,132,434	43,945	...	05:35	Morning
534,938	32,583	...	17:34	Afternoon
321,453	4,032	...	14:32	Afternoon

# Uploads

Views	Likes	Title	Length	Time of Day
2,132,434	43,945	...	05:35	Morning
534,938	32,583	...	17:34	Afternoon
321,453	4,032	...	14:32	Afternoon

# Uploads

Views	Likes	Title	Length	Time of Day
2,132,434	43,945	...	05:35	Morning
534,938	32,583	...	17:34	Afternoon
321,453	4,032	...	14:32	Afternoon



# Uploads

Views	Likes	Title	Length	Time of Day
2,132,434	43,945	...	05:35	Morning
534,938	32,583	...	17:34	Afternoon
321,453	4,032	...	14:32	Afternoon
		...	02:32	Night

# THE WALL STREET JOURNAL.

English Edition ▾ | October 15, 2019 | Print Edition | Video

Calvin Lui ▾

EXPLORE JOURNAL REPORTS

[Home](#)[World](#)[U.S.](#)[Politics](#)[Economy](#)[Business](#)[Tech](#)[Markets](#)[Opinion](#)[Life & Arts](#)[Real Estate](#)[WSJ. Magazine](#)Search 

◆ WSJ NEWS EXCLUSIVE

## Drug Distributors Are in Talks to Pay \$18 Billion to Settle Opioid Litigation

Three major drug distributors are in talks to pay \$18 billion to settle sweeping litigation brought by state and local governments blaming them for fueling the opioid crisis.

## Pence, Giuliani Say They Won't Give Congress Documents in Trump Impeachment Inquiry

The office of Vice President Mike Pence declined a records request from House committees leading an impeachment inquiry into President Trump, and Rudy Giuliani, President Trump's personal lawyer, said he doesn't intend to comply with a congressional subpoena.

- Ex-Rep. Sessions Subpoenaed Over Interactions With Giuliani



LIVE ANALYSIS

### Candidates Disagree on Breaking Up Big Tech

Sen. Elizabeth Warren faced pushback for her proposal to break up Amazon, Google and Facebook for having too big a share of the online marketplace. Although other candidates agreed that large tech companies should be reined in, they disagreed over what to do about it.

- Warren Pressed on Health-Care Plan at Debate
- Who's In and Who's Out of the 2020 Presidential Race
- What the Candidates Are (and Aren't) Saying

WOMEN IN THE WORKPLACE

### A Lot of Women Work in Health



**So You Make \$100,000? It Still Might Not Be Enough to Buy a Home.**



**Elite M.B.A. Programs Report Steep Drop in Applications**



US EUROPE ASIA FX RATES FUTURES



DJIA	<b>27024.80</b>	<b>237.44</b>	<b>0.89%</b>
S&P 500	2995.68	29.53	1.00%
Nasdaq	8148.71	100.06	1.24%
Russell 2000	1523.30	17.87	1.19%
DJ Total Mkt	30615.45	301.03	0.99%

[View All Market Data →](#)

## OPINION

### Warren and Sanders, Policy Mates

By *The Editorial Board / Review & Outlook*

**Trump Speaks Loudly, Carries a Toothpick**

By *William A. Galston / Politics & Ideas*

### When China Won't Play Ball

By *Erica V.*

# Headlines

Views	Title	Text	Date	Publisher
327,435	...	...	02/16/2019	NYTimes
239,425	...	...	03/01/2019	WSJ
40,241	...	...	08/15/2019	69420.com

# Headlines

Views	Title	Text	Date	Publisher
327,435	...	...	02/16/2019	NYTimes
239,425	...	...	03/01/2019	WSJ
40,241	...	...	08/15/2019	69420.com

# Headlines

Views	Title	Text	Date	Publisher
327,435	...	...	02/16/2019	NYTimes
239,425	...	...	03/01/2019	WSJ
40,241	...	...	08/15/2019	69420.com

How about identifying Fake News?



# Marketing

User ID	Purchase 1	Purchase 2	Purchase 3	Purchase 4
abcde-1234	\$32	\$98	\$100	\$233
beigh-2432	\$4353	\$4234	\$945	\$123
kjasfg-2341	\$654	\$84	\$12	\$5

Is this data in a usable form?

# Marketing

User ID	Purchase 1	Purchase 2	Purchase 3	Purchase 4
abcde-1234	\$32	\$98	\$100	\$233
beigh-2432	\$4353	\$4234	\$945	\$123
kjasfg-2341	\$654	\$84	\$12	\$5

Can we augment this dataset?

# Marketing

Time	Gender	Age	Location	Avg. Spend
2:59 PM	Female	22	Pittsburgh	\$52
11:00 AM	Male	39	Dallas	\$30
10:32 PM	Female	27	New York	\$97

# Marketing

Time of Day	Gender	Age	Location	Avg. Spend
Afternoon	Female	22	Pittsburgh	\$52
Morning	Male	39	Dallas	\$30
Night	Female	27	New York	\$97

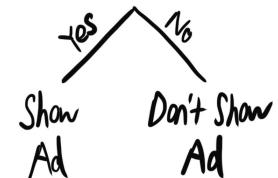
# Marketing

Time of Day	Gender	Age	Location	Avg. Spend
Afternoon	Female	22	Pittsburgh	\$52
Morning	Male	39	Dallas	\$30
Night	Female	27	New York	\$97

# Marketing

Time of Day	Gender	Age	Location	Avg. Spend
Afternoon	Female	22	Pittsburgh	\$52
Morning	Male	39	Dallas	\$30
Night	Female	27	New York	\$97

Spend > \$50?



# Marketing

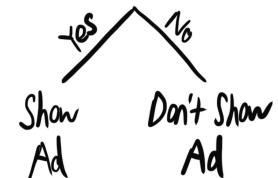
Time of Day	Gender	Age	Location	Avg. Spend
Afternoon	Female	22	Pittsburgh	\$52
Morning	Male	39	Dallas	\$30
Night	Female	27	New York	\$97

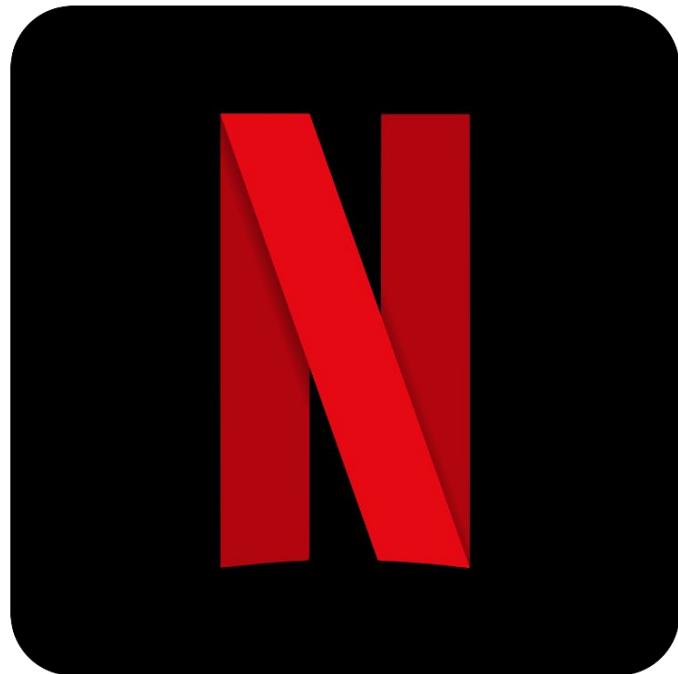
Cluster

# Marketing

Cluster	Time of Day	Gender	Age	Location	Avg. Spend
College Grad	Afternoon	Female	22	Pittsburgh	\$52
Jobless Dad	Morning	Male	39	Dallas	\$30
Party Monster	Night	Female	27	New York	\$97

Spend > \$50?





# Movies

User	Movie	Genre	Length	Rating
Calvin	Avengers	Action	01:30:00	5
Yvonne	Finding Nemo	Family	01:00:00	4
Matthew	Avengers	Action	01:30:00	2

# Movies

Movie Ratings	Zora	Sophie	Jordan	Ernie	Christie
<b>Harold and Kumar Escape..</b>	8	4	?	?	4
<b>Ted</b>	?	?	8	10	4
<b>Straight Outta Compton</b>	8	10	?	?	6

# Movies

## Collaborative Filtering

---

- Movies
- Songs
- Classes?

Movie Ratings	Zora	Sophie	Jordan	Ernie	Christie
<b>Harold and Kumar Escape..</b>	8	4	?	?	4
<b>Ted</b>	?	?	8	10	4
<b>Straight Outta Compton</b>	8	10	?	?	6

Search 22,431 datasets[Feedback](#)[Filter](#)[PUBLIC](#)[YOUR DATASETS](#)[FAVORITES](#)Sort by: **Hottest****17K Mobile Strategy Games**

Tristan

2 months

8 MB

10.0

1 File (CSV)



76

**New York City Airbnb Open Data**

Dgomonov

2 months

2 MB

10.0

2 Files (other, CSV)



815

**Forest Fires in Brazil**

Luís Gustavo Modelli

2 months

31 KB

9.4

1 File (CSV)



95

**Food.com Recipes and Interactions**

Shuyang Li

4 days

261 MB

10.0

8 Files (CSV, other)



43

**UFC-Fight historical data from 1993 to 2019**

Rajeev Warrier

3 months

2 MB

9.7

4 Files (CSV)



73

**Africa Economic, Banking and Systemic Crisis Data**

36

**tinyurl.com/btg-ml-f19**