HITCHHIKERS GUIDE TO ROCKING MACHINE LEARNING IN 2019

Session for NewBies

Sho Fola Soboyejo

Apr 26th, 2019

@shoreason

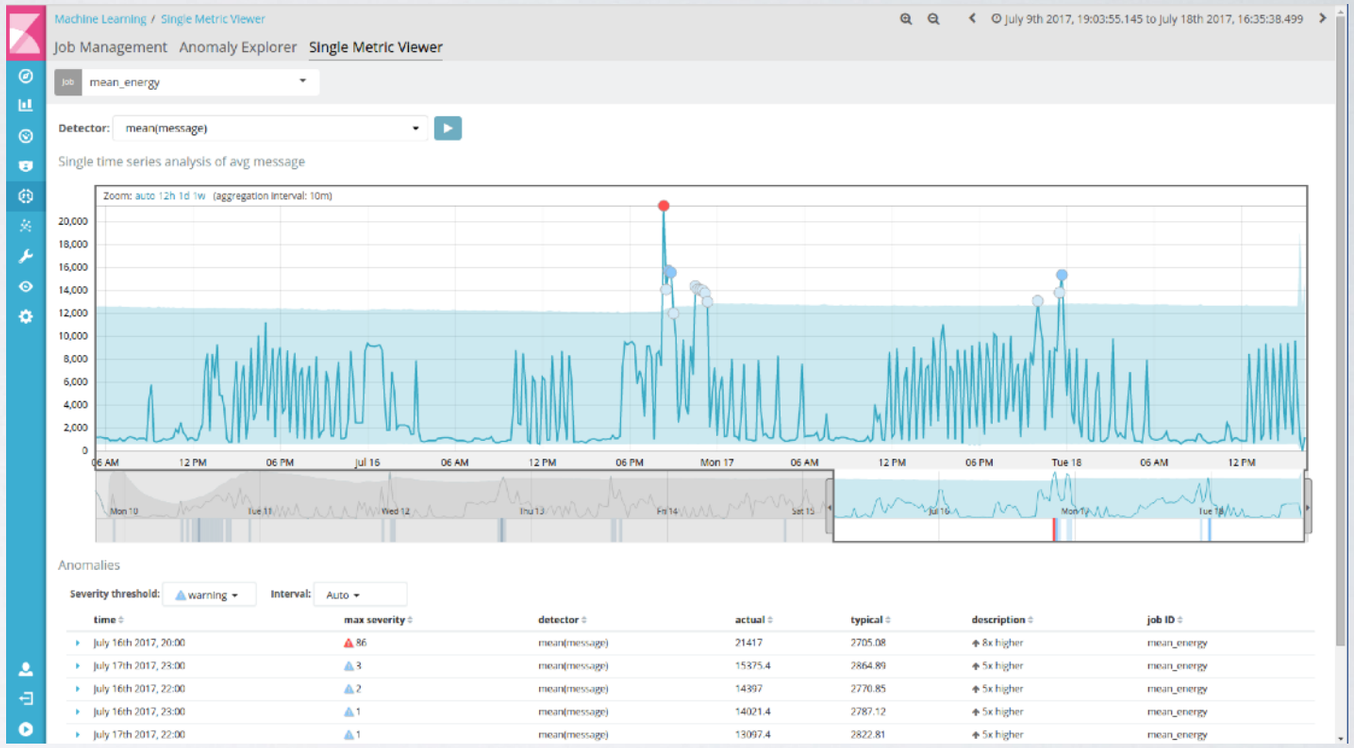
10,000 hours

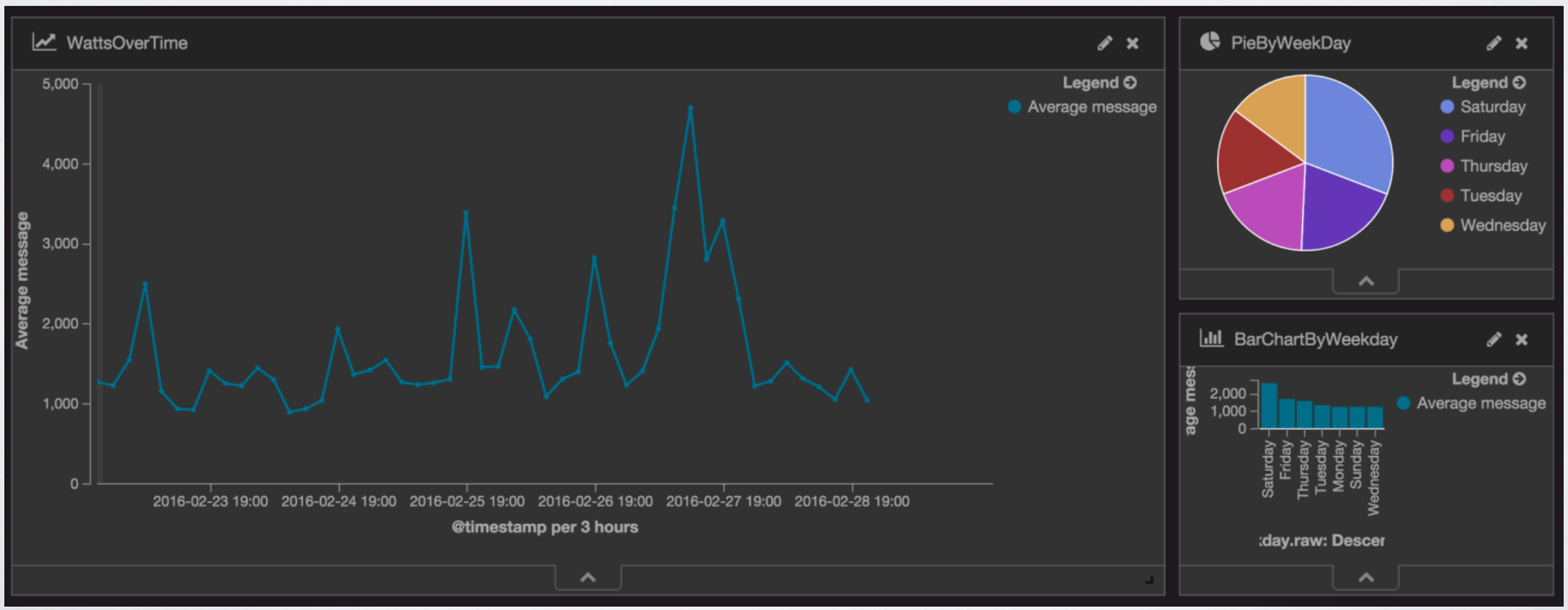
DON'T PANIC

Photo by: Douglas Adams

PERSONAL APPLICATIONS

Energy utilization anomaly detection





PERSONAL APPLICATIONS

Recommendation Engine for nearby fun activities

	1	2
Kings Island	Kings Island	The Beach waterpark
The Beach waterpark	Kings Island	The Beach waterpark
Cininnati Aquarium	Creation Meseum	Cininnati Aquarium
Cincinnati Arts meseum	Creation Meseum	Cininnati Aquarium
Paul Brown Stadium	Paul Brown Stadium	cinebistro

	3	4
Kings Island	Great Wolf Lodge	Entertrainment Junction
The Beach waterpark	Great Wolf Lodge	Entertrainment Junction
Cininnati Aquarium	Cincinnati Arts meseum	Cincinnati Children's Meseum
Cincinnati Arts meseum	Cincinnati Arts meseum	Cincinnati Children's Meseum
Paul Brown Stadium	Entertrainment Junction	Pump it Up

	5
Kings Island	Pump it Up
The Beach waterpark	Pump it Up
Cininnati Aquarium	Great Wolf Lodge
Cincinnati Arts meseum	Great Wolf Lodge
Paul Brown Stadium	Jumping Jacks

	6
Kings Island	Jumping Jacks
The Beach waterpark	Jumping Jacks
Cininnati Aquarium	COSI
Cincinnati Arts meseum	COSI
Paul Brown Stadium	Regal Cinemas Deerfield Towne Center 16

	7	8
Kings Island	COSI	Liberty Center
The Beach waterpark	COSI	Liberty Center
Cininnati Aquarium	Smale Park	Kings Island
Cincinnati Arts meseum	Smale Park	Kings Island
Paul Brown Stadium	Liberty Center	Run, Jump and Play

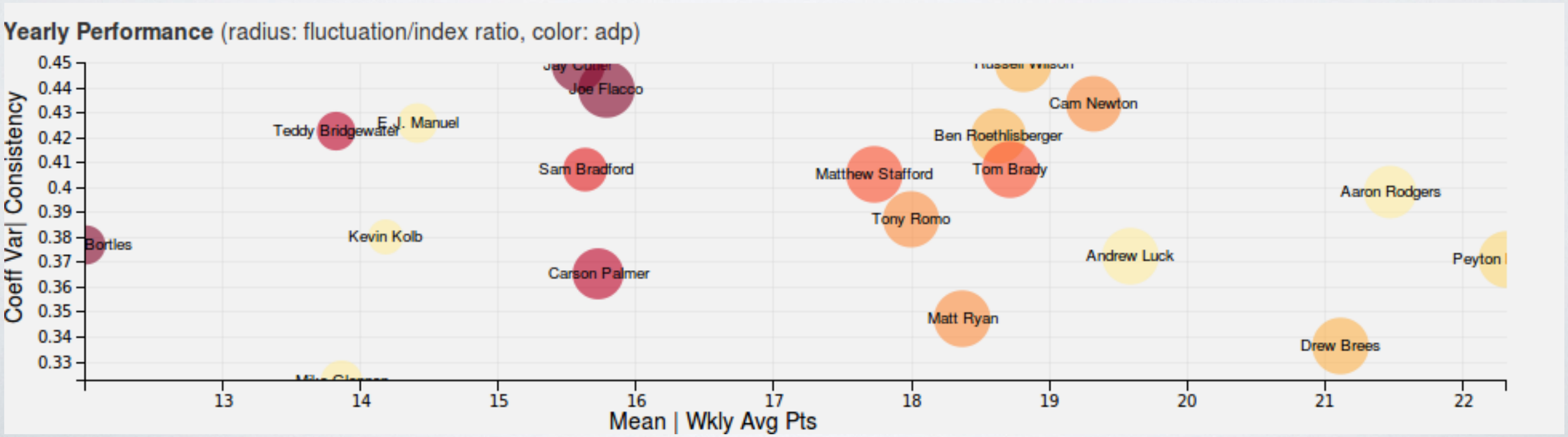
	9	10
Kings Island	Run, Jump and Play	Smale Park
The Beach waterpark	Run, Jump and Play	Smale Park
Cininnati Aquarium	The Beach waterpark	Voice of America Park
Cincinnati Arts meseum	The Beach waterpark	Voice of America Park
Paul Brown Stadium	Great Wolf Lodge	COSI

PERSONAL APPLICATIONS

Creating clusters and tiers for fantasy football players plus predicting performance

```
{0: 0.0, 1: 0.1111111111111111, 2: 0.07692307692307693, 3: 0.0,
number of estimated clusters : 5
```

	Player	ATT	YDS	Y/A	LG	20+	TD	REC	TGT	YDS.1	\
0	David Johnson (ARI)	293	1,239	4.2	58	6	16	80	120	879	
2	Le'Veon Bell (PIT)	261	1,268	4.9	44	4	7	75	94	616	
4	DeMarco Murray (TEN)	293	1,287	4.4	75	4	9	53	67	377	
5	Devonta Freeman (ATL)	227	1,079	4.8	75	7	11	54	65	462	
7	Mark Ingram (NO)	205	1,043	5.1	75	4	6	46	58	319	
14	Todd Gurley (LAR)	278	885	3.2	24	2	6	43	58	327	
15	Bilal Powell (NYJ)	131	722	5.5	35	4	3	58	75	388	
104	De'Anthony Thomas (KC)	4	29	7.3	23	1	0	7	9	35	
109	Terrell Watson (PIT)	9	28	3.1	8	0	1	1	1	5	
110	Malcolm Brown (LAR)	18	39	2.2	10	0	0	3	3	46	
112	Adrian Peterson (NO)	37	72	1.9	13	0	0	3	6	8	
190	Samkon Gado (TEN)	0	0	0.0	0	0	0	0	0	0	
204	Shane Bannon (KC)	0	0	0.0	0	0	0	0	0	0	
257	Corey Clement (PHI)	0	0	0.0	0	0	0	0	0	0	
258	James Conner (PIT)	0	0	0.0	0	0	0	0	0	0	
259	Brian Hill (ATL)	0	0	0.0	0	0	0	0	0	0	
266	Austin Ekeler (LAC)	0	0	0.0	0	0	0	0	0	0	
267	Tion Green (DET)	0	0	0.0	0	0	0	0	0	0	



PERSONAL APPLICATIONS

Text Summarizer to summarize news articles

Algorithmia
Gensim
✓ Naive 1.0
Naive 2.0
All

- Algo in the works: 3

Content to Summarize

Algorithm of choice Naive 1.0

Content

It's been a disappointing start for Johnson, playing in a new system that doesn't appear to be tailored to his skill set.

A touchdown helped salvage his opening day fantasy performance, but in Week 2, he was held under 50 rushing yards and saw just two targets in a blowout loss to the Rams. New head coach Steve Wilks promised Johnson would be used more as a receiver, though the lack of success by the offense as a whole is equally to blame.

If you can find a running back-needy owner who is willing to pay close to draft day price for Johnson, it might be best to move him now before everyone realizes how bad the Cardinals' attack will be this season.

Submit

naive_one with summary ratio of 81.69440242057489 boiled it down to:
It's been a disappointing start for Johnson, playing in a new system that doesn't appear to be tailored to his skill set.

80/20 Rule

CHOOSING A PATH

What are you saying No to?

LITERACY

Learning to know

VS

COMPETENCE

Learning to Do

ON GETTING STARTED

- Being clear where you are starting from

- Know where you are headed before you start

Developer

Business Decision Maker

Data Scientist

Data Platform Engineer

IS THIS YOU?

- Learning a new instrument
- Top Down = Starting with a song
- Bottom Up = Learn A lot of Theory

PLAYING THE WHOLE GAME

- David Perkins
- Learning things on as needed basis

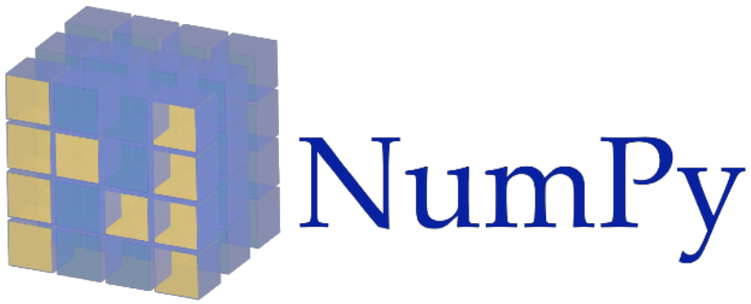
WORK ON THE HARD PARTS

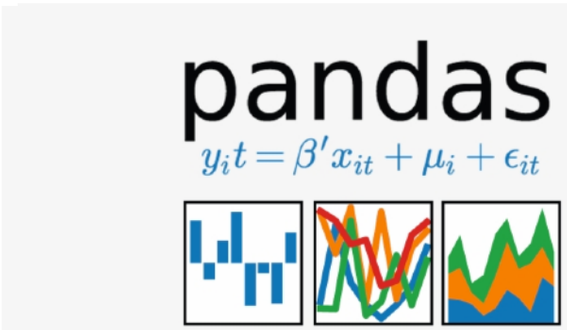
- Zone of discomfort
- Circle of 5ths & Riffs
- Getting and cleaning data
- Setting up your cloud env
- Moving from Python Notebook to CMD Line


RESOURCES

- Fast.ai - Intro to ML Course
- Udacity ML Course
- Machine Learning is fun - Medium Blog (Literacy)
- Deep Learning - Online book (Literacy)
- Follow Siraj Raval, Sentdex and Corey Schafer on YouTube

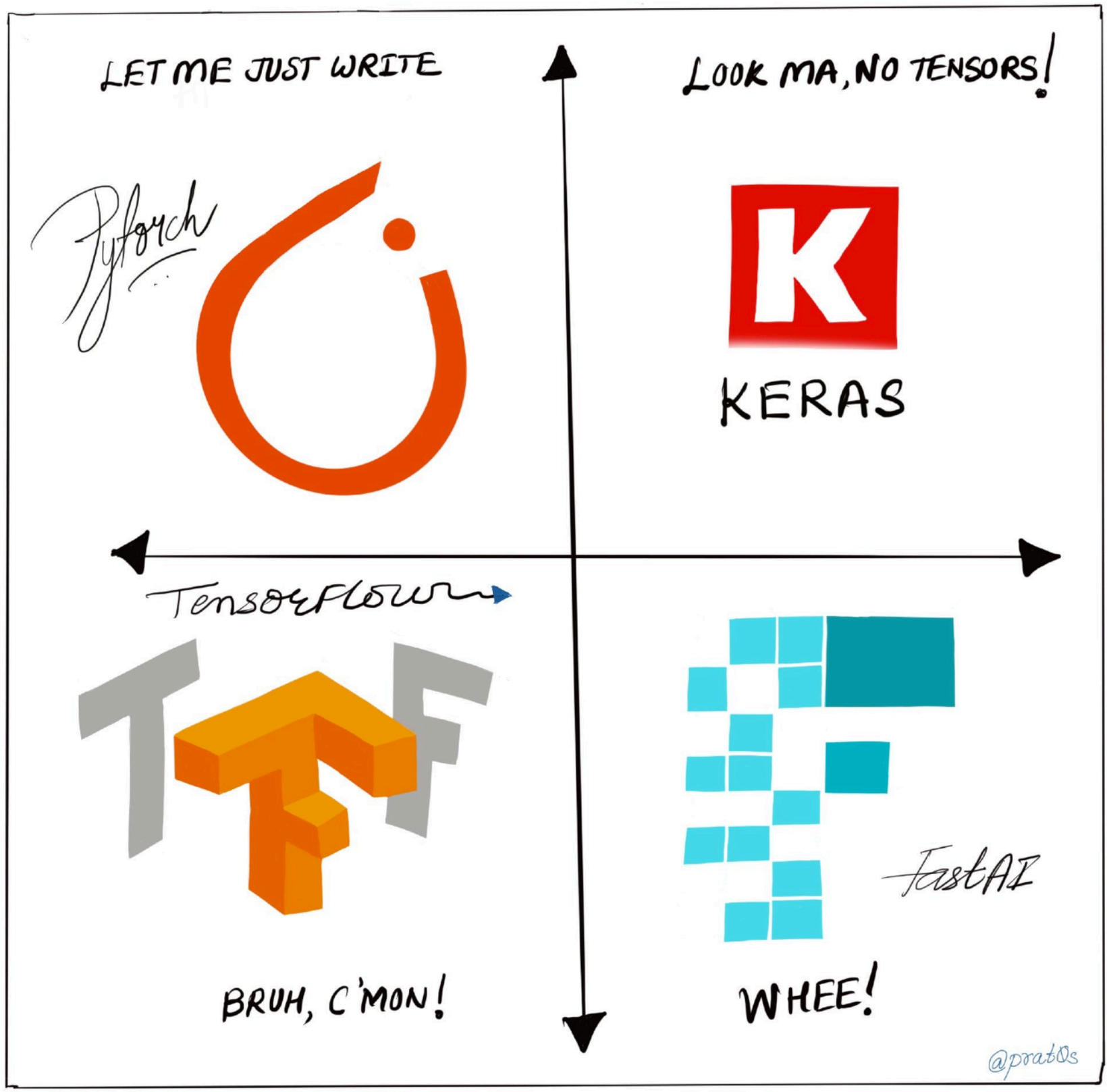
LANGUAGES vs TOOLS





Machine Learning

Neural Networks & Deep Learning



PLATFORMS

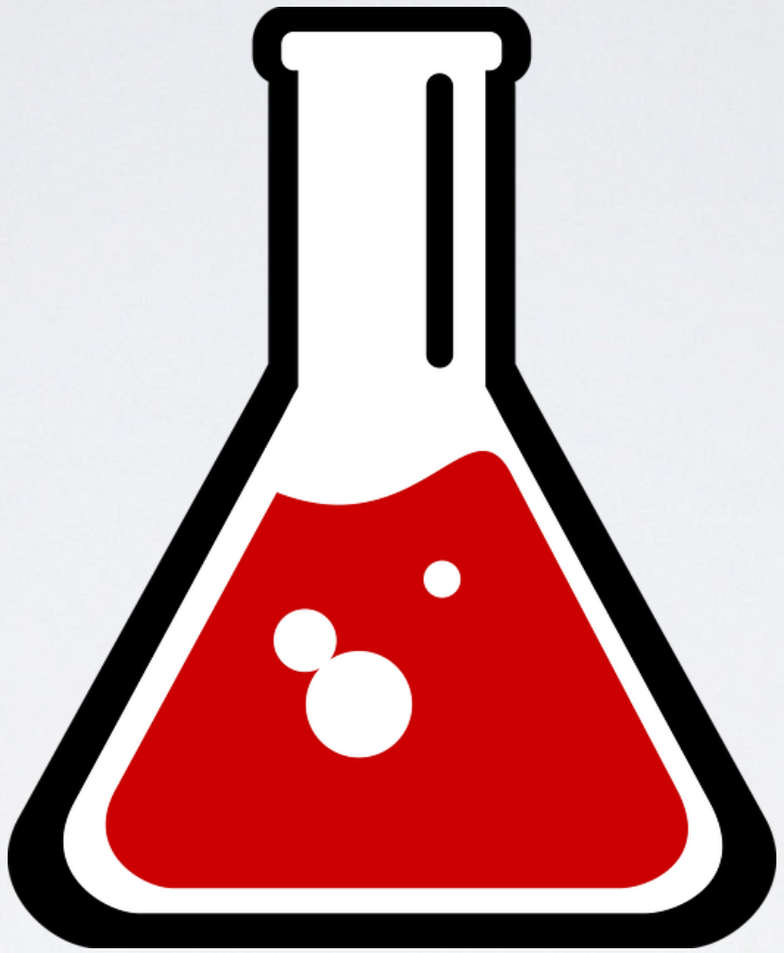
Google Cloud Platform

Paperspace

FINDING A PROJECT AND DATA

- Kaggle
- UCI ML Repository
- Fast.ai Datasets
- Your own personal project
- Hackathons

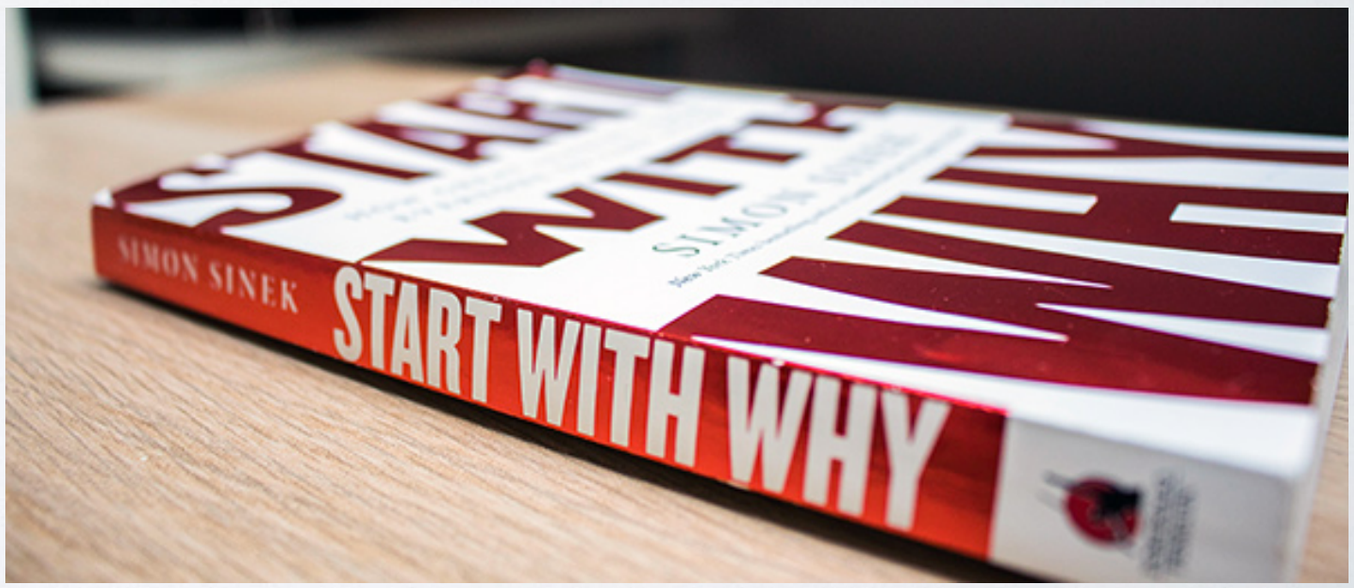
ACHIEVING SUCCESS



RIGHT MIX OF EXPERIENCE AND MOTIVATION

FINDING MOTIVE

- **Staying competitive (in your industry)**
- Improving your productivity
- Being in the driver's seat
- Managing Bias



Cranfield University Blogs

TOM SIMONITE BUSINESS 03.13.19 10:00 AM

CHINA IS CATCHING UP TO THE US IN AI RESEARCH—FAST

THE VERGE

China is about to overtake America in AI research

China will publish more of the most-cited 50 percent of papers than America for the first time this year

By James Vincent | Mar 14, 2019, 7:03am EDT

WIRED

The New York Times

Trump Signs Executive Order Promoting Artificial Intelligence

Feb. 11, 2019

FINDING MOTIVE

- Staying competitive (in your industry)
- **Improving your productivity**
- Being in the driver's seat
- Managing Bias

Cranfield University Blogs

cia documents reveal iot-specific televisions can be used to secretly record conversations .
cybercriminals who initiated the attack managed to commandeer a large number of internet-connected devices in current use .
cia documents revealed that microwave ovens can spy on you - maybe if you personally don't suffer the consequences of the sub-par security of the iot .

Internet of Things (IoT) security breaches have been dominating the headlines lately . WikiLeaks's trove of CIA documents revealed that internet-connected televisions can be used to secretly record conversations . Trump's advisor Kellyanne Conway believes that microwave ovens can spy on you - maybe she was referring to microwave cameras which indeed can be used for surveillance . And don't delude yourself that you are immune to IoT attacks , with 96 % of security professionals responding to a new survey expecting an increase in IoT breaches this year . Even if you personally don't suffer the consequences of the sub-par security of the IoT , your connected gadgets may well be unwittingly cooperating with criminals . Last October , Internet service provider Dyn came under an attack that disrupted access to popular websites . The cybercriminals who initiated the attack managed to commandeer a large number of internet-connected devices (mostly DVRs and cameras) to serve as their helpers . As a result , cybersecurity expert Bruce Schneier has called for government regulation of the IoT , concluding that both IoT manufacturers and their customers don't care about the security of the 8.4 billion internet-connected devices in current use . Whether because of government regulation or good old-fashioned self-interest , we can expect increased investment in IoT security technologies . In its recently-released TechRadar report for security and risk professionals , Forrester Research discusses the outlook for the 13 most relevant and important IoT security technologies , warning that " there is no single , magic security bullet that can easily fix all IoT security issues . " Based on Forrester's analysis , here's my list of the 6 hottest technologies for IoT security : IoT network security : Protecting and securing the network connecting IoT devices to back-end systems on the internet . IoT network security is a bit more challenging than traditional network security because there is a wider range of communication protocols , standards , and device capabilities , all of which pose significant issues and increased complexity . Key capabilities include traditional endpoint security features such as antivirus and antimalware as well as other features such as firewalls and intrusion prevention and detection systems . Sample vendors : Bayshore Networks , Cisco , Darktrace , and Senrio . IoT authentication : Providing the ability for users to authenticate an IoT device , including managing multiple users of a single device (such as a connected car) , ranging from simple static password/pins to more robust authentication mechanisms such as two-factor

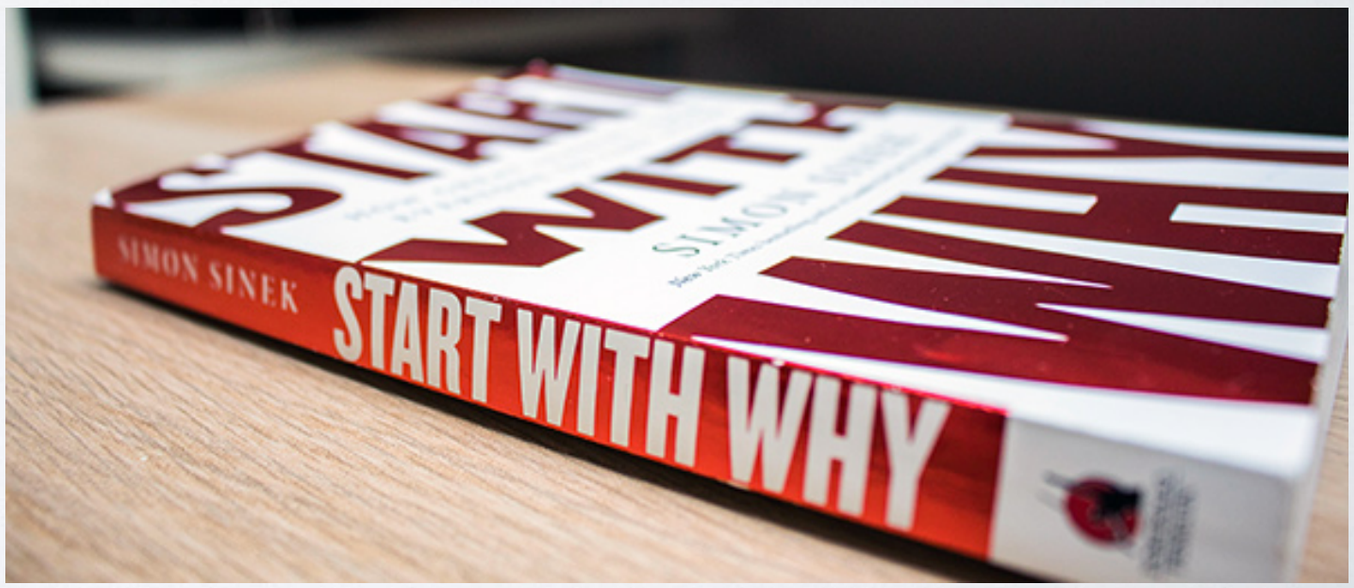
Numpy Error

This is the source of my error. Any thoughts on this? tab

```
input_data = np.zeros((568411, 15074, 84), dtype='float32')
```

FINDING MOTIVE

- Staying competitive (in your industry)
- Improving your productivity
- **Being in the driver's seat**
- Managing Bias

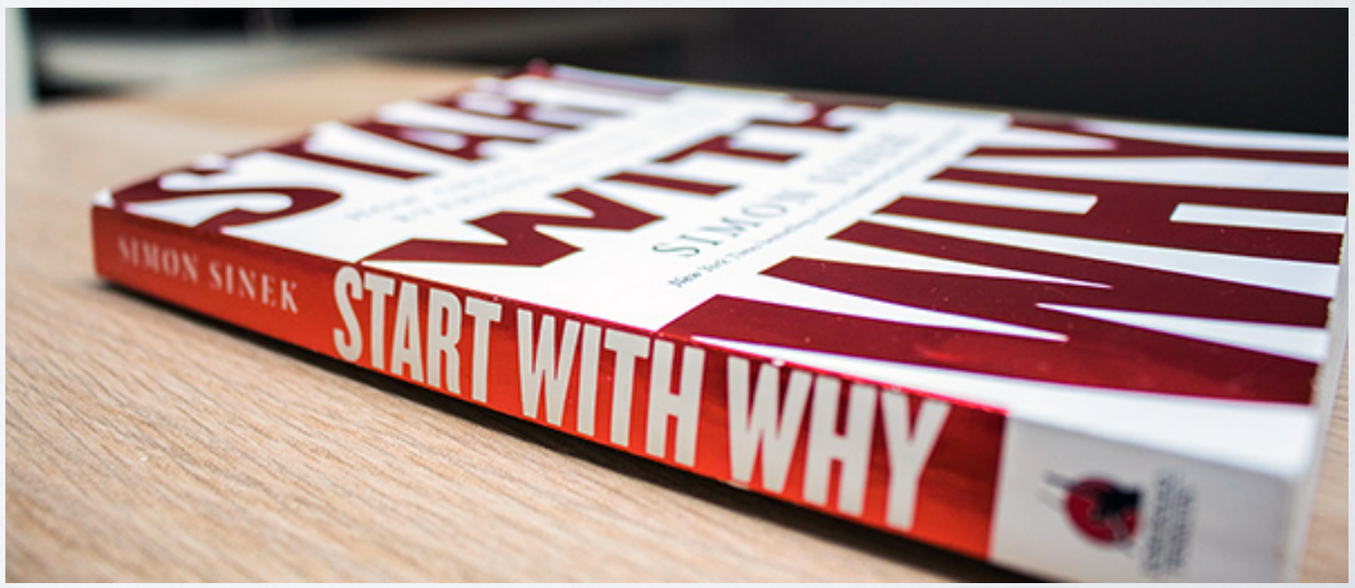


Cranfield University Blogs

“Predictive analytics will have an effect on who gets hired, is approved for a loan or sees an Ad”

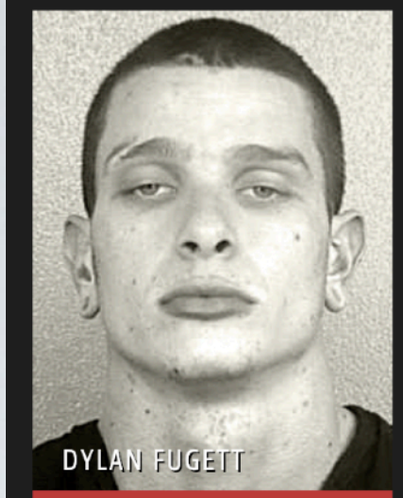
FINDING MOTIVE

- Staying competitive (in your industry)
- Improving your productivity
- Being in the driver's seat
- **Managing Bias**



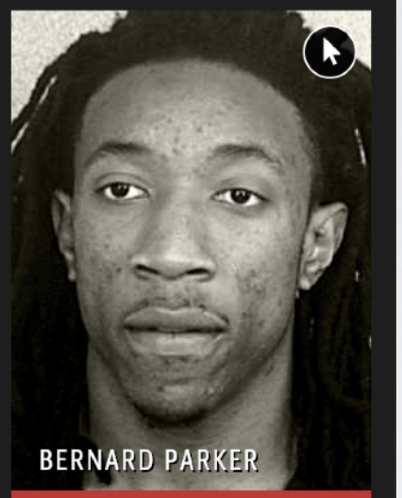
Cranfield University Blogs

Two Drug Possession Arrests



DYLAN FUGETT

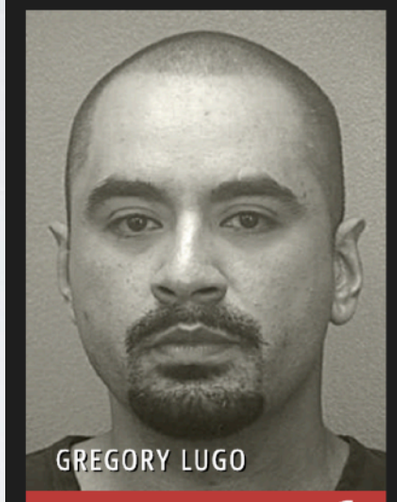
LOW RISK 3



BERNARD PARKER

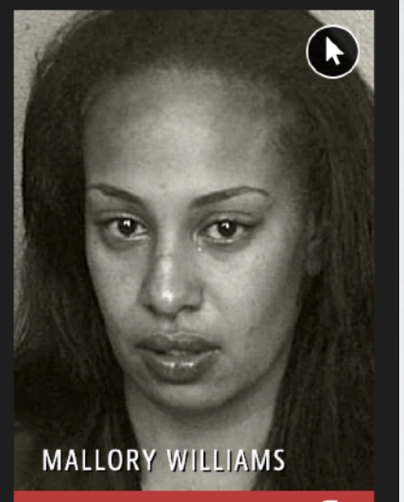
HIGH RISK 10

Two DUI Arrests



GREGORY LUGO

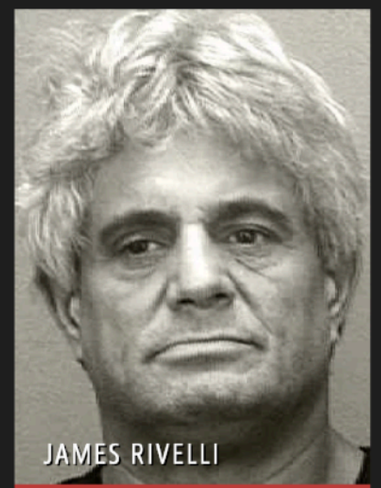
LOW RISK 1



MALLORY WILLIAMS

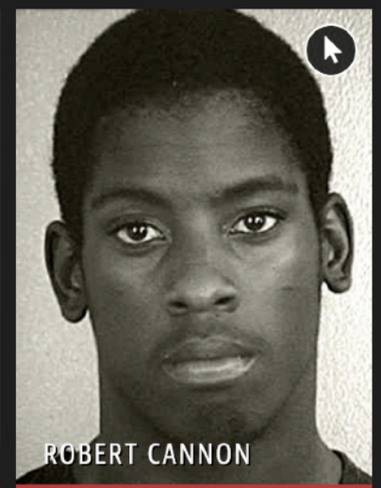
MEDIUM RISK 6

Two Shoplifting Arrests



JAMES RIVELLI

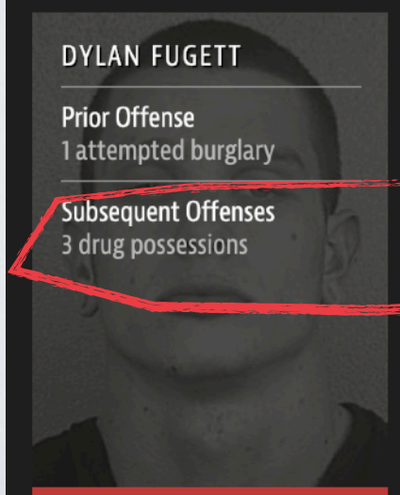
LOW RISK 3



ROBERT CANNON

MEDIUM RISK 6

Two Drug Possession Arrests

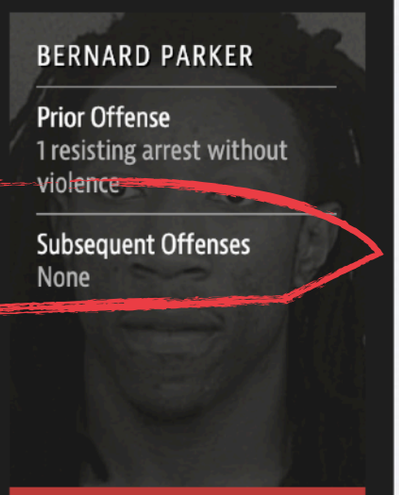


DYLAN FUGETT

Prior Offense
1 attempted burglary

Subsequent Offenses
3 drug possessions

LOW RISK 3



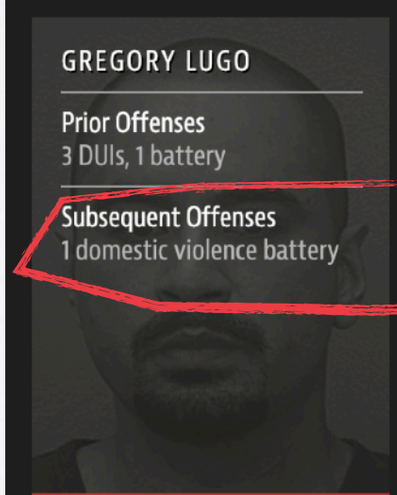
BERNARD PARKER

Prior Offense
1 resisting arrest without violence

Subsequent Offenses
None

HIGH RISK 10

Two DUI Arrests

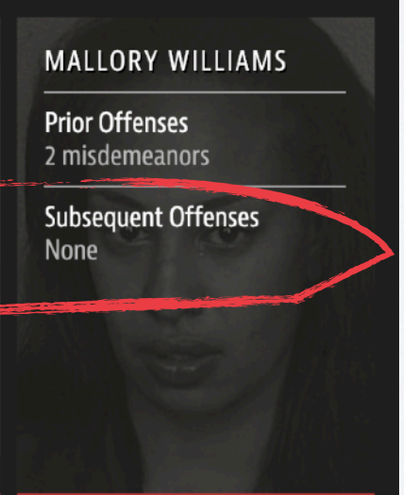


GREGORY LUGO

Prior Offenses
3 DUIs, 1 battery

Subsequent Offenses
1 domestic violence battery

LOW RISK 1



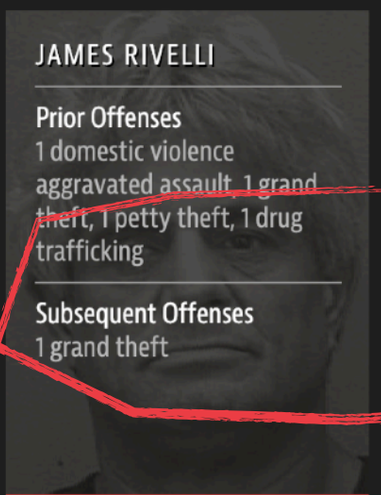
MALLORY WILLIAMS

Prior Offenses
2 misdemeanors

Subsequent Offenses
None

MEDIUM RISK 6

Two Shoplifting Arrests

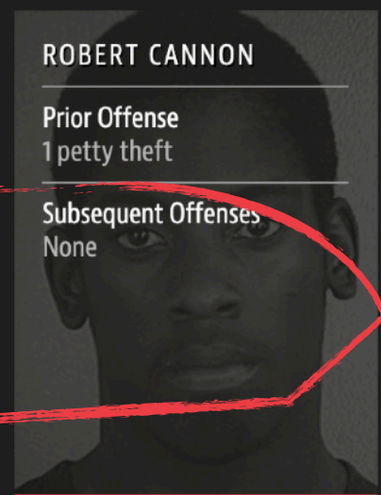


JAMES RIVELLI

Prior Offenses
1 domestic violence aggravated assault, 1 grand theft, 1 petty theft, 1 drug trafficking

Subsequent Offenses
1 grand theft

LOW RISK 3



ROBERT CANNON

Prior Offense
1 petty theft

Subsequent Offenses
None

MEDIUM RISK 6

English | Turkish | Spanish | Detect language

She is a doctor.
He is a nurse.

31/5000

English | Turkish | Spanish | Translate

O bir doktor.
O bir hemşire.

English | Turkish | Spanish | Turkish - detected

O bir doktor.
O bir hemşire

28/5000

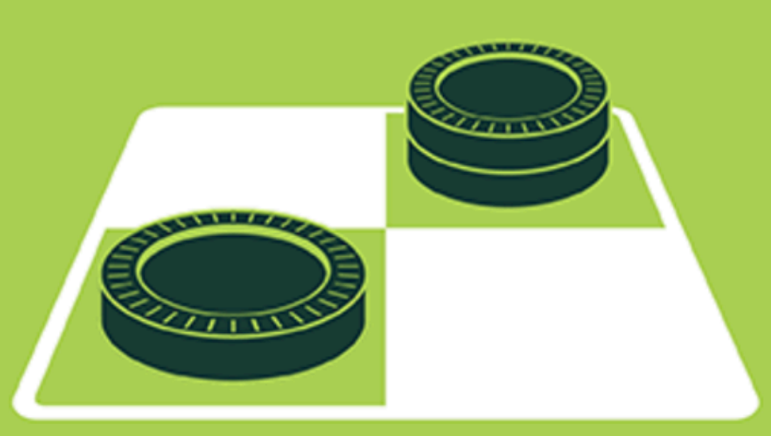
English | Turkish | Spanish | Translate

He is a doctor.
She is a nurse

The Timeline

ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



MACHINE LEARNING

Machine learning begins to flourish.

DEEP LEARNING

Deep learning breakthroughs drive AI boom.

1950's 1960's 1970's 1980's 1990's 2000's 2010's

Nvidia

10 Algorithms To Know

Support Vector Machines

Random Forest

Naive Bayes

Logistic Regression

Linear Regression

K-Nearest Neighbors

K-Means Clustering

Principal Component Analysis

Linear Discriminant Analysis

Neural Networks

"99% OF THE ECONOMIC VALUE LIES IN SUPERVISED LEARNING"

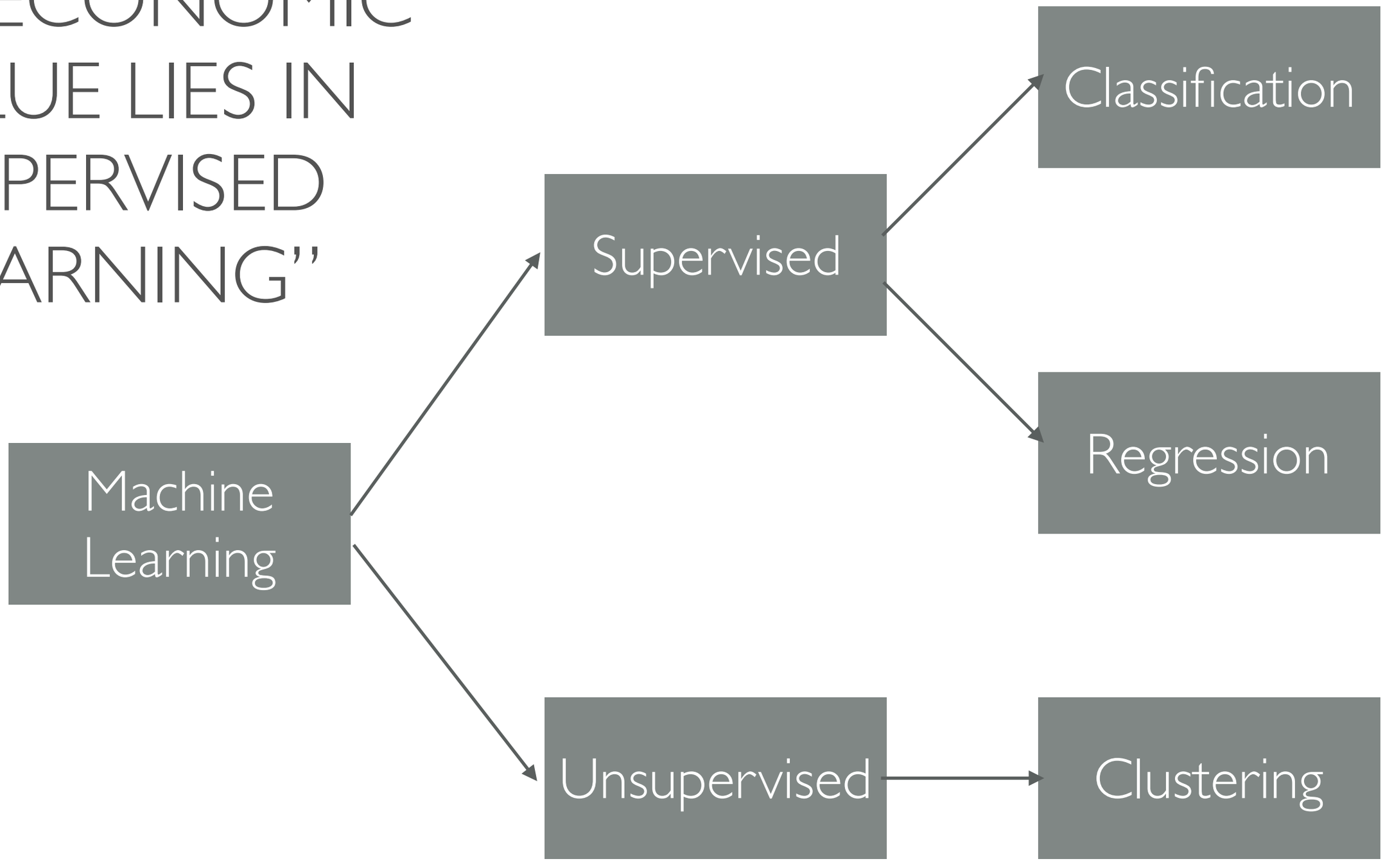


Photo by: Douglas Adams

Frequently Bought Together

Price for all three: $94.23

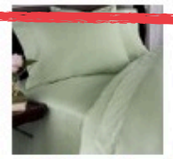
Add all three to Cart | Add all three to Wish List

Show availability and shipping details

- **This item:** 1500 Thread Count Queen 4pc Bed Sheet Set Egyptian Quality Deep Pocket White by Choice Linen $25.75
- Chezmoi Collection White Goose Down Alternative Comforter, Full/Queen with Corner Tab by Chezmoi Collection $44.49
- Dream Supreme Plus Gel Fiber-Filled Pillows, Standard (Set of 2) by Dream Supreme Plus $23.99

Customers Who Bought This Item Also Bought

Page 1 of 11



1500 Thread Count Queen 4pc Bed Sheet Set Egyptian Quality Deep Pocket Sage
(216)
$25.75

Dream Supreme Plus Gel Fiber-Filled Pillows, Standard (Set of 2)
(920)
$23.99

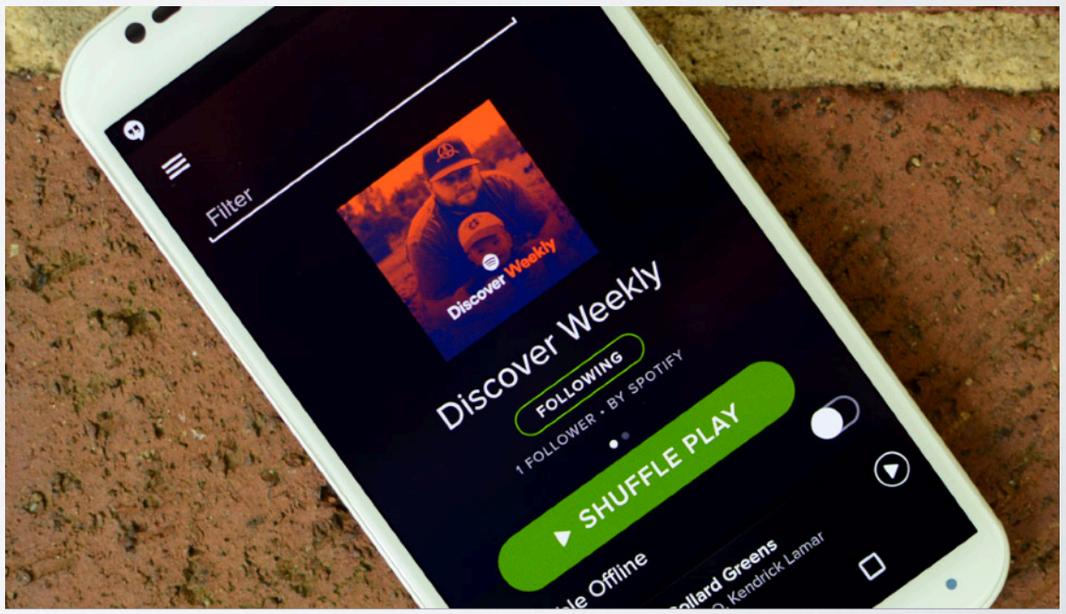
Royal Hotel's 1200 Thread Count Queen Size Siberian Goose Down Alternative Comforter 100% Egyptian ...
(32)
$84.99

Chezmoi Collection White Goose Down Alternative Comforter, Full / Queen with Corner Tab
(242)
$44.49

3pc Egyptian Duvet Cover Set 1200 Thread Count Queen Size White
(139)
$23.49

Queen Size White Goose Feather and Goose Down Pillows - Set of 2
(39)
$26.24

Clara Clark Goose Down Alternative Double Fill Comforter (Duvet) Queen Size, White / White
(95)
$36.99



ARTIFICIAL INTELLIGENCE (OR LACK OF IT)

VS

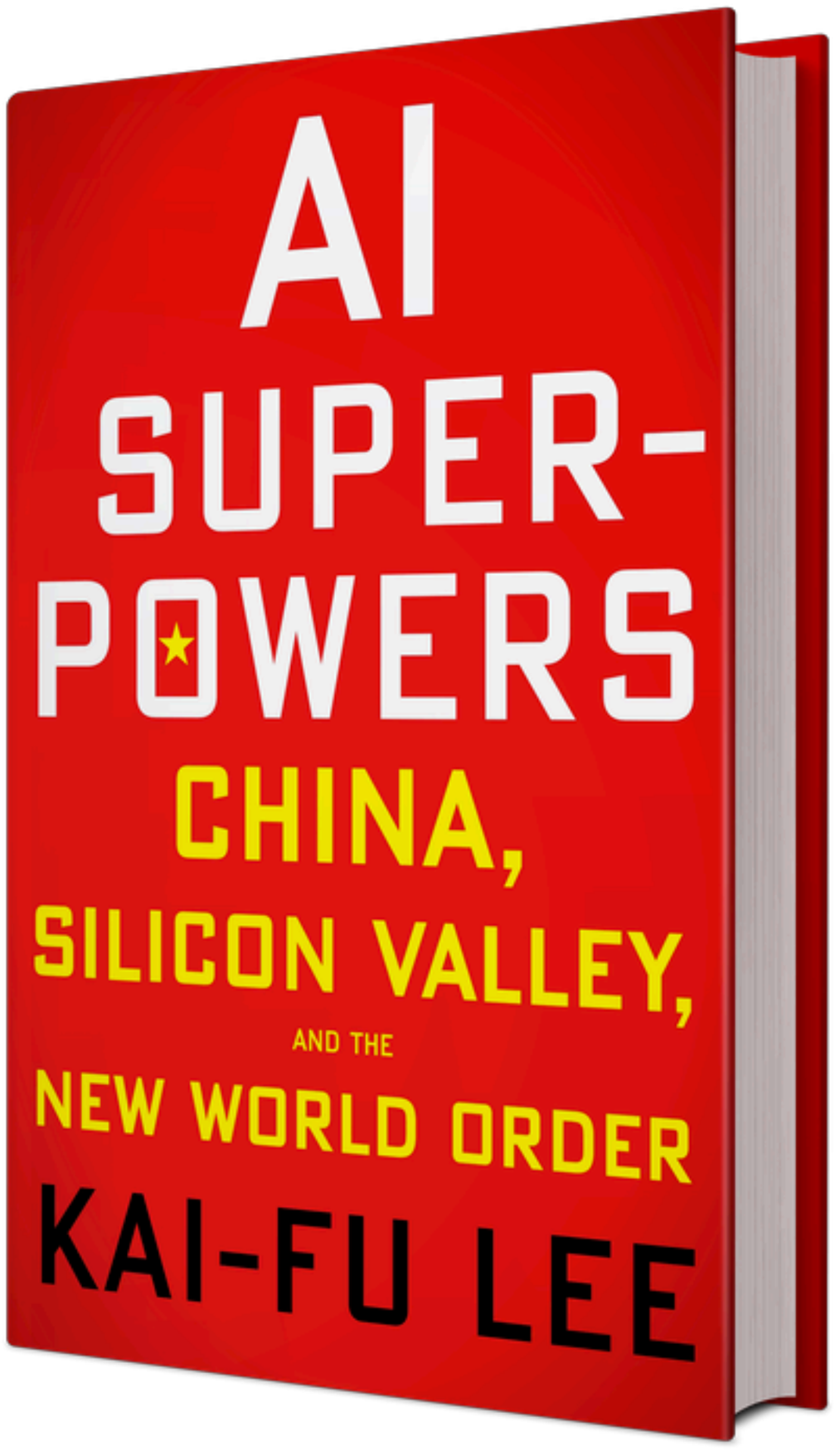
AI Fail

Human Fail

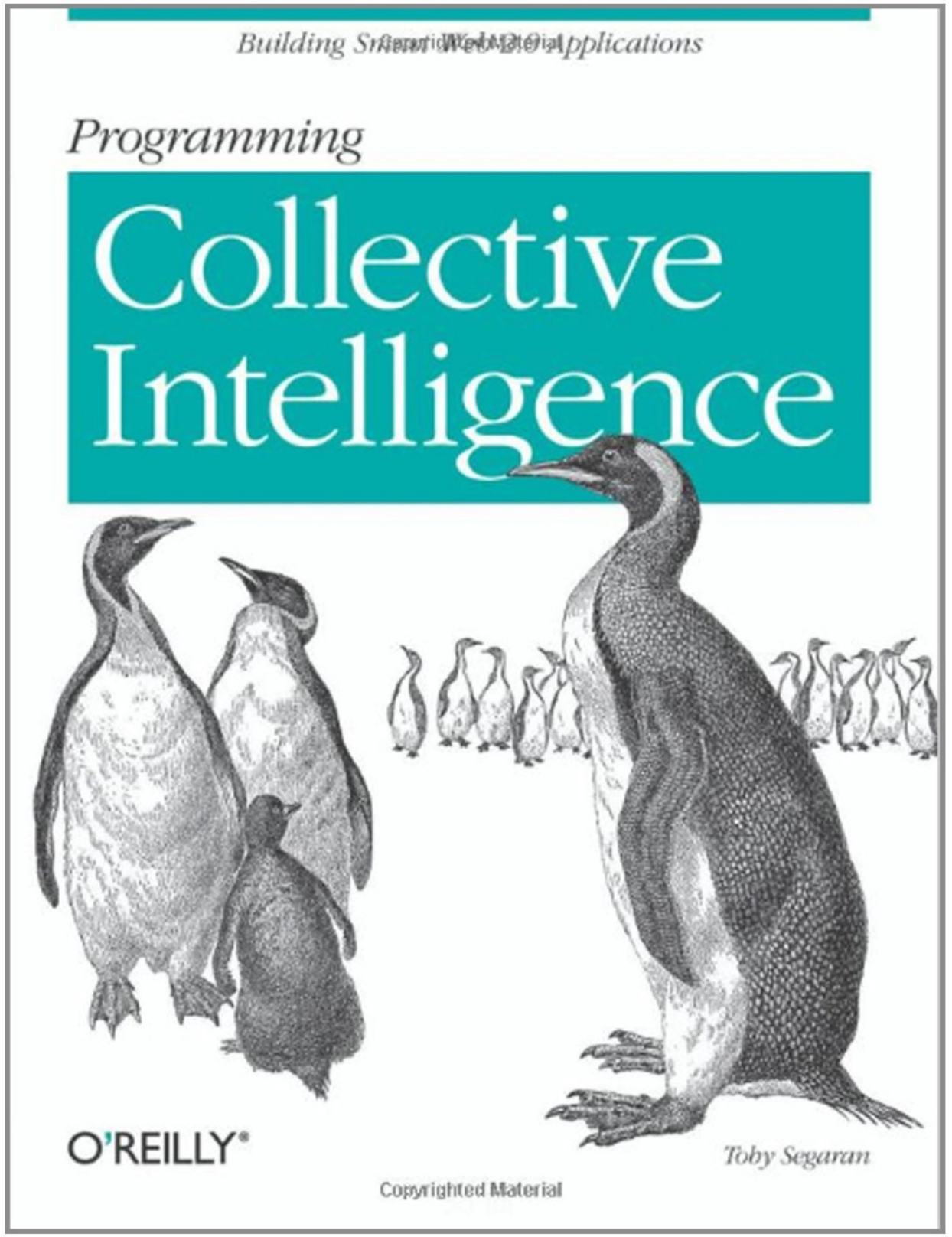
MORE APPLICATIONS

- Product Defect Detection Using Image Recognition
- Customer Interaction Bots
- Fraud, Anomaly and Bad Actor Detection
- Emotion Detection (Mental Health & Counseling)
- Self Driving Cars

Discover Magazine


AI
SUPER-
POWERS
CHINA,
SILICON VALLEY,
AND THE
NEW WORLD ORDER
KAI-FU LEE

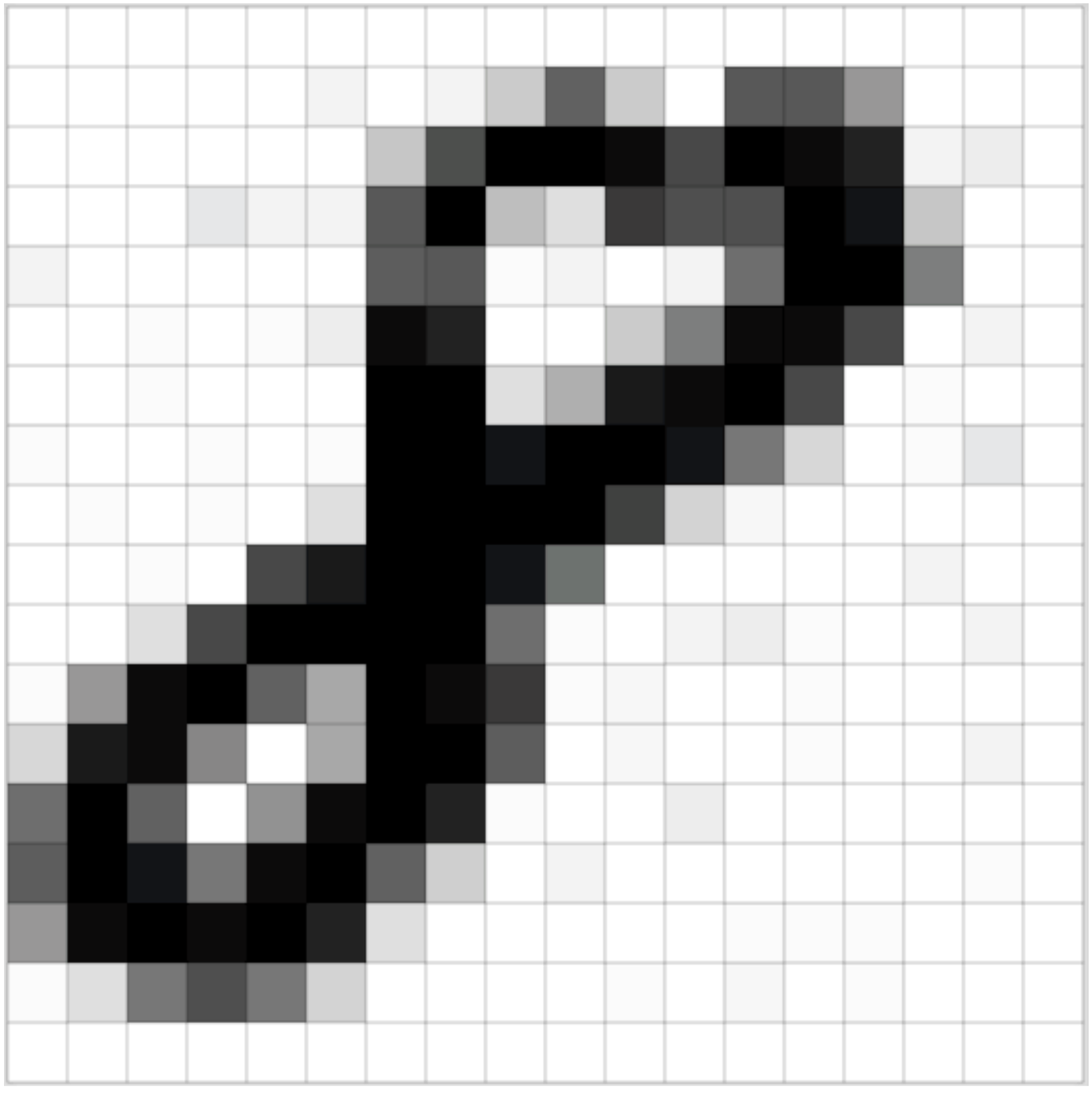
Literacy


Building Smart Web 2.0 Applications
Programming
Collective Intelligence
O'REILLY®
Toby Segaran
Copyrighted Material

Competency

HANDS-ON TIME

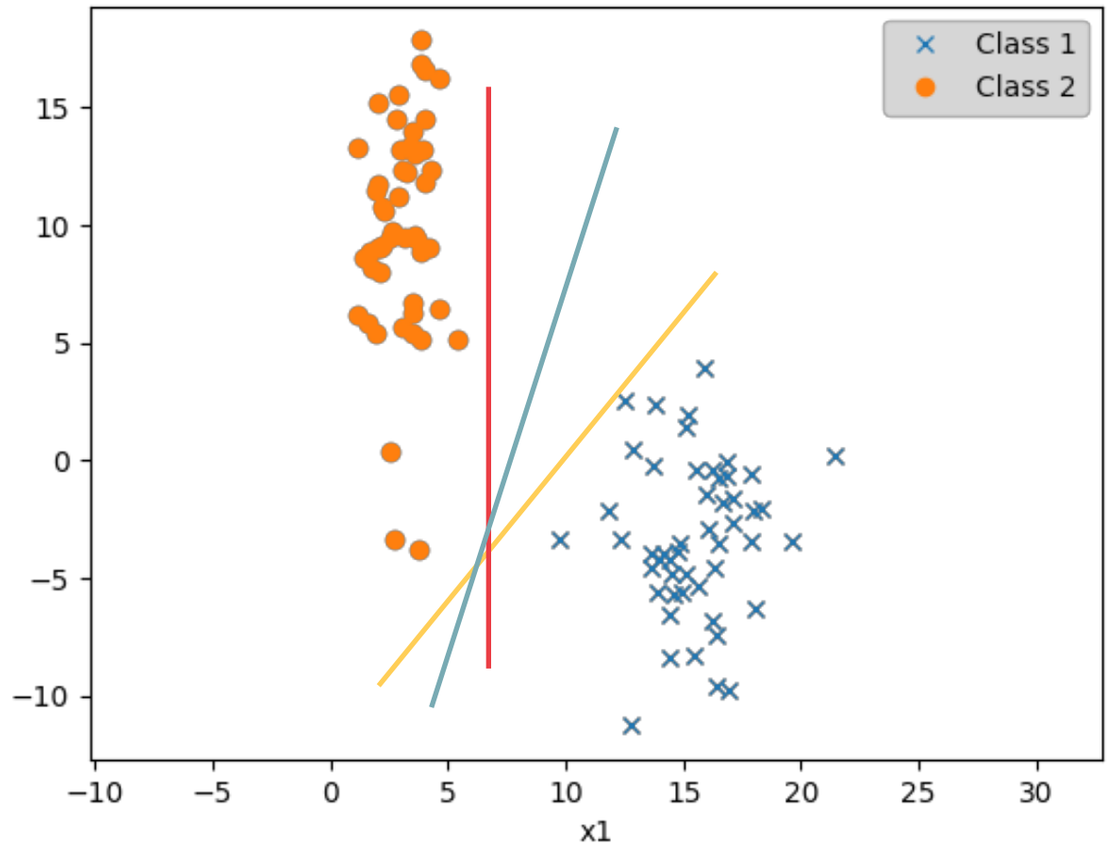
Black Box: Sci-kit Learn Example



Credit: Adam Geitgey

Support Vector Machines

Great for classification problems with numerical features and relatively few samples



Credit: Adedapo Alabi

Key Steps

Initialize: clf = svm.SVC(gamma=0.001, C=100)
Train: clf.fit(x, y)
Test: clf.score(x_test, y_test)
Predict: clf.predict(test)

Scikit Learn Example

```
import matplotlib.pyplot as plt
import numpy as np
from sklearn import datasets
from sklearn import svm

digits = datasets.load_digits()

clf = svm.SVC(gamma=0.001, C=100)
x, y = digits.data[:-10], digits.target[:-10]
clf.fit(x, y)

# test = digits.data[-6]
# test = np.array(test).reshape(1, -1)
# print('Prediction: ', clf.predict(test));

print(digits.images[-6])
plt.imshow(digits.images[-6], cmap=plt.cm.gray_r, interpolation="nearest")
plt.show()

```

DECEMBER 2019

Photo by: Douglas Adams

DECEMBER 2019

@shoreason

www.linkedin.com/in/shofola/

github.com/shoreason