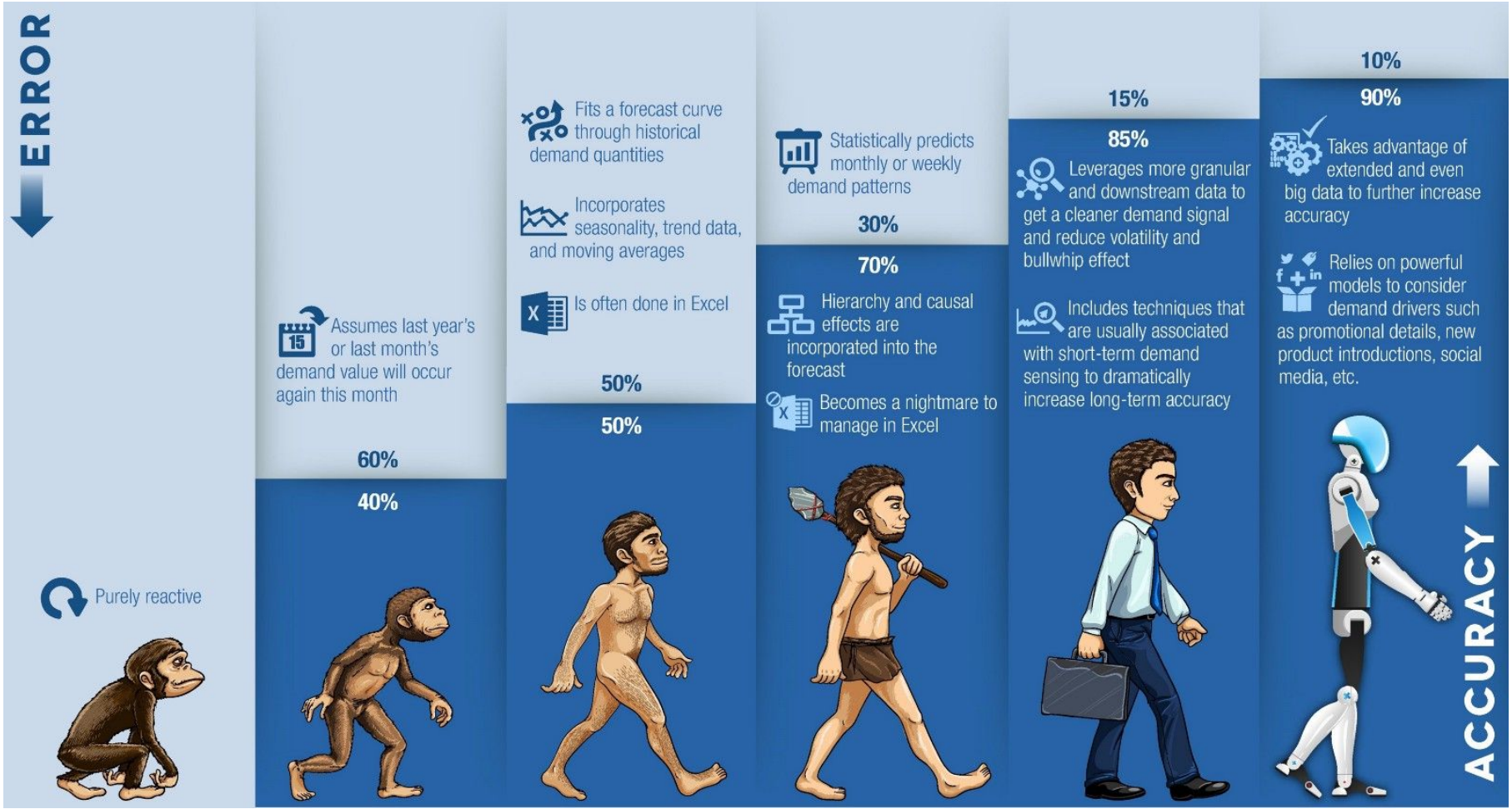
Deep Learning School

Lecture 3

Introduction To Machine Learning

Astakhov Anton

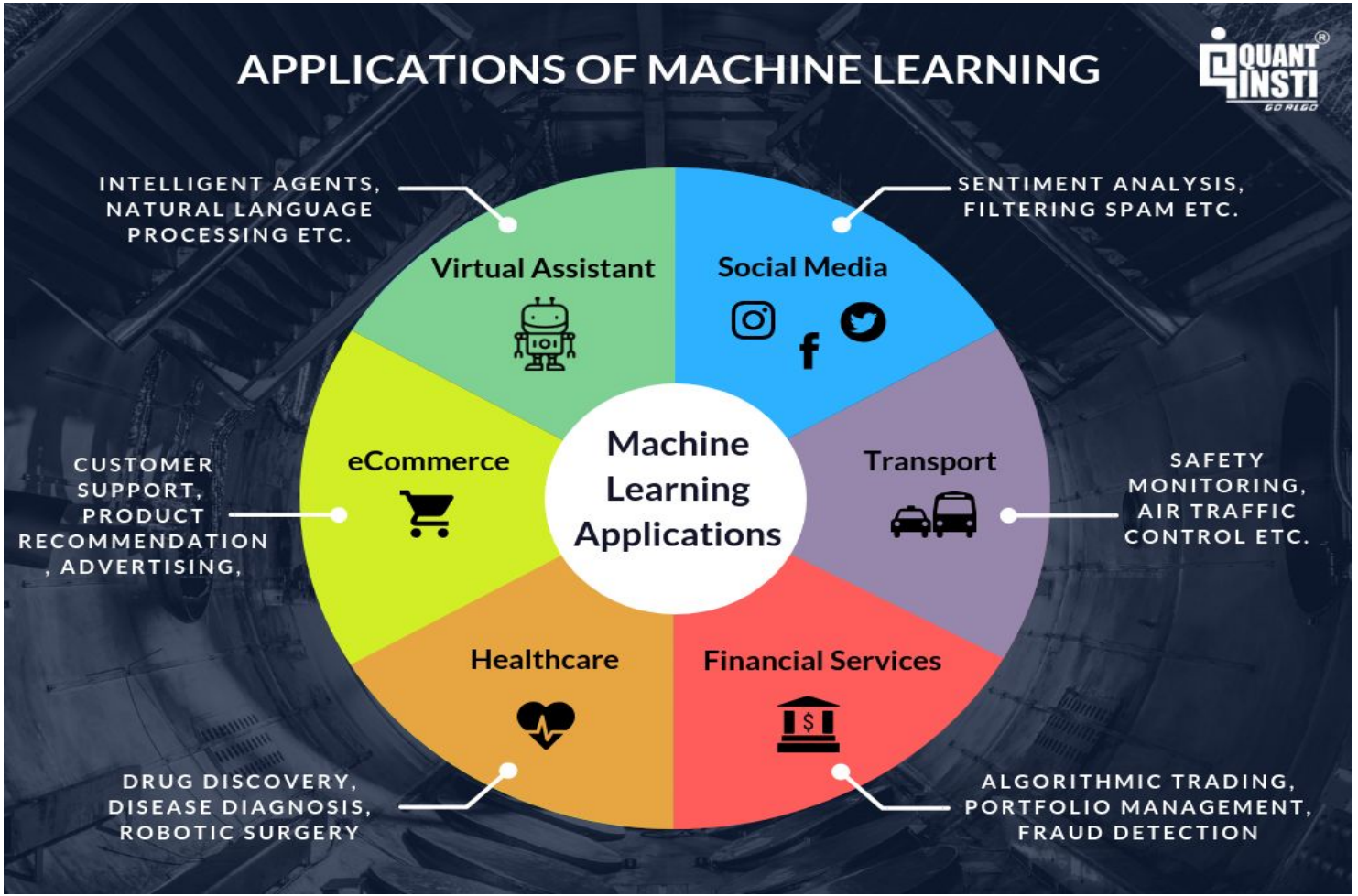


https://cdn-images-1.medium.com/max/1600/1*ZGf62foEavtwM0SVo5lYXg.jpeg

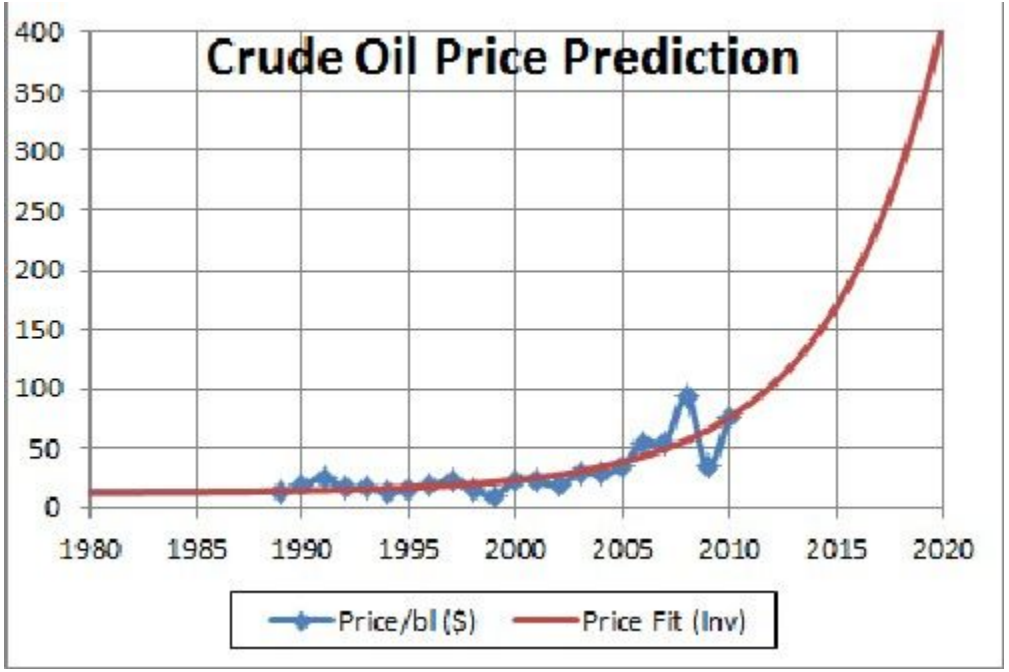
Lecture plan

1. Application
2. Process
3. Types
4. Short history
5. Tools

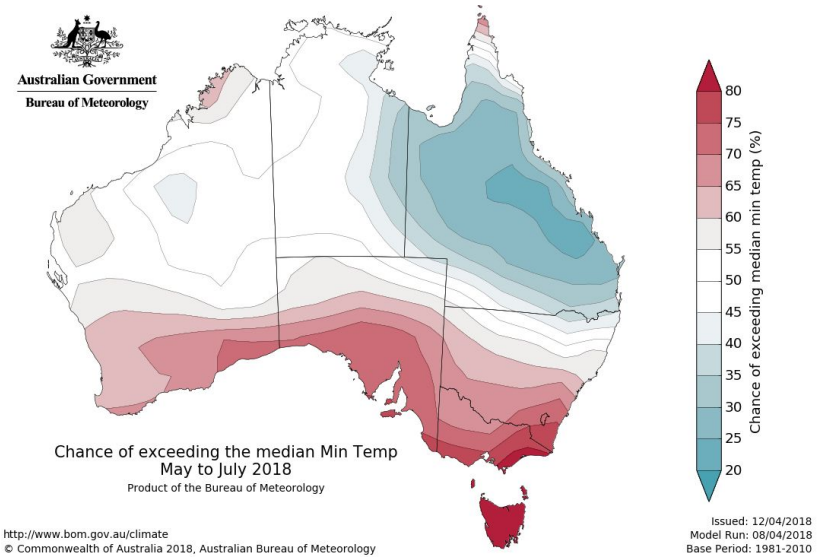
Application of Machine Learning



https://www.quantinsti.com/wp-content/uploads/2018/10/Applications-of-Machine-Learning-1024x768.png

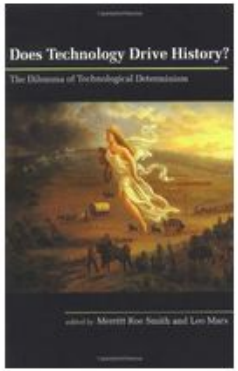

Crude Oil Price Prediction
400
350
300
250
200
150
100
50
0
1980
1985
1990
1995
2000
2005
2010
2015
2020
Price/bl ($)
Price Fit (Inv)

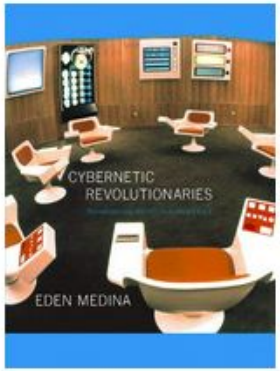

Good Credit Score

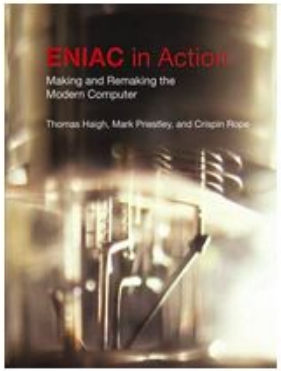

Australian Government
Bureau of Meteorology
Chance of exceeding the median Min Temp
May to July 2018
Product of the Bureau of Meteorology
http://www.bom.gov.au/climate
© Commonwealth of Australia 2018, Australian Bureau of Meteorology
80
75
70
65
60
55
50
45
40
35
30
25
20
Chance of exceeding median min temp (%)
Issued: 12/04/2018
Model Run: 08/04/2018
Base Period: 1981-2010

We Have Recommendations for You

Sign in to see personalized recommendations



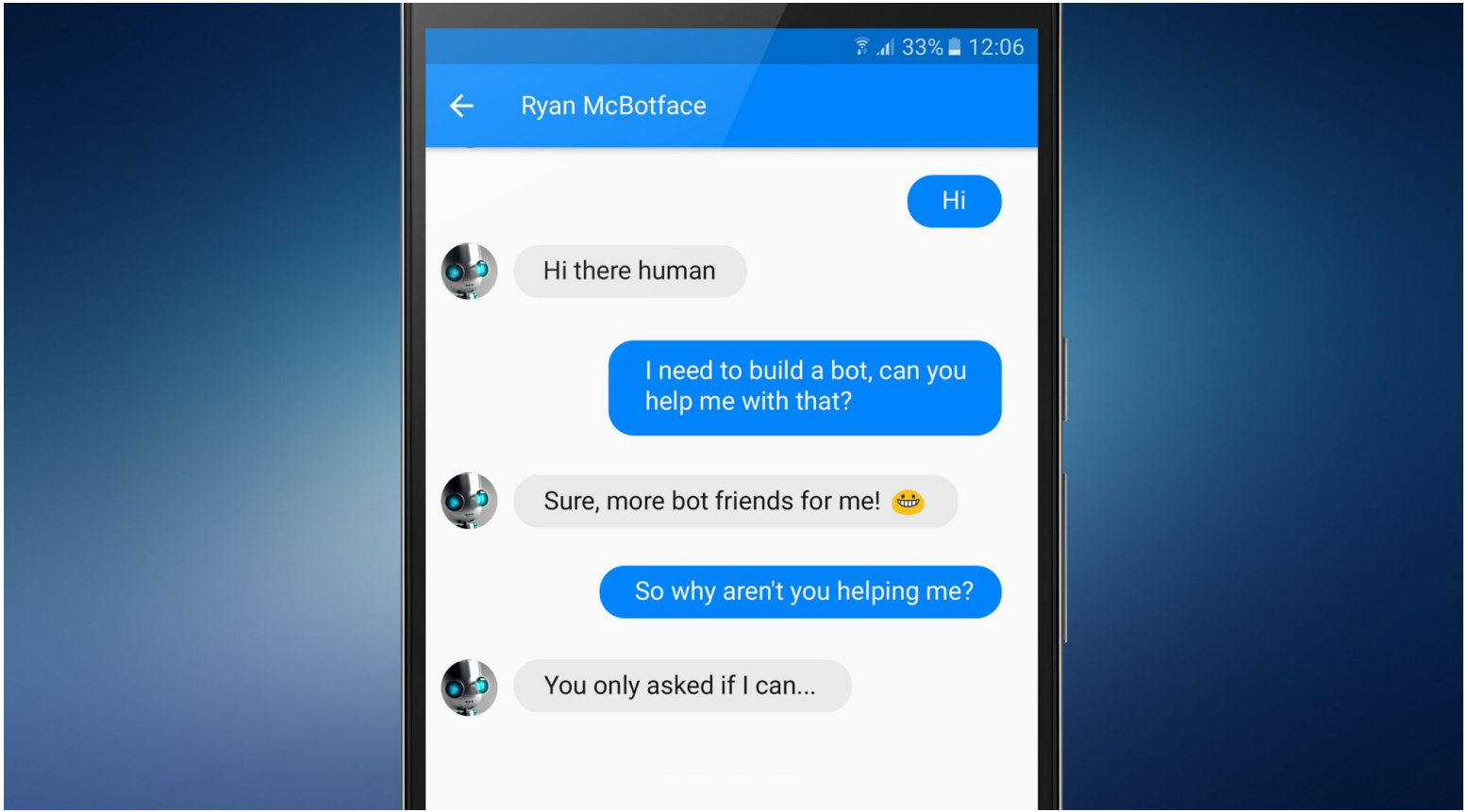




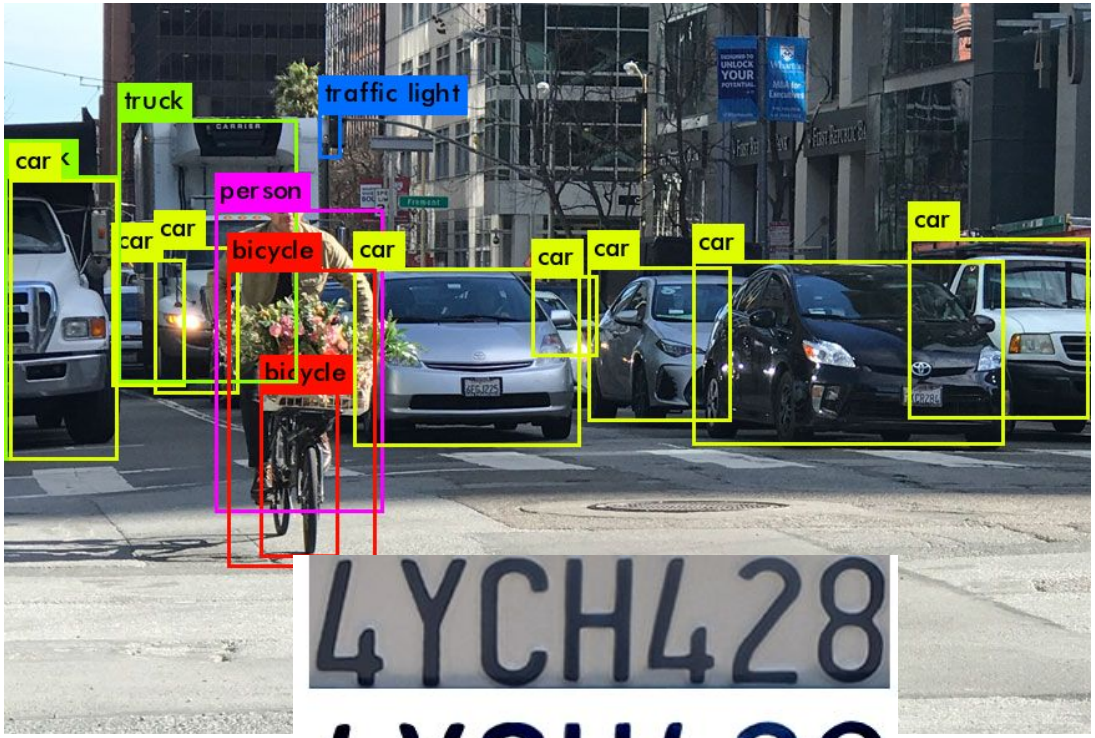
Google

rus

russian
russian doll
rust
russia today
rusprofile
russianpost
ruski
rusbase
rusline
russia study



https://cdn-images-1.medium.com/max/2000/1*yLsDfWKps2xWLV2aeZO40g.png


truck
traffic light
car
person
car
car
bicycle
car
car
car
car
bicycle
car


4YCH428
4YCH428
4YCH428

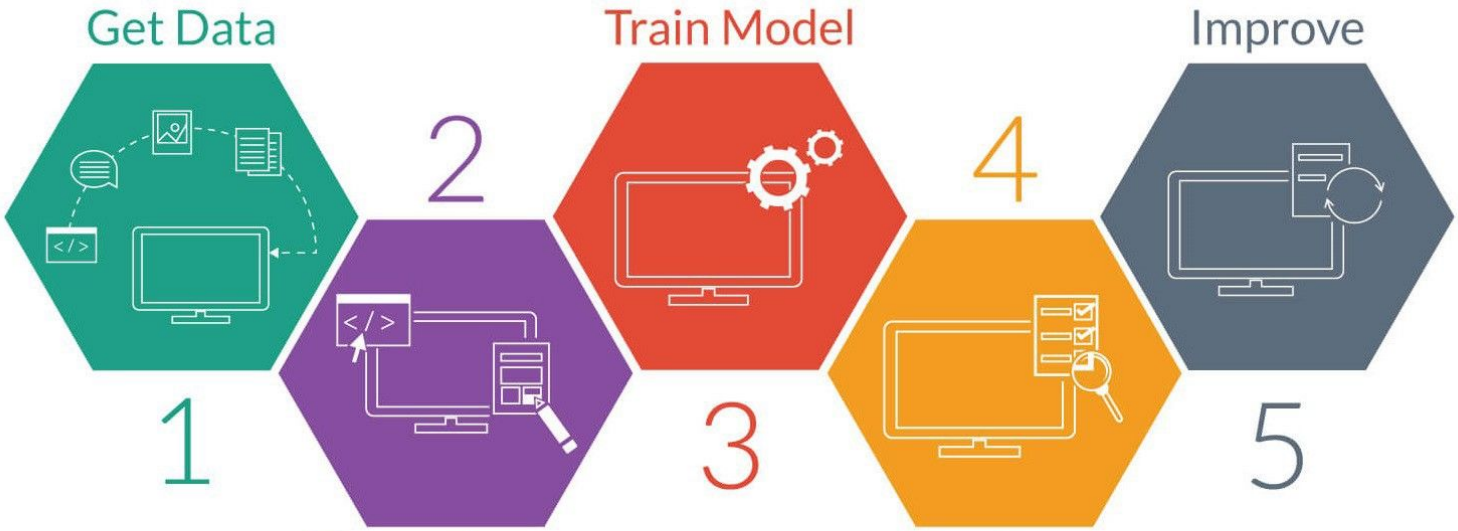

Fetch! FUN
Brittany?
Sporty, strong desire to please, sensitive & devoted
© Getty Images/Microsoft Fetch


35
42
30
34
39

Process of Machine Learning

- X - the set of the objects
- Y - the set of the answers to these objects
- $a : X \to Y$ - decison function
- $X^l = (x_i, y_i)_{i=1}^l$ - the data for training
- $X^k = (x_i, y_i)_{i=1}^k$ - the data for testing
- $x_i \in R^n$ where n - the number of features
- $y_i \in R$
- w - the parameters of the model
- $Q(a(w), X)$ - the loss function on objects X

$$X^l = (x_i, y_i)_{i=1}^{l}$$



$$a(w)$$

$$X^k = (x_i, y_i)_{i=1}^{k}$$

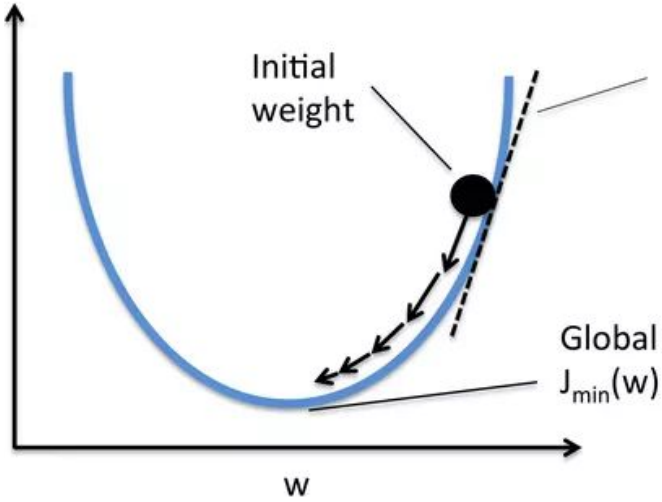
1. Gender
2. Family income
3. Work experience
4. Credit history
5. Age
6. The level of education
7. Profession
8. Place of residence
9. Property Owned

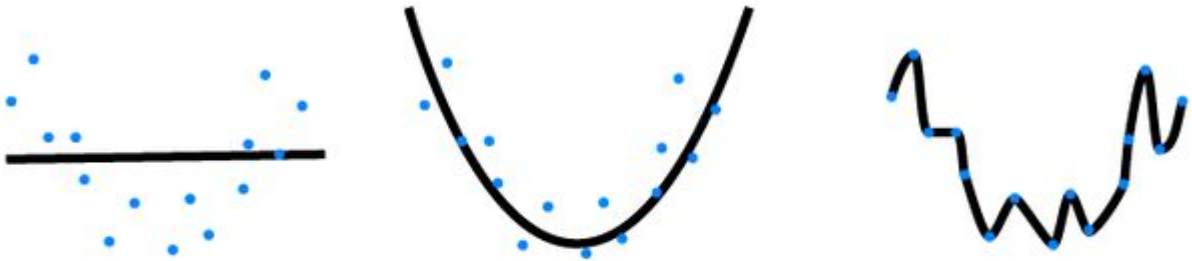
$Q(a(w), X^l) = \frac{1}{l} \sum_{i=1}^{l} L(a(w), x_i)$,
where $L(a(w), x_i)$ - the loss function on the object x_i.

$L(a(w), x_i) = [a(x_i) \neq y_i]$ - indicator(for classification)

$L(a(w), x_i) = (a(x_i) - y_i)^2$ - quadratic error(for regression)

$L(a(w), x_i) = |a(x_i) - y_i|$ - absolute error(for regression)

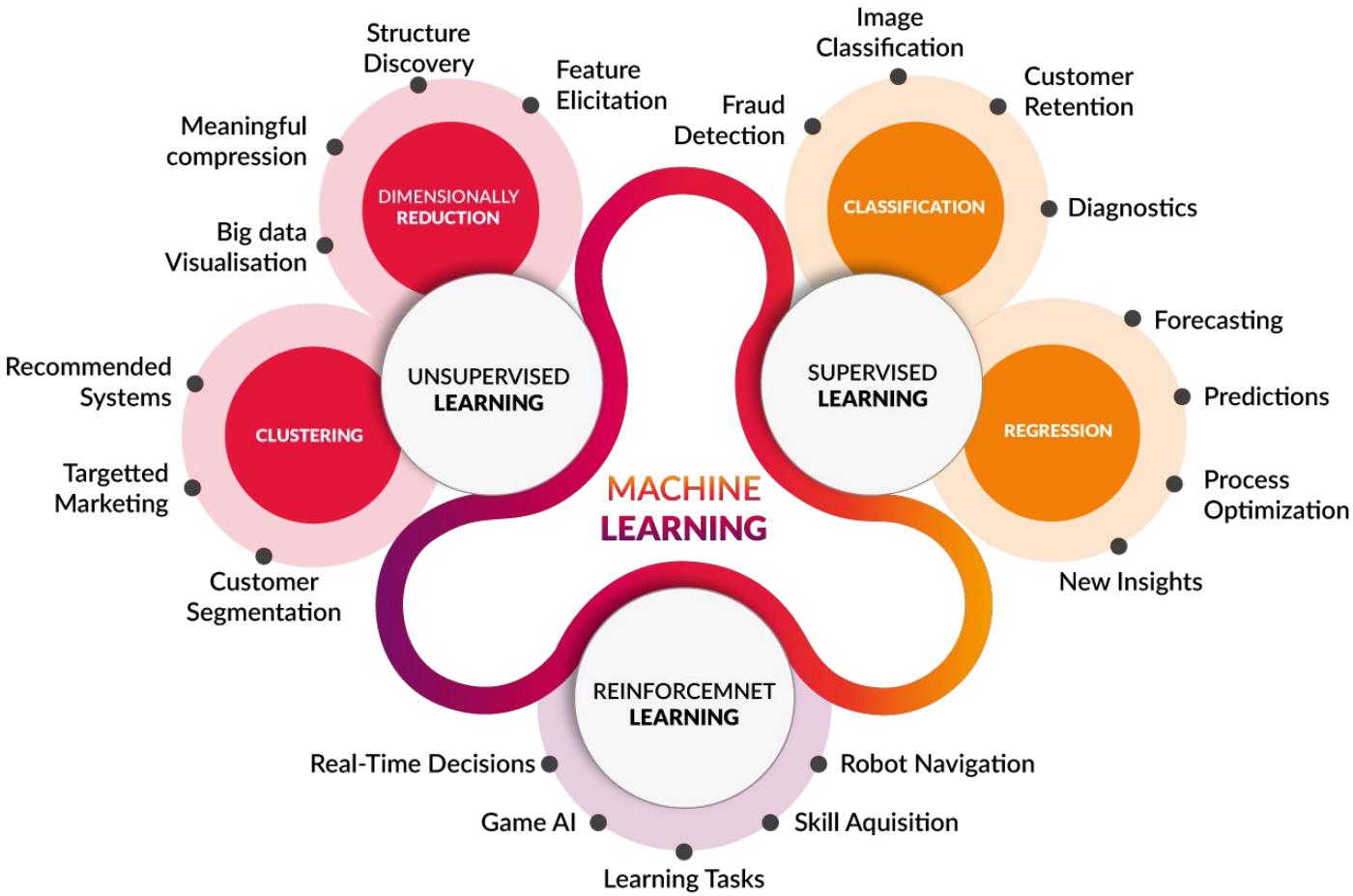




$$Q(a(w), X) = \frac{1}{l} \sum_x L(a(w), x)$$

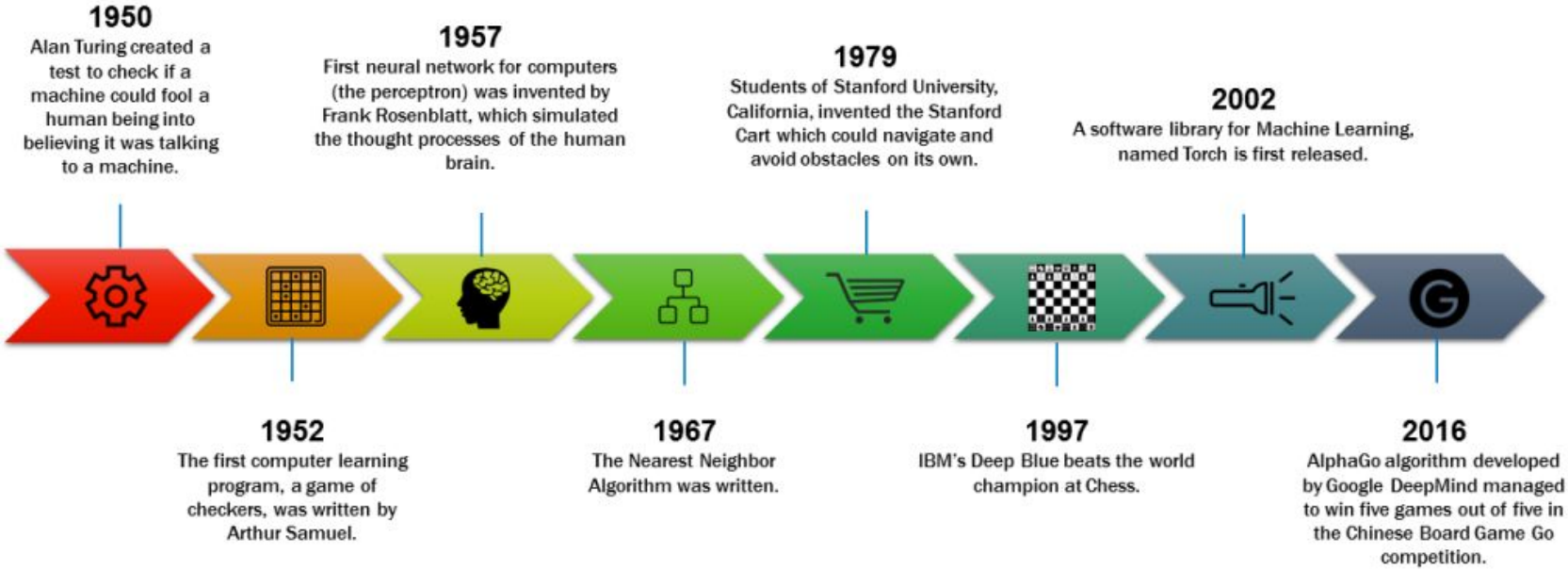
$$Q(a(w), X^k) \gg Q(a(w), X^l) \text{ - overfitting}$$

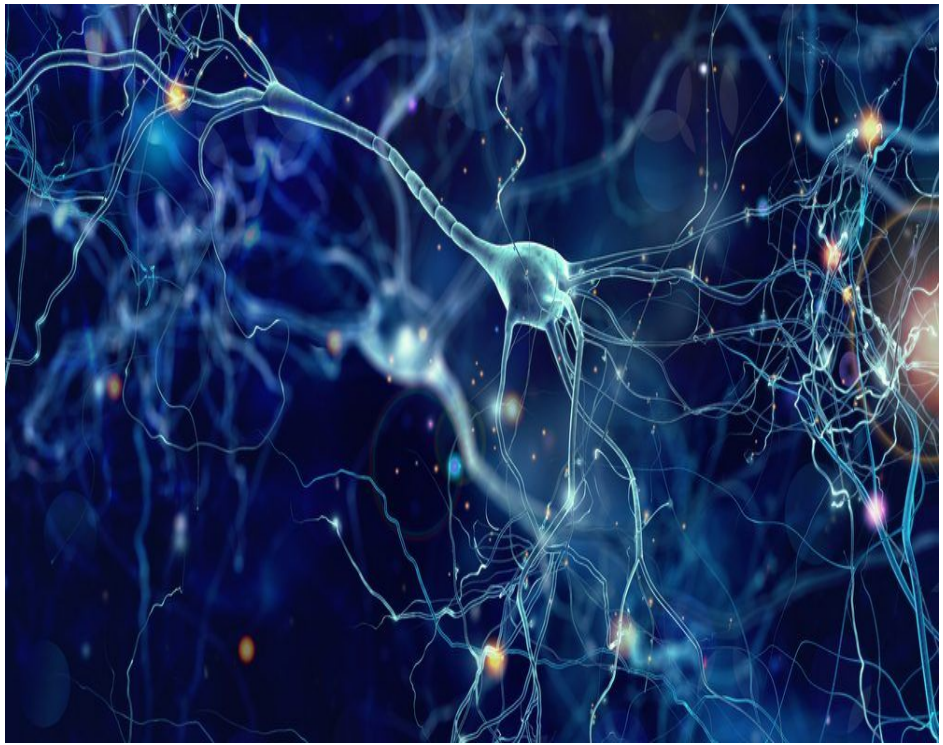
Types of Machine Learning

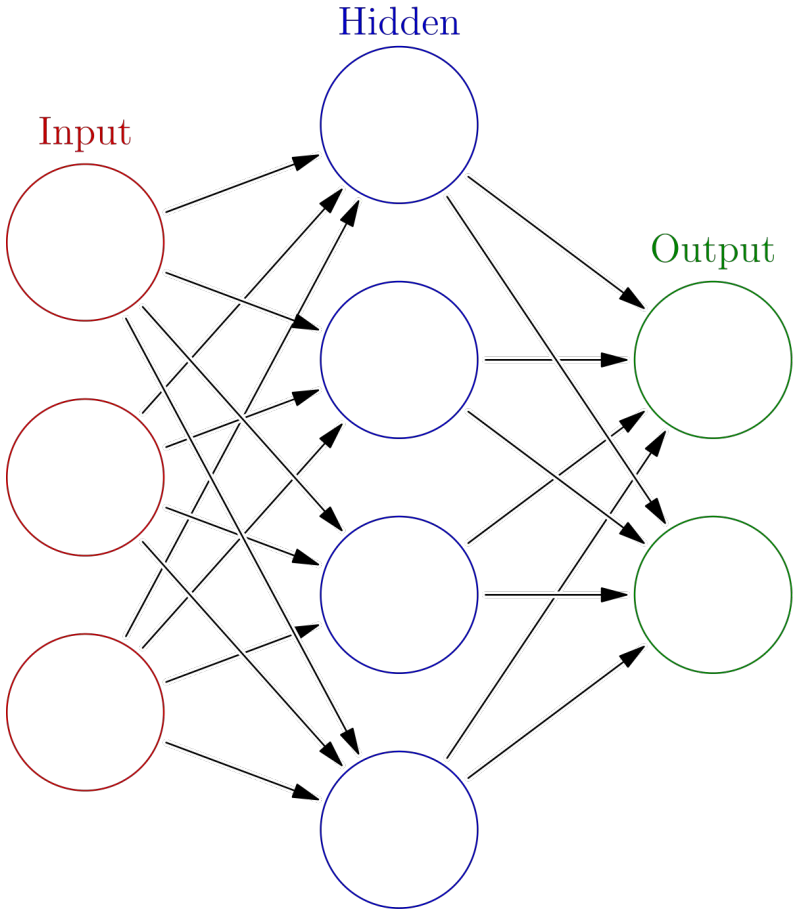


https://cdn-images-1.medium.com/max/2000/1*8wU0hfUY3UK_D8Y7tbIyFQ.png

Short history of Machine Learning


1950
Alan Turing created a test to check if a machine could fool a human being into believing it was talking to a machine.
1952
The first computer learning program, a game of checkers, was written by Arthur Samuel.
1957
First neural network for computers (the perceptron) was invented by Frank Rosenblatt, which simulated the thought processes of the human brain.
1967
The Nearest Neighbor Algorithm was written.
1979
Students of Stanford University, California, invented the Stanford Cart which could navigate and avoid obstacles on its own.
1997
IBM's Deep Blue beats the world champion at Chess.
2002
A software library for Machine Learning, named Torch is first released.
2016
AlphaGo algorithm developed by Google DeepMind managed to win five games out of five in the Chinese Board Game Go competition.



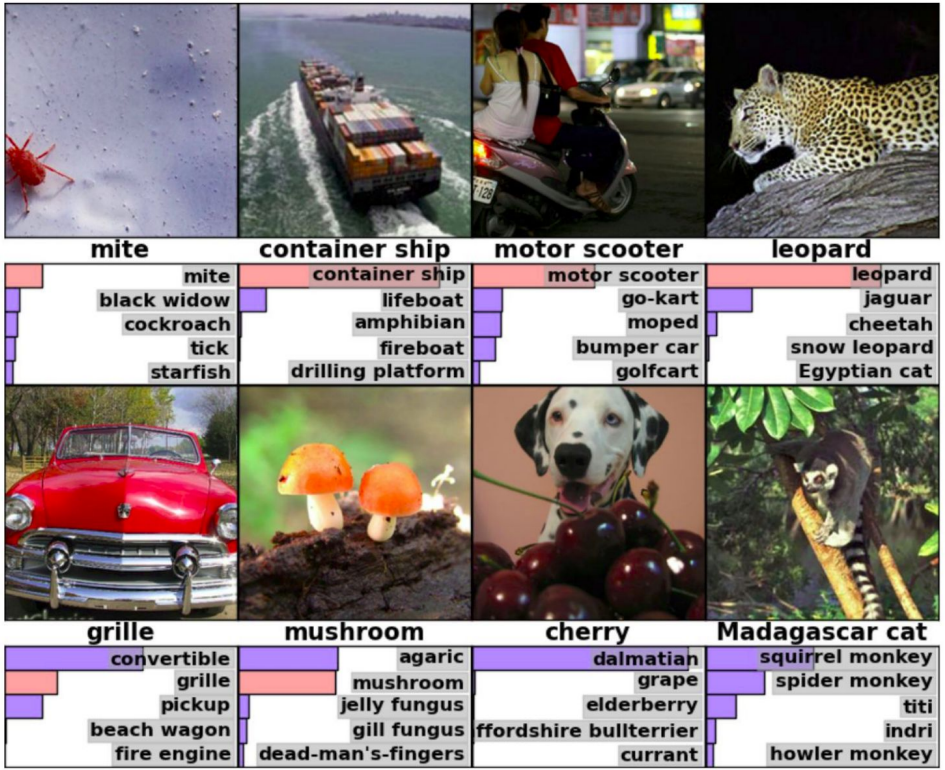


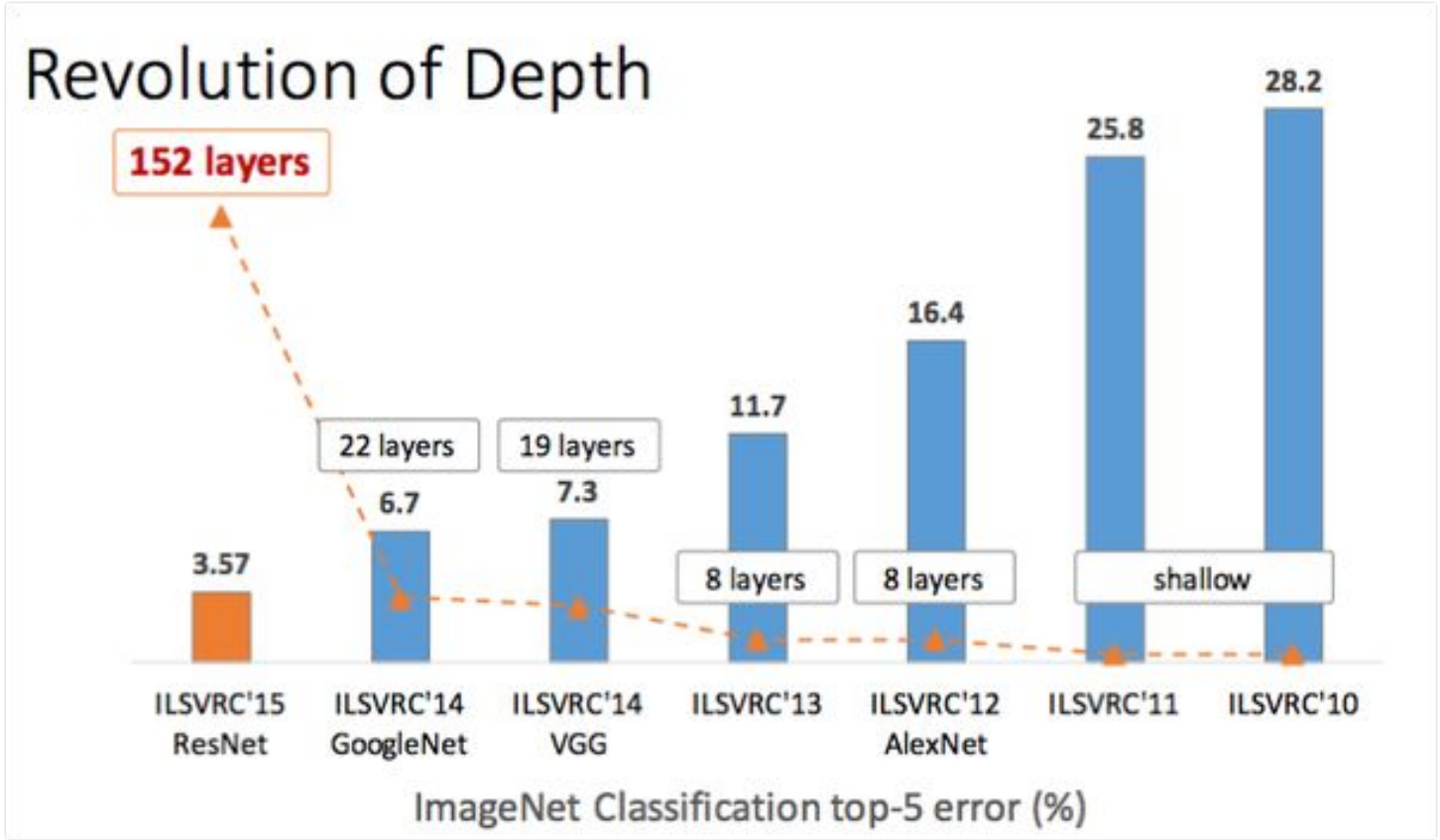
https://alsnewstoday.com/wp-content/uploads/2017/04/shutterstock_576832348-1000x480.jpeg

ImageNet Challenge

IMAGENET

- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.



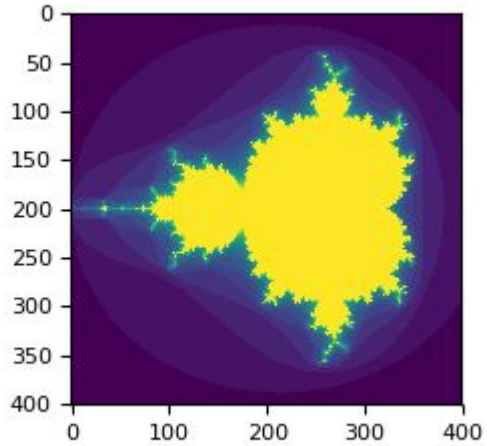

Revolution of Depth
152 layers
22 layers
19 layers
8 layers
8 layers
shallow
3.57
6.7
7.3
11.7
16.4
25.8
28.2
ILSVRC'15 ResNet
ILSVRC'14 GoogleNet
ILSVRC'14 VGG
ILSVRC'13
ILSVRC'12 AlexNet
ILSVRC'11
ILSVRC'10
ImageNet Classification top-5 error (%)

Tools

Jupyter notebook

NumPy

```
>>> import numpy as np
>>> import matplotlib.pyplot as plt
>>> def mandelbrot( h,w, maxit=20 ):
...     """Returns an image of the Mandelbrot fractal of size (h,w)."""
...     y,x = np.ogrid[ -1.4:1.4:h*1j, -2:0.8:w*1j ]
...     c = x+y*1j
...     z = c
...     divtime = maxit + np.zeros(z.shape, dtype=int)
...
...     for i in range(maxit):
...         z = z**2 + c
...         diverge = z*np.conj(z) > 2**2            # who is diverging
...         div_now = diverge & (divtime==maxit)  # who is diverging now
...         divtime[div_now] = i                  # note when
...         z[diverge] = 2                        # avoid diverging too much
...
...     return divtime
>>> plt.imshow(mandelbrot(400,400))
>>> plt.show()
```



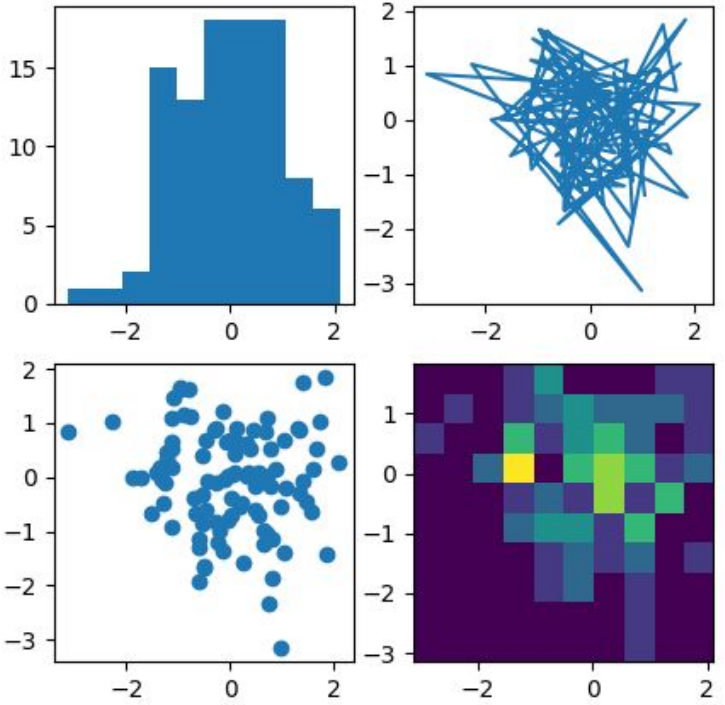
Matplotlib

```
import matplotlib.pyplot as plt
import numpy as np

np.random.seed(19680801)
data = np.random.randn(2, 100)

fig, axs = plt.subplots(2, 2, figsize=(5, 5))
axs[0, 0].hist(data[0])
axs[1, 0].scatter(data[0], data[1])
axs[0, 1].plot(data[0], data[1])
axs[1, 1].hist2d(data[0], data[1])

plt.show()
```

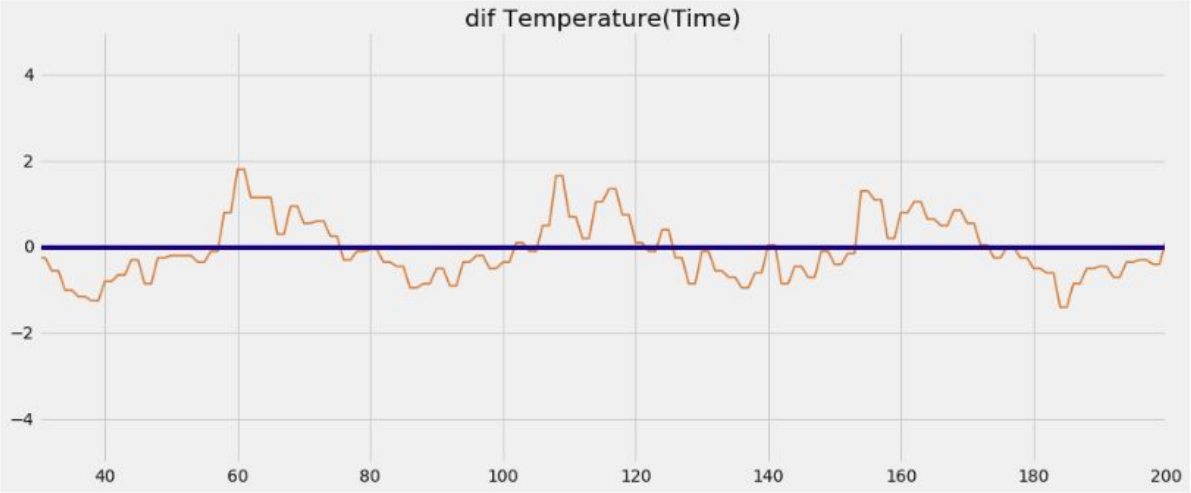


Pandas

```
Test_Data.head(1)
```

	Id	Consumption	Temperature	Time	DailySeasonality	WeeklySeasonality
5183	5183	4317.386273	14.75	5183	47	191

```
plt.figure(figsize=(15, 6))
plt.plot(Learn_Data_Temp.Time, Learn_Data_Temp.Temperature,color='peru',linewidth=1.7)
c=Learn_Data_Temp.Time
plt.xlim(30,200)
plt.ylim(-5,5)
plt.plot(c,-9.88869197*10**(-7)*c-1.46170425*10**(-3), color='navy')
plt.title("dif Temperature(Time)")
plt.show()
```



PYTORCH

Neural Machine Translation in Linear Time

Nal Kalchbrenner **Lasse Espeholt** **Karen Simonyan** **Aäron van den Oord** **Alex Graves** **Koray Kavukcuoglu**
Google Deepmind, London UK
nalk@google.com

Abstract

We present a novel neural network for processing sequences. The ByteNet is a one-dimensional convolutional neural network that is composed of two parts, one to encode the source sequence and the other to decode the target sequence. The two network parts are connected by stacking the de-