Machine Learning - What, Where and How?

Kurian Benoy
kurian.benoy@aot-technologies.com

MACHINE LEARNING

WHAT, WHERE AND HOW?

Online knowledge sharing

Session By: Kurian Benoy

Venue: Kapixi

Date & Time: 29th July, Thu @4pm

About Me

- SE - Data Scientist
- Kaggle 2x Expert
- FOSSASIA Open TechNights Winner
- Open source Enthusiast

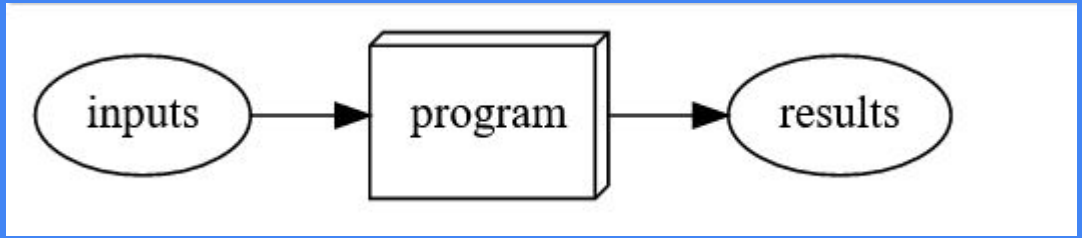
formsflow.ai

നാട്ടിൽ അവിടെയാ?

Kolenchery(Ernakulam)

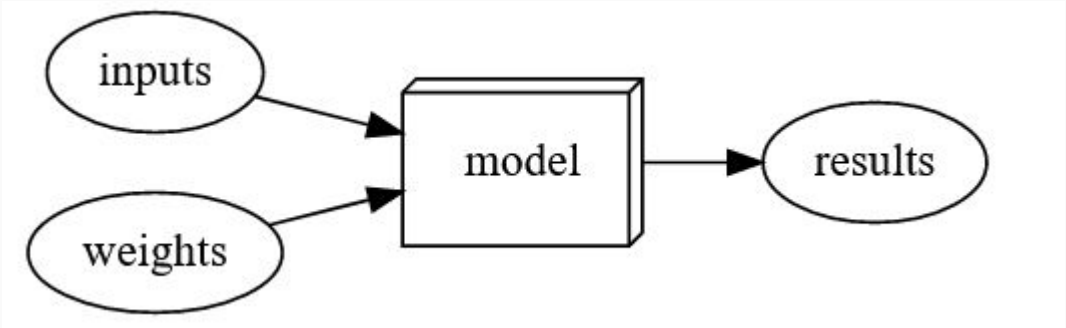
https://kurianbenoy.com/

What is Machine Learning?




```
def fib(x):
    """ calculate fibonaci series """
    if x == 0:
        return 0
    elif (x == 1) or (x == 2):
        return 1
    else:
        return fib(x-1) + fib(x-2)
```

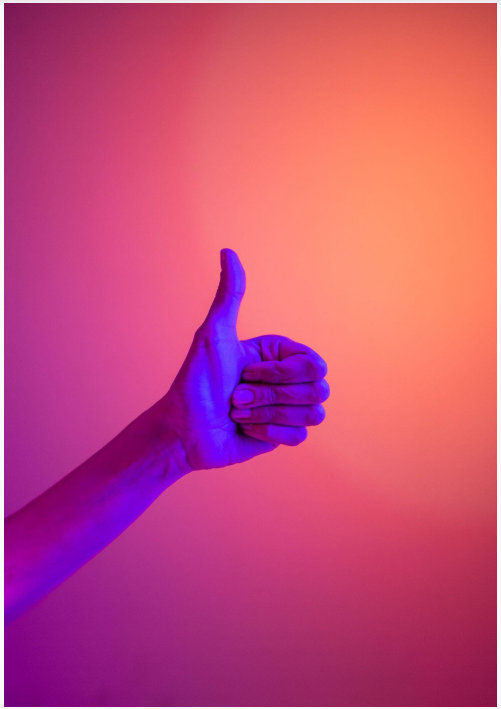
Learning



Arthur Samuel(1959)

Machine Learning - The field of study that gives computers the ability to learn without being explicitly programmed.

What is ML already good at?



cottonbro

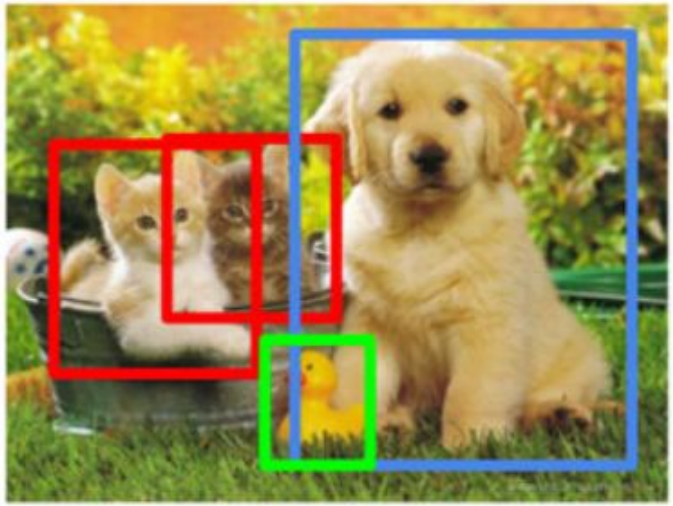
Computer Vision

- Object Detection
- Image Classification
- Multi-label Image Classification
- GANs

Classification

CAT

Object Detection



CAT, DOG, DUCK

https://www.datacamp.com/community/tutorials/object-detection-guide

Natural Language Processing(NLP)

- Translation
- Text Analytics
- Summarising
- Search Engines

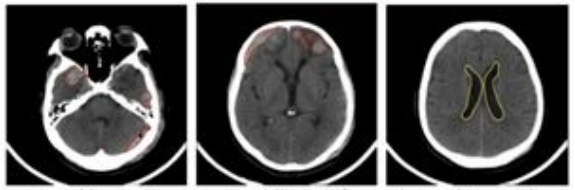
Combine Text & Images

- Input images with output captions

qQuant Analysis Report

Patient ID 346pd2ids
Accession Number 20202020.1
Age 49 years
Study Date 14 Oct 19

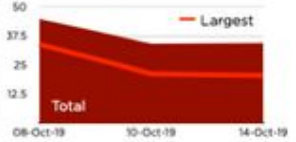
Type of Scan Non-contrast CT head
Analysis Automated quantification and prgression measurement of intracranial hyperdensities, lateral ventricles and midline shift



Intracranial hyperdensity | Midline shift | Lateral Ventricles

Intracranial hyperdensity and midline shift

Date	Hyperdensity		Midline shift	
	Total	Largest	Shift	Side
08 Oct 19	44.96	34.16	6mm	Left
10 Oct 19	34.37	21.42	4mm	Left
14 Oct 19	34.69	20.75	3mm	Left



Lateral Ventricles

Date	L Ventricle		R Ventricle		Total
	Volume	%	Volume	%	Volume
08 Oct 19	12.46	39.7	18.94	60.3	31.4
10 Oct 19	15.64	48.3	16.76	51.7	32.4
14 Oct 19	18.00	56.1	13.93	43.9	31.93

Study Time 08 Oct 19, 15:03:47, 10 Oct 19, 18:32:29, 14 Oct 19, 06:42:16,
Analysis Time 14 Oct 19, 11:02:09
qQuant Version 1.0.0

qure.ai

Tabular Data

- Time series data (forecasting sales data)
- Churn Prediction

https://arxiv.org/pdf/2106.03253.pdf

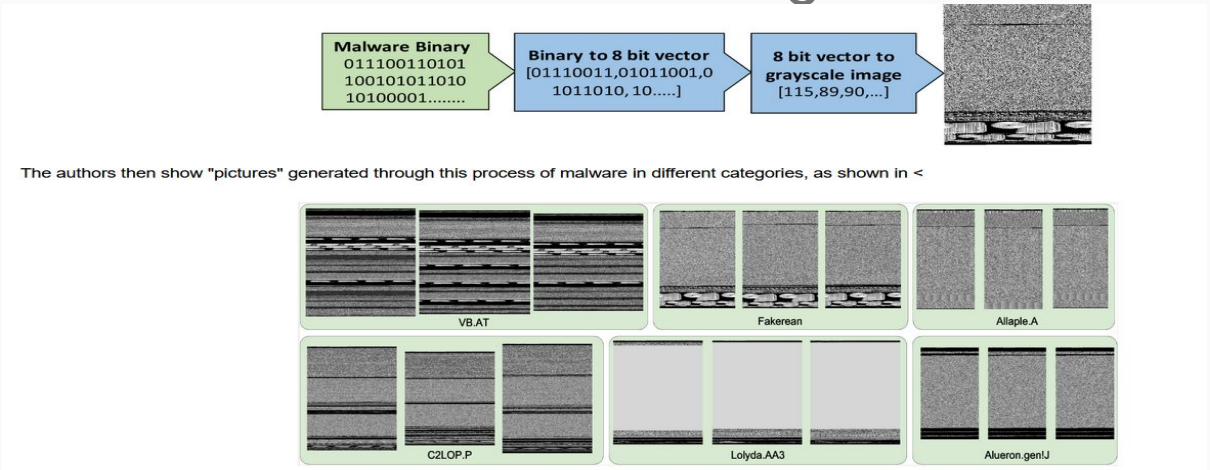
Recommendation Systems

YouTube

Youtube logo, Netflix

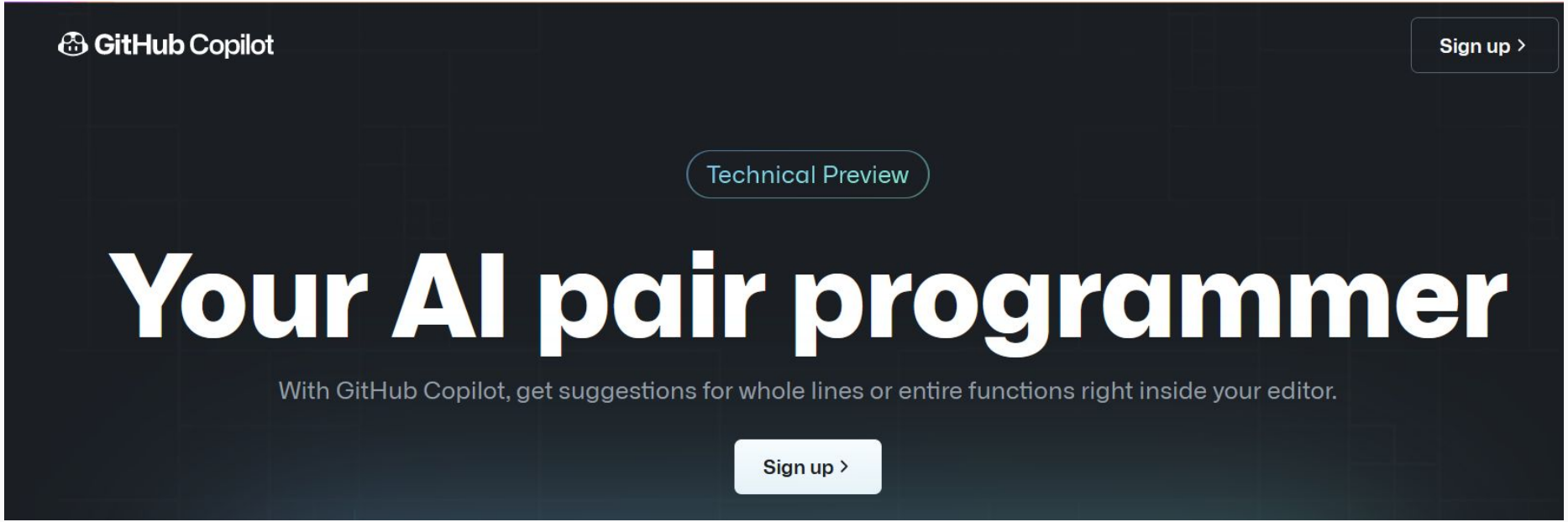
Other Data Types

- Domain specific data like protein chains
- Identifying Audio Data
- Malware identification algorithm



Where is it used?

Github Copilot



Kurian Benoy Retweeted

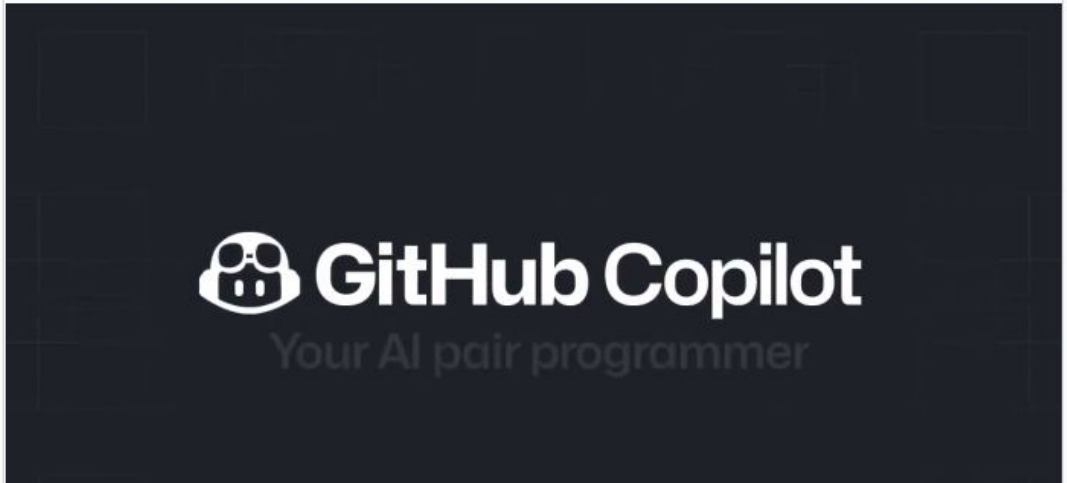
Hamel Husain @HamelHusain · 29 Jun

This is perhaps the most significant ML product GitHub has released in its history.

Meta: only ~1 FT ML person was involved from GitHub, rest was design SWE etc. Modeling work done primarily by OpenAI.

Don't underestimate transfer learning and hosted model hubs/apis

> **GitHub** @github · 29 Jun
>
> Meet GitHub Copilot - your AI pair programmer. copilot.github.com



https://twitter.com/HamelHusain

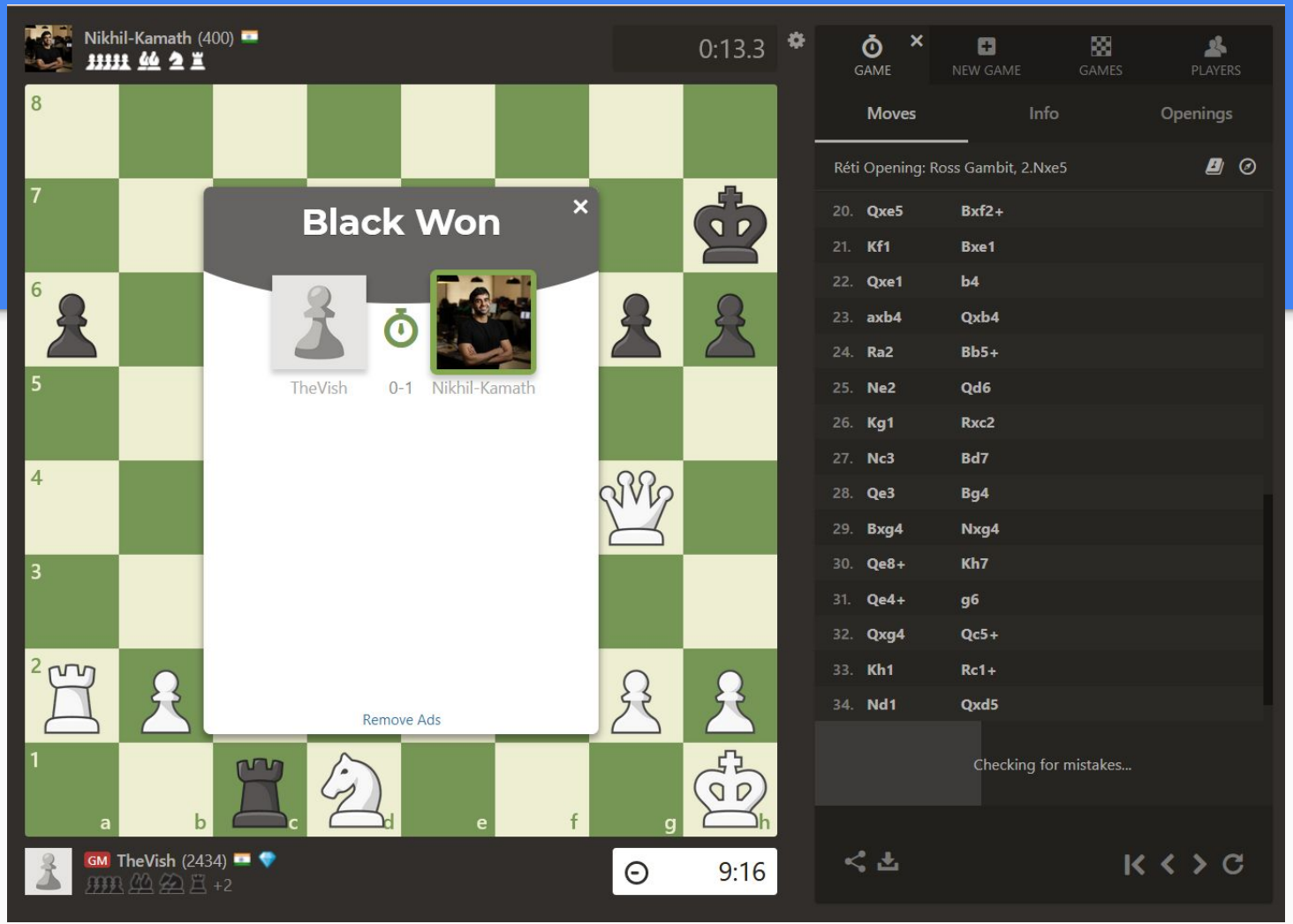
People Counter

fauxels

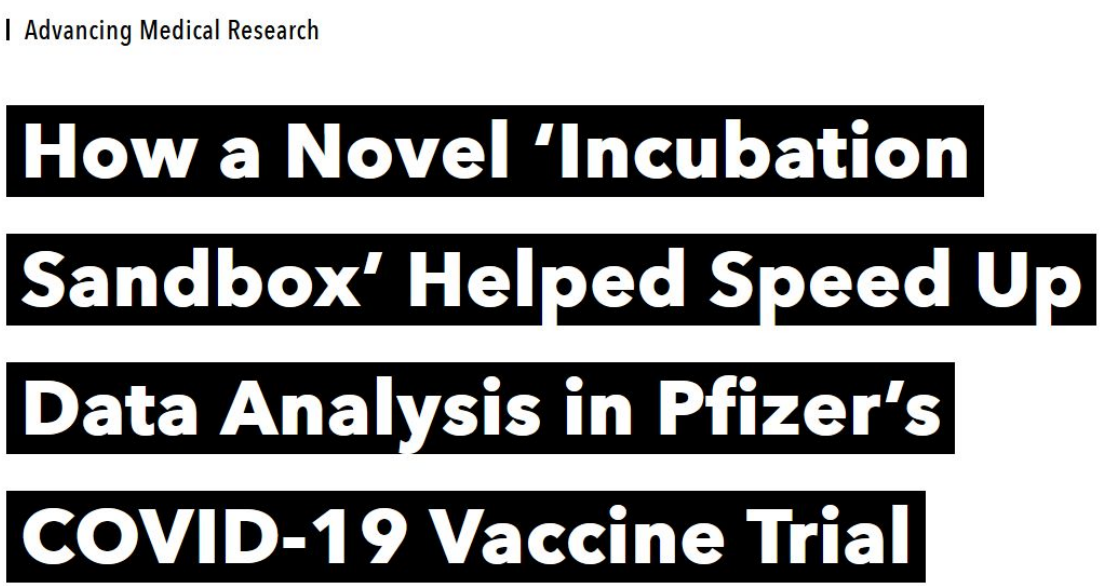
Alpha Go

Youtube, DeepMind

Youtube, DeepMind

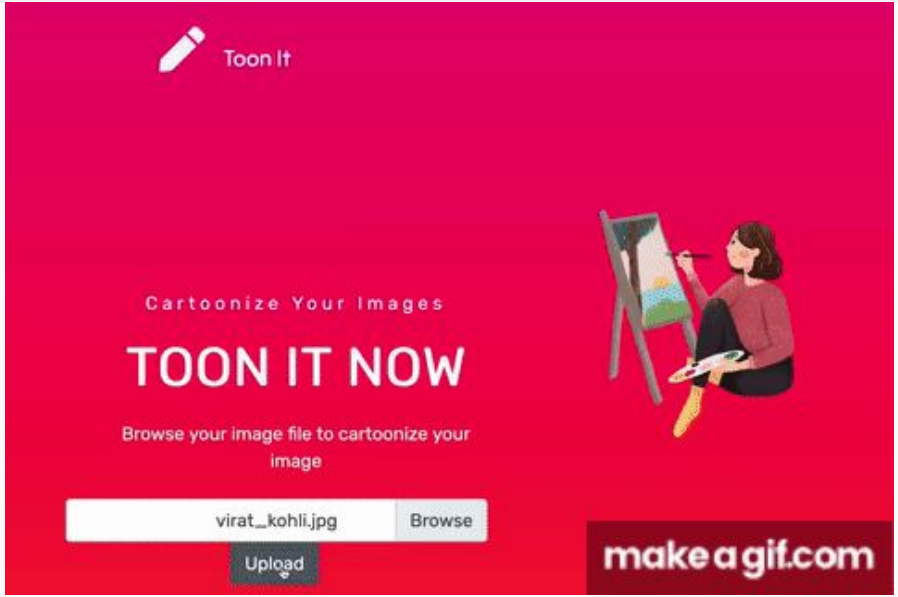

Nikhil-Kamath (400)
0:13.3
GAME
NEW GAME
GAMES
PLAYERS
Moves
Info
Openings
Réti Opening: Ross Gambit, 2.Nxe5
20. Qxe5 Bxf2+
21. Kf1 Bxe1
22. Qxe1 b4
23. axb4 Qxb4
24. Ra2 Bb5+
25. Ne2 Qd6
26. Kg1 Rxc2
27. Nc3 Bd7
28. Qe3 Bg4
29. Bxg4 Nxg4
30. Qe8+ Kh7
31. Qe4+ g6
32. Qxg4 Qc5+
33. Kh1 Rc1+
34. Nd1 Qxd5
Checking for mistakes...
Black Won
TheVish
0-1
Nikhil-Kamath
Remove Ads
GM TheVish (2434) +2
9:16

Covid-19 Response



This Person Does Not Exist (thispersondoesnotexist.com/)

Cartoonizer



https://github.com/Toon-It/Cartoonizer/blob/master/demo.md

Deoldify (https://deoldify.ai/)

How to use ML?

The Drivetrain Approach

Many accurate models are of no use to anyone

Many inaccurate models are highly useful to many

Designing great data products - https://www.oreilly.com/radar/drivetrain-approach-data-products/


What outcome am I trying to achieve?
Defined Objective
Levers
what inputs can we control
Data
what data we can collect
Models
how the levers influence the objective

Case Study: Recommendation Systems

Objective:

Levers:

Data Collection:

Building Models:

Important Things to ponder upon

Based on:

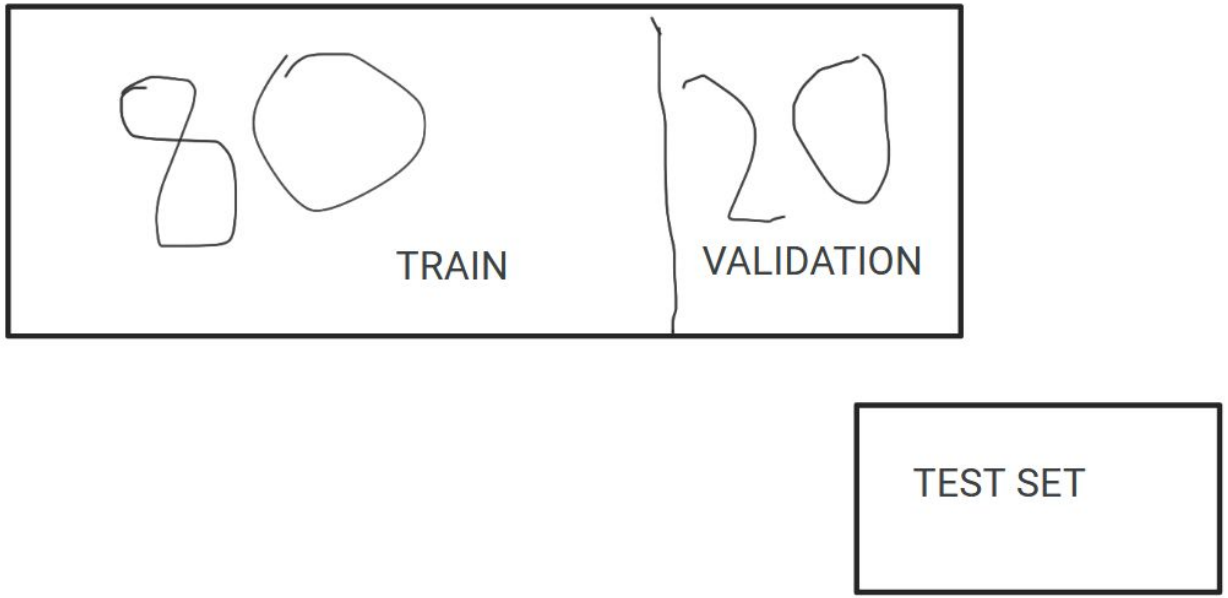
[1] Rules for Machine Learning (https://developers.google.com/machine-learning/guides/rules-of-ml/)
[2] Deep Learning for Coders, Jeremy Howards, Sylvain Gugger (https://github.com/fastai/fastbook)

Don’t be afraid to launch a product without ML

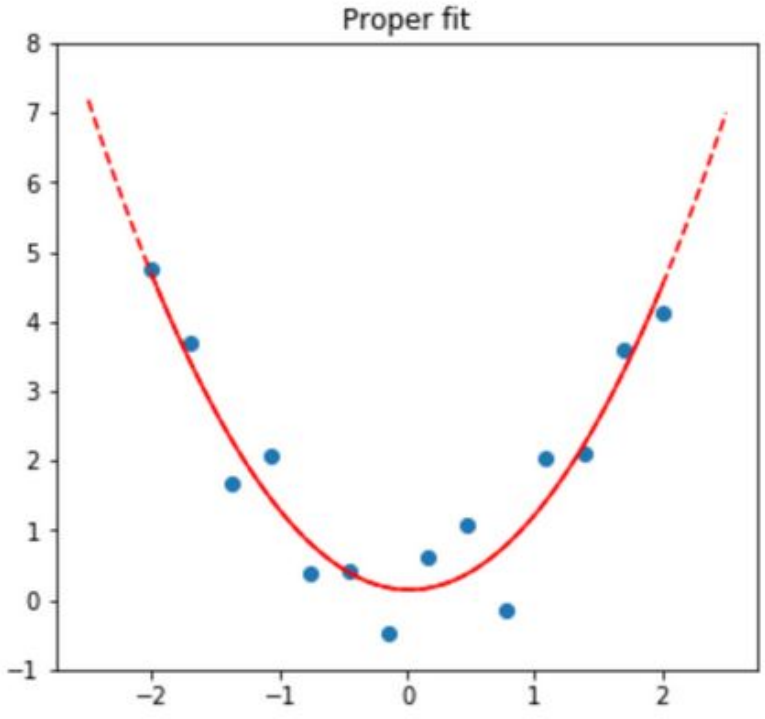
Data

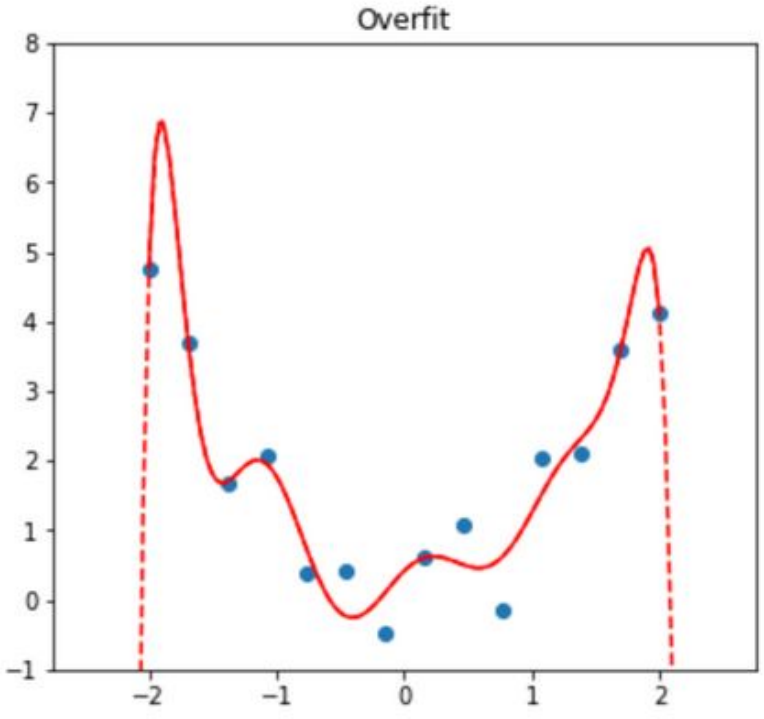
- Do I need a lot of Data?
- Can it be be applied with transfer learning?
- Labelled Data

Dataset splitting



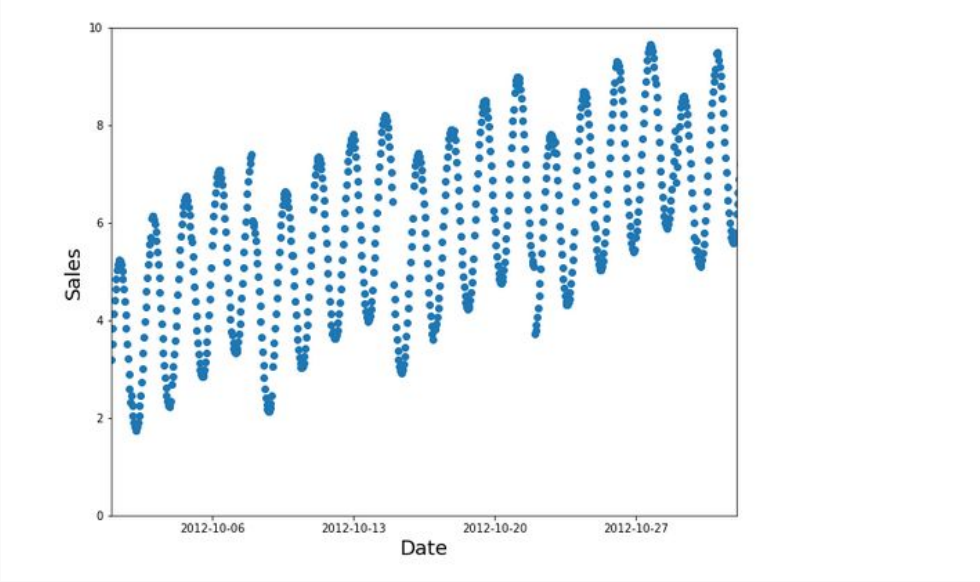
Always be bestfitting





Training Validation Split

- Building generalisable machine learning models



Use Baseline models

- Simple models like logistic regression, CNN with 2-3 layers
- Don’t spend a lot of time on fine tuning with big models

Read: A Recipe for Training Neural Networks

[1] https://karpathy.github.io/2019/04/25/recipe/

Build your first pipeline first correctly

Always getting the pipeline of productionizing based on ML is the hard thing.

Don’t care about accuracy too much

AI-ML CoE External
7 members

Chat Files Tasks

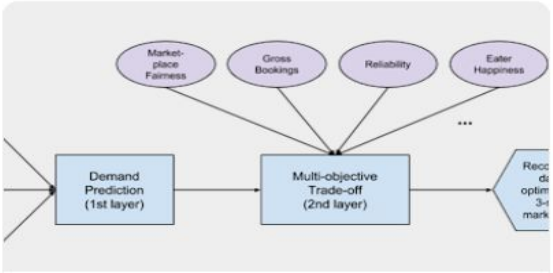
Kurian Benoy Jun 23, 12:54 PM

How Uber Eats recommends foods based on marketplace

https://eng.uber.com/uber-eats-recommending-marketplace/

Another interesting article on uber eats blogs:

https://eng.uber.com/uber-eats-query-understanding/



Food Discovery with Uber Eats: Recommending fc
eng.uber.com

👍 1

Rajeev Chandrasekharan Jun 23, 3:18 PM

Good one

https://towardsdatascience.com/7-underrated-channels-to-follow-on-youtube-251a3aedde37

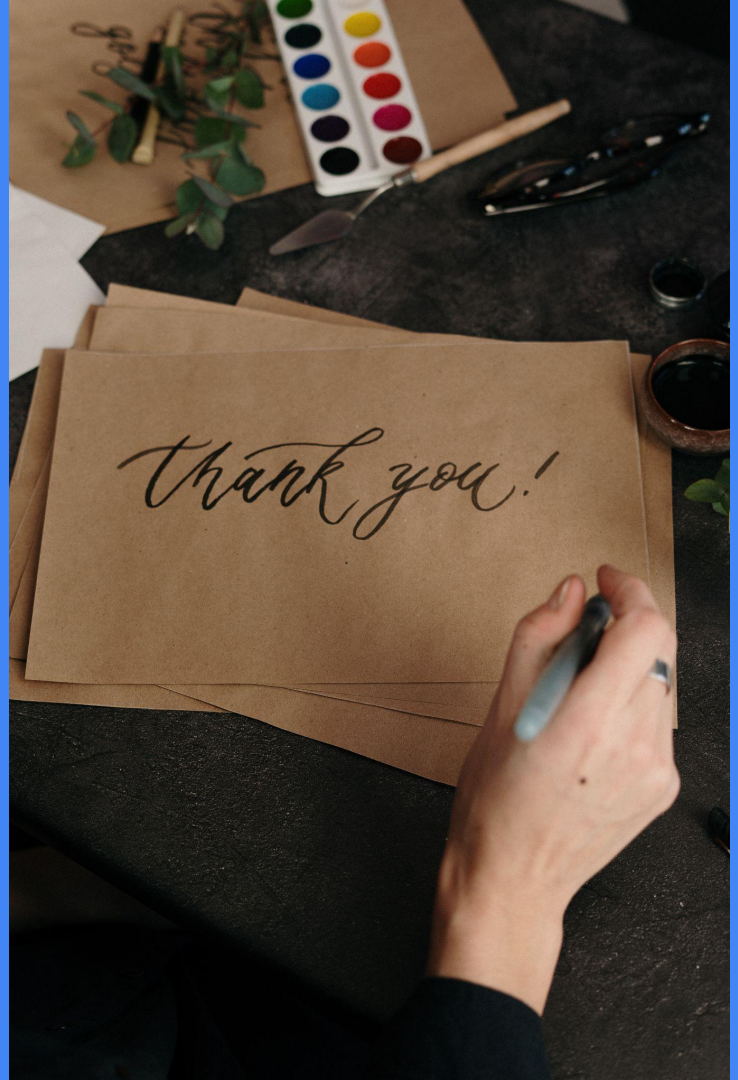
Jump to bottom

30 Days of ML

Machine learning beginner → Kaggle competitor in 30 days. Non-coders welcome.

Starts August 2nd!

Sign Up for the Challenge


thank you!

Appendix

Supervised learning - labelled data

Unsupervised learning - unlabelled data

(fraud detection)

Reinforcement learning - based on continuous feedback

Learning Machine Learning

https://github.com/eclipse/deeplearning4j

https://www.tensorflow.org/js

- Best course IMO: Deep learning for Coders

Difference b/w AI / ML / Deep Learning