Mani Sarkar
github: **neomatrix369**
twitter: **@theNeomatrix369**
blogs: **https://medium.com/@neomatrix369**

From backend development to machine learning

—

my journey from Software Engineering to ML/Data Science

3rd July 2020

hashtag: #abhishektalks

About me

Mani Sarkar

More about me

Freelance Software, Data, ML Engineer

Java / JVM

Cloud / Infra / DevOps

Polyglot developer

Code quality, testing, performance, DevOps, deep affinity for AI/ML/DL, NN...

LJC, Devoxx, developer communities

Strengthening teams and helping them *accelerate*

JCP member, F/OSS projects: @adoptopenjdk @graalvm @truffleruby

Java Champion, Oracle Groundbreaker Ambassador, Software Crafter, Blogger, Speaker

@theNeomatrix369

I'm a big fan of...

It's important to have <u>heros</u> and <u>mentors</u>!

<u>Nikola Tesla</u> and his fine work

<u>Richard Feynman</u>, his books and quotes!

<u>Sir Arthur Conan Doyle</u> and his work (I have all his 65+ series on DVD)

Agenda

About the talk

Hello, hello!

- *Timeline: Backend dev to ML Engineer*
- *Demos*
- *References to other talks*
- *Resources*
- *Summary and Closing*
- *Q&A*
- ***Appendix** section: more good stuff for later*

@theNeomatrix369
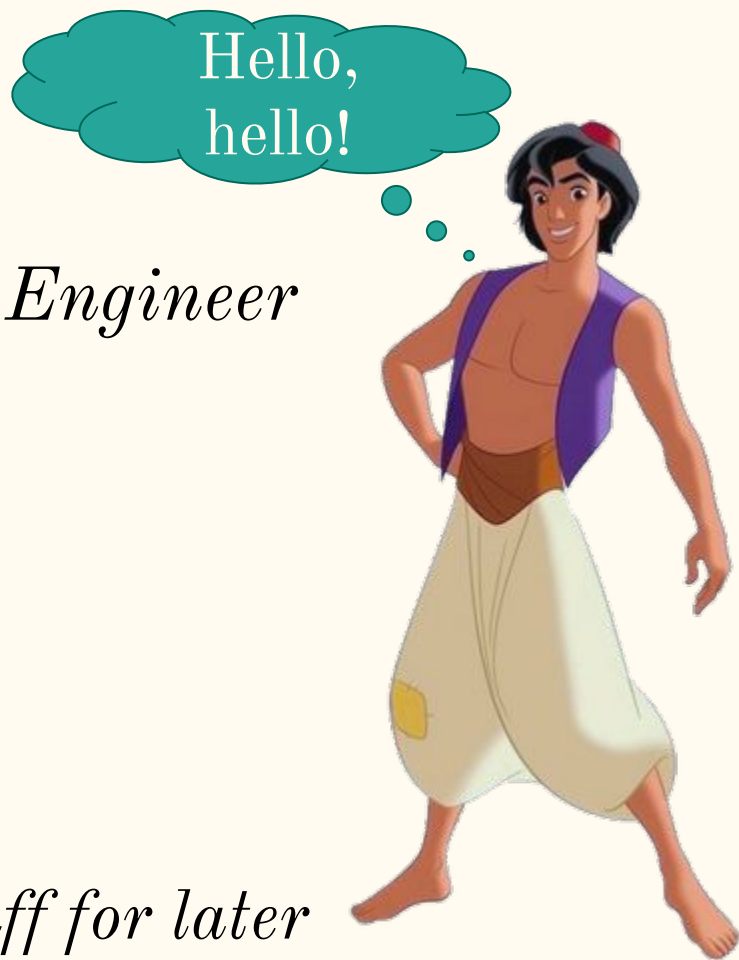
Presentation slides: *live*

https://bit.ly/backend-dev-to-ml-slides

Thank You!

So honoured!

- **Abhishek Thakur**, for organising this session, and giving me a chance to present at this forum
- And to "**you**", for sparing your valuable time and trusting me

Disclaimer

- ***YMMV***
- Might have rough edges and **inaccuracies**
- Sharing our **learnings** over the past years
- Gathered ideas from **different sources**
- **Sharing ideas** and **experiences**

Citation

The respective authors and creators are, and remain the true <u>owners of the images and other artifacts</u> used in this presentation.

Thank you for your creations!

Introduction

Summarising

~20 years till 2017

Dos

QuickBasic

VBA macros

First PC: 8086/88

Object Pascal / Delphi

Games: PacMan, Digger, **WITWI Carmen Sandiego**, Prince of Persia, Myst, among others

SmallTalk

Windows

Scala

R

Java

Linux

Bash shell

macOS

During the first half of my career I have been writing Desktop, command-line, REST-enabled and server-side apps on the Windows platform, and later doing the same on Linux and macOS.

@theNeomatrix369

Skipping a lot of details,
lots to talk about those
~20 years of my career

~2009 / ~2010:

Throwing myself into the deep-end of tech!

Stopped using Windows (secondary OS on the partition)

Very serious meetup go-er - sometimes *4 times a week (sometimes weekends)*

Throw myself in the deep-end: Linux CLI and macOS

Reading a lot

Introduction to Git

Getting more serious about Java

Interest and learning about VMs, Interpreters, Compilers, chipsets, GPUs

Building OpenJDK (using make and bash)

Running VMs: VirtualPC, vmWare

@theNeomatrix369

~2013 - ~2017:

A new phase begins

OpenJDK =>
AdoptOpenJDK

Core Java

Scala

JVM internals

Performance Tuning / GC

Shell scripting

Helping
Conference
organisers

Workshops and
speaking at
conferences

VBA Macros

Writing blogs

Agile methodology

Code quality

Tests and TDD

Craftmanship (crafter)

Code coverage

Fuzzy testing

Refactoring code

Mutation testing

Property-based testing

I was learning what it takes to become a *better professional!*

- Speak online or at conferences
- Write blogs or articles or papers
- Push code to git (personal or F/OSS projects)
- Review code, books, articles, etc...
- ....just help others become better at what they do
- Collaborate with other developers
- Be organised, practical and develop pragmatism
- Read and write a lot of code
- Rewrite any or all of the above (build muscle memory)
- and keep doing them iteratively, and more...

@theNeomatrix369

~2017 / 2018: another important phase

Switched to Freelancing

Best Practices

Moved roles
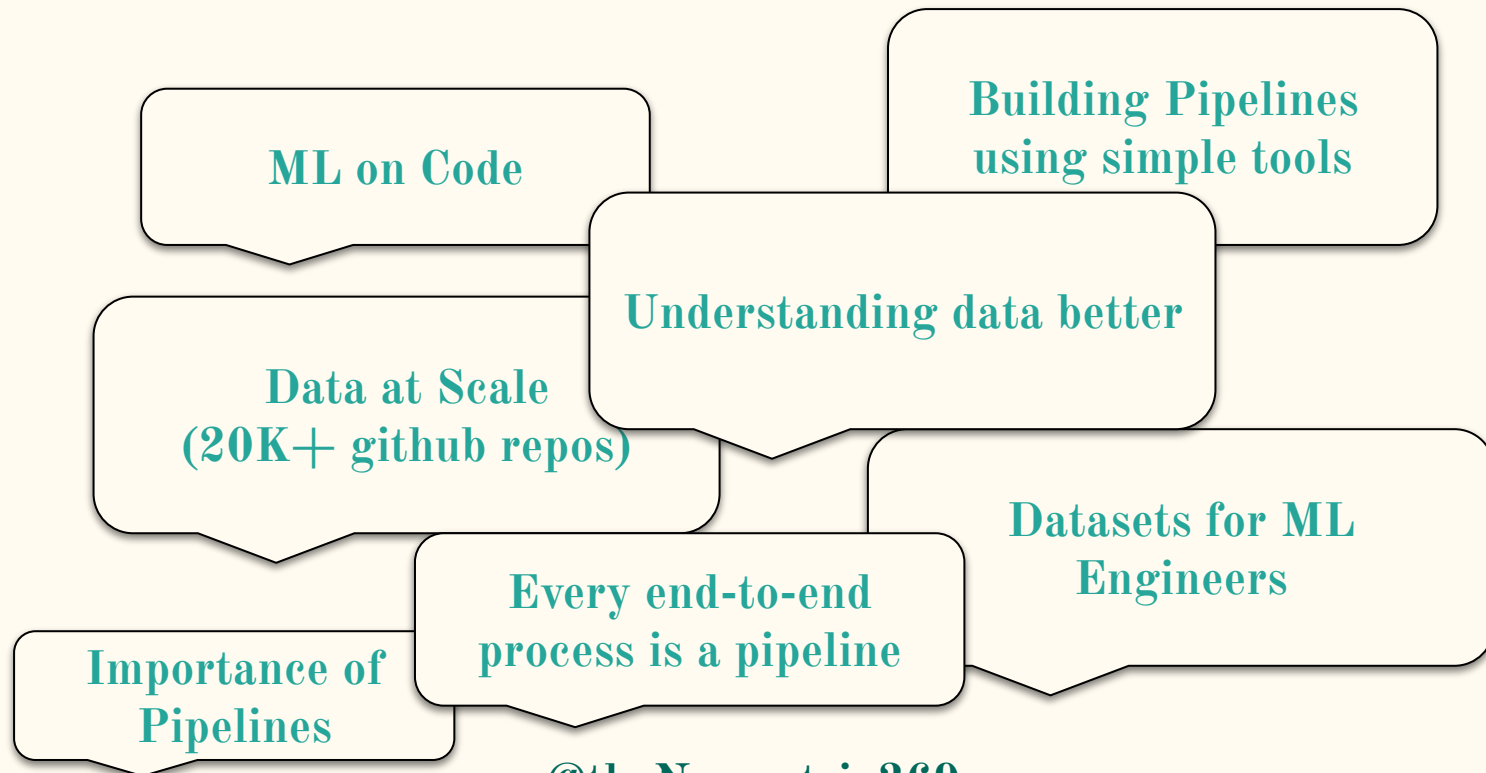
Small/Startup clients only

Small teams only

Accelerate development process

MVPs and PoCs

Development methodologies

@theNeomatrix369

Working on a ML/Data Engineering project

ML on Code

Building Pipelines using simple tools

Understanding data better

Data at Scale (20K+ github repos)

Datasets for ML Engineers

Every end-to-end process is a pipeline

Importance of Pipelines

@theNeomatrix369

AI/ML/DL: Learn and collaborate
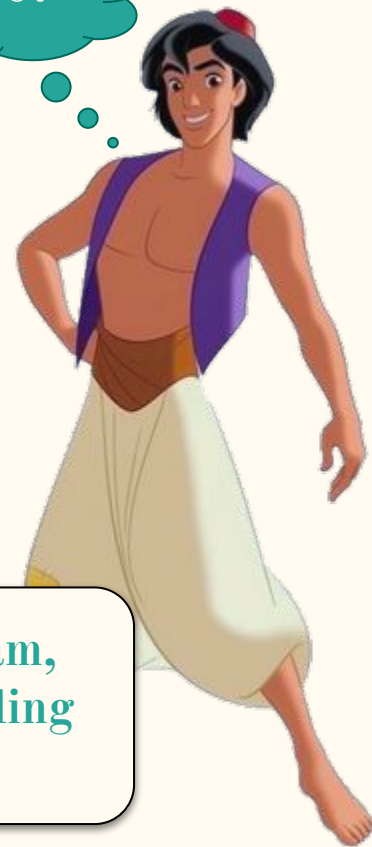
Bravo!

Joining Meetups

Attending Conferences

Participating in Study groups / AI Labs

Focus moved to more AI/ML/DL topics

WhatsApp, Telegram, Slack channels, Mailing list, etc...

@theNeomatrix369

Gave talks at UCL's CREST Open Workshops

See **awesome-graal** work

Discussing and developing
ML on Code ideas

My talks at the
CREST workshops:
[1] [2]

@theNeomatrix369

~2018 till date: AI/ML/DL fun begins

Thanks to Yolande Poirier for posting all those AI/ML/DL links on @java between 2016 and 2018

AI/ML/DL git repo

https://github/neomatrix369/awesome-ai-ml-dl

Please watch repo to follow latest updates and commits!

Please *star*, __watch,__ fork and share the repo, as it motivates me to continue with my work

AI/ML/DL artefacts

Cheatsheets

Notebooks

Presentations

Examples

Virgilio

Guides

Docker
containers

Study Notes

Things to know list

@theNeomatrix369

Thanks to the <u>Meet-a-Mentor</u> initiative by <u>LJC</u> during 2018-19, for holding the <u>ML Study Group</u>!

Talk on data

During this time I attended the study group
I paired with Jeremie Charlet to write a talk
on data, see Do we know our data, as well as
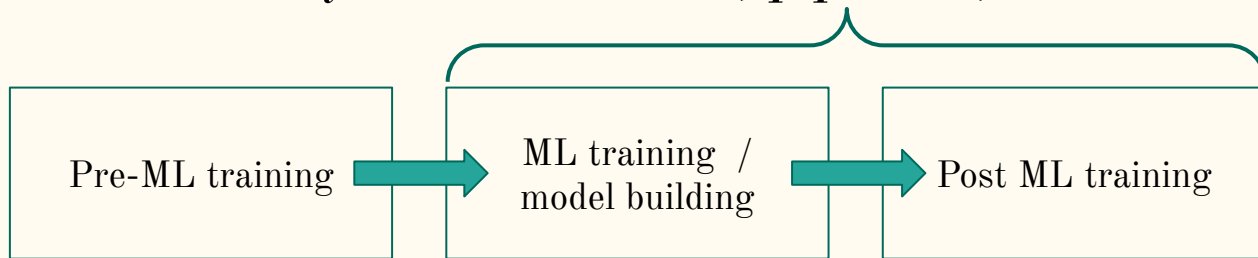we know our tools? talk!

Yes, yes I was there!
Recommend it.

Developing ideas about data and ML

Mostly <u>solved</u>: AutoML, pipelines, etc...

Pre-ML training	ML training / model building	Post ML training

Developing ideas about data and ML

So much to achieve...

Mostly <u>unsolved</u>: manual, human intervention, some tools

| Data Collection | → | Data Preparation | → | Feature Engineering |
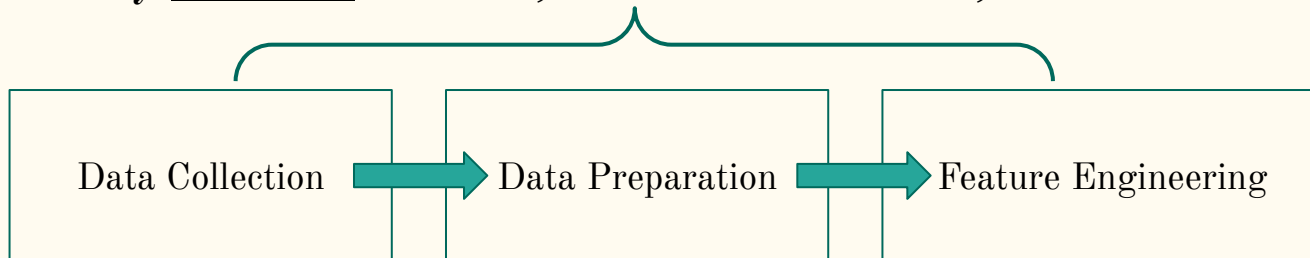
Developing ideas about data and ML

EDA, due diligence

*Very important steps, more important than **just building model***

| Data Collection | → | Data Preparation | → | Feature Engineering |
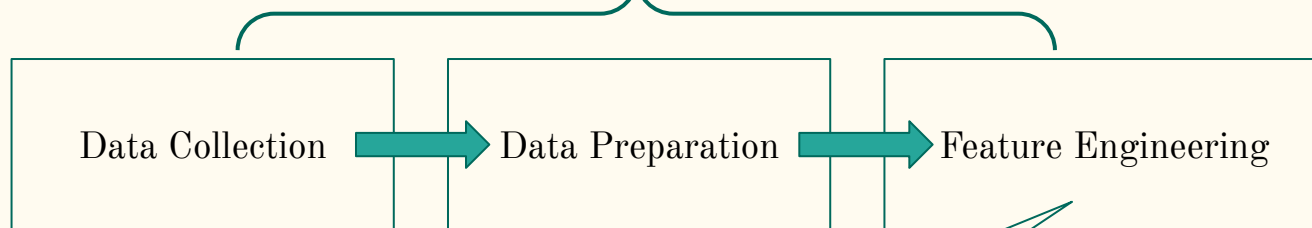
When done well gives reliable and robust models

Critical thinking

Thanks to <u>Yaz</u> for the Tensorflow meetups in London, UK during 2018 and 2019

My Study notes from Tensorflow and other topics

Rewrote a Keras post: BoW, Classification

Please have a look at all the TPU notebooks on each of the links!

Followed on by a post

on TPUs

Enjoyed CPU vs GPU vs TPU comparisons

How to harness the Powers of the Cloud TPU

Thanks to Professor Ajit Jaokar for his meetups and the AI Labs in London, UK during 2018 and 2019

Meetups and AI Labs

- Machine Learning is Fun! See Book | Tutorial / Blogs
- Better NLP library launch, see presentations:
  - First presentation: launch of Better NLP
  - Follow-up presentation of Better NLP

And many others in my journey...

Awesome AI/ML/DL repo

We are cross-linking resources with MWML & Virgilio

Thanks to the contributors for your pull requests

Collaboration with MWML

Working with Virgilio to make better guides

All things AI/ML/DL I'm excited about...

NLP

Metrics: tracking and analysis

Bayesian concepts

ML on Code

Maths

Visualisation

Model and experiment result analysis

Feature Engineering

Probability

Data, Numbers and distributions

Uncertainty concepts

Some of the fundamental ideas behind modelling perfection

Notebooks/kernels…

- A good way to learn anything…
- Here's a list of notebooks/kernels I rewrote/refactored/improved:
  - Kaggle kernels
  - Google Colab notebooks
    - Tensorflow GPU & TPU | TPU
    - Generative Image: CPU | GPU | TPU
  - Resources on AI/ML/DL

Have fun!

Richard Feynman
@ProfFeynman

Follow

If you're not having fun, you're not learning.
There's a pleasure in finding things out.

10:45 AM - 10 Oct 2018

1,426 Retweets 3,736 Likes

13   1.4K   3.7K
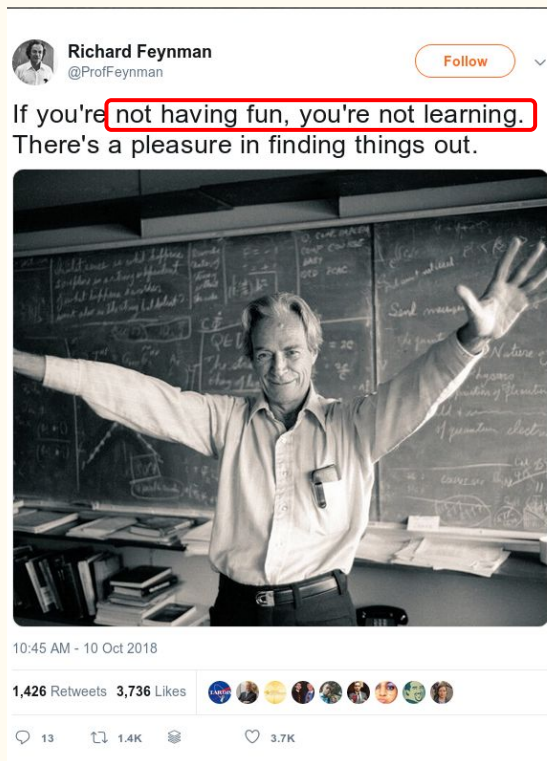
@theNeomatrix369

Participating in Competitions

It's so much fun!

- Late 2019: participating in competitions
- Become an Kaggle Discussion Expert two weeks ago
- Won my first Bronze medal at the recent Kaggle competition earlier this week
- See other results from online participations
- **Thanks:** Carlo, Marco, Abhishek and many others for your collaborations and inspirations

@theNeomatrix369

Demo: RoBERTa

About the demo: RoBERTa

- **Code on GitHub:**
  https://github.com/neomatrix369/chatbot-conversations
- Simple chatbot like app written in Python
- Runs as a Flask app (REST endpoint server)
- Demonstrates some NLP capabilities
- Download Roberta models
- Uses RoBERTa Base (others can be used as well)
- We wrote a similar one in Java using Eliza

File Edit View Search Terminal Tabs Help

./run-helidon-world.sh ✕   satyasai@sai-XPS-15-9560: ~/git-repos/ai-ml-dl-stuff/ch... ✕

```
Other world: were discussing you -- not me.
Helidon: We were discussing you -- not me.
Other world: We were discussing you -- not
Helidon: We were discussing you -- not me.
Other world: We were him -- not me.
Helidon: I rather hear from you first.
Other world: I must hear you first.
Helidon: You say must hear I first?
Other world: You say hear it first
Helidon: We were discussing you -- not me.
Other world: We were discussing you -- -
Helidon: Oh, I ?
Other world: Oh, and
Helidon: Please go on.
Other world: Please move on.
Helidon: I'm all ears!
Other world: I'm all set
Helidon: Can you elaborate on that?
Other world: Can you on...
Helidon: What are your feelings now?
Other world: What your options now
Helidon: Why do you ask?
Other world: How you ask
Helidon: What comes to mind when you ask that?
Other world: What comes to mind when hearing that
Helidon: Please go on.
Other world: To go on.
Helidon: I rather hear from you first.
Other world: I rather from the first.
Helidon: Do you say rather from the first for some special reason?
```

File Edit View Search Terminal Tabs Help

./run-helidon-world.sh ✕   satyasai@sai-XPS-15-9560: ~/git-repos/ai-ml-dl-stuff/ch... ✕

```
Other world: Why do you ask?
Roberta: How you ask? [confidence: 1.294%]

127.0.0.1 - - [02/Jul/2020 23:32:04] "GET /send?message=Why%20do%20y

Other world: What comes to mind when you ask that?
Roberta: What comes to mind when hearing that? [confidence: 31.04%]

127.0.0.1 - - [02/Jul/2020 23:32:12] "GET /send?message=What%20comes

Other world: Please go on.
Roberta: To go on. [confidence: 13.759%]

127.0.0.1 - - [02/Jul/2020 23:32:20] "GET /send?message=Please%20go%

Other world: I rather hear from you first.
Roberta: I rather from the first. [confidence: 63.393%]

127.0.0.1 - - [02/Jul/2020 23:32:28] "GET /send?message=I%20rather%2

Other world: Do you say rather from the first for some special reaso
Roberta: And you say rather from the first for some reason? [confide

127.0.0.1 - - [02/Jul/2020 23:32:36] "GET /send?message=Do%20you%20s
eason? HTTP/1.1" 200 -

Other world: We were discussing you -- not me.
Roberta: We were - you -- me. [confidence: 5.095%]

127.0.0.1 - - [02/Jul/2020 23:32:44] "GET /send?message=We%20were%20
```

```
                                                      ./run-connecting-wor
ile  Edit  View  Search  Terminal  Tabs  Help
        ./run-helidon-world.sh              ×   satyasai@sai-XPS-15-9560: ~/git-repos/ai-ml-dl-stuff/ch...  ×
elidon => Roberta: You say must hear I first?

oberta => Helidon: You say hear it first?

elidon => Roberta: We were discussing you -- not me.

oberta => Helidon: We were discussing you -- -

elidon => Roberta: Oh, I ?

oberta => Helidon: Oh, and ?

elidon => Roberta: Please go on.

oberta => Helidon: Please move on.

elidon => Roberta: I'm all ears!

oberta => Helidon: I'm all set

elidon => Roberta: Can you elaborate on that?

oberta => Helidon: Can you on...

elidon => Roberta: What are your feelings now?

oberta => Helidon: What your options now?

elidon => Roberta: Why do you ask?
```

I'm not calling these sensible conversations. LOL!

@theNeomatrix369

54

Puzzles: RoBERTa

- Did you notice anything unusual about the chatbots?
- What can we do to make it better?
- Can we use other models?
- Can we add more bots to the network?
- Can we make them make pass the Turing test?
- Is it a good idea?
- Any other ideas come to mind?

Demo: NLP Profiler

About the demo: NLP Profiler

- **Notebook & code on GitHub:**
  https://github.com/neomatrix369/awesome-ai-ml-dl/blob/master/examples/better-nlp/notebooks/jupyter/nlp_profiler.ipynb
  (https://github.com/neomatrix369/awesome-ai-ml-dl/blob/master/examples/better-nlp/library/org/neomatrix369/nlp_profiler.py)
- Simple python library-like module
- Liken to `pandas-profiling` but a lot simpler
- Tried to emulate Pandas' `describe()` function
- Can run on CLI (with changes) or `import` into a notebook
- Illustrates some use cases using a simple dataset
- More features can be built on top of the existing ones
- Can be integrated into the Better NLP library

```
[5]: profiled_text_dataframe = apply_text_profiling(text_dataframe, 'text')
     profiled_text_dataframe
```

	text	sentiment_polarity_score	sentiment_polarity	sentiment_subjectivity_score	sentiment_subjectivity	spellcheck_score
0	I love ⚽ very much 😄.	0.380000	Positive	0.43	Objective/subjective	1.000000
1	2833047 people live in this area. It is not a ...	-0.106818	Negative	0.55	Objective/subjective	0.968802
2	2833047 ... 111... people live ...	0.136364	Positive	0.50	Objective/subjective	1.000000
3	... sentence doesn't seem to too many commas,...	0.375000	Positive	0.75	Pretty subjective	0.923887
4	Todays date is 04/28/2020 for format ...	0.000000	Neutral	0.00	Very objective	0.711513

We can make better use of these continuous values!

Fuzzy mapping of scores to human-readable terms

@theNeomatrix369

Tab bar: nlp_profiler.ipynb ✕ | nlp_profiler-granular.ipynb ✕ | better_nlp_summarisers.ipyr ✕ | better_nlp_spacy_texacy_ex ✕

Toolbar: Code ⌄   Python 3 ◯

```
[5]: profiled_text_dataframe = apply_text_profiling(text_dataframe, 'text')
     profiled_text_dataframe
```

[5]:

spellcheck_score	spelling_quality	sentences_count	characters_count	spaces_count	words_count	duplicates_count	chars_excl_
1.000000	Good	2	21	5	4	0	
0.968802	Quite good	3	56	11	11	2	
1.000000	Good	2	42	7	6	0	
0.923887	Quite good	2	74	11	13	0	
0.711513	Pretty good	2	64	8	9	0	

May not be very accurate (~70%)

See https://en.wikipedia.org/wiki/Words_of_estimative_probability - how we map the probability scores to English words

Puzzles: NLP Profiler

- What are the limitations of the NLP Profiler?
- What can we do to make it better?
- Can we make it more accurate?
- If we have scaling issues how do we tackle it?
- Any other ideas come to mind?

Throwing oneself at the deep-end

Build sample apps (similar to the demos), use simple ideas, improve, share, rinse and repeat!

How do I learn?

See previous talks &

Appendix section

Summary

- Timeline of my career progression
- Demos & projects
- Things I'm excited about in AI/ML/DL
- Components of AI/ML/DL
- Links to "Previous talks" and other "Learning Resources" to follow
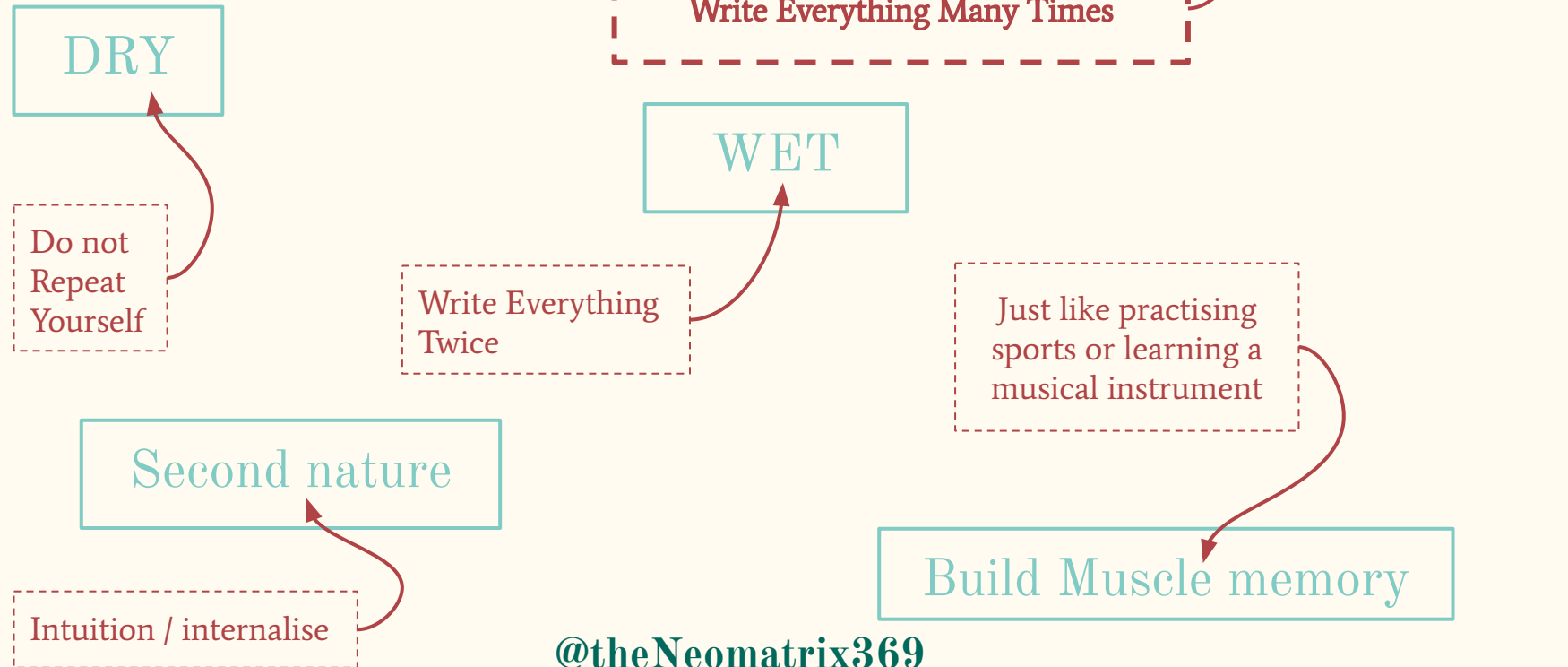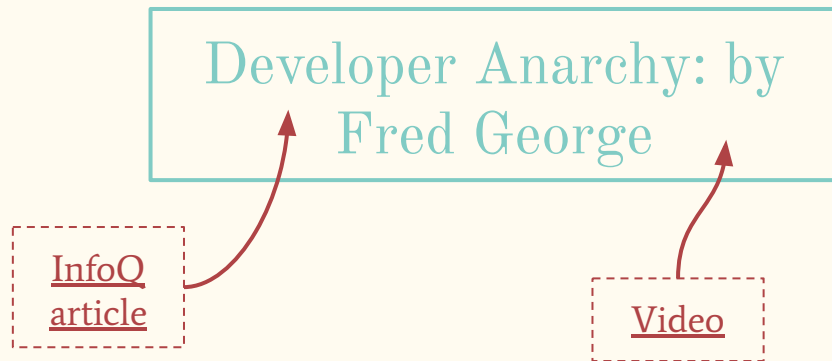- Masters & Grandmasters: thought exercises in the Appendix section

@theNeomatrix369

Previous talks

- I recently gave a talk on ["nn" things every Java developer should know about AI/ML/DL](#)
- [Naturally, getting productive, my journey with Grakn and Graql](#)
- [Do we know our data as well as our tools?](#)
- [Java N.n: What to know? How to learn?](#)
- Some of my other talks a can be found [here](#) and [here](#) (and others on [Slideshare](#))

Repetition Theory

WEMT

Write Everything Many Times

DRY

WET

Do not Repeat Yourself

Write Everything Twice

Just like practising sports or learning a musical instrument

Second nature

Build Muscle memory

Intuition / internalise

@theNeomatrix369

Repetition Theory: WEMT (Write Everything Many Times)

Developer Anarchy: by Fred George

InfoQ article

Video

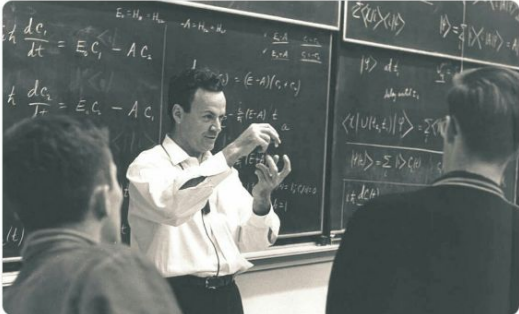# Repetition Theory: Richard Feynman quote

Richard Feynman
@ProfFeynman

Follow

The FEYNMAN method of learning:

1) Choose a concept and study it quickly
2) Teach it to someone, like a child, who is unfamiliar with the topic
3) Revise to fill in the gaps
4) Review and simplify
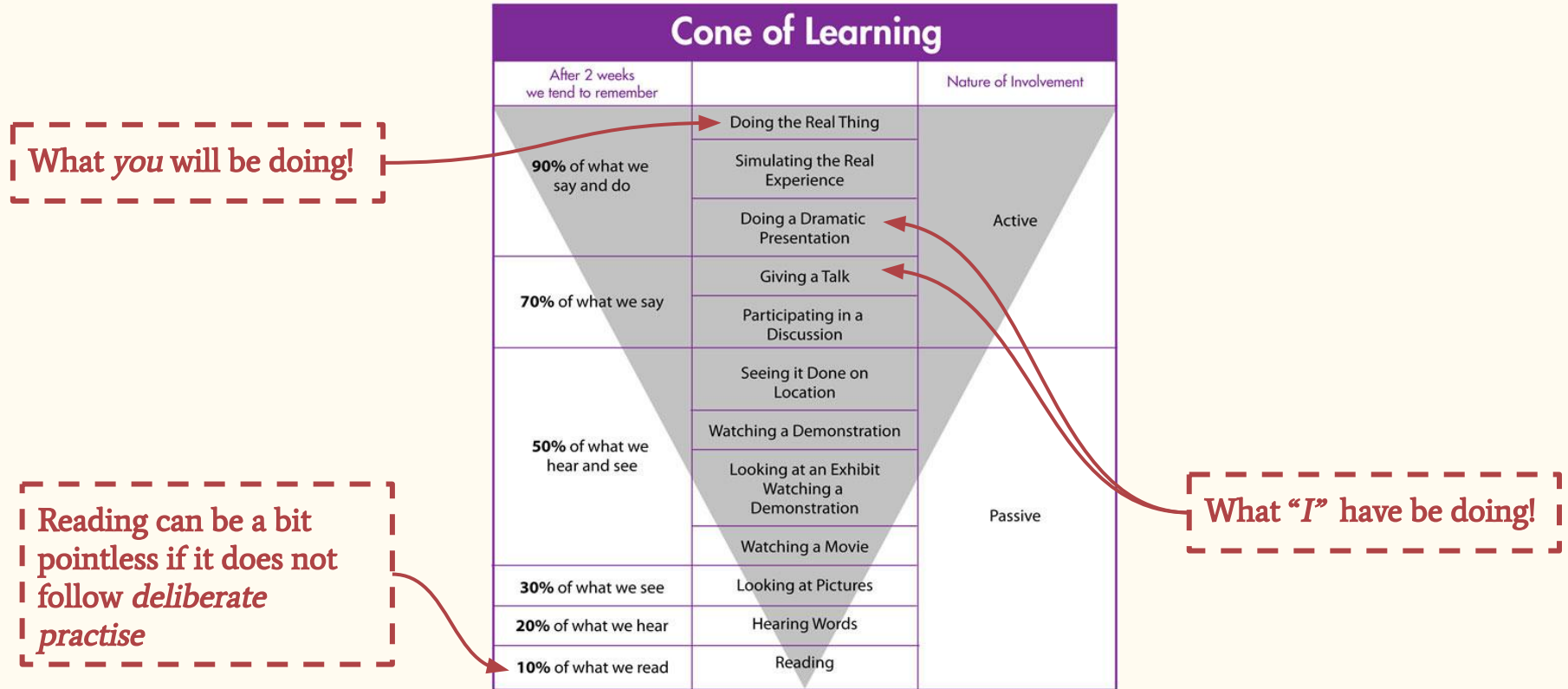(repeat step 2 and 4 until you can explain the concept in your own words)

9:20 am - 25 Sep 2018

1,377 Retweets 3,369 Likes

20   1.4K   3.4K

@theNeomatrix369

Cone of Learning

Cone of Learning

After 2 weeks we tend to remember		Nature of Involvement
90% of what we say and do	Doing the Real Thing	Active
	Simulating the Real Experience	
	Doing a Dramatic Presentation	
70% of what we say	Giving a Talk	
	Participating in a Discussion	
50% of what we hear and see	Seeing it Done on Location	
	Watching a Demonstration	
	Looking at an Exhibit Watching a Demonstration	Passive
	Watching a Movie	
30% of what we see	Looking at Pictures	
20% of what we hear	Hearing Words	
10% of what we read	Reading	

What *you* will be doing!

What *"I"* have be doing!

Reading can be a bit pointless if it does not follow *deliberate practise*

Insight

Writing code is Deep Work
Software Design is Deep Work

Learning websites

- [Awesome AI/ML/DL](#)
- [Better NLP](#)
- [Virgilio](#) | [GitHub](#)
- [MadeWithML](#)

All you need to know, pick and choose, make your own collections / checklists / playlists of resources to study and practice from.

Coherence with the brain

- Brain indexes knowledge
  - when you are resting
    - sleep and rest ⟵ **Very important!!!**
  - let the brain or the mind index away
- Don't overload,
  - instead interleave
  - fill the gaps
  - learn incrementally & iteratively

@theNeomatrix369

Book references

- Code Complete
- Clean Code
- Refactoring (Martin Fowler and Joshua Kerievsky)
- Four rules of simple design

Book references

- [Approaching (Almost) Any Machine Learning Problem](#)
- [How to think like a Scientist!](#)
- [Deep Work book](#)
- Getting Things Done: [Book](#) | [Free Resources](#)

Other Resources

> So much to know and learn, my, my!

- No need to feel overwhelmed
- Kagglers, read this for inspiration
- See Appendix section of the slide deck

@theNeomatrix369

Contact and keep in touch

- twitter: **@theNeomatrix369**
- medium: **https://medium.com/@neomatrix369**
- github: **https://github.com/neomatrix369/**
- linkedin: **https://www.linkedin.com/in/mani-sarkar/**
- youtube: **channel** | **playlists**
- kaggle: **https://kaggle.com/neomatrix369**
- about me: **https://neomatrix369.wordpress.com/about**

Thank You!

- **Abhishek Thakur**, for organising this session, and giving me a chance to present at this forum
- And to <u>you</u>, for sparing your valuable time and trusting in me

Closing

The pathway to <u>mastery</u> or <u>championing</u> a subject means we must be different and take <u>different approaches</u>, in addition to the ones that have <u>already been taken</u>!

Appendix

Learning by example

DL4J example

- [Github](#)
- [Blog post](#)

NLP examples

- Example 1
  - [Github](#)
  - [Blog post](#)
- Example 2
  - [Blog post](#)
- Example 3
  - [Blog post](#)
- [Better NLP](#)

Jupyter Notebook example

- Example 1
  - Github
  - Blog: Exploring NLP concepts using Apache OpenNLP inside a Jupyter notebook
- Example 2
  - Blog post
- Example 3
  - Github
  - Blog post

@theNeomatrix369

Apache Zeppelin example

- [Github](#)
- Blog posts
  - [Apache Zeppelin: stairway to notes* haven!](#)
  - [Running Apache Zeppelin on Oracle Cloud Infrastructure](#)

graql-to-english, english-to-graql example

- [Presentation](#)
- [Github](#)

grCuda example

- Blog posts
  - [grCUDA: A Polyglot Language Binding for CUDA in GraalVM. NVIDIA Developer Blog, November 2019.](#)
  - [grCUDA: A Polyglot Language Binding. Presentation at Oracle CodeOne 2019, September 2019.](#)
  - [Simplifying GPU Access. Presentation at NVIDIA GTC 2020, March 2020](#)
  - [Optimizing Machine Learning Performance at Netsuite with GraalVM and NVIDIA GPUs](#)
- [Github](#)

graalPython examples

- Blog posts
  - [Introduction to the Python implementation for GraalVM](#)
  - [Moving from Jython to GraalVM](#)
  - [Running Python on GraalVM](#)
- [Github](#)

How do I learn?
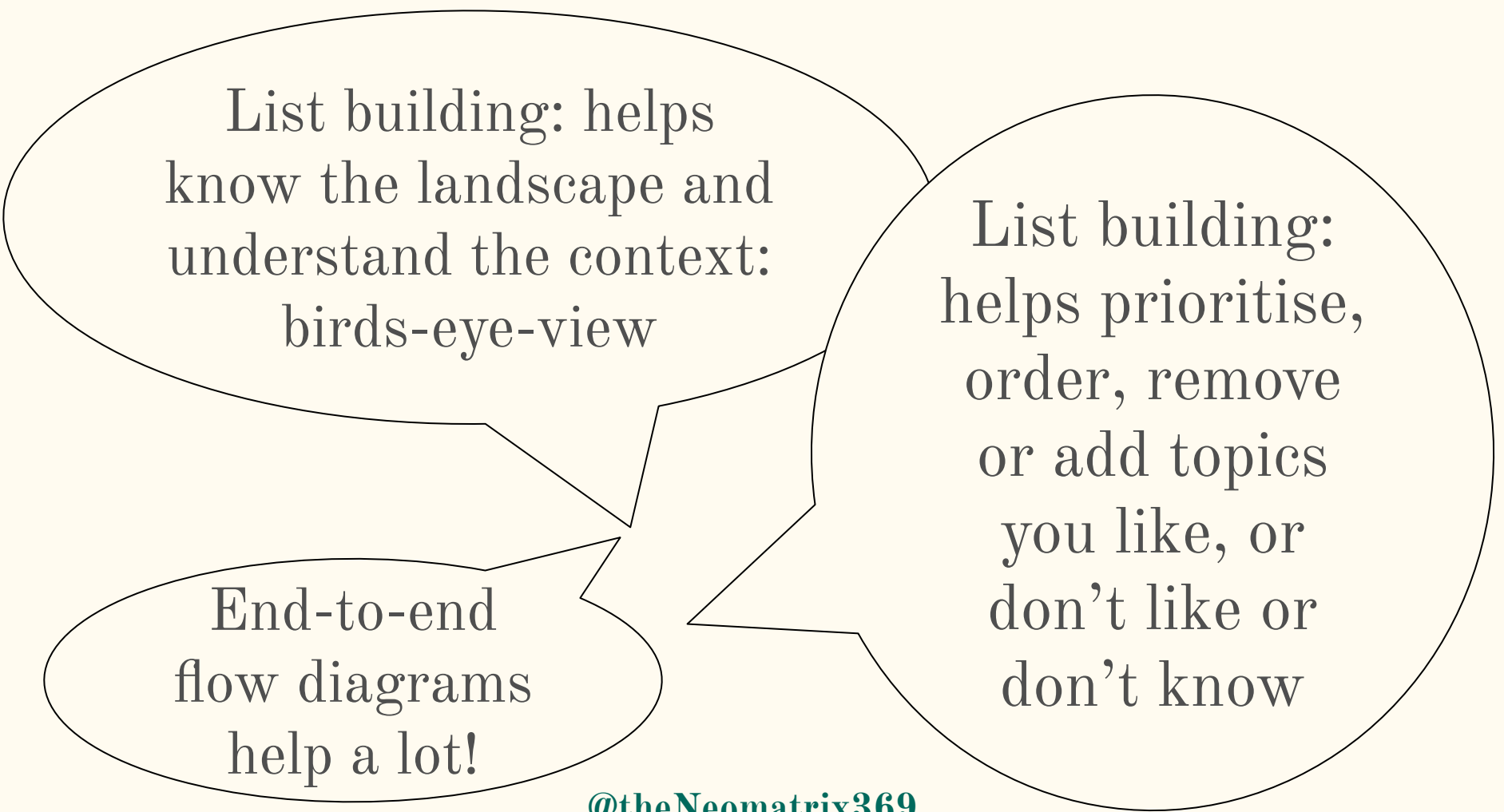
@theNeomatrix369

One may find these methods *unconventional* or *non-mainstream* <u>but they do work and give good results</u>!

How to build and how not to?

How **not to build** a minimum viable product
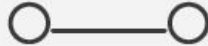
How **to build** a minimum viable product

@theNeomatrix369

Useful advice

See what [Subhaditya](#) say about intuition

- His talk on Abhishek's talk series: [Talks # 2: Subhaditya Mukherjee - Image restoration using Deep Learning: Dehazing](#)
- And this blog post on ["imposter syndrome"](#)

From KDM Delhi Kaggle Expert & Kaggle Master

- Sound advice from : [Talks # 2: Subhaditya Mukhe1st Anniverary Meetup #3 || AMA with Mayank Kumar & Siddharth Yadavrjee](#)

Masters and Grandmasters

Why do they inspire us?

- Start early: the amazing performance by little Souparnika Nair

- ...and keep going

- ...meet the masters (the droplets returns to the ocean)

What's a common thread in all of them, in all of their performances?
(thought exercise)

Master of grandmasters?

[Britain's Got Talent 2020 Souparnika Nair: 10 Year Old Singer Full Audition S14E07](#)

What does it take to become such a talented person? Why?
(thought exercise)

The <u>journey</u> continues, the road maybe long, it's full of <u>learnings that help us grow</u>, many a times a winding points but the <u>milestones keep moving</u> - we are living <u>in such unprecedented times</u>!

Feynman quotes

Being wrong isn't bad...

Me believe same!

t⏎ You Retweeted

Richard Feynman @ProfFeynman · 19h
Being wrong isn't a bad thing like they teach you in school. It is an
opportunity to learn something.

💬 22   🔁 1.2K   ❤️ 4.1K   ↥

@theNeomatrix369

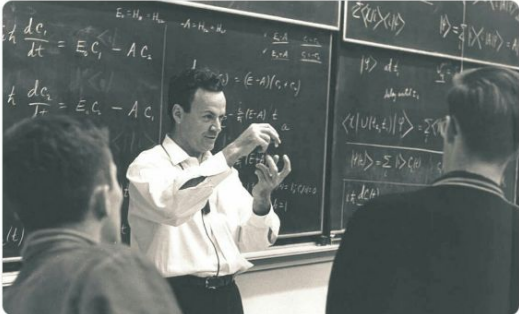# Repetition Theory: Richard Feynman quote

Richard Feynman
@ProfFeynman

The FEYNMAN method of learning:

1) Choose a concept and study it quickly
2) Teach it to someone, like a child, who is unfamiliar with the topic
3) Revise to fill in the gaps
4) Review and simplify
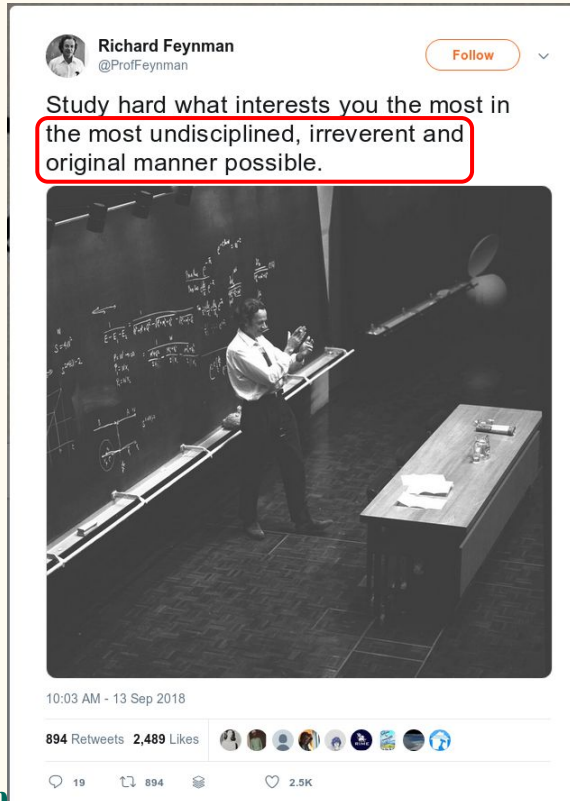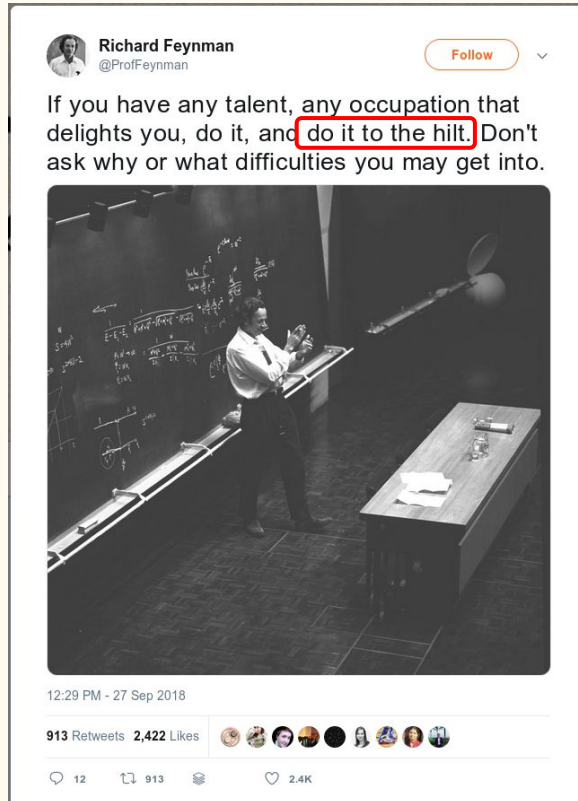(repeat step 2 and 4 until you can explain the concept in your own words)

9:20 am - 25 Sep 2018

1,377 Retweets 3,369 Likes

20   1.4K   3.4K

@theNeomatrix369

More Richard Feynman quotes...

Richard Feynman
@ProfFeynman
Follow

If you have any talent, any occupation that delights you, do it, and do it to the hilt. Don't ask why or what difficulties you may get into.

12:29 PM - 27 Sep 2018

913 Retweets 2,422 Likes

12   913   2.4K

Richard Feynman
@ProfFeynman
Follow

Study hard what interests you the most in the most undisciplined, irreverent and original manner possible.

10:03 AM - 13 Sep 2018

894 Retweets 2,489 Likes

19   894   2.5K

@theNeoMatrix999

Other fun stuff

@theNeomatrix369

Comedians really know their data!

55 seconds of fun

Link to the video

@theNeomatrix369