<Vincent Tatan>
vintatan@google.com

Intro to Machine Learning

Wait, who am I?

Tak Kenal maka Tak Sayang,
Tak Sayang, maka Tak Tanya
Tak Tanya, maka Tak Tahu
-- Vincent Tatan --

Meet Vincent

Safe Browsing Analyst (Machine Learning)
Google Trust & Safety

Medium : towardsdatascience.com/@vincentkernn
Linkedin : linkedin.com/in/vincenttatan/
Data Podcast: https://datacast.simplecast.com/

Path to Google

Visa
Data & Architecture Engineer
Jun 17 - Aug 19

Google
Data Analyst, Machine Learning
Aug 19 - Present

Lazada Group
Data Scientist Intern
Dec 16 - Apr 17

SMU
B.Sc., Management Information Systems and Services
Aug 13 - July 17

Google: Trust and Safety

To **prevent phishing @ scale** with **Data Analytics and ML**

So, what do I do at Google?

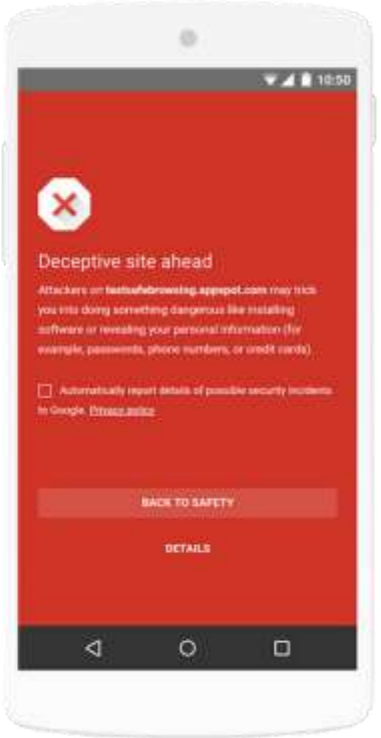
SAFETY AND SECURITY

The geeky detective-work that protects you online, automatically

Trust and Safety

Protect more than 3 billion devices worldwide

1. **Google notifies your browsers** to prevent **phishing** and **malware**.
2. Using machine learning-based detection, we contributed to **99.9% accuracy in spam detection**
3. So if you see this, **beware!**



What is machine learning?

Machine Learning

Computational methods using **experience** to improve **performance**

Machine Learning

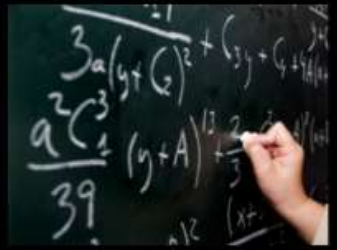
Using **Computer** and **data** to **achieve objective**

- **Computer** → Algorithm, complexity analysis, theoretical guarantees.
- **Data analysis**→ Statistics, probability
- **Achieve Objective** → Understanding the problem, simulation, evaluation, etc

Machine Learning

What society thinks I do

What stock holders think I do



What my manager thinks I do

What product teams think I do

What I think I do

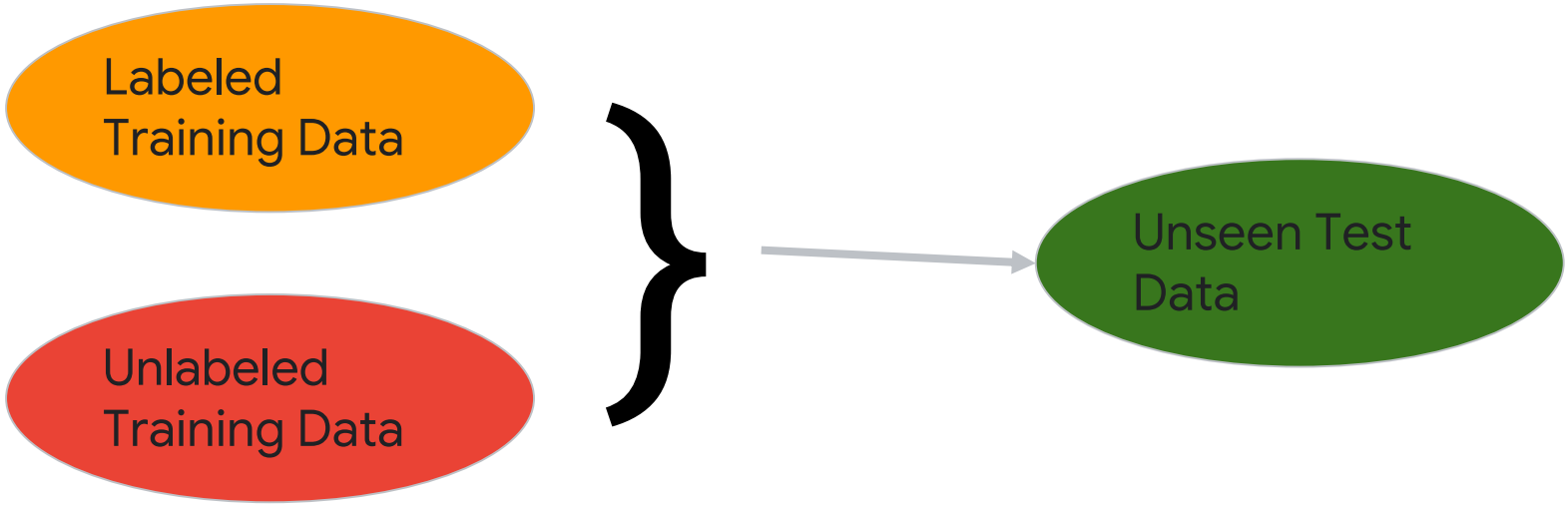
What I really do

Supervised vs Unsupervised

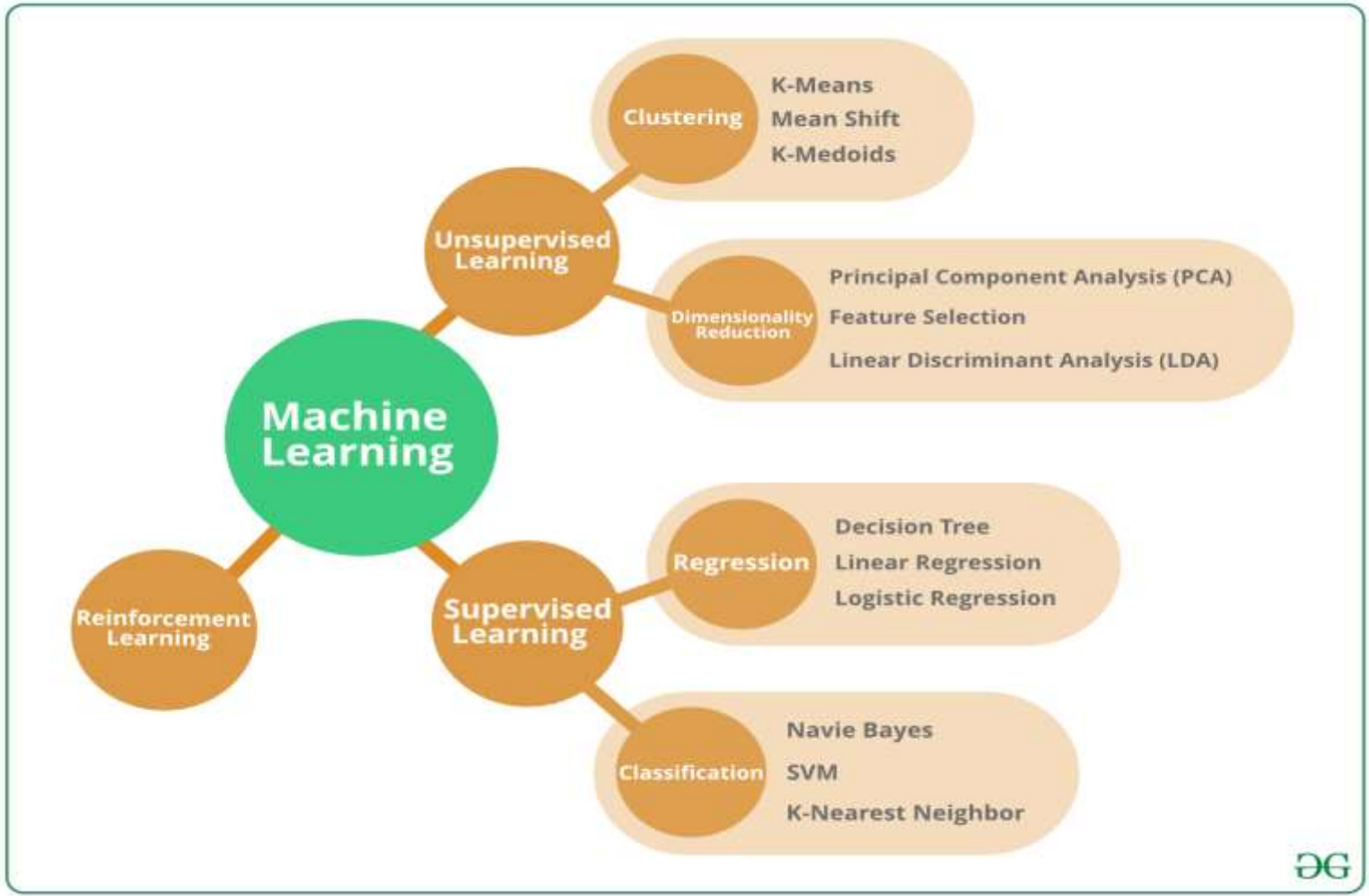
Unsupervised Learning : No labeled data. Finding patterns/insights

Supervised Learning: Most common learning scenarios

Semi Supervised Learning : With labeled and unlabeled training data



Why? Training data might imply same distributions.

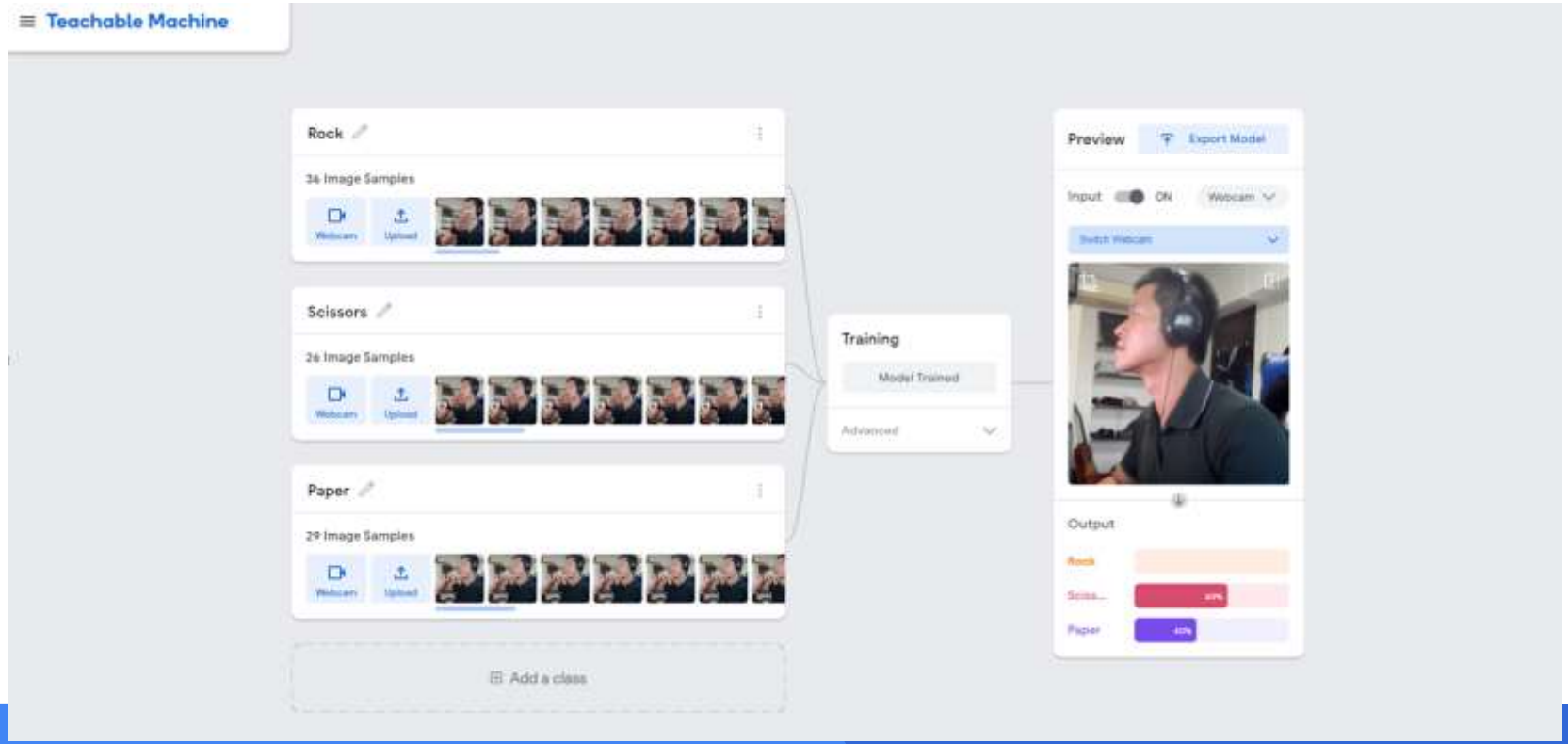


Real World Impact

What are the applications of AI and ML?

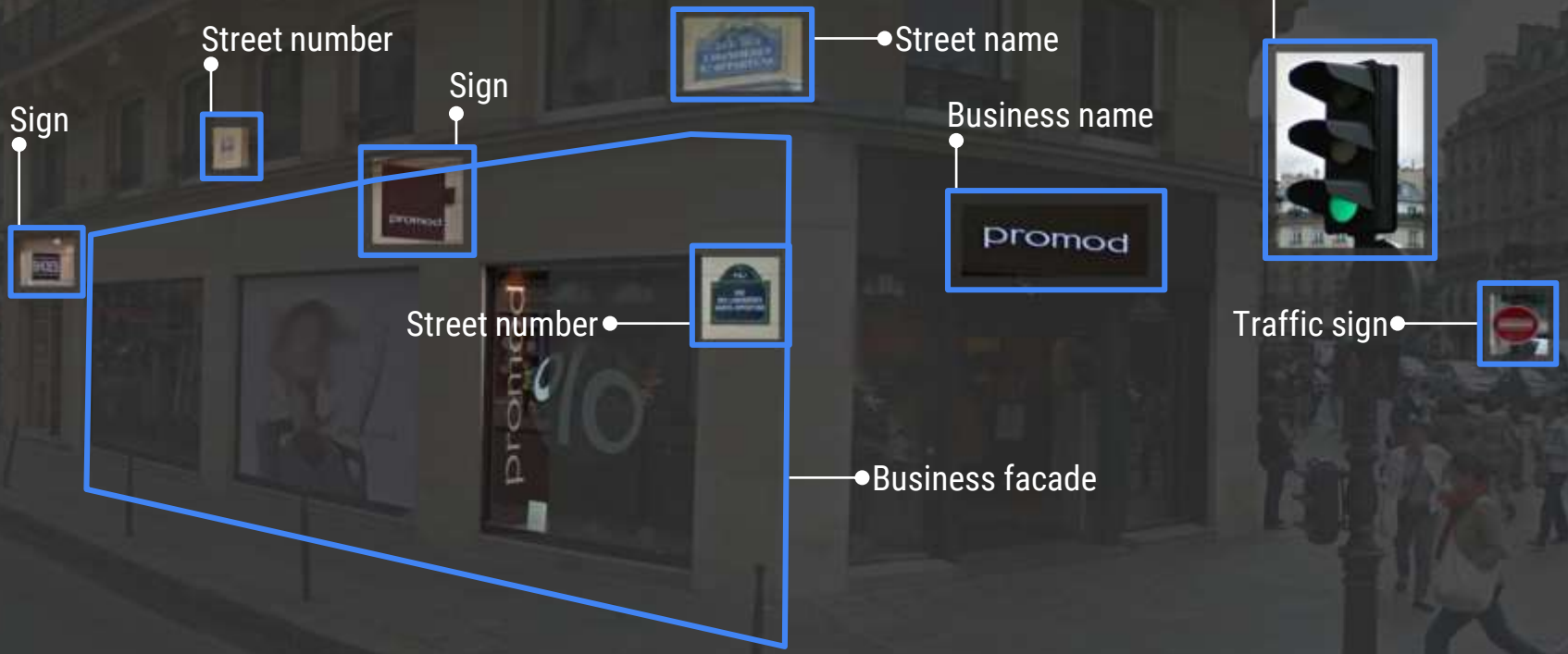
Important Resources (Teachable Machine)

https://teachablemachine.withgoogle.com/train/image




1
4

Street View

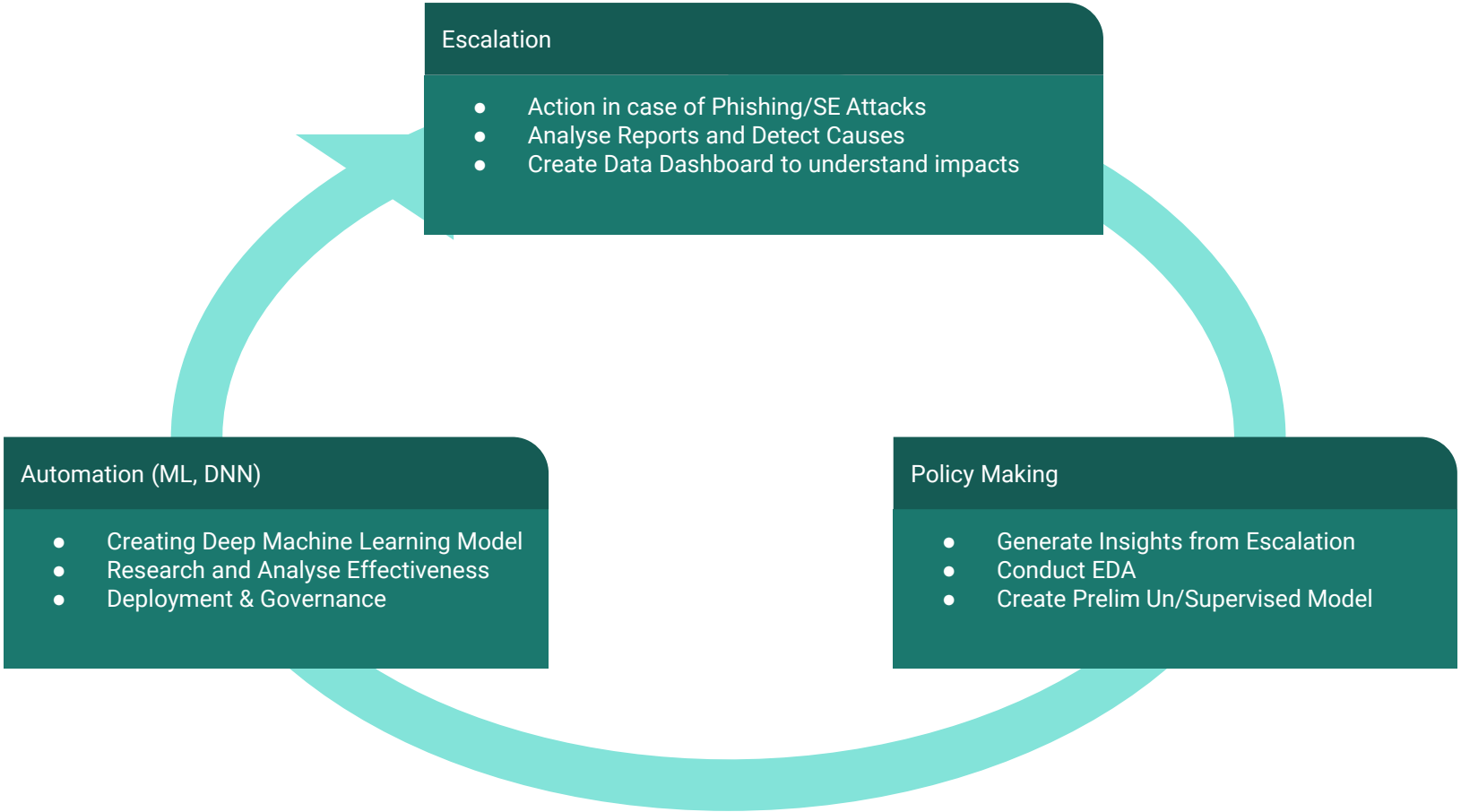

Sign
Street number
Sign
Street name
Business name
promod
Traffic light
Street number
Traffic sign
Business facade


Google Translate
EXIT

What do I do mostly?

My Life at Google...

Focus Work: The cycle of Data Project



ML Pipeline

Frame the Problem

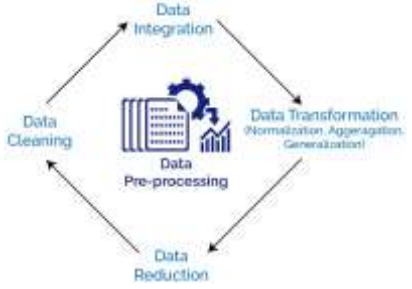
What is your goal?

Who are your stakeholders?

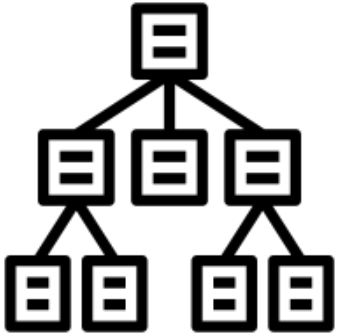
How do you add value to them?

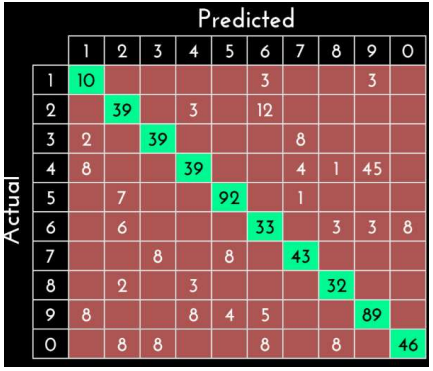
ML Pipeline

Data Collection + Preprocessing



Model Training and Evaluation



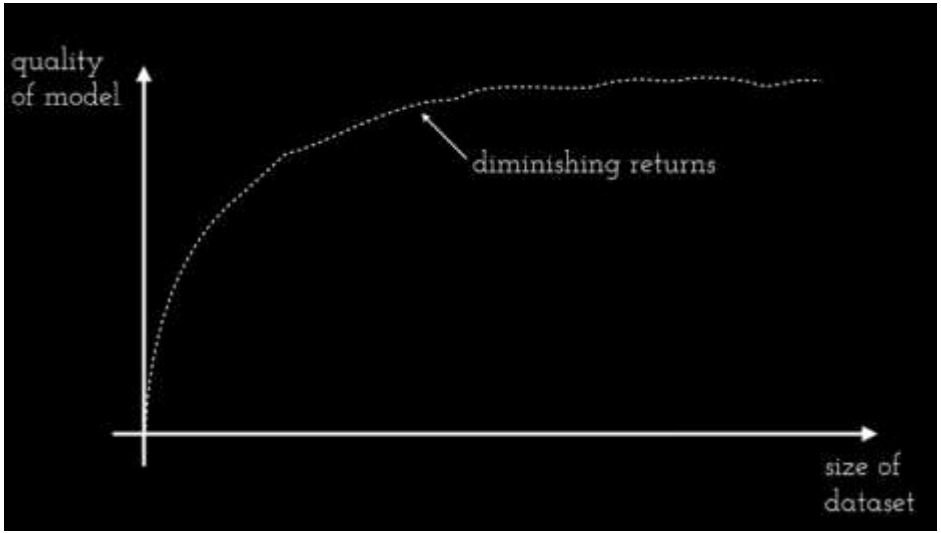


Machine Learning Operations (MLOps)

Data Collection

More data beats smarter algorithms

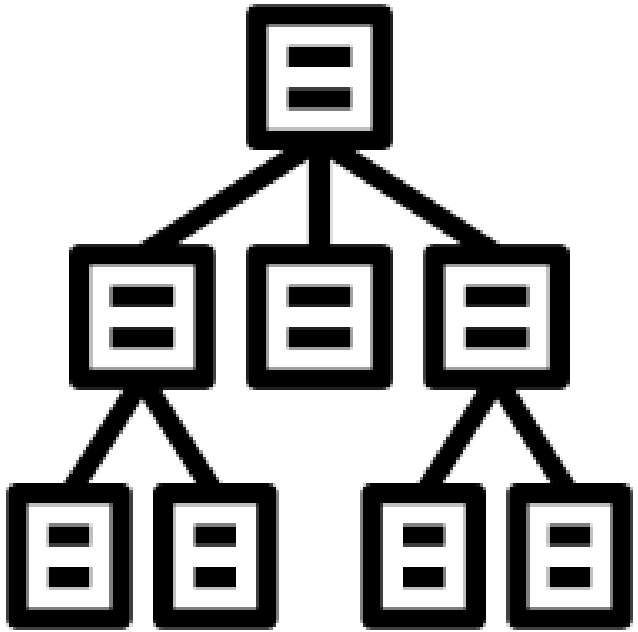
1. **But it is not practical**
2. Data is **expensive**. Money and time to collect labels
3. Big data might **be overkill**



Model Training

Based on different use cases

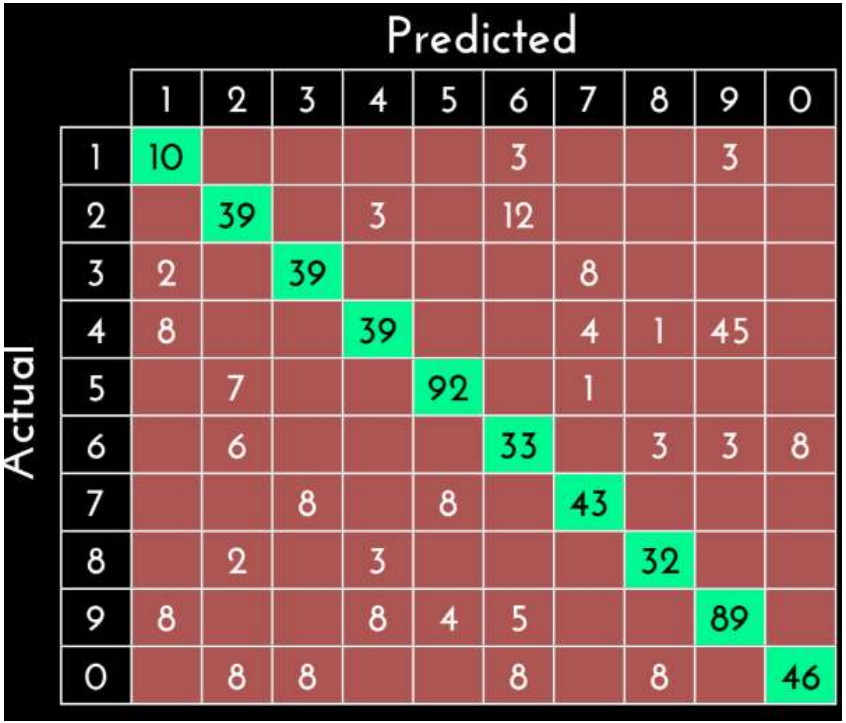
1. **Regression: n dim-Polynomial?**
2. **Classification:** Decision tree, logistic regression SVM
3. Each of the algorithm has multiple characteristics:
    a. Susceptible to outliers
    b. Explainability



Model Evaluation

Is it useful?

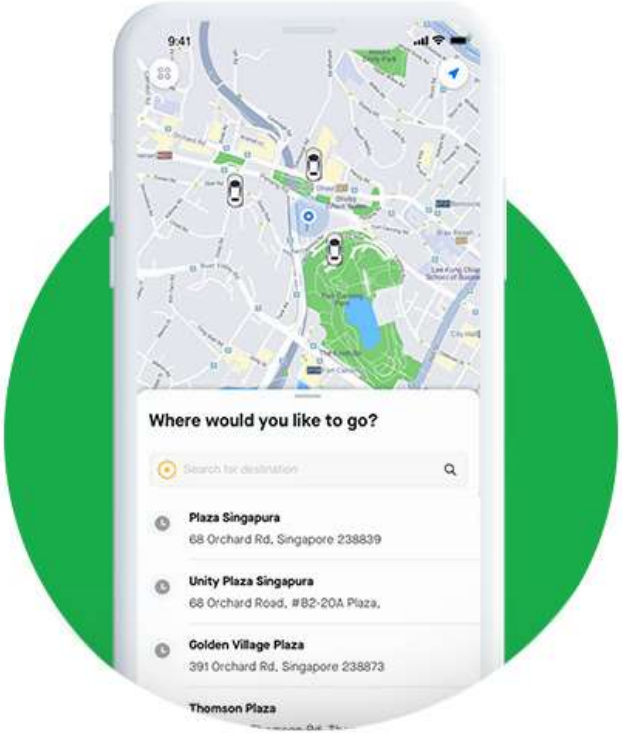
1. **Regression:** Root Mean Squared Error (RMSE)
2. **Classification:** Confusion metrics, AUC, Precision, Recall, F1
3. **Complexity, explainability, latency** (time and space)
4. **Eager/Lazy learners**



ML Ops

Operating real ML for real Use Case

1. **Model Push**
2. **Model Validation**
3. **Monitoring/Anomaly Detection**

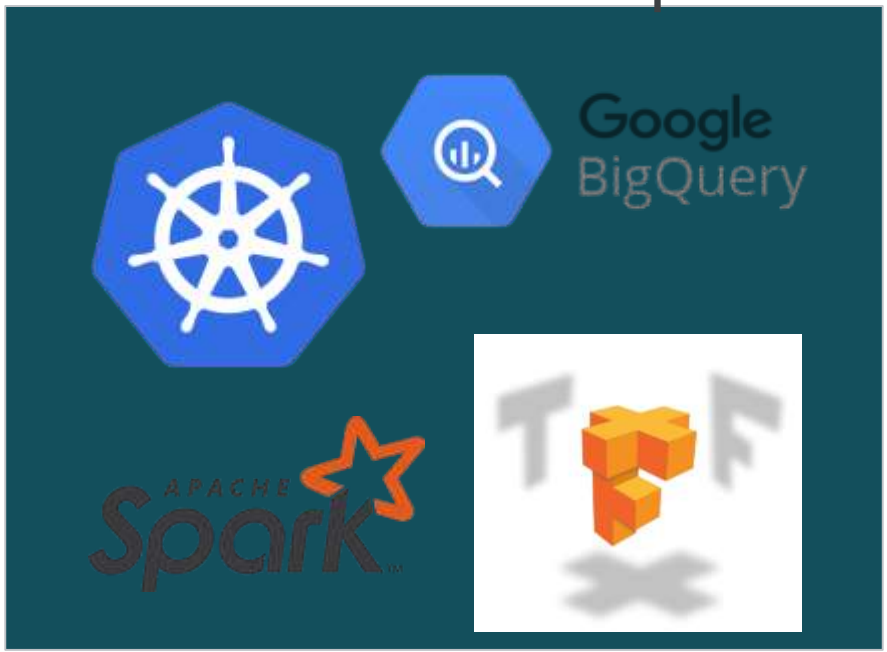


ML Tech Stack

Development + IDE

Language + Library

Data + ML Ops



Analytics / ML Trend

How Analytics enter/menyurupi our lives?

Analytics Development in Indonesia

Infrastructure & Business Growth

Education & Political Supports

Crisis & Innovation Movement

4G Technology is out in Indonesia

First time 4G is out in Indonesia. High smartphone adoption with large digital market.

Large Growth from now "unicorn" Tech

Gojek Launches Android and iOS apps for 4 services: transportation, courier, and shopping

Traveloka becomes the leading choice of online flights and hotel bookings.

UI opens the first big data curriculum

The first analytics curriculum opened by tertiary institutions

The expansion of Jakarta Smart City supported by Ahok for Government and Startup Collaboration

Strong Political Support of Indonesia Tech-Education Movement

Nadiem was appointed as the minister of Education

Merdeka Hackathon and Merdeka belajar is held

Sandiaga Uno and Anies greatly supports these initatives.

Corona Virus

Corona virus affects livelihood while promoting the attractiveness of data science / IT sector

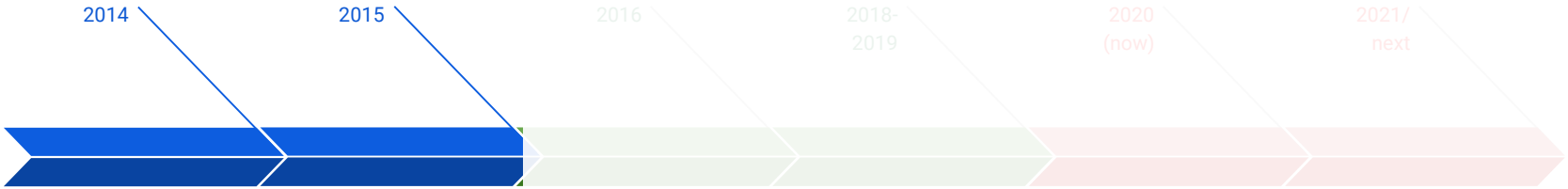
Data is the new Electricity

The unity of political movements, businesses, startups, and capital labour to boost data analytics movement in Indonesia

The coming of 5G in Indonesia

Analytics Development in Indonesia

Infrastructure & Business Growth



2014

4G Technology is out in Indonesia

First time 4G is out in Indonesia. High smartphone adoption with large digital market.

2015

Large Growth from now "unicorn" Tech

Gojek Launches Android and iOS apps for 4 services: transportation, courier, and shopping

Traveloka becomes the leading choice of online flights and hotel bookings.

Education & Political Supports

2016

UI opens the first big data curriculum

The first analytics curriculum opened by tertiary institutions

The expansion of Jakarta Smart City supported by Ahok for Government and Startup Collaboration

2018-2019

Strong Political Support of Indonesia Tech-Education Movement

Nadiem was appointed as the minister of Education

Merdeka Hackathon and Merdeka belajar is held

Sandiaga Uno and Anies greatly supports these initatives.

Crisis & Innovation Movement

2020 (now)

Corona Virus

Corona virus affects livelihood while promoting the attractiveness of data science / IT sector

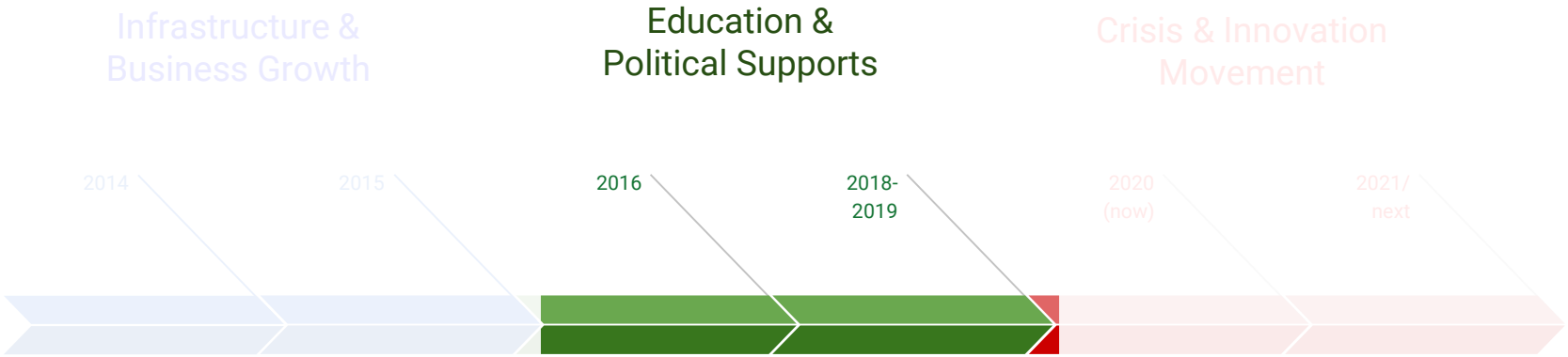
2021/ next

Data is the new Electricity

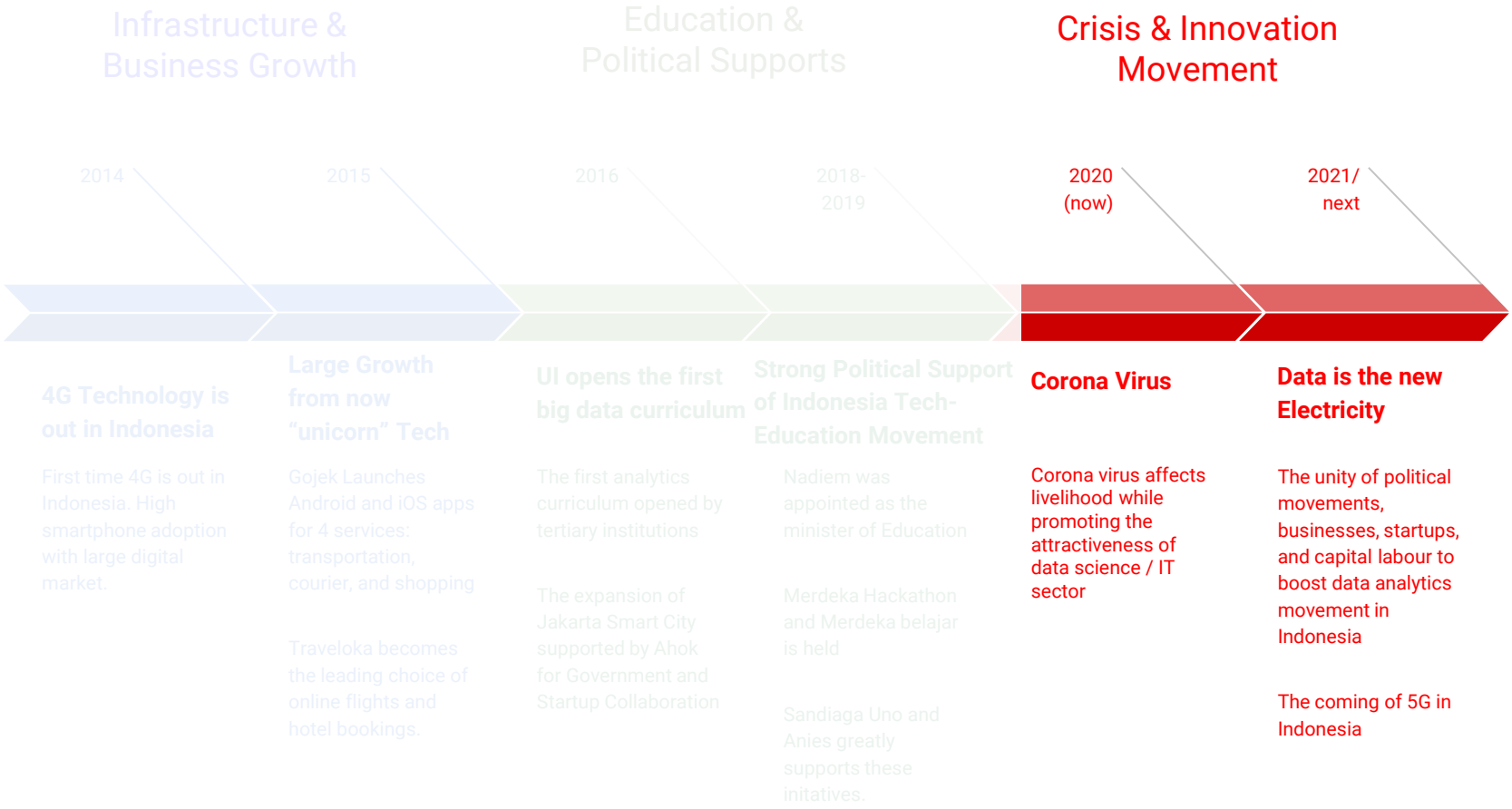
The unity of political movements, businesses, startups, and capital labour to boost data analytics movement in Indonesia

The coming of 5G in Indonesia

Analytics Development in Indonesia



Analytics Development in Indonesia

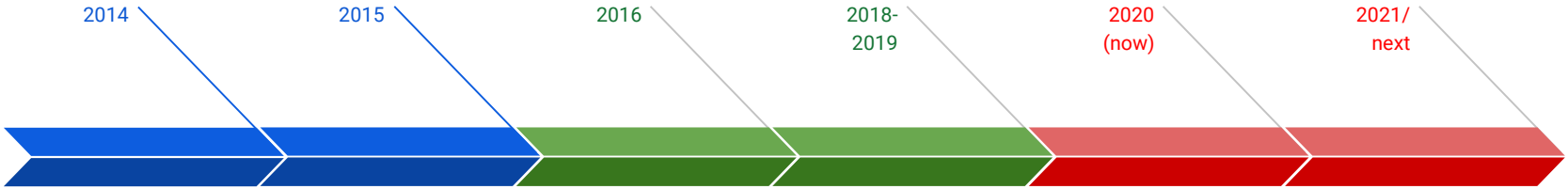


Analytics Development in Indonesia

Infrastructure & Business Growth

Education & Political Supports

Crisis & Innovation Movement



2014 — 4G Technology is out in Indonesia

First time 4G is out in Indonesia. High smartphone adoption with large digital market.

2015 — Large Growth from now "unicorn" Tech

Gojek Launches Android and iOS apps for 4 services: transportation, courier, and shopping

Traveloka becomes the leading choice of online flights and hotel bookings.

2016 — UI opens the first big data curriculum

The first analytics curriculum opened by tertiary institutions

The expansion of Jakarta Smart City supported by Pak Ahok for Government and Startup Collaboration

2018-2019 — Strong Political Support of Indonesia Tech-Education Movement

Nadiem was appointed as the minister of Education

Merdeka Hackathon and Merdeka belajar is held

Pak Sandiaga Uno and Pak Anies greatly support these initatives.

2020 (now) — Corona Virus

Corona virus affects livelihood while promoting the attractiveness of data science / IT sector

2021/ next — Data is the new Electricity

The unity of political movements, businesses, startups, and capital labour to boost data analytics movement in Indonesia

The coming of 5G in Indonesia

How can you excel in ML?

I’m super excited! What’s next!

Data Analytics is hard... Here's how you can excel

What I have learnt over 3 years as Data Analysts/Engineers at Google, Visa, and Lazada ★

Vincent Tatan
Oct 20 · 8 min read ★

Upskill yourself (source)

https://bit.ly/2rBUP0Y

Ace your Data Analytics Interviews

My experience Interview as a Data Analyst/ Scientist at Google, Visa and many more

Vincent Tatan
Sep 16 · 9 min read ★

https://bit.ly/2Pc2axe

Contribute more!

Data Science expands a lot, share your knowledge

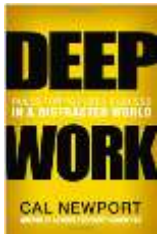
1. **Read more**: Keep experimenting with your learning styles (kinesthetic, auditory, visual)
2. **Write more**: Write articles and share them!
3. **Speak more**: Teach your fellow peers or any conferences out there!

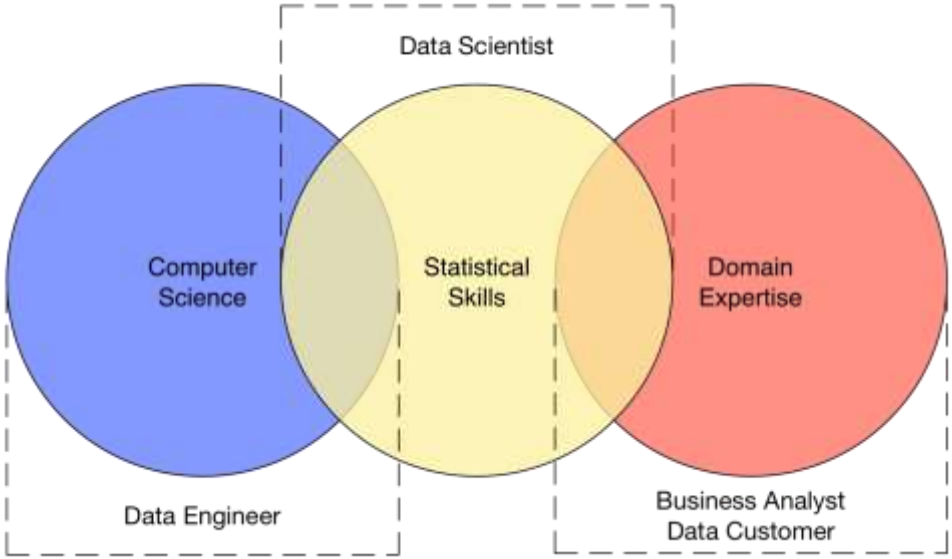
Generalize and Specialize

Follow it to the T

1. **Strength:** Utilize your biggest strengths
2. **Communicate:** Communicate your strength and impacts more.
3. **Learn in T**: SQL & Python/R are the breadth, then domain knowledge is your depth

Read Deep Work





Smile!

Data Science is Fun

1. **Play:** Tough, so have fun.
2. **Hack:** Use Saturdays to learn with friends.
3. **Celebrate impacts:** Data science is about building impacts. Start small and celebrate!

This is the key to excel

Contribute, Prepare, and Smile

Questions?