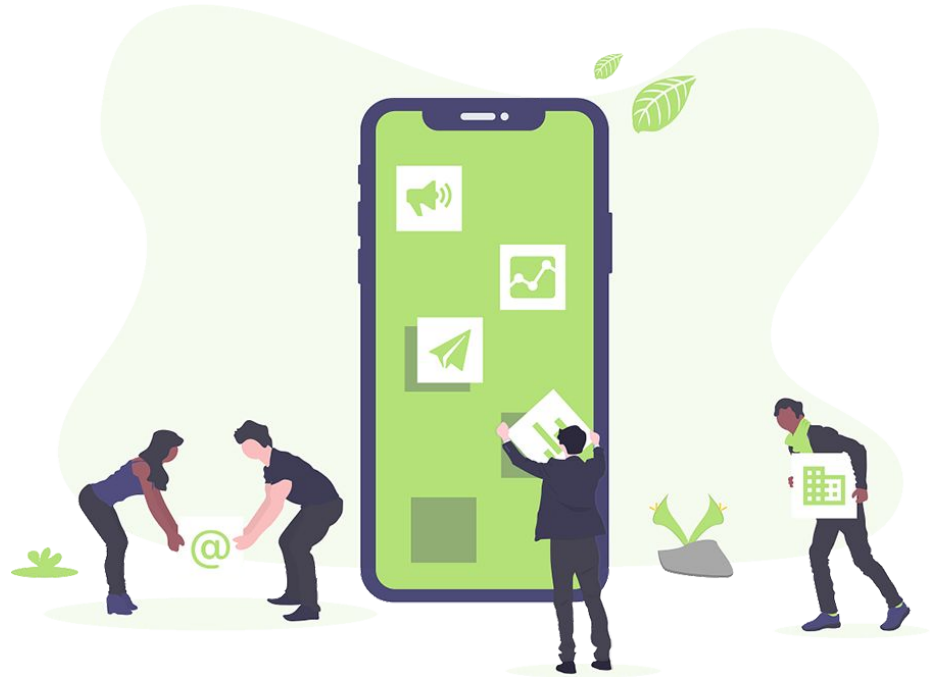# Machine Learning:

An Industry Perspective

Hello!

I am Bramantyo Erlangga (bram)

Data Engineer at Go-Jek

Former Data Scientist at Prism and Midtrans

Former Solution Architect at Ericsson

1.
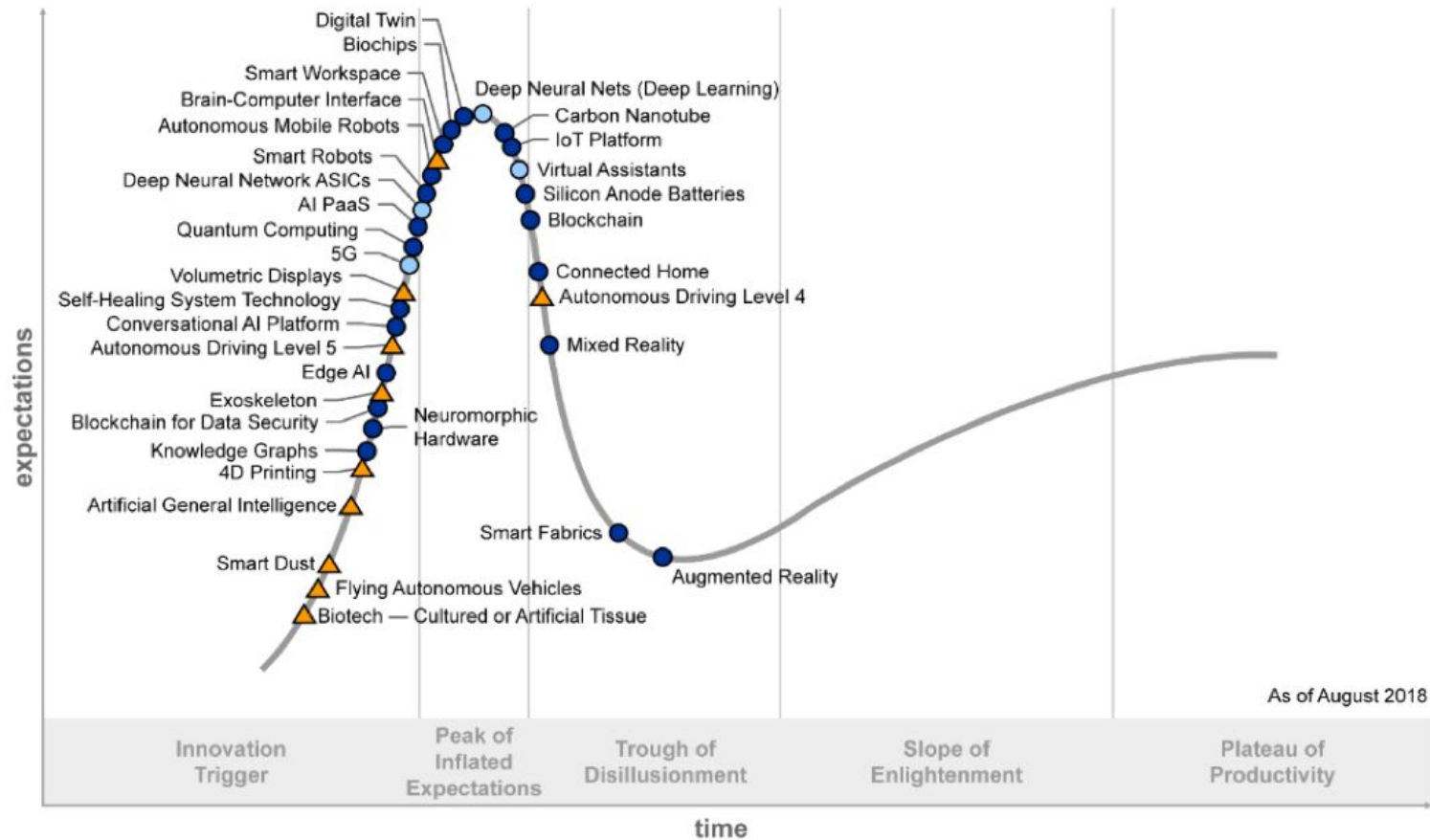# The Hype

Fear of Missing Out

> " Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it....
>
> **Dan Ariely**

Gartner Hype Cycle for Emerging Technologies, 2018

- Digital Twin
- Biochips
- Smart Workspace
- Brain-Computer Interface
- Autonomous Mobile Robots
- Smart Robots
- Deep Neural Network ASICs
- AI PaaS
- Quantum Computing
- 5G
- Volumetric Displays
- Self-Healing System Technology
- Conversational AI Platform
- Autonomous Driving Level 5
- Edge AI
- Exoskeleton
- Blockchain for Data Security
- Knowledge Graphs
- 4D Printing
- Artificial General Intelligence
- Smart Dust
- Flying Autonomous Vehicles
- Biotech — Cultured or Artificial Tissue
- Neuromorphic Hardware
- Deep Neural Nets (Deep Learning)
- Carbon Nanotube
- IoT Platform
- Virtual Assistants
- Silicon Anode Batteries
- Blockchain
- Connected Home
- Autonomous Driving Level 4
- Mixed Reality
- Smart Fabrics
- Augmented Reality

expectations

time

| Innovation Trigger | Peak of Inflated Expectations | Trough of Disillusionment | Slope of Enlightenment | Plateau of Productivity |

As of August 2018

Plateau will be reached:

○ less than 2 years   ○ 2 to 5 years   ● 5 to 10 years   ▲ more than 10 years   ⊗ obsolete before plateau

© 2018 Gartner, Inc.

5

5

Sometimes

- Some company still clueless on how to embrace machine learning towards their organization

- Remember, "The most valuable commodity is information"
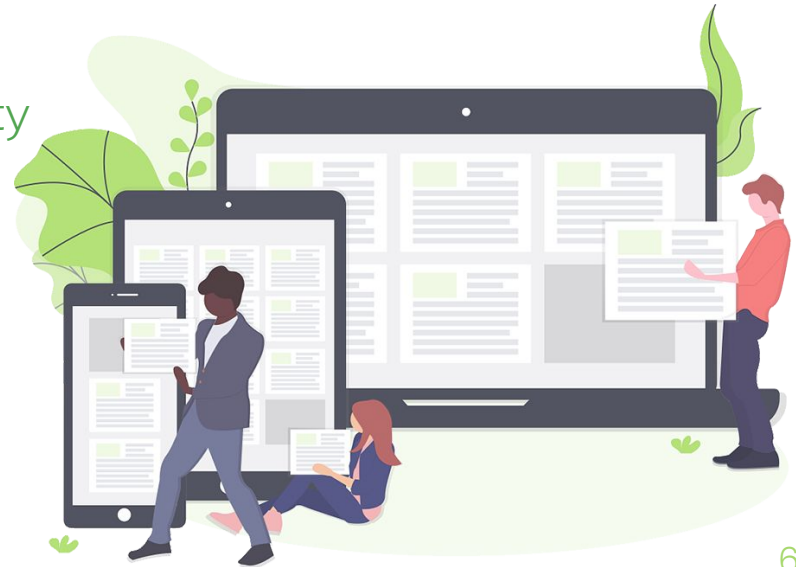
2.
It's All About Value

Value, value, value

It's not always

- The most sophisticated algorithm
- The most advanced methodologies
- The most powerful infrastructure

It's all about value

- Focus on solving problems
- Focus on give added values to the organization (increase revenue, reduce OPEX, new features, etc)
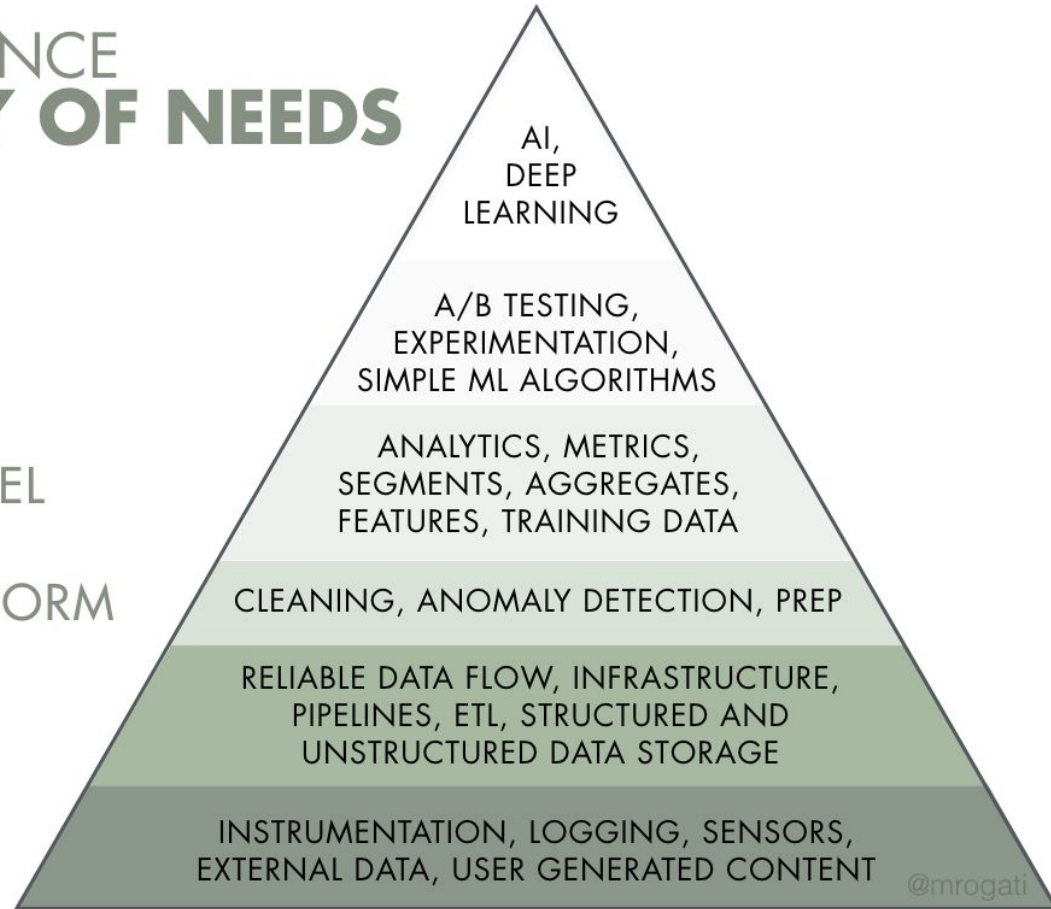
THE DATA SCIENCE
HIERARCHY OF NEEDS

AI,
DEEP
LEARNING

LEARN/OPTIMIZE

A/B TESTING,
EXPERIMENTATION,
SIMPLE ML ALGORITHMS

AGGREGATE/LABEL

ANALYTICS, METRICS,
SEGMENTS, AGGREGATES,
FEATURES, TRAINING DATA

EXPLORE/TRANSFORM

CLEANING, ANOMALY DETECTION, PREP

MOVE/STORE

RELIABLE DATA FLOW, INFRASTRUCTURE,
PIPELINES, ETL, STRUCTURED AND
UNSTRUCTURED DATA STORAGE

COLLECT

INSTRUMENTATION, LOGGING, SENSORS,
EXTERNAL DATA, USER GENERATED CONTENT

@mrogati

Data Preparation

- Because most of data scientist time is to prepare the data before implements any ML algo
- Then it's mandatory to focus on good data infrastructure before focusing on ML research
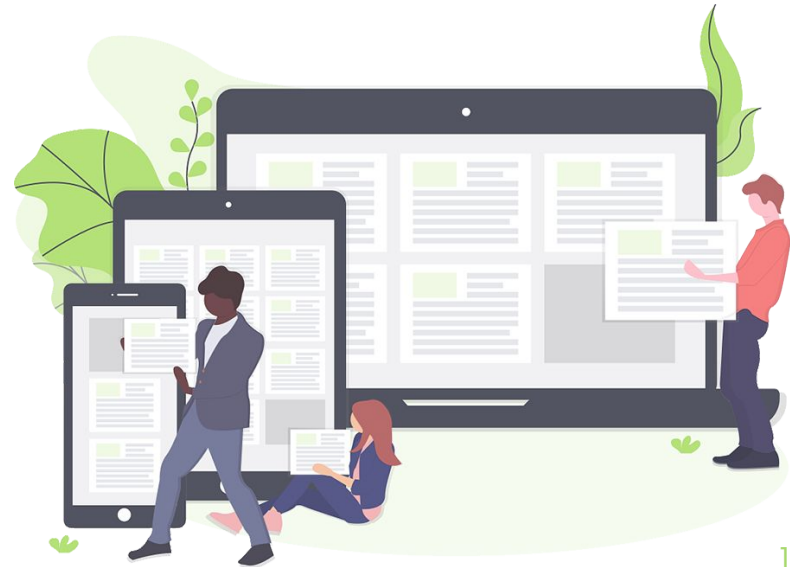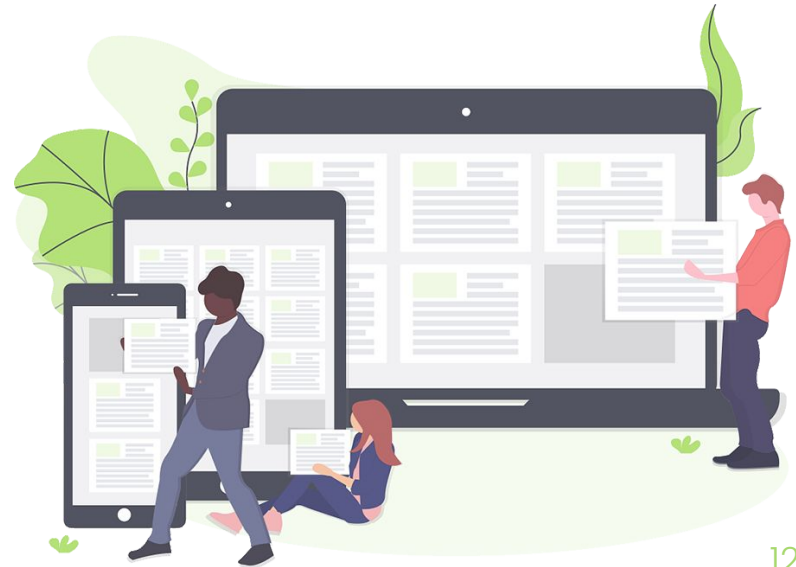
Examples of Application

- Fraud Prevention
- Customer Segmentation
- Recommendation Engine
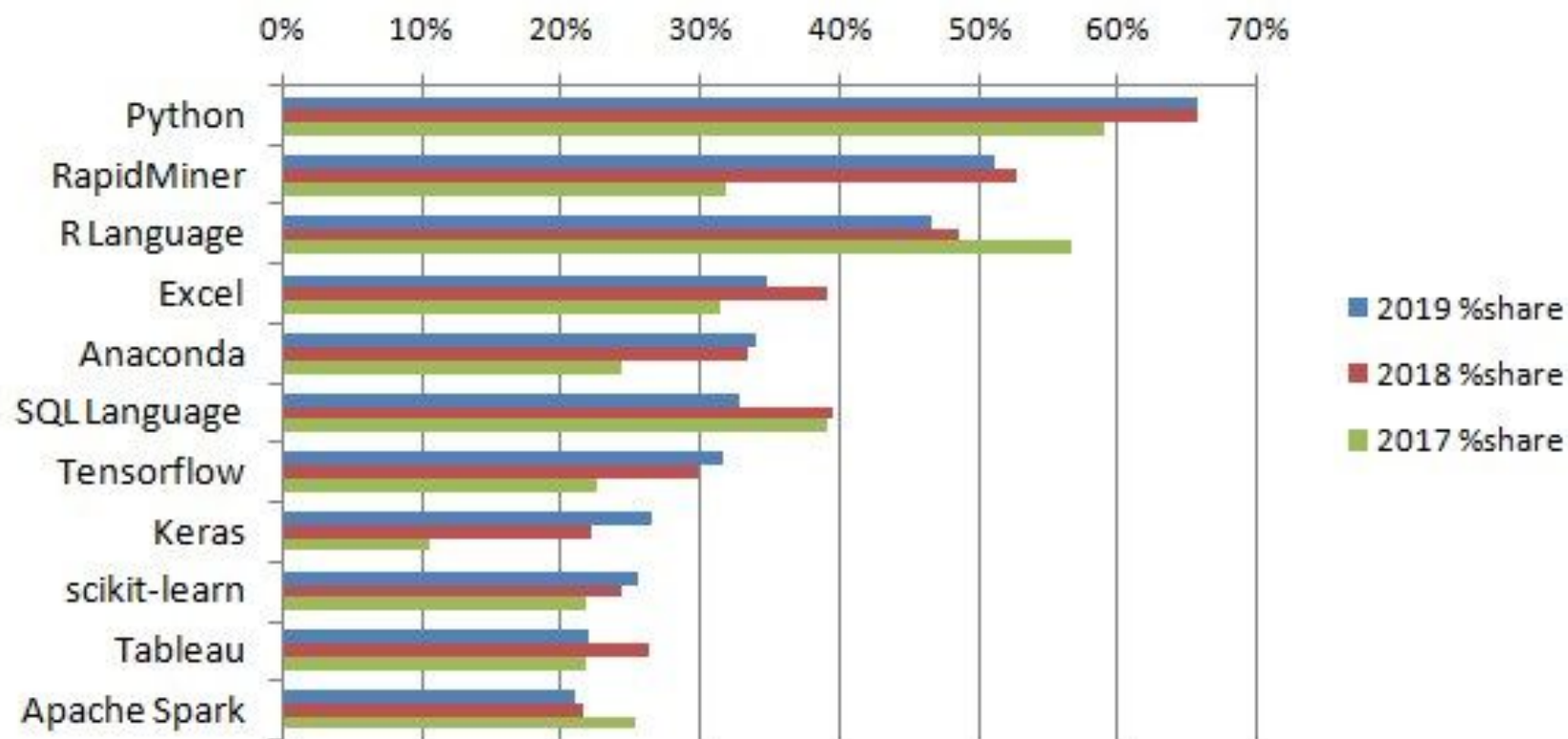
More in
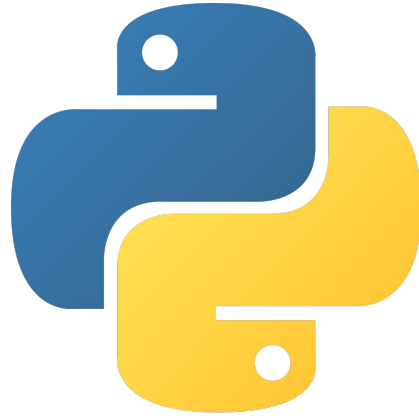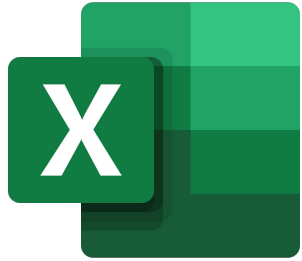https://blog.gojekengineering.com/data-science/home
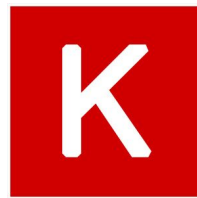
3.
Where to Start

Be a Data Scientist

Top Analytics, Data Science, Machine Learning Software 2017-2019, KDnuggets Poll

Legend:
- 2019 %share
- 2018 %share
- 2017 %share

Categories (top to bottom): Python, RapidMiner, R Language, Excel, Anaconda, SQL Language, Tensorflow, Keras, scikit-learn, Tableau, Apache Spark
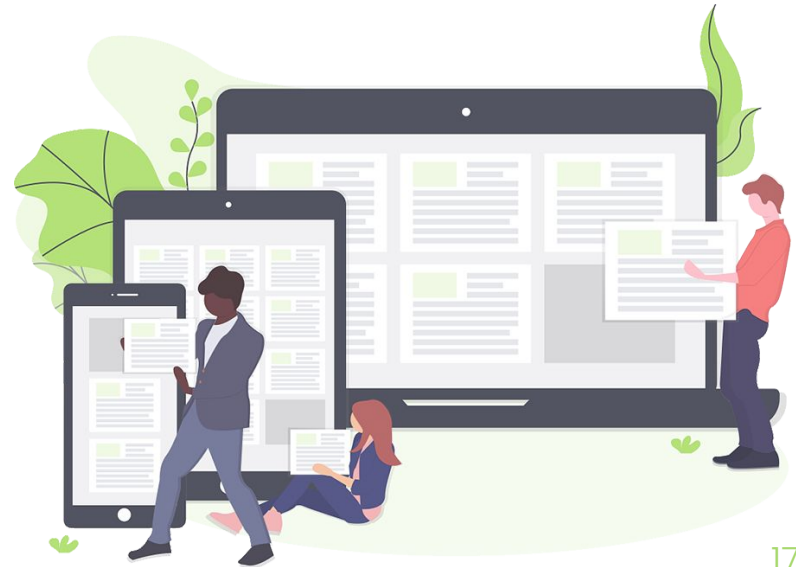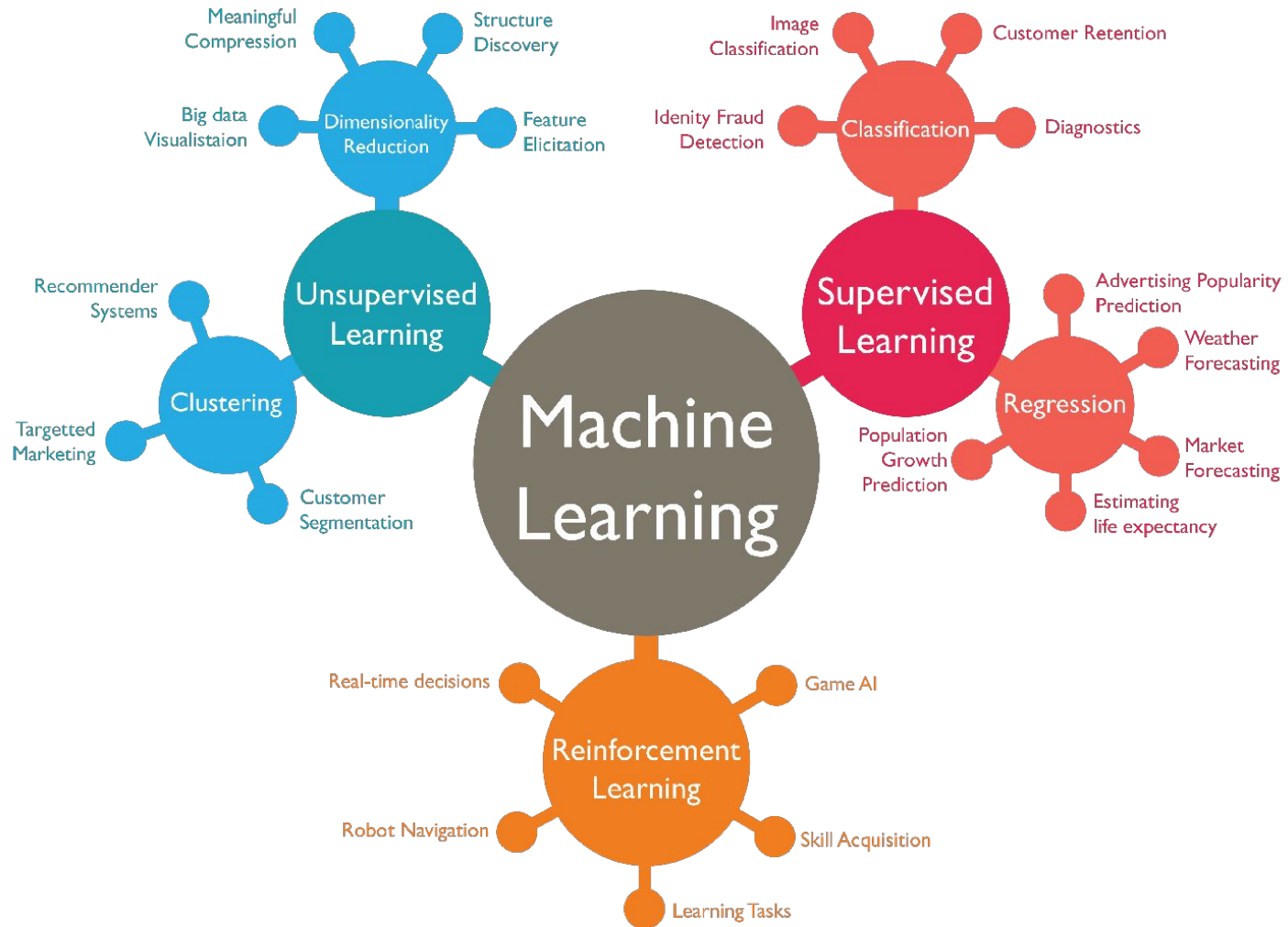
Start with concept

- Understanding concept of Machine Learning is a Must.
- Tools and skills can be taught and learned, but to understand the concept is a basic requirements in industry.

Machine Learning

Unsupervised Learning
- Dimensionality Reduction
  - Meaningful Compression
  - Structure Discovery
  - Big data Visualistaion
  - Feature Elicitation
- Clustering
  - Recommender Systems
  - Targetted Marketing
  - Customer Segmentation

Supervised Learning
- Classification
  - Image Classification
  - Customer Retention
  - Idenity Fraud Detection
  - Diagnostics
- Regression
  - Advertising Popularity Prediction
  - Weather Forecasting
  - Market Forecasting
  - Estimating life expectancy
  - Population Growth Prediction

Reinforcement Learning
- Real-time decisions
- Game AI
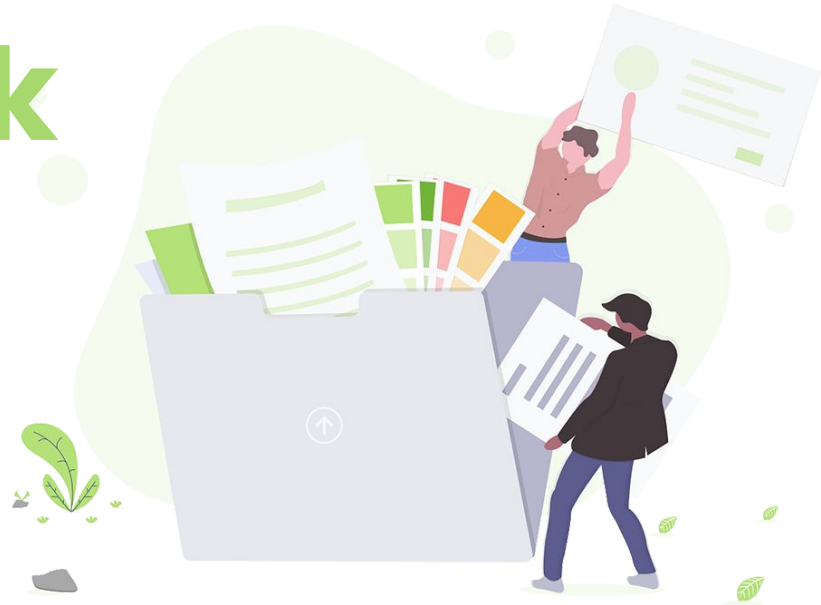- Robot Navigation
- Skill Acquisition
- Learning Tasks

5.
Our Data Stack

Some glimpse

Our Stack (ML)

Usual Problem (ML)

- Fraud analysis (CC, promo fraud, fake bookings, etc) (gopay/goride/gobiz/etc)
- Inappropriate content detection (gofood/gobiz)
- Tagging, searching related to intent (gofood)
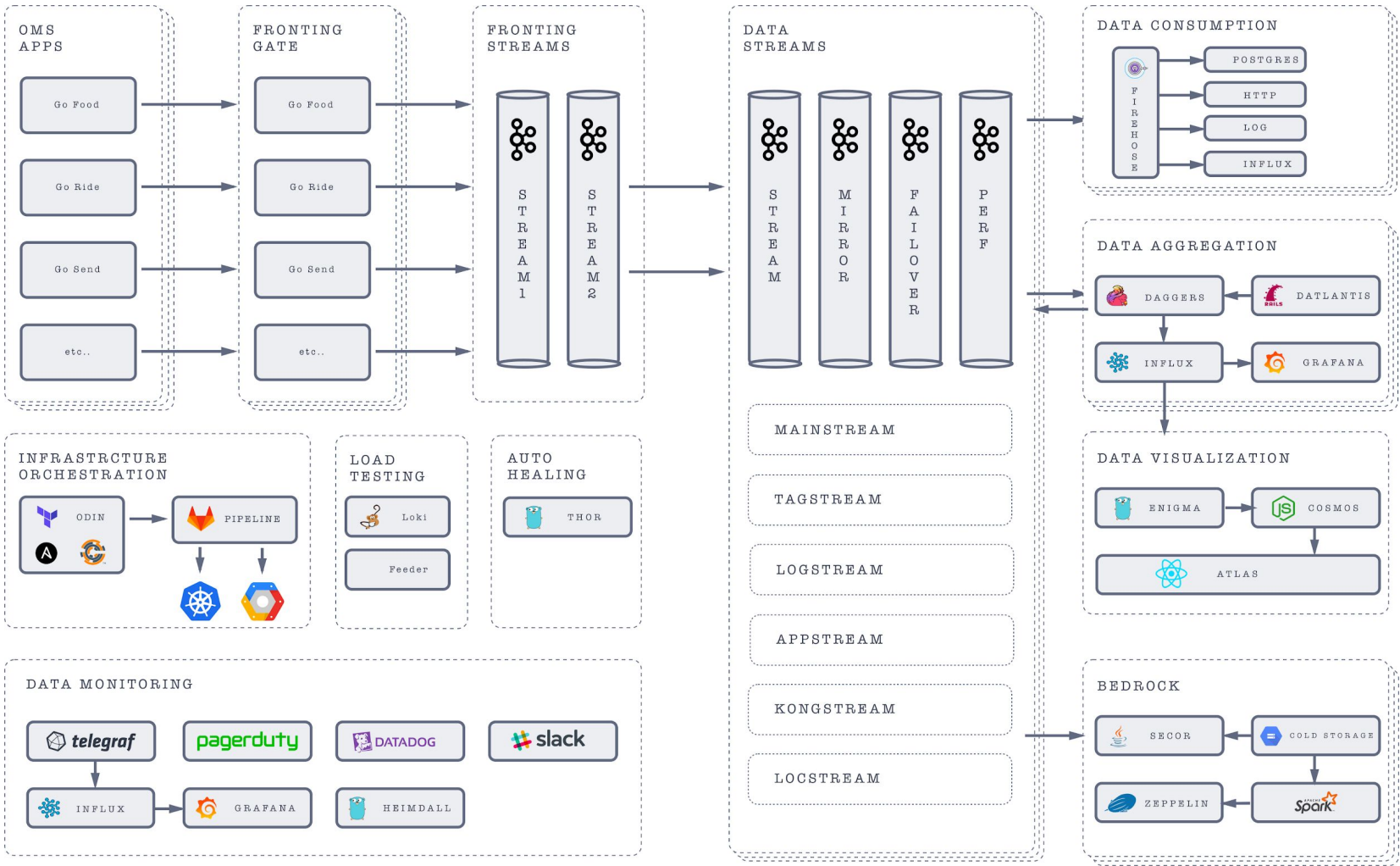- Recommendation system (gofood)

Our Stack (BI)

Usual Problem (BI)

- Executive Management Dashboard
- Monitoring dashboard
- OKR dashboard

Our Stack (DE)

OMS APPS

- Go Food
- Go Ride
- Go Send
- etc..

FRONTING GATE

- Go Food
- Go Ride
- Go Send
- etc..

FRONTING STREAMS

- STREAM 1
- STREAM 2

DATA STREAMS

- STREAM
- MIRROR
- FAILOVER
- PERF

- MAINSTREAM
- TAGSTREAM
- LOGSTREAM
- APPSTREAM
- KONGSTREAM
- LOCSTREAM

DATA CONSUMPTION

FIREHOSE

- POSTGRES
- HTTP
- LOG
- INFLUX

DATA AGGREGATION

- DAGGERS
- DATLANTIS
- INFLUX
- GRAFANA

DATA VISUALIZATION

- ENIGMA
- COSMOS
- ATLAS

INFRASTRCTURE ORCHESTRATION

- ODIN
- PIPELINE

LOAD TESTING

- Loki
- Feeder

AUTO HEALING

- THOR

BEDROCK

- SECOR
- COLD STORAGE
- ZEPPELIN
- SPARK

DATA MONITORING

- telegraf
- pagerduty
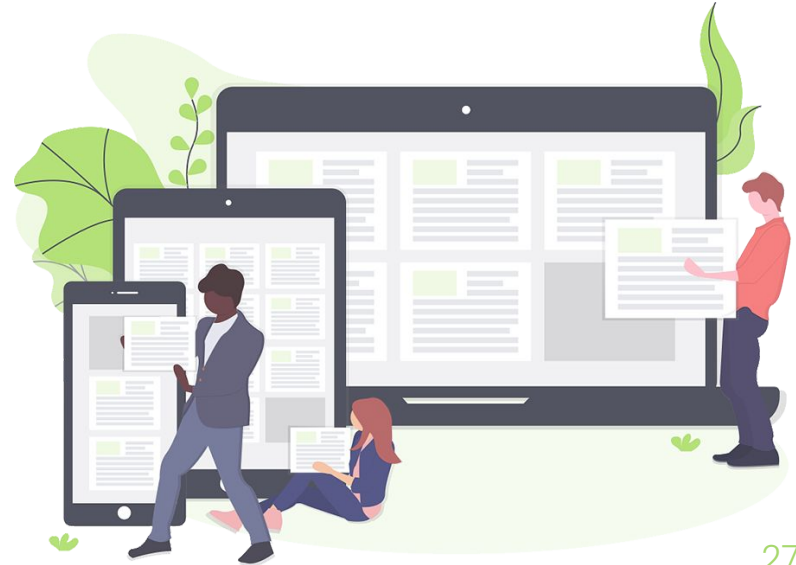- DATADOG
- slack
- INFLUX
- GRAFANA
- HEIMDALL

Usual Problem (DE)

- Data ingestion (from multiple source, kafka, db, queue system, salesforce, clevertap, etc)
- Data warehousing

Training Ground

- https://www.kaggle.com/ (IMHO, have the best resources to start)

Thanks!

We Are Hiring

https://www.gojek.io/careers/

You can find me at:

- bramantyo.erlangga@gmail.com

Credits

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by SlidesCarnival
- Photographs by Unsplash
- Illustrations by Undraw.co