

I. BIG DATA



- 3V (volume, velocity, variety)
- 5V (+ veracity, value)
- 7V (+ variability, visualization)
- 10V (+ volatility, vulnerability, validity)

Welcome to the new vocabulary

Geopbyte*

This will be our digital universe tomorrow...

10^{30}

10^{27}

Brontobyte

A 1BB hard drive would cover the earth 23,000 times

Yotabyte

This is our digital universe today = 250 trillion of DVDs

10^{24}

10^{21}

Zettabyte

1.3 ZB of network traffic by 2016

Exabyte

10^{18}

10^{15}

Petabyte

The CERN Large Hadron Collider generates 1PB per second

Terabyte

10^{12}

10^9

Gigabyte

Megabyte

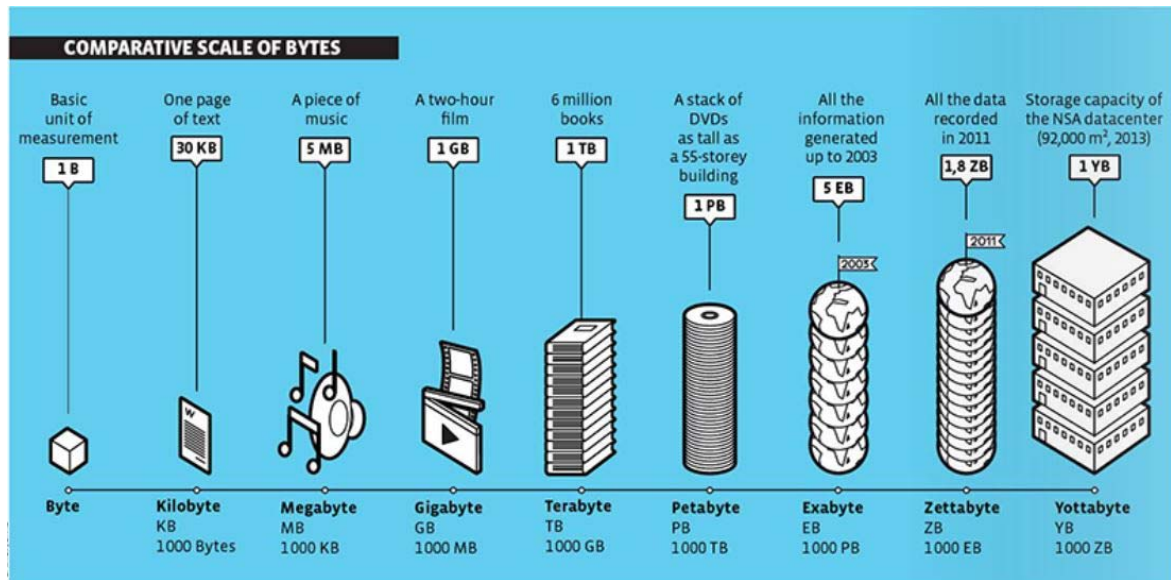
10^6

1 EB of data is created on the internet each day = 250 million DVDs worth of information. The proposed Square Kilometer Array telescope will generated an EB of data per day

500TB of new data per day are ingested in Facebook databases



*The terms Gegobyte and Geobyte are also used in the literature



(waterfordtechnologies.com)

Asal Data

Evolusi Teknologi | IOT | Media Sosial | Lain-lain

blog.pitsco.com



Telephone

Mobile-phone

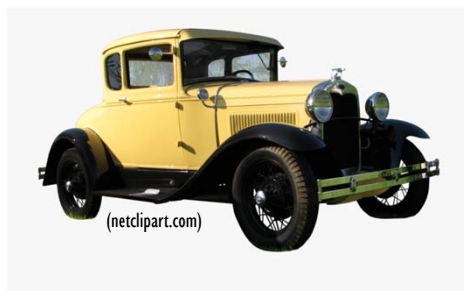
Asal Data

Evolusi Teknologi | IOT | Media Sosial | Lain-lain



Asal Data

Evolusi Teknologi | IOT | Media Sosial | Lain-lain



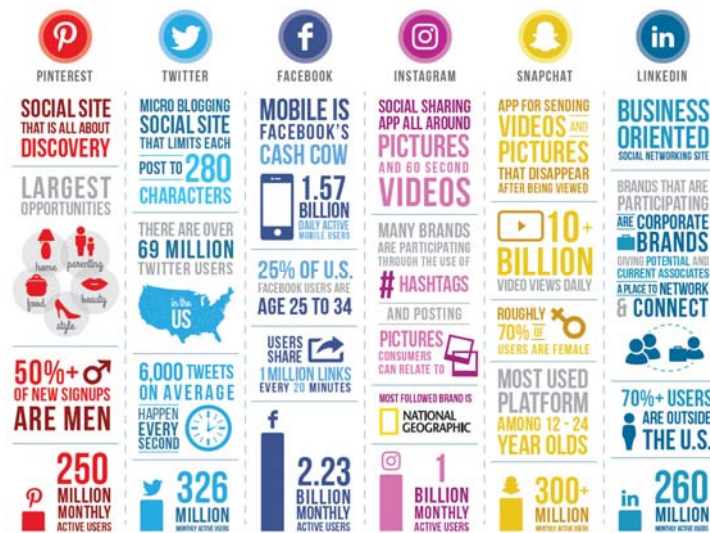
Asal Data

Evolusi Teknologi | IOT | Media Sosial | Lain-lain



Asal Data

Evolusi Teknologi | IOT | Media Sosial | Lain-lain



Asal Data

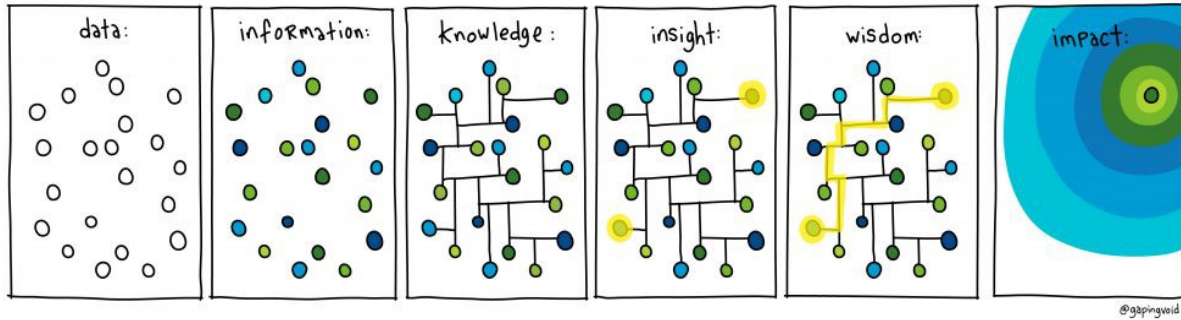
Evolusi Teknologi | IOT | Media Sosial | Lain-lain



M asalah :
D ata O besity!

Too M uch D ata, N ot E nough I n s i g h t s... .

(twitter.com/gapingvoid)



**"IF YOU TORTURE THE DATA LONG ENOUGH,
IT WILL CONFESS TO ANYTHING"**

Ronald Coase, British Economist



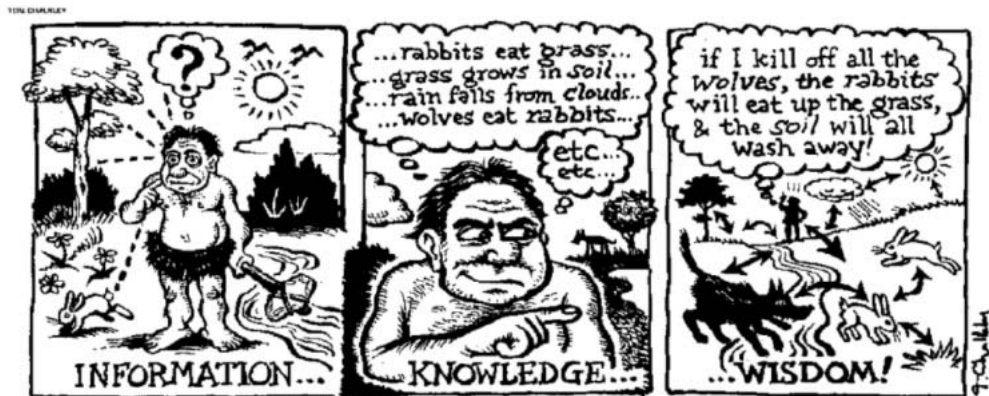
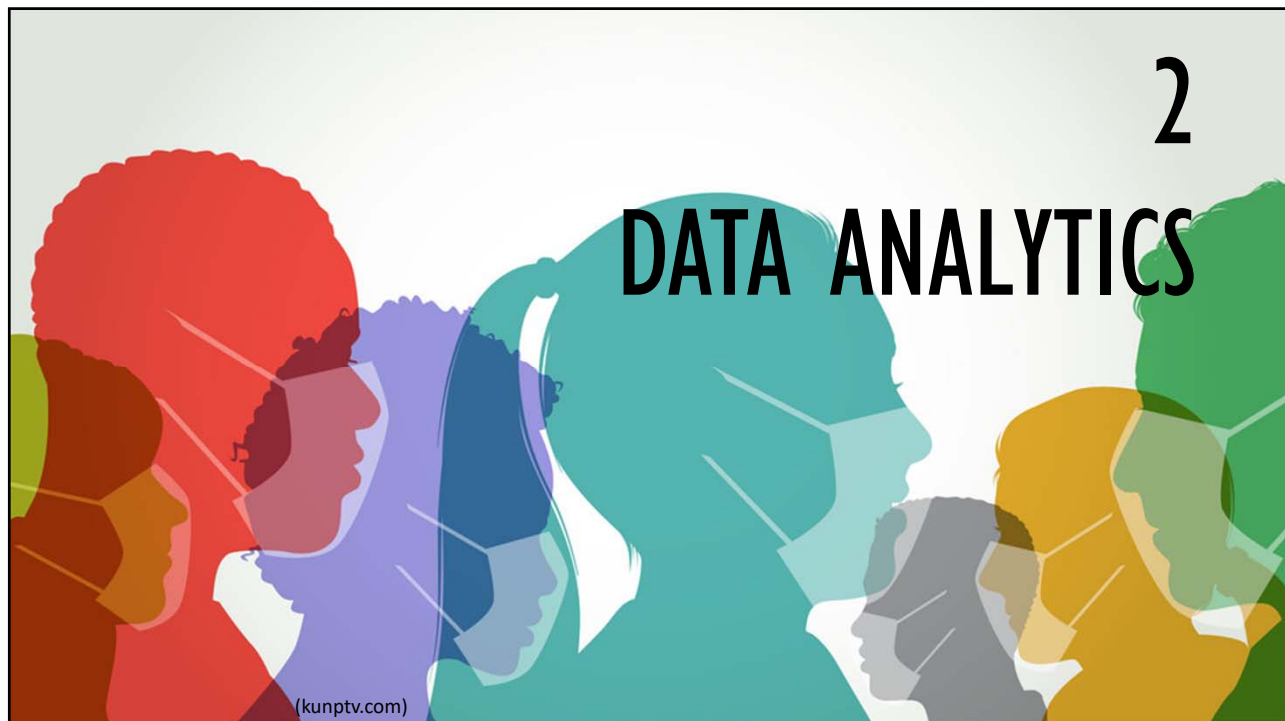
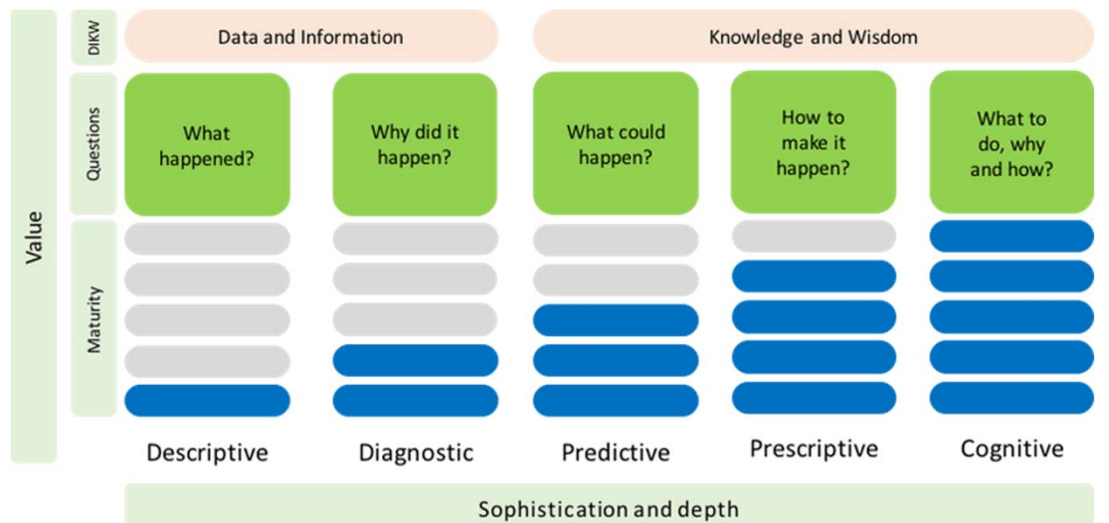


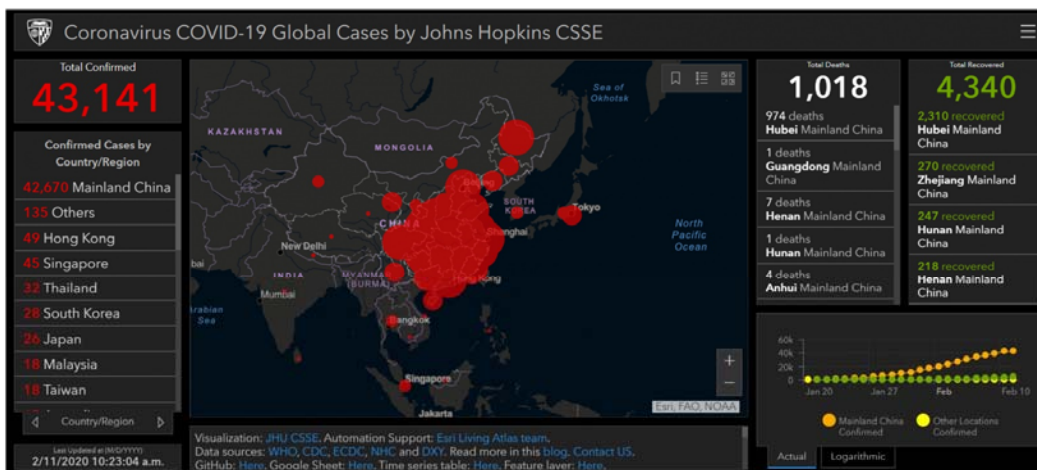
Image originally published in the December 1982 issue of THE FUTURIST. Used with permission from the World Future Society, 7910 Woodmont Avenue, Suite 450, Bethesda, Maryland 20814. Telephone: 301/656-8274; Fax: 301/951-0394; <http://www.wfs.org>



(melsatar.blog/2017/07/30/the-evolution-of-analytics)



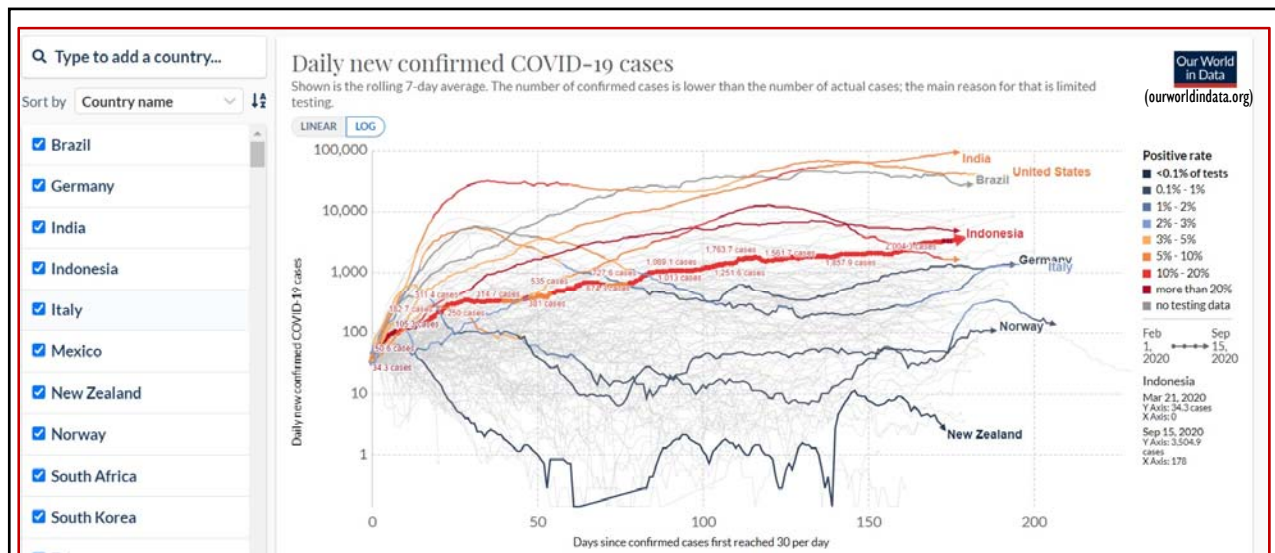
Five Types of Analytics



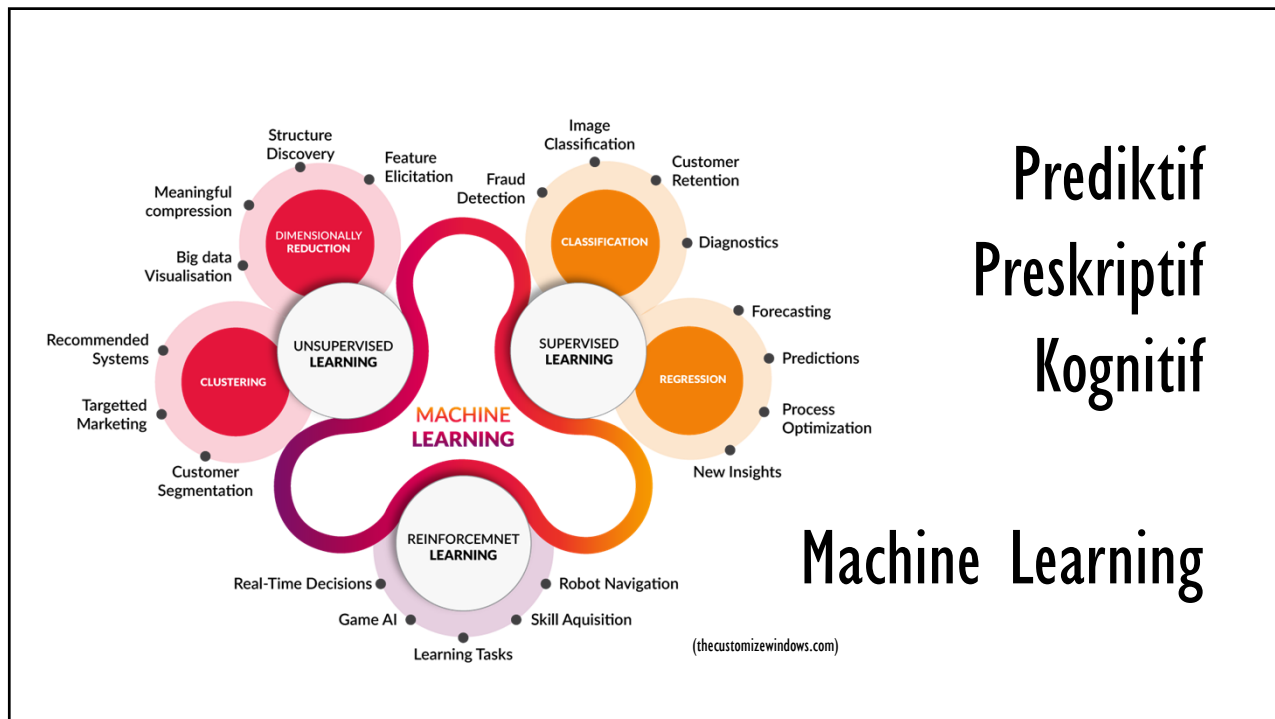
Reference: Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*; published online Feb 19. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1).

<https://systems.jhu.edu/research/public-health/ncov/>

Deskriptif

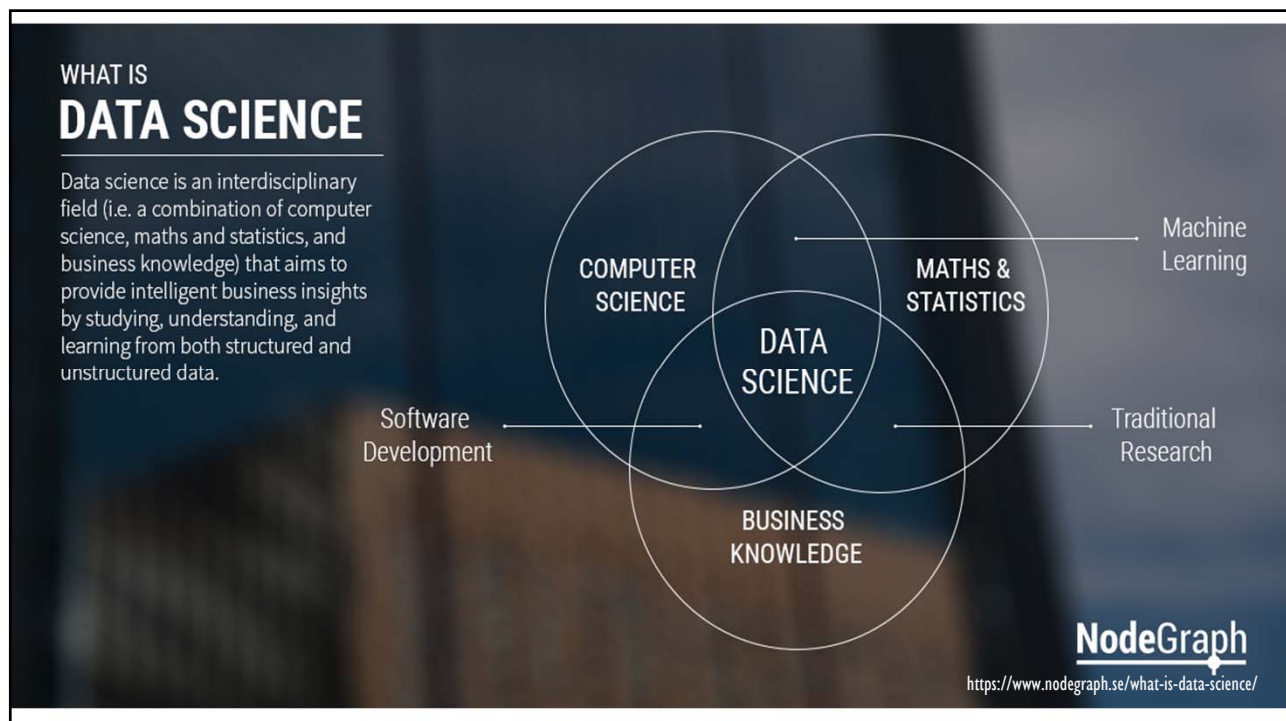
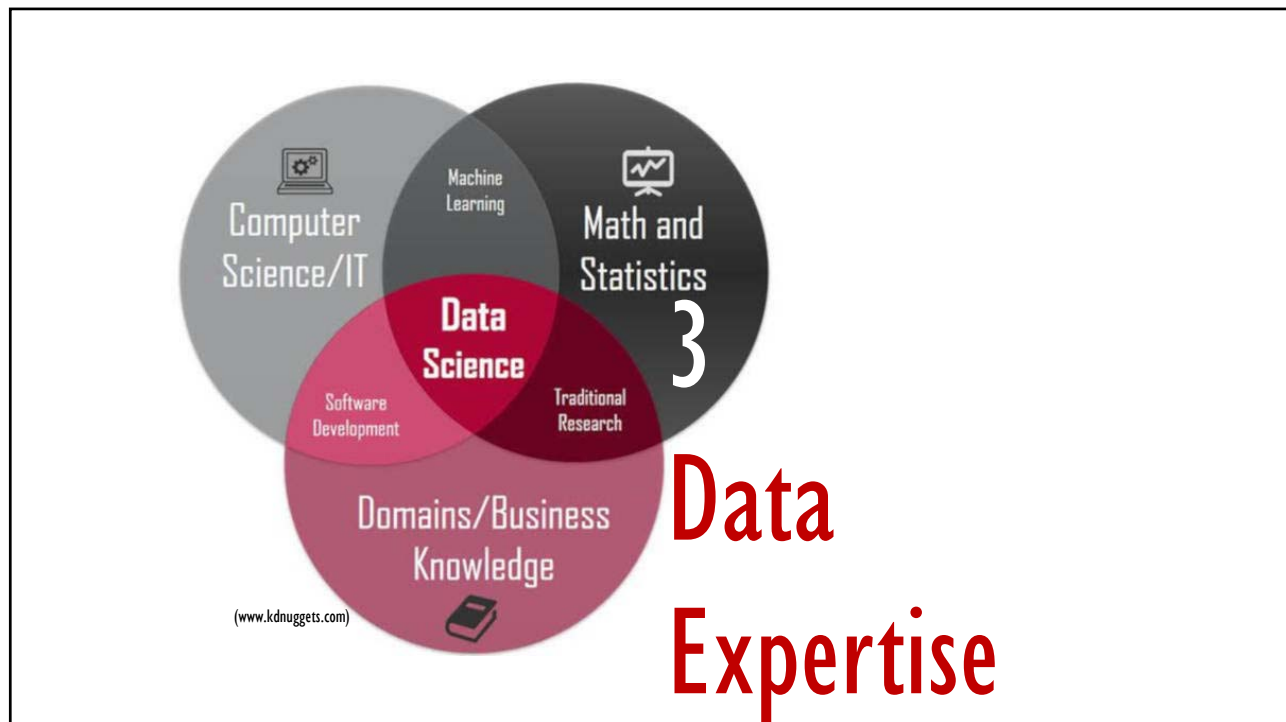


Diagnostik



Prediktif
Preskriptif
Kognitif

Machine Learning



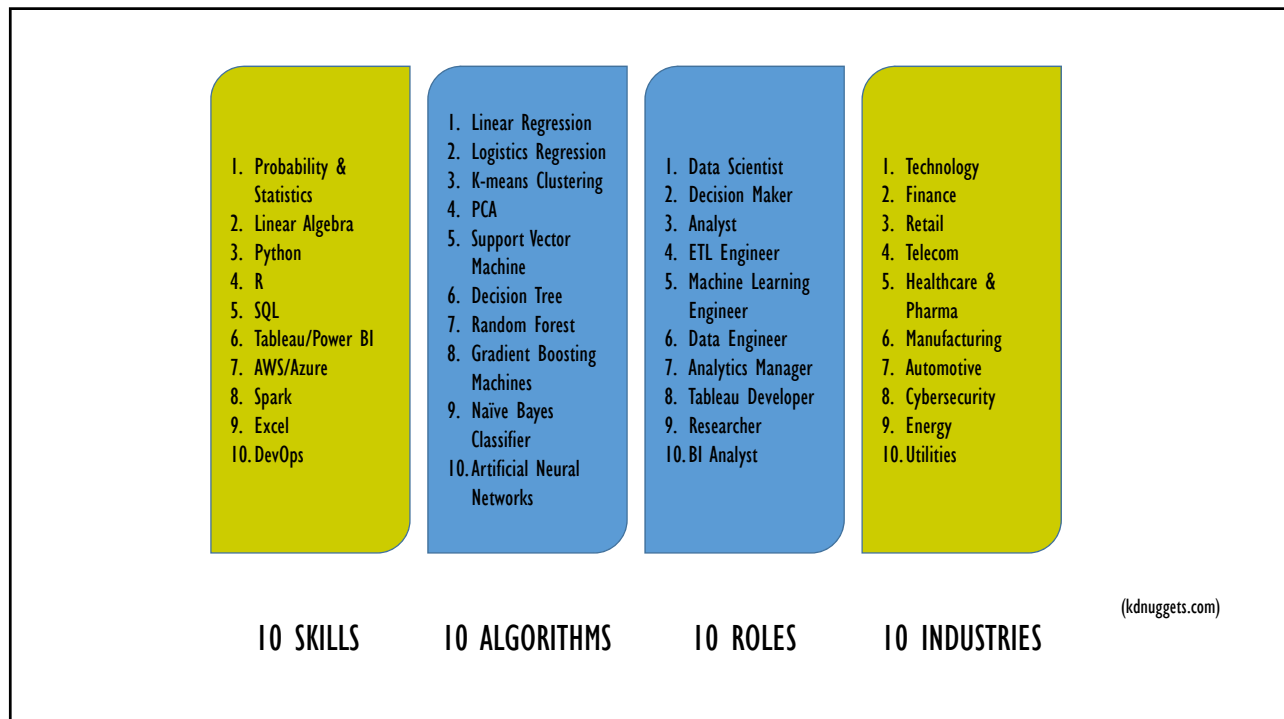
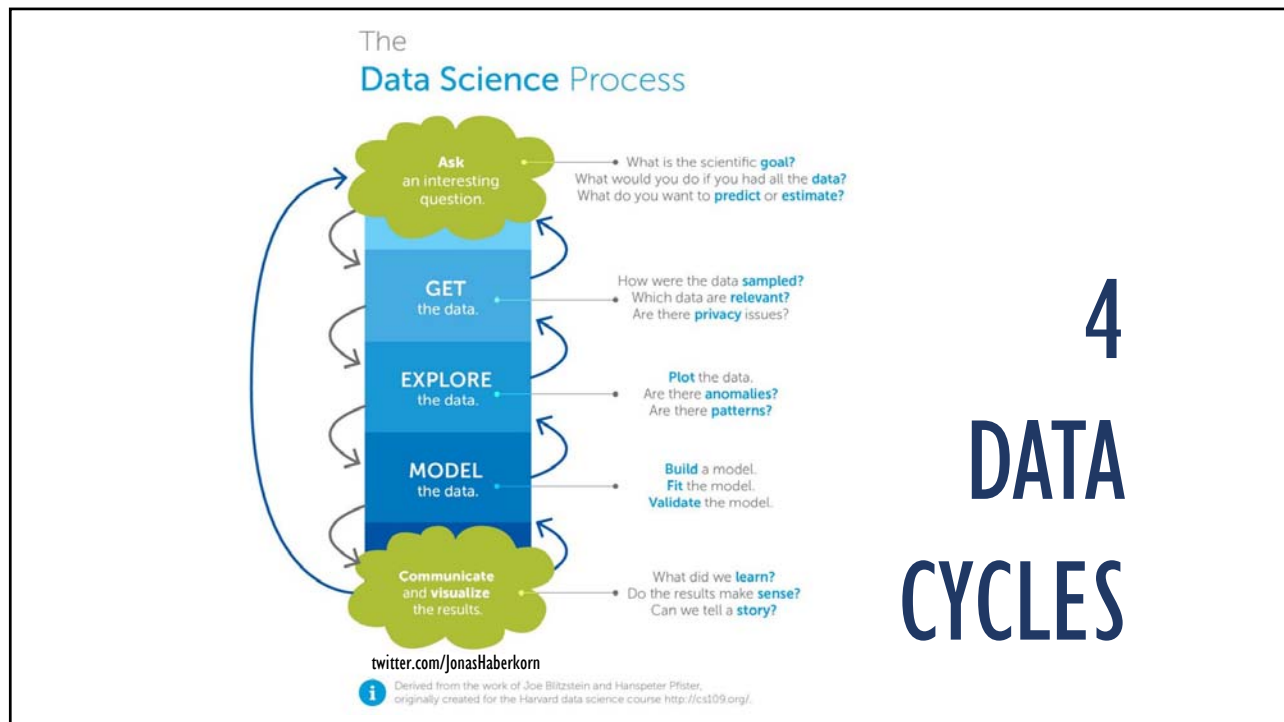
Data Scientist also known as Data Managers, statisticians.	Data Engineers also known as database administrators and data architects.	Data Analysts also known as business Analysts.
A data scientist will be able to take data science projects from end to end. They can help store large amounts of data, create predictive modelling processes and present the findings.	They are versatile generalists who use computer science to help process large datasets. They typically focus on coding, cleaning up data sets, and implementing requests that come from data scientists.	They typically help people from across the company understand specific queries with charts.
Skills: Mathematics, Programming, Communication	Skills: Programming, Mathematics, Big data	Skills: Statistics, Communication, Business knowledge
Will use programmes such as: SQL, Python, R	Will use programmes such as: Hadoop, NoSQL, and Python	Will use programmes such as: Excel, Tableau, SQL

23
(elu.nl)

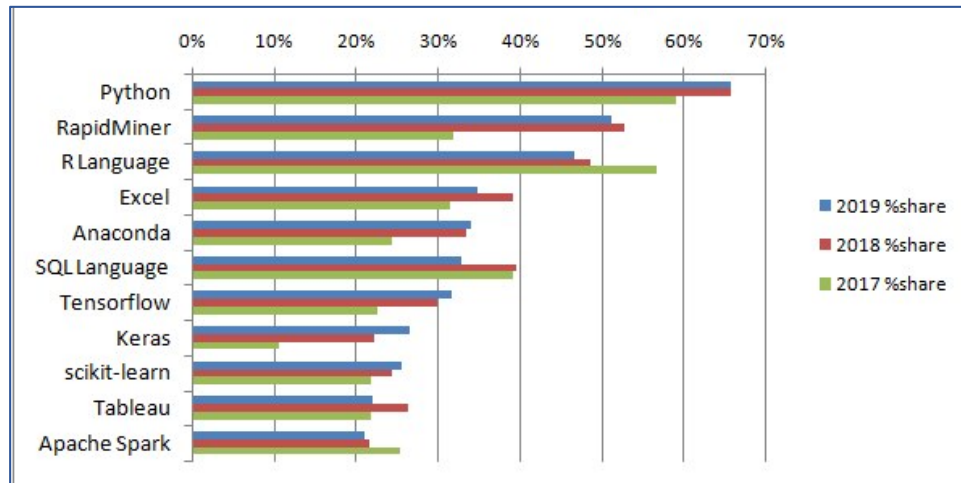
	Data Analyst	Machine Learning Engineer	Data Engineer	Data Scientist
Programming Tools	Very important	Very important	Very important	Very important
Data Visualization and Communication	Very important	Somewhat important	Somewhat important	Very important
Data Intuition	Somewhat important	Very important	Somewhat important	Very important
Statistics	Somewhat important	Very important	Somewhat important	Very important
Data Wrangling	Not that important	Not that important	Very important	Very important
Machine Learning	Not that important	Very important	Not that important	Very important
Software Engineering	Not that important	Somewhat important	Very important	Somewhat important
Multivariable Calculus and Linear Algebra	Not that important	Very important	Not that important	Somewhat important

● Not that important
 ● Somewhat important
 ● Very important

24
<http://blog.hazemshokry.me/>



Contemporary Data Tools



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

(Kdnuggets Poll — kdnuggets.com)

