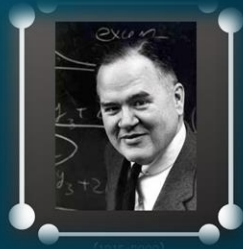


Data Science

Here is where CNC's presentation begins !! ☺
Chrisenia Nadia Carelsz



Sejarah Data Science



Semua berawal dari **John Tukey** yang pada tahun 1962 menulis pergeseran dan perkembangan dunia statistik. Beliau berkata "... *as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt...I have come to feel that my central interest is in data analysis...*".

Tukey menjelaskan bahwa gabungan ilmu statistik dan ilmu komputer dapat mempersingkat waktu penyajian hasil analisis.

Pada tahun 2012 Harvard Business Review dalam artikel "**Data Scientist: The Sexiest Job of the 21st Century**", DJ.Patil meng-claim telah menggunakan istilah ini pada tahun 2008 dengan Jeff Hammerbacher untuk mendefinisikan pekerjaan mereka di **LinkedIn** dan **Facebook**. Dia menambahkan bahwa Data Scientist adalah "**a new breed**", and "**shortage of data scientists is becoming a serious constraint in some sectors**", tapi menjelaskan **role** yang lebih **business-oriented**.



What is Data Science ?

01

Data

is a new currency

02

Data

Kumpulan fakta, seperti angka, kata, pengukuran, pengamatan, atau bahkan deskripsi sesuatu.

03

Science

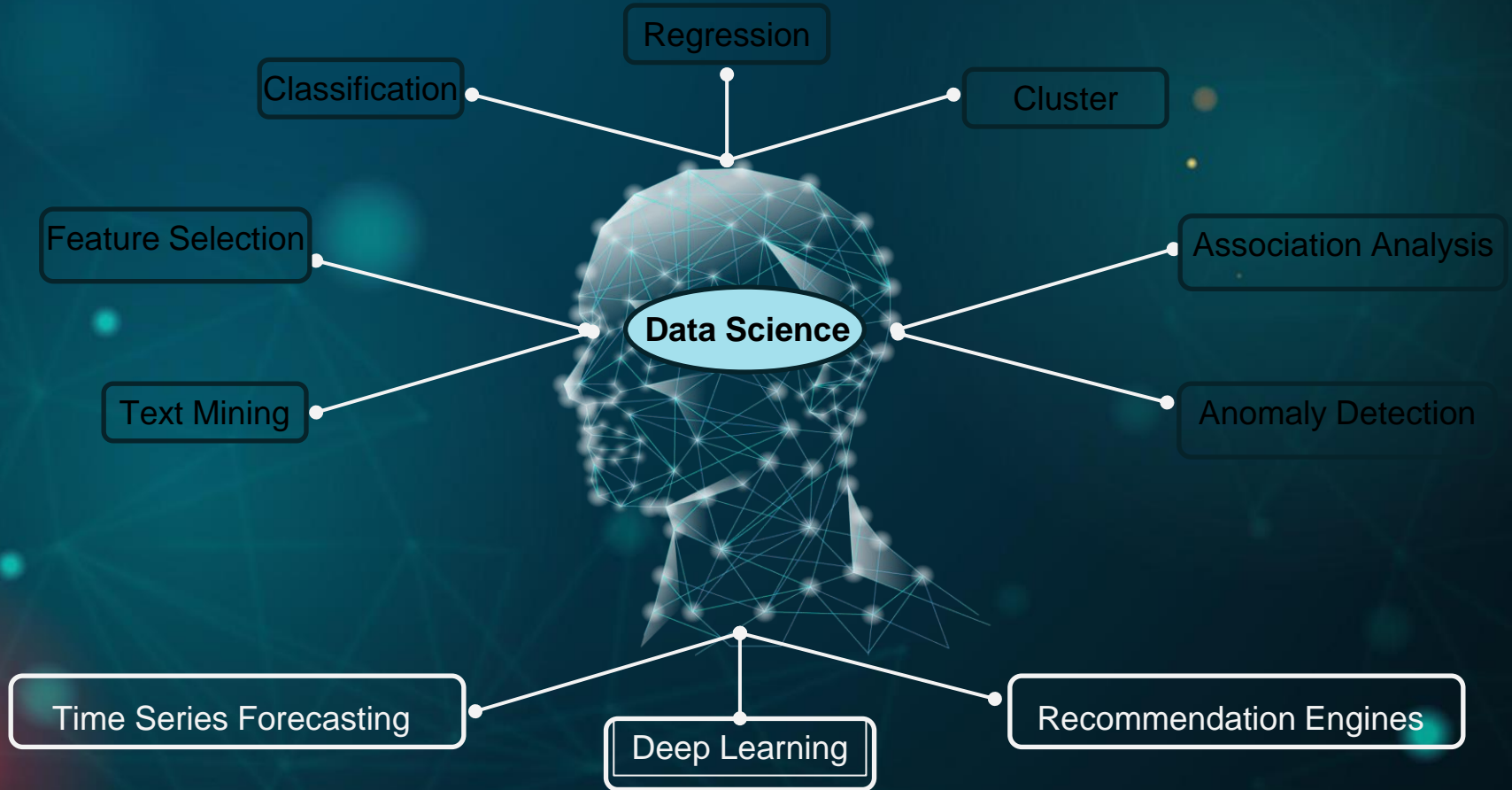
Proses observasi, pengumpulan, dan analisis terhadap fakta-fakta.

04

Data Science

Bidang yang menggunakan metode, proses, algoritma dan sistem ilmiah untuk mengekstraksi pengetahuan dan insights dari data dalam berbagai bentuk, baik terstruktur maupun tidak terstruktur,

Data Science Tasks



Sumber Data

Kaggle

Salah satu situs web untuk Data Science dan Machine Learning yang menyediakan sekitar 6000 dataset dalam format CSV.

Pusat dataset yang menyediakan dataset yang dapat diunduh secara gratis. Terdapat sekitar 400 dataset.

UCI

Machine Learning Repository

Pusat data terbuka milik Pemerintah AS yang terdiri dari berbagai kategori beberapa diantaranya yaitu Pertanian, Konsumen, Ekosistem, Pendidikan, Energi, Keuangan dan Sains.

data.gov

Paket Software Pemrograman Data Science



Pandas

Merupakan sebuah software library Python yang digunakan untuk melakukan manipulasi dan analisis data



SciPy

Merupakan library dasar untuk scientific computing



NumPy

Merupakan Add-On dari Python yang mendukung untuk operasi multidimensional arrays and matrices dalam skala besar



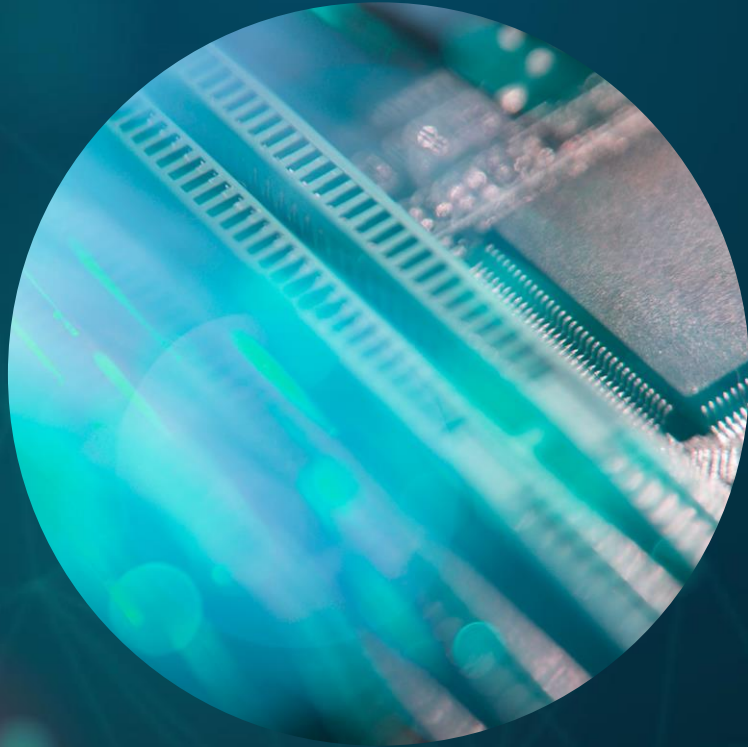
Matplotlib

Merupakan tools yang digunakan untuk membuat visualisasi data 2D.



Seaborn

Merupakan varian pemrograman Python untuk melakukan visualisasi data dengan menggunakan library dari matplotlib.



THANK YOU !!