

# THE EVOLUTION OF DATA SCIENCE

Low Wai Leong  
Lim Jou Hui  
Liu, Hongyang  
Muhammad Amirul Samsul Bahar



# DATA SCIENCE



0101

**BINARY NUMBER SYSTEM** — 1703  
 -Gottfried Leibniz, considered the “first computer scientist”, formalized this system  
 -Laid the groundwork for the field of CS and digital circuitry

2020

TODAY

**ANALYTICS 4.0**

- Quantum analytics
- Automated analytics & better decision
- Machine learning, AI & deep learning change knowledge work
- Chat-bots, meeting assistant, neural machine translation, smart reply etc. was created

**ANALYTICS 1.0****ANALYTICS 2.0****ANALYTICS 3.0**

kaggle

**KAGGLE** — 2010  
 -Kaggle is an online community for data scientists  
 -Offers machine learning competitions  
 -Offers a public data platform

R Studio

**R STUDIO BETA** — 2011  
 -An integrated development environment (IDE) for R

**ANALYTICS 3.0**

-Data enriched-Offerings  
 -Company invested more in analytics to attract viewers  
 -Prescriptive analytics  
 -Better search algorithms, recommendations, and highly targeted ads  
 -Majority of the non-tech firms still suffers due to no knowledge on big data

**QUANTUM COMPUTER**

-Google’s quantum computer achieved quantum supremacy

**BOTTLENECKS**

-Data openness & Data quality  
 -Data privacy

2000

**SMARTPHONE**

-The mobile phone slowly being replaced with 3G smartphone  
**FLASH DRIVE**  
 -Flash drive (pen drives) appeared with storage capacity ranges from 8 to 64GB

**ANALYTICS 2.0**

-Big data & Open Source community is formed  
 -Predictive analytics is coined by SPSS  
 -Structured & unstructured data  
 -New class of database (NoSQL)

**GOOGLE ANALYTICS**

-Used to track website activity

**HADOOP**

-Solve problems involving massive amounts of data  
 -Process both structured & unstructured data

**TWITTER** — 2006

-A social networking service which later becomes an important data source

**TRADITIONAL STATISTICAL METHOD & BIG DATA**

-Traditional method works well with small data sets. Scalability is a huge problem in data mining & developing analytics model

2000S

1990

**THINKPAD TABLET**

-IBM introduced the 1st ThinkPad tablet

**MAGNETO-OPTICAL DISK**

-A hybrid storage Magneto-Optical disk appeared

**ANALYTICS 1.0**

-Business intelligence  
 -Uprising of data warehouse  
 -Huge repositories like eCDW  
 -Descriptive & Diagnostic analytics  
 -Potential capabilities of data only utilized within organization & no explanations or predictions were made

**DATA SCIENCE** — 1997

-The term “Data Science” was coined  
 -Data science is only a concept & hypothesis, which is studied & discussed by a few scholars  
 -No further exploration on the collection, processing & storage of data

1990S

1980

**MOBILE COMPUTER & MOBILE PHONE**

-Mobile computer were introduced, & 1st portable mobile phone appeared in the world

**DATA WAREHOUSE**

-Data warehouse is born  
 -Store current & historical data in a single place that can be used to create analytical reports for separate departments in a company  
 -Very technical & expensive maintenance by expensive IT staff

**COMPUTER TECHNOLOGY & STATISTICS BEGIN TO CONVERGE**

-Data mining is more of a science than a methodology, and it leads to the development of data science, but with very little change

1980S

1970

**MICROPROCESSORS**

-Microprocessors minimize the size of computer  
 -Microcomputer hits the market

**FLOPPY DISK**

Stores 110 kilobytes of data

**RELATIONAL DATABASE**

-The term was invented by E. F. Codd at IBM  
 -This databases allow users to write in SQL  
 -Analyze data on demand & easy to work with  
 -Rigid & unable to translate unstructured data

**BI SYSTEM**

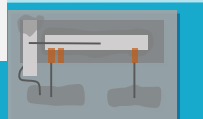
-IBM & Siebel develop the 1st comprehensive BI system (acquired by Oracle)  
 -BI software began giving structure to vast amounts of data

**DATALOLOGY**

-The science of data & data processes and its place in education  
 -There are no tools, no paradigm & no new science to support it

1970S

1960

**COMPUTER CHIP**

-Integrated circuit made of transistors  
 -Computer chips transformed room-sized vacuum tube machines into the micro computers.

**ELECTRONIC CALCULATOR**

-1st electronic calculator to perform work of 50k people by hand  
 -Storage problem & new computers are expensive

**Statistics vs Mathematical Statistics Vs Data science**

-No clear distinction between 3 areas.  
 -The research scope of statistics needed to be dramatically enlarged & redirected.

1960S

1950

**VACUUM TUBE COMPUTER**

-A.k.a 1st gen computer  
 -Widely used until mid 1950s

**TRANSISTOR COMPUTER**

-A.k.a 2nd gen computer  
 -Transistors allow computers to be cheaper, smaller, & more efficient

**TURING TEST**

-A.k.a. Imitation Game” was designed to test if machines can think as humans do  
 - Doors to the field of AI

1950S