Department of Statistics MALAYSIA

For More Info : www.dosm.gov.my
www.mycensus.gov.my

@StatsMalaysia
@MyCensus2020

DOSM/BPIP/2.2020/Series 56

NEWSLETTER

Prepared by:
VERONICA S. JAMILAT
Principal Assistant Director, BPIP

DOSM BIG DATA ANALYTICS INITIATIVES

The Department of Statistics (DOSM) is moving towards Big Data Analytics (BDA) in **producing more reliable and timeliness official statistics**. The objectives of the implementation of STATSBDA are (1) to modernise the data collection; (2) to allow high accuracy; (3) to reduce respondents' burden; (4) to use as supplements for existing data in production of certain statistics; and (5) to produce new statistics indicators. The overall landscape of STATSBDA project is focusing mainly on the project deliverables which are establishing big data infrastructure, development of STATSBDA modules and **building skill sets in data science field.** There are **5 modules** develop in this project which are:

Trade By Enterprise Characteristic (TEC)

Integration of **Malaysia Statistical Business Register** (MSBR) with **trade database** in order to identify the enterprises that are engaged in international markets and what are their characteristics. **Export Imports statistics by states** is a **new statistics generated** from this module has been produced .

Public Maturity Assessment on Official Statistics (PMAOS) & Real Time News on Official Statistics (RTOS)

The **analysis and assessment of the degree of "happiness"** of Malaysia community towards to official statistics published by DOSM through online **social media.**

Price Intelligence (PI)

The modernization of data collection mainly consists of the adoption of web scraping techniques to **scrape price data from related websites for Consumer Price Index (CPI)** compilation and improving the quality of the data.

Statistics Classification (BizCode@Stats)

BizCode@Stats is **Mobile application** development for statistical code and classifications allow users to access the application via handheld devices.

Real Time Business Status (RTBS)

RTBS is part of STATSBDA initiative, will **integrate dat**a from companies **Commission of Malaysia (CCM)** to DOSM environment to enable the organization to have direct access of information of the business entities.

[Source: statsbda.dosm.gov.my]

①

POPULATION & HOUSING CENSUS MALAYSIA 2020
YOUR DATA IS OUR FUTURE

#MyCensus2020
7.7.2020
Make Sure You're COUNTED

HARI STATISTIK NEGARA
20 OCT

2016-2030

Some fundamentals on Big Data

Big Data, Artificial Intelligence, Machine Learning, and Deep Learning have become the latest hot topics and often seem to be used interchangeably. They're behind the revolutionary new business, marketing applications and many more new technology applications.

So, what is the difference between them?
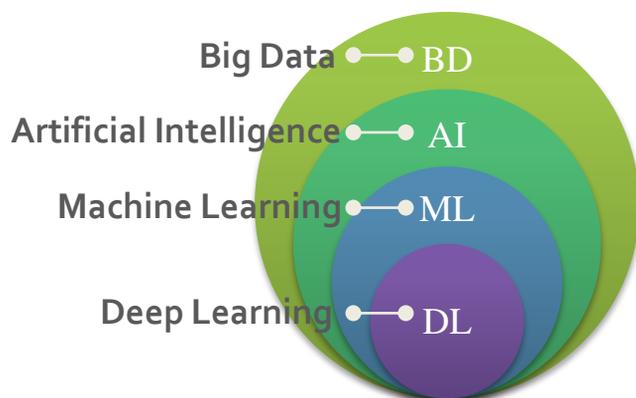
Big Data usually refers to data that **can't be effectively processed with traditional applications** due the challenge of capturing, storing, transferring, querying, and updating data in such large amounts. The **dimensions of Big Data** also known as **5 V's** are: **Volume** (the size of the data), **Variety** (data from different sources or in different formats), **Velocity** (the speed at which data is generated and at which it needs to be available for processing), **Veracity** (trustworthiness of the data sources) and **Value** (usefulness of the data).

Since the Big Data concept refers to the storage and fast processing of large amounts of data, it often uses analytics involving Artificial Intelligence (AI), Machine Learning and Deep Learning. These terms are different and related to each other but often used interchangeably.

The **AI** field can be defined as a **branch of computer science that aims to create intelligent machines**. That is, machines that emulate human performance typically by learning, understanding complex concepts, getting to conclusions or engaging in dialogues with people.

Machine Learning refers to one of the **branches of AI** that is based around the idea that we should really just be able to give. **machines access to data and let them learn for themselves**

Deep Learning refers to a part of **a broader family of Machine Learning methods based on learning data representations focused on Neural Networks**, as opposed to task-specific algorithms. A Neural Network or an Artificial Neural Networks (ANN) are computing systems inspired by the biological networks of nerves and neurons that constitute our human brain. Researchers are looking to apply this concept in many other future applications to more complex tasks such as automatic language translation, medical diagnoses, marketing and numerous other important social and business problems.

Big Data & COVID-19 Pandemic

The new strain of Coronavirus, COVID-19 need no introduction as the virus is spreading around the world and become a worldwide pandemic, causing deaths and major disruption to the global economy. While front liners risk themselves battling the outbreak, other professionals such as scientists, policymakers and researchers in various fields are also taking their part to contain the outbreak.

The COVID-19 pandemic has brought advanced **big data analytics too**ls with entities from all sectors of the healthcare industry seeking to **monitor and reduce the impact of this virus**. Artificial intelligence, machine learning, and deep learning have been increasingly used by scientists, researchers and developers to track and contain COVID-19.
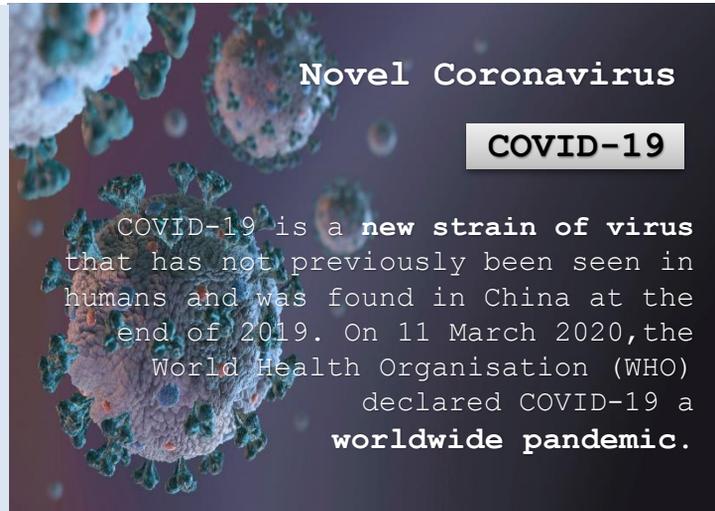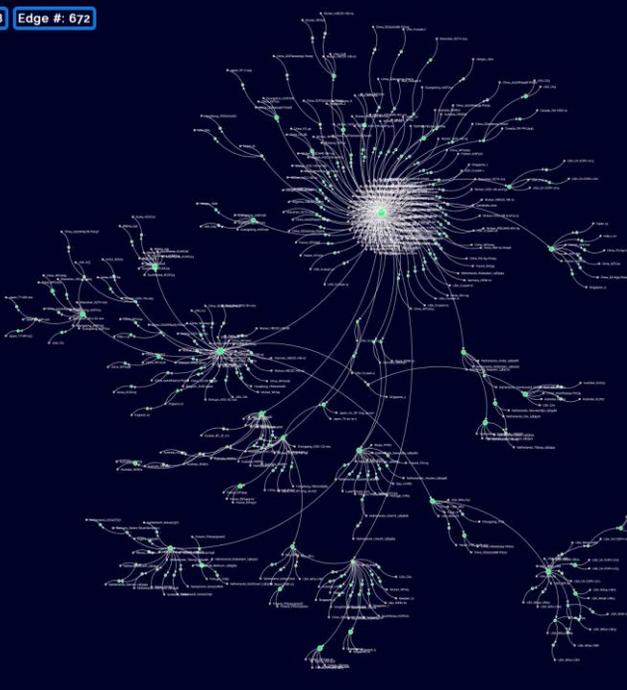
Novel Coronavirus

COVID-19

COVID-19 is a **new strain of virus** that has not previously been seen in humans and was found in China at the end of 2019. On 11 March 2020, the World Health Organisation (WHO) declared COVID-19 a **worldwide pandemic.**

Some **big data components and advanced big data analytics tools** used for COVID-19 research are:

1) **Biomedical research** – Artificial Intelligence can eliminate many false tracks and allow us to identify potential targets.
2) **Natural language processing** – Machine learning is able to cater multi-factor learning on how people are bearing the burdens and stresses of the pandemic .
3) **Virology research** – Deep learning is applied to gain a more comprehensive understanding of the virus.

Evolution of a virus [Source: graphen.ai]

③

Malaysia coping COVID-19 using Big Data Analytics

The COVID-19 pandemic hit hard almost all countries across the world. Started in Wuhan, China and now the US and European countries recorded the highest cases and seems not slowing down at the moment. Malaysia on the other hand, showing an acceleration in the beginning of the second wave, and slowly recorded a stable cases for the past 2 weeks. According to Ministry of Health Malaysia (MOH), the implementation of Movement Control Order (MCO) has shown to flatten the curve. In order to cope with this outbreak, Selangor state which recorded highest number of COVID-19 case in Malaysia, initiated a **dedicated task force** in battling the outbreak.

The **use of Big Data and Machine Learning** would then allow the state to conduct communal screenings and contact tracing in a more efficient and organised manner. The government of Malaysia take a step ahead by launching an application called **MySejahtera** developed to assist in monitoring this outbreak in the country by enabling users to assess their health risk against COVID-19. Data captured using this app enable the government to conduct complex research using Big Data Analytics to assist the government to contain this pandemic in this country.

Assess COVID-19 health risk for you and your family

Get COVID-19 health guidelines

Get guidance on actions to be taken based on COVID-19 risk category

Get information about the nearest health facilities

Receive the latest information and guideline from a trusted source

HELP US TO HELP YOU

KEEP YOU AND YOUR FAMILY SAFE FROM COVID-19

Strategic Collaboration: NSC, MOH, MAMPU & MCMC

[Source: mysejahtera.malaysia.gov.my]

POPULATION & HOUSING CENSUS MALAYSIA 2020
YOUR DATA IS OUR FUTURE

#MyCensus2020
7.7.2020

HARI STATISTIK NEGARA
20 OCT 2016-2030