Knowledge Sharing: What's So Exciting About Data Science?

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

DR NILAM NUR BINTI AMIR SJARIF
UNIVERSITI TEKNOLOGI MALAYSIA

innovative ● entrepreneurial ● global

©Copyright@nilamnur20222023

2

UTM

Have you heard about Data Science before?

Voting code **8513 4695**

https://www.menti.com/dzo6wetr6y

innovative ● entrepreneurial ● global

©Copyright@nilamnur20222023

3

Is Data Science a Good Career Choice in 2022 & Beyond?

YES NO

innovative ● entrepreneurial ● global

©Copyright@nilamnur20222023

Source:medium.com

4

4

Data Science for Everyone - Data Science is a 21st Century Job Skill for Every Discipline

innovative ● entrepreneurial ● global

©Copyright@nilamnur20222023

5

5

DATA SCIENTIST
THE SEXIEST JOB OF THE 21st CENTURY

— *Harvard Business Review*

Source: successatschool.org

©Copyright@nilamnur20222023
6

6

Prediction

By 2026, 27.9 % = 11.5 million data science jobs will be created globally in the field.

Growth

Market : **$37.9 billion** in 2019 to **$230.80 billion** by 2026

Data science jobs are a top pick for Gen Z

*"The most promising career" by **LinkedIn** and the "best job" by **Glassdoor***

©Copyright@nilamnur20222023
7

7

3

UTM

HOME   ABOUT US   LEADERSHIP TEAM   RESEARCH EXPERTS   BRIEFINGS   CAREERS   CONTACT US

KNOWLEDGE STORE   REPORTS   RESEARCH INSIGHT   RESOURCE CENTER   EVENTS   360 QUADRANTS

The Data Science Platform market size is projected to grow from USD 95.3 billion in 2021 to 322.9 USD billion in 2026, at a Compound Annual Growth Rate (CAGR) of 27.7%

Major effects from COVID 19 on Global Economy: directly impacting production and demand, causing supply chain and market disruption

Data Science Platform software are adopted for Banking, Financial Services, and Insurance (BFSI) , Retail and eCommerce, Telecom and IT, Media and Entertainment, Healthcare and Life Sciences, Government and Defense, Manufacturing, Transportation and Logistics, Energy and Utilities, Other (travel and hospitality, and education and research).

[349 Pages Report] The Data Science Platform market size is projected to grow from USD 95.3 billion in 2021 to 322.9 USD billion in 2026, at a Compound

Major Data Science Platform key players include IBM(US), Google(US), Microsoft(US), SAS(US), AWS(US), MathWorks (US), Cloudera (US), Teradata (US), TIBCO (US), Alteryx (US), RapidMiner (US), Databricks (US), Snowflake (US), H2O.ai (US), Altair (US), Anaconda (US), SAP (US),

Source: marketsandmarkets.com

innovative ● entrepreneurial ● global

8

8

UTM

Is Data Science Demand in Indonesia??

innovative ● entrepreneurial ● global

9

9
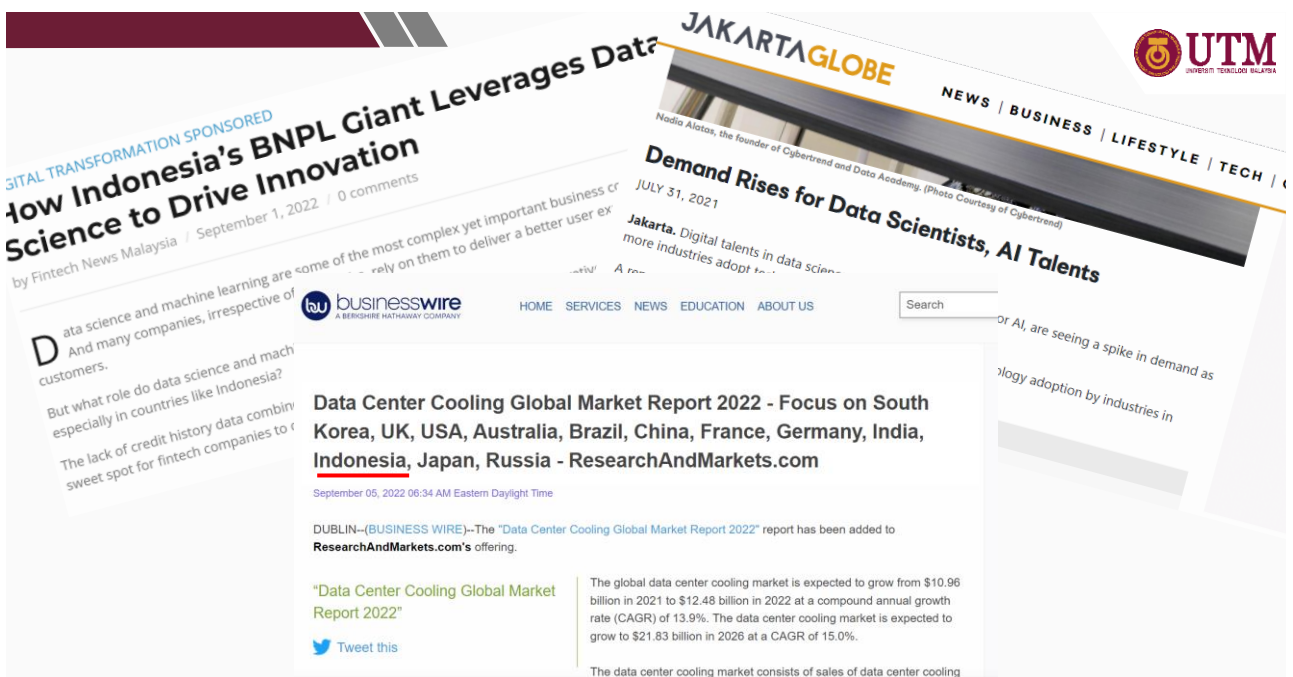
Source: datamites.com

innovative ● entrepreneurial ● global   ©Copyright@nilamnur20222023   10

10

innovative ● entrepreneurial ● global   ©Copyright@nilamnur20222023   11

11

Data Scientist average salary change by experience in Indonesia

0 - 2 Years		12,100,000 IDR
2 - 5 Years	▲ +26%	15,300,000 IDR
5 - 10 Years	▲ +32%	20,200,000 IDR
10 - 15 Years	▲ +18%	23,800,000 IDR
15 - 20 Years	▲ +11%	26,300,000 IDR
20+ Years	▲ +6%	28,000,000 IDR

Percentage increase and decrease are relative to the previous value

Source: salaryexplorer.com

innovative ● entrepreneurial ● global   ©Copyright@nilamnur20222023   13

Data Scientist average salary difference by education level in Indonesia

Bachelor's Degree		14,600,000 IDR
Master's Degree	▲ +33%	19,500,000 IDR
PhD	▲ +42%	27,700,000 IDR

Percentage increase and decrease are relative to the previous value

Source: salaryexplorer.com

BL Bukalapak
4.5 ★
Data Scientist ⓘ
See 18 salaries from this location

4 open jobs

IDR 15,506,870 / mo
IDR 11M   IDR 18M

Tokopedia
4.5 ★
Data Scientist ⓘ
See 17 salaries from this location

194 open jobs

IDR 11,000,000 / mo
IDR 10M   IDR 17M

Telkom Indonesia
4.3 ★
Data Scientist ⓘ
See 11 salaries from this location

IDR 14,000,000 / mo
IDR 12M   IDR 25M

Astra International
4.3 ★
Data Scientist ⓘ
See 10 salaries from this location

2 open jobs

IDR 9,874,628 / mo
IDR 9M   IDR 12M

Telkomsel
4.5 ★
Data Scientist ⓘ
See 8 salaries from this location

IDR 26,567,737 / mo
IDR 21M   IDR 33M

Can I Do Data Science without Computer Science?

"Data science doesn't care about what you majored in or if you even got a degree. It's what you do with data that matters." – DJ Patil (Chief data scientist of the United States)

16

How To Become a Data Scientist Without a Degree

Perfect Your Prerequisites

Refine Your Data Science Skills

Practice With Data Science Projects

Practice Your Interview Skills

DATA SCIENCE

Build a Portfolio

PORTFOLIO

innovative ● entrepreneurial ● global

17

17

Data Science Career Path
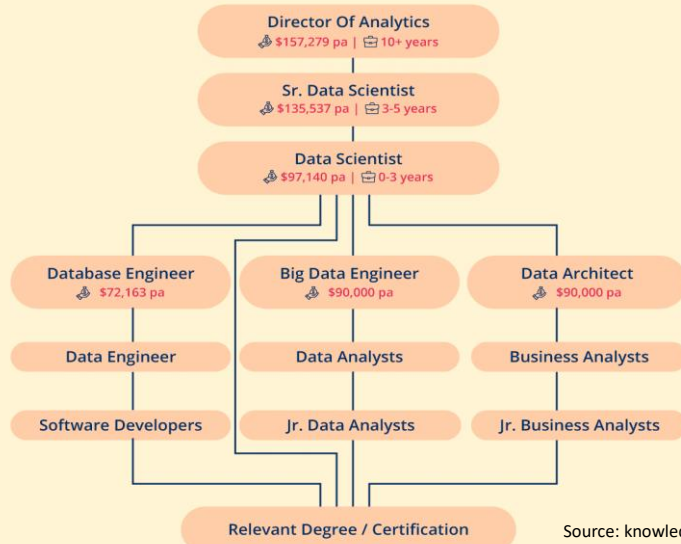
innovative ● entrepreneurial ● global

©Copyright@nilamnur20222023

18

18

What is Data Science?

innovative ● entrepreneurial ● global

©Copyright@nilamnur20222023

19

19

What is Data Science?

Source: techtarget.com
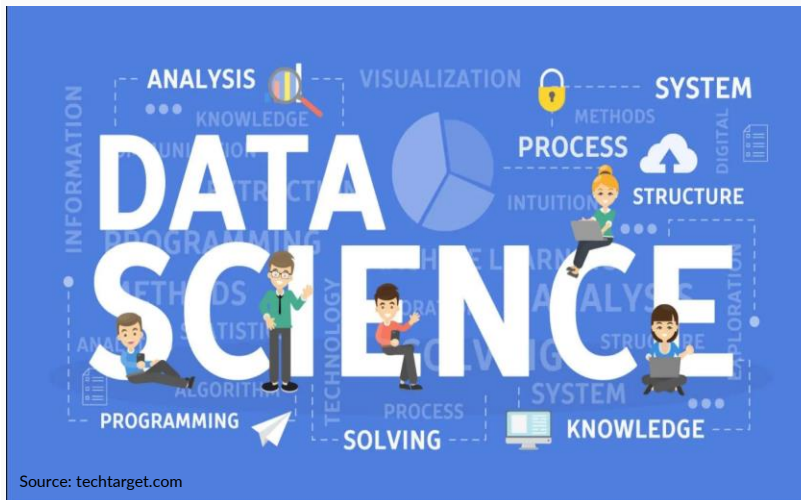
Defined as an interdisciplinary field of study that uses data for various research and reporting purposes to derive insights and meaning out of that data.

innovative ● entrepreneurial ● global   ©Copyright@nilamnur20222023   20

20

Data Science Terms'
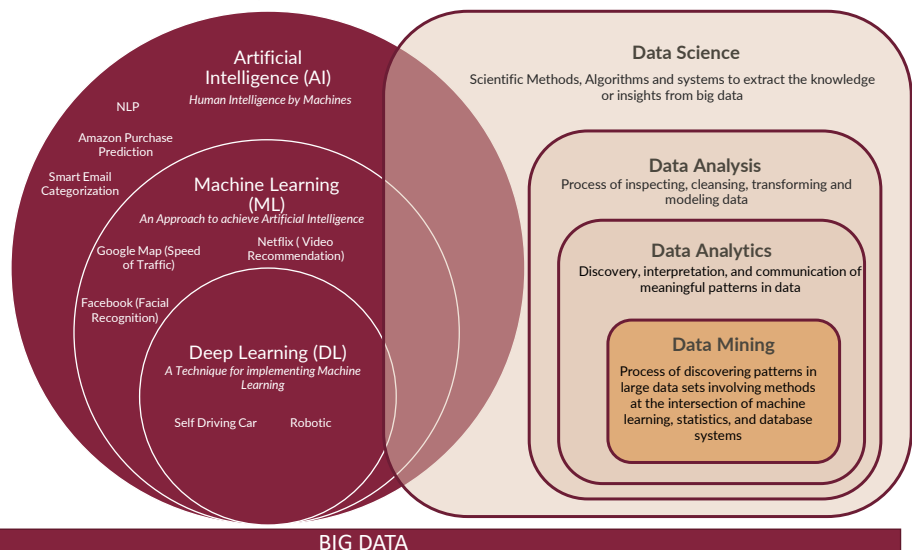
innovative ● entrepreneurial ● global   ©Copyright@nilamnur20222023   21

21

10

"As more and more organizations attempt using data for the first time, there will be plenty of new opportunities that only a person trained in data science will be able to leverage." – Prof Rafael Irizarry, Applied Statistics at Harvard University

22

What is Data?

The world today produces **2.5 quintillion bytes** of data

» **306.4 billion emails**

» **5 million Tweets**

» **95 million videos and photos sharing in social media**

According to **recent estimates**, by 2020, our digital ecosystem will be approximately produces **44 zettabytes of data**, projected to reach 180 zettabytes by 2025.

The business demand for **DATA, INFORMATION** and **ANALYTICS**

Too Much Data, Too Little Information

Data refers to a collection of facts usually obtained as the result of experiences, observations, or experiments

23

11
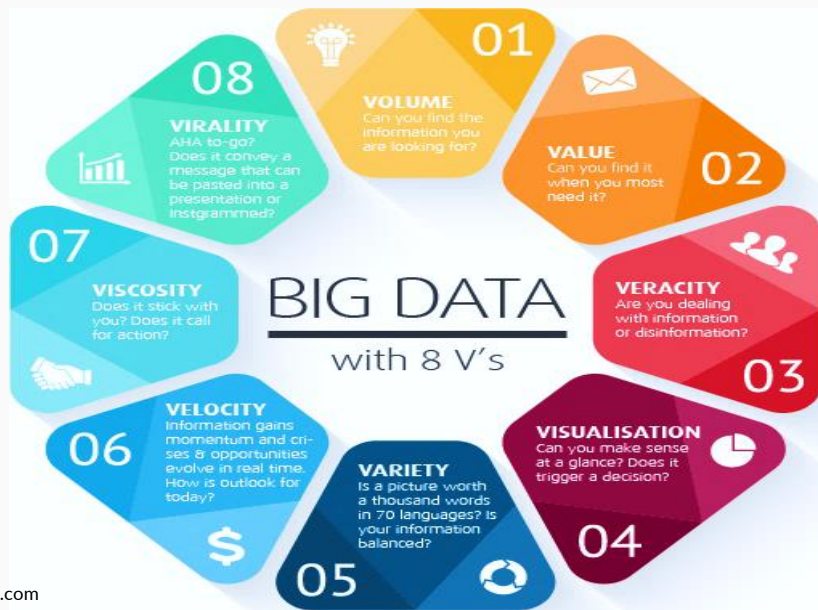
Source: zarantec.com

innovative ● entrepreneurial ● global

24

24

Data is the new Oil

"Data has to be captured, analyzed and acted upon"

Source: Clive Humby
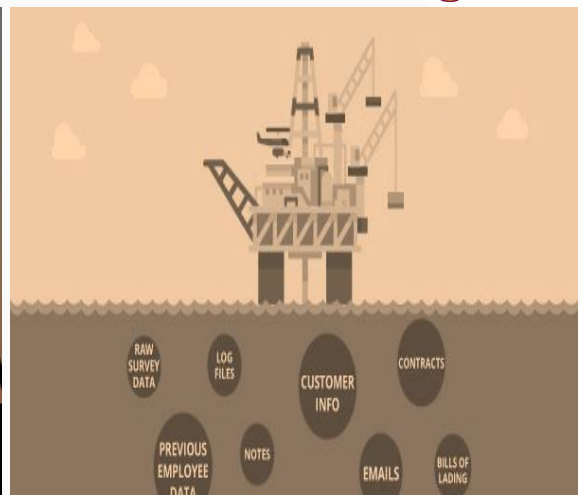
"If the 'crude' data can be extracted, refined and pipe to where it can impact decision in real time, its values with soar and if data can properly shared across an entire ecosystems and made accessible in the places where analytics are most useful, then it will become a true charger, altering the way we live, work, learn and play"

innovative ● entrepreneurial ● global

25

25
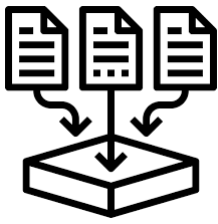
Data
Raw, random and unorganized

Information
Data that has been organized, structured and processed. And is what you used to gain knowledge

Knowledge
what we know, accumulation of past experience and insight. which interpret, and assign meaning to, produce actionable information that can lead to impact

innovative ● entrepreneurial ● global

©Copyright@nilamnur20222023

26

26

The Concept of Data, Information and Knowledge : Scenario

Let's take a look at how a foodie would approach the concept

Do you have any idea how you can differentiate between data, information and knowledge for this soup??

innovative ● entrepreneurial ● global

©Copyright@nilamnur20222023

27

13

Data: A collection of ingredients sitting on the counter. They include carrots, onions, leeks, garlic, and potatoes from the farmer's market, and a package of chicken, a box of rice and some cans of broth from grocery store. In a data science concept, this is like a source of data from different operational systems.

Information: Then you get everything ready by washing, peeling, and cutting up the vegetables, cutting up the chicken and opening the cans of broth. You put it all in the pot and turn on the heat where it cooks and become soup. In data science world, the data has been move to the data science life cycle steps including clean a missing data, data aggregation, data extraction and data transformation, modeling, evaluation etc. And it is transform into information

Knowledge: Now the soup is ready to put into a bowl and eaten. In data science, business people/stakeholder consume the information in reports to gain knowledge to help them make informed business decisions

innovative ● entrepreneurial

©Copyright@nilamnur20222023

28

SKILLS FOR DATA SCIENCE

Skills You'll Need!

Entry Level Skills

Python | R | SQL
Tableau | Advanced Excel
Statistics | Data Visualisation | SAS

On-the Job Skills

Machine Learning & AI | Deep Learning
Data Mining | Data Wrangling | Hadoop
Apache Spark | Predictive Modelling | NLP

Soft Skills

Communication Skills | Curiosity | Collaboration | Story Telling
Project Management | Business Acumen

Source: knowledgehut.com

innovative ● entrepreneurial ● global

©Copyright@nilamnur20222023

29

29

What is Data Science Life Cycle?

A data science life cycle is **an iterative set of data science steps** you take to deliver a project or analysis.

Source: sudeep.co
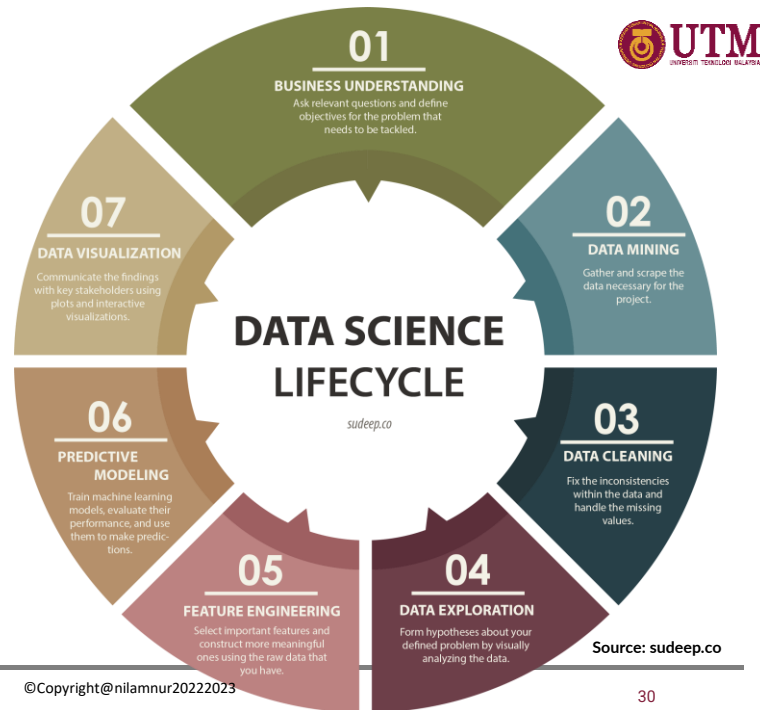
©Copyright@nilamnur20222023
30

30

Machine Learning

Supervised Learning

✓ The analysis done to predict the future event or other data or trends.
✓ It uses the supervised learning functions which are used to predict the target value
✓ Input data is label
✓ Use training dataset
✓ Use for prediction

Unsupervised Learning

✓ basically used to produce correlation, cross-tabulation, frequency etc.
✓ These technologies are used to determine the similarities in the data and to find existing patterns.
✓ to develop the captivating subgroups in the major part of the data available.
✓ Input data is unlabel
✓ Use just input dataset
✓ Used for analysis for summarization and transformation of the data into meaningful information

©Copyright@nilamnur20222023
31

31

#1 CLASSIFICATION

Assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target **class** for each case in the data

#2 REGRESSION

Regression, used primarily as a form of planning and modeling, is used to identify the likelihood of a certain variable, given the presence of other variables.

#3 TIME SERIES ANALYSIS

It is a sequence of well defined data points measured at consistent time interval

#4 CLUSTERING

Clustering is very similar to classification, but involves grouping chunks of data together based on their similarities

#5 SUMMARIZATION

It involves technique for finding a compact description of a dataset

#6 ASSOCIATION RULES

Association is related to tracking patterns, but is more specific to dependently linked variables. In this case, you'll look for specific events or attributes that are highly correlated with another event or attribute

#7 SEQUENCE DISCOVERY ANALYSIS

It is about finding a sequence of activity

innovative ● entrepreneurial ● global

©Copyright@nilamnur20222023

32

32

ML Modeling

Machine Learning Technique
- Supervised Learning
  - Classification
    - Decision Tree
    - ANN
    - KNN
    - Random Forest
    - SVM
    - Naïve Bayes
  - Regression
    - Logistic Regression
    - Linear Regression
  - Time Series Analysis
    - ARIMA
- Unsupervised Learning
  - Clustering
    - Kmeans Clustering
    - Simple Text Clustering
  - Summarization
  - Association Rules
  - Sequence discovery
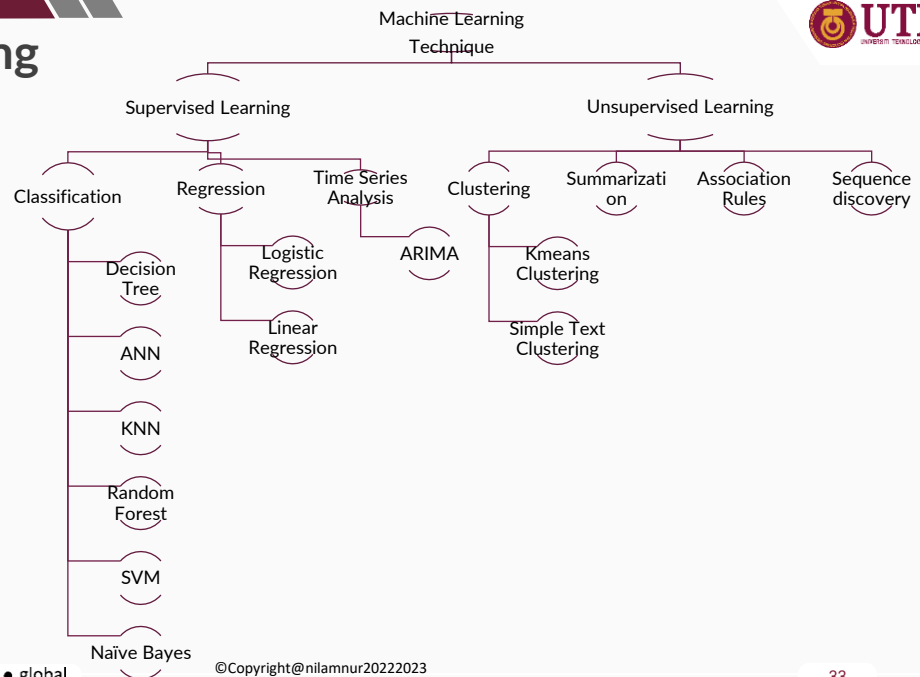
innovative ● entrepreneurial ● global

©Copyright@nilamnur20222023

33

33

Best Public Datasets for Machine Learning and Data Science : https://pub.towardsai.net/best-datasets-for-machine-learning-data-science-computer-vision-nlp-ai-c9541058cf4f
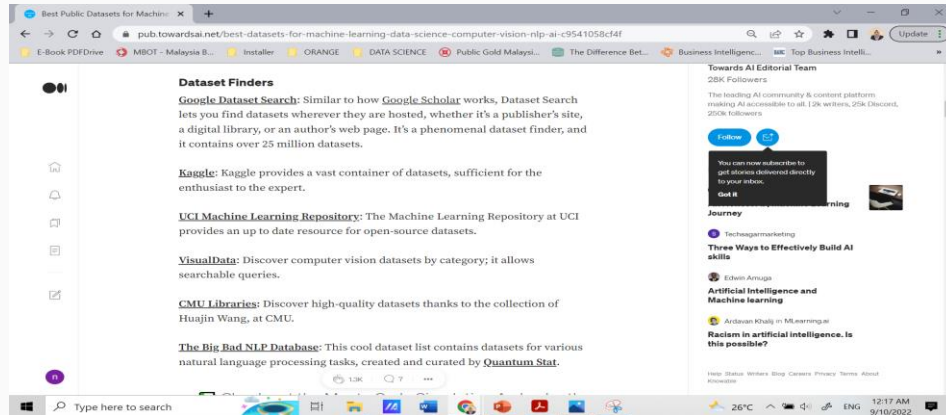
34

Data Science Tools & Platform

Molina-Solana, Miguel & Ros, María & Ruiz, M. & Gómez-Romero, Juan & Martin-Bautista, Maria. (2017). Data science for building energy management: A review. Renewable and Sustainable Energy Reviews. 70. 598-609. 10.1016/j.rser.2016.11.132.

35
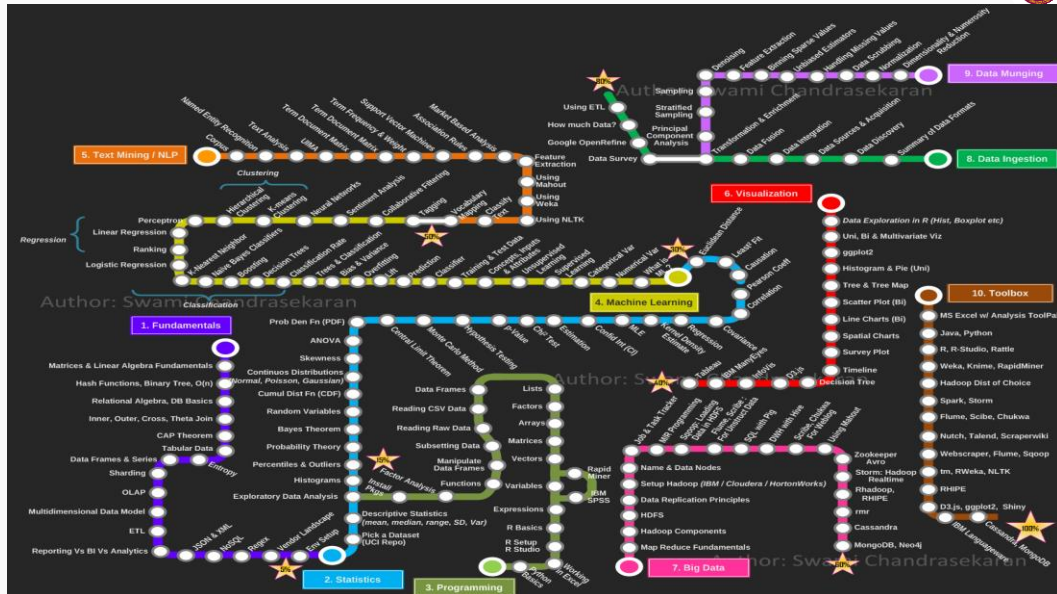
Source: aponia.com/; cheat sheet by swamichandrasekaran

innovative ● entrepreneurial ● global

36

36

Data Science is Everywhere

Manufacturing

Predicting Potential Problems
Monitoring System
Automating Manufacturing Units
Maintenance Scheduling
Anomaly Detection

E-Commerce

Identifying Consumer
Recommending products
Analyzing Reviews

Healthcare

Medical Image Analysis
Drug Discovery
Bioinformatics
Virtual Assistants

Sports

Sports Analytics
Game Outcome Prediction
(Movement Football, Basketball etc)
Team Recommendation Systems

Finance

Customer Segmentation
Strategic Decision Making
Algorithmic Trading
Risk Analytics

**Banking**

Fraud Detection
Credit Risk Modeling
Customer Life Value
Customer Support
Managing Customer Data

Transportation

Self Driving Car
Enhance Driving Experience
Car Monitoring System
Enhancing the safety of the
passenger

Agriculture

detect weeds and diseases,
predicting yield and quality of
crops, predictions regarding
livestock production

All industries and organizations alike are awashed with data in this pro-tech age and data is being created in an exponential manner, so Data Science is being used by every industry in different proportion as Data Science leads to Smarter Decision-Making.

innovative ● entrepreneurial ● global

37

37

Top Data Science Companies

©Copyright@nilamnur20222023

38

38

TOP BIG DATA ANALYTICS COMPANIES INDONESIA – GoodFirm List

CYBERTREND

XQ Informatics

Atreus Global

MasterSystem ICT Solutions

Telkom Indonesia

Purple Analytics

Eureka AI

AiSensum

Jovenindo

tokopedia

bukalapak

Arfadia

Dattabot

Cxrus Solutions

Nutanix

TELKOMSEL

ASTRA international

©Copyright@nilamnur20222023

39

39

19

So, What's Exciting About Data Science?

innovative ● entrepreneurial ● global   ©Copyright@nilamnur20222023   40

40

What's So Exciting About Data Science?

1#
"Data science was a way to become a detective."

2#
"The excitement in data science is in the journey toward achieving three significant kinds of results: discovery, insights, and innovation."

3#
"Data science is the future, and it is better to be on the cutting-edge than left behind."

4#
"Data Scientists and Machine Learning Engineers are in great demand and at the forefront of coming AI revolution"

innovative ● entrepreneurial ● global   ©Copyright@nilamnur20222023   41

41

What's So Exciting About Data Science?

5#
"Data science makes me feel powerful!"

6#
"Being able to obtain, process, and build models out of data, which will ultimately be used for their predictive capabilities [...] is incredibly exciting in my view"

7#
"Data science is our means of taming unstructured information and gathering insight."

8#
"Data scientists help improve how humans make decisions and how algorithms optimize outcomes"

innovative ● entrepreneurial ● global

©Copyright@nilamnur20222023

42

42

So, Get Ready for the Sexiest Job of 21st Century

Source: BW Education
innovative ● entrepreneurial ● global

©Copyright@nilamnur20222023

43

43

THANK YOU

DR NILAM NUR BINTI AMIR SJARIF
Email: nilamnur@utm.my

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

In the Name of God for Mankind

utm.my   univteknologimalaysia   utmofficial

44