

APB Rembau e-Bulletin

e-ISSN: 2682-776X

Edition: 11/2023

EDITORIAL BOARD

PATRON

Prof. Dr. Yamin Yasin

COORDINATOR

Nur Faathinah Mohammad Roshdan

CHIEF EDITOR

Assoc. Prof. Dr. Soo Kum Yoke, Carolyn

EDITORIAL COMMITTEE

Ooi Sing Ee

Khairon Nisa Shafeei

Shahrul Muhazad Shahrudin

October 2022 - February 2023

PEMBERSIHAN DATA DALAM DUNIA DATA SAINS

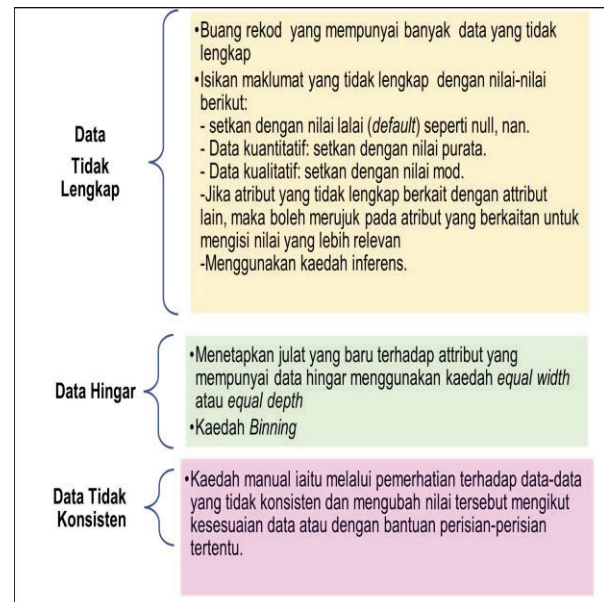
Ditulis oleh: Rozianiwati Yusof, Sri Yusmawati Mohd Yunus, Normaziah Abdul Rahman, Norhafizah Hashim, Nor Azlina Aziz Fadzillah

Bidang data sains semakin popular sejak era pendigitalan. Dengan penggunaan peralatan teknologi yang semakin canggih, ia mencetuskan letupan data dalam pelbagai bidang. Penghasilan data yang banyak telah menjadikan data sains semakin penting.

Data merupakan sesuatu fakta yang menerangkan sesuatu objek dan boleh wujud dalam pelbagai bentuk seperti nilai kuantitatif, kualitatif, imej dan teks. Data yang terhasil tidak kira daripada mana-mana sumber seperti mesin, peranti pintar atau soal selidik akan menghasilkan data mentah yang kotor (*noisy data*) iaitu data yang tidak lengkap, hingar dan tidak konsisten. Keadaan data tersebut akan mempengaruhi hasil analisis. Oleh itu, proses pembersihan data amat penting bagi menghasilkan keputusan analisa yang lebih tepat dan berkualiti.

Terdapat pelbagai cara untuk membersihkan data di mana ia bergantung kepada keadaan data tersebut. Data yang tidak lengkap boleh berlaku disebabkan masalah ketika memasukkan data samada secara sengaja atau tidak sengaja, masalah pada mesin atau data tersebut tidak penting pada waktu

itu. Data hingar merupakan data yang salah atau di luar daripada julat nilai yang betul, contohnya umur yang dimasukkan sebagai nilai negatif. Manakala data yang tidak konsisten merupakan percanggahan nilai dalam suatu atribut seperti gred markah A, B, C diletakkan nilai 1, 2, 3 atau atribut tarikh lahir bercanggah dengan atribut umur. Rajah di bawah menunjukkan cara-cara mengendalikan masalah-masalah tersebut:



Rajah 1: Kaedah Pembersihan Data

Terdapat banyak perisian yang ada dapat membantu proses pembersihan data ini. Antaranya Microsoft Excel, WEKA, MATLAB, Pustaka daripada pengaturcaraan Python dan juga R. Pengguna boleh memilih perisian yang ada berdasarkan kepada kepakaran masing-masing. Walau bagaimanapun, adalah wajib untuk melakukan pembersihan data bagi

memastikan keputusan analisa yang terhasil adalah lebih tepat dan berkualiti.

Rujukan:

Laman Web “Machine Learning Mastery: Making Developers Awesome at Machine Learning”
:https://machinelearningmastery.com/

Buku “Data Mining Concept and Techniques”, 3rd Edition, Jiawei Han & Micheline Kamber



PUBLIC SECTOR CORRUPTION IN MALAYSIA

Ditulis oleh: Nor Syamaliah Ngah & Azlina Mohd Hussain

The prevalence of public sector corruption in Malaysia in its more visible forms, which include bribery, embezzlement, fraud, cronyism, bid-rigging, money laundering, extortion, and secret commissions, have been widely discussed and debated at one point or another in public discussions and debates on national news and the social media.

The Global Corruption Barometer published by Transparency International, which consisted of a survey for Malaysia, between the time frame from July 2019 to June 2020 where two different governments were in power during that period exhibited, inter alia, that 71% Malaysians think that government corruption is an issue for real concern. Among the public institutions surveyed, Parliament (36%), the Police (30%) and Government officials (28%) were ranked the highest for perception of corruption, respectively. The survey also showed that 39% of all respondents strongly believed that corruption is on the rise.

The Global Corruption Barometer also reported that 67% of Malaysians were confident that the government was doing a