

Latihan 1 – Getting To Know Your Data

8 September – 14 September

IF4041 Ilmu Data dan Penggalan Data

Penjelasan Terkait Data

- Data ini akan digunakan untuk melakukan penelitian terkait hubungan kelahiran seorang bayi dan kondisi dari ibu yang mengandung (i.e., kebiasaan, perilaku, dan demografi).
- Dataset variables:
 - fage – father's age in years.
 - mage – mother's age in years.
 - mature – maturity status of mother.
 - weeks – length of pregnancy in weeks.
 - premie – whether the birth was classied as premature (premie) or full-term.
 - visits – number of hospital visits during pregnancy.
 - gained – weight gained by mother during pregnancy in pounds.
 - weight – weight of the baby at birth in pounds.
 - lowbirthweight – whether baby was classied as low birthweight (low) or not (not low).
 - gender – gender of the baby, female or male.
 - habit – status of the mother as a nonsmoker or a smoker.
 - marital – whether mother is married or not married at birth.
 - whitemom – whether mom is white or not white.

Tugas: Lakukan Observasi Terhadap Data

- Gunakan Jupyter Notebook
- Observasi central tendency, variance, correlation, data dispersion, dll. Gunakan juga Histogram, Boxplot, Q Plot, Q-Q plot, Scatter Plot untuk menjelaskan karakteristik data.
- Lakukan perhitungan dissimilarity atau distance, untuk objek ke 1, 51, 101, 151, ..., 951 (20 object). Buatlah dari dissimilarity matrix-nya.
- Buatlah penjelasan dari aspek yang menarik dari data berdasarkan 2 hal di atas.
- Tuliskan problem atau pertanyaan Anda terhadap data/cara observasinya yang menurut Anda perlu didiskusikan.

Deliverable

- Dokumen: berisi ulasan observasi dan penjelasannya, serta pertanyaan untuk diskusi
- Source code