

# Rekayasa Data Terapan

## Tugas Kelompok 1 (TK-01)

### Basic Text Processing

Setelah mempelajari materi yang diberikan secara flipped classroom, maka kerjakanlah Latihan berikut ini:

#### 1. Regular expression

- a. Pelajari tutorial berikut:

<https://www.kaggle.com/sohier/introduction-to-regular-expressions>

<https://stackabuse.com/using-regex-for-text-manipulation-in-python/>

- b. Kerjakan soal dari link berikut: (at least 5 problem terjawab)

<https://www.kaggle.com/sohier/exercises-for-intro-to-regular-expressions/>

#### 2. Text feature extraction

- a. Pelajari tutorial berikut ini:

<https://towardsdatascience.com/introduction-to-natural-language-processing-for-text-df845750fb63>

- b. Dengan menggunakan data 1 buah dokumen teks yang Anda definisikan sendiri, buatlah matriks TF-IDF nya.

Dikumpulkan dalam bentuk .ipynb dan file txt yang digunakan untuk soal no 2, di mana output matriks dari keluaran program di-insert sebagai image di notebook, dan dipresentasikan di kelas pada saat kuliah