**Tugas 2**

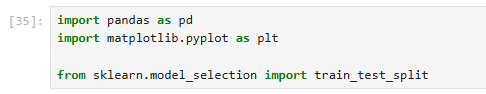
**Syahri Ghifari Maulidi - 0110222217**

1 Teknik Informatika, STT Terpadu Nurul Fikri, Depok

\*E-mail: syah22217ti@student.nurulfikri.ac.id

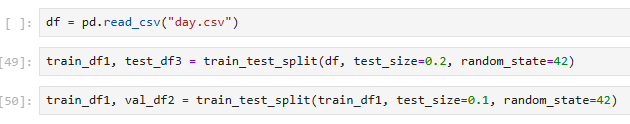
**Abstract.** Laporan ini adalah latihan mandiri yang harus dikerjakan dalam kurun waktu tertentu, di dalamnya menjelaskan apa yang harus dilakukan dan dikerjakan.

**1. Import Panda and Matplotlib.**



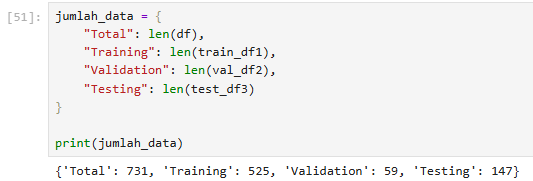
pandas digunakan untuk membaca dan mengolah data dalam format CSV. Dataset day.csv dibaca dan disimpan ke dalam variabel df.

**2. Membagi dataset**



Fungsi train\_test\_split dari sklearn membagi dataset menjadi dua bagian. Sebanyak 80% data untuk training dan 20% data untuk testing. random\_state=42 digunakan agar pembagian data selalu konsisten (tidak berubah-ubah).Dari data Training, diambil lagi 10% untuk Validation. Validation digunakan sebagai data uji coba sebelum model diterapkan ke data Testing.

## **2.1 Jumlah data dan pemanggilan data**



Menampilkan jumlah data pada setiap bagian (Training, Validation, Testing). Sebuah dictionary yang menunjukkan jumlah data per kategori. Untuk membuktikan dataset memang berhasil dibagi, ditampilkan 5 baris teratas dari setiap set. Tiga tabel yang masing-masing berisi data Training, Validation, dan Testing.

