

## **Laporan Tahap 1 – Collecting Data**

Nama : Habil Nasruddin

Kelompok : 4

Kelas : C

a. Sumber Data

Dataset yang digunakan dalam tahap ini berasal dari data penduduk. Data dikumpulkan dari sumber terbuka (misalnya open data pemerintah atau data dummy yang disusun sendiri). Dataset memiliki ukuran lebih dari 1000 baris sesuai dengan ketentuan tugas.

b. Proses Collecting Data

1. Mengunduh atau menyusun dataset mentah dan menyimpannya dalam file `raw_data.csv`.
2. Melakukan pembersihan awal seperti penghapusan data duplikat atau perbaikan format sederhana.
3. Menyimpan dataset hasil olahan awal dalam file `data_penduduk.csv`.
4. Menyimpan kode Python untuk tahap ini dalam file `01_collecting.py` atau `01_collecting.ipynb`.

c. Hasil Collecting Data

File yang dihasilkan dari tahap ini adalah sebagai berikut:

1. `raw_data.csv` → dataset mentah hasil pengumpulan.
2. `data_penduduk.csv` → dataset hasil olahan tahap awal.
3. `01_collecting.py` / `01_collecting.ipynb` → script pengumpulan data.

d. Kesimpulan

Tahap 1 (Collecting Data) telah selesai dilakukan dengan hasil berupa dataset mentah, dataset hasil olahan, serta script Python. Seluruh file ini akan digunakan sebagai dasar untuk tahap berikutnya yaitu Integrasi Data, Analisis Data, dan Validasi Data.