

Tugas 1: Laporan Praktikum Mandiri 2

Noer Muhammad Ayub - 0110222142

¹ Teknik Informatika, STT Terpadu Nurul Fikri, Depok

² Sistem Informasi, STT Terpadu Nurul Fikri, Depok

*E-mail: noer22142ti@student.nurulfikri.ac.id

1. Menghubungkan Google Drive

```
[1] ✓ 1m from google.colab import drive
drive.mount('/content/gdrive')
Mounted at /content/gdrive
```

Penjelasan

1. `from google.colab import drive`: Mengimpor modul `drive` dari paket `google.colab` yang memungkinkan kita mengakses Google Drive dalam sesi Colab.
2. `drive.mount('/content/gdrive')`: Perintah ini menghubungkan Google Drive kamu ke Colab, sehingga kamu bisa membuka dan menggunakan file Drive langsung dari folder `/content/gdrive`.

2. Menyimpan alamat file

```
[2] ✓ 0s path = "/content/gdrive/MyDrive/praktikum/praktikum2/data/day.csv"
```

`path = "/content/gdrive/MyDrive/praktikum/praktikum2/data/day.csv"`

Baris kode ini menyimpan alamat lengkap path sebuah file CSV yang ada di Google Drive ke dalam variabel bernama `path`.

3. Membaca Dataset

```
[3]
✓ 1s
#membaca dataset
import pandas as pd

df = pd.read_csv(path)
print("Jumlah total data:", len(df))
print("\n5 baris pertama dataset:")
print(df.head())
```

Jumlah total data: 731

5 baris pertama dataset:

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	\
0	1	2011-01-01	1	0	1	0	6	0	
1	2	2011-01-02	1	0	1	0	0	0	
2	3	2011-01-03	1	0	1	0	1	1	
3	4	2011-01-04	1	0	1	0	2	1	
4	5	2011-01-05	1	0	1	0	3	1	

	weathersit	temp	atemp	hum	windspeed	casual	registered	\
0	2	0.344167	0.363625	0.805833	0.160446	331	654	
1	2	0.363478	0.353739	0.696087	0.248539	131	670	
2	1	0.196364	0.189405	0.437273	0.248309	120	1229	
3	1	0.200000	0.212122	0.590435	0.160296	108	1454	
4	1	0.226957	0.229270	0.436957	0.186900	82	1518	

	cnt
0	985
1	801
2	1349
3	1562
4	1600

import pandas as pd

Mengimpor pustaka **pandas** dan memberi alias **pd**, yang sering digunakan untuk manipulasi data.

df = pd.read_csv(path)

Membaca file CSV yang lokasinya sudah disimpan di variabel **path** dan memasukkannya ke dalam objek **DataFrame df**. **DataFrame** adalah struktur data seperti tabel dengan baris dan kolom.

print("Jumlah total data:", len(df))

Mencetak jumlah total baris data dalam dataset (**len(df)** memberikan jumlah baris)

print("\n5 baris pertama dataset:")

print(df.head())

Mencetak 5 baris pertama dari dataset menggunakan fungsi **head()**, agar kita bisa melihat contoh data dan struktur kolomnya.

4. Split Dataset

```
[4]
✓ 1s
#Split dataset
from sklearn.model_selection import train_test_split
```

train_test_split

digunakan untuk membagi dataset menjadi dua bagian utama: data untuk pelatihan (training) dan data untuk pengujian (testing). Dan juga mengimpor fungsi **train_test_split** dari modul **model_selection** di pustaka **scikit-learn** **sklear**.

5. Split Training

```
[5]  
✓ 0s # Split Training (80%) dan Testing (20%)  
train_df, test_df = train_test_split(df, test_size=0.2, random_state=42)
```

df

dataset asli yang akan dibagi.

train_df

Data untuk pelatihan (training) yang berisi 80% dari total data.

test_df

Data untuk pengujian (testing) yang berisi 20% dari total data.

test_size=0.2: 20%

data dialokasikan untuk testing, sisanya (80%) untuk training.

random_state=42

memberikan seed untuk pengacakan, supaya pembagian data ini bisa diulang dan hasilnya selalu sama.

6. Membagi Ulang Training

```
[6]  
✓ 0s # Dari Training, ambil 10% sebagai Validation  
train_df, val_df = train_test_split(train_df, test_size=0.1, random_state=42)
```

train_df

data training sebelumnya (80% dari total data).

val_df

validation set yang berisi 10% data training awal (untuk validasi model saat pelatihan).

test_size=0.1

ambil 10% dari data training sebagai data validasi.

random_state=42

supaya hasil pembagian bisa direproduksi dengan cara yang sama.

7. Menampilkan Hasil

```
[7]
✓ Os
#Tampilkan hasil
print("\nJumlah Data Training:", len(train_df))
print("Jumlah Data Validation:", len(val_df))
print("Jumlah Data Testing:", len(test_df))

print("\n=== Training Data (5 teratas) ===")
print(train_df.head())

print("\n=== Validation Data (5 teratas) ===")
print(val_df.head())

print("\n=== Testing Data (5 teratas) ===")
print(test_df.head())
```

```
Jumlah Data Training: 525
Jumlah Data Validation: 59
Jumlah Data Testing: 147

=== Training Data (5 teratas) ===
   instant  dteday  season  yr  mnth  holiday  weekday  workingday  \
657      658  2012-10-19     4    1    10         0         5         1
163      164  2011-06-13     2    0     6         0         1         1
305      306  2011-11-02     4    0    11         0         3         1
111      112  2011-04-22     2    0     4         0         5         1
538      539  2012-06-22     3    1     6         0         5         1

   weathersit  temp  atemp  hum  windspeed  casual  registered
657         2  0.563333  0.537896  0.815000  0.134954      753      4671
163         1  0.635000  0.601654  0.494583  0.305350      863      4157
305         1  0.377500  0.390133  0.718750  0.082092      370      3816
111         2  0.226667  0.221054  0.720582  0.210531      177      1506
538         1  0.777500  0.724121  0.573750  0.182842      964      4859
```

Kode ini digunakan untuk menampilkan ringkasan dan contoh isi dari data yang telah dibagi menjadi:

- Training data
- Validation data
- Testing data

```
cnt
657 5424
163 5020
305 4186
111 1683
538 5823

=== Validation Data (5 teratas) ===
   instant  dteday  season  yr  mnth  holiday  weekday  workingday  \
325      326  2011-11-22     4    0    11         0         2         1
410      411  2012-02-15     1    1     2         0         3         1
92        93  2011-04-03     2    0     4         0         0         0
47        48  2011-02-17     1    0     2         0         4         1
508      509  2012-05-23     2    1     5         0         3         1

   weathersit  temp  atemp  hum  windspeed  casual  registered  \
325         3  0.416667  0.421696  0.962500  0.118792      69      1538
410         1  0.348333  0.351629  0.531250  0.181600     141      4028
92          1  0.378333  0.378767  0.480000  0.182213     1651     1598
47          1  0.435833  0.428658  0.505000  0.230104     259      2216
508         2  0.621667  0.584612  0.774583  0.102000     766      4494

   cnt
325 1607
410 4169
92  3249
47  2475
508 5260

=== Testing Data (5 teratas) ===
   instant  dteday  season  yr  mnth  holiday  weekday  workingday  \
```

```

=== Testing Data (5 teratas) ===
instant      dteday  season  yr  mnth  holiday  weekday  workingday  \
703         704  2012-12-04      4   1   12      0        2          1
33          34  2011-02-03      1   0    2      0        4          1
300         301  2011-10-28      4   0   10      0        5          1
456         457  2012-04-01      2   1    4      0        0          0
633         634  2012-09-25      4   1    9      0        2          1

weathersit    temp    atemp    hum  windspeed  casual  registered  \
703          1  0.475833  0.469054  0.733750  0.174129    551      6055
33           1  0.186957  0.177878  0.437826  0.277752     61     1489
300          2  0.330833  0.318812  0.585833  0.229479    456     3291
456          2  0.425833  0.417287  0.676250  0.172267   2347     3694
633          1  0.550000  0.544179  0.570000  0.236321    845     6693

cnt
703  6606
33   1550
300  3747
456  6041
633  7538

```