

# Tugas 1: Praktikum 1 dan Tugas 1

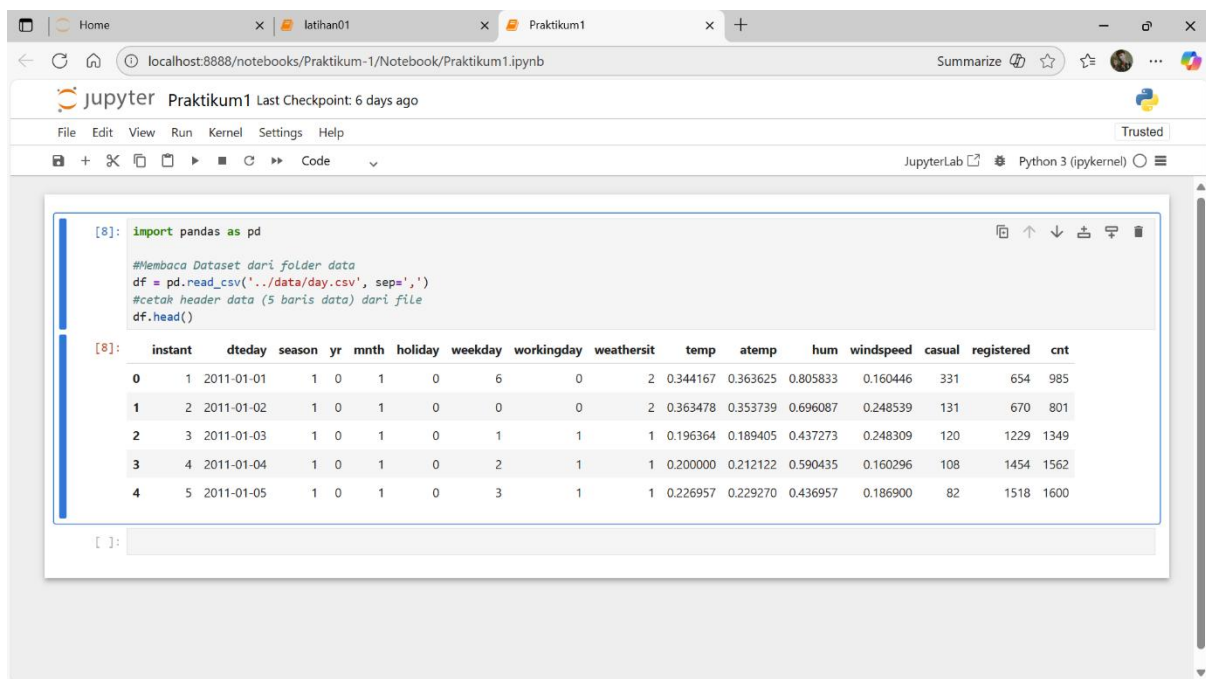
Vikri Putra Permana - 0110222070 <sup>1\*</sup>

<sup>1</sup> Teknik Informatika, STT Terpadu Nurul Fikri, Depok

<sup>2</sup> Sistem Informasi, STT Terpadu Nurul Fikri, Depok

\*E-mail: vikripp10@gmail.com

## 1. Praktikum 1



Pada praktikum pertama ini saya belajar bahasa pemrograman Python menggunakan tools Jupyter Notebook. Saya belajar untuk mengambil data dan membaca data yang berada pada folder **Data** dan didalamnya ada file **day.csv** yang merupakan data untuk mengerjakan praktikum pertama. Saya ingin menjelaskan apa fungsi dari barisan kode pada Screenshootan diatas:

Code : **import pandas as pd**

Saya mulai dengan mengimpor **pandas**, library Python buat ngolah data tabel. Aliasnya **pd** supaya lebih ringkas.

Code : **df = pd.read\_csv('../data/day.csv', sep=',')**

Di sini saya baca file bernama **day.csv** yang ada di folder **data**. Karena ini file CSV, pemisah antar datanya adalah koma, jadi saya tulis **sep=','**.

Data hasil pembacaan saya simpan ke variabel **df** dalam bentuk DataFrame.

Code : **df.head()**

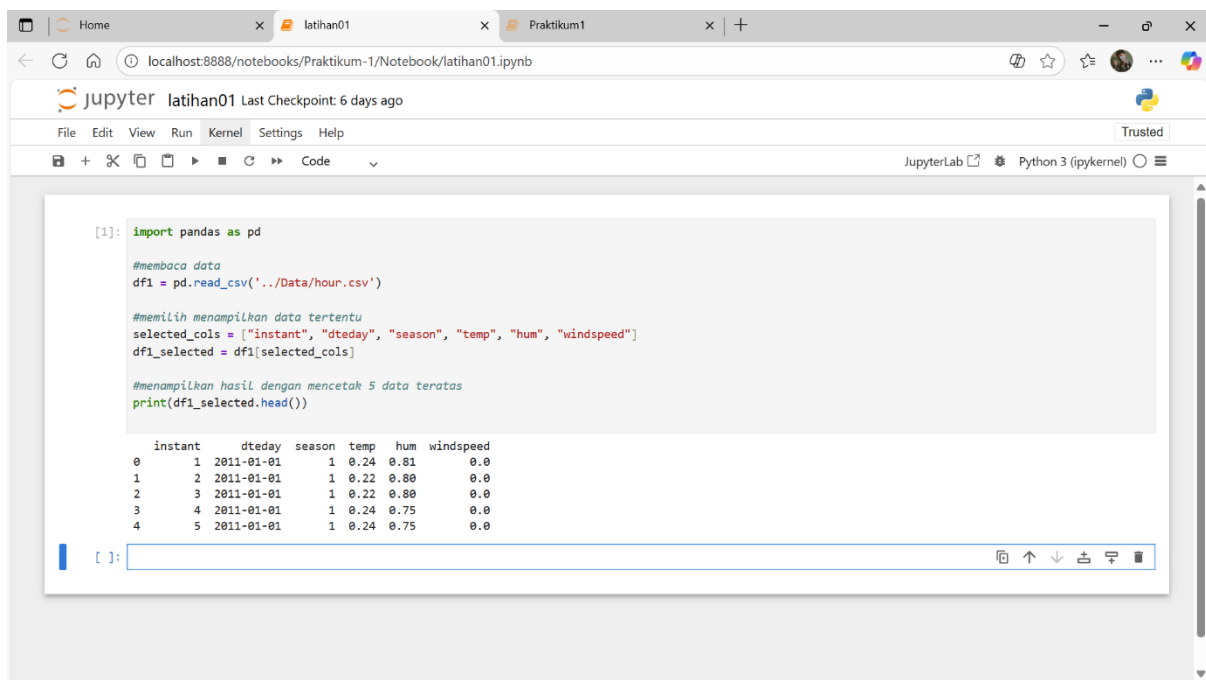
**head()** dipakai buat nunjukin **5 baris pertama** dari dataset. Ini berguna untuk ngecek struktur data, nama kolom, sama isi datanya secara cepat.

Output dari **df.head()** nunjukin 5 baris pertama tabel dengan kolom-kolom berikut:

```
[8]:
```

|   | instant | dteday     | season | yr | mnth | holiday | weekday | workingday | weathersit | temp     | atemp    | hum      | windspeed | casual | registered | cnt  |
|---|---------|------------|--------|----|------|---------|---------|------------|------------|----------|----------|----------|-----------|--------|------------|------|
| 0 | 1       | 2011-01-01 | 1      | 0  | 1    | 0       | 6       | 0          | 2          | 0.344167 | 0.363625 | 0.805833 | 0.160446  | 331    | 654        | 985  |
| 1 | 2       | 2011-01-02 | 1      | 0  | 1    | 0       | 0       | 0          | 2          | 0.363478 | 0.353739 | 0.696087 | 0.248539  | 131    | 670        | 801  |
| 2 | 3       | 2011-01-03 | 1      | 0  | 1    | 0       | 1       | 1          | 1          | 0.196364 | 0.189405 | 0.437273 | 0.248309  | 120    | 1229       | 1349 |
| 3 | 4       | 2011-01-04 | 1      | 0  | 1    | 0       | 2       | 1          | 1          | 0.200000 | 0.212122 | 0.590435 | 0.160296  | 108    | 1454       | 1562 |
| 4 | 5       | 2011-01-05 | 1      | 0  | 1    | 0       | 3       | 1          | 1          | 0.226957 | 0.229270 | 0.436957 | 0.186900  | 82     | 1518       | 1600 |

## 2. Tugas 1



The screenshot shows a Jupyter Notebook window with the following code in cell [1]:

```
[1]: import pandas as pd

#membaca data
df1 = pd.read_csv('../Data/hour.csv')

#memilih menampilkan data tertentu
selected_cols = ["instant", "dteday", "season", "temp", "hum", "windspeed"]
df1_selected = df1[selected_cols]

#menampilkan hasil dengan mencetak 5 data teratas
print(df1_selected.head())
```

The output of the code is displayed below the cell:

|   | instant | dteday     | season | temp | hum  | windspeed |
|---|---------|------------|--------|------|------|-----------|
| 0 | 1       | 2011-01-01 | 1      | 0.24 | 0.81 | 0.0       |
| 1 | 2       | 2011-01-01 | 1      | 0.22 | 0.80 | 0.0       |
| 2 | 3       | 2011-01-01 | 1      | 0.22 | 0.80 | 0.0       |
| 3 | 4       | 2011-01-01 | 1      | 0.24 | 0.75 | 0.0       |
| 4 | 5       | 2011-01-01 | 1      | 0.24 | 0.75 | 0.0       |

Pada tugas kali ini saya belajar mandiri menggunakan bahasa pemrograman Python menggunakan tools Jupyter Notebook. Saya belajar untuk mengambil data dan baca data yang berada pada folder Data dan didalamnya ada file hour.csv yang merupakan data untuk tugas pertama. Saya ingin menjelaskan apa fungsi dari barisan code yang ada pada SSan diatas:

Pertama, saya mulai dengan mengimpor library pandas:

```
import pandas as pd
```

Kenapa pakai pandas? Karena pandas ini library di Python yang memang sering banget dipakai buat ngolah data tabel. Jadi mirip Excel versi Python lah. Nah, biar gampang dipanggil lagi nanti, saya kasih alias jadi pd.

Setelah itu, saya baca file CSV-nya:

```
dfl = pd.read_csv('../Data/hour.csv')
```

File CSV yang saya baca ini namanya **hour.csv** dan letaknya ada di folder **Data**. Simbol **..** di depan itu maksudnya naik satu folder dari posisi sekarang. Data yang sudah dibaca itu otomatis jadi DataFrame (tabel versi pandas), dan saya simpan ke variabel **dfl**.

Setelah itu, saya hanya ingin menampilkan beberapa data saja :

```
selected_cols = ["instant", "dteday", "season", "temp", "hum", "windspeed"]
```

dengan kode itu saya membuat daftar nama kolom yang mau dipakai atau yang mau saya tampilkan ('instant', 'dteday', 'season', 'temp', 'hum', 'windspeed') lalu disimpan ke variabel 'selected\_cols'.

Setelah itu, saya bikin DataFrame baru yang isinya hanya kolom-kolom tadi:

```
dfl_selected = dfl[selected_cols]
```

Jadi, kalau **dfl** masih penuh dengan banyak kolom, sekarang **dfl\_selected** lebih ringkas karena cuma nyimpen kolom yang saya butuhkan aja.

Terakhir, saya cek hasilnya dengan menampilkan 5 baris pertama:

```
print(dfl_selected.head())
```

Kenapa cuma 5 baris? Karena kalau langsung tampil semua bisa panjang banget. Dengan head(), kita cukup lihat sampel data teratas buat memastikan data sudah terbaca dengan benar.

Output-nya kelihatan seperti ini:

|   | instant | dteday     | season | temp | hum  | windspeed |
|---|---------|------------|--------|------|------|-----------|
| 0 | 1       | 2011-01-01 | 1      | 0.24 | 0.81 | 0.0       |
| 1 | 2       | 2011-01-01 | 1      | 0.22 | 0.80 | 0.0       |
| 2 | 3       | 2011-01-01 | 1      | 0.22 | 0.80 | 0.0       |
| 3 | 4       | 2011-01-01 | 1      | 0.24 | 0.75 | 0.0       |
| 4 | 5       | 2011-01-01 | 1      | 0.24 | 0.75 | 0.0       |