

Tugas 1: Praktikum dan Latihan Mandiri 01

Revani – 0110224111¹

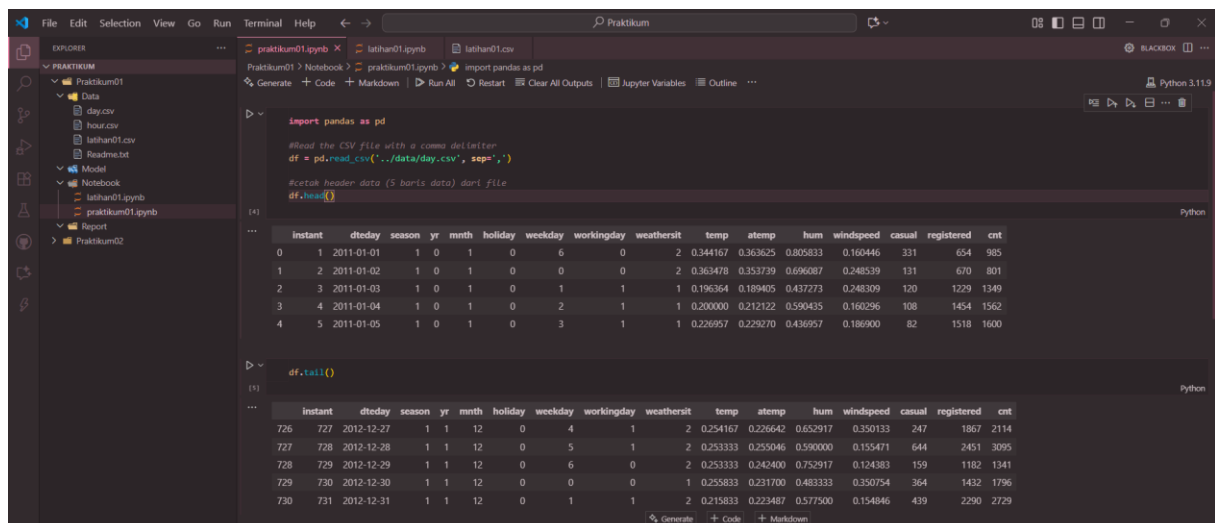
¹ Teknik Informatika, STT Terpadu Nurul Fikri, Depok

*E-mail: 0110224111@student.nurulfikri.ac.id

Abstract. Praktikum dan latihan mandiri ini bertujuan untuk memahami proses dasar pengolahan data menggunakan Python dengan library Pandas. Dataset yang digunakan adalah *hour.csv* dan *day.csv*. Langkah yang dilakukan mencakup membaca dataset, menyeleksi kolom penting (*instant*, *dteday*, *season*, *temp*, *hum*, *windspeed*), menampilkan hasilnya, serta menyimpan dataframe baru ke file latihan01.csv. Praktikum ini membuktikan bahwa pemahaman terhadap struktur data dan pemilihan atribut relevan sangat penting sebagai pondasi analisis *machine learning*.

1. Praktikum Mandiri

Pada praktikum mandiri ini, langkah awal yang dilakukan adalah membaca dataset *day.csv* melalui perangkat lunak Visual Studio Code dengan menggunakan *extensions* Jupyter Notebook untuk memastikan data dapat diakses dengan baik. Berikut ini adalah hasil praktikum yang saya kerjakan pada praktikum mandiri 01.



```
import pandas as pd

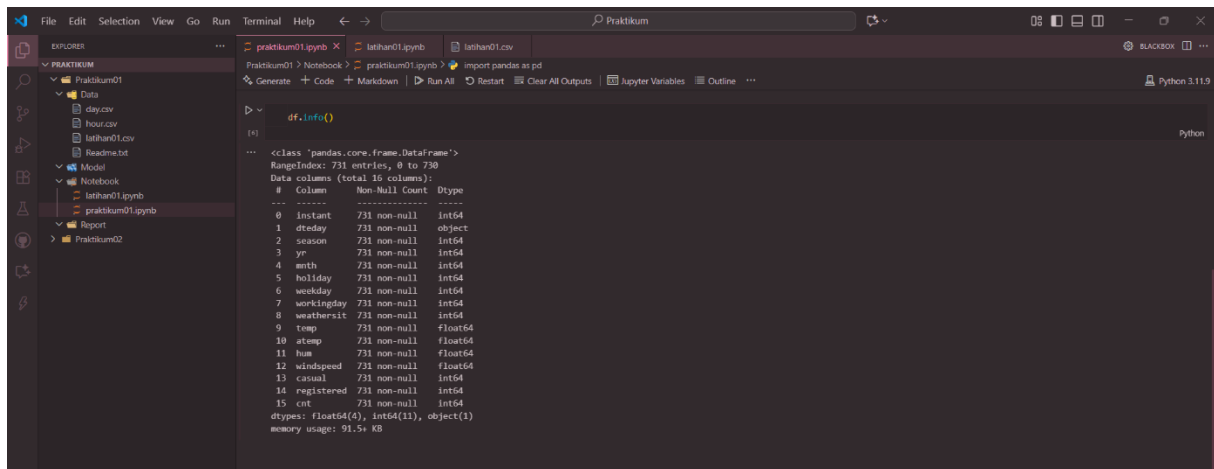
# Read the CSV file with a comma delimiter
df = pd.read_csv("../data/day.csv", sep=',')

# Cetak header data (5 baris data) dari file
df.head()
```

	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

```
df.tail()
```

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
726	727	2012-12-27	1	1	12	0	4	1	2	0.254167	0.226642	0.652917	0.350133	247	1867	2114
727	728	2012-12-28	1	1	12	0	5	1	2	0.253333	0.255046	0.590000	0.155471	644	2451	3095
728	729	2012-12-29	1	1	12	0	6	0	2	0.253333	0.242400	0.752917	0.124383	159	1182	1341
729	730	2012-12-30	1	1	12	0	0	0	1	0.255833	0.231700	0.483333	0.350754	364	1432	1796
730	731	2012-12-31	1	1	12	0	1	1	2	0.215833	0.223487	0.577500	0.154846	439	2290	2729



Gambar 1. Pada bagian ini ditunjukkan proses membaca dataset *day.csv* dengan Pandas di Visual Studio Code menggunakan *extensions* Jupyter Notebook.

import pandas as pd

Perintah ini digunakan untuk memanggil library Pandas yang digunakan untuk mengolah data berbasis tabel (dataframe).

df = pd.read_csv('./data/day.csv', sep=',')

Perintah ini digunakan untuk membaca file *day.csv* yang ada di folder data. Hasil pembacaan disimpan dalam variabel *df* sebagai sebuah dataframe.

df.head()

Perintah ini digunakan untuk menampilkan beberapa baris pertama dari sebuah dataframe. Output yang ditampilkan menunjukkan 5 baris pertama dataset dengan kolom seperti *instant*, *dteday*, *season*, *temp*, *hum*, *windspeed*, *casual*, *registered*, dan *cnt*.

df.tail()

Perintah ini digunakan untuk menampilkan beberapa baris terakhir dari sebuah dataframe. Output yang ditampilkan yaitu dari baris ke-727 sampai 731 yang merupakan data terakhir.

df.info()

Perintah ini digunakan untuk menampilkan ringkasan informasi dari dataframe. Output yang ditampilkan adalah jumlah data dengan 731 baris dan 16 kolom, dengan tipe data int, float dan object, serta semua kolom punya 731 *non-null values* yang berarti tidak ada data kosong. Jadi dapat dipahami bahwa dataset rapi dan siap untuk dianalisis lebih lanjut tanpa perlu pembersihan data besar-besaran.

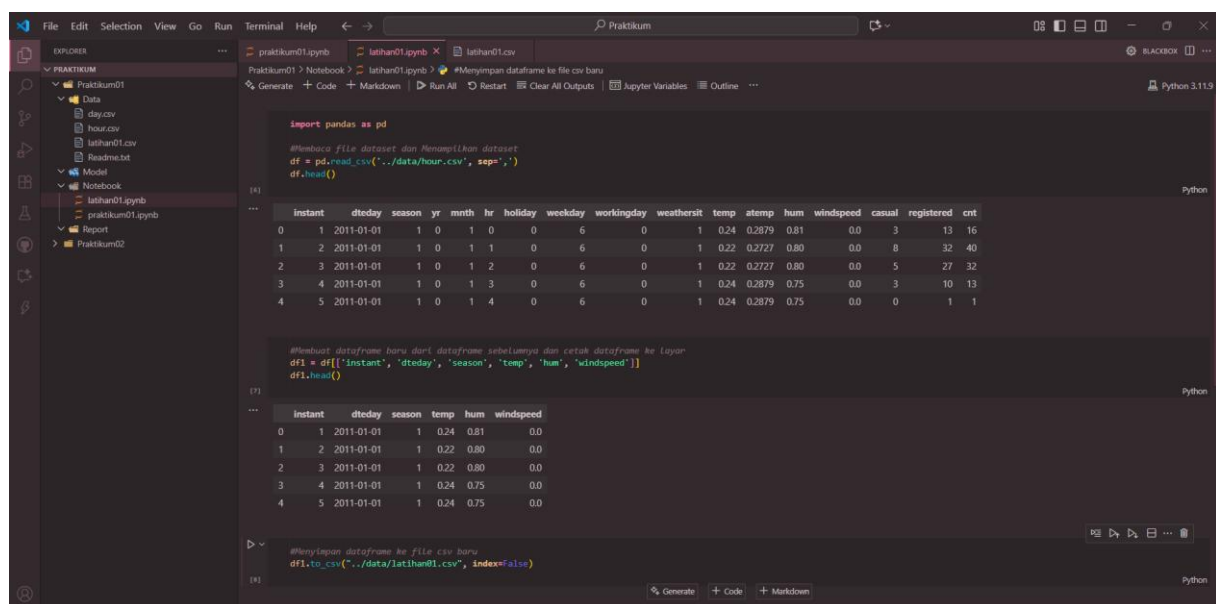
Kesimpulan dan Hasil Implementasi

Dari implementasi ini, saya berhasil membaca dataset *day.csv* yang dimuat ke dalam Python menggunakan Pandas tanpa kendala dan mengecek isinya menggunakan fungsi

df.head(), df.tail(), dan df.info(). Hasilnya menunjukkan bahwa data lengkap, konsisten dan tidak ada nilai kosong. Praktikum ini menunjukkan bahwa memahami struktur awal data penting dilakukan agar kita mengetahui kondisi dataset dan atribut yang relevan, sehingga proses analisis maupun penerapan *machine learning* dapat berjalan lebih efektif.

2. Latihan Mandiri

Pada latihan mandiri ini, langkah awal yang dilakukan adalah membaca dataset *hour.csv* melalui perangkat lunak Visual Studio Code dengan menggunakan *extensions* Jupyter Notebook untuk memastikan data dapat diakses dengan baik. Selanjutnya dibuat sebuah dataframe baru dengan memilih kolom-kolom yang dianggap relevan sesuai instruksi, yaitu *instant*, *dteday*, *season*, *temp*, *hum*, dan *windspeed*. Dataframe hasil seleksi tersebut kemudian ditampilkan untuk memastikan data sesuai dengan yang diharapkan. Tahap terakhir adalah menyimpan dataframe baru tersebut ke dalam file *latihan01.csv* di direktori *data*. Berikut ini adalah hasil latihan yang saya kerjakan pada latihan mandiri 01.



Gambar 2. Bagian ini merupakan hasil dari tugas latihan mandiri, di mana langkah yang dilakukan adalah membaca dataset *hour.csv*.

```
import pandas as pd
```

Perintah ini digunakan untuk memanggil library Pandas yang digunakan untuk mengolah data berbasis table (dataframe).

```
df = pd.read_csv('../data/hour.csv', sep=',')
```

Perintah ini membaca file *hour.csv* yang ada di folder *data*. Hasil pembacaan disimpan dalam variabel *df* sebagai sebuah dataframe.

```
df.head()
```

Perintah ini digunakan untuk menampilkan beberapa baris pertama dari sebuah dataframe. Output yang ditampilkan menunjukkan dataset berisi data peminjaman sepeda per jam beserta kondisi cuaca. Ini langkah awal untuk memastikan file berhasil terbaca.

```
df1 = df[['instant', 'dteday', 'season', 'temp', 'hum', 'windspeed']]  
df1.head()
```

Perintah ini untuk membuat dataframe baru df1 dengan kolom *instant*, *dteday*, *season*, *temp*, *hum*, dan *windspeed*. Output yang ditampilkan menunjukkan dataframe lebih sederhana dan fokus pada variabel yang relevan.

```
df1.to_csv("../data/latihan01.csv", index=False)
```

Langkah terakhir yaitu menyimpan dataframe df1 ke file baru bernama latihan01.csv di direktori data. Dengan cara ini, hasil olahan bisa dipakai lagi tanpa harus mengulang proses dari awal.

Kesimpulan dan Hasil Implementasi

Dari latihan mandiri ini dapat disimpulkan bahwa proses pengolahan data dengan Pandas sangat membantu dalam proses pengolahan data berbasis table, mulai dari membaca dataset, menyeleksi kolom, hingga menyimpan hasil. Pembuatan dataframe baru membantu menyederhanakan data agar lebih fokus pada variabel yang relevan, sehingga memudahkan analisis selanjutnya. Implementasi ini juga menunjukkan pentingnya pemahaman struktur awal dataset dan pemilihan atribut relevan merupakan Langkah penting sebelum masuk ke tahap analisis data yang lebih kompleks, termasuk penerapan algoritma *machine learning*.

Link GitHub (Praktikum dan Latihan Mandiri)

https://github.com/revani18/ML_2025_Revani_3AI01.git