

# Tugas 1: Praktikum Mandiri 1 Machine Learning

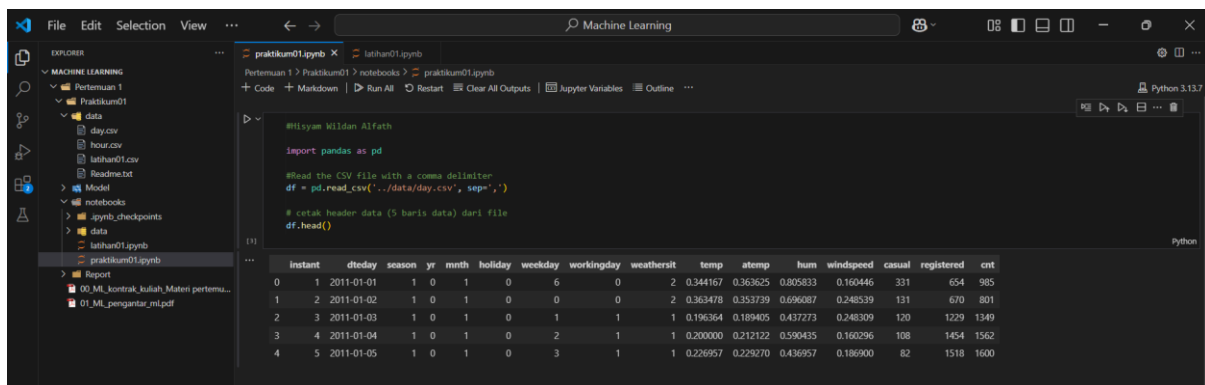
**Hisyam Wildan Alfath - 0110222206**

Teknik Informatika, STT Terpadu Nurul Fikri, Depok

E-mail: [hisy22206ti@student.nurulfikri.ac.id](mailto:hisy22206ti@student.nurulfikri.ac.id)

**Abstract.** Laporan ini berisi dokumentasi dari langkah-langkah belajar praktikum machine learning. Dataset yang digunakan ada dua, yaitu day.csv yang berisi data penyewaan sepeda per hari, dan hour.csv yang berisi data penyewaan sepeda per jam. Pada praktikum pertama, kita menggunakan library pandas di Python untuk mengelola data. Pertama, kita mengimpor pandas, lalu membaca data dari file hour.csv ke dalam dataframe, dan menampilkan 5 baris pertama dari data tersebut. Selanjutnya, kita membuat dataframe baru yang bernama df1 dengan memilih beberapa kolom penting dari data asli, yaitu instant, dteday, season, temp, hum, dan windspeed. Terakhir, data di df1 disimpan ke file CSV baru bernama latihan01.csv supaya hanya data penting saja yang disimpan untuk analisis berikutnya.

## 1. Praktikum 01



```
#Hisyam Wildan Alfath

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	month	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	0.248339	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

Penjelasan Kode:

- `import pandas as pd`  
Baris ini mengimpor library pandas dengan alias pd. Library ini sangat populer untuk memproses data berbentuk tabel seperti CSV.
- `df = pd.read_csv('../data/day.csv', sep=',')`  
Kode ini membaca file CSV yang bernama day.csv dari folder data. File CSV tersebut dipisahkan dengan tanda koma (,), sesuai argumen `sep=','`. Hasil pembacaan disimpan ke dalam variabel df berupa DataFrame, yaitu struktur data tabular di pandas.
- `df.head()`

Fungsi `head()` digunakan untuk menampilkan lima baris pertama dari data yang telah dimuat ke dalam DataFrame `df`. Ini untuk melihat secara cepat seperti apa data dan kolom-kolom yang ada.

Penjelasan Output:

Output dari kode `df.head()` menampilkan lima baris pertama data dengan beberapa kolom, antara lain:

- `instant`: Nomor urut data.
- `dteday`: Tanggal.
- `season`, `yr`, `mnth`, `holiday`, `weekday`, `workingday`, `weathersit`: Informasi terkait waktu, musim, hari kerja, hari libur, dan kondisi cuaca.
- `temp`, `atemp`, `hum`, `windspeed`: Data numerik untuk suhu, suhu terasa, kelembapan, dan kecepatan angin.
- `casual`, `registered`, `cnt`: Jumlah pengguna sepeda harian baik kasual maupun terdaftar, serta totalnya.

Kesimpulan:

File CSV berhasil dimuat ke dalam DataFrame `pandas` dan data dapat dengan mudah ditampilkan serta diolah. Struktur data sudah lengkap dengan kolom-kolom informatif, sehingga siap untuk analisis lebih lanjut. Menampilkan lima baris pertama dari data membantu pengguna memahami susunan dan jenis data sebelum melakukan analisis lanjutan seperti visualisasi atau pemodelan statistik.

## 2. Praktikum Mandiri 01

```
import pandas as pd

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

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

	instant	dteday	season	yr	month	hr	holiday	weekday	workingday	weatherit	temp	atemp	hum	windspeed	casual	registered	cnt
0	1	2011-01-01	1	0	1	0	0	6	0	1	0.24	0.2879	0.81	0.0	3	13	16
1	2	2011-01-01	1	0	1	1	0	6	0	1	0.22	0.2727	0.80	0.0	8	32	40
2	3	2011-01-01	1	0	1	2	0	6	0	1	0.22	0.2727	0.80	0.0	5	27	32
3	4	2011-01-01	1	0	1	3	0	6	0	1	0.24	0.2879	0.75	0.0	3	10	13
4	5	2011-01-01	1	0	1	4	0	6	0	1	0.24	0.2879	0.75	0.0	0	1	1

```
# Membaca dataset
df = pd.read_csv('../data/day.csv')

# Membuat dataframe baru dengan kolom yang dipilih
selected_cols = ['instant', 'dteday', 'season', 'temp', 'hum', 'windspeed']
df1 = df[selected_cols]

# Menampilkan dataframe baru
print(df1.head())

# Menyimpan dataframe ke file CSV baru
df1.to_csv('../data/latihan01.csv', index=False)
```

	instant	dteday	season	temp	hum	windspeed
0	1	2011-01-01	1	0.344157	0.805833	0.168446
1	2	2011-01-02	1	0.363428	0.696087	0.248539
2	3	2011-01-03	1	0.196364	0.432723	0.248309
3	4	2011-01-04	1	0.200000	0.590435	0.168296
4	5	2011-01-05	1	0.226957	0.436957	0.186900

### Penjelasan Kode:

- import pandas as pd  
Mengimpor library pandas untuk pemrosesan data berbasis tabel (DataFrame).
- Membaca file CSV  
`df = pd.read_csv('../data/hour.csv', sep=',')`  
`df.head()`  
Kode ini membaca file CSV bernama hour.csv dengan delimiter koma, lalu menampilkan 5 baris pertama menggunakan `df.head()`.
- Membaca file CSV lain  
`df = pd.read_csv('../data/day.csv')`  
Baris ini membaca file CSV lain yaitu day.csv, juga dengan pandas.
- Membuat DataFrame baru dengan kolom spesifik  
`selected_cols = ['instant', 'dteday', 'season', 'temp', 'hum', 'windspeed']`  
`df1 = df[selected_cols]`  
Kode ini membuat list berisi nama kolom yang dipilih, lalu mengambil subset dari DataFrame hanya dengan kolom tersebut dan menyimpannya ke `df1`.
- Menampilkan data dari DataFrame baru

```
print(df1.head())
```

Menampilkan 5 baris pertama dari DataFrame hasil seleksi kolom. Output menunjukkan data yang lebih ringkas hanya pada kolom yang dipilih.

- Menyimpan DataFrame ke file CSV baru

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

Kode ini menyimpan DataFrame baru df1 ke file CSV baru bernama latihan01.csv tanpa menyertakan index baris.

Penjelasan Output:

- Output pertama dari df.head() menampilkan lima baris pertama dari file hour.csv, data yang terdiri dari berbagai atribut waktu dan data pengguna sepeda tiap jam.
- Output dari print(df1.head()) menampilkan lima baris pertama dari data harian sepeda, namun hanya pada kolom yang sudah dipilih: 'instant', 'dteday', 'season', 'temp', 'hum', dan 'windspeed'. Data yang ditampilkan menjadi lebih sederhana, memudahkan fokus ke analisis variable tertentu.

Kesimpulan:

Implementasi pengolahan data menggunakan pandas berhasil memilih dan menyimpan variabel penting dari dataset penyewaan sepeda seperti instant, dteday, season, temp, hum, dan windspeed. Proses ini menyederhanakan dataset yang awalnya kompleks sehingga memudahkan analisis selanjutnya dan memungkinkan fokus pada variabel yang relevan. Seleksi kolom, pembuatan DataFrame baru, dan penyimpanan ke file CSV baru membuktikan fleksibilitas dan efisiensi pandas dalam pengolahan data tabular. Implementasi ini sangat membantu mempercepat persiapan data untuk tahap eksplorasi, visualisasi, dan pemodelan prediktif berikutnya.

Link github praktikum 01:

[https://github.com/HisyamWildan/TI03\\_HisyamW.A\\_0110222206/blob/main/Pertemuan%201/Praktikum%20dan%20Tugas%20Mandiri%2001/notebooks/praktikum01.ipynb](https://github.com/HisyamWildan/TI03_HisyamW.A_0110222206/blob/main/Pertemuan%201/Praktikum%20dan%20Tugas%20Mandiri%2001/notebooks/praktikum01.ipynb)

Link github praktikum mandiri 01:

[https://github.com/HisyamWildan/TI03\\_HisyamW.A\\_0110222206/blob/main/Pertemuan%201/Praktikum%20dan%20Tugas%20Mandiri%2001/notebooks/latihan01.ipynb](https://github.com/HisyamWildan/TI03_HisyamW.A_0110222206/blob/main/Pertemuan%201/Praktikum%20dan%20Tugas%20Mandiri%2001/notebooks/latihan01.ipynb)