

Tugas 1: Praktikum Machine Learning (Tugas/Latihan Mandiri)

Shapiere Januar Rafiansyah – 0110224191¹

¹ Teknik Informatika, STT Terpadu Nurul Fikri, Depok

*E-mail: shapierejanuarr@gmail.com

Abstract. Pada praktikum mandiri ini, fokus utama adalah 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 berikutnya.

Pada praktikum mandiri ini, langkah awal yang dilakukan adalah membaca dataset *hour.csv* melalui 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 codingan yang saya kerjakan pada praktikum mandiri.

```
[3]: import pandas as pd

#Membaca Dataset dari folder data
df=pd.read_csv("../data/day.csv")
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

```
[4]: 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

```
[5]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 731 entries, 0 to 730
Data columns (total 16 columns):
#   Column             Non-Null Count  Dtype  
---  -
0   instant            731 non-null   int64  
1   dteday             731 non-null   object  
2   season             731 non-null   int64  
3   yr                 731 non-null   int64  
4   mnth               731 non-null   int64  
5   holiday            731 non-null   int64  
6   weekday            731 non-null   int64  
7   workingday         731 non-null   int64  
8   weathersit          731 non-null   int64  
9   temp               731 non-null   float64 
10  atemp              731 non-null   float64 
11  hum                 731 non-null   float64 
12  windspeed          731 non-null   float64
```

Gambar 1. Pada bagian ini ditunjukkan proses membaca dataset `day.csv` menggunakan Pandas di Jupyter Notebook.

```
import pandas as pd
```

Langkah awal ini memanggil library **Pandas** yang digunakan untuk mengolah data berbasis tabel (dataframe).

```
df = pd.read_csv("../data/day.csv")
```

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

```
df.head()
```

Output menunjukkan 5 baris pertama dataset. Terlihat kolom seperti *instant*, *dteday*, *season*, *temp*, *hum*, *windspeed*, *casual*, *registered*, dan *cnt*. Fungsi utamanya: untuk mengecek apakah dataset berhasil dimuat dengan benar

```
df.tail()
```

Output menampilkan baris ke-727 sampai 731 (data terakhir). Hal ini memastikan bahwa dataset terbaca penuh, dari awal sampai akhir. Dataset berisi periode tahun 2011–2012 dengan total 731 hari.

```
df.info()
```

Output `info()` memberikan ringkasan struktur dataset:

- Jumlah data: **731 baris, 16 kolom**.
- Tipe data: 11 kolom bertipe integer, 4 kolom bertipe float, dan 1 kolom (tanggal *dteday*) bertipe object/string.
- Semua kolom punya **731 non-null values**, artinya tidak ada data kosong (missing values). Dari sini bisa dipahami bahwa dataset rapi, siap untuk dianalisis lebih lanjut tanpa perlu pembersihan data besar-besaran.

Kesimpulan Hasil Implementasi

Dari implementasi ini, saya berhasil membaca dataset `day.csv` ke dalam dataframe menggunakan Pandas dan mengecek isinya melalui fungsi `head()`, `tail()`, dan `info()`. Hasilnya menunjukkan bahwa dataset memiliki 731 entri harian selama tahun 2011–2012 dengan 16 atribut lengkap tanpa missing values. Praktikum ini membuktikan bahwa memahami struktur awal data merupakan langkah penting sebelum masuk ke tahap analisis atau penerapan algoritma machine learning, karena dari sinilah kita tahu kualitas data dan atribut apa saja yang relevan untuk dipakai

```
[13]: import pandas as pd

#Menampilkan dataset
df = pd.read_csv("../data/hour.csv")
df.head()
```

	instant	dteday	season	yr	mnth	hr	holiday	weekday	workingday	weathersit	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

```
[14]: #Buat dataframe baru dan cetak dataframe
df1 = df[['instant', 'dteday', 'season', 'temp', 'hum', 'windspeed']]
df1.head()
```

	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

```
[15]: #Simpan dataframe ke file csv baru
df1.to_csv("../data/latihan01.csv", index=False)
```

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

```
import pandas as pd

df = pd.read_csv("../data/hour.csv")
df.head()
```

Bagian ini membaca dataset `hour.csv` dan menampilkan beberapa baris pertama. Dari output terlihat 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()
```

Di sini dibuat dataframe baru `df1` dengan kolom tertentu: *instant*, *dteday*, *season*, *temp*, *hum*, dan *windspeed*. Outputnya menunjukkan dataframe lebih sederhana dan fokus pada variabel yang relevan.

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

Tahap terakhir adalah menyimpan dataframe `df1` ke file baru bernama `latihan01.csv`. Dengan cara ini, hasil olahan bisa dipakai lagi tanpa harus mengulang proses dari awal.

Kesimpulan Hasil Implementasi

Dari latihan mandiri ini dapat disimpulkan bahwa proses pengolahan data dengan Pandas cukup efektif untuk membaca, memfilter, dan menyimpan ulang dataset. Pembuatan dataframe baru membantu menyederhanakan data agar lebih fokus pada variabel yang relevan, sehingga memudahkan analisis selanjutnya. Implementasi ini juga menunjukkan pentingnya manajemen data sejak tahap awal sebelum masuk ke pemodelan machine learning yang lebih kompleks.

Link GitHub (Tugas dan Latihan Mandiri)

<https://github.com/Shapiere/Praktikum-MachineLearning>