

# ▼ TUGAS PRAKTIKUM I PEMBELAJARAN MESIN

(Introduction to Machine Learning and Google Colab)

Adiba Justinian || 24060120130080 || Lab ML C1

---

## Import Library yang Dibutuhkan

Pada tugas kali ini, hanya menggunakan library pandas. Pandas merupakan sebuah library di Python yang berlisensi BSD dan open source yang menyediakan struktur data dan analisis data yang mudah digunakan.

```
#import library yang dibutuhkan
import pandas
```

## Memuat Dataset

Untuk memuat dataset, pertama-tama dilakukan dengan mendefinisikan url dari dataset yang diambil dari alamat repository UCI Machine Learning dan pada tugas ini dataset yang digunakan adalah dataset klasifikasi Bunga Iris. Kemudian, didefinisikan atribut dari dataset tersebut yang terdiri dari 5 kolom. Empat kolom pengukuran bunga dalam centimeter dan kolom kelima adalah spesies bunga yang diamati. Lalu, untuk membaca dataset digunakan fungsi read\_csv().

```
#mendefinisikan url dari dataset
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"

#mendefinisikan names yang berisi atribut dari dataset
names = ['sepal-lenght-in-cm', 'sepal-widht-in-cm', 'petal-lenght-in-cm', 'petal-widht-in-cm']

#membaca dataset
dataset = pandas.read_csv(url, names=names)
```

## Melakukan Summary Dataset

### I. Menentukan Dimensi dari Dataset

Dimensi dari dataset merupakan gambaran singkat mengenai banyaknya jumlah baris yang menunjukkan banyaknya sampel data dan jumlah kolom yang menunjukkan atribut data dari

dataset terkait. Pada hasil dibawah, dimensi dari dataset terdefinisi memiliki 150 sampel data dan 5 atribut data.

```
#menentukan dimensi dari dataset
print(dataset.shape)
```

```
(150, 5)
```

## II. Melihat Isi Dataset

Selanjutnya, untuk melihat atau menampilkan 20 baris data pertama (teratas) dari dataset dilakukan dengan perintah seperti di bawah.

```
#menampilkan 20 data teratas dari dataset
print(dataset.head(20))
```

	sepal-lenght-in-cm	sepal-widht-in-cm	petal-lenght-in-cm	\
0	5.1	3.5	1.4	
1	4.9	3.0	1.4	
2	4.7	3.2	1.3	
3	4.6	3.1	1.5	
4	5.0	3.6	1.4	
5	5.4	3.9	1.7	
6	4.6	3.4	1.4	
7	5.0	3.4	1.5	
8	4.4	2.9	1.4	
9	4.9	3.1	1.5	
10	5.4	3.7	1.5	
11	4.8	3.4	1.6	
12	4.8	3.0	1.4	
13	4.3	3.0	1.1	
14	5.8	4.0	1.2	
15	5.7	4.4	1.5	
16	5.4	3.9	1.3	
17	5.1	3.5	1.4	
18	5.7	3.8	1.7	
19	5.1	3.8	1.5	

	petal-widht-in-cm	class
0	0.2	Iris-setosa
1	0.2	Iris-setosa
2	0.2	Iris-setosa
3	0.2	Iris-setosa
4	0.2	Iris-setosa
5	0.4	Iris-setosa
6	0.3	Iris-setosa
7	0.2	Iris-setosa
8	0.2	Iris-setosa
9	0.1	Iris-setosa
10	0.2	Iris-setosa

```

11          0.2  Iris-setosa
12          0.1  Iris-setosa
13          0.1  Iris-setosa
14          0.2  Iris-setosa
15          0.4  Iris-setosa
16          0.4  Iris-setosa
17          0.3  Iris-setosa
18          0.3  Iris-setosa
19          0.3  Iris-setosa

```

### III. Distribusi Kelas Data

Kode di bawah digunakan untuk menampilkan distribusi kelas data dengan fungsi `groupby('class').size()` sehingga akan menampilkan distribusi kelas data pada dataset klasifikasi Bunga Iris seperti di bawah ini.

```

print(dataset.groupby('class').size())

class
Iris-setosa      50
Iris-versicolor  50
Iris-virginica   50
dtype: int64

```

### IV. Ringkasan Statistik

Perintah di bawah digunakan untuk mendeskripsikan statistik data meliputi count, mean, standard deviasi, nilai min, nilai max, dan quartile.

```

print(dataset.describe())

```

	sepal-lenght-in-cm	sepal-widht-in-cm	petal-lenght-in-cm	\
count	150.000000	150.000000	150.000000	
mean	5.843333	3.054000	3.758667	
std	0.828066	0.433594	1.764420	
min	4.300000	2.000000	1.000000	
25%	5.100000	2.800000	1.600000	
50%	5.800000	3.000000	4.350000	
75%	6.400000	3.300000	5.100000	
max	7.900000	4.400000	6.900000	

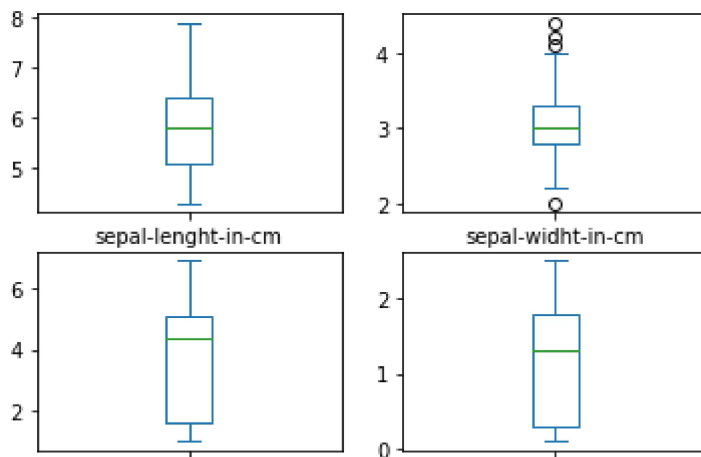
  

	petal-widht-in-cm
count	150.000000
mean	1.198667
std	0.763161
min	0.100000
25%	0.300000
50%	1.300000
75%	1.800000
max	2.500000

## V. Visualisasi Data Menggunakan Plot Univariat

Untuk membuat visualisasi data menggunakan plot, maka dilakukan import library matplotlib.pyplot terlebih dahulu. Plot univariat adalah plot dari masing-masing variabel individu yang inputnya numerik. Karena variabel inputnya numerik, maka dibuat jenis plot box dengan kode seperti di bawah yang kemudian ditampilkan dengan perintah plt.show()

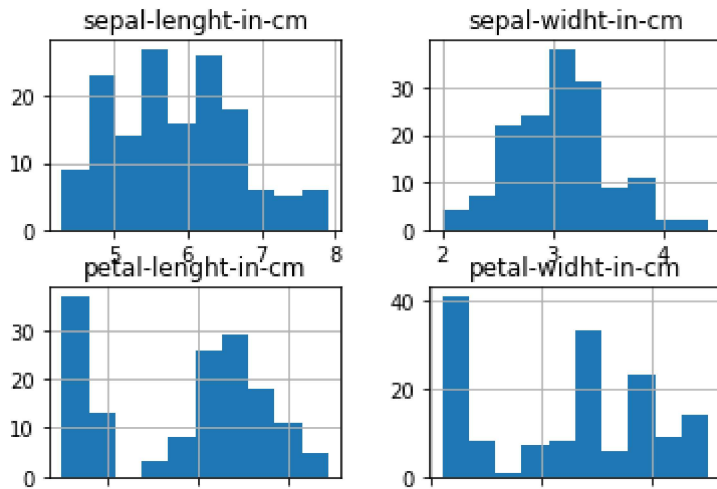
```
#membuat plot univariat
import matplotlib.pyplot as plt
dataset.plot(kind='box', subplots=True, layout=(2,2), sharex=False, sharey=False)
plt.show()
```



## VI. Membuat Histogram

Membuat histogram dari masing-masing variabel input untuk mendapatkan ide tentang distribusi dataset menggunakan fungsi dataset.hist(). Kemudian, ditampilkan menggunakan plt.show().

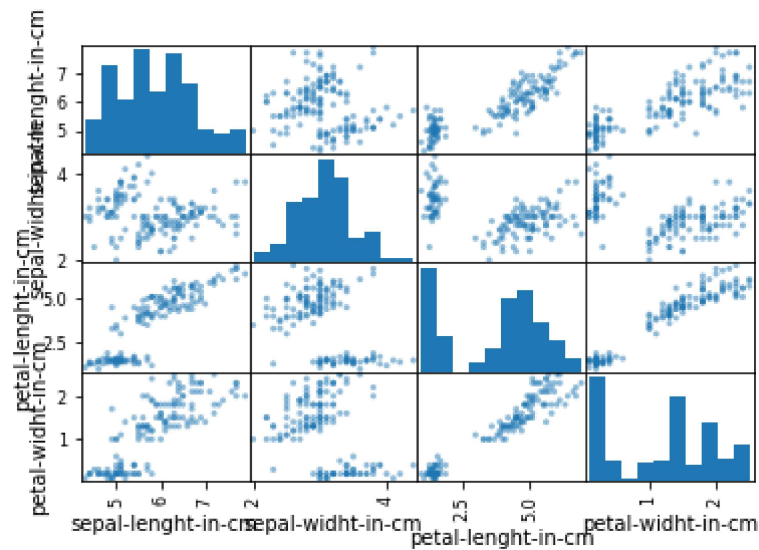
```
#membuat histogram
dataset.hist()
plt.show()
```



## VII. Visualisasi Data Menggunakan Plot Multivariat

Plot Multivariat digunakan untuk melihat interaksi antar variabel. Dalam membuat visualisasi data menggunakan plot multivariat, pertama-tama dilakukan dengan melihat scatterplots dari semua pasang atribut. Hal ini untuk membantu melihat hubungan terstruktur antara variabel input. Lalu, scatter matrix akan memplot semua variabel numerik yang kita miliki dalam kumpulan data satu sama lain. Kemudian, hasilnya akan ditampilkan menggunakan perintah `plt.show()`.

```
from pandas.plotting import scatter_matrix
scatter_matrix(dataset)
plt.show()
```



[Colab paid products](#) - [Cancel contracts here](#)

