
Praktikum Machine Learning

Nama : Fayza Aulia

NIM : 24060120120010

Lab : A2

Praktikum Pertemuan 1

Machine Learning :

Introduction to Machine Learning and Google Colab

▼ Melakukan Import Library dan Memuat Dataset

```
import pandas
url = "http://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
names = ['sepal-length', 'sepal-width', 'petal-length', 'petal-width', 'class']
dataset = pandas.read_csv(url, names=names)
```

```
print(dataset)
```

```
↗      sepal-length  sepal-width  petal-length  petal-width      class
0           5.1           3.5           1.4           0.2  Iris-setosa
1           4.9           3.0           1.4           0.2  Iris-setosa
2           4.7           3.2           1.3           0.2  Iris-setosa
3           4.6           3.1           1.5           0.2  Iris-setosa
4           5.0           3.6           1.4           0.2  Iris-setosa
..           ...           ...           ...           ...      ...
145          6.7           3.0           5.2           2.3  Iris-virginica
```

146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

[150 rows x 5 columns]

▼ Melakukan Summary Dimensi dari Dataset

Dimensi dari dataset yang dimaksud adalah gambaran singkat mengenai banyaknya jumlah baris yang menunjukkan banyaknya sampel data dan jumlah kolom yang menunjukkan atribut data dari dataset terkait. Kemudian, digunakan fungsi **shape** untuk melihat dimensi dari dataset ini.

```
print(dataset.shape)
```

```
(150, 5)
```

1. Dari hasil diatas terlihat bahwa dataset.shape memiliki jumlah sample data baris 150 serta kolom 5 sebagai atribut.

▼ Melihat Isi Dataset di 20 Baris Pertama

2. Melihat isi dataset dari 20 teratas didapat untuk melihat n baris pertama dari dataset digunakan fungsi head(n). Berikut adalah 20 baris pertama dari Iris Dataset.

```
print(dataset.head(20))
```

	sepal-length	sepal-width	petal-length	petal-width	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa

2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa
10	5.4	3.7	1.5	0.2	Iris-setosa
11	4.8	3.4	1.6	0.2	Iris-setosa
12	4.8	3.0	1.4	0.1	Iris-setosa
13	4.3	3.0	1.1	0.1	Iris-setosa
14	5.8	4.0	1.2	0.2	Iris-setosa
15	5.7	4.4	1.5	0.4	Iris-setosa
16	5.4	3.9	1.3	0.4	Iris-setosa
17	5.1	3.5	1.4	0.3	Iris-setosa
18	5.7	3.8	1.7	0.3	Iris-setosa
19	5.1	3.8	1.5	0.3	Iris-setosa

▼ Distribusi Kelas Data

3. Menampilkan data grup tertentu dan ukurannya (jumlah baris dengan class yang sama). Untuk melakukannya digunakan fungsi `groupby('class').size()`. Berikut adalah hasil untuk distribusi kelas pada dataset:

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

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

▼ Ringkasan Statistik

4. Pada ringkasan ini dilakukan untuk mendeskripsikan statistik data digunakan fungsi `describe()`. Hasil tampilannya meliputi count, mean, standard deviasi, nilai min, nilai max, dan quartile.

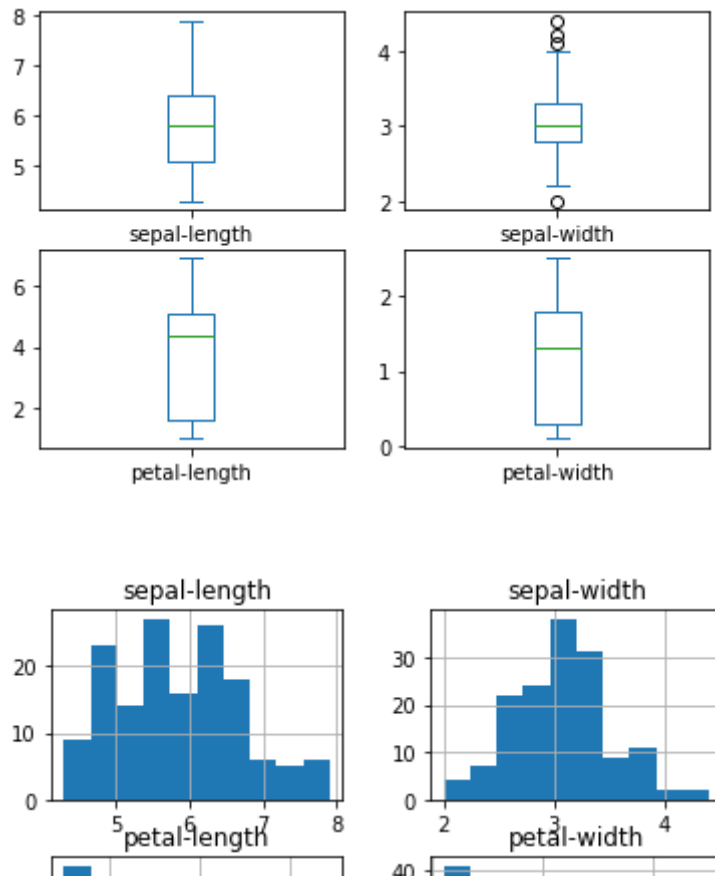
```
print(dataset.describe())
```

	sepal-length	sepal-width	petal-length	petal-width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

▼ Visualisasi Data Menggunakan Plot Univariat

Plot univariat adalah plot dari masing-masing variabel individu. Mengingat bahwa variabel inputnya numerik, kita bisa membuat jenis plot box.

```
import matplotlib.pyplot as plt
dataset.plot(kind='box', subplots=True, layout=(2,2), sharex=False, sharey=False)
plt.show()
print("\n")
dataset.hist()
plt.show()
```

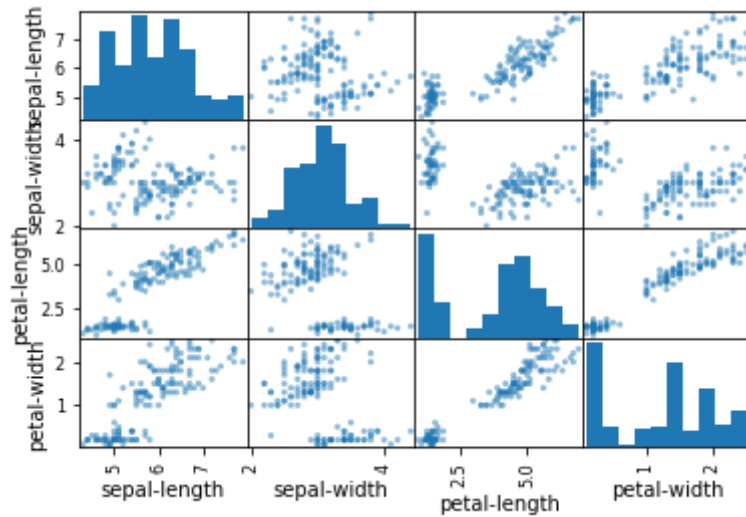


5. Mengeplot dataset menjadi gambar berbentuk box. Dimana bagian-bagian dari box tersebut memiliki arti seperti garis vertikal paling atas adalah batas maksimum dan vertikal paling bawah adalah batas minimum, kemudian garis horizontal adalah whiskers yang dimulai dari kuartil 3 ke batas maksimum dan kuartil 1 ke batas minimum, garis vertikal di tengah box adalah median. Data tersebut didapat dari fungsi `dataset.plot(kind="box", subplots=True, layout=(2,2), sharex=False, sharey=False)`. Lalu untuk gambar dibawahnya mengeplot dataset menjadi bentuk histogram yang didapat dari fungsi `dataset.hist()`

▼ Visualisasi Data Dengan Plot Multivariat

Kita juga bisa membuat histogram untuk masing-masing variabel. Hal ini dapat membantu kita untuk mendapatkan ide tentang distribusi.

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



6. Pada visualisasi kali ini dilakukan plot dataset dengan multivariat yang menggunakan scatter plot. Dimana penggunaan titik-titik untuk mewakili nilai untuk dua variabel numerik yang berbeda. Jika dua data tersebut sama, maka akan menghasilkan histogram.

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

✓ 0s completed at 4:15 PM

