

# Modul Praktikum

## **Pembelajaran Mesin** ***(Machine Learning)***



### **Pertemuan 1** **Python Introduction**

# Python Introduction

Python adalah bahasa pemrograman interpretatif multiguna. Tidak seperti bahasa lain yang susah untuk dibaca dan dipahami, python lebih menekankan pada keterbacaan kode agar lebih mudah untuk memahami sintaks. Python dapat digunakan untuk Web Programming, GUI Development, Scientific and Numeric, dan System Administration.

## Python 2 or Python 3

Terdapat 2 versi python yang sering diributkan mana yang sebaiknya digunakan. Namun, dibalik 2 versi tersebut, keduanya memiliki kelebihan dan kekurangan masing-masing. Python 2 adalah python yang muncul dahulu, sedangkan python 3 adalah yang terbaru. Untuk versi pada praktikum ini menggunakan python 3, karena merupakan versi yang terbaru.

## Pre-Installed Python?

Pada beberapa sistem operasi semisal Ubuntu dan OS X, bahasa pemrograman Python sudah *pre-installed* (bawaan sistem operasi), namun untuk Windows harus melakukan instalasi terlebih dahulu.

Untuk mengetahui apakah Python sudah ter-*install* dalam sistem operasi, berikan perintah `python --version` pada *command line*. Berikut ini contoh pada Ubuntu:

```
$ python --version
Python 2.7.6
```

Perhatikan bahwa perintah python diketik setelah simbol \$ pada *command line* Linux / OS X, sedangkan pada Windows diketik setelah simbol > .

```
> python --version
Python 2.7.6
```

Versi Python yang digunakan pada contoh adalah Python 2.7.6. Jika sistem operasi anda belum terinstall Python, maka output tidak menunjukkan versi Python yang terinstall.

## Installing Python with Conda

Untuk kemudahan praktikum, pada modul ini Anda akan “dipaksa” untuk melakukan instalasi Python versi 3 menggunakan Conda, sekalipun Python telah terinstall pada platform yang anda gunakan, kecuali jika Anda sudah terbiasa menggunakan Python.

Conda merupakan *package manager* untuk mengelola Python beserta modul-modul yang dibutuhkan. Secara *default*, *package manager* bawaan Python adalah pip (Pip Installs Packages). Praktikum ini akan menggunakan Conda dan pip secara bergantian, tergantung kebutuhan saat praktikum berlangsung.

Anda dapat melakukan instalasi Conda *full-packages* (Anaconda) atau *minimal-packages* (Miniconda). Ukuran Anaconda sebesar 455MB (Anaconda 4.1.1 for Linux) dengan ruang instalasi sekitar 1.8GB, sedangkan Miniconda (Miniconda 4.1.1 for Linux) hanya 34MB dengan ruang instalasi berkisar 169MB.

Peserta praktikum diharapkan melakukan instalasi Miniconda untuk Python versi 3.

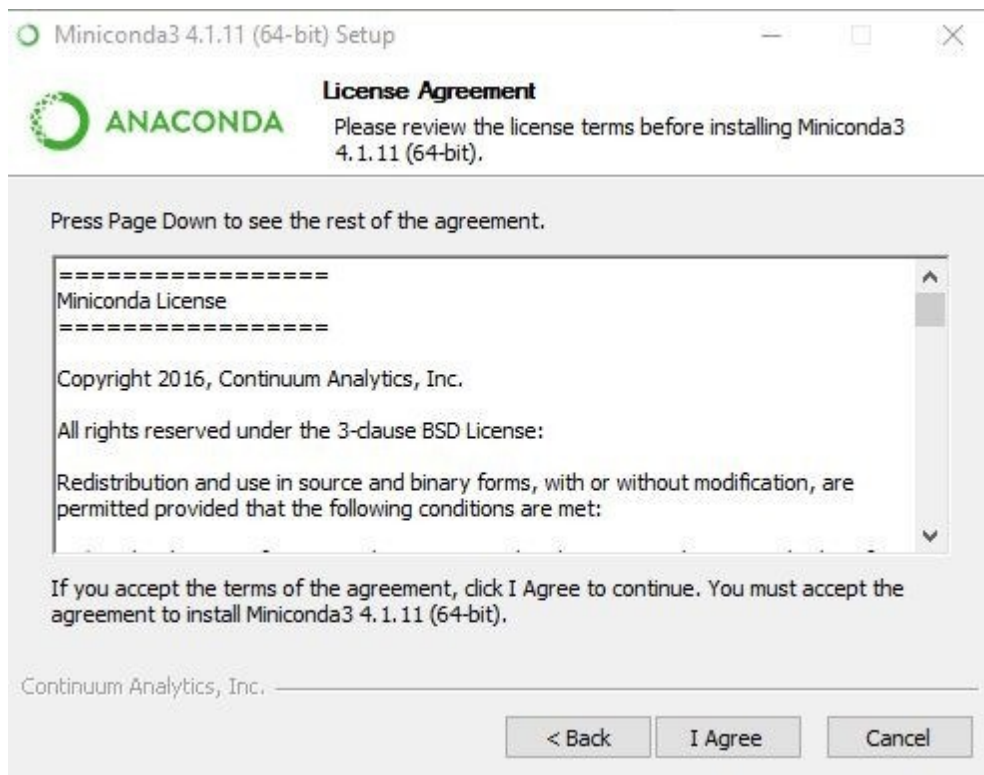
- Bukalah tautan berikut ini : <http://conda.pydata.org/miniconda.html> (Pilih yang versi terbaru)
- Unduh installer untuk Python 3.5 sesuai dengan sistem operasi yang anda gunakan.
- Lakukan Instalasi

### Installation on Windows

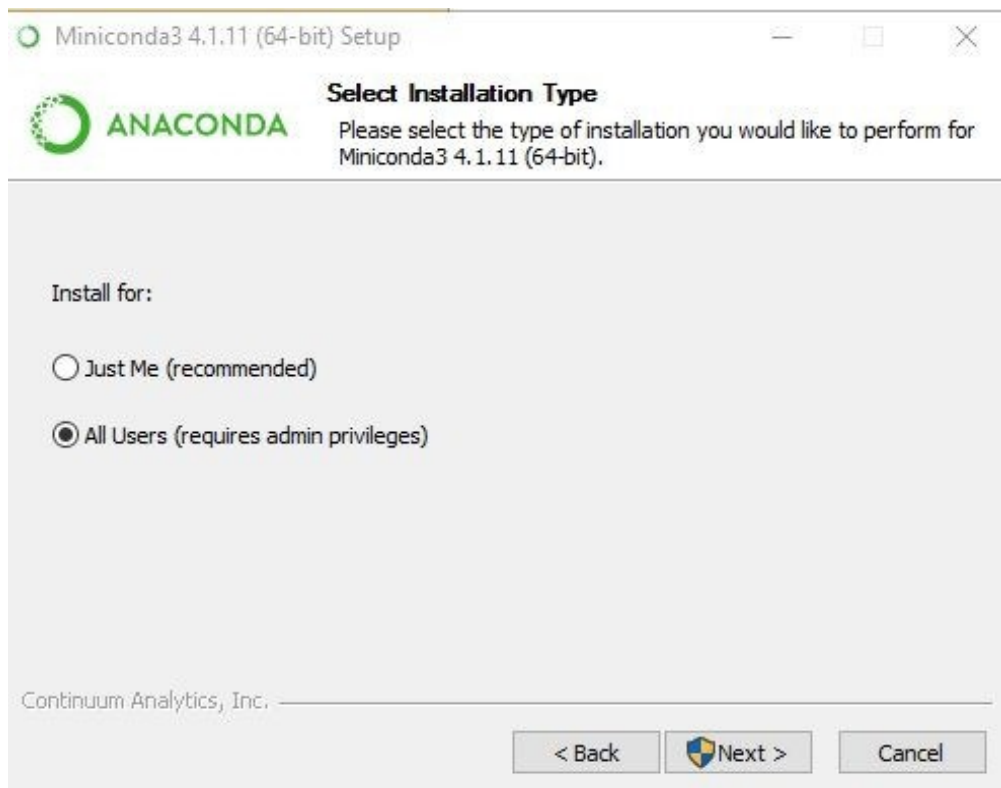
- Buka installer yang telah diunduh
- Klik Next



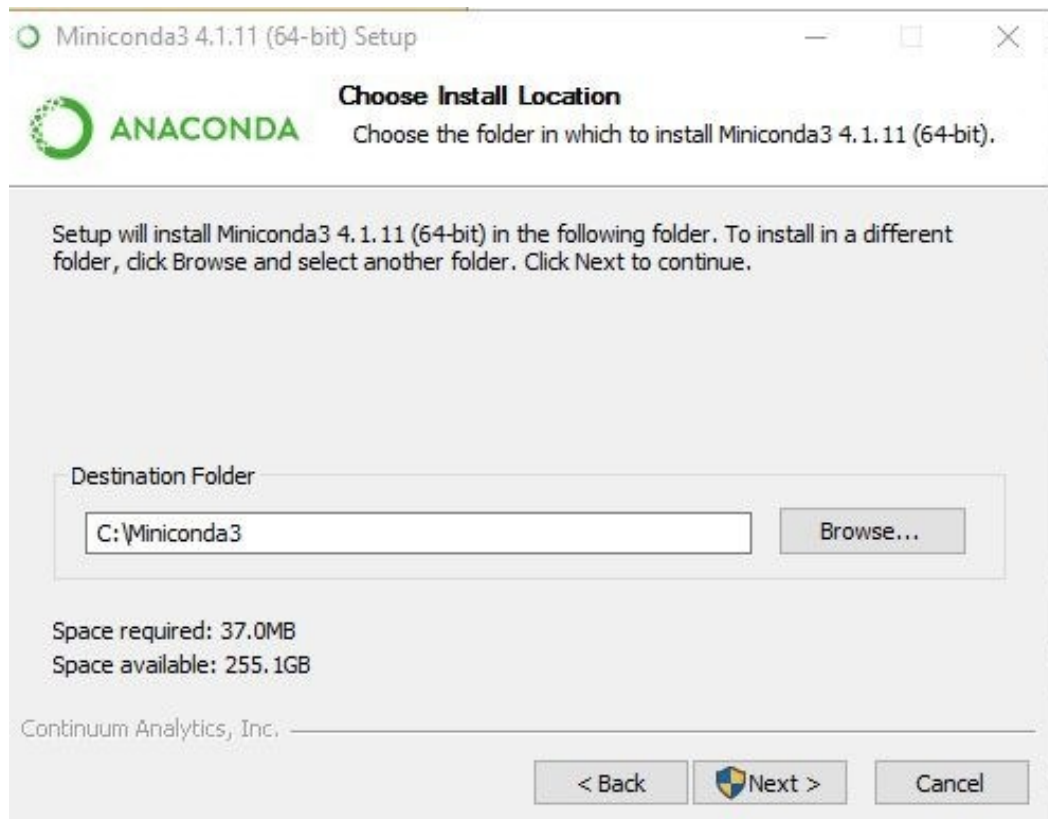
- Klik Agree untuk menyetujui License Agreement



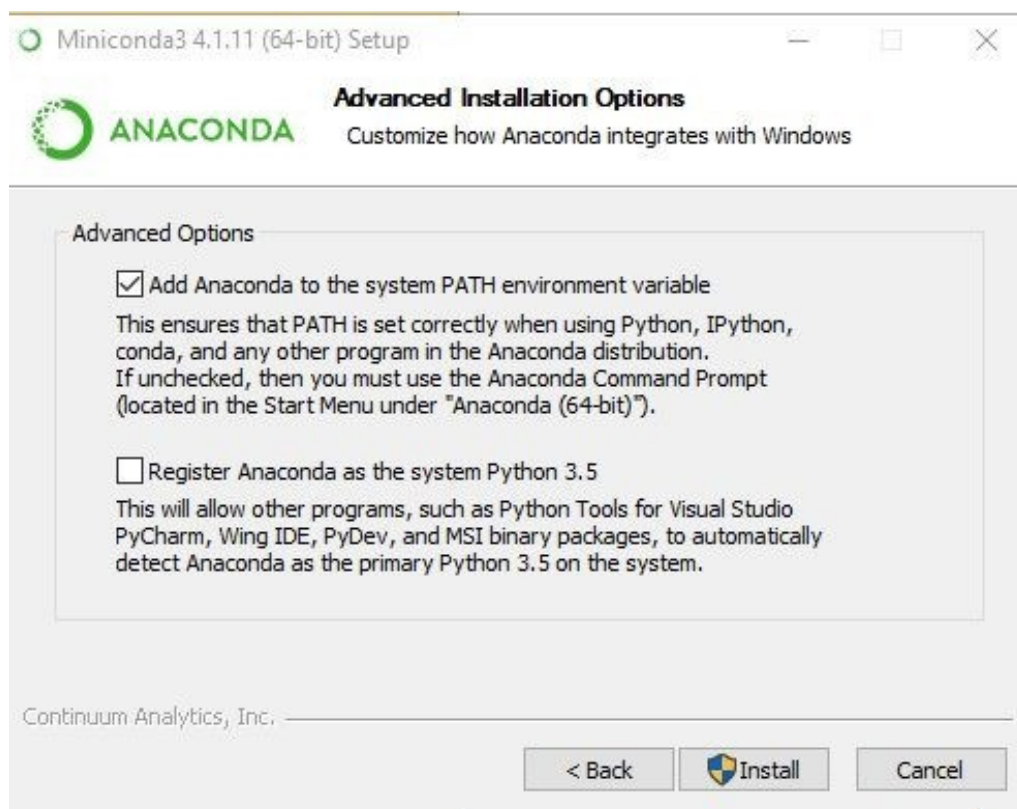
- Anda dapat memilih Just Me atau All Users



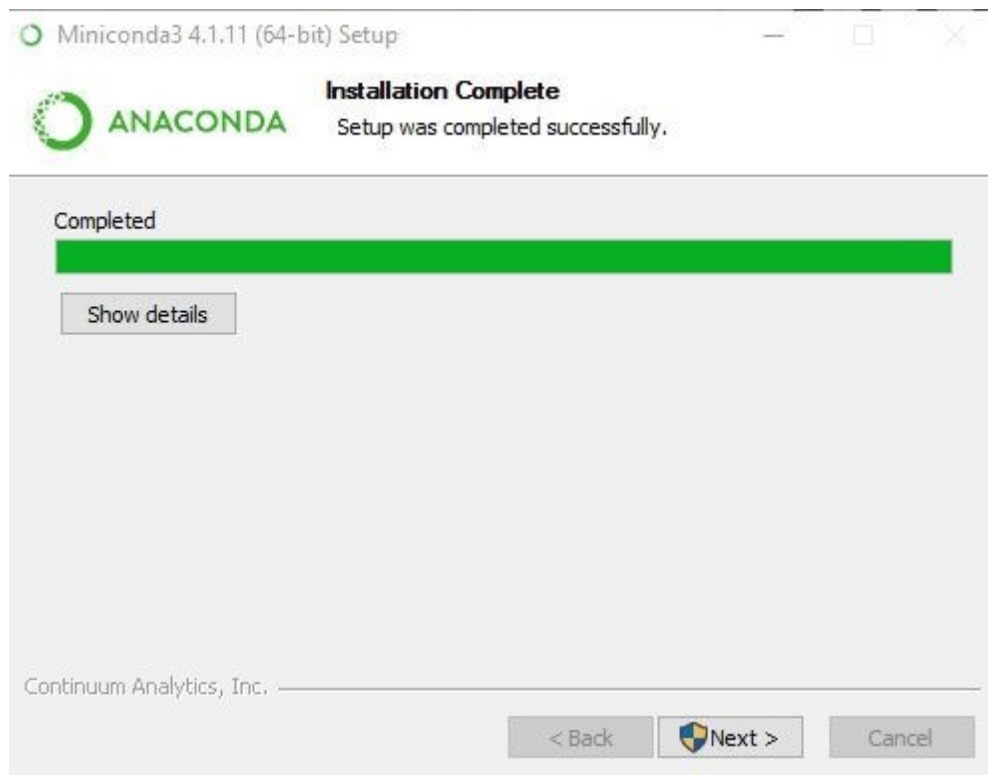
- Pilih lokasi untuk instalasi



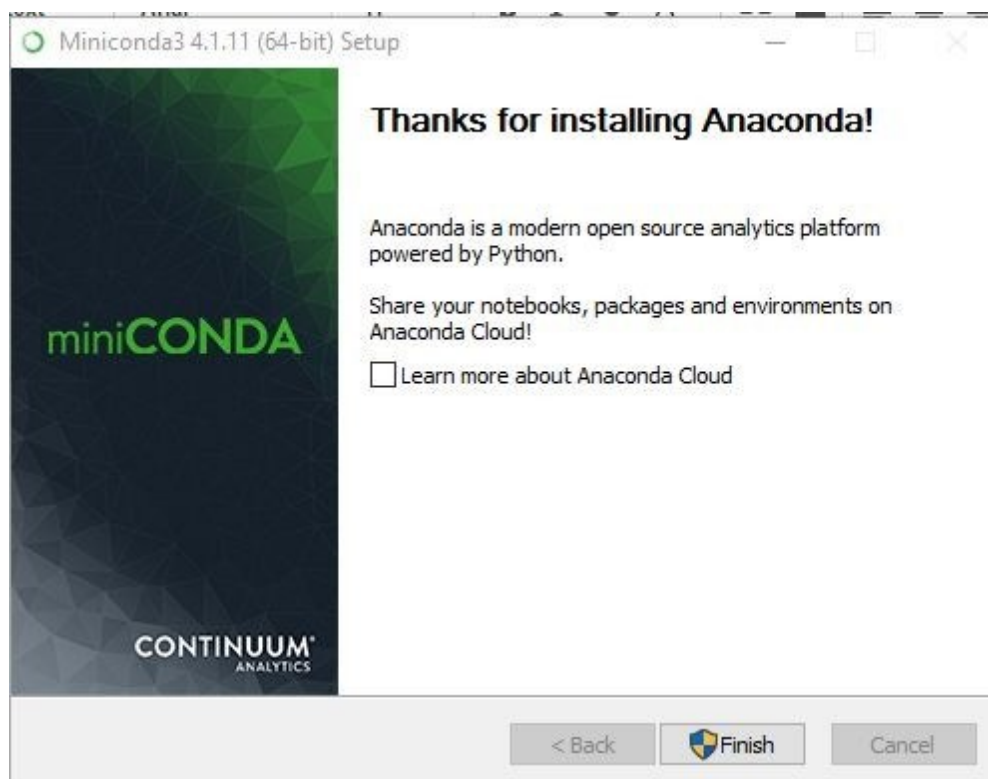
- Pilih kedua checkbox



- Tunggu sampai instalasi selesai



- Instalasi selesai



## Installation on Linux

- Unduh installer

```
katiklabs@DESKTOP-LEES0AL:~$ wget https://repo.continuum.io/miniconda/Miniconda3-latest-Linux-x86_64.sh
--2016-10-26 14:03:35-- https://repo.continuum.io/miniconda/Miniconda3-latest-Linux-x86_64.sh
Resolving repo.continuum.io (repo.continuum.io)... 54.225.185.77, 184.73.190.72, 23.21.228.27, ...
Connecting to repo.continuum.io (repo.continuum.io)|54.225.185.77|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 33987431 (32M) [application/octet-stream]
Saving to: 'Miniconda3-latest-Linux-x86_64.sh'

100%[=====>] 33,987,431 144KB/s in 2m 52s

2016-10-26 14:06:30 (193 KB/s) - 'Miniconda3-latest-Linux-x86_64.sh' saved [33987431/33987431]
```

- Lakukan instalasi dengan perintah

```
$ bash Miniconda3-latest-Linux-x86_64.sh
```

```
katiklabs@DESKTOP-LEES0AL:~$ bash Miniconda3-latest-Linux-x86_64.sh

Welcome to Miniconda3 4.1.11 (by Continuum Analytics, Inc.)

In order to continue the installation process, please review the license
```

- Tekan enter untuk menyelesaikan dialog lisensi penggunaan, sampai muncul *Do you approve the license terms? [yes|no]*

```
Do you approve the license terms? [yes|no]
>>>
```

- Sesuaikan lokasi instalasi. Tekan Enter untuk mengonfirmasi.

```
Please answer 'yes' or 'no':
>>> yes

Miniconda3 will now be installed into this location:
/home/katiklabs/miniconda3

- Press ENTER to confirm the location
- Press CTRL-C to abort the installation
- Or specify a different location below

[/home/katiklabs/miniconda3] >>>
```

- Setelah instalasi selesai, ketik yes untuk menambahkan lokasi instalasi ke PATH, supaya lokasi instalasi miniconda dapat dikenali sistem.



```
katiklabs@DESKTOP-LEES0AL: ~
- Press ENTER to confirm the location
- Press CTRL-C to abort the installation
- Or specify a different location below

[/home/katiklabs/miniconda3] >>> /home/katiklabs/miniconda3
PREFIX=/home/katiklabs/miniconda3
installing: python-3.5.2-0 ...
installing: conda-env-2.5.2-py35_0 ...
installing: openssl-1.0.2h-1 ...
installing: pycosat-0.6.1-py35_1 ...
installing: pyyaml-3.11-py35_4 ...
installing: readline-6.2-2 ...
installing: requests-2.10.0-py35_0 ...
installing: ruamel_yaml-0.11.14-py35_0 ...
installing: sqlite-3.13.0-0 ...
installing: tk-8.5.18-0 ...
installing: xz-5.2.2-0 ...
installing: yaml-0.1.6-0 ...
installing: zlib-1.2.8-3 ...
installing: conda-4.1.11-py35_0 ...
installing: pycrypto-2.6.1-py35_4 ...
installing: pip-8.1.2-py35_0 ...
installing: wheel-0.29.0-py35_0 ...
installing: setuptools-23.0.0-py35_0 ...
Python 3.5.2 :: Continuum Analytics, Inc.
creating default environment...
installation finished.
Do you wish the installer to prepend the Miniconda3 install location
to PATH in your /home/katiklabs/.bashrc ? [yes/no]
[no] >>> yes
```

- Exit console kemudian buka lagi.

## Running Python

Untuk menjalankan Python ada banyak cara yang dapat dilakukan. Bisa menggunakan shell, terminal atau IDE (Integrated Development Environment). Berikut ini adalah langkah-langkah untuk menjalankan Python pada terminal di Windows.

1. Buka terminal/command prompt.
2. Ketikkan `python` pada command line.
3. Untuk keluar, bisa menggunakan shortcut `ctrl + c` atau menuliskan perintah `exit()`

## Check The Versions of Libraries

```
>>>import sys    #python version
>>>print('Python: {}'.format(sys.version))
Python: 2.7.13 |Anaconda custom (64-bit)| (default, May 11
2017, 13:17:26) [MSC v.1500 64 bit (AMD64)]

>>>import scipy
>>>print('scipy: {}'.format(scipy.__version__))
scipy: 0.19.1

>>>import numpy
>>>print('numpy: {}'.format(numpy.__version__))
numpy: 1.13.1
```



```
>>>import matplotlib
>>>print('matplotlib: {}'.format(matplotlib.__version__))
matplotlib: 2.0.2
>>>import pandas
>>>print('pandas: {}'.format(pandas.__version__))
pandas: 0.20.3

>>>import sklearn
>>>print('sklearn: {}'.format(sklearn.__version__))
sklearn: 0.19.0
```

## Data Yang Digunakan

Dalam kegiatan praktikum pembelajaran mesin ini, dataset yang akan digunakan adalah dataset klasifikasi Bunga Iris. Dataset ini berisi 150 pengamatan bunga Iris. Terdapat empat kolom pengukuran bunga dalam centimeter. Kolom kelima adalah spesies bunga yang diamati. Dataset ini merupakan dataset yang paling umum digunakan, sehingga sering disebut sebagai dataset “hello word” dalam *machine learning*. Informasi detail mengenai dataset tersebut dapat diakses melalui [https://en.wikipedia.org/wiki/Iris\\_flower\\_data\\_set](https://en.wikipedia.org/wiki/Iris_flower_data_set).

## Memuat Dataset

Untuk memuat dataset dapat langsung diambil dari alamat repository UCI Machine Learning. Berikut ini adalah script untuk memuat dataset:

```
>>>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)
```

## 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 contoh dataset untuk task klasifikasi, kolom terakhir pada umumnya menunjukkan kelas datanya. Berikut ini adalah script untuk menentukan dimensi dari dataset:

```
>>>print(dataset.shape)
(150, 5)
```

Hal tersebut berarti dataset tersebut terdiri dari 150 sampel data dan 5 atribut data.

## II. Melihat Isi Dataset

Berikut perintah untuk melihat 20 baris data pertama:

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

Output:

	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

## III. Distribusi Kelas Data

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

Output:

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

## IV. Ringkasan Statistik

Berikut adalah perintah untuk mendeskripsikan statistik data meliputi count, mean, standard deviasi, nilai min, nilai max, dan quartile:

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

Output:

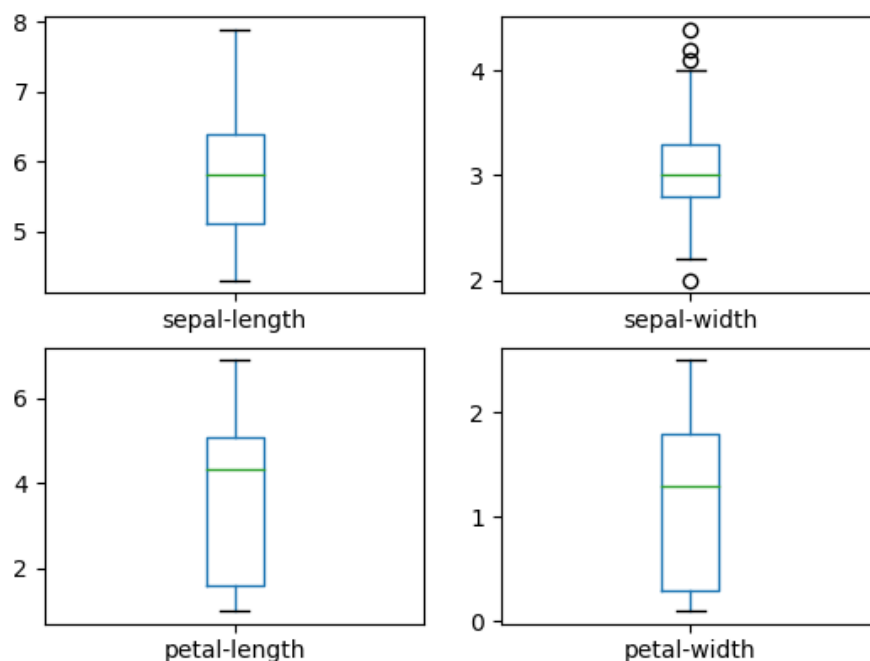
	sepal-length	sepal-width	petal-length	petal-width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000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

## V. 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.

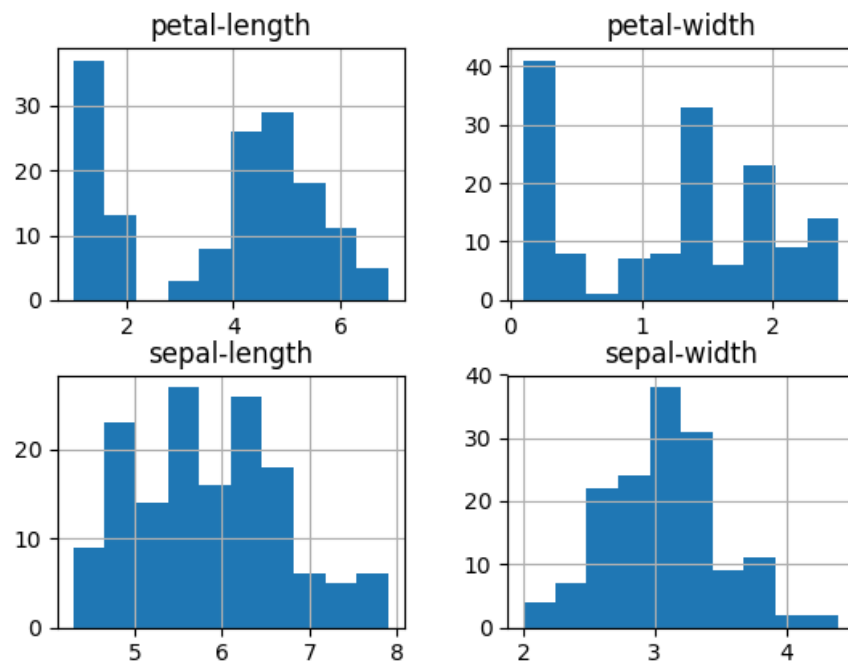
```
>>>import matplotlib.pyplot as plt
>>>dataset.plot(kind='box', subplots=True, layout=(2,2),
sharex=False, sharey=False)
>>>plt.show()
```

Output:



Selanjutnya juga bisa membuat histogram masing-masing variabel input untuk mendapatkan ide tentang distribusi.

```
>>>dataset.hist()
>>>plt.show()
```

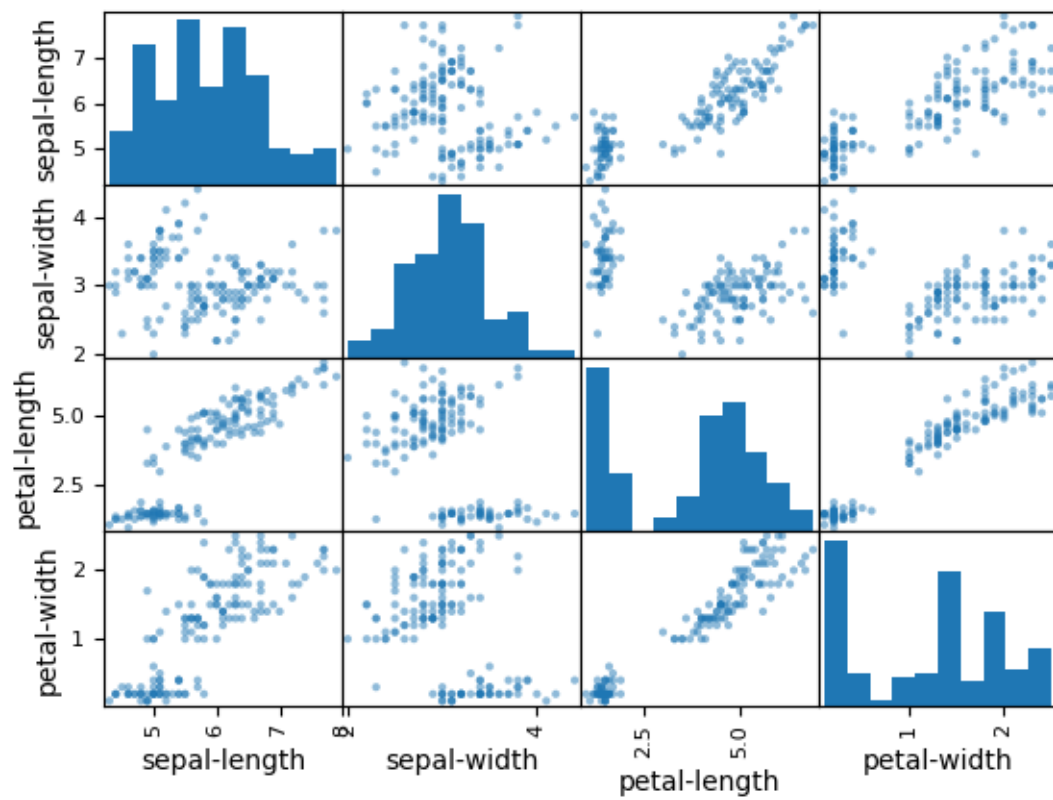


## VI. Visualisasi Data Menggunakan Plot Multivariat

Selanjutnya kita bisa melihat interaksi antar variabel. Pertama, kita lihat scatterplots dari semua pasang atribut. Hal ini dapat membantu melihat hubungan terstruktur antara variabel input.

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

Output:



## Latihan

Cari dataset dataset lain dari repository UCI Machine Learning dan lakukan keenam langkah untuk menampilkan *summary dataset*!

## References

1. <http://www.belajarpython.com/>
2. <https://www.quora.com/topic/Python-programming-language-1>
3. [https://medium.com/@haydar\\_ai/learning-data-science-day-1-environment-and-python-9076c82f0298#.h9sjjomkw](https://medium.com/@haydar_ai/learning-data-science-day-1-environment-and-python-9076c82f0298#.h9sjjomkw)