

# Python for Data Mining

## Preparing the Tools

Python menjadi salah satu bahasa pemrograman yang sangat digemari untuk proses data science/data mining dikarenakan kemudahan penggunaan, ketersediaan library yang lengkap dan komunitas yang baik.

## Python

Python bersifat *open source*, Anda dapat dengan mudah mengunduhnya di

<https://www.python.org/downloads/>



Looking for a specific release?

Python releases by version number:

Release version	Release date	Click for more	
<a href="#">Python 3.7.4</a>	July 8, 2019	<a href="#">Download</a>	<a href="#">Release Notes</a>
<a href="#">Python 3.6.9</a>	July 2, 2019	<a href="#">Download</a>	<a href="#">Release Notes</a>
<a href="#">Python 3.7.3</a>	March 25, 2019	<a href="#">Download</a>	<a href="#">Release Notes</a>
<a href="#">Python 3.4.10</a>	March 18, 2019	<a href="#">Download</a>	<a href="#">Release Notes</a>
<a href="#">Python 3.5.7</a>	March 18, 2019	<a href="#">Download</a>	<a href="#">Release Notes</a>
<a href="#">Python 2.7.16</a>	March 4, 2019	<a href="#">Download</a>	<a href="#">Release Notes</a>

**Download :** Pada kesempatan ini kita akan menggunakan python versi **3.8.3**. Pada bagian Release Version, scroll ke bawah, hingga Anda menemukan Python 3.8.3., kemudian klik tautan download.

Link Python 3.8.3 : <https://www.python.org/downloads/release/python-383/>

## Files

Version	Operating System	Description	MD5 Sum	File Size	GPG
<a href="#">Gzipped source tarball</a>	Source release		a7c10a2ac9d62de75a0ca5204e2e7d07	24067487	<a href="#">SIG</a>
<a href="#">XZ compressed source tarball</a>	Source release		3000cf50aaa413052aef82fd2122ca78	17912964	<a href="#">SIG</a>
<a href="#">macOS 64-bit installer</a>	Mac OS X	for OS X 10.9 and later	dd5e7f64e255d21f8d407f39a7a41ba9	30119781	<a href="#">SIG</a>
<a href="#">Windows help file</a>	Windows		4aeedbd7cc8dd90d61e7cfdda9cb9422	8568303	<a href="#">SIG</a>
<a href="#">Windows x86-64 embeddable zip file</a>	Windows	for AMD64/EM64T/x64	c12ffe7f4c1b447241d5d2aedc9b5d01	8175801	<a href="#">SIG</a>
<a href="#">Windows x86-64 executable installer</a>	Windows	for AMD64/EM64T/x64	fd2458fa0e9ead1dd9fbc2370a42853b	27805800	<a href="#">SIG</a>
<a href="#">Windows x86-64 web-based installer</a>	Windows	for AMD64/EM64T/x64	17e989d2fecf7f9f13cf987825b695c4	1364136	<a href="#">SIG</a>
<a href="#">Windows x86 embeddable zip file</a>	Windows		8ee09403ec0cc2e89d43b4a4f6d1521e	7330315	<a href="#">SIG</a>
<a href="#">Windows x86 executable installer</a>	Windows		452373e2c467c14220efeb10f40c231f	26744744	<a href="#">SIG</a>
<a href="#">Windows x86 web-based installer</a>	Windows		fe72582bbca3dbe07451fd05ece1d752	1325800	<a href="#">SIG</a>

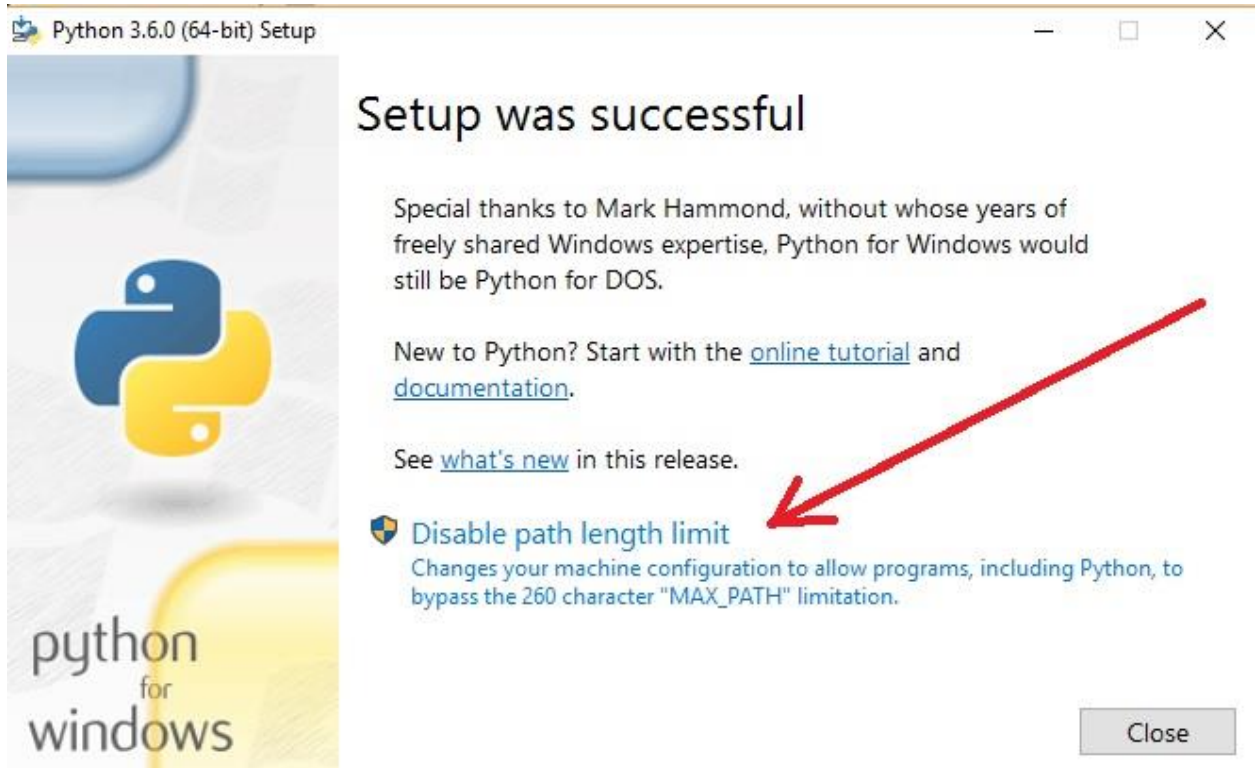
Pilih installer sesuai operating sistem Anda, perhatikan tipe bit installer.

## Install :

1. Pada saat instalasi beri tanda centang pada **Add Python 3.8 to PATH**. Kemudian klik **Install Now**



2. Saat instalasi berakhir, **jika** muncul pesan **Disable path length limit**, pastikan anda memilihnya dengan cara klik.



## Buat Environment (Windows)

Untuk menghindari konflik versi yang berbeda pada setiap library, Anda memerlukan membuat environment Python. Langkah-langkah membuat environment dalam Python adalah sebagai berikut :

1. Buat folder baru, misalnya dalam contoh kali ini adalah di **E:\DataMining**
2. Buka command prompt, masuk ke direktori folder yang baru Anda buat. Misalnya jika di direktori E, ketik **E:** dan tekan Enter.

```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.18362.356]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\Satria>E:

E:\>
```

3. Setelah masuk ke direktori, ketik **python -m venv E:\DataMining**

```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.18362.356]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\Satria>E:

E:\>python -m venv E:\DataMining

E:\>
```

4. Setelah installasi berhasil, maka folder yang Anda buat akan berisi data seperti di bawah ini :

> This PC > Data (E:) > DataMining				
Name		Date modified	Type	Size
✦	Include	30/09/2019 13:10	File folder	
	Lib	30/09/2019 13:10	File folder	
	Scripts	30/09/2019 13:10	File folder	
	pyvenv.cfg	30/09/2019 13:10	CFG File	1 KB

5. Selamat! Anda telah berhasil membuat environment di Python

# Library di Python

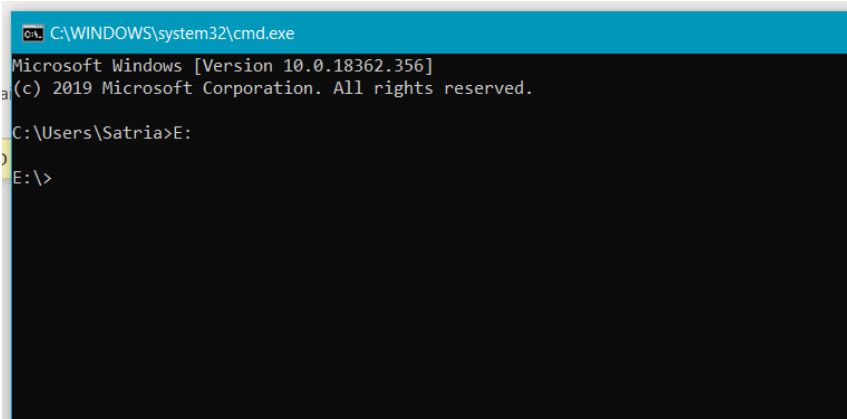
---

Alasan utama Python menjadi tool yang paling sering digunakan oleh data scientist adalah banyaknya library *machine learning/artificial intelligence* terkini. Anda dapat dengan mudah menggunakannya dengan **Python Package Indexing (PIP)**.

Beberapa library yang akan kita gunakan adalah Numpy (pemrosesan matriks, aljabar linier), Pandas (pemrosesan dataset), Matplotlib (visualisasi data), Seaborn (wrapping dari Matplotlib, visualisasi data dengan kode lebih ringkas), Scikit-learn (kumpulan implementasi algoritma data mining) dan Jupyter Notebook (web application untuk python code editor).

Adapun langkah-langkahnya adalah sebagai berikut :

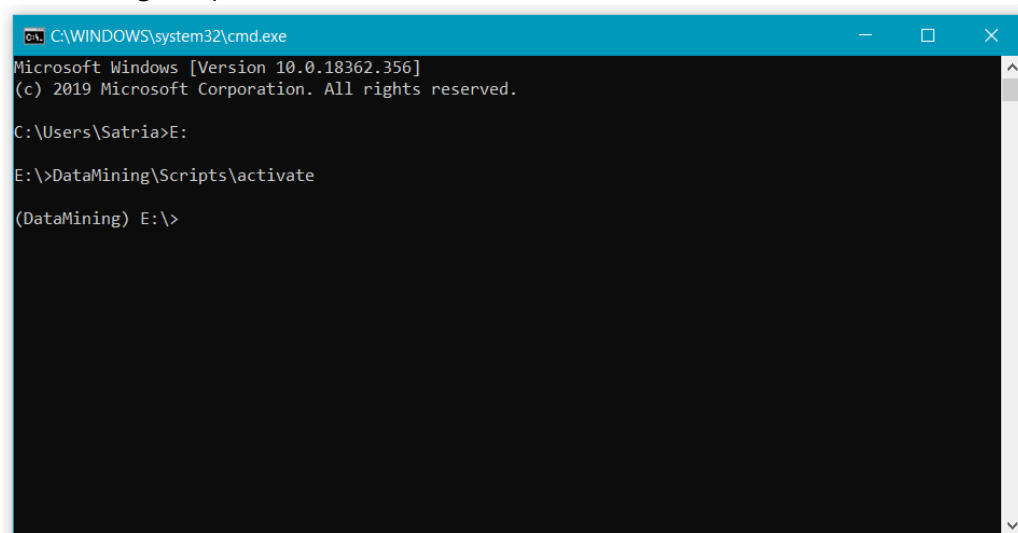
1. Buka command prompt, pindah ke direktori environment Anda.



```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.18362.356]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\Satria>E:
E:\>
```

2. Kemudian aktifkan environment yang Anda buat dengan cara mengetikkan DataMining\Scripts\activate dan tekan Enter



```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.18362.356]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\Satria>E:
E:\>DataMining\Scripts\activate
(DataMining) E:\>
```

3. Sebelum memulai instalasi library, update pip dengan command berikut pada command prompt : ***python -m pip install --upgrade pip***

4. Jika command prompt telah berubah seperti gambar di atas, maka environment Anda telah aktif. Kemudian Anda bisa menginstall library yang dibutuhkan satu persatu. Untuk instalasi library, Anda harus terkoneksi ke internet. Ketik command di bawah ini satu persatu:

***pip install numpy***

***pip install matplotlib***

***pip install pandas***

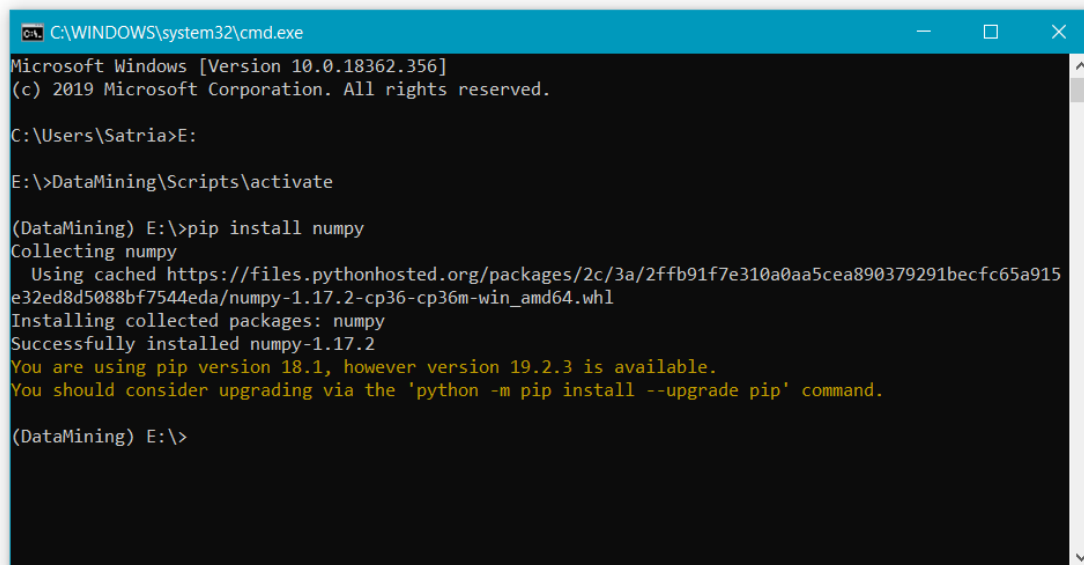
***pip install seaborn***

***pip install scikit-learn***

***pip install jupyterlab***

***pip install Flask***

Contoh instalasi package yang berhasil dapat dilihat pada gambar di bawah



```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.18362.356]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\Satria>E:

E:\>DataMining\Scripts\activate

(DataMining) E:\>pip install numpy
Collecting numpy
  Using cached https://files.pythonhosted.org/packages/2c/3a/2ffb91f7e310a0aa5cea890379291becfc65a915e32ed8d5088bf7544eda/numpy-1.17.2-cp36-cp36m-win_amd64.whl
Installing collected packages: numpy
Successfully installed numpy-1.17.2
You are using pip version 18.1, however version 19.2.3 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.

(DataMining) E:\>
```

5. Jika Anda telah selesai menggunakan environment, ketik **deactivate** pada command prompt dan tekan Enter



*Jika koneksi internet Anda tidak stabil, biasanya instalasi akan memunculkan pesan error. Ulangi command yang error pada koneksi internet yang stabil.*