

Lembar Kerja Praktikum
Mata Kuliah : Data Mining (KOM 1338)

Semester Genap 2025/2026
Pertemuan 1

A. Mengenal Python untuk Data Mining

Python adalah salah satu *scripting language* bisa dijalankan melalui berbagai sistem operasi. Saat ini, Python juga merupakan bahasa yang populer bagi bidang *data science* dan analisis. Hal ini dikarenakan oleh dukungan bahasa Python terhadap banyak *library* yang ada di dalamnya serta menyediakan fungsi analisis data dan fungsi *machine learning*, *deep learning* *data preprocessing tools*, serta visualisasi data

Python banyak dipakai pada data mining karena ekosistem library yang lengkap untuk:

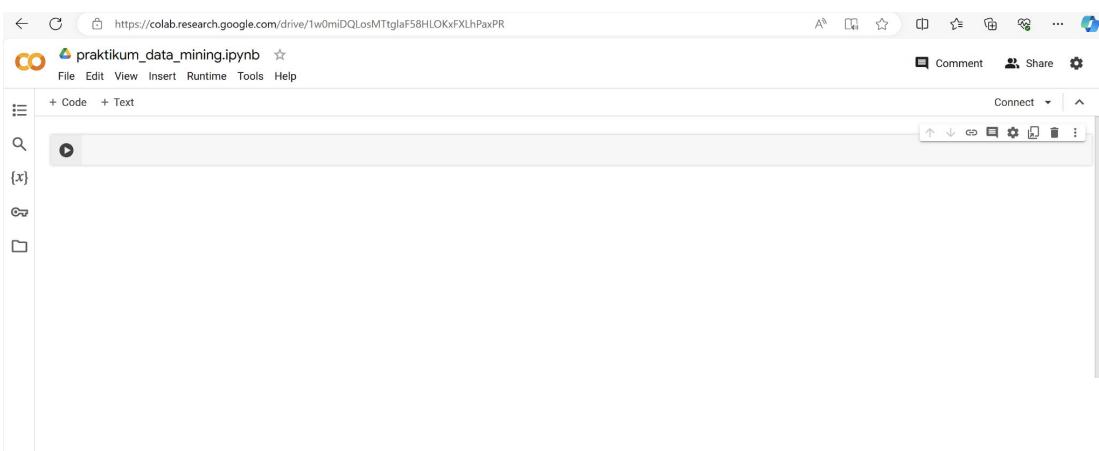
- manipulasi data (NumPy, Pandas),
- visualisasi (Matplotlib, Seaborn, Plotly),
- machine learning (Scikit-learn),
- deep learning (PyTorch/TensorFlow, opsional),
- serta dapat berjalan di CPU maupun GPU (terutama via Colab).

Untuk dapat menggunakan Python, ada dua alternatif cara:

- menggunakan **notebook** pada Google Collab (<https://colab.research.google.com/>).
Langkah:

- Buka Google Colab Collab (<https://colab.research.google.com/>).
- Login/Sign-In ke Akun Google/Google Apps Anda.
- Klik New Notebook.
- Ubah judul notebook: DM_P1_NamaNIM.ipynb
- Pastikan runtime bisa jalan: klik Runtime → Run all (nanti setelah kode diisi).

Gambar 1 adalah tampilan Google Collab.

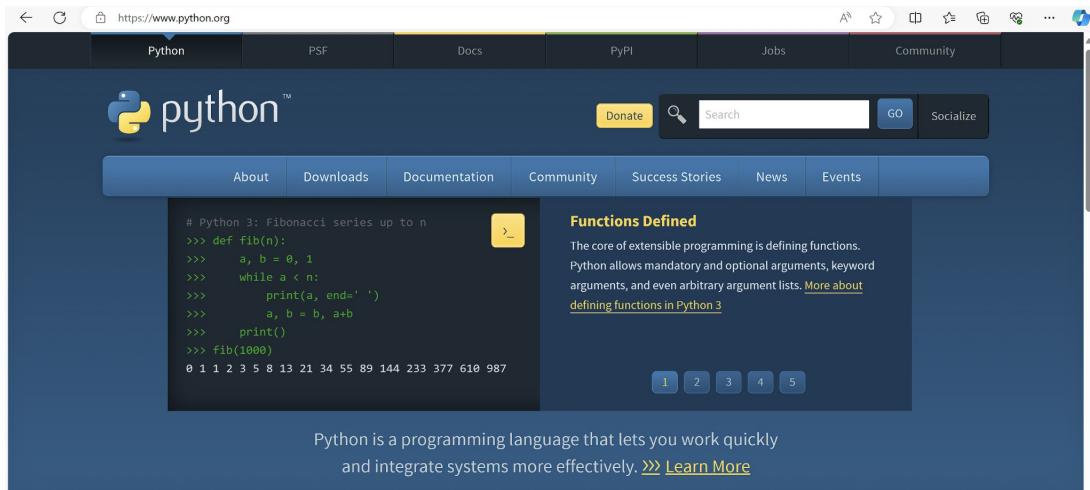


Gambar 1 . Tampilan Google Collab

- **Instalasi Python**

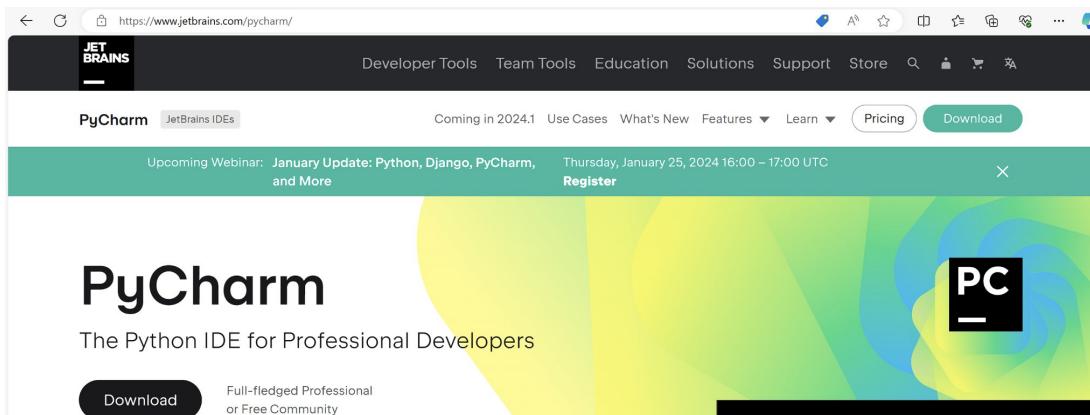
Alamat resmi Python dapat diakses pada alamat : <https://www.python.org/>. Karena Python dapat dijalankan pada berbagai sistem operasi, kita dapat menentukan format file instalasi

Python sesuai dengan sistem operasi yang digunakan. Gambar 3 menampilkan halaman utama website Python. IDLE adalah salah IDE yang dapat digunakan sebagai teks editor untuk menjalankan Python.



Gambar 3. Tampilan halaman resmi Python

Selain IDLE, Pycharm juga dapat digunakan untuk menjalankan Python yang dapat diakses pada alamat <https://www.jetbrains.com/pycharm/>. Gambar 4 menampilkan halaman utama dari Pycharm



Gambar 4. Halaman Utama PyCharm IDE

Hingga saat ini Python masih populer digunakan di bidang data sains dan data mining setidaknya dikarenakan beberapa alasan :

1. Memiliki ketersediaan *open source library, framework, tools* yang memberikan kemudahan pengguna. Beberapa diantaranya adalah Scikit-learn, tensorflow, torch, keras, theano dan lain-lain
2. Kode bahasa python relatif mudah dipahami dan simple sehingga memudahkan belajar bagi pengguna pemula
3. Selain untuk pemrosesan data, juga dapat digunakan untuk membuat antarmuka (GUI).

Tugas 1: Silakan lakukan salah satu dari dua cara di atas untuk menjalankan lingkungan Python, lalu tangkap/screen-capture dan isikan hasilnya ke lembar di bawah ini!

Library Python untuk Data Mining

Aktivitas / pekerjaan data mining dan library Python yang dapat digunakan pada data mining dijelaskan pada Tabel 1

Tabel 1. Aktivitas / Data mining dan Library Python yang digunakan

No	Aktivitas	Library
1	Pengolahan Row Data	numpy, pandas
2	Eksplorasi Data	pandas, matplotlib, seaborn
3	Pre-Processing Data	pandas
4	Penggunaan Teknik Klasifikasi	Scikit-learn
5	Penggunaan Teknik Clustering	Scikit-learn
6	Penggunaan Teknik Asosiasi	apyori , pycaret
7	Penggunaan Teks Mining	nltk, textblob,
8	Data Warehouse	olapy
9	Melakukan Evaluasi	sklearn.metrics
10	Melakukan Visualisasi	matplotlib, seaborn, plotly

Tugas 2: Ikuti Langkah-langkah berikut untuk mencoba lingkungan Python serta library-nya, tangkap/screenshot hasilnya lalu laporkan !

Aktivitas 1 — “Hello Data Mining” di Notebook

Buat cell di Colab (atau Jupyter lokal), lalu jalankan:

```
print("Hello Data Mining!")  
import sys  
print("Python version:", sys.version)
```

Aktivitas 2 — Import Library Inti

Jalankan kode di bawah ini:

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns


from sklearn.model_selection import
train_test_split

from sklearn.preprocessing import
StandardScaler

from sklearn.metrics import accuracy_score,
classification_report, confusion_matrix

from sklearn.linear_model import
LogisticRegression
```

Cek versi (pastikan tidak ada error):

```
import sklearn

print("numpy:", np.__version__)
print("pandas:", pd.__version__)
print("sklearn:", sklearn.__version__)
```

kompetisi Data Mining Tingkat Nasional dan Internasional

a. Gemastik

GEMASTIK atau Pagelaran Mahasiswa Nasional Bidang Teknologi Informasi dan Komunikasi, merupakan program Pusat Prestasi Nasional, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. Program ini ditujukan untuk meningkatkan kompetensi mahasiswa Indonesia. Gemastik sampai saat ini dilaksanakan secara rutin setiap tahun. Terdapat berbagai bidang/divisi yang dilombakan pada ajang gemastik, antara lain :

1. Divisi Pemrograman
2. Divisi Keamanan Siber
3. **Disis Penambangan Data (Data Mining)**
4. Divisi Desain Pengalaman Pengguna
5. Divisi Animasi
6. Divisi Kota Cerdas
7. Divisi Karya Tulis TIK
8. Divisi Pengembangan Perangkat Lunak
9. Divisi Pengembangan Piranti Cerdas
10. Divisi Pengebangaan Bisnis TIK
11. Divisi Pengembangan Aplikasi Permainan

Informasi mengenai gemastik dapat dilihat pada alamat: <https://gemastik.kemdikbud.go.id/>

b. Satria Data

Satria Data merupakan program yang diselenggarakan oleh Kementerian setiap tahun untuk memperkenalkan statistika dan sains data, menumbuhkan minat terhadap sains data serta meningkatkan keahlian dan pengetahuan di bidang statistika dan sains data. Terdapat empat kategori di ajang satria data ini diantaranya

1. National Statistic Competition
2. Statistic Essay Competition
3. Statistic Infographics Competition
4. **Big Data Challenge**

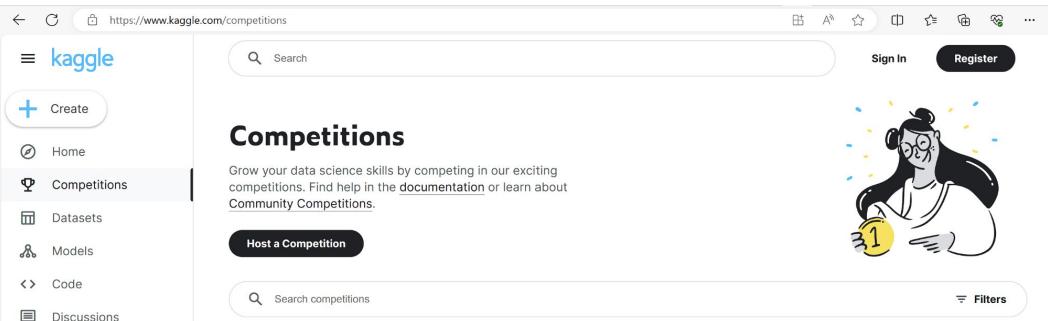
Kategori yang relevan dengan perkuliahan Data Mining adalah Big Data Challenge (BDC).

Informasi detail mengenai Satria Data dapat dilihat pada alamat

<https://satriadata.kemdikbud.go.id/>

B. Kaggle Competition

Mahasiswa akan mendapatkan tugas secara berkelompok untuk berkompetisi menyelesaikan permasalahan dengan pendekatan data mining pada Kaggle (Gambar 2) (<https://www.kaggle.com/competitions>). Setiap Mahasiswa akan mendapatkan penugasan dari dosen untuk menyelesaikan permasalahan dan dapat melihat hasilnya pada scoreboard.



Gambar 2. Halaman Kaggle Competition

Aktivitas 3: Setiap mahasiswa membuat akun di Kaggle Competition, pastikan nama dan username yang digunakan sesuai dengan nama resmi, serta gunakan email Apps IPB. Tampilkan screenshot hasil registrasi di lembar jawaban!

LEMBAR JAWABAN

Nama: Muhammad Farhadh

NIM: G6401231080

Nama Asprak/NIM Asprak:

1. Ryan Muhammad Syahran G6401221026
2. Farhan Lado Anggaraksa Mascahyanto G6401221047

TUGAS 1: Python environment dengan nix flake. Editor: Neovim, Shell: Bash/zsh

```
daming = pkgs.mkShell {
  packages = [
    (python.withPackages (py:[
      py.numpy
      py.pandas
      py.matplotlib
      py.scikit-learn
      py.seaborn
      py.plotly
      py.textblob

      # Manual Derivation
      # py.pycaret
      # py.apyori
      # py.olap
    ]))
  ];
}
```

TUGAS 2

Aktivitas 1:

```
lele 33 ~/dev/semester6/daming/prak1:
2036 o cat aktivitas1.py \
&& echo 'output:' \
&& python aktivitas1.py
print("hello data mining!")
import sys
print("python version:", sys.version)
output:
hello data mining!
python version: 3.12.12 (main, Oct 9 2025, 11:07:00) [GCC 15.2.0]
```

Aktivitas 2:

```
lele ~ ~/dev/semester6/daming/prak1:  
1801 o cat aktivitas2.py \  
&& echo 'output:' \  
&& python aktivitas2.py  
import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
  
from sklearn.preprocessing import StandardScaler  
from sklearn.model_selection import train_test_split  
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix  
from sklearn.linear_model import LogisticRegression  
  
import sklearn  
print("numpy:", np.__version__)  
print("pandas:", pd.__version__)  
print("sklearn", sklearn.__version__)  
  
output:  
numpy: 2.3.5  
pandas: 2.3.3  
sklearn 1.7.2
```

Aktivitas 3:

Settings

Control over your Kaggle account and all communications

Account Notifications

Your email address

2905muhammad@apps.ipb.ac.id

Change email

Your username

muhammadfarhadh1 (not editable)

Your account number

32024702

Phone verification

Your account is not verified. Verifying your account with a phone number allows you to do more on Kaggle, and helps prevent spam and other abuse.

Phone verify

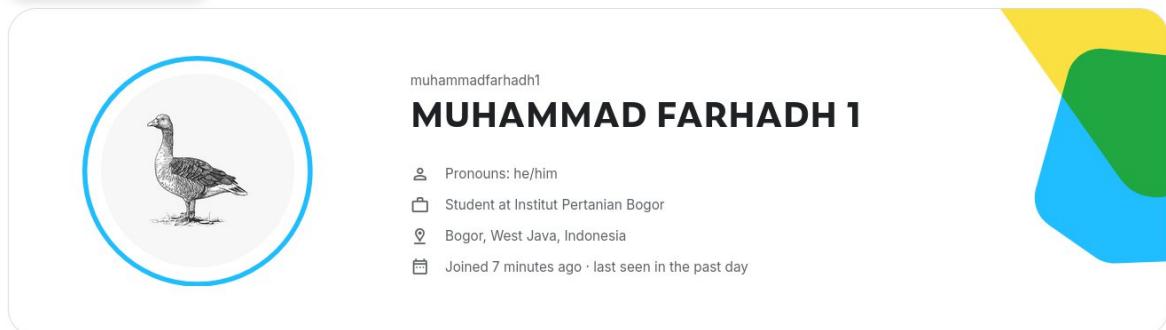
Identity verification

You have not verified your identity using Persona, a trusted 3rd-party service. Verifying your identity allows you to join competitions that require identity verification. [Learn More](#)

Verify my account

Edit your public profile

Settings Your Work Progression



A screenshot of a Kaggle user profile page. The profile picture is a circular illustration of a duck. The username is "muhammadfarhadh1" and the name is "MUHAMMAD FARHADH 1". Below the name, there are four profile details: "Pronouns: he/him", "Student at Institut Pertanian Bogor", "Bogor, West Java, Indonesia", and "Joined 7 minutes ago · last seen in the past day". To the right of the profile is a decorative graphic of overlapping colored shapes (yellow, green, blue). At the bottom of the profile section, there are "About", "Bio", "Follow", and "Contact" buttons.

About

Bio

Halo, Saya menggunakan WhatsApp!

Follow

Contact