

Tugas 1 Data Sains dan Analisis

1. Ambil data dari Kaggle dengan API
*menggunakan instruksi yang ada di [V.2]Pertemuan 5_Contoh Kasus_Microcredential

```

tampildata1.py - py - Visual Studio Code
tampldata1.py > ...
1 import pandas as pd

NOVO\py> pip install numpy
Requirement already satisfied: numpy in c:\users\lenovo\anaconda3\lib\site-packages (1.21.5)
PS C:\Users\LENOVO\py> pip install opendatasets
Collecting opendatasets
  Downloading opendatasets-0.1.22-py3-none-any.whl (15 kB)
Requirement already satisfied: click in c:\users\lenovo\anaconda3\lib\site-packages (from opendatasets) (8.0.4)
Requirement already satisfied: kaggle in c:\users\lenovo\anaconda3\lib\site-packages (from opendatasets) (1.5.13)
Requirement already satisfied: tqdm in c:\users\lenovo\anaconda3\lib\site-packages (from opendatasets) (4.64.1)
Requirement already satisfied: colorama in c:\users\lenovo\appdata\roaming\python\python39\site-packages (from click->opendatasets) (0.4.4)
Requirement already satisfied: six>=1.10 in c:\users\lenovo\appdata\roaming\python\python39\site-packages (from kaggle->opendatasets) (1.15.0)
Requirement already satisfied: urllib3 in c:\users\lenovo\anaconda3\lib\site-packages (from kaggle->opendatasets) (1.26.11)
s (from python-slugify->kaggle->opendatasets) (1.3)
Requirement already satisfied: idna<4,>=2.5 in c:\users\lenovo\anaconda3\lib\site-packages (from requests->kaggle->opendatasets) (3.3)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\lenovo\anaconda3\lib\site-packages (from requests->kaggle->opendatasets) (2.0.4)
Installing collected packages: opendatasets
Successfully installed opendatasets-0.1.22
PS C:\Users\LENOVO\py>
  History restored

PS C:\Users\LENOVO\py> C:/Users/LENOVO/anaconda3/Scripts/activate
PS C:\Users\LENOVO\py> conda activate base
PS C:\Users\LENOVO\py> pip install kaggle
Requirement already satisfied: kaggle in c:\users\lenovo\anaconda3\lib\site-packages (1.5.13)
Requirement already satisfied: urllib3 in c:\users\lenovo\anaconda3\lib\site-packages (from kaggle) (1.26.11)
-----
piantc/plantpathology-apple-dataset      PlantPathology Apple Dataset      813MB  2020-04-24 13:45:22      1035      37  0.
5882353
aqeeljajja/whitefly-leave-classification  AgriPK dataset for Whitefly affected leaves      8GB   2021-12-14 17:58:25      36        2  0.
8125
aqeeljajja/white-fly-cotton-sample       White_Fly_Cotton_Sample            122MB 2021-12-10 12:26:04      5         1  0.
125
  
```

```

tampildata1.py - py - Visual Studio Code
tampldata1.py > ...
1 import pandas as pd

Requirement already satisfied: colorama in c:\users\lenovo\appdata\roaming\python\python39\site-packages (from click->opendatasets) (0.4.4)
Requirement already satisfied: six>=1.10 in c:\users\lenovo\appdata\roaming\python\python39\site-packages (from kaggle->opendatasets) (1.15.0)
Requirement already satisfied: urllib3 in c:\users\lenovo\anaconda3\lib\site-packages (from kaggle->opendatasets) (1.26.11)
s (from python-slugify->kaggle->opendatasets) (1.3)
Requirement already satisfied: idna<4,>=2.5 in c:\users\lenovo\anaconda3\lib\site-packages (from requests->kaggle->opendatasets) (3.3)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\lenovo\anaconda3\lib\site-packages (from requests->kaggle->opendatasets) (2.0.4)
Installing collected packages: opendatasets
Successfully installed opendatasets-0.1.22
PS C:\Users\LENOVO\py>
  History restored

PS C:\Users\LENOVO\py> C:/Users/LENOVO/anaconda3/Scripts/activate
PS C:\Users\LENOVO\py> conda activate base
PS C:\Users\LENOVO\py> pip install kaggle
Requirement already satisfied: kaggle in c:\users\lenovo\anaconda3\lib\site-packages (1.5.13)
Requirement already satisfied: urllib3 in c:\users\lenovo\anaconda3\lib\site-packages (from kaggle) (1.26.11)
-----
piantc/plantpathology-apple-dataset      PlantPathology Apple Dataset      813MB  2020-04-24 13:45:22      1035      37  0.
5882353
aqeeljajja/whitefly-leave-classification  AgriPK dataset for Whitefly affected leaves      8GB   2021-12-14 17:58:25      36        2  0.
8125
aqeeljajja/white-fly-cotton-sample       White_Fly_Cotton_Sample            122MB 2021-12-10 12:26:04      5         1  0.
125
aqeeljajja/white-fly-detection-complete  White_Fly_Detection_Complete      8GB   2022-01-02 19:28:45      1         1  0.
125
PS C:\Users\LENOVO\py> kaggle datasets download piantc/plantpathology-apple-dataset --unzip
Downloading plantpathology-apple-dataset.zip to C:\Users\LENOVO\py
100% |#####| 813M/813M [02:26:00:00, 6.30MB/s]
PS C:\Users\LENOVO\py>
  
```

Menampilkan data:

The screenshot shows the Visual Studio Code interface with a Python file named `tampildata1.py` open. The script imports `pandas` and `numpy`, sets a `dataset` variable to `"train.csv"`, reads the CSV file into a `df` object, and prints the first 10 rows using `df.head(10)`. The output in the terminal shows a table with 5 columns: `image_id`, `healthy`, `multiple_diseases`, `rust`, and `scab`. The rows are indexed from 0 to 9.

```

1 import pandas as pd
2 import numpy as df
3
4 dataset = "train.csv"
5 df = pd.read_csv(dataset)
6
7 print(df.head(10))

```

	image_id	healthy	multiple_diseases	rust	scab
0	Train_0	0	0	0	1
1	Train_1	0	1	0	0
2	Train_2	1	0	0	0
3	Train_3	0	0	1	0
4	Train_4	1	0	0	0
5	Train_5	1	0	0	0
6	Train_6	0	1	0	0
7	Train_7	0	0	0	1
8	Train_8	0	0	0	1
9	Train_9	1	0	0	0

[Done] exited with code=0 in 1.456 seconds

Isi data dalam csv dan ouput

The screenshot shows the Visual Studio Code interface with the `train.csv` file open. The file contains 12 rows of data, including a header row. The output in the terminal is identical to the previous screenshot, showing the first 10 rows of the CSV file.

```

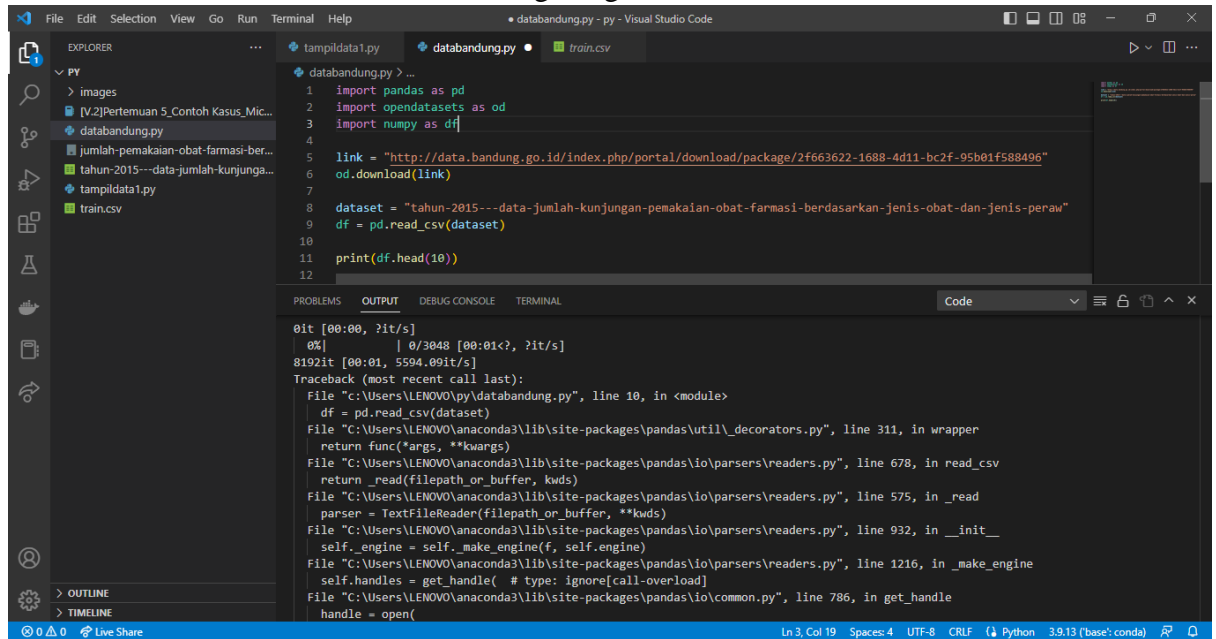
1 image_id,healthy,multiple_diseases,rust,scab
2 Train_0,0,0,0,1
3 Train_1,0,1,0,0
4 Train_2,1,0,0,0
5 Train_3,0,0,1,0
6 Train_4,1,0,0,0
7 Train_5,1,0,0,0
8 Train_6,0,1,0,0
9 Train_7,0,0,0,1
10 Train_8,0,0,0,1
11 Train_9,1,0,0,0
12 Train_10,0,0,1,0

```

	image_id	healthy	multiple_diseases	rust	scab
0	Train_0	0	0	0	1
1	Train_1	0	1	0	0
2	Train_2	1	0	0	0
3	Train_3	0	0	1	0
4	Train_4	1	0	0	0
5	Train_5	1	0	0	0
6	Train_6	0	1	0	0
7	Train_7	0	0	0	1
8	Train_8	0	0	0	1
9	Train_9	1	0	0	0

[Done] exited with code=0 in 1.456 seconds

2. Ambil data dari Portal Satu Data Bandung dengan API



The screenshot shows a Visual Studio Code editor with a Python file named `databandung.py`. The script imports `pandas`, `opendatasets`, and `numpy`. It defines a `link` variable pointing to a URL on the `data.bandung.go.id` portal. The script then uses `od.download(link)` to fetch the data, stores it in a `dataset` variable, and reads it into a `df` using `pd.read_csv(dataset)`. Finally, it prints the first 10 rows of the dataset using `print(df.head(10))`.

The output window shows the execution of the script, which results in a `FileNotFoundError` exception. The traceback indicates that the error occurred in the `pd.read_csv(dataset)` call at line 10 of `databandung.py`. The error message is `FileNotFoundError: [Errno 2] No such file or directory: 'C:\\Users\\LENOVO\\py\\databandung.py', line 10, in <module> df = pd.read_csv(dataset)`.

```
File "C:\\Users\\LENOVO\\py\\databandung.py", line 10, in <module>
    df = pd.read_csv(dataset)
File "C:\\Users\\LENOVO\\anaconda3\\lib\\site-packages\\pandas\\util\\decorators.py", line 311, in wrapper
    return func(*args, **kwargs)
File "C:\\Users\\LENOVO\\anaconda3\\lib\\site-packages\\pandas\\io\\parsers\\readers.py", line 678, in read_csv
    return _read(filepath_or_buffer, kwds)
File "C:\\Users\\LENOVO\\anaconda3\\lib\\site-packages\\pandas\\io\\parsers\\readers.py", line 575, in _read
    parser = TextFileReader(filepath_or_buffer, **kwargs)
File "C:\\Users\\LENOVO\\anaconda3\\lib\\site-packages\\pandas\\io\\parsers\\readers.py", line 932, in __init__
    self._engine = self._make_engine(f, self.engine)
File "C:\\Users\\LENOVO\\anaconda3\\lib\\site-packages\\pandas\\io\\parsers\\readers.py", line 1216, in _make_engine
    self._handles = get_handle( # type: ignore[call-overload]
File "C:\\Users\\LENOVO\\anaconda3\\lib\\site-packages\\pandas\\io\\common.py", line 786, in get_handle
    handle = open(
```

Masih terjadi eror pada saat pengambilan data dan menampilkan data