



Hands-On

Hands-On ini digunakan pada kegiatan Microcredential Associate Data Scientist 2021

Pertemuan 5

Pertemuan 5 (lima) pada Microcredential Associate Data Scientist 2021 menyampaikan materi mengenai Mengumpulkan Data, Menelaah Data dengan metode Statistik

Pengambilan Data dari API Kaggle

Salah satu portal yang menyediakan dataset untuk project Data Science adalah Kaggle (<https://www.kaggle.com/>). Pada latihan ini, silakan peserta mengunduh dataset mengenai bunga Iris dengan menggunakan kata kunci: "iris species" yang disediakan oleh UCI Machine Learning (UCIML)

1. Install Modul kaggle:

In [1]:

```
# Install modul kaggle secara inline (di dalam notebook)
!pip install kaggle
```

```
Collecting kaggle
  Downloading kaggle-1.5.12.tar.gz (58 kB)
    Preparing metadata (setup.py): started
    Preparing metadata (setup.py): finished with status 'done'
Requirement already satisfied: six>=1.10 in c:\users\ramat\miniconda3\envs\mcds\lib\site-packages (from kaggle) (1.16.0)
Requirement already satisfied: certifi in c:\users\ramat\miniconda3\envs\mcds\lib\site-packages (from kaggle) (2021.10.8)
Requirement already satisfied: python-dateutil in c:\users\ramat\miniconda3\envs\mcds\lib\site-packages (from kaggle) (2.8.2)
Collecting requests
  Downloading requests-2.26.0-py2.py3-none-any.whl (62 kB)
Requirement already satisfied: tqdm in c:\users\ramat\miniconda3\envs\mcds\lib\site-packages (from kaggle) (4.62.3)
Collecting python-slugify
  Downloading python_slugify-5.0.2-py2.py3-none-any.whl (6.7 kB)
Collecting urllib3
  Downloading urllib3-1.26.7-py2.py3-none-any.whl (138 kB)
Collecting text-unidecode>=1.3
  Downloading text_unidecode-1.3-py2.py3-none-any.whl (78 kB)
Collecting idna<4,>=2.5
  Downloading idna-3.3-py3-none-any.whl (61 kB)
Collecting charset-normalizer~=2.0.0
```

```

  Downloading charset_normalizer-2.0.7-py3-none-any.whl (38 kB)
Requirement already satisfied: colorama in c:\users\ramat\miniconda3\envs\mcds\lib\site-packages (from tqdm->kaggle) (0.4.4)
Building wheels for collected packages: kaggle
  Building wheel for kaggle (setup.py): started
  Building wheel for kaggle (setup.py): finished with status 'done'
  Created wheel for kaggle: filename=kaggle-1.5.12-py3-none-any.whl size=73051 sha256=ebd50fa41a30612527920c68cff9fdabbe2892af3805d416cad7e770b06c3a65
  Stored in directory: c:\users\ramat\appdata\local\pip\cache\wheels\62\d6\58\5853130f941e75b2177d281eb7e44b4a98ed46dd155f556dc5
Successfully built kaggle
Installing collected packages: urllib3, text-unidecode, idna, charset-normalizer, requests, python-slugify, kaggle
Successfully installed charset-normalizer-2.0.7 idna-3.3 kaggle-1.5.12 python-slugify-5.0.2 requests-2.26.0 text-unidecode-1.3 urllib3-1.26.7

```

In [2]:

```
# Install modul kaggle secara eksternal melalui anaconda prompt:
```

```

Administrator: Anaconda Prompt (Anaconda3)
(base) C:\WINDOWS\system32 pip install kaggle
Collecting kaggle
  Using cached kaggle-1.5.12-py3-none-any.whl
Requirement already satisfied: six>=1.10 in c:\programdata\anaconda3\lib\site-packages (from kaggle) (1.15.0)
Requirement already satisfied: requests in c:\programdata\anaconda3\lib\site-packages (from kaggle) (2.25.1)
Requirement already satisfied: tqdm in c:\programdata\anaconda3\lib\site-packages (from kaggle) (4.59.0)
Requirement already satisfied: python-dateutil in c:\programdata\anaconda3\lib\site-packages (from kaggle) (2.8.1)
Requirement already satisfied: python-slugify in c:\programdata\anaconda3\lib\site-packages (from kaggle) (5.0.2)
Requirement already satisfied: certifi in c:\programdata\anaconda3\lib\site-packages (from kaggle) (2020.12.5)
Requirement already satisfied: urllib3 in c:\programdata\anaconda3\lib\site-packages (from kaggle) (1.26.4)
Requirement already satisfied: text-unidecode>=1.3 in c:\programdata\anaconda3\lib\site-packages (from python-slugify->kaggle) (1.3)
Requirement already satisfied: idna<3,>=2.5 in c:\programdata\anaconda3\lib\site-packages (from requests->kaggle) (2.10)
Requirement already satisfied: chardet<5,>=3.0.2 in c:\programdata\anaconda3\lib\site-packages (from requests->kaggle) (4.0.0)
Installing collected packages: kaggle
Successfully installed kaggle-1.5.12

```

2. Create Token API kaggle:

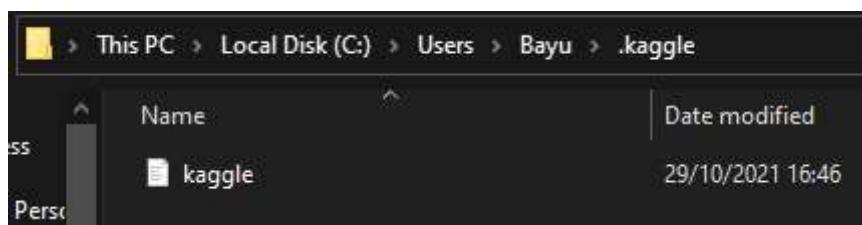
Home Competitions (1) Datasets Code (1) Discussion Followers Notifications Account

API

Using Kaggle's beta API, you can interact with Competitions and Datasets to download data, make submissions, and more via command line. [Read the docs](#)

[Create New API Token](#) [Expire API Token](#)

1. Login Kaggle.com
2. Kemudian pada menu Profile --> Account
3. Klik Create New Api Token
4. Maka akan terdownload file kaggle.json



Kaggle API secara default mengasumsikan bahwa file kaggle.json tersebut berada di dalam folder:

- ~/kaggle/ (Linux/Mac)
- C:\Users\kaggle\ (Windows)

Jika folder tersebut belum ada:

1. Buat folder di direktori C:\Users\kaggle\
2. letakkan file kaggle.json kedalam folder tersebut

3. Download Dataset dari Kaggle:

```
usage: kaggle datasets [-h]
                      {list,files,download,create,version,init,metadata,status} ...

optional arguments:
  -h, --help            show this help message and exit

commands:
  {list,files,download,create,version,init,metadata, status}
    list                 List available datasets
    files                List dataset files
    download             Download dataset files
    create               Create a new dataset
    version              Create a new dataset version
    init                Initialize metadata file for dataset creation
    metadata             Download metadata about a dataset
    status               Get the creation status for a dataset
```

Dokumentasi Kaggle Commands selengkapnya [Disini](#)

In [2]:

```
# Mencari dataset yang tersedia di kaggle --> pilih data provider dari UCIML
!kaggle datasets list -s Iris
```

ref	size	lastUpdated	downloadCount	voteCount	usabilityRating	title
uciml/iris	4KB	2016-09-27 07:38:05	223681	2660	0.7941176	Iris Species
arshid/iris-flower-dataset	1010B	2018-03-22 15:18:06	40269	367	0.8235294	Iris Flower Dataset
vikrishnan/iris-dataset	999B	2017-08-03 16:00:44	2888	26	0.7647059	Iris Dataset
therohk/ireland-historical-news	52MB	2021-09-25 10:52:48	2977	157	1.0	Irish Times - Waxy-Wany News
chuckyin/iris-datasets	1KB	2017-03-10 09:35:43	1756	14	0.7352941	Iris datasets
rmatman/iris-dataset-json-version	1KB	2018-04-06 20:21:31	5605	43	0.75	Iris Dataset (JSON Version)
parulpandey/palmer-archipelago-antarctica-penguin-data	67MB	2020-06-29 20:15:18	1851	40	0.8235294	Palmer Archipelago (Antarctica) penguin data
fleanend/birds-songs-numeric-dataset	1KB	2017-11-09 07:34:35	16997	57	0.4117647	Birds' Songs Numeric Dataset
kamrankausar/iris-data	25MB	2019-04-01 09:09:46	703	25	0.9411765	iris_data
jeffheaton/iris-computer-vision	1KB	2017-11-30 10:26:01	1101	13	0.64705884	Iris Computer Vision
styven/iris-dataset	5MB	2020-11-24 21:23:29	302	9	0.875	Iris dataset
arslanali4343/iris-species	1KB	2017-11-04 14:10:12	786	8	0.29411766	Iris Species

2KB	2020-07-02 06:09:09	48	13	0.5625
olgabelitskaya/flower-color-images				Flower Color Images
50MB	2020-10-01 22:48:07	8333	161	0.75
naureenmohammad/mmu-iris-dataset				MMU iris dataset
30MB	2020-07-25 18:38:33	639	19	0.5625
rutujavaidya/iris-dataset				Iris Dataset
1KB	2021-07-25 17:37:14	31	6	0.4117647
shantanuss/iris-flower-dataset				IRIS flower dataset
1KB	2020-01-18 19:43:18	190	3	0.9411765
ashishs0ni/iris-dataset				Iris_dataset
1KB	2018-08-05 14:26:19	584	7	0.64705884
jodx666/iris-is				IRIS is
981B	2018-02-16 09:11:35	339	6	0.3125

In [3]:

```
# Download dan ekstrak dataset, secara default akan berada dalam satu direktori dengan
!kaggle datasets download uciml/iris --unzip
```

Downloading iris.zip to E:\Online Course\Microcredential Data Science\Code\Pertemuan 5

```
0%|          | 0.00/3.60k [00:00<?, ?B/s]
100%|#####| 3.60k/3.60k [00:00<00:00, 3.70MB/s]
```

Atau bisa juga menggunakan link dari kaggle

Latihan (1)

Silahkan Download sebuah dataset menggunakan API Kaggle

In [4]:

```
#Latihan (1)
#Langkah nya seperti contoh diatas
```

PENGGUNAAN LIBRARY PANDAS dan NUMPY

Pada materi ini, peserta sudah mendapatkan pemahaman mengenai data dan dataset. Penggunaan library pada Python memberikan kemudahan dalam proses data understanding. Beberapa library yang digunakan adalah library Pandas dan Numpy.

Latihan (2)

Lakukan import Library Pandas dan Library Numpy

In [5]:

```
#Latihan(2)
#Import Library Pandas

import pandas as pd

#Import Library Numpy

import numpy as np
```

DATAFRAME

DataFrame adalah struktur data 2 dimensi yang berbentuk tabular (mempunyai baris dan kolom). Hampir semua data tidak hanya memiliki 1 kolom tetapi lebih dari 1 kolom, sehingga lebih cocok menggunakan pandas DataFrame untuk mengolahnya.

Penggunaan dataframe pada Python dengan menggunakan syntaks: df.

Latihan (3)

Panggil file (load dataset) dengan format .csv untuk dataset mengenai bunga Iris yang sudah peserta unduh dari Kaggle, dan akan disimpan di dalam dataframe df. Lalu tampilkan 5 baris awal dataset dengan function head()

```
In [6]: #Latihan(3)
#Panggil file (load file bernama Iris.csv) dan simpan dalam dataframe Lalu tampilkan
file_path = "Iris.csv"
df = pd.read_csv(file_path)

df.head()
```

```
Out[6]:   Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm Species
0  1        5.1          3.5         1.4         0.2  Iris-setosa
1  2        4.9          3.0         1.4         0.2  Iris-setosa
2  3        4.7          3.2         1.3         0.2  Iris-setosa
3  4        4.6          3.1         1.5         0.2  Iris-setosa
4  5        5.0          3.6         1.4         0.2  Iris-setosa
```

Telaah Data

Pada telaah data, dapat dilakukan untuk mengetahui:

- tipe data dari setiap kolom
- deskripsi statistik data

Latihan (4)

Tampilkan tipe data dari kolom yang ada pada dataset

```
In [7]: #Latihan(4)
#Tampilkan tipe data dari kolom yang ada pada dataset

df.dtypes
```

```
Out[7]:  Id          int64
SepalLengthCm    float64
SepalWidthCm     float64
PetalLengthCm    float64
PetalWidthCm     float64
```

```
Species          object  
dtype: object
```

Latihan (5)

Apakah tipe Data dari kolom berikut ini: (silakan diisi pada cell di bawah ini)

```
In [9]:  
#Latihan (5)  
#Tipe Data dari kolom yang ada di dataset  
  
#Kolom "Id" memiliki tipe data = Numerikal Integer 64 bits  
#Kolom "SepalLengthCm" memiliki tipe data = Numerikal Float 64 bits  
#Kolom "Species" memiliki tipe data = Kategorikal Objek
```

Latihan (6)

Hitunglah ukuran (jumlah baris dan kolom) dari dataset. Dengan menggunakan method function

```
In [11]:  
#Latihan (6)  
#Hitung ukuran (jumlah baris dan kolom) dari dataset  
  
df.shape  
  
Out[11]: (150, 6)
```

Latihan (7)

Berapakah jumlah baris, dan jumlah kolom pada dataset? (silakan diisi pada cell di bawah ini)

```
In [11]:  
#Latihan (7)  
  
#Jumlah Baris pada dataset adalah = 150 Baris  
  
#Jumlah kolom pada dataset adalah = 6 Kolom
```

Latihan (8)

Tampilkan data yang hanya berisi kolom "Id" dan kolom "Species" dalam bentuk dataframe.

```
In [23]:  
#Latihan (8)  
#Tampilkan data untuk kolom "Id" dan kolom "Species" dalam bentuk dataframe  
  
# df.loc[:, ['Id', 'Species']]  
df[['Id', 'Species']]
```

```
Out[23]:  
      Id   Species  
0     1   Iris-setosa  
1     2   Iris-setosa
```

Id		Species
2	3	Iris-setosa
3	4	Iris-setosa
4	5	Iris-setosa
...
145	146	Iris-virginica
146	147	Iris-virginica
147	148	Iris-virginica
148	149	Iris-virginica
149	150	Iris-virginica

150 rows × 2 columns

Latihan (9)

Tampilkan data dengan dataframe, dan data yang ditampilkan adalah data pada baris dengan indeks 0 (nol) sampai dengan indeks 9 (sembilan)

In [32]:

```
#Latihan (9)
#Tampilkan data dengan dataframe, dan data yang ditampilkan adalah baris dengan indeks 0 sampai dengan indeks 9
# df.iLoc[:10, :]
df[:10]
```

Out[32]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	1.5	0.1	Iris-setosa

Latihan (10)

Tampilkan data hanya kolom "Id" dan kolom "Species" dengan dataframe, dan yang ditampilkan adalah data pada baris dengan indeks 11 (sebelas) sampai dengan indeks 15 (limabelas)

```
In [29]: #Latihan (10)
#Tampilkan data hanya kolom "Id" dan kolom "Species", pada baris dengan indeks 0 (no
df.loc[11:15, ['Id', 'Species']]
```

Out[29]:

	Id	Species
11	12	Iris-setosa
12	13	Iris-setosa
13	14	Iris-setosa
14	15	Iris-setosa
15	16	Iris-setosa

Latihan (11)

Pada DataFrame dapat menampilkan beberapa baris pertama/terakhir dari dataset yang di load. Gunakan Method head() dan tail().

Latihan: Tampilkan data pada 8 (delapan) baris pertama dari dataset, dengan dataframe.

```
In [33]: #Latihan (11)
#Tampilkan data pada 8 (delapan) baris pertama dari dataset, dengan dataframe
df.head(8)
```

Out[33]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa

Latihan (12)

Tampilkan data pada 3 (tiga) baris terakhir dari dataset, dengan dataframe.

```
In [34]: #Latihan (12)
#Tampilkan data pada 3 (tiga) baris terakhir dari dataset, dengan dataframe
df.tail(3)
```

Out[34]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
--	-----------	----------------------	---------------------	----------------------	---------------------	----------------

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

Deskripsi Statistik Data

DataFrame method `describe()` menampilkan statistik dasar setiap kolom data yang bertipe numerik, mencakup banyaknya data (count), rerata aritmetik (mean), simpangan baku (std), nilai terkecil (min), kuartil pertama (25%), kuartil kedua/median (50%), kuartil ketiga (75%), dan nilai terbesar (max).

Latihan (13)

Hitung korelasi dari dataset. Dengan menggunakan method function

In [35]:

```
#Latihan (13)
#Hitung korelasi dataset

df.corr()
```

Out[35]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
Id	1.000000	0.716676	-0.397729	0.882747	0.899759
SepalLengthCm	0.716676	1.000000	-0.109369	0.871754	0.817954
SepalWidthCm	-0.397729	-0.109369	1.000000	-0.420516	-0.356544
PetalLengthCm	0.882747	0.871754	-0.420516	1.000000	0.962757
PetalWidthCm	0.899759	0.817954	-0.356544	0.962757	1.000000

Latihan (14)

Berdasarkan pada perhitungan korelasi di Latihan (11), apakah yang dapat Bapak/Ibu simpulkan sementara? Silakan tuliskan simpulan sementara Bapak/Ibu pada cell di bawah ini.

In [18]:

```
#Latihan (14)
#Simpulan Sementara Hasil Korelasi di Latihan (13)

<.....Berikan Kesimpulan disini.....>
```

Latihan (15)

Hitung korelasi untuk kolom berikut ini: PetalLengthCm, PetalWidthCm

In [36]:

```
#Latihan (15)
```

```
#Hitung korelasi dataset untuk kolom PetalLengthCm, PetalWidthCm
df[['PetalLengthCm', 'PetalWidthCm']].corr()
```

Out[36]:

	PetalLengthCm	PetalWidthCm
PetalLengthCm	1.000000	0.962757
PetalWidthCm	0.962757	1.000000

Latihan (16)

Method "describe" secara otomatis melakukan komputasi statistik untuk semua continuous variable. Secara default "describe" melakukan ignore terhadap variabel bertipe objek.

Komputasi statistik yang dilakukan terdiri dari: count, mean, std, min, max, 25%, 75%, max.

Latihan: Gunakan method describe pada dataset yang sudah di load untuk semua continuous variabel. (Dataset Iris.csv)

In [37]:

```
#Latihan (16)
# Penggunaan Metode describe untuk komputasi statistik

df.describe()
```

Out[37]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.300000	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

Latihan (17)

Gunakan method describe pada dataset yang sudah di load untuk data bertipe objek. (Dataset Iris.csv)

In [40]:

```
#Latihan (17)
#Gunakan method describe pada dataset yang sudah di Load untuk data bertipe objek

df.describe(include=[object])
```

Out[40]:

	Species
count	150

Species	
unique	3
top	Iris-setosa
freq	50

Latihan 18

Gunakan method describe pada dataset yang sudah di load untuk semua type data (continous variabel dan type object).

In [39]:

```
#Latihan (18)
#Gunakan method describe pada dataset yang sudah di Load untuk semua type data

df.describe(include='all')
```

Out[39]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
count	150.000000	150.000000	150.000000	150.000000	150.000000	150
unique	NaN	NaN	NaN	NaN	NaN	3
top	NaN	NaN	NaN	NaN	NaN	Iris-setosa
freq	NaN	NaN	NaN	NaN	NaN	50
mean	75.500000	5.843333	3.054000	3.758667	1.198667	NaN
std	43.445368	0.828066	0.433594	1.764420	0.763161	NaN
min	1.000000	4.300000	2.000000	1.000000	0.100000	NaN
25%	38.250000	5.100000	2.800000	1.600000	0.300000	NaN
50%	75.500000	5.800000	3.000000	4.350000	1.300000	NaN
75%	112.750000	6.400000	3.300000	5.100000	1.800000	NaN
max	150.000000	7.900000	4.400000	6.900000	2.500000	NaN

Latihan (19)

Hitunglah nilai mean dari dataset.

In [41]:

```
#Latihan (19)
#Hitung nilai Mean dari dataset

df.mean()
```

C:\Users\ramat\miniconda3\envs\mcds\lib\site-packages\ipykernel_launcher.py:4: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=True') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.
after removing the cwd from sys.path.

Out[41]:

Id	75.500000
SepalLengthCm	5.843333
SepalWidthCm	3.054000

```
PetalLengthCm      3.758667  
PetalWidthCm      1.198667  
dtype: float64
```

Latihan (20)

Hitung nilai mean dari dataset untuk kolom PetalLengthCm.

```
In [42]:  
#Latihan (20)  
#Hitung nilai Mean untuk kolom PetalLengthCm  
  
df.PetalLengthCm.mean()
```

```
Out[42]: 3.7586666666666666
```

Latihan (21)

Carilah nilai minimal dari dataset untuk kolom SepalWidthCm.

```
In [43]:  
#Latihan (21)  
#Cari nilai minimal untuk kolom SepalWidthCm  
  
df.SepalWidthCm.min()
```

```
Out[43]: 2.0
```

Method Groupby

Method groupby memungkinkan analisis dilakukan secara per kelompok nilai atribut tertentu.

Latihan (22)

Hitunglah nilai mean dari dataset untuk kolom SepalLengthCm per Species dengan menggunakan metode groupby.

```
In [54]:  
#Latihan (22)  
#Hitung nilai mean dari dataset untuk SepalLengthCm per Species dengan metode groupby  
  
df.groupby(['Species']).mean()[['SepalLengthCm']]
```

```
Out[54]:
```

SepalLengthCm

Species	SepalLengthCm
Iris-setosa	5.006
Iris-versicolor	5.936
Iris-virginica	6.588

Method Value Count

value_counts() menghasilkan frekuensi setiap nilai unik di dalam kolom, dan yang tertinggi count-nya adalah merupakan modus pada kolom tersebut.

Latihan (23)

Hitunglah frekuensi pada kolom 'Species' dengan menggunakan metode value_counts().

In [60]:

```
#Latihan (23)
#Hitung frekuensi pada kolom 'Species' dengan menggunakan metode value_counts()

df.Species.value_counts()
```

Out[60]:

```
Iris-setosa      50
Iris-versicolor 50
Iris-virginica  50
Name: Species, dtype: int64
```

Latihan (24)

Tampilkan perhitungan frekuensi pada kolom 'Species' dengan menggunakan metode value_counts() dalam bentuk dataframe.

In [85]:

```
#Latihan (24)
#Perhitungan frekuensi pada kolom 'Species' dengan menggunakan metode value_counts()

df[['Species']].value_counts().reset_index(name='counts')
```

Out[85]:

	Species	counts
0	Iris-setosa	50
1	Iris-versicolor	50
2	Iris-virginica	50

Latihan (25)

Hitunglah frekuensi pada kolom 'PetalLengthCm' dengan menggunakan metode value_counts() dan dalam bentuk dataframe.

In [89]:

```
#Latihan (25)
# Hitung frekuensi pada kolom 'PetalLengthCm' dengan menggunakan metode value_counts()

df[['PetalLengthCm']].value_counts().reset_index(name='counts')
```

Out[89]:

	PetalLengthCm	counts
0	1.5	14
1	1.4	12

PetalLengthCm counts

2	4.5	8
3	5.1	8
4	1.3	7
5	1.6	7
6	5.6	6
7	4.0	5
8	4.9	5
9	4.7	5
10	5.0	4
11	4.8	4
12	4.4	4
13	4.2	4
14	1.7	4
15	5.8	3
16	4.1	3
17	5.7	3
18	5.5	3
19	6.1	3
20	4.6	3
21	3.9	3
22	6.7	2
23	1.2	2
24	5.4	2
25	5.3	2
26	5.2	2
27	5.9	2
28	1.9	2
29	3.3	2
30	3.5	2
31	4.3	2
32	6.0	2
33	6.6	1
34	6.4	1
35	6.3	1
36	1.0	1
37	1.1	1

PetalLengthCm counts

	PetalLengthCm	counts
38	3.8	1
39	3.7	1
40	3.6	1
41	3.0	1
42	6.9	1