Tugas Mandiri Pertemuan 5_Andrean Yonathan_Institut Teknologi Sepuluh Nopember

Latihan (1)

Silahkan Download sebuah dataset menggunakan API Kaggle

In [1]:

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

```
Requirement already satisfied: kaggle in c:\users\user\appdata\local\programs\python\python39\lib\site-packages
(1.5.12)
Requirement already satisfied: python-slugify in c:\users\user\appdata\local\programs\python\python39\lib\site-
packages (from kaggle) (5.0.2)
Requirement already satisfied: python-dateutil in c:\users\user\appdata\local\programs\python\python39\lib\site
-packages (from kaggle) (2.8.1)
Requirement already satisfied: urllib3 in c:\users\user\appdata\local\programs\python\python39\lib\site-package
s (from kaggle) (1.26.6)
Requirement already satisfied: certifi in c:\users\user\appdata\local\programs\python\python39\lib\site-package
s (from kaggle) (2021.5.30)
Requirement already satisfied: six>=1.10 in c:\users\user\appdata\local\programs\python\python39\lib\site-packa
ges (from kaggle) (1.16.0)
Requirement already satisfied: tqdm in c:\users\user\appdata\local\programs\python\python39\lib\site-packages
(from kaggle) (4.62.2)
Requirement already satisfied: requests in c:\users\user\appdata\local\programs\python\python39\lib\site-packag
es (from kaggle) (2.26.0)
Requirement already satisfied: text-unidecode>=1.3 in c:\users\user\appdata\local\programs\python\python39\lib
\site-packages (from python-slugify->kaggle) (1.3)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\user\appdata\local\programs\python\python3
9\lib\site-packages (from requests->kaggle) (2.0.6)
Requirement already satisfied: idna<4,>=2.5 in c:\users\user\appdata\local\programs\python\python39\lib\site-pa
ckages (from requests->kaggle) (3.2)
Requirement already satisfied: colorama in c:\users\user\appdata\local\programs\python\python39\lib\site-packag
es (from tqdm->kaggle) (0.4.4)
```

```
WARNING: You are using pip version 21.2.4; however, version 21.3.1 is available.
You should consider upgrading via the 'C:\Users\USER\AppData\Local\Programs\Python\Python39\python.exe -m pip i
nstall --upgrade pip' command.
```

In [2]:

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

```
ref                                                     title                                          size  l
astUpdated          downloadCount  voteCount  usabilityRating
------------------------------------------------------  --------------------------------------------  -----  -
------------------  -------------  ---------  ---------------
uciml/iris                                              Iris Species                                    4KB  2
016-09-27 07:38:05         225795       2679  0.7941176
arshid/iris-flower-dataset                              Iris Flower Dataset                           1010B  2
018-03-22 15:18:06          40549        370  0.8235294
vikrishnan/iris-dataset                                 Iris Dataset                                   999B  2
017-08-03 16:00:44           2921         26  0.7647059
therohk/ireland-historical-news                         Irish Times - Waxy-Wany News                   52MB  2
021-09-25 10:52:48           2982        157  1.0
chuckyin/iris-datasets                                  Iris datasets                                   1KB  2
017-03-10 09:35:43           1764         14  0.7352941
rtatman/iris-dataset-json-version                       Iris Dataset (JSON Version)                     1KB  2
018-04-06 20:21:31           5629         43  0.75
parulpandey/palmer-archipelago-antarctica-penguin-data  Palmer Archipelago (Antarctica) penguin data   11KB  2
020-06-09 10:14:54          10010        114  0.9705882
conorrot/irish-weather-hourly-data                      Irish Weather (hourly data)                    67MB  2
020-06-29 20:15:18           1861         40  0.8235294
saurabh00007/iriscsv                                    Iris.csv                                        1KB  2
017-11-09 07:34:35          17111         57  0.4117647
jillanisofttech/iris-dataset-uci                        Iris dataset uci                                1KB  2
021-11-06 15:11:47             36         12  1.0
fleanend/birds-songs-numeric-dataset                    Birds' Songs Numeric Dataset                   25MB  2
019-04-01 09:09:46            706         25  0.9411765
kamrankausar/iris-data                                  iris_data                                       1KB  2
017-11-30 10:26:01           1117         13  0.64705884
jeffheaton/iris-computer-vision                         Iris Computer Vision                            5MB  2
020-11-24 21:23:29            304          9  0.875
styven/iris-dataset                                     Iris dataset                                    1KB  2
017-11-04 14:10:12            791          8  0.29411766
arslanali4343/iris-species                              Iris Species                                    2KB  2
020-07-02 06:09:09             57         13  0.5625
olgabelitskaya/flower-color-images                      Flower Color Images                            50MB  2
020-10-01 22:48:07           8362        161  0.75
naureenmohammad/mmu-iris-dataset                        MMU iris dataset                               30MB  2
020-07-25 18:38:33            644         19  0.5625
rutujavaidya/iris-dataset                               Iris Dataset                                    1KB  2
021-07-25 17:37:14             35          6  0.4117647
shantanuss/iris-flower-dataset                          IRIS flower dataset                             1KB  2
020-01-18 19:43:18            195          3  0.9411765
ashishs0ni/iris-dataset                                 Iris_dataset                                    1KB  2
018-08-05 14:26:19            598          7  0.64705884
```

In [3]:

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

```
Downloading iris.zip to C:\Users\USER\OneDrive - Universitas Diponegoro\Andrean\Microcredential - Associate Dat
a Science\Pertemuan 5
```

```
  0%|          | 0.00/3.60k [00:00<?, ?B/s]
100%|##########| 3.60k/3.60k [00:00<00:00, 921kB/s]
```

Latihan (2)

Lakukan import Library Pandas dan Library Numpy

In [4]:

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

import pandas as pd

#Import Library Numpy

import numpy as np
```

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 [5]:

```
#latihan(3)
#Panggil file (load file bernama Iris.csv) dan simpan dalam dataframe Lalu tampilkan 5 baris awal dataset denga

#load file Iris.csv
df = pd.read_csv('Iris.csv')

#tampilkan 5 baris awal
df.head()
```

Out[5]:

	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

Latihan (4)

Tampilkan tipe data dari kolom yang ada pada dataset

In [6]:

```
#latihan(4)
#Tampilkan tipe data dari kolom yang ada pada dataset

dfType = df.dtypes
dfType
```

Out[6]:

```
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 [7]:

```
#Latihan (5)
#Tipe Data dari kolom yang ada di dataset

#Kolom  "Id" memiliki tipe data =  <isikan jawaban di sini>
#Kolom "SepalLengthCm" memiliki tipe data =  <isikan jawaban di sini>
#Kolom "Species" memiliki tipe data =  <isikan jawaban di sini>

dfType_list = dfType.index

for i in range(6):
    df_list = dfType_list[i]
    df_type = dfType[i]
    print(f'Kolom {df_list} memiliki tipe data = {df_type}')
```

```
Kolom Id memiliki tipe data = int64
Kolom SepalLengthCm memiliki tipe data = float64
Kolom SepalWidthCm memiliki tipe data = float64
Kolom PetalLengthCm memiliki tipe data = float64
Kolom PetalWidthCm memiliki tipe data = float64
Kolom Species memiliki tipe data = object
```

Latihan (6)

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

In [8]:

```
#Latihan (6)
#Hitung ukuran (jumlah baris dan kolom) dari dataset

df_size = df.shape
df_size
```

Out[8]:

```
(150, 6)
```

Latihan (7)

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

In [9]:

```
#Latihan (7)

#Jumlah Baris pada dataset adalah =  <isikan jawaban di sini>

#Jumlah kolom pada dataset adalah =  <isikan jawaban di sini>

print('Jumlah baris pada dataset adalah = %d \nJumlah kolom pada dataset adalah = %d' %df_size)
```

```
Jumlah baris pada dataset adalah = 150
Jumlah kolom pada dataset adalah = 6
```

Latihan (8)

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

In [10]:

```
#Latihan (8)
#Tampilkan data untuk kolom "Id" dan kolom "Species" dalam bentuk dataframe

df_id_species = df[['Id', 'Species']]
df_id_species.head()
```

Out[10]:

	Id	Species
0	1	Iris-setosa
1	2	Iris-setosa
2	3	Iris-setosa
3	4	Iris-setosa
4	5	Iris-setosa

Latihan (9)

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

In [11]:

```
#Latihan (9)
#Tampilkan data dengan dataframe, dan data yang ditampilkan adalah baris dengan indeks 0 (nol) sampai dengan in

df.iloc[:10]
```

Out[11]:

	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 [12]:

```
#Latihan (10)

df_id_species.iloc[11:16]
```

Out[12]:

	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 [13]:

```
#Latihan (11)
#Tampilkan data pada 8 (delapan) baris pertama dari dataset, dengan dataframe

df.head(8)
```

Out[13]:

	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 [14]:

```
#Latihan (12)
#Tampilkan data pada 3 (tiga) baris terakhir dari dataset, dengan dataframe

df.tail(3)
```

Out[14]:

	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

Latihan (13)

Hitung korelasi dari dataset. Dengan menggunakan method function

In [15]:

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

df.corr()
```

Out[15]:

	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 (13), apakah yang dapat Bapak/Ibu simpulkan sementara? Silakan tuliskan simpulan sementara Bapak/Ibu pada cell di bawah ini.

Simpulan Sementara Hasil Korelasi di latihan (13)

variabel SepalLengthCm, PetalLengthCm, dan PetalWidthCm ketiganya memiliki korelasi positif kuat antar variabel

Latihan (15)

Hitung korelasi untuk kolom berikut ini: PetalLengthCm, PetalWidthCm

In [16]:

```
#Latihan (15)
#Hitung korelasi dataset untuk kolom PetalLengthCm, PetalWidthCm

df[['PetalLengthCm', 'PetalWidthCm']].corr()
```

Out[16]:

	PetalLengthCm	PetalWidthCm
PetalLengthCm	1.000000	0.962757
PetalWidthCm	0.962757	1.000000

Latihan (16)

Method "describe" secara otomatis melakukan komputasi statistik untuk semua continous variable. Secara default "describe" melakukan ignore terhadap variabel bertype 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 continous variabel. (Dataset Iris.csv)

In [17]:

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

df.describe()
```

Out[17]:

	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 bertype objek. (Dataset Iris.csv)

In [18]:

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

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

Out[18]:

	Species
count	150
unique	3
top	Iris-virginica
freq	50

Latihan 18

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

In [19]:

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

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

Out[19]:

	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-virginica
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 [20]:

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

print('Mean dari dataset adalah:\n')
df.mean()
```

```
Mean dari dataset adalah:
```

Out[20]:

```
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 [21]:

```
#Latihan (20)
#Hitung nilai Mean untuk kolom PetalLengthCm

petalLengthmean = df['PetalLengthCm'].mean()
print(f'Nilai mean untuk kolom PetalLengthCm adalah {petalLengthmean}')
```

```
Nilai mean untuk kolom PetalLengthCm adalah 3.7586666666666693
```

Latihan (21)

Carilah nilai minimal dari dataset untuk kolom SepalWidthCm.

In [22]:

```
#Latihan (21)
#Cari nilai minimal untuk kolom SepalWidthCm

minSepalWidth = df['SepalWidthCm'].min()
print(f'Nilai minimal untuk kolom SepalWidthCm {minSepalWidth}')
```

```
Nilai minimal untuk kolom SepalWidthCm 2.0
```

Latihan (22)

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

In [23]:

```
#Latihan (22)
#Hitung nilai mean dari dataset untuk SepalLengthCm per Species dengan metode groupby

dfSpecies = df.groupby('Species')['SepalLengthCm'].mean()
dfSpecies
```

Out[23]:

```
Species
Iris-setosa        5.006
Iris-versicolor    5.936
Iris-virginica     6.588
Name: SepalLengthCm, dtype: float64
```

Latihan (23)

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

In [24]:

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

df['Species'].value_counts()
```

Out[24]:

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

Latihan (24)

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

In [27]:

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

df['Species'].value_counts().to_frame()
```

Out[27]:

	Species
Iris-virginica	50
Iris-versicolor	50
Iris-setosa	50

Latihan (25)

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

In [31]:

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

df['PetalLengthCm'].value_counts()
```

Out[31]:

```
1.5    14
1.4    12
5.1     8
4.5     8
1.6     7
1.3     7
5.6     6
4.0     5
4.9     5
4.7     5
5.0     4
1.7     4
4.8     4
4.4     4
4.2     4
4.1     3
5.7     3
5.5     3
6.1     3
3.9     3
4.6     3
5.8     3
5.2     2
1.9     2
6.0     2
1.2     2
4.3     2
5.3     2
5.4     2
3.3     2
6.7     2
3.5     2
5.9     2
3.6     1
3.8     1
1.0     1
3.0     1
6.3     1
6.6     1
3.7     1
1.1     1
6.4     1
6.9     1
Name: PetalLengthCm, dtype: int64
```

In [32]:

```
#dalam bentuk dataframe
df['PetalLengthCm'].value_counts().to_frame()
```

Out[32]:

	PetalLengthCm
1.5	14
1.4	12
5.1	8
4.5	8
1.6	7
1.3	7
5.6	6
4.0	5
4.9	5
4.7	5
5.0	4
1.7	4
4.8	4
4.4	4
4.2	4
4.1	3
5.7	3
5.5	3
6.1	3
3.9	3
4.6	3
5.8	3
5.2	2
1.9	2
6.0	2
1.2	2
4.3	2
5.3	2
5.4	2
3.3	2
6.7	2
3.5	2
5.9	2
3.6	1
3.8	1
1.0	1
3.0	1
6.3	1
6.6	1
3.7	1
1.1	1
6.4	1
6.9	1