

Nama : Astuti novita wulandari

Nim : 211001007

Pemograman python C

Task 5 Modul 2

1.

Sign in required We can't sync your favorites, history, password until you sign in to verify your identity

Olahdata Iris.ipynb

File Edit View Insert Runtime Tools Help Last edited on April 1

+ Code + Text

```
[ ] import pandas as pd
```

Load DataSet dan Cek Data

```
[ ] #Load data set iris Unclean
df=pd.read_csv('iris_unclean.csv')

[ ] #Tampilkan DataSet
df
```

	SepallengthCm	SepalwidthCm	PetalLengthCm	PetalWidthCm	Species
0	NaN	3.5	1.4	0.2	Iris-setosa
1	4.9	2000.0	1.4	0.2	Iris-setosa
2	4.7	3.2	-1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica

2.

Sign in required We can't sync your favorites, history, password until you sign in to verify your identity

Olahdata Iris.ipynb

File Edit View Insert Runtime Tools Help Last edited on April 1

+ Code + Text

```
[ ] ...
145 6.7 3.0 5.2 2.3 Iris-virginica
146 6.3 2.5 5.0 1.9 Iris-virginica
147 6.5 3.0 5.2 2.0 Iris-virginica
148 6.2 3.4 5.4 2.3 Iris-virginica
149 5.9 3.0 5.1 1.8 Iris-virginica
150 rows x 5 columns

[ ] #Hitung jumlah nilai null pada Dataset
df.isna().sum()
```

	SepallengthCm	SepalwidthCm	PetalLengthCm	PetalWidthCm	Species
2	2	0	0	0	0
dtype:	int64				

Handle Missing value dengan imputasi Mean

Imputasi adalah pilihan penanganan missing data yang paling bijak dari pada membuang sebagian observasi atau variabel yang mengandung missing value, mengingat bahwa data sangat mahal dan berharga.

Sebelumnya terlihat bahwa ada dua data yang hilang pada SepallengthCm.

```
[ ] #Cari nilai mean dari SepallengCm
df['SepallengCm'].mean()
```

3.

Sign in required We can't sync your favorites, history, password until you sign in to verify your identity

Olahdata Iris.ipynb

File Edit View Insert Runtime Tools Help Last edited on April 1

+ Code + Text

Handle Missing value dengan imputasi Mean

Imputasi adalah pilihan penanganan missing data yang paling bijak dari pada membuang sebagian observasi atau variabel yang mengandung missing value, mengingat bahwa data sangat mahal dan berharga.

Sebelumnya terlihat bahwa ada dua data yang hilang pada Sepal.LengthCm.

```
[ ] #Cari nilai mean dari Sepal.LengthCm
df['Sepal.LengthCm'].mean()
```

5.856756756756757

#Mengganti missing Value dengan mean(), kemudian masukkan ke variabel dfDataBaru=df['Sepal.LengthCm'].fillna(df['Sepal.LengthCm'].mean())

```
[ ] #Cetak Data
dfDataBaru
```

0	5.856757
1	4.900000
2	4.700000
3	4.600000
4	5.000000
...	...
145	6.700000
146	6.300000
147	6.500000
148	6.200000
149	5.900000

Name: Sepal.LengthCm, Length: 150, dtype: float64

4.

Sign in required We can't sync your favorites, history, password until you sign in to verify your identity

Olahdata Iris.ipynb

File Edit View Insert Runtime Tools Help Last edited on April 1

+ Code + Text

#Gabung Data Baru menjadi DataFrame
df2=pd.DataFrame({'Sepal.LengthCm':dfDataBaru,'Sepal.WidthCm':df['Sepal.WidthCm'],'Petal.LengthCm':df['Petal.LengthCm'],'Petal.WidthCm':df['Petal.WidthCm'],'Species':df['Species']})
df2

	Sepal.LengthCm	Sepal.WidthCm	Petal.LengthCm	Petal.WidthCm	Species
0	5.856757	3.5	1.4	1.4	Iris-setosa
1	4.900000	2.0000	1.4	1.4	Iris-setosa
2	4.700000	3.2	-1.3	-1.3	Iris-setosa
3	4.600000	3.1	1.5	1.5	Iris-setosa
4	5.000000	3.6	1.4	1.4	Iris-setosa
...
145	6.700000	3.0	5.2	5.2	Iris-virginica
146	6.300000	2.5	5.0	5.0	Iris-virginica
147	6.500000	3.0	5.2	5.2	Iris-virginica
148	6.200000	3.4	5.4	5.4	Iris-virginica
149	5.900000	3.0	5.1	5.1	Iris-virginica

150 rows x 5 columns

```
[ ] #Cek jumlah baris dan kolom
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column                Non-Null Count  Dtype
...
```

