

Pokok Bahasan II

Data Pre-processing with Python

Kode Pokok Bahasan: TIK.RPL03.001.002.01

Deskripsi Pokok Bahasan:

Membahas bagaimana pengolahan data sebelum diproses menggunakan bahasa Python. Mempersiapkan sebuah data mentah, agar dapat sesuai dengan spesifikasi untuk tahap processing data.

No	Elemen Kompetensi	Indikator Kinerja	Jml Jam	Hal
1	Memahami data preprocessing menggunakan Python.	1.1 Menganalisa data preprocessing dari data yang disediakan	1	
2	Mengimplementasikan preprocessing data pada console Jupyter Notebooks.	1.1 Mampu memproses data dengan Missing Values 1.2 Mampu menangani missing Value	2	

TUGAS PENDAHULUAN

Hal yang harus dilakukan dan acuan yang harus dibaca sebelum praktikum :

1. Menginstal Anaconda Python pada PC masing-masing praktikan.
2. Menginstal Jupyter Notebooks pada Anaconda Python untuk masing-masing praktikan.

DAFTAR PERTANYAAN

1. Apa yang dimaksud dengan Missing Value?
Missing Value adalah hilangnya beberapa data yang telah diperoleh.
2. Adakah perbedaan penanganan Missing Value antara R dan Python, berikan Alasannya?
R: Mean dan Algoritma K-Means merupakan metode yang dapat digunakan untuk imputasi missing data.
Python: Kita dapat gunakan fungsi dropna() untuk menghapus data yang tidak memiliki nilai dan fungsi fillna() untuk mengisi nilai yang kosong.

TEORI SINGKAT

Salah satu Python IDE yang banyak digunakan adalah Jupyter Notebook yang mana mulai digunakan pada tahun 2004. Jupyter Notebook merupakan aplikasi web berdasarkan struktur

server-client yang memungkinkan kita melakukan manipulasi pada notebook yang sedang dijalankan.

Jupyter Notebook adalah singkatan dari 3 bahasa pemrograman yaitu Ju (Julia), Py (Python), dan R. Ia merupakan tools yang populer digunakan untuk melakukan pengolahan data bagi seorang data scientist yang memungkinkan untuk mengintegrasikan antara kode dengan output di dalam satu dokumen secara interaktif.

LAB SETUP

Hal yang harus disiapkan dan dilakukan oleh praktikan untuk menjalankan praktikum modul ini.

1. Menginstall library yang dibutuhkan untuk mengerjakan modul.
2. Menjalankan Jupyter Notebooks.

ELEMEN KOMPETENSI I

Deskripsi:

Memahami data pre-processing menggunakan Python.

Kompetensi Dasar:

1. Menganalisa data pre-processing dari data yang telah disediakan.

Latihan 1.1.1

Penjelasan Singkat :

Pada latihan ini anda akan diminta untuk melakukan analisis terhadap data yang telah disediakan.

Apakah data yang disediakan, termasuk data yang perlu di pre-processing? Jika ya, jelaskan alasannya!

ELEMEN KOMPETENSI 2

Deskripsi:

Mengimplementasikan pre-processing data.

Kompetensi Dasar:

1. Mampu memproses data dengan Missing Values
2. Mampu menangani missing Value

Latihan 1.2.1

Penjelasan Singkat:

Pada latihan ini praktikan diminta untuk memproses missing value dari data yang telah disediakan. Bagaimana cara agar missing value tersebut tidak menjadi kendala dalam tahap processing menggunakan python.

Langkah-Langkah Praktikum:

1. Letakan dataset dalam sebuah folder, dataset adalah data train.csv yang didownload dari <https://www.kaggle.com/c/titanic/data>

2. Buka anaconda prompt dan jalankan perintah berikut ini :

(base) C:\Users\Section>

(base) C:\Users\Section>>cd documents

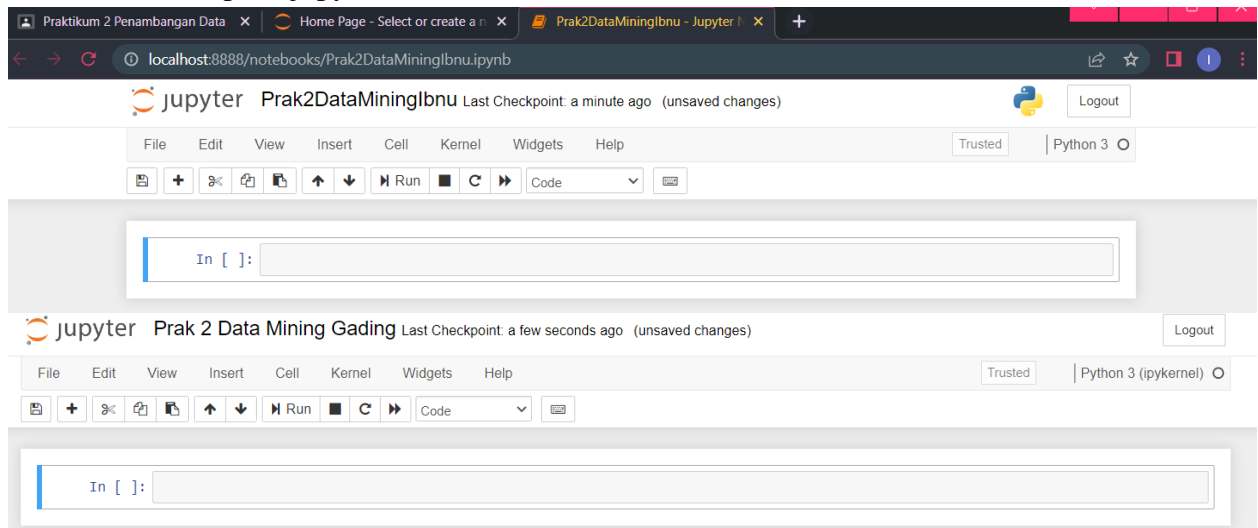
(base) C:\Users\Section\Documents >cd Aslab Data Mining

(base) C:\Users\Section\Documents\Aslab Data Mining >jupyter notebook

```
Jupyter Notebook (Anaconda3) (1)
[I 15:57:33.744 NotebookApp] JupyterLab extension loaded from C:\Users\ibnuu\Anaconda3\lib\site-packages\jupyterlab
[I 15:57:33.744 NotebookApp] JupyterLab application directory is C:\Users\ibnuu\Anaconda3\share\jupyter\lab
[I 15:57:33.749 NotebookApp] Serving notebooks from local directory: C:\Users\ibnuu
[I 15:57:33.749 NotebookApp] The Jupyter Notebook is running at:
[I 15:57:33.749 NotebookApp] http://localhost:8888/?token=eef415c3de0519e9a7f10564cf7eda66926c88409f4854e0
[I 15:57:33.750 NotebookApp] or http://127.0.0.1:8888/?token=eef415c3de0519e9a7f10564cf7eda66926c88409f4854e0
[I 15:57:33.750 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 15:57:33.779 NotebookApp]

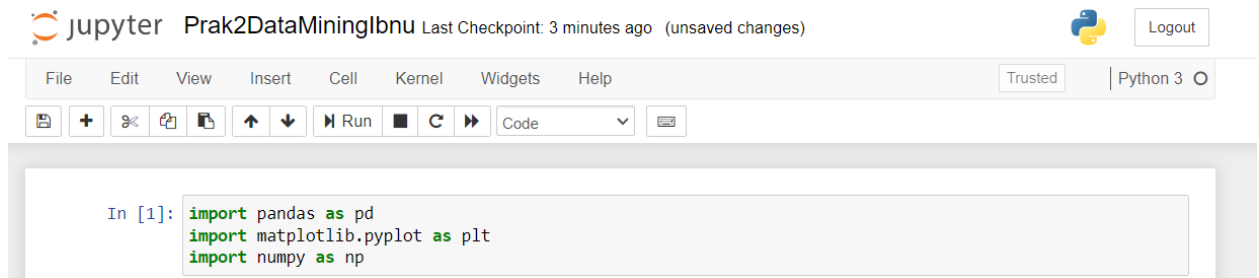
To access the notebook, open this file in a browser:
file:///C:/Users/ibnuu/AppData/Roaming/jupyter/runtime/nbserver-18216-open.html
Or copy and paste one of these URLs:
http://localhost:8888/?token=eef415c3de0519e9a7f10564cf7eda66926c88409f4854e0
or http://127.0.0.1:8888/?token=eef415c3de0519e9a7f10564cf7eda66926c88409f4854e0
[I 15:57:39.615 NotebookApp] Creating new notebook in
[I 15:57:40.537 NotebookApp] Kernel started: a0f46fbc-c7a5-4cc7-afcb-8a635b517211
```

3. Buat file baru pada jupyter notebook



4. Import Library

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```



Jika library belum diinstall maka install terlebih dahulu library tersebut dengan perintah berikut

```
In [2]: #install library
! pip install pandas
! pip install matplotlib
! pip install numpy

Requirement already satisfied: pandas in c:\users\sectio\anaconda3\lib\site-packages (1.0.5)
Requirement already satisfied: numpy>=1.13.3 in c:\users\sectio\anaconda3\lib\site-packages (from pandas) (1.21.4)
Requirement already satisfied: python-dateutil>=2.6.1 in c:\users\sectio\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2017.2 in c:\users\sectio\anaconda3\lib\site-packages (from pandas) (2022.1)
Requirement already satisfied: six>=1.5 in c:\users\sectio\anaconda3\lib\site-packages (from python-dateutil>=2.6.1->pandas) (1.16.0)
```

5. Mengambil data dari file .csv `data_namapraktikan=pd.read_csv('train.csv')` `data_namapraktikan.head(5)`

```
In [5]: data_ibnu=pd.read_csv("D:/File Kuliah Semester 5/Penambangan Data/Prak-2/titanicraw.csv", sep=",")
data_ibnu.head(5)
```

Out[5]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

6. Adakah missing value pada variabe Age? Y

7. Mengecek data secara umum dengan sintaks `data_namapraktikan.info()`

- ada berapa baris dan kolom?
- apa fitur/kolom yang ada pada data?
- apa tipe data dari masing-masing kolom?
- fitur/kolom apa yang mengandung data null?
- kunjungi <https://www.kaggle.com/c/titanic> dan dapatkan informasi lengkap dari setiap kolom yang ada.

```
In [6]: data_ibnu.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   PassengerId     891 non-null    int64
1   Survived        891 non-null    int64
2   Pclass          891 non-null    int64
3   Name            891 non-null    object
4   Sex             891 non-null    object
5   Age            714 non-null    float64
6   SibSp           891 non-null    int64
7   Parch          891 non-null    int64
8   Ticket          891 non-null    object
9   Fare            891 non-null    float64
10  Cabin           204 non-null    object
11  Embarked        889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

8. Mengambil nama kolom dan tipe data dari masing-masing kolom

Selain dengan `data_frame.info()`, pengambilan informasi kolom dapat dilakukan dengan

`#data_frame.columns`

`#data_frame.dtypes`

Dimana `data_frame` adalah nama variabel yang menyimpan data dengan tipe data

`DataFrame`

Cek apa representasi dari data bertipe object pada kolom/fitur Name, Sex, dan Ticket?

```
In [7]: data_ibnu.columns
```

```
Out[7]: Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
              'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
              dtype='object')
```

```
In [8]: data_ibnu.dtypes
```

```
Out[8]: PassengerId    int64
Survived              int64
Pclass                int64
Name                  object
Sex                   object
Age                   float64
SibSp                 int64
Parch                 int64
Ticket                object
Fare                  float64
Cabin                 object
Embarked              object
dtype: object
```

9. Mengidentifikasi nilai mean, median, min value, max value quartile dan standar deviasi pada data bertipe numerik.

```
In [9]: data_ibnu.describe()
```

```
Out[9]:
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

10. Cara lain mengecek missing value selain dengan `data_frame.info()` adalah dengan cara mengambil nama kolom dan menghitung jumlah data pada kolom yang bernilai null dengan `data[nama_kolom].isnull().sum()`

Kolom mana yang mengandung banyak nilai null?

```
In [10]: daftar_kolom = data_ibnu.columns
print("Jumlah Sel pada Kolom Bernilai Null: ")
for kolom in daftar_kolom:
    print(f"{kolom}: {data_ibnu[kolom].isnull().sum()}")
```

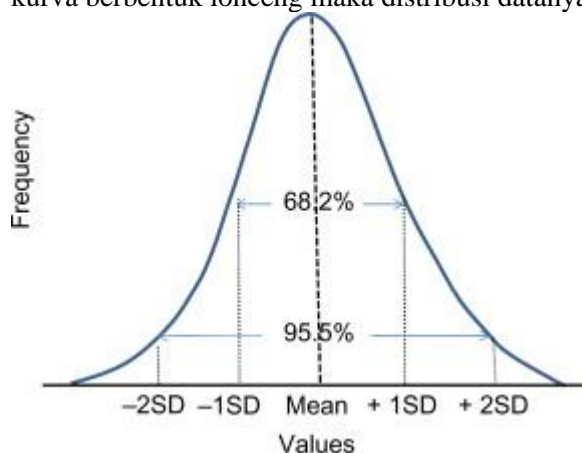
```
Jumlah Sel pada Kolom Bernilai Null:
PassengerId: 0
Survived: 0
Pclass: 0
Name: 0
Sex: 0
Age: 177
SibSp: 0
Parch: 0
Ticket: 0
Fare: 0
Cabin: 687
Embarked: 2
```

11. Data kolom Cabin mengandung banyak sekali nilai null, kita bisa mengabaikan kolom tersebut dengan cara tidak menyertakan data kolom Cabin dengan cara membuangnya. copy data kedalam variabel baru `data_1` dan panggil fungsi drop kolom. Pada fungsi drop parameter `axis` di set = 1. Apa maksudnya?

```
data_ibnu_v1 = data_ibnu.copy(deep = True)
data_ibnu_v1 = data_ibnu_v1.drop(['Cabin'], axis=1)
data_ibnu_v1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  -
0   PassengerId     891 non-null   int64
1   Survived        891 non-null   int64
2   Pclass          891 non-null   int64
3   Name            891 non-null   object
4   Sex             891 non-null   object
5   Age            714 non-null   float64
6   SibSp           891 non-null   int64
7   Parch           891 non-null   int64
8   Ticket          891 non-null   object
9   Fare            891 non-null   float64
10  Embarked        889 non-null   object
dtypes: float64(2), int64(5), object(4)
memory usage: 76.7+ KB
```

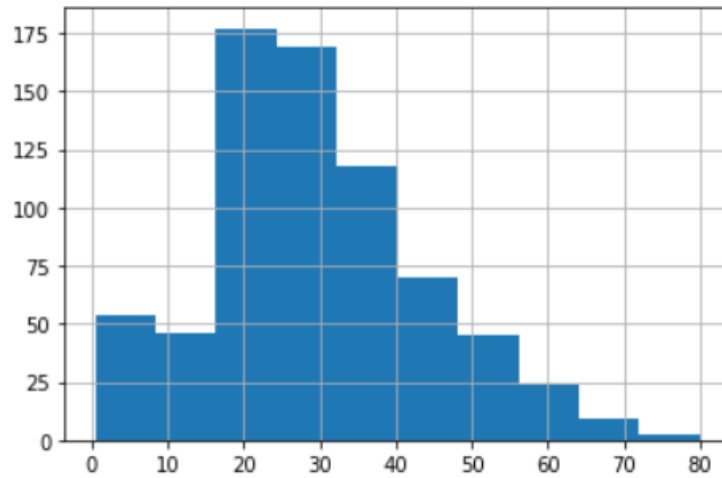
Data yang digunakan untuk membangun model seharusnya berdistribusi normal. Cara mudah untuk mengidentifikasi distribusi data adalah dengan membandingkan kurva distribusi dengan kurva gaussian yang berbentuk lonceng seperti gambar dibawah ini. Jika bentuknya tidak mirip dengan kurva berbentuk lonceng maka distribusi datanya tidak normal. (SD : standard deviation).



Cara memvisualisasikan distribusi data dengan mudah dilakukan dengan fungsi histogram.

```
In [12]: data_ibnu.Age.hist()
```

```
Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x24929236a90>
```



Latihan 1.2.2

Penjelasan Singkat:

Pada latihan ini praktikan diminta untuk mampu menangani missing value dengan data baru

Lakukan penanganan missing value pada variabel age dan salary sesuai langkahlangkah yang telah dipelajari dari data berikut :

1. Cara paling mudah untuk membuang nilai missing value adalah dengan memanggil fungsi dropna()


```
In [14]: data_ibnu_v2 = data_ibnu_v1.copy(deep=True)
data_ibnu_v2.dropna(inplace=True)
data_ibnu_v2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 712 entries, 0 to 890
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  -
0   PassengerId      712 non-null    int64
1   Survived         712 non-null    int64
2   Pclass          712 non-null    int64
3   Name             712 non-null    object
4   Sex              712 non-null    object
5   Age             712 non-null    float64
6   SibSp           712 non-null    int64
7   Parch           712 non-null    int64
8   Ticket           712 non-null    object
9   Fare            712 non-null    float64
10  Embarked         712 non-null    object
dtypes: float64(2), int64(5), object(4)
memory usage: 66.8+ KB
```

Bandingkan dengan data sebelum dilakukan pemanggilan fungsi dropna().

Apa yang terjadi setelah pemanggilan fungsi dropna()?

Data Dihilangkan

2. Melakukan pemberian nilai pada data yang missing value (imputasi)

Ada beberapa cara yang dapat dilakukan untuk melakukan pemberian nilai missing value:

- Memberikan nilai konstan
- Nilai yang diambil secara acak dari kolom yang sama
- Mean, median atau mode dari kolom tersebut
- Diestimasi dari model prediksi lain

```
In [15]: data_ibnu_v3 = data_ibnu.copy(deep = True)
data_ibnu_v3['Age'] = data_ibnu_v3['Age'].fillna(data_ibnu_v3.Age.mean())
data_ibnu_v3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  -
0   PassengerId  891 non-null    int64
1   Survived     891 non-null    int64
2   Pclass       891 non-null    int64
3   Name         891 non-null    object
4   Sex          891 non-null    object
5   Age          891 non-null    object
6   SibSp        891 non-null    int64
7   Parch        891 non-null    int64
8   Ticket       891 non-null    object
9   Fare         891 non-null    float64
10  Cabin        204 non-null    object
11  Embarked     889 non-null    object
dtypes: float64(1), int64(5), object(6)
memory usage: 83.7+ KB
```

Perhatikan pada kolom Age, apakah masih ada data na? Perhatikan apa yang berubah dari data pada kolom Age?

Data Age Ter isi

3. Mengubah representasi data kolom Sex dan Embark.
Lakukan pengisian data missing value pada kolom Embark dengan salah satu nilai selain 'nan' terlebih dahulu.

```
In [18]: data_ibnu_v5 = data_ibnu_v3.copy(deep = True)
data_ibnu_v5.Embarked.unique()
data_ibnu_v5.Embarked.fillna('S', inplace=True)
data_ibnu_v5.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  -
0   PassengerId  891 non-null    int64
1   Survived     891 non-null    int64
2   Pclass       891 non-null    int64
3   Name         891 non-null    object
4   Sex          891 non-null    object
5   Age          891 non-null    object
6   SibSp        891 non-null    int64
7   Parch        891 non-null    int64
8   Ticket       891 non-null    object
9   Fare         891 non-null    float64
10  Cabin        204 non-null    object
11  Embarked     891 non-null    object
dtypes: float64(1), int64(5), object(6)
memory usage: 83.7+ KB
```

Cek nilai unik pada masing-masing kolom, kemudian petakan data pada kolom tersebut menjadi angka.

```
In [19]: print(data_ibnu_v5.Embarked.unique())
print(data_ibnu_v5.Sex.unique())

['S' 'C' 'Q']
['male' 'female']
```

Ubah representasi dari ('S' \Rightarrow 0, 'C' \Rightarrow 1, 'Q' \Rightarrow 2)

Ubah representasi male \Rightarrow 0 dan female \Rightarrow 1

```
In [20]: data_ibnu_v5['Sex'] = data_ibnu_v5['Sex'].map({'male':0, 'female':1})
data_ibnu_v5['Embarked'] = data_ibnu_v5['Embarked'].map({'S':0, 'C':1, 'Q':2})
print(data_ibnu_v5.Embarked.unique())
print(data_ibnu_v5.Sex.unique())

[0 1 2]
[0 1]
```

- Setelah dilakukan beberapa preprocessing, tampilkan kembali beberapa baris data untuk mengecek ulang representasi data hasil proses cleaning

```
In [21]: data_ibnu_v5.head()
Out[21]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	0	22	1	0	A/5 21171	7.2500	NaN	0
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	1	38	1	0	PC 17599	71.2833	C85	1
2	3	1	3	Heikinen, Miss. Laina	1	26	0	0	STON/O2 3101282	7.9250	NaN	0
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	1	35	1	0	113803	53.1000	C123	0
4	5	0	3	Allen, Mr. William Henry	0	35	0	0	373450	8.0500	NaN	0

- Nama dan Tiket bisa diabaikan dengan melakukan drop kolom. Lakukan drop kolom pada kedua kolom tersebut.

```
In [22]: data_ibnu_v6 = data_ibnu_v5.copy(deep = True)
data_ibnu_v6 = data_ibnu_v6.drop(['Name', 'Cabin'], axis=1)
data_ibnu_v6.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  -
0   PassengerId      891 non-null    int64
1   Survived         891 non-null    int64
2   Pclass           891 non-null    int64
3   Sex              891 non-null    int64
4   Age              891 non-null    object
5   SibSp            891 non-null    int64
6   Parch            891 non-null    int64
7   Ticket           891 non-null    object
8   Fare             891 non-null    float64
9   Embarked         891 non-null    int64
dtypes: float64(1), int64(7), object(2)
memory usage: 69.7+ KB
```

Lakukan penanganan missing value pada variabel Age dan Salary sesuai langkahlangkah yang telah dipelajari dari data berikut :

	A	B	C	D	
1	Name	Is_localplayer	Age	Salary	
2	Krmencik	No	29	1500	
3	Andritany	Yes	30	850	
4	Rico	Yes	30		
5	Behrens	No		1750	
6	Hansamu	Yes	27	900	
7	Yusuf	No		1400	
8	Hanif	Yes		880	
9	Kudela	No	35	2000	
10	Rio	Yes		600	
11	Abimanyu	Yes	23	750	
12	Firza	Yes	23	750	

Output

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

In [2]: data_ibnu=pd.read_excel("D:/File Kuliah Semester 5/Penambangan Data/Prak-2/LatihanPrak2.xlsx")
data_ibnu.head(5)
```

```
Out[2]:
```

	Name	Is_localplayer	Age	Salary
0	Krmencik	No	29.0	1500.0
1	Andritany	Yes	30.0	850.0
2	Rico	Yes	30.0	NaN
3	Behrens	No	NaN	1750.0
4	Hansamu	Yes	27.0	900.0

```
In [3]: data_ibnu.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11 entries, 0 to 10
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Name            11 non-null    object
1   Is_localplayer  11 non-null    object
2   Age             7 non-null     float64
3   Salary         10 non-null    float64
dtypes: float64(2), object(2)
memory usage: 480.0+ bytes
```

```
In [4]: data_ibnu.columns

Out[4]: Index(['Name', 'Is_localplayer', 'Age', 'Salary'], dtype='object')
```

```
In [5]: data_ibnu.dtypes

Out[5]: Name            object
Is_localplayer  object
Age             float64
Salary         float64
dtype: object
```

```
In [6]: data_ibnu.describe()
```

```
Out[6]:
```

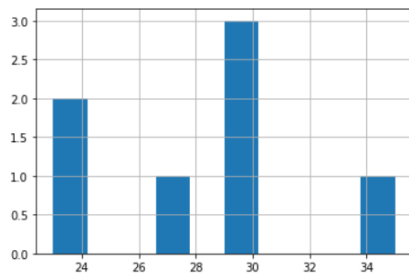
	Age	Salary
count	7.000000	10.000000
mean	28.142857	1138.000000
std	4.259443	484.648785
min	23.000000	600.000000
25%	25.000000	775.000000
50%	29.000000	890.000000
75%	30.000000	1475.000000
max	35.000000	2000.000000

```
In [7]: daftar_kolom = data_ibnu.columns
print("Jumlah Sel pada Kolom Bernilai Null: ")
for kolom in daftar_kolom:
    print(f"{kolom}: {data_ibnu[kolom].isnull().sum()}")
```

```
Jumlah Sel pada Kolom Bernilai Null:
Name: 0
Is_localplayer: 0
Age: 4
Salary: 1
```

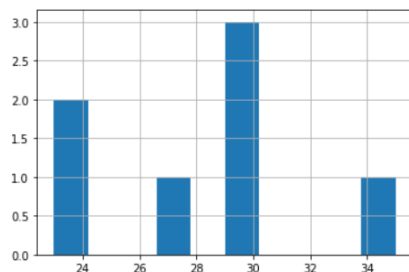
```
In [8]: data_ibnu.Age.hist()
```

```
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x1fcee6fdc40>
```



```
In [8]: data_ibnu.Age.hist()
```

```
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x1fcee6fdc40>
```



```
In [9]: data_ibnu_v3 = data_ibnu.copy(deep = True)
data_ibnu_v3['Age'] = data_ibnu_v3['Age'].fillna(data_ibnu_v3.Age.mean())
data_ibnu_v3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11 entries, 0 to 10
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  ---
0    Name         11 non-null    object
1    Is_localplayer 11 non-null    object
2    Age          11 non-null    object
3    Salary       10 non-null    float64
dtypes: float64(1), object(3)
memory usage: 480.0+ bytes
```

```
In [11]: daftar_kolom = data_ibnu_v3.columns
print("Jumlah Sel pada Kolom Bernilai Null: ")
for kolom in daftar_kolom:
    print(f"{kolom}: {data_ibnu_v3[kolom].isnull().sum()}")
```

```
Jumlah Sel pada Kolom Bernilai Null:
Name: 0
Is_localplayer: 0
Age: 0
Salary: 1
```

```
In [16]: data_ibnu_v5 = data_ibnu_v3.copy(deep = True)
data_ibnu_v5.Salary.unique()
data_ibnu_v5['Salary'] = data_ibnu_v5['Salary'].fillna(data_ibnu_v5.Salary.mean())
data_ibnu_v5.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11 entries, 0 to 10
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Name             11 non-null    object
1   Is_localplayer   11 non-null    object
2   Age              11 non-null    object
3   Salary           11 non-null    object
dtypes: object(4)
memory usage: 480.0+ bytes
```

```
In [17]: daftar_kolom = data_ibnu_v5.columns
print("Jumlah Sel pada Kolom Bernilai Null: ")
for kolom in daftar_kolom:
    print(f"{kolom}: {data_ibnu_v5[kolom].isnull().sum()}")
```

```
Jumlah Sel pada Kolom Bernilai Null:
Name: 0
Is_localplayer: 0
Age: 0
Salary: 0
```

```
In [18]: data_ibnu_v5.head(11)
```

Out[18]:

	Name	Is_localplayer	Age	Salary
0	Krmencik	No	29	1500
1	Andritany	Yes	30	850
2	Rico	Yes	30	<bound method Series.mean of 0 1500.0\n1 ...
3	Behrens	No	<bound method Series.mean of 0 29.0\n1 ...	1750
4	Hansamu	Yes	27	900
5	Yusuf	No	<bound method Series.mean of 0 29.0\n1 ...	1400
6	Hanif	Yes	<bound method Series.mean of 0 29.0\n1 ...	880
7	Kudela	No	35	2000
8	Rio	Yes	<bound method Series.mean of 0 29.0\n1 ...	600
9	Abimanyu	Yes	23	750
10	Firza	Yes	23	750

CEK LIST

Elemen Kompetensi	No Latihan	Penyelesaian	
		Selsai	Tidak Selsai
1	1.1.1	✓	
2	1.2.1	✓	
	1.2.2	✓	

FORM UMPAN BALIK

Elemen Kompetensi	Tingkat Kesulitan	Tingkat Ketertarikan	Waktu Penyelesaian dalam menit
-------------------	-------------------	----------------------	--------------------------------

<p>Memahami data preprocessing menggunakan Python.</p>	<div><input type="checkbox"/> Sangat Mudah</div> <div><input type="checkbox"/> Mudah</div> <div><input checked="" type="checkbox"/> Biasa</div> <div><input type="checkbox"/> Sulit</div> <div><input type="checkbox"/> Sangat Sulit</div>	<div><input checked="" type="checkbox"/> Sangat Tertarik</div> <div><input type="checkbox"/> Cukup Tertarik</div> <div><input type="checkbox"/> Tertarik</div> <div><input type="checkbox"/> Tidak Tertarik</div>	
<p>Mengimplementasikan pre-processing data pada console Jupyter Notebooks</p>	<div><input type="checkbox"/> Sangat Mudah</div> <div><input type="checkbox"/> Mudah</div> <div><input checked="" type="checkbox"/> Biasa</div> <div><input type="checkbox"/> Sulit</div> <div><input type="checkbox"/> Sangat Sulit</div>	<div><input checked="" type="checkbox"/> Sangat Tertarik</div> <div><input type="checkbox"/> Cukup Tertarik</div> <div><input type="checkbox"/> Tertarik</div> <div><input type="checkbox"/> Tidak Tertarik</div>	