



Praktikum Data Mining

Minggu Ke-3



Faris Saifullah
3124640034

Input

```
#no.1
import pandas as pd
dataset = pd.read_csv('titanic.csv')
print(dataset)
```

Analisa:
Menampilkan dataset dari file data titanic.csv

Output

```
PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3
..          ...      ...    ...
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         0       3

Name      Sex  Age  SibSp  \
0      Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2      Heikkinen, Miss. Laina    female  26.0      0
3  Futrelle, Mrs. Jacques Heath (Lily May Peel)    female  35.0      1
4      Allen, Mr. William Henry    male  35.0      0
..          ...      ...    ...
886      Montvila, Rev. Juozas    male  27.0      0
887      Graham, Miss. Margaret Edith    female  19.0      0
888      Johnston, Miss. Catherine Helen "Carrie"    female   NaN      1
889      Behr, Mr. Karl Howell    male  26.0      0
890      Dooley, Mr. Patrick    male  32.0      0

Parch  Ticket  Fare  Cabin  Embarked
0      0   A/5 21171   7.2500   NaN      S
1      0   PC 17599  71.2833   C85      C
2      0  STON/O2. 3101282   7.9250   NaN      S
3      0   113803  53.1000  C123      S
4      0   373450   8.0500   NaN      S
..     ...    ...    ...    ...
886     0   211536  13.0000   NaN      S
887     0   112053  30.0000  B42      S
888     2   W./C. 6607  23.4500   NaN      S
889     0   111369  30.0000  C148      C
890     0   370376   7.7500   NaN      Q

[891 rows x 12 columns]
```

Input

```
#no.2
import pandas as pd
dataset = pd.read_csv('titanic.csv')
rows, cols = dataset.shape
print(f"Jumlah baris: {rows}")
print(f"Jumlah kolom: {cols}")
```

Output

```
Jumlah baris: 891
Jumlah kolom: 12
```

Analisa:
Menampilkan total baris dan kolom
yang ada didalam file titanic.csv



Input

```
#no.3
import pandas as pd
dataset = pd.read_csv('titanic.csv')
data = dataset[['Age', 'Fare']]
print(data)
```

Analisa:
Menampilkan dataset dari kolom age dan fare
Yang ada di file data titanic.csv

Output

	Age	Fare
0	22.0	7.2500
1	38.0	71.2833
2	26.0	7.9250
3	35.0	53.1000
4	35.0	8.0500
..
886	27.0	13.0000
887	19.0	30.0000
888	NaN	23.4500
889	26.0	30.0000
890	32.0	7.7500

[891 rows x 2 columns]

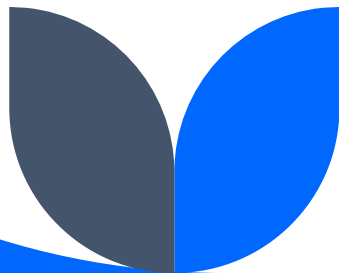
Input

```
#no.4
import pandas as pd
dataset = pd.read_csv('titanic.csv')
class_data = dataset['Survived']
print(class_data)
```

Analisa:
Menampilkan data penumpang yang selamat
Pada file titanic.csv

Output

```
0      0
1      1
2      1
3      1
4      0
..
886    0
887    1
888    0
889    1
890    0
Name: Survived, Length: 891, dtype: int64
```



Input

```
#no.5
import pandas as pd
dataset = pd.read_csv('titanic.csv')
mean_age_per_class = dataset.groupby('Survived')['Age'].mean()
def fill_missing_age(row):
    if pd.isnull(row['Age']):
        return mean_age_per_class[row['Survived']]
    else:
        return row['Age']

dataset['Age'] = dataset.apply(fill_missing_age, axis=1)
print(dataset)
```

Analisa:

Menampilkan rata-rata penumpang yang selamat

Atau tidak berdasarkan usia.

Melakukan pengisian terhadap usia yang kosong dengan

Usia rata-rata

Output

```
PassengerId  Survived  Pclass  \
0             1         0       3
1             2         1       1
2             3         1       3
3             4         1       1
4             5         0       3
..          ...      ...     ...
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         0       3

Name      Sex      Age  \
0      Braund, Mr. Owen Harris    male  22.000000
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.000000
2      Heikkinen, Miss. Laina    female  26.000000
3  Futrelle, Mrs. Jacques Heath (Lily May Peel)    female  35.000000
4      Allen, Mr. William Henry    male  35.000000
..      ...      ...     ...
886      Montvila, Rev. Juozas    male  27.000000
887      Graham, Miss. Margaret Edith    female  19.000000
888  Johnston, Miss. Catherine Helen "Carrie"    female  30.626179
889      Behr, Mr. Karl Howell    male  26.000000
890      Dooley, Mr. Patrick    male  32.000000

SibSp  Parch      Ticket    Fare Cabin Embarked
0      1      0      A/5 21171   7.2500   NaN      S
1      1      0      PC 17599  71.2833   C85      C
2      0      0  STON/O2. 3101282   7.9250   NaN      S
3      1      0      113803  53.1000  C123      S
4      0      0      373450   8.0500   NaN      S
..      ...      ...     ...     ...     ...
886      0      0      211536  13.0000   NaN      S
887      0      0      112053  30.0000  B42      S
888      1      2      W./C. 6607  23.4500   NaN      S
889      0      0      111369  30.0000  C148      C
890      0      0      370376   7.7500   NaN      Q

[891 rows x 12 columns]
```

Input

```
#no.6
import pandas as pd
min_val = data.min()
max_val = data.max()
new_min = 0
new_max = 1
normalized_data = (data - min_val) * (new_max - new_min) / (max_val - min_val) + new_min
print(normalized_data.head())
```

Output

	Age	Fare
0	0.271174	0.014151
1	0.472229	0.139136
2	0.321438	0.015469
3	0.434531	0.103644
4	0.434531	0.015713

Analisa:

Mengisi usia yang hilang menggunakan rata-rata pada kolom age dan fare
Yang dinormalisasikan ke rentang 0-1 nilai setiap fitur mencerminkan posisi relatifnya antara nilai minimum dan maksimum dalam dataset



Input

```
#no.7
import pandas as pd
mean_val = data.mean()
std_val = data.std()
z_score_normalized_data = (data - mean_val) / std_val
print(z_score_normalized_data.head())
```

Output

	Age	Fare
0	-0.592148	-0.502163
1	0.638430	0.786404
2	-0.284503	-0.488580
3	0.407697	0.420494
4	0.407697	-0.486064

Analisa:

Mengisi usia yang hilang menggunakan rata-rata pada kolom age dan fare

ormalisasi ini mengubah nilai fitur menjadi nilai dengan distribusi mean 0 dan deviasi standar 1

Untuk mengukur seberapa jauh nilai masing-masing fitur dari rata-ratanya.



Input

```
#no.8
import pandas as pd
import numpy as np
sigmoidal_normalized_data = 1 / (1 + np.exp(-data))
print(sigmoidal_normalized_data.head())
```

Output

	Age	Fare
0	1.0	0.999290
1	1.0	1.000000
2	1.0	0.999639
3	1.0	1.000000
4	1.0	0.999681

Analisa:

Mengisi usia yang hilang menggunakan rata-rata pada kolom age dan fare

Normalisasi Sigmoidal ini membuat nilai berada dalam rentang (0, 1),

tetapi hasil yang sangat mendekati 1 menunjukkan bahwa data asli mungkin memiliki rentang nilai yang sangat tinggi.

