

# Laporan praktikum Data Mining

Agung Dwi Nugroho  
3122600006

# dataset titanic.csv, dan tampilkan



```
import pandas as pd
```

```
dataset = pd.read_csv('titanic.csv')  
dataset
```

Analisis :  
Menampilkan data csv  
menjadi table



PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C

# rows, cols jumlah baris dan kolom pada dataset, dan tampilkan

```
[5] rows,cols = dataset.shape  
  
print("jumlah baris : ",rows)  
print("jumlah kolom : ",cols)
```

```
⇒ jumlah baris : 891  
   jumlah kolom : 12
```

Analisis :  
Menampilkan banyak row dan  
kolom data csv

# data ambil dataset kolom fitur (Age, Fare), dan tampilkan

```
data = dataset.loc[:,['Age', 'Fare']]  
data
```

Analisis :  
Menampilkan data dengan hanya memilih Age dan Fare saja



	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

# class ambil dataset kolom kelas (Survived)

```
[7] Class = dataset['Survived']  
Class
```

Analisis :  
Menampilkan data survived

Survived	
0	0
1	1
2	1
3	1
4	0

# Lakukan pengisian missing value pada fitur Age dengan nilai mean dari masing-masing class

```
data['Age'] = data['Age'].fillna(data.groupby(Class)['Age'].transform('mean'))  
data
```

Analisis :

Mengisi data value yang kosong dengan rata-rata value dari nilai yang lain

	Age	Fare
0	22.000000	7.2500
1	38.000000	71.2833
2	26.000000	7.9250
3	35.000000	53.1000
4	35.000000	8.0500

# Min-Max (0-1)

```
[9] data_min_max_manual = (data - data.min()) / (data.max() - data.min())  
  
data_normalisasi_min_max_manual = pd.DataFrame(data_min_max_manual, columns=['Age', 'Fare'])  
data_normalisasi_min_max_manual
```

Analisis :  
Membuat normalisasi data  
dengan metode min max dengan  
rumus  $(data - data_{kecil}) / (data_{besar} - data_{kecil})$

	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

# Z-Score

```
data_z_score_manual = (data - data.mean()) / data.std()
```

```
data_normalisais_z_score_manual = pd.DataFrame(data_z_score_manual, columns=['Age', 'Fare'])  
data_normalisais_z_score_manual
```

Analisis :  
Membuat normalisasi data  
dengan metode z score dengan  
rumus  $(data - data \text{ rata rata}) / std$

	Age	Fare
0	-0.595670	-0.502163
1	0.634089	0.786404
2	-0.288230	-0.488580
3	0.403509	0.420494
4	0.403509	-0.486064



# Sigmoidal

```
import numpy as np

data_sigmoidal = 1 / (1 + np.exp(-data))

data_normalisasi_sigmoidal = pd.DataFrame(data_sigmoidal, columns=['Age', 'Fare'])
data_normalisasi_sigmoidal
```

Analisis :  
Membuat normalisasi data  
dengan metode sigmoidal dengan

	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

Link google collab :

<https://colab.research.google.com/drive/1awB6g87Q2I8kQZ-H3Ak5DKmCJfuP9zs6?usp=sharing>

