

BAB 2

TRANSFORMASI DATA

Pada beberapa kasus, variabel cenderung memiliki nilai rentang yang sangat besar. Nilai rentang yang sangat besar ini akan mempengaruhi hasil pengolahan data. Untuk mengatasi masalah ini, data harus ditransformasi terlebih dahulu. Normalisasi sering digunakan. Nilai rentang yang besar akan menjadi rentang yang tidak terlalu besar. Normalisasi terdapat beberapa metode. Metode normalisasi data yang paling sering digunakan yaitu:

1. Min-Max Normalization

Min-max merupakan metode normalisasi data dengan menskalakan data diantara 0 dan 1. Metode ini menggunakan rumus:

$$X'_i = \frac{X_i - \text{Min}(X)}{\text{Max}(X) - \text{Min}(X)}$$

Misalnya data $X = [7 \ 10 \ 15 \ 20 \ 25]$

- $X'_1 = \frac{7-7}{25-7} = 0$
- $X'_2 = \frac{10-7}{25-7} = 0,1667$
- $X'_3 = \frac{15-7}{25-7} = 0,4444$
- $X'_4 = \frac{20-7}{25-7} = 0,7222$
- $X'_5 = \frac{25-7}{25-7} = 1$

Perbandingan data sebelum dan sesudah di normalisasi ditunjukkan pada Tabel 1. X adalah data sebelum dinormalisasi dan X' adalah data setelah dinormalisasi. Rentang data X berada diantara 7 dan 25 sedangkan setelah dinormalisasi rentang data menjadi diantara 0 dan 1.

Tabel 1. Perbandingan data sebelum dan sesudah dinormalisasi

X	X'
7	0
10	0,1667
15	0,4444
20	0,7222
25	1

Untuk mempermudah perhitungan, perhitungan dapat dilakukan pada MATLAB. Berikut contoh source codenya

```
v = [7 10 15 20 25];
for i = 1:length(v)
    nor(i) = (v(i)-min(v))/(max(v)-min(v));
end
disp(nor)
```

Hasil :

Command Window					
	7	10	15	20	25
	0	0.1667	0.4444	0.7222	1.0000

2. Z-Score Standardization

Metode Z-Score Standardization merupakan metode yang menskalakan selisih antara nilai pada data dan rata-ratanya dengan nilai standar deviasinya. Metode ini menggunakan rumus:

$$X'_i = \frac{X_i - \text{Mean}(X)}{\sigma}$$

Dengan

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \text{Mean}(X))^2}{n}}$$

Keterangan: n = Banyak data

σ = standar deviasi

Misalnya data X = [7 10 15 20 25]

- Mencari nilai rata-rata

$$\begin{aligned} \text{Mean}(X) &= \frac{7 + 10 + 15 + 20 + 25}{5} \\ &= \frac{77}{5} \\ &= 15,4 \end{aligned}$$

- Mencari standar Deviasi

$$\sigma = \sqrt{\frac{(7 - 15,4)^2 + (10 - 15,4)^2 + (15 - 15,4)^2 + (20 - 15,4)^2 + (25 - 15,4)^2}{5}}$$

$$\sigma = \sqrt{\frac{70,56 + 29,16 + 0,16 + 21,16 + 92,16}{5}}$$

$$\sigma = \sqrt{\frac{213,2}{5}}$$

$$\sigma = 6,5299$$

$$\circ X'_1 = \frac{7-15,4}{6,5299} = -1,2864$$

$$\circ X'_i = \frac{10-15,4}{6,5299} = -0,827$$

$$\circ X'_i = \frac{15-15,4}{6,5299} = -0,0613$$

$$\circ X'_i = \frac{20-15,4}{6,5299} = 0,7044$$

$$\circ X'_i = \frac{25-15,4}{6,5299} = 1,4720$$

Perbandingan data sebelum dan sesudah di normalisasi ditunjukkan pada Tabel 2. X adalah data sebelum dinormalisasi dan X' adalah data setelah dinormalisasi. Rentang data X berada diantara 7 dan 25 sedangkan setelah dinormalisasi rentang data menjadi diantara 1,4720 dan -1,2864.

Tabel 2. Perbandingan Data Sebelum dan Sesudah dinormalisasi menggunakan Z-Score

X	X'
7	-1,2864
10	-0,827
15	-0,0613
20	0,7044
25	1,4720

Untuk mempermudah perhitungan dapat dilakukan pada MATLAB dengan source code sebagai berikut:

```
A = [7 10 15 20 25];
rata2 = mean(A);
c = 0;
for i = 1:length(A)
    d(i) = (A(i)-rata2)^2;
    c = c+d(i)
    sd = sqrt(c/length(A));
end
fprintf('Standar Deviasi = %.4f\n Data Baru = ',sd);
for i = 1:length(A)
    X(i) = (A(i)-rata2)/sd;
end
disp(X)
```

Hasil :

```
Command Window
Standar Deviasi = 6.5299
Data Baru =      -1.2864      -0.8270      -0.0613      0.7044      1.4702
```

Normalisasi Data pada MATLAB

MATLAB menyediakan *function* yang dapat digunakan untuk memudahkan pengguna dalam menormalisasi data dengan beberapa metode. *Function* ini adalah `normalize`.

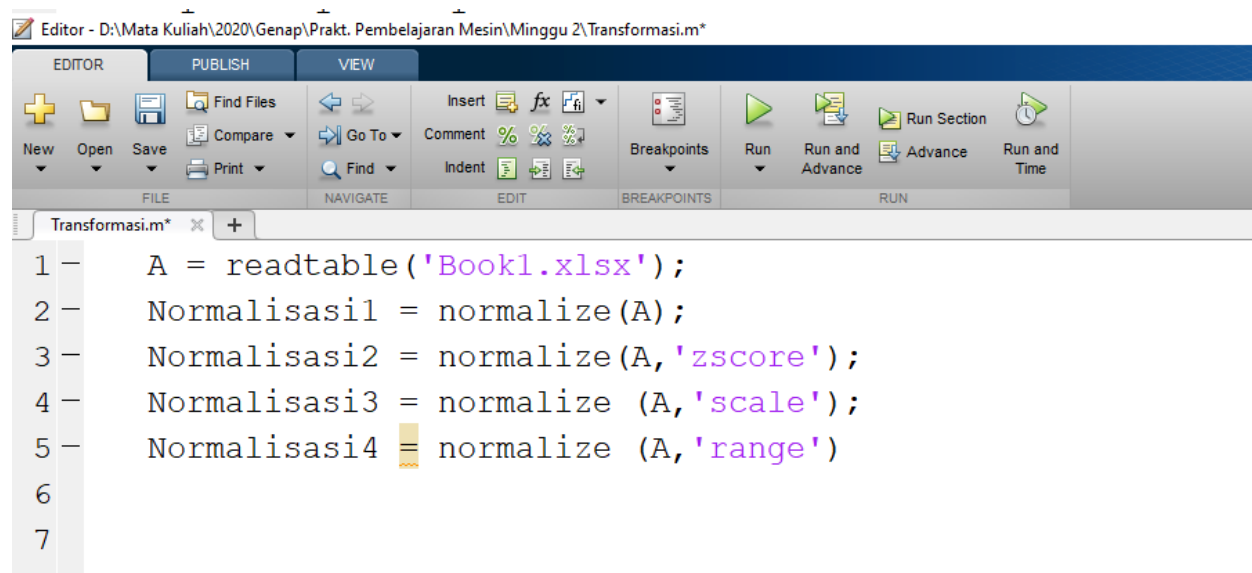
```
Normalisasi = normalize(A)
Normalisasi = normalize(A,method)
```

Pada syntax diatas A merupakan data yang akan dinormalisasi, secara default metode yang digunakan untuk menormalisasi data pada MATLAB adalah metode Z-Score Standardization. method yang digunakan dalam function `normalize` akan dijelaskan pada Tabel 3.

Tabel 3. Metode normalisasi pada MATLAB

Metode	Penjelasan
'zscore'	Digunakan untuk data dengan rata-rata = 0 dan standar deviasi = 1
'scale'	Menskalakan data berdasarkan standar deviasi
'range'	Menskalakan data pada jarak anatar 1 dan 0.

Sebagai contoh penggunaan function normalisasi dilakukan dengan source code sebagai berikut:



```
Editor - D:\Mata Kuliah\2020\Genap\Prakt. Pembelajaran Mesin\Minggu 2\Transformasi.m*
EDITOR  PUBLISH  VIEW
+ New  Open  Save  Find Files  Compare  Print  Go To  Find  Insert  Comment  Indent  Breakpoints  Run  Run and Advance  Run Section  Advance  Run and Time
FILE  NAVIGATE  EDIT  BREAKPOINTS  RUN

Transformasi.m*  +
1 - A = readtable('Book1.xlsx');
2 - Normalisasi1 = normalize(A);
3 - Normalisasi2 = normalize(A,'zscore');
4 - Normalisasi3 = normalize (A,'scale');
5 - Normalisasi4 = normalize (A,'range')
6
7
8
```

Keterangan dan hasil source code diatas sebagai berikut :

1. Membaca data

Hasil :

```
>> Transformasi
```

```
A =
```

```
3×3 table
```

x1	x2	x3
—	—	—
1	2	3
4	5	6
7	8	9

2. Normlasisasi

Hasil

```
>> Transformasi
```

```
Normalisasi1 =
```

```
3×3 table
```

x1	x2	x3
—	—	—
-1	-1	-1
0	0	0
1	1	1

3. Normalisas dengan method 'zscore'

Hasil :

```
>> Transformasi
```

```
Normalisasi2 =
```

```
3×3 table
```

x1	x2	x3
—	—	—
-1	-1	-1
0	0	0
1	1	1

4. Normalisas dengan method 'scale'

```
>> Transformasi
```

```
Normalisasi3 =
```

```
3×3 table
```

x1	x2	x3
0.33333	0.66667	1
1.3333	1.6667	2
2.3333	2.6667	3

5. Normalisas dengan method 'range'

```
>> Transformasi
```

```
Normalisasi4 =
```

```
3×3 table
```

x1	x2	x3
0	0	0
0.5	0.5	0.5
1	1	1

Normalisasi Data pada Phyton

Ada 3 macam normalisasi yang menggunakan Phyton, yaitu :

1. Simple Feature Scaling

Simple Feature Scaling merupakan metode normalisasi data dengan menskalakan data diantara 0 dan 1. Metode ini menggunakan rumus:

$$X'_i = \frac{X_i}{\text{Max}(X)}$$

Misalnya data X = [7 10 15 20 25]

$$\circ X'_1 = \frac{7}{25} = 0.28$$

$$\circ X'_2 = \frac{10}{25} = 0,4$$

- $X'_3 = \frac{15}{25} = 0,6$
- $X'_4 = \frac{20}{25} = 0,8$
- $X'_5 = \frac{25}{25} = 1$

Normalisasi setiap atribut dapat menerapkan kode berikut. Nilai atribut yang akan dinormalisasi atribut 'umur' dan 'gaji'.

```
df["Umur"] = df["Umur"] / df["Umur"].max()
df["Gaji"] = df["Gaji"] / df["Gaji"].max()
```

2. Min-Max

Penjelasan sama seperti bagian diatas.

```
df["Umur"] = (df["Umur"] - df["Umur"].min()) / (df["Umur"].max() - df["Umur"].min())
df["Gaji"] = (df["Gaji"] - df["Gaji"].min()) / (df["Gaji"].max() - df["Gaji"].min())
```

3. Z score

Penjelasan sama seperti bagian diatas.

```
df["Umur"] = (df["Umur"] - df["Umur"].mean()) / df["Umur"].std()
df["Gaji"] = (df["Gaji"] - df["Gaji"].mean()) / df["Gaji"].std()
```

Normalisasi :

1. Langkah pertama adalah import library terlebih dahulu, library yang digunakan adalah dan Pandas.

```
import pandas as pd
from sklearn import preprocessing
```

2. Baca data

Misalkan data yang digunakan dengan nama "shopping data.csv". Untuk nilai dan atribut data ini buatlah sendiri (bebas)

```
data = pd.read_csv("shopping_data.csv")
df = pd.DataFrame(data)
```

3. Normalisasi

```
min_max_scaler = preprocessing.MinMaxScaler()
np_scaled = min_max_scaler.fit_transform(df)
df_normalized = pd.DataFrame(np_scaled)
```

LATIHAN :

1. Cobalah semua source code diatas

TUGAS :

1. Carilah data bebas
2. Normalisasikan data tersebut dengan menggunakan matlab dan Phyton.
3. Buatlah laporan yang terdiri dari print screen hasil dan penjelasannya
4. Berilah nama “ NIM_Normalisasi.pdf
5. Tugas terdiri dari Laporan, file matlab, dan file phyton, dan data aslinya dan dikumpulkan dengan nama “Tugas Tranformasi_NIM.Zip
6. Tugas dikumpulkan paling lambat hari Kamis / 23 Februari 2023 pukul 23.59 Wib

***** Selamat Mengerjakan *****