

# Modul Praktikum Data Mining



## **Tim Penyusun:**

Dr. Rakhmat Arianto, S.ST., M.Kom

Ir. Rudy Ariyanto, ST., M.Cs

Prof. Dr. Eng. Rosa Andrie Asmara, ST., MT

**Jurusan Teknologi Informasi**

**Sistem Informasi Bisnis**

**Politeknik Negeri Malang**

**Februari 2025**

## DAFTAR ISI

DAFTAR ISI .....	2
JOB SHEET 2 Pengumpulan Data.....	3
Pendahuluan .....	3
Tujuan Praktikum .....	3
Visual Objectives .....	<b>Error! Bookmark not defined.</b>
Peralatan yang dibutuhkan .....	3
Praktikum .....	3
Pengumpulan Data Secara Manual .....	<b>Error! Bookmark not defined.</b>
Pengumpulan Data Menggunakan API .....	<b>Error! Bookmark not defined.</b>
Latihan.....	<b>Error! Bookmark not defined.</b>
Tugas Praktikum .....	<b>Error! Bookmark not defined.</b>

## JOBSHEET 5

### Menentukan Objek Data

#### Pendahuluan

Modul ini menjelaskan proses Menentukan Objek Data dengan menggunakan metode Correlation dan implementasi dari metode Sampling Slovin yang dilengkapi dengan langkah-langkah detail serta gambar untuk memudahkan pemahaman.

#### Tujuan Praktikum

Setelah menyelesaikan praktikum ini, mahasiswa mampu:

- Memahami tentang Metode Correlation.
- Memahami tentang Metode Sampling Slovin.

#### Peralatan yang dibutuhkan

Beberapa peralatan yang dibutuhkan dalam menyelesaikan praktikum ini adalah:

- Aplikasi Microsoft Excel
- Google Colab
- Google Drive
- Koneksi Internet
- Browser Web

#### Praktikum

##### Implementasi Metode Correlation Menggunakan Ms. Excel

Lakukan praktikum sesuai tahapan berikut:

- a. Buka aplikasi web browser
- b. Unduh file contoh praktikum pada [Hitung Korelasi.xlsx](#)
- c. Pada file Excel tersebut telah terdapat data sebagai berikut:

ID	Luas Tanah (m <sup>2</sup> )	Jumlah Kamar	Jarak ke Pusat Kota (km)	Harga (Juta Rupiah)
1	100	3	5	500
2	150	4	10	450
3	80	2	2	400
4	120	3	8	350
5	200	5	15	700
6	90	2	7	300
7	130	3	12	380
8	110	3	6	480
9	140	4	9	420
10	95	2	4	320

- d. Berdasarkan data tersebut, maka jumlah keseluruhan data adalah 10 data
- e. Hitung korelasi dari setiap kolom terhadap Kolom **Harga** dengan menggunakan rumus:

$$r_{xy} = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{(n \sum x_i^2 - (\sum x_i)^2)(n \sum y_i^2 - (\sum y_i)^2)}}$$

Dimana:

- r = nilai korelasi
  - x = variabel x
  - y = variabel y
- f. Hitung Korelasi antara **Harga** dengan **Luas Tanah** maka:
- X = Luas Tanah dan Y = Harga
- g. Buat tabel dataset sesuai dengan kebutuhan dari rumus Correlation

ID	x	y	xy	x <sup>2</sup>	y <sup>2</sup>
1	100	500	50000	10000	250000
2	150	450	67500	22500	202500
3	80	400	32000	6400	160000
4	120	350	42000	14400	122500
5	200	700	140000	40000	490000
6	90	300	27000	8100	90000
7	130	380	49400	16900	144400
8	110	480	52800	12100	230400
9	140	420	58800	19600	176400
10	95	320	30400	9025	102400
Total	1215	4300	549900	159025	1968600

- h. Hitung nilai r sesuai dengan rumus, maka akan didapatkan nilai korelasinya  $r = 0,743321541$
- i. Intepretasi dari nilai r tersebut adalah
- Harga dengan Luas Tanah memiliki korelasi positif yang sangat erat
- Harga akan naik jika luas tanah semakin besar

**LATIHAN**

1. Salin data dari Ms. Excel ke Google SpreadSheet
2. Lakukan perhitungan untuk korelasi dari:
  - a. Harga dengan Jumlah Kamar
  - b. Harga dengan Jarak Ke Pusat Kota
3. Intepretasikan setiap hasil korelasi yang didapatkan!
4. Kumpulkan hasil pengerjaan dengan mengirimkan linknya

## Implementasi Metode Correlation menggunakan Python

- Buka Web Browser
- Masuk pada Google Colab
- Ketikkan perintah berikut ini

```
import pandas as pd

# Data dalam bentuk dictionary
data = {
    'Luas Tanah (m²)': [100, 150, 80, 120, 200, 90, 130, 110, 140, 95],
    'Jumlah Kamar': [3, 4, 2, 3, 5, 2, 3, 3, 4, 2],
    'Jarak ke Pusat Kota (km)': [5, 10, 2, 8, 15, 7, 12, 6, 9, 4],
    'Harga (Juta Rupiah)': [500, 450, 400, 350, 700, 300, 380, 480, 420, 320]
}

# Membuat DataFrame
df = pd.DataFrame(data)

# Menghitung matriks korelasi
correlation_matrix = df.corr()

# Menampilkan korelasi antara fitur dan target
print(correlation_matrix['Harga (Juta Rupiah)'])
```

- Maka hasil yang didapatkan

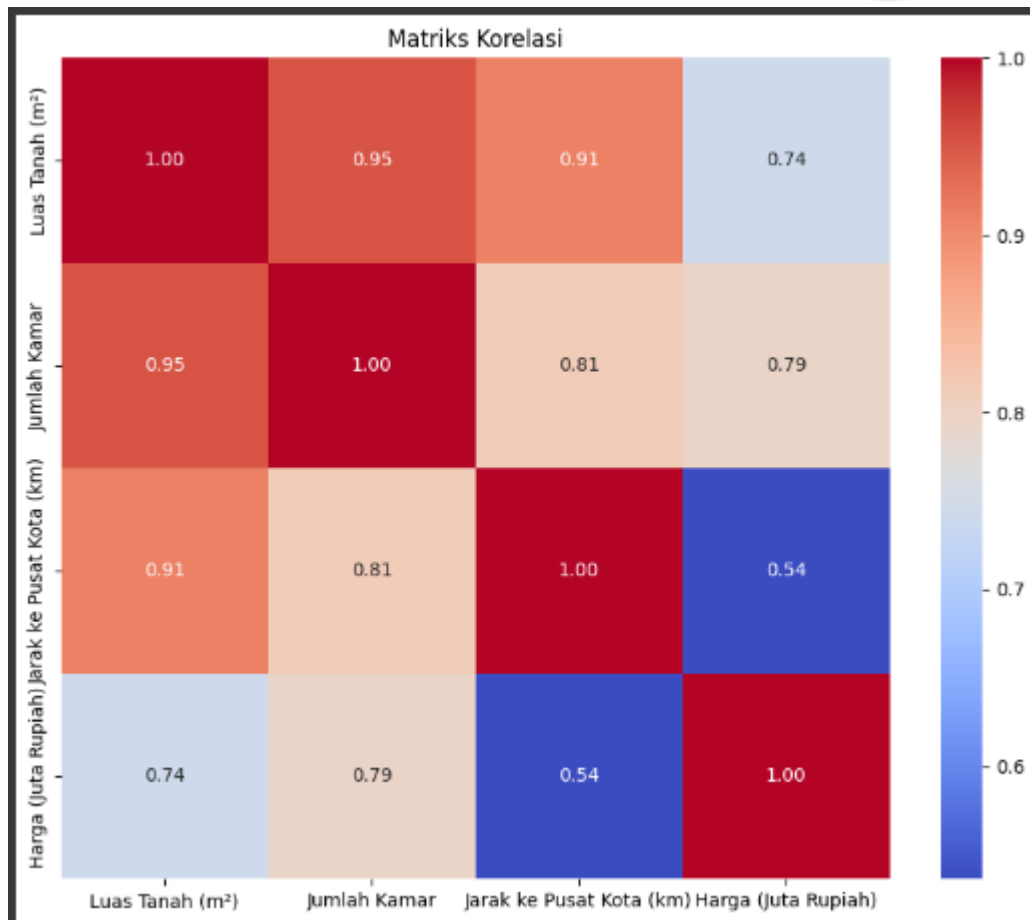
```
Luas Tanah (m²)      0.743322
Jumlah Kamar         0.794792
Jarak ke Pusat Kota (km) 0.536362
Harga (Juta Rupiah)   1.000000
Name: Harga (Juta Rupiah), dtype: float64
```

- Pada kode berikutnya, ketikkan kode berikut:

```
import matplotlib.pyplot as plt
import seaborn as sns

# Visualisasi matriks korelasi menggunakan heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Matriks Korelasi')
plt.show()
```

- Maka hasil yang didapatkan adalah:



### LATIHAN

1. Tambahkan satu kolom pada data menggunakan Python dengan nama “Gangguan Listrik”
2. Buatlah data “Gangguan Listrik” dengan nilai antara 1 sampai dengan 5 yang menunjukkan frekuensi terjadi gangguan listrik
3. Isikan kolom Gangguan Listrik dengan sifat **Korelasi Negatif**
4. Ulangi perhitungan Korelasi dengan kolom yang baru!

## Implementasi Metode Sampling Slovin Menggunakan Python

- Buka Google Colabs
- Ketikkan perintah berikut:

```
import pandas as pd
# Parameter Slovin
N = 100 # Ukuran populasi (100)
e = 0.05 # Tingkat kesalahan 5%

# Hitung ukuran sampel
n = slovin_sample_size(N, e)
print(f"Ukuran sampel yang dibutuhkan: {n}")

# Buat DataFrame dummy dengan 100 data
data = {'value': range(1, 101)}
df = pd.DataFrame(data)

# Ambil sampel acak dari DataFrame
sample_df = df.sample(n=n, random_state=42) # random_state untuk hasil yang konsisten

# Tampilkan sampel
print("Sampel yang diambil:")
sample_df
```

- Maka akan menghasilkan sebagai berikut:

Ukuran sampel yang dibutuhkan: 80  
Sampel yang diambil:

value
83
84
53
54
70
71
45
46
44
45
...
57
58
75
76
32
33
94
95
59
60

80 rows x 1 columns



**LATIHAN**

1. Buatlah data dengan menggunakan 5 kolom dan 1000 baris
2. Lakukan metode Sampling Slovin dengan menggunakan Python
3. Jelaskan masing-masing baris perintah Python yang tersusun!