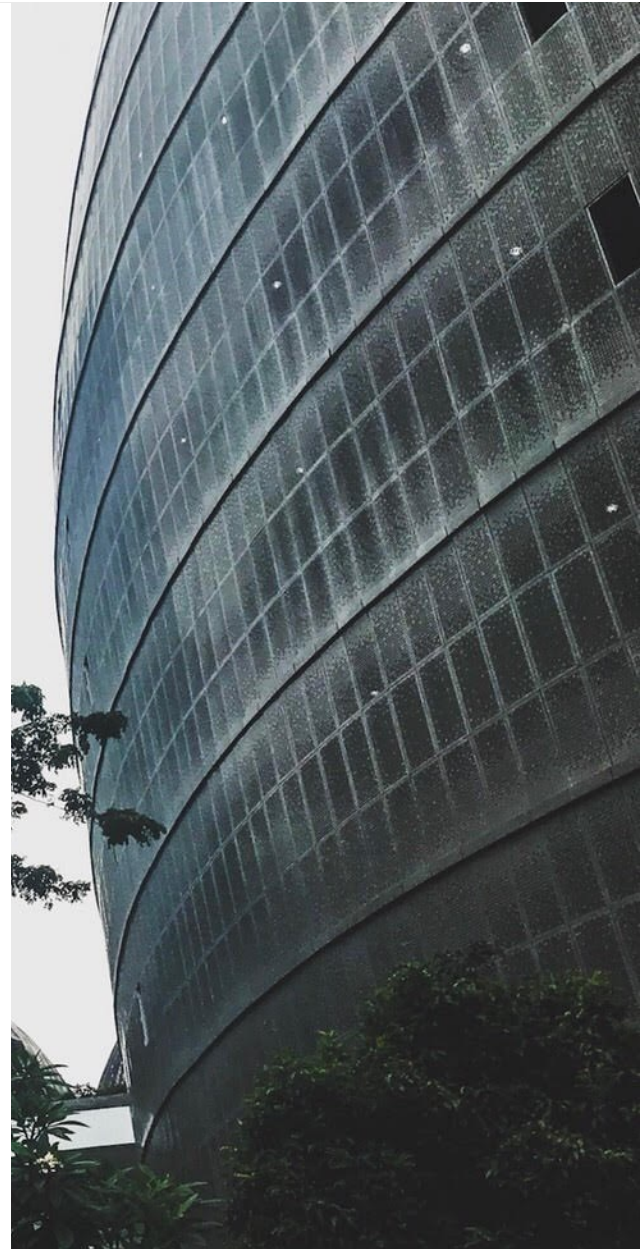


# MODUL PRAKTIKUM

IS411 – DATA MODELING  
PROGRAM SARJANA S1 SISTEM INFORMASI  
FAKULTAS TEKNIK DAN INFORMATIKA

---



---

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA**

Gedung B Lantai 5, Kampus UMN  
Jl. Scientia Boulevard, Gading Serpong, Tangerang, Banten-15811 Indonesia  
Telp: +62-21.5422.0808 (ext. 1803), email: [ict.lab@umn.ac.id](mailto:ict.lab@umn.ac.id), web: [umn.ac.id](http://umn.ac.id)

# MODUL 6

## STATISTIC FOR DATA SCIENCE



### DESKRIPSI TEMA

1. Measure of Variability
2. Range
3. Quartile, interquartile, percentiles
4. Using formula to calculating variance and standard deviation for the sample and population
5. Probability Distribution on statistics

### CAPAIAN PEMBELAJARAN MINGGUAN (SUB-CAPAIAN PEMBELAJARAN)

Students are able to implement Statistics Methods for Data Science (C<sub>3</sub>)

### PENUNJANG PRAKTIKUM

1. Anaconda Navigator
  2. Jupyter Notebook
- (+ Perlengkapan Apd/Alat Pelindung Diri Yang Harus Digunakan, Jika Ada)

### LANGKAH-LANGKAH PRAKTIKUM

#### Import Library

1. Lakukan langkah yang sama seperti minggu sebelumnya untuk meng-import library Numpy dan Pandas.

#### Import Data

2. Gunakan dataset **winequality-red.csv** dan masukkan ke dalam dataframe menggunakan Pandas. Berikan nama **dataset\_red** (Petunjuk: gunakan delimiter=; ).
3. Tampilkan informasi dataset dan isi data.

#### Menampilkan Statistik Data

Beberapa function perhitungan statistik:

- a. Mean, digunakan untuk menghitung rata-rata. Dalam dataset Pandas, Mean dapat dihitung menggunakan function `mean()`. Mean bisa digunakan secara spesifik pada 1 kolom atau keseluruhan dataset.
4. Menampilkan Mean dari salah satu kolom.

```
1 #Menampilkan mean dari salah satu kolom
2 print('mean dari fixed acidity : ', dataset_red['fixed acidity'].mean())
```

5. Menampilkan Mean dari keseluruhan data.

```
1 #Menampilkan mean dari keseluruhan data
2 print(dataset_red.mean())
```

- b. Median, digunakan untuk menemukan nilai tengah dan menentukan bagian atas atau bawah dari kumpulan data, menggunakan function median(). Sama seperti Mean, Median juga dapat digunakan pada dataset Pandas secara spesifik atau keseluruhan dataset.

6. Menampilkan Median salah satu kolom.

```
1 #Menampilkan median salah satu kolom
2 print('median dari fixed acidity : ', dataset_red['fixed acidity'].median())
3
```

7. Menampilkan Median keseluruhan data.

```
1 #Menampilkan median semua data
2 print(dataset_red.median())
```

8. Jika data Median yang ingin dicari adalah per baris, dapat menggunakan atribut axis =1.

```
1 #Jika data median yang ingin dicari adalah per row, dapat menggunakan atribut axis = 1
2 print(dataset_red.median(axis=1))
```

- c. Mode, digunakan untuk menghitung data yang paling banyak muncul, menggunakan function mode(). Berikut contoh penggunaannya pada dataset Pandas.

9. Menampilkan Modus salah satu kolom.

```
1 #Menampilkan modus dari salah satu kolom
2 print('mode dari fixed acidity : ', dataset_red['fixed acidity'].mode())
```

10. Menampilkan Modus keseluruhan data.

```
1 #Menampilkan modus semua data
2 print(dataset_red.mode())
```

- d. Standard Deviation, ~~digunakan untuk menghitung banyaknya variasi data terhadap mean dari~~ tersebut. Semakin sedikit angka standard deviation artinya data semakin mendekati mean, sebaliknya. Untuk menghitung standard deviation pada pandas dataset dapat menggunakan function `std()`.

11. Menampilkan Standard Deviation salah satu kolom.

```
1 #Menampilkan Standard Deviation salah satu kolom
2 print('standard deviation dari fixed acidity : ', dataset_red['fixed acidity'].std())
```

12. Menampilkan Standard Deviation keseluruhan data.

```
1 #Menampilkan Standard Deviation semua data
2 print(dataset_red.std())
```

- e. Variance, digunakan untuk menghitung penyebaran data, menggunakan function `var()`. Merupakan kuadrat dari standard deviasi, dan menjadi kovarian, atau hubungan dari salah satu variable ke variable lainnya.

13. Menampilkan Variance salah satu kolom.

```
1 #Menampilkan Variance salah satu kolom
2 print('variance dari fixed acidity : ', dataset_red['fixed acidity'].var())
```

14. Menampilkan Variance keseluruhan data.

```
1 #Menampilkan Variance semua data
2 print(dataset_red.var())
```

- f. Quartile, adalah persentase 25%, 50%, dan 75% yang membagi data menjadi 4 bagian dengan rata. Jika dibandingkan dengan percentile, quartile dapat dituliskan sebagai berikut. Q0 = 0%, Q1 = 25%, Q2 = 50% (sama dengan nilai median), Q3 = 75%, Q4 = 100%.

15. Menampilkan Q1 – Q4

```
1 print('Q1 dari fixed acidity : ', np.quantile(dataset_red['fixed acidity'], 0.25))
2 print('Q2 dari fixed acidity : ', np.quantile(dataset_red['fixed acidity'], 0.5))
3 print('Q2 atau median dari fixed acidity : ', dataset_red['fixed acidity'].median())
4 print('Q3 dari fixed acidity : ', np.quantile(dataset_red['fixed acidity'], 0.75))
```

- g. Interquartile range (IQR) dapat dihitung dengan rumus sederhana  $Q3 - Q1$ , atau dengan bantuan dari `scipy.stats` library.

16. Menghitung Interkuartil



```

1 from scipy.stats import iqr
2
3 print('Menggunakan rumus Q3 - Q1 : '
4       ,np.quantile(dataset_red['fixed acidity'], 0.75) - np.quantile(dataset_red['fixed acidity'], 0.25))
5 print('Menggunakan bantuan library scipy.stats',iqr(dataset_red['fixed acidity']))

```

17. Cara lain untuk melihat data dengan perhitungan statistik secara lengkap adalah menggunakan **describe()**, dan tambahkan attribute **include = 'all'**

```

1 dataset_red.describe(include='all')

```

Selain menggunakan function yang disediakan dari Pandas, Python juga mempunyai library untuk menghitung statistik menggunakan library statistics. Function dari library ini meliputi **mean()**, **mode()**, **median()** setara dengan Quartile 2, **median\_low()** atau setara dengan Quartile 1, **median\_high()** atau setara dengan Quartile 3, **stdev()**, dan **variance()**.

Function statistics tidak bisa digunakan untuk keseluruhan dataset seperti Pandas, tetapi harus spesifik per kolom dataset.

```

1 import statistics as st
2
3 print('Mean dari Fixed Acidity : ', st.mean(dataset_red['fixed acidity']))
4 print('Q1 dari Fixed Acidity : ', st.median_low(dataset_red['fixed acidity']))
5 print('Median dari Fixed Acidity : ', st.median(dataset_red['fixed acidity']))
6 print('Q3 dari Fixed Acidity : ', st.median_high(dataset_red['fixed acidity']))
7 print('Modus dari Fixed Acidity : ', st.mode(dataset_red['fixed acidity']))
8 print('Standard Deviation dari Fixed Acidity : ', st.stdev(dataset_red['fixed acidity']))
9 print('Variance dari Fixed Acidity : ', st.variance(dataset_red['fixed acidity']))

```

## Challenge

1. Lakukan langkah – langkah yang sama dengan dataset Red Wine menggunakan dataset **winequality-white.csv**. Untuk kolom yang digunakan adalah **Density**.
2. Pada fungsi describe() tambahkan dengan parameter selain yang ada di contoh.

## PENGUMPULAN

1. File yang dikumpulkan terdiri dari:
  - a. File project (.ipynb)
  - b. File PDF berisi screenshot output dan jawaban (jika ada pertanyaan)
2. File di-compress (.zip) dan diberi nama **KODEMATAKULIAH\_KELAS\_NIM\_NAMA\_WEEK KE-XX.zip** (contoh: IS5411\_A\_13110310017\_Monika Evelin Johan\_Week-01.zip).

## REFERENSI

Deitel, P., & Deitel, H. (2020). Intro to Python for Computer Science and Data Science. Pearson Education.