

Nama : Afina Putri Dayanti
NIM : 825200049
Prodi : Sistem Informasi

UAS Big Data

Pilihan Ganda

1	C	21	B, C, E
2	D	22	A, C, E
3	B	23	A, B, C, E
4	A	24	B
5	B	25	B
6	A, B, D, E	26	A, C, D
7	D	27	C
8	C	28	A, D
9	D	29	B
10	A, C, D, E	30	A, D
11	C	31	B, C
12	E	32	B
13	B, D, E	33	A
14	D	34	A
15	A, E	35	A, E
16	D	36	A
17	A	37	E
18	B	38	A
19	A	39	A
20	B	40	A, B, C

Essay

41. Berikut ini merupakan sampel data pasien Covid-19 yang bertahan hidup dan meninggal yang ditunjukkan oleh variabel survival (hidup = 0, meninggal = 1). Asumsikan bahwa Anda memiliki 10000 sampel data dengan jumlah pasien hidup 8000 orang dan meninggal 2000 orang. Anda diminta untuk melakukan analisis pada data berikut!

Sebutkan dan jelaskan teknik analisis Big Data yang Anda gunakan!

Jawab : **Teknik analisis data kuantitatif** karena data tersebut bersifat angka atau data numerik dengan jumlah data yang sangat besar sehingga teknik kuantitatif merupakan analisis data yang tepat untuk dihitung secara tepat dengan perhitungan rumus statistik. Data-data kuantitatif tersebut nantinya dapat berupa survei, arsip data, peringkat, dan sebagainya.

42. Tulis Python code untuk:

- a. Load data set tersebut dengan menggunakan Pandas. Lokasi penyimpanan file csv ditentukan sendiri.

```
import pandas as pd
datasetName = "Downloads/DataCovid19.csv"
dataCovid19 = pd.read_csv(datasetName)
```

```
In [1]: import pandas as pd
datasetName = "Downloads/DataCovid19.csv"
dataCovid19 = pd.read_csv(datasetName)
```

```
In [2]: dataCovid19
```

```
Out[2]:
```

	Detak jantung	SpO2	Pasang_02	Temperatur	Respiratory Rate	Survival
0	120	97.711213	0	37.0	20	0
1	3	99.000000	0	36.3	20	0
2	102	99.000000	0	36.3	20	0
3	1	100.000000	0	36.0	20	0
4	88	100.000000	0	36.0	20	0
5	1	97.711213	0	36.0	20	0
6	89	97.711213	0	36.0	20	0
7	1	97.000000	0	37.0	22	0
8	110	99.000000	0	36.1	23	0
9	106	97.711213	0	37.5	20	0
10	120	97.711213	0	36.5	20	0
11	110	95.665794	0	36.4	22	1
12	90	100.000000	0	36.5	20	1
13	90	100.000000	0	36.5	20	1
14	122	95.665794	0	37.0	22	1
15	140	92.000000	0	37.6	20	1

- b. Tampilkan 5 (lima) record terakhir dari data set tersebut.

```
dataCovid19.tail()
```

```
In [3]: dataCovid19.tail()
```

```
Out[3]:
```

	Detak jantung	SpO2	Pasang_02	Temperatur	Respiratory Rate	Survival
11	110	95.665794	0	36.4	22	1
12	90	100.000000	0	36.5	20	1
13	90	100.000000	0	36.5	20	1
14	122	95.665794	0	37.0	22	1
15	140	92.000000	0	37.6	20	1

```
In [ ]:
```

- c. Tampilkan jumlah baris dan jumlah kolom dari data set tersebut

```
dataCovid19.shape
```

```
In [5]: dataCovid19.shape
```

```
Out[5]: (16, 6)
```

- d. Tampilkan data pasien yang memiliki temperature tertinggi dan terendah.

```
longest = dataCovid19[dataCovid19['Temperatur']==dataCovid19['Temperatur'].max()]  
shortest = dataCovid19[dataCovid19['Temperatur']==dataCovid19['Temperatur'].min()]
```

```
In [6]: longest = dataCovid19[dataCovid19['Temperatur']==dataCovid19['Temperatur'].max()]  
shortest = dataCovid19[dataCovid19['Temperatur']==dataCovid19['Temperatur'].min()]  
longest.append(shortest)
```

Out[6]:

	Detak jantung	SpO2	Pasang_02	Temperatur	Respiratory Rate	Survival
15	140	92.000000	0	37.6	20	1
3	1	100.000000	0	36.0	20	0
4	88	100.000000	0	36.0	20	0
5	1	97.711213	0	36.0	20	0
6	89	97.711213	0	36.0	20	0

- e. Tampilkan data pasien dengan temperatur di atas 36 dan respiratory rate di atas 20.

```
dataCovid19[(dataCovid19['Temperatur']>36) & (dataCovid19['Respiratory Rate']>20)]
```

```
In [8]: dataCovid19[(dataCovid19['Temperatur']>36) & (dataCovid19['Respiratory Rate']>20)]
```

Out[8]:

	Detak jantung	SpO2	Pasang_02	Temperatur	Respiratory Rate	Survival
7	1	97.000000	0	37.0	22	0
8	110	99.000000	0	36.1	23	0
11	110	95.665794	0	36.4	22	1
14	122	95.665794	0	37.0	22	1

- f. Tampilkan data 20 pasien dengan nilai rata-rata SpO2 tertinggi.

```
dataCovid19['SpO2'].sort_values(ascending=False)[0:20]
```

```
In [19]: dataCovid19['SpO2'].sort_values(ascending=False)[0:20]
```

```
Out[19]: 3    100.000000  
4    100.000000  
12   100.000000  
13   100.000000  
1    99.000000  
2    99.000000  
8    99.000000  
0    97.711213  
5    97.711213  
6    97.711213  
9    97.711213  
10   97.711213  
7    97.000000  
11   95.665794  
14   95.665794  
15   92.000000  
Name: SpO2, dtype: float64
```

- g. Ubah nama kolom pada data set tersebut menjadi `detak_jantung`, `nilai_SpO2`, `O2_pasang`, `nilai_temperatur`, `respiratory_rate`, dan `survival_rate`.
`dataCovid19.columns = ['detak_jantung', 'nilai_SpO2', 'O2_pasang', 'nilai_temperatur', 'respiratory_rate', 'survival_rate']`

```
In [21]: dataCovid19.columns = ['detak_jantung', 'nilai_SpO2', 'O2_pasang', 'nilai_temperatur', 'respiratory_rate', 'survival_rate']
dataCovid19.columns
```

```
Out[21]: Index(['detak_jantung', 'nilai_SpO2', 'O2_pasang', 'nilai_temperatur',
               'respiratory_rate', 'survival_rate'],
              dtype='object')
```

```
In [22]: dataCovid19
```

```
Out[22]:
```

	detak_jantung	nilai_SpO2	O2_pasang	nilai_temperatur	respiratory_rate	survival_rate
0	120	97.711213	0	37.0	20	0
1	3	99.000000	0	36.3	20	0
2	102	99.000000	0	36.3	20	0
3	1	100.000000	0	36.0	20	0
4	88	100.000000	0	36.0	20	0
5	1	97.711213	0	36.0	20	0
6	89	97.711213	0	36.0	20	0
7	1	97.000000	0	37.0	22	0
8	110	99.000000	0	36.1	23	0
9	106	97.711213	0	37.5	20	0
10	120	97.711213	0	36.5	20	0
11	110	95.665794	0	36.4	22	1
12	90	100.000000	0	36.5	20	1
13	90	100.000000	0	36.5	20	1
14	122	95.665794	0	37.0	22	1
15	140	92.000000	0	37.6	20	1