

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
In [2]: df = pd.read_csv("case_cost_prediction_val.csv")
df.head()
```

Out[2]:

	row_id	tglpelayanan	kddati2	tkp	peserta	a	b	c	cb	d	...	km	ko	kp	kt	ku	s	sa	sb	sc	sd
0	1	2020-02-01 00:00:00	257	40	305959	0	0	2	0	0	...	0	0	0	0	0	0	0	0	0	1
1	2	2014-08-01 00:00:00	228	30	214545	0	1	0	0	0	...	0	0	0	0	0	0	0	0	3	0
2	3	2014-03-01 00:00:00	345	30	95774	0	0	1	0	0	...	0	0	0	0	0	0	0	0	0	0
3	4	2017-05-01 00:00:00	6	30	194254	0	1	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	5	2015-09-01 00:00:00	119	40	1231866	0	1	1	0	0	...	0	0	0	0	0	0	0	0	0	0

5 rows × 34 columns

```
In [5]: to_drop = ['row_id', 'peserta',
                  'row_id',
                  'tglpelayanan',
                  'kddati2',]
```

```
In [6]: df.drop(to_drop, inplace=True, axis=1)
```

```
In [7]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24633 entries, 0 to 24632
Data columns (total 30 columns):
#   Column      Non-Null Count  Dtype
---  ---
0    tkp         24633 non-null   int64
1    a           24633 non-null   int64
2    b           24633 non-null   int64
3    c           24633 non-null   int64
4    cb          24633 non-null   int64
5    d           24633 non-null   int64
6    ds          24633 non-null   int64
7    gd          24633 non-null   int64
8    hd          24633 non-null   int64
9    i1          24633 non-null   int64
10   i2          24633 non-null   int64
11   i3          24633 non-null   int64
12   i4          24633 non-null   int64
13   kb          24633 non-null   int64
14   kc          24633 non-null   int64
15   kg          24633 non-null   int64
16   ki          24633 non-null   int64
17   kj          24633 non-null   int64
18   kk          24633 non-null   int64
19   kl          24633 non-null   int64
20   km          24633 non-null   int64
21   ko          24633 non-null   int64
22   kp          24633 non-null   int64
23   kt          24633 non-null   int64
24   ku          24633 non-null   int64
25   s           24633 non-null   int64
26   sa          24633 non-null   int64
27   sb          24633 non-null   int64
28   sc          24633 non-null   int64
29   sd          24633 non-null   int64
dtypes: int64(30)
memory usage: 5.6 MB
```

```
In [8]: df.head()
```

Out[8]:

	tkp	a	b	c	cb	d	ds	gd	hd	i1	...	km	ko	kp	kt	ku	s	sa	sb	sc	sd
0	40	0	0	2	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
1	30	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	3	0
2	30	0	0	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	30	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	40	0	1	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

5 rows × 30 columns

```
In [9]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24633 entries, 0 to 24632
Data columns (total 30 columns):
#   Column  Non-Null Count  Dtype
---  -
0    tkp      24633 non-null    int64
1     a      24633 non-null    int64
2     b      24633 non-null    int64
3     c      24633 non-null    int64
4    cb      24633 non-null    int64
5     d      24633 non-null    int64
6    ds      24633 non-null    int64
7    gd      24633 non-null    int64
8    hd      24633 non-null    int64
9    i1      24633 non-null    int64
10   i2      24633 non-null    int64
11   i3      24633 non-null    int64
12   i4      24633 non-null    int64
13   kb      24633 non-null    int64
14   kc      24633 non-null    int64
15   kg      24633 non-null    int64
16   ki      24633 non-null    int64
17   kj      24633 non-null    int64
18   kk      24633 non-null    int64
19   kl      24633 non-null    int64
20   km      24633 non-null    int64
21   ko      24633 non-null    int64
22   kp      24633 non-null    int64
23   kt      24633 non-null    int64
24   ku      24633 non-null    int64
25    s      24633 non-null    int64
26   sa      24633 non-null    int64
27   sb      24633 non-null    int64
28   sc      24633 non-null    int64
29   sd      24633 non-null    int64
dtypes: int64(30)
memory usage: 5.6 MB
```

```
In [10]: x = df.drop(["tkp"], axis=1)
x.head(11)
```

Out[10]:

	a	b	c	cb	d	ds	gd	hd	i1	i2	...	km	ko	kp	kt	ku	s	sa	sb	sc	sd
0	0	0	2	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
1	0	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	3	0
2	0	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	0	1	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
5	0	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
6	0	1	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
7	0	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	1	1	3	
8	0	1	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
9	0	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

11 rows × 29 columns

```
In [11]: y = df["tkp"]
y.head(11)
```

Out[11]:

0	40
1	30
2	30
3	30
4	40
5	30
6	40
7	30
8	30
9	40
10	30

Name: tkp, dtype: int64

```
In [12]: from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB

modelnb = GaussianNB()
```

```
In [13]: nbtrain = modelnb.fit(x, y)
df.head(11)
```

```
Out[13]:
```

	tkp	a	b	c	cb	d	ds	gd	hd	l1	...	km	ko	kp	kt	ku	s	sa	sb	sc	sd
0	40	0	0	2	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
1	30	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	3	0
2	30	0	0	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	30	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	40	0	1	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
5	30	0	0	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
6	40	0	1	0	0	1	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
7	30	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	1	1	3
8	30	0	1	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
9	40	0	0	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
10	30	0	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

11 rows × 30 columns

```
In [14]: x_test = df.drop(["tkp"], axis=1)
x_test.head(11)
```

```
Out[14]:
```

	a	b	c	cb	d	ds	gd	hd	l1	l2	...	km	ko	kp	kt	ku	s	sa	sb	sc	sd
0	0	0	2	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
1	0	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	3	0
2	0	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	0	1	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
5	0	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
6	0	1	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
7	0	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	1	1	3
8	0	1	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
9	0	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

11 rows × 29 columns

```
In [15]: y_uji = df["tkp"]
y_uji.head(11)
```

```
Out[15]:
```

0	40
1	30
2	30
3	30
4	40
5	30
6	40
7	30
8	30
9	40
10	30

Name: tkp, dtype: int64

```
In [16]: Y_predict = nbtrain.predict(x_test)
print("Prediksi Naive Bayes : ",Y_predict)
```

Prediksi Naive Bayes : [40 40 40 ... 40 40 40]

```
In [17]: from sklearn.metrics import accuracy_score
accuracy= accuracy_score(y_uji, Y_predict)
print("Akurasi Naive Bayes : ",accuracy)
```

Akurasi Naive Bayes : 0.5405756505500751

```
In [18]: from sklearn.metrics import classification_report
print(classification_report(y_uji, Y_predict))
```

	precision	recall	f1-score	support
30	0.68	0.14	0.24	12250
40	0.52	0.93	0.67	12383
accuracy			0.54	24633
macro avg	0.60	0.54	0.45	24633
weighted avg	0.60	0.54	0.46	24633