

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

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

Out[2]:
```

	visit_id	kdkc	dati2	typeppk	jkpst	umur	jnspelsep	los	cmg	severitylevel	...	proc58_62	proc63_67	proc68_70	proc71_73	proc74_75	proc76_77	proc
0	1	102	1	SD	P	40	2	0	Q	0	...	0	0	0	0	0	0	
1	2	102	1	SD	L	12	2	0	Q	0	...	0	0	0	0	0	0	
2	3	102	1	SD	L	50	2	0	K	0	...	0	0	0	0	0	0	
3	4	102	1	SD	P	41	2	0	K	0	...	0	0	0	0	0	0	
4	5	102	1	SD	P	67	1	3	I	1	...	0	0	0	0	0	0	

5 rows × 52 columns

```
In [3]: to_drop = ['visit_id', 'kdkc', 'cmg', 'diagprimer', 'typeppk']

In [4]: df.drop(to_drop, inplace=True, axis=1)

In [5]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49762 entries, 0 to 49761
Data columns (total 47 columns):
#   Column                Non-Null Count  Dtype
---  -
0   dati2                 49762 non-null  int64
1   jkpst                 49762 non-null  object
2   umur                  49762 non-null  int64
3   jnspelsep             49762 non-null  int64
4   los                   49762 non-null  int64
5   severitylevel         49762 non-null  int64
6   dx2_a00_b99           49762 non-null  int64
7   dx2_c00_d48           49762 non-null  int64
8   dx2_d50_d89           49762 non-null  int64
9   dx2_e00_e90           49762 non-null  int64
10  dx2_f00_f99           49762 non-null  int64
11  dx2_g00_g99           49762 non-null  int64
12  dx2_h00_h59           49762 non-null  int64
13  dx2_h60_h95           49762 non-null  int64
14  dx2_i00_i99           49762 non-null  int64
15  dx2_j00_j99           49762 non-null  int64
16  dx2_koo_k93           49762 non-null  int64
17  dx2_l00_l99           49762 non-null  int64
18  dx2_m00_m99           49762 non-null  int64
19  dx2_n00_n99           49762 non-null  int64
20  dx2_o00_o99           49762 non-null  int64
21  dx2_p00_p96           49762 non-null  int64
22  dx2_q00_q99           49762 non-null  int64
23  dx2_r00_r99           49762 non-null  int64
24  dx2_s00_t98           49762 non-null  int64
25  dx2_u00_u99           49762 non-null  int64
26  dx2_v01_y98           49762 non-null  int64
27  dx2_z00_z99           49762 non-null  int64
28  proc00_13             49762 non-null  int64
29  proc14_23             49762 non-null  int64
30  proc24_27             49762 non-null  int64
31  proc28_28             49762 non-null  int64
32  proc29_31             49762 non-null  int64
33  proc32_38             49762 non-null  int64
34  proc39_45             49762 non-null  int64
35  proc46_51             49762 non-null  int64
36  proc52_57             49762 non-null  int64
37  proc58_62             49762 non-null  int64
38  proc63_67             49762 non-null  int64
39  proc68_70             49762 non-null  int64
40  proc71_73             49762 non-null  int64
41  proc74_75             49762 non-null  int64
42  proc76_77             49762 non-null  int64
43  proc78_79             49762 non-null  int64
44  proc80_99             49762 non-null  int64
45  proce00_e99           49762 non-null  int64
46  procv00_v89           49762 non-null  int64
dtypes: int64(46), object(1)
memory usage: 17.8+ MB
```

```
In [6]: x = df.drop(["jpkst"], axis=1)
x.head(11)
```

```
Out[6]:
```

	dati2	umur	jnspelsep	los	severitylevel	dx2_a00_b99	dx2_c00_d48	dx2_d50_d89	dx2_e00_e90	dx2_f00_f99	...	proc58_62	proc63_67	proc68_70	proc7
0	1	40	2	0	0	0	0	0	0	0	...	0	0	0	
1	1	12	2	0	0	0	0	0	0	0	...	0	0	0	
2	1	50	2	0	0	0	0	0	0	0	...	0	0	0	
3	1	41	2	0	0	0	0	0	0	0	...	0	0	0	
4	1	67	1	3	1	0	0	0	0	0	...	0	0	0	
5	1	5	2	0	0	0	0	0	0	0	...	0	0	0	
6	1	66	1	1	1	0	0	0	0	0	...	0	0	0	
7	1	71	1	4	1	0	0	0	0	0	...	0	0	0	
8	1	44	2	0	0	0	0	0	0	0	...	0	0	0	
9	1	30	1	4	1	0	0	0	0	0	...	0	0	0	
10	1	66	2	0	0	0	0	0	0	0	...	0	0	0	

11 rows × 46 columns

```
In [7]: y = df["jpkst"]
y.head(11)
```

```
Out[7]: 0    P
1    L
2    L
3    P
4    P
5    L
6    P
7    L
8    L
9    P
10   L
Name: jpkst, dtype: object
```

```
In [8]: from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB

modelnb = GaussianNB()
```

```
In [9]: nbtrain = modelnb.fit(x, y)
df.head(11)
```

```
Out[9]:
```

	dati2	jpkst	umur	jnspelsep	los	severitylevel	dx2_a00_b99	dx2_c00_d48	dx2_d50_d89	dx2_e00_e90	...	proc58_62	proc63_67	proc68_70	proc71_73
0	1	P	40	2	0	0	0	0	0	0	...	0	0	0	0
1	1	L	12	2	0	0	0	0	0	0	...	0	0	0	0
2	1	L	50	2	0	0	0	0	0	0	...	0	0	0	0
3	1	P	41	2	0	0	0	0	0	0	...	0	0	0	0
4	1	P	67	1	3	1	0	0	0	0	...	0	0	0	0
5	1	L	5	2	0	0	0	0	0	0	...	0	0	0	0
6	1	P	66	1	1	1	0	0	0	0	...	0	0	0	0
7	1	L	71	1	4	1	0	0	0	0	...	0	0	0	0
8	1	L	44	2	0	0	0	0	0	0	...	0	0	0	0
9	1	P	30	1	4	1	0	0	0	0	...	0	0	0	0
10	1	L	66	2	0	0	0	0	0	0	...	0	0	0	0

11 rows × 47 columns

In [11]:

x_test = df.drop(["jkpst"], axis=1)
x_test.head(11)

Out[11]:

	dati2	umur	jnsplsep	los	severitylevel	dx2_a00_b99	dx2_c00_d48	dx2_d50_d89	dx2_e00_e90	dx2_f00_f99	...	proc58_62	proc63_67	proc68_70	proc7
0	1	40	2	0	0	0	0	0	0	0	...	0	0	0	
1	1	12	2	0	0	0	0	0	0	0	...	0	0	0	
2	1	50	2	0	0	0	0	0	0	0	...	0	0	0	
3	1	41	2	0	0	0	0	0	0	0	...	0	0	0	
4	1	67	1	3	1	0	0	0	0	0	...	0	0	0	
5	1	5	2	0	0	0	0	0	0	0	...	0	0	0	
6	1	66	1	1	1	0	0	0	0	0	...	0	0	0	
7	1	71	1	4	1	0	0	0	0	0	...	0	0	0	
8	1	44	2	0	0	0	0	0	0	0	...	0	0	0	
9	1	30	1	4	1	0	0	0	0	0	...	0	0	0	
10	1	66	2	0	0	0	0	0	0	0	...	0	0	0	

11 rows × 46 columns

In [12]:

y_uji = df["jkpst"]
y_uji.head(11)

Out[12]:

0	P
1	L
2	L
3	P
4	P
5	L
6	P
7	L
8	L
9	P
10	L

Name: jkpst, dtype: object

In [13]:

Y_predict = nbtrain.predict(x_test)
print("Prediksi Naive Bayes : ",Y_predict)

Prediksi Naive Bayes : ['L' 'L' 'L' ... 'L' 'P' 'L']

In [14]:

from sklearn.metrics import accuracy_score
accuracy= accuracy_score(y_uji, Y_predict)
print("Akurasi Naive Bayes : ",accuracy)

Akurasi Naive Bayes : 0.5093444797234837

In [15]:

Menghitung nilai akurasi dari klasifikasi naive bayes
from sklearn.metrics import classification_report
print(classification_report(y_uji, Y_predict))

	precision	recall	f1-score	support
L	0.47	0.95	0.63	22169
P	0.79	0.16	0.26	27593
accuracy			0.51	49762
macro avg	0.63	0.55	0.45	49762
weighted avg	0.65	0.51	0.43	49762

In []: