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

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

Out[2]:

	Kata	Count
0	obat	49957
1	anak	36150
2	nyeri	30711
3	darah	29588
4	perut	28098

In [4]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1089 entries, 0 to 1088
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   Kata    1089 non-null   object
 1   Count   1089 non-null   int64 
dtypes: int64(1), object(1)
memory usage: 17.1+ KB
```

In [5]:
```
x = df.drop(["Kata"], axis=1)
x.head(11)
```

Out[5]:

	Count
0	49957
1	36150
2	30711
3	29588
4	28098
5	27902
6	25498
7	24185
8	22418
9	22122
10	21587

In [6]:
```
y = df["Kata"]
y.head(11)
```

Out[6]:
```
0       obat
1       anak
2      nyeri
3      darah
4      perut
5       haid
6      minum
7      hamil
8       mata
9     kepala
10     gatal
Name: Kata, dtype: object
```

In [7]:
```
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB

modelnb = GaussianNB()
```

In [8]:
```
nbtrain = modelnb.fit(x, y)
df.head(11)
```

Out[8]:

	Kata	Count
0	obat	49957
1	anak	36150
2	nyeri	30711
3	darah	29588
4	perut	28098
5	haid	27902
6	minum	25498
7	hamil	24185
8	mata	22418
9	kepala	22122
10	gatal	21587

In [9]:
```
x_test = df.drop(["Kata"], axis=1)
x_test.head(11)
```

Out[9]:

	Count
0	49957
1	36150
2	30711
3	29588
4	28098
5	27902
6	25498
7	24185
8	22418
9	22122
10	21587

In [10]:
```
y_uji = df["Kata"]
y_uji.head(11)
```

Out[10]:
```
0       obat
1       anak
2      nyeri
3      darah
4      perut
5       haid
6      minum
7      hamil
8       mata
9     kepala
10     gatal
Name: Kata, dtype: object
```

In [11]:
```
Y_predict = nbtrain.predict(x_test)
print("Prediksi Naive Bayes : ",Y_predict)
```

```
Prediksi Naive Bayes :  ['obat' 'anak' 'nyeri' ... 'hdl' 'bisoprolol' 'bisoprolol']
```

In [12]:
```
from sklearn.metrics import accuracy_score
accuracy= accuracy_score(y_uji, Y_predict)
print("Akurasi Naive Bayes : ",accuracy)
```

```
Akurasi Naive Bayes :  0.6740128558310376
```

In [13]:
```
from sklearn.metrics import classification_report
print(classification_report(y_uji, Y_predict))
```

```
                    precision    recall  f1-score   support

           abdomen       0.12      1.00      0.22         1
             abses       0.50      1.00      0.67         1
               abu       0.50      1.00      0.67         1
              acid       0.50      1.00      0.67         1
              acne       0.17      1.00      0.29         1
         acyclovir       1.00      1.00      1.00         1
              aids       1.00      1.00      1.00         1
               air       1.00      1.00      1.00         1
              akar       0.25      1.00      0.40         1
           albumin       0.17      1.00      0.29         1
             alergi       1.00      1.00      1.00         1
         alerginya       0.17      1.00      0.29         1
              alis       1.00      1.00      1.00         1
           alkohol       0.50      1.00      0.67         1
           alpukat       0.12      1.00      0.22         1
           amandel       1.00      1.00      1.00         1
           ambeien       1.00      1.00      1.00         1
```

In [ ]: