```
import pandas as pds
import numpy as npy
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from scipy.spatial.distance import hamming
```

```
diabetes = pds.read_csv("diabetes.csv")
diabetes.head(6)
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPe
0	6	148	72	35	0	33.6	
1	1	85	66	29	0	26.6	
2	8	183	64	0	0	23.3	
3	1	89	66	23	94	28.1	
4	0	137	40	35	168	43.1	
5	5	116	74	0	0	25.6	

```
diabetes.describe()
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.0
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.3
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.0
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.6
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.1

```
print("Number of rows in the dataset : ", diabetes.shape[0])
training_data, testing_data = train_test_split(diabetes, test_size=0.2, random_state=25)
```

```
Number of rows in the dataset :  768
```

```
training_data.describe()
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	
count	614.000000	614.000000	614.000000	614.000000	614.000000	614.0
mean	3.842020	121.293160	69.587948	20.688925	81.203583	32.
std	3.409438	32.329746	18.992518	15.932647	116.372311	7.
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
25%	1.000000	99.000000	64.000000	0.000000	0.000000	27.3
50%	3.000000	117.000000	72.000000	23.000000	34.000000	32.0

```
model = GaussianNB()
model.fit(training_data.iloc[:,0:8], training_data.iloc[:,8])
```

```
GaussianNB(priors=None, var_smoothing=1e-09)
```

```
predicted_data = model.predict(testing_data.iloc[:,0:8])
hamming_distance = hamming(predicted_data, testing_data.iloc[:, 8])
accuracy = 1-hamming_distance
print("Accuracy (in percentage) : ", accuracy*100)
```

```
Accuracy (in percentage) :  76.62337662337663
```

✓ 0s completed at 19:48