```
[7] from google.colab import files
    from google.colab import drive
    drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn import neighbors
import pandas as pd
```

```
[9] df = pd.read_csv('/content/drive/MyDrive/ML/diabetes.csv')
    df.head(5)
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

```
[10] from sklearn.model_selection import train_test_split
     from sklearn.metrics import accuracy_score
     X_train, X_test, y_train, y_test = train_test_split(df.iloc[:,0:8], df.iloc[:,8], test_size=0.9, random_state= 42)
     print('Training dataset size:', len(y_train))
     print('Testing dataset size:', len(y_test))
```

```
Training dataset size: 76
Testing dataset size: 692
```

Với K = 3

```
[11] model = neighbors.KNeighborsClassifier(n_neighbors = 3, p = 2)
     model.fit(X_train, y_train)
     y_pred = model.predict(X_test)
     print("Accuracy score: %.2f %%" %(100*accuracy_score(y_test, y_pred)))
```

```
Accuracy score: 71.24 %
```

Nhận xét: Chỉ 71.24 % tổng số mẫu là được phân loại chính xác

Với K = 7

```
model = neighbors.KNeighborsClassifier(n_neighbors = 7, p = 2)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
print("Accuracy score: %.2f %%" %(100*accuracy_score(y_test, y_pred)))
```

```
Accuracy score: 68.93 %
```

Nhận xét: Chỉ 68.93 % tổng số mẫu là được phân loại chính xác