```python
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.metrics import accuracy_score
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler


df=pd.read_csv('diabetes.csv')
print(df.head())
```

```
   Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin
BMI  \
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


   DiabetesPedigreeFunction  Age  Outcome
0                     0.627   50        1
1                     0.351   31        0
2                     0.672   32        1
3                     0.167   21        0
4                     2.288   33        1
```

```python
df.drop('Outcome',axis=1,inplace=True)
print(df.head())
```

```
   Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin
BMI  \
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


   DiabetesPedigreeFunction  Age
0                     0.627   50
1                     0.351   31
2                     0.672   32
```

```
3                       0.167   21
4                       2.288   33
```

```
print(df.isnull().sum())
```

```
Pregnancies                 0
Glucose                     0
BloodPressure               0
SkinThickness               0
Insulin                     0
BMI                         0
DiabetesPedigreeFunction    0
Age                         0
dtype: int64
```

```
X_std=StandardScaler().fit_transform(df)
```

```
kmeans=KMeans(n_clusters=2,random_state=42)
kmeans.fit(X_std)
```

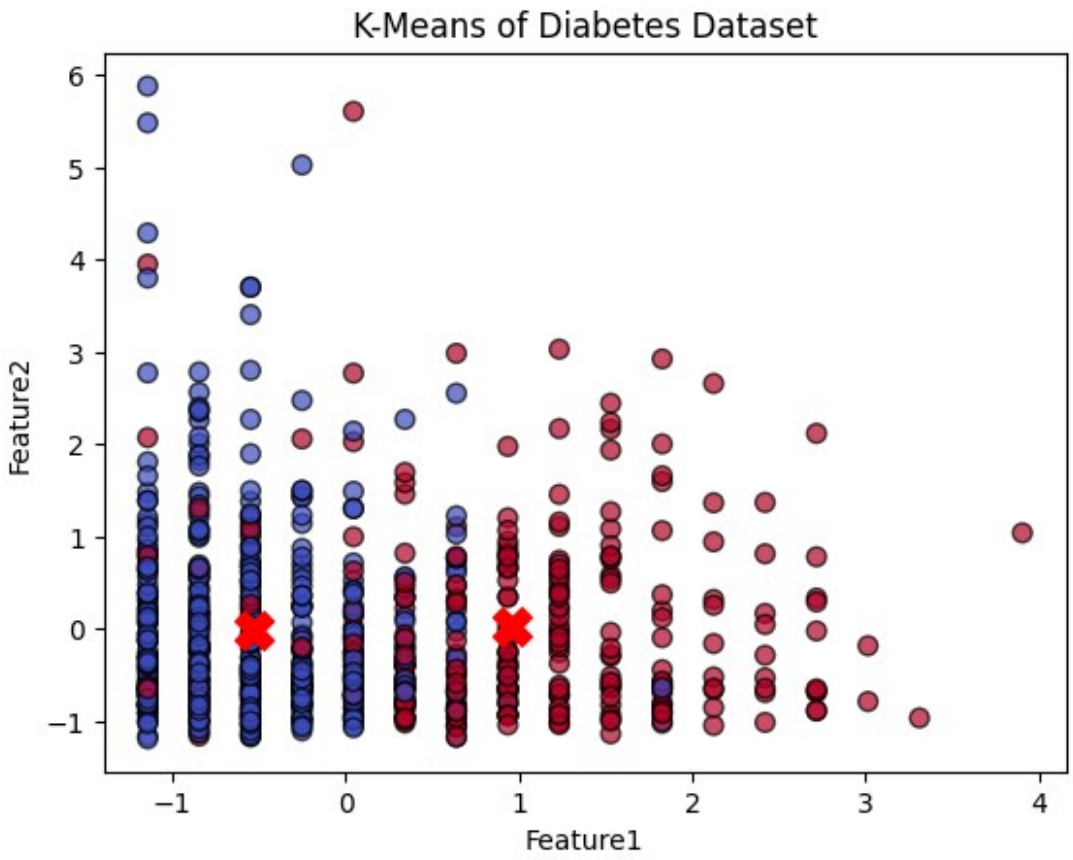
```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
```

```
KMeans(n_clusters=2, random_state=42)
```

```
labels=kmeans.labels_
centroids=kmeans.cluster_centers_
x_axis_feature = 0 # Change this to visualize different features on
the x-axis
y_axis_feature = 6  # Change this to visualize different features on
the y-axis
```

```
labels=kmeans.labels_
centroids=kmeans.cluster_centers_
x_axis_feature = 0 # Change this to visualize different features on
the x-axis
y_axis_feature = 6  # Change this to visualize different features on
the y-axis

plt.scatter(X_std[:, x_axis_feature], X_std[:, y_axis_feature],
c=labels, cmap='coolwarm', edgecolors='k', s=50, alpha=0.7)
plt.scatter(kmeans.cluster_centers_[:, x_axis_feature],
kmeans.cluster_centers_[:, y_axis_feature],
            c='red', marker='X', s=200, label='Centroids')
plt.xlabel('Feature1')
plt.ylabel('Feature2')
plt.title('K-Means of Diabetes Dataset')
plt.show()
```


K-Means of Diabetes Dataset
Feature2
Feature1
6
5
4
3
2
1
0
−1
−1
0
1
2
3
4