```
import numpy as np
import pandas as pd

dataset = pd.read_csv("diabetes.csv")
df = pd.DataFrame(dataset)
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
```

```
df.shape
```

```
(768, 9)
```

```
X = dataset.drop(columns='Outcome')
y = dataset['Outcome']

from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2

chi2_selector = SelectKBest(chi2, k=4)
X_kbest = chi2_selector.fit_transform(X, y)

selected_features = chi2_selector.get_support(indices=True)
selected_feature_names = X.columns[selected_features]
print("Top 4 features selected using Chi-Squared test:",
selected_feature_names)
```

```
Top 4 features selected using Chi-Squared test: Index(['Glucose',
'Insulin', 'BMI', 'Age'], dtype='object')
```

```
from sklearn.linear_model import LogisticRegression
from sklearn.feature_selection import RFE

model = LogisticRegression(max_iter=1000)
```

```
rfe = RFE(estimator=model, n_features_to_select=3)

rfe.fit(X, y)

rfe_selected_features = rfe.get_support(indices=True)
rfe_selected_feature_names = X.columns[rfe_selected_features]
print("Top 3 features selected using RFE:",
rfe_selected_feature_names)
```

```
Top 3 features selected using RFE: Index(['Pregnancies', 'BMI',
'DiabetesPedigreeFunction'], dtype='object')
```