

Diabetes Prediction

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score

diabetes_dataset = pd.read_csv('/content/diabetesdata.csv')
```

```
# printing the first 5 rows of the dataset
diabetes_dataset.head()
```

```
↗
```

	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

```
# number of rows and Columns in this dataset
diabetes_dataset.shape
```

```
(768, 9)
```

```
# separating features and Target
X = diabetes_dataset.drop(columns='Outcome', axis = 1)
Y = diabetes_dataset['Outcome']
```

```
# train test split
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, stratify=Y, random_state=2)
```

```
print(X.shape, X_train.shape, X_test.shape)
```

```
(768, 8) (614, 8) (154, 8)
```

Model Training

KNN Classifier in Sklearn

```
classifier = KNeighborsClassifier(p=1)
```

```
classifier.fit(X_train, Y_train)
```

```
KNeighborsClassifier(p=1)
```

Model Evaluation

```
y_pred = classifier.predict(X_test)
```

```
accuracy = accuracy_score(Y_test, y_pred)
```

```
print(accuracy*100)
```

```
77.92207792207793
```

