

```
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
```

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

```
data.head()
```



```
data.describe()
#for statistical measures
```

```
X = data.drop(columns="Outcome",axis=1)
Y = data["Outcome"]
```

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

```
#classifier object
clf = DecisionTreeClassifier(criterion='entropy',max_depth=3)
#train clf
clf =clf.fit(X_train,Y_train)
#predict
y_pred =clf.predict(X_test)
```

```
#Accuracy
print(metrics.accuracy_score(Y_test, y_pred))
```

```
0.7402597402597403
```

