import pandas as pd  
from sklearn.tree import DecisionTreeClassifier  
from sklearn.model\_selection import train\_test\_split  
from sklearn import metrics

col\_names = ["Pregnancies", "Glucose", "BloodPressure", "SkinThickness", "Insulin", "BMI", "DiabetesPedigreeFunction", "Age", "Outcome"]

pima = pd.read\_csv("./data/diabetes.csv")  
pima.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

feature\_cols = ["Pregnancies", "Glucose", "BloodPressure", "SkinThickness", "Insulin", "BMI", "DiabetesPedigreeFunction", "Age"]  
  
x = pima[feature\_cols]  
y = pima.Outcome

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size=0.3, random\_state=0)

clf = DecisionTreeClassifier()  
  
clf = clf.fit(x\_train, y\_train)  
  
y\_pred = clf.predict(x\_test)

print("Accuracy: {:.2f}%".format(metrics.accuracy\_score(y\_test, y\_pred) \* 100))

Accuracy: 73.16%