**จากโจทย์ ให้train model ที่สามารถแยกประเภท (classify) การมีบุตรยากของเพศชายว่าเป็น normal/weak โดยใช้ Decision Tree**

**Import Scikit-Learn: DecisionTreeClassifier, LabelEncoder**

**Import: Graphviz**

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.tree import DecisionTreeClassifier, export\_text, export\_graphviz

from sklearn.metrics import accuracy\_score

from sklearn.preprocessing import LabelEncoder

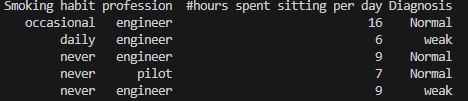
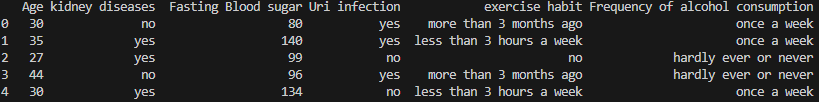
import graphviz

**Pre-Process**

**Load Dataset แล้วลอง Display เพื่อตรวจสอบ**

df = pd.read\_csv('fertility2.csv')

print(df.head())



(139, 10)

**Age = อายุ**

**kidney diseases = ภาวะโรคไต**

**Fasting Blood sugar = ระดับนำตาลในเลือด**

**Uri infection = การติดเชื้อทางเดินปัสสาวะ**

**exercise habit = รวมถึงพฤติกรรมการใช้ชีวิต**

**Frequency of alcohol consumption = การดื่มสุรา**

**Smoking habit profession = การสูบบุหรี่**

**#hours spent sitting per day = จำนวนชั่วโมงที่นั่งอยู่กับที่**

**Profession = อาชีพ**

**Diagnosis = ว่าปกติหรืออ่อนแอ**

**Process**

**แปลง Raw Data จาก String to Int และจัดการ feature บางตัวให้เป็นค่าที่นับได้เพื่อใช้ในการเทรนและลอง Display เพื่อจรวจสอบ**

label\_encoder = LabelEncoder()

for col in ['Age', 'kidney diseases', 'Fasting Blood sugar', 'Uri infection', 'Frequency of alcohol consumption', 'profession', '#hours spent sitting per day']:

    df[col] = label\_encoder.fit\_transform(df[col])

df['exercise habit'] = df['exercise habit'].map({

    'more than 3 months ago': 0,

    'less than 3 hours a week': 1,

    'no': 2

})

df = pd.get\_dummies(df, columns=['Frequency of alcohol consumption'], prefix='alcohol', drop\_first=True)

df = pd.get\_dummies(df, columns=['Smoking habit'], prefix='smoking', drop\_first=True)

print(df.head())

**ให้ X = feature, Y = class**

X = df.drop('Diagnosis', axis=1)

y = df['Diagnosis']

**แบ่ง Data ออกเป็นส่วนสำหรับการ trainและ test โดยที่ให้สัดส่วนของ test = 0.2**

# Split the data into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

**เทรนโมเดลและ Prediction**

clf = DecisionTreeClassifier()

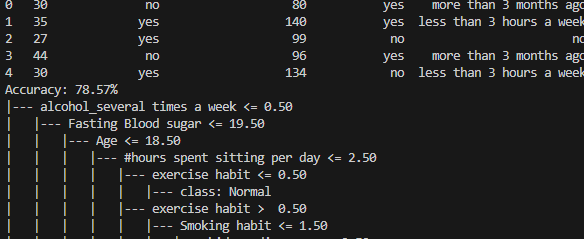
clf.fit(X\_train, y\_train)

y\_pred = clf.predict(X\_test)

**หาความแม่นยำและลอง Display ออกมา**

accuracy = accuracy\_score(y\_test, y\_pred)

print(f'Accuracy: {accuracy \* 100:.2f}%')



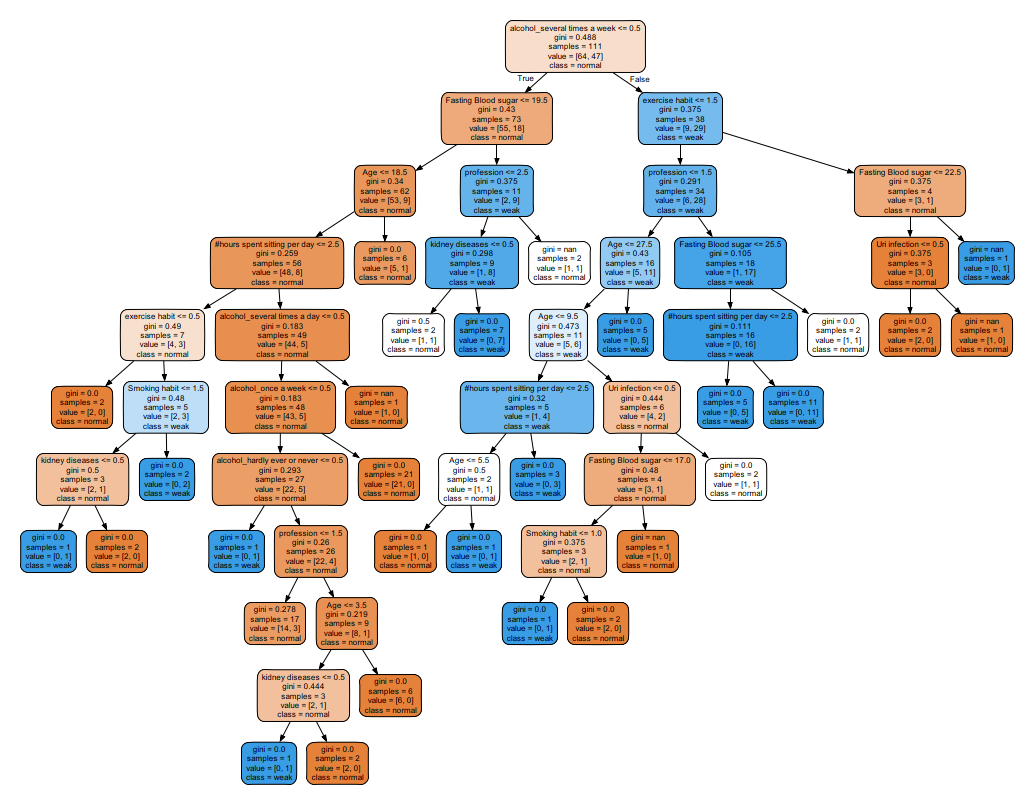
**Display tree ออกมา จะได้เป็นไฟล์ .png ดังรูป**

dot\_data = export\_graphviz(clf, out\_file=None, feature\_names=list(X.columns), class\_names=['normal', 'weak'], filled=True, rounded=True)

graph = graphviz.Source(dot\_data)

graph.render("fertility\_tree2", format="png", cleanup=True)

graph.view("fertility\_tree2")



<https://github.com/TanapatButsai/AIAssignment01_6410451059.git>