

ผลการทดสอบและประสิทธิภาพโมเดลตรวจหาภาวะมีบุตรยาก

ตรวจสอบ datasets

```
RangeIndex: 139 entries, 0 to 138
Data columns (total 10 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Age                                    139 non-null    int64
1   kidney diseases                       139 non-null    object
2   Fasting Blood sugar                  139 non-null    int64
3   Uri infection                        139 non-null    object
4   exercise habit                       139 non-null    object
5   Frequency of alcohol consumption    139 non-null    object
6   Smoking habit                       139 non-null    object
7   profession                           139 non-null    object
8   #hours spent sitting per day         139 non-null    int64
9   Diagnosis                            139 non-null    object
dtypes: int64(3), object(7)
```

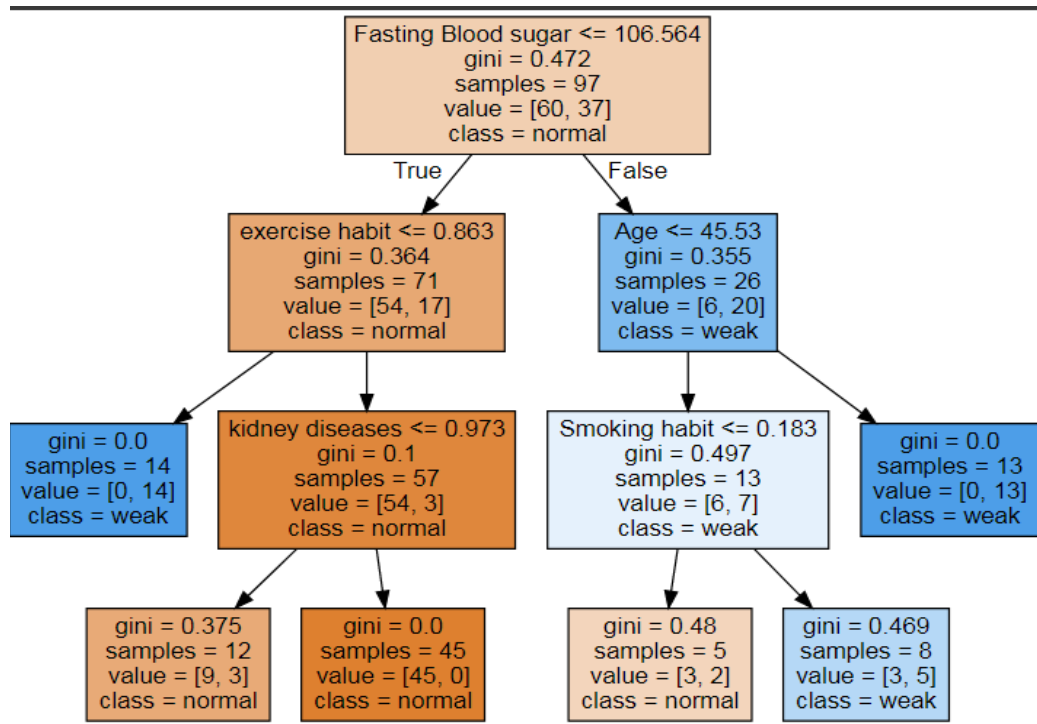
- ข้อมูลทั้งหมดมี 139 แถว, 10 หลัก
- ไม่มีข้อมูลหาย หรือเป็นค่าว่าง

ขั้นตอนการฝึก

1. แบ่ง Feature และ Target
 - a. **Target** = ผลการวินิจฉัย
 - b. **Feature** = อายุ, โรคไต, น้ำตาลในเลือด, การติดเชื้อทางเดินปัสสาวะ, การออกกำลังกาย, ความถี่การดื่มแอลกอฮอล์, ความถี่สูบบุหรี่, อาชีพ, จำนวนชั่วโมงนั่งอยู่กับที่
2. แปลงข้อความเป็นตัวเลข (Label encoding)
3. แบ่ง train set 70%, test set 30%
4. ฝึกฝนโมเดล
 - parameter ที่ใช้ในโมเดล
 - criterion = 'gini'
 - max_depth = 5
 - ccp_alpha = 0.001
 - max_depth = 5
 - min_samples_split = 5
 - min_samples_leaf = 5
 - max_features = 'sqrt'
 - splitter = 'random'

ผลลัพธ์และประสิทธิภาพ

- แผนภาพต้นไม้



- ลำดับความสำคัญของ Feature แต่ละตัว

ลำดับความสำคัญของ feature	
exercise habit	: 0.5744051445
Fasting Blood sugar	: 0.3041657073
Age	: 0.0788433598
kidney diseases	: 0.0337159104
Smoking habit	: 0.0088698780
Uri infection	: 0.0000000000
Frequency of alcohol consumption	: 0.0000000000
profession	: 0.0000000000
#hours spent sitting per day	: 0.0000000000

Feature ที่ถูกใช้ในการตัดสินใจจากมากไปน้อย คือ การออกกำลังกาย, น้ำตาลในเลือด, อายุ, โรคไต, การสูบบุหรี่ (ใช้ทั้งหมด 5 features)

- Training set score และ Test set score

```
Training set score: 0.9175257731958762
Test set score: 0.9761904761904762
```

- confusion matrix

```
[[26  0]
 [ 1 15]]
```

	Positive	Negative
Positive	26 (ทายว่าปกติแล้วถูก)	0 (ทายว่าปกติแล้วผิด)
Negative	1 (ทายว่าป่วยแล้วผิด)	15 (ทายว่าป่วยแล้วถูก)