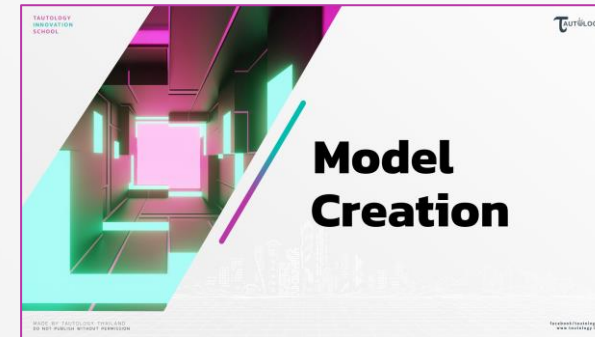


CLASSIFICATION TREE

BY TAUTOLOGY

Classification Tree



Introduction

Introduction

What is
Classification Tree?

Data for
Classification Tree

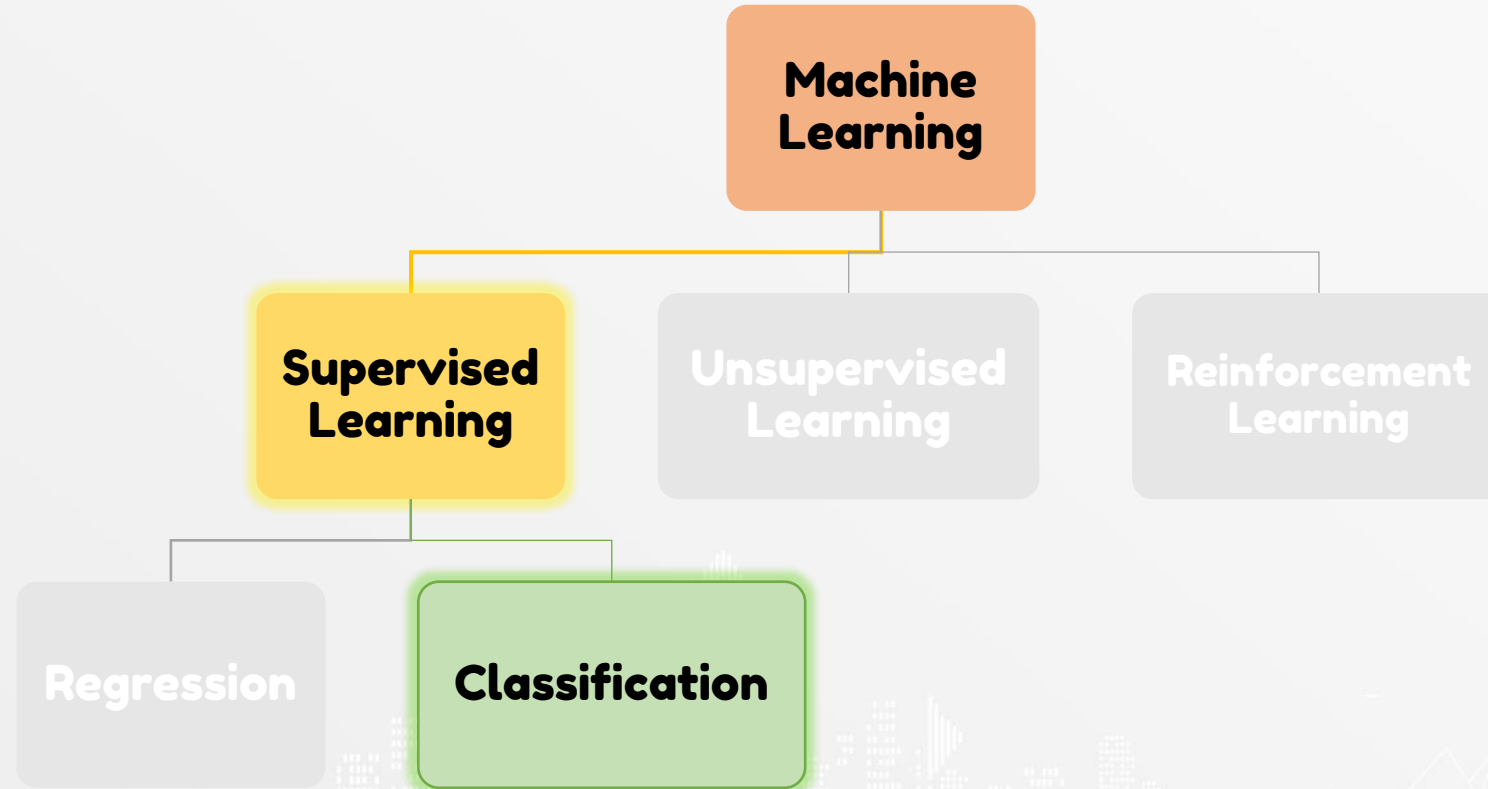
Pros & Cons

Real World
Application

What is Classification Tree?

Classification Tree เป็นหนึ่งใน algorithm ประเภท supervised learning ที่ใช้สำหรับแก้ปัญหา classification โดยมีหลักการทำงาน คือ การสร้างชุดของกฎเพื่อจำแนกประเภทของข้อมูล

What is Classification Tree?



What is Classification Tree?

sex	BMI	target
0	26	diabetes
1	26	normal
1	28	diabetes
1	30	diabetes
0	28	normal
0	30	normal

sex \leq 0.5

BMI \leq 27

predict = 'diabetes'

BMI > 27

predict = 'normal'

sex > 0.5

BMI \leq 27

predict = 'normal'

BMI > 27

predict = 'diabetes'

What is Classification Tree?

sex ≤ 0.5

BMI ≤ 27

predict = 'diabetes'

=> If sex ≤ 0.5 and BMI ≤ 27 , then 'diabetes'

BMI > 27

predict = 'normal'

=> If sex ≤ 0.5 and BMI > 27 , then 'normal'

sex > 0.5

BMI ≤ 27

predict = 'normal'

=> If sex > 0.5 and BMI ≤ 27 , then 'normal'

BMI > 27

predict = 'diabetes'

=> If sex > 0.5 and BMI > 27 , then 'diabetes'

Introduction

**What is
Classification Tree?**



**Data for
Classification Tree**



Pros & Cons

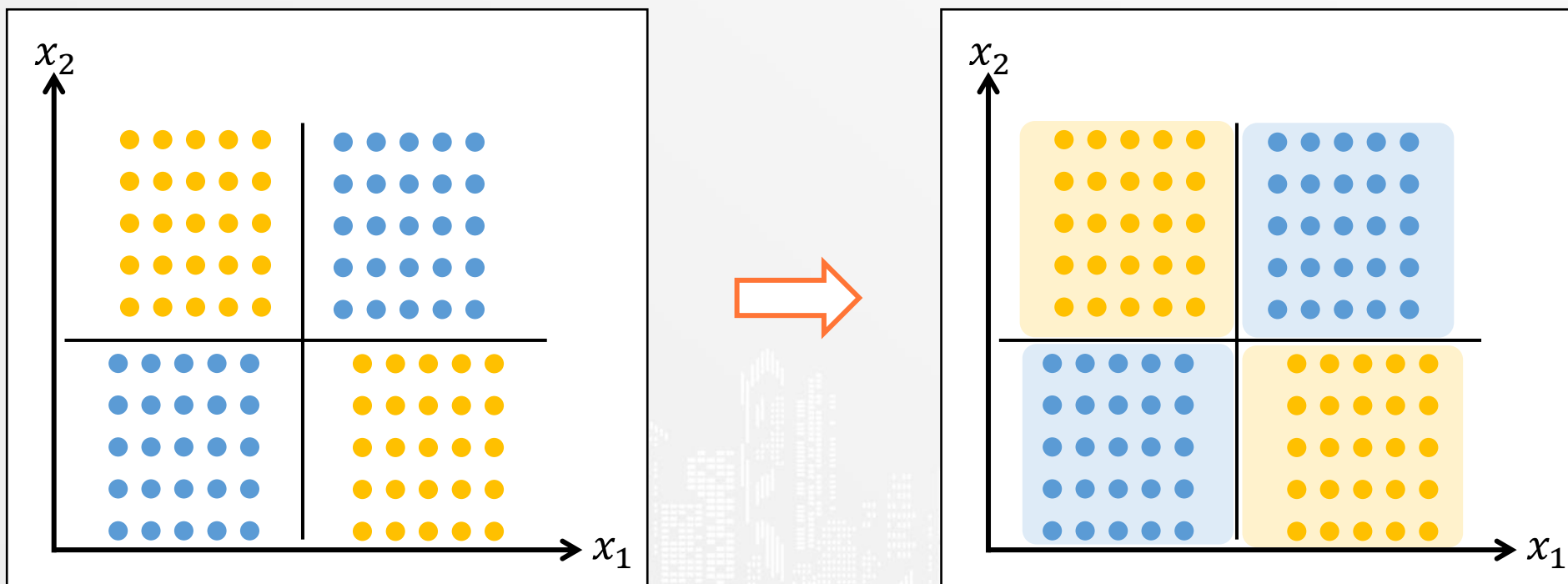


**Real World
Application**



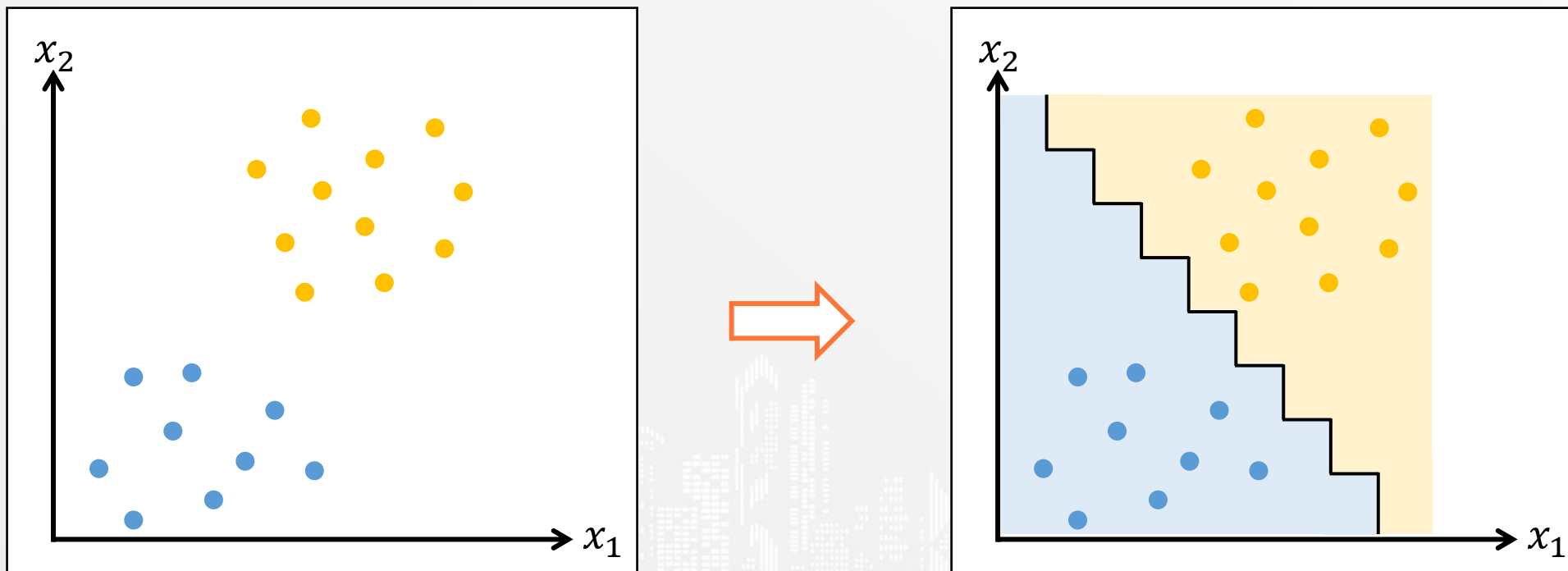
Data for Classification Tree

ตัวอย่างของข้อมูลที่เหมาะสมกับ Classification Tree



Data for Classification Tree

ตัวอย่างของข้อมูลที่ไม่เหมาะกับ Classification Tree



Introduction

**What is
Classification Tree?**



**Data for
Classification Tree**



Pros & Cons



**Real World
Application**



Pros & Cons

ข้อดี

- หลักการของ algorithm เรียบง่าย
- สามารถตีความผลลัพธ์ได้ของ model ได้ง่าย (model อยู่ในรูปของกฎ)

ข้อเสีย

- ง่ายต่อการเกิด overfitting
- การเปลี่ยนแปลงข้อมูลเพียงเล็กน้อยใน training อาจส่งผลให้ model เปลี่ยนแปลงอย่างมาก

ข้อจำกัด

- decision boundary ที่ได้จะขนานกับแกนเสมอ

Introduction

**What is
Classification Tree?**



**Data for
Classification Tree**



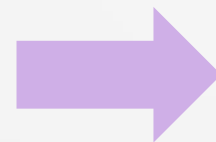
Pros & Cons



**Real World
Application**



Real World Application



การจำแนกผู้ป่วยโรคหัวใจ

โดยพิจารณาจาก อายุ เพศ ความดัน
โลหิต คอเลสเตอรอล ประเภทการ
เจ็บหน้าอก เป็นต้น

อ้างอิง : [2023, Ozcan & Peker] A classification and regression tree
algorithm for heart disease modeling and prediction

Real World Application



การระบุ spam e-mail

โดยพิจารณาจาก e-mail ที่ถูกส่ง
เข้าของ University Utara
Malaysia's Computer Center ใน
1 สัปดาห์

อ้างอิง : [2022, Abdulrahman & Salim] Using Decision Tree
Algorithms in Detecting Spam Emails Written in Malay: A
Comparison Study

Real World Application



การจำแนกประเภทลูกค้าที่สามารถกู้เงินได้

โดยพิจารณาจาก เพศ อายุ ประเภท
ของบริษัทที่ทำงาน อาชีพ ระดับ
การศึกษา การแต่งงาน รายได้
จำนวนปีที่จะกู้

อ้างอิง : [2004, Xiu Li et al.] Applications of Classification Trees to
Consumer Credit Scoring Methods in Commercial Banks

Introduction

**What is
Classification Tree?**



**Data for
Classification Tree**



Pros & Cons



**Real World
Application**

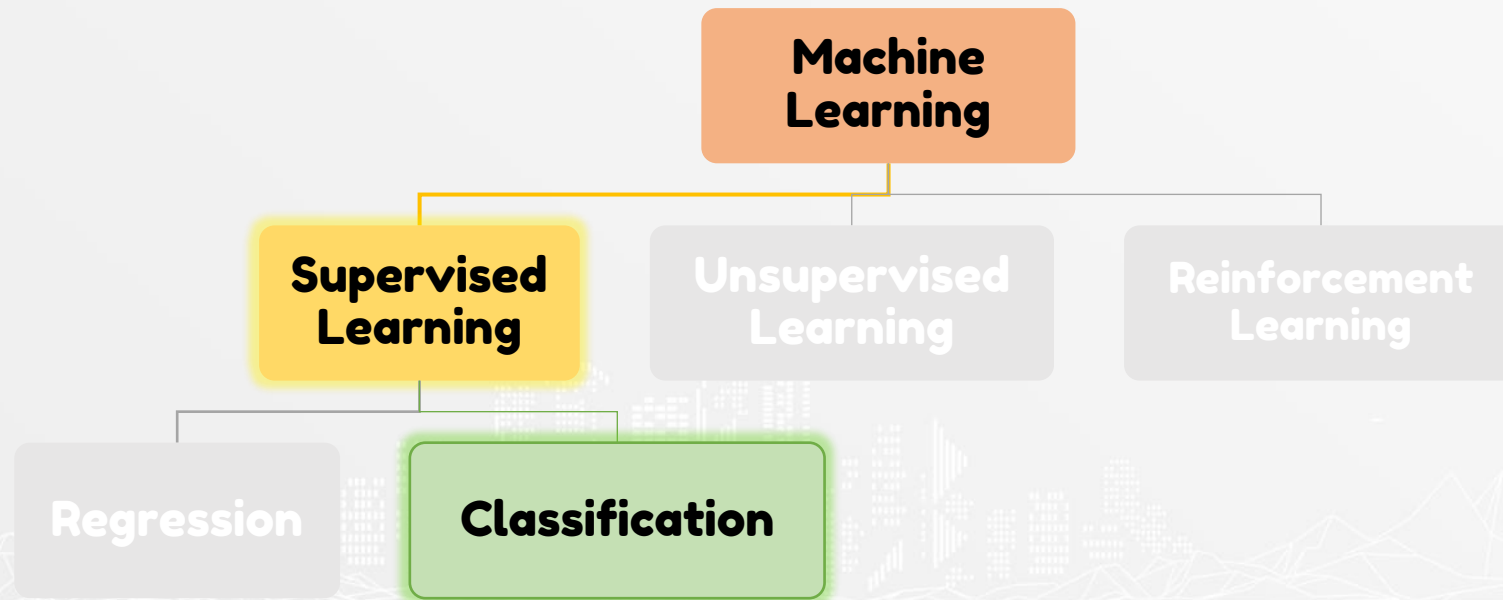


Classification Tree



Classification Tree

Classification Tree เป็นหนึ่งใน algorithm ประเภท **supervised learning**



Concept of Supervised Learning

Data \Rightarrow **Model** \Rightarrow **Prediction**

Model Creation

Model Creation

Assumption

Real Face of the
Model

How to Create Model
(Math)

How to Create Model
(Code)

Further Reading

Assumption

- No Missing Features

Model Creation

Assumption



Real Face of the
Model



How to Create Model
(Math)



How to Create Model
(Code)



Further Reading



Real Face of the Model

Classification Tree คือ ชุดของกฎเพื่อจำแนกประเภทของข้อมูล

sex \leq 0.5

BMI \leq 27

predict = 'diabetes'

=> If sex \leq 0.5 and BMI \leq 27, then 'diabetes'

BMI > 27

predict = 'normal'

=> If sex \leq 0.5 and BMI > 27, then 'normal'

sex > 0.5

BMI \leq 27

predict = 'normal'

=> If sex > 0.5 and BMI \leq 27, then 'normal'

BMI > 27

predict = 'diabetes'

=> If sex > 0.5 and BMI > 27, then 'diabetes'

Model Creation

Assumption



**Real Face of the
Model**



**How to Create Model
(Math)**



**How to Create Model
(Code)**



Further Reading



How to Create Model (Math)

- ☐ Step 1 : พิจารณา unique values ของ feature ทุกตัวใน dataset
- ☐ Step 2 : ตั้งคำถามจาก unique values
- ☐ Step 3 : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ classification tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น ภายใต้เงื่อนไขที่กำหนด

How to Create Model (Math)

ตัวอย่างการคำนวณ Classification Tree

sex	BMI	target
0	26	diabetes
1	26	normal
1	28	diabetes
1	30	diabetes
0	28	normal
0	30	normal

ตารางแสดงข้อมูลผู้ป่วยที่เป็นโรคเบาหวาน

How to Create Model (Math)

☑ **Step 1** : พิจารณา unique values ของ feature ทุกตัวใน dataset

sex	BMI	target
0	26	diabetes
1	26	normal
1	28	diabetes
1	30	diabetes
0	28	normal
0	30	normal

→ `unique_values(sex) = {0, 1}`

→ `unique_values(BMI) = {26, 28, 30}`

How to Create Model (Math)

☑ **Step 2** : ตั้งคำถามจาก unique values

$\text{unique_values}(\text{sex}) = \{0, 1\}$

$\text{unique_values}(\text{BMI}) = \{26, 28, 30\}$



Question1 : $\text{sex} \leq 0.5$?

Question2 : $\text{BMI} \leq 27$?

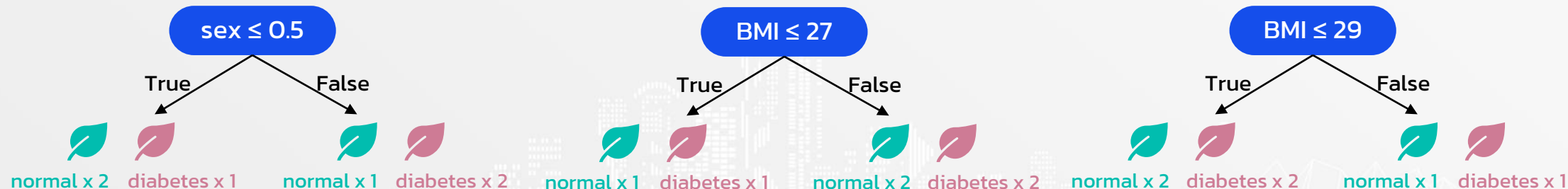
Question3 : $\text{BMI} \leq 29$?

How to Create Model (Math)

✓ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ classification tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น ภายใต้เงื่อนไขที่กำหนด

Root Node

normal x 3 diabetes x 3

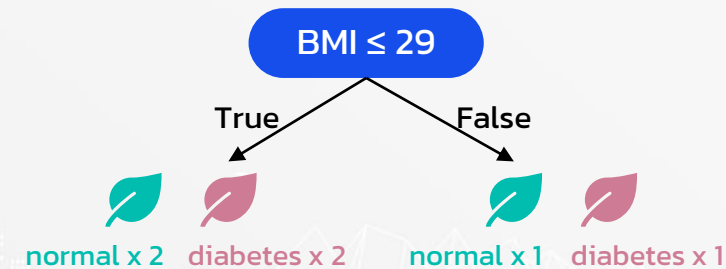
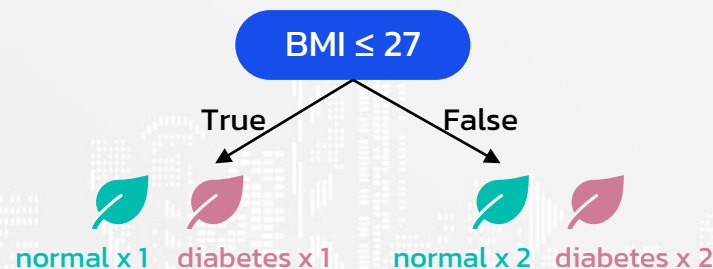
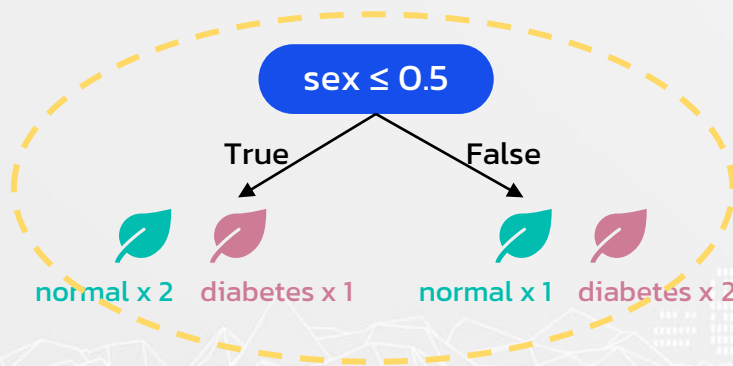


How to Create Model (Math)

✓ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ classification tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น ภายใต้เงื่อนไขที่กำหนด

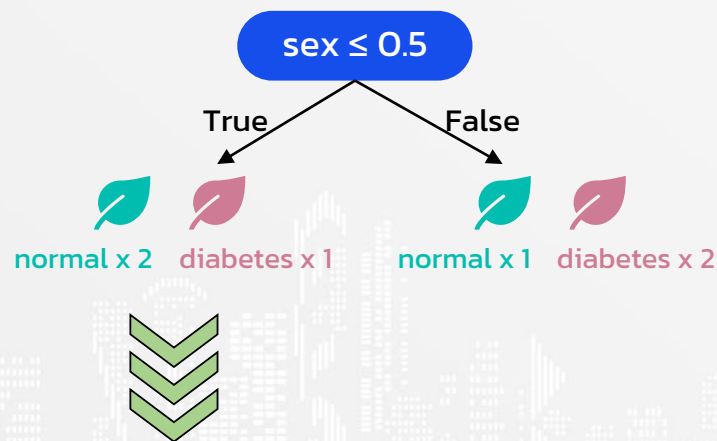
Root Node

normal x 3 diabetes x 3



How to Create Model (Math)

☑ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ classification tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น ภายใต้เงื่อนไขที่กำหนด



How to Create Model (Math)

☑ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ classification tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น ภายใต้เงื่อนไขที่กำหนด

Secondary Node (left)

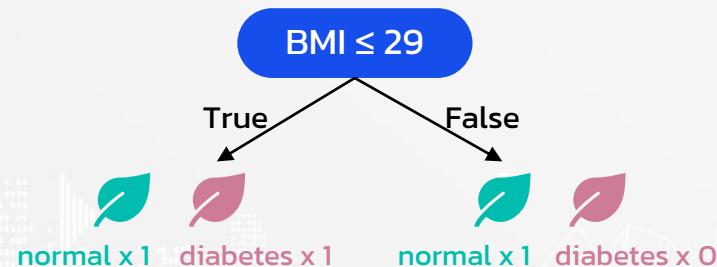
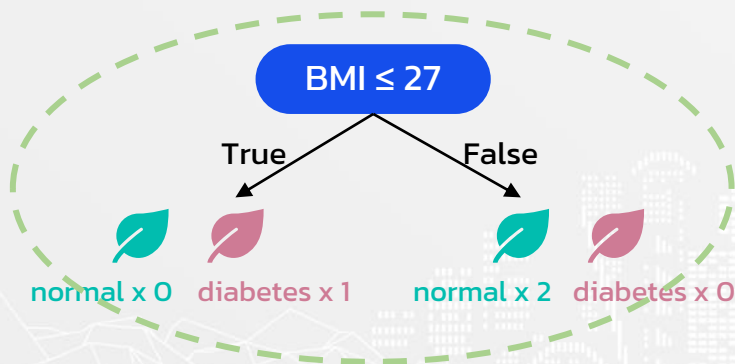


How to Create Model (Math)

✓ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ classification tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น ภายใต้เงื่อนไขที่กำหนด

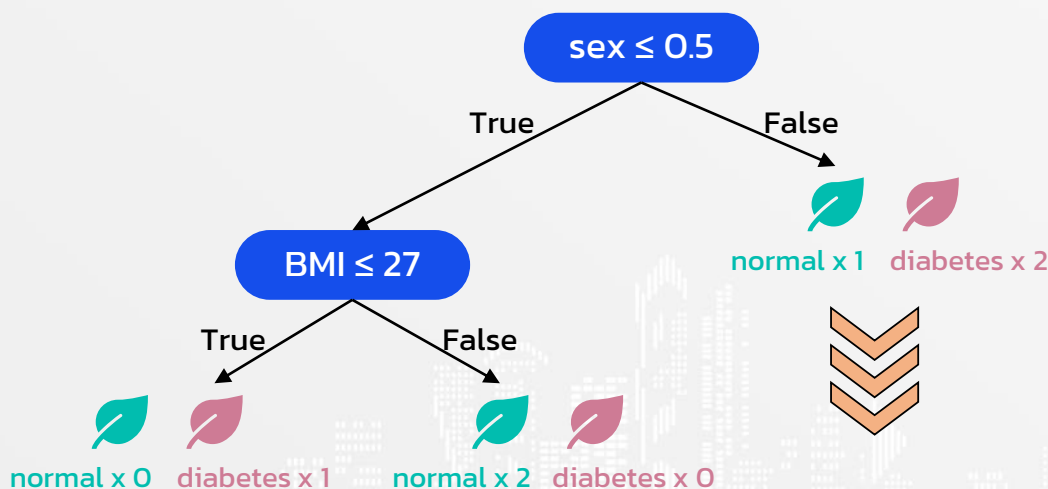
Secondary Node (left)

normal x 2 diabetes x 1



How to Create Model (Math)

☑ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ classification tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น ภายใต้เงื่อนไขที่กำหนด



How to Create Model (Math)

✓ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ classification tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น ภายใต้เงื่อนไขที่กำหนด

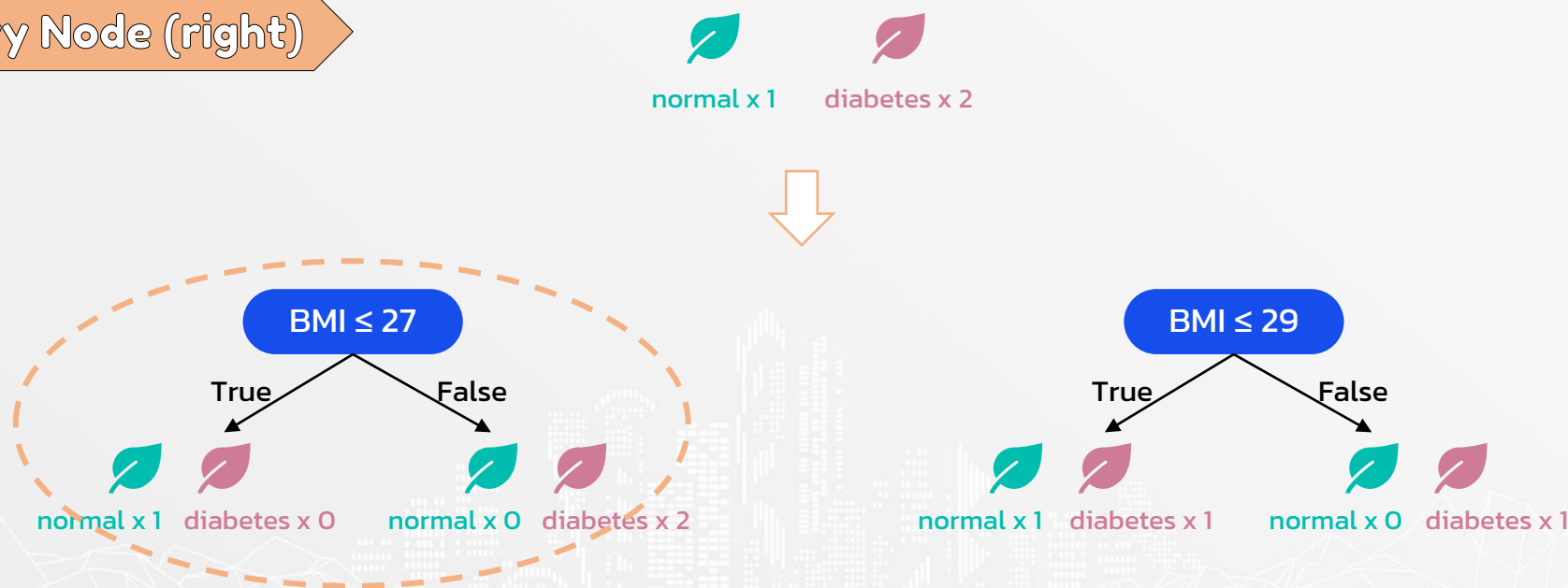
Secondary Node (right)



How to Create Model (Math)

✓ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ classification tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น ภายใต้เงื่อนไขที่กำหนด

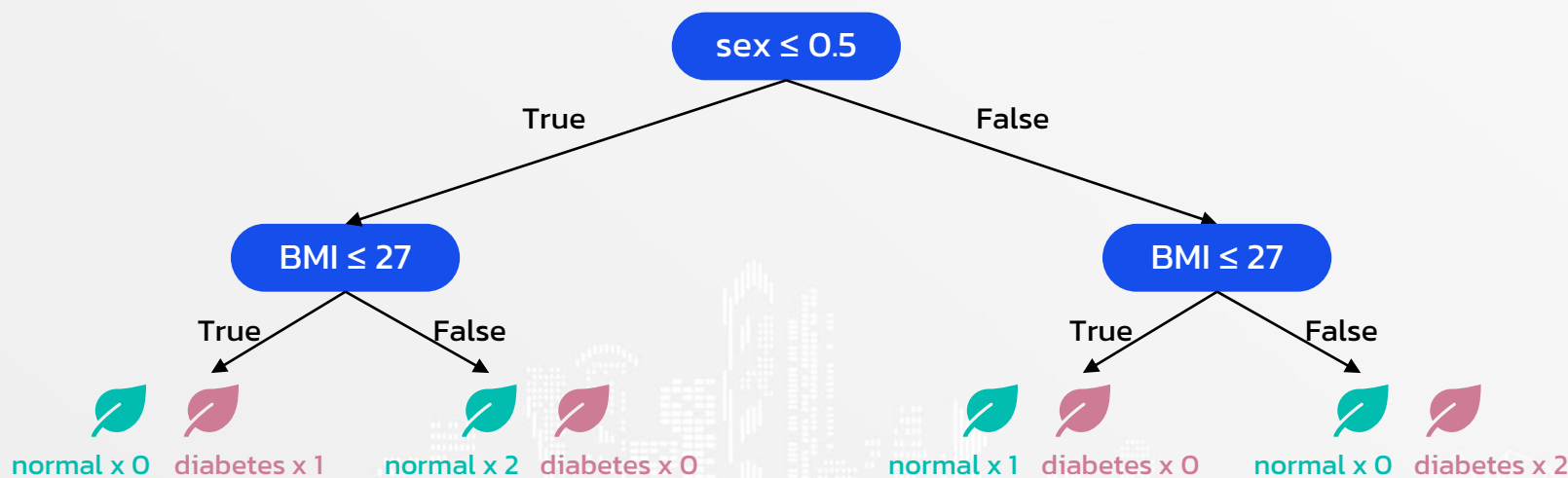
Secondary Node (right)



How to Create Model (Math)

✓ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ classification tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น ภายใต้เงื่อนไขที่กำหนด

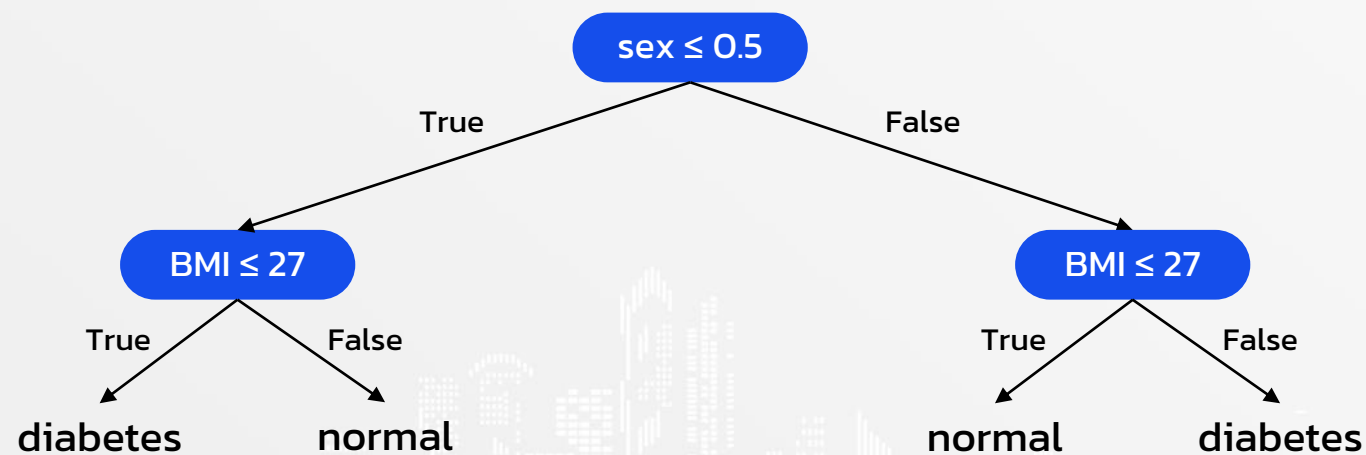
Full Tree



How to Create Model (Math)

✓ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ classification tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น ภายใต้เงื่อนไขที่กำหนด

Full Tree



How to Create Model (Math)

☑ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ classification tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น ภายใต้เงื่อนไขที่กำหนด

If sex ≤ 50 and BMI ≤ 27 , then 'diabetes'

If sex ≤ 0.5 and BMI > 27 , then 'normal'

If sex > 0.5 and BMI ≤ 27 , then 'normal'

If sex > 0.5 and BMI > 27 , then 'diabetes'

Model Creation

Assumption



**Real Face of the
Model**



**How to Create Model
(Math)**



**How to Create Model
(Code)**



Further Reading



How to Create Model (Code)

ตัวอย่าง Code สำหรับ Classification Tree

sex	BMI	target
0	26	diabetes
1	26	normal
1	28	diabetes
1	30	diabetes
0	28	normal
0	30	normal

ตารางแสดงข้อมูลผู้ป่วยที่เป็นโรคเบาหวาน

How to Create Model (Code)

- Code สำหรับสร้าง model จากข้อมูลของเราโดยที่

$$X = \begin{bmatrix} 0 & 26 \\ 1 & 26 \\ 1 & 28 \\ 1 & 30 \\ 0 & 28 \\ 0 & 30 \end{bmatrix}, \quad y = \begin{bmatrix} \text{diabetes} \\ \text{normal} \\ \text{diabetes} \\ \text{diabetes} \\ \text{normal} \\ \text{normal} \end{bmatrix}$$

```
1 clf = DecisionTreeClassifier()  
2 clf.fit(X, y)
```

DecisionTreeClassifier()

How to Create Model (Code)

```
1 r = export_text(clf, feature_names=list(X.columns))
```

```
1 print(r)
```

```
| --- sex <= 0.50  
|   | --- BMI <= 27.00  
|   |   | --- class: diabetes  
|   | --- BMI > 27.00  
|   |   | --- class: normal  
| --- sex > 0.50  
|   | --- BMI <= 27.00  
|   |   | --- class: normal  
|   | --- BMI > 27.00  
|   |   | --- class: diabetes
```


How to Create Model (Code)



Code for this section



Open File
Model Creation.ipynb

Model Creation

Assumption



**Real Face of the
Model**



**How to Create Model
(Math)**



**How to Create Model
(Code)**



Further Reading



Further Reading

- Gini Impurity
- Entropy
- Information Gain
- ID3 Algorithm

Model Creation

Assumption



**Real Face of the
Model**



**How to Create Model
(Math)**



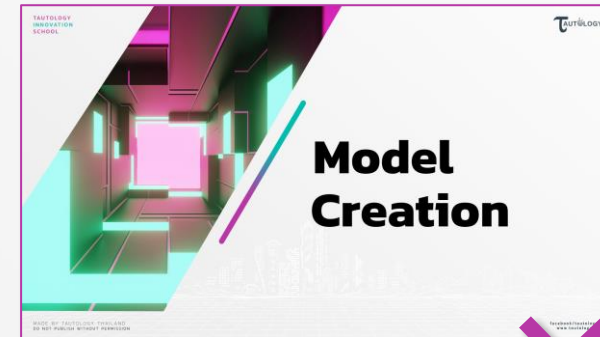
**How to Create Model
(Code)**



Further Reading



Classification Tree



Prediction

Prediction

Classification Tree คือ ชุดของกฎเพื่อจำแนกประเภทของข้อมูล

sex \leq 0.5

BMI \leq 27

predict = 'diabetes'

=> If sex \leq 0.5 and BMI \leq 27, then 'diabetes'

BMI > 27

predict = 'normal'

=> If sex \leq 0.5 and BMI > 27, then 'normal'

sex > 0.5

BMI \leq 27

predict = 'normal'

=> If sex > 0.5 and BMI \leq 27, then 'normal'

BMI > 27

predict = 'diabetes'

=> If sex > 0.5 and BMI > 27, then 'diabetes'

Prediction

If sex ≤ 0.5 and BMI ≤ 27 , then 'diabetes'

If sex ≤ 0.5 and BMI > 27 , then 'normal'

If sex > 0.5 and BMI ≤ 27 , then 'normal'

If sex > 0.5 and BMI > 27 , then 'diabetes'

Prediction

1-Sample

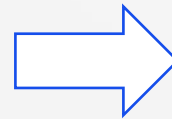
Multi-Sample

Code

1-Sample

ตัวอย่างการคำนวณ \hat{y}

sex	BMI
0	29



\hat{y}
?

1-Sample

sex	BMI		\hat{y}
0	29	→	normal

If sex \leq 0.5 and BMI \leq 27, then 'diabetes'

If sex \leq 0.5 and BMI $>$ 27, then 'normal'

If sex $>$ 0.5 and BMI \leq 27, then 'normal'

If sex $>$ 0.5 and BMI $>$ 27, then 'diabetes'

Prediction

1-Sample



Multi-Sample



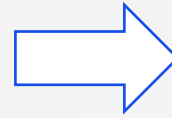
Code



Multi-Sample

ตัวอย่างการคำนวณ \hat{y}

sex	BMI
0	29
0	26
1	30
1	28



\hat{y}
?
?
?
?

Multi-Sample

If sex ≤ 0.5 and BMI ≤ 27 , then 'diabetes'

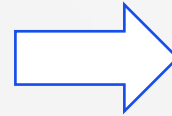
If sex ≤ 0.5 and BMI > 27 , then 'normal'

If sex > 0.5 and BMI ≤ 27 , then 'normal'

If sex > 0.5 and BMI > 27 , then 'diabetes'

Multi-Sample

sex	BMI
0	29
0	26
1	30
1	28



\hat{y}
normal
diabetes
diabetes
diabetes

Prediction

1-Sample



Multi-Sample



Code



Code

ตัวอย่าง code สำหรับการคำนวณ \hat{y}

sex	BMI
0	29
0	26
1	30
1	28



\hat{y}
?
?
?
?

Code

- Code สำหรับสร้าง model จากข้อมูลของเราโดยที่

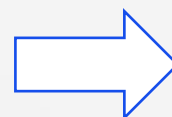
$$X = \begin{bmatrix} 0 & 29 \\ 0 & 26 \\ 1 & 30 \\ 1 & 28 \end{bmatrix}$$

```
1 clf.predict(X)  
  
array(['normal', 'diabetes', 'diabetes', 'diabetes'], dtype=object)
```

Code

ดังนั้น เราจะได้ \hat{y} สำหรับข้อมูลชุดนี้คือ

sex	BMI
0	29
0	26
1	30
1	28



\hat{y}
normal
diabetes
diabetes
diabetes

Code



Code for this section



Open File
Model Creation.ipynb

Prediction

1-Sample



Multi-Sample



Code



Classification Tree



Workshop

AI in Healthcare

- Abstract
- Why this project important?
- Who this project for?
- Hepatitis C Dataset
- What we learn from this project?

Abstract

สร้าง model เพื่อวินิจฉัยผู้ป่วยโรคไวรัสตับอักเสบ C (Hepatitis C) โดย feature ที่นำมาใช้ คือ ข้อมูลทั่วไปของผู้ป่วย และ ผลตรวจการทำงานของตับ



Why this project important?



- สามารถสร้างระบบสำหรับตรวจโรคไวรัสตับอักเสบบี c ที่ทำงานได้ตลอด 24 ชั่วโมง
- สามารถนำไปต่อยอดกับการวินิจฉัยโรคอื่น ๆ
- สามารถใช้เป็นพื้นฐานสำหรับการแพทย์ทางไกล

Who this project is for?

- ✦ ผู้บริหารโรงพยาบาล
- ✦ บุคลากรทางการแพทย์
- ✦ นักวิเคราะห์ข้อมูล



Hepatitis C Dataset



<https://www.kaggle.com/datasets/fedesoriano/hepatitis-c-dataset>

Hepatitis C Dataset

Feature

- Age : อายุ (ปี)
- Sex : เพศ (m = ชาย, f = หญิง)
- ALB : ปริมาณโปรตีน Albumin ในตับ
- ALP : ปริมาณเอนไซม์ Alkaline phosphatase ในตับ
- ALT : ปริมาณเอนไซม์ Alanine transaminase ในตับ
- AST : ปริมาณเอนไซม์ Aspartate transaminase ในตับ
- BIL : สาร Bilirubin ในตับ

Hepatitis C Dataset

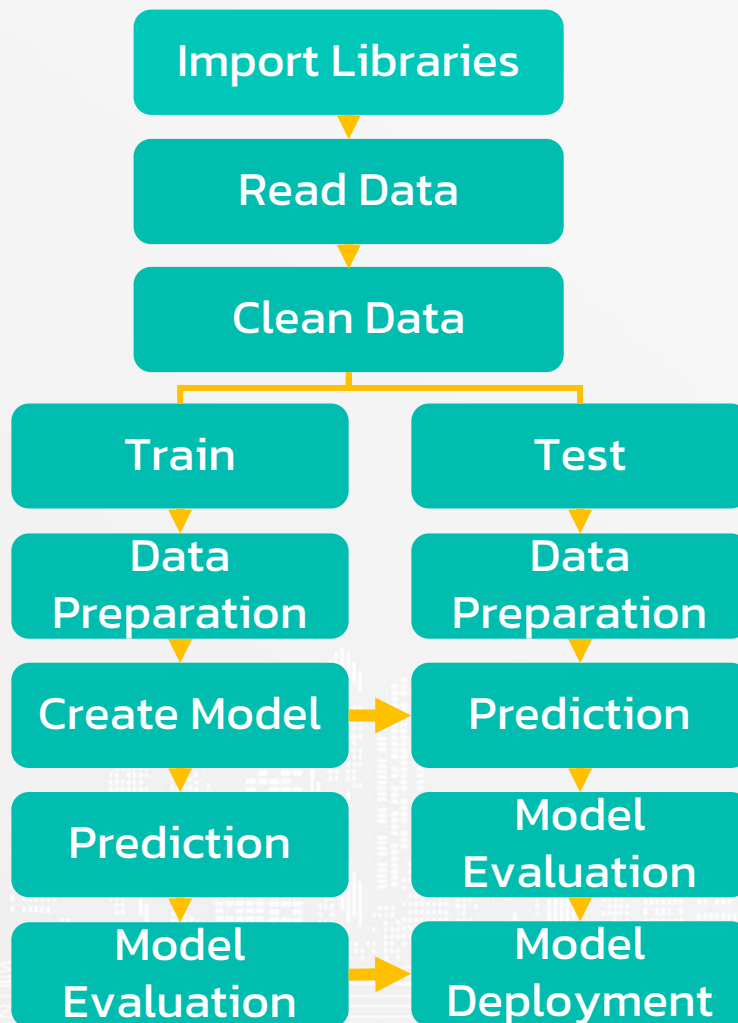
Feature

- CHE : ปริมาณเอนไซม์ Cholinesterase ใน serum
- CHOL : ปริมาณ Cholesterol ในตับ
- CREA : ปริมาณ Creatinine ในตับ
- GGT : ปริมาณโปรตีน Gamma glutamic transpeptidase ในตับ
- PROT : ปริมาณโปรตีน Prothrombin ในตับ

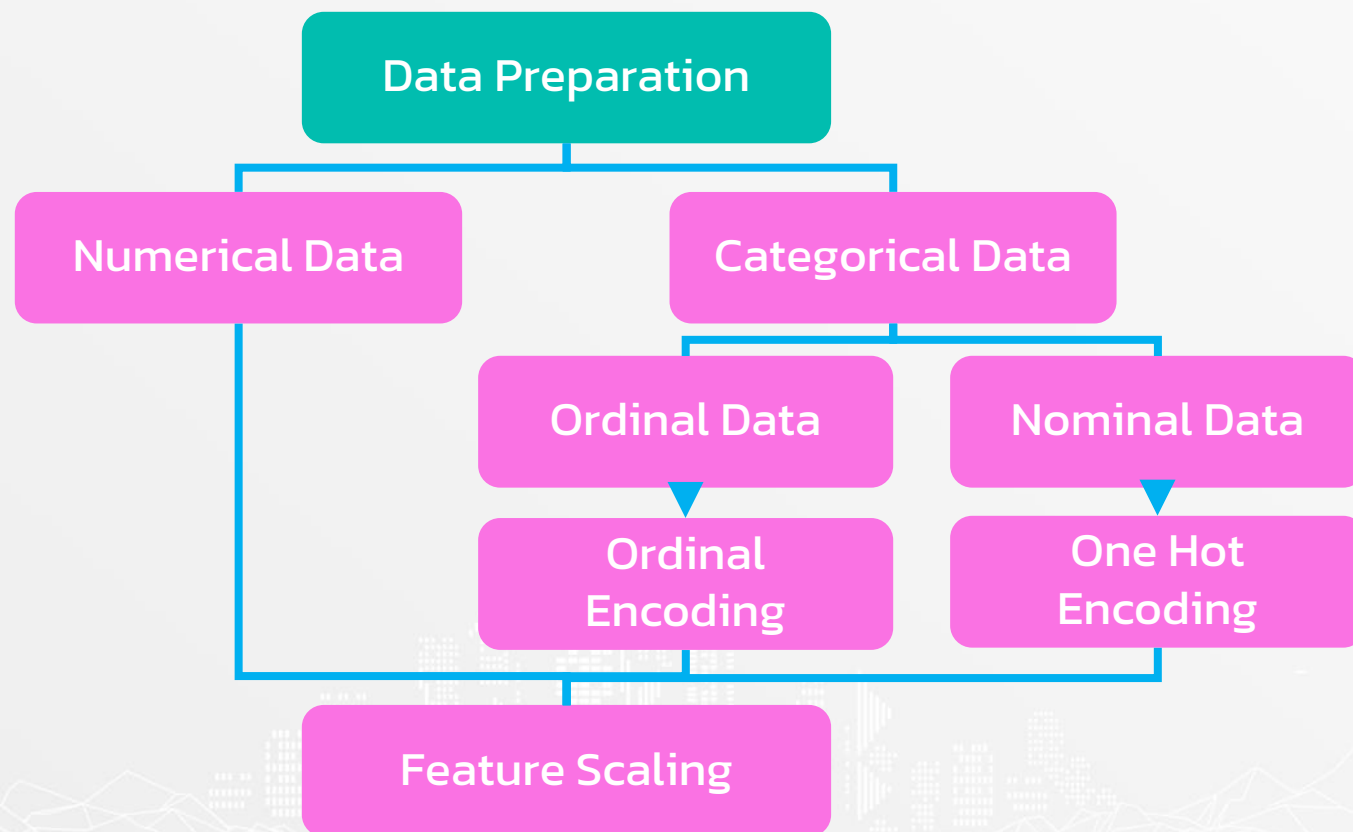
Target

- target : การเป็นโรคไวรัสตับอักเสบ c (0 = ไม่เป็น, 1 = เป็น)

What we learn from this project?



Data Preparation



File



02. HEPATITIS C



hepatitis_c_model.pickle



hepatitis_c_mc.ipynb



hepatitis_c_md.ipynb



hepatitis_c_dataset.csv

Classification Tree

