

# REGRESSION TREE

BY TAUTOLOGY

# Regression Tree



# Introduction

# Introduction

What is  
Regression Tree?

Data for Regression  
Tree

Pros & Cons

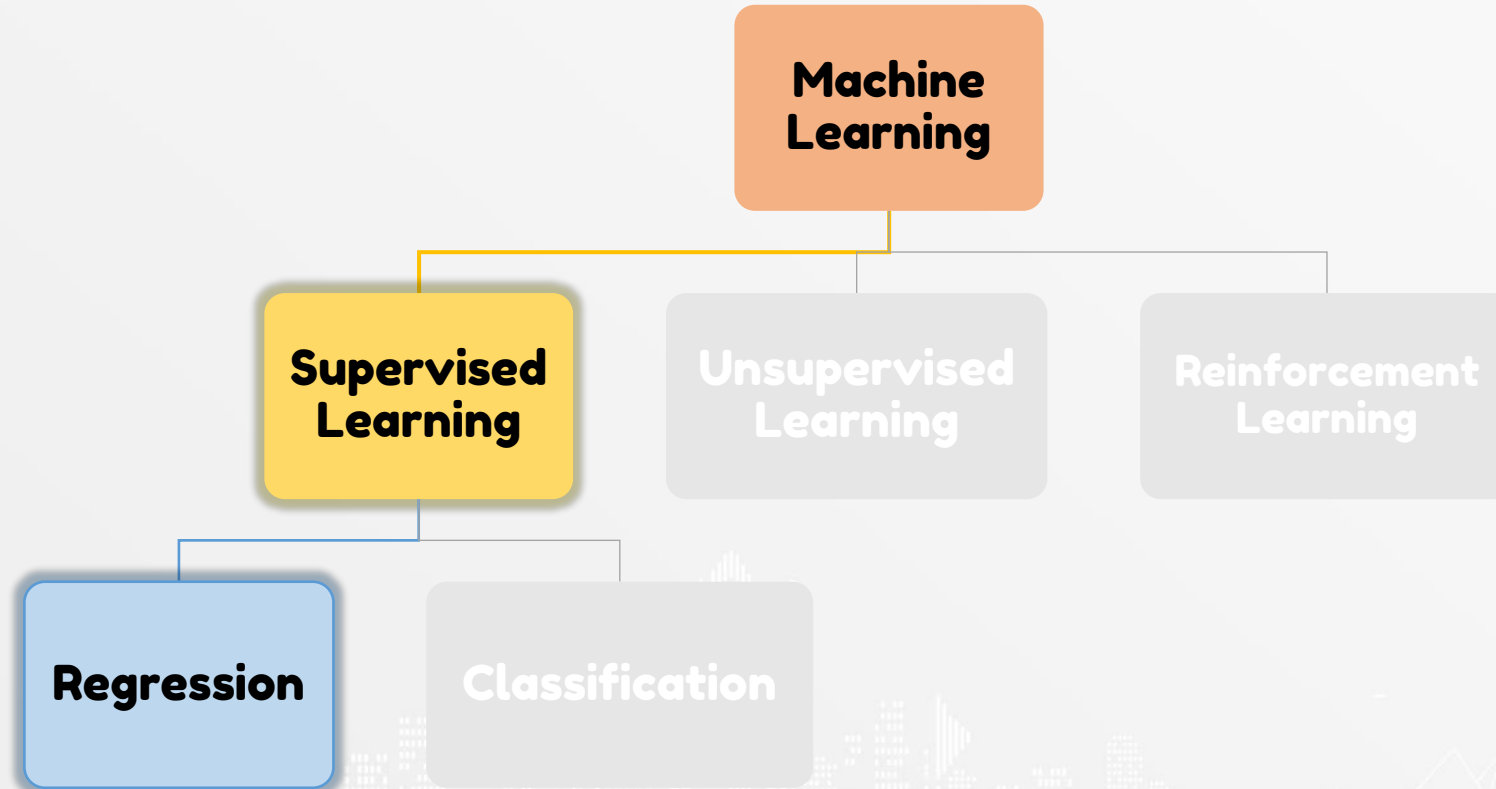
Real World  
Application

# What is Regression Tree?

**Regression Tree** เป็นหนึ่งใน algorithm ประเภท supervised learning ที่ใช้สำหรับแก้ปัญหา regression โดยมีหลักการทำงานคือ การสร้างชุดของกฎเพื่อประมาณค่าตัวแปรตาม

regression

# What is Regression Tree?





# What is Regression Tree?

ตัดแปรต้น

ตัดแปรตาม

area	price (ล้านบาท)
120	2.0
150	2.0
250	3.5
280	3.5
500	20.0
550	20.0

**area  $\leq$  390**area  $\leq$  200

price = 2.0

area &gt; 200

price = 3.5

**area > 390**

price = 20.0

# What is Regression Tree?

**area  $\leq 390$**

area  $\leq 200$

price = 2.0

$area \leq 200$

=> If area  $\leq 390$  and area  $\leq 200$ , then price=2.0

area  $> 200$

price = 3.5

=> If area  $\leq 390$  and area  $> 200$ , then price=3.5

**area  $> 390$**

price = 20.0

=> If area  $> 390$ , then price=20.0



# Introduction

**What is  
Regression Tree?**



**Data for Regression  
Tree**



**Pros & Cons**



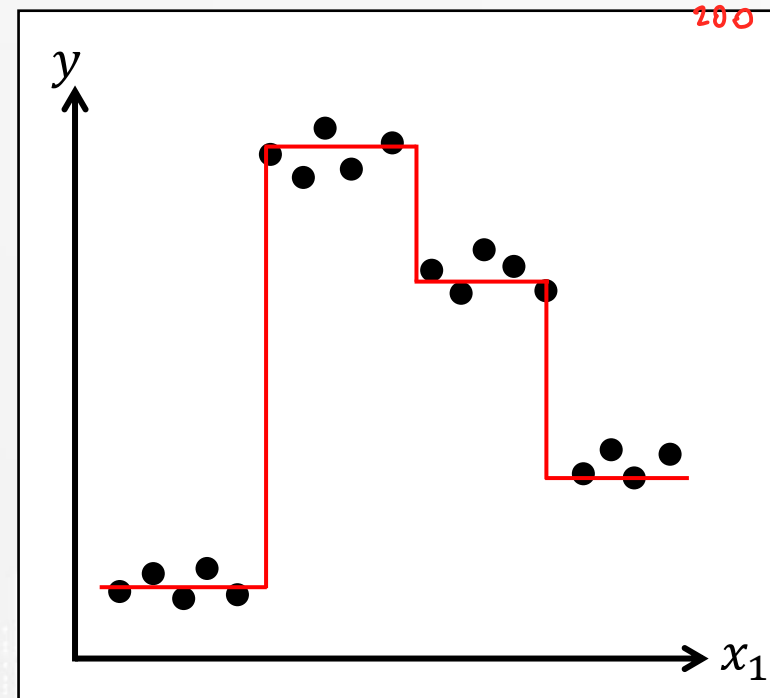
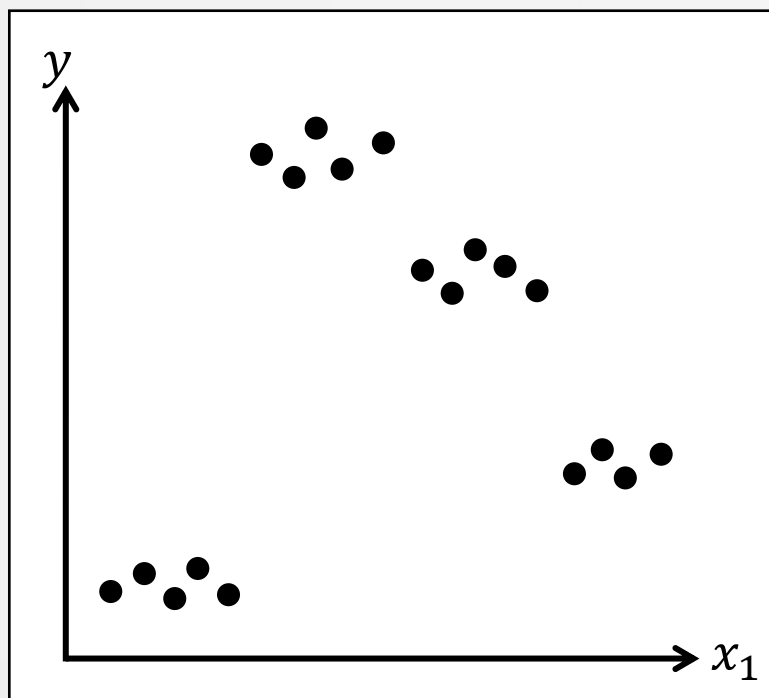
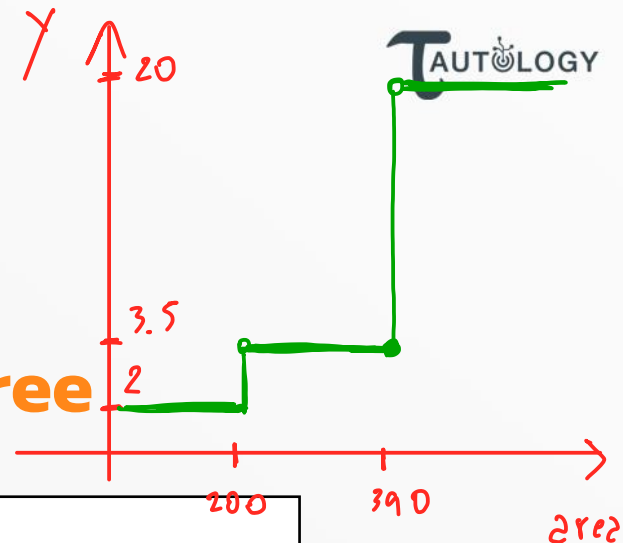
**Real World  
Application**



# Data for Regression Tree

step function

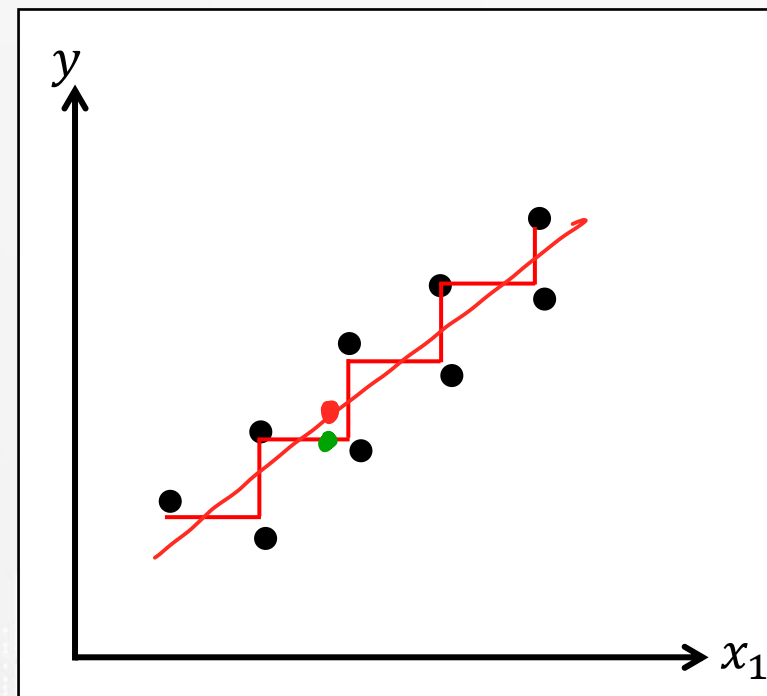
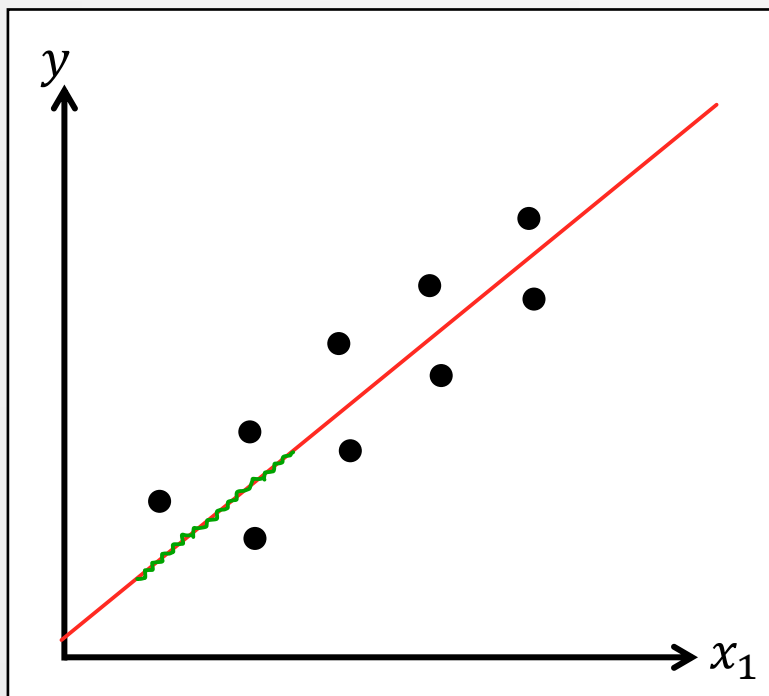
ตัวอย่างของข้อมูลที่เข้ากับ Regression Tree



# Data for Regression Tree

not step function

ตัวอย่างของข้อมูลที่ไม่เหมาะกับ Regression Tree



# Introduction

**What is  
Regression Tree?**



**Data for Regression  
Tree**



**Pros & Cons**



**Real World  
Application**



# Pros & Cons

## ข้อดี

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

## ข้อเสีย

- ง่ายต่อการเกิด overfitting
- การเปลี่ยนแปลงข้อมูลเพียงเล็กน้อยใน training อาจส่งผลให้ model เปลี่ยนแปลงอย่างมาก

## ข้อจำกัด

- ค่าพยากรณ์ที่ได้จะเป็นค่าเฉลี่ยของข้อมูลที่อยู่ในกฎเดียวกัน

# Introduction

**What is  
Regression Tree?**



**Data for Regression  
Tree**



**Pros & Cons**

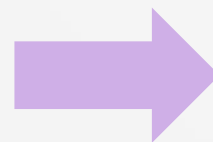


**Real World  
Application**





# Real World Application



**การทำนายราคา Bitcoin**  
โดยพิจารณาจากราคา open high  
low close

อ้างอิง : [2019, Rathan et al.] Crypto-Currency price prediction  
using Decision Tree and Regression techniques

# Real World Application



**การทำนายราคาตั๋วเครื่องบิน**  
โดยพิจารณาจากสายการบิน วันที่  
การเดินทาง เวลาเดินทาง ระยะเวลา  
เดินทาง จำนวนจุดพัก เป็นต้น

อ้างอิง : [2020, Joshi et al.] Airline Prices Analysis and Prediction  
Using Decision Tree Regressor

# Introduction

**What is  
Regression Tree?**



**Data for Regression  
Tree**



**Pros & Cons**



**Real World  
Application**

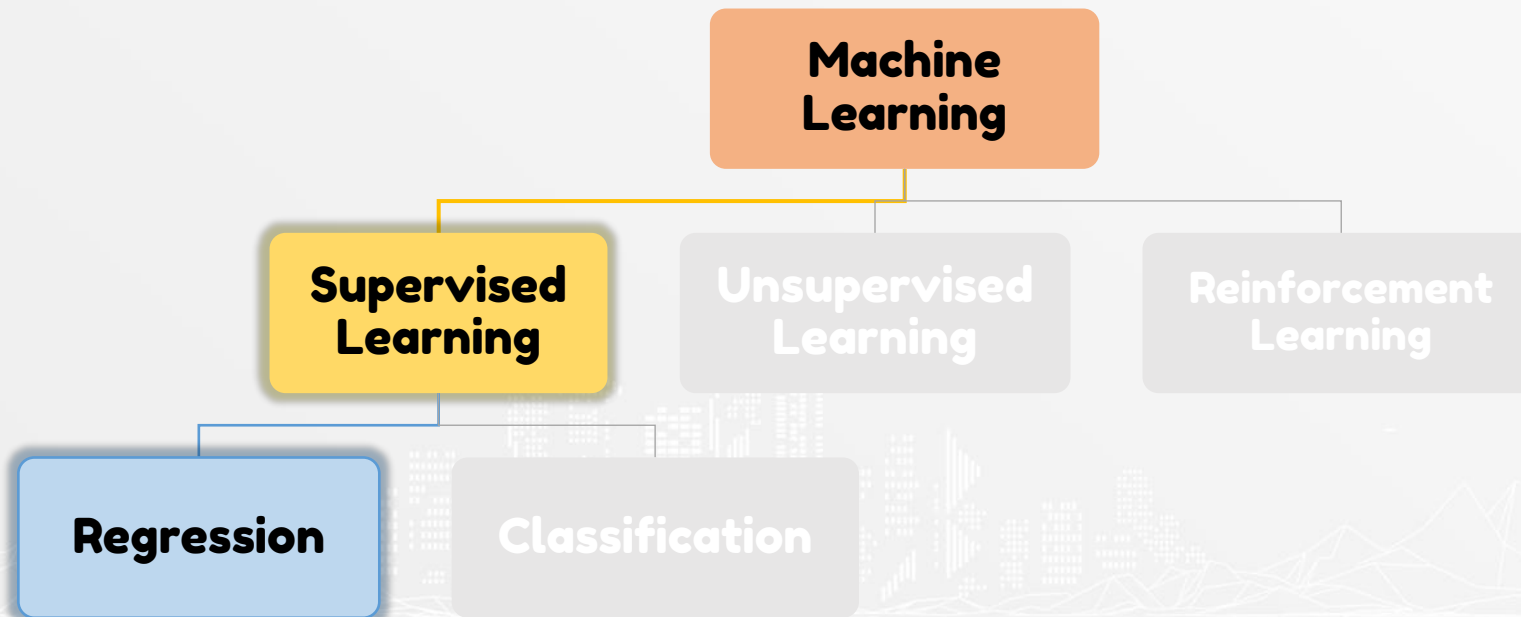


# Regression Tree



# Regression Tree

**Regression 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)

# Assumption

- No Missing Features (perfect case : ឯង feature ၁၀၀/၁၀၀)

# Model Creation

**Assumption**



Real Face of the  
Model



How to Create Model  
(Math)



How to Create Model  
(Code)



# Real Face of the Model

Regression Tree คือ ชุดของกฎเพื่อประมาณค่าตัวแปรตาม

**area  $\leq$  390**

area  $\leq$  200

price = 2.0

$\Rightarrow$  If area  $\leq$  390 and area  $\leq$  200, then price=2.0

area  $>$  200

price = 3.5

$\Rightarrow$  If area  $\leq$  390 and area  $>$  200, then price=3.5

**area  $>$  390**

price = 20.0

$\Rightarrow$  If area  $>$  390, then price=20.0

# Model Creation

**Assumption**



**Real Face of the  
Model**



**How to Create Model  
(Math)**



**How to Create Model  
(Code)**





# How to Create Model (Math)

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

# How to Create Model (Math)

## ตัวอย่างการคำนวณ Regression Tree

area	price
120	2.0
150	2.0
250	3.5
280	3.5
500	20
550	20

ตารางแสดงข้อมูลพื้นที่และราคาของบ้าน

# How to Create Model (Math)

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

area	price
120	2.0
150	2.0
250	3.5
280	3.5
500	20
550	20

120

`unique_values(area) = {120, 150, 250, 280, 500, 550}`

# How to Create Model (Math)

☑ **Step 2** : ตั้งคำถามจาก unique values

unique\_values(area)  
= {120, 150, 250, 280, 500, 550}



Question1 : area  $\leq$  135 ?

Question2 : area  $\leq$  200 ?

Question3 : area  $\leq$  265 ?

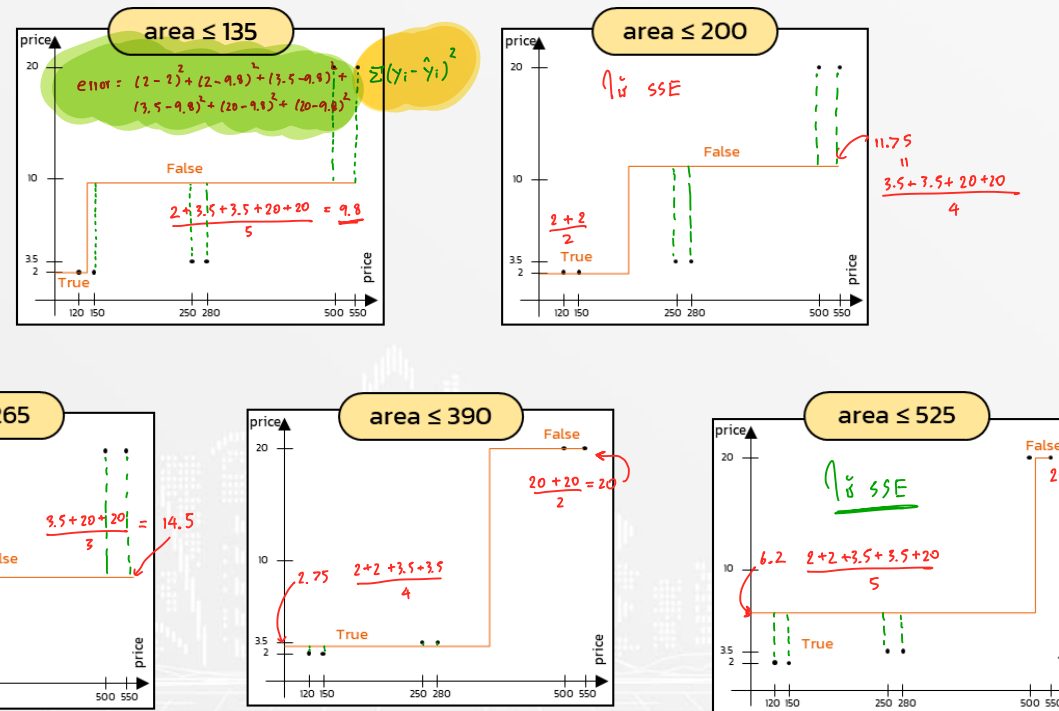
Question4 : area  $\leq$  390 ?

Question5 : area  $\leq$  525?

# How to Create Model (Math)

✓ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ regression tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น

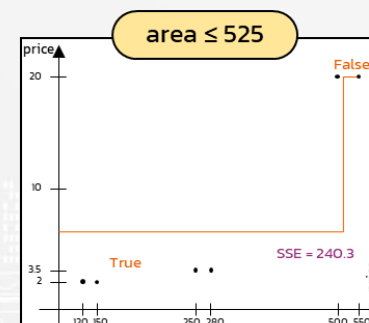
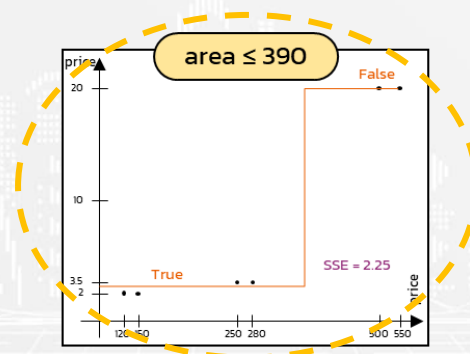
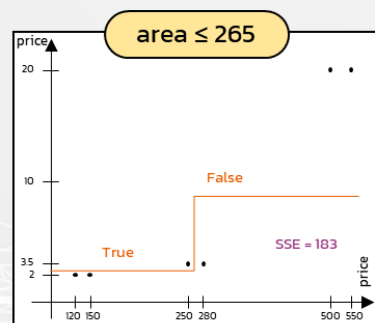
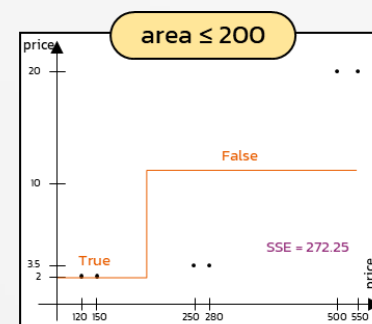
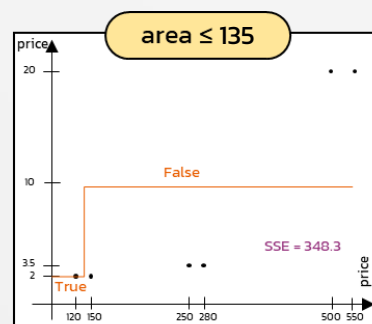
Root Node



# How to Create Model (Math)

✓ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ regression tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น

Root Node





# How to Create Model (Math)

☑ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ regression tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น

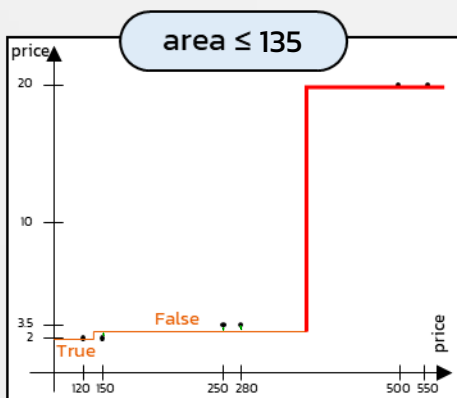


# How to Create Model (Math)

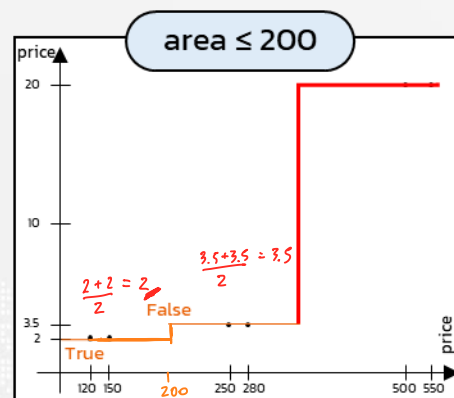
✓ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ regression tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น

Secondary Node

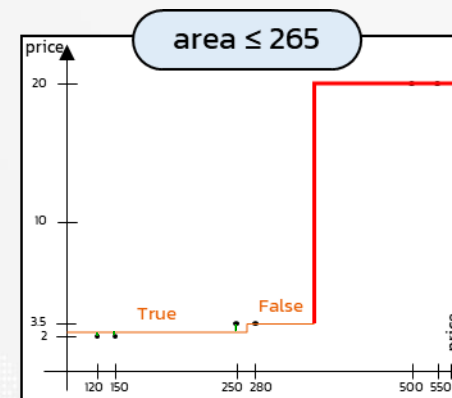
$$SSE > 0$$



$$SSE = 0$$



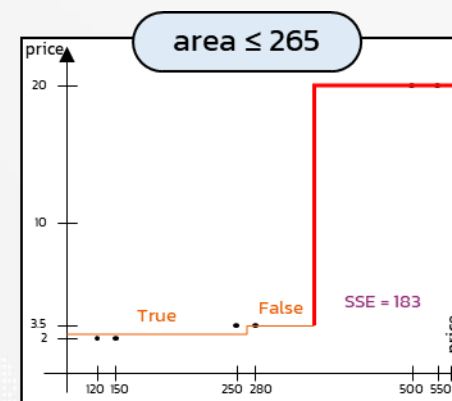
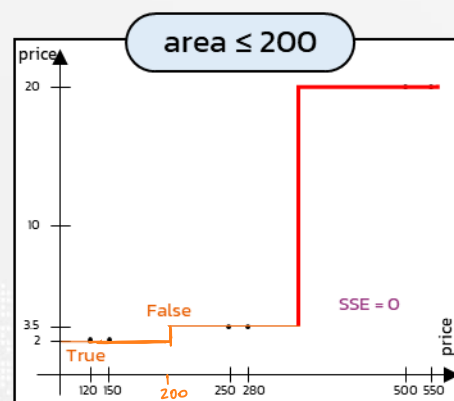
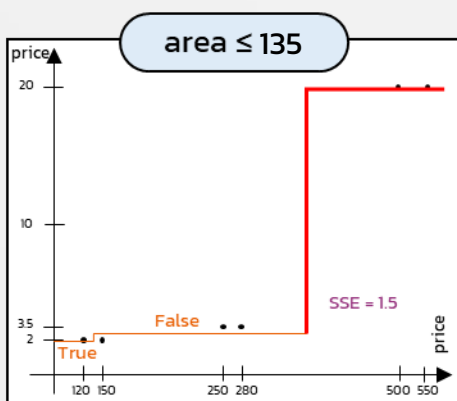
$$SSE > 0$$



# How to Create Model (Math)

✓ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ regression tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น

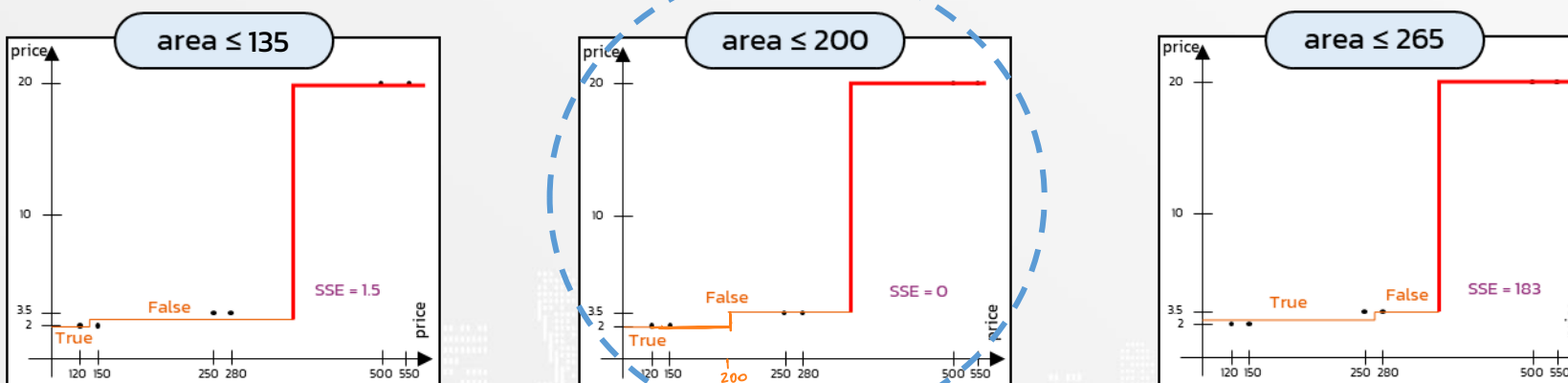
Secondary Node



# How to Create Model (Math)

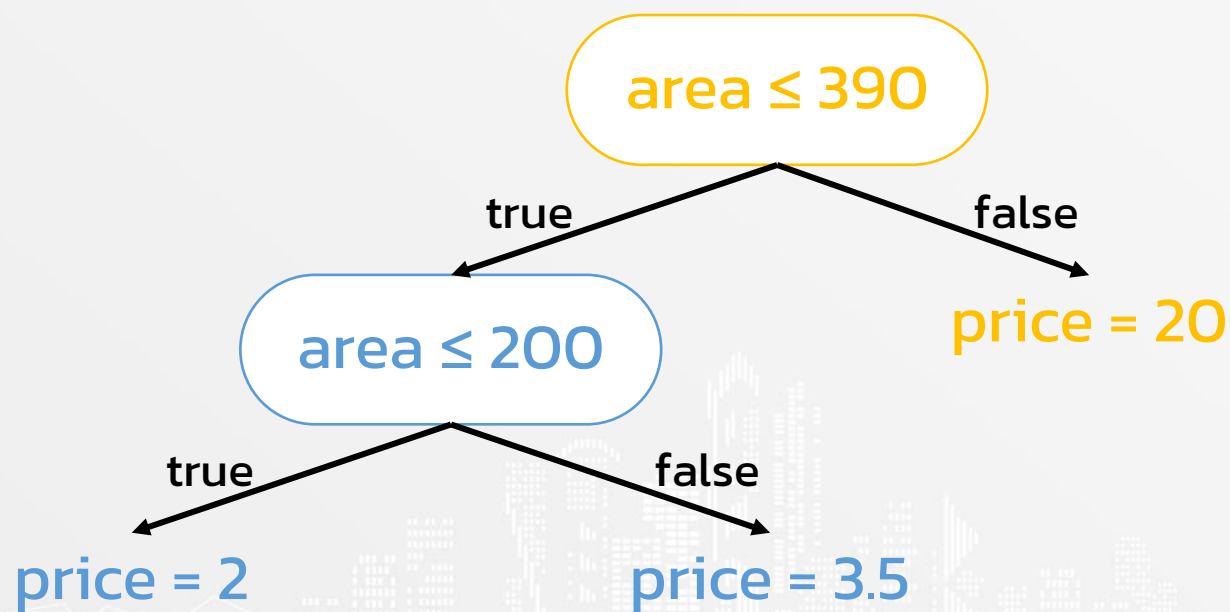
✓ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ regression tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น

Secondary Node



# How to Create Model (Math)

☑ **Step 3** : ในแต่ละชั้น, ตั้งคำถามที่ทำให้ regression tree มีความสามารถในการพยากรณ์มากยิ่งขึ้น



# Model Creation

**Assumption**



**Real Face of the  
Model**



**How to Create Model  
(Math)**



**How to Create Model  
(Code)**



# How to Create Model (Code)

## ตัวอย่าง Code สำหรับ Regression Tree

area	price
120	2.0
150	2.0
250	3.5
280	3.5
500	20
550	20

ตารางแสดงข้อมูลพื้นที่และราคาของบ้าน



# How to Create Model (Code)

- Code สำหรับสร้าง model จากข้อมูลของเราโดยที่

$$X = \begin{bmatrix} 120 \\ 150 \\ 250 \\ 280 \\ 500 \\ 550 \end{bmatrix}, \quad y = \begin{bmatrix} 2 \\ 2 \\ 3.5 \\ 3.5 \\ 20 \\ 20 \end{bmatrix}$$

```
1 reg = DecisionTreeRegressor()  
2 reg.fit(X, y)
```

DecisionTreeRegressor()

# How to Create Model (Code)

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

```
1 print(r)
```

```
| --- area <= 390.00  
|   | --- area <= 200.00  
|   |   | --- value: [2.00]  
|   | --- area > 200.00  
|   |   | --- value: [3.50]  
| --- area > 390.00  
|   | --- value: [20.00]
```

# 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)**



# Regression Tree



# Prediction

# Prediction

Regression Tree คือ ชุดของกฎเพื่อประมาณค่าตัวแปรตาม

**area  $\leq$  390**

area  $\leq$  200

price = 2.0

$\Rightarrow$  If area  $\leq$  390 and area  $\leq$  200, then price=2.0

area  $>$  200

price = 3.5

$\Rightarrow$  If area  $\leq$  390 and area  $>$  200, then price=3.5

**area  $>$  390**

price = 20.0

$\Rightarrow$  If area  $>$  390, then price=20.0



# Prediction

If area  $\leq 390$  and area  $\leq 200$ , then price=2.0

If area  $\leq 390$  and area  $> 200$ , then price=3.5

If area  $> 390$ , then price=20.0

# Prediction

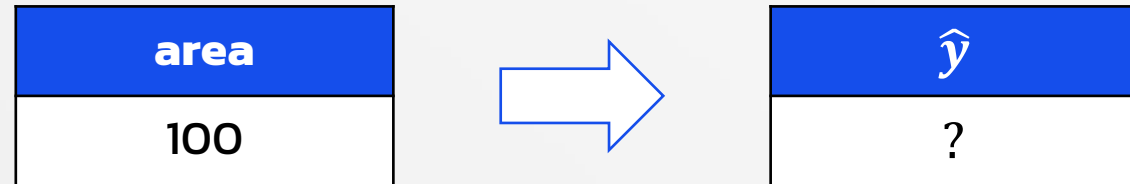
1-Sample

Multi-Sample

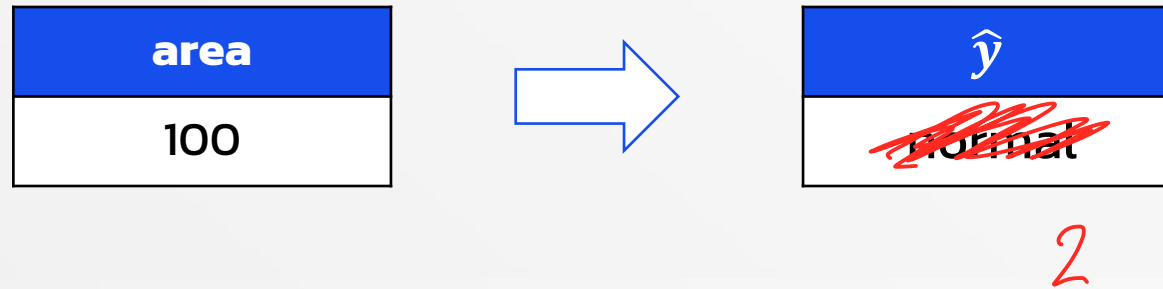
Code

# 1-Sample

## ตัวอย่างการคำนวณ $\hat{y}$



# 1-Sample



If  $\text{area} \leq 390$  and  $\text{area} \leq 200$ , then  $\text{price}=2.0$

If  $\text{area} \leq 390$  and  $\text{area} > 200$ , then  $\text{price}=3.5$

If  $\text{area} > 390$ , then  $\text{price}=20.0$

# Prediction

**1-Sample**



**Multi-Sample**



**Code**



# Multi-Sample

## ตัวอย่างการคำนวณ $\hat{y}$

area	$\hat{y}$
160	?
220	?
300	?
480	?

# Multi-Sample

If area  $\leq 390$  and area  $\leq 200$ , then price=2.0

If area  $\leq 390$  and area  $> 200$ , then price=3.5

If area  $> 390$ , then price=20.0



# Multi-Sample

area
160
220
300
480



$\hat{y}$
2.0
3.5
3.5
20.0

# Prediction

**1-Sample**



**Multi-Sample**



**Code**



# Code

## ตัวอย่าง code สำหรับการคำนวณ $\hat{y}$

area	$\hat{y}$
160	?
220	?
300	?
480	?

# Code

- Code สำหรับสร้าง model จากข้อมูลของเราโดยที่

$$X = \begin{bmatrix} 0 & 29 \\ 0 & 26 \\ 1 & 30 \\ 1 & 28 \end{bmatrix} \quad \begin{bmatrix} 160 \\ 220 \\ 300 \\ 480 \end{bmatrix}$$

```
1 reg.predict(X)
```

```
array([ 2. ,  3.5,  3.5, 20. ])
```

# Code

ดังนั้น เราจะได้  $\hat{y}$  สำหรับข้อมูลชุดนี้คือ

area
160
220
300
480



$\hat{y}$
2.0
3.5
3.5
20.0

# Code



Code for this section



Open File  
**Model Creation.ipynb**

# Prediction

**1-Sample**



**Multi-Sample**



**Code**





# Regression Tree



<u>machine model</u>	
parameter	no parameter
Linear Regression	CT
Logistic — " —	RT
NN	KNN
DL	NB
SVC, SVR, LDA	
GP	

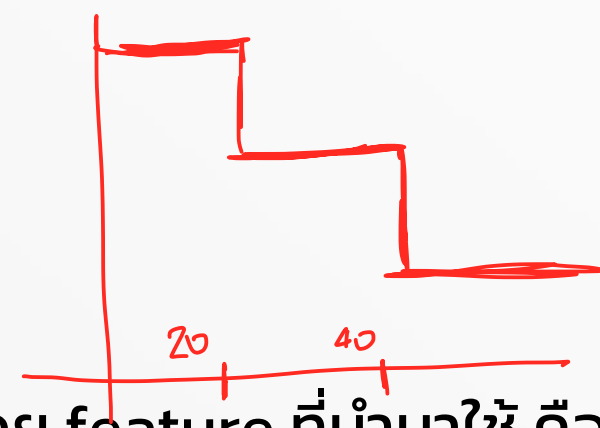
hyper-parameter  $\Rightarrow$  mr config model

# Workshop

# AI in Civil Engineering

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

# Abstract



สร้าง model เพื่อประเมินความแข็งแรงของคอนกรีต โดย feature ที่นำมาใช้ คือ ข้อมูล ส่วนผสมของคอนกรีตนั้น ๆ เช่น

- ปริมาณซีเมนต์
- ปริมาณน้ำ
- อายุของคอนกรีต





# Why this project important?



- สามารถสร้างระบบประเมินคุณภาพของ  
สิ่งก่อสร้างที่ทำงานได้ 24 ชั่วโมง
- สามารถต่อยอดกับการทดสอบความ  
แข็งแรงของวัสดุประเภทอื่น ๆ เช่น เหล็ก

# Who this project is for?

- ✦ วิศวกรโยธา
- ✦ ผู้ตรวจสอบคุณภาพสิ่งก่อสร้าง
- ✦ นักวิเคราะห์ข้อมูล





# Concrete Dataset



<https://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength>

# Concrete Dataset

## Feature

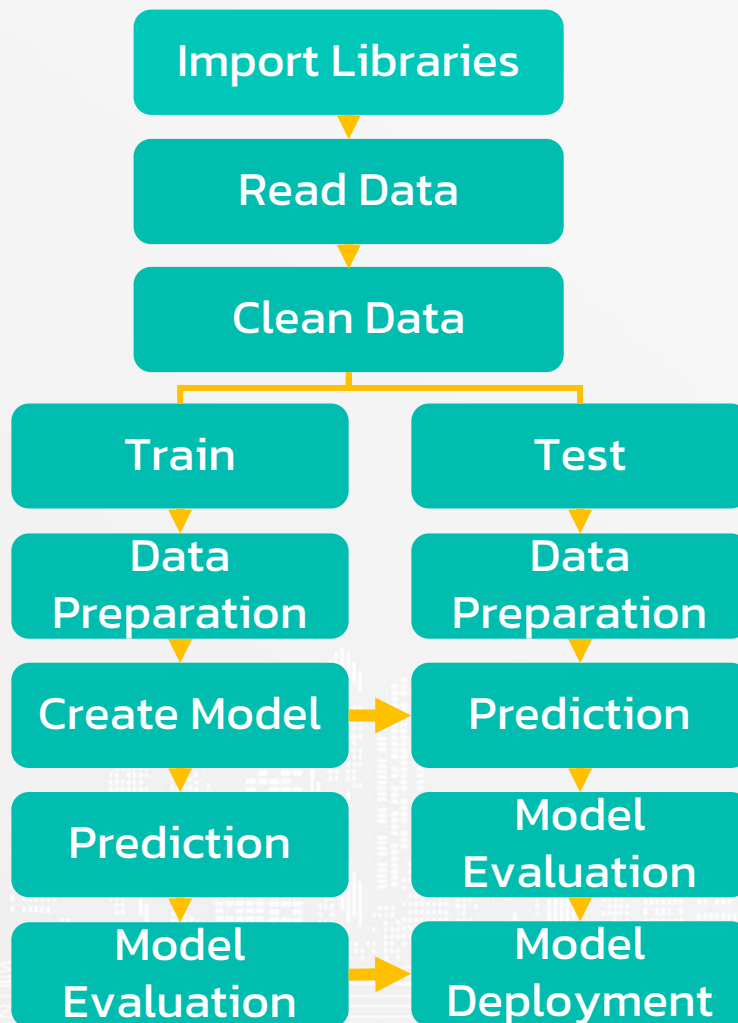
- Cement - ซีเมนต์
- Blast Furnace Slag - ตะกรันเตาหลอม
- Fly Ash - เถ้าลอย
- Water - น้ำ
- Superplasticizer - สารลดน้ำประเภทพิเศษ
- Coarse Aggregate - มวลรวมหยาบ
- Fine Aggregate - มวลรวมละเอียด
- Age - อายุของสิ่งก่อสร้าง

ตัวแปรต้นสัมพันธ์กับตัวแปรตาม

## Target

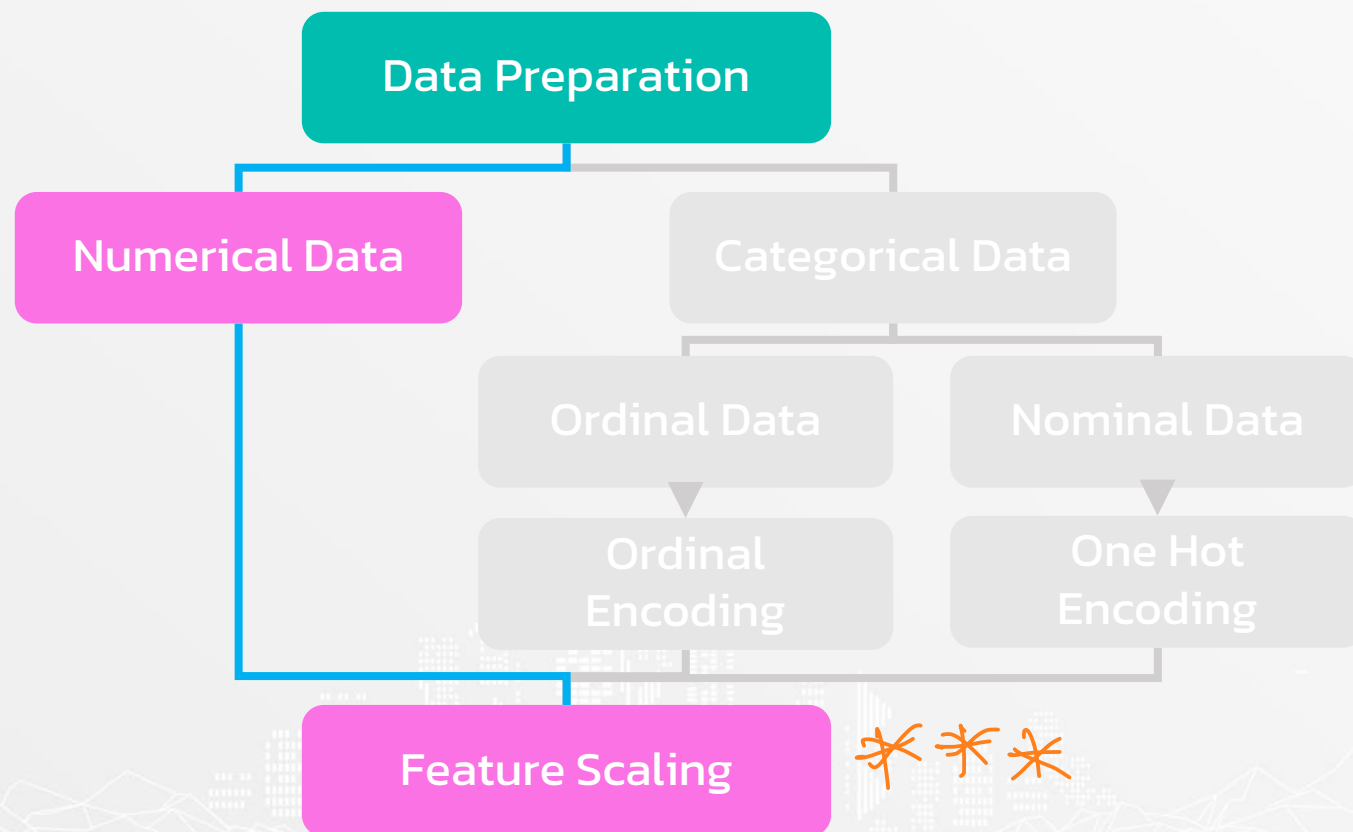
- Concrete compressive strength - ความแข็งแรงของคอนกรีต

# What we learn from this project?



1. USE CASE  
2. RT

# Data Preparation



# File



## 01. CONCRETE STRENGTH



**concrete\_strength\_model.pickle**



**concrete\_strength\_mc.ipynb**



**concrete\_strength\_md.ipynb**



**concrete\_dataset.xlsx**

# Regression Tree

