

TAUTOLOGY  
INNOVATION  
SCHOOL



# MODEL EVALUATION FOR REGRESSION

BY TAUTOLOGY

MADE BY TAUTOLOGY THAILAND  
DO NOT PUBLISH WITHOUT PERMISSION

facebook/tautologyai  
www.tautology.live

# Model Evaluation

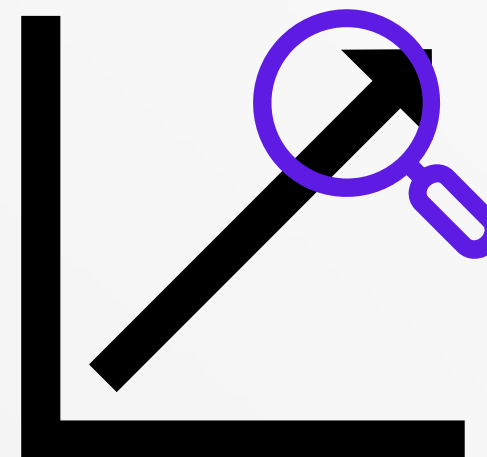
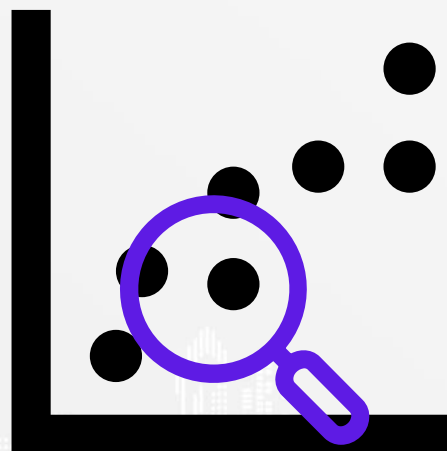
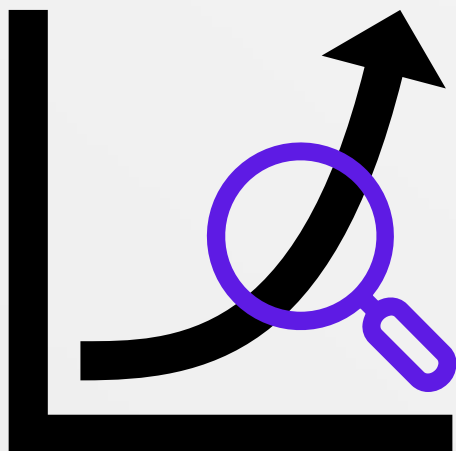
What is Model  
Evaluation?

Why need Model  
Evaluation?

Model Evaluation  
for Regression

# What is Model Evaluation?

Model Evaluation คือการวัดประสิทธิภาพของโมเดล



# Model Evaluation

**What is Model  
Evaluation?**



**Why need Model  
Evaluation?**



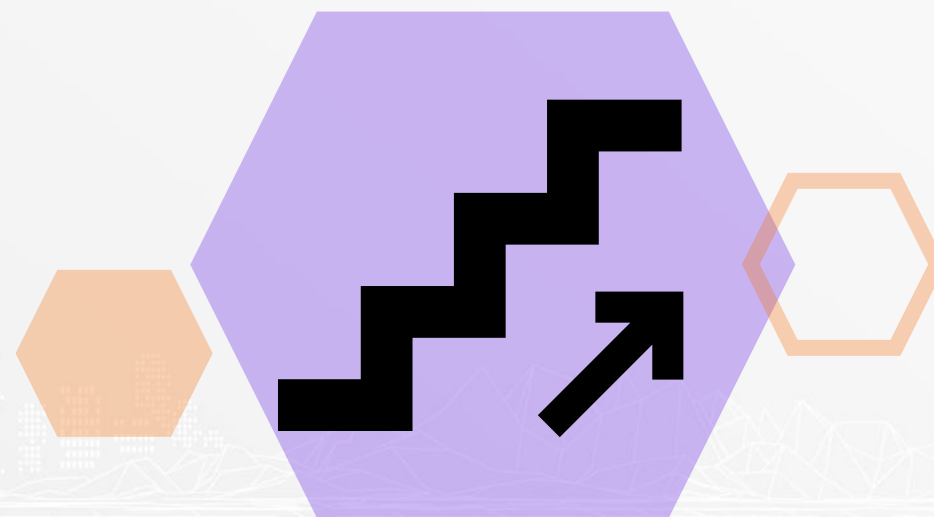
**Model Evaluation  
for Regression**



# Why need Model Evaluation?

$model_1(x_2 \text{ ตัว}), model_2(x_3 \text{ ตัว})$

- เพื่อเลือก model ที่ดีที่สุด ผ่านการเปรียบเทียบประสิทธิภาพ
- เพื่อวิเคราะห์ model แล้วนำไปปรับปรุง และพัฒนาต่อ
- เพื่อวัดประสิทธิภาพของ model ก่อนนำไปใช้งานจริง



# Model Evaluation

**What is Model  
Evaluation?**



**Why need Model  
Evaluation?**

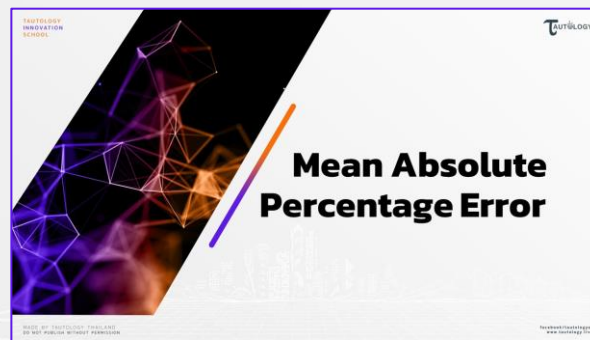
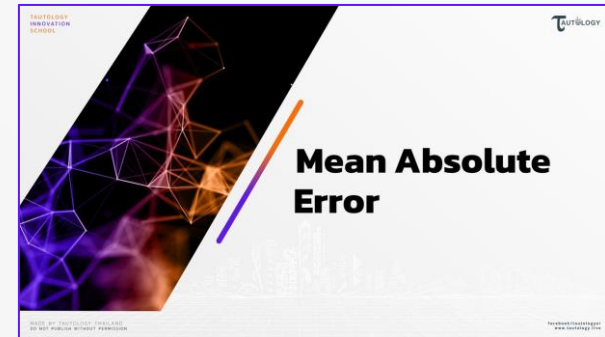
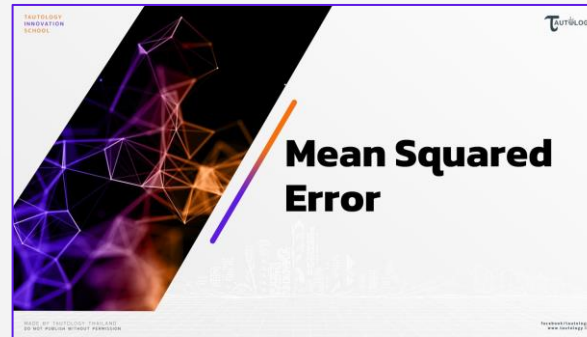


**Model Evaluation  
for Regression**





# Model Evaluation for Regression



R-squared

# $R^2$ Score



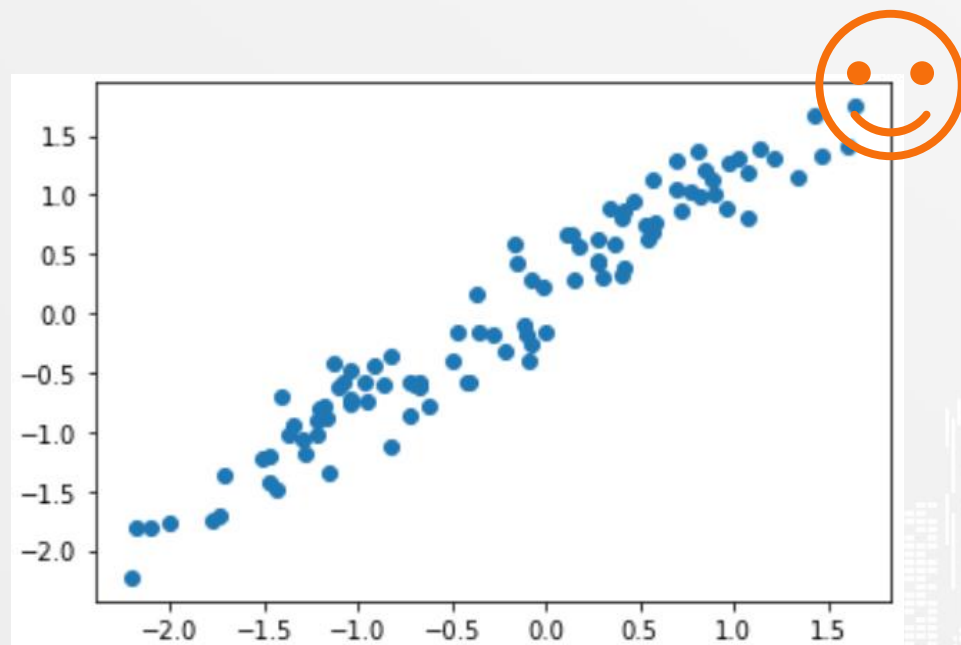
# $R^2$ score

- What is  $R^2$  score?
- Formula
- Step to Calculate  $R^2$
- Example
- Code

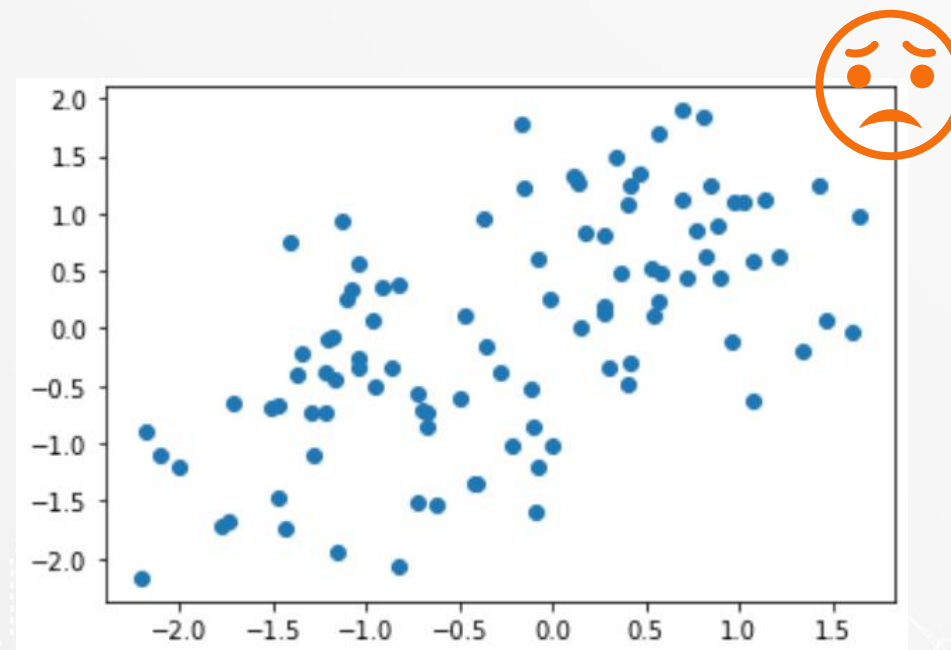
# What is $R^2$ score?

$(-\infty, 1]$

$R^2$  score คือ ค่าที่บอกความสัมพันธ์ระหว่างค่าจริง และค่าพยากรณ์



**Model 1**



**Model 2**

# Formula

 $[-\infty, 1]$ 

$$R^2 = 1 \iff \sum_{i=1}^n (y_i - \hat{y}_i)^2 = 0 \iff \forall i, y_i = \hat{y}_i$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

target ของ

- $y_i$  คือ sample ที่  $i$
- $\hat{y}_i$  คือ ค่าที่พยากรณ์ได้จากโมเดลของ sample ที่  $i$
- $\bar{y}$  คือ ค่าเฉลี่ยของข้อมูล (target)
- $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$

# Step to calculate $R^2$

1. เก็บค่า  $y_i$  และ  $\hat{y}_i$
2. หาค่า  $\bar{y}$
3. วัดประสิทธิภาพของ model ตามสูตรของ  $R^2$

# Example

## 1. เก็บค่า $y_i$ และ $\hat{y}_i$

	$y_i$	$\hat{y}_i$
0	1168	1204.183
1	1488	1498.152
2	1232	1199.06
3	949	947.087
4	439	438.018
5	262	275.159
6	897	873.342

ตารางแสดงข้อมูลของราคาบ้านจริง และราคาบ้านที่พยากรณ์ได้จากโมเดล  
โดยใช้ feature ที่ใช้คือจำนวนห้องและพื้นที่ของบ้าน



# Example

2. หาค่า  $\bar{y}$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

$$\bar{y} = \frac{1188+1468+\dots+897}{7}$$

$$\bar{y} = 919.29$$

# Example

## 3. วัดประสิทธิภาพของ model ตามสูตรของ $R^2$

	$y_i$	$\hat{y}_i$
0	1168	1204.183
1	1488	1498.152
2	1232	1199.06
3	949	947.087
4	439	438.018
5	262	275.159
6	897	873.342

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$R^2 = 1 - \frac{(1168 - 1204.183)^2 + \dots + (897 - 873.342)^2}{(1168 - 919.29)^2 + \dots + (897 - 919.29)^2}$$

$$R^2 = 0.997$$

# Code

	Actual_SalePrice	Predicted_SalePrice
0	1168.0	1204.18303571
1	1488.0	1498.15178571
2	1232.0	1199.06026786
3	949.0	947.08705357
4	439.0	438.01785714
5	262.0	275.15848214
6	897.0	873.34151786

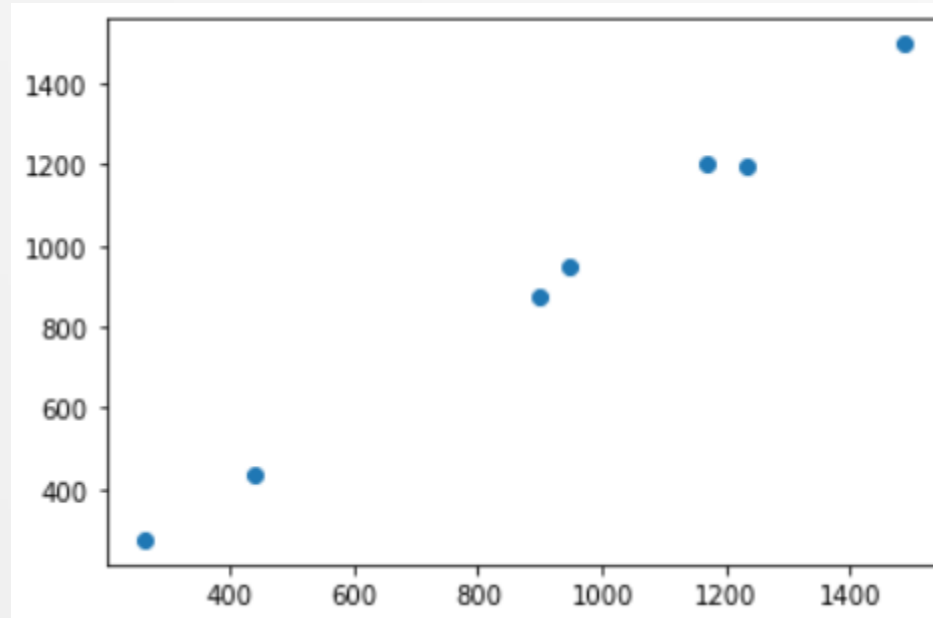
ตารางแสดงข้อมูลของราคาบ้านจริง และราคาบ้านที่พยากรณ์ได้จากโมเดล  
โดยใช้ feature ที่ใช้คือจำนวนห้องและพื้นที่ของบ้าน

# Code

```
1 r2_score(y_true, y_pred)
```

```
0.9971801836617127
```

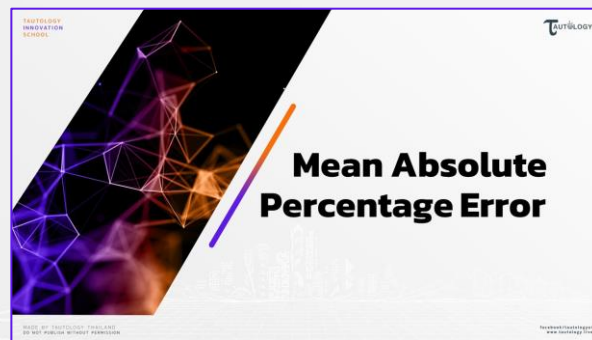
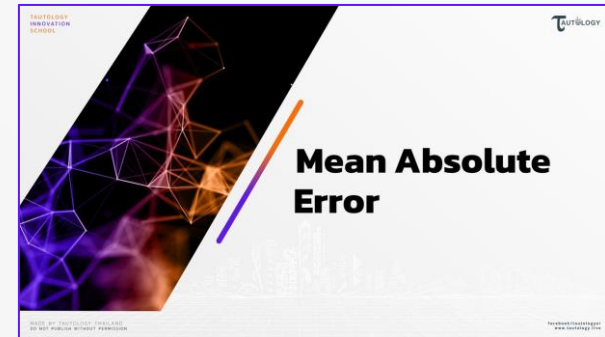
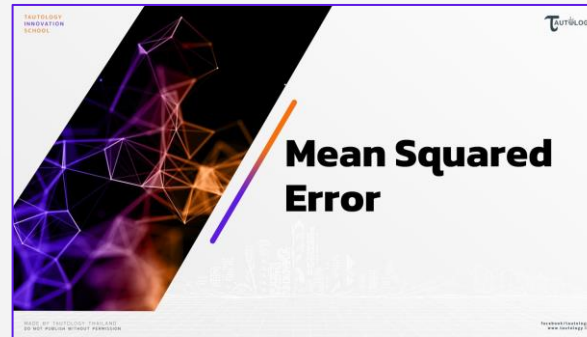
# Code



$$R^2 = 0.99718$$



# Model Evaluation for Regression



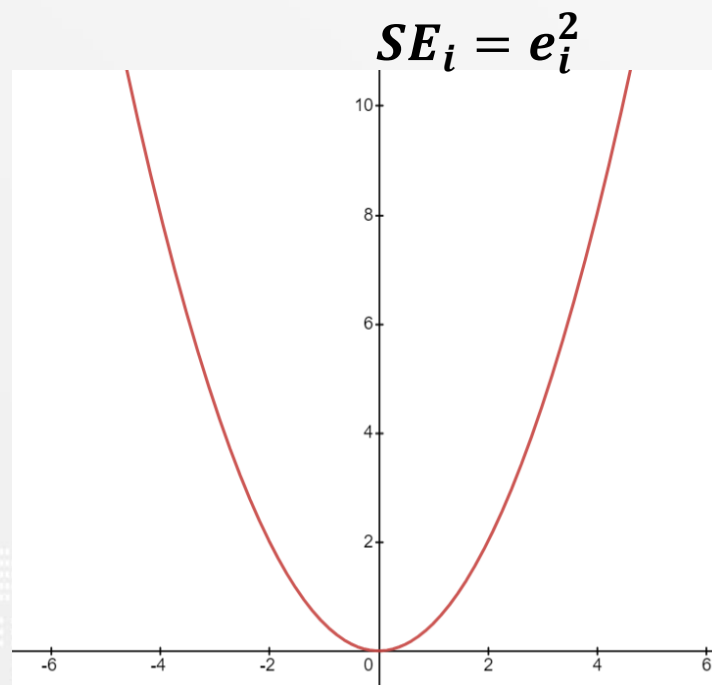
# Mean Squared Error

# Mean Squared Error

- What is Mean Squared Error?
- Formula
- Step to Calculate MSE
- Example
- Code

# What is Mean Squared Error?

Mean Squared Error (MSE) คือ ค่าเฉลี่ยของ error (ผลต่างของค่าจริงและค่าพยากรณ์) ยกกำลังสอง



$$e_i = y_i - \hat{y}_i$$

# Formula

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- $y_i$  คือ sample ที่  $i$
- $\hat{y}_i$  คือ ค่าที่พยากรณ์ได้จากโมเดลของ sample ที่  $i$
- $n$  คือ จำนวน sample



# Step to Calculate MSE

1. เก็บค่า  $y_i$  และ  $\hat{y}_i$
2. วัดประสิทธิภาพของ model ตามสูตรของ mean squared error ( $MSE$ )

# Example

## 1. เก็บค่า $y_i$ และ $\hat{y}_i$

	$y_i$	$\hat{y}_i$
0	1168	1204.183
1	1488	1498.152
2	1232	1199.06
3	949	947.087
4	439	438.018
5	262	275.159
6	897	873.342

ตารางแสดงข้อมูลของราคาบ้านจริง และราคาบ้านที่พยากรณ์ได้จากโมเดล  
โดยใช้ feature ที่ใช้คือจำนวนห้องและพื้นที่ของบ้าน

# Example

## 2. วัดประสิทธิภาพของ model ตามสูตรของ mean squared error ( $MSE$ )

	$y_i$	$\hat{y}_i$
0	1168	1204.183
1	1488	1498.152
2	1232	1199.06
3	949	947.087
4	439	438.018
5	262	275.159
6	897	873.342

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$MSE = \frac{(1168 - 1204.183)^2 + \dots + (897 - 873.342)^2}{7}$$

$$MSE = 462.113$$

# Code

	Actual_SalePrice	Predicted_SalePrice
0	1168.0	1204.18303571
1	1488.0	1498.15178571
2	1232.0	1199.06026786
3	949.0	947.08705357
4	439.0	438.01785714
5	262.0	275.15848214
6	897.0	873.34151786

ตารางแสดงข้อมูลของราคาบ้านจริง และราคาบ้านที่พยากรณ์ได้จากโมเดล  
โดยใช้ feature ที่ใช้คือจำนวนห้องและพื้นที่ของบ้าน

# Code

mse = 400    แปลว่าอะไร?

โดยเฉลี่ยแล้ว  $(y - \hat{y})^2 = 400 \rightarrow y - \hat{y} = 20$

```
1 mean_squared_error(y_true, y_pred)
```

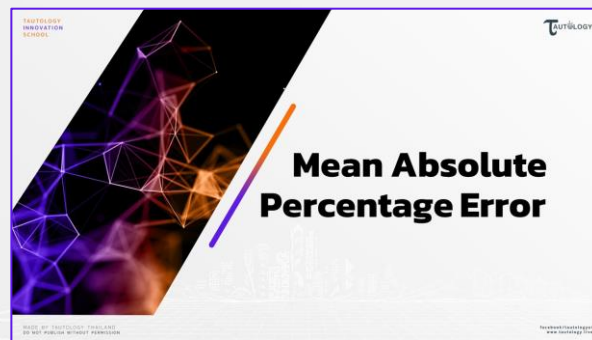
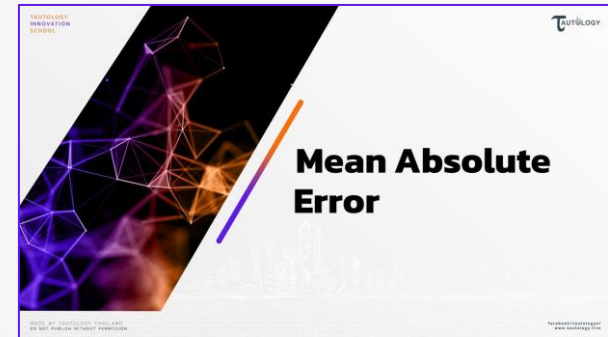
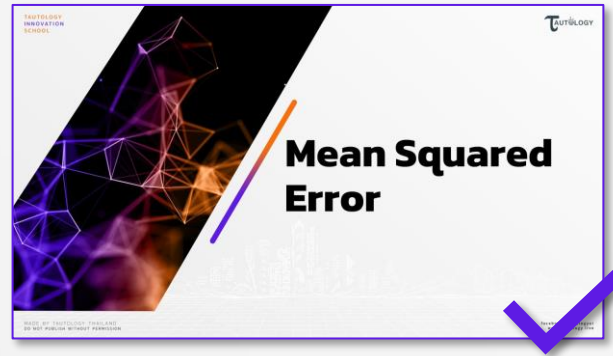
```
462.1128826530673
```

$\hat{y} = 1000 \in [980, 1020]$

$\hat{y} = 200$ , mse = 100  $\Rightarrow y - \hat{y} = 10$   
(190, 210)



# Model Evaluation for Regression



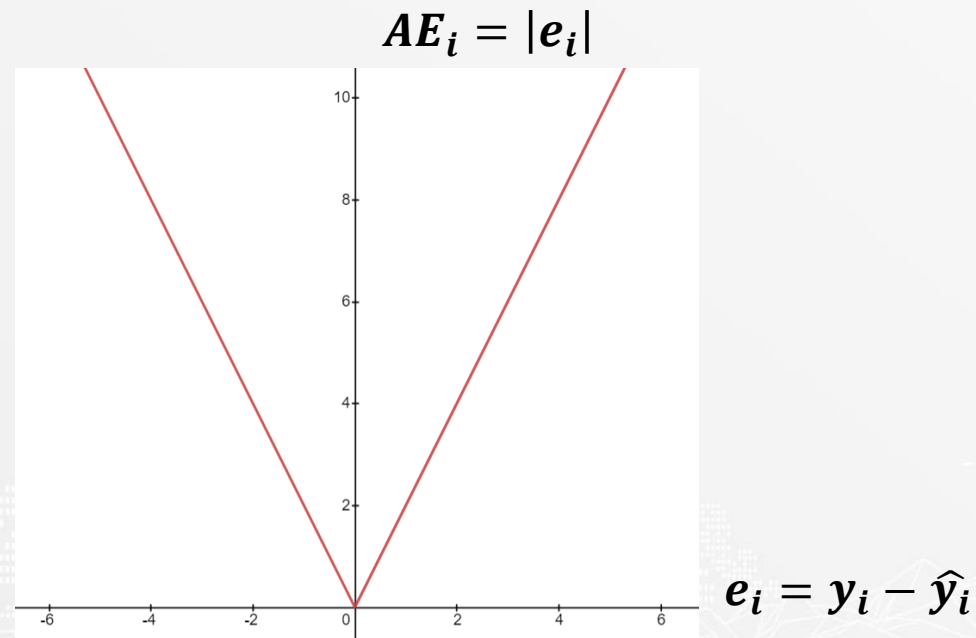
# Mean Absolute Error

# Mean Absolute Error

- What is Mean Absolute Error?
- Formula
- Step to Calculate MAE
- Example
- Code

# What is Mean Absolute Error?

Mean Absolute Error (MAE) คือ ค่าเฉลี่ยของ absolute ของ error (ผลต่างของค่าจริงและค่าพยากรณ์)



# Formula

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- $y_i$  คือ sample ที่  $i$
- $\hat{y}_i$  คือ ค่าที่พยากรณ์ได้จากโมเดลของ sample ที่  $i$
- $n$  คือ จำนวน sample

# Step to Calculate MAE

1. เก็บค่า  $y_i$  และ  $\hat{y}_i$
2. วัดประสิทธิภาพของ model ตามสูตรของ mean absolute error ( $MAE$ )



# Example

## 1. เก็บค่า $y_i$ และ $\hat{y}_i$

	$y_i$	$\hat{y}_i$
0	1168	1204.183
1	1488	1498.152
2	1232	1199.06
3	949	947.087
4	439	438.018
5	262	275.159
6	897	873.342

ตารางแสดงข้อมูลของราคาบ้านจริง และราคาบ้านที่พยากรณ์ได้จากโมเดล  
โดยใช้ feature ที่ใช้คือจำนวนห้องและพื้นที่ของบ้าน

# Example

## 2. วัดประสิทธิภาพของ model ตามสูตรของ mean absolute error ( $MAE$ )

	$y_i$	$\hat{y}_i$
0	1168	1204.183
1	1488	1498.152
2	1232	1199.06
3	949	947.087
4	439	438.018
5	262	275.159
6	897	873.342

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$MAE = \frac{1}{7} \{|1168 - 1204.183| + \dots + |897 - 873.342|\}$$

$$MAE = 16.998$$

# Code

	Actual_SalePrice	Predicted_SalePrice
0	1168.0	1204.18303571
1	1488.0	1498.15178571
2	1232.0	1199.06026786
3	949.0	947.08705357
4	439.0	438.01785714
5	262.0	275.15848214
6	897.0	873.34151786

ตารางแสดงข้อมูลของราคาบ้านจริง และราคาบ้านที่พยากรณ์ได้จากโมเดล  
โดยใช้ feature ที่ใช้คือจำนวนห้องและพื้นที่ของบ้าน

# Code

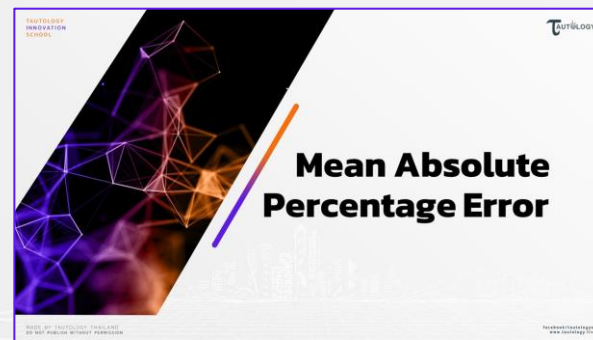
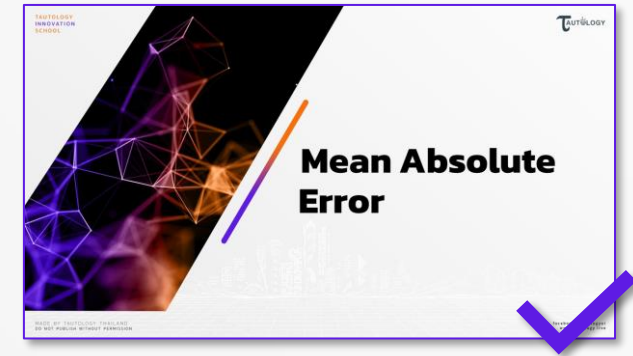
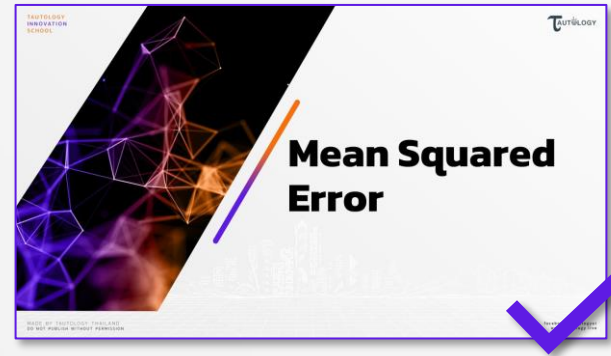
$m a e = 17 \rightarrow$  โดยเฉลี่ยแล้ว  $|y - \hat{y}| = 17$

$\hat{y} = 100 \Rightarrow y \in [100 - 17, 100 + 17]$   
 $[83, 117]$

```
1 mean_absolute_error(y_true, y_pred)
```

```
16.998086734694034
```

# Model Evaluation for Regression



# Mean Absolute Percentage Error

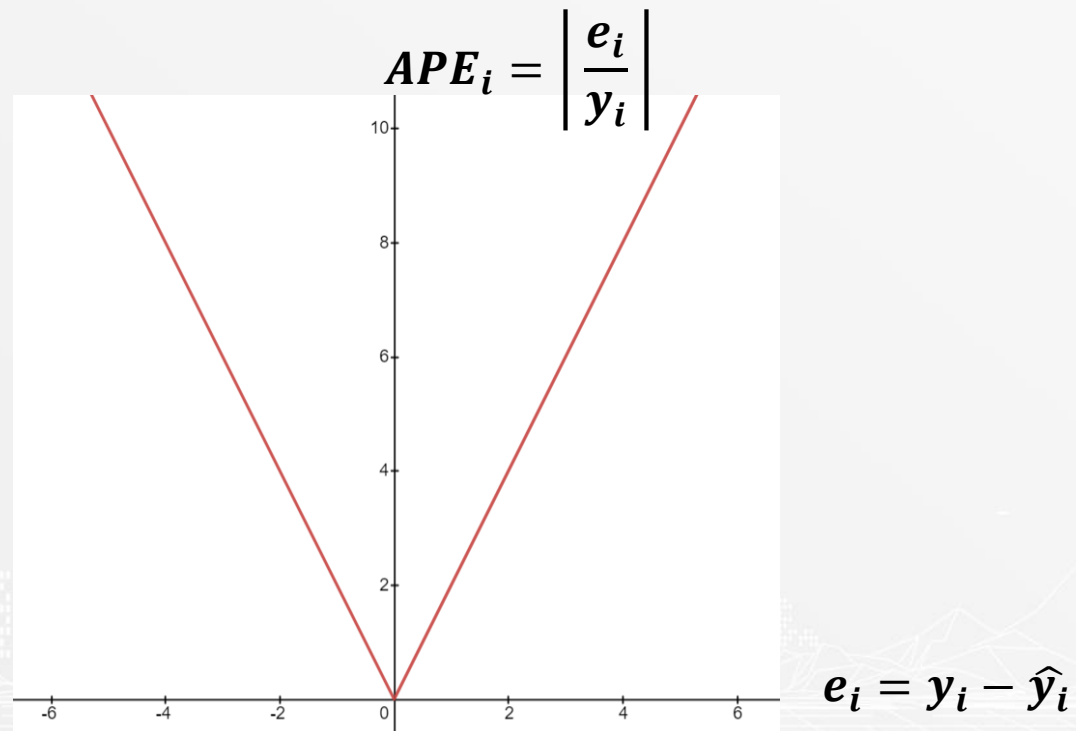


# Mean Absolute Percentage Error

- What is Mean Absolute Percentage Error?
- Formula
- Step to Calculate MAPE
- Example
- Code

# What is Mean Absolute Percentage Error?

Mean Absolute Percentage Error (MAPE) คือ ค่าเฉลี่ยของ absolute ของ อัตราส่วนระหว่าง error (ผลต่างของค่าจริงและค่าพยากรณ์) และข้อมูลจริง



# Formula

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

- $y_i$  คือ sample ที่  $i$
- $\hat{y}_i$  คือ ค่าที่พยากรณ์ได้จากโมเดลของ sample ที่  $i$
- $n$  คือ จำนวน sample

# Step to Calculate MAPE

1. เก็บค่า  $y_i$  และ  $\hat{y}_i$
2. วัดประสิทธิภาพของ model ตามสูตรของ mean absolute percentage error (*MAPE*)

# Example

## 1. เก็บค่า $y_i$ และ $\hat{y}_i$

	$y_i$	$\hat{y}_i$
0	1168	1204.183
1	1488	1498.152
2	1232	1199.06
3	949	947.087
4	439	438.018
5	262	275.159
6	897	873.342

ตารางแสดงข้อมูลของราคาบ้านจริง และราคาบ้านที่พยากรณ์ได้จากโมเดล  
โดยใช้ feature ที่ใช้คือจำนวนห้องและพื้นที่ของบ้าน

# Example

2. วัดประสิทธิภาพของ model ตามสูตรของ mean absolute percentage error (*MAPE*)

	$y_i$	$\hat{y}_i$
0	1168	1204.183
1	1488	1498.152
2	1232	1199.06
3	949	947.087
4	439	438.018
5	262	275.159
6	897	873.342

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

$$MAPE = \frac{1}{7} \left\{ \left| \frac{1168 - 1204.183}{1168} \right| + \dots + \left| \frac{897 - 873.342}{897} \right| \right\}$$

$$MAPE = 0.021$$



# Code

	Actual_SalePrice	Predicted_SalePrice
0	1168.0	1204.18303571
1	1488.0	1498.15178571
2	1232.0	1199.06026786
3	949.0	947.08705357
4	439.0	438.01785714
5	262.0	275.15848214
6	897.0	873.34151786

ตารางแสดงข้อมูลของราคาบ้านจริง และราคาบ้านที่พยากรณ์ได้จากโมเดล  
โดยใช้ feature ที่ใช้คือจำนวนห้องและพื้นที่ของบ้าน

# Code

$mape = 0.02 \leadsto \text{error } 2\%$

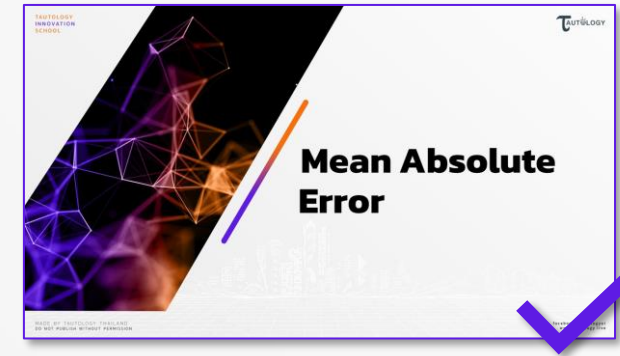
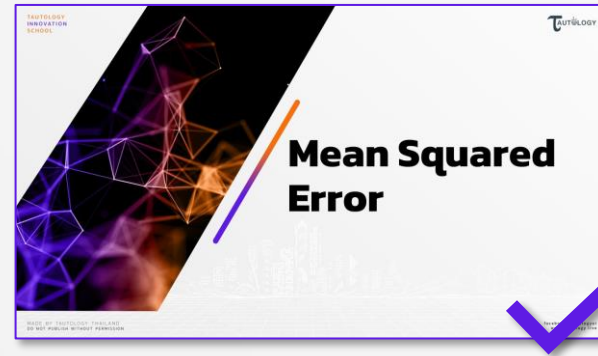
โดยเฉลี่ยแล้ว  $\left| \frac{y - \hat{y}}{y} \right| = 2\%$

$\hat{y} = 100, y \in [100 - 2\%, 100 + 2\%] = [98, 102]$

```
1 mean_absolute_percentage_error(y_true, y_pred)
```

```
0.02076988136170835
```

# Model Evaluation for Regression



# Conclusion

# Conclusion

Name	Formula
$R^2$	$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$ $(-\infty, 1]$
$MSE$	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ $mse = 9$ $\hat{y} = 100$ $y \in [97, 103]$
$MAE$	$MAE = \frac{1}{n} \sum_{i=1}^n  y_i - \hat{y}_i $ $mae = 5$ $\hat{y} = 100$ $y \in [95, 105]$
$MAPE$	$MAPE = \frac{1}{n} \sum_{i=1}^n \left  \frac{y_i - \hat{y}_i}{y_i} \right $ $mape = 0.02$ $error = 2\%$ $\hat{y} = 100$ $y \in [98, 102]$

MAE

ขนาดได้ปริมาณที่คาดเคลื่อน

MAPE

ขนาดได้ % ที่คาดเคลื่อน

$$mape = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

ถ้ามี  $y_j$  ใกล้ 0

ขนาดอย่างนั้น  $y_j = 0.0001$

$$\hat{y}_j = 1$$

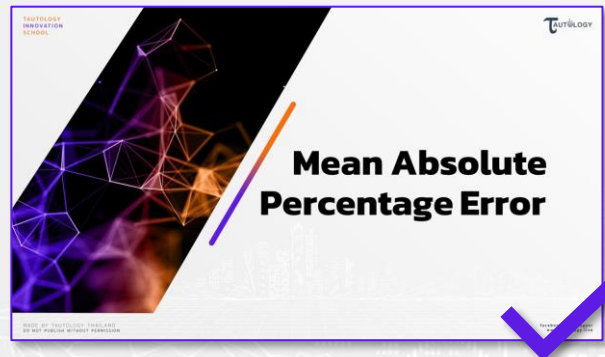
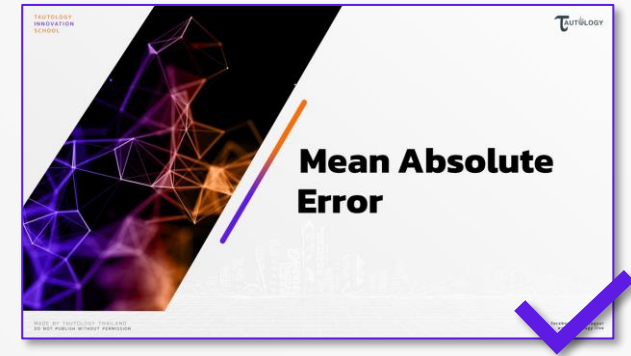
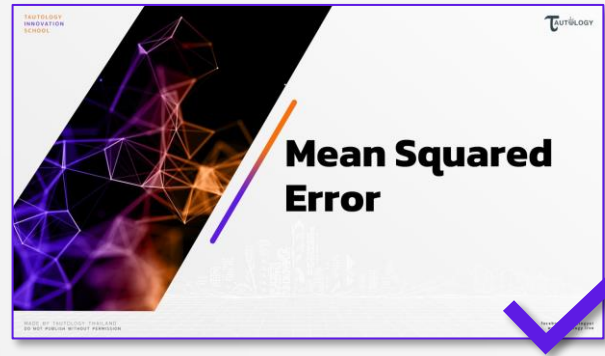
$$mape_j = \left| \frac{0.0001 - 1}{0.0001} \right| = 10,000$$

$$mape_{i \neq j} = 0.02$$

$$mape = 3000$$



# Model Evaluation for Regression



# Model Evaluation

**What is Model  
Evaluation?**



**Why need Model  
Evaluation?**



**Model Evaluation  
for Regression**

