

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DO NOT PUBLISH WITHOUT PERMISSION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

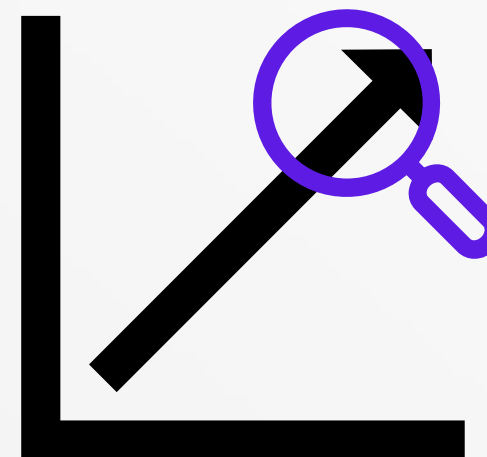
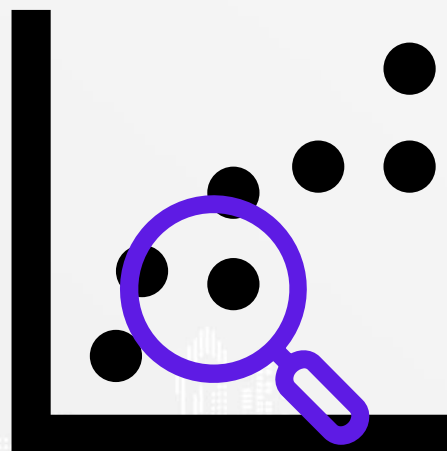
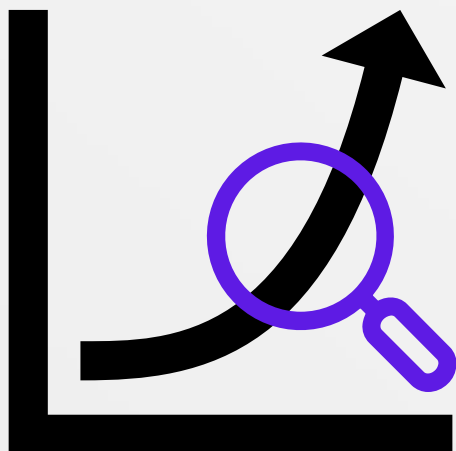
What is Model  
Evaluation?

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?**

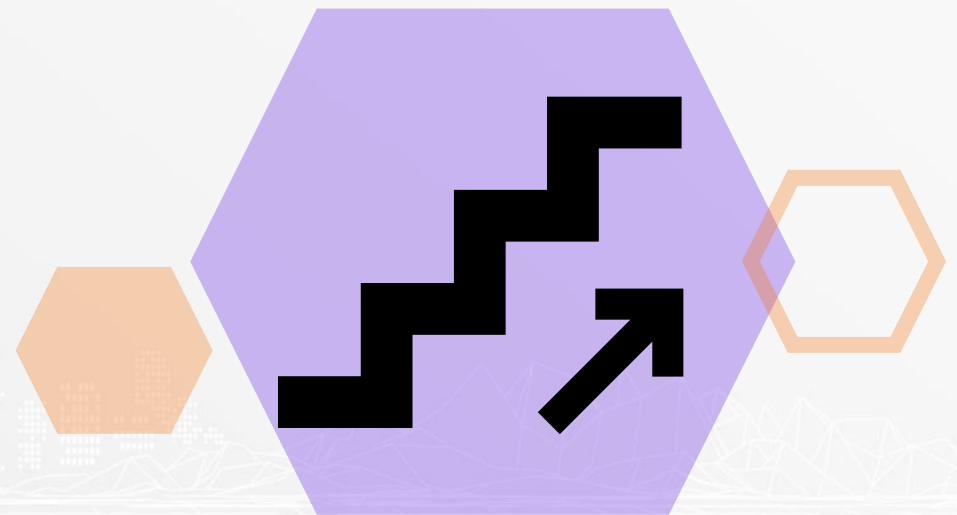


**Model Evaluation  
for Regression**



# Why need Model Evaluation?

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



# Model Evaluation

**What is Model  
Evaluation?**



**Why need Model  
Evaluation?**

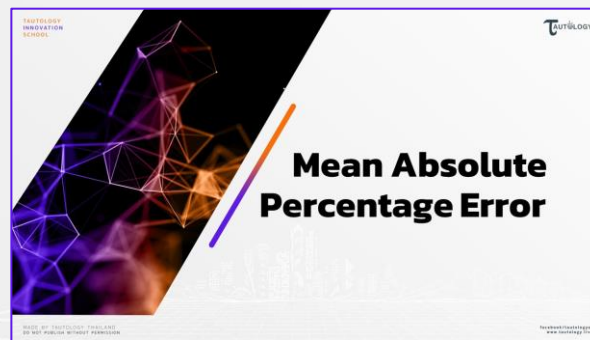
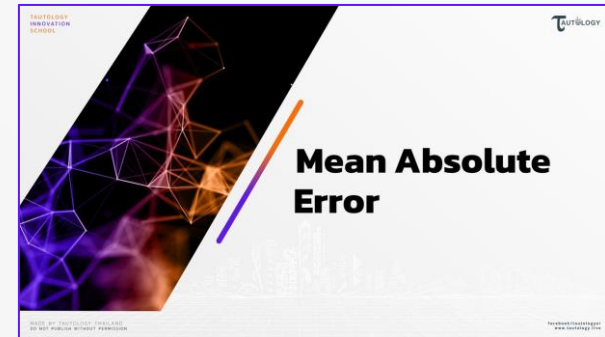
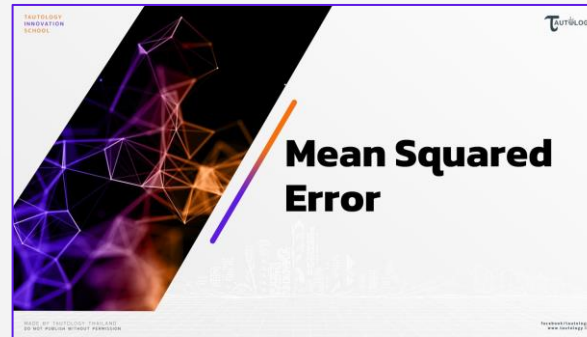


**Model Evaluation  
for Regression**





# Model Evaluation for Regression



# $R^2$ Score



# $R^2$ score

- What is  $R^2$  score?

- Formula

- Step to Calculate  $R^2$

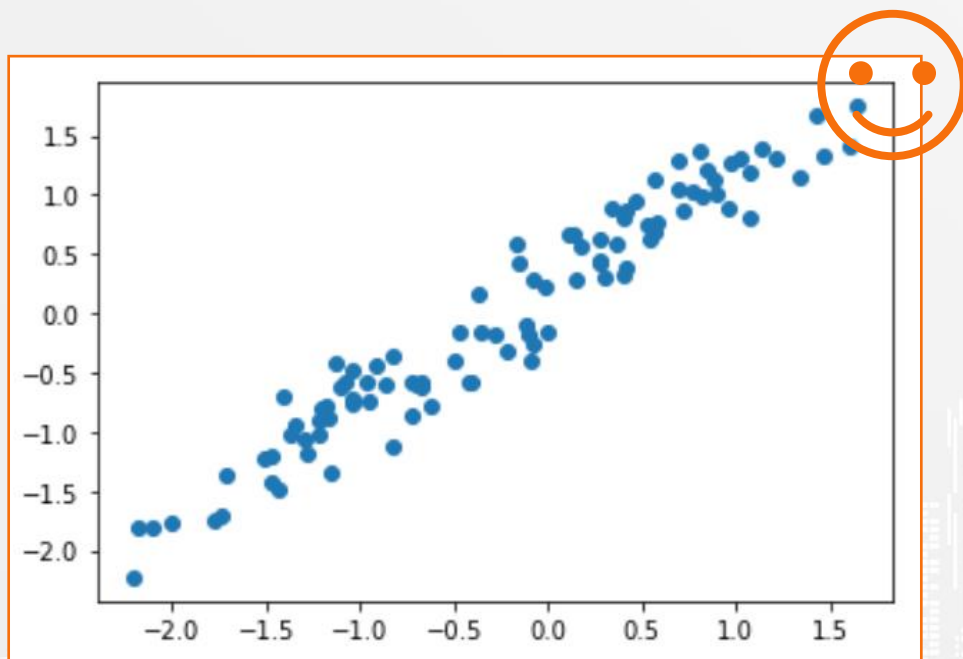
- Example

- Code

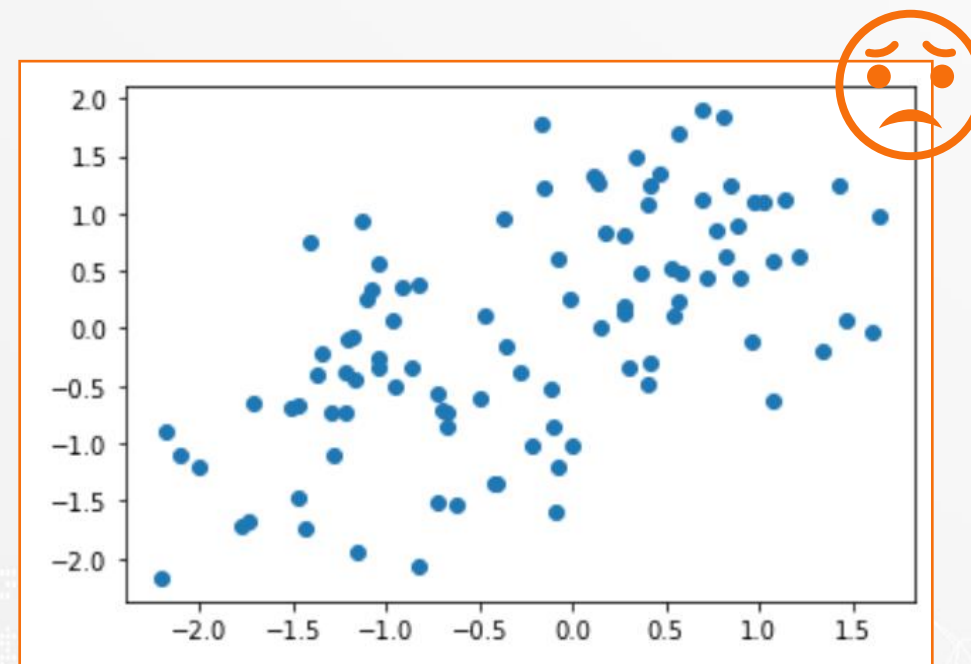
- Interpretation

# What is $R^2$ score?

$R^2$  score คือ ค่าที่บอกความสัมพันธ์ระหว่างค่าจริง และค่าพยากรณ์



**Model 1**



**Model 2**

# Formula

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

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

# 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 $\hat{y}_i$

	$y$	$\hat{y}$
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 $R^2$

	$y$	$\hat{y}$
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

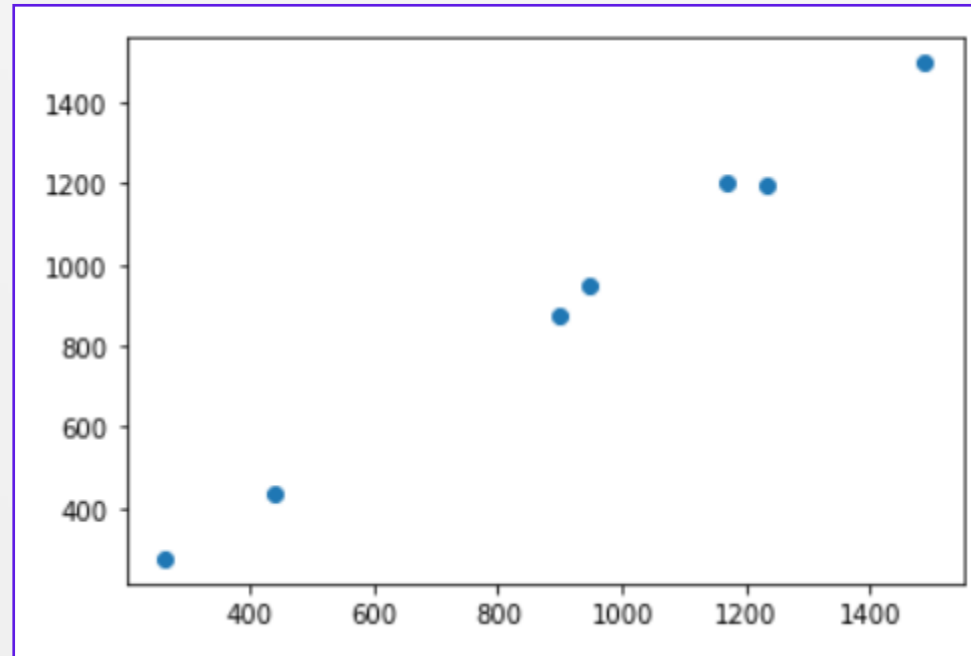
ตารางแสดงข้อมูลของราคาบ้านจริง และราคาบ้านที่พยากรณ์ได้จากโมเดล  
โดยใช้ feature ที่ใช้คือจำนวนห้องและพื้นที่ของบ้าน

# Code

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

```
0.9971801836617127
```

# Code



$$R^2 = 0.99718$$



# Interpretation

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

◆  $R^2 = 1$  ◆

↳  $\forall i, y_i = \hat{y}_i$

↳ ค่าที่ทำนาย = ค่าจริง

◆  $R^2 = 0$  ◆

↳  $\forall i, \hat{y}_i = \bar{y}$

↳ ประสิทธิภาพของ model **เท่ากับ** การพยากรณ์ด้วยค่าเฉลี่ยของ  $y$

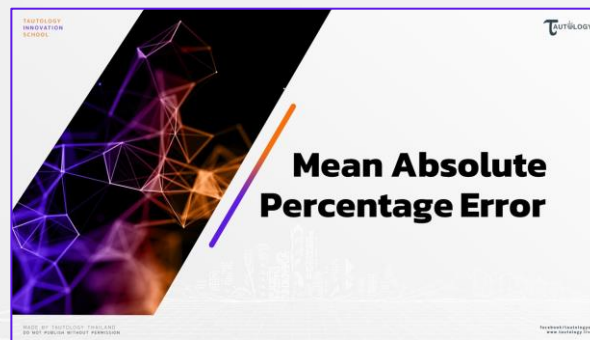
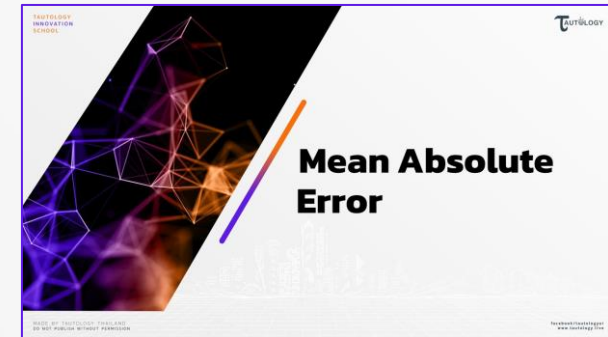
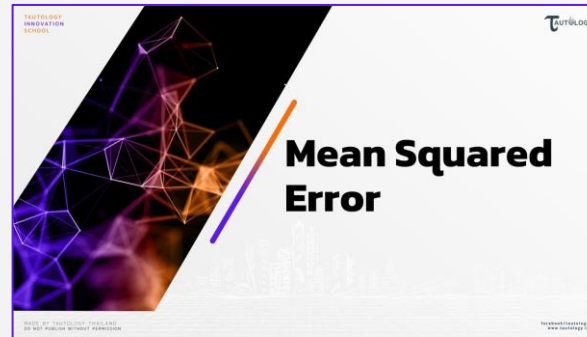
◆  $R^2 < 0$  ◆

↳ ประสิทธิภาพของ model **แย่กว่า** การพยากรณ์ด้วยค่าเฉลี่ยของ  $y$

# Interpretation

$R^2$  ที่ดีต้องมากกว่า 0  
และยิ่งใกล้ 1 จะยิ่งดี

# Model Evaluation for Regression



# Mean Squared Error

# Mean Squared Error

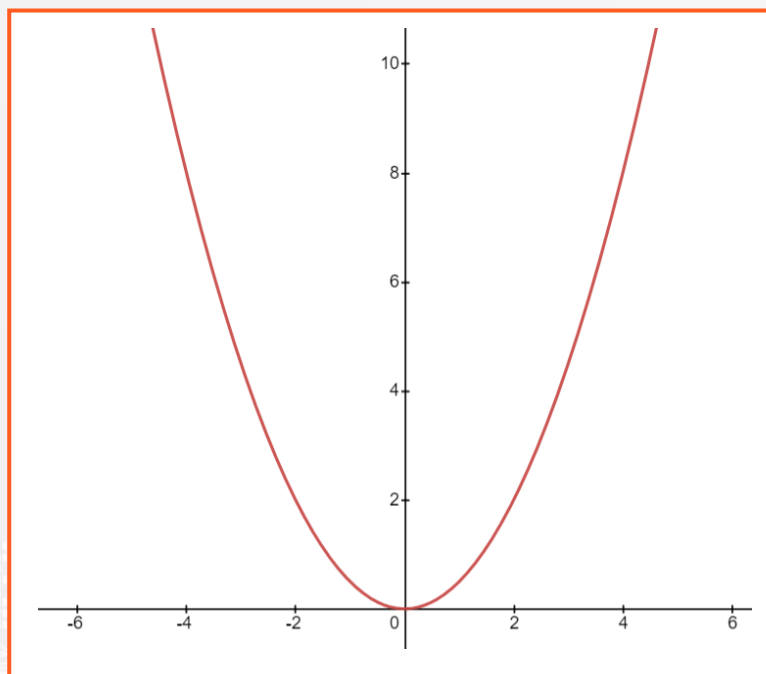
- What is Mean Squared Error?
- Formula
- Step to Calculate MSE
- Example
- Code
- Interpretation



# What is Mean Squared Error?

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

$$SE_i = \varepsilon_i^2$$



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

# Formula

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

- $y_i$  คือ target ของ 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 $\hat{y}_i$

	$y$	$\hat{y}$
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 ( $MSE$ )

	$y$	$\hat{y}$
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

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

```
462.1128826530673
```

# Interpretation

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

◆  $MSE = 400$  ◆

↳ ค่าเฉลี่ย  $(y_i - \hat{y}_i)^2 = 400$

↳ โดยเฉลี่ยแล้ว  $y_i - \hat{y}_i = \pm 20$

↳ โดยเฉลี่ยแล้ว  $y_i = \hat{y}_i \pm 20$

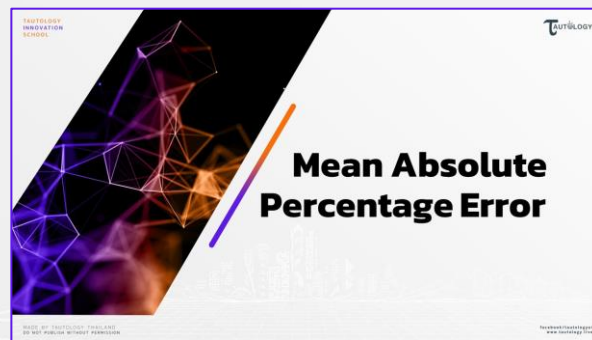
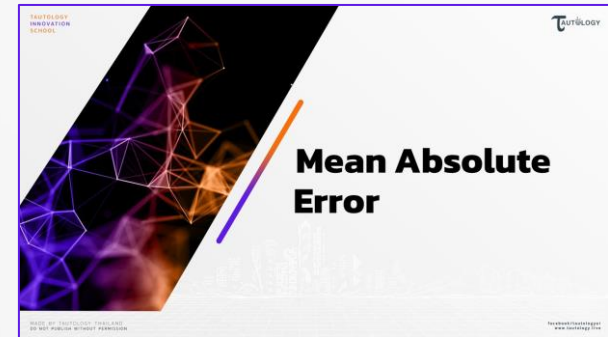
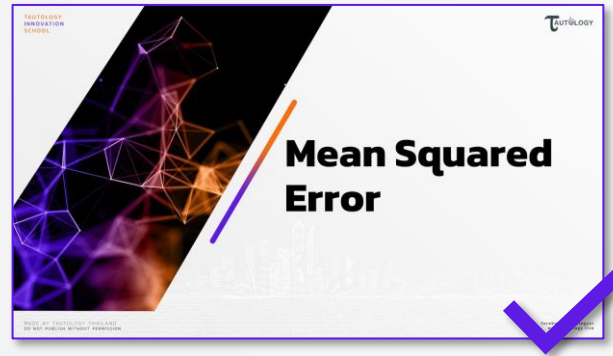
↳ ตัวอย่าง สมมติ  $\hat{y}_i = 1000$

- $y_i = \hat{y}_i \pm 20 = 1000 \pm 20$
- $y_i \in [1000 - 20, 1000 + 20]$
- $y_i \in [980, 1020]$

# Interpretation

$$y_i = \hat{y}_i \pm \sqrt{MSE}$$
$$y_i = [\hat{y}_i - \sqrt{MSE}, \hat{y}_i + \sqrt{MSE}]$$

# Model Evaluation for Regression



# Mean Absolute Error

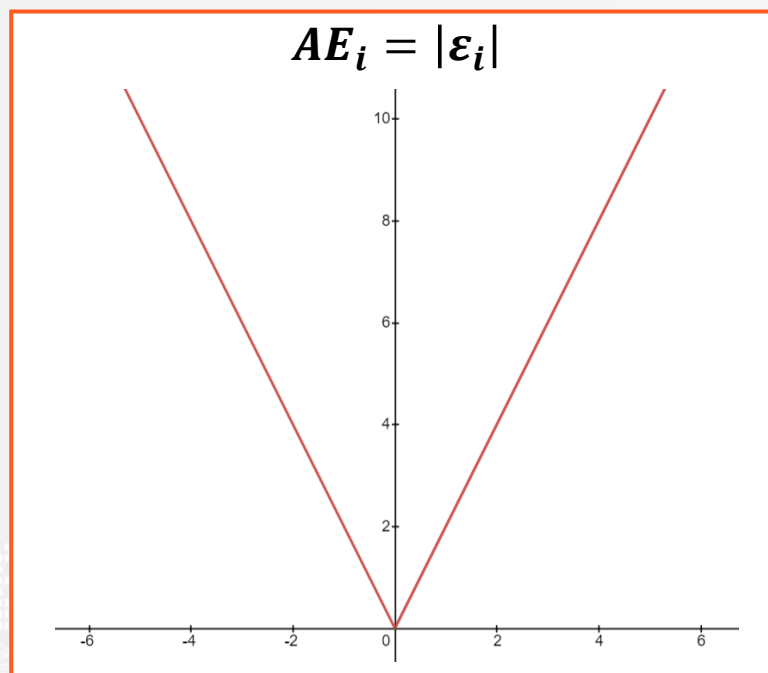


# Mean Absolute Error

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

# What is Mean Absolute Error?

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



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

# Formula

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

- $y_i$  คือ target ของ 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 $\hat{y}_i$

	$y$	$\hat{y}$
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 ( $MAE$ )

	$y$	$\hat{y}$
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

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

```
16.998086734694034
```

# Interpretation

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

◆  $MAE = 17$  ◆

↳ ค่าเฉลี่ย  $|y_i - \hat{y}_i| = 17$

↳ โดยเฉลี่ยแล้ว  $y_i - \hat{y}_i = \pm 17$

↳ โดยเฉลี่ยแล้ว  $y_i = \hat{y}_i \pm 17$

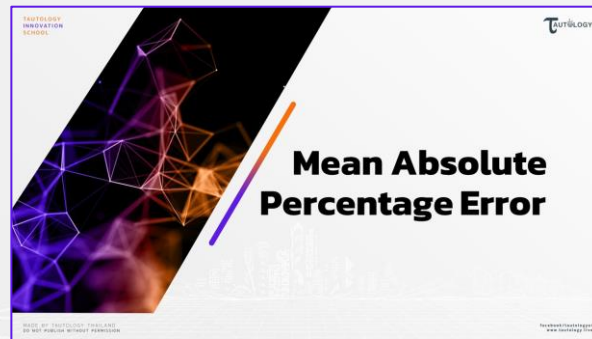
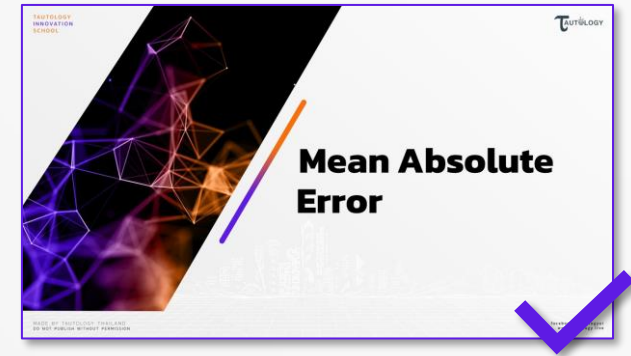
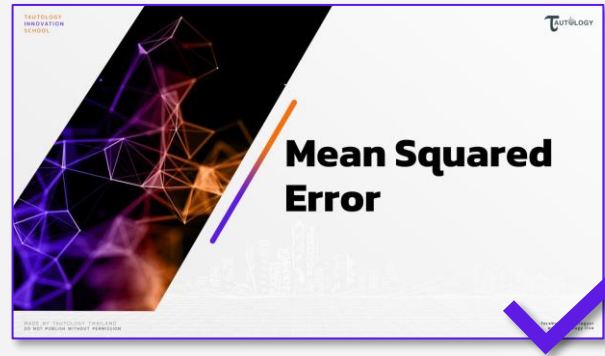
↳ ตัวอย่าง สมมติ  $\hat{y}_i = 1000$

- $y_i = \hat{y}_i \pm 17 = 1000 \pm 17$
- $y_i \in [1000 - 17, 1000 + 17]$
- $y_i \in [983, 1017]$

# Interpretation

$$y_i = \hat{y}_i \pm MAE$$
$$y_i \in [\hat{y}_i - MAE, \hat{y}_i + MAE]$$

# Model Evaluation for Regression



# Mean Absolute Percentage Error

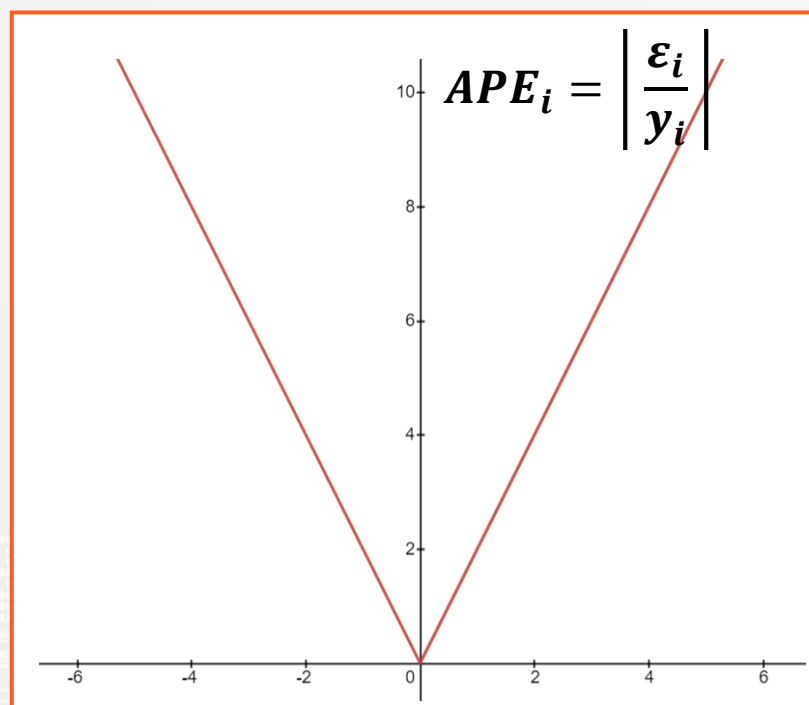


# Mean Absolute Percentage Error

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

# What is Mean Absolute Percentage Error?

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



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

# Formula

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

- $y_i$  คือ target ของ 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 $\hat{y}_i$

	$y$	$\hat{y}$
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*)

	$y$	$\hat{y}$
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

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

```
0.02076988136170835
```

# Interpretation

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

◆  $MAPE = 0.021$  ◆

↳  $MAPE = 0.021 \times 100 = 2.1\%$

↳ โดยเฉลี่ยแล้ว  $\hat{y}_i$  จะคลาดเคลื่อนอยู่  $\pm 2.1\%$

↳ ตัวอย่าง สมมติ  $\hat{y}_i = 1000$

- $y_i = \hat{y}_i \pm 2.1\% = 1000 \pm 2.1\%$
- $y_i \in [1000 - 2.1\%, 1000 + 2.1\%]$
- $y_i \in [979, 1021]$

# Interpretation

$$y_i = \hat{y}_i \pm MAPE \times 100\%$$
$$y_i \in [\hat{y}_i - MAPE(\%), \hat{y}_i + MAPE(\%)]$$

# Interpretation

★ ข้อควรระวัง ★

$$y = [0.0001, 4, 11, 6, 8]$$

$$\hat{y} = [1, 7, 8, 5, 4]$$

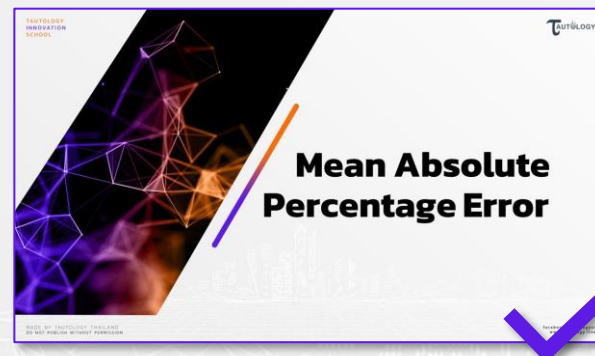
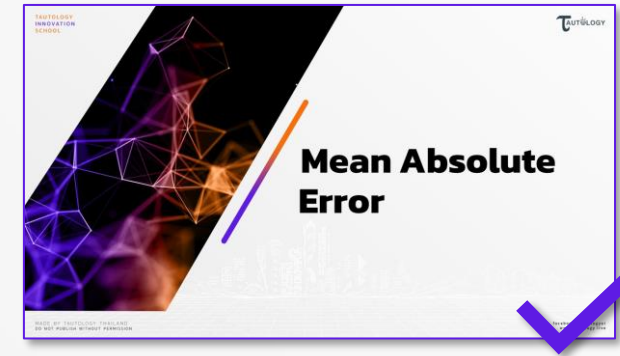
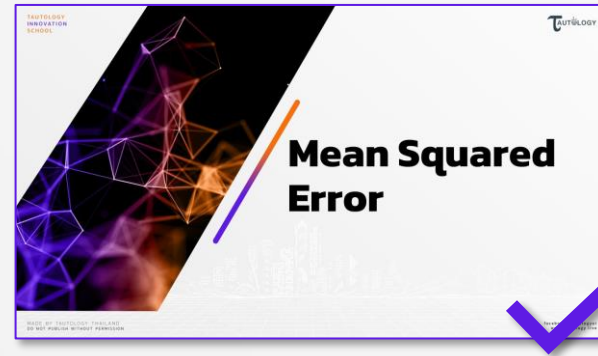
$$\begin{aligned} MAPE &= \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \\ &= \frac{1}{5} \left( \left| \frac{0.0001 - 1}{0.0001} \right| + \left| \frac{4 - 7}{4} \right| + \left| \frac{11 - 8}{11} \right| + \left| \frac{6 - 5}{6} \right| + \left| \frac{8 - 4}{8} \right| \right) \\ &= \frac{1}{5} (9999 + 0.75 + 0.2727 + 0.1667 + 0.5) \\ &= 2000.1379 = 200013.79\% \end{aligned}$$

# Interpretation

✱ ข้อควรระวัง ✱

สำหรับการใช้ *MAPE* ควรระวังว่า  
“ค่า  $y_i$  ไม่ควรใกล้เคียง 0”

# Model Evaluation for Regression





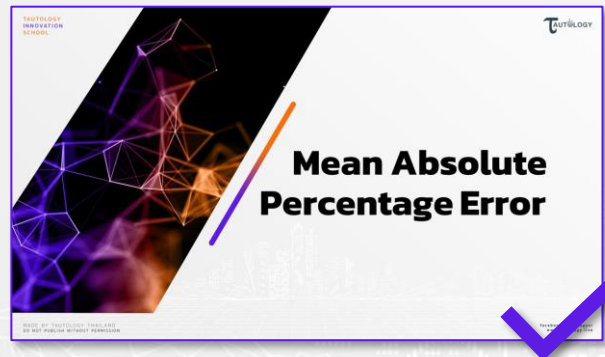
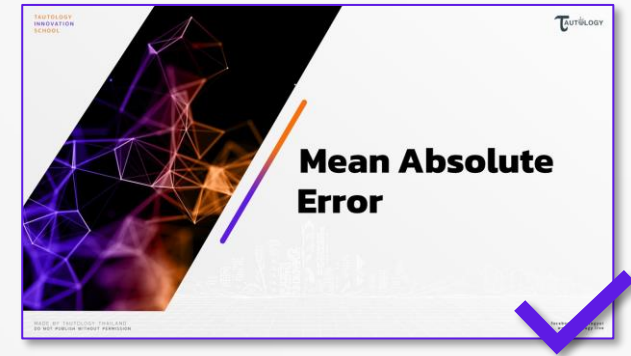
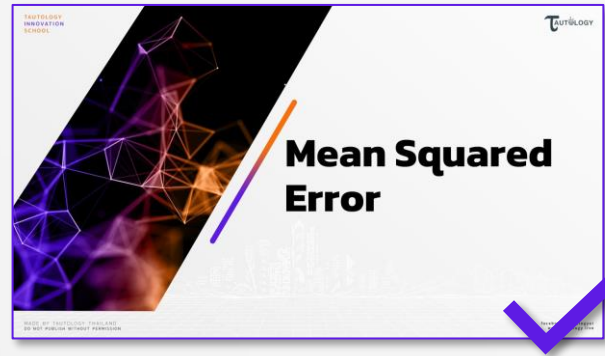
# Conclusion



# Conclusion

Name	Formula	How to use
$R^2$	$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$	ใช้บอกประสิทธิภาพของ model เมื่อเทียบกับการพยากรณ์ด้วยค่าเฉลี่ยของ $y$
$MSE$	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$	ใช้ในบริบทของ Cost function
$MAE$	$MAE = \frac{1}{n} \sum_{i=1}^n  y_i - \hat{y}_i $	ใช้บอกระยะความคลาดเคลื่อนของการพยากรณ์
$MAPE$	$MAPE = \frac{1}{n} \sum_{i=1}^n \left  \frac{y_i - \hat{y}_i}{y_i} \right $	ใช้บอกความคลาดเคลื่อนของการพยากรณ์ในหน่วย %

# Model Evaluation for Regression



# Model Evaluation

**What is Model  
Evaluation?**



**Why need Model  
Evaluation?**



**Model Evaluation  
for Regression**

