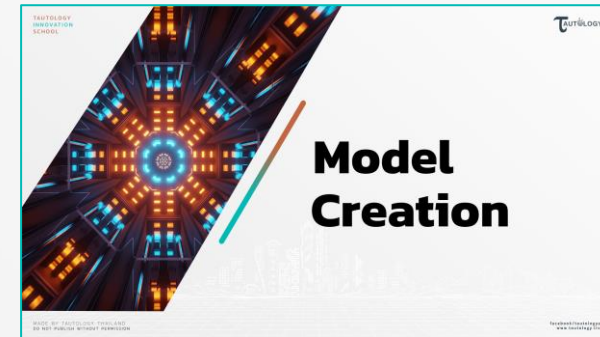


K-NEAREST NEIGHBORS

BY TAUTOLOGY

K-Nearest Neighbors



Introduction

Introduction

What is KNN?

Data for KNN

Pros & Cons

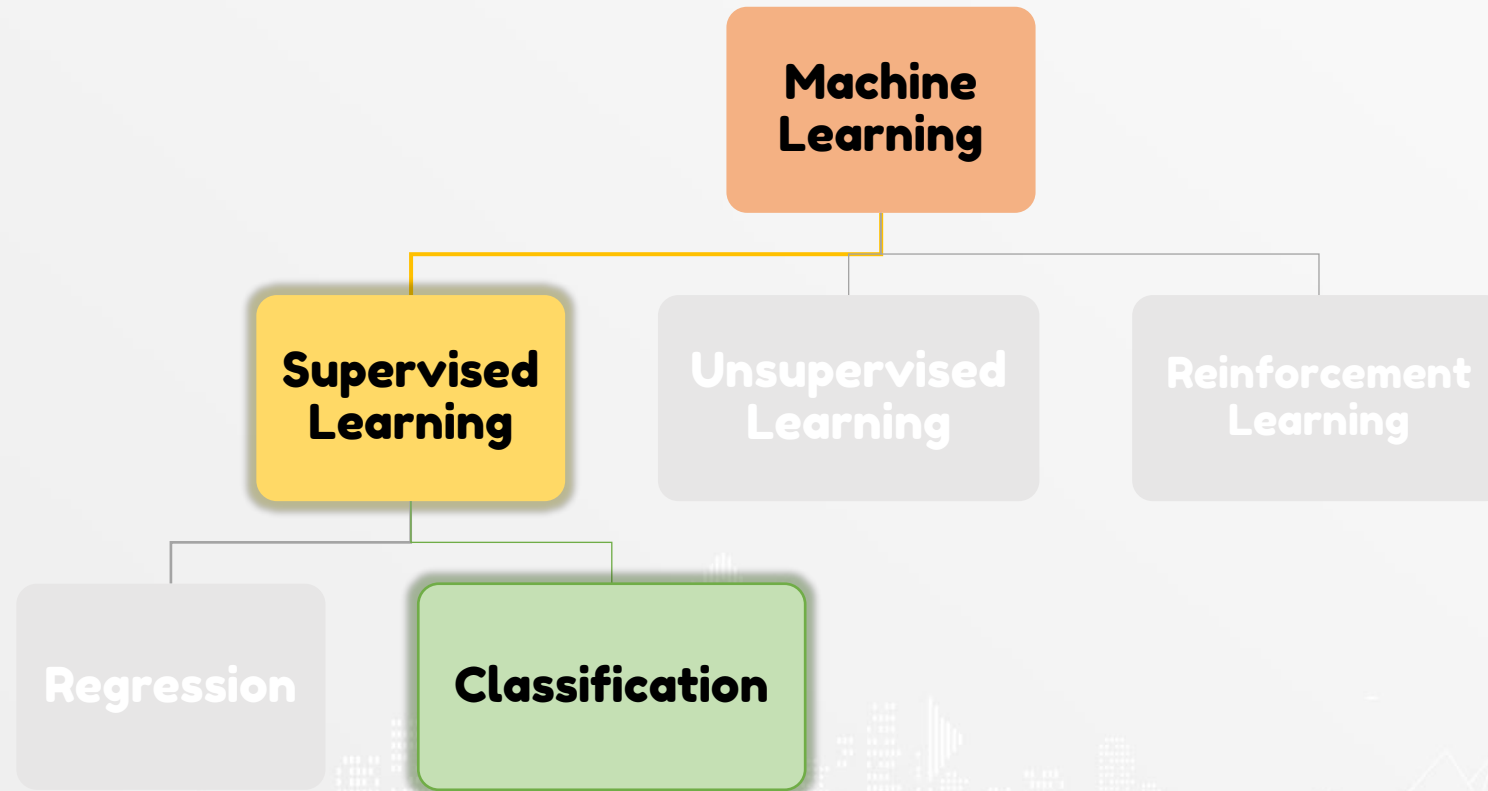
Real World
Application

What is KNN?

k Nearest Neighbors (KNN) เป็นหนึ่งใน algorithm ประเภท supervised learning ที่ใช้สำหรับแก้ปัญหา classification โดยมีหลักการทำงานคือ การระบุประเภทของข้อมูลตัวที่สนใจโดยพิจารณาข้อมูลเพื่อนบ้านที่ใกล้ที่สุด k ตัว

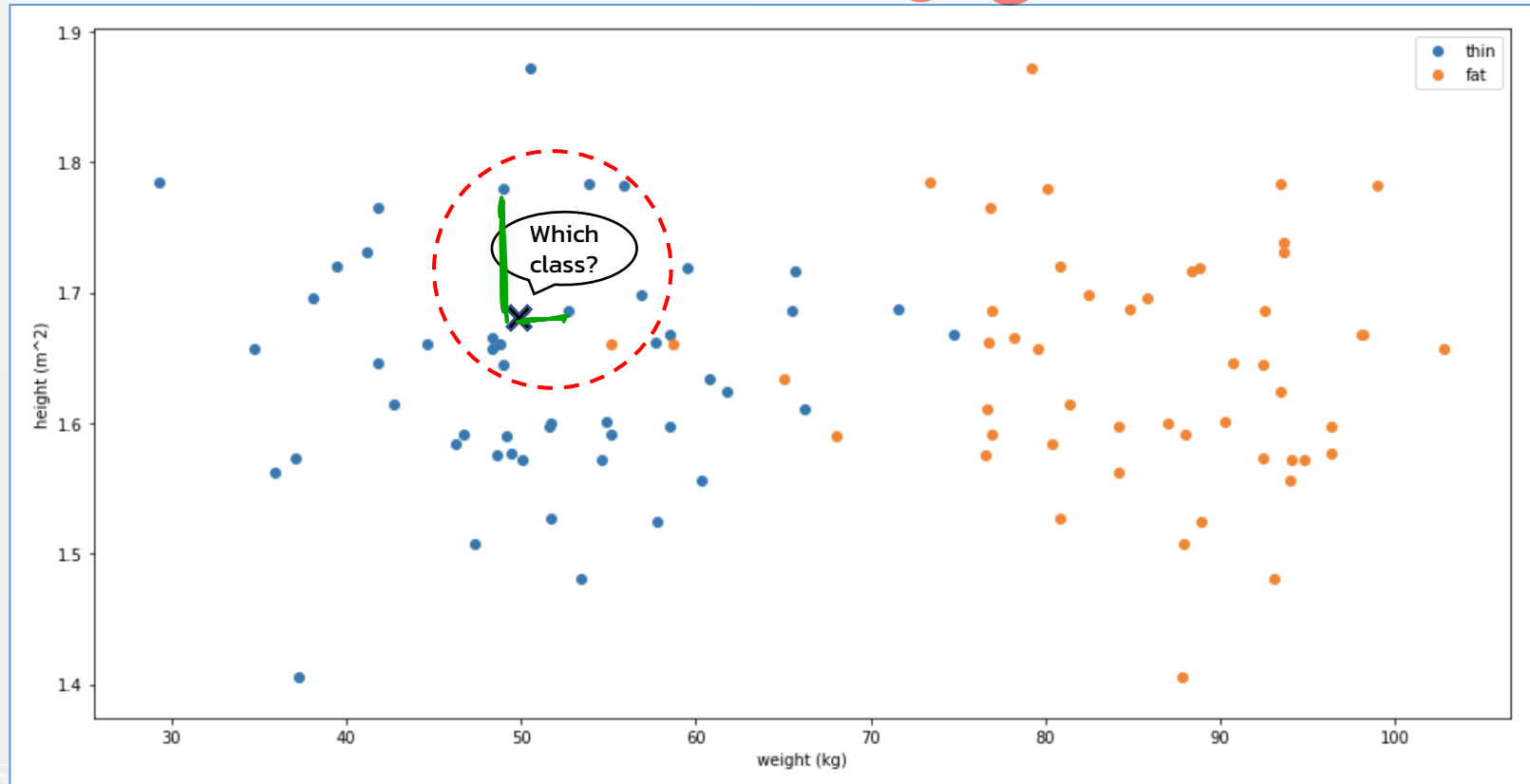
regression

What is KNN?



What is KNN?

$$\sqrt{(x_1 - \square)^2 + (x_2 - \triangle)^2 + (x_3 - \star)^2 + (x_4 - \diamond)^2}$$



Introduction

What is KNN?



Data for KNN



Pros & Cons

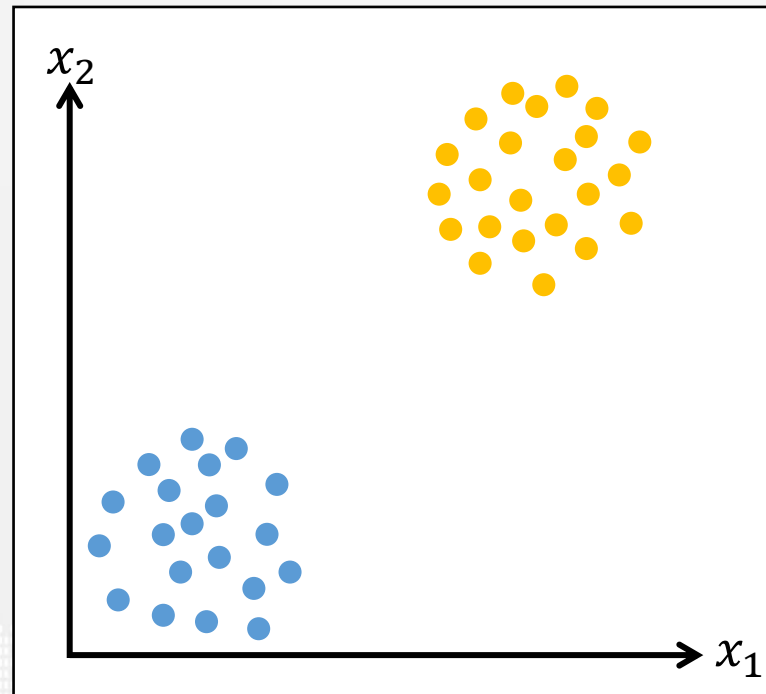


**Real World
Application**



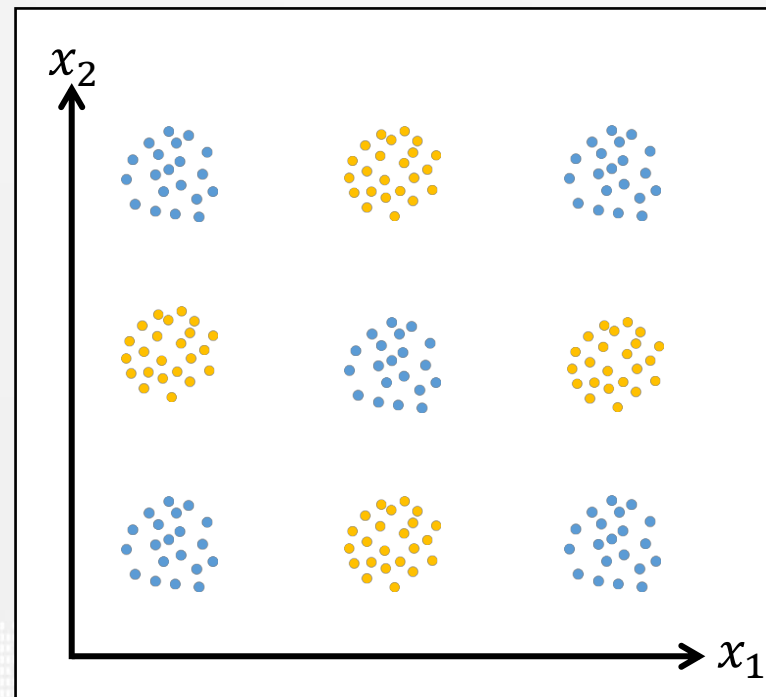
Data for KNN

ตัวอย่างของข้อมูลที่เหมาะสมกับ KNN



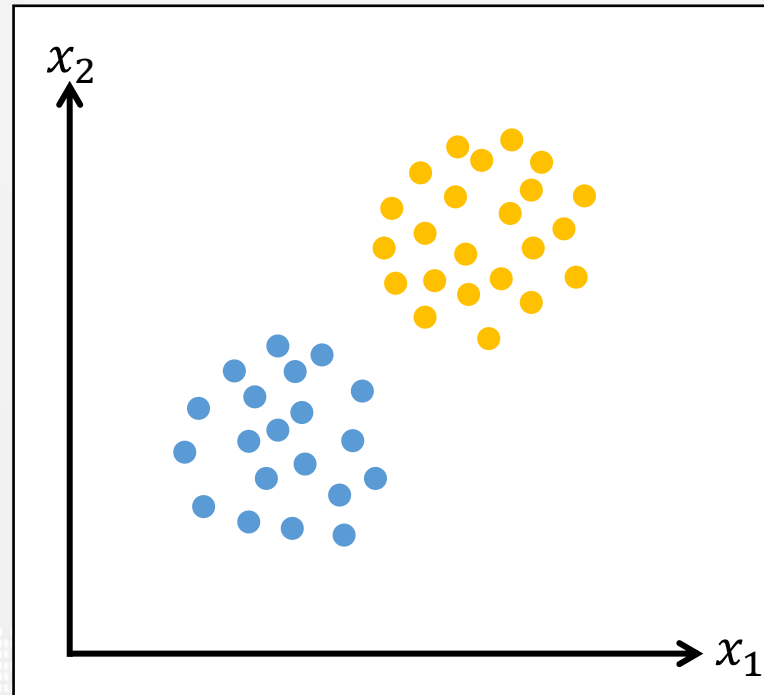
Data for KNN

ตัวอย่างของข้อมูลที่เหมาะสมกับ KNN



Data for KNN

ตัวอย่างของข้อมูลที่ไม่เหมาะกับ KNN



Introduction

What is KNN?



Data for KNN



Pros & Cons



**Real World
Application**



Pros & Cons

ข้อดี

- เป็น algorithm ที่เรียบง่าย & ง่ายต่อการทำความเข้าใจ

LR + LoR + NN + DL

\Rightarrow มันพิจารณา. ลำดับของ feature ได้ (ผ่าน w)

(T + RT \Rightarrow มีการเลือก feature

\Rightarrow algorithm พวกนี้มันได้ผลลัพธ์
feature selection เบื้องต้น

ข้อเสีย

- Curse of dimensionality

(ถ้า feature เยอะ = model จะ งง)

โชน feature ไปเพื่อ \Rightarrow ~~งง~~

\Rightarrow ใช้หมด

ข้อจำกัด

- การเลือกค่า k

ถ้า data ซ้ำซ้อน, การมี feature เยอะ ๆ มีโอกาสสูงมาก ที่ KNN จะผิดพลาด



Introduction

What is KNN?



Data for KNN



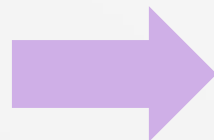
Pros & Cons



**Real World
Application**



Real World Application

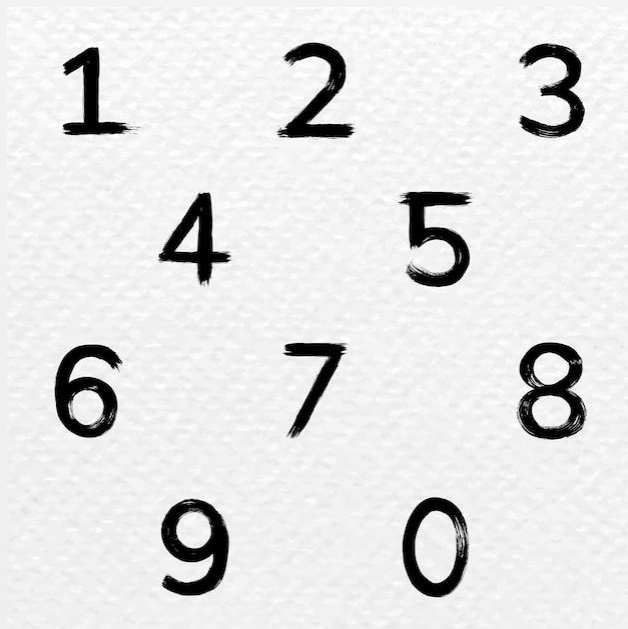


การจำแนกผู้ป่วยโรคมะเร็งเต้านม

โดยใช้พิจารณาจาก id ของผู้ป่วย ความ
สม่ำเสมอของรูปร่างเซลล์ ความสม่ำเสมอ
ของขนาดเซลล์ ความหนาของก้อนเนื้อ
เป็นต้น ซึ่งเป็นข้อมูลผู้ป่วยมะเร็งเต้านม
ของ Wisconsin-Madison

อ้างอิง : [2000, Sarkar & Leong] Application of K-nearest
neighbors algorithm on breast cancer diagnosis problem

Real World Application



จำแนกตัวเลข 0-9 จากลายมือ
โดยพิจารณาจากรูปภาพการเขียน
ตัวเลข 0-9 ของผู้เขียน 750 คน

อ้างอิง : [2014, Babu et al.] Handwritten Digit Recognition Using K-Nearest Neighbour Classifier

Introduction

What is KNN?



Data for KNN



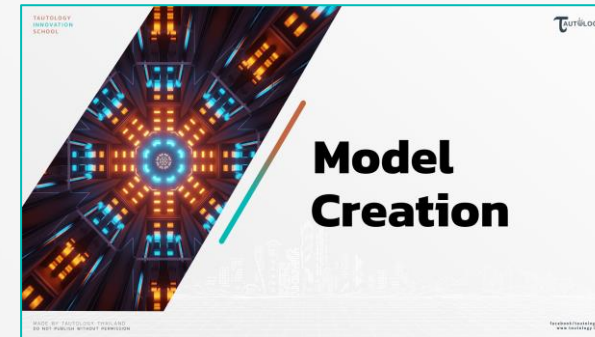
Pros & Cons



**Real World
Application**

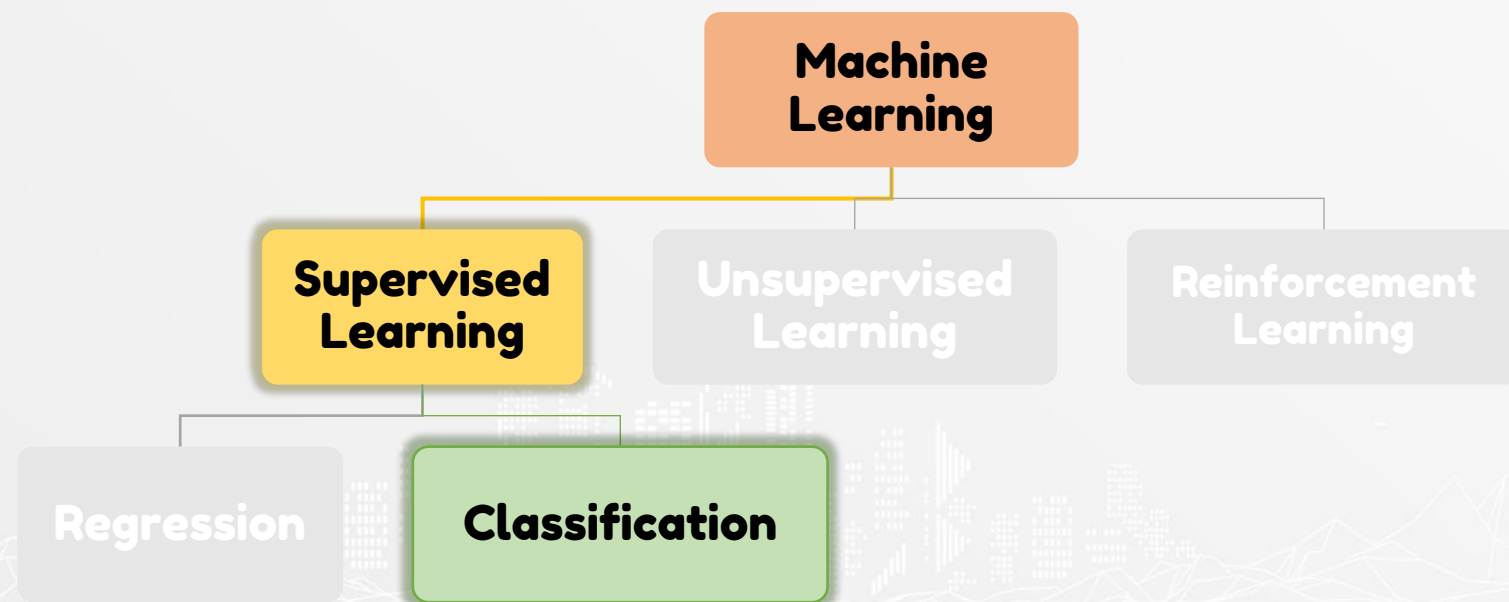


K-Nearest Neighbors



KNN

KNN 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**

(feature อย่างใด อย่างหนึ่ง)

Model Creation

Assumption



Real Face of the
Model



How to Create Model
(Math)



How to Create Model
(Code)

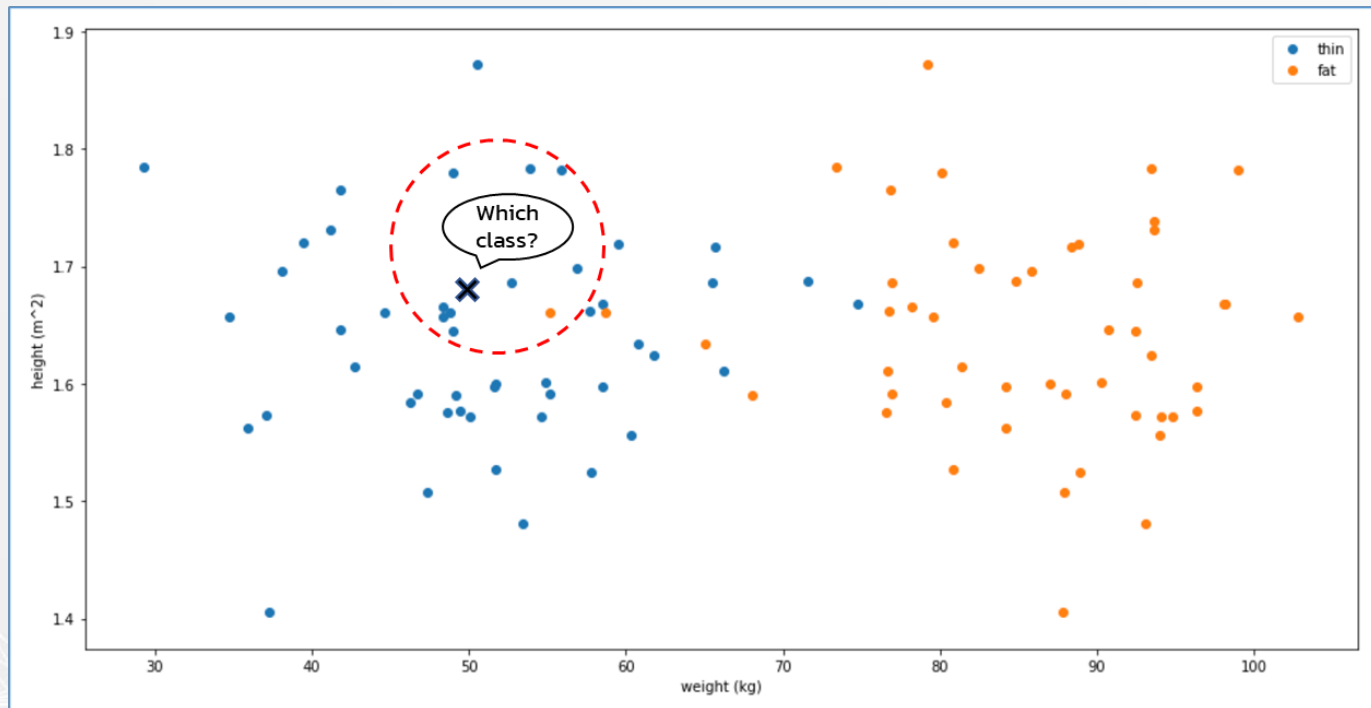


Further Reading



Real Face of the Model

KNN คือ การจำแนกประเภทของข้อมูล โดยใช้ความคล้ายคลึงกันของคุณลักษณะ (feature similarity) ในการจำแนก



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)

- Calculation Example
- How to Choose K

Calculation Example $K = \sqrt{n}$

- ☐ Step 1 : เลือก K
- ☐ Step 2 : วัดระยะห่างระหว่างข้อมูลของแต่ละ sample
- ☐ Step 3 : เรียงลำดับระยะทางที่ได้จากขั้นตอนที่ 2 จากน้อยไปมาก
- ☐ Step 4 : หาผลลัพธ์จากเพื่อนบ้านที่ใกล้ที่สุด K ค่า และทำการ
Majority Vote (ดูเสียงส่วนใหญ่)

Calculation Example

ตัวอย่างการคำนวณ KNN

weight	temperature	covid-19
45	39.9	yes
51	36.5	no
60	36.6	no
42	38.7	yes
44	36.5	yes

ตารางแสดงผู้ป่วยโรค covid-19

Calculation Example

☑ **Step 1**: เลือก K

เลือก $k = 3$

Calculation Example

☑ **Step 2** : วัดระยะห่างระหว่างข้อมูลของแต่ละ sample

	Weight (kg)	Temperature (°C)	covid-19
1	45	39.9	yes
2	51	36.5	no
3	60	36.6	no
4	42	38.7	yes
5	44	36.5	yes



Distance
4.17
3.04
12
6.23
4.03

X

weight = 48
temperature = 37.0

weight = 48

temperature = 37.0

$$\begin{aligned}\text{ระยะห่าง X 1} &= \sqrt{(48 - 45)^2 + (37.0 - 39.9)^2} \\ &= \sqrt{3^2 + (-2.9)^2} \\ &= \sqrt{9 + 8.41} = \sqrt{17.41}\end{aligned}$$

$$\begin{aligned}\text{ระยะห่าง X 2} &= \sqrt{(48 - 51)^2 + (37 - 36.5)^2} \\ &= \sqrt{(-3)^2 + (0.5)^2} = \sqrt{9.25}\end{aligned}$$

Calculation Example

☑ **Step 3** : เรียงลำดับระยะทางที่ได้จากขั้นตอนที่ 2 จากน้อยไปมาก

Distance	Distance
4.17	3.04
3.04	4.03
12	4.17
6.23	6.23
4.03	12

Calculation Example

☑ **Step 3** : เรียงลำดับระยะทางที่ได้จากขั้นตอนที่ 2 จากน้อยไปมาก

Weight (kg)	Temperature (°C)	covid-19	Distance
45	39.9	yes	4.17
51	36.5	no	3.04
60	36.6	no	12
42	38.7	yes	6.23
44	36.5	yes	4.03



Weight (kg)	Temperature (°C)	covid-19	Distance
51	36.5	no	3.04
44	36.5	yes	4.03
45	39.9	yes	4.17
42	38.7	yes	6.23
60	36.6	no	12

Calculation Example

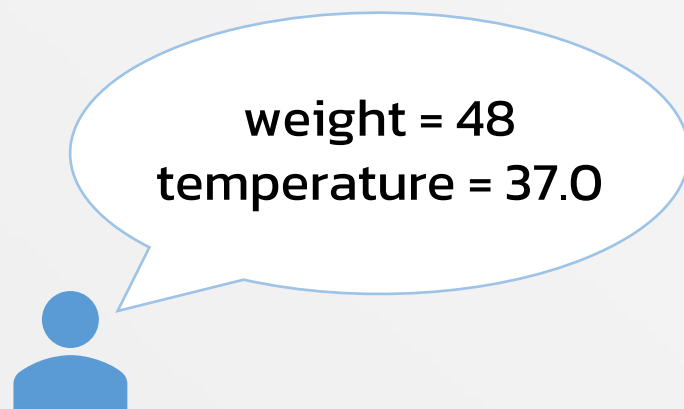
☑ **Step 4**: หาผลลัพธ์จากเพื่อนบ้านที่ใกล้ที่สุด K ค่า และทำการ **Majority Vote**

Weight (kg)	Temperature (°C)	covid-19	Distance
51	36.5	no	3.04
44	36.5	yes	4.03
45	39.9	yes	4.17
42	38.7	yes	6.23
60	36.6	no	12

K = 3

Calculation Example

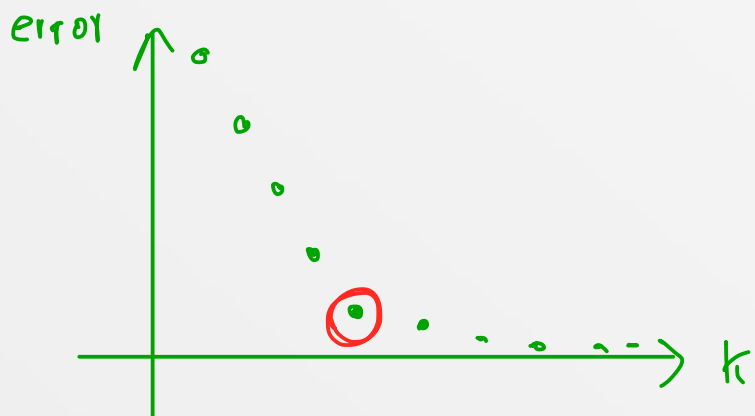
☑ **Step 4** : หาผลลัพธ์จากเพื่อนบ้านที่ใกล้ที่สุด K ค่า และทำการ Majority Vote



How to Create Model (Math)

✓ • Calculation Example

- How to Choose K



① $k = \sqrt{n}$

② $k \leftarrow$ elbow method

③ $k \leftarrow$ Domain Expert + Experiment

dataset \rightarrow train | valid | test
or k that minimizes error

How to Choose K

$$K \approx \sqrt{n}$$

How to Create Model (Math)

- ✓• **Calculation Example**
- ✓• **How to Choose K**

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

weight	temperature	covid-19
45	39.9	yes
51	36.5	no
60	36.6	no
42	38.7	yes
44	36.5	yes

ตารางแสดงผู้ป่วยโรค covid-19

How to Create Model (Code)

- Code สำหรับสร้าง model จากข้อมูลของเราโดยที่

$$X = \begin{bmatrix} 45 & 39.9 \\ 51 & 36.5 \\ 60 & 36.6 \\ 42 & 38.7 \\ 44 & 36.5 \end{bmatrix}, \quad y = \begin{bmatrix} yes \\ no \\ no \\ yes \\ yes \end{bmatrix}$$

```
1 clf = KNeighborsClassifier(n_neighbors=3)
2 clf.fit(X, y)
```

```
KNeighborsClassifier(n_neighbors=3)
```

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

- Elbow Method
- Cross Validation

Model Creation

Assumption



**Real Face of the
Model**



**How to Create Model
(Math)**



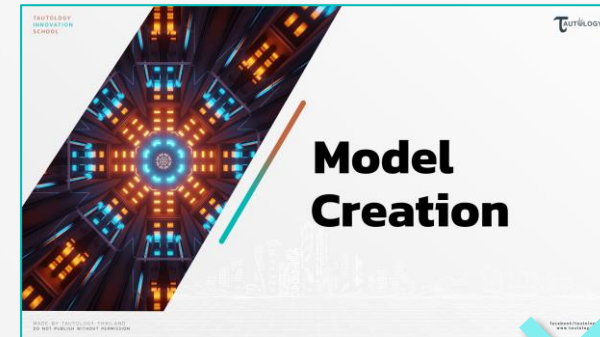
**How to Create Model
(Code)**



Further Reading



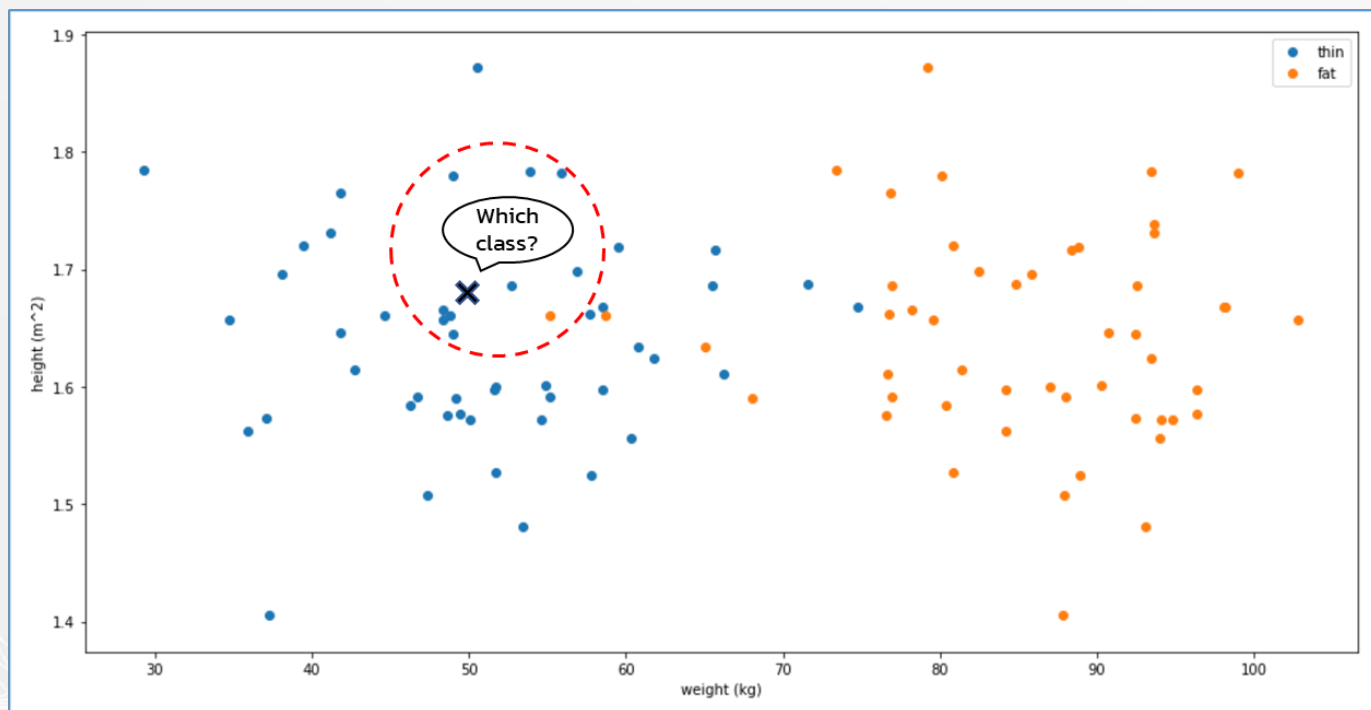
K-Nearest Neighbors



Prediction

Prediction

KNN คือ การจำแนกประเภทของข้อมูล โดยใช้ความคล้ายคลึงกันของคุณลักษณะ (feature similarity) ในการจำแนก



Prediction

1-Sample

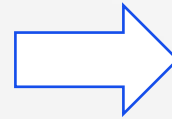
Multi-Sample

Code

1-Sample

ตัวอย่างการคำนวณ \hat{y}

weight	temperature
44	37.0



\hat{y}
?

frzin อยู่ข้าง 7
ของคนเข้ามาหัด 7 = ชะนำง

1-Sample

Weight (kg)	Temperature (°C)	covid-19
44	36.5	yes
42	38.7	yes
45	39.9	yes
51	36.5	no
60	36.6	no

Distance
0.5
2.62
3.07
7.02
16.01

weight = 44
temperature = 37.0

$K = 3$

1-Sample



Prediction

1-Sample



Multi-Sample



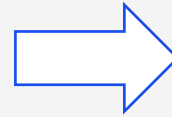
Code



Multi-Sample

ตัวอย่างการคำนวณ \hat{y}

weight	temperature
44	37.0
49	38.0
45	37.2
52	38.7



\hat{y}
?
?
?
?

Multi-Sample

Weight (kg)	Temperature (°C)	covid-19
44	36.5	yes
42	38.7	yes
45	39.9	yes
51	36.5	no
60	36.6	no

Distance
0.5
2.62
3.07
7.02
16.01

weight = 44
temperature = 37.0

$K = 3$

Multi-Sample

weight	temperature
44	37.0
49	38.0
45	37.2
52	38.7



\hat{y}
yes
?
?
?

Multi-Sample

Weight (kg)	Temperature (°C)	covid-19
51	36.5	no
45	39.9	yes
44	36.5	yes
42	38.7	yes
60	36.6	no

Distance
2.5
4.43
5.22
7.03
11.09

weight = 49
temperature = 38.0

$K = 3$

Multi-Sample

weight	temperature
44	37.0
49	38.0
45	37.2
52	38.7



\hat{y}
yes
yes
?
?

Multi-Sample

weight	temperature
44	37.0
49	38.0
45	37.2
52	38.7



\hat{y}
yes
yes
no
no

Prediction

1-Sample



Multi-Sample



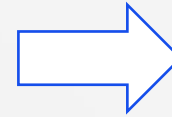
Code



Code

ตัวอย่าง code สำหรับการคำนวณ \hat{y}

weight	temperature
44	38.0
49	37.0
45	37.2
52	38.7



\hat{y}
?
?
?
?

Code

- Code สำหรับสร้าง model จากข้อมูลของเราโดยที่

$$X = \begin{bmatrix} 44 & 38.0 \\ 49 & 37.0 \\ 45 & 37.2 \\ 52 & 38.7 \end{bmatrix}$$

```
1 clf.predict(X)
```

```
array(['yes', 'yes', 'no', 'no'], dtype=object)
```

Code

ดังนั้น เราจะได้ \hat{y} สำหรับข้อมูลชุดนี้คือ

weight	temperature
44	38.0
49	37.0
45	37.2
52	38.7



\hat{y}
yes
yes
no
no

Code



Code for this section



Open File
Model Creation.ipynb

Prediction

1-Sample



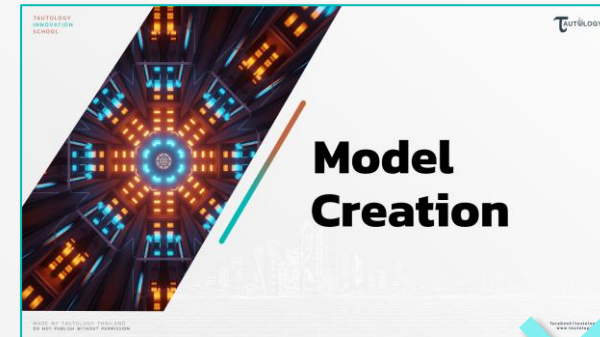
Multi-Sample



Code



K-Nearest Neighbors



Workshop

AI in Diagnosing Alzheimer's

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

Abstract

สร้าง model เพื่อวินิจฉัยโรคอัลไซเมอร์ โดยพิจารณาจากภาพ MRI ของสมอง



Why this project important?



- สามารถสร้างระบบสำหรับตรวจโรคอัลไซเมอร์ที่ทำงานได้ตลอด 24 ชั่วโมง
- สามารถนำไปต่อยอดกับการวินิจฉัยโรคอื่น ๆ
- สามารถใช้เป็นพื้นฐานสำหรับการแพทย์ทางไกล
- สามารถนำไปประยุกต์ใช้กับงานที่มีลักษณะใกล้เคียงกันได้ เช่น การตรวจจับอาวุธในสนามบิน

Who this project is for?

- ✦ ผู้บริหารโรงพยาบาล
- ✦ บุคลากรทางการแพทย์
- ✦ นักวิเคราะห์ข้อมูล



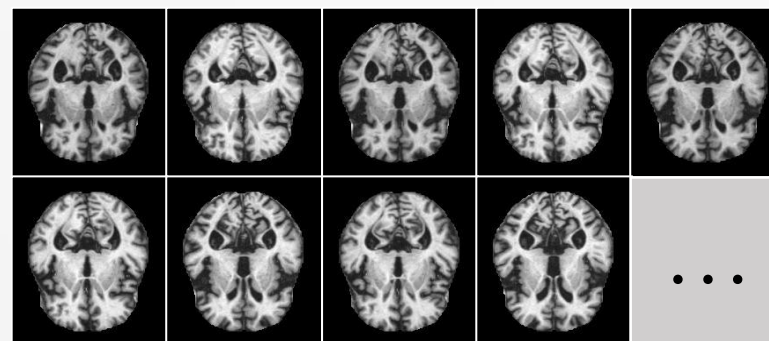
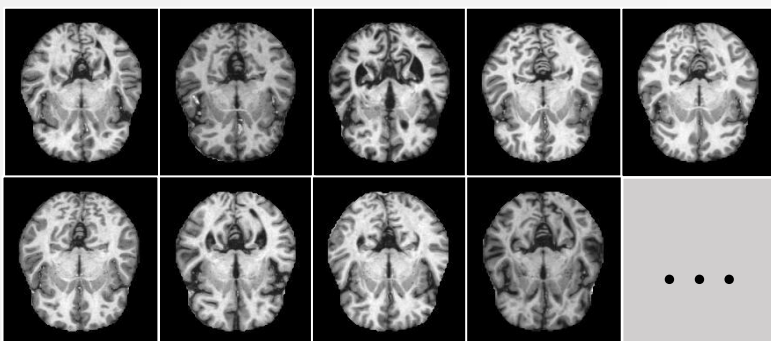
Alzheimer Dataset



<https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images>

Alzheimer Dataset

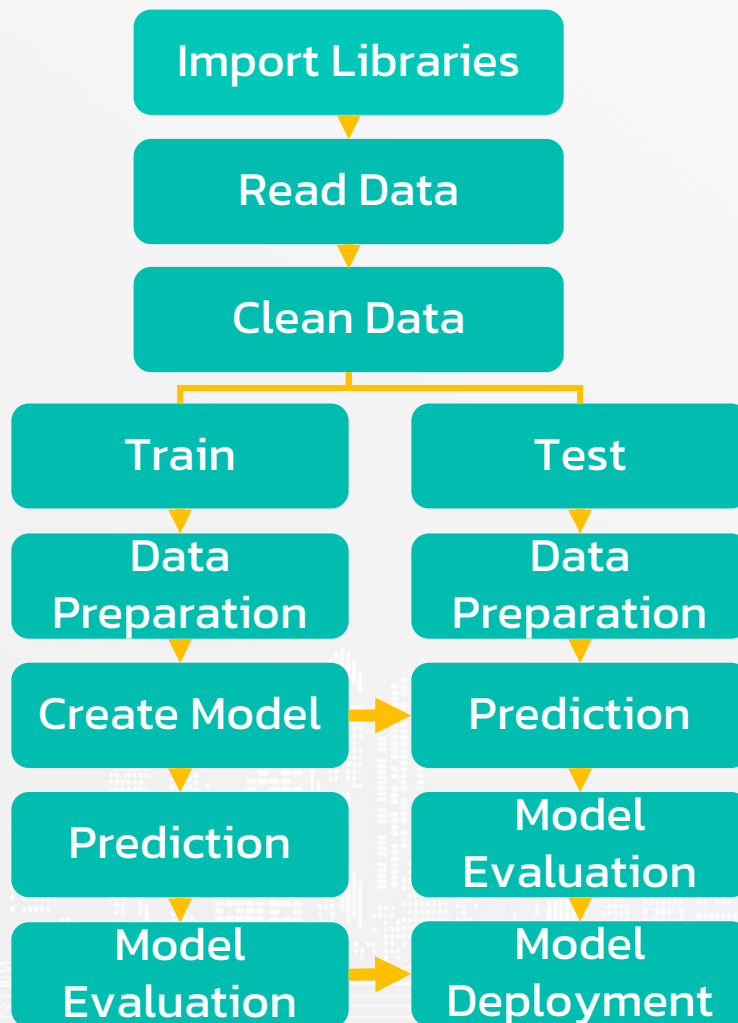
Feature



Target

- การเป็นโรคอัลไซเมอร์ (1 = เป็น, 0 = ไม่เป็น)

What we learn from this project?



point: ขั้นตอนคือการดึง
dataset ในชีวิต
(ณ.วันที่เรามีความรู้)

ทำด.เข้าใจ
dataset

choose
Algo —

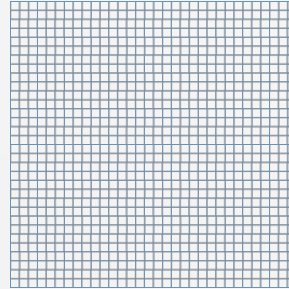
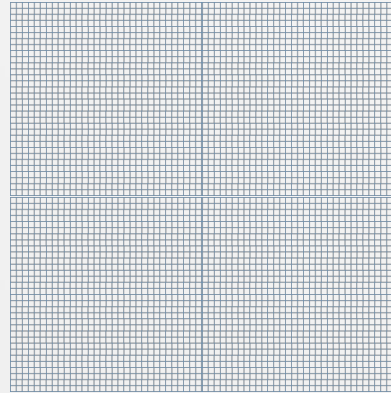
config
Hyper —

Image to CSV

```
1 classes = ['MildDemented', 'ModerateDemented', 'NonDemented', 'VeryMildDemented']
```

```
1 width = 176
2 height = 208
3
4 X = np.empty([0, width*height])
5 y = np.empty([0, 1])
6
7 for _class in tqdm(classes):
8     img_path = glob('dataset/' + _class + '/*')
9     for path in tqdm(img_path):
10         img = Image.open(path)
11         img = img.resize([width, height])
12         img = np.array(img)
13         img = img.reshape(1, -1)
14         X = np.vstack([X, img])
15         if _class == 'NonDemented':
16             y = np.vstack([y, False])
17         else:
18             y = np.vstack([y, True])
```

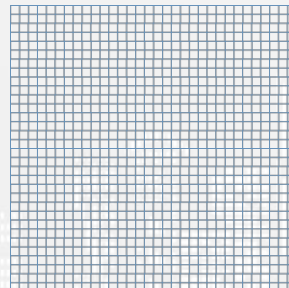
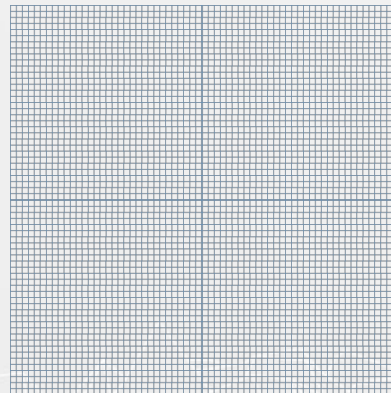
Image to CSV



176x208



เป็น



176x208



ไม่เป็น

Image to CSV

x_1	x_2	x_3	...	x_{36608}	y
0.0	0.0	0.0	...	0.0	1.0
0.0	0.0	0.0	...	0.0	1.0
0.0	0.0	0.0	...	0.0	1.0
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
0.0	0.0	0.0	...	0.0	1.0

X

y

Image to CSV

```
1 columns = [f'pixel_{i}' for i in range(width*height)]  
2  
3 data = pd.DataFrame(X, columns=columns)  
4 data['label'] = y  
5  
6 data.to_csv('alzheimer_dataset.csv', index=False)
```

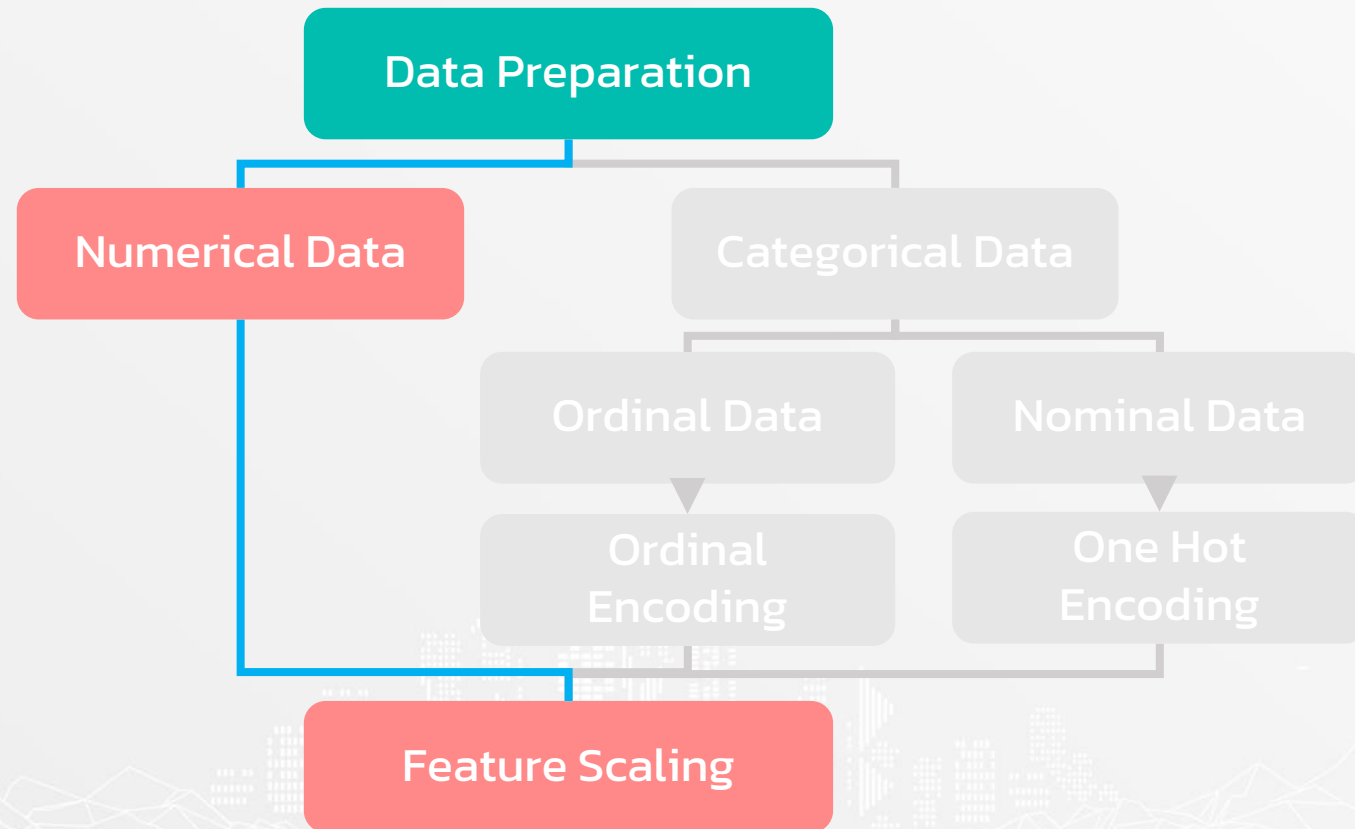
Read Data

```
1 data = pd.read_csv('../image_to_csv/alzheimer_dataset.csv')
2
3 data
```

	pixel_0	pixel_1	pixel_2	pixel_3	pixel_4	pixel_5	pixel_6	pixel_7	pixel_8	pixel_9	...	pixel_36606	pixel_36607	label
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0
...
5116	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0
5117	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0
5118	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0
5119	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0
5120	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0

5121 rows × 36609 columns

Data Preparation



File



05. ALZHEIMER



alzheimer_model.pickle



alzheimer_mc.ipynb



alzheimer_md.ipynb



Imgage_to_csv

K-Nearest Neighbors

