



Alzheimer Classification

นายปัญญาวัตร สุวรรณทัต 6614450042

Data Processing

```
1 dataSet.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4100 entries, 0 to 4099
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   HN                     4100 non-null   int64  
1   Education              4100 non-null   object  
2   age                    4100 non-null   float64 
3   hypertension           4100 non-null   int64  
4   heart_disease          4100 non-null   int64  
5   family_history         4100 non-null   object  
6   Occupation             4100 non-null   object  
7   Triglycerides          4100 non-null   object  
8   Blood Sugar            4100 non-null   float64 
9   BMI                    3936 non-null   float64 
10  smoking_status         4100 non-null   object  
11  alzheimer              4100 non-null   int64  
dtypes: float64(3), int64(4), object(5)
memory usage: 384.5+ KB
```

4100 rows x 12 columns

```
1 dataSet.isna().sum()

HN                0
Education         0
age               0
hypertension      0
heart_disease     0
family_history    0
Occupation        0
Triglycerides     0
Blood Sugar       0
BMI               164
smoking_status    0
alzheimer         0
dtype: int64
```

Data Processing

Check BMI ที่ Class == 1

```
1 len(dataSet[(dataSet['alzheimer'] == 1)])
```

239

```
1 len(dataSet[(dataSet['BMI'].isna()) & (dataSet['alzheimer'] == 1)])
```

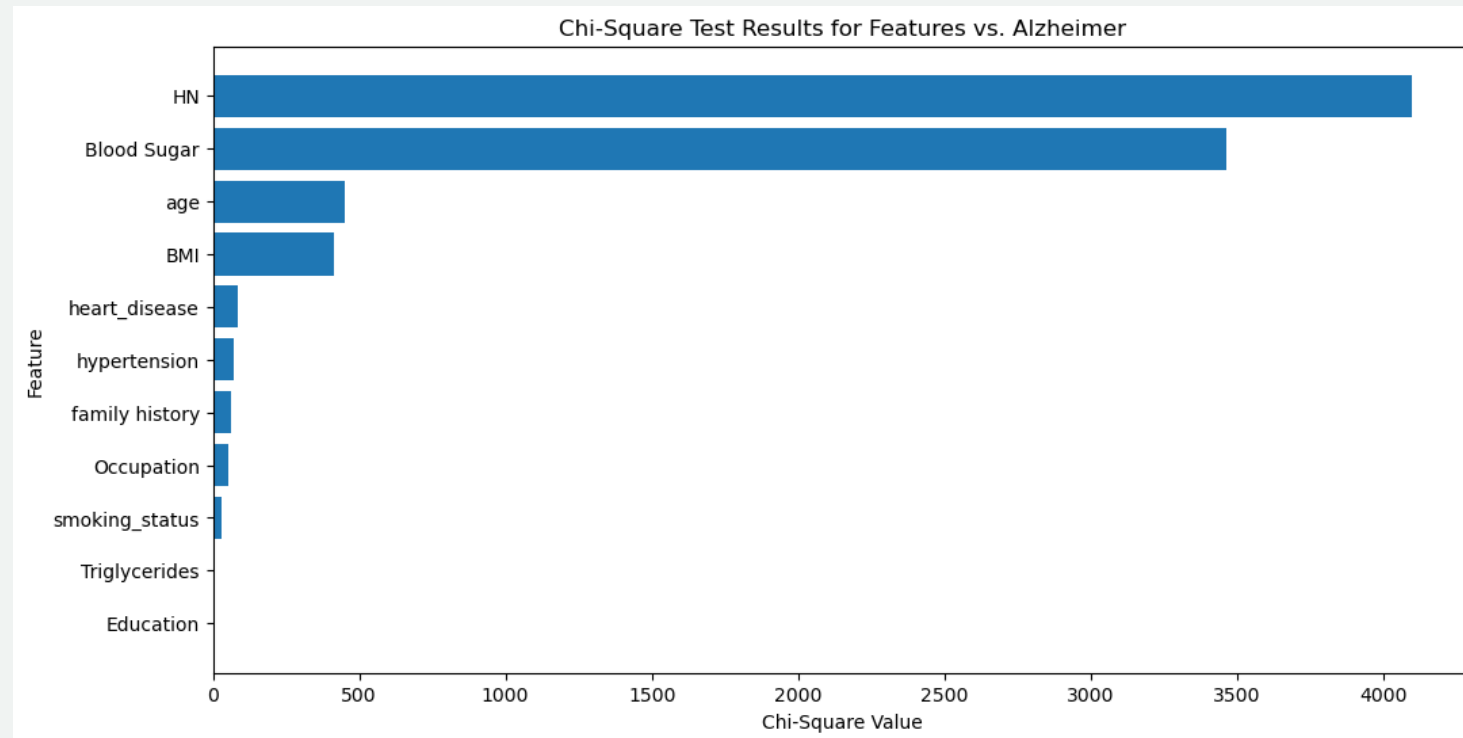
37

```
1 len(dataSet[(dataSet['BMI'].isna() == False) & (dataSet['alzheimer'] == 1)])
```

202

Data Processing

Chi-Square หาความสัมพันธ์ของ Feature และ Class



Data Processing

MARK: จากการตรวจสอบ พิจารณา Top 5 ความสัมพันธ์ระหว่าง Feature กับ Class พบว่า BMI ค่อนข้างมีผลกับ Class จึงตัดสินใจแก้ไข BMI None ด้วยค่าเฉลี่ย

```
1 average_bmi = dataSet['BMI'].mean()
2 dataSet['BMI'].fillna(average_bmi, inplace=True)
```

```
1 dataSet.isna().sum()
```

```
HN          0
Education   0
age         0
hypertension 0
heart_disease 0
family history 0
Occupation  0
Triglycerides 0
Blood Sugar 0
BMI         0
smoking_status 0
alzheimer   0
dtype: int64
```

MARK: ตรวจสอบ Class Target จำนวนเหมาะสมกันหรือไม่

```
1 print("Target == 1 ->", len(dataSet.loc[dataSet["alzheimer"] == 1, "alzheimer"]))
```

```
Target == 1 -> 239
```

```
1 print("Target == 0 ->", len(dataSet.loc[dataSet["alzheimer"] == 0, "alzheimer"]))
```

```
Target == 0 -> 3861
```

Data Processing

Before LabelEncode

	HN	Education	age	hypertension	heart_disease	family history	Occupation	Triglycerides	Blood Sugar	BMI	smoking_status	alzheimer
0	11046	Secondary School	67.0	0	1	Yes	Private	Normal	228.69	36.600000	formerly smoked	1
1	51676	Bachelor	61.0	0	0	Yes	Freelance	High	202.21	28.680056	never smoked	1
2	35112	Secondary School	80.0	0	1	Yes	Private	High	105.92	32.500000	never smoked	1
3	67182	Bachelor	49.0	0	0	Yes	Private	Normal	171.23	34.400000	smokes	1
4	1665	Bachelor	79.0	1	0	Yes	Freelance	High	174.12	24.000000	never smoked	1

After LabelEncode

	HN	Education	age	hypertension	heart_disease	family history	Occupation	Triglycerides	Blood Sugar	BMI	smoking_status	alzheimer
0	11046	1	67.0	0	1	1	4	1	228.69	36.600000	1	1
1	51676	0	61.0	0	0	1	1	0	202.21	28.680056	2	1
2	35112	1	80.0	0	1	1	4	0	105.92	32.500000	2	1
3	67182	0	49.0	0	0	1	4	1	171.23	34.400000	3	1
4	1665	0	79.0	1	0	1	1	0	174.12	24.000000	2	1

Data Processing

Train Test Split = Test 30%, Train 70%

SMOTE เนื่องจาก Class Imbalance

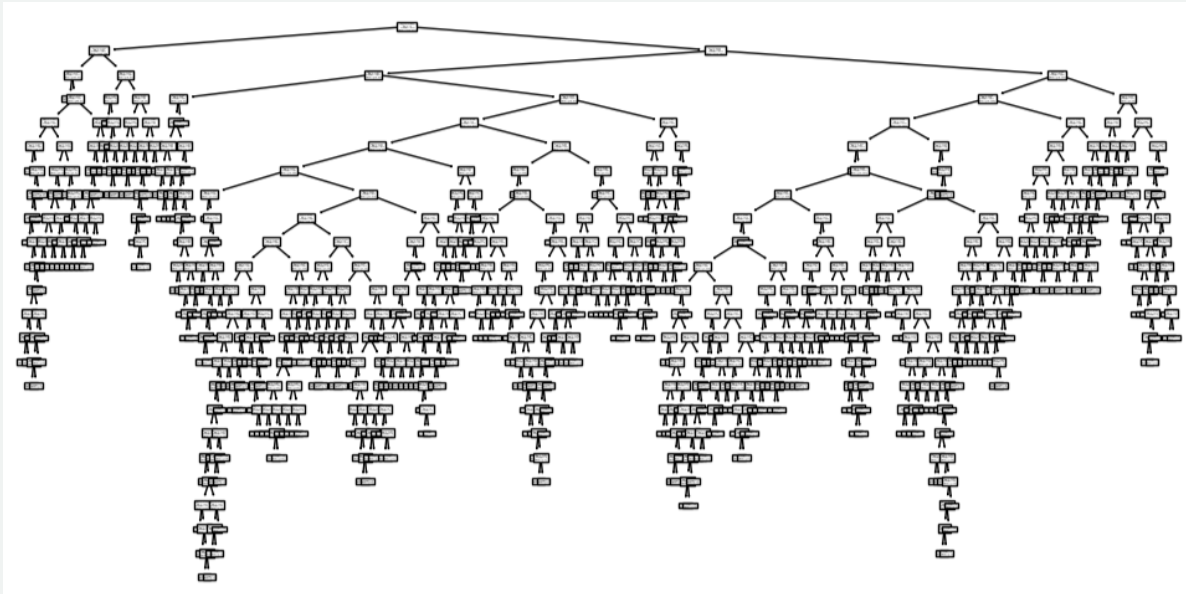
ทำ Scaler

Decision Tree

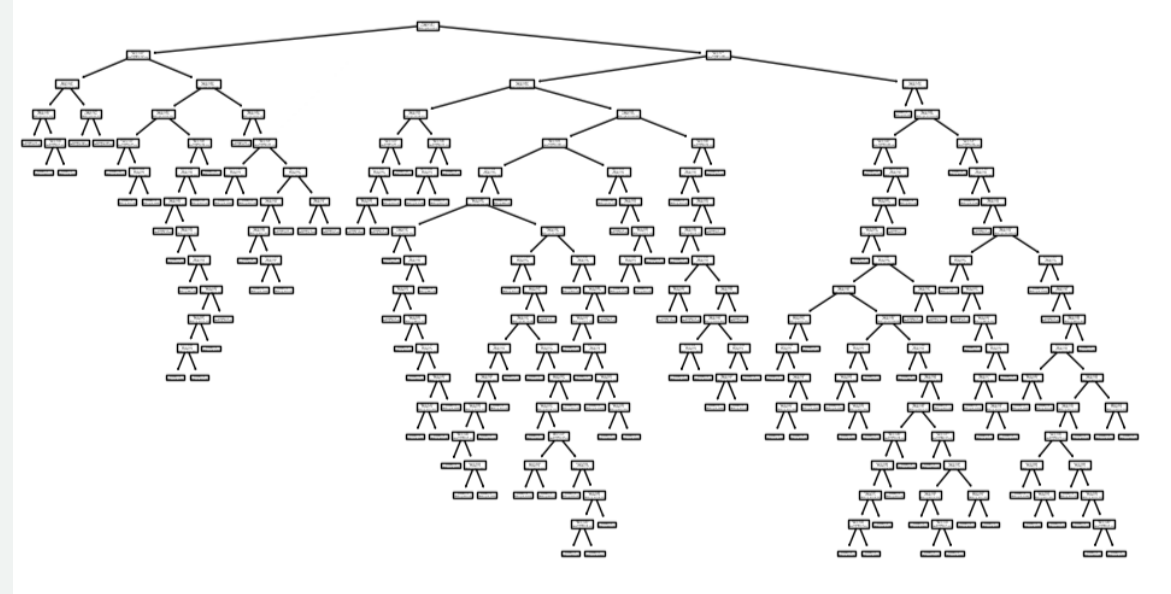
Decision Tree

Parameter

criterion="gini",class_weight='balanced' random_state=42



Visualize Decision Tree(Training)



Visualize Decision Tree(Test)

Decision Tree

GridSearchCV

```
param_grid = {  
    'max_depth': [2, 4, 6, 8, 10],  
    'min_samples_split': [2, 5, 10],  
    'min_samples_leaf': [1, 2, 4],  
    'max_features': ['auto', 'sqrt', 'log2'],  
    'criterion': ['gini', 'entropy']  
}
```

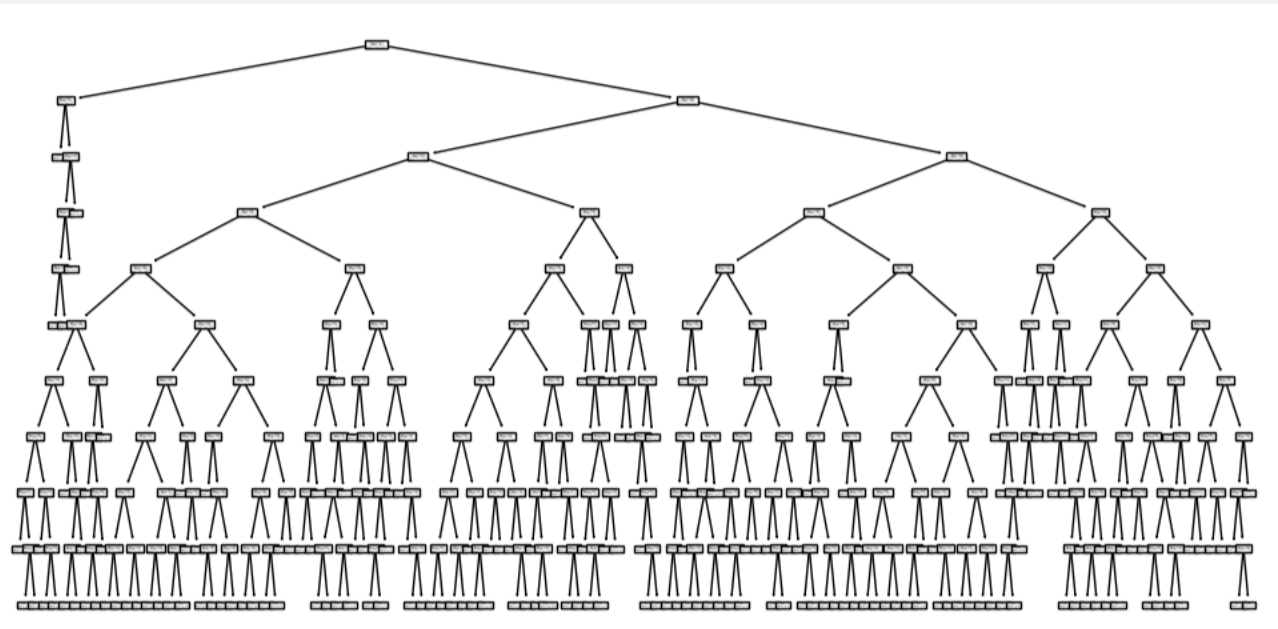
```
Best Parameters: {'criterion': 'gini', 'max_depth': 10, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 10}
```

```
Best Accuracy: 0.8648832737514895
```

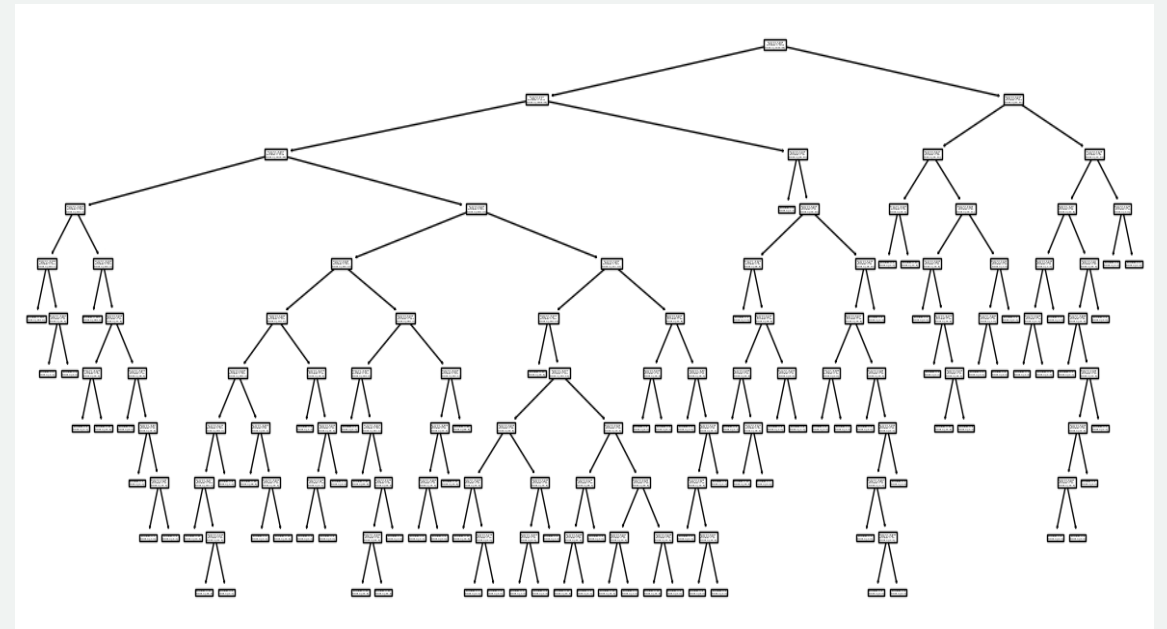
Decision Tree

Best Parameter

Best Parameters: {'criterion': 'gini', 'max_depth': 10, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 10}



Visualize Decision Tree(Training)



Visualize Decision Tree(Test)

Precision: ความแม่นยำในการทำนาย Class + ถูกต้องกี่ครั้ง คิดเป็น % จากทั้งหมด

Recall: ความแม่นยำในการตรวจสอบ Class + จากตัวอย่างข้อมูล n ตัว

F1-Score: ค่าเฉลี่ยความถูกต้องระหว่าง Precision และ Recall

Support: จำนวนตัวอย่างที่อยู่ในแต่ละคลาสในชุดข้อมูลทดสอบ

Decision Tree

Parameter

criterion="gini",class_weight='balanced' random_state=42

col_0	0	1
alzheimer		
0	1020	125
1	65	20

```
Classification Report:
              precision    recall  f1-score   support

    0           0.94        0.89        0.91       1145
    1           0.14        0.24        0.17         85

 accuracy          0.85       1230
 macro avg         0.54        0.56        0.54       1230
 weighted avg      0.88        0.85        0.86       1230
```

Best Parameter

criterion: 'gini', 'max_depth': 10,
'max_features': 'sqrt', 'min_samples_leaf': 1,
'min_samples_split': 10

col_0	0	1
alzheimer		
0	922	223
1	50	35

```
Classification Report:
              precision    recall  f1-score   support

    0           0.95        0.81        0.87       1145
    1           0.14        0.41        0.20         85

 accuracy          0.78       1230
 macro avg         0.54        0.61        0.54       1230
 weighted avg      0.89        0.78        0.82       1230
```

Precision: ความแม่นยำในการทำนาย Class + ถูกต้องกี่ครั้ง คิดเป็น % จากทั้งหมด

Recall: ความแม่นยำในการตรวจสอบ Class + จากตัวอย่างข้อมูล n ตัว

F1-Score: ค่าเฉลี่ยความถูกต้องระหว่าง Precision และ Recall

Support: จำนวนตัวอย่างที่อยู่ในแต่ละคลาสในชุดข้อมูลทดสอบ

Decision Tree

FillNA

Parameter

criterion="gini",class_weight='balanced' random_state=42

col_0	0	1
alzheimer		
0	1020	125
1	65	20

Classification Report:					
	precision	recall	f1-score	support	
0	0.94	0.89	0.91	1145	
1	0.14	0.24	0.17	85	
accuracy			0.85	1230	
macro avg	0.54	0.56	0.54	1230	
weighted avg	0.88	0.85	0.86	1230	

Best Parameter

criterion': 'gini', 'max_depth': 10, 'max_features': 'sqrt',
'min_samples_leaf': 1, 'min_samples_split': 10

col_0	0	1
alzheimer		
0	922	223
1	50	35

Classification Report:					
	precision	recall	f1-score	support	
0	0.95	0.81	0.87	1145	
1	0.14	0.41	0.20	85	
accuracy			0.78	1230	
macro avg	0.54	0.61	0.54	1230	
weighted avg	0.89	0.78	0.82	1230	

Drop NA BMI

Parameter

criterion="gini",class_weight='balanced' random_state=42

col_0	0	1
alzheimer		
0	1008	106
1	53	14

Classification Report:					
	precision	recall	f1-score	support	
0	0.95	0.90	0.93	1114	
1	0.12	0.21	0.15	67	
accuracy			0.87	1181	
macro avg	0.53	0.56	0.54	1181	
weighted avg	0.90	0.87	0.88	1181	

Best Parameter

criterion': 'gini', 'max_depth': 10, 'max_features': 'log2',
'min_samples_leaf': 4, 'min_samples_split': 5

col_0	0	1
alzheimer		
0	903	211
1	40	27

Classification Report:					
	precision	recall	f1-score	support	
0	0.96	0.81	0.88	1114	
1	0.11	0.40	0.18	67	
accuracy			0.79	1181	
macro avg	0.54	0.61	0.53	1181	
weighted avg	0.91	0.79	0.84	1181	

The background is a light gray with several decorative elements: a purple circle with a blue flower-like pattern in the top left; a large purple circle with radial lines in the top right; a pink circle with wavy lines in the bottom left; and a blue circle with a dashed pattern in the bottom right. Scattered throughout are various colored dots in shades of pink, purple, blue, and orange.

KNN

KNN

Parameter Default

```
{'algorithm': 'auto', 'leaf_size': 30, 'metric': 'minkowski', 'metric_params': None, 'n_jobs':  
None, 'n_neighbors': 5, 'p': 2, 'weights': 'uniform'}
```

GridSearchCV

```
param_grid = {'algorithm': 'auto', 'leaf_size': 30,  
'metric': 'minkowski', 'metric_params': None, 'n_jobs':  
None, 'n_neighbors': 1, 'p': 2, 'weights': 'uniform'}
```

```
Best K: 1  
Best Accuracy: 0.9026246479254685
```

Precision: ความแม่นยำในการทำนาย Class + ถูกต้องกี่ครั้ง คิดเป็น % จากทั้งหมด

Recall: ความแม่นยำในการตรวจสอบ Class + จากตัวอย่างข้อมูล n ตัว

F1-Score: ค่าเฉลี่ยความถูกต้องระหว่าง Precision และ Recall

Support: จำนวนตัวอย่างที่อยู่ในแต่ละคลาสในชุดข้อมูลทดสอบ

KNN

Parameter Default

```
{'algorithm': 'auto', 'leaf_size': 30, 'metric':  
'minkowski', 'metric_params': None, 'n_jobs': None,  
'n_neighbors': 5, 'p': 2, 'weights': 'uniform'}
```

col_0	0	1
alzheimer		
0	939	206
1	58	27

```
Classification Report:  
              precision    recall  f1-score   support  
  
   0       0.94         0.82         0.88         1145  
   1       0.12         0.32         0.17           85  
  
 accuracy          0.79         0.79         0.79         1230  
 macro avg         0.53         0.57         0.52         1230  
 weighted avg      0.88         0.79         0.83         1230
```

Best Parameter

```
param_grid = {'algorithm': 'auto', 'leaf_size': 30,  
'metric': 'minkowski', 'metric_params': None,  
'n_jobs': None, 'n_neighbors': 1, 'p': 2, 'weights':  
'uniform'}
```

col_0	0	1
alzheimer		
0	1006	139
1	67	18

```
Classification Report:  
              precision    recall  f1-score   support  
  
   0       0.94         0.88         0.91         1145  
   1       0.11         0.21         0.15           85  
  
 accuracy          0.83         0.83         0.83         1230  
 macro avg         0.53         0.55         0.53         1230  
 weighted avg      0.88         0.83         0.85         1230
```


Precision: ความแม่นยำในการทำนาย Class + ถูกต้องกี่ครั้ง คิดเป็น % จากทั้งหมด
Recall: ความแม่นยำในการตรวจสอบ Class + จากตัวอย่างข้อมูล n ตัว
F1-Score: ค่าเฉลี่ยความถูกต้องระหว่าง Precision และ Recall
Support: จำนวนตัวอย่างที่อยู่ในแต่ละคลาสในชุดข้อมูลทดสอบ

KNN

Fit

Parameter Default

{'algorithm': 'auto', 'leaf_size': 30, 'metric': 'minkowski', 'metric_params': None, 'n_jobs': None, 'n_neighbors': 5, 'p': 2, 'weights': 'uniform'}

Classification Report:					
	precision	recall	f1-score	support	
0	0.94	0.82	0.88	1145	
1	0.12	0.32	0.17	85	
accuracy			0.79	1230	
macro avg	0.53	0.57	0.52	1230	
weighted avg	0.88	0.79	0.83	1230	

col_0	0	1
alzheimer		
0	939	206
1	58	27

Best Parameter

param_grid = {'algorithm': 'auto', 'leaf_size': 30, 'metric': 'minkowski', 'metric_params': None, 'n_jobs': None, 'n_neighbors': 1, 'p': 2, 'weights': 'uniform'}

Classification Report:					
	precision	recall	f1-score	support	
0	0.94	0.88	0.91	1145	
1	0.11	0.21	0.15	85	
accuracy			0.83	1230	
macro avg	0.53	0.55	0.53	1230	
weighted avg	0.88	0.83	0.85	1230	

col_0	0	1
alzheimer		
0	1006	139
1	67	18

Drop NA BMI

Parameter Default

{'algorithm': 'auto', 'leaf_size': 30, 'metric': 'minkowski', 'metric_params': None, 'n_jobs': None, 'n_neighbors': 5, 'p': 2, 'weights': 'uniform'}

Classification Report:					
	precision	recall	f1-score	support	
0	0.96	0.81	0.88	1114	
1	0.12	0.42	0.18	67	
accuracy			0.79	1181	
macro avg	0.54	0.61	0.53	1181	
weighted avg	0.91	0.79	0.84	1181	

col_0	0	1
alzheimer		
0	900	214
1	39	28

Best Parameter

param_grid = {'algorithm': 'auto', 'leaf_size': 30, 'metric': 'minkowski', 'metric_params': None, 'n_jobs': None, 'n_neighbors': 1, 'p': 2, 'weights': 'uniform'}

Classification Report:					
	precision	recall	f1-score	support	
0	0.95	0.85	0.90	1114	
1	0.11	0.30	0.16	67	
accuracy			0.82	1181	
macro avg	0.53	0.58	0.53	1181	
weighted avg	0.90	0.82	0.86	1181	

col_0	0	1
alzheimer		
0	949	165
1	47	20

Neural Network

Neural Network

Parameter

hidden_layer_sizes=(10, 5), max_iter=1000, random_state=42

GridSearchCV

```
param_grid = {  
    'hidden_layer_sizes': [(64, 32, 16), (24, 12, 6), (12, 6, 3)],  
    'activation': ['relu'],  
    'solver': ['adam'],  
    'max_iter': [100, 500, 700, 1000],  
    'learning_rate': ['adaptive'],  
    'batch_size': [64, 32, 16],  
    'learning_rate_init': [0.001, 0.01, 0.1]  
}
```

Precision: ความแม่นยำในการทำนาย Class + ถูกต้องกี่ครั้ง คิดเป็น % จากทั้งหมด

Recall: ความแม่นยำในการตรวจสอบ Class + จากตัวอย่างข้อมูล n ตัว

F1-Score: ค่าเฉลี่ยความถูกต้องระหว่าง Precision และ Recall

Support: จำนวนตัวอย่างที่อยู่ในแต่ละคลาสในชุดข้อมูลทดสอบ

Neural Network

Parameter

hidden_layer_sizes=(10, 5),
max_iter=1000, random_state=42

col_0	0	1
alzheimer		
0	967	178
1	52	33

Classification Report:					
	precision	recall	f1-score	support	
0	0.95	0.84	0.89	1145	
1	0.16	0.39	0.22	85	
accuracy			0.81	1230	
macro avg	0.55	0.62	0.56	1230	
weighted avg	0.89	0.81	0.85	1230	

Best Parameter

batch_size=32, hidden_layer_sizes=(64, 32, 16),
learning_rate='adaptive', max_iter=500,
random_state=42, 'learning_rate_init': 0.001

col_0	0	1
alzheimer		
0	1024	121
1	61	24

Classification Report:					
	precision	recall	f1-score	support	
0	0.94	0.89	0.92	1145	
1	0.17	0.28	0.21	85	
accuracy			0.85	1230	
macro avg	0.55	0.59	0.56	1230	
weighted avg	0.89	0.85	0.87	1230	

Precision: ความแม่นยำในการทำนาย Class + ถูกต้องกี่ครั้ง คิดเป็น % จากทั้งหมด

Recall: ความแม่นยำในการตรวจสอบ Class + จากตัวอย่างข้อมูล n ตัว

F1-Score: ค่าเฉลี่ยความถูกต้องระหว่าง Precision และ Recall

Support: จำนวนตัวอย่างที่อยู่ในแต่ละคลาสในชุดข้อมูลทดสอบ

Neural Network

FillNA

Parameter

hidden_layer_sizes=(10, 5), max_iter=1000, random_state=42

col_0	0	1
alzheimer		
0	967	178
1	52	33

Classification Report:					
	precision	recall	f1-score	support	
0	0.95	0.84	0.89	1145	
1	0.16	0.39	0.22	85	
accuracy			0.81	1230	
macro avg	0.55	0.62	0.56	1230	
weighted avg	0.89	0.81	0.85	1230	

Best Parameter

batch_size=32, hidden_layer_sizes=(64, 32, 16), learning_rate='adaptive', max_iter=500, random_state=42, learning_rate_init= 0.001

col_0	0	1
alzheimer		
0	1024	121
1	61	24

Classification Report:					
	precision	recall	f1-score	support	
0	0.94	0.89	0.92	1145	
1	0.17	0.28	0.21	85	
accuracy			0.85	1230	
macro avg	0.55	0.59	0.56	1230	
weighted avg	0.89	0.85	0.87	1230	

Drop NA BMI

Parameter

hidden_layer_sizes=(10, 5), max_iter=1000, random_state=42

col_0	0	1
alzheimer		
0	903	211
1	38	29

Classification Report:					
	precision	recall	f1-score	support	
0	0.96	0.81	0.88	1114	
1	0.12	0.43	0.19	67	
accuracy			0.79	1181	
macro avg	0.54	0.62	0.53	1181	
weighted avg	0.91	0.79	0.84	1181	

Best Parameter

batch_size=32, hidden_layer_sizes=(64, 32, 16), learning_rate='adaptive', max_iter=500, random_state=42, learning_rate_init= 0.01

col_0	0	1
alzheimer		
0	1017	97
1	49	18

Classification Report:					
	precision	recall	f1-score	support	
0	0.95	0.91	0.93	1114	
1	0.16	0.27	0.20	67	
accuracy			0.88	1181	
macro avg	0.56	0.59	0.57	1181	
weighted avg	0.91	0.88	0.89	1181	

Compare Best of Model

Precision: ความแม่นยำในการทำนาย Class + ถูกต้องกี่ครั้ง คิดเป็น % จากทั้งหมด
Recall: ความแม่นยำในการตรวจสอบ Class + จากตัวอย่างข้อมูล n ตัว
F1-Score: ค่าเฉลี่ยความถูกต้องระหว่าง Precision และ Recall
Support: จำนวนตัวอย่างที่อยู่ในแต่ละคลาสในชุดข้อมูลทดสอบ

Compare Best of Model

Decision Tree

Best Parameter

criterion': 'gini', 'max_depth': 10,
'max_features': 'sqrt', 'min_samples_leaf': 1,
'min_samples_split': 10

col_0	0	1
alzheimer		
0	922	223
1	50	35

Classification Report:				
	precision	recall	f1-score	support
0	0.95	0.81	0.87	1145
1	0.14	0.41	0.20	85
accuracy			0.78	1230
macro avg	0.54	0.61	0.54	1230
weighted avg	0.89	0.78	0.82	1230

KNN

Parameter Default

{'algorithm': 'auto', 'leaf_size': 30,
'metric': 'minkowski', 'metric_params':
None, 'n_jobs': None, 'n_neighbors': 5,
'p': 2, 'weights': 'uniform'}

col_0	0	1
alzheimer		
0	939	206
1	58	27

Classification Report:				
	precision	recall	f1-score	support
0	0.94	0.82	0.88	1145
1	0.12	0.32	0.17	85
accuracy			0.79	1230
macro avg	0.53	0.57	0.52	1230
weighted avg	0.88	0.79	0.83	1230

Neural Network

Parameter

hidden_layer_sizes=(10, 5), max_iter=1000,
random_state=42

col_0	0	1
alzheimer		
0	967	178
1	52	33

Classification Report:				
	precision	recall	f1-score	support
0	0.95	0.84	0.89	1145
1	0.16	0.39	0.22	85
accuracy			0.81	1230
macro avg	0.55	0.62	0.56	1230
weighted avg	0.89	0.81	0.85	1230