Mahidol University
Wisdom of the Land

Heart Disease Prediction

Bundit Rotyoon 6636060

Outline

- Introduction
- Exploring data
- Select and Train Model
- Result
- Conclusion

Introduction

Heart Disease Dataset

Introduction

Heart Disease Dataset

- เริ่มเก็บตั้งแต่ 1988

Introduction

Heart Disease Dataset

- เริ่มเก็บตั้งแต่ 1988
- ข้อมูลจาก 4 แหล่ง
  Cleveland, Hungary, Switzerland, and Long Beach V
- จริงๆ มี 76 Attribute แต่เผยแพร่แค่ 14

Introduction

Heart Disease Dataset

- เริ่มเก็บตั้งแต่ 1988
- ข้อมูลจาก 4 แหล่ง
  Cleveland, Hungary, Switzerland, and Long Beach V
- จริงๆ มี 76 Attribute แต่เผยแพร่แค่ 14
- Labeled Data (0 = no disease and 1 = disease)

Introduction

Mahidol University
Wisdom of the Land

Labeled Data

(0 = no disease and 1 = disease)

Introduction

Mahidol University
Wisdom of the Land

Labeled Data

(0 = no disease and 1 = disease)

Classification Problem

Introduction

Labeled Data

(0 = no disease and 1 = disease)

Classification Problem

Machine Learning Technique

Exploring data

Heart Disease Dataset

- 1,025 records
- 14 attributes

Exploring data

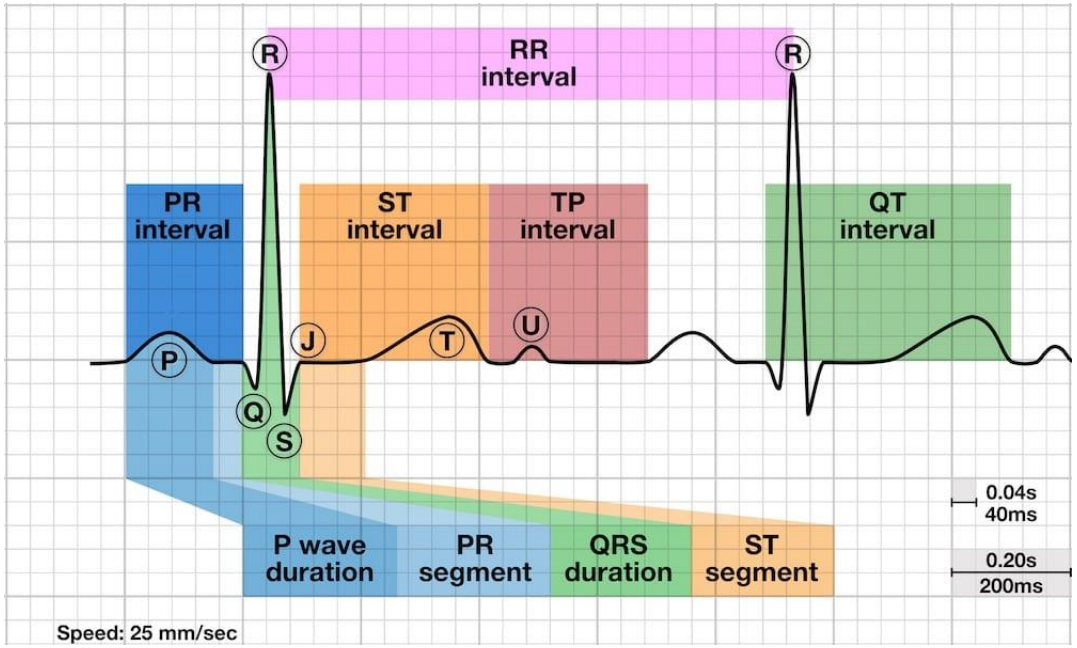
14 attributes

- age: age (age in years)
- sex: sex (1 = male; 0 = female)
- cp: chest pain type (4 values)
- trestbps: resting blood pressure
- chol: The person's cholesterol measurement in mg/dl
- fbs: fasting blood sugar > 120 mg/dl (1 = true; 0 = false)
- restecg: restecg: resting electrocardiographic results
    - Value 0: showing probable or definite left ventricular hypertrophy by Estes' criteria
    - Value 1: normal
    - Value 2: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
- thalach: maximum heart rate achieved
- exang: exercise induced angina (1 = yes; 0 = no)

Exploring data

14 attributes

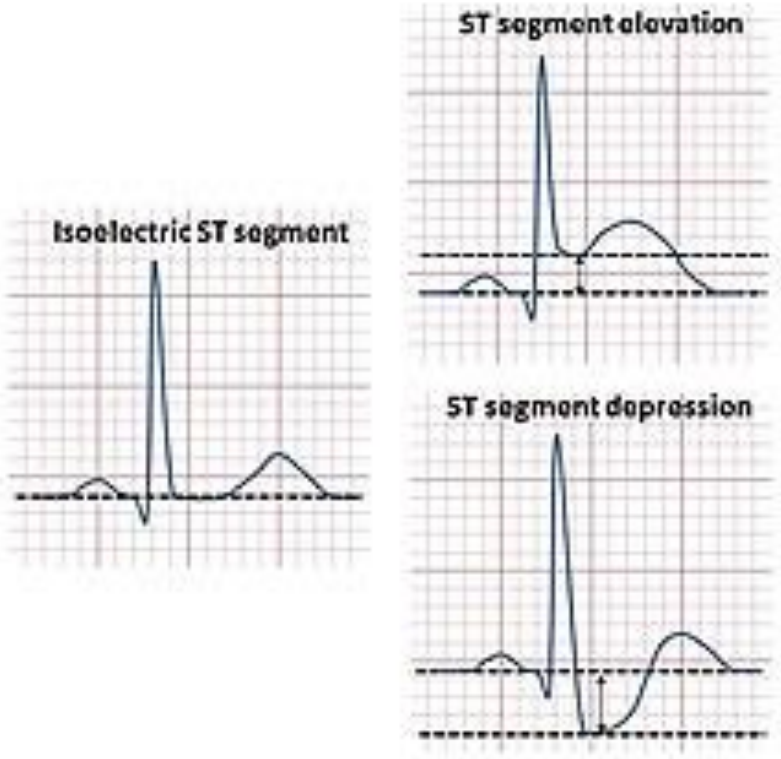
- age: age (age in years)
- sex: sex (1 = male; 0 = female)
- cp: chest pain type (4 values)
- trestbps: resting blood pressure
- chol: The person's cholesterol measurement in mg/dl
- fbs: fasting blood sugar > 120 mg/dl (1 = true; 0 = false)
- restecg: restecg: resting electrocardiographic results
    - Value 0: showing probable or definite left ventricular hypertrophy by Estes' criteria
    - Value 1: normal
    - Value 2: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
- thalach: maximum heart rate achieved
- exang: exercise induced angina (1 = yes; 0 = no)



Exploring data

14 attributes

- oldpeak: oldpeak = ST depression induced by exercise relative to rest
- slope: the slope of the peak exercise ST segment
- ca: number of major vessels (0-3) colored by flourosopy
- thal: A blood disorder called thalassemia
    - Value 0: NULL (dropped from the dataset previously
    - Value 1: fixed defect (no blood flow in some part of the heart)
    - Value 2: normal blood flow
    - Value 3: reversible defect (a blood flow is observed but it is not normal)
- target: 0 = no disease and 1 = disease



Exploring data

- ไม่มี Missing Value

```
df.isnull().sum()
```

```
age         0
sex         0
cp          0
trestbps    0
chol        0
fbs         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
target      0
dtype: int64
```

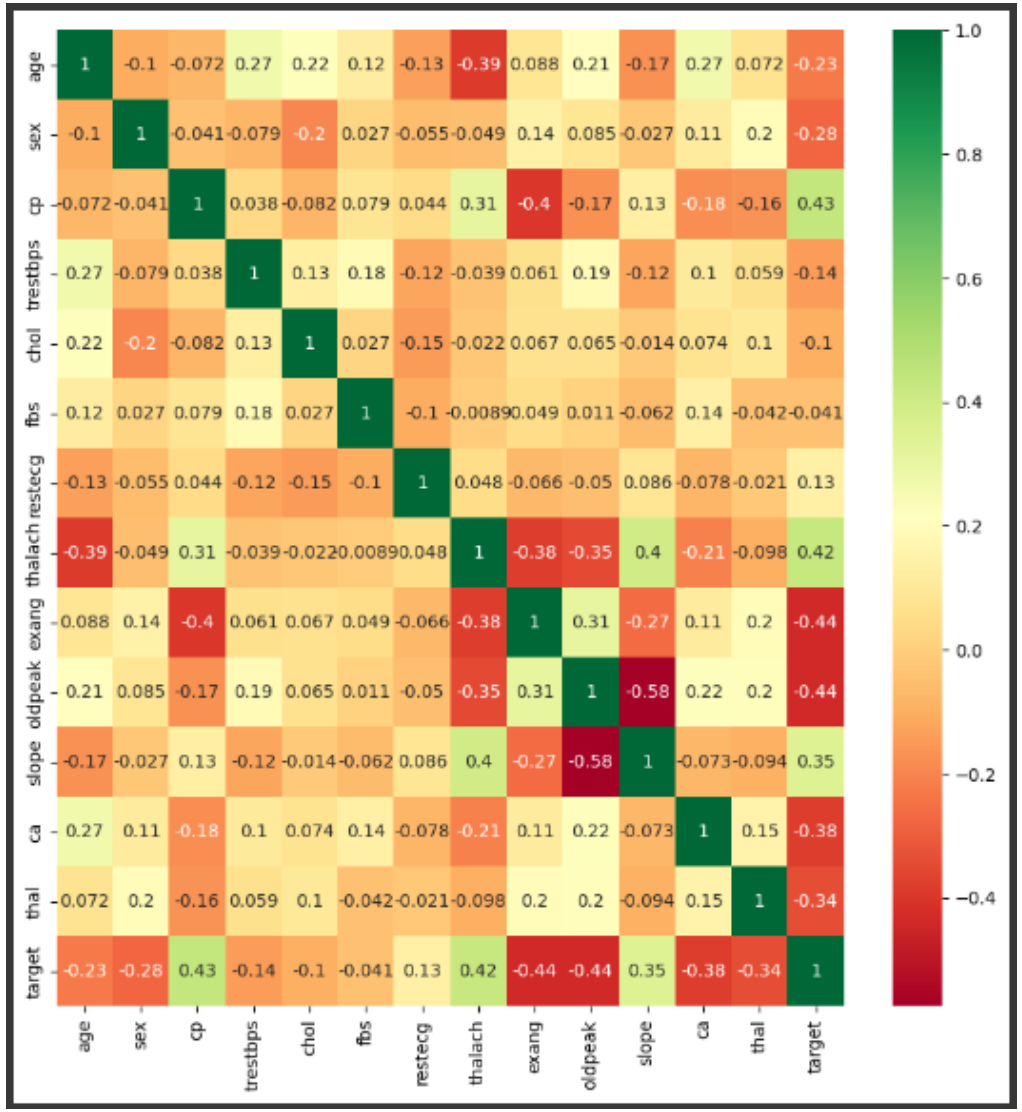
Exploring data

- ไม่มี Missing Value
- Balanced Data

```
df['target'].value_counts()
```

```
1    526
0    499
Name: target, dtype: int64
```

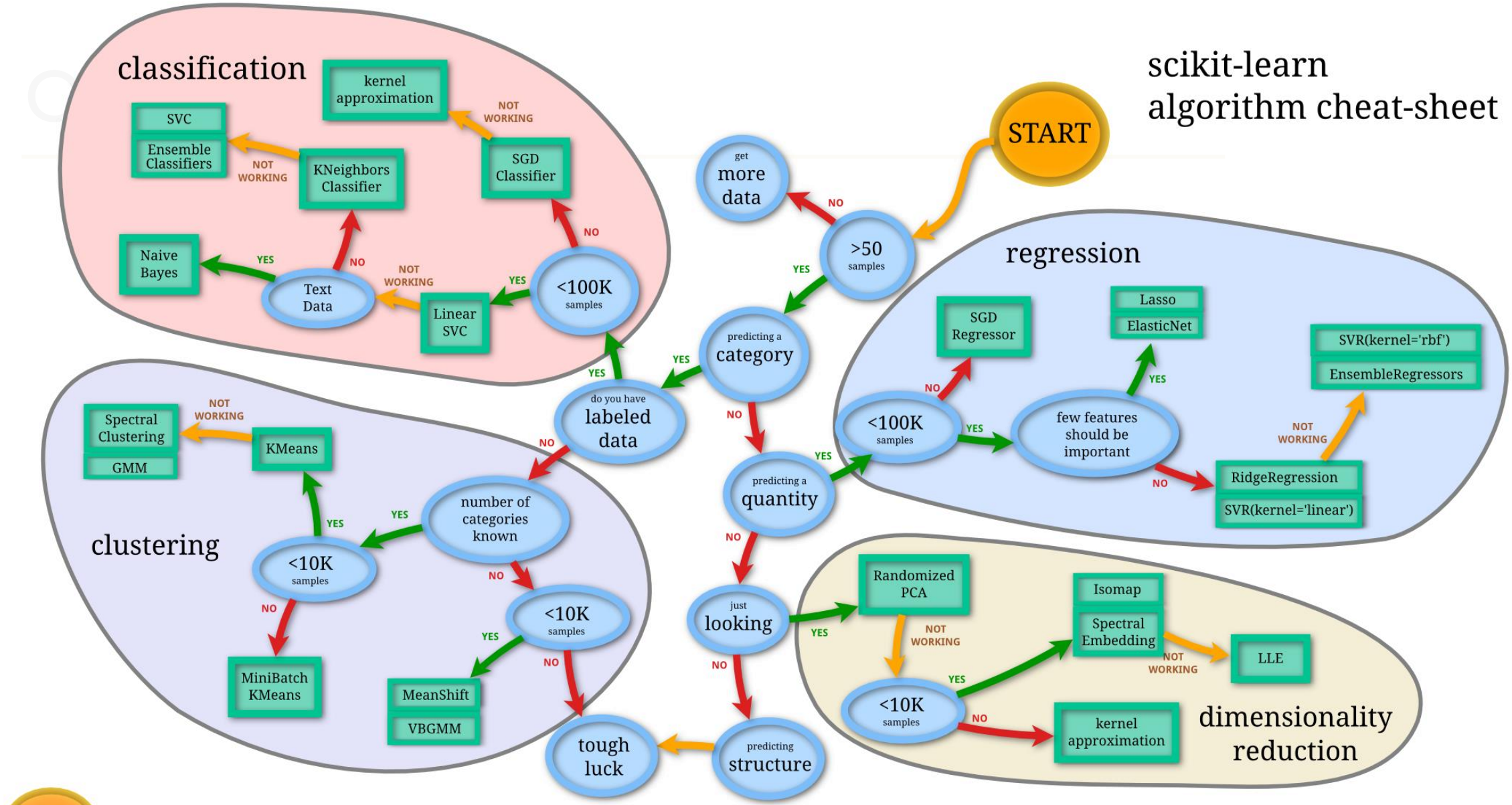
Exploring data



- ไม่มีความสัมพันธ์
- ไม่มีการตัด Feature

Select and Train Model

scikit-learn algorithm cheat-sheet



Select and Train Model

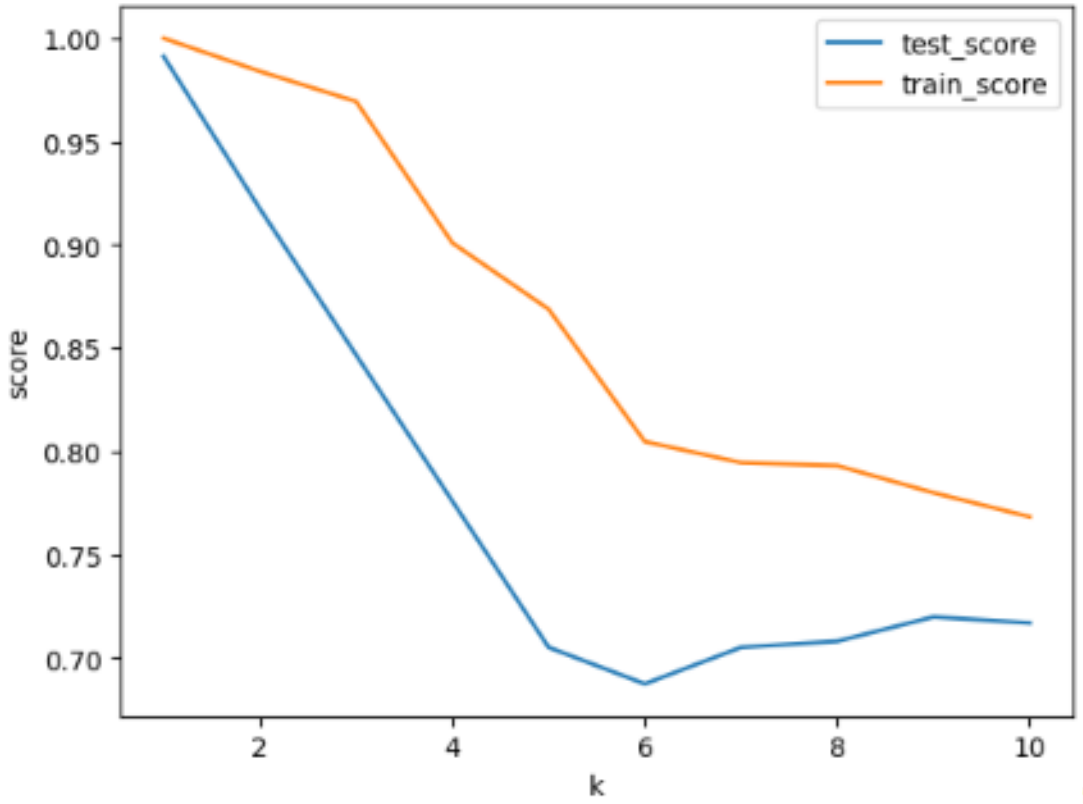
- K-Neighbors Classifier

Select and Train Model

- K-Neighbors Classifier
- เลือกค่า k

Select and Train Model

- K-Neighbors Classifier
- เลือกค่า k = 1



Select and Train Model

- K-Neighbors Classifier
- เลือกค่า k = 1
- Train-Test Split (33% Test)

Select and Train Model

- K-Neighbors Classifier
- เลือกค่า k = 1
- Train-Test Split (33% Test)
- Train Model

Result

K-Neighbors Classifier

- Accuracy = 0.96

Classification Report:

	precision	recall	f1-score	support
0	0.95	0.96	0.95	165
1	0.96	0.95	0.96	174

Confusion Matrix:

```
[[159   6]
 [  9 165]]
```

Result

เปรียบเทียบ Accuracy และ False Positive เพิ่มเติมกับอีก 3 models

- Support Vector Machine

Accuracy: 0.71

Classification Report:

	precision	recall	f1-score	support
0	0.71	0.67	0.69	165
1	0.70	0.74	0.72	174
accuracy			0.71	339
macro avg	0.71	0.71	0.71	339
weighted avg	0.71	0.71	0.71	339

Confusion Matrix:
[[111 54]
[45 129]]

- Random Forest Classifier

Accuracy: 0.94

Classification Report:

	precision	recall	f1-score	support
0	0.95	0.94	0.94	165
1	0.94	0.95	0.95	174
accuracy			0.94	339
macro avg	0.94	0.94	0.94	339
weighted avg	0.94	0.94	0.94	339

Confusion Matrix:
[[155 10]
[9 165]]

- Decision Tree

Accuracy: 0.88

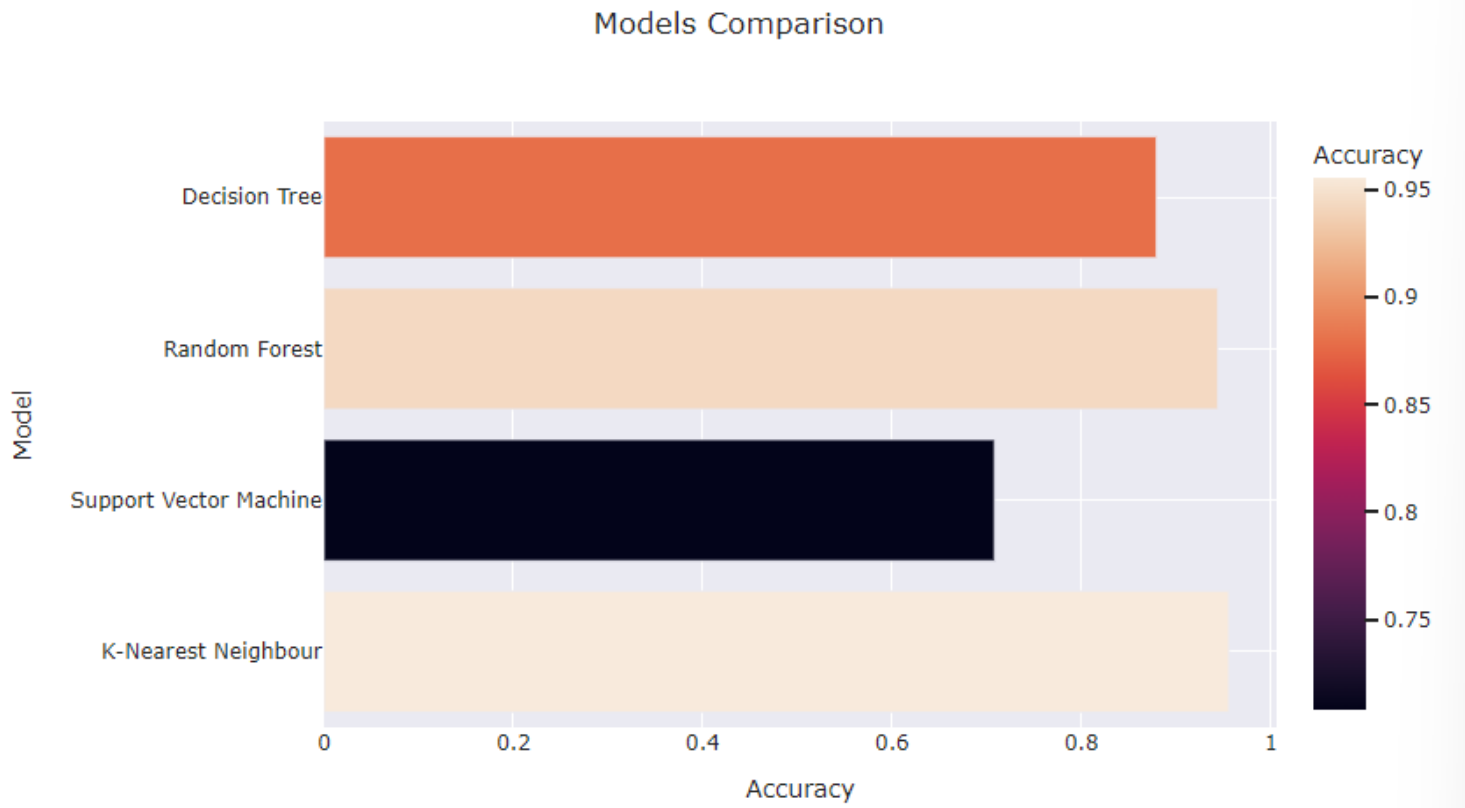
Classification Report:

	precision	recall	f1-score	support
0	0.91	0.84	0.87	165
1	0.86	0.92	0.89	174
accuracy			0.88	339
macro avg	0.88	0.88	0.88	339
weighted avg	0.88	0.88	0.88	339

Confusion Matrix:
[[138 27]
[14 160]]

Result

เปรียบเทียบ Accuracy และ False Positive เพิ่มเติมกับอีก 3 models



Conclusion

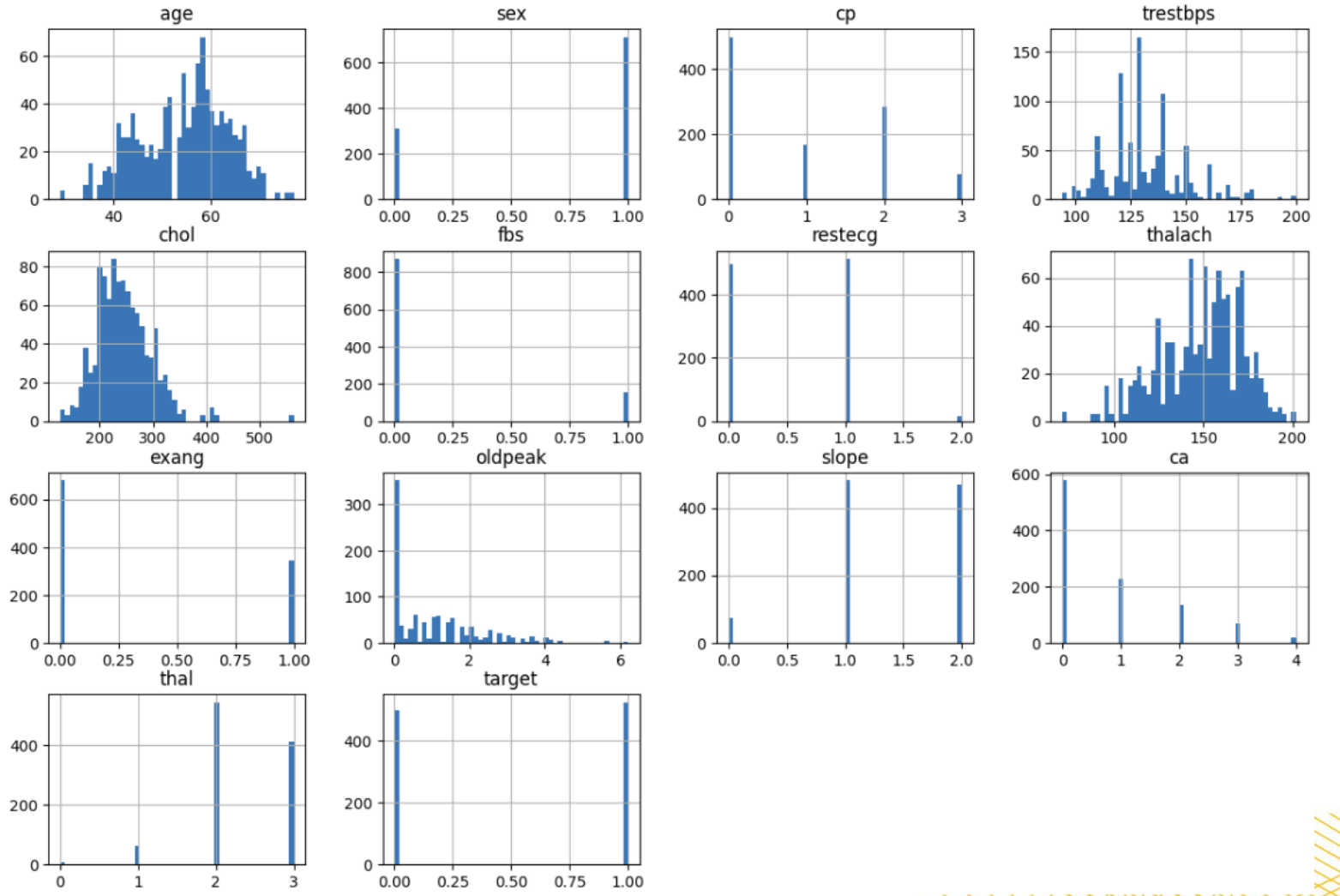
- ค่า Accuracy เท่ากับ 0.96 ค่อนข้างสูงมากเนื่องจากเลือก classifier และ parameter ได้เหมาะสม

- แต่ข้อสังเกตจาก confusion matrix ในช่อง false positive เท่ากับ 9 ค่อนข้างน่าเป็นห่วง เนื่องจากผู้ป่วยที่เป็นโรคหัวใจแต่ได้รับผลว่าไม่เป็นโรคอาจจะมีความระมัดระวังต่อพฤติกรรมที่เสี่ยงต่อโรคน้อยกว่า ซึ่งนำไปสู่อันตรายถึงชีวิตได้ แต่ก็ต่ำที่สุดในทั้ง 4 models

Mahidol University
Wisdom of the Land

Thank You for Your Attention

Q & A


age
sex
cp
trestbps
chol
fbs
restecg
thalach
exang
oldpeak
slope
ca
thal
target

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
count	1025.000000	1025.000000	1025.000000	1025.000000	1025.00000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000
mean	54.434146	0.695610	0.942439	131.611707	246.00000	0.149268	0.529756	149.114146	0.336585	1.071512	1.385366	0.754146	2.323902	0.513171
std	9.072290	0.460373	1.029641	17.516718	51.59251	0.356527	0.527878	23.005724	0.472772	1.175053	0.617755	1.030798	0.620660	0.500070
min	29.000000	0.000000	0.000000	94.000000	126.00000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	48.000000	0.000000	0.000000	120.000000	211.00000	0.000000	0.000000	132.000000	0.000000	0.000000	1.000000	0.000000	2.000000	0.000000
50%	56.000000	1.000000	1.000000	130.000000	240.00000	0.000000	1.000000	152.000000	0.000000	0.800000	1.000000	0.000000	2.000000	1.000000
75%	61.000000	1.000000	2.000000	140.000000	275.00000	0.000000	1.000000	166.000000	1.000000	1.800000	2.000000	1.000000	3.000000	1.000000
max	77.000000	1.000000	3.000000	200.000000	564.00000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000	3.000000	1.000000

```
1    526
0    499
Name: target, dtype: int64
```