



**MATPLOTLIB**

The background features a light gray crumpled paper texture. In the top left, there are blue and purple abstract shapes and three pills (red/white, blue/white, black/white). In the top right, there is an orange heart with a white ECG line. On the right side, there is a large orange pill outline. In the bottom left, there is an orange abstract shape and a blister pack of orange and white pills. In the bottom right, there are blue and purple abstract shapes and a small blue circle.

# DIABETES

---

**[ CLASSIFICATION ]**

# Business Understanding

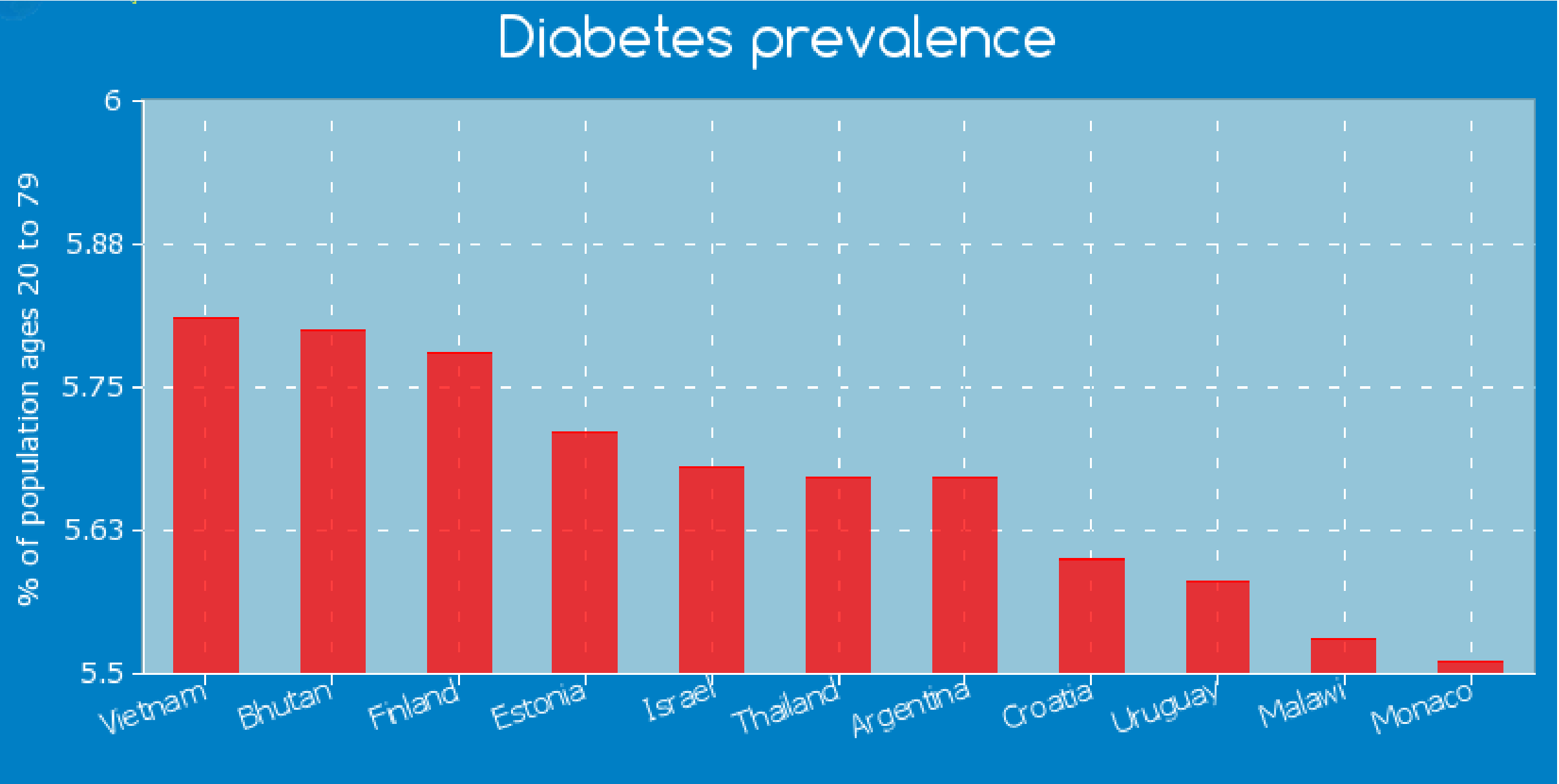
- ลดระยะเวลาและขั้นตอนในกระบวนการวินิจฉัย
- ตรวจสอบความเสี่ยงของการเป็นโรคเบาหวานได้ง่ายขึ้น
- ช่วยประกอบการตัดสินใจในการไปพบแพทย์
- ช่วยแนะนำการปรับพฤติกรรมที่เสี่ยงต่อการเป็นโรคเบาหวาน

## ประโยชน์ต่อวงการการแพทย์

วงการแพทย์ ผู้ที่ต้องการเข้ารับการวินิจฉัยและวงการธุรกิจเกี่ยวกับการดูแลสุขภาพ



# The Diabetes prevalence of Thailand



**5.67%**

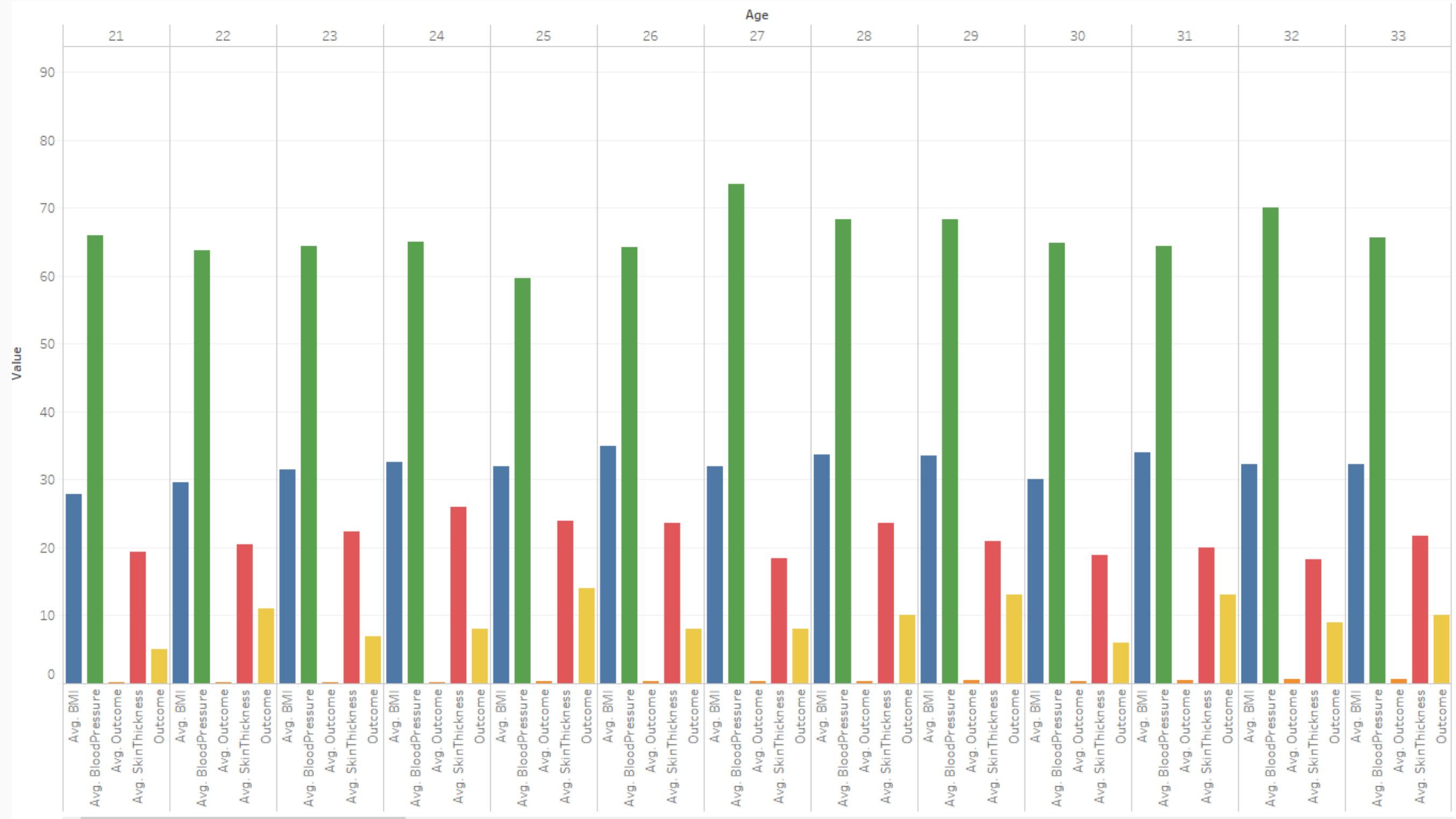
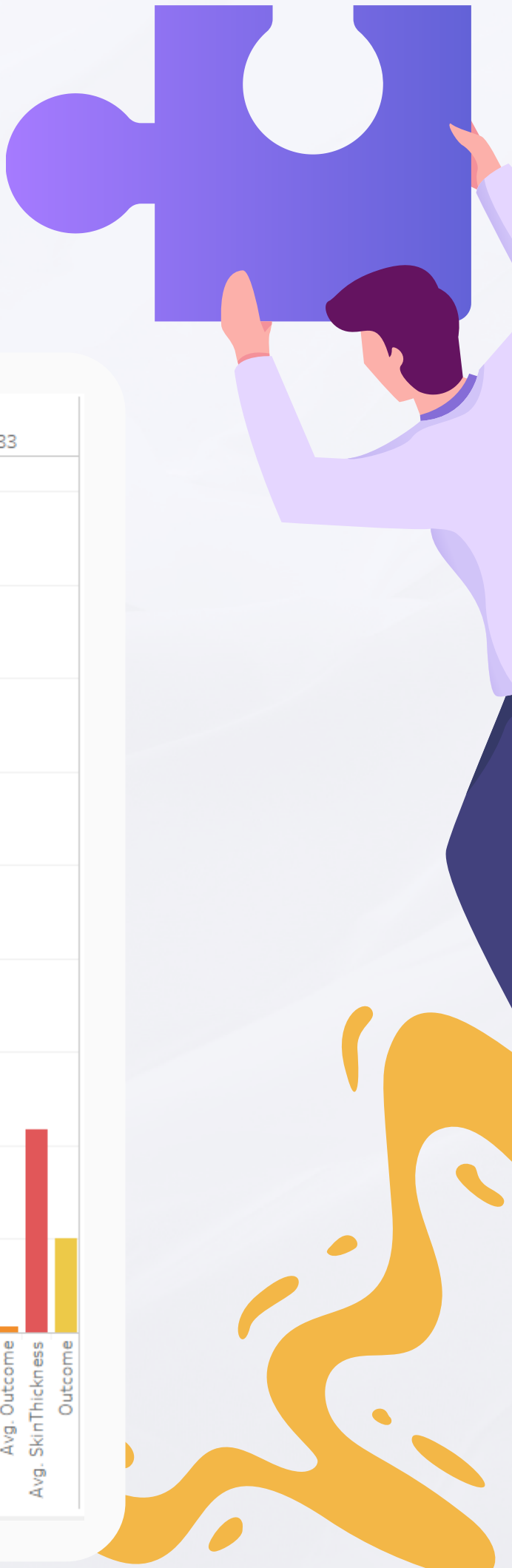
**of Thailand People have  
Diabetes**

**(% of population  
ages 20 to 79)**



# Data Understanding

- หาความสัมพันธ์ระหว่างข้อมูลที่ได้มาใน Data set และเลือกข้อมูลที่จะนำมาใช้



03

# Data Preparation & Exploration

Original Data

Data with remove  
SkinThickness & Pregnancies  
Column & Outlier

Data with remove  
SkinThickness & Pregnancies Column



# First Data

0s

```
1 Data = pd.read_csv('https://docs.google.com/spreadsheets/d/e/2PACX-1vRD6aYw4M_2XLqFYtLeq8vne6ZhNiHk0AfZdZvz64MNgC7SugGjKzn7')
2 Data
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
...	...	...	...	...	...	...	...	...	...
763	10	101	76	48	180	32.9	0.171	63	0
764	2	122	70	27	0	36.8	0.340	27	0
765	5	121	72	23	112	26.2	0.245	30	0
766	1	126	60	0	0	30.1	0.349	47	1
767	1	93	70	31	0	30.4	0.315	23	0

768 rows × 9 columns



# นำเข้าข้อมูลที่ได้และทำการตัดคอลัมน์ Pregnancies และ SkinThickness

```
[234] 1 Diabetes_data_drop = Diabetes_data.drop(['SkinThickness', 'Pregnancies'], axis=1)
      2 Diabetes_data_drop
```

	Glucose	BloodPressure	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	148	72	0	33.6	0.627	50	1
1	85	66	0	26.6	0.351	31	0
2	183	64	0	23.3	0.672	32	1
3	89	66	94	28.1	0.167	21	0
4	137	40	168	43.1	2.288	33	1
...	...	...	...	...	...	...	...
763	101	76	180	32.9	0.171	63	0
764	122	70	0	36.8	0.340	27	0
765	121	72	112	26.2	0.245	30	0
766	126	60	0	30.1	0.349	47	1
767	93	70	0	30.4	0.315	23	0

768 rows x 7 columns

```
1 Diabetes_data = pd.read_csv('/content/cleanedlaew_insurance.csv')
2 Diabetes_data
```

	Glucose	BloodPressure	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	89	66	94	28.1	0.167	21	0
1	137	40	168	43.1	2.288	33	1
2	78	50	88	31.0	0.248	26	1
3	197	70	543	30.5	0.158	53	1
4	189	60	846	30.1	0.398	59	1
...	...	...	...	...	...	...	...
387	181	88	510	43.3	0.222	26	1
388	128	88	110	36.5	1.057	37	1
389	88	58	16	28.4	0.766	22	0
390	101	76	180	32.9	0.171	63	0
391	121	72	112	26.2	0.245	30	0

392 rows x 7 columns

```
1 from sklearn.model_selection import train_test_split
2 x_train, x_test, y_train, y_test = train_test_split(Diabetes_data_drop.iloc[:, :-1],
3                                                    Diabetes_data_drop.iloc[:, -1:],
4                                                    test_size=0.3, random_state=12)
```

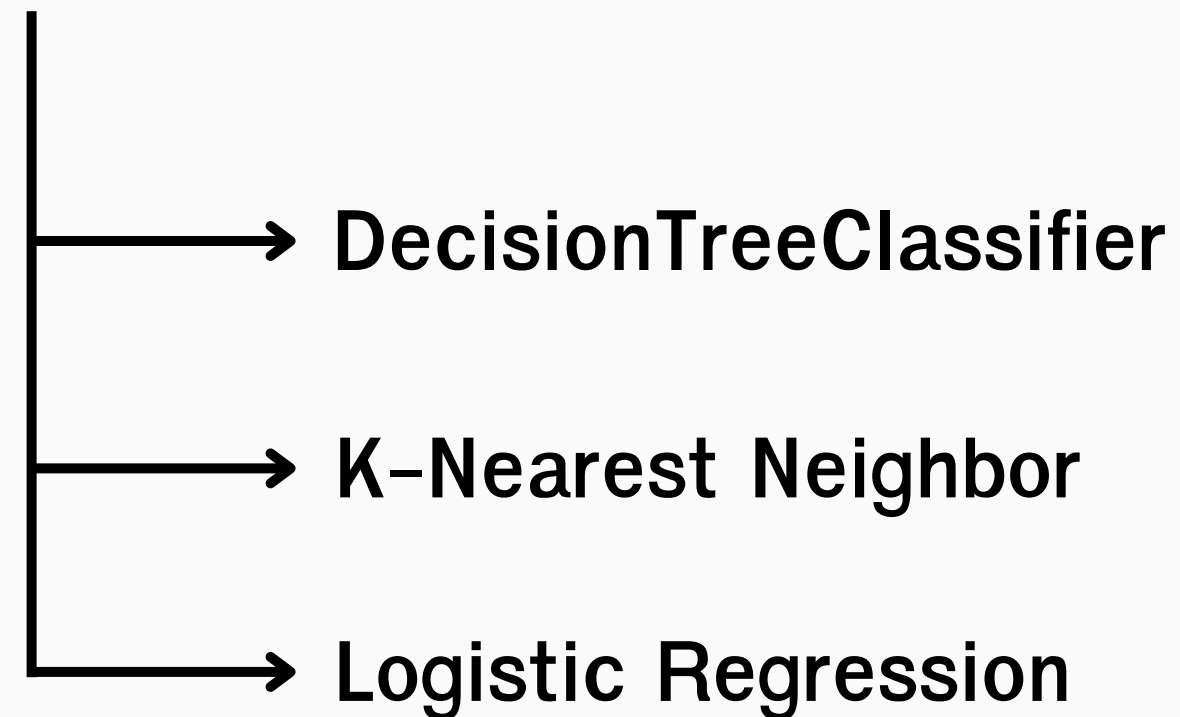
## ทำการสุ่มกลุ่ม Train และ Test



03

# Modeling

## Supervised Learning: Classification





# **Supervised Learning**

## **[Classification]**

### **[Decision Tree & K-Nearest Neighbor]**

---

**Example 1**

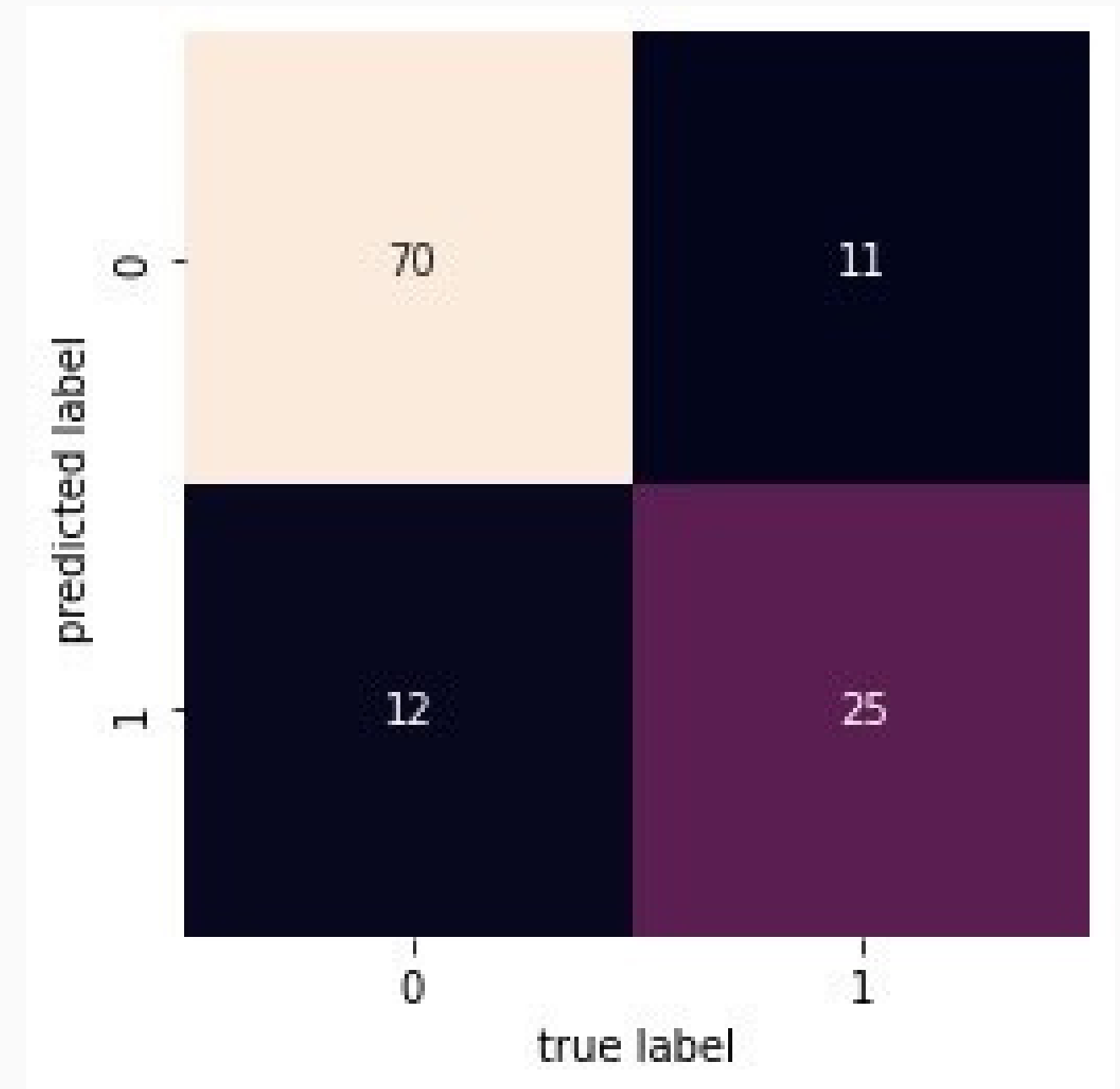
# Decision Tree Classifier

→ Accuracy: 81%

# K-Nearest Neighbor

→ Accuracy: 78%

	precision	recall	f1-score	support
0	0.86	0.85	0.86	82
1	0.68	0.69	0.68	36
accuracy			0.81	118
macro avg	0.77	0.77	0.77	118
weighted avg	0.81	0.81	0.81	118





# **Supervised Learning** **[Classification]**

## **[Decision Tree & K-Nearest Neighbor]**

---

**Example 2**

# Cleaning Data

```
1 Data = Data.drop(['SkinThickness', 'Pregnancies'], axis=1)
2 Data
```

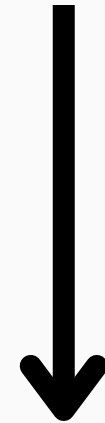
	Glucose	BloodPressure	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	148	72	0	33.6	0.627	50	1
1	85	66	0	26.6	0.351	31	0
2	183	64	0	23.3	0.672	32	1
3	89	66	94	28.1	0.167	21	0
5	116	74	0	25.6	0.201	30	0
...	...	...	...	...	...	...	...
763	101	76	180	32.9	0.171	63	0
764	122	70	0	36.8	0.340	27	0
765	121	72	112	26.2	0.245	30	0
766	126	60	0	30.1	0.349	47	1
767	93	70	0	30.4	0.315	23	0

690 rows x 7 columns

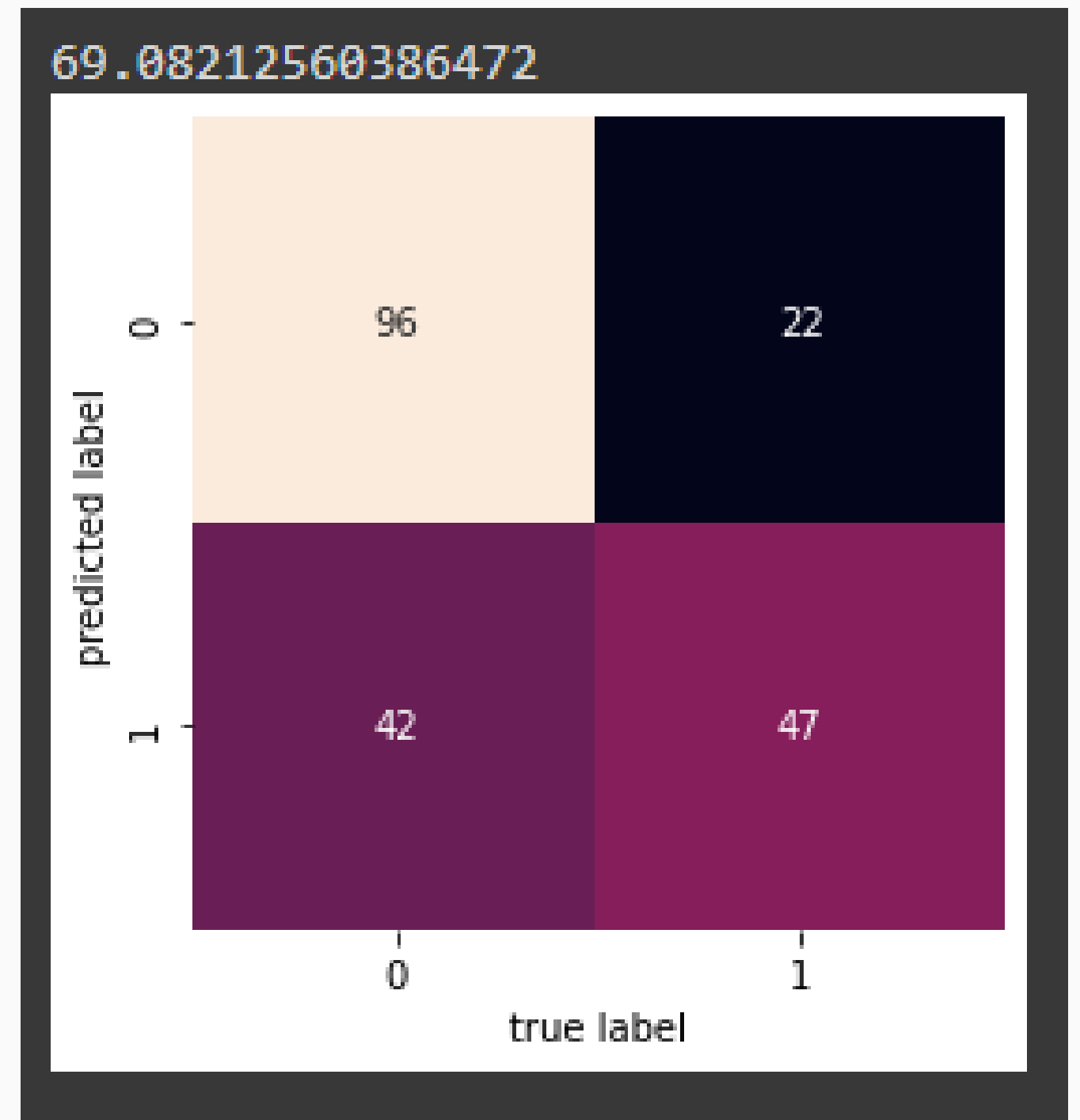
**Remove**  
SkinThickness &  
Pregnancies Column  
&  
All Outlier

# Accuracy

***Decision Tree***  
**[ Percentage Accuracy ]**



**69.08%**





# **Supervised Learning [Classification]**

## **LogisticRegression**

---

**Example 3**



# LogisticRegression Model

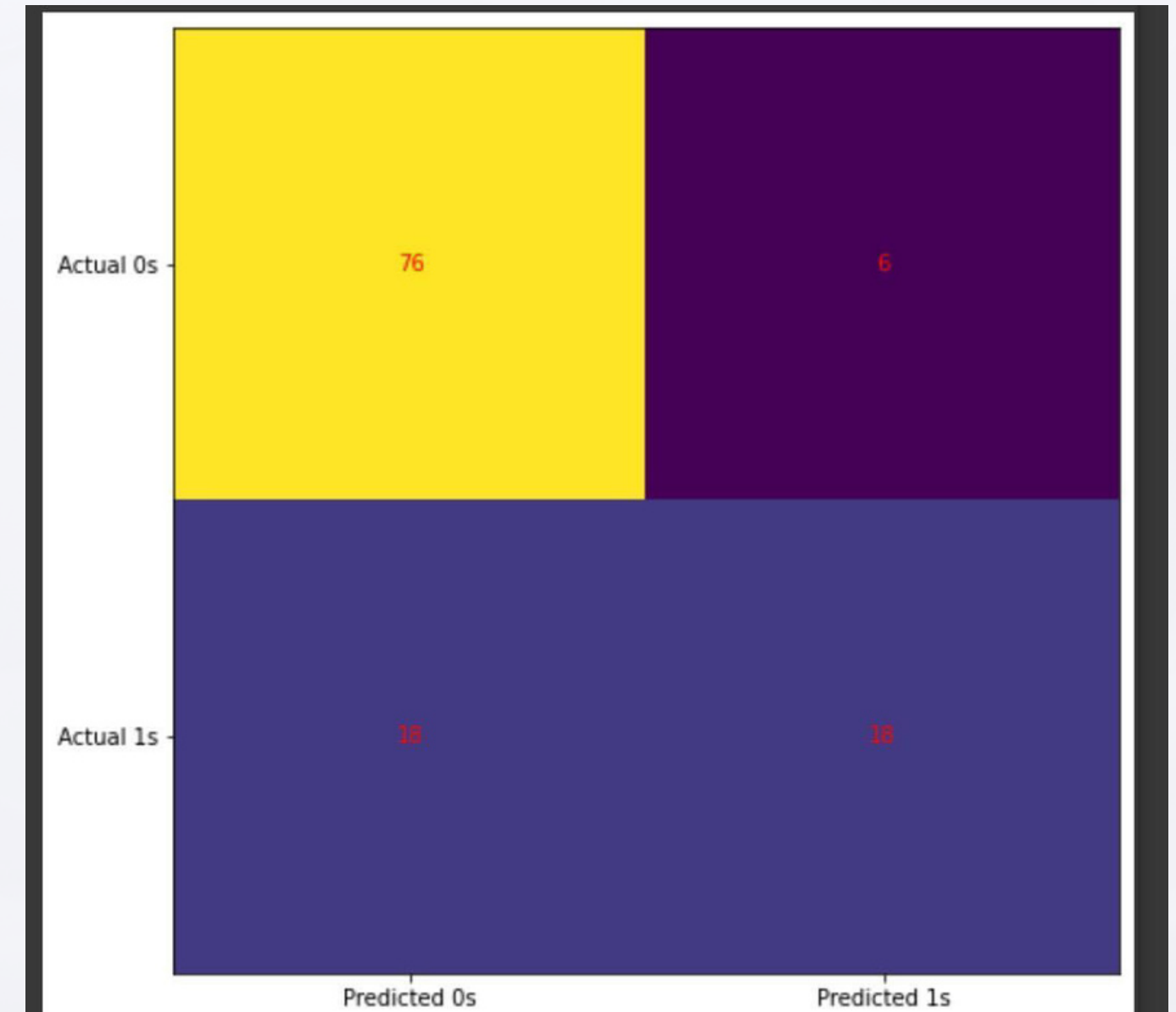
04

## Evaluation

Accuracy: 80 %  
True:94 Wrong:24

	precision	recall	f1-score	support
0	0.81	0.93	0.86	82
1	0.75	0.50	0.60	36
accuracy			0.80	118
macro avg	0.78	0.71	0.73	118
weighted avg	0.79	0.80	0.78	118

0 - ไม่เป็นโรคเบาหวาน  
1 - เป็นโรคเบาหวาน



# Evaluation

## Modeling analysis

- จากการทำ Modeling ได้ผลสรุปว่า

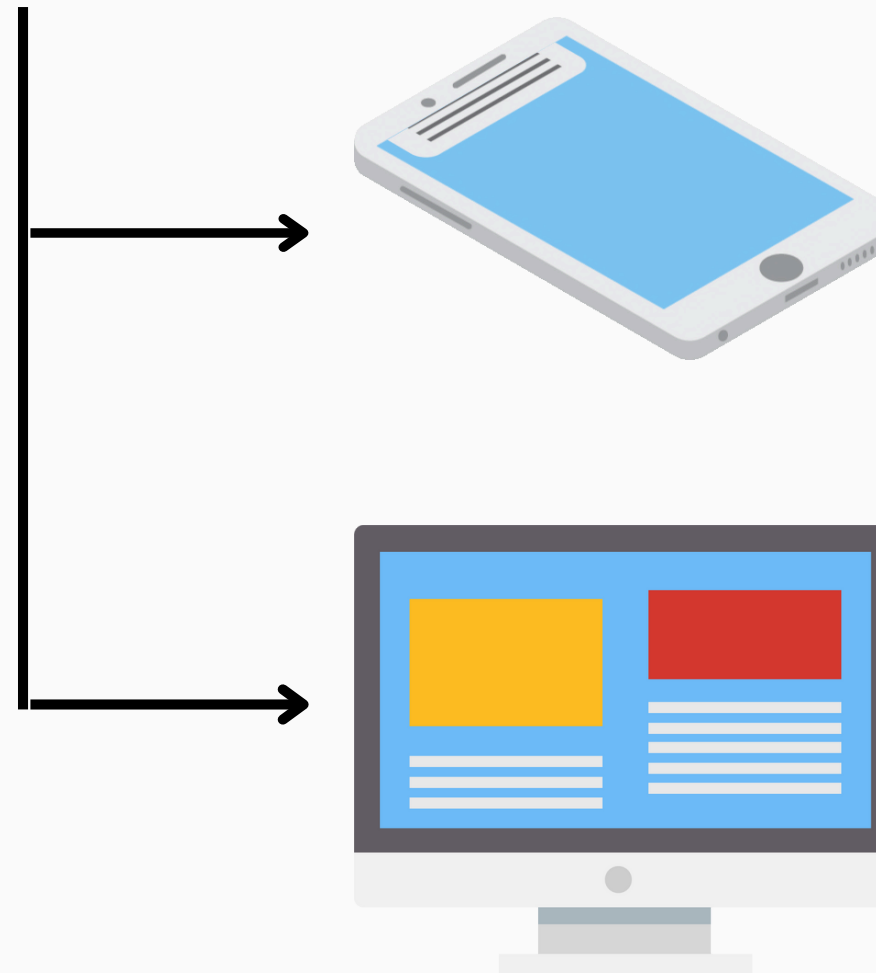
Model	Accuracy
Decision Tree Classifier	81.00%
K-Nearest Neighbor	77.96%
Decision Tree Classifier	69.08%
LogisticRegression	80.00%

	precision	recall	f1-score	support
0	0.81	0.93	0.86	82
1	0.75	0.50	0.60	36
accuracy			0.80	118
macro avg	0.78	0.71	0.73	118
weighted avg	0.79	0.80	0.78	118

	precision	recall	f1-score	support
0	0.86	0.85	0.86	82
1	0.68	0.69	0.68	36
accuracy			0.81	118
macro avg	0.77	0.77	0.77	118
weighted avg	0.81	0.81	0.81	118

## DecisionTreeClassifier

# Deployment



Web Application / Website / Mobile Application

# Example Application

---



<https://www.figma.com/proto/rUGeuYIKJbLZHS8nXIWHWI/BMI-Cal?node-id=1%3A2&scaling=scale-down&page-id=0%3A1&starting-point-node-id=1%3A2>



## TEAM MEMBERS

- ปี่ใหม่
- วิน
- อาร์เธอร์
- เกรซ
- เอ็ม



# HAVE FUN WITH CODING

