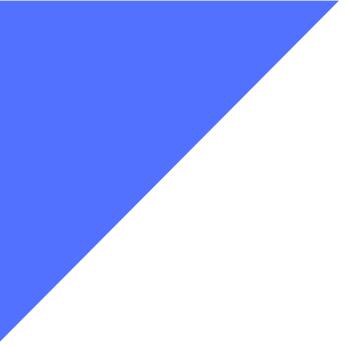


PYTORCH PRESENT

"DIABEST"





#CRISP-DM

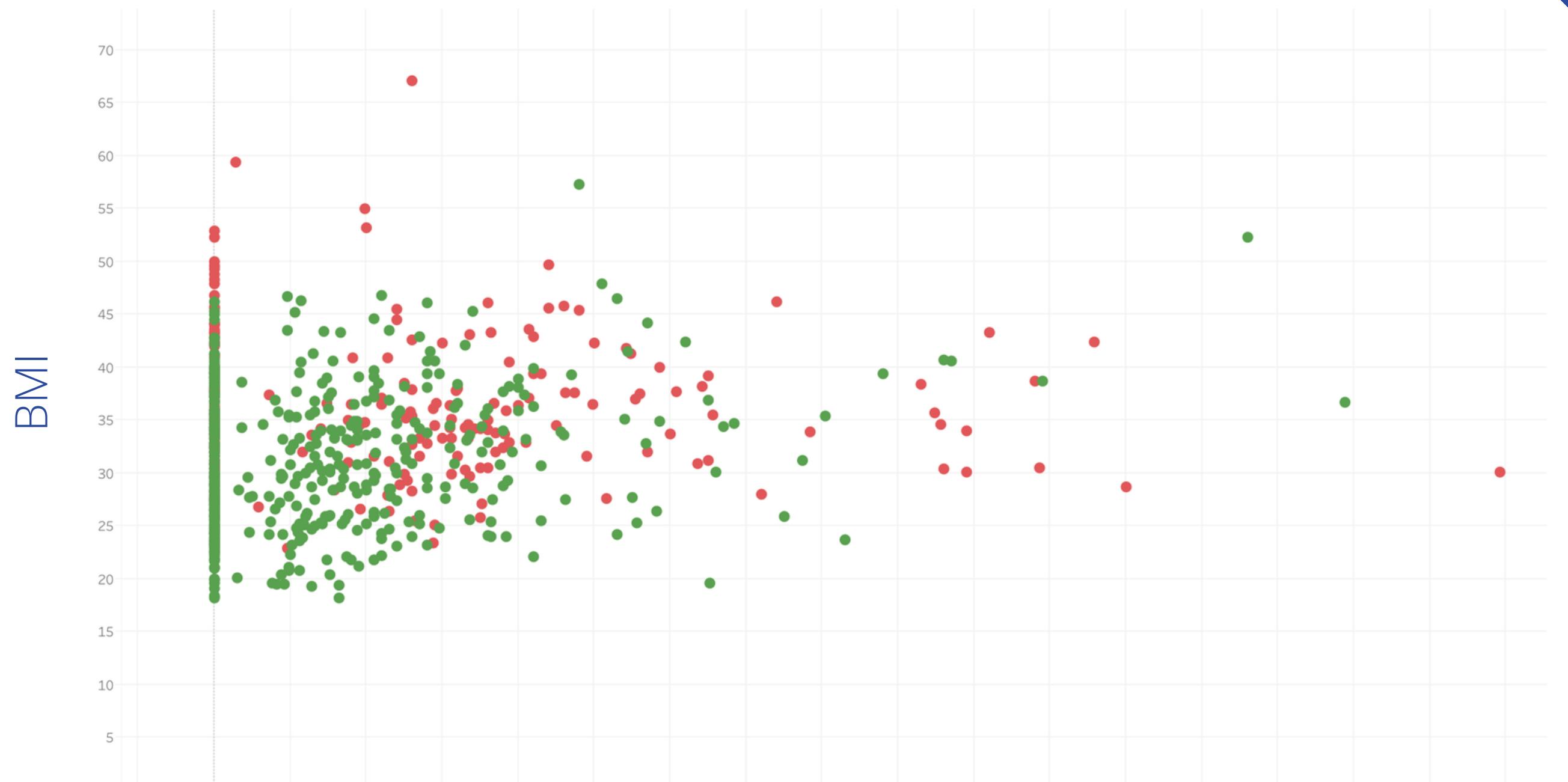
BUSINESS UNDERSTANDING

- Test kit ทั่วไปวัดได้เฉพาะ Glucose
- ไม่สามารถตรวจเป็นประจำได้
- การตรวจแบบ
ละเอียด มีค่าใช้จ่าย สูง



DATA UNDERSTANDING

แผนภาพแสดงการเปรียบเทียบระหว่างฮอร์โมนอินซูลินกับค่า BMI



ប្រព័ន្ធអង់គំ DATA CLEANING

Glucose	5
BloodPressure	35
SkinThickness	227
Insulin	374
BMI	11
DiabetesPedigreeFunction	0
Age	0
Pregnancies	0
Outcome	0

Data Preparation

```
Drop columns which feature importance < 10%  
df_dropfea = raw.drop(['Age','BMI','DiabetesPedigreeFunction','Glucose'],axis=1)  
  
Pregnancies BloodPressure SkinThickness Insulin Outcome  
0 6 72 35 0 1  
1 1 66 29 0 0  
2 8 64 0 0 1  
3 1 66 23 94 0  
4 0 40 35 168 1  
... ... ... ... ...  
763 10 76 48 180 0  
764 2 70 27 0 0  
765 5 72 23 112 0  
766 1 60 0 0 1  
767 1 70 31 0 0  
768 rows x 5 columns
```

Drop feature importance < 10%

```
drop insulin  
df_dropinsu=raw.drop(['Insulin'],axis=1)  
df_dropinsu  
  
Pregnancies Glucose BloodPressure SkinThickness BMI DiabetesPedigreeFunction Age Outcome  
0 6 148 72 35 33.6 0.627 50 1  
1 1 85 66 29 26.6 0.351 31 0  
2 8 183 64 0 23.3 0.672 32 1  
3 1 89 66 23 28.1 0.167 21 0  
4 0 137 40 35 43.1 2.288 33 1  
... ... ... ... ... ... ...  
763 10 101 76 48 32.9 0.171 63 0  
764 2 122 70 27 36.8 0.340 27 0  
765 5 121 72 23 26.2 0.245 30 0  
766 1 126 60 0 30.1 0.349 47 1  
767 1 93 70 31 30.4 0.315 23 0  
768 rows x 8 columns
```

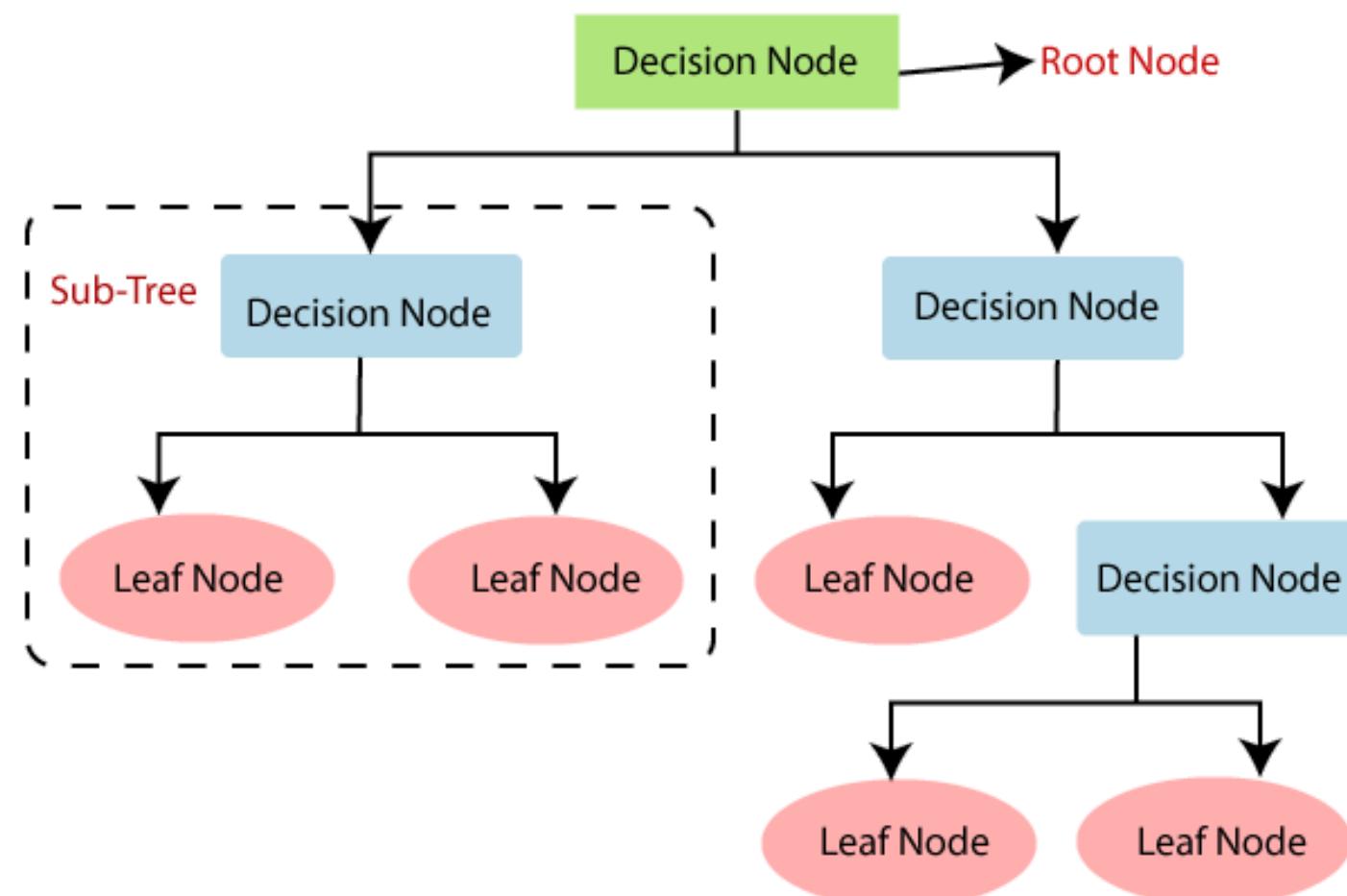
Drop Insulin

```
drop all NaN value  
df_dropna = df2.dropna()  
df_dropna  
  
Glucose BloodPressure SkinThickness Insulin BMI DiabetesPedigreeFunction Age Pregnancies Outcome  
3 89.0 66.0 23.0 94.0 28.1 0.167 21 1 0  
4 137.0 40.0 35.0 168.0 43.1 2.288 33 0 1  
6 78.0 50.0 32.0 88.0 31.0 0.248 26 3 1  
8 197.0 70.0 45.0 543.0 30.5 0.158 53 2 1  
13 189.0 60.0 23.0 846.0 30.1 0.398 59 1 1  
... ... ... ... ... ... ...  
753 181.0 88.0 44.0 510.0 43.3 0.222 26 0 1  
755 128.0 88.0 39.0 110.0 36.5 1.057 37 1 1  
760 88.0 58.0 26.0 16.0 28.4 0.766 22 2 0  
763 101.0 76.0 48.0 180.0 32.9 0.171 63 10 0  
765 121.0 72.0 23.0 112.0 26.2 0.245 30 5 0  
392 rows x 9 columns
```

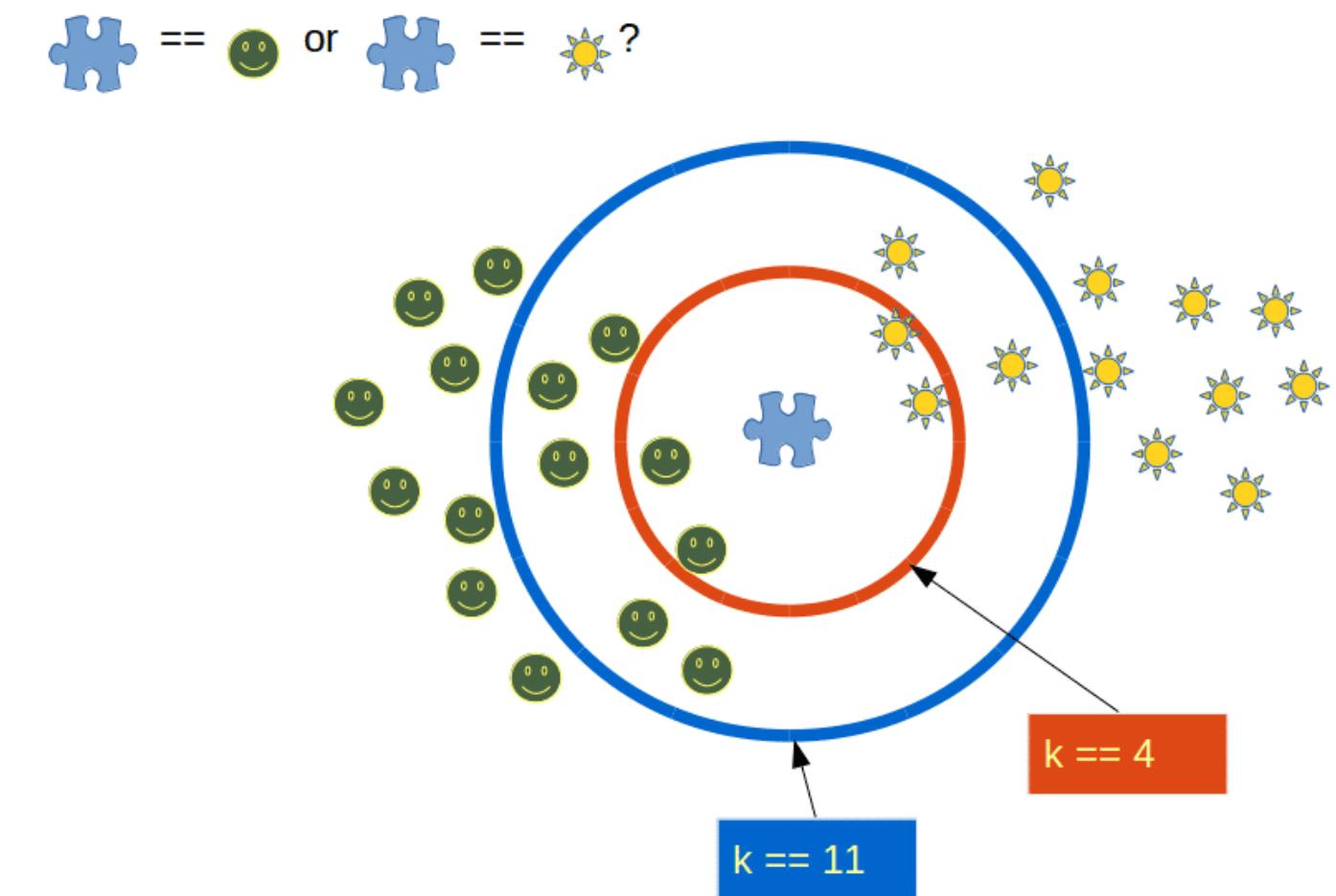
Drop NaN Value

MODELING

CLASSIFICATION



DECISION TREE



KNN

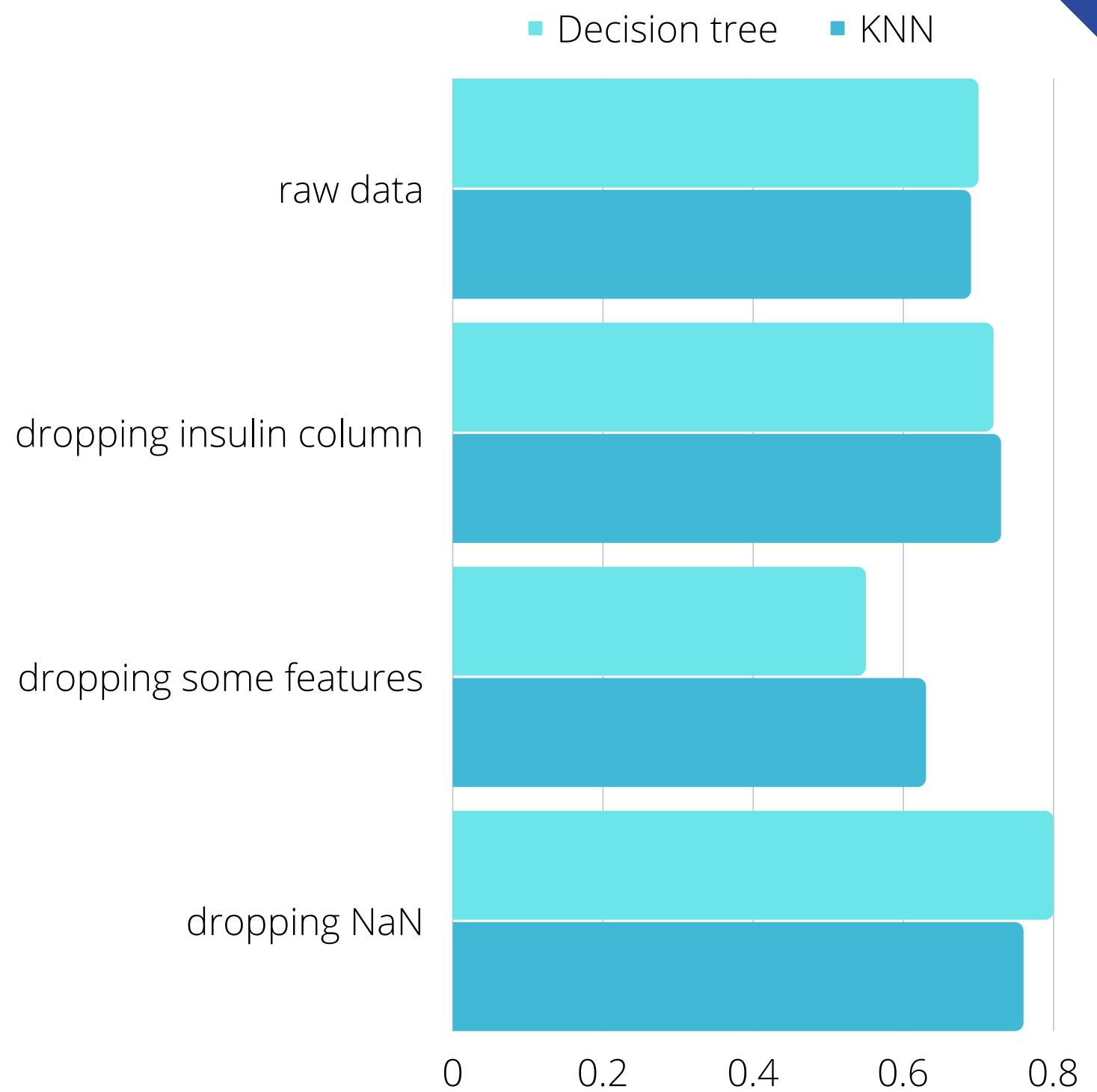
EVALUATION

DECISION TREE - ACCURACY

- raw data : 0.70
- drop insulin : 0.72
- drop columns which feature importance < 10% : 0.55
(เหลือ Glucose, Diabetes, BMI, Age)
- drop NaN : 0.80 (392 rows)

KNN - ACCURACY

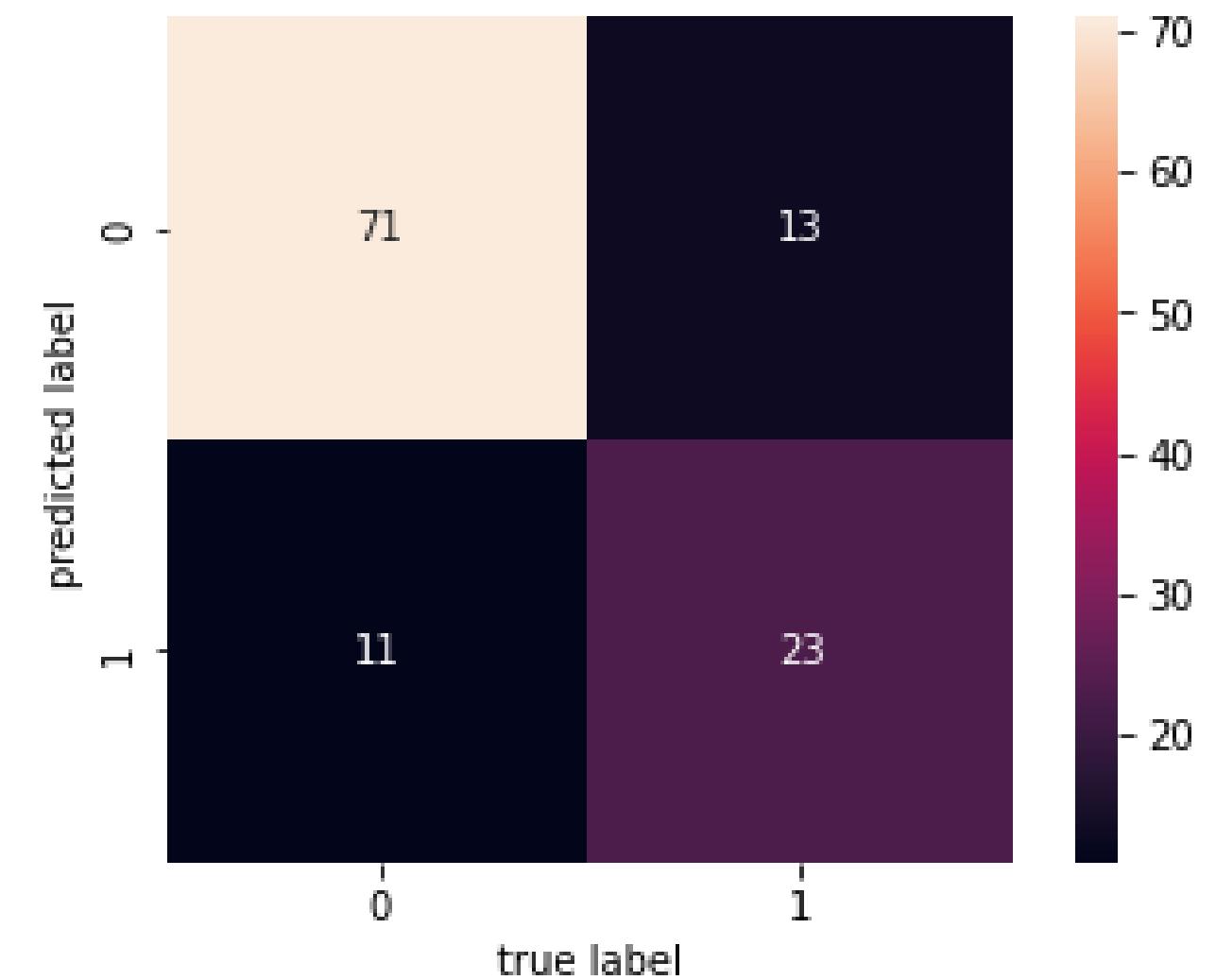
- raw data : 0.69
- drop insulin : 0.73
- drop columns which feature importance < 10% : 0.63
- drop NaN : 0.76 (392 rows)



DEPLOYMENT

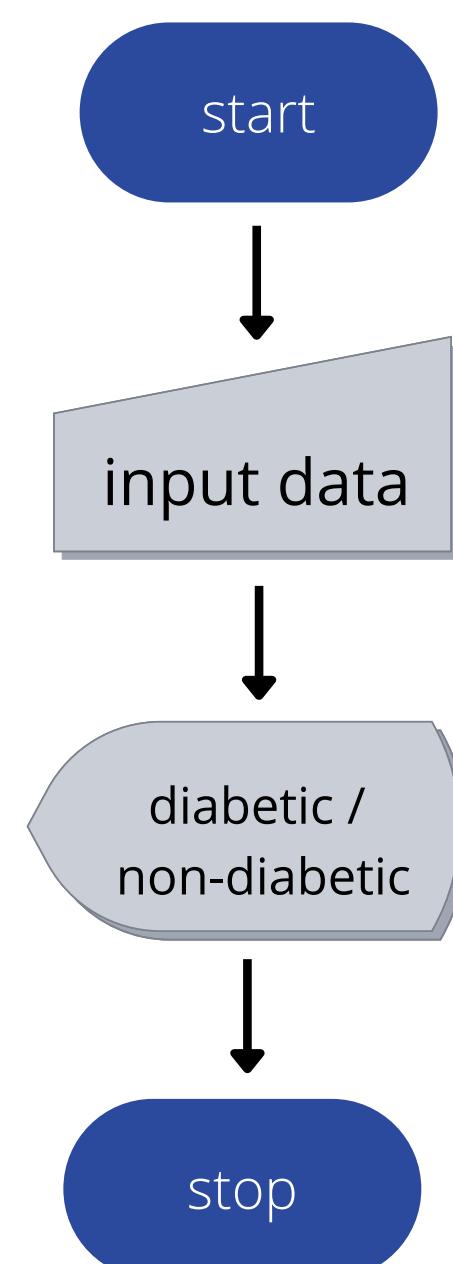
DECISION TREE - DROP NAN

	precision	recall	f1-score	support
0	0.85	0.87	0.86	82
1	0.68	0.64	0.66	36
accuracy			0.80	118
macro avg	0.76	0.75	0.76	118
weighted avg	0.79	0.80	0.79	118



MODEL ของเราตอบໄວຍ້ ລູກຄ້າຍ່າງໄຣ?

"ລູກຄ້າຈະໄດ້ຮັບການວິນິຈອັຍໂຮຄເບາ
ຫວານຈາກຂໍ້ມູນທີ່ປ່ອນ"



REFERENCES

- [HTTPS://WWW.HEALTHITANSWERS.NET/WHY-CLEANING-UP-YOUR-DIRTY-DATA-IS-CRITICAL-TO-CLINICAL-INTEROPERABILITY-AND-FINANCIAL-SURVIVAL/](https://www.healthitanswers.net/why-cleaning-up-your-dirty-data-is-critical-to-clinical-interoperability-and-financial-survival/)
- [HTTPS://WWW.ASC.OHIO-STATE.EDU/GOEL.1//STATLEARN/PROJECTS/PRESENTATIONS/DIABETES_PIMAINDIANS.PDF](https://www.asc.ohio-state.edu/goel.1//statlearn/projects/presentations/diabetes_pimaindians.pdf)
- [HTTPS://SCI-HUB.SE/10.2337/DIACARE.12.5.309](https://sci-hub.se/10.2337/DIACARE.12.5.309)





Thank you

Research + Canva

Coding

Coding

Research + Decorate

Canva

