



# APPENDICITIS PREDICTION

**METHODOLOGY CRISP-DM**



Pisit Janthawee (Data Scientist)

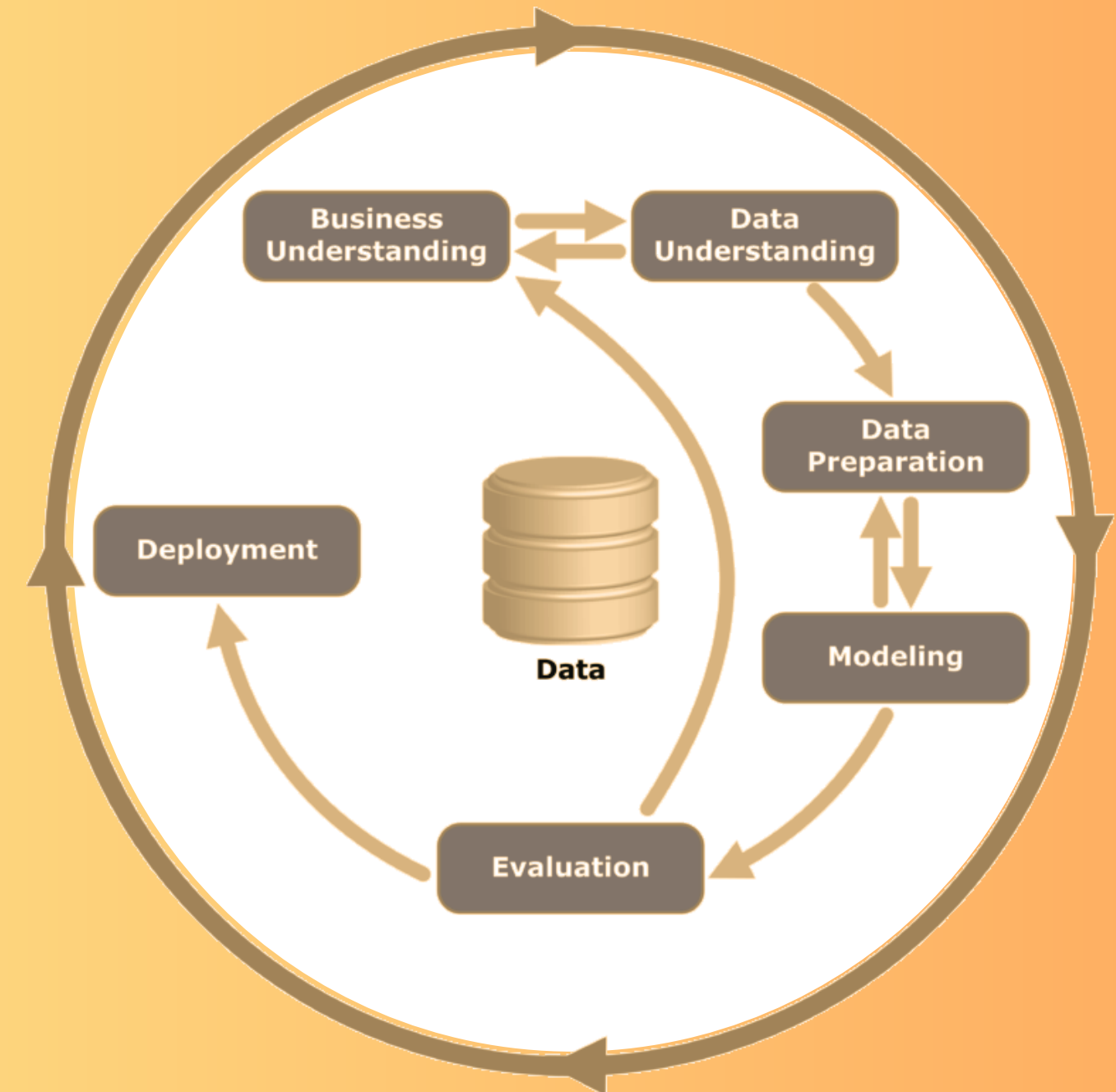
Bachelor of Science in Computer  
Science, Bangkok University

Created Date: 29 May 2023

# METHODOLOGY

## CRISP-DM

(CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING)



# BUSINESS UNDERSTANDING



OVERLAB!

**HARD TO DIAGNOSE!**  
**VARIOUS LABS TO DO**  
**HIGH COST**



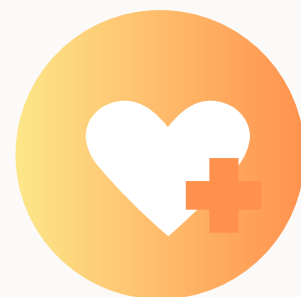
**PROBLEM?**

Appendicitis can present with a wide range of symptoms, making it difficult to accurately **diagnose** without additional medical tests.



**BUSINESS TYPE**

Healthcare, Medical



**FAST DIAGNOSE**  
**FAST TREATMENT**

- **Medical personnel**
  - Doctor diagnoses faster
  - Reduce diagnosis time
  - Diagnosed at the last delayed or missed, which can lead to complications
- **Patient**
  - Reduce the chance of getting into complications.
  - Helps reduce costs for patients
  - Blood, urine, X-ray
  - Save time
  - Reduce unnecessary steps



**GOAL**

- Prediction Tool for appendicitis suspecting
- Decision support system
- Improve healthcare outcomes

# DATA UNDERSTANDING

01

## VERIFY DATA QUALITY

How clean/dirty/missing is the data? Document any quality issues.

02

## DATA

Info: 500 rows x 23 columns

Target: Disease = 1, No disease = 0

Contains: Demographic, Laboratory, CC (Chief Complaint)

Nation: Thai

03

## ANALYSIS

- What are the factors of disease?
- The relationship between related disease A and disease B leading to the Main disease?
- Recurrence?

04

## TARGET

Classes

Disease

1

No disease

0



# DATA PREPARATION

## SELECT DATA

Appendicitis is among the commonest childhood diseases, between 10 and 19 years of age.



## CLEAN DATA

- Remove laboratory test results from erroneous values.
- Data Imputation



## CONSTRUCT DATA

Create related conditions from Text data (Chief complaint)

Migration of pain

Rebound tenderness

Cough tenderness

Dysuria

Stool

Tenderness in right lower quadrant (RLQ)

Nausea/vomiting

Anorexia

AS / PAS Score

# MODELING



# 01

## DIMENSIONALITY REDUCTION

What is the factor of disease



# 02

## OVER-SAMPLING

SMOTE (Synthetic Minority Over-sampling Technique)

Solved: Class imbalance

After performing SMOTE oversampling, Will the correlation change?

# 03

## SELECT MODELING

- Ensemble Methods:
  - Random Forest
  - Gradient Boosting Classifier
  - AdaBoost Classifier
- Linear Models:
  - Logistic Regression
  - GaussianNB (though it's technically a Naive Bayes classifier, it falls under this category due to its linear nature)
- Tree-based Models:
  - Decision Tree Classifier

- Nearest Neighbors:
  - KNeighbors Classifier
- LightGBM and XGBoost:
  - LGBM Classifier
  - XGB Classifier
- Dummy Classifier:
  - Dummy Classifier (most frequent strategy)



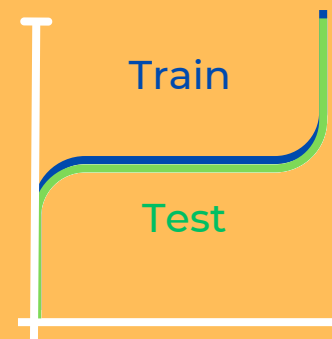
# EVALUATION

N

Number of  
features

Factors

Helps reduce costs for patients



Learning  
curve

Generalization ability

Can model able to predict on Unseen data?



Sensitivities,  
Specificities,  
Accuracy,  
AUROC

Ensure the reliability

Diagnostic accuracy, patient safety,  
treatment decisions



# DEPLOYMENT

User interface design Entering patient information is easy to use.

Appendicitis Prediction

Alvarado and Pediatric Appendicitis scores are automatically calculated from your conditions.

Age

Sex ☐ Male ☒ Female

Height

Body Weight

Body Temperature

Alcohol ☒ No ☐ Yes

Smoking ☐ No ☒ Yes

Dysuria อาการปัสสาวะแสบขัด ปวดตอนปัสสาวะ? ☒ No ☐ Yes

Anorexia อาการเบื่ออาหาร? ☐ No ☒ Yes

Nausea/vomiting คลื่นไส้ / อาเจียน? ☒ No ☐ Yes

Rebound tenderness ความปวดของการกดหน้าท้อง ☐ No ☒ Yes

Peritonitis/abdominal guarding ผ่านหน้าปวดหน้าท้อง ☒ No ☐ Generalized ☐ Localized

Tenderness in right lower quadrant (RLQ) ความปวดหน้าท้องขวาล่าง ☐ No ☒ Yes

Cough tenderness อาการไอ? ☐ No ☒ Yes

Probability of disease

Disease 85%

No Disease 15%

Estimated

age	sex	height	body_weight	body_temperature	alcohol	smoking
18	female	160	50	36.5	no	yes

Seaborn Plot

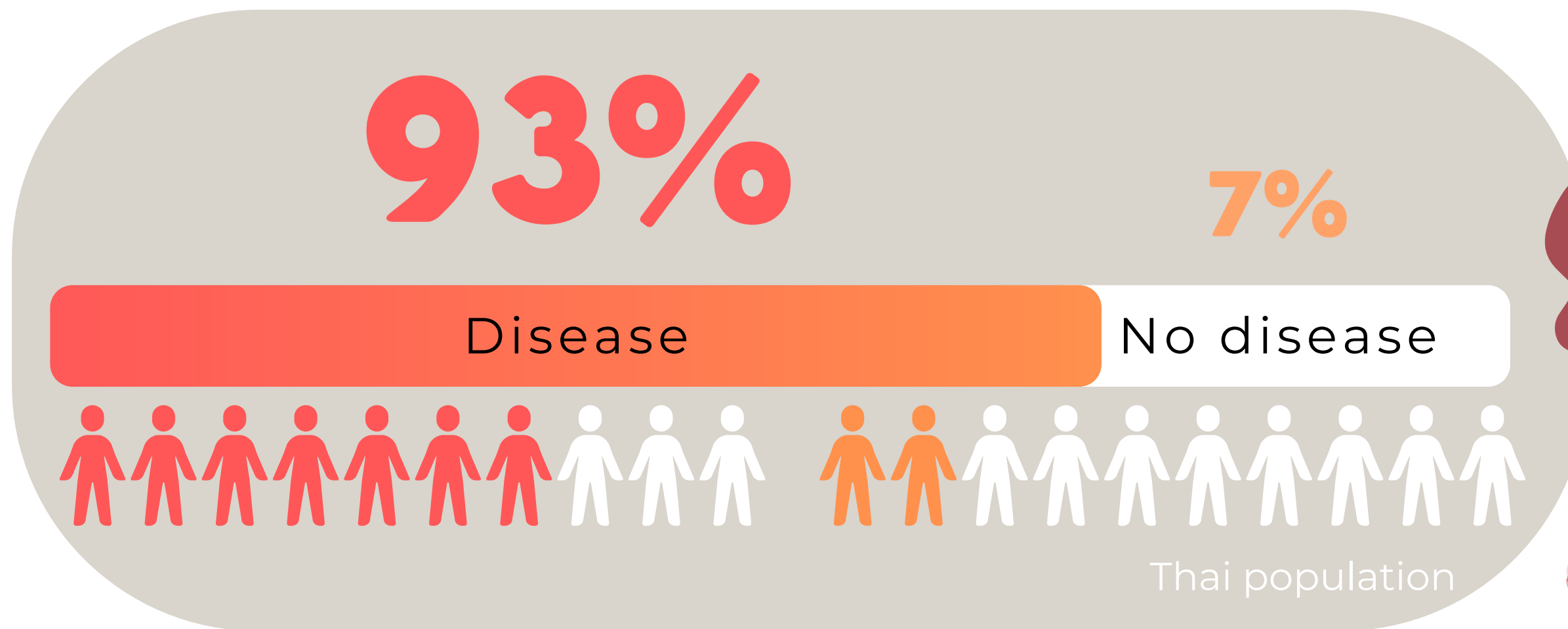
AlvaradoScore(A5) = 7 for Appendicitis Population

PediatricAppendicitisScore(PAS) = 6 for Appendicitis Population

Flag



# EXPECTED RESULT



↓

Treatment decisions  
Antibiotics / Surgical



# THANKS YOU

## CONTACT



099-465-5516



pisit.jant@gmail.com



Bangkok, Bangkok City,  
Thailand



Portfolio

<https://pisit-janthawee.mystrikingly.com/>



LinkedIn

<https://www.linkedin.com/in/pisit-janthawee/>



Github

<https://github.com/Pisit-Janthawee>



**Pisit Janthawee (Data Scientist)**  
Bachelor of Science in Computer Science,  
Bangkok University