



# Kaggle : ASA Classification

系所 : 大數據科技及管理研究所

年級 : 碩一

姓名 : 張哲安

教授 : 張詠淳教授



# 資料預處理

## 1 體重

0kg轉換為NA值

## 2 身高

身高單位轉換為公尺

## 3 麻醉方式

合併相同的麻醉方式類別





## 建立類別欄位

## 抗凝血劑、心律不整、心衰竭用藥

# 高血壓、糖尿病藥物

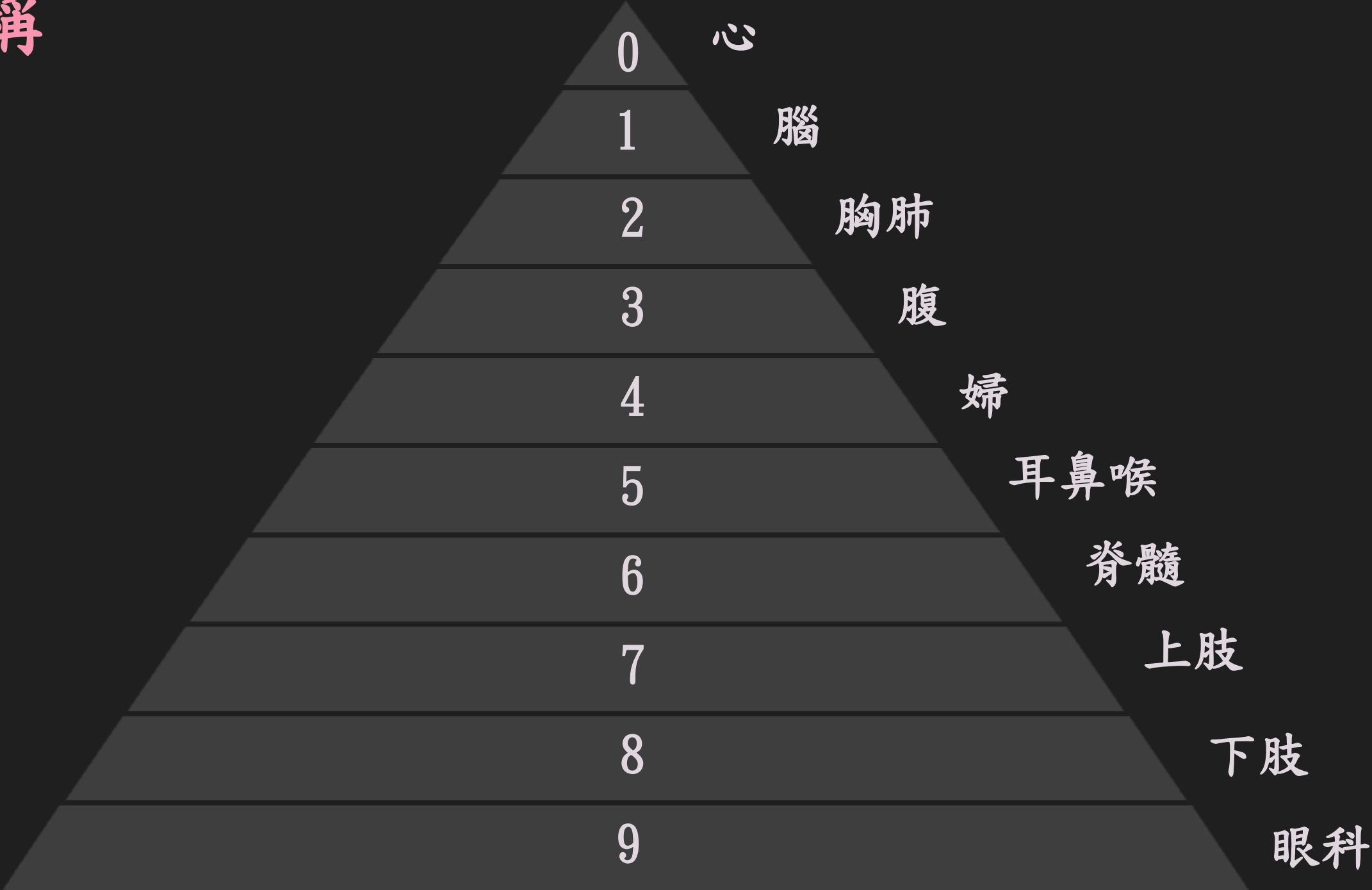
# 癲癇藥物

# 止吐藥物

# 嗎啡類止痛藥物

## 建立0-8的藥物使用組合

手術名稱



# 檢驗值

1

Troponin I, NA, K, WBC, HB, PLT,  
INR, CREA, O, GLU, CRP, ALB

2

用CREA、年齡、性別計算 eGFR

3

所有抽血是否為異常



# 身上管路及擺放位置

計算身上總管路數

PICC

CVC

Chest tube

Port-A

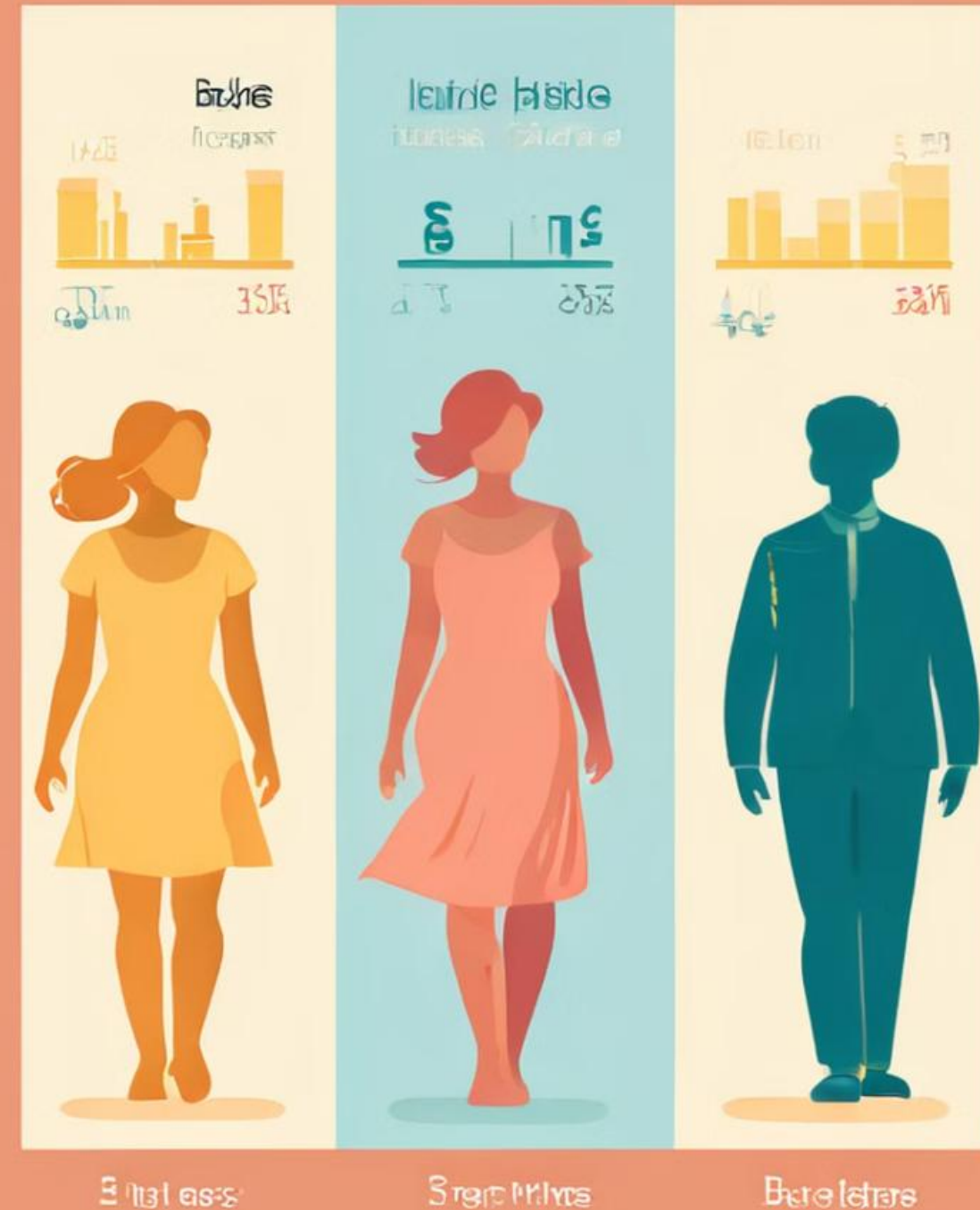
Sheath

Ileostomy

共 30 種

Baill

Baelbalhatt s



# 手術名稱、藥物使用、管路使用

相同內容合併

加總分別每250、1300、760進行類別分組

各組數量大約相等

空值或未出現在訓練資料中的數據歸類到最後一類



# 缺失值填補及BMI與年齡分組

KNN填補

$K = 5$  填補身高和體重

BMI分組

小於30

30-40

大於40

年齡分組

小於30

30-45

46-65

66-75

大於75歲





# 特徵轉換與標轉化

## Label Encoder

麻醉方式、是否為ICU

病人、病人來源進行

Label Encoder轉換，把

空值當作一類

## Target Encoder

所有類別特徵進行

Target Encoder

## StandardScaler

所有特徵進行標準化

# 模型訓練策略

StratifiedKFold : 切 10 份

權重調整 : 每類balance

BayesSearchCV : 250組參數

特徵重要性篩選 : 篩選前80%重要性特徵

# Catboost

- 空值、類別特徵多、文本

- Categorical to Numerical

- Ordered Target Encode

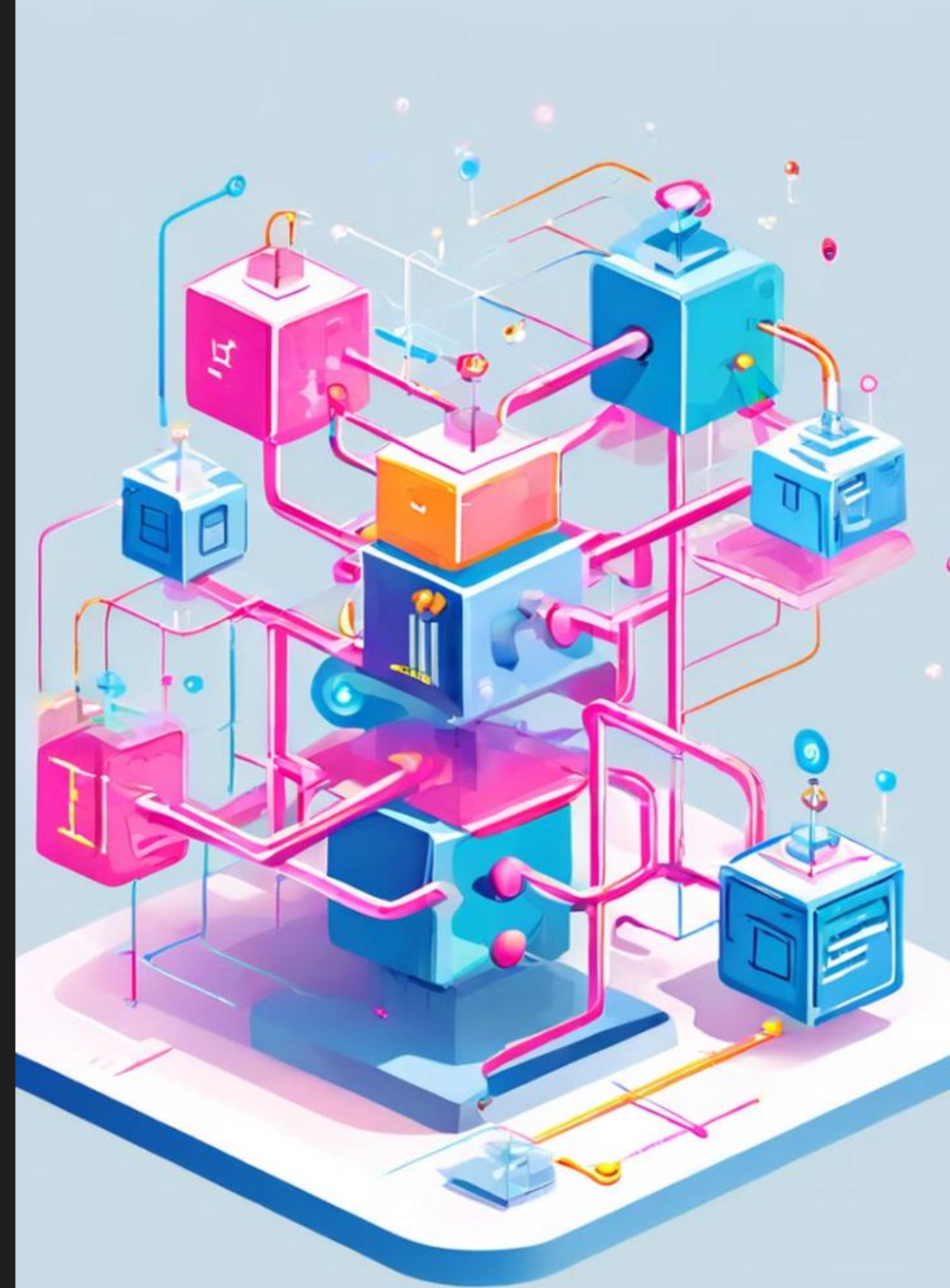
- 看n-1筆資料進行數值轉換

- Text to Numerical

- BoW、NaiveBayes

- Symmetric Tree

提高模型泛化能力





# 最終重要性篩選後特徵

## ■ 文本

Lab\_Values  
properties\_display  
Surgery\_Name  
Medication\_Usage  
Catheter\_Use

## ■ 類別

Gender  
Patient\_Source  
Anesthesia\_Method  
ICU\_Patient

## ■ 數值

Age  
WEIGHT

## ■ 類別

AGE\_GROUP  
Surgery\_Label  
Medication\_Group  
Catheter\_Group  
medication\_category  
CREAT  
INR  
NA  
ESRD  
GLU  
WBC  
LAB\_OK

## ■ 數值

TOLTAL\_DISEASE

END