

# 生病小助手

組員：B0829005 黃淞蜂

B0829010 柯名軒

B0829016 邱弘璽

B0829025 林宥任

指導老師：吳世琳教授

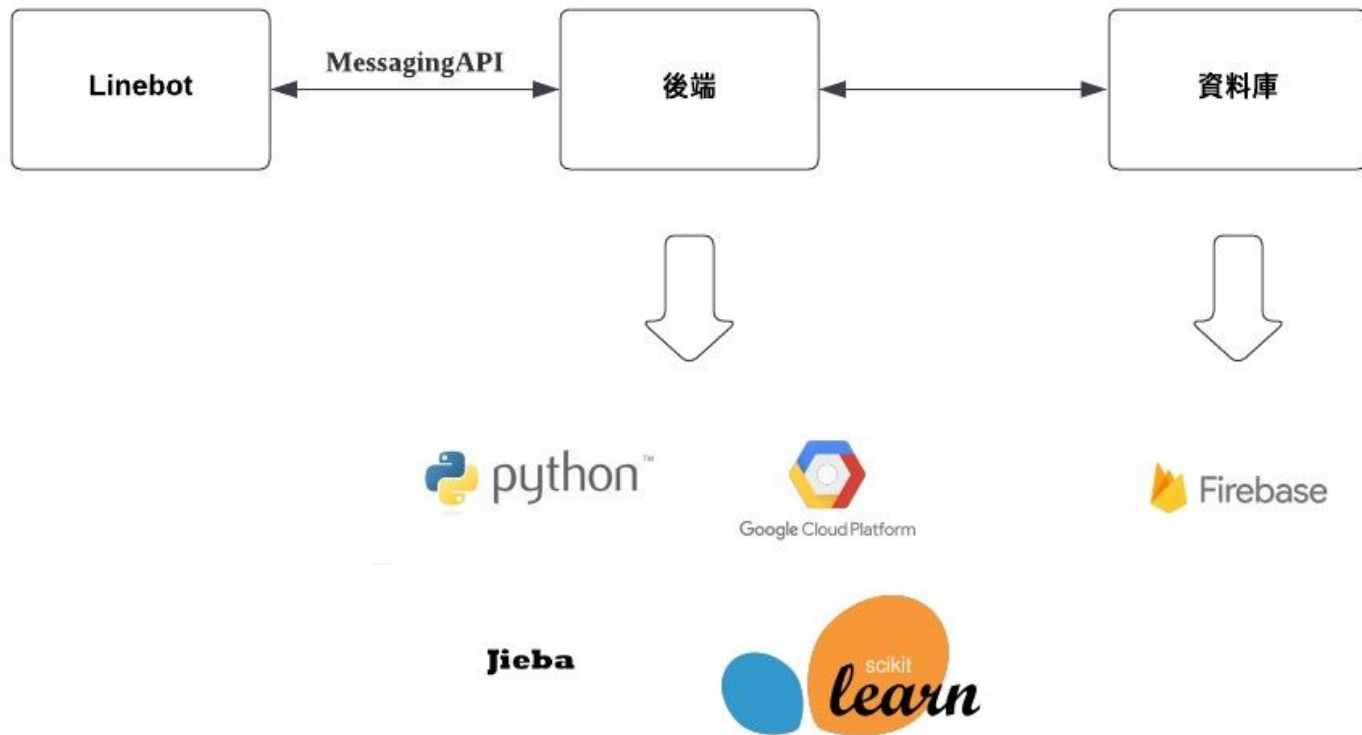
# 背景

在現今的疫情下，人們去往診所和醫院就診的意願大幅下降，因而選擇購買成藥來面對病情，也有人是懷著小病吃藥大病才需就診的心態，同時也有人常常因為不了解病情而跑錯診所致使耽誤就診，因此我們開發出了此套系統期望能幫助這些人來了解自己得了什麼病和就診什麼科別。

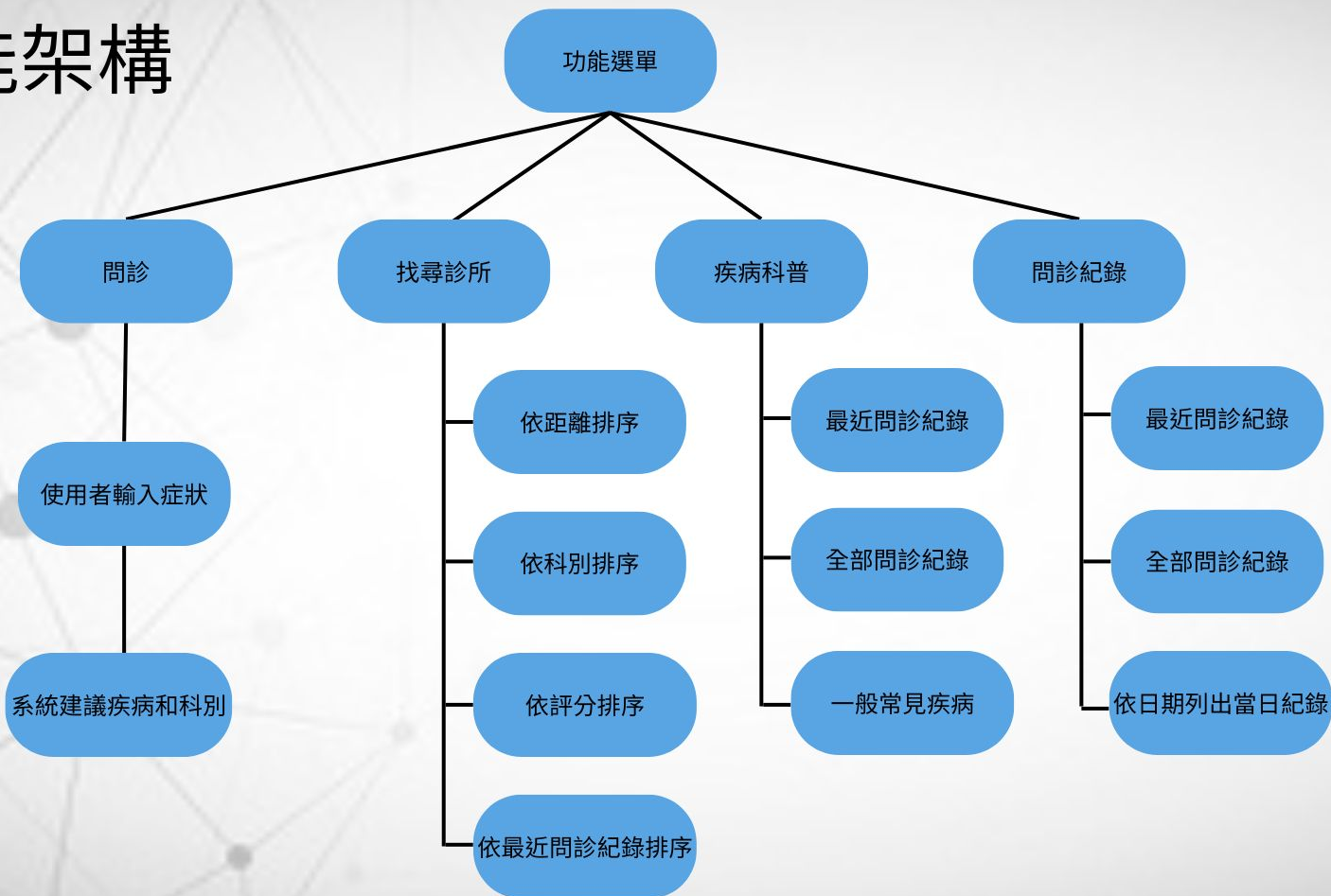
# 目的

透過使用LineBot來讓使用者能夠方便的透過Line得知自己的病症和建議科別，以便在疫情期間能讓使用者能降低出入醫院和診所的次數。

# 技術架構



# 功能架構



# 資料來源

Kaggle是全世界公認的資料集，其內容為疾病

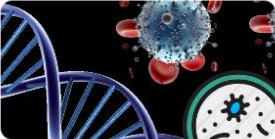
dataset.csv (632.2 kB)

DetailCompactColumn

10 of 18 columns

▲ Disease	▲ Symptom_1	▲ Symptom_2	▲ Symptom_3	▲ Symptom_4	▲ Symptom_5
Diseases that may be present	the symptoms experienced during the disease	the symptoms experienced during the disease	the symptoms experienced during the disease	the symptoms experienced during the disease	the symptoms experienced during the disease
41 unique values	vomiting 17% fatigue 14% Other (3408) 69%	vomiting 18% fatigue 8% Other (3648) 74%	fatigue 15% high_fever 7% Other (3870) 79%	high_fever [null] 8% Other (4194) 7% Other (4194) 85%	[null] 8% headache 7% Other (3408) 85%
Fungal infection	itching	skin_rash	dischromic_patches		
Fungal infection	itching	skin_rash	nodal_skin_eruptions		
Fungal infection	itching	skin_rash	nodal_skin_eruptions	dischromic_patches	
Allergy	continuous_sneezing	shivering	chills	watering_from_eyes	
Allergy	shivering	chills	watering_from_eyes		
Allergy	continuous_sneezing	chills	watering_from_eyes		
Allergy	continuous_sneezing	shivering	watering_from_eyes		
Allergy	continuous_sneezing	shivering	chills		
Allergy	shivering	chills	watering_from_eyes		
Allergy	continuous_sneezing	chills	watering_from_eyes		
Allergy	continuous_sneezing	shivering	watering_from_eyes		
Allergy	continuous_sneezing	shivering	chills		
Allergy	continuous_sneezing	shivering	chills	watering_from_eyes	
GERD	stomach_pain	acidity	ulcers_on_tongue	vomiting	cough
GERD	stomach_pain	ulcers_on_tongue	vomiting	cough	chest_pain
GERD	stomach_pain	acidity	vomiting	cough	chest_pain
GERD	stomach_pain	acidity	ulcers_on_tongue	cough	chest_pain
GERD	stomach_pain	acidity	ulcers_on_tongue	vomiting	chest_pain
GERD	stomach_pain	acidity	ulcers_on_tongue	vomiting	cough
GERD	acidity	ulcers_on_tongue	vomiting	cough	chest_pain
GERD	stomach_pain	ulcers_on_tongue	vomiting	cough	chest_pain
GERD	stomach_pain	acidity	vomiting	cough	chest_pain
GERD	stomach_pain	acidity	ulcers_on_tongue	cough	chest_pain
Chronic cholestasis	itching	vomiting	yellowish_skin	nausea	loss_of_appetite
Chronic cholestasis	vomiting	yellowish_skin	nausea	loss_of_appetite	abdominal_pain

古模型，都是使用此資



Usability ⓘ

9.71

License

CC BY-SA 4.0

Expected update frequency

Monthly

304/healthcare-chatbot

ct=Training.csv

and how rows and columns are

[e-symptom-description-dataset](#)

# 資料處理

我們將這筆資料整理成對應的表格，並以0、1表示疾病案例所擁有的症狀

未來資料要更新，會新增疾病案例，將其擁有的特徵值(症狀)設為1

[illegible]

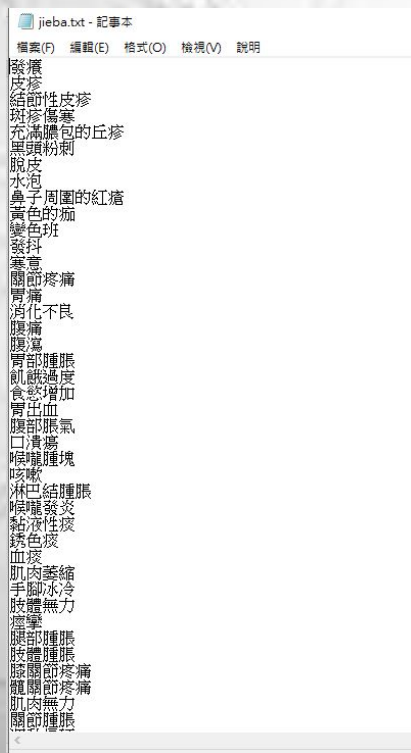


## 相關技術使用



# Jieba

透過自定義辭典來加入疾病相關詞彙並利用Jieba進行斷詞。



← 自定義辭典

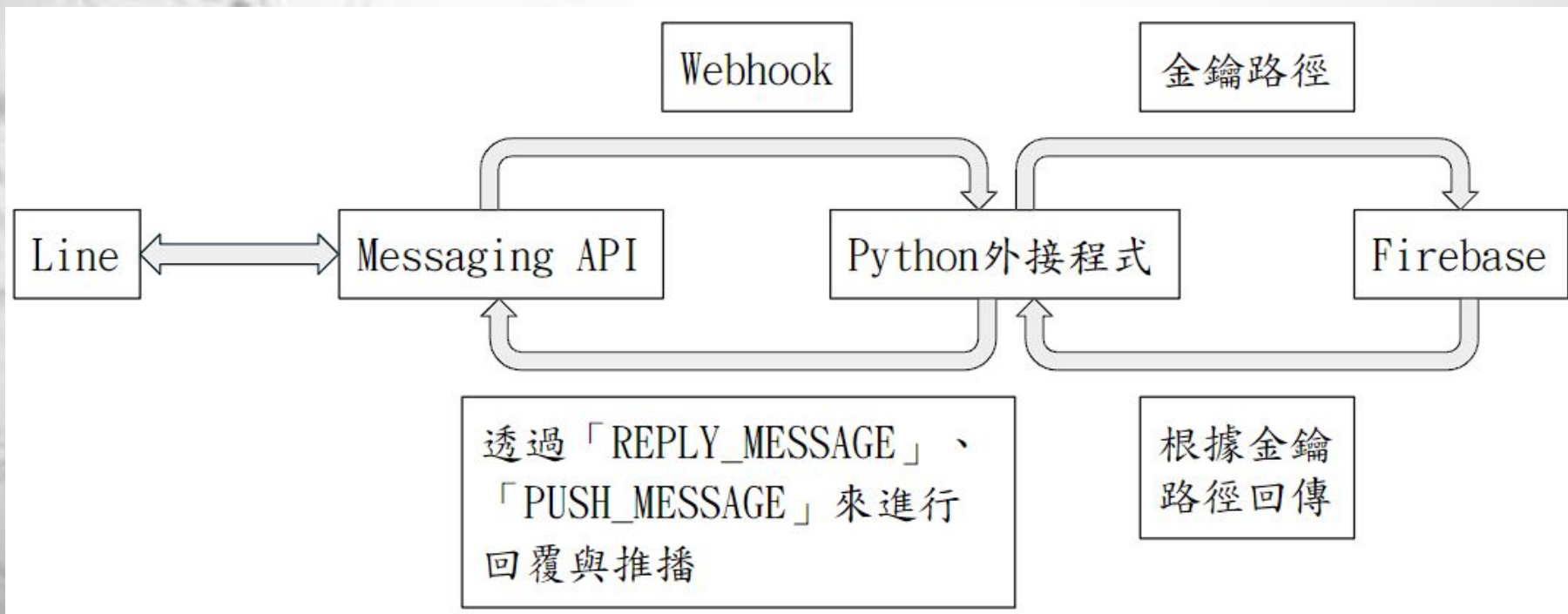
系統操作者可進入FTP後端dictionary資料夾中，並在userdict.txt、jieba.txt、同義詞.csv中加入新詞重新上傳後，就能辨識出新詞

```
----- word_segmentation -----  
['感覺', '有點', '頭痛', '和', '嘔吐']  
感覺 有點 頭痛 和 嘔吐
```

← 斷詞結果

# LineBot 連接後台進行疾病推測

LineBot為使用介面主體，與使用者對接，透過問答來獲取所需的資料。



# Firebase

資料庫, 儲存每個用戶的問診相關訊息。

使用者的UserID

問診時間

當次問診的內容

clinic-smart	U20d91dd5d33bab6b6d91db43c2a8...	2022-11-30 14:56:56.251926
+ 新增集合	+ 新增文件	+ 新增集合
U13c495274e0575e5c103e90799292...	2022-11-30 14:50:39.665673	+ 新增欄位
U145ce5af8a265edc9dd638c08e195...	2022-11-30 14:56:56.251926 >	一開始輸入訊息: "頭痛"
U20d91dd5d33bab6b6d91db43c2a84... >	2022-11-30 14:58:26.329404	使用者Line ID: "U20d91dd5d33bab6b6d91db43c2a84344"
	2022-11-30 14:58:46.202337	問診時間: "2022-11-30 14:56:56.251926"
	2022-11-30 14:59:47.446835	推薦科別: "耳鼻喉科"
	2022-11-30 15:00:58.799451	是否進行二分法: "True"
	2022-11-30 15:03:14.174244	最終判斷疾病: "普通感冒"
	2022-11-30 15:03:47.501326	模糊比對後的詞: "['頭痛']"
	2022-11-30 16:33:54.094516	



問診紀錄



# 關鍵性技術

# 伯努利貝氏分類器

用在二元的特徵，比方說特徵是否出現、特徵大小、特徵長短等等這種二元的分類。模型的特徵值只能是 1 和 0 的二元特徵。

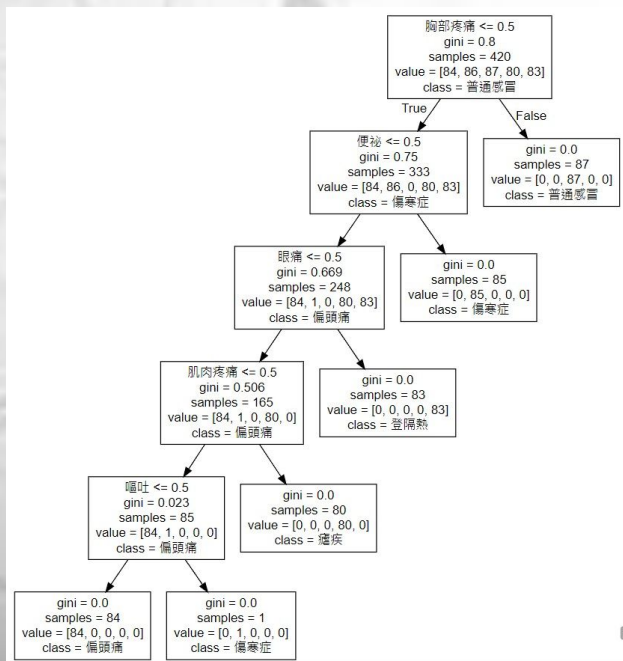
以貝氏定理為基礎，將多筆疾病的資料，其症狀為特徵值，計算出在特定的症狀下，為所有疾病的機率，再挑選出機率最高的前幾項疾病。

給定症狀下疾病的機率 =  $P(\text{疾病} | \text{特徵向量(症狀)}) =$

$$P(y | x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^n P(x_i | y)}{P(x_1, \dots, x_n)}$$

# 決策樹

輸入由貝氏分類器計算出的前幾項最高機率的疾病，從整理好的資料中找出每筆資料的特徵，並透過其特徵將其分支，直到有單一疾病能夠透過特徵被區分。



症狀 $\leq 0.5$ :判斷症狀是否存在，來決定它的走向

gini:判斷每個症狀之間是否為平均分布，越小越平均，反之則越亂

sample:為value總和，即樣本數

value:每個疾病的對應樣本數

class:疾病名稱



# DEMO展示

## 結論

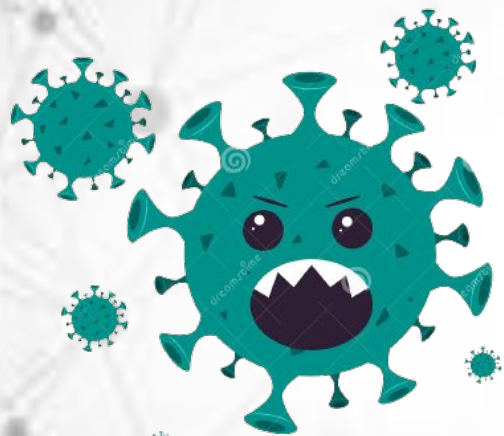
1. 目前市面上大多的問診系統多採用篩選的方式來得出結果，而本系統採用計算特定症狀下各種疾病的機率去告知使用者，相較於市面上的系統來說，我們所得出的疾病較能讓使用者進行判斷。
2. 由於本系統採用的資料集為國外的資料集。未來如果有機會和台灣的醫院或診所合作，能夠拿到更加詳細的資料後，可以進一步的提升系統的準確性，也能使本系統更加的貼近台灣人普遍的情況。
3. 本系統在提供疾病預估的同時，還記錄了使用者曾輸入過的症狀，往後如果有機會跟醫院或是診所合作，亦可以提供在醫生看診前，將事先記錄好的問診紀錄傳送到醫生端，讓醫生能盡早了解病人的症狀。



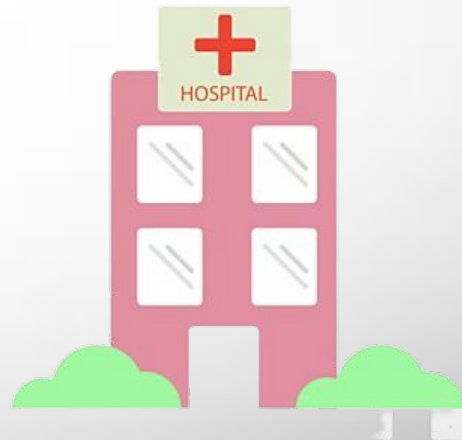
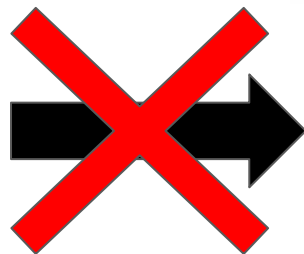


Thanks For Your Listening

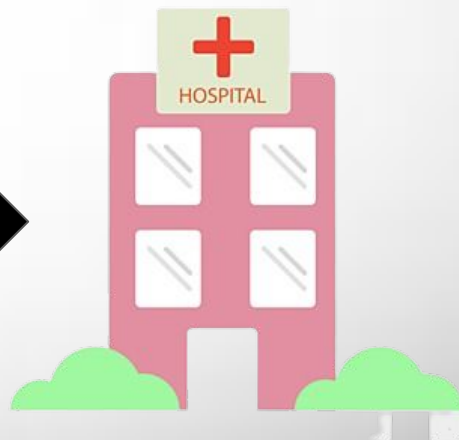
# 背景



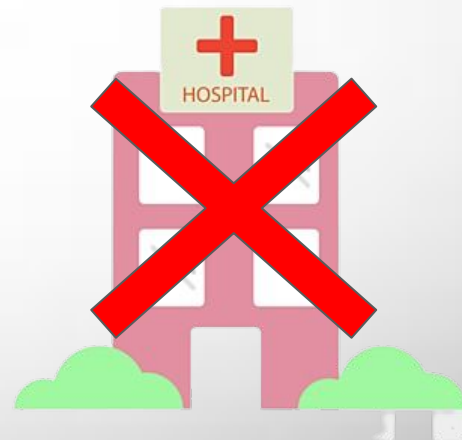
**COVID-19**



# 背景



背景



# 技術架構

