

# 以機器學習預測氣喘病患未來氣喘狀況之App

## An app for predicting asthma patient's future condition with machine learning

資工4A 陳韋霖 資工4A 陳昱銘

指導教授：曾新穆

### Introduction

氣喘是一種因體質或外在刺激導致的慢性呼吸道疾病，發作可能非常快速且危險，故患者以尖峰呼氣流速計做規律記錄是監控病情的重要方法，也是醫師臨床診斷的重要參考。本專題以：

- 一、氣喘的突發危險性
- 二、不易考量的環境刺激因素
- 三、傳統紙筆紀錄的不便

為出發點，開發出一款方便記錄生理數據並加入環境因子，以機器學習預測未來氣喘病況的應用程式。

### About Data

- 一、病患生理資料：  
尖峰呼氣流量測量值、正常值、各種身體症狀、用藥及治療情形、基本資料。
- 二、天氣資料：  
平均氣溫、最高溫、最低溫及平均濕度
- 三、空汙資料：  
各種空污指標物質測量數據
- 四、預測指標：

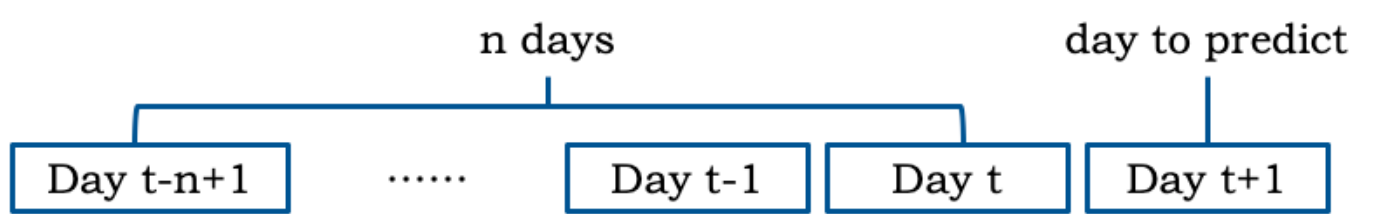
- 綠燈區  $PEFR > 80\%$

黃燈區  $80\% > PEFR > 60\%$

紅燈區  $PEFR < 60\%$

### Methodology

根據PEFR值區間，綠燈區與黃燈區為一類，紅燈區為一類，為二分類問題。根據專業醫師指出，氣喘較有效之觀察天數區間為3至5天，意即氣喘病患鄰近天數內的情形有一定的相關性，故我們使用不同連續天數的資料集，實驗前n天資料預測下一天。最後以不同n的設定、不同的不平衡資料處理方法、不同模型選擇，以f1-score為評估指標進行實驗比較。



imbalance	n	Decision Tree	Random Forest	Logistic Regression	SVM
None	1	0.517	0.493	0.694	0.493
	2	0.712	0.493	0.829	0.493
	3	0.559	0.494	0.690	0.494
SMOTE	1	0.559	0.594	0.628	0.655
	2	0.606	0.630	0.802	0.637
	3	0.520	0.548	0.658	0.588
SMOTE + ENN	1	0.561	0.671	0.598	0.649
	2	0.580	0.629	0.759	0.634
	3	0.506	0.545	0.654	0.590
Adjust weights	1	0.528	0.493	0.575	0.616
	2	0.598	0.493	0.786	0.613
	3	0.512	0.494	0.654	0.564

根據上表實驗數據，選擇綜合表現最佳的羅吉斯回歸及支持向量機模型，以n等於2，不平衡處理採用SMOTE的設定再進行優化，評估指標則是PEFR紅燈區的召回值（recall）。

	Logistic Regression	SVM
recall	0.700	0.800
f1-score	0.913	0.900

最後根據實驗結果，在f1-score沒有顯著差異下，選擇召回值較高的支持向量機模型作為最終app上部署的模型。

### App

主要架構為Home按鈕及四個主要功能按鈕。

- 一、個人資料設定：基本資料填寫。
- 二、生理資料輸入：昨日及今日的生理資料填寫。
- 三、氣喘情況預測：進行預測並顯示結果。
- 四、天氣空汙查看：顯示氣溫、濕度、空污指標。

