

# 金融科技

## 第 13 組\_final\_project

### 一、 動機

在本學期上課時，教授有提到，在 Fintech 領域上，原先人們認為會最先對股票或期貨等金融業有所衝擊，然而，保險業才是首當其衝。

在保險業領域中，機器學習的應用雖沒有金融業如此直觀（預測賺或賠），但能夠應用的方面非常廣泛。包含上課提到的發生意外之後，意外險透過車禍畫面評估肇事原因，或是預估風險來制定保險金額，以客觀且固定的模型來避免主觀因素的人為判斷錯誤，更能夠得到公準的保險評估。

### 二、 本次專案簡介

本次資料來源是 Kaggle 上針對醫療保險預估的資料集，其資料為一家醫療保險公司收集其近 1000 位客戶自願提供的健康相關參數之數據，其目的為針對該客戶每年該付多少醫療保險費用，建立一客觀的預測模型。

### 三、 實驗流程、方法及結果

1. 針對資料缺失值、統計量進行檢查及評估
2. 對各欄位資料進行視覺化分析(包含 correlation,

multivariable distribution)

3. 建立各類模型，並選定一最佳分類模型
4. 最佳分類模型之參數最佳化
5. 評估模型分數並匯出模型
6. 建立一 App 方便評估客戶資料

本次選定最佳分類模型為 RandomForest Classification，經過一系列處理後預估分數達 93% 正確率。

#### 四、 討論

本次資料集選定資料量偏少，分布資料也不夠均勻，可能導致模型訓練時掉入區域最佳解，且在切分資料時，若沒均勻切分訓練資料各類別的比例，很容易造成 overfitting，因此未來在針對切分資料集時需更加小心。