



2021 企業疊合數據競賽

團隊名稱：

團隊成員：

下次會更好 (團隊名稱)

(隊長) 簡辰穎

(隊員) 蔡睿芸

題目名稱：

消費者族群購物周期及商品預測

熱門巧克力和餅乾口味變化預測

創造出 AI 新時代股票投資的新方式

快篩試劑需求量預估方式

使用者訂閱行為預測

安全庫存預測

疊合題目 1 「提出一解決方案，運用 AI 與大數據，改善或優化金融市場客戶需求之問題」

疊合題目 2 「提出一數據模型，預測消費市場變化及消費者食品口味趨勢」

疊合題目 3 「提出一新的演算模型或商模機制，協助我國的商圈及實體店家，掌握疫情逐漸復甦後的商業機會」

中華民國 110 年 10 月 03 日

※申請團隊保證申請文件所列資料及附件均屬實※

※若有偽造不實者或侵權行為，申請團隊須負完全之法律責任※

一、參賽團隊基本資料

(團隊成員須與官網上報名隊伍相同，請於繳件時於系統確認，不符者主辦單位有權取消其領獎資格)

參賽團隊介紹

| 隊伍名稱 | 下次會更好 | | | |
|--------|------------|-----|-----------|--|
| 團隊介紹 | 持續進步，下次會更好 | | | |
| 團隊成員介紹 | | | | |
| N | 角色 | 姓名 | 服務單位或就讀學校 | 專長與經歷 |
| 1 | 隊長 | 簡辰穎 | 國立陽明交通大學 | <ul style="list-style-type: none">● 資訊管理與財務金融學系● 參與蝦皮 2020 I'm the best coder 競賽● 參與路博邁 x 交大金融科技數據應用黑客松● 參與 2020 Shopee Code League |
| 2 | 隊員 | 蔡睿芸 | 國立台北教育大學 | <ul style="list-style-type: none">● 資訊科學系● 曾參與蝦皮 2020 I'm the best coder 競賽● 擁有 Microsoft AI-900 證照● 參與 2020 Shopee Code League |

二、構想說明 (頁數建議 3 ~ 8 頁)

(一) 提案摘要及目的

1. 出題企業遇到的痛點：
 - A. 大量資料，無法直接以視覺化的方式得出結論
 - B. 企業永遠無法知道客戶的消費需求為何，因此需要透過大量的數據分析來進行預測，以得到相對合理的結果
 - C. 若無法提前預知應該儲備多少貨物量，等到儲備量完全沒有時才進行貨物訂購，將會大幅提升銷售成本
2. 解題構想的摘要
 - A. 由於我們所要預測的安全庫存量並未在資料中發現，因此希望以非監督式學習的方式建構模型，將具有相似特點的資料進行比對，用以分析生產線所需準備的安全庫存量
 - B. 將資料分組，建立叢集模型
3. 產品或服務對企業或產業環境的幫助
 - A. 若能夠透過所提供的資料，正確預測出安全庫存量，不僅可以幫助企業提前備好所需的貨物，避免客戶需要購買的商品沒貨，還可以幫助他們管理與供應商之間的關係，供應商能夠再得知企業的貨物數量低於安全庫存量時，主動詢問企業是否需要供貨
 - B. 擁有自己的庫存能夠使廠商批量購買商品時，在價格談判方面也可以有優勢
 - C. 可通過比對自己目前的庫存量與所預測的安全庫存量，來大大減少訂購與交貨之間的時間
4. 技術面的突破與創新
 - A. 運用機器學習的方式進行預測，取代過去透過直覺經驗的方式來決定供貨商的行銷方針
 - B. 針對貨物數量的重要性與產值的重要性進行重點分析，從而輔助協助企業達到合理配置庫存，相比過往純粹由人工進行觀察和分析，不只可以減少一些人為分析的錯誤，也能降低企業的人力成本

(二) 數據運用

- 企業數據採用：主辦方所提供的電子庫存採購分類資料集
- 我們選擇了使用 python 語言作資料整理，運用 pandas、numpy 等套件進行清理資料，並以 matplotlib 將資料可視化，便於接下來的分析
- 首先，我們先進行了缺失值檢查，檢查資料中是否存在缺失值
- 接著，由於資料數較少，不利於模型的訓練，因此我們通過 data engineering，利用現有資料衍生出更多的資料以利於後續的模型訓練
- 最後，進行 feature selection，我們將資料中我們認為對於預測結果「最可能有直接關係」的特徵選出，作為模型訓練的主要特徵來進行分組

(三) 提案應用與技術架構

提案應用：

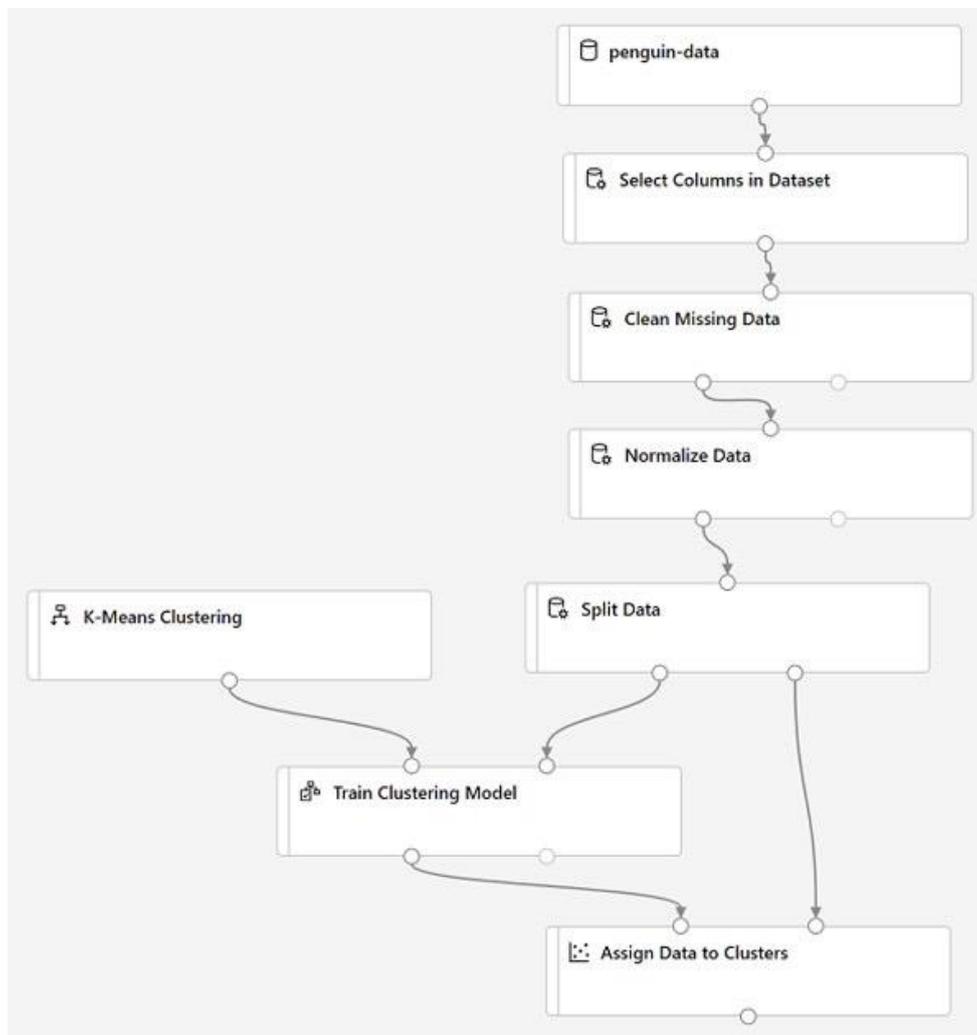
- 我們預期透過給予資料中的訂購數量、數量重要性以及產值重要性三個主要特徵進行安全庫存量進行分析
- 將模型部署至雲端服務上，供企業或廠商使用
- 當客戶訂購的商品數量越多且其數量重要性、產值重要性皆為 A 時，我們可以猜想這項商品可能對於客戶而言需求量大且非常重要，所需的安全庫存量相對來說也會需要較多，以防沒有貨物供給給顧客，造成顧客損失等情形
- 我們整體的服務設計流程是希望能跟 SCM(供應鏈管理系統)一同應用，在透過我們的模型預測出適合的安全庫存量以後，搭配供應鏈管理系統時時監測庫存量與安全庫存量，並即時向供應商進貨

技術架構：

- 以向量化的特徵定義 n 維度座標，以便於接下來 K-Means Clustering 的視覺理解
- 由於選用非監督式學習，我們以 K-Means Clustering 為主要演算法，重複設立每個組別的 centroid，以找到最合適的 K value

- 由於欄位中的資料多為名稱而非類別，因此我們利用遮罩將部分類別的內容修正為數值，以便於資料叢集的處理
- 由於我們目前的資料量過少，即使是經過 data engineering，訓練出的模型也極有可能發生過度擬合，使得模型的預測效果不合適於真實世界中的應用。有鑑於此，我們期望未來能夠找到更多相關 open data，以解決資料量不足的問題，並進一步讓模型預測能達到更好的效果

技術流程(參考)：



(四) 預期成果與效益

- 解決庫存量不足才進行補貨，造成成本增加的問題

- 預期以服務的形式部署到 linebot 展現成果，店家可以直接在 line 上看到由模型預測出的安全庫存量
- 若能夠進一步取得企業目前的庫存量資料，我們甚至也可以直接在 line 上取得各商品的現行庫存數據，並在模型預測出安全庫存量為多少時，當發現目前庫存量小於安全庫存量時，發送訊息提醒企業應該要進行進貨，避免庫存不足的問題
- 我們期望未來能更進一步的與其他的雲端服務並用，讓此模型的功能得到更深的推廣

(五) 其他有利審查項目 (選填，無則免填)

如相關產業實績、競賽得獎證明、隊員學經歷、學術、技術證明等，請重點摘要條列。