

# Instacart 的消費者購物分析

## Instacart Market Basket Analysis

指導教授：莊坤達

專題成員：曾鈺安

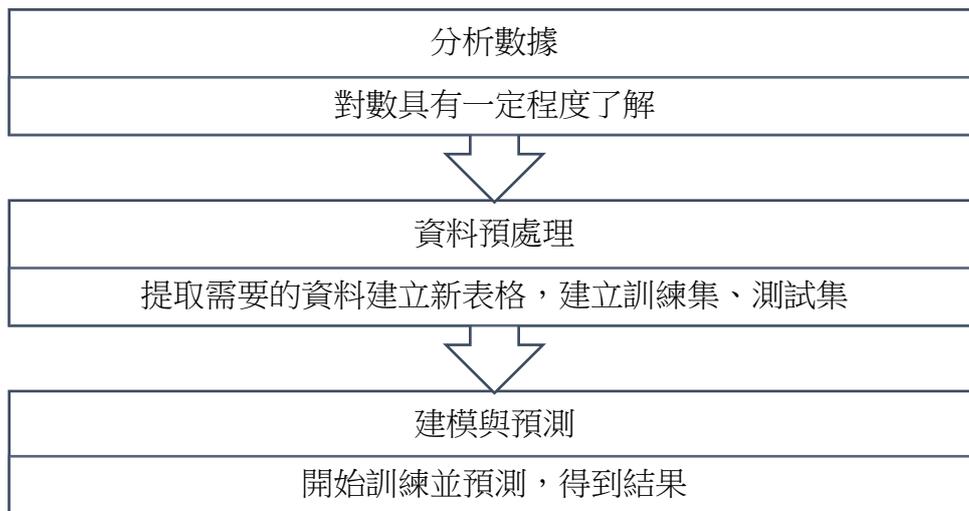
開發工具：python 3.9

測試環境：Linux

### 一、簡介：

以 Kaggle 競賽上的題目為題，目標是根據 InstaCart 此生鮮雜貨代買代送平台所提供的消費者訂單記錄，來預測每一位消費者在曾購買過的商品名單裡，可能會再次購買那些商品，InstaCart 平台即可以依據此數據來對消費者進行個人化的購買推薦。

以下為流程圖：



過程可分為三部分，分別是數據分析、資料預處理、建模與預測。首先根據 InstaCart 所提供的官方數據，我們可以分析消費者的購買習慣，例如：購買頻率、購買時間、熱門商品……等等，讓我們對資料有一定程度上的了解。

在分析完數據之後，就可以開始將我們需要的資料進行整理，合成新的表格，以方便之後做進一步的處理。

接著，可以開始依據訓練集 training 建立模型，並將結果進行 testing 來得到預測結果。因為我們將此題目定義為一個二元分類問題：對每一位消費者而

言，他們購買過的每一項產品是否會再度購買，結果為是/否。所以若 testing 完購買的機率大於一定值，我們則判斷結果為1，即會購買，否則為0，即不會購買。

## 二、測試結果：

因為此為 Kaggle 所提供之競賽題目，所以可以將最後的預測結果存在 submission.csv，並上傳至 Kaggle 取得成績，這次 Public Score 為0.37930。依照 Kaggle 平台給出的分數可以大概了解本次機器學習的好壞。

