

# 社群媒體分析 期末報告

PTT超商版與行動支付版\_分析報告 情緒探勘與主題模型應用

# 第22組

N124320007 藍筱琦 N124320014 陳柏豪 N124320021 永瑞華 N124020013 胡馨予

# 目錄

壹、分析議題簡介	
二、研究目的	
貳、工作流程設計	
一、資料前處理	
二、情緒分析	3
參、詞彙相關性觀察	13
一、轉換為文件詞矩陣(DTM)	
二、相關性矩陣熱圖	
三、關聯式文字雲 (shiny)	
五丶word2vec	
肆、主題模型	
一、LDA主題模型	
二、GuidedLDA 主題模型	21
伍、總結	22
陸、影片連結	25

# 壹、分析議題簡介

## 一、研究動機

隨著台灣零售業的快速發展,便利商店作爲重要的零售通路,涵蓋食品、飲料、咖啡等多元商品,並在消費者日常生活中扮演不可或缺的角色。根據未來流通研究所2024資料顯示,2023年台灣零售業規模達4兆7,298億元,便利商店業銷售商品結構集中度(CR3)高達87.4%。近年來,便利商店的商品結構呈現動態變化,例如食品銷售佔比持續攀升至34.2%,顯示食品特化策略的成功。

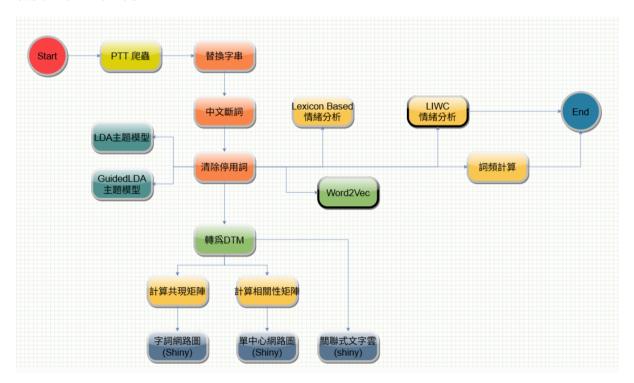
目前對便利商店商品結構及消費者行為的分析仍相對有限,難以全面掌握市場趨勢 及消費者需求,傳統市場調查在時效性和成本上也存在限制。因此,PTT作為台灣重 要的網路討論平台,其CVS版匯集大量消費者真實討論,能提供即時且深入的市場洞 察。本研究希望透過社群文字探勘技術,分析消費者對便利商店商品的情緒及主題, 提供更具洞察力的市場資訊,進一步提升商品設計及行銷策略的精準度。

#### 二、研究目的

- (一)瞭解消費者情緒傾向:透過分析 2024 年 4 月 1 日至 2025 年 3 月 31 日期間的便利 商店(CVS)及行動支付(MobilePay)相關數據,量化消費者對便利商店商品的 正面與負面情緒,並探討影響情緒的因素,例如商品特價活動、促銷策略(如買一 送一)、商品品質(如厚乳拿鐵、霜淇淋等)。
- (二)挖掘消費者關注主題:使用 TarFlow 工作流程系統進行文字探勘,分析消費者對 便利商店商品的主要關注點,例如特價、美食、小資族偏好、聯名商品(如濃翠、 霜淇淋)等,並識別消費者對便利商店促銷活動及熱門商品的需求及偏好。
- (三)提供市場洞察:根據分析結果,提出具體建議,協助便利商店制定更精準的促銷 策略及商品設計,並為便利商店及行動支付平台提供消費者行為洞察,提升消費者 滿意度及品牌忠誠度。

# 貳、工作流程設計

本研究分析流程主要透過 TarFlow 平台進行,包含資料前處理、中文斷詞處理、DTM建立、LDA/GuidedLDA 主題模型建構及關聯文字雲等模組。情緒分析部分採用兩種方法:一為 Lexicon-Based 字典法,結合 NTUSD(台灣大學情緒字典)及自訂詞庫;二為 LIWC 語意分類系統,分析正負情緒、焦慮、憤怒與悲傷等情緒維度。資料來源為 PTT 超商(CVS)板、行動支付(MobilePay)版,分析期間為 2024 年 4 月 1日至 2025 年 3 月 31 日,總計收集 1,040 則與超商商品相關的留言,作為分析基礎,相關流程如下所示:



#### 一、資料前處理

#### 1.資料匯入

讀入「PTT爬蟲」的資料,以下為分析資訊說明:

■ 分析期間:2024.4.1~2025.3.31

■ 資料來源:CVS(超商) + MobilePay(行動支付)

■ 使用工具:TarFlow工作流程系統

■搜尋關鍵字:

超商

特價

美食

小資

愛珍食

買一送一

咖啡

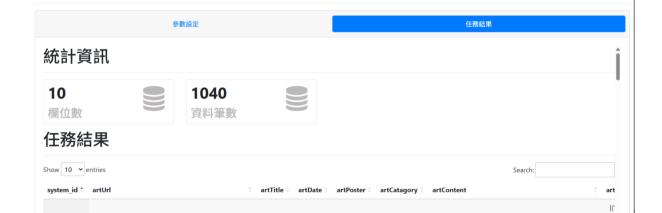
厚乳拿鐵

濃翠

霜淇淋

**註 PTT 爬蟲** <sup>☑</sup> (4)





"(<sub>₩</sub>

[商品]7-

# 2. 資料清理

■ 使用「替換字串」,將符號左邊的文字替換成右邊的文字,如下所示

\n\n>> °

n>>

統一超商>>小七

7-11>>小七

小7>>小七

7仔>>小七

7 eleven>>小七

7eleven>>小七

FamilyMart>>全家

family mart>>全家

Hi-Life>>萊爾富

hilife>>萊爾富

OK便利商店>>OK超商

幸替換字串 ② (7)



×

# 3.設定「中文斷詞」定義詞彙

將斷詞斷不好的字詞,如「午后」、「時光」應爲品牌名稱「午后時光」等,設定字詞如下:

7-11 10000

可爾必思 500

新口味 500

午后時光500

多多綠茶 500

商品名稱 500

廠商名稱500

注意事項 500

五捨六入 500

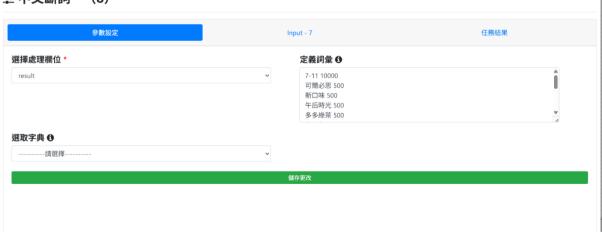
純喫茶 500

白玉抹茶 500

自問自答 500

生巧克力 500

**幸中文**斷詞<sup>②</sup> (8)



×

# 4. 清除停用字

排除不具分析價值的字詞,有助於後續情緒分析與主題模型的準確性。

### 設定如下:

● 語言選擇: 中文 (Chinese)

● **預設處理**: 開啟內建停用詞庫,並將文字轉爲小寫,排除單字元、英文字母、數字、特殊符號、HTML標籤與換行符號。

**自定義停止詞:**為更貼合超商評論語境,我們額外加入以下常見但語意薄弱或冗餘的詞彙作爲排除對象:

商品名稱、價格、便利商店、廠商名稱、評分、心得、規格、內容物、熱量、刪除、跳過、清楚、今天、好像、裡面、完全、知道、看看、東西、分爲、看到、區域型、註明、標示、如有、項目、現在、這款、這種、最近、表示、直接、一點、之後、一下、買來、目前、還蠻、算是、附近、那種、一次、作者、看板、標題、未滿、商品。

# 芸清除停用詞 ♂ (9)

参數設定	Input - 8	任務結果
語言*	使用預設停止詞	
Chinese	是	•
是否清除單字元 ❸	是否轉為小寫英文	
是	是	V
清除英文字母 *	清除數字 *	
是	是	•
清除換行符號 *	清除特殊標點符號 *	
是	- 是	•
清除html tag *	自定義停止詞	
是	商品名稱 價格	ê
	便利商店廠商名稱	

#### 二、情緒分析

本次使用兩種方法進行情緒分析,分別為 Lexicon-Based 字典法 以及 LIWC 分類法, 透過內建與自訂的情緒詞彙,統計使用者留言中的正負面情緒分布與常見情緒詞。

### (一)Lexicon-Based 字典情緒分析

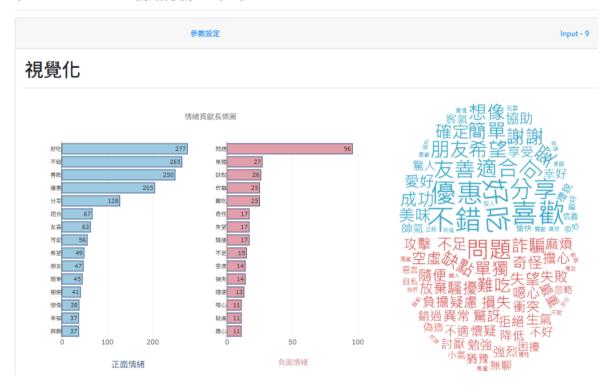
我們載入預設情緒字典,並依研究需求自訂正面詞彙(如「喜歡」、「好吃」、「優惠」) 與負面詞彙(如「問題」、「難吃」、「單獨」、「失望」)進行辨識與計數。 從分析結果可見,出現次數最多的正面情緒詞爲「好吃」、「不錯」、「喜歡」、「優惠」、「分享」等,顯示消費者對商品口味與促銷活動普遍持正向評價。負面情緒詞中則以「問題」、「單獨」、「缺點」、「難吃」等爲主,反映部分使用者對產品品質或體驗仍有不滿。視覺化文字雲顯示,正面情緒詞聚集在「適合」、「朋友」、「開心」、「簡單」等語彙附近,負面詞則包含「失望」、「損失」、「欺騙」、「抱怨」等,呈現情緒分布的兩極化特性。

× ^

# E Lexicon Based 情緒分析 (20)

參數設定	Input - 9	任務結果
语言 *	載入預設情緒字典 *	
Chinese	是	•
定義正面詞彙 ❸	定義負面詞彙 ❶	
以換行符號區隔,e.g.	以換行符號區隔,e.g.	
詞彙	詞彙	
開心	生氣	
高興	悲傷	
		,
	儲存更改	

# 幸 Lexicon Based 情緒分析 <sup>☑</sup> (20)





# (二) LIWC 分類情緒分析

使用 LIWC 模組,對情緒進行更細緻分類。從圓餅圖統計可見,整體情緒用字比例中:

- 正面情緒(positive\_count)佔比 38.4%
- 負面情緒 (negative\_count) 爲 7.78%
- 其餘為中性或情緒相關詞(如 affect\_count \ anger \ sadness \ anxiety)

與 Lexicon-Based 結果一致,留言整體偏向正面,代表消費者對超商聯名商品多抱持好感,但仍須留意部分留言中提及的負面情緒,例如「價格」、「難吃」、「問題」、「失望」等,可能與產品體驗或期待落差有關。



# 幸 LIWC 情緒分析 <sup>☑</sup> (17)



# (三)、詞頻分析與文字雲視覺化

完成斷詞與淸除停用詞後,我們進行詞頻統計,觀察在所有留言中最常出現的詞彙。這個步驟能讓我們快速掌握討論焦點與熱門話題,也為後續的主題分類與情緒判斷提供依據。

本次設定保留前 200 個高頻詞,從文字雲結果可以看出,最常被提到的詞包括:

- 品牌與店名相關:「全家」、「小七」、「超商」
- 產品與體驗:「霜淇淋」、「口味」、「好吃」、「咖啡」、「蛋糕」
- 行銷活動與支付:「優惠」、「活動」、「推薦」、「付款」、「感覺」

這些熱門詞彙反映出消費者對超商推出的聯名商品,特別是在**口味體驗與優惠活動**方面的討論熱度較高,也與後續的情緒與主題模型結果相互呼應。



# 幸 詞頻計算 ᠍(14)



### (四)小結論

透過本研究對 PTT 超商板的文本分析,我們發現消費者對超商聯名商品的關注主要集中在甜點類產品、口味體驗與促銷活動。從詞頻統計與文字雲結果來看,「霜淇淋」、「好吃」、「優惠」、「推薦」等詞彙頻繁出現,顯示出產品本身的感受與活動誘因是留言的核心。

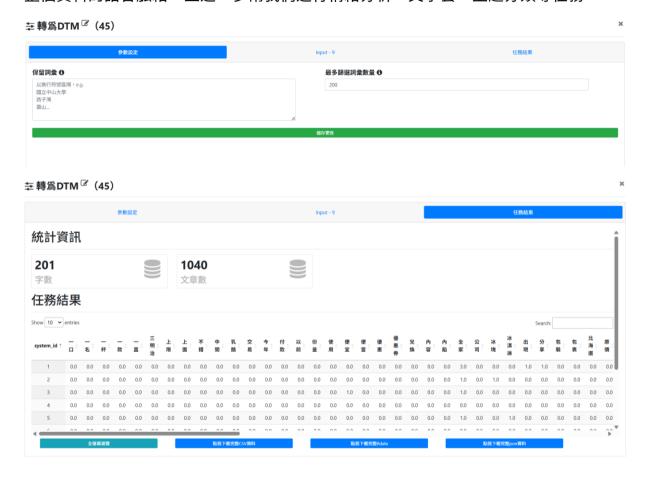
在情緒分析部分,整體留言以正面情緒為主,使用者普遍對產品感到滿意,尤其是口味與設計,但仍有部分留言提及「難吃」、「失望」、「問題」等詞,反映對商品品質與定價的個別負面反饋。

主題模型分析方面,LDA讓我們掌握留言的大致分類輪廓,而 GuidedLDA 則進一步聚焦在研究團隊指定的「咖啡」、「霜淇淋」、「店員」、「麵包」等品項與服務互動,呈現出更有意義的主題分群

# 參、詞彙相關性觀察

# 一、轉換爲文件詞矩陣(DTM)

我們將「文字」變成「可以分析的數據」,因此設定最多詞彙數量為 200 個高頻詞(根據 出現次數排序),讓系統知道每篇留言到底出現了哪些詞、出現幾次,讓系統可以理解 整個資料的語言脈絡,並進一步幫我們進行情緒分析、文字雲、主題分類等任務



#### 二、相關性矩陣熱圖

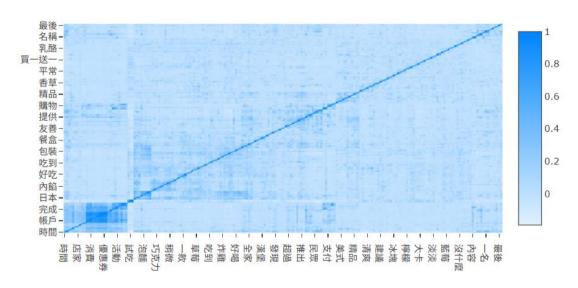
這張相關性矩陣熱圖視覺化了消費者評論中常見詞彙之間的關聯程度,幫助我們理解 哪些詞彙會在同一段文字中共同出現,揭示消費者心智中的語意關聯。

横軸與縱軸列出了本次分析中出現頻率較高的詞彙,例如「時間」、「超商」、「門市」、「優惠」、「問題」等。顏色的深淺則代表相關性的強弱:

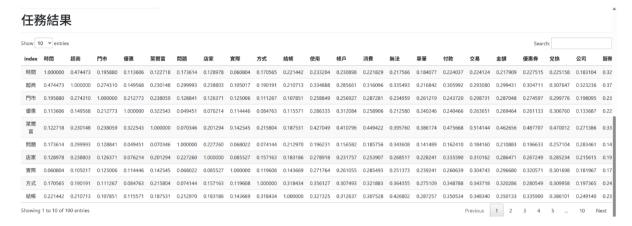
● 深藍色代表關聯性高(相關係數接近1),

- 淺藍色代表關聯性低(相關係數接近0)。對角線上的深色為詞彙與自身的相關性(永遠為1),這是數據矩陣的自然特性。從圖中可以觀察到幾個顯著的詞彙關聯:
- ●「超商」與「付款」、「兌換」、「優惠券」等詞彙有明顯關聯,顯示消費者常在評論中將 超商與促銷行為連結起來。
- ●「萊爾富」與「帳戶」、「消費」的相關性偏高,反映出其數位化支付或會員制度被頻繁 提及。
- ●「問題」與「公司」、「客服」也呈現中等以上的相關性,代表負面經驗常與品牌服務連結。
- ●「方式」與「結帳」、「交易」間的高關聯也揭示了流程與支付的討論頻繁程度。 這些關聯揭示了消費者在描述購物體驗時,會自然地將特定詞彙組合使用,反映出他 們在特定情境下的心理連結與經驗重點,對於未來的行銷策略擬定或服務流程設計, 皆具有參考價值。

熱圖



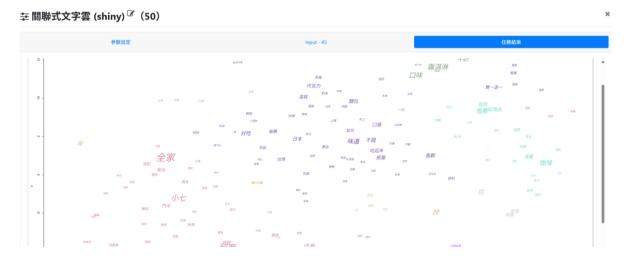
詞彙相關性熱圖



相關性DTM矩陣

# 三、關聯式文字雲 (shiny)

這張關聯式文字雲圖透過詞彙的分群與空間分布,展示了消費者在討論超商聯名商品時所關注的焦點與情緒傾向。圖中詞彙依照語意相似度與共現頻率被自動分成五個群組,每個群組以不同顏色標示,代表不同的關注面向或主題。



從圖中可觀察到以下幾個明顯的主題群組:

- 品牌與通路(粉紅色區塊):如「全家」、「小七」、「超商」、「門市」、「支付」、「消費者」等,代表消費者會明確提及品牌名稱與購買通路,顯示品牌力與通路印象對消費者影響深遠。
- 2. 口味與食感體驗(紫色區塊):包括「好吃」、「味道」、「口感」、「不錯」、「起司」、「蛋糕」、「飯糰」等,顯示消費者會細緻描述商品的味覺與感受,對產品本身品質具高度關注。

- 3. 促銷與優惠(藍綠色區塊):如「買一送一」、「推薦試用品」、「優惠」、「拿鐵」、「咖啡」、「特價」等,顯示價格促銷活動對吸引消費者具有強烈吸引力。
- 4. 新品與包裝設計(右上綠色區塊):包含「新品」、「推出」、「霜淇淋」、「口味」、「包裝」、「試吃」等,代表消費者對新奇感與視覺呈現亦相當在意。
- 5. 負面與問題類(右下與左下分布):雖不明顯聚集成群,但如「問題」、「客服」、「退 貨」等詞彙也隱約分布,提醒企業需注意使用者體驗中的痛點。

整體而言,此文字雲不僅揭示了熱門討論詞彙,也讓我們了解消費者關注的多元面向,從品牌認知、促銷誘因到口味偏好與產品創新,對後續產品開發與行銷策略皆有重要參考價值。

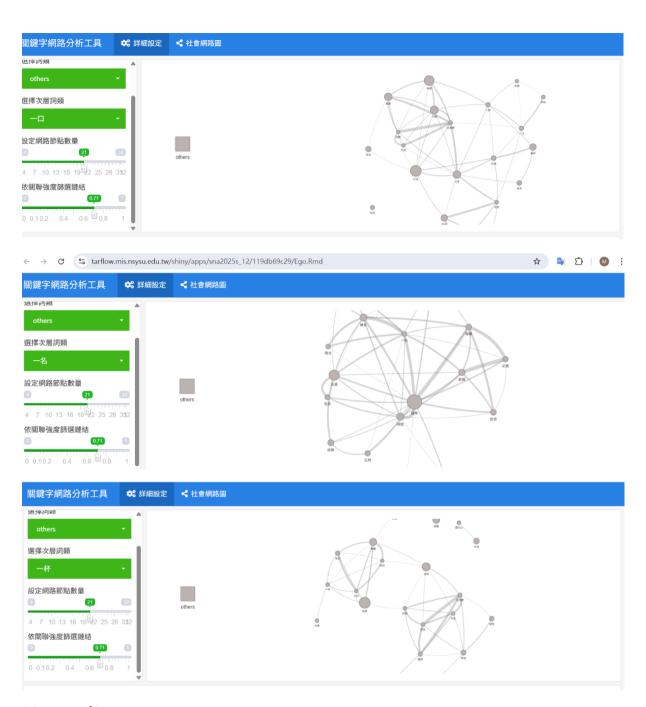
参數設定	Input - 45	任務結果
群數 <b>*</b>	文字雲顯示字數 *	
0	1000	
代次數(最少250次)	距離計算公式 *	
000	euclidean	v
<b>合演算法</b> ◆		
ingle	~	

#### 四、詞彙相關性矩陣分析

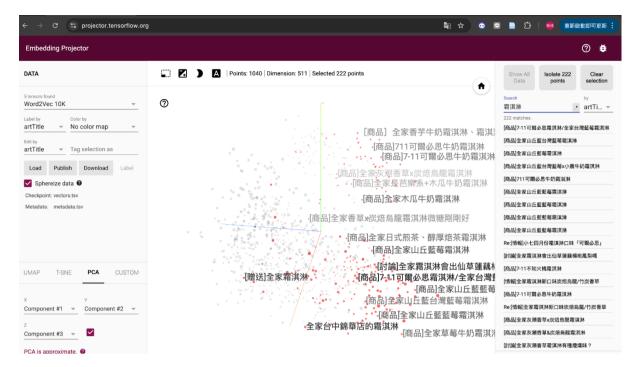
在了解了詞彙之間的共現次數後,我們進一步計算詞與詞之間的「相關性」,也就是除了同時出現的頻率之外,它們在整體留言中是否存在統計上的連動關係。

這次分析同樣針對出現頻率最高的 200 個詞進行計算。從結果可以看到,像是「使用」 和「帳戶」、「付款」和「優惠券」這些詞的相關性數值偏高,代表使用者在留言中常常會 一起提到這些詞,語意上也可能彼此相關。

我們也將這些相關性結果透過熱圖呈現,顏色越深表示詞彙之間的關聯越強。從圖中可以明顯看出,與「交易」、「付款」、「帳戶」、「活動」相關的詞彙多半集中在一起,顯示這些主題在 PTT 討論中通常是連結出現的焦點。

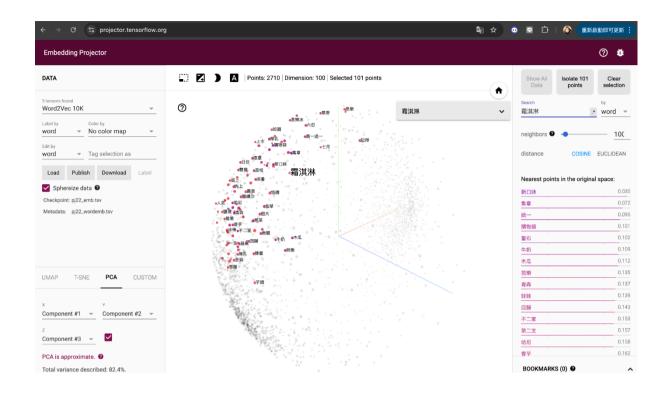


五、word2vec



使用 Word2Vec 模型觀察「霜淇淋」這個詞在語意空間中與哪些詞語「語意相近」,亦即經常在類似語境中出現,藉此理解使用者提及「霜淇淋」時的關聯概念與情境語義。

- 1.「霜淇淋」的語意鄰近詞(右側淸單,依Cosine距離排序):
- 新口味、集章、統一、購物袋、聖石、牛奶、木瓜、芒樂、青茶、……
- 這些詞反映出:
- 「霜淇淋」常與新品、季節限定、或活動促銷(如集章)共同出現
- 搭配食材詞彙如「牛奶」「木瓜」「芒樂」「靑茶」等,顯示與口味或原料強關聯
- 品牌名稱如「統一」、IP名稱如「娃娃」,指出其可爲聯名或品牌策略核心之一
- 2. 嵌入分佈位置(圖中左上):
- 圖中顯示「霜淇淋」與其他食品、促銷、商品名稱聚集,構成一塊語意密集區
- 同群詞彙如「特價」「第二支」「口感」「小資」「聯名」等,說明「霜淇淋」與生活感、日常 體驗、實惠訴求有高度連結



# 3. 嵌入分佈位置(圖中左下):

圖中顯示「霜淇淋」的各種口味,以及與其他品牌聯名

# 小結

透過 Word2Vec 分析,「霜淇淋」不僅是一項單一商品,而是一個涵蓋「新奇口味」、「品牌活動」、「季節限定」、「實惠體驗」等語意的多層次語境詞。它在消費者心中所聯想的語彙與情境反映了便利商店如何透過此類商品打造品牌記憶點與社群熱度。

# 肆、主題模型

#### 一、LDA主題模型

為探索使用者在 PTT 超商(CVS)板、行動支付(MobilePay)版中針對超商商品的潛在討論主題,先建構 LDA主題模型,進行文本主題分類與語意結構辨識來觀察主題分類的效果

# 我們設定的方式:

● 我們讓系統分成6個主題

- 排除太常出現或太少見的詞(太雜或太冷門的都不要)
- 每個主題抓出 20個代表詞,來幫助我們看主題是什麼
- (一) Intertopic Distance Map(主題距離圖)顯示各主題分佈與彼此關係,其中主題 2、3、6 聚集較近,推測語意有交集,例如皆與小七、超商或食物類型有關。
- (二) LDA 模型的 Top-30 重要詞彙中,出現頻率較高的詞包括「活動」、「推廣」、「小七」、「牛奶」、「門市」、「內容物」、「便當」、「推薦」、「取餐」、「口味」、「甜點」、「巧克力」等,顯示 PTT 討論多集中在產品內容、口感體驗、品牌形象與促銷活動等面向。



Interropic Dictance May (via multidimensional scaling)

A = 1

Top-30 Most Selient Terms 1

1 2 2 2 4 4 5 1 5 1 5 1 1 2

1 2 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

1 2 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

1 2 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

1 2 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

1 2 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

1 2 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

1 2 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

1 2 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

1 2 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

1 2 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

1 2 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

1 2 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

1 2 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

1 2 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

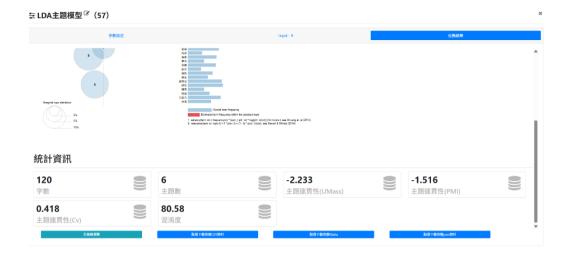
1 2 2 4 6 5 1 5 1 1 3

Top-30 Most Selient Terms 1

1 2 2 4 6 5 1 5 1 1 1 3

Top-30 Most Selient Terms 1

Top-30 Most Selient



# 二、GuidedLDA 主題模型

為了深入探討消費者在評論中關注的焦點,我們使用 GuidedLDA 進行主題模型分析,並透過種子字詞引導主題分類,使模型更貼近實務觀察。

# 以下爲本次模型設定重點:

● 主題數:4個主題(實際視覺化結果顯示5個主題)

主題種子字:霜淇淋、店員、咖啡、麵包

詞彙頻率下限:40;上限比例:0.5(過濾過低或過高頻詞)

每主題保留關鍵詞數量:20

● 模型迭代次數:50

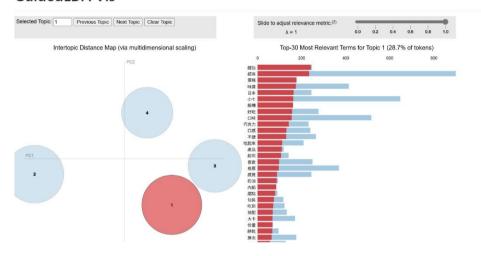
主題模型視覺化分析與主題命名

每張圖左側為主題距離圖(Intertopic Distance Map),用於表示不同主題的相似程度;右側則為該主題的Top 30 關鍵詞,紅藍條代表關鍵詞在主題內的重要性與特異性。

# 主題 1:「便當主食群」

- 關鍵詞如:便當、飯糰、小七、日本、咖哩、味道、好吃
- 說明:此主題以主食類商品為核心,消費者對超商提供的主食產品(如飯糰、便當)及其風味、口感表現出高度關注,並常搭配品牌名稱共同提及。
- 結論:商品的風味品質是影響滿意度的主要因素。

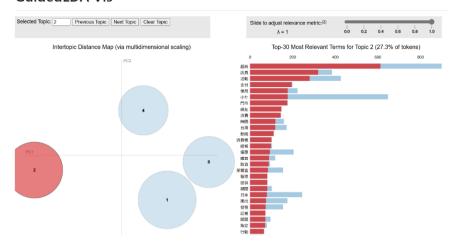
#### GuidedLDA Vis



# 主題 2:「服務與支付體驗」

- 關鍵詞如:超商、店員、活動、支付、消費者、時間、結帳、服務
- 説明:此主題聚焦於門市互動與購買流程體驗,包括人員態度、支付便捷性與 促銷活動。
- 結論:人員服務品質與流程效率是影響購物體驗的重要環節。

#### **GuidedLDA Vis**



# 主題3:「甜點與口感體驗」

- 關鍵詞如:霜淇淋、口味、試吃、濃郁、巧克力、牛奶、焦糖
- 說明:本主題描述消費者對甜點產品的品評與感受,特別著重於口味細節與創 新口感的喜好。
- 結論:甜點創新與試吃體驗是促使消費者回購的關鍵。

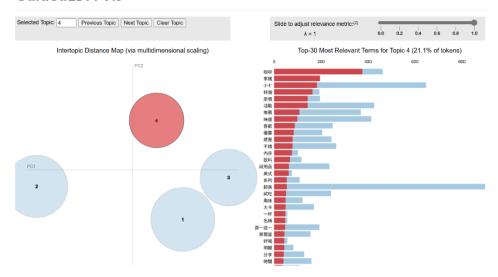
#### **GuidedLDA Vis**



# 主題 4:「咖啡與優惠組合」

- 關鍵詞如:咖啡、拿鐵、小七、活動、原價、優惠、特價、試用品、買一送一
- 說明:此主題聚焦於咖啡相關商品,搭配促銷優惠方案與品牌活動,是消費者 在日常消費中常見的關注點。
- 結論:咖啡搭配促銷活動可有效提升消費意願。

#### **GuidedLDA Vis**



# 小結

透過 GuidedLDA 模型,我們成功將大量評論分為四大主題群組,涵蓋食品品質、服務體驗、甜點口味與優惠活動等面向,並結合種子詞設定進行語意引導。此分析不僅有助於理解消費者關注重點,也能為後續商品設計與行銷活動提供具體依據。



# 伍、總結

本研究以PTT CVS版與行動支付相關討論為資料來源,運用文字探勘與情緒分析技術,針對2024年4月1日至2025年3月31日間的便利商店消費者評論進行系統性分析,並對應三項研究目的,歸納以下結論:

# 一、消費者情緒傾向具明顯變異,促銷與特定商品爲主要情緒驅動因子

本研究透過Lexicon-based情緒分析與LIWC詞彙分群,發現消費者對便利商店的整體情緒以正面居多,尤以「買一送一」、「厚乳拿鐵」、「霜淇淋」等具代表性的促銷商品為主,其正向情緒占比較高,顯示優惠活動成功引起消費者情緒共鳴。然而,在促銷排隊、口味爭議或供應不足等情境中,也出現負向情緒波動,顯示情緒並非一面倒的樂觀,而受限於商品體驗與行銷執行品質。因此,促銷雖具高度影響力,但須搭配完善的配套與供應鏈安排,方能維持情緒正向導向。

## 二、消費者關注主題集中於價格、創新商品與生活情境連結

透過Word2Vec語意關聯與GuidedLDA主題模型分析,顯示消費者關注以「特價」、「美食」、「小資」、「聯名」為核心,呈現出一種「高性價比+體驗感」的消費思維。例如,濃翠茶類商品與厚乳拿鐵常與「療癒」、「下午茶」等語意共現,反映便利商店商品已不再只是功能性滿足,而是逐漸嵌入日常生活儀式感與情緒價值。這也顯示品牌若能進一步在包裝、命名與主題活動上強化情境連結,將有助於深化顧客參與與品牌印象。

# 三、整合分析結果提供具體市場洞察與策略建議

綜合上述情緒與主題分析,本研究指出便利商店應持續發展具情緒感染力的產品策略,例如期間限定、聯名商品等能激發消費者社群討論與期待感的行銷作法;同時也建議可透過社群輿情監測,卽時掌握消費者反饋與情緒轉折,以利動態調整促銷活動與商品供應。此外,行動支付平台也可依據使用者情緒與偏好進行個人化行銷推播,提升使用體驗與忠誠度。

# 陸、影片連結

https://youtu.be/UyiNU\_NEHk0?si=-TLbmIjpSedk5Jqy