

**國立中山大學資訊管理系碩士班**

**社群媒體分析**

**第一次讀書會報告**

**第十組成員：**

**N094020030 陳詠琳**

**N114020004 蔡志強**

**N114020006 黃銘輝**

**N114020007 陳嘉忻**

**N1140200012 黃延平**

**11221828030 范瑞洋**

**11221828031 王上豪**

**M121020012 凃宥安**

**指導教授：黃三益 教授**

**助教：蔡易航、蔡睿澤、張宸瑜、呂育真**

**中華民國 113 年 03 月**

|  |
| --- |
| 生 育 觀 測 |
| 動機與結論：  生兒育女本就屬於較極具情緒化的過程，涉及到未來父母的憂慮、期待、喜悅和壓力等複雜情緒。因此，進行情緒分析的研究對於了解並幫助未來父母應對這些情緒至關重要。透過情緒分析，可以深入探究在懷孕、分娩和育兒過程中的情緒變化，揭示潛在的情緒模式和影響因素。這有助於提供更有效的情緒支持和輔導，促進父母的心理健康和家庭和睦。在研究的結果最後可以發現PTT媽寶版觀察到民眾對於生小孩的看法是正面大於負面。 |
| 資料範圍、來源、關鍵字：  資料來源：PTT 媽寶版；工作流程名稱：project\_1\_確認版  資料範圍：2021/01/01~2023/12/31  搜尋關鍵字：生小孩  PTT爬蟲任務結果：總共有334篇與生小孩相關資料 |
| 系統流程： |
| 分析過程：  替換字串：  將文章中出現的網址進行替換    任務結果：取代了58965筆    中文斷字設定：  依據斷詞結果，反覆修正、清洗斷詞。  會用在生小孩的特殊用語做為斷詞，讓分析可以更為準確。  參數設定：    任務結果    詞性篩選：  參數設定選擇保留Noun、Proper Noun、Verb、Adjective。  以下所得到結果：        停用詞：  由任務結果及詞頻計算結果去觀察文字，同樣反覆清洗及加入停用字。  參數設定：利用自定義停止詞進行去除雜訊文字。    任務結果  觀察內容是否較符合所要資訊後，再進行下一步驟。    詞頻計算  篩選詞彙數量設定為200    任務結果      Lexicon Based情緒分析：    任務結果  一篇文章最大正向情緒81個,最大負向情緒有72個    合併資料：  因要進行正、負面情緒分析，所以需要利用system\_id欄位做為key值，利用已分析出來的Lexicon Based情緒分析與PTT原始文章來源進行join。    正向情緒篩選  參數設定條件式設置sentiment\_value > 0 ，挑出正向情緒字詞。    任務結果  移除113筆，保留221筆。正向情緒字詞。    負向情緒篩選  參數設定條件式設置sentiment\_value < 0 ，挑出負向情緒字詞。    任務結果  移除252筆，保留82筆。負向情緒字詞。    每日情緒總合  合併匯總group by 日期artDate，匯總函數用sum，  原因是我們要看到每天的情緒值sentiment\_value的加總    任務結果  這裡可以看到每個日期下，每天所有文章的情緒加總分數。    儀表板-折線圖  利用折線圖可以看到新手爸媽或是準備生小孩的民眾，針對生小孩的疑問和討論議題來對照出他們心情的喜怒哀樂(每日正負情緒彙總走向)。    儀表板-文字雲  利用文字雲來看詞頻計算的結果  生小孩的議題中可以清楚看到生育過程的可怕、經歷、長大等，為群眾最重視迫切第一時間會去收集了解的資料，所以也就會出現在文字雲當中。  儀表板-正向直條圖  利用直條圖來看正面情緒分析，在2021年7月以及2022年2月兩個時期的正面情緒篇數最多。  特別看到的是2021年7月1日，當天正負面兩者趨勢、討論都來到特別的多。        儀表板-負向直條圖  利用直條圖來看負面情緒分析，在2021年7月1日當天的負面情緒篇數最多。  如上述所言，當天出現出一篇文章，引發民眾熱意討論。  因此才會特別出現2021年7月1日正負篇數兩者評論均呈現高峰狀況。    結論：  利用PTT媽寶版資料，分析新手爸媽以及待產中的民眾對於「生小孩」的看法是正面大於負面。    (正面篇數分析圖) (負面篇數分析圖) |