

近代婚姻和諧討論議題觀測

動機與結論：

近年來，世界各地的離婚率持續上升成為全球關注的社會問題。現代社會中，婚姻面臨著前所未有的挑戰，了解引起婚姻問題的主要因素並探索解決途徑，對於提升家庭和諧、促進社會穩定具有重要意義。也希望可以為班上有婚姻關係的同學提供一些具體關於和諧的指引。

我們小組研讀了Samantha Joel等人在《美國國家科學院院刊》上發表的論文《Machine learning uncovers the most robust self-report predictors of relationship quality across 43 longitudinal couples studies》(機器學習在 43 項縱貫夫妻研究中發現了最可靠的關係品質自我報告預測因素)，探討了影響戀愛關係品質的關鍵因素。這項大規模研究使用機器學習技術(如隨機森林算法)，分析了來自 29 個實驗室的 43 個雙向縱向數據集，目的是量化關係品質的可預測性並識別穩定預測關係品質的因素。

這篇論文薈萃分析了四十三項研究，總共調查了96對戀人和夫婦，有將近一百位研究者共同署名：

想要回答一個人們總在問自己的問題：決定婚姻關係好壞的最重要因素到底是什麼？所有這些因素被分成了兩類，一類是描寫個人的因素，一類是在這個關係中的行為模式。比如性格特點、年齡、收入、受教育程度等就是個人因素，說的是這個人是個什麼人；而像是否支持對方、是否信任對方、有沒有愛，這些則是關係中的行為模式。

因此，我們希望藉由這堂社群媒體分析的課程，透過PTT(批踢踢實業坊)這個台灣最具影響力的社群討論區之一，探勘其中的婚姻版。此版是用戶分享婚姻經驗、討論婚姻問題的熱門板塊。通過對PTT婚姻版上用戶發帖內容的探勘和分析，可以更好地了解台灣當前社會中婚姻面臨的主要挑戰和問題。

同時透過探勘技術，系統性地分析PTT婚姻版的發帖內容，識別出影響婚姻和諧的關鍵因素。我們將收集並分析2023/06/01到2023/12/31近半年的發帖量，運用自然語言處理和情感分析等技術，探討用戶在婚姻版上關注的主要議題和表達的情感趨勢。通過這些分析，我們希望能夠揭示影響婚姻和諧的主要因素，找到婚姻幸福的關鍵！！！！！！

資料範圍、來源、關鍵字：

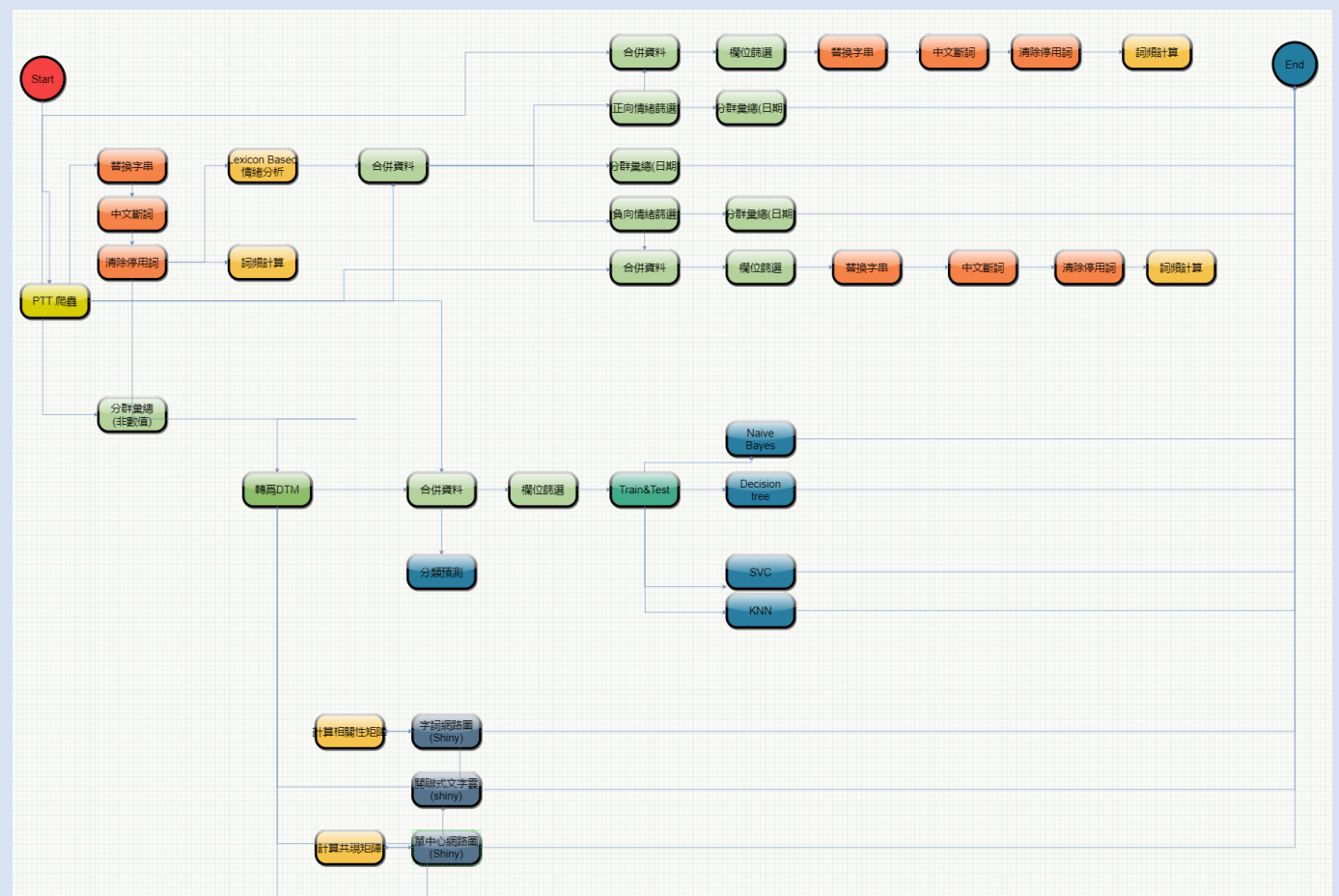
資料來源:PTT 婚姻版

資料範圍：2023/06/01~2023/12/31

搜尋關鍵字：

PTT爬蟲任務結果：總共有5493篇與婚姻相關資料

系統流程:



1. Ptt爬蟲設定：選定婚姻版與結婚版並設定搜尋起始截止日期。

PTT 爬蟲 (4)

參數設定

任務結果

選擇看板 *

yearny(白學)

Lifeismoney(省錢)

Loan(貸款)

MacShop(蘋果買賣)

MakeUp(美妝)

marriage(婚姻)

medstudent(醫學生)

skitstudent(圖文)

搜尋起始日期

2023/06/01

搜尋結束日期

2023/12/31

儲存更改

搜尋關鍵字 ①

以換行區隔, e.g.
國立中山大學
西子灣
...

排除關鍵字 ①

以換行區隔, e.g.
壽山動物園
猴子
...

儲存更改

2. 替換字串：將ptt特有的一些程式碼替換成標點符號，以方便後續的處理。

替換字串 (6)

參數設定

Input - 4

任務結果

選擇處理欄位 *

artContent

選擇替換規則檔案 ①

-----請選擇-----

儲存更改

替換字串設定 ①

```
\n>>,  
\n\n>>,  
Sent from JPTT on my .">>  
https://[a-zA-Z0-9-\.]+\.(?!\d+)?(?!\[^\s\])?>>
```

儲存更改

3. 清除停用詞：將文章內特定鄉民用語、轉載新聞以及網頁程式碼等，後續處理不會使用的字詞刪除。

清除停用詞 (10)

參數設定

Input - 7

任務結果

語言 *

Chinese

是否清除單字元 ①

是

清除英文字母 *

否

清除換行符號 *

是

清除html tag *

是

使用預設停止詞

是

是否轉為小寫英文

是

清除數字 *

是

清除特殊標點符號 *

是

自定義停止詞

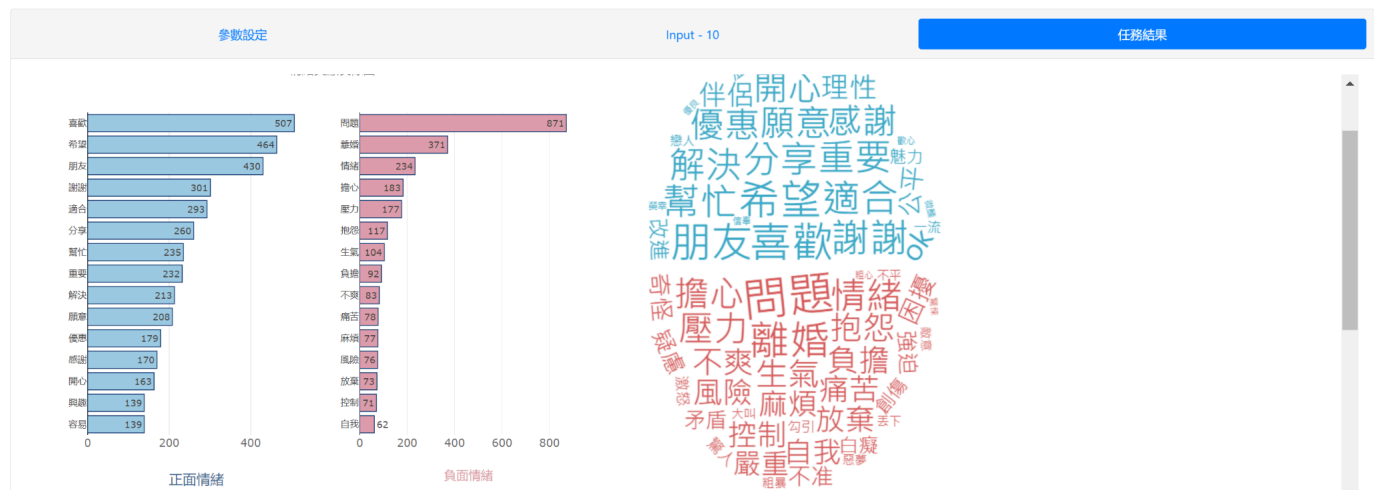
記者
郵寄
我媽
最後
這種

儲存更改

分析過程：

1. 使用文字雲：使用Lexicon Based內建的情緒字典分析，觀察現代人對婚姻正面及負面的態度有哪些議題在板上討論，並將參數設定笑死及XD視為嘲笑的意思，故將其定義在負面詞彙。

Lexicon Based 情緒分析 (24)



> 版上文章可以整體來說對婚姻的感受是抱持著正向的情緒內容比較多，

正面字詞分析：

正面字詞包括：理解、分享、希望、適合、朋友、喜歡、謝謝...等等

婚姻內的神隊友表現，認為喜歡像知心朋友般的相處、彼此會分享每天的感受、互相幫忙以及分工是重要的、同時時常表達對彼此的感謝，都是正面情緒的來源

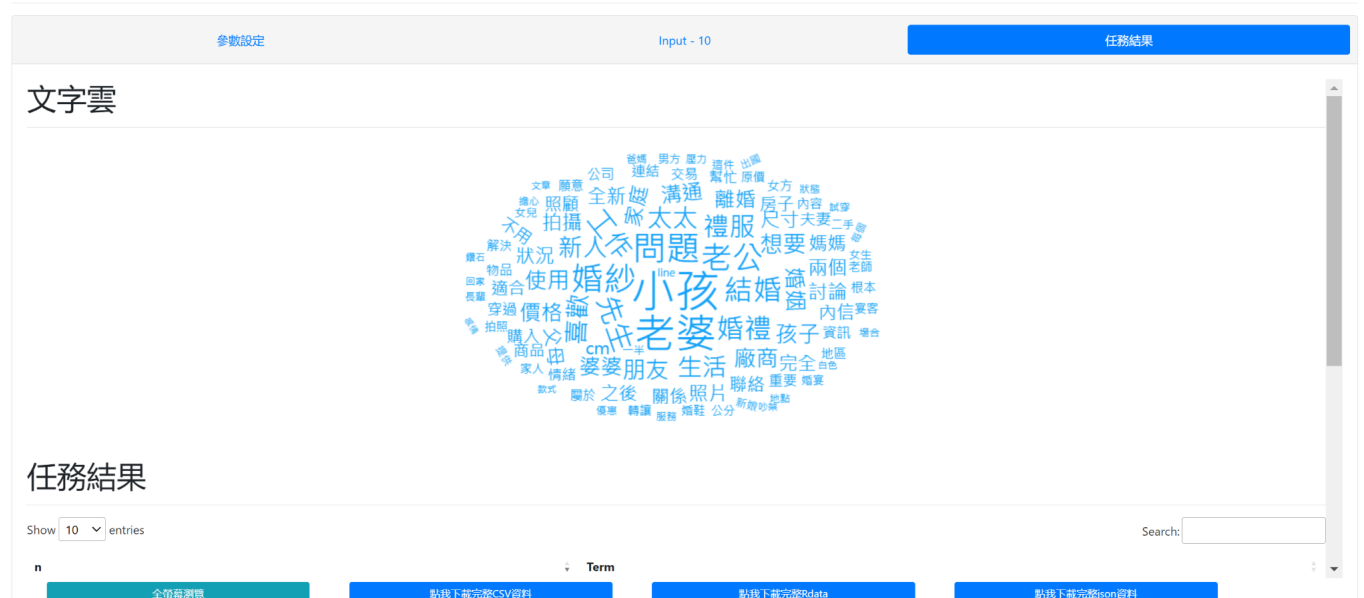
負面字詞分析：

負面字詞包括：問題、困難、壓力、離婚、抱怨、憤怒、不快樂、憤怒...等等

婚姻內的豬隊友表現，大致上發文都是婚姻內的問題、想要離婚、多數的在文章中婚姻內的負面狀況都是豬隊友的抱怨、壓力過大、擔心等字詞

綜合柱狀圖與字詞分析發現，關於婚姻的討論中正面情緒多於負面情緒。雖然存在一些負面的情緒和詞彙，但整體上正面的情緒佔據了更大比重。而不論正面還是反面的文章討論，都是在彼此之間相處時的情緒穩定或者不穩定內這些議題打轉。

2. 詞頻計算：進一步探勘這些發文內討論的內容最多圍繞在哪些議題上



婚姻相關詞彙：結婚、婚姻、婚紗、禮服、離婚、結婚禮物

這些詞彙表明討論主要集中在婚姻相關的各個方面，內容可能是包括結婚禮儀、婚紗禮服以及離婚問題。

家庭成員：老婆、老公、小孩、孩子

這些詞彙顯示討論中涉及了許多關於家庭成員的內容，特別是關於配偶和孩子的議題。

生活方面：問題、生活、朋友、公司、工作

這些詞彙表明討論涉及了日常生活中的各個方面，包括生活中的問題、朋友關係以及工作和公司相關的話題。

情感表達：喜歡、感情、關係、愛、抱怨

這些詞彙顯示討論中有不少關於感情和關係的內容，既有積極的情感表達如“喜歡”和“愛”，也有負面的情緒如“抱怨”。

其他常見詞彙：想要、需要、使用、情況、討論

這些詞彙表示了人們在討論中的一些普遍主題和表達方式。

其中最大的詞頻內容是小孩、老婆以及老公、工作問題等，且因為老婆的詞頻最大，可推測偏男性發文居多

3. 轉為DTM進行議題分類，使用決策樹來判斷文章內探討的議題。

統計資訊

201

字數



2186

文章數



任務結果

Show 10 entries

Search:

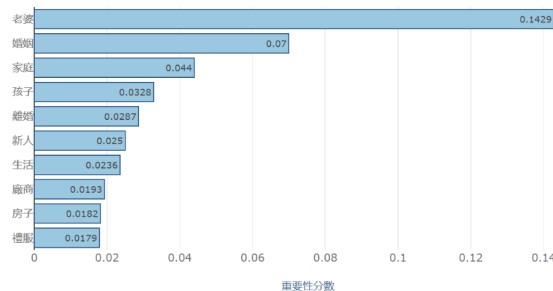
system_id ^	cm	fb	line	xd	一下	一半	年	一次	一直	一種	一點	上班	不想	不用	不錯	之後	也許	二手	交易	人生	介紹	以下	以後	伴娘	使用	價格	優惠	先生	兒子	內信	內容
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	2.0	1.0	1.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	2.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0
5	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

經過合併資料>欄位切割>訓練和測試後以決策樹來判斷哪些上述議題是文章內探討的核心議題。

Decision tree (58)

視覺化

參數重要性長條圖



1. 老婆(0.1429)這是最重要的議題，顯示討論中「老婆」是核心關注。
2. 結婚(0.070)結婚也是討論中的重要議題，顯示了對婚姻狀態和婚禮的重視。
3. 家庭(0.044)家庭是另一個主要的議題，顯示了對家庭整體狀況的關注。
4. 孩子(0.0328)孩子也是討論中的重要部分，顯示育兒和家庭關係的重要性。
5. 關係(0.0297)關係議題的得分較高，顯示對人際關係的關注，包括夫妻關係、親子關係等。
6. 生活(0.025)生活問題也是一個關注點，顯示了對日常生活品質的討論。
7. 老公(0.0236)雖然"老公"在重要性上低於"老婆"，但也是一個顯著的議題。
8. 感情(0.0193)感情議題顯示了對情緒交流和心理狀態的重視。
9. 壓力(0.0183)壓力顯示了對生活和工作的挑戰和壓力的討論。
10. 工作(0.0179)工作問題也被討論，顯示了職業和工作狀態的影響。

討論的核心議題集中在「老婆」、「結婚」、「家庭」和「孩子」上，顯示這些主題在討論中佔據重要位置。同時，關係、生活、感情和壓力等也顯示了對日常生活和心理狀態的關注。內容是小孩和工作的問題最主要在闡述的文章主角多數還是圍繞與老婆相關的話題

接著才是探討婚姻、家庭與小孩的內容，推斷為這些內容多為男性上來版上尋求解答在這些事項老婆的反應是主要核心。

4. 使用關聯式文字雲分類：根據提供的關聯式文字雲圖，對文章內容進行更具體的分析 and 分類

家庭與婚姻(左下方區域):
 主要詞彙:老婆、老公、小孩、小孩、家庭、朋友、結婚、離婚
 討論主題:夫妻關係、育兒問題、家庭生活、朋友關係。討論主要集中在家庭成員之間的互動和關係。

日常生活與工作(右下方區域):
 主要詞彙:工作、生活、先生、妻子、問題、希望、忙碌、討論
 討論主題:工作壓力、生活品質、夫妻角色分工和日常問題的討論。可能是對日常生活和工作狀況的討論或抱怨。

發現跟"結婚"這個儀式相關的問題都圍繞在婚紗、婚顧廠商、婚禮價格等字詞上與"婚姻"相關的字詞則有老婆老公、小孩、工作等等

5. 進一步分析"婚姻":使用社會網絡圖對"婚姻"進行分析

117(32), 19061-19071.

組成員：

N114020010 黃崇祐

N114020012 黃延平

N114320021 花秀旻