

元智大學  
資訊管理學系碩士班  
碩士論文

YouTube 流行影片主題模型分析：影片描述與留言之  
比較

Topic Model Analysis of Popular YouTube Videos:  
Comparison of Video Descriptions and Comments

研 究 生：黃紹瑋

指導教授：楊錦生 博士

中華民國一一三年七月

YouTube 流行影片主題模型分析：影片描述與留言之  
比較

Topic Model Analysis of Popular YouTube Videos:  
Comparison of Video Descriptions and Comments

研 究 生：黃紹瑋 Student: Shao-Wei Huang

指導教授：楊錦生 博士 Advisor: Dr. Chin-Sheng Yang

元智大學  
資訊管理學系碩士班  
碩士論文  
A Thesis

Submitted to Department of Information Management  
College of Informatics  
Yuan Ze University  
in Partial Fulfillment of the Requirements  
for the Degree of  
Master of Science  
in  
Information Management

July 2024

Chungli, Taiwan, Republic of China

中華民國一一三年七月

## 論文口試委員審定書



# YouTube 流行影片主題模型分析：影片描述與留言之比較

學生：黃紹瑋

指導教授：楊錦生 博士

元智大學資訊管理學系碩士班

## 摘要

近年來社群媒體蓬勃發展，平台如 Facebook、Twitter、Instagram 等吸引了全球數十億用戶，這些平台不僅改變了社會互動方式，也促進了新工作型態與商業模式的誕生，吸引了大量學術研究。採用 YouTube 作為本研究的資料集，旨在透過分析 YouTube 影片的元數據資料和文本資料，探討影響影片熱門度的各種因素。研究目標包括分析影片元數據、標題和描述、觀眾評論等對熱門度的影響，並比較不同特徵組合和機器學習模型。首先，從文本特徵出發，使用主題模型（Latent Dirichlet Allocation，簡稱 LDA）進行文本分析。接著，將這些文本特徵與影片的元數據特徵進行比較。然後，通過多個特徵的組合來進行更全面的分析。最後，利用機器學習模型（邏輯回歸、支持向量機、決策樹、隨機森林和梯度提升機）對影片的熱門度進行評估與分析。根據實驗結果，集成分類器的表現通常優於單一分類器，尤其是梯度提升機在寬鬆資料集中的優異表現，而隨機森林則在嚴謹資料集中較為突出。總體來看，元數據特徵對影片熱門度的影響較大，而文本特徵的運用可以進一步提升影片的熱門程度。

關鍵字：社群媒體、主題模型、文本特徵、機器學習

# **Topic Model Analysis of Popular YouTube Videos: Comparison of Video Descriptions and Comments**

Student : Shao-Wei Huang      Advisor : Dr. Chin-Sheng Yang

Department of Information Management

College of Informatics

Yuan Ze University

## **ABSTRACT**

In recent years, social media has experienced rapid development, with platforms like Facebook, Twitter, and Instagram attracting billions of users worldwide. These platforms have not only changed the way people interact socially but also promoted the emergence of new types of work and business models, drawing significant academic research interest. Using YouTube as the dataset for this study, the aim is to explore various factors influencing video popularity by analyzing YouTube video metadata and textual data. The research objectives include analyzing the impact of video metadata, titles and descriptions, and audience comments on popularity, and comparing different feature combinations and machine learning models. First, starting from textual features, Latent Dirichlet Allocation (LDA) is employed for text analysis. Then, these textual features are compared with the video metadata features. Subsequently, a more comprehensive analysis is conducted by combining multiple features. Finally, machine learning models (logistic regression, support vector machine, decision tree, random forest, and gradient boosting machine) are utilized to evaluate and analyze video popularity. According to the experimental results, ensemble classifiers generally outperform individual classifiers, with gradient boosting machines showing excellent performance in lenient datasets, while random forests excel in rigorous datasets. Overall, metadata features have a significant impact on video popularity, and the use of textual features can further enhance video popularity.

**Keywords:** Social Media, Topic Model, Text Features, Machine Learning

## 誌謝

白駒過隙歲月如梭，在元智的碩士生活即將告一段落，感謝我的指導教授楊錦生老師。在碩士求學過程中給予我許多的幫助，在每週的開會中引導我探討相關的文獻、協助我訂定碩士論文的題目以及在論文中遇到的資料問題解惑和實驗結果的指點。在楊老師循序漸進的教導與幫助下，我才能在碩士生活中學習到許多專業知識並運用在學術上或實務上。

再來感謝我的實驗室夥伴，云瑄、宣甫、鈺均、庭葦、明祺。謝謝他們在這段期間幫助了我許多，在我遇到問題時提供我許多寶貴意見、鼓勵。也感謝實驗室的大家常常一起聊天吃飯，使實驗室充滿歡笑。期盼未來每一位都能繼續相互協助，共同進步。還想謝謝：楊毅、敬萱、均庭、彥庭、文齊等，大家常常一起討論論文進度，督促對方共同努力，在學業上互相討論和鼓勵，在生活上也常一起吃飯聊天，能認識你們是很幸運的事，希望大家畢業後都能一路順遂，保持聯絡，順利平安。

最後，我要感謝我的家人和朋友們，給予我全力的支持和鼓勵。有你們的支持陪伴和無條件的幫助下，才能讓我順利完成學業，希望未來能讓我愛的人都感到驕傲，成為有能力照顧自己的人，到達人生下一個階段，謝謝你們。

黃紹瑋 謹誌

中華民國一一三年七月

## 目錄

書名頁.....	i
論文口試委員審定書.....	ii
摘要.....	iii
ABSTRACT.....	iv
誌謝.....	v
表目錄.....	viii
圖目錄.....	x
第一章 緒論.....	1
1.1 研究背景.....	1
1.2 研究動機與目的.....	2
1.3 論文架構.....	3
第二章 文獻探討.....	4
2.1 不同社群媒體上用戶生成資料的流行度.....	4
2.2 影響社群媒體用戶生成資料流行度的相關研究.....	5
2.3 主題模型相關研究.....	10
第三章 研究方法.....	11
3.1 網頁爬蟲.....	12
3.2 資料前處理.....	12
3.3 特徵萃取.....	13
3.4 統計分析.....	16
3.5 機器學習.....	17
3.6 分析與評估.....	18

第四章 實驗結果與設計.....	20
4.1 資料集.....	20
4.2 統計檢定結果分析.....	21
4.3 實驗與結果.....	34
第五章 結論與未來展望.....	55
5.1 結論.....	55
5.2 未來展望.....	56
參考文獻.....	57
附錄 A.....	59
附錄 B.....	81
附錄 C.....	101





## 表目錄

表 -1 相關文獻匯總表 .....	5
表 -2 YouTube 影片的元數據 .....	14
表 -3 運動類別的頻道資訊 .....	20
表 -4 人物&部落客類別的頻道資訊 .....	21
表 -5 娛樂類別的頻道資訊 .....	21
表 -6 運動類別連續值流行度統計檢定結果(寬鬆樣本).....	23
表 -7 運動類別的類別值流行度統計檢定結果(寬鬆樣本).....	24
表 -8 人物&部落客類別的連續值流行度統計檢定結果(寬鬆樣本).....	25
表 -9 人物&部落客類別的類別值流行度統計檢定結果(寬鬆樣本).....	26
表 -10 娛樂類別的連續值流行度統計檢定結果(寬鬆樣本).....	27
表 -11 娛樂類別的類別值流行度統計檢定結果(寬鬆樣本).....	27
表 -12 運動類別的連續值流行度統計檢定結果(嚴謹樣本).....	29
表 -13 運動類別的類別值流行度統計檢定結果(嚴謹樣本).....	30
表 -14 人物&部落客類別的連續值流行度統計檢定結果(嚴謹樣本).....	31
表 -15 人物&部落客類別的類別值流行度統計檢定結果(嚴謹樣本).....	32
表 -16 娛樂類別的連續值流行度統計檢定結果(嚴謹樣本).....	33
表 -17 娛樂類別的類別值流行度統計檢定結果(嚴謹樣本).....	34
表 -18 特徵組合 .....	35
表 -19 LDA 主題模型文檔表 .....	36
表 -20 運動類別特徵組合 .....	37
表 -21 人物&部落客類別特徵組合 .....	39
表 -22 娛樂類別特徵組合 .....	42

表 -23 運動類別特徵組合 .....	45
表 -24 人物&部落客類別特徵組合 .....	48
表 -25 娛樂類別特徵組合 .....	50
表 -26 寬鬆與嚴謹樣本綜合評估表 .....	54



## 圖目錄

圖表 1 研究方法架構圖.....	11
-------------------	----



# 第一章 緒論

## 1.1 研究背景

隨著社群媒體的快速發展，社群媒體經歷了爆炸性的增長，成為全球數十億人日常生活中不可或缺的一部分。從最初的 Facebook<sup>1</sup> 和 Twitter<sup>2</sup>，到後來的 Instagram<sup>3</sup>、Threads<sup>4</sup>，社群媒體平台以其多樣化的功能和便捷的使用體驗吸引了大量用戶。這些社群媒體也大大改變了我們生活的環境，從觀看者到影片的創作者，發展了許多不同的工作型態與商業模式，也因此增加了許多學者對於社群媒體的研究探討。

社群媒體是個充滿討論和互動的社群空間。研究這些社群媒體的動態有助於理解觀眾對於內容的評論、分享和興趣喜好，進而發覺內容是如何在社群媒體上傳播、共享和討論的。例如，在 Twitter 上，推文的轉發和評論量可以反映出社群媒體上用戶生成資料的流行度，在 Instagram 上，發布的圖文互動率則是衡量受眾參與度的重要指標(Li & Zhang, 2017)，在 YouTube 上，透過影片的標題、影片描述等因素都對影片流行度有顯著影響 (Mishra & Singh, 2016)。

社群媒體創作者花費大量的時間為他們發布的內容創建吸引人的標題和介紹。如何吸引觀眾的注意力並引起他們的好奇心和興趣，使觀眾願意點擊和分享內容，是用戶生成資料流行度的關鍵因素。Instagram 的圖片和影片標題，以及 Twitter 的簡短推文，都需要在有限的字數內傳達出強烈的吸引力來吸引觀看者，而影音平台中透過影片的標題、影片描述等因素也都對影片流行度有顯著影響，因此對創作者而言找出用戶生成資料流行度的關鍵因素也為他們必須煩惱的課題 (Mishra & Singh, 2016)。

在眾多社群媒體平台中，YouTube<sup>5</sup> 因其獨特的影片分享和內容創作功能而脫穎而出。自 2005 年創立以來，YouTube 已成為全球最大且最受歡迎的影音分享平台，其用戶數量超過 20 億，並且每天都有超過 10 億小時的影片被觀看(Jia et al., 2023)。觀眾可以在 YouTube 上選擇不同主題和類型的影片來觀看；對於創作者而言，這是一個競爭激烈的環境。因此，如何使用戶生成資料成為流行影片，找尋關鍵方法、深入探討觀看次數的變動，與解析社群動態發展至關重要。

為了在 YouTube 上能夠獲得巨量的觀看次數，以及大量影片的點讚和留言，是任何 YouTube 頻道主都期望的，因為在 YouTube 提供的分潤機制下，越多的人觀賞影片，頻道主所能獲得的分潤就越多，然而對於頻道主而言找尋有效方法來製作用戶生成資料是需要花費大量時間試錯，才可能找到方法讓用戶生成資料成為

---

<sup>1</sup> <https://www.facebook.com>

<sup>2</sup> <https://x.com/?lang=zh-Hant&mx=2>

<sup>3</sup> <https://www.instagram.com/>

<sup>4</sup> <https://www.threads.net/?hl=zh-tw>

<sup>5</sup> <https://www.youtube.com>

流行影片。因此，眾多學者（Kim et al., 2019；Wulczyn et al., 2017；Mishra et al., 2016）也將 YouTube 作為研究主題，對此展開深入的研究，期望找到有效的策略方法，來增加影片的瀏覽量、點閱數，使其用戶生成資料在 YouTube 上成為流行影片。

## 1.2 研究動機與目的

在當今百花爭鳴、遍地開花的數位時代中，YouTube 已成為最能展現影音創作者才華的主要平台之一，無論是長影音或短影音內容都能在此獲得廣泛的觀眾關注。過去的研究顯示，影片的標題和描述文本不僅影響觀眾的點擊率和觀看時間，還可能直接影響影片的熱門度（Mishra & Singh, 2016）。在後續流行度與熱門度為相同詞彙，因為在英文文獻上流行度與熱門度都是統稱 Popularity。隨著社群媒體的不斷發展，對於文字描述如何在影片熱門度形成過程中發揮作用的具體機制和影響仍然需要進一步的深入研究。

近年來的研究開始關注影音內容的多元化特徵及其與觀眾互動的關聯。例如，Borrajó et al. (2021) 指出，影片的描述文本在提高觀眾互動和社群參與方面發揮了關鍵作用。同樣地，Tufekci et al. (2020) 的研究強調了社群媒體平台上文字內容如何塑造形成觀眾對影片的感受和反應。這些研究突顯了文字描述在影音內容中不僅僅是信息傳達的工具，更是影響觀眾互動和社群參與的重要因素。

過去的研究中主要都是用影片元數據(Metadata)或單一文本特徵進行流行度分析(Borrajó et al. 2021；Koller et al. 2022)，本研究旨在透過分析影片的元數據和 YouTube 影片中的三個重要文本內容(影片標題、描述、觀眾留言)，深入探討影響影片熱門度的各種因素。具體而言，本研究將從 YouTube 頻道主和其觀眾的文字行為、用戶生成資料特徵，進行主題模型分析和統計分析與機器學習模型等角度進行綜合探討。

具體研究目標包括：

1. 分析影片元數據資料對熱門度的影響：探討不同類型影片的時長、標題長度與字數、影片的類別、影片上傳日期及 Tag 標籤數量等如何影響影片的熱門度。
2. 分析影片標題和影片描述對熱門度的影響：將每部影片標題和影片描述進行主題模型分析，探索影片標題和描述背後的主題特徵如何影響影片的熱門度。
3. 分析影片中的觀眾評論對熱門度的影響：將每部影片的前 50 條留言儲存為文本，並進行主題模型分析，研究觀眾留言的主題特徵如何影響影片的熱門度。
4. 不同特徵組合對熱門度的影響：透過影片元數據資料與三種文本資料(影片標題、影片描述、影片留言)進行特徵組合，並使用五種不同的機器學習模型進行實證分析，深入探討不同組合特徵對影片熱門度的相關影響。

本研究目的是為影音平台上的創作者提供數據參考，從影片元數據資料和文本資料中找出能幫助創作者優化影片的內涵策略，並提升影片的可見度和流行度。

### 1.3 論文架構

論文將分為五個章節，分別為第一章的緒論，主要說明研究背景、研究動機與目的、研究架構。第二章則進行文獻探討，先從社群媒體流行度及其重要性探討，再來討論影響社群媒體流行度的關鍵因素和主題模型研究相關的參考文獻資料。第三章為研究方法，章節中將介紹研究時所用的方法和技術，包含資料蒐集、網頁爬蟲、資料前處理、熱門度計算與特徵萃取、統計分析與機器學習方法，以及效能評估指標。第四章為實驗結果，在實驗結果中針對不同結果進行分析探討與方法總結。最後第五章為結論與未來展望，將本研究成果做總結並從本研究中找尋並提供未來展望。



## 第二章 文獻探討

在數位時代，社群媒體已成為人們日常生活中不可或缺的一部分。隨著使用者數量的急劇增長，社群媒體上的文字內容對於品牌營銷、政治活動以及社會互動的影響力日益顯著。社群媒體不僅是一個交流的平台，更是一個資訊傳播和影響力展示的重要渠道。通過對社群媒體文字內容的分析，我們可以了解用戶的情感和行為，制定更有效的策略來影響目標受眾。其中，主題模型（Topic Modeling）作為一種常用的自然語言處理技術，能夠有效地從大量的文字內容中提取潛在的主題，幫助我們更深入地理解文字內容的結構和潛在意圖。因此，探討社群媒體文字內容的影響及其相關因素，不僅具有理論意義，更具有重要的實踐價值。

### 2.1 不同社群媒體上用戶生成資料的流行度

社群媒體用戶生成資料的流行度對於任何平台都是至關重要的指標，流行度的高低對品牌發展也具有重要意義。在數位時代，品牌若能夠善加利用社群媒體平台，便能迅速提升其知名度和市場影響力。Chakraborty et al. (2020) 在 Facebook 上探討了使用數據探勘方法來預測社群媒體的績效指標，並評估其對品牌建設的影響。他們的研究表明，社群媒體的績效指標，如流行度，能顯著提高品牌知名度和市場影響力(Chakraborty & Basu, 2020)。此外，Khan et al. (2021) 對現有社群媒體(Twitter、Facebook、Instagram)預測研究進行了綜合回顧，強調了流行度預測在商業和營銷領域的應用價值，幫助企業制定更有效的市場策略(Khan & Alghazzawi, 2021)。不僅如此，社群媒體流行度的影響力還延伸至政治和公共領域。Smith et al. (2021) 的研究表明，社群媒體活動可以用來預測重要事件的結果，如選舉。他們以 2020 年美國總統選舉為例，展示了 Twitter、Facebook 社群媒體活動數據如何用來預測選舉結果，進一步證明了社群媒體流行度在政治和公共領域的影響力(Smith & Lee, 2021)。

社群媒體平台中，YouTube 自 2005 年開始成為眾人日常影音娛樂的重要平台，YouTube 提供了開放式的平台，讓任何人都可以輕鬆地成為影音創作者去上傳和分享自己影片的內容。這也大大降低了影音製作的進入門檻，開啟了大傳播時代使許多人都成為創作者爭相加入其中，並且在 YouTube 等平台中引起了大量的參與和廣大的回響(Jia et al., 2023)。在 YouTube 成為主要的社群媒體平台後，人們可以在此分享和觀看彼此的日常生活和對不同事物的觀點。影音創作者花費大量時間來製作不同類型的影片，進而提供觀看者獲得不同的知識或娛樂休閒外，也逐漸變成許多創作者的主要賺錢管道來源。YouTube 作為全球最大的影音共享平台，其內容分析成為研究的熱點。多樣式的研究從不同角度來深入探討分析影片內容的類別和觀看用戶行為模式。例如，通過研究分析 YouTube 上影片的標題、描述以及縮圖來觀察影片的流行度 (Choudhury et al., 2018)。利用數據分析和機器學習技術，探討了影片的內容特徵如何影響其在社群媒體上的表現。

## 2.2 影響社群媒體用戶生成資料流行度的相關研究

影響社群媒體用戶生成資料流行度的因素可以分為多個層面，不同平台影響的因素也不相同，例如在 Instagram 上，Gao et al. (2018) 研究了影響 Instagram 圖片和影片流行度的關鍵因素，建立了預測模型，結果顯示發布時間、使用的標籤和內容的視覺效果是主要影響因素(Gao & Wu, 2018)。在 Twitter 上，Rao et al. (2018) 通過情感分析測量 Twitter 用戶的正面或負面影響，研究發現流行用戶對其受眾的情感影響顯著，這些影響進而影響貼文流行度(Rao & Srivastava, 2018)。在 YouTube 上，Mishra et al. (2016) 通過回歸分析研究 YouTube 影片流行度的預測，結果顯示影片的標題、描述和標籤等因素對其流行度有顯著影響(Mishra & Singh, 2016)。

對於 YouTube 研究學者而言影片觀看數為影片的熱門指標，但當影片觀看數多後，要如何將其應用去計算影片的熱門度，則是重要的探討議題，因為一個影片的好壞絕大程度建構在影片的熱門或不熱門上，而如何去計算出影片的熱門度就是大家所去探討的方向，在 Koller 和 Grabner (2022)除了深入探討縮圖和標題對影音觀看量的影響外，提出了一種計算影片熱門度的方法來做熱門度的計算，本研究也將參考利用此方式作為熱門度計算方法(Koller & Grabner, 2022)。而我們也彙整了之前在社群媒體上用戶生成資料的相關研究在表 1。

表 -1 相關文獻匯總表

文獻	研究問題	分析對象	分析方法	變數特徵	結果
Jia et al. (2023)	平台創作者和觀眾之間的互動。	YouTube	質性分析方法	文本分析和案例研究	平台將繼續發展，對全球文化和社會產生深遠影響。
Borrajó et al. (2021)	影片描述對觀眾參與度的影響。	YouTube	實證研究分析	關鍵詞豐富度、內容長度、點擊率、評論數量和觀看時間	精心設計的影片描述可以顯著提高觀眾參與度



文獻	研究問題	分析對象	分析方法	變數特徵	結果
Tufekci et al. (2020)	社群媒體在現代抗議活動中的角色和影響。	Twitter	質性分析和量化數據分析	用戶行為、信息傳播模式以及抗議活動的組織結構	社交媒體需要與傳統的抗議策略結合，才能實現持久的社會變革。
Chakraborty et al. (2020)	使用數據探勘方法來預測社群媒體的績效指標。	Facebook	支持向量機、生命週期貼為消費者模型	類別、頁面總讚數、發布月份、發布小時、發布星期	流行度對品牌建設有顯著影響
Khan et al. (2021)	社群媒體預測研究的綜合回顧，分析不同方法模型預測流行度的應用。	Twitter、Facebook、Instagram	回歸演算法、分類演算法、關聯規則或聚類	文本分析和案例研究	流行度預測在商業和營銷領域具有重要應用價值
Ding et al. (2017)	通過分析用戶貼文和圖片特徵來預測社群媒體中貼文的流行度。	Instagram	決策樹、隨機森林、線性回歸、自適應增強、K 近鄰、多層感知器回歸、TensorFlow 深度回歸	關注數、關注者數、評論數、日期、愛心數、類型、影片時間、影片觀看次數、圖片特徵是具有類別對象的布靈值	用戶的互動行為、貼文內容的質量和圖片的視覺吸引力是影響貼文流行度的重要因素。

文獻	研究問題	分析對象	分析方法	變數特徵	結果
Rao et al. (2018)	通過情感分析測量 Twitter 用戶的正面或負面影響。	Twitter	情緒分析 使用了 LIWC 詞典，皮爾森相關分析、斯皮爾曼等級相關分析和格蘭傑因果分析	皮爾森相關性分析:回覆+提及+轉推。斯皮爾曼等級相關性:總觀眾數、積極用戶數和消極用戶數。 格蘭傑因果分析:關注人數、轉發和提及	Twitter 上的流行用戶對其受眾的情感影響顯著，這些影響進而影響貼文流行度。
Gao et al. (2018)	研究影響 Instagram 圖片和影片流行度的因素，並建立預測模型。	Instagram	線性回歸、局部多項式回歸、支持向量機、線性支持向量機、k 近鄰、隨機森林、決策樹、樸素貝葉斯（核）	愛心數、評論數和影片瀏覽數	影響 Instagram 內容流行度的主要因素包括發布時間、使用的標籤和內容的視覺效果。
Koller et al. (2022)	研究縮圖和標題對點擊率的影響。	YouTube	使用自然語言處理模型、熱門成功定義	影片類別、評論數、按讚數、標題長度、發布日期	優質的縮圖和吸引人的標題能顯著提高流行度。

文獻	研究問題	分析對象	分析方法	變數特徵	結果
Kim et al. (2019)	通過機器學習方法分析縮圖的視覺屬性，預測 YouTube 品牌頻道的流行度。	YouTube	支持向量機、隨機森林、極限梯度提升、線性回歸計算平均絕對百分比誤差	觀看次數、時長、標題長度、訂閱者數量、色彩豐富度、亮度、物體複雜度、元素複雜度	縮圖的視覺屬性，如顏色、構圖和文字，對影片的流行度有顯著影響。
Wulczyn et al. (2017)	使用主題模型識別 YouTube 上的辱罵性評論，分析其對影片流行度的影響。	YouTube	LDA 主題模型	影片標題、影片描述、按讚數、3 業評論數、留言按讚數	辱罵性評論會對影片的流行度產生負面影響。
Mishra et al. (2016)	通過回歸分析研究 YouTube 影片流行度的預測。	YouTube	逐步回歸模型	影片 ID、上傳者、年齡、分類、長度、觀看次數、評分、評論、相關 ID	影片的標題、描述和標籤等因素對其流行度有顯著影響，回歸模型能有效預測影片流行度。

文獻	研究問題	分析對象	分析方法	變數特徵	結果
Choudhury et al. (2018)	通過分析 YouTube 流行影片的縮圖來研究不同文化對影片流行度的影響。	YouTube	深度學習模型	在 MS-COCO 數據集上對其進行預訓練	不同文化背景的用戶對縮圖的偏好存在差異，這些差異影響影片的全球流行度。
Li et al. (2017)	針對不同類別的貼文進行流行度預測，分析各類別的特定影響因素。	Instagram	RSVR、LSVR、RFR、MLPR	類別混合、特定類別	不同類別的貼文受不同因素影響，精細化的流行度預測模型能提高預測準確性。
Smith et al. (2021)	研究了社群媒體活動對政治事件結果的預測。	Twitter、Facebook	情感分析	貼文數量、轉發次數、評論數量	社群媒體數據能夠準確預測 2020 年美國總統選舉的結果，顯示了社群媒體在選舉預測中的潛力。

在過去的相關文獻中可以發覺，在社群媒體用戶生成資料的流行度中，不管是 Twitter、Facebook、Instagram、YouTube 都有許多學者對不同平台做研究，然而可以發現 YouTube 是被單獨研究最多的社群媒體平台，並且可以發現在變數特徵中最常使用的都是 YouTube 的元數據特徵如影片類別、評論數、按讚數、觀看數等等，而在機器學習中常見的分析方法模型是使用線性回歸、決策樹、支持向量機與梯度提升機等回歸和分類模型為主。

## 2.3 主題模型相關研究

主題模型 (Topic Modeling) 是一種用於統計分析的技術模型，用於將大量的文本資料中找尋識別隱含的主題文字。Blei 等 (2003) 提出的 Latent Dirichlet Allocation (LDA) 是最常使用的主題模型之一，它假設文檔是由多個主題的混合組成，而每個主題又是由多個詞相依組成 (Blei et al., 2003)。LDA 已被廣泛應用於文字探勘和信息檢索上，證明其在找尋文本潛在結構方面的有效性，並且主題模型在社群媒體的數據分析應用中也受到了廣泛的關注。目前對於 YouTube 影片內容去做主題模型分析中，也已有了多項研究應用在 LDA 和其他主題模型技術。Wulczyn et al. (2017) 使用主題模型識別 YouTube 上的辱罵性評論，分析其對影片流行度的影響，結果顯示，辱罵性評論會對影片的流行度產生負面影響 (Wulczyn & Dixon, 2017)。也因此本研究將主題模型與情感分析運用到影片的標題和描述以及留言中來做應用。



### 第三章 研究方法

本章將介紹研究流程與研究方法以及研究架構，圖 1 為此章的研究方法架構圖，YouTube 的資料收集使用 YouTube API 來爬蟲抓取分為結構化的 YouTube 元數據資料(metadata)和非結構化的 YouTube 文字資料，文字資料分為三大類：影片標題、影片描述、影片留言，並將這些資料進行資料前處理、特徵萃取中會介紹元數據特徵和主題特徵，最後會進行機器學習和統計分析與效能評估；接著在後續小節中會一一做介紹。

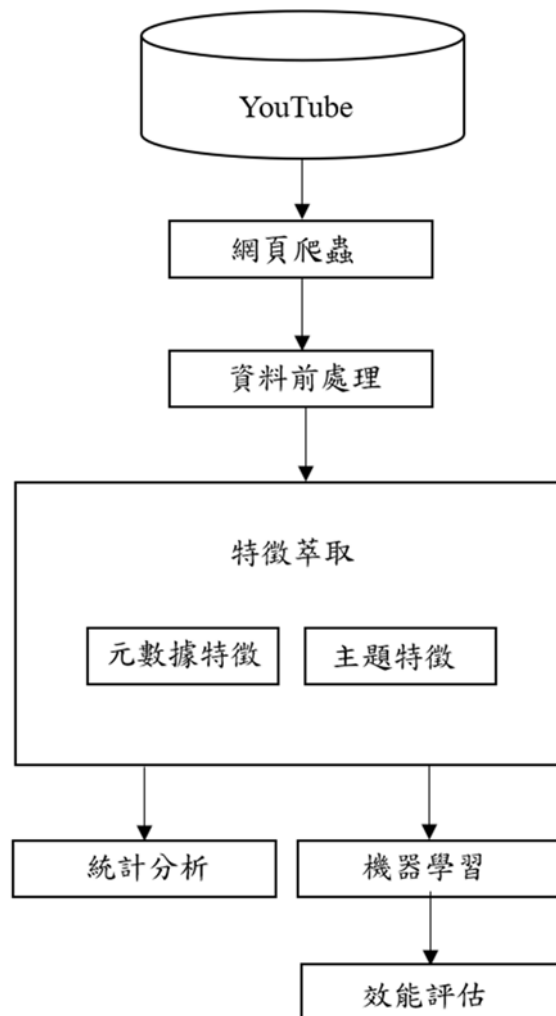


圖 1 研究方法架構圖

### 3.1 網頁爬蟲

本研究使用 YouTube API 來爬蟲抓取頻道內容，在建立 YouTube API 之前需要先去 Google Cloud 中建立一個專案，並且新增一個 YouTube API 憑證，以用於爬蟲所需，在 YouTube API 中最常使用的抓取介面方法有很多，例如影片的基本資訊需要藉由 Snippet 來取得，通常要取得跟影片中的固定資訊都需要這個抓取介面來獲得。根據影片固定資訊可以回傳的內容包含：影片編號(id)、影片的發佈時間 (published At)、頻道序號 (channel Id)、頻道名稱 (channel Title)、影片標題 (title)、影片描述 (description)、影片的類別(category Id)、影片的網址(video url)、影片的縮圖網址、影片 Tag 標籤資料 (tags)、影片的留言(comment)，至於影片的統計資料通常會藉由 statistics 的方法來取得，通常要取得跟影片有關的數值統計資料都會需要使用到這個介面，根據這部影片的資料，我們可以得到的內容包含：影片的觀看次數 (view Count)、喜歡數 (like Count)、留言次數 (comment Count)、影片時長(duration)，並且會抓取使用這個 API 時的當下影片次數為基準。

本研究從 YouTube 排名網站 SOCIAL BLADE<sup>6</sup>提供的 YouTube 頻道分類中選了三類 (運動類、娛樂類、人物&部落客類)，各從三類裡依照不同粉絲訂閱數等級取出了各 10 個 YouTube 頻道，三個類型加總共 30 個 YouTube 頻道資訊，而抓取時間都是從 2022 年 1 月 1 日開始抓取至 2023 年 10 月左右，並且只抓取長影片內容並不包含直播和 shorts 短影音內容。

### 3.2 資料前處理

蒐集下來的 YouTube 資料分為兩大部分，一部分為 YouTube 的元數據資料，而另一部分為 YouTube 的文字資料，分為影片標題、影片描述、影片留言共三部分，然而因為 YouTube 中留言資料數量太多，因此我們參照 Wulczyn 等人(2017)在他們的研究中將每部影片的前三頁評論合併為一個評論文本的做法(Wulczyn et al., 2017)。在本研究因為影片留言數量太多，因此選擇將每部影片中的前 50 個影片留言合併為一個文本資料。

本研究使用了由台灣中央研究院資訊科學研究所開發的中文斷詞系統 CKIP<sup>7</sup> (Chinese Knowledge and Information Processing)，其中的 CKIP Tagger 是一款專門用於繁體中文文本的分詞和詞性標註工具。這對於處理中文自然語言處理(NLP)任務，如文本分析、機器翻譯和信息檢索等，具有重要的應用價值。

停用詞 (stopwords) 是在文本處理過程中被過濾掉的常見詞彙，這些詞通常不會對文本的主要信息產生實質性影響，如“的”、“是”、“在”等。停用詞中使用 cnsstopwords<sup>8</sup>，在中文文本分析中，去除停用詞是常見的預處理步驟之一，其目

<sup>6</sup> <https://socialblade.com>

<sup>7</sup> <https://ckip.iis.sinica.edu.tw>

<sup>8</sup> [https://cishell.wiki.cns.iu.edu/Stopword-List\\_2785749.html](https://cishell.wiki.cns.iu.edu/Stopword-List_2785749.html)

的是提高後續分析的準確性和效率。本研究中會將 YouTube 的文字資料內容進行 CKIP Tagger 斷詞處理，並使用 cnsstopwords 停用詞來過濾掉常見的詞彙，並確保所有文本不為空值。

在研究中，我們會將 YouTube 文字資料中的文本三大類：影片標題、影片描述、影片留言，在資料前處理中，會將文本資料使用中央研究院的 CKIP Tagger 進行斷詞和停用詞處理。

本研究中為了找出影片的熱門度計算方式，因此使用了 Koller 和 Grabner (2022) 提供了一種計算影片熱門度的方法，他表明了 YouTube 影片的成功通常取決於它獲得的觀看次數 (Koller & Grabner, 2022)。

因此我們熱門度計算方式為：

$$\frac{\text{觀看影片次數}}{\text{影片頻道者總體訂閱數}} \quad \text{式 1}$$

而本研究在判定影片熱門不熱門時，也使用了兩種計算方法將資料分為兩部分，其中一部分為寬鬆樣本資料：做法是將頻道主的影片資料依照熱門度計算公式出來的各影片成績使用二分法去區分影片的熱門或不熱門；而另一做法為嚴謹樣本資料：將各個頻道主資料去做分割，依照熱門度計算公式計算出來的各影片成績，依照四分位法計算，取出成績的前 25% 記錄為熱門影片，後 25% 記錄為不熱門影片。

### 3.3 特徵萃取

特徵萃取 (Feature Extraction) 是自然語言處理 (NLP) 中的關鍵步驟，用意是從文本中提取出有用的特徵，以便進行後續的數據分析和建模。YouTube 影片的文字資料和 YouTube 的影片元數據資料中我們也使用了不同的特徵萃取技術。

#### 3.3.1 元數據特徵

將 YouTube 的影片的元數據資料轉換為 YouTube 的影片數值資料中會使用到中文情感字典 CVAW (The Chinese Valence-Arousal Words) 進行情感分析。在本研究中，使用了 CVAW 中文維度型情感字典 (Lin and Yu, 2020)。該字典包含詞彙的正面和負面情感程度。首先，對文本進行斷詞處理，然後從情感字典中提取包含在其中的情感詞彙以及它們的情感分數。每個詞彙的情感分數範圍是 1 至 9，以 5 作為正面和負面之間的分界線。而本研究在情感特徵最大值和最小值的計算上，在處理每條留言或切換到新影片時，程式碼會比較當前留言的情感分數與影片的最大值和最小值，如果當前留言的情感分數大於影片目前的最大值，則更新最大值，如果當前留言的情感分數小於影片目前的最小值，則更新最小值。而在計算情感平



均值上，平均值是在每部影片的所有留言的情感分數總和的基礎上計算的。首先，對於每條留言，我們將其情感詞彙的分數加總起來，得到該留言的情感總分。然後，將每條留言的情感總分相加，得到整部影片的情感分數總和。最後，將整部影片的情感分數總和除以留言總數，得到的就是該影片的平均情感分數。這個平均值反映了整部影片留言中的平均情感傾向。影片總分為所有留言的情感分數總和；留言總數為該影片的所有留言數量。

平均值公式為：

$$\text{平均值} = \frac{\text{影片總分}}{\text{留言總數}} \quad \text{式 2}$$

將 YouTube 的影片的元數據資料將其轉換為 YouTube 的影片數值資料。如下表 2：

表 -2 YouTube 影片的元數據

變數中文名稱	變數英文名稱	計算方式
影片類別	Category Id	YouTube 提供的影片類別編號。
影片時長	duration	影片的長度以秒數計算。
影片按讚數	Like Count	抓取時間當下影片的按讚總數。
影片留言數	Comment Count	抓取時間當下影片的留言總數。
Tags 數量多寡	Number of Tags	影片的Tags總數。
影片上傳星期	weekday	一週七日中發布的日子。
影片上傳早上、晚上	Day night	24小時制，前12小時為am，後12小時為pm。
影片上傳間隔	Time interval	以日為單位計算與前一部影片上傳的間隔時間。
影片上傳是否過十四天	time_14days	創作者更新頻率以14天為基準，是否在十四天以內更新。
影片標題長度	Title length	計算標題中的字元數量。
影片標題詞數	Word count	將標題進行中文斷詞，保留每個詞的詞性信息

		(比如名詞、動詞等)，計算其中包含的詞數。
是否為延伸續集	Has episode	檢查影片標題中是否包含(ep, 第i集, 續集)中的任何一個關鍵字。
是否與網紅或專家合作	Has feat	檢查影片標題中是否包含 (@, feat., ft., 合作)中的任何一個關鍵字。
是否為實測主題	Has measured	檢查影片標題中是否包含(實測, 測試, 實驗, 數據, 數據收集, 測量, 試用)中的任何一個關鍵字。
是否為日常主題	Has daily	檢查影片標題中是否包含(日常, 生活, 分享, 體驗, 日記, 紀錄, vlog)中的任何一個關鍵字。
情感特徵平均值	CVAW average	平均情感分數是影片的總情感分數除以影片的留言數量。
情感特徵最大值	CVAW max	在處理每條留言或回覆時, 程式碼會比較當前留言的情感分數與影片的最大值 (Max)。
情感特徵最小值	CVAW min	在處理每條留言或回覆時, 程式碼會比較當前留言的情感分數與影片的最小值 (Min)。

### 3.3.2 主題模型

Blei 等人於 2003 年提出了隱含狄利克雷分布 (Latent Dirichlet allocation, 簡稱 LDA), 這是一種無監督學習方法, 用於推測文檔的主題分佈。該方法將每個文檔視為一組潛在主題的混合, 通過對文檔進行主題概率分佈建模, 能夠對文檔進行摘要表達。

TF-IDF (詞頻-逆文檔頻率) 是一種常用的文本分析和信息檢索的統計方法。這種方法改進了詞袋模型 (Bag-of-Words, BoW), 透過單詞在在單個文件中出現的頻率 (詞頻, TF) 和該詞語在整個文檔集出現的情況 (逆文檔頻率, IDF) 來評估

詞語的重要性。TF-IDF 這兩項指標的乘積可以有效地減少常見詞彙對文本表示的影響，並能加以呈現那些能夠在不同文檔內容中區分的關鍵詞。簡單來說，TF-IDF 幫助我們找出每篇文章中最重要詞，讓我們更容易理解文本的主要內容。TF-IDF 值通過將詞頻和逆文檔頻率相乘來獲得，其計算公式為： $TF-IDF(t, d, D) = TF(t, d) \times IDF(t, D)$ 。t 代表詞語，d 代表文件，D 代表所有文本的總數。

主題模型中各主題的權重分析是一種評估主題模型結果的方法，它能夠讓我們更深入地理解文檔集合中的特徵詞彙和主題之間的關係。這種分析首先對每個主題中的詞彙進行分析，這有助於瞭解每個主題代表的主題類型，以及它們的特徵詞彙有哪些，接著我們計算每個主題在整個文檔集合中的分佈情況，這可以幫助我們了解每個主題在整個文檔集合中的重要性，然後我們使用某種方法(例如主題在所有文檔中出現的頻率)來計算每個主題的權重，這可以通過主題在文檔集合中的相對出現頻率或其他指標來實現。對於每個文檔，我們回推分析其與每個主題的關聯性，這有助於我們評估每個主題的重要性的影響。

通過主題模型運用於我們 YouTube 文字資料中的文字資料三大文本類別：影片標題、影片描述、影片留言，我們先將各文本創建一個 TF-IDF 矩陣，來從 TF-IDF 矩陣中過濾減少常見詞語並提取每篇文本的重要詞語，再運用於主題模型中，為了確定最佳的主題數量，程式迭代不同的主題數量範圍，使用 LDA 模型計算每個主題數量對應的一致性得分 (coherence)，並選擇最佳主題數量。隨後，使用最佳的主題數量來訓練 LDA 模型，通過指定適當的迭代次數來提高模型的效果和穩定性。通過文本分析和主題建模，幫助使用者深入瞭解留言的主題分佈和關鍵詞語，而產生出來的主題模型結果能夠更好地理解各文本中的主題類型和主題之間的關係，從而為進一步的研究和分析提供重要的參考依據。

### 3.4 統計分析

當前研究將 YouTube 影片的元數據資料轉換為類別值資料和連續值資料，並對這些資料進行敘述統計與顯著性檢定。在統計分析過程中，首先對連續值資料進行平均值、標準差等計算。接著，通過使用 T 檢定 (T-test) 來比較熱門與不熱門影片特定變數的平均值是否存在顯著差異，並評估兩者因素對這些差異的影響，從而進一步理解這些變數對於影片熱門程度的影響。最後，對結果進行解釋，以全面了解數據特性並辨別可能存在的差異。

對於類別值資料，將使用卡方檢定 (Chi-square test) 中的獨立性檢定來評估不同類別變數之間的關聯性。卡方檢定用於測試兩個類別變數之間是否獨立，即它們是否彼此相關。本研究中卡方檢定用於比較類別值資料中不同變數與影片的熱門程度之間的關係。具體來說，它將考慮每個變數在不同熱門程度下的分佈情況，並評估這些變數是否與影片熱門程度有關聯性。在進行卡方檢定時，常用的是建立一個列聯表 (contingency table)，表中包含了不同變數類別在熱門和非熱門影片中的分佈情況。然後，檢定統計量 (卡方值) 將根據這些分佈情況來計算，從而評估變

數與影片熱門程度之間是否存在顯著的關聯性。這種方法可以幫助分析者理解哪些變數與影片的熱門程度有較強的相關性。在統計分析中，P 值的計算和解釋是評估觀察到的差異是否具有統計學上顯著意義的關鍵步驟。P 值代表在零假設（即沒有實際差異或效應）的情況下，觀察到目前數據或更極端情形的概率。

總體來說，本研究主要探討熱門與不熱門之間的比較，T 檢定主要用於比較兩組資料的平均值是否有顯著差異，而卡方獨立性檢定則用於評估兩個類別變數之間的相關性或獨立性。在統計分析中，這兩種檢定方法能夠提供更深入全面的數據理解，幫助我們更好地了解數據中存在的模式和趨勢，從而更準確地評估資料。

### 3.5 機器學習

在機器學習模型的使用上，本研究使用了 scikit-learn 的機器學習方法進行分類模型，模型總共有 5 種，包含單一分類器 (Single classifiers)：邏輯式回歸 (Logistic Regression)、決策樹 (Decision Tree)、支持向量機 (Support Vector Machine)，以及集成分類器 (Ensemble classifiers)：隨機森林 (Random Forest)、梯度提升機 (Gradient Boosting Machine)。

在模型測試的基礎上，模型選擇是預測分析過程中的重要步驟。使用不同機器學習模型，並比較它們的性能來選擇最適合的模型。首先，透過評估指標如準確度、精確度和召回率，比較了解各個模型在模型測試階段的表現。考慮了模型的預測能力、泛化能力和適應性，以確定哪類模型在分析 YouTube 熱門流行趨勢預測問題上表現最佳。

#### 3.5.1 單一分類器

##### ■ 邏輯回歸 (Logistic regression) (Menard, 2010)

這是一種用於解決二元分類問題的機器學習算法。雖然名為回歸，但它是處理分類的演算法，主要是為了尋找出一條直線能夠將所有資料清楚地切分開來並做分類，並假設其變數相互獨立，專門解決二元分類的問題，常應用於統計學上的分析方法。

##### ■ 決策樹 (Decision Tree) (Lantz, 2019)

這是一種基於樹狀結構的監督式學習算法，類似於流程圖的樹狀結構圖，樹內部的每一個節點代表對一個特徵或屬性，通過選擇最佳屬性來拆分數據，使得每個子節點中的數據更加純淨，直到達到滿足停止條件的葉節點。而樹的葉節點代表一種分類結果，給出每個類別的占比，機率最大的類別作為這個樹葉節點的預測值。

##### ■ 支持向量機 (Support Vector Machine) (Géron, 2019)

這是一種主要用於解決分類問題的機器學習算法。其核心思想是在數據與數據之間找到一個最佳的分隔邊界，這個分隔邊界可以將不同類別的數據分開支持向量機通過最大化邊界來確定最佳分類超平面，從而實現對數據的有效分類。簡而

言之，支持向量機是通過尋找最佳分隔邊界來將不同類別的數據分開的一種機器學習算法。

### 3.5.2 集成分類器

#### ■ 隨機森林 (Random Forest) (Hastie& Friedman, 2009)

這是一種用於解決分類問題的集成學習方法。它是通過構建多個決策樹來進行分類，隨機森林通過在數據集的不同子集上訓練多個決策樹，然後將這些決策樹的預測結果進行集成，最終得到一個更加穩定和準確的預測結果。

#### ■ 梯度提升機 (Gradient Boosting Machine) (Chen & Guestrin, 2016)

這是一種通過將多個弱學習器 (通常是決策樹) 逐步組合成一個強學習器的機器學習模型。其核心思想是通過不停反覆的迭代，並且每次迭代都會將新的弱學習器添加到集成模型中，並根據前一輪的結果調整樣本的權重，以使新的學習器能夠更好的跟之前的數據擬合。通過這種方式，梯度提升機能夠逐步提升整個模型的性能，並在預測準確度上取得很好的表現。

## 3.6 分析與評估

為了充分利用所有 YouTube 的特徵資料，本研究採用了十折交叉驗證方法。這種驗證方法將資料集分成十份，輪流選擇其中九份作為訓練集，並將剩餘的一份作為測試集。進行十次訓練和測試後，將每次的結果取平均值，以確保模型的性能。

在模型評估方面採用準確率 (Accuracy)、精確率 (Precision)、召回率 (Recall)、調和平均數 (F1-score)、AUC，作為分類模型評估指標。將驗證方法應用在前述的模型中，用於分析不同特徵組合對影片熱門度的影響，從而評估每個模型的效能。

分類的結果分為四種象限：

1. True Positive (TP)：實際上是熱門影片且被預測為熱門影片
2. False Positive (FP)：實際上是不熱門影片但被誤判為熱門影片
3. False Negative (FN)：實際上是熱門影片但被誤判為不熱門影片
4. True Negative (TN)：實際上是不熱門影片且被預測為不熱門影片

■ 準確率 (Accuracy)：預測熱門影片成功的比率。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

式 3

■ 精確率 (Precision)：判斷為熱門影片的情況下，有多少實際為熱門影片的比率。

$$Precision = \frac{TP}{TP + FP} \quad \text{式 4}$$

■ 召回率 (Recall)：在真實的熱門影片情況下，有多少被正確判斷出來是熱門影片的比率。

$$Recall = \frac{TP}{TP + FN} \quad \text{式 5}$$

■ 調和平均數 (F1-score)：精確率與召回率的調和平均。

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad \text{式 6}$$

■ AUC (Area Under the Curve)：ROC 曲線之下所覆蓋的面積，若 AUC 值越趨近於 1 的分類器，表示分類器效能越好、模型對正例和負例的區分能力越強，正確率則越高。如果 AUC 值等於 0.5，表示分類器的效果與隨機猜測相當。相反，如果 AUC 值為 1，則表示分類器能夠完美區分正例和負例。

## 第四章 實驗結果與設計

### 4.1 資料集

本研究所使用的資料樣本，使用 Python Requests 函式庫和 YouTube API，替換相應的 API 金鑰和影片 ID，獲取需要的 YouTube 頻道影片數據，如標題、觀看次數、按讚數和留言數等…。在資料抓取的時間區域則是從 2022 年 1 月 1 日開始到使用 API 抓取當日，依序抓取 YouTube 頻道者發佈的影片。抓取的資料樣本包含了三個 YouTube 類別（運動類別、人物&部落客類別、娛樂類別），每個類別各 10 位，總共 30 位 YouTube 頻道創作者。資料樣本內容如表 3、表 4、表 5。

在資料樣本中運動類別所抓取的合計影片數量為：2,343；人物&部落客類別所抓取的合計影片數量為：1,259；娛樂類別所抓取的合計影片數量為：1,383 筆樣本數，總計資料樣本數為 4,985 筆影片資料。

而在實驗數據中，本研究依照資料前處理中所述：影片熱門度的計算方法，將全部資料分為兩部分，其中一部分為寬鬆資料樣本：是將頻道主的影片資料利用二分法將流行度排序，從中間區分前半段為影片的熱門後半段為影片的不熱門；另一部分為嚴謹資料樣本：為了將邊界樣本移除，將頻道主的影片資料使用四分位法區分，第一分位為熱門影片，第四分位為不熱門影片。

表 -3 運動類別的頻道資訊

頻道名稱	頻道訂閱數	頻道影片抓取數
HU 理 HU 途	6 萬	180
轟酷看籃球	6.89 萬	310
likemikesports	10.3 萬	340
球給彭尊	12.3 萬	107
GM 冠名播出	15.4 萬	394
夢多 TV	18.9 萬	128
豹子腿	21.6 萬	351
台南 Josh	37.1 萬	322
Petta 哥哥	75.4 萬	91
館長	109 萬	120

表 -4 人物&amp;部落客類別的頻道資訊

頻道名稱	頻道訂閱數	頻道影片抓取數
布萊恩的世界	6.52 萬	117
阿鑫	9.5 萬	75
陸子玄 Lois	13.2 萬	90
琳妲	19.3 萬	96
Blaire Chen	57.8 萬	66
壹加壹	13.2 萬	90
一隻阿圓	82.2 萬	181
滴妹	146 萬	86
千千進食中	181 萬	135
Joeman	257 萬	311

表 -5 娛樂類別的頻道資訊

頻道名稱	頻道訂閱數	頻道影片抓取數
油條兄弟	9.26 萬	199
孫腫來了	16.7 萬	160
進藤喜彥	29.2 萬	91
安安邊緣子	30.8 萬	96
林嘉凌 薔薔	46.5 萬	132
恩熙俊	53.1 萬	115
The DoDo Men	76.2 萬	76
Hook	138 萬	91
HowFun	149 萬	139
黃氏兄弟	219 萬	284

## 4.2 統計檢定結果分析

本節將說明資料樣本的統計檢定中的觀察結果。本研究將三大類頻道中各頻道主 YouTube 的影片資料去做敘述性統計、T 檢定、卡方檢定，並且再依照各類別頻道去做整理。因篇幅關係，敘述性統計資料結果檢視於附錄 A、附錄 B、附錄 C。

將 YouTuber 的影片資料中的連續值資料執行敘述性統計和 T 檢定；類別值資料執行敘述性統計和卡方檢定。在統計檢定中只使用影片發布前的資料來作為觀察影片是否熱門的主要因素，因此在 YouTuber 的影片資料中，將已知發布影片後



才產生的資料去做刪除，包括了：按讚數、評論數、觀看數，這些數值都是影片發布後才產生的資料，不屬於在這次統計檢定的觀察中。

#### 4.2.1 寬鬆樣本資料統計檢定結果分析

根據表 6 與表 7 顯示，可以觀察到運動類別連續值流行度統計檢定結果(寬鬆樣本)中，影片時長裡較為顯著  $p<0.05$  中有 8 個創作者，不顯著中則是 2 個創作者，所以在頻道顯著比例中為 0.8，熱門影片趨勢中顯示在熱門與不熱門的比較中，有 5 個創作者發布的熱門影片為較短影音，3 個創作者發布的熱門影片為較長影音。情感特徵平均值裡較為顯著  $p<0.05$  中有 4 個創作者，不顯著中則是 6 個創作者，所以在頻道顯著比例中為 0.4，熱門影片趨勢中顯示在情感特徵平均值的比較中，有 2 個創作者的值較不熱門影片小，有 2 個創作者的值較不熱門影片大。情感特徵最大值裡較為顯著  $p<0.05$  中有 9 個創作者，不顯著中則是 1 個創作者，所以在頻道顯著比例中為 0.9，熱門影片趨勢中顯示在熱門影片創作者中有 9 個都是極為正向的。情感特徵最小值裡較為顯著  $p<0.05$  中有 7 個創作者，不顯著中則是 3 個創作者，所以在頻道顯著比例中為 0.7，熱門影片趨勢中顯示在熱門影片創作者中有 7 個都是極為負向的。標籤數量裡較為顯著  $p<0.05$  中有 5 個創作者，不顯著中則是 5 個創作者，所以在頻道顯著比例中為 0.5，熱門影片趨勢中顯示共有 5 個創作者的標籤數量是較多的。影片標題長度裡較為顯著  $p<0.05$  中有 5 個創作者，不顯著中則是 5 個創作者，所以在頻道顯著比例中為 0.5，熱門影片趨勢中顯示有 5 個創作者的標題長度是較長的。影片標題詞數中較為顯著  $p<0.05$  中有 7 個創作者，不顯著中則是 3 個創作者，所以在頻道顯著比例中為 0.7，熱門影片趨勢中顯示有 7 個創作者的標題詞數是較多的。影片上傳間隔裡較為顯著  $p<0.05$  中有 3 個創作者，不顯著中則是 7 個創作者，所以在頻道顯著比例中為 0.3，熱門影片趨勢中顯示有 3 個創作者發布的影片與前一部影片相比間隔天數較長。

運動類別的類別值流行度統計檢定結果(寬鬆樣本)中，影片類別裡較為顯著  $p<0.05$  中有 5 個創作者，不顯著中則是 5 個創作者，所以在頻道顯著比例中為 0.5，熱門影片趨勢中顯示在熱門與不熱門的比較中，有 4 個創作者的影片類別為 17(體育)，1 個創作者的影片類別為 20(遊戲)。是否為延伸續集裡較為顯著  $p<0.05$  中有 3 個創作者，不顯著中則是 7 個創作者，所以在頻道顯著比例中為 0.3，熱門影片趨勢中顯示 1 個創作者在影片不是續集時影片較熱門，2 個創作者在影片有續集時影片較熱門。是否為實測主題裡較為顯著  $p<0.05$  中有 0 個創作者，不顯著中則是 10 個創作者，所以在頻道顯著比例中為 0，熱門影片趨勢中顯示為實測影片對於熱門與否較不顯著。是否為日常主題裡較為顯著  $p<0.05$  中有 2 個創作者，不顯著中則是 8 個創作者，所以在頻道顯著比例中為 0.2，熱門影片趨勢中顯示有 2 個創作者在拍攝日常主題中的影片較熱門。是否與網紅或專家合作裡較為顯著  $p<0.05$  中有 4 個創作者，不顯著中則是 6 個創作者，所以在頻道顯著比例中為 0.4，熱門

影片趨勢中顯示有 4 個創作者在與網紅或專家拍攝的影片較熱門。影片上傳星期裡較為顯著  $p<0.05$  中有 1 個創作者，不顯著中則是 9 個創作者，所以在頻道顯著比例中為 0.1，熱門影片趨勢中顯示有 1 個創作者在星期五上傳的影片較熱門。影片上傳早上、晚上裡較為顯著  $p<0.05$  中有 3 個創作者，不顯著中則是 7 個創作者，所以在頻道顯著比例中為 0.3，熱門影片趨勢中顯示有 3 個創作者在早上發布的影片較熱門。影片上傳是否超過 14 天裡較為顯著  $p<0.05$  中有 1 個創作者，不顯著中則是 9 個創作者，所以在頻道顯著比例中為 0.1，熱門影片趨勢中顯示有 1 個創作者在影片上傳間隔超過 14 天情況下影片還較為熱門。

從上觀察到在運動類別的影片時長中，較短時間的影片較容易成為熱門影片，而在影片的留言評論中，不管是正向或負向留言都是偏向兩邊的極性分佈，也因此可以認為運動頻道中的留言中越正向或負向的留言，都將使影片更趨向熱門，對於影片發布者來說，影片標題字詞數量越多越好，盡量讓標題內容豐富詳細，更有助於增加對於觀眾的吸引力。運動類別值中，可以觀察到在卡方檢定中，其顯著比例都不高，然而還是可以看出在寬鬆樣本資料中的影片類別是有相對顯著比例的，可以看到影片類別中主要是 17(體育)、20(遊戲)這兩類為主要熱門影片類別，並且與他人合作拍攝影片也較能夠更吸引觀眾的眼球。

表 -6 運動類別連續值流行度統計檢定結果(寬鬆樣本)

	* $p<0.05$	不顯著	頻道顯著比例	熱門影片趨勢
影片時長	8	2	0.8	5 短 3 長
情感特徵平均值	4	6	0.4	2 小 2 大
情感特徵最大值	9	1	0.9	9 正向
情感特徵最小值	7	3	0.7	7 負向
標籤數量	5	5	0.5	5 數量多
影片標題長度	5	5	0.5	5 個長
影片標題詞數	7	3	0.7	7 個多
影片上傳間隔	3	7	0.3	3 個長

表 -7 運動類別的類別值流行度統計檢定結果(寬鬆樣本)

	*p<0.05	不顯著	頻道顯著比例	熱門影片趨勢
影片類別	5	5	0.5	4 個類別 17 1 個類別 20
是否為延伸續集	3	7	0.3	1 個沒續集 2 個有續集
是否為實測主題	0	10	0	0 個實測
是否為日常主題	2	8	0.2	2 個日常主題
是否與網紅或專家 合作	4	6	0.4	4 個有合作
影片上傳星期	1	9	0.1	1 個星期五
影片上傳早上、晚 上	3	7	0.3	3 個都是早上
影片上傳是否超過 14 天	1	9	0.1	1 個影片上傳間隔 超過 14 天

根據表 8 與表 9 顯示，可以觀察到人物&部落客類別連續值流行度統計檢定結果(寬鬆樣本)中，影片時長裡，熱門影片趨勢中顯示在熱門與不熱門的比較中，有 2 個創作者發布的熱門影片為較短影音，5 個創作者發布的熱門影片為較長影音。情感特徵平均值裡，熱門影片趨勢中顯示在情感特徵平均值的比較中，有 2 個創作者的值較不熱門影片小，有 2 個創作者的值較不熱門影片大。情感特徵最大值裡，熱門影片趨勢中顯示在熱門影片創作者中有 5 個都是極為正向的。情感特徵最小值裡，熱門影片趨勢中顯示在熱門影片創作者中有 6 個都是極為負向的。標籤數量裡，熱門影片趨勢中顯示共有 3 個創作者的標籤數量是較多的。影片標題長度裡，熱門影片趨勢中顯示有 4 個創作者的標題長度是較長的，有 1 個創作者的標題長度是較短的。影片標題詞數中，熱門影片趨勢中顯示有 4 個創作者的標題詞數是較多的。影片上傳間隔裡，熱門影片趨勢中顯示有 1 個創作者發布的影片與前一部影片相比間隔天數較長。

人物&部落客類別的類別值流行度統計檢定結果(寬鬆樣本)中，影片類別裡，熱門影片趨勢中顯示在熱門與不熱門的比較中，有 1 個創作者的影片類別為 22(人物與網誌)。是否為延伸續集裡，熱門影片趨勢中顯示 1 個創作者在影片不是續集時影片較熱門，2 個創作者在影片有續集時影片較熱門。是否為實測主題裡，熱門影片趨勢中顯示為實測影片對於熱門與否較不顯著。是否為日常主題裡，熱門影片趨勢中顯示為日常影片對於熱門與否較不顯著。是否與網紅或專家合作裡，熱門影片趨勢中顯示有 2 個創作者在與網紅或專家拍攝的影片較熱門。影片上傳星期裡，熱門影片趨勢中顯示有 4 個創作者分別在星期一，星期五，星期日和星期三或六

上傳的影片較熱門。影片上傳早上、晚上裡，熱門影片趨勢中顯示有 1 個創作者在早上發布的影片較熱門，有 1 個創作者在晚上發布的影片較熱門。影片上傳是否超過 14 天裡，熱門影片趨勢中顯示影片上傳間隔超過 14 天情況下影片對於熱門與否較不顯著。

人物&部落客類別連續值中影片時長、情感特徵最大和最小值較為顯著，人物&部落客影片中若影片時長較長較容易成為熱門影片，而觀眾對於影片的留言情感極性，也影響著影片的熱門程度。人物&部落客類別類別值中，在卡方檢定裡其顯著比例都偏低，然而還是可以看出在寬鬆樣本資料中的影片上傳星期、是否與網紅或專家合作是有相對顯著比例的，可以看到影片類別中主要是 22(人物與網誌)，在發布影片的日子也很廣泛平日假日都有，而創作者要如何吸引觀眾，則會建議影片創作者能找人合作或朝續集方面去拍攝，並且建議創作者盡量不要拍攝與實測或日常相關的影片，以及影片發布的間隔不要超過 14 天。

表 -8 人物&部落客類別的連續值流行度統計檢定結果(寬鬆樣本)

	*p<0.05	不顯著	頻道顯著比例	熱門影片趨勢
影片時長	7	3	0.7	2 短 5 長
情感特徵平均值	4	6	0.4	2 小 2 大
情感特徵最大值	5	5	0.5	5 正向
情感特徵最小值	6	4	0.6	6 負向
標籤數量	3	7	0.3	3 數量多
影片標題長度	5	5	0.5	4 個長 1 個短
影片標題詞數	4	6	0.4	4 個多
影片上傳間隔	1	9	0.1	1 個長

表 -9 人物&部落客類別的類別值流行度統計檢定結果(寬鬆樣本)

	*p<0.05	不顯著	頻道顯著比例	熱門影片趨勢
影片類別	1	9	0.1	1 個類別 22
是否為延伸續集	3	7	0.3	1 個沒續集 2 個有續集
是否為實測主題	0	10	0	0 個實測
是否為日常主題	0	10	0	0 個日常主題
是否與網紅或專家合作	2	8	0.2	2 個有合作
影片上傳星期	4	6	0.4	4 個分別星期一， 日，五，三/六
影片上傳早上、晚上	2	8	0.2	1 個早上 1 個晚上
影片上傳是否超過 14 天	0	10	0	0 個影片上傳間隔 超過 14 天

根據表 10 與表 11 顯示，可以觀察到娛樂類別連續值流行度統計檢定結果(寬鬆樣本)中，影片時長裡，熱門影片趨勢中顯示在熱門與不熱門的比較中，有 1 個創作者發布的熱門影片為較短影音，3 個創作者發布的熱門影片為較長影音。情感特徵平均值裡，熱門影片趨勢中顯示在情感特徵平均值的比較中，有 2 個創作者的值較不熱門影片小，有 2 個創作者的值較不熱門影片大。情感特徵最大值裡，熱門影片趨勢中顯示在熱門影片創作者中有 5 個都是極為正向的。情感特徵最小值裡，熱門影片趨勢中顯示在熱門影片創作者中有 7 個都是極為負向的。標籤數量裡，熱門影片趨勢中顯示共有 3 個創作者的標籤數量是較多的，1 個創作者的標籤數量是較少的。影片標題長度裡，熱門影片趨勢中顯示有 3 個創作者的標題長度是較長的。影片標題詞數中，熱門影片趨勢中顯示有 4 個創作者的標題詞數是較多的。影片上傳間隔裡，熱門影片趨勢中顯示有 2 個創作者發布的影片與前一部影片相比間隔天數較長。

娛樂類別的類別值流行度統計檢定結果(寬鬆樣本)中，影片類別裡，熱門影片趨勢中顯示在熱門與不熱門的比較中，有 1 個創作者的影片類別為 24(娛樂)，1 個創作者的影片類別為 19(旅遊與活動)。是否為延伸續集裡，熱門影片趨勢中顯示延伸續集影片對於熱門與否較不顯著。是否為實測主題裡，熱門影片趨勢中顯示為實測影片對於熱門與否較不顯著。是否為日常主題裡，熱門影片趨勢中顯示為日常影片對於熱門與否較不顯著。是否與網紅或專家合作裡，熱門影片趨勢中顯示有 1 個創作者在與網紅或專家拍攝的影片較熱門。影片上傳星期裡，熱門影片趨勢中顯示有 3 個創作者在星期五時上傳的影片較熱門。影片上傳早上、晚上裡，熱門影片趨

勢中顯示有 2 個創作者在早上發布的影片較熱門。影片上傳是否超過 14 天裡，熱門影片趨勢中顯示影片上傳間隔超過 14 天情況下影片對於熱門與否較不顯著。

在娛樂類別連續值統計中，情感特徵最小值較為顯著。可以發覺在娛樂類別影片中，觀眾對於影片的留言情感極性，也影響著影片的熱門程度。娛樂類別類別值中，可以觀察到在卡方檢定中，其顯著比例都偏低，然而影片上傳星期、影片上傳早上、晚上是相對顯著比例的，可以看到影片類別中主要是 24(娛樂)、19(旅遊與活動)，而影片上傳星期、影片上傳早上、晚上這兩個變數在娛樂類別中，會建議創作者影片發布時間為星期五早上，並且建議創作者盡量不要拍攝與實測或日常和影片續集相關的影片，以及影片發布的間隔不要超過 14 天。

表 -10 娛樂類別的連續值流行度統計檢定結果(寬鬆樣本)

	*p<0.05	不顯著	頻道顯著比例	熱門影片趨勢
影片時長	4	6	0.4	1 短 3 長
情感特徵平均值	4	6	0.4	2 小 2 大
情感特徵最大值	5	5	0.5	5 正向
情感特徵最小值	7	3	0.7	7 負向
標籤數量	4	6	0.4	3 多;1 少
影片標題長度	3	7	0.3	3 個長
影片標題詞數	4	6	0.4	4 個多
影片上傳間隔	2	8	0.2	2 個長

表 -11 娛樂類別的類別值流行度統計檢定結果(寬鬆樣本)

	*p<0.05	不顯著	頻道顯著比例	熱門影片趨勢
影片類別	2	8	0.2	1 個類別 24 1 個類別 19
是否為延伸續集	0	10	0	0 個續集
是否為實測主題	0	10	0	0 個實測
是否為日常主題	0	10	0	0 個日常主題
是否與網紅或專家合作	1	9	0.1	1 個有合作
影片上傳星期	3	7	0.3	3 個星期五
影片上傳早上、晚上	2	8	0.2	2 個都是早上
影片上傳是否超過 14 天	0	10	0	0 個影片上傳間隔超過 14 天

#### 4.2.2 嚴謹樣本資料統計檢定結果分析

根據表 12 與表 13 顯示，可以觀察到運動類別連續值流行度統計檢定結果(嚴謹樣本)中，影片時長裡較為顯著  $p<0.05$  中有 8 個創作者，不顯著中則是 2 個創作者，所以在頻道顯著比例中為 0.8，熱門影片趨勢中顯示在熱門與不熱門的比較中，有 5 個創作者發布的熱門影片為較短影音，3 個創作者發布的熱門影片為較長影音。情感特徵平均值裡較為顯著  $p<0.05$  中有 2 個創作者，不顯著中則是 8 個創作者，所以在頻道顯著比例中為 0.2，熱門影片趨勢中顯示在情感特徵平均值的比較中，有 1 個創作者的值較不熱門影片小，有 1 個創作者的值較不熱門影片大。情感特徵最大值裡較為顯著  $p<0.05$  中有 9 個創作者，不顯著中則是 1 個創作者，所以在頻道顯著比例中為 0.9，熱門影片趨勢中顯示在熱門影片創作者中有 9 個都是極為正向的。情感特徵最小值裡較為顯著  $p<0.05$  中有 8 個創作者，不顯著中則是 2 個創作者，所以在頻道顯著比例中為 0.8，熱門影片趨勢中顯示在熱門影片創作者中有 8 個都是極為負向的。標籤數量裡較為顯著  $p<0.05$  中有 5 個創作者，不顯著中則是 5 個創作者，所以在頻道顯著比例中為 0.5，熱門影片趨勢中顯示共有 5 個創作者的標籤數量是較多的。影片標題長度裡較為顯著  $p<0.05$  中有 6 個創作者，不顯著中則是 4 個創作者，所以在頻道顯著比例中為 0.6，熱門影片趨勢中顯示有 6 個創作者的標題長度是較長的。影片標題詞數中較為顯著  $p<0.05$  中有 7 個創作者，不顯著中則是 3 個創作者，所以在頻道顯著比例中為 0.7，熱門影片趨勢中顯示有 7 個創作者的標題詞數是較多的。影片上傳間隔裡較為顯著  $p<0.05$  中有 3 個創作者，不顯著中則是 7 個創作者，所以在頻道顯著比例中為 0.3，熱門影片趨勢中顯示有 3 個創作者發布的影片與前一部影片相比間隔天數較長。

運動類別的類別值流行度統計檢定結果(寬鬆樣本)中，影片類別裡較為顯著  $p<0.05$  中有 4 個創作者，不顯著中則是 6 個創作者，所以在頻道顯著比例中為 0.4，熱門影片趨勢中顯示在熱門與不熱門的比較中，有 3 個創作者的影片類別為 17(體育)，1 個創作者的影片類別為 20(遊戲)。是否為延伸續集裡較為顯著  $p<0.05$  中有 4 個創作者，不顯著中則是 6 個創作者，所以在頻道顯著比例中為 0.4，熱門影片趨勢中顯示 1 個創作者在影片不是續集時影片較熱門，3 個創作者在影片有續集時影片較熱門。是否為實測主題裡較為顯著  $p<0.05$  中有 0 個創作者，不顯著中則是 10 個創作者，所以在頻道顯著比例中為 0，熱門影片趨勢中顯示為實測影片對於熱門與否較不顯著。是否為日常主題裡較為顯著  $p<0.05$  中有 3 個創作者，不顯著中則是 7 個創作者，所以在頻道顯著比例中為 0.3，熱門影片趨勢中顯示有 3 個創作者在拍攝日常主題中的影片較熱門。是否與網紅或專家合作裡較為顯著  $p<0.05$  中有 4 個創作者，不顯著中則是 6 個創作者，所以在頻道顯著比例中為 0.4，熱門影片趨勢中顯示有 4 個創作者在與網紅或專家拍攝的影片較熱門。影片上傳星期裡較為顯著  $p<0.05$  中有 2 個創作者，不顯著中則是 8 個創作者，所以在頻道顯著比例中為 0.2，熱門影片趨勢中顯示有 1 個創作者在星期五上傳的影片較熱門，1

個創作者在星期二或六上傳的影片較熱門。影片上傳早上、晚上裡較為顯著  $p < 0.05$  中有 3 個創作者，不顯著中則是 7 個創作者，所以在頻道顯著比例中為 0.3，熱門影片趨勢中顯示有 3 個創作者在早上發布的影片較熱門。影片上傳是否超過 14 天裡較為顯著  $p < 0.05$  中有 0 個創作者，不顯著中則是 10 個創作者，所以在頻道顯著比例中為 0，熱門影片趨勢中顯示創作者在影片上傳間隔超過 14 天情況下影片對於熱門與否較不顯著。

從上觀察到在運動類別的影片時長中，較短時間的影片較容易成為熱門影片，而在影片的留言評論中，不管是正向或負向留言都是偏向兩邊的極性分佈，也因此可以認為運動頻道中的留言中越正向或負向的留言，都將使影片更趨向熱門，對於影片發布者來說，影片標題字詞數量越多越好，盡量讓標題內容豐富詳細，更有助於增加對於觀眾的吸引力。運動類別值中，可以觀察到在卡方檢定中，其顯著比例都不高，然而還是可以看出在嚴謹樣本資料中的影片類別是有相對顯著比例的，可以看到影片類別中主要是 17(體育)、20(遊戲)這兩類為熱門影片類別，也會建議影片發布者盡量以續集形式呈現，能拍攝些日常的影片或與他人合作拍攝影片都將能夠更吸引觀眾的眼球，而其中特別的是影片發布時間建議為星期五早上發布。

表 -12 運動類別的連續值流行度統計檢定結果(嚴謹樣本)

	* $p < 0.05$	不顯著	頻道顯著比例	熱門影片趨勢
影片時長	8	2	0.8	5 短 3 長
情感特徵平均值	2	8	0.2	1 大 1 小
情感特徵最大值	9	1	0.9	9 正向
情感特徵最小值	8	2	0.8	8 負向
標籤數量	5	5	0.5	5 數量多
影片標題長度	6	4	0.6	6 個長
影片標題詞數	7	3	0.7	7 個多
影片上傳間隔	3	7	0.3	3 個長



表 -13 運動類別的類別值流行度統計檢定結果(嚴謹樣本)

	*p<0.05	不顯著	頻道顯著比例	熱門影片趨勢
影片類別	4	6	0.4	3 個類別 17 1 個類別 20
是否為延伸續集	4	6	0.4	1 個沒續集 3 個有續集
是否為實測主題	0	10	0	0 個實測
是否為日常主題	3	7	0.3	3 個日常主題
是否與網紅或專家合作	4	6	0.4	4 個有合作
影片上傳星期	2	8	0.2	2 個分別星期二/ 六和五
影片上傳早上、晚上	3	7	0.3	3 個都是早上
影片上傳是否超過 14 天	0	10	0	0 個影片上傳間 隔超過 14 天

根據表 14 與表 15 顯示，可以觀察到人物&部落客類別連續值流行度統計檢定結果(嚴謹樣本)中，影片時長裡，熱門影片趨勢中顯示在熱門與不熱門的比較中，有 2 個創作者發布的熱門影片為較短影音，4 個創作者發布的熱門影片為較長影音。情感特徵平均值裡，熱門影片趨勢中顯示在情感特徵平均值的比較中，有 1 個創作者的值較不熱門影片小，有 2 個創作者的值較不熱門影片大。情感特徵最大值裡，熱門影片趨勢中顯示在熱門影片創作者中有 8 個都是極為正向的。情感特徵最小值裡，熱門影片趨勢中顯示在熱門影片創作者中有 7 個都是極為負向的。標籤數量裡，熱門影片趨勢中顯示共有 4 個創作者的標籤數量是較多的。影片標題長度裡，熱門影片趨勢中顯示有 4 個創作者的標題長度是較長的，有 1 個創作者的標題長度是較短的。影片標題詞數中，熱門影片趨勢中顯示有 3 個創作者的標題詞數是較多的。影片上傳間隔裡，熱門影片趨勢中顯示有 2 個創作者發布的影片與前一部影片相比間隔天數較長。

人物&部落客類別的類別值流行度統計檢定結果(嚴謹樣本)中，影片類別裡，熱門影片趨勢中顯示在熱門與不熱門的比較中，影片類別對於熱門與否較不顯著。是否為延伸續集裡，熱門影片趨勢中顯示 1 個創作者在影片不是續集時影片較熱門，1 個創作者在影片有續集時影片較熱門。是否為實測主題裡，熱門影片趨勢中顯示為實測影片對於熱門與否較不顯著。是否為日常主題裡，熱門影片趨勢中顯示為日常影片對於熱門與否較不顯著。是否與網紅或專家合作裡，熱門影片趨勢中顯示有 1 個創作者在與網紅或專家拍攝的影片較熱門。影片上傳星期裡，熱門影片

趨勢中顯示有 4 個創作者分別在星期一，星期五，星期日和星期三或六上傳的影片較熱門。影片上傳早上、晚上裡，熱門影片趨勢中顯示有 1 個創作者在早上發布的影片較熱門。影片上傳是否超過 14 天裡，熱門影片趨勢中顯示影片上傳間隔超過 14 天情況下影片對於熱門與否較不顯著。

人物&部落客類別連續值中影片時長、情感特徵最大和最小值較為顯著，人物&部落客影片中若影片時長較長較容易成為熱門影片，而觀眾對於影片的留言情感極性，也影響著影片的熱門程度，而從中會發覺比較特別的是，少數在人物&部落客類別中的創作者影片上傳的間隔時間增長，也會增加影片的吸引力。人物&部落客類別類別值中，在卡方檢定裡其顯著比例都偏低，然而還是可以看出在嚴謹樣本資料中的影片上傳星期是相對顯著比例的，發布日期遍佈在一、三、五、六、日中，表示人物&部落客類別創作者的影片發布日中並沒有明顯因為假日發布而比較多人觀看，並建議創作者盡量不要拍攝與實測或日常相關的影片，以及影片發布的間隔不要超過 14 天。

表 -14 人物&部落客類別的連續值流行度統計檢定結果(嚴謹樣本)

	*p<0.05	不顯著	頻道顯著比例	熱門影片趨勢
影片時長	6	4	0.6	2 短 4 長
情感特徵平均值	3	7	0.3	1 小 2 大
情感特徵最大值	8	2	0.8	8 正向
情感特徵最小值	7	3	0.7	7 負向
標籤數量	4	6	0.4	4 數量多
影片標題長度	5	5	0.5	4 長 1 短
影片標題詞數	3	7	0.3	3 個多
影片上傳間隔	2	8	0.2	2 個長

表 -15 人物&部落客類別的類別值流行度統計檢定結果(嚴謹樣本)

	*p<0.05	不顯著	頻道顯著比例	熱門影片趨勢
影片類別	0	10	0	0 個類別
是否為延伸續集	2	8	0.2	1 個沒續集 1 個有續集
是否為實測主題	0	10	0	0 個實測
是否為日常主題	0	10	0	0 個日常主題
是否與網紅或專家合作	1	9	0.1	1 個有合作好
影片上傳星期	4	6	0.4	4 個分別星期一，日，五，三/六
影片上傳早上、晚上	1	9	0.1	1 個早上
影片上傳是否超過 14 天	0	10	0	0 個影片上傳間隔超過 14 天

根據表 16 與表 17 顯示，可以觀察到娛樂類別連續值流行度統計檢定結果(嚴謹樣本)中，影片時長裡，熱門影片趨勢中顯示在熱門與不熱門的比較中，有 1 個創作者發布的熱門影片為較短影音，4 個創作者發布的熱門影片為較長影音。情感特徵平均值裡，熱門影片趨勢中顯示在情感特徵平均值的比較中，有 2 個創作者的值較不熱門影片小，有 2 個創作者的值較不熱門影片大。情感特徵最大值裡，熱門影片趨勢中顯示在熱門影片創作者中有 6 個都是極為正向的。情感特徵最小值裡，熱門影片趨勢中顯示在熱門影片創作者中有 7 個都是極為負向的。標籤數量裡，熱門影片趨勢中顯示共有 3 個創作者的標籤數量是較多的，1 個創作者的標籤數量是較少的。影片標題長度裡，熱門影片趨勢中顯示有 3 個創作者的標題長度是較長的。影片標題詞數中，熱門影片趨勢中顯示有 3 個創作者的標題詞數是較多的，有 1 個創作者的標題詞數是較少的。影片上傳間隔裡，熱門影片趨勢中顯示有 3 個創作者發布的影片與前一部影片相比間隔天數較長。

娛樂類別的類別值流行度統計檢定結果(嚴謹樣本)中，影片類別裡，熱門影片趨勢中顯示在熱門與不熱門的比較中，有 1 個創作者的影片類別為 24(娛樂)，1 個創作者的影片類別為 19(旅遊與活動)。是否為延伸續集裡，熱門影片趨勢中顯示延伸續集影片對於熱門與否較不顯著。是否為實測主題裡，熱門影片趨勢中顯示為實測影片對於熱門與否較不顯著。是否為日常主題裡，熱門影片趨勢中顯示為日常影片對於熱門與否較不顯著。是否與網紅或專家合作裡，熱門影片趨勢中顯示有 2 個

創作者在與網紅或專家拍攝的影片較熱門。影片上傳星期裡，熱門影片趨勢中顯示有 3 個創作者在星期五時上傳的影片較熱門。影片上傳早上、晚上裡，熱門影片趨勢中顯示有 2 個創作者在早上發布的影片較熱門，有 1 個創作者在晚上發布的影片較熱門。影片上傳是否超過 14 天裡，熱門影片趨勢中顯示影片上傳間隔超過 14 天情況下影片對於熱門與否較不顯著。

在娛樂類別連續值統計中，情感特徵最大值和情感特徵最小值較為顯著。可以發覺在娛樂類別影片中，觀眾對於影片的留言情感極性，也影響著影片的熱門程度，娛樂類別的影片時長中，較長時間的影片較容易成為熱門影片，而從表中會發覺比較特別的是，少數在娛樂類別中作者影片上傳的間隔時間增長，也會增加影片的吸引力。

娛樂類別類別值中，可以觀察到在卡方檢定中，其顯著比例都偏低，然而影片上傳星期、影片上傳早上、晚上和是否與網紅或專家合作是相對顯著比例的，可以看到影片類別中主要是 24(娛樂)、19(旅遊與活動)，而是否與網紅或專家合作、影片上傳早上、晚上這兩個變數在娛樂類別中，會建議創作者能與他人合作外，也建議創作者影片發布時間為星期五早上，並且建議創作者盡量不要拍攝與實測或日常和影片續集相關的影片，以及影片發布的間隔不要超過 14 天。

表 -16 娛樂類別的連續值流行度統計檢定結果(嚴謹樣本)

	*p<0.05	不顯著	頻道顯著比例	熱門影片趨勢
影片時長	5	5	0.5	1 短 4 長
情感特徵平均值	4	6	0.4	2 小 2 大
情感特徵最大值	6	4	0.6	6 正向
情感特徵最小值	7	3	0.7	7 負向
標籤數量	4	6	0.4	3 多;1 少
影片標題長度	3	7	0.3	3 個長
影片標題詞數	4	6	0.4	3 多;1 少
影片上傳間隔	3	7	0.3	3 個長

表 -17 娛樂類別的類別值流行度統計檢定結果(嚴謹樣本)

	*p<0.05	不顯著	頻道顯著比例	熱門影片趨勢
影片類別	2	8	0.2	1 個類別 24 1 個類別 19
是否為延伸續集	0	10	0	0 個續集
是否為實測主題	0	10	0	0 個實測
是否為日常主題	0	10	0	0 個日常主題
是否與網紅或專家合作	2	8	0.2	2 個有合作
影片上傳星期	3	7	0.3	三個星期五
影片上傳早上、晚上	3	7	0.3	2 個早上 1 個晚上
影片上傳是否超過 14 天	0	10	0	0 個影片上傳 間隔超過 14 天

### 4.3 實驗與結果

本節將針對 YouTube 資料樣本經過監督式學習後的成果進行實驗結果的說明，並且本研究將資料中的影片元數據、影片標題、影片描述、留言評論這四大資料指標去進行特徵組合，並產生十種特徵組合，如表 18，並且將三大類別（運動、人物&部落客、娛樂）中的各十位創作者，都將依照這十種特徵組合去分別進行單一分類器和集成分類器的訓練，最後會將同一類別中的十位創作者的模型評估指標分數去做平均值的統計計算，最後以同一類別十種特徵組合的平均分數來做呈現，並且將訓練結果依照本研究挑選的評估指標去進行評估研究。

而每個類別中同一種特徵在不同模型訓練結果中，該評估指標中表現最好的會以黑色粗體標記，方便進行比較。

表 -18 特徵組合

特徵組合內容	特徵組合名稱
元數據	meta
標題	Title
描述	description
評論	comment
元數據 + 標題	meta+ title
元數據 + 描述	meta+ description
元數據 + 評論	meta+ comment
元數據 + 標題 + 評論	meta+ title+ comment
元數據 + 標題 + 描述	meta+ title+ description
元數據 + 標題 + 描述 + 評論	meta+ title+ description+ comment

#### 4.3.1 主題模型結果

本研究將各類別中的創作者影片的標題、描述以及觀眾的留言這三個文本特徵組合去進行主題模型分析，首先我們會將文本進行文本預處理，包含使用了CKIP斷詞和移除停用詞（cnsstopwords），接著使用 TF-IDF 向量器將文檔術語矩陣（DTM）轉換為數值陣列，並且使用 LDA 主題模型自動去計算找出最佳的主題數量後，再使用最佳主題數來訓練 LDA 模型，使每一個主題中的單詞都賦予權重，而這些權重也暗示了哪些詞在主題中佔主導地位。接著本研究會將這些生成的主題去分配給原本的文檔，使每個文檔會有個主題的比例類別。如下表 19，我們使用館長的影片描述(寬鬆資料集)生成的主題模型為例，總共生成 6 個主題，接著將主題依序分配給每部影片的影片描述中，計算每個主題的權重，從下表可以看出 Document 0 中 Topic 2 的權重為最高，為了篇幅關係因此取其中片段展示。然而因為在影片描述中，同個頻道主在影片描述中的文字描述可能是一系列的相差不多，而本研究使用自動生成主題模型方法，因此在呈現的主題中可能會有相似的主題呈現。

表 -19 LDA 主題模型文檔表

Document	Topic_0	Topic_1	Topic_2	Topic_3	Topic_4	Topic_5
0	0.04023	0.04023	0.79882	0.04026	0.04023	0.04023
1	0.03998	0.03998	0.27781	0.56226	0.03998	0.03998
2	0.04023	0.04023	0.79882	0.04026	0.04023	0.04023
3	0.04023	0.04023	0.79882	0.04026	0.04023	0.04023
4	0.03803	0.03843	0.03896	0.80852	0.03803	0.03803
5	0.03755	0.03724	0.04316	0.80757	0.03724	0.03724

#### 4.3.2 寬鬆樣本資料實驗與結果

在觀察表 20 中可以看出在單一分類器中，三種分類器中比較起來邏輯回歸的效果呈現是相對最佳的，在單一特徵中元數據與單一文本特徵(標題、描述、評論)相比元數據中的準確率、精確率、召回率、F1、AUC 效果最好。在集成分類器中，兩種分類器比較起來隨機森林表現較好，然而兩種分類器表現差異不大，在單一特徵元數據時，隨機森林效果較佳，但當是單一文本特徵時，梯度提升機的效果較佳，然而當元數據與單一文本特徵組合做結合時，可以看出在集成分類器上，則是隨機森林的效果呈現上為較佳的分類器。

從表中可以觀察出，全數模型的評估指標中 AUC 有最高的分數。在單一特徵中，元數據特徵與標題、描述、評論相比，元數據效果最好；三種文本特徵相比則是評論特徵效果最好。組合特徵中，元數據特徵依序加入三種不同的單一文本特徵去做特徵組合時，可以發覺當文本特徵與元數據進行組合時，都比原來單一文本特徵效果好，而當全部特徵(元數據、評論、標題、評論)與元數據特徵相比，效果並沒有什麼不同，這也表明了運動類別中，元數據特徵，本身就較為影響影片的熱門程度。

通過這些觀察，我們可以得出結論：元數據在預測運動類別影片的熱門程度上是重要且有用的特徵，而集成分類器中的隨機森林在結合多種特徵時表現最好。

表 -20 運動類別特徵組合

Model	元數據				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.72</b>	<b>0.72</b>	<b>0.72</b>	<b>0.70</b>	<b>0.79</b>
決策樹	0.68	0.66	0.67	0.65	0.67
支持向量機	0.60	0.60	0.70	0.60	0.70
隨機森林	<b>0.75</b>	<b>0.74</b>	<b>0.78</b>	<b>0.74</b>	<b>0.82</b>
梯度提升機	0.74	0.73	0.76	0.73	0.80

Model	描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.63	0.62	0.64	0.62	0.67
決策樹	0.62	0.62	0.63	0.62	0.62
支持向量機	0.63	0.64	0.62	0.62	0.67
隨機森林	0.64	0.65	0.64	0.63	0.70
梯度提升機	<b>0.66</b>	<b>0.67</b>	<b>0.67</b>	<b>0.66</b>	<b>0.71</b>

Model	標題				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.63	0.63	0.64	0.62	0.65
決策樹	0.63	0.64	0.63	0.62	0.64
支持向量機	<b>0.63</b>	<b>0.65</b>	<b>0.62</b>	<b>0.62</b>	<b>0.67</b>
隨機森林	0.64	0.64	0.65	0.64	0.70
梯度提升機	<b>0.65</b>	<b>0.68</b>	<b>0.64</b>	<b>0.64</b>	<b>0.71</b>

Model	評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.64</b>	<b>0.63</b>	<b>0.68</b>	<b>0.64</b>	<b>0.69</b>
決策樹	0.65	0.66	0.67	0.65	0.66
支持向量機	0.63	0.64	0.62	0.62	0.67
隨機森林	0.66	0.67	0.66	0.65	0.71
梯度提升機	<b>0.69</b>	<b>0.71</b>	<b>0.68</b>	<b>0.68</b>	<b>0.74</b>

Model	元數據 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.71</b>	<b>0.72</b>	<b>0.71</b>	<b>0.70</b>	<b>0.78</b>



決策樹	0.70	0.69	0.72	0.69	0.71
支持向量機	0.60	0.60	0.70	0.60	0.70
隨機森林	0.73	0.73	0.77	0.73	0.81
梯度提升機	<b>0.74</b>	<b>0.74</b>	<b>0.78</b>	<b>0.74</b>	<b>0.81</b>

Model	元數據 + 描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.71</b>	<b>0.72</b>	<b>0.71</b>	<b>0.69</b>	<b>0.79</b>
決策樹	0.69	0.69	0.70	0.68	0.69
支持向量機	0.60	0.60	0.70	0.60	0.70
隨機森林	<b>0.73</b>	<b>0.74</b>	<b>0.76</b>	<b>0.72</b>	<b>0.81</b>
梯度提升機	0.73	0.73	0.76	0.73	0.81

Model	元數據 + 標題				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.73</b>	<b>0.74</b>	<b>0.73</b>	<b>0.71</b>	<b>0.81</b>
決策樹	0.72	0.70	0.70	0.69	0.71
支持向量機	0.61	0.61	0.78	0.63	0.71
隨機森林	<b>0.77</b>	<b>0.76</b>	<b>0.80</b>	<b>0.76</b>	<b>0.83</b>
梯度提升機	0.76	0.75	0.78	0.75	0.82

Model	元數據 + 標題 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.72</b>	<b>0.73</b>	<b>0.72</b>	<b>0.70</b>	<b>0.79</b>
決策樹	0.70	0.68	0.70	0.67	0.70
支持向量機	0.60	0.60	0.70	0.60	0.70
隨機森林	0.73	0.72	0.76	0.73	0.81
梯度提升機	0.73	0.72	0.76	0.73	0.81

Model	元數據 + 標題 + 描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.72</b>	<b>0.72</b>	<b>0.71</b>	<b>0.70</b>	<b>0.79</b>
決策樹	0.68	0.67	0.67	0.65	0.68
支持向量機	0.60	0.60	0.70	0.60	0.70
隨機森林	<b>0.72</b>	<b>0.74</b>	<b>0.75</b>	<b>0.72</b>	<b>0.81</b>
梯度提升機	0.73	0.73	0.75	0.72	0.80

Model	元數據 + 標題 + 描述 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.72</b>	<b>0.73</b>	<b>0.72</b>	<b>0.70</b>	<b>0.80</b>
決策樹	0.70	0.69	0.69	0.68	0.70
支持向量機	0.60	0.60	0.70	0.60	0.70
隨機森林	<b>0.74</b>	<b>0.74</b>	<b>0.78</b>	<b>0.74</b>	<b>0.82</b>
梯度提升機	0.73	0.73	0.75	0.72	0.81

在觀察表 21 中可以看出在單一分類器中，三種分類器中比較起來邏輯回歸的效果呈現是相對最佳的，在單一特徵中元數據與單一文本特徵(標題、描述、評論)相比元數據中的準確率、精確率、F1、AUC 效果最好，而召回率方面則是支持向量機表現最佳，達到了 0.77，優於邏輯回歸的召回率。在集成分類器中，兩種分類器比較起來梯度提升機表現較好，單一特徵中元數據在隨機森林中的精確率、召回率和 F1 分數表現較好，但當單一文本特徵或將其與元數據特徵進行組合時，梯度提升機在準確率、精確率、召回率、F1 分數和 AUC 方面都表現較好。

從表中可以觀察出，全部混合模型的評估指標中召回率有最高的分數。在單一特徵中，元數據特徵與標題、描述、評論相比，元數據效果最好；三種文本特徵相比則是描述特徵效果最好。組合特徵中，元數據特徵依序加入三種不同的單一文本特徵去做特徵組合時，可以發覺組合特徵都比原本的單一文本特徵時效果更好，而當全部特徵(元數據、評論、標題、評論)與元數據特徵相比，效果也變好獲得提升，這也表明了在一人物&部落客類別中，結合元數據和文本特徵可以有效性去提升模型的整體效果，並且當文本特徵在與元數據特徵進行結合使用時能影響影片的熱門程度。

通過這些觀察，我們可以得出結論：在一人物&部落客類別中，結合元數據和文本特徵可以提升模型的效果，文本特徵與元數據特徵結合使用時能更好地預測影片的熱門程度，這也意謂著除了元數據外，文本特徵也對熱門影片程度上具有重要作用。

表 -21 人物&部落客類別特徵組合

Model	元數據				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.64</b>	<b>0.65</b>	0.67	<b>0.63</b>	<b>0.71</b>
決策樹	0.63	0.64	0.66	0.62	0.64
支持向量機	0.62	0.60	<b>0.77</b>	0.63	0.71
隨機森林	0.67	<b>0.69</b>	<b>0.72</b>	<b>0.68</b>	<b>0.72</b>
梯度提升機	0.67	0.67	0.70	0.65	0.73

Model	描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.59</b>	<b>0.59</b>	<b>0.71</b>	<b>0.63</b>	<b>0.62</b>
決策樹	0.56	0.59	0.58	0.56	0.56
支持向量機	0.52	0.50	0.57	0.51	0.53
隨機森林	<b>0.59</b>	<b>0.61</b>	<b>0.63</b>	<b>0.60</b>	<b>0.61</b>
梯度提升機	0.58	0.59	0.61	0.59	0.61

Model	標題				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.60</b>	<b>0.59</b>	<b>0.66</b>	<b>0.60</b>	<b>0.60</b>
決策樹	0.58	0.60	0.59	0.58	0.58
支持向量機	0.52	0.50	0.56	0.50	0.53
隨機森林	0.56	0.58	0.59	0.57	0.59
梯度提升機	<b>0.59</b>	<b>0.61</b>	<b>0.61</b>	<b>0.59</b>	<b>0.61</b>

Model	評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.53</b>	<b>0.50</b>	<b>0.62</b>	<b>0.54</b>	<b>0.57</b>
決策樹	0.59	0.61	0.59	0.58	0.59
支持向量機	0.52	0.50	0.56	0.50	0.53
隨機森林	0.58	0.60	0.60	0.58	0.61
梯度提升機	<b>0.60</b>	<b>0.61</b>	<b>0.60</b>	<b>0.59</b>	<b>0.62</b>

Model	元數據 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.64	0.65	0.67	0.63	0.71
決策樹	0.66	0.66	0.68	0.64	0.66
支持向量機	0.62	0.60	<b>0.77</b>	0.63	0.71
隨機森林	0.68	0.70	0.68	0.66	0.72
梯度提升機	<b>0.70</b>	<b>0.70</b>	<b>0.75</b>	<b>0.69</b>	<b>0.74</b>

Model	元數據 + 描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.64	0.65	0.68	0.63	0.72
決策樹	0.66	0.66	0.66	0.64	0.65
支持向量機	0.62	0.60	<b>0.77</b>	0.63	0.71

隨機森林	0.66	0.67	0.74	0.67	0.71
梯度提升機	<b>0.67</b>	<b>0.68</b>	<b>0.72</b>	<b>0.67</b>	<b>0.72</b>

Model	元數據 + 標題				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.65	0.66	0.70	0.65	0.72
決策樹	0.63	0.64	0.65	0.61	0.63
支持向量機	0.62	0.60	<b>0.77</b>	0.63	0.71
隨機森林	0.65	0.68	0.70	0.65	0.71
梯度提升機	<b>0.67</b>	<b>0.66</b>	<b>0.70</b>	<b>0.65</b>	<b>0.72</b>

Model	元數據 + 標題 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.66	0.66	0.70	0.65	0.72
決策樹	0.64	0.66	0.67	0.63	0.65
支持向量機	0.62	0.60	<b>0.77</b>	0.63	0.65
隨機森林	0.67	0.68	0.69	0.65	0.72
梯度提升機	<b>0.69</b>	<b>0.70</b>	<b>0.72</b>	<b>0.67</b>	<b>0.76</b>

Model	元數據 + 標題 + 描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.66	0.67	0.71	0.65	0.72
決策樹	0.64	0.66	0.66	0.63	0.65
支持向量機	0.62	0.60	<b>0.77</b>	0.63	0.71
隨機森林	<b>0.66</b>	<b>0.66</b>	<b>0.72</b>	<b>0.66</b>	<b>0.72</b>
梯度提升機	0.67	0.67	0.69	0.65	0.73

Model	元數據 + 標題 + 描述 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.65	0.67	0.70	0.65	0.73
決策樹	0.64	0.67	0.66	0.63	0.65
支持向量機	0.62	0.60	<b>0.77</b>	0.63	0.71
隨機森林	0.68	0.69	0.72	0.67	0.74
梯度提升機	<b>0.69</b>	<b>0.70</b>	<b>0.72</b>	<b>0.68</b>	<b>0.75</b>

在觀察表 22 中可以看出在單一分類器中，三種分類器中比較起來邏輯回歸的效果呈現是相對最佳的，在單一特徵中元數據中顯示邏輯回歸在準確率和 AUC 中效果較佳，決策樹在精確率和 F1 效果較好，支持向量機則是召回率表現最好，而在單一文本特徵以及與元數據特徵進行特徵組合後，整體上邏輯回歸在準確率、精確率、召回率、F1、AUC 都是相對有較好的表現。在集成分類器中，兩種分類器比較起來隨機森林表現較好，然而單一文字特徵組合(標題、評論)時梯度提升機效果有些許的提升略優於隨機森林，但當文字特徵與元數據特徵進行組合時，可以看出在集成分類器上，則還是隨機森林的效果呈現上為較佳的分類器。

從表中可以觀察出，全數模型的評估指標中 AUC 有最高的分數。在單一特徵中，元數據特徵與標題、描述、評論相比，元數據效果最好；三種文本特徵相比則是標題特徵效果最好。組合特徵中，元數據特徵依序加入三種不同的單一文本特徵去做特徵組合時，與元數據特徵相比，除了準確率與精確率有些為提升，其餘評估指標並沒有明顯提升，這也表明了娛樂類別中，元數據特徵本身就較為影響影片的熱門程度。

通過這些觀察，我們可以得出結論：在娛樂類別中，元數據特徵在預測影片熱門程度上具有重要作用，雖然文本特徵（特別是標題）在某些情況下能夠提升模型效果，但整體來說元數據特徵的影響更大。

表 -22 娛樂類別特徵組合

Model	元數據				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.65</b>	0.66	0.66	0.63	<b>0.70</b>
決策樹	0.65	<b>0.67</b>	0.67	<b>0.64</b>	0.66
支持向量機	0.59	0.60	<b>0.71</b>	0.60	0.66
隨機森林	<b>0.66</b>	<b>0.67</b>	<b>0.70</b>	<b>0.66</b>	<b>0.75</b>
梯度提升機	0.66	0.68	0.68	0.65	0.71

Model	描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.63</b>	<b>0.64</b>	<b>0.68</b>	<b>0.64</b>	<b>0.67</b>
決策樹	0.62	0.66	0.60	0.60	0.62
支持向量機	0.63	0.63	0.66	0.62	0.62
隨機森林	<b>0.62</b>	<b>0.65</b>	<b>0.63</b>	<b>0.61</b>	<b>0.66</b>
梯度提升機	0.62	0.64	0.63	0.61	0.65

Model	標題				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.64</b>	<b>0.65</b>	<b>0.67</b>	<b>0.65</b>	<b>0.68</b>

決策樹	0.63	0.65	0.61	0.61	0.63
支持向量機	0.64	0.64	0.67	0.63	0.63
隨機森林	0.63	0.65	0.63	0.62	0.68
梯度提升機	<b>0.63</b>	<b>0.65</b>	<b>0.65</b>	<b>0.63</b>	<b>0.70</b>

Model	評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.63</b>	<b>0.64</b>	<b>0.68</b>	<b>0.64</b>	<b>0.65</b>
決策樹	0.59	0.63	0.57	0.58	0.60
支持向量機	0.63	0.63	0.67	0.63	0.63
隨機森林	0.59	0.60	0.58	0.58	0.64
梯度提升機	<b>0.62</b>	<b>0.64</b>	<b>0.62</b>	<b>0.61</b>	<b>0.65</b>

Model	元數據 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.66</b>	<b>0.67</b>	<b>0.67</b>	<b>0.64</b>	<b>0.70</b>
決策樹	0.63	0.64	0.61	0.61	0.63
支持向量機	0.59	0.60	0.71	0.60	0.66
隨機森林	0.63	0.63	0.63	0.61	0.68
梯度提升機	<b>0.65</b>	<b>0.66</b>	<b>0.65</b>	<b>0.63</b>	<b>0.71</b>

Model	元數據 + 描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.65</b>	<b>0.65</b>	<b>0.65</b>	<b>0.62</b>	<b>0.71</b>
決策樹	0.61	0.63	0.63	0.60	0.62
支持向量機	0.59	0.60	0.71	0.60	0.66
隨機森林	<b>0.66</b>	<b>0.68</b>	<b>0.66</b>	<b>0.64</b>	<b>0.71</b>
梯度提升機	0.65	0.66	0.66	0.64	0.69

Model	元數據 + 標題				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.67</b>	<b>0.69</b>	<b>0.66</b>	<b>0.64</b>	<b>0.70</b>
決策樹	0.65	0.65	0.65	0.63	0.65
支持向量機	0.59	0.60	0.71	0.60	0.66
隨機森林	0.66	0.67	0.67	0.64	0.75
梯度提升機	<b>0.67</b>	<b>0.66</b>	<b>0.69</b>	<b>0.66</b>	<b>0.73</b>

Model	元數據 + 標題 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.66</b>	<b>0.68</b>	<b>0.66</b>	<b>0.64</b>	<b>0.71</b>
決策樹	0.63	0.63	0.62	0.60	0.62
支持向量機	0.59	0.60	0.71	0.60	0.66
隨機森林	<b>0.66</b>	<b>0.66</b>	<b>0.66</b>	<b>0.64</b>	<b>0.71</b>
梯度提升機	0.64	0.64	0.66	0.63	0.70

Model	元數據 + 標題 + 描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.67</b>	<b>0.69</b>	<b>0.67</b>	<b>0.65</b>	<b>0.70</b>
決策樹	0.61	0.62	0.62	0.59	0.61
支持向量機	0.59	0.60	0.71	0.60	0.66
隨機森林	<b>0.68</b>	<b>0.68</b>	<b>0.71</b>	<b>0.67</b>	<b>0.73</b>
梯度提升機	0.64	0.65	0.66	0.63	0.71

Model	元數據 + 標題 + 描述 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.67</b>	<b>0.68</b>	<b>0.66</b>	<b>0.64</b>	<b>0.70</b>
決策樹	0.62	0.62	0.62	0.60	0.62
支持向量機	0.59	0.60	0.71	0.60	0.66
隨機森林	<b>0.66</b>	<b>0.66</b>	<b>0.67</b>	<b>0.65</b>	<b>0.71</b>
梯度提升機	0.64	0.64	0.65	0.62	0.70

#### 4.3.3 嚴謹樣本資料實驗與結果

在觀察表 23 中可以看出在單一分類器中，三種分類器中比較起來邏輯回歸的效果呈現是相對最佳的，在單一特徵中元數據中邏輯回歸在 AUC 成績上表現最佳，而在單一文本特徵中決策樹在準確率、精確率、召回率、F1 展現出最好的效果，然而當文本特徵組合(標題、描述、評論)與元數據進行結合時，會發現在邏輯回歸中的 AUC 值為 0.89 相較於決策樹更高。在集成分類器中，兩種分類器比較起來隨機森林表現較好，然而兩種分類器表現差異不大，儘管在單一文本特徵組合時在集成分類器上梯度提升機的效果較好，但當文本特徵與元數據做結合時，可以看出在集成分類器上，則還是隨機森林的效果呈現上為較佳的分類器。

從表中可以觀察出，全數模型的評估指標中 AUC 有最高的分數。在單一特徵中，元數據特徵與標題、描述、評論相比，元數據效果最好；三種文本特徵相比則

是標題特徵效果最好。組合特徵中，元數據特徵依序加入三種不同的單一文本特徵去做特徵組合時，與文本特徵相比可以發覺效果都有明顯提升，而當全部特徵(元數據、評論、標題、評論)與元數據特徵相比效果有些微提升，這也表明了運動類別中，元數據特徵，本身就較為影響影片的熱門程度，然而文本特徵也能夠稍微對整體提供些許幫助進而提升整體效果。

通過這些觀察，我們可以得出結論：元數據在預測運動類別影片的熱門程度上是非常重要的且有用的特徵，但文本特徵也能對模型效果提供一定提升。

表 -23 運動類別特徵組合

Model	元數據				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.80	0.81	0.82	0.79	<b>0.88</b>
決策樹	0.81	0.81	0.82	0.80	0.82
支持向量機	0.64	0.62	0.79	0.65	0.71
隨機森林	<b>0.85</b>	<b>0.84</b>	<b>0.88</b>	<b>0.83</b>	<b>0.93</b>
梯度提升機	0.83	0.83	0.86	0.82	0.90

Model	描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.68	0.69	0.71	0.67	0.71
決策樹	<b>0.76</b>	<b>0.77</b>	<b>0.75</b>	<b>0.75</b>	<b>0.75</b>
支持向量機	0.69	0.71	0.67	0.66	0.73
隨機森林	0.76	0.76	0.76	0.74	0.80
梯度提升機	<b>0.76</b>	<b>0.76</b>	<b>0.77</b>	<b>0.75</b>	<b>0.81</b>

Model	標題				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.71	0.74	0.71	0.70	0.78
決策樹	<b>0.74</b>	<b>0.76</b>	<b>0.74</b>	<b>0.73</b>	<b>0.74</b>
支持向量機	0.69	0.71	0.67	0.66	0.72
隨機森林	0.76	0.76	0.77	0.75	0.84
梯度提升機	<b>0.78</b>	<b>0.79</b>	<b>0.77</b>	<b>0.76</b>	<b>0.83</b>

Model	評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.68	0.68	0.73	0.68	0.73
決策樹	<b>0.73</b>	<b>0.74</b>	<b>0.73</b>	<b>0.72</b>	<b>0.74</b>



支持向量機	0.69	0.71	0.67	0.66	0.72
隨機森林	0.75	0.74	0.77	0.74	0.80
梯度提升機	<b>0.74</b>	<b>0.75</b>	<b>0.75</b>	<b>0.73</b>	<b>0.81</b>

Model	元數據 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.79	0.79	0.81	0.77	<b>0.89</b>
決策樹	<b>0.84</b>	<b>0.82</b>	<b>0.84</b>	<b>0.82</b>	0.84
支持向量機	0.64	0.62	0.79	0.65	0.71
隨機森林	<b>0.86</b>	<b>0.84</b>	<b>0.89</b>	<b>0.85</b>	<b>0.93</b>
梯度提升機	0.85	0.83	0.87	0.83	0.91

Model	元數據 + 描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.80</b>	<b>0.81</b>	<b>0.82</b>	<b>0.79</b>	<b>0.89</b>
決策樹	0.82	0.82	0.82	0.80	0.82
支持向量機	0.64	0.62	0.79	0.65	0.71
隨機森林	<b>0.85</b>	<b>0.85</b>	<b>0.90</b>	<b>0.85</b>	<b>0.92</b>
梯度提升機	0.84	0.83	0.87	0.83	0.91

Model	元數據 + 標題				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.79</b>	<b>0.80</b>	<b>0.80</b>	<b>0.77</b>	<b>0.89</b>
決策樹	0.81	0.81	0.81	0.79	0.81
支持向量機	0.64	0.62	0.79	0.65	0.71
隨機森林	0.83	0.82	0.84	0.81	0.93
梯度提升機	<b>0.83</b>	<b>0.83</b>	<b>0.86</b>	<b>0.82</b>	<b>0.90</b>

Model	元數據 + 標題 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.79	0.80	0.81	0.78	<b>0.89</b>
決策樹	<b>0.82</b>	<b>0.82</b>	<b>0.83</b>	<b>0.80</b>	0.83
支持向量機	0.64	0.62	0.79	0.65	0.71
隨機森林	<b>0.87</b>	<b>0.85</b>	<b>0.87</b>	<b>0.85</b>	<b>0.94</b>
梯度提升機	0.85	0.83	0.87	0.84	0.92

Model	元數據 + 標題 + 描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.80	0.82	0.82	0.79	<b>0.89</b>
決策樹	0.83	0.82	0.83	0.81	0.83
支持向量機	0.64	0.62	0.79	0.65	0.71
隨機森林	<b>0.84</b>	<b>0.83</b>	<b>0.87</b>	<b>0.83</b>	<b>0.92</b>
梯度提升機	0.83	0.83	0.88	0.83	0.91

Model	元數據 + 標題 + 描述 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.80	0.81	0.82	0.79	<b>0.89</b>
決策樹	<b>0.84</b>	<b>0.83</b>	<b>0.84</b>	<b>0.82</b>	0.85
支持向量機	0.64	0.62	0.79	0.65	0.71
隨機森林	<b>0.87</b>	<b>0.85</b>	<b>0.89</b>	<b>0.85</b>	<b>0.93</b>
梯度提升機	0.86	0.84	0.87	0.84	0.91

在觀察表 24 中可以看出在單一分類器中，三種分類器中比較起來邏輯回歸的效果呈現是相對最佳的，在單一特徵中元數據與單一文本特徵(標題、描述、評論)相比元數據中的準確率、精確率、召回率、F1、AUC 效果最好。在集成分類器中，兩種分類器比較起來隨機森林表現較好，單一特徵中元數據在隨機森林中的準確率、精確率、召回率、F1、AUC 表現較好，但當元數據特徵或與文本數據特徵進行組合時，有些組合中的梯度提升機效果會優於隨機森林。

從表中可以觀察出，全數模型的評估指標中 AUC 有最高的分數。在單一特徵中，元數據特徵與標題、描述、評論相比，元數據效果最好；三種文本特徵相比則是標題特徵效果最好。組合特徵中，元數據特徵依序加入三種不同的單一文本特徵去做特徵組合時，可以發覺組合特徵都比原本的單一文本特徵效果好，而當全部特徵(元數據、評論、標題、評論)與元數據特徵相比，可以發覺在單一分類器中反而是決策樹在準確率、精確率、召回率、F1 上表現較好，而在集成分類器中，隨機森林和梯度提升機的準確率、精確率、召回率、F1、AUC 中效果都有明顯提升，這也表明了在一人物&部落客類別中，結合元數據和文本特徵可以提升模型的效果，文本特徵在與元數據特徵結合使用時能影響影片的熱門程度。

通過這些觀察，我們可以得出結論：在一人物&部落客類別中，結合元數據和文本特徵可以顯著提升模型的效果，且文本特徵在與元數據特徵結合使用時能更好地預測影片的熱門程度，這也意謂著除了元數據外，文本特徵也對熱門影片程度上具有重要作用。

表 -24 人物&amp;部落客類別特徵組合

Model	元數據				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.79</b>	<b>0.80</b>	<b>0.78</b>	<b>0.77</b>	<b>0.82</b>
決策樹	0.76	0.75	0.79	0.75	0.76
支持向量機	0.74	0.73	0.76	0.71	0.81
隨機森林	<b>0.79</b>	<b>0.79</b>	<b>0.80</b>	<b>0.77</b>	<b>0.85</b>
梯度提升機	0.79	0.79	0.79	0.77	0.81

Model	描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.67</b>	<b>0.66</b>	<b>0.73</b>	<b>0.67</b>	<b>0.72</b>
決策樹	0.66	0.67	0.71	0.66	0.66
支持向量機	0.62	0.60	0.67	0.61	0.67
隨機森林	<b>0.69</b>	<b>0.69</b>	<b>0.72</b>	<b>0.68</b>	<b>0.73</b>
梯度提升機	0.67	0.65	0.70	0.65	0.70

Model	標題				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.69</b>	<b>0.68</b>	<b>0.71</b>	<b>0.66</b>	<b>0.76</b>
決策樹	0.66	0.63	0.69	0.64	0.66
支持向量機	0.62	0.60	0.66	0.60	0.66
隨機森林	<b>0.70</b>	<b>0.68</b>	<b>0.72</b>	<b>0.68</b>	<b>0.74</b>
梯度提升機	0.68	0.66	0.70	0.66	0.73

Model	評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.63</b>	<b>0.62</b>	<b>0.69</b>	<b>0.62</b>	<b>0.68</b>
決策樹	0.60	0.55	0.61	0.56	0.60
支持向量機	0.63	0.61	0.66	0.61	0.66
隨機森林	<b>0.64</b>	<b>0.61</b>	<b>0.62</b>	<b>0.59</b>	<b>0.70</b>
梯度提升機	0.63	0.60	0.62	0.59	0.68

Model	元數據 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.77</b>	<b>0.77</b>	<b>0.75</b>	<b>0.74</b>	<b>0.82</b>
決策樹	0.78	0.77	0.81	0.77	0.79
支持向量機	0.74	0.73	0.76	0.71	0.81
隨機森林	0.78	0.74	0.77	0.73	0.85
梯度提升機	<b>0.80</b>	<b>0.79</b>	<b>0.81</b>	<b>0.78</b>	<b>0.83</b>

Model	元數據 + 描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.77</b>	<b>0.77</b>	<b>0.77</b>	<b>0.75</b>	<b>0.81</b>
決策樹	0.76	0.74	0.78	0.74	0.76
支持向量機	0.74	0.73	0.76	0.71	0.81
隨機森林	0.77	0.73	0.79	0.74	0.87
梯度提升機	<b>0.80</b>	<b>0.79</b>	<b>0.81</b>	<b>0.78</b>	<b>0.84</b>

Model	元數據 + 標題				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.79</b>	<b>0.78</b>	<b>0.78</b>	<b>0.76</b>	<b>0.82</b>
決策樹	0.76	0.78	0.80	0.76	0.77
支持向量機	0.74	0.73	0.76	0.71	0.81
隨機森林	<b>0.80</b>	<b>0.80</b>	<b>0.80</b>	<b>0.78</b>	<b>0.89</b>
梯度提升機	0.76	0.77	0.77	0.75	0.82

Model	元數據 + 標題 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.78</b>	<b>0.78</b>	<b>0.77</b>	<b>0.75</b>	<b>0.82</b>
決策樹	0.79	0.79	0.81	0.78	0.79
支持向量機	0.74	0.73	0.76	0.71	0.81
隨機森林	0.78	0.73	0.73	0.71	0.86
梯度提升機	<b>0.80</b>	<b>0.79</b>	<b>0.80</b>	<b>0.78</b>	<b>0.84</b>

Model	元數據 + 標題 + 描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.78</b>	<b>0.77</b>	<b>0.77</b>	<b>0.75</b>	<b>0.81</b>
決策樹	0.75	0.75	0.77	0.74	0.76
支持向量機	0.74	0.73	0.76	0.71	0.81

隨機森林	<b>0.80</b>	<b>0.78</b>	<b>0.80</b>	<b>0.77</b>	<b>0.89</b>
梯度提升機	0.77	0.77	0.77	0.75	0.84

Model	元數據 + 標題 + 描述 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.78	0.77	0.77	0.75	0.81
決策樹	<b>0.80</b>	<b>0.78</b>	<b>0.82</b>	<b>0.78</b>	<b>0.80</b>
支持向量機	0.74	0.73	0.76	0.71	0.81
隨機森林	0.80	0.76	0.79	0.76	0.86
梯度提升機	<b>0.83</b>	<b>0.84</b>	<b>0.83</b>	<b>0.82</b>	<b>0.86</b>

在觀察表 25 中可以看出在單一分類器中，三種分類器中比較起來邏輯回歸的效果呈現是相對最佳的，在單一分類器中邏輯回歸在準確率、精確率、召回率、F1、AUC 的表現上最好，只有在文本特徵中(標題)裡決策樹效果更好一些。在集成分類器中，兩種分類器比較起來梯度提升機表現較好，雖然隨機森林在元數據中的效果最佳，然而當元數據特徵和文本特徵進行特徵組合時，梯度提升機則普遍展現較好的效果。

從表中可以觀察出，全數模型的評估指標中 AUC 有最高的分數。在單一特徵中，元數據特徵與標題、描述、評論相比，元數據效果最好；三種文本特徵相比則是標題特徵效果最好。組合特徵中，元數據特徵依序加入三種文本特徵去做特徵組合時，與元數據特徵相比，當全部特徵都加入時(元數據、評論、標題、評論)在單一分類器上效果有些微的提升，而在集成分類器上則有較明顯的提升，這也表明了娛樂類別中，文本特徵在集成分類器上對影片的熱門程度產生影響。

通過這些觀察，我們可以得出結論：在娛樂類別中，文本特徵對影片的熱門程度有一定的影響，但元數據特徵依然是關鍵，特別是在集成分類器中結合文本特徵時效果最佳。

表 -25 娛樂類別特徵組合

Model	元數據				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.75</b>	<b>0.76</b>	<b>0.79</b>	<b>0.75</b>	<b>0.82</b>
決策樹	0.73	0.72	0.73	0.69	0.74
支持向量機	0.64	0.59	0.67	0.57	0.68
隨機森林	<b>0.78</b>	<b>0.77</b>	<b>0.79</b>	<b>0.75</b>	<b>0.86</b>
梯度提升機	0.72	0.71	0.74	0.69	0.82

Model	描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.68</b>	<b>0.61</b>	<b>0.63</b>	<b>0.60</b>	<b>0.73</b>
決策樹	0.66	0.66	0.70	0.66	0.66
支持向量機	0.67	0.59	0.61	0.58	0.69
隨機森林	<b>0.68</b>	<b>0.67</b>	<b>0.69</b>	<b>0.66</b>	<b>0.74</b>
梯度提升機	0.65	0.65	0.67	0.63	0.68

Model	標題				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	0.71	0.65	0.66	0.63	0.75
決策樹	<b>0.71</b>	<b>0.71</b>	<b>0.72</b>	<b>0.69</b>	<b>0.71</b>
支持向量機	0.68	0.60	0.61	0.58	0.69
隨機森林	0.69	0.67	0.67	0.64	0.75
梯度提升機	<b>0.73</b>	<b>0.73</b>	<b>0.73</b>	<b>0.70</b>	<b>0.77</b>

Model	評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.69</b>	<b>0.62</b>	<b>0.65</b>	<b>0.62</b>	<b>0.71</b>
決策樹	0.63	0.62	0.65	0.61	0.63
支持向量機	0.68	0.60	0.61	0.58	0.69
隨機森林	0.64	0.64	0.63	0.61	0.71
梯度提升機	<b>0.67</b>	<b>0.67</b>	<b>0.66</b>	<b>0.64</b>	<b>0.70</b>

Model	元數據 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.76</b>	<b>0.77</b>	<b>0.80</b>	<b>0.75</b>	<b>0.82</b>
決策樹	0.74	0.75	0.75	0.71	0.76
支持向量機	0.64	0.59	0.67	0.57	0.68
隨機森林	<b>0.74</b>	<b>0.73</b>	<b>0.77</b>	<b>0.71</b>	<b>0.81</b>
梯度提升機	0.75	0.73	0.77	0.72	0.79

Model	元數據 + 描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.76</b>	<b>0.77</b>	<b>0.81</b>	<b>0.76</b>	<b>0.84</b>
決策樹	0.71	0.72	0.70	0.67	0.73
支持向量機	0.64	0.59	0.67	0.57	0.68

隨機森林	0.73	0.72	0.77	0.70	0.82
梯度提升機	<b>0.74</b>	<b>0.73</b>	<b>0.77</b>	<b>0.71</b>	<b>0.80</b>

Model	元數據 + 標題				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.76</b>	<b>0.77</b>	<b>0.80</b>	<b>0.75</b>	<b>0.83</b>
決策樹	0.76	0.77	0.75	0.73	0.77
支持向量機	0.64	0.59	0.67	0.57	0.68
隨機森林	<b>0.76</b>	<b>0.78</b>	<b>0.79</b>	<b>0.74</b>	<b>0.85</b>
梯度提升機	0.76	0.76	0.77	0.72	0.86

Model	元數據 + 標題 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.76</b>	<b>0.77</b>	<b>0.80</b>	<b>0.75</b>	<b>0.82</b>
決策樹	0.78	0.78	0.78	0.74	0.79
支持向量機	0.64	0.59	0.67	0.57	0.68
隨機森林	0.77	0.74	0.76	0.72	0.85
梯度提升機	<b>0.77</b>	<b>0.78</b>	<b>0.79</b>	<b>0.74</b>	<b>0.86</b>

Model	元數據 + 標題 + 描述				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.75</b>	<b>0.76</b>	<b>0.79</b>	<b>0.75</b>	<b>0.83</b>
決策樹	0.76	0.75	0.77	0.73	0.77
支持向量機	0.64	0.59	0.67	0.57	0.68
隨機森林	0.73	0.70	0.74	0.69	0.80
梯度提升機	<b>0.78</b>	<b>0.79</b>	<b>0.80</b>	<b>0.75</b>	<b>0.84</b>

Model	元數據 + 標題 + 描述 + 評論				
	準確率	精確率	召回率	F1	AUC
邏輯回歸	<b>0.76</b>	<b>0.77</b>	<b>0.81</b>	<b>0.76</b>	<b>0.83</b>
決策樹	0.79	0.78	0.80	0.76	0.79
支持向量機	0.64	0.59	0.67	0.57	0.68
隨機森林	0.75	0.72	0.73	0.69	0.86
梯度提升機	<b>0.79</b>	<b>0.82</b>	<b>0.81</b>	<b>0.78</b>	<b>0.87</b>

#### 4.3.4 寬鬆與嚴謹樣本結果的綜合討論

寬鬆樣本資料實驗與結果以及嚴謹樣本資料實驗與結果中，我們可以看到嚴謹樣本中在模型指標的效果都優於寬鬆樣本，嚴謹樣本中，三種文本特徵時標題特徵效果最好；寬鬆樣本中，三種文本特徵時，標題、描述、留言特徵都有相對好的時候，接著會討論兩種樣本間共同的趨勢和差異。

首先，我們來看寬鬆樣本。在單一分類器中，邏輯回歸在準確率、精確率、召回率、F1、AUC 等指標上表現最佳。相比於單一文本特徵（如標題、描述、評論），元數據特徵在多個指標上有較好的效果。在集成分類器中，隨機森林在單一特徵元數據時表現較好，但在組合特徵（元數據與標題、描述、評論）後，梯度提升機的效果較佳。在所有模型中，AUC 指標得分最高。這些結果表明，元數據特徵在預測影片熱門程度上是關鍵因素，特別是在集成分類器中。而文本特徵（尤其是標題）在某些情況下能夠提升模型效果，但整體來看，元數據特徵的影響更大。

接著，我們來看嚴謹樣本。在單一分類器中，邏輯回歸在多個指標上表現較佳，尤其在 AUC 上最高。在單一特徵中，文本特徵在決策樹指標上也有不錯表現。在集成分類器中，隨機森林在單一特徵中效果較好，但在組合特徵後，梯度提升機的效果明顯提升。在所有模型中，AUC 指標得分最高。這些結果表明，元數據特徵在這些樣本中仍然是影響影片熱門程度的重要因素。而文本特徵與元數據特徵進行組合時能夠顯著提升模型的效果。

無論是在寬鬆樣本還是嚴謹樣本中，邏輯回歸在單一分類器中的表現都相對最佳，特別是在元數據的準確率、精確率、召回率、F1、AUC 等指標上。這可能是因為邏輯回歸是一種線性模型，對於這種問題，它可能能夠找到一個良好的決策邊界來區分不同的類別。然而，當我們將更多的文本特徵加入模型時，邏輯回歸的表現可能會受到影響，因為這可能會使問題變得更加複雜，超出了邏輯回歸的處理能力。在集成分類器中，隨機森林的表現普遍較佳，無論是在寬鬆樣本還是嚴謹樣本中。這可能是因為隨機森林是一種集成方法，它通過組合多個決策樹的預測來提高預測的準確性和穩定性。此外，隨機森林還能夠處理大量的特徵並識別出最重要的特徵，這可能是它在這種問題上表現良好的原因。而當單一文本特徵（標題、描述、評論）與元數據進行結合時，梯度提升機會展現出較好的效果，這可能是因為梯度提升機是一種強大的集成方法，它通過逐步增強弱預測器（如決策樹）來改進模型的預測性能，當我們將更多的特徵加入模型時，梯度提升機可能能夠更好地捕捉這些特徵之間的複雜交互，從而提高預測的準確性。

最後，從兩種實驗結果中我們都可以觀察到，在兩組樣本中，我們可以看到以下一些共同點和不同之處。共同點：元數據特徵在多數情況下都表現較好，尤其是在集成分類器中都優於單一分類器。AUC 是所有模型中的主要評估指標，顯示出模型在區分熱門和非熱門影片方面的能力。差異之處：嚴謹樣本中的文本特徵在結合元數據的情況下能夠提升模型效果，尤其是在梯度提升機中的表現。寬鬆樣本中，



隨機森林在單一特徵元數據時表現較佳，但集成分類器中梯度提升機的優勢更加明顯。這些結果顯示，無論是在寬鬆還是嚴謹的樣本分析中，元數據特徵的重要性都是不可忽視的。同時，文本特徵的適當使用可以進一步提升模型的效果，尤其是在與元數據特徵結合使用時。

總體來說如表 26，這些分析結果顯示在單一特徵中元數據都擁有最好的結果，在文本特徵中則是影片標題有相對較好結果，並且當進行混合特徵時，將元數據與三種文本特徵進行組合會有最好的結果，在單一分類器中邏輯回歸有最好的效能，而集成分類器中，雖然隨機森林相對較好，但其實與梯度提升機差異不大，相對單一分類器中都擁有更好的效果。因此這些結果可以幫助我們理解如何使用機器學習方法來分析影片的熱門程度，並揭示了影片元數據特徵和文本特徵的組合在這一過程中的重要性。

表 -26 寬鬆與嚴謹樣本綜合評估表

	單一特徵	文本特徵	混合特徵	單一分類器	集成分類器
運動寬鬆資料	元數據	留言評論	全部特徵	邏輯回歸	隨機森林
運動嚴謹資料	元數據	影片標題	全部特徵	決策樹	隨機森林
人物寬鬆資料	元數據	影片描述	全部特徵	邏輯回歸	梯度提升機
人物嚴謹資料	元數據	影片標題	全部特徵	邏輯回歸	隨機森林
娛樂寬鬆資料	元數據	影片標題	全部特徵	邏輯回歸	隨機森林
娛樂嚴謹資料	元數據	影片標題	全部特徵	邏輯回歸	梯度提升機

## 第五章 結論與未來展望

### 5.1 結論

本研究主要探討四大方面：(1) 分析影片元數據對影片熱門度的影響。(2)分析影片標題和影片描述文本對影片熱門度的影響。(3)分析影片中的觀眾評論文本對影片熱門度的影響。(4) 比較不同特徵組合對影片熱門度的影響。

從第一點在機器學習實驗結果中可以發現，元數據在單一特徵中都擁有最好的結果，並且在統計檢定中元數據變數中的影片時長，在不同類別頻道中幾乎都擁有顯著結果。第二點在利用主題模型生成的文本特徵進行機器學習中可以發現，影片標題在多數時候擁有最好的文本特徵結果，並且在統計檢定中影片標題的長度和詞數，在不同類別頻道中幾乎都擁有顯著結果。第三點在機器學習中可以發現，影片留言在運動類別寬鬆資料中有較好的文本特徵，在統計檢定中，可以發現影片留言在經過情感特徵計算後，留言評論中越正向或越負向的留言幾乎都擁有顯著性的結果。最後第四點中，在比較不同特徵組合時，機器學習模型可以發現在三種類別中，當全部特徵都加入時(元數據+標題+描述+留言)擁有最好的混合特徵結果，並且當利用主題模型生成的文本特徵與元數據結合時，在嚴謹資料集中可以進一步提升模型預測效果。

採用 YouTube 作為本研究資料集，將蒐集的資料以前面的熱門度計算方法將資料分成兩種：寬鬆樣本資料和嚴謹樣本資料，以標題和描述與留言，這三種非結構化的文本資料去做 LDA 主題模型分析，並與結構化的元數據資料總共這四種資料去做十種的特徵組合，使用三種的單一分類器：邏輯回歸、決策樹、支持向量機，與兩種的集成分類器：隨機森林、梯度提升機，總共五種機器學習模型相互深入評估與比較其文本數據與影片原始數據對 YouTube 影片熱門度之影響。

透過統計檢定中的顯著性分析可以看出來，在影片時長中可以發覺在運動類別的影片時長較短的在頻道顯著比例中較高，而在人物 & 部落客類別和娛樂類別中，則是較長的影片擁有較高的頻道顯著結果。而在情感特徵最大、情感特徵最小值中，可以發覺在三個類別中，留言評論的情感特徵波動越大，包括極正向和極負向越有較好的頻道顯著比例。在影片標題的長度與字數多寡，在三個類別中幾乎也都是擁有較長或較多詞與組成的標題有較好的顯著比例。以上這些連續值變數在所有頻道中都具有較為顯著的差異，而在類別值變數中，所有變數都較無特別顯著差異，然而創作者還是可以從不同類別中找尋些微的頻道顯著比例類別以此提供 YouTube 影片創作者一些參考依據，讓創作者能在相似類型的頻道中，深入瞭解各變數間的相互作用，以及各變數間如何影響影片的熱門程度，並依照較為顯著的變數去做影片的加強，來增加對觀眾的吸引力，創作出在相同影片類型中較容易成為熱門影片的影片。

經過主題模型將標題、描述、留言三大文字文本去進行 LDA 主題模型分析與

權重計算後，並在將各自結果投入到機器學習中。根據機器學習中以十折交叉驗證在單一分類器和集成分類器模型訓練的實驗結果，綜合三個類別十個不同創作者的平均值表示，結合表 18 的特徵組合，並以準確率 (Accuracy)、精確率 (Precision)、召回率 (Recall)、調和平均數 (F1-score)、AUC，作為模型評估指標。結果發現集成分類器的結果皆較優於單一分類器，在單一分類器中，不管是嚴謹或寬鬆資料集中幾乎都是邏輯回歸為較好的分類結果，而在集成分類器中，寬鬆資料集中則是梯度提升機有較好的分類結果，而嚴謹資料集中則是隨機森林有較好的分類結果，然而仔細比較會發現其實在集成分類器中兩者模型結果差異不大，皆只有些微差距。

總體來看，三種類別中，運動類別不管在統計分析或機器學習中效果都較好，可能原因是其影片資料量也最多。利用主題模型生成的文本特徵與元數據結合時，在嚴謹資料集中可以進一步提升模型預測效果，三種文本中影片標題的效果最好。在單一特徵中，元數據的特徵效果都明顯優於標題、描述、留言這三大文字文本特徵，而在所有特徵組合中當全部特徵都加入時(元數據+標題+描述+留言)，可以發覺在單一或集成分類器上特徵效果都有些微的提升，也因此當全部特徵都加入時，文本特徵可以提升影片的熱門度效果。

## 5.2 未來展望

以下為本研究的未來展望：

### 1. 資料集的選擇

本研究在社群媒體平台中只使用 YouTube 做為資料集，並對其元數據和評論、標題、留言，將文本數據使用主題模型分析後，進行機器學習，找尋影響影片熱門度的變數進行分析，而在未來可以使用其他不同平台的資料集，比如 Instagram、Twitter、Facebook 等，運用本研究的方法，是否也有相同的結果。

### 2. 改善資料量不平衡

在影片類別頻道主選擇方面，可以盡量讓不同類別中的影片總體數量相當，可以增加不同類別中的比較。

### 3. 主題模型的應用與缺失

本研究在生成的主題模型中，在主題質量上不同主題間可能存在關鍵詞的高度重疊，導致主題不清晰，或有些主題可能包含的關鍵詞沒有實際意義或沒有明顯的主題。未來能在理解主題模型的缺陷和局限性後，在實際應用中加以注意和克服此問題。

## 參考文獻

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Borrajo, E., Colomo-Palacios, R., & Soto-Acosta, P. (2021). Enhancing Audience Engagement on YouTube Through Video Description: An Empirical Study. *Computers in Human Behavior*, 119, Article 106706.
- Chakraborty, A., Ghosh, S., & Basu, S. (2020). Predicting Social Media Performance Metrics and Evaluation of the Impact on Brand Building: A Data Mining Approach. *Journal of Business Research*, 114, 50-62.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM.
- Choudhury, M. D., & Kiciman, E. (2018). Mi YouTube es Su YouTube? Analyzing the Cultures using YouTube Thumbnails of Popular Videos. *Proceedings of the International Conference on Web and Social Media*, 189-198.
- Ding, J., Chen, X., & Zhao, Y. (2017). Popularity Prediction of Posts in Social Networks Based on User, Post and Image Features. *IEEE Transactions on Knowledge and Data Engineering*, 29(2), 448-461.
- Gao, J., Liu, H., & Wu, J. (2018). Popularity Prediction of Images and Videos on Instagram. *International Journal of Multimedia Information Retrieval*, 7(1), 13-23.
- Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly Media.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Jia, Y. (2023). YouTube: A Phenomenon of Video Sharing. *Journal of Broadcasting & Electronic Media*, 67(1), 67-82.

Khan, A. S., Ahmad, N., & Alghazzawi, D. M. (2021). Social Media Prediction: a Literature Review. *IEEE Access*, 9, 53793-53815.

Kim, J., & Lee, K. (2019). Visual Attributes of Thumbnails in Predicting Top YouTube Brand Channels: A Machine Learning Approach. *IEEE Transactions on Multimedia*, 21(7), 1772-1783.

Koller, T., & Grabner, H. (2022). Who Wants to Be a Click-Millionaire? On the Influence of Thumbnails and Captions. In *2022 26th International Conference on Pattern Recognition (ICPR)*, 629–635. IEEE.

Lantz, B. (2019). *Machine learning with R: Expert techniques for predictive modeling* (3rd ed.). Packt Publishing.

Li, R., & Zhang, C. (2017). Category Specific Post Popularity Prediction. In *Proceedings of the 26th International Conference on World Wide Web Companion* (pp. 841-849).

Menard, S. (2010). *Logistic regression: From introductory to advanced concepts and applications*. Sage Publications.

Mishra, A., & Singh, V. K. (2016). Popularity Prediction of Videos in YouTube as Case Study: A Regression Analysis Study. *International Journal of Computer Applications*, 145(5), 22-26.

Rao, T., & Srivastava, S. (2018). Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers. *International Journal of Information Management*, 40, 98-108.

Smith, J., Johnson, A., & Lee, R. (2021). Predicting Elections Results using Social Media Activity: A Case Study: USA Presidential Election 2020. *Journal of Political Communication*, 38(2), 145-160.

Tufekci, Z. (2020). *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press.

Wulczyn, E., Thain, N., & Dixon, L. (2017). Utilizing Topic Modelling To Identify Abusive Comments On YouTube. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 1-16.

## 附錄 A

### 運動類別敘述統計資料

#### 附錄 A - 1、GM 冠名播出連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	709.70	2646.52	172.28	3378.57	736.76	4532.87	178.39	3706.06
影片按讚數	2713.51	1480.53	1056.10	947.40	3058.14	902.26	1219.27	531.80
影片留言數	315.59	148.14	192.69	130.50	378.42	84.33	237.05	101.79
影片觀看數	138140.62	71846.96	35426.23	25501.68	160036.11	50958.25	38718.97	19026.53
情感特徵平均值	0.09	0.09	0.08	0.16	0.09	0.09	0.09	0.20
情感特徵最大值	2.54	1.83	1.40	1.70	2.79	1.43	1.57	1.14
情感特徵最小值	-1.81	-1.36	0.95	1.05	-1.95	-1.16	1.10	1.16
Tags 數量多寡	2.29	2.08	0.92	0.47	2.39	2.00	1.02	0.00
影片熱門度	0.90	0.47	0.23	0.17	1.04	0.33	0.25	0.12
影片標題長度	41.95	36.73	13.40	12.35	42.90	35.28	13.73	10.11
影片標題詞數	17.22	16.75	4.85	4.09	17.10	16.18	4.69	3.58
影片上傳間隔	2.32	1.92	1.40	1.02	2.37	1.77	1.45	0.89

附錄 A - 2、GM 冠名播出類別值

變數名稱	類別值	寬鬆資料集				嚴謹資料集			
		熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	17	170	86.29%	185	93.91%	87	87.88%	96	96.97%
	20	27	13.71%	12	6.09%	12	12.12%	3	3.03%
影片上傳早上、晚上	早上	84	42.64%	68	34.52%	42	42.42%	29	29.29%
	晚上	113	57.36%	129	65.48%	57	57.58%	70	70.71%
是否為日常主題	否	195	98.98%	192	97.46%	97	97.98%	98	98.99%
	是	2	1.02%	5	2.54%	2	2.02%	1	1.01%
是否為延伸續集	否	196	99.49%	146	74.11%	98	98.99%	53	53.54%
	是	1	0.51%	51	25.89%	1	1.01%	46	46.46%
是否與網紅或專家合作	否	56	28.43%	133	67.51%	26	26.26%	75	75.76%
	是	141	71.57%	64	32.49%	73	73.74%	24	24.24%
是否為實測主題	否	192	97.46%	190	96.45%	98	98.99%	97	97.98%
	是	5	2.54%	7	3.55%	1	1.01%	2	2.02%
影片上傳是否過十四天	否	197	100.00%	197	100.00%	99	100.00%	99	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳星期	五	23	11.68%	17	8.63%	11	11.11%	6	6.06%
	一	24	12.18%	24	12.18%	14	14.14%	12	12.12%
	六	36	18.27%	28	14.21%	21	21.21%	12	12.12%
	日	30	15.23%	32	16.24%	13	13.13%	19	19.19%

	四	27	13.71%	48	24.37%	11	11.11%	33	33.33%
	二	34	17.26%	21	10.66%	20	20.20%	8	8.08%
	三	23	11.68%	27	13.71%	9	9.09%	9	9.09%

附錄 A - 3、HU 理 HU 途連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	4064.94	5197.36	4511.28	4135.73	4499.56	4533.69	5265.00	4292.88
影片按讚數	789.26	366.87	350.71	230.57	958.51	283.38	373.30	140.21
影片留言數	113.42	44.17	111.96	80.42	126.87	33.11	97.39	46.91
影片觀看數	60993.86	29859.22	16438.34	8381.31	72813.69	22961.22	15718.50	5482.45
情感特徵平均值	0.04	0.09	0.14	0.19	0.04	0.12	0.14	0.22
情感特徵最大值	1.69	1.09	0.89	1.04	1.82	0.98	0.89	0.72
情感特徵最小值	-1.46	-0.73	0.84	0.83	-1.64	-0.61	0.97	0.97
Tags 數量多寡	3.00	2.32	1.36	1.31	3.22	2.13	1.17	1.46
影片熱門度	1.02	0.50	0.27	0.14	1.21	0.38	0.26	0.09
影片標題長度	39.54	36.36	12.69	12.21	37.71	33.56	13.61	12.60
影片標題詞數	20.44	18.42	6.50	6.68	19.56	16.36	6.43	6.14
影片上傳間隔	2.73	2.34	1.94	1.87	2.69	2.40	1.73	2.16



附錄 A - 4、HU 理 HU 途類別值

變數名稱	類別值	寬鬆資料集				嚴謹資料集			
		熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	17	90	100.00%	90	100.00%	45	100.00%	45	100.00%
影片上傳早上、晚上	早上	36	40.00%	25	27.78%	16	35.56%	17	37.78%
	晚上	54	60.00%	65	72.22%	29	64.44%	28	62.22%
是否為日常主題	否	89	98.89%	89	98.89%	44	97.78%	44	97.78%
	是	1	1.11%	1	1.11%	1	2.22%	1	2.22%
是否為延伸續集	否	71	78.89%	68	75.56%	40	88.89%	39	86.67%
	是	19	21.11%	22	24.44%	5	11.11%	6	13.33%
是否與網紅或專家合作	否	84	93.33%	86	95.56%	43	95.56%	42	93.33%
	是	6	6.67%	4	4.44%	2	4.44%	3	6.67%
是否為實測主題	否	90	100.00%	89	98.89%	45	100.00%	45	100.00%
	是	0	0.00%	1	1.11%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	90	100.00%	90	100.00%	45	100.00%	45	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳星期	五	7	7.78%	9	10.00%	2	4.44%	8	17.78%
	一	14	15.56%	8	8.89%	6	13.33%	3	6.67%
	六	5	5.56%	9	10.00%	2	4.44%	3	6.67%
	日	23	25.56%	13	14.44%	15	33.33%	9	20.00%
	四	14	15.56%	22	24.44%	7	15.56%	9	20.00%

	二	15	16.67%	13	14.44%	8	17.78%	5	11.11%
	三	12	13.33%	16	17.78%	5	11.11%	8	17.78%

附錄 A - 5、likemikesport 連續值

	寬鬆資料集				嚴謹資料集			
變數名稱	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	602.26	395.08	1239.73	671.44	441.52	351.58	775.54	526.87
影片按讚數	392.33	67.38	347.93	41.86	606.46	42.53	381.93	25.27
影片留言數	54.68	7.91	54.39	7.35	82.25	5.47	63.24	5.86
影片觀看數	48372.51	7622.24	36078.57	3853.54	70759.09	4378.15	39579.12	1809.40
情感特徵平均值	0.12	0.13	0.15	0.32	0.13	0.11	0.12	0.27
情感特徵最大值	1.15	0.55	0.79	0.72	1.40	0.41	0.86	0.59
情感特徵最小值	-0.65	-0.18	0.44	0.47	-0.75	-0.13	0.41	0.36
Tags 數量多寡	4.45	3.52	5.65	3.79	4.69	3.84	5.83	3.59
影片熱門度	0.47	0.07	0.35	0.04	0.69	0.04	0.38	0.02
影片標題長度	71.82	73.55	23.00	19.20	73.33	75.65	22.94	17.90
影片標題詞數	34.69	36.36	12.11	11.55	34.88	37.45	11.59	11.50
影片上傳間隔	2.29	2.11	1.90	1.27	2.24	1.95	2.17	1.18

附錄 A - 6、likemikesport 類別值

變數名稱	類別值	寬鬆資料集				嚴謹資料集			
		熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	17	170	100.00%	170	100.00%	85	100.00%	85	100.00%
影片上傳早上、晚上	早上	109	64.12%	116	68.24%	50	58.82%	60	70.59%
	晚上	61	35.88%	54	31.76%	35	41.18%	25	29.41%
是否為日常主題	否	158	92.94%	162	95.29%	78	91.76%	81	95.29%
	是	12	7.06%	8	4.71%	7	8.24%	4	4.71%
是否為延伸續集	否	133	78.24%	143	84.12%	63	74.12%	79	92.94%
	是	37	21.76%	27	15.88%	22	25.88%	6	7.06%
是否與網紅或專家合作	否	40	23.53%	36	21.18%	15	17.65%	19	22.35%
	是	130	76.47%	134	78.82%	70	82.35%	66	77.65%
是否為實測主題	否	168	98.82%	170	100.00%	85	100.00%	85	100.00%
	是	2	1.18%	0	0.00%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	170	100.00%	170	100.00%	85	100.00%	85	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳星期	五	36	21.18%	34	20.00%	17	20.00%	16	18.82%
	一	28	16.47%	27	15.88%	17	20.00%	14	16.47%
	六	10	5.88%	16	9.41%	5	5.88%	9	10.59%
	日	11	6.47%	11	6.47%	6	7.06%	7	8.24%
	四	22	12.94%	25	14.71%	13	15.29%	10	11.76%

	二	30	17.65%	22	12.94%	10	11.76%	11	12.94%
	三	33	19.41%	35	20.59%	17	20.00%	18	21.18%

附錄 A - 7、peeta 哥哥連續值

	寬鬆資料集				嚴謹資料集			
變數名稱	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1152.33	993.43	396.13	360.58	1302.22	968.30	439.04	375.91
影片按讚數	5642.67	1558.11	4261.87	628.93	7908.65	1188.04	4973.85	450.32
影片留言數	305.73	87.74	215.27	48.55	409.43	64.13	236.02	29.06
影片觀看數	335646.40	85668.70	203043.90	32902.35	477855.87	59268.04	196088.16	21649.47
情感特徵平均值	0.26	0.24	0.13	0.16	0.25	0.26	0.14	0.19
情感特徵最大值	3.05	2.08	1.55	0.96	3.21	2.06	1.65	0.97
情感特徵最小值	-1.30	-0.95	0.66	1.04	-1.56	-1.14	0.78	1.43
Tags 數量多寡	3.42	4.87	3.54	3.26	3.04	5.48	3.55	2.95
影片熱門度	0.45	0.11	0.27	0.04	0.63	0.08	0.26	0.03
影片標題長度	56.29	51.61	19.10	15.51	58.70	48.65	22.10	14.20
影片標題詞數	28.51	28.50	5.91	6.55	28.91	28.43	6.26	5.89
影片上傳間隔	7.44	7.35	1.59	3.59	7.65	6.61	1.70	2.19

附錄 A - 8、peeta 哥哥類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	17	45	100.00%	46	100.00%	23	100.00%	23	100.00%
影片上傳早上、晚上	早上	1	2.22%	1	2.17%	1	4.35%	0	0.00%
	晚上	44	97.78%	45	97.83%	22	95.65%	23	100.00%
是否為日常主題	否	40	88.89%	38	82.61%	21	91.30%	19	82.61%
	是	5	11.11%	8	17.39%	2	8.70%	4	17.39%
是否為延伸續集	否	36	80.00%	39	84.78%	19	82.61%	21	91.30%
	是	9	20.00%	7	15.22%	4	17.39%	2	8.70%
是否與網紅或專家合作	否	26	57.78%	34	73.91%	11	47.83%	19	82.61%
	是	19	42.22%	12	26.09%	12	52.17%	4	17.39%
是否為實測主題	否	45	100.00%	46	100.00%	23	100.00%	23	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	44	97.78%	45	97.83%	22	95.65%	23	100.00%
	是	1	2.22%	1	2.17%	1	4.35%	0	0.00%
影片上傳星期	一	0	0.00%	1	2.17%	0	0.00%	1	4.35%
	六	1	2.22%	3	6.52%	0	0.00%	2	8.70%
	日	44	97.78%	42	91.30%	23	100.00%	20	86.96%

附錄 A - 9、台南 josh 連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	2403.91	3437.40	4635.03	4479.34	2666.73	5041.70	4881.98	4656.73
影片按讚數	4273.89	1173.41	6303.08	705.55	6056.02	699.06	8509.12	479.01
影片留言數	776.34	167.82	1014.10	146.07	1126.70	90.20	1324.57	109.25
影片觀看數	240077.16	75953.24	214811.46	32230.65	323299.26	49605.86	279212.71	21162.53
情感特徵平均值	0.09	0.13	0.11	0.24	0.10	0.13	0.11	0.29
情感特徵最大值	2.90	1.84	1.58	1.22	3.32	1.35	1.77	1.01
情感特徵最小值	-2.69	-1.07	5.37	0.76	-3.58	-0.79	7.42	0.67
Tags 數量多寡	3.84	3.60	0.67	0.68	3.81	3.37	0.61	0.73
影片熱門度	0.65	0.20	0.58	0.09	0.87	0.13	0.75	0.06
影片標題長度	32.01	30.02	8.66	9.47	30.63	27.57	9.53	8.43
影片標題詞數	18.87	17.25	4.94	5.58	18.33	16.26	5.52	5.44
影片上傳間隔	2.49	2.27	2.34	1.68	2.59	2.28	2.84	1.77

附錄 A 10、台南 josh 類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	2	2	1.24%	0	0.00%	2	2.47%	0	0.00%
	17	149	92.55%	142	88.20%	70	86.42%	63	77.78%
	19	3	1.86%	2	1.24%	3	3.70%	2	2.47%
	20	2	1.24%	16	9.94%	1	1.23%	15	18.52%
	22	5	3.11%	1	0.62%	5	6.17%	1	1.23%
影片上傳早上、晚上	早上	46	28.57%	26	16.15%	31	38.27%	10	12.35%
	晚上	115	71.43%	135	83.85%	50	61.73%	71	87.65%
是否為日常主題	否	153	95.03%	157	97.52%	75	92.59%	81	100.00%
	是	8	4.97%	4	2.48%	6	7.41%	0	0.00%
是否為延伸續集	否	117	72.67%	142	88.20%	66	81.48%	81	100.00%
	是	44	27.33%	19	11.80%	15	18.52%	0	0.00%
是否與網紅或專家合作	否	155	96.27%	157	97.52%	77	95.06%	80	98.77%
	是	6	3.73%	4	2.48%	4	4.94%	1	1.23%
是否為實測主題	否	160	99.38%	159	98.76%	81	100.00%	80	98.77%
	是	1	0.62%	2	1.24%	0	0.00%	1	1.23%
影片上傳是否過十四天	否	160	99.38%	161	100.00%	80	98.77%	81	100.00%
	是	1	0.62%	0	0.00%	1	1.23%	0	0.00%
影片上傳星期	五	22	13.66%	26	16.15%	14	17.28%	13	16.05%

	一	41	25.47%	26	16.15%	17	20.99%	11	13.58%
	六	15	9.32%	24	14.91%	11	13.58%	13	16.05%
	日	18	11.18%	14	8.70%	10	12.35%	8	9.88%
	四	21	13.04%	21	13.04%	7	8.64%	12	14.81%
	二	25	15.53%	29	18.01%	10	12.35%	15	18.52%
	三	19	11.80%	21	13.04%	12	14.81%	9	11.11%

附錄 A - 11、豹子腿連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1073.01	1453.18	365.65	1504.74	1144.66	1822.98	384.02	1783.87
影片按讚數	1961.12	731.07	862.85	518.94	2425.57	426.09	915.52	213.70
影片留言數	315.00	117.32	405.52	237.47	297.17	50.48	242.81	62.35
影片觀看數	133335.50	37328.42	70754.16	20240.15	178251.85	20191.69	76181.32	10887.78
情感特徵平均值	0.31	0.25	0.24	0.25	0.31	0.19	0.21	0.23
情感特徵最大值	3.87	2.48	2.72	1.68	3.85	1.63	1.79	1.14
情感特徵最小值	-1.45	-1.26	0.93	2.95	-1.41	-0.85	0.81	0.83
Tags 數量多寡	45.42	41.86	15.12	20.41	44.77	37.82	15.80	20.88
影片熱門度	0.62	0.17	0.33	0.09	0.83	0.09	0.35	0.05
影片標題長度	70.54	58.14	15.55	17.56	70.66	51.40	15.52	16.74



影片標題詞數	35.99	31.39	6.75	8.69	35.58	28.24	6.50	8.68
影片上傳間隔	2.37	2.07	1.40	1.40	2.51	1.86	1.38	0.97

附錄 A 12、豹子腿類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	2	5	2.86%	6	3.41%	2	2.27%	4	4.55%
	17	167	95.43%	163	92.61%	86	97.73%	79	89.77%
	19	0	0.00%	5	2.84%	0	0.00%	4	4.55%
	20	0	0.00%	1	0.57%	0	0.00%	1	1.14%
	22	0	0.00%	1	0.57%	0	0.00%	0	0.00%
	24	1	0.57%	0	0.00%	0	0.00%	0	0.00%
	25	1	0.57%	0	0.00%	0	0.00%	0	0.00%
	26	1	0.57%	0	0.00%	0	0.00%	0	0.00%
影片上傳早上、晚上	早上	74	42.29%	81	46.02%	36	40.91%	33	37.50%
	晚上	101	57.71%	95	53.98%	52	59.09%	55	62.50%
是否為日常主題	否	150	85.71%	157	89.20%	76	86.36%	81	92.05%
	是	25	14.29%	19	10.80%	12	13.64%	7	7.95%
是否為延伸續集	否	162	92.57%	168	95.45%	82	93.18%	85	96.59%

	是	13	7.43%	8	4.55%	6	6.82%	3	3.41%
是否與網紅或專家合作	否	141	80.57%	161	91.48%	67	76.14%	85	96.59%
	是	34	19.43%	15	8.52%	21	23.86%	3	3.41%
是否為實測主題	否	166	94.86%	172	97.73%	81	92.05%	87	98.86%
	是	9	5.14%	4	2.27%	7	7.95%	1	1.14%
影片上傳是否過十四天	否	175	100.00%	176	100.00%	88	100.00%	88	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳星期	五	30	17.14%	26	14.77%	19	21.59%	15	17.05%
	一	36	20.57%	36	20.45%	17	19.32%	16	18.18%
	六	20	11.43%	17	9.66%	12	13.64%	8	9.09%
	日	27	15.43%	25	14.20%	15	17.05%	11	12.50%
	四	21	12.00%	26	14.77%	9	10.23%	14	15.91%
	二	20	11.43%	26	14.77%	5	5.68%	13	14.77%
	三	21	12.00%	20	11.36%	11	12.50%	11	12.50%

附錄 A - 13、球給彭尊連續值

	寬鬆資料集				嚴謹資料集			
變數名稱	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	854.81	1814.52	247.61	2961.31	888.52	2671.67	273.13	3686.90
影片按讚數	923.34	241.02	635.82	149.22	1247.44	142.37	752.16	69.77

影片留言數	133.47	38.98	125.04	43.28	157.33	18.85	117.07	16.88
影片觀看數	80678.57	16847.30	64464.61	8713.36	114692.74	9830.33	76260.50	4089.78
情感特徵平均值	0.27	0.23	0.19	0.23	0.29	0.20	0.19	0.27
情感特徵最大值	2.82	1.32	2.77	1.08	3.66	1.07	3.55	1.31
情感特徵最小值	-1.02	-0.50	0.74	0.54	-1.27	-0.24	0.82	0.37
Tags 數量多寡	10.38	8.44	6.03	6.93	10.41	8.52	4.53	6.67
影片熱門度	0.66	0.14	0.52	0.07	0.93	0.08	0.62	0.03
影片標題長度	33.66	31.48	9.68	12.50	33.37	30.78	8.28	14.45
影片標題詞數	19.40	16.87	5.11	4.01	19.41	16.22	5.14	3.39
影片上傳間隔	6.64	6.09	2.14	3.80	6.59	5.93	2.22	4.08

附錄 A 14、球給彭尊類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	17	45	84.91%	43	79.63%	25	92.59%	24	88.89%
	22	8	15.09%	11	20.37%	2	7.41%	3	11.11%
影片上傳早上、晚上	早上	53	100.00%	45	83.33%	27	100.00%	19	70.37%
	晚上	0	0.00%	9	16.67%	0	0.00%	8	29.63%
是否為日常主題	否	34	64.15%	49	90.74%	13	48.15%	27	100.00%

	是	19	35.85%	5	9.26%	14	51.85%	0	0.00%
是否為延續集	否	53	100.00%	54	100.00%	27	100.00%	27	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
是否與網紅或專家合作	否	44	83.02%	43	79.63%	25	92.59%	22	81.48%
	是	9	16.98%	11	20.37%	2	7.41%	5	18.52%
是否為實測主題	否	53	100.00%	54	100.00%	27	100.00%	27	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	53	100.00%	52	96.30%	27	100.00%	26	96.30%
	是	0	0.00%	2	3.70%	0	0.00%	1	3.70%
影片上傳星期	五	2	3.77%	1	1.85%	1	3.70%	0	0.00%
	一	12	22.64%	7	12.96%	7	25.93%	3	11.11%
	六	2	3.77%	1	1.85%	1	3.70%	0	0.00%
	日	1	1.89%	1	1.85%	1	3.70%	1	3.70%
	四	35	66.04%	39	72.22%	16	59.26%	19	70.37%
	二	1	1.89%	2	3.70%	1	3.70%	1	3.70%
	三	0	0.00%	3	5.56%	0	0.00%	3	11.11%

附錄 A - 15、夢多 TV 連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1302.61	1060.66	629.63	546.38	1446.63	960.78	625.57	345.00
影片按讚數	1581.80	438.86	1544.45	190.06	2267.56	317.00	1927.07	110.80
影片留言數	135.63	34.45	122.78	20.15	193.31	24.56	148.72	11.97
影片觀看數	90238.38	18222.25	86167.56	8867.97	134822.34	10790.03	104605.49	4422.13
情感特徵平均值	0.27	0.33	0.13	0.16	0.28	0.35	0.15	0.16
情感特徵最大值	2.42	2.25	1.40	1.25	2.53	2.18	1.46	1.07
情感特徵最小值	-0.93	-0.64	0.65	0.56	-1.07	-0.45	0.81	0.39
Tags 數量多寡	0.11	0.09	0.62	0.75	0.09	0.00	0.53	0.00
影片熱門度	0.48	0.10	0.46	0.05	0.71	0.06	0.55	0.02
影片標題長度	36.20	26.73	14.06	9.09	39.69	25.16	16.14	7.59
影片標題詞數	18.75	14.22	6.92	5.41	19.81	14.03	6.94	5.36
影片上傳間隔	5.72	5.05	5.11	4.30	4.91	5.25	4.64	4.57

附錄 A 16、夢多 TV 類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	2	0	0.00%	1	1.56%	0	0.00%	0	0.00%
	17	51	79.69%	39	60.94%	24	75.00%	16	50.00%
	19	11	17.19%	24	37.50%	8	25.00%	16	50.00%
	24	2	3.13%	0	0.00%	0	0.00%	0	0.00%
影片上傳早上、晚上	早上	57	89.06%	62	96.88%	28	87.50%	30	93.75%
	晚上	7	10.94%	2	3.13%	4	12.50%	2	6.25%
是否為日常主題	否	60	93.75%	57	89.06%	32	100.00%	30	93.75%
	是	4	6.25%	7	10.94%	0	0.00%	2	6.25%
是否為延伸續集	否	61	95.31%	64	100.00%	29	90.63%	32	100.00%
	是	3	4.69%	0	0.00%	3	9.38%	0	0.00%
是否與網紅或專家合作	否	53	82.81%	56	87.50%	26	81.25%	28	87.50%
	是	11	17.19%	8	12.50%	6	18.75%	4	12.50%
是否為實測主題	否	64	100.00%	64	100.00%	32	100.00%	32	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	59	92.19%	62	96.88%	30	93.75%	30	93.75%
	是	5	7.81%	2	3.13%	2	6.25%	2	6.25%
影片上傳星期	五	20	31.25%	17	26.56%	10	31.25%	10	31.25%
	一	4	6.25%	1	1.56%	3	9.38%	1	3.13%

	六	24	37.50%	15	23.44%	11	34.38%	7	21.88%
	日	9	14.06%	15	23.44%	4	12.50%	6	18.75%
	四	5	7.81%	5	7.81%	3	9.38%	3	9.38%
	二	1	1.56%	3	4.69%	1	3.13%	2	6.25%
	三	1	1.56%	8	12.50%	0	0.00%	3	9.38%

附錄 A - 17、館長連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	3550.33	9623.30	4878.84	6272.96	2227.70	7363.57	1314.56	5456.48
影片按讚數	3194.80	538.13	2869.03	143.39	5070.73	456.77	2979.31	86.64
影片留言數	240.97	41.62	205.16	35.73	355.17	35.90	205.34	28.75
影片觀看數	182660.73	18041.82	213287.66	5162.32	307550.37	13951.77	244422.09	3747.62
情感特徵平均值	0.18	0.08	0.14	0.21	0.20	0.11	0.13	0.20
情感特徵最大值	2.49	1.53	1.16	1.11	3.21	1.59	1.01	1.27
情感特徵最小值	-4.06	-1.39	18.51	1.25	-6.84	-1.15	26.09	0.93
Tags 數量多寡	27.97	34.13	13.56	9.81	28.43	32.67	14.21	10.55
影片熱門度	0.17	0.02	0.20	0.00	0.28	0.01	0.22	0.00
影片標題長度	47.67	29.33	13.36	7.21	51.97	28.00	11.79	6.12
影片標題詞數	23.82	14.38	6.48	4.26	25.80	14.13	5.66	4.15
影片上傳間隔	7.82	2.80	5.31	2.58	8.33	3.33	4.24	2.81

附錄 A - 18、館長類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	17	45	75.00%	15	25.00%	23	76.67%	9	30.00%
	20	7	11.67%	44	73.33%	0	0.00%	20	66.67%
	22	4	6.67%	0	0.00%	3	10.00%	0	0.00%
	25	4	6.67%	1	1.67%	4	13.33%	1	3.33%
影片上傳早上、晚上	早上	53	88.33%	19	31.67%	30	100.00%	12	40.00%
	晚上	7	11.67%	41	68.33%	0	0.00%	18	60.00%
是否為日常主題	否	44	73.33%	57	95.00%	22	73.33%	30	100.00%
	是	16	26.67%	3	5.00%	8	26.67%	0	0.00%
是否為延伸續集	否	35	58.33%	60	100.00%	13	43.33%	30	100.00%
	是	25	41.67%	0	0.00%	17	56.67%	0	0.00%
是否與網紅或專家合作	否	22	36.67%	34	56.67%	7	23.33%	19	63.33%
	是	38	63.33%	26	43.33%	23	76.67%	11	36.67%
是否為實測主題	否	60	100.00%	60	100.00%	30	100.00%	30	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	54	90.00%	60	100.00%	28	93.33%	30	100.00%
	是	6	10.00%	0	0.00%	2	6.67%	0	0.00%
影片上傳星期	五	46	76.67%	12	20.00%	27	90.00%	7	23.33%
	一	2	3.33%	8	13.33%	1	3.33%	5	16.67%



	六	3	5.00%	7	11.67%	1	3.33%	3	10.00%
	日	2	3.33%	7	11.67%	0	0.00%	3	10.00%
	四	4	6.67%	7	11.67%	1	3.33%	3	10.00%
	二	1	1.67%	11	18.33%	0	0.00%	6	20.00%
	三	2	3.33%	8	13.33%	0	0.00%	3	10.00%

附錄 A - 19、轟酷看籃球連續值

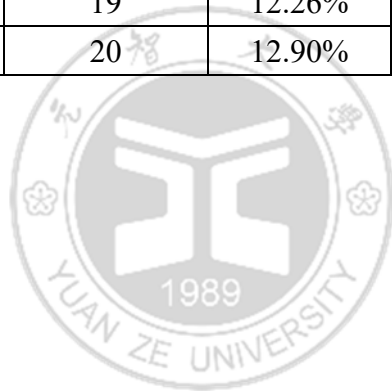
變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	629.01	540.69	107.94	135.18	638.40	509.60	103.62	139.79
影片按讚數	1316.64	626.56	357.98	194.62	1517.03	478.10	341.06	98.93
影片留言數	263.97	102.99	174.48	61.09	327.63	84.04	205.22	49.47
影片觀看數	69895.03	30305.37	16515.59	8876.43	82683.86	22803.44	12763.98	5073.58
情感特徵平均值	0.06	0.07	0.11	0.12	0.06	0.07	0.13	0.12
情感特徵最大值	2.40	1.75	1.32	0.92	2.74	1.60	1.49	0.73
情感特徵最小值	-1.96	-1.58	1.04	1.17	-2.12	-1.46	1.00	0.95
Tags 數量多寡	0.00	0.01	0.00	0.08	0.00	0.01	0.00	0.11
影片熱門度	1.01	0.44	0.24	0.13	1.20	0.33	0.19	0.07
影片標題長度	61.83	46.07	15.31	11.99	64.99	47.94	14.63	10.69

影片標題詞數	31.34	24.23	6.88	4.92	33.22	24.85	6.99	4.94
影片上傳間隔	2.34	2.64	1.37	1.87	2.36	2.65	1.53	2.27

附錄 A - 20、轟酷看籃球類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	17	128	82.58%	81	52.26%	72	92.31%	56	71.79%
	20	27	17.42%	73	47.10%	6	7.69%	21	26.92%
	22	0	0.00%	1	0.65%	0	0.00%	1	1.28%
影片上傳早上、晚上	早上	137	88.39%	125	80.65%	69	88.46%	60	76.92%
	晚上	18	11.61%	30	19.35%	9	11.54%	18	23.08%
是否為日常主題	否	153	98.71%	154	99.35%	77	98.72%	78	100.00%
	是	2	1.29%	1	0.65%	1	1.28%	0	0.00%
是否為延伸續集	否	147	94.84%	150	96.77%	72	92.31%	73	93.59%
	是	8	5.16%	5	3.23%	6	7.69%	5	6.41%
是否與網紅或專家合作	否	142	91.61%	152	98.06%	72	92.31%	77	98.72%
	是	13	8.39%	3	1.94%	6	7.69%	1	1.28%
是否為實測主題	否	154	99.35%	152	98.06%	78	100.00%	76	97.44%
	是	1	0.65%	3	1.94%	0	0.00%	2	2.56%

影片上傳是否過十四天	否	155	100.00%	154	99.35%	78	100.00%	77	98.72%
	是	0	0.00%	1	0.65%	0	0.00%	1	1.28%
影片上傳星期	五	13	8.39%	15	9.68%	7	8.97%	9	11.54%
	一	23	14.84%	30	19.35%	10	12.82%	10	12.82%
	六	24	15.48%	20	12.90%	13	16.67%	9	11.54%
	日	20	12.90%	24	15.48%	11	14.10%	10	12.82%
	四	23	14.84%	27	17.42%	11	14.10%	16	20.51%
	二	24	15.48%	19	12.26%	12	15.38%	14	17.95%
	三	28	18.06%	20	12.90%	14	17.95%	10	12.82%



## 附錄 B

### 人物&部落客類別敘述統計

#### 附錄 B - 1、Blaire Chen 連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1214.06	840.70	339.66	211.55	1288.94	865.59	394.45	214.09
影片按讚數	3390.30	1268.67	1895.35	882.08	4087.53	655.47	2354.49	353.21
影片留言數	189.58	62.48	105.20	38.11	217.94	43.88	130.24	20.87
影片觀看數	186954.94	48851.39	90820.07	26571.25	243360.76	27173.88	96522.95	8725.91
情感特徵平均值	0.67	0.69	0.24	0.29	0.63	0.58	0.24	0.31
情感特徵最大值	5.63	4.48	3.48	3.33	6.08	3.26	4.06	2.12
情感特徵最小值	-1.13	-0.73	1.16	0.82	-1.22	-0.78	1.46	0.93
Tags 數量多寡	10.30	8.82	4.86	3.57	11.12	9.29	5.62	3.62
影片熱門度	0.32	0.08	0.16	0.05	0.42	0.05	0.17	0.02
影片標題長度	45.24	43.15	9.14	8.61	43.00	43.53	8.07	8.40
影片標題詞數	25.45	25.00	5.47	5.01	23.47	24.76	4.43	4.31
影片上傳間隔	9.94	9.21	11.78	8.90	8.76	6.71	11.60	6.90

附錄 B - 2、Blaire Chen 類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	22	33	100.00%	33	100.00%	17	100.00%	17	100.00%
影片上傳早上、晚上	早上	22	66.67%	27	81.82%	11	64.71%	14	82.35%
	晚上	11	33.33%	6	18.18%	6	35.29%	3	17.65%
是否為日常主題	否	28	84.85%	26	78.79%	14	82.35%	14	82.35%
	是	5	15.15%	7	21.21%	3	17.65%	3	17.65%
是否為延伸續集	否	27	81.82%	33	100.00%	14	82.35%	17	100.00%
	是	6	18.18%	0	0.00%	3	17.65%	0	0.00%
是否與網紅或專家合作	否	16	48.48%	26	78.79%	10	58.82%	13	76.47%
	是	17	51.52%	7	21.21%	7	41.18%	4	23.53%
是否為實測主題	否	33	100.00%	33	100.00%	17	100.00%	17	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	25	75.76%	27	81.82%	15	88.24%	16	94.12%
	是	8	24.24%	6	18.18%	2	11.76%	1	5.88%
影片上傳星期	五	2	6.06%	2	6.06%	2	11.76%	0	0.00%
	一	14	42.42%	14	42.42%	8	47.06%	6	35.29%
	六	3	9.09%	1	3.03%	0	0.00%	1	5.88%
	日	3	9.09%	2	6.06%	2	11.76%	1	5.88%
	四	6	18.18%	11	33.33%	2	11.76%	7	41.18%

	二	4	12.12%	3	9.09%	2	11.76%	2	11.76%
	三	1	3.03%	0	0.00%	1	5.88%	0	0.00%

附錄 B - 3、Joeman 連續值

	寬鬆資料集				嚴謹資料集			
變數名稱	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1889.04	1109.04	798.05	550.85	2013.05	889.96	815.20	328.27
影片按讚數	14041.57	4729.13	9364.26	2303.82	18327.85	3348.88	11094.24	1359.78
影片留言數	1242.21	445.17	1358.90	269.69	1537.08	320.79	1607.11	179.53
影片觀看數	980304.38	347169.48	411759.27	115631.25	1272683.94	249177.27	396541.92	61980.74
情感特徵平均值	0.24	0.18	0.12	0.13	0.25	0.15	0.11	0.10
情感特徵最大值	4.76	3.15	3.14	1.99	5.30	2.88	3.51	2.15
情感特徵最小值	-3.02	-2.01	5.93	3.08	-3.14	-1.50	6.46	1.04
Tags 數量多寡	28.99	28.83	0.08	2.34	28.99	28.63	0.11	3.28
影片熱門度	0.38	0.14	0.16	0.04	0.50	0.10	0.15	0.02
影片標題長度	52.05	45.80	16.95	14.18	52.59	43.23	17.43	13.00
影片標題詞數	23.80	21.17	6.15	5.69	23.73	20.05	6.35	5.27
影片上傳間隔	2.72	2.33	1.61	1.38	2.86	2.23	1.73	1.12

附錄 B 4、Joeman 類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	22	155	100.00%	156	100.00%	78	100.00%	78	100.00%
影片上傳早上、晚上	早上	138	89.03%	128	82.05%	69	88.46%	63	80.77%
	晚上	17	10.97%	28	17.95%	9	11.54%	15	19.23%
是否為日常主題	否	144	92.90%	146	93.59%	74	94.87%	73	93.59%
	是	11	7.10%	10	6.41%	4	5.13%	5	6.41%
是否為延伸續集	否	60	38.71%	119	76.28%	26	33.33%	67	85.90%
	是	95	61.29%	37	23.72%	52	66.67%	11	14.10%
是否與網紅或專家合作	否	62	40.00%	122	78.21%	29	37.18%	70	89.74%
	是	93	60.00%	34	21.79%	49	62.82%	8	10.26%
是否為實測主題	否	149	96.13%	145	92.95%	74	94.87%	73	93.59%
	是	6	3.87%	11	7.05%	4	5.13%	5	6.41%
影片上傳是否過十四天	否	155	100.00%	156	100.00%	78	100.00%	78	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳星期	五	20	12.90%	37	23.72%	6	7.69%	24	30.77%
	一	75	48.39%	18	11.54%	40	51.28%	1	1.28%
	六	11	7.10%	25	16.03%	5	6.41%	10	12.82%
	日	8	5.16%	4	2.56%	4	5.13%	1	1.28%
	四	9	5.81%	26	16.67%	6	7.69%	16	20.51%

	二	11	7.10%	29	18.59%	6	7.69%	18	23.08%
	三	21	13.55%	17	10.90%	11	14.10%	8	10.26%

附錄 B - 5、一隻阿圓連續值

	寬鬆資料集				嚴謹資料集			
變數名稱	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1581.40	1148.12	800.64	439.73	1785.37	1114.59	991.06	425.10
影片按讚數	4568.42	2046.37	2524.79	642.84	6001.87	1695.28	2786.50	497.78
影片留言數	417.57	187.24	435.50	142.22	570.89	148.89	554.81	100.22
影片觀看數	319002.21	137487.75	155743.34	40003.16	410136.61	107236.61	173956.99	33741.34
情感特徵平均值	0.51	0.54	0.20	0.21	0.52	0.55	0.23	0.24
情感特徵最大值	5.69	4.53	2.64	2.33	6.22	4.47	3.12	2.85
情感特徵最小值	-1.88	-0.89	5.38	0.79	-2.67	-0.80	7.44	0.74
Tags 數量多寡	5.79	5.16	3.29	3.50	6.15	4.28	2.93	3.13
影片熱門度	0.39	0.17	0.19	0.05	0.50	0.13	0.21	0.04
影片標題長度	45.91	38.99	15.42	10.58	46.22	39.46	14.19	11.47
影片標題詞數	25.91	24.24	4.36	5.47	25.78	24.30	4.14	6.15
影片上傳間隔	3.99	3.78	1.43	1.18	3.98	3.70	1.54	1.30



附錄 B - 6、一隻阿圓類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	19	13	14.44%	10	10.99%	11	23.91%	6	13.04%
	22	37	41.11%	44	48.35%	14	30.43%	24	52.17%
	24	40	44.44%	37	40.66%	21	45.65%	16	34.78%
影片上傳早上、晚上	早上	88	97.78%	90	98.90%	45	97.83%	46	100.00%
	晚上	2	2.22%	1	1.10%	1	2.17%	0	0.00%
是否為日常主題	否	85	94.44%	90	98.90%	43	93.48%	46	100.00%
	是	5	5.56%	1	1.10%	3	6.52%	0	0.00%
是否為延伸續集	否	86	95.56%	89	97.80%	44	95.65%	46	100.00%
	是	4	4.44%	2	2.20%	2	4.35%	0	0.00%
是否與網紅或專家合作	否	71	78.89%	76	83.52%	32	69.57%	37	80.43%
	是	19	21.11%	15	16.48%	14	30.43%	9	19.57%
是否為實測主題	否	90	100.00%	91	100.00%	46	100.00%	46	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	90	100.00%	91	100.00%	46	100.00%	46	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳星期	五	1	1.11%	0	0.00%	0	0.00%	0	0.00%
	一	0	0.00%	6	6.59%	0	0.00%	4	8.70%
	日	52	57.78%	34	37.36%	31	67.39%	15	32.61%

	四	1	1.11%	3	3.30%	1	2.17%	2	4.35%
	二	0	0.00%	2	2.20%	0	0.00%	1	2.17%
	三	36	40.00%	46	50.55%	14	30.43%	24	52.17%

附錄 B - 7、千千進食中連續值

	寬鬆資料集				嚴謹資料集			
變數名稱	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1217.85	3912.10	539.25	2301.17	1263.00	4741.21	620.41	1992.82
影片按讚數	9725.55	2233.68	4019.52	1758.27	12270.09	1001.09	4113.90	286.92
影片留言數	480.54	99.35	283.13	165.74	598.50	27.56	297.45	15.13
影片觀看數	702317.06	141323.51	298955.99	102808.63	901956.79	64597.12	300258.12	17981.94
情感特徵平均值	0.32	0.39	0.10	0.16	0.32	0.39	0.10	0.19
情感特徵最大值	4.44	2.85	1.89	1.61	4.53	2.32	2.03	1.37
情感特徵最小值	-1.47	-0.64	1.45	0.58	-1.76	-0.50	1.95	0.58
Tags 數量多寡	0.09	0.00	0.73	0.00	0.00	0.00	0.00	0.00
影片熱門度	0.39	0.08	0.17	0.06	0.50	0.04	0.17	0.01
影片標題長度	39.24	27.57	8.17	12.82	36.97	21.94	7.40	9.64
影片標題詞數	24.58	16.49	4.13	7.19	23.68	12.65	4.13	4.95
影片上傳間隔	5.10	5.12	2.72	2.25	5.65	4.65	2.92	1.76

附錄 B - 8、千千進食中類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	22	67	100.00%	61	89.71%	34	100.00%	29	85.29%
	24	0	0.00%	7	10.29%	0	0.00%	5	14.71%
影片上傳早上、晚上	早上	31	46.27%	11	16.18%	16	47.06%	4	11.76%
	晚上	36	53.73%	57	83.82%	18	52.94%	30	88.24%
是否為日常主題	否	65	97.01%	66	97.06%	33	97.06%	33	97.06%
	是	2	2.99%	2	2.94%	1	2.94%	1	2.94%
是否為延伸續集	否	60	89.55%	67	98.53%	30	88.24%	34	100.00%
	是	7	10.45%	1	1.47%	4	11.76%	0	0.00%
是否與網紅或專家合作	否	57	85.07%	58	85.29%	30	88.24%	32	94.12%
	是	10	14.93%	10	14.71%	4	11.76%	2	5.88%
是否為實測主題	否	67	100.00%	67	98.53%	34	100.00%	34	100.00%
	是	0	0.00%	1	1.47%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	66	98.51%	68	100.00%	33	97.06%	34	100.00%
	是	1	1.49%	0	0.00%	1	2.94%	0	0.00%
影片上傳星期	五	48	71.64%	11	16.18%	27	79.41%	0	0.00%
	一	0	0.00%	3	4.41%	0	0.00%	1	2.94%
	六	17	25.37%	4	5.88%	6	17.65%	0	0.00%
	日	0	0.00%	44	64.71%	0	0.00%	28	82.35%

	四	0	0.00%	2	2.94%	0	0.00%	1	2.94%
	三	2	2.99%	4	5.88%	1	2.94%	4	11.76%

附錄 B - 9、布萊恩的世界連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	835.59	755.10	327.04	442.32	845.93	713.93	284.39	454.74
影片按讚數	307.21	107.41	166.49	47.86	404.80	74.27	178.91	29.15
影片留言數	49.95	21.47	28.46	19.00	64.93	17.80	28.88	21.66
影片觀看數	23832.45	5670.81	18928.91	2479.04	34484.30	3589.73	21424.39	1213.49
情感特徵平均值	0.27	0.33	0.14	0.20	0.26	0.34	0.12	0.23
情感特徵最大值	1.73	1.46	1.03	0.84	1.86	1.27	1.22	0.86
情感特徵最小值	-0.49	-0.23	0.43	0.28	-0.67	-0.21	0.49	0.31
Tags 數量多寡	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
影片熱門度	0.37	0.09	0.29	0.04	0.53	0.06	0.33	0.02
影片標題長度	37.12	35.34	16.95	15.50	40.50	36.03	17.88	14.71
影片標題詞數	17.10	16.41	5.47	5.96	18.97	16.97	5.02	6.13
影片上傳間隔	6.16	5.27	3.15	2.71	5.80	5.10	2.76	2.28

附錄 B - 10、布萊恩的世界類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	22	58	100.00%	59	100.00%	30	100.00%	30	100.00%
影片上傳早上、晚上	早上	39	67.24%	41	69.49%	18	60.00%	21	70.00%
	晚上	19	32.76%	18	30.51%	12	40.00%	9	30.00%
是否為日常主題	否	53	91.38%	56	94.92%	28	93.33%	29	96.67%
	是	5	8.62%	3	5.08%	2	6.67%	1	3.33%
是否為延伸續集	否	57	98.28%	57	96.61%	30	100.00%	28	93.33%
	是	1	1.72%	2	3.39%	0	0.00%	2	6.67%
是否與網紅或專家合作	否	56	96.55%	57	96.61%	28	93.33%	29	96.67%
	是	2	3.45%	2	3.39%	2	6.67%	1	3.33%
是否為實測主題	否	58	100.00%	59	100.00%	30	100.00%	30	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	57	98.28%	59	100.00%	30	100.00%	30	100.00%
	是	1	1.72%	0	0.00%	0	0.00%	0	0.00%
影片上傳星期	五	5	8.62%	8	13.56%	3	10.00%	5	16.67%
	一	9	15.52%	11	18.64%	4	13.33%	6	20.00%
	六	2	3.45%	2	3.39%	2	6.67%	1	3.33%
	日	15	25.86%	13	22.03%	8	26.67%	5	16.67%
	四	12	20.69%	10	16.95%	6	20.00%	7	23.33%

	二	6	10.34%	4	6.78%	5	16.67%	2	6.67%
	三	9	15.52%	11	18.64%	2	6.67%	4	13.33%

附錄 B - 11、阿鑫連續值

	寬鬆資料集				嚴謹資料集			
變數名稱	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1012.51	876.13	239.69	219.41	1057.58	904.00	242.85	238.31
影片按讚數	888.30	386.03	345.88	163.27	1056.68	276.21	321.40	115.80
影片留言數	166.30	96.97	220.60	223.31	252.00	58.95	284.27	134.46
影片觀看數	64420.38	24986.08	21661.65	9760.87	80709.89	17084.26	18413.56	6106.65
情感特徵平均值	0.25	0.26	0.16	0.16	0.29	0.26	0.20	0.15
情感特徵最大值	2.15	1.78	1.26	1.00	2.49	1.58	1.56	0.72
情感特徵最小值	-0.79	-0.49	0.42	0.44	-0.84	-0.49	0.38	0.48
Tags 數量多寡	8.68	4.11	4.99	5.29	9.42	3.05	4.68	4.88
影片熱門度	0.68	0.26	0.23	0.10	0.85	0.18	0.19	0.06
影片標題長度	54.22	44.61	24.34	17.13	52.37	46.63	24.81	20.52
影片標題詞數	25.54	23.00	8.71	8.14	26.11	23.89	10.04	7.58
影片上傳間隔	9.16	8.53	6.61	4.75	9.47	8.74	7.63	6.18

附錄 B 12、阿鑫類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	22	37	100.00%	38	100.00%	19	100.00%	19	100.00%
影片上傳早上、晚上	早上	1	2.70%	0	0.00%	1	5.26%	0	0.00%
	晚上	36	97.30%	38	100.00%	18	94.74%	19	100.00%
是否為日常主題	否	29	78.38%	30	78.95%	16	84.21%	17	89.47%
	是	8	21.62%	8	21.05%	3	15.79%	2	10.53%
是否為延伸續集	否	35	94.59%	36	94.74%	18	94.74%	19	100.00%
	是	2	5.41%	2	5.26%	1	5.26%	0	0.00%
是否與網紅或專家合作	否	30	81.08%	30	78.95%	18	94.74%	14	73.68%
	是	7	18.92%	8	21.05%	1	5.26%	5	26.32%
是否為實測主題	否	37	100.00%	38	100.00%	19	100.00%	19	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	33	89.19%	35	92.11%	17	89.47%	17	89.47%
	是	4	10.81%	3	7.89%	2	10.53%	2	10.53%
影片上傳星期	日	9	24.32%	12	31.58%	5	26.32%	6	31.58%
	四	28	75.68%	25	65.79%	14	73.68%	13	68.42%
	三	0	0.00%	1	2.63%	0	0.00%	0	0.00%

附錄 B - 13、陸子玄連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	724.84	684.44	156.01	221.72	775.96	657.57	178.61	232.75
影片按讚數	1156.53	444.78	1954.08	79.42	1673.96	396.09	2656.75	68.14
影片留言數	141.36	92.22	126.00	24.12	176.43	87.22	169.28	24.00
影片觀看數	57674.36	10672.09	138818.18	2492.08	94618.17	8600.74	188719.99	1365.92
情感特徵平均值	0.31	0.38	0.13	0.14	0.34	0.38	0.12	0.15
情感特徵最大值	2.85	2.83	0.94	1.06	3.09	2.60	0.93	0.86
情感特徵最小值	-0.98	-0.73	0.59	0.50	-0.96	-0.56	0.74	0.41
Tags 數量多寡	4.93	7.04	6.00	8.72	5.43	6.61	6.00	8.62
影片熱門度	0.44	0.08	1.05	0.02	0.72	0.07	1.43	0.01
影片標題長度	29.93	27.76	12.30	11.41	33.26	30.39	14.38	14.07
影片標題詞數	17.38	16.22	5.58	5.50	18.48	17.22	5.66	6.45
影片上傳間隔	7.18	7.42	1.27	1.48	7.30	7.65	0.70	1.67

附錄 B - 14、陸子玄類別值

變數名稱	類別值	寬鬆資料集				嚴謹資料集			
		熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比



影片類別	10	0	0.00%	1	2.22%	0	0.00%	1	4.35%
	19	0	0.00%	2	4.44%	0	0.00%	1	4.35%
	22	45	100.00%	42	93.33%	23	100.00%	21	91.30%
影片上傳早上、晚上	早上	44	97.78%	45	100.00%	23	100.00%	23	100.00%
	晚上	1	2.22%	0	0.00%	0	0.00%	0	0.00%
是否為日常主題	否	43	95.56%	44	97.78%	23	100.00%	23	100.00%
	是	2	4.44%	1	2.22%	0	0.00%	0	0.00%
是否為延伸續集	否	45	100.00%	45	100.00%	23	100.00%	23	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
是否與網紅或專家 合作	否	31	68.89%	34	75.56%	13	56.52%	16	69.57%
	是	14	31.11%	11	24.44%	10	43.48%	7	30.43%
是否為實測主題	否	44	97.78%	44	97.78%	22	95.65%	22	95.65%
	是	1	2.22%	1	2.22%	1	4.35%	1	4.35%
影片上傳是否過十四 天	否	45	100.00%	45	100.00%	23	100.00%	23	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳星期	五	42	93.33%	42	93.33%	21	91.30%	21	91.30%
	六	2	4.44%	1	2.22%	1	4.35%	0	0.00%
	日	1	2.22%	1	2.22%	1	4.35%	1	4.35%
	三	0	0.00%	1	2.22%	0	0.00%	1	4.35%

附錄 B - 15、壹加壹連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1215.59	2557.53	336.87	1584.29	1322.73	3968.31	224.09	656.03
影片按讚數	6190.69	1530.80	2293.90	1543.99	7476.69	176.23	2482.60	46.18
影片留言數	337.14	80.29	170.05	85.42	418.85	14.15	189.00	8.24
影片觀看數	354677.39	86463.39	106847.30	77657.08	436961.81	17947.88	83886.57	4414.56
情感特徵平均值	0.28	0.18	0.13	0.27	0.29	0.09	0.10	0.35
情感特徵最大值	4.41	2.09	7.22	1.80	3.69	1.42	1.52	1.55
情感特徵最小值	-1.27	-1.09	0.48	1.34	-1.24	-1.13	0.34	1.80
Tags 數量多寡	25.02	16.08	4.37	8.94	25.69	10.19	4.13	7.69
影片熱門度	0.53	0.13	0.16	0.12	0.66	0.03	0.13	0.01
影片標題長度	30.31	35.27	6.79	10.08	30.08	37.38	6.32	9.19
影片標題詞數	20.18	20.22	4.20	4.07	19.96	20.69	3.83	4.68
影片上傳間隔	6.59	6.18	4.12	4.40	6.65	5.08	4.05	2.21

附錄 B - 16、壹加壹類別值

變數名稱	類別值	寬鬆資料集				嚴謹資料集			
		熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	19	5	9.80%	7	13.73%	3	11.54%	0	0.00%

	22	44	86.27%	44	86.27%	21	80.77%	26	100.00%
	24	1	1.96%	0	0.00%	1	3.85%	0	0.00%
	26	1	1.96%	0	0.00%	1	3.85%	0	0.00%
影片上傳早上、晚上	早上	1	1.96%	8	15.69%	0	0.00%	0	0.00%
	晚上	50	98.04%	43	84.31%	26	100.00%	26	100.00%
是否為日常主題	否	49	96.08%	48	94.12%	25	96.15%	26	100.00%
	是	2	3.92%	3	5.88%	1	3.85%	0	0.00%
是否為延伸續集	否	47	92.16%	23	45.10%	24	92.31%	0	0.00%
	是	4	7.84%	28	54.90%	2	7.69%	26	100.00%
是否與網紅或專家合作	否	49	96.08%	46	90.20%	26	100.00%	23	88.46%
	是	2	3.92%	5	9.80%	0	0.00%	3	11.54%
是否為實測主題	否	49	96.08%	50	98.04%	25	96.15%	26	100.00%
	是	2	3.92%	1	1.96%	1	3.85%	0	0.00%
影片上傳是否過十四天	否	48	94.12%	47	92.16%	25	96.15%	26	100.00%
	是	3	5.88%	4	7.84%	1	3.85%	0	0.00%
影片上傳星期	五	1	1.96%	6	11.76%	0	0.00%	0	0.00%
	一	1	1.96%	1	1.96%	1	3.85%	1	3.85%
	六	22	43.14%	11	21.57%	12	46.15%	0	0.00%
	日	0	0.00%	25	49.02%	0	0.00%	24	92.31%
	四	4	7.84%	1	1.96%	0	0.00%	0	0.00%
	二	2	3.92%	3	5.88%	0	0.00%	1	3.85%
	三	21	41.18%	4	7.84%	13	50.00%	0	0.00%

附錄 B - 17、琳姐連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	907.33	721.38	242.32	387.84	912.75	581.88	267.40	432.57
影片按讚數	2446.54	614.38	1922.26	276.71	3652.42	420.25	2081.08	179.42
影片留言數	299.10	108.50	224.30	105.50	355.08	62.88	240.50	70.53
影片觀看數	180521.71	30329.21	138729.20	15473.49	276034.63	17657.38	140780.21	7140.49
情感特徵平均值	0.50	0.43	0.31	0.27	0.49	0.39	0.26	0.23
情感特徵最大值	3.34	2.45	1.49	1.52	3.81	1.86	1.85	1.03
情感特徵最小值	-0.91	-0.44	0.79	0.32	-1.15	-0.35	1.04	0.39
Tags 數量多寡	24.00	21.00	0.00	8.02	24.00	19.00	0.00	9.96
影片熱門度	0.94	0.16	0.72	0.08	1.43	0.09	0.73	0.04
影片標題長度	58.88	48.17	16.72	16.16	58.08	44.96	15.90	15.19
影片標題詞數	34.08	28.54	5.79	9.89	34.38	25.54	5.46	10.29
影片上傳間隔	7.19	6.52	3.23	3.16	8.17	6.00	3.57	2.80

附錄 B - 18、琳姐類別值

變數名稱	類別值	寬鬆資料集				嚴謹資料集			
		熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比

影片類別	22	48	100.00%	48	100.00%	24	100.00%	24	100.00%
影片上傳早上、晚上	早上	5	10.42%	6	12.50%	3	12.50%	5	20.83%
	晚上	43	89.58%	42	87.50%	21	87.50%	19	79.17%
是否為日常主題	否	46	95.83%	47	97.92%	23	95.83%	23	95.83%
	是	2	4.17%	1	2.08%	1	4.17%	1	4.17%
是否為延伸續集	否	48	100.00%	48	100.00%	24	100.00%	24	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
是否與網紅或專家合作	否	26	54.17%	36	75.00%	12	50.00%	18	75.00%
	是	22	45.83%	12	25.00%	12	50.00%	6	25.00%
是否為實測主題	否	48	100.00%	48	100.00%	24	100.00%	24	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	46	95.83%	48	100.00%	22	91.67%	24	100.00%
	是	2	4.17%	0	0.00%	2	8.33%	0	0.00%
影片上傳星期	五	7	14.58%	5	10.42%	2	8.33%	4	16.67%
	一	1	2.08%	3	6.25%	0	0.00%	2	8.33%
	六	4	8.33%	5	10.42%	4	16.67%	2	8.33%
	日	35	72.92%	30	62.50%	18	75.00%	12	50.00%
	四	0	0.00%	1	2.08%	0	0.00%	1	4.17%
	二	0	0.00%	1	2.08%	0	0.00%	1	4.17%
	三	1	2.08%	3	6.25%	0	0.00%	2	8.33%

附錄 B - 19、滴妹連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	787.93	750.77	203.32	291.20	784.23	721.95	181.81	180.85
影片按讚數	11952.40	4043.60	8384.23	1442.04	16655.55	3073.73	9466.80	787.80
影片留言數	619.09	185.53	755.46	116.12	647.73	149.64	337.93	103.57
影片觀看數	593091.98	174391.37	448039.09	56085.71	820079.05	129336.77	536993.17	30179.02
情感特徵平均值	0.35	0.39	0.14	0.13	0.31	0.42	0.12	0.13
情感特徵最大值	3.57	3.35	1.22	1.78	3.60	3.24	1.01	1.49
情感特徵最小值	-1.29	-0.91	0.73	0.62	-1.51	-0.81	0.80	0.56
Tags 數量多寡	2.98	2.51	2.75	1.99	3.55	2.18	3.39	0.85
影片熱門度	0.41	0.12	0.31	0.04	0.56	0.09	0.37	0.02
影片標題長度	43.56	39.63	20.46	13.60	46.41	38.64	25.60	11.44
影片標題詞數	23.14	22.21	5.86	4.38	23.23	22.64	6.80	4.51
影片上傳間隔	7.67	7.63	2.17	2.09	7.73	7.77	2.68	1.48

附錄 B - 20、滴妹類別值

變數名稱	類別值	寬鬆資料集				嚴謹資料集			
		熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比

影片類別	22	43	100.00%	43	100.00%	22	100.00%	22	100.00%
影片上傳早上、晚上	早上	43	100.00%	43	100.00%	22	100.00%	22	100.00%
	晚上	0	0.00%	0	0.00%	0	0.00%	0	0.00%
是否為日常主題	否	39	90.70%	41	95.35%	19	86.36%	21	95.45%
	是	4	9.30%	2	4.65%	3	13.64%	1	4.55%
是否為延伸續集	否	41	95.35%	41	95.35%	20	90.91%	22	100.00%
	是	2	4.65%	2	4.65%	2	9.09%	0	0.00%
是否與網紅或專家 合作	否	35	81.40%	37	86.05%	18	81.82%	19	86.36%
	是	8	18.60%	6	13.95%	4	18.18%	3	13.64%
是否為實測主題	否	43	100.00%	41	95.35%	22	100.00%	21	95.45%
	是	0	0.00%	2	4.65%	0	0.00%	1	4.55%
影片上傳是否過十四 天	否	42	97.67%	42	97.67%	21	95.45%	22	100.00%
	是	1	2.33%	1	2.33%	1	4.55%	0	0.00%
影片上傳星期	五	2	4.65%	0	0.00%	2	9.09%	0	0.00%
	六	41	95.35%	41	95.35%	20	90.91%	22	100.00%
	日	0	0.00%	1	2.33%	0	0.00%	0	0.00%
	二	0	0.00%	1	2.33%	0	0.00%	0	0.00%

## 附錄 C

### 娛樂類別敘述統計

附錄 C - 1、 Hook 連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1158.04	1251.74	275.34	1119.54	1182.52	1529.65	299.25	1528.20
影片按讚數	35141.18	22269.70	12413.33	9089.50	43003.61	19126.65	12553.29	10713.20
影片留言數	1172.20	769.70	563.34	440.61	1473.74	739.30	584.98	539.08
影片觀看數	1737463.96	918542.65	455166.49	317631.43	2072384.17	697594.17	406217.55	317249.35
情感特徵平均值	0.37	0.49	0.12	0.23	0.41	0.50	0.12	0.27
情感特徵最大值	5.46	5.75	1.69	2.95	5.79	5.86	1.79	3.72
情感特徵最小值	-4.22	-2.94	10.28	7.40	-2.32	-2.00	1.34	2.11
Tags 數量多寡	20.71	14.89	4.91	7.61	19.70	11.70	4.50	7.62
影片熱門度	1.26	0.67	0.33	0.23	1.50	0.51	0.29	0.23
影片標題長度	48.38	45.00	9.13	15.06	45.91	41.78	8.78	17.81
影片標題詞數	24.22	21.76	3.46	5.88	22.91	20.91	2.70	6.63
影片上傳間隔	7.40	7.35	1.78	2.63	8.00	6.52	1.51	2.79



附錄 C - 2、 Hook 類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	20	1	2.22%	0	0.00%	0	0.00%	0	0.00%
	24	44	97.78%	46	100.00%	23	100.00%	23	100.00%
影片上傳早上、晚上	早上	44	97.78%	41	89.13%	23	100.00%	18	78.26%
	晚上	1	2.22%	5	10.87%	0	0.00%	5	21.74%
是否為日常主題	否	42	93.33%	42	91.30%	21	91.30%	22	95.65%
	是	3	6.67%	4	8.70%	2	8.70%	1	4.35%
是否為延伸續集	否	45	100.00%	46	100.00%	23	100.00%	23	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
是否與網紅或專家合作	否	8	17.78%	16	34.78%	6	26.09%	11	47.83%
	是	37	82.22%	30	65.22%	17	73.91%	12	52.17%
是否為實測主題	否	45	100.00%	45	97.83%	23	100.00%	23	100.00%
	是	0	0.00%	1	2.17%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	45	100.00%	46	100.00%	23	100.00%	23	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳星期	五	0	0.00%	3	6.52%	0	0.00%	3	13.04%
	一	2	4.44%	3	6.52%	1	4.35%	1	4.35%
	六	1	2.22%	0	0.00%	0	0.00%	0	0.00%
	日	41	91.11%	37	80.43%	22	95.65%	17	73.91%

	三	0	0.00%	3	6.52%	0	0.00%	2	8.70%
	二	1	2.22%	0	0.00%	0	0.00%	0	0.00%

附錄 C 3、HowFun 連續值

	寬鬆資料集				嚴謹資料集			
變數名稱	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	593.28	1085.29	664.34	1886.68	564.34	1615.86	630.20	2484.83
影片按讚數	12998.86	6003.19	7294.91	2524.27	16652.54	4309.06	8647.70	1962.38
影片留言數	791.30	371.70	797.71	290.48	865.49	261.40	502.27	173.54
影片觀看數	632852.06	198013.17	536976.08	69639.16	903809.63	142649.97	649444.80	53188.46
情感特徵平均值	0.14	0.14	0.12	0.11	0.13	0.16	0.10	0.12
情感特徵最大值	3.30	2.35	1.68	1.14	3.44	2.29	2.00	1.22
情感特徵最小值	-1.97	-1.49	1.00	0.83	-2.09	-1.30	0.92	0.81
Tags 數量多寡	1.12	1.59	2.69	3.30	0.89	1.83	2.10	3.57
影片熱門度	0.42	0.13	0.36	0.05	0.61	0.10	0.44	0.04
影片標題長度	23.42	22.47	13.39	9.03	20.11	23.34	8.51	9.14
影片標題詞數	11.58	11.90	7.76	5.18	9.74	12.74	4.93	5.30
影片上傳間隔	5.80	3.93	5.78	3.15	6.09	3.74	6.55	3.17

附錄 C - 4、HowFun 類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	19	0	0.00%	1	1.43%	0	0.00%	1	2.86%
	20	2	2.90%	8	11.43%	2	5.71%	7	20.00%
	22	9	13.04%	21	30.00%	5	14.29%	12	34.29%
	24	58	84.06%	40	57.14%	28	80.00%	15	42.86%
影片上傳早上、晚上	早上	55	79.71%	52	74.29%	28	80.00%	22	62.86%
	晚上	14	20.29%	18	25.71%	7	20.00%	13	37.14%
是否為日常主題	否	69	100.00%	69	98.57%	35	100.00%	35	100.00%
	是	0	0.00%	1	1.43%	0	0.00%	0	0.00%
是否為延伸續集	否	62	89.86%	67	95.71%	33	94.29%	34	97.14%
	是	7	10.14%	3	4.29%	2	5.71%	1	2.86%
是否與網紅或專家合作	否	66	95.65%	70	100.00%	34	97.14%	35	100.00%
	是	3	4.35%	0	0.00%	1	2.86%	0	0.00%
是否為實測主題	否	69	100.00%	70	100.00%	35	100.00%	35	100.00%
	是	0	0.00%	0	0.00%	33	94.29%	35	100.00%
影片上傳是否過十四天	否	65	94.20%	69	98.57%	2	5.71%	0	0.00%
	是	4	5.80%	1	1.43%	6	17.14%	10	28.57%
影片上傳星期	五	15	21.74%	17	24.29%	7	20.00%	5	14.29%
	一	13	18.84%	10	14.29%	2	5.71%	1	2.86%

	六	3	4.35%	3	4.29%	3	8.57%	0	0.00%
	日	4	5.80%	1	1.43%	5	14.29%	6	17.14%
	二	9	13.04%	13	18.57%	3	8.57%	5	14.29%
	三	7	10.14%	7	10.00%	9	25.71%	8	22.86%

附錄 C - 5、The DoDo Men 連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1339.92	1320.97	408.19	613.80	1288.79	1464.79	434.88	764.37
影片按讚數	11769.29	6087.95	4324.05	2702.77	13561.32	5137.74	4479.85	2123.16
影片留言數	863.05	355.61	1287.43	218.34	1126.68	331.21	1745.13	214.61
影片觀看數	518093.50	251357.24	160052.55	56209.48	637108.95	206354.74	142921.65	43733.38
情感特徵平均值	0.45	0.48	0.16	0.18	0.42	0.52	0.13	0.21
情感特徵最大值	4.77	4.84	1.83	2.44	4.65	5.00	1.89	2.85
情感特徵最小值	-1.66	-1.43	1.02	0.78	-1.61	-1.53	0.68	1.00
Tags 數量多寡	34.55	33.45	13.66	15.08	35.26	36.53	12.99	16.23
影片熱門度	0.68	0.33	0.21	0.07	0.84	0.27	0.19	0.06
影片標題長度	34.37	35.32	13.89	12.83	34.00	35.21	13.26	14.79
影片標題詞數	20.50	20.45	7.35	6.08	20.89	20.32	7.80	5.65
影片上傳間隔	7.74	7.68	2.97	0.90	7.79	7.53	3.47	0.61

附錄 C - 6、The DoDo Men 類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	24	38	100.00%	38	100.00%	19	100.00%	19	100.00%
影片上傳早上、晚上	早上	1	2.63%	1	2.63%	1	5.26%	0	0.00%
	晚上	37	97.37%	37	97.37%	18	94.74%	19	100.00%
是否為日常主題	否	37	97.37%	35	92.11%	18	94.74%	16	84.21%
	是	1	2.63%	3	7.89%	1	5.26%	3	15.79%
是否為延伸續集	否	38	100.00%	38	100.00%	19	100.00%	19	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
是否與網紅或專家合作	否	36	94.74%	37	97.37%	18	94.74%	18	94.74%
	是	2	5.26%	1	2.63%	1	5.26%	1	5.26%
是否為實測主題	否	38	100.00%	37	97.37%	19	100.00%	18	94.74%
	是	0	0.00%	1	2.63%	0	0.00%	1	5.26%
影片上傳是否過十四天	否	37	97.37%	38	100.00%	18	94.74%	19	100.00%
	是	1	2.63%	0	0.00%	1	5.26%	0	0.00%
影片上傳星期	六	0	0.00%	1	2.63%	0	0.00%	0	0.00%
	五	1	2.63%	0	0.00%	1	5.26%	0	0.00%
	二	1	2.63%	0	0.00%	1	5.26%	0	0.00%
	三	36	94.74%	37	97.37%	17	89.47%	19	100.00%

附錄 C - 7、安安邊緣子連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1109.25	1206.88	503.64	960.13	1242.21	1426.83	597.30	1288.22
影片按讚數	3143.23	1479.83	1467.57	599.38	4030.25	1109.67	1591.31	501.18
影片留言數	326.00	243.21	201.58	266.27	363.83	230.00	226.89	320.99
影片觀看數	196545.56	77564.25	100046.49	31110.19	257922.33	51651.38	111370.21	22804.06
情感特徵平均值	0.27	0.30	0.13	0.11	0.24	0.32	0.10	0.11
情感特徵最大值	3.19	2.58	1.29	0.98	3.48	2.58	1.47	1.03
情感特徵最小值	-1.39	-0.99	0.79	0.38	-1.61	-0.97	0.94	0.40
Tags 數量多寡	31.81	38.67	13.11	19.03	29.46	37.42	10.45	21.14
影片熱門度	0.64	0.25	0.32	0.10	0.84	0.17	0.36	0.07
影片標題長度	54.90	50.67	23.24	18.05	58.54	50.38	23.66	15.95
影片標題詞數	27.33	26.75	5.02	5.27	28.63	26.71	5.71	6.22
影片上傳間隔	6.56	7.15	2.82	3.41	6.21	7.38	2.30	4.04

附錄 C - 8、安安邊緣子類別值

變數名稱	類別值	寬鬆資料集				嚴謹資料集			
		熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比

影片類別	10	1	2.08%	1	2.08%	1	4.17%	1	4.17%
	20	2	4.17%	0	0.00%	1	4.17%	0	0.00%
	22	30	62.50%	38	79.17%	14	58.33%	17	70.83%
	24	15	31.25%	9	18.75%	8	33.33%	6	25.00%
影片上傳早上、晚上	早上	0	0.00%	1	2.08%	0	0.00%	1	4.17%
	晚上	48	100.00%	47	97.92%	24	100.00%	23	95.83%
是否為日常主題	否	38	79.17%	41	85.42%	20	83.33%	20	83.33%
	是	10	20.83%	7	14.58%	4	16.67%	4	16.67%
是否為延伸續集	否	47	97.92%	42	87.50%	24	100.00%	20	83.33%
	是	1	2.08%	6	12.50%	0	0.00%	4	16.67%
是否與網紅或專家合作	否	21	43.75%	27	56.25%	10	41.67%	16	66.67%
	是	27	56.25%	21	43.75%	14	58.33%	8	33.33%
是否為實測主題	否	48	100.00%	48	100.00%	24	100.00%	24	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	47	97.92%	45	93.75%	24	100.00%	21	87.50%
	是	1	2.08%	3	6.25%	24	100.00%	3	12.50%
影片上傳星期	六	40	83.33%	35	72.92%	0	0.00%	15	62.50%
	日	3	6.25%	4	8.33%	22	91.67%	3	12.50%
	二	0	0.00%	3	6.25%	0	0.00%	1	4.17%
	三	5	10.42%	6	12.50%	0	0.00%	5	20.83%

附錄 C 9、林嘉凌 薔薔連續值

	寬鬆資料集				嚴謹資料集			
變數名稱	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1248.29	944.08	296.08	274.71	1297.48	871.24	315.87	221.07
影片按讚數	8777.86	2502.86	4145.14	1605.81	11014.91	1336.67	4676.08	721.66
影片留言數	604.86	174.09	428.19	132.74	779.24	96.70	506.82	58.29
影片觀看數	500764.91	135510.89	215130.43	72412.97	660812.48	74877.67	198542.22	38170.06
情感特徵平均值	0.49	0.43	0.19	0.10	0.48	0.42	0.20	0.10
情感特徵最大值	5.50	3.57	2.53	1.62	5.69	3.05	2.79	1.32
情感特徵最小值	-2.32	-1.18	1.97	1.09	-2.95	-0.77	2.38	0.48
Tags 數量多寡	4.88	6.71	4.82	6.01	4.85	8.21	3.78	6.28
影片熱門度	1.08	0.29	0.46	0.16	1.42	0.16	0.43	0.08
影片標題長度	48.76	43.12	13.36	14.83	49.94	37.48	14.08	12.40
影片標題詞數	24.94	19.82	6.86	7.99	26.00	17.21	7.50	6.73
影片上傳間隔	5.02	5.55	2.86	3.01	4.82	5.18	2.76	2.64

附錄 C - 10、林嘉凌 薔薔類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比



影片類別	2	0	0.00%	2	3.03%	0	0.00%	2	6.06%
	19	43	65.15%	16	24.24%	19	57.58%	4	12.12%
	22	4	6.06%	23	34.85%	1	3.03%	16	48.48%
	23	1	1.52%	2	3.03%	1	3.03%	2	6.06%
	24	17	25.76%	21	31.82%	11	33.33%	7	21.21%
	26	1	1.52%	2	3.03%	1	3.03%	2	6.06%
影片上傳早上、晚上	早上	53	80.30%	57	86.36%	24	72.73%	31	93.94%
	晚上	13	19.70%	9	13.64%	9	27.27%	2	6.06%
是否為日常主題	否	53	80.30%	51	77.27%	25	75.76%	25	75.76%
	是	13	19.70%	15	22.73%	8	24.24%	8	24.24%
是否為延伸續集	否	66	100.00%	63	95.45%	33	100.00%	31	93.94%
	是	0	0.00%	3	4.55%	0	0.00%	2	6.06%
是否與網紅或專家合作	否	33	50.00%	38	57.58%	18	54.55%	25	75.76%
	是	33	50.00%	28	42.42%	15	45.45%	8	24.24%
是否為實測主題	否	64	96.97%	65	98.48%	32	96.97%	33	100.00%
	是	2	3.03%	1	1.52%	1	3.03%	0	0.00%
影片上傳是否過十四天	否	65	98.48%	65	98.48%	33	100.00%	33	100.00%
	是	1	1.52%	1	1.52%	0	0.00%	0	0.00%
影片上傳星期	五	21	31.82%	13	19.70%	14	42.42%	5	15.15%
	一	4	6.06%	6	9.09%	2	6.06%	3	9.09%
	六	11	16.67%	6	9.09%	6	18.18%	4	12.12%
	日	22	33.33%	16	24.24%	7	21.21%	6	18.18%

	四	6	9.09%	8	12.12%	2	6.06%	5	15.15%
	二	2	3.03%	6	9.09%	2	6.06%	3	9.09%
	三	0	0.00%	11	16.67%	0	0.00%	7	21.21%

附錄 C - 11、油條兄弟連續值

	寬鬆資料集				嚴謹資料集			
變數名稱	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	175.30	184.64	128.15	249.35	175.76	154.32	153.14	146.10
影片按讚數	983.87	179.50	1579.93	102.65	1514.90	105.48	2097.00	48.28
影片留言數	183.93	62.09	120.94	36.11	229.84	42.08	135.16	25.09
影片觀看數	53805.82	8208.41	98447.13	3971.66	86512.48	4729.66	130994.01	2007.35
情感特徵平均值	0.20	0.19	0.10	0.13	0.21	0.16	0.10	0.12
情感特徵最大值	2.09	1.33	0.90	0.69	2.24	1.03	1.00	0.51
情感特徵最小值	-1.03	-0.59	0.66	0.41	-1.16	-0.55	0.79	0.38
Tags 數量多寡	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
影片熱門度	0.58	0.09	1.06	0.04	0.93	0.05	1.41	0.02
影片標題長度	23.29	23.74	9.12	10.71	24.36	22.68	8.70	7.82
影片標題詞數	14.28	15.70	5.39	7.11	14.84	15.32	5.30	5.20
影片上傳間隔	3.35	3.92	1.81	2.56	3.34	3.92	2.03	3.21

附錄 C - 12、油條兄弟類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	24	99	100.00%	100	100.00%	100.00%	50	50	100.00%
影片上傳早上、晚上	早上	98	98.99%	94	94.00%	98.00%	49	47	94.00%
	晚上	1	1.01%	6	6.00%	2.00%	1	3	6.00%
是否為日常主題	否	99	100.00%	99	99.00%	100.00%	50	50	100.00%
	是	0	0.00%	1	1.00%	0	0.00%	0	0.00%
是否為延伸續集	否	99	100.00%	99	99.00%	100.00%	50	50	100.00%
	是	0	0.00%	1	1.00%	0	0.00%	0	0.00%
是否與網紅或專家合作	否	87	87.88%	96	96.00%	88.00%	44	49	98.00%
	是	12	12.12%	4	4.00%	12.00%	6	1	2.00%
是否為實測主題	否	97	97.98%	96	96.00%	98.00%	49	50	100.00%
	是	2	2.02%	4	4.00%	2.00%	1	0	0.00%
影片上傳是否過十四天	否	99	100.00%	98	98.00%	100.00%	50	48	96.00%
	是	0	0.00%	2	2.00%	0.00%	0	2	4.00%
影片上傳星期	五	33	33.33%	35	35.00%	38.00%	19	16	32.00%
	一	9	9.09%	6	6.00%	6.00%	3	3	6.00%
	六	3	3.03%	0	0.00%	0.00%	0	3	6.00%
	日	1	1.01%	3	3.00%	6.00%	3	0	0.00%
	四	9	9.09%	13	13.00%	6.00%	3	5	10.00%

	二	31	31.31%	36	36.00%	30.00%	15	20	40.00%
	三	13	13.13%	7	7.00%	14.00%	7	3	6.00%

附錄 C - 13、孫腫來了連續值

	寬鬆資料集				嚴謹資料集			
變數名稱	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	826.06	621.04	701.38	1275.93	768.05	193.58	226.73	231.66
影片按讚數	1737.36	137.76	4088.82	81.62	3027.68	82.85	5508.12	42.09
影片留言數	308.44	22.69	822.32	26.73	525.65	13.85	1120.20	15.66
影片觀看數	75883.46	4832.30	127232.34	2916.20	133070.48	2403.88	161376.03	1051.56
情感特徵平均值	0.44	0.30	0.25	0.39	0.46	0.30	0.23	0.47
情感特徵最大值	3.74	1.73	2.42	1.29	4.55	1.50	2.81	1.07
情感特徵最小值	-2.04	-0.81	4.24	1.05	-2.96	-0.73	5.72	1.01
Tags 數量多寡	2.88	1.26	1.00	1.65	3.03	0.53	0.80	1.15
影片熱門度	0.45	0.03	0.76	0.02	0.80	0.01	0.97	0.01
影片標題長度	22.79	23.44	7.58	18.99	23.53	25.43	8.37	23.38
影片標題詞數	14.24	14.21	4.70	12.44	14.95	15.33	5.06	15.19
影片上傳間隔	5.46	3.28	2.57	2.40	5.85	2.68	2.52	1.95

附錄 C - 14、孫腫來了類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	24	80	100.00%	80	100.00%	40	100.00%	40	100.00%
影片上傳早上、晚上	早上	77	96.25%	61	76.25%	40	100.00%	33	82.50%
	晚上	3	3.75%	19	23.75%	0	0.00%	7	17.50%
是否為日常主題	否	80	100.00%	76	95.00%	40	100.00%	38	95.00%
	是	0	0.00%	4	5.00%	0	0.00%	2	5.00%
是否為延伸續集	否	80	100.00%	80	100.00%	40	100.00%	40	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
是否與網紅或專家合作	否	76	95.00%	77	96.25%	36	90.00%	40	100.00%
	是	4	5.00%	3	3.75%	4	10.00%	0	0.00%
是否為實測主題	否	80	100.00%	80	100.00%	40	100.00%	40	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	80	100.00%	80	100.00%	40	100.00%	40	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳星期	五	69	86.25%	25	31.25%	36	90.00%	6	15.00%
	一	2	2.50%	7	8.75%	1	2.50%	6	15.00%
	六	1	1.25%	6	7.50%	1	2.50%	1	2.50%
	日	4	5.00%	6	7.50%	1	2.50%	3	7.50%
	四	1	1.25%	11	13.75%	1	2.50%	5	12.50%

	二	2	2.50%	14	17.50%	0	0.00%	11	27.50%
	三	1	1.25%	11	13.75%	0	0.00%	8	20.00%

附錄 C - 15、恩熙俊連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1430.60	1003.24	652.12	273.00	1705.10	984.59	716.95	266.67
影片按讚數	5259.70	2827.76	4039.21	881.31	6998.03	2510.34	5109.01	704.07
影片留言數	459.40	897.88	400.75	1639.39	514.66	922.31	278.51	1852.75
影片觀看數	338028.12	133375.12	339240.69	19765.87	484237.38	117180.28	429970.08	13775.11
情感特徵平均值	0.20	0.20	0.07	0.20	0.20	0.19	0.09	0.14
情感特徵最大值	3.11	3.51	1.88	4.93	3.54	3.64	2.38	5.78
情感特徵最小值	-1.56	-1.36	0.55	0.56	-1.72	-1.39	0.61	0.60
Tags 數量多寡	36.14	35.86	8.31	9.67	35.21	37.03	8.13	9.51
影片熱門度	0.64	0.25	0.64	0.04	0.91	0.22	0.81	0.03
影片標題長度	37.16	30.76	9.37	6.58	39.14	30.62	11.24	6.77
影片標題詞數	19.42	17.60	3.95	3.38	19.97	17.69	3.87	3.78
影片上傳間隔	6.04	5.88	2.00	2.01	6.17	5.86	2.17	2.15

附錄 C - 16、恩熙俊類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比
影片類別	24	57	100.00%	58	100.00%	29	100.00%	29	100.00%
影片上傳早上、晚上	早上	16	28.07%	17	29.31%	8	27.59%	9	31.03%
	晚上	41	71.93%	41	70.69%	21	72.41%	20	68.97%
是否為日常主題	否	45	78.95%	52	89.66%	23	79.31%	25	86.21%
	是	12	21.05%	6	10.34%	6	20.69%	4	13.79%
是否為延伸續集	否	57	100.00%	58	100.00%	29	100.00%	29	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
是否與網紅或專家合作	否	39	68.42%	56	96.55%	14	48.28%	28	96.55%
	是	18	31.58%	2	3.45%	15	51.72%	1	3.45%
是否為實測主題	否	52	91.23%	54	93.10%	25	86.21%	26	89.66%
	是	5	8.77%	4	6.90%	4	13.79%	3	10.34%
影片上傳是否過十四天	否	57	100.00%	58	100.00%	29	100.00%	29	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳星期	六	44	77.19%	47	81.03%	20	68.97%	24	82.76%
	三	13	22.81%	11	18.97%	9	31.03%	5	17.24%

附錄 C - 17、進藤喜彥連續值

變數名稱	寬鬆資料集				嚴謹資料集			
	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1237.38	1056.74	386.89	353.85	1240.65	1011.87	414.01	218.73
影片按讚數	5075.89	2589.63	2379.28	613.65	6234.52	2295.17	2662.26	513.36
影片留言數	347.58	195.61	188.86	60.51	394.52	169.83	184.10	42.07
影片觀看數	259616.51	125545.17	79478.26	21556.45	317709.78	108403.87	71689.68	14821.43
情感特徵平均值	0.51	0.54	0.17	0.18	0.51	0.53	0.17	0.16
情感特徵最大值	4.70	4.17	1.77	1.65	4.91	3.78	1.74	1.26
情感特徵最小值	-1.00	-0.98	0.46	0.64	-0.92	-0.87	0.41	0.47
Tags 數量多寡	12.82	12.89	4.07	5.31	12.96	12.26	4.17	5.33
影片熱門度	0.89	0.43	0.27	0.07	1.09	0.37	0.25	0.05
影片標題長度	46.00	47.63	13.67	14.14	46.30	50.00	11.33	13.79
影片標題詞數	26.40	26.67	6.93	7.60	26.83	27.57	6.00	7.23
影片上傳間隔	7.60	7.26	1.88	1.78	7.70	7.00	1.66	1.65

附錄 C - 18、進藤喜彥類別值

變數名稱	類別值	寬鬆資料集				嚴謹資料集			
		熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比



影片類別	24	45	100.00%	46	100.00%	23	100.00%	23	100.00%
影片上傳早上、晚上	早上	44	97.78%	46	100.00%	23	100.00%	23	100.00%
	晚上	1	2.22%	0	0.00%	0	0.00%	0	0.00%
是否為日常主題	否	35	77.78%	29	63.04%	18	78.26%	13	56.52%
	是	10	22.22%	17	36.96%	5	21.74%	10	43.48%
是否為延伸續集	否	42	93.33%	46	100.00%	21	91.30%	23	100.00%
	是	3	6.67%	0	0.00%	2	8.70%	0	0.00%
是否與網紅或專家合作	否	42	93.33%	45	97.83%	23	100.00%	23	100.00%
	是	3	6.67%	1	2.17%	0	0.00%	0	0.00%
是否為實測主題	否	45	100.00%	46	100.00%	23	100.00%	23	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳是否過十四天	否	45	97.78%	46	100.00%	22	95.65%	23	100.00%
	是	44	2.22%	46	100.00%	1	4.35%	0	0.00%
影片上傳星期	五	1	0.00%	0	0.00%	0	0.00%	0	0.00%
	一	0	6.67%	1	2.17%	2	8.70%	1	4.35%
	六	3	88.89%	1	2.17%	21	91.30%	22	95.65%
	二	40	2.22%	44	95.65%	0	0.00%	0	0.00%
	三	1	2.22%	0	0.00%	0	0.00%	0	0.00%

附錄 C - 19、黃氏兄弟連續值

	寬鬆資料集				嚴謹資料集			
變數名稱	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差	熱門平均值	不熱門平均值	熱門標準差	不熱門標準差
影片時長	1273.97	1453.96	463.38	2153.22	1318.51	1295.28	489.95	902.18
影片按讚數	22834.96	7948.21	11967.43	4107.43	29831.69	5821.10	12733.59	3864.26
影片留言數	1054.89	424.36	921.48	298.13	1443.39	306.87	1137.08	211.02
影片觀看數	776274.56	239444.52	509972.45	110562.47	1052154.48	152344.76	605279.43	82818.78
情感特徵平均值	0.32	0.39	0.17	0.18	0.27	0.40	0.15	0.19
情感特徵最大值	7.90	7.86	5.53	5.27	8.36	7.30	7.09	6.17
情感特徵最小值	-1.96	-1.29	1.66	0.98	-2.34	-1.10	1.95	1.07
Tags 數量多寡	30.87	23.37	9.51	8.86	32.49	20.77	10.33	9.18
影片熱門度	0.35	0.11	0.23	0.05	0.48	0.07	0.28	0.04
影片標題長度	56.96	49.87	18.48	20.76	59.03	42.86	20.82	21.40
影片標題詞數	30.23	27.14	5.95	8.16	30.48	24.63	6.84	9.61
影片上傳間隔	2.80	2.60	0.85	1.17	2.96	2.51	0.84	1.31

附錄 C 20、黃氏兄弟類別值

		寬鬆資料集				嚴謹資料集			
變數名稱	類別值	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比	熱門出現次數	熱門百分比	不熱門出現次數	不熱門百分比

影片類別	10	3	2.11%	1	0.70%	3	4.23%	1	1.41%
	24	139	97.89%	141	99.30%	68	95.77%	70	98.59%
影片上傳早上、晚上	早上	142	100.00%	129	90.85%	71	100.00%	60	84.51%
	晚上	0	0.00%	13	9.15%	0	0.00%	11	15.49%
是否為日常主題	否	132	92.96%	133	93.66%	67	94.37%	65	91.55%
	是	10	7.04%	9	6.34%	4	5.63%	6	8.45%
是否為延伸續集	否	135	95.07%	140	98.59%	68	95.77%	70	98.59%
	是	7	4.93%	2	1.41%	3	4.23%	1	1.41%
是否與網紅或專家合作	否	82	57.75%	92	64.79%	36	50.70%	53	74.65%
	是	60	42.25%	50	35.21%	35	49.30%	18	25.35%
是否為實測主題	否	142	100.00%	141	99.30%	71	100.00%	71	100.00%
	是	0	0.00%	1	0.70%	0	0.00%	0	0.00%
影片上傳是否過十天	否	142	100.00%	142	100.00%	71	100.00%	71	100.00%
	是	0	0.00%	0	0.00%	0	0.00%	0	0.00%
影片上傳星期	五	36	25.35%	18	12.68%	26	36.62%	8	11.27%
	一	36	25.35%	44	30.99%	13	18.31%	23	32.39%
	六	27	19.01%	21	14.79%	18	25.35%	10	14.08%
	日	2	1.41%	7	4.93%	1	1.41%	3	4.23%
	四	2	1.41%	8	5.63%	1	1.41%	7	9.86%
	二	3	2.11%	5	3.52%	2	2.82%	3	4.23%
	三	36	25.35%	39	27.46%	10	14.08%	17	23.94%