

中華民國第 62 屆中小學科學展覽會  
作品說明書

---

高級中等學校組 行為與社會科學科

第三名 探究精神獎

052701

基於神經網路分析封城對民眾攻擊性情緒之影響：  
以美國為例

學校名稱：臺北市立建國高級中學

作者： 高二 許澤厚	指導老師： 周麗芬 潘威歷
---------------	---------------------

關鍵詞：攻擊性情緒、大數據分析、封城

# 摘要

疫情期間各地家暴、離婚負面社會現象顯著增加，而攻擊性情緒可能是造成此現象的重要原因之一。如何快速掌握人們情緒變動，提供有效配套措施，為當務之急。本研究利用人工智慧神經網路模型，訓練其分辨攻擊性情緒，快速分析大量社群媒體資料。以此模型分析全美各地在不同時間發出的推文，即時判斷疫情下各地攻擊性情緒的變化，發現疫情開始後，整體攻擊性情緒急劇增加，且封城下的民眾相較未封城者呈更高的攻擊性情緒。再來，本研究以固定效應模型的統計大數據分析，證實封城與攻擊性增加之因果關係，為理解疫情中人們攻擊性情緒變化的成因作出重要貢獻。提供心理學或精神醫學大範圍、快速、低成本的研究方法，以快速掌握人們情緒模式，研擬對策。

## 壹、前言

### 一、背景簡介

在疫情時代之下，一個地區封城與否的決定，對於對抗疫情至關重要。然而，長期的居家生活，除了可能造成身體上的不適，對心理上的影響亦不容小覷。在封城的政策下，許多的活動被迫取消，人們必須待在家中，隨之而來的便是深刻的挫折與孤寂感(Killgore et al., 2021; Muñoz-Fernández & Rodríguez-Meirinhos, 2021; Shah et al., 2020)。而這些挫折感，很可能透過攻擊性的形式展現出來(Breuer & Elson, 2017)。封城中，人們必須與自己的家人近距離的接觸，導致此時的攻擊性，更容易造成像是離婚，家暴等等摧毀性的行為，對社會造成立即的負擔(Moreira & da Costa, 2020; Sher, 2020)。在近期 Jinan Usta et al.(Usta et al., 2021) 的研究指出，封城中女性受到家庭暴力的次數劇增。黎巴嫩的社福機構接到性別暴力的投訴電話是疫情前的兩倍；在英格蘭以及蘇格蘭，Manuel Serrano-Alarcón et al.(Serrano-Alarcón et al., 2022) 從問卷發現人的心理健康與封城與否是有關係的，同樣在得病的威脅下，提早解封的英格蘭人比起蘇格蘭人擁有更好的精神健康狀況；BBC 的新聞報導(Savage, 2020)也指出，美國比起去年同期離婚率增加了 34%；在世界各地如中國、英國與瑞典也發現了相似的趨勢。疫情下攻擊性情緒，是全體人類必須共同正視，並解決的問題。

### 二、相關文獻

## （一）封城下的負面社會現象的心理研究

從個人層面來看，研究顯示封城中人們的自殺傾向、焦慮症、惡夢、憂鬱症、孤寂感皆有增加(Groarke et al., 2020; Majumdar et al., 2020; Marroquín et al., 2020; Niedzwiedz et al., 2021; Pesonen et al., 2020; Wang et al., 2020)。從社會的角度，封城中的負面社會現象包含離婚、性暴力與家暴(Sanga & McCrary, 2020; Usta et al., 2021)。這些所有的負面社會現象，都是封城下亟待解決的問題。

有心理研究嘗試比較封城前後的精神狀況差異，並發現封城後人們精神狀態的惡化(Niedzwiedz et al., 2021)。然而因為傳統心理問卷方法採樣效率的限制，此研究僅能採樣封城前後兩個時間點的資料。因缺乏不同時間點下的資料，因此難以歸咎造成這些結果的原因。

William Killgore 等人所著的 *Increasing Aggression During Covid 19 Lockdowns* 的論文中(Killgore et al., 2021)，該文作者透過問卷調查，發現人們在封城中的攻擊性有增加的趨勢。他們在 2020 年四到九月，每個月分別對 1000 名左右的受試者，六個月總共超過 6000 名受試者，進行心理學問卷 BPAQ( Buss Perry Aggression Questionnaire) 調查。這篇研究顯示，言語攻擊性、肢體攻擊性、憤怒以及敵意，四種攻擊性傾向在封城一段時間後，封城中的人們都顯著的比沒有封城的人們高。然而，因為問卷研究資料量有限的情況下，此研究也僅能觀察出各種攻擊性情緒在封城中的變化，並未能分析出因果關係。再者，因為問卷研究執行上的限制，他們必須在每次採樣時更換受試者，造成更大的誤差。

## （二）人工智慧自然語言處理相關研究

現在新興的人工智慧自然語言處理科技，可以快速的分析文本，理解語意，達到幫助問卷研究的作用。過去有些研究透過自然語言處理 (natural language processing, NLP) 針對 Twitter，探討封城前後的特定詞彙出現的頻率變化來理解封城對於人們情緒的影響(Su et al., 2020)。然而，若純粹研究字詞出現的頻率，不看單字出現時的前後文，則無法像神經網路模型一樣理解句子的情緒含義(Catelli et al., 2022; Kotelnikova, 2020)。而其他利用神經網路技術的研究，專注在訓練模型理解語意，也並未將此技術推廣至社會科學實證研究的領域(Hu et al., 2021; Jang et al., 2022; Ridhwan & Hargreaves, 2021; Shamrat et al., 2021)。

### (三) 目前問題

傳統的問卷心理學研究存在許多缺陷。首先，因為尋找受試者的困難，他們僅能以月為單位進行研究，資料的精細度大大的侷限。再者，問卷調查能觸及到的人也有限，無法有效率地做大規模的調查。不只如此，時間與金錢上的消耗也相當可觀（例如 Killgore et al 花了半年，每個月尋找受試者，給予金錢補償）。因為疫情來得突然，傳統問卷調查無法取得疫情發生以前的資料作為對比，因此並未分析出封城與攻擊性增加之間的關聯，僅觀察比較了 2020 年中封城與未封城下受試者的情緒變化。

在自然語言處理方面，目前的研究大多專注在優化模型理解大量數位語料語意的過程（表 1）。然而，現今的研究顯少將自然語言處理技術導入社會科學與統計實證當中，兩個領域間存在一定的距離。跨領域的研究雖有廣大的可能性，其中卻從在許多需要克服的難關，必須仰賴實際的實作與研究，才能繼續向前。

表 1 各類文獻比較與統整，呈現本研究的貢獻價值與創意性。此表只是舉例統整，並未列出所有文獻。

研究探討面向 文獻資料種類	研究疫情下的 各種情緒	利用人工智慧 理解大量文獻 中的情緒	研究封城與各 種負面社會現 象的相關性	利用統計模型 與大數據判斷 封城與負面社 會現象的因果 關係
(Killgore et al., 2021)	√	x	√	x
(Su et al., 2020)	x	x	√	x
(Niedziedz et al., 2021)	x	x	√	x
(Majumdar et al., 2020)	x	x	√	x
(Hu et al., 2021)	√	√	x	x
(Ridhwan & Hargreaves, 2021)	√	√	x	x
(Jang et al., 2022)	√	√	x	x
(Sanga & McCrary, 2020)	x	x	√	√
本研究探討問題	√	√	√	√

### 三、研究目標與貢獻

就以上問題，本研究的核心目標與貢獻可以分為四個部分。

1. 訓練出可以即時辨別推特上攻擊性情緒的神經網路模型。
2. 以神經網路模型分析推特資料，取得不同地點與封城狀態的攻擊性情緒波動。
3. 分析出且提供一個比問卷研究資料時間刻度更細，地理範圍更廣，數量更大，且更容易

取得的情緒資料集。

4. 用大數據分析深刻了解封城時間長度與大眾攻擊性增加之間的關聯，提供政策執行者有利的參考。

## 貳、研究設備及器材

### 一、硬體：

- (一) CPU: MacBook Pro
- (二) GPU: GeForce GTX 1080

### 二、軟體：

- (一) 工作環境：Jupyter Notebook, Google Colab
- (二) 深度學習模組：Pytorch
- (三) 機器學習整理軟體：Weights and Biases
- (四) Python 套件: Tweepy, Matplotlib, Pandas
- (五) 統計分析軟體：STATA 17.0.

## 參、研究過程及方法

### 一、推特資料的爬文與採樣

美國每一個州都有高度的行政獨立性，因此封城時間不盡相同，州內的政策也有高度的統一性，因此，本研究以州為單位來取得資料。推特用戶的資料中，有些會標注其所在的州與城市。仿照臨床資料的受試者概念，本研究從推特隨機採樣（自行使用 Tweepy 套件爬文）11,455 名的受試者（這些受試者在 Twitter 帳號上面有透露其所在地區的資訊）作為研究的對象<sup>1</sup>。取得受試者樣本後，依照 Twitter API（自行申請 Twitter 開發者套件）對各個公開用戶爬文的功能，

---

<sup>1</sup> 推特對於每一位研究者帳號，若向推特申請提出申請，可利用 Twitter API 伺服器取得公開推文。只需以約 50 行程式碼及推特提供給研究者的密碼即可。

取得受試用戶 2019 年至 2020 年疫情爆發川普總統宣佈國家緊急狀態後六個月所有的推文文字。此資料共計 1,281,362 筆。其中，被採樣的用戶來自美國的 50 個不同的州。在接下來的研究當中，本研究利用各州不同的封城狀態資料分析攻擊性與封城之間的關聯性。

本研究在推特隨機採樣的使用者來自全美各地，圖 1 顯示各州的人口數、採樣人數與資料總數，可看見各州採樣的使用者的地理分佈大致上實際人口相似。為了解推文中最初的情緒，避免任何人工偏誤，因此本研究納入所有受試者發出的推文。

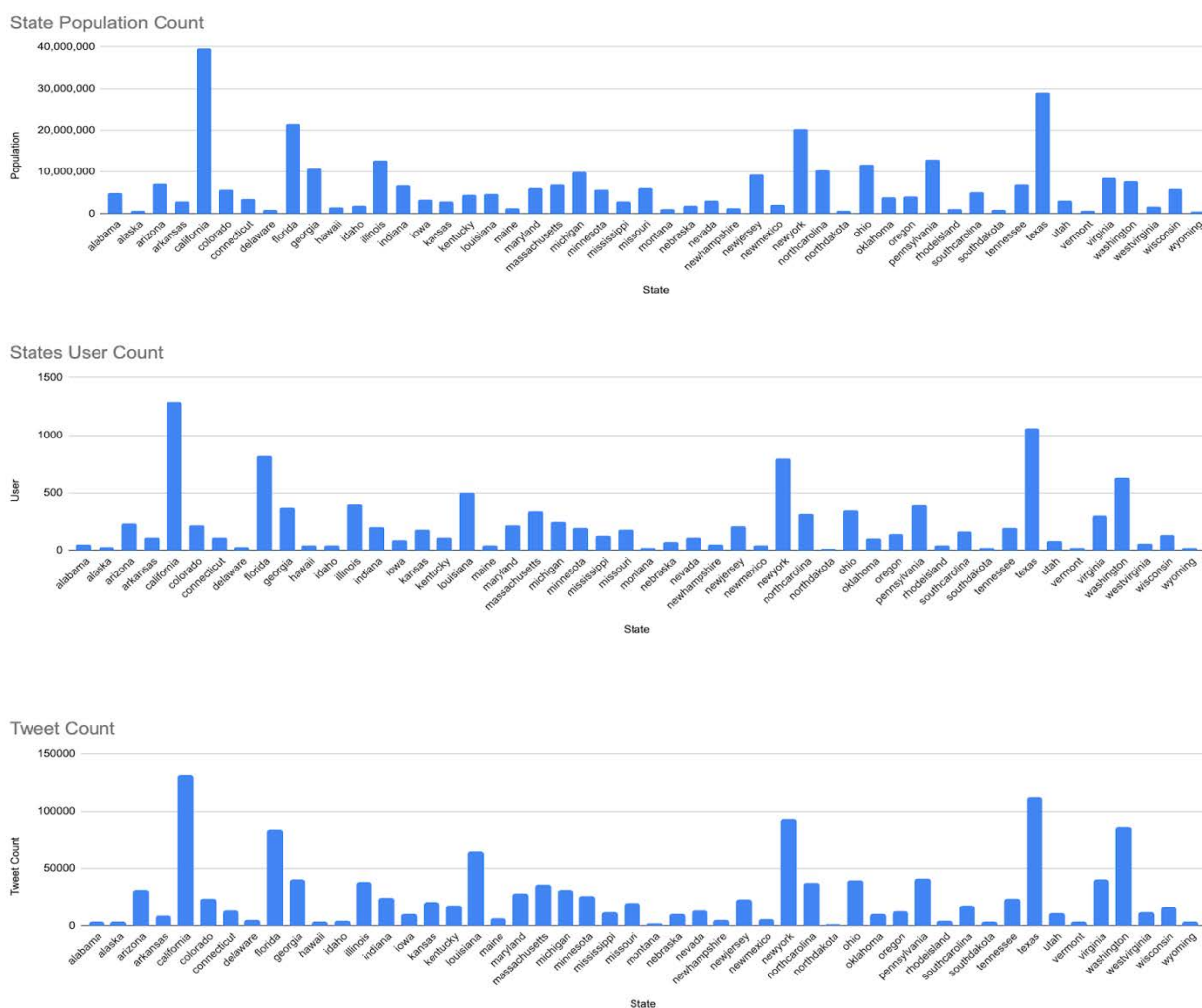


圖 1 研究資料樣本分布，圖中橫軸為照字母排序的美國各州名稱，縱軸上方為各州人口數量，中間為研究採樣的用戶數量，下方為採樣用戶從 2019/1/1 到 2020/10/30 發文的數量。

## 二、 訓練資料的搜集與整合

考慮到每一篇推文 (Tweet) 可以擁有多種情緒，此研究對於每一種情緒都分別分析訓練資料，分別訓練模型分類器。以下介紹不同情緒的資料收集過程。攻擊性情緒造成社會影響劇烈，因此本研究將三種網路常見攻擊性情緒作為分析的對象(表 2 及表 3)。

### (一) 憤怒 (Anger) 訓練資料集

本研究選用 Go Emotions 資料集來訓練辨別憤怒的情緒分類模型(Dorottya Demszky, 2020)。作為最大的人工標註的情緒資料集之一，Go Emotions 包含 58000 筆 Reddit 上的語句。對於每一個語句，Go Emotions 都標註了其是否和有包括憤怒在內的 28 種常見情緒。本研究從其中選取的 6000 筆標註好的語句來訓練本研究的二元分類模型，另外選取了 1000 筆資料用於測試模型的表現。根據 Go Emotions，憤怒定義為「強烈的不悅或敵意」。

### (二) 攻擊性語言 (Offensive Language) 資料集

攻擊性語言是包含不當字眼且可能傷及弱勢族群的言論。我們選用“Automated Hate Speech Detection and the Problem of Offensive Language”(AHSD) (Davidson et al., 2017)這個資料集作為訓練資料。此資料集裡面包含 24,802 筆人工標註的推文。本研究從其中選取 7750 筆作為訓練資料，以及 613 筆作為測試資料。本研究有將攻擊性語言(Offensive Language) 與仇恨言論(Hate Speech) 做出區隔。(例如，許多青少年說髒話時可能是以一種嬉鬧態度。在本研究中，只要有不當字眼且由讓人冒犯之虞，都可以算是攻擊性語言，然而，仇恨言論是更蓄意且針對性的。)

### (三) 仇恨言論 (Hate Speech) 訓練資料集

仇恨言論定義為「用來羞辱，辱罵，貶低特定族群的公開言論」。欲訓練仇恨言論的分類模型，本研究使用 AHSD 資料以及“Large Scale Crowd Sourcing”(Founta et al., 2018)來做訓練。本研究將兩資料集合併訓練，總共有 6450 筆訓練資料與 639 筆測試資料。兩資料集對仇恨言論的定義是相同的。

表 2 本研究中攻擊性情緒之定義整理

情緒	來源	定義
憤怒 Anger	(Dorotyya Demszky, 2020)	強烈的不悅或敵意  A strong feeling of displeasure or antagonism.
攻擊性語言 Offensive Language	(Davidson et al., 2017)	包含不當字眼且可能傷及弱勢族群的言論  Speech that contains unacceptable language (Profanity)  And is potentially harmful to a disadvantaged group.
仇恨言論 Hate Speech	(Davidson et al., 2017; Founta et al., 2018)	用來羞辱，辱罵，貶低特定族群的公開言論  The language that expresses hatred towards a targeted group  And is intended to be derogatory, insult, and humiliate.

表 3 各種情緒範例推文

情緒	範例推文
憤怒	<i>I hope this all ends soon. This is hell</i>
攻擊性語言	<i>Are people really this stupid?</i>
仇恨言論	<i>@user The rot starts from the top.....Trumps wankers are all racist.....F*ck them all!</i>



### 三、 機器學習模型介紹：BERT

最早期也最直觀的情緒分類方法，即是用文字出現的頻率對應相關的情緒，此方法稱作 Bag of Words。然而，這個做法的缺陷，就是無法了解文字之間的關係所構築出的含義。近年來神經網路模型 (Neural Network Models) 的發展克服了這樣的缺陷。神經網路模型可以想像成諸多的資料點為神經元，中間有不同權重的邊相連，定義兩個神經元間的重要或相關程度，邊的權重就是模型所謂的「參數」，其決定了送入的資料如何被處理。Google 在 2018 年發表的 BERT ( Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018) 是目前理解自然語言廣泛使用的模型。BERT 是由 Transformer 組成。Transformer 是由編碼器 (Encoder) 與解碼器 (Decoder) 組成，編碼器將得到的資訊化成特徵值，傳送給神經元，而解碼器解讀神經元輸出的資訊，得到最終想要的輸出。BERT 不同於先前的神經網路模型的地方在於，它預先使用大量的文句資料（整個英文維基百科全書），做了克漏字與語句預測，使其對於語言具備基本的認知。相較於先前的遞迴神經網路 (RNN) 只讓模型在看完句子後預測接續的字詞，BERT 的克漏字的訓練讓模型能夠理解整篇文章的背景與先備知識。BERT 在預訓練 (Pre-Training) 過後，固定了基礎模型的參數，對於語言有了基本的認知。往後在訓練時可以依照各實驗的特定需求，加上少量新的參數，不需要調整原始模型的參數，使訓練的時間與成本顯著降低，此步驟稱為 Fine-Tuning。

本研究當中，我們選用 BERT 家族當中，不分英文大小寫的 bert-base-uncased 模型。鑑於 BERT 模型優異的能力，符合我們研究的需求，因此我們目前以此模型作為情緒分析的模型。

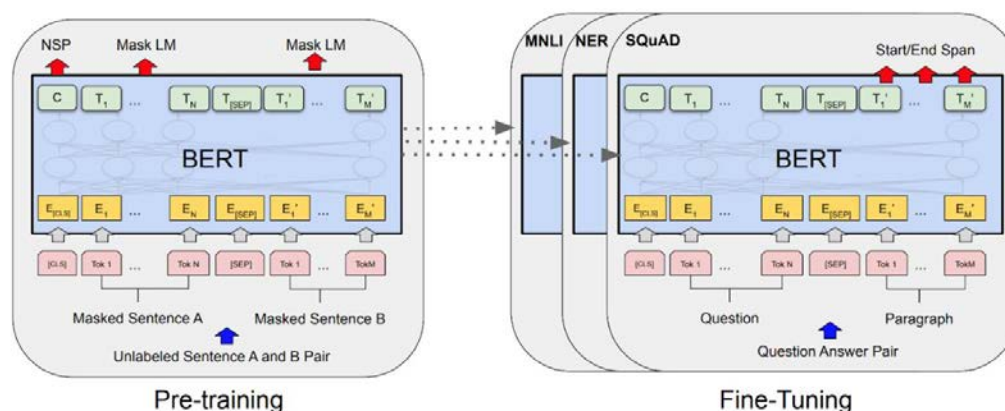


圖 2 BERT 模型概念圖

## 四、 模型訓練過程

### (一) 背景知識

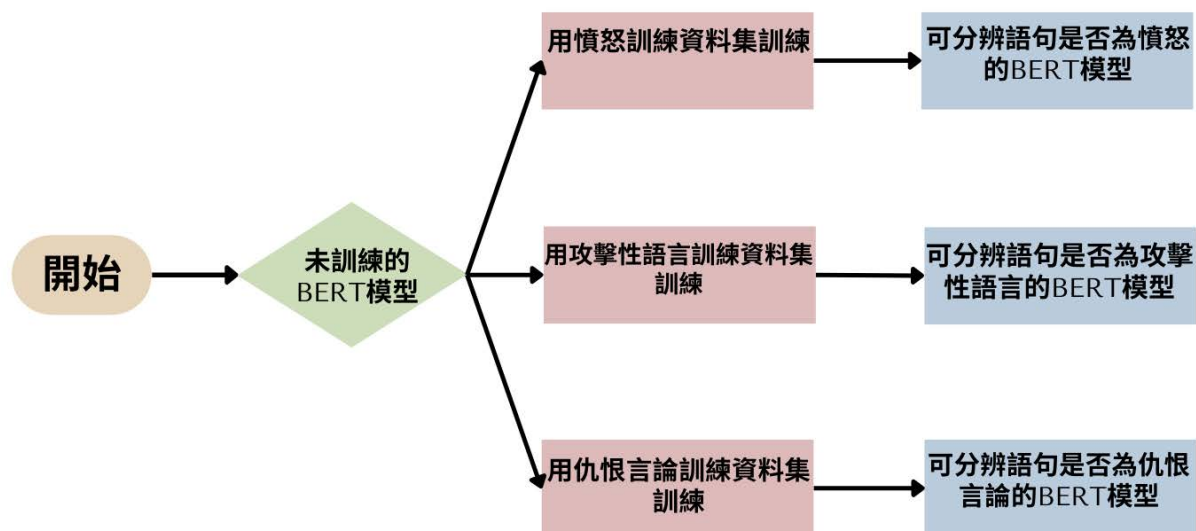


圖 3 訓練模型流程圖（不包括優化細節）

在深度學習中，最重要的目的即是藉由讓模型學習大量標注過特定情緒的資料，訓練其分辨該情緒的能力。本研究利用此流程訓練上述 BERT 神經網路模型分辨文字中的情緒(圖 3)。在每一個訓練回合當中，會算一個損失函數 (Loss Function)，若模型越不準確，則損失函數值越高。訓練的過程便是藉由調整參數，使損失函數值降低，提高模型準確度。訓練模型的最終目的就是找到損失函數發生最小值時的模型參數。在這個過程當中，訓練參數 (Hyperparameter)，也就是不包含模型內神經元的權重參數，而是用於指示模型學習的參數，亦會影響訓練成果。例如 Learning Rate 與 dropout layer 等等，都包含在訓練參數當中。因此，在訓練模型的第一步，便是找到適合的訓練參數。在訓練的過程中，必須將資料集分成訓練資料集 (Training Set) 與測試資料集 (Testing Set)，在模型由訓練資料集學習後，測試資料集可測試模型在接觸首次見到的文章時的表現狀況。

## (二) 訓練參數的尋找過程

訓練參數 (Hyperparameter) 即是模型內部神經網路的參數以外，決定機器學習環境的各種外在參數，例如學習速率、一次要看多少資料。訓練參數對於每一個任務皆會相異，且因為理解自然語言的複雜程度極高，因此難以直接推算出最適合的訓練參數，只能透過模型不斷的自動嘗試來找到適宜的訓練參數。以下兩張圖(圖 4 及圖 5)是我利用 Weights and Biases 機器學習軟體，自動嘗試各種不同訓練參數在訓練資料集的過程。

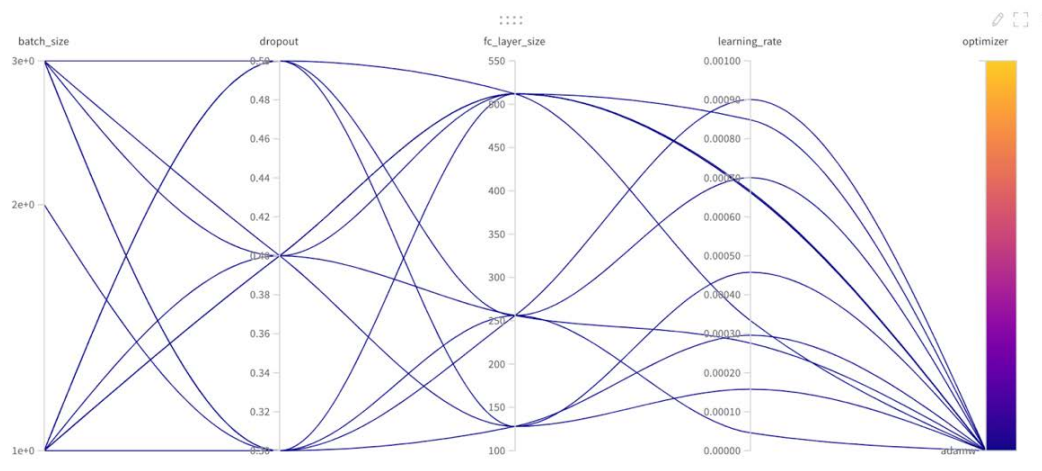


圖 4 不同訓練組合的嘗試

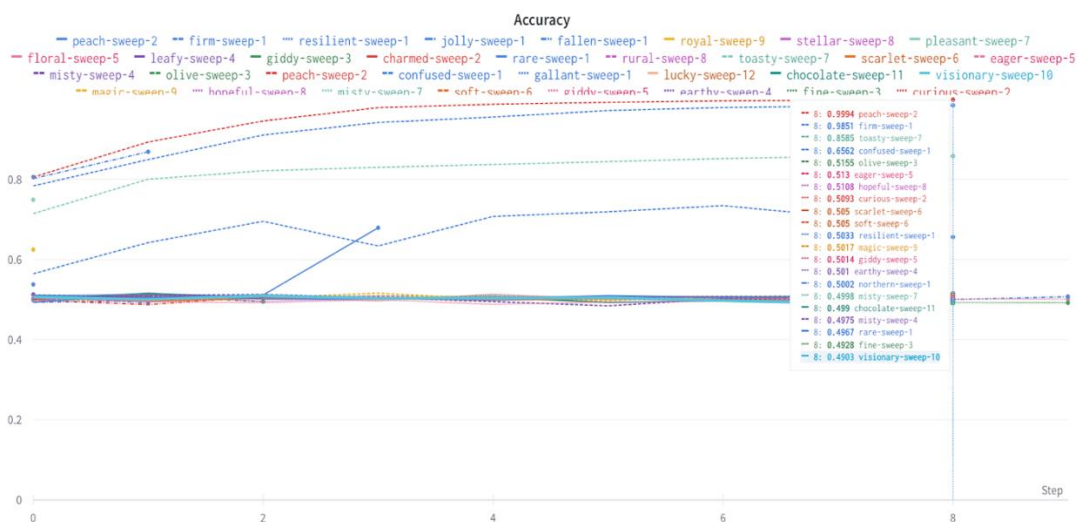


圖 5 不同訓練參數的準確度變化情形，圖中橫軸為訓練回合數，縱軸為模型準確度，其中每一條線都是一次訓練模型的嘗試。

由圖 5 可以看出，有些訓練參數使模型無法有效學習，不論訓練多少回合(Epoch) 準確率都只有 50%左右。有些訓練參數訓練出來的模型，準確率隨著回合數增多穩定增加。訓練參數的目的即是讓模型以適當的速率學習，達到高的準確性及穩定性即可，並沒有絕對的標準。經過嘗試，將適宜的訓練參數用於訓練模型。

### (三) 模型的訓練成果

此部分對於經過訓練資料集訓練的模型進行兩階段的檢驗。第一階段檢驗其在測試資料集上的表現，而第二階段檢驗其在我們從 Twitter 採樣的資料集的表現。因為分類情緒的神經網路模型是基於訓練資料集訓練的，因此可以預期其準確度在跟訓練資料集來源相同的測試資料集上會很好，而更重要需要檢驗的是模型是否可以準確的預測本研究從推特上取得的資料中的情緒。我們為了嚴謹驗證模型的好壞，因此由本研究採樣的推特資料集中，隨機抽取了 1080 筆資料，請兩位標註者依據本研究對三種攻擊性情緒的定義手動標註語句當中是否含有憤怒、攻擊性語言、以及仇恨言論的情緒。手動標注好的資料首先進行標註者間的共識度檢定，以確保手動標注的穩定及正確性（表 5），再用模型預測此 1080 筆資料的情緒，進行模型準確性的驗證。Cohen Kappa 值是檢定標註者間共識度的常用指標(McHugh, 2012)，其數值介於 0 到 1 之間，大於 0.8 即可視為極高的共識度。

機器學習中，要檢驗一個模型的好壞，除了大眾熟知的準確率，學界也常用一個很重要的指標：F1 Score。F1 是將精確率 (Precision) 和召回率(Recall) 做調和平均的結果，只考慮真實陽性與偽陽性錯誤及偽陰性錯誤之間的比率，避免陽性（具有攻擊性）的推文比例很低時，準確率形成誤判(表 4、表 6 與圖 6)。

表 4 模型在測試資料集下的表現

情緒類型	準確性參數	模型好壞參數 F1 score
憤怒	0.851	0.847
攻擊性語言	0.975	0.970
仇恨言論	0.938	0.933

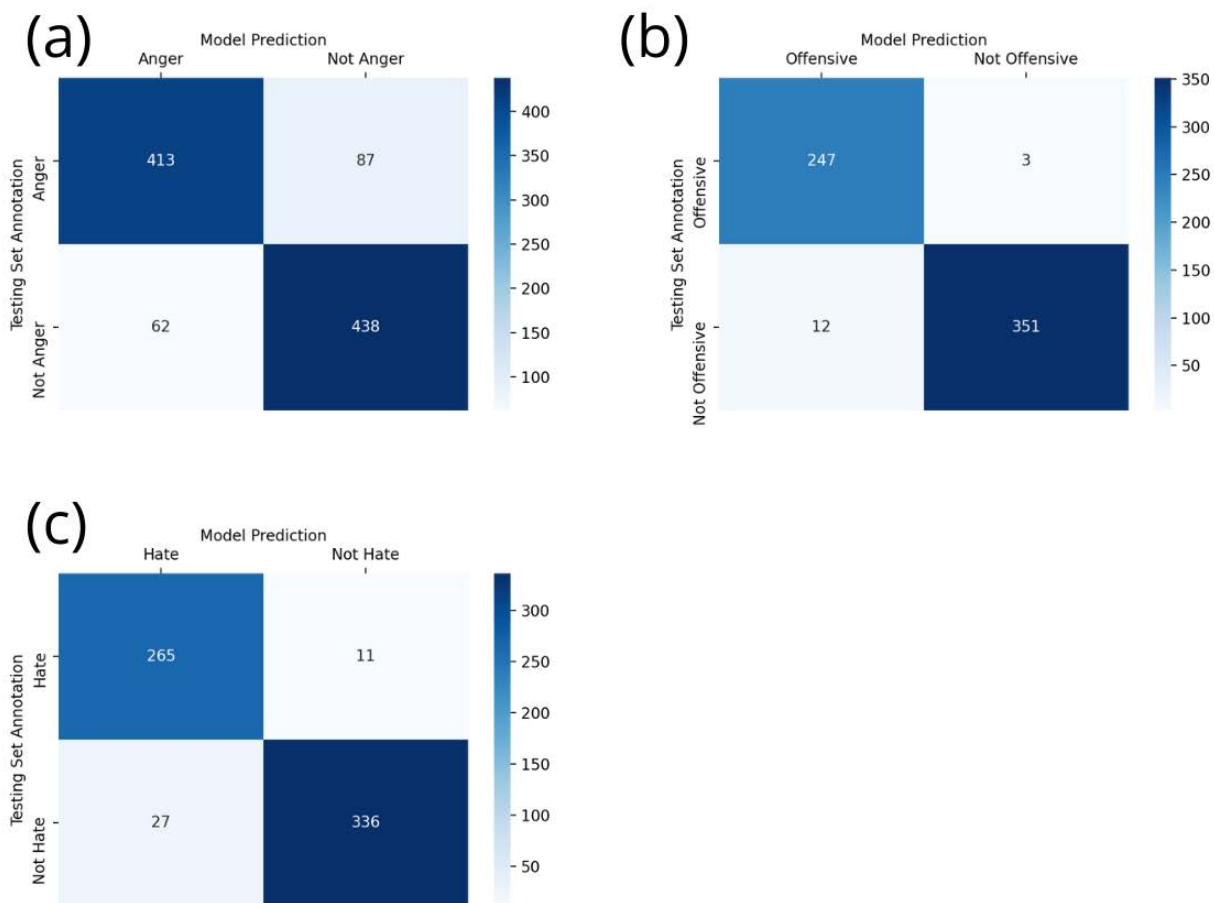


圖 6 在測試資料集上的模型判斷矩陣，左上和右下分別是模型判斷正確的部分，顏色越深代表越高的準確率，此圖為三個模型分類三種攻擊性情緒的結果 a：憤怒、b：攻擊性語言、c：仇恨言論

表 5 標註者之間的共識程度

情緒類型	共識度指標 Cohen Kappa <sup>2</sup>
憤怒	0.928
攻擊性語言	0.937
仇恨言論	0.890

<sup>2</sup> Kappa 值的計算方式如下  $k = \frac{P_o - P_e}{1 - P_e}$  其中， $P_o$  為觀察機率， $P_e$  為期望機率

表 6 模型在本研究取樣的推特資料上的表現

情緒類型	準確性參數	模型好壞參數 F1 score
憤怒	0.937	0.839
攻擊性語言	0.985	0.880
仇恨言論	0.987	0.839

由嚴謹的檢驗可以知道，我們的情緒分類模型不但在原始訓練的資料上能有好的表現，就算到未看過的推文當中，也可以準確的預測其情緒（表 4 與表 6）。

## 五、 資料分析方法

我們利用訓練好偵測三種攻擊性情緒的神經網路模型逐一分析推文是否含有三種攻擊性情緒。分析好後對於每種攻擊性情緒即有一個顯示地理及時間情緒分佈資料集。如此一來，我們用有了各州各星期攻擊性推文的數量。然而，數量的變動容易被當次採樣總數影響，為了瞭解推文當中的攻擊性強度，本研究計算含有攻擊性的推文佔整體隨機採樣的推文的的比例。此資料集中，在美國的每一州當中，對於從 2019 年 1 月至 2020 年 10 月的 92 個星期，都計算該種攻擊性情緒所佔當州當週推文的的比例。

### （一） 即時了解疫情下封城與未封城中人們情緒變化

本研究的神經網路模型，可以即時大量的分析社群媒體上的情緒，此部分研究呈現封城與未封城的人們攻擊性情緒的即時變動。我們採週為單位做分析，預先計算每一州於每一週的攻擊性推文百分比，再確認美國每一個州是否為封城狀態，將封城與未封城的州分成兩組，將兩組數據分別總和，以此兩組的數據代表封城與未封城人們的攻擊性變化。由圖表呈現 92 周內不同地區人們即時情緒波動。

### （二） 觀測有封城的州當中，封城前後的攻擊性情緒變化

相較前一個部分比較同時間封城與未封城的人們攻擊性情緒的變化，這部分分析著重在封城前後的情緒差異。比較封城前後可以明顯看出封城手段介入後，造成的影響。因為每一州封城開始的時間並不相同，因此我們以封城開始的那一週為基準對齊每個州的資料。更精確的來說，我們將封城開始的那一週定為第 0 週，封城後一週為第 1 週，封城前一週為第 -1 週，以此類推。利用這樣的分析，我們可以精確知道封城開始後攻擊性情緒如何變化。目前分析尚在視覺化的階段，尚未利用統計學方式證明封城與攻擊性情緒的因果關係。實際因果關係的驗證會利用嚴謹的統計方式證明。

### （三） 利用統計模型驗證封城與攻擊性情緒的因果關係

要在真實世界大數據當中分析因果關係，相較於實驗室中的數據，必須克服無法隨機選取控制組與對照組的問題。

隨機實驗是最傳統的因果關係分析方法，其中，將受試者隨機分成兩組，一組為實驗組，一組為控制組。然而，在真實世界的大數據當中，往往因可行性或倫理因素的影響，無法將隨即分配對照組與實驗組。（例如，在本研究中，即無法隨機抽取 5000 位受試者對其施於封城政策。）當實驗者想要知道欲觀察因子  $X$  對應變變因  $Y$  的影響，可能有許多相關因子  $Z_1, \dots, Z_k$  影響，而在控制組與實驗組無法隨機選取時，這些  $Z$  變數在兩組之間的差異，就會造成更大的誤差。

有些社會科學家利用多變數統計回歸( Multivariate regression) 嘗試解決這樣的問題。多變數統計回歸當中，由欲觀察的因子  $X$ ，以及其他相關的影響因子  $Z_1, \dots, Z_k$  一同預測應變變因  $Y$ 。雖然此方法可以考慮選取的  $Z$  變數的影響，然而，真實世界中可能造成影響的  $Z$  變數實在太多了，要一一找出並取得  $Z$  變數的相關資料是近乎不可能的事。

計量經濟學中的差異中之差異 (Difference in difference, DID) 可以是一個解決此類問題的好方法。DID 比較控制組與實驗組在事件前  $T_1$  時間點與事件後  $T_2$  時間點的  $Y$  值。觀察兩組數據在  $T_1$  的差異  $D_{T_1}$  以及在  $T_2$  的差異  $D_{T_2}$ ，透過分析  $D_{T_1}$  與  $D_{T_2}$  的差異，即可看到在考慮到控制組與對照組本身差異的情形下，事件造成的影響。換言之，紛雜的  $Z$  變數的確會造成控制組與實驗組的差異，然而，這些  $Z$  變數的影響已經顯現在事件前控制組與實驗組的差異  $D_{T_1}$  當中，比較  $D_{T_1}$  與  $D_{T_2}$  的差異，顧名思義「差異中之差異」。DID 因為固定了各種  $Z$  變因的影響，即可逼近

客觀的因果關係推論。在時間序列的資料 (Time series data) 當中 (如本研究連續 92 週的攻擊性資料)，並不只有 T1 和 T2 兩個時間點，可利用固定效應模型 (Fixed effect Models) 實現 DID。固定效應模型是在回歸模型的基礎上控制各個時間 (如星期) 與地理 (如州) 資料群 (Data group) 的平均。每個資料群的平均可以捕捉各種雜亂的 Z 變數的影響，如影響一個州的數據平均的因素，可能包含其人民的種族組成、政治傾向、次文化等等。透過減去平均，即可控制雜亂 Z 變數的影響，觀察 X 對 Y 的淨影響。固定效應模型比起傳統統計回歸，「更能達到客觀衡量因果關係的功效」(Kempf-Leonard, 2004)。

固定效應必須建立在統計回歸上。因為本研究處理的是一段時間、空間下攻擊性推文的數量，是一種計數資料，因此選用專門處理計數資料的 Poisson 回歸搭配固定效應來實現作 DID 精神的因果推論(Hilbe, 2014)。Poisson 回歸是廣義線性回歸的一種，與一般的線性回歸不同的點是其在應變變因上加入自然對數函數，更能有效契合計數資料。最初式子如下：

$$\ln(\text{Aggressive Tweet Count}_{s,t}) = \alpha_0 + \alpha_1 \text{under lockdown} + \mu_{state} + \sigma_{time} + \varepsilon$$

$\text{Aggressive Tweet Count}_{s,t}$  是某個州 (State, 簡寫 s) 在某個時間 (Time, 簡寫 t) 下的攻擊性推文數量。 $\alpha_0$  是標準統計模型當中的常數項。 $\alpha_1$  就是封城狀態 *under lockdown* 對於攻擊性推文數量影響的係數。*Under lockdown* 為一個二元變數，同時也是本實驗欲測定的解釋變數 (Explanatory Variable)，在未封城時其值為 0，在封城時其值為 1。 $\mu_{state}$  和  $\sigma_{time}$  分別是空間與時間的控制效應 (如前段所述)，其值域並未有明確限制，視其代表的資料群而定 (例如，在某一州當中的平均，或某一星期的平均)。 $\varepsilon$  即為誤差項，是所有統計模型都具備的。

然而，就如前面所述，攻擊性推文數量的增加容易受總數的影響，要知道攻擊性強度，就必須考慮總數，也就是計算攻擊性推文數量與採樣推文總數的比例。為了用 Poisson 計數模型達成比例的計算，本研究將每週每州的攻擊性推文總數的對數值  $\ln(\text{Total}_{s,t})$  加在回歸式子後方，不影響回歸的進行，卻能考慮總數。計量經濟學上，此方法正是設計給本研究這類的狀況，並且實作可以由 STATA 的 *exposure()* 功能支援 (StataCorp., 2021)。最終統計回歸式如下：



$$\ln(\text{Aggressive Tweet Count}_{s,t}) = \alpha_0 + \alpha_1 \text{under lockdown} + \mu_{state} + \sigma_{time} + \varepsilon + \ln(\text{Total}_{s,t})$$

要更深入理解為什麼加入 $\ln(\text{Total}_{s,t})$  就可以估算攻擊性推文比例，可以將 $\ln(\text{Total}_{s,t})$  在式子兩端移項，得到估算比例的式子：

$$\ln\left(\frac{\text{Aggressive Tweet Count}_{s,t}}{\text{Total}_{s,t}}\right) = \alpha_0 + \alpha_1 \text{under lockdown} + \mu_{state} + \sigma_{time} + \varepsilon$$

我們的統計模型能控制著各種可能影響結果的因子，捕捉客觀的因果關係效應。此方法已經獲得學界廣泛的認可，若待側變因與應變變因以此模型分析出正向的關係，不但可以紮實的證實正相關，也可以檢驗因果關係。

## 肆、研究結果

### 一、即時了解疫情下封城與未封城中人們情緒變化

利用神經網路模型即時分析整理各州攻擊性情緒變動，比較封城與未封城的地區的攻擊性情緒差異。自從美國於 2020/3/13 因應疫情宣布緊急狀態後，總共有 42 個州宣布封城，其中 40 個州的封城開始時間落在 3/20 與 4/4 之間。圖 7 顯示疫情爆發嚴重的前六個月封城與未封城的人們攻擊性的即時變化。而圖 8 將時間尺度拉長，與 2019 年的資料相比較，可以看見疫情發生後整體上升的趨勢。

#### (一) 憤怒：

憤怒的情緒在 2020 年 5 月顯著的升高，其中，封城下的人們憤怒升高的比未封城的人明顯。自從各地疫情開始嚴重的 2020 年四月的九個禮拜內，封城下憤怒情緒都較未封城高 2~3 個百分

點。圖 8 顯示 2019 年憤怒推文比例大多在 20% 上下浮動。至 2020 年 5 月，一度飆升至 34.0%。而進入到 2020 年的夏天，封城與未封城的人們攻擊性情緒稍歇，且往後幾個月呈現相似的趨勢，然而，其憤怒推文比例仍比疫情前的 2019 年時高。

### （二） 攻擊性語言：

攻擊性語言推文比例在 2020 年 4 到 5 月快速增加，持續約 7 週。2019 年時，攻擊性語言推文在 6% 上下浮動，而在 2020 年四月後的 7 週，封城與未封城的攻擊性語言快速分別升至 12.2% 與 9.8%。即便封城及未封城的比例在高峰後緩慢的下降，封城中的攻擊性語言推文比例仍持續比為封城者高 2~3 個百分點。

### （三） 仇恨言論：

至於仇恨言論，在 2020/4/1 後 7 週達到高峰。對於封城中的人，從疫情開始前的 3% 左右，直升到 7.8%，而為封城者則是到 6.0%。夏天時仇恨言論推文比例亦有下降，最後達到相對穩定的值。跟前述兩種攻擊性情緒相似的是，在高峰後下降至的穩定值，仍然明顯地比疫情前高。

對於三種不同的攻擊性情緒，其強度皆於 2020/4/1 後的 6~8 週達到峰值。其中，封城下的攻擊性情緒皆比沒有封城的高。不論封城或未封城，即時的變化趨勢大致相同，封城下的變化幅度尤其大。經過高峰之後，三種攻擊性情緒強度皆有和緩，推測是因為人們逐漸習慣新的生活，抑或感到疲乏。然而，穩定後的攻擊性情緒強度，仍然比疫情發生前高。

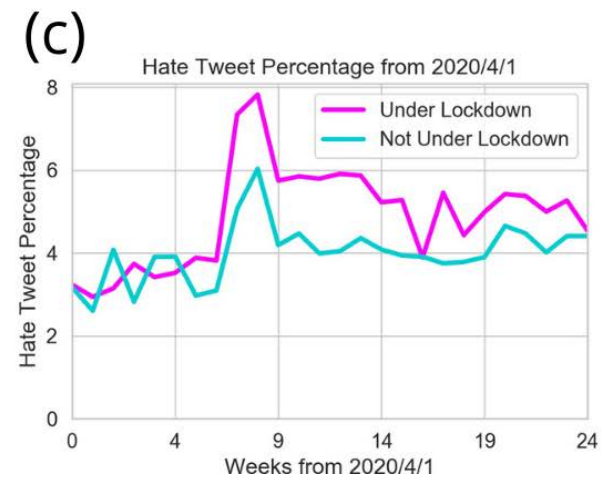
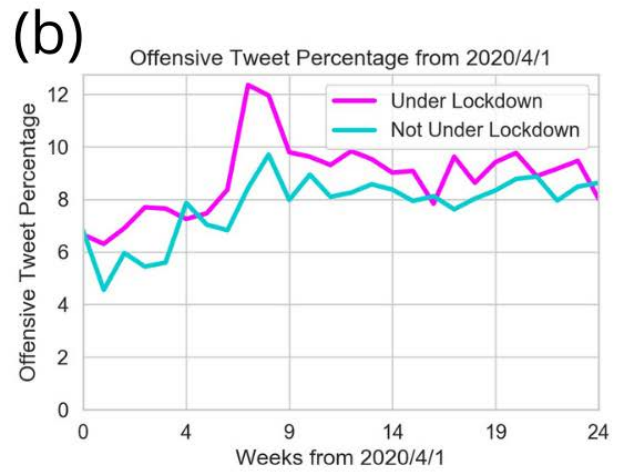
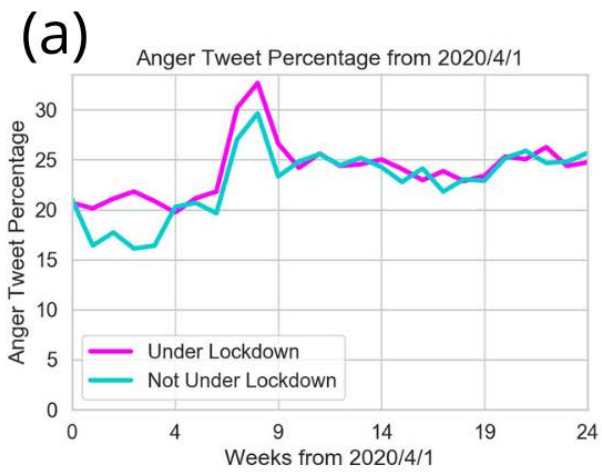
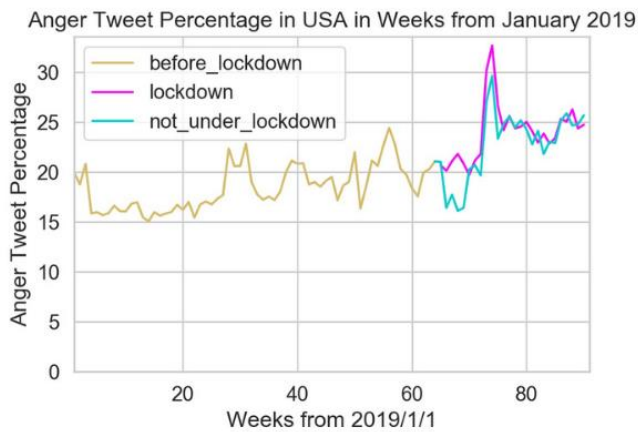
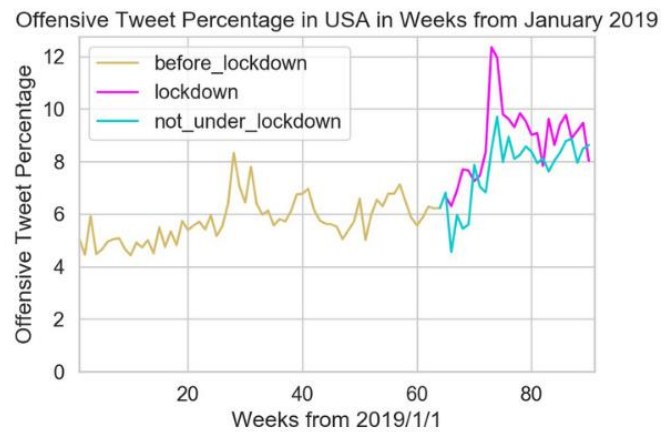


圖 7 橫軸為 2020/4/1 算起之週數，縱軸為攻擊性情緒百分比。紅線：封城下，藍線：未封城  
 a：憤怒、b：攻擊性語言、c：仇恨言論。

(a)



(b)



(c)

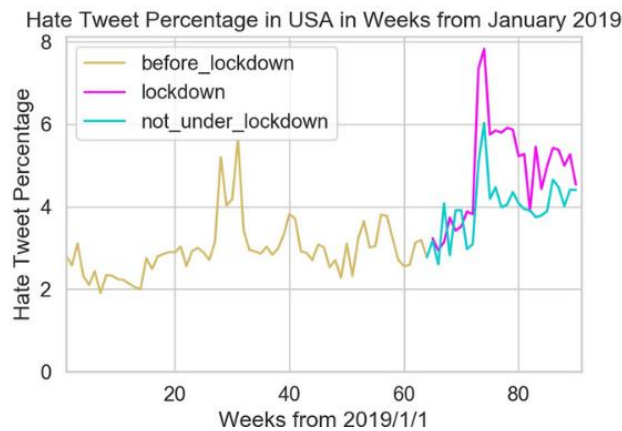


圖 8 橫軸為從 2019/1/1 算起之週數，縱軸為攻擊性情緒推文百分比。紅線：封城下，藍線：未封城，黃線：2019 年直到疫情前， a：憤怒、b：攻擊性語言、c：仇恨言論。此表顯示封城中的峰值並非正常四季的變化。

## 二、 觀測有封城的州當中，封城前後的攻擊性情緒變化

相較於前一種分析比較不同時間點封城與未封城的人們情緒變化情形。此部分分析在封城前後攻擊性推文百分比的變化。以封城開始時間為 0 將各州資料對齊後，可以看見封城後的攻擊性情緒明顯增加。表 7 顯示三種攻擊性情緒的變化情形。

表 7 封城前後攻擊性情緒推文的平均比例

Average 情緒類型	封城前	封城後
憤怒	18.5%	23.8%
攻擊性語言	5.8%	8.8%
仇恨言論	3.0%	4.9%

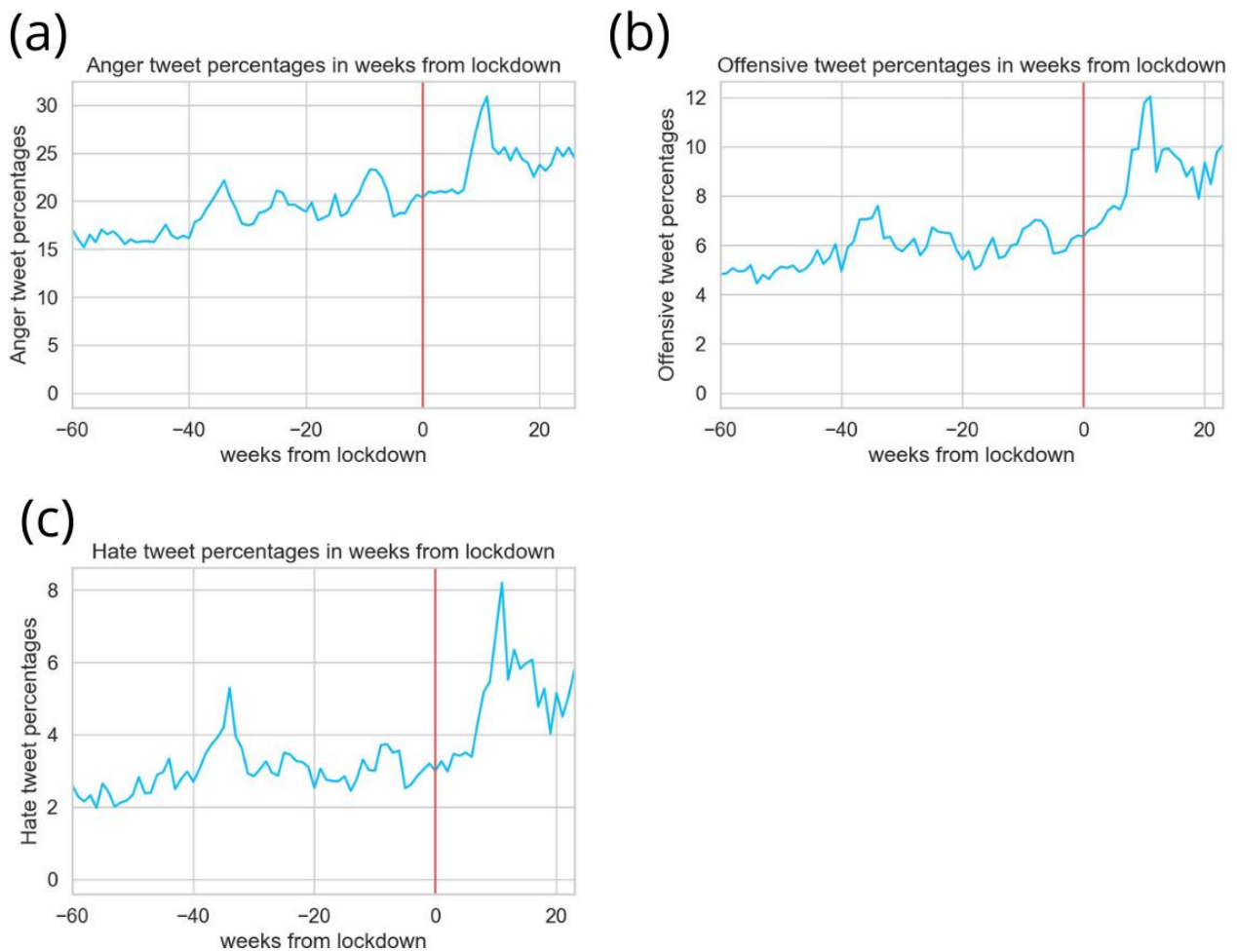


圖 9 橫軸為距離封城起始時間的週數，縱軸為攻擊性情緒百分比，鉛直紅線為封城開始時刻。  
 (a：憤怒、b：攻擊性語言、c：仇恨言論)。三種攻擊性情緒在封城後都急速上升

比較封城開始週前後的攻擊性情緒變化。圖 9 可以明顯看出，封城後的攻擊性情緒顯著增加。憤怒從原本 15~25%快速上升至高點 32%，攻擊性語言從原本 5~7%上升至 12%，仇恨言論由 3% 上升至 8%。這樣視覺化的呈現讓我們對封城與攻擊性情緒增加之間的可能關聯有一個基礎的認識，接下來的因果統計分析，將給出更確切的驗證。

### 三、 利用統計模型驗證封城與攻擊性情緒的因果關係

本研究利用 Poisson 回歸實作 DID(Difference in Difference)檢驗以應證封城與攻擊性情緒增加的因果關係。結果顯示，不論哪一種攻擊性情緒，封城皆與攻擊性情緒增加有關。在對數線性模型 (Log-Linear Model, e.g. Poisson) 當中，回歸得到的係數 $\alpha_1$  表示每一單位的 X 增加對應到  $\ln(Y)$  的增加，然而，因為對數函數的存在，使得係數 $\alpha_1$ 除了正負外的實際意義難以解讀。因應這個問題，發生率比例 (Incidence Rate Ratio, IRR) 即是欲觀測的 X 變數之係數指數化的結果，為 $e^{\alpha_1}$ ，可表示每一單位的 X 增加對應到的 Y 增加。

在表 8 三種攻擊型情緒當中，攻擊性推文比例與封城與否的 IRR 皆大於 1 (憤怒：1.049、攻擊性語言：1.168，仇恨言論：1.114) 意味著在其餘條件相似的情形下，封城中的憤怒、攻擊性語言以及仇恨言論之情緒強度比未封城的分別高出 4.9%、16.8%、11.4%。這樣的增加皆具備堅強的統計顯著性 (憤怒：P=0.002\*\*, 攻擊性語言：P<0.001\*\*\*, 仇恨言論：P=0.005\*\*)。在前面提及的時間與空間控制效應下，絕大多數可能影響結果的因素已經得到控制並消除。因此，本實驗可以驗證封城狀態對於攻擊性情緒增加的淨影響，發現其具備統計顯著的因果關係。

表 8 研究數據統計分析結果 \*: P<0.05, \*\*:P<0.01, \*\*\*:P<0.001

攻擊性推文比例	IRR	標準誤差	P Value	[95% 信賴區間]	
憤怒	1.050	.016	P=.002**	1.018	1.082
攻擊性語言	1.168	.032	P<.001***	1.107	1.232
仇恨言論	1.114	.043	P=.005**	1.032	1.202

## 伍、討論

### 一、 主要貢獻

#### (一) 訓練模型即時了解人們的攻擊性情緒，提供執政者有力參考

了解即時攻擊性情緒的變化是預測疫情下社會衝擊的重要起步。廣大的社群媒體資料提供人們一個新的平台，利用機器學習與大數據快速即時的了解各地的情緒變化，有效提供社會資源調度一個新的參考點，幫助注重個體傳統的心理學更有效率的了解到有需要的族群。

藉由分析各州在不同時間點的推文，我們可以捕捉攻擊性情緒在長段時間內的浮動、高山和低谷。利用各州不同的封城狀態，本研究發現了封城與未封城者攻擊性情緒的異同。在長時間的攻擊性情緒變化當中，有一些明顯的峰值值得探討。這些攻擊性情緒的高峰有可能反應當時社會事件。

從 2020 年 4 月開始至同年 5 月是多數州實施封城之時，三種攻擊性情緒都飆升至遠遠超過 2019 年任何時間所取得之值。同樣是這段時間，當日疫情的死亡數字從 3 月底不到 100 人指數性成長至 4 月中的 2000 人。這導致人們的情緒與生活受到莫大的影響。隨著封城與疫情漸漸得到控制，接下來幾個月的當日死亡人數持續下降，直到 6 月漸趨穩定。人們攻擊性情緒的變化趨勢與代表當時疫情嚴重程度的死亡人數趨勢雷同，其峰值也落在死亡率高且社會動蕩不安之時。

經過大約兩個月的封城，攻擊性情緒強度下降且漸趨平緩。其中，造成的原因可能為使人降低焦慮與攻擊性的「疫情疲勞」(Pandemic Fatigue) 現象 (Bodenmann et al., 2010)。根據世界衛生組織所發表的報告「在危機開始之時，多數人可以進入一個高亢的應對狀態：一系列可以幫助在高壓狀態短期生存的精神與物理適應系統。然而，當危機持續不散，他們就必須採取不一樣的應對模式，結果可能導致疲乏與缺乏動力。」(Organization, 2020)。就算整體的攻擊性有趨緩的趨勢，仍然可以看到封城下的攻擊性情緒比未封城者高。

相較於傳統問卷研究只能事後花大量時間分析，我們的神經網路模型能即時大量分析各地的情緒，觀察到重要的攻擊性趨勢變化，使人們可以做更好的防治與應變，並且提供執政者與公衛人員至關重要的資訊。

## （二） 了解封城與攻擊性情緒之間的因果關係，拓展人們對封城心理影響的認識

本研究利用統計方式應證了封城與攻擊性情緒增加的因果關係。計數資料專用的 Poisson 回歸模型幫助我們分析某個時間某個地點下的攻擊性情緒。時間與空間的控制效應將可能影響結果的因素去除，使研究結果更為準確。嚴謹的統計分析幫助我們測定封城對於攻擊性增加的淨影響，得到統計顯著的因果推論（憤怒： $P=0.002^{**}$ ，攻擊性語言： $P<0.001^{***}$ ，仇恨言論： $P=0.005^{**}$ ）。我們的發現提供人們一個更全面的視野，了解封城較未知卻極為重要的心理層面影響。

## （三） 建立比問卷調查時間與空間範圍更大的攻擊性情緒資料集

除了方法效率上的貢獻，本研究利用模型分析出來的地理化情緒資料，顯著的比心理學問卷的資料更精確，收集範圍更廣，資料量更大，也更容易取得。首先，我們的資料當中包含確切的日期，比起問卷調查因為更為耗時只能達到月的刻度精確許多。第二，利用 Twitter 爬文的功能，我們可以根據各州的人口數篩選相對應數量的受試者，使各地的人能夠更均等的被採樣。在起初的問卷當中，研究者僅指出受試者來自全美各地，並未註明其分佈狀況，根據研究中倚賴受試者自行上門的方法合理推測，受試者來自的地方無法像網路爬文一樣分佈均勻。第三，我們的資料量比問卷調查更大，更完整。由我們採樣的 11,455 名用戶，1,281,362 比推文就比問卷採樣的 5928 名受試者多非常多。而且不像問卷因為追蹤受試者的困難，每個月必須換受試者，我們的資料集中有每一位用戶從 2019/1 到 2020/10 的所有推文，連續性大大進步。最後，我們的資料當中擁有疫情發生前一年的資料，是問卷得到的。這部分資料提供關鍵的比對，讓我們看出疫情下攻擊性的增加已經超越正常四季的波動。綜合來說，我們分析出的資料集比問卷資料時間上更精確，地理上更均勻，數量上更龐大，更容易取的，且提供了關鍵的資料，得到問卷調查無法看出的發現(表 9)。



表 9 大數據與問卷資料研究法的比較

	本研究資料集	問卷資料集
時間精確度	天	月
地理	可以根據各州人口精確採樣	可能需倚賴各州參與問卷調查的踴躍程度
資料集大小	11,455 位用戶與 1,281,362 筆推文	5928 名受試者
可取得之資料廣泛度	任何時間與地點 (包括實驗開始之前)	根據實驗規劃與經費侷限的時間與地點，難以取得實驗開始前的資料做比對

#### (四) 對政策的幫助

統計數據對政策影響的評估，對於疫情下的政策決定至關重要(Dattner et al., 2022)。利用統計大數據，發現封城與攻擊性情緒增加的關係，可以提供國家施行封城精神健康配套措施的指引。針對疫情對精神健康的影響，已有從下家暴保護(Chandan et al., 2020)、學校輔導(Pincus et al., 2020)、以及心理諮商(Kavoor et al., 2020)等層面著手。

本研究透過即時的偵測攻擊性情緒地理的分佈，能幫助這些政策執行的更精準且有效率。本研究因果關係的測定，也能帶來封城攻擊性影響更多的重視與政策上的能見度，使有意義封城配套計畫，實際幫助到社會。

## 二、 與先前研究比較與突破

先前學者用問卷觀察出封城下攻擊性情緒變化，卻因資料範圍與數量的限制，難以偵測統計上的因果關係。而本研究的研究利用控制效應模型的計量經濟學方法從大數據中測出了封城與攻擊性情緒增加的因果關係。相較於傳統問卷研究一次只能針對一個問題設計問卷，本研究中的社群媒體資料可以從中分析出多種情緒與特徵，提供更大範圍且精確的資訊。

先前自然語言處理（N L P, Natural Language Processing）的研究大多專注在如何使機器理解語意的課題上。然而，鮮少有研究將此技術帶至社會科學的實證應用面。受到兩者的啟發，本研究利用N L P技術訓練模型分析了封城與攻擊性情緒的關係，利用資訊科技使傳統心理學與公衛研究更加豐富且有效率。

### 三、 未來發展與應用

即時的情緒分析對於包含軍事、政治、商業等等領域都有廣大的應用。例如，在上海嚴重的封城當中，藉由社群媒體的資料就可以真實地了解民眾當下的情緒反應。又如烏俄戰爭，透過即時的情緒分析，可以透過網路深入戰區，觀測當下的戰況和公眾的輿論。執政者若能了解各地民眾的情緒，就有機會做政策的調整與應變。

了解封城與攻擊性的因果關係，關乎人們的心理健康。在做疫情下封城政策時，政府往往需考慮眾多的影響因素，若能對於封城的影響有更全面的了解，就能做出更理性，對人民更好的政策(Williams et al., 2020)。

## 陸、 結論

本研究突破了時間與空間限制的藩籬，使任何需要了解群眾情緒的專家，可以在做細部問卷調查之前，快速的了解大範圍群眾情緒變化，不但大大的提升效率，也取代問卷耗時耗能資料少的劣勢，用龐大的資料彌補心理學問卷資訊不足。社群媒體是廣大群眾抒發自己情緒的地方，藉由了解社群媒體上的情緒，可以探討的社會現象遠不止封城。本研究的方法在未來的研究當中，用自然語言處理從心理學的角度出發，可以幫助人們快速且有效率地察覺、了解，並面對社會脈動。更可以幫助政策制定者甚至執法者，預測或者迅速回應變動中的社會氛圍，以有效維持社會安全網絡。

## 柒、 參考資料

- Bodenmann, G., Meuwly, N., Bradbury, T. N., Gmelch, S., & Ledermann, T. (2010). Stress, anger, and verbal aggression in intimate relationships: Moderating effects of individual and dyadic coping. *Journal of Social and Personal Relationships*, 27(3), 408-424.
- Breuer, J., & Elson, M. (2017). *Frustration-aggression theory*. Wiley Blackwell.

- Catelli, R., Pelosi, S., & Esposito, M. (2022). Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian. *Electronics, 11*(3), 374.
- Chandan, J. S., Taylor, J., Bradbury-Jones, C., Nirantharakumar, K., Kane, E., & Bandyopadhyay, S. (2020). COVID-19: a public health approach to manage domestic violence is needed. *The Lancet Public Health, 5*(6), e309.
- Dattner, I., Gal, R., Goldberg, Y., Goldshtein, I., Huppert, A., Kenett, R. S., Manor, O., Pfeffermann, D., Schechtman, E., & di Serio, C. (2022). The role of statisticians in the response to COVID-19 in Israel: a holistic point of view. *Israel Journal of Health Policy Research, 11*(1), 1-9.
- Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. Proceedings of the International AAAI Conference on Web and Social Media,
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Dorottya Demszky, D. M.-A., Jeongwoo Ko, Alan Cowen, Gaurav Nemade, Sujith Rav. (2020). Go Emotions: A Dataset of Fine-Grained Emotions. *Association of Computational Linguistics (ACL)*.
- Founta, A. M., Djouvas, C., Chatzakou, D., Leontiadis, I., Blackburn, J., Stringhini, G., Vakali, A., Sirivianos, M., & Kourtellis, N. (2018). Large scale crowdsourcing and characterization of twitter abusive behavior. Twelfth International AAAI Conference on Web and Social Media,
- Groarke, J. M., Berry, E., Graham-Wisener, L., McKenna-Plumley, P. E., McGlinchey, E., & Armour, C. (2020). Loneliness in the UK during the COVID-19 pandemic: Cross-sectional results from the COVID-19 Psychological Wellbeing Study. *PloS one, 15*(9), e0239698.
- Hilbe, J. M. (2014). *Modeling count data*. Cambridge University Press.
- Hu, T., Wang, S., Luo, W., Zhang, M., Huang, X., Yan, Y., Liu, R., Ly, K., Kacker, V., & She, B. (2021). Revealing Public Opinion Towards COVID-19 Vaccines With Twitter Data in the United States: Spatiotemporal Perspective. *Journal of medical Internet research, 23*(9), e30854.
- Jang, H., Rempel, E., Roe, I., Adu, P. A., Carenini, G., & Janjua, N. Z. (2022). Tracking Public Attitudes toward COVID-19 Vaccination on Tweets in Canada: Using Aspect-based Sentiment Analysis. *Journal of medical Internet research*.
- Kavoor, A. R., Chakravarthy, K., & John, T. (2020). Remote consultations in the era of COVID-19 pandemic: Preliminary experience in a regional Australian public acute mental health care setting. *Asian Journal of Psychiatry, 51*, 102074.
- Kempf-Leonard, K. (2004). Encyclopedia of social measurement.
- Killgore, W. D., Cloonan, S. A., Taylor, E. C., Anlap, I., & Dailey, N. S. (2021). Increasing aggression during the COVID-19 lockdowns. *Journal of affective disorders reports, 5*, 100163.
- Kotelnikova, A. (2020). Comparison of Deep Learning and Rule-based Method for the Sentiment Analysis Task. 2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon),
- Majumdar, P., Biswas, A., & Sahu, S. (2020). COVID-19 pandemic and lockdown: cause of sleep disruption, depression, somatic pain, and increased screen exposure of office workers and students of India. *Chronobiology international, 37*(8), 1191-1200.
- Marroquín, B., Vine, V., & Morgan, R. (2020). Mental health during the COVID-19 pandemic: Effects of stay-at-home policies, social distancing behavior, and social resources. *Psychiatry research, 293*, 113419.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica, 22*(3), 276-282.
- Moreira, D. N., & da Costa, M. P. (2020). The impact of the Covid-19 pandemic in the precipitation of intimate partner violence. *International journal of law and psychiatry, 71*, 101606.

- Muñoz-Fernández, N., & Rodríguez-Meirinhos, A. (2021). Adolescents' concerns, routines, peer activities, frustration, and optimism in the time of COVID-19 confinement in Spain. *Journal of Clinical Medicine, 10*(4), 798.
- Niedzwiedz, C. L., Green, M. J., Benzeval, M., Campbell, D., Craig, P., Demou, E., Leyland, A., Pearce, A., Thomson, R., & Whitley, E. (2021). Mental health and health behaviours before and during the initial phase of the COVID-19 lockdown: longitudinal analyses of the UK Household Longitudinal Study. *J Epidemiol Community Health, 75*(3), 224-231.
- Organization, W. H. (2020). *Pandemic fatigue: Reinvigorating the public to prevent COVID-19: Policy considerations for member states in the WHO European region.*
- Pesonen, A.-K., Lipsanen, J., Halonen, R., Elovainio, M., Sandman, N., Mäkelä, J.-M., Antila, M., Béchar, D., Ollila, H. M., & Kuula, L. (2020). Pandemic dreams: network analysis of dream content during the COVID-19 lockdown. *Frontiers in psychology, 2569.*
- Pincus, R., Hannor-Walker, T., Wright, L., & Justice, J. (2020). COVID-19's effect on students: How school counselors rise to the rescue. *NASSP Bulletin, 104*(4), 241-256.
- Ridhwan, K. M., & Hargreaves, C. A. (2021). Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore. *International Journal of Information Management Data Insights, 1*(2), 100021.
- Sanga, S., & McCrary, J. (2020). The impact of the coronavirus lockdown on domestic violence. *Available at SSRN 3612491.*
- Savage, M. (2020). Why the pandemic is causing spikes in break-ups and divorces. <https://www.bbc.com/worklife/article/20201203-why-the-pandemic-is-causing-spikes-in-break-ups-and-divorces>
- Serrano-Alarcón, M., Kentikelenis, A., Mckee, M., & Stuckler, D. (2022). Impact of COVID-19 lockdowns on mental health: Evidence from a quasi-natural experiment in England and Scotland. *Health economics, 31*(2), 284-296.
- Shah, S. G. S., Nogueras, D., Van Woerden, H. C., & Kiparoglou, V. (2020). The COVID-19 pandemic: a pandemic of lockdown loneliness and the role of digital technology. *Journal of medical Internet research, 22*(11), e22287.
- Shamrat, F., Chakraborty, S., Imran, M., Muna, J. N., Billah, M. M., Das, P., & Rahman, O. (2021). Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm. *Indones. J. Electr. Eng. Comput. Sci, 23*(1).
- Sher, L. (2020). The impact of the COVID-19 pandemic on suicide rates. *QJM: An International Journal of Medicine, 113*(10), 707-712.
- StataCorp. (2021). *Stata 17 Base Reference Manual.* College Station, TX: Stata Press.
- Su, Y., Xue, J., Liu, X., Wu, P., Chen, J., Chen, C., Liu, T., Gong, W., & Zhu, T. (2020). Examining the impact of COVID-19 lockdown in Wuhan and Lombardy: a psycholinguistic analysis on Weibo and Twitter. *International journal of environmental research and public health, 17*(12), 4552.
- Usta, J., Murr, H., & El-Jarrah, R. (2021). COVID-19 Lockdown and the increased violence against women: understanding domestic violence during a pandemic. *Violence and gender, 8*(3), 133-139.
- Wang, X., Hegde, S., Son, C., Keller, B., Smith, A., & Sasangohar, F. (2020). Investigating mental health of US college students during the COVID-19 pandemic: Cross-sectional survey study. *Journal of medical Internet research, 22*(9), e22817.
- Williams, G. A., Diez, S. M. U., Figueras, J., & Lessof, S. (2020). Translating evidence into policy during the COVID-19 pandemic: bridging science and policy (and politics). *Eurohealth, 26*(2), 29-33.

## 【評語】 052701

本研究使用 BERT 進行訓練，辨識並標記 twitter 中發文的三類情緒（憤怒，攻擊，仇恨），以分析封城前後這些情緒詞出現的變化，並且利用統計模型分析封城與否和這些情緒詞出現頻率的因果關係。本研究所強調的優勢在於取代傳統的問卷調查方法，資料收集可以更大量、更即時，以節省人力與金錢的方式進行研究。

本研究進行的時機和所探討的議題和當今蔓延全球的疫情有關，所使用的研究方法奠基於近年來資料科學的發展，審查委員對本研究有以下問題、建議：

1. 此研究的主要假說為何？在作品中無明確呈現。若本研究的議題與假設是”封城是否造成負面情緒上升”，這似乎是必然的結果，不知進一步探討此議題的理論基礎為何？
2. 本研究中提到：“透過問卷收集到的資料較無法推論因果關係”，不知此說法的背後基礎為何？如果本研究欲驗證透過分析網路上的貼文可取代問卷有效性的話，需要有問卷資料作為比較
3. 並非所有人都會使用社群媒體、發表貼文，使用社群媒體者的特質（例如：是否喜愛發表負面言詞者比較會貼文）、會被公開發表的貼文內容屬性（例如：在社群媒體上的貼文是否大多以正向的貼文為主，而負面貼文通常會比較少數）等因

素，都可能對本研究的資料來源造成偏差。請問作者在研究中可如何避免、減輕此問題？是否可以在統計及分析方法上做調整？

4. 根據本研究的資料顯示，在封城期間與未封城期間所得到的數據有著類似的起伏 (fluctuation)，看起來似乎在不同的季節性也會影響貼文內容。另外，不同城市的疫情嚴重程度(疫情越嚴重的城市越有可能封城)、居民的屬性(越崇尚自由的城市越不可能封城)等因素，也是會影響本研究結果的混淆變項，建議作者對這些可能性進行評估和討論。
5. 本作品的撰寫上還有更加精進的空間，如有些文句並不完整，也有錯字、參考資料不完整、格式不正確的問題。

## 作品簡報

# 基於神經網路分析封城對民眾攻擊性 情緒之影響：以美國為例

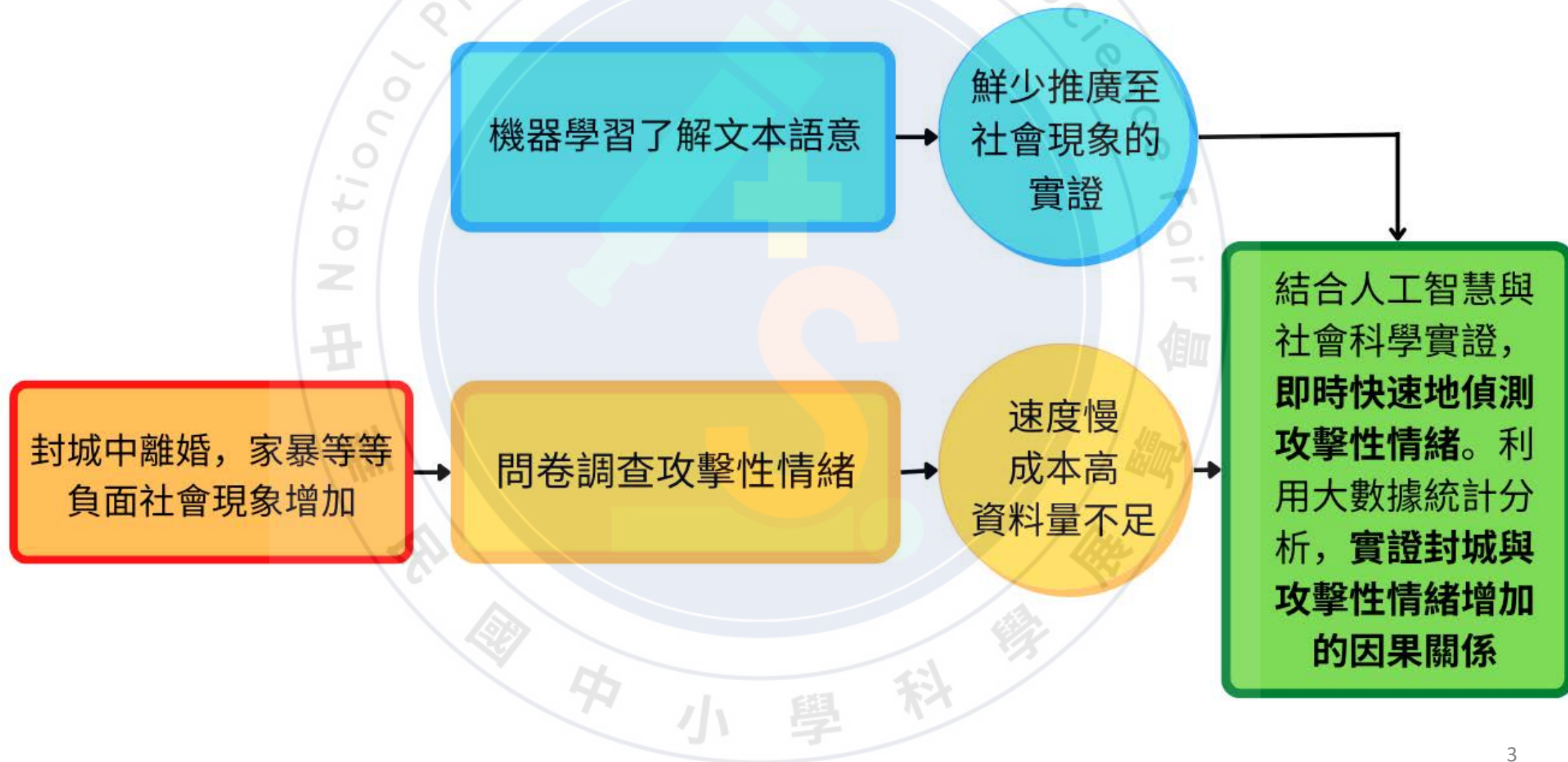
組別：高級中等學校組  
科別：行為與社會科學科



# 文獻探討與比較

研究探討面向 文獻資料種類	研究疫情下的各種情緒	利用人工智慧理解大量文獻中的情緒	研究封城與各種負面社會現象的相關性	利用統計模型與大數據判斷封城與負面社會現象的因果關係
(Killgore et al., 2021)	√	x	√	x
(Su et al., 2020)	x	x	√	x
(Niedzwiedz et al., 2021)	x	x	√	x
(Majumdar et al., 2020)	x	x	√	x
(Hu et al., 2021)	√	√	x	x
(Ridhwan & Hargreaves, 2021)	√	√	x	x
(Jang et al., 2022)	√	√	x	x
(Sanga & McCrary, 2020)	x	x	√	√
<b>本研究探討問題</b>	√	√	√	√

# 動機與目的



# 研究流程及方法

1. Twitter 資料採樣
2. 訓練情緒分類模型
3. 分析 Twitter 資料採樣中的情緒
4. 利用統計回歸檢定封城與攻擊性情緒的因果關係

# 資料採樣

Twitter 地理標注的功能

11,455位美國不同州的使用者

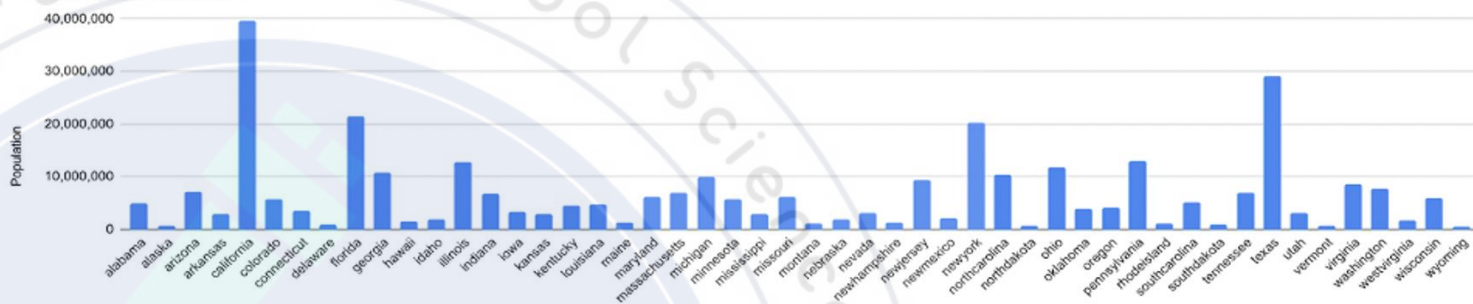
於2019/1/1~2020/10/1的

1,281,362 筆

帶有時空資訊的推文

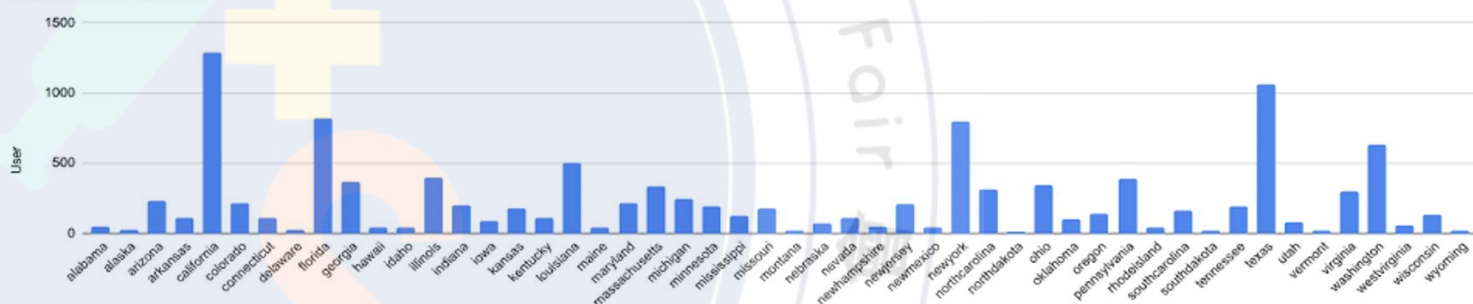
人口數量

State Population Count



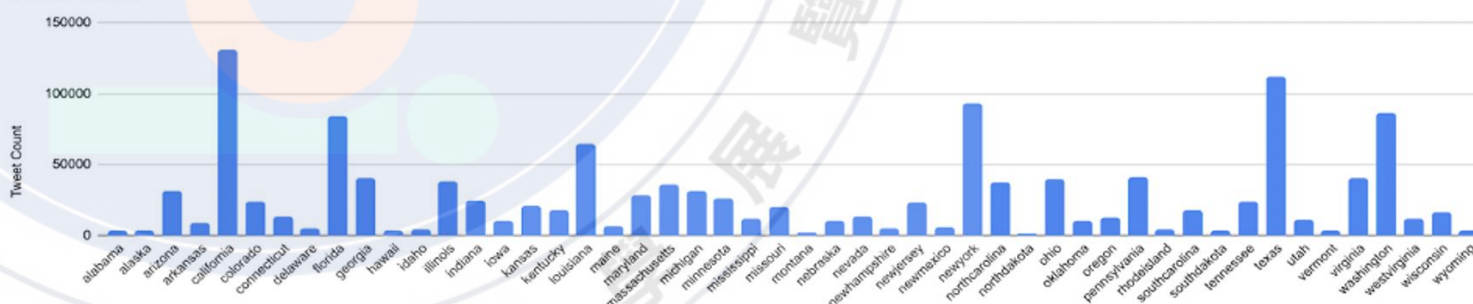
使用者數量

States User Count



推文數量

Tweet Count



州名

# 訓練模型

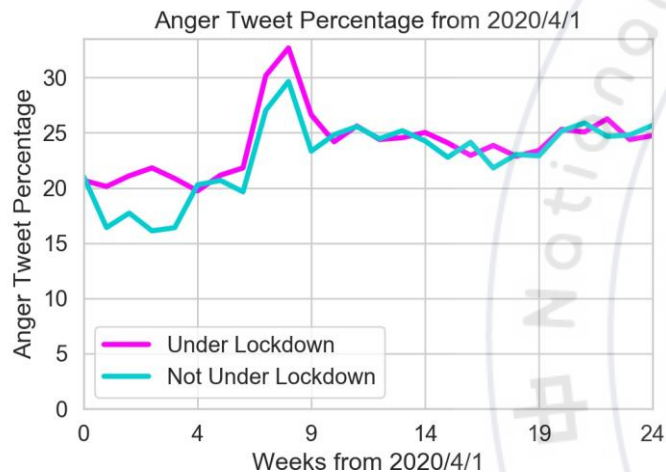


## 模型表現分析

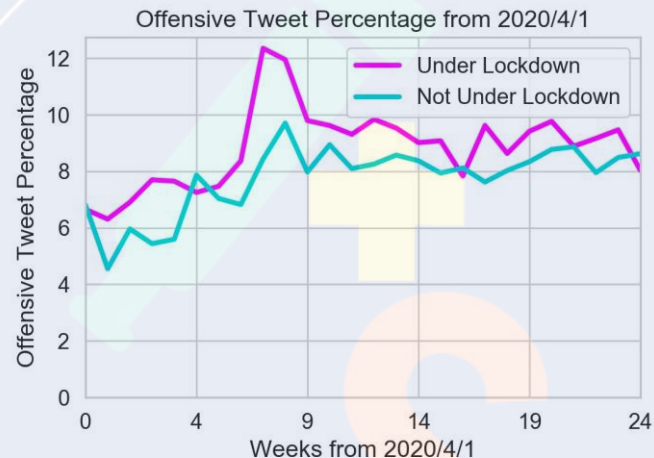
情緒類型	準確性參數	模型好壞參數F1 score
憤怒 (Anger)	0.851	0.847
攻擊 (Offensive Language)	0.975	0.970
仇恨 (Hate Speech)	0.938	0.933

# 即時分析攻擊性情緒變化

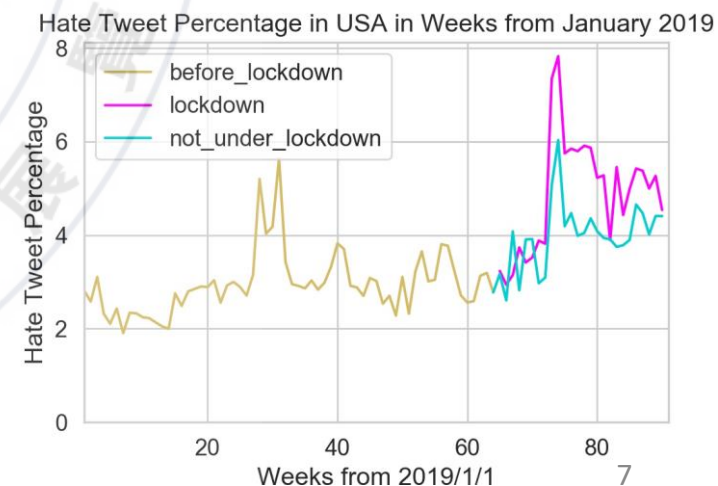
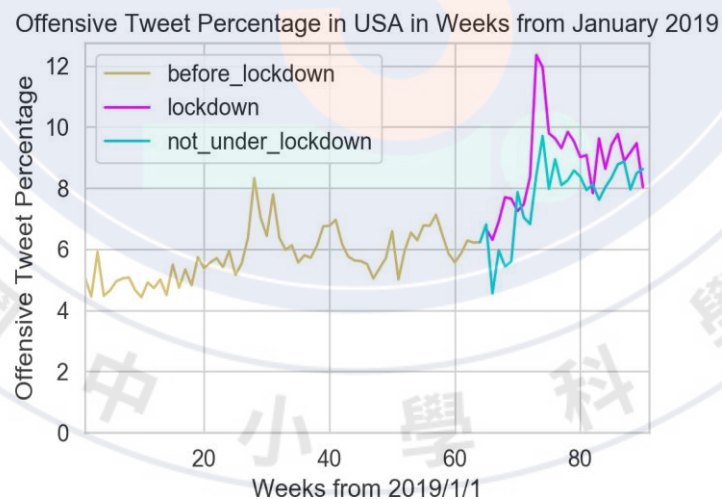
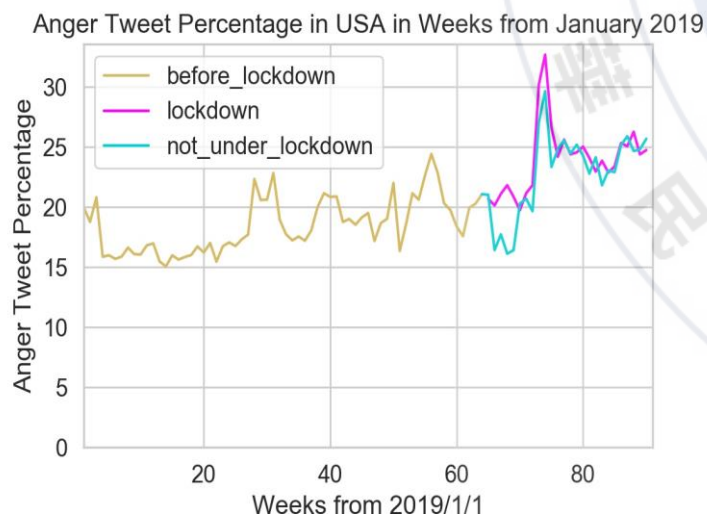
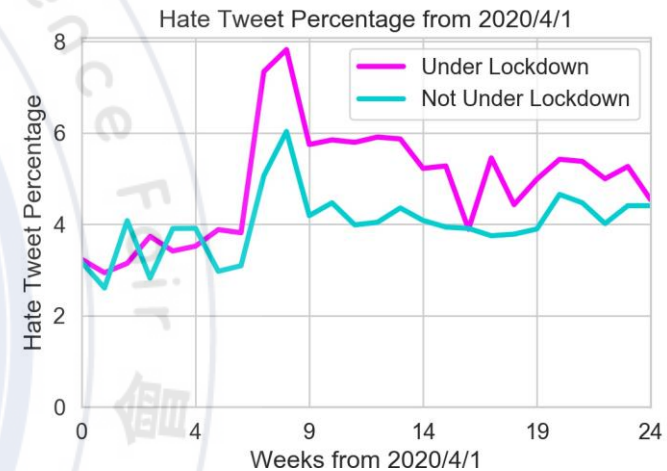
## 憤怒



## 攻擊性



## 仇恨



# 固定效應模型(Fixed Effect Models)

- **問題**：大規模觀察性的資料，難以取得傳統對照組，如何由其推論因果關係？
- **實際解決方法**：對於每一州與每個星期，控制在特定 時間 空間下，對於應變變因有影響的因子
- **核心概念**：控制可能造成影響的變因，測量封城狀態對攻擊性情緒增加的淨影響
- **舉例**：若某一州的政治傾向造成其人民特別有攻擊性，則此因素反映在該州的數據的平均當中，因此當我們在回歸等式中考慮平均，即可消彌該因素及其他相似因素的影響

# 分析封城與攻擊性增加的因果關係

- 基礎柏松 ( Poisson ) 固定效應模型 :

$$\ln(\text{Aggressive Tweet Count}_{s,t}) = \alpha_0 + \alpha_1 \text{under lockdown} + \mu_{state} + \sigma_{time} + \varepsilon$$

- 加上  $\ln(\text{total})$  項使其估計攻擊性推文比例

$$\ln(\text{Aggressive Tweet Count}_{s,t}) = \alpha_0 + \alpha_1 \text{under lockdown} + \mu_{state} + \sigma_{time} + \varepsilon + \ln(\text{Total}_{s,t})$$

- 可得

$$\ln\left(\frac{\text{Aggressive Tweet Count}_{s,t}}{\text{Total}_{s,t}}\right) = \alpha_0 + \alpha_1 \text{under lockdown} + \mu_{state} + \sigma_{time} + \varepsilon$$



# 因果關係分析結果

正向因果關係

高度統計顯著性

攻擊性推文比例	相關係數	標準誤差	P 值	[95% 信賴區間]	
憤怒	1.050	.016	P=.002**	1.018	1.082
攻擊	1.168	.032	P<.001***	1.107	1.232
仇恨	1.114	.043	P=.005**	1.032	1.202

# 結果與應用

**結果1**：即時分析封城下民眾的攻擊性情緒變化

**應用**：為社會資源分配、政策執行等領域，提供即時的參考資訊

**結果2**：證實封城與攻擊性情緒增加之間的**因果關係**

**應用**：幫助人們更認識封城對於心理層面的影響，使執政者在封城前做出更完善的評估與配套措施

# 結論

透過即時分析封城中的社群媒體資料，證實封城的確會造成攻擊性情緒的增加。  
利用機器學習與大數據分析提供心理學一個新的研究視角。

# 參考文獻

1. Callaway, B. and P.H. Sant' Anna, *Difference-in-differences with multiple time periods*. Journal of Econometrics, 2021. 225(2): p. 200-230
2. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
3. Killgore, W.D., et al., *Increasing aggression during the COVID-19 lockdowns*. Journal of affective disorders reports, 2021. 5: p. 100163