

# 預測企業員工評論正負向及評論文本分析

闕元甫  
會計所  
國立台灣大學  
R11722019

陳柏言  
資訊管理所  
國立台灣大學  
R11725004

胡予瑄  
資訊管理所  
國立台灣大學  
R11725015

林宜璇  
資訊管理所  
國立台灣大學  
R11725039

楊佳真  
資訊管理所  
國立台灣大學  
R11725040

## 1 THE PURPOSE OF OUR PROJECT

考量到求職者在選擇企業時，多少會希望透過內部員工對公司的評論與評價，來了解該公司的狀況，但企業員工之評論往往非常多，難以迅速了解企業內部狀況；另外，對於企業而言，了解員工心聲可以協助改善工作環境，以增進公司治理及未來收益，而員工評論可以部分反映出員工對企業的看法，因此了解評論並進行相關的分析至關重要。

故本專案藉由建立文字探勘機器學習模型，預測評論(review)是屬於正向或負向，並透過評論從中挖掘工作者或企業員工的關注面向；此外，也會透過情緒用字分析，了解員工評論的情緒用字分布狀況，期望以上預測結果與評論分析能提供求職者、企業及平台方參考。

## 2 OUR SOLUTION

### 2.1 資料收集

本專案採用 Kaggle 公開資料集 - Employer Review about their organization，該資料集為 json 資料格式，資料共有 145209 筆，其中 5 個欄位，分別為 ReviewTitle、CompleteReview、URL、Rating、ReviewDetails（包含填寫時間、是否在職的資訊。

### Json (原始資料欄位)

	英文變數名稱	中文變數名稱	型態
1	ReviewTitle	評論標題	String
2	CompleteReview	論文內容	String
3	URL	網址 (內含員工所屬之企業名稱)	String
4	Rating	評分等級	Numeric
5	ReviewDetails	詳細資料	填寫時間 Date 仍在職 String

表一、Employer Review about their organization 資料集  
原始欄位及資料格式

首先，本專案將 json 資料格式改為 DataFrame 形式，並且將 URL 欄位中的企業名稱擷取出來形成新欄位 Company，另外將 ReviewDetails 中的填寫時間移除，僅保留仍在職作為欄位。

### DataFrame (整理後欄位)

	英文變數名稱	中文變數名稱	型態
1	ReviewTitle	評論標題	String
2	CompleteReview	評論內容	String
3	Company	員工所屬之企業名稱	String
4	isCurrentEmployee	仍在職	Boolean
5	Rating	評分等級	Numeric

表二、Employer Review about their organization 資料集  
整理後欄位

### 2.2 資料前處理

在資料前處理的部分，首先，本專案將 14 萬多筆資料中的 Rating 欄位，進行加總平

均，若 Rating 值高於其平均值，將該 Rating 欄位改為 positive；Rating 值低於其平均值，則改為 negative。

考量資料筆數過多的問題，因此採用分層抽樣方式取出 10% 資料進行分類模型訓練，接著再用分層抽樣的方式抽 10% 作為測試資料，剩下 90% 作為測試資料跟驗證資料。

完成資料切分後，針對文字進行 tokenization、normalization、lower case、stemming、stopword removal，以及移除阿拉伯文的處理。最後將所有文本轉成向量 (TF-IDF)。

### 2.3 評論正負向分類預測

在 Muhammad Saqlain 等人<sup>[1]</sup>的研究中，他們將員工評論的文本資料轉換為 tf-idf 的文件表達方式，並使用分類器來預測員工對公司的滿意與否，因此本專案參考該篇研究的作法，同樣將評論的文本資料轉換為 tf-idf 向量。在評論正負向分類預測的部分中，本專案使用資料集裡的「CompleteReview」欄位來轉成 tf-idf 向量。此外，為了新增 label 欄位（代表員工評論的正向或負向），將「Rating」欄位做平均，如果「Rating」欄位大於等於 4.053 則定義為正向的評論、小於 4.053 則定義為負向的評論。

在將「CompleteReview」欄位來轉成 tf-idf 向量的過程中，本專案先對該欄位做了 tokenization、lower casing、stemming 以及 stop word removal，建立了一個原始字典、刪除阿拉伯文字及明顯無意義的長字。之後計算出 tf-idf 並將它轉為 unit vector。

在字典特徵選取的部分本專案採用了 Log-likelihood ratio 及 Chi-square 的方法，各挑選出了 2000、4000、5000、6000、8000 個字後去分類。

Kowsari 等人<sup>[2]</sup>整理了文字探勘分類領域的常見作法，包括資料前處理、各種分類器模型的介紹。因此我們在看完他們的 survey paper 後，本專案選擇了 SVM、Random Forest、XGBoost 及 Naive Bayes 來進行分類任務並記錄了對應的 Accuracy、Precision、Recall 及 F1 Score 的結果進行比較，在這篇報告中會以準確度的部分進行結果表現說明。

### 2.4 評論主題分群

評論除了可以幫助本專案了解一家企業的工作環境、工作性質與企業文化等，也可以讓本專案知道大多工作者（企業員工）的關注面向，而求職者便可以透過評論，依據大多工作者的關注面向，去評斷一間企業是否符合自己的需求，或自己是否適合在某產業的企業下工作。另外，對於企業而言，瞭解員工的關注面向，也可以協助自身根據這些關注面向改善企業的工作環境等，提升企業自身長期的收益。然而，員工的評論往往非常多，求職者或企業無法迅速從眾多評論中瞭解員工的關注面向。

在上述動機下，機器學習中的分群 (clustering) 可以幫助找出內容、性質或主題相似的評論，在這之後，針對不同主題 (群) 的評論進行特徵選擇，透過所挑選的特徵字，瞭解該群的主題內容或討論面向，因此在這次的專案中，首先會透過分群的方式將眾多評論進行分群，並搭配特徵選擇，針對各群評論挑選具有代表性的字，以瞭解各群評論所探討的主題或關注面向為何。

Sumbal Riaz 等人<sup>[3]</sup>提到，企業要從龐大的資料中萃取出消費者的意向具有困難性，因此他們針對了消費者評論的進行 K-means 分群及相關的分析，以幫助企業從評論中挖掘出消費者的消費意向。

在本專案中，同樣採取 K-means 分群方法，將「CompleteReview」欄位中的各個評論

轉換成 TF-IDF 單位向量後，接著會使用 sklearn 套件中的 K-means 進行分群，並搭配 Elbow method，分別測試不同 K 值（分群數 2 ~ 10 群）下的 RSS 值，觀察 RSS 值下降趨緩時的 K 值為何，以此挑合適的分群數，接著，會使用該分群數，將眾多評論進行 K-means 分群，並針對分群後的各群評論進行特徵（字）選擇。

在本專案中，為挑選正指標字、避免挑選到負指標字，本專案選擇使用 PMI (Point Mutual Information) 值作為特徵選擇方法，並且為了解決使用 PMI 時傾向選擇罕見字的問題，本專案會將 PMI 值乘上字的出現頻率，以此選擇常出現、又具有代表性的字，而在本專案中，會挑選分數最高的 10 個字，來了解各群評論的內容或主題，以此探討員工的關注面向。

Yeonjae Jung 等人<sup>[4]</sup>在研究中，提到透過員工評論探討員工工作滿意因素鮮少被研究，因此他們使用了 LDA (Latent Dirichlet Allocation) 方法，從員工評論中挖掘員工工作滿意的因素。

因此在本專案中，除了使用分群及特徵選擇，探討評論的主題，也會使用基於線性代數的 SVD (singular value decomposition) 以及基於機率的 LDA 尋找評論的主題或關注面向，並與分群的結果進行比較，觀察是否會有新的主題出現。

## 2.5 評論情緒用字分析

由於一般常見的情緒分析的方法，僅將文字分成正向與負向，對於分析後所得到的結果較為不精確，所以為了解決上述問題，故於此次的專案中，透過 NRC 字典(裡面包含信任、參與感、歡樂、驚喜、恐懼、難過、憤怒、厭惡等情緒)，將所有高達十四萬篇的評論加以分析，只要該篇評論內的單字有在字典內出現，則該篇評論的各情緒數量則會因該英文單

字所對應到的情緒而有所變動，例如當一篇評論內存在 aback 此單字，則在 NRC 字典中，可以看到該字典對應到的十個情緒皆為零，所以該篇評論的各情緒數量則不會有所變動，然而如果是存在 abandon 此單字時，在字典內可以看到該單字所對應到的十個情緒中，anger、fear、negative、sadness 此四個情緒皆顯示為一，而其餘的情緒則顯示為零，代表該單字與上述四個情緒係屬相關，所以此時，該篇評論的該四個情緒的數量則會分別加一。

透過上述的做法，逐一計算出十四萬篇評論的十種情緒數量。一開始的做法是於迴圈中，建立長度為十的 list，然後一個一個去看字典內的字其對應到的情緒所顯示的數值，在執行的過程中，發現到此種作法會過於緩慢，進而對於後續的操作與分析造成負面的影響，所以在經過一番思考後，決定將字典內的每一列轉換成 numpy.array 的形式，而每篇評論各情緒的數量也使用相同作法，只要看到字典內有相同的單字，就將字典內轉換成 numpy.array，並與 numpy.array 形式的評論各情緒數量進行加總，最後再將該評論的十種情緒數量存入 metric 中即可。藉由這種方式，就不需要每一個評論內的每一個字都要去掃字典內該字的十種情緒，使運算的效率大幅提升，但或許是由於篇數過大的緣故，所以儘管提升了執行效率，依然要花上一段時間來完成。

得到所有評論的情緒用字情形後(亦即將所有評論的各情緒數量存入 metric)，為了能使結果一目瞭然，所以透過圖表的呈現方式來達到上述的目的。由於圖表的繪製，不須要評論的內文與公司等等一些資訊，所以使用篩選的功能，擷取出所需要的資訊，再建立一個新的二維 list 將其儲存。因為想要查看所有評論的各情緒數量分布，所以使用 sum() 函數以快速的加總並取得最終的結果。此處需要說明一下沒有分析 negative 與 positive 此兩個情

緒的原因，由於這兩個情緒過於粗略，所以難以去詳細了解評論的情緒狀況，所以在篩選的過程中，決定將其排除在外，以更能專注於較為詳細的情緒資訊。從所儲存的 list 可以看到 anger、anticipation、disgust、fear、joy、sadness、surprise 以及 trust 的情緒總共分別為 21747、307713、10395、32271、259021、28520、152000 以及 402307。

取得所有評論的各情緒數量後，為了能使圖表能夠有降冪的效果，所以再對二維 list 進行 `sort_values()`，並且參數內加入 `ascending = False`，以符合呈現的目的。此外，大致看過所有評論的各情緒數量後，可以明顯看到正向情緒的數量顯著大於負向情緒的數量，所以為了能於圖表中加以區隔，所以透過固定情緒數量前四多的四種情緒的顏色為 #138808，至於後四者則固定為 #7F7F7F。藉由顏色的區分，再加上圓餅圖內面積與角度的差異與長條圖內長度的差別，能夠更為清楚的看出其中的不同之處。

為了更進一步探討同產業內不同企業間評論的情緒差異，所以選取其中三種產業，分別是科技業、顧問業與金融業。科技業選擇 Microsoft、Google 與 Meta；顧問業選擇 Deloitte 與 KPMG；至於金融業則是選擇 JPMorgan-Chase 與 HSBC。這一部份的操作其實大同小異，差別只在於選取的標的不同，像是如果是要選擇 Microsoft，則直接下指令 `[ 'Company' ] == 'Microsoft'` 即可，其他公司則依此類推，而後續製作圓餅圖與長條圖的方式，由於與製作整體十四萬篇評論的作法相似，所以直接將該部分複製即可得到最終的結果。

### 3 SYSTEM OUTCOMES

#### 3.1 員工評論正負向預測

首先，觀察以 Log-likelihood ratio 特徵選擇的字進行的分類結果，可以發現當選擇

8000 字的時候，所有的分類器有最好的結果，其中，SVM 的準確率最高 (65.6%)。而 XGBoost 有較差的結果表現。(如表三)

觀察以 Chi-square 特徵選擇的字進行的分類結果，可以發現當選擇 4000 字的時候 Naive Bayes 的準確率最高 (65.6%)。(如表四)

在以 Log-likelihood ratio 進行特徵選擇時可以發現當選擇的字數越多的時候，模型表現結果越好，因此，建議以 Log-likelihood ratio 進行特徵選擇時，可以盡量使用多一點的字典字數；反觀以 Chi-square 來進行特徵選擇時，則發現當模型中使用的特徵越多時，模型表現反而會下降，在字數的使用上應該是有一個臨界值，超過該數值以後，準確率會因此下降，因此，若以 Chi-square 進行特徵選擇時，則可以不用使用過多的字典字數。

LLR	2000 字	4000 字	5000 字	6000 字	8000 字
RF	63.4%	63.3%	63.3%	63.2%	65%
XGB	60.7%	60.7%	60.7%	59.9%	64.5%
SVM	62.6%	63.7%	64.1%	64%	65.6%
NB	63.9%	62.6%	62.8%	62.8%	64.7%

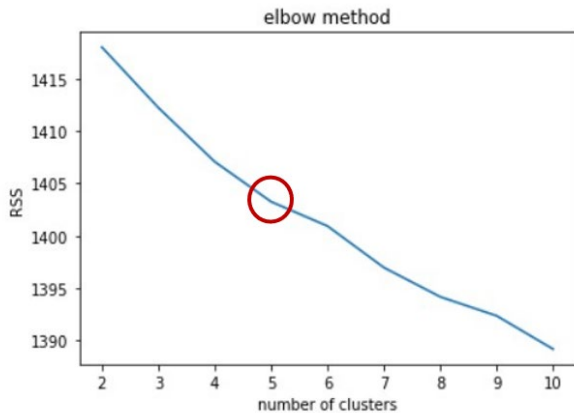
表三、LLR 特徵選擇與各分類器的模型表現 (accuracy 的部分)

Chi-square	2000 字	4000 字	5000 字	6000 字	8000 字
RF	63.3%	63.9%	64.7%	63.7%	64.3%
XGB	65%	64%	64%	64%	63.4%
SVM	63.7%	64.3%	64.8%	64.9%	64.6%
NB	63.6%	65.6%	65.3%	65.2%	64.9%

表四、Chi-square 特徵選擇與各分類器的模型表現 (accuracy 的部分)

#### 3.2 探討員工關注面向 (詞)

使用 K-means 搭配 Elbow method 後，執行結果如下圖所示：



圖一、K-means Elbow method

根據上圖所示，分群數為 5 群時，RSS 值下降有變緩的趨勢，因此本專案選擇以 5 群作為合適的分群數。接著，設定 K 值（分群數）為 5 群，進行 K-means 分群，並針對分群後的 5 群評論，計算每個字的 count \* PMI 值，挑選數值最高的前 10 個字，作為各群的代表字，以瞭解各群的主題。以下是 K-means 分 5 群的結果：

```
Topic 0:
['exposure', 'projects', 'growth', 'opportunities', 'good', 'place', 'work', 'great', 'life', 'balance']
Topic 1:
['experience', 'skills', 'hgs', 'work', 'working', 'best', 'place', 'nice', 'environment', 'good']
Topic 2:
['lot', 'challenges', 'technology', 'everyday', 'learning', 'day', 'technologies', 'things', 'learn', 'new']
Topic 3:
['typical', 'management', 'learned', 'day', 'job', 'hardest', 'enjoyable', 'part', 'workers', 'co']
Topic 4:
['worked', 'products', 'hr', 'much', 'services', 'employee', 'quality', 'project', 'bank', 'process']
```

圖二、K-means 分 5 群結果

根據上圖分群結果，可以發現到主題 1（群 1）中的評論，提及了工作與生活的平衡，而主題 2、主題 3 和主題 4，則分別提到了工作環境、工作上的學習以及工作中的管理，由此可見，企業內部的環境與工作生活是企業員工會關注的面向。

另外，為了找出更多的評論主題面向，本專案也使用 K-means 將評論分成 10 群進行分析。由於分 10 群的結果與分 5 群的結果有部分主題重複，因此以下呈現較無重複的主題內容：

```
Topic 0:
['hike', 'salary', 'career', 'security', 'growth', 'excellent', 'good', 'work', 'life', 'balance']
Topic 1:
['usa', 'press', 'cost', 'writing', 'claims', 'time', 'tasks', 'thing', 'needs', 'accenture']
Topic 2:
['typical', 'management', 'learned', 'hardest', 'job', 'day', 'enjoyable', 'workers', 'co', 'part']
Topic 3:
['company', 'training', 'oracle', 'new', 'got', 'one', 'learn', 'opportunity', 'technology', 'best']
Topic 4:
['endeavors', 'feedback', 'experience', 'us', 'glad', 'hear', 'thank', 'hi', 'wish', 'hgs']
Topic 5:
['enjoy', 'place', 'location', 'hard', 'work', 'good', 'environment', 'working', 'fun', 'nice']
```

圖三、K-means 分 10 群結果（呈現主題較無重複的部分）

根據上圖的分群結果，可以發現評論也提到了工作薪水，因此工作的報酬也是員工們會關注的面向。

除了上述分群搭配特徵選擇的作法，我們也使用了 SVD 及 LDA 進行評論的主題分析，以觀察結果是否與分群有所不同。在本專案中我們同樣依據 Elbow method 的結果，挑選 5 個主題進行觀察，首先是 SVD 主題分析的部分，以下為分析結果：

```
Topic 0:
['life', 'environment', 'job', 'culture', 'management', 'great', 'place', 'company', 'work', 'good']
Topic 1:
['job', 'balance', 'enjoyable', 'place', 'great', 'day', 'workers', 'learned', 'part', 'co']
Topic 2:
['company', 'great', 'part', 'lot', 'job', 'life', 'balance', 'new', 'learn', 'things']
Topic 3:
['learning', 'workers', 'nice', 'also', 'co', 'opportunities', 'company', 'best', 'place', 'great']
Topic 4:
['balance', 'working', 'team', 'good', 'life', 'company', 'place', 'best', 'environment', 'nice']
```

圖四、SVD 分 5 群結果

根據上圖 SVD 的分析結果，可以發現主題分析的結果與分群有所不同，主題 1 和主題 2 的評論分別提到了企業的文化及工作的地點，可見員工也會關注這兩個面向。

接著是 LDA 主題分析的部分，以下為分析結果：

```
Topic 0:
['different', 'company', 'projects', 'learning', 'employees', 'new', 'management', 'time', 'good', 'team']
Topic 1:
['business', 'one', 'years', 'experience', 'banking', 'company', 'hdfc', 'employees', 'management', 'bank']
Topic 2:
['experience', 'customers', 'daily', 'hi', 'handling', 'process', 'team', 'customer', 'hgs', 'management']
Topic 3:
['life', 'environment', 'culture', 'working', 'great', 'place', 'management', 'company', 'good', 'work']
Topic 4:
['new', 'workers', 'day', 'working', 'co', 'enjoyable', 'learned', 'job', 'work', 'part']
```

圖五、LDA 分 5 群結果

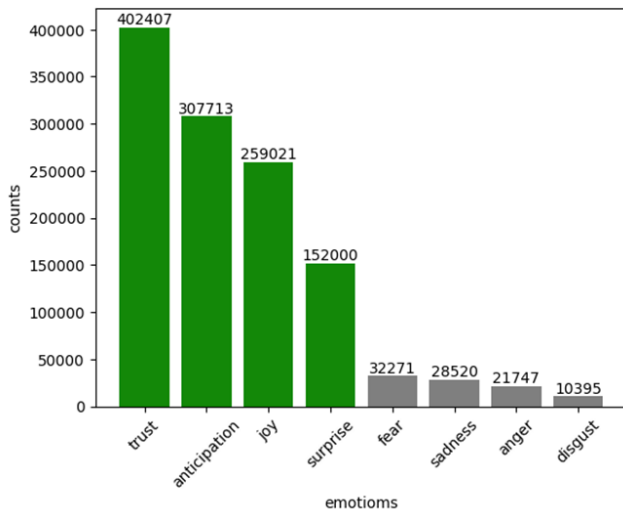
根據上圖 LDA 的分析結果，我們發現主題 4 涵蓋了前面分群的結果，評論提到了工



作生活、工作環境與企業文化，可見這三個面向是員工會關注的重點。

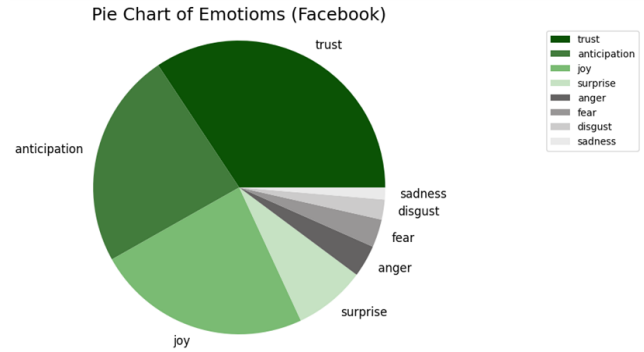
### 3.3 進行評論情緒用字統計

第三個部分是進行評論文本的情緒分析，使用的是 NRC 情緒字典，藉此以觀察企業員工的用字情緒的分布狀況。以下會透過長條圖與圓餅圖，來呈現整體公司評論的情緒用字情況，以及同產業不同公司間的比較與差異。

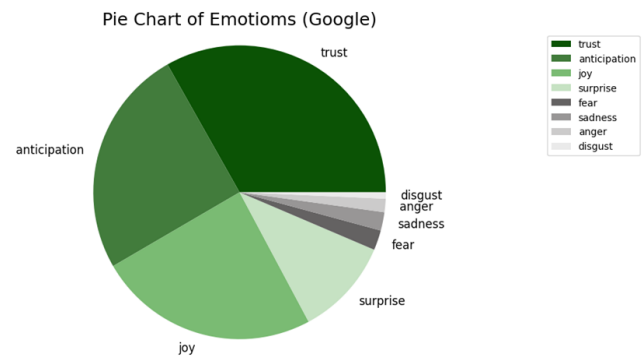


圖六、情緒用字分布圖

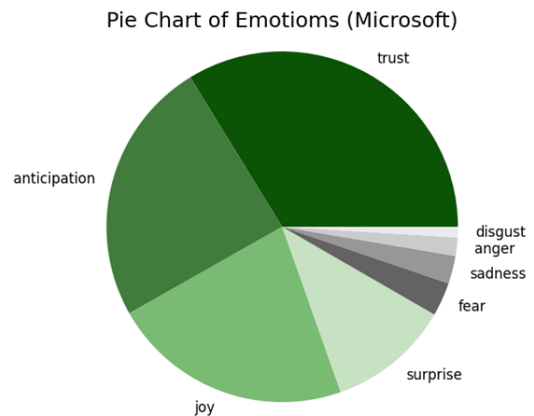
從圖二的結果可知，整體評論內容以正向情緒為主，比重由大到小排序為信任、參與感、歡樂、驚喜的情緒；相反的，負向情緒用詞的比例低，比重由大到小排序為恐懼、難過、憤怒、厭惡的情緒，但比例之間的差距落差不大。可見所取得之公司，其所獲得的評論皆以正向為主，代表整體而言，大部分的公司皆有不錯的評價，因此，如能針對差評所述的內容去做調整與改善，相信將使公司獲得更多人的青睞。



圖七、Facebook 情緒用字分布圖



圖八、Google 情緒用字分布圖

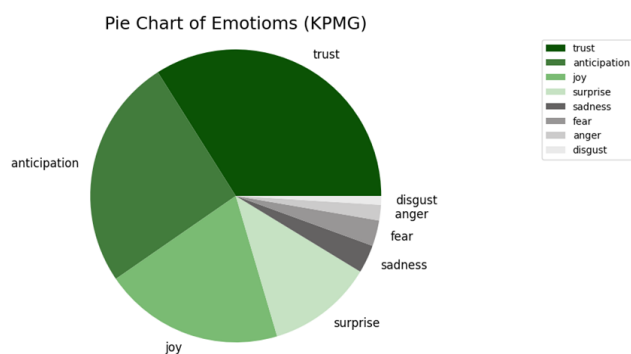


圖九、Microsoft 情緒用字分布圖

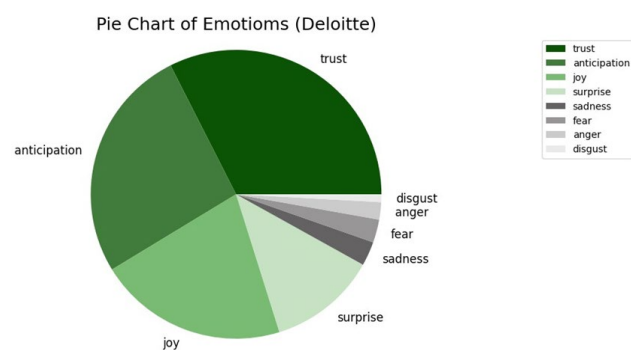
從以上三張圖(圖三、圖四與圖五)可以看到 Microsoft、Google 與 Meta 與三家公司評論的情緒相似度相當高，本專案認為是由於此三家公司皆屬於科技業，且服務性質內容相似，因此造就工作環境和企業文化相近。然而，因為此三家企業所從事的營運內容有本質上的不同，Google 是以搜尋引擎為主、而 Meta 是以社群軟體為主，至於 Microsoft 則是以電

腦系統為主，因而造就不同的工作環境和壓力，企業文化也會較為不相似，導致員工對企業的觀感存在些微上的差異。

至於負向情緒的部分，可以發現 Google 的負向情緒為三者中最低，本專案覺得會有此結果可以歸功於其給予員工的多項福利<sup>[5]</sup>，相比 Microsoft 的長工作時間與辦公室政治<sup>[6]</sup>以及 Meta 修改員工福利所造成員工的不滿<sup>[7]</sup>，Google 理應擁有三者中最佳的表現。



圖十、KPMG 情緒用字分布圖

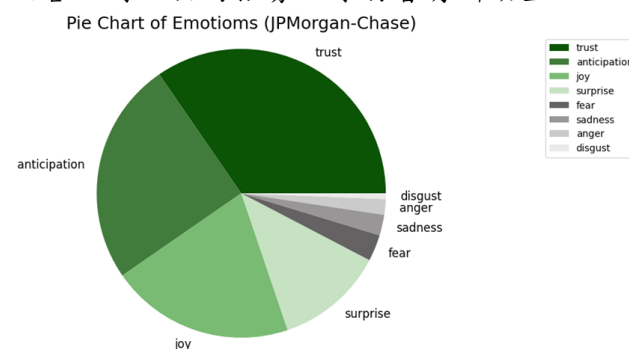


圖十一、Deloitte 情緒用字分布圖

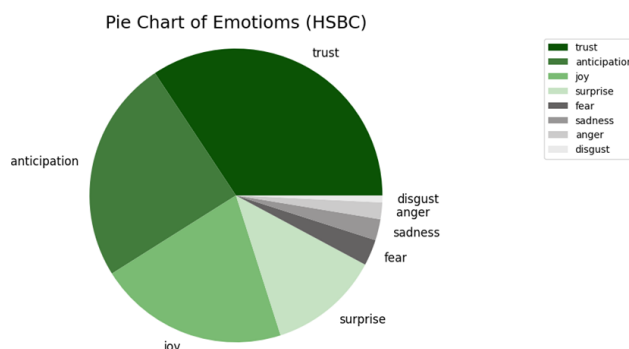
根據上圖(圖六與圖七)的結果，可以看到其實以上兩家公司評論的情緒非常類似，由於皆為會計師事務所，所以無論是公司環境，抑或是工作模式等等皆雷同，所以會有此結果實屬合理且可預期。然而令本專案意外的是占大多數的情緒皆是以正向為主，與校園內學生的想法有所出入，本專案認為是由於本研究係以國外為主，而非台灣的四大事務所，所以較

不會出現像本國一樣在忙季時，必須要熬夜工作的情況，此外，由於國外的顧問業佔事務所的比例比台灣還要高，意謂審計服務的比例相對較低，因此繁雜且須人工翻憑證的情形也較少，故會有較高比例的正向評論，實屬合情合理。

針對負向情緒的部分，由於不可能一家公司皆無負評，而以上兩家也只有少部分負面的評論，因此可以不用太過於擔憂，但是如能從中發現公司所存在可以改善的問題，並加以改善，對於公司招募人才將會有所助益。



圖十二、JPMorgan-Chase 情緒用字分布圖



圖十三、HSBC 情緒用字分布圖

最後則是檢視兩家皆屬金融業的公司評論的情緒狀況。從上圖(圖八與圖九)可以發現兩家公司評論的情緒比例並無相差太多，本專案認為金融業所從事的工作性質類似，且環境也極其相似，因此差別只在於公司給予員工的福利，而造成評論的情緒有所不同。

至於負面情緒的部分，本專案認為最主要的原因仍為工作上的壓力與績效的考核，但是

隨著近年來，金融業不斷朝向數位化發展，因此過去重複且毫無經濟附加價值的事情將逐漸被流程自動化所取代，進而減少員工將時間花在相對無意義的事情上，反而有更多的時間投入較需要專業知識的工作，如此一來，將能減少員工的壓力，對於員工之於公司的負面情緒也能因此減少。

## 4 CONCLUSIONS

### 4.1 評論正負向預測

使用了 4 種分類器觀看預測表現，其中使用 NB、SVM 的準確率最高 (65.6%)。雖然在提升模型效能上，試著以搭配不同特徵選擇方式及字數挑選上的差異來進行改善，但發現分類表現仍不理想，未來可以朝更多不同特徵選擇方法，或不同的文字向量表示法像是 word2vec，或是透過調整模型參數，甚至使用深度學習模型多加嘗試，以提升分類表現。

### 4.2 員工關注面向分析

分析員工評論的關注面向，可以幫住求職者透過這些面向審視企業是否符合需求，進而挑選到適合自己的工作，另外，企業瞭解員工的關注面向，也有助於改善不足之處，給予員工更好的工作環境。

本專案透過分群、SVD 及 LDA 方法，發現員工會關注生活與工作的平衡、工作地點和環境、薪水報酬、工作上的學習、公司內部的管理以及公司文化等面向。

求職者可以針對企業的文化、工作地點、工作環境與管理、工作報酬等面向，觀察一間公司是否在這些面向上符合自己的需求，除此之外，求職者也可以注意工作與生活平衡的問題，審視企業所提供的工作是否會佔據生活太多時間。

而對於企業而言，由於員工會關注企業文化、薪水、工作環境與企業管理，因此企業應

在這些面向上多加注意，並分析自身在這些面向上需要改善的地方，進一步調整，以給予員工良好的工作環境，有助於提升員工的表現，提升長期的企業收益，並且可以增加員工對於公司的正向評論。

### 4.3 員工情緒用字分析

透過使用 NRC 情緒字典，可以從評論中分析出情緒字的分布情況，其中正面的用字較多，而在三家科技業公司中，Google 的正向情緒用字比例最高，Facebook 的憤怒、厭惡用字比例較多。至於在顧問業與金融業內所挑選的公司，則沒有太大的差別。

本專案認為由於所找的公司皆為該產業內較具代表性的企業，所以推測公司的員工對於企業的向心力也會較強，因此大部分的情緒仍是以正向為主，如果日後能夠挑選較為有名的公司，並且加以分析與研究，或許就能看出顯著的差異。

此外，由於本專案的分類依據係以 rating 作為標準，所以依此去判斷為正向或負向的評論可能過於簡易，因此，如能將其改以情緒分析後正負向情緒的數量作為評判標準，或許會更為精確，進而提升準確率。然而因為各篇評論的字數皆不相同，所以藉由除以該篇評論字數後的數值等其他類似的方法，將能排除因評論字數不同所導致情緒數量的差異，亦即能達到去除評論字數規模化的效果。

## 5 REFERENCES

[1] Muhammad Saqlain Rehan, Furqan Rustam, Saleem Ullah, Safdar Hussain, Arif Mehmood & Gyu Sang Choi (2022). Employees reviews classification and evaluation (ERCE) model using supervised machine learning approaches, Journal of Ambient Intelligence and Humanized Computing, volume 13, pages 3119–3136



[2] Kowsari K, Jafari Meimandi K, Heidarysafa M, Mendu S, Barnes L, Brown D. Text Classification Algorithms: A Survey. Information. 2019; 10(4):150. <https://doi.org/10.3390/info10040150>

[3] Sumbal Riaz, Mehvish Fatima, M. Kamran & M. Wasif Nisar, Opinion mining on large scale data using sentiment analysis and k-means clustering, Cluster Comput 22 (Suppl 3), 7149–7164 (2019).

[4] Yeonjae Jung & Yongmoo Suh, Mining the voice of employees: A text mining approach to identifying and analyzing job satisfaction factors from online employee reviews, Decision Support Systems, volume 123, August 2019, 113074

[5] Google 令人眼紅的十四個大福利

[6] 微軟員工盤點公司五大優缺點- 國際 - 中時新聞網

[7] 職員疫後重返 Meta 辦公室 卻被告知削減多項員工福利