

本次個人作業中，我主要使用 Jaccard Coefficient 與 Association Rule 為主要選擇相關詞的方式，並去除「鴻海」本身。在計算 MI 值時，由於鴻海新聞中大多都有「鴻海」二字。因此，在計算某關鍵字與「鴻海」的交集時通常就是該關鍵字的 DF 值。因此，MI 值大多都為 $\text{LOG}(1/\text{DF}(\text{「鴻海」})) = -3.31429$ ，所以較無判斷的顯著性。

term	MI
INCJ	-3.3142887
人力	-3.3142887
人工智慧	-3.3142887
上漲	-3.3142887
下挫	-3.3142887
下滑	-3.3142887
中鋼	-3.3142887
元件	-3.3142887
升息	-3.3142887
友達	-3.3142887
支付	-3.3142887
日月光	-3.3142887
日經新聞	-3.3142887

在計算 Jaccard Coefficient 時，我先利用 Python 計算每個關鍵字的 df 值、與鴻海共同出現的篇數（鴻海與該關鍵字的交集篇數）、鴻海出現新聞與該關鍵字出現新聞的聯集篇數。藉由這些數字，我們可以利用

$\frac{N(\text{鴻海} \cap \text{關鍵字})}{N(\text{鴻海} \cup \text{關鍵字})}$ 去計算出該關鍵字的 Jaccard Coefficient。藉由這些數字，我們可以

做出初步的關鍵字選擇，也可以作為 Support，最為選擇「鴻海」關鍵字的門檻。在此，我利用 Jaccard Coefficient 選出了「夏普、集團、日本、台灣、報導、郭台銘、蘋果、技術、台積電、指數、合作、美國、外資、大陸、企業、法人、金額、手機、上漲、機會」這 20 個與「鴻海」相關的字詞。

計算「鴻海→某關鍵字」的 Confidence 時，我們可以利用 $\frac{N(\text{鴻海} \cap \text{關鍵字})}{N(\text{鴻海})}$

去計算出該關鍵字的 Confidence。藉由這些數字，我們可以做出進一步的關鍵字選擇，也可以配合 Support 當作選擇關鍵字的門檻。

在此，若未設定 Support 門檻會得到與 Jaccard Coefficient 相同的結果。由於這一題 $\text{DF}(\text{鴻海})$ 與 $\text{DF}(\text{鴻海 OR 關鍵字})$ 大多相同，因此計算出來的 Confidence 與 Jaccard Coefficient 大致相同，藉此選出的關鍵字也相同。

若設定 Support $\geq 5\%$ 的門檻（大約等於所有 Support 的平均），「鴻海→關鍵字」中我選出了「日月光、月營收、INCJ、平台、萬元、友達、富士康、戴正吳、和碩、盤面、財報、召開、業績、總裁、供應鏈、開盤、個股、股東、社長、權值股」這 20 個與「鴻海」相關的字詞。

term	鴻海→x	JACOB
日月光	0.05043647	0.05043647
月營收	0.0528613	0.0528613
INCJ	0.0557711	0.0557711
平台	0.05868089	0.05868089
萬元	0.06013579	0.06004843
友達	0.06401552	0.06401552
富士康	0.06983511	0.06976744
戴正吳	0.07274491	0.07274491
和碩	0.07419981	0.07419981
盤面	0.07613967	0.07613967
財報	0.09408341	0.09408341
召開	0.10038797	0.10038797
業績	0.10475267	0.10470189
總裁	0.10669253	0.10658915
供應鏈	0.10766246	0.10766246
開盤	0.10766246	0.10766246
個股	0.11008729	0.11008729
股東	0.11348206	0.11348206
社長	0.11542192	0.11542192
權值股	0.11784675	0.11784675

若設定 Support $\geq 6\%$ 的門檻（剛好選出 20 個 Confidence=1 關鍵字的 Support 門檻），「關鍵字→鴻海」我選出了「夏普、蘋果、台積電、指數、外資、法人、上漲、銷售、權值股、社長、股東、個股、供應鏈、開盤、召開、財報、盤面、和碩、戴正吳、友達」這 20 個與「鴻海」相關的字詞。

term	x→鴻海	JACOB
夏普	1	0.41367604
蘋果	1	0.24199806
台積電	1	0.23278371
指數	1	0.22890398
外資	1	0.19641125
法人	1	0.18622696
上漲	1	0.16197866
銷售	1	0.13482056
權值股	1	0.11784675
社長	1	0.11542192
股東	1	0.11348206
個股	1	0.11008729
供應鏈	1	0.10766246
開盤	1	0.10766246
召開	1	0.10038797
財報	1	0.09408341
盤面	1	0.07613967
和碩	1	0.07419981
戴正吳	1	0.07274491
友達	1	0.06401552