

電影產業個人推薦化系統

利用NLP推薦系統分析電影產業，打造個人化推薦，提升用戶體驗並創造商業價值

電影產業痛點



- OTT 平台競爭激烈，用戶忠誠度低

現今影音串流平台眾多，消費者選擇性高



- 傳統推薦方式不夠精準，影響觀看率

傳統推薦較大眾化、無法吻合消費者個人喜好



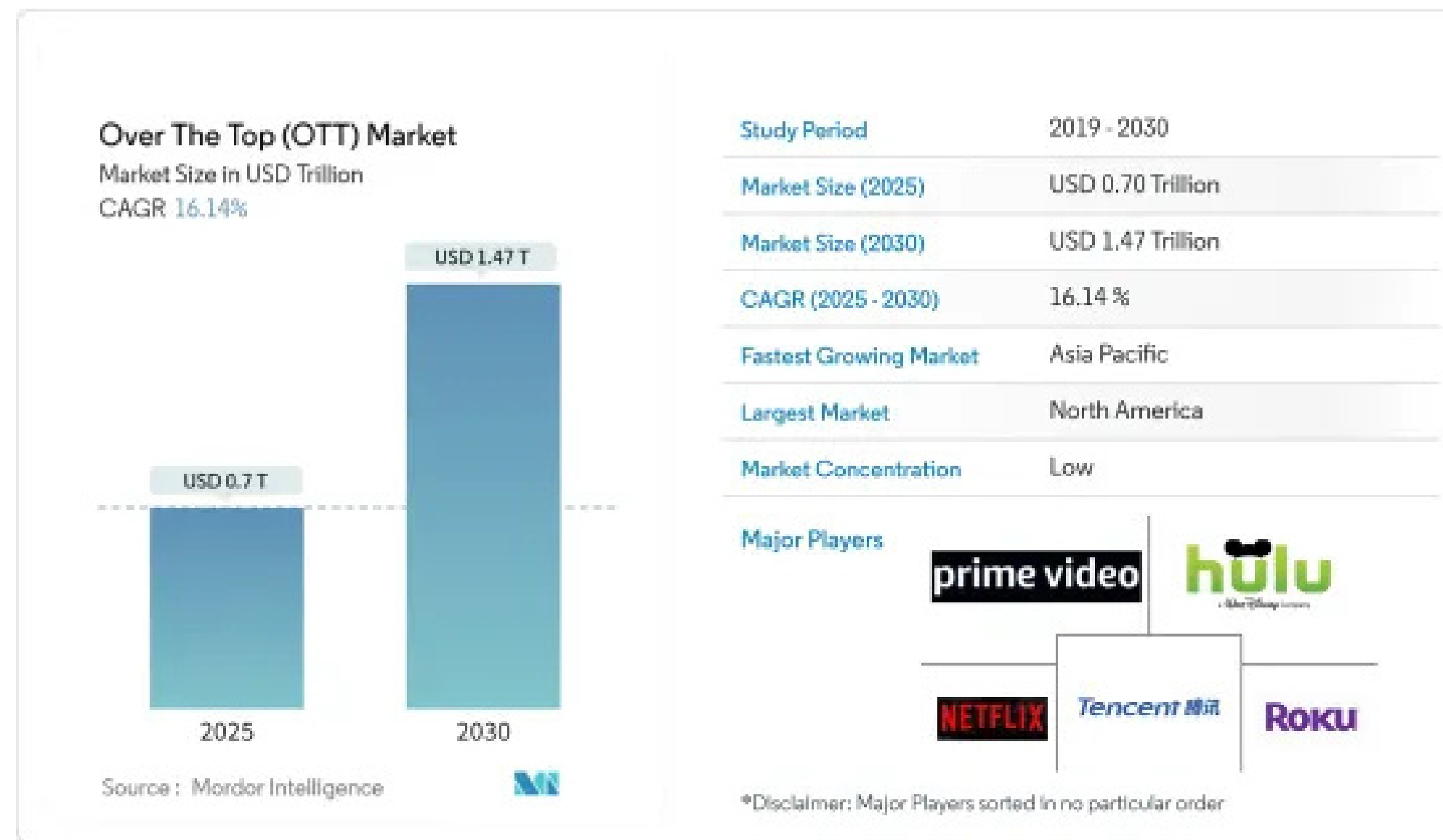
- 影視行銷方式不夠有效，難以精準投放廣告

無法精確地投放廣告於有興趣之消費者

市場機會

全球OTT規模不斷擴大

- OTT市場規模預計2025年達到7000億美元和2030年更將達到1.47兆美元
- 影視數據分析市場不斷地成長



資料來源：日商環球訊息有限公司

技術架構與方法

資料集	The Movies Dataset	
	credits.csv (電影角色、工作人員名單)	keywords.csv (電影關鍵詞)
	links.csv links_small.csv (電影IMDB、TMDB編號)	ratings.csv ratings_small.csv (使用者對電影的評分)
	movies_metadata.csv (電影主資料)	
使用技術	TF-IDF、cosine similarity、基於內容的推薦系統	

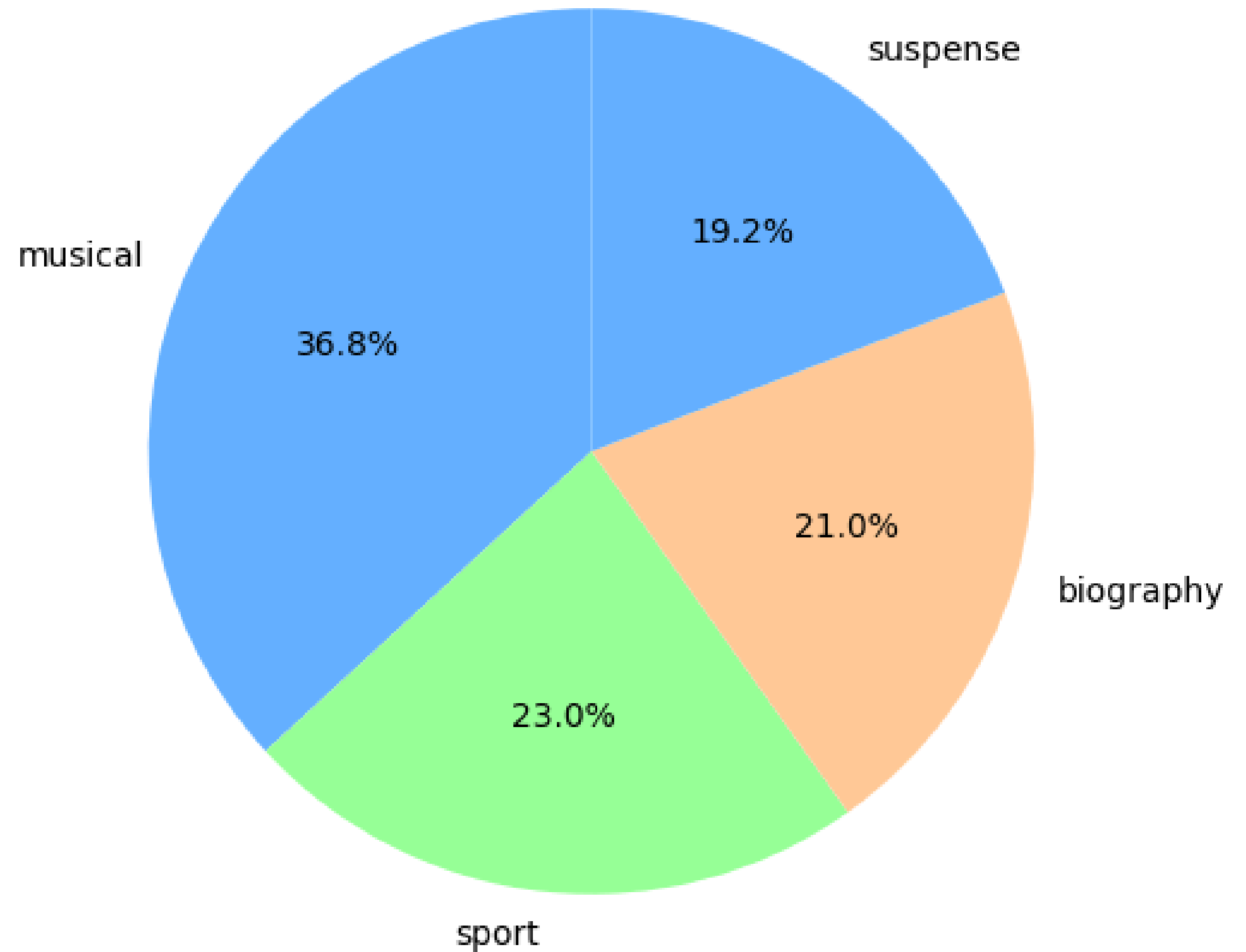
結果呈現

電影關鍵詞之圖形視覺化

電影關鍵詞較具有特殊性且詞彙次數出現次數超過100次以上的有：

- 1.musical
- 2.sport
- 3.biography
- 4.suspense

TOP 4 Movies Keywords



結果呈現

基於電影關鍵詞之推薦系統

search_movies_names：搜尋欲搜尋的電影名稱
get_top_k_similar_movies：找出前五個推薦的電影

```
search_movies_names
```

	id	original_title
147	568	Apollo 13
897	62	2001: A Space Odyssey
1043	116356	La vie sexuelle des Belges 1950-1978
1161	389	12 Angry Men
1318	11674	101 Dalmatians
...
45123	71576	Buenos Aires 100 KM
45148	438910	Konstruktor krasnogo tsveta -1993
45325	190817	La Fille du 14 juillet
45329	43656	13 Moons
45381	264269	Vladieniye 18

568 rows × 2 columns

```
movie_idx,search_info,top_movies_name = get_top_k_similar_movies("")

print("搜尋的電影名稱為：\n",movie_idx,search_info,"\n")
print("Top 5 推薦電影：\n",top_movies_name)
```

搜尋的電影名稱為：
1893 Star Wars: Episode I - The Phantom Menace

Top 5 推薦電影：

	電影id	電影名稱	Cos相似度分數
0	69315	Battlestar Galactica: Razor	0.794426
1	15467	Kismet Konnection	0.794426
2	313297	Kubo and the Two Strings	0.794426
3	70703	The Captains	0.794426
4	25926	Mind Meld: Secrets Behind the Voyage of a Life...	0.794426



未來展望

可改善目標

- 利用flask來建立python flask API以應用落地。
 - 可使用其他技術(例如：Word2Vec或BERT)來比較推薦準確結果。
-

未來目標

- 加入使用者對電影的評分預測並結合現有的推薦系統來達成**混合式推薦系統**，解決冷處理問題。
- 可考慮使用**深度學習**來建立推薦系統，可以更精確地達成電影的推薦。