

在相似判決書中，我們需要比對詞嵌入向量及詞頻的方式，何種的效果比較好，以人工檢查的方式評估效果。

詞嵌入是使用風險詞加上下文作為輸入資料，以此能夠獲取文章中的語意及情境，將找到的 10 筆相似判決書輸出到 CSV 檔中。

|    | A                                | B | C | D | E | F | G |
|----|----------------------------------|---|---|---|---|---|---|
| 1  | 0                                |   |   |   |   |   |   |
| 2  | 事實一、林家瑋(一)前於民國101年間，因違反戶籍法及竊盜案件， |   |   |   |   |   |   |
| 3  | 事實及理由一、按刑事訴訟法第449條第1項規定之案件，檢察官   |   |   |   |   |   |   |
| 4  | 理由一、非常上訴意旨略以：「一、按判決不適用法則或適用不     |   |   |   |   |   |   |
| 5  | 事實為有罪之陳述，經裁定改由受命法官獨任依簡式審判程序審     |   |   |   |   |   |   |
| 6  | 事實一、詹正豪前因竊盜及毒品等案件，經本院於民國（下同）     |   |   |   |   |   |   |
| 7  | 事實一、詹正豪前因竊盜及毒品等案件，經本院於民國（下同）     |   |   |   |   |   |   |
| 8  | 事實及理由一、本件犯罪事實及證據，除證據部分應補充「被告     |   |   |   |   |   |   |
| 9  | 事實一、謝清林前於民國96年間因竊盜、傷害、妨害自由等案件    |   |   |   |   |   |   |
| 10 | 事實一、施純金前於民國98年間，因竊盜案件，經原審法院98年   |   |   |   |   |   |   |
| 11 | 事實及理由一、犯罪事實：向貴興(一)前於民國101年間，因竊盜第 |   |   |   |   |   |   |

詞頻只使用風險詞經過 Jieba 斷詞，以 cos 尋找相似判決書，cos 的值愈低代表愈相似。

|    | A                              | B | C | D | E | F | G |
|----|--------------------------------|---|---|---|---|---|---|
| 1  | 事實一、林柏孝前因酒後駕車之公共危險案件，經本院以99年度科 |   |   |   |   |   |   |
| 2  | 事實及理由一、本件犯罪事實及證據，均引用如附件檢察官起訴書  |   |   |   |   |   |   |
| 3  | 理由壹、本件非常上訴理由稱：「一、按判決不適用法則或適用不  |   |   |   |   |   |   |
| 4  | 事實及理由一、本案犯罪事實及證據，除證據欄所載：「告訴人」  |   |   |   |   |   |   |
| 5  | 事實一、鐘○盛前因施用毒品案件，經臺灣士林地方法院以103年 |   |   |   |   |   |   |
| 6  | 事實及理由一、犯罪事實：林永興基於意圖為自己不法所有之竊盜  |   |   |   |   |   |   |
| 7  | 事實一、羅光興基於意圖為自己不法所有之竊盜犯意，於民國105 |   |   |   |   |   |   |
| 8  | 事實一、簡新春前於民國99年間，因竊盜案件，經臺灣桃園地方  |   |   |   |   |   |   |
| 9  | 事實及理由一、事實：(一)王騏賓有下列前科及徒刑執行之情形： |   |   |   |   |   |   |
| 10 | 事實及理由一、犯罪事實：吳凱群前於民國99年間，因傷害案件  |   |   |   |   |   |   |

我們針對每一個案件類型，各輸入 5 筆資料，詞嵌入相似判決書和詞頻相似判決書各 50 筆共計 100 筆判決書進行人工檢查，以人工檢查方式，查看是否相似，以此做為評估標準。

|       |     |       |  |       |     |       |
|-------|-----|-------|--|-------|-----|-------|
| 毒品    | EMB | TFIDF |  | 過失傷害  | EMB | TFIDF |
| TOTAL | 32  | 26    |  | TOTAL | 27  | 26    |
|       |     |       |  |       |     |       |
| 偽造文書  | EMB | TFIDF |  | 公共危險  | EMB | TFIDF |
| TOTAL | 26  | 16    |  | TOTAL | 30  | 34    |
|       |     |       |  |       |     |       |
| 賭博    | EMB | TFIDF |  | 竊盜    | EMB | TFIDF |
| TOTAL | 23  | 22    |  | TOTAL | 27  | 21    |