

# 第一次會議記錄

1. 跟mentor要看看資料
2. 文字擷取相關資料：結巴、stanford NLP
3. 題目簡報提到的Scordcard是什麼意思
4. 找影片修改軟體
5. 有哪些新聞網站需要爬

財金新聞相關網站：鉅亨網、工商時報、Money DJ、

step

1. 資料建模
  - a. 結巴 (斷詞)
    - i. 可以先試著斷一些例句的詞
    - ii. 轉換成對應的向量
    - iii. 斷詞拿出來，先根據順序做編號
    - iv. TFIDF
  - b. 清理資料
  - c. 做模型
    - i. Baseline 要比什麼比較好
    - ii. 想辦法用機器做到跟人一樣
    - iii.
2. 資料爬蟲
  - a. 爬新新聞
  - b. EDA
3. 股價分析(Bonus)
4. 做網站
5. 做影片
6. 判斷股價

4/1 ~4/6 看影片、找切割字相關套件、找架網站的資料

4/6 有資料了，開始對照新聞是否有重複資料

4/8 20:30 線上討論，開始建模

4/8~4/29 完成資料建模

4/29~5/19 開始資料爬蟲

5/19~ 5/26 開始做網站

5/26~6/18 做股價分析or緩衝期

6/18 交影片

7/2 報告