金融大數據 2023/4/28 筆記

* V1

Spam → 垃圾訊息

Ham → 正常訊息

* 此重要最重要的是Review Text欄位(不可以刪除)
* [S1E10] Making Sense of Text Analytics | 5 Minutes With Ingo

<https://www.youtube.com/watch?v=FsP3Z5ieos8>

Tokenize 斷詞斷句

Sentiment

* Transform Cases 大小寫轉換

SVM → 只能用兩類

30% Testing Data也要做斷詞斷句處理

* 重要步驟

Nominal to Text

30% Testing Data也要做斷詞斷句處理Process Documents to Data，並且要給word，讓它知道如何斷句。