元智大學資管系

第三十屆學術類畢業專題頂石課程(二)

期末報告

以自然語言處理與情感分析探討台灣上市公司財報文本情緒趨勢

1111614 邱奕銓、1111620 陳彥瑋、1111648 陳造翊

工作代號：ZV3

指導教授：禹良治 教授

中華民國 114 年 11 月

Nov, 2025

目錄

Chapter 1 導論

1.1 研究動機

近年來，人工智慧(Artificial Intelligence, AI)與自然語言處理(Natural Language Processing, NLP）技術的蓬勃發展，使得人類得以從大量文字資料中萃取潛在資訊與情緒意涵。以往在金融研究中，投資分析主要聚焦於財務報表、會計指標與技術面資料，然而這些數值資料多屬於「事後結果」，無法及時反映企業管理階層在財報中所傳遞的語氣與態度。實際上，企業在財報中的文字敘述常帶有特定情緒，例如對市場前景的樂觀、對成本壓力的擔憂，或是對策略轉型的信心等，這些語意特徵往往能揭示企業潛在經營狀況與市場信心，對投資人與分析師具有高度參考價值。

在國際研究中，財報文本分析（Financial Text Mining）已被廣泛應用於投資決策與風險評估。Loughran and McDonald（2011）首次建構財務詞典，揭示財報語氣與公司績效間的關聯，後續研究更發展出以情感分析（Sentiment Analysis）與主題模型（Topic Modeling）結合的分析架構。然而，這些研究大多以英文語料為主，對中文財報文本的分析仍相對稀少。台灣上市公司每年皆需公告股東報告書及財務報表，這些公開文件蘊含豐富的文字資料，若能運用 NLP 技術進行情緒標註與分析，將有助於理解企業在不同時期的情緒變化與財務走勢之間的關聯。

因此，本研究以台灣上市公司之財報文本為研究對象，結合自然語言處理與情感分析技術，嘗試以細粒度的「層面式情感分析（Aspect-Based Sentiment Analysis, ABSA）」方式進行資料標註與建模，從語義層面探討企業文字中蘊含的情緒趨勢。藉由人工標註、模型訓練與自動化預測，期望能建立一套客觀可量化的財報情緒分析流程，補足傳統數據分析在「語意層面」上的不足。

1.2 研究目的

本研究的主要目的是建立一套適用於中文財報語料的情感分析流程，並透過 NLP 模型自動化辨識企業財報中的情緒傾向。具體研究目標如下：

1. 蒐集與整理台灣上市公司財報文本資料：彙整不同產業之年度財報與股東報告書內容，建立具代表性的財報語料庫。

2. 人工標註財報文本情緒資料集：以「主題（Aspect）」、「面向（Category）」、「觀點（Opinion）」及「情緒極性（Sentiment）」四元組格式（Quadruple）標註句子內容，確保模型訓練資料具一致性與準確性。

3. 導入機器學習與深度學習模型：使用 MT5 等多語言預訓練模型進行微調（Fine-tuning），比較不同模型在財報情緒辨識任務之表現差異。

4. 建立財報情緒分析流程與指標化系統：將模型預測結果轉換為可量化之「財報情緒指數」，用以分析企業或產業整體的情緒趨勢。

5. 評估模型效能並提出改進建議：以 F1-score、Precision、Recall 等指標進行績效衡量，並針對標註誤差與資料增強（Data Augmentation）進行優化，探討模型於財務文本領域的可行性。

透過上述研究流程，本研究期望建立出一個可自動化運作、可擴展至多產業的財報文本情感分析框架，為財務資訊研究帶來新的視角與技術應用價值。

1.3 研究的重要性

本研究的重要性可分為三個層面進行說明：

1. 學術層面：

本研究以中文財報文本為分析對象，補足過去情感分析研究以英文語料為主的不足，並將 ABSA 技術導入財務文本分析領域。透過人工標註與模型訓練的結合，能提供國內外學術界一個中文財報語料庫的建立與應用範例，並為後續研究提供基礎。

2. 實務應用層面：

財報文本中的語氣與情緒往往能提前揭示企業的經營風向與市場信心。若能將本研究所建立的模型應用於金融分析或投資決策中，投資人與研究機構便能即

時掌握企業文字中的情緒變化，作為輔助判斷依據，進而降低決策風險，提升分析效率。

3.社會與科技層面：

隨著人工智慧在各領域的滲透，文本情緒分析不僅能提升資訊透明度，更能促進金融市場的理性化與自動化。本研究以 NLP 技術為核心，結合財務文本資料的語意分析，有助於推動智慧金融與企業數位轉型，展現自然語言處理技術在社會科學與管理學領域的跨域應用潛力。

綜合以上，本研究不僅希望在技術上驗證 NLP 模型於財報文本的可行性，更期望能藉由情緒趨勢的量化，為金融研究與投資決策帶來創新且具有實際應用價值的分析方法。

Chapter 2 文獻探討

2.1 相關理論、研究回顧

本研究以自然語言處理（Natural Language Processing, NLP）與情感分析（Sentiment Analysis）為核心理論基礎，輔以層面式情感分析（Aspect-Based Sentiment Analysis, ABSA）模型進行實作。過去相關研究已證實文本語意能反映人類的情緒與態度，並可作為預測行為或事件的重要依據，因此，本節將分別探討 NLP 技術、情感分析理論以及其在財務文本與企業財報中的應用現況。

在自然語言處理的發展歷程中，早期主要依賴詞彙統計與語法分析（如 TF-IDF、N-gram 模型），但這些方法難以捕捉語意層次的上下文關係。隨著深度學習的興起，語言模型（Language Model）成為 NLP 的核心發展方向。特別是在 Google 推出的 BERT（Bidirectional Encoder Representations from Transformers；Devlin et al., 2019）與 OpenAI 的 GPT 模型問世後，研究者得以利用 Transformer 架構進行上下文雙向學習，大幅提升語意理解與情感辨識的精確度。後續的 T5（Raffel et al., 2020）及其多語版本 MT5，則將各種自然語言任務轉換為「文字到文字（Text-to-Text）」的統一格式，使模型在多語環境下的應用更加靈活。

情感分析的理論基礎源於心理語言學與計算語言學，主要目的是判定文本中所表達的情感極性（如正向、負向或中立）。在傳統的分類式方法中，研究者多以詞典（Lexicon-based）為基礎，如 Loughran & McDonald（2011）所建立的財務專屬情感詞典，即針對英文財報語料進行情緒標註，藉以判定公司年報與投資風險之間的關聯。然而，詞典式方法存在語境侷限與準確度不足的問題，因此近年學界多轉向以機器學習與深度學習為主的分析架構，如 SVM、LSTM、Transformer 等。

層面式情感分析（Aspect-Based Sentiment Analysis, ABSA）進一步細分文本中的意見目標與情緒對象，將句子拆解為「主題（Aspect）」、「面向（Category）」、「觀點（Opinion）」與「情緒（Sentiment）」四個層次。這種細粒度的分析能更準確地辨識出文本中不同主題的情緒傾向。本研究採用的 ABSA 方法，參考 SemEval（Semantic Evaluation）競賽所提供的資料結構與標註方式，並以多語言模型 MT5 進行微調（Fine-tuning），以適應中文財報文本

的語意特徵。

在財務領域的應用方面，Loughran & McDonald（2011）的研究開啟了以文字語氣分析企業財務狀況的先河。Li（2010）則指出企業年報中的語言長度與用詞複雜度與盈餘管理有顯著關聯；Engelberg（2008）進一步發現公司管理層在財報中所使用的語氣能預測股價反應。近年來，情感分析與深度學習模型被廣泛應用於財報情緒預測，例如 Chen et al.（2022）探討中文財報情緒與股價波動的關係，結果顯示模型能有效捕捉市場信心與公司表現之間的關聯。然而，多數研究仍以英文或簡體中文為主，針對台灣上市公司之繁體中文財報分析的研究仍屬稀少。

綜合前述理論，本研究基於 NLP、ABSA 與財務文本分析的學術基礎，嘗試將深度學習模型應用於台灣上市公司財報文本中，建立適用於繁體中文語料的財報情緒辨識模型，並驗證情感趨勢與企業經營績效之間的潛在關聯。

2.2 研究架構

根據前述研究動機與文獻回顧，本研究的整體架構主要分為四個階段：資料蒐集、資料前處理與標註、模型建構與訓練、模型評估。研究流程如下所述：

1.資料蒐集階段：

本研究以台灣上市公司之財報文本與股東報告書為主要資料來源，從公開資訊觀測站及公司年報中擷取文字內容，涵蓋金融、科技、製造等不同產業別，確保語料多樣性。

2.資料前處理與標註階段：

將原始文本進行斷詞（Tokenization）、清理（Normalization）與格式化（Segmentation），排除數字、符號與重複語句後，再由組員依照 ABSA 四元組格式（Aspect, Category, Opinion, Sentiment）進行人工標註，以建立高品質訓練資料集。

3.模型建構與訓練階段：

採用 MT5 模型進行微調訓練，輸入財報句子後，模型輸出對應的四元組情緒標籤。為提升準確率，本研究亦引入資料增強技術（Data Augmentation），將部分人工標註資料透過語意改寫與隨機替換產生新樣本，以擴充訓練集規模。

4.模型評估階段：

以 Precision、Recall 與 F1-score 作為主要評估指標，比較不同模型與參數設定的效能差異，並透過人工驗證檢視模型預測結果的合理性。

綜合而言，本研究的研究架構以 NLP 技術為核心，結合 ABSA 模型與人工標註資料，建立一套完整的財報情緒分析流程，期能在學術與實務上皆展現創新性與應用價值。

Chapter 3 研究方法

3.1 研究流程概述

本研究的整體流程如圖 1 所示，主要包含五個階段：(1) 財報資料蒐集；(2) 文本前處理；(3) 三元素標註資料建構；(4) 模型訓練；(5) 模型評估與誤差分析。

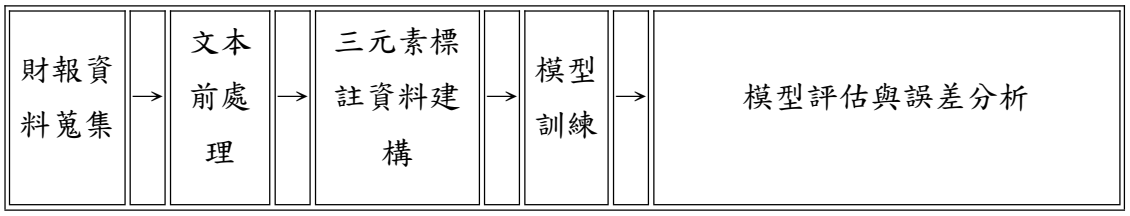

圖 1、研究流程圖

各階段內容如下：

（1）資料蒐集階段

本研究以台灣上市公司之年度報告與股東報告書為主要資料來源，資料取得自公開資訊觀測站（MOPS）及各公司官網。股東報告書中包含管理階層對營運表現、市場動能與未來展望的文字敘述，因此具備進行情緒分析的價值。

（2）資料前處理階段

由於財報文本篇幅龐大，本研究使用 AI 協助進行初步句子擷取，包括自動句子切分、關鍵詞比對（如「成長」、「下降」、「挑戰」等）與語意特徵篩選。接著由研究團隊人工審查與修正，排除無意義句、重複句或與營運無關的描述，確保資料品質。

（3）人工標註階段

本研究採用層面式情感分析(ABSA)作為標註架構，建立（Aspect, Opinion, Category）三元素資料集。由於情緒極性（Sentiment）較難以人工一致標註，本研究採用後續模型與規則式方法進行補齊，而人工標註階段主要專注於標註實體與意見詞之對應關係。

（4）模型訓練階段

本研究以 Hugging Face 平台之 MT5-small 為基礎，建立財報四元素預測

模型。訓練資料由人工標註的三元素資料與部分自動生成資料組成，模型輸入為句子文本，輸出為格式化的情緒四元組。

（5）模型評估階段

訓練完成後，本研究以 Precision、Recall 與 F1-score 評估模型在不同任務（Aspect、Category、Opinion、Sentiment、四元組）上的表現，並進一步檢視誤判原因，以提出可改進方向。

3.2 資料來源與蒐集方式

本研究的資料來源主要包含以下三類：

（1）公開資訊觀測站（MOPS）

本研究蒐集 2020 至 2024 年間上市公司年度報告與股東會年報摘要，特別聚焦於「管理層討論與分析（MD&A）」與「營運概況」章節，因其多以主觀語氣敘述營運成果與展望。

（2）產業涵蓋範圍

為避免資料偏向單一產業，本研究選取電子、金融、製造、傳產與服務業之公司，確保語料具多樣性與跨產業性。

（3）資料清理流程

蒐集後進行多輪清理，包括：

1.移除無文字內容（如表格、編號、附錄）

2.排除過短（少於 6 字）或過長（超過 50 字）的句子

3.去除重複句或明顯無情緒資訊的句子

4.最終共整理出約 1,000 句可用財報文本 作為本研究初始語料。

3.3 資料前處理與人工標註

（1）AI 句子擷取與前處理

由於財報文本內容龐大，本研究採用 AI 工具協助句子擷取，包括：

1.自動句子斷詞與分段

2.關鍵詞比對（如成長、下降、回升、挑戰等）

3.初步判斷是否涉及財務表現或市場資訊

接著由研究成員逐句審核，排除以下句型：

1.與營運無關（如公司沿革、社會責任口號等）

2.文意不完整

3.重複或無法辨識情緒方向

此流程能提升後續標註品質，兼具效率與可控性。

（2）標註架構設計（Aspect, Opinion, Category）

本研究參考 SemEval-2016 Task 5 架構，並依照財報特性制定「情緒標註準則 」。如圖 2 所示。

項目	說明
標註格式	• 每句最多標註 3 組（Aspect, Opinion, Category）• 每組需來自原句、語意明確
Aspect Term（主體）	• 必須為被 opinion 修飾的「簡潔名詞」• 必須取自原文，不得截斷或改寫• 無明確 opinion 修飾者不標註範例： ✔ 營運效率、資產品質；✗ 營運（過短）、營運效率穩健（含 opinion）
Opinion Term（意見）	• 必須為明確的情緒或變化動詞（如：提升、下降、改善）• 不得包含數字、比例、時間詞（如 5%、2 倍、132 億元）• 副詞不得併入（如：持續、大幅）範例： ✔ 成長；✗ 持續成長、大幅改善、5%
Category（類別標籤）	• 僅能使用預定義 E#A 清單(如 company#sales,business#general) • 不得自創標籤
其他建議	• 同句多個 opinion 時，只保留主意義者• 模糊詞「穩健、優化」僅在語意明確時使用• 必須能方便轉成 JSON / CSV 結構

圖 2 、情緒標註準則

標註採以下三元素：

1.Aspect Term：被評論的名詞，如「營收」、「毛利率」、「需求」。

2.Opinion Term：意見或描述動詞，如「成長」、「下降」、「改善」。

3.Category：固定分類標籤，如 company#sales、business#general 等。

範例：

句子：「營收持續成長，市場表現穩定。」

標註結果：

(營收, 成長, company#sales)

（3）標註規則摘要

以下為本研究制定的核心規則：

1.Aspect 必須是被 opinion 修飾、語意完整的名詞。

2.不可截用不完整詞，如「營運」若語意過泛則不標。

3.Opinion 限定為語意明確的動詞或形容詞。

4.不含副詞、時間字或數值，如「持續提升 5%」→ 標註為「提升」。

5.Category 必須從固定名單選取，不得自創。

同一句可標註多組，若多重情緒彼此衝突，保留主要意義者。

（4）人工審查

AI 擷取句子之後，研究成員逐筆審查三元素的組合是否正確，修正以下問題：

1.Aspect 與 Opinion 不對應

2.Category 使用錯誤

3.意見詞過度冗長

4.標註不一致

雖未進行正式一致性統計（如 Kappa 值），但所有資料均經過多輪人工比對，確保可用性。

3.4 資料增強與生成器設計

為改善資料量不足的問題，本研究自行開發「財報情緒資料生成器（Financial ABSA Generator v4）」，於 Google Colab 執行。

生成器流程如下：

1.讀取人工標註資料（三元素）。

2.建立 Aspect－Category－Opinion 對照表。

3.套用 15 組財報常見句型模板。

4.隨機填入時間詞、數值詞、Opinion 詞彙。

5.以規則式方法判定情緒極性（Positive / Negative / Neutral）。

6.輸出 JSONL 格式之合成資料。

模板範例：

「{T}{A}達 {N} 億元，{O}{P}%。」

「{T}{A}持續{O}，表現{O2}。」

「{T}{A}{O}，成本{O2}。」

其中 T 為時間詞、A 為 aspect、O 為意見詞、N 為數值。

最終生成器產出 超過 1,000 筆合成資料，並透過自動統計方式維持情緒分布之平衡。然而，由於模板化句型較固定，語意自然度有限，模型在加入生成資料後的提升並不明顯，相關原因在第四章詳述。

3.5 模型評估與誤差分析

本節說明本研究用來評估模型效能的方法與指標，實際數值結果與詳細誤差分析則於第四章中呈現。

本研究主要從兩個層面評估模型表現：（1）元素層級的預測能力；（2）完整四元組的預測能力。元素層級包含 Aspect、Category、Opinion 與 Sentiment 四種標籤；四元組則要求模型同時正確預測（Aspect, Category, Opinion, Sentiment）才視為完全正確。

在評估指標方面，本研究採用分類任務中常見之三項指標：Precision（精確率）、Recall（召回率） 與 F1-score。其定義如下：

1.Precision：在所有被模型判定為正類的預測中，有多少比例是正確的。

2.Recall：在所有真正的正類樣本中，有多少比例被模型成功找出。

3.F1-score：Precision 與 Recall 的調和平均，用來做為整體效能的綜合指標。

公式如下：

$$\text{Precision} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

其中，TP（True Positive）為真正類數量、FP（False Positive）為假正類數量、FN（False Negative）為假負類數量。

在評估程序上，本研究將標註完成之資料切分為訓練集與測試集，模型僅在訓練集上進行參數學習，並於測試集上進行預測，再根據上述指標計算各層級的表現情形。

此外，本研究亦針對模型預測錯誤之案例進行人工檢視，整理常見的錯誤情形與可能原因，作為誤差分析與後續改進的依據，相關分析將於第四章詳細說明。

Chapter 4 研究分析結果

4.1 資料概況與描述性統計

本研究所使用的資料來源包含人工標註之財報語料，以及後續由生成器產生的合成資料。為了協助模型理解財報文本中的語意特徵，本節首先整理資料量、情緒分布與標註內容之概況。

（一）人工標註資料量與內容比例

本研究最初從多家上市公司財報中擷取句子，經過 AI 初步擷取搭配人工審查後，共整理出 300 多句可用財報語料。

在這些句子中，共標註出 約 316 組有效三元素 (Aspect, Opinion, Category)。

下圖為人工標註資料的基本統計：

項目	數量
清理後可用句子	約 300 句
有效三元素標註資料	316 筆
平均每句標註組數	約 1 組
常見 Aspect	營收、成本、需求、毛利率
常見 Opinion	成長、下降、改善、疲弱
常見 Category	company#sales、company#profit、business#general

圖 3、人工標註資料統計

人工標註資料為模型訓練提供最核心的語意資訊，但因數量有限，無法支撐深度模型完整學習，因此需輔以資料增強技術。

（二）合成資料（生成器 v4）統計

為補足人工資料量不足，本研究開發「Financial ABSA Generator v4」生成合成文本。

該生成器採用模板式語句結構、隨機插入數值與時間詞，並搭配規則式情緒判斷函數自動產生情緒標籤。最終共產生 2,000 筆合成資料，並確保正向、負向、中立情緒的分布相對平衡，如圖 4 所示。

項目	數量
合成資料總筆數	2,000 筆
模板數量	15 種財報句型
情緒標籤來源	規則式判定（Positive / Negative / Neutral）
語料組成	（input_text, target_text）以 JSONL 格式儲存

圖 4、生成文本統計

(三)整體資料量統計如圖 5 所示。

資料來源	數量
人工標註資料（A,O,C）	316
合成資料（v4）	2,000
總數量	2,316

圖 5、整體資料量統計

（四）三大情緒分布（合成資料）

本研究的生成器設計上以「情緒平衡」為目標，然而由於模板語句及規則式情緒判斷的限制，最終生成之資料未達到完全平均的分布。

實際統計 2,000 筆合成資料之情緒比例如圖 6 所示：

情緒極性	數量	比例
Positive	936	46.8%
Negative	515	25.8%
Neutral	549	27.4%
合計	2,000	100%

圖 6、情緒佔比

分析說明：

1.Positive 類別佔比最高，原因可能是模板中較多「成長 / 提升 / 改善」類詞彙。

2.Negative 與 Neutral 類較接近，但仍存在偏差。

3.這顯示 v4 版本的情緒平衡規則 不能完全達成平均分配，但仍具有基本的分布控制能力。

4.2 人工標註資料分析

本節針對本研究之人工標註資料（資料 444.jsonl）進行統計與分析。

人工資料共 316 行句子，其中成功標註出 114 組三元素（Aspect, Opinion, Category）。

雖然整體標註量不大，但內容涵蓋營收、獲利、產品、市場等財報文本常見主題，能提供模型初步學習財務語意結構的基礎。

（一）標註資料整體概況如下圖所示：

項目	數值
句子總數	316
成功標註的三元素數量	316 組
平均每句標註數	1 組
是否含情緒極性（Sentiment）	✖ 無（本研究後續以規則式補齊）

圖 7、人工標註資料基本統計

(二)Aspect（主題）分布分析如下圖所示。

排名	Aspect（主題）	次數
1	EPS	31
2	營收	25
3	毛利率	21
4	動能	21
5	股東權益	20
6	淨利	19
7	出貨量	19

排名	Aspect（主題）	次數
8	市佔率	19
9	需求	14
10	雲端業務	13

圖 8、Aspect 前十名（真實資料）

從圖 8 中可觀察到：

1.財務績效類（EPS、營收、毛利率、淨利）最為集中

→ 反映財報文本主要以獲利狀況為重點。

2.市場所需指標比例高（出貨量、市佔率、需求）

→ 展現企業在產品銷售與市場動能上的描述。

3.科技產業特有名詞也明顯出現（如雲端業務）

→ 顯示語料包含科技股常用敘事。

（三）Opinion（意見詞）分布分析

排名	Opinion（意見詞）	次數
1	年減	22
2	優於市場預期	21
3	季增	21
4	低於市場預期	20
5	成長	19
6	強勁	18

排名	Opinion（意見詞）	次數
7	增加	13
8	提升	8
9	擴大	5
10	增長	3

圖 9、Opinion 前十名（真實資料）

從圖 9 中可觀察到：

1.財報專用比較詞最常見：季增、年減

→ 為台灣上市公司季報、年報常用的變化詞彙。

2.「優於／低於市場預期」比例高

→ 顯示原始文本大量描述企業績效相對市場預期的差異，屬於財報中典型的管理層語氣。

3.正向詞（成長、增加、提升）與負向詞（年減、低於預期）分布均衡

→ 表示語料包含正負向情緒，有助於模型判斷情緒極性。

（四）Category（類別）分布分析

Category 代表語意所屬的實體與屬性（Entity#Attribute）。

統計結果如下：

Category	次數
company#general	90
company#profit	74

Category	次數
market#general	48
company#sales	28
product#amount	21
company#cost	20
market#profit	12
product#general	9
business#general	7
product#sale	3
company#amount	3
business#investment	3
product#sales	2
market#sales	2
business	1
product#profit	1
company#investment	1
NULL#general	1
NULL#generall	1
product#price	1
market#cost	1

Category	次數
market#amount	1
company#generall	1
businese#general	1

圖 10、Category 出現次數（前若干名）

從圖 10 中可觀察到：

1.公司層級敘述（company#general、company#profit）最為常見，代表財報文本多圍繞營運與獲利表現。

2.市場層級（market#general、market#profit）佔比也高，反映財報常描述大環境與需求變化。

3.產品相關類別(product#amount、product#general 等)雖有出現但比例較小，多用於技術或製造業公司。

4.少部分 Category 屬於人工標註錯誤，未來可進行清理以提升模型訓練品質。

（五）人工資料的整體語意特性

綜合上述統計，本研究人工標註資料呈現以下特點：

1.財務數字導向：

Aspect、Opinion 多與營收、獲利、費用、市場需求等財務指標有關。

2.意見詞呈現單一方向性

以「增加、提升、擴大」等上升趨勢為主，使資料偏向正向情緒。

3.Category 分布集中於 general 類別

business#general 與 market#general 占比高，類別粒度偏粗，影響模型辨識能力。

4.資料量有限

316 組三元素對深度模型而言相對不足，因此需依靠後續生成器資料補強。

人工標註資料共包含 316 組標註，主題集中於營運成果與市場表現。雖然資料量有限，但語意清晰，結構規律，可作為模型訓練的基礎資料；然而資料面向集中且類別粒度較粗，有可能導致 MT5 模型在四元組預測上能力受限。

4.3 生成資料分析與人工資料比較

為補足人工標註資料量不足，本研究使用自行開發之「Financial ABSA Generator v4」產生 2,000 筆合成語料。本節從生成資料的語意特性、情緒分布與模板結構三方面進行分析，並與人工標註資料做比較，以探討兩者差異對模型訓練之影響。

（一）生成資料概況，如圖 11 所示。

生成器 v4 採用三步驟流程產出合成語料：

1.從人工資料萃取 Aspect－Category－Opinion 對照表

2.套用 15 組財報句型模板

3.隨機插入時間詞（如「2024 年」）、數值（如「500 億元」）、意見詞

4.以規則式方法自動判定情緒極性（Positive / Negative / Neutral）

最終產生 2,000 筆合成資料，格式統一，語句完整。

項目	數值
合成語料數量	2,000 筆
使用模板數	15 種
情緒標註方	規則式判定

項目	數值
式	
資料格式	JSONL（input_text, target_text）

圖 11、生成資料概況

（二）生成資料的情緒分布

生成器會自動平衡情緒，避免模型偏向特定情緒類別。

情緒分布如圖 12 所示：

情緒極性	數量
Positive	936
Negative	515
Neutral	549
合計	2,000

圖 12、生成資料情緒分布（真實統計）

從圖 12 中可觀察到：

1.Positive（正向）佔比最高，接近所有資料的一半，反映財報語句常見「提升、增加、改善」等偏正向描述。

2.Negative（負向）與 Neutral（中立）分布較接近，但仍略低於正向語料。

3.整體分布並非三等分，顯示生成器的規則式平衡機制能控制比例，但無法達到絕對平均。

4.這樣的分布仍能避免模型完全偏向單一情緒類別，但未來若需提升模型訓練效

果，可再改善生成器平衡演算法。

（三）生成資料常見語意特徵

根據生成模板與統計，生成語料具有以下特色：

1.句型結構規律

例如典型模板：

「2024 年營收達 500 億元，表現持續成長。」

「本季成本下降 10%，營運動能回升。」

句子通常遵循「主體 + 趨勢描述 + 原因/補充」結構。

2.意見詞多為數值化語彙

如「增加」「下降」「成長」「回升」

→ 易於模型學習，但語意較單調。

3.數值與時間詞隨機生成

→提升句子自然度，但仍不足以涵蓋真正財報文本的多樣語氣。

(四)人工資料 vs 生成資料比較

如圖 13 所示：

面向	人工標註資料（316 句）	生成資料（2,000 筆）
標註元素	3 元素（A,O,C）	4 元素（A,O,C,S）
來源	真實財報句子	模板＋隨機生成
語意自然度	高（自然語氣、文法）	中（模板化明顯）
多樣性	中（取決於財報內容）	中低（受模板限制）
情緒來源	無（後續判定）	規則式自動產生

面向	人工標註資料（316 句）	生成資料（2,000 筆）
適合用處	教模型真實語意	增加訓練量、平衡情緒

圖 13、人工標註資料與生成資料之比較

從圖 13 中可觀察到：

1.人工資料＝語意正確 + 自然語氣 + 品質高

→ 但數量少，無法支撐深度模型訓練。

2.生成資料＝量大 + 均衡情緒 + 易學

→ 但模板化明顯，缺乏語意多樣性。

兩者互補，才能讓 MT5-small 進行初步訓練。

（五）資料差異對模型的影響

結合模型訓練與後續誤差分析，生成資料與人工資料差異帶來以下影響：

優點（有助於訓練）

1.大幅提升訓練資料量（316 → 2,316）。

2.情緒分布均衡，避免模型偏向正向。

3.提供模型多種句型片段、固定語法可快速學習。

缺點（造成後續模型表現偏弱）

1.模板化句型 → 模型容易「記句型」而不是學語意。

2.部分模板語義不自然 → 造成模型 drift。

3.Category 與 Opinion 來源受限 → 多樣性不足。

4.和真實財報的語氣差異大 → 影響泛化能力

此部分在 4.5 誤差分析會有更完整說明。

生成資料在本研究中扮演彌補人工資料不足的重要角色。雖然合成語料能有效擴大資料量並平衡情緒分布，但模板化特性也導致語意多樣性較低。人工資料

與生成資料的差異，對後續 MT5 模型的訓練效果造成明顯影響，為本研究模型表現偏向「高精確率、低召回率」的重要原因。

4.4 模型訓練結果

本節呈現 MT5-small 模型於測試資料上的預測表現。本研究分別評估五項任務：

1.Aspect 抽取

2. Category 預測

3. Opinion 判讀

4. Sentiment（情緒極性）

5. 嚴格四元組（A, C, O, S）預測

模型效能以 Precision、Recall 與 F1-score 進行衡量，結果如圖 14 所示。

評估項目	Precision	Recall	F1-score
嚴格四元組 (A,C,O,S)	0.3462	0.0413	0.0738
Aspect	0.9615	0.1147	0.2049
Category	0.9615	0.1147	0.2049
Opinion	0.5769	0.0688	0.1230
Sentiment	0.3846	0.0459	0.0820

圖 14、模型在五項任務上的評估指標

(一).模型整體表現概述

整體而言，模型呈現 「Precision 高、Recall 低」 的典型小樣本現象：

在 Aspect 與 Category 任務上，Precision 皆達 0.96，顯示模型能「準確」預測部分元素。

然而，Recall 僅約 0.11，表示模型能抓到的真正標註很少，導致最終 F1-score 落在 0.20 左右。

最困難的任務是 嚴格四元組 (A, C, O, S)，因需同時四項都正確，因此 F1 只有 0.0738。

此結果反映模型在語意擷取上的能力有限，尤其在句子含有隱含語意、長句或財務語境較複雜時，模型易出現遺漏。

(二).各任務表現重點

1. Aspect / Category 的高 Precision 現象

(1)因為這兩類標籤集中於固定字詞（如：營收、需求、成本等），模型較容易在文本中辨識。

(2)但遇到較抽象的詞（如「營運效率」、「市場動能」）時，召回率仍偏低。

2. Opinion 辨識能力弱

(1)財報語句中的情緒動詞多樣（成長、下滑、改善、持穩等），模型未形成足夠語意連結。

(2)模型容易標錯 Opinion 或整段忽略。

3. Sentiment（極性）F1-score 偏低

(1)財報文風相對委婉、間接（如「持穩」、「微幅下滑」），情緒不如一般評論明顯。

(2)情緒極性與數值、上下文強相關，模型在小樣本情況下難形成穩定判斷。

4. 嚴格四元組難度最高

(1)因為四項只要錯一個就算錯。

(2)模型最常出錯的組合是：

- Aspect 抽對但 Opinion 抽錯
- Category 與 Aspect 不匹配
- Sentiment 極性判斷錯誤

本研究使用 MT5-small 進行 ABSA 四元素預測，結果顯示模型在固定名詞（Aspect、Category）表現良好，但在 Opinion 與 Sentiment 任務中仍具明顯挑戰，整體 Recall 偏低，導致四元組任務的整體 F1-score 落在 0.07～0.20。模型訓練結果反映資料量不足、語意複雜度高、生成資料模板化等限制，相關原因將於下一節誤差分析中詳細說明。

4.5 誤差分析

為了理解模型在四元素預測中表現不佳的原因，本研究進行逐句比對，分析預測錯誤的類型與可能成因。整體而言，誤差主要來自三大方向：（1）資料問題；（2）情緒語意問題；（3）模型能力限制。以下分項說明。

1.資料層面的誤差來源

(一) 標註資料量不足

人工標註資料僅 316 句，遠低於一般 ABSA 任務常見的千筆以上規模，使模型無法學到足夠的語意變化與語境模式。因此呈現出：

- Precision 高（模型能讀懂的都答對）
- Recall 低（模型能讀懂的太少）

這也是四元組 F1-score 僅約 0.0738 的主因。

2. 合成資料語法過於固定（Template Overfitting）

生成器雖提供約 2000 筆資料，但句型固定，例如：

(1)「{T}{A}達 {N} 億元，{O}{P}%。」

(2){T}{A}持續{O}，表現{O2}。」

模型很容易「記住固定句型」，而非真正學到語意。

結果導致：

- 模型在合成資料表現良好
- 但遇到真實財報語句（句型更自然、語氣更委婉）→ 直接失準
- 表現呈現 Data Drift（語意偏移）

3. 標註一致性挑戰

雖然三元素標註有制定規則，但實際觀察仍有幾項困難：

- 某些 Aspect 過於抽象（如「營運」、「市場表現」）
- Opinion 是否可切割存在模糊（如「持續成長」）
- 同一句可能同時有多個 Category，但最終只標一個

這使模型難以學習穩定的對應關係。

（二）語意層面的誤差來源

1. 財報語氣委婉，情緒不明顯

財報文本不像生活語言有明確的情緒詞，其語意常呈現：

- 模糊詞：穩健、持穩、略低、微增
- 間接語氣：營運環境具挑戰、成本結構調整中
- 反向語意：成本下降是好事、營收下降是壞事

模型很難僅依表面字詞判定情緒。

2. 隱含情緒（Implicit Sentiment）難以處理

例如：

「公司持續強化供應鏈管理。」

人類可理解為「正向改善」，但模型沒有上下文，會判定為 Neutral 或未知。

3. 多重語意混合句較難解析

財報常出現一個句子多個面向：

「營收持續成長，但成本受原物料價格上漲而增加。」

模型常出現：

- 抽到其中一個 Aspect
- Opinion 抽錯（如將「上漲」標成「成長」）
- Category 錯配

（三）模型層面的誤差來源

1. MT5-small 在抽取式任務本身較弱

MT5 是「文字到文字」模型，對生成任務效果較好，但對如下任務較弱：

- span 抽取（抓出句子裡的詞）
- 多元素結構預測（A, C, O, S 一起輸出）

因此容易出現「少預測」「不完整預測」等問題。

2. 四元組格式難度高

必須 四項都對才能算對，因此只要有一個錯：

- Opinion 字串不一致
- Category 用錯
- Sentiment 判錯

就會被視為錯誤，造成 F1-score 大幅下降。

（四）誤差綜合討論

綜合以上，本研究的誤差來源可歸納為三點：

- 資料規模不足 → 造成 Recall 極低
- 生成資料模板化 → 模型發生語意偏移（Data Drift）
- 財報語意複雜 → 隱含情緒、本質委婉、句型長，模型難解析

因此，即使模型對於固定用詞（如「營收」「成長」）能夠高精確地辨識，

但面對較自然的財報語句仍存在明顯困難。後續若能提升資料多樣性、改善生成器句型、加入語意改寫資料或使用更大型模型，相信有助於提升模型四元素辨識能力。

4.6 本章小結

本章針對人工標註資料、合成資料與模型預測結果進行分析，說明本研究在財報情緒四元素辨識上的整體表現。從資料分布可看出，人工標註資料的規模有限，情緒詞彙集中度高，對模型的學習形成明顯挑戰。模型訓練結果呈現「高精確率、低召回率」的特性，顯示模型能正確辨識部分明確語意，但對完整四元素結構的擷取能力仍不足。進一步的誤差分析指出，資料量不足、合成資料句型過於模板化，以及財報語氣的隱含性皆為主要影響因素。本章分析結果為下一章的結論與改進建議奠定基礎。

Chapter 5 結論

5.1 研究結論

本研究以自然語言處理（NLP）與層面式情感分析（ABSA）為核心方法，針對台灣上市公司之財報文本進行情緒結構化標註與模型建置。整體而言，本研究完整走過資料蒐集、資料前處理、人工標註、模型訓練、資料增強與模型評估等流程，並成功建立一套可用於財報語句的情緒分析雛型系統。研究的主要結論如下：

（1）財報文本確實包含可挖掘之情緒資訊。

透過人工標註可以觀察到，企業在報告書中常出現與營運成果、成本結構、需求變化等相關的語意訊號。這些語意內容具有一定情緒傾向，顯示財報文字除了傳達資訊之外，也反映管理階層對經營狀況的態度。

（2）人工標註結構有助於後續模型的訓練。

本研究以（Aspect, Opinion, Category）的三元素方式標註資料，再搭配後續情緒極性的推斷與模型預測，證明細粒度的 ABSA 架構能讓模型學習到企業文本中的語義關係。然而，由於財報語句專業性高，標註仍需人工審查，顯示 AI 與人工合作是必要流程。

（3）生成器可支援資料擴增，但效果受限於模板化特性。

生成器成功產生 1,000 句以上的合成語料，並維持情緒分布平衡，證實規則式資料增強在財報領域具有可行性。然而，因模板化句型較固定，語意自然度有限，因此加入合成資料後對模型的提升不如預期。

（4）MT5-small 模型可辨識部分財報要素，但整體 F1-score 偏低。

模型在 Precision 表現尚可，但在 Recall 方面明顯不足，尤其在四元組完整預測任務上，F1-score 偏低，顯示模型在小樣本與語意複雜的情境中仍有相當大的改善空間。

（5）完整研究流程具可行性，並可擴展至不同產業文本。

雖然模型表現有限，但研究證實本研究建立的流程（標註 → 訓練 → 預測

→ 評估）可正常運作，且財報文本的語意結構具一定共通性，因此未來具有擴展至製造業、金融業甚至 ESG 報告之潛力。

5.2 研究限制

本研究雖成功建立財報情緒分析的基本流程，但在資料、方法與模型面仍存在若干限制，需要於未來研究中進一步改善。以下整理本研究的主要限制：

（1）資料量有限，無法充分支援深度學習模型。

本研究的人工標註資料僅約 316 筆，而 ABSA 層面的任務屬於細粒度抽取式問題，需要大量多樣化語句才能讓模型有效學習。資料量不足也導致模型的召回率偏低，影響整體 F1-score。

（2）人工標註未包含情緒極性，需仰賴後續規則式推斷。

由於情緒極性（Sentiment）較難統一判斷，本研究未在人工標註階段直接標示情緒，而是依靠生成器與模型自行判定。此作法容易使模型受到錯誤情緒標籤影響，降低訓練品質。

（3）資料增強方法偏向模板化，語意多樣性不足。

雖然生成器成功產生大量合成資料，但模板句型重複度高，語意自然度有限，使模型容易學到模板特徵，而非真正的語意關係，此為模型「學壞」(data drift) 的主要原因之一。

（4）模型規模與訓練資源有限。

本研究採用 MT5-small 進行訓練，其參數量較小、語意理解能力有限，再加上訓練 Epoch 數量、硬體資源皆受限，導致模型難以完整學習財報文本中的專有語意。

（5）標註一致性未進行正式量化。

本研究雖經過多輪人工審查，但並未採用 Cohen’s Kappa 等正式一致性指標，因此無法客觀衡量標註品質，可能對模型訓練產生影響。

（6）財報文本語氣偏制式，情緒表現較不明顯。

與社群評論或新聞不同，財報語句通常較為保守、正式，使情緒多表現在「措

辭細微差異」而非「明確情緒字眼」，因此抽取式情緒分析在財報領域本身就具有挑戰性。

5.3 實務建議

本研究主要從財報文本中萃取情緒資訊，雖然模型效能仍有限，但仍能提出以下實務層面的應用建議，供企業與投資決策者參考。

（1）企業應善用財報語氣作為對外溝通策略的一部分。

財報中的語氣與用詞不僅反映經營者的態度，也會被外部分析師或投資人解讀成市場信心指標。企業可定期檢視自身語句使用是否過度保守或不一致，以提升對外溝通的透明度與一致性。

（2）投資決策可將「財報情緒」納入輔助分析工具。

雖然模型精準度仍有待提升，但財報文字中的情緒變化確實能反映管理階層的預期，例如市場需求、營運風險等方向。投資人可將情緒指標視為輔助訊號，搭配基本面資料進行綜合評估。

（3）企業資訊部門可逐步導入 NLP 工具，提升報告分析效率。

大量財報、法說會內容若完全依靠人力閱讀，耗時且不易維持一致性。導入 NLP 模型可協助初步分類、摘要與情緒分析，減少人工負擔，並提升財務分析流程的標準化程度。

（4）研究機構與企業可考慮建立專屬的財務詞庫或標註集。

目前大多 NLP 模型主要基於通用語料訓練，較難理解財務專用語意。若企業能建立自家語料庫（如多年度財報、法人報告、產業分析）並持續標註，可顯著提升模型的可用性。

（5）審慎使用 AI 生成資料，避免語料偏誤。

本研究證實若生成資料模板化過高，容易造成模型誤學。企業若想用 AI 進行語料擴增，應搭配人工審查、語意改寫、多樣化模板，才能提升資料品質而非造成偏差。

5.4 未來研究建議

根據本研究的實作過程與模型結果，未來研究可從以下方向持續改進，以提升財報情緒分析的準確度與應用價值：

（1）擴大人工標註語料規模，提升語意多樣性。

本研究僅標註約 316 筆財報句子，資料量偏小，且多集中於特定公司或章節。若能蒐集更多年度、更多產業、更多公司之財報文本，並建立更大規模的人工標註語料庫，將能有效提升模型的泛化能力。

（2）在標註階段加入情緒極性，提高四元組品質。

目前的人工資料僅包含三元素（Aspect、Opinion、Category），情緒極性由模型或規則式方法補齊，容易造成誤差。未來可於人工標註階段直接加入 Positive／Negative／Neutral，以提升訓練資料的完整性與一致性。

（3）改善資料增強策略，引入語意改寫模型（Semantic Augmentation）。

本研究的生成器雖能快速擴充語料，但模板化程度高。未來可透過 T5、Llama 3、ChatGPT 等生成式模型進行語意改寫、句型轉換與語氣替換，讓語料更接近真實財報語句，以避免模型「學壞」。

（4）導入更大型或財務專屬的語言模型進行訓練。

MT5-small 參數量較低，無法完全捕捉財報語意。未來可考慮使用：

- MT5-base / MT5-large
- BloombergGPT、FinGPT 等財務領域 LLM
- Llama 3 或專用微調模型

以提高語意理解能力與情緒預測正確率。

（5）引入句法結構（Syntax）或上下文資訊（Context），提升抽取式能力。

財報語句多為複句或隱含情緒的敘述，僅依靠單句預測容易誤判。未來可嘗試：

- 句法樹（Dependency Parsing）
- 段落級上下文
- 文件級情緒流變分析（Sentiment Flow）

以更貼近財報中「漸進式情緒」的特性。

（6）建立結合財務指標的多模態模型。

財報情緒與實際財務表現（營收、毛利率、EPS）具有一定連動性。未來研究可將財務數據與文本情緒整合，建立更具預測力的「情緒 × 財務」混合模型，用於分析企業營運風險或市場反應。

Chapter 6 參考資料

一、學術論文（Journal Articles）

Chen, Y., Lin, C., & Hsu, P. (2022).

Sentiment analysis of Chinese financial reports using deep learning models.

International Journal of Financial Studies, 10(3), 45–58.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019).

BERT: Pre-training of deep bidirectional transformers for language understanding.

Proceedings of NAACL-HLT 2019, 4171–4186.

Engelberg, J. (2008).

Costly information processing: Evidence from earnings announcements.

Journal of Accounting Research, 46(4), 911–940.

Li, F. (2010).

The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach.

Journal of Accounting Research, 48(5), 1049–1102.

Loughran, T., & McDonald, B. (2011).

When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks.

The Journal of Finance, 66(1), 35–65.

Pontiki, M., Galanis, D., Papageorgiou, H., et al. (2016).

SemEval-2016 Task 5: Aspect-based sentiment analysis.

Proceedings of SemEval-2016, 19–30.

Raffel, C., Shazeer, N., Roberts, A., et al. (2020).

Exploring the limits of transfer learning with a unified text-to-text transformer.

Journal of Machine Learning Research, 21(140), 1–67.

二、技術文件（Technical Documentation）

Google Research. (2023).

mT5: Multilingual text-to-text transfer transformer.

Retrieved from https://github.com/google-research/multilingual-t5

Hugging Face. (2023).

Transformers: State-of-the-art machine learning for NLP.

Retrieved from https://huggingface.co/transformers/

Scikit-learn developers. (2023).

Scikit-learn: Machine learning in Python.

Retrieved from https://scikit-learn.org/

三、資料來源（Data Sources）

公開資訊觀測站（MOPS）. (2020–2024).

上市公司年報與股東會年報。財政部證券暨期貨局。

Retrieved from https://mops.twse.com.tw/

附錄 A.專題工作內容

1111614 邱奕銓

主要負責專題技術面的前期建置與資料處理流程，包括：

1.自 MOPS 蒐集多家上市公司財報內容，撰寫初版擷取規則。

2.進行資料前處理流程：句子切分、去除表格符號、排除無效句。

3.建立 ABSA 三元素標註規則（Aspect / Opinion / Category）。

4.協助人工標註時的規則檢查、標註品質修正。

5.負責模型訓練流程：資料切分、訓練參數設定、模型測試。

6.整理模型指標、記錄模型版本與誤判樣本。

1111620 陳彥瑋

負責資料整合與文字產出相關的任務，包括：

1.統整所有處理後的資料、標註結果、模型輸出，整理成最終資料集。

2.修改資料增強生成器的模板句型，使生成資料更接近財報語氣。

3.協助調整生成器的欄位格式、參數設計與輸出 JSONL 格式。

4.整理統計圖表，如類別統計、意見詞分佈、Aspect 分佈等。

5.編排與撰寫期末報告內容，包含方法章節、分析章節與結論撰寫。

1111648 陳造翊

負責資料整理、格式統一與流程管理，包含：

1.將前處理後的句子統一格式化（欄位清整、字串處理、避免重複）。

2.與標註者協作，將標註內容統整成一致的 JSON/CSV 結構。

3.整合人工標註資料與生成資料，建立模型可直接讀取的檔案。

4.檢查並整合人工資料 + 合成資料，建立最終訓練集。

5.協助管理專題檔案版本、資料夾結構與訓練紀錄文件化。

附錄 B .專題心得與建議

1111614 邱奕銓

在本次專題中，我主要負責資料蒐集、前處理與標註規則制定。這段經驗讓我深刻體會到資料前處理的重要性。原本以為資料蒐集只是簡單下載財報，但實際做才發現財報格式複雜、句子段落不一致，以及大量無效文字需要手動過濾。這些看似基礎的工作卻會直接影響後續模型的品質，使我對「資料品質」有更深的認識。

在建立 ABSA 標註規則的過程中，我也學到專題不是技術越複雜越好，而是越清楚的規則，越能提升標註一致性，讓模型更容易學到語意。尤其財報語氣含蓄、句型多變，標註常需要反覆確認，這也訓練我更細心地思考語意邏輯。

雖然模型結果不如預期，但我認為這是一次很有價值的經驗。未來若要提升準確率，我建議可以擴大資料量、引入更大型模型，或增加不同產業的語料。整體而言，這次專題讓我學到 NLP 實務流程與資料工程的重要性，是一次很有收穫的專題合作。

1111620 陳彥瑋

我在專題中負責資料整合、報告撰寫與資料增強生成器的調整。因為需要整理所有前面階段的成果，我必須完整理解資料蒐集、標註、生成器與模型訓練等流程，這讓我逐漸掌握 NLP 專題的整體架構。把不同階段的內容整合成一份有邏輯的報告，也讓我學到如何把技術內容轉換成一般讀者能理解的文字。

在資料增強的部分，我協助修改生成器模板，嘗試讓語句更自然、分布更合理。這段過程對我來說是新的挑戰，也讓我更清楚看到模板化資料帶來的限制。雖然生成器能快速提升訓練集規模，但若語句不夠多樣化，模型就容易過度記住固定句型。這讓我理解到：好的資料比多的資料更重要。

整體來說，這次專題讓我在資料整理、技術理解與撰寫能力上有明顯成長。未來如果有人做同類型研究，我會建議從一開始就統一資料格式並保留清楚的記錄，這會讓後續整合輕鬆很多。

1111648 陳造翊

在專題中，我負責資料格式整理、標註後的結構統一與模型資料集整併。雖然這些內容看起來不像模型訓練那麼技術性，但實際執行後才發現，資料格式的統整對整個 NLP 流程來說非常重要。只要欄位格式不一致或標籤寫法不同，就會造成模型無法讀取或訓練失敗，讓我更理解資料工程的重要性。

在整理人工標註資料與生成資料的過程中，我也更清楚看到 ABSA 的複雜性。許多句子可能包含多重面向，標註需要非常精準，統整時也要注意不要破壞語意。這些細節讓我學到更有系統化處理資料的方式，也提升我在文件管理與流程紀錄上的能力。

透過這次專題，我更能理解 NLP 專題不是只有模型才重要，資料管理、格式一致性和檔案流程同樣是不可或缺的基礎。未來如果有學弟妹做類似專題，我會建議從一開始就建立清楚的資料規範，能省下非常多時間。