



BERT 模型微調於行業統計分類之應用

隨 AI 迅速發展，政府統計作業面臨極大的挑戰與轉型契機，以企業面統計調查觀察經濟活動的行業分類為例，長期仰賴人工判讀，相當耗時費力，亟待善用 AI 工具，提升行業判定效率。本文介紹應用自然語言模型 BERT 所擁有雙向語意理解之優勢，協助進行行業判讀，提升統計分類作業之效率與品質。

關偉倫、趙明光（行政院主計總處綜合統計處專員、研究員）

壹、前言

我國行業統計分類係參照聯合國訂定之「國際標準行業分類」（International Standard Industrial Classification of All Economic Activities, ISIC），將各種經濟活動做有系統之編碼分類，現行分類層級共分大（第 11 版有 19 類）、中（88 類）、小（249 類）、細（522 類）4 個層別，主要目的係供統計分類參考使用，並利於國際間統

計資料之比較。

大部分統計調查中（例如工業及服務業普查、服務業營運及投資概況調查等），經濟活動問項為請受訪單位以自然語言方式填復，且其相關編碼分類須仰賴人工檢視判定。以 110 年工業及服務業普查為例，訪查家數約 147 萬家，共計約 149 萬筆資料須進行行業分類編碼（包含主、次要經濟活動），此項業務須投入大量人力以及時間，不僅如此，編碼

易因個人經驗或訓練時間不足（例如新進人員）等因素致品質不穩定，增加後續檢核校正時間。

近年來，受惠於運算能力與訓練資料量大幅提升，人工智慧（Artificial Intelligence, AI）發展迅速，在深度學習領域，特別是在語意辨識、圖像識別與自然語言生成方面表現優異，其中，自然語言模型在語意理解及文本分類方面已有突破性的發展，本文將介紹以自

然語言模型 BERT (Bidirectional Encoder Representations from Transformers) 為基礎訓練行業判讀模型，藉由解析統計調查之經濟活動描述，進行行業分類編碼自動化，以提升分類效率及品質，並做為綜合統計處（以下簡稱本處）115 年實施第 12 版行業統計分類邁向智能服務的前導性應用。

貳、自然語言模型 BERT 介紹

一、自然語言模型於政府統計之應用

自然語言處理 (Natural Language Processing, NLP) 技術係指透過大量文本訓練，能夠理解、生成並推理自然語言，代表模型包括 ChatGPT、Gemini、LLama 等，由於 NLP 技術的突破性發展，相關模型廣泛應用於文本分類、語意分析等任務。在行業分類領域，NLP 模型可有效處理由企業經濟活動描述，自動化進行行業分類，目前世界各國已逐步導入，以提升統

計作業效率。其中 BERT 因支援多國語言、可落地部署避免敏感資料洩露，且可透過微調快速建立模型及訓練成本低等特點，逐漸為一些國家統計局導入應用。

例如，新加坡統計局近年開發「智能分類引擎」(DOS Intelligent Classification Engine, DICE)，以 BERT 模型為基礎，自動將企業經濟活動描述對應至「新加坡標準行業分類 (SSIC)」代碼，準確率

高達 86%，大幅減少人工作業負擔。土耳其及挪威統計局則運用 BERT 模型，將商品描述依「個人消費用途別分類 (COICOP)」代碼分類作業自動化，土耳其另開發商品分類查詢介面，使用者只須輸入商品描述，系統即自動顯示分類。這些案例顯示自然語言模型已逐步成為協助政府統計之重要工具，本處也以 BERT 建構自動預測 4 碼行業分類的查詢系統 (表 1)。

表 1 國際與我國自然語言模型分類應用案例

	土耳其	新加坡	挪威	行政院主計總處
文獻年份	2022 年	2024 年	2025 年	2025 年
主要目的	模型依商品描述自動預測所屬個人消費用途別分類 (COICOP) 代碼	模型依經濟活動描述自動預測所屬新加坡標準行業分類 (SSIC) 代碼 (5 碼)	模型依家庭預算調查 (HBS) 之商品描述自動預測所屬個人消費用途別分類 (COICOP) 代碼	模型依經濟活動描述自動預測行業統計分類細類代碼 (4 碼)
語言模型	BERT	BERT	BERT	BERT
類別數量	420 個	1,023 個	300 個	522 個
準確度	-	86%	81%	90%
實際應用	開發商品分類查詢介面	開發智能分類引擎 (DICE) 查詢介面，並提供 API	-	開發行業智能查詢系統

資料來源：作者自行整理。

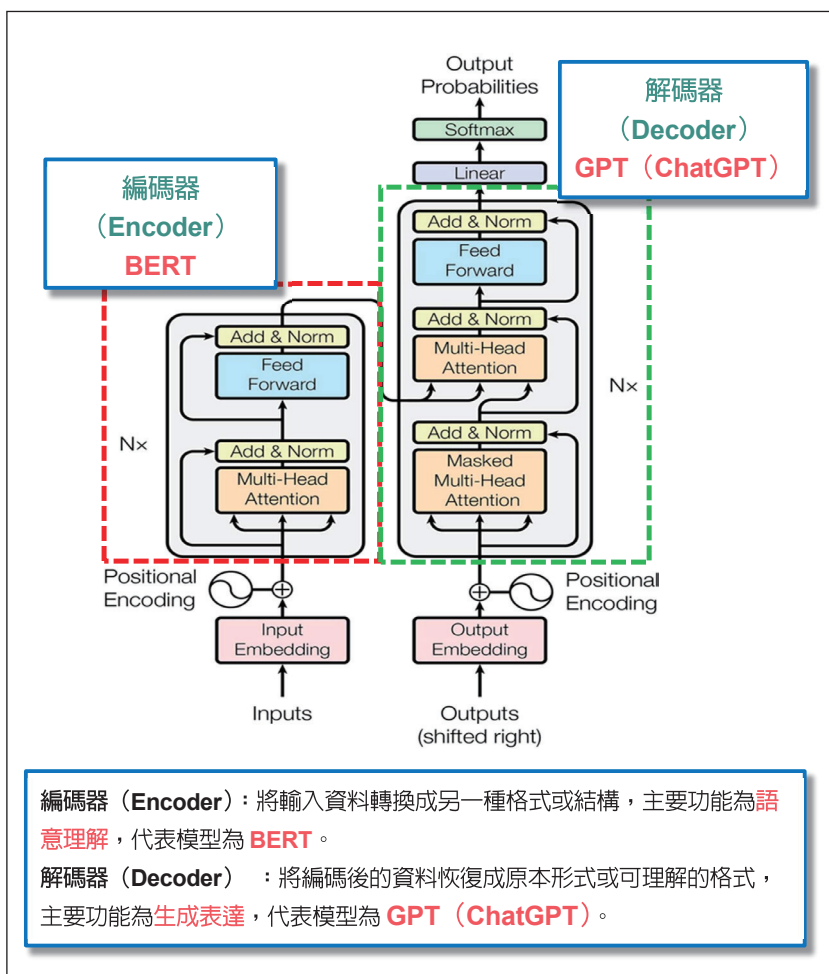
論述 》 統計 · 調查

表 2 自然語言模型 BERT 及 GPT 比較

	BERT	GPT (ChatGPT)
架構	雙向 Transformer 編碼器	單向 Transformer 解碼器
優勢	語意理解及分類	生成自然語言
模型性質	開源模型 (免費)	商用模型 (須付費)
主要應用	實體辨識命名、文本分類	對話生成、寫作輔助、程式碼生成

資料來源：作者自行整理。

圖 1 Transformer 架構



資料來源：Vaswani et al. (2017), "Attention Is All You Need".

二、BERT 模型－語意理解及文本分類上表現優異

上述案例中提到之 BERT 模型，是由 Google 團隊於 2018 年提出的自然語言模型，開創了 NLP 領域之新典範 (表 2)，其核心創新係在 Transformer 的編碼器 (Encoder) 架構下 (圖 1)，採用雙向閱讀理解模式，可同時從上下文中學習詞彙的語意表示，突破了傳統語言模型僅能單向 (由左至右或由右至左) 閱讀的限制，相較於傳統的長短期記憶網路 (LSTM)，BERT 模型因其雙向語意理解能力、預訓練及微調機制，在文本分類任務中表現優異。

BERT 模型訓練機制主要分為兩個階段，分別是「預訓練 (Pre-training)」與「微調 (Fine-tuning)」，以本文採用之 BERT-BASE-Chinese 模型為例，其在大量中文語料 (如中文維基百科) 上，分別透過「遮蔽語言模型」 (Masked Language Modeling, MLM) 及「預測下一句」 (Next Sentence Prediction,

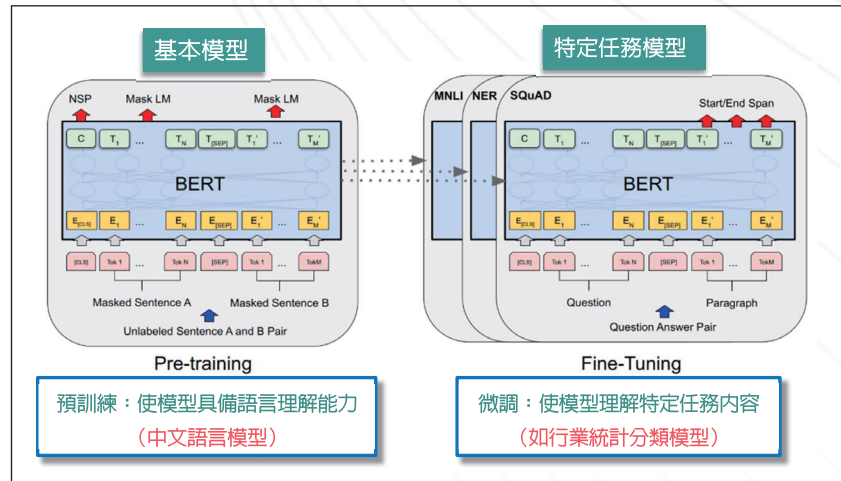
NSP) 進行「預訓練」，以學習語言結構。MLM 是隱藏句子中的某些詞，讓模型根據上下文來預測這些被遮蔽的詞，NSP 是判斷兩句話是否為連續語句，以學習句子之間的語意關聯。

由於 BERT-BASE-Chinese 模型已完成預訓練，具備中文語意理解能力，本處可用其接續進行「微調」，應用於各種特定任務，如中文文本分類，只須在特定任務上採用成對標註資料，即可快速微調模型參數，達到優異表現，適合應用於高準確度要求的行業分類任務 (圖 2)。

參、資料清洗及模型微調訓練

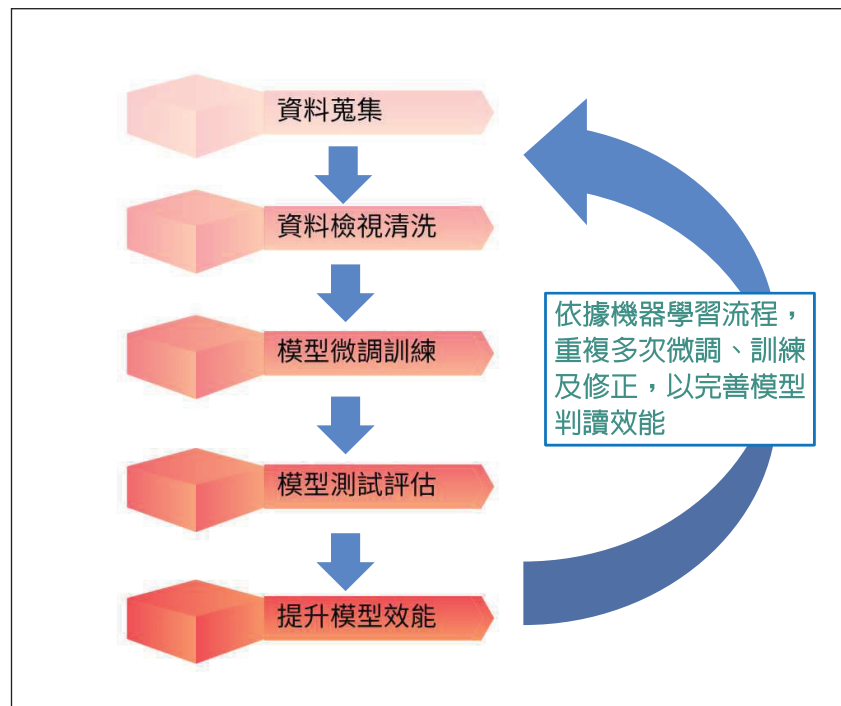
訓練 BERT 模型進行行業分類任務時，須經過嚴謹的訓練流程，為提升模型效能與分類準確度，其微調過程須重複循環進行，並根據每次測試結果持續改進，直至完善狀態，訓練過程包括「資料蒐集」、「資料檢視清洗」、「模型微調訓練」、「模型測試評估」及「提升模型效能」(圖 3)。

圖 2 BERT 模型「預訓練」與「微調」機制



資料來源：Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding".

圖 3 提升 BERT 模型行業判讀效能流程



資料來源：作者自行繪製。

論述》統計 · 調查

一、資料蒐集清洗

首先須蒐集大量且具代表性之經濟活動描述資料作為模型訓練基礎，若訓練資料量過低，模型將無法廣泛捕捉各種經濟活動描述，本次係採用 110 年工業及服務業普查行業代碼及主、次要經濟活動描述共約 149 萬筆資料，去除代碼與描述重複資料後，以近 17 萬筆資料進行訓練，每筆資料皆包含經濟活動描述及人工判別之行業細類 4 碼（標註編碼）。

在採用標註資料進行訓練前，須先檢視資料並修正錯誤（資料清洗），以確保品質，相關問題包括文字亂碼、缺漏、標註編碼錯誤與各業別訓練資料量差異過大等。其中針對文字亂碼或缺漏問題，透過大型語言模型 ChatGPT 可有效辨識並修正異常字元、補齊缺漏字等，提升資料可讀性。

針對原始標註編碼可能存在錯誤的情況，則利用 BERT 模型對每一筆訓練資料行業描述進行預測，若預測編碼與標註編碼不一致就須進一步確

認，確認方式包括三種，分別是「相似度比較」、「ChatGPT 判讀」及「人工判讀」，因資料量龐大，加以人力有限，故先以相似度比較及 ChatGPT 判讀初步篩選出疑似標註編碼有誤資料，再由人工進行判讀及修正該編碼（圖 4）。

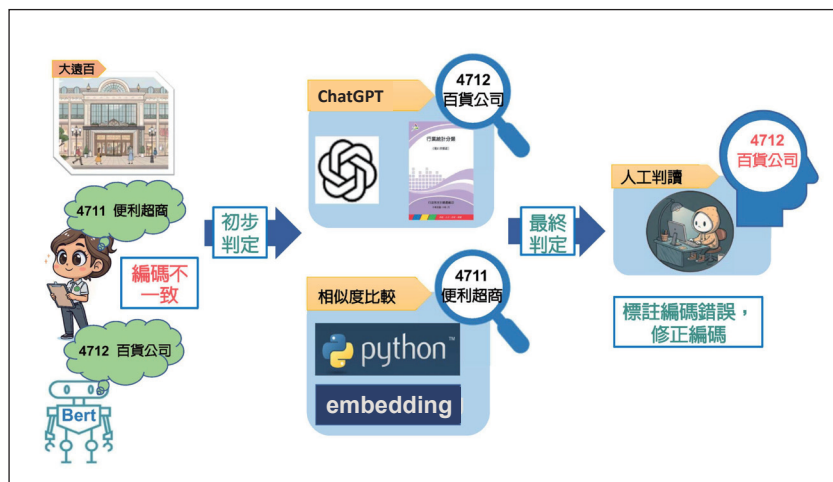
此外，有關各業別間訓練資料量差異較大或部份業別因非普查範圍致無訓練資料情形，舉例來說，餐飲業因經濟活動多元，可用於訓練之資料筆數相對較多，而部分產業別因經濟活動較為單純，訓練資料相對較少（例如石油及天然氣礦業），導致模型於判讀某

些業別時準確率偏低，我們亦運用 ChatGPT 依該業別行業資訊，生成對應經濟活動描述（如油氣井鑽探及擴展、石油與天然氣勘探及生產、油氣勘探與測試分析等），補足資料缺口，強化模型能力。

二、模型微調訓練

模型進行微調訓練前，先將資料集進行切分，其中 85% 資料集用於模型學習與參數調整之「訓練集」，其餘 15% 資料集為「測試集」，用以評估模型預測能力。此配置有助於模型訓練與效能驗證（下頁圖 5）。

圖 4 確認標註編碼正確性流程



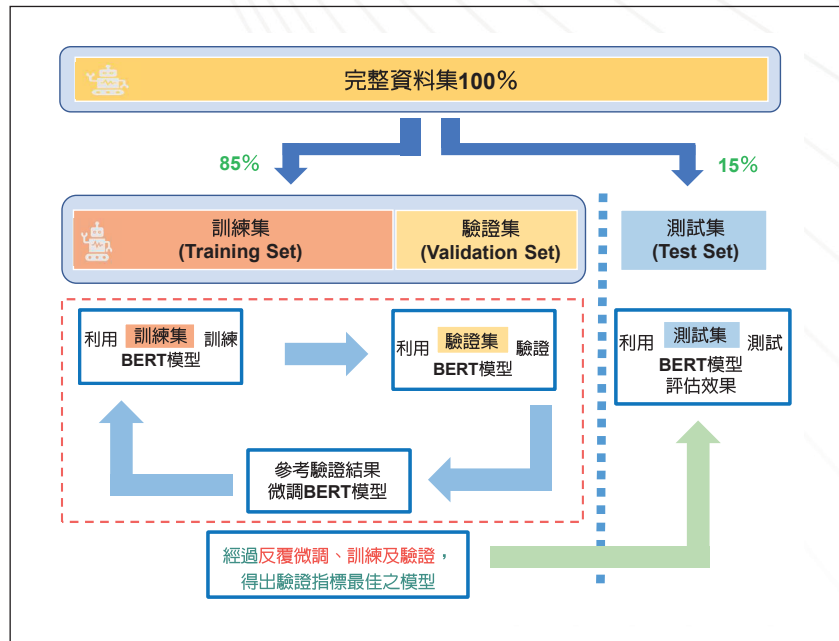
資料來源：作者自行繪製。

本研究採用 BERT-BASE-Chinese 模型，其具備良好的中文語意理解能力，在微調階段，模型透過最小化損失函數（如交叉熵損失）進行參數調整，使其能針對行業分類任務進行語意特徵學習。透過反覆訓練與驗證，模型逐步提升分類準確度，並可根據驗證結果調整超參數（如學習率、批次大小、訓練次數等），以達到最佳效能。

驗證評估模型常用指標包括準確度（Accuracy）、精確度（Precision）、召回率（Recall）與 F1 分數，透過這些指標可全面掌握模型在不同類別上的分類效能（圖 6）。

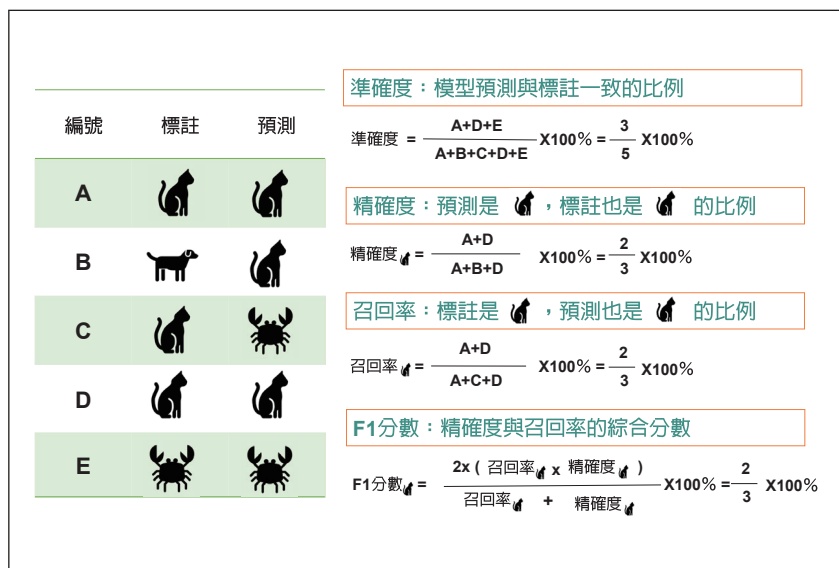
若模型表現未達預期則須進行修正，修正策略包括檢視資料標註錯誤並加以修正、強化資料（例如同義詞替換、大型語言模型擴充資料等），以及調整超參數，以提升模型學習與預測能力。這些步驟有助於建立一個高應用性之分類模型。

圖 5 資料集配置及微調過程示意圖



資料來源：作者自行繪製。

圖 6 準確度、精確度、召回率與 F1 分數



資料來源：作者自行繪製。

論述》統計 · 調查



肆、行業分類模型應用及效益

一、統計調查行業判讀自動化

在企業面統計調查方面，應用 BERT 模型可解析企業經濟活動內容，快速對應行業代碼。傳統以人工進行行業細類（4 碼）判讀，若平均每筆所需時間為 1 分鐘，以 110 年工業及服務業普查主、次要經濟活動資料約 149 萬筆為例，總花費時間達 24,833 小時，若改由 BERT 模型進行判讀，在硬體設備為 8 核心 CPU 及 32GB 記憶體環境下，可於 19 小時內完成，整體花費時間較人工編碼減少超過 99%，節省大量人力與時間成本。

此外，傳統人工編碼品質易受個人判斷能力等因素影響，導入 BERT 語言模型，透過雙向語意理解，顯著提升編碼一致性與準確性。以隨機抽取工業及服務業普查，包含生成約 4 萬筆測試資料為例，預測編碼準確率達 90% 以上。為擴大應用，本處規劃未來將提

供 API 功能，支援國內各企業面抽樣調查及普查，使其經濟活動描述可批次自動化判讀，提升作業效率與模型效用。

二、行業智能查詢系統

其次，為提升即時查詢便利性，行業智能查詢系統初步將建置於全國主計網，供主計人員於企業面調查（例如工業及服務業普查、受僱員工薪資調查、服務業營運及投資概況調查等）使用，只要有手機或電腦皆可即時查詢廠商經濟活

動所屬細類行業代碼，大幅減少行業判斷時間及提昇分類品質。使用者只要以自然語言形式（如「經營咖啡廳」、「提供雲端儲存服務」）輸入，系統即時回應建議可能的行業分類，展現 NLP 技術在公共服務介面上的創新應用（圖 7）。

伍、結論

在 NLP 領域中，BERT 模型自推出以來，憑藉其出色的語言理解能力及預訓練機制，成為眾多語言任務的基礎模

圖 7 行業智能查詢系統

行業智能查詢系統

請輸入該場所**主要**從事的經濟活動（如主要生產、製造的產品或提供的服務）：

搜尋輔導英文 送出 **A**

關於所詢「搜尋輔導英文」的行業細類，按行業統計分類第11次修正版（110年1月實施）原則，可能如下：

行業編號及名稱	行業定義	預測機率
8591 語言教育業	從事提供語言教育服務之行業；手語指導服務亦歸入本類。	57.00%
8595 課業、升學及就業補習教育業	從事提供國小、國中及高中課業、升學及就業考試（如公職考試、專業執照及證照考試）補習教育服務之行業。	36.69%
8891 兒童及少年之社會工作服務業	對兒童及少年從事居住照顧服務以外之社會工作服務之行業，如日間托兒服務、協助辦理收出養、協助社會資源轉介及諮詢等服務。	1.69%

B

行業統計分類係以「場所單位」（非以企業單位）實際從事的主要經濟活動分類，一個場所單位只能歸屬一個行業細類，若同時從事多種經濟活動，以附加價值最大的經濟活動判定。➡ 更多資訊請至 [行業統計分類專區查詢](#)。

本分類係供統計之用，各機關衍生應用於行政管理或管制等非統計用途（如勞動基準法適用類別），由權責機關依目的自行認定。

歡迎提供您使用本服務的滿意度或建議：

在此輸入您的意見 **C**

送出意見

A.輸入經濟活動：以自然語言方式輸入經濟活動，取得可能之行業代碼。
B.提供行業類別及其信心值：提供最可能之三種行業細類及其預測機率。
C.滿意度調查：使用者可填復滿意度及建議。

資料來源：作者自行繪製。

型，其主要優勢在於雙向編碼架構，能同時考慮上下文語意，顯著提升文本分類、問答系統等任務的準確性。此外，BERT 模型的微調機制，能針對特定任務進行訓練，快速適應不同應用場景，展現高度彈性與泛化能力。

本處基於 BERT 模型的優點，以 BERT 模型為基礎，微調訓練行業統計分類判讀模型，其中在資料清洗階段，亦藉助 ChatGPT 完善訓練資料；模型訓練結果透過「行業智能查詢系統」提供應用，做為 115 年實施第 12 版行業統計分類邁向智能服務的前導石，並期許本次經驗可外溢至更多統計應用層面，推動公共服務創新。

參考文獻

1. 黃河川 (2024)，以 LINE 建構 AI 行職業統計編碼機器人，主計月刊，819 期，98-101 頁。
2. 趙明光 (2024)，機器學習於政府統計之應用，主計月刊，818 期，72-77 頁。
3. Sanjit Kakarla, NameConrad Borchers, NameDanielle Thomas, NameShambhavi Bhushan, NameKenneth R. Koedinger (2025)，Comparing Few-Shot Prompting of GPT-4 LLMs with BERT Classifiers for Open-Response Assessment in Tutor Equity Training, Carnegie Mellon University. arXiv:2501.06658.
4. Boriska Toth, Statistics Norway (2025)。Use cases of generative AI at Statistics Norway ,and making the business case, Statistics Norway.
5. Malcolm Cai, Jun Wen Tay and Nicholas Chin (2024)。The Use of Artificial Intelligence and Machine Learning in DOS, Digital Services and Transformation Division, Department of Statistics, Singapore.
6. Abdulcebar On, Halil Zeybek, Ali Osman Bilgin, Kudret Elif Berkman, Mustafa Karamavus (2022)，Using Big Data Tools and Machine Learning Techniques to Assign Classification of Individual Consumption by Purpose (COICOP) Categories, Turkey Statistical Institute (TURKSTAT) 2022.01.24.
7. UNECE (2022)，Machine Learning for Official Statistics, March 2022.
8. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019)。BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)，4171-4186.
9. Liu, Y., Ott, M., Goyal, N., Du, J., Li, M., Ma, L., ... & Gao, J. (2019)。RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692.
10. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018)。Improving Language Understanding by Generative Pre-Training. OpenAI blog.
11. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017)。Attention Is All You Need. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, 6000-6010.
12. Kim, Y. (2014)。Convolutional Neural Networks for Sentence Classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)，1746-1751.
13. Sutskever, I., Vinyals, O., & Le, Q. V. (2014)。Sequence to Sequence Learning with Neural Networks. Advances in Neural Information Processing Systems, 27.
14. Joachims, T. (1998)。Text Categorization with Support Vector Machines: Learning with Many Relevant Features. Proceedings of ECML-98, 10th European Conference on Machine Learning, 137-142.
15. Hochreiter, S., & Schmidhuber, J. (1997)。Long short-term memory. Neural Computation, 9 (8)，1735-1780. ❖