

機器學習工程師職能基準

| | | | | |
|--------------------|-----|--|-------|------|
| 職能基準代碼 | | SMS2512-001v1 | | |
| 職能基準名稱 (擇一填寫) | 職類 | | | |
| | 職業 | 機器學習工程師 | | |
| 所屬 類別 | 職類別 | 科學、技術、工程、數學 / 數學及科學 | 職類別代碼 | SMS |
| | 職業別 | 資訊及通訊專業人員 / 軟體開發及程式設計師 | 職業別代碼 | 2512 |
| | 行業別 | 出版影音及資通訊業 | 行業別代碼 | J63 |
| 工作描述 | | 為協助企業判斷預測或制定決策，洞悉業務需求資料意義，運用大量計算資源處理特徵及標註，應用演算法建立並訓練模型，設定合適驗證指標以評估模型效能，並依據部署環境(軟硬體)完成整合測試。 | | |
| 基準級別 | | 5 | | |

| 主要職責 | 工作任務 | 工作產出 | 行為指標 | 職能 級別 | 職能內涵 (K=knowledge 知識) | 職能內涵 (S=skills 技能) |
|-------------------------|------------------------|---|--|----------|---|---|
| T1 釐清需求，獲取攸關資料並轉成機器學習所需 | T1.1 釐清客戶機器學習應用需求並擬定目標 | O1.1.1 機器學習應用需求釐清表 O1.1.2 機器學習應用業務目標 | P1.1.1 釐清客戶需求並聚焦機器學習之應用，擬定合理且明確業務目標。 | 5 | K01 機器學習演算法基礎知識(決策樹、KNN、SVM、線性回歸、K-means、PCA、Apriori、Q-Learning 等) K02 機器學習數學基礎(機率、線性代數、微積分) | S01 拆解問題能力 S02 業務溝通能力 |
| | T1.2 獲取業務攸關資料，並觀察及瞭解內容 | O1.2.1 資料範圍定義 O1.2.2 可視化資料變項分析表 | P1.2.1 明確且具體掌握業務需求，收集並了解攸關資料意義，善用探索式資料分析，透過視覺化統計工具洞悉資料關聯性(相關係數、分布狀況) | 4 | K03 資料庫基本原理 K04 資料庫架構 K05 資料處理與分析概論 K06 統計基本知識 | S03 探索式資料分析(EDA) S04 網路爬蟲技術 S05 資料視覺化能力 S06 資料整合能力 |

| 主要職責 | 工作任務 | 工作產出 | 行為指標 | 職能級別 | 職能內涵 (K=knowledge 知識) | 職能內涵 (S=skills 技能) |
|-----------|--------------------------|--|---|------|---|---|
| | | O1.2.3 資料關聯 (圖)表 | 等。 | | | S07 資料擷取、轉換和載入(ETL)能力 |
| | T1.3 整理、 標註並清洗 資料集 | O1.3.1 測試資料 集 訓練資料集 O1.3.2 標註參照 標準 O1.3.3 資料探索 與品質檢核 | P1.3.1 依據業務需求進行資料分割，分成測試資料及訓練資料，以利達成後續模型訓練成效之驗證。 P1.3.2 與領域專家共同設定標註參照標準及形式，並選擇合適的標註技術，完成符合業務需求的標註資料。 P1.3.3 透過探索式資料分析(EDA)，檢查資料是否異常、缺失或有噪音，並判斷資料是否保留、去除或補齊缺失值，以提高模型精準度。 | 4 | K03 資料庫基本原理 K04 資料庫架構 K05 資料處理與分析概論 K06 統計基本知識 K07 資料標註類型 K08 資料標註技術 | S03 探索式資料分析(EDA) S06 資料整合能力 S07 資料擷取、轉換和載入(ETL)能力 S08 資料分類套件操作能力(如： Scikit-Learn) |
| T2 執行特徵工程 | T2.1 確認執 行框架 | O2.1.1 機器學習 工作流程圖(包含 input/output) | P2.1.1 確認執行機器學習適用框架，並規劃工作流程圖 (包含 input/output)，以達成機器學習業務目標。 | 4 | K01 機器學習演算法基礎知識(決策樹、KNN、SVM、線性回歸、K-means、PCA、Apriori、Q-Learning 等) K09 深度學習演算法基礎知識 (DNN、CNN、RNN.....等) K10 應用程式介面(API) K11 機器學習資訊流概念 | S09 程式語言編碼能力 S10 機器學習框架的架設能力 S11 機器學習常用框架的編碼能力 S12 統計相關性檢定 |
| | T2.2 提取或 轉換特徵值 | O2.2.1 重要特徵 集 | P2.2.1 將原始資料(可能為文字或影像)轉化成更好表達問題本質的特徵，並根 | 5 | K06 統計基本知識 K12 特徵提取概念 | S13 特徵標準化方法 Standardization |

| 主要職責 | 工作任務 | 工作產出 | 行為指標 | 職能級別 | 職能內涵 (K=knowledge 知識) | 職能內涵 (S=skills 技能) |
|----------------|----------------------|-----------------------------|---|------|---|---|
| | (feature extraction) | O2.2.2 定量特徵集 | <p>據特徵與因變數之相關性找出重要特徵，提升模型效能或預測精準度。</p> <p>P2.2.2 識別並利用不同類型的特徵，將定性特徵(類別型特徵)轉換為定量特徵(數值型特徵)，作為輸入供演算法和模型使用。</p> | | K13 資料探勘概念 | <p>S14 歸一化方法</p> <p>Normalization(線型數據歸一化、零均值歸一化)</p> <p>S15 資料降維方法</p> <p>S16 定量特徵分析</p> |
| T3 建立並訓練 ML 模型 | T3.1 建構 ML 初步模型 | O3.1.1ML 初步模型 | P3.1.1 針對目標問題，界定機器學習類型(回歸預測、分類、分群)，並選擇合適演算法(演算法可交互使用)，解決特定問題。 | 5 | <p>K01 機器學習演算法基礎知識(決策樹、KNN、SVM、線性回歸、K-means、PCA、Apriori、Q-Learning 等)</p> <p>K09 深度學習演算法基礎知識(DNN、CNN、RNN.....等)</p> | <p>S09 程式語言編碼能力</p> <p>S10 機器學習框架的架設能力</p> <p>S11 機器學習常用框架的編碼能力</p> <p>S17 機器學習常用框架操作能力。如：scikit-learn、PyTorch、TensorFlow</p> |
| | T3.2 調校參數 | O3.2.1 超參數(Hyperparameters) | P3.2.1 依據目標問題，考量演算法運作細節，並利用參數之調整，產出「超」參數，讓 ML 模型能更精確判斷或預測。 | 5 | <p>K01 機器學習演算法基礎知識(決策樹、KNN、SVM、線性回歸、K-means、PCA、Apriori、Q-Learning 等)</p> <p>K09 深度學習演算法基礎知識(DNN、CNN、RNN.....等)</p> | <p>S10 機器學習框架的架設能力</p> <p>S11 機器學習常用框架的編碼能力</p> <p>S17 機器學習常用框架操作能力。如：scikit-learn、PyTorch、TensorFlow</p> <p>S18 超參數優化技術(包含學習因子 Learning rate、動量梯度下</p> |

| 主要職責 | 工作任務 | 工作產出 | 行為指標 | 職能級別 | 職能內涵 (K=knowledge 知識) | 職能內涵 (S=skills 技能) |
|-----------|---------------|--|--|------|--|---|
| | | | | | | 降因子 Gradient descent with momentum hyper-parameter、學習因子下降參數 Learning rate decay hyper-parameter、神經網絡層數、Adam 算法參數) |
| | T3.3 優化 ML 模型 | O3.3.1 最佳化 ML 模型 | P3.3.1 瞭解優化演算法及損失函數演算法，並經由分析結果，進行 ML 模型優化，穩健的機器學習模型。 | 5 | K01 機器學習演算法基礎知識(決策樹、KNN、SVM、線性回歸、K-means、PCA、Apriori、Q-Learning 等) K09 深度學習演算法基礎知識(DNN、CNN、RNN.....等) K14 優化演算法 Optimizer(批量梯度下降 BGD、隨機梯度下降 SGD、小批量梯度下降 MBGD.....等) K15 損失函數演算法(MSE、RMSE、Cross Entropy) | S10 機器學習框架的架設能力 S11 機器學習常用框架的編碼能力 S19 選擇優化工具(Optimizer) |
| T4 評估模型效能 | T4.1 設定模型驗證指標 | O4.1.1Confusion Matrix 混淆矩陣結果表 O4.1.2 交叉驗證 | P4.1.1 設定模型驗證指標，將「測試資料」置入模型，評估模型成效。 P4.1.2 藉由調整 ML 模型架構，以利 ML 模型能快速運行，並讓 ML 模型更 | 4 | K02 機器學習數學基礎(機率、線性代數、微積分) K16 驗證指標(分類指標、回歸指標.....等) | S09 程式語言編碼能力 S10 機器學習框架的架設能力 S17 機器學習常用框架操作能力。如：scikit-learn、 |

| 主要職責 | 工作任務 | 工作產出 | 行為指標 | 職能級別 | 職能內涵 (K=knowledge 知識) | 職能內涵 (S=skills 技能) |
|-----------|----------------------|---|---|------|--|---|
| | | 圖 | 合適導入軟硬體。 | | K17 混淆矩陣概念 Confusion Matrix K18 模型性能度量 performance measure(包含：準確率 accuracy、精確率 precision、召回率 recall、F1-score、F2-score、F0.5-score、ROC/AUC) | PyTorch、TensorFlow S20 交叉驗證能力(cross-validation) |
| | | | | | | |
| T5 導入應用系統 | T4.2 進行模型測試 | O4.2.1 測試報告 performance metrics O4.2.2 矩陣運算表 | P4.2.1 判斷合適的「實際資料」置入模型進行測試，區分 TP 跟 FP，避免擬合過度(over-fitting)或欠擬合(underfitting)。 | 4 | K02 機器學習數學基礎(機率、線性代數、微積分) K16 驗證指標(分類指標、回歸指標) K17 混淆矩陣概念 Confusion Matrix K19 泛化理論(generalization) | S09 程式語言編碼能力 S10 機器學習框架的架設能力 S17 機器學習常用框架操作能力。如：scikit-learn、PyTorch、TensorFlow S21 統計分析基礎能力 S22 判斷驗證結果能力 S23 執行測試能力 |
| | T5.1 協調部署 ML 模型並持續整合 | O5.1.1 建立可持續進行預測 ML 模型 | P5.1.1 了解部署環境(軟硬體)，提供擴充硬體或加速運算伺服器的建議，協助建構執行模型的基礎架構，確保模型能實現導入。 | 5 | K20 部署 ML 硬體環境知識 (Linux、Container、Embedded System、Edge、Cloud) | S02 業務溝通能力 S09 程式語言編碼能力 S24 程式語言編碼轉換能力 |
| | T5.2 掌握模型監控結果，並修正 | O5.2.1 回饋迴圈(Feedback Loop) | P5.2.1 即時掌握使用者回饋意見，持續修正模型及調整參數，或建立回饋迴圈，以模型具備可移植性或可擴展 | 5 | K20 部署 ML 硬體環境知識 (Linux、Container、Embedded System、Edge、Cloud) | S02 業務溝通能力 S25 整合測試能力 |

| 主要職責 | 工作任務 | 工作產出 | 行為指標 | 職能級別 | 職能內涵 (K=knowledge 知識) | 職能內涵 (S=skills 技能) |
|------|------|------|------|------|----------------------------|-------------------------|
| | 回饋 | | 性。 | | K21 機器學習模型生命周期 | |

職能內涵 (A=attitude 態度)

A01 持續學習

A02 主動積極

A03 正直誠實

A04 壓力容忍

A05 應對不確定性

A06 追求卓越

A07 團隊意識

A08 好奇開放

A09 謹慎細心

說明與補充事項

- 建議擔任此職類/職業之學歷/經歷/或能力條件：

- 資工、電機、電子、自動控制、機械等相關理工科系大學畢業。
- 具備程式語言能力。