

國立中正大學課程大綱

科目：資料科學導論

課程網路連結：<https://ecourse2.ccu.edu.tw/>（公告事項、習題、影片、照片等）

A. 授課老師：賴振耀 mthcyl@ccu.edu.tw

B. 上課時間、地點：二 B、四 B 電算中心 217 教室

C. Office Hours：一、三 14:30~15:30 數學館 416 研究室

課程大綱

週次	單元主題	主要授課內容
1	資料科學導論與開發環境	資料科學概述與基本流程； Jupyter Notebook / Google Colab； Visual Studio Code 簡介； GitHub 基本概念
2	NumPy（一）：陣列與向量化運算	ndarray、資料型態、shape、indexing、slicing、Boolean indexing、向量化運算
3	NumPy（二）：數值與統計運算	Broadcasting、aggregation、random、基本矩陣運算；平均數、變異數、相關係數等統計量的 NumPy 實作
4	pandas（一）：DataFrame 與資料操作	Series、DataFrame、CSV 資料讀取、index、資料選取、篩選、排序、loc / iloc、基本資料檢視
5	pandas（二）：資料整理與前處理	缺失值、重複值、資料型態轉換、groupby、aggregation、merge / join、類別資料處理
6	資料視覺化與探索式資料分析（EDA）	Matplotlib；直方圖、長條圖、散佈圖、箱型圖；敘述統計、相關分析；整合 NumPy 與 pandas 完成基本 EDA
7	AI 輔助程式開發工具（AI Coding Assistants）	AI-assisted programming；程式碼補全、生成、解釋、除錯與重構；Prompt → Code；AI 生成程式碼的測試與驗證；VS Code / GitHub 工作流程
8	AI 輔助 EDA 與資料取得	使用 AI Coding Assistant 協助資料整理與 EDA；公開資料平台；CSV / JSON；API 基本概念；比較人工與 AI 輔助分析流程；AI 輸出驗證
9	機器學習的資料前處理	Feature 與 Target；類別變數編碼；Feature Scaling；Training / Test Split；Data Leakage；scikit-learn 基本工作流程
10	監督式學習：決策樹（一）	Classification 基本概念；Decision Tree 結構；節點與分割；Gini Impurity、Entropy、Information Gain；scikit-learn 實作
11	監督式學習：決策樹（二）	Tree growth；max depth；overfitting；模型複雜度控制；pruning

週次	單元主題	主要授課內容
	*	概念；Decision Tree 視覺化與結果解釋
12	模型評估與驗證	Confusion Matrix；Accuracy、Precision、Recall、F1；類別不平衡問題；Validation；Cross-validation；基本 Hyperparameter Selection
13	非監督式學習：聚類分析	距離與相似度；Feature Scaling；K-means；Centroid；WCSS；K 值選擇；Elbow Method；Silhouette Score；Hierarchical Clustering 簡介
14	降維：主成分分析（PCA） *	高維資料與降維概念；標準化；共變異數；特徵值與特徵向量；Principal Components；Explained Variance；PCA 視覺化與 Python 實作
15	整合式資料科學流程	從問題定義、資料取得、資料清理、EDA、前處理、模型建立到模型評估；以完整案例整合前述方法
16	AI 輔助資料科學：除錯、驗證與批判	使用 AI Coding Assistant 完成資料分析；辨識 AI-generated code / analysis 中的 Data Leakage、錯誤評估、錯誤統計解讀與不合理資料處理；AI 使用倫理、資料隱私與可重現性
17	期末整合專題與討論	完整 Data Science Project；程式碼、分析結果與視覺化成果展示；說明 AI Coding Assistant 的使用方式、人工修改與驗證過程

*：暫定

參考書籍

1. J. VanderPlas, *Python Data Science Handbook*, O'Reilly Media. Free online edition and Jupyter notebooks available from the author's website.
2. G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, *An Introduction to Statistical Learning: with Applications in Python*, Springer.
3. A. Géron, *Hands-On Machine Learning with Scikit-Learn and PyTorch*, O'Reilly Media, 2025.
4. NumPy, pandas, Matplotlib, scikit-learn, GitHub Copilot official documentation.

評量方式

- (1) 課堂測驗：40%
- (2) 作業：20%
- (3) 課堂參與及實作：10%
- (4) 期末整合專題：30%