



《尊重智慧財產權，請使用正版教科書，勿非法影印書籍及教材，以免侵犯他人著作權》

開課班級：四資管二A

授課老師：余榮裕

學分數：3

課程大綱：

AI興起，作為基礎的大數據更為重要，而資料科學相關研究為該領域應用的基礎，其相關知識能力的培養涉及領域甚廣，為期使學生在修習此課程後，能具備資料科學研究領域之輪廓認知，並熟悉相關技術概念與整合。

本課程將從如何成為「資料科學家」開始，描述其工作及流程，逐步介紹並練習實作以下相關方法：1.機器學習方法；2.資料獲取(網路爬蟲、物聯網、API操作)；3.資料清理、儲存及調用；4.自然語言處理；5.資料視覺化等。

使學生善用整合相關方法及工具，從應用操作中認識理論，並建立相關領域學習興趣，期使能繼續深入研究，為此領域作出貢獻。

outline:

The topic of big data is in the ascendant, and data science-related research is the basis for the application of this field. The cultivation of related knowledge and abilities involves a wide range of fields. After taking this course, students will have the outline of the field of data science and be familiar with the related fields. Technical concept and integration. This course will start with how to become a "Data scientist", describe its work and process, and gradually introduce and practice the following related methods: 1. Machine learning methods; 2. Data acquisition (web crawlers, Internet of Things, API operations); 3. Data cleaning, storage and fetch; 4. Natural language processing; 5. Data visualization, etc. To enable students to make good use of relevant methods and tools, to understand theories from applied operations, and to establish learning interests in related fields, hoping to continue in-depth research and contribute to this field.

教學型態：

課堂教學

成績考核方式：

平時成績:40%

期中考:30%

期末考:30%

其它:%

本科目教學目標：

參考書目：

1. Python 資料科學學習手冊, 2/e (Python Data Science Handbook: Essential Tools for Working with Data, 2/e) 2. 資料科學基礎數學 (Essential Math for Data Science) 3.Data Science from Scratch | 用Python 學資料科學, 2/e (中文版) 4.東京大學資料科學家養成全書：使用Python動手學習資料分析



課程進度表：

週次	起訖月日	授課單元(內容)	備註
第1週	2.19~2.25	課程介紹-資料科學家的工作發掘議題，提問：能創造被動式收入的項目？收集、清理資料，爬蟲資料視覺化找出規律說故事更進一步：建立自己的資料模型設計演算法創建輔助使用者的工具、平台	8日正式上課。8~12日課程加退選，轉學(系)生、復學生及延修生選課，雙主修、輔系申請，12日申辦抵免學分截止日
第2週	2.26~3.03	表訂：機器學習方法介紹-Scikit-Learn、k最近鄰居法、迴歸、決策樹； 實務：Python環境與相關應用 講解內容參考：電子書：Python自學聖經：從程式素人到開發強者的技術與實戰大全！ 開發環境：Google Colab 程式練習參考： 1. 精通 Python 2. Github: Python 基礎必修課習題練習 加入 CodeCombat 班級，從遊戲中學習程式：;https://codecombat.com/students?_cc=LeafBagPan班級代碼：LeafBagPan；	
第3週	3.04~3.10	機器學習方法介紹-隱馬爾科夫模型(Hidden Markov Model, HMM)及循環神經網路(Recurrent Neural Network, RNN)	28日(日)孔子誕辰紀念日/教師節(放假),29日(一)補假
第4週	3.11~3.17	資料獲取-網路爬蟲	29日成績優異提前畢業者提出申請截止日
第5週	3.18~3.24	資料獲取-物聯網及API應用操作	6日(一)中秋節(放假)，10日(五)國慶日(放假)
第6週	3.25~3.31	資料清理-正則表達式、資料型別轉換、遺漏資料與異常值處理、時間序列資料處理及重塑資料	14日學生宿舍安全輔導暨複合式防災疏散演練。18日多益測驗
第7週	4.01~4.07	資料儲存及調用實作-關聯式資料庫及結構化查詢語言，以MySQL為例	24日(五)補假，25日(六)光復暨古寧頭大捷日(放假)。
第8週	4.08~4.14	資料儲存及調用實作-NoSQL/鍵值(Key-Value)資料庫，以 MongoDB 為例，進行 CRUD 操作	30日校課程委員會
第9週	4.15~4.21	期中考	3~9日期中考試
第10週	4.22~4.28	自然語言處理-方法、工具介紹與專題分組(自然語言處理 NLP), 大型語言模型 LLM)	13日教務會議,16日教師期中成績上網登錄截止日
第11週	4.29~5.05	表訂：自然語言處理-科學運算工具NumPy、Panda實務：安裝 Selenium，模擬人類自動化操作瀏覽器，抓取網站資料並寫入資料庫	
第12週	5.06~5.12	表訂：自然語言處理-機器學習工具Scikit-Learn實務：架設 Flask 服務	24~28體育運動週。24日校園路跑。27日運動大會夜間開幕，28日運動大會活動，29



			日101週年校慶活動日，照常上班
第13週	5.13~5.19	表訂：自然語言處理-資料建模和數據分析 競賽平台Kaggle 實務：Discord 介接及 Git 專案版控	
第14週	5.20~5.26	表訂：自然語言處理-全文檢索、文本分詞(Text Segmentation)、建立字典及轉換數字序列 實務：ChatGPT 及 Stable Diffusion 原理介紹及應用討論	12日申請停修課程截止日
第15週	5.27~6.02	表訂：自然語言處理-中文(多)語言處理Stanford CoreNLP, Stanza 實務：NumPy 及 Panda	
第16週	6.03~6.09	資料視覺化-Matplotlib, Dash, amCharts	22日校務會議。25日行憲紀念日(放假)
第17週	6.10~6.16	分組報告 期末報告，於第 17, 18 周報告，使用投影片口頭報告或佐以部份操作錄影，作業要求： 1.使用爬蟲程式擷取並清理資料，或連接某服務 API 2.使用資料庫：MongoDB、MySQL、MS-SQL 皆可 3.提供 Web Service，例：Flask、Node.js 或 .Net 皆可 4.介接 Discord 或 LineBot，提供使用者互動服務介面 特別加分，以下擇一：有使用 Panda 整理資料、使用統計預測模型、使用任一機器學習演算法或介接 ChatGPT API 提供客製化服務。 服務互動回應可以是：文字、聲音、圖檔、影片； 題目的發想：儘量限縮在一個清晰易懂的服務即可，不用包山包海使用 Panda 整理資料及統計預測模型，通常是數值，可以採用 Matplotlib 來輸出圖表，這也是加分項。	1日(四)開國紀念日(放假)
第18週	6.17~6.23	期末考	5~11日期末考試，10~11日學生退宿