

# Schedule for Daniel

## 關於我

- 姓名：葉瀚元 Stephen
- 居住地：台中市
- 年齡：22
- 聯絡資訊：
  - line id: wenyeh0214
  - 電話：0905056920
- 專業領域：資料科學、作業研究、創新創業、後端開發、AI 產品開發、LLM
- 熟悉程式語言：Python, SQL, HTML, Javascript
- 教育背景：
  - 高中：國立中興大學附屬高中
  - 大學：國立政治大學資訊管理學系 學士班 已畢業
  - 研究所：國立政治大學資訊管理學系 碩士班 錄取未就讀
- 經歷：
  - 2023/03~Now: 未來巢科技股份有限公司 RD 後端工程師 & AI 產品工程師 實習生
  - 2024/03~Now: 決策與數量分析實驗室 研究助理
  - 2023/09-2024/1: 政大資訊管理課程助教
  - 2022/09-2023/07: 政大數據分析社 社員 參與 Acer 宏碁公司企業專案
- 重要成就：
  - 專利：兩項與 AI 相關專利申請中
  - 學業表現：大學四年總平均 88.66 分，於班排名為 12/39 名
  - 重要獲獎紀錄：
    - 2023 X 創業家 - Pitch 擂台賽金牌獎
    - 2022 未來巢資料科學家培訓課程 - 第二名結業
    - 2022 政大創客松競賽 - 最佳 MVP 獎
    - 2022 政大數據分析社 數據分析競賽 - 最佳數據應用獎
- 自傳：

Daniel 以及 Daniel 爸爸媽媽您好，我是葉瀚元 Stephen，目前是政大資管系大四的畢業生，並且已錄取政大資管研究所，目前正在等待就讀中。我的專長領域是程式開發、資料科學，大學四年參與過數據分析

社，以及進修過各 AI 專業領域的大學課程，並且也自主進行過多項的資料科學專題，領域包含金融、行銷、製造、電商、網路資訊安全等等，也實際參與過與 Acer 宏碁的數據分析專案，對於資料科學的專案已有相當多經驗。我是一個對 AI, 新創抱有極大興趣的人，目前我任職於新創公司擔任 AI 產品開發工程師，並且成功為公司開發出兩項專利。因此，我在資料科學中有學術上與實戰中的經驗。另外，我也熱愛分享知識，在公司中也常常與夥伴舉辦工作坊分享最近的研究成果，我喜歡共同成長的感覺。

目前除了等待研究所以外，也積極存錢中，因為看到了自己得規劃以及想挑戰自我的下定決心，未來期望能到美國攻讀二碩，並於美國就業，這是我從小到大的夢想之一。

我與 Daniel 是在景美溪河堤旁籃球場認識，當下聽聞 Daniel 有出國攻讀電腦相關科系學位的規劃，我是充滿敬佩的。近幾年，因為企業發現資料科學對於企業營運有巨大幫助，且硬體已逐漸普及，因此資料科學領域的概念在近期 AI 爆發之前就已相當熱門，美國眾多知名大學也開出相當多相關多的 program 得以申請。而在 AI 爆發之後（黃仁勳），未來幾年將會是 AI 發展最為迅速得幾年，而企業端需要的人才也越來越多，相信未來的申請機會、工作機會將會越來越多。資料科學中，最中意的概念就是 AI，因此能夠通盤的了解這項專業領域，一者是有幫助 Daniel 更了解自己喜歡的電腦科學領域，二來是可以讓 Daniel 於 AI 快速發展中能有升學與就業的機會，最後是了解 AI 才不會對 AI 有不必要的恐懼，並且能與 AI 為伍，讓自己更具競爭力，有足夠知識面對未來的挑戰。

本堂課將資料科學專業領域，特別結合了 Daniel 有興趣的資訊安全概念，希望將完整的觀念能夠帶給 Daniel，並在過程中分享給這位優秀的年輕人我過去所學的、所做的、所想的，希望幫助 Daniel 找到自己的興趣，並抱有熱情的前往深造。我認為 Daniel 相當有潛力成為一位專業的電腦科學人才。

在此謝謝 Daniel 與 Daniel 爸爸媽媽  
有任何問題歡迎請以 line 聯繫我  
期望我們一起進步 學習過程合作愉快

祝順心 身體健康

葉瀚元 敬上

## 關於本堂課程

- 課程名稱：網路安全與數據科學基礎家教課程
- 課程日期：7/10~9/4（最少上完 12 堂，後續可彈性調整）
- 課程時間：每週三、五，早上 9:00~12:00，每週兩堂課，每堂課共三小時
- 課程費用：新台幣 900 800 元 整（每小時）
- 上課地點：國立政治大學達賢圖書館 3 樓視訊討論空間
- 課程內容：
  - 資料科學知識(80%)

- 網路資安觀念(5%)
- 專案實作經驗(10%)
- 其他後端開發與程式設計知識(5%)
- 課程目標：
  - 獲得基本 python 與活用資料科學相關工具能力
  - 獲得基礎性、綜覽性資料科學領域知識能力
  - 獲得獨立完成資料科學專案之能力
  - 保存本次家教課程作業、專案為作品集，有利未來申請
  - 獲得一個終生保固資料科學知識的良師益友(指的是本人😁)
- 上課方式：
  - 教材將會上傳至 Github，學生將本週教材下載至 Google drive
  - 以 Colab 為主要上課資源，並搭配手寫筆記
  - 本課程結束前，學生會完成一項專案，詳情請見下文「關於作業與專案」
- 其他事項：
  - 若上課時間有異動，則將會請前一個禮拜提前告知學生，也請同學一個禮拜前告知我下週是否上課時間有所異動，以便彼此安排
  - 課程內容將希望最大化將資料科學領域知識帶給同學，原則上以行事曆預先規劃之進度為主。然若需進行進度調整，會與同學達成共識後，調整上課內容。

## 關於作業與專案

### Homework:

- 每 1~3 週會有一次作業，依據上課進度決定
- 作業包含每週上課的觀念與實作
- 每堂課的最後 30 分鐘，將會有作業與專案的討論時間

### Project:

- 本堂課結束之前，學生將會完成一份與資料科學相關的專案
- 題目由學生自行挑選，並由我協助調整成有意義、足夠挑戰性的專案題目，並且在之後的課程輔導學生完成專案
- 專案將會在 class 5 之後展開，並於 9/4 結束
- 每堂課的最後 30 分鐘，將會有作業與專案的討論時間

## Schedule (Summer 2024)

## 1. Class 1 (7/10)

- a. 教材 Github 下載
- b. Google Colab 介紹
- c. Python Introduction (Python 簡介)
- d. Numpy & Pandas (Numpy 與 Pandas)
- e. 作業繳交方式與第一次作業介紹
  - i. Github 下載作業介紹 & 繳交作業方式
  - ii. HW1 - H-ZeroAccess 介紹

## 2. Class 2 (7/17)

- a. 網路資訊安全介紹
  - i. Security Management (安全管理)
  - ii. Malicious Software and Cyberattacks (惡意軟體與網路攻擊)
- b. Data Science 資料科學領域介紹
  - i. 資料科學領域介紹
  - ii. 資料工程 & 資料科學 & 資料分析 介紹
  - iii. 資料科學專案流程
- c. Visualization 資料視覺化介紹
  - i. 資料視覺化介紹與重要性
  - ii. 資料視覺化工具介紹 (Matplotlib & Seaborn)

## 3. Class 3 (7/19)

- a. Overview for Machine Learning & Deep Learning (機器學習與深度學習概述)
  - i. 機器學習 (Machine Learning) 模型介紹
  - ii. 深度學習 (Deep Learning) 模型介紹
- b. Regression Model (迴歸模型)
  - i. Regression Model Introduction (迴歸模型介紹)
  - ii. Linear Regression (MSE, Gradient Descent) (線性迴歸)
  - iii. Orange Workflow for Regression (Orange 迴歸工作流程)

## 4. Class 4 (7/24)

- a. Classification Model (分類模型)
  - i. Classification Model Introduction (分類模型介紹)
  - ii. Logistic Regression (邏輯迴歸)
  - iii. SVM (Support Vector Machine) (支持向量機)

- iv. Cross-Entropy (交叉熵)
- v. Evaluation (評估)
- vi. Orange Workflow for Classification (Orange 分類工作流程)
- b. Tree Model (樹模型)
  - i. Tree Model Introduction (樹模型介紹)
  - ii. Tree and Random Forest (決策樹與隨機森林)
  - iii. Entropy, Information Gain, Gini, Chi, Variance (熵、信息增益、基尼、卡方、變異數)
  - iv. Orange Workflow for Tree (Orange 樹模型工作流程)
- 5. Class 5 (7/26)
  - a. Clustering (聚類)
    - i. Clustering Model Introduction (聚類模型介紹)
    - ii. Distance (距離)
    - iii. K-means (K均值)
    - iv. Hierarchical Clustering (層次聚類)
    - v. DBScan (密度聚類)
    - vi. Orange Workflow for Clustering (Orange 聚類工作流程)
  - b. Problematic Data (問題數據)
    - i. Dimension Reduction, PCA (降維與主成分分析)
    - ii. Problematic Data (問題數據)
- 6. Class 6 (7/31)
  - a. Neural Network Basics (神經網絡基礎)
    - i. Basic Linear Algebra & Calculus Math Concept (基礎線性代數與微積分概念)
    - ii. Backpropagation Algorithm (反向傳播演算法)
    - iii. Activation Function (激活函數)
      - 1. Activation (激活)
      - 2. Optimizing Gradient-Descent (優化梯度下降)
    - iv. Overview for Variant Models (變體模型概述)
  - b. Tensorflow and Pytorch Introduction (TensorFlow 與 PyTorch 介紹)
- 7. Class 7 (8/2)
  - a. Convolution Neural Network (CNN) (卷積神經網絡)
    - i. Basic Concept (基本概念)
    - ii. LaNet (LaNet 模型)

- iii. Other Models (其他模型)
  - iv. Application in Cybersecurity: Static Analysis: Windows PE File and Image Analysis (在網絡安全中的應用：靜態分析：Windows PE 文件和圖像分析)
- 8. Class 8 (8/7)
  - a. Recurrent Neural Network (RNN) Introduction (循環神經網絡介紹)
    - i. Basic Concept (基本概念)
    - ii. Understanding LSTM Networks
      - 1. LSTM, GRU, ResNet
    - iii. Application in Cybersecurity: (在網絡安全中的應用)
      - 1. Dynamic Analysis: Malware Call and Sequence Analysis (動態分析：惡意軟體呼叫和序列分析)
      - 2. Text Classification with an RNN (使用 RNN 進行文本分類)
- 9. Class 9 (8/9)
  - a. Latent Space (潛在空間)
    - i. Basic Concept (基本概念)
    - ii. Auto-Encoder (N04) (自編碼器)
      - 1. Convolutional, Variational AE (VAE) (卷積、自變 AE)
      - 2. Denoising, Anomaly Detection (去噪、異常檢測)
- 10. Class 10 (8/14)
  - a. Natural Language Processing (NLP) (自然語言處理)
    - i. Introduction (介紹)
    - ii. Text Process (文本處理)
      - 1. Text Preprocessing (文本預處理)
      - 2. Feature Extraction (特徵提取)
- 11. Class 11 (8/16)
  - a. Language Model (語言模型)
    - i. Word2vec (CBOW, Skip-gram), fastText (supervised, unsupervised), Markov Chains (Word2vec、fastText、馬爾可夫鏈)
    - ii. Transformer, Self-Attention, BERT (Transformer、自注意力、BERT)
- 12. Class 12 (8/21)
  - a. Natural Language Understanding (NLU) (自然語言理解)
    - i. Introduction (介紹)
    - ii. Named Entity Recognition (NER) (命名實體識別)

- iii. Sentiment Analysis (情感分析)
  - iv. Text Classification (文本分類)
  - v. Others (其他)
  - b. Natural Language Generation (NLG) (自然語言生成)
    - i. Introduction (介紹)
    - ii. Text Generation (文本生成)
    - iii. Machine Translation (機器翻譯)
    - iv. Others (其他)
13. Class 13 (8/23)
- a. Language Model and Others (語言模型與其他)
    - Transfer Learning & Fine-Tuning (遷移學習與微調)
    - LoRA, Parameter-Efficient Fine-Tuning (PEFT) (LoRA、參數高效微調)
    - Classification on Imbalanced Data (不平衡數據上的分類)
14. Class 14 (8/28)
- a. Anomaly Detection Introduction (異常檢測介紹)
    - i. VAE (變分自編碼器)
    - ii. V. Chandola, A. Banerjee and V. Kumar, "Anomaly Detection: A Survey," ACM Computing Survey, vol. 41, no. 3, July 2009.
    - iii. Novelty and Outlier Detection (新穎性和異常檢測)
      - 1. One-Class SVM (單類支持向量機)
    - iv. Self-Organized Map (自組織映射)
15. Class 15 (8/30)
- a. Image Generation (圖像生成)
    - i. Introduction (介紹)
    - ii. Stable Diffusion (穩定擴散)
16. Class 16 (9/4)
- a. Information Retrieval (信息檢索)
    - i. Introduction (介紹)
    - ii. RAG Introduction (RAG 介紹)
    - iii. RAG Tools Introduction (RAG 工具介紹)
    - iv. Agent in LLM (大型語言模型中的代理)
    - v. Make Your Own LLM App! (製作你的 LLM 應用)