

## 國立清華大學課程大綱

科號 Course Number		學分 Credit	3	人數限制 Class Size	30
中文名稱 Course Title	資料科學與數據分析				
英文名稱 Course English Title	Data Science and Analytics				
任課教師 Instructor	莊鈞翔				
上課時間 Time		上課教室 Room			

**課程簡述(必填) (最多 500 個中文字) 本欄位資料會上傳教育部課程網**

**Brief Course Description (required) (50-200 words if possible, up to 1000 letters)**

資料科學與數據分析是從大量數據中提取有意義的訊息和模式的藝術和科學，它將用於辨識數據中之模式，同時以統計方法與相關技術相結合進行推理，包括認識大數據庫、視覺化技術和近代熱門的機器學習與人工智慧技術，本課程將提供學生其廣泛的背景知識，並且介紹相關領域最為人所熟知的概念和應用，學生可應用於心智科學或教育大數據等資料。成功完成本課程後，學生應能夠：

1. 展示有關資料科學和知識發現過程的背景知識。
2. 描述數據探究時涉及的方法、範圍和局限性。
3. 應用資料科學技術的實踐知能以及分析技能。
4. 應用機器學習等相關技術進行科學探索。
5. 設定專案目標、完成專案開發設計，並且展示其內容。
6. 具備執行數據科學與機器學習應用之基礎程式(Python, KNIME, or Matlab)。

**請輸入課程內容「中文暨英文關鍵字」至少 5 個，每個關鍵字至多 20 個中文，以半形逗點分隔 (必填)**

**Please fill in at least 5 course keywords (up to 40 letters for each keyword) and use commas to separate them.(required)**

數據科學(Data Science)、機器學習(Machine Learning)、統計(Statistics)、大數據(Big Data)、教育(Education)、心智科學(Mind Science)

**課程大綱 Detailed Course Syllabus**

● **課程說明(Course Description)**

本課程將涵蓋以下主題：

- 1.數據科學簡介：問題形成、數據挖掘概念、數據類型、挖掘和知識發現過程

( CRISP DM 方法，SAS SEMMA 方法 )，數據挖掘和知識發現之間的差異，可以發現什麼；發現模式的“有趣性”、“有用性”和“新穎性”的概念；應用領域概述，數據挖掘中之專業。

2. 視覺化：資料可視化技術及其在資料探索中的應用、視覺化方法。

3. 數據預處理和轉換：小型和大型資料集、資料採集、缺失數據、數據清洗技術；刪除敏感信息、法律問題的技術。

4. 分類與預測：分類與預測的問題與其技術。

5. 分群：聚類分析的問題；劃分方法、分層方法；異常值檢測和分析。

6. 挖掘時間序列數據簡介：趨勢分析和周期性分析；時間序列中的相似性搜尋等。

7. 部署：模型部署與應用至期末專案中。

### ● 指定用書(Text Books)

自編。

### ● 參考書籍(References)

Pang-Ning Tan, Michael Steinbach and Vipin Kumar (2006). Introduction to Data Mining, Addison-Wesley.

黃志勝，機器學習的統計基礎，旗標，2021。

### ● 教學方式(Teaching Method)

每週含課程內容講述與上機實作，配合學生程式能力(Python 或 Matlab)。

### ● 教學進度(Syllabus)

Week 1. 資料科學與數據分析簡介

Week 2. 統計基礎

Week 3. 數據與資料庫

Week 4. 數據視覺化

Week 5. 特徵選取

Week 6. 特徵萃取

Week 7. 非監督式學習

Week 8. 期中考

Week 9. 監督式學習

Week 10. 模型表現評估

Week 11. 分類

Week 12. 預測

Week 13. 深度學習理論

Week 14. 深度學習實踐

Week 15. 期末專題驗收與除蟲

## Week 16. 期末專題展示

註：每週皆有上機實作，使學生具備基礎程式能力。

### ● 成績考核(Evaluation)

作業(35%)

期中考(30%)

期末專案(35%)

### ● 可連結之網頁位址 相關網頁(Personal Website)