

國立中正大學通識教育課程教學大綱

開課學年度/學期	115 學年度第 1 學期		
課程名稱(中文)	生活中的資料科學		
課程名稱(英文)	Data Science in Real Life		
課 碼	7507026	學分數	2
授 課 方 式	請勾選(可複選)： ☒ 課堂授課 ☐ 網路教學 ☒ 分組討論 ☐ 校外教學 ☒ 其他 <u>電腦教室授課與實作</u>		
教學目標及範圍	當全世界各行各業數位化日益普及的趨勢之下，資料（Data）如同空氣般的充斥在您我生活的周遭。近幾年，在資訊與通信科技（Information and Communication Technology, ICT）快速發展的推波助瀾之下，人們除了已經能有效率的獲取（Acquisition）大量且多樣的資料之外，並透過發展資料分析模型與決策模式，成為協助人類在決策過程中的重要工具，進而形成當代重要領域學科：資料科學（Data Science）。資料科學是一門專業、多面向、跨領域的新興學科，而本課程則是設計為資料科學之基礎概念課程，修習本課程之學生，並不需要具備任何計算機科學或其相關知識基礎，事實上，我們更鼓勵來自非資訊及非數統背景領域的學生選修這門課程。 在現實生活中，許多定量（Quantitative）和以資料為中心（Data-Centric）的問題可以通過資料思維（Data-Driven）制訂決策模式來解決問題。但不幸的是，在我們的現實生活中所面臨各式各樣的資料，往往是混亂的！因此本課程將著重在訓練學生對資料特徵認識、蒐集、及前處理技巧，並透過案例實際操作（Hands-on）對資料具備一定的分析能力與推論程度，並將這些概念運用到管理現實生活中的分析，為您在未來具備解決真實生活問題之能力奠定基礎。 本課程是一門進入資料科學領域之前導（Prerequisite）課程，旨在幫助學生能快速掌握現實生活中的資料，本課程的目標是在以不偏廢相關基本學科（統計學與機器學習演算原理）的情況下，盡可能的提供概念性、基礎性、容易理解的課程內容，同時避免涉及進階程式設計技術與艱深模型理論，而是以簡單的機器學習模型範例來啟發學生對於資料科學之興趣，並實踐於日常生活中的應用，以做為本課程的聚焦方向。 本課程的學習目標如下： 1. 培養學生具備資料思維能力與素養。 2. 培養獨特的問題分析與演算能力。 3. 具備使用資料與邏輯，建構決策模式（Decision-Making）的基本能力。 4. 具備描述性分析的資料科學經驗。 5. 以機器學習演算法工具，完成資料分析體驗。 6. 啟發學生對資料科學領域之興趣，並實踐於生活中之應用。		

授課大綱
(須含週次表及每週課程進度說明)

週次	課程進度
一	課程與評量介紹
二	資料思維與生活
三	資料科學現況與商業應用
四	資料分析成熟度模型(Data Analytics Maturity Model)
五	資料型態與描述性分析(Descriptive Analytics)介紹
六	預測性分析(Predictive Analytics)原理簡介
七	指示性分析(Prescriptive Analytics)原理簡介
八	機器學習與傳統統計學
九	期中考 - 繳交專題提案書(Proposal)
十	資料準備階段(1) - 資料蒐集
十一	資料準備階段(2) - 連續型(Continuous)資料處理
十二	資料準備階段(3) - 離散型(Discrete)資料處理
十三	資料準備階段(4) - 資料正規化及標準化介紹
十四	資料準備階段(5) - 其他資料預處理(Data Preprocessing)技巧
十五	資料分析階段(1) - Orange-Data Mining 與內建演算法介紹
十六	資料分析階段(2) - 實際案例演練
十七	案例討論(Case Study)
十八	期末報告

第一週：課程與評量介紹

1. 教學內容：概觀性的介紹本課程內容及預期學習效果。
2. 教學目標：讓學生瞭解本課程內容範圍、規則與評分標準，以利於修課學生自我規劃學習策略。

第二週：資料思維與生活

1. 教學內容：藉由實際的案例來說明資料分析實現於生活中的應用，以深入淺出的方式，帶領學生自發性聯想資料思維與機器學習背後原理。藉由範例與背後物理意義的闡釋與定義，引導學生認識資料科學與機器學習之間的關連性與概念。
2. 教學目標：學生具備資料思維與機器學習基本概念，以及可以實踐在生活中應用的聯想。
3. 延伸閱讀：讓資料為你產生價值、架構你的資料策略 (哈佛商業評論，2017)。

第三週：資料科學現況與商業應用

1. 教學內容：以更廣義的思維簡介資料科學廣泛應用在各領域之現況，並前瞻性的介紹資料科學在當代商業分析 (Business Analytics, BA) 與商業智慧 (Business Intelligence, BI) 的角色與發展 (應用) 現況。

2. 教學目標：學生瞭解資料科學在商業可應用的模式與概念，以吸引學生對資料科學之興趣，利於學生未來學習規劃有進一步的認識。
3. 延伸閱讀：
 - (1) 陳昇瑋 & 溫怡玲 (2019)。人工智慧在台灣：產業轉型的契機與挑戰從數據到決策。台北市：天下文化。
 - (2) Gartner Top 10 Trends in Data and Analytics for 2020
(<https://www.gartner.com/smarterwithgartner/gartner-top-10-trends-in-data-and-analytics-for-2020/>)。

第四週：資料分析成熟度模型(Data Analytics Maturity Model)

1. 教學內容：本單元將藉由市場研究機構 Gartner 所提出來的資料分析成熟度模型，以及其他資料管理(治理)模型(例如：Carnegie Mellon 大學所提出的 DMM Model)之大數據(或開放資料)評估準則、授權模式、安全性原則，並說明決策者如何規劃資料分析工作，以及所需考慮的策略方向。
2. 教學目標：學生瞭解規劃資料分析所必備之工作內容以及具備資料管理概念。
3. 延伸閱讀：Getting Started with Data Literacy (Valerie Logan & Alan D. Duncan, 2018)

第五週：資料型態與描述性分析(Descriptive Analytics)介紹

1. 教學內容：本單元將介紹有關於資料本質(例如：型態、維度、類型、切片尺度、特徵、連續、非連續、次序、名目...等)表現；與此同時，本單元並進一步的說明，如何將現實生活中所蒐集到的原始資料(Raw Data)，透過描述性分析的技術，反射於描述問題事實的模式。
2. 教學目標：學生能將生活中所面臨的問題，將其轉換以資料描述性分析的呈現方式，做為陳述問題的能力。
3. 延伸閱讀：Turning Data into Actionable Insights (<https://towardsdatascience.com/>)

第六週：預測性分析(Predictive Analytics)原理簡介

1. 教學內容：本單元說明如何藉由資料與數值的分析，去預測問題變數(自變數與依變數)的規律性與規則(Rules)，並嘗試歸納出問題行為模式(Pattern)之概念技術。藉由這個演繹推理過程，找出主要導致此模式的原則或因素，進行預測性分析。最後，再以規則形式，並做為這個模式的敘述方式。除此之外，本單元也將介紹預測性分析常用的模式與工具。
2. 教學目標：培養學生具備洞悉問題本質與因果關係，並瞭解資料如何透過預測性分析模式及工具，呈現預測模式的方法。
3. 延伸閱讀：Turning Data into Actionable Insights (<https://towardsdatascience.com/>)

第七週：指示性分析(Prescriptive Analytics)原理簡介

1. 教學內容：基於前一單元演繹推理基礎，本單元著重於問題命題(Proposition)與演算法設計(Algorithm Design)思考訓練，並藉由案例的說明，介紹如何基於可預測性分析的結果，利用機器學習演算法，演算出可提供決策者決策之「建議性」結果之分析流程。
2. 教學目標：培養學生具備命題與歸納演算涵養，並有能力設計一個簡單演算法，來詮釋與推論大多數類似問題現象。
3. 延伸閱讀：Turning Data into Actionable Insights (<https://towardsdatascience.com/>)

第八週：機器學習與傳統統計學

1. 教學內容：本單元主要基於 5-7 週的教學總結，介紹在不同資料分析成熟度情境之下，機器學習演算法與傳統統計學支持資料科學所分別扮演的角色。另一方面，本單元也將從計量學概念（特別是資訊計量學），以實際生活案例讓學生瞭解如何將問題的陳述，拆解為計量單元，並以數學與統計的量化方法，應用於支持機器學習演算分析模式與其他形式的媒體上。
2. 教學目標：學生瞭解計量學基本運作原理，並具備分解問題為計量單元的能力，始能進而以電腦運算思維的方式，提出解決問題的方案。
3. 延伸閱讀：The Actual Difference Between Statistics and Machine Learning (<https://towardsdatascience.com/>)

第九週：期中考 - 繳交專題提案書(Proposal)

1. 教學內容：回顧並摘要過去 8 週課程內容，並經由在實際生活中遭遇到的問題（或創新應用），構思如何運用資料提出可行解決問題方案，提交期末專題提案書(Proposal)。
2. 教學目標：量測學生至目前為止對於課程內容的理解程度，做為後續課程內容調整或補救教學參考指標。
3. 延伸閱讀：Rochelle et al. (2017). *Designing with Data*. CA: O'Reilly Media.

第十週：資料準備階段(1) - 資料蒐集

1. 教學內容：本單元介紹專題提案書規劃預期結果（Outcome）及關心的變數（Variable）所需資料蒐集基本原理（例如：每日記錄、資料庫擷取、API 擷取、網路爬蟲...等）。除此之外，本單元亦同時介紹資料切分技巧（例如：特殊字元、HTML、JSON、XML...等解析方法）。另一方面，本單元將引介網路 Benchmark Dataset 資源，讓學生選擇有興趣的資料主題，用以練習資料蒐集與整理技巧。
2. 教學目標：透過實作與課堂介紹，讓學生學習資料蒐集技巧與整理程序。
3. 延伸閱讀：Joel Grus (2019). *Data Science from Scratch*, 2/e (Ch9 – Ch10) CA: O'Reilly Media.

第十一週：資料準備階段(2) - 連續型(Continuous)資料處理

1. 教學內容：本單元除了介紹連續型資料基本定義之外，並以實際案例說明連續型資料離散化（Discretization）切片尺度準則（例如：名目尺度、順序尺度、區間尺度、比例尺度...等），將連續型資料轉換成摘要性資料，以利於後續機器學習工具匯入使用。
2. 教學目標：本單元透過課堂講解與 Excel 函數練習的方式，讓學生瞭解連續型資料基本定義及熟悉資料離散化的技巧與能力。
3. 延伸閱讀：Data Preprocessing in Data Mining (Springer e-book)

第十二週：資料準備階段(3) - 離散型(Discrete)資料處理

1. 教學內容：本單元除了介紹離散型資料基本定義之外，並以實際案例說明離散型資料類別標籤化準則（例如：Ordinal vs. Nominal 處理），並予以編碼成適合機器學習演算的資料樣式（One-Hot Encoding），以利於後續機器學習工具匯入使用。
2. 教學目標：本單元透過課堂講解與作業練習的方式，讓學生瞭解離散型資料基本定義及熟悉類別資料編碼技巧與能力。

3. 延伸閱讀：Data Preprocessing in Data Mining (Springer e-book)

第十三週：資料準備階段(4) - 資料正規化及標準化介紹

1. 教學內容：當我們所蒐集到的資料表示單位尺度不同（例如：年收入 vs. 體重）這將影響資料分析的結果。為處理這個情況，本單元將介紹如何以正規化（Normalization）的方法（例如：最小值最大值正規化），以及採取標準化（Standardization）的方法（例如：Z 分數標準化），將原始資料按比例縮放於[0, 1]區間中，且不改變其原分佈方法，以提供一個具相同的基準進行資料分析。
2. 教學目標：學生將學習到資料正規化及標準化技巧，並具備對實驗資料集進行實作之能力。
3. 延伸閱讀：Data Preprocessing in Data Mining (Springer e-book)

第十四週：資料準備階段(5) - 資料預處理(Data Preprocessing)技巧

1. 教學內容：本單元介紹資料預處理技術（特別是：資料清理技術），包含：去除極端值(Outlier)、去除雜訊(Noisy data)、過濾重覆記錄、缺失值(Missing value)處理、資料特徵縮放、正規化與相關預處理函數具體操作練習。
2. 教學目標：經由課堂講解與作業練習的方式，瞭解資料預處理相關技巧，並具備整理成可匯入機器學習演算的正規化資料之能力。
3. 延伸閱讀：The Ultimate Guide to Data Cleaning (<https://towardsdatascience.com/>)

第十五週：資料分析階段(1) - Kaggle 與內建演算法介紹

1. 教學內容：本單元介紹機器學習實作平台：Orange-Data Mining 執行環境、具體操作（資料匯入）、資料預處理練習、及演算法適用特性與用法，並深入介紹分類、分群、及關聯法則演算法的操作及結果解釋。本單元亦將以 Kaggle Dataset 案例，提供實戰展示，及學生提案書內容討論與觀摩，以確保執行期末報告之品質。
2. 教學目標：學生至少熟悉機器學習工作平台其中一種演算法使用，並具備對實驗腳本結果解釋之能力。
3. 延伸閱讀：Orange-Data Mining Tutorials (<https://orangedatamining.com/getting-started/>)

第十六週：資料分析階段(2) - 實際案例演練

1. 教學內容：本週除讓學生實際上機練習與製作期末專題報告之外，也將開放期末專題報告演練與討論，藉由這些務實應用，激發學生對資料思維的創新與實踐具深刻省思。
2. 教學目標：熟稔機器學習工作平台操作技巧，培養實踐創意基本能力。

第十七週：案例討論(Case Study)

1. 教學內容：本週除讓學生實際上機練習與製作期末專題報告之外，亦開始進行學生分組專題報告（10-15 分鐘/組）。
2. 教學目標：對本課程學習深刻體驗，並強化表達能力。

第十八週：期末報告

1. 教學內容：本週進行學生分組專題報告（10-15 分鐘/組），並繳交期末專題報告。
2. 教學目標：對本課程學習深刻體驗，並強化表達能力。

教科書及 延伸閱讀	1. Data Science Author : John D. Kelleher & Brendan Tierney Publisher : The MIT Press; Illustrated edition (April 13, 2018) ISBN-10 : 0262535432 ISBN-13 : 9780262535434 2. Business Analytics, Volume I: A Data-Driven Decision-Making Approach for Business. Author : Amar Sahay Publisher : Business Expert Press ISBN-13 : 9781631573316 3. 機器學習工作平台 Orange-Data Mining : https://orangedatamining.com/ 4. Kaggle (Machine Learning and Data Science Community) : https://www.kaggle.com/datasets 5. 自編講義。
評 量 方 式	請勾選(可複選)，並填寫類別： ☒ 課堂參與， <u>A</u> 類 ☐ 期 中 考， ____ 類 ☐ 期 末 考， ____ 類 ☒ 小組報告， <u>C</u> 類 ☐ 小組討論， ____ 類 ☐ 書面報告， ____ 類 ☐ 課後作業， ____ 類 ☐ 平時測驗， ____ 類 ☐ 心得分享， ____ 類 ☐ 學習紀錄， ____ 類 ☒ 專題創作， <u>B</u> 類 ☐ 其他____， ____ 類 A 類佔 <u>20</u> %；B 類佔 <u>40</u> %；C 類佔 <u>40</u> % (類別可自行增加)
與聯合國永續發展 目標(SDGs)及 細項之對應 (請參閱 SDGs 對照表)	目標: <u>4</u> 細項: <u>4.3</u> <u>4.7</u> 目標: _____ 細項: _____ 目標: _____ 細項: _____ (至多三個目標，每個目標至多三個細項) 範例： 目標: <u>4</u> 細項: <u>4.3</u> <u>4.5</u> <u>4.7</u>

核心能力指標設定	通識課程 核心能力指標	說明	本課程能 培養學生 此項核心 能力者請 打✓(請複 選3~5項)
	(1)思考與創新	經由課程的訓練與引導設計，使學生能夠進行獨立性、批判性、系統性或整合性等面向的思考，或能以創意的角度來思考新事物。	✓
	(2)道德思辨與實踐	能夠對於社會、文化中相關的倫理或道德議題，進行明辨、慎思與反省，或能實踐在日常生活中。	
	(3)生命探索與生涯規劃	能夠主動探索自我的價值或生命的真諦，或能具體實踐在自我生涯的規劃或發展。	✓
	(4)公民素養與社會參與	能夠尊重民主與法治的精神、關心公共事務及議題，或能參與社會事務及議題的討論與決策。	
	(5)人文關懷與環境保育	能夠具備同理、關懷、尊重、惜福等人文素養，或能擴及到更為廣泛的環境及生態議題。	
	(6)溝通表達與團隊合作	能夠善用各種不同的表達方式進行有效的人際溝通，或能理解組織運作，與他人完成共同的事物或目標。	✓
	(7)國際視野與多元文化	能夠了解國際的情勢與脈動，具備廣博的世界觀，或能尊重或包容不同文化間的差異。	
	(8)美感與藝術欣賞	能夠領略各種知識、事物或領域中的美感內涵，或能據此促成具美感內涵之實踐力。	
	(9)問題分析與解決	能夠透過各種不同的方式發現問題，解析問題，或能進一步透過思考以有效解決問題。	✓
說明：課程符合指標內涵之部份內容，即可勾選。請依據課程內涵判定其符合程度，勾選項數以主要的3~5項為度。			

授課教師資料	姓名：鍾峰宜 ☒專任 學系(所，中心)：通識教育中心 職稱：助理研究員 ☐兼任 學系(所，中心)： 職稱： 學歷： 中正大學資訊管理研究所，資訊管理學博士 逢甲大學資訊工程研究所，工學碩士 經歷： 中正大學通識教育中心專任助研究員兼助理教授（現職） 中正大學圖書館讀者服務組組長（現職） 中正大學運動競技學系兼助理教授 研究領域： 圖書與資訊科學（Library & Information Science） 機器學習（Machine Learning）與商業智能（Business Intelligence） 大數據分析（Big Data Analytics）與資料科學（Data Science）
備註	請尊重智慧財產權，不得非法影印教師指定之科書籍！