

資料探勘

Ko-Wei Huang (黃科瑋)

elone.huang@kuas.edu.tw

這學期要上什麼？

會教大家 R 語言及



資料探勘與相關演算法



資料探勘 是什麼？

從大量資料中選取合適的資料，進行資料處理、轉換等工作後，再挖掘出有用資訊 (如關係或分組) 的過程

資料探勘 = 資料庫知識發掘 (Knowledge Discovery in Databases, 簡稱 KDD)

Example

年齡在十八歲到三十五歲之間的男生 -> 會購買不錯的顯卡

年齡在十八歲到三十五歲之間的女生 -> 會購買 iphone























R 語言 是什麼？

R 語言主要用於統計分析，統計圖表繪製以及資料相關分析演算法

R 語言內建多種統計及分析功能，也可安裝外部套件增加可用性

為什麼要學 R 語言？

IEEE SPECTRUM 2017 Top Programming Languages

Language Rank	Types	Spectrum Ranking
1. Python	 	100.0
2. C	  	99.7
3. Java	  	99.5
4. C++	  	97.1
5. C#	  	87.7
6. R		87.7
7. JavaScript	 	85.6
8. PHP		81.2
9. Go	 	75.1
10. Swift	 	73.7

你需要準備什麼 !?

安裝 R language software

統計 (會有一些數學邏輯)

筆電 (如果有的話...)

課程大綱

Part I. R Language

R 語言程式設計

R 語言繪圖

R 語言統計圖表

Part I. Data Mining

資料前處理

關聯規則探勘

資料分群

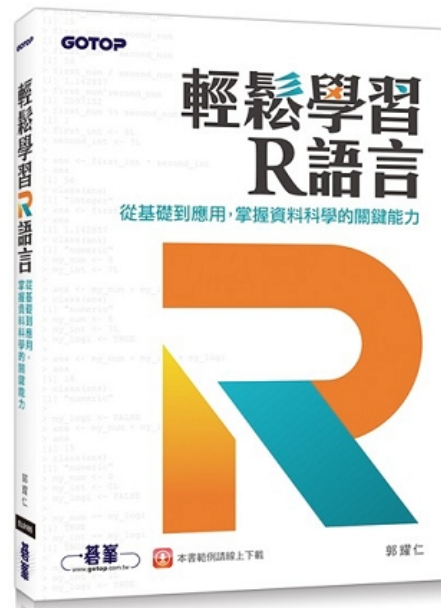
資料分類

Case Study

教科書

輕鬆學習 R 語言：從基礎到應用，掌握資料科學的關鍵能力

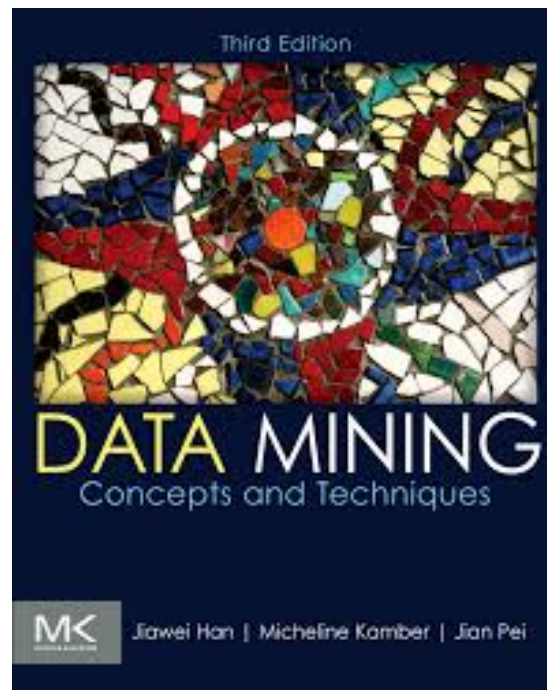
郭耀仁



教科書

Data Mining Concepts and Techniques 3rd

Jiawei Han et al.



成績計算方式

- 期中專題 : 25%
- 期中報告 : 25%
- 期末專題 : 25%
- 期中報告 : 25%

期中 / 期末報告

搜尋資料探勘相關期刊論文，整理成簡報檔並上台報告

期中報告時間：5/18

期末報告時間：6/29

Request：

1. 1-2 人一組
2. 至少是 SCI/SCIE Q3 以上

期中專題 (Brief ver. 4/27)

Input : 三篇英文新聞

Output :

1. 所有詞彙以及出現的字數統計
2. 算出 TF-IDF(term frequency-inverse document frequency)

並找出這三篇文章中重要的詞彙

Request :

1. 1-2 人一組
2. 使用 R 語言

期末專題 (Brief ver. 6/22)

Input：氣象局觀測資料（氣溫 / 雨量 / PM2.5 等）

Output：

1. 以現有的資料預測 2018-6/25 – 2018-6/29 的 PM2.5 數值
2. 找出氣象觀測資料中和 PM2.5 相關的規則（ex 下雨 -> PM2.5 低）

Request：

1. 1-2 人一組
2. 使用 R 語言

上課時間

每週五

9：10～12：00

如有需要我們可以調整上下課時間

課輔時間及聯絡方式

有問題的話隨時可以來我辦公室找我，我在 711

找不到人可以去 311 跟學長留言或寄信給我…

課程調整

因為 4 月 13 號我要出國出差，補課時間暫定為

日期	時間
3/9	12:00~12:30
3/16	12:00~12:30
3/23	12:00~12:30
3/30	12:00~12:30
4/13	12:00~12:30

Any Questions !?

有問題都可以提出來哦