商業智慧與巨量資料分析

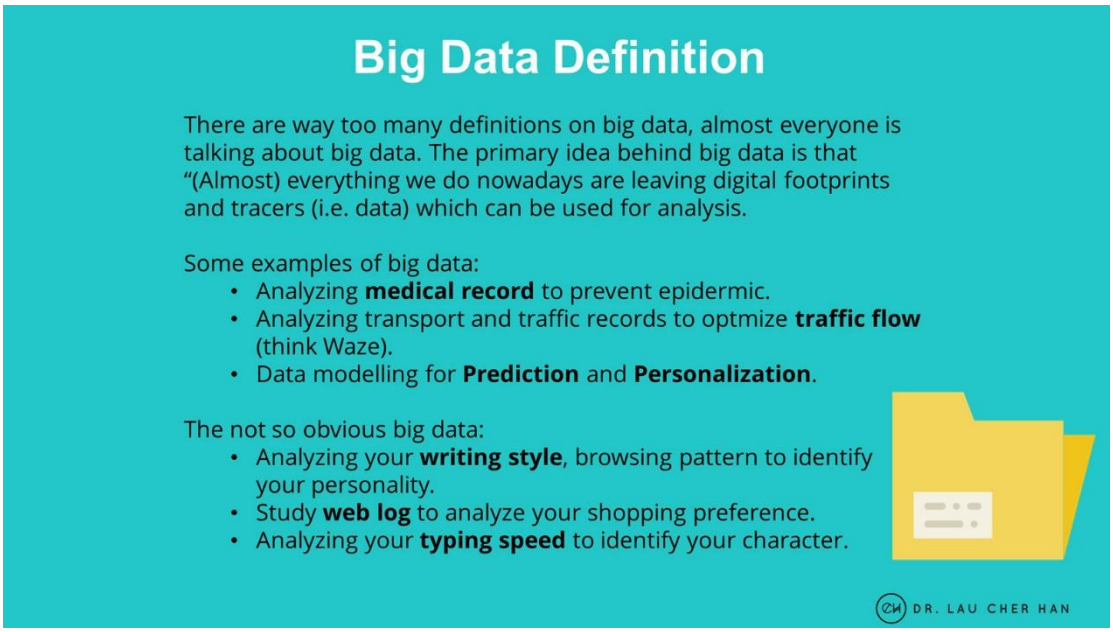
Business Intelligence and Big Data Analytics

(1)Big Data Analytics (Analysis): 大數據分析、巨量資料分析

Let's do "Big Data!" 你大數據了嗎?

- Big data definition

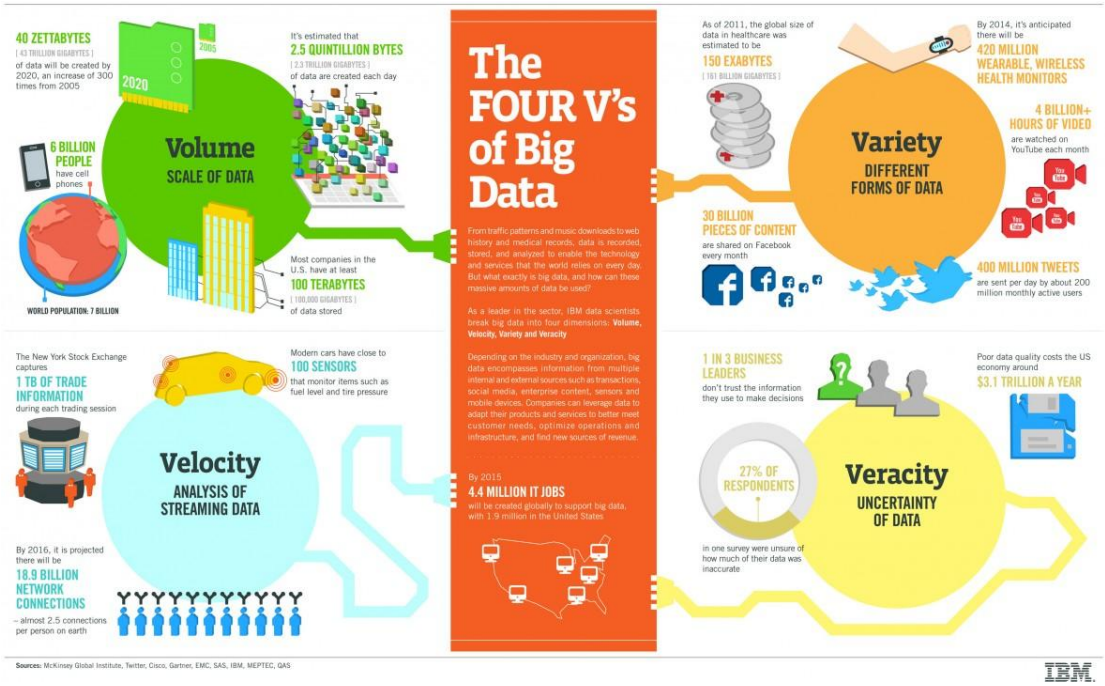
Big data analytics is the use of advanced analytic techniques against very large, diverse big data sets (龐大的大數據: 數字、文字、圖片、影片、聲音).



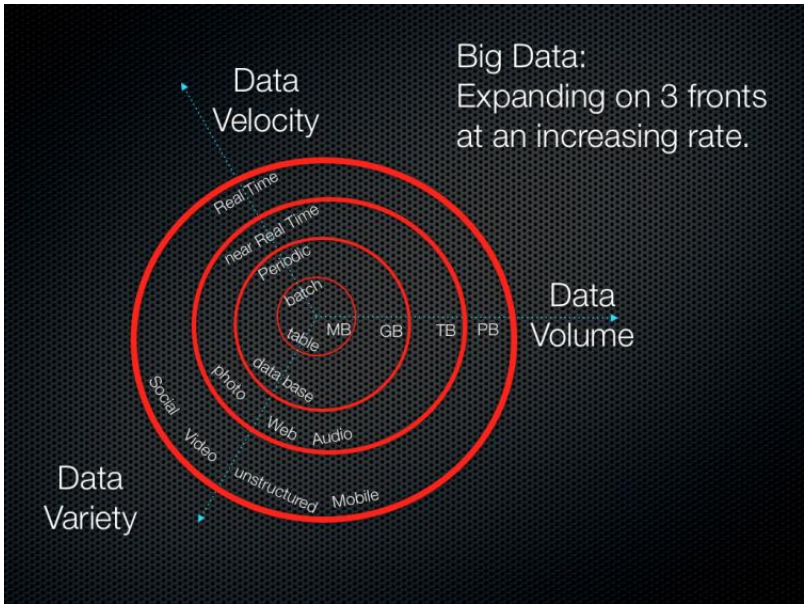
- Application domains

Some applications of big data in commerce include the following domains.

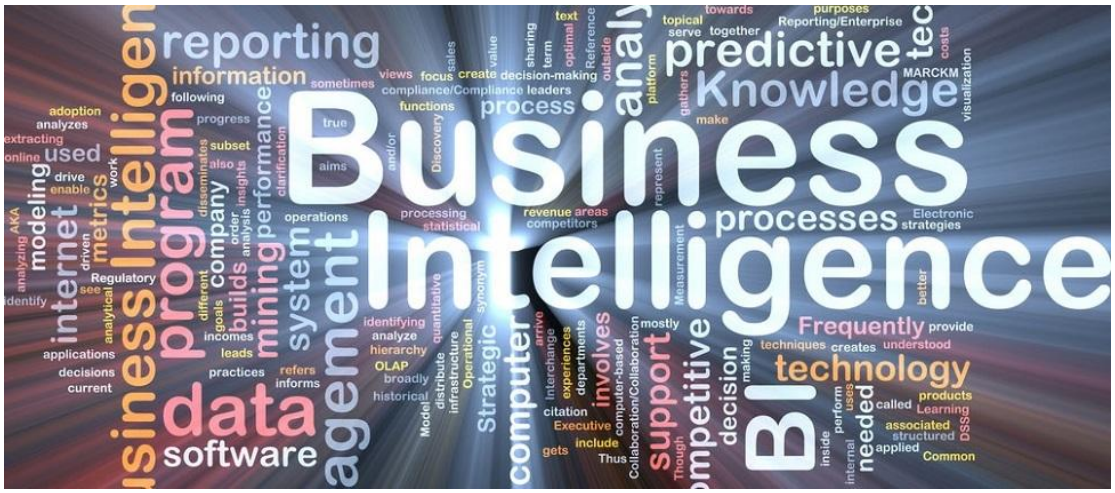
- The Four V's of Big Data:



The 3V's of Big Data

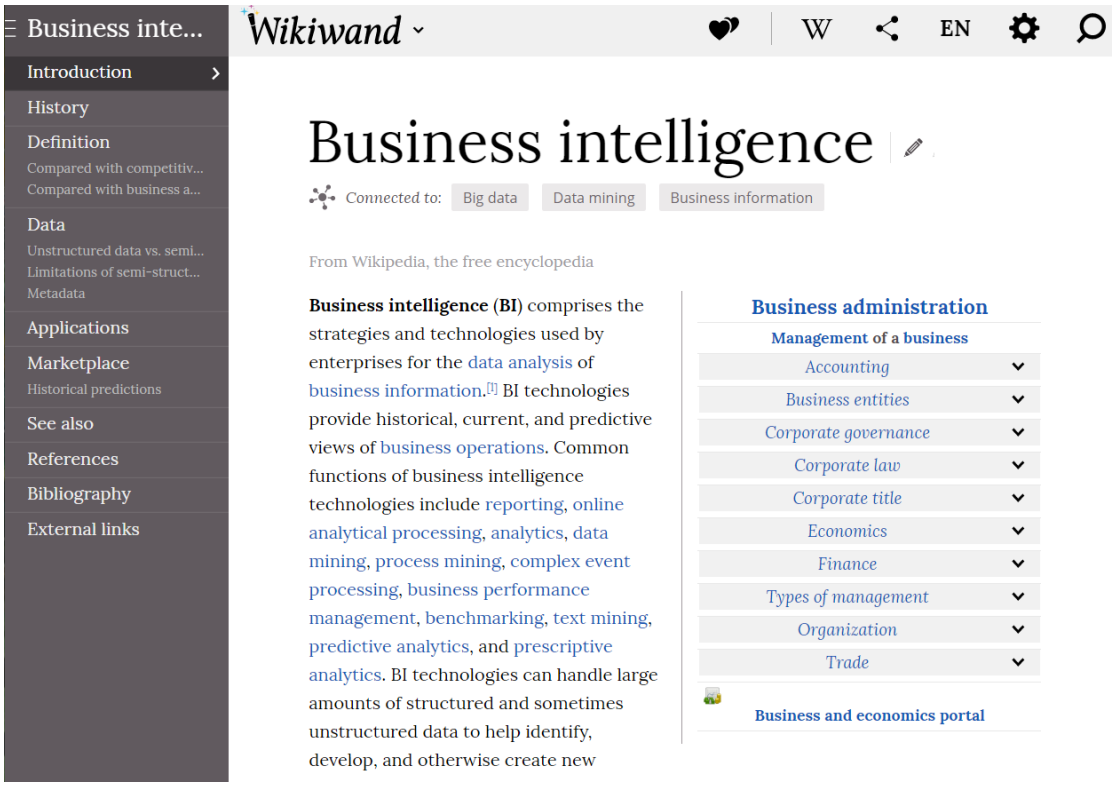


(2)Business Intelligence (BI)商業智慧、商業上的人工智慧(AI in business)



Source

Source



Business intelligence (BI) comprises the strategies and technologies used by enterprises for the data analysis of business information.

BI technologies provide historical, current, and predictive views of business operations. Common functions of business intelligence technologies include reporting, online analytical processing, analytics, data mining, process mining, complex event processing, business performance management, benchmarking, text mining(文字分析), predictive analytics, and prescriptive analytics.

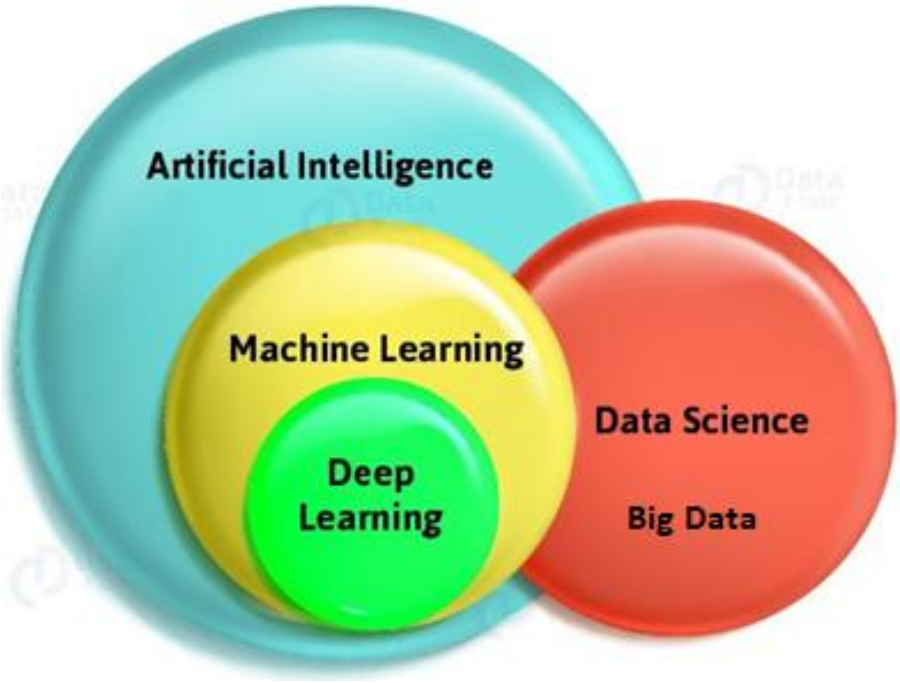
BI technologies can handle large amounts of structured and sometimes unstructured data to help identify, develop, and otherwise create new strategic business opportunities.

They aim to allow for the easy interpretation of these big data. Identifying new opportunities and implementing an effective strategy based on insights can provide businesses with a competitive market advantage and long-term stability.

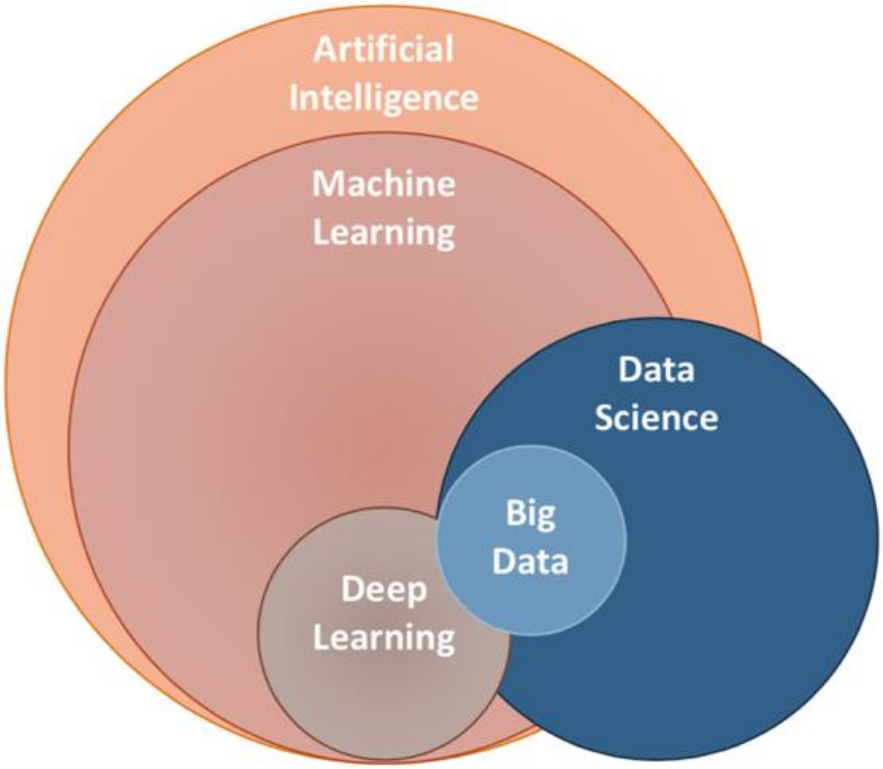
BI 商業上的人工智慧(AI in business)

BI is AI in business, so I can say that BI is a subset of AI.

Relationship among big data, AI, machine learning, and deep learning



來源



圖片來源

AI: 人工智慧
Machine learning: 機器學習
Deep learning: 深度學習 是一種進化版的 類神經網路
Big data: 大數據
Data Science: 數據科學 專門研究數據

- **Big Data and Machine Learning are complimentary (互補).**

Big data needs machine learning.
In order to have the ability of learning knowledge from data, big data needs the learning algorithms of the machine learning.

Machine learning needs big data
In order to be accurate, Machine Learning needs more data samples to learn, and we need the ability to handle the Big Data sets. We have to implement distributed computing using big data technologies like Apache Hadoop or Spark platform for machine learning algorithms to generate model.

Because AI has a more general (broad) definition, I will say that big data and machine learning should work together, and both are the subsets of artificial intelligence.

Why do we focus on Public Opinion Analysis 輿情大數據分析?

- Text data is around us. We can't escape from it.

Why we choose text? Because it is very import in our daily life. We read text from the internet via our smartphone and computer. The internet provides a lot of data and information in text format.

- We should focus in a specific application because we have limited time.

We have huge data such as numbers, photos, video, music, sound, text, and many others collected data from the internet, factory, and government.

The data from different areas should be processed using different techniques. We cannot study all of the techniques in limited time. So, we choose one of them-that is text 文字.

This semester we focus on the text data from the internet. Processing text data is also known as web mining or text mining.
Precisely speaking, in this course, we will learn text mining, deep learning, NLP, python, Keras, crawler, and so on.

After finishing this course, you can say that you have some experience in frontend and backend web techniques. You are able to add these new skills in your resume. Some jobs are related to these techniques: frond end and backend website engineer, data analysist.前端/後端工程師 數據分析師

What will we learn in this course?

You will familiar with these techniques:
你將會熟悉以下技術：

Python
Django website design (Python web framework)
jQuery, JavaScript (web)
HTML, Bootstrap4 (web)
NLP: Natural Language Processing (基礎自然語言處理)
Word embedding 詞嵌入
Deep Learning (MLP or CNN or BERT language model)

What project should we accomplish in class?

Public opinion analysis 輿情大數據分析系統

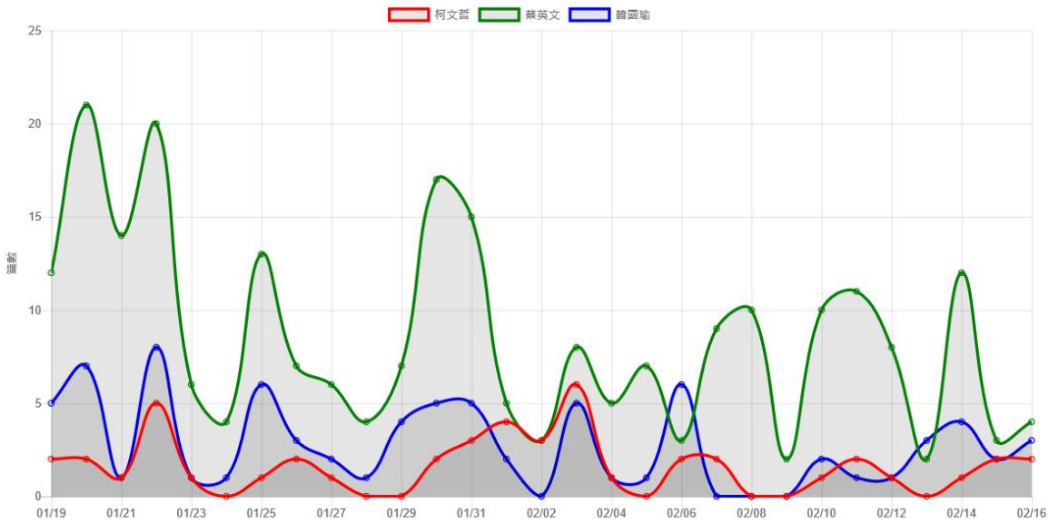
Here is the demonstration website:

http://163.18.23.20:8000/

總聲量變化比較

依據時間顯示總聲量的變化

在哪個時間點被新聞報導最多?



輿情大數據 | 政治人物聲量排行榜 | 政黨聲量排行榜 | 熱門關鍵詞分析 | 熱門人物排行分析 | NER熱門分析 | 你的關鍵詞熱門度分析 | 全文檢索與關聯新聞分析 | 你的關鍵詞情緒分析 | 最新新聞瀏覽與推薦 | 新聞或文章情緒分類 | 新聞分類新聞分類 | 會員獨享功能

政治人物或政黨聲量觀察

你關心的政治人物或政黨

排行榜(資料週期:資料截止時間的前4周)

網路聲量

網路聲量是什麼？ 計算新聞報導提及的文章則數，聲量越高表示能見度越高。

國民黨

總聲量:266篇

影響力:57% 19% 23%

民進黨

總聲量:223篇

影響力:63% 17% 19%

民眾黨

總聲量:63篇

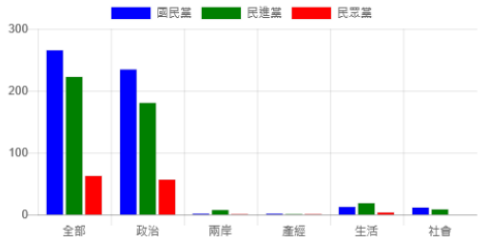
影響力:57% 19% 23%

« 1 2 3 »

總聲量分布情況

依據新聞類別統計網路聲量

在哪一類別的新聞中被報導最多篇?



全文檢索與你關心的關鍵詞關聯分析

對你想要了解的議題進行全文檢索，找出有哪些詞與你的關鍵詞一起出現?

輸入條件

關心哪個關鍵詞?: 蔡英文 韓國瑜

全文搜尋，可輸入多個關鍵詞或片段詞句，以空白隔開。

條件: ◉and ○or

新聞類別: ◉全部 ○政治 ○科技 ○運動 ○證卷 ○產經 ○娛樂 ○生活 ○國際 ○社會 ○文化 ○兩岸

最近多少周?: ○1 ◉2 ○3 ○4 ○6 ○8 ○12

以最新資料時間為準，往前推多少周?

查詢

這些詞與它同時出現喔!

以下新聞與它有關(只列出最多10篇，尚無頁次功能)

- 政治:鄭文燦：武漢肺炎對中共執政正當性是很大的挑戰
- 政治:傅崐萁恢復黨籍案 國民黨中常會通過
- 生活:早安世界》陸配子女入境政策反轉 陸委會道歉

所在的文章段落(最多顯示10段)

- 傅崐萁晚間發表聲明指出，他一直以來都是藍營的正藍戰將，不管是2004年、2008年、2012年的總統大選，甚至是最艱困的2016年及2020年總統選舉，他都是支持國民黨籍的總統候選人，這次2020大選，高雄市長韓國瑜得票更大贏總統蔡英文4.5萬票，這也表示他過去堅守在藍營及服務札根獲得花蓮鄉親的認同。

你關心的關鍵詞的情緒分析

可以了解媒體對該關鍵詞的情緒程度

輸入條件

關心哪個關鍵詞?　柯文哲

查找關鍵字，可輸入多個，空白隔開。主要以人名，產品，地理區域為主(搜尋斷詞後的詞語，並非全文搜尋)。

條件　◉and ○or

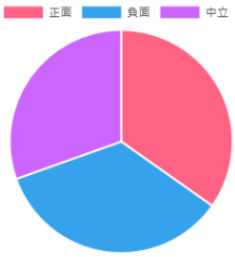
新聞類別　◉全部 ○政治 ○科技 ○運動 ○證卷 ○產經 ○娛樂 ○生活 ○國際 ○社會 ○文化 ○兩岸

最近多少周?　○1 ◉2 ○3 ○4 ○6 ○8 ○12

以最新資料時間為準，往前推多少周?

查詢

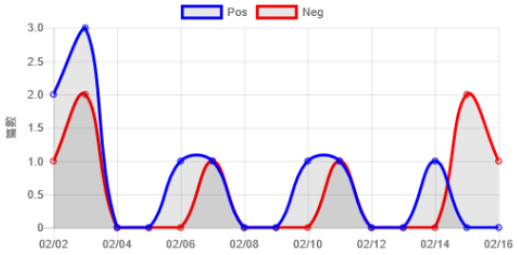
情緒分析:文章層級



pos: 8篇 neg: 8篇 obj: 7篇

正反面情緒變化

每個時間點的正反面報導的篇數(每天統計，6周含以上以5天為單位統計)



熱門人物

哪個人物被報導最多次?

篩選條件(資料週期:資料截止時間的前4周)

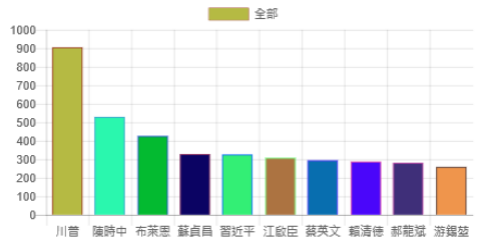
新聞類別　◉全部 ○政治 ○科技 ○運動 ○證卷 ○產經 ○娛樂 ○生活 ○國際 ○社會 ○文化 ○兩岸

新聞類別內定值為"全部"新聞

多少個熱門詞?　○5 ◉10 ○20 ○30 ○40 ○50 ○60

內定值為10

熱門關鍵字繪圖



熱門關鍵字(頻率次數)

- 川普,905
- 陳時中,529
- 布萊恩,426
- 蘇貞昌,328
- 習近平,326
- 江啟臣,308
- 蔡英文,295
- 賴清德,288
- 郝龍斌,281
- 游錫堃,259

What does a commercial public opinion look like?

熱文榜

1	2	3
棚拜 ◂ 86,769	富發牌 ◂ 66,640	蔡英文 T... ◂ 43,264
		
很久沒帶老婆出門吃飯了	春酒過後與粉絲的約會就來了大家不見不散楠哥哥來陪各位囉	確保防疫物資原料穩定，#台灣製造全力支援！台灣的產業在這次防疫物資的協助上，展現出台灣製...

MORE

1	2	3
NBA ◂ 2,397	老遊戲版 ◂ 2,272	八卦板 ◂ 2,107
		
[Live]2020NBA全明星賽	[LIVE]tvN愛的迫降最終回	Re:[問卦]你各位知道我現在很害怕恐慌嗎

MORE

1	2	3
文茜的... ◂ 27,143	天氣即... ◂ 15,453	三立新聞 ◂ 12,342
		
《封城之愛》武漢「封城」後，位於光谷的Wakanda咖啡店沒有歇業，7名員工自願留守，每...	今晚~週二(18日)晨，本次寒流最冷(17:40發佈)寒流影響即將進入第二天，隨著乾空氣...	高大成：「叫他回去祖國，有沒有得都他的事情。」（#金魚腦編）怕得病躲台灣？高大成狂嗆黃安...

MORE

How do we learn them (a lot of codes)?

Learning by doing

從做中學

Learning by examples

從實例中學

Learning step by step

逐步分解動作練習

中文教學目標(Chinese Teaching Objectives)

英文教學目標(English Teaching Objectives)

(1)培養學生熟悉巨量資料分析的基礎與進階理論與應用領域。

(2)培養學生熟悉巨量資料分析的程式開發工具與平台

(3)培養產業界所需之商業智慧與巨量資料分析專業人才

This course introduces an overview of theory and practice on business intelligence and big data analytics. In class, students will learn by doing. Students will be able to understand various applications and know how to use the modern platforms and development tools in big data analytics. This course can help learners to achieve the following objectives:

(1) To be familiar with the basic and advanced theory and application domains in big data analytics.
(2) To strengthen the ability of using big data analytics platform and programming tools.
(3) To develop the logical thinking and the skill of implementation of big data analytics and business intelligence in various application domains.

中文教學綱要(Chinese Course Description)

英文教學綱要(English Course Description)

- (I) Public opinion is the collective opinion of the people of a society or state on an issue. A public opinion analysis is an important application domain of big data analysis and artificial intelligence. This course will introduce how to implement a public opinion analysis system.
- (I) 本課程將介紹大數據分析的一個重要的應用場景—輿情分析(Public opinion analysis)。

- (II) A public opinion analysis system involves several technologies including: (1) Natural Language Processing (NLP)—abstract extraction, document similarity and recommendation, document classification and emotional classification (2) Machine learning and deep learning, and (3) Website design.
- (II) 課程會學習輿情分析用到的相關技術，包括機器學習與自然語言處理NLP，例如:文件摘要、文章相似度、文件推薦文件分類、情緒分類等技術。

- (III) In this course, you will learn how to use Python, Keras, and Django to implement a real world public opinion analysis system.
- (III) 使用的工具包含Python程式、Keras深度學習模型、Django網頁前端後端整合設計。

Syllabus

中文授課進度	English course content
(0) Python基礎: 常用資料結構 list []串列 (類似陣列) tuple () dict {}字典 set {}集合 數學運算 判斷 迴圈 自訂函數(方法) numpy套件 pandas套件	(0) Python introduction Useful data structure: list tuple dict set Arithmetic Selection Loop (for, while, for each) Define function Numpy package Pandas package
(I)自然語言處理NLP 1. 斷詞 2. 詞性分析 3. 詞頻統計 4. 全文檢索 5. 大型語言模型與詞嵌入 6. 文章相似度與新聞推薦系統	(I) Natural Language Processing (NLP) 1. Word segmentation 2. Part of speech analysis 3. Keyword frequency 4. Full text search 5. Large language model and Word embedding 6. Document similarity and recommendation
(II)網頁技術: 1. 網頁爬蟲 2. 前端網頁技術 3. Django後端網頁技術	(II) Websites design 1. Web page Crawler 2. Front end technology 3. Back end technology 4. Ajax

(III)機器學習與深度學習: 1. 深度學習模型訓練與測試 2. 新聞情緒分類 3. 新聞分類	(III) Machine Learning and Deep Learning 1. Deep learning 2. Sentiment classification 3. Document classification
(IV)實作專題 1. 期中專題製作 2. 期末題製作	(IV) Projects 1. Midterm Project 2. Term Project

1. 課程簡介與環境安裝	1. Course Introduction and package installation
2. Python簡介與新聞爬取	2. Python Introduction and News Crawl
3. 文字斷詞與詞頻文字雲	3. Word tokenize, word frequency and word could
4. 文章情緒分析	4. Document sentiment analysis
5. 新聞爬蟲使用BS4	5. Crawl news content using beautiful soup4
6. Django 網頁設計簡介與開發環境	6. Django website design and VS Code IDE
7. 新聞關鍵字分析與網頁建立	7. News keywords analysis on Django website
8. 中研院繁體中文斷詞與熱門人物分析	8. Ckiplab tokenization and top person analysis
9. 期中專案	9. Midterm project
10. 使用者自訂關鍵詞分析	10. User defined keyword analysis
11. 自訂關鍵詞關聯分析	11. User defined keyword association analysis
12. 自訂關鍵詞情緒分析	12. User defined keyword sentiment analysis
13. 相關新聞推薦系統-使用詞嵌入模型	13. News recommendation using word embedding model
14. 深度學習簡介	14. Introduction to deep learning
15. 新聞文本情緒分類: 使用深度學習	15. Document sentiment analysis using CNN deep learning model
16. 新聞文本類別分類: 使用深度學習	16. News document classification using CNN deep learning model
17. 輿情分析系統入口網站設計	17. Public opinion analysis website portal
18. 期末專案	18. Term project

Scoring/Grading 評分規定

1. (30%) Class exercises and home works 課堂操作 (每一節課都有任務須完成) 與 作業
2. (35%) Midterm project 期中專案 public opinion analysis website (I)線上新聞文字大數據分析網頁專案(Basic)
3. (35%) Term project 期末專案: public opinion analysis website (II)線上新聞文字大數據分析網頁專案(Complete)

You may do the project alone. Or you may group a [redacted]-person team to finish it. But I encourage you to finish the project alone.

1[redacted]人一起完成期中、期末專題。

Course Requirements / Rules

Attend class 出席上課
Concentrate in class, and do not use cellphone.專心聽講，不要滑手機
Do tutorial exercise in class.跟著老師完成課堂練習
Questions and comments are welcome.歡迎多問問題，多發表意見

Frequently asked questions

You may also have these questions.

常見的問題，你可能也會有這些疑問?

1. 這門課程是英文授課，我的英文不好，可以修這門課嗎?

可以喔。

因為課堂上使用的英文句型不複雜，大多可以聽得懂，比較需要適應的是專有名詞，老師第一次提到這些專有名詞時，，同時也會提及專有名詞對應的中文，，所以，你應該可以應付得來。

別擔心，大多數的同學英文都不盡理想，多給自己機會學習，英文可以慢慢變好。

更何況，課程內容大多為程式碼，程式碼不需要太多的講解，就算有些聽不懂，親自執行試一試，你就能理解了。

1. This course is taught in English. My English is not good. Can I take this course?

Okay.

Because the English sentences used in the classroom are not complicated, you can understand most of them. Only some technical terms (terminologies)(技術名詞、專有名詞) are difficult for us. But don't worry. When these terms are first mentioned, the teacher will also speak it in the corresponding Chinese. So, you should be able to understand these technical terms.

Most of us are not good at English. If we learn more, we will gradually improve our English ability.

What's more, the course content is mostly about programming code. The programming code does not require much explanation. Even if you don't understand what teacher says, you can also understand it by running the code by yourself.

2. 課堂上問問題一定得說英文嗎?

課堂上我們說英文，比手畫腳也 OK，別擔心。下課後，找老師討論，我們盡量說英文，不得已中英夾雜也可以。

2. Do I have to speak English in class?

In class, try to speak English as much as possible. However, if you don't know how to speak some words or sentences in English, in that case, you can speak Chinese. After class, you may speak Chinese.

3. 我沒有寫過 Python 程式，我適合修這門課嗎?

不建議。本課程適合有 Python 基礎者修課，如果你 Python 程式的基礎，這門課內容都是 Python 程式，那麼這門課對你而言，就太艱難了，你會挫折感很重的。如果你有 Python 程式的基礎，那就太好了，你一定可以學得很有心得。

3. I didn't write any Python program before. Is this course suitable for me?

It is Not suitable. If you have no Python foundation at all, this course will be too difficult for you and you may be frustrated.
If you have Python experiences, that's great. You can learn it well.

4. 我只有程式基礎，但是沒有寫過網頁程式，我適合修這門課嗎?

不建議。
希望有寫過網頁的經驗者來修課比較好。我們用到不少網頁技術，尤其大多是動態網頁 JavaScript，比較複雜些，需要有基礎才較能上手。

4. Although I have basic programming skill, I didn't write any web program. Is this course suitable for me?

It is Not suitable. If you are familiar with web programming, it will be easier to do mid-term and term projects for you. If you are not familiar with web programming, you may be frustraed, because we use a lot of webpage technologies, including JavaScript.

5. 這門課有期中與期末專案，我沒有把握，我無法完成怎麼辦?

只要將老師課堂展示的程式碼搞懂，每次上課都能掌握所有細節，期中期末專案，只要將資料集換成別的資料集，或是增加一些額外的分析展示功能，就可以了。

你若是很擔心負擔太重，你可以找一個技術比你好的同學組隊，兩人一組，一起分工合作，互相學習，這樣可以減輕負擔。

5. This course has mid-term and final term projects. What can I do if I can't complete it?

As long as you understand the tutorial code in the classroom, you can understand all the details for finishing the mid-term and term projects. By replacing with a different data set and doing some additional analysis, you can complete the project. If you worry about heavy loading, you may group a two-person team to work together.

6. 寫程式時遇到問題怎麼辦?

先問 Google, StackOverflow，ChatGPT 通常它可以給你很好的答案。
這門課也有一位很有經驗的學長擔任教學助教，他可以協助你。
老師每周也有 4 小時的解惑時間，歡迎你來預約討論。

6. What should I do if I have questions on how to write programs?

Ask Google, StackOverflow first, it usually gives you good answers.

There is a teaching assistant can help you. Additionally, the teacher also has 4 hours of discussion time each week. You are welcome to make an appointment for discussion.

7. 這門課適合怎樣的人選修?

想要學到比較整合性的、困難度較高的資訊技術的人
未來想要走技術路線的人
想要磨練一下自己寫程式能力的人
想要做一個數據(資料)分析師的人

對文字分析工作有興趣的人
對人工智慧在文字上的應用有興趣的人
想要做動態網頁的人
對網頁後端有興趣的人
想要應用 Python 程式的人
想要開發一個實用的系統的人
…

7. Who is eligible for this course?

Students who want to learn an integrated information technology,
Students who want to get IT jobs in the future,
Students who want to improve their programming skills,
Students who want to be a data (data) analyst,
Students who are interested in text analysis,
Students who are interested in the application of artificial intelligence in text,
Students who want to design dynamic web pages,
Students who are interested in design the backend of websites,
Students who want to master Python programs,
Students who want to develop a practical system,
…

8. 我是管理組，不是技術組的學生，適合修這門課嗎?

建議你必須有基礎的程式能力，其餘的好好跟著老師課堂上課，回家親自寫一次，應可以應付。

我見過不少程式也寫得不錯的管理組的學生，別太小看自己，你若對寫程式有一些基礎，想要開發自己，趁研究所這兩年，就是你最佳的學習時光。

8. I am a management group student, not a technical group student. Am I suitable for this course?

This course is suitable for you, if you have basic programming skills. Follow the teacher in class, write codes at home, and you can handle it.

I've seen a lot of students from the management group who also are good at programming. Do not underestimate yourself. The two years in NKUST is the best time for you to learn programming and develop yourself.

9. 修這門課需要怎樣的電腦配備?

基本的電腦就可以，不過，我們的開發環境比較大型，因此，RAM 最好有 8GB 以上。

9. What kind of computer equipment is required in this course?

A basic desktop or laptop computer is sufficient. Our development environment is a little bit large, so it is better to have more than 8GB of RAM in your computer.

10. 修這門課一定需要自己的筆電?

可以不用。

我們電腦教室的配備很高檔，使用教室的 PC 即可。課堂練習的程式碼，你可以拷貝到雲端備份。當然，你也可以攜帶你的筆電到課堂上來使用，不過若是你的筆電速度太慢，可能會趕不上課堂進度，建議你還是使用電腦教室的電腦比較適宜。

10. Do I need my own laptop for this course?

It’s not mandatory.

You are able to use the powerful PC in our computer classroom. You can back up your codes to the Google cloud.

Of course, you may use your own laptop (notebook) in the classroom.

How to contact the teaching assistant?

To be announced.

How to contact me?

黃承龍 國立高雄科技大學 資管系教授

Cheng-Lung Huang, Professor, Department of Information Management

E-Mail: clhuang@nkust.edu.tw

Office: C304(管理學院) (第一校區)

TEL: +886-7-6011000 Ext 34127

FAX: +886-7-6011042

高雄市 811 楠梓區卓越路二號

National Kaohsiung University of Science and Technology

2, Juoyue Rd., Nantz District, Kaohsiung 811, Taiwan.