國立高雄科技大學電腦與通訊工程系
資料探勘與最佳化實驗室

巨量資料分析應用與實作

(Big Data Analytics in Practice)

Syllabus

授課教師：江傳文

Instructor Information

- Instructor: Chuan-Wen Chiang (江傳文)
- Office: B429
- Phone: (07) 6011000 # 32012
- E-mail: ccw@nkust.edu.tw
- Office Hours:
  - Tue. 16:30 ~ 18:30
  - Thur. 10:00 ~ 12:00
  - **Other times by appointment**

Outline

- Instructor Information
- Synopsis
- Course Topics
- Textbook & References
- Grading Policies

Synopsis (1/4)

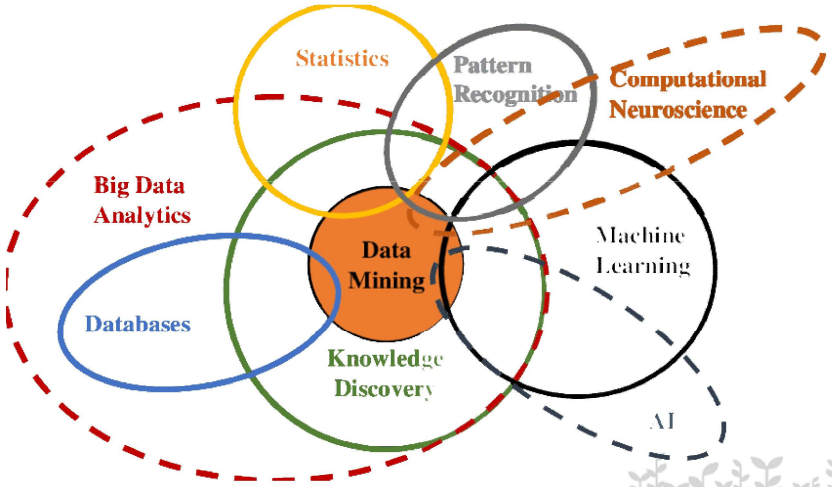
- What is big data?
  - Big data is a collection of very huge data sets with a great diversity of types so that it becomes difficult to process by using state-of-the-art data processing approaches or traditional data processing platforms.
  - Big data arises with many challenges, such as difficulties in data capture, data storage, data analysis and data visualization.

Synopsis (2/4)

- What is big data analytics?
  - Big data analytics refers to the process of collecting, organizing and analyzing large sets of data ("big data") to discover patterns and other useful information.
  - Big data analytics will help organizations to better understand the information contained within the data and will also help identify the data that is most important to the business and future business decisions.
  - Big data analysts basically want the knowledge that comes from analyzing the data.

Synopsis (3/4)

- Multidisciplinary nature of big data analytics



Synopsis (4/4)

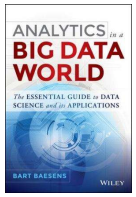
- The objective of this course is to provide students with a comprehensive introduction about techniques and tools required for data science and big data analytics. Students who successfully complete this course will be able to:
  - understand the advanced techniques in data science and big data analytics; and
  - know how to design, implement and troubleshoot massive data analytics algorithms using existing parallel program environment.

Course Topics

- Big Data: Opportunities and Challenges
- Overview of Data Mining
- Exploring Data
- Classification
- Cluster Analysis
- Association Analysis
- Case Study: Recommendation Systems
- Dimensionality Reduction
- Programming in Hadoop (MapReduce) and Spark

Textbook & References

- Textbook:
  - Bart Baesens, *Analytics in a Big Data World: The Essential Guide to Data Science and Its Applications*, John Wiley & Sons, 2014.



- References:
  - Journal Papers
  - J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of Massive Datasets (2e)*, Cambridge University Press, 2014.
  - P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Addison Wesley, 2005.

Grading Policies (1/3)

- Homework assignments: 20%
- Midterm examination: 35%
- Final examination: 35%
- Attendance and class participation: 10%

- **額外加分：**
  以巨量資料分析應用為主題參加校園創意發想提案競賽，
  - **晉級決賽：加學期總成績 3 分**
  - **榮獲優等：加學期總成績 5 分**
  - **榮獲特優：加學期總成績 8 分**

Grading Policies (2/3)

- Academic dishonesty will not be tolerated. Any student caught cheating will be dropped from the course.
- Browsing the Internet or working on any other computer project is prohibited.
- Mobile phones must be turned off while the student is in class.

Grading Policies (3/3)

- All students are strongly encouraged to attend every class. Students arriving late will be counted as being absent.
- Students who miss an examination/quiz without prior approval from the instructor will receive a grade of zero.
- Homework assignments will be due at the beginning of class. Late submission of homework will be accepted later on the same day and will result in 50% reduction of grade.

References (1/3)

1. C.L. Philip Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," *Information Sciences*, vol. 275, pp. 314–347, August 2014.
2. M. Bilal, L.O. Oyedele, J. Qadir, K. Munir, S.O. Ajayi, O.O. Akinade, H.A. Owolabi, H.A. Alaka, and M. Pasha, "Big Data in the construction industry: A review of present status, opportunities, and future trends," *Advanced Engineering Informatics*, vol. 30, no. 3, pp. 500–521, 2016.

References (2/3)

❑ 跨域思考專題報告成果節錄：(1/2)

References (3/3)

❑ 跨域思考專題報告成果節錄：(2/2)

國立高雄科技大學電腦與通訊工程系
資料探勘與最佳化實驗室

Welcome to the class!

授課教師：江傳文