



# Data Mining

## 資料探勘

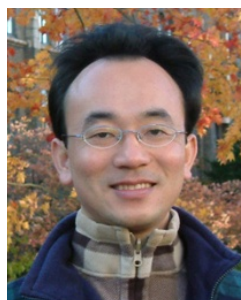
## Introduction to Data Mining

### (資料探勘導論)

1022DM01

MI4

Wed, 6,7 (13:10-15:00) (B216)



Min-Yuh Day

戴敏育

Assistant Professor

專任助理教授

Dept. of Information Management, Tamkang University

淡江大學 資訊管理學系

<http://mail.tku.edu.tw/myday/>

2014-02-19



# 淡江大學102學年度第2學期 課程教學計畫表

**Spring 2014 (2014.02 - 2014.06)**

- 課程名稱：資料探勘 (Data Mining)
- 授課教師：戴敏育 (Min-Yuh Day)
- 開課系級：資管四P (TLMXB4P)
- 開課資料：選修 單學期 2 學分 (2 Credits, Elective)
- 上課時間：週三 6,7 (Wed 13:10-15:00)
- 上課教室：B216

# 課程簡介

- 本課程介紹資料探勘 (Data Mining) 的基礎概念及應用技術。
- 課程內容包括
  - 資料探勘導論
  - 關連分析
  - 分類與預測
  - 分群分析
  - 文字探勘與網頁探勘
  - 海量資料分析
  - SAS企業資料採礦實務與認證 (SAS EM)
  - 資料探勘個案分析與實作

# Course Introduction

- This course introduces the **fundamental concepts** and **applications technology** of **data mining**.
- Topics include
  - Introduction to Data Mining
  - Association Analysis
  - Classification and Prediction
  - Cluster Analysis
  - Text and Web Mining
  - Big Data Analytics
  - **Data Mining Using SAS Enterprise Miner (SAS EM)**
  - **Case Study and Implementation of Data Mining**

# 課程目標 (Objective)

- 瞭解及應用 資料探勘基本概念與技術。
- Understand and apply the fundamental concepts and technology of data mining

# 課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
1	103/02/19	資料探勘導論 (Introduction to Data Mining)
2	103/02/26	關連分析 (Association Analysis)
3	103/03/05	分類與預測 (Classification and Prediction)
4	103/03/12	分群分析 (Cluster Analysis)
5	103/03/19	個案分析與實作一 (SAS EM 分群分析) : Case Study 1 (Cluster Analysis – K-Means using SAS EM)
6	103/03/26	個案分析與實作二 (SAS EM 關連分析) : Case Study 2 (Association Analysis using SAS EM)
7	103/04/02	教學行政觀摩日 (Off-campus study)
8	103/04/09	個案分析與實作三 (SAS EM 決策樹、模型評估) : Case Study 3 (Decision Tree, Model Evaluation using SAS EM)

# 課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
9	103/04/16	期中報告 (Midterm Project Presentation)
10	103/04/23	期中考試週 (Midterm Exam)
11	103/04/30	個案分析與實作四 (SAS EM 迴歸分析、類神經網路) : Case Study 4 (Regression Analysis, Artificial Neural Network using SAS EM)
12	103/05/07	文字探勘與網頁探勘 (Text and Web Mining)
13	103/05/14	海量資料分析 (Big Data Analytics)
14	103/05/21	期末報告 (Final Project Presentation)
15	103/05/28	畢業考試週 (Final Exam)

# 教學方法與評量方法

- 教學方法
  - 講述、討論、實作
- 評量方法
  - 實作、報告、上課表現



# 教材課本

- 講義 (Slides)
- 參考書籍
  - Applied Analytics Using SAS Enterprise Mining, Jim Georges, Jeff Thompson and Chip Wells, 2010, SAS
  - Decision Support and Business Intelligence Systems, Ninth Edition, Efraim Turban, Ramesh Sharda, Dursun Delen, 2011, Pearson
  - 決策支援與企業智慧系統，九版，Efraim Turban 等著，李昇暉審定，2011，華泰

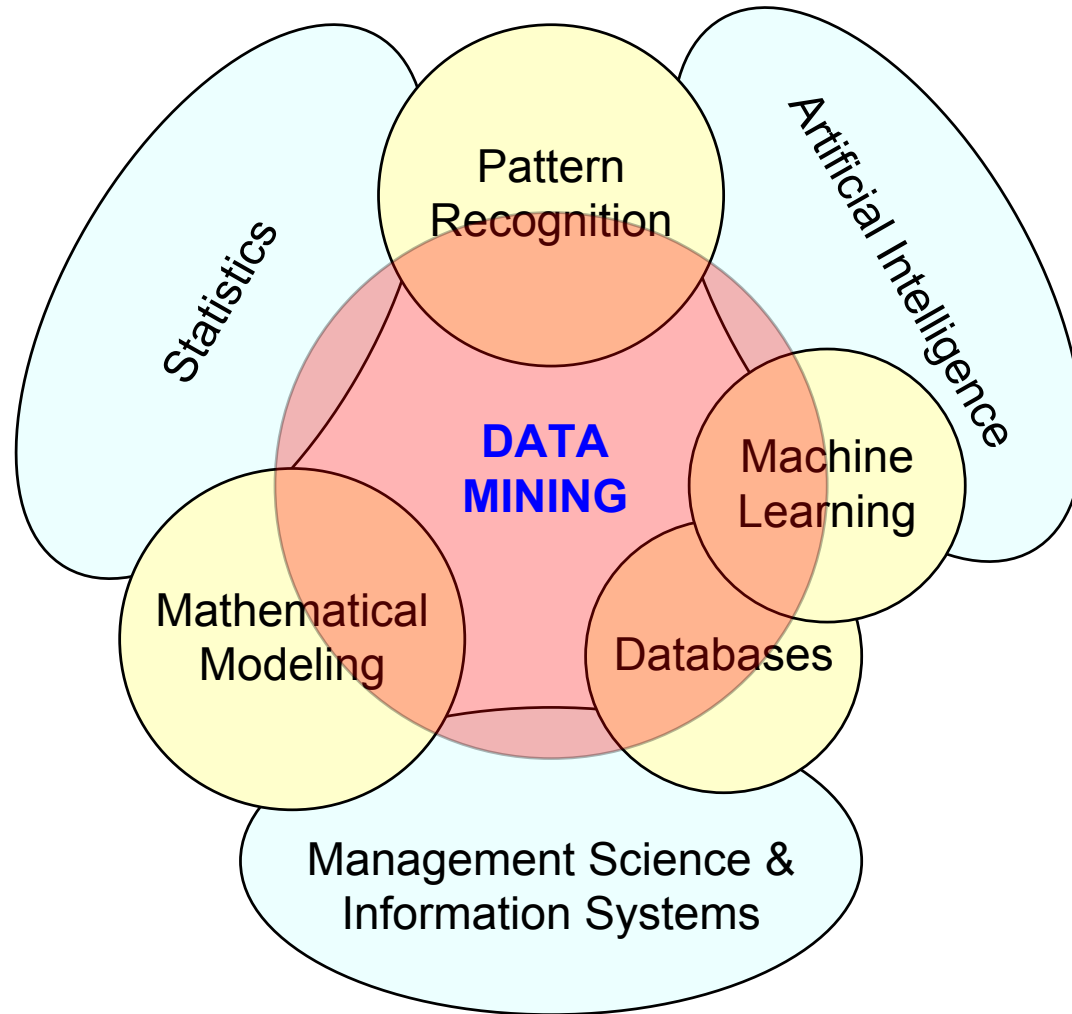
# 作業與學期成績計算方式

- 作業篇數
  - 3篇
- 學期成績計算方式
  - 期中評量：30 %
  - 期末評量：30 %
  - 其他（課堂參與及報告討論表現）：40 %

# Team Term Project

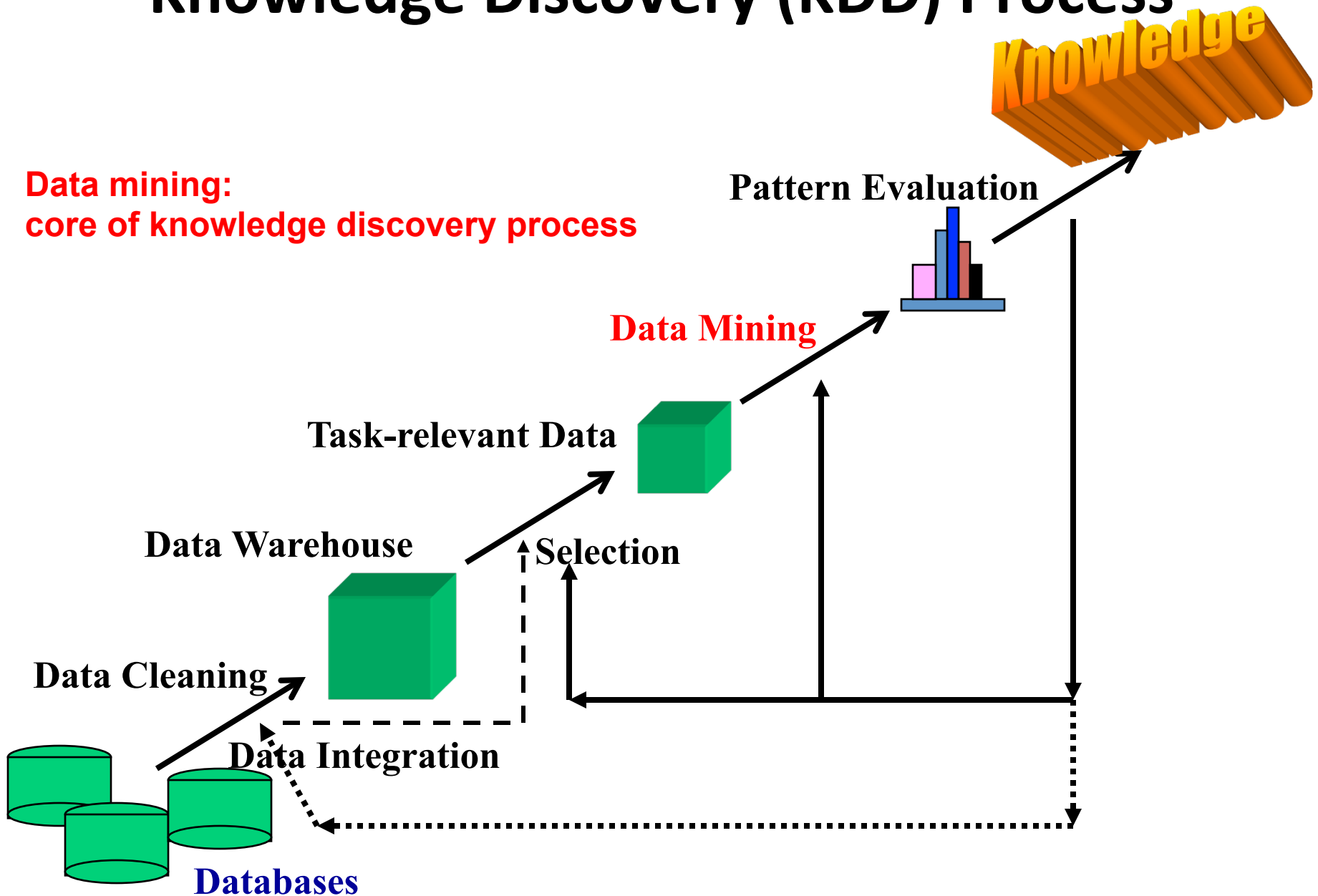
- Term Project Topics
  - Data mining
  - Web mining
  - Business Intelligence
- 3-4 人為一組
  - 分組名單於 2014/03/05 (三) 課程下課時繳交
  - 由班代統一收集協調分組名單

# Data Mining at the Intersection of Many Disciplines



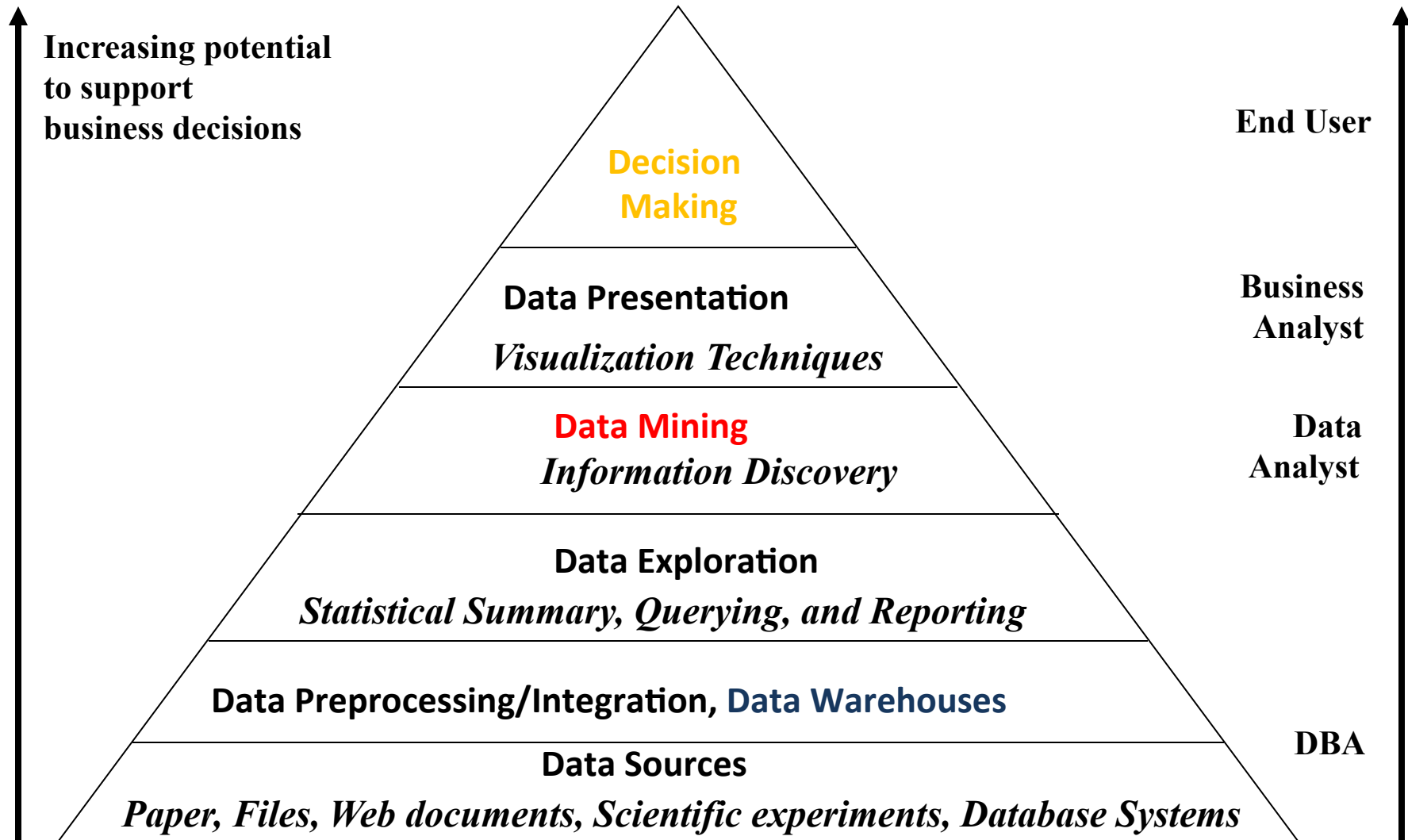
# Knowledge Discovery (KDD) Process

**Data mining:**  
core of knowledge discovery process

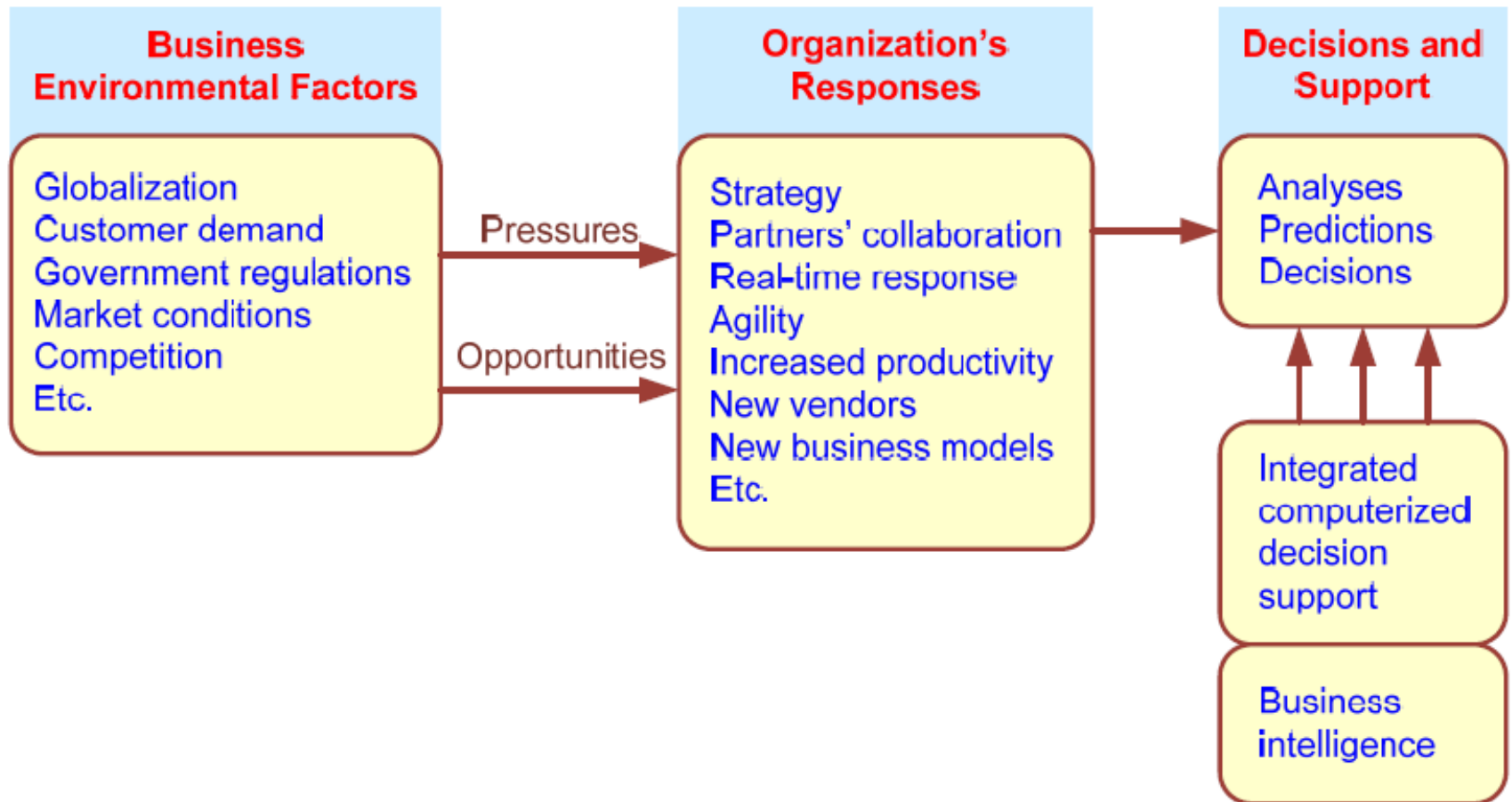


# Data Warehouse

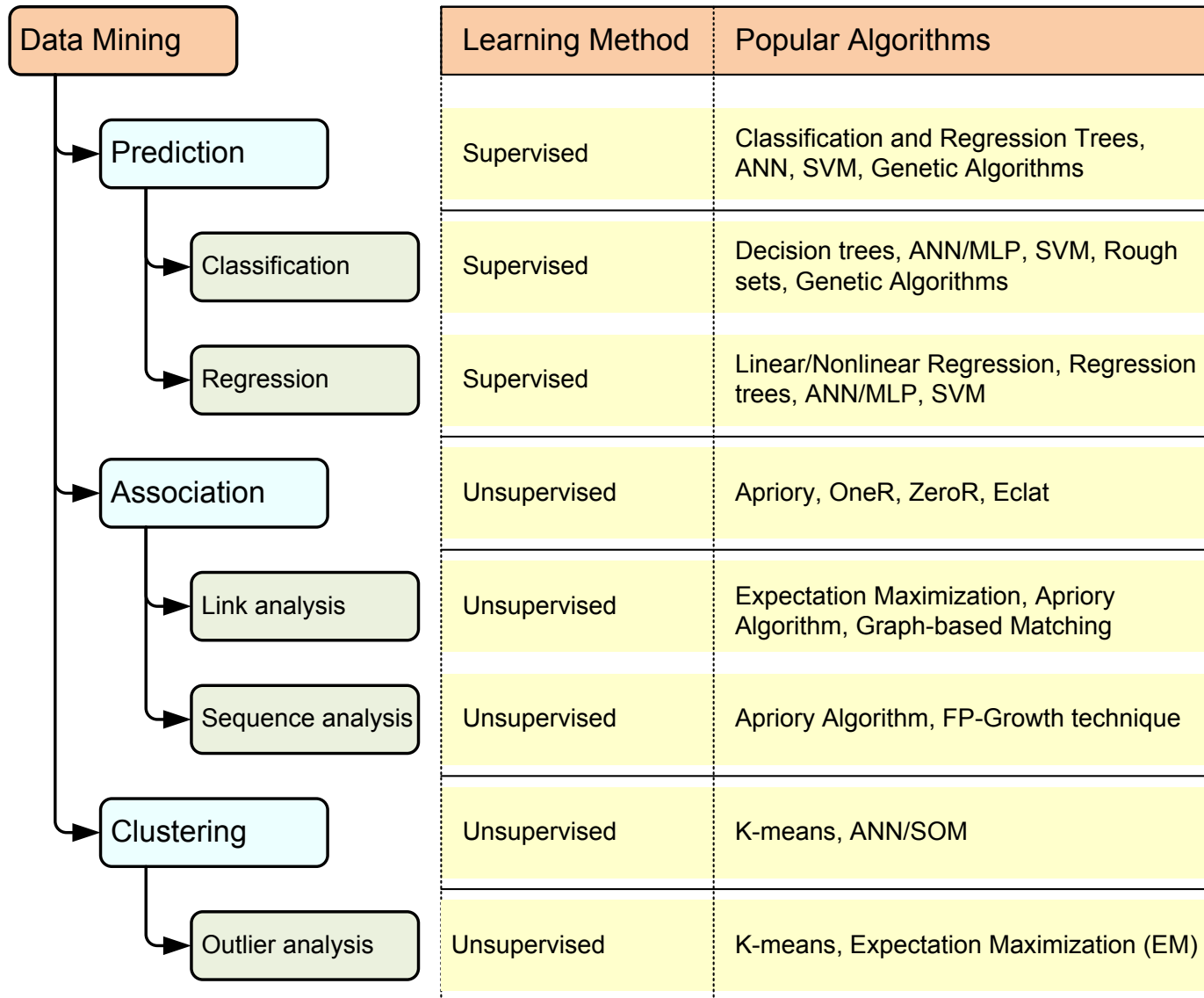
## Data Mining and Business Intelligence



# Business Pressures–Responses– Support Model



# A Taxonomy for Data Mining Tasks

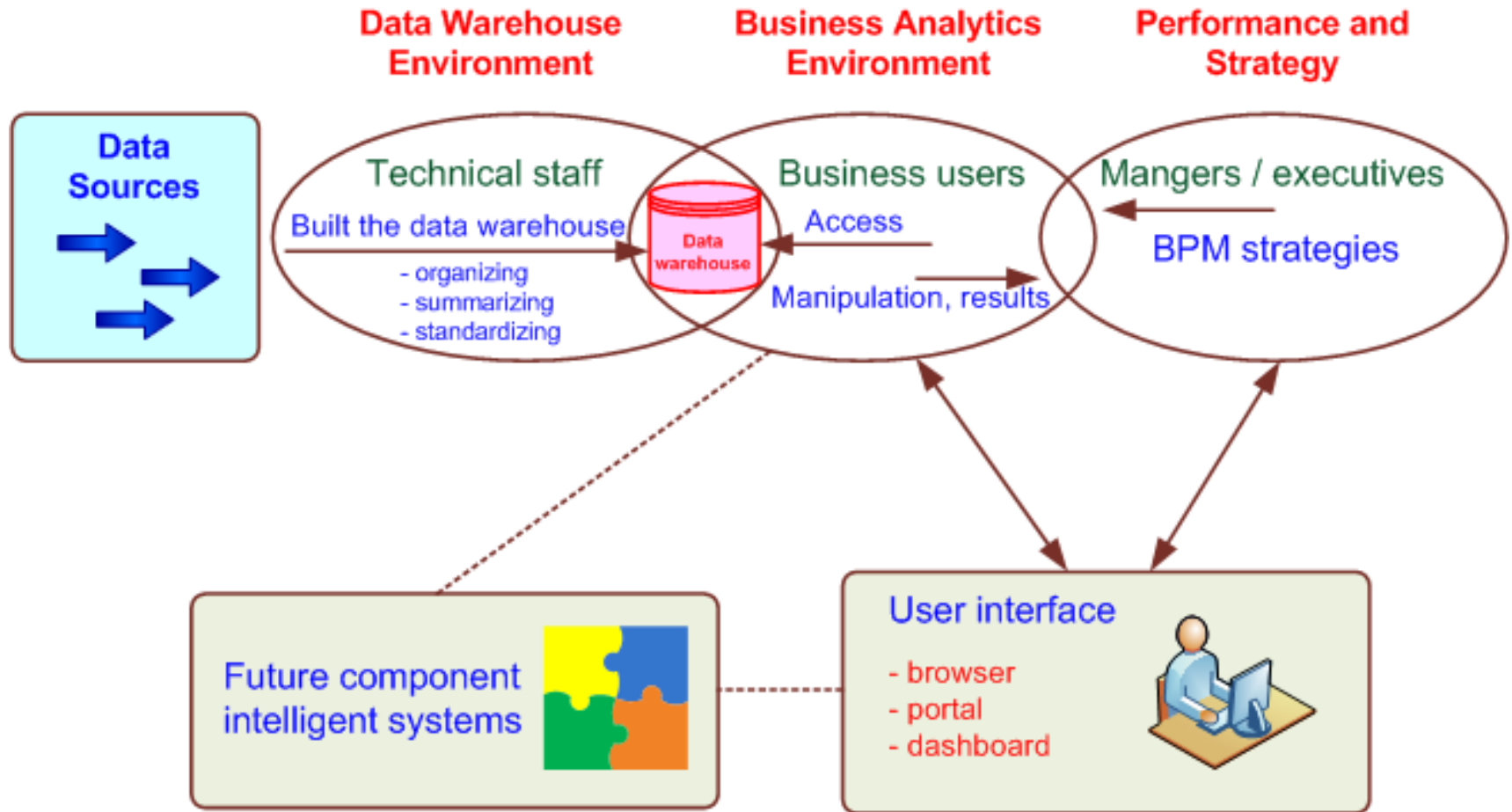




# The Evolution of BI Capabilities



# A High-Level Architecture of BI



# Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites

*Analyzing Data from Facebook, Twitter, LinkedIn,  
and Other Social Media Sites*



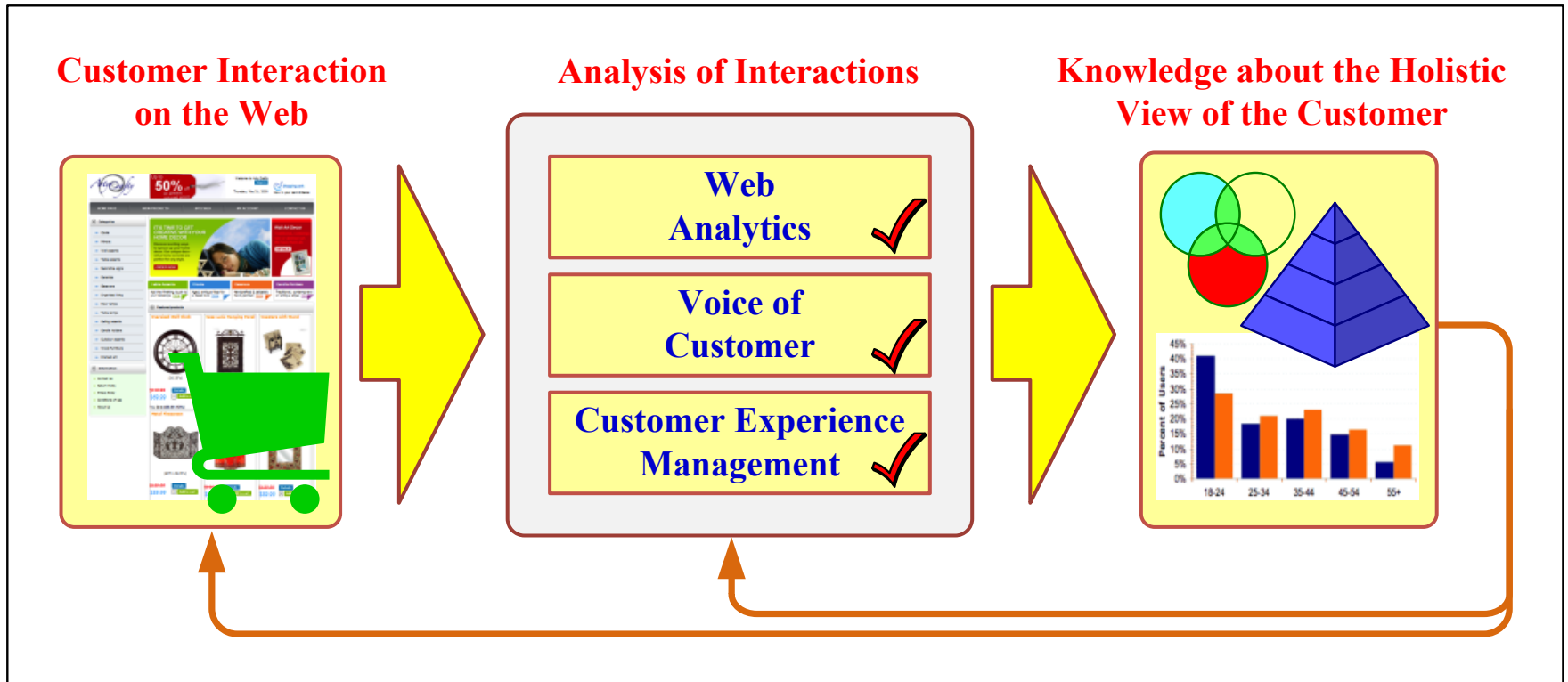
Mining the  
Social Web

O'REILLY®

*Matthew A. Russell*

# Web Mining Success Stories

- Amazon.com, Ask.com, Scholastic.com, ...
- Website Optimization Ecosystem



# Business Intelligence Trends

1. **Agile** Information Management (IM)
2. **Cloud** Business Intelligence (BI)
3. **Mobile** Business Intelligence (BI)
4. **Analytics**
5. **Big Data**

# Business Intelligence Trends: Computing and Service

- Cloud Computing and Service
- Mobile Computing and Service
- Social Computing and Service

# Business Intelligence and Analytics

- Business Intelligence 2.0 (BI 2.0)
  - Web Intelligence
  - Web Analytics
  - Web 2.0
  - Social Networking and Microblogging sites
- Data Trends
  - Big Data
- Platform Technology Trends
  - Cloud computing platform

# Business Intelligence and Analytics: Research Directions

## 1. Big Data Analytics

- Data analytics using Hadoop / MapReduce framework

## 2. Text Analytics

- From Information Extraction to Question Answering
- From Sentiment Analysis to Opinion Mining

## 3. Network Analysis

- Link mining
- Community Detection
- Social Recommendation



# Data Scientist:

## *The Sexiest Job of the 21st Century*

**Meet the people who  
can coax treasure out of  
messy, unstructured data.**

*by Thomas H. Davenport  
and D.J. Patil*

**W**hen Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

# Big Data 2014

活動辦法 | 第一屆東森盃Big x

www.ettoday.net/events/bigdata2014/rules.php

## 第一屆東森盃 BigData 校\*園\*爭\*霸\*戰

### 引爆資料科學家前哨戰

過關斬將 不戰不快  
火速組隊報名 成為最搶手巨量分析菁英!

最新消息 | 回首頁/關於活動 | 活動辦法 | 我要報名 | 分享到FaceBook | FAQ

#### 活動辦法

##### 一、活動時間與地點

**報名日期：**即日起至2014年3月10日（一）止

**起跑說明會：**2014年3月14日（五）

**地點：**臺大管理學院正大國際會議廳(台北市基隆路4段144巷50號)

**訓練課程：**2014年3月16日（日）、3月17日（一）、3月18日(二)、及3月19日(三) 共四梯次，每一梯次為期一天。

**訓練地點：**國立台北大學資訊中心1樓 C1F05教室(台北市中山區民生東路三段67號)

<http://www.ettoday.net/events/bigdata2014/>

# Summary

- This course introduces the **fundamental concepts** and **applications technology** of data mining.
- Topics include
  - Introduction to Data Mining
  - Association Analysis
  - Classification and Prediction
  - Cluster Analysis
  - Text and Web Mining
  - Big Data Analytics
  - **Data Mining Using SAS Enterprise Miner (SAS EM)**
  - **Case Study and Implementation of Data Mining**

# Contact Information

戴敏育 博士 (Min-Yuh Day, Ph.D.)

專任助理教授

淡江大學 資訊管理學系

電話：02-26215656 #2347

傳真：02-26209737

研究室：i716 (覺生綜合大樓)

地址：25137 新北市淡水區英專路151號

Email：myday@mail.tku.edu.tw

網址：<http://mail.tku.edu.tw/myday/>

