Course Orientation for Practices of Business Intelligence

1071BI01
MI4 (M2084) (2888)
Wed, 7, 8 (14:10-16:00) (B217)

Min-Yuh Day
Assistant Professor
Dept. of Information Management, Tamkang University

http://mail.tku.edu.tw/myday/
2018-09-12
• 課程名稱：商業智慧實務  
(Practices of Business Intelligence)

• 授課教師：戴敏育 (Min-Yuh Day)

• 開課系級：資管四P (TLMXB4P) (M2244) (2995)

• 開課資料：選修 單學期 2 學分 (2 Credits, Elective)

• 上課時間：週三 7, 8 (Wed 12:10-16:00)

• 上課教室：B217
Business Intelligence (BI)

1. Introduction to BI and Data Science
2. Descriptive Analytics
3. Predictive Analytics
4. Prescriptive Analytics
5. Big Data Analytics
6. Future Trends
淡江資管系（所）教育目標

一、精進資訊管理知能。

二、提升資訊科技專業。

三、獨立思考邏輯分析。

四、強化團隊合作能力。

五、重視企業資訊倫理。

六、培育全球化世界觀。
淡江資管系(所)核心能力

A. 問題分析與關鍵思考。
B. 企業基礎與實務知識。
C. 資訊系統運用。
D. 程式設計。
E. 網路系統規劃。
F. 資料庫設計與管理。
G. 資訊系統分析、設計與整合。
H. 專案管理。
課程簡介

• 本課程介紹商業智慧 (BI) 的基礎概念及技術實務

• 課程內容
  - 商業智慧、分析與資料科學、人工智慧、大數據與雲端運算
  - 描述性分析：數據的性質、統計模型與可視化、商業智慧與資料倉儲、
  - 預測性分析：資料探勘流程、方法與演算法、文本、網路與社群媒體分析
  - 處方性分析：最佳化與模擬
  - 社會網絡分析
  - 機器學習與深度學習
  - 自然語言處理
  - AI交談機器人與對話式商務
  - 商業分析的未來趨勢、隱私與管理考量
Course Introduction

• This course introduces the fundamental concepts and technology practices of business intelligence.

• Topics include

  – Business Intelligence, Analytics, and Data Science,
  – AI, Big Data, and Cloud Computing,
  – Descriptive Analytics: Nature of Data, Statistical Modeling, and Visualization, Business Intelligence and Data Warehousing,
  – Predictive Analytics: Data Mining Process, Methods, and Algorithms, Text, Web, and Social Media Analytics,
  – Prescriptive Analytics: Optimization and Simulation,
  – SNA, Machine and Deep Learning, NLP,
  – AI Chatbots and Conversational Commerce,
  – Future Trends in Analytics.
(Objective)

• Understand and apply the fundamental concepts and technology practice of business intelligence.
Business Intelligence (BI)

1. Introduction to BI and Data Science
2. Descriptive Analytics
3. Predictive Analytics
4. Prescriptive Analytics
5. Big Data Analytics
6. Future Trends
課程大綱 (Syllabus)

週次 (Week) 日期 (Date) 內容 (Subject/Topics)
1 2018/09/12 商業智慧實務課程介紹  
(Course Orientation for Practices of Business Intelligence)
2 2018/09/19 商業智慧、分析與資料科學  
(Business Intelligence, Analytics, and Data Science)
3 2018/09/26 人工智慧、大數據與雲端運算  
(ABC: AI, Big Data, and Cloud Computing)
4 2018/10/03 描述性分析I：數據的性質、統計模型與可視化  
(Descriptive Analytics I: Nature of Data, Statistical Modeling,  
and Visualization)
5 2018/10/10 國慶紀念日 (放假一天) (National Day) (Day off)
6 2018/10/17 描述性分析II：商業智慧與資料倉儲  
(Descriptive Analytics II: Business Intelligence and  
Data Warehousing)
週次 (Week) 日期 (Date) 內容 (Subject/Topics)
7 2018/10/24 預測性分析 I：資料探勘流程、方法與演算法
(Predictive Analytics I: Data Mining Process, Methods, and Algorithms)
8 2018/10/31 預測性分析 II：文本、網路與社群媒體分析
(Predictive Analytics II: Text, Web, and Social Media Analytics)
9 2018/11/07 期中報告 (Midterm Project Report)
10 2018/11/14 期中考試 (Midterm Exam)
11 2018/11/21 處方性分析：最佳化與模擬
(Prescriptive Analytics: Optimization and Simulation)
12 2018/11/28 社會網絡分析
(Social Network Analysis)
<table>
<thead>
<tr>
<th>週次 (Week)</th>
<th>日期 (Date)</th>
<th>內容 (Subject/Topics)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>2018/12/05</td>
<td>機器學習與深度學習 (Machine Learning and Deep Learning)</td>
</tr>
<tr>
<td>14</td>
<td>2018/12/12</td>
<td>自然語言處理 (Natural Language Processing)</td>
</tr>
<tr>
<td>15</td>
<td>2018/12/19</td>
<td>AI交談機器人與對話式商務 (AI Chatbots and Conversational Commerce)</td>
</tr>
<tr>
<td>16</td>
<td>2018/12/26</td>
<td>商業分析的未來趨勢、隱私與管理考量 (Future Trends, Privacy and Managerial Considerations in Analytics)</td>
</tr>
<tr>
<td>17</td>
<td>2019/01/02</td>
<td>期末報告 (Final Project Presentation)</td>
</tr>
<tr>
<td>18</td>
<td>2019/01/09</td>
<td>期末考試 (Final Exam)</td>
</tr>
</tbody>
</table>
教學方法與評量方法

• 教學方法
  - 講述、討論、賞析、模擬、實作、問題解決

• 評量方法
  - 紙筆測驗、實作、報告、上課表現
教材課本

- 講義 (Slides)
- iClass
- http://mail.tku.edu.tw/myday/teaching.htm#1071BI

參考書籍

- 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
  – Business Intelligence
  – AI Challenge Champion
  – Social Network Analysis (SNA)
  – FinTech
  – Short Text Conversation (STC)

• 3-4 人為一組
  – 分組名單於 2018/09/19 (三) 課程下課時繳交
  – 由班代統一收集協調分組名單

Source: https://www.amazon.com/Business-Intelligence-Analytics-Data-Science/dp/0134633288


Source: https://www.amazon.com/Practical-Machine-Learning-Python-Problem-Solvers/dp/1484232062

Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners,
Jared Dean,


Source: https://www.amazon.com/Network-Analysis-Applications-Lecture-Networks/dp/3319781952
Welcome to Colaboratory!

Colaboratory is a free Jupyter notebook environment that requires no setup and runs entirely in the cloud. See our FAQ for more info.

Getting Started

- Overview of Colaboratory
- Loading and saving data: Local files, Drive, Sheets, Google Cloud Storage
- Importing libraries and installing dependencies
- Using Google Cloud BigQuery
- Forms, Charts, Markdown, & Widgets
- TensorFlow with GPU
- Machine Learning Crash Course: Intro to Pandas & First Steps with TensorFlow

Highlighted Features

Seedbank

Looking for Colab notebooks to learn from? Check out Seedbank, a place to discover interactive machine learning examples.

TensorFlow execution

Colaboratory allows you to execute TensorFlow code in your browser with a single click. The example below adds two matrices.

\[
\begin{bmatrix} 1. & 1. & 1. \\ 1. & 2. & 3. \end{bmatrix} + \begin{bmatrix} 1. & 2. & 3. \end{bmatrix} = \begin{bmatrix} 2. & 3. & 4. \end{bmatrix}
\]
Evolution of Decision Support, Business Intelligence, and Analytics

During the 1970s, the primary focus of information systems support for decision making focused on providing structured, periodic reports that a manager could use for decision making (or ignore them). Businesses began to create routine reports to inform decision makers (managers) about what had happened in the previous period (e.g., day, week, month, quarter). Although it was useful to know what had happened in the past, managers needed more than this: They needed a variety of reports at different levels of granularity to better understand and address changing needs and challenges of the business. These were usually called management information systems (MIS). In the early 1970s, Scott-Morton first articulated the major concepts of DSS. He defined DSSs as "interactive computer-based systems, which help decision makers utilize data and models to solve unstructured problems" (Gorry and Scott-Morton, 1971). The following is another classic DSS definition, provided by Keen and Scott-Morton (1978):

"Decision support systems couple the intellectual resources of individuals with the capabilities of the computer to improve the quality of decisions. It is a computer-based support system for management decision makers who deal with semistructured problems." 

Note that the term "decision support system," like "management information system" and several other terms in the field of IT, is a content-free expression (i.e., it means different things to different people). Therefore, there is no universally accepted definition of DSS.

During the early days of analytics, data was often obtained from the domain experts using manual processes (i.e., interviews and surveys) to build mathematical or knowledge-based models to solve constrained optimization problems. The idea was to do the best with limited resources. Such decision support models were typically called operations research (OR). The problems that were too complex to solve optimally (using linear or nonlinear mathematical programming techniques) were tackled using heuristic methods such as simulation models. (We will introduce these as prescriptive analytics later in this chapter and in a bit more detail in Chapter 6.)

In the late 1970s and early 1980s, in addition to the mature OR models that were being used in many industries and government systems, a new and exciting line of models had emerged: rule-based expert systems. These systems promised to capture experts' knowledge in a format that computers could process (via a collection of if–then–else rules or heuristics) so that these could be used for consultation much the same way that one...
Organizations have to work smart. Paying careful attention to the management of BI initiatives is a necessary aspect of doing business. It is no surprise, then, that organizations are increasingly championing BI and under its new incarnation as analytics. Application Case 1.1 illustrates one such application of BI that has helped many airlines as well as, of course, the companies offering such services to the airlines.
A High-Level Architecture of BI

Organizations have to work smart. Paying careful attention to the management of BI initiatives is a necessary aspect of doing business. It is no surprise, then, that organizations are increasingly championing BI and under its new incarnation as analytics. Application Case 1.1 illustrates one such application of BI that has helped many airlines as well as, of course, the companies offering such services to the airlines.

**FIGURE 1.9** Evolution of Business Intelligence (BI).

**FIGURE 1.10** A High-Level Architecture of BI. (Source: Based on W. Eckerson, Smart Companies in the 21st Century: The Secrets of Creating Successful Business Intelligent Solutions. The Data Warehousing Institute, Seattle, WA, 2003, p. 32, Illustration 5.)

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
This idea of looking at all the data to understand what is happening, what will happen, and how to make the best of it has also been encapsulated by INFORMS in proposing three levels of analytics. These three levels are identified (informs.org/Community/Analytics) as descriptive, predictive, and prescriptive. Figure 1.11 presents a graphical view of these three levels of analytics. It suggests that these three are somewhat independent steps and one type of analytics applications leads to another. It also suggests that there is actually some overlap across these three types of analytics. In either case, the interconnected nature of different types of analytics applications is evident. We next introduce these three levels of analytics.
Although some researchers have distinguished business analytics professionals from data scientists (Davenport and Patil, 2012), as pointed out previously, for the purpose of understanding the overall analytics ecosystem, we treat them as one broad profession. Clearly, skill needs can vary between a strong mathematician to a programmer to a modeler to a communicator, and we believe this issue is resolved at a more micro/individual level rather than at a macro level of understanding the opportunity pool. We also take the widest definition of analytics to include all three types as defined by INFORMS—descriptive/reporting/visualization, predictive, and prescriptive as described earlier.

Figure 1.13 illustrates one view of the analytics ecosystem. The components of the ecosystem are represented by the petals of an analytics flower. Eleven key sectors or clusters in the analytics space are identified. The components of the analytics ecosystem are grouped into three categories represented by the inner petals, outer petals, and the seed (middle part) of the flower.

The outer six petals can be broadly termed as the technology providers. Their primary revenue comes from providing technology, solutions, and training to analytics user organizations so they can employ these technologies in the most effective and efficient manner. The inner petals can be generally defined as the analytics accelerators. The accelerators work with both technology providers and users. Finally, the core of the ecosystem comprises the analytics user organizations. This is the most important component, as every analytics industry cluster is driven by the user organizations.

The metaphor of a flower is well-suited for the analytics ecosystem as multiple components overlap each other. Similar to a living organism like a flower, all these petals grow and wither together. We use the terms components, clusters, petals, and sectors interchangeably to describe the various players in the analytics space. We introduce each of the industry sectors next and give some examples of players in each sector. The list of company names included in any petal is not exhaustive. The representative list of companies in each cluster is just to illustrate that cluster’s unique offering to describe where analytics talent may be used or hired away. Also, mention of a company’s name or its capability in one specific sector could appear in another petal.

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
Job Titles of Analytics

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
Although its value proposition is undeniable, to live up its promise, the data has to comply with some basic usability and quality metrics. Not all data is useful for all tasks, obviously. That is, data has to match with (have the coverage of the specifics for) the task for which it is intended to be used. Even for a specific task, the relevant data on hand needs to comply with the quality and quantity requirements. Essentially, data has to be analytics ready. So what does it mean to make data analytics ready? In addition to its rel

- **Relevance** to the problem at hand and the quality/quantity requirements, it also has to have a certain data structure in place with key fields/variables with properly normalized values. Furthermore, there must be an organization-wide agreed-on definition for common variables and subject matters (sometimes also called master data management), such as how you define a customer (what characteristics of customers are used to produce a holistic enough representation to analytics) and where in the business process the customer-related information is captured, validated, stored, and updated. Sometimes the representation of the data may depend on the type of analytics being employed. Predictive algorithms generally require a flat file with a target variable, so making data analytics ready for prediction means that data sets must be transformed into a flat-file format and made ready for ingestion into those predictive algorithms. It is also imperative to match the data to the needs and wants of a specific predictive algorithm and/or a software tool—for instance, neural network algorithms require all input variables to be numerically represented (even the nominal variables need to be converted).
Business Intelligence (BI) Infrastructure

- Operational Data
- Historical Data
- Machine Data
- Web Data
- Audio/Video Data
- External Data

Data Mart

Data Warehouse

Hadoop Cluster

Extract, transform, load

Casual users
- Queries
- Reports
- Dashboards

Power users
- Queries
- Reports
- OLAP
- Data mining

Business Intelligence and Data Mining

Increasing potential to support business decisions

- Decision Making
- Data Presentation
  - Visualization Techniques
  - Data Mining
    - Information Discovery
- Data Exploration
  - Statistical Summary, Querying, and Reporting
- Data Preprocessing/Integration, Data Warehouses
- Data Sources
  - Paper, Files, Web documents, Scientific experiments, Database Systems

Source: Jiawei Han and Micheline Kamber (2006), Data Mining: Concepts and Techniques, Second Edition, Elsevier
Business Insights with Social Analytics
Analyzing the Social Web: Social Network Analysis
Jennifer Golbeck (2013), Analyzing the Social Web, Morgan Kaufmann

AI Challenge Champion
科技大擂台 與AI對話 (正式賽)

報名時間：2017-12-27 ~ 2018-02-28

人工智能（AI）正加速改變全球產業、經濟與社會生活發展型態，亦成為各大產業的發展重點。各項AI技術研發項目中，尤以語音應用為最重要的技術，因為語音對話是人機互動最直覺、最人性化的方式，語意理解技術是AI智慧應用的核心。科技部舉辦台灣首屆「科技大擂台 與AI對話」，以獎勵賽的模式鼓勵創新者運用創意與技術來解決語音AI應用的挑戰。

一、競賽目的

1. 建置多情境的中文語音大數據，提升我國AI團隊技術。
2. 加速中文語音對話的核心技術開發。

二、賽程規劃

<table>
<thead>
<tr>
<th>2017</th>
<th>2018</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>即日起 - 2月底</td>
<td>8月</td>
<td>12月</td>
</tr>
<tr>
<td>報名</td>
<td>初賽</td>
<td>複賽</td>
</tr>
<tr>
<td>專家委員會審查晉級資格</td>
<td>決賽</td>
<td></td>
</tr>
</tbody>
</table>

三、參賽方式與資格

1. 初賽：本賽事採團體報名，團員人數以10名為限。團隊中至少有一位中華民國公民，且其他成員若非本國籍者需持有中華民國工作許可或我國學籍，非本國籍成員報名時須檢附有效中華民國工作證明或在學證明，上述參賽資格需於本賽事報名截止日前(含當日)取得。詳細內容請參考簡章內容。

https://fgc.stpi.narl.org.tw/activity/techai2018
The 14th NTCIR (2018 - 2019)

Evaluation of Information Access Technologies

January 2018 - June 2019

What's New

- **February 1, 2018**: Call for participation to the NTCIR-14 Kick-Off Event released.
- **February 1, 2018**: Call for participation to the NTCIR-14 QALab-PolInfo Kick-Off Event released.

December 5, 2017: The NTCIR-14 Task Selection Committee has selected the following six Tasks: Lifelig-3, OpenLiveQ-2, QA Lab-4, STC-3, WWW-2, CENTRE.

August 23, 2017: NTCIR-14 Call for Task Proposals released. (Closed.)

About Proceedings

After the NTCIR-14 conference, a post-proceedings of revised selected papers will be published in the Springer Lecture Notes on Computer Science (LNCS) series.

http://research.nii.ac.jp/ntcir/ntcir-14/index.html
Welcome to the top page of STC-3@NTCIR-14!

STC-3 offers three subtasks:

- **Chinese Emotional Conversation Generation (CECG) Subtask**
- Dialogue Quality (DQ) Subtask (for Chinese and English)
- Nugget Detection (ND) Subtask (for Chinese and English)

**Key dates for DQ and ND Subtasks**

Feb-Mar 2018 Crawling Chinese test data from Weibo

Oct 2017-Jan 2018 Training data translation into English

Apr-Jun, 2018 Test data translation into English

Jul-Aug 2018 Training/test data annotation

Aug 31, 2018 STC-3 task registrations due (CECG, DQ, ND)

Sep 1, 2018 Training data with annotations released

Nov 1, 2018 Test data released

Nov 30, 2018 Run submissions due

Dec 20, 2018 Results and draft overview released to participants

Feb 1, 2019 Participant papers due

Mar 1, 2019 Acceptance notification

Mar 20, 2019 All camera-ready papers due

Jun 2019 NTCIR-14 Conference@NII
Call for Participation

In recent years, there has been a rising tendency in AI research to enhance Human-Computer Interaction by humanizing machines. However, to create a robot capable of acting and talking with a user at the human level requires the robot to understand human cognitive behaviors, while one of the most important human behaviors is expressing and understanding emotions and affects. As a vital part of human intelligence, emotional intelligence is defined as the ability to perceive, integrate, understand, and regulate emotions. Though a variety of models have been proposed for conversation generation from large-scale social data, it is still quite challenging (and yet to be addressed) to generate emotional responses.

In this challenge, participants are expected to generate Chinese responses that are not only appropriate in content but also adequate in emotion, which is quite important for building an empathic chatting machine. For instance, if user says “My cat died yesterday”, the most appropriate response may be “It’s so sad, so sorry to hear that” to express sadness, but also could be “Bad things always happen, I hope you will be happy soon” to express comfort.

Previous Evaluation Challenge at NLPCC 2017

Overview of the NLPCC 2017 Shared Task: Emotion Generation Challenge
Short Text Conversation (NTCIR-13 STC2)

Retrieval-based retrieval-based method

Short Text Conversation (NTCIR-13 STC2) Generation-based method

Given a new post, can a fluent, coherent and useful comment be generated?

Summary

- This course introduces the fundamental concepts and technology practices of business intelligence.

- Topics include
  - Business Intelligence, Analytics, and Data Science,
  - AI, Big Data, and Cloud Computing,
  - Descriptive Analytics: Nature of Data, Statistical Modeling, and Visualization, Business Intelligence and Data Warehousing,
  - Predictive Analytics: Data Mining Process, Methods, and Algorithms, Text, Web, and Social Media Analytics,
  - Prescriptive Analytics: Optimization and Simulation,
  - SNA, Machine and Deep Learning, NLP,
  - AI Chatbots and Conversational Commerce,
  - Future Trends in Analytics.
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 #2846
傳真：02-26209737
研究室：B929
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/