Please consult Intellectual Property Rights before making a photocopy. Please use the textbook of copyrighted edition.

②國玄東華大學

教學計劃表 Syllabus

課程名稱(中文) Course Name in Chinese	大數據統計分析				學年/學期 Academic Year/Semester		106/2	
課程名稱(英文) Course Name in English	Statistical Analysis of Big Data							
科目代碼 Course Code	EC50430	系級 Department & Year	nt 碩士		開課單位 Course-Offering Department		經濟學系	
修別 Type	必修 Required	學分數/時間 Credit(s)/Hour(s)		·	3.0/3.0			
授課教師 Instructor	/林金龍							
先修課程 Prerequisite								

課程描述 Course Description

Big Data is a broad term for data sets so large or complex that traditional data processing applications are inadequate.

It offers promises for discovering subtle population patterns and heterogeneities that are not possible with small data.

Yet, the huge sample size and high dimensionality of Big Data create unique computational and statistical challenges,

including scalability and storage bottleneck, noise accumulation, spurious correlation, incidental endogeneity and measurement errors.

These challenges demand new computational and statistical methods.

This course focuses on the salient features of Big Data and reviews newly proposed data analytical and statistical methods to

meet the challenges. It consists of five parts. The first part overviews the main characteristics of Big Data

and the architecture for the analysis. Due to its huge sample size, the hardware and software are essential for

effective analysis of Big Data. The second part covers popular methods for data mining including A/B testing, crowdsourcing, data fusion and integration, genetic algorithms, machine learning, natural language processing,

signal processing, simulation, time series analysis, visualisation. tensors, multilinear subspace learning.

As almost financial data is in the format of time series, the third part focuses upon time series mining.

Text mining is the focus of the four part as it becomes more and more important for financial Big Data.

Popular textming techniques include information extraction, topic tracking, categorization, clustering, concept linkage,

information visualization, and association rule mining. We shall cover commonly used text mining algorithms including

k nearest neighbor, support vector machine, Bayesian classifier and K-mean clustering. Final part includes the empirical

application of Big Data analytics. One cannot really master Big Data technology unless he or she could complete

analyzing one real big dataset. The airline data includes on-time information of more than 120 millions domestic flights in US

between 1987 to 2008 and is a perfect place to start the journey. Also, students are required to analyze a real bank marketing dataset.

I choose R as the main software as it is free, powerful and very popular for the analysis of Big Data.

課程目標 Course Objectives											
圖示說明Ⅰ	Illustration :	高度相關	Highly	correl	ated C	中度相	關 Mode	erately	corre	lated	
		授課進	度表	Teachin	g Sched	lule & C	Content				
週次Week	內容 Subject/Topics							備註Remarks			
1	Introduction to Big data and R										
2	Big Data Basics (I)										
3	Big Data Basics (II)										
4	Pattern recognition and association (I)										
5	Pattern recognition and association (II)										
6	Classification (I)										
7	Classification (II)										
8	Classification (III)										
9	期中考試週 Midterm Exam										
10	Clustering (I)										
11	Clustering (II)										
12	Outlier dectect	Outlier dectection									
13	Time series mining (I)										
14	Time series mining (II)										
15	Time series mining (III)										
16	Finance applica	tions									
17	Project present	ation (I)									
18	Project present	Project presentation (II)									
		教	學策	略 Tea	aching	Strateg	ies				
✓ 課堂講	授 Lecture	✓	分組討	論Group	Discus	sion	多	觀實習	Field T	rip	
其他Mi	scellaneous:										
		學期成績計	算及多元	.評量方:	弋 Gradi	ng & As	sessmen	ıts			
配分項目 配分比例 多元評量方式 Assessme						1					
	Items	Percentage	測驗會考	實作觀察	口頭 發表	專題 研究	創作 展演	卷宗 評量	證照	其他	
平時成績 General Performance		50%		~							
期中考成績											
期末考成績	Final Exam										
作業成績 H Assignment	omework and/or s	50%			~	~					

其他 Miscellaneous ()										
評量方式補充說明										
Grading & Assessments Supplemental instructions										
Homework, class attendance and discussion 50%, Project 50%										
教科書與參考書目(書名、作者、書局、代理商、說明) Textbook & Other References(Title, Author, Publisher, Agents, Remarks, etc.)										
Main textbooks:										
Jiawei Han, Micheline Kamber, and Jian Pei, \textbf{Data Mining: Concepts and Techniques}, 3rd										
edition, Morgan Kaufmann Publishers, , 2012.										
Reference books: Ian H. Witten, Eibe Frank and Mark A. Hall: Data Mining: Practical Machine Learning Tools and Techniques, (Third Edition), Morgan Kaufmann Publishers, 2011, e-book available at NDHU library Michael W. Berry and Jacob Kogan, Text Mining Applications and Theory, John Wiley 2010. Yanchang Zhao, R and Data Mining: Examples and Case Studies, Academic Press, 2013, e-book available at NDHU library										
課程教材網址(教師個人網址請列在本校內之網址)										
Teaching Aids & Teacher's Website (Personal website can be listed here.)										
其他補充說明(Supplemental instructions)										