國立東華大學

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		105/2
課程名稱(英文) Course Name in English	Statistical Analysis of Big Data					
科目代碼 Course Code	FIN_52690	系級 Department & Year	碩士	開課單位 Course-Offering Department		財務金融學系
修別 Type	選修 Elective	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 \textit{large} or \textit{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.
\textit{Text mining} is the focus of the four part as it becomes more and more important for
financial Big Data.
Popular text ming 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 t0 2008 and is a perfect place to start the journey. Also, students are required to
extract some websites, say
 Twitter and examine if messages there can provide useful information for stock price prediction.

   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

授 課 進 度 表 Teaching Schedule & 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	Classifcation (I)	
7	Classifcation (II)	
8	Classifcation (III)	
9	期中考試週 Midterm Exam	
10	Clustering (I)	
11	Clustering (II}	
12	Outlier dectection	
13	Time series mining (I)	
14	Time series mining (II)	
15	Time series mining (III)	
16	Finance applications	
17	Project presentation (I)	
18	Project presentation (II)	

教 學 策 略 Teaching Strategies

☑ 課堂講授 Lecture　　　☑ 分組討論Group Discussion　　　☐ 參觀實習 Field Trip

☐ 其他Miscellaneous:

學期成績計算及多元評量方式 Grading & Assessments

配分項目 Items	配分比例 Percentage	多元評量方式 Assessments							
		測驗會考	實作觀察	口頭發表	專題研究	創作展演	卷宗評量	證照檢定	其他
平時成績 General Performance	50%		✔						
期中考成績 Midterm Exam									
期末考成績 Final Exam									
作業成績 Homework and/or Assignments	50%			✔	✔				

其他 Miscellaneous (_____)										

Homework, class attendance and discussion 50%, Project 50%

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