THE UNIVERSITY OF HONG KONG
FACULTY OF BUSINESS AND ECONOMICS

ECON 6086: Digital Economy and Big Data Analysis

GENERAL INFORMATION

Instructor: Dr. Ping YU
Email: pingyu@hku.hk
Office: KKL 1110
Phone: 2857-8358
Consultation times: 10:30-11:30am, Thursday
Teaching time: 9:00am – 12:30pm + 2:00pm – 5:30pm, Friday or Sunday
Teaching location: Cyberport Classroom H or KKL101

Tutor: Lily Wang
Email: lily959@hku.hk

** Please send your emails to me directly from your email account instead of using the email communication facility in Moodle.*

Pre-requisite(s): Applied Econometrics (ECON6001) or Econometric Theory (ECON6005)
Co-requisite(s): None
Mutually exclusive: None

Course website: Available through HKU Portal e-learning
Other important details: None

COURSE DESCRIPTION

This course introduces fundamental ideas, important methods and popular techniques in big data analysis and machine learning. Combining statistical theory, computational tools, and hands-on experience with real data, this course will provide students with a solid basis for handling big data in the practice of economics, finance, and management.

This course is a combination of theory and application, but the emphasis is put on application, and the theory is introduced only to aid understanding of key ideas. Anyway, a strong background in statistics and/or econometrics is an advantage (although not required) to this course. At minimum, the students should have learned least squares, standard error, t-statistic, p-value, and confidence interval before taking this course. These concepts will be mentioned only briefly in the first lecture; the remaining lectures will focus on some genuinely new techniques (i.e., not covered in the other courses of our MECON program).

The evaluation is based on four assignments and one project. The problems in the assignments are mainly from the textbook; the project asks the students to analyse a real dataset given by the teacher. Coding is a key ingredient to all these turn-ins; we will use the statistical package R via a front-end called RStudio for all programming. Both R and RStudio are free and open source.

COURSE OBJECTIVES

1. Provide students a broad overview of the most popular data science methods in economics and business studies
2. Enhance students' analytical ability to apply appropriate methods in different contexts
3. Equip students with a basic toolkit that can be directly used for their own research

Programme Learning Outcomes

PLO1.	Understanding of fundamental theories and new development in economics
PLO2.	Mastering of skills in analyzing economic data
PLO3.	Demonstration of ability to apply economic knowledge and analytical skills to address policy and

business problems

PLO4. Awareness of ethical concerns in economic issues

PLO5. Mastering of communication skills

COURSE LEARNING OUTCOMES (CLOs)

Course Learning Outcomes	Aligned Programme Learning Outcomes
CLO1. Gain a solid understanding of the principles of applying data science to social sciences.	PLO 1-3
CLO2. Demonstrate a solid grounding in recent developments in big data methods, including state-of-the-art machine learning techniques and their suitability to solve important economic, finance, and business problems.	PLO 1-4
CLO3. Demonstrate ability to address questions of interest by using applied data science and econometric techniques.	PLO 2-5
CLO4. Demonstrate facility with implementing the techniques covered in the course using statistical software on real-world datasets.	PLO 2-5

COURSE TEACHING AND LEARNING ACTIVITIES

Course Teaching and Learning Activities	Expected Study Hours	Study Load (% of study)
T&L1. Lectures	36	30
T&L2. Weekly problem sets	36	30
T&L3. Computer programming	24	20
T&L4. Tutorial/self-learning sessions	24	20
Total	120	100

Assessment Methods	Brief Description (Optional)	Weight	Aligned Course Learning Outcomes
A1. Four Problem Sets		80%	CLO 1-4
A2. One Project		20%	CLO 1-4
	Total	100%	

STANDARDS FOR ASSESSMENT

Course Grade Descriptors

A+, A, A-	Strong evidence of superb ability to fulfill the intended learning outcomes of the course at all levels of learning: describe, apply, evaluate and synthesis.
B+, B, B-	Strong evidence of ability to fulfill the intended learning outcomes of the course at all levels of learning: describe, apply, evaluate and synthesis.
C+, C, C-	Evidence of adequate ability to fulfill the intended learning outcomes of the course at low levels of learning; such as describe and apply, but not at high levels of learning such as evaluate and synthesis.
D+, D	Evidence of basic familiarity with the subject.
F	Little evidence of basic familiarity with the subject.

Assessment Rubrics for Each Assessment (Please provide us the details in a separate file if the space here is not enough)

Problem sets are mainly extracted from the textbook, including both analytical and empirical exercises. A few other exercises are not from the textbook, but are helpful to understand the course materials. The project asks the students to predict a quantitative or qualitative outcome using the techniques learned in this course or even out of this course. The best prediction will get the full score, and the worst prediction will get a pass score. Students can form a team up to five members to do the problem sets and the project. The team members for each of the four assignments and the project need not be the same, i.e., each student can join at most five teams.

COURSE CONTENT AND TENTATIVE TEACHING SCHEDULE

Lecture 1. Introduction to Statistical Learning
Lecture 2. Classification
Lecture 3. Clustering
Lecture 4: Model Selection and Regularization
Lecture 5: Principal Components Analysis
Lecture 6*: Text as Data
Lecture 7: Moving Beyond Linearity
Lecture 8: Tree-Based Methods
Lecture 9: Support Vector Machines
Lecture 10: Deep Learning

REQUIRED/RECOMMENDED READINGS & ONLINE MATERIALS (e.g. journals, textbooks, website addresses etc.)

Required: (**ISLR**) *An introduction to Statistical* Learning, 2nd edition. James, Witten, Hastie, and Tibshirani. Springer. 2021.

Useful references:

Data Science for Business (DSB): *What you need to know about data mining and data-analytic thinking.* Provost and Fawcett. O'Reilly. 2013.

Business Data Science (BDS): *combining machine learning and economics to optimize, automate, and accelerate business decisions*. Taddy. McGraw Hill. 2019.

The Elements of Statistical Learning (ESL): data mining, inference, and prediction, 2nd edition. Hastie, Tibshirani, and Freidman. 2017.

MEANS/PROCESSES FOR STUDENT FEEDBACK ON COURSE

○ conducting mid-term survey in additional to SFTL around the end of the semester

○ Online response via Moodle site

✓ Others: ___SFTL_____ (please specify)

COURSE POLICY (e.g. plagiarism, academic honesty, attendance, etc.)

1. This is an active learning course, and attendance and participation are extremely important. Please observe appropriate classroom etiquette and be considerate to others. In particular, laptop use should be limited to course-related activities, and cell phones are not allowed in class.
2. All the turn-ins must be typewritten.
3. Plagiarism and cheating are serious academic offenses, so copying other teams' answers is not permitted even with consent.

ADDITIONAL COURSE INFORMATION (e.g. e-learning platforms & materials, penalty for late assignments, etc.)

All course materials can be downloaded from Moodle. Late assignments and project are not acceptable for whatever reasons. To avoid any risk, start your assignments and project early (the assignments indicate clearly which problems can be solved after each lecture).