HKU BUSINESS SCHOOL
港大經管學院

Inspire·Empower·Lead

1. General Information	
Course Subject	ECON
Course Number	3225
Course Title	Big Data Economics
Academic Years	2023-2024
Grading Method	Letter

2. Instructors
Professor HAO, Yu Jasmine Office: Room 927 9/F K.K. Leung Building Email: haoyu@hku.hk Office: 39177759 Subclasses: 2A
Professor YU, Ping Office: Room 1110 11/F K.K. Leung Building Email: pingyu@hku.hk Office: 28578358 Subclasses: 2A

4. Course Description	
Course Description	This course introduces students to 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 economics, finance, and management. The primary focus of this course is on application instead of methodological rigor. Hence, the use of mathematics will be limited to an elementary level. However, students are expected to have a strong background in statistics and/or econometrics. In particular, the course will assume that participants have an understanding of statistical inference using t-tests and have prior experience of interpreting the results of multiple linear regression. We will review these topics briefly during the course. Because of the emphasis on hands-on experience, students are expected not to be scared by data and coding. Previous experience with statistical software and knowledge about computer programming is an advantage but not required. Homework assignments are designed to familiarize students with the necessary programing language. For programing, we will use the statistical package R via a front-end called RStudio. Both R and RStudio are free and open source.
Prerequisites	ECON2280: Introductory Econometrics, or ECON3284: Causal Inference
Free Elective	Yes

5. 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

6. Faculty Learning Goals
Goal 1: Acquisition and internalization of knowledge of the programme discipline
Goal 2: Application and integration of knowledge
Goal 3: Inculcating professionalism
Goal 4: Developing global outlook
Goal 5: Mastering communication skills
Goal 6: Cultivating leadership

7. Course Learning Outcomes						
Course Teaching and Learning Activities	Aligned Faculty Learning Goals					
	1	2	3	4	5	6
CLO1. Gain a solid understanding of the principles of applying data science to social sciences	✔	✔				
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.	✔	✔			✔	
CLO3. Demonstrate ability to address questions of interest by using applied data science and econometric techniques.	✔	✔		✔	✔	
CLO4. Demonstrate facility with implementing the techniques covered in the course using statistical software on real-world datasets	✔	✔	✔			

8. 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	Total: 100

9. Assessment Methods			
Assessment Methods	Description	Weight %	Aligned Course Learning Outcomes

9. Assessment Methods			
A1. Four Problem sets	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. 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.	80%	1,2,3,4
A2. One Final Project	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 three 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.	20%	1,2,3,4
A3. Final Exam		0%	

10. 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, implement, 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, implement, 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.

11. Course Content and Tentative Teaching Schedule			
Topic/ Session	Date	Content	Other information
1		Introduction to Statistical Learning	
2		Classification	
3		Clustering	
4		Model Selection and Regularization	
5		Principal Components Analysis	
6		Text as Data	
7		Moving Beyond Linearity	
8		Tree-Based Methods	
9		Support Vector Machines	
10		Deep Learning	

12. Required/Recommended Readings & Online Materials	
Reading	**Useful references**: **Data Science for Business**: What you need to know about data mining and data-analytic thinking. Provost and Fawcett. O'Reilly. 2013. **Business Data Science**: combining machine learning and economics to optimize, automate, and accelerate business decisions. Taddy. McGraw Hill. 2019.
Textbook	Required: (**ISLR2**) An introduction to Statistical Learning, 2nd edition. James, Witten, Hastie, and Tibshirani. Springer. 2021. (The digital version of this book is freely available on https://hastie.su.domains/ISLR2/ISLRv2_website.pdf)

13. Means / Processes for Student feedback on Course	
	Conducting mid-term survey in additional to SETL around the end of the semester
✔	Online response via Moodle site
✔	Others
	SETL

14. Course Policy

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.

15. Additional Course Information

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).