CUHK Business School

Course Syllabus

DSME6653BA: Web Analytics and Intelligence

Winter 2023

INSTRUCTOR 1: Professor Renyu (Philip) Zhang

E-mail: philipzhang@cuhk.edu.hk

Telephone: 852-3943-7763

Office Hours: Monday 1:00PM-2:00PM or by appointment

Office Location: Cheng Yu Tung Building 911

INSTRUCTOR 2: Professor Hongfan (Kevin) Chen

E-mail: kevinchen@cuhk.edu.hk

Telephone: 852-3943-7902

Office Hours: by appointment

Office Location: Cheng Yu Tung Building 936

COURSE MEETINGS

Meeting Times: Tuesday, 6:30PM-9:15PM

Class Will Meet on: December 5, 12, 19; January 2, 9, 16, 23, 30; February 6, 20, 27 (Final Exam)

Location: Wu Ho Man Yuen Building (WMY) 303

Lecture Format: In person

Zoom Room: 918 5858 4939, Pass code: 718897

Lecture Videos: Available on GitHub.

TEACHING ASSISTANT: Yilin Shi

Email: ylshi@link.cuhk.edu.hk

Office Hour: By appointment.

Office Hour Location: By appointment.

COURSE DESCRIPTION

This course aims to provide an introduction to the modern landscape of internet, online platforms, and digital economy, together with the fundamental analytics tools to analyze the business problems thereof. In the first half of the course, we will cover the methodological topics that are widely used in the online technology sector, such as web-scraping, A/B testing, instrumental variables, difference-in-difference, synthetic control, and machine-learning powered causal inference. In the second half, we will cover some demand forecasting models, the estimation procedures and the corresponding algorithms. We will also learn to apply data science tools in the field of revenue management.

The course is entirely hands-on. You will learn how to apply data analytics to a wide array of business models in the era of digital economy, with the help of the programming language Python. Cases will be drawn from the internet tech sector to showcase the power and challenges of web analytics. The emphasis will be on applying analytics techniques to internet economy, but not on the underlying mathematical theory.

LEARNING GOALS

You will

- Develop analytics models and methods (causal inference in particular) that can be used to improve business decision making in the internet technology sector.
- Understand the basic business models in the internet technology sector.
- Prepare for a career in the internet technology track as an entry-level colleague.

PREREQUISITES

There are no official prerequisite courses for Web Analytics and Intelligence. However, prior knowledge in statistics, machine learning, econometrics, and micro-economics will be helpful.

Knowledge of basic algebra (including functions such as the quadratic, exponential, and logarithmic) and simple logic is also assumed.

Though prior experiences in coding are not assumed, the course relies on the programming language Python as the platform for model building and implementation. So you should also not be averse to programming. Python will be the "official" programming languages for this course in the sense that all classroom demonstrations, homework solutions, and reading materials will be coded in Python (Jupyter Notebook). Having said that, feel free to use other languages (such as *R*, Stata,

C++, Java, MatLab, or Excel spreadsheet) if you feel more comfortable with.

Finally, you should not be averse to analytical thinking and quantitative analysis in general.

TEXTBOOKS

There is NO required textbook for this course. Below are some useful reference books.

- *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* 2nd Edition, 2018, by W. McKinney.
- *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, 2015, by Guido W. Imbens, Donald B. Rubin.
- *Mostly Harmless Econometrics*, 2009, by Joshua D. Angrist, Jörn-Steffen Pischke.
- *Causal Inference in Python: Applying Causal Inference in the Tech Industry*, 2023, by Matheus Facure.
- *Causal Inference: The Mixtape*, 2021, by Scott Cunningham. Link to the Book.
- *Causal Inference for the Brave and True*, 2022, by Matheus Facure Alves. Link to the Book.
- *Numerical Optimization*, 2006, by Jorge Nocedal and Stephen J. Wright.

COURSE WEBSITE/MATERIALS

Course materials, including slides, lecture notes, codes, and optional readings, will be distributed electronically through the GitHub course website. Furthermore, the homework problem sets will be distributed on Blackboard, where you should also submit your homework solutions. All lecture videos will be uploaded on GitHub.

The link to the GitHub of this course:

- `https://github.com/DSME-6653/Web-W2023`

The GitHub repository is private. Please note the invitation sent by the TA to join.

GRADING

At CUHK, we strive to create courses that challenge students intellectually and that meet high standards of academic excellence. For this course, I will give grades the students deserve, while maintaining rigorous academic standards. Your course grade will be based on

- **Class Participation** (20% of the final grade)
- **Problem Sets** (30% of the final grade)
- **Projects** (20% of the final grade for the first project + bonus 5% of the final grade for the second project)

- **Final Exam** (30% of the final grade)

Class Participation

20% of the final grade will be assigned on the basis of class participation and individual professional conduct thereof. Class participation includes class discussion of assignments and cases, presentation of an exercise solution, as well as active participation in lectures. I expect you to arrive to class on-time and be prepared, and to stay involved during class sessions.

In person attendance is mandatory. You will lose 3% of the final grade for each session you missed without prior permission (or were late for more than 15 minutes). Missing four sessions or more without prior permission will lead to a failing grade. Every conceivable effort should be made to avoid absences, late arrivals or early departures. In cases when these are unavoidable, they need to be communicated to us 2 hours in advance. If you miss a class session for any reason, you are required to watch the videos after class.

Your baseline Class Participation grade will be 12%. The other 8% will be given based on your classroom performance.

Problem Sets

There will be (roughly) 5 problem sets distributed throughout this course in total. All problem sets will be posted on Blackboard and are due 1 week after their distribution, when the reference solutions will be posted. You should submit before the deadline a Jupyter Notebook via Blackboard. At our discretion, late assignments will either not be accepted or will incur a grade penalty unless due to documented serious illness or family emergency. I will make exceptions for religious observance or civic obligation only when the assignment cannot reasonably be completed prior to the due date and the student makes arrangements for late submission with me in advance. The 3 problem sets with the highest grades will be counted, 10% of the final grade each.

You are allowed to discuss with each other about the questions in the problem sets, but you should submit your own solutions.

Projects

You will finish two projects in groups, one at the middle of this semester and the other at the end. Each project will count towards 10% of the final grade. We will follow up with the detailed instructions of the projects later in this course. Please stay tuned.

Each group has at most 2 students. Please send us your groups by the distribution of the first project. We recommend you form groups with some mixed skill sets.

Final Exam

The final exam of this course will be scheduled in the last session (February 27), at 6:30pm-9:00pm.

The exam will cover the course materials of the entire semester, and will be close-book and close-notes. You can take a two-sided and A4-sized cheat sheet. The use of a scientific non-programmable calculator is allowed in the final exam. However, you are not allowed to use cellphones, laptops, iPads, or any other electronic devices during the exam.

We follow the same grading practices as the CUHK Business School. The following grades may be awarded: A, A-, B+, B, B-, C+, C, C-, D+, D, F. In general, A indicates excellent work, B indicates good work, C indicates satisfactory work, and D indicates passable work and is the lowest passing grade. F indicates failure.

The process of assigning grades is intended be one of unbiased evaluation. This means that students are encouraged to respect the integrity and authority of the professor's grading system and discouraged from pursuing arbitrary challenges to it.

If a student feels that an inadvertent error has been made in the grading of an individual assignment or in assessing an overall course grade, a request to have that the grade be re-evaluated may be submitted. Students should submit such requests in writing to the professor within 7 calendar days of receiving the grade, including a brief written statement of why he or she believes that an error in grading has been made.

CLASS WORK

The process of building analytics models and implementing them in a programming language to solve a business problem in the internet technology sector is the most important and difficult problem-solving skill you will learn in this course. It involves developing a structure to conceptualize, formalize and (quantitatively) analyze a given problem. It seems deceptively simple to watch someone else do it, but the only way to learn this skill is by practicing it yourself. Therefore, this course involves a hands-on, in-class learning experience. Attending each class and bringing a laptop computer to class are essential. Preparation for each class involves reading and thinking about the problems/cases to be covered in class. The problems will be posted on GitHub one week in advance. The Jupyter Notebooks of the problems modeled and analyzed in class should be downloaded from GitHub before (not during) the class.

Classroom Norms: Cell phones and other electronic devices are a disturbance to both students and to me. All electronic devices (except laptops) must be turned off prior to the start of each class meeting.

Laptops: You are expected to bring a laptop to each class, unless otherwise instructed. But we will not use it throughout each class. Please close your laptop until you are asked to use it (and it is used for class exercises only).

Class recording: We will record the lecture of this course and post them on GitHub for students to review the materials covered in class.

Students with Disabilities: Please refer to the Support Services for Students with Disabili-

ties (https://www2.osa.cuhk.edu.hk/disability/en-GB/).

WECHAT GROUP

We will establish a WeChat group as the off-class online discussion platform for this course. All students are required to enter this group, and are encouraged to post and discuss any questions, suggestions, and/or comments about this course in the class WeChat group. Students who actively contribute to the discussions in our WeChat group may receive some extra credits in the final course grade.

ACADEMIC INTEGRITY

Integrity is critical to the learning process and to all that we do here at the CUHK Business School. As members of our community, all students agree to abide by the Academic Honesty policies of CUHK (see https://www.cuhk.edu.hk/policy/academichonesty/ for details), which includes a commitment to:

- Exercise integrity in all aspects of one's academic work including, but not limited to, the preparation and completion of exams, papers and all other course requirements by not engaging in any method or means that provides an unfair advantage.
- Clearly acknowledge the work and efforts of others when submitting written work as one's own. Ideas, data, direct quotations (which should be designated with quotation marks), paraphrasing, creative expression, or any other incorporation of the work of others should be fully referenced.
- Refrain from behaving in ways that knowingly support, assist, or in any way attempt to enable another person to engage in any violation of the Academic Honesty policies of CUHK. Our support also includes reporting any observed violations that are deemed to adversely affect the CUHK community.
- You may not submit the same work (or substantially similar work) to meet the requirements of more than one course without written consent of all instructors concerned.

COURSE EVALUATIONS

Course evaluations are important to us and to students who come after you. Please complete them thoughtfully.

COURSE OUTLINE

The course schedule below is tentative and subject to minor changes.

Important Dates

DATE AND TIME	EVENT
Tuesday, January 9	Project 1 distributed
Tuesday, February 20	Project 2 (bonus) distributed
Tuesday, February 20, 11:59PM	Project 1 due
Tuesday, February 27, 6:30PM	Final Exam
Sunday, March 3, 11:59PM	Project 2 due

Weekly Schedule

Session	Date	Instructor	TOPIC
1	December 5	Philip Zhang	Introduction
2	December 12	Philip Zhang	A/B Testing
3	December 19	Philip Zhang	Regression and Matching
4	January 2	Philip Zhang	Instrumental Variables and Panel Data
5	January 9	Philip Zhang	Machine Learning Powered Causal Inference
6	January 16	Hongfan (Kevin) Chen	Demand Model and Forecasting
7	January 23	Hongfan (Kevin) Chen	Least Square and Maximum Likelihood Estimation
8	January 30	Hongfan (Kevin) Chen	Convex Optimization and Algorithms
9	February 6	Hongfan (Kevin) Chen	Auctions and Customized Pricing
10	February 20	Hongfan (Kevin) Chen	Network Revenue Management
11	February 27	N.A.	Final Exam