CUHK Business School

Course Syllabus

DSME6756BA/B: Business Intelligence Techniques and Applications

Winter 2023

INSTRUCTOR: 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

COURSE MEETINGS

Meeting Times: Monday, 9:30AM-12:15PM or 2:30PM-5:15PM

Class Will Meet on: December 4, 11, 18; January 8, 15, 22, 29; February 5, 19, 26; Mar 02 (Final Exam)

Location: Morning: Wu Ho Man Yuen Building (WMY) 508; Afternoon: Yasumoto International Academic Park (YIA) LT6

Lecture Format: In person

Zoom Room: 984 6416 2488, Pass code: 263862 (Morning); 970 8075 7238, Pass code: 930454 (Afternoon)

Lecture Videos: Available on GitHub.

TEACHING ASSISTANT: Qinlu Hu

Email: qinlu.hu@link.cuhk.edu.hk

Office Hour: By appointment.

Office Hour Location: By appointment.

COURSE DESCRIPTION

Business analytics refers to the scientific process of transforming data into insights for making better business decisions. Analytics has enabled many businesses, non-profits, and governments to improve their routine activities, to identify long-run decision opportunities, and sometimes to rethink the whole of their activities. In this sense, business analytics has tactical and strategic value; it is an important factor of value creation.

This course introduces the basic concepts, principles and techniques of business analytics. In the first module of the course, we focus on analytics basics, such as descriptive analytics, data pre-processing, and coding with Generative AI. In the second module, we study predictive analytics, i.e., how to use data to develop insights and predictive capabilities using statistics, machine learning, deep learning, and forecasting techniques.

The course is entirely hands-on. You will learn how to apply business analytics to a wide array of business decision problems with the help of the programming language Python. Examples will be drawn from Kwai, Netflix, and Canadian Bank, among others.

The emphasis will be on model formulation, the implementation of analytics techniques, and the interpretation of results, but not on the underlying mathematical theory.

LEARNING GOALS

You will

- Think critically about data and learn how to analyze data to create business value.
- Develop business analytics models that can be used to improve decision making within an organization.
- Sharpen your ability to structure problems and to perform logical analyses.
- Practice translating practical decision problems into formal models, and investigate those models in an organized fashion.
- Strengthen your programming skills, focusing on how to use Python programming to support business decision making.

PREREQUISITES

There are no official prerequisite courses for Business Intelligence Techniques and Applications. However, prior knowledge in statistics, machine learning, and/or econometrics 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*, 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.
- *The Analytics Edge*, 2016, by Dimitris Bertsimas, Allison O'Hair, William Pulleyblank.
- *Data Science for Business: What You Need to Know about Data Mining and Data-Analytical Thinking*, 2013, by Foster Provost and Tom Fawcett.
- *Pattern Recognition and Machine Learning*, 2006, by Christopher M. Bishop.
- *The Elements of Statistical Learning* (2nd Edition), 2009, by Trevor Hastie, Robert Tibshirani, Jerome Friedman.
- *Deep Learning with PyTorch*, 2020, by Eli Stevens, Luca Antiga, and Thomas Viehmann.

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 Blackboard.

The link to the GitHub of this course:

- `https://github.com/DSME6756-2023/BA-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** (10% of the final grade)
- **Problem Sets** (30% of the final grade)
- **Projects** (20% of the final grade)

- **Final Exam** (40% of the final grade)

Class Participation

10% 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 2% 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 me 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 6%. The other 4% will be given based on your classroom performance.

Problem Sets

There will be (roughly) 1 problem set distributed after each class. 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 my 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 5 problem sets with the highest grades will be counted, 6% 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

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

The projects will be evaluated based on

- whether the analytics problem is clearly defined;
- whether the model is correctly formulated;
- whether the analysis and methodologies are rigorous;

- whether your results are reasonably and well-interpreted;
- whether your report is well written.

Final Exam

The final exam of this course will be scheduled on the Saturday of the last week in Term 2 (March 2), at 2:00pm-5:00pm. If you have any time conflicts, please let me know as soon as possible. We will finalize the final exam by the end of Week 1. 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 is the most important and difficult problem-solving skill. 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 Blackboard for students to review the materials covered in class.

Students with Disabilities: Please refer to the Support Services for Students with Disabilities (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.

DATA SCRAPING

Sometimes you need to get your own data from the web. In this case, you will need to directly scrap publicly available data from legal sources. Although we cannot cover data scraping for the interest of time, we will provide references and demos for data scrapping in the course GitHub.

COURSE OUTLINE

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

Important Dates

DATE AND TIME	EVENT
Monday, January 15	Project 1 distributed
Monday, February 5	Project 2 distributed
Monday, February 26, 11:59PM	Project 1 due
Saturday, March 2, 2:00PM	Final Exam
Sunday, March 3, 11:59PM	Project 2 due

Weekly Schedule

Session	Date	TOPIC
1	December 4	Introduction and Python Basics
2	December 11	Data Pre-processing and Coding with AI
3	December 18	Linear Models for Classification and Regression
4	January 8	Bias-Variance Trade-off
5	January 15	Classification and Regression Trees
6	January 22	Ensemble Learning
7	January 29	Clustering
8	February 6	Principal Component Analysis
9	February 19	Neural Networks
10	February 26	Generative Artificial Intelligence and Beyond
11	March 2 (2:00PM)	Final Exam