THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學
Opening Minds • Shaping the Future

# Machine Learning Empowered Business Intelligence (MLE.BI) – Syllabus

Spring 2024

(The most updated version: https://www.dropbox.com/scl/fi/nb4df0rw3uymdv2nj13of/MLE.BI-PolyU-MM6761.pdf?rlkey=kmfvdpth4jel1ipbxjtavsocy&dl=0)

## Basic Information

**Subject Code** : MM6761
**Subject Title** : Machine Learning Empowered Business Intelligence (hereafter, MLE.BI)
**Semester** : Semester 2, Academic Year 2023-24
**Credit Value** : 3
**Level** : PhD students across all years and all disciplines
**Instructor** : Dr YAO Dai, M827
**Email** : daiyao@polyu.edu.hk; dai@yaod.ai
**URL** : http://www.yaod.ai
**Office Hour** : By appointment

## Objectives

Recent development in machine learning, such as Autoencoders, deep neural networks (DNN), transformers, generative adversarial networks (GAN), and most recently, artificial general intelligence (AGI), has greatly impact the way researchers and practitioners analyze and understand data in different business contexts, in particular, marketing contexts. For example, applying Autoencoders to complex networks of users and brands, marketers now can project all users and brands into a common space and obtain their representations or embeddings, and perform a series of downstream tasks such as user segmentation, market structure analysis of the brands, as well as user-brand matching. For another example, with the help of DNNs, marketers now can decipher the emotion of a streamer in a live stream based on her facial expression, and quantify the effect of positive emotion, i.e., smile, on the number of tips viewers send into the live stream.

This course will provide a systematic discussion of the recent development in machine learning, and how it empowers the derivation of business intelligence. It will help students to build the abilities to appreciate the new techniques and criticize the emerging studies in the quantitative marketing literature using these new techniques, to develop their own research ideas, and to conduct high-quality research themselves in the future.

## Prerequisite and Preparation

The course is meant for PhD students across all years and all disciplines, though readings and discussions will be primarily marketing oriented. To understand the fundamentals of various machine learning algorithms, some basic understandings of linear algebra, matrix, probability,

statistics, as well as calculus are needed. In addition, as the majority of the course is empirical, some basic knowledge in economics and econometrics are demanded as well.

The format of the class will be primarily discussions. We will discuss in detail the core machine learning algorithm, the specific invention/advancement to the core algorithm, the research questions, the data, and the main results in each of the selected papers. We will also criticize (hopefully) all aspects of each paper ranging from the technique to data, to analysis, and to results. We will also discuss other potential extensions to the core algorithms.

## Learning Outcomes

Upon completion of the subject, students will be able to:

a) Understand the key techniques and core models in machine learning that are used to solve marketing problems.
b) Develop an ability to criticize research methodologies and findings in the quantitative marketing literature.
c) Cultivate and advance their own research interests in the quantitative marketing and/or related areas such as information system, economics, and computer science.

## Assessment

| Specific Assessment Methods / Tasks | Weighting (%) |
|---|---|
| **Continuous Assessment** | **100** |
| • In-class discussions | • 25 |
| • Take-home assignments | • 25 |
| • Paper presentations | • 25 |
| • Term paper | • 25 |
| **Total** | **100** |

*In-class discussions: 25%*
Each student must participate in the discussions in class actively. The performance in each class in terms of the frequency and quality of participation will be accumulated to determine the final grade for this component.

*Take-home assignments: 25%*
There will be 5 take-home assignments, and each is worth 5%. They will be released during class one week before the due week. Two types of problems will appear in the assignments: (a) describe the key invention in methodology in a paper and elaborate on how the invention helps to uncover the main results of the paper; (b) formally derive the core methodology in a paper and interpret the main results.

*Paper presentations: 25%*
Each student will give two presentations over the semester. Each presentation will be 20 minute long and focuses on a paper of their choice. Presentation time will be finalized by week 3 and presentation starts in week 3 as well. The choice of paper for presentation could be determined as the semester progresses.

*Term paper: 25%*
Each student will submit a term paper when the semester ends. The term paper should be focused on one or multiple research questions that are plausibly original, which may arise throughout the semester. Ideally, some new machine learning techniques will be developed to solve the research question(s). Of course, it is also fine to adopt any existing technique from the literature if adequate. The term paper also needs to include some results based on real or synthetic data.

## Student Study Effort Required

| | |
|---|---|
| **Class contact:** | |
| • Lectures and seminars | 39 Hrs. |
| **Other study effort:** | |
| • Preparation for lectures and seminars<br>• Working on homework, presentations, and term paper | 39 Hrs. (est.)<br>39 Hrs. (est.) |
| **Total student study effort** | 117 Hrs. |

## Schedule and Outline

| Week | Date | Topic | Activities and Deliverables |
|---|---|---|---|
| 1 | 19 January | Introduction | |
| 2 | 26 January | Traditional empirical marketing | **Jan 27, Add/Drop period ends** |
| 3 | 2 February | Key machine learning models | Take-home assignment #1 |
| 4 | 9 February | Feature extraction from text | |
| | 16 February | - | No class, Lunar New Year Break<br>Happy New Year! |
| 5 | 23 February | Feature extraction from others | Take-home assignment #2 |
| 6 | 1 March | Market structure analysis with unstructured data | |
| 7 | 8 March | Modeling generation of unstructured data | Take-home assignment #3 |
| 8 | 15 March | GenAI and business | |
| 9 | 22 March | Causal inference: experiments | Take-home assignment #4 |
| | 29 March | - | No class, Good Friday |
| 10 | 5 April | Causal inference: advanced methods | |
| 11 | 12 April | Spatio-temporal data mining | |
| 12 | 19 April | New development in structural models | Take-home assignment #5 |

## Declaration of Use of Generative Artificial Intelligence (GenAI)

You are *encouraged* to use Generative AI tools in all the assignments. If you use Generative AI tools in any of the assignments, please add a specific declaration section at the end of your submission like the following:

Declaration:

I/We declare that Generative AI tools have been used to prepare the submitted work. The Generative AI tools used and the manner in which they were used are as follows: