

Data Mining for Business Analytics

ISOM 3360 (L1 & L2): Fall 2023

Course Name	Data Mining for Business Analytics
Course Code	ISOM 3360
No. of Credit	3 Credits
Exclusion(s)	COMP 4331, IDEA 3560
Prerequisite(s)	ISOM 2010
Professor	Rong Zheng, ISOM
Contact	Office: LSK 4042 Tel: 2358 7642 Email: rzheng@ust.hk
Office Hours	by appt.
Course Schedule and Classroom	L1: Tue & Thur 12:00PM - 01:20PM LSK 1009 L2: Tue & Thur 09:00PM - 10:20AM LSK 1014 Lab1: Fri 09:30AM - 10:20AM LSK G005 Lab2: Thur 03:00PM - 03:50PM LSK G005 Lab3: Wed 05:00PM - 05:50PM LSK G005
Course Webpage	Accessible from Canvas
Teaching Assistant	Sophie GU (LSK 4065) Email: imsophie@ust.hk Tel: 2358 7653 Ray Pang (LSK 4065) Email: imncpang@ust.hk Tel: 2358 7653

TA Office Hours	By appointment
-----------------	----------------

Course Overview

"For every leader in the company, not just for me, there are decisions that can be made by analysis. These are the best kinds of decisions. They're fact-based decisions." --- Amazon's Chairman, Jeff Bezos

Businesses, governments, and individuals create massive collections of data as a byproduct of their activities. Increasingly, decision-makers rely on intelligent technology to analyze data systematically to improve decision-making. In many cases, automating the decision-making processes is required because of the sheer volume of data and the speed with which new data are generated.

The course will explain with real-world examples the uses and some details of various data analytics techniques. The emphasis is placed on understanding both the methods and their business applications. We will delve into the mechanics of how the methods work when necessary. You will establish analytical thinking to the problems and understand that the proper application of analytics is as much an art as it is a science.

After the completion of this course, the students are expected to achieve the following objectives.

1. Approach business problems data-analytically (intelligently). Think carefully & systematically about whether & how data can improve business performance. Be able to envision Big Data opportunities.
2. Know the basics of Big Data processes, techniques, & systems well enough to interact with data scientists, business analysts, marketers, and clients.
3. Gain hands-on experience in using Python to implement Big Data projects.

Lecture Notes and Readings

- Lecture notes are the primary source of materials for the class.
- Supplemental readings might be posted to Canvas or distributed in class to supplement the lecture notes.
- Supplemental book (optional):
Data Science for Business: What you need to know about data mining and data-analytic thinking, by Foster Provost, Tom Fawcett, O'Reilly Media, 2013 ISBN: 1449361323

Grading Decomposition

The grade breakdown is as follows:

- In-class exercise 10%
- Lab participation: 10%
- Individual Assignments: 10%
- Group Assignment: 15%
- Midterm quiz: 25%
- Final Exam: 30%

Lab Session

This is primarily a lecture-based course, but lab participation is an essential part of the learning process in the form of active practice. You are NOT going to learn without practicing the data analysis yourselves. During the lab session, I will expect you to be entirely devoted to the class by following the instructions. And you should actively link the empirical results you obtained during the lab to the concepts you learned in the lectures.

Assignments

Individual Assignment

There will be **2 individual** homework, each comprising short answer questions. Completed assignments must be handed in via Canvas.

Group Assignment

There will be **2 assignments** that are mini projects in nature. They are designed to train your practical skills to solve the real-world problems by analyzing the data. Python coding is essential in completing these assignments. Groups of **2-3 students** are expected to form a team along the semester to work on the assignments. Peer evaluation will be conducted for each assignment.

Assignments are due prior to the start of the lecture on the due date. Turn in your assignment early if there is any uncertainty about your ability to turn it in on the due date. Assignments up to 24 hours late will have their grade reduced by 25%; assignments up to one week late will have their grade reduced by 50%. After one-week, late assignments will receive no credit.

Exams

The mid-term quiz is tentatively scheduled on **October 17th Evening**. Let me know as early as possible if there is any unavoidable conflict.

Hardware and Software requirements

Computer is required for this class. You will need it to work on your assignments. The software we are going to use is **Anaconda Navigator** that includes **Jupyter notebook** and **Python 3**. You will get some guidance on the software installation once the class has started.

Tentative Schedule of Lectures and Labs

Please take note that this schedule is tentative and may be adjusted as the semester progresses.

Class Number	Date	Topics	Remarks
1	Sept. 5	Course Introduction	
2	Sept. 7	Overview of Data mining process	
3	Sept. 12	Data Preparation and data Visualization	
	Sept. 14	No Class (The second day following the Mid-Autumn Festival)	
4	Sept. 19	Decision tree learning	Assignment #1 (individual) release
5	Sept. 21		
6	Sept. 26		
7	Sept. 28	Model Selection and Evaluation	Assignment #1 Due Assignment Team Formation
8	Oct. 3	Prediction: Linear Regression	Assignment #2 (Group) release
9	Oct. 5	Prediction: Logistic Regression	
10	Oct. 10	Prediction: Naïve Bayes	
11	Oct. 12	Midterm Exam Review	
	Oct. 17	Midterm Quiz (7:00pm – 9:pm)	No regular class
12	Oct. 19	Text Mining	Assignment #2 Due

13	Oct. 24	Feature Selection	Assignment #3 (individual) release
14	Oct. 26	Relationship Mining: Association Rule	
15	Oct. 31	Relationship Mining: k-means	
16	Nov. 2	Prediction: k-nearest neighbor	Assignment #3 Due Assignment #4 (Group) release
17	Nov. 7	Application: Recommender System	
18	Nov. 9	Ensemble learning	
19	Nov. 14	Neural Network and Deep Learning	
20	Nov. 16	Neural Network and Deep Learning	
21	Nov. 21	Network Analytics	
22	Nov. 23	Search Engine Technology	
23	Nov. 28	TBD	
24	Nov. 30	Course Review	

Lab Session Schedule

Number	Date	Topics
1	Sept. 6/7/8	Introduction to Anaconda and Jupyter Notebook
2	Sept. 13/14/15	Data visualization and Data Preprocessing

3	Sept. 20/21/22	Decision tree (model building)
4	Sept.27/28/ 29	Decision tree II (CV and Grid Search)
5	Oct. 4/5/6	Model Evaluation (ROC, AUC)
6	Oct. 11/12/13	Linear Regression and Logistic Regression
7	Oct. 18/19/20	Naïve Bayes
8	Oct. 25/26/27	Text Mining
9	Nov. 1/2/3	Association Rule
10	Nov. 8/9/10	Clustering & KNN
11	Nov. 15/16/17	Collaborative Filtering
12	Nov. 22/23/24	Ensemble learning