Data Mining for Business Analytics

ISOM 3360 (L1 & L2): Fall 2024

Course Name	Data Mining for Business Analytics
Course Code	ISOM 3360
No. of Credit	3 Credits
Exclusion(s)	COMP 4331 & IEDA 3560
Prerequisite(s)	ISOM 2020 OR COMP 1021
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 1:30PM - 2:50PM LSK 1011 L2: Tue & Thur 3:00PM - 4:20PM LSK 1011 Lab1: Mon 1:30PM - 2:20PM LSK G005 Lab2: Wed 3:30PM - 4:20PM LSK G005 Lab3: Mon 4:30PM - 05:20PM LSK G005
Course Webpage	Accessible from Canvas
Teaching Assistant	Sophie GU (LSK 4065) Email: imsophie@ust.hk Tel: 2358 7653 Aaron Lee (LSK 4065) Email: imaaron@ust.hk Tel: 2358 7653
TA office hour	by appt.

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 place 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
3. Gain hands-on experience in using Python to implement Big Data

Lecture Notes and Readings

Lecture notes are the primary source of materials for the

- Supplemental readings might be posted to Canvas or distributed in class to supplement the lecture

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%
- Group Assignment: 25%
- 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

There will be **group assignments** that are mini projects in nature. They are designed to train your practical skills to solve 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 throughout 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 22nd 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. 3	Course Introduction	
2	Sept. 5	Overview of Data mining process	
3	Sept. 10	Data Preparation and data Visualization	
4	Sept. 12	Prediction: Linear Regression	
5	Sept. 17	Prediction: Logistic Regression	Team formation Due
6	Sept. 19	Model Selection and Evaluation	Assignment #1 release
7	Sept. 24	Model Selection and Evaluation	
8	Sept. 26	Decision tree learning	
	Oct. 1	No Class: National Day	
9	Oct. 3	Decision tree learning	Assignment #1 Due
10	Oct. 8		Assignment #2 release
11	Oct. 10	Prediction: Naïve Bayes	

12	Oct. 15	Prediction: Naïve Bayes	
13	Oct. 17	Midterm Exam Review	
	Oct. 22	Midterm Quiz (7:00pm – 9:00pm)	No regular class
14	Oct. 24	Text Mining	Assignment #2 Due Assignment #3 release
15	Oct. 29	Feature Selection	
16	Oct. 31	Relationship Mining: Association Rule	
17	Nov. 5	Relationship Mining: k-means	Assignment #3 Due Assignment #4 release
18	Nov. 7	Prediction: k-nearest neighbour	
19	Nov. 12	Application: Recommender System	
20	Nov. 14	Ensemble learning	
21	Nov. 19	Neural Network and Deep Learning	
22	Nov. 21	Network Analytics	
23	Nov. 26	GenAI and LLM	Assignment #4 Due
24	Nov. 28	Course Review	

Lab Session Schedule

1	Sept. 9/4	Introduction to Anaconda and Jupyter Notebook
2	Sept. 16/11	Data visualization and Data Preprocessing
3	Sept. 23/25	Linear Regression and Logistic Regression
4	Sept. 30/Oct. 2	Model Evaluation (CV and Grid Search)
5	Oct. 7/9	Model Selection (ROC and AUC)
6	Oct. 14/16	Decision Tree
7	Oct. 21/23	Naïve Bayes
8	Oct. 28/30	Text Mining
9	Nov. 4/6	Association Rule
10	Nov. 11/13	Clustering & KNN
11	Nov. 18/20	Collaborative Filtering
12	Nov. 25/27	Ensemble learning