ISOM 3360: Data Mining for Business Analytics
Spring 2024

Course Name	Data Mining for Business Analytics
Course Code	ISOM 3360, 3 Credits
Instructor	Yi Yang, ISOM
Contact	Office: LSK 4041 Email: imyiyang@ust.hk **Begin subject: [ISOM3360] …**
Office Hours	By appointment via Zoom
Course Schedule and Classroom	Lecture (L2): TuTh 01:30PM - 02:50PM, Rm 1007, LSK Bldg Lecture (L3): TuTh 09:00AM - 10:20AM, Rm 1014, LSK Bldg LA1: We 03:00PM - 03:50PM, G005, LSK Bldg LA2: Tu 05:00PM - 05:50PM, G005, LSK Bldg LA3: We 04:30PM - 05:20PM, G005, LSK Bldg LA4: Mo 09:00AM - 09:50AM, G005, LSK Bldg
Course Webpage	Accessible from Canvas
Teaching Assistant	Sophie Gu (imsophie@ust.hk) Ray Pang (imncpang@ust.hk)
TA Office Hours	By appointment

1. 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 CEO, Jeff Bezos

This course will change the way you think about data and its role in business.

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

In virtually every industry, data mining has been widely used across various business units such as marketing, finance and management to improve decision making. In this course, we discuss specific scenarios, including the use of data mining to support decisions in customer relationship management (CRM), market segmentation, credit risk management, e-commerce, financial trading and search engine strategies.

The course will explain with real-world examples the uses and some technical details of various data mining techniques. The emphasis primarily is on understanding the **business application** of data mining techniques, and secondarily on the variety of techniques. We will discuss the mechanics of how the methods work only if it is necessary to understand the general concepts and business applications. You will establish analytical thinking to the problems and understand that proper application of technology is as much an art as it is a science.

The course is designed for students with various backgrounds -- the class **does not** require any

technical skills or prior knowledge.

After taking this course you should:

1. Approach business problems data-analytically (intelligently). Think carefully & systematically about whether & how data can improve business performance.
2. Be able to interact competently on the topic of data mining for business intelligence. Know the basics of data mining processes, techniques, & systems well enough to interact with business analysts, marketers, and managers. Be able to envision data-mining opportunities.
3. Be able to identify the right BI techniques for various business problems. Gain hands-on experience in using Python and get ready for the job positions that require familiarities with the data analytics.

2. Lecture Notes and Readings

- Course Materials

All courses' materials (Lecture slides, assignments, and lab handouts) are available on Canvas course website.

- Supplemental books (optional):

  ◇ *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R*, by Galit Shmueli, Peter C. Bruce, Inbal Yahav, Nitin R. Patel, Kenneth C. Lichtendahl, ISBN: 1118879368

  ◇ *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

  ◇ *Learning Data Mining with Python*, by Robert Layton, ISBN: 1787126781

- Software: *Anaconda Navigator (for Win-64, OSX-64, and Linux-64)*
  ◇ Jupyter notebook, Python 3

3. Requirements and Grading

Your grades will be determined based on lab participation, homework assignments, group project, midterm exam, and final exam.

Component	Percentage
Lab	5%
Class Attendance/Participation	10%
Homework Assignments (2)	10%
Group Project	15%
Midterm Exam	30%
Final Exam	30%

4. Important Notes on the 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.

During the lab session, you will gain hands-on experience with Python - a popular programming language for programming beginners.

5. Homework Assignments, Term Project and Exams

There will be a total of **2 individual** homework assignments, each comprising questions to be answered and hands-on tasks. Completed assignments must be handed in via Canvas prior to the start of the class on the due date. Assignments will be graded and returned promptly.

Turn in your assignments 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.

You are expected to finish a term project. The term project is a teamwork, which means you need to first form a team. Each team includes up to 4 students. In this project, you will apply the data mining techniques you learned in the class to solve real-world problems. The deliverable is a written report summarizing what you have done and what you have achieved. More details will be provided later.

This course will have two closed-book exams. The midterm exam will test issues covered in the first half of the course. The final exam will cover the classes in the second half of the course. Review sessions will be scheduled to help you prepare for these examinations. The midterm exam is tentatively scheduled on **March 19 (7:00-8:30pm).** Let me know as early as possible if there is any unavoidable conflict. The final exam will be held during the final examination period; the date will be announced later in the semester.

Make-up exam policy: https://arr.ust.hk/reg/em/em_std_reg/reg_makeup.html

To quote, "If students wish the University to take into account illness or some other extenuating circumstances that have affected their performance in an examination, or ability to attend an examination, or to complete other assessment activities, they must report the circumstances of the case in writing and provide appropriate documentation to ARR, Academic Registry within one week of the scheduled date of the assessment activity. The Academic Registrar will review the case and make a recommendation to the relevant Dean, the Dean's designate or the Director of IPO."

Academic Integrity

Students at HKUST are expected to observe the Academic Honor Code at all times (see http://acadreg.ust.hk/generalreg.html for more information). Zero tolerance is shown to those who are caught cheating on any quiz or exam. In addition to receiving a zero mark on the quiz or exam involved, the final course grade will appear on your record with an X, to show that the grade resulted from cheating. This X grade will stay with your record until graduation. If you receive another X grade, you will be dismissed from HKUST.

Tentative Lecture Schedule

Week	Date	Topics	Remarks
1	Feb 1	Course Overview	
	Feb 6	Basic Concepts	
2	Feb 8	Data Preparation	
	Feb 13	**No class (Holiday break)**	
3	Feb 15	Decision Tree Learning	
	Feb 20	Decision Tree Learning Continued	Project Announcement
4	Feb 22	Model Evaluation	
	Feb 27	Model Evaluation Continued	Team Formation Deadline HW1 out
5	Feb 29	Identifying Good Problems for ML	
	Mar 5	Project Idea Meeting	Project Idea Report Due
6	Mar 7	Linear Model	HW1 due
	Mar 12	Neural Network	
7	Mar 14	**Midterm Review**	
	Mar 19	NLP and Sentiment Analysis	
8	Mar 21	Computer Vision and Image Recognition	Project Progress Report Due
	Mar 26	Association Rule Learning	HW2 out
9	Mar 28	**No class (Mid-term break)**	
	Apr 2	**No class (Mid-term break)**	
10	Apr 4	**No class (Mid-term break)**	
	Apr 9	k-means Clustering	
11	Apr 11	k-nearest Neighbor	HW2 Due
	Apr 16	Project Progress Meeting	
12	Apr 18	Project Progress Meeting	
	Apr 23	Recommendation Systems	
13	Apr 25	Ensemble Learning	
	Apr 30	Dimension Reduction	
14	May 2	Time series analysis	
	May 7	Advanced topics in data mining	Project Final Report Due
15	May 9	**Final Exam Review**	

Tentative Lab Schedule

Week	Date	Topics
1		Introduction to Anaconda and Jupyter Notebook
2		Cancelled for Public Holiday
3		Data Visualization and Data Preprocessing
4		Decision Tree
5		Overfitting/cross validation
6		Model evaluation / Cost benefit analysis
7		Linear Regression & Logistic Regression
8		Text Mining & Sentiment Analysis
9		Image Recognition
10		Cancelled for Study break
11		Association Rule
12		Clustering & k-nearest Neighbors
13		Recommendation System
14		Ensemble Learning