ISOM 3360: Data Mining for Business Analytics

Spring 2022

Course Name	Data Mining for Business Analytics
Course Code	ISOM 3360
No. of Credit	3 Credits
Zoom Meeting ID	484-862-792
Exclusion(s)	COMP 4331
Prerequisite(s)	ISOM 2010
Professor	Jing Wang, ISOM
Contact	Office: LSK 4044 Tel: 3469-2125 Email: jwang@ust.hk **Begin subject: [ISOM3360] …**
Office Hours	By appointment
Course Schedule and Classroom	Lecture (L3): Mo 04:30PM - 05:50PM & Fr 12:00PM - 01:20PM (LSK 1011) Lab 1: We 9:00am – 9:50am Lab 2: We 10:30am – 11:20am Lab 3: We 3:00pm – 3:50pm Lab 4: We 1:30pm – 2:20pm
Course Webpage	Accessible from Canvas
Teaching Assistant	Sophie Gu (LSK 4065) Tel: 2358-5728 Email: imsophie@ust.hk **Begin subject: [ISOM3360] …**
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 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
  - ✧ Vscode

3. Requirements and Grading

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

Component	**Percentage**
Lab	5%
Class Participation	10%
Homework Assignments (2)	15%
Group Project	20%
Midterm Exam	20%
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 very 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 about 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 **Mar 18 (In-class).** 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.

Tentative Schedule of Lectures and Labs

Week	Date	Topics	Remarks
1	Feb 4	Overview of the Course	
	Feb 7	Data Mining Basics	
	Feb 11	Data Understanding and Preparation	
2	Feb 14	Decision Tree Learning (I)	
	Feb 18	Decision Tree Learning (II)	Project Announcement
3	Feb 21	Model Selection and Evaluation Measures	Team Formation
	Feb 25	Cost-sensitive Classification	Homework 1 out
4	Feb 28	Linear Regression	
	Mar 4	Logistic Regression	
5	Mar 7	Naïve Bayes Classifier	Project Idea (Mar 9)
	Mar 11	Project Idea Meeting	Homework 1 due
6	Mar 14	Midterm Review	
	Mar 18	Midterm Exam	
7	Mar 21	Text Mining	
	Mar 25	Feature Selection	
8	Mar 28	Association Rule Learning	
	Apr 1	K-Means Clustering	Project Summary (Apr 2)
9	Apr 4	Project Progress Meeting	
	Apr 8	Project Progress Meeting	
10	Apr 11	K-Nearest Neighbors	Homework 2 out
	Apr 15	[No Class] Good Friday	
	Apr 18	[No Class] Easter Monday	
	Apr 22	Recommender System using Collaborative Filtering	
11	Apr 25	Ensemble Learning	
	Apr 29	Neural Networks and Deep Learning	Homework 2 due
12	May 2	[No Class] The day following the Labor Day	
	May 6	Final Exam Review	Project Final Report
13	May 9	[No Class] The day following the Birthday of the Buddha	

Lab Session Schedule (tentative)

Date	Topics
Feb-09	Introduction to Anaconda and Jupyter Notebook
Feb-16	Data Visualization and Data Preprocessing
Feb-23	Decision Tree
Mar-02	Overfitting/cross validation
Mar-09	Model evaluation / Cost benefit analysis
Mar-16	Linear Regression & Logistic Regression
Mar-23	Naive Bayes
Mar-30	Text Mining & Sentiment Analysis
Apr-06	Association Rule
Apr-13	Cancelled for Study break
Apr-20	Clustering & k-nearest Neighbors
Apr-27	Recommendation System
May-04	Ensemble Learning