

Data Mining for Business Analytics

ISOM 3360 (L1 & L2): Fall 2019

| | |
|-------------------------------|--|
| Course Name | Data Mining for Business Analytics |
| Course Code | ISOM 3360 |
| No. of Credit | 3 Credits |
| Exclusion(s) | COMP 4331 |
| Prerequisite(s) | ISOM 2010 |
| Professor | Rong Zheng, ISOM |
| Contact | Office: LSK 4042 Tel: 2358 7642 Email: rzheng@ust.hk |
| Office Hours | and by appt. |
| Course Schedule and Classroom | L1: Mon 10:30AM - 11:50AM (LSK 1034) Wed 10:30AM - 11:50AM (LSK 1034) L2: Mon 01:30PM - 02:50PM (LSK 1011) Fri 09:00AM - 10:20AM (LSK 1011) Lab1: Tue 10:30AM - 11:20AM (LSK G005) Lab2: Wed 09:00AM - 09:50AM (LSK G005) Lab3: Tue 12:00PM - 12:50PM (LSK G005) |
| Course Webpage | Accessible from Canvas |
| Teaching Assistant | Sophie GU (LSK 4049C) Tel: 2358 5428 imsophie@ust.hk |
| TA Office Hours | By appointment |

1. Course Overview

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 tools/techniques for various business problems.* Gain hands-on experience in using popular BI tools and get ready for the job positions that require familiarities with the BI tools.

2. Lecture Notes and Readings

- Lecture notes

For most classes I will hand out lecture notes, which will outline the primary material for the class. Other readings are intended to supplement the material we learn in class. They give alternative perspectives and additional details about the topics we cover:

- Supplemental readings posted to Canvas or distributed in class.

- 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

- Software: Anaconda Navigator

- Jupyter notebook
- Python 3

3. Requirements and Grading

The grade breakdown is as follows:

1. Lab participation: 10%
2. Homework (2): 15%
3. Group Project 15%
4. Midterm quiz: 25%
5. Final exam: 35%

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 Assignment and Exams

Homework

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

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.

Term project

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 3-4 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.

Exams

The mid-term quiz is tentatively scheduled on **October, 14**. 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

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

| Class Number | Date | Topics | Assignment Due Dates |
|--------------|-----------------|--|---|
| 1 | Sept. 2 | Course Introduction | |
| 2 | Sept. 4/6 | Overview of Data mining process | |
| 3 | Sept. 9 | Data Preparation and data Visualization | |
| 4 | Sept. 11/13 | Decision tree learning | |
| 5 | Sept. 16 | | Team Formation |
| 6 | Sept. 18/20 | Model Selection and Evaluation Measures | |
| 7 | Sept. 23 | Prediction: Linear Regression | Project Idea |
| 8 | Sept. 25/27 | Project Idea Meeting | |
| 9 | Sept. 30 | Prediction: Logistic Regression | |
| 10 | Oct.2/4 | Prediction: Naïve Bayes | Homework 1 Due (Oct. 5th) |
| | Oct. 7 | Chung Yeung Festival | |
| 11 | Oct. 9/11 | Midterm Exam Review | |
| | Oct. 14 | Midterm Quiz (evening, 7:30pm – 9:30pm) | |
| 12 | Oct. 16/18 | Text Mining | |
| 13 | Oct. 21 | Feature Selection | |
| 14 | Oct. 23/25 | Relationship Mining: Association Rule | |
| 15 | Oct. 30 /Nov. 1 | Relationship Mining: k-means | |
| 16 | Nov. 4 | Prediction: k-nearest neighbor | |
| 17 | Nov. 6/8 | Project Progress Meeting | |
| 18 | Nov. 9 | Application: Recommender System | |
| 19 | Nov. 11 | Ensemble learning | Homework 2 Due |

| | | | |
|----|------------|----------------------------------|----------------------|
| 20 | Nov. 13/15 | Neural Network and Deep Learning | |
| 21 | Nov. 18 | Network Analytics | |
| 22 | Nov. 20/22 | Search Engine Technology | |
| 23 | Nov. 25 | TBC | Project Final Report |
| 24 | Nov. 27/29 | Course Review | |

Lab Session Schedule

| Number | Date | Topics |
|--------|-------------|---|
| 1 | Sept. 3/4 | Introduction to Anaconda and Jupyter Notebook |
| 2 | Sept. 10/11 | Data visualization and Data Preprocessing |
| 3 | Sept. 17/18 | Decision tree |
| 4 | Sept. 24/25 | Decision tree II |
| | Oct. 1/2 | Cancelled for National Day |
| 5 | Oct. 8/9 | Model Evaluation |
| | Oct. 15/16 | Cancelled for Midterm Exam |
| 6 | Oct. 22/23 | ROC and Linear Regression and Logistic Regression |
| 7 | Oct. 29/30 | Naïve Bayes and Text Mining |
| 8 | Nov. 5/6 | Association Rule |
| 9 | Nov. 12/13 | Clustering & KNN |
| 10 | Nov. 19/20 | Collaborative Filtering |
| 11 | Nov. 26/27 | Ensemble learning |