

# Data Mining for Business Analytics

## ISOM3360: Summer 2022

Course Name	Data Mining for Business Analytics
Course Code	ISOM 3360 (3 Credits)
Exclusion	COMP 4331
Prerequisite	ISOM 2010
Instructor	Jiexin ZHENG (Jason)
Contact	Office: LSK 4063 Office Hours: By Appointment Email: <a href="mailto:jzhengas@connect.ust.hk">jzhengas@connect.ust.hk</a>
Course Schedule and Classroom	Lecture: Tue & Thu: 9:00am-11:50am (Zoom) Lab: Tue & Thu: 12:00pm-12:50pm (Zoom)
Course Webpage	Accessible from Canvas

### 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.

The course will explain with real-world examples of 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.

After taking this course, you should:

1. Approach business problems data-analytically (intelligently). Think carefully and systematically about whether and 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, and 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 data science tools and get ready for the job positions that require familiarities with the data science tools.

The detailed course schedule is shown below:

Week	Date	Topics	Assignments
1	June 21	C1 - Introduction	Group formation
		C2 - Overview of the Data Mining Process	
		LAB0 - Introduction to Anaconda and Jupyter notebook	
	June 23	C3 - Data Preparation	HW1 release
C4 - Decision Tree I			
	LAB1 - Data Exploration and Data Preprocessing		
2	June 28	C5 - Decision Tree II	
		C6 - Model Evaluation	
		LAB2 - Decision Tree	
	June 30	C7 - Model Evaluation ROC	Project release
C8 - Linear Regression			
	LAB3 - Overfitting and Cross-Validation		
3	Jul 05	C9 - Logistic Regression	HW1 due HW2 release
		LAB4 - Cost-Benefit Analysis and ROC	
	Jul 07	C10 - Naive Bayes	
		C11 - Naive Bayes Classifier Application	
	LAB5 - Linear Regression & Logistic Regression		
4	Jul 12	C12 - Association Rule Learning	Project proposal submission
		C13 - Clustering	
		LAB6 - Naive Bayes	
	Jul 14	C14 - K-Nearest Neighbor Classification	HW2 due HW3 release
C15 - Collaborative Filtering			
	LAB7 - Association Rule and K-Means Clustering		

5	Jul 19	Curriculum review and gusted speaker presentation	
	Jul 21	C16 - Network Analysis C17 - Ensemble Learning	
		LAB8 - KNN	
6	Jul 26	C18 - Text Mining	HW3 due
		LAB9 - Ensemble Learning	
	Jul 28	C19 - Neural Network and Deep Learning	
		LAB10 - Text Mining	
7	Aug 02	C20 - Latest Development in AI	
		LAB11 - Neural Network and Deep Learning	
	Aug 04	Project Presentation	Project due

## Lecture Notes and Readings

- All course materials (Lecture slides, assignments, and lab handouts) are available on the class website.
- Supplemental books (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

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

Lab Participation	5%
Class Participation	10%
Homework Assignments (× 3)	55% (15%, 20%, and 20%)
Group Project	20%
Presentation	10%
Total	100%

## **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. The lab participation is based on attendance, in which you need to attend at least ten labs in order to obtain the full mark.

## **Important Notes on the Class Participation**

I highly appreciate your in-class participation. I will expect you to actively ask questions and participate in group discussions. There will be several small in-class quizzes (MC questions) to help you consolidate your understanding of the class materials. These quizzes will also be counted toward your class participation score.

## **Homework Assignment and Term Project**

- Homework Assignment (55%)

There will be a total of **3** 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 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 (20%)

The term project is 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.

- Project Presentation (10%)

Each team will deliver a 15-min presentation (10-min project presentation + 5-min Q&A) in the last class. The purpose is to allow your classmates to comment on your work and exercise your insights on a big data project that engages in real situations. The assessment will mainly be based on your understanding of materials covered in class and your analytical mindset that revealed from your presentation.

### **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.