

**City University of Hong Kong
Course Syllabus**

**offered by Department of Management Sciences
with effect from Semester B 2021/22**

Part I Course Overview

Course Title: Data Mining

Course Code: MS6711

Course Duration: One Semester

Credit Units: 3

Level: P6

Medium of Instruction: English

Medium of Assessment: English

Prerequisites: MS5217 Statistical Data Analysis
(Course Code and Title)

Precursors: Nil
(Course Code and Title)

Equivalent Courses: Nil
(Course Code and Title)

Exclusive Courses: FB6711 Data Mining
(Course Code and Title)

Part II Course Details

1. Abstract

This course introduces students to a range of popular and practical data mining and machine learning algorithms relevant to business applications. Students are required to perform data analysis using the python programming language. Upon successful completion of this course, students will have acquired the core foundational knowledge in the field, and be well-prepared for a wide variety of careers in data-analytics.

2. Course Intended Learning Outcomes (CILOs)

(CILOs state what the student is expected to be able to do at the end of the course according to a given standard of performance.)

| No. | CILOs | Weighting (if applicable) | Discovery-enriched curriculum related learning outcomes (please tick where appropriate) | | |
|-----|--|------------------------------|---|----|----|
| | | | A1 | A2 | A3 |
| 1. | Demonstrate knowledge of state-of-the-art machine learning / data mining algorithms | 30% | ✓ | ✓ | ✓ |
| 2. | Define and formulate real-world data mining problems; prepare the data for data mining projects; execute algorithms and interpret outputs. | 30% | ✓ | ✓ | ✓ |
| 3. | Analyze real-world data using python. | 40% | ✓ | ✓ | ✓ |
| ... | | | | | |
| | | 100% | | | |

A1: Attitude

Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.

A2: Ability

Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to self-life problems.

A3: Accomplishments

Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.

3. Teaching and Learning Activities (TLAs)

(TLAs designed to facilitate students' achievement of the CILOs.)

| TLA | Brief Description | CILO No. | | | | | Hours/week (if applicable) |
|--------------------------|---|----------|---|---|---|---|----------------------------|
| | | 1 | 2 | 3 | 4 | 5 | |
| Lectures | Lecturer explains and discusses the concepts, algorithms, of data mining and how to evaluate the quality of the extracted information. | ✓ | ✓ | ✓ | ✓ | | |
| In-class activities | Lecturer will demonstrate the use of python in class on selected topics. Students are required to work individually or as a group on simulated or small real data using the programming language. Through these in-class exercises, the lecturer can identify the common problems that students have and give more elaboration as needed. Students can also identify the kinds of mistakes that they have made and learn how to correct them. | ✓ | | ✓ | ✓ | ✓ | |
| Out-of-Class assignments | Running the data mining algorithms on large data set is a very time consuming process. It is not possible to do it regularly in class. Students tackle focused problems based on large business data as out-of-class assignments. Students may work in small groups for these assignments so that they can discuss the problems and come up a solution together. | ✓ | | ✓ | ✓ | ✓ | |

| | | | | | | | | |
|---------|---|---|---|---|---|---|--|--|
| Project | <p>The ultimate aim of the course is to provide students with the specialist knowledge and training to run a business data mining task. Students are given a large data set with described business problem. They are asked to extract useful information related to the set of identified data mining goals for the problem. This is likely to be a semester-long activity.</p> <p>Students need to make use of everything they have learned in this course in order to achieve the goals. They are encouraged to form small groups for the project so that they can analyze the data and run the software together. They can always seek help and advice from the lecturer during the semester.</p> | ✓ | ✓ | ✓ | ✓ | ✓ | | |
|---------|---|---|---|---|---|---|--|--|

4. Assessment Tasks/Activities (ATs)

(ATs are designed to assess how well the students achieve the CILOs.)

| Assessment Tasks/Activities | CILO No. | | | | | Weighting | Remarks |
|---|----------|---|---|---|---|-----------|---------|
| | 1 | 2 | 3 | 4 | 5 | | |
| Continuous Assessment: <u>60</u> % | | | | | | | |
| Assignments | ✓ | ✓ | ✓ | ✓ | ✓ | 10% | |
| Project | ✓ | ✓ | ✓ | ✓ | ✓ | 50% | |
| Examination: <u>40</u> % (duration: <u>3 Hours</u> , if applicable) | | | | | | | |
| Examination | ✓ | ✓ | ✓ | ✓ | ✓ | 40% | |
| | | | | | | 100% | |

5. Assessment Rubrics

(Grading of student achievements is based on student performance in assessment tasks/activities with the following rubrics.)

| Assessment Task | Criterion | Excellent (A+, A, A-) | Good (B+, B, B-) | Fair (C+, C, C-) | Marginal (D) | Failure (F) |
|-----------------|---|--------------------------|---------------------|---------------------|-----------------|-----------------------------------|
| Assignments | Core concepts and ideas; use of appropriate statistical methods | High | Significant | Moderate | Basic | Not even reaching marginal levels |
| Examination | Core concepts and ideas; use of appropriate statistical methods | High | Significant | Moderate | Basic | Not even reaching marginal levels |
| Project | Core concepts and ideas; use of appropriate statistical methods | High | Significant | Moderate | Basic | Not even reaching marginal levels |

Part III Other Information (more details can be provided separately in the teaching plan)

1. Keyword Syllabus

(An indication of the key topics of the course.)

1. Introduction to data mining

Explain background of data mining / machine learning. Brief introduction to the history, current development, and future.

2. Data Mining Procedure

Demonstrate the steps of data mining, including problem definition, data preparation and cleaning, model execution and result interpretation. Show examples in python.

3. Data Mining Algorithms

Algorithms to be covered include clustering (K-means and nearest neighbor), linear regression, logistic regression, decision trees, tree ensembles (random forest, boosting), and neural networks. Understand the fundamental problem of bias-variance trade-off and the importance of cross-validation in machine learning.

4. Case Studies

Applications on real data.

5. Data Mining Software

Python.

2. Reading List

2.1 Compulsory Readings

(Compulsory readings can include books, book chapters, or journal/magazine articles. There are also collections of e-books, e-journals available from the CityU Library.)

Müller, Andreas C., and Sarah Guido. Introduction to machine learning with Python: a guide for data scientists. " O'Reilly Media, Inc.", 2016.

2.2 Additional Readings

(Additional references for students to learn to expand their knowledge about the subject.)

| | |
|----|--|
| 1. | Han, Jiawei, Kamber, Micheline and Pei, Jian. Data Mining: Concepts and Techniques 3 rd edition. Elsevier, 2011. This is a great overview book of data mining, from basic concept, data cleaning to algorithms and database. It covers many problems you might face in the real world. |
|----|--|

| | |
|----|--|
| 2. | <p>An Introduction to Statistical Learning with Applications in R, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani.</p> <p>This book provides a great introduction to machine learning algorithms. Concepts are well explained without too much technical details.</p> |
| 3. | <p>AIQ: How People and Machines Are Smarter Together. Nicholas Polson, James Scott. St. Martin's Press 2018.</p> <p>This book contains many interesting stories about how machine learning and data mining influence the business world. It is not a technical book.</p> |
| 4. | <p>Introduction to Data Mining, Tan P N, Steinbach M, Kumar V. Pearson Education. 2016</p> <p>This is another great book about data mining. It focuses more on algorithms.</p> |