City University of Hong Kong
Course Syllabus

offered by College/School/Department of Mathematics
with effect from Semester A 20 22 / 23

Part I Course Overview

Course Title:	**Statistical Modelling for Data Mining**
Course Code:	**MA6633**
Course Duration:	**1 semester**
Credit Units:	**3 CUs**
Level:	**P5**
Medium of Instruction:	**English**
Medium of Assessment:	**English**
Prerequisites: *(Course Code and Title)*	**MA5617 Statistical Data Analysis**
Precursors: *(Course Code and Title)*	**Nil**
Equivalent Courses: *(Course Code and Title)*	**Nil**
Exclusive Courses: *(Course Code and Title)*	**Nil**

Part II Course Details

1. Abstract

This course will aim to provide the student with mathematical foundation and statistical knowledge of data mining and machine learning techniques. Particular focus will be on the fundamental statistical properties and analysis of many popular techniques for learning, classification and prediction. The topics covered in this course will include elements of the following: Bayesian decision theory, model selection, linear Models for regression and classification, Bayesian networks, decision trees, association rule mining, and clustering.

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.	Understand the basic problems and mathematical challenges in data mining	20%	✓	✓	
2.	Ability to use various algorithms for supervised and unsupervised learning and understand their underlying principles	50%	✓	✓	
3.	Ability to choose a suitable combination of different models to explore data sets and solve real problems	30%	✓	✓	✓
		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	**6**	
teaching	Learning through teaching is primarily based on lectures.	✓	✓	✓				3 hours/week
Lab sessions and assignments	Learning through lab assignments allows students to develop hands-on skills of using statistical software to solve real problems	✓	✓	✓				Four 2-hour sessions in total

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	6		
Continuous Assessment: _40___%								
Lab assignments	✓	✓	✓				15%	
Mid-term quiz	✓	✓	✓				25%	
Examination	✓	✓	✓				60%	
Examination: _60___% (duration: 2 hrs, if applicable)								
							100%	

5. Assessment Rubrics

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

Applicable to students admitted in Semester A 2022/23 and thereafter

Assessment Task	Criterion	Excellent (A+, A, A-)	Good (B+, B)	Marginal (B-, C+, C)	Failure (F)
1. Lab assignments	Software usage and data analysis ability	High	Significant	Basic	Not even reaching marginal levels
2. Mid-term quiz	Problem solving based on comprehensive understanding	High	Significant	Basic	Not even reaching marginal levels
3. Examination	Problem solving based on comprehensive understanding	High	Significant	Basic	Not even reaching marginal levels

Applicable to students admitted before Semester A 2022/23

Assessment Task	Criterion	Excellent (A+, A, A-)	Good (B+, B, B-)	Fair (C+, C, C-)	Marginal (D)	Failure (F)
1. Lab assignments	Software usage and data analysis ability	High	Significant	Moderate	Basic	Not even reaching marginal levels
2. Mid-term quiz	Problem solving based on comprehensive understanding	High	Significant	Moderate	Basic	Not even reaching marginal levels
3. Examination	Problem solving based on comprehensive understanding	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.)

Bayesian decision theory; Linear and nonlinear models; Discriminant analysis; Clustering models; Naïve Bayes; Bayesian networks; Decision trees; Association rule.

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.)

1.	Course materials provided
2.	
3.	
…	

2.2 Additional Readings

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

1.	Pattern classification 2nd edition by Duda, Hart and Stork, Wiley-Interscience; 2nd edition
2.	Data Mining: Concepts, Models, Methods, and Algorithms by Mehmed Kantardzic, Wiley-IEEE Press; 2nd edition
3.	Pattern Recognition and Machine Learning by Bishop, Springer, 2006
…	