

**City University of Hong Kong  
Course Syllabus**

**offered by Department of Systems Engineering  
with effect from Semester A 2024 / 25**

**Part I Course Overview**

<b>Course Title:</b>	<u>Data Mining and Statistical Modeling</u>
<b>Course Code:</b>	<u>SYE8012</u>
<b>Course Duration:</b>	<u>One semester</u>
<b>Credit Units:</b>	<u>3</u>
<b>Level:</b>	<u>R8</u>
<b>Medium of Instruction:</b>	<u>English</u>
<b>Medium of Assessment:</b>	<u>English</u>
<b>Prerequisites:</b> <i>(Course Code and Title)</i>	<u>Nil</u>
<b>Precursors:</b> <i>(Course Code and Title)</i>	<u>Basic Probability and Statistics</u>
<b>Equivalent Courses:</b> <i>(Course Code and Title)</i>	<u>SEEM8012 Data Mining and Statistical Modeling(offered until 2021/22)/ ADSE8012 Data Mining and Statistical Modeling (offered until 2023/24)</u>
<b>Exclusive Courses:</b> <i>(Course Code and Title)</i>	<u>Nil</u>

## Part II Course Details

### 1. Abstract

This course focuses on data mining tools and techniques that are useful for a wide range of applications in manufacturing, service, logistics, health and medical, financial and banking, etc. We discuss four basic data mining operation steps: business objective identification, data preparation, knowledge discovery, and consolidation/implementation. We cover both supervised learning and unsupervised learning methods and algorithms, including regression, classification, forecasting, clustering, association rules, and market basket analysis etc. The methods will be illustrated with case studies in credit card fraud detection, telecommunication, express mail service, inventory management, customer relationship management, and bioinformatics.

### 2. Course Intended Learning Outcomes (CILOs)

No.	CILOs	Weighting (if applicable)	Discovery-enriched curriculum related learning outcomes		
			A1	A2	A3
1.	Recognize basic statistical learning, data mining, machine learning, and knowledge discovery and potential applications	15%	✓		
2.	Familiarize the operational steps on data mining and knowledge discovery	15%	✓		
3.	Recognize and apply supervised learning methods and algorithms and their applications.	20%	✓		
4.	Recognize and apply unsupervised learning methods and algorithms and their applications.	20%	✓		
5.	Demonstrate how data mining methods and algorithms can be applied to real life problems in various applications	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. Learning and Teaching Activities (LTAs)

LTA	Brief Description	CILO No.					Hours/week (if applicable)
		1	2	3	4	5	
Lecture	- large class activity - questions and discussion	✓	✓	✓	✓	✓	39 hours/sem

### 4. Assessment Tasks/Activities (ATs)

Assessment Tasks/Activities	CILO No.					Weighting	Remarks
	1	2	3	4	5		
Continuous Assessment: <u>100</u> %							
Group Work	✓	✓	✓	✓	✓	40%	
Individual Coursework	✓	✓	✓	✓		25%	
Test		✓	✓	✓	✓	35%	
Examination: <u>0</u> % (duration: _____, if applicable)						100%	

## 5. Assessment Rubrics

Applicable to students admitted before Semester A 2022/23 and in Semester A 2024/25 & thereafter

Assessment Task	Criterion	Excellent (A+, A, A-)	Good (B+, B, B-)	Fair (C+, C, C-)	Marginal (D)	Failure (F)
1. Group Work	Application of class materials and teamwork	High	Significant	Moderate	Basic	Not even reaching marginal levels
2. Individual Coursework	Application of class materials	High	Significant	Moderate	Basic	Not even reaching marginal levels
3. Test	Understanding of class materials	High	Significant	Moderate	Basic	Not even reaching marginal levels

Applicable to students admitted from Semester A 2022/23 to Summer Term 2024

Assessment Task	Criterion	Excellent (A+, A, A-)	Good (B+, B)	Marginal (B-, C+, C)	Failure (F)
1. Group Work	Application of class materials and teamwork	Excellent	Good	Marginal	Failure
2. Individual Coursework	Application of class materials	Excellent	Good	Marginal	Failure
3. Test	Understanding of class materials	Excellent	Good	Marginal	Failure

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

**1. Keyword Syllabus**

- Introduction to Data Mining
- Data Processing and Data Preparation
- Supervised Learning Methods
- Linear Methods for Prediction/Regression
- Linear Methods for Classification
- Model Assessment and Inferences
- Tree Models and Related Methods
- Neural Networks and SVM
- Forecasting and Time Series Modeling
- Unsupervised Learning Methods
- Clustering and Association Methods
- Data Mining Case Studies

**2. Reading List**

**2.1 Compulsory Readings**

1.	The Elements of Statistical Learning by Hastie, Tibshirani, and Friedman, Springer
2.	Lecture notes

**2.2 Additional Readings**

*NIL*