City University of Hong Kong
Course Syllabus

offered by Department of Computer Science
with effect from Semester A 2024/25

Part I Course Overview

Course Title:	Data Warehousing and Data Mining
Course Code:	CS5483
Course Duration:	One semester
Credit Units:	3 credits
Level:	P5
Medium of Instruction:	English
Medium of Assessment:	English
Prerequisites: *(Course Code and Title)*	CS3402 Database Systems or CS5481 Data Engineering or equivalent
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

The course introduces the subject of automatic/semi-automatic knowledge discovery from data. It gives an overview of the principles and applications of supervised/unsupervised learning, and reinforces the concepts with hands-on experience using some state-of-the-art software. The students will learn to apply different learning algorithms and evaluate them with appropriate statistical performance metrics and techniques that avoid overfitting. Students will also learn how corporations manage their data with dimensional models and OLAP for simple and fast data analysis.

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.	Identify and explain the main characteristics of different data warehousing and data mining techniques through observation of their operations.		✓		
2.	Critically evaluate the strengths and limitations of current data warehousing and data mining techniques.		✓		
3.	Apply the main algorithms in data warehousing and data mining in a computationally efficient way.			✓	
4.	Design new solutions for data warehousing and data mining problems by improving and combining current techniques.			✓	
		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 real-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)

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

LTA	Brief Description	CILO No.				Hours/week (if applicable)
		1	2	3	4	
Lecture	Students will learn data warehousing and data mining techniques, and their applications in different problem domains.	✓	✓			2 hours/ week
Tutorial	Students will work on a set of problems on the principles and applications of data warehousing and data mining, and present their solutions in the class.	✓	✓			1 hour/ week
Project	Students will complete two projects. The first project gives students an opportunity to step through the process of knowledge discovery from data. The second project requires students to gain deeper insights of the existing solutions and encourage them to create new designs.			✓	✓	

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		
Continuous Assessment: 50%						
Project 1 (Implementation of data mining algorithms.)			✓		15%	
Project 2 (Application of data warehousing and data mining techniques to real world problems.)				✓	15%	
Quiz	✓	✓			20%	
Examination^: 50% (duration: 2 hours)						
					100%	

^ For a student to pass the course, at least 30% of the maximum mark for the examination must be obtained.

5. Assessment Rubrics

(Grading of student achievements is based on student performance in assessment tasks/activities with the following 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. Project	1.1 Capacity for effectively implementing data warehousing and data mining algorithms in a computationally efficient way.	High	Significant	Moderate	Basic	Not even reaching marginal levels
	1.2 Capability to create new solutions for data warehousing and data mining problems by improving and combining different techniques.	High	Significant	Moderate	Basic	Not even reaching marginal levels
2. Quiz	2.1 Ability to explain in detail the principles of different data warehousing and data mining techniques	High	Significant	Moderate	Basic	Not even reaching marginal levels
	2.2 Capability to correctly apply suitable techniques to solve data warehousing and data mining problems.	High	Significant	Moderate	Basic	Not even reaching marginal levels
3. Examination	3.1 Capacity for explaining the main characteristics of different data warehousing and data mining techniques in depth.	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. Project	1.1 Capacity for effectively implementing data warehousing and data mining algorithms in a computationally efficient way.	High	Significant	Moderate to Basic	Not even reaching marginal levels
	1.2 Capability to create new solutions for data warehousing and data mining problems by improving and combining different techniques.	High	Significant	Moderate to Basic	Not even reaching marginal levels
2. Quiz	2.1 Ability to explain in detail the principles of different data warehousing and data mining techniques	High	Significant	Moderate to Basic	Not even reaching marginal levels
	2.2 Capability to correctly apply suitable techniques to solve data warehousing and data mining problems.	High	Significant	Moderate to Basic	Not even reaching marginal levels
3. Examination	3.1 Capacity for explaining the main characteristics of different data warehousing and data mining techniques in depth.	High	Significant	Moderate to 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.)

Data Mining

Supervised learning: classification, numerical prediction, decision-trees.

Unsupervised learning: association rule mining, clustering.

Performance evaluation: stratification, cross-validations, overfitting.

Data Warehousing

Dimension modeling, star schema, on-line analytical processing (OLAP)

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.	Han J. and Kamber M. Data Mining: Concepts and Techniques, 3rd Ed., Morgan Kaufmann (2011).
2.	Witten I., Frank E. and Hall M. Data Mining: Practical Machine Learning Tools and Techniques, 3rd Ed., Morgan Kaufmann (2011).

2.2 Additional Readings

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

1.	Hastie T., Tibshirani R. and Friedman J. The Elements of Statistical Learning, 2nd Ed., Springer (2009).
2.	VanderPlas, Jake. Python data science handbook: Essential tools for working with data. " O'Reilly Media, Inc.", 2016.
3.	Kimball R. and Ross M. The Data Warehouse Toolkit, 3rd Ed., Wiley (2013).