Course Title: Data Mining

Course Code: COMP 4331

No. of Credits: 3-credit

Any Pre/co-requisities: ELEC 2600 OR IEDA 2520 OR IEDA 2540 OR ISOM 2500 OR LIFS 3150 OR MATH 2411 OR MATH 2421 OR MATH 2431

Name: Prof. James KWOK

Email: jamesk@cse.ust.hk

Course Description

This course will provide an introduction to concepts and techniques in the field of data mining. Materials include an introduction to data warehousing and OLAP, data preprocessing and the techniques used to explore the large quantities of data for the discovery of predictive models and knowledge. The course will include techniques such as nearest neighbor, decision tress, neural networks, Bayesian networks and Naive Bayes, association analysis and clustering, as well as social networks and data mining applications in business and finance applications, and other emerging data mining subareas. Students learn the materials by attending lectures and implementing and applying different data analysis and mining techniques to large datasets throughout the semester.

List of Topics

- Introduction
- Data Preprocessing
    - data summarization, cleaning, integration, transformation, reduction and discretization
- Classification and Regression
    - decision trees, neural networks, Bayesian classification, nearest-neighbor classification, and various regression methods
- Cluster Analysis
    - partitioning, hierarchical and density-based clustering algorithms
- Association Analysis
    - popular frequent itemset mining algorithms
- Miscellaneous topics

<u>Textbooks</u>

Data Mining: Concepts and Techniques by Jiawei Han and Micheline Kamber

Intended Learning Outcomes (ILOs)

By the end of this course, students should be able to:

-1 Map a practical problem to a data mining problem.
-2 Propose solutions to data mining problems.
-3 Evaluate basic literature in the data mining area.
-4 Apply data mining to new practical problems.

Assessment and Grading

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided below, outlining the criteria used for evaluation.

Assessment Task	Contribution to Overall Course grade (%)
Assignments	25%
Midterm	25%
Final examination	50%

Assessments:

[List specific assessed tasks, exams, quizzes, their weightage, and due dates; perhaps, add a summary table as below, to precede the details for each assessment.]

Assessment Task	Contribution to Overall Course grade (%)	Due date
3 assignments	25%	24/11/2025
Midterm	25%	22/10/2025
Final examination	50%	19/12/2025

* Assessment marks for individual assessed tasks will be released within two weeks of the due date.

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
Assignments	ILO1, ILO2, ILO3. ILO4	This task assesses students' ability to map a practical problem to a data

		mining problem [ILO 1], apply data mining algorithms [ILO 4], propose solutions [ILO 2] and then also evaluate [ILO 3].
Midterm	ILO1. ILO2	This task requires the student to map a practical problem to a data mining problem [ILO 1], and then propose solutions [ILO 2].
Final examination	ILO1. ILO2	This task requires the student to map a practical problem to a data mining problem [ILO 1], and then propose solutions [ILO 2].

Grading Rubrics

1. Data Understanding & Preprocessing (25 points)

- Evaluates the student's ability to handle raw data, identify issues, and prepare it for modeling.

2. Modeling & Algorithm Application (25 points)

- Evaluates the technical execution of applying data mining algorithms and the rationale behind choices.

3. Evaluation & Analysis (25 points)

- Evaluates the depth of the student's analysis of the model's performance and their ability to interpret results.

4. Documentation, Presentation, & reproducibility (25 points)

- Evaluates the clarity, organization, and professionalism of the submitted work.

Final Grade Descriptors:

[As appropriate to the course and aligned with university standards]

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	Demonstrates a comprehensive grasp of subject matter, expertise in problem-solving. Exhibits a high capacity for scholarship, going beyond core requirements to achieve learning goals.]
B	Good Performance	Shows good knowledge and understanding of the main subject matter, competence in problem-solving, and the ability to analyze and evaluate issues. Displays high motivation to learn.
C	Satisfactory Performance	Possesses adequate knowledge of core subject matter, competence in dealing with familiar problems, and some capacity

		for analysis and critical thinking. Shows persistence and effort to achieve broadly defined learning goals.
D	Marginal Pass	Has threshold knowledge of core subject matter, potential to achieve key professional skills, and the ability to make basic judgments. Benefits from the course and has the potential to develop in the discipline.
F	Fail	Demonstrates insufficient understanding of the subject matter and lacks the necessary problem-solving skills. Shows limited ability to think critically or analytically and exhibits minimal effort towards achieving learning goals. Does not meet the threshold requirements for professional practice or development in the discipline.

Course AI Policy

The use of generative artificial intelligence tools to complete assessment tasks is allowed.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Students who have further questions about the feedback including marks should consult the instructor within five working days after the feedback is received.

Resubmission Policy

Resubmission not allowed

Required Texts and Materials

Reference book: Data Mining: Concepts and Techniques by Jiawei Han and Micheline Kamber, Morgan Kaufmann Publishers, 3rd Edition, 2011

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to Academic Integrity | HKUST – Academic Registry for the University's definition of plagiarism and ways to avoid cheating and plagiarism.

Additional Resources

Online course content to be published in HKUST canvas when the semester begins.