The Hong Kong University of Science and Technology

UG Course Syllabus (Spring 2025-26)

[Course Title] Big Data Mining and Management

[Course Code] COMP 4332/ RMBI 4310

[No. of Credits] 3

[Any pre-/co-requisites] COMP 4211 OR COMP 4331 OR ISOM 3360

Name: Yangqiu Song

Email: yqsong@ust.hk

Course Description

This is a project-oriented course. It will expose students to practical issues of large-scale and real-world data mining. Data mining is a process of extracting implicit, previously unknown, and potentially useful knowledge from data, and it is a critical task in many applications. This course will place emphasis on applications of data mining in areas such as business intelligence, which aims to uncover facts and patterns in large volumes of data for decision support. Application areas also include many other areas in science and engineering applications. This course builds on the basic knowledge gained in the introductory data-mining course and explores how to more effectively mine large volumes of real-world data and tap into large quantities of data. It will introduce new algorithms that can more effectively find hidden and profitable data patterns and knowledge. Working on real-world data sets, students will experience all steps of a data-mining project.

Intended Learning Outcomes (ILOs) (*Course Learning Outcomes)*

- Students will understand issues related to real-world data mining.
- Students will master tools and skills for large-scale data mining projects.
- Students will gain experience on recent topics in business intelligence and social media mining.

Course Prerequisites:

- Statistics and Probability
- Machine Learning/Pattern Recognition/Data Mining
- Programming

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 (%)**
Assignment/Projects	50%

Presentations	10%
Final Exam	40%
Total	100

Assessments:

Assessment Task	Contribution to Overall Course grade (%)	Due date
Assignment / Projects	50%	29/05/2025
Presentations	10%	29/05/2025
Final Exam	40%	29/05/2025

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
Assignment / Projects	ILO1, ILO2, ILO3	This task assesses students' ability to address real-world data mining challenges (ILO1), apply tools and techniques to large-scale projects (ILO2), and explore emerging topics in business intelligence or social media mining (ILO3).
Presentations	ILO1, ILO2, ILO3	This task assesses students' ability to articulate practical data mining issues (ILO1), demonstrate technical workflows for scalable solutions (ILO2), and critically analyze recent advancements in the field (ILO3).
Final Exam	ILO1, ILO2, ILO3	This task assesses students' ability to analyze theoretical and applied aspects of real-world data mining (ILO1), solve problems using learned tools and methodologies (ILO2), and reflect on contemporary trends in business intelligence and social media (ILO3).

Grading Rubrics

The project are graded based on clarity of problem introduction, methodological justification, experimental results, and discussions incorporating case studies, error analysis, and unique insights into the research problem.

The presentation grading rubric evaluates **content clarity, method description, experiments/components, insight sharing, and time management** on a 1-5 scale, with final scores averaged across instructor and TAs.

The grading criteria for assignments and final exams will be evaluated on a case-by-case basis.

Final Grade Descriptors:

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	Students demonstrate a deep understanding of real-world data mining challenges, expertly apply advanced tools to large-scale projects, and deliver innovative solutions for business intelligence or social media mining. Their work showcases cutting-edge pattern discovery, robust technical execution, and actionable insights exceeding project goals.
B	Good Performance	Students show strong competence in addressing practical data mining issues, effectively utilize large-scale tools, and explore recent trends with logical analysis. Minor gaps in complexity or originality may exist, but outcomes align well with requirements.
C	Satisfactory Performance	Students grasp basic real-world data mining concepts, apply tools to small-to-medium datasets with occasional errors, and engage superficially with modern topics. Work meets minimal standards but lacks depth, scalability, or persuasive insights.
D	Marginal Pass	Students exhibit fragmented understanding of data mining challenges, struggle technically with large datasets, and inadequately address current applications. Projects are incomplete, lack critical analysis, or rely heavily on template approaches.
F	Fail	Students fail to identify core data mining issues, demonstrate no functional tool proficiency, and ignore contemporary topics. Work is non-functional, irrelevant to real-world contexts, or absent, reflecting no engagement with course objectives.

Course AI Policy

Students may use generative AI tools to assist with project design, coding, or analysis, provided all AI-generated content is properly cited and its contributions are transparently explained in final submissions.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Feedback on assignments will include specific scoring criteria breakdowns, identified strengths, areas needing improvement, and detailed point deductions for each evaluated component. 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

Students may submit assignments multiple times prior to the deadline, with the final submission automatically designated for grading. Late submission will not be accepted except in exceptional circumstances requiring prior approval from the instructor.

Required Texts and Materials

There is no required text book.

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

Text Books:

- Introduction to Data Mining, by Pang-Ning Tan, Michael Steinbach and Vipin Kumar. Addison-Wesley. [DM]
- Community Detection and Mining in Social Media, by Lei Tang and Huan Liu. Morgan & Claypool Publishers.
- Social Network Data Analytics, by Charu C. Aggarwal, Springer
- Introduction to Data Mining with Case Studies, Second Edition, G.K. Gupta. PHI Learning, 2011. [DM-CS]
- Luna Dong, Divesh Srivastava: Big Data Integration. Morgan & Claypool, 2015. [BI]
- Bing Liu, Web Data Mining, Springer, 2011[BI]