

Academic Org: Dept of Computer Sci & Engg – Subject: Computer Science

**Course:** CSCI2750      **Course ID:** 014539      **Eff Date:** 2024-07-01      **Crse Status:** Active      **Apprv. Status:** Approved      [New Course]  
Data Mining and Information Processing 數據挖掘與訊息處理

This course is designed for students to learn the principles of data mining and information processing via a hands-on approach. The first half of the course focuses on the fundamentals of data collection and processing, including the mathematical foundation and the representation of audio and visual data. In the second half of the course, classical data mining methods will be introduced, along with their applications in datasets with meaningful physical interpretations. Students are advised to have a background in programming and a basic understanding of college-level mathematics.

本科旨在讓學生透過實作方法學習數據挖掘和訊息處理的原則。本科的前半部分重點介紹資料收集和處理的基礎知識，包括數學基礎以及音訊和視覺資料的表示。本科的後半部分，將介紹經典的數據挖掘方法，以及它們在具有有意義的物理解釋的資料集中的應用。建議學生有程式設計背景，並對工程數學有基本了解。

**Grade Descriptor:**      A

EXCELLENT – exceptionally good performance and far exceeding expectation in all or most of the course learning outcomes; demonstration of superior understanding of the subject matter, the ability to analyze problems and apply extensive knowledge, and skillful use of concepts and materials to derive proper solutions.

有關等級說明的資料，請參閱英文版本。

B

GOOD – good performance in all course learning outcomes and exceeding expectation in some of them; demonstration of good understanding of the subject matter and the ability to use proper concepts and materials to solve most of the problems encountered.

有關等級說明的資料，請參閱英文版本。

C

FAIR – adequate performance and meeting expectation in all course learning outcomes; demonstration of adequate understanding of the subject matter and the ability to solve simple problems.

有關等級說明的資料，請參閱英文版本。

D

MARGINAL – performance barely meets the expectation in the essential course learning outcomes; demonstration of partial understanding of the subject matter and the ability to solve simple problems.

有關等級說明的資料，請參閱英文版本。

F

FAILURE – performance does not meet the expectation in the essential course learning outcomes; demonstration of serious deficiencies and the need to retake the course.

有關等級說明的資料，請參閱英文版本。

**Equivalent Offering:**

**Units:** 3 (Min) / 3 (Max) / 3 (Acad Progress)

**Grading Basis:** Graded

**Repeat for Credit:** N

**Multiple Enroll:** N

**Course Attributes:**

**Topics:**

### COURSE OUTCOMES

**Learning Outcomes:**

- At the end of the course of studies, students will be able to:
1. understand and explain the ideas of data mining and information processing;
  2. organize raw data into suitable format for further analysis;
  3. analyze data and extract information with traditional techniques and algorithms;
  4. implement some algorithms learned from the course to solve real-world problems.

**Course Syllabus:**

Week 1: Course overview and a glimpse of topics

Week 2: Mathematics refresher  
Week 3: Elementary optimization  
Week 4: Data collection with physical signals  
Week 5: Representation of audio and image data  
Week 6: Blind source separation of audio and image data  
Week 7: The perceptron  
Week 8: Support vector machines  
Week 9: The decision tree and probabilistic classifiers  
Week 10: Pattern mining  
Week 11: Centroid-based clustering  
Week 12: Agglomerative clustering  
Week 13: Advanced topics

<b>Assessment Type:</b>	Examination	: 40%
	Homework or assignment	: 15%
	Lab reports	: 25%
	Test or quiz	: 20%

**Feedback for Evaluation:**

1. Quiz and examinations
2. Course evaluation and questionnaire
3. Question-and-answer sessions during class
4. Student consultation during office hours or online

**Required Readings:**

-

**Recommended Readings:**

- Christopher M. Bishop, "Pattern Recognition and Machine Learning," Springer 2006.
- Simon Haykin, "Neural Networks and Learning Machines", Pearson, 3rd Edition, 2008.
- Stephen Boyd and Lieven Vandenberghe, Introduction to Applied Linear Algebra – Vectors, Matrices, and Least Squares, Cambridge University Press, 2018.
- M. J. Zaki and W. Meirna Jr., Data Mining and Machine Learning: Fundamental Concepts and Algorithm,

Cambridge, 2020.

**OFFERINGS**

1. CSCI2750 Acad Organization=CSD; Acad Career=UG

**COMPONENTS**

LAB : Size=50; Final Exam=N; Contact=2  
LEC : Size=50; Final Exam=Y; Contact=2

**ENROLMENT REQUIREMENTS**

1. CSCI2750

**Enrollment Requirement Group:**

Prerequisite: (ENGG1110 or ESTR1002) and (ENGG1120 or ESTR1005)

**New Enrollment Requirement(s):**

Pre-requisite = (ENGG1110 or ESTR1002) and (ENGG1120 or ESTR1005)

Exclusion = CSCI3230 or CSCI3320

**Additional Information**

eLearning hrs for blended cls 0  
VTL-Onsite face-to-face hrs 0  
VTL-Online synch. hrs 0  
VTL-Online asynch. hrs 0  
No. of micro-modules 0  
Research components (UG) 0%

---

---

<END OF REPORT>