INFSCI 1530 Data Mining

Department of Computer Science, SCUPI

Spring 2025

Instructor

Ziyi Wang
Email: ziyi.wang@scupi.cn
Office: SCUPI Building N527
Office Hours: Friday 12:00-17:00 or by appointment

Teaching Assistants

Section 1: Wenhao Zheng 郑文豪 (email: 2023323045055@stu.scu.edu.cn)
Section 2: Yuyao Tang 唐宇瑶 (email: 1017170093@qq.com)

Lectures

Section 1:
Thursday 8:15-11:00
Location: SCUPI Building N212

Section 2:
Thursday 13:50-16:25
Location: SCUPI Building N212

Description

The ability to collect, store and process large amounts of detailed data in a variety of fields has led to a surge in the use of data in various decision-making tasks, ranging from governmental policy-making to drafting players in sports. Data literacy is thus important and in this first introductory course, we will focus on shifting the traditional mode of deterministic (yes/no) thinking to probabilistic thinking. We will see concepts from applied probability and statistics, while we will also explore a variety of data analysis methods including linear regression, matrix factorization, and network analysis.

Prerequisites

- INFSCI 0510 Data Analysis
- MATH 0280 Matrices and Linear Algebra

Course Objectives

- Explain the terminology and concepts of data mining
- Evaluate the possibilities and fundamental limitations of data mining
- Evaluate the relative advantages and disadvantages of major approaches to data mining
- Demonstrate the basic methods and techniques used in data mining
- Apply the basic methods and techniques to actual problems in data mining

Applicable ABET Outcomes

- Analyze a complex computing problem and apply principles of computing and other relevant disciplines to identify solutions.
- Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline.

- Communicate effectively in a variety of professional contexts.
- Recognize professional responsibilities and make informed judgments in computing practice based on legal and ethical principles.
- Function effectively as a member or leader of a team engaged in activities appropriate to the program's discipline.
- Apply computer science theory and software development fundamentals to produce computing-based solutions.

Textbook

The following textbooks are used as reference:

- Jiawei Han, Micheline Kamber, Jian Pei. *Data Mining: Concepts and Techniques, 3rd Edition*. Morgan Kaufmann Publishers / Elsevier.
- Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar. *Introduction to Data Mining, 2nd Edition*. Pearson.
- Charu C. Aggarwal. *Data Mining: The Textbook*. Springer.

Grading

Participation	5%
Homework	45%
Final Exam	50%

Conversion from Numerical Score to Letter Grade

100~90	89~85	84~80	79~76	75~73	72~70	69~66	65~63	62~61	60	<60
A	A–	B+	B	B–	C+	C	C–	D+	D	F

Communication

All lecture notes, assignments, and announcements will be published on Blackboard (https://pibb.scu.edu.cn/). Announcements and notifications for update on Blackboard will be sent to QQ / email group. It is the student's responsibility to regularly check Blackboard in a timely manner.

Class Policy

Participation

Class participation is expected and takes a share in your final grade. If a student has a valid reason to be absent from a class session, please notify the instructor beforehand.

Homework Assignment

The homework assignments of this course are semester-long project-like homework. Students are required to schedule weekly online meeting with the TAs to report their progress on the homework assignments. The TAs have the responsibility to supervise and provide guidance to the students.

Exam

There will be only one final exam. The exam will be CLOSED BOOK, CLOSED NOTES,

and CLOSED COMPUTER.

Academic Integrity

The principles of academic integrity require that a student to make sure that all work submitted is the student's own and created without the aid of impermissible technologies, materials, or collaborations. Academic integrity policy will be strictly followed.

Classroom Recording

To ensure the free and open discussion of ideas, students may not record classroom lectures, discussion and/or activities without the advance written permission of the instructor, and any such recording properly approved in advance can be used solely for the student's own private use.

Tentative Course Schedule

Week	Date	Topic
1	Feb 24	Introduction
2	Mar 3	Data Preparation
3	Mar 10	Regression
4	Mar 17	Classification
5	Mar 24	Cluster Analysis
6	Mar 31	Dimension Reduction
7	Apr 7	Anomaly Detection
8	Apr 14	Mid-term Homework Presentation
9	Apr 21	Association Rule Learning
10	Apr 28	Sequence Mining
11	May 5	Text Mining
12	May 12	Network Mining
13	May 19	Web Mining
14	May 26	Recommendation
15	Jun 2	Final Homework Presentation
16	Jun 9	Final Homework Presentation
17	Jun 16	Final Exam