Welcome to C S 488/508

Introduction to Data Mining

Tuan Le

College of Arts and Sciences | Computer Science

BE BOLD. Shape the Future.

About the Course

Course	Introduction to Data Mining
Prerequisite:	At least a C- in C S 272 and C S 278.
Credits:	3
Class Timing:	01:30 PM – 02:45 PM Monday & Wednesday
	Location: Jett Hall 109
	First Class: August 21, 2024
Important dates:	https://records.nmsu.edu/students/important-dates.html

Course Homepage

- https://nmsu.instructure.com/courses/1620462
- Schedule
- Announcements
- Slides
- Assignments
- Quizzes
- Team Project
- …

Teaching Team

Instructor:	Dr. Tuan Le
Office Hours:	10:30 AM – 12:00 PM Tuesday or by appointment
Meeting Zoom Link:	https://nmsu.zoom.us/j/4048892636 or SH 149
Email:	tuanle@nmsu.edu
Teaching Assistant:	Phuong Nguyen
Meeting Zoom Link:	https://nmsu.zoom.us/j/4661464430
Office Hours:	01:30 PM – 02:30 PM Tuesday or by appointment
Email:	ntphuong@nmsu.edu

Syllabus

Course Description

- This course introduces techniques for exploring large data sets and discovering patterns in them.
- Topics covered:
  - Data, Proximity
  - Classification
  - Clustering
  - Association Analysis
  - Anomaly Detection
  - Avoid False Discoveries

Textbook

Recommended textbook:

Introduction to Data Mining (2nd Edition) by Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar.
(https://www-users.cs.umn.edu/~kumar001/dmbook/index.php)

References:

- Data Mining: The Textbook by Charu C. Aggarwal
  http://www.charuaggarwal.net/Data-Mining.htm
- Data Mining: Concepts and Techniques (3rd Edition) by Jiawei Han, Micheline Kamber, and Jian Pei
  https://hanj.cs.illinois.edu/bk3/
- Pattern Recognition and Machine Learning, by Christopher M. Bishop
  https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf

Grading Item

Grading Item	%Weight
1. Attendance	5%
2. Quizzes	10%
3. Assignments	25%
4. Team Project	
4.1 Project proposal	5%
4.2 Midterm report	5%
4.3 Peer evaluated weighted grade (Final Presentation/Report)	15%
5. Final Exam	35%

Quizzes

- Expect a quiz every one or two weeks.
- Each quiz may contain 10-20 multiple-choice questions.
- Quizzes will be conducted on Canvas.
- No second attempt will be allowed on a missed quiz.
- %Weight: 10%.

Assignments

- 5 individual assignments are expected.
- No late submissions.
- %Weight: 25%.

Team Project

- Group: 4 people.
- Goal: solve some real-world data mining problems.
- You are expected to:
  - find an interesting real-world data mining problem.
  - figure out what data mining tasks needed to perform.
  - submit midterm and final project reports, and source code.
  - present your project.
- %Weight: 25%.

- **Team formation due: 09/04/2024, 11:55 PM**
- **Project Proposal due: 09/25/2024, 11:55 PM**
- More details later: example projects, report format and content, presentation, resources and references, due dates,…

Final Exam

- Closed book, closed notes, closed internet.
- You can bring one A4 size cheat sheet (one-side or both-sided, hand-written or printed).
- Calculator is allowed.
- Cell phone is NOT allowed.
- Blank papers/pencil/pen.
- %Weight: 35%.
- **1:00 PM – 3:00 PM, Monday (December 9, 2024).**

Course Objectives

1. Explain and recognize different data mining tasks such as data pre-processing, visualization, classification, regression, clustering, association rules, and anomaly detection.
2. Apply classical data mining / machine learning algorithms for classification, clustering, association rules, and anomaly detection.
3. Evaluate and compare the performance of different data mining/machine learning algorithms.
4. Utilize data mining algorithms to analyze data in real applications using a data mining tool.