CSCE 474/874: Introduction to Data Mining
Spring 2021

General Information

Instructor: Ashok Samal
Avery Hall, Room 359
Phone: 472-2217
Email: samal@cse.unl.edu
Office Hours: TR (12:15 – 1:15)

Teaching Assistant: Atharva Tendle (atharva.tendle@huskers.unl.edu).
Office Hours: M/T 10:00 am – 11:00 am (https://unl.zoom.us/j/96441158653)

Lectures: TR 11:00 am – 12:15 pm (Web Conferencing)

Text: Introduction to Data Mining (Second Edition) by Tan, Steinbach, Karpatne, and Kumar, Pearson.

Prerequisites: CSCE310, STAT 380 or permission or equivalent.

1. **Mastery** of a high-level programming language.
2. **Familiarity** with calculus and statistics.
3. **Exposure** to linear algebra.

Course Objectives: Data mining is the process of automatically finding useful *hidden* patterns in data. The objective of data mining is to use discovered patterns to help *explain* current behaviors or to *predict* the future outcomes. Upon completing this course, students will understand:

- **Mastery** of basic data mining algorithms, including association rule mining, clustering, prediction and classification.
- **Familiarity** with data warehousing and its applications.
- **Familiarity** with role of data mining in knowledge discovery.
- **Familiarity** with data preprocessing and reduction techniques.
- **Exposure** to applications of data mining to practical problems.

Communication

To keep our inboxes clean and to ensure emails don't get lost, all communication should be done using Piazza. Piazza is an online communication tool where instructors can post notes, and students can ask questions. The Piazza site for this course is located at piazza.com/unl/spring2021/csce474874. You should have received an invitation to join. If you have not, you may sign yourself up at: piazza.com/unl/spring2021/csce474874.

Again, any general or homework questions you might have, or if you have questions about a graded assignment, please use Piazza. When you post a question, the instructor and the TA will see your question and respond as needed. You are encouraged to check-in and respond to the questions posed by other students. Extra credit will be given to students who actively participate in this activity, particularly to students who answer the questions posed by a student.
Please remember to use professional language and letter salutations in your emails to the instructor and the TA (no "Hey" or any kind of text jargons).

Assignments

There will be four assignments for this course. Together they will count for 40% of the overall grade. The assignments are due at the class time on the indicated dates. We will use a flexible, slip date system for the assignments. Each student is given an automatic extension of four (4) calendar days. You can use these on any assignment(s) during the course. However, the total number of late days cannot exceed 4. Slack days cannot be used for the project. After you have used your ``late'' days, if you submit an assignment late, you will not get any credit for it. Then, it is better to submit a partially completed homework than a late one. If you have a special reason for being late, get permission well ahead of the due date.

The assignments will be done in pairs and your partner will be randomly assigned and will change for each assignment.

You are required to submit the assignment online on Canvas. Do not turn in any hard copies. You are required to submit all source code, executable, report, etc.

While it can be informative to discuss the assignments with others, you should write your final programs independently. Make sure that all your files are protected. You are also responsible if somebody copies your files and hands them in.

Attendance

Attendance in class is required. 5% of the overall grade is reserved for class attendance and participation.

Tests and Quizzes

There will be no exams in this class. However, there will be periodic quizzes (both announced and unannounced). 20% of the overall grade is devoted to quizzes.

ICDM/KDD Session Summary and Presentation

Each student will be required to study a paper (from a selected set) presented either at IEEE International Conference on Data Mining (ICDM) or ACM Conference on Knowledge

Discovery and Data Mining (KDD), two leading conferences in data mining and present a summary to the class. This will be worth 5% of the overall grade.

Group Project

Students are required to work on a group project on some aspect of data mining. You must select a project from a list assigned in the class. Each team should consist of three to four students. All project documents will be kept online in Blackboard. The group project that will be worth 30% of the overall grade.

Students with Disabilities

Students with disabilities, including mental health disabilities like depression and anxiety, are encouraged to contact the instructor for a confidential discussion of their individual academic accommodation needs. The University of Nebraska-Lincoln provides individualized accommodations to students with documented disabilities that may affect their ability to participate in course activities fully or meet course requirements. Students must register with the Services for Students with Disabilities (SSD) office, 232 Canfield Administration, 472-3787 voice or TTY to receive the accommodations.

UNL CSE Department Policies

The CSE Department expects all students in CSE courses to check their email regularly, so they do not miss important announcements.

The CSE Department's Student Resource Center is located in Avery 12: http://cse.unl.edu/src. The office hours by the TAs are also listed there.

The CSE Department has an anonymous contact form that you may use to voice your concerns about any problems in the course or department if you do not wish to be identified.

Academic Misconduct

All homework assignments, quizzes, exams, etc. must be your own work. No direct collaboration with fellow students, past or current, is allowed unless otherwise stated. The Computer Science & Engineering department has an Academic Integrity Policy. All students enrolled in any computer science course are bound by this policy. You are expected to read, understand, and follow this policy. Violations will be dealt with on a case by case basis and may result in a failing assignment or a failing grade for the course itself.

Those who share their code, and those who copy other's code will be penalized in the same way; both parties will be considered to have plagiarized.

The penalty for the violations are as follows:

- 1st violation — award a zero in homework or exam
- 2nd violation — award an F in the course
- 3rd violation — recommend the student be dismissed from the program

---

Final Grade

The distribution of points for the course is as follows.

Category	Total Points
Participation	50
ICDM/KDD Presentation	50
Quizzes	200
Homework	400
Project	300
Total	**1000**

The final grade will be based on the following scale.

≥ 95%: A	**≥ 87%: B+**	**≥ 77%: C+**	**≥ 67%: D+**	
≥ 90%: A-	**≥ 84%: B**	**≥ 74%: C**	**≥ 64%: D**	**< 60%: F**
	≥ 80%: B-	**≥ 70%: C-**	**≥ 60%: D-**	
