# CS 412 Introduction to Data Mining

## Course Description

This course is an introductory course on data mining. It introduces the basic concepts, principles, methods, implementation techniques, and applications of data mining, with a focus on (1) data preprocessing and preparation, (2) classification, (3) pattern discovery, and (4) cluster analysis.

In the first part of the course, which focuses on foundational topics in data analysis and pattern discovery, you will learn about data types, measurements, basic statistics, similarity and distance measures, data quality and preprocessing techniques, and dimensionality reduction). After that, you will learn why pattern discovery is important, the major tricks for efficient pattern mining, and how to apply pattern discovery in some exciting applications. The course provides you the opportunity to learn concepts, principles, and skills to practice and engage in scalable pattern discovery methods on massive data; discuss pattern evaluation measures; study methods for mining diverse kinds of frequent patterns, sequential patterns, and sub-graph patterns; and study constraint-based pattern mining, and explore their applications.

In the second part of the course, which focuses on classification methods, you will learn concepts and methodologies for classification, also known as supervised learning. We will introduce the basic concepts of classification and then study a set of typical classification methods, algorithms, and applications. This includes decision tree induction, Bayes classification methods, linear classifier, model evaluation and selection, ensemble methods, and additional concepts on classification such as multi-class classification, semi-supervised classification, active learning, and transfer learning. Moreover, we introduce a few advanced classification methods, including the Bayesian belief network, support vector machine (SVM), neural network and deep learning, pattern-based classification, k-nearest neighbors (KNN), and genetic algorithm. Programming assignments will enhance the learning.

In the third part of the course, which focuses on cluster analysis, you will learn concepts and methodologies for cluster analysis, also known as clustering, data segmentation, or unsupervised learning. We will introduce the basic concepts of cluster analysis and then study a set of typical clustering methodologies, algorithms, and applications. This includes partitioning methods, such as k-means, hierarchical methods, such as BIRCH, density-based methods, such as DBSCAN, and grid-based methods, such as CLIQUE. We will also discuss

methods for clustering validation. The learning will be enhanced by clustering software and programming assignments.

## Instructor

Ruby Tahboub
Email: rubyt@illinois.edu
Office: SC 2334

## Sections

This syllabus is intended for Online Master of Computer Science (MCS) offering

## Textbook and Readings

The lectures are designed to be self-contained and reference the 4th edition of the textbook Data Mining: Concepts and Techniques (4th ed.).

Publisher Waltham: Morgan Kaufmann.

## Elements of This Course

The course is comprised of the following elements:

- **Lecture Videos.** The concepts you need to know will be presented each week through short video lectures. You may stream these videos for playback within the browser by clicking on their titles or downloading them. You may also download the slides and videos.

- **In-video questions**. Some lecture videos have questions to help verify your understanding of the topics. These questions will automatically appear while watching the video if you stream the video through your browser. These questions **do not** contribute toward your final score in the class.

- **Lesson Quizzes.** Each week may contain one or multiple lessons. A lesson is a series of videos on a topic, concluding with a lesson quiz. Quiz attempts are unlimited, and the deadline for all quizzes is the last day of class. There is no time limit on how long you take to complete each attempt at the quiz. Each attempt may present a different selection of questions to you. Your highest score will be used when calculating your final score in the class.

- **Assignments.** This course has 7 programming assignments- two are designed around pattern discovery, three on cluster analysis, and two on classification. Please read the instructions for the respective weeks for more information about the assignments.

  **Proctored Exams.** There are **three proctored exams** on each topic. All exams will be proctored via a proctoring service called ProctorU.

## Grading Distribution

| Assignment Type | Occurrence | Weight (%) |
|---|---|---|
| Weekly Quizzes + Orientation Quiz | 25 + 1 | 19.49% in total |
| Programming Assignments (or MP) | 7 | 35% (5% each) |
| Course Part 1 Exam (Pattern Discovery) | 1 | 15.17% |
| Course Part 2 Exam (Cluster Analysis) | 1 | 15.17% |
| Course Part 3 Exam (Classification) | 1 | 15.17% |
| **Total** | | **100** |

## Grading Scale

| Letter Grade | Percent Needed | Letter Grade | Percent Needed |
|---|---|---|---|
| A+ | $\geq 95\%$ | B- | $70\% \leq x < 75\%$ |
| A | $90\% \leq x < 95\%$ | C | $65\% \leq x < 70\%$ |
| A- | $85\% \leq x < 90\%$ | D | $60\% \leq x < 65\%$ |
| B+ | $80\% \leq x < 85\%$ | F | $x < 60$ |
| B | $75\% \leq x < 80\%$ | | |

## Campus Policies

The University of Illinois at Urbana-Champaign is committed to providing a safe and welcoming campus environment free from discrimination based on sex, which includes sexual assault, sexual exploitation, stalking, sexual harassment, dating violence, and domestic violence (collectively referred to as sexual misconduct). See Campus Policies and ProceduresOpens for more information.

## Student Code

A student at the University of Illinois at the Urbana-Champaign campus is a member of a University community of which all members have at least the rights and responsibilities common to all citizens, free from institutional censorship; affiliation with the University as a student does not diminish the rights or responsibilities held by a student or any other community member as a citizen of larger communities of the state, the nation, and the world. See the University of Illinois Student Code for more information.

## Academic Integrity

It is the responsibility of each student to refrain from infractions of academic integrity, from conduct that may lead to suspicion of such infractions, and from conduct that aids others in such infractions. Students have been given notice of this Part by its publication. Regardless of whether a student has read this Part, a student is charged with knowledge of it. Ignorance is not a defense. For more information about academic integrity, see Article 1 – Student Rights and Responsibilities or visit the Students' Quick Reference Guide to Academic Integrity.

Academic integrity violations may result in dismissal from the University.

## AI Use Policy

1. **Permitted Uses**

- Assisting with brainstorming ideas and outlining assignments.

- Improving grammar, style, or formatting of written work.

- Debugging or generating code snippets, provided proper understanding is demonstrated.

- Exploring general concepts or receiving explanations related to course content.

2. **Restrictions**

- AI tools must not be used to produce work that is submitted without substantial modification and understanding by the student.

- Directly copying answers, solutions, or essays generated by AI tools is considered a violation of academic integrity.

- Students must cite AI tools if their contributions are significant (e.g., "Generated suggestions using ChatGPT on [date]").

3. **Accountability**

- Students are responsible for verifying the accuracy, relevance, and originality of AI-generated content.

- Misuse of AI tools to evade learning, misrepresent understanding, or commit plagiarism will result in academic penalties as outlined in the university's academic integrity policy.

# Mental Health

Diminished mental health, including significant stress, mood changes, excessive worry, substance/alcohol abuse, or problems with eating and/or sleeping, can interfere with optimal academic performance, social development, and emotional well-being. The University of Illinois offers various confidential services, including individual and group counseling, crisis intervention, psychiatric services, and specialized screenings at no additional cost. If you or someone you know experiences any of the above mental health concerns, it is strongly encouraged to contact or visit any of the University's resources provided below. Getting help is a smart and courageous thing to do – for yourself and for those who care about you.

Counseling Center: 217-333-3704, 610 East John Street, Champaign, IL 61820

McKinley Health Center:217-333-2700, 1109 South Lincoln Avenue, Urbana, Illinois 61801

## Sexual Misconduct Reporting Obligation

The University of Illinois is committed to combating sexual misconduct. Faculty and staff members are required to report any instances of sexual misconduct to the University's Title IX Office. In turn, an individual with the Title IX Office will provide information about rights and options, including accommodations, support services, the campus disciplinary process, and law enforcement options.

A list of the designated University employees who, as counselors, confidential advisors, and medical professionals, do not have this reporting responsibility and can maintain confidentiality can be found here.

Other information about resources and reporting is available here.

## Academic Integrity

The University of Illinois at Urbana-Champaign Student Code should also be considered as a part of this syllabus. Students should pay particular attention to Article 1, Part 4: Academic Integrity.

Academic dishonesty may result in a failing grade. Every student must review and abide by the Academic Integrity Policy. Ignorance is not an excuse for any academic dishonesty. It is your responsibility to read this policy to avoid any misunderstanding. Do not hesitate to ask the instructor(s) if you are ever in doubt about what constitutes plagiarism, cheating, or any other breach of academic integrity.

## Religious Observances

Illinois law requires the University to reasonably accommodate its students' religious beliefs, observances, and practices in regard to admissions, class attendance, and the scheduling of examinations and work requirements. You should examine this syllabus at the beginning of the semester for potential conflicts between course deadlines and any of your religious observances. If a conflict exists, you should notify your instructor.

## Disability-Related Accommodations

To obtain disability-related academic adjustments and/or auxiliary aids, students with disabilities must contact the course instructor and the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES, you may visit 1207 S. Oak St., Champaign, call 333-4603, email disability@illinois.edu or go to https://www.disability.illinois.edu

If you are concerned you have a disability-related condition that is impacting your academic progress; there are academic screening appointments available that can help diagnose a previously undiagnosed disability. You may access these by visiting the DRES website and selecting "Request an Academic Screening" at the bottom of the page.

## Family Educational Rights and Privacy Act (FERPA)

Any student who has suppressed their directory information pursuant to the Family Educational Rights and Privacy Act (FERPA) should self-identify with the instructor to ensure the protection of the privacy of their attendance in this course. See https://registrar.illinois.edu/academic-records/ferpa/for more information on FERPA.