CSE 572: Data Mining

Goal: This course will introduce fundamental concepts and techniques in data mining including classification, clustering, dimensionality reduction, and outlier detection. Students will learn the theory behind topics as well as gain hands-on experience implementing data mining techniques and applying them to real world problems and data. Students will learn how data mining is used in research and gain understanding and practice of the complete research process.

Prerequisites: Students are expected to have a working knowledge of basic probability theory, linear algebra, and the academic research process. In addition, proficiency in Python programming is required for the labs, homeworks, and final project.

Course information

Course topics: classification, clustering, dimensionality reduction, anomaly detection, deep learning, real world applications and considerations, and more

Semester: Spring 2023

Meeting times: Tu 10:30-11:45 am

Instructor: Dr. Hua Wei

Office: BYENG 486

Office hours: Tuesday 9:20-10:20 am, 11:45am-12:45PM (email if appointment needed outside of office hours)

Email: hua.wei@asu.edu

Location: PSH 153(ASU interactive map)

TA: Sandipan Choudhuri

TA email: schoud13@asu.edu

TA office hours: Mondays and Fridays from 3:00 pm to 4:00 pm (email if appointment needed outside of office hours)

TA office: https://asu.zoom.us/j/9077535235 (Before a physical room is finalized)

Website: Canvas (https://canvas.asu.edu/courses/162579)

Textbook

Recommended: Jiawei Han, Micheline Kamber and Jian Pei, Data Mining: Concepts and Techniques, 3rd ed. (book website).

Recommended (optional):

- Pattern Recognition and Machine Learning. Chris Bishop, 2006 (book website).
- Introduction to Data Mining (2nd edition). Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar, 2019 (book website).

Learning outcomes

By the end of this course, students will:

- Understand the history of data mining, key concepts/techniques, and real-world applications of data mining
- Understand, define, and compare/contrast approaches for classification
- Know techniques for preparing/preprocessing datasets, selecting the best model (including tuning model hyperparameters), and evaluate models using appropriate metrics
- Understand, define, and compare/contrast approaches for clustering data
- Understand, define, and compare/contrast approaches for dimensionality reduction
- Understand a wide range of advanced data mining techniques on different data as well as special considerations for real-world problems (e.g., class imbalance, noisy data, etc)

Course structure

Lectures will consist of a combination of **presentations** and **demos** by the instructor on course topics and hands-on programming **labs** in which students have the opportunity to practice using the concepts and methods taught in class. Students will have time to complete the labs in class, which will be due at the end of each class. This means that students are required to bring their laptop to class and ensure it is fully charged and they are ready to do assigned work in class.

There will also be assignments to be completed outside of class. There will be four **homework** assignments in which students will practice applying the concepts learned in class independently. There will also be required reading (textbook or research papers), discussion, peer review, or other assignments throughout the course.

Software and computer requirements

We will use Colab notebooks for all in-class labs and homework assignments (Chrome web browser and Google Drive account required). Students may use any IDE or environment they choose to develop their final project.

Student success

To be successful in this course, students should check the course page and announcements in Canvas daily, read all announcements and course messages, and complete assignments by the due date. Communicate regularly with the TA, instructor, and fellow students to ensure you have a good understanding of the course content; if you need help, visit the instructor or TA during office hours. You may choose to create a study and/or assignment schedule for yourself to stay on track. If you are reading the syllabus diligently as instructed and find this sentence, send me an email with the subject line “student success”. The first student to do so will receive additional credit. Review the ASU Student Resources page for more resources to help you succeed.

Prerequisites

Programming, linear algebra, probability, algorithm analysis, data structure.

Note: Assignments and projects should be implemented in Python.

Tentative schedule/objectives

Week	Unit
Week 2 - 3	Knowing your data Data exploration, visualization, normalization, feature extraction, etc.
Week 4 - 8	Supervised Learning logistic regression, decision trees, random forest, neural networks/deep learning, ensembles
Week 9	Fall Break (no class)
Week 10 - 12	Unsupervised Learning Clustering, Dimensionality reduction, Generative models
Week 13	Advanced topic: Text Mining & Representation learning
Week 14	Advanced topic: Reinforcement Learning
Week 15-16	Final project presentations

*Topics might be subject to changes

Grading

Grades reflect your performance on assignments and adherence to deadlines. Below is a breakdown of assignments, their assigned weights, and a brief description of each.

- Participation (5%)
- Homework assignments and quiz (35% in total, Quiz 5%, Assignment 1 10%, Assignment 2 10%, Assignment 3 10%)
- In-class lab assignments (20%)
- Final project total (40%)
    - Final project proposal (5%)
    - Final project literature review (5%)
    - Final project progress report (5%)
    - Final project presentation (15%)
    - Final project written report (10%)

Your grade will be determined based on the following grading scheme:

Grade	Percentage	Grade	Percentage
A+	100-97%	B-	<84-80%
A	<97-94%	C+	<80-77%
A-	<94-90%	C	<77-70%
B+	<90-87%	D	<70-60%
B	<87-84%	E	<60%

Absences and late or missed assignments

If the student must be absent from class, the student is responsible for obtaining lecture notes and the lab assignment from the instructor, TA, or Canvas and meeting assignment deadlines as posted.Excused absences for classes will be given without penalty to the grade in the case of (1) a university-sanctioned event [ACD 304-02]; (2) religious holidays [ACD 304-04]; a list of religious holidays can be found here https://eoss.asu.edu/cora/holidays]; (3) work performed in the line-of-duty according

[SSM 201-18]. Students who request an excused absences must follow the policy/procedure guidelines. Excused absences do not relieve students of responsibility for any part of the course work required during the period of absence.

Late assignments will not be accepted unless there are extenuating circumstances for which the student requires an extension. Notify the instructor **BEFORE** an assignment is due if an urgent situation arises and the assignment will not be submitted on time. Published assignment due dates are firm.

Please follow the appropriate University policies to request an accommodation for religious practices or to accommodate a missed assignment due to University-sanctioned activities.

End-user technology issues are not an adequate excuse for late submission or missing assignments in this course. Preparation should be made to ensure all course work can be completed including a back-up plan should your computer, internet connection, or other technology fail. It is recommended that you backup all of your work using Github, a portable USB drive, an external drive, or a cloud service such as Dropbox, Google Drive, or iCloud.

Communicating with instructor and TA

We have multiple ways for you to communicate with instructors and TAs. In general, the preferred order to interact with them is through Community Forum > Office Hours > Emails.

Community Forum

This course uses a discussion topic called "Community Forum" for general questions and comments about the course. Prior to posting a question or comment, check the syllabus, announcements, and existing posts to ensure it's not redundant. You are encouraged to respond to the questions of your classmates. Usually the TA would reply to your questions within two business days.

Email Questions

Only if you do not receive a response on the Community Forum or you feel your question is not appropriate for the public form (e.g., it is of a personal nature), you may email the TA or instructor. Send your email first to the TA. If you do not receive an answer after 24 hours or the TA is unable to answer your question, then you may email the instructor. When you email the instructor, show that you have already tried or considered 1) posting your question in the forum and 2) emailing the TA.

Chat

The Chat tool in Canvas allows students and teachers to interact in real time. Use Chat only for informal course-related conversations unless your instructor informs you otherwise. Chat is not ideal for questions about assignments; instructors are not required to monitor it and conversations may be buried or lost.

Email

ASU email is an official means of communication among students, faculty, and staff. Students are expected to read and act upon an email in a timely fashion. Students bear the responsibility of missed messages and should check their ASU-assigned email regularly. All instructor correspondence will be sent to your ASU email account.

Academic integrity

Policy regarding expected student behavior: Students in this class are expected to acknowledge and embrace the FSE student professionalism expectation located at: https://engineering.asu.edu/professionalism/

Students in this class must adhere to ASU's academic integrity policy, which can be found at https://provost.asu.edu/academic-integrity/policy). **Students are responsible for reviewing this policy and understanding each of the areas in which academic dishonesty can occur.** In addition, all engineering students are expected to adhere to the ASU Academic Integrity Honor Code. **All academic integrity violations will be reported to the Fulton Schools of Engineering Academic Integrity Office (AIO).** The AIO maintains a record of all violations and has access to academic integrity violations committed in all other ASU colleges/schools.

Specific rules for this class are as follows. All assignments and projects must be your own individual work unless specified as team efforts. You are encouraged to learn from each other but copying is strongly discouraged. All solutions turned in for credit are to be your individual work and should demonstrate your problem-solving skills. The instructor reserves the right to question a student orally or in writing and to use their evaluation of the student's understanding of the assignment and of the submitted solution as evidence of cheating. **Violators of this policy will be faced with severe penalties, which may range from deducted points to failure of the course.**

Policies of using Generative AI

Our class of Data Mining in SCAI at ASU is driven by an innovative mindset. We believe in embracing cutting-edge technologies and teaching students the ethical responsibilities associated with data mining and artificial intelligence (AI). By doing so, we aim to enhance the quality of education and remove any barriers that may exist for

students interested in the field of data mining. With the incorporation of generative AI into our curriculum, we are preparing our students for successful careers across various disciplines within computer science and engineering.

Therefore, in our class, we allow students to use generative AI for assignments, reports and presentations, as long as the students' clear note on using generative AI and provide the **full trace of interactions** with generative AI.

- For chatGPT, export your conversations into a html file.
- For Bard or others, please contact TA for detailed submissions.

While we accept contents with the help of generative AI, the final grades will be based on the quality of the code/report.

Below are some examples of plausible uses of generative AI:

- as a grammar tool
- as a dictionary/thesaurus
- improve sentences or for ideas, but not for substantial content creation
- as a document editing tool for suggested recommendations, but not substantial content changes
- assist with properly drafting citations using APA, MLA, etc.
- help as a search tool for reputable sources
- assist with data collection, but not draft final reporting documentation

Below are possible–but not exhaustive–items uses of generative AI that may costs you longer time or leads to lower grades:

- Utilizing AI to create the research question and generate an answer to the question
- Generating Sections, chapters or appendices without specific indications on each paragraph.
- Generating codes without curated information (input, output and the description of your algorithm) for your specific task

Harassment and Sexual Discrimination

Arizona State University is committed to providing an environment free of discrimination, harassment, or retaliation for the entire university community, including all students, faculty members, staff employees, and guests. ASU expressly prohibits discrimination, harassment, and retaliation by employees, students, contractors, or agents of the

university based on any protected status: race, color, religion, sex, national origin, age, disability, veteran status, sexual orientation, gender identity, and genetic information.

Title IX is a federal law that provides that no person be excluded on the basis of sex from participation in, be denied benefits of, or be subjected to discrimination under any education program or activity. Both Title IX and university policy make clear that sexual violence and harassment based on sex is prohibited. An individual who believes they have been subjected to sexual violence or harassed on the basis of sex can seek support, including counseling and academic support, from the university. If you or someone you know has been harassed on the basis of sex or sexually assaulted, you can find information and resources at https://sexualviolenceprevention.asu.edu/faqs.

As a mandated reporter, I am obligated to report any information I become aware of regarding alleged acts of sexual discrimination, including sexual violence and dating violence. ASU Counseling Services, https://eoss.asu.edu/counseling is available if you wish to discuss any concerns confidentially and privately. ASU online students may access 360 Life Services, https://goto.asuonline.asu.edu/success/online-resources.html.

Disability accommodations

Suitable accommodations are made for students having disabilities. Students needing accommodations must register with the ASU Disabilities Resource Center and provide documentation of that registration to the instructor. Students should communicate the need for an accommodation in enough time for it to be properly arranged. See ACD 304-08 Classroom and Testing Accommodations for Students with Disabilities.

Student copyright responsibilities

You must refrain from uploading to any course shell, discussion board, or website used by the course instructor or other course forum, material that is not the student's original work, unless the student first complies with all applicable copyright laws; faculty members reserve the right to delete materials on the grounds of suspected copyright infringement.

The contents of this course, including lectures and other instructional materials, are copyrighted materials. Students may not share outside the class, including uploading, selling or distributing course content or notes taken during the conduct of the course. Any recording of class sessions is authorized only for the use of students enrolled in this course during their enrollment in this course. Recordings and excerpts of recordings may not be distributed to others. (see ACD 304–06, "Commercial Note Taking Services" and ABOR Policy 5-308 F.14 for more information).

Policy against threatening behavior

Students, faculty, staff, and other individuals do not have an unqualified right of access to university grounds, property, or services (see SSM 104-02). Interfering with the peaceful conduct of university-related business or activities or remaining on campus grounds after a request to leave may be considered a crime. All incidents and allegations of violent or threatening conduct by an ASU student (whether on- or off-campus) must be reported to the ASU Police Department (ASU PD) and the Office of the Dean of Students.

Waiting for an absent instructor

In the event the instructor fails to indicate a time obligation, the time obligation will be 15 minutes for class sessions lasting 90 minutes or less, and 30 minutes for class sessions lasting more than 90 minutes. Students may be directed to wait longer by someone from the academic unit if they know the instructor will arrive shortly.

Syllabus disclaimer

Any information in this syllabus (other than grading and absence policies) may be subject to change with reasonable advance notice. Remember to check your ASU email and the course site often.