CSE 572: Data Mining

Room: Tempe Ed Lecture Hall EDC117*, Tuesday and Thursday 4:30PM to 5:45PM*
Zoom:

Instructor: Dr. Yanjie Fu, Associate Professor of Computer Science
E-mail: yanjie.fu@asu.edu; Phone: TBD; Blackboard: Canvas
Office: Tempe campus, BYENG 476
Office Hours: Tuesday and Thursday 3:30PM to 4:30 PM or by appointment via zoom

TA: Liangyi Huang; Email: lhuan139@asu.edu
TA Office hours: TBD
TA's Office: TBD

Grader 1: PRITHVI, SHIRKE, Email: pshirke1@asu.edu
Grader 2: Kartik Jawanjal, Email: kjawanja@asu.edu

Prerequisite(s): 1) Computer Science background in databases, algorithms, data structures, programming; 2) programming skills in python and other languages

Suggested Preparation Before Classes or During the First Two Weeks: 1, practice your python programming skills, for example, use python to implement the bubble sorting algorithm and use python to develop a course registration system. 2, get familiar with some machine learning libraries (e.g., pandas, scikit-learn, pytorch); 3, review your probability, statistics, hypothesis testing, and Baysian methods course materials; 4, review your database, data structure, algorithm course materials; 5, review your linear algebra, discrete math, linear regression course materials; 6, if you have learned optimization or operation research, review various optimization methods, such as, maximized likelihood estimation, maximize a posterior, MCMC, expectation-maximization. 7, select and practice a data science/ML/DM competition in Kaggle.com

Text Book: There is no required textbook. Most of the material (slides and handwritten notes) for the course will be provided on the course webpage.

Recommended Readings:
- ``Introduction to Data Mining'', by Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Addison Wesley, ISBN: 0-321-32136-7, 2005. (A CS algorithmic perspective. Easy to read.)
- ``Pattern Recognition and Machine Learning", by Christopher M. Bishop, Springer, ISBN: 0-387-31073-8. (A statistical perspective. Fundamental yet challenging.)

- ``Deep Learning'', by Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Available for free online (http://www.deeplearningbook.org/) or available for purchase. 2016.
- ``Reinforcement Learning: An Introduction'', by Richard S. Sutton and Andrew G. Barto. Second Edition. MIT Press, 2018

Course Description and Objectives:

(1) To introduce the ideas, intuitions, methodologies, algorithms, strengths, weaknesses, improvements, applications of classical machine learning tasks including supervised learning (e.g., regression, classification, ranking) and unsupervised learning (e.g., clustering)
(2) To discuss emerging machine learning topics, such model regularization, recommender systems, reinforcement learning, deep learning

Learning Programming Languages before Classes:

Python is a major language for AI and data science. To learn Python, check out: https://www.programiz.com/python-programming/first-program.

Grading Policy:

3-6 Homework Assignments:	20%
1-2 Exams	40%
Hands-on Project or ML Competition	35%
Attendance/paper reading/discussion:	5%

All grading will be determined based on a scale of: A+: [97-100%], A: [94-97%), A-: [90-94%), B+: [87-90%), B: [84-87%), B-: [80-84%), C+: [76-80%), C: [70-76%), D: [60- 70%), E: [0-60%). Final grades in the course may be curved at the instructor's discretion.

Class Policies:

1. Face-to-face lectures: the face-to-face mode is more informative and interactive. You should make every effort to attend each lecture. There will be one point penalty (out of a maximum of 5) for each class missed beyond three (regardless of the reason). To make up a missed exam, you need to contact the student service office and notify the instructor about your absence before the exam. Your participation points will be reduced if you are found playing your electronics, sleeping, chatting, being significantly late, etc.

2. Note that class sessions may be recorded, and recordings provided to enrolled students, instructors or instructional support personnel. If you have concerns about being recorded, please contact the course instructor. Recordings may be used to accommodate student absences.

3. The Late Submission policy is set to deduct 1/4 per day for late submissions.

4. The reading materials (PPT files) for the lecture (if any) should be available on Webcourses before about 11:59 PM on the day before the lecture. Please read these slides and print it out in Handouts format and bring to the classroom and take notes on them.

5. Assignments must be completed independently. While discussions and collaborations during class time are encouraged, homework and examination problems must be solved independently. Plagiarism in any assignment and examination is strictly prohibited. A score of zero will be given to **all involved parties**, and more actions may be taken which may result in dismissal from the university.

6. Issues related to a grade must be raised within 3 days after the grade is posted on the Canvas. Keep calm (not angry) and be polite when writing your emails.

7. If you copy and past materials from existing papers and technical posts to your final project reports with (>20% detection rates), I will file a claim to the student service office and initialize **the student misconduct investigation**.

8. Be honest with your written emails when you request an extension of due day, modification of scores, etc. In the past, I observed fake materials (e.g., a later modified & manipulated picture of homework submission), and non-verified reasons (e.g., unreliable network connectivity right at 11:50pm, operation system crash right at 11:55pm).

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

Anonymous Feedback/Comments/Suggestions?

Feel free to leave any comments and suggestions about how the instructor/grader can do better to help you learn this course, such as whether the lectures are clear, examples are helpful, questions are answered timely, etc. Please check this anonymous form. Your input is highly appreciated.

Schedule (non-restrictive):

Week 1: Introduction (Tutorial of Python Programming by TA)
Week 1&2: Understanding Data & Data Preprocessing (Tutorial of Pandas by TA)
Week 3: Supervised Learning I: Decision Trees
Week 4: Supervised Learning II: Ensemble Learning, Bagging, Random Forest
Week 5: Supervised Learning III: Generative Models & Naïve Bayesian Learning, Discriminative Models & Logistic Classification, KNN Nearest Neighbor
Week 6: Supervised Learning IV: Boosting, AdaBoosting
Week 7 & 8: Supervised Learning V: Hard Margin, Soft Margin, Non-linear SVM, Kernel Tricks,

Week 9 & 10: Unsupervised Learning I: Clustering Basics, K-means, Hierarchical Clustering
Week 10: Unsupervised Learning II: Density Based Clustering, Fuzzy Set & Fuzzy Logic, Fuzzy K-means
Week 11 (reading): Unsupervised Learning III: EM, Mixture Model, GMM
Week 11 & 12: Learning with Matrix Data I: RecSys, Content based Recommendation and Collaborative Filtering, Latent Factor Models and Factorization
Week 13: Learning to Rank
Week 14 & 15: Reinforcement Learning (reading: comparison of onpolicy, offpolicy, offline RL, inverse RL, Waterloo CS IRL)
Additional Lectures: If I still have time, I will quickly discuss deep sequential learning (RNN, LSTM, GRU, etc.), deep representation learning, deep generative learning (GANs, VAEs, Transformers), deep time series learning.

Important Dates (non-restrictive):

Project pitch due date: Week 5 Monday (Deliverables: a video)
Project progress check point 1 due date: Week 7 Monday (Deliverables: a video & a peer evaluation)
Form project team due date: Week 5 Monday after project pitch due day
Project progress check point 2 due date: Week 12 Monday (Deliverables: a video & a peer evaluation)
Final project report due date: Week 15 Thursday (Deliverables: a paper & a peer evaluation)
Mid-term Exam (online flexible time exam): The midterm exam covers supervised learning and unsupervised learning I & II. The midterm exam is scheduled around Week 9 after finishing unsupervised learning II.
Final Exam: arranged by ASU, but usually in the final exam week: Tuesday *4:30PM to 5:45PM.*

Project:

Check project descriptions.

Homework:

Check homework descriptions.

My Recommended Learning Strategy:

1, Attend the classes instead of being a home-stay student watching videos. Classrooms are more interactive and informative.
2, Read the slides and sometimes textbook chapters provided in the Modules before every lecture. Attend the face-to-face lectures. Watch the video recording after the lecture, combined with reading the notes.

3, During the learning, you need to create a checklist of your learning outcomes. For each ML algorithm:
- Understand the general idea and motivation?
- Understand key definitions/measures/concepts that are used the algorithm?
- Understand the key steps/components/calculation procedures of the algorithm?
- Understand the underlying optimization objective of the algorithm?
- Understand the strengths and weaknesses of the algorithm? In which case it work well or doesn't work?
- Understand the improvement of the algorithm?
- Play a toy data with the algorithm using ML programming libraries (e.g., Python, Scikit Learn, open-source code in the Gittub)
- Read the open-resource code of the algorithm to see if you can understand the code and then reproduce the algorithm from scratch by yourself?
- Try to use this algorithm or the problem-solving philosophy behind the algorithm to solve a challenge in your full-time/part-time job, your current Ph.D. research, or a data competition.

4, interactive learning by attending lectures, interacting with classmates, TA, and lectures, reading external materials, practicing programming
- Before asking a question, you are suggested to firstly recall what your learned in the classroom and check whether the related knowledge was discussed in the lecture or not.
- If your question was not discussed in the lecture, you are suggested to exploit what you learned in the lecture to perform inference to solve the question creatively.
- You can leverage Google and your external reading skills to find your answer to the question.
- You can ask your classmates and TA.
- In the face to face in-classroom learning, you can ask questions during the lecture, the break, or the office hours.

Email Policy:

This document sets forth guidelines for email communication in this course. Excessive emails are problematic for our fellow students, and the instructor. Please be sure you have a legitimate need to send an email.
- The instructor will answer email covering the following:
  a. Questions that arise from difficulty in understanding course material or content
  b. Requests for feedback on graded work
  c. Private issues related to your participation and progress in the course
- The instructor will NOT answer emails for the following:
  a. Questions already answered in the course information and contents on Canvas (please look on Canvas first)
  b. Lack of clear purpose of why the email was sent

   c. Questions unrelated to the course or of an inappropriate nature
   d. No signature that indicates who send the email
- The instructor will respond to emails sent on a given day no later than close of work on the next workday.
- If the subject of the question would be of general interest, the instructor will copy all other members of the class.
- You will be expected to have daily access to the internet and e-mail, since I will be e-mailing you constantly about assignment updates, additions and changes. YOU MUST UNIVERSITY E-MAIL FOR ALL CORRESPONDENCE RELATED TO THIS CLASS. The Professor will not communicate with you via e-mail addresses outside the University system. If you do not have a Knights E-mail account, you need to activate one immediately.
- If you do not own a computer, there are computers accessible to you in all computer labs and libraries, and most computer labs have computers connected to the internet.

Academic Integrity:

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. All engineering students are expected to adhere to the ASU Student Honor Code.

All work submitted for the course cannot have been submitted for any other course or any previous section of this same course. Student academic integrity violations are reported to the Fulton Schools of Engineering Academic Integrity Office (AIO). Withdrawing from this course will not absolve you of responsibility for an academic integrity violation and any sanctions that are applied. The AIO maintains a record of all violations and has access to academic integrity violations committed in all other ASU college/schools.

Unless explicitly allowed by your instructor, the use of generative AI tools on any course assignment or exam will be considered academic dishonesty and a violation of the ASU Academic Integrity Policy. Students confirmed to be engaging in non-allowable use of generative AI will be sanctioned according to the academic integrity policy and FSE sanctioning guidelines.

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, per the Student Services Manual, SSM 104–02
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.

Disability Accommodations

Suitable accommodations are made for students having disabilities. Students needing accommodations must register with the ASU Student Accessibility and Inclusive Learning Services office 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.

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

Photo requirement

Arizona State University requires each enrolled student and university employee to have on file with ASU a current photo that meets ASU's requirements (your "Photo"). ASU uses your Photo to identify you, as necessary, to provide you educational and related services as an enrolled student at ASU. If you do not have an acceptable Photo on file with ASU, or if you do not consent to the use of your Photo, your access to ASU resources, including access to classes (online or in person) may be negatively affected or denied.