CS 584: Data Mining

George Mason University

Department of Computer Science

Course title and number	CS 584: Data Mining
Term	Fall 2024
Class times and location	Tue 7:20pm-10:00pm, Planetary Hall 131
Piazza	https://piazza.com/gmu/fall2024/cs584001

1. Course Description and Prerequisites

The amount of data available for analysis continues to increase exponentially across a broad range of areas. This leads to the need for development of techniques to discover useful and interesting information from these large collections of data. This course aims to provide an overview of key data mining methods and techniques, including classification, regression, dimension reduction, clustering, association rule mining, recommendation, and text mining. The emphasis will be on developing basic skills for data processing, modeling, prediction, and performance evaluation. Besides, topics about dark side of data mining techniques and applications will also be introduced, like various types of unfairness and bias in different data mining applications.

Formally, you must have received a grade of C or better in CS 310 and STAT 344. Programming experience in Python is preferred, although Java or C will work as well (assignments will use the Python framework). Students should be familiar with probability and statistics concepts, as well as linear algebra. Please expect lots of programming in all the assignments and class projects.

2. Learning Outcomes

1) The ability to apply computing principles, probability and statistics relevant to the data mining discipline to analyze data.
2) A thorough understanding of model programming with data mining tools, algorithms for estimation, prediction, and pattern discovery.
3) The ability to analyze a problem, identifying and defining the computing requirements appropriate to its solution: data collection and preparation, functional requirements, selection of models and prediction algorithms, software, and performance evaluation.
4) The ability to understand performance metrics used in the data mining field to interpret the results of applying an algorithm or model, to compare methods and to reach conclusions about data.
5) The ability to communicate effectively to an audience the steps and results followed in solving a data mining problem.

The learning outcomes will be assessed based on a combination of homework assignments, exams, projects, and presentations.

3. Instructor Information

Instructor: Ziwei Zhu
Email: zzhu20 at gmu dot edu
Office: Engineering 4609
Office hours: 4:00pm-5:00pm every Tuesday
Attention: write '[CS 584]' in the subject if the email is related to this course

4. TA Information

TA: Amir Hossain Raj
Email: araj20@gmu.edu
Office hours: 12pm-1pm every Wednesday
Zoom link:
https://us05web.zoom.us/j/86853180370?pwd=cugC6rbUc3L4jQ5W2DAXzbpj13srjO.1

5. Textbook and/or Resource Material

- Pang-Ning Tan, Michael Steinbach, Anuj Karpatne and Vipin Kumar *Introduction to Data Mining (Second Edition)*. Website: https://www-users.cse.umn.edu/~kumar001/dmbook/index.php
- Jure Leskovec, Anand Rajaraman, and Jeff Ullman *Mining of Massive Datasets*. Website: http://www.mmds.org/

6. Grading Policies

Your final letter grade will be given based on:

Letter grade	Points (out of 100)
A+	[98, 100)
A	[94, 98)
A-	[90, 94)
B+	[86, 90)
B	[83, 86)
B-	[80, 83)
C	**[60, 80)**
F	**<60**

The overall course points will be determined using the following weights.

1) Six homework assignments: 35%
2) In-class quiz: 20%
3) Midterm exam: 20%
4) Final exam: 25%

Details are as follows:

Homework assignments: The first assignment accounts for 5%, and each of the other five assignments accounts for 6%. ***** Late submissions will incur a penalty of 10 points deducted for every hour past the deadline. *****

In-class Quiz: there will be 4 in-class quizzes, each will be 20 min, and each accounts for 5% towards the final grade.

Midterm Exam: will be a two-hour closed book exam. (One cheatsheet of standard US letter size paper is allowed, calculator is allowed but is not necessary).

Final Exam: will be a two-hour closed book exam. (One cheatsheet of standard US letter size paper is allowed, calculator is allowed but is not necessary)

***** A missed exam or quiz cannot be made up. All assignments must be performed individually. *****

7. Course Topics, Calendar of Activities, Project Milestones (subject to change)

Week	Topic	Assignments	Quiz
1 (08/27)	Introduction to Data Mining	HW0 out	
2 (09/03)	Classification Basics and K Nearest Neighbor	HW1 out	
3 (09/10)	Principle Component Analysis	HW0 due	Quiz 1 (w1, w2)
4 (09/17)	Linear Regression		
5 (09/24)	Logistic Regression and Neural Networks	HW1 due, HW2 out	Quiz 2 (w3, w4)
6 (10/01)	Deep Learning and Fairness in Machine Learning		Explain Quiz 1&2
7 (10/08)	**Midterm Exam**		
8 (10/15)	Explain midterm exam, Clustering	HW2 due, HW3 out	
9 (10/22)	Clustering		
10 (10/29)	Associate Rule Mining	HW4&5 out	Quiz 3 (w8, w9)
11 (11/05)	**Election Day, no class**	HW3 due	
12 (11/12)	Recommender Systems: Basics		
13 (11/19)	Recommender Systems: Advance		
14 (11/26)	Fairness and Bias in Recommender Systems		Quiz 4 (w10, 12, 13)
15 (12/03)	Final review	HW4&5 due	Explain Quiz 3&4
16 (TBA)	**Final Exam**		

Common Policies Affecting All Courses at George Mason University

Updated August 2024

These four policies affect students in all courses at George Mason University. This Course Policy Addendum must be made available to students in all courses (see Catalog Policy AP.2.5).

Additional policies affecting this course, and additional resources or guidance regarding these policies, may be provided to students by the instructor.

Academic Standards

Academic Standards exist to promote authentic scholarship, support the institution's goal of maintaining high standards of academic excellence, and encourage continued ethical behavior of faculty and students to cultivate an educational community which values integrity and produces graduates who carry this commitment forward into professional practice.

As members of the George Mason University community, we are committed to fostering an environment of trust, respect, and scholarly excellence. Our academic standards are the foundation of this commitment, guiding our behavior and interactions within this academic community. The practices for implementing these standards adapt to modern practices, disciplinary contexts, and technological advancements. Our standards are embodied in our courses, policies, and scholarship, and are upheld in the following principles:

- **Honesty:** Providing accurate information in all academic endeavors, including communications, assignments, and examinations.
- **Acknowledgement:** Giving proper credit for all contributions to one's work. This involves the use of accurate citations and references for any ideas, words, or materials created by others in the style appropriate to the discipline. It also includes acknowledging shared authorship in group projects, co-authored pieces, and project reports.
- **Uniqueness of Work:** Ensuring that all submitted work is the result of one's own effort and is original, including free from self-plagiarism. This principle extends to written assignments, code, presentations, exams, and all other forms of academic work.

Violations of these standards—including but not limited to plagiarism, fabrication, and cheating—are taken seriously and will be addressed in accordance with university policies. The process for reporting, investigating, and adjudicating violations is outlined in the university's procedures. Consequences of violations may include academic sanctions, disciplinary actions, and other measures necessary to uphold the integrity of our academic community.

The principles outlined in these academic standards reflect our collective commitment to upholding the highest standards of honesty, acknowledgement, and uniqueness of work. By adhering to these principles, we ensure the continued excellence and integrity of George Mason University's academic community.

Student responsibility: Students are responsible for understanding how these general expectations regarding academic standards apply to each course, assignment, or exam they participate in; students should ask their instructor for clarification on any aspect that is not clear to them.

Accommodations for Students with Disabilities

Disability Services at George Mason University is committed to upholding the letter and spirit of the laws that ensure equal treatment of people with disabilities. Under the administration of University Life, Disability Services implements and coordinates reasonable accommodations and disability-related services that afford equal access to university programs and activities. Students can begin the registration process with Disability Services at any time during their enrollment at George Mason University. If you are seeking accommodations, please visit https://ds.gmu.edu/ for detailed information about the Disability Services registration process. Disability Services is located in Student Union Building I (SUB I), Suite 2500. Email: ods@gmu.edu. Phone: (703) 993-2474.

Student responsibility: Students are responsible for registering with Disability Services and communicating about their approved accommodations with their instructor *in advance* of any relevant class meeting, assignment, or exam.

FERPA and Use of GMU Email Addresses for Course Communication

The Family Educational Rights and Privacy Act (FERPA) governs the disclosure of education records for eligible students and is an essential aspect of any course. **Students must use their GMU email account** to receive important University information, including communications related to this class. Instructors will not respond to messages sent from or send messages regarding course content to a non-GMU email address.

Student responsibility: Students are responsible for checking their GMU email regularly for course-related information, and/or ensuring that GMU email messages are forwarded to an account they do check.

Title IX Resources and Required Reporting

As a part of George Mason University's commitment to providing a safe and non-discriminatory learning, living, and working environment for all members of the University community, the University does not discriminate on the basis of sex or gender in any of its education or employment programs and activities. Accordingly, **all**
non-confidential employees, including your faculty member, have a legal requirement to report to the Title IX Coordinator, all relevant details obtained directly or indirectly about any incident of Prohibited Conduct (such as sexual harassment, sexual assault, gender-based stalking, dating/domestic violence). Upon notifying the Title IX Coordinator of possible Prohibited Conduct, the Title IX Coordinator will assess the report and determine if outreach is required. If outreach is required, the individual the report is about (the "Complainant") will receive a communication, likely in the form of an email, offering that person the option to meet with a representative of the Title IX office.

For more information about non-confidential employees, resources, and Prohibited Conduct, please see University Policy 1202: Sexual and Gender-Based Misconduct and Other Forms of Interpersonal Violence.

Questions regarding Title IX can be directed to the Title IX Coordinator via email to TitleIX@gmu.edu, by phone at 703-993-8730, or in person on the Fairfax campus in Aquia 373.

Student opportunity: If you prefer to speak to someone ***confidentially***, please contact one of Mason's confidential employees in Student Support and Advocacy (SSAC), Counseling and Psychological Services (CAPS), Student Health Services (SHS), and/or the Office of the University Ombudsperson.