
GEORGE MASON UNIVERSITY
SCHOOL OF COMPUTING
DEPARTMENT OF COMPUTER SCIENCE
CS484: DATA MINING
SPRING 2023
INSTRUCTOR: ERION PLAKU

Contents

1	General Information	2
2	Course Description	2
3	Course Outcomes	2
4	Textbook	2
5	Lectures	3
6	Assessment and Grade Brakedown	3
7	Lateness and Make-Up Policy	4
8	Student Support Resources on Campus	4
9	Academic Integrity	4
10	Recording and/or Sharing Class Materials	4
11	Privacy	5
12	Disability Accommodations	5
13	Diversity and Inclusion	5
14	Sexual Harassment, Sexual Misconduct, and Interpersonal Violence	6

1 General Information

Classroom	Innovation Hall
Day and Time	Thursday, 4:30pm-7:10pm
Schedule Type	Lecture, in person
Credit Hours	3
Required Prerequisites	((CS 310 ^C or 310 ^{XS}) and (STAT 344 ^C , 344 ^{XS} , 334 ^C or 334 ^{XS}))
Grading	This course is graded on the Undergraduate Regular scale
Instructor	Dr. Erion Plaku Associate Professor Office Nguyen Engineering Building 4423 Web https://cs.gmu.edu/~plaku/ E-mail plaku@gmu.edu Office Hours Thursday 2:00pm-4:00pm
Teaching Assistant	TBD

2 Course Description

GMU Catalog: Basic principles and methods for data analysis and knowledge discovery. Emphasizes developing basic skills for modeling and prediction and performance evaluation. Topics include system design; data quality, preprocessing, and association; event classification; clustering; biometrics; business intelligence; and mining complex types of data.

3 Course Outcomes

- The ability to apply computing principles, probability and statistics relevant to the data mining discipline to analyze data.
- A good understanding of model programming with data mining tools, algorithms for estimation, prediction, and pattern discovery.
- 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.
- The ability to understand performance metrics used in data mining to interpret the results of applying an algorithm or model, to compare methods and to reach conclusions about data.
- The ability to communicate effectively the approach and results followed in solving a data-mining problem.

4 Textbook

Title	“Introduction to Data Mining (Second Edition)”
Authors	Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar
Print ISBN	9780133128901, 0133128903
eText ISBN	9780134080284, 0134080289
Publisher	Pearson
Web	click here

Useful textbook (free): Jure Leskovec, Anand Rajaraman, and Jeff Ullman “Mining of Massive Datasets.” Website: mmdes.org

5 Lectures

[as time permits]

- Course Logistics & Introduction to Data Mining
- Data
 - Types of Data
 - Data Quality
 - Data Processing
 - Measures of Similarity and Dissimilarity
- Classification
 - Basic Concepts
 - General Framework for Classification
 - Decision Tree Classifier
 - Model Overfitting, Selection, Evaluation, and Comparison
 - Rule-Based Classifiers
 - Nearest Neighbors Classifiers
 - Naive Bayes Classifier
 - Logistic Regression
- Regression
 - Linear Regression
 - Adapting Decision Trees and Nearest Neighbors for Regression Analysis
- Ensemble Methods
- Artificial Neural Networks
- Deep Learning (introduction)
- Association Analysis
- Cluster Analysis
 - K -means Algorithm
 - Agglomerative Hierarchical Clustering
 - DBSCAN
 - Cluster Evaluation
 - Prototype-Based Clustering
 - Density-Based Clustering
 - Graph-Based Clustering
- Anomaly Detection
- Avoiding False Discoveries

6 Assessment and Grade Brakedown

Homeworks and Quizzes	10%
Midterm Exam	15%
Programming Projects	55%
Final Exam	20%

- All work is individual. You are not allowed to work with others on homeworks, quizzes, exams, or projects.
- Quizzes and exams will be in-person, and are closed book, closed notes. No access to electronic devices during quizzes or exams.
- Programming assignments should be done in Python. If you are not familiar with Python, please see the instructor as soon as possible.
- There will be three or four programming projects. You will have between two to three weeks for each project depending on its difficulty. A written report, describing your approach and results, will be required for each project.

7 Lateness and Make-Up Policy

- Every homework and programming project should be returned by the due date.
- You have 5 late days in total that you can use at any point in time without any penalty. You will receive 0 points on every homework or programming project that is returned late (after exhausting your 5 late days). Each homework or project will be due by the end of day (11:59pm). A submission of even one minute late will count as a day late.
- There will be no make-up for missed in-class quizzes or exams except under the following: If there is ever an emergency situation which prohibits you from attending class during quizzes or exams, you must alert the instructor and the appropriate student offices to make alternative arrangements. Doctor's notice or other proper documentation may be required to grant the exception.

8 Student Support Resources on Campus

Mason offers numerous student support resources, as detailed [here](#).

9 Academic Integrity

- All students must adhere to the **GMU Honor Code**. The University Honor Code is upheld and supported by the Office for Academic Integrity. Please see the Office for Academic Integrity for a full description of the code and the honor committee process. The principle of academic integrity is taken very seriously and violations are treated gravely.
- In addition to this honor code, the computer science department has further honor code policies regarding programming projects, as detailed [here](#).
- See also the **Statement on Academic Integrity** by the CS Department.
- Any deviation from the GMU or the CS Honor Code is considered an Honor Code violation.

10 Recording and/or Sharing Class Materials

- Sharing of instructor-created materials, particularly materials relevant to assignments or exams, to public online “study” sites is considered a violation of Mason's Honor Code: these include but are not limited to accessing exam/quiz/homeworks/project/or any assignment questions or answers for this class; uploading of any of the instructor's materials, quizzes, homeworks, projects, or exams; and uploading any of your own answers or finished work. For

more information, see the Office of Academic Integrity’s **summary of information about online study sites**. They also have a **short video** on this subject.

- All course materials posted to Blackboard or other course site are private to this class; by federal law, any materials that identify specific students (via their name, voice, or image) must not be shared with anyone not enrolled in this class.
- Videorecordings – whether made by instructors or students – of class meetings that include audio, visual, or textual information from other students are private and must not be shared outside the class.
- Live video conference meetings (e.g. Collaborate or Zoom) that include audio, textual, or visual information from other students must be viewed privately and not shared with others in your household or recorded and shared outside the class.

11 Privacy

- Student privacy is governed by the Family Educational Rights and Privacy Act (FERPA) and is an essential aspect of any course.
- Students must use their Mason email account to receive important University information, including communications related to this class. I will not respond to messages sent from or send messages to a non-Mason email address.

12 Disability Accommodations

- 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 <http://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
- **Covid-19 Note:** Students who have a Covid-related disability should contact the Disability Services office; DS will contact faculty using standard protocols about any students who require accommodations. Faculty are not expected to create accommodations for students outside of the Disability Services official guidelines.

13 Diversity and Inclusion

- George Mason University promotes a living and learning environment for outstanding growth and productivity among its students, faculty and staff. Through its curriculum, programs, policies, procedures, services and resources, Mason strives to maintain a quality environment for work, study and personal growth.
- An emphasis upon diversity and inclusion throughout the campus community is essential to achieve these goals. Diversity is broadly defined to include such characteristics as, but not limited to, race, ethnicity, gender, religion, age, disability, and sexual orientation. Diversity also entails different viewpoints, philosophies, and perspectives. Attention to these aspects of

diversity will help promote a culture of inclusion and belonging, and an environment where diverse opinions, backgrounds and practices have the opportunity to be voiced, heard and respected.

- For more information, please see
 - **Mason Diversity Statement** and
 - **Mason Non-Discriminatory Policy**

14 Sexual Harassment, Sexual Misconduct, and Interpersonal Violence

- George Mason University is committed to providing a learning, living and working environment that is free from discrimination and a campus that is free of sexual misconduct and other acts of interpersonal violence in order to promote community well-being and student success. We encourage students and employees who believe that they have been sexually harassed, sexually assaulted or subjected to sexual or interpersonal misconduct to seek assistance and support. University Policy 1202: Sexual Harassment and Misconduct speaks to the specifics of Mason’s process, the resources, and the options available to students and employees.
- **Notice of mandatory reporting of sexual or interpersonal misconduct:** As a faculty member, I am designated as a “Non-Confidential Employee,” and must report all disclosures of sexual assault, sexual harassment, interpersonal violence, stalking, sexual exploitation, complicity, and retaliation to Mason’s Title IX Coordinator per University Policy 1202. If you wish to speak with someone confidentially, please contact one of Mason’s confidential resources, such as Student Support and Advocacy Center (SSAC) at 703-993-3686 or Counseling and Psychological Services (CAPS) at 703-993-2380. You may also seek assistance or support measures from Mason’s Title IX Coordinator by calling 703-993-8730, or emailing titleix@gmu.edu.