Department of Computational and Data Science

4400 University Drive, MS 6C3, Fairfax, Virginia 22030
Phone: 703-993-3807, Fax: 703-993-9300
Web: cos.gmu.edu/cds

CDS 303 – SCIENTIFIC DATA MINING

3 CREDITS, FALL 2021

ONLINE

FACULTY

Name:	Mr. Matthew Valko
Course Forum:	Blackboard
Office Hours:	Wednesday 5—6pm or by appointment
Office Location:	Remote
Email Address:	mvalko@gmu.edu

PREREQUISITES/COREQUISITES

Recommended: CDS 101, CDS 130, or equivalent.

UNIVERSITY CATALOG COURSE DESCRIPTION

Data mining techniques from statistics, machine learning, and visualization to scientific knowledge discovery. Students will be given a set of case studies and projects to test their understanding of this field and provide a foundation for future applications in their careers.

COURSE OVERVIEW

Data mining is the process of discovering salient information in large data sets. Data mining is a subfield of computer science, involving methods from artificial intelligence, machine learning, statistics, and database systems. Data mining is an integral part of data science; which also focuses on deriving meaningful insight from data. In this class, students will become an understanding of this field and learn the basic and most important techniques, methods and tools associated with data mining, including frequent itemsets mining and association rules, classification, clustering, and

outlier detection. As there is no point in discovering useful patterns in data if no one ever hears about it, communication of results will be important and reoccurring theme.

COURSE DELIVERY METHOD

This is a 3-credit course that will be delivered online for consumption. Lectures, labs and activities will be online.

REQUIRED AND RECOMMENDED TEXTS

Tan, P., Steinbach, M., Karpatne, A., Kumar, V. (2018). Introduction to Data Mining.

O'Neil, C., & Schutt, R. (2013). Doing data science: Straight talk from the frontline. " O'Reilly Media, Inc.". (Highly Recommended)

Silge, J. & Robinson, D. (2017). Text Mining with R: A Tidy Approach.

Han J., Kamber M., Pei J., Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2011

COURSE SCHEDULE

The course will be taught as a combination of lectures, topic/problem-oriented discussion, and tutorials based on independent reading and class discussion.

Week	Week of	Topic	Notes
1	23-Aug	Introduction to Data Mining	
2			
3			
4		Classification	
5			
6			
7		Association Analysis	Mid-Term
8			
9			
10			
11		Cluster Analysis	
12			
13			

14		Anomaly Detection	
15	Dec 8th		Final Exam

Please note that this schedule is tentative, and it will change according to the needs, pace and interest of the class. Dates of midterm and quizzes might change. Students will be notified in time.

COURSE WEBSITE – MEANS OF COMMUNICATION

We will use Blackboard as the course management system. Through the Blackboard page you may obtain lecture notes, retrieve assignment data and, review links to additional materials, and receive special announcements. Assignments will be posted in Blackboard, and unless otherwise stated will also be submitted through Blackboard.

Please be aware that innocent remarks can be easily misconstrued. Sarcasm and humor can be easily taken out of context. When communicating, please be positive and diplomatic.

COURSE PERFORMANCE EVALUATION

Activity	Weight
Online "class" participation. Blackboard Discussions, Quizzes, discussions etc.	**25%**
Homework Assignments and Exercises	**35%**
Midterm Exam	**15%**
Final Exam	**25%**

ASSIGNMENTS

There will be assignments which will be posted in Blackboard.

Assignments are usually to be uploaded in Blackboard as PDF file - **DO NOT email assignments directly to the instructor.**

Due dates and method of delivery will also be posted in Blackboard.

Late assignment submission: Assignments submitted **after the due date will not be accepted**. Exceptions to this policy may be made given serious circumstances at the discretion of the instructor.

Please note: Deferred of term work is a privilege and not a right; there is no guarantee that a deferral will be granted. Please make sure you notify the instructor as soon as you know a deferral is required.

It is OK to discuss homework problems with each other, including solution approaches and methods. It is not OK to share homework write-ups (i.e., copy homework assignments). Violations of this policy will be considered violations of GMU's Honor Code.

Assignment questions will usually consist of calculations, short answers, algorithms in pseudo code, or small technical challenges. The questions are designed to test a student's understanding of the course material.

When applicable, the student is **strongly** encouraged to write out **both** the solution and the step-by-step solution logic in their assignment responses, so that instructors may assess the student's overall approach to and understanding of the assigned problems. Credit will be assigned on student assignments based upon whether the student's solution is correct (approximately 50% of the score), and whether the student's solution logic is correct (approximately 50% of the score).

EXAMS

Midterm and Final exam will be given in the class to test comprehension of the topics covered in the lecture, discussions and assignments.

The final exam is cumulative.

The exams will include calculations, short answers, multiple choice, and simple discussion questions. The questions will be based on concepts covered on the assignments and in the in-class questions.

At least one class period will be used for exam review and the results of the exam will be discussed within a week after the exam was given.

ATTENDANCE

Following lectures, doing "class" assignments, and participating in discussion is pivotal for comprehending the course materials; as many details are not present in the text of the slides but may show up on tests. Non- homework assignments and quizzes are a non-trivial part of the overall grade in the course and should not be ignored.

STUDENT EXPECTATIONS

ACADEMIC INTEGRITY

Students must be responsible for their own work, and students and faculty must take on the responsibility of dealing explicitly with violations. The tenet must be a foundation of our university culture. [See http://oai.gmu.edu].

HONOR CODE

Students must adhere to the guidelines of the George Mason University Honor Code [See https://catalog.gmu.edu/policies/honor-code-system/].

MASONLIVE/EMAIL (GMU EMAIL)

Students are responsible for the content of university communications sent to their George Mason University email account and are required to activate their account and check it regularly. All communication from the university, college, school, and program will be sent to students **solely** through their Mason email account. [See https://masonlivelogin.gmu.edu/login].

RESPONSIBLE USE OF COMPUTING

Students must follow the university policy for Responsible Use of Computing.

[See https://universitypolicy.gmu.edu/policies/responsible-use-of-computing/].

STUDENTS WITH DISABILITIES

Students with disabilities who seek accommodations in a course must be registered with the George Mason University Office of Disability Services (ODS) and inform their instructor, in writing, at the beginning of the semester [See http://ods.gmu.edu].

STUDENT SERVICES

STUDENT RESOURCES

For technical questions regarding Blackboard, see Courses Support and Blackboard Tutorials. If you still have questions, email courses@gmu.edu for assistance with Blackboard.

For technical questions regarding computer networking, see IT Services for Students. If you still have questions, email support@gmu.edu or call (703) 993-8870.

UNIVERSITY LIBRARIES

University Libraries provides resources for distance students. [See http://library.gmu.edu/distance].

WRITING CENTER

The George Mason University Writing Center staff provides a variety of resources and services (e.g., tutoring, workshops, writing guides, handbooks) intended to support students as they work to construct and share knowledge through writing. [See http://writingcenter.gmu.edu]. You can now sign up for an Online Writing Lab (OWL) session just like you sign up for a face-to-face session in the Writing Center, which means YOU set the date and time of the appointment! Learn more about the Online Writing Lab (OWL) (found under Online Tutoring).

COUNSELING AND PSYCHOLOGICAL SERVICES

The George Mason University Counseling and Psychological Services (CAPS) staff consists of professional counseling and clinical psychologists, social workers, and counselors who offer a wide range of services (e.g., individual and group counseling, workshops and outreach programs) to enhance students' personal experience and academic performance [See http://caps.gmu.edu].

Family Educational Rights and Privacy Act (FERPA)

The Family Educational Rights and Privacy Act of 1974 (FERPA), also known as the "Buckley Amendment," is a federal law that gives protection to student educational records and provides students with certain rights. [See http://registrar.gmu.edu/privacy].

Disclaimer and Copyright Notice

Disclaimer: Any typographical errors in this Course Outline are subject to change and will be announced in class.