CDS-303- Scientific Data Mining

LOCATION: Fairfax Campus, Nguyen Engineering Building, Room 1110

SCHEDULE: Monday and Wednesday 1:30 – 2:45

INSTRUCTOR: Holly Russo, PhD, hrusso@masonlive.gmu.edu

OFFICE HOURS: Monday and Wednesday (F2F) 10am – Noon

Contents

Course Description

Everyone makes some type of data-driven decision every day. Individuals decide what to wear based on weather data. Fortune 500 executives decide what to do next year based on business data. The likelihood that those data-driven decisions are correct is based on a variety of factors including:

- Quality of data

- Quality of analysis

- Understanding of the question or problem

- Communication of the analytic results

Most often, the person relying on the decision is not the one performing the data analysis. Meteorologists analyze data to predict the weather supporting general public decision makers: they must understand the specific questions that individuals have about the weather, and how to present the results in a way that the general public can understand. Likewise, data scientists analyze business data in support of executive decisions: they must understand what really matters to the executive, what the questions are and how to present the results in a way that the executive can understand and use.

You are most likely here because you want to help others make better decisions based on solid data and analysis. This course will help you understand the process of scientific data analysis to support decision makers, from the initial question to the final report and beyond.

We'll cover technical, communication and project management topics. You will have the opportunity to practice:

- Communicating with a 'client'

- Planning an analytic project

- Choosing and applying the correct analytic approach

- Developing a model, visualization, and final report

- Presenting interim and final results to the 'client'

The class structure will be a mix of lecture, quizzes, role playing and project.

Objectives

By the end of this course, you should be able to:

- Describe how the Scientific Method, the CRISP-DM process, and Agile project management can be applied on a project

- Restate a business problem as an analytic problem, in a way that makes sense to the non-technical stakeholders

- List who will benefit from your work, who may be inconvenienced, and how

- Develop a conceptual model of the analytic problem including assumptions

- Identify the type of data required for the problem, and the steps required to prepare it for analysis

- Describe the data using exploratory data analysis

- Identify the correct analytic, visualization, reporting, deployment and maintenance approaches for the problem

- Develop a modeling and testing proposal that describes your planned approach, timing and milestones, as well as a plan for reviewing the model for accuracy over time

- Develop and test a machine learning model based on a conceptual model, apply the model for analysis and produce a final report

- Understand the difference between a development and production environment, and considerations for deploying analytic models

Prerequisites

The level of skill you will need in various areas is as follows:

- Technical: You will be required to apply a machine learning algorithm to data. The software and/or coding language you use is up to you. If you choose to use Python or R, that is fine. Tableau and Excel are also fine. You will need coding, math and analytic skills to perform this task, however the level of skill required will vary based on the approach you choose.

- Communication: You will be required to elicit requirements and present results. This means you will need to think of clarifying questions to ask, develop review material and a final report, all in a way that a non-technical person would understand. To do this, you will need verbal, writing and data visualization skills.

Important Logistics

Currently, our courses are scheduled to be conducted face to face (F2F) 100%. In addition, GMU has recently updated its mask policy as follows:

Everyone, even those who are fully vaccinated, must wear a face covering when inside university property (buildings and vehicles). If you're unvaccinated, you must wear a mask outside, as well, when at an event or if physical distancing can't be maintained. Wear a mask consistently and correctly when:

- *In university facilities (indoor or enclosed space).*

- *Outside on university property if you can't maintain physical distance of at least six feet or if you're in close contact with others who are not fully vaccinated.*

- *Attending an event.*

- *Traveling to/from sites off university property for work or study if in a vehicle with other individuals.*

Follow CDC guidance and University Policy 1415 when choosing a mask and following guidance on when to wear it. Don't share face coverings.

As you arrive to each class, you will be required to show your green COVID check for that day. Please see me if you require additional accommodation.

In addition to F2F, we will use Blackboard for collaboration. Students are encouraged to collaborate with each other frequently and hold group sessions outside of class time using Teams, Zoom, Meet or other modes convenient to the group. Please see the Policies section of this syllabus for more in-depth information about masks, and required preparations in the event we must pivot to remote learning.

For communication with me, please use one of the following:

- Office hours: 10 am to Noon, Monday and Wednesday when class is in session (location TBD)

- Email: hrusso@masonlive.gmu.edu, I will do my best to respond by 8pm the following calendar day.

- Blackboard: be sure to add @hrusso to your forum post

Grading and Course Requirements

Grading will be based on both individual and team performance. Individual performance will consist of quizzes and participation in class discussions, while group grades will be based on midterm and final reviews of the group project.

Grade	Weight
Individual grade: Top 6 out of 9 total quizzes	30%
Individual grade: Participation	20%
Group grade: Project proposals (midterm grade)	20%
Group grade: Project presentation and report (final grade)	30%

The four grades are explained as follows:

Quizzes: On nine Mondays throughout the course, at the start of class, I will issue a short quiz based on the required reading for the week (note: the last two 'quizzes' will actually be short at-home assignments from your reading). Readings will be based on material that we will discuss in class, so the only way to be prepared is to do the readings. My focus in this class is on having you immediately applying what you learn, so think of each reading as a tutorial for the work you'll be doing. The quiz questions will be about applying what you read, not about memorizing it. Because I understand that life happens, I will take your top six of all quiz scores, so manage your time wisely and do not skip reading if you do not have to.

Participation: We all benefit from a vibrant class discussion, and that is what I expect from you all for this class. Above all else, the most important thing I look for is respect for everyone in the room. Specifically, the things I would like to see are that you:

- Actively support, engage and listen to your class peers: respect every question and comment, ask or comment as a follow on to others' comments and questions.

- Arrive fully prepared to every session: complete reading, prepare questions and comments, bring all materials and /or equipment needed for class.

- Play an active role in class discussions: pay attention (to the room, not your phone), ask questions and respond when called upon.

- Advance the depth of the class dialogue: present comments or questions that help increase everyone's understanding of the discussion topic. Please note that there are times when I may need to move us along to the next topic, however I will avoid this where possible.

- Promote an environment where everyone walks away feeling that they just had a great discussion from which they learned a lot.

Please understand that listening is just as important as speaking, and each person is stronger in one or the other. If you find that speaking comes easily to you, I want you to practice listening and responding directly to what was just said. Likewise, if you find that listening is much easier for you, I want you to practice speaking when you have comments or questions.

Project midterm and final grades: these will be your team grades. You will work together to solve an analytical problem using scientific data mining techniques. For the midterm, the team will submit a

project proposal to me, which we'll cover in class prior to the due date. For final reviews, you will present your work as a team to the class and submit a final report to me. The grade for the work will be the grade for everyone on the team. Note that the first bullet point in the discussion participation grade – actively support, engage and listen to your class peers – will be evaluated not only during full class discussions but also within your teams, so you must be an active member of your team.

Reading Material

Because I want you to treat each reading as a tutorial, there will not be a significant amount of reading. Instead, I want you to try out the process that you read about or answer the reading chapter questions. So, if your assigned reading is Chapter 1 of the *Hands-On Machine Learning with Scikit-Learn* book, then you may want to work through Example 1.1, and/or answer the questions at the end of Chapter 1. Trust me, this will help you score better grades on the reading quizzes.

We will use the following reading material:

- Geron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras and Tensorflow, 2nd Edition. Sebastopol, CA: O'Reilly Media Inc.

- Nwanganga, F. and Chapple, M. (2020). Practical Machine Learning in R. John Wiley & Sons Inc.

- Chapman, P., et. al. (1999). CRISP-DM 101 – Step-by-Step Data Mining. PDF guide retrieved from https://www.the-modeling-agency.com/crisp-dm.pdf (copy will be posted on Blackboard).

- Taber, A., et. al. Certified Analytic Professional Study Guide. Institute for Operations Research and the Management Sciences (INFORMS). Retrieved from https://www.certifiedanalytics.org/for_professionals.php (copy will be posted on Blackboard, register at this URL for free to receive updates and other helpful info).

- Agile Alliance. (2017). Agile Practice Guide. Newtown Square, PA: Project Management Institute.

- Provost, F. and Fawcett, T. (2013). Data Science for Business. Sebastopol, CA: O'Reilly Media Inc.

- Other PDFs: listed in a separate Class Readings document and uploaded to the applicable module on Blackboard.

I will assign page ranges or chapters from one or more of these each week. You may notice that both R and Python books are in this list. Elsewhere in this syllabus I stated that you may choose which programming language you want to use. I chose these particular R and Python books because in addition to actual coding, each has different and useful information on the process of machine learning in general. Therefore, I will be assigning readings from both books.

Course Schedule Overview

There are five different types of activities that comprise our schedule: quizzes (on readings), lectures, discussion, project work and presentations. We have already covered readings, quizzes and participation. Project presentations will be made by each team for the midterm and final grades, and I will be grading discussion participation during each presentation. I will cover more specifics on presentation requirements as we get closer to the date.

You will be doing most of your team project work as an assignment outside of class time, however I have allowed some class time for this as well. During the in-class project work, you'll be expected to get into your teams and focus on items that benefit from being together in person. I will circulate among the teams to answer questions and observe. We will work out the logistics of this together at the start of the semester, based on class size and COVID situation. **Important: Election Day is again a holiday, so the Columbus Day Monday class will not meet on the Tuesday and an extra Monday class is added to the end of the semester. Please pay close attention to the dates in the schedule below.**

Date	Focus	Quiz	Lecture	Discussion	Project work	Presentations
Monday, August 23, 2021	Intro, admin, outline, expectations		X			
Wednesday, August 25, 2021	Review, Q&A, refine expectations			X		
Monday, August 30, 2021	Module I: What is Scientific Data Mining	X	X	X		
Wednesday, September 1, 2021	Module II: Agile Project Management		X	X		
Monday, September 6, 2021	CAMPUS CLOSED, HOLIDAY					
Wednesday, September 8, 2021	Project Kickoff (in class)		X	X	X	
Monday, September 13, 2021	Module III: Business Problem Framing	X	X			
Wednesday, September 15, 2021	Module III: Business Problem Framing			X	X	
Monday, September 20, 2021	Module IV: Analytic Problem Framing	X	X			
Wednesday, September 22, 2021	Module IV: Analytic Problem Framing			X	X	
Monday, September 27, 2021	Module V: Data Gathering and Preparation	X	X			
Wednesday, September 29, 2021	Module V: Data Gathering and Preparation			X	X	
Monday, October 4, 2021	Module VI: Methodology Selection	X	X			
Wednesday, October 6, 2021	Module VI: Methodology Selection		X	X	X	
Monday, October 11, 2021	CAMPUS CLOSED, HOLIDAY: PROJECT PROPOSALS DUE					
Wednesday, October 13, 2021	Module VI: Methodology Selection		X	X	X	
Monday, October 18, 2021	Module VII: Model Building and Testing	X	X			
Wednesday, October 20, 2021	Module VII: Model Building and Testing		X			
Monday, October 25, 2021	Module VII: Model Building and Testing			X	X	
Wednesday, October 27, 2021	Project Day (in class)			X	X	
Monday, November 1, 2021	Module VIII: Solution Deployment	X	X			
Wednesday, November 3, 2021	Module VIII: Solution Deployment			X	X	
Monday, November 8, 2021	Module IX: Visualization and Presentation	X	X			
Wednesday, November 10, 2021	Module IX: Visualization and Presentation			X	X	
Monday, November 15, 2021	Module X: Development vs. Production	X	X			
Wednesday, November 17, 2021	Module X: Development vs. Production			X	X	
Monday, November 22, 2021	Project Day (in class)			X		X
Wednesday, November 24, 2021	CAMPUS CLOSED, HOLIDAY					
Monday, November 29, 2021	Module XI: Data Operations Management		X			
Wednesday, December 1, 2021	Module XI: Data Operations Management			X	X	
Monday, December 6, 2021	READING DAY: Project Work					
Wednesday, December 8, 2021	Project Day (in class)			X		X
Monday, December 13, 2021	Final presentations			X		X
Wednesday, December 15, 2021	Final presentations			X		X
Monday, December 20, 2021	Final presentations			X		X

Course Project

The course project is designed to give you a chance to work through a scientific data mining project from business problem through to final presentation. You will receive vignettes and break up into teams to work through the project. In addition to the vignettes, I will provide references to the data that should be used. For this project, each team will be required to prepare a project proposal to submit for the midterm, and a presentation with final report for the final. Each team will receive an overall grade for each of these deliverables.

In addition, your individual participation as a team member will make up part of your participation grade. At the midterm and end of the semester, I will hand out a brief questionnaire to be completed privately by each student and returned to me. The questionnaire will ask you to rate yourself and each team member on the following items:

- Participation in developing ideas and planning project

- Willingness to discuss the ideas of others

- Cooperation with other group members

- Interest and enthusiasm in project

- Participation in leading/facilitating discussion

- Group meeting attendance

- Completion of tasks by date due

Rating levels for each item will be 'always', 'most times', 'occasionally' and 'never'.

Policies

Undergraduate Course Repetition
Beginning fall 2018, there is a limit of three graded attempts for this course. A W does not count as a graded attempt. Please see AP. 1.3.4 in the University Catalog and consult with your academic advisor if you have any questions.

Academic Integrity
The integrity of the University community is affected by the individual choices made by each of us. Mason has an Honor Code with clear guidelines regarding academic integrity. Three fundamental and rather simple principles to follow at all times are that: (1) all work submitted be your own; (2) when using the work or ideas of others, including fellow students, give full credit through accurate citations; and (3) if you are uncertain about the ground rules on a particular assignment, ask for clarification. No grade is important enough to justify academic misconduct. Plagiarism means using the exact words, opinions, or factual information from another person without giving the person credit. Writers give credit through accepted documentation styles, such as parenthetical citation, footnotes, or endnotes. Paraphrased material must also be cited, using the appropriate format for this class. A simple listing of books or articles is not sufficient. Plagiarism is the equivalent of intellectual robbery and cannot be tolerated in the academic setting. If you have any doubts about what constitutes plagiarism, please see me.

DIVERSITY AND INCLUSION

Student and Faculty Names and Pronouns

Gender identity and pronoun use: If you wish, please share your name and gender pronouns with me and how best to address you in class and via email. I use She / Her / Her for myself, and you may address me as "Holly" or "Dr. Russo" in email and verbally.

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

Sexual Harassment, Sexual Misconduct, and Interpersonal Violence

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.

Religious Holidays

Students who miss classes, exams, or other assignments as a consequence of their religious observance will be provided a reasonable alternative opportunity, consistent with class attendance policies stated in the syllabus, to make up the missed work. Please inform me in advance about your planned observances to enable me to work with you to create a pathway to success.

TECHNOLOGY AND PRIVACY

Student Use of Electronic Devices

Cell phones, pagers, and other communicative devices are not allowed in this class. Please keep them stowed away and out of sight. If you have an extenuating circumstance that requires you to have your phone out, please notify me prior to the start of class. Laptops or tablets (e.g., iPads) may be permitted for the purpose of taking notes and team project work. Engaging in activities not related to the course (e.g., gaming, email, chat, etc.) will result in a significant deduction in your participation grade.

Basic Course Technology Requirements

Activities and assignments in this course will regularly use the Blackboard learning system, available at https://mymason.gmu.edu. Students are required to have regular, reliable access to a computer with an updated operating system (recommended: Windows 10 or Mac OSX 10.13 or higher) and a stable broadband Internet connection (cable modem, DSL, satellite broadband, etc., with a consistent 1.5

Mbps [megabits per second] download speed or higher. You can check your speed settings using the speed test on this website.) In addition to the requirements above, students are required to have a device with a functional camera and microphone. In an emergency, students can connect through a telephone call, but video connection is the expected norm.

Course Materials and Student Privacy

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.

Video recordings -- 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

Course Recordings

In the event that we must pivot to conducting some or all classes remotely, then some/all of our synchronous meetings in this class will be recorded to provide necessary information for students in this class. Recordings will be stored on Blackboard [or other secure site] and will only be accessible to students taking this course during this semester.

Privacy

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.

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. Some kinds of participation in online study sites also violate the Mason Honor code: these include accessing exam or quiz questions for this class; accessing exam, quiz, or assignment answers for this class; uploading of any of the instructor's materials or exams; and uploading any of your own answers or finished work. Always consult your syllabus and your professor before using these sites.

LockDown Browser

In the event that we must pivot to remote learning, you will need to use the LockDown Browser and a webcam for online exams. I strongly recommend that you set up the technology at the start of class so you don't have to scramble to do it if the situation suddenly changes.

The webcam can be built into your computer (internal webcam) or can be the type of webcam that plugs in with a USB cable (external webcam). Watch this short video to get a basic understanding of LockDown Browser and the webcam feature. A Quick Start Guide for Students is also available.

You will need the following system requirements for online exams:

- Windows: 10, 8, 7
- Mac: OS X 10.10 or higher
- iOS: 10.0+ (iPad only). Must have a compatible LMS integration [Details].
- Web camera (internal or external) & microphone
- A reliable internet connection

Prior to your first exam, you must install LockDown Browser following the step-by-step instructions. To ensure LockDown Browser and the webcam are set up properly, do the following:

- Start LockDown Browser, log into Blackboard and select this course.
- Locate and select the Help Center button on the LockDown Browser toolbar.
- Run the Webcam Check and, if necessary, resolve any issues or permissions your computer prompts.
- Run the System & Network Check. If a problem is indicated, see if a solution is provided in the Knowledge Base. Further troubleshooting is available through the ITS Support Center.
- Exit the Help Center and locate the practice quiz named [NOTE TO INSTRUCTOR: Create a brief practice quiz and insert name/location of quiz].
- Upon completing and submitting the practice quiz, exit LockDown Browser.

When taking an online exam that requires LockDown Browser and a webcam, remember the following guidelines:

- Ensure you're in a location where you won't be interrupted
- Turn off all other devices (e.g. tablets, phones, second computers) and place them outside of your reach
- Clear your desk of all external materials not permitted — books, papers, phones, other devices
- Before starting the test, know how much time is available for it, and that you've allotted sufficient time to complete it
- Remain at your computer for the duration of the test
- Make sure that your computer is plugged into a power source, or that battery is fully charged.
- If the computer or networking environment is different than what was used previously with the Webcam Check and System & Network Check in LockDown Browser, run the checks again prior to starting the test.

To produce a good webcam video, do the following:

- Do not wear a baseball cap or hat with a brim that obscures your face
- Ensure your computer or tablet is on a firm surface (a desk or table). Do NOT have the computer on your lap, a bed, or any other surface where the device (or you) are likely to move
- If using a built-in (internal) webcam, avoid tilting the screen after the webcam setup is complete
- Take the exam in a well-lit room and avoid backlighting, such as sitting with your back to a window

Remember that LockDown Browser will prevent you from accessing other websites or applications; you will be unable to exit the test until all questions are completed and submitted

CAMPUS LOGISTICS INCLUDING COVID-RELATED CONCERNS

Safe Return to Campus Statement

All students taking courses with a face-to-face component are required to follow the university's public health and safety precautions and procedures outlined on the university Safe Return to Campus webpage (https://www2.gmu.edu/safe-return-campus). Similarly, all students in face-to-face and hybrid courses must also complete the Mason COVID Health Check daily, seven days a week. The COVID Health Check system uses a color code system and students will receive either a Green, Yellow, or Red email response. Only students who receive a "green" notification are permitted to attend courses with a face-to-face component. If you suspect that you are sick or have been directed to self-isolate, please quarantine or get testing. Faculty are allowed to ask you to show them that you have received a Green email and are thereby permitted to be in class.

Campus Closure or Emergency Class Cancelation/Adjustment Policy

If the campus closes, or if a class meeting needs to be canceled or adjusted due to weather or other concern, students should check Blackboard [or other instruction as appropriate] for updates on how to continue learning and for information about any changes to events or assignments.

GMU MASK POLICY – UNIVERSITY POLICY 1415

The university can provide face coverings to employees and students. Employees should contact Environmental Health and Safety to request a face covering. Students can go to the Student Involvement Office in the HUB, Suite 2300, to get a mask.

Guidelines include:

- Everyone, even those who are fully vaccinated, must wear a face covering when inside university property (buildings and vehicles). If you're unvaccinated, you must wear a mask outside, as well, when at an event or if physical distancing can't be maintained.

    o Wear a mask consistently and correctly when:

    o In university facilities (indoor or enclosed space).

    o Outside on university property if you can't maintain physical distance of at least six feet or if you're in close contact with others who are not fully vaccinated.

    o Attending an event.

    o Traveling to/from sites off university property for work or study if in a vehicle with other individuals.

- Follow CDC guidance and University Policy 1415 when choosing a mask and following guidance on when to wear it.

- Don't share face coverings.

Under Policy 1415, exceptions to face covering requirements include when:

- Eating or drinking.

- Communicating with a person who is deaf or hard of hearing via lip reading.

- Someone gets an accommodation from the appropriate Mason office because of a disability.

- Students should contact Disability Services at 703-993-2474.

- Faculty, staff, and volunteers should contact Human Resources and Payroll to work with Employee Relations at 703-993-3878. Employees of contractors should work with their employer regarding accommodations.

- Someone experiences significant trouble breathing, or is unconscious or incapacitated.

- Someone needing university or medical services.

- In a resident's assigned space (no guests), in communal residence hall bathrooms when conducting personal grooming or hygiene and when eating, drinking, or sleeping.

- Alone in a private, enclosed space (e.g., an individual's office, private rehearsal space).

- When swimming or participating in water-based recreational or athletic activities.

- Actively engaging in required skills instruction or intra-squad scrimmages as part of an approved and sanctioned Intercollegiate Athletics team practice or when actively playing in an NCAA sanctioned athletic competition.

- Participating in a religious ritual.

- When participating in a performing art that requires the use of the mouth or unrestricted breathing (e.g., playing a wind instrument, singing, acting, or dance) as part of a Mason-sanctioned performance, or as part of a group rehearsal in a performance space. Masks (or singing masks) and/or wind instrument bell covers are required for classrooms, practice rooms, or other non-performance spaces where several people are present. (A singer's mask is a reusable face covering that helps to contain droplets while creating additional space around the mouth to enable the user to sing comfortably.)

- People with weakened immune systems (regardless of vaccination status) should talk to their health care providers about face coverings and avoiding exposure risks.

University Policy 1415 COVID Public Health and Safety Precautions – Face Coverings establishes Mason requirements for wearing face coverings.

All faculty, staff, students, contractors, employees of contractors, visitors and any other individual on university property age 2 or older must follow Policy 1415 and CDC guidance for face coverings while on University Property.