Data 201
Intro to Data Science

Fall 2024

Syllabus

Class Schedule:

Lecture: M W F, 10:00 AM – 10:50 AM (Section 01)
In-person: Andrews Hall, Room 201

Instructor:

Md Mahfuz Ibn Alam
Office: 2265 Integrated Science Building
Virtual Office: https://cwm.zoom.us/j/7866552578
E-mail: malam02@wm.edu
Phone: 757-221-4554
Office Hours: M F, 3:00 PM – 4:00 PM (In-person);
W, 4:00 PM – 5:00 PM (In-person);
T R, 4 PM – 5 PM (Zoom);
and by appointment.

Teaching Assistants:

Graduate Bullpen TA: Di Yang; Isuru Babarandage; Han Meng
Monday: 8 AM - 2 PM; 5 PM - 10 PM
Tuesday: 8 AM - 3 PM
Wednesday: 8 AM - 7 PM
Thursday: 8 AM - 10 AM; 12 PM - 7 PM
Friday: 9 AM - 3 PM
(ISC 3261)
Graduate TA: Vinay Viajayan (vvijayan@wm.edu) - R, 10:30 AM - 12:30 PM (ISC 1109)
Undergraduate TA: Courtney Maynard - T R, 2 PM - 3 PM, (ISC 3261)

Piazza:

Link: https://piazza.com/wm/fall2024/data20101f24
Access Code: wdnp7utpnhm

Course Overview:

In this course, you will learn the fundamentals of data processing and modeling in the context of Data Science. Emphasis will be placed on careful planning and deliberate decision-making when working with data and building models. Programming will be done in Python, and we will extensively use the scikit-learn collection (https://scikit-learn.org/stable/).

This course will introduce you to various supervised and unsupervised machine-learning techniques, including regression, classification, and clustering methods. By the end of the course, you are not expected to be an expert on any particular technique. Still, you should exhibit a solid high-level understanding of the goals of each method, be able to

determine when a specific type of model is more or less suitable to a real-world problem, and, most importantly, demonstrate keen attention to detail when working with data. A very strong emphasis will be placed on understanding why we are doing what we are doing.

Learning Objectives:

1. The course will provide you with a critical understanding of the essential aspects and techniques in data preprocessing, modeling, and data use.
2. You will learn to apply specific methods/algorithms and interpret the results.
3. You will work with real-world problems that are being engaged with by contemporary problem-solvers and decision-makers.

Course Schedule:

	What we will discuss/do in class	**What to prepare (before class)**
Lecture 1 (August 28th)	Introduction and Syllabus	
Lecture 2 (August 30th)	Getting Ready; Brief Review of Python	
Lecture 3 (September 4th)	Brief Review of Python	Read: Python for Data Analysis Chapter 5 (.pdf provided on BB)
Lecture 4 & 5 (September 6th & 9th)	Pandas Data Frames	Read: https://www.dunderdata.com/blog/selecting-subsets-of-data-in-pandas-part-1
Lecture 6 (September 11th)	Pandas Data Frames; Numpy Arrays	Read: Python for Data Analysis Section 4.1 and Appendix A excerpt (.pdf provided on BB)
Lecture 7 (September 13th)	Numpy Arrays	
Lecture 8 - 10 (September 16th, 18th & 20th)	Describing Data	Read: Python for Data Analysis Section 9.1 and 9.2 (.pdf provided on BB)
Lecture 11 & 12 (September 23rd & 25th)	Describing Data; Feature Scaling	
Lecture 13 (September 27th)	Intro to Modeling	
Lecture 14 & 15 (September 30th & October 2nd)	OLS	
Lecture 16 - 18 (October 4th, 7th & 9th)	Model Validation	Read Python for Data Analysis Section 13.1(.pdf provided on BB)

Lecture 19 & 20 (October 14th & 16th)	Regularization	
Lecture 21 (October 18th)	Q/A and Review for the Midterm	
Lecture 22 (October 21st)	Work Day Midterm	
Lecture 23 - 26 (October 23rd, 25th, 28th & 30th)	Dimensionality Reduction - PCA/tSNE	
Lecture 27 - 29 (November 1st, 4th & 6th)	Classification: Logistic Regression	
Lecture 30 & 31 (November 8th & 11th)	Classification: KNN	
Lecture 32 - 35 (November 13th, 15th, 18th & 20th)	Classification: Decision Trees/RF	
Lecture 36 (November 22nd)	Feature Importance	
Lecture 37 - 39 (November 25th, December 2nd & 4th)	Clustering	
Lecture 40 (December 6th)	Q/A and Review for the Final	

Important Dates:

1. **Labor Day (no classes) September 2nd.**
2. **The add-and-drop deadline is September 9th.**
3. **Fall Break (no classes) October 10th - October 13th.**
4. **The Midterm Exam is due October 22nd, 11:55 PM.**
5. **The withdrawal deadline is October 28th.**
6. **Thanksgiving (offices closed, remote (online) instruction) November 25th – 26th**
7. **Thanksgiving (no classes) November 27th - December 1st.**
8. **The last day of classes is December 7th.**
9. **The Final Exam is due December 13th, 11:55 PM.**

Assignment Due Dates:

HW 0	September 3rd, 11:55 PM
HW 1	September 13th, 11:55 PM
HW 2	September 24th, 11:55 PM
HW 3	October 4th, 11:55 PM
HW 4	October 17th, 11:55 PM
Midterm	October 22nd, 11:55 PM
HW 5	October 31st, 11:55 PM
HW 6	November 12th, 11:55 PM
HW 7	November 24th, 11:55 PM
HW 8	December 6th, 11:55 PM
Final	December 13th, 11:55 PM

Assignments: The assignments will be posted on Blackboard in a timely manner before the due date. You will have at least five days to complete each assignment. If you have any questions, please actively engage in Q/As on Piazza. Assignments will involve a lot of short answer or multiple-choice questions and developing code to determine the answer. In contrast, some assignments may require that you design more ample programming approaches. For such questions, you will write your own code to solve the problem and get feedback on how to improve. In this case, grading includes partial credit.

Midterm and Final Exam: An online midterm (due on October 22nd) and final (due on December 13th) will test your knowledge of the concepts and Machine Learning algorithms presented during the course. The midterm and the final are administered via Blackboard.

Artificial Intelligence: The use or incorporation of any AI-generated content (from ChatGPT, Bard, Dall-e, etc.) in graded assignments or exams is prohibited.

Quiz Dates:

QZ 1 (Lecture 3,4,5)	September 11th
QZ 2 (Lecture 6,7,8)	September 18th
QZ 3 (Lecture 9,10,11)	September 25th
QZ 4 (Lecture 12,13,14)	October 2nd
QZ 5 (Lecture 15,16,17)	October 9th
QZ 6 (Lecture 18,19,20)	October 18th
Midterm	October 22nd, 11:55 PM
QZ 7 (Lecture 23,24,25)	October 30th
QZ 8 (Lecture 26,27,28)	November 6th
QZ 9 (Lecture 29,30,31)	November 13th
QZ 10 (Lecture 32,33,34)	November 20th
Makeup (Lecture 35,36,37)	December 4th
Final	December 13th, 11:55 PM

Quizzes: Quizzes will involve mainly short answer or multiple-choice questions. Every Wednesday last 15-20 minutes will be booked for quizzes. The quizzes will be administered via Blackboard.

Participation: You are expected to attend class regularly and will need to do so to learn the material needed to meet expectations for assignments. We will have occasional classwork that will count towards participation. Participation does not necessarily mean volunteering in class – sometimes I will call on you randomly, sometimes I will use PollEverywhere to allow you to respond electronically, and you can participate on the course Piazza page by answering questions posed by other students. You also participate by being engaged – this means not working on other coursework doing class, and basically not doing anything else besides paying attention and taking notes.

Evaluation:

Class Participation	**5%**
Assignments	**35%**

Quizzes	**20%**
Midterm	**20%**
Final	**20%**

The lowest two score from the Assignments and Quizzes section will be dropped to accommodate an unexpected short-term illness or a class absence. **There will be extra credit opportunities in the Assignments, Quizzes, Midterm and Final. Keep a sharp eye.**

Letter Grades:

93 – 100 %	**A**	73 – 76.99%	**C**
90 – 92.99%	**A-**	70 – 72.99%	**C-**
87 – 89.99%	**B+**	67 – 69.99%	**D+**
83 – 86.99%	**B**	63 – 66.99%	**D**
80 – 82.99%	**B-**	60 – 62.99%	**D-**
77 – 79.99%	**C+**	0 – 59.99%	**F**

Expectations: All students are strongly encouraged to bring their laptops/notebook computers to class since we plan on having a significant amount of active coding. Do not forget pen and papers.

We will address instructor absence/illness by scheduling remote (online) classes for no more than a week. In extreme cases, a substitute instructor will be arranged.

Class Absences: I will take attendance at the end of every class. This is to find out if everyone is okay or not? If you miss a class or two, check the material covered in class posted on Blackboard. For additional help, you can make an appointment with the TA or the instructor. Long-term absences need to be documented via the Dean of Students Office.

Student Accessibility: William & Mary accommodates students with disabilities in accordance with federal laws and university policy. Any student who feels they may need an accommodation based on the impact of a learning, psychiatric, physical, or chronic health diagnosis should contact Student Accessibility Services staff at 757-221-2512 or at sas@wm.edu to determine if accommodations are warranted and to obtain an official letter of accommodation. For more information, please see www.wm.edu/sas.

Writing and Communication: Writing and Communication Center is located on the first floor of Swem Library, is a free service provided to W&M students. Trained consultants offer individual assistance with writing, presentation, and other communication assignments across disciplines and at any stage, from generating ideas to polishing a final product. The WCC's goal is to help you become a better writer and communicator. To make an appointment, visit the WCC webpage (www.wm.edu/wcc).

Honor Code: Academic integrity is at the heart of the university, and we all are responsible for upholding the ideals of honor and integrity. The student-led honor system is responsible for resolving any suspected violations of the Honor Code, and I will report all suspected instances of academic dishonesty to the honor system. The Student Handbook

(www.wm.edu/studenthandbook) includes your responsibilities as a student. Your full participation and observance of the Honor Code are expected. To read the Honor Code, see www.wm.edu/honor.

Student Success: Student success supports each student's personal growth, development, engagement, and belonging by providing holistic guidance as they navigate their own W&M journey. It encompasses the offices of Academic Wellbeing, Care Support Services, and Student Accessibility Services.

For academic support such as tutoring, time management, study skills, and academic coaching, please contact Academic Wellbeing at wm.edu/academicwellbeing (academicwellbeing@wm.edu).

For concerns about the wellbeing of a member of the William & Mary community or to seek assistance for interpersonal, academic, and wellness challenges, please contact Care Support Services at wm.edu/care (care@wm.edu).

For accommodation needs or questions, please contact Student Accessibility Services at wm.edu/sas (sas@wm.edu).

Student Health: William & Mary recognizes that students juggle different responsibilities and can face challenges that make learning difficult. There are many resources available at W&M to help students navigate emotional/psychological, physical/medical, material/accessibility concerns, including:

1. The W&M Counseling Center at (757) 221-3620. Services are free and confidential.
2. The W&M Health Center at (757) 221-4386.
3. To seek assistance for interpersonal, academic, and wellness challenges, please contact Care Support Services at wm.edu/care (care@wm.edu).
4. For a list of other resources available to students, see here or:

Additional Information: William & Mary values inclusiveness most highly and believes that diversity is critical to equity and the pursuit of academic excellence. Our goal is to create and sustain an environment in which diversity can thrive. A diverse faculty, student body, administration, and curriculum together foster learning and enhance excellence. We seek to promote an environment of inclusion and to maintain a safe, nurturing community that is respectful of our differences and what we share.

Possible Changes: Any changes to the following course schedule or due dates will be announced in class and on Blackboard ahead of time.

Books:

1. Python for Data Analysis, 3E (https://wesmckinney.com/book/)

2. Introduction to Data mining, 2E (https://www-users.cse.umn.edu/~kumar001/dmbook/index.php)

3. Introduction to Data mining, 3E (https://www.ceom.ou.edu/media/docs/upload/Pang-Ning_Tan_Michael_Steinbach_Vipin_Kumar_-_Introduction_to_Data_Mining-Pe_NRDK4fi.pdf)