Data 201
Intro to Data Science

Spring 2025

Syllabus

Class Schedule:
Lecture: T R, 12:30 PM – 01:50 AM (Section 01)
In-person: John E. Boswell Hall, Room 220

Instructor:
Md Mahfuz Ibn Alam
Office: 2265 Integrated Science Building
Virtual Office: https://cwm.zoom.us/j/7866552578
E-mail: malam02@wm.edu
Office Hours: T R, 2 PM – 3 PM (In-person);
W, 2:00 PM – 5:00 PM (In-person);
and by appointment.

Teaching Assistants:
Graduate TA: Kibiwott Koech
Email: kkoech@wm.edu
Office Hours: Tuesdays, 6-8 PM
Location: ISC 1291

Piazza:
Link: https://piazza.com/wm/spring2025/data20101s25
Access Code: kpoe1v7ifwm

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.

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.

Class Format:

The class will be in-person. Each class will take the following format:

- **Reading:** Before the class, you will be pointed to some reading materials (see "Reading Materials" in the course schedule) . Reading is not required but highly recommended to do before the class.
- **Summary/Elaboration/Q&A:** In the class, the instructor will summarize important points from the reading material, elaborating on details that were not included in the reading while fielding any questions.

Grading:

Your final grade will be dependent on:

- **Midterm:** 25%. An online midterm (due on March 19th) will test your knowledge of the concepts and Machine Learning algorithms presented during the course. The midterm will be administered via Blackboard.
- **Final:** 25%. An online final (due on May 11th) will test your knowledge of the concepts and Machine Learning algorithms presented during the course. The final will be administered via Blackboard.
- **Assignments:** 45%. There will be seven assignments in this class. Each assignment must be completed independently. 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 answers and developing code to determine the answer.
    - A1: Pandas and Numpy (6%).
    - A2: Statistics and Scaling (6%).
    - A3: Linear Regression (6%).
    - A4: Model Validation and Regularization (6%).
    - A5: Dimensionality Reduction (7%).
    - A6: Classification: Logistic Regression & KNN (7%).
    - A7: Classification: Decision Trees (7%).
- **Participation:** 5%. 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.

- There will be extra credit opportunities in the Assignments and Exams. Keep a sharp eye.

Late Day Policy for Assignments:

In case there are unforeseen circumstances that don't let you turn in your assignment on time, 5 late days total over the seven assignments will be allowed. Notes: (1) The late days cannot be used fractionally, e.g., submitting the assignment 1 hour late will incur 1 late day; (2) You may want to save the late days for assignments, which are generally harder than the others. Late days will be applied greedily and automatically (no need to contact the instructor). Assignments that are late beyond the allowed late days will be graded down by 5% per day. In the case of a serious illness or other excused absence, as defined by university policies (including providing necessary evidence), coursework submissions will be accepted late by the same number of days as the excused absence. However, students should inform the instructor as soon as they can, ideally before the due date.

Class Attendance Policy:

As we will have classworks for some classes, attendance is highly suggested.

Readings:

Students should be able to understand the course content just by following the lecture and by doing the readings. However, the following textbooks serve as good references.

- Introduction to Data mining, 1E [online][pdf];
- Introduction to Data mining, 2E [online];
- Python for Data Analysis, 3E [online];

Course Schedule:

#	Date	Topic	Reading Materials	Assignment Dates
1	01/23	Introduction and Syllabus		**A0 is Out**
2	01/28	Getting Ready and Python		
3	01/30	Python and Pandas Data Frames	Python for Data Analysis Chapter 5 (.pdf provided on BB)	
4	02/04	Pandas Data Frames		**A1 is Out**
5	02/06	Pandas Data Frames		
6	02/11	Numpy Arrays and Describing Data	Python for Data Analysis Section 4.1	

			and Appendix A excerpt (.pdf provided on BB) Python for Data Analysis Section 9.1 and 9.2 (.pdf provided on BB)	
7	02/13	Describing Data: Univariate		**A1 is Due; A2 is Out**
8	02/18	Describing Data: Univariate and Describing Data: Bivariate		
9	02/20	Describing Data: Bivariate and Feature Scaling		
10	02/25	Intro to Modeling and OLS		**A2 is Due; A3 is Out**
11	02/27	Intro to Modeling and OLS		
12	03/04	Model Validation	Python for Data Analysis Section 13.1(.pdf provided on BB)	
13	03/06	Model Validation and Regularization		**A3 is Due; A4 is Out**
14	03/18	Q/A and Review for the Midterm		**Midterm Out on 17th.** **Midterm Due on 19th.**
15	03/20	Regularization		
16	03/25	Dimensionality Reduction - PCA	https://dimensionality-reduction-293e465c2a3443e8941b016d.vercel.app/ ; https://setosa.io/ev/principal-component-analysis/	
17	03/27	Dimensionality Reduction - PCA and tSNE	https://distill.pub/2016/misread-tsne/	**A4 is Due; A5 is Out**
18	04/01	Dimensionality Reduction - tSNE		
19	04/03	Classification: Logistic Regression	Logistic Regression Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (.pdf provided on BB)	

20	04/08	Classification: Logistic Regression		**A5 is Due; A6 is Out**
21	04/10	Classification: Logistic Regression and Classification: KNN	Classification Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (.pdf provided on BB)	
22	04/15	Classification: KNN		
23	04/17	Classification: Decision Trees	Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition (Decision Trees)	**A6 is Due; A7 is Out;**
24	04/22	Classification: Decision Trees		
25	04/24	Classification: Random Forests	Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition (Ensemble Learning and Random Forests)	
26	04/29	Feature Selection		
27	05/01	Q/A and Review for the Final		**A7 is Due**

Important Dates:

1. **The first day of add-and-drop is January 21st.**
2. **The first day of classes is January 22nd.**
3. **The last day of add-and-drop is January 31st.**
4. **The first day of the withdrawal period is February 1st.**
5. **Spring Break (no classes) March 8th - March 16th.**
6. **The last day of the withdrawal period is March 24th.**
7. **The last day of classes is May 2nd.**
8. **Final grades due for all students by 9 a.m is May 20th.**

Artificial Intelligence:

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

Letter Grades:

95 – 100 %	**A**	72 – 74.99%	**C**
90 – 94.99%	**A-**	70 – 71.99%	**C-**
85 – 89.99%	**B+**	60 – 69.99%	**D+**
82 – 84.99%	**B**		
80 – 81.99%	**B-**		
75 – 79.99%	**C+**	0 – 59.99%	**F**

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 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:

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.

Name and Pronouns Statement:

If you wish, please share your name and gender pronouns with me and indicate how best to address you in class and via email. I use "he/him/his" for myself and you may address me as “Alam”, “Dr. Alam” in email and verbally.

Recording and/or sharing class materials:

Some kinds of participation in online study sites violate the Honor code: these include accessing exam or quiz questions for this class; accessing exam, quiz, or assignment answers for this class; uploading 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.

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.

Possible Changes:

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

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).

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).

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.