

DATS 6103.10- Introduction to Data Mining

Spring 2026

COURSE AND CONTACT INFORMATION

Section	10
Time	Monday, 6:10 – 8:40 pm, Jan 12-Apr 27
Location	Tompkins Hall (TOMP) 309

INSTRUCTOR

Name: Dr. Ning Rui

Email: ning.rui@gwu.edu

Office hours: Monday 5-6 pm ET at TOMP 309 & F 5-6 pm ET on Zoom, by appointment

Web conference URL: <https://gwu-edu.zoom.us/my/ningrui1>

COURSE DESCRIPTION

This course introduces the concepts of data mining as applied to data science, using Python as the programming language. We start with an overview of the python programming language itself. With weekly hands-on activities and homework exercises, we cover the concepts and the practical knowledge to perform data mining in real world situations, then use preprocessing data wrangling techniques to make them suitable for typical data analysis. We then use the data in a variety of data modeling and visualization schemes. As an introduction class, we aim to cover a broad spectrum of this emerging field, including how to integrate genAI tools in your data analytic workflow. Students will continue to use these concept and skills to the more advanced courses and in their own projects.

LEARNING OUTCOMES

As a result of completing this course, students will be able to

1. Apply data mining concepts and techniques to real-life problems;
2. Demonstrate knowledge of python programming and basic object-oriented programming concepts by creating python codes for common tasks;
3. Write python codes and leverage GenAI to perform data analysis including data pre-processing, data wrangling and model building of various machine learning algorithms;
4. Produce visual charts and graphs of real-world data using python programming;
5. Analyze data to find information that are relevant and consequential using python;
6. Synthesize knowledge gained through collaboration with peers on group projects;
7. Communicate data analysis results to a general audience through presentations to peers and instructor.

INDEPENDENT OR OUT-OF-CLASS LEARNING

Average minimum amount of out-of-class or independent learning expected per week is 5 hours, for our 2.5 hour of classroom meeting. You are strongly encouraged to make use of our office hours, the TA's help, as well as other resources including but limited to those offered by the library services to help you learn. Do not wait until it is too late to get help.

ATTENDANCE REQUIREMENT

Students are expected to attend all scheduled classes. Absences in a course may affect one's final grade. The student is responsible for all assigned work in the course regardless of excused or unexcused absences. Excessive absences can result in a failing grade.

COURSE PREREQUISITE

Although we go through the basic of python programming from the start, you will find the class more rewarding if you have good working knowledge of python or other programming experience yourself. If you feel the need to catch up more on the programming side, consider spending a few hours to complete an online course or two on python programming. Lynda.com (now owned by LinkedIn) access is included in your GWU account. Other useful resources include Udemy, EdX, Coursera, Dataquest, Datacamp, etc.

Here is a list of software you need.

- a. Install Python3 on your laptop.
- b. Install Jupyter Notebook on your laptop (for ipython, interactive-python).
- c. [Check to see if git is installed on your laptop](#). (Type “which git” on a terminal.)
- d. [Set up a Github account](#), and install GitHub Desktop for easy access to repositories from your local environment.
- e. Install [Visual Studio Code](#) for programming projects. Make sure you install and enable other useful extensions to interpret python, Markdown, Jupyter notebook, etc.

COURSE SCHEDULE

Date	Topics
Jan 12	Linear Algebra/Statistics/Probability key concept review Intro to main platforms – Blackboard, MS Teams/OneNote, Git Set up IDE (VS Code)
Jan 19	No Class – MLK Day
Jan 26	Python programming basic Variables and Data type Logic, Conditional, Loops
Feb 2	Quiz 1 Object Oriented Programming Recursive loops, Functions Class / Objects
Feb 9	Basic Scientific Programming Numpy, Pandas, Scipy
Feb 16	No Class – Presents’ Day
Feb 23	Quiz 2 Data wrangling, Data imputation, Categorical encoding Data visualization, Matplotlib, Seaborn
Mar 2	Data preprocessing & cleaning; Missing data treatment Filter, Sort, Group By Dimensionality Reduction, PCA, Feature Selection and ranking Mid-term Exam released (Submit by Mar 16)
Mar 9	No Class – Spring Break
Mar 16	Quiz 3 EDA, Significance testing (t-test, ANOVA, ANCOVA, chi-square test, non-parametric test)
Mar 23	Causal inference, a/b testing, difference-in-differences, instrumental variables, propensity score matching, double ML Final Project Proposal Due
Mar 30	Quiz 4 <i>Unsupervised Learning:</i> Apriori Algorithm, K-Means Clustering, Hierarchical Clustering

Apr 6	<i>Supervised Learning:</i> Decision Tree, Random Forest Regression (linear and logistic)
Apr 13	Quiz 5 KNN, Naïve Bayes classifier, Support Vector Machine Ensemble models: Gradient Boosting vs. AdaBoost
Apr 20	Neural Networks, Transformers, LLM, GenAI
Apr 27	Last Class: Final Project Presentation

TEXTBOOKS AND COURSE RESOURCES

These are some free online textbooks for your reference. We do not use one single textbook. The following are a few recommended electronic books that can be accessed via [Safari Tech Books](#) by O'Reilly.

1. McKinney, W. (2017). [Python for Data Analysis](#). O'Reilly Media, Inc. (PDA, Safari Tech)
2. Matthes, E. (2019). [Python Crash Course, 2nd ed.](#) (Optional reference on Git in Appendix D). O'Reilly Media, Inc. (PCC, Safari Tech)
3. Lubanovic, B. (2019). [Introducing Python](#). O'Reilly Media, Inc. (IP, Safari Tech)
4. (Optional. General Reference on Pandas) Harrison, M., Petrou, T. (2020). [Pandas 1.x Cookbook](#). O'Reilly Media, Inc. (P1C, Safari Tech)
5. Hajba, G. S. (2018). [Website Scraping with Python: Using BeautifulSoup and Scrapy](#). O'Reilly Media, Inc. (WSP, Safari Tech)
6. Raschka, G., Mirjalili, V. (2019). [Python Machine Learning](#) - Third Edition. O'Reilly Media, Inc. (PML, Safari Tech)
7. Ciaburro, G., Joshi, P. (2019). [Python Machine Learning Cookbook](#) - Second Edition. O'Reilly Media, Inc. (PMLC, Safari Tech)

LEARNING ASSIGNMENTS

Homework assignments: These assessments are used to reinforce learning and give feedback on areas for improvement. The homework is assigned roughly on a weekly basis. Follow the instructions on each homework, submit your work as a .py python file on blackboard. Most assignments will be due before the next class period. Homework assignments are open book, but **should be completed independently without exchanging information with any other students.**

Discussion Boards: Throughout the course, you will interact with classmates to discuss various topics on data science, python, and programming in general. Posts and responses will be based on concepts learned in lectures and readings, or other times to expand on broader areas that will help you grow as a data scientist and a competent programmer. You are expected to post original ideas and respond substantively to peers.

Quizzes: There will be short quizzes during some class periods. They are designed to test your basic skills you learned in the class sections and homework you completed the 1-2 weeks before. Keeping up with lectures and homework will make you successful in the quizzes. You may refer to any class notes or reference materials independently, but **are not allowed to interact or exchange any information with any other students in person, via email or any social media platforms.**

Class Project: Teams are formed in groups of 4 students. Each team will present their project at the end of semester (15 minutes), on a topic of your choosing. A one-page proposal of the project, accompanied by a one-page description of the task delegation among the team members, should be submitted by **Oct 30**. Students will build communication skills as well as python data mining skills.

Mid-term and Final Exams: Two take-home exams will be given to ensure students are gaining knowledge from the coding exercises and from class lectures. You may look up your class notes or reference materials independently, but **are not allowed to interact or exchange any information with any other students in person, via email or any social media platforms.** If your submission is found to be largely identical to someone else's work for any parts or the entire problem set, your submission will receive a mark of zero and a report will be filed to Office of Student Rights and Responsibilities (SRR).

GRADING

Your final grade will be determined by:

- ✓ Attendance (10%)
- ✓ Assignments (10%)
- ✓ Discussion boards (6%)
- ✓ Quizzes (16%): The highest 4 scores out of the 5 quizzes count toward your grade
- ✓ Midterm Exam (15%)
- ✓ Final Exam (18%)
- ✓ Final Project (25%)
 - Proposal (2%)
 - Presentation (7%)
 - Programming and Data mining (6%)
 - Summary report (6%)
 - Git collaboration (4%)
 - Presentation and git collaboration grades are earned individually. Other three scores are earned as a team. Upon evidence of unfair/disproportionate amount of work among teammates, I reserve the right to delegate the earned points by the team differently should such unfortunate situation arises.

Letter Grades

A	A-	B+	B	B-	C+	C	C-	D	F
≥ 93	90 – 92	87 – 89	83 – 86	80 – 82	77 – 79	73 – 76	70 – 72	60 – 69	< 60

UNIVERSITY POLICIES

OBSERVANCE OF RELIGIOUS HOLIDAYS

Students must notify faculty during the first week of the semester in which they are enrolled in the course, or as early as possible, but no later than three weeks prior to the absence, of their intention to be absent from class on their day(s) of religious observance. If the holiday falls within the first three weeks of class, the student must inform faculty in the first week of the semester. For details and policy, see “Religious Holidays” at provost.gwu.edu/policies-procedures-and-guidelines.

ACADEMIC INTEGRITY CODE

Academic integrity is an essential part of the educational process, and all members of the GW community take these matters very seriously. As the instructor of record for this course, my role is to provide clear expectations and uphold them in all assessments. Violations of academic integrity occur when students fail to cite research sources properly, engage in unauthorized collaboration, falsify data, and otherwise violate the Code of Academic Integrity. If you have any questions about whether or not particular academic practices or resources are permitted, you should ask me for clarification. If you are reported for an academic integrity violation, you should contact the Office of Student Rights and Responsibilities (SRR) to learn more about your rights and options in the process. Consequences can range from failure of assignment to expulsion from the university and may include a transcript notation. For more information, please refer to the SRR website (<https://studentconduct.gwu.edu/academic-integrity>), email rights@gwu.edu, or call 202-994-6757.

USE OF GENERATIVE AI TOOLS

1. Work submitted for evaluation in this course is represented as your own intellectual product. You may not submit content (e.g., ideas, text, code, images) for performance evaluation that was generated, in whole or in part, by Generative Artificial Intelligence tools (such as ChatGPT or other genAI tools or large language models). Doing so without instructor’s explicit permission constitutes cheating under the [Code of Academic Integrity](#) and is therefore prohibited. Examples of conduct that is prohibited unless explicitly permitted by the instructor include: A student types a prompt into a genAI tool and pastes *all or part of the generated content into their answer on an out-of-class assessment or test*; a student types a prompt into a genAI tool and incorporates all or part of the generated content into a paper or report without proper attribution to the genAI tool.

2. You are permitted to use genAI tools to generate content that is not submitted to the instructor for evaluation. For example, using genAI tools to study for exams and homework assignments is permitted.
3. The use of genAI tools during any *assessment* (e.g., exam, quiz) whether taken in the classroom or elsewhere, constitutes cheating under the [Code of Academic Integrity](#) and is therefore prohibited.

USE OF ELECTRONIC COURSE MATERIALS AND CLASS RECORDINGS

Students are encouraged to use electronic course materials, including recorded class sessions, for private personal use in connection with their academic program of study. Electronic course materials and recorded class sessions should not be shared or used for non-course related purposes unless express permission has been granted by the instructor. Students who impermissibly share any electronic course materials are subject to discipline under the Student Code of Conduct. Please contact the instructor if you have questions regarding what constitutes permissible or impermissible use of electronic course materials and/or recorded class sessions. Please contact Disability Support Services at disabilitysupport.gwu.edu if you have questions or need assistance in accessing electronic course materials.

ACADEMIC SUPPORT

WRITING CENTER

GW's Writing Center cultivates confident writers in the University community by facilitating collaborative, critical, and inclusive conversations at all stages of the writing process. Working alongside peer mentors, writers develop strategies to write independently in academic and public settings. Appointments can be booked online at gwu.mywconline.com.

ACADEMIC COMMONS

Academic Commons provides tutoring and other academic support resources to students in many courses. Students can schedule virtual one-on-one appointments or attend virtual drop-in sessions. Students may schedule an appointment, review the tutoring schedule, access other academic support resources, or obtain assistance at academiccommons.gwu.edu.

SUPPORT FOR STUDENTS OUTSIDE THE CLASSROOM

DISABILITY SUPPORT SERVICES (DSS) 202-994-8250

Any student who may need an accommodation based on the potential impact of a disability should contact Disability Support Services at disabilitysupport.gwu.edu to establish eligibility and to coordinate reasonable accommodations

COUNSELING AND PSYCHOLOGICAL SERVICES 202-994-5300

GW's Colonial Health Center offers counseling and psychological services, supporting mental health and personal development by collaborating directly with students to overcome challenges and difficulties that may interfere with academic, emotional, and personal success. healthcenter.gwu.edu/counseling-and-psychological-services.

SAFETY AND SECURITY

- In an emergency: call GWPD 202-994-6111 or 911
- For situation-specific actions: review the Emergency Response Handbook at: safety.gwu.edu/emergency-response-handbook
- In an active violence situation: Get Out, Hide Out, or Take Out. See go.gwu.edu/shooterprep
- Stay informed: safety.gwu.edu/stay-informed