

DATS 2103: Data Mining for Data Science with Python

Syllabus for Fall 2025

Class Info

Time: Mon Wed 11:10-12:25 & 2:20-3:35

Office Hours: Tues 2-4 & Thurs 11-1

Additional office hours available by appointment

Instructor Info

Instructor: Angelica M. Walker, Ph.D.

Office: Samson Hall Room 307

Email: angelica.walker@gwu.edu

Course Overview

This course introduces undergraduate students to data mining in Python, aimed for students within the GW Data Science Program. This includes practical applications of data science applied to a variety of fields, including astronomy, biology, physics, economics, and more. Students will gain a comprehensive understanding of data mining techniques, including data pre-processing, data analysis, visualization, and model evaluation. Key topics include regression, classification, clustering, predictive modeling, statistical inference, dimension reduction, cross-validation, multiple hypothesis testing, and application of linear regression, logistic regression, polynomial regression, K -nearest neighbors, generalized additive models, random forests, support vector machines, and neural networks. At the end of the course, students will have a thorough understanding of many data mining methods and be capable of applying them to any data set.

Course Prerequisites

CSCI 1012, DATS 1001, MATH 1232, and (STAT 1051 or STAT 1053 or STAT 1111 or STAT 1127).

Course Materials

- *An Introduction to Statistical Learning with Applications in Python* ([Link](#))

Learning Outcomes

Successful students will become familiar with:

- data pre-processing and basic analysis techniques
- applications of Python, Jupyter Lab, and scikit-learn in data science
- core concepts of regression and classification
- utilizing supervised and unsupervised statistical learning techniques for prediction and inference
- how to select appropriate machine learning methods and evaluate their performances

Schedule

Class Date	Lecture Topic	Quiz/Exam	Assignments
M August 25	Intro to Statistical Learning		
W August 27			
M September 1	Labor Day, No Class		
W September 3	Linear Regression		
M September 8		Quiz 1	
W September 10	Classification		Homework 1 Due
M September 15		Quiz 2	
W September 17	Resampling Methods	Quiz 3	Homework 2 Due
M September 22	Regularization		Lab 1 Due
W September 24		Quiz 4	
M September 29	Moving Beyond Linearity		
W October 1		Quiz 5	
M October 6	Tree Based Methods		Homework 3 Due
W October 8		Quiz 6	
M October 13	Review for Midterm		
W October 15	In Class Midterm	In Class Midterm	
M October 20	Support Vector Machines		Lab 2 Due
W October 22		Quiz 7	
M October 27	Deep Learning		Homework 4 Due
W October 29			
M November 3	Survival Analysis	Quiz 8	
W November 5			Homework 5 Due
M November 10	Unsupervised Learning	Quiz 9	Lab 3 Due
W November 12			
M November 17	Multiple Hypothesis Testing	Quiz 10	
W November 19			Final Project Out
November 24,26	Thanksgiving Break, No Class		
M December 1	Multiple Hypothesis Testing	Quiz 11	Homework 6 Due
W December 3	Final Project Work In Class		Lab 4 Due
M December 8			
M December 15	Final Exam Day, No Class		Final Project Due

*This schedule is subject to change.

Evaluation

Category	Percentage
Quizzes (Best 9 Scores)	15%
Labs (4)	25%
Homework (6)	20%
Midterm (1)	15%
Final Project (1)	25%

Students will be evaluated by quizzes at the end of each subject of instruction, corresponding to chapters in the textbook. There will be eleven quizzes in total, but only the top nine scores will be used for grading. Two of the lowest quiz grades will be dropped at the end of the semester. These quizzes account for 15% of the final grade and are designed to be simple and completed within five to ten minutes. The four in class labs may be completed in small groups no larger than three students within the same section and accounts for 25% of the final grade. These labs are due at 11:59 pm on the dates given above. There are six homework assignments that account for 20% of the final grade and must be completed alone. These are due at 11:59 pm on the dates given above. This course also contains one midterm worth 15% of the final grade. There is no final exam, a final project worth 25% of the final grade will be due on the final exam day. This final project must be completed alone and will contain an oral component.

Course Policies

Work Load

This 3-credit course includes 2.5 hours of instructional time and 5-10 hours of independent learning per week.

Attendance

Attendance is not explicitly required, however make up quizzes will not be available if missed. I strongly recommend coming to class every day.

Late Assignments

Assignments will be penalized by 20% each day it is late, up to 5 days. No partial credit is given for assignments submitted 5 days after the deadline. Extensions may be granted under extenuating circumstances if requested in advanced.

Generative Artificial Intelligence

By submitting work for evaluation in this course, you represent it as your own intellectual product. You may not submit for evaluation any content (e.g., ideas, text, code, images) that was generated, in whole or in part, by Generative Artificial Intelligence tools (including, but not limited to, ChatGPT and other large language models). Doing so in this course constitutes cheating under the [George Washington University Code of Academic Integrity](#). Additionally, the instructor does not consent for any course materials to be used as input or training data for Generative Artificial Intelligence tools unless explicitly specified.

Data Science Help Desk

Visit the Data Science Helpdesk for technical assistance, available to undergraduate and graduate students in the program, or taking our DATS classes. Technical support is available throughout the semester in Room 304, Samson Hall. For current schedule, please visit the <https://dats.columbian.gwu.edu/>. You can also schedule a remote support session by sending an email to dshelpdesk@gwu.edu.

University Policies

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 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 Conflict Education and Student Accountability (CESA), formerly known as 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, refer to the CESA website at students.gwu.edu/code-academic-integrity or contact CESA by email cesa@gwu.edu or phone 202-994-6757.

University policy on 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 provost.gwu.edu/policies-procedures-and-guidelines.

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. Contact the instructor if you have questions regarding what constitutes permissible or impermissible use of electronic course materials and/or recorded class sessions. Contact Disability Support Services at disabilitysupport.gwu.edu if you have questions or need assistance in accessing electronic course materials.

Academic Commons

Academic Commons is the central location for academic support resources for GW students. To schedule a peer tutoring session for a variety of courses visit go.gwu.edu/tutoring. Visit academiccommons.gwu.edu for study skills tips, finding help with research, and connecting with other campus resources. For questions email academiccommons@gwu.edu.

GW Writing Center

GW 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 writingcenter.gwu.edu/appointments.

Disability Support Services (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.

Student Health Center (202-994-5300, 24/7)

The Student Health Center (SHC) offers medical, counseling/psychological, and psychiatric services to GW students. More information about the SHC is available at healthcenter.gwu.edu. Students experiencing a medical or mental health emergency on campus should contact GW Emergency Services at 202-994-6111, or off campus at 911.

GW Campus Emergency Information

GW Emergency Services: 202-994-6111

For situation-specific instructions, refer to GW's Emergency Procedures guide.

GW Alert

GW Alert is an emergency notification system that sends alerts to the GW community. GW requests students, faculty, and staff maintain current contact information by logging on to alert.gwu.edu. Alerts are sent via email, text, social media, and other means, including the Guardian app. The Guardian app is a safety app that allows you to communicate quickly with GW Emergency Services, 911, and other resources. Learn more at safety.gwu.edu.

Protective Actions

GW prescribes four protective actions that can be issued by university officials depending on the type of emergency. All GW community members are expected to follow directions according to the specified protective action. The protective actions are Shelter, Evacuate, Secure, and Lockdown. Learn more at safety.gwu.edu/gw-standard-emergency-statuses.