DATS 6103 – Introduction to Data Mining

"Linear algebra is everywhere in the world of learning from data."
– Gilbert Strang

Course and Contact Information

Term		Fall 2025
Section	11 (CRN: 32545)	12 (CRN: 33719)
Time	Tue (6:10–8:40 p.m.)	Tue (3:30–6:00 p.m.)
Location	1776 G C-103	1957 E 211

Instructor

Name: Dr. Sushovan Majhi
Office: Samson Hall 313 (easier to enter from Corcoran 2nd floor, through the end-of-hallway stairs)
Email: s.majhi@gwu.edu
Office hours: By appointment at calendly.com/sushovan4 (In-person or conference calls)

Teaching Assistant

Name: Aishwarya Maddula
Email: aishwaryam@email.gwu.edu
Office hours: Monday 4 pm–6 pm

Course Description

This course introduces the basic concepts of data mining as applied to data science, using **Python** as the programming language. **Python Bootcamp** is a prerequisite for the course; we start with only advanced topics of the programming language. This is followed by a more thorough treatment of **Linear Algebra**, **Numpy**, and **Pandas**. With weekly hands-on activities (called **labs**) and homework exercises, we delve into the **fundamental concepts** and the **practical know-how** to perform data mining in real-world situations. We use various data **wrangling** and **visualization** techniques to make them suitable for typical exploratory data analysis (**EDA**). We then use the data in a variety of machine-learning (**ML**) schemes. As an introductory class, we aim to cover a broad spectrum of this emerging field instead of a specialized area. Students will continue to use these concepts and skills in the more advanced courses further into the program.

Learning Outcomes

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

1. How to see the world of data science through the lenses of matrices;
2. Apply data mining concepts and techniques to real-life problems;
3. Demonstrate knowledge of Python programming and basic object-oriented programming concepts by creating Python code for common tasks;
4. Write Python code to perform data analysis, including data pre-processing, data wrangling and model building of various machine learning algorithms;
5. Produce visual charts and graphs of real-world data using Python programming.

6. Analyze data to find information that is relevant and consequential using Python.
7. Synthesize knowledge gained through collaboration with peers on group projects;
8. Communicate data analysis results to a general audience through presentations to peers and the instructor.

Course Prerequisites

We don't go through the basics of Python programming. You will find the class more rewarding if you have a 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 completing an online course or two on Python programming. In addition to Python Bootcamp, I **strongly** recommend completing one of the following:

- (**Trinket Tutorials**) https://hourofpython.com/#tutorials
- (**Official Tutorials**) https://docs.python.org/3/tutorial/index.html
- (e**Book**) **Python for Data Analysis** (Ch 1–3) https://wesmckinney.com/book/

Here is a list of software you need.

1. Install the latest (stable) **Python 3** on your laptop.
    a. Install Jupyter Notebook on your laptop (for ipython, interactive-python).
    b. Follow the guide if you are using VS Code: https://code.visualstudio.com/docs/datascience/jupyter-notebooks
2. Check to see if git is installed on your laptop. (Type "which git" on a terminal.)
3. Sign up for a free GitHub account if you don't already have one.

Below are optional, but strongly recommended.

4. I personally use Visual Studio Code for all my programming projects.
    a. If you use VS Code or another IDE, make sure you install and enable other useful extensions to interpret Python, Markdown, Jupyter Notebook, etc.

Course Format and Labs

The primary format of this course is a lecture followed by a lab or discussion. In general**, lectures with labs** are the main learning tools. These labs are an important part of the course. It is the student's responsibility to complete the in-class labs, which will be graded.

Textbook

The second half of the course will loosely follow the following book(s):

1. (The *Python* version of) **An Introduction to Statistical Learning** by Gareth M. James, Daniela Witten, Trevor Hastie, and Robert Tibshirani
   (This book is good for coding exercises.)
    - Download a free online copy from here: https://www.statlearning.com/
2. **Python for Data Analysis, 3E** by Wes McKinney
   The e-book is available here: https://wesmckinney.com/book/

Course Schedule

Date	Mod	Topics	Reading	Assignment(s) Due
Aug 26	Advanced Python	Course Introduction Getting to know each other Python classes & objects *Lab*: *Python problem solving*	Python Basics	GW Bootcamp
Sep 2	Linear Algebra & Numpy	Vectors, vector operations Matrices, basic matrix operations *Lab*: Linear Algebra Problem Solving		Read: week-1 HW Python
Sep 9		Tensors, tensor algebra Introduction to Numpy Ndarrays and their basic operations *Lab*: Problem-solving using Numpy		Read: week-2 HW Linear Algebra
Sep 16		Numpy: linalg module Random matrices Diagonalization, Matrix decompositions (QR, SVD) PCA, LU, Orthonality, Eigenvalues Applications *Lab*: Solve systems of equations, image compression		
Sep 23	Pandas, EDA	Pandas, data frames Acquiring data (API calls, CSV) Data Imputation, Data wrangling *Lab*: Data wrangling exercise		HW Numpy
Sep 30		Matplotlib, Seaborn, Basic EDA *Lab*: Plotting, Exploratory data analysis (EDA)		
Oct 7		**Exam 1**		HW Pandas
Fall Break				
Oct 14	ML	Introduction to Statistical Learning Types of learning Noise & Loss Functions Training vs Testing Cross-validation Bias-Variance tread-off Curse of dimensionality Evaluation metrics		Exam 1

Oct 21	Machine Learning	Recap of Linear and Logistic Regression Scoring *Lab*: Applying to data		
Oct 28		Decision Tree Regression Random Forest *Lab*: Applying to data		HW Regression
Nov 4		K-NN, concepts and applications *Lab*: Applying to data		HW DT
Nov 11		Support Vector Machine (SVM/SVC) Concepts Linear vs Non-linear kernels *Lab*: Applying to data		HW 10
Nov 18		Unsupervised Learning K-Means clustering *Lab*: Applying to data		HW 11
Dec 2		Discriminant Analysis, Naive Bayes		
Dec 11 (R)		**Make-up session**		Final Project

Independent or Out-of-Class Learning

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

Exams

There is an in-class exam on **Oct 7**.

Project

There will be a final project towards the end of the semester.

Assignments

Assignments will be due weekly (sometimes bi-weekly) and will be **individual**. More details will be posted on Blackboard.

Participation
Student participation in the course is key to fully grasping the class material. The following factors will be taken into account when calculating the participation grade:

- Being present in class
- Asking questions
- Making contributions to class discussions and when called upon
- Staying on task during lab exercises

Readings
Most class days have a corresponding reading assigned, which is indicated within the class syllabus. It is the student's responsibility to complete these readings **before** class on the date they appear on the schedule.

Labs
Students are expected to be present during labs and to complete all labs during class lectures. Labs will be due at the end of class and will be submitted on Blackboard. Labs will be graded and constitute 30% of the overall grade.

Grading
Your final grade will be determined by:

- ✔ Assignments (30%)
- ✔ In-class Labs (30%)
- ✔ Final Project (20%)
- ✔ Exam(s) (20%)

Letter Grades

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

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 the 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 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 at 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