Hood College

Department of Computer Science

Spring 2026

Course Title:

BIFX-546: Machine Learning for Bioinformatics

Instructor: Dr. Sarangan (Ravi) Ravichandran, Ph.D.

Credits: 3

Class Time: Wednesdays, 5:30 PM – 8:15 PM

Location: TBD

Office Hours: By appointment (email: ravichandran@hood.edu)

Course Description

This course introduces the principles and practice of data science and machine learning using *Data Science from Scratch, 2nd Edition* by Joel Grus. We will work directly with the book's code, using guided walkthroughs to unpack key implementations and connect them to the underlying concepts. I will provide accompanying coding notebooks for each topic, and students will build on the starter code by adding functions, writing tests, and applying methods to real-world problems in biomedical, public health, and scientific contexts. The course emphasizes hands-on coding, mathematical understanding, and ethical data use. Each session combines lecture, structured code walkthroughs, and practical exercises.

Course Learning Outcomes

- Implement data science algorithms from scratch using Python.
- Conduct exploratory data analysis (EDA) and visualize patterns in real-world datasets.
- Apply statistical reasoning to biomedical and health data.
- Build and evaluate supervised and unsupervised learning models.
- Apply responsible data science practices, including reproducibility and fairness.

Textbook and Resources

Primary Text: Joel Grus, Data Science from Scratch: First Principles with Python (2nd Edition), O'Reilly Media.

Supplementary References:

- Wes McKinney, Python for Data Analysis (O'Reilly)
- Open datasets from NCBI, Kaggle, and CDC
- Python libraries: NumPy, pandas, matplotlib, scikit-learn, and Jupyter Notebook

Grading and Evaluation

This course is graded entirely through milestone-based deliverables (total = 100%):

- Self-Intro & Goals: 5%
- Problem Definition & Project Proposal: 10%
- Check-In #1 (Pre-Midterm): 10%
- Project Midterm Presentation (EDA milestone): 20%
- Check-In #2 (Pre-Final): 10%
- Final Demo & Discussion (Modeling milestone): 30%
- GitHub Repository Quality: 15%

Project structure (two major milestones):

- Midterm milestone (20%): EDA + problem framing + code demo and reasoning (Project Midterm Presentation).
- Final milestone (30%): modeling + evaluation + interpretation + limitations (Final Demo & Discussion).
- Milestone due dates in the project handout are labeled by class meeting (e.g., "Class 3"). These correspond directly to the week numbers in the 16-week course schedule (e.g., Class 3 = Week 3). The course spans 16 calendar weeks with **15 class meetings** due to a holiday/no-class week. All deadlines labeled 'Class X' refer to the **Xth class meeting**."

Class Schedule (15 Class Meetings)

All hands-on coding exercises will be completed using **Google Colab**, a free cloud-based Jupyter notebook environment provided by Google. It allows you to run Python code directly in your browser—no installation required. You only need a Google account to access, edit, and save your work in the cloud.

The semester spans 16 calendar weeks with 15 class meetings due to a holiday/no-class week. In the schedule below, the numbered rows (1–15) indicate class meetings; the "Holiday — No Class" row is unnumbered and indicates no meeting.

Class Meeting	Topic	Book Chapter(s)	In-Class Focus
1	Course Introduction: What is Data Science?	1–2	Setup Python & Jupyter, course overview
2	Python Fundamentals	2	Data structures, loops, functions, debugging
3	Visualizing Data	3	Plotting data using matplotlib & seaborn

4	Linear Algebra Foundations	4	Implementing vector and matrix operations
5	Statistics and Probability	5–6	Descriptive stats, sampling, probability models
6	Hypothesis Testing and Inference	7	P-values, confidence intervals, bootstrap methods
7	Optimization and Gradient Descent	8	Implement gradient descent manually
Holiday - No Class	---	---	Holiday - No Class
8	Midterm Exam (in-class)	-	Coding + conceptual test
9	Linear Regression	14-15	Model fitting and evaluation
10	Logistic Regression	16	Binary classification on biomedical datasets
11	Decision Trees and Random Forests	17	Tree-based models, feature importance
12	Neural Networks from Scratch	18	Build a simple feed-forward neural net
13	Clustering and Unsupervised Learning	20	K-means, hierarchical clustering
14	Natural Language Processing; Data Ethics & Reproducibility	21, 26	TF-IDF, bag-of-words, biomedical abstracts; Bias, fairness, reproducible notebooks
15	Final Project Demos	-	Presentations and review

Policies

- Attendance: Participation is expected in each weekly session.
- Late Work: Late submissions accepted only with prior approval.
- Collaboration: Students may discuss ideas but must submit their own code and analysis.
- Academic Integrity: All external sources must be cited. Plagiarism will result in disciplinary action per Hood College policies.

Instructor Contact

Dr. Sarangan (Ravi) Ravichandran

Email: ravichandran@hood.edu

Office Hours: By appointment