OEAS 895 Advanced Data Science Techniques in Ocean, Earth and Environmental Sciences
3 credits, Spring 2023

Dr Sophie Clayton, sclayton@odu.edu
Office hours: 13:00 - 15:00 M, OCNPS 423
Class times: 9:30 - 10:45 T/Th, OCNPS 403

Course description

The Ocean, Earth and Environmental Sciences are quickly moving from being a data-poor to data-rich disciplines, with many scientific and industry-related applications enabled by the analysis, synthesis and statistical modeling of large and diverse environmental data sets.

This is an advanced computational analysis course designed to introduce students to data management and analysis methods commonly used in data science applications. The data analysis portion of the course will be primarily based on machine learning methods. The course will also give an overview of a selection of scientific databases which host freely available oceanographic data and output from numerical model simulations. ***This course is not discipline specific*** and will be useful for any students who want to work with data efficiently and gain experience in data management, proper techniques in developing analytical pipelines and applying machine learning to their research.

The class will meet two days a week, Tuesday and Thursday. Classes will consist of a combination of lectures, discussions and practical coding exercises where collaboration and teamwork will be encouraged. The outcome of the course will be an individual capstone project where each student applies the techniques learned during the course to undertake a data analysis project based on their own research interests using at least 2 different data sources from open scientific databases, and may include data that they have generated themselves. Students will be expected to publish the code developed and results of their project in a public GitHub repository.

Learning objectives

1. Understand FAIR data principles and how to apply them when generating, sharing and accessing data.
2. Develop a working knowledge of existing ocean and earth science databases and how to efficiently access data from them, including via APIs.

3. Students will develop their own data analysis toolbox using, but not limited to, Python and shell scripts.
4. Understand and use version control (e.g. git), environments (e.g. conda) and code repositories (e.g. GitHub) to manage and share code.
5. Understand the underlying principles of machine learning techniques for regression and classification, including supervised and unsupervised learning and apply them to a targeted research question.
6. Understand the process of model evaluation and optimization and commonly used metrics for reporting model performance.

Pre-requisites

This course is an advanced data analysis course. Students must have working knowledge of at least one programming language (e.g. MATLAB, R, python) and some familiarity with using the command line. The course will be taught using python, so ideally, students will have at least participated in either the summer 2022 COS data science bootcamp, or one of the recent ODU python workshops. Students should have a basic working knowledge of statistics and calculus.

Software

The course will be taught using python. Students are encouraged to use the Anaconda package manager, but this is not required. We will use git and GitHub for version control, and Slack for class communications. The software packages used in this class are all available for free.

Capstone Project

The goal of the final capstone project is to assess students ability to combine and apply the skills learned in class in the context of a real-world research problem. The class will mostly focused on tools for data analysis, visualization and developing and evaluating machine learning models, so this will be the focus of the capstone project. Students must have the dataset(s) and general scope of their capstone project approved by the instructor prior to spring break.

Grading Summary

Problem sets	50%
Capstone project report	20%
Capstone project repository	15%
Capstone project presentation	15%

Attendance

You are expected to attend the class meetings in person and fully participate in class. If you know that you will be absent for a specific reason, please inform me as soon as you can.

Grades and expectations

The grading policy for this class is non-competitive; there will be no curve. If everyone in the class does well, everyone will get an A.

A	93 – 100%	C	73 – 76%
A-	90 – 92%	C-	70 – 72%
B+	87 – 89%	D+	67 – 69%
B	83 – 86%	D	63 – 66%
B-	80 – 82%	D-	60 – 62%
C+	77 – 79%	F	$\leq$ 59%

Academic Integrity and Classroom Conduct

Old Dominion University is committed to students' personal and academic success. In order to achieve this vision, students, faculty, and staff work together to create an environment that provides the best opportunity for academic inquiry and learning. All students must be honest and forthright in their academic studies. Your work in this course and classroom behavior must align with the expectations outlined in the Code of Student Conduct, which can be found at www.odu.edu/oscai. The following behaviors along with classroom disruptions violate this policy, corrupt the educational process, and will not be tolerated.

Cheating: Using unauthorized assistance, materials, study aids, or other information in any academic exercise.

Plagiarism: Using someone else's language, ideas, or other original material without acknowledging its source in any academic exercise.

Fabrication: Inventing, altering or falsifying any data, citation or information in any academic exercise.

Facilitation: Helping another student commit, or attempt to commit, any Academic Integrity violation, or failure to report suspected Academic Integrity violations to a faculty member.

Academic dishonesty will be reported to the Office of Student Conduct & Academic Integrity and may result in sanctions up to and including expulsion from the University.

Course Accommodations

Students are encouraged to self-disclose disabilities that have been verified by the Office of Educational Accessibility by providing Accommodation Letters to their instructors early in the semester in order to start receiving accommodations. Accommodations will not be made until the Accommodation Letters are provided to instructors each semester.

The Office of Educational Accessibility is located at 1021 Student Success Center and its phone number is (757) 683-4655. Additional information is available on the OEA website.

Course schedule:

Modifications may need to be made to this schedule as the semester progresses.

Week	Topic	Assignment
1 (1/10)	Open Science framework FAIR data Version control (git, github)	PS 1 – git and github
2 (1/17)	Initial data access and exploration Basic plotting in python Oceanographic databases and repositories	PS 2 – exploratory data analysis
3 (1/24)	Oceanographic toolboxes Mapping toolboxes	PS 3 – building a function for accessing data using an API
4 (1/31)	NO CLASSES	
5 (2/7)	Building packages and sharing code Collaborative workspaces	PS 4 – documenting and sharing your code (builds on PS3)
6 (2/14)	Machine Learning overview Introduction to scikit-learn	
7 (2/21)	Supervised Learning Overview of algorithms Training and testing algorithms	PS 5 - building a simple classification model with scikit-learn
8 (2/28)	Unsupervised learning Clustering Classification	PS 6 – building a regression model with scikit-learn
9 (3/7)	NO CLASSES – SPRING BREAK	Deadline for capstone project approval
10 (3/14)	Model evaluation Cross-validation (dealing with small training sets)	
11 (3/21)	Capstone project development	PS 7 – project outline
12 (3/28)	Machine Learning applications in oceanography Paper discussion (student-led)	
13 (4/4)	Capstone project hacking	
14 (4/11)	Capstone project hacking	PS 8 – project and code review (peer evaluation)
15 (4/18)	Data analysis project hacking In-class capstone presentations	Capstone Project Due (published github repository)