CEE 498EDS: Environmental Data Science, Fall 2020

Basic Course Information

- Department: Civil and Environmental Engineering
- Title: CEE 498EDS: Environmental Data Science
- Credits: 3 for Undergraduates, 4 for Graduate Students
- Semester: Fall 2020
- Meeting time and location: TBD

Basic Instructor Information

- Instructor: Prof. Christopher Tessum, PhD
- Office: 3213 Newmark Civil Engineering Laboratory
- Office hours: TBD
- email: ctessum@illinois.edu
- Names and contact information for teaching assistants: TBD

Description of the course

Welcome to Environmental Data Science! This semester, you will learn to leverage data to study environmental problems, identify patterns, and make actionable insights. This course combines training in digital and computer tools—including distributed computing, exploratory data analysis, and statistical modeling and deep learning—with application of those tools to environmental issues.

By the end of the semester, you will be able to:

1. Use software tools for data processing and visualization, machine learning, and deep learning to
2. Retrieve, manipulate, and analyze environmental data; and
3. Make inferences and predictions about the environment.

This course will help you to gain the skills and tools necessary to make the most of the great increases in the amount and quality of environmental data that is being collected and stored.

Because data science methods are used across a number of different industries and instructional materials are readily available, this course will include readings and video lectures from across the internet. We will focus our face-to-face time on learning aspects of environmental data science that differ from data science as used by other fields, and on applying data science concepts to solving environmental problems. This course will be structured around semester-long projects; students will choose project topics at the beginning of the semester and will apply the concepts learned in the class to their projects as the semester progresses.

Prerequisites

CEE 202; CEE 330; CS 101 or equivalent.

Course Requirements and Assessment Overview

- Grades will be assigned based on several types of deliverables:
  - Mini quizzes and assignments on readings and video lectures, due before class: 20% of total grade

    - Homework problem sets: 20% of final grade
    - Midterm exam: 5% of final grade
    - Final exam: 15% of final grade
    - Course project: 40% of final grade: 5% each for each of 5 checkpoints, 5% for midterm presentation, and 10% for final presentation.
- Graduate students are expected to register for 4 credits and undergraduates are expected to register for 3 credits. Correspondingly, course projects for graduate students are expected to include a machine learning component that is more complex than linear regression, whereas for undergraduates this is optional.
- Grades will be assigned according to the attached rubrics.
- Letter grades will be assigned according to the following scale:
    - 90%-100%: A
    - 80%-89%: B
    - 70%-79%: C
    - 60%-69%: D
    - <60%: F

Learning Resources

- Students are expected to bring a laptop to class. Laptops can be loaned from the library.
- There is no required textbook. Some supplmental textbooks which students may find useful are:
    - Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython
    - Hands-On Machine Learning with Scikit-Learn & TensorFlow

Course Schedule

- Week 1: Open Reproducible Science
    - Reading: https://www.earthdatascience.org/courses/intro-to-earth-data-science/open-reproducible-science/get-started-open-reproducible-science/
    - In class lecture/tutorial:
        - FAIR Principles
        - Git/Github
        - Jupyter
- Week 2: Data science for the environment
    - Reading:
        - https://arxiv.org/pdf/1906.05433.pdf
        - https://pangeo.io/use_cases/index.html
        - https://www.kaggle.com/
        - https://earthengine.google.com/case_studies/
        - Others from specific environmental disciplines
    - In class:
        - Students present on potential projects
- Week 3: Programming review
    - Reading/Lecture: https://developers.google.com/edu/python
    - In class:
        - Python & Jupyter exercises
- Week 4: Big data

    - In class:
        - Lecture:
            - Cloud / High-performance computing
            - Pangeo
            - Earth engine
        - Practice and discussion
        - Choose project groups and topics

- Week 6: Exploratory data analysis (EDA)

    - Reading:
        - https://mlcourse.ai/articles/topic1-exploratory-data-analysis-with-pandas/
        - https://mlcourse.ai/articles/topic2-visual-data-analysis-in-python/
        - https://mlcourse.ai/articles/topic2-part2-seaborn-plotly/
        - And/or accompanying video lectures
    - In class:
        - Statistics review lecture
        - Students present initial EDA for their projects

- Week 7: Geospatial data

    - Reading: https://github.com/geopandas/scipy2018-geospatial-data
    - In class lecture:
        - raster vs. vector formats
        - joins and boolean operations

- Week 8: Spatial statistics

    - Reading: http://pysal.org/notebooks/intro
    - In class dicussion:
        - Applicability of spatial statistics to student projects

- Week 9: Mid-way project presentations

    - Students should be able to access, characterize, and visualize the data for their projects.

- Week 10: Supervised learning

    - Reading/lecture:
        - https://developers.google.com/machine-learning/crash-course/framing/video-lecture
        - https://developers.google.com/machine-learning/crash-course/descending-into-ml/video-lecture
        - https://developers.google.com/machine-learning/crash-course/reducing-loss/video-lecture
        - https://developers.google.com/machine-learning/crash-course/first-steps-with-tensorflow/toolkit
        - https://developers.google.com/machine-learning/crash-course/generalization/video-lecture
        - https://developers.google.com/machine-learning/crash-course/training-and-test-sets/video-lecture
        - https://developers.google.com/machine-learning/crash-course/validation/check-your-intuition
    - In class:
        - Application to student projects

- Week 11: Unsupervised learning
    - Reading/Lecture: https://mlcourse.ai/articles/topic7-unsupervised/
    - In class:
        - Application to student projects
- Week 12: Deep learning
    - Reading/lectures:
        - https://developers.google.com/machine-learning/crash-course/introduction-to-neural-networks/video-lecture
        - https://developers.google.com/machine-learning/crash-course/training-neural-networks/video-lecture
        - https://developers.google.com/machine-learning/crash-course/multi-class-neural-networks/video-lecture
    - In class:
        - Hyperparameter optimization
        - Application to student projects
- Week 13-14: Projects
- Week 15-16: Final exam; final project presentations