EAS 8803- Machine Learning for Earth & Environmental Sciences

UG 4803- Machine Learning for Earth & Environmental Sciences

Instructor: Dr. Ali Sarhadi
Email: sarhadi@gatech.edu

Office: ES&T 3244
Class Date, Time & Location: [M W, 9:30-10:45, Clough Room123]
Office Hours: [Day, Time]

Course description and learning objectives

This course provides a comprehensive introduction to a broad range of machine learning techniques, with a particular emphasis on their applications in Earth systems and environmental sciences. Students will engage with both foundational and advanced algorithms, applying them to analyze complex environmental datasets and solve critical real-world challenges. Through hands-on learning, students will develop the practical skills necessary to design, implement, and validate data-driven models, equipping them for advanced research in graduate studies.

By the end of the course, students will be able to:

- Learn key machine learning algorithms, understanding their theoretical foundations and practical applications relevant to earth systems and environmental sciences.
- Analyze and process environmental datasets, using advanced Python-based machine learning tools and libraries.
- Design, implement, and validate data-driven models, applying machine learning techniques to enhance understanding, predictability, and forecasting of Earth and environmental systems.
- Develop and present a comprehensive research project, showcasing the application of machine learning techniques to real-world environmental or earth system problems.

Course structure

Classes will consist of lectures, hands-on coding sessions in Python, and collaborative discussions. Lectures will cover the theoretical aspects of machine learning, supplemented with practical coding examples to help students learn key steps in the development of machine learning models. Assignments will assess students' understanding of fundamental concepts and their coding proficiency in solving provided problems. All students will be required to implement code individually. For the final project, students will form groups of 2 or 3 to collaborate on a research topic, culminating in a final report and presentation that demonstrates the application of machine learning to an environmental or Earth system issue.

Prerequisites

No prior experience with machine learning is required. However, a solid foundation in mathematics, particularly in statistics and linear algebra, is recommended. While the course focuses on applying machine learning to Earth systems and environmental sciences rather than diving into deep mathematical theory, having a mathematical background will help students grasp the underlying analytical methods more effectively.

A basic knowledge of Python programming is preferred, but not mandatory. Students unfamiliar with Python are strongly encouraged to either complete an introductory Python course or utilize the many online tutorials available for self-study. Resources for self-guided learning will also be provided to help you get up to speed. Students may choose to use R or MATLAB to complete assignments if preferred; however, they must write and implement their own scripts independently to meet the assignment requirements.

Students need to have access to a personal computer or laptop capable of running the codes.

Course materials

The course will provide the following materials to support your learning:

- **Lecture slides**: Slides will be uploaded to Canvas after each lecture for reference and review.
- **Practical code examples**: For each topic, practical coding examples will be provided to help students implement the concepts learned in class. These codes will be available on Canvas and will serve as a guide for assignments and projects.
- **Optional textbooks**: The following books are recommended for students who wish to explore additional resources and deepen their understanding:
    - *An Introduction to Machine Learning* by Gopinath Rebala, Ajay Ravi, and Sanjay Churiwala (2019), Springer International Publishing.
    - *Introduction to Machine Learning with Python: A Guide for Data Scientists* by Andreas C. Müller and Sarah Guido (2017), O'Reilly Media.
    - Chollet, F. Deep Learning with Python. (2018), Manning Publications.
    - Chollet, F., & Allaire, J. J. *Deep Learning with R*. (2018). Manning Publications.

Grading

- **Assignments (50%)**:
  Throughout the semester, multiple assignments will be given to students via Canvas. These assignments will cover both the fundamental concepts taught in class and practical

coding exercises. Each assignment will focus on applying machine learning methods to solve real-world issues, offering students the opportunity to practice coding and optimize solutions using the techniques covered in the course.

- **In-class activities (10%)**
  In-class activities will include group-based paper reviews, where students will collaborate to analyze and discuss key findings from assigned research papers. Each group will then share their insights with the class in a jigsaw format, encouraging peer learning and discussion. Active participation and engagement during these sessions will also contribute to the in-class activity grade.

- **Research proposal, final report, and presentation (40%)**:

- **Proposal (10%)**:
  Students will develop an original research proposal on a topic related to the application of machine learning in environmental sciences. The proposal must be unique and not part of another course or research project, though proposals related to a student's thesis work are encouraged.
  The proposal should include:
  - A clear statement of the research question, its machine learning component, and its contribution to the literature.
  - A description of the data required, including data acquisition plans or an actual dataset.
  - A detailed methodology for data analysis, including the proposed machine learning algorithm(s), justification, alternative approaches, and potential challenges.
    The proposal should not exceed 1 single-spaced pages, using 12pt font.

- **Final report (20%)**:
  The final report will build on the proposal and include detailed results of the data analysis, including the methods applied and an evaluation of the model's performance. The report should also address challenges faced during the project and possible improvements. The report should not exceed 5 single-spaced pages.

- **Presentation (10%)**:
  Students will present their project during the last class, with a 10-minute presentation followed by 3 minutes for questions. The presentation should summarize the research question, methodology, results, and conclusions.

The grading scale will be based on the overall class performance. I reserve the right to adjust the scale if necessary to ensure a fair and balanced distribution of grades across the class.

Tentative course outline

This is a tentative schedule, and topics may be adjusted as the course progresses.

Week	Topics	Details
1	Machine Learning: Introduction	Overview of machine learning concepts and applications in Earth and environmental systems
2	Feature Engineering and Dimensionality Reduction	Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA)
3	Unsupervised Learning (Clustering, Pattern Mining)	K-means, DBSCAN, and hierarchical clustering
4	Supervised Learning (Linear and Multivariate Regressions)	Linear, polynomial, Ridge Regression, LASSO, K-Nearest Neighbors (KNN), and decision trees
5	Bayesian Learning and Probabilistic Modeling	Bayesian Regression and Naïve Bayes
6	Ensemble Learning and Support Vector Machines	Random Forest, Gradient Boosting, and Support Vector Machines (SVMs)
7	Neural Networks and Deep Learning	Introduction to neural networks, feedforward networks, and backpropagation
8	Advanced Deep Learning	Convolutional Neural Networks, Recurrent Neural Networks (RNNs), and Autoencoders
9	Break	--
10	Generative AI and Advanced Architectures	Generative Adversarial Networks (GANs), Diffusion Models, and Transformers
11	Interpretable Machine Learning and Causal Inference	Model interpretability, Shapley values, and causal inference in environmental models
12	Reinforcement Learning	Reinforcement learning concepts, with applications in environmental decision-making and control
13	Project presentation	Presentation and discussion of student projects on applied machine learning in environmental sciences

Online course evaluation process

Students are encouraged to provide professional and constructive feedback on the quality of instruction in this course by completing the online course evaluations via the Georgia Tech Course Instructor Opinion Survey (CIOS). Detailed guidance on how to provide feedback in a thoughtful and respectful manner is available through the CIOS portal. You will be notified when the evaluation period opens, and evaluations can be submitted through the email invitation you receive, directly in Canvas under the CIOS tab, or via the Georgia Tech CIOS website at https://www.academiceffectiveness.gatech.edu/surveys/cios. Summaries of course evaluation results are available to students through this same platform.