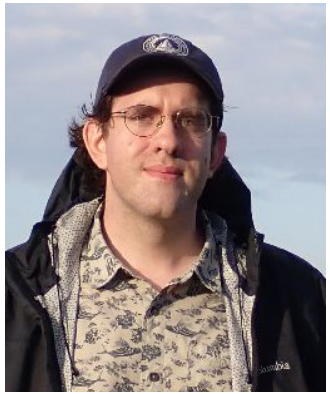
Intermediate Machine Learning in Python for Environmental Science Problems

AMS Committee on AI Applications to Environmental Science

AMS American Meteorological Society

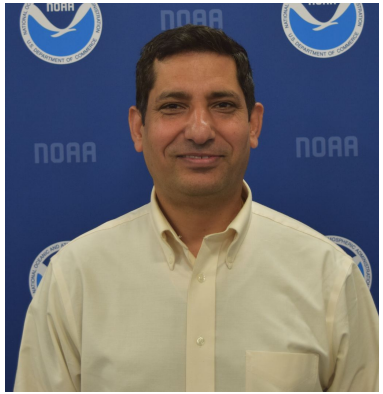
105th AMS Annual Meeting
January 12, 2025
8:00 AM - 3:30 PM Central

The Team



Evan Krell
ekrell@islander.tamucc.edu

Kara Lamb
kl3231@columbia.edu



Praveen Singh
praveen.singh@noaa.gov
Lynker @ NOAA/NCEP/EMC

Christian Duff
cduff4@islander.tamucc.edu

TEXAS A&M UNIVERSITY CORPUS CHRISTI | CONRAD BLUCHER INSTITUTE FOR SURVEYING AND SCIENCE

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Lynker
An Employee-Owned Company

TEXAS A&M UNIVERSITY CORPUS CHRISTI | CONRAD BLUCHER INSTITUTE FOR SURVEYING AND SCIENCE

Coastal Dynamics Lab

ARTIFICIAL INTELLIGENCE · RISK COMMUNICATION · ENVIRONMENTAL SCIENCE
ai2es.org

LEAP

Coastal Dynamics Lab

ARTIFICIAL INTELLIGENCE · RISK COMMUNICATION · ENVIRONMENTAL SCIENCE
ai2es.org

Motivation

This course is aimed at practitioners using ML for environmental science applications.

Intermediate: we assume basic familiarity with developing ML models using Python.

- At least some experience developing simple ML models like Random Forest or Multilayer Perceptron
- Basic Python skills for working with data: `numpy`, `pandas`, `matplotlib`

There is a wealth of beginner ML material online, but focusing on simple toy problems that don't address challenges that are common in environmental science.

This course focuses on the *next steps* for better working with environmental data for more challenging environmental science problems.

Topics

Handling Imbalanced Data

- Extreme events: it is trivial to achieve high performance on average by simply always predicting no storm… you'll be right most of the time!
- How to we encourage ML systems to achieve skill prediction of rare events?

Hyperparameter Tuning

- There are so many options when configuring ML models: number of layers, learning rate, batch size, loss function, etc.
- Are there strategies and tools to efficiently select these?

Model Evaluation

- A model with overall high performance might might not be suited to the target application → proper evaluation crucial to ensure a model is used appropriately

Explainable Artificial Intelligence

- What has the model learned? Is the model learning physically-sound strategies or relying on spurious correlations?

Physics-informed ML

- Can we exploit domain knowledge about the physical system to guide ML systems toward learning meaningful prediction strategies?

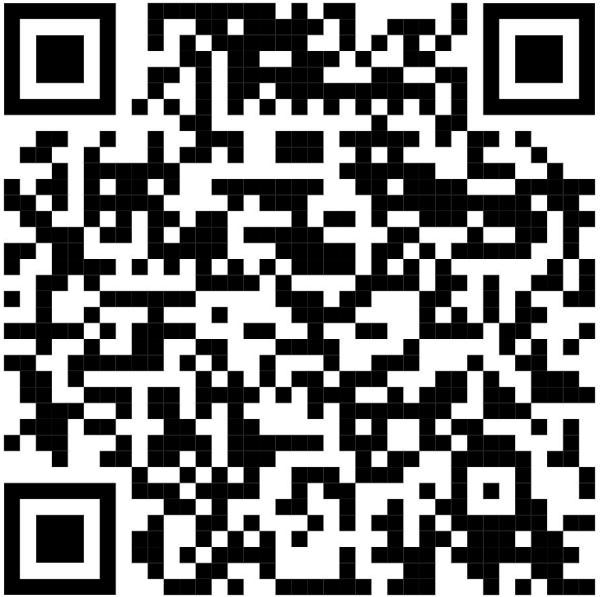
Agenda

Activity	Content Description	Estimated Time
Introduction	Introduction to the course and instructors, and setting up the coding environment.	15 minutes
Follow-Along Lectures	The instructors will teach by interleaving lecture with code demonstrations that the students will follow along with on their laptops.	1 hour, 45 minutes
Break		15 minutes
Follow-Along Lectures		1 hour, 45 minutes
Lunch		45 minutes
Follow-Along Lectures		1 hour, 45 minutes
Break		15 minutes
Hands-on Exercise	Participants will work together in small groups to solve a problem that combines concepts from all lectures	1 hour, 15 minutes

Lecture sequence

Introduction	Evan Krell
Imbalanced data	Praveen Singh
Hyperparameter tuning	Christian Duff
Model evaluation	Evan Krell
Explainable AI	Evan Krell
Physics-informed ML	Kara Lamb

Course Resources



github.com/ekrell/ams_ai_shortcourse_2025

GitHub repository contains all lecture material, including Google Colab notebooks that students can follow-along with.

Use this time to navigate to the repo, and make sure you can access Google Colab (requires Google account).