Machine Learning in Earth & Environmental Sciences
Course Syllabus – Mohammed Ombadi (ombadi@umich.edu)
Fall Semester

Instructor: Dr. Mohammed Ombadi (ombadi@umich.edu); He/Him

Credits: 3.0
Lecture: Mon., Wed. 3 – 4:30 am in TBD.
Office Hours: Thursday 1 – 2 pm; Location: TBD. Also, by appointment in 1539 CSRB North Campus.
Course Website: via Canvas

Course Description

The rapid increase in environmental data acquisition over the last few decades has provided us with an unprecedented opportunity to draw insights from big data on the behavior of environmental systems. This course aims to introduce students to statistical methods, ranging in complexity from autoregression to machine learning models.

The course covers the basic theory behind machine learning and provides hands-on experience in building machine learning models. Students will learn to apply these models for both prediction and hypothesis formulation purposes. The methods will be taught through example applications in environmental sciences, with a specific focus on climate and hydrologic applications.

Examples include short-term forecasts of temperature and precipitation, streamflow forecasting in selected hydrologic basins, understanding the relative contributions of temperature and precipitation in snowmelt trends, regional clustering of precipitation patterns and trends, and climatic teleconnections in regulating regional precipitation patterns.

Learning Outcomes

After successfully completing this course, students will be able to:

- Identify the basic statistical features of environmental data.
- Understand the differences between parametric/non-parametric and linear/nonlinear models and select the most appropriate statistical models based on the research problem and/or availability of data.
- Conduct exploratory analysis on data and develop hypotheses regarding variable interactions.
- Develop data-driven models for prediction and evaluate their adequacy.
- Gain an understanding of the basic theory behind machine learning models.

Teaching Method and Philosophy

Teaching Methods and Philosophy: The teaching method for this course combines lectures, small group work, and active (experiential) learning. Lectures will introduce concepts, methodologies, and examples of applications in climate and environmental sciences.

Small group work is primarily focused on the final project, where each group of 2 to 4 students will apply some of the methods introduced in the course to a problem of their own interest. Experiential learning is integrated into homework assignments, where students will receive Python codes and data files. They will be tasked with applying the methods and answering a set of questions.

Class Climate & Inclusivity

Diversity is not only appreciated but celebrated in this classroom. The teaching staff maintains a zero-tolerance policy towards any form of discrimination. Students are expected to demonstrate respect, civility, and considerate conduct. Additionally, recognizing differences in language, culture, and personal viewpoints is encouraged. Feedback on issues related to diversity in the classroom is welcomed.

Required Computing Software

Coding scripts for the data analysis methods will be provided in Python. Students are welcome to use R or MATLAB to complete assignments if they prefer, but they must write their own scripts.

Grade Breakdown

Participation in class / Quizzes	10%
Homework Assignments	60%
Project	30%

Course Schedule

Lecture	Topic	Application Example
1	**Review of basic statistics:** Random variables, measures of central tendency (mean, median, mode), variance, higher moments (skewness and kurtosis), sample vs population statistics,	Summarize the statistics of temperature and precipitation in Ann Arbor, Michigan
2	**Probability distributions and Model-evaluation metrics:** continuous and discrete probability distributions, Distributions for modelling extreme values, Model evaluation metrics (Mean Squared Error, Mean Absolute Error, Bias …etc.), Metrics for categorical variables.	Explore changes in global mean surface temperature (compare two periods), Explore the relationship between snowmelt, precipitation and temperature over selected watersheds in Western US
3	**Hypothesis testing and statistical dependence:** one- and two-sided hypothesis testing, statistical significance, confidence interval, Pearson	Explore trends in global mean surface temperature.

	correlation, nonparametric correlations (e.g., Spearman), partial correlation	
4	**Feature Selection and Dimensionality Reduction:** Types of feature selection, Feature selection for continuous and categorical variables, Singular Value Decomposition.	**Hydrology:** Selecting features of continuous, meteorological variables for prediction of streamflow **Ecology:** Selecting categorical features associated with different types of animals.
5	**Clustering algorithms:** Supervised vs Unsupervised learning. K-means Clustering Algorithm	**Synthetic Data:** 2-Dimensional, 2-classes dataset **Climate Science:** Using clustering for climate classification based on temperature and precipitation.
6	**Clustering (continued**): Hierarchical (agglomerative method) and Density-based (DBSCAN) algorithms	**Hydro-climatology:** clustering for classifying trends in groundwater **Hydrology:** Determining the optimum location of environmental sensors
7	**Simple and Multiple linear regression:** Maximum likelihood parameter estimation, loss functions (L1 and L2 norms), model evaluation and goodness of fit.	**Hydrology:** Prediction of streamflow based on different meteorological predictors. Build models with different number of inputs and evaluate their goodness of fit.
8	**Choice of Loss Function, Regularization techniques (**Ridge and Lasso**)**	**Climate Science:** Using regression for assessing trends in mean global temperature
9	**Gradient Descent, Anatomy of a Neural Network**	
10	**Intro to machine learning:** Gradient descent and backpropagation, feedforward neural networks, Multilayer Perceptron	**Hydrology:** Streamflow Prediction
11	**Considerations in building an ML model:** Selection of optimizers, choice of normalization, training-testing data split	
12	**Decision Trees:** Classification Trees, Gini Impurity, Regression Trees, Random Forests	**Hydrology:** Streamflow prediction using both

		continuous and categorical meteorological variables.
13	**Interpretability of Decision Trees, Feature Importance, Shapley Values**	**Environmental Engineering and Water Quality:** Understanding Drivers mediating flood impacts on salinity **Climate:** Environmental Drivers of Wildfire
14	**Recurrent Neural Networks (RNNs):** Structure of recurrent neural networks; Building a simple RNN.	**Climate:** Temperature prediction at sub-daily times scales using RNN.
Students start forming groups and discuss project topics with instructors		
15	**LSTM (continued)**	**Hydrology:** Streamflow Prediction
16	**Convolutional Neural Network (CNNs):** Enforcing simple physics-based constraints, nonlinear dynamics-based neural networks	Climate: Cloud Classification
17	**CNNs (Continued):** considerations of selecting the Kernel size, pooling layers and other parameters of the network.	
18	**Causal Inference:** Granger and predictive causality, Transfer Entropy, Graph-based methods	Evaluating the relative contribution of environmental variables in driving evapotranspiration
19	**Graphical models and Graph Neural Networks**	Applications to streamflow and water quality modeling
20	**Bias-variance tradeoff, K-fold cross validation, Hyperparameter Tuning (**GridSearchCV and RandomizedSearchCV)	
21	**Physics-informed Machine Learning (Objectives and approaches)**	**Environmental Science:** Modeling lake water temperature
22	**Autoencoders:** simple autoencoder, types of autoencoders, representation learning, reconstruction of attractors using autoencoders	