ABE 498: Data Science for ABE, Fall 2023

THIS SYLLABUS IS SUBJECT TO CHANGE! Please check back throughout the course.

Course Information

- Department: Agricultural and Biological Engineering
- Title: ABE 498: Applied Data Science for Agricultural and Biological Engineering
- Credits: 3
- Semester: Fall 2023
- Meeting time and location: 2-3:20 Tuesday and Thursday, Room 248
- First day of instruction: August 22
- Last day of instruction: December 5
- Important links:
  PrairieLearn (for classes, assignments and exams)
  CampusWire (for announcements and discussion)

Instructor Information

- Instructor: Prof. Mei Tessum, PhD, MBBS
- Office: 360 Agricultural Engineering Science Building
- Office hours: 1-2 Tuesday and Thursday
- Website: https://abe.illinois.edu/directory/mtessum

Course Description

In this course, students will learn to leverage data to study agricultural and biological engineering problems, identify patterns, and make actionable insights. This course includes training in (1) exploratory data analysis including data profiling, missing data, description, and data visualization; (2) data processing techniques including singular value decomposition, dimensionality reduction, and fourier and wavelet transforms; (3) machine learning techniques including regression, classification, feature selection, clustering, and neural networks.

This course differs from other available data science courses in that it focuses on agricultural and biological engineering problems and the methods used to solve them. In particular, this course emphasizes tools and skills for working with datasets that are unique to agricultural and biological engineering, nature

resources, and environmental science, such as spatiotemporal data, genetic data, etc.

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

1. Use software tools for data description and visualization, data processing, and machine learning to
2. Retrieve, manipulate, and analyze data; and
3. Make inferences and predictions about the built and natural environment to solve agricultural and biological engineering problems.

This course will include readings and video lectures from across the internet, and face-to-face time on hands-on practice with the concepts and tools in classroom.

Prerequisites

- MATH 225; MATH 285; CS 101; ABE 341 or ABE 440 or CEE 202 or IE 300 or STAT 400 or equivalent.

Course Structure

This course is structured as a series of modules, with each module containing recorded lectures, readings, and quizzes to be completed before each class meeting. Class meetings will be in-person, and will focus on hands-on application of the material that was covered in recorded lectures and readings through the use of in-class worksheets. The in-class worksheets will be completed in groups and will be graded. Near the beginning of the semester, students will also choose a topic for a project, which they will work on throughout the semester, applying the concepts that we learn in class. Additionally, students will complete homework assignments and exams.

Course Requirements and Assessment Overview

- Grades will be assigned based on several types of deliverables:
    - Pre-lecture quizzes: 5% of final grade
    - In-class worksheets: 11% of final grade.
    - Homework: 20% of final grade
    - Exams: 39% of final grade (13% each).
    - Course project: 25% of final grade
        - 3% for project selection and introduction,

        - 3% for exploratory analysis,
        - 3% for preliminary modeling,
        - 3% for report rough draft,
        - and 13% for final report and presentation.
- Letter grades will be assigned according to the following scale:
    - 97-100: A+
    - 93-96.9: A
    - 90-92.9: A-
    - 87-89.9: B+
    - 83-86.9: B
    - 80-82.9: B-
    - 77-79.9: C+
    - 73-76.9: C
    - 70-72.9: C-
    - 67-69.9: D+
    - 63-66.9: D
    - 60-62.9: D-
    - Below 59.5: F

Homework and Exams

Homework and Exams will be done through PrairieLearn. In assigning these types of homework and exams, I'm placing emphasis on **mastery**. The idea is to keep doing questions until you master the underlying concept or method. Once you do, you should be able to answer these questions very quickly.

Important: When you log in to PrairieLearn, choose "Log in with Illinois" rather than "Log in with Google" or "Log in with Microsoft". The UIUC login is the only one that will work.

➢ Homework

For the homework, I try to encourage preparation for class before a module starts, so if you finish all of the questions completely before the first meeting time for the module, you will receive 110% of the available points. Questions finished between the first and last meeting times of the module receive 100%, and questions finished up to two weeks after the module ends can receive 80%.

Note that new homework assignments are assigned most weeks, so if you don't stay ahead, it can be easy to fall behind.

➢ Exams

Exams are also administered using PrairieLearn. For exams, partial credit usually isn't given, but you can try each problem more than once, with a decreasing number of points possible for each try.

Learning Resources

- **Students are highly encouraged to bring a laptop to class**, and have use of one and an internet connection for homework and pre-class activities. If there is no laptop access, the computers in room 210 are available.
- There is no required textbook to purchase. However, much of the course follows the book *Data Driven Science and Engineering* by Brunton and Kutz. The book includes a freely available pdf version and lecture videos, is also available for sale in hardcover and ebook formats.
- The course also includes content from the Julia learning resources, which is also a good resource for learning Julia in general. The free course *Introduction to Computational thinking* from MIT is particularly well done.
- For students that want to continue beyond the material covered in this course, there is *Parallel Computing and Scientific Machine Learning* (another course from MIT) and *Deep Learning* by Goodfellow, Bengio, and Courville.

Policies

Inclusive Environment

The effectiveness of this course is dependent upon the creation of an encouraging and safe classroom environment. Exclusionary, offensive or harmful speech (such as racism, sexism, homophobia, transphobia, etc.) will not be tolerated and in some cases subject to University harassment procedures. We are all responsible for creating a positive and safe environment that allows all students equal respect and comfort. I expect each of you to help establish and maintain and environment where you and your peers can contribute without fear of ridicule or intolerant or offensive language.

If you witness or experience racism, discrimination, micro-aggressions, or other offensive behavior, you are encouraged to bring this to the attention of the course director if you feel comfortable. You can also report these behaviors to the Bias Assessment and Response Team (BART) (https://bart.illinois.edu/). Based on your report, BART members will follow up and reach out to students to make sure they have the support they need to be healthy and safe. If the reported behavior also violates university policy, staff in the Office for Student Conflict Resolution may respond as well and will take appropriate action.

Accommodations

To obtain disability-related academic adjustments and/or auxiliary aids, students should contact both the instructor and the Disability Resources and Educational Services (DRES) as soon as possible. You can contact DRES at 1207 S. Oak Street, Champaign, (217) 333-1970, or via email at disability@illinois.edu.

Participation

Active participation in the online learning environment is vital to your success in this course. Depending on your course, you may be asked to engage in online discussions and other interactive learning environments that invite your active participation and involvement with other students and your instructor.

Student Commitment

By registering for this course, you commit to self-motivated study, participation in online course activities, and timely submission of all assignments. Furthermore, you commit to accessing the course website and checking email at least four days per week (daily for 4-week courses).

The University guidelines for course credit hours are posted here. In summary, students are expected spend at least 6 hours per week working outside of class times on readings, assigned lectures, assignments, and test preparation.

It is my goal for this class to follow these guidelines so let me know if you think it does not (keeping in mind that the guidelines are for the *minimum* effort requirements).

Deadlines

If you are unable to meet a particular deadline, it is your responsibility to make prior arrangements with the instructor for that given week. Otherwise, work

submitted later than 1 day late will receive 10% penalty, and work submitted later than 2 days late will not be considered for grading unless consent has been given by the instructor. PrairieLearn assignments have separate policies for late submission which are shown on the PrairieLearn website.

Regrades

Requests for regrading homework and exams must be submitted in writing within one week of receiving the initial grade, and must include a written explanation of the reason for the regrade request.

Instructor Responses

➢ Instructor Feedback Turnaround Time

Questions posted to CampusWire generally will be answered within 2 days. Questions may be answered on weekends but this should not be expected.

Assignments submitted online will generally be reviewed and graded by the course instructor within 5 business days. Exams, essays, and term papers will generally be graded within 10 business days.

➢ Contacting the instructor

For the fastest response, the best way to contact the instructor is by attending office hours or posting questions to the CampusWire.

The instructor will not respond to phone calls and may not respond to emails. If you must send an email, include a subject line that identifies the course number and nature of your question. Please don't be offended if you are asked to repost your question on CampusWire to allow the instructional team to efficiently answer all questions.

Academic Integrity

Academic dishonesty will not be tolerated. Examples of academic dishonesty include the following:

- Cheating
- Fabrication
- Facilitating infractions of academic integrity
- Plagiarism

- Bribes, favors, and threats
- Academic interference
- Examination by proxy
- Grade tampering
- Non-original works

Should an incident arise in which a student is thought to have violated academic integrity, the student will be processed under the disciplinary policy set forth in the Illinois Academic Integrity Policy. If you do not understand relevant definitions of academic infractions, contact your instructor for an explanation within the first week of class.

Giving and receiving advice on homework assignments is acceptable and encouraged. However, it is expected that help be given in general terms and in the form of natural language sentences (for example, English) rather than in the form of mathematical equations, algorithms, computer code, or anything else that could be copied and pasted into the recipient's answer. Similarly, students are encouraged to consult the Internet, but copying and pasting code from the Internet and submitting it for the class is not acceptable. The work that each student submits is expected to be their own, written with their own hand or typed on their own keyboard. For group work, work can be submitted by a single member but must include substantial contributions of all group members. Please contact the instructor to discuss instances of non-contributing group members.

Copyright

➢ Student Content

Participants in University of Illinois courses retain copyright of all assignments and posts they complete; however, all materials may be used for educational purposes within the given course. In group projects, only the portion of the work completed by a particular individual is copyrighted by that individual. The University of Illinois may request that students' materials be shared with future courses, but such sharing will only be done with the students' consent. The information that students submit during a course may, however, be used for the purposes of administrative data collection and research. No personal information is retained without the students' consent.

➢ Non-student Content

Student Behavior

➢ Student Conduct

Students are expected to behave in accordance with the penal and civil statutes of all applicable local, state, and federal governments, with the rules and regulations of the Board of Regents, and with university regulations and administrative rules.

For more information about the student code and handbook, see the CITL course policies page.

➢ Netiquette

In any social interaction, certain rules of etiquette are expected and contribute to more enjoyable and productive communication. The following are tips for interacting online via email or discussion board messages, adapted from guidelines originally compiled by Chuq Von Rospach and Gene Spafford (1995):

- Remember that the person receiving your message is someone like you, deserving and appreciating courtesy and respect.
- Be brief; succinct, thoughtful messages have the greatest effect.
- Your messages reflect on you personally; take time to make sure that you are proud of their form and content.
- Use descriptive subject headings in your emails.
- Think about your audience and the relevance of your messages.

- Be careful when you use humor and sarcasm; absent the voice inflections and body language that aid face-to-face communication, internet messages are easy to misinterpret.
- When making follow-up comments, summarize the parts of the message to which you are responding.
- Avoid repeating what has already been said; needless repetition is ineffective communication.
- Cite appropriate references whenever using someone else's ideas, thoughts, or words.

Communications

➢ Course Questions

Questions pertaining to the course should be posted on CampusWire. You can get to this forum from the course home page. Posting questions here allows everyone to benefit from the answers. If you have a question, someone else is probably wondering the same thing. Also, participants should not hesitate to answer questions posed by peers if they know the answers and the instructor has not yet responded. This not only expedites the process but also encourages peer interaction and support.

➢ Personal and Grade-Related Questions

Questions of a personal nature can be sent as a direct message on CampusWire, or if desired may be sent to the instructor's email address. If you need to send an email, include a subject that identifies the course number and nature of your question.

➢ Emergencies

If you have an emergency that will keep you from participating in the course, please notify your instructor on CampusWire or using email. Provide callback information in your message (if necessary). You should also notify your program director of any emergencies.

Sexual Misconduct Policy and Reporting

The University of Illinois is committed to combating sexual misconduct. Faculty and staff members are required to report any instances of sexual misconduct to the university's Title IX and Disability Office. In turn, an individual with the Title IX and

Disability Office will provide information about rights and options, including accommodations, support services, the campus disciplinary process, and law enforcement options.

A list of the designated university employees who, as counselors, confidential advisors, and medical professionals, do not have this reporting responsibility and can maintain confidentiality, can be found in the Confidential Resources section. Other information about resources and reporting is available at wecare.illinois.edu.

Student Wellness Resources

The University of Illinois strives to promote student success through the support of student psychological and emotional well-being. Please take advantage of the resources listed on the Student Affairs website.

Schedule

Modules

Module	Start Date	Contact Hours
0. Introduction and motivating problems	8/22/2023	1.3
1. Linear algebra review and intro to the Julia Language	8/24/2023	4.0
2. Open reproducible science	9/5/2023	2.7
3. Singular value decomposition and principle component analysis	9/12/2023	4.0
4. Fourier and wavelet transforms	9/21/2023	4.0
5. Regression	10/5/2023	4.0
6. Machine learning	10/19/2023	2.7
7. Neural networks	10/26/2023	2.7
8. Convolutional neural networks	11/2/2023	2.7
9. Data-driven dynamical systems	11/9/2023	2.7
Fall break	11/16/2023	2.7
10. Fairness in machine learning	11/28/2023	1.3
11. Final projects	11/30/2023	2.7

Homework

Title	Assigned	Deadline for 110% Credit	Deadline for 100% Credit	Deadline for 80% Credit
HW1: Linear algebra review and intro to the Julia Language	8/20/2023	Thu 8/24/2023, 12:00 CDT	Fri 9/8/2023, 17:00 CDT	Fri 9/22/2023, 17:00 CDT
HW2: Open reproducible science	8/29/2023	Tue 9/5/2023, 12:00 CDT	Fri 9/22/2025, 17:00 CDT	Fri 10/6/2023, 17:00 CDT
HW3: Singular value decomposition and principle component analysis	9/5/2023	Tue 9/12/2023, 12:00 CDT	Fri 9/29/2023, 17:00 CDT	Fri 10/13/2023, 17:00 CDT
HW4: Fourier and wavelet transforms	9/14/2023	Thu 9/21/2023, 12:00 CDT	Fri 10/13/2023, 17:00 CDT	Fri 10/27/2023, 17:00 CDT
HW5: Regression	9/28/2023	Thu 10/5/2023, 12:00 CDT	Fri 10/20/2023, 17:00 CDT	Fri 11/3/2023, 17:00 CST
HW6: Machine learning	10/12/2023	Thu 10/19/2023, 12:00 CDT	Fri 11/3/2023, 17:00 CDT	Fri 11/17/2023, 16:00 CST
HW7: Neural networks	10/19/2023	Thu 10/26/2023, 12:00 CDT	Fri 11/10/2023, 17:00 CST	Fri 11/24/2023, 17:00 CST
HW8: Convolutional Neural networks	10/26/2023	Thu 11/2/2023, 12:00 CDT	Fri 11/17/2023, 17:00 CST	Fri 12/1/2023, 17:00 CST

Title	Assigned	Deadline for 110% Credit	Deadline for 100% Credit	Deadline for 80% Credit
HW9: Data-driven dynamical systems	11/2/2023	Thu 11/9/2023, 12:00 CST	Fri 11/24/2023, 17:00 CST	Fri 12/8/2023, 17:00 CST

Exams

- Exam 1: Computational thinking, Thu 9/28/2023
- Exam 2: Signal processing, Tue 10/17/2023
- Exam 3: Machine learning: Thu 11/23/2023
- Final Presentation: Thu 11/30/2023 and Tue 12/5/2023

Modules

Module 0: Introduction and motivating problems

Module 0 Overview: In this module we will get to know each other and cover the format of the course, its contents, and expectations.

Module 0 Class sessions:

- Tue 8/22/2023, Introduction

Module 1: Linear algebra review and intro to the Julia Language

Module 1 Overview: we review linear algebra and introduce Julia programming language. You should already be familiar with linear algebra, so we will only briefly review it here. You're not expected to know anything about the Julia language before starting this class, but you are expected to have completed a basic computer programming class (similar to CS101) using some computing language.

Module 1 Learning Objectives: By the end of this module, you should be able to:

- Operate the Julia language programming interface, using Pluto notebooks
- Use variables, arrays, conditional statements, loops, and functions to process data using Julia
- Use Julia's built-in and library functions to operate on text and data

- Solve systems of equations using linear algebra in the Julia language
- Debug Julia programs to fix programming errors

Module 1 Class sessions:

- Thu 8/24/2023, Julia basics
- Tue 8/29/2023, Julia basics 2
- Thu 8/31/2023, Linear algebra review

Module 2: Open reproducible science

Module 2 Overview: This module covers tools and methods for ensuring your work is correct, understandable, and reproducible.

Module 2 Learning Objectives: By the end of this module, you should be able to:

- Apply the theory of 'tidy data' to wrangle a tabular dataset into tidy format, using for example the `groupby` and `combine` functions in DataFrames.jl
- Evaluate an unfamiliar dataset with exploratory statistical analysis, using for example the `filter` and `select` functions in DataFrames.jl as well as array indexing and basic descriptive statistics
- Use tools to deal with miss data
- Create exploratory visualizations for tabular, array, and image data using Plots.jl and StatsPlots.jl

Module 2 Class sessions:

- Tue 9/5/2023, Data Wrangling
- Thu 9/7/2023, Visualization

Module 3: Singular value decomposition and principal component analysis

Module 3 Overview: SVD and PCA fundamental algorithms for data processing and analysis. We will learn how they work and how they can be applied to gain insight from data.

Module 3 Learning Objectives: By the end of this module, you should be able to:

- Apply the SVD and PCA algorithms to create a low-rank approximation of a dataset

- Interpret the results of the algorithms in a given context, including the significance of the resulting values and how much of the variance in the original dataset is represented in the low-rank approximation

Module 3 Class sessions:

- Tue 9/12/2023, Singular Value Decomposition
- Thu 9/14/2023, Principal Components Analysis 1
- Tue 9/19/2023, Principal Components Analysis 2

Module 4: Fourier and wavelet transforms

Module 4 Overview: Fourier and wavelet transforms are powerful methods for coordinate transformation, data compression, and feature engineering and are used in almost every field of science and engineering.

Module 4 Learning Objectives: By the end of this module, you should be able to:

- Apply the FFT, Gabor transform, and Wavelet transform algorithms to determine the frequency spectra of a dataset
- Interpret the results of the algorithms in a given context, including the significance of the resulting values

Module 4 Class sessions:

- Thu 9/21/2023, Fourier Series
- Tue 9/26/2023, Fourier Transforms
- Thu 9/28/2023, Exam 1: Computational thinking
- Tue 10/3/2023, Spectrograms and Wavelets

Module 5: Regression

Module 5 Overview: In this module, we will learn how to use regression to predict the value of a dependent variable given a set of independent variables.

Module 5 Learning Objectives: By the end of this module, you should be able to:

- Apply the gradient descent algorithm to minimize error between a model prediction and observations
- Design and implement a linear regression model to predict a dependent variable in a dataset when given independent variables
- Apply regularization to the model to avoid overfitting

- Apply feature selection and engineering and coordinate transformation to a dataset to improve regression performance

Module 5 Class sessions:

- Thu 10/5/2023, Regression
- Tue 10/10/2023, Regularization
- Thu 10/12/2023, Model and feature selection
- Tue 10/17/2023, Exam 2: Signal processing

Module 6: Machine learning

Module 6 Overview: In this module, we learn about two popular machine learning algorithms: k-means and decision trees.

Module 6 Learning Objectives: By the end of this module, you should be able to:

- Implement the k-means algorithm to divide a dataset into clusters
- Design and implement a decision tree model to predict a dependent variable in a dataset when given independent variables

Module 6 Class sessions:

- Thu 10/19/2023, k-Means clustering
- Tue 10/24/2023, Classification trees

Module 7: Neural networks

Module 7 Overview: In this module, we will learn how to implement and use fully-connected neural networks.

Module 7 Learning Objectives: By the end of this module, you should be able to:

- Train a neural network to for regression and classification
- Identify and debug common problems with neural network training

Module 7 Class sessions:

- Thu 10/26/2023, Neural networks 1
- Tue 10/31/2023, Neural networks 2

Module 8: Convolutional neural networks

Module 8 Overview: In this module, we will learn how to implement and use convolutional neural networks.

Module 8 Learning Objectives: By the end of this module, you should be able to:

- Train a convolutional neural network to for regression and classification
- Identify and debug common problems with convolutional neural network training

Module 8 Class sessions:

- Thu 11/2/2023, CNN 1
- Tue 11/7/2023, CNN 2

Module 9: Data-driven dynamical systems

Module 9 Overview: In this module, we will apply the machine learning techniques we have learned so far to dynamical systems and the differential equations that describe them.

Module 9 Learning Objectives: By the end of this module, you should be able to:

- Apply gradient descent to t the parameters of a system of differential equations to observed data
- Implement a Neural ODE to make data-driven predictions of the evolution of a dynamical system

Module 9 Class sessions:

- Thu 11/9/2023, Parameter fitting for dynamical systems
- Tue 11/14/2023, Neural ordinary differential equations
- Thu 11/16/2023, Fall break
- Tue 11/21/2023, Fall break
- Thu 11/23/2023, Exam 3: Machine learning

Module 10: Bias in machine learning

Module 10 Overview: In this module, we will learn how to detect and minimize this bias in machine learning models.

Module 10 Learning Objectives: By the end of this module, you should be able to:

- Use disaggregated testing to detect bias in machine learning models
- Design and construct models to minimize any detected bias

Module 10 Class sessions:

- Tue 11/28/2023, Fairness in machine learning

Module 11: Final projects

Module 11 Overview: In this module we will present the results of our semester projects.

Module 11 Class sessions:

- Thu 11/30/2023, Final project presentations
- Tue 12/5/2023, Final project presentations