ENGR 103: Computational Methods of Data Science for Engineers

CRN 15175 | updated 1/8/2024

Spring 2024 Semester - Purdue University
LMBS 3285 - Monday 12:30-1:20 p.m.

Professor Sean Brophy
Office: ARMS 1217
Phone: 765-496-3316
Email: sbrophy@purdue.edu

Professor Michael Witt
Office: WALC 3053T
Phone: 765-494-8703
Email: mwitt@purdue.edu

Gloria Yutong Zhang (TA)
Office: LMBS 3285
Email: zhan3606@purdue.edu
Office hours: Tuesday 7-9 pm

Leo Y. Lin (TA)
Office: WALC 3053T
Email: lin1736@purdue.edu
Office hours: Wednesday 1-2pm & Thursdays 5-6pm or by appointment

Course Description

This course develops students' ability to design and use algorithms associated with the various components of a data science ecosystem using python. We will explore the major systems associated with the flow of big data into knowledge discovered through an inquiry process. Students will leave this course with a better understanding of the various professions associated with data science in engineering disciplines. Learning activities will focus on developing computational thinking abilities relevant to transforming data into knowledge used to make new discoveries and decisions for innovation.

The course structure is designed to align with the learning objectives for the ENGR 132 Transforming Ideas to Innovations II course that are associated with learning to model systems with numerical methods and empirical data to support analysis of system behavior and performance. Therefore, students will be learning to program in both MATLAB and python to solve problems presented in both courses. Unique to ENGR 103 will be the focus on the workflows associated with common data science problems in engineering, which can be generalized to other inquiry methods used in the sciences and business.

Learning Resources, Technology & Texts

- Python Crash Course, 3rd Edition, by Eric Matthes - available online without cost to students through Purdue Libraries.
- Google Colab will be our development environment. Students should bring a laptop computer to class and be prepared to work in small teams, in person, and online.

Learning Outcomes

- Develop ability to evaluate system performance with data and numerical models.
- Generate effective and efficient Python code to process and analyze data.
- Design experiments to analyze system performance using data.
- Develop code to process real time data acquisition systems.
- Model natural phenomena using computer simulations.
- Analyze and translate problems into algorithms composed of logical constructs and be able to generate programming-language-independent system charts and flow diagrams to represent those algorithms.
- Communicate results from an inquiry around large or big data.

Course Prerequisites

- All students enrolled in this section of ENGR 103 belong to the Engineering in the World of Data Learning Community that offers a variety of social and educational activities throughout the year. The learning community strives to increase students' awareness of data-related behaviors and ethics in the fields of engineering through combined assignments, learning and social activities, and other shared experiences.
- Completion of ILS 103 Introduction to Data Lifecycle Management.
- Previous programming experience is not required.

Course Structure

The primary modality for the class is face-to-face. The first half of the semester presents a python tutorial in eight modules, and the second half focuses on student-led, team-based projects. The major outcome for this course is to help students become familiar with generating python code to process data and generate knowledge. This requires spending time at the keyboard trying out the concepts. During the tutorial, there will be a Pre-Class Assignment (PCA) designed to help students become familiar with the topic for each week that will be due before each class meeting. During each class, we will perform an In-Class Assignment (ICA) to practice these concepts together and ask questions. ICAs will be submitted at the end of each class meeting. The other major outcome is to learn how to engage in an inquiry process like a data scientist. Four-person teams will work together on two projects to manage the process of making sense of a data set to answer a question. The team's Google Drive notebook will provide the report for their work along with a short presentation for each project. Students are required to participate in at least two learning community events during the semester.

Class Schedule

This is a *tentative* schedule for the semester that is subject to change, for example, to accommodate guest speakers or to address relevant, emerging issues and current events. The current schedule will be updated and maintained in the calendar in Brightspace.

Dates	Topics
January 8	**Welcome & syllabus review; Module 1** - Chapters 1, 2, 8
January 15	**Martin Luther King Jr. Day - No class**
January 22	**Module 2** - Chapters 3, 4, 5
January 29	**Module 3** - Chapters 6 & 7
February 5	**Module 4** - Chapters 9, 10, 11
February 12	**Module 5** - matplotlib, numpy
February 19	**Module 6** - data wrangling & pandas
February 26	**Module 7** - machine learning
March 4	**Guest lecture: Dr. Mohamed Yakout**, Senior Manager, Product Knowledge Team, Amazon.com
March 11	**Spring break - No class**
March 18	**Module 8** - deep learning
March 25	**Team project 1 work session**
April 1	**Team project 1 work session**
April 8	**Team project 2 work session**
April 15	**Team project 2 work session**
April 22	**Demos & year-end learning community celebration**

Assignments and Grading

Grades are based on 100 points available through 8 pre-class assignments (PCAs) and 8 in-class assignments (ICAs), two team-based projects, and participation in at least two learning community events. PCAs must be submitted before the class session, and attendance is mandatory for ICA submissions. Late assignments will not be accepted without prior approval of an instructor. There is no final exam.

Pre-class assignments	8 PCAs (4 points each)	32
In-class assignments	8 ICAs (4 points each)	32
Learning community participation	Attend 2 Learning Community events (3 points each)	6
Team projects	2 projects (15 points each)	30
Total		100

The following grading scale will be used:

Grade	Value
A	94-100 points
A-	90-93 point
B+	87-89 points
B	84-86 points
B-	80-83 points
C+	77-79 points
C	74-76 points
C-	70-73 points
D	60-69 points
F	Below 60 points

Communication

Brightspace will be the primary course website. Other platforms such as Gradescope and CATME may be used for submission of assignments and peer evaluation. Instructors may be contacted via Brightspace or by the email addresses listed above. The TA is available in-person during posted office hours or by appointment to meet online or in person at other times. A Discord server has been established for the Engineering in the World of Data Learning Community that can be used for informal communication among faculty and students, e.g., information about related seminars and events, social activities, current news and articles related to data, forming study groups, etc. Students are expected to read and respond to their @purdue.edu email on a frequent basis.

Professional Expectations

Everyone in our community helps shape the environment to be positive and productive for everyone else. Behaving professionally includes arriving for class on time and being prepared; focusing during class on ENGR 103; controlling your behavior to minimize distractions to those around you; and engaging with others in a respectful and professional manner. Instructors can deduct points from your semester total for behavior that is disruptive to your class or to your team's dynamics and performance. Students who are more experienced programmers are encouraged to help students who are beginner programmers.

Attendance

Students are expected to attend all class meetings. Attending class and actively participating is the best way to succeed in this class. In keeping with the Class Attendance and Absence Reporting Policy, only the instructor can excuse a student from classes or course responsibilities. For this course, excusable absences are limited to documented illness; extended absences noted as documented by the Office of the Dean of Students; or Purdue sanctioned events (e.g., conferences, club activities, varsity athletics, etc.) that do not conflict with an exam.

Be aware, having grounds for an excusable absence does not necessarily guarantee you will obtain instructor permission to miss class or course responsibilities. In the case of foreseeable events (e.g., Purdue sanctioned events with known dates, scheduled medical procedures, etc.), students are expected to provide proper documentation of the event prior to its occurrence with sufficient time for the instructor to make a decision about the suitability of the absence. Instructors reserve the right to decline to excuse an absence, with the understanding that such a decision must be consistent with the Purdue policy referenced above. In all cases where a student is requesting an excused absence, the student is expected to make a good faith effort to notify the course instructor in a timely fashion (prior to a class is considered timely, after the fact is generally not). The instructor may request documentation to corroborate your request and assign additional work to make up.

Please note: students having an excessive number of unexcused absences (more than 10% of the regularly scheduled class meetings) will automatically be dropped one letter grade regardless of their class standing.

Academic Integrity and Artificial Intelligence (AI)

Students are expected to follow the Purdue Honor Pledge and University policies related to plagiarism and academic honesty. Our learning community discussed and decided together that AI is a tool, much like a shovel or a calculator, and like any tool, it can be used properly or improperly. We decided that proper uses of AI assist (but do not replace) the process of learning, original thinking and work. We will be considerate of what data has been used to train generative models that may belong to others, be cautious of inaccuracies that may occur from

AI-assisted results, and be sure to clearly attribute what work has been done using AI in assignments.

Important University Policies and Resources

Students are expected to read and to all university policies, in particular those that are linked in Brightspace relating to Academic Integrity, Nondiscrimination, Accessibility, Mental Health and Wellness, Basic Needs Security, and Emergency Preparedness.