COMPSCI 390B Harnessing Data Science for Societal Good

Semester Spring 2024

Credits: 3

Instructor: Abhidip Bhattacharyya

abhidipbhatt@umass.edu

Office hours: TBD

Course Description: This is a project-based course in which students will explore using large-scale datasets and data analysis to address real-world problems of societal or industrial relevance in domains such as sustainability, health, and work with different techniques of data analysis and processing.

Each semester, the course will offer one or more real-world datasets and a selection of sample problems and students will define a project based on these datasets to address a real-word problem in a group setting. Students will explore modern data processing tools and software systems to build data processing pipelines for their chosen project. Throughout the course, students will be expected to present their project ideas, develop project proposals outlining their implementation plans, and conclude with a final presentation and report submission. This course provides students with the opportunity to integrate their analytical and collaborative skills for real-world problem-solving.

Pre-requisite: Undergraduates with a C or better in COMPSCI 240 (or STATISTIC 315 or 515) and COMPSCI 230.

Learning Objectives:
- Learn data analysis and processing tools necessary to address the chosen problem. Basics and theory behind some of the processes will be taught. However, there will be scope for exploration.
- Use datasets and data analysis for real-world societal good applications in domains such as energy, sustainability, and health.
- Leverage computing skills for solving real-world problems in a goal-oriented environment.
- Survey a problem domain that needs to be addressed.
- Practice collaboration skills through a group project.
- Critical analysis and reasoning about project results.
- Effective technical communication skills to present project results through written reports, a poster, and presentations.

Sample Project Idea:

A sample project idea is as follows: you have solar energy data from an energy plant, such as daily yield, monthly yield over some time. You have weather data for that region. Moreover, you

have data for past energy requirements of the region. The project could try to analyze the data and find answers to the following questions:

- What is the pattern for energy requirement vs. solar energy production of the area?
- What is the impact of weather in the production of solar energy?
- What is the impact of weather on energy requirements?
- Given the past data of productions, requirements and weather conditions can we predict the future state?

Attendance Policy: Students are expected to attend every lecture and participate in class discussions. Class participation will count for the grade. Makeups will be offered for valid and documented absences.

Textbook: None. Required readings will be free electronic materials provided by the instructor, including conference papers, textbook chapters etc.

Grade Weight:

5% Proposal Presentation

10% Written Project Proposal

5% Midway Progress Presentation

10% Written Mid-semester Progress Report

10% Final Project Presentation

20% Written Final Project Report

10% Intra-Group Feedback and Individual Evaluation

10% Peer Feedback (Inter-group Feedback)

20% In-class exercises and homeworks

Grade Scale:
Grading is on a letter scale, listed below. Grades will be rounded to the nearest integer.

A: 93-100%

A-: 90-92%

B+: 87-89%

B: 84-86%

B-: 80-83%

C+: 77-79%

C: 74-76%

C-: 70-73%

D+: 65-69%

D: 60-64%

F: below 59.5%

Minor revisions may be made to the grading rubric at the discretion of the instructor.

Course Outline:

week	topic	assignment
1	Lecture 1: Introduction, course description Lecture 2: Computing for common good concept; Project ideas	
2	Lecture 3: Review of some basics of statistics ● Distributions and statistics ● Hypothesis testing ● Confidence interval Lecture 4: Project-specific discussion and group creation; assign readings	
3	Lecture 5: Working with data ● Pandas ● Scikit learn ● Numpy Lecture 6: Working on proposals and discussions ● Identify the problem and why it is important ● What resources to be ventured ● Initial work plan towards solution	
4	Lecture 7: Classification and regression ● Linear regression ● Logistic regression Lecture 8: Proposal presentations - group 1 ● (Alternate topic for lecture Support vector machine)	Proposal due

5	Lecture 9: Proposal presentations - group 2 Lecture 10: Guest lecture on computing for common good (with elevator pitch from students)	
6	Lecture 11: (discussion section) Peer feedback ● Help another group predicting issues ● Provide constructive feedback Lecture 12: Advanced tools ● Pytorch	Peer feedback due
7	Lecture 13: Training a model ● Model training and optimization ● Loss function optimization and back propagation Lecture 14: (discussion section) ● How would you address some of the problems raised by the other group	
8	Spring break	
9	Lecture 15: Advanced topics ● Huggingface models ● Embedding and LMs Lecture 16: Progress presentations - group 1 ● (Alternate topic: LLMs and Vision-language models)	
10	Lecture 17: Progress presentations - group 2 Lecture 18: Guest lecture on computing for common good	Progress report due
11	Lecture 19: (discussion section) What can go wrong and how to address - group 1 Lecture 20: (discussion section) What can go wrong and how to address - group 2	
12	Lecture 21: Individual interview grading ● Instructor will meet individual students and assess their contribution in the group project Lecture 22: Individual interview grading ● Instructor will meet individual students and assess their contribution in the group project	
13	Lecture 23: Work in groups ● Planning the presentation	

	● Solving issues for integration ● Bring everything into the pipeline Lecture 24: Help for final presentations	
14	Lectures 25 and 26: Final presentations Dry-run (group presentation) Poster and demo session during final exam week.	Final report due

Academic Honesty Policy:
Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst. Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructor should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent (https://www.umass.edu/honesty/).

Accommodations Statement:
The University of Massachusetts Amherst is committed to providing an equal educational opportunity for all students. If you have a documented physical, psychological, or learning disability on file with Disability Services (DS), you may be eligible for reasonable academic accommodations to help you succeed in this course. If you have a documented disability that requires an accommodation, please notify me within the first two weeks of the semester so that we may make appropriate arrangements. For further information, please visit Disability Services (https://www.umass.edu/disability/)

Title IX Statement (Mandated Reporter Version): **(Please use the applicable version)**
In accordance with Title IX of the Education Amendments of 1972 that prohibits gender-based discrimination in educational settings that receive federal funds, the University of Massachusetts Amherst is committed to providing a safe learning environment for all students, free from all forms of discrimination, including sexual assault, sexual harassment, domestic violence, dating violence, stalking, and retaliation. This includes interactions in person or online through digital platforms and social media. Title IX also protects against discrimination on the basis of

pregnancy, childbirth, false pregnancy, miscarriage, abortion, or related conditions, including recovery. There are resources here on campus to support you. A summary of the available Title IX resources (confidential and non-confidential) can be found at the following link: https://www.umass.edu/titleix/resources. You do not need to make a formal report to access them. If you need immediate support, you are not alone. Free and confidential support is available 24 hours a day / 7 days a week / 365 days a year at the SASA Hotline 413-545-0800.

For purposes of Title IX reporting, I am a considered a "responsible employee" at UMass (https://www.umass.edu/titleix/about). **That means that if you tell me about a situation involving sexual assault, sexual harassment, domestic violence, dating violence, stalking, and retaliation, I must share that information with the Title IX Coordinator**. Making a report to the Title IX Coordinator is my legal obligation, meets the University's goal of providing members of our community with supportive resources they might need, and enables the University to obtain a more accurate picture of the extent of sexual violence in our community. **It will be completely up to you to determine if and how you want to work with the Title IX Coordinator's office. You will not be in trouble for reporting to me that you have experienced any of these situations, and the law prohibits retaliation against anyone who participates in a Title IX process.**