CSE 587:Data Intensive Computing

Spring 2024 , University at Buffalo

Today's Agenda

- Class Rules
- Class logistics

Course Staff

Professors
Shamsad Parvin
shamsadp@buffalo.edu
351 Davis Hall

Teaching/Grader Assistants:

Name	Email	Office Hour
kartikeya Sign	ksingh35@buffalo.edu	TBD/Piazza
Jialing cai	jialingc@buffalo.edu	TBD/Piazza

Name	Email
Sanket miland	sgadhave@buffalo.edu
Jaskirat Singh	jaskira2@buffalo.edu

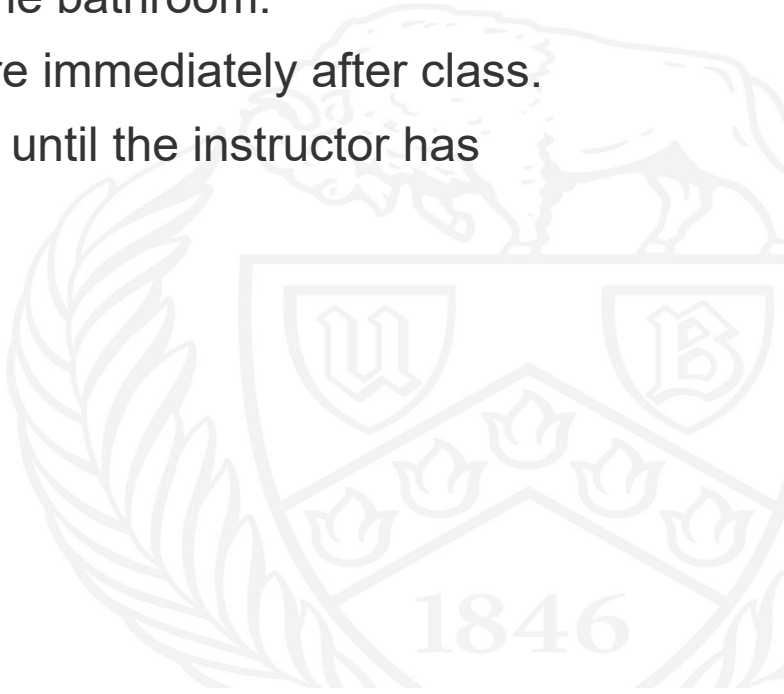
Class Rules

- Paying attention in class
- Not coming to class late or leaving early
- Not talking with other classmates while the instructor or another student is speaking.
- Turning off electronic devices including cell phones, pagers, and beeper watches
- Focusing on class material during class time.

Sleeping, talking to others, doing work for another class, reading the newspaper, checking email, and exploring the Internet are unacceptable and can be disruptive

Class Rules

- Interrupt me If You don't understand what we are talking about.
- Do not interrupt us for Whether you can use the bathroom.
- We will **not** answer questions about the lecture immediately after class.
- Not packing book bags or backpacks to leave until the instructor has dismissed class.



Logistics

- **Brightspace course**
  - All course materials, links, schedule, extra resources
- **Course Forum (Piazza)**
- **https://piazza.com/buffalo/spring2024/cse487**

Access code : DICSPSecC

  - All discussion for the course is hosted here – check regularly
- Brightspace
  - Quiz, Project and Assignment submission, grades
- **Please keep class discussions on Piazza**
- **Always include [CSE 487] or [CSE 587] in the subject line when emailing Also mention the section name Section A**

Grading

Score (x)	Letter Grade	Quality Points
95% ≤ x ≤ 100%	A	4
90% ≤ x < 95%	A-	3.67
85% ≤ x < 90%	B+	3.33
80% ≤ x < 85%	B	3
75% ≤ x < 80%	B-	2.67
70% ≤ x < 75%	C+	2.33
65% ≤ x < 70%	C	2
60% ≤ x < 65%	C-	1.67
55% ≤ x < 60%	D	1
0% ≤ x < 55%	F	0

Grade Breakdown:

- Homework: 10%
- Quiz: 10%
- Project: 40%
- Midterms: 20%
- Final Exam: 20%

Textbook

- There are two recommended texts:
- 1. Doing Data Science: Straight Talk from the Frontline, 1st Edition Author(s): Cathy O'Neil and Rachel Schutt ISBN: 978-1449358655 Publisher: O'Reilly Media
- 2. Joel Grus. Data Science from Scratch: First Principles with Python 1st Edition, 978-1491901427, O'Reilly media. 2015.

- We will be using many other references and online sources throughout the semester. Links will be posted on the piazza later. .

HomeWorks (10%)

- Two/Three homework assignments that cover lecture and reading material
- Submit all assignments in Brightspace and All homework should be done individually.
- The deadlines are hard and will not be extendable.
- Your written assignments/Homework will need to be clearly written, succinct, and accurate.
- We will allow multiple submission windows, and consider the last submission
- **487 Vs. 587** : If you registered in CSE 487, you only need to submit HW 1 and 2 (but not HW 3); for 587, you need to submit three HWs.

Quiz (10%)

- There will be ten quizzes in total.
- Quizzes will be given to test students' knowledge On the course.
- Quizzes are timed (10-15 minutes).
- The best seven quizzes will be counted toward the final grade.
- The quizzes will be conducted in the class.
- Quiz dates will not be announced in advance.
- It will be a surprise quiz. (No retake for the quiz)
- You will have single attempt for the quiz.
- Quizzes will be multiple choice/true false or fill in the blank

Exam (Mid-term + Final Exam) (20%,20%)

- There will be two exams during the course.
- These exams will cover the mandatory readings and topics discussed in class.
- The exam will be a Closed book paper and a pen-based exam.
- The first will be an "in-class" midterm exam.
- The second will be a final exam during the University's final examination period at the end of the semester (The final Exam date cannot be changed).

Project (40%)

The details of project will be discussed later

The project will be group project (2 or 3 students)

Office hours

TAs will hold office hours on weekdays (Mon-Fri) at different times.

You can find the TAs Office hour in Piazza

We will hold Two Workshops Day before the project Due Date

Late Policy

- **No late submission will be accepted for Homework**. (The entire homework and project schedule will be posted on the schedule page in Brightspace, so please plan accordingly.
- You will lose 20% of the points for a Project phase submission for every 24 hours it is late. (Project only)
- We will grant no-penalty extensions due to extreme circumstances (e.g., medical emergencies).
- If something comes up, please contact the TA as soon as possible.

Academic Integrity

Collaboration, AI, Extra Resources

Do…

- Work together to brainstorm ideas
- Explain concepts to each other
- Discuss course content
- Include a list of your collaborators on all submitted work

Do Not…

- Write solutions when working together
- Describe the details of solutions to problems
- Leave your work in a place where it is accessible to another student

Academic Integrity

- Don’t cheat! You will be caught and punished
- All assignments will be checked for plagiarism / unauthorized collaboration!

  (See the course syllabus for more details)

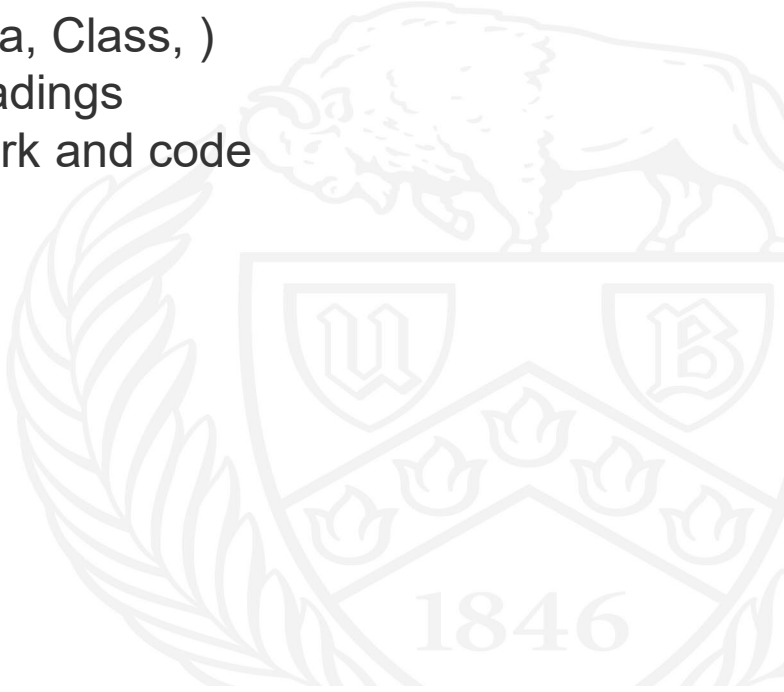
The Homework and Project must be your Own Work

- Plagiarism is **not** tolerated. See UB policy on Academic Integrity
- Please ask instructors and TAs, if you are unsure.
- If caught: 0 on the assignment, -5% on your final grade

Resource Policy

Do…

- Use materials provided by course staff (Piazza, Class,)
- Use materials from the course textbook or readings
- **Cite** all materials you reference for written work and code
- ChatGpt is prohibited in this course



Resource Policy

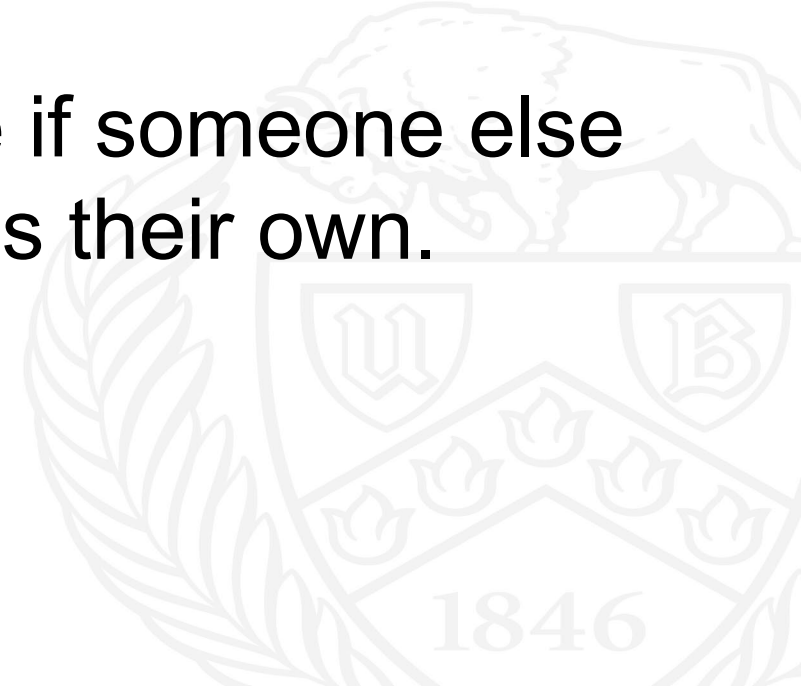
Do NOT…

- Reference random videos on YouTube that “helped you solve the problem”
- Hire “private tutors”
  - Save the money from Chegg
  - If you’re not doing the work yourself, you’re not learning
- Reference exact solutions found online

If you are caught using unauthorized resources in project, you get an F

Other Ways to Get an F

You are liable/punishable if someone else submits your work as their own.



Ways to Avoid an F

Don’t Cheat…

Ways to Avoid an F (amnesty policy)

Don’t Cheat…but we understand mistakes are made.

We will grant amnesty for any AI violation **IF** you tell us about it **BEFORE** we discover it

Why does Academic Integrity Matter?

But it doesn't JUST hurt you…it also hurts the credibility of UB and its graduates!

How should you assess success in this course?

Not by grade…

By new concepts you learn about data-intensive computing

By new skills you develop to solve data-related problems

By new knowledge you gain about data applications, python libraries, MR, streaming data, etc.

…but do this and the grade will come too

What is the course about?

Foundational concepts in data intensive computing

Useful tools

Go from small data to big data

Go from big data to streaming data

Identifying a problem

Data Acquisition

Understanding the data

Extracting features

Analysis

Visualizing

What is the course about?

Foundational concepts in data intensive computing

Useful tools

Go from small data to big data

Go from big data to streaming data

- Python
- Hadoop
- MapReduce
- Spark

What is the course about?

Foundational concepts in data intensive computing

Useful tools

Go from small data to big data

Go from big data to streaming data

Go from small, structured data, ie excel tables to big unstructured data (mainly text)

Big data data structures and algorithms (Hadoop, MapReduce)

What is the course about?

Foundational concepts in data intensive computing

Useful tools

Go from small data to big data

Go from big data to streaming data

New challenges with streaming data

What is it? Social media and enterprise data

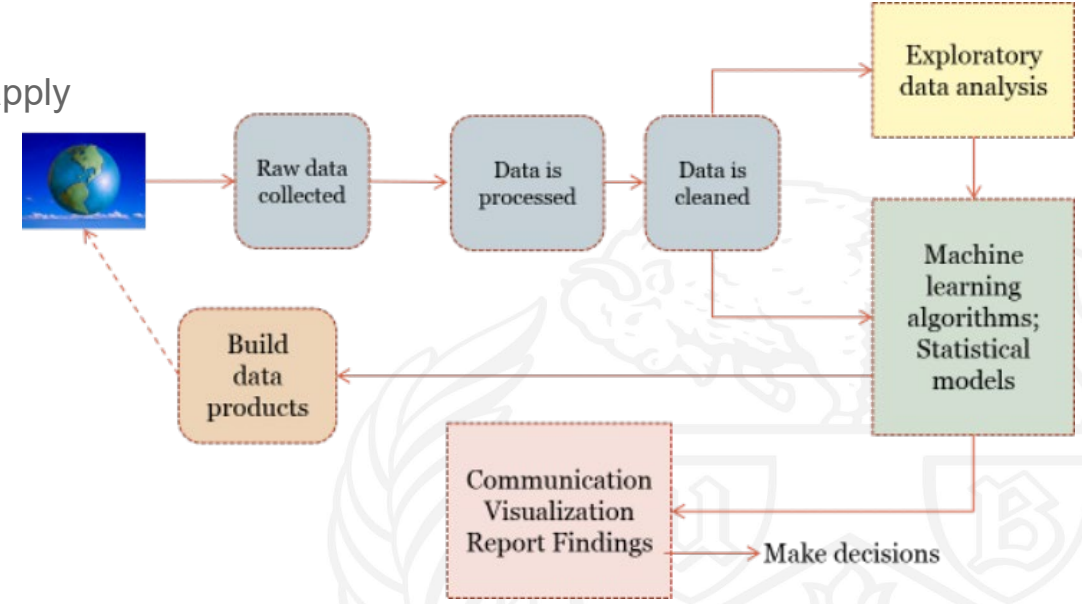
How do we characterize it and manage it (ie Spark streaming)

What will you learn?

Basic data analytics processes and how to apply them

Big data infrastructures and algorithms

Newer challenges (and how to handle them)



Questions?