Big Data and Data Science CSCE 4300

Instructor Information

Name: Dr. Zeenat Tariq. Pronouns: she/her, Email: zeenat.tariq@unt.edu,

Virtual Office: Available on Teams on Tuesday from 11:00 AM to 12:00 PM

Teaching Assistant(s) TA:

Assistant 1:

Poli Nemkova, PhD Candidate

Email: poli.nemkova@unt.edu (please add course number in the subject line)
Zoom link: https://unt.zoom.us/my/nemkova
Office hours (please email me in advance if you plan to join):
Monday: 1-3pm (via Zoom)
Friday: 3-4pm (via Zoom)
If none of those timeslots work for you, please email me to find another appointment time.

Instructional Assistant:

Name: Sai Rushitha Bhimavarapu

Office hours are scheduled as follows:

- **Online on Tuesday and Thursday from 11:00 A.M to 1:00 P.M through Microsoft Teams /Email**: SaiRushithaBhimavarapu@my.unt.edu
- If necessary, you can schedule in-person office hours through an appointment.

Course Description, Structure, and Objectives

This course provides a comprehensive introduction to the field of Big Data and Data Science. Students will explore both foundational and advanced topics, including the nature of structured and unstructured data, real-world applications, computational strategies, data visualization, and the ethical considerations of data science. Special attention is given to the impact of Big Data on

scientific discovery, engineering solutions, and business innovation, as well as the tools and infrastructures used in large-scale data analysis.

Students will gain knowledge of scalable computing resources and platforms, including Amazon Web Services (AWS), Hadoop, Spark, and familiarity with Cassandra. The course emphasizes a broad understanding of Big Data and Data Science concepts, methods for data acquisition, cleaning, and analysis, and the application of analytics techniques in cloud and distributed computing environments. Distributed and parallel computing concepts are introduced to support the understanding of platforms such as Hadoop and Spark.

Upon successful completion of this course, students will be able to:

- Demonstrate an understanding of data science and its transformative role across disciplines such as science, engineering, and industry (including market-driven applications).
- Analyze and interpret Big Data to generate insights that support informed decision-making and innovation.
- Apply Big Data tools and techniques to acquire, assess, clean, and prepare large datasets for analytical use.
- Demonstrate familiarity with scalable computing resources and Big Data ecosystems, including cloud-based platforms (AWS), distributed data processing frameworks (Hadoop and Spark), and related computing environments, with a conceptual understanding of distributed and parallel computing.
- Utilize commonly used data science tools and environments such as Jupyter Notebook, Python, mpi4py, and command-line interfaces for data analysis tasks.
- Apply data science, software development, and computing principles to address real-world challenges in both industry and academic research, including model selection and validation, predictive modeling, and parameter tuning.

Required/Recommended Materials

This course does not have formal prerequisites. However, students will benefit from familiarity with programming (e.g., Python), foundations of data structures, and introductory statistics. Knowledge of these topics is strongly recommended to support your understanding of the course material and to successfully complete assignments.

The course includes video lectures, readings, tutorials, assignments, and exams, all scheduled on the course calendar. Detailed explanations and clarifications will be provided through discussion boards and other course platforms, and opportunities for interaction are built into each module to help students fully engage with the material. Additional resources, such as online tutorials or reference materials, may also be useful to support your learning.

Since this course includes digital components**,** reliable internet access is essential. Students must be able to access the Canvas Learning Management System to view course materials and participate fully in all course activities.

How to Succeed in this Course

I am committed to being as accessible and supportive as possible throughout this course. All lectures and video content for each module will be posted online no later than each Monday. If Monday is a university holiday, materials will be posted by Tuesday end of the day. A discussion board will be available for each module, where students are encouraged to ask questions and engage with peers. Discussion board posts will generally receive responses by the end of the week.

Students are encouraged to participate actively and respond to their peers' questions in a professional and respectful manner. Discussions should remain focused and maintain appropriate conduct. General course questions should be posted on the Q&A forum of the discussion board so that all students can view the discussion and benefit from shared information. For individual questions or concerns, students may contact the instructor via email.

Assignment feedback is typically returned within approximately one week of the due date. If additional time is needed, an announcement will provide an updated timeline.

How to Reach Me?

You can connect with me via email or Teams. Teams may be used for one-on-one meetings, and if you are using your UNT email, no link or special access is needed. However, all official communications, requests, or private concerns should be sent via email. Responses to emails are typically provided within 24–48 hours on weekdays; if a reply is not received within this timeframe due to high traffic or other circumstances, a polite follow-up is welcome. Weekend replies are not guaranteed.

Office hours provide an opportunity to ask for clarification or receive support with understanding course material. Additional office hours may be offered as the semester progresses to provide extra support when needed. I encourage you to connect with me and/or your TA whenever you need guidance or clarification.

Use Your Resources! We all need additional support from time to time. Don't suffer in silence or feel you have to go at this alone. University life is filled with challenges! UNT has so many great resources to support you.

Office of Disability Access

The University of North Texas makes reasonable academic accommodation for students with disabilities. Students seeking reasonable accommodation must first register with the Office of Disability Access (ODA) to verify their eligibility. If a disability is verified, the ODA will provide you with a reasonable accommodation letter to be delivered to the faculty to begin a private discussion regarding your specific needs in a course. You may request reasonable accommodations at any time; however, ODA notices of reasonable accommodation should be provided as early as possible in the semester to avoid any delay in implementation. Note that students must obtain a new letter of reasonable accommodation for every semester and must meet with each faculty member prior to implementation in each class. Students are strongly encouraged to deliver letters of reasonable accommodation during faculty office hours or by appointment. Faculty members have the authority to ask students to discuss such letters during their designated office hours to protect the privacy of the student. For additional information, refer to the Office of Disability Access website (http://www.unt.edu/oda). You may also contact ODA by phone at (940) 565-4323.

Additional Resources I encourage you to take advantage of are:

- UNT's Counseling and Testing Services can provide psychological counseling and academic testing
- UNTWell provides FREE individual and group counseling as well as a vocational assessment
- UNT Food Pantry is a great resource if you or someone you know experiences food insecurity
- Academic Success Center can support you in your academics

Our Learning Community

My goal is to foster a learning environment where every student feels welcome, respected, and valued. Your unique experiences and perspectives are important contributions to our learning community, and I encourage you to share them openly.

This course relies heavily on collaboration and group discussions. Your active participation not only enhances your own learning but also enriches the experience for your peers. To ensure everyone benefits, please engage fully, contribute thoughtfully, and meet the deadlines provided.

When interacting with both classmates and me, I ask that you maintain an open mind and communicate with respect. We may not always agree, and that's okay. Our differences can lead to meaningful conversations and deeper understanding when approached with curiosity and mutual respect.

As a shared foundation, I encourage all students to review the UNT Code of Student Conduct to ensure we are all aligned in our expectations for a positive and productive classroom experience.

Assessing Your Work

Course Activities & Assessments (100 points total)

- Quizzes (4 @ 5 points each, 20 points total)
- Assignments (10 @ 3 to 4 points each, 39 points total)
- Exams (2 @ 20 points each, 40 points total)
- Software Installations (1 point total)

Grading

- A: 90-100% (Outstanding, excellent The student performs well above the minimum criteria.)
- B: 80-89% (Good, impressive The student performs above the minimum criteria.)
- C: 70-79% (Solid, college-level The student meets the criteria of the assignment.)
- D: 60-69% (Below average The student fails to meet the minimum criteria.)
- F: 59 and below (Sub-par The student fails to complete the assignment.)

Grade-Related Policies

Late Work

I will not accept late work in this course. All work turned in after the deadline will receive a grade of zero unless the student has a university-excused absence and provides documentation within 48 hours of the missed deadline.

Turnaround Time

We make every effort to return your graded assignments within one week after the assignment due date. Delays will be updated during class or via Canvas announcements.

Grade Disputes

If you wish to dispute a grade, you must wait at least 24 hours after the grade is posted before contacting the TA or instructor. This period allows you to carefully review the assignment guidelines and thoughtfully reflect on the work you submitted.

Grade concerns should first be directed to the TA or instructional assistant who graded the assessment, within five calendar days of the grade being posted. If a resolution cannot be reached at that level, you may then contact the instructor by email to request a virtual meeting. Please note that grades will not be discussed via email.

For a productive meeting, you must come prepared with specific evidence supporting your request for a grade review. If you miss a scheduled meeting or do not contact the instructor within seven calendar days of the grade being posted, you will lose the opportunity to dispute the grade.

Extra Credit

To differentiate between student levels, bonus points are structured differently for undergraduate and graduate students to reflect varying academic expectations. Eligibility criteria and point allocation will be clearly communicated in the course materials. Course activities and assessments are designed to support learning and performance for all students.

Usage of GenAI Tools:

Throughout the semester, you will or may use specific Generative AI (GenAI) tools for certain assignments, with guidance on responsible use. These assignments help build ethical resilience and GenAI literacy, preparing you for careers in a GenAI-oriented workforce.

> In this course, Generative AI (GenAI) tools (e.g., Claude, ChatGPT, Gemini) may be used **by the instructor** to enhance materials, streamline tasks, generate prompts, create scenarios, draft syllabi, build study guides, and analyze performance. I will always disclose how GenAI is used, and I expect transparency from students as well.
>
> For students, **unauthorized use of GenAI tools for assignments, exams, or coursework is strictly prohibited**. Using GenAI content without proper credit or substituting your own work with GenAI undermines the learning process and is considered a violation of academic integrity under the UNT Honor Code. Tools such as Grammarly, spellcheck, predictive text, speech-to-text, or translation tools are also not allowed if they blur authorship or misrepresent independent work.
>
> Any use of GenAI tools for coursework must follow the rules outlined above, and students should seek clarification if they are unsure about what is permitted. The goal is to foster the development of independent thinking, critical skills, and ethical digital literacy.

Important Course Disclaimer

I reserve the right to modify course policies, the course calendar, assignment or project point values, and due dates as necessary throughout the semester. Any changes will be communicated in advance through official course announcements.

Course Requirements/Schedule (Tentative)

Provided below tentative list of topics along with assignments. This list can change during the course, and you will be notified of such changes wherever possible. The weekly plan may be adjusted as needed to align with the university calendar, including Spring Break.

Unit 0: Course Overview, Programming Environment, and Course Expectations

Week	***Topic***	***Assignment Due***	***Points Possible***	***% of Final Grade***
Week 1	*Syllabus* *Course Overview and Expectations: Programming Environments*	*Assignment-0.1 (Jupyter NB)* *Assignment-0.2 (VM and VM File share)*	*0.25 pts.* *0.75 pts.*	*0.25%* *0.75%*

Unit 1: Introduction to Python, Data Structures, and Visualization

Week	***Topic***	***Assignment Due***	***Points Possible***	***% of Final Grade***
Week 2	*Intro. Python Programming Data Structures*	*Assignment-1*	*4 pts.*	*4%*
Week 3	***Quiz-1*** *2D Visualization , 3D Visualization*	***In canvas*** *Assignment-2*	*5 pts..* *4 pts.*	*5%* *4%*
Week 4	*DataFrames: Pandas and PySpark*	*Assignment-3*	*4 pts.*	*4%*

Unit 2: Data Science, Machine Learning Algorithms

Week	***Topic***	***Assignment Due***	***Points Possible***	***% of Final Grade***
Week 5	*Statistics for Data Science: Introduction to Machine Learning* *Polynomial Regression*	*Assignment-4*	*4 pts.*	*4%*
Week 6	***Quiz-2*** *Logistic Regression, Confusion Matrix*	***In Canvas*** *Assignment-5*	*5 pts.* *4 pts.*	*5%* *4%*

Week 7	***Exam – 1 (Weeks 1 through 6)***	***In Canvas***	*20 pts.*	*20%*
Week 8	*Image Processing using KNN Classification*	*Assignment-6*	*4 pts.*	*4%*

Unit 3: Big Data

Week	***Topic***	***Assignment Due***	***Points Possible***	***% of Final Grade***
Week 9	*Introduction to Big Data Hadoop Distributed Computing* ***Quiz-3***	*Assignment-7 In Canvas*	*4 pts.* *5 pts.*	*4%* *5%*

Unit 4: Parallel Computing

Week	***Topic***	***Assignment Due***	***Points Possible***	***% of Final Grade***
Week 10	*Intro. to parallel computing Installing and Programming with MPI4PY*	*Assignment-8*	*4 pts.*	*4%*
Week 11	***Quiz-4*** *Introduction to Deep Learning Dimensionality Reduction / PCA Clustering / PyTorch*	***In Canvas*** *Assignment-9 Assignment-10*	*5 pts.* *4 pts.* *3 pts.*	*5%* *4%* *3%*

Unit 5: Amazon Web Services (AWS)

Week	***Topic***	***Assignment Due***	***Points***	***% of Final***

			Possible	***Grade***
Week 12	*Amazon Web Services Installation and configuration*	*Possible Bonus Activity*		
Week 13	*Hadoop in AWS*			
Week 14	*Cassandra in AWS*	*Possible Bonus Activity*		
Week 15	*Spark in AWS*			
Week 16	***Exam-2 (Weeks 8 through 15)***	***In canvas***	*20 pts.*	*20%*
	TOTAL Points		**100**	**100%**

Eagle Alert

Students will be notified by Eagle Alert if there is a campus closing that will impact a class, and note that the proposed course schedule is subject to change. For more information on emergency notification and procedures, please review the Emergency Notifications and Procedures Policy (PDF).

Course Evaluation

Student Perceptions of Teaching (SPOT) is the student evaluation system for UNT and allows students the ability to confidentially provide constructive feedback to their instructor and department to improve the quality of student experiences in the course.

Academic Integrity

You are expected to adhere to the university's Academic Integrity Policy (PDF) (https://policy.unt.edu/sites/default/files/06.049_Standard%20Syllabus%20Policy%20Statements_supplement. pdf).

Syllabus Policy

Grades are based on mastery of the content. As a rule, I do not grade on a "curve" because that is a comparison of your outcomes to others. I do, however, encourage you to find opportunities to learn with and through others. Explore Navigate's Study Buddy (https://navigate.unt.edu) tool to join study groups. Maximize your learning with our coaching staff at the Learning Center. Focus on areas where you are struggling in this course by attending scheduled study group sessions with me the week before each exam. Forward together!

or

Every student in my class can improve by doing their own work and trying their hardest with access to appropriate resources. Students who use other people's work without citations will be violating UNT's Academic Integrity Policy. Please read and follow this important set of guidelines for your academic success (https://policy.unt.edu/policy/06-003). If you have questions about this or any UNT policy, please email me or come discuss this with me during my office hours.

Attendance and Participation

This is a fully online course. In this course, "attendance" means **regularly accessing the course, completing readings and tutorials, participating in discussion boards, and submitting assignments on time**. All course materials, including quizzes and assignments, will be made available in a timely manner. Students are encouraged to follow the materials carefully. While participation in discussions is not graded, engaging thoughtfully with posts and responding to peers can help you better understand the course content.

If you are unable to meet course deadlines due to an emergency, please notify me as soon as possible. Consistent engagement and timely communication are important for your success in this course.

If you have questions or concerns, feel free to reach out to the instructor or course TA. See the Canvas pages for TA office hours and contact information.