Syllabus
CSCE 5300-002: Introduction to Big Data and Data Science (grad)
Fall 2021 at University of North Texas
Lectures: Saturday 5:30 pm – 8:20 pm BLB 180 UNT Main Campus

Instructor
Name: Abdelnasser Ouda, Ph.D.
Office Location: F286 Discovery Park
Office Hour: Thursday 2:30pm- 4:00 pm.
Email: abdelnasser.ouda-@unt.edu

Graders
Sai Kiran Rasthapuram SaiKiranRasthapuram@my.unt.edu
Naga Maheswara Reddy Gunta nagamaheswarareddygunta@my.unt.edu

<table>
<thead>
<tr>
<th>Name</th>
<th>Office Hours</th>
<th>Room</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sai Kiran Rasthapuram</td>
<td>Tuesday 2:00 am– 4:00 pm</td>
<td>F232</td>
<td><a href="https://unt.zoom.us/j/87811615434?pwd=bzB5UFFjc0R2RGVuQldNdHNFSWVDUT09">https://unt.zoom.us/j/87811615434?pwd=bzB5UFFjc0R2RGVuQldNdHNFSWVDUT09</a></td>
</tr>
<tr>
<td></td>
<td>Friday 12:00 am–2:00 pm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naga Maheswara Reddy Gunta</td>
<td>Tuesday 12:00 pm–2:00 pm</td>
<td>F232</td>
<td><a href="https://unt.zoom.us/j/81770429043?pwd=cmkxZ29sa2xwVVZ5YXhIdENYM3kydz09">https://unt.zoom.us/j/81770429043?pwd=cmkxZ29sa2xwVVZ5YXhIdENYM3kydz09</a></td>
</tr>
<tr>
<td></td>
<td>Wednesday 11:30 am-1:30 pm</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Communication Expectations: You will be expected to regularly check university email and attend class regularly. When you miss a class, you are expected to check the course calendar shortly after class for assignments, quizzes, and other materials. Questions that were not answered in class are best asked before or after class. For quick questions, email is preferred and you can expect a response within 24 hours during the work week (M-F). With regard to a group project, it is beneficial to meet with multiple people simultaneously (rather than separate individual meetings discussing the same project) so please coordinate with your group to schedule a project-related meeting.
Course Description

Introduction to Big Data and Data Science includes an overview of the field, technical challenges, computational approaches, practical applications, structured and unstructured data processing, empirical methods in computer science, data analytics and learning, data visualization, privacy and ethics. Emphasis will be on Big Data and its effect on other topics within Data Science, its technical characteristics, and state-of-the-art Big Data analytics architectures and tools.

Course Structure

Time: Sat 5:30pm - 8:20pm

Location: BLB 080

BLB – Business Leadership Building, 1307 W Highland St., Denton, 76207

Credit hours: 3

Tentative topics

- Statistics for Data Science
- Data Visualization
- Basic data structures and data frames
- Structured and Unstructured Datasets
- Image processing
- Machine learning concepts
- Deep learning concepts
- Distributed Computing: High Performance Computing, Cloud computing
- Big Data / parallel processing architectures (e.g. Spark/Hadoop)

Course Prerequisites

None

Programming Environment

Python

Course Objectives
By the end of the course, students will be able to:

1. Understand the advances in Big Data era, challenges, and opportunities for improving outcomes.
2. Use Data Science and Big Data tools to obtain, assess, and prepare data for analysis.
3. Articulate key advances in contemporary Data Science and describe the skillsets needed to be successful in a data science career.
4. Manage collections of data, create automated processes for analysis, use collaborative tools, and rapidly report quantitative findings.
5. Understand application of data science tools across domain areas. These include model selection and validation, predictive modeling and parameter tuning.
6. Become familiar with parallel and distributed computing environments and the capabilities they offer to support data analytics.

Materials

All materials (readings, videos, tutorials, assignments, and exams) will be accessible online and posted on the course calendar on the respective class day at the latest. Readings will all be fairly dense, so please search for additional resources (e.g. wikipedia, coursera lectures) as needed. All attempts will be made to provide sufficient resources for everyone.

Textbook: No textbook is required. Students are encouraged to consult online sources which will be referenced throughout the course and linked in the course calendar where appropriate.

For example,

- The Python 3 tutorial documentation: https://docs.python.org/3/tutorial/ (Links to an external site.)
- The scikit-learn documentation: https://scikit-learn.org/stable/documentation.html (Links to an external site.)

Course communication: We will be using Canvas Discussion

- Feel free to use the forum to ask questions, ask about partners for problem sets, or to make comments that the rest of the class might find useful.
• The forum is primarily for timely, supplementary communications; the course calendar will be the definitive source of requirements and course expectations.
• Contact the instructor if you are not available to access the discussion forum after the first week of the course.

Minimum Technology Requirements

• Computers are **required for exams**
  o [Canvas Technical Requirements (Links to an external site.)](https://clear.unt.edu/supported-technologies/canvas/requirements)
  o **In-class exams**: You will need to bring a laptop on exam days. These will be done individually on your computer in class. You will be expected to connect to the UNT wireless network.
  o Quizzes and exams will use the Canvas quiz system
• Computers are **optional during class, outside of exam times**
  o There will be occasional in-class activities along with lectures that computers are not required but may help in understanding and applying the material.
• Students will be expected to Download and install Anaconda Python version 3 and be able to open a Jupyter notebook.

Course Requirements

**Readings/Tutorials**: These will be available on the course calendar. Expectations of what is learned will be discussed in class and, as will be clear in the first few weeks, assignments will test your knowledge on the most important aspects of the readings/tutorials only.

**Assignments**: Assignments are designed to engage you in your learning, so you can begin to apply these principles in practice and tailor them to your needs. Unless otherwise specified, assignments are due at midnight exactly one week (or two weeks if there is a quiz) after they are assigned (allowing for questions in class on the day the assignment is due). Results should be presentable, with appropriate comments for someone to follow what you have done. Assignments can be done individually or in a group of no more than 3 people, but must be submitted to Canvas individually unless otherwise specified. Wherever groupwork is involved annotate your project document or include comments covering your partners information (Name and email) so, I would know that its a collaborative work and not a copy. Notably, low-point assignments are more for learning and self-evaluation than assessment.
Concept Paper: There are far more tools and techniques than we can possibly cover, many of which may be particularly relevant to your interests. Also, the ability to distill complex topics into a form useful for the audience is a critical skill to develop. **Students will be expected to present one concept, tool, or technique which goes beyond what is covered in the course.**

Submit a two-page paper and a 5 min pre-recorded presentation of your concept. Schedule a meeting with the instructional team for project selection.

**Quizzes/Exams:** There will be six quizzes and a final exam. Quizzes/Exams will resemble assignments but are targeted to develop your data science skills. Wherever possible, a new data set with a slightly different request for tasks will be involved. **Quizzes/Exams must be taken in the classroom.** Time will be limited, and all work will be individual. You are strongly encouraged to attempt to solve the tasks iteratively and incrementally - write code that works first, but works poorly, and improve from there, rather than attempting to write perfect code first. Quizzes/Exams are cumulative, but the emphasis will be on the newer material. Quiz/Exam days will be fixed and notified by next week. Prior arrangements can potentially be made without loss of points, but have to be discussed in advance.

**Missed quizzes/exams:** Quizzes/Exams cannot be missed without prior arrangements or later documented proof of extenuating circumstances.

**Grading**

Grades are determined by a simple points system, with a total of **at least** 100 pts given though more than 100 points are likely. The expected distribution of points is given below, with the exact scale determined by point values given for each assignment, presentation, competition, project, or quiz/exam - this is subject to minor modification based on actual points given. Note, due to the nature of the course, quizzes/exams are a significant means of establishing your final grade, so please complete the assignments in a timely way and study appropriately prior to each quiz/exam.

- Assignments: 40 pts (4 pts each) (individual submission)
- Concept Paper: 5 pts (individual work)
- Quizzes: 5 pts (2.5 pts each) (individual work)
- Project 30 pts (group work)
- Final Exam: 20 pts

**Grading Scale:** A=90, B=80-89.9, C=70-79.9, D=60-69.9, F=0-59.9 pts. No exceptions. If class grades are low (e.g., I expect most students will end with A’s and B’s), extra quizzes or assignments will be given to add points to the class. (Note, these grades are based on points and not percentages, so if 120 points are given, you only need 90/120 for an A!)
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. Spot evaluations will be available a few weeks before the end of the semester. Will notify the days soon!

Course Policies

Examination Policy

Exams will be on the computer using the Canvas quiz system. You need to bring a laptop on the appropriate exam days if in person. Exams are open book, open note, and open internet unless otherwise specified in advance, however, no communication with others other than the instructor and the TA is allowed in any form (e.g., email, chatting, etc.). Exams must be taken at the same time whether remote (with camera on) or in-person unless special accommodations have been made through the Office of Disability Accommodation (ODA). Other accommodations must be given by prior arrangement with the instructor, otherwise documentation proving an extenuating circumstance will have to be provided after the missed exam. Time will be limited, and all work will be individual. **Anyone involved in cheating (copying others’ solutions, helping others cheat, etc. - with an exception to group projects) will be reported to the university.** You are strongly encouraged to attempt to solve the tasks iteratively and incrementally - write code that works first, but works poorly, and improve from there, rather than write perfect code top to bottom. Exams will focus on the most recent material but are expected to be cumulative in scope.

Technical errors during exams

If during an online quiz or exam there is a technical error which affects your ability to complete the assignment, you are immediately to let the quiz or exam proctor know and the instructor will discuss ways to allow you to resume the test without giving an unfair advantage. In the event of any unexpected server outage or any unusual technical difficulty which prevents students from completing a time sensitive assessment activity, the instructor will extend the time windows and provide an appropriate accommodation based on the situation. Students should immediately report any problems to the instructor.

Late Policy
When assignments and project work are turned in after the due date, this places an undue burden on the instructor and TA, especially when this policy is abused. As a compromise, if the assignment or project work is turned in prior to grading there will be no reduction in points, however, grading can occur any time after the due date (including the following morning!). A request for missing submissions may be sent at the discretion of the TA or instructor, but is not guaranteed, and a reduction of points may also occur in a way that is consistent for the rest of the class.

**Attendance Policy**

You are expected to attend class meetings regularly and to abide by the attendance policy established for the course. It is important that you communicate with the professor and the instructional team prior to being absent, so you, the professor, and the instructional team can discuss and mitigate the impact of the absence on your attainment of course learning goals. Please inform the professor and instructional team if you are unable to attend class meetings because you are ill, in mindfulness of the health and safety of everyone in our community.

**COVID-19**

If you are experiencing any symptoms of COVID-19 ([Links to an external site.](https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html)) please seek medical attention from the Student Health and Wellness Center (940-565-2333 or askSHWC@unt.edu) or your health care provider PRIOR to coming to campus. UNT also requires you to contact the UNT COVID Team at COVID@unt.edu for guidance on actions to take due to symptoms, pending or positive test results, or potential exposure.

**Syllabus Change Policy**

Any substantial changes to the syllabus after the first week will be notified suitably. Approximate point values for assignments or exams are expected to vary but will be fixed when the assignment or exam is given.

**UNT Policies**

**Academic Integrity Policy**

Academic Integrity Standards and Consequences. According to UNT Policy 06.003, Student Academic Integrity, academic dishonesty occurs when students engage in behaviors including, but not limited to cheating, fabrication, facilitating academic
dishonesty, forgery, plagiarism, and sabotage. A finding of academic dishonesty may result in a range of academic penalties or sanctions ranging from admonition to expulsion from the University.

**ADA Policy**

UNT makes reasonable academic accommodations for students with disabilities. Students seeking accommodations must first register with the Office of Disability Accommodation (ODA) to verify their eligibility. If a disability is verified, the ODA will provide a student with an accommodation letter to be delivered to faculty to begin a private discussion regarding one’s specific course needs. Students may request accommodations at any time, however, ODA notices of 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 accommodation for every semester and must meet with each faculty member prior to implementation in each class. For additional information see the [ODA website](https://disability.unt.edu/).

**Emergency Notification & Procedures**

UNT uses a system called Eagle Alert to quickly notify students with critical information in the event of an emergency (i.e., severe weather, campus closing, and health and public safety emergencies like chemical spills, fires, or violence). In the event of a university closure, please refer to Blackboard for contingency plans for covering course materials.

**Retention of Student Records**

Student records pertaining to this course are maintained in a secure location by the instructor of record. All records such as exams, answer sheets (with keys), and written papers submitted during the duration of the course are kept for at least one calendar year after course completion. Course work completed via the Blackboard online system, including grading information and comments, is also stored in a safe electronic environment for one year. Students have the right to view their individual record; however, information about student’s records will not be divulged to other individuals without proper written consent. Students are encouraged to review the Public Information Policy and the Family Educational Rights and Privacy Act (FERPA) laws and the University’s policy. See UNT Policy 10.10, Records Management and Retention for additional information.

**Acceptable Student Behavior**
Student behavior that interferes with an instructor’s ability to conduct a class or other students’ opportunity to learn is unacceptable and disruptive and will not be tolerated in any instructional forum at UNT. Students engaging in unacceptable behavior will be directed to leave the classroom and the instructor may refer the student to the Dean of Students to consider whether the student's conduct violated the Code of Student Conduct. The University's expectations for student conduct apply to all instructional forums, including University and electronic classroom, labs, discussion groups, field trips, etc. Visit UNT’s [Code of Student Conduct](https://deanofstudents.unt.edu/conduct) to learn more.

**Access to Information - Eagle Connect**

Students’ access point for business and academic services at UNT is located at: [my.unt.edu](http://my.unt.edu). All official communication from the University will be delivered to a student’s Eagle Connect account. For more information, please visit the website that explains Eagle Connect and how to forward e-mail [Eagle Connect](https://it.unt.edu/eagleconnect).

**Student Evaluation Administration Dates**

Student feedback is important and an essential part of participation in this course. The student evaluation of instruction is a requirement for all organized classes at UNT. The survey will be made available during weeks 13, 14 and 15 of the long semesters to provide students with an opportunity to evaluate how this course is taught. Students will receive an email from "UNT SPOT Course Evaluations via IASystem Notification" (no-reply@iasystem.org) with the survey link. Students should look for the email in their UNT email inbox. Simply click on the link and complete the survey. Once students complete the survey, they will receive a confirmation email that the survey has been submitted. For additional information, please visit the [SPOT website](http://spot.unt.edu/) or email spot@unt.edu.

**Getting Help**

**Technical Assistance**

**UIT Help Desk** (http://www.unt.edu/helpdesk/index.htm)

**Email:** helpdesk@unt.edu

**Phone:** 940-565-2324

**In Person:** Sage Hall, Room 330
Walk-In Availability: 8am-9pm

Telephone Availability:
- Sunday: noon-midnight
- Monday-Thursday: 8am-midnight
- Friday: 8am-8pm
- Saturday: 9am-5pm

Laptop Checkout: 8am-7pm

Student Support Services
- Registrar (https://registrar.unt.edu/registration)
- Financial Aid (Links to an external site.) (https://financialaid.unt.edu/)
- Student Legal Services (Links to an external site.) (https://studentaffairs.unt.edu/student-legal-services)
- Career Center (Links to an external site.) (https://studentaffairs.unt.edu/career-center)
- Multicultural Center (Links to an external site.) (https://edo.unt.edu/multicultural-center)
- Counseling and Testing Services (Links to an external site.) (https://studentaffairs.unt.edu/counseling-and-testing-services)
- Student Affairs Care Team (Links to an external site.) (https://studentaffairs.unt.edu/care)
- Student Health and Wellness Center (Links to an external site.) (https://studentaffairs.unt.edu/student-health-and-wellness-center)
- Pride Alliance (Links to an external site.) (https://edo.unt.edu/pridealliance)

Academic Support Services
- Academic Resource Center (Links to an external site.) (https://clear.unt.edu/canvas/student-resources)
- Academic Success Center (Links to an external site.) (https://success.unt.edu/asc)
- UNT Libraries (Links to an external site.) (https://library.unt.edu/)
- Writing Lab (Links to an external site.) (http://writingcenter.unt.edu/ (Links to an external site.))
- MathLab (Links to an external site.) (https://math.unt.edu/mathlab)