Introduction to Big Data and Data Science

Syllabus
CSCE 5300: Introduction to Big Data and Data Science (grad)
Spring 2022 at University of North Texas

Instructor
Name: Ravi Vadapalli, Ph.D.
Office Location: Discovery Park E180Q
Office Hours: by appointment via Zoom
Email: Ravi.Vadapalli@unt.edu

TA/Graders
Section 4 (Saturday, 11:30am)
Divya Birudharaju, DivyaGeethanjali.Birudharaju@unt.edu
Brahmeswararao Donepudi, BrahmeswararaoDonepudi@my.unt.edu
Vijya Maheesh Reddy Duggimpudi, VijayaManeeshareddy.Duggimpudi@unt.edu

Section 5 (Wednesday, 5:30pm)
Vineel Krishnamsetty, VineelKrishnamsetty@my.unt.edu
Sharat Chandra Reddi, Reddi, SharathChandra.Reddi@unt.edu
Mani Renukuntala, ManiRenukuntla@my.unt.edu

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.

Class Meetings:
Section 004, Sat 11:30am - 2:20pm at BLB 080
Section 005, Wed 5:30pm - 8:20p at DP 150K
BLB – Business Leadership Building, 1307 W Highland St., Denton, TX 76201
DP - Discovery Park, 3940 N. Elm Street, Denton, 76207

Credit hours: 3
Tentative topics

- Descriptive Statistics
- Data Visualization
- Data Structures and Data Frames
- Image processing
- Machine Learning Concepts
- Deep Learning Concepts
- Clustering and Regression
- Big Data / parallel processing
- Hadoop, Spark, and Parallel Computing
- Google Cloud Project

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 big data analytics.

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/
- The scikit-learn documentation: https://scikit-learn.org/stable/documentation.html
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 (LAPTOP is Mandatory for in-class assignments, NO EXCEPTIONS)

- Computers are required for exams
  - Canvas Technical Requirements (https://clear.unt.edu/supported-technologies/canvas/requirements)
  - 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.
  - Quizzes and exams will use the Canvas quiz system.
- Computers are mandatory during class.
  - There will be occasional in-class activities along with lectures and working on your own laptop may help in understanding and applying the material covered in the class.
- Students will be expected to Download and install Anaconda Python version 3 and be able to open a Jupyter notebook.

Google Cloud Project (GCP). At a suitable time during the course, you will be introduced to Google Cloud platform for course related work and assignments.

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 on date specified in the assignment. Please make sure the results are presentable, with appropriate comments for someone to follow what you have done.

Submission Requirements: Submit both PDF and .ipynb files (for python programs) for full credit. Late submissions incur penalties - see Late Policy below for details.

Concept Paper: (EXTRA CREDIT). There are far more tools and techniques than we can possibly cover, many of which may be particularly relevant to your interests. Also, your ability to distill complex topics into a form useful for your audience (or customers) is a critical skill to
Students have the option to write concept papers (format and expectations will be discussed at appropriate time during the course). This extra credit activity requires you to provide evidence of additional work done beyond the required course objectives. Submit up to two-page paper of your concept. Schedule a meeting with the instructional team for project selection.

**Quizzes/Exams:** There will be up to six quizzes and two exams. 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 must be taken in the classroom. Time will be limited, and all work must be individual. Quizzes/Exams are cumulative, but the emphasis will be on the newer material. Quiz/Exam days will be fixed and notified by Feb 3, 2022.

**Programming Assignments:** 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.

**Missed quizzes/exams:** Quizzes/Exams cannot be missed without prior arrangements or later documented proof of extenuating circumstances such as a health or unforeseen emergencies.

**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 (e.g., those doing extra credit work). The expected distribution of points is given below, with the exact scale determined by point values given for each assignment, project, or quiz/exam - 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. Concept Paper is for EXTRA CREDIT.

- Assignments: 44 pts (4 pts each)
- Quizzes: 30 pts (5 pts each)
- Exams: 22 pts (11 points per Exam)
- Attendance: 4 pts
- Concept Paper: 5 pts (EXTRA CREDIT)
- Additional Points: up to 10 (for spot quiz, class participation, etc.)

**Grading Scale:** A=90, B=80-89.9, C=70-79.9, D=60-69.9, F=0-59.9 pts. No exceptions.

**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 other than the instructor and the TAs 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 explore appropriate accommodations 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 TAs, 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.

- Assignments submitted up to 24 hours late: 10% penalty
- Assignments submitted late by 24 hours but less than 48 hours: 25% penalty
- Assignments submitted late by 48 hours but less than 72 hours: 50% penalty
- Assignments submitted late by 72 hours or more: 100% penalty (ZERO grade for that assignment)
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 prior to being absent, so you and the professor can discuss and mitigate the impact of the absence on your attainment of course learning goals. Please inform the professor 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 (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. 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).