INFO 4501: PRINCIPLES OF DATA SCIENCE AND ANALYSIS

Fall 2019

Instructor:	Lian MO, Ph.D.	**Lecture Time:**	W: 7:00PM - 9:30PM
Email:	lian.mo@unt.edu	**Place:**	zoom.us
Office:	zoom.us	**Phone:**	(786)368-6789

Course Pages

- Course management: Canvas - https://unt.instructure.com/courses/17716

- Lecture delivering: zoom.us - https://unt.zoom.us/meeting/512583583

Office Hours

- Th: 9:00PM - 9:30PM. https://unt.zoom.us/j/604647444

Textbooks

1. *Introduction to Computation and Programming Using Python: With Application to Understanding Data*, 2nd ed. by John V. Guttag, The MIT Press, 2016. ISBN: 9780262529624. (**required**)

2. *Principle of Data Science: Learn the techniques and math you need to start making sense of your data*, by Sinan Ozdemir, Packt Publishing, 2016. ISBN: 9781785887918. (**optional**)

3. *Principles and Techniques of Data Science*, by Sam Lau, Deb Nolan, and Joey Gonzalez, Online book: https://www.textbook.ds100.org/ch/01/lifecycle_intro.html. (**optional**)

Software
The following software will be used in this course.

- Python

Objectives

Combining data, computation, and inferential thinking, **data science is** redefining how people and organizations solve challenging problems and understand their world. Through a strong emphasizes on data centric computing, quantitative critical thinking, and exploratory data analysis, this course introduces the fundamentals of data science and data analytics including question formulation, data collection and cleaning, visualization, statistical inference, predictive modeling, and decision making. It provides the required foundational knowledge and practice to students to successfully integrate automatics methods and tools for qualitative and quantitative analysis. The topics include computational programming using Python; processing and analyzing data; algorithms for machine learning methods including regression, classification and clustering; statistical concepts of measurement error and prediction.

The objectives of this course:

1. Students will be able to understand the fundamental principles, techniques, algorithms, and tools for data analytics.

2. Students will be able to select and use techniques, algorithms and tools for processing and analyzing real-world data to support decision making and critical thinking.

3. Students will be able to apply computational and inferential thinking to program data analytics applications using Python.

Topics

The following is the tentative schedule with the covered topics. Actual schedule may be adjusted according to progress:

1. Introduction to Data Science

2. Python Preliminaries

3. Algorithms and Data Structures in Python I

4. Algorithms and Data Structures in Python II

5. Object Oriented Programming, Testing, Debugging, Exceptions, Assertions

6. Python Classes and Inheritance

7. Understanding Program Efficiency

8. Stochastic Thinking, Random Walk

9. Monte Carlo Simulation, Confidence Intervals

10. Sampling and Standard Error

11. Understanding Experimental Data

12. Machine Learning

13. Clustering, Classification

14. Classification, Statistical Sins

Technical Assistance

UIT Help Desk: http://www.unt.edu/helpdesk/index.htm. The University of North Texas provides student technical support in the use of Zoom and Canvas and supported resources. The student help desk may be reached at:

- Email: helpdesk@unt.edu

- Phone: 940.565-2324

- In Person: Sage Hall, Room 130

Hours:

- Monday-Thursday 8am-midnight

- Friday 8am-8pm

- Saturday 9am-5p

- Sunday 8am-midnight

Canvas technical requirements:

- https://clear.unt.edu/supported-technologies/canvas/requirements.

Minimal Technical Skills and Resources Needed

Each student needs an Internet connected computer that can run Zoom software. The computer should include a video camera, a microphone and a speaker for attending the online lectures and participating online discussions. Students are strongly encouraged to test their system for running Zoom and Canvas software at least one day before the first lecture.

Communication

Students can email their questions to the instructor and the teacher assistant (TA). They are also encouraged to talk to the instructor and TA during the office hours. Emails are normally respond within 24 hours, and all assignments,

Grading Policy

Grading will be based on project assignments and weekly assignments to be assigned as the course proceeds and a comprehensive final exam.. The exam is open-book, open-notes exam.

Grades will be computed as follows:

- Project Assignments: 40

- Weekly Assignments: 50

- Final Exam: 10

- Grading Scale: A: 90-100; B: 80-89; C: 70-79; D: 60-69; F: 59 or below.

Incompletes

A grade of incomplete (I) will be given only for a justifiable reason (such as a serious illness or military service) and only if you are passing the course. It is your responsibility to contact the instructor to request an incomplete and discuss requirements for completing the course. If you do not remove the incomplete within the timeframe agreed upon with the instructor or within one calendar year, you will receive a grade of an F. Please refer to http://essc.unt.edu/registrar/academic-record-incomplete.html for more information.

Withdrawal

A grade of withdraw (W) or withdraw-failing (WF) will be given depending on your participation and grades to date. If you simply disappear and do not file a formal UNT withdrawal form, you may receive a grade of an F.

ADA Statement

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 Accommodation (ODA) to verify their eligibility. If a disability is verified, the ODA will provide you with a reasonable accommodation letter to be delivered to 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 see the Office of Disability Accommodation website at http://www.unt.edu/oda. You may also contact them by phone at 940.565.4323.

Class Policy

- **Attendance Policy:** You are expected to attend class via GotoMeeting. You are responsible for announcements and assignments given in class. If you miss a class, it is up to you to obtain notes and any other information that was provided in the class. Those who do not attend class or review the recorded lectures in a timely manner can count on doing poorly in this course.

- **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 accommodation for students with disabilities. Students seeking accommodation 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 ones 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 at 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.

- **APA Style:** When doing the assignment, it is important to provide details of all the sources of information that you have used to prepare your work. All written assignments should follow APA (American Psychological Association) style to ensure that all sources are cited completely, correctly, and with consistency. The purpose of APA style is to (a). give credit to the author whose ideas or research you have used, (b). provide the exact location for sources of information used in the text of your paper, and (c). enable your reader to verify information you have provided or to explore your topic in greater depth. Consult the Publication Manual of the American Psychological Association, 6th edition.

- **Academic Dishonesty:** Students caught cheating or plagiarizing will receive a "0" for that particular assignment or exam (or specify alternative sanction, such as course failure). Additionally, the incident will be reported to the Office of Student Rights and Responsibilities, which may impose for further penalty. According to the UNT catalog, the term "cheating" includes, but is not limited to: (a). use of any unauthorized assistance in taking quizzes, tests, or examinations; (b). dependence upon the aid of sources beyond those authorized by the instructor in writing papers, preparing reports, solving problems, or carrying out other assignments; (c). the acquisition, without permission, of tests or other academic material belonging to a faculty or staff member of the university; (d). dual submission of a paper or project, or resubmission of a paper or project to a different class without express permission from the instructor(s); or (e). any other act designed to give a student an unfair advantage. The term "plagiarism" includes, but is not limited to: (a). the knowing or negligent use by paraphrase or direct quotation of the published or unpublished work of another person without full and clear acknowledgment; and (b). the knowing or negligent unacknowledged use of materials prepared by another person or agency engaged in the selling of term papers or other academic materials.

- **Acceptable Student Behavior:** Student behavior that interferes with an instructors 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 Center for Student Rights and Responsibilities 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. The Code of Student Conduct can be found at `www.unt.edu/csrr`.

- **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 students Eagle Connect account. For more information, please visit the website that explains Eagle Connect and how to forward e-mail: `eagleconnect.unt.edu/`

- **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 [insert administration dates] 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 at `http://spot.unt.edu/` or email: spot@unt.edu.

- **Sexual Assault Prevention:** UNT is committed to providing a safe learning environment free of all forms of sexual misconduct, including sexual harassment sexual assault, domestic violence, dating violence, and stalking. Federal laws (Title IX and the Violence Against Women Act) and UNT policies prohibit discrimination on the basis of sex, and therefore prohibit sexual misconduct. If you or someone you know is experiencing sexual harassment, relationship violence, stalking, and/or sexual assault, there are campus resources available to provide support and assistance. UNTs Survivor Advocates can assist a student who has been impacted by violence by filing protective orders, completing crime victims compensation applications, contacting professors for absences related to an assault, working with housing to facilitate a room change where appropriate, and connecting students to other resources available both on and off campus. The Survivor Advocates can be reached at SurvivorAdvocate@unt.edu or by calling the Dean of Students Office at 940-565- 2648. Additionally, alleged sexual misconduct can be non-confidentially reported to the Title IX Coordinator at oeo@unt.edu or at (940)565 2759.

Important Notice for F-1 Students taking Distance Education Courses

- **Federal Regulation:** To read detailed Immigration and Customs Enforcement regulations for F-1 students taking online courses, please go to the Electronic Code of Federal Regulations website at `http://www.ecfr.gov/`. The specific portion concerning distance education courses is located at Title 8 CFR 214.2 Paragraph (f)(6)(i)(G).

  The paragraph reads:

  (G) For F-1 students enrolled in classes for credit or classroom hours, no more than the equivalent of one class or three credits per session, term, semester, trimester, or quarter may be counted toward the full course of study requirement if the class is taken on-line or through distance education and does not require the students physical attendance for classes, examination or other purposes integral to completion of the class. An online or distance education course is a course that is offered principally through the use of television, audio, or computer transmission including open broadcast, closed circuit, cable, microwave, or satellite, audio conferencing, or computer conferencing. If the F-1 students course

of study is in a language study program, no on-line or distance education classes may be considered to count toward a students full course of study requirement.

- **University of North Texas Compliance:** To comply with immigration regulations, an F-1 visa holder within the United States may need to engage in an on-campus experiential component for this course. This component (which must be approved in advance by the instructor) can include activities such as taking an on-campus exam, participating in an on-campus lecture or lab activity, or other on-campus experience integral to the completion of this course. If such an on-campus activity is required, it is the students responsibility to do the following:

  1. Submit a written request to the instructor for an on-campus experiential component within one week of the start of the course.
  2. Ensure that the activity on campus takes place and the instructor documents it in writing with a notice sent to the International Student and Scholar Services Office. ISSS has a form available that you may use for this purpose.

  Because the decision may have serious immigration consequences, if an F-1 student is unsure about his or her need to participate in an on-campus experiential component for this course, s/he should contact the UNT International Student and Scholar Services Office (telephone 940-565-2195 or email: internationaladvising@unt.edu) to get clarification before the one-week deadline.

- **Student Verification:** UNT takes measures to protect the integrity of educational credentials awarded to students enrolled in distance education courses by verifying student identity, protecting student privacy, and notifying students of any special meeting times/locations or additional charges associated with student identity verification in distance education courses. See UNT Policy 07-002 Student Identity Verification, Privacy, and Notification and Distance Education Courses at https://policy.unt.edu/policy/07-002.

- **Use of Student Work:** A student owns the copyright for all work (e.g. software, photographs, reports, presentations, and email postings) he or she creates within a class and the University is not entitled to use any student work without the students permission unless all of the following criteria are met:

  - The work is used only once.
  - The work is not used in its entirety.
  - Use of the work does not affect any potential profits from the work.
  - The student is not identified.
  - The work is identified as student work.

  If the use of the work does not meet all of the above criteria, then the University office or department using the work must obtain the students written permission.

Advice for Doing Programming Assignments:

- **Start early.** You will hit snags, and you can count on a programming assignment taking longer than you think it should.

- **Finish early.** Do not wait until the deadline to finish up.

- **Test thoroughly.** The most common mistake among inexperienced programmers is to assume that their programs work after very limited testing. Try different kinds of inputs. Try unusual inputs. Be sure that you have exercised every part of the program. I recommend that you not do tests by hand. Instead, create a script that does a collection of tests automatically. That way, after each change, you

can run the script and see that everything is still working, or at least find out what is not working. You can even create the test script before you write the program. Remember that fixing one thing sometimes ruins something else that used to work, so you want to rerun tests that worked before.

- **Do not give up.** You are going to encounter problems. Count on it. You are not the only one who has difficulties writing computer programs. Do not throw in the towel at the first snag. If something is wrong, first determine what it is, then

  x it. If you cannot see how to diagnose or

  x the problem, ask for help.

- **Do what the assignment asks for.** Do not make up your own assignment. If you decide to make temporary modifications to aid in development, be sure that you go back and repair them before turning in the assignment. Reread the assignment before turning in your work. Pay attention to details.