INFO5802: Seminar in Research and Research Methodology

Spring 2021

Instructor:	Junhua Ding, Ph.D.	**Lecture Time:**	Monday: 2:30pm - 5:20pm
Email:	junhua.ding@unt.edu	**Place:**	Zoom Meeting/NTDP K150
Office:	Discovery Park: Room E292G	**Phone:**	940-565-2186

Course Information

- INFO5082 (Sections 202): Seminar in Research and Research Methodology, 3 Credit Hours
- Course management at Canvas: https://unt.instructure.com/
- No pre-requisite.

Office Hours

- By appointment

Textbooks

1. *Quantitative and Statistical Research Methods: From Hypothesis to Results*, by William E. Martin, and Krista D. Bridgmon, Jossey-Bass, 2012. ISBN: 978-0-470-63182-9. (**required**)

Software

The following software might be used in this course.

- IBM SPSS or GNU PSPP.
- Tableau, Power BI.
- Microsoft Excel, MySQL.
- R/RStudio.
- Python/Python Jupyter Notebook/Lab, and Python Frameworks.
- Tensorflow, PyTorch

Important Note

Due to the continuing impact of COVID-19, it is important that you read all Announcements posted on Canvas and UNT Website. While attendance is expected as outlined in this syllabus, it is important for all of us to be mindful of the health and safety of everyone in our community, especially given concerns about **COVID-19**. Please contact me if you are unable to attend class because you are ill, or unable to attend class due to a related issue regarding **COVID-19**. It is important that you communicate with me prior to being absent so I may make a decision about accommodating your request to be excused from class. 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 (phone number: 940-565-2333 or email: askSHWC@unt.edu) or your health care provider PRIOR to

coming to campus. UNT also requires you to contact the UNT COVID Hotline (phone number: 844-366-5892 or email: COVID@unt.edu) for guidance on actions to take due to symptoms, pending or positive test results, or potential exposure. While attendance is an important part of succeeding in this class, your own health, and those of others in the community, is more important. Face coverings are required in all UNT facilities. Students are expected to wear face coverings during this class. If you are unable to wear a face covering due to a disability, please contact the Office of Disability Access to request an accommodation. UNT face covering requirements are subject to change due to community health guidelines. Any changes will be communicated via the instructor.

Objectives

This course is designed for students to enhance their technical proficiency and research experience in data science. It will train each student to become a qualified data scientist for solving real-world problems. Students will learn quantitative and statistical research methods and apply them into directed research study in selected data science topics.

The objectives of this course:

1. Students will be able to conduct data science research using quantitative and statistical research methods.
2. Students will be able to develop an area of research interest in data science, choose a specific topic, and formulate research questions from which research proposals are developed.
3. Students will be able to conduct data science research with data from a variety of disciplines. The research will cover the entire life cycle of data including data specification and acquisition, exploratory data analysis, hypothesis development, data modeling and management, data analysis model development and validation, feature learning and value extraction, data visualization and reporting, and other special topics such as data security and quality.
4. Students will be able to use data analysis results to support decision-making and critical thinking.

Topics

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

1. Introduction to Research Methods
2. Overview of Quantitative Research
3. Research and Statistical Designs
4. Diagnosing Study Data
5. One-Way Analysis of Variance (ANOVA)
6. Repeated-Measures Analysis of Variance
7. Factorial Analysis of Variance
8. Analysis of Covariance
9. Nonparametics
10. Data Science Research
    (a) Big Data Analysis
    (b) Machine Learning

11. Data Science Project Development
    (a) Requirements analysis and Data Specification;
    (b) Data Collection, Cleaning, and Management;
    (c) Exploratory Data Analysis and Hypothesis Development;
    (d) Data Modeling and Transformation;
    (e) Data Analysis, Mining, and Learning;
    (f) Data Visualization; Verification and Validation;
    (g) Result Interpretation and Decision Making;
    (h) Product Deployment and Maintenance;
    (i) Project Management;
    (j) Security and Privacy;
    (k) Ethics.

Teaching Philosophy

This course will not be taught with traditional lectures. Students will do directed research in data science. After we complete the introduction of research methods and quantitative research methods in the three weeks. Students will conduct quantitative research in several small projects, which cover some special topics. Then each student will work on a selected data science project. Students will report their progresses and results in the class, and the instructor will closely work with students through the entire project development process to ensure students appropriately use quantitative and statistical research methods in their research. This class is reading, writing, and on-class discussion intensive.

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, quizzes, projects, and papers should be graded within 10 days after the submission deadline.

Grading Policy

Students are required to attend the class on time, complete all assignments, readings and projects on time, and participate class discussions. Grading will be based on assignments to be assigned as the course proceeds. Grades will be computed as follows:

- Reading Assignments and Presentations: 40%
    1. Presentations and demonstration of SPSS projects from the textbook: 20%
    2. Reading assignment report and presentation: 20 %
- Research Project: 50%
    1. Research Project Proposal: 5%
    2. Data Collection and Management: 5%
    3. Exploratory Data Analysis: 10%
    4. Hypothesis Development: 5%
    5. Data Modeling, Transformation, and Description: 5%
    6. Data Analysis Model Development and Validation: 10%
    7. Feature Representation and Learning: 5%
    8. Value Extraction: 5%
    9. Data Visualization: 5%
    10. Decision Making and Result Interpretation: 5%
    11. Report and Presentation: 10%
- Class Participation and Discussion: 10%

  2 points will be reduced for late, early leave or absent of one class until all 10 points will be deducted from the final points.
- **Grading Scale**: A: 90-100; B: 80-89; C: 70-79; D: 60-69; F: 59 or below.

All reading assignments and project requirements are post in Canvas, and students are required to uploaded their work into Canvas. Late submission of work will be graded with reduced points. The final grade is calculated based on grade points of assignments and presentations, project, class participation and discussion. NO final exam is given this time.

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

- **Prohibition of Discrimination, Harassment, and Retaliation:** As members of the UNT community, we have all made a commitment to be part of an institution that respects and values the identities of the students and employees with whom we interact. UNT does not tolerate identity-based discrimination, harassment, and retaliation. According UNT Policy 16.004, UNT prohibits discrimination and harassment because of race, color, national origin, religion, sex, sexual orientation, gender identity, gender expression, age, disability, genetic information, veteran status, or any other characteristic protected under applicable federal or state law in its application and admission processes; educational programs and activities; employment policies, procedures, and processes; and university facilities. The University takes active measures to prevent such conduct and investigates and takes remedial action when appropriate.

- **Attendance Policy:** You are expected to attend class via Zoom. 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.

  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.

- **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 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 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.

- **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 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 deanofstudents.unt.edu/conduct.

- **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: 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. UNT's Survivor Advocates can assist a student who has been impacted by violence by filing protective orders, completing crime victim's 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.

1 Appendix: Reading List

1. Alex Krizhevsky, Ilya Sutskever, & Geoffrey E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*,https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

2. Lingfei Wu, Dashun Wang & James A. Evan, *Large teams develop and small teams disrupt science and technology*, https://www.nature.com/articles/s41586-019-0941-9.

3. Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, & Dan Steinberg, *Top 10 algorithms in data mining*, http://39.104.72.142:802/algorithms/10Algorithms-08.pdf.

4. Malvina Nissim, Rik van Noord, & Rob van der Goot*Fair is Better than Sensational: Man is to Doctor as Woman is to Doctor*, https://arxiv.org/abs/1905.09866.

5. Sergio Oramas, Vito Claudio Ostuni, Tommaso Di Noia, Xavier Serra, & Eugenio Di Sciascio, *Sound and Music Recommendation with Knowledge Graphs*, https://dl.acm.org/doi/10.1145/2926718.

6. Hovy and Spruitt, *The Social Impact of Natural Language Processing*, https://aclweb.org/anthology/P16-2096.

7. Lei Zhang, Shuai Wang, & Bing Liu, *Deep Learning for Sentiment Analysis : A Survey*, https://arxiv.org/abs/1801.07883.

8. Haohan Wang, Bhiksha Raj, *On the Origin of Deep Learning*, https://arxiv.org/abs/1702.07800.

9. Sepp Hochreiter, Juergen Schmidhuber, *Long Short-Term Memory* https://www.bioinf.jku.at/publications/older/2604.pdf

10. Savolainen, R. (2018). *Pioneering models for information interaction in the context of information seeking and retrieval.* Journal of Documentation, 74(5), 966–986. https://libproxy.library.unt.edu:2147/10.1108/JD-11-2017-0154

11. Kelly, D., and Sugimoto, C. R. (2013). *A systematic review of interactive information retrieval evaluation studies, 1967–2006*, Journal of the Association for Information Science and Technology, 64 (4), (745-770). https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.22799.

12. Fokkens et al. (2013), *Offspring from Reproduction Problems: What Replication Failure Teaches Us:* https://www.aclweb.org/anthology/P13-1166.

13. Azzopardi, L. (2011). *The economics in interactive information retrieval.* SIGIR 2011, July 24-28, 2011, Beijing, China. https://doi.org/10.1145/2009916.2009923

14. Edwards, A., and Kelly, D. (2017). *Engaged or frustrated? Disambiguating emotional state in search.* SIGIR 2017, August 07-11, 2017, Tokyo, Japan. https://doi.org/10.1145/3077136.3080818

15. Lewis et al. (2011), *CrisisMT: Developing A Cookbook for MT in Crisis Situations.* https://www.aclweb.org/anthology/W11-2164 (should be paired with Bird 2009, Natural Language Processing and Linguistic Fieldwork: https://www.mitpressjournals.org/doi/pdf/10.1162/coli.35.3.469)

16. Devlin et al. (2019), *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.* https://arxiv.org/pdf/1810.04805.pdf (should be paired with Mikolov et al. 2013, *Efficient Estimation of Word Representations in Vector Space.* https://arxiv.org/pdf/1301.3781.pdf) doi:10.1080/07421222.1996.11518099,

17. Richard Y. Wang, Diane M. Strong, *Beyond Accuracy: What Data Quality Means to Data Consumers.* Journal of Management Information Systems, Vol.12(4), page 5-33, 1996. doi:10.1080/07421222.1996.11518099, https://doi.org/10.1080/07421222.1996.11518099.

18. Q. Li, H. Peng, J. Li, C. Xia, R. Yang, L. Sun, P. Yu, L. He, *A Survey on Text Classification: From Shallow to Deep Learning*, arXiv:arXiv:2008.00364, 2020.

19. M. Oetzel, S. Spiekermann,*A systematic methodology for privacy impact assessments: A design science approach.* European Journal of Information Systems. 23. 10.1057/ejis.2013.18. 2014.

20. C. Tan, et al. *A Survey on Deep Transfer Learning.* arXiv: abs/1808.01974, 2018.

21. J.L. Ogburn, *The Imperative for Data Curation.* Libraries and the Academy, vol.10(2), pages 241-246. doi:10.1353/pla.0.0100. 2010.

22. Mei Kobayashi, Koichi Takeda, *Information retrieval on the web*, ACM Computing Surveys, https://doi.org/10.1145/358923.358934, 2000.