DTSC 3010.020 - Introduction to Data Science, Spring 2026

Instructor: Pretom Roy Ovi, PhD
Email: pretomroy.ovi@unt.edu
Office hours: By Appointment

TA: Mayeesha, Tasmiah Tahsin
Email: TasmiahTahsinMayeesha@my.unt.edu

Class Schedule: Thursday 5:30pm -8:20 pm
Class room: NTDP B190

Welcome to UNT!

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. UNT's full Non-Discrimination Policy can be found in the UNT Policies section of the syllabus.

Course Description

The course is designed to introduce students to the concepts, theories, and techniques that are involved in different stages of the data lifecycle: data access, collection and management, data description and summarization, data modeling including traditional machine learning algorithms, and evaluation. The course aims to present a broad view of data science in industry and research, which prepares students with a solid base to develop specializations in data science. Students will acquire an integrated skill set that covers web and database technologies, programming, statistics, and machine learning, while gaining hands-on experience using Python for applied data analysis and data-driven research.

Course Prerequisites or Other Restrictions

MATH 1650, MATH 1680, CSCE 1030, CSCE 1040; or permission of instructor.

Course Objectives

After successfully completing this course, you will be able to:

1. Have a thorough understanding of stages in data lifecycle.
2. Build a mindset of managing and integrating different sources of information for decision-making through data.
3. Understand the generation of different types of data and how to obtain data from webpages and application databases.
4. Handle basic algorithms in machine learning for data modeling, including supervised methods and unsupervised clustering.
5. Foster critical thinking of ethical issues in data science projects.
6. Master the use of Python/R for data scrapping, data manipulation, exploratory data analysis and data modeling.
7. Learn how to document and report data science projects.

Materials

Textbooks:

1. Irizarry, Rafael A. (2019). Introduction to Data Science: Data Analysis and Prediction Algorithms with R. Available at: https://rafalab.github.io/dsbook/
2. Cielen, D., Meysman, A., & Ali, M. (2016). Introducing data science: big data, machine learning, and more, using Python tools. Manning Publications Co.. Available at: https://livebook.manning.com/book/introducing-data-science/about-this-book/
3. An introduction to statistical learning: Pdf link:https://drive.google.com/file/d/1_PqXnKi7pm19FDWmJG6VOYXkf_kXeBIe/view

Software

The following software and computing environments might be used in this course. You may use Google Colab, Anaconda for Python Jupyter Notebook, JupyterLab, Python Frameworks such as NumPy, SciPy, Pandas.

Required Technology:

Laptop - We will do live programming exercises during most classes, so bring your laptop and be prepared to write code. Any reasonably current operating system can be used. If you don't have access to a laptop, contact me before the first class.

Teaching Philosophy

The instructor will explain concepts supplemented with case studies to guide students' understanding of stages in the data lifecycle. The instructor takes heuristic methods to guide students' thinking about questions and strategies in working with data. The instructor will work together with students on in-class coding practices to understand how to use R for data manipulation, data scrapping, data visualization, and data modeling. The instructor will monitor the progress of students and is open to suggestions from students. Students are expected to participate actively in class, work together and help each other to achieve success in class. The students should submit their assignments on time to achieve satisfactory class performance. Interaction between the student and the instructor/TA is guaranteed and strongly encouraged. Students who don't have knowledge and experience in statistics or programming are expected to spend extra hours on this course.

Notes

Our time together in class is precious. To use it effectively, you must come to class on time and be prepared. Being prepared for class means that you have:

1. Completed the required readings.
2. Attempted the installation of required packages and read the documentations for hands-on practice in class.
3. Computers are used when needed to accomplish a class objective. No phones or tablet devices are permitted during our class meetings. Researchers have found that such devices present an irresistible distraction and detract from the cooperative learning environment.
4. We will use Canvas Discussions as an online forum that you can use to ask/answer questions, get clarifications, point out my mistakes, etc. Be sure to check it regularly.

Here is my suggested general strategy for working on assignments:

1. Start early – don't wait. That will give you time to work through the problems and get help as needed.
2. When you run into a problem, first trying to solve the problem by yourself. You can also search Q&A sites (e.g. StackOverflow) for solutions to technical questions, however search answers to assignment questions or sharing assignments online to get help is cheating.
3. You are encouraged to post the problem and how you learn to solve it on discussion board. You will get bonus points for actively participate in sharing information.
4. You are welcome to post your question in discussion board. I will be monitoring and will respond as soon as I am able, usually within a day (longer during weekends, travel, etc.).

ASSESSMENT & GRADING

A student's grade is composed of the following:

- In-class participation (10%)
- Assignments (35%)
- Term Project including project proposal, project update, final presentation and report submission (35%)
- Exam (20%)

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

COURSE CALENDAR

The contents of the course are organized into 16 weeks. Please refer to Table below for preliminary schedule. It is subject to change based on the student's outcome and will be updated according to progress.

Week	Topics
Week 1: 01/15/2026	Introduction and Course Outline
Week 2: 01/22/2026	Data Sources and Data Types
Week 3: 01/29/2026	Get Familiar with Python
Week 4: 02/05/2026	Exploratory Data Analysis
Week 5: 02/12/2026	Working with Text Data
Week 6: 02/19/2026	Supervised Method: Decision Tree
Week 7: 02/26/2026	Supervised Method: Linear Regression & Naive Bayes Classification
Week 8: 03/05/2026	**Project Proposal**
Week 9: 03/12/2026	Spring Break

Week 10: 03/19/2026	Classification Method: Logistic Regression & Random Forest
Week 11: 03/26/2026	Unsupervised Method: K-means Clustering, Association, Anomaly Detection
Week 12: 04/02/2026	**Project Update**
Week 13: 04/09/2026	Unsupervised Method: Hierarchical Clustering, Dimensionality Detection
Week 14: 04/16/2026	Data Ethics and Review
Week 15: 04/23/2026	Project Review (Team meeting)
Week 16: 04/30/2026	**Final Project Presentation & Report Submission**
Week 17: 05/07/2026	**Final Exam**