CS 301: Introduction to Data Science

1 Course Information

Instructor: Lingxiao Wang
Department of Data Science
Email: lw324@njit.edu

TA: TBD

Times and Locations: Tuesday 8:30 - 11:20 am, KUPF 117

Office Hourrs: TBD

2 Course Description

This course provides a comprehensive introduction to the field of data science, equipping students with the foundational knowledge, methodologies, and practical skills needed to analyze and interpret real-world datasets. Students will learn how to collect, process, analyze, and visualize data using widely adopted tools in the data science ecosystem, while developing the ability to draw meaningful insights from diverse data sources. In addition to technical skills, the course examines the broader impact of data science on society and everyday life. Topics include ethical and legal considerations, societal implications, and the role of education in navigating a world shaped by data. Through case studies and hands-on activities, students will not only apply data science techniques but also critically evaluate their consequences.

Prerequisites: CS 114 and (MATH 333 or MATH 341) with a grade C or better.

Textbook
There is no required textbook.

Recommended Textbooks

- Data Mining: Practical Machine Learning Tools and Techniques. By Ian Witten, Eibe Frank, Mark Hall, 2017.

- Python for Data Analysis. By Wes McKinney, O'Reilly Media, 2018.
- An Introduction to Statistical Learning with Applications in Python. By Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer, 2023.
- Understanding Machine Learning: From Theory to Algorithms. By Shalev-Shwartz, Shai, and Shai Ben-David, Cambridge university press, May 19, 2014.

3 Brief List of Topics to be Covered:

- Data Preprocessing (e.g, data acquisition, data manipulation, dealing with data issues)
- Data Interpretation (e.g., statistics, visualization, exploratory data analysis)
- Data Mining Techniques
- Modern Topics: Artificial Intelligence, Large Language Model, Reinforcement Learning
- Social Impact of Data Science
- Potential Risks of Data Science

4 Grading

Grades will be computed based on the following factors

- Homework (20%)
- Project (25%)
- Midterm Exam (25%)
- Final Exam (30%)

We use the grade cutoff points defined in Table 1. The final grade might be curved depending on students' overall performance.

5 Homeworks

There will be about 5 homework assignments during the semester as we cover the corresponding material. Unless stated otherwise, you are encouraged to discuss homework problems with other students, but each student must submit their own answers and independently write their own code for the programming

Table 1: Grade cutoff points

Letter Grade	Point Range
A	[90,100)
B+	[85,90)
B	[78,85)
C+	[70,78)
C	[65,70)
D	[60,65)
F	[0,60)

portion. You must also indicate on each assignment with whom you collaborated and cite any additional sources used, including websites. Any student found cheating in this course will receive a failing grade.

Late Policy: Homework and projects are worth full credit if submitted by the due date. Late submissions will not be accepted.

6 Exam

There will be one midterm and one final exam. The exams are closed-book exams. There are no late submissions or makeup for exams.

Academic Integrity

"Academic Integrity is the cornerstone of higher education and is central to the ideals of this course and the university. Cheating is strictly prohibited and devalues the degree that you are working on. As a member of the NJIT community, it is your responsibility to protect your educational investment by knowing and following the academic code of integrity policy that is found at: NJIT Academic Integrity Code.

Please note that it is my professional obligation and responsibility to report any academic misconduct to the Dean of Students Office. Any student found in violation of the code by cheating, plagiarizing or using any online software inappropriately will result in disciplinary action. This may include a failing grade of F, and/or suspension or dismissal from the university. If you have any questions about the code of Academic Integrity, please contact the Dean of Students Office at dos@njit.edu"

Generative AI

In this class, the use of Generative AI tools like ChatGPT is generally not allowed unless specifically required by the instructor for a particular assignment.

The final responsibility for any work submitted lies with the student. If Generative AI tools produce incorrect results and these are submitted, the student will be graded accordingly. Mistakes made by the AI are the student's responsibility.