Spring 2024

CS 301-102: Introduction to Data Science

Kaustubh Vijaykumar Pethkar

CS 301 - Introduction to Data Science
Course Syllabus for Spring 2024

Instructor: Kaustubh Pethkar

Office Hours:

- Tuesday, Friday - 4:00 pm to 4:40 pm
- Wed: 5:30 pm to 6:00 pm
- Fri: 1:45 pm to 2:25 pm (Webex/zoom)

TA: Swastik Biswas (sb2785@njit.edu)
Office hours:

- Tuesdays - GITC 4324, 5:30 PM to 7:30 PM
- Fridays - GITC 4324, 12:00 PM to 1 PM

YWCC Tutoring: https://computing.njit.edu/undergraduate-tutoring-1

Course Overview
This is an introductory data science course that introduces students to what data science is and what data scientists do. The course will explore some real-world datasets and teach students how to engineer them for their appropriate use in data driven decision making as well as in machine learning. There are two parts in this course. The first part focuses on data analysis and visualization, and the second part focuses on the use of data in decision making and in machine learning.

Prerequisite: CS 114, Math 333

Topics to Be Covered:
The list of topics to be covered includes the following:

- Data Analysis and Preprocessing
- Data Visualization and Dimensionality Reduction
- Regression
- Classification
- Clustering
- Tree Based Algorithms
- Networks in Data Science

Course Outcome:
By the end of the course, students will be able to:

- Perform analysis on noisy and large datasets.
- Engineer the data for their best use.
- Build predictive models.
- Perform correlation and clustering analysis.
- Identify appropriate machine learning models for a particular problem.
- Explain the outcome of a model.

Course Resources

There is no specific textbook required for this course. However, the majority of the materials will come from the following books, and you are highly advised to read the relevant chapters of the books along with the lecture notes:

- Data Mining: Practical Machine Learning Tools and Techniques - by Ian Witten, Eibe Frank, Mark Hall
- Python for Data Analysis (2nd Edition) - by Wes McKinney
- Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications - by Laura Igual, Santi Seguí

Class Attendance

Class attendance is mandatory. Getting to class late or leaving early counts as absence. If you are unable to attend for some valid reasons, you must submit proper documentation to the dean of students. You will be excused once they deem the reason as valid. Students who miss more than 5 classes (unexcused) will have their grade reduced by a letter grade.

Assignments

Assignments must be submitted through Canvas by the due date. It will not be accepted late except for special circumstances (such as jury duty or medical problem), for which you have to provide documentation to the dean of student's office, and they must approve it.

Quizzes

Quizzes will be given to test fundamental concepts. There will be two types of quizzes: regular quizzes and pop quizzes. Pop quizzes will be given at the beginning of the class or at the end of the class. When given at the beginning, the focus will be on assessing your grasp of the previous lecture's content. Conversely, if a pop quiz takes place at the end of the class, it will test your understanding of the material covered during that day's lecture.

Project

There will be two projects in this course. Students will work individually in their first project and in groups in their second project. The project requirements will be discussed in class and will be posted on canvas.

Exams

There are two exams in this course: Midterm and Final. You must bring ID to all exams. Students with special needs are advised to make arrangements with the Office of Accessibility Resources and Services. There are no makeup exams. If you miss an exam because of a documented special circumstance, you may receive a grade based on the other exam or based on the average performance on the other parts of the course.

Class Participation

Asking and answering questions, taking quizzes, solving problems — individually or in groups — is a regular part of class meetings. Cell phones must be turned off during class. During class time you may not play games, text, email, browse the web or engage in other activities that are not part of the class. Any violation will be reported to the dean of student's office.

Course Communication

Canvas (canvas.njit.edu) will be used to post lecture notes and to submit homework. For any other communications, we will use **Discord**. You will receive the invitation to join the discord server as soon as the semester starts. It is highly recommended that you install discord on your phone for easy access.

Overall Course Score Formula

Attendance	5%

Assignment	15%
Quiz	15%
Project 1	5%
Project 2	10%
Midterm Exam	20%
Final Exam	30%

The letter grade is based on the overall course score.

Grade Formula						
Grade	**A**	**B+**	**B**	**C+**	**C**	**D**
Overall Course Score Cutoff	90	85	80	75	70	60

Grade Appeals

If you believe that you deserve more credit than you have been awarded on a particular problem, you may request, **at the time it is returned or within 48 hours of the grade being posted**, that it be re-graded. Your entire assignment will be re-graded, which may result in points being **added or subtracted.**

University Code on Academic Integrity

Read the University Code on Academic Integrity (njit.edu/policies/sites/policies/files/academic-integrity-code.pdf). It describes infractions of academic integrity and penalties for violations, including, for the most serious violations, an XF grade in the course or expulsion. **All work that you represent as your own must, in fact, be your own.** Work done by others must be given proper credit.

I reserve the right to make small changes to this syllabus; if there is any modification, you will be informed during the semester.