CS 301: Introduction to Data Science Syllabus

Instructor: Akm Islam (azi3@njit.edu) Office Hours: Tuesday and Friday (3:10pm - 4:50pm) Location: GITC 2125

TA: TBD Office Hours: TBD

Course Overview: This is an introductory data science course that introduces students to the field of data science and the role of data scientists. The course will explore real-world datasets and teach students how to prepare and engineer them for use in data-driven decision-making and machine learning. The course is divided into three parts: the first focuses on data analysis and visualization, while the second and third parts cover the application of data in classical machine learning and deep learning.

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

Brief list of topics to be covered:

- 1. Data Analysis and Preprocessing
- 2. Data Visualization and Exploratory Data Analysis
- 3. Regression Techniques
- 4. Classification Methods
- 5. Tree-Based Models and Ensemble Learning
- 6. Deep Learning and Large Language Models
- 7. Clustering Algorithms

Educational objectives for the course:

Upon completing the course, the student will be able to:

- 1. Perform data analysis and preprocessing on noisy datasets to derive actionable insights.
- 2. Apply data visualization techniques and dimensionality reduction methods to interpret and present data effectively.
- 3. Build and evaluate regression and classification models tailored to specific use cases.
- 4. Develop and optimize tree-based algorithms and ensemble learning techniques to enhance predictive performance.
- 5. Utilize deep learning models, including Large Language Models, to solve complex problems.
- 6. Interpret and explain the outcomes of machine learning models to support decision-making.
- 7. Implement clustering techniques to discover patterns and groupings in data.

Instructional Materials:

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, 4th edition Ian Witten, Eibe Frank, Mark Hall, ISBN-13: 9780128042915, Morgan Kaufmann, 2016.
- Python for Data Analysis, 2nd edition Wes McKinney, ISBN-13: 9781491957660, O'Reilly Media, 2017.
- Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications, 1st edition Laura Igual, Santi Seguí, ISBN-13: 9783319500164, Springer, 2017.

Class Attendance (5% of grade): Class attendance is mandatory. Attendance will be taken randomly, and the final score will be based on the recorded attendance. Methods of taking attendance may include participation in Canvas activities, electronically submitting in-class work, calling students by name, passing a signature list around, or any combination of these methods. This may vary from week to week. Arriving late to class or leaving early will count as an absence. If you are unable to attend for valid reasons, you must submit proper documentation to the dean of students. You will be excused once the reason is deemed valid. Students who miss more than three unexcused classes will have their grade reduced by one letter grade.

Assignments (15% of grade): This course includes six assignments, all of which must be submitted via Canvas by the designated due date. Late submissions will incur a penalty of 0.5% per hour beyond the deadline. Submissions will not be accepted more than 48 hours after the due date.

Note: The assignment with the lowest score will be dropped.

Quizzes (10% of grade): There will be weekly quizzes based on the materials covered during that week. These quizzes are designed to reinforce your understanding of the course content and help you stay engaged with the material. Quizzes will be administered using a lockdown browser to ensure academic integrity. You must clearly show your environment before taking each quiz and ensure that your desk is empty with no noise around you. Note: The two quizzes with the lowest scores will be dropped.

Project (15% of grade): This course includes a group project with three milestones to guide your progress. Detailed requirements will be posted on Canvas in advance. All group members are expected to contribute equally to each milestone. Meeting deadlines and maintaining effective collaboration will be key to your group's success.

Exams (55% of grade): There are two exams in this course: a midterm worth 25% and a partially cumulative final exam worth 30%. Both exams will be administered using a lockdown browser with a camera to ensure academic integrity. Please test the lockdown browser setup beforehand. You are also required to bring your ID to both exams.

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 or set to silent mode 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.

Course Communication: <u>Canvas</u> will be used for posting lecture notes and submitting homework. Lecture-related questions will be addressed during office hours, while the TA will be your primary contact for assignment-related queries. For email communication, include the course number and section in the subject line, and allow 24-48 hours for a response.

Letter Grades: The letter grade is based on the overall course score.

Grade Formula						
Grade	A	B +	В	C +	С	D
Overall Course Score Cutoff	90	85	80	75	70	65

If the overall performance falls below the expected average, adjustments may be applied to determine the final letter grade.

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 or copied must be given proper credit.

Use of Generative AI Tools: The use of generative AI tools in any form to directly assist with assignments, quizzes, exams, or any other course-related work is strictly prohibited. These tools may be used for clarifying concepts, supplementing course materials, and practicing problems, but they should not replace your own work. If you do use generative AI tools, you must disclose this and provide the full transcript of your interaction. Unauthorized use of AI tools will lead to disciplinary action, which may include a failing grade for the assignment or course and additional consequences as per the institution's academic integrity policies. For any questions about appropriate use or assistance with coursework, please contact me or utilize available academic support resources.

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