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

Instructor: Dr. Yajuan Li Email: <u>yajuan.li@njit.edu</u> Office: GITC 4309 Office Hour: Monday 4:00-5:30 pm (Office), Friday 10:00-10:30 am (Webex)

Course Overview:

The course is an introductory data science course that focuses on how to develop principled analytics and implementations on a variety of large data sets. The course will focus on learning models, formalism, and algorithmic techniques that are popular in data science and heavily used in practice. The students will be introduced to data science tools, such as performing data analysis with Excel and Python. Extra attention will be paid to strengthen theoretical as well as development/programming skills of the students in performing data analyses using real world small and large-scale datasets.

The primary objectives of the course are:

- Establishing quantitative view and mastering scientific approaches for analyzing large scale datasets.
- Learn data science algorithms and applications.
- Implementation skill of data science algorithms.
- Understand analysis, metrics, visualization and navigation of results.
- Learn how to use existing data science tools (e.g., python, excel)

The outcomes of the course are:

- An ability to perform predictive modeling in various data science applications.
- An ability to perform semi-supervised modeling in various data science applications.
- An ability to perform correlation and clustering analysis in various data science applications.
- An ability to perform analysis on noisy, high dimensional, large-scale datasets.
- An ability to develop end to end solutions to real world data science problems formalizing the problem, identifying appropriate modeling techniques, and developing solutions.
- An ability to implement real-world large-scale data science problems and evaluate its outcome in a principled manner.

Prerequisite: <u>CS 114</u> and (<u>MATH 333</u> or <u>MATH 341</u>) with a grade C or better.

Textbooks:

- Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications by Laura Igual, et. Al.
- Data Mining: Practical Machine Learning Tools and Techniques Frank & Witten
- Applied Statistics for Engineers and Scientists (3E 14) Devore & Farnum

Attendance

Class attendance is mandatory. Getting to class late or leaving early counts as half an absence. If you have special circumstances that prevents from attending class, you must provide documentation to the dean of students for your absence to be excused.

Class Participation

Active participation in class, including asking and answering questions and engaging in classwork activities, is an integral component of our course. Classwork assignments will be assigned during lectures and will contribute to your class participation grade. You have the option to work on these assignments individually or in groups (up to a maximum of two individuals). When collaborating in a group, please designate one representative to submit the work through the Canvas platform, ensuring that both names are included on the submitted assignment to receive credit. It's important to note that late submissions will not be accepted. All classwork should be completed during the class session unless there are exceptional circumstances, such as jury duty or a medical issue. In such cases, verification from the Dean of Students is required.

Please ensure that you come prepared with your personal laptop or device for the purpose of completing assigned classwork.

During class, kindly ensure that mobile phones are powered off. Additionally, it is expected that during class hours, activities such as gaming, texting, emailing, web browsing, or any other unrelated activities are refrained from, and the focus remains on the class content.

Late Policies

 Homework and project must be submitted through Canvas by the specified due date, unless otherwise specified in class. It will not be accepted late.
Students can get extended deadline if they have special/emergency reasons verified from Dean of students. <u>https://www.njit.edu/dos/student-excusals</u>

There will be NO EXCEPTION to these late policies. Please manage your time appropriately.

Exam Policies

There will be one midterm and one final exam. Be sure that you will be present for all your exams. Respondus LockDown Brower and Monitor may be used to monitor and proctor the test, and all students are required to have a working webcam to take the exam. More information about Respondus Lockdown Browser can be found at https://web.respondus.com/student-help/

You must bring a student ID to all exams. There are no late submission or makeup for exams. Students who have special/emergency reasons to reschedule exams must apply and get approval of testing accommodation from the Office of Accessibility Resources and Services before the exams. <u>https://www.njit.edu/accessibility/requesting-testing-accommodations</u>

Grading Scale:

Grade	А	B+	В	C+	С	D	F
Overall Course Score	≥ 90	[85,90)	[80,85)	[75,80)	[70,75)	[60,70)	< 60

Students' final grades are based on the above tentative grading scale. I reserve the right to change the scale as the need arises.

Grading Breakdown:

- Attendance: 5%
- Homework/Class Participation: 25%
- Project: 20%
- Midterm Exam: 20%
- Final Exam: 30%

Tentative Course Topics (Subject to changes according to progress)

- Introduction to data science
- Probability theory and statistics
- Basic Python tools for data science
- Regression algorithms, e.g., Linear regression
- Classification algorithms, e.g., Naïve Bayes
- Clustering algorithms, e.g., K-means clustering
- Association Rule, e.g., Apriori
- Methods in Artificial Intelligence
- Assessing predictive models

Collaboration and Honor Code

Each student is responsible for his/her own work. Students may discuss problems together but must write up their own solutions. When writing up the solutions, students should write the names of people, if any, with whom they discussed the assignment. Note that copying homework or programming assignments, in full or in part is forbidden. Students found cheating or plagiarizing will be immediately referred to the Dean of Students and the NJIT Committee on Professional Conduct and subject to Disciplinary Probation, a permanent marking on the record, possible dismissal, and an "F" grade in the course. All submitted assignments will be checked for similarities, and plagiarism and guilty students identified. In the exam, each student is required to sign the Honor Code Agreement "On my honor, I pledge that I have not violated the provision of the NJIT Student Honor Code."

University Policy on 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: <u>http://www5.njit.edu/policies/sites/policies/files/academic-integrity-</u> <u>code.pdf</u>. 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 <u>dos@njit.edu</u>"

*Students will be notified in class of any changes to the syllabus.