

Spring 2020

CS 301-002: Introduction to Data Science

Senjuti Basu Roy

Follow this and additional works at: <https://digitalcommons.njit.edu/cs-syllabi>

Recommended Citation

Basu Roy, Senjuti, "CS 301-002: Introduction to Data Science" (2020). *Computer Science Syllabi*. 84.
<https://digitalcommons.njit.edu/cs-syllabi/84>

This Syllabus is brought to you for free and open access by the NJIT Syllabi at Digital Commons @ NJIT. It has been accepted for inclusion in Computer Science Syllabi by an authorized administrator of Digital Commons @ NJIT. For more information, please contact digitalcommons@njit.edu.

CS 301, Introduction to Data Science, Spring 2020

Instructor: Senjuti Basu Roy

Location: CKB 206, **Class sessions:** Monday/Wednesday, 1 pm – 2:20 pm, **Instructor Office:** GITC 4302
Office Hours: Monday/Wednesday (11:45 – 12:45 pm) /by appointment

(email: senjutib@njit.edu)

CS 301 Webex

<https://njit.webex.com/njit/j.php?MTID=mac466805df703dedd10d39f162c1c2a4>

Senjuti Basu Roy invites you to join this Webex meeting.

Meeting number (access code): 924 921 970

Meeting password: 1234

Every Monday, Wednesday, from Monday, March 23, 2020, to no end date

1:00 pm | Eastern Daylight Time (New York, GMT-04:00) | 1 hr 20 mins

Teaching Assistant: Mahsa Asadi, email: ma2266@njit.edu , office hour: M/W 2:30-3:30 pm (PhD Lab, GITC 4th Floor)

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 be designed in two parts - during the first part of the course, we will focus on learning models, formalism, and algorithmic techniques that are popular in data science and heavily used in practice. In the second part of the course, 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:

1. An ability to perform predictive modeling in various data science applications;

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

Prerequisite: CS 114, Math 333

Textbooks

Data Mining: Practical Machine Learning Tools and Techniques Frank & Witten

Applied Statistics for Engineers and Scientists (3E 14) Devore & Farnum

Reference Books: *Introduction to Data Science, A python approach to concept, techniques and applications*, by Laura Igual, et. al

Course Content:

There will be several aspects to this course:

1. Instructor lectures
2. Individual assignments: There will be four homework assignments, involving both theory and implementation.
3. Final project: A class project will be assigned and will be completed individually. Additionally, each student must present his or her final project during the presentation week.
4. Midterm and Final Exam
5. Class participation

Course Schedule: (may change)

Dates	Schedule
Week 1	Introduction to Data Science, Understanding different types of data
Week 2	Statistical techniques related to data science

Week 3 Analyzing outcomes of data science processes

Data fitting and outlier analyses

Week 4

Week 5 Data discretization

Week 6 Dimensionality reduction

Week 7 **Midterm**, Assessing predictive models

Week 8 Linear supervised algorithms

Week 9 Non-linear supervised algorithms

Week 10 Semi-supervised algorithms

Week 11	Correlation and clustering techniques
Week 12	Search algorithms for intelligent systems
Week 13	Emerging topics in data science
Week 14	Student Presentation
Week 15	Final Exam

Class Etiquette: It is sometimes useful to distance ourselves from technology to obtain an environment of quality in-class discussion. Hence, kindly minimize the use of personal laptop computers in the classroom unless appropriate. Please turn off all cell-phones and PDAs unless taking notes. If placed on vibrate mode, please ensure that the ‘buzz’ does not disturb your neighbors during the sessions. Please bring a pen (black or blue) and a writing pad to class to take notes.

Come prepared to the class in the following manner: Read the assigned reading several days prior to the session on that topic.

All assignments are to be submitted to canvas on the date due, **unless otherwise specified in class**. If a handout is needed, make sure that your name is on each sheet that is handed in and that any hand-in with multiple pages is stapled. You should as a backup also upload your submissions to appropriate slots on Canvas to facilitate grade recording. Failure to submit /uploads to canvas prior to deadline will result in automatic denial of grade for that work. To keep the grading process on time, extensions are not available and should not be requested.

Grading Scheme*:

I reserve the right to make small adjustments to grade weights, or to add small assignments as the need arises.

Item	Grade (% of final grade)	Due date
Homework Assignments (total 4)	7.5*4=30%	Home works would be posted on Fridays. It would be due on Fridays after 1-week.

HW-1 posted: Jan 31

HW-2 posted: Feb 14

HW-3 posted: March 6

HW-4 Posted: March 20

Posted : April 3

Final Project+ presentation	14+4=18%	Due: May 6, Presentation May 4, 6
Midterm Exam	18%	March 4 TBD
Final Exam	30%	
Class participation and attendance	4%	

All exams are open book and notes. However, Internet access is not allowed.

***20% late submission penalty if submitted within one day of deadline. Beyond that, late submission is not entertained. Unless otherwise stated, all submissions are due by 11:59 pm eastern on the due date.**

An extra credit assignment (could also be a quiz or an exam) may be given to make up for 5% of the course total. The exact date of the extra credit is not decided yet.

Grading Scale

A <= 20%, B+<= 20%, B <= 20%, C+/C/D/F/W the rest of the class

Honor and Policy

- 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."