

Fall 2020

## CS 301-001: 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 301-001: Introduction to Data Science" (2020). *Computer Science Syllabi*. 144.  
<https://digitalcommons.njit.edu/cs-syllabi/144>

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](mailto:digitalcommons@njit.edu).

# CS 301, Introduction to Data Science, Fall 2020

**Instructor: Senjuti Basu Roy**

**Webex link:**

<https://njit.webex.com/njit/j.php?MTID=m599e64f3ec3fe8d10f7237da6797bf09>

**Class sessions:** Monday/Wednesday, 12:30 – 1:50 pm

**Instructor Office:** GITC 4302

**Office Hours:** Monday/Wednesday (2 pm – 3 pm) /by appointment on (use the same Webex link)

(email: [senjutib@njit.edu](mailto:senjutib@njit.edu))

**Teaching Assistant: TBD, email: TBD, office hour: M/W 3-4 pm (GITC 4111)**

**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:

- a. An ability to perform predictive modeling in various data science applications;
- b. An ability to perform semi-supervised modeling in various data science applications;
- c. An ability to perform correlation and clustering analysis in various data science applications;
- d. An ability to perform analysis on noisy, high dimensional, large scale datasets;
- e. An ability to develop end to end solutions to real world data science problems –formalizing the problem, identifying appropriate modeling

- techniques, and developing solutions;
- f. 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**

<b>Data Mining: Practical Machine Learning Tools and Techniques</b>	Frank & Witten	
<b>Applied Statistics for Engineers and Scientists (3E 14)</b>	Devore & Farnum	

**Reference Books:**

<b>Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications</b>	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.
4. 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.
5. Midterm and Final Exam
6. Class participation

Course Schedule: (may change)

<b>Dates</b>	<b>Schedule</b>
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

Week 4	Data fitting and outlier analyses
Week 5	Data discretization
Week 6	Dimensionality reduction
Week 7	<b>Midterm</b> , Assesing 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	<b>Final Exam</b>

### 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: Sept 18  HW-2 posted: Oct 9  HW-3 posted: Nov 6  HW-4 Posted: Nov 27
Final Project+ presentation	15%	Posted : Nov 9  Due: December 9 Presentation TBD
Midterm Exam	20%	October 12
Final Exam	30%	TBD
Class participation and attendance	5%	

**All exams are open book and notes.**

**\*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

*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:*

*<http://www5.njit.edu/policies/sites/policies/files/academic-integrity-code.pdf>.*

*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](mailto:dos@njit.edu)”*