New Jersey Institute of Technology
Department of Computer Science

CS 301 Introduction to Data Science

This 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. 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.

Course (learning) outcomes

- An ability to perform predictive 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 outcomes in a principled manner.

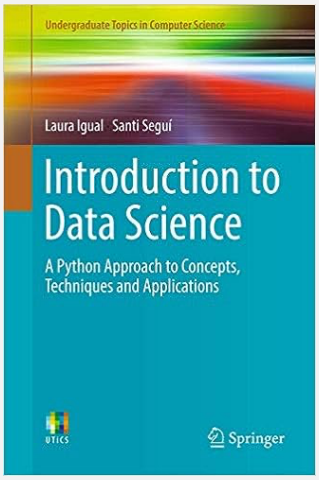
Fall 2023 Section 001
Mon. - Thurs 10:00 – 11:20pm, Mech. Eng. Bldg, Rm. 224
Prerequisite(s): CS 114 and Math 333, 341 or equivalent

Michael Renda
E-Mail: michael.renda@njit.edu
Phone: N/A
Webex:
https://njit.webex.com/meet/mr787njit.edu
Office: GITC 5714
Office Hours: Mondays
1:00 – 1:40 pm
Thursdays
1:00 – 1:40 pm
Fridays
1:00 – 2:25 pm
And by Appointment

Grader: Richa Mishra
E-Mail: rm957@njit.edu
Office Hours: GITC 4403
Wednesday: 11am – 1pm
Thursday: 4 – 6pm

All Email communication between students and faculty should be accomplished using NJIT Email accounts.

Required Textbook(s):



Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications by Laura Igual, et. Al.

Available without charge in PDF format at this link.

Course organization

- The slides for each lecture are available before the class. A good practice is to read from your book the material to be taught in class and to come prepared.
- Homework assignments will be given containing exercises on selected topics.
- You will work in groups on a project which will include programming and a written / oral presentation.

Attendance and Participation:

Class attendance is mandatory and a component of your course grade. See **Grading** below. 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.

Experience shows that students who do not attend classes do not perform well in the midterm and final exams. If you miss a class, be sure to consult one of your classmates about the content of the lecture and visit the course area in Canvas to get notes, exercises, assignments, and announcements.

Many classes will begin with a brief oral quiz on the topics covered in the previous class. These quizzes will count towards the course grade. See **Grading** below.

Classroom Conduct Policies:

- **Turn off cell phones during class**
- **No food or drink are allowed in class**
- **No surfing the Internet, instant messaging, or visiting any social network during lecture**
- **Raise your hand and wait to be recognized**

Communications with the Instructor

Communications with the instructor should be conducted via email (michael.renda@njit.edu) or on Canvas. I will strive to respond as quickly as possible. At a minimum, you should expect to hear back from me within one business day.

All emails **must** originate from an NJIT email account and include the Course and Section Number in the Subject line. Example:

Subject: CS301 001 Question regarding project deliverables

Improperly formatted emails will be returned to the sender.

Other Course Materials:

Other course materials:

- We will be using the Python programming language for this course, specifically Jupiter notebook. Therefore, make sure to have **Anaconda** or **Mini-Conda** platform downloaded on your machine.
- Google Colab might be used as well.
- You will be expected to bring your personal computing device to each class session.
- The following link provides helpful information for various data science topics: https://pantelis.github.io/cs301/

Instructor's Syllabus Statement

This syllabus is subject to change due to student interests, special needs, cancellations, or instructor's decision.

Finally, Welcome. Explore your text options, make sure you have reliable technology, explore the online technology we will be using, and enjoy the adventure.

Homework Assignments

Homework must be submitted through Canvas by the specified due date in the specified format. It will not be accepted late except for an excused absence from the Dean of Students (with a valid, documented reason) that also must be approved by the Instructor.

Project

Each student will work in groups on a project, consisting of a written and an oral presentation.

Project Groups

Groups of *at most* 3 students are required. You can choose the classmates you want to work with. You can also work by yourselves if you prefer. In order to form a group you have to fill a form available in class or in my office during office hours. If you do not express any preference, I will put you in a group randomly.

Submissions and Late policy

The homework assignments and all project deliverables should be submitted on or before the day and time they are due through Canvas. **Late submissions will not be accepted or will get penalties.**

Make-up Exams

Make-up exams require an excused absence from the Dean of Students (with a valid, documented reason) and need to be approved by the Instructor.

Peer Tutoring

YWCC maintains an active program of peer tutors. The tutors for any course have received a grade of A for that course. Many of the undergraduate tutors offer assistance with CS 301. You can find a complete list of tutors and their availability at Undergraduate Tutoring.

Grading

The midterm, the assignments, the project, and the final exam contribute to the course grade as follows:

Attendance	5%	
Participation	5%	
Assignments	20%	
Midterm	20%	
Project	20%	Deliverables and demonstrations
Final	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

Course Outline

Introduction to data science
Probability theory and statistics
Basic Python tools for data science, e.g., data preprocessing
Regression models, e.g., Linear regression
Classification models, e.g., Naïve Bayes
Clustering, e.g., K-means clustering
Decision Tree, Random Forest
Association Rule Mining, e.g., Apriori
Natural Language Processing (NLP)
Methods in Artificial Intelligence, e.g., Neural Networks

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:

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"*

Collaboration and Individual Responsibility

You are encouraged to study and to work on assignments together with others; collaboration is a basic learning technique. You may not take credit for the work of others. You must understand and be able to explain all work that you submit.

Accommodations

If you need accommodations due to a disability please contact Scott Janz, Associate Director of the Office of Accessibility Resources & Services (OARS), Kupfrian Hall 201, to discuss your

specific needs. A Letter of Accommodation Eligibility from the OARS authorizing your accommodations will be required.