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.

Michael Renda
E-Mail: michael.renda@njit.edu
Phone: N/A
Office: GITC 2120
Office Hours: Mondays
           10:00 – 11:30 am
           Wednesdays
           10:00 – 11:30 am
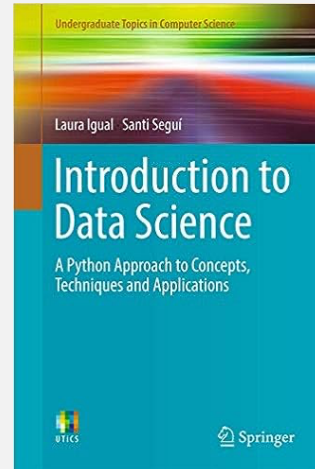   And by Appointment
Grader: Atharv Pallav Rajbanshi
E-Mail: ar2699@njit.edu
Office Hours: Friday, 12:00 – 2:00pm
           GITC 4324

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.
- Two projects will be assigned as described below.

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 quiz on the topics covered in the previous class.

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
Canvas will be used to post lecture notes, assignments and supplementary materials. 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.

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.

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.

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 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. Any violation will be reported to the Dean of Students.

Project

There will be two projects in this course. Students will work individually in their first project and in groups in their second project. The project requirements will be discussed in class and will be posted on Canvas.

Exams

There are two exams in this course: Midterm and Final. You must bring ID to all exams. Students with special needs are advised to make arrangements in advance with the Office of Accessibility Resources and Services (OARS). There are no makeup exams. If you miss an exam because of a documented special circumstance, you may receive a grade based on the other exam or based on the average performance on other parts of the course.

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.**

Grading

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

Attendance	5%	
Participation	5%	
Assignments	15%	
Midterm	20%	
Project 1	15%	
Project 2	15%	
Final	25%	

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

Grade Appeals

If you believe that you deserve more credit than you have been awarded on a particular assignment, 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.

Course Outline

• **Introduction to data science**   o **Basic python tools**
• **Statistics**
• **Data preparation**
• **Regression models, e.g., Linear regression**
• **Evaluating predictive models**
• **Classification models, e.g., Decision Tree, Random Forest, Naïve Bayes**
• **Clustering, e.g., K-means clustering**
• **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. All work that you represent as your own must, in fact, be your own. 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.

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