



Department of Computer Science

CSCI 7090 (A/O1E) Data Science and Machine Learning Project

Syllabus

Vijayalakshmi (Viji) Ramasamy, Ph.D.,

Assistant Professor

Office: IT 2323 (Statesboro Campus)

Email: vramasamy@georgiasouthern.edu

Class Time:	<i>TR: 3:30 – 4:45 PM</i>
CRN, Credits & Location:	<i>CSCI 7090 A (CRN: 14286)</i> Statesboro Campus: Info Technology Bldg, Room 2207 <i>CSCI 7090 01E (CRN: 14183)</i> Online Synchronous — Zoom meeting link on the D2L course website
Number of Credits:	3 Cr
Office Hours:	<i>TR 9:30 am to 11:30 am</i> or by appointment
Final Exam/Paper:	<i>Thursday, May 7, 2026 — 3:00 – 5:00 PM</i> (refer to Final Exam Schedule)

Contents

1	Course Description	2
2	Course Goals & Objectives	2
3	Learning Objectives	2
4	Prerequisites	3
5	Required Text & Course Materials	3
6	Performance Evaluation	3
6.1	Grading Components	3
6.2	Letter Grades	3
6.3	Grading Policies	3
7	Course Policies	4
7.1	Communication	4
7.2	Taking Notes	4
7.3	Participation	4
7.4	Attendance & Absences	5
7.5	Evaluation	5
7.6	Submitting Work	5
7.7	Late Submissions/Presentations	5
8	General Instructions	5
9	Tentative Course Schedule	6

1 Course Description

The Data Science and Machine Learning Project course will help students understand the conceptual underpinnings of the broad field of data science, which touches on aspects of statistics and machine learning. The students will then apply the process, methods, and computing tools to specialized research problems identified by the instructor. You will select a Data Science and Machine Learning topic from a domain of your choice, collect data, perform a literature review of research, analyze previous research related to your topic, analyze the benefits and societal impact, and evaluate the computational techniques. You will present your findings from all stages of the work. The final course product will be a paper that can be submitted to a professional conference and a poster. This course is offered in *face-to-face/online synchronous* and *blended* modes, and most project work is guided, but it can be done independently or in teams of two. In-class meetings will involve lectures, research, development of your paper, presentations, and discussion of future progress with the instructor. It is estimated that you will spend 60% of class hours in class and 40% on your own reviewing literature, project development, and preparing presentations and papers.

2 Course Goals & Objectives

The Data Science and Machine Learning Project course intends to prepare the students to embark on their journey as a Data Scientist and a Machine Learning Engineer. Students will select a research problem/topic from the Data Science and Machine Learning of their choice, perform a detailed and systematic literature review related to their topic, document their understanding, acquire data, perform data wrangling, analyze various machine learning models, and compare and evaluate them. Students will document and present their findings from all stages of their work. The final course product will be a paper edited using LaTeX software for typesetting technical documents. The final paper can be submitted to a peer-reviewed conference and a poster. The students will have a firm understanding of writing research articles in the Data Science and Machine Learning landscape and several types of relevant and most recent research concepts. They will also implement, train, and apply machine learning models in various application domains and understand the caveats of real-world data and settings.

3 Learning Objectives

Upon completion of the course, you will be able to:

- **Develop Research Skills.** Gain a solid understanding of key topics such as web scraping, data characteristics, and systematic literature review methods. Learn to write research articles in LaTeX, collaborate using GitHub repositories, and improve literature reviews to meet academic standards.
- **Apply Data Collection and Wrangling Techniques.** Master the data collection, preprocessing, and visualization processes to prepare datasets for analysis.
- **Utilize Natural Language Processing (NLP) Techniques.** Explore the basics of NLP techniques and their application in extracting insights from text-based datasets.
- **Perform Statistical Analysis and Implement Predictive Analytics.** Leverage statistical modeling and visualization tools to extract insights from data, communicate findings clearly, and develop predictive models using machine learning.
- **Develop Team Collaboration and Communication Skills.** Work collaboratively on team projects that involve literature reviews, data wrangling, ML techniques, and deliverables such as conference papers and poster presentations.
- **Enhance Technical Presentation Skills.** Deliver professional presentations on literature reviews, data wrangling, and final projects, ensuring clear and concise communication of technical work to prepare for academic and professional success in roles in academia or industry.
- **Foster Ethical and Responsible Use of Data.** Understand the importance of ethical considerations and responsible practices when collecting, analyzing, and interpreting data.

These objectives ensure that students develop a comprehensive understanding of data science and machine

learning, both in theory and practice.

4 Prerequisites

This course implicitly requires knowledge of Python programming, Statistics, and some Data Science, Data Mining, or Machine Learning.

5 Required Text & Course Materials

The textbook and reference books provide many examples of code and tools.

1. **Textbook:** *Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*, Ramesh Sharda, Dursun Delen, and Efraim Turban, 11th Ed., Pearson, 2020.

2. **Reference Books:**

- (a) *Python for Data Analysis*, 3E, O'Reilly Media, 2017. Available at wesmckinney.com/book.
- (b) *Data Science from Scratch*, Joel Grus, O'Reilly Media, 2015. Companion code at github.com/joelgrus/data-science-from-scratch.
- (c) *Practical Statistics for Data Scientists: 50 Essential Concepts*, Peter Bruce and Andrew Bruce, O'Reilly Media, 2017. Companion code at github.com/gedeck/practical-statistics-for-data-scientists.
- (d) *Python Data Science Handbook*, Jake VanderPlas, O'Reilly Media, 2016. Freely available at jakevdp.github.io/PythonDataScienceHandbook.
- (e) *Natural Language Processing with Python*, Steven Bird, Ewan Klein, and Edward Loper, O'Reilly Media, 2009. Reference PDF at tjzhifei.github.io/resources/NLTK.pdf.
- (f) *Mining of Massive Datasets (Python Edition)*, Jure Leskovec, Anand Rajaraman, and Jeffrey D. Ullman, Stanford InfoLab, 2010.
- (g) *Introduction to Machine Learning with Python*, Andreas Müller and Sarah Guido, O'Reilly Media, 2017.

All course material, including the PDF version of the textbook, research articles, and essential Power-Point presentations, can be found in the D2L Brightspace LMS (Learning Management System) and is accessible via the web. Course modules provide further details about the assignments (such as quizzes and assignments) to be completed for the specific week. The course assignments include investigating the topic of Data Science and Machine Learning to prepare the students for project presentations and research article writing. The tentative project and assignment due dates are posted below and on D2L.

6 Performance Evaluation

6.1 Grading Components

The table below shows the grade distribution for various categories and grading components. The Team Project comprises four deliverables that sum to 50% of the course grade; each row is listed separately so that assistive technology can announce every due date and weight individually.

6.2 Letter Grades

At the end of the course, letter grades will be assigned based on your cumulative score percentages as follows: **A** = [90–100]; **B** = [80–89]; **C** = [70–79]; **D** = [60–69]; **F** = below 60.

6.3 Grading Policies

Papers and presentations will be submitted on D2L, with formal feedback within one week. Presentation feedback will also occur during class presentations. If the work is insufficient for publication, a grade below B will be given. Students may resubmit their work for higher grades. No grade higher than 92% will be awarded on resubmissions.

#	Category	Component	% of Grade	Date Due
1	Participation	Quizzes, Discussions, Case Studies	10%	Weekly
2	Team Project	Literature Review of selected topic	10%	16-Feb-26
2	Team Project	Data Wrangling and Visualization	10%	16-Mar-26
2	Team Project	DS/ML Techniques, Prediction Analysis	10%	13-Apr-26
2	Team Project	Conference Paper and Poster Presentation	20%	27-Apr-26
<i>Team Project subtotal</i>			<i>50%</i>	
3	Tests	Test 1	10%	17-Feb-26
3	Tests	Test 2	10%	7-Apr-26
4	Final Exam	Final Exam / Final Paper	20%	7-May-26
Total			100%	

7 Course Policies

7.1 Communication

- All course communication will be accomplished using announcements in D2L Brightspace and during lectures. Regularly check the course website and your Georgia Southern e-mail for announcements.
- Email is the best way to communicate with your instructor outside of class. Please include the following marker in the subject line when using email: *CSCI 7090*. This marker shows that the email is from this class and will help me fight spam. If you do not include the marker, I may not read your email.
- I would prefer answering your questions regarding the assignments in *Course Discussions*. Posting your questions may help all the students get a clear picture and reduce my time answering the same question multiple times. If you need assistance with your homework, come to the office hours.

7.2 Taking Notes

Taking notes during lectures is essential because programming activities will often depend on what you wrote in your notes. Be selective and focus on writing sample code, diagrams, and “notes to self.” The lecture sessions include activities like project demonstrations, hands-on programming, and implementation of Python code learned using Google Colaboratory (Colab for short), a web IDE for Python that enables Machine Learning with storage on the cloud.

7.3 Participation

Participation means:

- Keeping up with the reading.
- Attending all in-person lectures.
- Hands-on programming with the instructor during the lectures to appreciate and demonstrate learning activity.
- Asking questions in class and weekly discussions.
- Volunteering answers to questions in class and weekly D2L discussion.
- Encouraging others to get involved in team-based learning activities.
- Contributing to online discussions on weekly assignments (ask good questions, help provide answers; do not post your code).

- **Quizzes.** Regular weekly and in-class online quizzes will be available throughout the course on D2L. These quizzes reinforce learning the key concepts in the course material. Your two lowest quiz grades will be dropped.

7.4 Attendance & Absences

Attendance at lectures and discussions is mandatory because each lecture builds on the previous ones. You are expected to attend all lectures and arrive on time for lectures and exams. If you miss a class meeting due to unforeseen circumstances, email me before class starts to let me know you will not be there. Please submit any work due that day on D2L and coordinate with me to make up any missing classwork. If a student is absent from *three class sessions in a row* without giving prior notice or providing documentation, the instructor may require the student to obtain permission from the department chair before returning to class.

7.5 Evaluation

- There will be TWO tests and a Final Exam (see Tentative Course Schedule for the exact dates).
- No make-ups for missed tests.
- If you miss a test due to an emergency, you must provide written documentation supporting your emergency claim to be excused. **If you are excused for a test absence, you must make it up before it is discussed in class.** This is usually the next lecture period. Please check the *Tentative Course Schedule* for the dates.

7.6 Submitting Work

Assignment submissions and project milestones are all online via our course website on D2L. You can submit as many times as you want. Extensions may be granted in rare cases when extenuating circumstances (such as serious illness or disability, a death in the family, an accident, etc.) exist and are supported by written documentation.

7.7 Late Submissions/Presentations

Assignments, including paper drafts and presentations, must be submitted on or before the due date, with a 10% additional grade deduction per day it is late (starting with the day after it is due). You may miss a class/presentation, but be sure to submit your D2L assignments by the due date. Then you may present the next class period without a drop in grade. You may also present on an earlier meeting date if you plan to miss a class. Presentation grades will not be awarded until the presentation is given in class.

8 General Instructions

Grading questions. If you have a question about a grade, you should contact me within one week of the day the graded work is returned to you. You lose the right to re-grading after that.

Academic dishonesty. Violations of the Georgia Southern University academic honesty policy (including cheating and plagiarism) are taken very seriously. All coursework should be done individually except for weekly lab and programming exercises unless otherwise directed by the instructor. Students may discuss assignments in general terms with other students and may receive assistance from the professor or classmates. However, the assignment must be done by each individual or each specified group. Otherwise, the work will be considered plagiarism and result in a zero score. In addition, all instances of academic misconduct will be reported to the Dean of Students for further disciplinary actions. More information on the academic dishonesty policy and procedure can be found at the [Georgia Southern University Student Code of Conduct](#).

Disability-Related Accommodations. Georgia Southern University is committed to providing reasonable accommodations to students with documented disabilities as required under federal law. Disabilities may include ADD or AD/HD, autism spectrum disorder, brain injury, chronic medical conditions, communication disorders, hearing loss, learning disabilities, mobility impairment, psychological disorders, visual impairment, or temporary injuries. The purpose of disability accommodation is to provide

equal access to the academic material and equal access to demonstrate mastery of the material. If you have a disability and need accommodations, please contact the Student Accessibility Resource Center (SARC). You will need to meet with a SARC staff member who can help you gather documentation of your disability or refer you to an appropriate resource for assessment. Once documentation of the disability is approved, SARC staff will provide you with an accommodation letter detailing the approved accommodations, which you should present to me so we can discuss and implement your accommodations. Disability accommodations work best starting at the beginning of the semester, but can be approved and started at any point in the semester. Accommodations start at the time the accommodation letter is presented to faculty within reasonable timelines; accommodations are not given retroactively. SARC on the Statesboro campus is located on the second floor of Cone Hall; the phone number is (912) 478-1566. SARC for Armstrong and Liberty campuses is located on the second floor of Memorial College Center; the phone number is (912) 344-2572.

Website: [Student Accessibility Resource Center](#).

Accessibility of this Syllabus. This syllabus is designed to meet WCAG 2.1 Level AA. If you require this document in an alternate format (large print, tagged Word document, Braille, or audio), please email the instructor at vramasamy@georgiasouthern.edu or contact SARC. Requests will be fulfilled within three business days.

Campus Technical Support. Information about technical support is located in the *Campus Technology* section of this D2L course module.

Student Support. Georgia Southern provides academic and student support on all campuses. Please see the *Student Support* page in this module.

Confidentiality. Georgia Southern University asks that you respect the rights of faculty and other students as you participate in the education process, including your use of D2L. Accordingly, when accessing D2L course materials, please respect the privacy and personal information of faculty, staff, and other students in academic work, such as class lists, discussion board postings, drafts of papers, and other work produced in the course.

9 Tentative Course Schedule

Key course milestones (Test 1, Test 2, Spring Break, project presentations, final exam) are indicated in bold text in addition to color so that their significance is conveyed without relying on color alone (WCAG 1.4.1).

Week	Date	Topic
Week 1	13, 15 Jan-26	Introduction, Syllabus and Course Resources, Project Scenario and Overview and Readings
Week 2	20, 22 Jan-26	Writing LaTeX articles and linking to GitHub, Review Readings
Week 3	27, 29 Jan-26	Web Scraping, Intro to Nature of Data
Week 4	03, 05 Feb-26	Systematic Literature Review
Week 5	10, 12 Feb-26	Review & Refining Literature Review
Week 6	17 Feb-26 19 Feb-26	Test 1 (milestone) Presentation: Literature Review (milestone)
Week 7	24, 26 Feb-26	Data Collection, Data Wrangling, and Visualization
Week 8	03, 05 Mar-26	Statistical Modeling, and Visualization
Week 9	10, 12 Mar-26	Natural Language Processing (NLP)
Week 10	17, 19 Mar-26	Spring Break (no class)
Week 11	26 Mar-26	Data Wrangling and Visualization
Week 12	31 Mar, 2 Apr-26	Data Mining Process, Methods, and Algorithms — Overview
Week 13	07 Apr-26 09 Apr-26	Test 2 (milestone) Presentation: Data Wrangling and Visualization (milestone)
Week 14	14, 16 Apr-26	DS/ML Techniques and Empirical Analysis
Week 15	21, 23 Apr-26	ML Techniques for Predictive Analytics
Week 16	28, 30 Apr-26	Final Presentation and Final Exam Review (milestone)
Week 17	07 May-26	Final Exam / Final Paper (milestone)