

# CS620/DASC600 – Introduction to Data Science and Analytics

## Fall 2020

### COURSE INFORMATION

- Instructor : Dr. Sampath Jayarathna, Web: <http://www.cs.odu.edu/~sampath/>
- Contact : Office: 3109, Email: [sampath@cs.odu.edu](mailto:sampath@cs.odu.edu) , Phone: (757) 683-7787
- Office Hours : Tuesday, 2.00 PM – 3.00 PM, or email me for an appointment
- PLE : <http://ple.odu.edu/courses/202010/cs620> (Note that the course is offered via ODU PLE)
- Recitation : Only for live session CRN 19924. Tuesday, 4.30 PM – 5.30 PM (Online Zoom).  
Although scheduled as a synchronous class for students registered for live session CRN 19924 (i.e., a class with an assigned days and times), due to the scarcity of suitable classrooms on the ODU campus under the constraints of safe distancing constraints, the course will be delivered via Zoom. Recordings will be made available, but students registered for the live session are expected to attend at the scheduled time.
- Website : <http://www.cs.odu.edu/~sampath/courses/f20/cs620>
- Piazza : <https://piazza.com/odu/fall2020/cs620/home>
- Blackboard : <https://www.blackboard.odu.edu/>
- Prerequisites : There are no specific course prerequisites for this course. But, I expect you to be comfortable coding in Python and its associated libraries, and knowledge in linear algebra and statistics.

### WHAT IS THIS COURSE ABOUT?

Data science is an interdisciplinary blend of the analytical, computational, and statistical skills necessary to extract knowledge from large and complex sets of data. The proliferation of such data has led to an acute shortage of students with data science skills in the local, national, and global economy.

This course will introduce students to this rapidly growing field of Data Science and equip them with some of its basic principles and tools as well as its general mindset. Students will learn concepts, techniques and tools they need to deal with various facets of data science practices. Cross-listed with DASC 600.

### WHAT WILL YOU GET FROM THIS COURSE?

- Define and explain the key concepts and models relevant to data science.
- Understand the processes of data science: identifying the problem to be solved, data collection, preparation, modeling, evaluation and visualization.
- Develop an appreciation of the many techniques for data modeling
- Be comfortable using commercial and open source tool such as python and associated libraries for data analytics and visualization.

## REQUIRED/OPTIONAL MATERIALS:

- **Required textbook.** No textbook is required. All the key course content will be documented in slides, which will be available in the course website after each lecture.
- **List of optional but recommended materials.** You may find some of these optional textbooks helpful, though none are required:
  - A Hands-On Introduction to Data Science by Chirag Shah, Cambridge University Press, April 2, 2020
  - Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, By William McKinney, O'Reilly; 2 edition (October 20, 2017)
  - Data Science from Scratch: First Principles with Python By Joel Grus, O'Reilly 1st edition, 2015

## COURSE STRUCTURE AND TENTATIVE COURSE SCHEDULE

**Structure:** Course is divided into two units with each unit having several modules. Unit-I is the programming unit that covers python, and its associated libraries required for data science activities. Unit-II is the core unit that covers data science concepts and techniques.

This is a hands-on course with a number of programming activities, assignments, and a final data project. Participation in discussion forum (piazza) is mandatory and will also be graded. See discussion participation section for more details. The course project is a major component of this course, which will enable students to apply their knowledge acquired in the course to develop and implement a data science application on one or more of the technologies covered in Unit-I and II.

**Duration:** Unit-I is covered in 6 weeks, Unit-II is covered in 7 weeks. The course project overlaps with Unit-I and II and is spread over 12 weeks. See the topics below for more information.

**Topics:** The tentative topics are as follows.

Unit-I:

Week 1: Syllabus and Introductions, Python Workshop  
Week 2: Pandas  
Week 3: NumPy  
Week 4: Data Wrangling  
Week 5: Unstructured and Semi-Structured Data  
Week 6: NoSQL  
Week 7: Weka

Unit-II:

Week 8: Text Data Analysis and Inference  
Week 9: Pattern Mining  
Week 10: Machine Learning on Data  
Week 11: Machine Learning on Data  
Week 12: Evaluations  
Week 13: Delivering Results  
Week 14: Recommender Systems

**Week 15: Written Exam**

**Module Activity:**

Watching module overview video from the instructor

Reading instruction material  
Completing course activities and submitting all required work to piazza.  
Completing homework assignment (Note that not all modules have homework assignments)  
Participation in piazza discussion forum. This is mandatory and will be graded, see discussion participation section for details

**Online meeting with instructor:**

Meeting with instructor to discuss project at various stages of completion (see Project section for details)  
Presenting the project along with demo (see Project section for details)

**WHAT YOU CAN EXPECT FROM ME:**

I am committed to supporting students with disabilities. If you have challenges related to these issues or others, I want to work with you to help you succeed. Please contact me, since only you can properly communicate your situation to me.

**WHAT YOU CAN GIVE TO THE CLASS:**

It is extremely important for you to be engaged in the course; otherwise, you will wonder what happened to your tuition dollars. So, I encourage you to setup online meetings, ask questions and actively participate at the piazza discussion forum.

**COMMUNICATION**

**Piazza:** All questions will be fielded through Piazza. The primary benefit is that for many questions everyone can see the answer and other students can answer as well. I will endorse good student responses. Additionally, I expect you to actively participate in online discussions at Piazza. You can post public or private messages that can only be seen by the instructor. You will be signed up with your odu email, but you may switch to another email.

**Blackboard:** Blackboard will be used primarily for grade dissemination.

**Email:** If you send email to me, please be sure to include your name and the course number in the body of the e-mail. You should also use an appropriate subject line that looks like “CS620-HW1” etc. Failure to follow these guidelines may result in delayed response. *Again, email should only be used in rare instances*, I will probably point you back to Piazza if you have a question related to course materials and/or relevant to other students in the class.

**COURSE ACTIVITIES**

The scores you receive on the various graded tasks in the class will be weighted as follows:

45%	Data project
20%	Written Exam
25%	Homework Assignments
10%	in-class activities + discussion forum interactions
100%	Your Total Score for the class

**Written Exam:** Written examination will be a short question answer type exam and will be scheduled during the last week of the class. On the week before written exam, I will post a study guide that will help students to prepare for the written examination.

**Homework:** We will have several homework assignments, worth 25% of your overall grade.

**Data Project:** The data project is an opportunity to tackle a more challenging data science activity. Details, requirements and submission information will be on the project section. For the project, you are required to *individually* work on a dataset of your choosing that is interesting, significant, and relevant to Data Science. The ultimate goal of your data project is to apply the techniques learn in each week of the class towards your dataset (exploration, wrangling, machine learning, visualization). We are going to use Google Colab (Colaboratory) (<https://colab.research.google.com/>), a free Jupyter notebook environment that requires no setup and runs entirely in the cloud. With Colaboratory you can write and execute code, save and share your analyses, and access powerful computing resources, all for free from your browser.

**Class Activities/Discussion Forum interactions:** Class activities and participation in the discussion are both important to your success in the course. As one measure of your participation and course preparation, we will have class activities related to lecture topics to supplement the learning.

## GRADES

Final course grades are based on the overall average. You are guaranteed a grade based on a 10% window (e.g., 90-100% is an A). Overall class grade (not the individual grade) windows may be increased in size if the instructor finds it appropriate. Final score in % will be rounded to the nearest whole number. Assigning + or – grades may be made at instructor’s discretion.)

A: 90-100, B: 80-89, C: 70-79, D: 60-69, Fail (Grade F): 0-59

**Grading correction:** The assignment or exam grading correction requests should be sent to the instructor within 1 week of receiving the grade, or before the end of the semester, whichever comes first. After that, your grade will not be adjusted. If you find a mistake in grading, please let the instructor know. Your grade will not be lowered.

There is no separate grading scale for PhD students, but PhD students will typically be held to a higher standard.

## COURSE POLICIES:

**Attendance:** Since this is an on-line course, there is no mandatory attendance policy. However, students are expected to actively participate in the piazza discussions, class activities, homework submissions, and Google Colab project writing. Each of these components is graded and counted towards the final grade.

**Online Classroom Conduct (Netiquette):** Students are expected to follow good Netiquette rules. Netiquette is the accepted behavior for online participation. The following is a list of general guidelines for this course:

- Check your grammar and spelling
- Keep your comments focused on the topic
- Strive to write succinctly and clearly
- Share your knowledge and include supportive evidence for your comments
- Do not use all capital letters as that is viewed as shouting
- Avoid flaming—disrespectful language is unacceptable

Select the link to find more information on [Netiquette](#).

**Tests, Make-ups, and Late Policies:** Due dates will be set to give ample time for completion of the assignments and will not be extended save for the unexpected and unlikely major, long-lived catastrophe. Start projects and homework early--last minute computer malfunctions will not be accepted as a reason for delaying a due date. Changes to a submission's due dates will be avoided because they are unfair to those students who have organized their time to complete the assigned work. Individual accommodations will be discussed if you have a valid medical excuse.

Unless otherwise specified by the instructor, only the written exam will be comprehensive, covering material from the entire course. There are no makeups or rescheduling of exam unless you have a plausible reason with appropriate document or verification. Rescheduling of exams must be arranged at least 3 days in advance. An exam missed without an acceptable excuse will be recorded as a grade of zero (0).

For Homework assignments, each late submission will incur a 5 points penalty per day. A missed submission without an acceptable excuse will be recorded as a grade of zero (0). No submission will be accepted after 3rd day and will be recorded as a grade of zero (0).

There will be no makeup for homework assignments or class activities.

## ACADEMIC OFFENSES

By attending Old Dominion University you have accepted the responsibility to abide by the honor code. If you are uncertain about how the honor code applies to any course activity, you should request clarification from the instructor. The honor pledge is as follows:

"I pledge to support the Honor System of Old Dominion University. I will refrain from any form of academic dishonesty or deception, such as cheating or plagiarism. I am aware that as a member of the academic community, it is my responsibility to turn in all suspected violators of the Honor Code. I will report to a hearing if summoned."

Scholarly dishonesty, especially plagiarism, will not be tolerated. Plagiarism is defined as "Failing to credit sources used in a work product to pass off the work as one's own. Attempting to receive credit for work performed by another, including papers obtained in whole or in part from individuals or other sources." Students found to have engaged in plagiarism will be punished severely, typically earning an automatic F in the course and being reported to the Office of Student Conduct and Academic Integrity.

**Homework Assignments Collaboration Clarification:** To clarify, your homework assignment is yours alone and you are expected to complete each independently. Your solution should be written by you without the direct aid or help of anyone else. However, I believe that collaboration and team work are important for facilitating learning, so I encourage you to discuss problems and general problem approaches (but not actual solutions) with your classmates. If you do have a chat with another student about a problem, you must inform me by writing a note on your submission (e.g., Bob pointed me to the relevant section for problem 3). The basic rule is that no student should explicitly share a solution with another student (and thereby circumvent the basic learning process), but it is okay to share general approaches, directions, and so on. If you feel like you have an issue that needs clarification, feel free to contact me.

## DISABILITY RESOURCES

In compliance with PL94-142 and more recent federal legislation affirming the rights of disabled individuals, provisions will be made for students with special needs on an individual basis. The student must have been identified as special needs by the university and an appropriate letter must be provided to the course

instructor. Provision will be made based upon written guidelines from the University's <https://www.odu.edu/educationalaccessibility> . All students are expected to fulfill all course requirements.

Students are encouraged to self-disclose disabilities that have been verified by the Office of Educational Accessibility by providing Accommodation Letters to their instructors early in the semester in order to start receiving accommodations. Accommodations will not be made until the Accommodation Letters are provided to instructors each semester. For additional information visit the Office of Educational Accessibility online or at 1525 Webb Center.

## OTHER POLICIES

**University Email Policy:** The Old Dominion University e-mail system is the official electronic mail system for distributing course-related Communications, policies, Announcements and other information. In addition, the University e-mail user ID and password are necessary for authentication and access to numerous electronic resources (online courses, faculty Web pages, etc.)

**Withdrawal:** A syllabus constitutes an agreement between the student and the course instructor about course requirements. Participation in this course indicates your acceptance of its teaching focus, requirements, and policies. Please review the syllabus and the course requirements as soon as possible. If you believe that the nature of this course does not meet your interests, needs or expectations, if you are not prepared for the amount of work involved - or if you anticipate that the class meetings, assignment deadlines or abiding by the course policies will constitute an unacceptable hardship for you - you should drop the class by the drop/add deadline, which is located in the ODU Schedule of Classes. For more information, please visit the Office of the Registrar.