

# CS620/DASC600 – Introduction to Data Science and Analytics

## Fall 2019

### 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
- Schedule : Tuesday, Room: DRGS 1117, Time: 4.20 PM – 7.00 PM
- Website : <http://www.cs.odu.edu/~sampath/courses/f19/cs620>
- Piazza : <https://piazza.com/odu/fall2019/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 learning new programming languages/tools/APIs, and knowledge in linear algebra and statistics.

### WHAT IS THIS COURSE ABOUT?

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:
  - 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

- **Bring Your Own Device (BYOD).** You must have a computing device (Laptop, Tablet, or Phablet), we will do some activities in class and you should have a device in class to fully participate.

## TENTATIVE COURSE SCHEDULE

**Topics:** The tentative topics are as follows. Topics and specific course activities may change as needed. PowerPoint slides will be available on the course web page after each lecture.

Week 1 (Aug 27): Syllabus and Introductions, Python Workshop

Week 2 (Sep 3): Pandas

Week 3 (Sep 10): NumPy

Week 4 (Sep 17): Data Wrangling

Week 5 (Sep 24): Unstructured and Semi-Structured Data

Week 6 (Oct 1): NoSQL

**Week 7 (Oct 8): Mid-Term Exam**

Week 8 (Oct 15): *No Class – Fall Break*

Week 9 (Oct 22): Text Data Analysis and Inference

Week 10 (Oct 29): Machine Learning on Data

Week 11 (Nov 5): Machine Learning on Data

Week 12 (Nov 12): Evaluations

Week 13 (Nov 19): Delivering Results

Week 14 (Nov 26): Recommender Systems + Final Revision

**Week 15 (Dec 3): Final Exam**

## WHAT YOU CAN EXPECT FROM ME:

I have an open-door policy i.e., office visits. My posted office hours are times when I will make concerted effort to be available. Occasionally administrative meetings or emergencies may interfere with these posted times. The open-door policy is: if my door is open, I am in and welcome walk-in visitations. 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 come and talk to 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 fall asleep and wonder what happened to your tuition dollars. So, I encourage you to ask questions during lecture and actively participate at the piazza forum. For the first few weeks, when asking a question at the class, state your name so that I know who you are.

**Cell phones and Tardiness:** You may have cell phones in class, but they must be on mute, or airplane mode and not answered until the end of class. You are expected to arrive on time so that you do not cause a disruption in the middle of class. I would like to start the class at the scheduled time. If you cannot make it on time or want to leave early for some reason, please let me know. Persistent tardiness will be noted.

## 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:

25%	Final Exam (Final is comprehensive), <b>Tuesday, December 03 from 5.00 pm to 7.00 PM</b>
25%	Mid-term Exam, <b>Tuesday, October 08, from 5.00 pm to 6.30 PM</b>
25%	Homework Assignments (5)
15%	Data project
10%	in-class activities + Attendance
100%	Your Total Score for the class

**Final Exam:** The final exam is comprehensive, closed books and will be held on **Tuesday, December 03 from 5.00 pm to 7.00 PM**. You may bring one standard 8.5" by 11" piece of paper with any notes you deem appropriate or significant (front and back).

**Midterm Exam:** The midterm exam will be held on **Tuesday during class time**. For both exams, no iPads, iPhones, Blackberries, Android phones/tablets are allowed. Standard calculators are allowed.

**Homework:** We will have 5 homework assignments, each worth 5% 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 course website. 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.

**In Class Activities/Attendance:** Attendance in class 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 in class activities related to lecture topics to supplement the learning. I will ask you to bring a computing device (laptop, tablet).

- Will receive points (0-2) for each week of class for attendance. Attendance rubric:
  - 0 - unexcused absence (excused absences must be requested before class)
  - 1 - late to class or not prepared to participate
  - 2 - on time to class and actively participating in-class activities

## 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:** Bring any assignment or exam grading correction requests to the instructor within 1 week of receiving the grade, or before the end of the quarter, 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.

## ATTENDANCE, MAKE-UPS AND LATE POLICIES:

**All project reports, homework assignments, are due at the beginning of class in all required forms (e.g., paper and/or submit on blackboard) on the 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.

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.

Unless otherwise specified by the instructor, only the final exam will be comprehensive, covering material from the entire course. There are no makeups or rescheduling of exams unless you have a plausible reason with appropriate document or verification for absence. Rescheduling of exams must be arranged at least one week in advance. An exam missed without an acceptable excuse will be recorded as a grade of zero (0). Please also be aware that no electronic devices are allowed during the exam. Having another exam on the same day is not an acceptable excuse for requesting to reschedule the exam.

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.