

COMP 4449—Data Science Capstone

Course Overview

After a brief recap of all the stages required to put together a successful data science project, two sets of challenges of increasing complexity are presented as midterm and term projects. These two projects are implemented, documented, tested, and presented by the student or student team.

Objectives

- Understand how and why data science projects work and how to build, steer, refine, and get the most of a project
- Integrate prior knowledge on data science and engineering regarding how to apply it in real-world contexts
- Assess the value of data science products or services in enterprisedecision-making
- Learn to design, develop, test, and present “full-cycle” data science products

Textbooks and Materials

- [1] McKinney, W. (2018). *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython* (2nd ed.). O'Reilly.
- [2] Zhang, A. (2018). *Data analytics: Practical guide to leveraging the power of algorithms, data science, data mining, statistics, Big Data, and predictive analysis to improve business, work, and life*. Kindle Edition.
- [3] Braschler, M., Stadelmann, T., & Stockinger, K. (Editors). (2019). *Applied data science: Lessons learned for the data-driven business*. Springer.

Optional Reading

- [1] Raschka, S. (2015). *Python machine learning: Unlock deeper insights into machine learning with this vital guide to cutting-edge predictive analytics*. Packt Publishing.
- [2] Provost, F., & Fawcett, T. (2019). *Data science for business*. O'Reilly.
- [3] Kazil, J., & Jarmul, K. (2016). *Data wrangling with Python*. O'Reilly.
- [4] Casella, G., & Berger, R.L. (2017). *Statistical inference* (2nd ed.). Cengage Learning.
- [5] Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- [6] Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.
- [7] Barber, D. (2012). *Bayesian reasoning and machine learning*. Cambridge University Press.

[8] Harrington, P. (2012). *Machine learning in action*. Manning Publications.

External Online Reading

- **Python tutorials:** www.learnpython.org; docs.python.org/tutorial; hackr.io/tutorials/learn-python
- **Data analysis tutorials:** pythonprogramming.net/data-analysis-tutorials; www.datacamp.com/community/tutorials; www.dataquest.io/python-tutorials-for-data-science
- **Visual analysis tutorials:** ieevis.org/year/info/tutorials; realpython.com/tutorials/data-viz

Course Resources

- **Python Notebooks** will be located in the digital learning platform with examples, hands-on projects, and additional materials.

Grading

Assignment/Assessment	Points	Weight on Final Grade
Midterm Project Presentation	100	35%
Term Project Presentation	100	65%

Grading Scale

A 93-100
A- 90-92.99
B+ 86-89.99
B 83-85.99
B- 80-82.99
C+ 76-79.99
C 73-75.99
C- 70-72.99
D+ 66-69.99
D 63-65.99
D- 60-62.99
F < 60

Assignment and Assessment Information

Grading is based on the presentation of an individual midterm project and a two-person term project. Both submissions consist of a Notebook that includes all the required files and code and an Overleaf report.

1. Midterm project reports should be a 3- to 4-page description of the complete experience (dataset description, data preparation and analysis, results and discussion), together with a slideshow adequate for a 4- to 6-minute presentation in class. Your presentation

materials are to be submitted to faculty immediately after your presentation. Please find the midterm project presentation rubric in the class Toolbox.

2. Term project reports are due immediately after your presentation and should be a 12- to 15-page description that includes (among others) the following topics:
 - Research question: What problem are you solving? What is the usefulness of the project?
 - What dataset and metadata was collected, and why?
 - Required data munging and wrangling procedures
 - Statistical and visual data exploration
 - Data analysis and modeling
 - Model evaluation and visualization
 - Discussion, conclusion, further work
 - A slideshow adequate for a 10– to 15-minute comprehensive presentation in class; please find the term project presentation rubric in the class Toolbox.
 -

Weekly Schedule

Week 1

Readings:

- Zhang, Chapter 7
- Braschler et al., Chapters 1 and 2

Additional material: Notebooks with ungraded projects to solve in class.

Week 2

Readings:

- Zhang, Chapters 21–26
- Braschler et al., Chapters 8 and 9

Additional material: Notebooks with examples and with midterm project statements

Week 3

Readings:

- McKinney, Chapters 7 and 8

Additional material: Notebooks with examples and learning material

Week 4

Work on Midterm Project.

Week 5

Midterm project presentation and materials due during the Week 5 live session.

Week 6

Readings:

- Zheng, Chapter 8, 9, 10, and 27
- McKinney, Chapters 9 and 10

Additional material: Notebooks with examples and learning material and term project statements

Week 7

Work on Term Project.

Week 8

Work on Term Project.

Week 9

Work on Term Project.

Week 10

Term project presentation and materials due during the Week 10 live session.

Attendance Policy

Attendance at all live session meetings is mandatory.

Program Mission

Our MS in data science provides students with a broad course of study in programming, algorithms, statistics, and data management, as well as a depth of understanding in specific fields such as data mining, machine learning, and parallel systems. Graduates of the data science program go on to work in a wide variety of careers, including business, government, education, and the natural sciences.

Honor Code and Academic Integrity

All students are expected to abide by the [University of Denver Honor Code](#). These expectations include the application of academic integrity and honesty in your class participation and assignments. Violations of these policies include, but are not limited to

- Plagiarism, including any representation of another's work or ideas as one's own in academic and educational submissions
- Cheating, including any actual or attempted use of resources not authorized by the instructor(s) for academic submissions
- Fabrication, including any falsification or creation of data, research or resources to support academic submissions

Violations of the Honor Code may have serious consequences including, but not limited to, a zero for an assignment or exam, a failing grade in the course, and reporting of violations to the Office of Student Conduct.

Diversity, Inclusiveness, Respect

DU has a core commitment to fostering a diverse learning community that is inclusive and respectful. Our diversity is reflected by differences in race, culture, age, religion, sexual orientation, socioeconomic background, and myriad other social identities and life experiences. The goal of inclusiveness, in a diverse community, encourages and appreciates expressions of different ideas, opinions, and beliefs, so that conversations and interactions that could potentially be divisive turn instead into opportunities for intellectual and personal enrichment.

A dedication to inclusiveness requires respecting what others say, their right to say it, and the thoughtful consideration of others' communication. Both speaking up AND listening are valuable tools for furthering thoughtful, enlightening dialogue. Respecting one another's individual differences is critical in transforming a collection of diverse individuals into an inclusive, collaborative and excellent learning community. Our core commitment shapes our core expectation for behavior inside and outside of the classroom.