

Elements of Data Science: A First Course

Fall 2018

Time: Mondays, 7:00 - 9:30 PM

Instructor: Bryan R. Gibson

Textbook: Python Data Science Handbook, Jake VanderPas

(Free at: <https://github.com/jakevdp/PythonDataScienceHandbook>)

Prerequisite(s):

- Introductory programming class as well as basic familiarity with Python 3.
- Linear algebra; concepts such as vectors and matrices as well as basic data structures such as arrays, hashes, trees, etc.

Course Description

This course is designed as an introduction to elements that constitutes the skill set of a data scientist. The course will focus on the utility of these elements in common tasks of a data scientist, rather than their theoretical formulation and properties. The course provides a foundation of basic theory and methodology with applied examples to analyze large engineering, business, and social data for data science problems. Hands-on experiments with Python will be emphasized. The programming language utilized will depend on the course section, which provides the opportunity to focus on a specific programming language while covering the same skills.

Topics include:

- Data Cleaning, Exploration and Visualization
- Classification, Regression and Clustering
- Dimensionality Reduction
- Model Evaluation and Model Selection
- Feature Engineering and Feature Selection
- Statistical Analysis and Hypothesis Testing
- Natural Language Processing and Topic Modeling
- Data processing and delivery using ETL and APIs
- Time Series Analysis
- Recommendation Engines
- Image Recognition

Assignments and Grading

Participation	10%
Homework Assignments (Four, equally weighted at 10% each)	40%
Midterm Exam	25%
Final Exam	25%

TOTAL	100%
--------------	-------------

Quality of Performance	Letter Grade	Range %	GPA/Quality Pts.
Excellent - work is of exceptional quality	A+	98 - 100	4.33
	A	93 - 97.9	4.0
	A-	90 - 92.9	3.67
Good - work is above average	B+	87 - 89.9	3.33
Satisfactory	B	83 - 86.9	3.0
Below Average	B-	80 - 82.9	2.67
Poor	C+	77 - 79.9	2.33
	C	73 - 76.9	2.0
	C-	70 - 72.9	1.67
	D	65 - 69.9	1.0
	D-	60 - 64.9	0.67
Failure	F	< 60	0.0

Weekly Outline

Week	Topic	Readings	ToDo
1	Introduction to Data Science Problems and Tools		
2	Data Processing and Delivery: ETL and APIs		
3	Data Exploration and Visualization		
4	Statistical Modeling and Hypothesis Testing		
5	Classification and Regression		

6	Feature Engineering and Application: Natural Language Processing		
7	Midterm Exam		
8	Feature Selection, Model Evaluation and Selection		
9	Project Reporting and Application: Time Series		
10	Dimensionality Reduction and Application: Image Recognition		
11	Clustering and Topic Modeling		
12	Data Cleaning and Management		
13	Application: Recommendation Engine		
14	Review, summary and Final Exam		