DATA170
Exploring Data
Syllabus

Instructor: Dr. Durell Bouchard
Office Hours: by appointment
Office: Trexler 365-C
E-Mail: bouchard@roanoke.edu
Phone: 375-4901

Course Objectives

The amount of data we generate when interacting with the digital world is staggering. In one second, users create 347,222 posts on Instagram, upload 500 hours of video to YouTube, and order 6,659 packages from Amazon. To understand the vast quantities of information generated, we need sophisticated algorithms to simplify, analyze, and visualize the data. And the companies collecting the digital information need data scientists who understand the algorithms. This class will introduce you to the tools and techniques required to become a data scientist. You will learn to apply machine learning to large datasets to explore data and make predictions.

Intended Learning Outcomes: At the end of the course, the successful student will be able to

1. write programs that use machine learning to make predictions.
2. correctly format and manage data.
3. report and interpret the performance of prediction models.

Course Content

Prerequisites: CPSC120

Project: The course will culminate in a project that uses machine learning to create a model that can make predictions from data. This project is designed to allow you to put together all of the skills and techniques you learn throughout the semester to explore a dataset that interests you.

Assignments: We will have regular small programming assignments that are designed to reinforce class concepts. These assignments are an opportunity for you to demonstrate that you are ready to apply what you have learned to the project.

Activities: Programming activities during class give you a structured experience in data cleaning and analysis. The activities connect the reading and lectures to the practice of data science and prepare you for assignments.

Co-curricular: The Department of Mathematics, Computer Science, and Physics is offering a series of lectures designed to engage the campus community in discussions of ongoing research, novel applications, and other issues that face these disciplines. You may submit to Inquire up to two papers reflecting on a talk you attend for extra credit.

Grading: Course grades are assigned based on the following weights and scale:

Grade Weights

Category	Weight
Project	34%
Activities	32%
Assignments	32%

Grade Scale

Grade	Range	Grade	Range
A	93-100	C	73-76
A-	90-92	C-	70-72
B+	87-89	D+	67-69
B	83-86	D	63-66
B-	80-82	D-	60-62
C+	77-79	F	0-59

Course Policies

Attendance Policy: If you have a temperature of 100.4 or higher or other COVID symptoms, don't come to class. Call Health Services IMMEDIATELY. Do not come to class or go to any public area on campus. For your absence to be excused, you must permit Health Services to notify me that you have consulted them about COVID symptoms. If Health Services informs you that you should isolate and not attend class for multiple days, tell me to make a plan to keep you current in the course. All absences caused by consultation with Health Services about coronavirus symptoms or isolation ordered by Health Services will be excused. Still, you will need to do the work and graded assignments even if we extend your deadline.

Class attendance is vital to your success in this course; the material covered during missed sessions is the responsibility of the student. Conversations in class illuminate the published class materials and are subject to evaluation on subsequent tests and quizzes. If you anticipate being unable to attend class, email me before class to be excused.

Masks: The College has issued a mask mandate for the semester that requires masks to be worn in common indoor spaces such as our classroom. You must wear a mask in this class. If you arrive without a mask, you will not be allowed to stay and may lose credit for attendance or in-

class work. The bookstore sells masks if you need to make a quick purchase. If the mandate is extended, you will be required to continue to wear a mask.

Late Work: If you anticipate being unable to meet a deadline, email me before the deadline to request an extension. Unexcused late work will receive no credit.

Academic Integrity: Collaboration is a fundamental part of learning. You are encouraged to discuss and learn from one another on the activities. However, unless expressly stated otherwise, all work on assignments and the project should be solely your own. It is accepted that you have read and understood the standards for academic integrity at Roanoke College. If you are ever uncertain about how the policy pertains to any assignments in this course, please ask me for clarification.

Subject Tutoring: Subject Tutoring, located on the lower level of Fintel Library (Room 5), is open 4-9 PM, Sunday-Thursday. Subject Tutors are highly trained, current students who offer free, one-on-one (and small group) tutorials in over 80 courses taught at Roanoke College, including: Business, Economics, Mathematics, INQ 240, Modern Languages, Lab Sciences, and Social Sciences. Check out all available subjects and schedule 30- or 60-minute appointments at <www.roanoke.edu/tutoring>. If you have a question, feel free to stop by, or contact us at subject_tutoring@roanoke.edu or 540-375-2590. See you soon!

Accessible Education Services: Accessible Education Services (AES) is located in the Goode-Pasfield Center for Learning and Teaching in Fintel Library. AES provides reasonable accommodations to students with documented disabilities. To register for services, students must self-identify to AES, complete the registration process, and provide current documentation of a disability along with recommendations from the qualified specialist. Please contact Becky Harman, Assistant Director of Academic Services for Accessible Education, at 540-375-2247 or by email at aes@roanoke.edu to schedule an appointment. If you have registered with AES in the past and would like to receive academic accommodations for this semester, please contact Becky Harman at your earliest convenience to schedule an appointment and/or obtain your accommodation letter for the current semester.

Diversity: I consider this classroom to be a place where you will be treated with respect, and I welcome individuals of all ages, backgrounds, beliefs, ethnicities, genders, gender identities, gender expressions, national origins, religious affiliations, sexual orientations, ability – and other visible and nonvisible differences. All members of this class are expected to contribute to a respectful, welcoming, and inclusive environment for every other member of the class.

Preferred Name/Pronoun: I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester to make appropriate changes to my records.

Course Schedule

This course expects you to spend at least 12 hours of work each week inside and outside of class.

Date	Topic
Tuesday, January 18	Introduction
Thursday, January 20	Python Review
Tuesday, January 25	File I/O
Thursday, January 27	Nearest Neighbor
Tuesday, February 1	Numpy Arrays
Thursday, February 3	Array Operations
Tuesday, February 8	Array Indexing
Thursday, February 10	Array Shaping
Tuesday, February 15	Fancy indexing
Thursday, February 17	Pandas
Tuesday, February 22	Matplotlib
Thursday, February 24	Data Preparation
Tuesday, March 1	Perceptron
Thursday, March 3	Neural Network
Spring Break	
Tuesday, March 15	Hyper Parameters
Thursday, March 17	Cross Validation
Tuesday, March 22	Regression
Thursday, March 24	Feature Selection
Tuesday, March 29	Naive Bayes
Thursday, March 31	SVM
Tuesday, April 5	Decision Tree
Thursday, April 7	Dimension Reduction
Tuesday, April 12	Class Imbalance
Thursday, April 14	Kaggle Competitions
Tuesday, April 19	Ensemble Learning
Thursday, April 21	Tensor Flow
Thursday, April 28 (8:30 AM - 11:30 AM)	Presentations