INTRODUCTION TO DATA SCIENCE

Course Number: CAP4770

Section: ONLN (20952)

Credit Hours: 3

Academic Term: Spring 2022

Class meeting time and location: This class is facilitated 100% online.

INSTRUCTOR INFORMATION

INSTRUCTOR NAME: Daisy Zhe Wang, Ph.D.

PHONE: 352-294-6677

EMAIL: daisyw@cise.ufl.edu

OFFICE HOURS: TBD (online through Zoom Conferences).

COURSE TA OR COORDINATOR:
Ira Harmon, *iharmon1@ufl.edu*, office hour TBD (online through Zoom Conferences)

COURSE INFORMATION

COURSE WEBSITE: http://elearning.ufl.edu

COURSE COMMUNICATIONS: Contact the instructor using the Canvas Mail Tool. Use email for ONLY personal questions such as grades, special circumstances, needed accommodations.

Expect a response within 2 business days. All e-mail correspondence to the course instructor must originate from your ufl.edu account, have your full name in the body of the e-mail, and contain your course and section number in the subject line.

REQUIRED TEXTBOOK:

DM = Han, J., Kamber, M., & Pei, J. (2012). *Data mining concepts and techniques* (3rd ed.). Waltham, MA: Elsevier.

RECOMMENDED TEXTBOOKS:

MMD = Leskovec, J., Rajaraman, A., & Ullman, J. (2014). *Mining of massive datasets* (2nd ed.). Cambridge, United Kingdom: Cambridge University Press. Retrieved from http://www.mmds.org/

SUPPLEMENTAL RESOURCES:

DSS = Grus, J. (2015). *Data science from scratch*. Sebastopol, CA: O'Reilly Media Inc. Retrieved from Google Scholar https://ipfs.io/ipfs/QmNz7iBVwYraL85DSRa4kC4xyM3h1WyN8RkLDSypPCcu4W/ Data%20Science%20from%20Scratch-%20First%20Principles%20with%20Python.pdf

PDA = McKinney, W. (2013). *Python for data analysis*. Sebastopol, CA: O'Reilly Media, Inc. Retrieved from https://proquest.safaribooksonline.com/9781449323592

NLTK = Bird, S., Klein, E., and Loper, E. (2009). *Natural language processing with python - analyzing text with the natural language toolkit* (1st ed. updated). Available from http://www.nltk.org/book/

MATERIALS AND SUPPLIES FEES: None

COURSE DESCRIPTION: Introduces the basics of data science including programming for data analytics, file management, relational databases, classification, clustering, and regression; lays the foundation for big data applications ranging from social networks to medical and business informatics.

PREREQUISITE KNOWLEDGE AND SKILLS: Data Structures and Algorithms (COP3530) or equivalent. Optional but preferred: Information and Database Systems I (CIS 4301) and an introductory course to probabilities and statistics. Previous experience in Python programming required.

COURSE OBJECTIVES:

By the end of this course, students will be able to:

1.   Implement the basic data science techniques including programming in Python and MapReduce for small and big data manipulation and analytics.

2.   Analyze topics including data preparation, data analytics including pattern mining, classification, clustering, data visualization, and parallel computing platforms.

3.   Apply advanced data analytics including text retrieval, information extraction, graph analysis, and knowledge bases.

During the course, we will also make an effort to introduce key application areas of data science including business intelligence, social sciences, life sciences, biomedical informatics, and legal analytics.

COURSE POLICIES

ATTENDANCE POLICY: Requirements for class attendance and make-up exams, assignments, and other work in this course are consistent with university policies that can be found on the [Attendance Policies page](#).

MAKE-UP POLICY:

I am extremely sympathetic if you have some conflict that will make it difficult for you to attend a quiz or get an assignment completed on time. To be fair to everyone in the class, these rules are always followed to the letter and without exception.

1.   If you have some conflict and feel like you may need an extra day on the assignment or need to take the quiz a day or two later, it must be cleared with me no fewer than one week (seven days) before the assignment is due or the quiz will be held. However, if you do not clear it a week in advance, you will receive a zero, with only a two exceptions, given below.

2.   If you are ill at the time of a quiz or right before an assignment is due and so you miss the one week window, or if there is a death in your immediate family, I will allow a late assignment or a make-up provided (a) you can give me proof of the circumstances, and (b) you let me know before the quiz is held or the assignment is due.

3.   If you simply don't turn in the assignment or don't show up for the quiz, the only valid excuse is a note from a doctor given as proof that you were injured or ill at the due date to such an extent that it would be unreasonable for you to send me an email or leave a message letting me know of your illness or injury. In any other case, the result is a zero on the assignment.

REGRADE REQUESTS POLICY:

Any request for an actual regrade must be made in writing within one week of the time that the grading is returned, with no exceptions. All regrade requests must be prepared using a word processing program; a copy should then be emailed to me and TA, along with the original graded work. On your regrade request, carefully describe why you feel that you were scored unfairly and/or incorrectly. Even if you discussed the grading issue

orally with someone, the written discussion must be self-contained and be able to be evaluated based only on what is written on the assignment or exam.

COURSE TECHNOLOGY: This course is facilitated 100% online through Canvas. You may access Canvas from UF's e-Learning webpage: http://elearning.ufl.edu/. Please contact the UF Help Desk, http://helpdesk.ufl.edu, if you have any technical difficulties with Canvas. Please review Accessibility within Canvas for accommodation information.

This course will be using Python, Virtual Machine, and Amazon Web Services (AWS). AWS accounts and credits will be available to each student.

ONLINE COURSE EVALUATION: Students are expected to provide feedback on the quality of instruction in this course by completing online evaluations at https://evaluations.ufl.edu. Evaluations are typically open during the last two or three weeks of the semesters, but students will be given specific times when they are open. Summary results of these assessments are available to students at https://evaluations.ufl.edu/results.

UF AND COLLEGE POLICIES

UNIVERSITY POLICY ON ACCOMMODATING STUDENTS WITH DISABILITIES (REQUIRED): Students with disabilities requesting accommodations should first register with the Disability Resource Center (352-392-8565, www.dso.ufl.edu/drc) by providing appropriate documentation. Once registered, students will receive an accommodation letter which must be presented to the instructor when requesting accommodation. Students with disabilities should follow this procedure as early as possible in the semester.

UNIVERSITY POLICY ON ACADEMIC CONDUCT: UF students are bound by The Honor Pledge which states, "We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honesty and integrity by abiding by the Honor Code. On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: "On my honor, I have neither given nor received unauthorized aid in doing this assignment." The Honor Code specifies a number of behaviors that are in violation of this code and the possible sanctions. Furthermore, you are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have any questions or concerns, please consult with the instructor or TA in this class.

CLASS DEMEANOR OR NETIQUETTE: All members of the class are expected to follow rules of common courtesy in all email messages, threaded discussions, and chats. Review the [Netiquette Guide for Online Courses](#) for expected student behavior.

SOFTWARE USE: All faculty, staff, and students of the university are required and expected to obey the laws and legal agreements governing software use. Failure to do so can lead to monetary damages and/or criminal penalties for the individual violator. Because such violations are also against University policies and rules, disciplinary action will be taken as appropriate. We, the members of the University of Florida community, pledge to uphold ourselves and our peers to the highest standards of honesty and integrity.

COMMITMENT TO A SAFE AND INCLUSIVE LEARNING ENVIRONMENT: The Herbert Wertheim College of Engineering values broad diversity within our community and is committed to individual and group empowerment, inclusion, and the elimination of discrimination. It is expected that every person in this class will treat one another with dignity and respect regardless of gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture.

If you feel like your performance in class is being impacted by discrimination or harassment of any kind, please contact your instructor or any of the following:

- Your academic advisor or Graduate Program Coordinator
- Robin Bielling, Director of Human Resources, 352-392-0903, rbielling@eng.ufl.edu
- Curtis Taylor, Associate Dean of Student Affairs, 352-392-2177, taylor@eng.ufl.edu
- Toshikazu Nishida, Associate Dean of Academic Affairs, 352-392-0943, nishida@ufl.edu

SEXUAL DISCRIMINATION, HARASSMENT, ASSAULT, OR VIOLENCE: If you or a friend has been subjected to sexual discrimination, sexual harassment, sexual assault, or violence contact the Office of Title IX Compliance, located at Yon Hall Room 427, 1908 Stadium Road, (352) 273-1094, title-ix@ufl.edu.

GETTING HELP AND RESOURCES

For issues with technical difficulties for Canvas, please contact the UF Help Desk at http://helpdesk.ufl.edu or (352) 392-HELP (4357).

Any requests for make-ups due to technical issues MUST be accompanied by the ticket number received from the Help Desk when the problem was reported to them. The ticket number will document the time and date of the problem. You MUST e-mail your instructor within 24 hours of the technical difficulty if you wish to request a make-up.

Other resources are available at Distance Learning's Getting Help for:

- Counseling and Wellness resources
- Disability resources
- Resources for handling student concerns and complaints
- Library Help Desk support

Should you have any complaints with your experience in this course, please visit Distance Learning's Student Complaint Process to submit a complaint.

GRADING POLICIES

METHODS BY WHICH STUDENTS WILL BE EVALUATED AND THEIR GRADE DETERMINED (REQUIRED):

Participation: online and zoom discussions: 5%

Labs: 3 different assignments (8% each) and 1 setup lab (1%): 25%

Exams: Midterm and Final (25% each): 50%

Final Project: Groups of two: 20%

COURSE GRADING POLICY: Grades will be determined based on your performance on the following activities:

Assignment	Points or percentage
Labs	25%
Midterm	25%
Final	25%
Final Project	20%

Assignment	Points or percentage
Participation	5%

Note: Assignments, quizzes and exercises will not be graded but will be taken into account in participation.

GRADING SCALE: For more information, review [Frequently Asked Questions for Minus Grades](#).

Percent	Grade
90.0-100.0	A
87.5-89.9	A-
85.0-87.4	B+
80.0-84.9	B
77.5-79.9	B-
75.0-77.4	C+
70.0-74.9	C
67.5-69.9	C-
65.0-67.4	D+
60.0-64.9	D
0-59.9	E

COURSE SCHEDULE

A WEEKLY SCHEDULE OF TOPICS AND ASSIGNMENTS:

Week	Date	Topic	Reading	Assignment
1	1.5-1.11	Course Introduction & Overview	DM Chap 1	Lab 1

Week	Date	Topic	Reading	Assignment
2	1.12-1.18	Data Exploration	DM Chap 2	Lab 2-1
3	1.19-1.25	Data Wrangling	DM Chap 3	Lab 2-2
4	1.26-2.1	Pattern Mining	DM Chap 6	Lab 3-1
5	2.2-2.8	Classification	DM Chap 8	Lab 3-2
6	2.9-2.15	Midterm Review	DM Chap 8	Exercise 1&2
7	2.16-2.22	Midterm		Lab 4-1
8	2.23-3.1	Map-Reduce	MMD Chap 2	Lab 4-2
9	3.2-3.4 3.7-3.11	Project Discussion Spring Break		Project Proposal
10	3.14-3.18	Clustering	MMD Chap 7	Project Discussion
11	3.21-3.25	Text Analysis	Lecture Notes	Project Discussion
12	3.28-4.1	Graph Analysis	MMD Chap 5	Project Discussion
13	3.4-4.8	Advanced Classification	DM Chap 9	Project Discussion
14	4.11-4.15	Final Exam Review and Project Presentations	Exercise 3&4	Project Discussion

Week	Date	Topic	Reading	Assignment
15	4.18-4.20	Final Exam		
16	4.25-4.27	Final Project Write-ups		

Disclaimer: This syllabus represents my current plans and objectives. As we go through the semester, those plans may need to change to enhance the class learning opportunity. Such changes, communicated clearly, are not unusual and should be expected.

Last update: 1/3/22