

CAP 4770: Introduction to Data Science

Instructor Information

Instructor	Email	Office Location & Hours
Dr. Zhe Jiang	zhe.jiang@ufl.edu	[TBD]
Teaching Assistants	Email	Office Location & Hours
Zelin Xu	zelin.xu@ufl.edu	[TBD]
Shiyu Xie	shiyu.xie@ufl.edu	[TBD]

General Information

Description

With the wide availability of huge amount data being collected in multiple sectors, our society is transitioning from a data poor era to a data rich era. An imminent need exists to turn such data into useful information and knowledge. Data science is a field across multiple disciplines such as computer science, statistics, and business analytics that studies how to automatically extract useful patterns and make predictions from a large amount of data. The objective of the course is to introduce fundamental concepts & techniques in data science. The course will primarily focus on the data science foundations, and the advantages and disadvantages of various methods for different data characteristics.

Topics to cover include data preprocessing, data exploration, relationship mining, classification/prediction, clustering, outlier detection, deep learning, etc. There will also be invited speakers on various data science applications. The specific topics are subject to adjustments through semester by the instructor.

Expectations and Goals

The course is suitable for undergraduate students with solid statistics knowledge and Python programming skills, who are strongly interested in learning data science. The goal is to learn how various data science techniques work (instead of using a Python library as a blackbox) and what are the advantages or disadvantages of different methods.

Pre-requisites

- The students have already taken Linear Algebra, Probabilities, Data Structure and Algorithm.
- The students should be fluent in Python programming.

Course Materials

Textbook

Required: Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar, 2nd Edition, ISBN 0-13-312890-3.

Optional Materials

- Data Mining: Concepts and Techniques, Book by Jiawei Han.

Topics to Be Covered

Tentative

Introduction to Data Science

Exploring Data

Sample Similarity Measures

Classification Techniques: I

Classification Techniques: II

Classification Techniques: III

Association Rule Mining

Clustering Techniques

Anomaly Detection

Deep Learning Techniques: I

Deep Learning Techniques: II

Deep Learning Techniques: III

Deep Learning Techniques: IV

(There may be invited guest lectures by other data science researchers in the field.)

Grading Policy

Note: All homework assignments are done in teams of two students. The students will spend the first week to decide a teammate.

Assignment (Tentative)

Percentage

Homework: Question-Answer

30%

Homework: Lab Programming

30%

Exam 1

15%

Exam 2

25%

Final Points

Grade

90 - 100

A

87 - 89

A-

84 - 86

B+

80 - 83

B

75 - 79	B-
72 - 74	C+
70 - 71	C
68 - 69	C-
65 - 67	D+
63 - 64	D
60 - 62	D-
<60	E

Attendance Policy, Class Expectations, and Make-Up Policy

Attendance is REQUIRED for this class. Excused absences must in compliance with university policies in the Graduate Catalog (<http://gradcatalog.ufl.edu/content.php?catoid=10&navoid=2020#attendance>) and require appropriate documentation.

Students are expected to have done the reading before class and actively participate during lectures and discussions (e.g., by asking questions or by volunteering their opinions). This is important to do well in this course.

In the event that you will miss any of the assignments or news presentation, please email the course instructor as soon as possible to determine the appropriate make-up action. You are strongly encouraged to make arrangements prior to any known absences. Non-valid excuses will result in a score of zero for the assignment or exam. All make-up actions must be completed before the tests or assignments are graded and returned to the class. No make-up actions will be taken after this time.

Other Issues

Students Requiring Accommodations

Students with disabilities requesting accommodations should first register with the Disability Resource Center (352-392-8565, <https://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.

Course Evaluation

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at <https://gatorevals.aa.ufl.edu/students/>. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via <https://ufl.bluera.com/ufl/>. Summaries of course evaluation results are available to students at <https://gatorevals.aa.ufl.edu/public-results>.

University Honesty Policy

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 honor 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 (<https://www.dso.ufl.edu/sccr/process/student-conduct-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 TAs in this class.