

INST 414 – Data Science Techniques

Section 0101

Meeting Days and Times: Online lecture, in-person meeting Thursday 11-12:15

Location: SQH1117

Instructor: Prof. Jennifer Golbeck, TA Jennifer Proctor

E-mail: jgolbeck@umd.edu

Office: 2117K Hornbake

Office Hours: TBA

Catalog Description

Pre-requisite: INST 314 Statistics for Information Science

Restrictions: Must be in BSIS program; Permission of BSIS program.

This course will explore approaches to extract insights from large-scale datasets. The course will cover the complete analytical funnel from data extraction and cleaning to data analysis and insights interpretation and visualization. The data analysis component will focus on techniques in both supervised and unsupervised learning to extract information from datasets. Topics will include clustering, classification, and regression techniques. Through homework assignments, a project, exams and in-class activities, students will practice working with these techniques and tools to extract relevant information from structured and unstructured data.

Extended Course Description

This course explores the application of data science techniques to unstructured, real-world datasets including social media and open data sources. The course will focus on techniques and approaches that allow the extraction of information relevant for experts and non-experts in a wide range of areas including smart cities, transportation or public safety.

This course will explore approaches to extract insights from large-scale datasets. The course will cover the complete analytical funnel from data extraction and cleaning to data analysis and insights interpretation and visualization. The data analysis component will focus on techniques in both supervised and unsupervised learning to extract information from datasets. Topics will include clustering, classification, and regression techniques. Through homework assignments, a project, exams and in-class activities, students will practice working with these techniques and tools to extract relevant information from structured and unstructured data.

Student Learning Outcomes:

Upon completing this course, students will be able to:

- Collect and clean large-scale datasets
- Articulate the math behind supervised and unsupervised techniques
- Execute supervised and unsupervised machine learning techniques
- Select and evaluate various types of machine learning techniques
- Interpret the results coming out of the models
- Critically evaluate the accuracy of different algorithms and the appropriateness of a given approach

Textbooks and Readings

There will be some free online materials. In addition, you need this textbook:

[Cady, Field. *The Data Science Handbook*. John Wiley & Sons, 2017.](#)

<https://amzn.to/2N7GMG9>

Required Technology

- **Laptop** – We will do live exercises in class. Please bring your laptop to class. If you do not have one, contact the professor before the first class.
- **Plain Text Editor** - Sublime Text, TextWrangler, BB Edit are all good, free options

Course Activities

Homework Assignments

We will have regular homework assignments. You may work with your classmates to figure out the underlying concepts but are expected to work *individually* to answer the specific problems

that are assigned. Timely submission of the completed assignments is essential. The due date of each assignment will be stated clearly in the assignment description. If an assignment due date is a religious holiday for you, please let the instructor know at least one week in advance, so an alternate due date can be set.

Group Project

For your group project you will form teams of 2-4 people and prepare a data-related analytic project. This involves identifying a question, finding or developing a dataset, creating appropriate measures, conducting analyses, and preparing an appropriate information product based on the results. The purpose of the project is to have you go through the steps and processes completing a high-impact data analytics project.

The project will be graded on your ability to articulate an appropriate question, prepare the data, identify and perform reasonable methodology and study design, justify the appropriateness of certain machine learning approaches, articulate and conduct evaluations, analyze and interpret the results and create appropriate visualizations.

Exams

In this course, the assignments provide you with opportunity to experiment with and learn about the ideas, concepts, and techniques associated with data science. The midterm exam will complement this by providing you with feedback about how well you have learned them and whether you have successfully developed the ability to apply those concepts and techniques in different settings.

Quizzes

You are required to watch the video lectures before the in-person class meeting. The quizzes are designed to be easy tests of your comprehension of the lecture material. There will be no extensions or re-dos on the quizzes.

Grading

Your final grade for the course is computed as weighted averages of your scores on the individual elements, converted to a letter grade:

A+ 97-100* B+ 87-89.99 C+ 77-79.99 D+ 67-69.99

A 93-96.99 B 83-86.99 C 73-76.99 D 63-66.99 F 0-59.99

A- 90-92.99 B- 80-82.99 C- 70-72.99 D- 60-62.99

* Note: To receive an A+ you must have demonstrated significant contributions to the class in addition to achieving this numeric grade.

COURSE SCHEDULE

This schedule is for planning purposes and may change.

See ELMS/Canvas for current information and deadlines.

University Policies

For university course policies, review go.umd.edu/ug-policy