UNIVERSITY OF MARYLAND

Course Information

Course Title: Data Science Techniques
Course Number: INST 414
Term: *Fall / 2024*
Location: ATL 1113

Instructor: Zubin Jelveh (zjelveh@umd.edu)
TAs:
- Marilyn Harbert (mharbert@umd.edu)
- Phuong (Kem) Nguyen-Le (nlpa@umd.edu)

UTA:
- Josephine Whittington (jwhitt@terpmail.umd.edu)

Course Description

The purpose of this course is to provide you with a strong foundation in the key techniques used in the field of data science. As the amount of data available to public- and private-sector organizations continues to grow, it has become critical to have data scientists who can leverage this data to generate insights and enable data-driven decision-making, while ensuring that these decisions are based on fair and ethical principles. This course will emphasize not only the technical aspects of data science but also the crucial considerations of fairness and equity in applying these techniques to real-world problems.

In this course, you will learn foundational data science techniques including basic probability and statistics, supervised machine learning algorithms (such as Naive Bayes, linear and logistic regression, decision trees, and random forests), feature engineering, working with Python libraries, principles of model selection and evaluation (including cross-validation), and considerations around fairness and bias in machine learning applications. We'll also touch on text mining techniques and explore the emerging field of generative AI, providing you with an introduction to these cutting-edge areas of data science. Through homework assignments, projects, exams, and in-class activities, you will gain practical experience applying these techniques to real-world datasets.

By the end of the course, you will have built intuition for working with tabular data and be equipped with a toolkit to start building predictive models and extracting impactful insights from data. The skills you learn here will enable you to contribute meaningfully to data science projects and prepare you for more advanced courses in machine learning, artificial intelligence, and data analytics.

Note: This course primarily focuses on foundational supervised learning techniques.

Note: Knowledge of Python is assumed.

Learning Outcomes

After successfully completing this course you will be able to:

- Review of basic probability and related concepts like uncertainty
- Basics of standard machine learning algorithms (linear regression, logistic regression, decision trees, and random forest)
- Model selection (e.g. cross-validation)
- Feature engineering
- Fairness considerations

Required Resources

This course does not stick to particular textbooks. Readings and course materials will be provided on the elms course site. No purchases required.

Laptop – We will do live exercises in class in python. Please bring your laptop to class. If you do not have one, contact the professor before the end of the first week.

Letter Grade Cutoffs

In the table below, g_i is the grade for i-th student

A+	$97 \leq g_i \leq 100^*$	B+	$87 \leq g_i < 90$	C+	$77 \leq g_i < 80$	D+	$67 \leq g_i < 70$	F	$g_i < 60$
A	$93 \leq g_i < 97$	B	$83 \leq g_i < 87$	C	$73 \leq g_i < 77$	D	$63 \leq g_i < 67$		
A-	$90 \leq g_i < 93$	B-	$80 \leq g_i < 83$	C-	$70 \leq g_i < 73$	D-	$60 \leq g_i < 63$		

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

Communications:

ELMS - Official course site for materials, assignments, announcements, gradings, etc. Make sure that your email & announcement notifications (including changes in assignments and/or due dates) are enabled in ELMS so you do not miss any messages.

Emails - Administrative requests, quick clarifications, etc. Please prefix the subject line with [INST414]. If you have not received a reply within 2 days, please email again.

Office Hours - Complex technical questions

Grading Structure

Assignment	Overall Percentage %
4 Homework Assignments	**30%**
10 weekly quizzes	**20%**
Midterm Exam	**25%**
Final Project or Exam	**25%**
Total	**100%**

Course Outline

Week #	Topic
1- (8/29)	Data Science Background
2 - (9/5)	Probability
3 - (9/12)	Conditional Probability / Problem Set #1 Out
4 - (9/19)	Bayes Theorem / Making Predictions
5 - (9/26)	Performance Metrics / Naive Bayes / Problem Set #1 Due (9/26)
6 - (10/3)	Linear and Logistic Regression / Problem Set #2 Out
7 - (10/10)	Overfitting / Decision Trees
8 - (10/17)	Random Forest / Gradient Boosting / Problem Set #2 Due (10/17)
9 - (10/24)	Midterm (No Class)
10 - (10/31)	Fairness / Problem Set #3 Out
11 - (11/7)	Sources of Bias, Part I
12 - (11/14)	Sources of Bias, Part II / Problem Set #3 Due (11/14)
13 - (11/21)	Text-Mining / Problem Set #4 Out
14 - (11/28)	No Class
15 - (12/5)	Text-Mining/Gen-AI / Problem Set #4 Due (12/3)

Note: This is a tentative schedule, and subject to change as necessary – monitor the course ELMS page for current deadlines.

Resources & Accommodations

Accessibility and Disability Services

Students with disabilities should inform me of their needs at the beginning of the semester. Please also contact UMD's Accessibility and Disability Service (https://counseling.umd.edu/ads/). ADS will make arrangements with the student and me to determine and implement appropriate academic accommodations. Inclusion is one of the iSchool's core values, and we have attempted to make all materials and assignments accessible to people with varying abilities. However, if there is something else I can do to make the class more accessible please schedule a time to come talk to me. This will benefit not only yourself but also future students!

Getting Help

Taking personal responsibility for your own learning means acknowledging when your performance does not match your goals and doing something about it. I hope you will come talk to me so that I can help you find the right approach to success in this course, and I encourage you to visit **UMD's Student Academic Support Services website** to learn more about the wide range of campus resources available to you.

In particular, everyone can use some help sharpening their communication skills (and improving their grade) by visiting **UMD's Writing Center** and schedule an appointment with the campus Writing Center.

You should also know there are a wide range of resources to support you with whatever you might need (**UMD's Student Resources and Services website** may help). If you feel it would be helpful to have someone to talk to, visit **UMD's Counseling Center** or **one of the many other mental health resources on campus**.

UMD Policies

It is our shared responsibility to know and abide by the University of Maryland's policies that relate to all courses, which include topics like:

- Academic integrity
- Student and instructor conduct
- Accessibility and accommodations
- Attendance and excused absences
- Grades and appeals
- Copyright and intellectual property

Please visit www.ugst.umd.edu/courserelatedpolicies.html for the Office of Undergraduate Studies' full list of campus-wide policies and follow up with me if you have questions.