INST 414 Data Science Techniques
Syllabus

Class Meets: Monday 9:00 - 11:45 am. BLG4 3321

Instructor: Wei Ai (aiwei@umd.edu)
Office Hours: Monday 11:45 am - 12:30 pm.

TA: TBD
Office Hours: TBA

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.

Learning Outcomes
- 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
- Explain the results coming out of the models
- Critically evaluate the accuracy of different algorithms and the appropriateness of a given approach

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. Please bring your laptop to class. If you do not have one, contact the professor before the end of the first week.

Grades

- Class Participation: 10%
- Four Homework Assignments: 10% x 4 = 40%
- Kaggle Competition: 20%
- Final Exam (Take Home): 30%

Letter Grade Cutoffs

A+	97-100*	B+	87-89.99	C+	77-79.99	D+	67-69.99	F	0-59.99
A	93-96.99	B	83-86.99	C	73-76.99	D	63-66.99		
A-	90-92.99	B-	80-82.99	C-	70-72.99	D-	60-62.99		

Communications

ELMS - Official course site for materials, assignments, announcements, gradings, etc.

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.

Piazza - General questions, discussions, seeking help, etc.

Office Hours - Complex technical questions.

Tentative Schedule

Week 1 (1/27): Course Overview and Introduction to Data Mining

Week 2 (2/3): Math Basics and Probability Review

Week 3 (2/10): Mining Itemsets

Week 4 (2/17): Mining Matrices

Week 5 (2/24): Mining Sequences

Week 6 (3/2): Text Analysis and NLP

Week 7 (3/9): Information Retrieval

Week 8 (3/16): Spring Break (No Class)

Week 9 (3/23): Machine Learning Overview

Week 10 (3/30): Supervised Machine Learning

Week 11 (4/6): Supervised Machine Learning (cont.)

Week 12 (4/13): Unsupervised Machine Learning

Week 13 (4/20): Unsupervised Machine Learning (cont.)

Week 14 (4/27): Deep Learning & Reinforcement Learning

Week 15 (5/4): Brief Introduction to Causal Inference

Week 16 (5/11): Data Science Applications

Campus 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 go.umd.edu/ug-policy for the Office of Undergraduate Studies' full list of campus-wide policies and follow up with me if you have questions.

Accessibility and Learning Support

Students with disabilities should inform me of their needs at the beginning of the semester. Please also contact the Accessibility and Disability Support Office (http://www.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!

Get Some 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 tutoring.umd.edu to learn more about the wide range of campus resources available to you. In particular, everyone can use some help sharpen their communication skills (and improving their grade) by visiting ter.ps/writing 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 (see go.umd.edu/assistance), and if you just need someone to talk to, visit counseling.umd.edu or one of the many other resources on campus.

Most services free because you have already paid for it, and everyone needs help… all you have to do is ask for it.