

SYLLABUS

Course description

Data mining is a practical discipline that combines computer science, statistics, math, and optimization techniques to analyze data and gather valuable knowledge from it. This course is designed to study fundamental data mining concepts and provide hands-on experience with several methods. The students will develop a broad and deep background in data mining and crucial skills to solve practical data science challenges. Problems will involve the analysis of real databases coming from various fields, such as food science, astronomy, human resources, social sciences, and banking, among others. Students are expected to have previous knowledge in the Python programming language.

Course Outcomes

By the end of this course, you will be able to identify fundamental principles, techniques, and data mining applications. You will also apply computational and statistical methods to visualize, explore, and prepare data for posterior analysis. In addition, you will be able to translate real-life problems and frame them under supervised or unsupervised paradigms. Later you can apply different classification, prediction, or clustering approaches, where you will be able to evaluate them empirically and choose the best one with clear foundations.

Grading

The grading scale will break down as follows:

A =	93–100%	C =	73–76%
A- =	90–92%	C- =	70–72%
B+ =	87–89%	D+ =	67–69%
B =	83–86%	D =	63–66%
B- =	80–82%	D- =	60–62%
C+ =	77–79%	F =	Below 60%

There will be one mid-term exam, take-home assignments (individual), and one final project

(proposal, presentation, and report). The grading breakdown is as follows:

Assignment	Weight
Class Participation	5%
Midterm Exam	25%
Homework Assignments	30%
Final Project Presentation	10%
Final Project Paper	30%

Evaluation Activities

Take-home assignments will help students gain skills and feel more confident about the topics reinforced in the assignment. The final project will be an open ended capstone project, intended to cover a broader spectrum of contents, implementing a data mining solution for a real problem with real data. We expect a thorough analysis and creative solutions to the problem. The final project can be done individually or in teams. The teams should be determined before midterms, if applicable. All details will be provided with the announcement of every course activity.

Class attendance and participation

We base the learning process of this class on in-class discussion and participation. Attendance is mandatory and preparation of the course material is highly recommended. That includes coming/connecting to the class on time. Classes will combine theory and practice with hands-on activities.

Schedule and Materials:

The course material is approximate and subject to change!

Week 1: 1/11 Introduction to Data Mining	<ul style="list-style-type: none"> • Lecture 1 ➞ (https://northeastern-my.sharepoint.com/:p:/g/personal/k_ni_northeastern_edu/ER7Mtlxcb4VlpDUYh1SDIe=jhZucB) • Assignment 1 ➞ (https://classroom.github.com/a/gU04OEFK) • Assignment 1: Files (https://northeastern.instructure.com/courses/140676/files/fo) • Reading ➞ (http://infolab.stanford.edu/~ullman/mmds/ch1n.pdf)
--	---

Week 2: 1/18 Data Analysis And Summarization	
Week 3: 1/25 Data Preprocessing and Engineering	
Week 4: 2/1 Frequent Item Sets and Parameter Estimation	
Week 5: 2/8 Unsupervised Machine Learning	
Week 6: 2/15 Supervised Machine Learning	
Week 7: 2/22 Deep Neural Networks	<ul style="list-style-type: none">• Project: Last Day to Change Teams
Week 8: 3/1 Midterm Exam	
Week 9: 3/8 Spring Break	<ul style="list-style-type: none">• No Class
Week 10: 3/15 Applying	<ul style="list-style-type: none">• Project Proposals Due

Machine Learning to Big Data	
Week 11: 3/22 Other Forms of Machine Learning	
Week 12: 3/29 Recommendation Sciences	
Week 13: 4/5 Special Topics -- <ul style="list-style-type: none">• Mining Relational (Graphs)• Infinite (Streams) Data	
Week 14: 4/12 Industry Day	<ul style="list-style-type: none">• Industry Topics• Project Discussions• Parting Lecture Advice
Week 15: 4/19 Project Presentations	
Week 16 4/26 Project Writeup Submissions Due	

ics

Recommended Books

courses/140676/pages/course- (https://northeastern.instructure.com/courses/140676/pages/recommended-
books)